RC22981 (W0311-060) November 7, 2003
Mathematics

IBM Research Report

Generalized Linear Latent Variable Modeling for Multi-Group Studies

Jens C. Eickhoff
Department of Biostatistics and Medical Informatics
University of Wisconsin at Madison
Madison, WI 53726-2397

Yasuo Amemiya
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Generalized Linear Latent Variable Modeling for Multi-Group Studies

Jens C. Eickhoff
Department of Biostatistics and Medical Informatics
University of Wisconsin – Madison
Madison, WI 53726-2397
eickhoff@biostat.wisc.edu

Yasuo Amemiya
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598
yasuo@us.ibm.com

Abstract

Latent variable modeling is commonly used in behavioral, social, and medical science research. The models used in such analysis relate all observed variables to latent common factors. In many applications, the observations are highly non-normal or discrete, e.g., polytomous responses or counts. The existing approaches for non-normal observations are applicable only for polytomous outcomes, and use models unsuitable for multi-group analysis. We propose a new generalized linear model approach for latent variable analysis that can handle a broad class of non-normal and discrete observations, and that furnishes meaningful interpretation and inference in multi-group studies through maximum likelihood analysis. A Monte Carlo EM algorithm is proposed for parameter estimation. The convergence assessment and standard error estimation are addressed. An application of this new approach in a substance abuse prevention study is presented.

Keywords and Phrases: Structural equation analysis; Exponential family distributions; Monte Carlo EM algorithm.

1 INTRODUCTION

Latent variable modeling is a widely used statistical method in multivariate situations where some concepts or variables are not directly observable. Special cases include factor analysis and the more general structural equations analysis which is also known as covariance structure analysis. Statistical methods for analyzing covariances and other relationships between latent and observed variables were historically originated by psychometricians. Today, latent variable models are extensively used in the behavioral, social, and medical sciences (see, e.g., Bentler 1995; Bollen 1989; Jöreskog and Sörbom 1996).

Most latent variable models are based on the assumption that the observed variables are continuous with a multivariate normal distribution. However in many applications, due to the nature of the problem or the design of the questionnaires, some or all observed outcome variables are in a non-normal form such as ordered categorical variables. Specifically, in behavioral and social science studies, data are frequently collected based on Likert scales (e.g., "disagree", "neutral", "agree"). Because of its importance in many application, there has been a lot of attention in latent variable modeling with ordinal outcomes and it remains an area of active research. Two main approaches for analyzing latent variable models with ordinal response have evolved.

The first approach assumes that each ordinal outcome variable is generated by an underlying unobserved continuous variables which is usually assumed to be normally distributed. In this approach, direct maximum likelihood estimation becomes computationally infeasible in models involving higher dimensional latent variables since it requires maximization over multiple integrals. To reduce the computational burden of multiple integration, several versions of multi-stage estimation procedures have been proposed. Christoffersson (1975) and Muthén (1978) have used a multi-stage estimation procedure to fit factor analysis model with dichotomous outcome variables. In the first stage of this procedure, the multivariate model is partitioned into bivariate sub-models. Then the thresholds and polychoric correlations in the bivariate sub-models are estimated. This requires only the evaluation of bivariate integrals. In the final step, the parameters are estimated by minimizing a weighted least squares (WLS) function where the weight matrix is the estimated asymptotic covariance matrix of

the polychoric correlations. Muthén (1984) proposed a multi-stage procedure for structural equation models with continuous and polytomous responses. Similar multi-stage procedures have been developed by Lee and Poon (1987), Jöreskog (1990), and Lee, Poon, and Bentler (1992). All these procedures use a weighted least squares estimator based on first and second-order sample information using polychoric and polyserial correlations. Today, the underlying variable approach with multi-stage WLS estimation procedures is widely used to analyze structural equation models with ordinal responses since it has been implemented in popular software packages including LISREL/PRELIS (Jöreskog and Sörbom 1996), EQS (Bentler 1995), LISCOMP (Muthén 1987), and Mplus (Muthén and Muthén 1998). However, the parameterization used in this approach is somewhat inflexible and there are limitations for multi-sample analysis. Specifically, this parameterization does not allow direct assessment of group characteristics. Another drawback is that multi-stage WLS estimation procedures can perform poorly in situations of low prevalence, small sample size, and large number of indicator variables. In particular, it has been demonstrated that multi-stage WLS procedures can experience problems such as instability, bias, non-convergence, and non-positive definiteness of weight matrices (Reboussin and Liang 1998). Finally, the underlying variable approach is limited to finite categorical outcomes while infinite outcomes, such as Poisson counts, cannot be accomplished.

The second approach for ordinal response data is the item response theory designed for dichotomous outcomes with a single factor. In this approach, the conditional response probability for each dichotomous outcome is expressed as a function of the single factor (see, e.g., Bock and Aitkin 1981; Jöreskog and Moustaki 2001). The item response approach has been used widely in measurement theory and has been implemented in a number of computer packages, such as TESTFACT (Wilson et al. 1991).

In this paper, we propose a generalization of the second approach. Specifically, we incorporate the generalized linear model concept into the latent variable framework. We introduce a Generalized Linear Latent Variable Model (GLLVM) approach that can furnish a wide class of latent variable models and can incorporate various types of outcome variables. The parameterization used in our GLLVM approach is flexible and particularly appropriate for coherent multi-sample analyses. We propose a computationally feasible Monte Carlo EM al-

gorithm to compute full maximum likelihood estimates. This article is organized as follows. The general model and motivation for our approach is given in Section 2. The maximum likelihood estimation via Monte Carlo EM algorithm is described in Section 3. This section includes subsections on how to compute standard errors and how to determine the convergence of the Monte Carlo EM algorithm. In Section 4 an example from a substance abuse intervention study is discussed, where the approach proposed in this article was successfully applied to perform a multi-group analysis. Finally, a brief summary is given in Section 5.

2 THE MODEL

Consider a set of G groups which may represent different treatment groups, sex groups, etc. Let $\boldsymbol{y}_i^{(g)} = \left(y_{1i}^{(g)}, \cdots, y_{pi}^{(g)}\right)'$ denote a set of p observed variables for the g^{th} group, measured on the i^{th} individual, $i = 1, \cdots, N^{(g)}$, where $N^{(g)}$ denotes the number of observations within group g. We assume independence between the groups and i.i.d. samples within each group. To motivate our model we assume that the expectations of the outcome variables are functionally related to the q dimensional latent variable $\boldsymbol{f}_i^{(g)}$ with a density function $p_{f^{(g)}}\left(\,\cdot\,;\boldsymbol{\gamma}_{f^{(g)}}\right)$, where $\boldsymbol{\gamma}_{f^{(g)}}$ denotes an unconstrained latent variable distribution parameter for the g^{th} group. We assume that $y_{1i}^{(g)}, \cdots, y_{pi}^{(g)}$ are conditionally independent given $\boldsymbol{f}_i^{(g)}$ and that each observed variable given $\boldsymbol{f}_i^{(g)}$ is from a regular exponential family with canonical parameterization. That is,

$$p_{y_k^{(g)}}\left(y_{ki}^{(g)} \mid \boldsymbol{f}_i^{(g)}; \eta_{ki}^{(g)}, \varphi_k^{(g)}\right) = \exp\left\{\frac{\eta_{ki}^{(g)} y_{ki}^{(g)} - b\left(\eta_{ki}^{(g)}\right)}{a\left(\varphi_k^{(g)}\right)} + c\left(y_{ki}^{(g)}, \varphi_k^{(g)}\right)\right\} \tag{1}$$

for $k = 1, \cdots, p$, where

$$\begin{aligned} \mathrm{E}\left(y_{ki}^{(g)}\right) &= b'\left(\eta_{ki}^{(g)}\right), \\ \mathrm{Var}\left(y_{ki}^{(g)}\right) &= a\left(\varphi_k^{(g)}\right) b''\left(\eta_{ki}^{(g)}\right). \end{aligned}$$

