Constructs May or May Not Be Latent: Studies on Two Domains of Structural Equation Modeling

Gyeongcheol Cho

Department of Psychology

McGill University, Montreal, Canada

August 2023

A dissertation submitted to McGill University

in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Table of Contents

Abstract

Psychological research often involves empirical testing of hypothetical relationships among variables including constructs or conceptual variables (e.g., intelligence). Although constructs are not as directly measurable as physical properties (e.g., weight), structural equation modeling (SEM) makes it possible to test such relationships involving constructs based on the data of observed variables.

Researchers in psychology have typically used SEM while assuming all constructs in the model are latent, which indicates that the constructs are considered real entities that exist independently of observed variables. An SEM domain that considers every construct to be latent and represents it by a (common) factor is called factor-based SEM. However, assuming that some entity is latent does not guarantee that it indeed exists as latent by nature. Clearly, some constructs, including socioeconomic status and genes, cannot be seen as latent but rather correspond to a summary or cluster of relevant observed variables. Another SEM domain, called component-based SEM, has emerged to deal with such constructs, where constructs are represented as composite indexes of observed variables, termed components. Therefore, it would be valuable for researchers to study both SEM domains and be able to strategically select, for each research context, the domain that best aligns with their theoretical assumptions about constructs of interest and their interrelationships.

This dissertation aims to make theoretical and methodological contributions to the two SEM domains. It begins with a systematic comparison of the two SEM domains along with an illustrative example. Then, it presents three novel SEM methods—structured factor analysis (SFA; Cho and Hwang, 2022), convex generalized structured component analysis (convex GSCA; Cho and Hwang, under review), and deep learning generalized structured component analysis (DL-GSCA; Cho and Hwang, in press)—to resolve two long-standing problems in each SEM domain. Jöreskog (1978)’s covariance-based approach, which has

been considered the de-facto standard method for factor-based SEM, has two limitations: the occurrence of improper solutions (e.g., negative variance estimates) and the lack of a statistical tool to deal with the factor score indeterminacy problem (i.e., an infinite number of factor scores being possibly true). As an alternative, SFA guarantees the convergence of its algorithm to proper solutions and enables an estimation of the probability distribution of all the factor scores that are possibly true given the data matrix. On the other hand, convex GSCA and DL-GSCA are new extensions of generalized structured component analysis (GSCA; Hwang and Takane, 2004), which overcomes two limitations of component-based SEM— removing information on indicators' scale from data and restricting a functional form of components to be linear. Convex GSCA analyzes the raw data of indicators without standardizing their scores, allowing the generated composite indexes and their network to be interpreted with respect to the original indicators' scale. DL-GSCA combines deep learning neural networks with GSCA in a single framework to find the optimal functional form for each component, maximizing its predictive power for target outcome variables. This dissertation provides the technical underpinnings of the three proposed methods in detail and demonstrates their practical utility through both simulated and real data analyses.

Abrégé

La recherche en psychologie implique souvent des tests empiriques de relations hypothétiques entre des variables, y compris des construits ou variables conceptuelles (par exemple, l'intelligence). Bien que les construits ne soient pas aussi directement mesurables que les propriétés physiques (par exemple, le poids), la modélisation par équations structurelles (SEM) permet de tester de telles relations impliquant des construits sur la base des données de variables observées.

Les chercheurs en psychologie ont généralement utilisé la SEM en supposant que tous les construits du modèle sont latents, ce qui indique que les construits sont considérés comme des entités réelles qui existent indépendamment des variables observées. Un domaine de la SEM qui considère chaque construit comme latent et le représente par un facteur (commun) est appelé SEM basée sur les facteurs. Cependant, supposer qu'une entité est latente ne garantit pas qu'elle existe réellement comme latente par nature. Il est clair que certains construits, y compris le statut socioéconomique et les gènes, ne peuvent pas être considérés comme latents mais correspondent plutôt à un résumé ou à un ensemble de variables observées pertinentes. Un autre domaine de la SEM, appelé SEM basée sur les composants, a émergé pour traiter de tels construits, où les construits sont représentés comme des indices composités de variables observées, appelés composants. Ainsi, il serait précieux pour les chercheurs d'étudier les deux domaines de la SEM et de pouvoir choisir stratégiquement, pour chaque contexte de recherche, le domaine qui s'aligne le mieux avec leurs hypothèses théoriques sur les construits qui les intéressent et leurs interrelations.

Cette thèse vise à apporter des contributions théoriques et méthodologiques aux deux domaines de la SEM. Elle commence par une comparaison systématique des deux domaines de la SEM accompagnée d'un exemple illustratif. Ensuite, elle présente trois nouvelles méthodes de la SEM : l'analyse factorielle structurée (SFA ; Cho et Hwang, 2022), l'analyse

structurée générale des composants convexes (convexe GSCA ; Cho et Hwang, en révision), et l'analyse structurée générale des composants en apprentissage profond (DL-GSCA ; Cho et Hwang, sous presse) afin de résoudre deux problèmes de longue date dans chaque domaine de la SEM. L'approche basée sur la covariance de Jöreskog (1978), qui a été considérée comme la méthode standard de facto pour la SEM basée sur les facteurs, a deux limitations : l'apparition de solutions incorrectes (par exemple, des estimations de variance négative) et l'absence d'un outil statistique pour traiter le problème de l'indétermination du score factoriel (c'est-à-dire, un nombre infini de scores factoriels pouvant être vrais). En alternative, la SFA garantit la convergence de son algorithme vers des solutions correctes et permet une estimation de la distribution de probabilité de tous les scores factoriels qui peuvent être vrais étant donné la matrice de données. D'autre part, convexe GSCA et DL-GSCA sont de nouvelles extensions de l'analyse structurée généralisée des composants (GSCA ; Hwang et Takane, 2004), qui surmontent deux limitations de la SEM basée sur les composants : enlevant l'information sur l'échelle des indicateurs à partir des données et limitant une forme fonctionnelle des composants à être linéaire. Convexe GSCA analyse les données brutes des indicateurs sans standardiser leurs scores, permettant aux indices composités générés et à leur réseau d'être interprétés par rapport à l'échelle des indicateurs originaux. DL-GSCA combine les réseaux neuronaux d'apprentissage profond avec GSCA dans un seul cadre pour trouver la forme fonctionnelle optimale pour chaque composant, maximisant son pouvoir prédictif pour les variables de résultat cibles. Cette thèse fournit les bases techniques des trois méthodes proposées en détail et démontre leur utilité pratique à travers des analyses de données simulées et réelles.

Acknowledgements

This dissertation has come to fruition due to the relentless support and countless sacrifices made by an extraordinary group of individuals. I am humbled to take this moment to express my deep appreciation for their unwavering commitment and assistance.

I am deeply indebted to my supervisor, Professor Heungsun Hwang, who, over the course of six years, granted me the freedom to navigate the depths of my research. Professor Hwang, your approach of treating me as a peer and valuing my insights during our academic dialogues was empowering. You illuminated the path towards transforming a seedling of an idea into a robust academic paper and instilled in me the highest standards of scholarly writing. Your guidance fostered an environment where I could generate a multitude of research ideas, nurture them, and eventually shape them into individual papers. In particular, the paper that comprises the second chapter of this dissertation was five years in the making, while the remaining two papers have been under development for three and two years respectively. Without the expansive freedom in research and your insightful feedback, the realization of these papers would have been impossible. I express my sincerest gratitude for your enduring support throughout my Ph.D. and for molding me into an independent quantitative psychologist.

My heartfelt thanks also go out to the members of the Department of Psychology at McGill University. Professors Carl F. Falk, Jessica Kay Flake, and Milica Miočević, your willingness to lend a helping hand and share your invaluable resources whenever I found myself in academic straits was incredibly beneficial. To all the staff members in our department, I have consistently been moved by your dedication to students in every interaction we have had. The former department chair, John E. Lydon, and my lab mate, Mairead Shaw, you've been instrumental in introducing me to the values of diversity and tolerance in Canada. My lunch companion, Sally Xie, your assistance over the past five years

has been manifold and significant. To all those with whom I have had the privilege of sharing this enlightening Ph.D. journey, I thank you wholeheartedly.

In addition, I owe a great deal of gratitude to OpenAI for ushering in a new epoch of conversational AI. The advent of ChatGPT 3.5 was transformative; it became an indispensable ally in my academic pursuits. I have consulted with it daily to refine my research ideas and polish my writings, and its contributions to the finalization of this dissertation were immeasurable. I am inspired by your mission to develop artificial general intelligence for the greater good of humanity, and I aspire to make significant contributions to this noble cause in the future.

Finally, my deepest appreciation extends to my cherished family. I express profound gratitude and sincere apologies to my devoted wife, Younyoung Choi. Your steadfast commitment to our family over the past six years has been an unwavering source of strength, transforming the completion of this dissertation from a distant dream into a tangible reality. I cannot express enough appreciation for your selfless dedication in managing our home and caring for our child, allowing me to devote myself wholly to this work. To my parents, Hyosook Kim and Deokhee Cho, your constant support and heartfelt prayers have enabled me to strike a delicate balance between academic pursuits and family responsibilities. Your resolute belief in me has been a driving force throughout this journey. In essence, this achievement is not merely mine but a testament to the love and unwavering support of my incredible family.

Contribution of Authors

This dissertation follows a manuscript-based format, incorporating two manuscripts submitted for publication and one published in a peer-reviewed journal. These works collectively form a unified research project focused on the technical advancements within two domains of structural equation modeling (SEM). One of the manuscripts was also presented at an international conference. The details of the manuscripts are as follows:

- Chapter 2. SFA: A data matrix-based alternative approach to SEM
    - Publication [1]: **Cho, G**., & Hwang, H. (2023). Structured factor analysis: A data matrix-based alternative approach to structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(3), 364–377. https://doi.org/10.1080/10705511.2022.2126360
- Chapter 3. GSCA accommodating convex components: A knowledge-based multivariate method with interpretable composite indexes
    - Manuscript [2]: **Cho, G.**, & Hwang, H. Generalized structured component analysis accommodating convex components: A knowledge-based multivariate method with interpretable composite indexes. Under the second round of review at *Psychometrika*.
    - Conference Presentation: **Cho, G.**, & Hwang, H. (June 2021). *Generalized structured component analysis accommodating convex components*. Presented at the Annual Meeting of Statistical Society of Canada, Virtual.
- Chapter 4. Deep learning GSCA: An interpretable artificial neural network model with composite indexes
    - Manuscript [3]: **Cho, G.**, & Hwang, H. (in press). Deep learning generalized structured component analysis: An interpretable artificial neural network

model with composite indexes. *Structural Equation Modeling: A Multidisciplinary Journal*.

The article [1] and the manuscripts [2] and [3] are included in full in Chapters 2, 3, and 4 of this dissertation, respectively.

I am the first and corresponding author for [1] and [2], and the first author for [3]. Throughout all stages of these research endeavors, my doctoral supervisor Professor Heungsun Hwang provided critical input and revisions. Each draft of these manuscripts was meticulously reviewed and edited by him. Also, he proposed an initial methodological idea for [3]. Outside of these acknowledged contributions, I was responsible for the complete body of work presented in this dissertation, which includes the technical development of primary algorithms, programming, simulation studies, data analysis, and the manuscript writing.

List of Tables

List of Figures

Chapter 1. Introduction and Background

1.1. Two Domains of Structural Equation Modeling

Psychological research often aims to test hypotheses about the relationships among variables including *constructs*. A construct refers to a conceptual variable that describes or explains a cluster of covarying human behaviors or natural/social phenomena (Binning, 2015; Edwards & Bagozzi, 2000). For example, if a specific human abnormal behavior tends to co-occur with other behaviors of the same kind, researchers may hypothesize something psychopathological for describing or explaining the covariations of those behaviors and may give it a name that denotes a psychotic syndrome or a mental disorder. Other examples of constructs that can be found in psychological studies are intelligence, basic emotions, personality, and socioeconomic status. Constructs are not as directly measurable as physical properties, such as age or height, thereby making it impossible to directly collect the data of those constructs. Consequently, to conduct an empirical study involving theoretical constructs, researchers need to find relevant observed variables that can serve as measures or *indicators* of constructs and to draw on a statistical methodology specialized for analyzing the indicators' data. This methodology is called *structural equation modeling* (SEM).

SEM allows for the statistical testing or exploration of the hypothetical relationships between variables involving constructs through the analysis of their indicators' data. It typically formalizes two classes of models: *measurement* and *structural*. The measurement model shows how constructs are associated with their respective indicators, whereas the structural model represents the interconnected network of variables including constructs and observed variables that are not used to measure the constructs. If constructs could be treated as observed variables whose data can be directly collected, the structural model would be identical to a generalized multivariate regression model, including the seemingly unrelated

model (Zellner, 1962) and the simultaneous equation model (Goldberger, 1964). In this respect, the distinguishing feature of SEM lies in the measurement model.

Conventionally, SEM researchers formalize the causal relationships between constructs and their indicators in the measurement model, assuming that every construct is *latent* in its indicators' cluster. This assumption means that each construct is considered as a real entity that exists independently of its indicators and causes the indicators to covary, though it is unobservable. This type of construct is called a hypothetical or latent construct in the literature (e.g., Heise, 1972; e.g., MacCorquodale & Meehl, 1948). Latent constructs are typically assumed to be *unique* sources of their indicators' covariations, under which the indicators become conditionally independent of each other given the scores of the latent constructs—*local independence*. Under this assumption, each latent construct is represented by a *common factor* of indicators, which refers to a statistical proxy accounting for the shared variance among the indicators. Subsequently, there exists a residual variance of each indicator that is unexplained by the common factors. This unexplained variance is attributed to a random measurement error, represented by a statistical proxy called a *unique factor*, which captures the variance specific to the indicator. By default, each unique factor is assumed to be uncorrelated to other common and unique factors, though the zero-correlations between unique factors can be partially relaxed if necessary. The measurement model with common and unique factors is called the *reflective model*, and indicators in the reflective model are called *reflective* or *effect indicators* (e.g., Bollen & Bauldry, 2011). The SEM domain having the reflective model as its measurement model is called *factor-based SEM* (e.g., Tenenhaus, 2008).

In psychological literature, the term SEM typically refers to this factor-based SEM domain. Popular SEM programs such as AMOS, Mplus, and the *lavaan* R package only covers factor-based SEM. SEM textbooks frequently used in psychology, including Bollen

(1989), Kleine (2016), and Hoyle (2014), also only deal with factor-based SEM. As a result, psychologists have adopted factor-based SEM exclusively for their empirical studies involving constructs, as highlighted in Rhemtulla et al.'s (2020) literature review of articles published in six major APA journals from 2016 to 2017. Jöreskog's (1973, 1978) covariance-based approach, factor score regression (Croon, 2002; Rosseel & Loh, 2022; Skrondal & Laake, 2001), and Bollen's (1996, 2019) model-implied instrument variable method are representative in this SEM domain.

However, applying factor-based SEM, as noted above, entails the assumption that every theoretical construct of interest exists by nature as a latent variable, which requires a strong justification. If some constructs are not latent in their indicators but incorrectly represented by common factors with a reflective measurement model, the specified measurement model can never be true, thereby resulting in biased estimates of structural model parameters (Cho, Sarstedt, et al., 2022; Hwang, Cho, Jung, et al., 2021). Under this circumstance, every statistical procedure in the factor-based domain, such as the χ^2 test for identifying the true model or the t-test for testing the estimates of individual model parameters, would not be valid anymore. Thus, before applying the factor-based SEM, researchers should justify how every construct of interests in their model can be considered a latent variable.

To justify the presence of latent constructs in the model, however, there are two issues that researchers would find difficult to address. First, there are many constructs whose existence cannot clearly be separated from their indicators. Socioeconomic status is a representative example of such constructs. Its indicators—income, educational level, and occupational prestige—conceptually constitute three dimensions of socioeconomic status itself (American Psychological Association, 2007), so that one cannot say socioeconomic status and its indicators are distinct entities that have causal effects on one another. A gene is

another exemplary construct that consists of its indicators, as each gene is defined as a biological cluster of multiple single nucleotide polymorphisms (SNPs) located within the gene.

Second, it is essentially not possible to obtain empirical evidence to verify the latent existence of constructs even when they are truly latent by nature. The strong covariation observed between indicators of a construct cannot act as sufficient evidence to verify the existence of the construct as an underlying cause of the indicators since, as is well known, covariation does not imply causation. For instance, if the severity of three pathological symptoms (depressive mood, somatic discomfort, and social avoidance) tends to strongly covary, one may think of a latent, psychopathological construct, named depression, that makes those symptoms co-occur. However, one cannot exclude the possibility that those symptoms may be simply associated with each other without having any underlying causal structure, or their strong covariation may result from the fact that they mutually affect each other (e.g., Borsboom, 2017). In the same vein, a perfect goodness-of-fit of a reflective model in factor-based SEM cannot provide conclusive evidence that constructs in the model are latent, as the reflective model still may produce a perfect goodness-of-fit for the data generated from an entirely disparate model that does not involves any latent constructs (e.g., Hayduk, 2014).

There is an alternative SEM domain that can avoid the two issues above, called *component-based SEM* (Rigdon, 2012; Tenenhaus, 2008). This SEM domain does not assume that constructs of interest are latent. It simply treats each construct as a summary or label of its covarying indicators' cluster. For example, for a depression construct for the three covarying symptoms above (depressive mood, somatic discomfort, and social avoidance), component-based SEM regards the depression construct as a descriptive variable that is made to refer to its symptom cluster efficiently. In the literature, this type of descriptive construct is

referred to by various names such as abstractive construct, intervening variable, emergent variable, or synthetic variable (e.g., Cole et al., 1993; MacCorquodale & Meehl, 1948; Nimon et al., 2010). Component-based SEM seeks to create composite indexes of indicators to represent the constructs, each of which can serve a summary of its indicators' cluster (e.g., Bollen, 2011; Cho, Sarstedt, et al., 2022). As in principle component analysis, the composite indexes are called components in component-based SEM.

In line with its perspective on the relationship between constructs and their indicators, component-based SEM employs a distinct measurement model, which consists of two sub-models: *weighted relation* and *component-measurement*. The weighted relation model defines each component as a deterministic function of its indicators (e.g., weighted sums of indicators), whereas the component-measurement model shows which cluster of indicators each component summarizes. Generalized structured component analysis (GSCA; Hwang & Takane, 2004, 2014) and partial least squares path modeling (PLSPM; Lohmöller, 1989; Wold, 1973, 1985) are two representative approaches in the component-based domain.

1.2. Model Specification in Two SEM Domains

To aid in understanding of how constructs are distinctively represented in the two SEM domains, this section will provide a brief explanation of their respective model specifications with a simple illustration. For the sake of clarity, all common factors, components, and indicators in the models are assumed to be standardized. Also, for the specification of the component model, this section will focus on the model proposed by Cho and Choi (2020), as it can be considered an underlying population model of any of the component-based SEM methods (Cho & Choi, 2020).

Suppose that a researcher hypothesizes that employees' health status may affect their job satisfaction level. If the two variables could be considered observed variables and their

data could be directly collected, the researcher's hypothesis could be easily tested under certain assumptions with the following simple regression model,

$$\textit{health status} = b_1 \textit{job satisfaction} + \zeta, \tag{1.1}$$

where b_1 is a path coefficient signifying the effect of *health status* on *job satisfaction*, ξ is an error term in (1.1), and *cov*(*health status*, ζ) = 0. If the estimate of b_1 is statistically significant, we may conclude that there exists a non-zero effect of *health status* on *job satisfaction*. However, the two variables (*job satisfaction* and *health status*) are actually theoretical constructs that are not as directly measurable as observed variables, so their data cannot be directly collected and thus, simple regression analysis cannot be applied. Let us assume that each of the two constructs can be measured with three relevant observed variables that can serve as the indicators. Under this condition, SEM would be indispensable to test the researcher's hypothesis.

Factor-based SEM represents the two constructs by two common factors of the six indicators, under the assumption that each construct is latent in its indicators as the unique source of their covariations. Let $\mathbf{z} = [z_1, z_2, z_3, z_4, z_5, z_6]'$ denote a vector of six indicators, where a set of z_1, z_2, and z_3 is for *health status*, a set of z_4, z_5, and z_6 is for *job satisfaction*. The covariance matrix of $\mathbf{z}$ is denoted by $\mathbf{\Sigma}$. Let $\mathbf{h} = [h_1, h_2]'$ denote a vector of two common factors for the two sets of indicators, where h_1 is a common factor of z_1, z_2, and z_3 and h_2 is a common factor of z_4, z_5, and z_6. Let $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6]'$ denote a vector of six unique factors, where ε_j is a unique factor of z_j (j = 1, 2, ⋯, 6). Each unique factor is assumed to be uncorrelated with any other factors at default (i.e., *cov*($\boldsymbol{\varepsilon}$, $\mathbf{h}$) = $\mathbf{0}$ and *cov*(ε_j, ε_k) = 0 for $j \neq k$ and k = 1, 2, ⋯, 6). Let $\mathbf{\Lambda} = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{2,4} & \lambda_{2,5} & \lambda_{2,6} \end{bmatrix}$ denote a matrix of factor loadings, where $\lambda_{p,j}$ is a loading parameter relating h_p to z_j (p = 1 or 2). Then, the relationship between $\mathbf{h}$, $\boldsymbol{\varepsilon}$, and $\mathbf{z}$ can be modeled as

$$\mathbf{z} = \mathbf{\Lambda}'\mathbf{h} + \boldsymbol{\varepsilon}, \tag{1.2}$$

which is the reflective model in factor-based SEM. This model represents the causal relationship between latent constructs, measurement errors, and indicators.

The common factors **h** in (1.2) are considered to be equivalent to the latent constructs of interest in factor-based SEM (i.e., **h** = [*health status*, *job satisfaction*]; Rigdon, 2012). Under this assumption, the equation (1.1) can be re-written as

$$h_2 = b_1 h_1 + \zeta, \tag{1.3}$$

which is the structural model in factor-based SEM. Given **Σ** and (1.2), *cov*(h_1, h_2) can be identified, from which b_1 in (1.3) also can be identified (e.g., refer to Bollen, 1989, pp. 238–251). The parameter b_1 in (1.3) represents the causal effect of the latent construct *health status* on the endogenous latent construct *job satisfaction*. If *health status* has no causal effect on *job satisfaction*, b_1 becomes zero. Figure 1.1 visualizes the reflective and structural models of factor-based SEM via a single diagram.

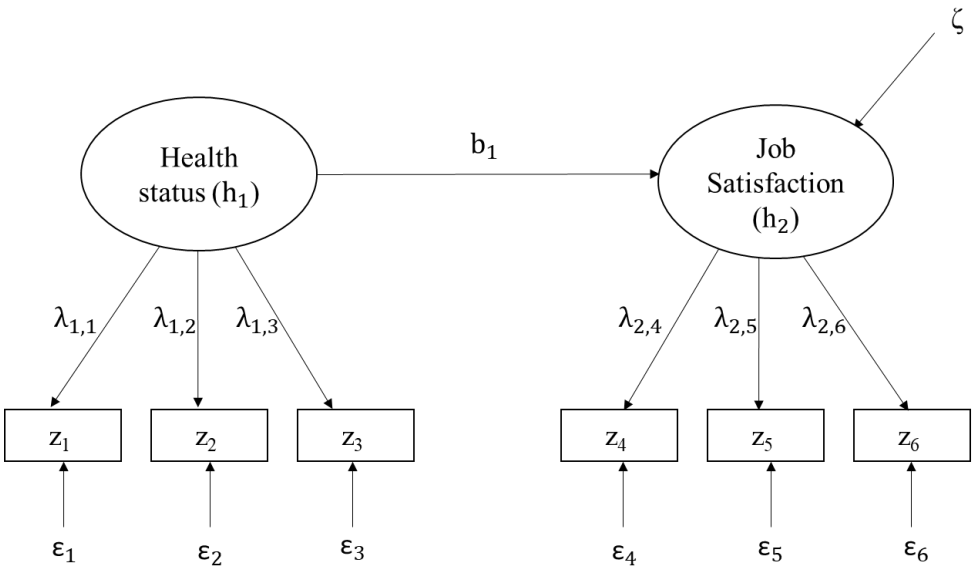


Figure 1.1. An illustration of a factor-based structural equation model. Squares denote indicators, circles represent factors, single-headed arrows correspond to loadings and path coefficients.

Conversely, component-based SEM represents the two constructs by two components, under the assumption that the constructs are descriptive variables that refer to the two

indicator clusters. As mentioned above, each component corresponds to a composite index summarizing the cluster of its indicators, not their underlying cause. Let $\boldsymbol{\gamma} = [\gamma_1, \gamma_2]'$ denote a vector of components for the two sets of indicators, where γ_1 is a component of z_1, z_2, and z_3 and γ_2 is a component of z_4, z_5, and z_6. Let $\mathbf{e} = [e_1, e_2, e_3, e_4, e_5, e_6]'$ denote a vector of residuals, where a residual e_j is a part of z_j that cannot be explained by the component of z_j. Let $\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{2,1} & w_{3,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{4,2} & w_{5,2} & w_{6,2} \end{bmatrix}'$ denote a matrix of weights, where $w_{j,p}$ is assigned to z_j for forming γ_p. Let $\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{2,4} & c_{2,5} & c_{2,6} \end{bmatrix}$ denote a matrix of component loadings, where $c_{p,j}$ is a loading parameter relating γ_p to z_j. Then, the relationships between $\boldsymbol{\gamma}$ and $\mathbf{z}$ can be written as

$$\boldsymbol{\gamma} \equiv \mathbf{W}'\mathbf{z}, \tag{1.4}$$

$$\mathbf{z} = \mathbf{C}'\boldsymbol{\gamma} + \mathbf{e}. \tag{1.5}$$

The two models (1.4) and (1.5) are the weighted relation and component-measurement models, respectively. The weighted relation model (1.4) indicates by which set of indicators each component is defined, whereas the component measurement model (1.5) shows which cluster of indicators each component aims to explain as their summary index.

Component-based SEM equates $\boldsymbol{\gamma}$ in (1.4) with the constructs of interests, that is, $\boldsymbol{\gamma}$ = [*health status*, *job satisfaction*]. Under this assumption, the equation (1.1) can be re-written as

$$\gamma_2 = b_1\gamma_1 + \zeta, \tag{1.6}$$

which is the structural model in component-based SEM. Given $\boldsymbol{\Sigma}$ and the model equations (i.e., (1.4), (1.5), and (1.6)), the path coefficient b_1 is identified at the point where γ_1 maximizes its explanatory power for γ_2 while γ_1 and γ_2 serve as the best summary of their respective indicator clusters (Cho & Choi, 2020). The parameter b_1 in (1.6) summarizes the causal effects of the indicators of γ_1 on the indicators of γ_2, which is illustrated in Appendix A1. If none of the indicators of *health status* has a causal effect on any of the indicators of *job*

satisfaction, b_1 becomes zero. Figure 1.2 displays the weighted relation, component-measurement, and structural models of component-based SEM via a single diagram.

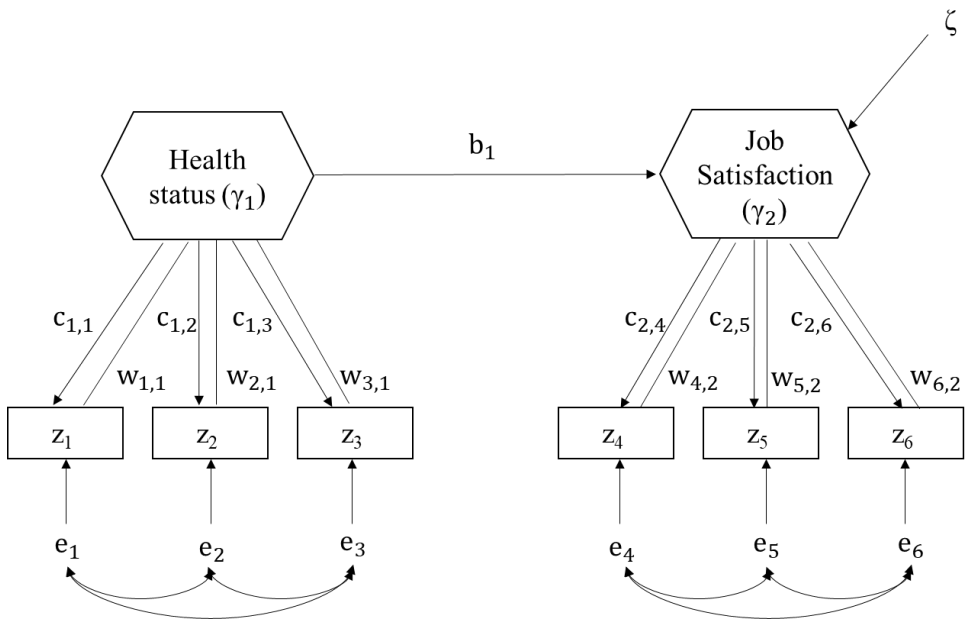


Figure 1.2. An illustration of a component-based structural equation model. Hexagons denote components, double-headed arrows signify correlations, and straight lines represent weights.

In summary, the two SEM domains share a common goal of testing the structural relationship between constructs of interest while deriving statistical representations for these constructs. However, they conceptualize the constructs differently, resulting in their distinct statistical representations and measurement models.

1.3. Discussion About the Superior Domain

In the literature, several researchers had argued for the universal superiority of factor-based SEM over component-based SEM (e.g., Henseler, 2012, p. 402), with some advocating for the abandonment of the latter (e.g., Antonakis et al., 2010; Rönkkö & Evermann, 2013, p. 443). The primary rationale behind this stance was the failure of component-based SEM to account for measurement errors in indicators, which consequently led to biased estimates of model parameters and inflated type I error rate. Implicit in their assumptions was the belief

that component-based SEM shares the same population models as factor-based SEM and therefore aim to estimate parameters in (1.3), rather in (1.6) (e.g., Henseler, 2012; Reinartz et al., 2009). This misconception, which had prevailed even since the initial development stage of component-based methods (e.g., Lohmöller, 1989; Wold, 1982), influenced a substantive body of simulation studies that compared component-based SEM with factor-based SEM in parameter recovery using data generated exclusively from population factor models (e.g., Areskoug, 1982; Goodhue et al., 2006, 2012; Henseler, 2012; Lu et al., 2011; Reinartz et al., 2009). Consequently, these studies inadvertently conveyed a misimpression regarding the biased nature of component-based SEM estimators.

It was not until the 2010s that methodologists began to realize the possibility that component-based SEM might have its unique population model, featuring different statistical representations for constructs, and could not be fairly evaluated under population factor models (e.g., Hwang, Malhotra, et al., 2010; Rigdon, 2012). Integrating diverse perspectives on the population component model (e.g., Cho & Choi, 2020; Dijkstra, 2013b, 2017), Cho et al. (2022) analytically proved the fundamental difference in the population models of the two SEM domains and the necessity of evaluating each domain based on the population models it assumes. In their subsequent comprehensive simulation study, they empirically revealed that each SEM domain excels at recovering parameters of the model whose construct representations align with their respective assumptions.

In conclusion, drawing on the findings of earlier research, there is no universally preferable SEM domain; rather, one domain may be more suited to a researcher's specific theory than the other, depending on the context.

1.4. How to Choose an SEM Domain

To determine which domain may be better aligned with their theories, researchers should begin with clearly defining constructs of interest, particularly in terms of their relationship with the respective indicators. As discussed in the previous sections, the factor-based SEM is preferable if constructs in researchers' hypothesis are theoretically defined as latent variables that explain the covariations of the indicators as their underlying cause. In this case, indicators merely serve as measures of each latent construct, allowing one set of indicators to be *interchangeable* with another, provided that the covariation of the new set can also be attributed to the latent construct of interest (e.g., Bollen, 2011; Bollen & Lennox, 1991). Conversely, component-based SEM is preferable if researchers define each construct as a conceptual tool designed to simply describe an indicator cluster of interest. Under this scenario, indicators of a construct cannot be readily exchanged with others, as the indicators themselves represent sub-domains of the construct.

Table 1.1 illustrates how the two constructs, health status and job satisfaction, can be defined distinctively in accordance with the assumptions of each SEM domain, leading to association with different sets of indicators. In this illustration, when the two constructs are considered latent, they denote an individual's overall, subjective perception of the target psychological property. The items serving as their respective indicators essentially ask about the same psychological attribute despite their varying expressions. Replacing these items with new ones featuring different synonyms or phrases would be relatively straightforward. In contrast, when the two constructs are descriptive, each serves as a comprehensive label encompassing various target properties that are related to one another. Since each indicator is directly linked to a specific facet of a construct, researchers must exercise caution when altering the composition of their indicators.

Table 1.1. An illustration of distinct definitions and indicator sets for identically named constructs.

Construct		Latent	Descriptive/Summary
Health Status	Definition	An individual's perception of their holistic health condition	An individual's health status that encompasses physical, mental, and nutritional well-being.
	Indicators	· How satisfied are you with your health status? · How much control do you feel you have over your health? · How much do you feel that your health interferes with your daily life?	· How often do you experience physical discomfort or pain? · How often do you experience symptoms of depression or anxiety? · How often do you consume a balanced and nutritious diet?
Job Satisfaction	Definition	An individual's subjective evaluation of their overall contentment with their occupational experience	An individual's satisfaction with their job, including compensation, working conditions, job duties, and social relationships with colleagues and supervisors.
	Indicators	· I am satisfied with the work I am presently engaged in. · I find my current work enjoyable. · Unless there are compelling reasons to the contrary, I would like to continue to perform the work I am presently engaged in.	· I am content with my current salary. · I am satisfied with the conditions of my workplace. · I am satisfied with the nature of my present job duties. · I am content with my colleagues and supervisor in my workplace.

In practice, however, it may be unclear to determine which type of construct would be more appropriate for a given indicator cluster of interest, making the choice of the SEM domain challenging. For instance, Table 1.2 displays the correlation matrix of six observed variables, representing children's performance in each school subject (Spearman, 1904). The table shows that children's performance in six school subjects is generally highly correlated. Spearman (1904), the inventor of a factor analysis, hypothesized the existence of a latent construct called general intelligence, which explains the correlations between the six observed variables as their underlying common cause. Conversely, researchers may opt not to

make such an additional assumption and instead treat a general intelligence as an umbrella term that broadly encompasses the children's diverse cognitive abilities measured on the tests (e.g., van der Maas et al., 2014). Given that both approaches can be considered theoretically plausible, researchers may be uncertain about which SEM domain is more appropriate for testing their research hypothesis involving the intelligence construct.

Table 1.2. The correlation matrix for children's performance in six school subjects.

	Classics	French	English	Math	Pitch	Music
Classics	1					
French	0.83	1				
English	0.78	0.67	1			
Math	0.7	0.67	0.64	1		
Pitch	0.66	0.65	0.54	0.45	1	
Music	0.63	0.57	0.51	0.51	0.4	1

In such cases, researchers may empirically examine the relationships between constructs and their indicators. For instance, when researchers have data for multiple, validated indicators of a construct of interest, they can create several subsets of indicators and use each set to contemplate a competing SEM model having a distinct reflective measurement model for the construct, under the assumption that the construct is latent. If the construct truly exists separately from their indicators as a latent variable, allowing the indicators to be exchangeable, the correlation estimates between the construct and the others should not vary depending on the choice of indicator subsets unless sampling error occurs (Widaman, 2018). Consequently, if a change in indicator composition for a construct greatly alters its relevant correlation estimates, it can provide strong evidence to reject the latent existence of the construct, and researchers may consider employing an alternative component-based SEM accordingly. When the component-based SEM is applied, a change in indicators is expected to lead to a change in correlation estimates associated with the

construct, because indicators themselves constitute the construct in component-based SEM. However, since there is no clear criterion for determining a significant difference in correlation estimates between competing models, it would be necessary to develop such a statistical test for identifying construct types in the structural equation model.

1.5. Objectives and Overview of the Dissertation

In Chapter 1, this dissertation introduced the two SEM domains—factor-based and component-based—and systematically compared them using examples to illustrate their distinct applications. In essence, both domains represent comprehensive, complementary SEM approaches, each grounded in unique premises regarding constructs and their relationships with indicators. Considering their respective utility in empirical research, it is important to continue technical development in both SEM domains. As such, this dissertation aims to identify two long-standing problems within each SEM domain and present advanced SEM methods capable of addressing these issues.

Chapter 2 delves into two persistent issues in factor-based SEM and offers a new technical solution to address these challenges. The chapter begins with a succinct overview of Jöreskog's (1973, 1978) covariance-based approach (JCA), which has been widely regarded as the standard method in factor-based SEM (e.g., Bollen, 2019). As the name suggests, this method primarily relies on the covariance matrix of indicators throughout the analysis process, including identification, parameter estimation, and model evaluation. However, this characteristic results in two significant limitations: the emergence of improper solutions (e.g., negative variance estimates) and the absence of a statistical tool to tackle the factor score indeterminacy problem (i.e., an infinite number of factor score sets possibly being true given the data).

The first limitation poses a challenge for applied researchers, as the occurrence of improper solutions renders all parameter estimates in the model unreliable, preventing researchers from drawing conclusions about their model specification (e.g., McDonald, 2004; Newsom, 2014). The second limitation makes the method less than ideal for researchers who require probabilistic inferences about individuals' true factor scores (e.g., the estimation of the probability that an individual's true factor score is higher than another's). To address these issues, Chapter 2 introduces an alternative factor-based SEM method, termed *structured factor analysis* (SFA; Cho and Hwang, 2023). Unlike JCA, SFA incorporates the score matrix of indicators and factors as the central component of the entire analysis process, which precludes the emergence of improper solutions and facilitates probabilistic inferences about individuals' true factor scores. Empirical data analyses are conducted to demonstrate the relative advantages of SFA compared to JCA. This chapter was published in the journal *Structural Equation Modeling: A Multidisciplinary Journal* in 2023.

Chapters 3 and 4 address two enduring limitations within component-based SEM and suggest innovative technical extensions of GSCA designed to surmount these obstacles. These two chapters commence with a brief description of GSCA. Although PLSPM, also known as PLS-SEM, tends to be more prevalent in application studies compared to GSCA (e.g., Cho, Schlaegel, et al., 2022), the focus here is primarily on GSCA, given its global, interpretable criterion for generating composite indexes, which makes the technical extension of the method more feasible. In contrast, PLSPM lacks such a criterion, leaving it uncertain how optimal the composite indexes generated from PLSPM are. In addition, in several simulation studies, GSCA has demonstrated comparable or even superior results to PLSPM in terms of parameter recovery and prediction accuracy (Cho et al., 2023; Cho, Sarstedt, et al., 2022; Cho & Choi, 2020; Hwang, Malhotra, et al., 2010).

In Chapter 3, the focus shifts to a limitation in GSCA's treatment of indicator scores, which involves standardizing all indicators' scores to have zero means and unit variances. In line with this standardization procedure, GSCA generates composite indexes in such a way that their scores are also standardized to have zero means and unit variances. While this simplification aids in the interpretation of the GSCA model, it prevents the generation of composite indexes that are interpretable on the indicators' original scales. To tackle this problem, Chapter 3 proposes a novel extension of GSCA, named *convex generalized structured component analysis* (convex GSCA). Unlike the conventional GSCA, this method retains the raw data for indicators when their measurement scales are identical within their respective blocks. It then generates composite indexes, known as *convex components*, which carry non-negative weights for their respective indicators, summing to one. The chapter elucidates how the scores of convex components and their associated path coefficients in the structural model can be interpreted in terms of the original scales of their indicators. The practical utility of convex GSCA is demonstrated through simulation and real data analyses. This chapter was first submitted to the journal *Psychometrika* in December 2021 and is currently in its second round of review.

Chapter 4 concentrates on another limitation of GSCA, namely its restriction on the functional form of components to be linear. In other words, every composite index in GSCA is always defined as a weighted sum of its indicators. Such linear constraint may potentially inhibit components from capturing non-linear association between their indicators, suggesting that the components might not serve as the optimal summary index of their indicators, thereby diminishing their predictive power for other components. To overcome this limitation, the chapter introduces a new method, termed *deep learning generalized structured component analysis* (DL-GSCA). This approach integrates deep learning and GSCA into a single framework, allowing for the identification of the best functional form of components in a

data-driven manner. It aims to maximize the predictive power of components while ensuring the interpretability of their interconnected relationship. Through the application of real and simulated data analyses, the chapter demonstrates the relative advantage of DL-GSCA over GSCA, particularly when non-linear associations exist between indicators per component. This chapter has been recently accepted for publication in the journal *Structural Equation Modeling: A Multidisciplinary Journal* as of July 2023.

Chapter 5 provides a comprehensive summary of the preceding chapters, highlighting the significance and implications of the proposed methods. It also offers a critical evaluation of potential limitations inherent in the current studies and outlines prospective directions for future research in this area.

References

American Psychological Association. (2007). *Report of the APA Task Force on Socioeconomic Status*. https://www.apa.org/pi/ses/resources/publications

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010

Areskoug, B. (1982). The first canonical correlation: Theoretical PLS analysis and simulation experiments. In H. Wold & K. G. Jöreskog (Eds.), *Systems under indirect observation: causality, structure, prediction* (pp. 95–118). North Holland.

Binning, J. F. (2015). Construct. In *Britannica*. https://www.britannica.com/science/construct

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. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, *35*(2), 359–372. https://doi.org/10.2307/23044047

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., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314. https://doi.org/10.1037/0033-2909.110.2.305

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Brandt, V., Zhang, Y., Carr, H., Golm, D., Correll, C. U., Arrondo, G., Firth, J., Hassan, L., Solmi, M., & Cortese, S. (2023). First evidence of a general disease ("d") factor, a common factor underlying physical and mental illness. *World Psychiatry*, *22*(2), 335–337. https://doi.org/10.1002/wps.21097

Cho, G., & Choi, J. Y. (2020). An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation models. *Behaviormetrika*, *47*(1), 243–272. https://doi.org/10.1007/s41237-019-00098-0

Cho, G., & Hwang, H. (2023). Structured factor analysis: A data matrix-based alternative approach to structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(3), 364–377. https://doi.org/10.1080/10705511.2022.2126360

Cho, G., Kim, S., Hwang, H., Lee, J., Sarstedt, M., & Ringle, C. M. (2022). A comparative study of the predictive power of component-based approaches to structural equation modeling. *European Journal of Marketing*. https://doi.org/10.1108/EJM-07-2020-0542

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

Cho, G., Schlaegel, C., Hwang, H., Choi, Y., Sarstedt, M., & Ringle, C. M. (2022). Integrated generalized structured component analysis: On the use of model fit criteria in international management research. *Management International Review*. https://doi.org/10.1007/s11575-022-00479-w

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons

of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, *114*(1), 174–184. https://doi.org/10.1037/0033-2909.114.1.174

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.

Dijkstra, T. K. (2013). *Composites as factors, generalized canonical variables revisited*. https://doi.org/10.13140/RG.2.1.3426.5449

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

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. https://doi.org/10.1037/1082-989X.5.2.155

Fatima, S., & Sheikh, H. (2014). Socioeconomic status and adolescent aggression: The role of executive functioning as a mediator. *The American Journal of Psychology*, *127*(4), 419–430. https://doi.org/10.5406/amerjpsyc.127.4.0419

Goldberger, A. S. (1964). *Econometric theory.* New York: John Wiley & Sons.

Goodhue, D. L., Lewis, W., & Thompson, R. (2006). PLS, small sample size, and statistical power in MIS research. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, *8*, 202b-202b. https://doi.org/10.1109/HICSS.2006.381

Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS Have advantages for small sample size or non-normal data? *MIS Quarterly*, *36*(3), 981–1001. https://doi.org/10.2307/41703490

Hayduk, L. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement*, *74*(6), 905–926. https://doi.org/10.1177/0013164414527449

Heise, D. R. (1972). Employing nominal variables, induced variables, and block variables in path analyses. *Sociological Methods & Research*, *1*(2), 147–173. https://doi.org/10.1177/004912417200100201

Henseler, J. (2012). Why generalized structured component analysis is not universally preferable to structural equation modeling. *Journal of the Academy of Marketing Science*, *40*(3), 402–413. https://doi.org/10.1007/s11747-011-0298-6

Hoyle, R. H. (2014). *Handbook of structural equation modeling* (1st ed.). The Guilford Press.

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. (2004). Generalized structured component analysis. *Psychometrika*, *69*(1), 81–99. https://doi.org/10.1007/BF02295841

Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling*. Chapman and Hall/CRC Press.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, *23*(2), 121–145. https://doi.org/10.1111/j.2044-8317.1970.tb00439.x

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system.

In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 255–284). Seminar Press.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4), 443–477. https://doi.org/10.1007/BF02293808

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press. https://psycnet.apa.org/record/2015-56948-000

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares.* Physica. https://doi.org/10.1007/978-3-642-52512-4

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

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*(2), 95–107. https://doi.org/10.1037/h0056029

McDonald, R. P. (2004). Respecifying improper structures. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(2), 194–209. https://doi.org/10.1207/s15328007sem1102_3

Newsom, J. T. (2014). *Improper solutions in SEM*. https://web.pdx.edu/~newsomj/semclass/ho_improper.pdf

Nimon, K., Henson, R. K., & Gates, M. S. (2010). Revisiting interpretation of canonical correlation analysis: A tutorial and demonstration of canonical commonality analysis. *Multivariate Behavioral Research*, *45*(4), 702–724. https://doi.org/10.1080/00273171.2010.498293

Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy

of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, *26*(4), 332–344. https://doi.org/10.1016/j.ijresmar.2009.08.001

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30–45. https://doi.org/10.1037/met0000220

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, *45*(5–6), 341–358. https://doi.org/10.1016/j.lrp.2012.09.010

Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, *16*(3), 425–448. https://doi.org/10.1177/1094428112474693

Rosseel, Y., & Loh, W. W. (2022). A structural after measurement (SAM) approach to structural equation modeling. *Psychological Methods*. https://doi.org/10.1037/met0000503

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563–575. https://doi.org/10.1007/BF02296196

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Quality Management and Business Excellence*, *19*(7–8), 871–886. https://doi.org/10.1080/14783360802159543

van der Maas, H. L. J., Kan, K.-J., & Borsboom, D. (2014). Intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence*, *2*(1), 12–15. https://doi.org/10.3390/jintelligence2010012

Widaman, K. F. (2018). On common factor and principal component representations of data:

Implications for theory and for confirmatory replications. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 829–847. https://doi.org/10.1080/10705511.2018.1478730

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) Modelling: Some current developments. In P. R. Krishnaiah (Ed.), *Multivariate analysis–III* (pp. 383–407). Academic Press. https://doi.org/10.1016/B978-0-12-426653-7.50032-6

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction, part II* (pp. 1–54). North Holland.

Wold, H. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences, Vol. 6* (pp. 581–591). Wiley.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368. https://doi.org/10.1080/01621459.1962.10480664

Chapter 2. Structured Factor Analysis: A Data Matrix-Based Alternative Approach to Structural Equation Modeling

Publication: Cho, G., & Hwang, H. (2023). Structured factor analysis: A data matrix-based alternative approach to structural equation modeling, *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(3), 364–377, https://doi.org/10.1080/10705511.2022.2126360

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.

