Maximum Likelihood Estimation of Two-level Latent Variable Models with Mixed Continuous and Polytomous Data

Sik-Yum Lee
Department of Statistics,
The Chinese University of Hong Kong, Shatin, Hong Kong
email: sylee@sparc2.sta.cuhk.edu.hk

Jian-Qing Shi
Department of Computing Science
The University of Glasgow, Glasgow G12 8QQ, UK
email: shi@dcs.gla.ac.uk

SUMMARY. Two-level data with hierarchical structure and mixed continuous and polytomous data are very common in biomedical research. In this paper, we propose a maximum likelihood approach for analyzing a latent variable model with these data. The maximum likelihood estimates are obtained by a Monte-Carlo EM algorithm which involves the Gibbs sampler for approximating the E-step and the M-step, and the bridge sampling for monitoring the convergence. The approach is illustrated by a two-level data set concerning the development and preliminary findings from an AIDS preventative intervention for Filipina commercial sex workers, where relation between some latent quantities is investigated.

KEY WORDS: Factor analysis; Mixed continuous and polytomous data; Monte-Carlo EM algorithm; Two-level latent variable models.

1. Introduction

In much medical and other applied research, substantive theory and models involve latent variables that represent theoretical concepts or constructs that cannot be assessed by one single measurement on each individual under study. Examples of latent variables include health status, personality type, level of anxiety, etc. The relationships among latent variables and observed variables were originally analyzed by means of the factor analysis model (see Lawley and Maxwell, 1971). Recently, this useful model has been extended to various latent variable models, see Bentler (1992), Jöreskog and Sörbom (1996), Sammel and Ryan (1996), Sammel, Ryan and Legler (1997), among others. These models, which are motivated by different objectives, have been applied extensively in practice, see Bollen and Long (1993), Catalano and Ryan (1992), Legler and Ryan (1997) and references therein.

Owing to the nature of the variables or the design of questionnaires, polytomous variables that are defined through underlying continuous variables with some threshold parameters are very common in practice. Typical examples are that a subject is asked to respond on a scale like 'getting worse, no change, getting better' after a medical treatment, or 'unsatisfactory, no opinion, satisfactory' on the service of a hospital, etc. Recently, a lot of attention has been focused on the analysis of latent variable models with mixed polytomous and continuous variables, see for example Lee, Poon and Bentler (1995), Sammel, Ryan and Legler (1997), and Reboussin and Liang (1998), among others.

Methods developed in the work cited above depend on the assumption of independence among observations. However, in many applications, it is common to encounter hierarchically structured data that are collected from units that are nested within a large clusters (see Aitkin and Longford, 1986; Goldstein, 1986; among others). Examples could include random samples of patients from within random samples of hospitals, or individuals from within random samples of families. For such multilevel data, individuals within a cluster are expected to share certain common influential factors and to produce correlated observations. Hence, the assumption of independence is not realistic. The need to develop statistical theory for multilevel models that take into consideration the correlated structure is well recognized in latent variable modeling; see McDonald and Goldstein (1989), and Lee (1990). Computationally, Longford and Muthen (1992) derived a scoring type algorithm for maximum likelihood (ML) estimation of a special model with an invariant within-groups structure. Recently, Lee and Poon (1998) formulated the ML estimation of a more general model as an incomplete-data problem by treating random vectors at the group level as hypothetical missing data; and developed a procedure on the basis of the EM algorithm. These methods are all dealing with continuous data only. Since multilevel latent variable models with a mixture of continuous and polytomous data are much more difficult to handle, they cannot be analyzed by existing packages and so far very limited result has been achieved. The objective of this paper is to establish a ML procedure for a two-level factor analysis model with mixture of polytomous and continuous data.

