Generalized Structured Component Analysis with Uniqueness Terms for Accommodating Measurement Error

Heungsun Hwang[1*], Yoshio Takane[2], Kwanghee Jung[3]

[1]McGill University, Canada, [2]University of Victoria, Canada, [3]Texas Tech University, United States

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

Author contribution statement

HH contributed to conducing 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.

Keywords

generalized structured component analysis, Uniqueness, measurement error, Bias Correction, Structural Equation Modeling

Abstract

Word count: 136

Generalized structured component analysis (GSCA) is a component-based approach to structural equation modeling, 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 GSCAM, 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 GSCAM and existing methods. These methods are also applied to fit a substantively well-established model to real data.

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: No

1

2

3

Generalized Structured Component Analysis with Uniqueness Terms for Accommodating

Measurement Error

Heungsun Hwang

McGill University

Yoshio Takane

University of Victoria

Kwanghee Jung

Texas Tech University

November 8, 2017

Requests for reprints should be sent to: Heungsun Hwang, Department of Psychology, McGill

University, 2001 McGill College Avenue, Montreal, QC H3A 1G1, Canada. Email:

heungsun.hwang@mcgill.ca.

Generalized Structured Component Analysis with Uniqueness Terms for Accommodating Measurement Error

Abstract

Generalized structured component analysis (GSCA) is a component-based approach to structural equation modeling, 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.

1. Introduction

41 Structural equation modeling (SEM) involves the specification and testing of the relationships

42 between variables that are observed (indicators) and unobserved (latent variables). Two

43 approaches have been proposed for SEM: Factor-based vs. component-based (e.g., Fornell &

44 Bookstein, 1982; Jöreskog & Wold, 1982; Rigdon, 2012; Tenenhaus, 2008). The former includes

45 covariance structure analysis (CSA) (Jöreskog, 1970, 1973), and the latter includes partial least

46 squares path modeling (PLSPM) (Lohmöller, 1989; Wold, 1966, 1973, 1982) and generalized

47 structured component analysis (GSCA) (Hwang & Takane, 2004, 2014). As their names imply,

48 the two approaches become divergent in how they approximate latent variables in these sub-

49 models. That is, factor-based SEM assumes that common factors may approximate latent

50 variables as in common factor analysis, whereas component-based SEM posits that weighted

51 composites of indicators may serve as proxies for latent variables as in principal component

52 analysis. In this regard, the two approaches are conceptually different, and choices on them

53 would likely be application-dependent and should be theoretically decided in advance,

54 considering how to conceptualize latent variables in the application as well as how to validate the

55 relation between latent variables and their proxies (factors or components) (Rigdon, 2012).

56 Nonetheless, in comparison with factor-based SEM, perhaps the most common criticism

57 of component-based SEM has been that it has no mechanism to formally take into account errors

58 in indicators, which appear practically inevitable in the social sciences (e.g., Bentler & Huang,

59 2014). This also leads components to take over these errors to some extent, although extracting a

60 weighted composite from a set of indicators can play a role in reducing the errors implicitly

61 (Gleason, 1973). It is well known that errors in independent (observed or latent) variables will

62 likely result in biased parameter estimates (e.g., Bollen, 1989, Chapter 5).

63　　　　　To deal with this problem, a bias-correction method, called consistent partial least

64　　squares (PLSc) (Dijkstra, 2010; Dijkstra & Henseler, 2015), has been proposed in the context of

65　　PLSPM. The basic premise of PLSc is that the true measurement model is the so-called basic

66　　design (Wold, 1982), representing a unidimensional confirmatory factor analytic model, where

67　　each latent variable/factor underlies at least two indicators with each indicator loading on one

68　　and only one latent variable. Under this assumption, PLSc begins to apply PLSPM to estimate

69　　component weights for a set of indicators per latent variable and then obtains a correction

70　　constant for the set of indicators based on their component weights and sample correlations. The

71　　loadings for the indicators are estimated by multiplying their component weight estimates by the

72　　correction constant. The correlations among latent variables are also estimated using the

73　　correction constants for all sets of indicators, which are subsequently used for estimating path

74　　coefficients. Conceptually, it is somewhat arbitrary whether PLSc falls into component-based

75　　SEM because it has little interest in the specification and estimation of components per se, and

76　　simply utilizes their weight estimates to obtain parameter estimates of factor-based SEM. In

77　　practice, the assumption of the basic design can be restrictive, leading to the exclusion of cross

78　　loadings that have been well accepted in numerous structural equation models (Asparouhov &

79　　Muthén, 2009). For example, a classical model involving cross loadings is a multitrait-

80　　multimethod model, where trait and method latent variables underlie each indicator (Campbell &

81　　Fiske, 1959). Another example is latent growth curve models (Meredith & Tisak, 1990; Duncan

82　　et al., 2006), where indicators are typically assumed to load on multiple latent variables, each of

83　　which captures a different trajectory of change over time.

84　　　　　To our knowledge, no attempts have been made to incorporate errors in indicators or

85　　develop a bias-correction strategy in the context of GSCA. Thus, in this paper, we propose to

86   extend GSCA to explicitly account for errors in indicators. Specifically, we aim to combine a

87   unique part of each indicator into GSCA. As postulated in common factor analysis or factor-

88   based SEM, adding such a unique part may be seen as accounting for measurement error in each

89   indicator. We shall call this proposed extension "$GSCA_M$," standing for generalized structured

90   component analysis with measurement errors incorporated. $GSCA_M$ will provide parameter

91   estimates comparable to those from factor-based SEM. Whereas PLSc involves two separate

92   estimation steps, $GSCA_M$ has a single estimation procedure where a least squares criterion is

93   consistently minimized to estimate all model parameters. In addition, $GSCA_M$ does not require

94   the basic design assumption in model specification and parameter estimation.

95   The paper is organized as follows. Section 2 discusses the technical underpinnings of

96    $GSCA_M$, including model specification and parameter estimation. Section 3 conducts a

97   simulation study to evaluate the performance of $GSCA_M$ and two existing methods - CSA and

98   PLSc. Section 4 presents an application to show the empirical usefulness of $GSCA_M$ as

99   compared to the existing methods. The final section summarizes the implications of the proposed

100   method.

2. Method

2.1. Model

103   As with GSCA, $GSCA_M$ involves three sub-models – measurement, structural, and weighted

104   relation. The measurement model is used to specify the relationships between indicators and

105   latent variables, whereas the structural model is to specify the relationships among latent

106   variables. The weighted relation model is used to express a latent variable as a weighted

107   composite of indicators. Unlike GSCA, however, $GSCA_M$ contemplates both common and

108   unique parts of each indicator in the measurement model, and expresses a latent variable as a

109 weighted composite of indicators with their unique parts removed in the weighted relation

110 model.

111   Let $\mathbf{Z} = [\mathbf{z}_1,\dots,\mathbf{z}_J]$ denote an N by J matrix of indicators, where N is the number of

112 observations, and $\mathbf{z}_j$ is the jth indicator ($j = 1,\dots, J$). Let $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1,\dots, \boldsymbol{\gamma}_P]$ denote an N by P matrix

113 of latent variables, where $\boldsymbol{\gamma}_p$ is the pth latent variable ($p = 1,\dots, P$). Assume that all indicators and

114 latent variables are normalized such that their lengths are equal to one (i.e., $\mathbf{z}_j{}'\mathbf{z}_j = \boldsymbol{\gamma}_p{}'\boldsymbol{\gamma}_p = 1$).

115 Let $\mathbf{C}$ denote a P by J matrix of loadings relating latent variables to indicators. Let $\mathbf{U}$ denote an N

116 by J matrix of unique variables. Let $\mathbf{D}$ denote a J by J diagonal matrix of unique loadings. Let $\mathbf{E}_1$

117 denote an N by J matrix of residuals for indicators. Let $\mathbf{B}$ denote a P by P matrix of path

118 coefficients connecting latent variables among themselves, and $\mathbf{E}_2$ denote an N by P matrix of

119 residuals for latent variables. The three sub-models of GSCA$_M$ are given as follows.

120 $$\mathbf{Z} = \boldsymbol{\Gamma}\mathbf{C} + \mathbf{U}\mathbf{D} + \mathbf{E}_1 \tag{1}$$

121 $$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\mathbf{B} + \mathbf{E}_2 \tag{2}$$

122 $$\boldsymbol{\Gamma} = (\mathbf{Z} - \mathbf{U}\mathbf{D})\mathbf{W}. \tag{3}$$

123 In the measurement model (1), $\boldsymbol{\Gamma}\mathbf{C}$ and $\mathbf{U}\mathbf{D}$ represent common and unique parts of indicators,

124 respectively, which are uncorrelated with $\mathbf{E}_1$. We also assume that $\boldsymbol{\Gamma}$ is uncorrelated with $\mathbf{U}$

125 ($\mathbf{U}'\boldsymbol{\Gamma} = \boldsymbol{\Gamma}'\mathbf{U} = \mathbf{0}$) and $\mathbf{U}$ is orthonormalized ($\mathbf{U}'\mathbf{U} = \mathbf{I}_J$, where $\mathbf{I}_J$ is the identity matrix of order J).

126 In the measurement model, the $\mathbf{C}$ matrix contains fixed values (e.g., zeros) to accommodate

127 hypothesized relationships between indicators and their latent variables as in confirmatory factor

128 analytic models. If this matrix has no fixed values, (1) may be seen as the fixed exploratory

129 factor analytic model (de Leeuw, 2004, 2008; Young, 1941). The structural model (2) remains

130 the same as that for GSCA or the reticular action model (McArdle & McDonald, 1984). The

131   weighted relation model (3) shows that a latent variable is defined as a weighted composite of

132   indicators with their unique parts eliminated.

