# Nonlinear Generalized Structured Component Analysis

**Heungsun Hwang**

McGill University

**Yoshio Takane**

McGill University

**November 10, 2009**

The work reported in this paper was supported by Grant 290439 and Grant 10630 from the Natural Sciences and Engineering Research Council of Canada to the first and second authors, respectively. The authors thank two anonymous reviewers for their comments that helped improve the quality of the manuscript. Requests for reprints should be sent to: Heungsun Hwang, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada. Email: heungsun.hwang@mcgill.ca.

## Abstract

Generalized Structured Component Analysis (GSCA) represents component-based structural equation modeling. Currently, GSCA is geared only for the analysis of quantitative data. In this paper, GSCA is extended to deal with qualitative data through data transformation. In particular, the optimal scaling approach is adopted for data transformation as it can be readily coupled with the GSCA estimation procedure. An alternating least-squares algorithm is developed that involves two phases for estimation of model and data parameters. Two empirical applications are presented to demonstrate the usefulness of the proposed method.

Keywords: Generalized structured component analysis, qualitative data, optimal scaling, alternating least squares.

## 1. Introduction

Structural Equation Modeling (SEM) is employed for the specification and analysis of interdependencies among observed variables and underlying theoretical constructs, often called latent variables. Generalized Structured Component Analysis (GSCA) (Hwang & Takane, 2004) was recently proposed as a *component-based* approach to SEM, in which latent variables are analogous to components or weighted composites of observed variables (Tenenhaus, 2008).

GSCA can be of use for substantive researchers and practitioners in psychology and various scientific disciplines for many reasons. First, GSCA does not require the assumption of multivariate normality of observed variables, which is usually violated in practice (Micceri, 1989). Second, it is shown to perform well in small samples (Hwang, Malhotra, Kim, Tomiuk, & Hong, in press), which appear practically inevitable in many studies. Third, GSCA does not yield inadmissible solutions such as negative variance estimates and correlations greater than one in absolute value. Inadmissible solutions are often difficult to circumvent effectively and obscure the interpretation of results. Fourth, this approach results in unique estimates of latent variable scores which can be used for various subsequent analyses. Finally, it enables the provision of overall model-fit measures for theory testing and model comparison.

A Monte Carlo simulation study was recently conducted to evaluate the relative performance of GSCA with respect to two extant approaches to SEM (i.e., covariance structure analysis and partial least squares) under a variety of experimental conditions (Hwang et al., in press). The results of the Monte Carlo analysis provide guidelines for substantive researchers with respect to the conditions under which GSCA is to be preferred over the two extant approaches. Specifically, it is suggested that GSCA be used in lieu of partial least squares for general SEM purposes. Moreover, GSCA is recommended over covariance structure analysis unless the researcher is assured that his/her model is correctly specified. This study supplies additional compelling evidence as to why GSCA would be considered a viable alternative approach to SEM.

Over the past several years, GSCA has been rapidly extended to enhance its data-analytic capability and generality. For instance, GSCA can readily handle constrained multi-group comparisons and higher-order latent variables (Hwang & Takane, 2004). It has been combined with fuzzy clustering simultaneously to accommodate cluster-level heterogeneity of respondents (Hwang, DeSarbo, & Takane, 2007). Moreover, multilevel GSCA was proposed to take into account the hierarchical structures of both observed and latent variables (Hwang, Takane, & Malhotra, 2007). Furthermore, GSCA has been extended for dealing with multicollinearity of both observed and latent variables (Hwang, 2009a). Recently, GSCA has been generalized to investigate various types of interaction effects among latent variables (Hwang, Ho, & Lee, 2009). In addition, GSCA was lately implemented into a user-friendly, web-based software program, named *GeSCA* (Hwang, 2009b) to facilitate the wide adoption of the approach by SEM users.

Despite these theoretical and empirical developments to date, GSCA methodology is still geared only for the analysis of quantitative data. As the collection of qualitative data abounds in psychology and other disciplines, it would be necessary to extend GSCA to analyze qualitative data as well. Thus, the objective of this paper is to propose a nonlinear version of GSCA which deals with qualitative data effectively. In particular, the proposed method seeks to handle qualitative data through a certain type of data transformation called *optimal scaling* (cf. Bock, 1960; Gifi, 1990; Young, 1981). Optimal scaling generally represents the procedure of *quantifying qualitative data* through assigning numerical values to qualitative data such that the association between the data and the model is maximized while strictly maintaining the measurement characteristics of the data (Bock, 1960). Accordingly, if there is a technique geared for the analysis of quantitative data, the same technique can also be utilized for the analysis of qualitative data because the qualitative data become quantitative through this procedure. As will be shown later, optimal scaling is particularly attractive in the context of GSCA because it is well coupled with the least-squares estimation procedure of GSCA.

The paper is organized as follows. In Section 2, the technical underpinnings of GSCA are briefly discussed to facilitate the derivation of the proposed method. In Section 3, the proposed nonlinear extension of GSCA is discussed in full detail. A detailed description of the estimation procedures for this method is provided. In Section 4, two applications are presented to illustrate the empirical usefulness of the proposed method. The final section briefly summarizes and discusses the implications of the proposed method.

