

**Structured factor analysis: A data matrix-based alternative approach to
structural equation modeling**

Gyeongcheol Cho and Heungsun Hwang
McGill University, Montreal, Canada

October 02, 2022

Author Note

- The authors are grateful to Claes Fornell for providing the ACSI data. They also thank the two reviewers for their constructive comments that helped improve the quality of the manuscript.
- Requests for reprints should be sent to: Gyeongcheol Cho, Department of Psychology, McGill University, 2001 McGill College Avenue, Montreal, QC H3A 1G1, Canada.
Email: gyeongcheol.cho@mail.mcgill.ca.
- The Version of Record of this manuscript has been published and is available in *Structural Equation Modeling: A Multidisciplinary Journal*, November 17, 2022, <https://www.tandfonline.com/doi/full/10.1080/10705511.2022.2126360>.

**Structured factor analysis: A data matrix-based alternative approach to
structural equation modeling**

Abstract

Jöreskog's covariance-based approach (JCA; Jöreskog, 1978) has been considered a standard method for structural equation modeling. However, JCA is prone to the occurrence of improper solutions and cannot make probabilistic inferences about the true factor scores. To address the enduring issues of JCA, we propose a data matrix-based alternative, termed structured factor analysis (SFA). Given a data matrix of indicators, SFA begins by estimating both measurement model parameters and factor scores by minimizing a single cost function via an alternating least squares algorithm, which mathematically guarantees convergence to proper solutions. It then employs the factor score estimates to estimate structural model parameters. Once all parameters are estimated, SFA further estimates the probability distribution of the factor scores that can generate the data matrix of indicators, which can be used for probabilistic inferences about the true factor scores. We investigate SFA's performance and empirical utility through simulated and real data analyses.

Keywords: Structured factor analysis, structural equation modeling, factor score, candidate factor score distribution, measurement.

1. Introduction

Structural equation modeling (SEM) can be used for examining the relationships between latent variables and indicators. Although there are various statistical methods for estimating structural equation models, including factor score regression (Croon, 2002; Skrondal & Laake, 2001), model-implied instrumental variable methods (Bollen, 1996, 2019), consistent partial least squares (Dijkstra, 2011, 2013), and generalized structured component analysis with measurement errors incorporated (Hwang et al., 2017), Jöreskog's covariance-based approach (Jöreskog, 1978), denoted by JCA herein, has been considered a de facto standard method for many reasons including its versatility in model specification and evaluation and statistically desirable properties in model estimation (e.g., Bollen, 2019).

As its name implies, JCA carries out all main SEM steps, including model identification, estimation, and evaluation, based on the covariance matrix of indicators. Specifically, in this approach, the covariance matrix of indicators is reformulated as a function of parameters in the model, called the implied covariance matrix, and the model's identification is ensured by checking whether the model parameters uniquely determine the implied covariance matrix (Bollen, 1989, p. 89). If the model is identified, JCA aims to estimate the model parameters by minimizing a cost function that represents the discrepancy between the sample and implied covariance matrices. JCA typically utilizes the maximum likelihood (ML) estimator, assuming the multivariate normality of indicators. The ML estimator is a full-information estimator that estimates model parameters simultaneously, using all information in model equations (Fomby et al., 2012, Chapter 22). It is known to be asymptotically unbiased and efficient, thereby being conceived of as the most optimal estimator in SEM when every endogenous variable in the model is continuous (e.g., Bollen, 2019). Once the parameters are estimated, JCA provides

overall model fit measures that assess the magnitude of the discrepancy between the sample and implied covariance matrices, such as the χ^2 statistic, root mean square error of approximation (Steiger, 2016), comparative fit index (Bentler, 1990), and standard root mean square residuals (Bentler, 1995), for evaluating the model and/or comparing competing models.

Despite its merits, JCA has two long-standing limitations. First, JCA occasionally suffers from improper solutions, such as negative error variance estimates or correlation estimates over ± 1 (e.g., Bentler & Chou, 1987; Chen, Bollen, Paxton, Curran, & Kirby, 2001). Given an improper solution, researchers can hardly consider the other parameter estimates reliable and conduct further model evaluation or comparison with confidence (e.g., McDonald, 2004; Newsom, 2014). Second, JCA does not provide statistical tools for making a probabilistic inference about the true latent variable or factor scores. In practice, researchers may be interested not only in the relationships between latent variables but also in the probability for a score interval to contain an individual's true factor score. JCA does not provide information on this probability, focusing solely on testing the relationships between latent variables.

The common source of these limitations is JCA's reliance on the implied covariance matrix of indicators. The implied covariance matrix is entirely defined by model parameters, indicating that individual factor scores are not treated as parameters to be estimated and are completely ignored in JCA's cost function. Consequently, it is impossible to obtain probabilistic information on the factor scores from the cost function. Moreover, as JCA's algorithm disregards the factor scores while updating model parameters at each iteration, it can update the model parameters to some values that won't be admissible if the factor scores are also to be updated prior to the other model parameters. For instance, although it is theoretically impossible that the variances of latent variables (or any other variables) are negative, JCA's algorithm may update

these variances to negative ones if such updating can reduce the value of its cost function. If JCA's cost function were also defined based on the factor scores and its algorithm was to update the factor scores before updating their variances, it would be essentially impossible for the algorithm to produce such improper solutions. De Jonckere and Rosseel (2022) recently suggested imposing additional constraints that force JCA's algorithm to avoid improper solutions. However, their constrained procedure still tends to yield improper solutions at times, as shown in their simulation study.

In this paper, we propose a new data matrix-based approach to SEM, named *structured factor analysis* (SFA), whose cost function reflects the discrepancy between the data matrices of indicators and their predicted values. SFA carries out two SEM stages sequentially. In the first stage, SFA begins by specifying the process of generating indicators' data as the measurement model. It then estimates the measurement model parameters (i.e., factor loadings and factor variances and covariances) and factor scores concurrently in such a way that they jointly minimize a single cost function—an average residual variance or (in-sample) prediction error for the indicators' scores. The factor scores estimated from SFA are called *candidate factor scores* in the sense that they represent a set of factor scores that can be considered potential candidates for the true factor scores given a data matrix of indicators. In the second stage, SFA specifies the process of generating the true latent variable scores as the structural model and estimates its model parameters (i.e., path coefficients and error variances and covariances) based on the candidate factor scores obtained from the first stage without the need of modifying the candidate factor scores or their covariance matrix.

SFA can prevent the occurrence of improper solutions as it simultaneously estimates the measurement model parameters and factor scores in the first stage. Moreover, it allows for

inferring the true factor scores probabilistically. Once all model parameters are estimated, SFA can additionally estimate the probability distribution of the candidate factor scores based on the cost function used in the first stage. From this probability distribution, we can obtain an individual's 95% candidate factor score interval for each latent variable, which contains 95% of the individual's factor scores that can generate a given data matrix of indicators. We can also utilize the distribution for statistically testing the difference in two individuals' factor scores or calculating the probability of one's factor score being greater or less than the others. Thus, SFA can be used as an alternative when JCA suffers from improper solutions and/or researchers are interested in making probabilistic inferences about individuals' true factor scores.

The remainder of the paper is organized as follows. In Section 2, we describe the two stages of SFA. In Section 3, we discuss the property and estimation of the candidate factor score distribution. Note that as one reviewer has suggested, we provide succinct yet essential information about both sections while relegating all technical details to Supplementary Material. In Section 4, we conduct Monte Carlo simulation studies to investigate how SFA performs as compared to JCA and whether SFA's candidate factor score distribution behaves as expected. In Section 5, we illustrate an application of SFA to a real dataset. In the final section, we discuss the implications of the proposed method and its potential extensions.

2. The Proposed Method

SFA carries out two stages sequentially, each of which involves its own model specification, identification, estimation, and evaluation. In the first stage, researchers are to specify the measurement model that represents the data-generating process of indicators under the assumption that the true latent variable scores are the underlying causes of the indicators' scores.

SFA estimates the parameters of the specified measurement model as well as factor scores and allows for statistical tests of the goodness-of-fit of the measurement model. If the measurement model with the factor score estimates may be acceptable, SFA can move on to the second stage. In this stage, researchers are to specify the structural model that represents the score-generating process of latent variables. SFA estimates the parameters of the structural model using the factor scores estimated from the first stage and evaluates the goodness-of-fit of the structural model.

The theoretical derivation of both stages is based on the random matrix theory, which involves a set of matrix-valued random variables. As stated above, we here present the most essential information on the stages, which is all based on realized counterparts of random matrices. For simplicity, we further assume that every realized matrix is representative of the population, which means that the sample mean vector and covariance matrix of the realized matrix are equivalent to their population counterparts, although this assumption can be relaxed. A fully detailed theoretical description of the stages, including several theorems and their proofs, is provided in Appendix S1.

Stage 1: Measurement model for the data-generating process of indicators

Model specification

Let $\mathbf{Z}$ denote an N by J matrix consisting of N individuals' scores on J indicators, whose mean vector is a zero vector and covariance matrix is denoted by $\mathbf{\Sigma}$. Let $\mathbf{H}_{true}$ denote an N by P matrix of the true latent variable scores for N individuals, whose mean vector is a zero vector and covariance matrix is denoted by $\mathbf{\Phi}$. Each latent variable is equivalent to a common factor that causes their respective indicators to covary. Let $\mathbf{E}_{true}$ denote an N by J matrix of the true unique factor scores for N individuals, whose mean vector is a zero vector and covariance matrix is denoted by $\mathbf{\Theta}$. Let $\mathbf{\Lambda}$ denote a P by J matrix of loadings that quantify the causal effects of P

latent variables on J indicators. For simplicity, both $\mathbf{Z}$ and $\mathbf{H}_{true}$ are assumed to be standardized.

In Stage 1, SFA formulates the measurement model to describe how $\mathbf{Z}$ is generated from $\mathbf{H}_{true}$.

The measurement model is given as

$$\mathbf{Z} = \mathbf{H}_{true}\mathbf{\Lambda} + \mathbf{E}_{true}. \quad (1)$$

Based on prior theory, researchers are to specify which elements of $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Theta}$ in the measurement model are non-zero parameters and ensure whether the specified measurement model is identified. The identification rules for the measurement model are equivalent to those for the confirmatory factor analysis model in JCA (e.g., refer to Bollen 1989, pp. 238–251).

Parameter estimation

The measurement model (1) can be re-expressed as

$$\mathbf{Z} = \mathbf{F}_{true}\mathbf{L}, \quad (2)$$

where $\mathbf{F}_{true} = [\mathbf{H}_{true}, \mathbf{E}_{true}]$, $\mathbf{L} = [\mathbf{\Lambda}; \mathbf{I}_J]$, and $\mathbf{I}_J$ is the identity matrix of order J . We call $\mathbf{F}_{true}$ a matrix of the *true factor scores* including both common and unique factor scores for N

individuals. Let $\mathbf{\Delta} \equiv \begin{bmatrix} \mathbf{\Phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{bmatrix}$ denote the covariance matrix of $\mathbf{F}_{true}$. The measurement model parameters include $\mathbf{L}$ and $\mathbf{\Delta}$. SFA seeks to simultaneously estimate $\mathbf{L}$, $\mathbf{\Delta}$, and $\mathbf{F}_{true}$ given $\mathbf{Z}$.

However, even if we knew the values of $\mathbf{L}$ and $\mathbf{\Delta}$, it would be impossible to obtain the precise value of $\mathbf{F}_{true}$ from $\mathbf{Z}$ because the number of scores in $\mathbf{F}_{true}$ to be estimated (i.e., NT) is greater than that of observed scores in $\mathbf{Z}$ (i.e., NJ), where $T = P + J$ (e.g., Mulaik, 2009; Steiger, 1979). This is the factor score indeterminacy problem (de Leeuw, 2017).

Accordingly, instead of aiming to obtain an unbiased estimate of $\mathbf{F}_{true}$, SFA contemplates a matrix of the candidate factor scores, denoted by $\mathbf{F}$, which satisfies $\mathbf{Z} = \mathbf{FL}$, $mean(\mathbf{F}) = \mathbf{0}$, and $cov(\mathbf{F}) = \mathbf{\Delta}$, and can thus be considered a potential candidate for $\mathbf{F}_{true}$ given $\mathbf{Z}$, where $mean()$ and $cov()$ transform an input matrix into its sample mean vector and covariance matrix, respectively

(refer to Appendix S1 for more details). SFA then aims to obtain unbiased estimates of $\mathbf{L}$ and $\mathbf{\Lambda}$ as well as an estimate of $\mathbf{F}$. Specifically, let $\mathbf{Z}_{std}$ denote the standardized counterpart of $\mathbf{Z}$, whose covariance matrix is denoted by $\mathbf{S}$. Let $\hat{\mathbf{Z}}$ denote a matrix of the predicted values of $\mathbf{Z}_{std}$ based on the estimated model. To estimate $\mathbf{L}$, $\mathbf{\Lambda}$, and $\mathbf{F}$, SFA seeks to minimize the following cost function.

$$\begin{aligned}\rho &= (JN)^{-1} SS(\mathbf{Z}_{std} - \hat{\mathbf{Z}}) \\ &= (JN)^{-1} SS(\mathbf{Z}_{std} - \mathbf{FL}),\end{aligned}\tag{3}$$

subject to $mean(\mathbf{F}) = \mathbf{0}$ and $cov(\mathbf{F}) = \mathbf{\Lambda}$, where $SS(\mathbf{X}) = tr(\mathbf{X}'\mathbf{X})$ for any matrix $\mathbf{X}$. The value of the cost function (3) can be interpreted as the average residual variance or (in-sample) prediction error for the standardized indicator scores. This indicates that SFA aims to simultaneously estimate the measurement model parameters and a matrix of the candidate factor scores in such a way that they maximize explanatory power for the standardized indicator scores.

There is no closed-form solution for the constrained minimization problem (3). Thus, we develop an alternating least squares (ALS) algorithm, which divides the model parameters into several groups and updates each group alternately with the remaining groups fixed. A detailed description of this algorithm is provided in Appendix S4. After the model parameters are estimated, their standard errors or 95% confidence intervals are calculated for testing their statistical significance. As SFA does not assume any distributional assumption on indicators, it employs a resampling technique, such as the bootstrap method (Efron, 1979, 1982), to obtain these statistics without recourse to a distributional assumption.

If the ALS algorithm minimizes (3) under the identified measurement model, it provides unbiased estimates of the measurement model parameters, as proved in Appendix S5. The convergence of the ALS algorithm has been mathematically proven (de Leeuw et al., 1976).

Moreover, the proposed algorithm does not result in improper solutions as its cost function (3) is built on individual factor scores rather than their covariance matrix. Obviously, a set of individual factor scores cannot have negative variances, a negative-definite covariance matrix, or correlations with $\mathbf{Z}_{std}$ greater than one in absolute value.

Despite its desirable properties, the algorithm may be computationally more costly than JCA's algorithm when the sample size is large as it needs to estimate N individuals' factor scores in addition to the model parameters. Thus, we propose a supplementary procedure to alleviate the algorithm's potential computational burden in Appendix S6. This procedure indicates that even if the cost function (3) is defined based on a data matrix of indicators, SFA only needs the sample covariance matrix of the indicators for estimating the model parameters. Thus, even when researchers only have the sample covariance matrix in hand, they can still apply SFA if they are only interested in estimating the model parameters.

Model evaluation

SFA provides an overall goodness-of-fit index, termed the in-sample prediction error for observed variables (IPE_O), for evaluating the measurement model using the estimates of the model parameters and factor scores. The IPE_O is defined as

$$\text{IPE}_O = SS(\mathbf{Z}_{std} - \hat{\mathbf{F}}\hat{\mathbf{L}})/SS(\mathbf{Z}_{std}), \quad (4)$$

where $\hat{\mathbf{F}}$ and $\hat{\mathbf{L}}$ are the estimated candidate factor score and loading matrices, respectively. This index is equivalent to the value of (3) computed based on $\hat{\mathbf{F}}$ and $\hat{\mathbf{L}}$. The value of $1 - \text{IPE}_O$ can also be interpreted as the average R^2 for the indicators. The IPE_O value is zero if and only if $\hat{\mathbf{L}} = \mathbf{L}$ and $\hat{\mathbf{F}}$ satisfies $\text{mean}(\hat{\mathbf{F}}) = \mathbf{0}$, $\mathbf{Z}_{std} = \hat{\mathbf{F}}\hat{\mathbf{L}}$, and $\text{cov}(\hat{\mathbf{F}}) = \mathbf{\Lambda}$, given that the measurement model is identified (see Appendix S5). As $\hat{\mathbf{F}}$ obtained from SFA always satisfies $\text{mean}(\hat{\mathbf{F}}) = \mathbf{0}$, a positive value of IPE_O indicates that $\hat{\mathbf{L}} \neq \mathbf{L}$, $\mathbf{Z}_{std} \neq \hat{\mathbf{F}}\hat{\mathbf{L}}$, or $\text{cov}(\hat{\mathbf{F}}) \neq \mathbf{\Lambda}$.

SFA can conduct a statistical test of the null hypothesis that $\hat{\mathbf{L}} = \mathbf{L}$, $\mathbf{Z}_{std} = \hat{\mathbf{F}}\hat{\mathbf{L}}$, and $cov(\hat{\mathbf{F}}) = \mathbf{\Lambda}$ by using the Bollen-Stine (B-S) bootstrap method (Bollen & Stine, 1993). Let $\hat{\mathbf{\Sigma}}$ denote the estimated implied covariance matrix of the indicators. Let $\mathbf{Z}_{imp}$ denote a modified data matrix generated under the null hypothesis, which can be obtained by $\mathbf{Z}_{imp} = \mathbf{Z}_{std}\mathbf{S}^{-1/2}\hat{\mathbf{\Sigma}}^{1/2}$. The bootstrap method is applied to $\mathbf{Z}_{imp}$ to estimate the sampling distribution of the IPE_O. If the value of IPE_O is greater than a critical or cut-off value, e.g., the $(1 - \alpha)$ th percentile of the estimated sampling distribution, we may reject the null hypothesis. In addition, SFA can offer other traditional goodness-of-fit indexes, such as GFI (Jöreskog & Sorbom, 1986) and SRMR (Bentler, 1995), which are computed based on the sample and implied covariance matrices of the indicators. Lastly, the bootstrap standard errors or 95% confidence intervals of the model parameter estimates are used to test the statistical significance of the estimates.

2.1. Stage 2: Structural model for the score-generating process of latent variables

Model specification

Let $\mathbf{H}_{X,true}$ and $\mathbf{H}_{Y,true}$ denote matrices of the true scores of exogenous and endogenous latent variables, respectively. Let $\mathbf{B}_X$ and $\mathbf{B}_Y$ denote matrices of path coefficients that quantify the causal effects of exogenous latent variables on endogenous latent variables and those between endogenous latent variables, respectively. Let $\mathbf{Q}_{true}$ denote a matrix of the true scores of structural errors for $\mathbf{H}_{Y,true}$, whose mean vector is a zero vector and covariance matrix is denoted by $\mathbf{\Psi}$. The structural model of SFA is defined as

$$\mathbf{H}_{Y,true} = \mathbf{H}_{X,true}\mathbf{B}_X + \mathbf{H}_{Y,true}\mathbf{B}_Y + \mathbf{Q}_{true}. \quad (5)$$

Similar to Stage 1, based on prior theory, researchers are to predetermine which elements of $\mathbf{B}_X$, $\mathbf{B}_Y$, and $\mathbf{\Psi}$ in (5) are non-zero parameters and check if the specified structural model is

identified. The rules for the identification are the same as those used for the path analysis model in JCA, which can be found in Bollen (1989, pp. 88–104) or Dijkstra (2017).

Parameter estimation

Let $\hat{\mathbf{H}}_X$ and $\hat{\mathbf{H}}_Y$ denote matrices of the candidate factor scores for the exogenous and endogenous latent variables estimated from Stage 1. In Stage 2, SFA estimates the structural model parameters ($\mathbf{B}_X$, $\mathbf{B}_Y$, and Ψ) while treating $[\hat{\mathbf{H}}_X, \hat{\mathbf{H}}_Y]$ as the input data. It utilizes a limited-information estimator, which successively applies ordinary least squares (OLS) or two-stage least squares (2SLS) to each equation for an endogenous latent variable (Lance et al., 1988). The proposed estimator draws on 2SLS if endogeneity occurs in the equation, and on OLS otherwise. Although this estimator can be less efficient than a full-information estimator, such as feasible generalized least squares (FGLS) or three-stage least squares (3SLS), it can be more robust to model misspecification (Wooldridge, 2010, pp. 252–254). A detailed description of the estimator is provided in Appendix S7. Once all the structural model parameters are estimated, their standard errors or confidence intervals are estimated based on a set of the latent variable covariance matrices estimated from the bootstrap samples.

Model evaluation

SFA provides a goodness-of-fit index, termed the in-sample prediction error for latent variables (IPE_L), for evaluating the structural model. The IPE_L is defined as

$$\text{IPE}_L = SS(\hat{\mathbf{H}}_Y - (\hat{\mathbf{H}}_X \hat{\mathbf{B}}_X + \hat{\mathbf{H}}_Y \hat{\mathbf{B}}_Y)) / SS(\hat{\mathbf{H}}_Y). \quad (6)$$

This index represents the average residual variance for all endogenous latent variables unexplained by the fitted structural model. The value of $1 - \text{IPE}_L$ is equivalent to the average R^2 for the endogenous latent variables. SFA can also provide GFI and SRMR for the structural model, which are calculated based on the discrepancy between the covariance matrix of the

latent variables estimated from Stage 1 and the implied covariance matrix of the latent variables estimated from Stage 2.

3. Candidate Factor Score Distribution

As discussed in the previous section, SFA obtains an estimate of a matrix of the candidate factor scores $\mathbf{F}$ and uses this estimate, denoted by $\hat{\mathbf{F}}$, to estimate the parameters of the measurement and structural models. However, SFA does not recommend using $\hat{\mathbf{F}}$ as a point estimate of $\mathbf{F}_{true}$ as there exist an infinite number of $\mathbf{F}$ s owing to the factor score indeterminacy problem, so that there is no possibility that a single estimate of $\mathbf{F}$ is equivalent to $\mathbf{F}_{true}$. Instead, SFA derives the probability distribution of all possible $\mathbf{F}$ s and uses this distribution to infer $\mathbf{F}_{true}$ *a posteriori* given $\mathbf{Z}$. We term this distribution the *candidate factor score distribution*.

As in Section 2, we only provide basic information about the candidate factor score distribution. A complete theoretical description of the candidate factor score distribution is available in Appendix S8. Let $\mathcal{F}_{N,T}$ denote the set of all possible N by T matrices of candidate factor scores. In a nutshell, the candidate factor score distribution is the uniform distribution on $\mathcal{F}_{N,T}$ with two parameter matrices $\mathbf{W}$ and $\mathbf{G}$. The parameter matrix $\mathbf{W}$ denotes a J by T matrix of weights that determines the center of the candidate factor score distribution as $\mathbf{ZW}$, whereas $\mathbf{G}$ denotes a T by T covariance matrix of measurement errors when $\mathbf{ZW}$ is used as a measurement of $\mathbf{F}_{true}$. Using $\mathbf{ZW}$ as a measurement of $\mathbf{F}_{true}$ can be justified in that $\mathbf{ZW}$ is the only part of $\mathbf{F}_{true}$ that can be inferred from $\mathbf{Z}$, as shown in Appendix S9, and can be considered the best linear predictor for $\mathbf{F}_{true}$ given $\mathbf{Z}$ (e.g., Bartholomew, 1981) because $\mathbf{W}$ is equivalent to the weight matrix obtained by regressing $\mathbf{F}_{true}$ on $\mathbf{Z}_{std}$ (Thurstone, 1934). The matrix $\mathbf{ZW}$ is called *a matrix of expected candidate factor scores* in SFA. The square roots of $\mathbf{G}$'s diagonal entries are the

standard deviations of measurement errors when $\mathbf{ZW}$ is used to measure $\mathbf{F}_{true}$. These standard deviations of errors are called the *standard errors of measurement* in SFA (Leong & Huang, 2016), which refer to the standard amount of error that is expected to occur when a measurement is used to quantify the true amount of a particular quantity.

As fully described in Appendices S11 and S12, SFA estimates the candidate factor score distribution by minimizing the same cost function (3) in a least squares sense. More specifically, let $\hat{\mathcal{F}}_{N,T}$ denote an estimate of $\mathcal{F}_{N,T}$, which can be expressed with $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Lambda}}$ obtained from the first stage. Then, SFA randomly samples a prescribed number of $\hat{\mathbf{F}}$ s from $\hat{\mathcal{F}}_{N,T}$ and uses a set of the sampled $\hat{\mathbf{F}}$ values as an estimate of the candidate factor score distribution.

Once SFA estimates the candidate factor score distribution, it also estimates the expected candidate factor scores and the standard errors of measurement and uses these estimates for measuring $\mathbf{F}_{true}$. Moreover, SFA can obtain an individual's 95% candidate factor score interval for each latent variable from the estimated candidate factor score distribution, which contains 95% of all candidate factor scores for the individual.

4. Simulation Studies

We conduct two simulation studies to examine SFA's performance. In the first simulation study, we compare the performance of SFA and JCA via maximum likelihood (JCA-ML) in terms of parameter recovery and the frequency of improper solutions. In the second study, we examine the accuracy of the estimated standard errors of measurement and the coverage probability of the 95% candidate factor score intervals.

