

Fuzzy Clusterwise Generalized Structured Component Analysis

Heungsun Hwang

McGill University

Wayne S. DeSarbo

Pennsylvania State University

Yoshio Takane

McGill University

November 13, 2006

The work reported in this paper was supported by Grant 290439 and Grant A6394 from the Natural Sciences and Engineering Research Council of Canada to the first and third authors, respectively. We wish to thank Terry Duncan for generously providing us with his alcohol use data. We also wish to thank the Editor, Associate Editor, and two anonymous reviewers for their constructive comments which helped improve the overall quality and readability of this 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) was recently introduced by Hwang and Takane (2004) as a component-based approach to path analysis with latent variables. The parameters of GSCA are estimated by pooling data across respondents under the implicit assumption that they all come from a single, homogenous group. However, as has been empirically demonstrated by various researchers across a number of areas of inquiry, such aggregate analyses can often mask the true structure in data when respondent heterogeneity is present. In this paper, GSCA is generalized to a fuzzy clustering framework so as to account for potential group-level respondent heterogeneity. An alternating least-squares procedure is developed and technically described for parameter estimation. A small-scale Monte Carlo study involving synthetic data is carried out to compare the performance between the proposed method and an extant approach. In addition, an empirical application concerning alcohol use among adolescents from US northwestern urban areas is presented to illustrate the usefulness of the proposed method. Finally, a number of directions for future research are provided.

Keywords: Generalized structured component analysis, fuzzy clustering, respondent heterogeneity, alternating least squares, latent curve models, alcohol use.

1. Introduction

Generalized Structured Component Analysis (GSCA) (Hwang & Takane, 2004) was recently proposed for path analysis with latent variables which are defined as weighted composites of observed variables. As was demonstrated, GSCA does not suffer from the occurrence of improper solutions and the problem of latent score indeterminacy. Moreover, it does not require the assumption of multivariate normality of observed variables and large samples for parameter estimation. Furthermore, the method employs a global optimization criterion that is consistently minimized as a criterion for the estimation of model parameters. This, in turn, serves to facilitate the calculation of measures of overall model fit which are usefully employed for model evaluation, development, and comparison purposes.

The parameters of GSCA are currently estimated by pooling the data across respondents under the tacit assumption that all respondents are sampled from a single homogenous population, i.e., an aggregate sample analysis. However, it is often more realistic to assume that respondents come from heterogeneous sub-groups characterized, for example, by different behaviors, attitudes, and/or preferences. Such group- or cluster-level respondent heterogeneity has been discussed from several different theoretical perspectives; for example, consumer belief structures have been hypothesized to vary across different market segments according to the expectancy value models (Bagozzi, 1982); the trajectories of antisocial behavior may be characterized by two distinct developmental pathways such as ‘life-course persistent’ and ‘adolescent-limited’ (Moffitt, 1993); and the process of brand choice decision can vary across different groups of consumers (Kamakura, Kim, & Lee, 1996). Moreover, in the situations where respondent

heterogeneity is present, an aggregate-level path analysis with latent variables that ignores the heterogeneity is likely to yield misleading results. For instance, Jedidi, Jagpal, and DeSarbo (1997) illustrated serious masking problems with such aggregate analyses when heterogeneity in consumer behavior was not taken into account.

One convenient manner for accounting for this heterogeneity in path analysis with latent variables is to adopt a two-step procedure. In the first step, some variety of cluster analysis is used to identify clusters of respondents; and, in the second step, path analysis with latent variables is applied to each of the resulting clusters. The latter step is identical to the multi-group comparison procedures given the clusters obtained via the first step. Although this two-step approach is simple to implement, it is unattractive to rely solely on clustering methods for identifying the cluster structures in the data with no reference to path analysis with latent variables. Moreover, the two disparate procedures typically optimize different objective criteria. Hence, the combined use of clustering methods and path analysis with latent variables in a single framework is recommended (cf. DeSarbo & Cron, 1988; Wedel & Kamakura, 1998).

In this paper, GSCA is extended to take into account cluster-level respondent heterogeneity. Specifically, GSCA is combined with fuzzy clustering (e.g., Bezdek, 1974a; Bezdek, Coray, Gunderson, & Watson, 1981; Dunn, 1974; Hathaway & Bezdek, 1993; Manton, Woodbury, & Tolley, 1994; Wedel & Steenkamp, 1989) in a unified framework. Fuzzy clustering is an overlapping clustering method which permits respondents to belong partially to multiple clusters. There are several reasons for adopting the fuzzy clustering approach. First, the fuzzy clustering algorithm is attractive in the context of the proposed method because it is easily compatible with the

distribution-free optimization procedure of GSCA. This compatibility is beneficial in combining two separate optimization criteria for GSCA and fuzzy clustering into a single one. Moreover, due to the difficulty of identifying a clear boundary between clusters in real-world problems, the partial classification of fuzzy clustering appears more attractive than the deterministic classification of non-overlapping clustering methods such as K -means (McBratney & Moore, 1985; Wedel & Kamakura, 1998). Furthermore, the fuzzy clustering approach offers other major advantages over traditional clustering methods. First, the fuzzy clustering algorithm is computationally more efficient because dramatic changes in the value of cluster membership are less likely to occur in estimation procedures (McBratney & Moore, 1985). Second, fuzzy clustering has been shown to be less afflicted by local optima problems (Heiser & Groenen, 1997). Finally, the memberships for any given set of respondents indicate whether there is a second best cluster almost as good as the best cluster – a result which traditional clustering methods cannot uncover (Everitt, Landau, & Leese, 2001).

In an effort to accommodate cluster-level respondent heterogeneity in GSCA, the proposed method appears to pursue the same goal as Finite Mixture Covariance Structure Analysis (FMCSA) (Jedidi et al., 1997) which aims to identify clusters of relatively homogeneous respondents; and, within each of those clusters, the parameters of the path-analytic model with latent variables are estimated by Covariance Structure Analysis (CSA) (Bock & Bargmann, 1966; Jöreskog, 1970, 1973). Under the assumption of multivariate normality of observed variables conditional upon cluster membership, FMCSA seeks to maximize a likelihood function typically by using the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). Under the additional

assumptions of large sample sizes and correct model specification, FMCSA provides efficient estimates of parameters. Moreover, FMCSA offers diverse information criteria such as AIC and BIC for model comparison. The capability of FMCSA to offer information criteria may help one determine the appropriate number of clusters in a more familiar manner than the proposed method.

FMCSA is technically based on CSA. As such, it is not free from the limitations which are inherent to CSA - for example, inadmissible or improper solutions, indeterminate latent score estimates, stringent assumptions such as no excessive kurtosis of observed variables and large sample size which, in turn, is required for valid estimation (Fornell & Bookstein, 1982; Hwang & Takane, 2002). In addition to those limitations stemming from CSA, FMCSA suffers from shortcomings of its own. In particular, FMCSA requires very large samples for parameter estimation. It is known that a minimum sample size of $J(J+1)/2$ in each cluster is needed in order to obtain a positive definite covariance matrix within each cluster, where J is the number of observed variables (Wedel & Kamakura, 1998). In practice, the large sample size can be burdensome when the number of clusters becomes large. Moreover, for identification, FMCSA assumes that all data within unknown clusters follow multivariate normal distributions, which is a more stringent distributional assumption than what is required in CSA. Furthermore, finite mixture procedures usually require very intensive computation and suffer from slow convergence due to the EM approach or gradient based estimation procedures (DeSarbo, Grewal, & Hwang, 2006). Lastly, the information criteria for model selection are based on the regularity properties of the likelihood function which do not typically hold in finite mixture models (McLachlan & Peel, 2000).