2.1. 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 2.2, we describe the two stages of SFA. In Section 2.3, 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 Appendix B. In Section 2.4, 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 2.5, we illustrate an application of SFA to a real dataset. In Section 2.6, we discuss the implications of the proposed method and its potential extensions.

2.2. 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 B.

2.2.1. 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

$$\mathbf{Z} = \mathbf{H}_{true}\mathbf{\Lambda} + \mathbf{E}_{true}. \tag{2.1}$$

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 (2.1) can be re-expressed as

$$\mathbf{Z} = \mathbf{F}_{true}\mathbf{L}, \tag{2.2}$$

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 $\mathbf{\Lambda}$. SFA seeks to simultaneously estimate $\mathbf{L}$, $\mathbf{\Lambda}$, and $\mathbf{F}_{true}$ given $\mathbf{Z}$. However, even if we knew the values of $\mathbf{L}$ and $\mathbf{\Lambda}$, 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 $cov(\mathbf{F}) = \mathbf{\Lambda}$, 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 B1 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 $\mathbf{S}$. $\mathbf{Z}_{std}$ Let $\widehat{\mathbf{Z}}$ 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.

$$\begin{aligned} \rho &= (JN)^{-1} SS(\mathbf{Z}_{std} - \widehat{\mathbf{Z}}) \\ &= (JN)^{-1} SS(\mathbf{Z}_{std} - \mathbf{FL}), \end{aligned} \tag{2.3}$$

subject to $mean(\mathbf{F}) = \mathbf{0}$ and $cov(\mathbf{F}) = \mathbf{\Delta}$, where $SS(\mathbf{X}) = tr(\mathbf{X}'\mathbf{X})$ for any matrix $\mathbf{X}$. The value of the cost function (2.3) 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 (2.3). 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 B4. 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 (2.3) under the identified measurement model, it provides unbiased estimates of the measurement model parameters, as proved in Appendix B5. 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 (2.3) 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 B6. This

procedure indicates that even if the cost function (2.3) 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

$$IPE_O = SS(\mathbf{Z}_{std} - \hat{\mathbf{F}}\hat{\mathbf{L}})/SS(\mathbf{Z}_{std}), \tag{2.4}$$

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 (2.3) computed based on $\hat{\mathbf{F}}$ and $\hat{\mathbf{L}}$. The value of $1 - IPE_O$ 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 $mean(\hat{\mathbf{F}}) = \mathbf{0}$, $\mathbf{Z}_{std} = \hat{\mathbf{F}}\hat{\mathbf{L}}$, and $cov(\hat{\mathbf{F}}) = \mathbf{\Delta}$, given that the measurement model is identified (see Appendix B5). As $\hat{\mathbf{F}}$ obtained from SFA always satisfies $mean(\hat{\mathbf{F}}) = \mathbf{0}$, 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 $cov(\hat{\mathbf{F}}) \neq \mathbf{\Delta}$.

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 $cov(\hat{\mathbf{F}}) = \mathbf{\Delta}$ 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.

2.2.2. 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

$$\mathbf{H}_{Y,true} = \mathbf{H}_{X,true}\mathbf{B}_X + \mathbf{H}_{Y,true}\mathbf{B}_Y + \mathbf{Q}_{true}. \quad (2.5)$$

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 (2.5) 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 $\widehat{\mathbf{H}}_X$ and $\widehat{\mathbf{H}}_Y$ 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 $\mathbf{\Psi}$) while treating [$\widehat{\mathbf{H}}_X$, $\widehat{\mathbf{H}}_Y$] 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 B7. 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

$$IPE_L = SS(\widehat{\mathbf{H}}_Y - (\widehat{\mathbf{H}}_X\widehat{\mathbf{B}}_X + \widehat{\mathbf{H}}_Y\widehat{\mathbf{B}}_Y))/SS(\widehat{\mathbf{H}}_Y). \quad (2.6)$$

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

2.3. Candidate Factor Score Distribution

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

As in Section 2.2, 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 B8. Let $\mathrm{F}_{N,T}$ 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 $\mathrm{F}_{N,T}$ with two parameter matrices **W** and **G**. The parameter matrix **W** denotes a J by T matrix of weights that determines the center of the candidate factor score distribution as **ZW**, whereas **G** denotes a T by T covariance matrix of measurement errors when **ZW** is used as a measurement of $\mathbf{F}_{true}$. Using **ZW** as a measurement of $\mathbf{F}_{true}$ can be justified in that **ZW** is the only part of $\mathbf{F}_{true}$ that can be inferred from **Z**, as shown in Appendix B9, and can be considered the best linear predictor for $\mathbf{F}_{true}$ given **Z** (e.g., Bartholomew, 1981) because **W** is equivalent to the weight matrix obtained by regressing $\mathbf{F}_{true}$ on $\mathbf{Z}_{std}$ (Thurstone, 1934). The matrix **ZW** is called *a matrix of expected candidate factor scores* in SFA. The square roots of **G**'s diagonal entries are the standard deviations of measurement errors when **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 B11 and Appendix B12, SFA estimates the candidate factor score distribution by minimizing the same cost function (2.3) in a least squares sense. More specifically, let $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$ denote an estimate of $\boldsymbol{\mathcal{F}}_{N,T}$, which can be expressed with $\widehat{\mathbf{L}}$ and $\widehat{\boldsymbol{\Delta}}$ obtained from the first stage. Then, SFA randomly samples a prescribed number of $\widehat{\mathbf{F}}$s from $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$ and uses a set of the sampled $\widehat{\mathbf{F}}$ 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.

2.4. 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.

2.4.1. Simulation Study 1

As depicted in Figure 2.1, 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 pth latent variable ($p = 1, 2, \cdots, P$), whereas $\mathbf{z}_{\cdot j}$ denotes N individuals' scores on the jth indicator ($j = 1, 2, \cdots, J$). 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 2.1.

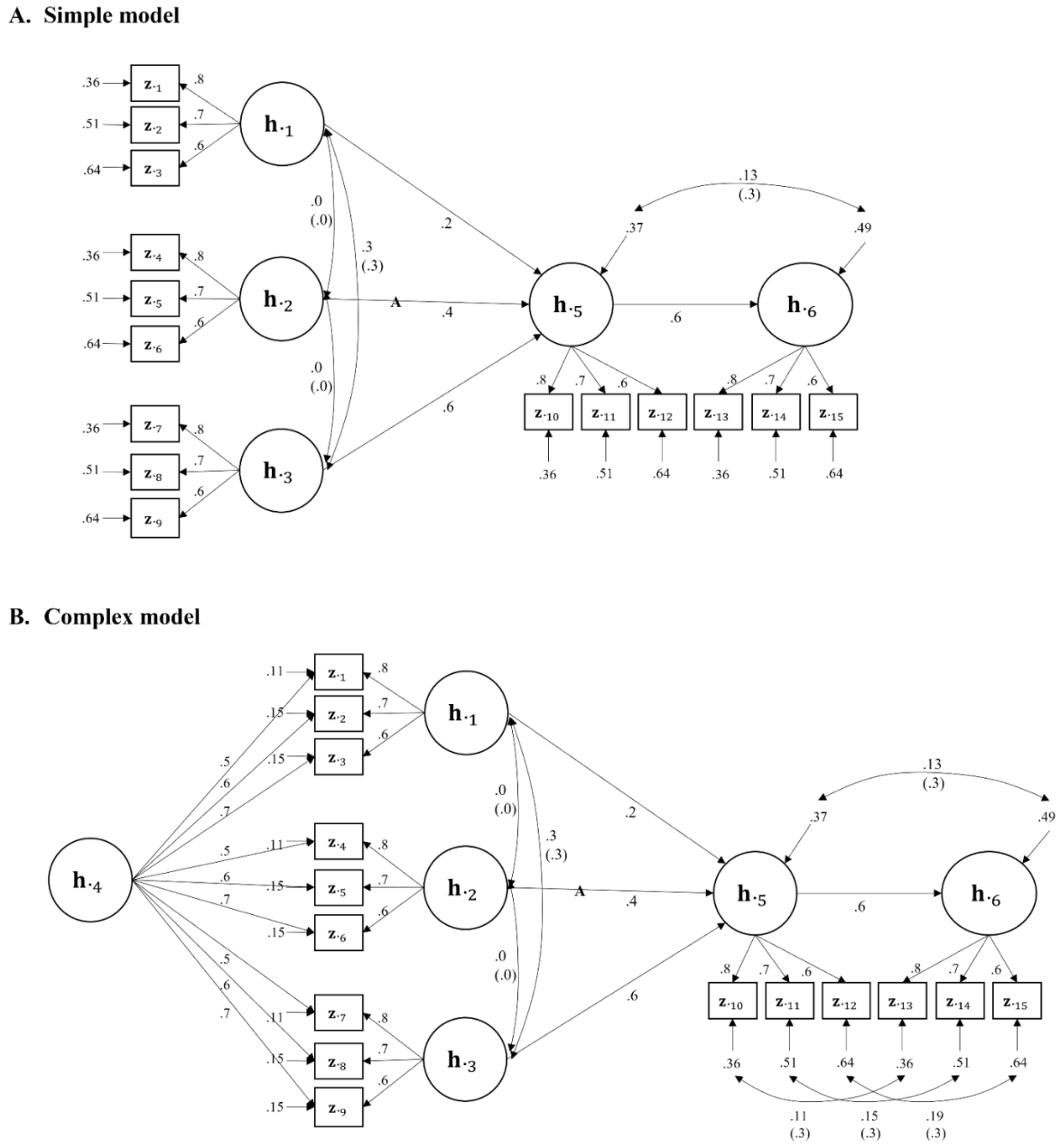


Figure 2.1. The two data generating models used for the simulation studies. Circles and squares signify latent variables and indicators, respectively. The arrow labeled A denotes the path coefficient excluded in the misspecified model. The values in the parentheses denote correlations.

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 (2.1). 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 $cov(\mathbf{F}_{true}) = \mathbf{\Delta}$, 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 2.1, 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

[1] This MATLAB package is available at https://sfaprime.wordpress.com/.

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 $\mathbf{h}_{\cdot 4}$ is assumed to be uncorrelated with $\mathbf{h}_{\cdot 1}$, $\mathbf{h}_{\cdot 2}$, and $\mathbf{h}_{\cdot 3}$ in the model.

Table 2.1 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.

Table 2.1. 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

Estimator	*N*	Correct model				Misspecified model			
		Simple		Complex		Simple		Complex	
		Normal	Non-normal	Normal	Non-normal	Normal	Non-normal	Normal	Non-normal
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

Tables 2.2 and 2.3 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, Malhotra, et al., 2010). The average biases and RMSE values for the same sets of model parameters under each distributional condition are provided in Table S2.1 to Table S2.4. Furthermore, the bias and RMSE values of the two estimators for each individual parameter per condition are provided in Table S2.5 to Table S2.8.

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.

Table 2.2. 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 2.3. 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.

2.4.2. 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 2.4 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 $N = 1000$.

Table 2.4. 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					
			$\mathbf{h}_{\cdot 1}$	$\mathbf{h}_{\cdot 2}$	$\mathbf{h}_{\cdot 3}$	$\mathbf{h}_{\cdot 4}$	$\mathbf{h}_{\cdot 5}$	$\mathbf{h}_{\cdot 6}$	$\mathbf{h}_{\cdot 1}$	$\mathbf{h}_{\cdot 2}$	$\mathbf{h}_{\cdot 3}$	$\mathbf{h}_{\cdot 4}$	$\mathbf{h}_{\cdot 5}$	$\mathbf{h}_{\cdot 6}$
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

2.5. 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 2.2, 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.

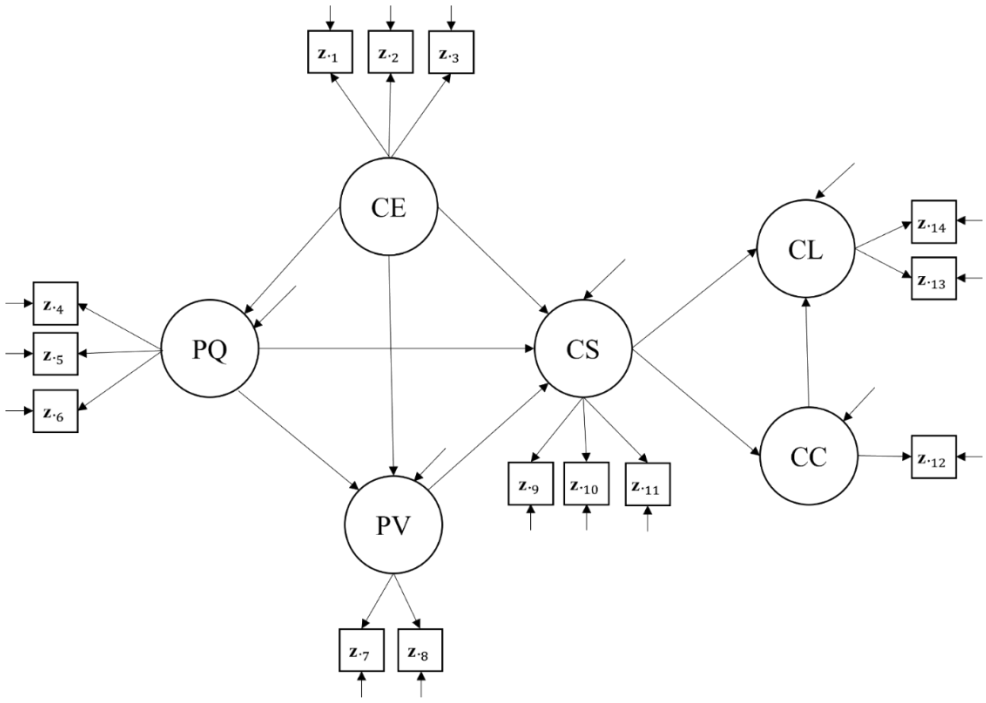


Figure 2.2. The American customer satisfaction (ACSI) model.

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 2.5 and 2.6 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)(Fornell et al., 1996).

Table 2.5. The loading estimates, their standard errors (SE), and 95% confidence intervals (CI) in the ACSI model obtained from SFA and JCA-ML.

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

Table 2.6. The path coefficient estimates, their standard errors (SE), and 95% confidence intervals (CI) in the ACSI model obtained from SFA and JCA-ML.

	SFA			JCA-ML		
Path	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

	SFA			JCA-ML		
Path	Estimate	SE	95% CI	Estimate	SE	95% CI
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

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 B11. For illustration, we show the marginal distributions of the candidate factor score estimates for three companies, labeled ID1, 2, and 21, in Figure 2.3. 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]).

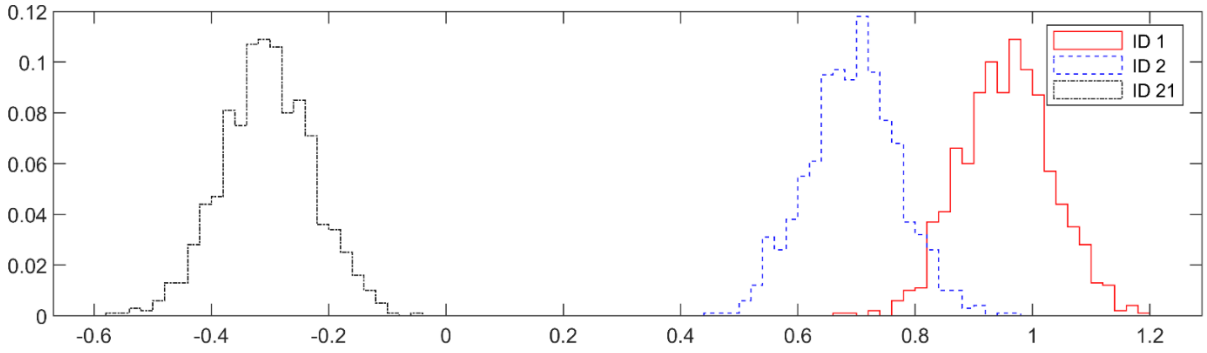


Figure 2.3. Marginal distributions of the candidate factor scores of customer satisfaction for three companies labeled ID1, ID2, and ID21.

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 2.4 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.

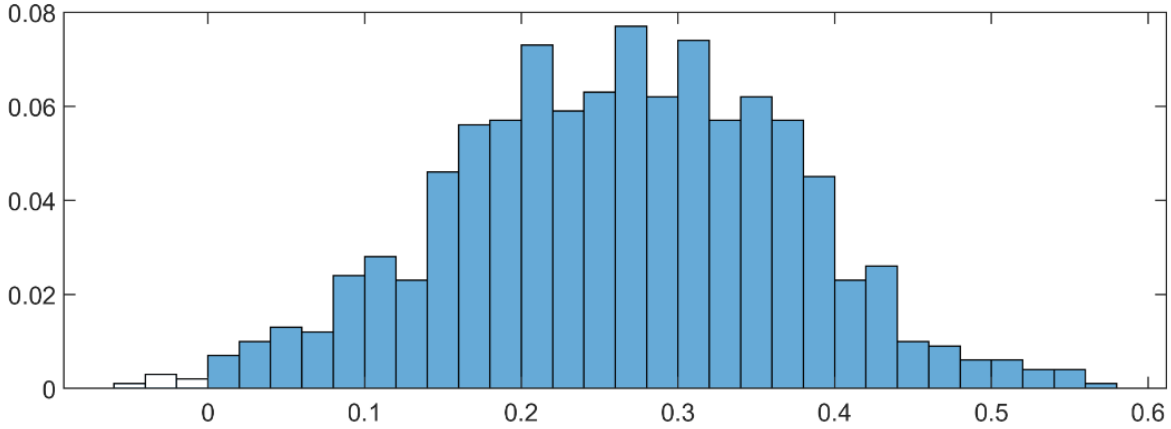


Figure 2.4. The distribution of the differences in the candidate factor scores of customer satisfaction between two companies ID1 and ID2. The colored area represents the probability for ID1 to have a higher customer satisfaction score than ID2.

2.6. 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, Sarstedt, 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-fledge 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, Cho, Jung, 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://deleeuwpdx.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.

Chapter 3. Generalized Structured Component Analysis Accommodating Convex Components: A Knowledge-Based Multivariate Method with Interpretable Composite Indexes

Manuscript: Cho, G., & Hwang, H. Generalized structured component analysis accommodating convex components: A knowledge-based multivariate method with interpretable composite indexes. Under review at *Psychometrika*.

Abstract

Generalized structured component analysis (GSCA) is a multivariate method for examining theory-driven relationships between variables including components. GSCA can provide the deterministic component score for each individual once model parameters are estimated. As the traditional GSCA always standardizes all indicators and components, however, it could not utilize information on the indicators' scale in parameter estimation. Consequently, its component scores could just show the relative standing of each individual for a component, rather than the individual's absolute standing in terms of the original indicators' measurement scales. In the paper, we propose a new version of GSCA, named *convex GSCA*, which can produce a new type of unstandardized components, termed convex components, which can be intuitively interpreted in terms of the original indicators' scales. We investigate the empirical performance of the proposed method through the analyses of simulated and real data.

Keywords: Generalized structured component analysis, convex component, multivariate analysis, composite index, interpretability

3.1. Introduction

Generalized structured component analysis (GSCA; Hwang & Takane, 2004, 2014) is a multivariate method that allows for specifying and testing path-analytic relationships between observed variables and components (i.e., weighted sums of observed variables). Observed variables forming components are called composite indicators (Bollen & Bauldry, 2011). Given a theory-driven model, GSCA constructs components from composite indicators such that the components can explain the total variances of all dependent variables in the model as much as possible.

As in many component analysis techniques, GSCA has typically assumed that all components and indicators were standardized to have zero means and unit variances. This traditional, standardized version of GSCA shall be called GSCA$_{std}$ hereafter. GSCA$_{std}$ begins by standardizing indicators prior to estimating parameters and updates component weights in such a way that they produce standardized components during the estimation process. Such standardization can be useful for the interpretation and comparison of GSCA$_{std}$'s estimates because the GSCA$_{std}$ model is equivalent to a system of multiple regression equations for standardized components and indicators, indicating that its loadings and path coefficients can be interpreted as standardized regression coefficients.

Nonetheless, the conventional standardization of components makes it difficult to interpret *component scores* in terms of the original indicators' measurement scales. The standardized component score for an individual merely shows the individual's relative location to the other individuals in the sample and the absolute score itself is not interpretable. This is less attractive to researchers who are interested in the absolute level of a component for each individual. For example, if a standardized component is used to measure the level of life satisfaction, an individual's component score can inform whether s/he has a relatively lower or higher level of life satisfaction than the others. However, it cannot tell exactly what

the level of her/his life satisfaction is, reflecting whether s/he is satisfied or dissatisfied with her/his life.

Moreover, if indicators for each component are measured on the same scale, which is often observed in practice, standardizing the indicators may not be recommended because it can eliminate "the natural and relevant variability present" (Naik & Khattree, 1996) in each of the indicators, forcing them to have the same variance, although their variances may not be the same in reality. For illustration, suppose that we made two versions of test batteries to assess children's intelligence, both of which were measured on a 0 to 100 scale. Three children took these tests and obtained {49, 50, 51} for Test 1 and {0, 50, 100} for Test 2. The results show that Test 1 almost fails to differentiate the children's intelligence levels, whereas Test 2 differentiates their intelligence level very well, indicating that the difference in score variability between the two tests is interpretable and contains meaningful information. However, when we standardize these scores, such information disappears since both score sets become identical (i.e., {–1, 0, 1}). If $GSCA_{std}$ is applied to the tests, the same standardized weight values (i.e., .5) will be assigned to the two tests.

To obtain unstandardized component scores from original indicators, $GSCA_{std}$ applies an additional rescaling of weight estimates after convergence (Hwang & Takane, 2014, Chapter 2). As will be discussed in more detail in Section 3.2.2, each indicator's weight estimate is rescaled by dividing it by the indicator's standard deviation. Subsequently, unstandardized component scores are obtained by pre-multiplying the rescaled weights by their indicators' original scores.

However, this rescaling procedure has two issues. Firstly, the procedure is carried out while keeping the variances of components fixed to one. Thus, the variances of the resultant unstandardized components are likely to be different from those of the original indicators, so that it is not guaranteed that the unstandardized component scores would vary within the

same range of the original indicators. Secondly, the rescaling procedure tends to have indicators with relatively small variances influence the construction of their unstandardized component more heavily. In the above example, as the sample standard deviations of the two test batteries were 1 and 50, the unstandardized weights obtained from this ad-hoc rescaling procedure would be .5 and .01 for Tests 1 and 2, respectively. This indicates that Test 1 is 50 times more influential for forming children's unstandardized component scores than Test 2, even though Test 2 differentiates children's intelligence levels much better than Test 1. In Section 3.2.2, we will explain why this issue occurs in the rescaling procedure.

To address these issues, we propose a different version of GSCA, named *convex GSCA* or GSCA$_{cvx}$ for short, which can estimate unstandardized components of original indicators. Specifically, GSCA$_{cvx}$ obtains an unstandardized component as a convex combination of original indicators, termed a *convex component*, if the indicators for the component have the same measurement scale. A convex combination of a set of vectors refers to a special linear combination whose weights are non-negative and summed up to one (Lay et al., 2015, Chapter 8). As will be shown in Section 3.3, a convex component's scores are within the same range of its indicators' scores. This property of the convex component facilitates the interpretation of its component scores with reference to the indicators' scales. Moreover, GSCA$_{cvx}$ avoids the unnecessary standardization of indicators when they are on the same measurement scales, allowing for utilizing information on their variances in parameter estimation.

The remaining sections of the paper are organized as follows. In Section 3.2, we briefly describe GSCA$_{std}$ and explain its ad-hoc procedure of computing unstandardized components and the procedure's limitation. In Section 3.3, we introduce a convex component and explain its six properties. In Section 3.4, we present the GSCA$_{cvx}$ model that accommodates convex components and propose an iterative algorithm for estimating model

parameters. We also provide a set of overall goodness-of-fit and cross-validation indexes for model evaluation and comparison. In Section 3.5, we conduct a Monte-Carlo simulation study to examine GSCA$_{cvx}$'s parameter recovery. In Section 3.6, we apply GSCA$_{cvx}$ to real data to demonstrate its practical usefulness. In Section 3.7, we summarize the previous sections and discuss the method's implications and prospective extensions.

3.2. Traditional GSCA with Standardized Variables

3.2.1. Model and Parameter Estimation

GSCA$_{std}$ involves three sub-models—weighted relation, component measurement, and structural models (Hwang & Takane, 2004, 2014). Let $\mathbf{z}_{std} = [z_{std,1}, z_{std,2}, \cdots, z_{std,J}]'$ denote a J by 1 random vector of standardized indicators, where $z_{std,j}$ is the jth standardized indicator, i.e., $E(z_{std,j}) = 0$ and $var(z_{std,j}) = 1$ ($j = 1, 2, \cdots, J$). The mean of $\mathbf{z}_{std}$ is a zero vector, and the correlation matrix of $\mathbf{z}_{std}$ is denoted by $\mathbf{\Sigma}_{std}$. Let $\boldsymbol{\gamma}_{std} = [\gamma_{std,1}, \gamma_{std,2}, \cdots, \gamma_{std,P}]'$ denote a P by 1 random vector of standardized components, where $\gamma_{std,p}$ is the pth standardized component, i.e., $E(\gamma_{std,p}) = 0$, $var(\gamma_{std,p}) = 1$ ($p = 1, 2, \cdots, P$). Let $\mathbf{W}_{std}$ denote a J by P matrix consisting of component weights assigned to indicators. Let $\mathbf{C}_{std}$ denote a P by J matrix of loadings relating components to indicators. Let $\mathbf{B}_{std}$ denote a P by P matrix of path coefficients relating components to each other. Let $\boldsymbol{\xi} = [\xi_1, \xi_2, \cdots, \xi_J]'$ denote a J by 1 random vector of errors in the component measurement model, where ξ_j is an error for the jth indicator. Let $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \cdots, \zeta_P]'$ denote a P by 1 random vector of errors in the structural model, where ζ_p is an error for the pth component. The three sub-models of GSCA$_{std}$ are expressed as follows.

$$\boldsymbol{\gamma}_{std} \equiv \mathbf{W}_{std}'\mathbf{z}_{std} \text{ (weighted relation model)} \quad (3.1)$$

$$\mathbf{z}_{std} = \mathbf{C}_{std}'\boldsymbol{\gamma}_{std} + \boldsymbol{\xi} \text{ (component measurement model)} \quad (3.2)$$

$$\boldsymbol{\gamma}_{std} = \mathbf{B}_{std}'\boldsymbol{\gamma}_{std} + \boldsymbol{\zeta} \text{ (structural model).} \quad (3.3)$$

The weighted relation model (3.1) shows that (standardized) components are defined as a linear combination of standardized indicators. The component measurement and structural models (3.2) and (3.3) express the directional relationships between the indicators and components and those among the components, respectively. As (3.2) and (3.3) can be seen as systems of linear regression equations, their model parameters, including loadings and path coefficients, can be interpreted in the same manner as standardized regression coefficients. The three sub-models are combined into the following equation,

$$\begin{aligned} &[\mathbf{z}_{std}; \boldsymbol{\gamma}_{std}] = [\mathbf{C}_{std}, \mathbf{B}_{std}]'\boldsymbol{\gamma}_{std} + [\boldsymbol{\xi}; \boldsymbol{\zeta}] \\ &\leftrightarrow [\mathbf{I}_J, \mathbf{W}_{std}]'\mathbf{z}_{std} = [\mathbf{C}_{std}, \mathbf{B}_{std}]'\mathbf{W}_{std}'\mathbf{z}_{std} + [\boldsymbol{\xi}; \boldsymbol{\zeta}] \\ &\leftrightarrow \mathbf{V}_{std}'\mathbf{z}_{std} = \mathbf{A}_{std}'\mathbf{W}_{std}'\mathbf{z}_{std} + \mathbf{e}, \end{aligned} \tag{3.4}$$

where $\mathbf{I}_J$ is the identity matrix of order J, $\mathbf{V}_{std} \equiv [\mathbf{I}_J, \mathbf{W}_{std}]$, $\mathbf{A}_{std} \equiv [\mathbf{C}_{std}, \mathbf{B}_{std}]$, $\mathbf{e} \equiv [\boldsymbol{\xi}; \boldsymbol{\zeta}]$, and a semicolon within brackets is an operator to vertically concatenate two vectors in the array. The equation (3.4) is called the GSCA$_{std}$ model.

Let $\mathbf{1}_Q$ denote a column vector of Q ones. Let $SS(\mathbf{X}) \equiv tr(\mathbf{X}'\mathbf{X})$ for any matrix $\mathbf{X}$. Let *vecdiag*() denote an operator that returns a column vector stacking the diagonal elements of a square matrix one below another. GSCA$_{std}$ estimates model parameters ($\mathbf{W}_{std}$ and $\mathbf{A}_{std}$) by minimizing the following objective function

$$\begin{aligned} &f_{std}(\mathbf{W}_{std}, \mathbf{A}_{std}) \\ &\quad = tr(E(\mathbf{e}_{std}\mathbf{e}_{std}')) \\ &\quad = E(SS([\mathbf{z}_{std}; \boldsymbol{\gamma}_{std}]' - \mathbf{z}_{std}'\mathbf{W}_{std}\mathbf{A}_{std}), \\ &\quad = E(SS(\mathbf{z}_{std}'([\mathbf{I}_J, \mathbf{W}_{std}] - \mathbf{W}_{std}\mathbf{A}_{std})) \\ &\quad = tr((\mathbf{V}_{std} - \mathbf{W}_{std}\mathbf{A}_{std})'\boldsymbol{\Sigma}_{std}(\mathbf{V}_{std} - \mathbf{W}_{std}\mathbf{A}_{std})) \end{aligned} \tag{3.5}$$

subject to $vecdiag(\mathbf{W}_{std}'\boldsymbol{\Sigma}_{std}\mathbf{W}_{std}) = \mathbf{1}_p$. Thus, GSCA$_{std}$ estimates the model parameters by minimizing the sum of error variances for all variables in the model given $\boldsymbol{\Sigma}_{std}$. In general, $\boldsymbol{\Sigma}_{std}$ is replaced with the sample correlation matrix of indicators, denoted by $\mathbf{S}_{std}$. The objective function (3.5) also shows that GSCA$_{std}$ aims to create components that explain the total variances of variables in the model rather than their covariances, as with PCA or other

component-based methods. The error terms in the $GSCA_{std}$ model are not considered independent entities that cause the variation of indicators but simply treated as residuals that are unexplained by independent components. Thus, $GSCA_{std}$ typically makes no assumptions about the correlation structure of the error terms of indicators, leaving them freely correlated. This is distinct from the common factor model, where the error terms are typically assumed to be uncorrelated. Nonetheless, no error covariances between different blocks of indicators may be assumed in some special cases of GSCA (Cho et al., 2020; Cho, Sarstedt, et al., 2022).

Note that (3.1) defines a component as a weighted sum of indicators, which is also the case in PCA. However, this equation itself is not identified because there would exist infinitely different ways of deciding the component weights. Thus, we need a certain rule or criterion to determine the component weights. PCA's criterion is one of the most widely used ones in statistics that the weights are to be determined in such a way that their corresponding components explain the maximum total variance of the indicators. The regression coefficients of indicators on their component are (component) loadings. These relationships between components and their indicators are expressed in the component measurement model (3.2). Thus, GSCA can have confirmatory PCA (Takane, Kiers, & de Leeuw, 1995) as a special case when it considers (3.1) and (3.2) only.

As the minimization problem (3.5) cannot be solved in closed form, an alternating least squares (ALS) algorithm was developed for iteratively finding the minimum point of (3.5). In the ALS algorithm, $\mathbf{W}_{std}$ and $\mathbf{A}_{std}$ are updated alternately with the other fixed until the difference in (3.5) between consecutive iterations decreases beyond a pre-specified tolerance level (e.g., 10^{-5}) (see Hwang & Takane, 2014, Chapter 2, for a full description of the ALS algorithm). Let $\widehat{\mathbf{\Gamma}}_{std}$ denote an N by P matrix of the standardized score estimates of components, $\mathbf{D}_{std}$ denote an N by J matrix of the standardized scores of indicators, and N is the number of cases in the sample. Let us suppose that we obtain the estimates of $\mathbf{W}_{std}$ and

$\mathbf{A}_{std}$ that minimize (3.5), denoted by $\widehat{\mathbf{W}}_{std}$ and $\widehat{\mathbf{A}}_{std}$. Then, a matrix of standardized component scores is obtained by

$$\widehat{\mathbf{\Gamma}}_{std} \equiv \mathbf{D}_{std}\widehat{\mathbf{W}}_{std}. \tag{3.6}$$

3.2.2. Unstandardized Weight Estimates in GSCA$_{std}$

Let $\mathbf{D} = \mathbf{1}_N\hat{\boldsymbol{\mu}}' + \mathbf{D}_{std}\widehat{\boldsymbol{\Delta}}_z$ denote an N by J matrix of the unstandardized scores of indicators, where $\hat{\boldsymbol{\mu}}$ is a J by 1 sample mean vector and $\widehat{\boldsymbol{\Delta}}_z$ is a diagonal matrix whose entries are sample standard deviations of unstandardized indicators. Conventionally, unstandardized component means are subsequently computed by transforming $\widehat{\mathbf{W}}_{std}$ as follows. As it follows from (3.6) that

$$\begin{aligned} &\widehat{\mathbf{\Gamma}}_{std} = (\mathbf{D} - \mathbf{1}_N\hat{\boldsymbol{\mu}}')\widehat{\boldsymbol{\Delta}}_z^{-1}\widehat{\mathbf{W}}_{std} \\ &\leftrightarrow \mathbf{1}_N\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Delta}}_z^{-1}\widehat{\mathbf{W}}_{std} + \widehat{\mathbf{\Gamma}}_{std} = \mathbf{D}\widehat{\boldsymbol{\Delta}}_z^{-1}\widehat{\mathbf{W}}_{std} \\ &\leftrightarrow \mathbf{1}_N\hat{\boldsymbol{\mu}}'\widehat{\mathbf{W}}_{uni} + \widehat{\mathbf{\Gamma}}_{std} = \mathbf{D}\widehat{\mathbf{W}}_{uni}, \end{aligned} \tag{3.7}$$

where $\widehat{\mathbf{W}}_{uni} \equiv \widehat{\boldsymbol{\Delta}}_z^{-1}\widehat{\mathbf{W}}_{std}$, GSCA$_{std}$ computes unstandardized component scores, denoted here by $\widehat{\mathbf{\Gamma}}_{uni}$, as $\widehat{\mathbf{\Gamma}}_{uni} \equiv \mathbf{D}\widehat{\mathbf{W}}_{uni}$ (Hwang & Takane, 2014, p. 26).

As shown in the last line of (3.7), however, $\widehat{\mathbf{\Gamma}}_{uni}$ can be simply seen as a variant of standardized component scores whose means are only relocated *a posteriori* by $\mathbf{1}_N\hat{\boldsymbol{\mu}}'\widehat{\mathbf{W}}_{uni}$ in that $\widehat{\mathbf{\Gamma}}_{std}$ remains standardized irrespective of the sample variances of the original indicators. Consequently, it is not guaranteed that the scores of $\widehat{\mathbf{\Gamma}}_{uni}$ are within the same range of the unstandardized scores of their indicators, which will be empirically shown in Section 3.5. Also, as illustrated in Section 3.1, GSCA$_{std}$ tends to assign smaller unstandardized weights to original indicators with relatively large variances in forming $\widehat{\mathbf{\Gamma}}_{uni}$. That is because minimizing (3.5) involves imposing a relatively large penalty on an original indicator with a relatively large variance, which is shown in Appendix C1. This disproportionate penalization for original indicators can inadvertently amplify the influence of an original indicator with a small variance on GSCA$_{std}$'s parameter estimation. Such an approach could be deemed

unsuitable when one aims to obtain an unstandardized component of original indicators on a single scale.

3.3. Convex Component and Its Six Properties

Let γ_p denote the pth component ($p = 1, 2, \cdots, P$) that is assumed to have the mean τ_p and variance ϕ_p. Let $\mathbf{z}_p$ denote a J_p by 1 vector of indicators for γ_p, where J_p is the number of indicators for $\boldsymbol{\gamma}_p$. We call the vector $\mathbf{z}_p$ *a block of indicators* for γ_p, which is assumed to have the mean vector $\boldsymbol{\mu}_p$ and covariance matrix $\boldsymbol{\Sigma}_p$. Let $\mathbf{w}_p$ denote a J_p by 1 vector of weights for $\mathbf{z}_p$. Let $\mathbf{0}_{k\times l}$ denote a k by l matrix of zeros, where k and l are any scalars. If $\boldsymbol{\gamma}_p$ is defined as a convex component, it can be expressed as

$$\gamma_p \equiv \mathbf{w}_p'\mathbf{z}_p \text{ subject to } \mathbf{w}_p'\mathbf{1}_{Jp} = 1 \text{ and } \mathbf{w}_p \geq \mathbf{0}_{Jp\times 1}. \tag{3.8}$$

A convex component has six useful properties as follows.

Proposition 1. *A convex component has scores within the range of its indicators' scores.*

Proposition 2. *Each score of a convex component corresponds to a component score of an individual whose scores for indicators are all the same as the component score.*

Proposition 3. *The mean of a convex component is not fixed to zero but is determined by weights within the range of its indicators' means.*

Proposition 4. *The standard deviation of a convex component is not fixed to one but is determined by weights within the range from 0 to the maximum standard deviation of its indicators.*

Proposition 5. *Given a linearly independent set of indicators' scores, a set of convex component scores has a unique set of weights that are nonnegative and summed up to one.*

Proposition 6. *The path coefficient of a convex component on an outcome variable indicates the expected amount of change in the outcome variable for a unit change in each indicator of the convex component while holding other variables fixed.*

We provide proofs for the six propositions in Appendix C2. The first four properties make a convex component' scores, mean, and standard deviation interpretable with reference to its indicators' scale when its indicators are on the same scale. The fifth property allows interpreting weight parameters as the contribution rates of indicators to forming their component. The last property allows for interpreting the path coefficient of a convex component with respect to its indicators' scale. We here illustrate these properties with an example of (major) depression.

Let us assume that depression can be represented by a convex component (γ) with three symptom-related indicators (z_1 = depressed affect, z_2 = somatic discomfort, and z_3 = interpersonal problem), which are commonly rated on a seven-point Likert scale (0 = "none", 1 = "minimal", 2 = "mild", 3 = "moderate", 4 = "moderately severe", 5 = "severe", and 6 = "extremely severe"). It is generally considered safe to treat ordinal variables with five or more categories as continuous (Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993). Then, this depression component serves as a summary index whose score indicates the overall severity level of the three depressive symptoms for each individual. Specifically, once weight parameters are estimated, a score set of depression component is obtained given a dataset of its indicators. Proposition 1 indicates that all individuals' scores of depression component will be within the range of the measurement scale of its indicators (e.g., [0, 6]). Proposition 2 implies that each individual's score of depression component within the range can be interpreted as the depression level of an individual whose indicators' scores are all the same as the depression component score. For example, if a patient's depression component score is 3, it implies that their depression level can be considered equivalent to that of depression of a patient whose symptom levels are all moderate (i.e., 3), suggesting that their depression is generally moderate. By Propositions 3 and 4, the means and the standard deviations of depression component are determined by

weight parameter estimates within the range of its indicators' original scales (e.g., [0, 6]) as well, which can also be interpreted in relation to those scales. For instance, if the mean of depression component scores turns out to be 5, it means that the average depression level of patients in the sample can be considered equivalent to the depression level of a patient whose symptom levels are all severe, or that the patients' depression is severe on average. Also, if the standard deviation of depression component scores turns out to be 1, it implies that the depression severity levels of patients in the sample were one-unit lower or higher than the moderate level on average.

By Proposition 5, it is guaranteed that once a set of depression component scores is obtained with a set of weight estimates, any other set of weight estimates does not exist that makes the same score set of depression component while satisfying the constraint in (3.8). As these weight estimates are always non-negative and summed up to one, they can be interpreted as the indicators' contribution 'rates' of forming the convex component. For example, suppose that the weight estimates for z_1, z_2 and z_3 are .41, .24, and .35, respectively. It indicates that when the severity level of depression component increases by one unit due to a one-unit increase in all the three symptom-related indicators, the contribution rates of z_1, z_2 and z_3 to the one-unit increase of depression severity are 41%, 24%, and 35%, respectively. Such interpretation was not applicable to weight of standardized components, as their values can be negative and not necessarily summed up to one. Note that this proposition is satisfied only if a linearly independent set of indicators' scores is given as a dataset. A set of indicators' scores being linearly independent means that a score vector of an indicator cannot be expressed as a linear combination of score vectors of the other indicators, which further implies that sample covariance matrix of the indicators is positive definite.

By Proposition 6, the path coefficient of a convex component on an outcome variable can be interpreted as an aggregate effect of the indicators of the convex component on the

outcome variable, given that the structural model holds. For example, let's consider a situation where a path coefficient of a depression component on employment earnings for the year of depression reported is identified -$5000 (e.g., Dobson et al., 2021). This would suggest that a one unit increase across all depression symptoms, such as a shift in all depression symptom levels from mild to moderate, would be associated with a $5000 loss for the individual experiencing depression. Such an interpretation was not feasible for path coefficients of standardized components.

3.4. Convex GSCA

3.4.1. Model Specification

Convex GSCA ($GSCA_{cvx}$) introduces a convex component with original indicators into the GSCA model. The $GSCA_{cvx}$ model also consists of three sub-models: weighted relation, component measurement, and structural models (Hwang & Takane, 2004, 2014). Let $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_P]'$ denote a P by 1 random vector of components. Each component (γ_p) can be either a convex or standardized component. If a block of indicators ($\mathbf{z}_p$) has the same measurement unit within the block, γ_p is defined as a convex component as expressed in (3.8). Otherwise, γ_p is defined as a standardized component, whose indicators ($\mathbf{z}_p$) are also assumed to be standardized such that $\tau_p = 0$, $\phi_p = 1$, $\boldsymbol{\mu}_p = \mathbf{0}_{Jp\times 1}$, and $vecdiag(\boldsymbol{\Sigma}_p) = \mathbf{1}_{Jp}$. Let $\mathbf{W}$ denote a J by P matrix consisting of component weights assigned to $\mathbf{z}$. Let $\mathbf{C}$ denote a P by J matrix of loadings relating $\boldsymbol{\gamma}$ to $\mathbf{z}$. Let $\mathbf{B}$ denote a P by P matrix of path coefficients relating $\boldsymbol{\gamma}$ to each other. Let $\mathbf{c}_0$ and $\mathbf{b}_0$ denote the column vectors of intercepts for the component measurement and structural models, respectively. The three sub-models of $GSCA_{cvx}$ are expressed as follows.

$$\boldsymbol{\gamma} \equiv \mathbf{W}'\mathbf{z} \text{ (weighted relation model)} \tag{3.9}$$

$$\mathbf{z} = \mathbf{c}_0 + \mathbf{C}'\boldsymbol{\gamma} + \boldsymbol{\xi} \text{ (component measurement model)} \tag{3.10}$$

$$\boldsymbol{\gamma} = \mathbf{b}_0 + \mathbf{B}'\boldsymbol{\gamma} + \boldsymbol{\zeta} \text{ (structural model).} \tag{3.11}$$

In GSCA$_{cvx}$, the weighted relation model (3.9) shows that each component is defined as a weighted sum of standardized or unstandardized indicators. As GSCA$_{cvx}$ may involve unstandardized variables, intercept terms ($\mathbf{c}_0$ and $\mathbf{b}_0$) are newly included into the component measurement and structural model (3.10) and (3.11). Each model parameter in (3.10) and (3.11)—intercepts, loadings, and path coefficients—can be interpreted in the same manner as the intercepts and regression coefficients in linear regression model with unstandardized variables. The three sub-models are combined into the following equation,

$$\begin{aligned} &[\mathbf{z}; \boldsymbol{\gamma}] = [\mathbf{c}_0; \mathbf{b}_0] + [\mathbf{C}, \mathbf{B}]'\boldsymbol{\gamma} + [\boldsymbol{\xi}; \boldsymbol{\zeta}] \\ &\leftrightarrow [\mathbf{I}_J, \mathbf{W}]'\mathbf{z} = [\mathbf{c}_0; \mathbf{b}_0] + [\mathbf{C}, \mathbf{B}]'\mathbf{W}'\mathbf{z} + [\boldsymbol{\xi}; \boldsymbol{\zeta}] \\ &\leftrightarrow \mathbf{V}'\mathbf{z} = \mathbf{a}_0 + \mathbf{A}'\mathbf{W}'\mathbf{z} + \mathbf{e}, \end{aligned} \tag{3.12}$$

where $\mathbf{a}_0 \equiv [\mathbf{c}_0; \mathbf{b}_0]$, $\mathbf{V} \equiv [\mathbf{I}_J, \mathbf{W}]$, $\mathbf{A} \equiv [\mathbf{C}, \mathbf{B}]$, and $\mathbf{e} \equiv [\boldsymbol{\xi}; \boldsymbol{\zeta}]$. The equation (3.12) is called the GSCA$_{cvx}$ model. If every indicator and component is standardized, the GSCA$_{cvx}$ model (3.12) becomes identical to the GSCA$_{std}$ model (3.4).

3.4.2. Parameter Estimation

Let $\boldsymbol{\sigma}_p$ denote a J_p by 1 vector of standard deviations (SD) of $\mathbf{z}_p$. If the pth component is defined as standardized ones, $\boldsymbol{\sigma}_p$ is equivalent to $\mathbf{1}_{Jp}$. Let $\mathbf{O}_z$ denote a J by J diagonal matrix whose jth element is $J_p^{-1}\mathbf{1}_{Jp}'\boldsymbol{\sigma}_p$ if the jth indicator in the pth block is a dependent variable and zero otherwise. Let $\mathbf{O}_\gamma$ denote a P by P diagonal matrix whose pth element is $J_p^{-1}\mathbf{1}_{Jp}'\boldsymbol{\sigma}_p$ if the pth component is a dependent variable and zero otherwise. Let $\mathbf{O} \equiv blkdiag(\mathbf{O}_z, \mathbf{O}_\gamma)$. GSCA$_{cvx}$ estimates parameters by minimizing the following objective function

$$\begin{aligned} &f_{cvx}(\mathbf{W}, \mathbf{A}, \mathbf{a}_0) \\ &= tr(\mathbf{O}E(\mathbf{e}\mathbf{e}')\mathbf{O}) \\ &= E(SS(([\mathbf{z}; \boldsymbol{\gamma}]' - (\mathbf{a}_0' + \mathbf{z}'\mathbf{W}\mathbf{A}))\mathbf{O})), \end{aligned} \tag{3.13}$$

subject to $\mathbf{w}_p'\boldsymbol{\Sigma}_p\mathbf{w}_p = 1$ or $\mathbf{1}_{Jp}'\mathbf{w}_p = 1$ ($p = 1, 2, \cdots, P$). The objective function (3.13) shows that components in GSCA$_{cvx}$ are constructed such that they can minimize the "weighted" sum

of error variances for all dependent variables under the constraints. Specifically, the objective function (3.13) penalizes each prediction error for dependent variables differentially by dividing it by the average SD of the corresponding block of indicators. This prevents prediction errors for a block of indicators with large variances from dominating the estimation of parameters.