4.1. Simulation Study 1

As depicted in Figure 1, we specify two data generating models (Models A and B), which are different in model complexity. The vector $\mathbf{h}_{\cdot p}$ denotes N individuals' scores on the p th latent variable ($p = 1, 2, \dots, P$), whereas $\mathbf{z}_{\cdot j}$ denotes N individuals' scores on the j th indicator ($j = 1, 2, \dots, J$). In Model A, each group of three indicators loads on only one latent variable and all unique factors are uncorrelated with one another. On the other hand, in Model B, the first three groups of indicators ($\mathbf{z}_{\cdot 1}$ to $\mathbf{z}_{\cdot 9}$) additionally load on another latent variable ($\mathbf{h}_{\cdot 4}$), whose variance represents a common method variance (Podsakoff et al., 2003) for the nine indicators. This latent variable is set to be uncorrelated with the other latent variables. Moreover, the unique factors of three pairs of indicators ($\mathbf{z}_{\cdot 9}$ and $\mathbf{z}_{\cdot 13}$, $\mathbf{z}_{\cdot 10}$ and $\mathbf{z}_{\cdot 14}$, and $\mathbf{z}_{\cdot 11}$ and $\mathbf{z}_{\cdot 15}$) are assumed to be correlated. The two models share the same structural model, which imitates the one used in Bentler and Speckart (1979). In the structural model, three latent variables ($\mathbf{h}_{\cdot 1}$, $\mathbf{h}_{\cdot 2}$, and $\mathbf{h}_{\cdot 3}$) are exogenous, whose effects on an outcome latent variable ($\mathbf{h}_{\cdot 6}$) are mediated by another latent variable ($\mathbf{h}_{\cdot 5}$). Two errors in the structural model are set to be correlated. The prescribed parameter values of the two data generating models are also presented in Figure 1.

=== Insert Figure 1 about here ===

We consider six different sample sizes ($N = 30, 60, 120, 250, 500$, and 1000). We include a very small sample size ($N = 30$), where JCA-ML is more likely to result in improper solutions (e.g., Wolf, Harrington, Clark, & Miller, 2013), in order to examine whether SFA can be used as an alternative to JCA-ML in such a condition. We also consider two distributions of factor scores: normal and non-normal. Whereas the normal distribution is a multivariate distribution with a skewness of 0 and a kurtosis of 3, the non-normal distribution is set to have a skewness of 1.25 and a kurtosis of 3.75, which have been used to evaluate the relative performance of SEM methods in the literature (e.g., Hwang et al., 2010). We generate 2000 samples per experimental

condition based on Mattson's (1997) procedure, whereby we randomly generate $\mathbf{H}_{true}$ and $\mathbf{E}_{true}$ per sample and then generate $\mathbf{Z}$ based on (1). To assess the bias of the SFA and JCA-ML estimators given a representative sample, we randomly generate a set of representative $\mathbf{H}_{true}$ and $\mathbf{E}_{true}$ satisfying $mean(\mathbf{F}_{true}) = \mathbf{0}$ and $cov(\mathbf{F}_{true}) = \mathbf{\Lambda}$, and then generate a sample of $N = 30$ by $\mathbf{Z} = \mathbf{H}_{true}\mathbf{\Lambda} + \mathbf{E}_{true}$ such that $\mathbf{Z}$ satisfies $mean(\mathbf{Z}) = \mathbf{0}$ and $cov(\mathbf{Z}) = \mathbf{\Sigma}$.

Furthermore, we consider two specifications of each data generating model. One is the correctly specified model, which is equivalent to the data generating model, and the other is a misspecified model, where one path coefficient is incorrectly removed from the data generating model. In Figure 1, the path coefficient omitted incorrectly is labeled A.

We use the R package *lavaan* (version 0.5-16) (Rosseel, 2012) for JCA-ML and the MATLAB package *SFA Prime* (version 0.9)¹ for SFA. To avoid non-convergence in JCA-ML, we additionally apply the standard bounded estimation method with marker indicators for JCA-ML, as suggested by De Jonckere and Rosseel (2022). We set the first indicator of each latent variable as the marker indicator. For SFA, we fix the sign of the first loading for each latent variable to be positive. The default tolerance level of *lavaan* (2.220×10^{16}) is employed for JCA-ML, whereas 10^{-13} is used for SFA's ALS algorithm. The maximum number of iterations is 50000 for both estimators. For each combination of the experimental conditions, we compute the (empirical) biases and root mean squared errors (RMSE) of the estimators. In the calculation of these measures for each estimator, any sample involving non-convergence or improper solutions is excluded. We treat negative variance estimates and negative definite covariance matrices of common and unique factors as improper solutions, as in the *lavInspect* function of the *lavaan* package. We also consider standardized loading estimates over ± 1 as improper solutions, as $\mathbf{h}_{.4}$ is assumed to be uncorrelated with $\mathbf{h}_{.1}$, $\mathbf{h}_{.2}$, and $\mathbf{h}_{.3}$ in the model.

¹ This MATLAB package is available at <https://sfaprime.wordpress.com/>.

Tables 1 presents the proportions of the samples involving non-convergence or improper solutions per condition for SFA and JCA-ML. As expected, SFA always produces proper solutions without a convergence problem, regardless of the experimental conditions. On the other hand, JCA-ML results in improper solutions at times. The relative ratio of converging to improper solutions increases when the sample size is small and/or the model is complex. For instance, when $N = 30$ and the model is complex, JCA-ML produces improper solutions in around 64% of the samples regardless of the conditions. This ratio gradually decreases when the model is simple and/or when the sample size becomes large. The distribution of the factor scores does not seem to make substantial differences in the occurrence of improper solutions.

=== Insert Table 1 about here ===

Tables 2 and 3 provide the average biases and RMSE values of SFA and JCA-ML for each set of model parameters (e.g., loadings and path coefficients) per condition under the two model specifications. We only report the results that are aggregated over the normal and non-normal distributions as the relative performance of the two estimators was not substantially different between the two distributional conditions. This may be because SFA does not require any distributional assumption for parameter estimation and JCA-ML seems robust to the violation of its distributional assumption (i.e., normality) in recovering model parameters (e.g., Cassel et al., 1999; Hwang et al., 2010). The average biases and RMSE values for the same sets of model parameters under each distributional condition are provided in Table S1 to Table S4. Furthermore, the bias and RMSE values of the two estimators for each individual parameter per condition are provided in Table S5 to Table S8.

Under the correct model specification, both estimators are biased for all model parameters to some extent when the sample size is very small (e.g., $N = 30$) regardless of model

complexity. However, in general, SFA tends to show comparable or smaller bias than JCA-ML. For instance, when $N = 30$, the average biases of the SFA and JCA-ML estimators, respectively, are .01 and .01 for the loadings and .01 and .03 for the path coefficients under the simple model, while those are .05 and .08 for the loadings and .03 and .10 for the path coefficients under the complex model. The average biases of both estimators become smaller with the sample size and close to zero when $N = 1000$ in all conditions. Given a representative sample, the average biases of both estimators are close to zero, suggesting that the two estimators are virtually unbiased given a representative sample.

Conversely, on average, SFA tends to show larger RMSE values than JCA-ML under correct model specification. This suggests that the SFA estimator is less efficient than the JCA-ML estimator as their biases are generally comparable. This efficiency difference is more pronounced when the sample size is small; for instance, when $N = 30$, SFA generally have larger RMSE values for the path coefficients than JCA-ML in both correct (data generating) models. The results are consistent with the literature that a full-information estimator is typically more efficient than a limited-information estimator under correct model specification (e.g., Fomby et al., 2012, Chapter 22). Nonetheless, the RMSE differences between SFA and JCA-ML decrease with the sample size and become negligible when $N \geq 500$.

On the contrary, under model misspecification, SFA generally tends to produce less biased and more accurate estimates than JCA-ML regardless of model complexity and sample size. The average biases and RMSE values of the two estimators decrease with the sample size. However, on average, JCA-ML always tends to provide biased estimates of the loadings, latent variable covariances, and path coefficients, leading to larger average RMSE values for the model parameters than SFA. For instance, when $N = 1000$ and the model is complex, the average biases

of the SFA and JCA-ML estimators, respectively, are .00 and .04 for the loadings, .00 and .09 for the latent variable covariances, and .00 and .02 for the path coefficients. In a representative sample, the JCA-ML estimator is still biased for both measurement and structural model parameters. This is consistent with that a misspecification in part of the model may lead the JCA-ML estimator to be biased for entire model parameters (e.g., Devlieger & Rosseel, 2017).

Taken together, SFA is free from the occurrence of improper solutions regardless of the experimental conditions, whereas JCA-ML suffers from convergence to improper solutions in several conditions. Moreover, SFA and JCA-ML generally show a similar pattern of bias across the two model specifications, even though our simulation design may be less favorable to SFA because it is applied to all samples, many of which lead to the improper solution problem in JCA-ML. This suggests that SFA can be an alternative to JCA-ML when JCA-ML fails to converge to proper solutions. Between the two methods, in general, when the model is correctly specified, JCA-ML tends to provide more accurate estimates with smaller RMSE values, whereas when the model is misspecified, SFA tends to yield more accurate estimates. Nonetheless, it is noteworthy to mention that when the model is correct, the RMSE differences of the methods virtually disappear when $N \geq 500$, whereas when the model is incorrect, the differences remain regardless of the sample size.

4.2. Simulation Study 2

In the second simulation study, we draw on the same samples generated in the previous study and apply SFA to fit the correctly specified model to each sample. For each sample, we randomly generate 500 matrices of candidate factor score estimates to estimate the candidate factor score distribution, from which we calculate 95% candidate factor score intervals for each individual and examine their coverage probability. We calculate the proportion of the 95%

candidate factor score intervals that contain the true factor score of each latent variable per sample and average it over the samples per experimental condition. Furthermore, we estimate the expected candidate factor scores and the standard errors of measurement from the candidate factor score distribution and subsequently investigate whether the estimated standard errors of measurement are close to the true standard errors of measurement when using the estimated expected candidate factor scores to measure $\mathbf{F}_{true}$. We calculate the average RMSE value of the standard error of measurement for each latent variable over the samples per condition.

Table 4 provides the average proportions of the 95% candidate factor score intervals and the average RMSE values for the standard errors of measurement in each condition. When the sample size is very small (i.e., $N = 30$), approximately 63% to 79% of the candidate factor score intervals contain the true factor score of each latent variable on average. The proportions rapidly increase as the sample size increases and become close to 95% when $N \geq 500$, regardless of the other conditions. Moreover, when $N = 30$, on average, the RMSE values for the standard errors of measurement range from .16 to .40. However, they rapidly decrease with the sample size and become close to zero when $N = 1000$.

=== Insert Table 4 about here ===

5. Empirical Illustration

We illustrate how to apply SFA in practice and utilize the estimated candidate factor distribution for making probabilistic inferences about the true factor scores. Fornell et al. (1996) proposed a structural equation model, named the American customer satisfaction index (ACSI) model, to measure customers' average satisfaction levels on major American companies in different sectors and industries. As displayed in Figure 2, they assumed the causal relationships between a

focal latent variable, customer satisfaction (CS), and its five antecedent or outcome latent variables—customer expectation (CE), perceived quality (PQ), perceived value (PV), customer complaints (CC), and customer loyalty (CL). This model contains fourteen indicators to measure the six latent variables. Refer to Fornell et al. (1996) for more information on the latent variables and their indicators.

=== Insert Figure 2 about here ===

We analyze company-level data from the ACSI database collected in 2002. The sample size is 152. We apply SFA and JCA-ML to fit the model to the dataset and obtain their parameter estimates. We estimate the standard errors and 95% confidence intervals of SFA's parameter estimates based on 1000 bootstrap samples. We use the *SFA Prime* and *lavaan* packages for SFA and JCA-ML, respectively, applying the same values for the maximum number of iterations and tolerance levels as those used in the simulation studies.

SFA provides that $IPE_O = .013$, indicating that the measurement model with the candidate factor score estimates explains 98.7% of the total variance of the fourteen indicators. However, the critical value of IPE_O at $\alpha = .05$ obtained from the B-S bootstrap method is .003, leading to a rejection of the measurement model with the candidate factor score estimates. This suggests that some part of the measurement model may be misspecified or the candidate factor score estimates may fail to recover the population covariance matrix. In this case, researchers should revise the measurement model more thoroughly before moving on to the next SFA stage. For illustration purposes, nonetheless, we here proceed with the second stage, assuming that the measurement model is correctly specified. SFA provides that $IPE_L = .305$, indicating that the structural model accounts for 69.5% of the variances of all the endogenous latent variables on average. When

applying JCA-ML, we find that the ACSI model may also be rejected as a whole ($\chi^2(69) = 1768.16, p = .000$).

Tables 5 and 6 exhibit the loading and path coefficient estimates, and their standard errors and 95% confidence intervals obtained from SFA and JCA-ML. SFA's loading estimates are all statistically significant and large, and most of its path coefficient estimates are statistically significant and generally consistent with the hypothesized relationships in the ACSI model. In contrast, JCA-ML fails to provide the standard errors and 95% confidence intervals of its parameter estimates, making it impossible to test the statistical significance of the estimates. Moreover, some of JCA-ML's estimates appear counterintuitive theoretically or substantively. For instance, some loading estimates for perceived value and customer loyalty are equal to one, indicating that the corresponding indicators contain no measurement error. Also, the path coefficient estimate relating customer satisfaction to customer loyalty is negative, indicating that more satisfied consumers are less likely to remain. This is inconsistent with that the two variables are expected to be positively correlated (Fornell et al., 1996).

=== Insert Tables 3 and 4 about here ===

We also estimate SFA's candidate factor score distribution for customer satisfaction. To estimate this distribution per company, specifically, we derive 1000 sets of candidate factor score estimates based on the resampling method described in Appendix S11. For illustration, we show the marginal distributions of the candidate factor score estimates for three companies, labeled ID1, 2, and 21, in Figure 3. ID1 shows a higher level of customer satisfaction (mean = .996, 95% candidate factor score interval = [.813, 1.108]) than ID2 (mean = .704, 95% candidate factor score interval = [.542, .843]) and ID21 (mean = -.270, 95% candidate factor score interval = [-.462, -.151]).

=== Insert Figure 3 about here ===

However, the 95% candidate factor score intervals of ID1 and ID2 overlap slightly, indicating that there may still be a chance that the two companies have the same level of customer satisfaction. Thus, we further calculate the probability for ID1 to have a higher level of customer satisfaction than ID2. Figure 4 displays the distribution of the differences in customer satisfaction scores between them. The blue area of the histogram amounts to the probability that the differences between ID1 and ID2 are greater than zero. In our example, this probability is equal to 99.4%, indicating that 99.4% of ID1's candidate factor scores are greater than their respective candidate factor scores of ID2. The result helps researchers conclude with great confidence that ID1 has a higher level of customer satisfaction than ID2.

=== Insert Figure 4 about here ===

6. Discussion

We proposed a new data matrix-based SEM approach, named structured factor analysis (SFA), which can circumvent the two enduring issues of Jöreskog's covariance-based approach (JCA; Jöreskog, 1978): the occurrence of improper solution and the lack of a formal procedure for making probabilistic inferences about the true factor scores. SFA begins by estimating measurement model parameters and (candidate) factor scores simultaneously while preventing the occurrence of improper solutions. In the first simulation study, we demonstrated that the SFA estimator always converged to proper solutions even in an experimental condition where the JCA-ML estimator yielded improper solutions in more than 60% of the samples. Furthermore, in the same condition, the bias of the SFA estimator was smaller than or comparable to that of the

JCA-ML estimator on average. The results suggest that SFA can be an alternative when JCA-ML produces improper solutions.

In addition, SFA can estimate the probability distribution of all factor scores that can generate a given data matrix of indicators, under the assumption that the true factor scores are representative of the population. In the second simulation study, we showed that the coverage probability of SFA's candidate factor score interval converged to a prescribed rate (e.g., 95%) when the sample was sufficiently large (e.g., $N \geq 500$). Moreover, in a real data analysis, we illustrated how the estimated candidate factor score distribution could be utilized for making probabilistic inferences on the true factor scores. We could obtain the 95% candidate factor score intervals for a latent variable (i.e., customer satisfaction) for all observations and calculate the probability for an observation to have a higher level of the latent variable than another observation. JCA does not enable such probabilistic inferences about the true factor scores.

Although SFA is developed as an alternative to JCA, it can also be considered a data matrix-based alternative to two-stage covariance-based approaches such as factor score regression (Croon, 2002; Lu et al., 2011; Skrondal & Laake, 2001) or structural-after-measurement methods (Rosseel & Loh, 2022). Similar to SFA, these existing two-stage approaches begin by estimating measurement model parameters as well as individual factor scores (or weights for obtaining factor scores) and subsequently estimate remaining structural model parameters based on the estimated factor scores. Nonetheless, SFA has clear advantages over the two-stage covariance-based approaches. First, the two-stage covariance-based approaches do not completely overcome the occurrence of improper solutions (e.g., Cho et al., 2022) and have no mechanism for making probabilistic inferences about the true factor scores. Second, they rely on conventional formulas that estimate factor scores as weighted sums of

indicators' scores (e.g., Anderson & Rubin, 1956; Bartlett, 1937; Thurstone, 1934). As the sample covariance matrix of these factor scores can never be true, the existing approaches have to apply an additional bias correction/avoiding procedure (Croon, 2002; Skrondal & Laake, 2001) for obtaining unbiased path coefficient estimates in the second stage. In contrast, SFA seeks to estimate candidate factor scores whose covariance matrix is equivalent to the covariance matrix of the true factor scores, thereby enabling itself to provide unbiased path coefficient estimates without adopting any additional procedure.

Furthermore, SFA can be seen as an extension of matrix decomposition factor analysis (MDFA), which is a class of data matrix-based approaches to exploratory factor analysis (Adachi & Trendafilov, 2018; de Leeuw, 2004; Sočan, 2003; Unkel & Trendafilov, 2010). MDFA aims to minimize the same cost function as SFA's to estimate measurement model parameters and factor scores concurrently. However, it does not allow researchers to specify which parameters are free or fixed based on their prior knowledge, simply pre-setting the factor covariance matrix to be diagonal and all loadings to be non-zero. Conversely, SFA permits specifying and estimating a broad array of both measurement and structural models, thus being able to stand as a full-fledged method for SEM.

Despite its usefulness, SFA can be technically extended in various ways to enhance its generality. For example, SFA is currently developed to deal with continuous variables only. Thus, it is important to extend SFA to handle discrete variables as well, which are not uncommon to encounter in practice. Moreover, it is meaningful to expand SFA to accommodate more complex models, including those with latent interactions (e.g., Marsh et al., 2013), random intercepts (Maydeu-Olivares & Coffman, 2006), and/or components (Gu et al., 2019; Hwang et al., 2021).

Furthermore, although SFA provides a statistical testing procedure for its goodness-of-fit index IPE_O based on the B-S bootstrap method, it does not offer a procedure for identifying which parts of the measurement model might be wrong. In JCA, the modification index (MI; Sörbom, 1989) provides such information and serves as a reference on how to modify the model to improve its fit, under the normality assumption. Thus, it may be useful to develop an index akin to the MI for SFA.

Besides SFA's technical refinement, further empirical studies are needed to investigate its performance from more diverse perspectives. In our simulation studies, we have considered a limited number of data generating models, although they could still be considered complex involving many factors, a bi-factor structure, and/or correlated unique factors. Thus, it would be desirable to examine whether SFA can show similar, impressive performance in a greater variety of data generating models. Also, we focused on evaluating the performance of SFA and JCA, as SFA is proposed as an alternative to JCA. However, there are many other SEM methods (e.g., Bollen, 2019; Hwang et al., 2017; Lance et al., 1988; Oldenburg, 2020; Rosseel & Loh, 2022), so it can be important to compare SFA to other SEM methods through simulation studies.

In closing, SFA represents a novel data matrix-based approach to SEM that integrates the measurement of factor scores and the estimation of model parameters into a single framework. Although more technical and empirical studies can be needed to further improve its data-analytic flexibility and establish its practical utility as we have discussed some of them above, we believe that SFA can make a significant contribution to broadening SEM's applicability. Obviously, SFA can be a sensible choice if researchers want to avoid the occurrence of improper solutions and/or make probabilistic inferences about the true factor scores. Moreover, even after using other SEM methods, researchers may still consider applying SFA to evaluate their model's prediction

accuracy for indicators based on the IPE_O. Finally, it will be essential to develop an *R* package or user-friendly software for SFA to facilitate its wide adoption by researchers and practitioners.

References

- Adachi, K., & Trendafilov, N. T. (2018). Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika*, 83(2), 407–424. <https://doi.org/10.1007/s11336-017-9600-y>
- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In Jerzy Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (5th ed., pp. 111–150). University of California Press.
- Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, 34(1), 93–99. <https://doi.org/10.1111/j.2044-8317.1981.tb00620.x>
- Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, 28, 97–104.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bentler, P. M. (1995). *EQS structural equations program manual* (Vol. 6). Multivariate Software.
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78–117. <https://doi.org/10.1177/0049124187016001004>
- Bentler, P. M., & Speckart, G. (1979). Models of attitude–behavior relations. *Psychological Review*, 86(5), 452–464. <https://doi.org/10.1037/0033-295X.86.5.452>
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. <https://doi.org/10.1002/9781118619179>
- Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, 61(1), 109–121. <https://doi.org/10.1007/BF02296961>

- Bollen, K. A. (2019). Model implied instrumental variables (MIIVs): An alternative orientation to structural equation modeling. *Multivariate Behavioral Research*, 54(1), 31–46.
<https://doi.org/10.1080/00273171.2018.1483224>
- Bollen, K. A., & Stine, R. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Sage Publications.
- Cassel, C., Hackl, P., & Westlund, A. H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, 26(4), 435–446.
<https://doi.org/10.1080/02664769922322>
- Chen, F., Bollen, K. A., Paxton, P. M., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, 29(4), 468–508. <https://doi.org/10.1177/0049124101029004003>
- Cho, G., Sarstedt, M., & Hwang, H. (2022). A comparative evaluation of factor- and component-based structural equation modeling methods under (in)consistent model specifications. *British Journal of Mathematical and Statistical Psychology*, 75(2), 220–251.
<https://doi.org/10.1111/bmsp.12255>
- Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 195–223). Erlbaum.
- De Jonckere, J., & Rosseel, Y. (2022). Using bounded estimation to avoid nonconvergence in small sample structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(3), 412–427. <https://doi.org/10.1080/10705511.2021.1982716>
- de Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models:*

- Theory and applications* (pp. 121–134). Springer Netherlands. https://doi.org/10.1007/978-1-4020-1958-6_7
- de Leeuw, J. (2017). *Factor analysis as matrix decomposition and approximation: Theory*. <http://deleeuwpx.net/pubfolders/factor/factor.pdf>
- de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41(4), 471–503. <https://doi.org/10.1007/BF02296971>
- Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 13(Suppl 1), 31–38. <https://doi.org/10.1027/1614-2241/a000130>
- Dijkstra, T. K. (2011). *Consistent partial least squares estimators for linear and polynomial factor models*. <https://doi.org/10.13140/RG.2.1.3997.0405>
- Dijkstra, T. K. (2013). *The simplest possible factor model estimator*. <https://doi.org/10.13140/RG.2.1.3605.6809>
- Dijkstra, T. K. (2017). A perfect match between a model and a mode. In H. Latan & R. Noonan (Eds.), *Partial least squares path modeling: Basic concepts, methodological issues and applications* (pp. 55–80). Springer. https://doi.org/10.1007/978-3-319-64069-3_4
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM. <https://doi.org/10.1137/1.9781611970319>
- Fomby, T. B., Johnson, S. R., & Hill, R. C. (2012). *Advanced econometric methods*. Springer. <https://doi.org/10.1007/978-1-4419-8746-4>

- Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American customer satisfaction index: Nature, purpose, and findings. *Journal of Marketing*, 60(4), 7. <https://doi.org/10.2307/1251898>
- Gu, F., Yung, Y.-F., & Cheung, M. W.-L. (2019). Four covariance structure models for canonical correlation analysis: A COSAN modeling approach. *Multivariate Behavioral Research*, 54(2), 192–223. <https://doi.org/10.1080/00273171.2018.1512847>
- Hwang, H., Cho, G., Jung, K., Falk, C. F., Flake, J., & Jin, M. J. (2021). An approach to structural equation modeling with both factors and components: Integrated generalized structured component analysis. *Psychological Methods*, 26(3), 273–294. <https://doi.org/10.1037/met0000336>.
- Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, M. A., & Hong, S. (2010). A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research*, 47(4), 699–712. <https://doi.org/10.2139/ssrn.1585305>
- Hwang, H., Takane, Y., & Jung, K. (2017). Generalized structured component analysis with uniqueness terms for accommodating measurement error. *Frontiers in Psychology*, 8, 2137. <https://doi.org/10.3389/fpsyg.2017.02137>
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4), 443–477. <https://doi.org/10.1007/BF02293808>
- Jöreskog, K. G., & Sorbom, D. (1986). *PRELIS: A program for multivariate data screening and data summarization*. Scientific Software, Mooresville.
- Lance, C. E., Cornwell, J. M., & Mulaik, S. A. (1988). Limited information parameter estimates for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research*, 23(2), 171–187. https://doi.org/10.1207/s15327906mbr2302_3

- Leong, F. T. L., & Huang, J. L. (2016). Standard error of measurement. In *Britannica*.
<https://www.britannica.com/science/standard-error-of-measurement>
- Lu, I. R. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, 28(3), 258–268.
<https://doi.org/10.1016/j.ijresmar.2011.03.006>
- Marsh, H. W., Wen, Z., Hau, K.-T., & Nagengast, B. (2013). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course, 2nd ed.* (pp. 267–308). IAP Information Age Publishing.
- Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, 32(4), 355–373.
https://doi.org/10.1207/s15327906mbr3204_3
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344–362. <https://doi.org/10.1037/1082-989X.11.4.344>
- McDonald, R. P. (2004). Respecifying improper structures. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(2), 194–209. https://doi.org/10.1207/s15328007sem1102_3
- Mulaik, S. A. (2009). *Foundations of factor analysis* (2nd ed.). Chapman and Hall/CRC Press.
<https://doi.org/10.1201/b15851>
- Newsom, J. T. (2014). *Improper solutions in SEM*.
https://web.pdx.edu/~newsomj/semclass/ho_improper.pdf
- Oldenburg, G. (2020). Structural Equation Modeling. *The Mathematica Journal*.
<https://doi.org/10.3888/tmj.22-5>

- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rosseel, Y., & Loh, W. W. (2022). A structural after measurement (SAM) approach to structural equation modeling. *Manuscript under review*. Retrieved from <https://osf.io/w9bmf>
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66*(4), 563–575. <https://doi.org/10.1007/BF02296196>
- Sočan, G. (2003). *The incremental value of minimum rank factor analysis*. Groningen: University of Groningen.
- Sörbom, D. (1989). Model modification. *Psychometrika, 54*(3), 371–384. <https://doi.org/10.1007/BF02294623>
- Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika, 44*(2), 157–167. <https://doi.org/10.1007/BF02293967>
- Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) Handout. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(6), 777–781. <https://doi.org/10.1080/10705511.2016.1217487>
- Thurstone, L. L. (1934). The vectors of mind. *Psychological Review, 41*(1), 1–32. <https://doi.org/10.1037/h0075959>
- Unkel, S., & Trendafilov, N. T. (2010). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review, 78*(3), 363–382.