133   GSCA$_M$ integrates the sub-models into a single equation, as follows.

134
$$[\mathbf{Z}, \mathbf{\Gamma}] = \mathbf{\Gamma}[\mathbf{C}, \mathbf{B}] + [\mathbf{UD}, \mathbf{0}] + [\mathbf{E}_1, \mathbf{E}_2]$$
$$\mathbf{\Psi} = \mathbf{\Gamma A} + \mathbf{S} + \mathbf{E},$$
(4)

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

136   **2.2. Parameter Estimation**

137   The parameters of GSCA$_M$ ($\mathbf{\Gamma}$, $\mathbf{A}$, $\mathbf{U}$, and $\mathbf{D}$) are estimated by minimizing the following least

138   squares criterion

139
$$\phi = \mathrm{SS}(\mathbf{\Psi} - \mathbf{\Gamma A} - \mathbf{S}),$$
(5)

140   subject to $\mathbf{\gamma}_p ' \mathbf{\gamma}_p = 1$ or equivalently diag$(\mathbf{\Gamma' \Gamma}) = \mathbf{I}_P$, $\mathbf{U' \Gamma} = \mathbf{0}$, and $\mathbf{U'U} = \mathbf{I}_J$, where SS$(\mathbf{X}) =$

141   tr$(\mathbf{X'X})$, and $\mathbf{I}_P$ is the identity matrix of order P.

142   A simple iterative algorithm is developed to minimize (5). This algorithm begins by

143   assigning initial values to the parameters. Then, it alternates several steps until convergence,

144   each of which updates a set of parameters in a least squares sense, with the other sets fixed. A

145   detailed description of the algorithm is provided in the Appendix.

146   We can apply GSCA to obtain initial values for $\mathbf{\Gamma}$, $\mathbf{C}$, and $\mathbf{B}$, although any other initial

147   values can be considered. Then, those for $\mathbf{U}$ and $\mathbf{D}$ may be obtained as described in Steps 3 and 4

148   in the Appendix. We can employ the bootstrap method (Efron, 1979) to estimate the standard

149   errors or confidence intervals of the parameter estimates without resorting to a distributional

150   assumption such as multivariate normality of indicators. When the number of observations is

151   smaller than that of indicators ($N < J$), rank$(\mathbf{U}) < J$ and the constraint $\mathbf{U'U} = \mathbf{I}_J$ cannot be

152   fulfilled. In this situation, we may apply Unkel and Trendafilov's (2013) algorithm to update $\mathbf{U}$,

153 subject to the new constraint $\mathbf{U'UD} = \mathbf{D}$ (also see Trendafilov & Unkel, 2011). We assume that

154 all indicators and latent variables are normalized, which still results in standardized parameter

155 estimates except for latent variable scores that are normalized. The standardized latent variable

156 scores are obtained by multiplying the normalized scores by $\sqrt{N}$.

157      $GSCA_M$ can provide a measure of overall model fit, called FIT. The FIT indicates the

158 total variance of all variables explained by a particular model specification. It is given by

159
$$\text{FIT} = 1 - \frac{\text{SS}(\mathbf{\Psi} - \mathbf{\Gamma A} - \mathbf{S})}{\text{SS}(\mathbf{\Psi})} . \tag{6}$$

160 The values of the FIT range from 0 to 1. The larger this value, the more variance in the variables

161 is accounted for by the specified model. Moreover, it can provide separate model fit measures for

162 the measurement and structural models, as follows.

163
$$\text{FIT}_M = 1 - \frac{\text{SS}(\mathbf{Z} - \mathbf{\Gamma C} - \mathbf{UD})}{\text{SS}(\mathbf{Z})} , \tag{7}$$

164
$$\text{FIT}_S = 1 - \frac{\text{SS}(\mathbf{\Gamma} - \mathbf{\Gamma B})}{\text{SS}(\mathbf{\Gamma})} . \tag{8}$$

165 The FIT_M shows how much the variance of indicators is explained by a measurement model,

166 whereas the FIT_S indicates how much the variance of latent variables is accounted for by a

167 structural model. Both measures range from 0 to 1 and can be interpreted in a manner similar to

168 the FIT.

169   **3. Simulation Studies**

170 We conducted simulation studies to evaluate the performance of $GSCA_M$ as compared to existing

171 methods, including GSCA, PLSc, and CSA. In particular, we focused on comparing $GSCA_M$ to

172 these methods in parameter recovery.

173 **3.1. Simulation Study 1**

174   We specified a structural equation model that consisted of three latent variables and three

175   indicators per latent variable. Figure 1 displays the specified model along with their

176   unstandardized and standardized parameter values.

177   _____

178   Insert Figure 1 about here

179   _____

180   For this study, we considered four levels of sample size: $N = 100, 200, 500,$ and 1000. At

181   each sample size, we generated 1000 random samples from a multivariate normal distribution

182   with zero means and the covariance matrix implied by the unstandardized parameters of the

183   correct model, based on a CSA formulation. We used the maximum likelihood method for CSA.

184   For PLSc, we used Mode A and the path weighting scheme that is preferred over the other

185   schemes (centroid and factorial) in estimating component weights (Esposito Vinzi et al., 2010).

186   We used the same random seed for all the methods for each sample to have them run with the

187   same initial values.

188   Table 1 exhibits the relative biases (expressed as percentages), standard deviations, and

189   mean square errors of the standardized loadings and path coefficients estimated from the four

190   methods over the different sample sizes. In the calculations of these properties, we removed any

191   sample involving non-convergence within 300 iterations or convergence to improper solutions.

192   CSA encountered such convergence problems across all the sample sizes. The number of the

193   samples omitted with the convergence problems under CSA was as follows: 66 ($N = 100$), 46 (N

194    $= 200$), 31 ($N = 500$), and 31 ($N = 1000$). PLSc was faced with the problems only when $N \leq 200$.

195   Specifically, 23 and 2 samples were omitted when $N = 100$ and 200, respectively. Conversely,

196   GSCA and GSCA$_M$ did not encounter non-convergence or the occurrence of improper solutions

197   across all the sample sizes.

198   We regarded relative bias greater than 10% in absolute value as indicative of an

199   unacceptable degree of bias (e.g., Bollen et al., 2007; Lei & Wu, 2012). As shown in Table 1,

200   GSCA provided positively biased loading estimates and negatively biased path coefficient

201   estimates across all the sample sizes. This is expected because the simulated data were generated

202   based on a CSA formulation, assuming that a latent variable was equivalent to a common factor.

203   In this case, component-based approaches to SEM, such as GSCA and PLSPM, are known to

204   overestimate loadings and underestimate path coefficients (e.g., Dijkstra, 2010; Hwang et al.,

205   2010; Sarstedt et al., 2016; Velicer & Jackson, 1990). Conversely, GSCA$_M$, PLSc, and CSA

206   tended to result in unbiased estimates of both loadings and path coefficients across the sample

207   sizes. When $N = 100$, however, the path coefficient estimates of GSCA$_M$ showed larger relative

208   biases (4 - 5%) than those from PLSc and CSA, although they decreased rapidly with the sample

209   size, approaching essentially zero when $N = 1000$.

210   The standard deviations of the estimates from the four methods became smaller with the

211   sample size. However, GSCA provided smaller standard deviations than the other methods. This

212   was particularly salient when the sample size was small (e.g., $N = 100$), which is consistent with

213   the literature (Hwang et al., 2010). When the sample size was small, all the methods tended to

214   show similar mean square errors of the loading estimates, whereas GSCA tended to provide

215   larger mean square errors of the path coefficient estimates. As the sample size increased, the

216   mean square errors of all the parameter estimates obtained from GSCA$_M$, PLSc, and CSA

217   approached zero, while those from GSCA remained slightly larger.

218   _____

219   Insert Table 1 about here

220   _____

3.2. Simulation Study 2

222   The first simulation study was useful to evaluate how GSCA$_M$ performed as compared to

223   different methods. Nonetheless, this study considered a model with equal loadings, which might

224   be too ideal in reality. Thus, we conducted another simulation study to compare the performance

225   of the methods under a model with the same structure but unequal loadings varying from .4 to .8.

226   We also compared their performance given a misspecification of the model. Figure 2 displays

227   both correct and misspecified models along with their unstandardized and standardized

228   parameter values. The misspecified model involved two cross loadings and an additional path

229   coefficient, as indicated by dashed arrows in Figure 2.

230   _____

231   Insert Figure 2 about here

232   _____

233   For both specifications, we considered the same four levels of sample size, at each of

234   which 1000 random samples were generated from a multivariate normal distribution with zero

235   means and the covariance matrix implied by the unstandardized parameters of the correct model,

236   based on a CSA formulation. Again, we used the maximum likelihood method for CSA, and

237   Mode A and the path weighting scheme for PLSc. We applied all the four methods to estimate

238   the parameters of the correct model, whereas applied only GSCA, GSCA$_M$ and CSA to estimate

239   the parameters of the misspecified model because as stated earlier, PLSc was not designed for

240   models involving cross loadings.

241   Table 2 presents the relative biases (expressed as percentages), standard deviations, and

242   mean square errors of the standardized loadings and path coefficients estimated from the four

243   methods under the correct model specification. GSCA did not encounter non-convergence or the

244   occurrence of improper solutions across all the sample sizes. Conversely, the numbers of the

245   samples omitted with the convergence problems under CSA at the different sample sizes were 72

246   ($N = 100$), 41 ($N = 200$), 30 ($N = 500$), and 23 ($N = 1000$), whereas those under PLSc were 38 (N

247    $= 100$) and 3 ($N = 200$). One sample was removed under GSCA$_M$ only when $N = 100$.