The present work is motivated by a study (Morisky et al. 1998) of the effects of establishment policies, knowledge and attitudes on condom use among Filipina commercial sex workers (CSWs). It has been suggested that the nature of commercial sex work promotes the spread of AIDS and other sexually transmitted diseases. Promotion of safer sexual practice among CSWs is an important issue. The study of Morisky et al. (1998) concerned the development and preliminary findings from an AIDS preventative intervention for Filipina CSWs. The data set was collected from female CSWs in 97 establishments (bars, night clubs, Karaoke TV and massage parlours) in cities of Philippines. The entire questionnaire consisted of 134 items, covering the areas of demographics knowledge, attitudes, beliefs, behaviours, self-efficacy for condom use, and social desirability. Latent psychological determinants such as CSWs risk behaviors, knowledge and attitudes associated with AIDs and condom use are important issues to be assessed and will be used for modifying the planned intervention. The primary goal of our analysis is to apply a latent variable model, namely the factor analysis model, to determine the relationships between latent quantities of CSWs. The manifest variables that were used as indicators for latent quantities were measured in terms of polytomous and continuous scales. Also, as emphasized by Morisky et al (1998), establishments policies on their CSWs condom use practices exert a strong influence on CSWs. Hence, it is necessary to develop a new methodology to take into account the influence of establishments and the discrete nature of the polytomous data in the ML analysis.

Owing to the complexity of the proposed model, it is not feasible to obtain the ML estimate via direct maximization of the observed data log-likelihood function. Inspired by the useful strategy suggested in Rubin (1991), Meng and Schilling (1996), and Shi and Lee (2000), we will treat the latent factor scores and unobserved measurements of the continuous variables that underlie the polytomous variables as hypothetical missing quantities, and use a Monte Carlo-EM (MCEM, see Wei and Tanner, 1990) algorithm to produce the ML estimate. Since closed form solution can be derived, the computation is not heavy. The organization of the paper is as follows. A description of the model is given in Section 2. Section 3 develops ML estimation of the model with the MCEM algorithm. An illustrative example based on the AIDS data set is provided in Section 4. A discussion is given in Section 5. Some technical details are given in the Appendix.

2. A Latent Variable Model with Mixed Type Variables

Consider a data set with two levels, in which there are G distinct groups, indexed by $g = 1, \cdots, G$, and within each group, there are N_g p-variate random vectors $\boldsymbol{u}_{gi}$, $i = 1, \cdots, N_g$. At the first level, we assume that conditional on the group mean $\boldsymbol{v}_g$, random observations in each group have the following structure:

$$\boldsymbol{u}_{gi} = \boldsymbol{v}_g + \boldsymbol{v}_{gi} = \boldsymbol{v}_g + \boldsymbol{\Lambda}_{wg}\boldsymbol{\xi}_{gi} + \boldsymbol{\delta}_{gi}, \quad g = 1, \cdots, G, \quad i = 1, \cdots, N_g \qquad (1)$$

where $\mathbf{\Lambda}_{wg}$ is a $p \times r_1$ matrix of factor loadings, $\boldsymbol{\xi}_{gi}$ is a $r_1 \times 1$ random vector of latent factor scores, and $\boldsymbol{\delta}_{gi}$ is a $p \times 1$ random vector of error measurements. Following the usual assumptions for latent variable models, it is assumed that $\boldsymbol{\xi}_{gi}$ and $\boldsymbol{\delta}_{gi}$ are independently distributed as $N(\mathbf{0},\ \mathbf{\Gamma}_{wg})$ and $N(\mathbf{0},\ \mathbf{\Psi}_{wg})$, respectively, where $\mathbf{\Psi}_{wg}$ is a diagonal matrix. For different g, N_g can be different, so we are dealing with a model with an unbalanced design. At the second level, we assume that the group mean $\boldsymbol{v}_g$ has the structure

$$\boldsymbol{v}_g = \mathbf{\Lambda}_b \boldsymbol{\xi}_g + \boldsymbol{\delta}_g, \quad g = 1, \cdots, G, \tag{2}$$