On the other hand, no specific form of distributions for observed variables (within each cluster) needs to be *a priori* assumed in the proposed method. As will be demonstrated in a later section regarding Monte Carlo analyses with synthetic data, the proposed method seems to perform well in small samples; and it is less afflicted by the problem of convergence compared to the finite mixture approach. Furthermore, as one reviewer denotes, the concept of partial membership underlying the proposed method (Zadeh, 1965) appears more appealing than that of hard classification underlying FMCSA and all latent class procedures which provide posterior probabilities of memberships due to the lack of sufficient information from data for unique classification (also see Wedel & Kamakura, 1998).

The paper is organized as follows. In Section 2, the technical underpinnings of GSCA are briefly discussed so as to facilitate the derivation of the proposed method. In Section 3, the technical details are provided which underlie the generalization of GSCA to a fuzzy clustering framework. An alternating least-squares optimization procedure for parameter estimation is presented. In Section 4, a small-scale Monte Carlo simulation study is presented to evaluate the performance of the proposed method and the finite mixture approach. In Section 5, an application is provided in relation to alcohol use among US adolescents in an effort to demonstrate the usefulness of the proposed method in the analysis of longitudinal data. The final section is devoted to implications with a focus on several additional aspects of the proposed method.

2. Generalized Structured Component Analysis

Let $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N]'$ denote an N by J matrix of observed variables, where \mathbf{z}_i is a J by 1 vector for respondent i ($i = 1, \dots, N$). As its name explicitly suggests, GSCA is

a *component-based* approach to path analysis with latent variables (cf. Chin 1998; Velicer & Jackson, 1990). Thus, GSCA defines latent variables as components or weighted composites of observed variables through the following equation:

$$\boldsymbol{\eta}_i = \mathbf{W}' \mathbf{z}_i, \quad (1)$$

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 consisting of component weights assigned to observed variables. Moreover, GSCA involves two additional equations for model specifications. One is for the measurement model which specifies the relationships between observed and latent variables; and the other is for the structural model which expresses the relationships among latent variables. Specifically, in GSCA, the measurement model is given by:

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

where \mathbf{C} is a T by J matrix of loadings relating latent variables to observed variables and $\boldsymbol{\varepsilon}_i$ is a J by 1 vector of residuals for \mathbf{z}_i . When $\mathbf{C} = \mathbf{0}$, (2) indicates the formative measurement model in which all latent variables are formed by their observed variables as given in (1). On the other hand, when $\mathbf{C} \neq \mathbf{0}$, (2) represents the reflective measurement model where all observed variables are influenced by their latent variables. In addition, only certain loadings in \mathbf{C} can be equal to zeros so as to include both formative and reflective models at the same time (see Hwang & Takane (2004) for an example). The structural model is defined by:

$$\boldsymbol{\eta}_i = \mathbf{B}' \boldsymbol{\eta}_i + \boldsymbol{\xi}_i, \quad (3)$$

where \mathbf{B} is a T by T matrix of path coefficients connecting latent variables among themselves and $\boldsymbol{\xi}_i$ is a T by 1 vector of residuals for $\boldsymbol{\eta}_i$.

Then, the general GSCA model is derived from combining these three equations into the following single framework:

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

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. Though (4) is the original GSCA model proposed in Hwang and Takane (2004), this model can also be re-written as follows:

$$\mathbf{u}_i = \mathbf{T} \mathbf{u}_i + \mathbf{e}_i, \quad (5)$$

where $\mathbf{u}_i = \begin{bmatrix} \mathbf{I} \\ \mathbf{W}' \end{bmatrix} \mathbf{z}_i$, and $\mathbf{T} = \begin{bmatrix} \mathbf{0} & \mathbf{C}' \\ \mathbf{0} & \mathbf{B}' \end{bmatrix}$. As shown in (5), it is noteworthy that the GSCA

model is essentially of the same form as the Reticular Action Model (RAM) (McArdle & McDonald, 1984), which is mathematically the most compact specification amongst various CSA formulations including the LISREL (Jöreskog, 1973) and the Bentler-Weeks (Bentler & Weeks, 1980) models. With respect to the RAM, the main difference in model specification is that GSCA defines latent variables as components, i.e., $\boldsymbol{\eta}_i = \mathbf{W}' \mathbf{z}_i$ in (1).

The parameters of GSCA (\mathbf{W} and \mathbf{A}) are estimated such that the sum of squares of all residuals (\mathbf{e}_i) is as small as possible. This is equivalent to minimizing the following optimization criterion:

$$\phi = \sum_{i=1}^N \text{SS}(\mathbf{V}' \mathbf{z}_i - \mathbf{A}' \mathbf{W}' \mathbf{z}_i) = \text{SS}(\mathbf{ZV} - \mathbf{ZWA}), \quad (6)$$

with respect to \mathbf{W} and \mathbf{A} , where $SS(\mathbf{M}) = \text{trace}(\mathbf{M}'\mathbf{M})$. An Alternating Least Squares (ALS) algorithm (de Leeuw, Young, & Takane, 1976) was developed to minimize this criterion. The ALS algorithm alternates two main steps until convergence. In the first step, for fixed \mathbf{W} , \mathbf{A} is updated in the least-squares sense. In the second step, \mathbf{W} is updated in the least-squares sense for fixed \mathbf{A} . A technical description of the ALS algorithm is provided in the Appendix.

As seen from (6), the parameters of GSCA are estimated by aggregating or pooling the data across all respondents. The aggregate-sample GSCA is not suitable for investigating whether there exist heterogeneous sub-groups of the population which evoke qualitatively distinct response characteristics within the framework of path analysis with latent variables. In the subsequent section, an extension of the GSCA methodology is proposed that enables to accommodate cluster-level respondent heterogeneity.