To help understand the role of **O** in (3.13), we illustrate how **O** is determined based on the standard deviations of indicators. This will also explain the characteristic of the objective function described above. Figure 3.1 presents an illustrative GSCA$_{cvx}$ model involving two convex components (γ_1 and γ_2), each measured by three indicators that share the same scale, while the scales of two indicator blocks differ. Let us assume that $\boldsymbol{\sigma}_1$ = [1; 2; 3] and $\boldsymbol{\sigma}_2$ = [100; 200; 300], indicating that the differences in the overall magnitude of indicators' variances between the two blocks arises from the difference in scale. In this case, without **O** in (3.13) (i.e., **O** = **I**), the value of (3.13) would predominantly rely on the error variances for $\mathbf{z}_2$ and γ_2, implying that the error variances for $\mathbf{z}_1$ would be rarely considered in parameter estimation due to their scale. However, GSCA$_{cvx}$ determines **O** = *blkdiag*($\mathbf{O}_z$, $\mathbf{O}_\gamma$), where $\mathbf{O}_z$ = *blkdiag*(2, 2, 2, 200, 200, 200)$^{-1}$ and $\mathbf{O}_\gamma$ = *blkdiag*(0, 200^{-1}), and then uses it to penalize the error variances for $\mathbf{z}_2$ and γ_2 to adjust their effects on (3.13). For instance, given A = **0** and $\mathbf{a}_0 = E([\mathbf{z}; \boldsymbol{\gamma}])$, there are substantial differences in error variances between $\mathbf{z}_1$ and $\mathbf{z}_2$ (i.e., [1^2; 2^2; 3^2] for $\mathbf{z}_1$ and [100^2; 200^2; 300^2] for $\mathbf{z}_2$), but their error variances contribute equally to the value of (3.13) (i.e., ($1^2 + 2^2 + 3^2$)/2^2 = [100^2; 200^2; 300^2]/200^2). This suggests that introducing **O** into (3.13) enables GSCA$_{cvx}$ to consider prediction errors for both $\mathbf{z}_1$ and $\mathbf{z}_2$ during the parameter estimation process.

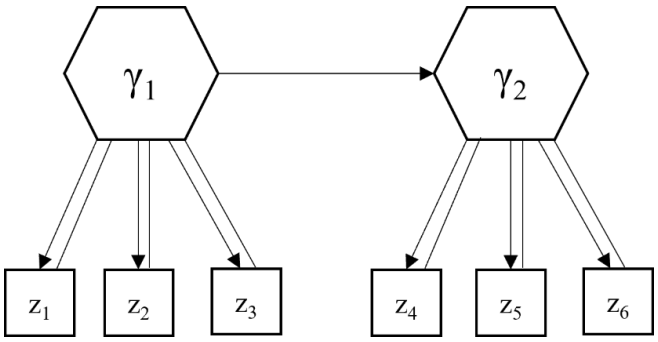


Figure 3.1. An illustrative GSCA$_{cvx}$ model. Hexagons represent components, squares denote indicators, straight lines indicate weights, single-headed arrows denote loadings and path coefficients. All intercepts and error terms are omitted to make the figure concise.

Conversely, as illustrated above, the objective function (3.13) does not impose different penalties on indicators within the same block to take into account potential differences in their variances. Furthermore, the objective function (3.13) is *partially scale-invariant*, which means that the minimum value of (3.13) does not vary with a linear change of measurement scales of each block of indicators that share the same scale (e.g., a scale range from 1 – 10 to 0 – 100), leading to the same weight estimates. This property is distinct from a property of (*full) scale invariant* (Swaminathan & Algina, 1978) in that changing the measurement scales of each indicator differentially (e.g., standardization) is not considered. The proof for the property is provided in Appendix C3.

As the minimum point of (3.13) cannot be found in closed form, we developed an ALS algorithm for iteratively finding its minimum point. A detailed description of the ALS algorithm is provided in Appendix C4. Note that we do not constrain the weights for convex components to be non-negative in (3.13) to make the method more flexible. In some cases, researchers may wish to examine which indicators contribute to forming a component in the opposite direction to the other indicators and may be excluded during model re-specification. The negative weight of an indicator for a convex component may signify that the indicator is not suitable to form the component along with other indicators. As discussed in Appendix C4,

the ALS algorithm allows for the imposition of the additional non-negativity constraints on weights, forcing the weights to be always positive.

3.4.3. Model Evaluation Indexes

GSCA$_{std}$ provides four overall goodness-of-fit measures, including FIT, AFIT, GFI, and SRMR, and one overall cross-validation index, out-of-bag prediction error (OPE). The FIT indicates the average explained variance of all variables in the model, whereas the AFIT is an adjusted version of FIT that takes into account the number of model parameters and sample size (Hwang & Takane, 2014, pp. 26–29). The GFI and SRMR evaluate the discrepancy between the sample and implied covariance matrices (Cho et al., 2020). The OPE aims to measure the average out-of-sample prediction error of the model for all variables via a bootstrapping-based cross validation and can be used for comparing models in terms of predictive generalizability (Cho et al., 2019). Whereas the GFI and SRMR can be used for GSCA$_{cvx}$ without modification, the FIT, AFIT, and OPE need to be modified for GSCA$_{cvx}$ because these measures were developed only for the condition where all variables are standardized. We revised FIT and OPE such that they can be applied for the GSCA$_{cvx}$ model with both standardized and unstandardized variables, taking into account the variances of dependent variables only.

We propose a modified version of FIT, termed FIT for unstandardized dependent variables (FIT^{UD}), as follows.

$$\text{FIT}^{\text{UD}} = 1 - \frac{SS(([\mathbf{D}, \mathbf{D}\hat{\mathbf{W}}] - (\mathbf{1}_N \hat{\mathbf{a}}_0' + \mathbf{D}\hat{\mathbf{W}}\hat{\mathbf{A}}))\hat{\mathbf{O}})}{SS(([\mathbf{D}, \mathbf{D}\hat{\mathbf{W}}] - \mathbf{1}_N \hat{\boldsymbol{\mu}}'[\mathbf{I}_J, \hat{\mathbf{W}}])\hat{\mathbf{O}})}. \quad (3.14)$$

The FIT^{UD} indicates the proportion of the explained variance of all dependent variables (including dependent convex components) to their weighted total variance. If every component and indicator is standardized, FIT^{UD} = $\frac{T}{T_Y}$ FIT, where $T \equiv P + J$ and T_Y is the

total number of dependent variables in the model. Also, we provide the following two local fit measures of FIT^{UD}

$$\text{FIT}_{\text{M}}^{\text{UD}} = 1 - \frac{SS((\mathbf{D} - (\mathbf{1}_N \hat{\mathbf{c}}_0' + \mathbf{D}\hat{\mathbf{W}}\hat{\mathbf{C}}))\hat{\mathbf{O}}_{\text{z}})}{SS((\mathbf{D} - \mathbf{1}_N \hat{\boldsymbol{\mu}}')\hat{\mathbf{O}}_{\text{z}})}, \tag{3.15}$$

$$\text{FIT}_{\text{S}}^{\text{UD}} = 1 - \frac{SS((\mathbf{D}\hat{\mathbf{W}} - (\mathbf{1}_N \hat{\mathbf{b}}_0' + \mathbf{D}\hat{\mathbf{W}}\hat{\mathbf{B}}))\hat{\mathbf{O}}_{\gamma})}{SS((\mathbf{D}\hat{\mathbf{W}} - \mathbf{1}_N \hat{\boldsymbol{\mu}}'\hat{\mathbf{W}})\hat{\mathbf{O}}_{\gamma})}, \tag{3.16}$$

where $\widehat{\mathbf{O}}_{\text{z}}$ and $\widehat{\mathbf{O}}_{\gamma}$ are sample analogies of $\mathbf{O}_{\text{z}}$ and $\mathbf{O}_{\gamma}$. We refer to "local fit" as the goodness-of-fit of GSCA's sub-models. The $\text{FIT}_{\text{M}}^{\text{UD}}$ and $\text{FIT}_{\text{S}}^{\text{UD}}$ can be used for evaluating the component measurement and structural models, respectively. The $\text{FIT}_{\text{M}}^{\text{UD}}$ indicates the proportion of the explained variance of all dependent indicators to their weighted total variance, whereas the $\text{FIT}_{\text{S}}^{\text{UD}}$ indicates the proportion of the explained variance of all dependent (convex) components to their weighted total variance.

Moreover, we propose a revised version of OPE, termed OPE for dependent variables (OPE^{UD}), to evaluate the predictive generalizability of models involving convex components, as follows.

$$\text{OPE}^{\text{UD}} = \frac{1}{K}\sum_{k=1}^{K} \frac{SS(([\mathbf{D}_k^*, \mathbf{D}_k^*\hat{\mathbf{W}}_k] - (\mathbf{1}_{N_k}\hat{\mathbf{a}}_0' + \mathbf{D}_k^*\hat{\mathbf{W}}_k\hat{\mathbf{A}}_k))\hat{\mathbf{O}}_k)}{SS(([\mathbf{D}_k^*, \mathbf{D}_k^*\hat{\mathbf{W}}_k] - \mathbf{1}_{N_k}\hat{\boldsymbol{\mu}}_k'[\mathbf{I}_J, \hat{\mathbf{W}}_k])\hat{\mathbf{O}}_k)}., \tag{3.17}$$

where $\widehat{\mathbf{W}}_k$, $\widehat{\mathbf{A}}_k$, $\hat{\mathbf{a}}_k$, and $\hat{\boldsymbol{\mu}}_k$ are the parameter estimates obtained from the kth bootstrap sample (k = 1,2, ⋯, K), $\widehat{\mathbf{O}}_k$ is the penalty term that rescales prediction errors for all dependent variables in the kth bootstrap sample, $\mathbf{D}_k^*$ is the kth test sample consisting of observations that are not included in the kth bootstrap sample, and N_k is the number of observations in the kth test sample. As shown in (3.17), the bootstrap sampling procedure generates pairs of mutually exclusive samples (bootstrap and test samples), over which a specified GSCA model is cross-validated (for a detailed description of OPE's computation, refer to Cho et al., 2019). The OPE^{UD} represents the weighted average out-of-sample prediction error of the model for dependent variables. The value of the OPE^{UD} ranges from 0 to infinity, where 0 means that a

specified model perfectly predicts every dependent variable, and a value over 1 indicates that the prediction accuracy of a specified model is worse than that of the null model, where all dependent variables are predicted by their sample means. Again, when every variable is standardized, $\text{OPE}^{\text{UD}} = \frac{T}{T_{\text{Y}}}\text{OPE} - \frac{(T - T_{\text{Y}})}{T_{\text{Y}}}$. In addition, we provide the following two local cross-validation indexes of OPE^{UD}

$$\text{OPE}_{\text{M}}^{\text{UD}} = \frac{1}{K}\sum_{k=1}^{K}\frac{SS((\mathbf{D}_k^* - (\mathbf{1}_{N_k}\hat{\mathbf{c}}_{0,k}' + \mathbf{D}_k^*\hat{\mathbf{W}}_k\hat{\mathbf{C}}_k))\hat{\mathbf{O}}_{\text{z},k})}{SS((\mathbf{D}_k^* - \mathbf{1}_{N_k}\hat{\boldsymbol{\mu}}_k')\hat{\mathbf{O}}_{\text{z},k})}, \tag{3.18}$$

$$\text{OPE}_{\text{S}}^{\text{UD}} = \frac{1}{K}\sum_{k=1}^{K}\frac{SS((\mathbf{D}_k^*\hat{\mathbf{W}}_k - (\mathbf{1}_{N_k}\hat{\mathbf{b}}_{0,k}' + \mathbf{D}_k^*\hat{\mathbf{W}}_k\hat{\mathbf{B}}_k))\hat{\mathbf{O}}_{\gamma,k})}{SS((\mathbf{D}_k^*\hat{\mathbf{W}}_k - \mathbf{1}_{N_k}\hat{\boldsymbol{\mu}}_k'\hat{\mathbf{W}}_k)\hat{\mathbf{O}}_{\gamma,k})}. \tag{3.19}$$

where $\hat{\mathbf{O}}_{\text{z},k}$ and $\hat{\mathbf{O}}_{\gamma,k}$ are the penalty terms that rescale prediction errors for dependent indicators and components, respectively, in the *k*th bootstrap sample. The $\text{OPE}_{\text{M}}^{\text{UD}}$ and $\text{OPE}_{\text{S}}^{\text{UD}}$ can be used for evaluating the predictive generalizability of the component measurement and structural models, respectively.

3.5. Simulated Data Analysis

We conduct a simulation study to examine the parameter recovery of the proposed method. Figure 3.2 depicts the population GSCA_{cvx} model used in our simulation study. The population model involves four convex components, each of which is measured by four composite indicators. Indicators per block had different mean vectors: the mean vectors of indicators are [6, 5, 4, 3] for γ_1, [5.5, 4.5, 3.5, 2.5] for γ_2, [5, 4, 3, 2] for γ_3, and [4.5, 3.5, 2.5, 1.5] for γ_4, respectively.

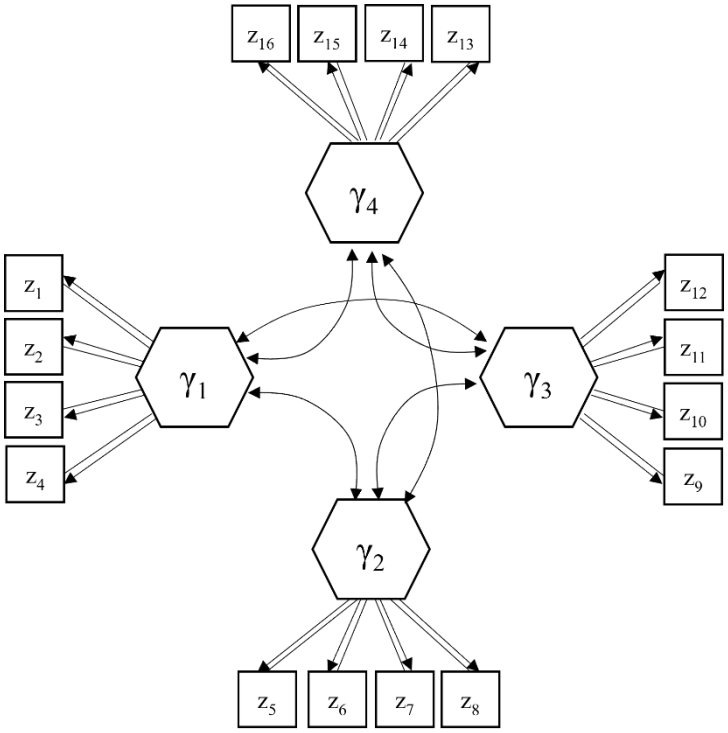


Figure 3.2. The population GSCA$_{cvx}$ model used in the simulation study. Double-headed arrows represent correlations. All intercepts and error terms are omitted to make the figure concise.

We manipulate four experimental factors: the variances of indicators, correlations between indicators per component, distribution of indicators, and correlations among components. We consider the variances of indicators because this is a unique piece of information for the proposed method to use for creating components as compared to GSCA$_{std}$. The other three factors have been frequently considered in testing the performance of GSCA (e.g., Cho, Sarstedt, et al., 2022; Cho & Choi, 2020; Hwang, Malhotra, et al., 2010). Specifically, we consider three levels of the variances of indicators per component: [1, 1, 1, 1], [1 2, 3, 4], and [1, 4, 9, 16]. We take into account three correlation matrices of indicators per component, which are provided in Table 3.1 (Cho & Choi, 2020). We consider two distributions of indicators: normal and non-normal. The normal distribution has a skewness of 0 and a kurtosis of 3, whereas the non-normal distribution has a skewness of 1.25 and a kurtosis of 3.75 as in Hwang et al. (2010). Lastly, we consider three levels of correlations among components (0, .2, and .4) as in Cho et al. (2022). In total, we consider 54 population

GSCA models with convex components (3 levels of indicators' variances × 2 types of indicators' distribution × 3 levels of indicators' correlations × 3 levels of components' correlations).

Table 3.1. Three conditions of the correlation patterns of four indicators per component in the simulation study.

	Condition 1				Condition 2				Condition 3			
	z_1	z_2	z_3	z_4	z_1	z_2	z_3	z_4	z_1	z_2	z_3	z_4
z_1	1				1				1			
z_2	.24	1			.50	1			.49	1		
z_3	.24	.20	1	.13	.43	.47	1		.56	.74	1	
z_4	.17	.21	.13	1	.30	.23	.45	1	.66	.48	.69	1

Per population model, we consider five sample sizes (N = 100, 200, 400, 800, and 1500), for each of which 1000 samples are randomly generated from the multivariate distribution with the population mean vector and covariance matrix of indicators. The procedure of deriving the population covariance matrix of indicators from the prescribed parameter values of a population GSCA$_{cvx}$ model is explained in Appendix C5. We apply GSCA$_{cvx}$[2] to each sample and obtain parameter estimates.

As parameter recovery measures, we empirically compute the absolute bias and root mean squared error (RMSE) of each parameter estimator. These measures are defined as

$$\text{Absolute bias} = \left|\mathrm{E}(\hat{\theta}) - \theta\right| \approx \left|\frac{1}{1000}\left(\sum_{i=1}^{1000}\hat{\theta}_i\right) - \theta\right| \tag{3.20}$$

$$\text{RMSE} = \sqrt{\mathrm{E}(\hat{\theta} - \theta)^2} \approx \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_i - \theta)^2}\ , \tag{3.21}$$

where θ is the value of each parameter, $\hat{\theta}$ is the estimator of θ, and $\hat{\theta}_i$ is the estimate of θ obtained from the ith sample. We focus here on reporting the average absolute bias and RMSE values of the estimators of weights, loadings, intercepts, component means, and

[2] The MATLAB code is available at https://osf.io/y75kg/?view_only=0d02aea6aaaa4aa29d405172544aae7d.

component variances over the population models per sample size, as the sample size is the only factor that substantially influences the absolute bias and RMSE values of the estimators. The results for each population model are provided in Supplementary Material.

Table 3.2 shows the average absolute bias and RMSE values of the estimators per sample size. In all sample sizes, the absolute biases of the weight, loading, and component mean estimators are small and close to zero on average. For example, when $N = 100$, the average absolute biases of the weight, loading, and component mean estimators are .002, .022, and .008, respectively. They continue to decrease and approach zero when the sample size increases. The average RMSE values of the same estimators show a similar pattern. When $N = 100$, the average RMSE values are around .047, .134, and .216, respectively, and becomes close to zero as the sample size increases. The average absolute bias and RMSE values of the intercept and component variance estimators are relatively large, compared to those of the other parameter estimators in the same condition. For instance, when $N = 100$, the average absolute biases of the intercept and component variance estimators are .107 and .178, respectively, and their average RMSE values are .668 and .859, respectively. However, both of them also decrease with the sample size and become close to zero. Taken together, $GSCA_{cvx}$ estimators are empirically unbiased on average, improving their parameter recovery as the sample size increases.

Table 3.2. The average absolute bias and RMSE values of the estimators of weights, loadings, intercepts, component means, and component variances per sample size.

	Absolute Bias					RMSE				
N	Weights	Loadings	Intercepts	Component Means	Component Variances	Weights	Loadings	Intercepts	Component Means	Component Variances
100	0.002	0.022	0.107	0.008	0.178	0.047	0.134	0.668	0.216	0.859
200	0.001	0.011	0.052	0.005	0.078	0.030	0.092	0.460	0.146	0.515
400	0.001	0.005	0.026	0.003	0.037	0.021	0.064	0.319	0.102	0.346
800	0.000	0.003	0.013	0.002	0.018	0.014	0.045	0.223	0.072	0.240
1500	0.000	0.002	0.008	0.002	0.010	0.010	0.033	0.162	0.052	0.174

3.6. Illustration with Empirical Data

To illustrate its empirical utility, we apply GSCA$_{cvx}$ to American customer satisfaction index (ACSI) data. The ACSI model (Fornell et al., 1996) is built on the established theories and has been used to produce index scores for customer satisfaction in the United States since 1994. The present ACSI data are comprised of 774 customers' responses for fourteen items: z_1 = expectation for overall quality, z_2 = expectation for reliability, z_3 = expectation for customization, z_4 = overall quality, z_5 = reliability, z_6 = customization, z_7 = price given quality, z_8 = quality given price, z_9 = perceived overall satisfaction, z_{10} = fulfilment of expectations, z_{11} = distance to the ideal, z_{12} = complaint behavior, z_{13} = repurchase intention, z_{14} = price tolerance. Twelve of the items (z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10}, z_{11}, and z_{13}) are measured on a 10-point Likert scale (e.g., 1 = "very negative" and 10 = "very positive"). Within the interval [1, 5], a smaller point reflects a stronger negative response, whereas within the interval [6, 10], a larger point indicates a stronger positive response. On the other hand, z_{12} is a binary variable (1 = formally complained and 0 = otherwise) and z_{14} is a composite of two price tolerance measures in different metrics, which is expressed as a percentage ranging from 0 to 50 (the higher, the more tolerant). The means, covariances, minimums, and maximums of the items are provided in Table 3.3. Refer to Fornell et al. (1996) for more detailed information on the items.

Table 3.3. Sample covariances (in upper triangular), correlations (in lower triangular), variances (in diagonal), means, minimums, and maximums of the fourteen indicators in the ACSI example.

	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}	z_{11}	z_{12}	z_{13}	z_{14}
z_1	5.81	3.61	2.71	2.92	2.91	2.10	1.96	2.79	3.01	2.73	3.23	–.14	2.62	11.78
z_2	.65	5.38	2.88	2.69	2.92	2.36	1.48	2.63	2.79	2.30	2.64	–.13	2.05	7.65
z_3	.43	.47	6.90	2.03	2.15	2.56	1.49	2.18	2.16	1.96	2.41	–.12	1.74	7.77
z_4	.53	.50	.33	5.31	4.77	3.76	2.52	4.15	4.87	4.35	4.42	–.32	3.94	16.86
z_5	.49	.51	.33	.84	6.11	3.98	2.48	4.37	5.23	4.80	5.01	–.36	4.29	17.20
z_6	.33	.39	.37	.62	.61	6.93	2.46	3.65	3.95	3.67	3.75	–.23	3.18	13.58

	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}	z_{11}	z_{12}	z_{13}	z_{14}
z_7	.31	.25	.22	.42	.39	.36	6.67	3.43	3.05	2.88	2.78	–.14	2.42	11.31
z_8	.47	.46	.33	.72	.71	.56	.53	6.18	4.74	4.42	4.70	–.29	3.60	16.60
z_9	.50	.48	.33	.85	.85	.60	.47	.77	6.19	5.06	5.09	–.35	4.55	19.55
z_{10}	.45	.40	.30	.75	.78	.56	.44	.71	.81	6.27	4.96	–.29	3.98	16.71
z_{11}	.51	.43	.35	.73	.77	.54	.41	.72	.78	.75	6.93	–.32	4.74	20.79
z_{12}	–.17	–.17	–.13	–.40	–.42	–.25	–.16	–.34	–.41	–.34	–.36	.12	–.33	-1.55
z_{13}	.37	.30	.22	.58	.58	.41	.32	.49	.62	.54	.61	–.33	8.79	35.21
z_{14}	.31	.21	.19	.47	.45	.33	.28	.43	.51	.43	.51	–.29	.76	241.11
Mean	7.34	7.75	6.67	7.66	7.59	7.39	5.96	7.12	7.59	6.82	6.76	.14	7.73	31.82
Min	1	1	1	1	1	1	1	1	1	1	1	0	1	0
Max	10	10	10	10	10	10	10	10	10	10	10	1	10	50

Figure 3.3 depicts the relationships among the six components and their indicators. The 14 items are used as composite indicators of the following six components: γ_1 = customer expectations (CE), γ_2 = perceived quality (PQ), γ_3 = perceived value (PV), γ_4 = customer satisfaction (CS), γ_5 = customer complaints (CC), and γ_6 = customer loyalty (CL). We represent all the constructs by convex components with unstandardized indicators except for the customer loyalty. As two indicators (z_{13} and z_{14}) for customer loyalty are not measured on the same scale, we set this component as a standardized one with the indicators standardized.

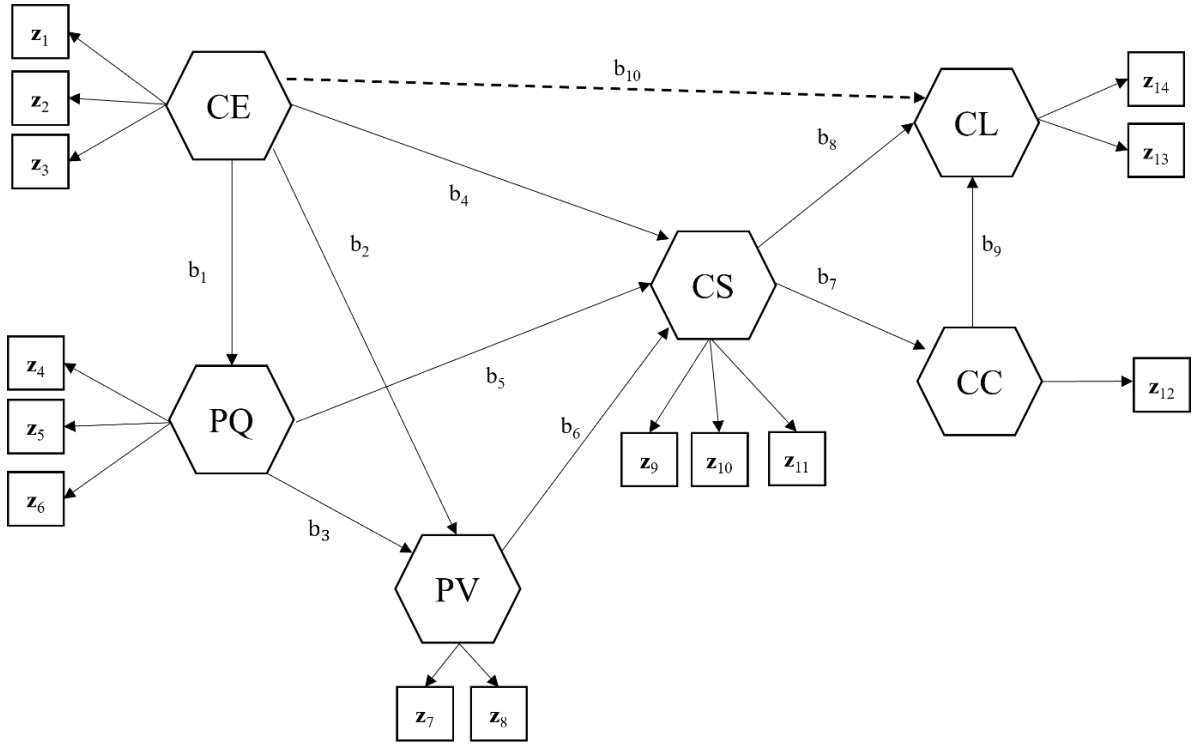


Figure 3.3. The ACSI model. The dashed line labeled b_{10} signifies an incorrectly specified path coefficient. All weights and error terms are omitted to make the figure concise. CE = customer expectations, PQ = perceived quality, PV = perceived value, CS = customer satisfaction, CC = customer complaints, CL = customer loyalty.

We use 4000 bootstrap samples for computing the standard error and 95% confidence interval of each parameter estimate. For comparison, we also apply $GSCA_{std}$ to the same data and compute unstandardized weight estimates and unstandardized component scores based on the procedure discussed in Section 3.2. As customer satisfaction is the focal component in the ACSI model, we concentrate on interpreting the scores of customer satisfaction, its statistics, and the relevant model parameters.

The model fitted by $GSCA_{cvx}$ shows $FIT^{UD} = .714$, indicating that the ACSI model accounts for 71.4% of the weighted total variance of all dependent variables in the model. It also provides GFI = .987 and SRMR = .022, pointing to an acceptable level of model fit (Cho et al., 2020). In addition, it provides that $FIT_M^{UD} = .802$ and $FIT_S^{UD} = .438$. This indicates that the component measurement model explains 80.2% of the weighted total variance of all

dependent indicators, whereas the structural model explains 43.8% of the weighted total variance of all dependent components.

Table 3.4 provides the weight and loading estimates, and their standard errors and 95% confidence intervals obtained from $GSCA_{cvx}$, along with the intercept estimates in the measurement model. The unstandardized weight estimates obtained from $GSCA_{std}$ are also provided for comparison. Overall, all the weight and loading estimates obtained from $GSCA_{cvx}$ are large and statistically significant, indicating that all the indicators contribute to forming their components, which in turn, explain the variances of their indicators well. Among the three indicators (z_9, z_{10}, and z_{11}) for customer satisfaction, z_9 (perceived overall satisfaction) are the largest contributor (w_9 =.422, SE = .015, 95% CI = [.393, .454]). This indicates that when each of the three indicators equally increases, leading to an increase in customer satisfaction, the contribution rate of $\mathbf{z}_9$ for the increase in customer satisfaction was 42.2%, which is greater than those of the two others (z_{10} = 25.4% and z_{11} = 32.4%). Similarly, the unstandardized weight estimate of $\mathbf{z}_9$ obtained from $GSCA_{std}$ is the largest among the three (w_9 = .188, w_{10} = .107, and w_{11} = .131). In contrast, it is uncertain how to interpret the unstandardized weight estimates obtained from $GSCA_{std}$.

Table 3.4. The weights, loading, and intercept estimates of the fourteen indicators in the ACSI model and their standard errors (SE) and 95% confidence intervals (CI) obtained from $GSCA_{cvx}$, along with the unstandardized weight estimates obtained from $GSCA_{std}$ ($\widehat{\mathbf{W}}_{uni}$).

Indicator	Component	Weights			$\widehat{\mathbf{W}}_{uni}$	Loadings			Intercepts
		Estimate	SE	95% CI		Estimate	SE	95% CI	($\hat{\mathbf{c}}_0$)
z_1	CE	.345	.013	[.320, .372]	.180	1.008	.025	[.957, 1.054]	.018
z_2		.337	.014	[.309, .366]	.188	.982	.027	[.925, 1.033]	.616
z_3		.317	.013	[.292, .343]	.128	1.011	.036	[.937, 1.077]	–.674
z_4	PQ	.387	.018	[.353, .425]	.184	.979	.013	[.953, 1.004]	.260
z_5		.342	.017	[.307, .374]	.170	1.043	.013	[1.018, 1.071]	–.303
z_6		.271	.007	[.257, .285]	.101	.976	.024	[.927, 1.021]	.012
z_7	PV	.404	.010	[.384, .423]	.154	.960	.020	[.921, .997]	–.427
z_8		.596	.010	[.577, .616]	.293	1.027	.013	[1.002, 1.051]	.289

Indicator	Component	Weights			$\widehat{\mathbf{W}}_{uni}$	Loadings			Intercepts ($\hat{\mathbf{c}}_0$)
		Estimate	SE	95% CI		Estimate	SE	95% CI	
z_9	CS	.422	.015	[.393, .454]	.188	1.004	.011	[.982, 1.025]	.433
z_{10}		.254	.013	[.229, .279]	.107	.965	.016	[.933, .996]	–.052
z_{11}		.324	.012	[.300, .348]	.131	1.022	.016	[.990, 1.053]	–.524
z_{12}	CL	1.000	.000	[1.000, 1.000]	2.909	1.000	.000	[1.000, 1.000]	.000
z_{13}	CC	.610	.015	[.579, .639]	.206	.956	.004	[.949, .963]	.000
z_{14}		.453	.016	[.424, .484]	.029	.920	.007	[.906, .932]	.000

Table 3.5 presents the path coefficient estimates and their standard errors and 95% confidence intervals obtained from GSCA$_{cvx}$. Overall, the patterns of all the path coefficient estimates are consistent with those from previous studies (e.g., Hwang & Takane, 2014, Chapter 2). For instance, perceived quality and perceived value have statistically significant influences on customer satisfaction (b_5 =.723, SE = .033, 95% CI = [.659, .786]; b_6 = .275, SE = .035, 95% CI = [.204, .344]). Customer satisfaction have statistically significant effects on customer complaints (b_7 = −.059, SE = .006, 95% CI = [−.072, −.047]) and customer loyalty (b_8 = .252, SE = .015, 95% CI = [.222, .279]). Each individual path coefficient estimate is indicative of the expected change of the dependent variable for a one-unit change in indicators of a predictor component. For instance, the estimate of the path coefficient, b_8 = .252, implies that a one-unit increase in z_9 (perceived overall satisfaction), z_{10} (expectation fulfillment), and z_{11} (distance to the ideal) would be associated with an increase of .252 unit in customer loyalty. The R^2 value is .331 for perceived quality, .511 for perceived value, .812 for customer satisfaction, .164 for customer complaints, and .404 for customer loyalty. Also, the intercept estimates for the dependent components in the same order as above are 3.014, .793, –.501, .558, and –1.756.

Table 3.5. The path efficient estimates and their standard errors (SE) and 95% confidence intervals (CI) obtained from $GSCA_{cvx}$.

		Estimate	SE	95% CI
b_1	CE → PQ	.626	.037	[.551, .696]
b_2	CE → PV	.134	.039	[.058, .209]
b_3	PQ → PV	.646	.038	[.573, .721]
b_4	CE → CS	.045	.026	[−.005, .095]
b_5	PQ → CS	.723	.033	[.659, .786]
b_6	PV → CS	.275	.035	[.204, .344]
b_7	CS → CC	−.059	.006	[−.072, −.047]
b_8	CS → CL	.252	.015	[.222, .279]
b_9	CC→ CL	−.267	.104	[−.471, −.064]

Table 3.6 presents the estimated means, standard deviations, and ranges of unstandardized component scores obtained from $GSCA_{cvx}$ and $GSCA_{std}$. As expected, the individual scores of each convex component obtained from $GSCA_{cvx}$ are within the range of their indicators' scores. The individual scores of customer expectation, perceived quality, perceived value, and customer satisfaction all range from 1 to 10 and those of customer complaint were between 0 and 1, which are equivalent to the ranges of their indicators' measurement scales. The mean of customer satisfaction from $GSCA_{cvx}$ is 7.125, indicating that the average satisfaction level in the sample is moderately positive or equivalent to the satisfaction level of a customer whose indicator scores are all 7.125. This mean of customer satisfaction appears to be congruent with the means of its original indicators (7.585, 6.824, and 6.760). The standard deviation of customer satisfaction is 2.353, suggesting that the scores of customer satisfaction are somewhat widely spread out from the mean. This standard deviation value also seems to conform to those of its original indicators (2.489, 2.504, and 2.632).

Table 3.6. The means, standard deviations (SD), and ranges of the unstandardized component scores estimated from $GSCA_{cvx}$ and $GSCA_{std}$. The last component (CL) is defined as a standardized component in $GSCA_{cvx}$.

	$GSCA_{cvx}$			$GSCA_{std}$		
	Mean	SD	Range	Mean	SD	Range
CE	7.265	2.014	[1.000, 10.000]	3.633	1.000	[.496, 4.961]
PQ	7.564	2.194	[1.000, 10.000]	3.444	1.000	[.455, 4.546]
PV	6.652	2.223	[1.000, 10.000]	3.008	1.000	[.448, 4.475]
CS	7.125	2.353	[1.000, 10.000]	3.037	1.000	[.425, 4.253]
CC	0.137	0.344	[0.000, 1.000]	.398	1.000	[.000, 2.909]
CL	0.000	1.000	[–2.311, .998]	2.518	1.000	[.206, 3.516]

On the contrary, unstandardized components' scores obtained from $GSCA_{std}$ are not always within the range of their indicators' scores. Some scores of customer expectation, perceived quality, perceived value, and customer satisfaction are smaller than 1, which is the minimum value of their indicators on the scale. Moreover, the means of unstandardized components are also far from those of their original indicators. For instance, the mean of customer satisfaction obtained from $GSCA_{std}$ is just 3.037, even though its indicators' means are around 7 as stated above. Thus, it is questionable whether the mean of customer satisfaction obtained from $GSCA_{std}$ can be a good representation of the average level of customer satisfaction in the sample. Furthermore, all the standard deviations of unstandardized components are fixed to one, even though none of their indicators have standard deviations being around 1.

To illustrate the usage of OPE^{UD} as a model comparison criterion, we additionally contemplate two misspecified models of the ACSI model, while assuming the original ACSI model as the true model (denoted by Model 1). One misspecified model (Model 2) is an under-specified one, where a path coefficient (b_6) is omitted from Model 1. The other misspecified model (Model 3) is an over-specified one that includes an additional path coefficient from customer expectation to customer loyalty in Model 1, as displayed in Figure 3.3. We apply $GSCA_{cvx}$ to fit the three models to the data and compute their OPE^{UD} values

based on 4000 bootstrap samples. Model 1 provides the smallest OPE^{UD} value (Model 1 = .2883, Model 2 = .2901, and Model 3 = .2887), indicating that the original ACSI model has the highest predictive generalizability among the three models. The OPE^{UD} value of Model 2 is larger than that of Model 1 (.2901 > .2883), suggesting that excluding a path coefficient (b_6) from Model 1 rather decreases the prediction accuracy of the model. On the other hand, the OPE^{UD} value of Model 3 is larger than that of Model 1 (.2887 > .2883), indicating that specifying an additional path coefficient (b_{10}) to Model 1 is not helpful to improve the predictive generalizability of the model.

3.7. Concluding Remarks

We proposed convex GSCA that can accommodate a new type of unstandardized components, named convex components. A convex component is defined as a convex combination of original indicators whose weights are all non-negative and summed up to one. Every individual score of a convex component is always within the range of its indicators' scores and can be interpreted as a construct's specific level of a person who has the same score for all its indicators as his/her component score. Moreover, the means and standard deviations of convex components are estimated along with other parameters through a single optimization procedure, which can also be interpreted in terms of indicators' scales. Thus, introducing convex components to the GSCA model will enhance the practical utility of component scores and their summary statistics, for instance, in investigating individuals' levels of a construct or comparing the average levels of a construct between groups.

We developed an alternating least squares (ALS) algorithm for estimating parameters of the convex GSCA model, which does not require standardizing blocks of indicators that have the same measurement scales within the blocks. The algorithm not only enables information on the variances of each block of indicators to be additionally utilized in

parameter estimation, but also prevents indicators with small variances from influencing more heavily the construction of an unstandardized component than those with large variances. Furthermore, its objective function is partially scale-invariant, indicating that the minimum value of the objective function remains unchanged with a linear change in the measurement scale of each block of indicators, giving rise to the same weight estimates.

We evaluated the parameter recovery of the proposed method in a simulation study and further illustrated the merits of the proposed method via a real data analysis. In the simulation study, the proposed method empirically produced unbiased parameter estimates on average under nine GSCA models with convex components and its accuracy was further improved with large sample size. In the real data analysis, the patterns of the parameter estimates were consistent with those from previous studies, and the benefits of convex components were pronounced, compared to the unstandardized components obtained from the conventional ad-hoc procedure of rescaling weight estimates. Unlike these unstandardized components, convex components' weight estimates were interpretable, all their individual scores fell within the range of indicators' measurement scales or their scores, and their estimated means and standard deviations were congruous with those of their indicators. Therefore, we are confident to recommend that researchers employ the method when they are interested in the GSCA model with unstandardized components of original indicators.

Note that as an anonymous reviewer pointed out, researchers may still want to consider standardizing observed variables that are measured on the same scale. We recommend considering this option only if researchers are not interested in unstandardized component scores. If researchers apply GSCA to estimate the scores of unstandardized components after standardizing indicators of the same scale, an indicator with a small variance can be assigned a relatively large unstandardized weight, leading to a potentially inflated influence of the indicator on the estimation of the component scores, as shown in

Section 3.2. This issue does not occur when researchers keep the original scales of indicators and apply convex GSCA with convex components.

In future research, we may consider incorporating convex components into various extensions of GSCA, which deal with more complex analyses, for instance, those of involving higher-order components (Hwang & Takane, 2014, Chapter 3), missing observations (Hwang & Takane, 2014, Chapter 3), multilevel components (Hwang, Takane, et al., 2007), components with categorical indicators (Hwang & Takane, 2010), component interaction terms (Hwang, Cho, Jin, et al., 2021; Hwang, Ho, et al., 2010), or factors (Hwang, Cho, Jung, et al., 2021). Such additional extensions will improve the usefulness of GSCA, placing components on their indicators' scales while having their means and variances free parameters to be estimated along with others.

References

Altman, A., & Gondzio, J. (1999). Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. *Optimization Methods and Software*, *11*(1–4), 275–302. https://doi.org/10.1080/10556789908805754

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Boyd, S. P., & Vandenberghe, L. (2018). *Introduction to applied linear algebra: vectors, matrices, and least squares*. Cambridge university press. https://doi.org/10.1017/9781108583664

Cho, G., & Choi, J. Y. (2020). An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation models. *Behaviormetrika*, *47*(1), 243–272. https://doi.org/10.1007/s41237-019-00098-0

Cho, G., Hwang, H., Sarstedt, M., & Ringle, C. M. (2020). Cutoff criteria for overall model fit indexes in generalized structured component analysis. *Journal of Marketing Analytics*, *8*, 189–202. https://doi.org/10.1057/s41270-020-00089-1

Cho, G., Jung, K., & Hwang, H. (2019). Out-of-bag prediction error: A cross validation index for generalized structured component analysis. *Multivariate Behavioral Research*, *54*(4), 505–513. https://doi.org/10.1080/00273171.2018.1540340

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

Dobson, K. G., Vigod, S. N., Mustard, C., & Smith, P. M. (2021). Major depressive episodes and employment earnings trajectories over the following decade among working-aged

Canadian men and women. *Journal of Affective Disorders*, *285*, 37–46. https://doi.org/10.1016/j.jad.2021.02.019

Floudas, C. A., & Visweswaran, V. (1995). Quadratic optimization. In R. Horst & P. M. Pardalos (Eds.), *Handbook of global optimization* (pp. 217–269). Springer US. https://doi.org/10.1007/978-1-4615-2025-2_5

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), 7–18. https://doi.org/10.2307/1251898

Frank, M., & Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, *3*(1-2), 95–110. https://doi.org/10.1002/nav.3800030109

Hwang, H., Cho, G., Jin, M. J., Ryoo, J. H., Choi, Y., & Lee, S.-H. (2021). A knowledge-based multivariate statistical method for examining gene-brain-behavioral/cognitive relationships: Imaging genetics generalized structured component analysis. *PloS One*, *16*(3), e0247592. https://doi.org/10.1371/journal.pone.0247592

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., Ho, M.-H. R., & Lee, J. (2010). Generalized Structured Component Analysis with Latent Interactions. *Psychometrika*, *75*(2), 228–242. https://doi.org/10.1007/s11336-010-9157-5

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. (2004). Generalized structured component analysis. *Psychometrika*,

69(1), 81–99. https://doi.org/10.1007/BF02295841

Hwang, H., & Takane, Y. (2010). Nonlinear generalized structured component analysis. *Behaviormetrika*, *37*(1), 1–14. https://doi.org/10.2333/bhmk.37.1

Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling*. Chapman and Hall/CRC Press.

Hwang, H., Takane, Y., & Malhotra, N. (2007). Multilevel generalized structural component analysis. *Behaviormetrika*, *34*(2), 95–109. https://doi.org/10.2333/bhmk.34.95

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, *48*, 398–407. https://doi.org/10.2307/2095231

Lawson, C. L., & Hanson, R. J. (1974). *Solving least squares problems*. Prentice Hall.

Lay, D. C., Lay, S. R., & McDonald, J. J. (2015). *Linear algebra and its applications* (5th ed.). Pearson Education.

Naik, D. N., & Khattree, R. (1996). Revisiting olympic track records: Some practical considerations in the principal component analysis. *The American Statistician*, *50*(2), 140–144. https://doi.org/10.1080/00031305.1996.10474361

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education : Theory and Practice*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Sullivan, G. M., & Artino, A. R. J. (2013). Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/JGME-5-4-18

Swaminathan, H., & Algina, J. (1978). Scale freeness in factor analysis. *Psychometrika*, *43*(4), 581–583. https://doi.org/10.1007/BF02293816

Vanderbei, R. J., & Carpenter, T. J. (1993). Symmetric indefinite systems for interior point

methods. *Mathematical Programming*, *58*(1), 1–32. https://doi.org/10.1007/BF01581257

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology / Psychologie Canadienne*, *34*, 390–400. https://doi.org/10.1037/h0078865

Chapter 4. Deep learning Generalized Structured Component Analysis: An Interpretable Artificial Neural Network Model with Composite Indexes

Manuscript: Cho, G., & Hwang, H. (in press). Deep learning generalized structured component analysis: An interpretable artificial neural network model with composite indexes. *Structural Equation Modeling: A Multidisciplinary Journal*.

Abstract

Generalized structured component analysis (GSCA) is a multivariate method for specifying and examining interrelationships between observed variables and components. Despite its data-analytic flexibility honed over the decade, GSCA always defines every component as a linear function of observed variables, which can be less optimal when observed variables for a component are nonlinearly related, often reducing the component's predictive power. To address this issue, we combine deep learning and GSCA into a single framework to allow a component to be a nonlinear function of observed variables without specifying the exact functional form in advance. This new method, termed deep learning generalized structured component analysis (DL-GSCA), aims to maximize the predictive power of components while their directed or undirected network remains interpretable. Our real and simulated data analyses show that DL-GSCA produces components with greater predictive power than those from GSCA in the presence of nonlinear associations between observed variables per component.

Keywords: Generalized structured component analysis, deep learning, nonlinear component, composite index, interpretability

4.1. Introduction

Generalized structured component analysis (GSCA; Hwang & Takane, 2004, 2014) is a multivariate method for examining path-analytic relationships between observed variables and components. GSCA estimates parameters via an iterative least squares algorithm, generating components that minimize the sum of (in-sample) prediction errors for all dependent variables in the model that researchers specify based on prior knowledge and theory. In this regard, GSCA's component may be considered an aggregated measure of observed variables, or a *composite index* created under a certain rule that a component is to explain the total variance of its observed variables and to be highly related to other components (Cho, Sarstedt, et al., 2022).

GSCA can include a broad array of component analysis methods as special cases, including (constrained) principal component analysis (Takane et al., 1995), (generalized) canonical correlation analysis (e.g., Carroll, 1968; Kettenring, 1971; Tenenhaus et al., 2017), principal covariate regression (de Jong & Kiers, 1992), redundancy analysis (van den Wollenberg, 1977), extended redundancy analysis (Takane & Hwang, 2005), canonical regression analysis (van der Leeden, 1990, p. 47), hierarchical structural component analysis (e.g., Choi et al., 2020), and partial or global least squares path modeling (Hwang et al., 2020; Hwang & Cho, 2020), particularly under the unidimensionality assumption that one component is extracted from each set of observed variables (refer to Hwang & Takane, 2014, Chapter 2).