where $\mathbf{\Lambda}_b$ is a $p \times r_2$ matrix of factor loadings, $\boldsymbol{\xi}_g$ is a $r_2 \times 1$ random vector of latent factor scores with distribution $N(\mathbf{0},\ \mathbf{\Gamma}_b)$, and $\boldsymbol{\delta}_g$ is a $p \times 1$ random vector of error measurements with distribution $N(\mathbf{0},\ \mathbf{\Psi}_b)$, where $\mathbf{\Psi}_b$ is a diagonal matrix. Similarly, $\boldsymbol{\xi}_g$ and $\boldsymbol{\delta}_g$ are assumed to be independent. Moreover, the first level latent vectors $\boldsymbol{\xi}_{gi}$ and $\boldsymbol{\delta}_{gi}$ are also assumed to be independent of the second level latent vectors $\boldsymbol{\xi}_g$ and $\boldsymbol{\delta}_g$. Thus, the among-groups covariance matrix $\mathbf{\Sigma}_b$ and the within-groups covariance matrix $\mathbf{\Sigma}_{wg}$ are respectively given by $\mathbf{\Sigma}_b = \mathbf{\Lambda}_b \mathbf{\Gamma}_b \mathbf{\Lambda}_b^{'} + \mathbf{\Psi}_b$, and $\mathbf{\Sigma}_{wg} = \mathbf{\Lambda}_{wg} \mathbf{\Gamma}_{wg} \mathbf{\Lambda}_{wg}^{'} + \mathbf{\Psi}_{wg}$, while the covariance matrix of $\mathbf{u}_{gi}$ is $\mathbf{\Sigma}_g = \mathbf{\Sigma}_b + \mathbf{\Sigma}_{wg}$. The matrices of unknown parameters in this model are $\mathbf{\Lambda}_b, \mathbf{\Gamma}_b, \mathbf{\Psi}_b, \mathbf{\Lambda}_{wg}, \mathbf{\Gamma}_{wg}$ and $\mathbf{\Psi}_{wg},\ g = 1, \cdots, G$.

To study the model with mixed polytomous and continuous variables, we suppose without loss of generality that the observations are $(\boldsymbol{x}_{gi}{}'; \boldsymbol{z}_{gi}{}')'$, where $\boldsymbol{x}_{gi} = (x_{1gi}, \cdots, x_{rgi})'$ is a continuous vector, $\boldsymbol{z}_{gi} = (z_{1gi}, \cdots, z_{sgi})'$ is a discrete polytomous vector, and $r + s = p$. A probit model is used to model $\boldsymbol{z} = (z_1, \cdots, z_s)'$ with its underlying vector $\boldsymbol{y} = (y_1, \cdots, y_s)'$ as follows:

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix} \quad \text{if} \quad \begin{array}{c} \alpha_{1,z_1} < y_1 \leq \alpha_{1,z_1+1} \\ \vdots \\ \alpha_{s,z_s} < y_s \leq \alpha_{s,z_s+1} \end{array}, \tag{3}$$

where z_k is an integral value in $\{0, 1, \cdots, b_k\}$, $\alpha_{k,0} = -\infty$ and $\alpha_{k,b_k+1} = \infty$. For the k-th variable, there are $b_k + 1$ categories which are defined by thresholds α_{kj}. Let $\boldsymbol{u}_{gi} = (\boldsymbol{x}_{gi}'; \boldsymbol{y}_{gi}')'$. We assume that $\boldsymbol{u}_{gi}$ satisfies the model defined by (1) and (2). Various computational methods have been proposed to deal with single-level latent variable models with this kind of data. For example, Muthen (1984) and Lee, Poon and Bentler (1995) proposed multistage estimation procedures based on classical gradient methods; Shi and Lee (1998) developed a Bayesian approach with the Gibbs sampler (Geman and Geman, 1984) to estimate thresholds, parameters in the covariance matrix as well as scores of latent variables; and Shi and Lee (2000) established a MCEM (Wei and Tanner, 1990) algorithm for obtaining ML estimates. However, these methods cannot be applied to two-level models.

It has been pointed out by Shi and Lee (1998, 2000) that the variance and thresholds associated with each polytomous variable are not identifiable, and some identification conditions on these parameters are required. In this paper, we will fix $\alpha_{k,1}$ and $\alpha_{k,b_k}, k = 1, \cdots, s$, at some preassigned values. This is basically equivalent to using the range $\alpha_{k,b_k} - \alpha_{k,1}$ as a measure for dispersion of the polytomous variable z_k. So, thresholds are used to control dispersion of the polytomous variable. As pointed out by Shi and Lee (1998, 2000), the choice of preassigned values for fixed thresholds only changes the scale of the covariance matrices, but not the correlations of latent variables. Moreover, to identify $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_{wg}$, we follow the common practice in factor analysis to fix some appropriate elements in $\boldsymbol{\Lambda}_b$ and $\boldsymbol{\Lambda}_{wg}$ at preassigned values. Let $\boldsymbol{\theta}$ be the vector that contains all unknown structural parameters in $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_{wg}$, and let $\boldsymbol{\alpha}$ be the vector that contains all the unknown thresholds. Statistical analysis will be based on the identified set of parameters $\{\boldsymbol{\theta}, \boldsymbol{\alpha}\}$.