248   As expected, all the loading estimates from GSCA except those of two high loadings (.8)

249   were positively biased, whereas all the path coefficient estimates were negatively biased,

250   regardless of the sample sizes. Conversely, overall, GSCA$_M$, PLSc, and CSA tended to result in

251   unbiased estimates of a majority of loadings and all path coefficients across the sample sizes.

252   However, the estimates of two low loadings (.4 or .5) under these methods remained similar in

253   magnitude and positively biased even when $N = 1000$. As in the first simulation study, when $N =$

254   100, the path coefficient estimates of GSCA$_M$ showed larger relative biases than those from

255   PLSc and CSA, although they decreased rapidly with the sample size, approaching zero when N

256    $= 1000$.

257   The standard deviations of the estimates from the four methods became smaller with the

258   sample size. GSCA provided smaller standard deviations than the other methods. The mean

259   square errors of all the parameter estimates obtained from GSCA$_M$, PLSc, and CSA remained

260   similar in magnitude across the sample sizes and most of them, except for those for the estimates

261   of the two loadings, approached zero when the sample size increased. On the other hand, those

262   from GSCA remained larger and only a few approached zero, although they gradually decreased

263   with the sample size.

267   Table 3 shows the relative biases (expressed as percentages), standard deviations, and

268   mean square errors of the standardized loadings and path coefficients estimated from GSCA,

269    $GSCA_M$ and CSA under the incorrect model specification. CSA suffered severely from non-

270   convergence or convergence to improper solutions across all the sample sizes. The numbers of

271   the samples omitted under CSA were 371 ($N = 100$), 216 ($N = 200$), 131 ($N = 500$), and 100 ($N =$

272   1000), whereas those under $GSCA_M$ were 54 ($N = 100$), 12 ($N = 200$), 1 ($N = 500$), and 1 ($N =$

273   1000). Again, GSCA had no such problems across all the sample sizes.

274   Overall, $GSCA_M$ and CSA tended to produce unbiased estimates of most of the loadings

275   and all the path coefficients across the sample sizes. However, their estimates of two low

276   loadings (.4 or .5) remained similar in magnitude and positively biased even when $N = 1000$.

277   GSCA resulted in positively biased loading estimates except those of two high loadings (.8)

278   regardless of the sample sizes. Conversely, it produced path coefficient estimates with an

279   acceptably small amount of bias, although the amount of bias on average remained unchanged

280   over the sample sizes.

281   The standard deviations of the estimates from the three methods became smaller with the

282   sample size. Again, GSCA provided smaller standard deviations than the other methods. The

283   mean square errors of all the parameter estimates obtained from $GSCA_M$ and CSA were similar

284   in magnitude across the sample sizes and most of them, except for those for the estimates of the

285   two loadings, approached zero when the sample size increased. On the other hand, the mean

286   square errors of the loading estimates from GSCA remained larger and only a few approached

287   zero, although they gradually decreased with the sample size. The mean square errors of the path

288   coefficient estimates from GSCA were comparable to those from GSCA$_M$ and CSA across the

289   sample sizes.

290   _____

291   Insert Table 3 about here

292   _____

293   To summarize, GSCA$_M$ was found to recover the parameters equally well to CSA in both

294   simulation studies that generated data within the factor-analytic framework. When the basic

295   design held for the specified model, PLSc also performed equally well to GSCA$_M$ and CSA. In

296   general, GSCA$_M$ was less likely to suffer from non-convergence or the occurrence of improper

297   solutions than CSA and PLSc. In particular, CSA tended to suffer from these problems when the

298   sample size was small and/or the model was misspecified, which was consistent with the

299   literature (e.g., Anderson & Gerbing, 1984; Boomsma, 1982, 1985). Conversely, GSCA largely

300   resulted in biased parameter estimates in these simulation studies that were based on the

301   assumption of factor-analytic models.

302   **4. An Empirical Application**

303   The present example came from the American customer satisfaction index (ACSI) (Fornell et al.,

304   1996) database. The ACSI has been widely used to assess four different levels of customer

305   satisfaction (national-, sector-, industry-, and company-level) in the United States over the past

306   two decades. This example was company-level data collected in 2002 for 152 companies ($N =$

307   152). Figure 2 displays the ACSI model. We did not display the residual terms associated with

308   all endogenous variables to make the figure concise. As depicted in Figure 2, the ACSI model

309   contains fourteen indicators: z_1 = customer expectations about overall quality, z_2 = customer

310   expectations about reliability, z_3 = customer expectations about customization, z_4 = overall

311   quality, z_5 = reliability, z_6 = customization, z_7 = price given quality, z_8 = quality given price, z_9 =

312   overall customer satisfaction, z_{10} = confirmation of expectations, z_{11} = distance to ideal product

313   or service, z_{12} = formal or informal complaint behaviour, z_{13} = repurchase intention, and z_{14} =

314   price tolerance. The measures and scales of the indicators are described in Fornell et al. (1996).

315   This model also involves six latent variables that underlie the fourteen indicators, as follows: CE

316   = customer expectations, PQ = perceived quality, PV = perceived value, CS = customer

317   satisfaction, CC = customer complaints, and CL = customer loyalty. The specified relationships

318   in the ACSI model were well derived from previous theories, and their detailed conceptual

319   derivations can be found in Fornell et al. (1996).

320   _____

321   Insert Figure 3 about here

322   _____

323   We applied GSCA$_M$, CSA, and PLSc to fit the ACSI model to the data. We used the R

324   packages lavaan (version 0.5-16) (Rosseel, 2012) to apply CSA and wrote MATLAB codes for

325   GSCA$_M$ and PLSc. As in the simulation study, we utilized maximum likelihood for CSA, and

326   Mode A and the path weighting scheme for PLSc.

327   Note that in the ACSI model, only a single indicator (z_{12}) loads on customer complaints.

328   As discussed earlier, PLSc was developed based on the basic design requiring at least two

329   indicators per latent variable. When there is only one indicator for a latent variable, PLSc cannot

330   estimate its loading and the path coefficients involving the latent variable because the correction

331   constant for the indicator becomes zero (see Dijkstra & Henseler, 2015). Thus, we used the

332   PLSPM estimate of the component weight for the indicator z_{12}, which was equal to one, by

333   fixing the correction constant to one. This was also the case when estimating the path

334   coefficients involving customer complaints (b_7 and b_9), indicating that they might be suboptimal

335   estimates of the path coefficients.

336   Tables 4 and 5 provide the estimates of the standardized loadings and path coefficients of

337   the ACSI model obtained from the methods. CSA yielded a number of improper solutions,

338   including the standardized loading and path coefficient estimates greater than one in absolute

339   value. In addition, the loading estimate for z_1 (customer expectations about overall quality) was

340   almost zero. This was inconsistent with that the indicator is expected to be highly and positively

341   related to customer expectations (Fornell et al., 1996). PLSc also produced improper solutions,

342   although it resulted in fewer than CSA. We have tried different schemes, but continued to have

343   the same problem. Moreover, the positive signs of the path coefficient estimates involving

344   customer complaints (b_7 and b_9) were substantively contradictory, suggesting that more satisfied

345   customers tended to complain more frequently ($b_7 = .46$) and more frequent complainers were

346   likely to be more loyal customers ($b_9 = .44$). These counterintuitive signs were provided, albeit

347   the signs of all the loading estimates remained positive as expected, indicating that the latent

348   variables were not likely to be sign-reversed.

349   Conversely, $GSCA_M$ resulted in neither improper solutions nor estimates that made little

350   substantive sense. It provided that FIT = .85, indicating that the ACSI model accounted for about

351   85% of the variance of all the variables. Moreover, $GSCA_M$ provided that $FIT_M = .98$ and FIT_S

352     $= .57$. This indicates that the measurement model of the ACSI accounted for about 98% of the

353   variance of the indicators, whereas the structural model explained about 57% of the variance of

354   the latent variables. As also shown in Table 3, all the loading estimates were large and positive.

355   The interpretations of the path coefficient estimates appeared to be generally consistent with

356 those reported in the literature (e.g., Anderson & Fornell, 2000; Fornell et al., 1996).

357 Specifically, customer expectations had a statistically significant impact on perceived quality (b_1

358 $= .93$, 95% CI $= .90 \sim .95$), but had statistically non-significant effects on perceived value ($b_2 =$

359 $-.11$, 95% CI $= -.48 \sim .22$) and customer satisfaction ($b_3 = -.05$, 95% CI $= -.18 \sim .05$). These non-

360 significant effects were also discussed in previous studies (e.g., Johnson et al., 2001). Perceived

361 quality had statistically significant effects on perceived value ($b_4 = .93$, 95% CI $= .54 \sim 1.33$) and

362 customer satisfaction ($b_5 = .80$, 95% CI $= .66 \sim .98$). Perceived value had a statistically

363 significant influence on customer satisfaction ($b_6 = .27$, 95% CI $= .21 \sim .34$). Customer

364 satisfaction had statistically significant effects on customer complaints ($b_7 = -.46$, 95% CI $= -.63$

365 $\sim -.30$) and customer loyalty ($b_8 = .50$, 95% CI $= .39 \sim .61$). Customer complaints had a

366 statistically significant effect on customer loyalty ($b_9 = -.45$, 95% CI $= -.56 \sim -.37$). We used 100

367 bootstrap samples for the estimation of the 95% confidence intervals of the $GSCA_M$ estimates.