## 2. Generalized Structured Component Analysis

Let $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N]'$ denote an *N* by *J* matrix of observed variables, where $\mathbf{z}_i$ is a *J* by 1 vector for a single respondent *i* ($i = 1, \cdots, N$). In *linear* GSCA, $\mathbf{z}_i$ is quantitative, and typically standardized to have zero mean and unit variance. As stated earlier, GSCA defines latent variables as weighted composited of observed variables as follows:

$$\boldsymbol{\eta}_i = \mathbf{W}'\mathbf{z}_i, \tag{1}$$

where $\boldsymbol{\eta}_i$ is a $T$ by 1 vector of latent variables for respondent $i$, and $\mathbf{W}$ is a $J$ by $T$ matrix of component weights for observed variables. In GSCA, the measurement model is given by:

$$\mathbf{z}_i = \mathbf{C}'\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

where $\mathbf{C}$ is a $T$ by $J$ matrix of loadings for observed variables, and $\boldsymbol{\varepsilon}_i$ is a $J$ by 1 vector of residuals for $\mathbf{z}_i$. The structural model is given by:

$$\boldsymbol{\eta}_i = \mathbf{B}'\boldsymbol{\eta}_i + \boldsymbol{\xi}_i, \tag{3}$$

where $\mathbf{B}$ is a $T$ by $T$ matrix of path coefficients among latent variables, and $\boldsymbol{\xi}_i$ is a $T$ by 1 vector of residuals for $\boldsymbol{\eta}_i$. Then, GSCA combines these three equations into a single one as follows:

$$\begin{aligned} \begin{bmatrix} \mathbf{z}_i \\ \boldsymbol{\eta}_i \end{bmatrix} &= \begin{bmatrix} \mathbf{C}' \\ \mathbf{B}' \end{bmatrix} \boldsymbol{\eta}_i + \begin{bmatrix} \boldsymbol{\varepsilon}_i \\ \boldsymbol{\xi}_i \end{bmatrix} \\ \begin{bmatrix} \mathbf{I} \\ \mathbf{W}' \end{bmatrix} \mathbf{z}_i &= \begin{bmatrix} \mathbf{C}' \\ \mathbf{B}' \end{bmatrix} \mathbf{W}'\mathbf{z}_i + \begin{bmatrix} \boldsymbol{\varepsilon}_i \\ \boldsymbol{\xi}_i \end{bmatrix} \\ \mathbf{V}'\mathbf{z}_i &= \mathbf{A}'\mathbf{W}'\mathbf{z}_i + \mathbf{e}_i, \end{aligned} \tag{4}$$

where $\mathbf{V} = [\mathbf{I}, \mathbf{W}]$, $\mathbf{A} = [\mathbf{C}, \mathbf{B}]$, $\mathbf{e}_i = \begin{bmatrix} \boldsymbol{\varepsilon}_i \\ \boldsymbol{\xi}_i \end{bmatrix}$, and $\mathbf{I}$ is an identity matrix.

To estimate the unknown parameters $\mathbf{W}$ and $\mathbf{A}$, the following least-squares criterion is minimized

$$\varphi = \sum_{i=1}^{N} \mathrm{SS}(\mathbf{V}'\mathbf{z}_i - \mathbf{A}'\mathbf{W}'\mathbf{z}_i) = \mathrm{SS}(\mathbf{ZV} - \mathbf{ZWA}), \tag{5}$$

with respect to $\mathbf{W}$ and $\mathbf{A}$, subject to an identification constraint, diag($\mathbf{W}'\mathbf{Z}'\mathbf{ZW}$) = $\mathbf{I}$, where SS($\mathbf{X}$) = tr($\mathbf{X}'\mathbf{X}$). An Alternating Least Squares (ALS) algorithm (de Leeuw, Young, & Takane, 1976) was developed to minimize this criterion. A detailed description of the ALS algorithm will be provided in the next section.

GSCA currently provides two measures of overall model fit – FIT (Hwang & Takane, 2004) and AFIT (Hwang, DeSarbo, & Takane, 2007). The FIT indicates the total variance of all endogenous variables explained by a model specification. It is given by FIT $= 1 - \left[\mathrm{SS}(\mathbf{ZV} - \mathbf{ZWA}) / \mathrm{SS}(\mathbf{ZV})\right]$. The values of FIT range from 0 to 1. The larger this value, the more variance in the variables is accounted for by the specified model. The AFIT was developed to take model complexity into account. It is given by AFIT $= 1 - (1 - \mathrm{FIT})\dfrac{d_0}{d_1}$, where $d_0 = NJ$ is the degrees of freedom for the null model ($\mathbf{W} = \mathbf{0}$ and $\mathbf{A} = \mathbf{0}$) and $d_1 = NJ - G$ is the degrees of freedom for the model being tested, where $G$ is the number of free parameters.

## 3. Nonlinear Generalized Structured Component Analysis

Young (1981) described six types of measurement as combinations of measurement level (nominal, ordinal, and numerical) and measurement process (discrete and continuous). Each measurement type involves its own measurement characteristics, and restrictions (refer to Young, 1981). The proposed nonlinear GSCA can accommodate all six types of measurement through the adoption of optimal scaling.

The proposed method consists of two estimation phases. One phase is equivalent to the current ALS estimation procedure for model parameters (**W** and **A**). The other is the optimal scaling procedure in which qualitative data are transformed to quantitative data in a way that they agree maximally with their model predictions while preserving the measurement characteristics of the data. Thus, in the optimal scaling phase, the original data matrix **Z** is assumed *qualitative*, and iteratively replaced by its transformed, quantitative one, denoted by **S** hereafter. Consequently, the proposed method divides an entire set of parameters into two subsets: model parameters (**W** and **A**) and data parameters (**S**).