<https://doi.org/10.1111/j.1751-5823.2010.00120.x>

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety.

Educational and Psychological Measurement, 76(6), 913–934.

<https://doi.org/10.1177/0013164413495237>

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). The MIT Press.

Table 1. Percentages of the samples involving non-convergence or improper solutions per condition.

Estimator	N	Correct model				Misspecified model			
		Simple		Complex		Simple		Complex	
		Normal	Non-normal	Normal	Non-normal	Normal	Non-normal	Normal	Non-normal
SFA	30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	120	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	250	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	500	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
JSA-ML	30	0.34	0.33	0.64	0.64	0.28	0.27	0.60	0.60
	60	0.13	0.12	0.36	0.38	0.08	0.08	0.31	0.33
	120	0.02	0.02	0.16	0.18	0.00	0.01	0.16	0.17
	250	0.00	0.00	0.05	0.05	0.00	0.00	0.07	0.09
	500	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02
	1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2. The average bias and RMSE values of the SFA and JCA-ML estimators per condition under correct model specification.

Model complexity	N	Estimator	Loadings		(Co)variances of unique factors		Covariances of latent variables		Path coefficients		(Co)variances of structural errors	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Simple	30	SFA	0.01	0.15	0.03	0.20	0.01	0.23	0.01	0.32	0.05	0.22
		JCA-ML	0.01	0.16	0.03	0.22	0.03	0.23	0.03	0.28	0.08	0.31
	60	SFA	0.00	0.10	0.01	0.14	0.01	0.15	0.00	0.16	0.02	0.15
		JCA-ML	0.00	0.10	0.01	0.14	0.01	0.15	0.01	0.16	0.02	0.16
	120	SFA	0.00	0.07	0.01	0.10	0.00	0.11	0.00	0.11	0.01	0.11
		JCA-ML	0.00	0.07	0.00	0.10	0.00	0.10	0.00	0.11	0.01	0.11
	250	SFA	0.00	0.05	0.00	0.07	0.00	0.07	0.00	0.07	0.00	0.07
		JCA-ML	0.00	0.05	0.00	0.07	0.00	0.07	0.00	0.07	0.01	0.07
	500	SFA	0.00	0.03	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05
		JCA-ML	0.00	0.03	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05
	1000	SFA	0.00	0.02	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
		JCA-ML	0.00	0.02	0.00	0.03	0.00	0.03	0.00	0.04	0.00	0.04
	∞	SFA	0.00		0.00		0.00		0.00		0.00	
		JCA-ML	0.00		0.00		0.00		0.00		0.00	
Complex	30	SFA	0.05	0.21	0.02	0.12	0.04	0.26	0.03	0.44	0.05	0.29
		JCA-ML	0.08	0.23	0.02	0.13	0.08	0.25	0.10	0.35	0.10	0.38
	60	SFA	0.02	0.14	0.01	0.09	0.02	0.19	0.01	0.35	0.03	0.21
		JCA-ML	0.03	0.15	0.01	0.09	0.04	0.17	0.04	0.22	0.04	0.21
	120	SFA	0.01	0.09	0.00	0.06	0.00	0.13	0.01	0.15	0.02	0.15
		JCA-ML	0.01	0.09	0.00	0.06	0.01	0.11	0.01	0.14	0.01	0.13
	250	SFA	0.00	0.06	0.00	0.04	0.00	0.09	0.01	0.10	0.01	0.11
		JCA-ML	0.00	0.06	0.00	0.04	0.00	0.08	0.00	0.09	0.01	0.09
	500	SFA	0.00	0.04	0.00	0.03	0.00	0.06	0.00	0.07	0.01	0.07
		JCA-ML	0.00	0.04	0.00	0.03	0.00	0.06	0.00	0.07	0.01	0.07
	1000	SFA	0.00	0.03	0.00	0.02	0.00	0.04	0.00	0.05	0.00	0.05
		JCA-ML	0.00	0.03	0.00	0.02	0.00	0.04	0.00	0.05	0.00	0.05
	∞	SFA	0.00		0.00		0.00		0.00		0.00	
		JCA-ML	0.00		0.00		0.00		0.00		0.00	

Note: The results on the rows in $N = \infty$ are obtained from one sample representative of the population.

Table 3. The average bias and RMSE values of the SFA and JCA-ML estimators per condition under model misspecification.

Model complexity	N	Estimator	Loadings		(Co)variances of unique factors		Covariances of latent variables		Path coefficients		(Co)variances of structural errors	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Simple	30	SFA	0.01	0.15	0.03	0.20	0.01	0.23	0.01	0.24	0.04	0.23
		JCA-ML	0.01	0.16	0.02	0.22	0.07	0.25	0.02	0.37	0.12	0.48
	60	SFA	0.00	0.10	0.01	0.14	0.01	0.15	0.00	0.16	0.01	0.16
		JCA-ML	0.00	0.11	0.01	0.15	0.06	0.18	0.00	0.21	0.03	0.21
	120	SFA	0.00	0.07	0.01	0.10	0.00	0.11	0.00	0.11	0.01	0.11
		JCA-ML	0.00	0.07	0.00	0.10	0.07	0.15	0.01	0.13	0.01	0.13
	250	SFA	0.00	0.05	0.00	0.07	0.00	0.07	0.00	0.07	0.00	0.08
		JCA-ML	0.00	0.05	0.00	0.07	0.07	0.12	0.01	0.08	0.00	0.08
	500	SFA	0.00	0.03	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05
		JCA-ML	0.00	0.03	0.00	0.05	0.07	0.10	0.01	0.06	0.00	0.06
	1000	SFA	0.00	0.02	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
		JCA-ML	0.00	0.02	0.00	0.03	0.07	0.09	0.01	0.04	0.00	0.04
	∞	SFA	0.00		0.00		0.00		0.00		0.00	
		JCA-ML	0.00		0.00		0.07		0.00		0.00	
Complex	30	SFA	0.05	0.21	0.02	0.12	0.04	0.26	0.03	0.30	0.06	0.31
		JCA-ML	0.08	0.24	0.02	0.13	0.11	0.26	0.06	0.36	0.11	0.50
	60	SFA	0.02	0.14	0.01	0.09	0.02	0.19	0.01	0.20	0.02	0.20
		JCA-ML	0.04	0.16	0.01	0.09	0.10	0.20	0.03	0.22	0.03	0.25
	120	SFA	0.01	0.09	0.00	0.06	0.00	0.13	0.00	0.13	0.01	0.13
		JCA-ML	0.04	0.11	0.01	0.06	0.09	0.15	0.03	0.14	0.01	0.13
	250	SFA	0.00	0.06	0.00	0.04	0.00	0.09	0.00	0.09	0.01	0.09
		JCA-ML	0.04	0.08	0.00	0.04	0.09	0.13	0.02	0.09	0.01	0.09
	500	SFA	0.00	0.04	0.00	0.03	0.00	0.06	0.00	0.06	0.00	0.06
		JCA-ML	0.04	0.06	0.00	0.03	0.09	0.11	0.02	0.07	0.00	0.06
	1000	SFA	0.00	0.03	0.00	0.02	0.00	0.04	0.00	0.04	0.00	0.04
		JCA-ML	0.04	0.06	0.00	0.02	0.09	0.10	0.02	0.05	0.00	0.04
	∞	SFA	0.00		0.00		0.00		0.00		0.00	
		JCA-ML	0.04		0.00		0.09		0.02		0.00	

Note: The results on the rows in $N = \infty$ are obtained from one sample representative of the population.

Table 4. The average proportion (%) of the 95% candidate factor score intervals that contain the true score of each latent variable and the average RMSE value for the standard errors of measurement per condition.

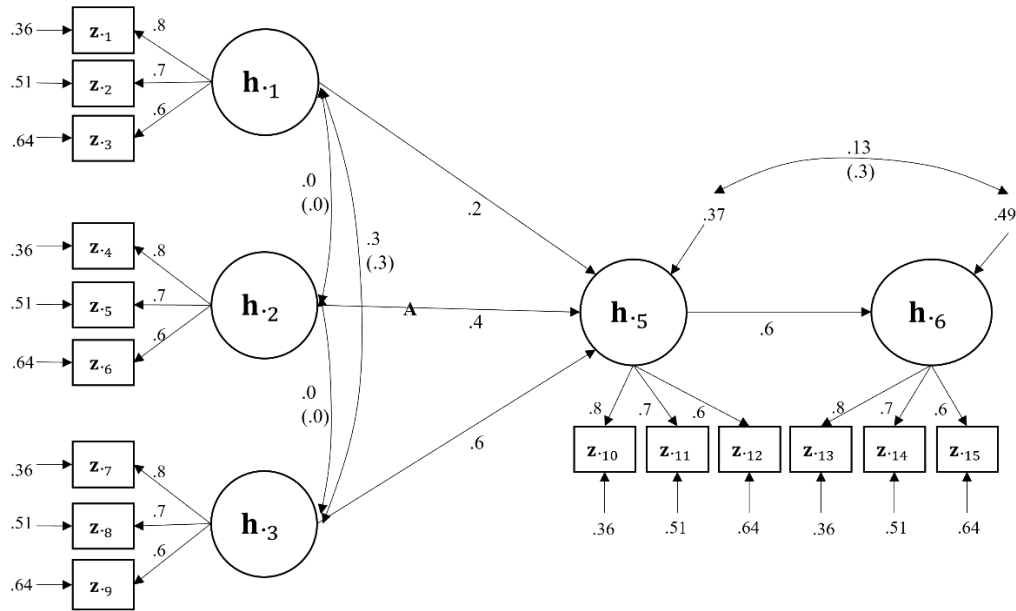
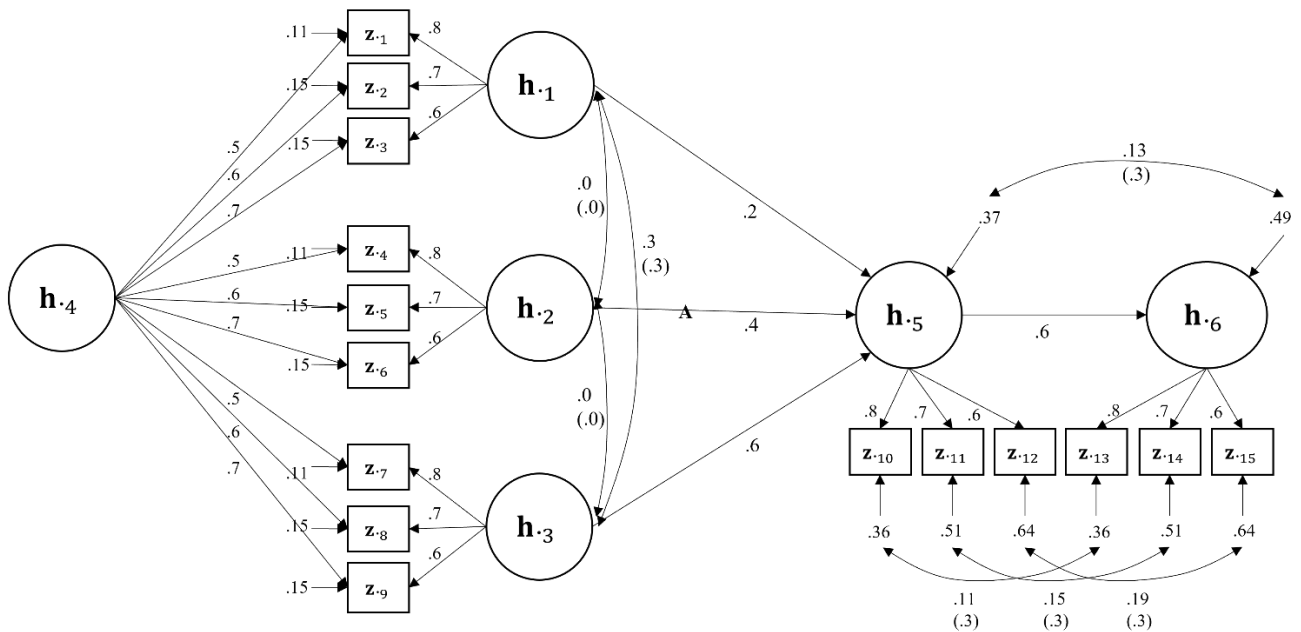
Model complexity	Distribution	N	Average proportion						Average RMSE					
			$\mathbf{h}_1$	$\mathbf{h}_2$	$\mathbf{h}_3$	$\mathbf{h}_4$	$\mathbf{h}_5$	$\mathbf{h}_6$	$\mathbf{h}_1$	$\mathbf{h}_2$	$\mathbf{h}_3$	$\mathbf{h}_4$	$\mathbf{h}_5$	$\mathbf{h}_6$
Simple	Normal	30	0.74	0.75	0.78	-	0.79	0.78	0.22	0.21	0.18	-	0.16	0.18
		60	0.86	0.86	0.88	-	0.88	0.88	0.12	0.11	0.09	-	0.08	0.09
		120	0.91	0.91	0.92	-	0.92	0.92	0.05	0.05	0.04	-	0.04	0.04
		250	0.94	0.94	0.94	-	0.94	0.94	0.03	0.02	0.02	-	0.02	0.02
		500	0.94	0.94	0.95	-	0.94	0.95	0.01	0.01	0.01	-	0.01	0.01
		1000	0.95	0.95	0.95	-	0.95	0.95	0.01	0.01	0.00	-	0.01	0.01
	Non-Normal	30	0.75	0.75	0.78	-	0.78	0.78	0.22	0.22	0.19	-	0.17	0.19
		60	0.85	0.86	0.88	-	0.88	0.88	0.12	0.11	0.10	-	0.09	0.09
		120	0.91	0.91	0.92	-	0.92	0.92	0.05	0.05	0.05	-	0.04	0.04
		250	0.93	0.94	0.94	-	0.94	0.94	0.03	0.02	0.02	-	0.02	0.02
		500	0.94	0.94	0.94	-	0.94	0.94	0.01	0.01	0.01	-	0.01	0.01
		1000	0.95	0.95	0.95	-	0.95	0.95	0.01	0.01	0.00	-	0.00	0.00
Complex	Normal	30	0.63	0.63	0.63	0.62	0.72	0.70	0.34	0.33	0.32	0.40	0.23	0.25
		60	0.78	0.79	0.78	0.79	0.86	0.85	0.20	0.19	0.19	0.21	0.11	0.12
		120	0.87	0.88	0.88	0.88	0.91	0.91	0.10	0.10	0.10	0.10	0.05	0.06
		250	0.92	0.92	0.92	0.92	0.94	0.94	0.05	0.05	0.05	0.05	0.03	0.03
		500	0.93	0.93	0.93	0.93	0.94	0.94	0.03	0.03	0.02	0.03	0.01	0.01
		1000	0.94	0.94	0.94	0.94	0.95	0.95	0.01	0.01	0.01	0.01	0.01	0.01
	Non-Normal	30	0.63	0.64	0.64	0.64	0.71	0.70	0.34	0.33	0.32	0.39	0.24	0.25
		60	0.78	0.78	0.78	0.79	0.86	0.85	0.20	0.20	0.19	0.21	0.12	0.12
		120	0.87	0.87	0.87	0.88	0.91	0.91	0.10	0.10	0.10	0.10	0.06	0.06
		250	0.92	0.91	0.92	0.91	0.94	0.93	0.05	0.05	0.05	0.05	0.03	0.03
		500	0.93	0.93	0.93	0.93	0.94	0.94	0.02	0.02	0.02	0.03	0.01	0.01
		1000	0.94	0.94	0.94	0.94	0.95	0.95	0.01	0.01	0.01	0.01	0.01	0.01

Table 5. The loading estimates, their standard errors (SE), and 95% confidence intervals (CI) in the ACSI model obtained from SFA and JCA-ML.

Latent variable	Indicator	SFA			JCA-ML		
		Estimate	SE	95% CI	Estimate	SE	95% CI
CE	z _{·1}	.926	.013	[.898, .952]	.935	NA	NA
	z _{·2}	.960	.001	[.941, .976]	.962	NA	NA
	z _{·3}	.930	.013	[.903, .951]	.919	NA	NA
PQ	z _{·4}	.974	.005	[.963, .983]	.970	NA	NA
	z _{·5}	.971	.008	[.953, .985]	.965	NA	NA
	z _{·6}	.940	.008	[.922, .955]	.941	NA	NA
PV	z _{·7}	.942	.013	[.914, .968]	.944	NA	NA
	z _{·8}	.995	.001	[.991, .997]	1.000	NA	NA
CS	z _{·9}	.992	.002	[.989, .995]	.467	NA	NA
	z _{·10}	.975	.005	[.964, .983]	.398	NA	NA
	z _{·11}	.918	.015	[.883, .945]	.463	NA	NA
CL	z _{·12}	1	0	[1.000, 1.000]	1.000	NA	NA
CC	z _{·13}	.953	.010	[.933, .972]	.934	NA	NA
	z _{·14}	.984	.009	[.962, .996]	1.000	NA	NA

Table 6. The path coefficient estimates, their standard errors (SE), and 95% confidence intervals (CI) in the ACSI model obtained from SFA and JCA-ML.

Path	SFA			JCA-ML		
	Estimate	SE	95% CI	Estimate	SE	95% CI
CE → PQ	.942	.011	[.918, .961]	.936	NA	NA
CE → PV	−.260	.237	[−.722, .218]	−.271	NA	NA
PQ → PV	1.090	.244	[.574, 1.545]	1.149	NA	NA
CE → CS	−.183	.072	[−.335, −.049]	−2.050	NA	NA
PQ → CS	.958	.086	[.814, 1.150]	2.551	NA	NA
PV → CS	.238	.040	[.149, .307]	−.229	NA	NA
CS → CC	−.460	.081	[−.618, −.298]	−1.000	NA	NA
CS → CL	.497	.056	[.377, .601]	−95.610	NA	NA
CC → CL	−.452	.049	[−.551, −.355]	−96.290	NA	NA

Figure 1. The two data generating models used for the simulation studies.**A. Simple model****B. Complex model**

Note: Circles and squares signify latent variables and indicators, respectively. The arrow labeled A denotes the path coefficient excluded in the misspecified model. The values in the parentheses denote correlations.

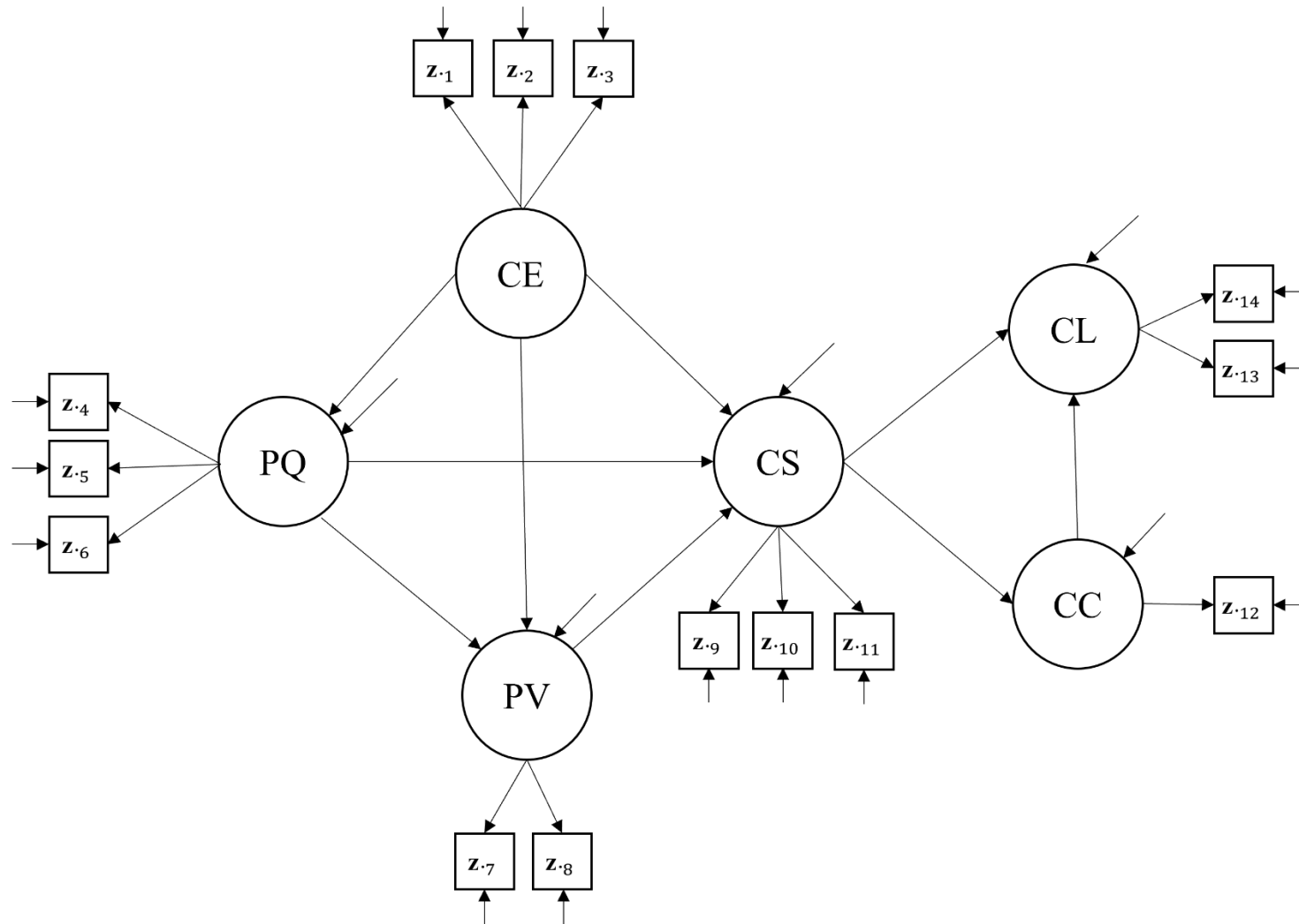
Figure 2. The American customer satisfaction (ACSI) model.

Figure 3. Marginal distributions of the candidate factor scores of customer satisfaction for three companies labeled ID1, ID2, and ID21.

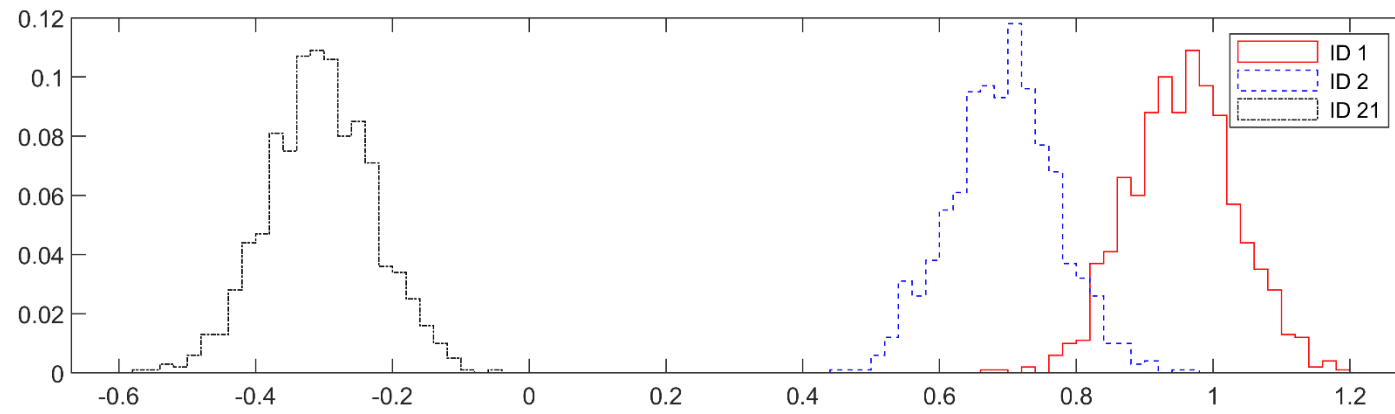
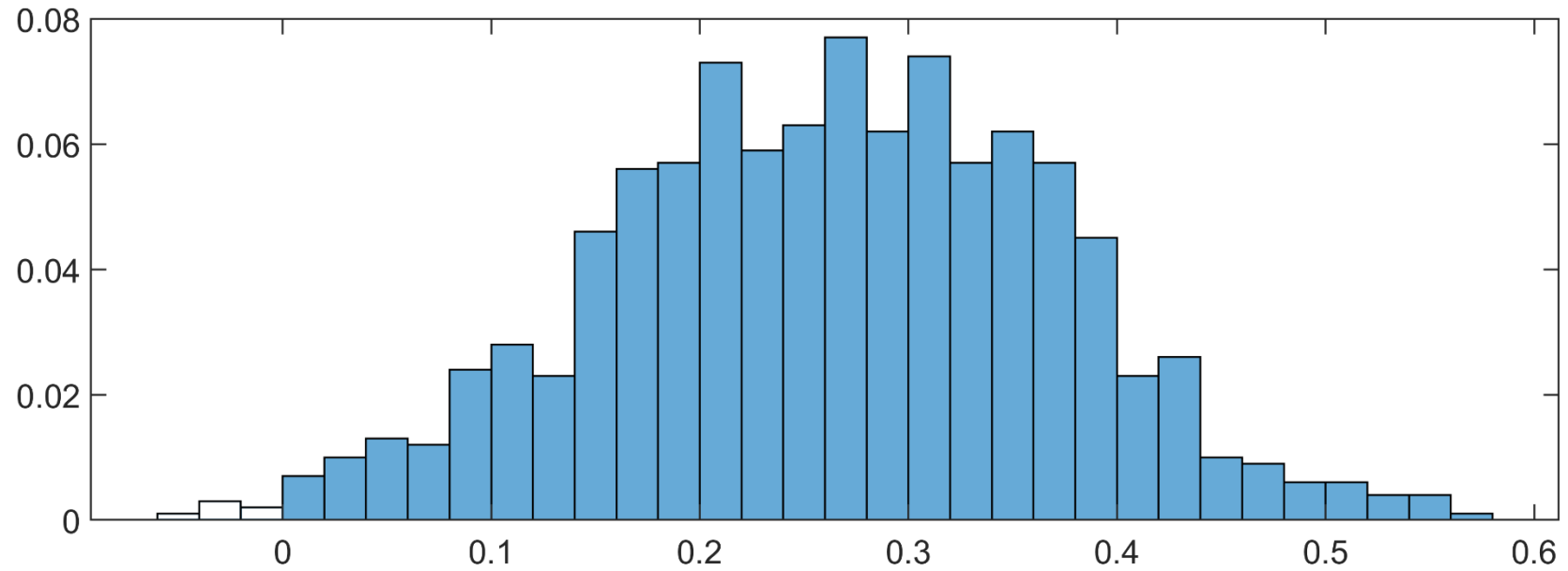


Figure 4. The distribution of the differences in the candidate factor scores of customer satisfaction between two companies ID1 and ID 2.