The canonical link function $g_k(\cdot)$ relates $\mathrm{E}\left(y_{ki}^{(g)}\right)$ to the linear function $\alpha_k^{(g)} + \boldsymbol{\beta}_k^{(g)\prime}\boldsymbol{f}_i^{(g)}$, i.e.,

$$g_k\left(\mathrm{E}\left(y_{ki}^{(g)}\right)\right) = \alpha_k^{(g)} + \boldsymbol{\beta}_k^{(g)\prime}\boldsymbol{f}_i^{(g)} = \eta_{ki}^{(g)}. \tag{2}$$

The general formulation of the GLLVM approach in (1) allows for various type of outcome variables, e.g., normal, gamma, Poisson, binomial, etc., though outcomes don't have to come from the same exponential family. For example, let $y_{ki}^{(g)} \mid \boldsymbol{f}_i^{(g)}$ be a dichotomous variable with categories 0 and 1 (e.g., 0 = "disagree", 1 = "agree"). In this case the canonical link function is given by

$$g_k\left(\mathrm{E}\left(y_{ki}^{(g)}\right)\right) = \ln\left(\frac{\mathrm{E}\left(y_{ki}^{(g)}\right)}{1-\mathrm{E}\left(y_{ki}^{(g)}\right)}\right),$$

where

$$\mathrm{E}\left(y_{ki}^{(g)}\right) = \frac{1}{1+\exp\left\{-\left(\alpha_k^{(g)} + \boldsymbol{\beta}_k^{(g)\prime}\boldsymbol{f}_i^{(g)}\right)\right\}}.$$

This model is equivalent to the model that uses an underlying variable for each dichotomous response, and its threshold (see, e.g., Christoffersson 1975; Muthén 1984). With the restriction that the underlying variable for each outcome has mean zero and variance one, the threshold parameter in the latter model is given by $\alpha_k^{(g)}/\sqrt{1-\boldsymbol{\beta}_k^{(g)\prime}\boldsymbol{\Sigma}_{f^{(g)}}\boldsymbol{\beta}_k^{(g)}}$, and the latent variable slope parameter is given by $-\boldsymbol{\beta}_k^{(g)}/\sqrt{1-\boldsymbol{\beta}_k^{(g)\prime}\boldsymbol{\Sigma}_{f^{(g)}}\boldsymbol{\beta}_k^{(g)}}$. However, the threshold model with common mean and variance over groups limits its use in multi-group analysis where comparisons of group characteristics may be of interest.

Model (1) contains the factor indeterminacy inherent in this type of latent variable models. That is, the same model can be expressed using transformed parameters and factors. To remove this indeterminacy we use a parameterization suitable for multi-group analysis (see, e.g., Fuller 1987; Wall and Amemiya 2000). With possible re-ordering of p outcome variables, $k = 1, \cdots, p$, we assume that, for the first q outcome variables,

$$\left(g_1\left(\mathrm{E}\left(y_{1i}^{(g)}\right)\right), \cdots, g_q\left(\mathrm{E}\left(y_{qi}^{(g)}\right)\right)\right)' = \mathbf{0}_q + \boldsymbol{I}_q\boldsymbol{f}_i^{(g)},$$

where $g_k(\,\cdot\,)$ is given in (2) and q denotes the dimension of the latent variable. This is a interpretable and meaningful identification parameterization where the group characteristic parameters $\boldsymbol{\gamma}_{f^{(g)}}$ are unrestricted. Therefore, differences between groups can be assessed by comparing $\boldsymbol{\gamma}_{f^{(g)}}$ over different groups.

In our model, the parameters $\alpha_k^{(g)}$, $\boldsymbol{\beta}_k^{(g)}$, and $\varphi_k^{(g)}$ describe the measurement properties for the k^{th} case outcome variable. For many studies, the same instrument is administered to all groups. For this typical multi-group analysis situation, it is reasonable to assume that

the measurement properties are invariant over groups, i.e., for $k = q+1, \cdots, p$,

$$\begin{aligned} \alpha_k^{(1)} &= \cdots = \alpha_k^{(G)} = \alpha_k, \\ \boldsymbol{\beta}_k^{(1)} &= \cdots = \boldsymbol{\beta}_k^{(G)} = \boldsymbol{\beta}_k, \\ \varphi_k^{(1)} &= \cdots = \varphi_k^{(G)} = \varphi_k. \end{aligned} \tag{3}$$

Thus, for the brevity of the presentation, we assume that model (1) satisfies the assumptions of common measurement properties (3) over groups. However, our approach and methodology can be used without these assumptions.

3 MAXIMUM LIKELIHOOD ESTIMATION VIA MONTE CARLO EM ALGORITHM

3.1 Monte Carlo EM Algorithm

We consider now full maximum likelihood estimation for the parameters of the GLLVM in (1) which includes the measurement parameter $\alpha_k, \boldsymbol{\beta}_k, \varphi_k$ and the latent variable distribution parameter $\boldsymbol{\gamma}_{f^{(g)}}$. Let $\boldsymbol{y} = \left(\boldsymbol{y}^{(1)\prime}, \cdots, \boldsymbol{y}^{(G)\prime}\right)'$ and $\boldsymbol{f} = \left(\boldsymbol{f}^{(1)\prime}, \cdots, \boldsymbol{f}^{(G)\prime}\right)'$. Moreover, for $k = 1, \cdots, p$, let

$$\boldsymbol{\theta}_k = \left(\alpha_k, \boldsymbol{\beta}', \varphi\right)', \tag{4}$$

$$\boldsymbol{\psi} = \left(\boldsymbol{\theta}_1', \cdots, \boldsymbol{\theta}_p', \boldsymbol{\gamma}_{f^{(1)}}', \cdots, \boldsymbol{\gamma}_{f^{(G)}}'\right)'.$$

For future reference we also define the g^{th} group specific parameter

$$\boldsymbol{\psi}^{(g)} = \left(\boldsymbol{\theta}_1', \cdots, \boldsymbol{\theta}_p', \boldsymbol{\gamma}_{f^{(g)}}'\right)', \tag{5}$$

where the measurement parameter $\boldsymbol{\theta}_k$ is common for all G groups.