368 _____

369 Insert Tables 4 and 5 about here

370 _____

371 To summarize, in this application, CSA and PLSc yielded improper solutions that were

372 problematic to interpret. The improper solutions may have occurred for reasons. For example, a

373 few latent variables underlie only two indicators each in the ACSI model, the sample size was

374 relatively small, the correlation between customer expectations and perceived quality, which was

375 equivalent to the standardized path coefficient between them (b_1), was quite large ($> |.90|$), or a

376 combination of these issues (e.g., Chen et al., 2001). In addition, both CSA and PLSc provided

377 estimates that were substantively counterintuitive. For the same data, conversely, $GSCA_M$ did

378 not result in improper solutions and its estimates were generally consistent with the hypothesized

379 relationships in the literature.

5. Concluding Remarks

381 We proposed an extension of GSCA, named GSCA$_M$, to explicitly accommodate errors in

382 indicators. As with GSCA, GSCA$_M$ can be viewed as a component-based approach to SEM in

383 that it still approximates a latent variable by a component. Unlike GSCA, however, GSCA$_M$

384 considers both common and unique parts of indicators as in factor-based SEM, and estimates a

385 component of indicators with their unique parts excluded. In this way, GSCA$_M$ deals with

386 measurement errors in indicators, yielding parameter estimates comparable to those from factor-

387 based SEM. In addition, it does not require a distributional assumption, such as multivariate

388 normality of indicators, for parameter estimation because it estimates parameters via least

389 squares. As a component-based approach, furthermore, it can avoid factor score indeterminacy

390 (e.g., Guttman, 1955; Schönemann & Wang, 1972), enabling to provide unique latent variable

391 scores.

392   In the simulation studies, GSCA$_M$ performed equally well to CSA and PLSc in parameter

393 recovery, when the model was correctly specified to satisfy the basic design assumption.

394 Conversely, when the model was misspecified to contain additional cross loadings and path

395 coefficients, only GSCA$_M$ and CSA could be applied to fit the model; and GSCA$_M$ tended to

396 recover parameters equally to CSA. In the real data application, GSCA$_M$ was the only method

397 that involved no improper solutions.

398   Although we do not venture into generalizing the results of our analyses, GSCA$_M$ may be

399 a promising alternative to CSA, when researchers have difficulty to address such issues as non-

400 convergence or convergence to improper solutions, or are interested in obtaining unique

401   individual latent variable scores for subsequent analyses or modeling of these scores. This can

402   contribute to widening the scope and applicability of GSCA. Nonetheless, it would be fruitful to

403   apply the proposed method to a wide range of real-world problems to investigate its performance

404   more thoroughly.

405

Appendix

The proposed algorithm repeats the following steps until convergence.

<u>Step 1</u>: Update $\mathbf{\Gamma}$ (or equivalently $\mathbf{W}$) for fixed $\mathbf{A}$, $\mathbf{U}$, and $\mathbf{D}$. We can re-express (5) as follows.

$$
\begin{aligned}
\phi &= \mathrm{SS}\big([\mathbf{Z}, \mathbf{\Gamma}] - \mathbf{\Gamma}\mathbf{A} - [\mathbf{UD}, \mathbf{0}]\big) \\
&= \mathrm{SS}\big([\mathbf{Z} - \mathbf{UD}, \mathbf{0}] + \mathbf{\Gamma}[\mathbf{0}, \mathbf{I}_P] - \mathbf{\Gamma}\mathbf{A}\big) \\
&= \mathrm{SS}\big((\mathbf{Z} - \mathbf{UD})[\mathbf{I}_J, \mathbf{0}] - \mathbf{\Gamma}[\mathbf{A} - [\mathbf{0}, \mathbf{I}_P]]\big) \\
&= \mathrm{SS}\big((\mathbf{Z} - \mathbf{UD})\mathbf{M} - \mathbf{\Gamma}\mathbf{H}\big),
\end{aligned}
\tag{A1}
$$

where $\mathbf{H} = \mathbf{A} - [\mathbf{0}, \mathbf{I}_P]$, and $\mathbf{M} = [\mathbf{I}_J, \mathbf{0}]$. The least squares estimate of $\mathbf{\Gamma}$ is obtained by

$$
\hat{\mathbf{\Gamma}} = (\mathbf{Z} - \mathbf{UD})\mathbf{M}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1},
\tag{A2}
$$

indicating $\hat{\mathbf{W}} = \mathbf{M}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}$. Each latent variable is updated by $\hat{\mathbf{\gamma}}_p = \mathbf{Z}_p^* \mathbf{w}_p$, where $\mathbf{Z}_p^*$ includes

the columns of $\mathbf{Z} - \mathbf{UD}$ corresponding to a set of indicators associated with the pth latent

variable, and $\mathbf{w}_p$ contains the elements of the pth column of $\mathbf{W}$ assigned to $\mathbf{Z}_p^*$. Subsequently, it

is normalized to satisfy $\hat{\mathbf{\gamma}}_p' \hat{\mathbf{\gamma}}_p = 1$. This can be done by dividing $\mathbf{w}_p$ by $\sqrt{\mathbf{w}_p' \mathbf{Z}_p^{*'} \mathbf{Z}_p^* \mathbf{w}_p}$.

<u>Step 2:</u> Update $\mathbf{A}$ for fixed $\mathbf{\Gamma}$, $\mathbf{U}$, and $\mathbf{D}$. This is equivalent to minimizing

$$
\begin{aligned}
\phi &= \mathrm{SS}\big(\mathbf{\Psi} - \mathbf{S} - \mathbf{\Gamma}\mathbf{A}\big) \\
&= \mathrm{SS}\big(\mathrm{vec}(\mathbf{\Psi}^*) - (\mathbf{I}_{J+P} \otimes \mathbf{\Gamma})\mathrm{vec}(\mathbf{A})\big),
\end{aligned}
\tag{A3}
$$

where $\mathbf{\Psi}^* = \mathbf{\Psi} - \mathbf{S}$. Let $\mathbf{a}$ denote the vector of free elements in $\mathrm{vec}(\mathbf{A})$. Let $\mathbf{\Phi}$ denote the matrix of

the columns of $\mathbf{I} \otimes \mathbf{\Gamma}$ corresponding to the free elements in $\mathrm{vec}(\mathbf{A})$. Then, the least squares

estimate of $\mathbf{a}$ is obtained by

$$
\hat{\mathbf{a}} = (\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathrm{vec}(\mathbf{\Psi}^*).
\tag{A4}
$$

The updated $\mathbf{A}$ is reconstructed from $\hat{\mathbf{a}}$.

<u>Step 3:</u> Update $\mathbf{U}$ for fixed $\mathbf{\Gamma}$, $\mathbf{A}$, and $\mathbf{D}$. This is equivalent to minimizing

$$\begin{aligned}
\phi &= \mathrm{SS}(\boldsymbol{\Psi} - \boldsymbol{\Gamma}\mathbf{A} - \mathbf{S}) \\
&= \mathrm{SS}([\mathbf{Z} - \mathbf{UD}, (\mathbf{Z} - \mathbf{UD})\mathbf{W}] - (\mathbf{Z} - \mathbf{UD})\mathbf{WA}), \\
&= \mathrm{SS}((\mathbf{Z} - \mathbf{UD})(\mathbf{V} - \mathbf{WA}))
\end{aligned} \tag{A5}$$

425 with respect to $\mathbf{U}$, subject to $\mathbf{U}'\boldsymbol{\Gamma} = \mathbf{0}$ and $\mathbf{U}'\mathbf{U} = \mathbf{I}_J$, where $\mathbf{V} = [\mathbf{I}, \mathbf{W}]$. We can update $\mathbf{U}$ based on

426 Trendafilov et al.'s (2013) procedure. Let $\boldsymbol{\Gamma} = \mathbf{QR} = \left[\tilde{\boldsymbol{\Gamma}}, \boldsymbol{\Gamma}_\perp\right]\begin{bmatrix}\mathbf{I}\\\mathbf{O}\end{bmatrix}$ denote the QR decomposition

427 of $\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}_\perp$ is an N by $N - P$ orthonormal basis matrix of the null space of $\boldsymbol{\Gamma}$. Minimizing

428 (A5) is equivalent to minimizing

$$\begin{aligned}
\phi^* &= \mathrm{SS}(\mathbf{Q}'(\mathbf{Z} - \mathbf{UD})(\mathbf{V} - \mathbf{WA})) \\
&= \mathrm{SS}\left(\begin{bmatrix}\tilde{\boldsymbol{\Gamma}}'(\mathbf{Z} - \mathbf{UD})(\mathbf{V} - \mathbf{WA}) \\ \boldsymbol{\Gamma}_\perp'(\mathbf{Z} - \mathbf{UD})(\mathbf{V} - \mathbf{WA})\end{bmatrix}\right) \\
&= \mathrm{SS}\left(\begin{bmatrix}\tilde{\boldsymbol{\Gamma}}'\mathbf{Z}(\mathbf{V} - \mathbf{WA}) \\ \boldsymbol{\Gamma}_\perp'(\mathbf{Z} - \mathbf{UD})(\mathbf{V} - \mathbf{WA})\end{bmatrix}\right) \\
&= \mathrm{tr}(\mathbf{V}'\mathbf{Z}'\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Gamma}}'\mathbf{Z}\mathbf{V}) + \mathrm{tr}(\mathbf{Z}'\boldsymbol{\Gamma}_\perp\boldsymbol{\Gamma}_\perp'\mathbf{Z}) + \mathrm{tr}(\mathbf{A}'\mathbf{W}'\mathbf{Z}'\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Gamma}}'\mathbf{Z}\mathbf{WA}) + \mathrm{tr}(\mathbf{D}^2) \\
&\quad - 2\mathrm{tr}(\mathbf{V}'\mathbf{Z}'\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Gamma}}'\mathbf{Z}\mathbf{WA}) - 2\mathrm{tr}(\mathbf{U}'\boldsymbol{\Gamma}_\perp\boldsymbol{\Gamma}_\perp'\mathbf{Z}\mathbf{D}).
\end{aligned} \tag{A6}$$