3. Fuzzy Clusterwise Generalized Structured Component Analysis

Let C denote the prescribed number of clusters of respondents. Let \mathbf{U}_c denote a diagonal matrix consisting of fuzzy membership values of respondents in cluster c , u_{ci} ($c = 1, \dots, C$), which satisfy the two assumptions of fuzzy clustering, i.e., $0 \leq u_{ci} \leq 1$ and

$\sum_{c=1}^C u_{ci} = 1$ (Bezdek, 1981). Let m denote the user prescribed fuzzy weight scalar, often called the ‘fuzzifier’ (Bezdek, 1974a), which influences the degree of fuzziness of the solution. The method proposed herein aims to identify fuzzy clusters of respondents that are relatively homogeneous and, at the same time, obtain GSCA parameters of a hypothesized path analytic model with latent variables within each of the clusters. This

problem of combining GSCA and fuzzy clustering in a single framework is equivalent to minimizing the following optimization criterion:

$$\varphi = \sum_{c=1}^C \sum_{i=1}^N u_{ci}^m \text{SS}(\mathbf{V}_c' \mathbf{z}_i - \mathbf{A}_c' \mathbf{W}_c' \mathbf{z}_i) = \sum_{c=1}^C \text{SS}(\mathbf{Z} \mathbf{V}_c - \mathbf{Z} \mathbf{W}_c \mathbf{A}_c)_{\mathbf{U}_c^m}, \quad (7)$$

with respect to \mathbf{U}_c , \mathbf{W}_c , and \mathbf{A}_c , subject to $\sum_{c=1}^C u_{ci} = 1$, where $\text{SS}(\mathbf{M})_{\mathbf{H}} = \text{trace}(\mathbf{M}' \mathbf{H} \mathbf{M})$. It

is easy to see that (7) reduces to (6) when $C = 1$. Thus, the proposed method subsumes GSCA as a special case where there is only one cluster (i.e., an aggregate sample analysis). We shall call this proposed method Fuzzy Clusterwise GSCA (FCGSCA) hereafter.

To minimize (7) for parameter estimation, two optimization steps are alternated until convergence as follows:

Step 1: The GSCA parameters \mathbf{W}_c and \mathbf{A}_c are updated for fixed \mathbf{U}_c . The first step is equivalent to minimizing:

$$\begin{aligned} \varphi &= \sum_{c=1}^C \text{SS}((\mathbf{U}_c^m)^{1/2} (\mathbf{Z} \mathbf{V}_c - \mathbf{Z} \mathbf{W}_c \mathbf{A}_c)) \\ &= \sum_{c=1}^C \text{SS}(\mathbf{Z}_c \mathbf{V}_c - \mathbf{Z}_c \mathbf{W}_c \mathbf{A}_c), \end{aligned} \quad (8)$$

where $\mathbf{Z}_c = (\mathbf{U}_c^m)^{1/2} \mathbf{Z}$. This criterion is essentially equivalent to the sum of (6) across C clusters. Thus, the ALS algorithm for GSCA is used to update \mathbf{W}_c and \mathbf{A}_c within each cluster (see the Appendix).

Step 2: The membership parameter u_{ci} is updated for fixed \mathbf{W}_c and \mathbf{A}_c .

Let $e_{ci} = \text{SS}(\mathbf{V}_c' \mathbf{z}_i - \mathbf{A}_c' \mathbf{W}_c' \mathbf{z}_i)$. Then, u_{ci} is updated by:

$$\hat{u}_{ci} = \frac{1}{\sum_k^c \left(\frac{e_{ci}}{e_{ki}} \right)^{1/(m-1)}}. \quad (9)$$

Formula (9) can be derived as follows. Minimizing (7) under the membership constraint is equivalent to minimizing:

$$L = \sum_{c=1}^C \sum_{i=1}^N u_{ci}^m e_{ci} - \lambda \left(\sum_{c=1}^C u_{ci} - 1 \right), \quad (10)$$

where λ is a Lagrangian multiplier. Solving $\frac{\partial L}{\partial u_{ci}} = m u_{ci}^{m-1} e_{ci} - \lambda = 0$ for u_{ci} yields:

$$\hat{u}_{ci} = \left(\frac{\lambda}{m e_{ci}} \right)^{1/(m-1)}. \quad (11)$$

Using $\frac{\partial L}{\partial \lambda} = \sum_{c=1}^C u_{ci} - 1 = 0$ and (11) leads to:

$$\hat{\lambda} = \left(\left(\sum_{c=1}^C 1/(m e_{ci}) \right)^{1/(m-1)} \right)^{1-m}. \quad (12)$$

Then, (9) is obtained by inserting (12) into (11) (see also Wedel & Steenkamp, 1989).

The proposed algorithm monotonically decreases the value of criterion (7) which, in turn, is also bounded from below. The algorithm is therefore convergent. However, it does not guarantee that the convergence point is the global minimum. To safeguard against local minima, we operationalize two alternative methods of obtaining starting values. With rational starts, the fuzzy c-means algorithm (Bezdek, 1974a; Dunn, 1974) is first applied to \mathbf{Z} , and the resultant memberships are used as initial values for u_{ci} .

Secondly, we can repeat the estimation procedures with many random initial starts for the parameters, compare the obtained criterion values after convergence, and choose the solution associated with the smallest criterion value.

When N is large relative to J , the first step of the algorithm may be made more efficient by adopting the following procedure. Let $\mathbf{Z}_c = \mathbf{Q}_c \mathbf{R}_c$ be portion of the QR decomposition of \mathbf{Z}_c pertaining to the column space of \mathbf{Z}_c , where \mathbf{Q}_c is an N by J orthonormal matrix, so that $\mathbf{Q}_c' \mathbf{Q}_c = \mathbf{I}$, and \mathbf{R}_c is a J by J upper-triangular matrix. Then, (8) can be rewritten as follows:

$$\begin{aligned}\varphi &= \sum_{c=1}^C \text{SS}(\mathbf{Q}_c \mathbf{R}_c \mathbf{V}_c - \mathbf{Q}_c \mathbf{R}_c \mathbf{W}_c \mathbf{A}_c) \\ &= \sum_{c=1}^C \text{SS}(\mathbf{Q}_c (\mathbf{R}_c \mathbf{V}_c - \mathbf{R}_c \mathbf{W}_c \mathbf{A}_c)) \\ &= \sum_{c=1}^C \text{SS}(\mathbf{R}_c \mathbf{V}_c - \mathbf{R}_c \mathbf{W}_c \mathbf{A}_c).\end{aligned}\tag{13}$$

It is computationally more efficient to minimize (13) instead of (8) because the size of \mathbf{R}_c is usually much smaller than \mathbf{Z}_c .

In FCGSCA, the overall fit of a hypothesized model is measured by:

$$\text{FIT} = 1 - \frac{\sum_{c=1}^C \text{SS}(\mathbf{Z}_c \mathbf{V}_c - \mathbf{Z}_c \mathbf{W}_c \mathbf{A}_c)}{\sum_{c=1}^C \text{SS}(\mathbf{Z}_c \mathbf{V}_c)}.\tag{14}$$

