

A multivariate latent class profile analysis for longitudinal data with a latent group variable

Jung Wun Lee^a, Hwan Chung^{1,b}

^aDepartment of Statistics, University of Connecticut, USA;

^bDepartment of Statistics, Korea University, Korea

Abstract

In research on behavioral studies, significant attention has been paid to the stage-sequential process for multiple latent class variables. We now explore the stage-sequential process of multiple latent class variables using the multivariate latent class profile analysis (MLCPA). A latent profile variable, representing the stage-sequential process in MLCPA, is formed by a set of repeatedly measured categorical response variables. This paper proposes the extended MLCPA in order to explain an association between the latent profile variable and the latent group variable as a form of a two-dimensional contingency table. We applied the extended MLCPA to the National Longitudinal Survey on Youth 1997 (NLSY97) data to investigate the association between of developmental progression of depression and substance use behaviors among adolescents who experienced Authoritarian parental styles in their youth.

Keywords: categorical latent variable, latent class analysis, latent stage-sequential process, longitudinal data, recursive EM algorithm

1. Introduction

Latent class analysis (LCA) is one type of finite mixture model that can be applied for a set of discrete response random variables. It summarizes the structure of population distribution by defining several partitions of population (latent classes) which cannot be observed directly. The latent class membership may be discovered by identifying a small number of representative response patterns of manifest variables. The LCA framework has been expanded to be utilized for more complicated data structures such as repeatedly measured data (Chung *et al.*, 2011), vectorized joint structure (Jeon *et al.*, 2017), and a hierarchical group-outcome structure (Lee and Chung, 2017).

A new type of LCA has been proposed in this article to deal with a latent group variable in a multivariate latent class profile analysis (MLCPA). Suppose that we have several sets of categorical item response variables that are repeatedly measured across the time stages. One set of item variables defines a discrete latent group variable via conventional LCA framework called a 'latent group variable.' The other sets of item response variables discover a sequence of multiple latent class variables to divide the population into homogeneous subgroups of those who have similar sequential patterns of latent class membership. We call this sequence of latent class variables a 'latent profile variable'. Individuals with same latent profile membership will show similar sequential patterns for each latent class variable over time. Further, our proposed model reveals the association between the latent profile

¹ Corresponding author: Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea.
E-mail: hwanch@korea.ac.kr

variable and the latent group variable using the conditional prevalence of latent profiles given in the latent group membership.

Parenting styles had been defined as four major categories: indulgent, permissive, authoritative, and authoritarian with respect to the characteristics of parent's distance and responsibilities toward children (Maccoby and Martin, 1983). King *et al.* (2016) suggested that the authoritarian parental style had a positive effect on children's drug-taking behavior and/or depression. For the real data analysis, we focused on the individuals who experienced authoritarian parental style in their youth in order to explore stage-sequential processes of drug-taking behavior and its association with depression symptoms. We used three sets of item variables measured over three time periods (2000, 2002, and 2004) from the National Longitudinal Survey on Youth 1997 (NLSY97) data. A set of item response variables for depression was measured in 2000, and additional two sets of categorical response variables related with alcohol drinking and cigarette/marijuana smoking were obtained over three time periods from 2000 to 2004. The proposed model can be used to discover the meaningful subgroup of population in terms of depression (latent group variable) and the joint latent-stage sequential process (latent profile variable) of alcohol drinking and cigarette/marijuana smoking. We also investigate the association between the latent profile variable and the latent group variable in terms of conditional prevalence. The covariate effects such as gender or race on the prevalence of latent profiles will be examined in forms of baseline multinomial logistic regression.

The rest of this article is as follows. The description of our proposed model and the estimation methods for the model parameters are presented in Sections 2 and 3. In Section 3.3, we examined the parameter estimation and inference procedure through empirical simulation; the simulation results are available in Appendix. In Section 4, we illustrate the practical usefulness of our new model by analyzing data from the NLSY97 using discrete item variables related to adolescents depression and drug-taking behavior such as alcohol drinking, cigarette smoking, and marijuana use. In Section 5, we summarize the paper and discuss future research areas.

2. Model

2.1. Latent class analysis

A LCA is a classical methodology that divides the population into homogenous subgroups with respect to response patterns for manifest items. It postulates that a distribution of a set of categorical random variable is a mixture of finite classes with respective response patterns. Suppose there are P categorical manifest items Z_1, \dots, Z_P . The responses of each manifest item for the i^{th} individual are obtained as a P -dimensional vector $\mathbf{z}_i = [z_{i1}, \dots, z_{iP}]$, where z_{ip} can take any value from $1, \dots, r_p$ for $p = 1, \dots, P$. Let the latent class variable D have G categories, then the observed-data likelihood of LCA for the i^{th} individual can be written as

$$\begin{aligned} P(\mathbf{Z} = \mathbf{z}_i) &= \sum_{d=1}^G P(D = d, \mathbf{Z} = \mathbf{z}_i) = \sum_{d=1}^G P(D = d) P(\mathbf{Z} = \mathbf{z}_i | D = d) \\ &= \sum_{d=1}^G P(D = d) \prod_{p=1}^P P(Z_p = z_{ip} | D = d) = \sum_{d=1}^G \delta_d \prod_{p=1}^P \prod_{h=1}^{r_p} \phi_{ph|d}^{I(z_{ip}=h)}, \end{aligned} \quad (2.1)$$

where $I(z_{ip} = h)$ is the indicator function which is 1 when $z_{ip} = h$ and 0 otherwise. The likelihood of LCA given in (2.1) is constructed under local independence assumptions, implying that the manifest items are conditionally independent when a latent class membership is given. Here, $\phi_{ph|d} = P(Z_p = h | D = d)$, referred as the primary measurement parameter, explains the relationship between the

latent class and the p^{th} manifest item, and $\delta_d = P(D = d)$ represents the proportion of latent class membership d . Since all parameters in (2.1) are conditional probabilities, the sum-to-one and non-negative constraints are explicit ($\sum_{d=1}^G \delta_d = 1$ and $\sum_{h=1}^{r_p} \phi_{ph|d} = 1$ for $p = 1, \dots, P$ and $d = 1, \dots, G$).

2.2. Multivariate latent class profile analysis

MLCPA has been introduced to explain the longitudinal patterns of latent class membership when there are multiple latent class variables (Lee *et al.*, 2019). In MLCPA, each set of manifest items measures a categorical latent variable, and the sequential patterns of identified latent classes are summarized by a latent profile variable. As a result, observations with the same latent profile membership will share common sequential patterns of joint latent class membership. In this manner, MLCPA provides a statistical tool that allows researchers to discover a meaningful subgroup based on the representative sequential pattern of unobservable memberships for several latent class variables.

Let C_{jt} denote the j^{th} latent class variable having K_j nominal categories at time stage t for $j = 1, \dots, J$ and $t = 1, \dots, T$. For each time stage, a vector of J latent class variables $\mathbf{C}_t = [C_{1t}, \dots, C_{Jt}]'$ can be represented as a contingency table with $\prod_{j=1}^J K_j$ cells, showing all possible combinations of class memberships. Therefore, T -sequences of J latent class variables will be written in a contingency table with $(\prod_{j=1}^J K_j)^T$ cells. Among all possible combinations of sequential patterns, MLCPA discovers the representative sequential patterns and categorizes them as latent profiles. Let the latent profile variable U have S nominal categories describing the most common stage-sequential patterns of J latent class memberships over T points in time. Let $\mathbf{Y}_t = [\mathbf{Y}_{1t}, \dots, \mathbf{Y}_{Jt}]$, where $\mathbf{Y}_{jt} = [Y_{1jt}, \dots, Y_{M_jjt}]'$ be a set of J vectors of discrete responses to M_j items to measure the j^{th} latent class variable at stage t , where each variable Y_{m_jjt} can take any value from 1 to r_{m_j} for $m_j = 1, \dots, M_j$ and $j = 1, \dots, J$. Then, the complete-data likelihood for the i^{th} individual (the joint probability of the latent profile $U = u$, the latent class membership $\mathbf{c} = [\mathbf{c}_1, \dots, \mathbf{c}_T]$, and the responses $\mathbf{y}_i = [\mathbf{y}_{i1}, \dots, \mathbf{y}_{iT}]$) will be

$$\begin{aligned} L_i^* &= P(U = u, \mathbf{C} = \mathbf{c}, \mathbf{Y} = \mathbf{y}_i) \\ &= P(U = u)P(\mathbf{C} = \mathbf{c} | U = u)P(\mathbf{Y} = \mathbf{y}_i | \mathbf{C} = \mathbf{c}) \\ &= P(U = u) \prod_{t=1}^T \prod_{j=1}^J \left[P(C_{jt} = c_{jt} | U = u) \prod_{m_j=1}^{M_j} P(Y_{m_jjt} = y_{im_jjt} | C_{jt} = c_{jt}) \right] \\ &= \gamma_u \prod_{t=1}^T \prod_{j=1}^J \left[\eta_{c_{jt}|u}^{(j,t)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{k_{m_j}} \left(\rho_{m_jk|c_{jt}}^{(j,t)} \right)^{I(y_{im_jjt}=k)} \right], \end{aligned} \quad (2.2)$$

where $I(y_{im_jjt} = k)$ is the indicator function which is 1 when $y_{im_jjt} = k$ and 0 otherwise. Note that we assume $P(\mathbf{Y} = \mathbf{y}_i | \mathbf{C} = \mathbf{c}, U = u) = P(\mathbf{Y} = \mathbf{y}_i | \mathbf{C} = \mathbf{c})$ in (2.2). This assumption states that the latent profile variable is identified only by the latent class variable. The meaning of the profile can be interpreted only by the sequential patterns of latent class memberships; in addition, responses to the manifest items only work for identifying the latent class variable.