430 Let us reparametrize $\mathbf{U}$ as $\mathbf{U} = \boldsymbol{\Gamma}_\perp\tilde{\mathbf{U}}$, where $\tilde{\mathbf{U}}$ is an $N - P$ by J orthonormal matrix with $\tilde{\mathbf{U}}'\tilde{\mathbf{U}} = \mathbf{I}_J$,

431 thereby keeping the constraint $\mathbf{U}'\mathbf{U} = \mathbf{I}_J$ satisfied. Then, updating $\mathbf{U}$ reduces to updating $\tilde{\mathbf{U}}$ by

432 maximizing $\mathrm{tr}(\mathbf{U}'\boldsymbol{\Gamma}_\perp\boldsymbol{\Gamma}_\perp'\mathbf{Z}\mathbf{D}) = \mathrm{tr}(\tilde{\mathbf{U}}'\boldsymbol{\Gamma}_\perp'\mathbf{Z}\mathbf{D})$ in (A6), subject to $\tilde{\mathbf{U}}'\tilde{\mathbf{U}} = \mathbf{I}_J$. This problem can be

433 solved as follows: Let $\mathrm{SVD}(\boldsymbol{\Gamma}_\perp'\mathbf{Z}\mathbf{D}) = \mathbf{P}\boldsymbol{\Delta}\mathbf{T}'$. Then, $\hat{\tilde{\mathbf{U}}} = \mathbf{PT}'$ (e.g., ten Berge, 1993, p. 34).

434 Finally, $\mathbf{U}$ is updated by $\hat{\mathbf{U}} = \boldsymbol{\Gamma}_\perp\hat{\tilde{\mathbf{U}}}$.

435 <u>Step 4:</u> Update $\mathbf{D}$ for fixed $\boldsymbol{\Gamma}$, $\mathbf{A}$, and $\mathbf{U}$. Minimizing (5) with respect to $\mathbf{D}$ is equivalent to

436 minimizing

$$\phi = \mathrm{SS}(\mathbf{Z}(\mathbf{V} - \mathbf{WA}) - \mathbf{UD}(\mathbf{V} - \mathbf{WA})). \tag{A7}$$

438 Then, the least squares estimate of $\mathbf{D}$ is obtained by

439
$$\hat{\mathbf{D}} = \text{diag}(\mathbf{U'Z}).$$
(A8)

440

References

441

442 Anderson, E. W., & Fornell, C. (2000). Foundations of the American customer satisfaction

443   index. *Total Quality Management*, 11, S869-S882.

444 Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence,

445   improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory

446   factor analysis. *Psychometrika*, 49, 155-173.

447 Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural*

448   *Equation Modeling*, 16, 397- 438.

449 Bentler, P. M., & Huang, W. (2014). On components, latent variables, PLS and simple methods:

450   Reactions to Rigdon's rethinking of PLS. *Long Range Planning*, 47, 38-145.

451 Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

452 Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable

453   models under misspecification: two-stage least squares (2SLS) and maximum likelihood

454   (ML) estimators. *Sociological Methods & Research,* 36, 48-86.

455 Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis

456   models. In K.G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation:*

457   *Causality, structure, prediction, part I* (pp. 149–173). Amsterdam: North-Holland.

458 Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL

459   maximum likelihood estimation. *Psychometrika*, 50, 2, 229-242.

460 Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the

461   multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

462 Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in

463   structural equation models: Causes, consequences, and strategies. *Sociological Methods*

464   *and Research*, 29, 468-508.

465 de Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In K. van

466   Montfort, J. Oud, & A. Satorra (Eds.), *Recent developments on structural equation*

467   *models: Theory and applications* (pp. 121-134). Dordrecht: Kluwer Academic Publishers.

468 de Leeuw, J. (2008). Factor analysis as matrix decomposition. Preprint series: Department of

469   Statistics, University of California, Los Angeles.

470 Dijkstra, T. K. (2010). Latent variables and indices: Herman Wold's basic design and partial

471   least squares. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.),

472   *Handbook of partial least squares: Concepts, methods and applications* (pp. 23-46).

473   Berlin: Springer-Verlag.

474 Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for

475   linear structural equations. *Computational Statistics and Data Analysis*, 81, 10-23.

476 Duncan, T. D., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable*

477   *growth curve modeling: Concepts, issues, and applications*. Mahwah NJ: Lawrence

478   Erlbaum Associates.

479 Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, 7, 1-26.

480 Esposito Vinzi, V., Trinchera, L., & Amato, S. (2010). PLS path modeling: from foundations to

481   recent developments and open issues for model assessment and improvement. In V.

482   Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (eds.), *Handbook of partial least*

483   *squares. Concepts, methods, and applications* (pp. 47-82). Berlin: Springer-Verlag.

484   Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS

485   applied to consumer exit-voice theory. *Journal of Marketing Research*, 19, 440-452.

486   Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American

487   customer satisfaction index: Nature, purpose, and findings. *Journal of Marketing*, 60, 7-

488   18.

489   Gleason, T. C. (1973). Improving the metric quality of questionnaire data. *Psychometrika*, 38,

490   393-410.

491   Guttman, L. (1955). The determinacy of factor score matrices with implications for five other

492   basic problems of common-factor theory. *British Journal of Statistical Psychology*, 8, 65-

493   81.

494   Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, M. A., & Hong, S. (2010). A comparative study

495   on parameter recovery of three approaches to structural equation modeling. *Journal of*

496   *Marketing Research*, 47, 699-712.

497   Hwang, H., & Takane, Y. (2004), Generalized structured component analysis. *Psychometrika*,

498   69, 81-99.

499   Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-*

500   *based approach to structural equation modeling*. Boca Raton, FL: Chapman & Hall/CRC

501   Press.

502   Johnson, M. D., Gustafsson, A., Andreassen, T., Lervik, L., & Cha, J. (2001). The evolution and

503   future of national customer satisfaction index models. *Journal of Economic Psychology*,

504   22, 217-245.

505   Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57,

506   409-426.

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

508   In A. S. Goldberger and O. D. Duncan (Eds.), *Structural equation models in the social*

509   *sciences* (pp. 85-112). New York: Academic Press.

510   Jöreskog, K. G. & Wold, H. (1982). The ML and PLS techniques for modeling with latent

511   variables: Historical and comparative Aspects. In H. Wold & K. G. Jöreskog (Eds.),

512   *Systems under Indirect observation: causality, structure, prediction, part I* (pp. 263-270).

513   Amsterdam: North-Holland.

514   Lei, P.-W., & Wu, Q. (2012). Estimation in structural equation modeling. In R. H. Hoyle (Ed.),

515   *Handbook of structural equation modeling*. New York: The Guilford Press.

516   Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. New York:

517   Springer-Verlag.

518   McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action

519   model for moment structures. *British Journal of Mathematical and Statistical*

520   *Psychology*, 37, 234-251.

521   Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122.

522   Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple

523   methods. *Long Range Planning*, 45, 341-358.

524   Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of*

525   *Statistical Software*, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/

526   Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation

527   issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, 69,

528   3998-4010.

529 Schönemann, P. H., & Wang, M.-M. (1972). Some results on factor indeterminacy.

530   *Psychometrika*, 37, 61-91.

531 ten Berge, J. M. F. (1993). *Least squares optimization in multivariate analysis*. Leiden: DSWO

532   Press, Leiden University.

533 Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Quality*

534   *Management and Business Excellence*, 19, 871-886.

535 Tenenhaus, M., Amato, S., & Esposito Vinzi, V. (2004). A global goodness-of-fit index for PLS

536   structural equation modelling. *Proceedings of the XLII SIS Scientific Meeting*, Vol.

537   Contributed Papers, CLEUP, Padova, 739–742.

538 Trendafilov, N.T., & Unkel, S. (2011). Exploratory factor analysis of data matrices with more

539   variables than observations. *Journal of Computational Graphical Statistics*, 20, 874-

540   891.

541 Trendafilov, N.T., Unkel, S., & Krzanowski, W. (2013). Exploratory factor and principal

542   component analyses: some new aspects. *Statistics and Computing*, 23, 209–220.

543 Unkel, S., & Trendafilov, N. T. (2013). Zig-zag exploratory factor analysis with more variables

544   than observations. *Computational Statistics*, 28, 107-125.

545 Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis:

546   Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25,

547   1-28.

548 Wold, H. (1965). A fixed-point theorem with econometric background, I-II. *Arkiv for Matematik*,

549   6, 209-240.

550 Wold, H. (1966). Estimation of principal components and related methods by iterative least

551   squares. In P. R. Krishnaiah (ed.), *Multivariate analysis* (pp. 391-420). New York:

552   Academic Press.

553   Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) modeling: Some

554   current developments. In P. R. Krishnaiah (ed.), *Multivariate analysis* (pp. 383-487).