Let $\mathbf{z}_j$ and $\mathbf{s}_j$ denote $N$ by 1 vectors of observed, qualitative variable $j$ and its optimally scaled counterpart, respectively ($j$ = 1, ···, $J$). In the proposed method, the following least-squares optimization criterion is minimized

$$\phi = \mathrm{SS}(\mathbf{SV} - \mathbf{SWA}) = \mathrm{SS}(\mathbf{\Psi}^* - \mathbf{\Gamma}^*\mathbf{A}), \qquad (6)$$

with respect to **W**, **A**, and **S**, subject to such restrictions as $\mathrm{diag}(\mathbf{\Gamma}^{*}{}'\mathbf{\Gamma}^*) = \mathbf{I}$, $\mathbf{s}_j{}'\mathbf{s}_j = 1$, and $\mathbf{s}_j = \omega(\mathbf{z}_j)$, where $\mathbf{\Psi}^* = \mathbf{SV}$, $\mathbf{\Gamma}^* = \mathbf{SW}$, and $\omega$ refers to a transformation of the original variable to the optimally scaled counterpart on the basis of the measurement characteristics of the data. In practice, **Z** can be composed of observed variables of different measurement types concurrently, for example, some variables are discrete-nominal, others discrete-ordinal, and others discrete-numerical. Thus, the proposed method should be capable of applying a data transformation to each variable separately.

An Alternating Least Squares (ALS) algorithm is developed to minimize (6). This algorithm alternates two major phases. One is the model estimation phase where model parameters (**W** and **A**) are updated in the least-square sense, conditional upon the optimally scaled, quantitative data (**S**). The other is the optimal scaling phase in which the data parameters (**S**) are updated in the least-square sense, given fixed model parameters.

The model estimation phase is equivalent to adopting the original ALS algorithm for estimation of the model parameters **W** and **A**. Specifically, for fixed **S**, this phase repeats the following two steps until convergence.

<u>Step 1</u>: Update **A** for fixed **W**. Criterion (6) can be re-written as:

$$\phi = \mathrm{SS}(\mathrm{vec}(\mathbf{\Psi}^*) - \mathrm{vec}(\mathbf{\Gamma}^*\mathbf{A})) = \mathrm{SS}(\mathrm{vec}(\mathbf{\Psi}^*) - (\mathbf{I} \otimes \mathbf{\Gamma}^*)\mathrm{vec}(\mathbf{A})), \tag{7}$$

where vec(**X**) indicates a supervector formed by stacking all columns of **X** one below another, and $\otimes$ indicates a Kronecker product. Let **a** denote the vector formed by eliminating fixed elements from vec(**A**). Let **Φ** denote the matrix formed by eliminating the columns of $\mathbf{I} \otimes \mathbf{\Gamma}^*$ corresponding to the fixed elements in vec(**A**). Then, the least-squares estimate of **a** is obtained by:

$$\hat{\mathbf{a}} = (\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathrm{vec}(\mathbf{\Psi}^*) . \tag{8}$$

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

<u>Step 2</u>: Update **W** for fixed **A**. Let $\mathbf{w}_t$ denote the *t*-th column of unknown component weights in **W**, which is shared by the *p*-th column in **V**, where $p = t + J$ ($t = 1, \cdots, T$ ). Let $\mathbf{\Lambda} = \mathbf{WA}$. Let $\mathbf{V}_{(-p)}$ denote **V** whose *p*-th column is the vector of zeros. Let $\mathbf{\Lambda}_{(-t)}$ denote a product matrix of **W** whose *t*-th column is the vector of zeros and **A** whose *t*-th row is the zero vector. Let $\mathbf{m}_{(p)}$ denote a *1* by *J*+*T* vector whose elements are all zeros except the *p*-th element being unity. Let $\mathbf{a}_{(t)}$ denote the *t*-th row of **A**. To update $\mathbf{w}_t$, (6) can be re-expressed as:

$$\phi = \sum_{t=1}^{T} \mathrm{SS}((\mathbf{\beta}'\otimes\mathbf{S})\mathbf{w}_t - \mathrm{vec}(\mathbf{S\Delta})) . \tag{9}$$

In (9), **β** and **Δ** are given by:

$$\boldsymbol{\beta} = \mathbf{m}_{(p)} - \mathbf{a}_{(t)}, \tag{10}$$

and

$$\boldsymbol{\Delta} = \boldsymbol{\Lambda}_{(-t)} - \mathbf{V}_{(-p)}. \tag{11}$$

Let $\boldsymbol{\theta}_t$ denote the vector formed by eliminating any fixed elements from $\mathbf{w}_t$. Let $\boldsymbol{\Xi}$ denote the matrix formed by eliminating the columns of $\boldsymbol{\beta}'\otimes\mathbf{S}$ corresponding to the fixed elements in $\mathbf{w}_t$. Then, the least-squares estimate of $\boldsymbol{\theta}_t$ is obtained by:

$$\hat{\boldsymbol{\theta}}_t = (\boldsymbol{\Xi}'\boldsymbol{\Xi})^{-1}\boldsymbol{\Xi}'\,\mathrm{vec}(\mathbf{S}\boldsymbol{\Delta}). \tag{12}$$

The updated $\mathbf{w}_t$ is recovered from $\hat{\boldsymbol{\theta}}_t$. Once $\mathbf{W}$ is obtained, latent variables are updated by $\boldsymbol{\Gamma}^* = \mathbf{SW}$, and normalized to satisfy $\mathrm{diag}(\boldsymbol{\Gamma}^{*}{}'\boldsymbol{\Gamma}^{*}) = \mathbf{I}$.

Next, given $\mathbf{W}$ and $\mathbf{A}$, the optimal scaling phase updates each variable $\mathbf{s}_j$. The optimal scaling phase consists of two steps. In one step, the model prediction of $\mathbf{s}_j$ is obtained in such a way that it minimizes (6). In the other, $\mathbf{s}_j$ is transformed to maximize the relation between $\mathbf{s}_j$ and the model prediction while respecting its measurement characteristics. Specifically, these two steps are as follows.