3. ML Estimation via the EM Algorithm

Let $\boldsymbol{X} = \{\boldsymbol{x}_{gi} : g = 1, \cdots, G; i = 1, \cdots, N_g\}$ and $\boldsymbol{Z} = \{\boldsymbol{z}_{gi} : g = 1, \cdots, G; i = 1, \cdots, N_g\}$ be the observed data matrices. Owing to the nature of polytomous variables, the observed-data log-likelihood function involves intractable multiple integrals; hence direct maximization of this function is very difficult. Our basic idea is to solve this computational difficulty by augmenting the observed data $(\boldsymbol{X}, \boldsymbol{Z})$ with hypothetical missing data $\boldsymbol{Y} = \{\boldsymbol{y}_{gi} : g = 1, \cdots, G; i = 1, \cdots, N_g\}$ and $\boldsymbol{F} = \{(\boldsymbol{v}_g, \boldsymbol{\xi}_g, \boldsymbol{\xi}_{gi}) : g = 1, \cdots, G; i = 1, \cdots, N_g\}$ in the estimation process. The ML estimate of $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ will be obtained via the EM algorithm (Dempster, Laird and Rubin, 1977). We note that when the complete data $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{F})$ are available, the relevant statistical analysis will be comparatively less difficult. It can be shown that the complete-data log-likelihood function is proportional to $L_1 + L_2 + L_3$, where

$$\begin{aligned}
L_1 &= -\frac{1}{2}\sum_{g=1}^{G}\{\log|\boldsymbol{\Psi}_b| + \log|\boldsymbol{\Gamma}_b| + (\boldsymbol{v}_g - \boldsymbol{\Lambda}_b\boldsymbol{\xi}_g)'\boldsymbol{\Psi}_b^{-1}(\boldsymbol{v}_g - \boldsymbol{\Lambda}_b\boldsymbol{\xi}_g) + \boldsymbol{\xi}_g{}'\boldsymbol{\Gamma}_b^{-1}\boldsymbol{\xi}_g\}, \\
L_2 &= -\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\{\log|\boldsymbol{\Psi}_{wg}| + \log|\boldsymbol{\Gamma}_{wg}| + \qquad\qquad (4) \\
&\quad +(\boldsymbol{u}_{gi} - \boldsymbol{v}_g - \boldsymbol{\Lambda}_{wg}\boldsymbol{\xi}_{gi})'\boldsymbol{\Psi}_{wg}^{-1}(\boldsymbol{u}_{gi} - \boldsymbol{v}_g - \boldsymbol{\Lambda}_{wg}\boldsymbol{\xi}_{gi}) + \boldsymbol{\xi}_{gi}{}'\boldsymbol{\Gamma}_{wg}^{-1}\boldsymbol{\xi}_{gi}\}, \\
L_3 &= \sum_{g=1}^{G}\sum_{i=1}^{N_g}\sum_{k=1}^{s}\{\log\ I(y_{kgi} \in (\alpha_{k,z_{kgi}}, \alpha_{k,z_{kgi}+1}])\},
\end{aligned}$$

$I(y \in (a, b])$ is an indicator function which takes the value of 1 if $a < y \leq b$ and zero otherwise. The observed-data log-likelihood function is much more complicated than the complete-data log-likelihood function which involves no integral. The standard EM algorithm (Dempster et al. 1977) consists of the following steps: (i) E-step: Evaluate the conditional expectation of the complete-data log-likelihood function given $(\boldsymbol{X}, \boldsymbol{Z})$ and current values of $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$; and (ii) M-step: Update $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ by maximizing the conditional expectation obtained in the E-step. Implementation of these steps is discussed as follows.

