
Faculty of Social Sciences

Faculty Publications

Generalized Structured Component Analysis with Uniqueness Terms for
Accommodating Measurement Error

Heungsun Hwang, Yoshio Takane, and Kwanghee Jung

December 2017

© 2017 Hwang et al. This is an open access article distributed under the terms of the
Creative Commons Attribution License. <http://creativecommons.org/licenses/by/4.0>

This article was originally published at:
<https://doi.org/10.3389/fpsyg.2017.02137>

Citation for this paper:

Hwang, H.; Takane, Y.; & Jung, K. (2017). Generalized structured component
analysis with uniqueness terms for accommodating measurement error. *Frontiers in
Psychology*, 8, article 2137. DOI: 10.3389/fpsyg.2017.02137



Generalized Structured Component Analysis with Uniqueness Terms for Accommodating Measurement Error

Heungsun Hwang^{1*}, Yoshio Takane² and Kwanghee Jung³

¹ Department of Psychology, McGill University, Montreal, QC, Canada, ² Department of Psychology, University of Victoria, BC, Canada, ³ Department of Educational Psychology and Leadership, Texas Tech University, Lubbock, TX, United States

OPEN ACCESS

Edited by:

Yanyan Sheng,
Southern Illinois University
Carbondale, United States

Reviewed by:

Edward E. Rigdon,
Georgia State University,
United States
Sunho Jung,
Kyung Hee University, South Korea

*Correspondence:

Heungsun Hwang
heungsun.hwang@mcgill.ca

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 26 July 2017

Accepted: 23 November 2017

Published: 06 December 2017

Citation:

Hwang H, Takane Y and Jung K (2017)
Generalized Structured Component
Analysis with Uniqueness Terms for
Accommodating Measurement Error.
Front. Psychol. 8:2137.
doi: 10.3389/fpsyg.2017.02137

Generalized structured component analysis (GSCA) is a component-based approach to structural equation modeling (SEM), where latent variables are approximated by weighted composites of indicators. It has no formal mechanism to incorporate errors in indicators, which in turn renders components prone to the errors as well. We propose to extend GSCA to account for errors in indicators explicitly. This extension, called GSCA_M, considers both common and unique parts of indicators, as postulated in common factor analysis, and estimates a weighted composite of indicators with their unique parts removed. Adding such unique parts or uniqueness terms serves to account for measurement errors in indicators in a manner similar to common factor analysis. Simulation studies are conducted to compare parameter recovery of GSCA_M and existing methods. These methods are also applied to fit a substantively well-established model to real data.

Keywords: generalized structured component analysis, uniqueness, measurement error, bias correction, structural equation modeling

INTRODUCTION

Structural equation modeling (SEM) involves the specification and testing of the relationships between variables that are observed (indicators) and unobserved (latent variables). Two approaches have been proposed for SEM: Factor-based vs. component-based (e.g., Fornell and Bookstein, 1982; Jöreskog and Wold, 1982; Tenenhaus, 2008; Rigdon, 2012). The former includes covariance structure analysis (CSA; Jöreskog, 1970, 1973), and the latter includes partial least squares path modeling (PLSPM; Wold, 1966, 1973, 1982; Lohmöller, 1989) and generalized structured component analysis (GSCA; Hwang and Takane, 2004, 2014). As their names imply, the two approaches become divergent in how they approximate latent variables in these sub-models. That is, factor-based SEM assumes that common factors may approximate latent variables as in common factor analysis, whereas component-based SEM posits that weighted composites of indicators may serve as proxies for latent variables as in principal component analysis. In this regard, the two approaches are conceptually different, and choices on them would likely be application-dependent and should be theoretically decided in advance, considering how to conceptualize latent variables in the application as well as how to validate the relation between latent variables and their proxies (factors or components; Rigdon, 2012).

Nonetheless, in comparison with factor-based SEM, perhaps the most common criticism of component-based SEM has been that it has no mechanism to formally take into account errors in indicators, which appear practically inevitable in the social sciences (e.g., Bentler and Huang, 2014).

This also leads components to take over these errors to some extent, although extracting a weighted composite from a set of indicators can play a role in reducing the errors implicitly (Gleason, 1973). It is well-known that errors in independent (observed or latent) variables will likely result in biased parameter estimates (e.g., Bollen, 1989, Chapter 5).

To deal with this problem, a bias-correction method, called consistent partial least squares (PLSc; Dijkstra, 2010; Dijkstra and Henseler, 2015), has been proposed in the context of PLSPM. The basic premise of PLSc is that the true measurement model is the so-called basic design (Wold, 1982), representing a unidimensional confirmatory factor analytic model, where each latent variable/factor underlies at least two indicators with each indicator loading on one and only one latent variable. Under this assumption, PLSc begins to apply PLSPM to estimate component weights for a set of indicators per latent variable and then obtains a correction constant for the set of indicators based on their component weights and sample correlations. The loadings for the indicators are estimated by multiplying their component weight estimates by the correction constant. The correlations among latent variables are also estimated using the correction constants for all sets of indicators, which are subsequently used for estimating path coefficients. Conceptually, it is somewhat arbitrary whether PLSc falls into component-based SEM because it has little interest in the specification and estimation of components *per se*, and simply utilizes their weight estimates to obtain parameter estimates of factor-based SEM. In practice, the assumption of the basic design can be restrictive, leading to the exclusion of cross loadings that have been well-accepted in numerous structural equation models (Asparouhov and Muthén, 2009). For example, a classical model involving cross loadings is a multitrait-multimethod model, where trait and method latent variables underlie each indicator (Campbell and Fiske, 1959). Another example is latent growth curve models (Meredith and Tisak, 1990; Duncan et al., 2006), where indicators are typically assumed to load on multiple latent variables, each of which captures a different trajectory of change over time.

To our knowledge, no attempts have been made to incorporate errors in indicators or develop a bias-correction strategy in the context of GSCA. Thus, in this paper, we propose to extend GSCA to explicitly account for errors in indicators. Specifically, we aim to combine a unique part of each indicator into GSCA. As postulated in common factor analysis or factor-based SEM, adding such a unique part may be seen as accounting for measurement error in each indicator. We shall call this proposed extension “GSCA_M,” standing for GSCA with measurement errors incorporated. GSCA_M will provide parameter estimates comparable to those from factor-based SEM. Whereas PLSc involves two separate estimation steps, GSCA_M has a single estimation procedure where a least squares criterion is consistently minimized to estimate all model parameters. In addition, GSCA_M does not require the basic design assumption in model specification and parameter estimation.

The paper is organized as follows. Section Method discusses the technical underpinnings of GSCA_M, including model

specification and parameter estimation. Section Simulation Studies conducts a simulation study to evaluate the performance of GSCA_M and two existing methods—CSA and PLSc. Section An Empirical Application presents an application to show the empirical usefulness of GSCA_M as compared to the existing methods. The final section summarizes the implications of the proposed method.

METHOD

Model

As with GSCA, GSCA_M involves three sub-models—measurement, structural, and weighted relation. The measurement model is used to specify the relationships between indicators and latent variables, whereas the structural model is to specify the relationships among latent variables. The weighted relation model is used to express a latent variable as a weighted composite of indicators. Unlike GSCA, however, GSCA_M contemplates both common and unique parts of each indicator in the measurement model, and expresses a latent variable as a weighted composite of indicators with their unique parts removed in the weighted relation model.