The log-likelihood function is given by

$$\ell\left(\boldsymbol{\psi} \mid \boldsymbol{y}\right) = \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \ln \int p_{y_k^{(g)}}\left(y_{ki}^{(g)} \mid \boldsymbol{f}_i^{(g)}; \boldsymbol{\theta}_k\right) dP_{\boldsymbol{f}_i^{(g)}}. \tag{6}$$

In general, direct maximization of (6) is computationally difficult, involving intractable multiple integrals which cannot be evaluated in closed form. To solve this difficulty, we treat the latent variable $\boldsymbol{f}^{(g)}$ as a missing variable, and utilize the EM approach (Dempster, Laird, and Rubin 1977). In this approach, the complete log-likelihood function

$$\ell^c\left(\boldsymbol{\psi} \mid \boldsymbol{y}, \boldsymbol{f}\right) = \sum_{g=1}^{G} \ell^c\left(\boldsymbol{\psi}^{(g)} \mid \boldsymbol{y}^{(g)}, \boldsymbol{f}^{(g)}\right), \tag{7}$$

does not involve integration. The E-step of the EM approach computes the conditional expectation of (7) given the observed variables evaluated at the current parameter estimate. Specifically, the E-step at iteration $(j+1)$ computes

$$Q\left(\boldsymbol{\psi} \mid \boldsymbol{\psi}_{(j)}\right) = \sum_{g=1}^{G} \mathrm{E}\left[\ell^c\left(\boldsymbol{\psi}^{(g)} \mid \boldsymbol{y}^{(g)}, \boldsymbol{f}^{(g)}\right) \mid \boldsymbol{y}^{(g)}; \boldsymbol{\psi}_{(j)}^{(g)}\right], \tag{8}$$

where $\boldsymbol{\psi}_{(j)}$ and $\boldsymbol{\psi}_{(j)}^{(g)}$ are the estimates of $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^{(g)}$ at iteration (j). Each conditional expectation in (8) is with respect to the conditional density

$$p_{f^{(g)}|y^{(g)}}\left(\boldsymbol{f}_i^{(g)} \mid \boldsymbol{y}_i^{(g)}; \boldsymbol{\psi}_{(j)}^{(g)}\right). \tag{9}$$

However, no closed form is generally available for this conditional density. In our approach, we rewrite (9) as

$$\begin{aligned}
& p_{f^{(g)}|y^{(g)}}\left(\boldsymbol{f}_i^{(g)} \mid \boldsymbol{y}_i^{(g)}; \boldsymbol{\psi}_{(j)}^{(g)}\right) \\
= & \frac{p_{y^{(g)}|f^{(g)}}\left(\boldsymbol{y}_i^{(g)} \mid \boldsymbol{f}_i^{(g)}; \boldsymbol{\psi}_{(j)}^{(g)}\right)}{\int p_{y^{(g)}|f^{(g)}}\left(\boldsymbol{y}_i^{(g)} \mid \boldsymbol{f}_i^{(g)}; \boldsymbol{\psi}_{(j)}^{(g)}\right) p_{f^{(g)}}\left(\boldsymbol{f}_i^{(g)}; \boldsymbol{\gamma}_{f^{(g)}(j)}\right) d\boldsymbol{f}_i^{(g)}} \quad p_{f^{(g)}}\left(\boldsymbol{f}_i^{(g)}; \boldsymbol{\gamma}_{f^{(g)}(j)}\right) \\
= & \frac{\exp\left\{\sum_{k=1}^{p}\left(\frac{\eta_{ki}^{(g)} y_{ki}^{(g)} - b\left(\eta_{ki}^{(g)}\right)}{a\left(\varphi_{k(j)}\right)} + c\left(y_{ki}^{(g)}, \varphi_{k(j)}\right)\right)\right\}}{\int \exp\left\{\sum_{k=1}^{p}\left(\frac{\eta_{ki}^{(g)} y_{ki}^{(g)} - b\left(\eta_{ki}^{(g)}\right)}{a\left(\varphi_{k(j)}\right)} + c\left(y_{ki}^{(g)}, \varphi_{k(j)}\right)\right)\right\} p_{f^{(g)}}\left(\boldsymbol{f}_i^{(g)}; \boldsymbol{\gamma}_{f^{(g)}(j)}\right) d\boldsymbol{f}_i^{(g)}} \\
\times & \; p_{f^{(g)}}\left(\boldsymbol{f}_i^{(g)}; \boldsymbol{\gamma}_{f^{(g)}(j)}\right) \\
= & \; w_i^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right) p_{f^{(g)}}\left(\boldsymbol{f}_i^{(g)}; \boldsymbol{\gamma}_{f^{(g)}(j)}\right)
\end{aligned} \tag{10}$$

and use this expression in our simulation based (Monte Carlo) E-step.

3.2 Monte Carlo E-Step

We use (10) to express the conditional expectation of the complete log-likelihood function given the observed variables at iteration $(j+1)$ as

$$\begin{aligned}
& Q\left(\boldsymbol{\psi} \mid \boldsymbol{\psi}_{(j)}\right) \\
= & \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \int \ell^{c}\left(\boldsymbol{\psi}^{(g)} \mid \boldsymbol{y}_{i}^{(g)}, \boldsymbol{f}_{i}^{(g)}\right) w_{i}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right) p_{f^{(g)}}\left(\boldsymbol{f}_{i}^{(g)} ; \boldsymbol{\gamma}_{f^{(g)}(j)}\right) d \boldsymbol{f}_{i}^{(g)} \\
= & \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \int\left(\frac{\eta_{ki}^{(g)} y_{ki}^{(g)}-b\left(\eta_{ki}^{(g)}\right)}{a\left(\varphi_{k}\right)}+c\left(y_{ki}^{(g)}, \varphi_{k}\right)+\ln p_{f^{(g)}}\left(\boldsymbol{f}_{i}^{(g)} ; \boldsymbol{\gamma}_{f^{(g)}}\right)\right) \\
& \times w_{i}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right) p_{f^{(g)}}\left(\boldsymbol{f}_{i}^{(g)} ; \boldsymbol{\gamma}_{f^{(g)}(j)}\right) d \boldsymbol{f}_{i}^{(g)}. \qquad (11)
\end{aligned}$$