This fit measure ranges from 0 to 1. The larger is the FIT value, the more variance of the endogenous variables is accounted for by the hypothesized model. However, FIT is obviously affected by model complexity, i.e., the more parameters estimated (e.g., more clusters) the larger FIT. We thus propose another fit measure which takes into account model complexity. This fit measure is called adjusted FIT or AFIT, and is given by:

$$\text{AFIT} = 1 - (1 - \text{FIT}) \frac{d_0}{d_1},\tag{15}$$

where $d_0 = NJ$ is the degrees of freedom for the null model ($\mathbf{V}_c = \mathbf{I}$, $\mathbf{W}_c = \mathbf{0}$ and $\mathbf{A}_c = \mathbf{0}$), $d_1 = NJ - G$ is the degrees of freedom for the model being tested, where G is the number of free parameters including the unknown elements in \mathbf{W}_c and \mathbf{A}_c and all u_{ci} 's. The AFIT measure takes into account the degrees of freedom available for evaluating the model. This measure typically favors simpler models over complex models given similar explanatory power. A model that maximizes AFIT is regarded as the most appropriate one among competing models. Note that AFIT is useful as long as $G < NJ$. This implies that AFIT can be used when C is always smaller than J because the number of u_{ci} 's becomes equivalent to NJ when $C = J$. In practice, this may not be a major concern because J is usually large in path analysis with latent variables. In addition to these various model fit measures, it is necessary to investigate how well the derived clusters are separated from each other. In fuzzy clustering, a number of so-called cluster validity measures (Bezdek, 1981; Roubens, 1982) are used to examine the separation status of clusters. Based on the analysis of synthetic data, Roubens (1982) concluded that the Fuzziness Performance Index (FPI) and the Normalized Classification Entropy (NCE) are the most useful cluster validity measures for fuzzy clustering. The FPI and NCE are given by:

$$\text{FPI} = 1 - (C \times \text{PC} - 1) / (C - 1), \quad (16)$$

where PC is the Partition Coefficient (Bezdek, 1974b), defined as $\text{PC} = \frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C u_{ci}^2$, and

$$\text{NCE} = \text{PE} / \log C, \quad (17)$$

where PE is the Partition Entropy (Bezdek, 1974b), defined as $\text{PE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C u_{ci} \log u_{ci}$.

The smaller the values of FPI and NCE are, the more distinctly separated the clusters are from each other. Note that NCE and FPI can be used when $C > 1$.

FCGSCA utilizes the bootstrap method (Efron, 1982) to calculate the standard errors of parameter estimates without recourse to certain distributional assumptions such as multivariate normality. The bootstrapped standard errors can be used to assess the reliability of the parameter estimates. The critical ratios (CR) (i.e., the parameter estimates divided by their standard errors) can be used to examine the significance of the parameter estimates (e.g., roughly speaking, a parameter estimate having a critical ratio greater than two in absolute value is considered significant at .05 level).

In FCGSCA, the value of the fuzzy weight m should be selected in advance. This is common to fuzzy clustering analysis (Bezdek et al., 1981). Though $1 < m < \infty$, values too close to 1 will result in a near hard partition with all memberships close to 0 or 1. Excessively large values will lead to disproportionate overlap with all memberships close to $1/C$ (Wedel & Steenkamp, 1989). Consequently, neither of these types of m is recommended (Arabie, Carroll, DeSarbo, & Wind, 1981). Although there have been some empirical heuristic procedures to determine the value of m (e.g., McBratney & Moore, 1985; Okeke & Karnieli, 2006; Wedel & Steenkamp, 1989), there seems to exist no theoretically justifiable manner of selecting m . In practice, $m = 2$ is the most popular choice in fuzzy clustering (Bezdek, 1981; Gordon, 1999; Hruschka, 1986; Wedel & Steenkamp, 1991). In FCGSCA, $m = 2$ is also used as the default value of the fuzzy weight.

Similarly to other fuzzy clustering methods, in FCGSCA, the number of clusters C has to be *a priori* determined as well. The number of clusters can be empirically

decided by examining how the values of FIT and AFIT change with different numbers of clusters. Also, the appropriate number of clusters should result in small values of FPI and NCE. In practice, non-statistical criteria for evaluating the usefulness and relevance of clusters (e.g., cluster size, potential, interpretability, etc.) will also play a crucial role in deciding C as well (Arabie & Hubert, 1994; Wedel & Kamakura, 1998).

4. A Monte Carlo Simulation

A small-scale Monte Carlo simulation study was conducted to compare the performance of FCGSCA and FMCSA. In particular, our study focused on how well the two approaches recovered clusterwise parameters and cluster membership in different sample sizes ($N = 50, 100, 200$, and 400). For this simulation study, a latent curve (LC) model (Meredith & Tisak, 1990) was specified for two clusters (i.e., $C = 2$). Unlike conventional path-analytic models with latent variables, in the LC model, loadings are fixed as basis functions (e.g., polynomials) so as to characterize a particular temporal pattern of repeated measures, and the means of latent variables are to be freely estimated. The means and variances of latent variables in the LC model represent the intra-individual temporal change and the inter-individual differences in the temporal change of repeated assessments, respectively. The specified LC model is displayed in Figure 1.

Insert Figure 1 about here

As depicted in Figure 1, two latent variables labeled I_c and S_c were specified for three repeated measures ($\mathbf{y}_{ci} = [y_{ci1}, y_{ci2}, y_{ci3}]'$) for each cluster. In both clusters, all loadings relating I_c to the repeated measures were fixed to 1, while those relating S_c to the repeated

measures were set to 0, 1, and 2. By fixing the loadings this way, the specified model assumed that \mathbf{y}_{ci} varied in a linear trend over three time occasions in both clusters. Moreover, in this model, only means of the latent variables (α_c and β_c) were free to be estimated while their variances were fixed to 0. Thus, this two-cluster LC model can be viewed as a semiparametric group-based trajectory model (Nagin, 1999) or latent class growth analysis (Muthén, 2001). Finally, this model assumed that a time-invariant predictor (x_{ci}) had different effects on the two latent variables across two clusters. However, the variances of the residuals of the repeated measures were invariant across two clusters.

Note, this LC model can be expressed as:

$$\mathbf{y}_{ci} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_c \\ \mathbf{S}_c \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \left(\begin{bmatrix} w_{c1} \\ w_{c2} \end{bmatrix} x_{ci} + \begin{bmatrix} e_4 \\ e_5 \end{bmatrix} \right) + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix},$$

because $\begin{bmatrix} \mathbf{I}_c \\ \mathbf{S}_c \end{bmatrix} = \begin{bmatrix} w_{c1} \\ w_{c2} \end{bmatrix} x_{ci} + \begin{bmatrix} e_4 \\ e_5 \end{bmatrix}$. Then, the mean vector and covariance matrix of \mathbf{y}_{ci} are

$$\text{given by } \boldsymbol{\mu}_{ci} = \mathbf{A} \left(\begin{bmatrix} w_{c1} \\ w_{c2} \end{bmatrix} x_{ci} + \begin{bmatrix} \alpha_c \\ \beta_c \end{bmatrix} \right) = \mathbf{A} \begin{bmatrix} w_{c1} & \alpha_c \\ w_{c2} & \beta_c \end{bmatrix} \begin{bmatrix} x_{ci} \\ 1 \end{bmatrix} = \mathbf{A} \mathbf{W}_c \mathbf{x}_{ci}^* \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \kappa & 0 \\ 0 & \kappa \end{bmatrix},$$

$$\text{respectively, where } \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \mathbf{W}_c = \begin{bmatrix} w_{c1} & \alpha_c \\ w_{c2} & \beta_c \end{bmatrix}, \text{ and } \mathbf{x}_{ci}^* = \begin{bmatrix} x_{ci} \\ 1 \end{bmatrix}.$$

For our simulation study, the values of the free parameters of the two-cluster LC model were chosen as follows: $\alpha_1 = 1.5$, $\beta_1 = 0$, $w_{11} = 1.9$, $w_{12} = 0$; $\alpha_2 = 3.0$, $\beta_2 = -1.5$, $w_{21} = 2.5$, $w_{22} = -1.2$; and $\kappa = 2$. For data generation, x_{ci} was drawn from the standard normal distribution once and was considered fixed for each sample size. Then,

y_{ci} were drawn from a multivariate normal distribution $N(\mu_{ci}, \Sigma)$ with different sample sizes. Moreover, the equal number of respondents was assigned to only one of the two clusters as in *K*-means. Thus, the size of each cluster was $N/2$. One hundred Monte Carlo samples were generated for each sample size.