Note: The colored area represents the probability for ID1 to have a higher customer satisfaction score than ID2.

Supplementary Material

Article title: Structured factor analysis: A data matrix-based alternative approach to structural equation modeling

Authors: Gyeongcheol Cho^{*} and Heungsun Hwang

Journal: Structural Equation Modeling: A Multidisciplinary Journal

List of Appendices

Appendix S1. A full description of the two stages in SFA

Appendix S2. Theorem 1 and its proof.

Appendix S3. Theorem 2 and its proof.

Appendix S4. The proposed ALS algorithm for the first stage of SFA

Appendix S5. Theorem 3 and its proof

Appendix S6. A supplementary procedure for the ALS algorithm

Appendix S7. A non-iterative estimation for the second stage of SFA

Appendix S8. A full description of the candidate factor score distribution

Appendix S9. Theorem 4 and its proof.

Appendix S10. Theorem 5 and its proof.

Appendix S11. The algorithm for estimating the candidate factor score distribution with **W** and **G**.

Appendix S12. Theorem 6 and its proof.

Tables S1 to S8 are provided in the file named *Supplementary Material (Tables).xlsx*.

^{*} Correspondence should be addressed to Gyeongcheol Cho, Department of Psychology, McGill University, 2001 McGill College Avenue, Montreal, QC, Canada, H3A 1G1 (email: gyeongcheol.cho@mail.mcgill.ca)

Appendix S1. A full description of the two stages in SFA

In this Appendix, we provide a full description of the two modeling stages in SFA based on the random matrix theory. In the first stage, researchers are to specify the measurement model that represents the data-generating process of indicators under the assumption that the true latent variable scores are the underlying causes of the indicators' scores. SFA estimates the parameters of the specified measurement model as well as factor scores and allows for statistical tests of the goodness-of-fit of the measurement model. If the measurement model with the factor score estimates may be acceptable, SFA can move on to the second stage. In this stage, researchers are to specify the structural model that represents the score-generating process of latent variables. SFA estimates the parameters of the structural model using the factor scores estimated from the first stage and evaluates the goodness-of-fit of the structural model.

As many of readers would not be familiarized with the theory of a random matrix, we start with briefly explaining the concept of a random matrix $\mathbb{X}$ and its relevant estimators here. Let $\mathbb{x}_n$ denote a set of random variables or a random vector for each individual n ($n = 1, 2, \dots, N$), whose population mean vector and covariance matrix are denoted by $\boldsymbol{\tau}_n$ and $\boldsymbol{\Xi}_n$, respectively (i.e., $\boldsymbol{\tau}_n \equiv E[\mathbb{x}_n]$ and $\boldsymbol{\Xi}_n \equiv E[(\mathbb{x}_n - E[\mathbb{x}_n])(\mathbb{x}_n - E[\mathbb{x}_n])']$). With $\{\mathbb{x}_1, \mathbb{x}_2, \dots, \mathbb{x}_N\}$, we can define a random matrix $\mathbb{X}$ for N individuals as $\mathbb{X} \equiv [\mathbb{x}_1, \mathbb{x}_2, \dots, \mathbb{x}_N]'$. For a random matrix $\mathbb{X}$, SFA employs two estimators of $\mathbb{X}$: *score mean vector* and *score covariance matrix*. The score mean vector and covariance matrix of $\mathbb{X}$ are defined as $mean(\mathbb{X})$ and $cov(\mathbb{X})$, respectively, where $mean(\mathbb{X}) \equiv (N^{-1}\mathbf{1}_N'\mathbb{X})$ and $cov(\mathbb{X}) \equiv N_0^{-1}(\mathbb{X} - \mathbf{1}_N mean(\mathbb{X}))'(\mathbb{X} - \mathbf{1}_N mean(\mathbb{X}))$ for any matrix $\mathbb{X}$ having N rows and $N_0 \equiv N - 1$ (see Appendix A in Bollen, 1989). If $\{\mathbb{x}_1, \mathbb{x}_2, \dots, \mathbb{x}_N\}$ is *i.i.d.*, then $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \dots = \boldsymbol{\tau}_N$ and $\boldsymbol{\Xi}_1 = \boldsymbol{\Xi}_2 = \dots = \boldsymbol{\Xi}_N$, indicating that the population mean vector and covariance matrix of $\mathbb{x}_n$ can be expressed without a subscript as $\boldsymbol{\tau}$ and $\boldsymbol{\Xi}$, respectively. Under this

condition, the expected score mean vector and covariance matrix of $\mathbb{X}$ becomes equivalent to the population mean vector and covariance matrix of $\mathbb{x}_n$, i.e., $E[\text{mean}(\mathbb{X})] = \boldsymbol{\tau}$ and $E[\text{cov}(\mathbb{X})] = \boldsymbol{\Xi}$ (Theorem 1), whose proof is provided in Appendix S2. This theorem suggests that $\text{mean}(\mathbb{X})$ and $\text{cov}(\mathbb{X})$ can serve as unbiased estimators for $\boldsymbol{\tau}$ and $\boldsymbol{\Xi}$, respectively, when the rows of $\mathbb{X}$ are *i.i.d.* random vectors. Also, under the same condition, $\text{mean}(\mathbb{X})$ can be seen as the sample mean vector of $\mathbb{x}_n$, as well as the score mean vector of $\mathbb{X}$, and similarly, $\text{cov}(\mathbb{X})$ can be considered the sample covariance matrix of $\mathbb{x}_n$ and the score covariance matrix of $\mathbb{X}$.

1.1. Stage 1: Modeling the data-generating process for indicators

Model specification

Suppose that we are interested in the levels of P latent variables of a certain group of N individuals. Let $\mathbb{h}_n$ denote a random vector of the n th individual's P latent variables, whose population mean vector and covariance matrix are denoted by $\boldsymbol{\alpha}$ and $\boldsymbol{\Phi}$, respectively, for all n ($n = 1, 2, \dots, N$). We assume that $\{\mathbb{h}_1, \mathbb{h}_2, \dots, \mathbb{h}_N\}$ is *i.i.d.* and every latent variable is standardized (i.e., $\boldsymbol{\alpha} = \mathbf{0}$ and $\text{diag}(\boldsymbol{\Phi}) = \mathbf{1}_P$), where $\mathbf{1}_k$ is a k by 1 vector of ones having N rows and $\text{diag}()$ is an operator that converts an input matrix into a column vector of its diagonal elements. If $\mathbb{h}_n$ is realized as a score vector, denoted by $\mathbf{h}_n$, it means that the n th individual has a specific level of the P latent variables that correspond to $\mathbf{h}_n$. Throughout the Stage 1, SFA assumes that $\mathbb{h}_n$ has been already realized for all n (i.e., $\mathbb{h}_1 = \mathbf{h}_1, \mathbb{h}_2 = \mathbf{h}_2, \dots, \mathbb{h}_N = \mathbf{h}_N$).

Let $\mathbf{H}_{\text{true}} \equiv [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N]'$ denote an N by P matrix that includes all the N individuals' true latent variable scores, from which the sample mean vector and covariance matrix of $\mathbb{h}_n$, denoted by $\boldsymbol{\alpha}_s$ and $\boldsymbol{\Phi}_s$ respectively, can be calculated by $\boldsymbol{\alpha}_s = \text{mean}(\mathbf{H}_{\text{true}})$ and $\boldsymbol{\Phi}_s = \text{cov}(\mathbf{H}_{\text{true}})$. For simplicity, the true latent variable scores in $\mathbf{H}_{\text{true}}$ are assumed to be standardized such that $\boldsymbol{\alpha}_s = \mathbf{0}$

and $\text{diag}(\Phi_s) = \mathbf{1}_P$. This matrix $\mathbf{H}_{true}$ is the target score matrix that SEM researchers eventually would like to measure.

However, researchers could have only a limited access to the true scores of latent variables in an epistemic sense (Borsboom, 2008), suggesting that they cannot have direct measures of $\mathbf{H}_{true}$. Accordingly, researchers typically seek to find J indicators of the P latent variables ($J > P$) and measure $\mathbf{H}_{true}$ indirectly through the N individuals' scores on the J indicators. Let $\mathbb{Z}$ denote an N by J random matrix of J indicators for the N individuals, whose realized value is denoted by $\mathbf{Z}$. This matrix $\mathbf{Z}$ is the data matrix researchers have collected for inferring $\mathbf{H}_{true}$. In Stage 1, SFA formulates a model that explains how this data matrix $\mathbf{Z}$ is generated from $\mathbf{H}_{true}$, which is called the *measurement model*.

As in other SEM approaches, SFA assumes that everyone's indicator scores in $\mathbf{Z}$ are generated from the scores of two factors: *common* and *unique*. The common factor corresponds to a latent variable that affects its multiple indicators simultaneously, whose scores are included in $\mathbf{H}_{true}$. The parameters quantifying the causal effects of each latent variable on its indicators are included in a P by J loading matrix, denoted by Λ , whose (p,j) th entry refers to the p th latent variable's effect on the j th indicator ($p = 1, 2, \dots, P; j = 1, 2, \dots, J$).

On the other hand, a unique factor in SFA corresponds to a random error that affects one indicator uniquely (Bollen, 1989, p. 233). Let $\mathbf{e}_{1\cdot}, \mathbf{e}_{2\cdot}, \dots, \mathbf{e}_{N\cdot}$ denote *i.i.d.* random vectors of J unique factors, where $E[\mathbf{e}_{n\cdot}] = \mathbf{0}$ and $\Theta \equiv E[\mathbf{e}_{n\cdot}\mathbf{e}_{n\cdot}']$ satisfying $\text{diag}(\Lambda'\Phi_s\Lambda + \Theta) = \mathbf{1}_J$ ($n = 1, 2, \dots, N$). Then, a random matrix of J unique factors for the N individuals, denoted by $\mathbb{E}$, can be defined as $\mathbb{E} \equiv [\mathbf{e}_{1\cdot}, \mathbf{e}_{2\cdot}, \dots, \mathbf{e}_{N\cdot}]'$. By Theorem 1, $E[\text{mean}(\mathbb{E})] = \mathbf{0}$ and $E[\text{cov}(\mathbb{E})] = \Theta$. In addition, we assume that $E[\mathbb{E}|\mathbf{H}_{true}] = E[\mathbb{E}]$, indicating that the scores of unique factors are determined

independently of the true scores of latent variables. It also implies that unique factor scores are expected to be uncorrelated with the true scores of latent variables (i.e., $E[\text{cov}(\mathbb{E}, \mathbf{H}_{true})] = \mathbf{0}$), where $\text{cov}(\mathbf{X}_1, \mathbf{X}_2) \equiv N_0^{-1}(\mathbf{X}_1 - \mathbf{1}_N \text{mean}(\mathbf{X}_1))'(\mathbf{X}_2 - \mathbf{1}_N \text{mean}(\mathbf{X}_2))$ for any matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ having N rows. Then, the measurement model can be expressed as

$$\mathbb{Z} = \mathbf{H}_{true}\mathbf{\Lambda} + \mathbb{E}. \quad (\text{A.1})$$

The measurement model (A.1) shows the hypothetical, probabilistic process of obtaining $\mathbf{Z}$ when researchers seek to measure $\mathbf{H}_{true}$ through $\mathbb{Z}$. For instance, suppose that J indicators are items of a self-reported questionnaire for P psychological latent variables. When N individuals respond to the J items without knowing their levels of the P psychological variables precisely, random errors (i.e., $\mathbb{E}$) can involve the individuals' response process to the items, so that $\mathbb{E}$ can be realized as an unknown constant, denoted by $\mathbf{E}_{true}$, and determines the resultant data matrix of measurements $\mathbf{Z}$ as $\mathbf{Z} = \mathbf{H}_{true}\mathbf{\Lambda} + \mathbf{E}_{true}$. Under (A.1), $E[\text{mean}(\mathbb{Z})] = \mathbf{0}$ and $E[\text{cov}(\mathbb{Z})] = \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda} + \mathbf{\Theta}$ (Theorem 2), whose proof is provided in Appendix S3. In the paper, $E[\text{cov}(\mathbb{Z})]$ is denoted by $\mathbf{\Sigma}_s$.

Based on a prior theory, researchers must specify which elements of $\mathbf{\Lambda}$, $\mathbf{\Phi}_s$, and $\mathbf{\Theta}$ in the measurement model (A.1) are non-zero parameters to be estimated and check whether the specified measurement model can be identified. The identification rules for the measurement model are equivalent to those for the confirmatory factor analysis model in JCA (e.g., refer to Bollen 1989, pp. 238–251).

Estimation algorithm

Suppose that the measurement model is correctly specified and $\mathbf{Z}$ is collected from N individuals, suggesting that $\mathbf{Z}$ is generated as $\mathbf{Z} = \mathbf{H}_{true}\mathbf{\Lambda} + \mathbf{E}_{true}$. For simplicity, let $\mathbf{F}_{true} \equiv [\mathbf{H}_{true}, \mathbf{E}_{true}]$, $\mathbf{L} \equiv [\mathbf{\Lambda}; \mathbf{I}_J]$, and $\mathbf{\Delta}_s \equiv \begin{bmatrix} \mathbf{\Phi}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{bmatrix}$, where $\mathbf{I}_J$ is the identity matrix of order J , followed by $\mathbf{Z} = \mathbf{F}_{true}\mathbf{L}$. We

call $\mathbf{F}_{true}$ a matrix of the *true factor scores* including the scores of both common and unique factors for N individuals. In Stage 1, SFA seeks to estimate $\mathbf{L}$, $\mathbf{\Lambda}_s$, and $\mathbf{F}_{true}$ given $\mathbf{Z}$. However, even if we knew the values of $\mathbf{L}$ and $\mathbf{\Lambda}_s$, it would be impossible to obtain the precise value of $\mathbf{F}_{true}$ from $\mathbf{Z}$ because the number of scores in $\mathbf{F}_{true}$ to be estimated (i.e., NT) is greater than the number of observed scores given in the dataset $\mathbf{Z}$ (i.e., NJ), where $T = P + J$ (e.g., Mulaik, 2009; Steiger, 1979). This is the factor score indeterminacy problem (de Leeuw, 2017).

Accordingly, instead of aiming to obtain an unbiased estimate of $\mathbf{F}_{true}$, SFA first contemplates a matrix of the candidate factor score matrix $\mathbf{F} \in \mathbb{R}^{N \times T}$, which satisfies $\mathbf{Z} = \mathbf{F}\mathbf{L}$, $mean(\mathbf{F}) = mean(\mathbf{F}_{true})$, and $cov(\mathbf{F}) = cov(\mathbf{F}_{true})$, and thus can be considered a potential candidate for $\mathbf{F}_{true}$ given $\mathbf{Z}$. To simplify the problem, SFA makes an additional assumption that $\mathbf{F}_{true}$ would be *representative of the population*. A score matrix $\mathbf{X}$, which is considered a realized value of a random matrix $\mathbb{X}$, is said to be representative (of the population) if $\mathbf{X}$ satisfies $mean(\mathbf{X}) = E[mean(\mathbb{X})]$ and $cov(\mathbf{X}) = E[cov(\mathbb{X})]$. The score matrix $\mathbf{F}_{true}$ being representative given $\mathbf{H}_{true}$ in Stage 1 implies that $\mathbf{E}_{true}$ is generated such that $mean(\mathbf{E}_{true}) = \mathbf{0}$, $cov(\mathbf{E}_{true}) = \mathbf{\Theta}$, and $cov(\mathbf{H}_{true}, \mathbf{E}_{true}) = \mathbf{0}$, thereby having $\mathbf{F}_{true}$ satisfy $mean(\mathbf{F}_{true}) = \mathbf{0}$ and $cov(\mathbf{F}_{true}) = \mathbf{\Lambda}_s$. Under this condition, $\mathbf{Z}$ also becomes representative (i.e., $mean(\mathbf{Z}) = \mathbf{0}$ and $cov(\mathbf{Z}) = \mathbf{\Sigma}_s$; see Appendix S3), from which $mean(\mathbf{F}_{true})$ and $cov(\mathbf{F}_{true})$ can be identified as $\mathbf{0}$ and $\mathbf{\Lambda}_s$, respectively.

Assuming that $\mathbf{F}_{true}$ is representative, SFA aims to obtain unbiased estimates of $\mathbf{L}$ and $\mathbf{\Lambda}$ as well as an estimate of $\mathbf{F}$. Specifically, let $\mathbf{Z}_{std}$ denote the standardized counterpart of $\mathbf{Z}$, whose sample mean vector is zero and sample covariance matrix is denoted by $\mathbf{S}$. Let $\hat{\mathbf{Z}}$ denote a matrix of the predicted values of $\mathbf{Z}_{std}$ based on the estimated model. SFA estimates $\mathbf{L}$ and $\mathbf{F}$ by minimizing the following cost function.

$$\begin{aligned}
\rho &= (JN)^{-1} SS(\mathbf{Z}_{std} - \hat{\mathbf{Z}}) \\
&= (JN)^{-1} SS(\mathbf{Z}_{std} - \mathbf{FL}),
\end{aligned} \tag{A.2}$$

subject to $mean(\mathbf{F}) = \mathbf{0}$ and $cov(\mathbf{F}) = \mathbf{\Lambda}_s$, where $SS(\mathbf{X}) \equiv tr(\mathbf{X}'\mathbf{X})$ for any matrix $\mathbf{X}$. The second constraint on $\mathbf{F}$ indicates that some entries of $cov(\mathbf{F})$ must be zeros if their corresponding parameters in $\mathbf{\Lambda}_s$ are fixed to zeros. The value of the cost function (A.2) can be interpreted as the average residual variance or (in-sample) prediction error for the standardized indicator scores. This indicates that SFA seeks to simultaneously estimate the measurement model parameters and a matrix of candidate factor scores in such a way that they maximize explanatory power for the standardized indicator scores. Once $\mathbf{L}$ and $\mathbf{F}$ are estimated, the estimate of $\mathbf{\Lambda}_s$ is obtained by $\hat{\mathbf{\Lambda}}_s = cov(\hat{\mathbf{F}})$.

There is no closed-form solution for minimizing (A.2) subject to the constraints. Thus, we developed an alternating least squares (ALS) algorithm, which divides model parameters into several groups and updates each group alternately with the remaining groups fixed. A detailed description of this algorithm is provided in Appendix S4. After the model parameters are estimated, their standard errors or 95% confidence intervals are calculated for testing their statistical significance. As SFA does not assume any distributional assumption on indicators, it employs a resampling technique, such as the bootstrap method (Efron, 1979, 1982), to obtain those statistics without recourse to a distributional assumption.

If the ALS algorithm minimizes (A.2) given the identified measurement model and the representative $\mathbf{F}_{true}$, it provides unbiased estimates of the measurement model parameters (Theorem 3), as proven in Appendix S5. Also, the proposed ALS algorithm mathematically guarantees the convergence of (A.2) (de Leeuw et al., 1976), which means that the decrease in (A.2) must become smaller than any positive value after some iterations. Moreover, the proposed

algorithm does not result in improper solutions as its cost function (A.2) is built on individual factor scores rather than their covariance matrix. Obviously, a set of individual factor scores cannot have negative variances, a negative-definite covariance matrix, or correlations with $\mathbf{Z}_{std}$ greater than one in absolute value.

Despite its desirable statistical properties, the algorithm may be computationally costly than JCA when the sample size is large as it needs to estimate N individuals' factor scores as well as the model parameters. Thus, we propose a supplementary procedure to alleviate the algorithm's potential computational burden in Appendix S6. This procedure indicates that even if SFA's cost function is defined based on a data matrix of indicators, SFA only needs the sample covariance matrix of the indicators for estimating the model parameters. Thus, albeit researchers only have the sample covariance matrix of indicators in hand, they can still apply SFA if they are interested in estimating the model parameters.

Model evaluation

SFA provides an overall goodness-of-fit index, termed the in-sample prediction error for observed variables (IPE_O), for evaluating the measurement model along with factor score estimates. IPE_O is defined as

$$\text{IPE}_O = SS(\mathbf{Z}_{std} - \hat{\mathbf{F}}\hat{\mathbf{L}})/SS(\mathbf{Z}_{std}), \quad (\text{A.3})$$

where $\hat{\mathbf{F}}$ is the estimate of the candidate factor score matrix and $\hat{\mathbf{L}}$ is the estimate of the loading matrix. This index is equivalent to the value of ρ given $\hat{\mathbf{F}}$ and $\hat{\mathbf{L}}$. The value of $1 - \text{IPE}_O$ can also be interpreted as the average R^2 for the indicators. The IPE_O value is zero if and only if $\hat{\mathbf{L}} = \mathbf{L}$ and $\hat{\mathbf{F}}$ satisfies $\text{mean}(\hat{\mathbf{F}}) = \mathbf{0}$, $\mathbf{Z}_{std} = \hat{\mathbf{F}}\hat{\mathbf{L}}$, and $\text{cov}(\hat{\mathbf{F}}) = \mathbf{\Lambda}_s$, under the condition that the measurement model is identified (see Appendix S5). As $\hat{\mathbf{F}}$ obtained from SFA always satisfies $\text{mean}(\hat{\mathbf{F}}) = \mathbf{0}$, the positive value of IPE_O indicates that $\hat{\mathbf{L}} \neq \mathbf{L}$, $\mathbf{Z}_{std} \neq \hat{\mathbf{F}}\hat{\mathbf{L}}$, or $\text{cov}(\hat{\mathbf{F}}) \neq \mathbf{\Lambda}_s$.

SFA can conduct a statistical test of the null hypothesis that $\hat{\mathbf{L}} = \mathbf{L}$, $\mathbf{Z}_{std} = \hat{\mathbf{F}}\hat{\mathbf{L}}$, and $cov(\hat{\mathbf{F}}) = \Delta_s$ by using the Bollen-Stine (B-S) bootstrap method (Bollen & Stine, 1993). Let $\hat{\Sigma}_s$ denote the estimated implied covariance matrix of the indicators. Let $\mathbf{Z}_{imp}$ denote a modified data matrix generated under the null hypothesis. The modified data matrix $\mathbf{Z}_{imp}$ can be obtained by $\mathbf{Z}_{imp} = \mathbf{Z}_{std}\mathbf{S}^{-1/2}\hat{\Sigma}_s^{1/2}$. The bootstrap method is applied to $\mathbf{Z}_{imp}$ to estimate the sampling distribution of the IPE_O . If the value of IPE_O is greater than a critical or cut-off value, e.g., the $(1 - \alpha)$ th percentile of the estimated sampling distribution, we may reject the null hypothesis. For more detailed information on the B-S bootstrapping method, you may refer to Bollen et al. (1993) and Kim et al. (2014).

Also, SFA offers traditional goodness-of-fit indexes, such as GFI (Jöreskog & Sorbom, 1986) and SRMR (Bentler, 1995), which are computed based on the sample and implied covariance matrices of the indicators. In contrast to IPE_O , these indexes only concentrate on evaluating the estimated measurement model. Lastly, the standard errors or 95% confidence intervals of individual parameter estimates are used to test the statistical significance of the estimates.

1.2. Stage 2 – modeling of the structural model

Model specification

Until now, we have just assumed that the true scores of latent variables (i.e., $\mathbf{H}_{true}$) exist without questioning how those scores are generated. In Stage 2, SFA considers $\mathbf{H}_{true}$ a value of a random matrix $\mathbb{H} \equiv [\mathbb{h}_1, \mathbb{h}_2, \dots, \mathbb{h}_N]'$ and additionally models the stochastic process of generating the scores of $\mathbb{H}$, which is called the *structural model*. We use $\mathbb{h}_p$ to denote the p th column of $\mathbb{H}$,

which corresponds to an N by 1 random vector of the p th latent variables for the N individuals.

By Theorem 1, $\mathbb{H}$ satisfies that $E[\text{mean}(\mathbb{H})] = \mathbf{0}$ and $E[\text{cov}(\mathbb{H})] = \mathbf{\Phi}$.