Evaluation of the expression above requires integration over $\boldsymbol{f}_i^{(g)}$. Note that evaluation of the denominator of $w_i^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right)$ itself also requires integration over $\boldsymbol{f}_i^{(g)}$. Numerical integration approximation such as Gauss-Hermite quadrature may be possible if the dimension of $\boldsymbol{f}^{(g)}$ is very small. However, for higher dimensional $\boldsymbol{f}_i^{(g)}$, any numerical integral approximation becomes computationally infeasible and unreliable (Meng and Schilling 1996). Our approach is to take advantage of (11), and use a version of Monte Carlo integration. At iteration $(j+1)$ of our EM algorithm, we draw a large number M of independent samples,

$$\widehat{\boldsymbol{f}}_{1 i}^{(g)}, \cdots, \widehat{\boldsymbol{f}}_{M i}^{(g)} \sim p_{f^{(g)}}\left(\cdot ; \boldsymbol{\gamma}_{f^{(g)}(j)}\right)$$

and approximate (11) by

$$\begin{aligned}
& \widehat{Q}\left(\boldsymbol{\psi} \mid \boldsymbol{\psi}_{(j)}\right) \\
= & \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{m=1}^{M} \ell^{c}\left(\boldsymbol{\psi}^{(g)} \mid \boldsymbol{y}_{i}^{(g)}, \widehat{\boldsymbol{f}}_{m i}^{(g)}\right) \widehat{w}_{i}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right) \\
= & \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \sum_{m=1}^{M}\left(\frac{\widehat{\eta}_{k m i}^{(g)} y_{k i}^{(g)}-b\left(\widehat{\eta}_{k m i}^{(g)}\right)}{a\left(\varphi_{k}\right)}+c\left(y_{k i}^{(g)}, \varphi_{k}\right)+\ln p_{f^{(g)}}\left(\widehat{\boldsymbol{f}}_{m i}^{(g)} ; \boldsymbol{\gamma}_{f^{(g)}}\right)\right) \\
& \times \widehat{w}_{m i}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right), \qquad (12)
\end{aligned}$$

where

$$\widehat{\eta}_{k m i}^{(g)}=\alpha_{k}+\boldsymbol{\beta}^{\prime} \widehat{\boldsymbol{f}}_{m i}^{(g)},$$

and

$$\begin{aligned}
\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right) &= \frac{\exp\left\{\sum_{k=1}^{p}\left(\frac{\widehat{\eta}_{kmi(j)}^{(g)}y_{ki}^{(g)}-b\left(\widehat{\eta}_{kmi(j)}^{(g)}\right)}{a\left(\varphi_{k(j)}\right)}+c\left(y_{ki}^{(g)},\varphi_{k(j)}\right)\right)\right\}}{\sum_{h=1}^{M}\exp\left\{\sum_{k=1}^{p}\left(\frac{\widehat{\eta}_{khi(j)}^{(g)}y_{ki}^{(g)}-b\left(\widehat{\eta}_{khi(j)}^{(g)}\right)}{a\left(\varphi_{k(j)}\right)}+c\left(y_{ki}^{(g)},\varphi_{k)(j)}\right)\right)\right\}} \\
&= \frac{p_{y^{(g)}|f^{(g)}}\left(\boldsymbol{y}_{i}^{(g)}\mid\widehat{\boldsymbol{f}}_{mi}^{(g)};\boldsymbol{\psi}_{(j)}^{(g)}\right)}{\sum_{h=1}^{M}p_{y^{(g)}|f^{(g)}}\left(\boldsymbol{y}_{i}^{(g)}\mid\widehat{\boldsymbol{f}}_{hi}^{(g)};\boldsymbol{\psi}_{(j)}^{(g)}\right)},
\end{aligned}$$

with

$$\widehat{\eta}_{kmi(j)}^{(g)}=\alpha_{k(j)}+\boldsymbol{\beta}_{(j)}'\widehat{\boldsymbol{f}}_{mi}^{(g)}.$$

Note that, for any given i and g, $\sum_{m=1}^{M}\widehat{w}_{mi}^{(g)}=1$.

As the Monte Carlo sample size M increases, $\widehat{Q}\left(\boldsymbol{\psi}\mid\boldsymbol{\psi}_{(j)}\right)$ approaches $Q\left(\boldsymbol{\psi}\mid\boldsymbol{\psi}_{(j)}\right)$. An adapted method of choosing M is a part of our algorithm, and is discussed in section 3.4.

3.3 Maximization Step

Because of the conditional independence of the observed variables given the latent variable, we can separate the parameter space of $\boldsymbol{\psi}$ into components corresponding to each outcome variable and the latent variable. Therefore expression (12) can be written as

$$\widehat{Q}\left(\boldsymbol{\psi}\mid\boldsymbol{\psi}_{(j)}\right)=\sum_{k=1}^{p}\widehat{Q}_{k}\left(\boldsymbol{\theta}_{k}\mid\boldsymbol{\psi}_{(j)}\right)+\sum_{g=1}^{G}\widehat{Q}_{f^{(g)}}\left(\boldsymbol{\gamma}_{f^{(g)}}\mid\boldsymbol{\psi}_{(j)}\right),$$

with

$$\begin{aligned}
\widehat{Q}_{k}\left(\boldsymbol{\theta}_{k}\mid\boldsymbol{\psi}_{(j)}\right) &= \sum_{g=1}^{G}\sum_{i=1}^{N^{(g)}}\sum_{m=1}^{M}\ln p_{y_{k}^{(g)}}\left(y_{ki}^{(g)}\mid\widehat{\boldsymbol{f}}_{mi}^{(g)};\boldsymbol{\theta}_{k}\right)\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right) \\
&= \sum_{g=1}^{G}\sum_{i=1}^{N^{(g)}}\sum_{m=1}^{M}\left(\frac{\widehat{\eta}_{km}^{(g)}y_{ki}^{(g)}-b\left(\widehat{\eta}_{km}^{(g)}\right)}{a\left(\varphi_{k}\right)}+c\left(y_{ki}^{(g)},\varphi_{k}\right)\right)\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right),
\end{aligned}$$

and

$$\widehat{Q}_{f^{(g)}}\left(\boldsymbol{\gamma}_{f^{(g)}}\mid\boldsymbol{\psi}_{(j)}\right)=\sum_{i=1}^{N^{(g)}}\sum_{m=1}^{M}\ln p_{f^{(g)}}\left(\widehat{\boldsymbol{f}}_{mi}^{(g)};\boldsymbol{\gamma}_{f^{(g)}}\right)\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right),$$

where $\boldsymbol{\theta}_k$ is defined in (4) and $\boldsymbol{\gamma}_{f^{(g)}}$ denotes the latent variable distribution parameter. Hence we can maximize (12) by maximizing each $\widehat{Q}_k$ with respect to $\boldsymbol{\theta}_k$ and each $\widehat{Q}_{f^{(g)}}$ with respect to $\boldsymbol{\gamma}_{f^{(g)}}$ separately. Note that $\widehat{Q}_k$ and $\widehat{Q}_{f^{(g)}}$ are in the form of a weighted likelihood. Thus each maximization can be carried out by modifying the existing maximum likelihood procedures. The $\widehat{Q}_k$ is a weighted generalized linear model likelihood, and can be maximized by modified iteratively re-weighted least squares with an additional weight $\widehat{w}_{mi}^{(g)}$. The $\widehat{Q}_{f^{(g)}}$ depend on the density function $p_{f^{(g)}}(\,\cdot\,;\boldsymbol{\gamma}_{f^{(g)}})$. A maximum likelihood algorithm for the distribution $p_{f^{(g)}}(\,\cdot\,;\boldsymbol{\gamma}_{f^{(g)}})$ can be modified to include the weight $\widehat{w}_{mi}^{(g)}$ for maximization of $\widehat{Q}_{f^{(g)}}$. If $\boldsymbol{f}^{(g)}$ is normally distributed, i.e., if $p_{f^{(g)}}(\,\cdot\,;\boldsymbol{\gamma}_{f^{(g)}})$ is $n\left(\,\cdot\mid \boldsymbol{\mu}_{f^{(g)}}, \boldsymbol{\Sigma}_{f^{(g)}}\right)$, then the closed form solutions for the next step estimate $\widehat{\boldsymbol{\mu}}_{(j+1)}$ and $\widehat{\boldsymbol{\Sigma}}_{(j+1)}$ are given by