Let $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_J]$ denote an N by J matrix of indicators, where N is the number of observations, and \mathbf{z}_j is the j th indicator ($j = 1, \dots, J$). Let $\mathbf{\Gamma} = [\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_P]$ denote an N by P matrix of latent variables, where $\boldsymbol{\gamma}_p$ is the p th latent variable ($p = 1, \dots, P$). Assume that all indicators and latent variables are normalized such that their lengths are equal to one (i.e., $\mathbf{z}_j' \mathbf{z}_j = \boldsymbol{\gamma}_p' \boldsymbol{\gamma}_p = 1$). Let \mathbf{C} denote a P by J matrix of loadings relating latent variables to indicators. Let \mathbf{U} denote an N by J matrix of unique variables. Let \mathbf{D} denote a J by J diagonal matrix of unique loadings. Let \mathbf{E}_1 denote an N by J matrix of residuals for indicators. Let \mathbf{B} denote a P by P matrix of path coefficients connecting latent variables among themselves, and \mathbf{E}_2 denote an N by P matrix of residuals for latent variables. The three sub-models of GSCA_M are given as follows.

$$\mathbf{Z} = \mathbf{\Gamma C} + \mathbf{UD} + \mathbf{E}_1 \quad (1)$$

$$\mathbf{\Gamma} = \mathbf{\Gamma B} + \mathbf{E}_2 \quad (2)$$

$$\mathbf{\Gamma} = (\mathbf{Z} - \mathbf{UD})\mathbf{W}. \quad (3)$$

In the measurement model (1), $\mathbf{\Gamma C}$ and \mathbf{UD} represent common and unique parts of indicators, respectively. We assume that $\mathbf{\Gamma}$ is uncorrelated with \mathbf{U} ($\mathbf{U}'\mathbf{\Gamma} = \mathbf{\Gamma}'\mathbf{U} = \mathbf{0}$) and \mathbf{U} is orthonormalized ($\mathbf{U}'\mathbf{U} = \mathbf{I}_J$, where \mathbf{I}_J is the identity matrix of order J). In the measurement model, the \mathbf{C} matrix contains fixed values (e.g., zeros) to accommodate hypothesized relationships between indicators and their latent variables as in confirmatory factor analytic models. If this matrix has no fixed values, (1) may be seen as the fixed exploratory factor analytic model (Young, 1941; de Leeuw, 2004, 2008). The structural model (2) remains the same as that for GSCA or the reticular action model (McArdle and McDonald, 1984). The weighted relation model (3) shows that a latent variable is defined as a weighted composite of indicators with their unique parts eliminated.

GSCA_M integrates the sub-models into a single equation, as follows.

$$\begin{aligned} [\mathbf{Z}, \boldsymbol{\Gamma}] &= \boldsymbol{\Gamma} [\mathbf{C}, \mathbf{B}] + [\mathbf{UD}, \mathbf{0}] + [\mathbf{E}_1, \mathbf{E}_2] \\ \boldsymbol{\Psi} &= \boldsymbol{\Gamma} \mathbf{A} + \mathbf{S} + \mathbf{E}, \end{aligned} \quad (4)$$

where $\boldsymbol{\Psi} = [\mathbf{Z}, \boldsymbol{\Gamma}]$, $\mathbf{A} = [\mathbf{C}, \mathbf{B}]$, $\mathbf{S} = [\mathbf{UD}, \mathbf{0}]$, and $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2]$. This is called the GSCA_M model.

Parameter Estimation

The parameters of GSCA_M ($\boldsymbol{\Gamma}$, \mathbf{A} , \mathbf{U} , and \mathbf{D}) are estimated by minimizing the following least squares criterion

$$\phi = SS(\boldsymbol{\Psi} - \boldsymbol{\Gamma} \mathbf{A} - \mathbf{S}), \quad (5)$$

subject to $\boldsymbol{\gamma}_p' \boldsymbol{\gamma}_p = 1$ or equivalently $\text{diag}(\boldsymbol{\Gamma}' \boldsymbol{\Gamma}) = \mathbf{I}_p$, $\mathbf{U}' \boldsymbol{\Gamma} = \mathbf{0}$, and $\mathbf{U}' \mathbf{U} = \mathbf{I}_j$, where $SS(\mathbf{X}) = \text{tr}(\mathbf{X}' \mathbf{X})$, and \mathbf{I}_p is the identity matrix of order p .

A simple iterative algorithm is developed to minimize (5). This algorithm begins by assigning initial values to the parameters. Then, it alternates several steps until convergence, each of which updates a set of parameters in a least squares sense, with the other sets fixed. A detailed description of the algorithm is provided in the Appendix in Supplementary Material.

We can apply GSCA to obtain initial values for $\boldsymbol{\Gamma}$, \mathbf{C} , and \mathbf{B} , although any other initial values can be considered. Then, those for \mathbf{U} and \mathbf{D} may be obtained as described in Steps 3 and 4 in the Appendix in Supplementary Material. We can employ the bootstrap method (Efron, 1979) to estimate the standard errors or confidence intervals of the parameter estimates without resorting to a distributional assumption such as multivariate normality of indicators. When the number of observations is smaller than that of indicators ($N < J$), $\text{rank}(\mathbf{U}) < J$ and the constraint $\mathbf{U}' \mathbf{U} = \mathbf{I}_j$ cannot be fulfilled. In this situation, we may apply Unkel and Trendafilov's (2013) algorithm to update \mathbf{U} , subject to the new constraint $\mathbf{U}' \mathbf{UD} = \mathbf{D}$ (also see Trendafilov and Unkel, 2011). We assume that all indicators and latent variables are normalized, which still results in standardized parameter estimates except for latent variable scores that are normalized. The standardized latent variable scores are obtained by multiplying the normalized scores by \sqrt{N} .

GSCA_M can provide a measure of overall model fit, called FIT. The FIT indicates the total variance of all variables explained by a particular model specification. It is given by

$$\text{FIT} = 1 - \frac{SS(\boldsymbol{\Psi} - \boldsymbol{\Gamma} \mathbf{A} - \mathbf{S})}{SS(\boldsymbol{\Psi})} \quad (6)$$

The values of the FIT range from 0 to 1. The larger this value, the more variance in the variables is accounted for by the specified model. Moreover, it can provide separate model fit measures for the measurement and structural models, as follows.

$$\text{FIT}_M = 1 - \frac{SS(\mathbf{Z} - \boldsymbol{\Gamma} \mathbf{C} - \mathbf{UD})}{SS(\mathbf{Z})}, \quad (7)$$

$$\text{FIT}_S = 1 - \frac{SS(\boldsymbol{\Gamma} - \boldsymbol{\Gamma} \mathbf{B})}{SS(\boldsymbol{\Gamma})} \quad (8)$$

The FIT_M shows how much the variance of indicators is explained by a measurement model, whereas the FIT_S indicates how much the variance of latent variables is accounted for by a structural model. Both measures range from 0 to 1 and can be interpreted in a manner similar to the FIT.

SIMULATION STUDIES

We conducted simulation studies to evaluate the performance of GSCA_M as compared to existing methods, including GSCA, PLSc, and CSA. In particular, we focused on comparing GSCA_M to these methods in parameter recovery.

Simulation Study 1

We specified a structural equation model that consisted of three latent variables and three indicators per latent variable. **Figure 1** displays the specified model along with their unstandardized and standardized parameter values.

For this study, we considered four levels of sample size: $N = 100, 200, 500$, and $1,000$. At each sample size, we generated 1,000 random samples from a multivariate normal distribution with zero means and the covariance matrix implied by the unstandardized parameters of the correct model, based on a CSA formulation. We used the maximum likelihood method for CSA. For PLSc, we used Mode A and the path weighting scheme that is preferred over the other schemes (centroid and factorial) in estimating component weights (Esposito Vinzi et al., 2010). We used the same random seed for all the methods for each sample to have them run with the same initial values.

Table 1 exhibits the relative biases (expressed as percentages), standard deviations, and mean square errors of the standardized loadings and path coefficients estimated from the four methods over the different sample sizes. In the calculations of these properties, we removed any sample involving non-convergence within 300 iterations or convergence to improper solutions. CSA encountered such convergence problems across all the sample sizes. The number of the samples omitted with the convergence problems under CSA was as follows: 66 ($N = 100$), 46 ($N = 200$), 31 ($N = 500$), and 31 ($N = 1,000$). PLSc was faced with the problems only when $N \leq 200$. Specifically, 23 and 2 samples were omitted when $N = 100$ and 200 , respectively. Conversely, GSCA and GSCA_M did not encounter non-convergence or the occurrence of improper solutions across all the sample sizes.