GSCA has been extended in various ways to deal with more complex data structures and analyses. For instance, GSCA was combined with fuzzy clustering to capture cluster-level heterogeneity in observations (Hwang, Desarbo, et al., 2007). It was also extended to accommodate interactions between components while addressing potential multicollinearity via regularization (e.g., Hwang, 2009; Hwang, Cho, Jin, et al., 2021; Hwang, Ho, et al., 2010).

GSCA and all these extensions to date commonly assume that each component is defined as a *linear* function (or a weighted sum) of its observed variables, also referred to as (composite) indicators (Bollen & Bauldry, 2011). This linearity assumption can be useful for the interpretation of individual component scores, particularly when all component weights for indicators are positive. For example, if an individual has a high score on a component relative to other individuals, the individual will also tend to have relatively high scores for the component's indicators (e.g., refer to Section 3.3).

However, defining a component as a linear function of indicators (i.e., a linear component) all the time can be rather too restrictive in practice, ignoring potential nonlinear associations between the indicators. For example, two scatterplots in Figure 4.1 exhibit the relationships among four indicators for the Human Development Index (HDI; UNDP, 1990), including countries' gross national income (GNI), life expectancy (LifeExp), expected schooling years (ExpSchl), and mean schooling years (MeanSchl), from the Human Development Report 2010 (UNDP, 2010). The scatterplots show clear curvilinear relationships between GNI and the other indicators. Likewise, it has been reported that two indicators for socioeconomic status (SES), i.e., schooling year and income, are nonlinearly related (e.g., Gensowski et al., 2011, pp. 13–15; P. V. Le, 2014; Park, 1994). In these cases, adopting a linear component for the HDI or SES indicators can lead to less explanatory power for in-sample data.

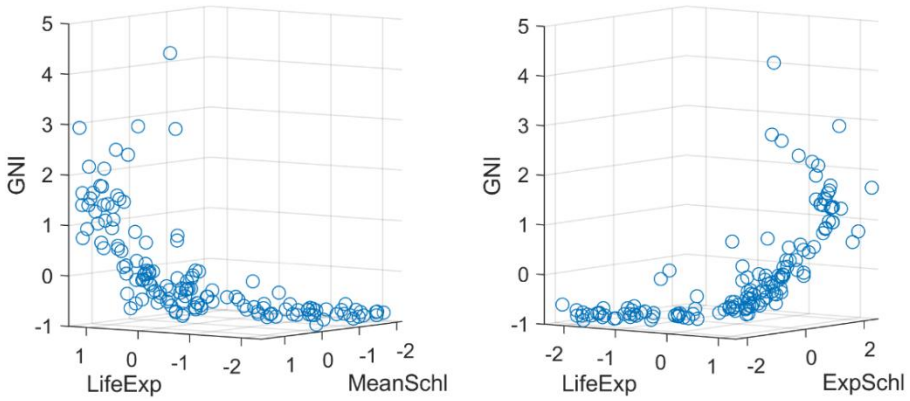


Figure 4.1. Scatterplots of four indicators for the Human Development Index (HDI).

In addition, linear components may have less predictive power for unseen observations or out-of-sample data. For example, Takane and Hwang (2007) investigated the relationships between two components—country-level food consumption and mortality rate. They showed that when both components were defined as linear ones, food consumption had substantially lower predictive power for mortality rate, as compared to when they were defined as nonlinear functions of their indicators.

We may consider adding polynomial and/or product terms of indicators to capture some nonlinear relationships between the indicators. Nonetheless, this way of modeling requires prior knowledge on what forms of nonlinear relationships between indicators is expected and which terms should be added to capture the nonlinearity sufficiently well. Deep learning (Lecun et al., 2015; Rosenblatt, 1962) can be useful to address such nonlinearity of indicators without relying on prior knowledge on their relationships. As in GSCA, deep learning begins by defining multiple components as linear deterministic functions of observed variables. Unlike GSCA, however, it typically further transforms the linear components in a nonlinear manner. This multivariate and vector-valued function, which transforms multiple input variables (e.g., observed variables) into multiple outcome variables (e.g., transformed components of the input variables), is called a layer. The outcome variables of a layer are fed into another layer, in which they are again combined into several linear

components and (nonlinearly) transformed. The operation of producing a function as a function of another function—function composition—is a key element that enables deep learning to successfully approximate nonlinear relationships between input and outcome variables (Urban & Gates, 2021). It is mathematically proven that any continuous function can be approximated by deep learning's function in a prescribed accuracy under some mild conditions—the universal approximation theorem (For the conditions in detail, Csaji, 2001; Cybenko, 1989; Z. Lu et al., 2017).

In the paper, therefore, we propose to combine deep learning (DL) with GSCA into a single framework. This extension, termed *deep learning generalized structured component analysis* (DL-GSCA), aims to conduct path-analytic modeling of indicators and components as in GSCA. At the same time, DL-GSCA relaxes the linearity assumption between indicators and components thereof, allowing components to be nonlinear functions of their indicators without the need of pre-determining their exact functional forms. It then estimates the functional form of each component from the data by utilizing DL's artificial neural networks, such that the component can maximize its predictive power for dependent variables in the model. In this regard, DL-GSCA is distinct from GSCA that always fixes the component's functional form to be linear. Nonetheless, DL-GSCA still assumes that the relationships among components remain linear to facilitate their interpretation as in GSCA.

The paper is organized as follows. In Section 4.2, we provide a technical account of DL-GSCA, including its model specification, parameter estimation, and model evaluation. We here focus on providing a rather general description of the proposed method to avoid excessive notational burden, while all technical details with formal mathematical notations are relegated to appendices. In Section 4.3, we present the application of DL-GSCA to the Human Development Index datasets to illustrate its relative explanatory and predictive performance to GSCA. In Section 4.4, we conduct a Monte-Carlo simulation study to

investigate whether DL-GSCA's comparative advantages over GSCA can be generalized to conditions where different forms of nonlinearity exist between indicators. In Section 4.5, we summarize and discuss the implications, limitations, and potential extensions of the proposed method.

4.2. The Proposed Method

4.2.1. Model Specification

We begin by describing GSCA's model specification to facilitate understanding of DL-GSCA's model specification. To specify path-analytic relationships between indicators and components, GSCA involves three sub-models: weighted relation, component measurement, and structural models. The weighted relation model indicates that components are defined as linear deterministic functions of their respective indicators. The weights assigned to indicators to form their components are called (component) weights. The component measurement model shows that indicators are explained by their components and these relationships are signified by (component) loadings. The structural model specifies path-analytic linear associations between components, represented by path coefficients.

Let $\boldsymbol{\gamma}$ and $\mathbf{z}$ denote random vectors of components and indicators, respectively, both of which are assumed to be standardized. Let $\mathbf{W}$, $\mathbf{C}$, and $\mathbf{B}$ denote matrices consisting of weights, loadings, and path coefficients, respectively. Let $\boldsymbol{\xi}$ denote a vector of errors for $\mathbf{z}$ in the component measurement model. Let $\boldsymbol{\zeta}$ denote a vector of errors for $\boldsymbol{\gamma}$ in the structural model. Then, the three sub-models of GSCA are expressed as

$$\boldsymbol{\gamma} = \mathbf{W}'\mathbf{z} \text{ (weighted relation model)} \quad (4.1)$$

$$\mathbf{z} = \mathbf{C}'\boldsymbol{\gamma} + \boldsymbol{\xi} \text{ (component measurement model)} \quad (4.2)$$

$$\boldsymbol{\gamma} = \mathbf{B}'\boldsymbol{\gamma} + \boldsymbol{\zeta} \text{ (structural model)}, \quad (4.3)$$

where $\mathbf{X}'$ denotes the transpose of a matrix $\mathbf{X}$.

DL-GSCA also involves the same sub-models. However, components in DL-GSCA are defined as any continuous (linear or nonlinear) deterministic functions of their indicators in the weighted relation model and are allowed to be nonlinearly related to their respective indicators in the component measurement model. Let f_W and f_C denote continuous functions of $\mathbf{z}$ and $\boldsymbol{\gamma}$, respectively. Then, DL-GSCA's three sub-models are given as

$$\boldsymbol{\gamma} = f_W(\mathbf{z}) \text{ (weighted relation model)} \tag{4.4}$$

$$\mathbf{z} = f_C(\boldsymbol{\gamma}) + \boldsymbol{\xi} \text{ (component measurement model)} \tag{4.5}$$

$$\boldsymbol{\gamma} = \mathbf{B}'\boldsymbol{\gamma} + \boldsymbol{\zeta} \text{ (structural model).} \tag{4.6}$$

If f_W and f_C are confined to be linear, (4.4) and (4.5) become equivalent to (4.1) and (4.2), respectively, indicating that DL-GSCA includes GSCA as a special case. A more detailed description of DL-GSCA's model specification is provided in Appendix D1.

4.2.2. Parameter Estimation

In DL-GSCA, we utilize DL for estimating f_W and f_C in a data-driven manner. Let h_W and h_C denote a set of DL's artificial neural networks that approximate f_W and f_C, respectively. Two hyperparameters—the number of hidden layers and the number of hidden units per hidden layer for each component—determine the basic forms of h_W and h_C, as discussed in detail in Appendix D2.

Given the matrix of standardized indicators and the two hyperparameters, we aim to estimate h_W, h_C, and $\mathbf{B}$ by minimizing a single optimization criterion that is equivalent to the average residual variance of all dependent variables in the model. We develop an alternating least squares (ALS) algorithm to minimize the optimization criterion iteratively. The ALS algorithm begins by dividing the model parameters into two sets and alternately updates each set of the parameters with the other set fixed until the criterion's difference between consecutive iterations becomes smaller than a prescribed tolerance level (e.g., .0001). Once h_W, h_C, and $\mathbf{B}$ are estimated, the standard errors and confidence intervals of path coefficient

estimates are obtained based on the bootstrap method (Efron, 1979, 1982). We provide a detailed description of the ALS algorithm in Appendix D3.

Prior to the implementation of the ALS algorithm, it is important to predetermine the values of the two hyperparameters. These values affect the potential capacity of h_W and h_C to approximate f_W and f_C (Lu et al., 2017). For example, by increasing the number of hidden layers and/or that of hidden units per layer for h_W and h_C, one can approximate a more intricate form of f_W and f_C, thereby reducing the bias in h_W and h_C. However, this approach can require estimating a large number of model parameters, which in turn tends to increase the variance in h_W and h_C. This phenomenon is often referred to as the bias-variance trade-off (Hastie et al., 2001, Chapter 7.3). If the increased variance outweighs the bias reduction, the model may overfit (e.g., Strang, 2019, p. 374) and fails to generate optimal components for predicting dependent variables in test samples. It is, therefore, essential to optimally calibrate the two hyperparameters of h_W and h_C to ensure the maximal predictive power of the components.

Nonetheless, it should be noted that there is, as yet, no universally agreed-upon method for determining hyperparameter values of deep learning methods (Mas & Flores, 2008). Within the context of DL-GSCA, we adopt the predictive feedforward search algorithm (Cho et al., 2022), which was originally proposed for variable selection in GSCA. This search algorithm gradually increases the hyperparameter values until any further increase results in a higher expected prediction error of the model. A detailed description of the search algorithm is presented in Appendix D4.

Lastly, it is worth noting that DL-GSCA's approach to estimating components can be categorized as unsupervised learning. This is due to its primary focus on "generating" components capable of accurately predicting their outcomes, including indicators and dependent components, without necessitating a pre-collected dataset of components. In fact,

if researchers do not specify the structural model (6), the DL-GSCA model becomes identical to a set of autoencoders (e.g., Q. V. Le, 2015; Rumelhart et al., 1986), each of which takes a specific set of indicators as both inputs and outputs and includes a component as a hidden unit in its hidden layer (refer to Figures B1 and B3).

4.2.3. Model Evaluation

DL-GSCA provides three types of indices for evaluating the overall performance of a specified model: FIT for dependent variables (FIT^D), test error for dependent variables (TE^D), and), and out-of-bag prediction error for dependent variables (OPE^D). We provide the formulae of these overall model fit indices and relevant local model fit indices that we will also discuss below in Appendix D5.

FIT^D is a rescaled version of the traditional goodness-of-fit measure FIT in GSCA (Hwang & Takane, 2014, pp. 26–30). It shows how much variance of dependent variables (both indicators and components) in the model is explained by components on average. This index is equivalent to the average of the R^2 values for all dependent variables, which ranges from 0 to 1. The larger the FIT^D value, the more variance is explained. Also, $1 - FIT^D$ is indicative of the average in-sample prediction error for dependent variables in the model. This in-sample prediction error indicates the model's estimated prediction error for the training sample. When every variable is specified as a dependent variable, FIT^D becomes equivalent to FIT. In addition, DL-GSCA provides two local goodness-of-fit indices: $FIT_M{}^D$ and $FIT_S{}^D$. The former shows how much variance of dependent indicators in the component measurement model is explained by components on average (i.e., average R^2 for all dependent indicators), whereas the latter shows how much variance of dependent components in the structural model is explained by other components on average (i.e., average R^2 for all dependent components).

On the other hand, TE^D assesses the average *out-of-sample* prediction error for dependent variables in the model. The value of TE^D indicates the ratio of the average prediction errors of a specified model and the null model that uses the training-sample means of indicators and components as their predicted scores in a test sample. A value of TE^D smaller than 1 indicates that the estimated model shows better predictive performance than the null model in the test sample. If the estimated model perfectly predicts dependent variables in the test sample, its TE^D value will be zero. DL-GSCA also provides $TE_M{}^D$ and $TE_S{}^D$ for respectively assessing the average out-of-sample prediction errors for dependent indicators in the component measurement model and for dependent components in the structural model, relative to the corresponding null models (i.e., in a test sample, the null component measurement model uses the training-sample means of indicators as the indicators' predicted scores and the null structural model uses the training-sample means of components as the components' predicted scores).

In practice, researchers may not have a designated test sample, or the sample size of their dataset may be too small. This can make it infeasible to set aside a portion of the dataset as a test sample and to calculate the TE^D value of the trained DL-GSCA model. To address this issue, DL-GSCA provides OPE^D, which is a rescaled version of the traditional cross-validation index OPE in GSCA (Cho et al., 2019). This index can estimate the expected prediction error of the DL-GSCA model, when a test sample is not available. The OPE^D value has the same interpretation as that of TE^D, but it is primarily used for comparing models (Cho et al., 2019). For example, when there are several competing models, the model with the smallest OPE^D value can be selected as the final one in terms of predictive generalizability. DL-GSCA also provides $OPE_M{}^D$ and $OPE_S{}^D$, which aim to estimate the expected prediction error of the measurement and structural models, respectively.

When a test sample is given, researchers can additionally evaluate the predictive power of individual predictor components through the use of $\Delta TE_{p,q}$, which assesses the contribution of one predictor component, denoted by component *p*, to prediction of its dependent component, denoted by component *q*, in the model. If the value of $\Delta TE_{p,q}$ is negative, including the predictor component in the model worsens the prediction of the dependent component. Conversely, if the value of $\Delta TE_{p,q}$ is positive, adding the predictor component contributes to better predicting the dependent component.

4.3. Empirical Application

The datasets we analyze come from two Human Development Reports (HDR) published by the United Nations Development Programme (UNDP, 2010, 2016) in 2010 and 2016. Under the premise that "the real wealth of nations" is people, not the real and financial assets accumulated (UNDP, 1990, p. 9), the UNDP created a composite index for measuring countries' development levels for humans, named the Human Development Index (HDI), and has published each country's HDI score and ranking annually since 1990. The HDI is evaluated in terms of three domains of human development—health, knowledge, and decent standard of living. Specifically, since the revision of the HDI formula in 2010, life expectancy at birth (LifeExp) has been used as the indicator for health, mean schooling years (MeanSchl) and expected schooling years (ExpSchl) as the indicators for knowledge, and gross national income per capital (GNI) as the indicator for decent standard of living.

Although the provision of the HDI plays a role in shifting a national paradigm of development from materialistic growth to human well-being in many countries (UNDP, 2016, p. 2), this index has been constantly criticized in two respects. First, many researchers point out that the formula used to calculate the HDI is rather arbitrary and difficult to justify because the formula has been heavily affected by the developers' opinions (e.g., Noorbakhsh,

1998; Nübler, 1995). Second, researchers are concerned that the HDI is not developed to sufficiently account for other important domains of human development, including ecological sustainability (e.g., Sagar & Najam, 1998), political freedom (e.g., Ranis et al., 2006), and subjective well-being (e.g., Blanchflower & Oswald, 2005). In fact, the HDI has been developed based solely on a group of experts' judgements on which change/growth in the country should be indicative of development. However, it seems to be of particular importance to take into account subjective well-being, which informs "whether progress has indeed occurred if the (partial) metrics of human development suggest it appears to have" (Hall & Helliwell, 2014, p. 11), reflecting people's actual opinions about which changes in the HDI domains they perceive as real-life developments in their country (e.g., Blanchflower & Oswald, 2005). Thus, some researchers suggest determining the weights for the HDI indicators based on their contributions to subjective well-being (e.g., Blanchflower & Oswald, 2005; Nübler, 1995).

DL-GSCA may be used for *statistically* addressing these issues with the HDI. It does not need to pre-determine a formula for combining the four HDI indicators or equivalently how the weights for the indicators are to be determined. Instead, DL-GSCA estimates the functional form between the HDI and its indicators in a data-driven fashion such that the index maximizes the total explained variance of dependent variables in the DL-GSCA model. Particularly, subjective well-being can be specified as a focal dependent variable for the HDI in the model. Then, DL-GSCA will estimate the weights for the HDI indicators in such a way that the index is highly associated with subjective well-being.

Figure 4.2 displays the HDI model that we specified for illustrative purposes. We consider the HDI an exogenous component (i.e., $\boldsymbol{\gamma}_1$ = HDI), which is associated with four indicators (i.e., $z_{1,1}$ = LifeExp, $z_{1,2}$ = MeanSchl, $z_{1,3}$ = ExpSchl, and $z_{1,4}$ = GNI). We also assume that the HDI influences two types of subjective well-being. One type is the average

level of satisfaction with the country's environment for development (SED; γ_2 = SED), which is measured by three indicators: $z_{2,1}$ = proportion of the people satisfied with healthcare quality (SatisHQ), $z_{2,2}$ = proportion of the people satisfied with education quality (SatisEQ), and $z_{2,3}$ = proportion of the people satisfied with state of living (SatisSL). The other type is overall life satisfaction (OLS; γ_3 = OLS), which is measured by a single indicator, $z_{3,1}$ = the average life satisfaction of the people living in a country (Cantril, 1965). We further assume that SED influences OLS, as it seems reasonable that an individual satisfied with his or her country's environment is more likely to be satisfied with his or her overall life (e.g., Silva et al., 2012).

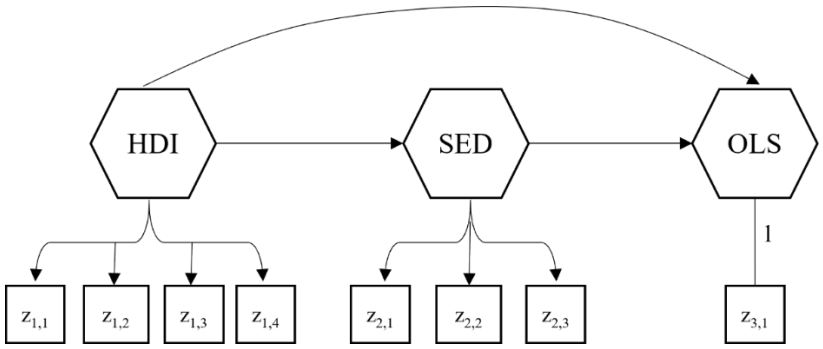


Figure 4.2. The Human Development Index (HDI) model specified for the empirical application. Hexagons and squares denote components and indicators, respectively. Each set of curved arrows shows which block of indicators is to be explained by which component. A straight line labelled with 1 indicates that OLS is set to be identical to $z_{3,1}$. Error terms are omitted.

We apply DL-GSCA to fit the model to the 2010 HDR data. Of 169 countries, we use 142 countries that have no missing observations (N = 142). We wrote a MATLAB code for DL-GSCA[3] for this analysis. We consider three levels of the number of hidden layers, denoted by L_p in Appendix D2, for HDI and SED (L_p = 0, 1, and 2). When the number of hidden layers is greater than zero, we consider four levels of hidden units per hidden layer, denoted by $R_p^{(l)}$ in Appendix D2 ($R_p^{(l)}$ = 2, 3, 4, and 5). We employ Cho et al.'s (2022) search algorithm with five-fold cross validation for deciding the two hyperparameters. We then

[3] The MATLAB code and the HDR data are available at https://osf.io/wdz79/?view_only=39cbbe18a7f442d381e2483b4bced177.

apply the ALS algorithm to each training sample, using ten different sets of initial values. The parameter estimates obtained from the set of initial values, which results in the largest FIT^D, are chosen as the final estimates. The tolerance level used in the study is .0001. For comparison, we also apply GSCA to the same data. The total number of bootstrap samples is 1000.

We choose one hidden layer for both HDI and SED and four and three hidden units per hidden layer for HDI and SED, respectively, based on the predictive feedforward search algorithm. We then proceed to apply the ALS algorithm with ten different sets of random initial values. Figure 4.3 displays how the value of the optimization criterion (D.6) changes as the ALS algorithm iterates through its two steps with a set of randomly assigned initial values. With each iteration, the value of (D.6) always decreases, indicating that the ALS algorithm is successfully finding parameter values that result in a smaller value of (D.6) than the previous iteration.

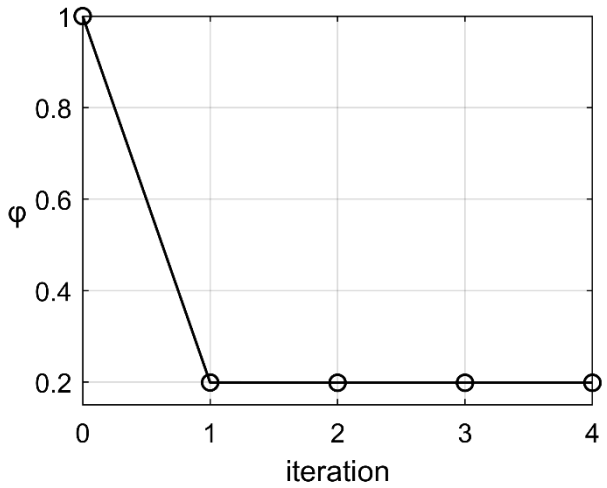


Figure 4.3. Plot of the optimization criterion value (D.6) versus the number of iterations.

DL-GSCA provides $FIT^D = .81$, indicating that the model explains 81% of the total variance of all dependent variables in the model. Also, DL-GSCA provides $FIT_M{}^D = .86$ and $FIT^D{}_S = .65$, indicating that the component measurement model accounts for 86% of the total

variance of all dependent indicators, whereas the structural model explains 65% of the total variance of all dependent components. Conversely, GSCA provides $FIT^D = .75$, $FIT_M^D = .79$, and $FIT_S^D = .60$. This indicates that DL-GSCA provides smaller in-sample prediction errors for both indicators and dependent components than GSCA.

Figure 4.4 displays the indicators' values predicted by DL-GSCA against their original values. The three scatterplots show that the HDI and SED components appear to capture the nonlinear relationships of their respective indicators sufficiently well, indicating that they can serve as a good summary of their indicators. The average R^2 value of the HDI and SED components for their respective indicators are .89 and .82, respectively.

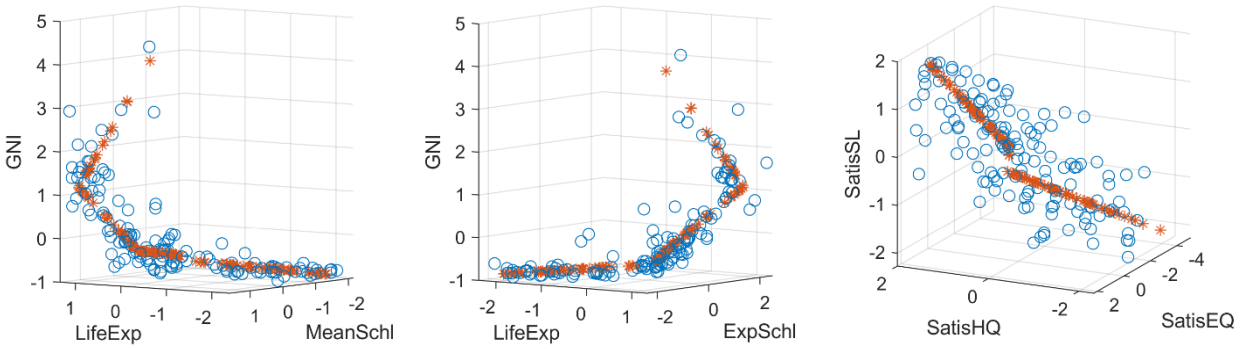
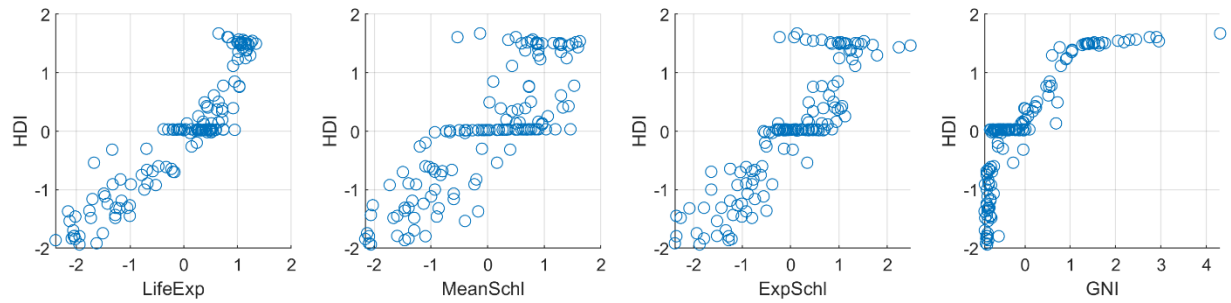


Figure 4.4. Scatterplots of indicators' scores (blue circles) and their predicted values (orange stars) obtained from DL-GSCA in the empirical application.

Figure 4.5 depicts the relationships among the HDI and SED components and their indicators obtained from DL-GSCA. The individual scores of the components are obtained by plugging their indicators' scores in the estimated weighted relation model (i.e., $\mathbf{g}_{n,p} = h_{w,p}(\mathbf{d}_{n,p})$ in Eq. (D.5) in Appendix D3). In general, an increase in the indicators' scores is associated with an increase in their component scores, indicating that an improvement in each domain of the HDI and SED is likely to be related to an increase in the level of human development and people's satisfaction with their living environment for development. It seems intriguing that a small amount of increase in GNI is linked to a steep growth in the HDI for countries with

extremely low levels of GNI, suggesting that an economic growth may contribute substantially to human development for these countries. On the other hand, for those with relatively high levels of GNI (i.e., greater than around a standardized score of 1.5 or equivalently 36,654 US dollars), an increase in GNI is associated with only a marginal increase in the HDI. This may be because GNI is negatively associated with the other domains of the HDI among the countries with high GNI levels, as displayed in the first two scatterplots of Figure 4.5.

(A) HDI



(B) SED

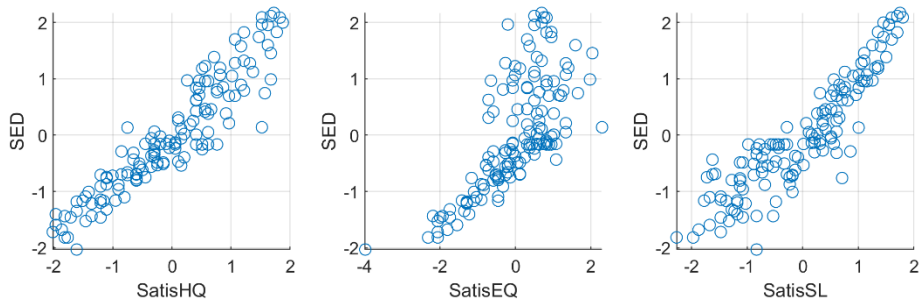


Figure 4.5. Scatterplots of two components (HDI and SED) and their indicators obtained from DL-GSCA in the empirical application.

In the structural model, the HDI has a positive and statistically significant effect on SED ($b_{1,2}$ = .76, SE = .04, 95% CI = [.69, .83]), indicating that a higher level of human development tends to lead to a higher level of people's satisfaction with their living environment. Also, SED has a positive and statistically significant effect on OLS ($b_{2,3}$ = .45, SE = .06, 95% CI = [.33, .58]), indicating that people satisfied with their living environment

are likely to have a high level of overall life satisfaction. Moreover, the HDI has a positive and statistically significant effect and on OLS ($b_{1,3}$ = .44, SE = .07, 95% CI = [.30, .56]), indicating that people living in a country with a relatively high level of human development tend to have a relatively high level of satisfaction with their life, controlling for satisfaction with their living environment. Furthermore, the HDI has a positive and statistically significant indirect effect on OLS, mediated by SED ($b_{1,2} \times b_{2,3}$ = .35, SE = .05, 95% CI = [.25, .46]). The total effect of the HDI on OLS (.79) is also positive and statistically significant (SE = .03, 95% CI = [.73, .84]).

To evaluate the model's predictive power, we evaluate its out-of-sample prediction error, using the 2016 HDR dataset as a test sample. We use 156 countries without missing observations. DL-GSCA provides TE^D = .24, indicating that on average, DL-GSCA's prediction error is 24% of the null model's prediction error. This value is smaller than that of GSCA (TE^D = .29). DL-GSCA provides $TE_M{}^D$ = .17 and $TE_S{}^D$ = .52, indicating that DL-GSCA's prediction errors for the measurement and structural models are 17% and 52% of the respective null models. They are also smaller than the counterparts of GSCA ($TE_M{}^D$ = .23 and $TE_S{}^D$ = .57). These results show the superior predictive performance of DL-GSCA to GSCA. In addition, all $\Delta TE_{p,q}$ values obtained from DL-GSCA are positive (i.e., $\Delta TE_{1,2}$ = .98, $\Delta TE_{2,3}$ = .17, and $\Delta TE_{1,3}$ = .09), indicating that all the predictor components contribute to predicting their dependent components.

For illustrative purposes, suppose that we do not have any additional dataset to use as a test sample. In such a case, as typically recommended in the literature (e.g., Hastie et al., 2001, p. 222), we might use about 75% of observations in the 2010 HDR dataset for model training and the rest for model testing. However, due to the limited number of countries in this dataset, we opt to use OPE^D, $OPE_M{}^D$, and $OPE_S{}^D$ values to evaluate the out-of-sample errors of the competing DL-GSCA and GSCA models, rather than TE^D, $TE_M{}^D$, and $TE_S{}^D$

values. DL-GSCA yields OPE^D = .20, whose value is smaller than that of GSCA (OPE^D = .27). This suggests that DL-GSCA can be chosen over GSCA in terms of predictive generalizability Additionally, DL-GSCA provides OPE_M^D = .15 and OPE_S^D = .36, both of which are also smaller than GSCA's counterparts (OPE_M^D = .23 and OPE_S^D = .43). This indicates that DL-GSCA's nonlinear components are expected to have smaller prediction error for outcome variables in both measurement and structural models than GSCA's linear components.

Taken together, DL-GSCA is applied to construct a composite index for human development. Without assuming any prescribed formula for calculating the HDI, DL-GSCA estimates the formula in a data-driven manner, which leads the HDI to be constructed to sufficiently capture potential nonlinear associations among its four indicators while predicting two types of subjective well-being. In addition, DL-GSCA's components, including the HDI, have higher in-sample and out-of-sample prediction powers than GSCA's components.

4.4. Simulation Study

We conduct a simulation study to examine the explanatory and predictive power of DL-GSCA when nonlinear relationships exist between indicators in the model. For this study, we design a population DL-GSCA model that imitates the structural model as the HDI model, as shown in Figure 4.6. We describe how to generate data based on the population model in Appendix D6.

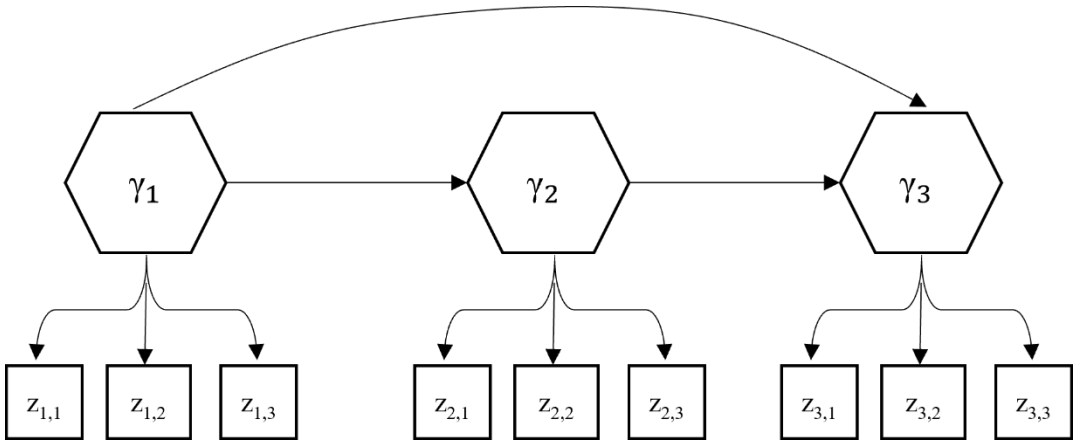


Figure 4.6. A population DL-GSCA model used in the simulation study.

We consider four levels of sample size (N = 100, 200, 500, and 1000), for each of which we generate 500 training samples. Additionally, we generate one test sample of N = 2000. Figure 4.7 displays the association patterns of each block of indicators in the test sample. In this simulation study, we employ 30% of each training sample as a validation sample to determine the values of the hyperparameters (i.e., L_p and $R_p^{(l)}$) from prescribed candidate values per hyperparameter (i.e., L_p = 0, 1, and 2; $R_p^{(l)}$ = 2, 3, 4, and 5). We use the same number of initial values and tolerance level used in the previous section.

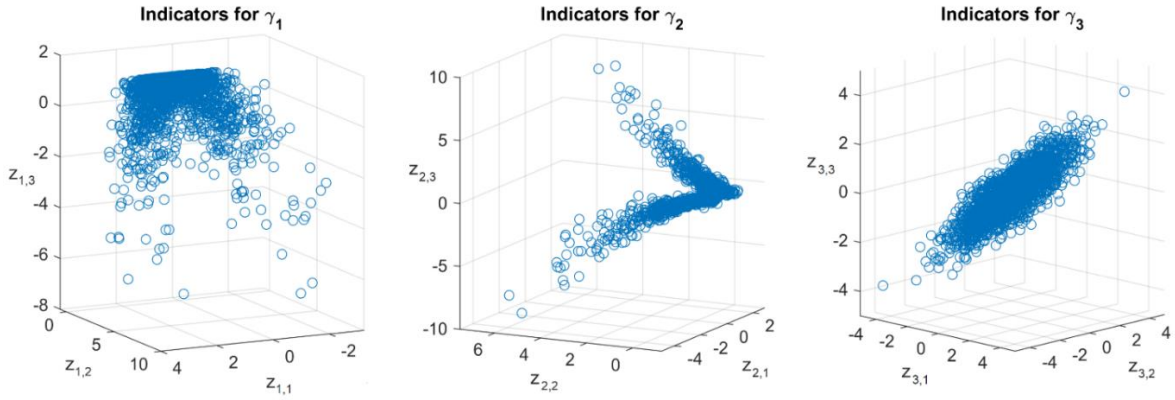


Figure 4.7. Scatterplots of indicators per component in the test sample for the simulation study.

We compute the values of FIT^D, $FIT_M{}^D$, and $FIT_S{}^D$ based on the final parameter estimates to evaluate the model's explanatory power (or in-sample prediction error) for each

training sample. We also calculate the TE^D, $TE_M{}^D$, and $TE_S{}^D$ values based on the test sample to assess the model's predictive power (or out-of-sample prediction error). We apply traditional GSCA as a benchmark for comparing DL-GSCA's performance in both in-sample and out-of-sample prediction errors.

Figure 4.8 displays the average values of FIT^D, $FIT_M{}^D$, and $FIT_S{}^D$ obtained from DL-GSCA and GSCA per sample size. As shown in the figure, DL-GSCA provides greater average FIT^D values than GSCA in all sample sizes, indicating that overall, the nonlinear components of DL-GSCA explain the variance of the dependent variables better than the linear components of GSCA. The difference in the average FIT^D values between the two methods remains unchanged across the sample sizes. The same patterns are observed in the average $FIT_M{}^D$ and $FIT_S{}^D$ values of the methods.

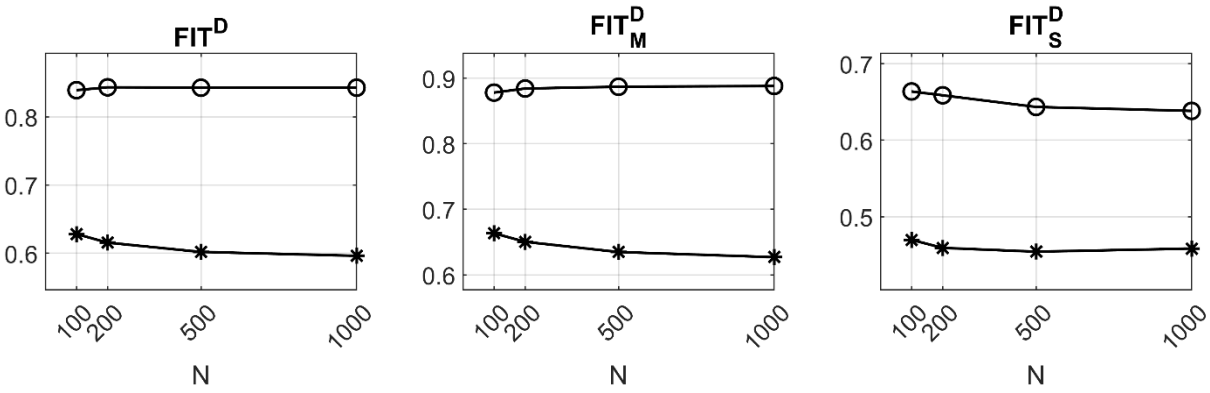


Figure 4.8. In-sample performance of DL-GSCA and GSCA in the simulation study. The average values of FIT^D, $FIT_M{}^D$, and $FIT_S{}^D$ obtained from DL-GSCA and GSCA per sample size. O = DL-GSCA and * = GSCA.

Figure 4.9 shows the average values of $1 - TE^D$, $1 - TE_M{}^D$, and $1 - TE_S{}^D$ from DL-GSCA and GSCA per sample size. DL-GSCA provides smaller average TE^D values than GSCA in all sample sizes, indicating that DL-GSCA generally outperforms GSCA in terms of predictive power. The TE^D values of both methods tend to decrease on average as the

sample size increases, but the difference in the values between the methods does not diminish. The average TE_M^D and TE_S^D values of the methods show the same patterns.

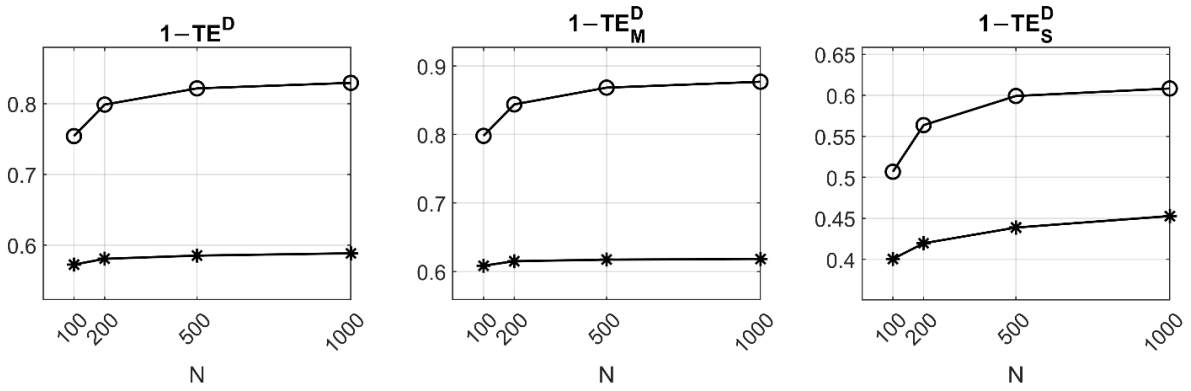
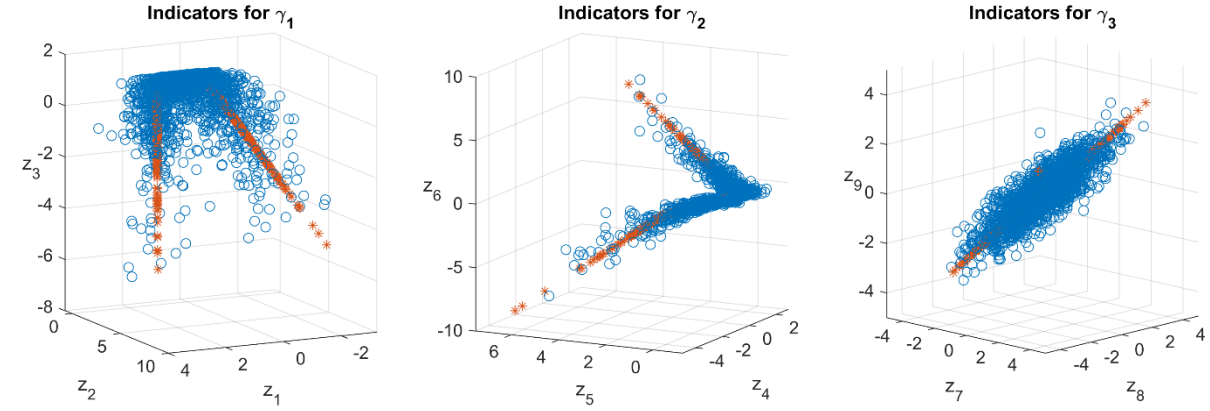


Figure 4.9. Out-of-sample performance of DL-GSCA and GSCA in the simulation study. The average values of 1 – TE^D, 1 – TE_M^D, and 1 – TE_S^D obtained from DL-GSCA and GSCA per sample size. O = DL-GSCA and * = GSCA.

To further compare the predictive performance of DL-GSCA and GSCA, we generate another training sample of N = 2000 and estimate their parameters from this sample. Then, we use the estimates to obtain the predicted values of the indicators in the test sample. Figure 4.10 displays the indicators' values predicted by DL-GSCA and GSCA against their original values in the test sample. It clearly shows that DL-GSCA's components can produce relatively accurate predictions of the three blocks of indicators, efficiently capturing their nonlinear and linear relationships. On the other hand, GSCA's components fail to make accurate predictions of the first two blocks of indicators ($z_{1,1}$ to $z_{2,3}$), leading its indicators' predicted values to deviate substantially from their original values. This indicates that GSCA is not optimal for predicting the nonlinear associations among the indicators for components 1 and 2, as expected.

(A) DL-GSCA



(B) GSCA

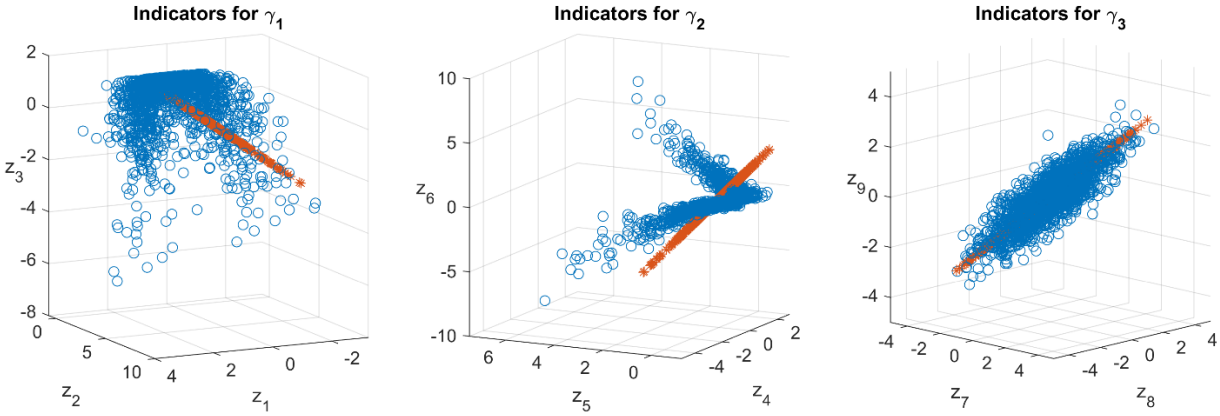


Figure 4.10. Scatterplots of indicators' scores (blue circles) and their predicted values (orange stars) per component obtained from DL-GSCA and GSCA in the simulation study.

In sum, our simulation study shows that when some indicators are nonlinearly related, on average, DL-GSCA results in smaller in-sample and out-of-sample prediction errors than GSCA in all conditions, speaking to DL-GSCA's superior performance to GSCA in terms of both explanatory and predictive power.

4.5. Concluding Remarks

We proposed a new extension of GSCA, termed deep learning GSCA (DL-GSCA), which incorporates DL's functions to produce components that can capture nonlinear associations among their indicators in a data-driven manner without the need of pre-specifying the exact

functional forms between the components and indicators. We also provided model evaluation indices to assess a model's explanation (in-sample) and prediction (out-of-sample) errors.

We conducted both real and simulated data analyses and demonstrated the superior performance of DL-GSCA to GSCA in the presence of nonlinear relationships among indicators per component. Moreover, we illustrated DL-GSCA's potential for creating the HDI while capturing nonlinear associations among the index's indicators and considering its hypothetical relationships with other variables. Unlike the traditional procedures for developing the HDI, in which a group of experts determine its computational formula without considering its relationships with other variables, DL-GSCA statistically estimated the formula for the HDI such that it could be a good summary of its indicators and good predictors for two types of subjective well-being.

DL-GSCA contributes substantially to broadening the scope of GSCA beyond linear modeling and improving prediction accuracy. Nonetheless, it has limitations as well. For example, the proposed method in its current form assumes that components are always linearly related in the structural model. We here intend to focus on capturing nonlinear associations among indicators per component in the component measurement model while keeping the interpretation of the relationships among components as simple as possible. However, some researchers may want to specify nonlinear associations between components while still defining components as linear functions of their indicators (e.g., Basco et al., 2021; Hwang, Ho, et al., 2010). For example, they may be interested in examining interaction effects of components, keeping the component scores easily interpretable (e.g., Hwang, Cho, Jin, et al., 2021). Thus, it may be necessary to incorporate DL's functions into the structural model as well, so that researchers can freely decide which sub-model(s) can be nonlinear.