FCGSCA and FMCSA were applied for fitting the specified LC model to each sample. The program Mplus 2.14 (Muthén & Muthén, 2001) was used for the maximum likelihood estimation under FMCSA, while a MATLAB code was written to implement the ALS estimation procedures of FCGSCA. FCGSCA and FMCSA involve different sets of parameters: the residual variances of the repeated measures and class probabilities are free parameters to be estimated in FMCSA, while the individual-level fuzzy memberships are free parameters in FCGSCA. The common parameters to both approaches are the mean values of two latent variables and the path coefficients of x_{ci} on the latent variables (i.e., α_c , β_c , w_{c1} , and w_{c2}). The same random starting values, given by Mplus, were used for the common parameters whereas the true parameter values were employed for the starting values for the other non-common parameters of the two approaches (e.g., class probability = .5 in FMCSA). For FCGSCA estimation, $m = 2$ was chosen as the fuzzy weight value. To evaluate the overall recovery of parameters under the two approaches, the Root Mean Square Error (RMSE) was computed. Besides the recovery of parameters, we also computed the average rate of misclassification of respondents under each sample size for the approaches. Furthermore, we investigated the occurrences of non-convergence in each approach.

Table 1 provides the mean estimates of the common parameters, the mean scores of RMSE, the average rate of misclassification, and the number of converged cases obtained from FCGSCA and FMCSA across the different sample sizes.

Insert Table 1 about here

As shown in the table, FCGSCA was found to converge in all samples across the different sample sizes. On the other hand, FMCSA suffered from the occurrences of non-convergence in all sample sizes, although they rapidly decreased with increasing sample size. Moreover, the mean scores of RMSE obtained from FCGSCA appeared to be consistently lower than those from FMCSA in all sample sizes. More specifically, the difference in RMSE between the two approaches was significant when $N = 50$ ($F_{1, 141} = 55.09$, $p = .000$), $N = 100$ ($F_{1, 180} = 25.67$, $p = .000$), and $N = 200$ ($F_{1, 188} = 8.68$, $p = .004$). However, no significant difference in RMSE was found between the approaches when $N = 400$ ($F_{1, 197} = 2.24$, $p = .14$). Finally, the misclassification rates from FCGSCA were much lower than that from FMCSA across all sample sizes, although they tended to decrease in both approaches as sample size increased.

In sum, it was found that the parameter recovery of FCGSCA was superior to that under FMCSA in small samples (< 400). FMCSA resulted in non-convergence, especially in small samples. Moreover, the classification of respondents under FCGSCA was more accurate than for FMCSA in all sample sizes. Although it is not reported to preserve space, the congruence coefficient (Tucker, 1951) was also calculated as an overall measure of similarity between parameters and their estimates. Even when $N = 50$,

the mean congruence coefficient from FCGSCA became larger than .90 which is an acceptable degree of congruence (Mulaik, 1972). Besides the superior performance compared to FMCSA, thus, the proposed method itself seemed to provide a sufficient level of parameter recovery in small samples. This also suggests that the default choice of the fuzzy weight, i.e., $m = 2$, is satisfactory in obtaining solutions from the proposed method.

5. Empirical Application

The present example is part of a longitudinal survey of substance use among adolescents from two northwestern urban areas in the US (Duncan, Duncan, Alpert, Hops, Stoolmiller, & Muthén, 1997). The sample consists of 632 adolescents measured on their use of alcohol over four points in time. Alcohol use was assessed by a single self-report item with 5 response options: (1) life time abstainers, (2) 6-month abstainers, (3) current use of less than four times a month, (4) current use of between 4 and 29 times a month, and (5) current use of 30 or more times a month. Five additional variables were measured once at the initial point in time: parental marital status, family status, socio-economic status (SES), gender, and age. Parental marital status was classified as follows: 0 = single and 1 = married or living in a committed relationship. Family status was categorized as follows: 0 = step or foster families and 1 = others. SES was calculated as the average of parental annual income and education level. Parental annual income was assessed based on a 16-point scale ranging from '6,000 dollars and below' to '50,000 dollars or more'. Education levels range from 'Grade level 6 or less' to 'Graduate level'. Male and female were coded as 0 and 1, respectively.

A latent curve model was specified for identifying both intra-individual change in alcohol use over time and inter-individual differences in temporal change; and simultaneously, for examining the effects of time-invariant exogenous variables on temporal change. From descriptive statistics, it was found that the mean levels of alcohol use increased monotonically over time ($t_1 = 2.23$, $t_2 = 2.46$, $t_3 = 2.65$, $t_4 = 2.94$). This suggests that there existed a linear trend of change in the consumption of alcohol over the four assessments. Moreover, the response variable of alcohol use was merely measured over four points in time. This relatively small number of measurements may be insufficient to reveal a complex non-linear temporal pattern in the response variable (MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997). In fact, previous studies with the same data also supported such a linear trend of change in alcohol use over the four measurement intervals (e.g., Duncan et al., 1997). Accordingly, a linear-trend LC model was specified for the data as depicted in Figure 2.

Insert Figure 2 about here

The interpretations of the parameters of the specified LC model parallel those of the model analyzed in the simulation study. That is, the means of I_c and S_c characterize the initial status and growth rate of the intra-individual change in alcohol use over time, respectively. The variances of I_c and S_c represent the inter-individual differences in the initial status and growth rate of the temporal change, respectively. In this model, both means and variances of the latent variables were free to be estimated. Moreover, the

specified model included the effects of the five time-invariant variables on the two latent variables involving temporal change.

FMCSA and FCGSCA were employed to fit this LC model to the data under different numbers of clusters. Again Mplus 2.14 (Muthén & Muthén, 2001) was used to estimate the parameters and calculate goodness-of-fit measures of the multiple-cluster LC models within the framework of FMCSA. A MATLAB program was written for the FCGSCA-based estimation.

In these multiple-cluster analyses, FMCSA failed to converge when the number of clusters was larger than six. The values of information criteria for the LC models with $C \leq 6$ were compared to investigate whether or not the number of clusters might be determined within the converged models. Table 2 provides the values of information criteria for the different multiple-cluster LC models.