2.3. Multivariate latent class profile analysis with latent group variable

Combining the LCA structure as group variable with MLCPA structure as an outcome, we propose MLCPA with latent group variable and refer it as GLCPA. The GLCPA postulates that the distribution of latent profile variable can be affected by another latent class variable which can be identified through the LCA structure. The proposed model is illustrated in Figure 1. A sequence of J latent class

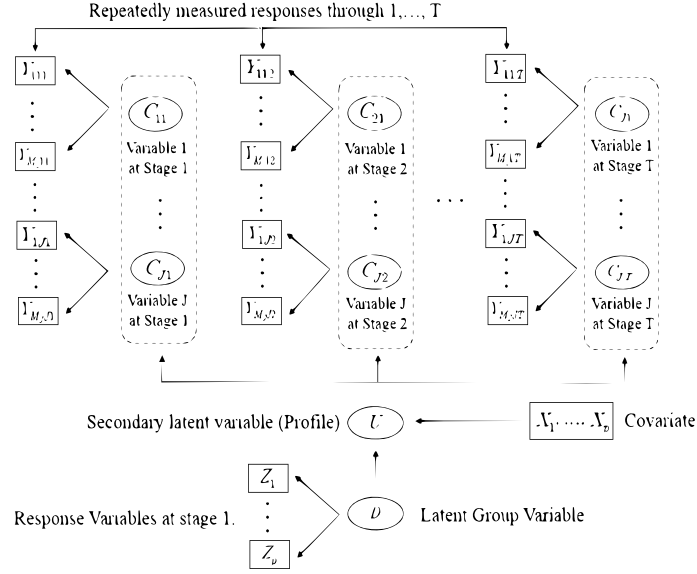


Figure 1: A diagram of multivariate latent class profile analysis with latent group variable (GLCPA).

variables $\mathbf{C}_t = [C_{1t}, \dots, C_{Jt}]'$ for $t = 1, \dots, T$ is associated through latent profile variable U , and each latent class variable C_{jt} is identified by the j^{th} set of manifest items $\mathbf{Y}_{jt} = [Y_{1jt}, \dots, Y_{M_{jt}t}]'$ at time stage t . Latent group variable D is the discrete latent class variable identified through the manifest items $\mathbf{Z} = [Z_1, \dots, Z_p]$. As discussed in Section 1, the distribution of outcome latent profile variable U is affected by the latent group membership $D = d$.

Using the notation given in (2.2), the complete-data likelihood of the GLCPA for the i^{th} observation is written as

$$\begin{aligned}
 L_i^* &= P(U = u, D = d, \mathbf{C} = \mathbf{c}, \mathbf{Y} = \mathbf{y}_i, \mathbf{Z} = \mathbf{z}_i) \\
 &= P(D = d) \prod_{p=1}^P P(Z_p = z_{ip} | D = d) \\
 &\quad \times P(U = u | D = d) \prod_{t=1}^T \prod_{j=1}^J \left[P(C_{jt} = c_{jt} | U = u) \prod_{m_j=1}^{M_j} P(Y_{m_jjt} = y_{im_jjt} | C_{jt} = c_{jt}) \right] \\
 &= \gamma_{u|d} \prod_{t=1}^T \prod_{j=1}^J \left[\eta_{c_{jt}|u}^{(j,t)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{k_{m_j}} \left(\rho_{m_jk|c_{jt}}^{(j,t)} \right)^{I(y_{im_jjt}=k)} \right] \delta_d \prod_{p=1}^P \prod_{h=1}^{r_p} \phi_{ph|d}^{I(z_{ip}=h)}. \tag{2.3}
 \end{aligned}$$

The complete likelihood in (2.3) consists of five parameters:

- (a) $\rho_{m_jk|c_{jt}}^{(j,t)} = P(Y_{m_jjt} = k | C_{jt} = c_{jt})$ denotes the probability of the response k to the $(m_j)^{\text{th}}$ item measuring the j^{th} latent class variable C_{jt} , for a given class c_{jt} of the j^{th} latent class variable C_{jt} at stage t .
- (b) $\phi_{ph|d} = P(Z_p = h | D = d)$ denotes the probability of the response h to the p^{th} item measuring the latent group variable D , for a given latent group membership d .

- (c) $\eta_{c_{jt}|u}^{(j,t)} = P(C_{jt} = c_{jt} | U = u)$ denotes the conditional probability of belonging to latent class c_{jt} , the class membership of the j^{th} latent class variable C_{jt} at stage t , when a latent profile variable U has a profile membership u .
- (d) $\gamma_{u|d} = P(U = u | D = d)$ denotes the probability that individual has a latent profile membership u , given latent group membership d .
- (e) $\delta_d = P(D = d)$ denotes the probability that individual has a latent group membership $D = d$.

The primary measurement parameters (ρ s and ϕ s) identify the underlying categorical latent class and group variables, respectively. The secondary measurement parameter η depicts the relationship between latent class variable C_{jt} and latent profile variable U . Each identified latent profile can be explained through a set of estimated secondary measurement parameters as individual sequential patterns of changing latent class membership.

The GLCPA assumes the following conditions: (1) the latent profile variable U is related with manifest items Y_{jt} only through the class membership of each latent class variable C_{jt} , (2) the response variables $Y_{m,jt}$ and Z_p are correlated only through the corresponding latent variables, (3) the latent class variables are correlated only through the latent profile variable, and (4) the latent group variable is related with latent class variables only through the latent profile variable. Here, conditions (1) and (3) assume the situation where the latent profile variable is identified only by the latent class variable, implying that the latent class variables C_{jt} for $j = 1, \dots, J$ and $t = 1, \dots, T$ be independent when the latent profile membership $U = u$ is given. Condition (2) states that a set of response variables $\mathbf{Y}_{jt} = [Y_{1,jt}, \dots, Y_{m,jt}]'$ corresponding to the j^{th} latent class variable at stage t are mutually independent if they are conditioned on $C_{jt} = c_{jt}$, the j^{th} latent class membership at t . Such conditional independence structure implies that any underlying association within a single set of response variables \mathbf{Y}_{jt} can be fully explained by the corresponding categorical latent variable C_{jt} . Same rule applies to the latent group variable and latent profile variable (the latent class variables C_{jt} for $j = 1, \dots, J$ and $t = 1, \dots, T$ to be independent when the latent profile membership $U = u$ is given). Condition (4) implies that the latent profile structure varies across latent group membership, but the latent class structure is invariant to the latent group membership.

The prevalence of the latent profile may also be affected by the individual's covariates. Figure 1 shows that we can construct the multinomial logistic regression model by treating the latent profile variable as a response variable. While the conventional multinomial logistic regression utilizes observed values of the response variable and covariates, the regression on unobservable latent profile membership relates the covariates with posterior probabilities, which will be discussed later. Suppose we have a vector of covariates $\mathbf{x}_i = [x_{i1}, \dots, x_{iq}]'$ for the i^{th} observation, then the latent profile can be written as a function of covariates in multinomial logistic regression form as:

$$L^*(\mathbf{x}_i) = \gamma_{u|d}(\mathbf{x}_i) \prod_{t=1}^T \prod_{j=1}^J \left[\eta_{c_{jt}|u}^{(j,t)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \left(\rho_{m_j k | c_{jt}}^{(j,t)} \right)^{I(y_{im_j jt}=k)} \right] \delta_d \prod_{p=1}^P \prod_{h=1}^{r_p} \phi_{ph|d}^{I(z_{ip}=h)}, \quad (2.4)$$

where $\gamma_{u|d}(\mathbf{x}_i) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_{u|d}) / \sum_{u=1}^S \exp(\mathbf{x}_i' \boldsymbol{\beta}_{u|d})$. Here, the vector of logistic regression coefficients $\boldsymbol{\beta}_{u|d} = [\beta_{1u|d}, \dots, \beta_{qu|d}]'$ is interpreted as the log-odds ratio that an individual belongs to a specific latent profile u versus to a baseline latent class, given the latent group membership $D = d$. The observed-data likelihood of the GLCPA can be derived by the marginal summation of (2.4) with

respect to all considered latent variables:

$$L(\mathbf{x}_i) = P(\mathbf{Y}_i = \mathbf{y}_i, \mathbf{Z}_i = \mathbf{z}_i | \mathbf{x}_i) = \sum_{d=1}^G \sum_{u=1}^S \sum_{c_{11}=1}^{K_1} \cdots \sum_{c_{JT}=1}^{K_J} L^*(\mathbf{x}_i). \quad (2.5)$$

3. Parameter estimation and model selection

In this section, we discuss the estimation and inference for the model parameters and the model selection procedure. We adopt the recursive expectation-maximization (EM) algorithm (Bartolucci *et al.*, 2010) to estimate the model parameters and to derive the Hessian matrix of the model to obtain the asymptotic standard errors of the parameter estimates. To determine the number of classes for each latent variable, we investigate the model with various number of classes and chose the most appropriate one using Akaike information criterion (AIC), Bayesian information criterion (BIC), and bootstrap p -value as our criteria.

3.1. Recursive expectation-maximization algorithm

The parameter estimation in finite mixture model can be considered as a missing data problem, because the latent class membership cannot be observed directly. EM algorithm (Dempster *et al.*, 1977) can be a useful tool to estimate parameters in GLCPA, especially in that it provides stable ML estimates in the proper parameter space when the appropriate initial values are given. The estimators can also be easily obtained in closed form when covariates are not included. The conditional distribution of the related response variables follows the multinomial distribution given the latent variables. Newton-Raphson can be used for estimating regression coefficients when the model contains covariates because the ML estimation can be treated as an estimation problem in baseline multinomial regression. The typical EM algorithm implements expectation step (E-step) and maximization step (M-step) for each iteration, and repeats these steps until the solutions satisfy the convergence threshold. We calculate the expectation of the conditional distributions of latent variables given the response variables in E-step, and update the parameter estimates using the estimators which maximize the expectation.