For simplicity, we assume that the entries of $\mathbb{h}_n$ are arranged such that $\mathbb{h}_n$ can be expressed as $\mathbb{h}_n = [\mathbb{h}_{X,n}, \mathbb{h}_{Y,n}]$, where $\mathbb{h}_{X,n}$ and $\mathbb{h}_{Y,n}$ denote a random vector of P_x exogenous latent variables and that of P_y endogenous latent variables, respectively ($n = 1, 2, \dots, N$). By the assumption, $\mathbf{\Phi}$ also can be expressed as $\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_{XX} & \mathbf{\Phi}_{XY} \\ \mathbf{\Phi}_{XY}' & \mathbf{\Phi}_{YY} \end{bmatrix}$, where $\mathbf{\Phi}_{XX} \equiv E[\mathbb{h}_{X,n} \mathbb{h}_{X,n}']$, $\mathbf{\Phi}_{YY} \equiv E[\mathbb{h}_{Y,n} \mathbb{h}_{Y,n}']$, and $\mathbf{\Phi}_{XY} \equiv E[\mathbb{h}_{X,n} \mathbb{h}_{Y,n}']$. Let $\mathbb{H}_X \equiv [\mathbb{h}_{X,1}, \mathbb{h}_{X,2}, \dots, \mathbb{h}_{X,N}]'$ and $\mathbb{H}_Y \equiv [\mathbb{h}_{Y,1}, \mathbb{h}_{Y,2}, \dots, \mathbb{h}_{Y,N}]'$, whose realized values are denoted by $\mathbf{H}_{X,true}$ and $\mathbf{H}_{Y,true}$, respectively. With $\mathbb{H}_X$ and $\mathbb{H}_Y$, $\mathbb{H}$ can be re-expressed as $\mathbb{H} = [\mathbb{H}_X, \mathbb{H}_Y]$, whose realized value is denoted by $\mathbf{H}_{true}$.

Let $\{\mathbb{q}_1, \mathbb{q}_2, \dots, \mathbb{q}_N\}$ denote a set of *i.i.d.* random vectors of P_y structural errors, where $\mathbb{q}_n$ is for the n th individual ($n = 1, 2, \dots, N$), $E[\mathbb{q}_n] = \mathbf{0}$, and $\mathbf{\Psi} \equiv E[\mathbb{q}_n \mathbb{q}_n']$. Let $\mathbb{Q} \equiv [\mathbb{q}_1, \mathbb{q}_2, \dots, \mathbb{q}_N]'$, whose realized value is denoted by $\mathbf{Q}_{true}$. By Theorem 1, $E[\text{mean}(\mathbb{Q})] = \mathbf{0}$ and $E[\text{cov}(\mathbb{Q})] = \mathbf{\Psi}$. We assume that $E[\mathbb{Q}|\mathbb{H}_X] = E[\mathbb{Q}]$, indicating that the scores of structural errors are determined independently of the scores of exogenous latent variables. It also implies that the scores of structural errors are expected to be uncorrelated with the scores of exogenous latent variables (i.e., $E[\text{cov}(\mathbb{Q}, \mathbb{H}_X)] = \mathbf{0}$). Let $\mathbf{B}_X$ and $\mathbf{B}_Y$ denote matrices of path coefficients that quantify the causal effects of $\mathbb{h}_{X,n}$ on $\mathbb{h}_{Y,n}$ and those between endogenous variables in $\mathbb{h}_{Y,n}$, respectively. We assume that $\text{diag}(\mathbf{B}_Y) = \mathbf{0}$ and $(\mathbf{I}_{P_y} - \mathbf{B}_Y)$ is invertible. Then, the structural model can be written as

$$\mathbb{H}_Y = \mathbb{H}_X \mathbf{B}_X + \mathbb{H}_Y \mathbf{B}_Y + \mathbb{Q} \quad (\text{A.4})$$

From (A.4), $\mathbb{H}_Y$ can be expressed as a function of $\mathbb{H}_X$ and $\mathbb{Q}$ (i.e., $\mathbb{H}_Y = (\mathbb{H}_X \mathbf{B}_X + \mathbb{Q})(\mathbf{I}_{P_y} - \mathbf{B}_Y)^{-1}$).

The structural model (A.4) describes how $\mathbf{H}_{true}$ are generated. For N individuals, the random matrices $\mathbb{H}_X$ and $\mathbb{Q}$ in (A.4) are initially realized as $\mathbf{H}_{X,true}$ and $\mathbf{Q}_{true}$, respectively, which determine the value of $\mathbb{H}_Y$ as $\mathbf{H}_{Y,true} = (\mathbf{H}_{X,true}\mathbf{B}_X + \mathbf{Q}_{true})(\mathbf{I}_{Py} - \mathbf{B}_Y)^{-1}$. Then, $\mathbf{H}_{true}$ is determined as $[\mathbf{H}_{X,true}, \mathbf{H}_{Y,true}]$. If $\mathbf{H}_{true}$ is representative, then $\mathbf{H}_{true}$ satisfies $mean(\mathbf{H}_{true}) = \mathbf{0}$ and $cov(\mathbf{H}_{true}) = \mathbf{\Phi}$.

Based on a prior theory, researchers must predetermine which elements of $\mathbf{B}_X$, $\mathbf{B}_Y$, and $\mathbf{\Psi}$ in the structural model (A.4) are non-zero parameters to be estimated and check whether the specified structural model can be identified. The rules for the identification are the same as those used for the path analysis model in JCA, which can be found in Bollen (1989, pp. 88–104) or Dijkstra (2017).

Parameter estimation

Let $\hat{\mathbf{H}}_X$ and $\hat{\mathbf{H}}_Y$ denote matrices of candidate factor score estimates for the exogenous and endogenous latent variables obtained from Stage 1. In Stage 2, SFA estimates the structural model parameters (i.e., $\mathbf{B}_X$, $\mathbf{B}_Y$, and $\mathbf{\Psi}$) while treating $[\hat{\mathbf{H}}_X, \hat{\mathbf{H}}_Y]$ as if they were the data of latent variables, assuming that $\hat{\mathbf{\Phi}}_s = \mathbf{\Phi}$. It utilizes a limited-information estimator, which successively applies ordinary least squares (OLS) or two-stage least squares (2SLS) to each equation of an endogenous latent variable (Lance et al., 1988). The proposed estimator draws on 2SLS if endogeneity occurs in the equation, and on OLS otherwise. Although this estimator can be less efficient than a full-information estimator, such as feasible generalized least squares (FGLS) or three-stage least squares (3SLS), it can be more robust to model misspecification (Wooldridge, 2010, pp. 252–254). Appendix S7 provides a detailed description of the estimator. Once all the structural model parameters are estimated, their standard errors or confidence intervals are

estimated based on a set of the estimates of latent variables' covariance matrix obtained from the bootstrap samples.

Model evaluation

SFA provides a goodness-of-fit index, termed the in-sample prediction error for latent variables (IPE_L), for evaluating the structural model. The IPE_L is defined as

$$IPE_L = SS(\hat{\mathbf{H}}_Y - (\hat{\mathbf{H}}_X \hat{\mathbf{B}}_X + \hat{\mathbf{H}}_Y \hat{\mathbf{B}}_Y)) / SS(\hat{\mathbf{H}}_Y). \quad (\text{A.5})$$

This index represents the average residual variance for all endogenous latent variables unexplained by the fitted structural model. The value of $1 - IPE_L$ is equivalent to the average R^2 for the endogenous latent variables. SFA can also provide GFI and SRMR for the structural model, which are calculated based on the discrepancy between the estimated and implied covariance matrices of the latent variables.

Appendix S2. Theorem 1 and its proof.

Theorem 1. Let $\mathbb{X} \equiv [\mathbb{x}_1, \mathbb{x}_2, \dots, \mathbb{x}_N]'$ denote a random matrix of N rows, where $\mathbb{x}_n$ is a random vector corresponding to the n th row of $\mathbb{X}$ ($n = 1, 2, \dots, N$). If $\{\mathbb{x}_1, \mathbb{x}_2, \dots, \mathbb{x}_N\}$ is *i.i.d.*, $\boldsymbol{\tau} \equiv E[\mathbb{x}_n]$, and $\boldsymbol{\Xi} \equiv E[(\mathbb{x}_n - E[\mathbb{x}_n])(\mathbb{x}_n - E[\mathbb{x}_n])']$, then $E[\text{mean}(\mathbb{X})] = \boldsymbol{\tau}$ and $E[\text{cov}(\mathbb{X})] = \boldsymbol{\Xi}$.

Proof. The expected value of $\text{mean}(\mathbb{X})$ can be expressed as $E[\text{mean}(\mathbb{X})] = E[(N^{-1}\mathbf{1}_N'\mathbb{X})] = N^{-1}\mathbf{1}_N'E[\mathbb{X}] = N^{-1}\mathbf{1}_N'E[[\mathbb{x}_1, \mathbb{x}_2, \dots, \mathbb{x}_N]'] = N^{-1}\mathbf{1}_N'[E[\mathbb{x}_1], E[\mathbb{x}_2], \dots, E[\mathbb{x}_N]]' = N^{-1}\mathbf{1}_N'\mathbf{1}_N\boldsymbol{\tau} = \boldsymbol{\tau}$,

indicating that $E[\text{mean}(\mathbb{X})] = \boldsymbol{\tau}$. Also, the expected value of $\text{cov}(\mathbb{X})$ can be expressed as $E[\text{cov}(\mathbb{X})]$

$$= E[N_0^{-1}(\mathbb{X} - \mathbf{1}_N\text{mean}(\mathbb{X}))(\mathbb{X} - \mathbf{1}_N\text{mean}(\mathbb{X}))'] = N_0^{-1}(E[\mathbb{X}'\mathbb{X}] - E[\mathbb{X}'\mathbf{1}_N\text{mean}(\mathbb{X})] -$$

$$E[\text{mean}(\mathbb{X})'\mathbf{1}_N'\mathbb{X}]) + E[\text{mean}(\mathbb{X})'\mathbf{1}_N'\mathbf{1}_N\text{mean}(\mathbb{X})]) = N_0^{-1}(E[\mathbb{X}'\mathbb{X}] - N(E[\text{mean}(\mathbb{X})'\text{mean}(\mathbb{X})])) =$$

$$N_0^{-1}(\sum_{i=1}^N E[\mathbb{x}_i\mathbb{x}_i'] - N^{-1}E[(\sum_{i=1}^N \mathbb{x}_i)(\sum_{i=1}^N \mathbb{x}_i)']) = N_0^{-1}(\sum_{i=1}^N E[\mathbb{x}_i\mathbb{x}_i'] - N^{-1}(\sum_{i=1}^N E[\mathbb{x}_i\mathbb{x}_i'] +$$

$$\sum_{i \neq k} E[\mathbb{x}_i\mathbb{x}_k'])) = N_0^{-1}(N_0 E[\mathbb{x}_n\mathbb{x}_n'] - N^{-1}\sum_{i \neq k} (E[(\mathbb{x}_i - E[\mathbb{x}_i])(\mathbb{x}_i - E[\mathbb{x}_k])'] + E[\mathbb{x}_i]E[\mathbb{x}_k]')) = N_0^{-1}$$

$$(N_0(\boldsymbol{\Xi} + \boldsymbol{\tau}\boldsymbol{\tau}') - N_0\boldsymbol{\tau}\boldsymbol{\tau}') = N_0^{-1}N_0\boldsymbol{\Xi} = \boldsymbol{\Xi}. \text{ Q.E.D.}$$

Appendix S3. Theorem 2 and its proof.

Theorem 2. $E[\text{mean}(\mathbb{Z})] = \mathbf{0}$ and $E[\text{cov}(\mathbb{Z})] = \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda} + \mathbf{\Theta}$ under (A.1).

Proof. The expected value of $\text{mean}(\mathbb{Z})$ can be expressed as $E[\text{mean}(\mathbb{Z})] = E[(N^{-1}\mathbf{1}_N'\mathbb{Z})] = E[(N^{-1}\mathbf{1}_N'(\mathbf{H}_{true}\mathbf{\Lambda} + \mathbb{E}))] = E[N^{-1}\mathbf{1}_N'\mathbb{E}] = E[\text{mean}(\mathbb{E})] = \mathbf{0}$, indicating that $E[\text{mean}(\mathbb{Z})] = \mathbf{0}$. Also, the expected value of $\text{cov}(\mathbb{Z})$ can be expressed as $E[N_0^{-1}(\mathbb{Z} - \text{mean}(\mathbb{Z}))'(\mathbb{Z} - \text{mean}(\mathbb{Z}))] = E[N_0^{-1}(\mathbf{H}_{true}\mathbf{\Lambda} + \mathbb{E} - \text{mean}(\mathbb{E}))'(\mathbf{H}_{true}\mathbf{\Lambda} + \mathbb{E} - \text{mean}(\mathbb{E}))] = N_0^{-1}(\mathbf{H}_{true}\mathbf{\Lambda})'(\mathbf{H}_{true}\mathbf{\Lambda}) + (\mathbf{H}_{true}\mathbf{\Lambda})'E[\mathbb{E} - \text{mean}(\mathbb{E})] + E[(\mathbb{E} - \text{mean}(\mathbb{E}))'](\mathbf{H}_{true}\mathbf{\Lambda}) + E[(\mathbb{E} - \text{mean}(\mathbb{E}))'(\mathbb{E} - \text{mean}(\mathbb{E}))] = \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda} + \mathbf{\Theta}$. Q.E.D.

Appendix S4. The proposed ALS algorithm for the first stage of SFA

Let us reparametrize $\mathbf{F} = \mathbf{U}\mathbf{V}_f$, where $\mathbf{U}$ is an N by T matrix satisfying $\mathbf{1}_N'\mathbf{U} = \mathbf{0}$ and $N_0^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_T$, and $\mathbf{V}_f$ is a T by T matrix satisfying $\mathbf{V}_f'\mathbf{V}_f = \Delta_s$. Then, (A.2) can be re-expressed as

$$\rho = (JN)^{-1}SS(\mathbf{Z}_{std} - \mathbf{U}\mathbf{V}_f\mathbf{L}), \quad (\text{A.6})$$

subject to $\mathbf{1}_N'\mathbf{U} = \mathbf{0}$, and $N_0^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_T$.

The ALS algorithm begins by assigning initial values to $\mathbf{V}_f$ and $\mathbf{L}$. The ALS algorithm alternately updates each of $\mathbf{U}$, $\mathbf{V}_f$, and $\mathbf{L}$ with the others fixed until no substantial decrease in (A.6) occurs between two consecutive iterations. One can utilize one of the existing SEM techniques to obtain the initial values of $\mathbf{L}$ and Δ_s , from which $\mathbf{V}_f$ can be obtained by the Cholesky decomposition of Δ_s . Otherwise, SFA draws on the following procedure to obtain the initial values of $\mathbf{L}$ and $\mathbf{V}_f$. Let $\mathbf{U}_\eta$ and $\mathbf{U}_\epsilon$ denote the matrix of the first P columns of $\mathbf{U}$ and that of the last J columns of $\mathbf{U}$, respectively. Let $\mathbf{V}_\eta$ denote a matrix of the first P by P diagonal block of $\mathbf{V}_f$ satisfying $\mathbf{V}_\eta'\mathbf{V}_\eta = \Phi_s$ and $\mathbf{V}_\epsilon$ denote a matrix of the last J by J diagonal block of $\mathbf{V}_f$ satisfying $\mathbf{V}_\epsilon'\mathbf{V}_\epsilon = \Theta$. The algorithm initially sets $\mathbf{V}_\eta$ to be $\mathbf{I}_P$ and all the non-zero entries of the loading matrix Λ to be ones. Then, it calculates a residual vector for each indicator by regressing each of the indicators on all the other indicators except for ones whose unique factors are correlated. Then, it standardizes the residual vectors and uses each of them for the initial value for each column of $\mathbf{U}_\epsilon$. The initial value of $\mathbf{V}_\epsilon$ is obtained by $\mathbf{V}_\epsilon = N_0^{-1}\mathbf{U}_\epsilon'\mathbf{Z}_{std}$, implying that the initial value of $\mathbf{V}_\epsilon$ is equivalent to the correlations between $\mathbf{Z}_{std}$ and $\mathbf{U}$. By subtracting $\mathbf{U}_\epsilon\mathbf{V}_\epsilon$ from $\mathbf{Z}_{std}$, the algorithm obtains a part of the data matrix that cannot be explained by the initial value of unique factors, denoted by $\mathbf{Z}_\eta$. Then, the initial value of Λ is rescaled such that $\text{diag}(\Lambda'\Phi_s\Lambda) =$

$N_0^{-1}diag(\mathbf{Z}_\eta'\mathbf{Z}_\eta)$. Lastly, the initial values of $\mathbf{V}_f$ and $\mathbf{L}$ are obtained by $\mathbf{V}_f = \begin{bmatrix} \mathbf{V}_\eta & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\epsilon \end{bmatrix}$ and $\mathbf{L} = [\mathbf{A}; \mathbf{I}_J]$, respectively.

Given the initial values of $\mathbf{V}_f$ and $\mathbf{L}$, the ALS algorithm repeats the following three steps at each iteration.

Step 1: Update $\mathbf{U}$ for fixed $\mathbf{V}_f$ and $\mathbf{L}$. As $\mathbf{Z}_{std}$ can be decomposed as

$$N_0^{-1/2}\mathbf{K}_z\mathbf{\Omega}_z\mathbf{O}_z', \quad (\text{A.7})$$

where $\mathbf{K}_z$ is an N by J matrix satisfying $\mathbf{1}_N'\mathbf{K}_z = \mathbf{0}$ and $\mathbf{K}_z'\mathbf{K}_z = \mathbf{I}_J$, $\mathbf{\Omega}_z^2$ is a J by J diagonal matrix of the eigenvalues of $\mathbf{S}$, and $\mathbf{O}_z$ is a J by J orthogonal matrix of the eigenvector of $\mathbf{S}$, (A.6) can be re-expressed as

$$\begin{aligned} \rho &= c_0((tr(\mathbf{S}) + tr(\mathbf{L}'\mathbf{\Delta}_s\mathbf{L})) - N_0^{-1}tr(2\mathbf{Z}_{std}'\mathbf{U}\mathbf{V}_f\mathbf{L})) \\ &= c_1 - 2c_0tr(N_0^{-1/2}\mathbf{U}'\mathbf{K}_z\mathbf{A}), \end{aligned} \quad (\text{A.8})$$

where $c_0 \equiv (JN)^{-1}N_0$, $c_1 \equiv c_0tr(\mathbf{S} + \mathbf{L}'\mathbf{\Delta}_s\mathbf{L})$, and $\mathbf{A} \equiv \mathbf{\Omega}_z\mathbf{O}_z'\mathbf{L}'\mathbf{V}_f'$. The rank of $\mathbf{A}$ is J because $rank(\mathbf{A}) = rank(\mathbf{\Omega}_z\mathbf{O}_z'\mathbf{L}'\mathbf{V}_f') = rank(\mathbf{L}') = rank([\mathbf{A}', \mathbf{I}_J]) = rank(\mathbf{I}_J) = J$. Matrix $\mathbf{A}$ can be decomposed into

$$\mathbf{A} = \mathbf{R}_A[\chi \quad \mathbf{0}'] \begin{bmatrix} \mathbf{Q}' \\ \mathbf{Q}_\perp' \end{bmatrix}, \quad (\text{A.9})$$

where $\mathbf{R}_A$ is a J by J orthogonal matrix of the left singular vectors of $\mathbf{A}$, $\mathbf{Q}$ is a T by J matrix of the right singular vectors of $\mathbf{A}$, $\mathbf{Q}_\perp$ is a T by P matrix of orthonormal columns satisfying $\mathbf{Q}'\mathbf{Q}_\perp = \mathbf{0}$, and χ is a J by J diagonal matrix of the non-zero singular values of $\mathbf{A}$. Then, $\mathbf{K}_z\mathbf{A}$ can be expressed as

$$\mathbf{K}_z\mathbf{A} = [\mathbf{R} \quad \mathbf{R}_\perp] \begin{bmatrix} \chi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Q}' \\ \mathbf{Q}_\perp' \end{bmatrix}, \quad (\text{A.10})$$

where $\mathbf{R} = \mathbf{K}_z \mathbf{R}_A$ and $\mathbf{R}_\perp$ is an N by $N - J - 1$ matrix of orthonormal columns satisfying $(\mathbf{K}_z \mathbf{R}_A)' \mathbf{R}_\perp = \mathbf{0}$ and $\mathbf{1}_N' \mathbf{R}_\perp = \mathbf{0}$. Given $\mathbf{K}_z \mathbf{A}$, $\mathbf{U}$ that minimizes (A.8) can be obtained by

$$\mathbf{U} = N_0^{-1/2}(\mathbf{R}\mathbf{Q}' + \mathbf{R}_\perp \mathbf{C}\mathbf{Q}_\perp'), \quad (\text{A.11})$$

where $\mathbf{C}$ is a matrix satisfying $\mathbf{C}'\mathbf{C} = \mathbf{I}_P$ (e.g., $[\mathbf{I}_{N-J-1}, \mathbf{0}_{(N-J-1) \times (P)}]$), at which point $\text{tr}(\mathbf{U}'\mathbf{K}_z \mathbf{A})$ is equivalent to $\text{tr}(\boldsymbol{\chi})$ (refer to de Leeuw (2017) for the proof). The column space of $\mathbf{R}_\perp$ is an orthogonal basis of the null space of $[\mathbf{K}_z \mathbf{R}_A, \mathbf{1}_N]'$, which can be obtained through the QR decomposition of $[\mathbf{K}_z \mathbf{R}_A, \mathbf{1}_N]$.

This step is largely based on de Leeuw (2017)'s algorithm. The difference is that we impose an additional constraint $\mathbf{1}_N' \mathbf{U} = \mathbf{0}$. Without the imposition of this constraint, $\mathbf{F} = \mathbf{U}\mathbf{V}_f$ cannot satisfy $\mathbf{1}_N' \mathbf{F} = \mathbf{0}$ and thus, $N_0^{-1} \mathbf{F}'\mathbf{F}$ cannot be interpreted as a covariance matrix anymore.

Step 2: Update $\mathbf{V}_f$ for fixed $\mathbf{U}$ and $\mathbf{L}$. Then, (A.8) can be re-written as

$$\begin{aligned} \rho &= N^{-1}N_0 + c_0(SS(\mathbf{V}_f \mathbf{L}) - 2\text{tr}(\mathbf{O}_z \boldsymbol{\Omega}_z \mathbf{R}_A \mathbf{Q}'(\mathbf{V}_f \mathbf{L}))) \\ &= c_2 + c_0(SS(\mathbf{Q}\mathbf{R}_A' \boldsymbol{\Omega}_z \mathbf{O}_z') - 2\text{tr}(\mathbf{O}_z \boldsymbol{\Omega}_z \mathbf{R}_A \mathbf{Q}'(\mathbf{V}_f \mathbf{L})) + SS(\mathbf{V}_f \mathbf{L})) \\ &= c_2 + c_0 SS(\mathbf{Q}\mathbf{R}_A' \boldsymbol{\Omega}_z \mathbf{O}_z' - \mathbf{V}_f \mathbf{L}) \\ &= c_2 + c_0 SS(\text{vec}(\mathbf{Q}\mathbf{R}_A' \boldsymbol{\Omega}_z \mathbf{O}_z') - (\mathbf{L}' \otimes \mathbf{I}_T) \text{vec}(\mathbf{V}_f)), \end{aligned} \quad (\text{A.12})$$

where $c_2 \equiv N^{-1}N_0 + c_0 SS(\mathbf{Q}\mathbf{R}_A' \boldsymbol{\Omega}_z \mathbf{O}_z')$. Let $\mathbf{v}_2$ denote a vector of non-zero elements in $\text{vec}(\mathbf{V}_f)$ and $\boldsymbol{\Gamma}_1$ denote a matrix of the columns of $(\mathbf{L}' \otimes \mathbf{I})$ corresponding to $\mathbf{v}_2$. Then, the least-squares estimate of $\mathbf{v}_2$ is obtained by

$$(\boldsymbol{\Gamma}_1' \boldsymbol{\Gamma}_1)^{-1} \boldsymbol{\Gamma}_1' \text{vec}(\mathbf{Q}\mathbf{R}_A' \boldsymbol{\Omega}_z \mathbf{O}_z'). \quad (\text{A.13})$$

The non-zero elements of $\mathbf{V}_f$ are updated by $\mathbf{v}_2$ if the smallest eigenvalue of $\mathbf{V}_f' \mathbf{V}_f$ remains positive. Note that the smallest eigenvalue of $\mathbf{V}_f' \mathbf{V}_f$ is considered positive only if its calculated value is greater than a small positive number (e.g., 10^{-10}), because the numerical calculation of

the smallest eigenvalue can be susceptible to a numerical error when its actual value is close to zero (e.g., Perla et al., 2020, Chapter 23).

Step 3: Update $\mathbf{L}$ for fixed $\mathbf{U}$ and $\mathbf{V}_f$. Let $\mathbf{Q}_\eta$ denote a P by J matrix of the first P rows of $\mathbf{Q}$ and $\mathbf{Q}_\epsilon$ denote a J by J matrix of the last J rows of $\mathbf{Q}$. Then, (A.6) can be re-expressed as

$$\begin{aligned}
 \rho &= c_2 + c_0 SS \left(\begin{bmatrix} \mathbf{Q}_\eta \\ \mathbf{Q}_\epsilon \end{bmatrix} \mathbf{R}_A' \mathbf{\Omega}_Z \mathbf{O}_Z' - \begin{bmatrix} \mathbf{V}_\eta & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\epsilon \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{I}_J \end{bmatrix} \right) \\
 &= c_2 + c_0 SS \left(\begin{bmatrix} \mathbf{Q}_\eta \mathbf{R}_A' \mathbf{\Omega}_Z \mathbf{O}_Z' \\ \mathbf{Q}_\epsilon \mathbf{R}_A' \mathbf{\Omega}_Z \mathbf{O}_Z' - \mathbf{V}_\epsilon \end{bmatrix} - \begin{bmatrix} \mathbf{I}_P \\ \mathbf{0} \end{bmatrix} \mathbf{V}_\eta \mathbf{\Lambda} \right) \\
 &= c_3 + c_0 SS(\mathbf{Q}_\eta \mathbf{R}_A' \mathbf{\Omega}_Z \mathbf{O}_Z' - \mathbf{V}_\eta \mathbf{\Lambda}) \\
 &= c_3 + c_0 SS(\text{vec}(\mathbf{Q}_\eta \mathbf{R}_A' \mathbf{\Omega}_Z \mathbf{O}_Z') - (\mathbf{I}_J \otimes \mathbf{V}_\eta) \text{vec}(\mathbf{\Lambda})),
 \end{aligned} \tag{A.14}$$

where $c_3 \equiv c_2 + c_0 SS(\mathbf{Q}_\epsilon \mathbf{R}_A' \mathbf{\Omega}_Z \mathbf{O}_Z' - \mathbf{V}_\epsilon)$. Let $\mathbf{v}_3$ denote a vector of non-zero elements in $\text{vec}(\mathbf{\Lambda})$

and $\mathbf{\Gamma}_2$ is a matrix of the columns of $(\mathbf{I}_J \otimes \mathbf{V}_\eta)$ corresponding to the non-zero elements in $\text{vec}(\mathbf{\Lambda})$.