<u>Step 3</u>: Update the model prediction of $\mathbf{s}_j$ for fixed $\mathbf{W}$ and $\mathbf{A}$. Let $\mathbf{S}_{(-j)}$ denote $\mathbf{S}$ whose $j$-th column is the vector of zeros. Let $\mathbf{S}_{(j)}$ denote an $N$ by $J$ matrix in which the $j$-th column is equal to $\mathbf{s}_j$ and the others are all zero vectors. Let $\boldsymbol{\Sigma} = \mathbf{V} - \mathbf{WA}$. Let $\boldsymbol{\Sigma}_{(-j)}$ denote $\boldsymbol{\Sigma}$ whose $j$-th row is the vector of zeros. Let $\boldsymbol{\sigma}_{(j)}$ denote the $j$-th row of $\boldsymbol{\Sigma}$. Then, (6) can be re-expressed as:

$$
\begin{aligned}
\phi &= \mathrm{SS}(\mathbf{S\Sigma}) \\
&= \mathrm{SS}((\mathbf{S}_{(j)} + \mathbf{S}_{(-j)})\mathbf{\Sigma}) \\
&= \mathrm{SS}(\mathbf{s}_j\boldsymbol{\sigma}_{(j)} - (-\mathbf{S}_{(-j)}\mathbf{\Sigma}_{(-j)})) \\
&= \mathrm{SS}(\mathbf{s}_j\boldsymbol{\sigma}_{(j)} - \mathbf{K}_{(-j)}),
\end{aligned}
\tag{13}
$$

where $\mathbf{K}_{(-j)} = -\mathbf{S}_{(-j)}\mathbf{\Sigma}_{(-j)}$. Then, the least-squares model prediction of $\mathbf{s}_j$ is obtained by:

$$
\hat{\mathbf{s}}_j = \mathbf{K}_{(-j)}\boldsymbol{\sigma}_{(j)}'(\boldsymbol{\sigma}_{(j)}\boldsymbol{\sigma}_{(j)}')^{-1}. \tag{14}
$$

Step 4: Transform $\mathbf{s}_j$ such that it is close to $\hat{\mathbf{s}}_j$ as much as possible while satisfying its measurement restrictions. Typically, $\mathbf{s}_j$ is updated by minimizing a least-squares fitting criterion (e.g., the (normalized) differences between $\mathbf{s}_j$ and $\hat{\mathbf{s}}_j$). This comes down to regressing $\hat{\mathbf{s}}_j$ onto the space of $\mathbf{z}_j$ which represents its measurement restrictions. The least-squares estimate of $\mathbf{s}_j$ may be generally expressed as

$$
\mathbf{s}_j = \mathbf{T}_j(\mathbf{T}'_j\,\mathbf{T}_j)^{-1}\mathbf{T}_j'\hat{\mathbf{s}}_j = \mathbf{T}_j\mathbf{q}_j, \tag{15}
$$

where $\mathbf{q}_j = (\mathbf{T}_j'\mathbf{T}_j)^{-1}\mathbf{T}_j'\hat{\mathbf{s}}_j$ is a vector of optimal scaling weights. In (15), $\mathbf{T}_j$ is constructed to satisfy the measurement restrictions imposed on the variable. For example, for a discrete-nominal variable, $\mathbf{T}_j$ is given as a dummy-coded indicator matrix which shows the category choices of respondents (cf. Gifi, 1990). In this case, once $\mathbf{T}_j$ is created as a dummy-coded indicator matrix, it remains fixed throughout iterations. This indicates that the data transformation for a discrete-nominal variable is simply equivalent to replacing the original variable by its appropriate dummy-coded indicator matrix. For a discrete-ordinal variable, $\mathbf{T}_j$ is iteratively updated by Kruskal's (1964) secondary monotonic transformation so as to satisfy the order restriction that observations should be order-preserving and tied observations remain tied. For a continuous-ordinal variable, $\mathbf{T}_j$

is iteratively constructed by Kruskal's (1964) primary monotonic transformation in order to assure the order restriction that observations should be order-preserving and tied observations become untied. Young (1981) provides detailed examples on the construction of $\mathbf{T}_j$ for different measurement types. The updated $\mathbf{s}_j$ is subsequently normalized to satisfy $\mathbf{s}_j'\mathbf{s}_j = 1$.

The proposed ALS algorithm repeats the two major phases until convergence of (6). Although the ALS algorithm is monotonically convergent, there is no guarantee that it always converges to the global minimum. To avoid potential convergence to a local minimum, this algorithm can be repeatedly applied to the data, varying starting values many times (e.g., 10-20 different sets of starting values). The smallest value of (6) is then regarded as the global minimum. Subsequently, the parameter estimates associated with the smallest criterion value are chosen as the final solutions. As will be shown in the next section, given many different sets of random starts, the proposed algorithm seems to converge to the same optimum with high frequency.

As in GSCA, nonlinear GSCA adopts the bootstrap method (Efron, 1982) to examine the significance of parameter estimates without recourse to the assumption of multivariate normality of observed variables.

# 4. Empirical Applications

## 4.1. Kempler's Size Judgement Data

The first example is Kempler's (1971) size judgement data (also see Takane, Young, & de Leeuw, 1980). In this example, four age groups consisting of 16-25 children in grades 1, 3, 5 and 7 were asked to judge the sizes of 100 rectangles as either 'large' or 'small'.