- **E-step.** We calculate the joint posterior probability of latent variables given the i^{th} observed responses and covariates as:

$$\begin{aligned} \theta_{i(u,d,c)} &= P(U = u, D = d, \mathbf{C} = \mathbf{c} | \mathbf{y}, \mathbf{z}, \mathbf{x}_i) \\ &= \frac{L^*(\mathbf{x}_i)}{L(\mathbf{x}_i)}, \end{aligned} \quad (3.1)$$

for $i = 1, \dots, n$, $u = 1, \dots, S$, $d = 1, \dots, G$, $c_{jt} = 1, \dots, K_j$, $j = 1, \dots, J$, and $t = 1, \dots, T$. Then, the expectation of the complete-data log-likelihood given in (2.4) can be expressed as

$$\begin{aligned} E\left(\sum_{i=1}^n \log L^*(\mathbf{x}_i)\right) &= \sum_{i=1}^n \theta_{i(u)} \log \delta_d + \sum_{i=1}^n \theta_{i(u,d)} \log \gamma_{u|d}(\mathbf{x}_i) + \sum_{i=1}^n \sum_{t=1}^T \sum_{j=1}^J \theta_{i(u,c_{jt})} \log \eta_{c_{jt}|u}^{(j,t)} \\ &\quad + \sum_{i=1}^n \sum_{p=1}^P \sum_{h=1}^{r_p} \theta_{i(d)} I(z_{ip} = h) \log \phi_{ph|d} + \sum_{i=1}^n \sum_{t=1}^T \sum_{j=1}^J \sum_{m_j=1}^{M_j} \sum_{k=1}^{r_{m_j}} \theta_{i(c_{jt})} I(y_{im_j t} = k) \log \rho_{m_j k | c_{jt}}^{(j,t)}, \end{aligned}$$

where marginal posterior probabilities can be computed as $\theta_{i(u,d)} = \prod_{j=1}^J \prod_{t=1}^T \sum_{c_{jt}=1}^{K_j} \theta_{i(u,d,c)} \theta_{i(u,c_{jt})} = \sum_{d=1}^G \prod_{j' \neq j} K_{j'} \prod_{t' \neq t} T_{t'} \sum_{c_{j't'}=1}^{K_{j'}} \theta_{i(u,d,c)}$, $\theta_{i(u)} = \sum_{d=1}^G \theta_{i(u,d)}$, $\theta_{i(d)} = \sum_{u=1}^S \theta_{i(u,d)}$, and $\theta_{i(c_{jt})} = \sum_{u=1}^S \sum_{d=1}^G \theta_{i(u,d,c_{jt})}$.

However, the computational complexity for calculating overall posterior was not negligible (especially in that as number of time stage T becomes larger) because its dimension is $(\prod_{j=1}^J K_j \times S \times G)^T$, which increases exponentially for T . Chang and Chung (2013) addressed such computational issue based on recursive formulas (Bartolucci *et al.*, 2010), and we also applied it to the E-step using the forward and backward probabilities to deal with computational complexity. Let α and λ represent the forward and backward probabilities:

$$\begin{aligned}\alpha_{it}(u, \mathbf{c}_t) &= P(\mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_t = \mathbf{y}_t, \mathbf{C}_t = \mathbf{c}_t | u) \\ &= \sum_{c_{1,t-1}=1}^{K_1} \cdots \sum_{c_{J,t-1}=1}^{K_J} \alpha_{i,t-1}(u, \mathbf{c}_{t-1}) \prod_{j=1}^J \left[\eta_{c_{jt}|u}^{(j,t)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \left(\rho_{m_j k | c_{jt}}^{(j,t)} \right)^{I(y_{im_j jt}=k)} \right], \\ \lambda_{it}(u, \mathbf{c}_t) &= P(\mathbf{Y}_{t+1} = \mathbf{y}_{t+1}, \dots, \mathbf{Y}_T = \mathbf{y}_T | \mathbf{c}_t, u) \\ &= \sum_{c_{1,t+1}=1}^{K_1} \cdots \sum_{c_{J,t+1}=1}^{K_J} \lambda_{i,t+1}(u, \mathbf{c}_{t+1}) \prod_{j=1}^J \left[\eta_{c_{j,t+1}|u}^{(j,t+1)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \left(\rho_{m_j k | c_{j,t+1}}^{(j,t+1)} \right)^{I(y_{im_j j(t+1)}=k)} \right].\end{aligned}$$

The forward and backward probabilities allow us to obtain the posterior probabilities of latent class memberships at single time point ($\mathbf{c}_t = [c_{1t}, \dots, c_{Jt}]$, $t = 1, \dots, T$) without considering the combination of whole latent variable sequences. The posterior probability of latent class memberships at stage t can be obtained as

$$\theta_{i(u,d,\mathbf{c}_t)} = \frac{\delta_d \prod_{p=1}^P \prod_{h=1}^{r_p} \phi_{ph|d}^{I(z_{ip}=h)} \gamma_{u|d}(\mathbf{x}_i) \alpha_{it}(u, \mathbf{c}_t) \lambda_{it}(u, \mathbf{c}_t)}{\sum_{g=1}^G \delta_g \prod_{p=1}^P \prod_{h=1}^{r_p} \phi_{ph|g}^{I(z_{ip}=h)} \sum_{s=1}^S \gamma_{s|g}(\mathbf{x}_i) \sum_{c_{1T}=1}^{K_1} \cdots \sum_{c_{JT}=1}^{K_J} \alpha_{iT}(s, \mathbf{c}_T)} \quad (3.2)$$

The dimension of the overall posterior probability is reduced to $(\prod_{j=1}^J K_j \times S \times G \times T)$. The reduction of dimension is noticeable when $T \geq 3$. Chang and Chung (2013) showed that the recursive approach in E-step results in a huge decrease in computational time through simulation study.

- **M-step.** The M-step maximizes the expected complete-data likelihood with respect to the model parameters. Since the sum of parameters that are used in measuring each latent variable is constrained to be one (for instance, $\sum_{d=1}^G \delta_d = 1$, $\sum_{u=1}^S \gamma_{u|d} = 1$, $d = 1, \dots, G$), we adopted Lagrange multiplier to obtain the ML estimator under such constraints. The resulting parameter estimators for the GLCPA without covariates are given as

$$\begin{aligned}\hat{\gamma}_{u|d} &= \frac{\sum_{i=1}^n \theta_{i(u,d)}}{\sum_{i=1}^n \theta_{i(d)}}, \quad \hat{\eta}_{c_{jt}|u}^{(j,t)} = \frac{\sum_{i=1}^n \theta_{i(u,c_{jt})}}{\sum_{i=1}^n \theta_{i(u)}}, \quad \hat{\delta}_d = \frac{\sum_{i=1}^n \theta_{i(d)}}{n}, \\ \hat{\phi}_{ph|d} &= \frac{\sum_{i=1}^n \theta_{i(d)} I(z_{ip} = h)}{\sum_{i=1}^n \theta_{i(d)}}, \quad \text{and} \quad \hat{\rho}_{m_j k | c_{jt}}^{(j,t)} = \frac{\sum_{i=1}^n \theta_{i(c_{jt})} I(y_{im_j jt} = k)}{\sum_{i=1}^n \theta_{i(c_{jt})}}.\end{aligned} \quad (3.3)$$

To deal with missing values in response variables, EM estimator for primary measurement parameters should be extended. Under the missing at random (MAR) assumption, the posterior probability $\theta_{i(c_{jt})}^{\text{obs}}$ and $\theta_{i(d)}^{\text{obs}}$ are calculated using only observed values. For the individuals whose response value is missing, the additional term is included in primary measurement parameter estimators that contain

information about provisional estimates $\hat{\rho}_{m_j k | c_{jt}}^{(j,t)*}$ and $\hat{\phi}_{ph | d}^*$ as

$$\hat{\rho}_{m_j k | c_{jt}}^{(j,t)} = \frac{\sum_{i \in \text{obs}_{m_j}^{(j,t)}} \theta_{i(c_{jt})}^{\text{obs}} I(y_{im_j jt} = k) + \sum_{i \in \text{mis}_{m_j}^{(j,t)}} \theta_{i(c_{jt})}^{\text{obs}} \hat{\rho}_{m_j k | c_{jt}}^{(j,t)*}}{\sum_{i=1}^n \theta_{i(c_{jt})}^{\text{obs}}},$$

and

$$\hat{\phi}_{ph | d} = \frac{\sum_{i \in \text{obs}_p} \theta_{i(d)}^{\text{obs}} I(z_{ip} = h) + \sum_{i \in \text{mis}_p} \theta_{i(d)}^{\text{obs}} \hat{\phi}_{ph | d}^*}{\sum_{i=1}^n \theta_{i(d)}^{\text{obs}}},$$

where $\text{obs}_{m_j}^{(j,t)}$ indicates the individuals who responded to the $(m_j)^{\text{th}}$ item measuring latent variable j at stage t , and $\text{mis}_{m_j}^{(j,t)}$ indicates the ones whose response values are missing.

To include the covariate effects on the distribution of latent profiles, $\gamma_{u|d}$ should be re-written as $\gamma_{u|d}(\mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta}_{u|d}) / \sum_{s=1}^S \exp(\mathbf{x}_i^T \boldsymbol{\beta}_{s|d})$, and thus the estimator for $\gamma_{u|d}$ in (3.3) is no longer available. In this case, we obtain β -estimate which maximizes complete-data log-likelihood by implementing the Newton-Raphson method for baseline multinomial logistic regression, where the fractional counts in the response variable are allowed.

3.2. Model diagnosis and selection

Model selection in a complex latent structure is crucial because the interpretability of the model significantly differed by the number of latent classes for each latent variable. Likelihood-ratio test statistics (LRT) cannot be used for comparing different models since the models with different number of latent classes are not in a nested relationship (Collins and Lanza, 2010). Alternatively, we adopted AIC and BIC to assess a relative model fit among candidate models with a different number of classes; consequently, and the model with smaller AIC (or BIC) is preferred. For the absolute model fit, we used the parametric bootstrap p -value; in addition, the model with a p -value larger than 0.05 will be considered as an appropriate model. The procedure for bootstrap p -value is as: (1) fit GLCPA and obtain ML estimates and $G^2 = -2(\log L_{\text{model}} - \log L_{\text{sat}})$, (2) generate the data set of n samples using ML estimates from step (1), (3) fit the model with the generated data set and obtain G^2 , (4) repeat the steps (2) and (3) 100 times; therefore, this will provide with empirical distribution of G^2 (5) to obtain p -value by calculating the proportion of bootstrap G^2 s that are greater than the observed G^2 from step (1).

Models with all possible combinations can be candidates since the model selection procedure can entail tedious trials and errors because each latent variable may have a different number of classes. Jeon *et al.* (2017) showed that the number of latent classes in a complex latent structure can be determined hierarchically. Therefore, one can determine the number of each latent variable and then work on finding the number of latent profiles with the number of latent classes that are fixed from the previous step. Based on this hierarchical process, we chose the number of latent classes for each latent variable based on AIC and BIC, then we calculated the bootstrap p -value to check if the model was appropriate for data. A p -value smaller than 0.05 was accepted as evidence of lack of fit. Bandeen-Roche *et al.* (1997) proved that the LCA has a marginalization property, showing that the model selection procedure can be done without considering the covariates. As a result, we consider the covariate effects on latent profiles after the model selection procedure is completed.