Insert Table 2 about here

As shown in Table 2, the values of AIC and the sample-size adjusted BIC (SABIC) monotonically decreased until $C = 6$, while the minimum value of BIC was obtained when $C = 1$. It seems unwise to determine the number of clusters based solely on either the AIC or BIC because the AIC tends to overestimate the number of clusters present whereas the BIC is likely to underestimate the number of clusters (Bauer & Curran, 2003; McLachlan & Peel, 2000). Although the SABIC is shown to be superior to the AIC and BIC in the context of latent class analysis (Yang, 1998), it appears unsafe to conclude that $C = 6$ was the appropriate number of clusters present in the data because there was no

means to ensure that seven- or higher-cluster LC models involved larger values of SABIC than the six-cluster LC model. The value of entropy (EN; Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993) also consistently increased until $C = 6$. Again, it was impossible to assure that this six-cluster model involved the true maximum EN value due to the non-convergence of FMCSA after $C = 6$. Thus, it was difficult to make a clear decision on the appropriate number of clusters for the data within the framework of FMCSA. This difficulty was largely due to the failure of convergence.

On the other hand, FCGSCA did not suffer from the problem of non-convergence. Table 2 also provides the values of FIT, AFIT, FPI, and NCE for different multiple-cluster LC models obtained from FCGSCA. It was shown that the values of FIT increased gradually beyond $C = 2$, suggesting that no substantial changes in FIT were obtained by having more than two clusters. This is consistent with a value of AFIT which was maximized when $C = 2$. Furthermore, the minimum values of both FPI and NCE were obtained at $C = 2$. Thus, $C = 2$ was adopted for further analyses.

Model fitting procedures of FCGSCA for the two-cluster LC model showed that the mean estimates of I_1 and S_2 were .77 (s.e. = .02, CR = 38.5) and .18 (s.e. = .01, CR = 18.0), respectively, in cluster 1. On the other hand, the mean estimates of I_2 and S_2 were 1.38 (s.e. = 0.02, CR = 69.0) and .08 (s.e. = .01, CR = 8.0), respectively, in cluster 2. Both clusters indicated significant levels of alcohol use at the initial status and also significant linear growth in alcohol use over the four time points. In addition, the two clusters involved substantively different temporal patterns of intra-individual change: The respondents in cluster 1 were likely to show relatively low levels of alcohol use at the initial assessment while their use of alcohol tended to increase at a higher rate over the

four time points. On the other hand, the respondents in cluster 2 involved a relatively high level of alcohol use at the first assessment, but a lower rate of increase in alcohol use over the four time points. Thus, cluster 1 can be characterized as ‘a low initial level of alcohol use accompanied by a high growth rate’ and cluster 2 as ‘a high initial level of alcohol use accompanied by a low growth rate’. FCGSCA also showed that the variance estimates of I_1 and S_1 were .11 (s.e. = .00, CR = 12.5) and .03 (s.e. = .00, CR = 12.9), respectively and those of I_2 and S_2 were 1.0 (s.e. = .06, CR = 18.1) and .01 (s.e. = .00, CR = 11.7), respectively. This indicates that both clusters involved substantial inter-individual differences in the initial status and the growth rate of alcohol use.

Table 3 provides the path coefficients of the five exogenous variables estimated from FCGSCA for each cluster. The first column of the table under the label of I_c exhibits the effects of the exogenous variables on alcohol use at the initial status, and the second column under the label of S_c displays their effects on the linear growth rate of alcohol use over time. The two clusters seemed to show distinct consequences of the exogenous variables from each other. For example, in cluster 1, family status had a significant and positive effect on I_1 ($w_{13} = .15$, s.e. = .07, CR = 2.1), indicating that adolescents living with other families rather than step or foster families displayed higher levels of alcohol use at the initial status. Also, age showed a significant and positive effect on I_1 ($w_{19} = .09$, s.e. = .01, CR = 9.0), indicating that older adolescents displayed higher levels of alcohol use at the initial status. On the other hand, in cluster 2, besides the similar effects of the two variables, parental marital status also had a significant and negative impact on I_2 ($w_{21} = -.25$, s.e. = .11, CR = -2.3), suggesting higher initial levels of alcohol use by adolescents living with single parents compared to those living with both parents.

In cluster 1, SES had a significant and negative impact on S_1 ($w_{16} = -.09$, s.e. = .02, CR = -4.5) suggesting that socially and economically more disadvantaged adolescents were likely to show a higher rate of increase in alcohol use than those less disadvantaged. Age exhibited significant and positive effects on S_1 ($w_{110} = .02$, s.e. = .00, CR = 6.3), indicating that older adolescents tended to increase use of alcohol at a higher rate compared to younger adolescents. On the other hand, in cluster 2, family status showed a significant and positive impact on S_2 ($w_{24} = .05$, s.e. = .02, CR = 2.5). It suggests that the alcohol use of adolescents living with other families rather than step or foster families increased at a higher rate during the study. When all respondents were assigned to the cluster associated with the highest membership value, the sizes of clusters 1 and 2 arrived at 293 (46%) and 339 (54%), respectively.

Insert Table 3 about here

To summarize, FMCSA and FCGSCA were employed to fit multiple-cluster LC models to alcohol use data. FMCSA failed to converge when the number of clusters became larger than 6. This, in turn, impeded the choice on the appropriate number of clusters. On the other hand, FCGSCA did not suffer from such a convergence problem, and enabled the determination that $C = 2$ represented the reasonable number of clusters based on model fit heuristics. In particular, the two-cluster LC model estimated by FCGSCA seemed to reveal substantively distinct developmental trajectories on alcohol use within the two groups of adolescents. That is, one cluster showed a lower level of alcohol use at the initial status but a higher rate of increase in alcohol use over time,

while the other cluster displayed a higher level of alcohol use at the initial status but a lower rate of growth in alcohol use. Moreover, there existed substantial inter-individual differences in such different temporal patterns of change in both clusters. Furthermore, the influences of the time-invariant exogenous variables on the two latent variables of temporal change were shown to be different across the two clusters. Therefore, the two-cluster LC model appears to be helpful in studying qualitatively different longitudinal processes on alcohol use and different consequences of its antecedents across two relatively heterogeneous subgroups of adolescents.

6. Conclusion

An extension of the aggregate sample GSCA was proposed that effectively captures cluster-level respondent heterogeneity. The proposed method simultaneously estimates a fuzzy classification of respondents into heterogeneous groups and estimates GSCA parameters to each of the groups in a unified way. The performance of the proposed method was compared to the extant finite mixture approach with synthetic data with different sample sizes. The analyses of synthetic data proved the superiority of the proposed method in recovering parameters within clusters and determining the number of clusters, especially in small samples. In addition, the usefulness of the proposed method was empirically demonstrated by the analysis of longitudinal data on substance use data of adolescents. The capability of the proposed method to capture respondent heterogeneity can contribute to strengthening the generality of GSCA, thereby giving rise to a broader range of applications in various fields of inquiry.