We regarded relative bias $> 10\%$ in absolute value as indicative of an unacceptable degree of bias (e.g., Bollen et al., 2007; Lei and Wu, 2012). As shown in **Table 1**, GSCA provided positively biased loading estimates and negatively biased path coefficient estimates across all the sample sizes. This is expected because the simulated data were generated based on a CSA formulation, assuming that a latent variable was equivalent to a common factor. In this case, component-based approaches to SEM, such as GSCA and PLSPM, are known to overestimate loadings and underestimate path coefficients (e.g., Velicer and Jackson, 1990; Dijkstra, 2010; Hwang et al., 2010; Sarstedt et al., 2016). Conversely, GSCA_M, PLSc, and CSA tended to result in unbiased estimates of both loadings and path coefficients across the sample sizes. When $N = 100$, however, the path coefficient estimates of

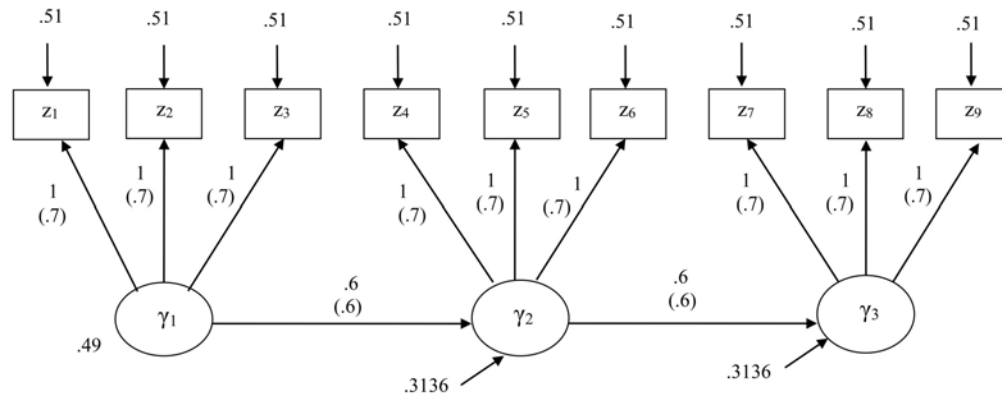


FIGURE 1 | The structural equation model specified for the first simulation study. Standardized parameters are given in parentheses.

GSCA_M showed larger relative biases (4–5%) than those from PLSc and CSA, although they decreased rapidly with the sample size, approaching essentially zero when $N = 1,000$.

The standard deviations of the estimates from the four methods became smaller with the sample size. However, GSCA provided smaller standard deviations than the other methods. This was particularly salient when the sample size was small (e.g., $N = 100$), which is consistent with the literature (Hwang et al., 2010). When the sample size was small, all the methods tended to show similar mean square errors of the loading estimates, whereas GSCA tended to provide larger mean square errors of the path coefficient estimates. As the sample size increased, the mean square errors of all the parameter estimates obtained from GSCA_M, PLSc, and CSA approached zero, while those from GSCA remained slightly larger.

Simulation Study 2

The first simulation study was useful to evaluate how GSCA_M performed as compared to different methods. Nonetheless, this study considered a model with equal loadings, which might be too ideal in reality. Thus, we conducted another simulation study to compare the performance of the methods under a model with the same structure but unequal loadings varying from 0.4 to 0.8. We also compared their performance given a misspecification of the model. **Figure 2** displays both correct and misspecified models along with their unstandardized and standardized parameter values. The misspecified model involved two cross loadings and an additional path coefficient, as indicated by dashed arrows in **Figure 2**.

For both specifications, we considered the same four levels of sample size, at each of which 1,000 random samples were generated from a multivariate normal distribution with zero means and the covariance matrix implied by the unstandardized parameters of the correct model, based on a CSA formulation. Again, we used the maximum likelihood method for CSA, and Mode A and the path weighting scheme for PLSc. We applied all the four methods to estimate the parameters of the correct model, whereas applied only GSCA, GSCA_M, and CSA to estimate the

parameters of the misspecified model because as stated earlier, PLSc was not designed for models involving cross loadings.

Table 2 presents the relative biases (expressed as percentages), standard deviations, and mean square errors of the standardized loadings and path coefficients estimated from the four methods under the correct model specification. GSCA did not encounter non-convergence or the occurrence of improper solutions across all the sample sizes. Conversely, the numbers of the samples omitted with the convergence problems under CSA at the different sample sizes were 72 ($N = 100$), 41 ($N = 200$), 30 ($N = 500$), and 23 ($N = 1000$), whereas those under PLSc were 38 ($N = 100$) and 3 ($N = 200$). One sample was removed under GSCA_M only when $N = 100$.

As expected, all the loading estimates from GSCA except those of two high loadings (0.8) were positively biased, whereas all the path coefficient estimates were negatively biased, regardless of the sample sizes. Conversely, overall, GSCA_M, PLSc, and CSA tended to result in unbiased estimates of a majority of loadings and all path coefficients across the sample sizes. However, the estimates of two low loadings (0.4 or 0.5) under these methods remained similar in magnitude and positively biased even when $N = 1,000$. As in the first simulation study, when $N = 100$, the path coefficient estimates of GSCA_M showed larger relative biases than those from PLSc and CSA, although they decreased rapidly with the sample size, approaching zero when $N = 1,000$.

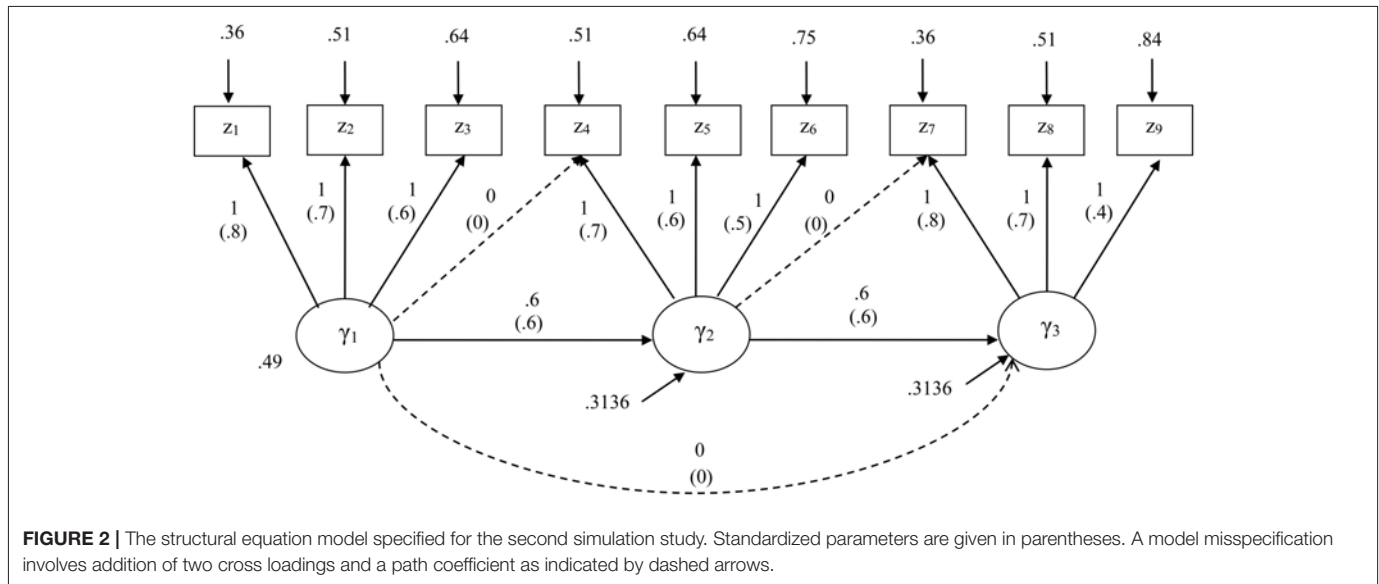
The standard deviations of the estimates from the four methods became smaller with the sample size. GSCA provided smaller standard deviations than the other methods. The mean square errors of all the parameter estimates obtained from GSCA_M, PLSc, and CSA remained similar in magnitude across the sample sizes and most of them, except for those for the estimates of the two loadings, approached zero when the sample size increased. On the other hand, those from GSCA remained larger and only a few approached zero, although they gradually decreased with the sample size.

Table 3 shows the relative biases (expressed as percentages), standard deviations, and mean square errors of the standardized loadings and path coefficients estimated from GSCA, GSCA_M, and CSA under the incorrect model specification. CSA suffered

TABLE 1 | Relative biases expressed as percentages (RB(%)), standard deviations (SD), and mean square errors (MSE) of standardized loadings and path coefficients obtained from GSCA, GSCA_M, PLS_c, and CSA over different sample sizes.