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_{f^{(g)}(j+1)} &= \frac{\sum_{i=1}^{N^{(g)}}\sum_{m=1}^{M}\widehat{\boldsymbol{f}}_{mi}^{(g)}\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right)}{\sum_{i=1}^{N^{(g)}}\sum_{m=1}^{M}N^{(g)}\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right)},\\
\widehat{\boldsymbol{\Sigma}}_{f^{(g)}(j+1)} &= \frac{\sum_{i=1}^{N^{(g)}}\sum_{m=1}^{M}\left(\widehat{\boldsymbol{f}}_{mi}^{(g)}-\boldsymbol{\mu}_{f^{(g)}}^{(j+1)}\right)\left(\widehat{\boldsymbol{f}}_{mi}^{(g)}-\boldsymbol{\mu}_{f^{(g)}}^{(j+1)}\right)'\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{(j)}^{(g)}\right)}{\sum_{i=1}^{N^{(g)}}\sum_{m=1}^{M}N^{(g)}\widehat{w}_{mi}^{(g)}\left(\boldsymbol{\psi}_{j}^{(g)}\right)}.
\end{aligned}$$

3.4 Adaptive method of checking convergence and choosing the Monte Carlo sample size

Assessing the convergence of the Monte Carlo EM algorithm is a challenging task. A method suggested in the literature is to plot each component of the parameter vector versus the iteration step j, and to terminate the algorithm when each plot shows only small fluctuations around an horizontal line. In addition to the difficulty associated with subjective examinations of plots, this method is limited to cases with only a few parameters. In our problem, the number parameters can be very large, even for a moderately sized model. Also, a plot of $\boldsymbol{\psi}_{(j)}$ versus the iteration step (j) does not by itself differentiate fluctuations over the iterations and the variability due to the Monte Carlo integration. In particular, the celebrated monotonicity property of the EM algorithm no longer holds and the change in the parameter values from iteration (j) to $(j+1)$ may be swamped by the Monte Carlo error. The Monte

Carlo error can be controlled by the Monte Carlo sample size. Therefore it makes sense to increase the Monte Carlo sample size over the EM algorithm iterations. Several ad-hoc strategies, where the Monte Carlo sample size is deterministically increased over iterations, have been proposed (see, e.g., Chan and Kuk 1997; McCulloch 1997; Booth and Hobert 1999).

We introduce a likelihood based approach, under which the Monte Carlo sample determination and convergence assessment are performed for each iteration at the same time. The basic idea behind our approach is to measure the distance between any two values of the parameter vector by the difference in log-likelihoods evaluated at them.

For an absolute distance criterion for the likelihood, we hypothetically consider performing the likelihood ratio test for the goodness of fit of the proposed model within a natural unrestricted model. Then, the asymptotic likelihood ratio goodness of fit test compares the difference in minus twice the log-likelihood evaluated at the maximum likelihood estimate under the proposed model and at the maximum likelihood estimate under the unrestricted model to a chi-squares distribution with degrees of freedom given by the difference in the number of parameters. For most values of degrees of freedom, a small difference in the test statistic value, e.g., $\delta_0 = 0.005$, 0.05, or even 0.5, has no practical effect on statistical judgment. However, the test statistic value typically increases with the sample size, because the approximate non-centrality parameter is of order $N = \sum N^{(g)}$, the sample size. Taking into account the effect of different sample sizes N, we consider the log-likelihood difference to be neglible if it is smaller in absolute value than $\sqrt{N}\delta = \delta_0/2$ for a fixed value δ. That is, the estimates from iteration (j) to $(j+1)$ are considered sufficient close to each other if

$$\left|\ell\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y}\right) - \ell\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right)\right| < \sqrt{N}\delta. \tag{13}$$

For example, the practical small difference of $\delta_0 = 0.5$ for likelihood ratio test is achieved for any $N \leq 10,000$ if we choose $\delta = 0.0025$. The same is achieved for any $N \leq 1,000$ if we choose $\delta = 0.008$. This gives us a way to set a criterion for the log-likelihood difference.

However, in the Monte Carlo EM algorithm, the log-likelihood functions in (13) are not explicitly available. Thus, we use Monte Carlo integration to obtain a probability upper

bound for the difference in (13). For this, let

$$\widehat{\boldsymbol{f}}_{1i}^{(g)}, \cdots, \widehat{\boldsymbol{f}}_{Mi}^{(g)} \sim p_{f^{(g)}}\left(\cdot \, ; \boldsymbol{\gamma}_{f^{(g)}(j+1)} \right)$$

be the independent Monte Carlo sample at iteration $(j+1)$ of the Monte Carlo EM algorithm described in Section 3.2. Then, an approximate value of the log-likelihood function evaluated at $\boldsymbol{\psi}_{(j+1)}$ is given by

$$\widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right) = \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \ln \widehat{p}_{y_k^{(g)}}\left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)}\right),$$

where

$$\widehat{p}_{y_k^{(g)}}\left(y_{ki}^{(g)}; \boldsymbol{\psi}_{(j+1)}\right) = \frac{1}{M} \sum_{m=1}^{M} p_{y_k^{(g)}}\left(y_{ki}^{(g)} \mid \widehat{\boldsymbol{f}}_{mi}^{(g)}; \boldsymbol{\theta}_{k(j+1)}\right)$$

and $\boldsymbol{\theta}_k$ is defined in (4). Using this notation, and the triangle inequality we note that

$$\begin{aligned} \left|\ell\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y}\right) - \ell\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right)\right| \;\leq\; & \left|\widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right) - \ell\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right)\right| \\ & + \left|\widehat{\ell}\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y}\right) - \ell\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y}\right)\right| \\ & + \left|\widehat{\ell}\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y}\right) - \widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right)\right|. \end{aligned} \tag{14}$$