3.1. Implementation of the EM Algorithm

The E-step involves the evaluation of conditional expectations of various complete-data sufficient statistics. Since the related distributions are very complicated, direct evaluation of these conditional expectations is very tedious and difficult. To solve the problem, we use the idea of the MCEM algorithm (Wei and Tanner, 1990) to approximate these expectations via sample means of a sequence of observations generated from the conditional distribution $[\boldsymbol{Y}, \boldsymbol{F}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}]$; here we use $[\cdot|\cdot]$ to denote a conditional distribution. The following Gibbs sampler (Geman and Geman, 1984) is used: At the t-th iteration with current values $\boldsymbol{Y}^{(t)}$ and $\boldsymbol{F}^{(t)}$, (i) generate $\boldsymbol{F}^{(t+1)}$ from $[\boldsymbol{F}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{Y}^{(t)}]$, and (ii) generate $\boldsymbol{Y}^{(t+1)}$ from $[\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{F}^{(t+1)}]$. Derivation of these conditional distributions is given in the Appendix. Geman and Geman (1984) showed that $(\boldsymbol{Y}^{(t)}, \boldsymbol{F}^{(t)})$ can be regarded as an observation from the conditional distribution $[\boldsymbol{Y}, \boldsymbol{F}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}]$ after a large number of iterations. Thus, after a sufficiently large number of burn-in iterations, a sequence of observations $\{(\boldsymbol{Y}^{(t)}, \boldsymbol{F}^{(t)}) : t = 1, \cdots, T\}$ is collected to evaluate conditional expectations of the required sufficient statistics.

The M-step updates unknown parameters by maximizing the conditional expectations obtained in the E-step. It can be seen from (4) that $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ are separable in the complete-data log-likelihood. So, threshold parameters can be updated by maximizing the conditional expectation of L_3. For all k, g and i, we are interested in the thresholds such that

$$I(y_{kgi} \in (\alpha_{k,z_{kgi}}, \alpha_{k,z_{kgi}+1}]) = 1. \tag{5}$$

For the k-th component of $\boldsymbol{y}_{gi}$, let $\{y^*_{k,n} : n = 1, \cdots, N\}$ be the order statistics of $\{y_{kgi} : g = 1, \cdots, G; i = 1, \cdots, N_g\}$, where $N = N_1 + \cdots + N_G$. Moreover, let m_{kh} be the total number of $\{z_{kgi} < h,\ g = 1, \cdots, G, i = 1, \cdots, N_g\}$ for $h = 1, \cdots, b_k$. Then any $\alpha_{k,h}$ that takes a value in the interval $[y^*_{k,m_{kh}}, y^*_{k,m_{kh}+1})$ satisfies (5). In this paper, we use the mid-point: $\tilde{\alpha}_{k,h} = (y^*_{k,m_{kh}+1} + y^*_{k,m_{kh}})/2$. To complete the M-step, we calculate $\tilde{\alpha}^{(t)}_{k,h}$ for every $\boldsymbol{Y}^{(t)}$. Then, using $\{\boldsymbol{Y}^{(t)}, t = 1, \cdots, T\}$ obtained in the E-step, $\alpha_{k,h}$ is updated by $(\tilde{\alpha}^{(1)}_{k,h} + \cdots + \tilde{\alpha}^{(T)}_{k,h})/T$.

To update parameters $(\boldsymbol{\Lambda}_b, \boldsymbol{\Psi}_b, \boldsymbol{\Gamma}_b)$ in the M-step, we first obtain ML estimates of parameters by maximizing L_1, and then replace the conditional expectations of the sufficient statistics involved in the ML estimates by appropriate sample means of the simulated observations obtained in the E-step. Since $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_G$ are identically and independently distributed (i.i.d.) as $N(\mathbf{0}, \boldsymbol{\Gamma}_b)$, the ML estimate of $\boldsymbol{\Gamma}_b$ is $G^{-1}(\boldsymbol{\xi}_1\boldsymbol{\xi}_1' + \cdots + \boldsymbol{\xi}_G\boldsymbol{\xi}_G')$. To update $\boldsymbol{\Gamma}_b$, the conditional expectation of each $\boldsymbol{\xi}_g\boldsymbol{\xi}_g^{'}$ is replaced by $(\boldsymbol{\xi}_g^{(1)}\boldsymbol{\xi}_g^{(1)'} + \cdots, \boldsymbol{\xi}_g^{(T)}\boldsymbol{\xi}_g^{(T)'})/T$, where $\boldsymbol{\xi}_g^{(t)}$ is the simulated observation obtained in the E-step.