3.3. Simulations

The simulation study was designed to check if our estimation method provides consistent parameter estimates and asymptotic standard errors. We generated datasets and calculated the ML estimates using a recursive EM algorithm. We constructed 95% confidence interval for each parameters based on parameter estimates and standard errors; in addition, the empirical coverage of the confidence intervals were calculated from 100 iterations. Standard errors of estimates were obtained through an asymptotic variance-covariance matrix, by taking the negative inverse of the Hessian matrix. Data was simulated to have two time stages with two latent variables, a group latent variable with two classes, and a sample size of 500. Each latent variable was measured by four binary item-response variables with a latent profile variable designed to have a two-profile structure. The true ρ -parameter values were set to be either strong (close to 0 or 1) (Table A.1) or mixed (Table A.2). The simulation results can be found in Appendix A.

4. Application to NLSY97 data

4.1. Data description

The NLSY97 is a longitudinal project that tracks the lives of a sample of American youth born between 1980–1984; 8,984 respondents ages from 14 to 17 were first interviewed in 1997. This ongoing project has been surveyed 10 times; respondents are now interviewed biennially. The response variables selected for the research are related with alcohol drinking, cigarette/marijuana smoking, and depression. Two sets of five items were used for measuring the two latent class variables, alcohol drinking and cigarette/marijuana smoking, respectively. An additional five items were used for measuring depression which serves as the latent group variable. Response variables related with depression were collected in 2000 when respondents were 17–20 years old, and the responses for alcohol drinking and cigarette/marijuana smoking were collected in 2000, 2002, and 2004.

To measure latent group variable (depression), we selected the following five survey questions: (a) How often you have been a nervous person in past month? (b) How often you felt calm and peaceful in past month? (c) How often you felt down and blue in past month? (d) How often you have been a happy person in past month? and (e) How often you depressed in last month? Response variables (b) and (d) were re-coded to be consistent in the manner that the higher response values implies more exposure to depression symptoms. In this way, we defined each binary manifest item indicating if the respondent had suffered that feeling at least one time or not, as Nervous, Not calm, Down and blue, Not happy, and Depressed, respectively.

For alcohol drinking, the following three survey items were selected and re-coded: (a) number of days drinking alcohol last 30 days (b) number of days having five or more drinks per day last 30 days (c) number of days drinking at schools or work per day last 30 days. The quantitative question (a) was used to create two binary manifest items on if one had drank alcohol in last 30 days (Current drinking), whether had ever drank for five or more days (Frequent drinking), and if one had drank for 20 or more days (Daily drinking). Questionnaire (b) and (c) were transformed into binary variables (Binge drinking and Drinking at school), having ‘yes’ if its value was higher than 0, ‘no’ otherwise.

The quantitative question for cigarette/marijuana smoking was transformed into two binary items of if one had ever smoked in the last 30 days (Current cigarette smoking), whether one had ever smoked in a daily manner for the last 30 days (Daily cigarette smoking), and if one had ever smoked 20 or more cigarettes per day in last 30 days (Heavy cigarette smoking). Finally, the variable Current marijuana smoking was ‘yes’ if one had ever smoked in last 30 days, and Frequent marijuana smoking was assigned to be ‘yes’ if one used marijuana more than 5 times in last 30 days. Table 1 shows

Table 1: Percentage of ‘yes’ responses and non-responses to the items of the latent group (depression) and class variables (alcohol drinking and cigarette/marijuana smoking) over three waves

Latent variable	Manifest item	Year					
		2000		2002		2004	
		Yes	Missing	Yes	Missing	Yes	Missing
Depression	Nervous	62.86	9.69				
	Not calm	6.34	9.63				
	Down and blue	69.05	9.74				
	Not happy	2.64	9.69				
	Depressed	37.95	9.69				
Alcohol drinking	Current drinking	42.51	9.48	51.04	11.11	52.21	17.11
	Frequent drinking	18.11	9.48	24.91	11.11	26.64	17.11
	Daily drinking	4.41	9.69	4.72	11.11	2.54	17.93
	Binge drinking	24.96	9.53	28.51	11.26	30.44	18.54
	Drinking at school	8.77	9.43	6.85	11.06	5.94	16.81
Cigarettes/ marijuana smoking	Current cigarette smoking	35.92	10.95	35.92	10.95	34.40	17.40
	Daily cigarette smoking	17.19	9.43	20.59	10.95	20.64	17.40
	Heavy cigarette smoking	13.03	9.58	15.12	10.95	15.17	17.40
	Current marijuana smoking	26.13	9.94	24.10	11.36	19.08	17.19
	Frequent marijuana smoking	11.52	9.44	12.43	11.01	9.69	16.59

Table 2: Goodness-of-fit measures for a series of LCA models with the different number of classes for each latent variable

Latent variable	Number of classes	AIC	BIC	Bootstrap <i>p</i> -value
Depression	2	7444.7	7506.1	0.00
	3	7365.1	7460.1	0.08
	4	7364.9	7493.5	0.18
	5	7367.6	7529.6	0.77
Alcohol drinking	2	18246.5	18320.1	0.00
	3	18092.4	18206.0	0.06
	4	18093.2	18247.0	0.42
	5	18105.2	18299.1	0.52
Cigarette/marijuana smoking	2	20802.1	20875.6	0.00
	3	19669.7	19783.3	0.04
	4	18931.7	19085.4	0.52
	5	18942.4	19136.3	0.54

LCA = latent class analysis; AIC = Akaike information criterion; BIC = Bayesian information criterion.

the percentages of respondents who responded ‘yes’ to the 15 binary response variables, and the proportion of non-responses.

By introducing the GLCPA to drug-taking behavior and depression measurement items, we expect to study the following properties of the population: (a) What kinds of latent classes may be found for alcohol drinking, cigarette/marijuana smoking, and depression? (b) What kinds of common sequential patterns of alcohol drinking and cigarette/marijuana smoking can be identified? (c) How does the prevalence of latent profiles of alcohol drinking and cigarette/marijuana smoking change as the latent group membership of depression is varied?

4.2. Model selection

We empirically fitted a series of conventional LCA models on each sets of response variable by increasing the number of latent classes from 2 to 5 and chose an appropriate model based on the AIC, BIC, and the model interpretability. Table 2 shows the goodness-of-fit statistics with the different num-

Table 3: The estimated item-response probabilities for the latent group variable, depression (ϕ -parameters)

Manifest item	Latent group for depression		
	Not depressed	Slightly depressed	Seriously depressed
Nervous	0.413	0.846	0.925
Not calm	0.021	0.000 [†]	0.772
Down and blue	0.423	0.961	0.949
Not happy	0.000 [†]	0.012	0.280
Depressed	0.062	0.603	0.764

[†] The estimated probabilities are constrained to be zero or one.

ber of classes for each latent variable. BIC selected a three-class model for depression and its bootstrap p -value supported the three-class model. Both AIC and BIC selected the three-class model for alcohol drinking and four-class model for cigarette/marijuana smoking, and their bootstrap p -values showed that these models were appropriate. For Alcohol drinking and Depression, ρ -parameters of third and fourth classes in four-class model were not noticeably different; therefore, the interpretation for the latent class became unclear. In addition, the redundant number of latent classes may cause an identifiability problem. As a result, we adopted three-class model although they showed a relatively small p -value.

The second step of model selection in GLCPA determined the number of latent profiles. As discussed in Section 3, we will treat the number of latent classes for each identified latent variable as given in the first step (three classes for depression and alcohol drinking and four classes for cigarette/marijuana smoking), and investigate a series of AIC and BIC values from various candidate models. We computed AIC and BIC values from different models whose number of latent profile are varied from 2 to 6. BIC showed the lowest value in the five-profile model. However, the class interpretations for the fourth and the fifth profile were obscure in a model with five profiles. Therefore, we adopted the four-profile structure as our final model. Given the selected latent structure, we tested if the primary measurement parameters can be equal across the time stages. This homogeneity assumption for ρ -parameter is critical in the longitudinal latent class model because the interpretation for each identified latent class is determined based on the ρ -parameter estimates. The meaning of each latent class should therefore be kept equal across the time for a clear interpretation of the latent profile variable. We used a likelihood-ratio test because the model with equal ρ -parameters over time is nested in the one with no constraints. We set ρ -parameters to be invariant over time because the null hypothesis (ρ -parameters for each latent class variable are equal across the time) was not rejected under $\alpha = 0.05$ ($\chi^2 = 88.42$ with $df = 70$ and p -value = 0.067).

Finally, we included the covariates on the prevalence of the latent profile by incorporating the baseline multinomial logistic regression in GLCPA. We used gender (male/female) and race (white/black/others) as covariates, and obtained the estimated odds ratios to investigate their effect on identified latent profiles.

4.3. Parameter estimates

Table 3 shows the estimates of item-response probabilities for the three identified latent classes of the latent group variable, depression. The first latent class has probabilities that are lower than 0.5 for all binary responses, meaning that individuals in that class do not have any depression symptoms. Thus, this subgroup can be named as ‘not depressed.’ The second group has high probabilities for nervous, down and blue, and depressed items. Thus, the second latent groups can be considered as a ‘slightly depressed’ group. The third group has high probabilities for all items, except not happy item, so this group can be identified as ‘seriously depressed’ group.

Table 4: The estimated item-response probabilities for the latent class variable, alcohol drinking (ρ -parameters)

Manifest item	Latent class for alcohol drinking		
	Non drinker	Current drinker	Heavy drinker
Current drinking	0.064	1.000 [†]	1.000 [†]
Frequent drinking	0.000 [†]	0.307	0.842
Daily drinking	0.000 [†]	0.000 [†]	0.158
Binge drinking	0.000 [†]	0.222	0.954
Drinking at school	0.000 [†]	0.117	0.189

[†] The estimated probabilities are constrained to be zero or one.

Table 5: The estimated item-response probabilities for the latent class variable, cigarette/marijuana smoking (ρ -parameters)

Manifest item	Latent class for cigarette/marijuana smoking			
	Non smoker	Marijuana smoker	Heavy cigarette smoker	Heavy cigarette/marijuana smoker
Current cigarette smoking	0.079	0.380	1.000 [†]	1.000 [†]
Daily cigarette smoking	0.000 [†]	0.000 [†]	0.867	0.947
Heavy cigarette smoking	0.000 [†]	0.000 [†]	0.524	0.581
Current marijuana smoking	0.039	1.000 [†]	0.112	1.000 [†]
Frequent marijuana smoking	0.000 [†]	0.443	0.000 [†]	0.737

[†] The estimated probabilities are constrained to be zero or one.