N	Parameters		RB (%)				SD				MSE			
			GSCA	GSCA _M	PLSc	CSA	GSCA	GSCA _M	PLSc	CSA	GSCA	GSCA _M	PLSc	CSA
100	Loadings	0.7	15.77	0.64	-1.11	-0.03	0.04	0.09	0.12	0.08	0.01	0.01	0.01	0.01
		0.7	15.91	1.04	-0.79	-0.14	0.04	0.09	0.11	0.08	0.01	0.01	0.01	0.01
		0.7	15.59	0.33	-1.29	-0.73	0.04	0.09	0.11	0.08	0.01	0.01	0.01	0.01
		0.7	16.00	1.49	-0.33	0.37	0.04	0.09	0.09	0.08	0.01	0.01	0.01	0.01
		0.7	15.80	0.56	-0.09	-0.29	0.04	0.09	0.09	0.07	0.01	0.01	0.01	0.01
		0.7	15.60	1.04	-1.07	-0.54	0.04	0.09	0.09	0.08	0.01	0.01	0.01	0.01
		0.7	15.80	0.34	-0.50	-0.46	0.04	0.09	0.11	0.08	0.01	0.01	0.01	0.01
		0.7	15.97	1.44	-1.19	0.26	0.04	0.09	0.12	0.08	0.01	0.01	0.01	0.01
		0.7	15.93	0.91	-0.64	0.03	0.04	0.09	0.11	0.08	0.01	0.01	0.01	0.01
	Paths	0.6	-25.00	-4.65	0.87	-0.60	0.08	0.10	0.10	0.11	0.03	0.01	0.01	0.01
0.6		-24.52	-4.08	1.42	-0.02	0.08	0.10	0.10	0.10	0.03	0.01	0.01	0.01	
200	Loadings	0.7	15.80	-0.01	-0.36	-0.43	0.03	0.06	0.08	0.05	0.01	0.00	0.01	0.00
		0.7	15.90	0.27	-0.51	-0.10	0.03	0.06	0.08	0.05	0.01	0.00	0.01	0.00
		0.7	16.06	0.80	-0.41	0.17	0.03	0.06	0.08	0.05	0.01	0.00	0.01	0.00
		0.7	15.94	0.24	0.01	-0.07	0.03	0.06	0.08	0.05	0.01	0.00	0.00	0.00
		0.7	15.94	0.60	-0.33	-0.17	0.03	0.06	0.08	0.05	0.01	0.00	0.00	0.00
		0.7	15.99	0.71	-0.31	-0.06	0.03	0.06	0.08	0.05	0.01	0.00	0.00	0.00
		0.7	16.00	0.39	-0.40	0.01	0.03	0.06	0.08	0.05	0.01	0.00	0.01	0.00
		0.7	15.96	0.50	-0.33	-0.03	0.03	0.06	0.08	0.05	0.01	0.00	0.01	0.00
		0.7	15.97	0.47	-0.36	0.01	0.03	0.06	0.08	0.05	0.01	0.00	0.01	0.00
	Paths	0.6	-25.10	-1.82	0.67	-0.18	0.06	0.07	0.07	0.07	0.03	0.01	0.01	0.01
0.6		-24.90	-1.63	0.87	0.18	0.06	0.07	0.07	0.07	0.03	0.01	0.01	0.00	
500	Loadings	0.7	16.03	0.14	-0.03	0.00	0.02	0.04	0.05	0.03	0.01	0.00	0.00	0.00
		0.7	16.01	0.20	-0.30	-0.13	0.02	0.04	0.05	0.03	0.01	0.00	0.00	0.00
		0.7	16.09	0.29	0.01	0.10	0.02	0.04	0.05	0.03	0.01	0.00	0.00	0.00
		0.7	16.06	0.33	-0.09	0.06	0.02	0.04	0.04	0.03	0.01	0.00	0.00	0.00
		0.7	15.96	0.11	-0.19	-0.11	0.02	0.04	0.04	0.03	0.01	0.00	0.00	0.00
		0.7	16.00	0.09	-0.04	-0.06	0.02	0.04	0.04	0.03	0.01	0.00	0.00	0.00
		0.7	16.03	0.17	-0.13	-0.14	0.02	0.04	0.05	0.04	0.01	0.00	0.00	0.00
		0.7	16.07	0.37	-0.34	-0.11	0.02	0.04	0.05	0.06	0.01	0.00	0.00	0.00
		0.7	16.04	0.07	0.21	-0.23	0.02	0.04	0.05	0.06	0.01	0.00	0.00	0.00
	Paths	0.6	-25.73	-1.02	-0.10	-0.35	0.04	0.04	0.05	0.04	0.03	0.00	0.00	0.00
0.6		-25.32	-0.45	-0.45	-0.02	0.04	0.05	0.05	0.06	0.02	0.00	0.00	0.00	
1,000	Loadings	0.7	16.04	0.13	-0.13	0.00	0.01	0.03	0.04	0.02	0.01	0.00	0.00	0.00
		0.7	16.01	0.04	-0.14	-0.13	0.01	0.03	0.03	0.03	0.01	0.00	0.00	0.00
		0.7	16.06	0.07	0.07	0.01	0.01	0.03	0.04	0.02	0.01	0.00	0.00	0.00
		0.7	16.01	0.01	-0.04	-0.03	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00
		0.7	15.94	-0.10	-0.17	-0.13	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00
		0.7	15.97	0.07	-0.21	-0.11	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00
		0.7	16.03	-0.04	0.04	-0.19	0.01	0.03	0.04	0.03	0.01	0.00	0.00	0.00
		0.7	16.14	0.30	0.09	0.20	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00
		0.7	16.07	0.23	-0.09	0.11	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00
	Paths	0.6	-25.60	-0.25	0.20	0.15	0.02	0.03	0.03	0.03	0.02	0.00	0.00	0.00
0.6		-25.67	-0.40	0.07	0.02	0.03	0.03	0.03	0.03	0.02	0.00	0.00	0.00	

A factor-based structural equation model with equal loadings was specified.



severely from non-convergence or convergence to improper solutions across all the sample sizes. The numbers of the samples omitted under CSA were 371 ($N = 100$), 216 ($N = 200$), 131 ($N = 500$), and 100 ($N = 1,000$), whereas those under $GSCA_M$ were 54 ($N = 100$), 12 ($N = 200$), 1 ($N = 500$), and 1 ($N = 1,000$). Again, GSCA had no such problems across all the sample sizes.

Overall, $GSCA_M$ and CSA tended to produce unbiased estimates of most of the loadings and all the path coefficients across the sample sizes. However, their estimates of two low loadings (0.4 or 0.5) remained similar in magnitude and positively biased even when $N = 1,000$. GSCA resulted in positively biased loading estimates except those of two high loadings (0.8) regardless of the sample sizes. Conversely, it produced path coefficient estimates with an acceptably small amount of bias, although the amount of bias on average remained unchanged over the sample sizes.

The standard deviations of the estimates from the three methods became smaller with the sample size. Again, GSCA provided smaller standard deviations than the other methods. The mean square errors of all the parameter estimates obtained from $GSCA_M$ and CSA were similar in magnitude across the sample sizes and most of them, except for those for the estimates of the two loadings, approached zero when the sample size increased. On the other hand, the mean square errors of the loading estimates from GSCA remained larger and only a few approached zero, although they gradually decreased with the sample size. The mean square errors of the path coefficient estimates from GSCA were comparable to those from $GSCA_M$ and CSA across the sample sizes.

To summarize, $GSCA_M$ was found to recover the parameters equally well to CSA in both simulation studies that generated data within the factor-analytic framework. When the basic design held for the specified model, PLSc also performed equally well to $GSCA_M$ and CSA. In general, $GSCA_M$ was less likely to suffer from non-convergence or the occurrence of improper solutions than CSA and PLSc. In particular, CSA tended to

suffer from these problems when the sample size was small and/or the model was misspecified, which was consistent with the literature (e.g., Boomsma, 1982, 1985; Anderson and Gerbing, 1984). Conversely, GSCA largely resulted in biased parameter estimates in these simulation studies that were based on the assumption of factor-analytic models.