The 100 different rectangles were created by factorial combinations of 10 height levels and 10 width levels each varying from 10 to 14.5 inches by a half-inch interval. For each age group, the number of children who judged a rectangle as large was counted, and then used as a measure of perceived largeness for the rectangle. Consequently, these data include qualitative variables only - two discrete-nominal variables representing 10 different levels of height and width; and four continuous-ordinal variables indicating the numbers of counts for the four age groups. The sample size was equal to the number of rectangles judged (i.e., N = 100).

Following Takane et al. (1980), we specified the same weighted addictive model for the data, in which two exogenous variables (height and width) were hypothesized to influence four endogenous variables (four age groups' perceived largeness of rectangles). It would be inappropriate to apply linear GSCA for fitting this model to the data since the data comprised all qualitative variables (either discrete-nominal or continuous-ordinal). Thus, we applied the proposed nonlinear GSCA for fitting the model to the data. In the optimal scaling phase, specifically, the two original discrete-nominal variables were converted to two 100 (rectangles) by 10 (categories) dummy-coded indicator matrices which were subsequently used as a new set of exogenous variables (i.e., 20 dummy-coded indicators in total) in the model. We applied Kruskal's (1964) primary monotonic transformation to the four original endogenous variables (see Young, 1980).

The specified model is depicted in Figure 1. In the figure, as shown in (15), each optimally scaled discrete-nominal variable (HEIGHT or WIDTH) is expressed as a weighted composite of the corresponding ten dummy-coded indicators (i.e., $\mathbf{z}_1 - \mathbf{z}_{10}$ for HEIGHT and $\mathbf{z}_{11} - \mathbf{z}_{20}$ for WIDTH). Thus, the optimal scaling weights for each discrete-nominal variable can be considered equivalent to component weights, and the optimally scaled discrete-nominal variables are comparable to latent variables in the context of GSCA, as given in (1). In addition, these optimally scaled discrete-nominal variables are hypothesized to affect four endogenous, continuous-ordinal variables ($\mathbf{s}_1$ – $\mathbf{s}_4$). All residual terms for the endogenous variables were removed to make the figure concise. It may thus be worthwhile to note that this weighted additive model can also be viewed as a structural equation model, where latent variables are formed as weighted composites of exogenous observed variables; and they are to influence endogenous observed variables. In particular, this type of a structural equation model can be called the extended redundancy analysis model (Takane & Hwang, 2005). GSCA subsumes the extended redundancy analysis model as a special case.

---

Insert Figure 1 about here

---

We repeated the proposed ALS algorithm with 20 sets of random starting values for parameter estimates. The ALS algorithm was found to converge to the same optimum throughout the repetitions. Nonlinear GSCA estimation provided FIT = .96 and AFIT = .96, indicating that the specified model accounted for about 96% of the total variance in the endogenous variables. Table 1 provides the parameter estimates and their 95% bootstrap confidence intervals calculated based on 200 bootstrap samples. For each bootstrap sample, we used a different set of random starting values for parameter estimates. Figure 2 displays the original observed values of the four endogenous variables on the x-axis and their monotonically transformed values on the y-axis.

Insert Table 1 and Figure 2 about here

As exhibited in Table 1, the optimal scaling weight estimates for HEIGHT turned out to be all significant except the weight for $\mathbf{z}_5$ that indicates the category of 12 inches high. Moreover, these weight estimates tended to be larger with increases in height. In addition, the optimal scaling weight estimates for WIDTH were all significant except that for $\mathbf{z}_{15}$ pointing to the category of 12 inches wide. Again, the weights estimates for WIDTH appeared larger when the level of width increased. Furthermore, HEIGHT and WIDTH had significant and positive effects on the four optimally scaled endogenous variables ($\mathbf{s}_1$ – $\mathbf{s}_4$). The impacts of HEIGHT on perceived largeness of rectangles tended to decrease with age, whereas those of WIDTH were likely to increase with age. This suggests that younger children tend to focus more heavily on the height than the width of rectangles in their size evaluations. This tendency appears to diminish as they become older. These findings seem to validate Kempler's (1971) hypothesis with respect to children's perception on the size of rectangles.

### 4.2. The Basic Health Indicator Data

Another example was analyzed to empirically compare the performance between linear GSCA and its proposed nonlinear version. This example is part of the so-called Basic Health Indicator data collected by the World Health Organization in 1999. From the entire database, five observed variables were employed in this application: (1) real gross domestic product (GDP) per capita adjusted for purchasing power parity in 1985 US dollars; (2) the average number of years of education given for females aged 25 years and above (FEUD); (3) the percentage of children immunized against measles in 1997 (Measles); (4) infant mortality rate (IMR), defined as the number of deaths per 1,000 live births between birth and exact age of one year in 1998; and (5) maternal mortality ratio (MMR), defined as the number of maternal deaths per 100,000 live births in 1990. The sample size was 50, indicating the total number of countries for which these variables were measured.

For illustrative purposes, we specified a model for these data, in which a latent variable was defined as a weighted composite of the first three observed variables (GDP, FEUD, and Measles); and in turn, influenced these observed variables. This latent variable was named ‘Social and Economic Growth (SEG)’. In addition, the latent variable was specified to affect two endogenous observed variables (IMR and MMR). The specified model is depicted in Figure 3. Only component loadings and path coefficients are displayed to make the figure concise.

---

Insert Figure 3 about here

---

At first, linear GSCA was applied to fit the model to the data. Linear GSCA estimation provided FIT = .79 and AFIT = .78, indicating that the specified model accounted for about 79% of the total variance in the data. Table 2 provides the parameter estimates and their 95% bootstrap confidence intervals calculated based on 200 bootstrap samples. The weight estimates of the three observed variables for the latent variable (i.e., GDP, FEDU, and Measles) were similar and significant, indicating that they contributed well to determining the latent variable. Moreover, all loading estimates for these variables were large and significant. This suggests that the latent variable was well defined to explain a large portion of the variances of the observed variables. However, the loading estimate for Measles were smaller than those for GDP and FEUD.  Furthermore, the latent variable SEG had significant and negative effects on IMR and MMR. This indicates that a high level of social and economic growth were likely to decrease both infant and maternal mortality rates.