Our approach is to bound each of the three terms of the right hand of (14) by $\delta_1 = \sqrt{N}\delta/3$. The third term on the right-hand side of (14) is observable, and its size can be assessed directly. The first two terms are of the same form corresponding to two consecutive iterations. For the first term, over the Monte Carlo simulation distribution,

$$\begin{aligned} & \widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right) - \ell\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right) \\ = \; & \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \ln \left(\frac{\widehat{p}_{y_k^{(g)}}\left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)}\right)}{p_{y_k^{(g)}}\left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)}, \boldsymbol{\gamma}_{f^{(g)}(j+1)}\right)} \right) \\ = \; & \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \left(\frac{\widehat{p}_{y_k^{(g)}}\left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)}\right)}{p_{y_k^{(g)}}\left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)}, \boldsymbol{\gamma}_{f^{(g)}(j+1)}\right)} - 1 \right) + O_p\left(\frac{1}{M}\right) \\ = \; & \frac{1}{M} \sum_{m=1}^{M} \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \left(\frac{p_{y_k^{(g)}}\left(y_{ki}^{(g)} \mid \widehat{\boldsymbol{f}}_{mi}^{(g)}; \boldsymbol{\theta}_{k(j+1)}\right)}{p_{y_k^{(g)}}\left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)}, \boldsymbol{\gamma}_{f^{(g)}(j+1)}\right)} - 1 \right) + O_p\left(\frac{1}{M}\right), \end{aligned}$$

as $M \longrightarrow \infty$. Hence, ignoring the terms $O_p\left(\frac{1}{M}\right)$,

$$\mathrm{E}\left[\left|\ell\left(\boldsymbol{\psi}_{(j+1)} \mid \mathbf{y}\right) - \widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right)\right|^2\right] \approx \frac{V}{M},$$

where

$$V = \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \frac{\mathrm{Var}\left[p_{y_k^{(g)}} \left(y_{ki}^{(g)} \mid \widehat{\boldsymbol{f}}_{mi}^{(g)}; \boldsymbol{\theta}_{k(j+1)} \right) \right]}{p_{y_k}^2 \left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)}, \boldsymbol{\gamma}_{f^{(g)}(j+1)} \right)}, \tag{15}$$

and the variance is with respect to the distribution $p_{f^{(g)}}\left(\cdot \, ; \boldsymbol{\gamma}_{f^{(g)}(j+1)} \right)$. By the Markov's inequality, for any δ_1,

$$P\left(\left| \ell\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y} \right) - \widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y} \right) \right| > \delta_1 \right) \leq \frac{V}{M \delta_1^2}. \tag{16}$$

Again, using the Monte Carlo integration, we can approximate (15) by

$$\widehat{V}^{(j+1)} = \frac{1}{M^{(j)}} \sum_{g=1}^{G} \sum_{i=1}^{N^{(g)}} \sum_{k=1}^{p} \left[\frac{\sum_{m=1}^{M^{(j)}} p_{y_k^{(g)}}^2 \left(y_{ki}^{(g)} \mid \widehat{\boldsymbol{f}}_{mi}^{(g)}; \boldsymbol{\theta}_{k(j+1)} \right)}{\left(\widehat{p}_{y_k^{(g)}} \left(y_{ki}^{(g)}; \boldsymbol{\theta}_{k(j+1)} \right) \right)^2} - M^{(j)} \right], \tag{17}$$

where $M^{(j)}$ was the Monte Carlo sample size at the previous iteration. The probability in (16) can be made small, e.g., at most $\epsilon = 0.01$, if we choose the Monte Carlo sample size $M_0^{(j+1)}$ satisfying

$$M_0^{(j+1)} \geq \frac{\widehat{V}^{(j+1)}}{\epsilon \delta_1^2}.$$

Similarly, the Monte Carlo sample size of $M_0^{(j)}$ in the previous iteration should have been chosen to satisfy

$$P\left(\left| \ell\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y} \right) - \widehat{\ell}\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y} \right) \right| > \delta_1 \right) \leq \epsilon,$$

i.e.,

$$M_0^{(j)} \geq \frac{\widehat{V}^{(j)}}{\epsilon \delta_1^2}.$$

Then we set $M^{(j+1)}$ to be the maximum of $M_0^{(j)}$ and $M_0^{(j+1)}$ and compute the Monte Carlo log-likelihood functions $\widehat{\ell}\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y} \right)$ and $\widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y} \right)$ using the Monte Carlo sample size $M^{(j+1)}$. If the third term of the right hand of (14) is less than $\sqrt{N}\delta/3$, then the Monte Carlo EM algorithm is considered to have converged. Otherwise, we go to iteration $(j+2)$ where the Monte Carlo sample size $M^{(j+1)}$ is used in the E-step.

In practice we start with a fixed Monte Carlo sample size for a few iterations without checking the convergence. After this "burn in" phase, the following steps are performed at iteration $(j+1)$:

1. Obtain $\boldsymbol{\psi}_{(j+1)}$ based on the sample size $M^{(j)}$.

2. Compute expression $\widehat{V}^{(j+1)}$ of (17).

3. Compute the smallest integer $M_0^{(j+1)}$, satisfying

$$M_0^{(j+1)} \geq \frac{\widehat{V}^{(j+1)}}{\epsilon \delta_1^2},$$

and smallest integer $M_0^{(j)}$, satisfying

$$M_0^{(j)} \geq \frac{\widehat{V}^{(j)}}{\epsilon \delta_1^2},$$

and set

$$M^{(j+1)} = \max\left\{M_0^{(j)}, M_0^{(j+1)}\right\}.$$

4. Compute $\widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right)$ and $\widehat{\ell}\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y}\right)$ based on $M^{(j+1)}$. If

$$\left|\widehat{\ell}\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{y}\right) - \widehat{\ell}\left(\boldsymbol{\psi}_{(j+1)} \mid \boldsymbol{y}\right)\right| < \frac{1}{3}\sqrt{N}\delta,$$

then stop the Monte Carlo EM algorithm and report $\boldsymbol{\psi}_{(j+1)}$ as the maximum likelihood estimate. Otherwise, go to iteration $(j+2)$ and return to step 1.

3.5 Estimation of Standard Errors

Let $\widehat{\boldsymbol{\psi}}$ denote the maximum likelihood estimate obtained by the Monte Carl EM, and $\widehat{\boldsymbol{\psi}}^{(g)}$ be the part of $\widehat{\boldsymbol{\psi}}$ corresponding to $\boldsymbol{\psi}^{(g)}$ as defined in (5). Then the empirical observed information matrix (see, e.g, Meilijson 1989) is given by

$$\widehat{\boldsymbol{I}}_e\left(\widehat{\boldsymbol{\psi}}\right) = \sum_{g=1}^{G}\sum_{i=1}^{N^{(g)}} \frac{\partial}{\partial \boldsymbol{\psi}^{(g)}} \ln p_{y^{(g)}}\left(\boldsymbol{y}_i^{(g)}; \widehat{\boldsymbol{\psi}}^{(g)}\right)\left[\frac{\partial}{\partial \boldsymbol{\psi}^{(g)}} \ln p_{y^{(g)}}\left(\boldsymbol{y}_i^{(g)}; \widehat{\boldsymbol{\psi}}^{(g)}\right)\right]'. \tag{18}$$