Moreover, DL-GSCA currently assumes that all dependent variables are continuous. Consequently, it is not suitable to estimate models with categorical dependent variables. For

example, in genomic modeling, a set of single nucleotide polymorphisms (SNPs) within a gene is typically categorical variables (e.g., Romdhani et al., 2015). To handle such cases, we may need to extend DL-GSCA to adopt a different activation function for categorical variables (e.g., sigmoid or softmax; see Nwankpa et al., 2021), minimizing a mixture of two optimization criteria, e.g., the mean squares error for continuous variables and cross-entropy for categorical variables.

Furthermore, DL-GSCA needs to accommodate various technical extensions of GSCA. For instance, it should be extended to deal with higher-order components (Hwang & Takane, 2014, pp. 99–110). In the empirical application, for example, the HDI may be considered a second-order component which is linked to the three domains of human development (i.e., health, knowledge, and decent standard of living) as its first-order components. Also, it would be beneficial to extend DL-GSCA to efficiently handle missing observations which frequently occur in practice (Graham, 2008). We may adopt GSCA's model-based imputation approach (Hwang & Takane, 2014, pp. 123–125), which treats missing values as parameters and estimates them along with the other model parameters by minimizing a single optimization criterion. In addition, we may consider combining DL-GSCA with regularization techniques, such as ridge and lasso, to avoid potential overfitting. DL-GSCA may be more susceptible to the overfitting issue than GSCA because the former will likely have a far larger number of parameters than the latter. We may borrow the same idea of regularized GSCA (Hwang, 2009; Hwang & Takane, 2014, Chapters 8 and 9) to shrink DL-GSCA's parameter estimates toward zero or to exact zero.

In closing, DL-GSCA represents a flexible, nonlinear multivariate method for estimating complex path-analytic models involving components. It aims to capture potential nonlinear associations between components and indicators while examining linear associations among components. Thus, DL-GSCA can substantially improve the predictive

power of its components while maintaining the interpretability of the relationships among the components. We concentrate on proposing DL-GSCA's general framework in the paper and will need to refine and extend the method in various ways to deal with a broad range of data-analytic issues, including those enumerated above. Moreover, it will be important to develop a software program for DL-GSCA to make it more accessible to researchers and practitioners. For example, DL-GSCA can be included in GSCA's free user-friendly software – GSCA Pro (Hwang et al., 2023). This can also contribute to the application of DL-GSCA to a greater variety of real-world problems and more thorough investigations of its practical usefulness.

Finally, we considered a maximum of two hidden layers per neural network (L_p) in our empirical analyses. This may give the impression that DL-GSCA can only handle a shallow neural network with a small number of hidden layers. However, in our DL-GSCA model, the overall structure is not represented merely by the layers of a single nested neural network. As described in Appendix D2, the model's architecture consists of several interconnected neural networks. Therefore, the total number of hidden layers in the entire model is always significantly higher than the number presented in any single nested network (L_p). In addition, DL-GSCA does not impose any limit on the number of hidden layers (L_p) in each individual network. Researchers, depending on the availability of computational resources, can consider a greater value of L_p than the ones employed in our empirical studies.

References

Aggarwal, C. C. (2018). An Introduction to neural networks. In C. C. Aggarwal (Ed.), *Neural Networks and Deep Learning: A Textbook* (pp. 1–52). Springer International Publishing. https://doi.org/10.1007/978-3-319-94463-0_1

Basco, R., Hair, J. F., Ringle, C. M., & Sarstedt, M. (2021). Advancing family business research through modeling nonlinear relationships: Comparing PLS-SEM and multiple regression. *Journal of Family Business Strategy*, 100457. https://doi.org/10.1016/j.jfbs.2021.100457

Blanchflower, D. G., & Oswald, A. J. (2005). Happiness and the human development index: The paradox of Australia. *Australian Economic Review*, *38*(3), 307–318. https://doi.org/10.1111/j.1467-8462.2005.00377.x

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. General considerations. *IMA Journal of Applied Mathematics*, *6*(1), 76–90. https://doi.org/10.1093/imamat/6.1.76

Cantril, H. (1965). *The pattern of human concerns*. Rutgers University Press.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th Annual Convention of the American Psychological Association*, *3*, 227–228.

Cho, G. (2023). *Constructs may or may not be latent: Studies on two domains of structural equation modeling*. McGill University, Montreal, Canada.

Cho, G., & Choi, J. Y. (2020). An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation

models. *Behaviormetrika*, *47*(1), 243–272. https://doi.org/10.1007/s41237-019-00098-0

Cho, G., Hwang, H., Sarstedt, M., & Ringle, C. M. (2022). A prediction-oriented specification search algorithm for generalized structured component analysis. *Structural Equation Modeling: A Multidisciplinary Journal*. https://doi.org/10.1080/10705511.2022.2057315

Cho, G., Jung, K., & Hwang, H. (2019). Out-of-bag prediction error: A cross validation index for generalized structured component analysis. *Multivariate Behavioral Research*, *54*(4), 505–513. https://doi.org/10.1080/00273171.2018.1540340

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

Choi, S., Lee, S., Huh, I., Hwang, H., & Park, T. (2020). HisCoM-G×E: Hierarchical structural component analysis of gene-based gene–environment interactions. *International Journal of Molecular Sciences*, *21*(18), 1–11. https://doi.org/10.3390/ijms21186724

Csaji, B. C. (2001). *Approximation with artificial neural networks* [Eötvös Loránd University]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.2647&rep=rep1&type=pdf

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314. https://doi.org/10.1007/BF02551274

de Jong, S., & Kiers, H. A. L. (1992). Principal covariates regression: Part I. Theory. *Chemometrics and Intelligent Laboratory Systems*, *14*(1), 155–164. https://doi.org/10.1016/0169-7439(92)80100-I

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

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, *13*(3), 317–322. https://doi.org/10.1093/comjnl/13.3.317

Gensowski, M., Heckman, J., & Savelyev, P. (2011). *The effects of education, personality, and IQ on earnings of high-ability men*. https://conference.iza.org/conference_files/CoNoCoSk2011/gensowski_m6556.pdf

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, *24*(109), 23–26. https://doi.org/10.1090/S0025-5718-1970-0258249-6

Graham, J. W. (2008). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Hall, J., & Helliwell, J. (2014). *Happiness and human development*. http://akwl.org/wp-content/uploads/2019/12/happiness_and_hd.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Springer. https://doi.org/10.1007/978-0-387-21606-5

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika*, *74*(3), 517–530. https://doi.org/10.1007/S11336-009-9119-Y

Hwang, H., & Cho, G. (2020). Global least squares path modeling: A full-information alternative to partial least squares path modeling. *Psychometrika*, *85*, 947–972. https://doi.org/10.1007/s11336-020-09733-2

Hwang, H., Cho, G., & Choo, H. (2023). *GSCA Pro (Version 1.2.1) [Computer software]*. http://www.gscapro.com.

Hwang, H., Cho, G., Jin, M. J., Ryoo, J. H., Choi, Y., & Lee, S.-H. (2021). A knowledge-based multivariate statistical method for examining gene-brain-behavioral/cognitive relationships: Imaging genetics generalized structured component analysis. *PloS One*, *16*(3), e0247592. https://doi.org/10.1371/journal.pone.0247592

Hwang, H., Desarbo, W. S., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, *72*(2), 181–198. https://doi.org/10.1007/s11336-005-1314-x

Hwang, H., Ho, M.-H. R., & Lee, J. (2010). Generalized structured component analysis with latent interactions. *Psychometrika*, *75*(2), 228–242. https://doi.org/10.1007/s11336-010-9157-5

Hwang, H., Sarstedt, M., Cheah, J.-H., & Ringle, C. M. (2020). A concept analysis of methodological research on composite-based structural equation modeling: Bridging PLSPM and GSCA. *Behaviormetrika*, *47*(1), 219–241. https://doi.org/10.1007/s41237-019-00085-5

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, *69*(1), 81–99. https://doi.org/10.1007/BF02295841

Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling*. Chapman and Hall/CRC Press.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, *58*(3), 433–451. https://doi.org/10.1093/biomet/58.3.433

Kumar, S. K. (2017). *On weight initialization in deep neural networks*. https://arxiv.org/abs/1704.08863

Le, P. V. (2014). *More schooling is not always better: Evidence from an Instrumental variables approach to educational reform in Vietnam*. https://fsppm.fulbright.edu.vn/cache/More Schooling Is Not Always Better-2014-12-19-08341367.pdf

Le, Q. V. (2015). A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks. *Google Brain*, *20*, 1–20. http://ai.stanford.edu/~quocle/tutorial2.pdf

Lecun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The expressive power of neural networks: A view from the width. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6231–6239). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/32cbf687880eb1674a07bf717761dd3a-Paper.pdf

Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, *29*(3), 617–663. https://doi.org/10.1080/01431160701352154

Noorbakhsh, F. (1998). The human development index: Some technical issues and alternative indices. *Journal of International Development*, *10*(5), 589–605. https://doi.org/10.1002/(SICI)1099-1328(199807/08)10:5<589::AID-JID484>3.0.CO;2-S

Nübler, I. (1995). The human development index revisited. *Intereconomics*, *30*(4), 171–176.

https://doi.org/10.1007/BF02928088

Nwankpa, C. E., Ijomah, W., Gachagan, A., & Marshall, S. (2021). Activation functions. *2nd International Conference on Computational Sciences and Technology*, 124–133. https://pure.strath.ac.uk/ws/portalfiles/portal/118946797/Nwankpa_etal_ICCST_2021_Activation_functions_comparison_of_trends_in_practice.pdf

Park, J. H. (1994). *Returns to schooling: A peculiar deviation from linearity* (No. 714; Working Papers, Issue 714). https://ideas.repec.org/p/pri/indrel/335.html

Ranis, G., Stewart, F., & Samman, E. (2006). Human development: Beyond the human development index. *Journal of Human Development*, *7*(3), 323–358. https://doi.org/10.1080/14649880600815917

Romdhani, H., Hwang, H., Paradis, G., Roy-Gagnon, M. H., & Labbe, A. (2015). Pathway-based association study of multiple candidate genes and multiple traits using structural equation models. *Genetic Epidemiology*, *39*(2), 101–113. https://doi.org/10.1002/gepi.21872

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books. https://apps.dtic.mil/sti/citations/AD0256582

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Sagar, A. D., & Najam, A. (1998). The human development index: a critical review. *Ecological Economics*, *25*(3), 249–264. https://doi.org/10.1016/S0921-8009(97)00168-7

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, *24*(111), 647–656. https://doi.org/10.2307/2004840

Silva, J., de Keulenaer, F., & Johnstone, N. (2012). *Environmental quality and life satisfaction* (OECD Environment Working Papers, Issue 44). https://doi.org/10.1787/5k9cw678dlr0-en

Strang, G. (2019). *Linear algebra and learning from data* (1st ed.). Wellesley-Cambridge Press.

Takane, Y., & Hwang, H. (2005). An extended redundancy analysis and its applications to two practical examples. *Computational Statistics and Data Analysis*, *49*(3), 785–808. https://doi.org/10.1016/j.csda.2004.06.004

Takane, Y., & Hwang, H. (2007). Regularized linear and kernel redundancy analysis. *Computational Statistics & Data Analysis*, *52*(1), 394–405. https://doi.org/10.1016/j.csda.2007.02.014

Takane, Y., Kiers, H. A. L., & de Leeuw, J. (1995). Component analysis with different sets of constraints on different dimensions. *Psychometrika*, *60*(2), 259–280. https://doi.org/10.1007/BF02301416

Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. F. (2017). Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods. *Psychometrika*, *82*(3), 737–777. https://doi.org/10.1007/s11336-017-9573-x

The MathWorks Inc. (2022). *MATLAB Optimization toolbox: User's Guide*. The MathWorks Inc. https://www.mathworks.com/help/pdf_doc/optim/optim.pdf

UNDP. (1990). *Human development report 1990: Concept and measurement of human development*. http://www.hdr.undp.org/en/reports/global/hdr1990

UNDP. (2010). *Human development report 2010: The real wealth of nations*. http://hdr.undp.org/en/content/human-development-report-2010

UNDP. (2016). *Human development report 2016: Human development for everyone*. http://hdr.undp.org/en/content/human-development-report-2016

Urban, C. J., & Gates, K. M. (2021). Deep learning: A primer for psychologists. *Psychological Methods*. https://doi.org/10.1037/met0000374

van den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical

correlation analysis. *Psychometrika*, *42*(2), 207–219. https://doi.org/10.1007/BF02294050

van der Leeden, R. (1990). *Reduced rank regression with structured residuals*. DSWO Press.

Zhangyang, W., Yingzhen, Y., Shiyu, C., Qing, L., & Huang, T. S. (2016). Learning a deep l∞ encoder for hashing. In S. Kambhampati (Ed.), *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence International Joint Conference on Artificial Intelligence* (pp. 2174–2180). AAAI Press.

Chapter 5. Concluding Remarks

5.1. Summary and Implications

A construct may or may not be latent, suggesting that it could represent an underlying reality causing its observed indicators to covary or merely serve as a summary or descriptive label of an indicator cluster (Binning, 2015). Despite these two distinct possibilities, many psychologists have treated psychological constructs of interest as inherently latent, rarely questioning their true latent status (e.g., Borsboom, 2008; Rhemtulla et al., 2020). Under this unexamined assumption, they have employed structural equation modeling (SEM) to study the relationships between latent constructs based on the data of their indicators. The SEM domain that characterizes each latent construct as a common factor of indicators through its reflective measurement model is known as factor-based SEM.

However, contrary to common expectations (e.g., Brandt et al., 2023), factor-based SEM per se cannot provide conclusive evidence for the existence of latent constructs. As illustrated in Chapter 1, the presence of alternative models, capable of explaining the covariance among indicators, limits the usefulness of the goodness-of-fit of a reflective model for verifying the latent nature of constructs (e.g., Hayduk, 2014). Additionally, some constructs used in psychology, such as socioeconomic status (SES) and genes, can be more suitably viewed as descriptive labels for their respective clusters of indicators. This suggests the need for an alternative SEM domain that does not require constructs to be latent, which is referred to as component-based SEM. This domain represents each construct as a composite or summary index of indicators through its weighted relation and component measurement model, utilizing these indexes to investigate the relationships between constructs.

Given that each SEM domain possesses its own advantages, researchers stand to benefit from equipping their statistical toolbox with both SEM domains, selecting the most

fitting one based on the research context. From a methodological perspective, continuing methodological development of both domains would be a valuable contribution to the field. Thus, this dissertation presented technical solutions to two enduring challenges in each SEM domain.

Chapter 2 of the dissertation introduced a new data matrix-based method, named structured factor analysis (SFA; Cho and Hwang, 2022). This technique was developed to address two major challenges in factor-based SEM: the issues of improper solution and factor score indeterminacy. Unlike conventional covariance-based approaches in this domain (Jöreskog, 1970, 1978), SFA simultaneously estimates both the parameters of the measurement model and the probability distribution of candidate factor score matrices, given the data matrix of indicators. This probability distribution can be used to infer individuals' true factor scores probabilistically (e.g., the estimation of the probability that an individual has a higher true factor score than another individual) while also quantifying the degree of factor score indeterminacy. Furthermore, since the factor variance-covariance matrix is directly derived from the candidate factor score matrix, SFA inhibits the emergence of improper solutions, such as negative factor variance estimates. Therefore, researchers can consider SFA as a viable alternative factor-based method in cases where they encounter improper solutions with other methods or when their research necessitates the probabilistic inference of individual true factor scores.

Chapters 3 and 4 of the dissertation presented two extensions of generalized structured component analysis (GSCA; Hwang and Takane, 2004)—convex generalized structured component analysis (convex GSCA; Cho and Hwang, under review at *Psychometrika*) and deep learning generalized structured component analysis (DL-GSCA; Cho and Hwang, accepted in *Structural equation modeling: A multidisciplinary Journal*). Each was designed to overcome a specific limitation of GSCA within the component-based

domain. Namely, composite indexes generated by GSCA face two main constraints: their scores are not directly interpretable in terms of the original scales of indicators, and they are restricted to linear functional forms, thereby limiting their predictive power for outcome variables. Convex GSCA offers a method for generating linear composite indexes whose scores can be interpreted on the original scales of indicators. DL-GSCA, in contrast, proposes a way to create nonlinear composite indexes that optimize their predictive power for targeted outcome variables, in a data-driven manner. Thus, when implementing component-based SEM, researchers might opt for convex GSCA if their main priority is model interpretability, and for DL-GSCA if their primary interest is model predictability.

5.2. Limitations and Future Research Directions

In sum, this dissertation delineated the two SEM domains, emphasizing their critical roles in the empirical investigation of psychological theories involving constructs, and proposed innovative SEM techniques to tackle the long-standing limitations of current methods in both SEM domains. However, this dissertation also has limitations. While individual chapters discussed the limitations of proposed methods specific to each SEM domain, this sub-section will pivot towards discussing their limitations within the overarching SEM framework.

Foremost, this dissertation stresses the critical need for accurate identification of construct types and the corresponding selection of SEM domain, but it does not offer a fully established statistical procedure for exploring the nature of constructs in the model. Misrepresenting constructs—by representing non-latent constructs as common factors or genuinely latent constructs as components—can lead to inaccurate outcomes in statistical testing and parameter estimates (Cho, Sarstedt, et al., 2022; Hwang, Cho, Jung, et al., 2021), underlining the importance for precise construct type clarification before SEM application. Additionally, a conceptual review of the constructs may not be sufficient to identify construct

types, as illustrated in studies concerning the American Customer Satisfaction Index (ACSI) model (Fornell, 1992; Fornell et al., 1996). Although the authors identified all key constructs to be latent in the ACSI model based on an in-depth conceptual review of the constructs, their measurement model turned out to be not statistically supported (Section 2.5). In contrast, the model representing the constructs as components led to acceptable goodness-of-fit values (Section 3.6).

This dissertation suggests potential empirical strategies for construct type identification. One such strategy involves comparing the goodness-of-fit values from two identical structural equation models with differing statistical representations, expecting those fit values to be within acceptable bounds if the models are true. While these fit values cannot serve as confirming evidence for a specified model, they might reject a model if the values fall outside of acceptable bounds, as shown in the ACSI model case. Another proposed strategy involves examining the variations in correlation estimates between a target construct and other variables in response to changes in their indicators. As elucidated in Chapter 1.4, alterations in the indicators of a target construct can differentially impact its correlations with other variables, contingent on its true nature.

Nonetheless, these suggestions are in their nascent stage and require rigorous empirical testing for their practical utility. Consequently, it is essential for future research to develop a comprehensive and reliable statistical procedure for discerning construct types. Such research should involve the establishment of statistical tests and a broad simulation study to gauge the effectiveness of these procedures in various research scenarios. This will not only solidify the theoretical underpinnings but also significantly enhance the practical applicability of the two SEM domains in empirical research.

Second, every method introduced in this dissertation necessitates that all constructs in the model be treated as either latent or summary/descriptive. This requirement may not suit

all practical situations, as researchers may wish to treat only some constructs in the model as latent and the others as summary or descriptive. For instance, in a study exploring the relationship between SES and adolescent aggression (Fatima & Sheikh, 2014), researchers might prefer to consider SES a summary and adolescent aggression a latent construct, thereby representing the former as a component and the latter as a common factor, respectively. However, proposed methods in either SEM domain can accommodate only one of these two statistical representations within a single structural equation model, possibly limiting their applicability in empirical research.

While the recently proposed method, integrated generalized structured component analysis, (IGSCA; Hwang, Cho, Jung, et al., 2021), attempts to address this limitation, it is restricted to estimating parameters of the basic structural equation model (e.g., recursive structure, absence of method factors, and uncorrelated errors) and lacks the ability to estimate the candidate factor score distribution for latent constructs. Considering the variety of methods available that can handle a broader range of models in each SEM domain, including those proposed in this dissertation, there is a substantial potential for future research to develop a statistical procedure that applies the appropriate SEM methods from each SEM domains to each construct based on its nature and subsequently combines the results to obtain unbiased parameter estimates for the entire model.

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

Aggarwal, C. C. (2018). An Introduction to neural networks. In C. C. Aggarwal (Ed.), *Neural Networks and Deep Learning: A Textbook* (pp. 1–52). Springer International Publishing. https://doi.org/10.1007/978-3-319-94463-0_1

Altman, A., & Gondzio, J. (1999). Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. *Optimization Methods and Software*, *11*(1–4), 275–302. https://doi.org/10.1080/10556789908805754

American Psychological Association. (2007). *Report of the APA Task Force on Socioeconomic Status*. https://www.apa.org/pi/ses/resources/publications

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.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010

Areskoug, B. (1982). The first canonical correlation: Theoretical PLS analysis and simulation experiments. In H. Wold & K. G. Jöreskog (Eds.), *Systems under indirect observation: causality, structure, prediction* (pp. 95–118). North Holland.

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.

Basco, R., Hair, J. F., Ringle, C. M., & Sarstedt, M. (2021). Advancing family business research through modeling nonlinear relationships: Comparing PLS-SEM and multiple regression. *Journal of Family Business Strategy*, 100457. https://doi.org/10.1016/j.jfbs.2021.100457

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

Binning, J. F. (2015). Construct. In *Britannica*. https://www.britannica.com/science/construct

Blanchflower, D. G., & Oswald, A. J. (2005). Happiness and the human development index: The paradox of Australia. *Australian Economic Review*, *38*(3), 307–318. https://doi.org/10.1111/j.1467-8462.2005.00377.x

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. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, *35*(2), 359–372. https://doi.org/10.2307/23044047

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., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314. https://doi.org/10.1037/0033-2909.110.2.305

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

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Boyd, S. P., & Vandenberghe, L. (2018). *Introduction to applied linear algebra: vectors, matrices, and least squares*. Cambridge university press. https://doi.org/10.1017/9781108583664

Brandt, V., Zhang, Y., Carr, H., Golm, D., Correll, C. U., Arrondo, G., Firth, J., Hassan, L., Solmi, M., & Cortese, S. (2023). First evidence of a general disease (“d”) factor, a common factor underlying physical and mental illness. *World Psychiatry*, *22*(2), 335–337. https://doi.org/10.1002/wps.21097

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. General considerations. *IMA Journal of Applied Mathematics*, *6*(1), 76–90. https://doi.org/10.1093/imamat/6.1.76

Cantril, H. (1965). *The pattern of human concerns*. Rutgers University Press.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of

variables. *Proceedings of the 76th Annual Convention of the American Psychological Association*, *3*, 227–228.

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., & Choi, J. Y. (2020). An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation models. *Behaviormetrika*, *47*(1), 243–272. https://doi.org/10.1007/s41237-019-00098-0

Cho, G., & Hwang, H. (2023). Structured factor analysis: A data matrix-based alternative approach to structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(3), 364–377. https://doi.org/10.1080/10705511.2022.2126360

Cho, G., Hwang, H., Sarstedt, M., & Ringle, C. M. (2020). Cutoff criteria for overall model fit indexes in generalized structured component analysis. *Journal of Marketing Analytics*, *8*, 189–202. https://doi.org/10.1057/s41270-020-00089-1

Cho, G., Hwang, H., Sarstedt, M., & Ringle, C. M. (2022). A prediction-oriented specification search algorithm for generalized structured component analysis. *Structural Equation Modeling: A Multidisciplinary Journal*. https://doi.org/10.1080/10705511.2022.2057315

Cho, G., Jung, K., & Hwang, H. (2019). Out-of-bag prediction error: A cross validation index for generalized structured component analysis. *Multivariate Behavioral Research*, *54*(4),

505–513. https://doi.org/10.1080/00273171.2018.1540340

Cho, G., Kim, S., Hwang, H., Lee, J., Sarstedt, M., & Ringle, C. M. (2023). A comparative study of the predictive power of component-based approaches to structural equation modeling. *European Journal of Marketing*, *57*(6), 1641–1661. https://doi.org/10.1108/EJM-07-2020-0542

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

Cho, G., Schlaegel, C., Hwang, H., Choi, Y., Sarstedt, M., & Ringle, C. M. (2022). Integrated generalized structured component analysis: On the use of model fit criteria in international management research. *Management International Review*, *62*, 569–609. https://doi.org/10.1007/s11575-022-00479-w

Choi, S., Lee, S., Huh, I., Hwang, H., & Park, T. (2020). HisCoM-G×E: Hierarchical structural component analysis of gene-based gene–environment interactions. *International Journal of Molecular Sciences*, *21*(18), 1–11. https://doi.org/10.3390/ijms21186724

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, *114*(1), 174–184. https://doi.org/10.1037/0033-2909.114.1.174

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.

Csaji, B. C. (2001). *Approximation with artificial neural networks* [Eötvös Loránd University].

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.2647&rep=rep1&type=pdf

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems, 2*(4), 303–314. https://doi.org/10.1007/BF02551274

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 Jong, S., & Kiers, H. A. L. (1992). Principal covariates regression: Part I. Theory. *Chemometrics and Intelligent Laboratory Systems, 14*(1), 155–164. https://doi.org/10.1016/0169-7439(92)80100-I

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://deleeuwpdx.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. (1989). Reduced form estimation, hedging against possible misspecification. *International Economic Review, 30*(2), 373–390. https://doi.org/10.2307/2526653

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. (2013a). *The simplest possible factor model estimator*. https://doi.org/10.13140/RG.2.1.3605.6809

Dijkstra, T. K. (2013b). *Composites as factors, generalized canonical variables revisited*. https://doi.org/10.13140/RG.2.1.3426.5449

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

Dobson, K. G., Vigod, S. N., Mustard, C., & Smith, P. M. (2021). Major depressive episodes and employment earnings trajectories over the following decade among working-aged Canadian men and women. *Journal of Affective Disorders, 285*, 37–46. https://doi.org/10.1016/j.jad.2021.02.019

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

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. https://doi.org/10.1037/1082-989X.5.2.155

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

Fatima, S., & Sheikh, H. (2014). Socioeconomic status and adolescent aggression: The role

of executive functioning as a mediator. *The American Journal of Psychology*, *127*(4), 419–430. https://doi.org/10.5406/amerjpsyc.127.4.0419

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, *13*(3), 317–322. https://doi.org/10.1093/comjnl/13.3.317

Floudas, C. A., & Visweswaran, V. (1995). Quadratic optimization. In R. Horst & P. M. Pardalos (Eds.), *Handbook of global optimization* (pp. 217–269). Springer US. https://doi.org/10.1007/978-1-4615-2025-2_5

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. (1992). A national customer satisfaction barometer: The Swedish experience. *Journal of Marketing*, *56*(1), 6–21. https://doi.org/10.2307/1252129

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), 7–18. https://doi.org/10.2307/1251898

Frank, M., & Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, *3*(1-2), 95–110. https://doi.org/10.1002/nav.3800030109

Gensowski, M., Heckman, J., & Savelyev, P. (2011). *The effects of education, personality, and IQ on earnings of high-ability men*. https://conference.iza.org/conference_files/CoNoCoSk2011/gensowski_m6556.pdf

Goldberger, A. S. (1964). *Econometric theory.* New York: John Wiley & Sons.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, *24*(109), 23–26. https://doi.org/10.1090/S0025-5718-1970-0258249-6

Goodhue, D. L., Lewis, W., & Thompson, R. (2006). PLS, small sample size, and statistical power in MIS research. *Proceedings of the 39th Annual Hawaii International*

Conference on System Sciences (HICSS'06), *8*, 202b-202b. https://doi.org/10.1109/HICSS.2006.381

Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS Have advantages for small sample size or non-normal data? *MIS Quarterly*, *36*(3), 981–1001. https://doi.org/10.2307/41703490

Graham, J. W. (2008). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

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

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

Hall, J., & Helliwell, J. (2014). *Happiness and human development*. http://akwl.org/wp-content/uploads/2019/12/happiness_and_hd.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Springer. https://doi.org/10.1007/978-0-387-21606-5

Hayduk, L. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement*, *74*(6), 905–926. https://doi.org/10.1177/0013164414527449

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

Heise, D. R. (1972). Employing nominal variables, induced variables, and block variables in path analyses. *Sociological Methods & Research*, *1*(2), 147–173. https://doi.org/10.1177/004912417200100201

Henseler, J. (2012). Why generalized structured component analysis is not universally preferable to structural equation modeling. *Journal of the Academy of Marketing Science*, *40*(3), 402–413. https://doi.org/10.1007/s11747-011-0298-6

Hoyle, R. H. (2014). *Handbook of structural equation modeling* (1st ed.). The Guilford Press.

Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika*, *74*(3), 517–530. https://doi.org/10.1007/S11336-009-9119-Y

Hwang, H., & Cho, G. (2020). Global least squares path modeling: A full-information alternative to partial least squares path modeling. *Psychometrika*, *85*, 947–972. https://doi.org/10.1007/s11336-020-09733-2

Hwang, H., Cho, G., & Choo, H. (2023). *GSCA Pro (Version 1.2.1) [Computer software]*. http://www.gscapro.com.

Hwang, H., Cho, G., Jin, M. J., Ryoo, J. H., Choi, Y., & Lee, S.-H. (2021). A knowledge-based multivariate statistical method for examining gene-brain-behavioral/cognitive relationships: Imaging genetics generalized structured component analysis. *PloS One*, *16*(3), e0247592. https://doi.org/10.1371/journal.pone.0247592

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., Desarbo, W. S., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, *72*(2), 181–198. https://doi.org/10.1007/s11336-005-1314-x

Hwang, H., Ho, M.-H. R., & Lee, J. (2010). Generalized structured component analysis with latent interactions. *Psychometrika*, *75*(2), 228–242. https://doi.org/10.1007/s11336-010-9157-5

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., Sarstedt, M., Cheah, J.-H., & Ringle, C. M. (2020). A concept analysis of methodological research on composite-based structural equation modeling: Bridging PLSPM and GSCA. *Behaviormetrika*, *47*(1), 219–241. https://doi.org/10.1007/s41237-019-00085-5

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, *69*(1), 81–99. https://doi.org/10.1007/BF02295841

Hwang, H., & Takane, Y. (2010). Nonlinear generalized structured component analysis. *Behaviormetrika*, *37*(1), 1–14. https://doi.org/10.2333/bhmk.37.1

Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling*. Chapman and Hall/CRC Press.

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

Hwang, H., Takane, Y., & Malhotra, N. (2007). Multilevel generalized structural component analysis. *Behaviormetrika*, *34*(2), 95–109. https://doi.org/10.2333/bhmk.34.95

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, *48*, 398–407. https://doi.org/10.2307/2095231

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of*

Mathematical and Statistical Psychology, *23*(2), 121–145. https://doi.org/10.1111/j.2044-8317.1970.tb00439.x

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 255–284). Seminar Press.

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.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, *58*(3), 433–451. https://doi.org/10.1093/biomet/58.3.433

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

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press. https://psycnet.apa.org/record/2015-56948-000

Kumar, S. K. (2017). *On weight initialization in deep neural networks*. https://arxiv.org/abs/1704.08863

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

Lawson, C. L., & Hanson, R. J. (1974). *Solving least squares problems*. Prentice Hall.

Lay, D. C., Lay, S. R., & McDonald, J. J. (2015). *Linear algebra and its applications* (5th ed.). Pearson Education.

Le, P. V. (2014). *More schooling is not always better: Evidence from an Instrumental*

variables approach to educational reform in Vietnam. https://fsppm.fulbright.edu.vn/cache/More Schooling Is Not Always Better-2014-12-19-08341367.pdf

Le, Q. V. (2015). A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks. *Google Brain*, *20*, 1–20. http://ai.stanford.edu/~quocle/tutorial2.pdf

Lecun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Leong, F. T. L., & Huang, J. L. (2016). Standard error of measurement. In *Britannica*. https://www.britannica.com/science/standard-error-of-measurement

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares.* Physica. https://doi.org/10.1007/978-3-642-52512-4

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

Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The expressive power of neural networks: A view from the width. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6231–6239). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/32cbf687880eb1674a07bf717761dd3a-Paper.pdf

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*(2), 95–107. https://doi.org/10.1037/h0056029

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.

Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, *29*(3), 617–663. https://doi.org/10.1080/01431160701352154

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

Naik, D. N., & Khattree, R. (1996). Revisiting olympic track records: Some practical considerations in the principal component analysis. *The American Statistician*, *50*(2), 140–144. https://doi.org/10.1080/00031305.1996.10474361

Newsom, J. T. (2014). *Improper solutions in SEM*. https://web.pdx.edu/~newsomj/semclass/ho_improper.pdf

Nimon, K., Henson, R. K., & Gates, M. S. (2010). Revisiting interpretation of canonical correlation analysis: A tutorial and demonstration of canonical commonality analysis. *Multivariate Behavioral Research*, *45*(4), 702–724.

https://doi.org/10.1080/00273171.2010.498293

Noorbakhsh, F. (1998). The human development index: Some technical issues and alternative indices. *Journal of International Development*, *10*(5), 589–605. https://doi.org/10.1002/(SICI)1099-1328(199807/08)10:5<589::AID-JID484>3.0.CO;2-S

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education : Theory and Practice*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Nübler, I. (1995). The human development index revisited. *Intereconomics*, *30*(4), 171–176. https://doi.org/10.1007/BF02928088

Nwankpa, C. E., Ijomah, W., Gachagan, A., & Marshall, S. (2021). Activation functions. *2nd International Conference on Computational Sciences and Technology*, 124–133. https://pure.strath.ac.uk/ws/portalfiles/portal/118946797/Nwankpa_etal_ICCST_2021_Activation_functions_comparison_of_trends_in_practice.pdf

Oldenburg, G. (2020). Structural Equation Modeling. *The Mathematica Journal*. https://doi.org/10.3888/tmj.22–5

Park, J. H. (1994). *Returns to schooling: A peculiar deviation from linearity* (No. 714; Working Papers, Issue 714). https://ideas.repec.org/p/pri/indrel/335.html

Perla, J., Sargent, T. J., & Stachurski, J. (2020). *Quantitative economics with Julia*. QuantEcon. https://julia.quantecon.org/

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

Ranis, G., Stewart, F., & Samman, E. (2006). Human development: Beyond the human

development index. *Journal of Human Development, 7*(3), 323–358. https://doi.org/10.1080/14649880600815917

Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, *26*(4), 332–344. https://doi.org/10.1016/j.ijresmar.2009.08.001

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30–45. https://doi.org/10.1037/met0000220

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, *45*(5–6), 341–358. https://doi.org/10.1016/j.lrp.2012.09.010

Romdhani, H., Hwang, H., Paradis, G., Roy-Gagnon, M. H., & Labbe, A. (2015). Pathway-based association study of multiple candidate genes and multiple traits using structural equation models. *Genetic Epidemiology*, *39*(2), 101–113. https://doi.org/10.1002/gepi.21872

Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods, 16*(3), 425–448. https://doi.org/10.1177/1094428112474693

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books. https://apps.dtic.mil/sti/citations/AD0256582

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. *Psychological Methods*. https://doi.org/10.1037/met0000503

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Sagar, A. D., & Najam, A. (1998). The human development index: a critical review. *Ecological Economics*, *25*(3), 249–264. https://doi.org/10.1016/S0921-8009(97)00168-7

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, *24*(111), 647–656. https://doi.org/10.2307/2004840

Silva, J., de Keulenaer, F., & Johnstone, N. (2012). *Environmental quality and life satisfaction* (OECD Environment Working Papers, Issue 44). https://doi.org/10.1787/5k9cw678dlr0-en

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

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

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

Strang, G. (2019). *Linear algebra and learning from data* (1st ed.). Wellesley-Cambridge Press.

Sullivan, G. M., & Artino, A. R. J. (2013). Analyzing and interpreting data from likert-type

scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/JGME-5-4-18

Swaminathan, H., & Algina, J. (1978). Scale freeness in factor analysis. *Psychometrika*, *43*(4), 581–583. https://doi.org/10.1007/BF02293816

Takane, Y., & Hwang, H. (2005). An extended redundancy analysis and its applications to two practical examples. *Computational Statistics and Data Analysis*, *49*(3), 785–808. https://doi.org/10.1016/j.csda.2004.06.004

Takane, Y., & Hwang, H. (2007). Regularized linear and kernel redundancy analysis. *Computational Statistics & Data Analysis*, *52*(1), 394–405. https://doi.org/10.1016/j.csda.2007.02.014

Takane, Y., Kiers, H. A. L., & de Leeuw, J. (1995). Component analysis with different sets of constraints on different dimensions. *Psychometrika*, *60*(2), 259–280. https://doi.org/10.1007/BF02301416

Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Quality Management and Business Excellence*, *19*(7–8), 871–886. https://doi.org/10.1080/14783360802159543

Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. F. (2017). Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods. *Psychometrika*, *82*(3), 737–777. https://doi.org/10.1007/s11336-017-9573-x

The MathWorks Inc. (2022). *MATLAB Optimization toolbox: User's Guide*. The MathWorks Inc. https://www.mathworks.com/help/pdf_doc/optim/optim.pdf

Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, *41*(1), 1–32. https://doi.org/10.1037/h0075959

UNDP. (1990). *Human development report 1990: Concept and measurement of human development*. http://www.hdr.undp.org/en/reports/global/hdr1990

UNDP. (2010). *Human development report 2010: The real wealth of nations*. http://hdr.undp.org/en/content/human-development-report-2010

UNDP. (2016). *Human development report 2016: Human development for everyone*. http://hdr.undp.org/en/content/human-development-report-2016

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

Urban, C. J., & Gates, K. M. (2021). Deep learning: A primer for psychologists. *Psychological Methods*. https://doi.org/10.1037/met0000374

van den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, *42*(2), 207–219. https://doi.org/10.1007/BF02294050

van der Leeden, R. (1990). *Reduced rank regression with structured residuals*. DSWO Press.

van der Maas, H. L. J., Kan, K.-J., & Borsboom, D. (2014). Intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence*, *2*(1), 12–15. https://doi.org/10.3390/jintelligence2010012

Vanderbei, R. J., & Carpenter, T. J. (1993). Symmetric indefinite systems for interior point methods. *Mathematical Programming*, *58*(1), 1–32. https://doi.org/10.1007/BF01581257

Widaman, K. F. (2018). On common factor and principal component representations of data: Implications for theory and for confirmatory replications. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 829–847. https://doi.org/10.1080/10705511.2018.1478730

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) Modelling: Some current developments. In P. R. Krishnaiah (Ed.), *Multivariate analysis–III* (pp. 383–407).

Academic Press. https://doi.org/10.1016/B978-0-12-426653-7.50032-6

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction, part II* (pp. 1–54). North Holland.

Wold, H. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences, Vol. 6* (pp. 581–591). Wiley.

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.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368. https://doi.org/10.1080/01621459.1962.10480664

Zhangyang, W., Yingzhen, Y., Shiyu, C., Qing, L., & Huang, T. S. (2016). Learning a deep l∞ encoder for hashing. In S. Kambhampati (Ed.), *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence International Joint Conference on Artificial Intelligence* (pp. 2174–2180). AAAI Press.

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology / Psychologie Canadienne*, *34*, 390–400. https://doi.org/10.1037/h0078865

Appendix A for Chapter 1

Appendix A1. An illustration of how a path coefficient in a component-based structural equation model summarizes causal effects between two indicator clusters

Figure A1.1 illustrates a hypothetical multivariate regression model. This model involves nine causal effects of the three indicators of γ_1 on the other three indicators of γ_2.

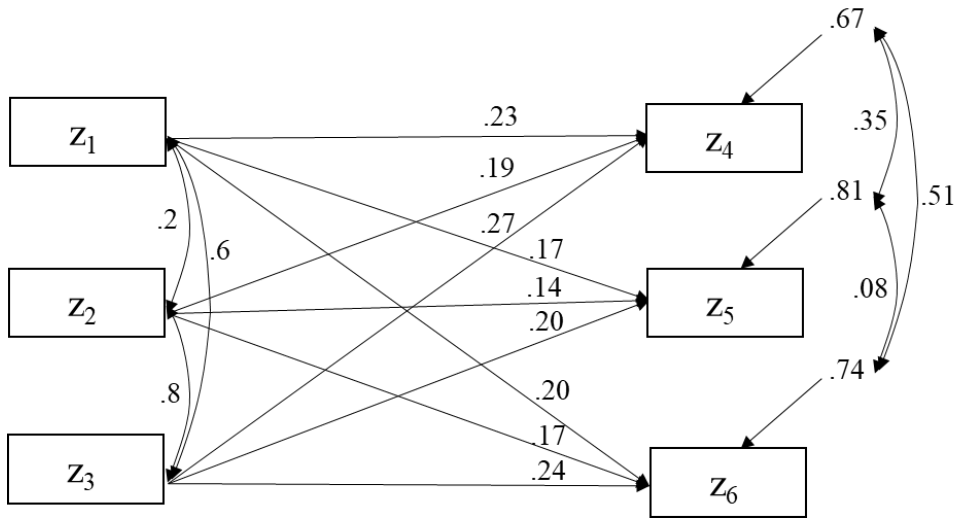


Figure A1.1. A hypothetical multivariate regression model. Squares signify indicators, single-headed arrows signify individual causal effects between variables, and double-headed arrows denote correlations.

According to Cho and Choi (2020), this regression model can be re-expressed as a component-based structural equation model with two components in Figure A1.2. This model condenses the nine causal effects between the two indicator clusters into a single path coefficient from γ_1 to γ_2. The two components fully mediate the causal effects of the indicators of γ_1 on the indicators of γ_2, so that the correlations between the two indicator clusters can be entirely explained by the two components. One can recover the nine original regression coefficient values from the component model by calculating the indirect effects of the indicators of γ_1 on the indicators of γ_2 though the path between γ_1 and γ_2 (e.g., the direct

effect of z_1 on z_4 = $w_{1,1} \times b_1 \times c_{2,4}$). With this regard, the path coefficient b_1 can be considered a summary of the causal effects between the two indicator clusters.

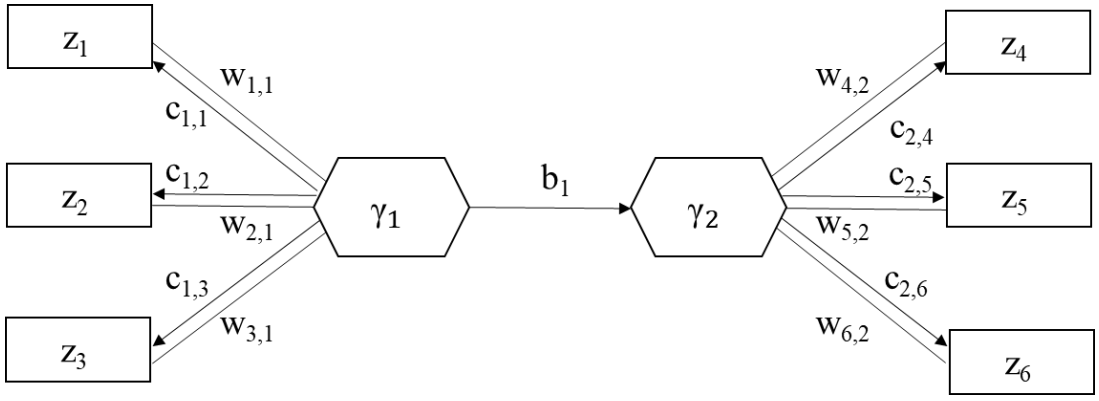


Figure A1.2. A component-based structural equation model that is equivalent to the multivariate regression model in Figure A1.1. Error terms and their correlations are omitted. The parameter values are as follows: b_1 = .6, $w_{1,1}$ = .40, $w_{2,1}$ = .32, $w_{3,1}$ = .46, $w_{4,2}$ = .44, $w_{5,2}$ = .33, $w_{6,2}$ = .39, $c_{1,1}$ = .83, $c_{1,2}$ = .68, $c_{1,3}$ = .97, $c_{2,4}$ = .96, $c_{2,5}$ = .72, and $c_{2,6}$ = .85.

Appendix B for Chapter 2

Appendix B1. 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 $\mathbb{x}_{n\cdot}$ denote a set of random variables or a random vector for each individual n (n = 1, 2, ⋯, N), whose population mean vector and covariance matrix are denoted by $\boldsymbol{\tau}_n$ and $\boldsymbol{\Xi}_n$, respectively (i.e., $\boldsymbol{\tau}_n \equiv E[\mathbb{x}_{n\cdot}]$ and $\boldsymbol{\Xi}_n \equiv E[(\mathbb{x}_{n\cdot} - E[\mathbb{x}_{n\cdot}])(\mathbb{x}_{n\cdot} - E[\mathbb{x}_{n\cdot}])']$). With $\{\mathbb{x}_{1\cdot}, \mathbb{x}_{2\cdot}, \cdots, \mathbb{x}_{N\cdot}\}$, we can define a random matrix $\mathbb{X}$ for N individuals as $\mathbb{X} \equiv [\mathbb{x}_{1\cdot}, \mathbb{x}_{2\cdot}, \cdots, \mathbb{x}_{N\cdot}]'$. 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 $mean(\mathbf{X}) \equiv (N^{-1}\mathbf{1}_N'\mathbf{X})$ and $cov(\mathbf{X}) \equiv N_0{}^{-1}(\mathbf{X} - \mathbf{1}_N mean(\mathbf{X}))'(\mathbf{X} - \mathbf{1}_N mean(\mathbf{X}))$ for any matrix $\mathbf{X}$ having N rows and $N_0 \equiv N - 1$ (see Appendix A in Bollen, 1989). If $\{\mathbb{x}_{1\cdot}, \mathbb{x}_{2\cdot}, \cdots, \mathbb{x}_{N\cdot}\}$ is *i.i.d.*, then $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_N$ and $\boldsymbol{\Xi}_1 = \boldsymbol{\Xi}_2 = \cdots = \boldsymbol{\Xi}_N$, indicating that the population mean vector and covariance matrix of $\mathbb{x}_{n\cdot}$ 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 $\mathbb{x}_{n\cdot}$, i.e., $E[mean(\mathbb{X})] = \boldsymbol{\tau}$ and $E[cov(\mathbb{X})] = \boldsymbol{\Xi}$ (Theorem 1), whose proof is provided in Appendix B2. This theorem suggests that $mean(\mathbb{X})$ and $cov(\mathbb{X})$ 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, $mean(\mathbf{X})$ can be seen as the sample mean vector of $\mathbb{x}_{n\cdot}$ as well as the score mean vector of $\mathbb{X}$, and similarly, $cov(\mathbf{X})$ can be considered the sample covariance matrix of $\mathbb{x}_{n\cdot}$ and the score covariance matrix of $\mathbb{X}$.