For each $g = 1, \cdots, G$, $\boldsymbol{v}_g$, given $\boldsymbol{\xi}_g$, is independently distributed as $N(\boldsymbol{\Lambda}_b\boldsymbol{\xi}_g, \boldsymbol{\Psi}_b)$. Moreover, since $\boldsymbol{\Psi}_b$ is diagonal, for $j = 1, \cdots, p, \boldsymbol{v}_{jg}$ are independent given $\boldsymbol{\xi}_g$. Hence, ML estimates of parameters involved in every row of $\boldsymbol{\Lambda}_b$ and every diagonal element of $\boldsymbol{\Psi}_b$ can be obtained via the regression model: $v_{jg} = \boldsymbol{\Lambda}_{bj}^{'}\boldsymbol{\xi}_g + \delta_{jg}, \quad g = 1, \cdots, G$, where $\boldsymbol{\Lambda}_{bj}^{'}$ is the j-th row of $\boldsymbol{\Lambda}_b$, δ_{jg} is independently distributed as $N(0, \psi_{bjj})$ and ψ_{bjj} is the (j, j)-th diagonal element of $\boldsymbol{\Psi}_b$. It can be shown via theory in regression that ML estimates of $\boldsymbol{\Lambda}_{bj}$ and $\boldsymbol{\Psi}_{bjj}$ are respectively given by $(\boldsymbol{AA}^{'})^{-1}\boldsymbol{AV}_j$ and $G^{-1}[\boldsymbol{V}_j^{'}\boldsymbol{V}_j - (\boldsymbol{AV}_j)^{'}(\boldsymbol{AA}^{'})^{-1}\boldsymbol{AV}_j]$, where $\boldsymbol{A} = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_G)$ and $\boldsymbol{V}_j^{\prime} = (v_{j1}, \cdots, v_{jG})$. To update the corresponding parameters in the M-step, conditional expectations of the sufficient statistics $\{\boldsymbol{AA}^{'}, \boldsymbol{AV}_j, \boldsymbol{V}_j^{'}\boldsymbol{V}_j\}$ are replaced by corresponding sample means of the simulated observations in the E-step.

We have the following facts about the unknown parameter matrices $\boldsymbol{\Lambda}_{wg}$ and $\boldsymbol{\Gamma}_{wg}$. For each $g = 1, \cdots, G$, (i) $\{\boldsymbol{\xi}_{gi}; i = 1, \cdots, N_g\}$ are i.i.d. as $N(\boldsymbol{0}, \boldsymbol{\Gamma}_{wg})$; and (ii) $\boldsymbol{v}_{gi} = \boldsymbol{u}_{gi} - \boldsymbol{v}_g$, given $\boldsymbol{\xi}_{gi}$, are independently distributed as $N(\boldsymbol{\Lambda}_{wg}\boldsymbol{\xi}_{gi}, \boldsymbol{\Psi}_{wg})$ for $i = 1, \cdots, N_g$, and $\boldsymbol{\Psi}_{wg}$ is diagonal. Thus, for every g and i, structures of the models relating to $\boldsymbol{\xi}_{gi}$ and $\boldsymbol{v}_{gi}$ are essentially the same as before. By exactly the same derivations, all unknown parameters in the within-groups covariance structure can be updated in the M-step.

Finally, the approach of Shi and Lee (2000) can be used to handle fixed parameters in $\boldsymbol{\Lambda}_b$ and $\boldsymbol{\Lambda}_{wg}$ for identifying the model. To save space, details are not presented.