Then, the least-squares estimate of $\mathbf{v}_3$ is obtained by

$$(\mathbf{\Gamma}_2' \mathbf{\Gamma}_2)^{-1} \mathbf{\Gamma}_2' \text{vec}(\mathbf{Q}_\eta \mathbf{R}_A' \mathbf{\Omega}_Z \mathbf{O}_Z'). \tag{A.15}$$

Then, the non-zero elements of $\mathbf{\Lambda}$ are updated by $\mathbf{v}_3$, from which $\mathbf{L}$ is reconstructed by $\hat{\mathbf{L}} = [\mathbf{\Lambda}; \mathbf{I}_J]$.

Upon convergence, $\mathbf{V}_\eta$ is obtained from the first P by P diagonal block of $\mathbf{V}_f$ and $\mathbf{V}_\epsilon$ is obtained from the last J by J diagonal block of $\mathbf{V}_f$. Let $\mathbf{\Gamma}_3$ denote a P by P diagonal matrix whose diagonal entries are the elements of $\text{diag}(\mathbf{V}_\eta' \mathbf{V}_\eta)^{0.5}$. Then, $\mathbf{V}_\eta$ is rescaled by post-multiplying $\mathbf{V}_\eta$ by $\mathbf{\Gamma}_3^{-1}$ such that the rescaled value of $\mathbf{V}_\eta$ can satisfy $\text{diag}(\mathbf{V}_\eta' \mathbf{V}_\eta) = \mathbf{1}_P$. Then, the least-squares estimate of $\mathbf{\Phi}_s$ is updated by $\hat{\mathbf{\Phi}}_s = \mathbf{V}_\eta' \mathbf{V}_\eta$. The rescaling of $\mathbf{V}_\eta$ is needed to ensure that the variances of latent variables are equal to one. To avoid the change of ρ by the rescaling, $\hat{\mathbf{\Lambda}}$ is also adjusted by pre-multiplying it by $\mathbf{\Gamma}_3$. Also, let $\mathbf{\Gamma}_4$ denote a J by J diagonal matrix whose diagonal entries are the elements of $\text{diag}(\hat{\mathbf{\Lambda}} \hat{\mathbf{\Phi}}_s \hat{\mathbf{\Lambda}} + \mathbf{V}_\epsilon' \mathbf{V}_\epsilon)^{0.5}$. Then, $\mathbf{V}_\epsilon$ is updated one more time by post-

multiplying it by Γ_4^{-1} such that the updated value of $\mathbf{V}_\epsilon$ can satisfy $\text{diag}(\hat{\Lambda}\hat{\Phi}_s\hat{\Lambda} + \mathbf{V}_\epsilon'\mathbf{V}_\epsilon) = \mathbf{1}_J$.

This makes the variances of $\hat{\mathbf{Z}}$ equal to one (i.e., $\text{diag}(\text{cov}(\hat{\mathbf{Z}})) = \mathbf{1}_J$). Then, the least-squares

estimate of Θ is obtained by $\hat{\Theta} = \mathbf{V}_\epsilon'\mathbf{V}_\epsilon$. Lastly, $\mathbf{V}_f$ is updated by $\mathbf{V}_f = \begin{bmatrix} \mathbf{V}_\eta & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\epsilon \end{bmatrix}$ and the least-

squares estimate of $\mathbf{F}$ is obtained by $\hat{\mathbf{F}} = \mathbf{U}\mathbf{V}_f$.

Appendix S5. Theorem 3 and its proof

Theorem 3. Suppose that the measurement model is identified and $\mathbf{F}_{true}$ satisfies $mean(\mathbf{F}_{true}) = \mathbf{0}$, and $cov(\mathbf{F}_{true}) = \Delta_s$. If the ALS algorithm minimize (A.2) given $\mathbf{Z}$, it provides unbiased estimates of the measurement model parameters.

Proof. Suppose that $\mathbf{F}_{true}$ satisfies $mean(\mathbf{F}_{true}) = \mathbf{0}$ and $cov(\mathbf{F}_{true}) = \Delta_s$. Then, $\mathbf{Z}$ satisfies that $mean(\mathbf{Z}) = \mathbf{0}$, $cov(\mathbf{Z}) = \Sigma_s$, and $\mathbf{Z} = \mathbf{Z}_{std}$, as $mean(\mathbf{Z}) = \mathbf{1}_N' \mathbf{Z} = (\mathbf{1}_N' \mathbf{F}_{true}) \mathbf{L} = mean(\mathbf{F}_{true}) \mathbf{L} = \mathbf{0} \mathbf{L} = \mathbf{0}$ and $cov(\mathbf{Z}) = N_0^{-1} \mathbf{Z}' \mathbf{Z} = \mathbf{L}' (N_0^{-1} \mathbf{F}_{true}' \mathbf{F}_{true}) \mathbf{L} = \mathbf{L}' \Delta_s \mathbf{L} = \Sigma_s$. Also, given the identified measurement model, $\hat{\mathbf{L}} = \mathbf{L}$ and $\hat{\Delta}_s = \Delta_s$ if and only if $\hat{\Sigma}_s = \Sigma_s$, where $\hat{\mathbf{L}}$ and $\hat{\Delta}_s$ are estimates of $\mathbf{L}$ and Δ , respectively, and $\hat{\Sigma}_s$ denotes the model-implied covariance matrix of indicators. We need to show that if and only if $\hat{\mathbf{L}} = \mathbf{L}$ and $\hat{\Delta}_s = \Delta_s$, (A.2) can be zero for some $\mathbf{F}$ subject to $\mathbf{1}_N' \mathbf{F} = \mathbf{0}$ and $\hat{\Delta}_s = N_0^{-1} \mathbf{F}' \mathbf{F}$.

Suppose that (A.2) is zero with a value of $\mathbf{F}$ satisfying its constraints, indicating that $\mathbf{Z} = \mathbf{F} \hat{\mathbf{L}}$, $mean(\mathbf{F}) = \mathbf{0}$, and $\hat{\Delta}_s = cov(\mathbf{F})$. Then, $\Sigma_s = N_0^{-1} \mathbf{Z}' \mathbf{Z} = N_0^{-1} (\mathbf{F} \hat{\mathbf{L}})' (\mathbf{F} \hat{\mathbf{L}}) = \hat{\mathbf{L}}' \hat{\Delta}_s \hat{\mathbf{L}} = \hat{\Sigma}_s$. Then, $\hat{\mathbf{L}} = \mathbf{L}$ and $\hat{\Delta}_s = \Delta_s$ by the identification assumption. Conversely, suppose that $\hat{\mathbf{L}} = \mathbf{L}$ and $\hat{\Delta}_s = \Delta_s$. We only have to show that there exists $\mathbf{F}$ that makes (A.2) zero subject to $mean(\mathbf{F}) = \mathbf{0}$ and $\hat{\Delta}_s = cov(\mathbf{F})$. By the assumption, $\Sigma_s = N_0^{-1} \mathbf{Z}' \mathbf{Z} = \mathbf{L}' \Delta_s \mathbf{L}$. As $\mathbf{A}$ in (A.8) satisfies $\mathbf{A} = \mathbf{\Omega}_z \mathbf{O}_z' \mathbf{L}' \mathbf{V}_f' = \mathbf{R}_A \chi \mathbf{Q}'$, then $\mathbf{A} \mathbf{A}' = \mathbf{\Omega}_z \mathbf{O}_z' \mathbf{L}' \Delta_s \mathbf{L} \mathbf{O}_z \mathbf{\Omega}_z = \mathbf{\Omega}_z \mathbf{O}_z' (\mathbf{O}_z \mathbf{\Omega}_z \mathbf{\Omega}_z' \mathbf{O}_z) \mathbf{O}_z \mathbf{\Omega}_z = \mathbf{\Omega}_z^4$, implying that $\chi = \mathbf{\Omega}_z^2$. When plugging (A.11) into (A.6) given $\hat{\mathbf{L}} = \mathbf{L}$ and $\hat{\Delta}_s = \Delta_s$, (A.2) can be re-expressed as

$$\rho = c_0((tr(\Sigma_s) + tr(\mathbf{L}' \Delta_s \mathbf{L}) - 2tr(\chi))). \quad (\text{A.16})$$

As the trace of Σ_s is equivalent to the sum of its eigenvalues, (A.16) can be written as

$$\rho = c_0((tr(\mathbf{\Omega}_z^2) + tr(\mathbf{\Omega}_z^2) - 2tr(\mathbf{\Omega}_z^2))) = 0. \quad (\text{A.17})$$

Q.E.D.

Appendix S6. A supplementary procedure for the ALS algorithm

As the cost function (A.2) is defined on $\mathbf{Z}_{std}$, the proposed ALS algorithm can be computationally less efficient as the sample size becomes large. This issue can be circumvented by replacing $\mathbf{Z}_{std}$ with $\mathbf{Z}_a$, where $\mathbf{Z}_a$ is a $(T + 1)$ by P data matrix with the same mean and covariance matrix as those of $\mathbf{Z}_{std}$ (i.e., $\mathbf{1}_{(T+1)}'\mathbf{Z}_a = \mathbf{0}$ and $T^{-1}\mathbf{Z}_a'\mathbf{Z}_a = \mathbf{S}$). This replacement makes the ALS algorithm computationally efficient, particularly when N is far larger than $(T + 1)$, because the number of scores to be estimated in $\mathbf{F}$ will be much smaller when $\mathbf{Z}_a$ is used than when $\mathbf{Z}_{std}$ is used (i.e., $(T+1)T < NT$). Let $\mathbf{\Gamma}_5 \equiv (\mathbf{I}_{(T+1)} - (T+1)^{-1}\mathbf{1}_{(T+1)}\mathbf{1}_{(T+1)}')[\mathbf{I}_{(J+1)}, \mathbf{0}_{(J+1) \times P}]'$ and $\mathbf{K}_a \equiv \mathbf{\Gamma}_5(\mathbf{\Gamma}_5'\mathbf{\Gamma}_5)^{-1/2}$. The data matrix $\mathbf{Z}_a$ can be obtained by

$$T^{1/2}\mathbf{K}_a\mathbf{\Omega}_z\mathbf{O}_z'. \quad (\text{A.18})$$

Once the ALS algorithm converges based on $\mathbf{Z}_a$, we can calculate the estimate of candidate factor scores for original observations by extracting $\mathbf{V}_f$ from $\hat{\mathbf{\Lambda}}_s$ and applying Step 1 of the ALS algorithm given $\mathbf{V}_f$ and $\hat{\mathbf{L}}$. We may also utilize this procedure when only $\mathbf{S}$ is available for parameter estimation. Specifically, we can obtain $\mathbf{\Omega}_z$ and $\mathbf{O}_z$ through the eigenvalue decomposition of $\mathbf{S}$, based on which we can obtain $\mathbf{Z}_a$ and apply the ALS algorithm to $\mathbf{Z}_a$ for estimating model parameters.

We can substitute $\mathbf{Z}_a$ for $\mathbf{Z}_{std}$ in the proposed algorithm because the minimum point of (A.2) does not depend on a change in $\mathbf{Z}_{std}$ unless such a change alters the sample mean vector and covariance matrix. To prove this, we need to show that the value of (A.2) is unaffected by a change in $\mathbf{K}_z$ in (A.7). As $\mathbf{K}_z$ is involved only in Step 1 of the algorithm, we only need to show that $\mathbf{K}_z$ does not affect the minimum point of (A.8). The minimum point of (A.8) is equivalent to $c_1 - 2c_0\text{tr}(\chi)$ (de Leeuw, 2017), where $c_0 \equiv (JN)^{-1}N_0$, $c_1 = c_0\text{tr}(\mathbf{S} + \mathbf{L}'\mathbf{\Lambda}_s\mathbf{L})$, and χ is a matrix of singular values of $\mathbf{A} = \mathbf{\Omega}_z\mathbf{O}_z'\mathbf{L}'\mathbf{V}_f$. Thus, $\mathbf{K}_z$ does not determine the minimum point of (A.8).

SFA uses $(T + 1)$ for the row size of $\mathbf{Z}_a$, because $(T + 1)$ is the minimum sample size required for the proposed ALS algorithm. In Step 1 of the ALS algorithm, the least-squares estimate of $\mathbf{U}$ is obtained by (A.11), where $\mathbf{C}$ in (A.11) must satisfy $\mathbf{C}'\mathbf{C} = \mathbf{I}_P$. As $\mathbf{C}$ is an $(N - J - 1)$ by P matrix, this condition can hold only when $N - J - 1 \geq P$. Thus, N must satisfy $N \geq P + J + 1 = T + 1$.

Appendix S7. A non-iterative estimation for the second stage of SFA

The structural model (A.4) can be re-expressed in the so-called *reduced-form equation* as follows.

$$\mathbb{H}_Y = \mathbb{H}_X \mathbf{\Pi} + \mathbb{Q}(\mathbf{I}_{P_Y} - \mathbf{B}_Y)^{-1}, \quad (\text{A.19})$$

where $\mathbf{\Pi} \equiv \mathbf{B}_X(\mathbf{I}_{P_Y} - \mathbf{B}_Y)^{-1}$. The matrix $\mathbf{\Pi}$ is called a matrix of reduced-form parameters. Since $E[N_0^{-1}(\mathbb{H}_X - \text{mean}(\mathbb{H}_X))'(\mathbb{Q} - \text{mean}(\mathbb{Q}))] = E[N_0^{-1}(\mathbb{H}_X - \text{mean}(\mathbb{H}_X))'(\mathbb{H}_Y(\mathbf{I}_{P_Y} - \mathbf{B}_Y) - \mathbb{H}_X \mathbf{B}_X - \text{mean}(\mathbb{H}_Y(\mathbf{I}_{P_Y} - \mathbf{B}_Y) - \mathbb{H}_X \mathbf{B}_X))] = E[N_0^{-1}(\mathbb{H}_X - \text{mean}(\mathbb{H}_X))'(\mathbb{H}_Y - \text{mean}(\mathbb{H}_Y))](\mathbf{I}_{P_Y} - \mathbf{B}_Y) - E[N_0^{-1}(\mathbb{H}_X - \text{mean}(\mathbb{H}_X))'(\mathbb{H}_X - \text{mean}(\mathbb{H}_X))]\mathbf{B}_X = \mathbf{\Phi}_{XY}(\mathbf{I}_{P_Y} - \mathbf{B}_Y) - \mathbf{\Phi}_{XX}\mathbf{B}_X = \mathbf{0}$, $\mathbf{\Pi}$ can be re-expressed as $\mathbf{\Pi} = \mathbf{\Phi}_{XX}^{-1}\mathbf{\Phi}_{XY}$, implying that $\mathbf{\Pi}$ can be identified from $\mathbf{\Phi}$. However, as researchers' interest is typically in the structural parameters $\mathbf{B}_X$ and $\mathbf{B}_Y$, not in the reduced one ($\mathbf{\Pi}$), researchers must examine whether all the non-zero elements of $\mathbf{B}_X$ and $\mathbf{B}_Y$ can be uniquely determined by $\mathbf{\Pi}$ or can be identified. When such a relationship between $\mathbf{\Pi}$ and a set of $\mathbf{B}_X$ and $\mathbf{B}_Y$ holds, $\mathbf{\Psi}$ can also be uniquely expressed as $\mathbf{\Psi} = [-\mathbf{B}_X; (\mathbf{I}_{P_Y} - \mathbf{B}_Y)]'\mathbf{\Phi}[-\mathbf{B}_X; (\mathbf{I}_{P_Y} - \mathbf{B}_Y)]$.

SFA's estimation algorithm starts with estimating $\mathbf{\Pi}$ given $\hat{\mathbf{H}}_X$ and $\hat{\mathbf{H}}_Y$ obtained from Stage 1. The least-squares estimate of $\mathbf{\Pi}$ is obtained by

$$\hat{\mathbf{\Pi}} = (\hat{\mathbf{H}}_X' \hat{\mathbf{H}}_X)^{-1} (\hat{\mathbf{H}}_X' \hat{\mathbf{H}}_{XY}). \quad (\text{A.20})$$

If no endogeneity occurs in the equation, SFA obtains OLS estimates for the path coefficients.

Otherwise, SFA seeks to estimate the path coefficients via 2SLS (e.g., Dijkstra, 1989). Let $\hat{\mathbf{\Pi}}_0 \equiv [\mathbf{I}_{P_X}, \hat{\mathbf{\Pi}}]$ and $\hat{\mathbf{B}} \equiv [\hat{\mathbf{B}}_X; \hat{\mathbf{B}}_Y]$. Let $\hat{\boldsymbol{\pi}}_q$ and $\hat{\mathbf{b}}_q$ denote the q th column vectors of $\hat{\mathbf{\Pi}}$ and $\hat{\mathbf{B}}$, respectively ($q = 1, 2, \dots, P_Y$). Let $\mathbf{u}_3$ denote a vector of non-zero elements in $\hat{\mathbf{b}}_q$. Let $\mathbf{\Gamma}_6$ denote a matrix of the columns of $\hat{\mathbf{\Pi}}_0$ corresponding to the non-zero elements in $\hat{\mathbf{b}}_q$. Then, the least-squares estimate of $\mathbf{u}_3$ is obtained by

$$\mathbf{u}_3 = (\mathbf{\Gamma}_6' \mathbf{\Gamma}_6)^{-1} \mathbf{\Gamma}_6' \hat{\boldsymbol{\pi}}_q. \quad (\text{A.21})$$

Then, the non-zero elements in the q th columns of $\hat{\mathbf{B}}_X$ and $\hat{\mathbf{B}}_Y$ are updated by $\mathbf{u}_3$. After all the path coefficients are estimated, the least-squares estimate of $\boldsymbol{\Psi}$ is obtained by $\hat{\boldsymbol{\Psi}} = [-\hat{\mathbf{B}}_X; \mathbf{I}_{p_y} - \hat{\mathbf{B}}_Y]' \hat{\boldsymbol{\Phi}}_s [-\hat{\mathbf{B}}_X; \mathbf{I}_{p_y} - \hat{\mathbf{B}}_Y]$.

Appendix S8. A full description of the candidate factor score distribution

As stated in Section The Proposed Method, SFA obtains an estimate of a matrix of the candidate factor scores $\mathbf{F}$ and uses this estimate, denoted by $\hat{\mathbf{F}}$, to estimate the parameters of the measurement and structural models. However, SFA does not recommend using $\hat{\mathbf{F}}$ as a point estimate of $\mathbf{F}_{true}$ as there exist an infinite number of $\mathbf{F}$ s owing to the factor score indeterminacy problem so that there is no possibility that a single estimate of $\mathbf{F}$ is equivalent to $\mathbf{F}_{true}$. Instead, SFA contemplates the set of all possible N by T matrices of candidate factor scores, denoted by $\mathcal{F}_{N,T}$, assuming $\mathbf{F}_{true}$ is representative given $\mathbf{Z}$ (i.e., $\mathbf{Z} = \mathbf{F}_{true}\mathbf{L}$, $mean(\mathbf{F}_{true}) = \mathbf{0}$, and $cov(\mathbf{F}_{true}) = \Delta_s$). Under this assumption, $\mathcal{F}_{N,T}$ can be expressed as $\mathcal{F}_{N,T} \equiv \{\mathbf{F} \in \mathbb{R}^{N \times T} \mid \mathbf{Z} = \mathbf{F}\mathbf{L}, mean(\mathbf{F}) = \mathbf{0}, \text{ and } cov(\mathbf{F}) = \Delta_s\}$, suggesting that if a factor score matrix does not have the sample mean vector of $\mathbf{0}$ or the sample covariance matrix of Δ_s , one does not necessarily consider the score matrix as a potential candidate for $\mathbf{F}_{true}$. SFA derives the probability distribution for all the possible values in $\mathcal{F}_{N,T}$ and uses it to infer $\mathbf{F}_{true}$ *a posteriori* given $\mathbf{Z}$.

Derivation of the candidate factor score distribution and its five properties

Let $\mathbb{F}$ denote a random matrix that takes on a value $\mathbf{F}$ in $\mathcal{F}_{N,T}$. As each element in $\mathcal{F}_{N,T}$ is called a matrix of candidate factor scores, we call $\mathbb{F}$ *a random matrix of candidate factors* and its probability distribution *the candidate factor score distribution*. Then, each column of $\mathbb{F}$, denoted by $\mathbf{f}_n$, correspond to a random vector of each individual's candidate factors, whose population covariance matrix is denoted by $\mathbf{G}_n$ ($n = 1, 2, \dots, N$).

The derivation of the candidate factor score distribution starts from the following theorem:

$\mathbf{F}_{true}$ in $\mathcal{F}_{N,T}$ can be re-expressed as $\mathbf{F}_{true} = \mathbf{Z}\mathbf{W} + \mathbf{M}_{true}$, where $\mathbf{W} \equiv \Sigma_s^{-1}[\mathbf{\Lambda}'\Phi_s, \Theta]$, $\mathbf{G} \equiv [\mathbf{I}_P, -\mathbf{\Lambda}]'(\Phi_s - \Phi_s\mathbf{\Lambda}\Sigma_s^{-1}\mathbf{\Lambda}'\Phi_s)[\mathbf{I}_P, -\mathbf{\Lambda}]$, and $\mathbf{M}_{true}$ is an N by T matrix satisfying $mean(\mathbf{M}_{true}) = \mathbf{0}$,

$\text{cov}(\mathbf{M}_{true}) = \mathbf{G}$, and $\text{cov}(\mathbf{Z}, \mathbf{M}_{true}) = \mathbf{0}$ (Theorem 4; Guttman, 1955), whose proof is provided in Appendix S9. The major implication of this theorem is that $\mathbf{F}_{true}$ in $\mathcal{F}_{N,T}$ can be decomposed into two parts: the deterministic part (i.e., $\mathbf{Z}\mathbf{W}$) that can be inferred from $\mathbf{Z}$ and its random part (i.e., $\mathbf{M}_{true}$) that can never be known from $\mathbf{Z}$. Unless $\mathbf{F}_{true}$ is given, any element of $\mathbf{M}_{true}$ cannot be estimated from $\mathbf{Z}$ because the two score matrices are uncorrelated (i.e., $\text{cov}(\mathbf{Z}, \mathbf{M}_{true}) = \mathbf{0}$). As $\mathbf{F}_{true}$ can be seen as an arbitrary value in $\mathcal{F}_{N,T}$ given $\mathbf{Z}$, the theorem also implies that any $\mathbf{F}$ in $\mathcal{F}_{N,T}$ can be re-expressed as $\mathbf{F} = \mathbf{Z}\mathbf{W} + \mathbf{M}$, where $\mathbf{M}$ is an N by T matrix in $\mathcal{M}_{N,T} \equiv \{\mathbf{M} \in \mathbb{R}^{N \times T} \mid \text{mean}(\mathbf{M}) = \mathbf{0}, \text{cov}(\mathbf{M}) = \mathbf{G}, \text{and } \text{cov}(\mathbf{Z}, \mathbf{M}) = \mathbf{0}\}$. The value of $\mathbf{F}$ depends solely on $\mathbf{M}$ given $\mathbf{Z}\mathbf{W}$, so that if $\mathbf{M} = \mathbf{M}_{true}$, then $\mathbf{F} = \mathbf{F}_{true}$. It means that if we know the value of $\mathbf{M}$ in $\mathcal{M}_{N,T}$ being equivalent to $\mathbf{M}_{true}$ given $\mathbf{Z}$, we can know the value of $\mathbf{F}_{true}$. However, as the value of $\mathbf{M}_{true}$ cannot be inferred from $\mathbf{Z}$ because they are uncorrelated, one cannot assume that a value in $\mathcal{M}_{N,T}$ is more likely to be $\mathbf{M}_{true}$ than the others. Accordingly, SFA considers $\mathbf{M}$ a realized value of a random matrix $\mathbb{M}$ that follows the uniform distribution on $\mathcal{M}_{N,T}$, which means that every $\mathbf{M}$ in $\mathcal{M}_{N,T}$ is assumed to be equally likely to be true. With the random matrix $\mathbb{M}$, we can re-express $\mathbb{F}$ as

$$\mathbb{F} = \mathbf{Z}\mathbf{W} + \mathbb{M}. \quad (\text{A.22})$$

The equation (A.22) implies that the candidate factor score distribution (i.e., the probability distribution of $\mathbb{F}$) has the following five properties: (a) $\mathbb{F}$ follows the uniform distribution on $\mathcal{F}_{N,T}$; (b) $E[\text{cov}(\mathbf{F}_{true} - \mathbb{F})] = 2\mathbf{G}$; (c) $E[\mathbb{F}] = \mathbf{Z}\mathbf{W}$; (d) $E[\text{cov}(\mathbf{F}_{true} - E[\mathbb{F}])] = \mathbf{G}$; (e) $N\theta^{-1}\sum_{i=1}^N \mathbf{G}_i = \mathbf{G}$ (Theorem 5). The proof of these properties is provided in Appendix S10.