Insert Table 2 about here

Next, the proposed nonlinear GSCA was used to fit the same model to the data. All five observed variables were now considered continuous-ordinal. We applied Kruskal’s (1964) primary monotonic transformation to the five variables. We repeated the proposed ALS algorithm with 20 sets of random starting values for parameter estimates. The ALS algorithm always converged to the same optimum. Figure 4 displays the original values of the five variables on the x-axis and their monotonically transformed values on the y-axis. Table 2 also presents the parameter estimates and their 95% bootstrap confidence intervals obtained from nonlinear GSCA.

Insert Figure 4 and Table 2 about here

Nonlinear GSCA estimation resulted in a much higher level of model fit (FIT = .94 and AFIT = .94), compared to its linear counterpart. The interpretations of these parameter estimates are basically the same as those obtained from linear GSCA. Nonetheless, all loading estimates for the transformed variables were larger than those for the original (non-transformed) ones. In particular, the loading for Measles was quite larger, indicating that it explained a greater portion of the variance of the latent variable. In addition, the estimates of path coefficients became larger than those from linear GSCA. Thus, the proposed method was able to enhance the associations among the variables substantially by taking into account the nonlinear relationships among them. This suggests that the proposed nonlinear GSCA is more beneficial in analyzing the interrelationships among the health-indicator variables compared to linear GSCA, due to its capability to accommodate potential nonlinear relationships among variables.

## 5. Concluding Remarks

In this paper, GSCA was extended to the analysis of qualitative variables. This proposed nonlinear version of GSCA integrates the current GSCA estimation and data transformation into a unified estimation procedure. In particular, this method adopts the optimal scaling approach to data transformation because the approach is well suited with the GSCA estimation. As demonstrated by empirical applications, the proposed method was effectively applied for the specification and analysis of path-analytic relationships among qualitative variables, taking into account their nonlinear associations.

Thus, the proposed method is a valuable extension of GSCA, which greatly contributes to improving the data-analytic capability of GSCA. Along with other extensions and refinements to date, the proposed method will serve to position GSCA as a more general and flexible approach to component-based SEM.

Despite its significant implications, the proposed method is not yet implemented into a software program, so that it will be difficult for substantive researchers and practitioners to employ the method in their research. It would thus be desirable to translate this new development in GSCA into the software program *GeSCA* in the near future.

## References


Bock, R. D. (1960). *Methods and applications of optimal scaling*. Psychometric Laboratory Report #25, University of North Carolina.

de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, 471-503.

Efron, B. (1982). *The Jackknife, The Bootstrap and Other Resampling Plans*. Philadelphia: SIAM.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: Wiley.

Hwang, H. (2009a). Regularized generalized structured component analysis. *Psychometrika*, 74, 517-530.

Hwang, H. (2009b). *GeSCA*. Available at http://www.sem-gesca.org.

Hwang, H., & Takane, Y. (2004), Generalized structured component analysis. *Psychometrika, 69, 81-99*.

Hwang, H., DeSarbo, S. W., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, 72, 181-198.

Hwang, H., Ho, M.-H., & Lee, J. (2009). Generalized structured component analysis with latent interactions. Paper submitted for publication.

Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, M. A., & Hong, S. (in press). A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research*.

Hwang, H., Takane, Y., & Malhotra, N. K. (2007). Multilevel generalized structured component analysis. *Behaviormetrika,* 34, 95-109.

Kempler, B. (1971). Stimulus correlates of area judgements: A psychophysiological developmental study. *Developmental Psychology*, 4, 158-163.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27

Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin*, 105, 156-166.

Takane, Y., & Hwang, H. (2005). An extended redundancy analysis and its applications to two practical examples. *Computational Statistics and Data Analysis*, 49, 785-808.

Takane, Y., Young, F. W., & de Leeuw, J. (1980). An individual differences additive model: An alternating least squares method with optimal scaling features. *Psychometrika*, 45, 183-209.

Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Quality Management & Business Excellence*, 19, 871-886.

Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, 46, 357-388.

Table 1. The parameter estimates and their 95% bootstrap confidence intervals in parentheses obtained from the proposed nonlinear GSCA for Kempler's data.