AN EMPIRICAL APPLICATION

The present example came from the American customer satisfaction index (ACSI; Fornell et al., 1996) database. The ACSI has been widely used to assess four different levels of customer satisfaction (national-, sector-, industry-, and company-level) in the United States over the past two decades. This example was company-level data collected in 2002 for 152 companies ($N = 152$). **Figure 3** displays the ACSI model. We did not display the residual terms associated with all endogenous variables to make the figure concise. As depicted in **Figure 3**, the ACSI model contains fourteen indicators: z_1 = customer expectations about overall quality, z_2 = customer expectations about reliability, z_3 = customer expectations about customization, z_4 = overall quality, z_5 = reliability, z_6 = customization, z_7 = price given quality, z_8 = quality given price, z_9 = overall customer satisfaction, z_{10} = confirmation of expectations, z_{11} = distance to ideal product or service, z_{12} = formal or informal complaint behavior, z_{13} = repurchase intention, and z_{14} = price tolerance. The measures and scales of the indicators are described in Fornell et al. (1996). This model also involves six latent variables that underlie the 14 indicators, as follows: CE = customer expectations, PQ = perceived quality, PV = perceived value, CS = customer satisfaction, CC = customer complaints, and CL = customer loyalty. The specified relationships in the ACSI model were well-derived from previous theories, and their detailed conceptual derivations can be found in Fornell et al. (1996).

We applied $GSCA_M$, CSA, and PLSc to fit the ACSI model to the data. We used the R packages lavaan (version 0.5-16) (Rosseev,

TABLE 2 | Relative biases expressed as percentages (RB(%)), standard deviations (SD), and mean square errors (MSE) of standardized loadings and path coefficients obtained from GSCA, GSCA_M, PLS_c, and CSA over different sample sizes.

N	Parameters		RB (%)				SD				MSE					
			GSCA	GSCA _M	PLSc	CSA	GSCA	GSCA _M	PLSc	CSA	GSCA	GSCA _M	PLSc	CSA		
100	Loadings	0.8	4.79	-4.43	-6.25	-5.23	0.03	0.09	0.11	0.08	0.00	0.01	0.01	0.01		
		0.7	16.13	0.99	-0.70	-0.04	0.04	0.08	0.11	0.08	0.01	0.01	0.01	0.01		
		0.6	31.38	9.97	8.07	8.98	0.04	0.08	0.12	0.08	0.04	0.01	0.02	0.01		
		0.7	15.97	2.29	-0.63	0.21	0.04	0.11	0.09	0.08	0.01	0.01	0.01	0.01		
		0.6	31.22	10.58	9.72	9.58	0.05	0.11	0.09	0.08	0.04	0.02	0.01	0.01		
		0.5	53.36	27.28	24.24	25.08	0.05	0.11	0.10	0.09	0.07	0.03	0.03	0.02		
		0.8	4.59	-4.58	-6.19	-5.39	0.03	0.09	0.11	0.08	0.00	0.01	0.01	0.01		
	Paths	0.7	16.17	1.69	-0.99	0.26	0.04	0.09	0.12	0.08	0.01	0.01	0.01	0.01		
		0.4	90.18	52.95	50.88	51.53	0.05	0.09	0.13	0.08	0.13	0.05	0.06	0.05		
		0.6	-26.52	-5.73	1.18	-0.35	0.08	0.10	0.11	0.11	0.03	0.01	0.01	0.01		
		0.6	-26.67	-5.40	2.12	0.07	0.08	0.10	0.10	0.10	0.03	0.01	0.01	0.01		
		200	Loadings	0.8	4.78	-4.96	-5.45	-5.59	0.02	0.06	0.07	0.06	0.00	0.01	0.01	0.01
				0.7	16.11	0.24	-0.51	-0.04	0.03	0.06	0.08	0.05	0.01	0.00	0.01	0.00
				0.6	32.03	10.52	9.15	9.92	0.03	0.06	0.08	0.05	0.04	0.01	0.01	0.01
0.7	15.87			0.60	-0.11	-0.04	0.03	0.07	0.07	0.06	0.01	0.01	0.00	0.00		
0.6	31.40			10.62	9.38	9.65	0.03	0.07	0.07	0.06	0.04	0.01	0.01	0.01		
0.5	54.06			26.72	25.28	25.80	0.03	0.07	0.07	0.06	0.07	0.02	0.02	0.02		
0.8	4.73			-4.53	-5.56	-5.05	0.02	0.06	0.08	0.05	0.00	0.01	0.01	0.00		
Paths	0.7	16.16	0.53	-0.39	-0.11	0.03	0.06	0.09	0.06	0.01	0.00	0.01	0.00			
	0.4	90.38	52.38	50.93	51.63	0.04	0.07	0.09	0.06	0.13	0.05	0.05	0.05			
	0.6	-26.65	-2.30	0.72	-0.05	0.06	0.07	0.08	0.08	0.03	0.01	0.01	0.01			
	0.6	-27.15	-2.33	1.05	0.25	0.06	0.07	0.07	0.07	0.03	0.01	0.01	0.00			
500	Loadings	0.8	4.93	-4.95	-5.10	-5.03	0.01	0.04	0.05	0.03	0.00	0.00	0.00	0.00		
		0.7	16.23	0.21	-0.33	-0.10	0.02	0.04	0.05	0.03	0.01	0.00	0.00	0.00		
		0.6	32.10	10.07	9.73	9.88	0.02	0.04	0.05	0.03	0.04	0.00	0.01	0.00		
		0.7	15.99	0.50	-0.13	0.09	0.02	0.04	0.04	0.03	0.01	0.00	0.00	0.00		
		0.6	31.42	9.90	9.53	9.63	0.02	0.04	0.04	0.04	0.04	0.01	0.01	0.00		
		0.5	54.10	25.82	25.66	25.68	0.02	0.04	0.04	0.04	0.07	0.02	0.02	0.02		
		0.8	4.73	-4.89	-5.25	-5.13	0.01	0.04	0.05	0.03	0.00	0.00	0.00	0.00		
		0.7	16.27	0.37	-0.37	0.09	0.02	0.04	0.05	0.03	0.01	0.00	0.00	0.00		
	Paths	0.4	90.60	51.88	52.13	51.73	0.02	0.04	0.06	0.04	0.13	0.04	0.05	0.04		
		0.6	-27.28	-1.18	-0.07	-0.35	0.04	0.05	0.05	0.05	0.03	0.00	0.00	0.00		
		0.6	-27.57	-0.67	0.55	0.22	0.04	0.05	0.05	0.05	0.03	0.00	0.00	0.00		
		1,000	Loadings	0.8	4.94	-4.95	-5.20	-5.06	0.01	0.03	0.03	0.02	0.00	0.00	0.00	0.00
				0.7	16.24	0.04	-0.16	-0.06	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00
				0.6	32.05	9.83	9.83	9.77	0.01	0.03	0.04	0.03	0.04	0.00	0.00	0.00
0.7	15.94			0.07	-0.07	-0.06	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00		
0.6	31.40			9.62	9.53	9.55	0.01	0.03	0.03	0.03	0.04	0.00	0.00	0.00		
0.5	54.06			25.80	25.38	25.52	0.02	0.03	0.03	0.03	0.07	0.02	0.02	0.02		
0.8	4.71			-5.14	-5.06	-5.18	0.01	0.03	0.03	0.02	0.00	0.00	0.00	0.00		
0.7	16.33			0.36	0.07	0.20	0.01	0.03	0.03	0.02	0.01	0.00	0.00	0.00		
Paths	0.4		90.73	52.10	51.55	51.88	0.02	0.03	0.04	0.03	0.13	0.04	0.04	0.04		
	0.6		-27.15	-0.32	0.23	0.10	0.03	0.03	0.03	0.03	0.03	0.00	0.00	0.00		
		0.6	-27.93	-0.47	0.12	0.07	0.03	0.03	0.03	0.03	0.00	0.00	0.00			

A factor-based structural equation model with unequal loadings was specified.

TABLE 3 | Relative biases expressed as percentages (RB(%)), standard deviations (SD), and mean square errors (MSE) of standardized loadings and path coefficients obtained from GSCA, GSCA_M, PLSc, and CSA over different sample sizes.