B1.1. 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 $\mathbb{h}_{n\cdot}$ denote a random vector of the nth individual's P latent variables, whose population mean vector and covariance matrix are denoted by $\boldsymbol{\alpha}$ and $\boldsymbol{\Phi}$, respectively, for all n ($n = 1, 2, \cdots, N$). We assume that $\{\mathbb{h}_{1\cdot}, \mathbb{h}_{2\cdot}, \cdots, \mathbb{h}_{N\cdot}\}$ is *i.i.d.* and every latent variable is standardized (i.e., $\boldsymbol{\alpha} = \mathbf{0}$ and $diag(\boldsymbol{\Phi}) = \mathbf{1}_P$), where $\mathbf{1}_k$ is a k by 1 vector of ones having N rows and $diag()$ is an operator that converts an input matrix into a column vector of its diagonal elements. If $\mathbb{h}_{n\cdot}$ is realized as a score vector, denoted by $\mathbf{h}_{n\cdot}$, it means that the nth 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., $\mathbb{h}_{1\cdot} = \mathbf{h}_{1\cdot}$, $\mathbb{h}_{2\cdot} = \mathbf{h}_{2\cdot}$, $\cdots$, $\mathbb{h}_{N\cdot} = \mathbf{h}_{N\cdot}$).

Let $\mathbf{H}_{true} \equiv [\mathbf{h}_{1\cdot}, \mathbf{h}_{2\cdot}, \cdots, \mathbf{h}_{N\cdot}]'$ 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 $\mathbb{h}_{n\cdot}$, denoted by $\boldsymbol{\alpha}_s$ and $\boldsymbol{\Phi}_s$ respectively, can be calculated by $\boldsymbol{\alpha}_s = mean(\mathbf{H}_{true})$ and $\boldsymbol{\Phi}_s = cov(\mathbf{H}_{true})$. For simplicity, the true latent variable scores in $\mathbf{H}_{true}$ are assumed to be

standardized such that $\boldsymbol{\alpha}_s = \mathbf{0}$ and $diag(\boldsymbol{\Phi}_s) = \mathbf{1}_P$. 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 $\boldsymbol{\Lambda}$, whose (p,j)th entry refers to the pth latent variable's effect on the jth indicator ($p = 1, 2, \cdots, P$; $j = 1, 2, \cdots, J$).

On the other hand, a unique factor in SFA corresponds to a random error that affects one indicator uniquely (Bollen, 1989, p. 233). Let $\mathbb{e}_{1\cdot}, \mathbb{e}_{2\cdot}, \cdots, \mathbb{e}_{N\cdot}$ denote *i.i.d.* random vectors of J unique factors, where $E[\mathbb{e}_{n\cdot}] = \mathbf{0}$ and $\boldsymbol{\Theta} \equiv E[\mathbb{e}_{n\cdot}\mathbb{e}_{n\cdot}']$ satisfying $diag(\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s\boldsymbol{\Lambda} + \boldsymbol{\Theta}) = \mathbf{1}_J$ ($n = 1, 2, \cdots, N$). Then, a random matrix of J unique factors for the N individuals, denoted by $\mathbb{E}$, can be defined as $\mathbb{E} \equiv [\mathbb{e}_{1\cdot}, \mathbb{e}_{2\cdot}, \cdots, \mathbb{e}_{N\cdot}]'$. By Theorem 1, $E[mean(\mathbb{E})] = \mathbf{0}$ and $E[cov(\mathbb{E})] = \boldsymbol{\Theta}$. 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., $E[cov([\mathbb{E}, \mathbf{H}_{true}])] = \mathbf{0}$), where $cov(\mathbf{X}_1, \mathbf{X}_2) \equiv N_0^{-1}(\mathbf{X}_1 - \mathbf{1}_N mean(\mathbf{X}_1))'(\mathbf{X}_2 - \mathbf{1}_N mean(\mathbf{X}_2))$ for any matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ having N rows. Then, the measurement model can be expressed as

$$\mathbb{Z} = \mathbf{H}_{true}\mathbf{\Lambda} + \mathbb{E}. \tag{B.1}$$

The measurement model (B.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 (B.1), $E[mean(\mathbb{Z})] = \mathbf{0}$ and $E[cov(\mathbb{Z})] = \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda} + \mathbf{\Theta}$ (Theorem 2), whose proof is provided in Appendix B3. In the paper, $E[cov(\mathbb{Z})]$ 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 (B.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 $\mathbf{L}$, $\mathbf{\Delta}_s$, and $\mathbf{F}_{true}$ given $\mathbf{Z}$. However, even if we knew the values of $\mathbf{L}$ and $\mathbf{\Delta}_s$, 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 $\mathbf{Z} = \mathbf{FL}$, $mean(\mathbf{F}) = mean(\mathbf{F}_{true})$, 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 $cov(\mathbf{F}_{true}) = \mathbf{\Delta}_s$. Under this condition, $\mathbf{Z}$ also becomes representative (i.e., $mean(\mathbf{Z}) = \mathbf{0}$ and $cov(\mathbf{Z}) = \mathbf{\Sigma}_s$; see Appendix B3), from which $mean(\mathbf{F}_{true})$ and $cov(\mathbf{F}_{true})$ can be identified as $\mathbf{0}$ and $\mathbf{\Delta}_s$, respectively.

Assuming that $\mathbf{F}_{true}$ is representative, SFA aims to obtain unbiased estimates of $\mathbf{L}$ and $\mathbf{\Delta}$ 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 $\widehat{\mathbf{Z}}$ 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.

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

subject to $mean(\mathbf{F}) = \mathbf{0}$ and $cov(\mathbf{F}) = \mathbf{\Delta}_s$, 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 $\mathbf{\Delta}_s$ are fixed to zeros. The value of the cost function (B.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 **L** and **F** are estimated, the estimate of $\mathbf{\Delta}_s$ is obtained by $\widehat{\mathbf{\Delta}}_s = cov(\widehat{\mathbf{F}})$.

There is no closed-form solution for minimizing (B.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 B4. 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 (B.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 B5. Also, the proposed ALS algorithm mathematically guarantees the convergence of (B.2) (de Leeuw et al., 1976), which means that the decrease in (B.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 (B.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 B6. 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

$$IPE_O = SS(\mathbf{Z}_{std} - \widehat{\mathbf{F}}\widehat{\mathbf{L}})/SS(\mathbf{Z}_{std}), \tag{B.3}$$

where $\widehat{\mathbf{F}}$ is the estimate of the candidate factor score matrix and $\widehat{\mathbf{L}}$ is the estimate of the loading matrix. This index is equivalent to the value of ρ given $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{L}}$. The value of 1 – IPE_O can also be interpreted as the average R^2 for the indicators. The IPE_O value is zero if and only if $\widehat{\mathbf{L}} = \mathbf{L}$ and $\widehat{\mathbf{F}}$ satisfies $mean(\widehat{\mathbf{F}}) = \mathbf{0}$, $\mathbf{Z}_{std} = \widehat{\mathbf{F}}\widehat{\mathbf{L}}$, and $cov(\widehat{\mathbf{F}}) = \mathbf{\Delta}_s$, under the condition that the measurement model is identified (see Appendix B5). As $\widehat{\mathbf{F}}$ obtained from SFA always satisfies $mean(\widehat{\mathbf{F}}) = \mathbf{0}$, the positive value of IPE_O indicates that $\widehat{\mathbf{L}} \neq \mathbf{L}$, $\mathbf{Z}_{std} \neq \widehat{\mathbf{F}}\widehat{\mathbf{L}}$, or $cov(\widehat{\mathbf{F}}) \neq \mathbf{\Delta}_s$.

SFA can conduct a statistical test of the null hypothesis that $\widehat{\mathbf{L}} = \mathbf{L}$, $\mathbf{Z}_{std} = \widehat{\mathbf{F}}\widehat{\mathbf{L}}$, and $cov(\widehat{\mathbf{F}}) = \mathbf{\Delta}_s$ by using the Bollen-Stine (B-S) bootstrap method (Bollen & Stine, 1993). Let $\widehat{\mathbf{\Sigma}}_s$ 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 $\mathbf{Z}_{imp} = \mathbf{Z}_{std}\mathbf{S}^{-1/2}\widehat{\mathbf{\Sigma}}_s^{1/2}$. 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 – α)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.

B1.2. 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 $\mathbb{H} \equiv [\mathbb{h}_{1\cdot}, \mathbb{h}_{2\cdot}, \cdots, \mathbb{h}_{N\cdot}]'$ and additionally models the stochastic process of generating the scores of $\mathbb{H}$, which is called the *structural model*. We use $\mathbb{h}_{\cdot p}$ to denote the pth column of $\mathbb{H}$, which corresponds to an N by 1 random vector of the pth latent variables for the N individuals. By Theorem 1, $\mathbb{H}$ satisfies that $E[mean(\mathbb{H})] = \mathbf{0}$ and $E[cov(\mathbb{H})] = \mathbf{\Phi}$.

For simplicity, we assume that the entries of $\mathbb{h}_{n\cdot}$ are arranged such that $\mathbb{h}_{n\cdot}$ can be expressed as $\mathbb{h}_n = [\mathbb{h}_{X,n\cdot}, \mathbb{h}_{Y,n\cdot}]$, where $\mathbb{h}_{X,n\cdot}$ and $\mathbb{h}_{Y,n\cdot}$ denote a random vector of P_x exogenous latent variables and that of P_y endogenous latent variables, respectively (n = 1, 2, ⋯, 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 $\mathbf{\Phi}_{XX} \equiv E[\mathbb{h}_{X,n\cdot}\mathbb{h}_{X,n\cdot}']$, $\mathbf{\Phi}_{YY} \equiv E[\mathbb{h}_{Y,n\cdot}\mathbb{h}_{Y,n\cdot}']$, and $\mathbf{\Phi}_{XY} \equiv E[\mathbb{h}_{X,n\cdot}\mathbb{h}_{Y,n\cdot}']$. Let $\mathbb{H}_X \equiv [\mathbb{h}_{X,1\cdot}, \mathbb{h}_{X,2\cdot}, \cdots, \mathbb{h}_{X,N\cdot}]'$ and $\mathbb{H}_Y \equiv [\mathbb{h}_{Y,1\cdot}, \mathbb{h}_{Y,2\cdot}, \cdots, \mathbb{h}_{Y,N\cdot}]'$, 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 $\{\mathbb{q}_{1\cdot}, \mathbb{q}_{2\cdot}, \cdots, \mathbb{q}_{N\cdot}\}$ denote a set of *i.i.d.* random vectors of P_y structural errors, where $\mathbb{q}_{n\cdot}$ is for the nth individual ($n = 1, 2, \cdots, N$), $E[\mathbb{q}_{n\cdot}] = \mathbf{0}$, and $\mathbf{\Psi} \equiv E[\mathbb{q}_{n\cdot}\mathbb{q}_{n\cdot}']$. Let $\mathbb{Q} \equiv [\mathbb{q}_{1\cdot}, \mathbb{q}_{2\cdot}, \cdots, \mathbb{q}_{N\cdot}]'$, whose realized value is denoted by $\mathbf{Q}_{true}$. By Theorem 1, $E[mean(\mathbb{Q})] = \mathbf{0}$ and $E[cov(\mathbb{Q})] = \mathbf{\Psi}$. 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., $E[cov(\mathbb{Q}, \mathbb{H}_X)] = \mathbf{0}$]. Let $\mathbf{B}_X$ and $\mathbf{B}_Y$ denote matrices of path coefficients that quantify the causal effects of $\mathbb{h}_{X,n\cdot}$ on $\mathbb{h}_{Y,n\cdot}$ and those between endogenous variables in $\mathbb{h}_{Y,n\cdot}$, respectively. We assume that $diag(\mathbf{B}_Y) = \mathbf{0}$ and $(\mathbf{I}_{Py} - \mathbf{B}_Y)$ is invertible. Then, the structural model can be written as

$$\mathbb{H}_Y = \mathbb{H}_X\mathbf{B}_X + \mathbb{H}_Y\mathbf{B}_Y + \mathbb{Q} \tag{B.4}$$

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

The structural model (B.4) describes how $\mathbf{H}_{true}$ are generated. For N individuals, the random matrices $\mathbb{H}_X$ and $\mathbb{Q}$ in (B.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 (B.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 $\widehat{\mathbf{H}}_X$ and $\widehat{\mathbf{H}}_Y$ 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 $\boldsymbol{\Psi}$) while treating $[\widehat{\mathbf{H}}_X, \widehat{\mathbf{H}}_Y]$ as if they were the data of latent variables, assuming that $\widehat{\boldsymbol{\Phi}}_s = \boldsymbol{\Phi}$. 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 B7 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

$$\text{IPE}_L = SS(\widehat{\mathbf{H}}_Y - (\widehat{\mathbf{H}}_X\widehat{\mathbf{B}}_X + \widehat{\mathbf{H}}_Y\widehat{\mathbf{B}}_Y))/SS(\widehat{\mathbf{H}}_Y). \quad \text{(B.5)}$$

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 estimated and implied covariance matrices of the latent variables.

Appendix B2. Theorem 1 and its proof

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

Proof. The expected value of $mean(\mathbb{X})$ can be expressed as $E[mean(\mathbb{X})] = E[(N^{-1}\mathbf{1}_N'\mathbb{X})] = N^{-1}\mathbf{1}_N'E[\mathbb{X}] = N^{-1}\mathbf{1}_N'E[[\mathbb{x}_1, \mathbb{x}_2, \cdots, \mathbb{x}_N]'] = N^{-1}\mathbf{1}_N'[E[\mathbb{x}_1], E[\mathbb{x}_2], \cdots, E[\mathbb{x}_N]]' = N^{-1}\mathbf{1}_N'\mathbf{1}_N\boldsymbol{\tau} = \boldsymbol{\tau}$, indicating that $E[mean(\mathbb{X})] = \boldsymbol{\tau}$. Also, the expected value of $cov(\mathbb{X})$ can be expressed as $E[cov(\mathbb{X})] = E[N_0^{-1}(\mathbb{X} - \mathbf{1}_N mean(\mathbb{X}))'(\mathbb{X} - \mathbf{1}_N mean(\mathbb{X}))] = N_0^{-1}(E[\mathbb{X}'\mathbb{X}] - E[\mathbb{X}'\mathbf{1}_N mean(\mathbb{X})] - E[mean(\mathbb{X})'\mathbf{1}_N'\mathbb{X}]) + E[mean(\mathbb{X})'\mathbf{1}_N'\mathbf{1}_N mean(\mathbb{X}))]) = N_0^{-1}(E[\mathbb{X}'\mathbb{X}] - N(E[mean(\mathbb{X})'mean(\mathbb{X})])) = N_0^{-1}(\sum_{i=1}^{N} E[\mathbb{x}_i\mathbb{x}_i'] - N^{-1}E[(\sum_{i=1}^{N} \mathbb{x}_i)(\sum_{i=1}^{N} \mathbb{x}_i)']) = N_0^{-1}(\sum_{i=1}^{N} E[\mathbb{x}_i\mathbb{x}_i'] - N^{-1}(\sum_{i=1}^{N} E[\mathbb{x}_i\,\mathbb{x}_i'] + \sum_{i \neq k} E[\mathbb{x}_i\mathbb{x}_k']) = N_0^{-1}(N_0E[\mathbb{x}_n\mathbb{x}_n'] - N^{-1}\sum_{i \neq k}(E[(\mathbb{x}_i - E[\mathbb{x}_i])(\mathbb{x}_i - E[\mathbb{x}_k])'] + E[\mathbb{x}_i]E[\mathbb{x}_k]') = N_0^{-1}(N_0(\boldsymbol{\Xi} + \boldsymbol{\tau}\boldsymbol{\tau}') - N_0\boldsymbol{\tau}\boldsymbol{\tau}') = N_0^{-1}N_0\boldsymbol{\Xi} = \boldsymbol{\Xi}$. Q.E.D.

Appendix B3. Theorem 2 and its proof

Theorem 2. $E[mean(\mathbb{Z})] = \mathbf{0}$ and $E[cov(\mathbb{Z})] = \mathbf{\Lambda}'\mathbf{\Phi}_s\mathbf{\Lambda} + \mathbf{\Theta}$ under (2.1).

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

Q.E.D.

Appendix B4. The proposed ALS algorithm for the first stage of SFA

Let us reparametrize $\mathbf{F} = \mathbf{U}\mathbf{V}_{\mathrm{f}}$, 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 $\mathbf{V}_{\mathrm{f}}$ is a T by T matrix satisfying $\mathbf{V}_{\mathrm{f}}'\mathbf{V}_{\mathrm{f}} = \mathbf{\Delta}_{\mathrm{s}}$. Then, (B.2) can be re-expressed as

$$\rho = (JN)^{-1} SS(\mathbf{Z}_{std} - \mathbf{U}\mathbf{V}_{\mathrm{f}}\mathbf{L}), \tag{B.6}$$

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 $\mathbf{V}_{\mathrm{f}}$ and $\mathbf{L}$. The ALS algorithm alternatingly updates each of $\mathbf{U}$, $\mathbf{V}_{\mathrm{f}}$, and $\mathbf{L}$ with the others fixed until no substantial decrease in (B.6) occurs between two consecutive iterations. One can utilize one of the existing SEM techniques to obtain the initial values of $\mathbf{L}$ and $\mathbf{\Delta}_{\mathrm{s}}$, from which $\mathbf{V}_{\mathrm{f}}$ can be obtained by the Cholesky decomposition of $\mathbf{\Delta}_{\mathrm{s}}$. Otherwise, SFA draws on the following procedure to obtain the initial values of $\mathbf{L}$ and $\mathbf{V}_{\mathrm{f}}$. 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 $\mathbf{V}_{\mathrm{f}}$ satisfying $\mathbf{V}_{\eta}'\mathbf{V}_{\eta} = \mathbf{\Phi}_{\mathrm{s}}$ and $\mathbf{V}_{\epsilon}$ denote a matrix of the last J by J diagonal block of $\mathbf{V}_{\mathrm{f}}$ satisfying $\mathbf{V}_{\epsilon}'\mathbf{V}_{\epsilon} = \mathbf{\Theta}$. The algorithm initially sets $\mathbf{V}_{\eta}$ to be $\mathbf{I}_P$ and all the non-zero entries of the loading matrix $\mathbf{\Lambda}$ 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 $\mathbf{\Lambda}$ is rescaled such that $diag(\mathbf{\Lambda}'\mathbf{\Phi}_{\mathrm{s}}\mathbf{\Lambda}) = N_0^{-1} diag(\mathbf{Z}_{\eta}'\mathbf{Z}_{\eta})$. Lastly, the initial values of $\mathbf{V}_{\mathrm{f}}$ and $\mathbf{L}$ are obtained by $\mathbf{V}_{\mathrm{f}} = \begin{bmatrix} \mathbf{V}_{\eta} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\epsilon} \end{bmatrix}$ and $\mathbf{L} = [\mathbf{\Lambda}; \mathbf{I}_J]$, 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

$$N_0^{1/2}\mathbf{K}_z\mathbf{\Omega}_z\mathbf{O}_z', \tag{B.7}$$

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 $\mathbf{S}$, (B.6) can be re-expressed as

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

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 $rank(\mathbf{A}) = rank(\mathbf{\Omega}_z\mathbf{O}_z'\mathbf{L}'\mathbf{V}_f') = rank(\mathbf{L}') = rank([\mathbf{\Lambda}', \mathbf{I}_J]) = rank(\mathbf{I}_J) = J$. Matrix $\mathbf{A}$ can be decomposed into

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

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

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

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 (B.8) can be obtained by

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

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 $tr(\mathbf{U}'\mathbf{K}_z\mathbf{A})$ is equivalent to $tr(\boldsymbol{\chi})$ (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, (B.8) can be re-written as

$$\begin{aligned}\rho &= N^{-1}N_0 + c_0(SS(\mathbf{V}_f\mathbf{L}) - 2tr(\mathbf{O}_z\boldsymbol{\Omega}_z\mathbf{R}_A\mathbf{Q}'(\mathbf{V}_f\mathbf{L})))\\ &= c_2 + c_0(SS(\mathbf{Q}\mathbf{R}_A'\boldsymbol{\Omega}_z\mathbf{O}_z') - 2tr(\mathbf{O}_z\boldsymbol{\Omega}_z\mathbf{R}_A\mathbf{Q}'(\mathbf{V}_f\mathbf{L})) + SS(\mathbf{V}_f\mathbf{L})\\ &= c_2 + c_0SS(\mathbf{Q}\mathbf{R}_A'\boldsymbol{\Omega}_z\mathbf{O}_z' - \mathbf{V}_f\mathbf{L})\\ &= c_2 + c_0SS(vec(\mathbf{Q}\mathbf{R}_A'\boldsymbol{\Omega}_z\mathbf{O}_z') - (\mathbf{L}' \otimes \mathbf{I}_T)vec(\mathbf{V}_f)),\end{aligned} \tag{B.12}$$

where $c_2 \equiv N^{-1}N_0 + c_0SS(\mathbf{Q}\mathbf{R}_A'\boldsymbol{\Omega}_z\mathbf{O}_z')$. Let $\boldsymbol{\iota}_2$ denote a vector of non-zero elements in $vec(\mathbf{V}_f)$ and $\boldsymbol{\Gamma}_1$ denote a matrix of the columns of $(\mathbf{L}' \otimes \mathbf{I})$ corresponding to $\boldsymbol{\iota}_2$. Then, the least-squares estimate of $\boldsymbol{\iota}_2$ is obtained by

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

The non-zero elements of $\mathbf{V}_f$ are updated by $\boldsymbol{\iota}_2$ 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., $10^{\wedge -10}$), 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, (B.6) can be re-expressed as

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

where $c_3 \equiv c_2 + c_0 SS(\mathbf{Q}_\epsilon \mathbf{R}_A' \boldsymbol{\Omega}_z \mathbf{O}_z' - \mathbf{V}_\epsilon)$. Let $\boldsymbol{\iota}_3$ denote a vector of non-zero elements in $vec(\boldsymbol{\Lambda})$ and $\boldsymbol{\Gamma}_2$ is a matrix of the columns of $(\mathbf{I}_J \otimes \mathbf{V}_\eta)$ corresponding to the non-zero elements in $vec(\boldsymbol{\Lambda})$. Then, the least-squares estimate of $\boldsymbol{\iota}_3$ is obtained by

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

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

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 $\boldsymbol{\Gamma}_3$ denote a P by P diagonal matrix whose diagonal entries are the elements of $diag(\mathbf{V}_\eta' \mathbf{V}_\eta)^{\circ 1/2}$. Then, $\mathbf{V}_\eta$ is rescaled by post-multiplying $\mathbf{V}_\eta$ by $\boldsymbol{\Gamma}_3^{-1}$ such that the rescaled value of $\mathbf{V}_\eta$ can satisfy $diag(\mathbf{V}_\eta' \mathbf{V}_\eta) = \mathbf{1}_p$. Then, the least-squares estimate of $\boldsymbol{\Phi}_s$ is updated by $\widehat{\boldsymbol{\Phi}}_s = \mathbf{V}_\eta' \mathbf{V}_\eta$. 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, $\widehat{\boldsymbol{\Lambda}}$ is also adjusted by pre-multiplying it by $\boldsymbol{\Gamma}_3$. Also, let $\boldsymbol{\Gamma}_4$ denote a J by J diagonal matrix whose diagonal entries are the elements of $diag(\widehat{\boldsymbol{\Lambda}} \widehat{\boldsymbol{\Phi}}_s \widehat{\boldsymbol{\Lambda}} + \mathbf{V}_\epsilon' \mathbf{V}_\epsilon)^{\circ 1/2}$. Then, $\mathbf{V}_\epsilon$ is updated one more time by post-multiplying it by $\boldsymbol{\Gamma}_4^{-1}$ such that the updated value of $\mathbf{V}_\epsilon$ can satisfy $diag(\widehat{\boldsymbol{\Lambda}} \widehat{\boldsymbol{\Phi}}_s \widehat{\boldsymbol{\Lambda}} + \mathbf{V}_\epsilon' \mathbf{V}_\epsilon) = \mathbf{1}_J$. This makes the variances of $\widehat{\mathbf{Z}}$ equal to one (i.e., $diag(cov(\widehat{\mathbf{Z}})) = \mathbf{1}_J$). Then, the least-squares estimate of $\boldsymbol{\Theta}$ is obtained by $\widehat{\boldsymbol{\Theta}} = \mathbf{V}_\epsilon' \mathbf{V}_\epsilon$. 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 $\widehat{\mathbf{F}} = \mathbf{U}\mathbf{V}_f$.

Appendix B5. 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 $cov(\mathbf{F}_{true}) = \mathbf{\Delta}_s$. If the ALS algorithm minimize (B.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 $cov(\mathbf{F}_{true}) = \mathbf{\Delta}_s$. Then, $\mathbf{Z}$ satisfies that $mean(\mathbf{Z}) = \mathbf{0}$, $cov(\mathbf{Z}) = \mathbf{\Sigma}_s$, 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 $cov(\mathbf{Z}) = N_0^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{L}'(N_0^{-1}\mathbf{F}_{true}'\mathbf{F}_{true})\mathbf{L} = \mathbf{L}'\mathbf{\Delta}_s\mathbf{L} = \mathbf{\Sigma}_s$. Also, given the identified measurement model, $\widehat{\mathbf{L}} = \mathbf{L}$ and $\widehat{\mathbf{\Delta}}_s = \mathbf{\Delta}_s$ if and only if $\widehat{\mathbf{\Sigma}}_s = \mathbf{\Sigma}_s$, where $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{\Delta}}_s$ are estimates of $\mathbf{L}$ and $\mathbf{\Delta}$, respectively, and $\widehat{\mathbf{\Sigma}}_s$ denotes the model-implied covariance matrix of indicators. We need to show that if and only if $\widehat{\mathbf{L}} = \mathbf{L}$ and $\widehat{\mathbf{\Delta}}_s = \mathbf{\Delta}_s$, (B.2) can be zero for some $\mathbf{F}$ subject to $\mathbf{1}_N'\mathbf{F} = \mathbf{0}$ and $\widehat{\mathbf{\Delta}}_s = N_0^{-1}\mathbf{F}'\mathbf{F}$.

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

$$\rho = c_0((tr(\mathbf{\Sigma}_s) + tr(\mathbf{L}'\mathbf{\Delta}_s\mathbf{L}) - 2tr(\mathbf{\chi})). \quad \text{(B.16)}$$

As the trace of $\mathbf{\Sigma}_s$ is equivalent to the sum of its eigenvalues, (B.16) can be written as

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

Q.E.D.

Appendix B6. A supplementary procedure for the ALS algorithm

As the cost function (B.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

$$T^{1/2}\mathbf{K}_a\mathbf{\Omega}_z\mathbf{O}_z'. \tag{B.18}$$

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 $\widehat{\mathbf{\Delta}}_s$ and applying Step 1 of the ALS algorithm given $\mathbf{V}_f$ and $\widehat{\mathbf{L}}$. 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 (B.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 (B.2) is unaffected by a change in $\mathbf{K}_z$ in (B.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 (B.8). The minimum point of (B.8) is equivalent to $c_1 - 2c_0tr(\mathbf{\chi})$ (de Leeuw, 2017), where $c_0 \equiv (JN)^{-1}N_0$, $c_1 = c_0tr(\mathbf{S} + \mathbf{L}'\mathbf{\Delta}_s\mathbf{L})$, and $\mathbf{\chi}$ 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 (B.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 **U** is obtained by (B.11), where **C** in (B.11) must satisfy $\mathbf{C}'\mathbf{C} = \mathbf{I}_P$. As **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 B7. A non-iterative estimation for the second stage of SFA

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

$$\mathbb{H}_{\mathrm{Y}} = \mathbb{H}_{\mathrm{X}}\mathbf{\Pi} + \mathbb{Q}(\mathbf{I}_{P_y} - \mathbf{B}_{\mathrm{Y}})^{-1}, \tag{B.19}$$

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

SFA's estimation algorithm starts with estimating $\mathbf{\Pi}$ given $\widehat{\mathbf{H}}_{\mathrm{X}}$ and $\widehat{\mathbf{H}}_{\mathrm{Y}}$ obtained from Stage 1. The least-squares estimate of $\mathbf{\Pi}$ is obtained by

$$\widehat{\mathbf{\Pi}} = (\widehat{\mathbf{H}}_{\mathrm{X}}'\widehat{\mathbf{H}}_{\mathrm{X}})^{-1}(\widehat{\mathbf{H}}_{\mathrm{X}}'\widehat{\mathbf{H}}_{\mathrm{XY}}). \tag{B.20}$$

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 $\widehat{\mathbf{\Pi}}_0 \equiv [\mathbf{I}_{P\mathrm{x}}, \widehat{\mathbf{\Pi}}]$ and $\widehat{\mathbf{B}} \equiv [\widehat{\mathbf{B}}_{\mathrm{X}}; \widehat{\mathbf{B}}_{\mathrm{Y}}]$. Let $\hat{\boldsymbol{\pi}}_q$ and $\hat{\mathbf{b}}_q$ denote the qth column vectors of $\widehat{\mathbf{\Pi}}$ and $\widehat{\mathbf{B}}$, respectively ($q = 1, 2, \cdots, P_{\mathrm{y}}$). Let $\boldsymbol{\iota}_3$ denote a vector of non-zero elements in $\hat{\mathbf{b}}_q$. Let $\mathbf{\Gamma}_6$ denote a matrix of the columns of $\widehat{\mathbf{\Pi}}_0$ corresponding to the non-zero elements in $\hat{\mathbf{b}}_q$. Then, the least-squares estimate of $\boldsymbol{\iota}_3$ is obtained by

$$\boldsymbol{\iota}_3 = \left(\boldsymbol{\Gamma}_6'\boldsymbol{\Gamma}_6\right)^{-1}\boldsymbol{\Gamma}_6'\hat{\boldsymbol{\pi}}_q. \tag{B.21}$$

Then, the non-zero elements in the qth columns of $\widehat{\mathbf{B}}_{\mathrm{X}}$ and $\widehat{\mathbf{B}}_{\mathrm{Y}}$ are updated by $\boldsymbol{\iota}_3$. After all the path coefficients are estimated, the least-squares estimate of $\boldsymbol{\Psi}$ is obtained by $\widehat{\boldsymbol{\Psi}} = [-\widehat{\mathbf{B}}_{\mathrm{X}}; \mathbf{I}_{Py} - \widehat{\mathbf{B}}_{\mathrm{Y}}]'\widehat{\boldsymbol{\Phi}}_{\mathrm{s}}[-\widehat{\mathbf{B}}_{\mathrm{X}}; \mathbf{I}_{Py} - \widehat{\mathbf{B}}_{\mathrm{Y}}]$.

Appendix B8. A full description of the candidate factor score distribution

As stated in Section 2.2, SFA obtains an estimate of a matrix of the candidate factor scores $\mathbf{F}$ and uses this estimate, denoted by $\widehat{\mathbf{F}}$, to estimate the parameters of the measurement and structural models. However, SFA does not recommend using $\widehat{\mathbf{F}}$ 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 $\boldsymbol{\mathcal{F}}_{N,T}$, 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 $cov(\mathbf{F}_{true}) = \boldsymbol{\Delta}_s$). Under this assumption, $\boldsymbol{\mathcal{F}}_{N,T}$ can be expressed as $\boldsymbol{\mathcal{F}}_{N,T} \equiv \{\mathbf{F} \in \mathbb{R}^{N \times T} | \ \mathbf{Z} = \mathbf{FL}, mean(\mathbf{F}) = \mathbf{0}, \text{ and } cov(\mathbf{F}) = \boldsymbol{\Delta}_s\}$, suggesting that if a factor score matrix does not have the sample mean vector of $\mathbf{0}$ or the sample covariance matrix of $\boldsymbol{\Delta}_s$, 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 $\boldsymbol{\mathcal{F}}_{N,T}$ 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 $\boldsymbol{\mathcal{F}}_{N,T}$. As each element in $\boldsymbol{\mathcal{F}}_{N,T}$ 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 $\mathbb{f}_{n\cdot}$, correspond to a random vector of each individual's candidate factors, whose population covariance matrix is denoted by $\mathbf{G}_n$ ($n = 1, 2, \cdots, N$).

The derivation of the candidate factor score distribution starts from the following theorem: $\mathbf{F}_{true}$ in $\boldsymbol{\mathcal{F}}_{N,T}$ can be re-expressed as $\mathbf{F}_{true} = \mathbf{ZW} + \mathbf{M}_{true}$, where $\mathbf{W} \equiv \boldsymbol{\Sigma}_s^{-1}[\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s, \boldsymbol{\Theta}]$, $\mathbf{G} \equiv [\mathbf{I}_P, -\boldsymbol{\Lambda}]'(\boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s)[\mathbf{I}_P, -\boldsymbol{\Lambda}]$, and $\mathbf{M}_{true}$ is an N by T matrix satisfying $mean(\mathbf{M}_{true}) = \mathbf{0}$, $cov(\mathbf{M}_{true}) = \mathbf{G}$, and $cov(\mathbf{Z}, \mathbf{M}_{true}) = \mathbf{0}$ (Theorem 4; Guttman, 1955), whose proof is provided in Appendix B9. The major implication of this theorem is that $\mathbf{F}_{true}$ in $\boldsymbol{\mathcal{F}}_{N,T}$

can be decomposed into two parts: the deterministic part (i.e., $\mathbf{ZW}$) 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., $cov(\mathbf{Z}, \mathbf{M}_{true}) = \mathbf{0}$). As $\mathbf{F}_{true}$ can be seen as an arbitrary value in $\boldsymbol{\mathcal{F}}_{N,T}$ given $\mathbf{Z}$, the theorem also implies that any $\mathbf{F}$ in $\boldsymbol{\mathcal{F}}_{N,T}$ can be re-expressed as $\mathbf{F} = \mathbf{ZW} + \mathbf{M}$, where $\mathbf{M}$ is an N by T matrix in $\boldsymbol{\mathcal{M}}_{N,T} \equiv \{\mathbf{M} \in \mathbb{R}^{N \times T} | \; mean(\mathbf{M}) = \mathbf{0}, cov(\mathbf{M}) = \mathbf{G}, \text{ and } cov(\mathbf{Z}, \mathbf{M}) = \mathbf{0}\}$. The value of $\mathbf{F}$ depends solely on $\mathbf{M}$ given $\mathbf{ZW}$, 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 $\boldsymbol{\mathcal{M}}_{N,T}$ 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 $\boldsymbol{\mathcal{M}}_{N,T}$ 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 $\boldsymbol{\mathcal{M}}_{N,T}$, which means that every $\mathbf{M}$ in $\boldsymbol{\mathcal{M}}_{N,T}$ is assumed to be equally likely to be true. With the random matrix $\mathbb{M}$, we can re-express $\mathbb{F}$ as

$$\mathbb{F} = \mathbf{ZW} + \mathbb{M}. \tag{B.22}$$

The equation (B.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 $\boldsymbol{\mathcal{F}}_{N,T}$; (b) $E[cov(\mathbf{F}_{true} - \mathbb{F})] = 2\mathbf{G}$; (c) $E[\mathbb{F}] = \mathbf{ZW}$; (d) $E[cov(\mathbf{F}_{true} - E[\mathbb{F}])] = \mathbf{G}$; (e) $N_0^{-1}\sum_{i=1}^{N} \mathbf{G}_i = \mathbf{G}$ (Theorem 5). The proof of these properties is provided in Appendix B10.

Property (a) means that every matrix of candidate factor scores in $\boldsymbol{\mathcal{F}}_{N,T}$ 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 $\boldsymbol{\mathcal{F}}_{N,T}$. Property (b) indicates that if one randomly chooses a value from $\boldsymbol{\mathcal{F}}_{N,T}$ 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 $E[\mathbb{F}]$ can be expressed as a matrix of weighted composite scores of indicators **ZW**. Accordingly, **ZW** is called *a matrix of expected candidate factor scores* in SFA. Property (d) means that if one uses **ZW** 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 $\boldsymbol{\mathcal{F}}_{N,T}$ as a measurement of $\mathbf{F}_{true}$. It suggests that it is more reasonable to use **ZW** as a measurement of $\mathbf{F}_{true}$ than any random value in $\boldsymbol{\mathcal{F}}_{N,T}$. Lastly, Property (e) indicates that each $\mathbb{f}_{n\cdot}$ has the covariance matrix that approximates **G** with $E[\mathbb{f}_{n\cdot}] = (\mathbf{z}_{n\cdot}'\mathbf{W})'$, where $\mathbf{z}_{n\cdot}$ is a vector of the nth individual's indicator scores in **Z**.

Note that researchers need to be cautious not to interpret the equation (B.22) as a data-generating process for $\mathbf{F}_{true}$. Unlike (B.1), the candidate factor score distribution is built on **Z** after **Z** has been generated as $\mathbf{Z} = \mathbf{F}_{true}\mathbf{L}$, for inferring $\mathbf{F}_{true}$ *a posteriori* given **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 (B.2) in a least-squares sense. Specifically, let $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$ denote an estimate of $\boldsymbol{\mathcal{F}}_{N,T}$, whose elements can be expressed with $\widehat{\mathbf{L}}$ and $\widehat{\boldsymbol{\Delta}}_s$ obtained from the first stage. Then, SFA randomly samples a prescribed number of $\widehat{\mathbf{F}}$ from $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$ and uses the set of the $\widehat{\mathbf{F}}$ values as an estimate of the candidate factor score distribution. A detailed explanation of this procedure is provided in Appendix B11. Under the assumption that $\widehat{\mathbf{L}} = \mathbf{L}$, $\widehat{\boldsymbol{\Delta}}_s = \boldsymbol{\Delta}_s$, and $\mathbf{F}_{true}$ is representative given **Z**, the estimate of the candidate factor score distribution with $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{G}}$ approximates the uniform distribution on $\boldsymbol{\mathcal{F}}_{N,T}$ with **W** and **G** (Theorem 6), whose proof is provided in Appendix B12.

Once it estimates the candidate factor score distribution, SFA estimates **ZW** and $diag(\mathbf{G})^{\circ 1/2}$ 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 **ZW** is the only part of $\mathbf{F}_{true}$ that can be inferred from **Z**, as shown in Theorem 4, and can be seen as the best linear predictor for $\mathbf{F}_{true}$ given **Z** (e.g., Bartholomew, 1981), as **W** in **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 $diag(\widehat{\mathbf{G}})^{\circ 1/2}$ obtained from SFA approximates the actual standard errors of measurement (i.e., $diag(cov(\mathbf{F}_{true} - \mathbf{Z}_{std}\widehat{\mathbf{W}}))^{\circ 1/2}$) when $\mathbf{Z}_{std}\widehat{\mathbf{W}}$ 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 B9. Theorem 4 and its proof

Theorem 4. $\mathbf{F}_{true}$ in $\boldsymbol{\mathcal{F}}_{N,T}$ can be re-expressed as $\mathbf{F}_{true} = \mathbf{ZW} + \mathbf{M}_{true}$, where $\mathbf{W} \equiv \boldsymbol{\Sigma}_s^{-1}[\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s, \boldsymbol{\Theta}]$, $\mathbf{G} \equiv [\mathbf{I}_P, -\boldsymbol{\Lambda}]'(\boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s)[\mathbf{I}_P, -\boldsymbol{\Lambda}]$, 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 $\mathbf{W} = \boldsymbol{\Sigma}_s^{-1}[\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s, \boldsymbol{\Theta}]$. Let us define $\mathbf{M}_{true}$ as $\mathbf{M}_{true} = \mathbf{F}_{true} - \mathbf{ZW}$. 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) $cov(\mathbf{M}_{true}) = [\mathbf{I}_P, -\boldsymbol{\Lambda}]'(\boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s)[\mathbf{I}_P, -\boldsymbol{\Lambda}]$. For (a), $mean(\mathbf{M}_{true}) = N^{-1}\mathbf{1}_N'\mathbf{M}_{true} = N^{-1}\mathbf{1}_N'(\mathbf{F}_{true} - \mathbf{ZW}) = N^{-1}\mathbf{1}_N'\mathbf{F}_{true} - N^{-1}(\mathbf{1}_N'\mathbf{Z})\mathbf{W} = \mathbf{0}$. For (b), $cov(\mathbf{Z}, \mathbf{M}_{true}) = N_0^{-1}\mathbf{Z}'\mathbf{M}_{true} = N_0^{-1}\mathbf{Z}'(\mathbf{F}_{true} - \mathbf{ZW}) = \mathbf{L}'\boldsymbol{\Delta}_s - [\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s, \boldsymbol{\Theta}] = \mathbf{0}$. 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 $\mathbf{F}_{true}\mathbf{L} = (\mathbf{ZW} + \mathbf{M}_{true})\mathbf{L} = \mathbf{Z}\boldsymbol{\Sigma}_s^{-1}[\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s\boldsymbol{\Lambda} + \boldsymbol{\Theta}] + [\mathbf{M}_1\boldsymbol{\Lambda} + \mathbf{M}_2] = \mathbf{Z} + [\mathbf{M}_1\boldsymbol{\Lambda} + \mathbf{M}_2] = \mathbf{Z}$, indicating that $\mathbf{M}_2 = -\mathbf{M}_1\boldsymbol{\Lambda}$, $\mathbf{G}_2 = \boldsymbol{\Lambda}'\mathbf{G}_1\boldsymbol{\Lambda}$, and $\mathbf{G}_{12} = -\mathbf{G}_1\boldsymbol{\Lambda}$. Let $\mathbf{H}_{true}$ denote a matrix of the first P columns of $\mathbf{F}_{true}$ and $\mathbf{W}_1 \equiv \boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s$. Then, $\mathbf{G}_1$ can be expressed as $\mathbf{G}_1 = N_0^{-1}\mathbf{M}_1'\mathbf{M}_1 = (\mathbf{H}_{true} - \mathbf{ZW}_1)'(\mathbf{H}_{true} - \mathbf{ZW}_1) = \boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\mathbf{W}_1 - \mathbf{W}_1'\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s + \mathbf{W}_1'\boldsymbol{\Sigma}_s\mathbf{W}_1 = \boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s^{-1}{}'\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s + \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_s\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s = \boldsymbol{\Phi}_s - \boldsymbol{\Phi}_s\boldsymbol{\Lambda}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_s$. As $\mathbf{G}_2 = \boldsymbol{\Lambda}'\mathbf{G}_1\boldsymbol{\Lambda}$ and $\mathbf{G}_{12} = -\mathbf{G}_1\boldsymbol{\Lambda}$, $N_0^{-1}\mathbf{M}_{true}'\mathbf{M}_{true} = [\mathbf{I}_P, -\boldsymbol{\Lambda}]'\mathbf{G}_1[\mathbf{I}_P, -\boldsymbol{\Lambda}] = \mathbf{G}$. Q.E.D.

Appendix B10. 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 $\boldsymbol{\mathcal{F}}_{N,T}$; (b) $E[cov(\mathbf{F}_{true} - \mathbb{F})] = 2\mathbf{G}$; (c) $E[\mathbb{F}] = \mathbf{ZW}$; (d) $E[cov(\mathbf{F}_{true} - E[\mathbb{F}])] = \mathbf{G}$; (e) $N_0^{-1}\sum_{i=1}^{N} \mathbf{G}_i = \mathbf{G}$.

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

Appendix B11. The algorithm for estimating the candidate factor score distribution with W and G

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

However, there exists an infinite number of $\mathbf{C}$ in (B.11) that minimizes (2.3) subject to $\mathbf{C}'\mathbf{C} = \mathbf{I}_P$ given $\widehat{\mathbf{W}}$, $\mathbf{R}_\perp$, $\mathbf{Q}_\perp$, and $\mathbf{V}_f$, so one can obtain an infinite number of $\widehat{\mathbf{F}}$s that result in the same value of ρ given $\widehat{\mathbf{\Delta}}_s$ and $\widehat{\mathbf{L}}$. Let $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$ denote a set of all possible values of $\widehat{\mathbf{F}}$ by changing the value of $\mathbf{C}$ in (B.11) given $\widehat{\mathbf{W}}$, $\mathbf{R}_\perp$, $\mathbf{Q}_\perp$, and $\mathbf{V}_f$. Let $\boldsymbol{\mathcal{C}}_{N-J-1,P}$ denote the set of all possible values of $\mathbf{C}$ with P orthonormal columns (i.e., $\boldsymbol{\mathcal{C}}_{N-J-1,P} \equiv \{\mathbf{C} \in \mathbb{R}^{(N-J-1)\times P} | \ \mathbf{C}'\mathbf{C} = \mathbf{I}_P\}$). Let $\mathbb{C}$ denote a random matrix that takes on $\mathbf{C}$ in $\boldsymbol{\mathcal{C}}_{N-J-1,P}$. SFA presumes that every $\mathbf{C}$ in $\boldsymbol{\mathcal{C}}_{N-J-1,P}$ is equally likely to be true, implying that $\mathbb{C}$ follows the uniform distribution on $\boldsymbol{\mathcal{C}}_{N-J-1,P}$ (For the mathematical definition of the uniform distribution of a random matrix with orthogonal columns, see Eaton, 1989, Chapter 2). Then, a random matrix $\widehat{\mathbb{F}}$ on the sample space $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$ can be defined as

$$\widehat{\mathbb{F}} = \mathbf{Z}_{std}\widehat{\mathbf{W}} + N^{1/2}\mathbf{R}_\perp\mathbb{C}\mathbf{Q}_\perp'\mathbf{V}_f. \tag{B.23}$$

SFA seeks to numerically obtain the probability distribution of $\widehat{\mathbb{F}}$ 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,

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

As $\mathbb{O}(\mathbb{O}'\mathbb{O})^{-1/2}$ is known to follow the uniform distribution on $\boldsymbol{\mathcal{C}}_{N-J-1,P}$ (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 (B.23), SFA obtains a value of $\widehat{\mathbb{F}}$ (i.e., $\widehat{\mathbf{F}}$). SFA repeats this resampling procedure multiple times (e.g., 1000) to obtain a set of $\widehat{\mathbf{F}}$ values that approximate to $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$.

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

Appendix B12. Theorem 6 and its proof

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

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

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

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

$$\mathbb{F} = \mathbf{Z}\mathbf{W} + \mathbb{D}[\mathbf{V}_{m1}, -\mathbf{V}_{m1}\boldsymbol{\Lambda}], \tag{B.25}$$

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 $\boldsymbol{\mathcal{D}}_{N,P}$ onto $\boldsymbol{\mathcal{F}}_{N,T}$.To prove this, we need to show the following three statements: (e) for any $\mathbf{D}$ in $\boldsymbol{\mathcal{D}}_{N,P}$, $(\mathbf{Z}\mathbf{W} + \mathbf{D}\mathbf{V}_m)$ is in $\boldsymbol{\mathcal{F}}_{N,T}$, (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 $\boldsymbol{\mathcal{D}}_{N,P}$, then, $\mathbf{D}_* = \mathbf{D}_{**}$, and (g) for any $\mathbf{F}$ in $\boldsymbol{\mathcal{F}}_{N,T}$, there exists $\mathbf{D}$ in $\boldsymbol{\mathcal{D}}_{N,P}$ such that $\mathbf{F} = (\mathbf{Z}\mathbf{W} + \mathbf{D}_*\mathbf{V}_m)$.