| | | | Estimate | 95% CI |
|---|---|---|---|---|
| Weights | HEIGHT | $\mathbf{z}_1$ | -1.43 | -1.66, -1.27 |
| | | $\mathbf{z}_2$ | -1.39 | -1.67, -1.16 |
| | | $\mathbf{z}_3$ | -.91 | -1.12, -.70 |
| | | $\mathbf{z}_4$ | -.57 | -.74, -.41 |
| | | $\mathbf{z}_5$ | -.20 | -.46, .03 |
| | | $\mathbf{z}_6$ | .33 | .52, .17 |
| | | $\mathbf{z}_7$ | .58 | .79, .37 |
| | | $\mathbf{z}_8$ | 1.08 | 1.34, .83 |
| | | $\mathbf{z}_9$ | 1.02 | 1.24, .80 |
| | | $\mathbf{z}_{10}$ | 1.48 | 1.72, 1.28 |
| | WIDTH | $\mathbf{z}_{11}$ | -1.56 | -1.84, -1.27 |
| | | $\mathbf{z}_{12}$ | -1.27 | -1.47, -1.01 |
| | | $\mathbf{z}_{13}$ | -.96 | -1.27, -.73 |
| | | $\mathbf{z}_{14}$ | -.53 | -.73, -.32 |
| | | $\mathbf{z}_{15}$ | -.10 | -.46, .14 |
| | | $\mathbf{z}_{16}$ | .33 | .64, .03 |
| | | $\mathbf{z}_{17}$ | .48 | .73, .21 |
| | | $\mathbf{z}_{18}$ | .99 | 1.29, .74 |
| | | $\mathbf{z}_{19}$ | 1.32 | 1.71, 1.09 |
| | | $\mathbf{z}_{20}$ | 1.30 | 1.56, 1.07 |
| Path Coefficients | HEIGHT $\rightarrow$ $\mathbf{s}_1$ | | .87 | .80, .97 |
| | HEIGHT $\rightarrow$ $\mathbf{s}_2$ | | .77 | .70, .88 |
| | HEIGHT $\rightarrow$ $\mathbf{s}_3$ | | .79 | .71, .91 |
| | HEIGHT $\rightarrow$ $\mathbf{s}_4$ | | .73 | .65, .82 |
| | WIDTH $\rightarrow$ $\mathbf{s}_1$ | | .44 | .35, .55 |
| | WIDTH $\rightarrow$ $\mathbf{s}_2$ | | .60 | .52, .70 |
| | WIDTH $\rightarrow$ $\mathbf{s}_3$ | | .58 | .49, .68 |
| | WIDTH $\rightarrow$ $\mathbf{s}_4$ | | .67 | .60, .79 |

Table 2. The parameter estimates and their 95% bootstrap confidence intervals obtained from linear and nonlinear GSCA for the Basic Health Indicator data.

| | | Linear GSCA | | Nonlinear GSCA | |
|---|---|---|---|---|---|
| | | Estimate | 95% CI | Estimate | 95% CI |
| Weights | GDP | .43 | .39, .49 | .35 | .33, .40 |
| | FEUD | .46 | .42, .52 | .35 | .33, .37 |
| | Measles | .33 | .19, .37 | .34 | .21, .35 |
| Loadings | GDP | .87 | .80, .93 | .98 | .95, 1.00 |
| | FEUD | .93 | .89, .96 | .94 | .90, 1.00 |
| | Measles | .62 | .35, .77 | .93 | .86, 1.00 |
| Path Coefficients | SEG → IMR | -.85 | -.91, -.80 | -.89 | -1.00, -.81 |
| | SEG → MMR | -.74 | -.82, -.67 | -.87 | -1.00, -.76 |

Figure 1. The specified model for Kempler's data.

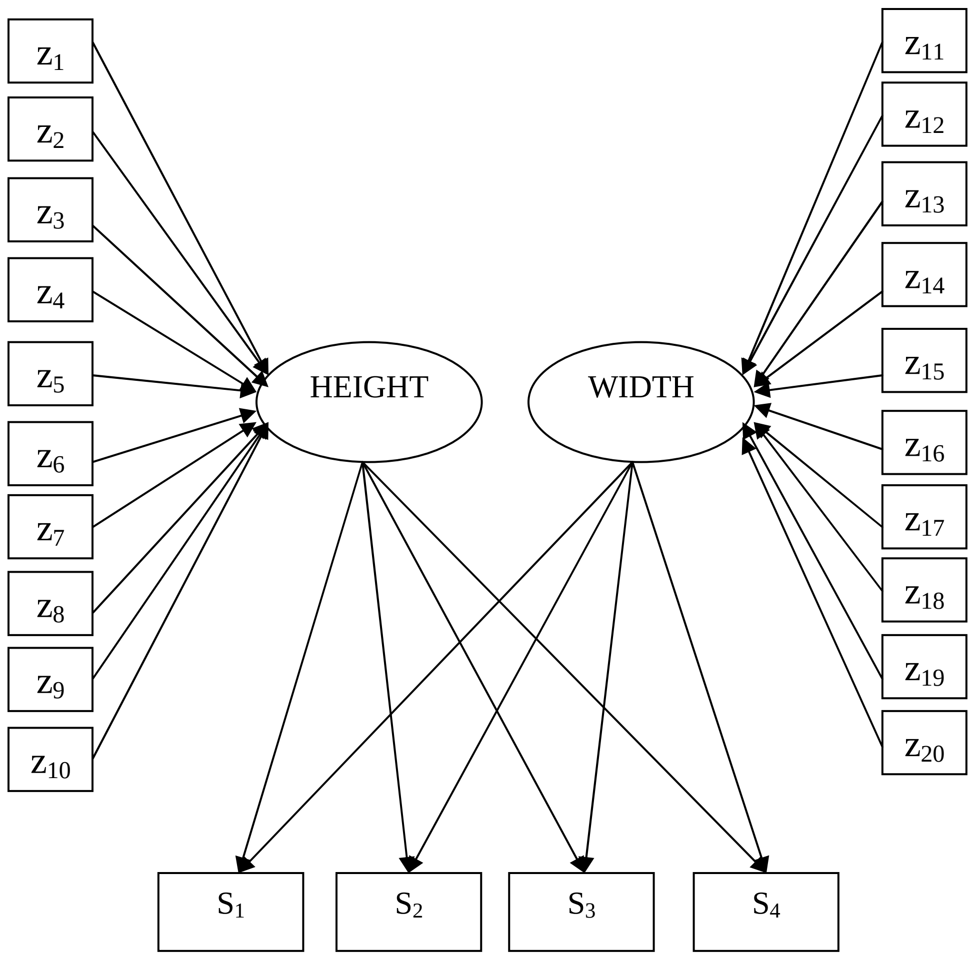

Figure 2. The monotonic transformations of the four endogenous variables ($\mathbf{s}_1 - \mathbf{s}_4$) in Kempler's data.