Note that, the gradient vector of the complete log-likelihood function given the observed data for observation i in group g can be written as

$$\frac{\partial}{\partial \boldsymbol{\psi}^{(g)}} \mathrm{E}\left[\ell_i^c\left(\boldsymbol{\psi} \mid \boldsymbol{y}_i^{(g)}, \boldsymbol{f}_i^{(g)}\right) \mid \boldsymbol{y}_i^{(g)}\right] = \frac{\partial}{\partial \boldsymbol{\psi}^{(g)}} \ln p_{y^{(g)}}\left(\boldsymbol{y}_i^{(g)}; \boldsymbol{\psi}^{(g)}\right),$$

i.e., the individual score functions of the observed data can be computed as a by-product of the EM algorithm. Therefore, expression (18) can be approximated by

$$\widehat{\boldsymbol{I}}_e\left(\widehat{\boldsymbol{\psi}}\right) \approx \sum_{g=1}^{G}\sum_{i=1}^{N^{(g)}} \frac{\partial}{\partial \boldsymbol{\psi}^{(g)}} \mathrm{E}\left[\ell_i^c\left(\widehat{\boldsymbol{\psi}} \mid \boldsymbol{y}_i^{(g)}, \boldsymbol{f}_i^{(g)}\right) \mid \boldsymbol{y}_i\right]\left[\frac{\partial}{\partial \boldsymbol{\psi}^{(g)}} \mathrm{E}\left[\ell_i^c\left(\widehat{\boldsymbol{\psi}} \mid \mathbf{y_i^{(g)}}, \boldsymbol{f_i^{(g)}}\right) \mid \boldsymbol{y}_i^{(g)}\right]\right]',$$

where

$$\frac{\partial}{\partial \boldsymbol{\psi}^{(g)}} \mathrm{E}\left[\ell_i^c\left(\widehat{\boldsymbol{\psi}} \mid \boldsymbol{y}_i^{(g)}, \boldsymbol{f}_i^{(g)}\right) \mid \boldsymbol{y}_i^{(g)}\right]$$

can be obtained in the same way as described in Section 3.2 using Monte Carlo integration. This Monte Carlo integration needs to be performed only once after the convergence of the Monte Carlo EM algorithm has been determined.

4 AN EXAMPLE

To demonstrate our multi-group GLLVM approach with a real life example, we analyze data from a substance abuse prevention study. The data are from the Project Family study, which was conducted at the Institute for Social and Behavioral Research at Iowa State University. Project Family investigations address the efficacy of universal competency-training interventions for families with young adolescents, factors influencing family participation in these interventions and strategies for the dissemination of validated preventive interventions (Spoth et al. 1998). Participants in the study were families of sixth grader enrolled in 33 rural schools of a Midwestern state. Sixth graders from these schools and their families were randomly assigned to one of three experimental conditions: the seven-session *Iowa Strengthening Families Program* (ISFP), the five session *Preparing for the Drug Free Years Program* (PDFY), and a minimal-contact control condition. Data were based on parent report and observational report items of in-home interviews. In this example, we analyze data which were collected one year after the initiation of the intervention programs. There were 134 families assigned to ISFP, 146 to PDFY, and 149 to the control group. Based on the knowledge of the underlying study, a confirmatory factor analysis models that involves two non-overlapping factors is used to analyze the data. The first factor is denoted as *Parent-Child Affective Quality* (PCAQ) and consists of four indicators. Two of them are based on a parent report while the remaining are based on an observational report. The parent report indicators are computed as averages over four 5-point Likert type questions. The observational report indicators are averages over three 7-point Likert type questions. Therefore all 4 indicators are treated as continuous responses. The second factor is denoted

as *Intervention Targeted Behaviors* (ITB). This factor consists of four parent reported indicators. Two of these parent reported indicators are averages over three 5-point Likert type questions. The other two indicators are both based on a single questions with 5-point Likert scale ranging from 1 = "strongly disagree" to 5 = "strongly agree". The indicator variables are denoted as follows: PosAffO = positive affect observational report (continuous), NegAffO = negative affect observational report (continuous), PosAffP = positive affect parent report (continuous), NegAffP = negative affect parent report (continuous), InvolveP = involving child in family activities and decisions parent report (continuous), CommP = communication with child parent report (continuous), RulConP = communications of substance use rules parents report (polytomous), and AngMgtP = anger management parents report (polytomous). The following confirmatory factor analysis model has been proposed

$$\begin{pmatrix} \text{PosAffO}_i^{(g)} \\ \text{InvolveP}_i^{(g)} \\ \text{NegAffO}_i^{(g)} \\ \text{PosAffP}_i^{(g)} \\ \text{NegAffP}_i^{(g)} \\ \text{CommP}_i^{(g)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \alpha_3 & \beta_{31} & 0 \\ \alpha_4 & \beta_{41} & 0 \\ \alpha_5 & \beta_{51} & 0 \\ \alpha_6 & 0 & \beta_{62} \end{pmatrix} \begin{pmatrix} 1 \\ f_{1i}^{(g)} \\ f_{2i}^{(g)} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i}^{(g)} \\ \epsilon_{2i}^{(g)} \\ \epsilon_{3i}^{(g)} \\ \epsilon_{4i}^{(g)} \\ \epsilon_{5i}^{(g)} \\ \epsilon_{6i}^{(g)} \end{pmatrix}$$

$$P\left(\text{RulConP}_i^{(g)} \leq c_1 \mid \boldsymbol{f}_i^{(g)}\right) = \frac{1}{1 + \exp\left\{-\left(\alpha_7 + \beta_{72} f_{2i}^{(g)}\right)\right\}}$$

$$P\left(\text{AngMgtP}_i^{(g)} \leq c_2 \mid \boldsymbol{f}_i^{(g)}\right) = \frac{1}{1 + \exp\left\{-\left(\alpha_8 + \beta_{82} f_{2i}^{(g)}\right)\right\}},$$

for g = ISFP, PDFY, Control, $c_1, c_2 = 1, \cdots, 4$, and $i = 1, \cdots, N^{(g)}$. We assume

$$\begin{pmatrix} f_{1i}^{(g)} \\ f_{2i}^{(g)} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{f_1}^{(g)} \\ \mu_{f_2}^{(g)} \end{pmatrix}, \begin{pmatrix} \sigma_{f_1}^{2(g)} & \sigma_{f_1 f_2}^{(g)} \\ \sigma_{f_2 f_1}^{(g)} & \sigma_{f_2}^{2(g)} \end{pmatrix}\right)$$

and $\boldsymbol{\epsilon}_i^{(g)} \sim N\left(\mathbf{0}, \boldsymbol{\Psi}^{(g)}\right)$, where $\boldsymbol{\Psi}^{(g)}$ is a diagonal matrix.

Note that this model allows to compare latent variable parameters of PCAQ and ITB across the three intervention groups which cannot be accomplished by using the underlying

response variable approach with a multi-stage WLS estimation procedure. We used the Monte Carlo EM algorithm as described in Section 3 to compute the maximum likelihood estimates. The initial Monte Carlo sample size in the first 10 iterations ("burn in" phase) was set to be $M = 50$, while the tolerance level of the likelihood ratio statistic, δ_0, was set to be 0.05, and the upper bound for the probability of the approximation error in (16), ϵ, was set to be 0.01. The Monte Carlo sample size increased steeply for about 80 iterations to about $10,000$ and stabilized after it. Convergence was determined after 108 iterations. The convergence of the Monte Carlo approximated log-likelihood function is monitored in Figure 1.