Property (a) means that every matrix of candidate factor scores in $\mathcal{F}_{N,T}$ is equally likely to be $\mathbf{F}_{true}$, suggesting that researchers cannot find a value that is more likely to be true than the

others in $\mathcal{F}_{N,T}$. Property (b) indicates that if one randomly chooses a value from $\mathcal{F}_{N,T}$ and uses it as a measurement of $\mathbf{F}_{true}$, the standard deviations of measurement error are expected to be $diag(2\mathbf{G})^{\circ 1/2}$, where $^{\circ 1/2}$ is a *Hadamard root* operator that takes elementwise square roots of the base vector. Property (c) shows that the center of the candidate factor score distribution $E[\mathbf{F}]$ can be expressed as a matrix of weighted composite scores of indicators $\mathbf{Z}\mathbf{W}$. Accordingly, $\mathbf{Z}\mathbf{W}$ is called a *matrix of expected candidate factor scores* in SFA. Property (d) means that if one uses $\mathbf{Z}\mathbf{W}$ as a measurement of $\mathbf{F}_{true}$, the standard deviations of measurement error are expected to be $diag(\mathbf{G})^{\circ 1/2}$, which are two times smaller than when one randomly chooses a value from $\mathcal{F}_{N,T}$ as a measurement of $\mathbf{F}_{true}$. It suggests that it is more reasonable to use $\mathbf{Z}\mathbf{W}$ as a measurement of $\mathbf{F}_{true}$ than any random value in $\mathcal{F}_{N,T}$. Lastly, Property (e) indicates that each $\mathbf{f}_{n\cdot}$ has the covariance matrix that approximates $\mathbf{G}$ with $E[\mathbf{f}_{n\cdot}] = (\mathbf{z}_{n\cdot}'\mathbf{W})'$, where $\mathbf{z}_{n\cdot}$ is a vector of the n th individual's indicator scores in $\mathbf{Z}$.

Note that researchers need to be cautious not to interpret the equation (A.22) as a data-generating process for $\mathbf{F}_{true}$. Unlike (A.1), the candidate factor score distribution is built on $\mathbf{Z}$ after $\mathbf{Z}$ has been generated as $\mathbf{Z} = \mathbf{F}_{true}\mathbf{L}$, for inferring $\mathbf{F}_{true}$ *a posteriori* given $\mathbf{Z}$ in a probabilistic manner.

Estimation of the candidate factor distribution and its statistics

SFA estimates the candidate factor score distribution by minimizing the same cost function (A.2) in a least-squares sense. Specifically, let $\hat{\mathcal{F}}_{N,T}$ denote an estimate of $\mathcal{F}_{N,T}$, whose elements can be expressed with $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Lambda}}_s$ obtained from the first stage. Then, SFA randomly samples a prescribed number of $\hat{\mathbf{F}}$ from $\hat{\mathcal{F}}_{N,T}$ and uses the set of the $\hat{\mathbf{F}}$ values as an estimate of the candidate factor score distribution. A detailed explanation of this procedure is provided in Appendix S11. Under

the assumption that $\hat{\mathbf{L}} = \mathbf{L}$, $\hat{\mathbf{\Delta}}_s = \mathbf{\Delta}_s$, and $\mathbf{F}_{true}$ is representative given $\mathbf{Z}$, the estimate of the candidate factor score distribution with $\hat{\mathbf{W}}$ and $\hat{\mathbf{G}}$ approximates the uniform distribution on $\mathcal{F}_{N,T}$ with $\mathbf{W}$ and $\mathbf{G}$ (Theorem 6), whose proof is provided in Appendix S12.

Once it estimates the candidate factor score distribution, SFA estimates $\mathbf{ZW}$ and $diag(\mathbf{G})^{o1/2}$ at default as well and uses their estimates as a single measurement of $\mathbf{F}_{true}$ and its standard deviations of measurement error, respectively. This choice makes sense in that $\mathbf{ZW}$ is the only part of $\mathbf{F}_{true}$ that can be inferred from $\mathbf{Z}$, as shown in Theorem 4, and can be seen as the best linear predictor for $\mathbf{F}_{true}$ given $\mathbf{Z}$ (e.g., Bartholomew, 1981), as $\mathbf{W}$ in $\mathbf{ZW}$ is equivalent to the one that can be obtained by regressing $\mathbf{F}_{true}$ on $\mathbf{Z}_{std}$ (Thurstone, 1934). In SFA, the standard deviations of measurement error are called *the standard errors of measurement* (Leong & Huang, 2016), which refer to the standard amount of error that is expected to occur when the measurement is used to quantify the true amount of a particular quantity. In Section Simulation Studies, we will investigate whether $diag(\hat{\mathbf{G}})^{o1/2}$ obtained from SFA approximates the actual standard errors of measurement (i.e., $diag(cov(\mathbf{F}_{true} - \mathbf{Z}_{std}\hat{\mathbf{W}}))^{o1/2}$) when $\mathbf{Z}_{std}\hat{\mathbf{W}}$ is used as a measurement of $\mathbf{F}_{true}$.

Moreover, SFA derives 95% candidate factor score intervals (95% CI) for each latent variable per individual from the estimate of candidate factor score distribution. As its name denotes, this interval shows a range of 95% factor scores that are possibly equivalent to $\mathbf{F}_{true}$ among the entire candidate factor scores. We will conduct a simulation study to examine that the coverage probability of the 95% CIs becomes close to .95 on average as the sample size becomes larger. In addition, in Section Empirical Illustration, we will demonstrate how to utilize the estimated candidate factor score distribution for various probabilistic inferences on $\mathbf{F}_{true}$.

Appendix S9. Theorem 4 and its proof.

Theorem 4. $\mathbf{F}_{true}$ in $\mathcal{F}_{N,T}$ can be re-expressed as $\mathbf{F}_{true} = \mathbf{Z}\mathbf{W} + \mathbf{M}_{true}$, where $\mathbf{W} \equiv \Sigma_s^{-1}[\Lambda'\Phi_s, \Theta]$, $\mathbf{G} \equiv [\mathbf{I}_P, -\Lambda]'(\Phi_s - \Phi_s\Lambda\Sigma_s^{-1}\Lambda'\Phi_s)[\mathbf{I}_P, -\Lambda]$, and $\mathbf{M}_{true}$ is an N by T matrix satisfying $mean(\mathbf{M}_{true}) = \mathbf{0}$, $cov(\mathbf{M}_{true}) = \mathbf{G}$, and $cov(\mathbf{Z}, \mathbf{M}_{true}) = \mathbf{0}$.

Proof. Suppose that $\mathbf{W} = \Sigma_s^{-1}[\Lambda'\Phi_s, \Theta]$. Let us define $\mathbf{M}_{true}$ as $\mathbf{M}_{true} = \mathbf{F}_{true} - \mathbf{Z}\mathbf{W}$. We need to show that $\mathbf{M}_{true}$ satisfies (a) $mean(\mathbf{M}_{true}) = \mathbf{0}$, (b) $cov(\mathbf{Z}, \mathbf{M}_{true}) = \mathbf{0}$, and (c) $cov(\mathbf{M}_{true}) = [\mathbf{I}_P, -\Lambda]'(\Phi_s - \Phi_s\Lambda\Sigma_s^{-1}\Lambda'\Phi_s)[\mathbf{I}_P, -\Lambda]$. For (a), $mean(\mathbf{M}_{true}) = N^{-1}\mathbf{1}_N'\mathbf{M}_{true} = N^{-1}\mathbf{1}_N'(\mathbf{F}_{true} - \mathbf{Z}\mathbf{W}) = N^{-1}\mathbf{1}_N'\mathbf{F}_{true} - N^{-1}(\mathbf{1}_N'\mathbf{Z})\mathbf{W} = \mathbf{0}$. For (b), $cov(\mathbf{Z}, \mathbf{M}_{true}) = N_0^{-1}\mathbf{Z}'\mathbf{M}_{true} = N_0^{-1}\mathbf{Z}'(\mathbf{F}_{true} - \mathbf{Z}\mathbf{W}) = \mathbf{L}'\Delta_s - [\Lambda'\Phi_s, \Theta] = \mathbf{0}$. For (c), let $\mathbf{M}_1$ and $\mathbf{M}_2$ denote a matrix of the first P columns of $\mathbf{M}_{true}$ and that of the last J columns of $\mathbf{M}_{true}$, respectively. Let $\mathbf{G}_1 \equiv N_0^{-1}\mathbf{M}_1'\mathbf{M}_1$, $\mathbf{G}_2 \equiv N_0^{-1}\mathbf{M}_2'\mathbf{M}_2$, and $\mathbf{G}_{12} \equiv N_0^{-1}\mathbf{M}_1'\mathbf{M}_2$. Then, the data matrix $\mathbf{Z}$ is equivalent to $\mathbf{F}_{true}\mathbf{L} = (\mathbf{Z}\mathbf{W} + \mathbf{M}_{true})\mathbf{L} = \mathbf{Z}\Sigma_s^{-1}[\Lambda'\Phi_s\Lambda + \Theta] + [\mathbf{M}_1\Lambda + \mathbf{M}_2] = \mathbf{Z} + [\mathbf{M}_1\Lambda + \mathbf{M}_2] = \mathbf{Z}$, indicating that $\mathbf{M}_2 = -\mathbf{M}_1\Lambda$, $\mathbf{G}_2 = \Lambda'\mathbf{G}_1\Lambda$, and $\mathbf{G}_{12} = -\mathbf{G}_1\Lambda$. Let $\mathbf{H}_{true}$ denote a matrix of the first P columns of $\mathbf{F}_{true}$ and $\mathbf{W}_1 \equiv \Sigma_s^{-1}\Lambda'\Phi_s$. Then, $\mathbf{G}_1$ can be expressed as $\mathbf{G}_1 = N_0^{-1}\mathbf{M}_1'\mathbf{M}_1 = (\mathbf{H}_{true} - \mathbf{Z}\mathbf{W}_1)'(\mathbf{H}_{true} - \mathbf{Z}\mathbf{W}_1) = \Phi_s - \Phi_s\Lambda\mathbf{W}_1 - \mathbf{W}_1'\Lambda'\Phi_s + \mathbf{W}_1'\Sigma_s\mathbf{W}_1 = \Phi_s - \Phi_s\Lambda\Sigma_s^{-1}\Lambda'\Phi_s - \Phi_s\Lambda\Sigma_s^{-1}\Lambda'\Phi_s + \Phi_s\Lambda\Sigma_s^{-1}\Sigma_s\Sigma_s^{-1}\Lambda'\Phi_s = \Phi_s - \Phi_s\Lambda\Sigma_s^{-1}\Lambda'\Phi_s$. As $\mathbf{G}_2 = \Lambda'\mathbf{G}_1\Lambda$ and $\mathbf{G}_{12} = -\mathbf{G}_1\Lambda$, $N_0^{-1}\mathbf{M}_{true}'\mathbf{M}_{true} = [\mathbf{I}_P, -\Lambda]'\mathbf{G}_1[\mathbf{I}_P, -\Lambda] = \mathbf{G}$. Q.E.D.

Appendix S10. Theorem 5 and its proof.

Theorem 5. The probability distribution of $\mathbb{F}$ has the following five properties: (a) $\mathbb{F}$ follows the uniform distribution on $\mathcal{F}_{N,T}$; (b) $E[\text{cov}(\mathbf{F}_{true} - \mathbb{F})] = 2\mathbf{G}$; (c) $E[\mathbb{F}] = \mathbf{Z}\mathbf{W}$; (d) $E[\text{cov}(\mathbf{F}_{true} - E[\mathbb{F}])] = \mathbf{G}$; (e) $N_0^{-1} \sum_{i=1}^N \mathbf{G}_i = \mathbf{G}$.

Proof. For (a), the random matrix $\mathbb{M}$ in (A.22) follows the uniform distribution on $\mathcal{M}_{N,T}$. Also, the equation (A.22) is a one-to-one mapping of $\mathcal{M}_{N,T}$ onto $\mathcal{F}_{N,T}$, which means that every $\mathbf{M}$ in $\mathcal{M}_{N,T}$ corresponds uniquely to each and every $\mathbf{F}$ in $\mathcal{F}_{N,T}$ though (A.22). Thus, the probability distribution of $\mathbb{F}$ is the uniform distribution on $\mathcal{F}_{N,T}$. For (b), $E[\text{cov}(\mathbf{F}_{true} - \mathbb{F})] = E[\text{cov}(\mathbf{M}_{true} - \mathbb{M})] = E[N_0^{-1}((\mathbf{M}_{true} - \mathbb{M}) - \mathbf{1}_N \text{mean}(\mathbf{M}_{true} - \mathbb{M}))'((\mathbf{M}_{true} - \mathbb{M}) - \mathbf{1}_N \text{mean}(\mathbf{M}_{true} - \mathbb{M}))] = E[N_0^{-1}(\mathbf{M}_{true} - \mathbb{M})'(\mathbf{M}_{true} - \mathbb{M})] = E[N_0^{-1}\mathbf{M}_{true}'\mathbf{M}_{true}] - E[N_0^{-1}\mathbb{M}'\mathbf{M}_{true}] - E[N_0^{-1}\mathbf{M}_{true}'\mathbb{M}] + E[N_0^{-1}\mathbb{M}'\mathbb{M}] = E[N_0^{-1}\mathbf{M}_{true}'\mathbf{M}_{true}] + E[N_0^{-1}\mathbb{M}'\mathbb{M}] = 2\mathbf{G}$. For (c), $E[\mathbb{F}] = E[\mathbf{Z}\mathbf{W} + \mathbb{M}] = \mathbf{Z}\mathbf{W} + E[\mathbb{M}]$. For any $\mathbf{M}$ in $\mathcal{M}_{N,T}$, $-\mathbf{M}$ is in $\mathcal{M}_{N,T}$, and both of $\mathbf{M}$ and $-\mathbf{M}$ are equally likely to be true, implying that $E[\mathbb{M}] = \mathbf{0}$. Thus, $E[\mathbb{F}] = \mathbf{Z}\mathbf{W}$. For (d), $E[\text{cov}(\mathbf{F}_{true} - E[\mathbb{F}])] = E[\text{cov}(\mathbf{F}_{true} - \mathbf{Z}\mathbf{W})] = E[\text{cov}(\mathbf{M}_{true})] = E[\mathbf{G}] = \mathbf{G}$. For (e), let $\mathbf{z}_n$ and $\mathbb{m}_n$ denote the n th column of $\mathbf{Z}'$ and that of $\mathbb{M}'$, respectively. Then, $\mathbf{G}$ can be re-expressed as $\mathbf{G} = N_0^{-1}\mathbb{M}'\mathbb{M} = E[N_0^{-1}\mathbb{M}'\mathbb{M}] = E[N_0^{-1}\sum_{i=1}^N \mathbb{m}_i \mathbb{m}_i'] = N_0^{-1}\sum_{i=1}^N E[\mathbb{m}_i \mathbb{m}_i'] = N_0^{-1}\sum_{i=1}^N E[(\mathbf{f}_i - \mathbf{W}'\mathbf{z}_i)(\mathbf{f}_i - \mathbf{W}'\mathbf{z}_i)'] = N_0^{-1}\sum_{i=1}^N E[(\mathbf{f}_i - E[\mathbf{f}_i])(\mathbf{f}_i - E[\mathbf{f}_i])'] = N_0^{-1}\sum_{i=1}^N \mathbf{G}_i$. Thus, $N_0^{-1}\sum_{i=1}^N \mathbf{G}_i = \mathbf{G}$. Q.E.D.

Appendix S11. The algorithm for estimating the candidate factor score distribution with $\mathbf{W}$ and $\mathbf{G}$.

SFA estimates the candidate factor score distribution with $\mathbf{W}$ and $\mathbf{G}$ given $\hat{\mathbf{\Lambda}}_s$ and $\hat{\mathbf{L}}$ by using the same cost function (A.2) as the one used in Stage 1. By the ALS algorithm, the least-squares estimate of $\mathbf{F}$ is obtained by $\hat{\mathbf{F}} = N_0^{-1/2}(\mathbf{K}_z \mathbf{R}_A \mathbf{Q}' + \mathbf{R}_\perp \mathbf{C} \mathbf{Q}_\perp) \mathbf{V}_f = \mathbf{Z}_{std}(\mathbf{O}_z \mathbf{\Omega}_z^{-1} \mathbf{R}_A \mathbf{Q}' \mathbf{V}_f) + N_0^{-1/2} \mathbf{R}_\perp \mathbf{C} \mathbf{Q}_\perp \mathbf{V}_f$, which means that the least-squares estimates of $\mathbf{W}$ and $\mathbf{G}$ can be obtained by $\hat{\mathbf{W}} = \mathbf{O}_z \mathbf{\Omega}_z^{-1} \mathbf{R}_A \mathbf{Q}' \mathbf{V}_f$ and $\hat{\mathbf{G}} = \mathbf{V}_f' \mathbf{Q}_\perp \mathbf{Q}_\perp' \mathbf{V}_f$, respectively.

However, there exists an infinite number of $\mathbf{C}$ in (A.11) that minimizes (A.2) subject to $\mathbf{C}'\mathbf{C} = \mathbf{I}_P$ given $\hat{\mathbf{W}}$, $\mathbf{R}_\perp$, $\mathbf{Q}_\perp$, and $\mathbf{V}_f$, so one can obtain an infinite number of $\hat{\mathbf{F}}$ s that result in the same value of ρ given $\hat{\mathbf{\Lambda}}_s$ and $\hat{\mathbf{L}}$. Let $\hat{\mathcal{F}}_{N,T}$ denote a set of all possible values of $\hat{\mathbf{F}}$ by changing the value of $\mathbf{C}$ in (A.11) given $\hat{\mathbf{W}}$, $\mathbf{R}_\perp$, $\mathbf{Q}_\perp$, and $\mathbf{V}_f$. Let $\mathcal{C}_{N-J-1,P}$ denote the set of all possible values of $\mathbf{C}$ with P orthonormal columns (i.e., $\mathcal{C}_{N-J-1,P} \equiv \{\mathbf{C} \in \mathbb{R}^{(N-J-1) \times P} \mid \mathbf{C}'\mathbf{C} = \mathbf{I}_P\}$). Let $\mathbb{C}$ denote a random matrix that takes on $\mathbf{C}$ in $\mathcal{C}_{N-J-1,P}$. SFA presumes that every $\mathbf{C}$ in $\mathcal{C}_{N-J-1,P}$ is equally likely to be true, implying that $\mathbb{C}$ follows the uniform distribution on $\mathcal{C}_{N-J-1,P}$ (For the mathematical definition of the uniform distribution of a random matrix with orthogonal columns, see Eaton, 1989, Chapter 2). Then, a random matrix $\hat{\mathbb{F}}$ on the sample space $\hat{\mathcal{F}}_{N,T}$ can be defined as

$$\hat{\mathbb{F}} = \mathbf{Z}_{std} \hat{\mathbf{W}} + N^{1/2} \mathbf{R}_\perp \mathbb{C} \mathbf{Q}_\perp' \mathbf{V}_f. \quad (\text{A.23})$$

SFA seeks to numerically obtain the probability distribution of $\hat{\mathbb{F}}$ by randomly generating multiple values of $\mathbb{C}$. Specifically, let $\mathbb{O}$ denote an $N - J - 1$ by P random matrix whose vectorized elements have the following distribution,

$$\text{vec}(\mathbb{O}) \sim \text{Normal}(\mathbf{0}, \mathbf{I}_{N-J-1} \otimes \mathbf{I}_P). \quad (\text{A.24})$$

As $\mathbb{O}(\mathbb{O}'\mathbb{O})^{-1/2}$ is known to follow the uniform distribution on $\mathcal{C}_{N-J-1, P}$ (Eaton, 1989, pp. 100–101), SFA randomly generates a value of $\mathbb{O}$, denoted by $\mathbf{O}$, and obtains a value of $\mathbb{C}$ by $\mathbf{C} = \mathbf{O}(\mathbf{O}'\mathbf{O})^{-1/2}$. By substituting the $\mathbf{C}$ value into (A.23), SFA obtains a value of $\hat{\mathbb{F}}$ (i.e., $\hat{\mathbf{F}}$). SFA repeats this resampling procedure multiple times (e.g., 1000) to obtain a set of $\hat{\mathbf{F}}$ values that approximate to $\hat{\mathcal{F}}_{N,T}$.

Note that if the resampling procedure is not sufficiently repeated, the resultant set of matrices of candidate factor score estimates may not approximate to $\hat{\mathcal{F}}_{N,T}$ due to the sampling error. Let $\{\hat{\mathbf{F}}^{(1)}, \hat{\mathbf{F}}^{(2)}, \dots, \hat{\mathbf{F}}^{(N_r)}\}$ denote a set of candidate factor score estimates obtained from the resampling procedure above, where $\hat{\mathbf{F}}^{(q)}$ is a matrix of candidate factor score estimates obtained from the q th sample ($q = 1, 2, \dots, N_r$), and N_r is the number of times the resampling procedure is repeated. In particular, if N_r is not sufficiently large, $N_r^{-1} \sum_{i=1}^{N_r} \hat{\mathbf{F}}^{(i)}$ may not be equivalent to $\mathbf{Z}_{std} \hat{\mathbf{W}}$, even though $E[\hat{\mathbb{F}}]$ is equivalent to $\mathbf{Z}_{std} \hat{\mathbf{W}}$. Accordingly, SFA additionally updates $\hat{\mathbf{F}}^{(q)}$ by $\hat{\mathbf{F}}_{updated}^{(q)} = (\hat{\mathbf{F}}^{(q)} - (N_r^{-1} \sum_{i=1}^{N_r} \hat{\mathbf{F}}^{(i)} - \mathbf{Z}_{std} \hat{\mathbf{W}}))$, when a small value is used for N_r .

Appendix S12. Theorem 6 and its proof

Theorem 6. If $\hat{\mathbf{L}} = \mathbf{L}$, $\hat{\mathbf{\Lambda}}_s = \mathbf{\Lambda}_s$, and $\mathbf{F}_{true}$ is representative given $\mathbf{Z}$, the estimate of the candidate factor score distribution with $\hat{\mathbf{W}}$ and $\hat{\mathbf{G}}$ approximates the uniform distribution on $\mathcal{F}_{N,T}$ with $\mathbf{W}$ and $\mathbf{G}$.

Proof. Suppose that $\hat{\mathbf{L}} = \mathbf{L}$, $\hat{\mathbf{\Lambda}}_s = \mathbf{\Lambda}_s$, $\mathbf{F}_{true}$ satisfies $N^{-1}\mathbf{1}_N'\mathbf{F}_{true} = \mathbf{0}$ and $N_0^{-1}\mathbf{F}_{true}'\mathbf{F}_{true} = \mathbf{\Lambda}_s$. Then, $\mathbf{Z}$ satisfies $\mathbf{1}_N'\mathbf{Z} = \mathbf{0}$ and $N_0^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{\Sigma}_s$. We need to show that (a) $\hat{\mathbf{W}}$ and $\hat{\mathbf{G}}$ are equivalent to $\mathbf{W}$ and $\mathbf{G}$, respectively, and (b) the probability distribution of $\hat{\mathbf{F}}$ on $\hat{\mathcal{F}}_{N,T}$ is equivalent to the uniform distribution on $\mathcal{F}_{N,T}$.

For (a), $\mathbf{A} = \mathbf{\Omega}_z\mathbf{\Omega}_z'\mathbf{L}'\mathbf{V}_f' = \mathbf{\Omega}_z^2\mathbf{Q}'$, implying that $\mathbf{Q}' = \mathbf{\Omega}_z^{-1}\mathbf{\Omega}_z'\mathbf{L}'\mathbf{V}_f'$. Then, $\hat{\mathbf{W}} = \mathbf{\Omega}_z\mathbf{\Omega}_z^{-1}\mathbf{R}_A\mathbf{Q}'\mathbf{V}_f = (\mathbf{\Omega}_z\mathbf{\Omega}_z^{-1}\mathbf{\Omega}_z^{-1}\mathbf{\Omega}_z')\mathbf{L}'(\mathbf{V}_f'\mathbf{V}_f) = \mathbf{\Sigma}_s^{-1}\mathbf{L}'\mathbf{\Lambda} = [\mathbf{\Sigma}_s^{-1}\mathbf{\Lambda}'\mathbf{\Phi}_s, \mathbf{\Sigma}_s^{-1}\mathbf{\Theta}] = [\mathbf{W}_1, \mathbf{W}_2] = \mathbf{W}$. Also, $\hat{\mathbf{G}} = \mathbf{V}_f'\mathbf{Q}_\perp\mathbf{Q}_\perp'\mathbf{V}_f = \mathbf{V}_f'(\mathbf{I}_T - \mathbf{Q}\mathbf{Q}')\mathbf{V}_f = \mathbf{V}_f'(\mathbf{I}_T - \mathbf{V}_f\mathbf{L}\mathbf{\Sigma}_s^{-1}\mathbf{L}'\mathbf{V}_f')\mathbf{V}_f = \mathbf{\Lambda}_s - \mathbf{\Lambda}_s\mathbf{L}\mathbf{\Sigma}_s^{-1}\mathbf{L}'\mathbf{\Lambda}_s = [\mathbf{I}_P, -\mathbf{\Lambda}]'(\mathbf{\Phi}_s - \mathbf{\Phi}_s\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}\mathbf{\Lambda}'\mathbf{\Phi}_s)[\mathbf{I}_P, -\mathbf{\Lambda}] = [\mathbf{I}_P, -\mathbf{\Lambda}]'\mathbf{G}_1[\mathbf{I}_P, -\mathbf{\Lambda}] = \mathbf{G}$.