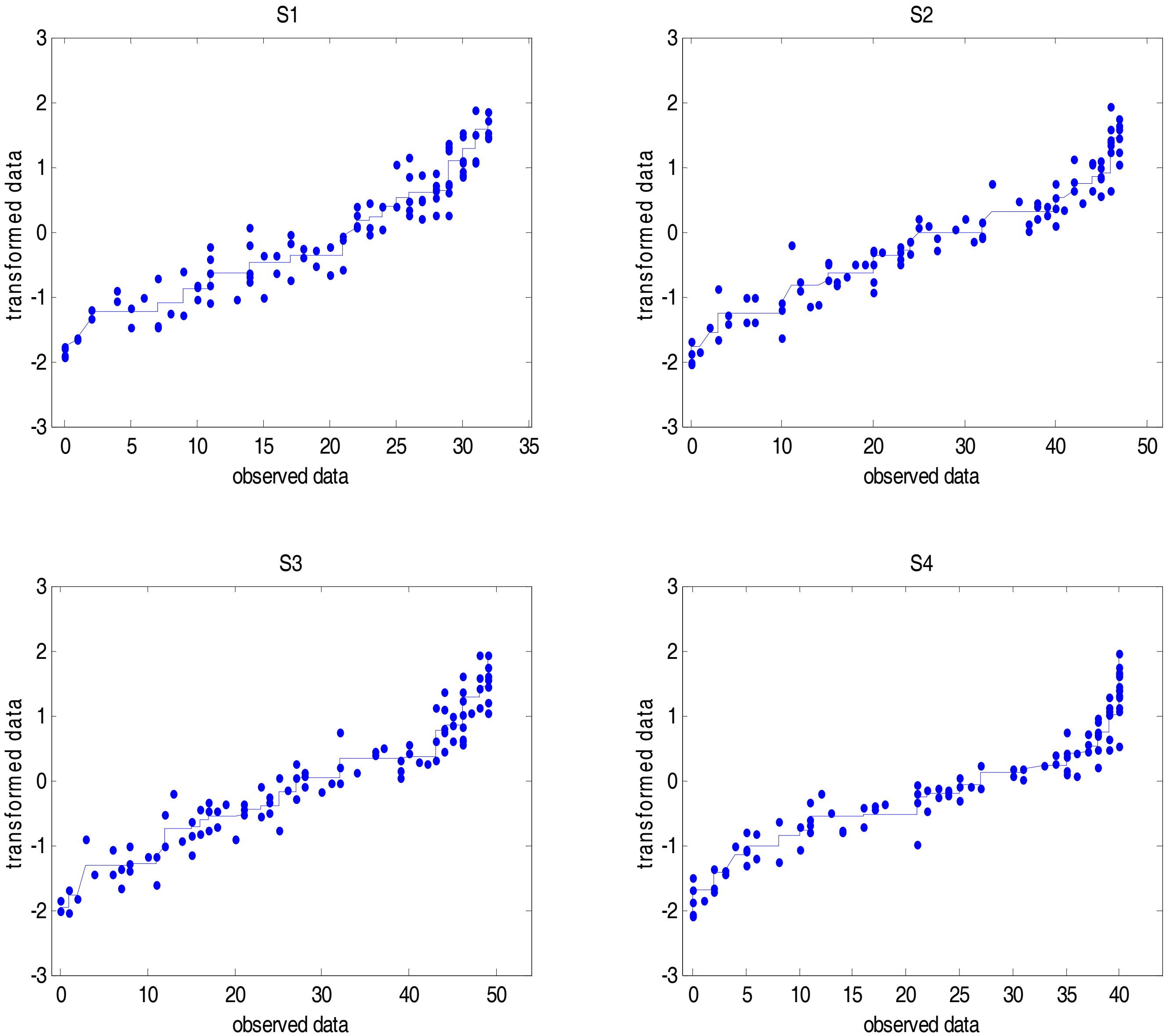

Figure 3. The specified model for the Basic Health Indicator data.

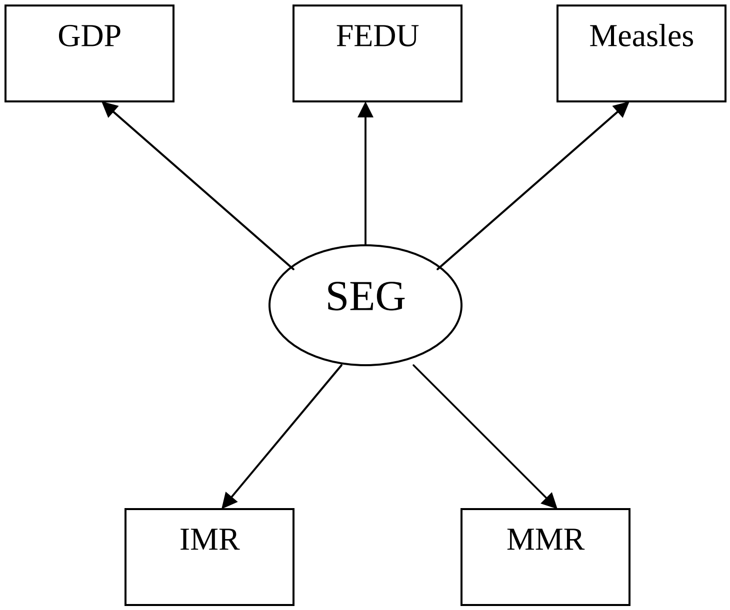

Figure 4.The monotonic transformations of the five variables in the Basic Health Indicator data.