Figure 1 here

The estimates for the parameters corresponding to the latent variable for the three intervention groups are summarized in Table 1. The estimated correlations between the two latent constructs PCAQ and ITB are 0.346 for ISFP, 0.435 for PDFY, and 0.363 for the control group which indicates a positive linear association between the *Parent-Child Affective Quality* and *Intervention Targeted Behaviors* for all three experimental conditions. It appears that the variability of the factors PCAQ and ITB for the ISFP group is higher than for the PDFY and Control group. The researchers were particularly interested in comparing the factor means of PCAQ and ITB between the three intervention groups. A test of equivalence between the factor means of the three groups provided a p value of 0.0002 for PCAQ and 0.0014 for ITB, respectively. The p values of the two-sided pairwise tests for the factor means are summarized in Table 2. The results indicate that there is a significant intervention effect on both factors for ISFP and PDFY.

Table 1 here

Table 2 here

5 SUMMARY

In this paper we have proposed a generalized linear latent variable modeling approach which allows outcome variables from any regular exponential family. The existing approaches for latent variable models with non-normal outcome variables are restricted to finite discrete response variables and can be considered lacking for multi-sample situations. The parameterization used in our approach is particularly useful for coherent multi-sample situations and provides meaningful parameter interpretation for those parameters which vary across groups. Additionally, our approach allows a flexible choice for the distribution of the latent variable. We use an EM algorithm to compute full maximum likelihood estimates. The E-step can be conveniently carried out using Monte-Carlo integration. The M-step can be computed using an iteratively re-weighted least square procedure. We also introduced a novel procedure for determining the convergence of the Monte EM algorithm. This procedure is described in a general framework and can be easily applied to any situation where a Monte Carlo EM algorithm is used. We conclude that the approach described in this article provides an important contribution for multi-group analysis of latent variable models with non-normal outcome variables.

Acknowledgments

This work was partially supported by National Institute on Drug Abuse Research Grant DA 07029-01A1 and National Institute of Mental Health Grant MH 49217-01A1.

References

Bentler, P.M. (1995), *EQS: Structural Equation Program Manual*. Los Angeles: BMDP Statistical Software.

Bock, R.D., and Aitkin, M. (1981), Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*. New-York: Wiley.

Booth, J.G., and Hobert, J.P. (1999), Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, 61, 265–285.

Chan, J.S., and Kuk, A.Y. (1997), Maximum Likelihood Estimation for Probit-linear Mixed Models With Correlated Random Effects. *Biometrics*, 53, 86–97.

Christoffersson, A. (1975), Factor analysis of dichotomized variables. *Psychometrika*, 40, 5–32.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Fuller, W.A. (1987). *Measurement Error Models*, New-York: Wiley.

Jöreskog, K. (1990), New developments in LISREL: analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, 24, 387–404.

Jöreskog, K., and Sörbom (1996), *LISREL 8: User's Reference Guide*. Scientific Software International, Chicago.

Jöreskog, K., and Moustaki, I. (2001), Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research*, 36, 347–387.

Lee, S.Y., and Poon, W.Y. (1987), Two-step estimation of multivariate polychoric correlations. *Communications in Statistics: Theory and Methods*, 16, 307–320.

Lee, S.Y., Poon, W.Y., and Bentler, P.M. (1992), Structural equation models with continuous and polytomous variables *Psychometrika*, 57, 89–105.

McCulloch, C.E. (1997), Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92, 162–170.

Meilijson, I. (1989), A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society, Series B*, 51, 127–138.

Meng, X.L., and Schilling, S. (1996), Fitting full-information item factor models and and empirical investigation of bridge sampling. *Journal of the American Statistical Association*. 91, 1254–1267.

Muthén, B. (1978), Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551–560.

Muthén, B. (1984), A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, 49, 115–132.

Muthén, B. (1987), *LISCOMP: Analysis of linear statistical equations using a comprehensive measurement model*. Mooreville, IN: Scientific Software, Inc.

Muthén, B., and Muthén, L. (1998), *Mplus User's Guide* Los Angeles, CA: Muthén & Muthén.

Reboussin, B.A., and Liang, K.Y. (1998), An estimating equations approach for the LISCOMP model. *Psychometrika*, 63, 165–182.

Spoth, R.L., Redmond, C., and Shin, C. (1998), Direct and indirect latent-variable parenting outcomes of two universal family-focused preventive interventions: Extending a public health oriented research base, *Journal of Consulting and Clinical Psychology*, 66, 385–399.

Wall, M.M., and Amemiya, Y. (2000), Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, 95, 929-940.

Wilson, D. Wood, R.L., and Gibbons, R. (1991), *TESTFACT 2*. Chicago: Scientific Software International.

Figure 1: Values of the Monte Carlo approximated negative log-likelihood function at each EM iteration

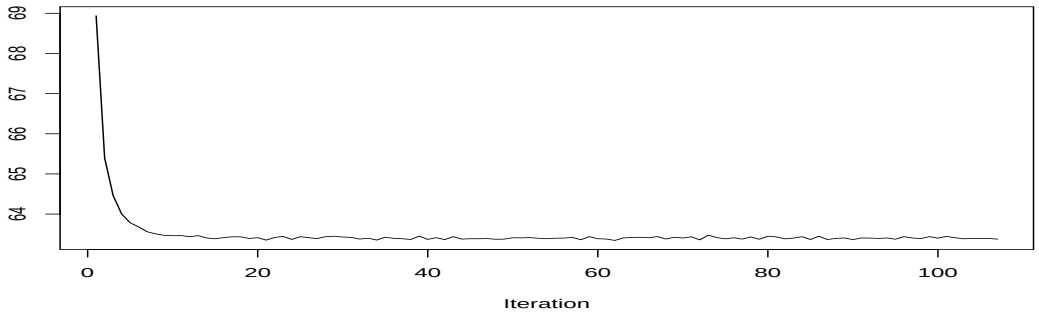

Table 1: Parameter estimates of latent variable means, variances, and covariances for the three intervention groups

Parameter	ISFP	PDFY	Control
μ_{PCAQ}	5.532	5.453	5.326
μ_{ITB}	3.772	3.929	3.695
σ^2_{PCAQ}	0.292	0.202	0.170
σ^2_{ITB}	0.242	0.207	0.231
$\sigma_{\text{PCAQ,ITB}}$	0.092	0.089	0.072

Table 2: p values for two-sided pairwise tests between ISFP, PDFY, and control group for PCAQ and ITB factor means

Contrast	PCAQ	ITB
ISFP vs. Control	< 0.0001	0.0353
PDFY vs. Control	< 0.0001	< 0.0001
ISFP vs. PDFY	0.1203	0.0031