3.2. Monitoring Convergence via Bridge Sampling

Owing to the complexity of the proposed model, it is non-trivial to determine convergence of the EM algorithm. Because of the variability introduced at the E-step, the log-likelihood function can still 'zigzag' along its iterates. Moreover, since the observed-data log-likelihood function cannot be evaluated analytically, we have to approximate it via some numerical methods. Hence, if we cannot compute the observed log-likelihood values with good accuracy, then we cannot judge whether the fluctuation is due to the numerical errors in computing the log-likelihood values, to the implementation errors, or to the non-convergence of the EM algorithm. To solve the above problem, the bridge sampling (Meng and Wong, 1996) is used to monitor convergence of the MCEM algorithm.

The main idea of bridge sampling is to use the ratio of two consecutive likelihood values to monitor changes of the function. As argued by Meng and Schilling (1996), the ratio of the likelihood values at consecutive iterations a and $a+1$ can be approximated by

$$R(\boldsymbol{\theta}^{(a+1)}, \boldsymbol{\alpha}^{(a+1)}) = \frac{\sum_{t=1}^{T} \left\{ \frac{p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}^{a,(t)}, \boldsymbol{F}^{a,(t)} | \boldsymbol{\theta}^{(a+1)}, \boldsymbol{\alpha}^{(a+1)})}{p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}^{a,(t)}, \boldsymbol{F}^{a,(t)} | \boldsymbol{\theta}^{(a)}, \boldsymbol{\alpha}^{(a)})} \right\}^{1/2}}{\sum_{t=1}^{T} \left\{ \frac{p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}^{a+1,(t)}, \boldsymbol{F}^{a+1,(t)} | \boldsymbol{\theta}^{(a)}, \boldsymbol{\alpha}^{(a)})}{p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}^{a+1,(t)}, \boldsymbol{F}^{a+1,(t)} | \boldsymbol{\theta}^{(a+1)}, \boldsymbol{\alpha}^{(a+1)})} \right\}^{1/2}}$$

where $(\boldsymbol{Y}^{a,(t)}, \boldsymbol{F}^{a,(t)})$ are random observations generated at the a-th iteration. To monitor the convergence of the algorithm, $R(\boldsymbol{\theta}^{(a+1)}, \boldsymbol{\alpha}^{(a+1)})$ will be plotted against $a+1$, the iteration index. Convergence is claimed if plots show a curve converging from above to one.

4. An Illustrative Example.

The proposed methodology is used to assess correlations between some latent quantities of the CSWs in the study of Morisky et al. (1998) as described in the Introduction. Nine manifest variables, of which the first three are continuous and the last six are polytomous with a five-point scale, were selected. The questions corresponding to the selected variables are given in Table 1. For brevity, we deleted those observations with missing entries in the analysis, and the remaining sample size is 755. The numbers of individuals in establishments varied from 1 to 58, so we have an unbalanced data set. The sample means and standard deviations of the continuous variables are $\{2.442, 1.180, .465\}$ and $\{5.299, 2.208, 1.590\}$, respectively. The cell freqencies of each polytomous variable are ranged from 12 to 348. To unify the scales of variables, the raw continuous data were standardized.

The resulting data set was analyzed using a two-level factor analysis model. Based on the meanings of the questions corresponding to the selected manifest variables, the among-groups and within-groups models were analyzed by factor analysis models with three factors ($r_1 = 3, r_2 = 3$), using the first three, the second three and the last three manifest variables as indicators for the first, second and third factors, respectively. Although other structures for the loading matrix can be considered, we choose the structure that gives non-overlapping latent factors for clear interpretation. These factors are allowed to be correlated. Hence, the following specifications with an invariant within-groups covariance structure are used: $\mathbf{\Lambda}_{wg} = \mathbf{\Lambda}_w, \mathbf{\Gamma}_{wg} = \mathbf{\Gamma}_w, \mathbf{\Psi}_{wg} = \mathbf{\Psi}_w$,

$$\mathbf{\Lambda}_w{}' = \begin{bmatrix} \lambda_{w11} & \lambda_{w21} & \lambda_{w31} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & \lambda_{w42} & \lambda_{w52} & \lambda_{w62} & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & \lambda_{w73} & \lambda_{w83} & \lambda_{w93} \end{bmatrix}, \ \mathbf{\Gamma}_w = \begin{bmatrix} 1^* & \gamma_{w12} & \gamma_{w13} \\ \gamma_{w12} & 1^* & \gamma_{w23} \\ \gamma_{w13} & \gamma_{w23} & 1^* \end{bmatrix},$$