Table 4 shows the primary measurement parameter estimates for the alcohol drinking variable. The ρ -estimates in the first class were all close or equal to 0, implying that individuals in the first class are not likely exposed to any alcohol drinking. Thus, the first class was named as ‘Non drinker.’ The second class showed the high probabilities for Current drinking item. As a result, this latent class was labeled as ‘Current drinker.’ The third class was labeled as ‘Heavy drinker’ because they had high probabilities for Current, Frequent, and Binge drinking items.

Table 5 shows the five classes of cigarette/marijuana smoking class variable and the estimated ρ -parameter estimates. The first latent class showed low probabilities for all items, meaning that individuals in the first class are ‘Non smoker.’ The second class was named as ‘Marijuana smoker’ because it showed high probabilities for Current mar. smoking item. Similarly, the third class was named as ‘Heavy cigarette smoker’ because it had high probabilities for the Current cigarette smoking, Daily cigarette smoking, and Heavy cigarette smoking items. The fourth class represented the most serious cigarettes and marijuana smokers: it showed the highest probabilities for all manifest items and thus classified as ‘Heavy cigarette/marijuana smoker.’

Table 6 shows the estimated secondary measurement parameters (η -parameters) for the latent profile variable. In Profile 1, η -parameter estimates for alcohol drinking and cigarette/marijuana smoking variables showed the highest probabilities for the ‘Non drinker’ and ‘Non smoker’ respectively, across all three time waves. This implies that Profile 1 presents ‘Not involved in any substance disorder’ subgroup. In Profile 2, the parameter estimates for alcohol drinking variable showed that individuals in this profile moved from ‘Current drinker’ class in 2000 towards ‘Heavy drinker’ in 2002, while they remained in the ‘Non smoker’ class for the cigarette/marijuana variable. As a result, Profile 2 can be named as ‘Heavy drinking advancer.’ In Profile 3, the prevalence of ‘Non drinker’ class for the alcohol drinking variable is the highest in 2000 and moved to ‘Heavy drinker’ across 2002 and 2004. For the cigarette/marijuana smoking variable, the adolescents in this profile had the highest probabilities in ‘Heavy cigarette smoker’ for all three waves. Consequently, Profile 3 was named as ‘Heavy drinking advancer/Heavy cigarette smoker.’ However, Profile 4 identified a subgroup of those who stayed in

Table 6: The estimated conditional probabilities of the latent class membership for a given latent profile membership (η -parameters)

Profile	Year	Latent class for alcohol drinking			Latent class for cigarette/marijuana smoking			
		Non drinker	Current drinker	Heavy drinker	Non smoker	Marijuana smoker	Heavy cigarette smoker	Heavy cigarette/marijuana smoker
1	00	0.891	0.109	0.000 [†]	0.975	0.000 [†]	0.025	0.000 [†]
	02	0.795	0.205	0.000 [†]	0.981	0.009	0.010	0.000 [†]
	04	0.690	0.272	0.038	0.950	0.011	0.039	0.000 [†]
2	00	0.319	0.347	0.334	0.594	0.381	0.013	0.012
	02	0.195	0.352	0.453	0.602	0.375	0.000 [†]	0.023
	04	0.167	0.352	0.481	0.636	0.302	0.023	0.039
3	00	0.483	0.233	0.284	0.304	0.057	0.532	0.107
	02	0.335	0.279	0.386	0.098	0.000 [†]	0.824	0.078
	04	0.339	0.223	0.438	0.146	0.000 [†]	0.818	0.036
4	00	0.161	0.173	0.666	0.059	0.170	0.148	0.623
	02	0.125	0.201	0.674	0.035	0.137	0.158	0.670
	04	0.053	0.182	0.765	0.030	0.084	0.270	0.616

[†] The estimated probabilities are constrained to be zero or one.

Table 7: The estimated odds ratio for the latent profile membership given a latent group variable (depression) and its 95% confidence interval (Profile 1 is the baseline)

Latent group for depression	Profile	Gender	Race	
		Male	Black	Others
Not depressed	2	1.074	1.237	0.881
		[0.615, 1.875]	[0.673, 2.275]	[0.377, 2.059]
	3	1.279	1.197	0.682
		[0.747, 2.189]	[0.657, 2.179]	[0.341, 1.363]
Slightly depressed	4	1.449	0.494	0.229
		[0.609, 3.444]	[0.177, 1.374]	[0.040, 1.312]
	2	1.370	0.559	0.580
		[0.865, 2.169]	[0.328, 0.953]	[0.312, 1.076]
Seriously depressed	3	1.392	0.738	0.757
		[0.902, 2.144]	[0.447, 1.219]	[0.452, 1.266]
	4	2.410	0.178	0.409
		[1.512, 3.838]	[0.082, 0.384]	[0.229, 0.729]
Seriously depressed	2	0.983	1.288	0.834
		[0.310, 3.111]	[0.344, 4.815]	[0.205, 3.392]
	3	0.950	1.045	0.902
		[0.314, 2.880]	[0.276, 3.955]	[0.259, 3.146]
Seriously depressed	4	4.212	0.154	0.217
		[1.003, 17.797]	[0.014, 1.650]	[0.036, 1.301]

‘Heavy drinker’ and ‘Heavy cigarette/marijuana smoker’ for alcohol drinking and cigarette/marijuana smoking variables, respectively for all three waves. Profile 4 was labeled as ‘Heavy substance user’ because it accurately represents adolescents who are seriously exposed to serious drug-taking behavior throughout the all time waves.

We incorporated the multinomial logistic regression into our model in order to examine the effect of individual characteristics on latent profile membership. Table 7 shows the estimated odds ratios and the 95% confidence intervals. Profile 1 was set as the baseline category; therefore, the estimated parameters represent the odds ratios of belonging to a certain latent profile compared to Profile 1. We considered gender (female was set to be the baseline) and race (White was set to be the baseline)

Table 8: Estimated prevalence of latent profile membership for a given depression group

Latent profile	Latent group for depression		
	Not depressed	Slightly depressed	Seriously depressed
1	0.512	0.337	0.325
2	0.219	0.261	0.248
3	0.200	0.248	0.297
4	0.069	0.154	0.130

as covariates, and the estimated coefficients were transformed into odds ratios for interpretation. No covariate effect had a significant effect for ‘Not depressed’ group. For the ‘Slightly depressed’ group, male adolescents have 2.41 times higher odds for Profile 4 compared to the baseline (Profile 1) than females. Similarly, Blacks have 0.559 times lower odds for Profile 2 versus the baseline, 0.178 times lower odds for Profile 4 versus the baseline than White counterparts. For ‘Seriously depressed’ group, male adolescents have 4.21 times higher odds for Profile 4 compared to the baseline than females.

Finally, Table 8 shows the γ -estimates which represent the prevalence of four latent profile subgroups given the depression latent group variable. The γ -estimates can be obtained by

$$\begin{aligned}\hat{P}(U = u | D = d, \mathbf{x}_i) &= \frac{1}{n} \sum_{i=1}^n \hat{\gamma}_{u|g}(\mathbf{x}_i) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\exp(\mathbf{x}_i \hat{\beta}_{u|d})}{\sum_{g=1}^G \exp(\mathbf{x}_i \hat{\beta}_{u|g})}.\end{aligned}$$

Profile 1 was the most prevalent class (0.512) among the four profiles when the depression group was ‘Not depressed,’ but decreases to 0.325 for the ‘Seriously depressed’ group. Profile 2 showed relatively consistent proportions throughout all depression levels, ranging from 0.219 to 0.261. However, Profiles 3 and 4 showed an increasing trend as the level of depression becomes severe, from ‘Not Depressed’ to ‘Seriously Depressed.’ The estimated γ -parameters provide the quantitative measures for the associations between two categorical latent variables in terms of the conditional probability of having a certain latent profile membership given the latent group membership. We could see that as individuals who are exposed to more severe depression levels are more likely to experience serious drug-taking behavior.

5. Conclusion

We suggested a new type of latent variable model to examine the complex structure of categorical latent variables, especially in cases where we need to study the longitudinal trends of latent variables identified through a repeated measured manifest item. The GLCPA uncovers the three types of categorical latent variables: the first are the latent class variables explaining the associations among manifest items, the second is the latent profile variable that examines the longitudinal patterns of one or more latent class variables through repeatedly measured manifest items, and the third is a single latent variable that can be treated as group variable. The GLCPA may specify the conditional probability of individuals belonging to a certain latent profile given the identified latent group membership. Consequently, our proposed model can systemically specify the effect of a latent group membership on the probability of having a certain sequential patterns. We expect that this methodology can be widely applied in educational and psychological studies.

Through the analysis of NLSY97 data, we found four representative sequential patterns of young adolescents drug-taking behavior. These four common patterns identify the subgroups of a population

not exposed to any type of substance use (Profile 1, ‘Not involved in any substance disorder’), who moved to severe alcohol drinker (Profile 2, ‘Heavy drinking advancer’), heavy cigarette smokers who were likely to transfer from non drinker to heavy drinker (Profile 3, ‘Heavy drinking advancer/Heavy cigarette smoker’), and adolescents with serious drug-taking behavior (Profile 4, ‘Heavy substance user’). The proportions of the four latent profiles varied by the individual level of depression symptoms. Our proposed model discovered that the probability of not being exposed to the any type of drug-taking behavior decreases as the level of depression symptoms increase and that the prevalence of the adolescents with severe drug-taking behaviors also increases. This does not imply a causal relationship; however, such trend provides a quantitative indication of positive association between depression symptoms and drug-taking behavior. A rigid causal inference between the latent profile variable and the group latent variable represents a future research topic. For the causal inference approach in the conventional latent class structure (Lanza *et al.*, 2013).

The EM algorithm is widely adopted for the parameter estimation of the finite mixture model due to difficulties with unobservable structures. An EM algorithm provided a stable ML estimation; however, the computational cost was relatively huge compared to the other estimation strategies with the burden of computational complexity also becoming worse if the number of time stage increases. The recursive method discussed in Subsection 3.1 significantly reduced computational complexity by skipping the calculation of redundant posterior terms from (3.1). For the simulation result in the latent class profile analysis, see Chang and Chung (2013) which showed the superiority of recursive EM estimation to conventional EM in time efficiency. EM algorithm also required appropriate initial values to guarantee the converged solution to be a global maximum. To achieve global maximum, we used 100 different sets of starting values and chose the one with the highest likelihood as a final solution, which requires another huge calculation and time cost. A deterministic annealing EM algorithm ensures that a global maximum can be adopted to avoid the difficulty of choosing an appropriate initial value (Chang and Chung, 2013; Lee and Chung, 2017). We have now made a program for GLCPA written in R language (version 3.6.0) that is available on request.