For (e), let $\mathbf{D}$ in $\boldsymbol{\mathcal{D}}_{N,P}$ be given. Then, $(\mathbf{ZW} + \mathbf{DV}_{\mathrm{m}})\mathbf{L} = \mathbf{Z}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}(\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda} + \boldsymbol{\Theta}) = \mathbf{Z}$, $mean(\mathbf{ZW} + \mathbf{DV}_{\mathrm{m}}) = N^{-1}\mathbf{1}_N'(\mathbf{ZW} + \mathbf{DV}_{\mathrm{m}}) = (N^{-1}\mathbf{1}_N'\mathbf{Z})\mathbf{W} + (N^{-1}\mathbf{1}_N'\mathbf{D})\mathbf{V}_{\mathrm{m}} = \mathbf{0}$, and $cov(\mathbf{ZW} + \mathbf{DV}_{\mathrm{m}}) = N^{-1}(\mathbf{ZW} + \mathbf{DV}_{\mathrm{m}})'(\mathbf{ZW} + \mathbf{DV}_{\mathrm{m}}) = [\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}, \boldsymbol{\Theta}]'\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}[\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}, \boldsymbol{\Theta}] + [\mathbf{V}_{\mathrm{m1}}, \mathbf{V}_{\mathrm{m2}}]'[\mathbf{V}_{\mathrm{m1}}, \mathbf{V}_{\mathrm{m2}}] = \begin{bmatrix} \boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}} + \mathbf{G}_1 & \boldsymbol{\Theta}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Theta} - \mathbf{G}_1\boldsymbol{\Lambda} \\ (\boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Theta} - \mathbf{G}_1\boldsymbol{\Lambda})' & \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Theta} + \boldsymbol{\Lambda}'\mathbf{G}_1\boldsymbol{\Lambda} \end{bmatrix} = \boldsymbol{\Delta}_{\mathrm{s}}$, as $\boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Theta} + \boldsymbol{\Lambda}'(\boldsymbol{\Phi}_{\mathrm{s}} - \boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}})\boldsymbol{\Lambda} = (\boldsymbol{\Sigma}_{\mathrm{s}} - \boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda})\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}(\boldsymbol{\Sigma}_{\mathrm{s}} - \boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}) = \boldsymbol{\Sigma}_{\mathrm{s}} - \boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda} = \boldsymbol{\Theta}$ and $\boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}(\boldsymbol{\Sigma}_{\mathrm{s}} - \boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}) - (\boldsymbol{\Phi}_{\mathrm{s}} - \boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}})\boldsymbol{\Lambda} = \boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda} - \boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda} = \mathbf{0}$. For (f), suppose that $(\mathbf{ZW} + \mathbf{D}_*\mathbf{V}_{\mathrm{m}}) = (\mathbf{ZW} + \mathbf{D}_{**}\mathbf{V}_{\mathrm{m}})$, where $\mathbf{D}_*$ and $\mathbf{D}_{**}$ are in $\boldsymbol{\mathcal{D}}_{N,P}$. Then, $(\mathbf{ZW} + \mathbf{D}_*\mathbf{V}_{\mathrm{m}}) - (\mathbf{ZW} + \mathbf{D}_{**}\mathbf{V}_{\mathrm{m}}) = (\mathbf{D}_* - \mathbf{D}_{**})[\mathbf{V}_{\mathrm{m1}}, \mathbf{V}_{\mathrm{m2}}] = \mathbf{0}$, thereby making $(\mathbf{D}_* - \mathbf{D}_{**})\mathbf{V}_{\mathrm{m1}} = \mathbf{0}$. As $\mathbf{G}_1$ is invertible (Guttman, 1955, theorem 4), $\mathbf{V}_{\mathrm{m1}}$ is also invertible, which implies that $\mathbf{D}_* = \mathbf{D}_{**}$.

For (g), let $\mathbf{F} = [\mathbf{H}, \mathbf{E}]$ be given in $\boldsymbol{\mathcal{F}}_{N,T}$, 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 $\boldsymbol{\mathcal{D}}_{N,P}$ that makes $\mathbf{ZW}+ \mathbf{DV}_{\mathrm{m}} = \mathbf{F}$. Let us assume that $\mathbf{D} = (\mathbf{H} - \mathbf{Z}_{std}\mathbf{W}_1)\mathbf{V}_{\mathrm{m1}}^{-1}$, where $\mathbf{W}_1 = \boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}$. It satisfies all of its constraints such that $mean(\mathbf{D}) = N^{-1}\mathbf{1}_N'\mathbf{D} = N^{-1}\mathbf{1}_N'(\mathbf{H} - \mathbf{ZW}_1)\mathbf{V}_{\mathrm{m1}}^{-1} = ((N^{-1}\mathbf{1}_N'\mathbf{H}) - (N^{-1}\mathbf{1}_N'\mathbf{Z})\mathbf{W}_1)\mathbf{V}_{\mathrm{m1}}^{-1} = \mathbf{0}$, $cov(\mathbf{D}) = N_0^{-1}\mathbf{D}'\mathbf{D} = N_0^{-1}(\mathbf{V}_{\mathrm{m1}}^{-1})'(\mathbf{H} - \mathbf{ZW}_1)'(\mathbf{H} - \mathbf{ZW}_1)\mathbf{V}_{\mathrm{m1}}^{-1} = (\mathbf{V}_{\mathrm{m1}}^{-1})'(\boldsymbol{\Phi}_{\mathrm{s}} - \boldsymbol{\Phi}_{\mathrm{s}}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}})\mathbf{V}_{\mathrm{m1}}^{-1} = (\mathbf{V}_{\mathrm{m1}}\mathbf{V}_{\mathrm{m1}}^{-1})'(\mathbf{V}_{\mathrm{m1}}\mathbf{V}_{\mathrm{m1}}^{-1}) = \mathbf{I}_P$, and $cov(\mathbf{Z}, \mathbf{D}) = N_0^{-1}\mathbf{Z}'(\mathbf{H} - \mathbf{Z}_{std}\mathbf{W}_1)\mathbf{V}_{\mathrm{m1}}^{-1} = (\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}} - \boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}})\mathbf{V}_{\mathrm{m1}}^{-1} = \mathbf{0}$. Also, $\mathbf{ZW} + \mathbf{DV}_{\mathrm{m}} = \mathbf{Z}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}[\boldsymbol{\Lambda}'\boldsymbol{\Phi}_{\mathrm{s}}, \boldsymbol{\Theta}] + (\mathbf{H} - \mathbf{ZW}_1)[\mathbf{I}_P, -\boldsymbol{\Lambda}] = [\mathbf{ZW}_1 + (\mathbf{H} - \mathbf{ZW}_1), \mathbf{Z}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}\boldsymbol{\Theta} - (\mathbf{H}\boldsymbol{\Lambda} - \mathbf{Z}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}(\boldsymbol{\Sigma}_{\mathrm{s}} - \boldsymbol{\Theta}))] = [\mathbf{H}, \mathbf{E}] = \mathbf{F}$. 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 $\boldsymbol{\mathcal{F}}_{N,T}$, $\mathbb{D}$ also follows the uniform distribution on $\boldsymbol{\mathcal{D}}_{N,P}$, 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 $\boldsymbol{\mathcal{C}}_{N-J-1,\,P}$ onto $\boldsymbol{\mathcal{D}}_{N,P}$. To prove this, we need to show the following three statements: (h) for any $\mathbf{C}$ in $\boldsymbol{\mathcal{C}}_{N-J-1,\,P}$, $N_0^{1/2}[\mathbf{R}_\perp\mathbf{C}]$ is in $\boldsymbol{\mathcal{D}}_{N,P}$, (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 $\boldsymbol{\mathcal{C}}_{N-J-1}$, then $\mathbf{C}_* = \mathbf{C}_{**}$ and (j) for any $\mathbf{D}$ in $\boldsymbol{\mathcal{D}}_{N,P}$, there exists $\mathbf{C}$ in $\boldsymbol{\mathcal{C}}_{N-J-1,\,P}$ such that $\mathbf{D} = \mathbf{R}_\perp\mathbf{C}$. For (h), let $\mathbf{C}$ in $\boldsymbol{\mathcal{C}}_{N-J-1,\,P}$ be given. Then, $\mathbf{C}$ satisfies $\mathbf{1}_N'(N_0^{1/2}\mathbf{R}_\perp\mathbf{C}) = N_0^{1/2}(\mathbf{1}_N'\mathbf{R}_\perp)\mathbf{C} = \mathbf{0}$, $N^{-1}(N_0^{1/2}\mathbf{R}_\perp\mathbf{C})'(N^{1/2}\mathbf{R}_\perp\mathbf{C}) = \mathbf{C}'(\mathbf{R}_\perp'\mathbf{R}_\perp)\mathbf{C} = \mathbf{I}_{T-r}$, $\mathbf{Z}'(N_0^{1/2}\mathbf{R}_\perp\mathbf{C}) = (\mathbf{K}_z\boldsymbol{\Omega}_z\mathbf{O}_z')'(N_0^{1/2}\mathbf{R}_\perp\mathbf{C}) = N_0^{1/2}\mathbf{O}_z\boldsymbol{\Omega}_z(\mathbf{K}_z'\mathbf{R}_\perp)\mathbf{C} = \mathbf{0}$. 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 $\boldsymbol{\mathcal{C}}_{N-r-1,\,T-r}$. 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 $\boldsymbol{\mathcal{D}}_{N,P}$ be given. Then, let us define $\mathbf{C}_+$ as $\mathbf{C}_+ \equiv N_0^{-1/2}\mathbf{R}_\perp'\mathbf{D}$. Let $\boldsymbol{\Gamma}_8$ 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 $[\boldsymbol{\Gamma}_8, \mathbf{R}_\perp][\boldsymbol{\Gamma}_8, \mathbf{R}_\perp]' = \mathbf{I}_N$, $N_0^{-1}\mathbf{D}'\mathbf{R}_\perp\mathbf{R}_\perp'\mathbf{D} = N_0^{-1}\mathbf{D}'(\mathbf{I}_N - \boldsymbol{\Gamma}_8\boldsymbol{\Gamma}_8')\mathbf{D} = N_0^{-1}\mathbf{D}'\mathbf{D} - N_0^{-1}(\boldsymbol{\Gamma}_8'\mathbf{D})'(\boldsymbol{\Gamma}_8'\mathbf{D}) = \mathbf{I}_P$. Also, $\mathbf{R}_\perp\mathbf{C}_+ = \mathbf{R}_\perp(\mathbf{R}_\perp'\mathbf{D}) = (\mathbf{I}_N - \boldsymbol{\Gamma}_8\boldsymbol{\Gamma}_8')\mathbf{D} = \mathbf{D}$.

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

$$\widehat{\mathbb{F}} = \mathbf{ZW} + \mathbb{D}\mathbf{Q}_\perp'\mathbf{V}_\mathrm{f}. \tag{B.26}$$

The function of $\mathbb{D}$ for $\widehat{\mathbb{F}}$ is a one-to-one mapping of $\boldsymbol{\mathcal{D}}_{N,P}$ onto $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$. To prove this, we only need to show that if $\widehat{\mathbf{F}}_* = \widehat{\mathbf{F}}_{**}$ for any $\widehat{\mathbf{F}}_*$ and $\widehat{\mathbf{F}}_{**}$ in $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$, then $\mathbf{D}_* = \mathbf{D}_{**}$ for any $\mathbf{D}_*$ and $\mathbf{D}_{**}$ in $\boldsymbol{\mathcal{D}}_{N,T}$. Let $\mathbf{D}_*$ and $\mathbf{D}_{**}$ in $\boldsymbol{\mathcal{D}}_{N,T}$ be given, having $\widehat{\mathbf{F}}_* = \mathbf{ZW} + \mathbf{D}_*\mathbf{Q}_\perp'\mathbf{V}_\mathrm{f}$ and $\widehat{\mathbf{F}}_{**} = \mathbf{ZW} + \mathbf{D}_{**}\mathbf{Q}_\perp'\mathbf{V}_\mathrm{f}$. Suppose that $\widehat{\mathbf{F}}_* = \widehat{\mathbf{F}}_{**}$. Then, $\widehat{\mathbf{F}}_* - \widehat{\mathbf{F}}_{**} = (\mathbf{ZW} + \mathbf{D}_*\mathbf{Q}_\perp'\mathbf{V}_\mathrm{f}) - (\mathbf{ZW} + \mathbf{D}_{**}\mathbf{Q}_\perp'\mathbf{V}_\mathrm{f}) = (\mathbf{D}_* - \mathbf{D}_{**})\mathbf{Q}_\perp'\mathbf{V}_\mathrm{f} = \mathbf{0}$. As $(\mathbf{D}_* - \mathbf{D}_{**})\mathbf{Q}_\perp'\mathbf{V}_\mathrm{f}\mathbf{V}_\mathrm{f}^{-1}\mathbf{Q}_\perp = \mathbf{0}\mathbf{V}_\mathrm{f}^{-1}\mathbf{Q}_\perp$, $\mathbf{D}_* = \mathbf{D}_{**}$. From the fact that the function of $\mathbb{C}$ for $\mathbb{D}$ is a one-to-one mapping of $\boldsymbol{\mathcal{C}}_{N-J-1,\,P}$ onto $\boldsymbol{\mathcal{D}}_{N,P}$ and the function of $\mathbb{D}$ for $\widehat{\mathbb{F}}$ (B.26) is a one-to-one mapping of $\boldsymbol{\mathcal{D}}_{N,P}$ onto $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$, it follows that (c) the function of $\mathbb{C}$ for $\widehat{\mathbb{F}}$ (B.23) is a one-to-one mapping of $\boldsymbol{\mathcal{C}}_{N-J-1,\,P}$ onto $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$. As $\mathbb{C}$ follows the uniform distribution on $\boldsymbol{\mathcal{C}}_{N-J-1}$, $\widehat{\mathbb{F}}$ follows the uniform distribution on $\widehat{\boldsymbol{\mathcal{F}}}_{N,T}$.

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

Appendix C for Chapter 3

Appendix C1. A proof of disproportional penalty imposition on indicators during the minimization of the objective function (3.5)

Let $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\Delta}_Z \mathbf{z}_{std}$ is a random vector of original indicators, where $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Delta}_Z$ denote a column mean vector of $\mathbf{z}$, the covariance matrix of $\mathbf{z}$, and a diagonal matrix consisting of each indicator's standard deviation, respectively. Let $\boldsymbol{\gamma}_{uni} \equiv \mathbf{W}_{uni}'\mathbf{z}$ denote a random vector of unstandardized components, where $\mathbf{W}_{uni}$ is a matrix of unstandardized weight parameters and $vecdiag(\mathbf{W}_{uni}'\boldsymbol{\Sigma}\mathbf{W}_{uni}) = \mathbf{1}_P$. Let $\mathbf{e}_{uni}$ is a random vector of prediction errors for $[\mathbf{z}; \boldsymbol{\gamma}_{uni}]$. Let $\boldsymbol{\Delta} \equiv blkdiag(\boldsymbol{\Delta}_Z, \mathbf{I}_P)$ is a diagonal matrix of penalty parameters for $\mathbf{e}_{uni}$, where $blkdiag()$ is an operator to convert input matrices into a block-diagonal matrix. Here, the penalty parameters refer to the parameters that rescale prediction error for each dependent variable in the model. Let $\mathbf{A}_{uni}$ denote a matrix of unstandardized loading and path coefficients in $GSCA_{std}$. Let $\mathbf{a}_{0,uni}$ denote a column vector of the unstandardized intercepts in $GSCA_{std}$. When $\mathbf{W}_{uni} = \boldsymbol{\Delta}_Z^{-1}\mathbf{W}_{std}$, $\mathbf{A}_{uni} = \mathbf{A}_{std}\boldsymbol{\Delta}$, and $\mathbf{a}_{0,uni} = ([\mathbf{I}_J, \mathbf{W}_{uni}] - \mathbf{W}_{uni}\mathbf{A}_{uni})'\boldsymbol{\mu}$, (3.5) is equivalent to the following objective function,

$$f_{uni}(\mathbf{W}_{uni}, \mathbf{A}_{uni}, \mathbf{a}_{0,uni}) = tr(\boldsymbol{\Delta}^{-1}E(\mathbf{e}_{uni}\mathbf{e}_{uni}')\boldsymbol{\Delta}^{-1}) \quad \text{(C.1)}$$

subject to $vecdiag(\mathbf{W}_{uni}'\boldsymbol{\Sigma}\mathbf{W}_{uni}) = \mathbf{1}_P$, which can be proved as follows.

$$
\begin{aligned}
& f_{std}(\mathbf{W}_{std}, \mathbf{A}_{std}) \\
& = tr(E(\mathbf{e}_{std}\mathbf{e}_{std}')) \\
& = E(SS(\mathbf{z}_{std}'\mathbf{V}_{std} - \mathbf{z}_{std}'\mathbf{W}_{std}\mathbf{A}_{std}))) \\
& = E(SS((\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Delta}_z^{-1}(\mathbf{V}_{std} - \mathbf{W}_{std}\mathbf{A}_{std}))) \\
& = E(SS((\mathbf{z} - \boldsymbol{\mu})'([\boldsymbol{\Delta}_z^{-1}, \boldsymbol{\Delta}_z^{-1}\mathbf{W}_{std}] - \boldsymbol{\Delta}_z^{-1}\mathbf{W}_{std}\mathbf{A}_{std}\boldsymbol{\Delta}\boldsymbol{\Delta}^{-1}))) \\
& = E(SS((\mathbf{z} - \boldsymbol{\mu})'([\mathbf{I}_J, \mathbf{W}_{uni}]\boldsymbol{\Delta}^{-1} - \mathbf{W}_{uni}\mathbf{A}_{uni}\boldsymbol{\Delta}^{-1}))) \\
& = E(SS((\mathbf{z} - \boldsymbol{\mu})'(\mathbf{V}_{uni} - \mathbf{W}_{uni}\mathbf{A}_{uni})\boldsymbol{\Delta}^{-1})) \\
& = E(SS((\mathbf{z}'\mathbf{V}_{uni} - (\mathbf{z}'\mathbf{W}_{uni}\mathbf{A}_{uni} + \mathbf{a}_{0,uni}'))\boldsymbol{\Delta}^{-1})), \\
& = tr(\boldsymbol{\Delta}^{-1}E(\mathbf{e}_{uni}\mathbf{e}_{uni}')\boldsymbol{\Delta}^{-1}) \\
& = f_{std}(\mathbf{W}_{uni}, \mathbf{A}_{uni}, \mathbf{a}_{0,uni}),
\end{aligned}
\tag{C.2}
$$

where $\mathbf{V}_{uni} \equiv [\mathbf{I}_J, \mathbf{W}_{uni}]$. The equivalence between (3.5) and (C.1) indicates that GSCA$_{std}$'s parameters are actually the standardized versions of $\mathbf{W}_{uni}$ and $\mathbf{A}_{uni}$ that are obtained by minimizing the sum of penalized error variances for the original indicators and unstandardized components. While minimizing (C.1), a relatively large penalty will be imposed on an indicator with a relatively large variance, potentially inflating the influence of an indicator with a small variance on GSCA$_{std}$'s parameter estimation.

Appendix C2. Proofs of the six propositions that characterize a convex component

Let us suppose that the pth component (γ_p) is a convex component defined with J_p indicators ($\mathbf{z}_p$), indicating that the sum of weights assigned to the indicators is equal to one (i.e., $\mathbf{1}_{Jp}'\mathbf{w}_p = 1$) and all the weights are non-negative (i.e., $\mathbf{w}_p \geq \mathbf{0}_{Jp \times 1}$). Let $z_{i,p}$ denote the ith random variable in $\mathbf{z}_p$ ($i = 1, 2, \cdots, J_p$), which takes a value in $\mathcal{z}_{i,p} \subset \mathbb{R}$. Let $w_{i,p}$ denote the ith element of $\mathbf{w}_p$ ($i = 1, 2, \cdots, J_p$).

Proposition 1. *A convex component has scores within the range of its indicators' scores.*

Proof. Let $m_1 \equiv inf\{inf\, \mathcal{z}_{1,p}, inf\, \mathcal{z}_{2,p}, \cdots, inf\, \mathcal{z}_{Jp,p}\}$ and $m_2 \equiv sup\{sup\, \mathcal{z}_{1,p}, sup\, \mathcal{z}_{2,p}, \cdots, sup\, \mathcal{z}_{Jp,p}\}$. Then, $m_1 = m_1\sum_{i=1}^{J_p} w_{i,p} = \sum_{i=1}^{J_p} m_1 w_{i,p} \leq \gamma_p = \sum_{i=1}^{J_p} z_{i,p} w_{i,p} \leq \sum_{i=1}^{J_p} m_2 w_{i,p} = m_2\sum_{i=1}^{J_p} w_{i,p} = m_2$.

Proposition 2. *Each score of a convex component corresponds to a component score of an individual whose scores for indicators are all the same as the component score.*

Proof. Let $g \in \mathcal{G}_p$ denote a value of γ_p, where $\mathcal{G}_p \subset \mathbb{R}$ is the set of all possible values γ_p can take in $\mathbb{R}$. If $\mathbf{z}_p = [g, g, \cdots, g]' = g\mathbf{1}_{Jp}$, then $\gamma_p = \mathbf{w}_p' g\mathbf{1}_{Jp} = g$.

Proposition 3. *The mean of a convex component is not fixed to zero but is determined by weights within the range of its indicators' means.*

Proof. $E(\gamma_p) = \mathbf{w}_p' E(\mathbf{z}_p) = \mathbf{w}_p'\boldsymbol{\mu}_p$. Thus, $E(\gamma_p)$ varies depending on $\mathbf{w}_p$ unless $\boldsymbol{\mu}_p = \mathbf{0}$. Let $\mu_{i,p}$ denote the ith element of $\mathbf{u}_p$. Let $m_3 \equiv inf\{\mu_{1,p}, \mu_{2,p}, \cdots, \mu_{Jp,p}\}$ and $m_4 \equiv sup\{\mu_{1,p}, \mu_{2,p}, \cdots, \mu_{Jp,p}\}$.

Then, $m_3 = m_3\sum_{i=1}^{J_p} w_{i,p} = \sum_{i=1}^{J_p} m_3 w_{i,p} \leq E(\gamma_p) = \mathbf{w}_p'\boldsymbol{\mu}_p = \sum_{i=1}^{J_p} \mu_{i,p} w_{i,p} \leq \sum_{i=1}^{J_p} m_4 w_{i,p} = m_4\sum_{i=1}^{J_p} w_{i,p} = m_4$.

Proposition 4. *The standard deviation of a convex component is not fixed to one but is determined by weights within the range from 0 to the maximum standard deviation of its indicators.*

Proof. $\text{var}(\boldsymbol{\gamma}_p)^{1/2} = (\mathbf{w}_p'\text{var}(\mathbf{z}_p)\mathbf{w}_p)^{1/2} = (\mathbf{w}_p'\boldsymbol{\Sigma}_p\mathbf{w}_p)^{1/2}$, indicating that the standard deviation of $\boldsymbol{\gamma}_p$ depends on $\mathbf{w}_p$. Let $\sigma_{k,l,p}$ denote the (k,l)th element of $\boldsymbol{\Sigma}_p$. Let $m_5 \equiv sup\{\sigma_{1,1p}, \sigma_{2,2,p}, \cdots, \sigma_{Jp,Jp,p}\}$.

Then, $\text{var}(\boldsymbol{\gamma}_p)^{1/2} = (\mathbf{w}_p'\boldsymbol{\Sigma}_p\mathbf{w}_p)^{1/2} = (\sum_{k=1}^{J_p}\sum_{l=1}^{J_p} w_{k,p}w_{l,p}\sigma_{k,l,p})^{1/2} \leq (\sum_{k=1}^{J_p}\sum_{l=1}^{J_p} w_{k,p}w_{l,p}m_5)^{1/2} =$

$(m_5\sum_{k=1}^{J_p}\sum_{l=1}^{J_p} w_{k,p}w_{l,p})^{1/2} = (m_5\sum_{k=1}^{J_p} w_{k,p}(\sum_{l=1}^{J_p} w_{l,p}))^{1/2} = (m_5\sum_{k=1}^{J_p} w_{k,p})^{1/2} = m_5^{1/2}$. Therefore, $0 <$

$var(\boldsymbol{\gamma}_p)^{1/2} \leq m_5^{1/2}$.

Proposition 5. *Given a linearly independent set of indicators' scores, a set of convex component scores has a unique set of weights that are nonnegative and summed up to one.*

Proof. Let $\mathbf{D}_p = [\mathbf{d}_{\cdot 1,p}, \mathbf{d}_{\cdot 2,p}, \cdots, \mathbf{d}_{\cdot Jp,p}]$ denote a N by J_p data matrix of $\mathbf{z}_p$, where N is the total number of individuals and $\mathbf{d}_{\cdot i,p}$ is the score set of $z_{i,p}$ ($i = 1, 2, \cdots, J_p$). Then, the score set of the pth convex component for N individuals, denoted by $\mathbf{g}_{\cdot p}$, can be expressed as $\mathbf{g}_{\cdot p} = \mathbf{D}_p\mathbf{w}_p$. Suppose that there exists a different set of weights, $\mathbf{w}_{p+} = [w_{1,p+}, w_{2,p+}, \cdots, w_{Jp,p+}]'$, such that $\mathbf{g}_{\cdot p} = \mathbf{D}_p\mathbf{w}_{p+}$ and $\mathbf{w}_{p+} \neq \mathbf{w}_p$. Then, $0 = \mathbf{g}_{\cdot p} - \mathbf{g}_{\cdot p} = \mathbf{D}_p\mathbf{w}_p - \mathbf{D}_p\mathbf{w}_{p+} = \mathbf{D}_p(\mathbf{w}_p - \mathbf{w}_{p+}) = \mathbf{d}_{\cdot 1,p}(w_{1,p} - w_{1,p+}) + \mathbf{d}_{\cdot 2,p}(w_{2,p} - w_{2,p+}) + \cdots + \mathbf{d}_{\cdot Jp,p}(w_{Jp,p} - w_{Jp,p+})$. By the assumption that $\{\mathbf{d}_{\cdot 1,p}, \mathbf{d}_{\cdot 2,p}, \cdots, \mathbf{d}_{\cdot Jp,p}\}$ is linearly independent, $w_{1,p} = w_{1,p+}, w_{2,p} = w_{2,p+}, \cdots, w_{Jp,p} = w_{Jp,p+}$, which contradicts the assumption. By the definition of a convex component, $\mathbf{w}_p'\mathbf{1}_{Jp} = 1$ and $\mathbf{w}_p \geq \mathbf{0}_{Jp\times 1}$.

Proposition 6. *The path coefficient of a convex component on an outcome variable indicates the expected amount of change in the outcome variable for a unit change in each indicator of the convex component while holding other variables fixed.*

Proof. Let $\gamma_q = b_{0,q} + b_{p,q}\gamma_p + \boldsymbol{\alpha}'\mathbf{x} + \zeta_q$ denote a structural model equation of the outcome variable γ_q on γ_p and a vector of covariates $\mathbf{x}$ for γ_q, where $b_{0,q}$ is an intercept for γ_q, $b_{p,q}$ is the path coefficient from γ_p to γ_q. $\boldsymbol{\alpha}$ is a vector of path coefficients of $\mathbf{x}$, and ζ_q is an error term for γ_q. As this model equation can be re-expressed as $\gamma_q = b_{0,q} + b_{p,q}\,\mathbf{w}_p'\mathbf{z}_p + \boldsymbol{\alpha}'\mathbf{x} + \zeta_q$, an expected change of γ_q for a one-unit change in every element of $\mathbf{z}_p$ with the values of $\mathbf{x}$ fixed can be expressed as $E((b_{0,q} + b_{p,q}\,\mathbf{w}_p'(\mathbf{z}_p + \mathbf{1}_p) + \boldsymbol{\alpha}'\mathbf{x} + \zeta_q) - (b_{0,q} + b_{p,q}\,\mathbf{w}_p'\mathbf{z}_p + \boldsymbol{\alpha}'\mathbf{x} + \zeta_q))$, which is equivalent to $E(b_{p,q}\mathbf{w}_p'(\mathbf{z}_p + \mathbf{1}_p) - b_{p,q}\mathbf{w}_p'\mathbf{z}_p) = E(b_{p,q}(\mathbf{w}_p'\mathbf{1}_p)) = E(b_{p,q}) = b_{p,q}$.

Appendix C3. A proof that the optimization function of convex GSCA is partially scale-invariant

Suppose that for each block of indicators that are on the same scale, the measurement scales are linearly transformed arbitrarily. Let $\mathbf{z}_{new} = \mathbf{\Omega}_z(\mathbf{z} + \mathbf{\lambda})$ denote a vector of rescaled indicators, where $\mathbf{\lambda}$ is a J by 1 constant vector for relocation and $\mathbf{\Omega}_z$ is a diagonal matrix for scalar multiplication for each indicator. This linear transformation of the measurement scales of indicators does not change the minimum value of the objective function (3.13) and the corresponding weight values, which can be proven as follows. Let $\mathbf{\gamma}_{new} = \mathbf{W}'\mathbf{z}_{new}$ denote a vector of components defined with rescaled indicators. Let $\mathbf{e}_{new}$ denote a vector of prediction errors for $[\mathbf{z}_{new}; \mathbf{\gamma}_{new}]$. Let $\mathbf{A}_{new}$ denote a matrix of unstandardized loading and path coefficients for $\mathbf{z}_{new}$. Let $\mathbf{a}_{0,new}$ denote a vector of unstandardized intercepts for $\mathbf{z}_{new}$. Let $\mathbf{O}_{new}$ denote a diagonal matrix of penalty parameters for prediction errors given $\mathbf{z}_{new}$. Let ω_p is the scalar multiplier that is applied the pth block of indicators. Let $\mathbf{\Omega}_\gamma \equiv blkdiag(\omega_1, \omega_2, \cdots, \omega_P)$ and $\mathbf{\Omega} \equiv blkdiag(\mathbf{\Omega}_z, \mathbf{\Omega}_\gamma)$.

When $\mathbf{A}_{new} = \mathbf{\Omega}_\gamma^{-1}\mathbf{A}\mathbf{\Omega}$, $\mathbf{a}_{0,new} = (\mathbf{\lambda}'(\mathbf{V} - \mathbf{W}\mathbf{A}) + \mathbf{a}_0)\mathbf{\Omega}$, and $\mathbf{O}_{new} \equiv \mathbf{\Omega}^{-1}\mathbf{O}$, the objective function (3.13) can be re-written as

$$
\begin{aligned}
& f_{cvx}(\mathbf{W}, \mathbf{A}, \mathbf{a}_0) \\
&= tr(\mathbf{O}E(\mathbf{e}\mathbf{e}')\mathbf{O}) \\
&= E(SS((\mathbf{z}'\mathbf{V} - (\mathbf{z}'\mathbf{W}\mathbf{A} + \mathbf{a}_0'))\mathbf{O}) \\
&= E(SS((\mathbf{z}'(\mathbf{V} - \mathbf{W}\mathbf{A}) - \mathbf{a}_0')\mathbf{O}) \\
&= E(SS((\mathbf{z}' + \mathbf{\lambda}' - \mathbf{\lambda}')\mathbf{\Omega}_z\mathbf{\Omega}_z^{-1}(\mathbf{V} - \mathbf{W}\mathbf{A}) - \mathbf{a}_0')\mathbf{\Omega}\mathbf{\Omega}^{-1}\mathbf{O}) \\
&= E(SS((\mathbf{z} + \mathbf{\lambda})'\mathbf{\Omega}_z(\mathbf{\Omega}_z^{-1}\mathbf{V} - \mathbf{\Omega}_z^{-1}\mathbf{W}\mathbf{A})\mathbf{\Omega} - (\mathbf{\lambda}'(\mathbf{V} - \mathbf{W}\mathbf{A}) + \mathbf{a}_0')\mathbf{\Omega})\mathbf{\Omega}^{-1}\mathbf{O}) \\
&= E(SS(\mathbf{z}_{new}'([\mathbf{\Omega}_z^{-1}, \mathbf{\Omega}_z^{-1}\mathbf{W}]\mathbf{\Omega} - \mathbf{\Omega}_z^{-1}\mathbf{W}\mathbf{A}\mathbf{\Omega}) - \mathbf{a}_{0,new}'))\mathbf{O}_{new}) \\
&= E(SS(\mathbf{z}_{new}'([\mathbf{I}_J, \mathbf{W}]\mathbf{\Omega}^{-1}\mathbf{\Omega} - \mathbf{W}\mathbf{\Omega}_\gamma^{-1}\mathbf{A}\mathbf{\Omega}) - \mathbf{a}_{0,new}'))\mathbf{O}_{new}) \\
&= E(SS(\mathbf{z}_{new}'(\mathbf{V} - \mathbf{W}\mathbf{A}_{new}) - \mathbf{a}_{0,new}'))\mathbf{O}_{new}) \\
&= E(SS(\mathbf{z}_{new}'\mathbf{V} - (\mathbf{z}_{new}'\mathbf{W}\mathbf{A}_{new} + \mathbf{a}_{0,new}'))\mathbf{O}_{new}) \\
&= tr(\mathbf{O}_{new}E(\mathbf{e}_{new}\mathbf{e}_{new}')\mathbf{O}_{new}) \\
&= f_{cvx}(\mathbf{W}, \mathbf{A}_{new}, \mathbf{a}_{0,new}).
\end{aligned}
\tag{C.3}
$$

The seventh equality in (C.3) holds because $\mathbf{\Omega}_z^{-1}\mathbf{W} = \mathbf{W}\mathbf{\Omega}_\gamma^{-1}$.

Appendix C4. A description of GSCA$_{cvx}$'s ALS algorithm

The objective function (3.13) can be re-written as

$$\begin{aligned} & f_{cvx}(\mathbf{W}, \mathbf{A}, \mathbf{a}_0) \\ & = E(SS((\mathbf{z}'(\mathbf{V} - \mathbf{WA}) - \mathbf{a}_0')\mathbf{O})) \\ & = E(\mathrm{SS}((\mathbf{z} - \boldsymbol{\mu})'\mathbf{LO} - (\mathbf{a}_0' - \boldsymbol{\mu}'\mathbf{L})\mathbf{O})), \\ & = E(tr(((\mathbf{z} - \boldsymbol{\mu})'\mathbf{LO} - (\mathbf{a}_0' - \boldsymbol{\mu}'\mathbf{L})\mathbf{O})'((\mathbf{z} - \boldsymbol{\mu})'\mathbf{LO} - (\mathbf{a}_0' - \boldsymbol{\mu}'\mathbf{L})\mathbf{O}))), \\ & = tr(\mathbf{OL}'\boldsymbol{\Sigma}\mathbf{LO}) - 2tr(\mathbf{O}(\mathbf{a}_0' - \boldsymbol{\mu}'\mathbf{L})'E(\mathbf{z} - \boldsymbol{\mu})'\mathbf{LO}) + SS((\mathbf{a}_0' - \boldsymbol{\mu}'\mathbf{L})\mathbf{O})) \\ & = tr(\mathbf{OL}'\boldsymbol{\Sigma}\mathbf{LO}) + SS((\mathbf{a}_0' - \boldsymbol{\mu}'\mathbf{L})\mathbf{O})), \end{aligned} \tag{C.4}$$

where $\mathbf{L} \equiv \mathbf{V} - \mathbf{WA}$. Let $\mathbf{S}$ denote the positive definite sample covariance matrix of indicators and $\widehat{\mathbf{O}}$ denote the sample analogy of $\mathbf{O}$. As $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, and $\mathbf{O}$ are typically not available, GSCA$_{cvx}$ replaces $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, and $\mathbf{O}$ in (C.4) with $\mathbf{S}$, $\hat{\boldsymbol{\mu}}$, and $\widehat{\mathbf{O}}$, respectively, as follows,

$$\begin{aligned} & f^*_{cvx}(\mathbf{W}, \mathbf{A}, \mathbf{a}_0) \\ & = tr(\widehat{\mathbf{O}}\mathbf{L}'\mathbf{SL}\widehat{\mathbf{O}}) + SS((\mathbf{a}_0' - \hat{\boldsymbol{\mu}}'\mathbf{L})\widehat{\mathbf{O}}), \end{aligned} \tag{C.5}$$

and applies the ALS algorithm to find the minimum point of (C.5) with respect to $\mathbf{W}$, $\mathbf{A}$, and $\mathbf{a_0}$ subject to $\mathbf{w}_p'\mathbf{S}_p\mathbf{w}_p = 1$ or $\mathbf{1}_{Jp}'\mathbf{w}_p = 1$ ($p = 1, 2, \cdots, P$), where $\mathbf{S}_p$ is an J_p by J_p sample covariance matrix of $\mathbf{z}_p$.

The proposed ALS algorithm begins by assigning random initial values to $\mathbf{A}$ and repeats two steps until convergence. In the first step, $\mathbf{W}$ and $\mathbf{a}_0$ are updated with $\mathbf{A}$ fixed. By solving $\frac{1}{2}\frac{\partial f^*_{cvx}}{\partial \mathbf{a}_0} = 0$, the least square estimates of $\mathbf{a}_0$ given $\mathbf{W}$ and $\mathbf{A}$ can be obtained as

$$\hat{\mathbf{a}}_0 = \hat{\boldsymbol{\mu}}(\mathbf{V} - \mathbf{WA}) \tag{C.6}$$

This implies that the least squares estimate of $\mathbf{a}_0$ can be expressed as a function of $\mathbf{W}$ and $\mathbf{A}$. In other words, if we can find the least square estimate of $\mathbf{W}$ given $\mathbf{A}$ under the constraint (C.6), we can obtain $\hat{\mathbf{a}}_0$ as well by (C.6). Inserting (C.6) into (C.5) makes (C.5) be simplified as

$$
\begin{aligned}
& f^*_{cvx}(\mathbf{W}, \mathbf{A}) \\
& = tr(\widehat{\mathbf{O}}(\mathbf{V} - \mathbf{WA})'\mathbf{S}(\mathbf{V} - \mathbf{WA})\widehat{\mathbf{O}}) \\
& = N^{-1} SS(\mathbf{D}_{ct}(\mathbf{V} - \mathbf{WA})\widehat{\mathbf{O}}),
\end{aligned} \tag{C.7}
$$

where $\mathbf{D}_{ct} \equiv \mathbf{D} - \mathbf{1}_N\hat{\boldsymbol{\mu}}'$. Let $\widehat{\mathbf{O}}_{\mathrm{Y}}$ denote a T by T_{Y} matrix consisting of all nonzero columns of $\widehat{\mathbf{O}}$, where $T \equiv P + J$ and T_{Y} is the number of dependent variables in the model. Let $\mathbf{I}_0 \equiv [\mathbf{I}_J, \mathbf{0}_{J\times P}]$ and $\mathbf{A}_{\mathrm{I}} \equiv \mathbf{A} - [\mathbf{0}_{P\times J}, \ \mathbf{I}_P]$. Let $\mathbf{W}_{-p}$ denote a J by $(P - 1)$ matrix formed by the columns of $\mathbf{W}$ except for its pth column. Let $\mathbf{A}_{\mathrm{I},-p}$ denote a $(P - 1)$ by T matrix formed by the rows of $\mathbf{A}_{\mathrm{I}}$ except for its pth row and $\mathbf{a}_p$ denote a row vector whose entries are the non-zero elements of the pth row of $\mathbf{A}_{\mathrm{I},-p}$ corresponding to $\mathbf{w}_p$. Let *vec*() denote an operator that returns a column vector obtained by stacking the columns of input matrix vertically. Given $\mathbf{A}$, (C.7) can be re-expressed as

$$
\begin{aligned}
& f^*_{cvx}(\mathbf{w}_p; \mathbf{A}, \mathbf{W}_{-p}) \\
& = N^{-1} SS(\mathbf{D}_{ct}(\mathbf{V} - \mathbf{WA})\widehat{\mathbf{O}}_{\mathrm{Y}}) \\
& = N^{-1} SS(\mathbf{D}_{ct}([\mathbf{I}_J, \mathbf{0}_{J\times P}] + [\mathbf{0}_{J\times J}, \ \mathbf{W}] - \mathbf{WA})\widehat{\mathbf{O}}_{\mathrm{Y}}) \\
& = N^{-1} SS(\mathbf{D}_{ct}(\mathbf{I}_0 - \mathbf{WA}_{\mathrm{I}})\widehat{\mathbf{O}}_{\mathrm{Y}}) \\
& = N^{-1} SS(\mathbf{D}_{ct}(\mathbf{I}_0 - \mathbf{W}_{-p}\mathbf{A}_{\mathrm{I},-p} - \mathbf{w}_p\mathbf{a}_p)\widehat{\mathbf{O}}_{\mathrm{Y}}) \\
& = N^{-1} SS(vec(\mathbf{D}_{ct}(\mathbf{I}_0 - \mathbf{W}_{-p}\mathbf{A}_{\mathrm{I},-p})\widehat{\mathbf{O}}_{\mathrm{Y}}) - ((\mathbf{a}_p\widehat{\mathbf{O}}_{\mathrm{Y}})' \otimes \mathbf{D}_{ct})\mathbf{w}_p)) \\
& = N^{-1} SS(\boldsymbol{\psi}_1 - \boldsymbol{\Xi}_1\mathbf{w}_p)),
\end{aligned} \tag{C.8}
$$

where $\boldsymbol{\psi}_1 \equiv vec(\mathbf{D}_{ct}(\mathbf{I}_0 - \mathbf{W}_{-p}\mathbf{A}_{\mathrm{I},-p})\widehat{\mathbf{O}}_{\mathrm{Y}})$ and $\boldsymbol{\Xi}_1 \equiv (\mathbf{a}_p\widehat{\mathbf{O}}_{\mathrm{Y}})' \otimes \mathbf{D}_{ct}$.

If γ_p is standardized component, the unstandardized least square estimate of $\mathbf{w}_p$ is obtained by

$$
\widehat{\mathbf{w}}_{p*} = (\boldsymbol{\Xi}_1'\boldsymbol{\Xi}_1)^{-1}\boldsymbol{\Xi}_1\boldsymbol{\psi}_1. \tag{C.9}
$$

Then, the standardized least square estimate of $\mathbf{w}_p$ is obtained by $\widehat{\mathbf{w}}_p = (\widehat{\mathbf{w}}_{p*}'\mathbf{S}_p\widehat{\mathbf{w}}_{p*})^{1/2}$ such that $\widehat{\mathbf{w}}_p$ can satisfy $\widehat{\mathbf{w}}_p'\mathbf{S}_p\widehat{\mathbf{w}}_p = 1$. If every element of $\widehat{\mathbf{w}}_p$ is forced to be positive, finding $\widehat{\mathbf{w}}_p$ that minimizes (C.8) becomes a well-known *nonnegative least squares problem* (NNLS;

Lawson & Hanson, 1974, Chapter 23), which should be solved numerically. For instance, the function *lsqnonneg* in MATLAB or the *nnls* package in R can be utilized under this condition.

If γ_p is a convex component, the ALS algorithm finds the solution for (C.8) subject to $\mathbf{1}_{Jp}'\mathbf{w}_p = 1$. This minimization is a linearly constrained least squares problem (Boyd & Vandenberghe, 2018, Chapter 16). As the product of the ranks of two matrices equals to the rank of the Kronecker product of the two matrices and $\mathbf{a}_p\widehat{\mathbf{O}}_\mathrm{Y}$ has one row, $(\mathbf{a}_p\widehat{\mathbf{O}}_\mathrm{Y})' \otimes \mathbf{D}_{ct}$ has linearly independent columns, thereby having the columns of $\begin{bmatrix}\boldsymbol{\Xi}_1 \\ \mathbf{1}_{Jp}'\end{bmatrix}$ are also linearly independent. Thus, there exists $\boldsymbol{\delta}$ satisfying

$$\begin{bmatrix}\boldsymbol{\Xi}_1'\boldsymbol{\Xi}_1 & \mathbf{1}_{Jp} \\ \mathbf{1}_{Jp}' & 0\end{bmatrix}\begin{bmatrix}\mathbf{w}_p \\ \boldsymbol{\delta}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Xi}_1'\boldsymbol{\psi}_1 \\ 1\end{bmatrix}, \tag{C.10}$$

where $\begin{bmatrix}\boldsymbol{\Xi}_1'\boldsymbol{\Xi}_1 & \mathbf{1}_{Jp} \\ \mathbf{1}_{Jp}' & 0\end{bmatrix}$ is invertible. Let $\boldsymbol{\Xi}_2 \equiv \begin{bmatrix}\boldsymbol{\Xi}_1'\boldsymbol{\Xi}_1 & \mathbf{1}_{Jp} \\ \mathbf{1}_{Jp}' & 0\end{bmatrix}$ and $\boldsymbol{\psi}_2 \equiv \begin{bmatrix}\boldsymbol{\Xi}_1'\boldsymbol{\psi}_1 \\ 1\end{bmatrix}$. Then, $\widehat{\mathbf{w}}_p$ can be obtained by the first J_p entries of $\boldsymbol{\Xi}_2^{-1}\boldsymbol{\psi}_2$, from which $\widehat{\mathbf{W}}$ is updated. If $\mathbf{w}_p$ is forced to be positive, minimizing (C.8) becomes a quadratic programming problem with a linear constraint and an inequality constraint (e.g., Floudas & Visweswaran, 1995; Frank & Wolfe, 1956). This problem does not have closed-form solution. Instead, it can be solved numerically via interior point methods (e.g., Altman & Gondzio, 1999; Vanderbei & Carpenter, 1993). For instance, the function *lsqlin* in MATLAB or the *quadprog* package in R can be utilized to minimize (C.8) numerically. This process repeats for every $\mathbf{w}_p$ (p = 1, 2, ⋯, P). Then $\hat{\mathbf{a}}_0$ is updated by (C.6).

In the second step, $\mathbf{A}$ and $\mathbf{a_0}$ are updated with $\mathbf{W}$ fixed. Given $\mathbf{W}$, (C.8) can be re-expressed as

$$\begin{aligned} & f_{cvx}^*(\mathbf{A};\mathbf{W}) \\ & = N^{-1}SS(vec(\mathbf{D}_{ct}\mathbf{V}\widehat{\mathbf{O}}_\mathrm{Y}) - (\widehat{\mathbf{O}}_\mathrm{Y}' \otimes (\mathbf{D}_{ct}\mathbf{W}))vec(\mathbf{A})) \\ & = N^{-1}SS(vec(\mathbf{D}_{ct}\mathbf{V}\widehat{\mathbf{O}}_\mathrm{Y}) - \boldsymbol{\Xi}_3\boldsymbol{\rho}), \end{aligned} \tag{C.11}$$

where $\boldsymbol{\Xi}_3$ is the matrix formed by the columns of ($\widehat{\mathbf{O}}_Y' \otimes (\mathbf{D}_{ct}\mathbf{W})$) corresponding to the nonzero elements in *vec*(**A**), and **ρ** is the column vector of the nonzero elements of *vec*(**A**). Then, the value of **ρ** that minimizes (C.11) is obtained by

$$\boldsymbol{\rho} = (\boldsymbol{\Xi}_3'\boldsymbol{\Xi}_3)^{-1}\boldsymbol{\Xi}_3' vec(\mathbf{D}_{ct}\mathbf{V}\,\widehat{\mathbf{O}}_Y) \tag{C.12}$$

from which the non-zero elements of $\widehat{\mathbf{A}}$ are updated. Then $\hat{\mathbf{a}}_0$ is updated by (C.6).