555   New York: Academic Press.

556   Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog &

557   H. Wold (Eds.), *Systems under indirect observations, part II* (pp. 1-54). Amsterdam:

558   North-Holland.

559   Young, G. (1941). Maximum likelihood estimation and factory analysis. *Psychometrika*, 6, 49-
560   53.
561

562 Table 1. Relative biases expressed as percentages (RB(%)), standard deviations (SD), and mean square errors (MSE) of standardized
563 loadings and path coefficients obtained from GSCA, GSCA$_M$, PLSc, and CSA over different sample sizes. A factor-based structural
564 equation model with equal loadings was specified.
565

N	Parameters		RB (%)				SD				MSE			
			GSCA	GSCA$_M$	PLSc	CSA	GSCA	GSCA$_M$	PLSc	CSA	GSCA	GSCA$_M$	PLSc	CSA
100	loadings	.7	15.77	.64	-1.11	-.03	.04	.09	.12	.08	.01	.01	.01	.01
		.7	15.91	1.04	-.79	-.14	.04	.09	.11	.08	.01	.01	.01	.01
		.7	15.59	.33	-1.29	-.73	.04	.09	.11	.08	.01	.01	.01	.01
		.7	16.00	1.49	-.33	.37	.04	.09	.09	.08	.01	.01	.01	.01
		.7	15.80	.56	-.09	-.29	.04	.09	.09	.07	.01	.01	.01	.01
		.7	15.60	1.04	-1.07	-.54	.04	.09	.09	.08	.01	.01	.01	.01
		.7	15.80	.34	-.50	-.46	.04	.09	.11	.08	.01	.01	.01	.01
		.7	15.97	1.44	-1.19	.26	.04	.09	.12	.08	.01	.01	.01	.01
		.7	15.93	.91	-.64	.03	.04	.09	.11	.08	.01	.01	.01	.01
	Paths	.6	-25.00	-4.65	.87	-.60	.08	.10	.10	.11	.03	.01	.01	.01
		.6	-24.52	-4.08	1.42	-.02	.08	.10	.10	.10	.03	.01	.01	.01
200	loadings	.7	15.80	-.01	-.36	-.43	.03	.06	.08	.05	.01	.00	.01	.00
		.7	15.90	.27	-.51	-.10	.03	.06	.08	.05	.01	.00	.01	.00
		.7	16.06	.80	-.41	.17	.03	.06	.08	.05	.01	.00	.01	.00
		.7	15.94	.24	.01	-.07	.03	.06	.08	.05	.01	.00	.00	.00
		.7	15.94	.60	-.33	-.17	.03	.06	.08	.05	.01	.00	.00	.00
		.7	15.99	.71	-.31	-.06	.03	.06	.08	.05	.01	.00	.00	.00
		.7	16.00	.39	-.40	.01	.03	.06	.08	.05	.01	.00	.01	.00
		.7	15.96	.50	-.33	-.03	.03	.06	.08	.05	.01	.00	.01	.00
		.7	15.97	.47	-.36	.01	.03	.06	.08	.05	.01	.00	.01	.00
	Paths	.6	-25.10	-1.82	.67	-.18	.06	.07	.07	.07	.03	.01	.01	.01
		.6	-24.90	-1.63	.87	.18	.06	.07	.07	.07	.03	.01	.01	.00
500	loadings	.7	16.03	.14	-.03	.00	.02	.04	.05	.03	.01	.00	.00	.00
		.7	16.01	.20	-.30	-.13	.02	.04	.05	.03	.01	.00	.00	.00
		.7	16.09	.29	.01	.10	.02	.04	.05	.03	.01	.00	.00	.00
		.7	16.06	.33	-.09	.06	.02	.04	.04	.03	.01	.00	.00	.00
		.7	15.96	.11	-.19	-.11	.02	.04	.04	.03	.01	.00	.00	.00

		.7	16.00	.09	-.04	-.06	.02	.04	.04	.03	.01	.00	.00	.00
		.7	16.03	.17	-.13	-.14	.02	.04	.05	.04	.01	.00	.00	.00
		.7	16.07	.37	-.34	-.11	.02	.04	.05	.06	.01	.00	.00	.00
		.7	16.04	.07	.21	-.23	.02	.04	.05	.06	.01	.00	.00	.00
	Paths	.6	-25.73	-1.02	-.10	-.35	.04	.04	.05	.04	.03	.00	.00	.00
		.6	-25.32	-.45	-.45	-.02	.04	.05	.05	.06	.02	.00	.00	.00
1000	loadings	.7	16.04	.13	-.13	.00	.01	.03	.04	.02	.01	.00	.00	.00
		.7	16.01	.04	-.14	-.13	.01	.03	.03	.03	.01	.00	.00	.00
		.7	16.06	.07	.07	.01	.01	.03	.04	.02	.01	.00	.00	.00
		.7	16.01	.01	-.04	-.03	.01	.03	.03	.02	.01	.00	.00	.00
		.7	15.94	-.10	-.17	-.13	.01	.03	.03	.02	.01	.00	.00	.00
		.7	15.97	.07	-.21	-.11	.01	.03	.03	.02	.01	.00	.00	.00
		.7	16.03	-.04	.04	-.19	.01	.03	.04	.03	.01	.00	.00	.00
		.7	16.14	.30	.09	.20	.01	.03	.03	.02	.01	.00	.00	.00
		.7	16.07	.23	-.09	.11	.01	.03	.03	.02	.01	.00	.00	.00
	Paths	.6	-25.60	-.25	.20	.15	.02	.03	.03	.03	.02	.00	.00	.00
		.6	-25.67	-.40	.07	.02	.03	.03	.03	.03	.02	.00	.00	.00

566
567
568

569 Table 2. Relative biases expressed as percentages (RB(%)), standard deviations (SD), and mean square errors (MSE) of standardized
570 loadings and path coefficients obtained from GSCA, GSCA$_M$, PLSc, and CSA over different sample sizes. A factor-based structural
571 equation model with unequal loadings was specified.
572

N	Parameters		RB (%)				SD				MSE			
			GSCA	GSCA$_M$	PLSc	CSA	GSCA	GSCA$_M$	PLSc	CSA	GSCA	GSCA$_M$	PLSc	CSA
100	loadings	.8	4.79	-4.43	-6.25	-5.23	.03	.09	.11	.08	.00	.01	.01	.01
		.7	16.13	.99	-.70	-.04	.04	.08	.11	.08	.01	.01	.01	.01
		.6	31.38	9.97	8.07	8.98	.04	.08	.12	.08	.04	.01	.02	.01
		.7	15.97	2.29	-.63	.21	.04	.11	.09	.08	.01	.01	.01	.01
		.6	31.22	10.58	9.72	9.58	.05	.11	.09	.08	.04	.02	.01	.01
		.5	53.36	27.28	24.24	25.08	.05	.11	.10	.09	.07	.03	.03	.02
		.8	4.59	-4.58	-6.19	-5.39	.03	.09	.11	.08	.00	.01	.01	.01
		.7	16.17	1.69	-.99	.26	.04	.09	.12	.08	.01	.01	.01	.01
		.4	90.18	52.95	50.88	51.53	.05	.09	.13	.08	.13	.05	.06	.05
	Paths	.6	-26.52	-5.73	1.18	-.35	.08	.10	.11	.11	.03	.01	.01	.01
		.6	-26.67	-5.40	2.12	.07	.08	.10	.10	.10	.03	.01	.01	.01
200	loadings	.8	4.78	-4.96	-5.45	-5.59	.02	.06	.07	.06	.00	.01	.01	.01
		.7	16.11	.24	-.51	-.04	.03	.06	.08	.05	.01	.00	.01	.00
		.6	32.03	10.52	9.15	9.92	.03	.06	.08	.05	.04	.01	.01	.01
		.7	15.87	.60	-.11	-.04	.03	.07	.07	.06	.01	.01	.00	.00
		.6	31.40	10.62	9.38	9.65	.03	.07	.07	.06	.04	.01	.01	.01
		.5	54.06	26.72	25.28	25.80	.03	.07	.07	.06	.07	.02	.02	.02
		.8	4.73	-4.53	-5.56	-5.05	.02	.06	.08	.05	.00	.01	.01	.00
		.7	16.16	.53	-.39	-.11	.03	.06	.09	.06	.01	.00	.01	.00
		.4	90.38	52.38	50.93	51.63	.04	.07	.09	.06	.13	.05	.05	.05
	Paths	.6	-26.65	-2.30	.72	-.05	.06	.07	.08	.08	.03	.01	.01	.01
		.6	-27.15	-2.33	1.05	.25	.06	.07	.07	.07	.03	.01	.01	.00
500	loadings	.8	4.93	-4.95	-5.10	-5.03	.01	.04	.05	.03	.00	.00	.00	.00
		.7	16.23	.21	-0.33	-.10	.02	.04	.05	.03	.01	.00	.00	.00
		.6	32.10	10.07	9.73	9.88	.02	.04	.05	.03	.04	.00	.01	.00
		.7	15.99	.50	-0.13	.09	.02	.04	.04	.03	.01	.00	.00	.00
		.6	31.42	9.90	9.53	9.63	.02	.04	.04	.04	.04	.01	.01	.00