N	Parameters		RB (%)			SD			MSE			
			GSCA	GSCA _M	CSA	GSCA	GSCA _M	CSA	GSCA	GSCA _M	CSA	
100	Loadings	0.8	2.90	-4.35	-4.89	0.04	0.08	0.07	0.00	0.01	0.01	
		0.7	13.76	0.71	-0.01	0.04	0.08	0.07	0.01	0.01	0.01	
		0.6	28.48	10.13	8.87	0.05	0.08	0.08	0.03	0.01	0.01	
		0.0	12.89	2.08	1.72	0.10	0.13	0.16	0.03	0.02	0.03	
		0.7	2.84	-0.80	-1.73	0.08	0.16	0.14	0.01	0.03	0.02	
		0.6	27.92	11.15	11.73	0.05	0.10	0.08	0.03	0.02	0.01	
		0.5	49.14	27.64	26.32	0.06	0.11	0.08	0.06	0.03	0.02	
		0.0	12.72	2.14	4.01	0.08	0.11	0.15	0.02	0.01	0.02	
		0.8	-4.31	-7.20	-9.75	0.06	0.13	0.13	0.01	0.02	0.02	
		0.7	15.90	2.26	2.67	0.04	0.09	0.08	0.01	0.01	0.01	
		0.4	89.50	53.35	53.88	0.05	0.09	0.09	0.13	0.05	0.05	
		Paths	0.6	-5.18	-6.68	-4.50	0.08	0.11	0.12	0.01	0.01	0.01
			0.0	-1.95	3.99	1.48	0.11	0.16	0.19	0.01	0.03	0.04
			0.6	-0.65	-8.78	-6.22	0.11	0.15	0.18	0.01	0.03	0.03
200	Loadings	0.8	2.80	-5.10	-5.66	0.02	0.06	0.05	0.00	0.01	0.00	
		0.7	13.73	0.17	-0.40	0.03	0.06	0.05	0.01	0.00	0.00	
		0.6	29.08	10.47	9.88	0.03	0.06	0.05	0.03	0.01	0.01	
		0.0	13.40	1.08	0.66	0.07	0.08	0.12	0.02	0.01	0.01	
		0.7	2.69	-0.53	-0.39	0.05	0.11	0.11	0.00	0.01	0.01	
		0.6	27.90	10.60	10.28	0.03	0.07	0.06	0.03	0.01	0.01	
		0.5	49.72	26.62	26.28	0.04	0.07	0.06	0.06	0.02	0.02	
		0.0	12.38	0.62	0.62	0.05	0.08	0.11	0.02	0.01	0.01	
		0.8	-3.60	-5.24	-5.76	0.04	0.10	0.11	0.00	0.01	0.01	
		0.7	15.69	0.63	0.63	0.03	0.06	0.06	0.01	0.00	0.00	
		0.4	89.35	52.40	51.75	0.04	0.06	0.06	0.13	0.05	0.05	
		Paths	0.6	-4.88	-2.32	-1.40	0.06	0.08	0.08	0.00	0.01	0.01
			0.0	-2.21	1.07	0.18	0.08	0.13	0.14	0.01	0.02	0.02
			0.6	-0.38	-3.02	-1.27	0.08	0.12	0.13	0.01	0.01	0.02
500	Loadings	0.8	3.01	-4.98	-4.96	0.02	0.04	0.03	0.00	0.00	0.00	
		0.7	13.84	0.19	-0.14	0.02	0.04	0.03	0.01	0.00	0.00	
		0.6	29.22	10.07	9.93	0.02	0.04	0.03	0.03	0.00	0.00	
		0.0	12.98	0.16	-0.33	0.04	0.05	0.07	0.02	0.00	0.01	
		0.7	3.39	0.29	0.53	0.03	0.07	0.07	0.00	0.00	0.00	
		0.6	27.85	9.90	9.82	0.02	0.04	0.04	0.03	0.01	0.00	
		0.5	49.68	25.82	25.64	0.02	0.04	0.04	0.06	0.02	0.02	
		0.0	12.39	0.13	-0.40	0.03	0.05	0.08	0.02	0.00	0.01	
		0.8	-3.48	-5.08	-4.63	0.03	0.06	0.07	0.00	0.01	0.01	
		0.7	15.77	0.40	0.23	0.02	0.04	0.04	0.01	0.00	0.00	
		0.4	89.53	51.95	52.03	0.02	0.04	0.04	0.13	0.04	0.04	
		Paths	0.6	-5.12	-1.32	-0.43	0.03	0.05	0.05	0.00	0.00	0.00
			0.0	-2.01	0.39	-0.19	0.05	0.08	0.08	0.00	0.01	0.01
			0.6	-0.68	-0.92	0.28	0.05	0.08	0.08	0.00	0.01	0.01
1,000	Loadings	0.8	3.00	-4.98	-5.04	0.01	0.03	0.02	0.00	0.00	0.00	
		0.7	13.87	0.04	-0.07	0.01	0.03	0.02	0.01	0.00	0.00	
		0.6	29.17	9.83	9.72	0.01	0.03	0.03	0.03	0.00	0.00	
		0.0	13.00	0.09	-0.27	0.03	0.04	0.05	0.02	0.00	0.00	

(Continued)

TABLE 3 | Continued

N	Parameters	RB (%)			SD			MSE		
		GSCA	GSCA _M	CSA	GSCA	GSCA _M	CSA	GSCA	GSCA _M	CSA
	0.7	3.37	-0.03	0.37	0.02	0.05	0.05	0.00	0.00	0.00
	0.6	27.78	9.65	9.53	0.02	0.03	0.03	0.03	0.00	0.00
	0.5	49.56	25.70	25.54	0.02	0.03	0.03	0.06	0.02	0.02
	0.0	12.45	0.22	-0.07	0.02	0.03	0.06	0.02	0.00	0.00
	0.8	-3.45	-5.30	-4.98	0.02	0.04	0.05	0.00	0.00	0.00
	0.7	15.81	0.36	0.21	0.01	0.03	0.03	0.01	0.00	0.00
	0.4	89.58	52.13	52.03	0.02	0.03	0.03	0.13	0.04	0.04
Paths	0.6	-4.92	-0.38	0.17	0.02	0.03	0.04	0.00	0.00	0.00
	0.0	-1.83	0.31	0.02	0.03	0.06	0.06	0.00	0.00	0.00
	0.6	-1.12	-0.62	-0.13	0.03	0.06	0.06	0.00	0.00	0.00

A misspecification of a factor-based structural equation model with unequal loadings was considered.