Appendix A: Simulation results

Table A.1 and A.2 shows that the average of parameter estimates, mean square errors, and 95% coverage probabilities for strong and mixed primary measurement parameters (ρ -parameters). Strong measurement parameters are close to 0 or 1, indicating a strong association between the response variable and the latent variable. However, some mixed measurement parameters, are close to 0.5 and represent the relatively weaker association. The average estimates from the EM algorithm were considerably similar with the true values along with the coverage probabilities of the 95% confidence intervals that are fairly close to 0.95 in both simulations. This implies that the parameter estimation and model identification are working properly.

Appendix B: Elements of the score function

Let Θ be a vector of all free parameters for the GLCPA. The score function $S(\Theta)$ is obtained by the first-ordered derivatives of the log-likelihood of the GLCPA given in (2.5) with respect to the model parameters Θ . Let β be the vectorized β -parameters in the GLCPA model. The elements of the first-derivative vector with respect to β (i.e., $\sum_{i=1}^n \partial \log L(\mathbf{x}_i) / \partial \beta$) are given by

$$\sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \beta_{qu|d}} = \sum_{i=1}^n x_{iq} [\theta_{i(u,d)} - \gamma_{u|d}(x_i) \theta_{i(d)}],$$

Table A.1: Average estimates (EST), mean square error (MSE), and coverage probability (CP) of 95% confidence intervals for parameter estimates (Strong)

Parameter	True	EST	MSE	CP	Parameter	True	EST	MSE	CP
$\rho_{11 1}^{(1,1)}$	0.90	0.901	0.0004	0.95	$\rho_{11 2}^{(1,1)}$	0.10	0.102	0.0003	0.97
$\rho_{21 1}^{(1,1)}$	0.90	0.902	0.0004	0.97	$\rho_{21 2}^{(1,1)}$	0.10	0.098	0.0004	0.97
$\rho_{31 1}^{(1,1)}$	0.90	0.897	0.0004	0.94	$\rho_{31 2}^{(1,1)}$	0.10	0.101	0.0004	0.97
$\rho_{41 1}^{(1,1)}$	0.90	0.904	0.0004	0.97	$\rho_{41 2}^{(1,1)}$	0.10	0.099	0.0004	0.96
$\rho_{11 1}^{(2,1)}$	0.10	0.097	0.0003	0.96	$\rho_{11 2}^{(2,1)}$	0.90	0.903	0.0005	0.95
$\rho_{21 1}^{(2,1)}$	0.10	0.101	0.0004	0.97	$\rho_{21 2}^{(2,1)}$	0.90	0.898	0.0004	0.92
$\rho_{31 1}^{(2,1)}$	0.10	0.101	0.0004	0.93	$\rho_{31 2}^{(2,1)}$	0.90	0.901	0.0004	0.94
$\rho_{41 1}^{(2,1)}$	0.10	0.102	0.0004	0.97	$\rho_{41 2}^{(2,1)}$	0.90	0.895	0.0004	0.94
$\rho_{11 1}^{(1,2)}$	0.10	0.099	0.0003	0.96	$\rho_{11 2}^{(1,2)}$	0.90	0.903	0.0004	0.96
$\rho_{21 1}^{(1,2)}$	0.10	0.104	0.0005	0.98	$\rho_{21 2}^{(1,2)}$	0.90	0.897	0.0005	0.97
$\rho_{31 1}^{(1,2)}$	0.10	0.097	0.0004	0.96	$\rho_{31 2}^{(1,2)}$	0.90	0.901	0.0003	0.94
$\rho_{41 1}^{(1,2)}$	0.10	0.097	0.0004	0.96	$\rho_{41 2}^{(1,2)}$	0.90	0.902	0.0008	0.96
$\rho_{11 1}^{(2,2)}$	0.90	0.896	0.0003	0.94	$\rho_{11 2}^{(2,2)}$	0.10	0.099	0.0004	0.97
$\rho_{21 1}^{(2,2)}$	0.90	0.900	0.0003	0.95	$\rho_{21 2}^{(2,2)}$	0.10	0.100	0.0004	0.95
$\rho_{31 1}^{(2,2)}$	0.90	0.896	0.0003	0.97	$\rho_{31 2}^{(2,2)}$	0.10	0.099	0.0003	0.95
$\rho_{41 1}^{(2,2)}$	0.90	0.899	0.0004	0.94	$\rho_{41 2}^{(2,2)}$	0.10	0.098	0.0003	0.92
$\eta_{1 1}^{(1,1)}$	0.80	0.799	0.0003	0.98	$\eta_{1 2}^{(1,1)}$	0.20	0.198	0.0004	0.98
$\eta_{1 1}^{(2,1)}$	0.20	0.201	0.0003	0.94	$\eta_{1 2}^{(2,1)}$	0.80	0.801	0.0004	0.96
$\eta_{1 1}^{(1,2)}$	0.20	0.202	0.0004	0.95	$\eta_{1 2}^{(1,2)}$	0.80	0.803	0.0003	0.99
$\eta_{1 1}^{(2,2)}$	0.80	0.802	0.0004	0.94	$\eta_{1 2}^{(2,2)}$	0.20	0.202	0.0004	0.97
$\beta_{11 1}$	-1.00	-1.023	0.0713	0.97	$\beta_{21 1}$	1.00	1.042	0.0456	0.95
$\beta_{11 2}$	1.00	1.031	0.0899	0.98	$\beta_{21 2}$	-1.00	-1.007	0.0392	0.97
$\phi_{11 1}$	0.90	0.904	0.0007	0.99	$\phi_{11 2}$	0.10	0.101	0.0008	0.97
$\phi_{21 1}$	0.90	0.900	0.0005	0.97	$\phi_{21 2}$	0.10	0.103	0.0007	0.96
$\phi_{31 1}$	0.90	0.906	0.0009	0.97	$\phi_{31 2}$	0.10	0.099	0.0007	0.95
$\phi_{41 1}$	0.90	0.901	0.0006	0.96	$\phi_{41 2}$	0.10	0.102	0.0008	0.98
δ_1	0.50	0.499	0.0005	0.95					

for $q = 1, \dots, p$, $u = 1, \dots, S - 1$, $d = 1, \dots, D$. Also, let $\boldsymbol{\rho}_{m_j|c_{jt}}^{(j,t)} = [\rho_{m_j1|c_{jt}}^{(j,t)}, \dots, \rho_{m_jr_{m_j}|c_{jt}}^{(j,t)}]^T$, and $\boldsymbol{\eta}_{t|s}^{(j,t)} = [\eta_{1t|s}^{(j,t)}, \dots, \eta_{K_{jt}|s}^{(j,t)}]^T$ be the vectorized ρ - and η -parameters, respectively for $m_j = 1, \dots, M_j$, $c_{jt} = 1, \dots, K_j$, $j = 1, \dots, J$, $t = 1, \dots, T$, and $s = 1, \dots, S$. The elements of the first-derivative vector with respect to $\boldsymbol{\rho}_{m_j|c_{jt}}^{(j,t)}$ and $\boldsymbol{\eta}_{t|s}^{(j,t)}$ are obtained by

$$\sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \rho_{m_j k|c_{jt}}^{(j,t)}} = \sum_{i=1}^n \frac{\theta_{i(c_{jt})} \zeta_{y_{im_j jk}}}{\rho_{m_j k|c_{jt}}^{(j,t)}},$$

$$\sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \eta_{c_{jt}|u}^{(j,t)}} = \sum_{i=1}^n \frac{\theta_{i(u, c_{jt})}}{\eta_{c_{jt}|u}^{(j,t)}}.$$

Table A.2: Average estimates (EST), mean square error (MSE), and coverage probability (CP) of 95% confidence intervals for parameter estimates (Mixed)