For (b), we need to show that (c) the function of $\mathbb{C}$ for $\hat{\mathbf{F}}$ (A.23) is a one-to-one mapping of $\mathcal{C}_{N-J-1,P}$ onto $\hat{\mathcal{F}}_{N,T}$ and (d) $\hat{\mathcal{F}}_{N,T} = \mathcal{F}_{N,T}$. To facilitate their proofs, we re-express model equations for $\mathbb{F}$ and $\hat{\mathbb{F}}$ (i.e., (A.22) and (A.23)) with a random matrix $\mathbb{D}$, which denotes an N by P random matrix that takes on a value $\mathbf{D}$ in $\mathcal{D}_{N,P}$, where $\mathcal{D}_{N,P} \equiv \{\mathbf{D} \in \mathbb{R}^{N \times P} | \text{mean}(\mathbf{D}) = \mathbf{0}, \text{cov}(\mathbf{D}) = \mathbf{I}_P, \text{and } \text{cov}(\mathbf{Z}, \mathbf{D}) = \mathbf{0}\}$. Then, (A.22) can be re-expressed as

$$\mathbb{F} = \mathbf{Z}\mathbf{W} + \mathbb{D}[\mathbf{V}_{m1}, -\mathbf{V}_{m1}\mathbf{\Lambda}], \quad (\text{A.25})$$

where $\mathbf{V}_{m1}$ is a P by P matrix satisfying $\mathbf{V}_{m1}'\mathbf{V}_{m1} = \mathbf{G}_1$. The function of $\mathbb{D}$ for $\mathbb{F}$ is a one-to-one mapping of $\mathcal{D}_{N,P}$ onto $\mathcal{F}_{N,T}$. To prove this, we need to show the following three statements: (e) for any $\mathbf{D}$ in $\mathcal{D}_{N,P}$, $(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m)$ is in $\mathcal{F}_{N,T}$, (f) if $(\mathbf{Z}\mathbf{W} + \mathbf{D}_*\mathbf{V}_m) = (\mathbf{Z}\mathbf{W} + \mathbf{D}_{**}\mathbf{V}_m)$ for $\mathbf{D}_*$ and $\mathbf{D}_{**}$

in $\mathcal{D}_{N,P}$, then, $\mathbf{D}_* = \mathbf{D}_{**}$, and (g) for any $\mathbf{F}$ in $\mathcal{F}_{N,T}$, there exists $\mathbf{D}$ in $\mathcal{D}_{N,P}$ such that $\mathbf{F} = (\mathbf{Z}\mathbf{W} + \mathbf{D}_*\mathbf{V}_m)$.

For (e), let $\mathbf{D}$ in $\mathcal{D}_{N,P}$ be given. Then, $(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m)\mathbf{L} = \mathbf{Z}\mathbf{\Sigma}_s^{-1}(\mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda} + \mathbf{\Theta}) = \mathbf{Z}$,
 $mean(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m) = N^{-1}\mathbf{1}_N'(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m) = (N^{-1}\mathbf{1}_N'\mathbf{Z})\mathbf{W} + (N^{-1}\mathbf{1}_N'\mathbf{D})\mathbf{V}_m = \mathbf{0}$, and $cov(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m)$
 $= N^{-1}(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m)'(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m) = [\mathbf{\Lambda}'\mathbf{\Phi}_s, \mathbf{\Theta}]\mathbf{\Sigma}_s^{-1}[\mathbf{\Lambda}'\mathbf{\Phi}_s, \mathbf{\Theta}] + [\mathbf{V}_{m1}, \mathbf{V}_{m2}][\mathbf{V}_{m1}, \mathbf{V}_{m2}] =$

$$\begin{bmatrix} \mathbf{\Phi}_s\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}\mathbf{\Lambda}'\mathbf{\Phi}_s + \mathbf{G}_1 & \mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}\mathbf{\Theta} - \mathbf{G}_1\mathbf{\Lambda} \\ (\mathbf{\Phi}_s\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}\mathbf{\Theta} - \mathbf{G}_1\mathbf{\Lambda})' & \mathbf{\Theta}\mathbf{\Sigma}_s^{-1}\mathbf{\Theta} + \mathbf{\Lambda}'\mathbf{G}_1\mathbf{\Lambda} \end{bmatrix} = \mathbf{\Lambda}_s$$
, as $\mathbf{\Theta}\mathbf{\Sigma}_s^{-1}\mathbf{\Theta} + \mathbf{\Lambda}'(\mathbf{\Phi}_s - \mathbf{\Phi}_s\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}\mathbf{\Lambda}'\mathbf{\Phi}_s)\mathbf{\Lambda} = (\mathbf{\Sigma}_s - \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda})\mathbf{\Sigma}_s^{-1}(\mathbf{\Sigma}_s - \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda}) = \mathbf{\Sigma}_s - \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda} = \mathbf{\Theta}$ and $\mathbf{\Phi}_s\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}(\mathbf{\Sigma}_s - \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda}) - (\mathbf{\Phi}_s - \mathbf{\Phi}_s\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}\mathbf{\Lambda}'\mathbf{\Phi}_s)\mathbf{\Lambda} = \mathbf{\Phi}_s\mathbf{\Lambda} - \mathbf{\Phi}_s\mathbf{\Lambda} = \mathbf{0}$. For (f), suppose that $(\mathbf{Z}\mathbf{W} + \mathbf{D}_*\mathbf{V}_m) = (\mathbf{Z}\mathbf{W} + \mathbf{D}_{**}\mathbf{V}_m)$, where $\mathbf{D}_*$ and $\mathbf{D}_{**}$ are in $\mathcal{D}_{N,P}$. Then, $(\mathbf{Z}\mathbf{W} + \mathbf{D}_*\mathbf{V}_m) - (\mathbf{Z}\mathbf{W} + \mathbf{D}_{**}\mathbf{V}_m) = (\mathbf{D}_* - \mathbf{D}_{**})[\mathbf{V}_{m1}, \mathbf{V}_{m2}] = \mathbf{0}$, thereby making $(\mathbf{D}_* - \mathbf{D}_{**})\mathbf{V}_{m1} = \mathbf{0}$. As $\mathbf{G}_1$ is invertible (Guttman, 1955, theorem 4), $\mathbf{V}_{m1}$ is also invertible, which implies that $\mathbf{D}_* = \mathbf{D}_{**}$.

For (g), let $\mathbf{F} = [\mathbf{H}, \mathbf{E}]$ be given in $\mathcal{F}_{N,T}$, where $\mathbf{H}$ and $\mathbf{E}$ are the matrix of the first P columns and that of the last J columns, respectively. We need to show that there exists $\mathbf{D}$ in $\mathcal{D}_{N,P}$ that makes $\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m = \mathbf{F}$. Let us assume that $\mathbf{D} = (\mathbf{H} - \mathbf{Z}_{std}\mathbf{W}_1)\mathbf{V}_{m1}^{-1}$, where $\mathbf{W}_1 = \mathbf{\Sigma}_s^{-1}\mathbf{\Lambda}'\mathbf{\Phi}_s$. It satisfies all of its constraints such that $mean(\mathbf{D}) = N^{-1}\mathbf{1}_N'\mathbf{D} = N^{-1}\mathbf{1}_N'(\mathbf{H} - \mathbf{Z}\mathbf{W}_1)\mathbf{V}_{m1}^{-1} = ((N^{-1}\mathbf{1}_N'\mathbf{H}) - (N^{-1}\mathbf{1}_N'\mathbf{Z})\mathbf{W}_1)\mathbf{V}_{m1}^{-1} = \mathbf{0}$, $cov(\mathbf{D}) = N_0^{-1}\mathbf{D}'\mathbf{D} = N_0^{-1}(\mathbf{V}_{m1}^{-1})'(\mathbf{H} - \mathbf{Z}\mathbf{W}_1)'(\mathbf{H} - \mathbf{Z}\mathbf{W}_1)\mathbf{V}_{m1}^{-1} = (\mathbf{V}_{m1}^{-1})'(\mathbf{\Phi}_s - \mathbf{\Phi}_s\mathbf{\Lambda}\mathbf{\Sigma}_s^{-1}\mathbf{\Lambda}'\mathbf{\Phi}_s)\mathbf{V}_{m1}^{-1} = (\mathbf{V}_{m1}\mathbf{V}_{m1}^{-1})'(\mathbf{V}_{m1}\mathbf{V}_{m1}^{-1}) = \mathbf{I}_P$, and $cov(\mathbf{Z}, \mathbf{D}) = N_0^{-1}\mathbf{Z}'(\mathbf{H} - \mathbf{Z}_{std}\mathbf{W}_1)\mathbf{V}_{m1}^{-1} = (\mathbf{\Lambda}'\mathbf{\Phi}_s - \mathbf{\Lambda}'\mathbf{\Phi}_s)\mathbf{V}_{m1}^{-1} = \mathbf{0}$. Also, $\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m = \mathbf{Z}\mathbf{\Sigma}_s^{-1}[\mathbf{\Lambda}'\mathbf{\Phi}_s, \mathbf{\Theta}] + (\mathbf{H} - \mathbf{Z}\mathbf{W}_1)[\mathbf{I}_P, -\mathbf{\Lambda}] = [\mathbf{Z}\mathbf{W}_1 + (\mathbf{H} - \mathbf{Z}\mathbf{W}_1), \mathbf{Z}\mathbf{\Sigma}_s^{-1}\mathbf{\Theta} - (\mathbf{H}\mathbf{\Lambda} - \mathbf{Z}\mathbf{\Sigma}_s^{-1}(\mathbf{\Sigma}_s - \mathbf{\Theta}))] = [\mathbf{H}, \mathbf{E}] = \mathbf{F}$. The proof of (g) is the finite sample analogy to Guttman's (1955) one given $N = \infty$. As $\mathbb{F}$ follows the

uniform distribution on $\mathcal{F}_{N,T}$, $\mathbb{D}$ also follows the uniform distribution on $\mathcal{D}_{N,P}$, which means that every $\mathbf{D}$ is equally likely to be true.

The random matrix $\mathbb{D}$ also can be re-parameterized again as $\mathbb{D} = N_0^{1/2}[\mathbf{R}_\perp \mathbb{C}]$ because $\mathbb{D} = N_0^{1/2}[\mathbf{R}_\perp \mathbb{C}]$ is a one-to-one mapping of $\mathcal{C}_{N-J-1, P}$ onto $\mathcal{D}_{N,P}$. To prove this, we need to show the following three statements: (h) for any $\mathbf{C}$ in $\mathcal{C}_{N-J-1, P}$, $N_0^{1/2}[\mathbf{R}_\perp \mathbf{C}]$ is in $\mathcal{D}_{N,P}$, (i) if $N_0^{1/2}[\mathbf{R}_\perp \mathbf{C}_*] = N_0^{1/2}[\mathbf{R}_\perp \mathbf{C}_{**}]$ for $\mathbf{C}_*$ and $\mathbf{C}_{**}$ in $\mathcal{C}_{N-J-1, P}$, then $\mathbf{C}_* = \mathbf{C}_{**}$ and (j) for any $\mathbf{D}$ in $\mathcal{D}_{N,P}$, there exists $\mathbf{C}$ in $\mathcal{C}_{N-J-1, P}$ such that $\mathbf{D} = \mathbf{R}_\perp \mathbf{C}$. For (h), let $\mathbf{C}$ in $\mathcal{C}_{N-J-1, P}$ be given. Then, $\mathbf{C}$ satisfies $\mathbf{1}_N' (N_0^{1/2} \mathbf{R}_\perp \mathbf{C}) = N_0^{1/2} (\mathbf{1}_N' \mathbf{R}_\perp) \mathbf{C} = \mathbf{0}$, $N^{-1} (N_0^{1/2} \mathbf{R}_\perp \mathbf{C})' (N_0^{1/2} \mathbf{R}_\perp \mathbf{C}) = \mathbf{C}' (\mathbf{R}_\perp' \mathbf{R}_\perp) \mathbf{C} = \mathbf{I}_{T-r}$, $\mathbf{Z}' (N_0^{1/2} \mathbf{R}_\perp \mathbf{C}) = (\mathbf{K}_Z \mathbf{\Omega}_Z \mathbf{O}_Z)' (N_0^{1/2} \mathbf{R}_\perp \mathbf{C}) = N_0^{1/2} \mathbf{O}_Z \mathbf{\Omega}_Z (\mathbf{K}_Z' \mathbf{R}_\perp) \mathbf{C} = \mathbf{0}$. For (i), let us assume that $N_0^{1/2}[\mathbf{R}_\perp \mathbf{C}_*] = N_0^{1/2}[\mathbf{R}_\perp \mathbf{C}_{**}]$ for $\mathbf{C}_*$ and $\mathbf{C}_{**}$ in $\mathcal{C}_{N-J-1, P}$. Then $N_0^{1/2}[\mathbf{R}_\perp \mathbf{C}_*] - N_0^{1/2}[\mathbf{R}_\perp \mathbf{C}_{**}] = N_0^{1/2} \mathbf{R}_\perp (\mathbf{C}_* - \mathbf{C}_{**}) = \mathbf{0}$, followed by $\mathbf{R}_\perp' \mathbf{R}_\perp (\mathbf{C}_* - \mathbf{C}_{**}) = \mathbf{R}_\perp' \mathbf{0}$ and thus $\mathbf{C}_* = \mathbf{C}_{**}$. For (j), let $\mathbf{D}$ in $\mathcal{D}_{N,P}$ be given. Then, let us define $\mathbf{C}_+$ as $\mathbf{C}_+ \equiv N_0^{-1/2} \mathbf{R}_\perp' \mathbf{D}$. Let $\mathbf{\Gamma}_8$ denote an N by $(J+1)$ matrix whose columns form an orthonormal basis for the column space of $[\mathbf{K}_Z, \mathbf{1}_N]$. As $[\mathbf{\Gamma}_8, \mathbf{R}_\perp][\mathbf{\Gamma}_8, \mathbf{R}_\perp]' = \mathbf{I}_N$, $N_0^{-1} \mathbf{D}' \mathbf{R}_\perp \mathbf{R}_\perp' \mathbf{D} = N_0^{-1} \mathbf{D}' (\mathbf{I}_N - \mathbf{\Gamma}_8 \mathbf{\Gamma}_8') \mathbf{D} = N_0^{-1} \mathbf{D}' \mathbf{D} - N_0^{-1} (\mathbf{\Gamma}_8' \mathbf{D})' (\mathbf{\Gamma}_8' \mathbf{D}) = \mathbf{I}_P$. Also, $\mathbf{R}_\perp \mathbf{C}_+ = \mathbf{R}_\perp (\mathbf{R}_\perp' \mathbf{D}) = (\mathbf{I}_N - \mathbf{\Gamma}_8 \mathbf{\Gamma}_8') \mathbf{D} = \mathbf{D}$.

As $\mathbb{D} = N_0^{1/2}[\mathbf{R}_\perp \mathbb{C}]$ is a one-to-one mapping of $\mathcal{C}_{N-J-1, P}$ onto $\mathcal{D}_{N,P}$, the model equation (A.23) can be re-expressed as

$$\hat{\mathbf{F}} = \mathbf{Z}\mathbf{W} + \mathbb{D}\mathbf{Q}_\perp' \mathbf{V}_f. \quad (\text{A.26})$$

The function of $\mathbb{D}$ for $\hat{\mathbf{F}}$ is a one-to-one mapping of $\mathcal{D}_{N,P}$ onto $\mathcal{F}_{N,T}$. To prove this, we only need to show that if $\hat{\mathbf{F}}_* = \hat{\mathbf{F}}_{**}$ for any $\hat{\mathbf{F}}_*$ and $\hat{\mathbf{F}}_{**}$ in $\mathcal{F}_{N,T}$, then $\mathbf{D}_* = \mathbf{D}_{**}$ for any $\mathbf{D}_*$ and $\mathbf{D}_{**}$ in $\mathcal{D}_{N,T}$.

Let $\mathbf{D}_*$ and $\mathbf{D}_{**}$ in $\mathcal{D}_{N,T}$ be given, having $\hat{\mathbf{F}}_* = \mathbf{Z}\mathbf{W} + \mathbf{D}_* \mathbf{Q}_\perp' \mathbf{V}_f$ and $\hat{\mathbf{F}}_{**} = \mathbf{Z}\mathbf{W} + \mathbf{D}_{**} \mathbf{Q}_\perp' \mathbf{V}_f$.

Suppose that $\hat{\mathbf{F}}_* = \hat{\mathbf{F}}_{**}$. Then, $\hat{\mathbf{F}}_* - \hat{\mathbf{F}}_{**} = (\mathbf{Z}\mathbf{W} + \mathbf{D}_* \mathbf{Q}_\perp' \mathbf{V}_f) - (\mathbf{Z}\mathbf{W} + \mathbf{D}_{**} \mathbf{Q}_\perp' \mathbf{V}_f) = (\mathbf{D}_* -$

$\mathbf{D}_{**})\mathbf{Q}_\perp'\mathbf{V}_f = \mathbf{0}$. As $(\mathbf{D}_* - \mathbf{D}_{**})\mathbf{Q}_\perp'\mathbf{V}_f\mathbf{V}_f^{-1}\mathbf{Q}_\perp = \mathbf{0}\mathbf{V}_f^{-1}\mathbf{Q}_\perp$, $\mathbf{D}_* = \mathbf{D}_{**}$. From the fact that the function of $\mathbb{C}$ for $\mathbb{D}$ is a one-to-one mapping of $\mathcal{C}_{N-J-1, P}$ onto $\mathcal{D}_{N, P}$ and the function of $\mathbb{D}$ for $\widehat{\mathbf{F}}$ (A.26) is a one-to-one mapping of $\mathcal{D}_{N, P}$ onto $\widehat{\mathcal{F}}_{N, T}$, it follows that (c) the function of $\mathbb{C}$ for $\widehat{\mathbf{F}}$ (A.23) is a one-to-one mapping of $\mathcal{C}_{N-J-1, P}$ onto $\widehat{\mathcal{F}}_{N, T}$. As $\mathbb{C}$ follows the uniform distribution on $\mathcal{C}_{N-J-1, P}$, $\widehat{\mathbf{F}}$ follows the uniform distribution on $\widehat{\mathcal{F}}_{N, T}$.

Lastly, for (d), we need to show that (k) any $\widehat{\mathbf{F}}$ in $\widehat{\mathcal{F}}_{N, T}$ is in $\mathcal{F}_{N, T}$ and (l) any $\mathbf{F}$ in $\mathcal{F}_{N, T}$ is in $\widehat{\mathcal{F}}_{N, T}$. For (k), let $\mathbf{D}$ in $\mathcal{D}_{N, T}$ be given, thereby having $\widehat{\mathbf{F}} = \mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{Q}_\perp'\mathbf{V}_f$. The estimate of the candidate factor score matrix $\widehat{\mathbf{F}}$ given $\mathbf{D}$ satisfies that $\widehat{\mathbf{F}}\mathbf{L} = \mathbf{Z}\mathbf{W}\mathbf{L} + \mathbf{D}\mathbf{Q}_\perp'\mathbf{V}_f\mathbf{L} = \mathbf{Z}\Sigma_s^{-1}\mathbf{L}'\Delta_s\mathbf{L} = \mathbf{Z}$, $mean(\widehat{\mathbf{F}}_*) = N^{-1}\mathbf{1}_N'\widehat{\mathbf{F}}_* = N^{-1}\mathbf{1}_N'(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{Q}_\perp'\mathbf{V}_f) = (N^{-1}\mathbf{1}_N'\mathbf{Z})\mathbf{W} + (N^{-1}\mathbf{1}_N'\mathbf{D})\mathbf{Q}_\perp'\mathbf{V}_f = \mathbf{0}$, and $cov(\widehat{\mathbf{F}}) = N_0^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} = N_0^{-1}(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{Q}_\perp'\mathbf{V}_f)'(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{Q}_\perp'\mathbf{V}_f) = \mathbf{W}'\Sigma_s\mathbf{W} + \mathbf{G} = \Delta_s$, implying that $\widehat{\mathbf{F}}_*$ is in $\mathcal{F}_{N, T}$. For (l), we need to show that for any $\mathbf{F}$ in $\mathcal{F}_{N, T}$, there exists $\mathbf{D}$ in $\mathcal{D}_{N, T}$, such that $\mathbf{F} = \mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{Q}_\perp'\mathbf{V}_f$. Let $\mathbf{F}_*$ in $\mathcal{F}_{N, T}$ be given. Let us define $\mathbf{D}_*$ as $\mathbf{D}_* = (\mathbf{F}_* - \mathbf{Z}\mathbf{W})\mathbf{V}_f^{-1}\mathbf{Q}_\perp$. The matrix $\mathbf{D}_*$ satisfies $mean(\mathbf{D}_*) = N^{-1}\mathbf{1}_N'\mathbf{D}_* = N^{-1}\mathbf{1}_N'(\mathbf{F}_* - \mathbf{Z}\mathbf{W})\mathbf{V}_f^{-1}\mathbf{Q}_\perp = \mathbf{0}$, $cov(\mathbf{D}_*) = N_0^{-1}\mathbf{D}_*'\mathbf{D}_* = \mathbf{Q}_\perp'\mathbf{V}_f^{-1}\mathbf{G}\mathbf{V}_f^{-1}\mathbf{Q}_\perp = \mathbf{Q}_\perp'\mathbf{V}_f^{-1}\mathbf{V}_f'\mathbf{Q}_\perp\mathbf{Q}_\perp'\mathbf{V}_f\mathbf{V}_f^{-1}\mathbf{Q}_\perp = \mathbf{I}_P$, and $cov(\mathbf{Z}, \mathbf{D}_*) = N_0^{-1}\mathbf{Z}'\mathbf{D}_* = N_0^{-1}\mathbf{Z}'(\mathbf{F}_* - \mathbf{Z}\mathbf{W})\mathbf{V}_f^{-1}\mathbf{Q}_\perp = N_0^{-1}(\mathbf{Z}'\mathbf{Z}\mathbf{W} - \mathbf{Z}'\mathbf{Z}\mathbf{W})\mathbf{V}_f^{-1}\mathbf{Q}_\perp = \mathbf{0}$, which means that $\mathbf{D}_*$ is in $\mathcal{D}_{N, T}$. Also, $\mathbf{D}_*$ satisfies $\mathbf{Z}\mathbf{W} + \mathbf{D}_*\mathbf{Q}_\perp'\mathbf{V}_f = \mathbf{Z}\mathbf{W} + (\mathbf{F}_* - \mathbf{Z}\mathbf{W})\mathbf{V}_f^{-1}\mathbf{Q}_\perp\mathbf{Q}_\perp'\mathbf{V}_f = \mathbf{Z}\mathbf{W} + (\mathbf{F}_* - \mathbf{Z}\mathbf{W})(\mathbf{I}_T - \mathbf{V}_f^{-1}\mathbf{Q}\mathbf{Q}'\mathbf{V}_f) = \mathbf{Z}\mathbf{W} + (\mathbf{F}_* - \mathbf{Z}\mathbf{W})(\mathbf{I}_T - \mathbf{V}_f^{-1}\mathbf{V}_f\mathbf{L}\mathbf{O}_z\mathbf{\Omega}_z^{-2}\mathbf{O}_z'\mathbf{L}'\mathbf{V}_f'\mathbf{V}_f) = \mathbf{Z}\mathbf{W} + (\mathbf{F}_* - \mathbf{Z}\mathbf{W})(\mathbf{I}_T - \mathbf{L}\Sigma_s^{-1}\mathbf{L}'\Delta_s) = \mathbf{Z}\mathbf{W} + (\mathbf{F}_* - \mathbf{Z}\mathbf{W})(\mathbf{I}_T - \mathbf{L}\mathbf{W}) = \mathbf{Z}\mathbf{W} + \mathbf{F}_* - \mathbf{Z}\mathbf{W} - \mathbf{F}_*\mathbf{L}\mathbf{W} + \mathbf{Z}\mathbf{W}\mathbf{L}\mathbf{W} = \mathbf{Z}\mathbf{W} + \mathbf{F}_* - \mathbf{F}_*\mathbf{L}\mathbf{W} = \mathbf{F}_*$, indicating that any $\mathbf{F}$ in $\mathcal{F}_{N, T}$ is in $\widehat{\mathcal{F}}_{N, T}$. From (k) and (l), it follows that (d) $\mathcal{F}_{N, T} = \widehat{\mathcal{F}}_{N, T}$. Q.E.D.

References

- Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, 34(1), 93–99. <https://doi.org/10.1111/j.2044-8317.1981.tb00620.x>
- Bentler, P. M. (1995). *EQS structural equations program manual* (Vol. 6). Multivariate Software.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. <https://doi.org/10.1002/9781118619179>
- Bollen, K. A., & Stine, R. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Sage Publications.
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, 6(1–2), 25–53. <https://doi.org/10.1080/15366360802035497>
- de Leeuw, J. (2017). *Factor analysis as matrix decomposition and approximation: Theory*. <http://deleeuwpxd.net/pubfolders/factor/factor.pdf>
- de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41(4), 471–503. <https://doi.org/10.1007/BF02296971>
- Dijkstra, T. K. (1989). Reduced form estimation, hedging against possible misspecification. *International Economic Review*, 30(2), 373–390. <https://doi.org/10.2307/2526653>
- Dijkstra, T. K. (2017). A perfect match between a model and a mode. In H. Latan & R. Noonan (Eds.), *Partial least squares path modeling: Basic concepts, methodological issues and applications* (pp. 55–80). Springer. https://doi.org/10.1007/978-3-319-64069-3_4
- Eaton, M. L. (1989). Group invariance applications in Statistics. In *Regional Conference Series in Probability and Statistics* (Vol. 1). Institute of Mathematical Statistics. <http://www.jstor.org/stable/4153172>
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM. <https://doi.org/10.1137/1.9781611970319>
- Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic basic problems of common-factor theory. *British Journal of Statistical Psychology*, 8(2), 65–81. <https://doi.org/10.1111/j.2044-8317.1955.tb00321.x>
- Jöreskog, K. G., & Sorbom, D. (1986). *PRELIS: A program for multivariate data screening and data summarization*. Scientific Software, Mooresville.
- Kim, H., & Millsap, R. (2014). Using the Bollen-Stine bootstrapping method for evaluating approximate fit indices. *Multivariate Behavioral Research*, 49(6), 581–596. <https://doi.org/10.1080/00273171.2014.947352>
- Lance, C. E., Cornwell, J. M., & Mulaik, S. A. (1988). Limited information parameter estimates

- for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research*, 23(2), 171–187. https://doi.org/10.1207/s15327906mbr2302_3
- Leong, F. T. L., & Huang, J. L. (2016). Standard error of measurement. In *Britannica*. <https://www.britannica.com/science/standard-error-of-measurement>
- Mulaik, S. A. (2009). *Foundations of factor analysis* (2nd ed.). Chapman and Hall/CRC Press. <https://doi.org/10.1201/b15851>
- Perla, J., Sargent, T. J., & Stachurski, J. (2020). *Quantitative economics with Julia*. QuantEcon. <https://julia.quantecon.org/>
- Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika*, 44(2), 157–167. <https://doi.org/10.1007/BF02293967>
- Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, 41(1), 1–32. <https://doi.org/10.1037/h0075959>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). The MIT Press.