Appendix C5. A procedure for deriving the population covariance matrix of indicators from the prescribed parameter values of the GSCA model with convex components

The proposed procedure imitates the one suggested by Cho and Choi's (2020) one, while simply replacing standardized components and correlation matrix of indicators with convex components and covariance matrix of indicators, respectively. Let $\mathbf{c}_p$ is a J_p by 1 vector of loadings for $\mathbf{z}_p$. Let $\boldsymbol{\xi}_p$ is a J_p by 1 vector of error terms for $\mathbf{z}_p$. Let $\boldsymbol{\Phi}_{std}$ denote a P by P correlation matrix of components. Let $\boldsymbol{\Delta}$ denote a P by P diagonal matrix whose pth entry is the standard deviation of γ_p, denoted by ϕ_p. Let $\boldsymbol{\Theta} = blkdiag(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \cdots, \boldsymbol{\Theta}_P)$ denote a J by J covariance matrix of errors in the measurement model, where $\boldsymbol{\Theta}_p$ is a J_p by J_p covariance matrix of errors for the pth block of indicators. Let $\boldsymbol{\tau} = [\tau_1, \tau_2, \cdots, \tau_P]'$ denote a P by 1 vector of component means.

Given the prescribed values of $\boldsymbol{\Sigma}_p$, $\boldsymbol{\mu}_p$, and $\boldsymbol{\Phi}_{std}$ (p = 1, 2, ⋯, P), $\mathbf{w}_p$ is obtained by $\mathbf{w}_p = (\boldsymbol{\Sigma}_p^{-1/2}\mathbf{u}_{1p}) / \mathbf{1}_p'\boldsymbol{\Sigma}_p^{-1/2}\mathbf{u}_{1p})$, where $\mathbf{u}_{1p}$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Sigma}_p$, indicating that $\mathbf{w}_p$ maximizes the sum of explained variances of $\mathbf{z}_p$ given $\boldsymbol{\Sigma}_p$ subject to $\mathbf{1}_p'\mathbf{w}_p = 1$. Then, τ_p and ϕ_p are calculated as $\tau_p = \mathbf{w}_p'\boldsymbol{\mu}_p$ and $\phi_p = \mathbf{w}_p'\boldsymbol{\Sigma}_p\mathbf{w}_p$, respectively, based on which $\mathbf{c}_p$ are obtained by $\mathbf{c}_p = \phi_p^{-2}\boldsymbol{\Sigma}_p\mathbf{w}_p$, implying that $\mathbf{c}_p$ is a vector of least-square loading values of $\mathbf{z}_p$ on γ_p. In turn, $\boldsymbol{\Theta}_p$ is obtained by $\boldsymbol{\Theta}_p = (\mathbf{I}_{Jp} - \mathbf{c}_p\mathbf{w}_p')\boldsymbol{\Sigma}_p(\mathbf{I}_{Jp} - \mathbf{w}_p\mathbf{c}_p')$. Then, all the block-diagonal elements of $\mathbf{W}$, $\mathbf{C}$, and $\boldsymbol{\Theta}$ can be filled in with $\mathbf{w}_p$, $\mathbf{c}_p$, and $\boldsymbol{\Theta}_p$ (p = 1, 2, ⋯, P), respectively. Also, $\boldsymbol{\tau}$ is computed by $\boldsymbol{\tau} = \mathbf{W}'\boldsymbol{\mu}$ and then, $\mathbf{c}_0$ is calculated as $\mathbf{c}_0 = \boldsymbol{\mu} - \mathbf{C}'\boldsymbol{\tau}$. Next, $\boldsymbol{\Phi}$ is derived by $\boldsymbol{\Phi} = \boldsymbol{\Delta}\boldsymbol{\Phi}_{std}\boldsymbol{\Delta}$. Finally, we obtain $\boldsymbol{\Sigma}$ by $\boldsymbol{\Sigma} = \mathbf{C}'\boldsymbol{\Phi}\mathbf{C} + \boldsymbol{\Theta}$. A more detailed explanation on each step of this procedure can be found in Cho and Choi (2020).

Appendix D for Chapter 4

Appendix D1. Model specification in DL-GSCA

Let $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_P]'$ denote a P by 1 random vector of components, where γ_p is the pth component ($p = 1, 2, \cdots, P$), and P is the total number of components. Let $\mathbf{z}_p = [z_{1,p}; z_{2,p}; \cdots; z_{Jp,p}]$ denote a J_p by 1 random vector of indicators for γ_p, called a block of indicators for γ_p, where $z_{i,p}$ is the ith indicator in $\mathbf{z}_p$ ($i = 1, 2, \cdots, J_p$) and J_p is the number of indicators for γ_p. Let $\mathbf{z} = [\mathbf{z}_1; \mathbf{z}_2; \cdots; \mathbf{z}_P]$ denote a J by 1 random vector of indicators, where J is the total number of indicators. Both indicators and components are assumed to be standardized, i.e., $E(z_{i,p}) = E(\gamma_p) = 0$ and $var(z_{i,p}) = var(\gamma_p) = 1$. Each component γ_p is defined as $\gamma_p = f_{w,p}(\mathbf{z}_p)$, where $f_{w,p}$: $\mathbb{R}^{Jp} \rightarrow \mathbb{R}$ denotes a continuous function of $\mathbf{z}_p$ to γ_p. Let $f_{c,p}$: $\mathbb{R} \rightarrow \mathbb{R}^{Jp}$ denote a continuous function of γ_p to $\hat{\mathbf{z}}_p$, where $\hat{\mathbf{z}}_p$ is the predicted value of $\mathbf{z}_p$ given γ_p. DL-GSCA assumes at default that γ_p is to explain the variance of $\mathbf{z}_p$. If $f_{w,p}$ and $f_{c,p}$ are linear, then $f_{w,p}(\mathbf{z}_p)$ and $f_{c,p}(\gamma_p)$ can be expressed as $f_{w,p}(\mathbf{z}_p) = \mathbf{w}_p'\mathbf{z}_p$ and $f_{c,p}(\gamma_p) = \mathbf{c}_p\gamma_p$ where $\mathbf{w}_p$ is a J_p by 1 vector of weight parameters for $\mathbf{z}_p$, and $\mathbf{c}_p$ is a J_p by 1 vector of weight parameters for $\mathbf{z}_p$. If γ_p is not assumed to explain $\mathbf{z}_p$, then $f_{c,p}(\gamma_p) = \mathbf{0}$. Let $\mathbf{b}_p = [b_{1,p}, b_{2,p}, \cdots, b_{P,p}]'$ denote a P by 1 vector of path coefficients relating $\boldsymbol{\gamma}$ to γ_p, where $b_{q,p}$ is non-zero if γ_q is assumed to influence γ_p and zero otherwise ($q = 1, 2, \cdots, P$ and $q \neq p$). Let $f_W(\mathbf{z}) \equiv [f_{w,1}(\mathbf{z}_1); f_{w,2}(\mathbf{z}_2); \cdots; f_{w,P}(\mathbf{z}_P)]$, $f_C(\boldsymbol{\gamma}) \equiv [f_{c,1}(\gamma_1), f_{c,2}(\gamma_2), \cdots, f_{c,P}(\gamma_P)]'$, and $\mathbf{B} \equiv [\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_P]$. Let $\boldsymbol{\xi} = [\xi_1; \xi_2; \cdots; \xi_P]$ denote a J by 1 vector of errors for $\mathbf{z}$ in the component measurement model, where ξ_p is a J_p by 1 vector of errors for $\mathbf{z}_p$. Let $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \cdots, \zeta_P]'$ denote a P by 1 vector of errors for $\boldsymbol{\gamma}$ in the structural model, where ζ_p is an error for γ_p. Then, the three sub-models of DL-GSCA are expressed as (4.4), (4.5), and (4.6). Furthermore, DL-GSCA combines the sub-models into a single equation as follows.

$$\begin{aligned} &[\mathbf{z}; \boldsymbol{\gamma}] = [f_C(\boldsymbol{\gamma}); \mathbf{B}'\boldsymbol{\gamma}] + [\boldsymbol{\xi}; \boldsymbol{\zeta}] \\ &\leftrightarrow [\mathbf{z}; f_W(\mathbf{z})] = f_A(\boldsymbol{\gamma}) + [\boldsymbol{\xi}; \boldsymbol{\zeta}] \\ &\leftrightarrow f_V(\mathbf{z}) = f_A(f_W(\mathbf{z})) + \boldsymbol{\varepsilon}, \end{aligned} \tag{D.1}$$

where $f_V(\mathbf{z}) = [\mathbf{z}; f_W(\mathbf{z})]$, $f_A(\boldsymbol{\gamma}) = [f_C(\boldsymbol{\gamma}); \mathbf{B}'\boldsymbol{\gamma}]$, and $\boldsymbol{\varepsilon} = [\boldsymbol{\xi}; \boldsymbol{\zeta}]$. This is called the DL-GSCA model.

Appendix D2. Approximations to DL-GSCA's f_W and f_C via deep learning's artificial neural networks

DL-GSCA utilizes deep learning (DL) for estimating each f_W and f_C in a data-driven manner. As f_W and f_C can be seen as function sets of $f_{w,p}$ and $f_{c,p}$ across components (p = 1, 2, ⋯, P), respectively, DL-GSCA splits f_W and f_C into their small ingredients (i.e., $f_{w,p}$ and $f_{c,p}$) and processes each of them separately. Let $h_{w,p}$ and $h_{c,p}$ denote DL's two artificial neural networks that approximate $f_{w,p}$ and $f_{c,p}$, respectively. As briefly discussed earlier, these functions are built based on function composition, indicating that $h_{w,p}$ and $h_{c,p}$ map their input arguments to the output through a sequence of functions. Specifically, $h_{w,p}$ is defined as $\gamma_p = h_{w,p}(\mathbf{z}_p) = h_{w,p}^{(Lp+1)}(h_{w,p}^{(Lp)}(\cdots(h_{w,p}^{(1)}(\mathbf{z}_p))))$, where $h_{w,p}^{(l)}$ is the lth function in a sequence of functions for $h_{w,p}$ (l =1, 2, ⋯, L_p+1) and L_p+1 is the total number of functions in the sequence for γ_p. Likewise, $h_{c,p}(\gamma_p)$ is defined as $\hat{\mathbf{z}}_p = h_{c,p}(\gamma_p) = h_{c,p}^{(Lp+1)}(h_{c,p}^{(Lp)}(\cdots(h_{c,p}^{(1)}(\gamma_p))))$, where $h_{c,p}^{(l)}$ is the lth function in a sequence of functions for $h_{c,p}$ (l =1, 2, ⋯, L_p+1). As stated earlier, $h_{w,p}^{(l)}$ and $h_{c,p}^{(l)}$ are called layers in DL. Once $h_{w,p}$ and $h_{c,p}$ are defined, h_W and h_C are defined as $h_W(\mathbf{z}) = [h_{w,1}(\mathbf{z}_1); h_{w,2}(\mathbf{z}_2);\cdots; h_{w,P}(\mathbf{z}_P)]$ and $h_C(\boldsymbol{\gamma}) = [h_{c,1}(\gamma_1); h_{c,2}(\gamma_2);\cdots; h_{c,P}(\gamma_P)]$, which approximate f_W and f_C, respectively. Taken together, DL-GSCA employs a total of 2P artificial neural networks to approximate f_W and f_C in (4.4) and (4.5).

There are three types of layers: input, hidden, and output layers. The input layer refers to a vector of input arguments, for instance, $\mathbf{z}_p$ for $h_{w,p}$ and γ_p for $h_{c,p}$, whereas the output layer refers to the last layer in a sequence of functions, such as, $h_{w,p}^{(Lp+1)}$ for $h_{w,p}$ and $h_{c,p}^{(Lp+1)}$ for $h_{c,p}$. The other functions between the input and output layers constitute hidden layers, for example, $h_{w,p}^{(Lp)}$, $h_{w,p}^{(Lp-1)}$, ⋯, $h_{w,p}^{(1)}$ for $h_{w,p}$ and $h_{c,p}^{(Lp)}$, $h_{c,p}^{(Lp-1)}$, ⋯, $h_{c,p}^{(1)}$ for $h_{c,p}$. Only the hidden and output layers are called computational layers, as the input layer simply transmits its input arguments to the next layer (Aggarwal, 2018, p. 6). The output values of each hidden layer are called hidden units in the hidden layer, and all indicators, components, and hidden

units are collectively called nodes. The number of hidden units in the lth hidden layer is denoted by $R_p^{(l)}$.

Figure D2.1 depicts an example of layers for $h_{w,p}$. The input layer includes three indicators for a component, whereas the output layer contains the component. There are two hidden layers, each of which contains four hidden units, signified by dotted hexagons. A hidden unit refers to an element of the output vector of each hidden layer. In this example, there are three computational layers for $h_{w,p}$.

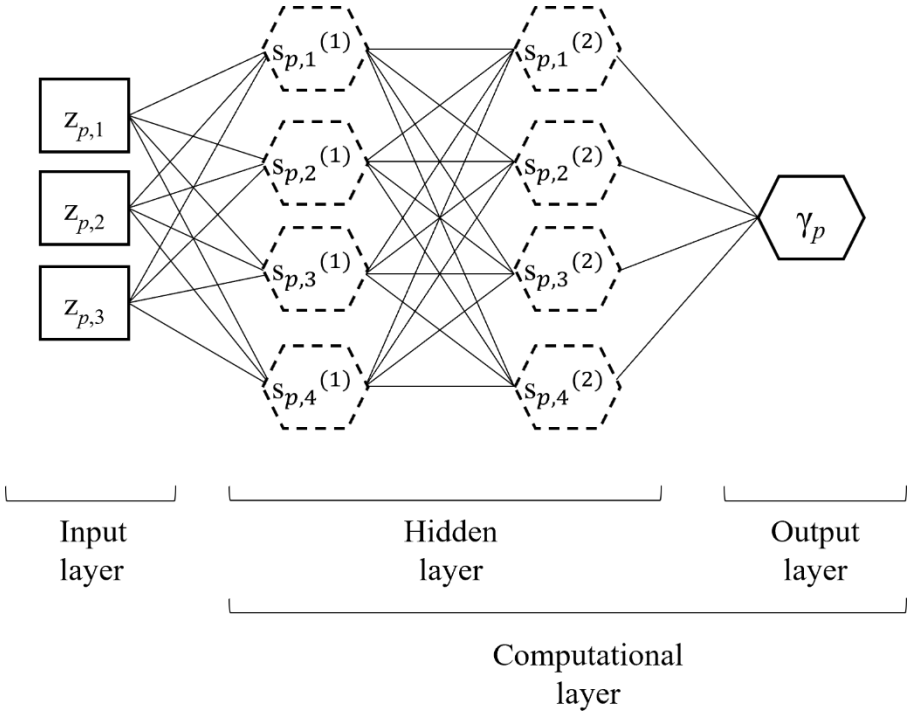


Figure D2.1. An example of layers for $h_{w,p}$ in DL-GSCA's weighted relation model. Dotted circles represent hidden units per hidden layer. Straight lines denote weight parameters.

Each node per computational layer is obtained by transforming the weighted sum of nodes in the previous layer in a linear or nonlinear fashion. Specifically, let $s_{p,a}^{(l)}$ denote the ath node in the lth computational layer for $h_{w,p}$ (a =1, 2, ⋯, $R_p^{(l)}$; l =1, 2, ⋯, L_p+1). Let $\mathbf{s}_p^{(l)} \equiv [s_{p,1}^{(l)}, s_{p,2}^{(l)}, \cdots, s_{p,Rp}^{(l)}]'$ and $\mathbf{s}_p^{(0)} \equiv \mathbf{z}_p$. Let $\mathbf{W}_p^{(l)}$ denote an $R_p^{(l-1)}$ by $R_p^{(l)}$ matrix of weights in the lth computational layer, where $R_p^{(0)} \equiv J_p$. Let $\boldsymbol{\beta}_{w,p}^{(l)}$ denote a $R_p^{(l)}$ by 1 vector of

biases/intercepts in the lth computational layer. Let $\alpha(\mathbf{x}) = [\alpha_1(x_1), \alpha_2(x_2), \cdots, \alpha_M(x_M)]'$ denote an activation function that transforms each entry of the input vector $\mathbf{x}$ linearly or nonlinearly, where $\mathbf{x} = [x_1, x_2, \cdots, x_M]'$ is any vector of size M and α_m is an activation function that is applied to x_m ($m = 1, 2, \cdots, M$). Then, $h_{w,p}^{(l)}$ in $h_{w,p}$ can be expressed as $\mathbf{s}_p^{(l)} = h_{w,p}^{(l)}(\mathbf{s}_p^{(l-1)}) = \alpha(\mathbf{W}_p^{(l)\prime}\mathbf{s}_p^{(l-1)} + \boldsymbol{\beta}_{w,p}^{(l)})$, indicating that a set of nodes in the lth computational layer for $h_{w,p}$ is defined as a linear or nonlinear function of a weighted sum of nodes in the (l–1)th computational layer for $h_{w,p}$.

We can consider several types of activation function, including identity, sigmoid, tangent hyperbolic (tanh), bounded linear unit (BLU), and rectified linear unit (ReLU). Refer to Nwankpa, Ijomah, Gachagan, and Marshall (2021) for the characteristics of these activation functions. The ReLU function is one of the most popular activation functions for hidden units in modern deep learning architectures because of its superior performance in multiple-layer cases (e.g., Aggarwal, 2018; Nwankpa et al., 2021). Thus, in DL-GSCA, by default, the ReLU function is used for hidden units, which can be expressed as

$$\alpha_m(x_m) = max(0, x_m). \tag{D.2}$$

Figure D2.2(a) displays the ReLU function, which is a continuous, piecewise, linear function. Based on the property of a piecewise linear function, $h_{w,p}$ involving the ReLU also becomes continuous and piecewise linear in $\mathbf{z}_p$(Strang, 2019, p. 375). This indicates that $h_{w,p}$ approximates $f_{w,p}$ by a linear combination of multiple line pieces. On the other hand, two activation functions, identity and BLU (Zhangyang et al., 2016), are recommended for the output layer, which can be written as

$$\alpha_m(x_m) = x_m \text{ (identity)} \tag{D.3}$$

$$\alpha_m(x_m) = max(min(x_m, \tau_{2,m}), \tau_{1,m}) \text{ (BLU)}, \tag{D.4}$$

where $\tau_{1,m}$ and $\tau_{2,m}$ are the lower and upper limits of x_m, respectively.[4] Figure D2.2(b) and Figure D2.2(c) exhibit the two activation functions. In DL-GSCA, the BLU function is preferred to the identity function for the output layer of $h_{w,p}$ because it can suppress the occurrence of extreme component scores from a test sample by truncating them to lie in the range of the component scores estimated from a training sample (e.g., [–3, 3]). Without such a truncation scheme, abnormally large component scores can be obtained from a test sample if this sample includes outliers, leading to large out-of-sample prediction error.

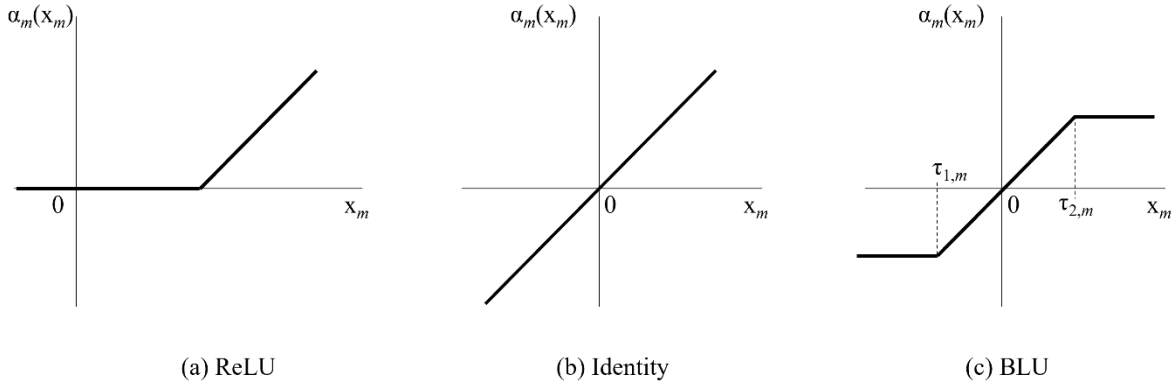


Figure D2.2. Three types of activation functions used in DL-GSCA.

Figure D2.3 depicts an example of layers for $h_{c,p}$, which consists of the same number of computational layers as those for $h_{w,p}$. The input layer in $h_{c,p}$ includes a component, whereas the output layer includes the predicted values of the component's indicators. Let $t_{p,a}^{(l)}$ denote the ath node in the lth computational layer for $h_{c,p}$ (a =1, 2, ⋯, $R_p^{(l)}$; l =1, 2, ⋯, L_p+1). Let $\mathbf{t}_p^{(l)} \equiv [t_{p,1}^{(l)}, t_{p,2}^{(l)}, \cdots, t_{p,Rp}^{(l)}]'$ and $\mathbf{t}_p^{(0)} \equiv \gamma_p$. Let $\mathbf{C}_p^{(l)}$ denote an $R_p^{(l-1)}$ by $R_p^{(l)}$ matrix of loadings, and $\boldsymbol{\beta}_{c,p}^{(l)}$ denote a $R_p^{(l)}$ by 1 vector of biases. Then, $h_{c,p}^{(l)}$ in $h_{c,p}$ can be expressed as $\mathbf{t}_p^{(l)} = h_{c,p}^{(l)}(\mathbf{t}_{p(l-1)}) = \alpha(\mathbf{C}_p^{(l)\prime}\mathbf{t}_p^{(l-1)} + \boldsymbol{\beta}_{c,p}^{(l)})$, indicating that a set of nodes in the lth computational layer for $h_{c,p}$ is defined as a linear or nonlinear function of a weighted sum of nodes in the (l–1)th computational layer for $h_{c,p}$. The ReLU function is again used for all

[4] In the original BLU function, $\tau_{1,m}$ and $\tau_{2,m}$ are restricted as $1 = \tau_{2,m} = -\tau_{1,m}$.

hidden layers of $h_{c,p}$, while the identity function is for the output layer. The BLU function is not used for the output layer because the values of $\hat{\mathbf{z}}_p$ are not expected to be abnormally extreme as the argument of $h_{c,p}$ (i.e., the score of γ_p) is already truncated not to have an extreme value.

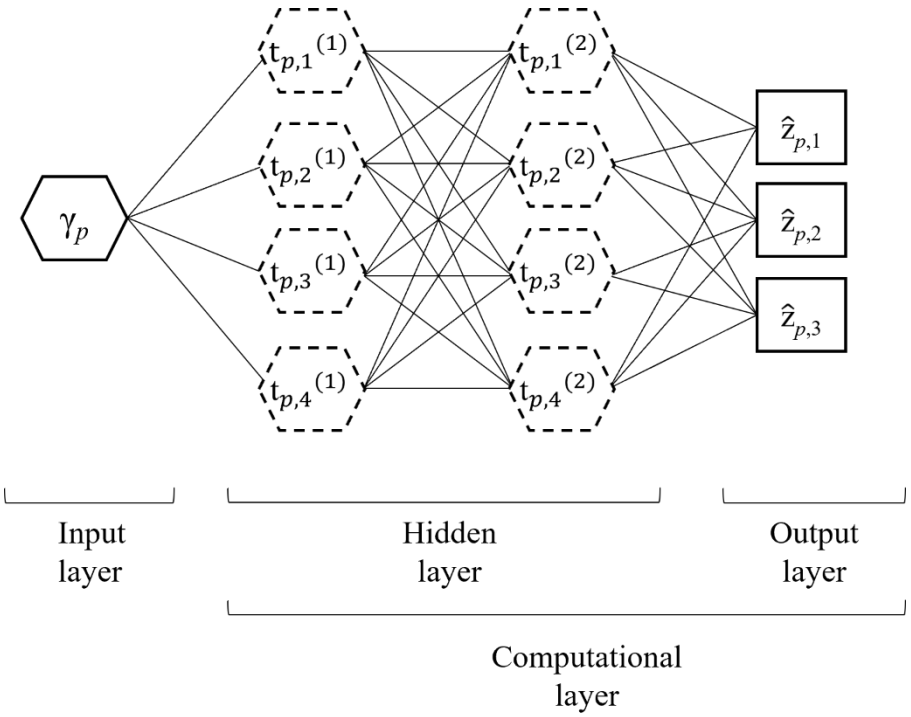


Figure D2.3. An example of layers for $h_{c,p}$ in DL-GSCA's component measurement model. Dotted circles represent hidden units per hidden layer in the measurement model. Straight lines denote loading parameters.

Appendix D3. Parameter estimation procedure for DL-GSCA

Given the data of indicators, DL-GSCA aims to estimate $h_{w,p}$, $h_{c,p}$, and $\mathbf{B}$ in such a way that components minimize the average unexplained variance of all dependent variables in the model. Let $\mathbf{d}_{n,p}$ denote a J_p by 1 vector of the standardized scores of $\mathbf{z}_p$ for the nth individual in a sample of N individuals. Let $\mathbf{d}_n \equiv [\mathbf{d}_{n,1}; \mathbf{d}_{n,2}; \cdots; \mathbf{d}_{n,P}]$ and $\mathbf{D} \equiv [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_N]'$. Let $g_{n,p}$ denote the nth individual's standardized score for γ_p, which can be obtained by

$$\mathbf{g}_{n,p} = h_{w,p}(\mathbf{d}_{n,p}). \tag{D.5}$$

Let $\mathbf{g}_n \equiv [g_{n,1}, g_{n,2}, \cdots, g_{n,P}]'$ and $\mathbf{G} \equiv [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N]'$. Let $\mathbf{e}_n$ denote a $J + P$ by 1 vector of residuals for $[\mathbf{d}_n; \mathbf{g}_n]$. Let $SS(\mathbf{X}) \equiv trace(\mathbf{X}'\mathbf{X})$, where $\mathbf{X}$ is any matrix. Let $blkdiag()$ denote an operator that converts a set of input matrices into a block-diagonal matrix that includes each input matrix as its diagonal block. Let $\mathbf{O}_z \equiv blkdiag(\mathbf{O}_{z,1}, \mathbf{O}_{z,2}, \cdots, \mathbf{O}_{z,P})$, where $\mathbf{O}_{z,p} = \mathbf{I}$ if $\mathbf{z}_p$ is explained by γ_p and $\mathbf{O}_{z,p} = \mathbf{0}$ otherwise. Let $\mathbf{O}_\gamma \equiv blkdiag(o_{\gamma,1}, o_{\gamma,2}, \cdots, o_{\gamma,P})$, where $o_{\gamma,p}$ is one if γ_p is explained by other components and zero otherwise. Let $\mathbf{O} \equiv blkdiag(\mathbf{O}_z, \mathbf{O}_\gamma)$. Let T denote the number of dependent variables in the model. Given $\mathbf{D}$, DL-GSCA seeks to minimize the following optimization criterion

$$\begin{aligned} &\varphi(h_{w,p}, h_{c,p}, \mathbf{B}) \\ &= (TN)^{-1} \sum_{n=1}^{N} SS(\mathbf{e}_n\mathbf{O}) \\ &= (TN)^{-1} \sum_{p=1}^{P} \left(SS((\mathbf{d}_{n,p} - h_{c,p}(g_{n,p}))\mathbf{O}_{z,p}) + SS(g_{n,p} - \mathbf{g}_n\mathbf{b}_p)o_{\gamma,p}) \right), \end{aligned} \tag{D.6}$$

subject to the standardization constraints on components. The criterion (D.6) is equivalent to the average (squared) prediction error for all the dependent variables in the model.

As this constrained optimization criterion cannot be minimized in closed form, we develop an alternating least squares (ALS) algorithm to minimize (D.6) iteratively. Specifically, the ALS algorithm begins by assigning initial values to the parameters in the following way. It utilizes the estimates of GSCA as the initial values of $\mathbf{G}$ and $\mathbf{B}$ and then, applies the He initialization (He et al., 2015) to assign random initial values to $h_{w,p}$ and $h_{c,p}$. The He initialization is more efficient for convergence than any other procedures when the

Relu function is used for deep learning (Kumar, 2017). Then, the algorithm alternates the following two steps until the difference in the value of (D.6) between consecutive iterations becomes smaller than a prescribed tolerance level (e.g., .0001).

Step 1. Update $h_{w,p}$, $h_{c,p}$, and $\mathbf{b}_p$ with $h_{w,q}$, $h_{c,q}$, and $\mathbf{b}_q$ fixed for all $q \in \mathbb{Q}_p$, where $\mathbb{Q}_p$ is a set of integers from 1 to P except for p. Let $\boldsymbol{\varsigma}_p$ denote the pth row of $\mathbf{B} - \mathbf{I}$, and $\boldsymbol{\Omega}_p$ denote $\mathbf{I} - \mathbf{B}$ with the pth row removed. Let $\boldsymbol{\pi}_p$ denote a row vector of the nonzero entries of $\boldsymbol{\varsigma}_p$. Let $\boldsymbol{\Phi}_p$ denote a matrix of the columns of $\boldsymbol{\Omega}_p$ that correspond to the non-zero entries of $\boldsymbol{\varsigma}_p$. Let $\boldsymbol{\Omega}_{p,z}$ denote a matrix of the columns of $\boldsymbol{\Omega}_p$ that correspond to the zero entries of $\boldsymbol{\varsigma}_p$. With the standardization constraints imposed on components, the objective function (D.6) can be re-expressed as

$$\begin{aligned}
&\varphi(h_W, h_C, \mathbf{B}) \\
&= (TN)^{-1} \sum_{n=1}^{N} \left(\sum_{k=1}^{P} \left(SS((\mathbf{d}_{n,k} - h_{c,k}(\mathrm{g}_{n,k}))\mathbf{O}_{\mathrm{z},k}) \right) + SS((\mathbf{I} - \mathbf{B})'\mathbf{g}_n\mathbf{O}_{\gamma}) \right) \\
&= (TN)^{-1} \sum_{n=1}^{N} \left(\sum_{k=1}^{P} \left(SS((\mathbf{d}_{n,k} - h_{c,k}(\mathrm{g}_{n,k}))\mathbf{O}_{\mathrm{z},k}) \right) + SS((\boldsymbol{\Omega}_p'\mathbf{g}_{n,-p} - \boldsymbol{\varsigma}_p'\mathrm{g}_{n,p})\mathbf{O}_{\gamma}) \right) \\
&= (TN)^{-1} \sum_{n=1}^{N} \left(SS([(\mathbf{d}_{n,p} - h_{c,p}(h_{w,p}(\mathbf{d}_{n,p})))\mathbf{O}_{\mathrm{z},p}, (\boldsymbol{\Phi}_p'\mathbf{g}_{n,-p} - \boldsymbol{\pi}_p'h_{w,p}(\mathbf{d}_{n,p}))\mathbf{O}_{\gamma}] \right) + \psi_1,
\end{aligned} \tag{D.7}$$

where $\psi_1 = (TN)^{-1} \sum_{n=1}^{N} \left(\sum_{q \in \mathbb{Q}_p} \left(\mathrm{SS}((\mathbf{d}_{n,q} - h_{c,q}(\mathrm{g}_{n,q}))\mathbf{O}_{\mathrm{z},q}) \right) + \mathrm{SS}(\boldsymbol{\Phi}_p'\mathbf{g}_{n,-p}\mathbf{O}_{\gamma}) \right)$ is a constant, indicating that $h_{c,p}$, $h_{w,p}$, and $\boldsymbol{\pi}_p$ can be estimated with $h_{w,q}$, $h_{c,q}$, and $\mathbf{b}_q$ fixed for all $q \in \mathbb{Q}_p$ in such a way that the pth component ($\mathrm{g}_{n,p}$) can minimize the sum of squared residuals for $\mathbf{d}_{n,p}$ and $\boldsymbol{\Phi}_p'\mathbf{g}_{n,-p}$. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) can be used to minimize (D.7) subject to the standardization constraints. We employ the *fminunc* function in MATLAB, which allows users to implement the BFGS algorithm without manually calculating the gradient of (D.7) (The MathWorks Inc., 2022). Once $h_{c,p}$, $h_{w,p}$, and $\boldsymbol{\pi}_p$ are updated, the non-zero elements of $\mathbf{b}_p$ are updated by the estimated $\boldsymbol{\pi}_p$.Then, h_W, h_C, and $\mathbf{B}$ are updated by $h_{w,p}$, $h_{c,p}$, and $\mathbf{b}_p$.

Step 2. Update $\mathbf{B}$ for h_W and h_C fixed. Let $\boldsymbol{\theta}$ denote a column vector of the non-zero elements of *vec*(**B**), where *vec*() is an operator to converts an input matrix into a column vector by

stacking the columns of the input matrix vertically in order. The operator $\otimes$ denotes the Kronecker product. Let $\boldsymbol{\Xi}$ denote a matrix of the columns of $\mathbf{O}_\gamma \otimes \mathbf{G}$ that correspond to the non-zero elements of $vec(\mathbf{B})$. We can re-write (D.6) as

$$\begin{aligned}
&\varphi(h_{w,p}, h_{c,p}, \mathbf{B}) \\
&= (TN)^{-1} \textstyle\sum_{n=1}^{N} \big(SS((\mathbf{g}_n - \mathbf{B}'\mathbf{g}_n)\mathbf{O}_\gamma) \big) + \psi_2 \\
&= (TN)^{-1} SS(\mathbf{GO}_\gamma - \mathbf{GBO}_\gamma) + \psi_2 \\
&= (TN)^{-1} SS(vec(\mathbf{GO}_\gamma) - (\mathbf{O}_\gamma \otimes \mathbf{G})vec(\mathbf{B})) + \psi_2 \\
&= (TN)^{-1} SS(vec(\mathbf{GO}_\gamma) - \boldsymbol{\Xi}\boldsymbol{\theta}) + \psi_2,
\end{aligned} \tag{D.8}$$

where $\psi_2 = (TN)^{-1} \sum_{n=1}^{N} \big(SS((\mathbf{d}_n - h_c(\mathbf{g}_n))\mathbf{O}_z) \big)$ is a constant. The least squares estimate of $\boldsymbol{\theta}$ can be obtained by

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{\Xi}'\boldsymbol{\Xi})^{-1}\boldsymbol{\Xi}'vec(\mathbf{GO}_\gamma), \tag{D.9}$$

from which the non-zero elements of $\mathbf{B}$ are updated.

DL-GSCA repeats the ALS algorithm with different random initial values to avoid potential convergence to local minima (Hwang & Takane, 2014, p. 24). Then, DL-GSCA selects a set of parameter estimates that leads to the smallest value of (D.6) as the final one.

Appendix D4. Predictive feedforward search algorithm used for tuning DL-GSCA's hyperparameters

Prior to applying the ALS algorithm, DL-GSCA should determine the values of two hyperparameters for each component (i.e., L_p and $R_p^{(l)}$). The predictive feedforward search algorithm (Cho, Hwang, et al., 2022), which was originally introduced for variable selection in GSCA, can be utilized for deciding on the values of the hyperparameters in DL-GSCA.

Given candidate sets of the hyperparameter values for a component, the search algorithm generates K pairs of training and validation samples. Then, the algorithm starts with the smallest candidate values of L_p and $R_p^{(l)}$. The DL-GSCA model with these values is fitted to a training sample, and the model's TE^D value is subsequently calculated from the corresponding validation sample. This procedure is reiterated for the remaining pairs of training and validation samples, leading to the calculation of the average TE^D value over K validation samples.

The algorithm proceeds with the next smallest $R_p^{(l)}$ value and calculates the model's average TE^D value again. Then, it compares the two average TE^D values from two different $R_p^{(l)}$ values to see if the TE^D value decreases with an increase in the $R_p^{(l)}$ value. If it does, the process continues with the next smallest $R_p^{(l)}$ value until the model's TE^D value no longer decreases with an increase in $R_p^{(l)}$, or there are no remaining $R_p^{(l)}$ candidates. The $R_p^{(l)}$ value resulting in the minimum average TE^D is selected as the optimal one for the smallest L_p value.

This process is then conducted for the next smallest L_p value with the aim of identifying the optimal $R_p^{(l)}$ value for that L_p. The algorithm subsequently compares the two average TE^D values from two different combinations of L_p and $R_p^{(l)}$ and determines whether an increase in the L_p value reduces the model's TE^D value. If so, the process proceeds with the next smallest L_p until the model's TE^D value no longer decreases with an increase in L_p or

there are no further L_p candidates. The L_p and $R_p^{(l)}$ pair that leads to the smallest average TE^D is selected as the optimal one for that component.

The search algorithm repeats the above procedures for all components. When the sample size is large, it may be too computationally costly to conduct *K*-fold cross validation. In this case, the algorithm is carried out based on the validation set approach that uses a prescribed percentage of the training sample (e.g., 70%) for training the model and the remaining one for calculating the model's TE^D value.

Appendix D5. Formulae for DL-GSCA's model evaluation indices

The overall goodness-of-fit index, $\mathrm{FIT}^{\mathrm{D}}$, is given as

$$\mathrm{FIT}^{\mathrm{D}} = 1 - \frac{\sum_{n=1}^{N} \mathrm{SS}(([\mathbf{d}_n, \mathbf{g}_n] - [h_C(\mathbf{g}_n), \mathbf{B}'\mathbf{g}_n])\mathbf{O})}{\sum_{n=1}^{N} \mathrm{SS}([\mathbf{d}_n, \mathbf{g}_n])\mathbf{O})}, \tag{D.10}$$

where $\widehat{\mathbf{B}}$ is the estimate of $\mathbf{B}$. The two local goodness-of-fit indices, $\mathrm{FIT}_{\mathrm{M}}^{\mathrm{D}}$ and $\mathrm{FIT}_{\mathrm{S}}^{\mathrm{D}}$, are provided as

$$\mathrm{FIT}_{\mathrm{M}}^{\mathrm{D}} = 1 - \frac{\sum_{n=1}^{N} \mathrm{SS}((\mathbf{d}_n - h_C(\mathbf{g}_n))\mathbf{O}_{\mathrm{z}})}{\sum_{n=1}^{N} \mathrm{SS}(\mathbf{d}_n\mathbf{O}_{\mathrm{z}})}, \tag{D.11}$$

$$\mathrm{FIT}_{\mathrm{S}}^{\mathrm{D}} = 1 - \frac{\sum_{n=1}^{N} \mathrm{SS}((\mathbf{g}_n - \mathbf{B}'\mathbf{g}_n)\mathbf{O}_{\gamma})}{\sum_{n=1}^{N} \mathrm{SS}(\mathbf{g}_n\mathbf{O}_{\gamma})}. \tag{D.12}$$

On the other hand, TE^{D} is defined as

$$\mathrm{TE}^{\mathrm{D}} = \frac{\sum_{n=1}^{N_{tt}} \mathrm{SS}(([\mathbf{d}_{n,tt}, \mathbf{g}_{n,tt}] - [h_{C,k}(\mathbf{g}_{n,tt}), \mathbf{B}'\mathbf{g}_{n,tt}])\mathbf{O})}{\sum_{n=1}^{N_{tt,k}} \mathrm{SS}([\mathbf{d}_{n,tt}, \mathbf{g}_{n,tt}])\mathbf{O})}, \tag{D.13}$$

where N_{tt} is the number of individuals in the test sample; $\mathbf{d}_{n,tt}$ is the nth individual's standardized scores in the test sample (n = 1,2, ⋯, N_{tt}); h_W, h_C, and $\widehat{\mathbf{B}}$ are the parameter estimates obtained from the training sample; and $\mathbf{g}_{n,tt} \equiv h_W(\mathbf{d}_{n,tt})$. Note that $\mathbf{d}_{n,tt}$ should be standardized using the sample means and standard deviations of indicators obtained from the training sample, not from the corresponding test sample, because h_W, h_C, and $\widehat{\mathbf{B}}$ are the estimates obtained from the standardized scores of indicators in the training sample. In addition, two local indices for evaluating the prediction errors of the component measurement and structural models are defined as follows.

$$\mathrm{TE}_{\mathrm{M}}^{\mathrm{D}} = \frac{\sum_{n=1}^{N_{tt}} \mathrm{SS}((\mathbf{d}_{n,tt} - h_C(\mathbf{g}_{n,tt}))\mathbf{O}_{\mathrm{z}})}{\sum_{n=1}^{N_{tt}} \mathrm{SS}(\mathbf{d}_{n,tt}\mathbf{O}_{\mathrm{z}})}, \tag{D.14}$$

$$\mathrm{TE}_{\mathrm{S}}^{\mathrm{D}} = \frac{\sum_{n=1}^{N_{tt}} \mathrm{SS}((\mathbf{g}_{n,tt} - \mathbf{B}'\mathbf{g}_{n,tt})\mathbf{O}_{\gamma})}{\sum_{n=1}^{N_{tt}} \mathrm{SS}(\mathbf{g}_{n,tt}\mathbf{O}_{\gamma})}. \tag{D.15}$$

Moreover, $\mathrm{OPE}^{\mathrm{D}}$ is defined as

$$\mathrm{OPE}^{\mathrm{D}} = \frac{1}{N_{boot}} \sum_{k}^{N_{boot}} \frac{\sum_{n=1}^{N_k} SS(([\mathbf{d}_{n,k}, \mathbf{g}_{n,k}] - [h_{C,k}(\mathbf{g}_{n,k}), \hat{\mathbf{B}}_k'\mathbf{g}_{n,k}])\mathbf{O})}{\sum_{n=1}^{N_k} SS(([\mathbf{d}_{n,k}, \mathbf{g}_{n,k}])\mathbf{O})}, \tag{D.16}$$

where N_{boot} is the number of in-bag and out-of-bag sample sets; N_k is the number of individuals in the kth out-of-bag sample; $\mathbf{d}_{n,k}$ is the nth individual's standardized scores in the kth out-of-bag sample; $h_{W,k}$, $h_{C,k}$, and $\hat{\mathbf{B}}_k$ are the parameter estimates obtained from the kth in-bag sample; and $\mathbf{g}_{n,k} \equiv h_W(\mathbf{d}_{n,k})$. Note that $\mathbf{d}_{n,k}$ should be standardized using the sample means and standard deviations of indicators obtained from the kth in-bag sample. The two local cross-validation indices are defined as follows.

$$\mathrm{OPE}_{\mathrm{M}}^{\mathrm{D}} = \frac{1}{N_{boot}} \sum_{k}^{N_{boot}} \frac{\sum_{n=1}^{N_k} SS((\mathbf{d}_{n,k} - h_{C,k}(\mathbf{g}_{n,k}))\mathbf{O}_{\mathrm{z}})}{\sum_{n=1}^{N_k} SS(\mathbf{d}_{n,k}\mathbf{O}_{\mathrm{z}})}, \tag{D.17}$$

$$\mathrm{OPE}_{\mathrm{S}}^{\mathrm{D}} = \frac{1}{N_{boot}} \sum_{k}^{N_{boot}} \frac{\sum_{n=1}^{N_k} SS((\mathbf{g}_{n,k} - \hat{\mathbf{B}}_k'\mathbf{g}_{n,k})\mathbf{O}_{\gamma})}{\sum_{n=1}^{N_k} SS(\mathbf{g}_{n,k}\mathbf{O}_{\gamma})}. \tag{D.18}$$

Lastly, $\Delta\mathrm{TE}_{p,q}$, which is used to evaluate the predictive power of each individual predictor for its dependent component, is defined as

$$\Delta \mathrm{TE}_{p,q} = \frac{\sum_{n=1}^{N_{tt}} (g_{n,q,tt} - \mathbf{b}_q' \mathbf{g}_{n,tt})^2 - \sum_{n=1}^{N_{tt}} (g_{n,q,tt} - (\mathbf{b}_q \mathbf{I}_{p0})' \mathbf{g}_{n,tt})^2}{\sum_{n=1}^{N_{tt}} (g_{n,q,tt} - \mathbf{b}_q' \mathbf{g}_{n,tt})^2}, \quad \text{(D.19)}$$

where $g_{n,q,tt}$ is the qth component's score in $\mathbf{g}_{n,tt}$, $\hat{\mathbf{b}}_q$ is the qth column of $\widehat{\mathbf{B}}$, and $\mathbf{I}_{p0}$ is the identity matrix of order P, whose pth diagonal entry is zero.

Appendix D6. Data generating procedure for the simulation study

Let $d_{n,p,i}$ denote the *n*th individual's score of the *i*th indicator for the *p*th component in the prototype model in Figure 4.6 ($n = 1, 2, \cdots, N$; $i = 1, 2, 3$; $p = 1, 2, 3$). To generate $d_{n,p,i}$, we begin by obtaining the implied correlation matrix of the GSCA model based on Cho and Choi's (2020) procedure, under the assumption that all indicators are linearly associated. This implied covariance matrix is provided in Table D6.1. We draw a sample of *N* individuals from a multivariate normal distribution with zero means and the covariance matrix. Let $y_{n,p,i}$ denote the *n*th individual's score of the *i*th indicator for the *p*th linear component. We then generate $d_{n,p,i}$ as follows: $d_{n,1,1} = y_{n,1,1}$, $d_{n,1,2} = (y_{n,1,2})^2$, $d_{n,1,3} = -(y_{n,1,3})^2$, $d_{n,2,1} = y_{n,2,1}$, $d_{n,2,2} = (y_{n,2,1} \times y_{n,2,2})$, $d_{n,2,3} = (y_{n,2,2} \times y_{n,2,3})$, $d_{n,3,1} = y_{n,3,1}$, $d_{n,3,2} = y_{n,3,2}$, and $d_{n,3,3} = y_{n,3,3}$.

Table D6.1. The correlation matrix of indicators used in the simulation study.

	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	$y_{3,1}$	$y_{3,2}$	$y_{3,3}$
$y_{1,1}$	1.000	0.800	0.800	0.433	0.433	0.433	0.303	0.303	0.303
$y_{1,2}$	0.800	1.000	0.800	0.433	0.433	0.433	0.303	0.303	0.303
$y_{1,3}$	0.800	0.800	1.000	0.433	0.433	0.433	0.303	0.303	0.303
$y_{2,1}$	0.433	0.433	0.433	1.000	0.800	0.800	-0.303	-0.303	-0.303
$y_{2,2}$	0.433	0.433	0.433	0.800	1.000	0.800	-0.303	-0.303	-0.303
$y_{2,3}$	0.433	0.433	0.433	0.800	0.800	1.000	-0.303	-0.303	-0.303
$y_{3,1}$	0.303	0.303	0.303	-0.303	-0.303	-0.303	1.000	0.800	0.800
$y_{3,2}$	0.303	0.303	0.303	-0.303	-0.303	-0.303	0.800	1.000	0.800
$y_{3,3}$	0.303	0.303	0.303	-0.303	-0.303	-0.303	0.800	0.800	1.000