Parameter	True	EST	MSE	CP	Parameter	True	EST	MSE	CP
$\rho_{11 1}^{(1,1)}$	0.90	0.898	0.0005	0.94	$\rho_{11 2}^{(1,1)}$	0.10	0.103	0.0003	0.98
$\rho_{21 1}^{(1,1)}$	0.90	0.900	0.0003	0.98	$\rho_{21 2}^{(1,1)}$	0.10	0.099	0.0004	0.97
$\rho_{31 1}^{(1,1)}$	0.70	0.703	0.0009	0.93	$\rho_{31 2}^{(1,1)}$	0.10	0.101	0.0004	0.96
$\rho_{41 1}^{(1,1)}$	0.70	0.700	0.0007	0.98	$\rho_{41 2}^{(1,1)}$	0.10	0.099	0.0004	0.96
$\rho_{11 1}^{(2,1)}$	0.10	0.096	0.0006	0.96	$\rho_{11 2}^{(2,1)}$	0.70	0.706	0.0005	0.94
$\rho_{21 1}^{(2,1)}$	0.10	0.101	0.0004	0.98	$\rho_{21 2}^{(2,1)}$	0.70	0.699	0.0004	0.92
$\rho_{31 1}^{(2,1)}$	0.30	0.301	0.0013	0.92	$\rho_{31 2}^{(2,1)}$	0.90	0.901	0.0014	0.94
$\rho_{41 1}^{(2,1)}$	0.30	0.300	0.0003	0.96	$\rho_{41 2}^{(2,1)}$	0.90	0.895	0.0009	0.94
$\rho_{11 1}^{(1,2)}$	0.10	0.099	0.0011	0.97	$\rho_{11 2}^{(1,2)}$	0.90	0.903	0.0006	0.95
$\rho_{21 1}^{(1,2)}$	0.10	0.104	0.0015	0.96	$\rho_{21 2}^{(1,2)}$	0.90	0.897	0.0006	0.94
$\rho_{31 1}^{(1,2)}$	0.30	0.299	0.0006	0.96	$\rho_{31 2}^{(1,2)}$	0.70	0.701	0.0010	0.94
$\rho_{41 1}^{(1,2)}$	0.30	0.300	0.0005	0.97	$\rho_{41 2}^{(1,2)}$	0.70	0.702	0.0008	0.97
$\rho_{11 1}^{(2,2)}$	0.90	0.896	0.0006	0.96	$\rho_{11 2}^{(2,2)}$	0.10	0.099	0.0004	0.94
$\rho_{21 1}^{(2,2)}$	0.90	0.900	0.0007	0.95	$\rho_{21 2}^{(2,2)}$	0.10	0.100	0.0005	0.95
$\rho_{31 1}^{(2,2)}$	0.70	0.697	0.0009	0.97	$\rho_{31 2}^{(2,2)}$	0.10	0.095	0.0008	0.95
$\rho_{41 1}^{(2,2)}$	0.70	0.697	0.0008	0.93	$\rho_{41 2}^{(2,2)}$	0.10	0.101	0.0010	0.93
$\eta_{1 1}^{(1,1)}$	0.80	0.799	0.0013	0.97	$\eta_{1 2}^{(1,1)}$	0.20	0.194	0.0012	0.98
$\eta_{1 1}^{(2,1)}$	0.20	0.196	0.0012	0.97	$\eta_{1 2}^{(2,1)}$	0.80	0.810	0.0018	0.96
$\eta_{1 1}^{(1,2)}$	0.20	0.190	0.0018	0.96	$\eta_{1 2}^{(1,2)}$	0.80	0.802	0.0014	0.97
$\eta_{1 1}^{(2,2)}$	0.80	0.796	0.0019	0.98	$\eta_{1 2}^{(2,2)}$	0.20	0.202	0.0012	0.97
$\beta_{11 1}$	-1.00	-1.105	0.1525	0.97	$\beta_{21 1}$	1.00	1.011	0.1301	0.94
$\beta_{11 2}$	1.00	1.082	0.0694	0.98	$\beta_{21 2}$	-1.00	-1.025	0.0622	0.96
$\phi_{11 1}$	0.90	0.901	0.0008	0.98	$\phi_{11 2}$	0.10	0.101	0.0015	0.96
$\phi_{21 1}$	0.90	0.900	0.0013	0.97	$\phi_{21 2}$	0.10	0.101	0.0012	0.95
$\phi_{31 1}$	0.70	0.700	0.0009	0.96	$\phi_{31 2}$	0.30	0.298	0.0011	0.96
$\phi_{41 1}$	0.70	0.698	0.0015	0.94	$\phi_{41 2}$	0.30	0.297	0.0008	0.98
δ_1	0.50	0.503	0.0007	0.94					

Here, $\zeta_{y_{im_jjt}k}$ is the indicator function which has the value of 1 if $y_{im_jjt} = k$, otherwise 0. Note that there are $r_{m_j} - 1$ and $K_j - 1$ free parameters in $\rho_{m_j|c_{jt}}^{(j,t)}$ and $\eta_{t|s}^{(j,t)}$, respectively. Therefore, the score function of the free parameters for $\rho_{m_j|c_{jt}}^{(j,t)}$ and $\eta_{t|s}^{(j,t)}$ can be obtained as:

$$\sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \rho_{m_j|c_{jt}}^{(j,t)}} \mathbf{A}_{r_{m_j}t}^T \quad \text{and} \quad \sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \eta_{t|s}^{(j,t)}} \mathbf{A}_{K_j}^T,$$

where \mathbf{A}_k is a $(k-1) \times k$ matrix, composed of an identity matrix in the first $k-1$ columns and a column vector of -1 in the last column for $m_j = 1, \dots, M_j$, $c_{jt} = 1, \dots, K_j$, $j = 1, \dots, J$, $t = 1, \dots, T$, and $s = 1, \dots, S$. Likewise, let $\phi_{p|d} = [\phi_{p1|d}, \dots, \phi_{pP|d}]^T$, and $\delta = [\delta_1, \dots, \delta_G]^T$ be the vectorized ϕ - and δ -parameters, respectively for $p = 1, \dots, P$, and $d = 1, \dots, G$. The elements of the first-derivative

vector with respect to $\phi_{p|d}$ and δ are obtained as:

$$\sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \phi_{ph|d}} = \sum_{i=1}^n \frac{\theta_{i(d)} \zeta_{z_{ip}h}}{\phi_{ph|d}},$$

$$\sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \delta_d} = \sum_{i=1}^n \frac{\theta_{i(d)}}{\delta_d}.$$

Here, $\zeta_{z_{ip}h}$ is the indicator function which has the value of 1 if $z_{ip} = h$, otherwise 0. Note that there are $r_p - 1$ and $G - 1$ free parameters in $\phi_{p|d}$ and δ , respectively. Therefore, the score function of the free parameters for $\phi_{p|d}$ and δ can be obtained by

$$\sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \phi_{p|d}} \mathbf{A}_{r_p}^T \quad \text{and} \quad \sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \delta} \mathbf{A}_G^T,$$

where \mathbf{A}_k is a $(k-1) \times k$ matrix, composed of an identity matrix in the first $k-1$ columns and a column vector of -1 in the last column for $p = 1, \dots, P$, $d = 1, \dots, G$.

Appendix C: Elements of the Hessian matrix

The Hessian matrix is the second derivatives of the log-likelihood with respect to all free parameters Θ . The second derivatives of log-observed data likelihood with respect to β and $\rho_{m_j j t | c_{jt}}^{(j,t)}$, $\eta_{u'}^{(j,t)}$, γ_d , and $\phi_{p|d}$ are obtained as

$$\begin{aligned} \sum_{i=1}^n \frac{\partial^2 \log L(x_i)}{\partial \beta_{qu|d} \partial \beta_{q'u'|d'}} &= \sum_{i=1}^n x_{iq} x_{iq'} \{ \zeta_{dd'} [\omega_{i(u,d)} (\zeta_{uu'} - \gamma_{u|d}(x_i)) - \omega_{i(u,d')} \gamma_{u|d}(x_i)] - \omega_{i(u',d')} \omega_{i(u,d)} \}, \\ \sum_{i=1}^n \frac{\partial^2 \log L(x_i)}{\partial \delta_d \partial \beta_{qu|d'}} &= \sum_{i=1}^n x_{iq} \frac{(\zeta_{dd'} - \theta_{i(d)}) \theta_{i(u,d')} - \gamma_{u|d'}(x_i) \theta_{i(d)} \theta_{i(d')}}{\partial \delta_d}, \\ \sum_{i=1}^n \frac{\partial^2 \log L(x_i)}{\partial \eta_{c_{jt}|u} \partial \beta_{qu|d'}} &= \sum_{i=1}^n \frac{x_{iq} \{ \zeta_{uu'} \theta_{i(c_{jt},u',d)} - \theta_{i(u',d)} \theta_{i(u,c_{jt})} - \gamma_{u'|d}(x_i) (\theta_{i(c_{jt},u,d)} - \theta_{i(d)} \theta_{i(u,c_{jt})}) \}}{\partial \eta_{c_{jt}|u}^{(j,t)}}, \\ \sum_{i=1}^n \frac{\partial^2 \log L(x_i)}{\partial \phi_{ph|d} \partial \beta_{qu|d'}} &= \sum_{i=1}^n \frac{x_{iq} \{ \theta_{i(u,d')} (\zeta_{dd'} - \theta_{i(d)}) - \gamma_{u|d'}(x_i) \theta_{i(d)} \theta_{i(d')} \} \zeta_{z_{ip}h}}{\phi_{ph|d}}, \\ \sum_{i=1}^n \frac{\partial^2 \log L(x_i)}{\partial \rho_{m_j j t | c_{jt}}^{(j,t)} \partial \beta_{qu|d'}} &= \sum_{i=1}^n \frac{x_{iq} \{ \theta_{i(c_{jt},u,d)} - \theta_{i(u,d)} \theta_{i(c_{jt})} - \gamma_{u|d}(x_i) [\theta_{i(c_{jt},d)} - \theta_{i(c_{jt})} \theta_{i(d)}] \} \zeta_{y_{im_j j t k}}}{\rho_{m_j k j t | c_{jt}}}, \end{aligned}$$

where $\omega_{i(u,d)} = \theta_{i(u,d)} - \gamma_{u|d}(x_i) \theta_{i(d)}$ for $d = 1, \dots, G$, $p = 1, \dots, P$, $h = 1, \dots, r_p$, $m_j = 1, \dots, M_j$, $k, k' = 1, \dots, r_{m_j}$, $c_{jt} = 1, \dots, K_j$, $u, u' = 1, \dots, S$, and $j, j' = 1, \dots, J$. $\beta_{S|d} = [\beta_{1S|d}, \dots, \beta_{LS|d}] = 0$, $u, u' = 1, \dots, S-1$, and $\zeta_{dd'} = 1$ if $d = d'$, 0 otherwise.

The second derivatives of log-observed data likelihood with respect to γ are:

$$\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \gamma_{u|d} \partial \gamma_{u'|d'}} = - \sum_{i=1}^n \frac{\theta_{i(u,d)} \theta_{i(u',d')}}{\gamma_{u|d} \gamma_{u'|d'}},$$

$$\begin{aligned}
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \delta_d \partial \gamma_{u|d'}} &= \sum_{i=1}^n \frac{(\zeta_{dd'} - \theta_{i(d)}) \theta_{i(u,d')}}{\delta_d \gamma_{u|d'}}, \\
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \eta_{c_{jt}|u}^{(j,t)} \partial \gamma_{u'|d}} &= \sum_{i=1}^n \frac{\zeta_{uu'} \theta_{i(c_{jt},u,d)} - \theta_{i(c_{jt},u)} \theta_{i(u',d)}}{\eta_{c_{jt}|u}^{(j,t)} \gamma_{u'|d}}, \\
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \phi_{ph|d} \partial \gamma_{u|d'}} &= \sum_{i=1}^n \frac{(\zeta_{dd'} - \theta_{i(d)}) \theta_{i(u,d')} \zeta_{z_{ip}h}}{\phi_{ph|d} \gamma_{u|d'}}, \\
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \rho_{m_j k | c_{jt}}^{(j,t)} \partial \gamma_{u|d}} &= \sum_{i=1}^n \frac{(\theta_{i(c_{jt},u,d)} - \theta_{i(c_{jt})} \theta_{i(u,d)}) \zeta_{y_{im_j} k}}{\rho_{m_j k | c_{jt}}^{(j,t)} \gamma_{u|d}},
\end{aligned}$$

for $d = 1, \dots, G$, $p = 1, \dots, P$, $h = 1, \dots, r_p$, $m_j = 1, \dots, M_j$, $k, k' = 1, \dots, r_{m_j}$, $c_{jt} = 1, \dots, K_j$, $u, u' = 1, \dots, S$, and $j, j' = 1, \dots, J$. Here, $\zeta_{y_{im_j} k}$ is an indicator function which has the value of 1 if $y_{im_j} = k$, 0 otherwise.