		.5	54.10	25.82	25.66	25.68	.02	.04	.04	.04	.07	.02	.02	.02
		.8	4.73	-4.89	-5.25	-5.13	.01	.04	.05	.03	.00	.00	.00	.00
		.7	16.27	0.37	-.37	.09	.02	.04	.05	.03	.01	.00	.00	.00
		.4	90.60	51.88	52.13	51.73	.02	.04	.06	.04	.13	.04	.05	.04
	Paths	.6	-27.28	-1.18	-.07	-.35	.04	.05	.05	.05	.03	.00	.00	.00
		.6	-27.57	-.67	.55	.22	.04	.05	.05	.05	.03	.00	.00	.00
1000	loadings	.8	4.94	-4.95	-5.20	-5.06	.01	.03	.03	.02	.00	.00	.00	.00
		.7	16.24	.04	-.16	-.06	.01	.03	.03	.02	.01	.00	.00	.00
		.6	32.05	9.83	9.83	9.77	.01	.03	.04	.03	.04	.00	.00	.00
		.7	15.94	.07	-.07	-.06	.01	.03	.03	.02	.01	.00	.00	.00
		.6	31.40	9.62	9.53	9.55	.01	.03	.03	.03	.04	.00	.00	.00
		.5	54.06	25.80	25.38	25.52	.02	.03	.03	.03	.07	.02	.02	.02
		.8	4.71	-5.14	-5.06	-5.18	.01	.03	.03	.02	.00	.00	.00	.00
		.7	16.33	.36	.07	.20	.01	.03	.03	.02	.01	.00	.00	.00
		.4	90.73	52.10	51.55	51.88	.02	.03	.04	.03	.13	.04	.04	.04
	Paths	.6	-27.15	-.32	.23	.10	.03	.03	.03	.03	.03	.00	.00	.00
		.6	-27.93	-.47	.12	.07	.03	.03	.03	.03	.03	.00	.00	.00

573
574
575

576 Table 3. Relative biases expressed as percentages (RB(%)), standard deviations (SD), and mean square errors (MSE) of standardized
577 loadings and path coefficients obtained from GSCA, GSCA$_M$, PLSc, and CSA over different sample sizes. A misspecification of a
578 factor-based structural equation model with unequal loadings was considered.
579

N	Parameters		RB (%)			SD			MSE		
			GSCA	GSCA$_M$	CSA	GSCA	GSCA$_M$	CSA	GSCA	GSCA$_M$	CSA
100	loadings	.8	2.90	-4.35	-4.89	.04	.08	.07	.00	.01	.01
		.7	13.76	.71	-0.01	.04	.08	.07	.01	.01	.01
		.6	28.48	10.13	8.87	.05	.08	.08	.03	.01	.01
		0	12.89	2.08	1.72	.10	.13	.16	.03	.02	.03
		.7	2.84	-.80	-1.73	.08	.16	.14	.01	.03	.02
		.6	27.92	11.15	11.73	.05	.10	.08	.03	.02	.01
		.5	49.14	27.64	26.32	.06	.11	.08	.06	.03	.02
		0	12.72	2.14	4.01	.08	.11	.15	.02	.01	.02
		.8	-4.31	-7.20	-9.75	.06	.13	.13	.01	.02	.02
		.7	15.90	2.26	2.67	.04	.09	.08	.01	.01	.01
		.4	89.50	53.35	53.88	.05	.09	.09	.13	.05	.05
	Paths	.6	-5.18	-6.68	-4.50	.08	.11	.12	.01	.01	.01
		0	-1.95	3.99	1.48	.11	.16	.19	.01	.03	.04
		.6	-0.65	-8.78	-6.22	.11	.15	.18	.01	.03	.03
200	loadings	.8	2.80	-5.10	-5.66	.02	.06	.05	.00	.01	.00
		.7	13.73	.17	-.40	.03	.06	.05	.01	.00	.00
		.6	29.08	10.47	9.88	.03	.06	.05	.03	.01	.01
		0	13.40	1.08	.66	.07	.08	.12	.02	.01	.01
		.7	2.69	-0.53	-.39	.05	.11	.11	.00	.01	.01
		.6	27.90	10.60	10.28	.03	.07	.06	.03	.01	.01
		.5	49.72	26.62	26.28	.04	.07	.06	.06	.02	.02
		0	12.38	0.62	.62	.05	.08	.11	.02	.01	.01
		.8	-3.60	-5.24	-5.76	.04	.10	.11	.00	.01	.01
		.7	15.69	0.63	.63	.03	.06	.06	.01	.00	.00
		.4	89.35	52.40	51.75	.04	.06	.06	.13	.05	.05
	Paths	.6	-4.88	-2.32	-1.40	.06	.08	.08	.00	.01	.01
		0	-2.21	1.07	.18	.08	.13	.14	.01	.02	.02

		.6	-.38	-3.02	-1.27	.08	.12	.13	.01	.01	.02
500	loadings	.8	3.01	-4.98	-4.96	.02	.04	.03	.00	.00	.00
		.7	13.84	.19	-.14	.02	.04	.03	.01	.00	.00
		.6	29.22	10.07	9.93	.02	.04	.03	.03	.00	.00
		0	12.98	.16	-.33	.04	.05	.07	.02	.00	.01
		.7	3.39	.29	.53	.03	.07	.07	.00	.00	.00
		.6	27.85	9.90	9.82	.02	.04	.04	.03	.01	.00
		.5	49.68	25.82	25.64	.02	.04	.04	.06	.02	.02
		0	12.39	.13	-.40	.03	.05	.08	.02	.00	.01
		.8	-3.48	-5.08	-4.63	.03	.06	.07	.00	.01	.01
		.7	15.77	.40	.23	.02	.04	.04	.01	.00	.00
		.4	89.53	51.95	52.03	.02	.04	.04	.13	.04	.04
	Paths	.6	-5.12	-1.32	-.43	.03	.05	.05	.00	.00	.00
		0	-2.01	.39	-.19	.05	.08	.08	.00	.01	.01
		.6	-.68	-.92	.28	.05	.08	.08	.00	.01	.01
1000	loadings	.8	3.00	-4.98	-5.04	.01	.03	.02	.00	.00	.00
		.7	13.87	.04	-.07	.01	.03	.02	.01	.00	.00
		.6	29.17	9.83	9.72	.01	.03	.03	.03	.00	.00
		0	13.00	.09	-.27	.03	.04	.05	.02	.00	.00
		.7	3.37	-.03	.37	.02	.05	.05	.00	.00	.00
		.6	27.78	9.65	9.53	.02	.03	.03	.03	.00	.00
		.5	49.56	25.70	25.54	.02	.03	.03	.06	.02	.02
		0	12.45	.22	-.07	.02	.03	.06	.02	.00	.00
		.8	-3.45	-5.30	-4.98	.02	.04	.05	.00	.00	.00
		.7	15.81	.36	.21	.01	.03	.03	.01	.00	.00
		.4	89.58	52.13	52.03	.02	.03	.03	.13	.04	.04
	Paths	.6	-4.92	-.38	.17	.02	.03	.04	.00	.00	.00
		0	-1.83	.31	.02	.03	.06	.06	.00	.00	.00
		.6	-1.12	-.62	-.13	.03	.06	.06	.00	.00	.00

580

33

581 Table 4. The estimates of standardized loadings of the ACSI model obtained from CSA, PLSc,
582 and GSCA$_M$.

583

Latent	Indicator	CSA	PLSc	GSCA$_M$
CE	z_1	.00	.91	.95
	z_2	-.91	.94	.97
	z_3	-.94	.96	.92
PQ	z_4	.96	.98	.97
	z_5	.97	.97	.96
	z_6	.94	.93	.97
PV	z_7	.93	.90	.96
	z_8	1.02	1.05	.99
CS	z_9	1.00	.98	.99
	z_{10}	.98	.94	.99
	z_{11}	.90	.95	.92
CC	z_{12}	1.00	1.00	1.00
CL	z_{13}	.93	.94	.96
	z_{14}	1.00	1.00	.99

584

585

586

587

588   Table 5. The estimates of standardized path coefficients of the ACSI model obtained from CSA,
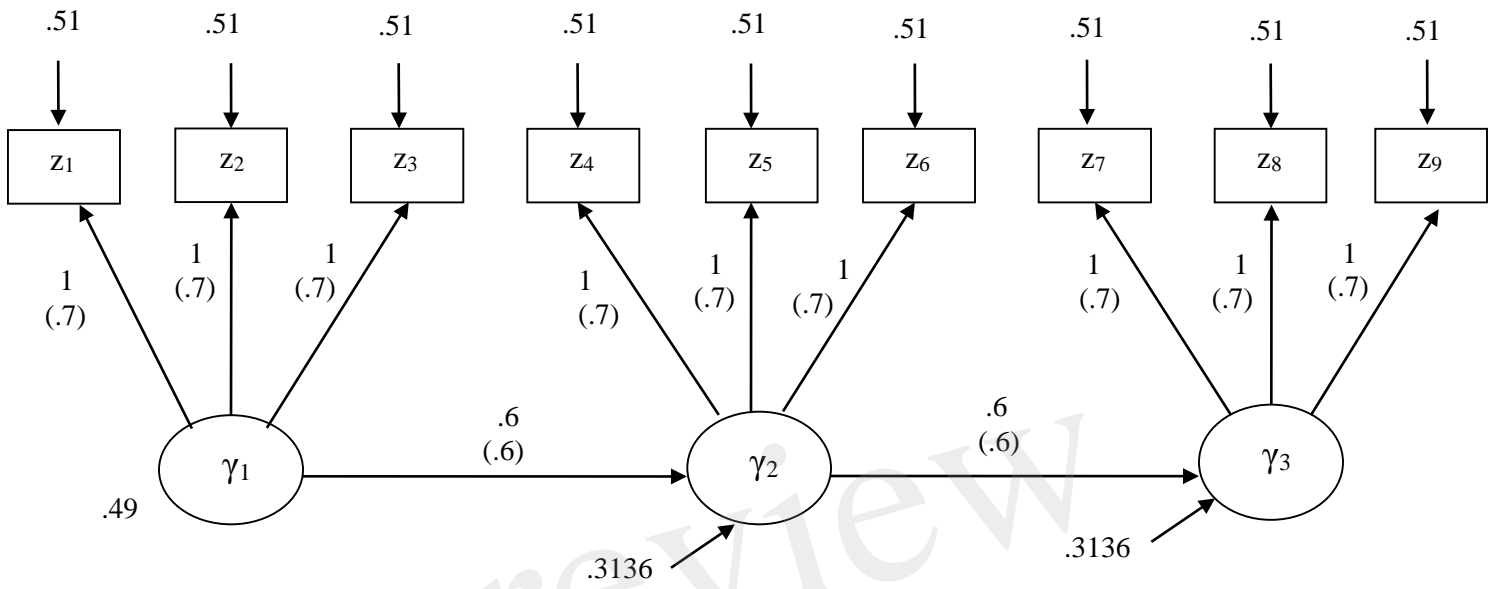589   PLSc, and GSCA$_M$.