The proposed method may be further refined and extended so as to enhance its data-analytic capability and the scope of its applicability. For instance, although it is not a

problem unique to this method, more formal rules are needed for determining the fuzzy weight. Similarly, a fruitful research area would be to investigate the influences of different values of the fuzzy weight on obtaining solutions in FCGSCA, although the default choice of the fuzzy weight value ($m = 2$) was found to be satisfactory from the simulation study. The proposed method is currently geared for the analysis of continuous variables. It may be effectively extended so as to deal with discrete variables through data transformations. In particular, the optimal scaling approach (Gifi, 1990; Young, 1981) is deemed promising because it can be readily coupled with the GSCA estimation procedure in the first step of the proposed estimation algorithm. Moreover, the proposed method can be viewed as a post-hoc classification approach in that it identifies clusters of respondents through the analysis of the data (Wind, 1978; Wedel & Kamakura, 1998). In many cases, respondent heterogeneity can also be addressed by classifying respondents into clusters *a priori* on the basis of external information about respondents such as demographic variables. This *a priori* classification approach may be incorporated into the proposed method for more sophisticated analyses. This extension may involve combining the proposed method with multilevel analysis (e.g., Bock, 1989; Hox, 1995; Raudenbush & Bryk, 2002; Goldstein, 1987) where within each of fuzzy clusters, the GSCA parameters are estimated, taking into account such *a priori* grouping information. All of these possibilities warrant further attention and provide the fodder for future theoretical and empirical work.

Appendix: The ALS Algorithm for GSCA

Let $\Psi = ZV$ and $\Gamma = ZW$. The ALS algorithm repeats the following steps until convergence is reached.

Step 1: Update A for fixed W . Criterion (6) can be re-written as:

$$\phi = SS(\text{vec}(\Psi) - \text{vec}(\Gamma A)) = SS(\text{vec}(\Psi) - (\mathbf{I} \otimes \Gamma) \text{vec}(A)), \quad (\text{A.1})$$

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

$$\hat{\mathbf{a}} = (\Phi' \Phi)^{-1} \Phi' \text{vec}(\Psi). \quad (\text{A.2})$$

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

Step 2: 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 $p = t + J$ ($t = 1, \dots, T$). Let $\Lambda = WA$. Let $V_{(-p)}$ denote V whose p -th column is the vector of zeros. Let $\Lambda_{(-t)}$ 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 I 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 also be re-written as:

$$\phi = \sum_{t=1}^T SS((\beta' \otimes Z) \mathbf{w}_t - \text{vec}(Z \Lambda)). \quad (\text{A.3})$$

In (A.3), β and Δ are defined as follows:

$$\boldsymbol{\beta} = \mathbf{m}_{(p)} - \mathbf{a}_{(t)}, \quad (\text{A.4})$$

and

$$\boldsymbol{\Delta} = \boldsymbol{\Lambda}_{(-t)} - \mathbf{V}_{(-p)}. \quad (\text{A.5})$$

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 $(\boldsymbol{\beta}' \otimes \mathbf{Z})$ corresponding to the fixed elements in \mathbf{w}_t . Then, the least-squares estimate of $\boldsymbol{\theta}_t$ is obtained by:

$$\hat{\boldsymbol{\theta}}_t = (\boldsymbol{\Xi}' \boldsymbol{\Xi})^{-1} \boldsymbol{\Xi}' \text{vec}(\mathbf{Z} \boldsymbol{\Delta}). \quad (\text{A.6})$$

The updated \mathbf{w}_t is recovered from $\hat{\boldsymbol{\theta}}_t$.

References

- Arabie, P., Carroll, J. D., DeSarbo, W. S., & Wind, J. (1981). Overlapping clustering: A new method for product positioning. *Journal of Marketing Research*, 18, 310-317.
- Arabie, P., & Hubert, L. (1994). Cluster analysis in marketing research. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 160-189). Oxford: Blackwell.
- Bagozzi, R. P. (1982). A field investigation of causal relations among cognition, affect, intentions, and behavior. *Journal of Marketing Research*, 19, 562-584.
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338-363.
- Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, 45, 3, 289-308.
- Bezdek, J. C. (1974a). Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1, 57-71.
- Bezdek, J. C. (1974b). Cluster validity with fuzzy set. *Journal of Cybernetics*, 3, 58-72.
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.
- Bezdek, J. C., Coray, C., Gunderson, R., & Watson, J. (1981). Detection and characteristics of cluster substructure. II. Fuzzy c-varieties and convex combinations thereof. *SIAM Journal on Applied Mathematics*, 40, 358-372.
- Bock, R. D. (1989). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.

- Bock, R. D., & Bargmann, R. E. (1966). Analysis of covariance structures. *Psychometrika*, 31, 507-534.
- Chin, W. W. (1998). Issues and opinion on structural equation modeling. *Management Information Systems Quarterly*, 22, 7-16.
- 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.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- DeSarbo, W. S., & Cron, W. L. (1988). A conditional mixture maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5, 249-289.
- DeSarbo, W. S., Grewal, R., & Hwang, H. (2006). A clusterwise bilinear multidimensional scaling methodology for marketing research: An application to the estimation of strategic groups. Unpublished manuscript.
- Duncan, T. E., Duncan, S. C., Alpert, A., Hops, H., Stoolmiller, M., & Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research*, 32, 275-318.
- Dunn, J. C. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3, 32-57.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: SIAM.

- Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis*, 4th ed. London: Arnold Press.
- Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 19, 440-452.
- Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: Wiley.
- Goldstein, H. I. (1987). *Multilevel models in educational and social research*. London: Oxford University.
- Gordon, A. D. (1999). *Classification*. London: Chapman & Hall/CRC.
- Hathaway, R. J., & Bezdek, J. C. (1993). Switching regression models and fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 1, 195-204.
- Heiser, W. J., & Groenen, P. J. F. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, 62, 63-83.
- Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: T-T Publikaties.
- Hruschka, H. (1986). Market definition and segmentation using fuzzy clustering methods. *International Journal of Research in Marketing*, 3, 117-134.
- Hwang, H., & Takane, Y. (2002). Structural equation modeling by extended redundancy analysis. In S. Nishisato, Y. Baba, H. Bozdogan and K. Kanefuji (Eds.), *Measurement and multivariate analysis* (pp. 115-124), Tokyo: Springer Verlag.
- Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, 69, 81-99.

- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39-59.
- Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57, 409-426.
- Jöreskog, K. G. (1973). A generating method for estimating a linear structural equation system. In A. S. Goldberger and O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences* (pp. 85-112). New York: Academic Press.
- Kamakura, W. A., Kim, B., & Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, 15, 152-172.
- MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32, 215-253.
- Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). *Statistical applications using fuzzy sets*. New York: John Wiley & Sons.
- McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
- McBratney, A. B., & Moore, A. W. (1985). Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 35, 165-185.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: John Wiley & Sons.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122.

- Moffitt, T. E. (1993). Adolescent-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100, 674-701.
- Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
- Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1- 33). Mahwah, NJ: Erlbaum.
- Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide (Second edition)*. Los Angeles, CA: Muthén & Muthén.
- Nagin, D. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, 4, 139-157.
- Okeke, F., & Karnieli, A. (2006). Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm. *Ecological Informatics*, 1, 117-124.
- Ramaswamy, V., DeSarbo, W. S., Reibstein, D., & Robinson, W. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12, 103-124.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Second edition. Thousand Oaks, CA: Sage Publications.
- Roubens, M. (1982). Fuzzy clustering algorithms and their cluster validity. *European Journal of Operational Research*, 10, 294-301.
- Tucker, L. R. (1951). A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984). Washington, D.C.: U.S. Department of the Army.

- Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting appropriate procedure. *Multivariate Behavioral Research*, 25, 1-28.
- Wedel, M., & Kamakura, W. A. (1998). *Market segmentation: Conceptual and methodological foundations*. Boston: Kluwer Academic Publishers.
- Wedel, M., & Steenkamp, J.-B. E. M. (1989). Fuzzy clusterwise regression approach to benefit segmentation. *International Journal of Research in Marketing*, 6, 241-258.
- Wedel, M., & Steenkamp, J.-B. E. M. (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research*, 28, 385-396.
- Wind, Y. (1978). Issues and advances in segmentation research. *Journal of Marketing Research*, 15, 317-337.
- Yang, C. C. (1998). Finite mixture model selection with psychometric applications. Unpublished doctoral dissertation. University of California, Los Angeles.
- Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, 46, 357-388.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353.

Table 1. The mean parameter estimates, the mean scores of RMSE, the average rate of misclassification, and the number of converged cases obtained from FCGSCA and FMCSA across different sample sizes in the simulation study.

	N = 50		N = 100		N = 200		N = 400	
Parameter	FCGSCA	FMCSA	FCGSCA	FMCSA	FCGSCA	FMCSA	FCGSCA	FMCSA
$\alpha_1 = 1.5$	1.48	1.57	1.47	1.57	1.43	1.56	1.45	1.52
$\beta_1 = 0$	0.02	-0.09	0.04	-0.04	0.03	-0.01	0.01	0.00
$w_{11} = 1.9$	1.88	1.78	1.86	1.87	1.93	1.93	1.95	1.86
$w_{12} = 0$	-0.01	0.11	0.05	0.03	-0.04	-0.03	-0.01	0.03
$\alpha_2 = 3.0$	3.01	2.34	3.01	2.77	3.06	3.01	3.06	2.98
$\beta_2 = -.15$	-0.16	-0.05	-0.15	-0.14	-0.16	-0.18	-0.15	-0.15
$w_{21} = 2.5$	2.44	2.03	2.49	2.36	2.43	2.50	2.44	2.51
$w_{22} = -.12$	-0.12	-0.02	-0.11	-0.07	-0.11	-0.11	-0.10	-0.11
RMSE	0.37	1.25	0.26	0.44	0.18	0.24	0.13	0.14
MR	4.0%	28.3%	2.7%	17.2%	2.2%	10.8%	1.3%	7.5%
NC	100	43	100	82	100	90	100	99

RMSE = Root mean square error
MR = Average misclassification rate
NC = Number of converged cases

Table 2. The values of model selection criteria for different multiple-cluster latent curve models for the alcohol use data obtained from FCGSCA and FMCSA.

	FCGSCA	FMCSA
C = 1	FIT = .9616 AFIT = .9615	AIC = 5939.200 BIC = 6023.729 SABIC = 5963.406
C = 2	FIT = .9727 AFIT = .9647 FPI = 0.6864 NCE = 0.7440	AIC = 5883.586 BIC = 6030.400 SABIC = 5925.628 EN = .641
C = 3	FIT = .9741 AFIT = .9608 FPI = 0.7903 NCE = 0.8058	AIC = 5861.249 BIC = 6070.347 SABIC = 5921.127 EN = .727
C = 4	FIT = .9759 AFIT = .9560 FPI = 0.8217 NCE = 0.8195	AIC = 5829.151 BIC = 6100.533 SABIC = 5906.865 EN = .727
C = 5	FIT = .9779 AFIT = .9492 FPI = 0.8309 NCE = 0.8211	AIC = 5779.771 BIC = 6113.437 SABIC = 5875.320 EN = .730
C = 6	FIT = .9788 AFIT = .9343 FPI = 0.8458 NCE = 0.8277	AIC = 5751.478 BIC = 6147.429 SABIC = 5864.863 EN = .798
C = 7	FIT = .9805 AFIT = .9074 FPI = 0.8442 NCE = 0.8228	No Convergence

Table 3. The path coefficient estimates and their standard errors in the parenthesis of the two-cluster latent curve model specified for the alcohol use data obtained from FCGSCA.

		I _c	S _c
Cluster 1	Parental marital status	-.15 (.13)	.09 (.05)
	Family status	.15 (.07)	.05 (.03)
	SES	-.09 (.08)	-.09 (.02)
	Gender	-.01 (.06)	.04 (.02)
	Age	.09 (.01)	.02 (.00)
Cluster 2	Parental marital status	-.25 (.11)	-.03 (.03)
	Family status	.30 (.08)	.05 (.02)
	SES	.02 (.06)	.02 (.02)
	Gender	.03 (.06)	.03 (.02)
	Age	.19 (.01)	.00 (.00)

Figure 1. The latent curve model specified for the Monte Carlo simulation study.

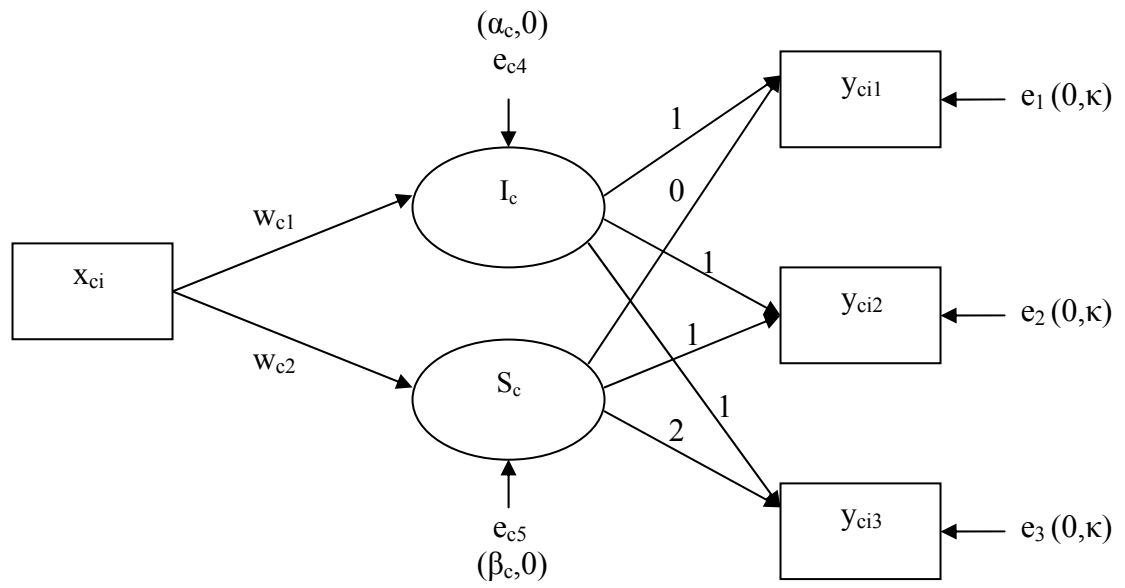


Figure 2. The latent curve model specified for the alcohol use data.