The second derivatives of log-observed data likelihood with respect to δ are:

$$\begin{aligned}
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \delta_d \partial \delta_{d'}} &= - \sum_{i=1}^n \frac{\theta_{i(d)} \theta_{i(d')}}{\delta_d \delta_{d'}}, \\
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \eta_{c_{jt}|u}^{(j,t)} \partial \delta_d} &= \sum_{i=1}^n \frac{\theta_{i(c_{jt},u,d)} - \theta_{i(u,c_{jt})} \theta_{i(d)}}{\eta_{c_{jt}|u}^{(j,t)} \delta_d}, \\
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \phi_{ph|d} \partial \delta_{d'}} &= \sum_{i=1}^n \frac{(\zeta_{dd'} - \theta_{i(d)}) \theta_{i(d')} \zeta_{z_{ip}h}}{\phi_{ph|d} \delta_{d'}}, \\
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \rho_{m_j k | c_{jt}}^{(j,t)} \partial \delta_d} &= \sum_{i=1}^n \frac{(\theta_{i(d,c_{jt})} - \theta_{i(c_{jt})} \theta_{i(d)}) \zeta_{y_{im_j} k}}{\rho_{m_j k | c_{jt}}^{(j,t)} \delta_d},
\end{aligned}$$

for $d = 1, \dots, G$, $p = 1, \dots, P$, $h = 1, \dots, r_p$, $m_j = 1, \dots, M_j$, $k, k' = 1, \dots, r_{m_j}$, $c_{jt} = 1, \dots, K_j$, $u, u' = 1, \dots, S$, and $j, j' = 1, \dots, J$. Here, $\zeta_{y_{im_j} k}$ is an indicator function which has the value of 1 if $y_{im_j} = k$, 0 otherwise.

The second derivatives of log-observed data likelihood with respect to η are:

$$\begin{aligned}
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \eta_{c_{jt}|u}^{(j,t)} \partial \eta_{c_{j't'}|u'}^{(j',t')}} &= \sum_{i=1}^n \frac{\theta_{i(u,c_{jt},c_{j't'})} \zeta_{uu'} (1 - \zeta_{j'j} \zeta_{t't}) - \theta_{i(u,c_{jt})} \theta_{i(u',c_{j't'})}}{\eta_{c_{jt}|u}^{(j,t)} \eta_{c_{j't'}|u'}^{(j',t')}}, \\
\sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \phi_{ph|d} \partial \eta_{c_{jt}|u}^{(j,t)}} &= \sum_{i=1}^n \frac{(\theta_{i(c_{jt},u,d)} - \theta_{i(u,c_{jt})} \theta_{i(d)}) \zeta_{z_{ip}h}}{\phi_{ph|d} \eta_{c_{jt}|u}^{(j,t)}}, \\
\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \rho_{m_j k | c_{jt}}^{(j,t)} \partial \eta_{c_{j't'}|u}^{(j',t')}} &= \sum_{i=1}^n \frac{(1 - \zeta_{j'j} \zeta_{t't}) \theta_{i(u,c_{j't'},c_{jt})} + \theta_{i(u,c_{j't'})} (\zeta_{c_{j't'}c_{jt}} - \theta_{i(c_{jt})}) \zeta_{y_{m_j} k}}{\eta_{c_{j't'}|u}^{(j',t')} \rho_{m_j k | c_{jt}}^{(j,t)}},
\end{aligned}$$

for $c_{jt} = 1, \dots, K_j$, $j = 1, \dots, J$, $t = 1, \dots, T$, $d = 1, \dots, G$, and $u, u' = 1, \dots, S$. Here, $\zeta_{j'j}$ is the indicator function whose value is 1 if $j = j'$ and 0 otherwise.

The second derivatives of log-observed data likelihood with respect to ϕ are:

$$\begin{aligned} \sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \phi_{ph|d'} \partial \phi_{p'h'|d'}} &= \sum_{i=1}^n \frac{\theta_{i(d)} \left\{ (1 - \zeta_{dd'}) + (1 - \zeta_{pp'}) - \theta_{i(d')} \right\} \zeta_{z_{ip}h} \zeta_{z_{ip'}h'}}{\phi_{ph|d} \rho_{p'h'|d'}}, \\ \sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \rho_{m_jk|c_{jt}}^{(j,t)} \partial \phi_{ph|d}} &= \sum_{i=1}^n \frac{(\theta_{i(c_{jt},d)} - \theta_{i(c_{jt})} \theta_{i(d)}) \zeta_{z_{ip}h} \zeta_{y_{im_j}jk}}{\phi_{ph|d} \rho_{m_jk|c_{jt}}}, \end{aligned}$$

for $p = 1, \dots, P$, $h = 1, \dots, r_p$, $d = 1, \dots, G$. Here, $\zeta_{z_{ip}h}$ is an indicator function which has the value of 1 if $z_{ip} = h$, otherwise 0.

The second derivatives of log-observed data likelihood with respect to ρ are:

$$\begin{aligned} \sum_{i=1}^n \frac{\partial^2 \log L(\mathbf{x}_i)}{\partial \rho_{m_jk|c_{jt}}^{(j,t)} \partial \rho_{m'_jk'|c'_{jt'}}^{(j',t')}} &= \sum_{i=1}^n \frac{\zeta_{y_{im_j}jk} \zeta_{y_{im'_j}j'k'}}{\rho_{m_jk|c_{jt}}^{(j,t)} \rho_{m'_jk'|c'_{jt'}}^{(j',t')}} \left[\theta_{i(c_{jt},c'_{jt'})} \left((1 - \zeta_{j'j} \zeta_{t't}) + \theta_{i(c_{jt})} \left\{ \zeta_{j'j} \zeta_{t't} \left[(1 - \zeta_{c_{jt}c'_{jt'}}) + (1 - \zeta_{m_jm'_j}) \right] - \theta_{i(c'_{jt'})} \right\} \right) \right], \end{aligned}$$

for $k = 1, \dots, r_{m_j}$, $m_j = 1, \dots, M_j$, $k, k' = 1, \dots, r_{m_j}$, $c_{jt} = 1, \dots, K_j$, $t, t' = 1, \dots, T$, and $j, j' = 1, \dots, J$. Here, $\zeta_{y_{im_j}jk}$ is an indicator function which has the value of 1 if $y_{im_j} = k$, 0 otherwise.

Since $\sum_{u=1}^S \gamma_{u|d}$, $\sum_{d=1}^G \delta_d$, $\sum_{c_{jt}=1}^{K_j} \eta_{c_{jt}|u}^{(j,t)}$, $\sum_{k=1}^{r_{m_j}} \rho_{m_jk|c_{jt}}^{(j,t)}$ are all 1, all existing parameters are not free, and thus we need to multiply the lagrange constraint matrix based on the chain rule. Note that there are $r_p - 1$, $r_{m_j} - 1$, $S - 1$, and $G - 1$ free parameters in $\phi_{p|d}$, $\rho_{m_jk|c_{jt}}$, γ_d , and δ , respectively. Therefore, the second derivatives of the free parameters can be obtained by

$$\left\{ \sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \pi^* \partial \theta^*} \right\} = \mathbf{A}_G \left\{ \sum_{i=1}^n \frac{\partial \log L(\mathbf{x}_i)}{\partial \pi \partial \theta} \right\} \mathbf{A}_K^T,$$

where \mathbf{A}_k is a $(k - 1) \times k$ matrix, composed of an identity matrix in the first $k - 1$ columns and a column vector of -1 in the last column. π and θ represent the vector of parameters where $\dim(\theta) = K$, $\dim(\pi) = G$ respectively, and π^* and θ^* vector of free parameters correspond to π , θ where $\dim(\theta^*) = K - 1$, $\dim(\pi^*) = G - 1$.

Acknowledgements

This work was supported by Korea University Research Program (K1706321 to Chung) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07045821 to Chung).

References

- Bandeem-Roche K, Miglioretti DL, Zeger SL, and Rathouz PJ (1997). Latent variable regression for multiple discrete outcomes, *Journal of the American Statistical Association*, **92**, 1375–1386.
- Bartolucci F, Farcomeni A, and Pennoni F (2010). An overview of latent Markov models for longitudinal categorical data, arXiv preprint arXiv:1003.2804

- Chang HC and Chung H (2013). Dealing with multiple local modalities in latent class profile analysis, *Computational Statistics & Data Analysis*, **68**, 296–310.
- Chung H, Anthony JC, and Schafer JL (2011). Latent class profile analysis: an application to stage sequential processes in early onset drinking behaviours, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **174**, 689–712.
- Collins LM and Lanza ST (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*, John Wiley & Sons, New York.
- Dempster AP, Laird NM, and Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 1–22.
- Jeon S, Lee J, Anthony JC, and Chung H (2017). Latent class analysis for multiple discrete latent variables: A study on the association between violent behavior and drug-using behaviors, *Structural Equation Modeling: A Multidisciplinary Journal*, **24**, 911–925.
- King KA, Vidourek RA, and Merianos AL (2016). Authoritarian parenting and youth depression: Results from a national study, *Journal of Prevention & Intervention in the Community*, **44**, 130–139.
- Lanza ST, Coffman DL, and Xu Shu (2013). Causal inference in latent class analysis, *Structural Equation Modeling: A Multidisciplinary Journal*, **20**, 361–383.
- Lee J, Chung H, and Jeon S (2019). Multivariate latent class profile analysis: Exploring the developmental progression of youth depression and substance use, *Journal of the Korean Statistical Society*, Manuscript submitted for publication.
- Lee JW and Chung H (2017). Latent class analysis with multiple latent group variables, *Communications for Statistical Applications and Methods*, **24**, 173–191.
- Maccoby EE and Martin JA (1983). Socialization in the context of the family: Parent-child interaction. In Paul H. Mussen (Eds), *Handbook of Child Psychology: Formerly Carmichael's Manual of Child Psychology*, Wiley, New York.

Received May 26, 2019; Revised September 16, 2019; Accepted September 17, 2019