590

	CSA	PLSc	GSCA$_M$
CE $\rightarrow$ PQ (b_1)	-.97	.95	.93
CE $\rightarrow$ PV (b_2)	1.35	-.28	-.11
CE $\rightarrow$ CS (b_3)	1.83	-.21	-.05
PQ $\rightarrow$ PV (b_4)	2.23	1.10	.93
PQ $\rightarrow$ CS (b_5)	2.90	1.01	.80
PV $\rightarrow$ CS (b_6)	-.13	.21	.27
CS $\rightarrow$ CC (b_7)	-.45	.46	-.46
CS $\rightarrow$ CL (b_8)	.47	.51	.50
CC $\rightarrow$ CL (b_9)	-.47	.44	-.45

591

592

593

594   Figure 1. The structural equation model specified for the first simulation study. Standardized
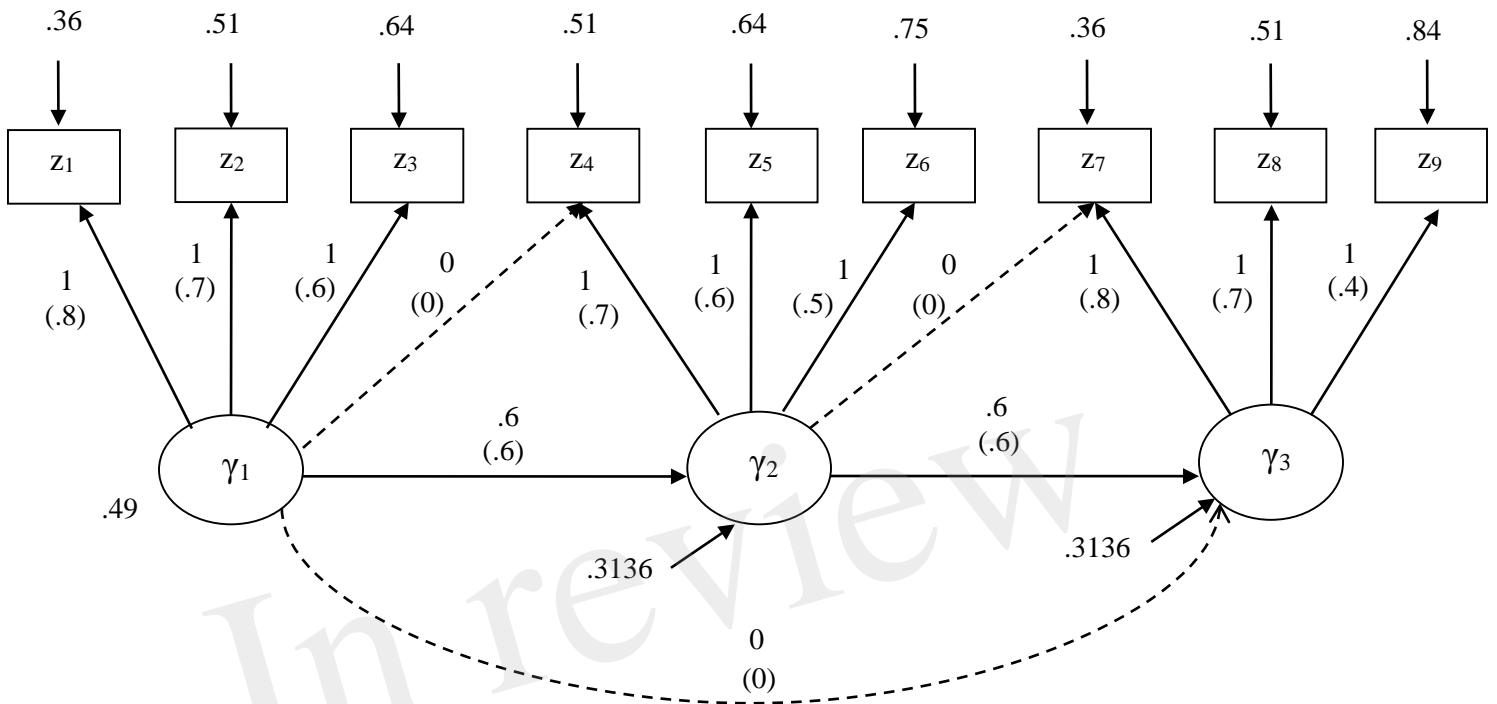595   parameters are given in parentheses.

596

597

598

599

Figure 2. The structural equation model specified for the second simulation study. Standardized

parameters are given in parentheses. A model misspecification involves addition of two cross

loadings and a path coefficient as indicated by dashed arrows.

603

604

605

606

607

608

609

610 z_1

611 z_2 CE

612

613 z_3 b_3 CC z_{12}

614   CS b_8

615 z_4 b_1 b_2 b_5 b_7 b_9

616 z_{13}

617 z_5 PQ b_6 z_9 z_{10} z_{11} CL

618 z_6 z_{14}

619   b_4 PV

620

621

622 z_7 z_8
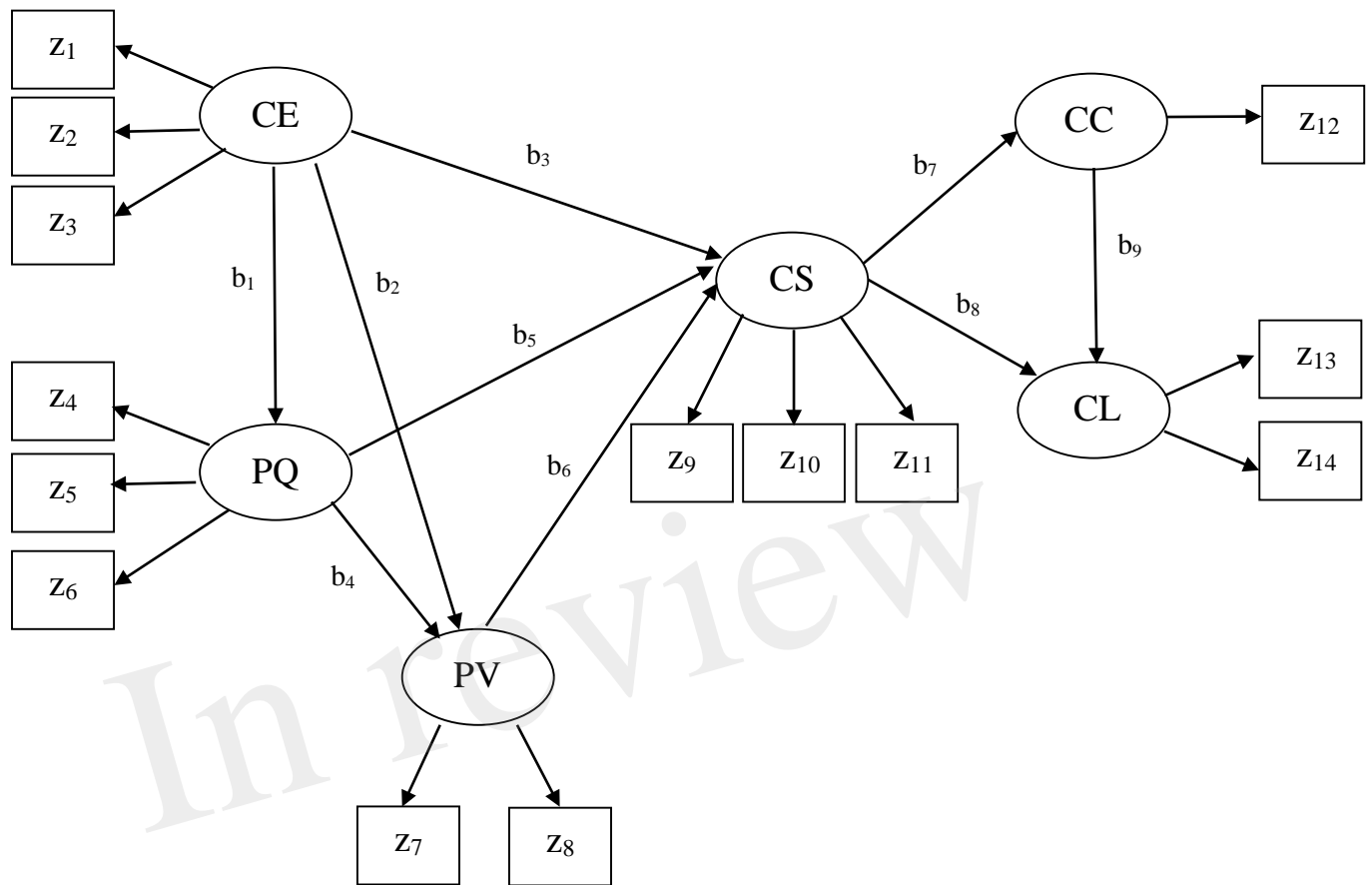
623

624

625

626

627

628 Figure 3. The American customer satisfaction index model. No residual terms are displayed.

38