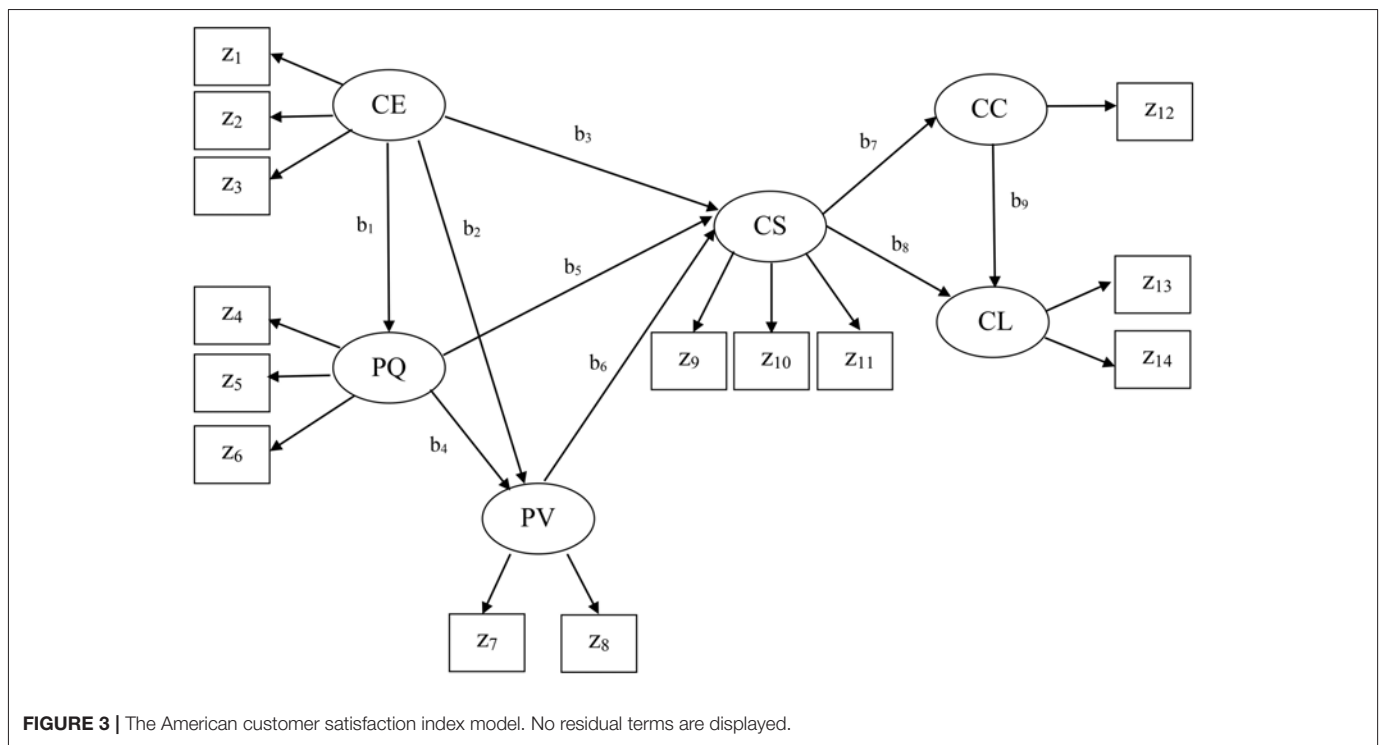


FIGURE 3 | The American customer satisfaction index model. No residual terms are displayed.

2012) to apply CSA and wrote MATLAB codes for GSCA_M and PLSc. As in the simulation study, we utilized maximum likelihood for CSA, and Mode A and the path weighting scheme for PLSc.

Note that in the ACSI model, only a single indicator (z_{12}) loads on customer complaints. As discussed earlier, PLSc was developed based on the basic design requiring at least two indicators per latent variable. When there is only one indicator for a latent variable, PLSc cannot estimate its loading and the path coefficients involving the latent variable because the correction constant for the indicator becomes zero (see Dijkstra and Henseler, 2015). Thus, we used the PLSPM estimate of the component weight for the indicator z_{12} , which was equal to

one, by fixing the correction constant to one. This was also the case when estimating the path coefficients involving customer complaints (b_7 and b_9), indicating that they might be suboptimal estimates of the path coefficients.

Tables 4, 5 provide the estimates of the standardized loadings and path coefficients of the ACSI model obtained from the methods. CSA yielded a number of improper solutions, including the standardized loading and path coefficient estimates greater than one in absolute value. In addition, the loading estimate for z_1 (customer expectations about overall quality) was almost zero. This was inconsistent with that the indicator is expected to be highly and positively related to customer expectations (Fornell et al., 1996). PLSc also produced improper solutions, although it

TABLE 4 | The estimates of standardized loadings of the ACSI model obtained from CSA, PLSc, and GSCA_M.

Latent	Indicator	CSA	PLSc	GSCA _M
CE	Z ₁	0.00	0.91	0.95
	Z ₂	-0.91	0.94	0.97
	Z ₃	-0.94	0.96	0.92
PQ	Z ₄	0.96	0.98	0.97
	Z ₅	0.97	0.97	0.96
	Z ₆	0.94	0.93	0.97
PV	Z ₇	0.93	0.90	0.96
	Z ₈	1.02	1.05	0.99
CS	Z ₉	1.00	0.98	0.99
	Z ₁₀	0.98	0.94	0.99
	Z ₁₁	0.90	0.95	0.92
CC	Z ₁₂	1.00	1.00	1.00
CL	Z ₁₃	0.93	0.94	0.96
	Z ₁₄	1.00	1.00	0.99

TABLE 5 | The estimates of standardized path coefficients of the ACSI model obtained from CSA, PLSc, and GSCA_M.

	CSA	PLSc	GSCA _M
CE → PQ (b ₁)	-0.97	0.95	0.93
CE → PV (b ₂)	1.35	-0.28	-0.11
CE → CS (b ₃)	1.83	-0.21	-0.05
PQ → PV (b ₄)	2.23	1.10	0.93
PQ → CS (b ₅)	2.90	1.01	0.80
PV → CS (b ₆)	-0.13	0.21	0.27
CS → CC (b ₇)	-0.45	0.46	-0.46
CS → CL (b ₈)	0.47	0.51	0.50
CC → CL (b ₉)	-0.47	0.44	-0.45

resulted in fewer than CSA. We have tried different schemes, but continued to have the same problem. Moreover, the positive signs of the path coefficient estimates involving customer complaints (b₇ and b₉) were substantively contradictory, suggesting that more satisfied customers tended to complain more frequently (b₇ = 0.46) and more frequent complainers were likely to be more loyal customers (b₉ = 0.44). These counterintuitive signs were provided, albeit the signs of all the loading estimates remained positive as expected, indicating that the latent variables were not likely to be sign-reversed.

Conversely, GSCA_M resulted in neither improper solutions nor estimates that made little substantive sense. It provided that FIT = 0.85, indicating that the ACSI model accounted for about 85% of the variance of all the variables. Moreover, GSCA_M provided that FIT_M = 0.98 and FIT_S = 0.57. This indicates that the measurement model of the ACSI accounted for about 98% of the variance of the indicators, whereas the structural model explained about 57% of the variance of the latent variables. As also shown in **Table 4**, all the loading estimates were large and

positive. The interpretations of the path coefficient estimates appeared to be generally consistent with those reported in the literature (e.g., Fornell et al., 1996; Anderson and Fornell, 2000). Specifically, customer expectations had a statistically significant impact on perceived quality (b₁ = 0.93, 95% CI = 0.90 ~ 0.95), but had statistically non-significant effects on perceived value (b₂ = -0.11, 95% CI = -0.48 ~ 0.22) and customer satisfaction (b₃ = -0.05, 95% CI = -0.18 ~ 0.05). These non-significant effects were also discussed in previous studies (e.g., Johnson et al., 2001). Perceived quality had statistically significant effects on perceived value (b₄ = 0.93, 95% CI = 0.54 ~ 1.33) and customer satisfaction (b₅ = 0.80, 95% CI = 0.66 ~ 0.98). Perceived value had a statistically significant influence on customer satisfaction (b₆ = 0.27, 95% CI = 0.21 ~ 0.34). Customer satisfaction had statistically significant effects on customer complaints (b₇ = -0.46, 95% CI = -0.63 ~ -0.30) and customer loyalty (b₈ = 0.50, 95% CI = 0.39 ~ 0.61). Customer complaints had a statistically significant effect on customer loyalty (b₉ = -0.45, 95% CI = -0.56 ~ -0.37). We used 100 bootstrap samples for the estimation of the 95% confidence intervals of the GSCA_M estimates.

To summarize, in this application, CSA and PLSc yielded improper solutions that were problematic to interpret. The improper solutions may have occurred for reasons. For example, a few latent variables underlie only two indicators each in the ACSI model, the sample size was relatively small, the correlation between customer expectations and perceived quality, which was equivalent to the standardized path coefficient between them (b₁), was quite large (>|0.90|), or a combination of these issues (e.g., Chen et al., 2001). In addition, both CSA and PLSc provided estimates that were substantively counterintuitive. For the same data, conversely, GSCA_M did not result in improper solutions and its estimates were generally consistent with the hypothesized relationships in the literature.

CONCLUDING REMARKS

We proposed an extension of GSCA, named GSCA_M, to explicitly accommodate errors in indicators. As with GSCA, GSCA_M can be viewed as a component-based approach to SEM in that it still approximates a latent variable by a component. Unlike GSCA, however, GSCA_M considers both common and unique parts of indicators as in factor-based SEM, and estimates a component of indicators with their unique parts excluded. In this way, GSCA_M deals with measurement errors in indicators, yielding parameter estimates comparable to those from factor-based SEM. In addition, it does not require a distributional assumption, such as multivariate normality of indicators, for parameter estimation because it estimates parameters via least squares. As a component-based approach, furthermore, it can avoid factor score indeterminacy (e.g., Guttman, 1955; Schönemann and Wang, 1972), enabling to provide unique latent variable scores.

In the simulation studies, GSCA_M performed equally well to CSA and PLSc in parameter recovery, when the model was correctly specified to satisfy the basic design assumption. Conversely, when the model was misspecified to contain additional cross loadings and path coefficients, only GSCA_M and CSA could be applied to fit the model; and GSCA_M

tended to recover parameters equally to CSA. In the real data application, GSCA_M was the only method that involved no improper solutions.

Although we do not venture into generalizing the results of our analyses, GSCA_M may be a promising alternative to CSA, when researchers have difficulty to address such issues as non-convergence or convergence to improper solutions, or are interested in obtaining unique individual latent variable scores for subsequent analyses or modeling of these scores. This can contribute to widening the scope and applicability of GSCA. Nonetheless, it would be fruitful to apply the proposed method to a wide range of real-world problems to investigate its performance more thoroughly.

REFERENCES

- Anderson, E. W., and Fornell, C. (2000). Foundations of the American customer satisfaction index. *Tot. Qual. Manag.* 11, S869–S882. doi: 10.1080/09544120050135425
- Anderson, J. C., and Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika* 49, 155–173. doi: 10.1007/BF02294170
- Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equat. Model.* 16, 397–438. doi: 10.1080/10705510903008204
- Bentler, P. M., and Huang, W. (2014). On components, latent variables, PLS and simple methods: reactions to Rigdon's rethinking of PLS. *Long Range Plann.* 47, 38–145. doi: 10.1016/j.lrp.2014.02.005
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.
- Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., and Chen, F. (2007). Latent variable models under misspecification: two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociol. Methods Res.* 36, 48–86. doi: 10.1177/0049124107301947
- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models," in *Systems under Indirect Observation: Causality, Structure, Prediction, Part I*, eds K. G. Jöreskog and H. Wold (Amsterdam: North-Holland), 149–173.
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika* 50, 229–242.
- Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., and Kirby, J. B. (2001). Improper solutions in structural equation models: causes, consequences, and strategies. *Sociol. Methods Res.* 29, 468–508. doi: 10.1177/0049124101029004003
- de Leeuw, J. (2004). "Least squares optimal scaling of partially observed linear systems," in *Recent Developments on Structural Equation Models: Theory and Applications*, eds K. van Montfort, J. Oud, and A. Satorra (Dordrecht: Kluwer Academic Publishers), 121–134.
- de Leeuw, J. (2008). *Factor Analysis as Matrix Decomposition*. Preprint Series: Department of Statistics, University of California, Los Angeles, CA.
- Dijkstra, T. K. (2010). "Latent variables and indices: Herman Wold's basic design and partial least squares," in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, eds V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (Berlin: Springer-Verlag), 23–46.
- Dijkstra, T. K., and Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Comput. Stat. Data Anal.* 81, 10–23. doi: 10.1016/j.csda.2014.07.008
- Duncan, T. D., Duncan, S. C., and Strycker, L. A. (2006). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications*. Mahwah NJ: Lawrence Erlbaum Associates.
- Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *Ann. Stat.* 7, 1–26.

AUTHOR CONTRIBUTIONS

HH contributed to conducting all research activities including technical development, empirical analyses, and manuscript writing; YT contributed to technical development and manuscript writing; and KJ contributed to empirical analyses and manuscript writing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02137/full#supplementary-material>

- Esposito Vinzi, V., Trinchera, L., and Amato, S. (2010). "PLS path modeling: from foundations to recent developments and open issues for model assessment and improvement" in *Handbook of Partial Least Squares. Concepts, Methods, and Applications*, eds V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (Berlin: Springer-Verlag), 47–82.
- Fornell, C., and Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *J. Mark. Res.* 19, 440–452.
- Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., and Bryant, B. E. (1996). The American customer satisfaction index: nature, purpose, and findings. *J. Mark.* 60, 7–18.
- Gleason, T. C. (1973). Improving the metric quality of questionnaire data. *Psychometrika* 38, 393–410.
- Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *Br. J. Stat. Psychol.* 8, 65–81.
- Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, M. A., and Hong, S. (2010). A comparative study on parameter recovery of three approaches to structural equation modeling. *J. Mark. Res.* 47, 699–712. doi: 10.1509/jmkr.47.4.699
- Hwang, H., and Takane, Y. (2004). Generalized structured component analysis. *Psychometrika* 69, 81–99. doi: 10.1007/BF02295841
- Hwang, H., and Takane, Y. (2014). *Generalized Structured Component Analysis: A Component-Based Approach to Structural Equation Modeling*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Johnson, M. D., Gustafsson, A., Andreassen, T., Lervik, L., and Cha, J. (2001). The evolution and future of national customer satisfaction index models. *J. Econ. Psychol.* 22, 217–245. doi: 10.1016/S0167-4870(01)00030-7
- Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika* 57, 409–426.
- Jöreskog, K. G. (1973). "A generating method for estimating a linear structural equation system," in *Structural Equation Models in the Social Sciences*, eds A. S. Goldberger and O. D. Duncan (New York, NY: Academic Press), 85–112.
- Jöreskog, K. G., and Wold, H. (1982). "The ML and PLS techniques for modeling with latent variables: historical and comparative aspects," in *Systems under Indirect Observation: Causality, Structure, Prediction, Part I*, eds H. Wold and K. G. Jöreskog (Amsterdam: North-Holland), 263–270.
- Lei, P.-W., and Wu, Q. (2012). "Estimation in structural equation modeling," in *Handbook of Structural Equation Modeling*, ed R. H. Hoyle (New York, NY: The Guilford Press), 164–180.
- Lohmöller, J.-B. (1989). *Latent Variable Path Modeling with Partial Least Squares*. New York, NY: Springer-Verlag.
- McArdle, J. J., and McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *Br. J. Math. Stat. Psychol.* 37, 234–251.
- Meredith, W., and Tisak, J. (1990). Latent curve analysis. *Psychometrika* 55, 107–122.
- Rigdon, E. E. (2012). Rethinking partial least squares path modeling: in praise of simple methods. *Long Range Plan.* 45, 341–358. doi: 10.1016/j.lrp.2012.09.010
- Rossee, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

- Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., and Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: where the bias lies! *J. Bus. Res.* 69, 3998–4010. doi: 10.1016/j.jbusres.2016.06.007
- Schönemann, P. H., and Wang, M.-M. (1972). Some results on factor indeterminacy. *Psychometrika* 37, 61–91.
- ten Berge, J. M. F. (1993). *Least Squares Optimization in Multivariate Analysis*. Leiden: DSWO Press, Leiden University.
- Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Qual. Manag. Bus. Excell.* 19, 871–886. doi: 10.1080/14783360802159543
- Trendafilov, N. T., and Unkel, S. (2011). Exploratory factor analysis of data matrices with more variables than observations. *J. Comput. Graph. Stat.* 20, 874–891. doi: 10.1198/jcgs.2011.09211
- Trendafilov, N. T., Unkel, S., and Krzanowski, W. (2013). Exploratory factor and principal component analyses: some new aspects. *Stat. Comput.* 23, 209–220. doi: 10.1007/s11222-011-9303-7
- Unkel, S., and Trendafilov, N. T. (2013). Zig-zag exploratory factor analysis with more variables than observations. *Comput. Stat.* 28, 107–125. doi: 10.1007/s00180-011-0275-z
- Velicer, W. F., and Jackson, D. N. (1990). Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. *Multivariate Behav. Res.* 25, 1–28.
- Wold, H. (1966). “Estimation of principal components and related methods by iterative least squares,” in *Multivariate Analysis*, ed P. R. Krishnaiah (New York, NY: Academic Press), 391–420.
- Wold, H. (1973). “Nonlinear iterative partial least squares (NIPALS) modeling: some current developments,” in *Multivariate Analysis*, ed P. R. Krishnaiah (New York, NY: Academic Press), 383–487.
- Wold, H. (1982). “Soft modeling: the basic design and some extensions,” in *Systems under Indirect Observations, Part II*, eds K. G. Jöreskog and H. Wold (Amsterdam: North-Holland), 1–54.
- Young, G. (1941). Maximum likelihood estimation and factory analysis. *Psychometrika* 6, 49–53.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hwang, Takane and Jung. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.