and $\mathbf{\Psi}_w = \text{diag}(\psi_{w1}, \cdots, \psi_{w9})$, where parameters with an asterisk were fixed at values of 1's and 0's. The among-groups covariance structure is exactly the same. To identify the model with respect to polytomous variables, α_{k1} and $\alpha_{k4}, k = 1, \cdots, 6$, were fixed at $\alpha_{kj} = \Phi^{-1}(m_k)$, where m_k are the observed cumulative marginal proportions of the categories with $z_k < j$. There are a total of 54 unknown parameters in this two-level model.

The proposed MCEM algorithm was used to obtain ML estimates of parameters. In the E-step, the length of burn-in for the Gibbs sampler is 20, and a total of $T = 80$ observations were simulated to approximate conditional expectations. The current parameters values obtained at the j-th iteration were used as the starting values in the Gibbs simulation at the $(j + 1)$-th iteration. To give some idea about the convergence, the sequence of likelihood ratios in the bridge sampling is displayed in Figure 1. Estimates of the thresholds and the structural parameters in the covariance structures at the iterations are presented in Figure 2. Apparently, the MCEM algorithm stabilized after about 60 iterations. To be conservative, parameter values at the 100th iteration were taken as ML estimates. The standard errors (SEs) of the ML estimates are computed via the Louis (1982) formula. Since L_3 is not differentiable with respect to the thresholds; for simpliticity, we just computed the SEs of the structural parameters. The results are presented in Table 2. The significance of the establishments influence is reflected by relatively large estimates of some parameters in the among-groups model: This result indicates that the problem on violation of the independence assumption cannot be solved without taking the among-groups model on v_g into account. It can be seen from the ML estimates of the parameters in the factor loading matrices that there are three non-overlapping factors in the within-groups covariance structure. The first factor can be roughly interpreted as 'aggressiveness' of workers, the second as 'worry about AIDS', while the last one as 'attitude to the risk of getting AIDS'. From the correlations of the first level factors that relating to the CSWs, 'aggressiveness' is negatively correlated with 'worry of getting AIDS' and 'attitude to the risk', while the last two factors are positively correlated. Clearly, these results carry the following meaningful interpretations: (i) the aggressive CSWs are less worry about AIDS and do not have serious attitude of risk in getting AIDS, and (ii) those CSWs who are more worry about AIDS are more likely to have serious attitude of risk in getting AIDS. It is interesting to note that, at the second level that relating to the policy of establishments, the signs of the correlations among the factors are completely different from those at the CSWs level. This may indicate the existence of difference on these latent quantities between CSWs and establishments.

The data have been reanalyzed using a single-level model that does not involve $\mathbf{\Lambda}_b$ and $\mathbf{\Psi}_b$. We observed substantial changes for some estimates, for example $\widehat{\lambda}_{w21}, \widehat{\lambda}_{w52}, \widehat{\lambda}_{w62}, \widehat{\psi}_{w2}, \widehat{\psi}_{w5}$ and $\widehat{\psi}_{w6}$ were changed to 0.667, 0.199, 0.522, 0.555, 0.974 and 0.649; while the factor correlations estimates $(\widehat{\gamma}_{w12}, \widehat{\gamma}_{w13}, \widehat{\gamma}_{w23})$ were changed to -0.274, -0.158 and 0.278. It seems that ignoring establishments influences can lead to different results.

To address the sensitivity to the choices of starting values of parameters, the number of burn-in iterations and T, we have reanalyzed the data with different starting values, 100 burn-in iterations and/or $T = 200$. We obtained similar results as before in all different settings. Moreover, to study the relation of fixed thresholds with estimates of parameters, the data were reanalyzed with fixed thresholds that are equal to $2^{1/2}$ times the previous values. As expected, for those parameters corresponding to polytomous variables, the new estimates in $\mathbf{\Lambda}_b$ and $\mathbf{\Lambda}_w$ are approximately increased by a factor of $2^{1/2}$, in $\mathbf{\Psi}_b$ and $\mathbf{\Psi}_w$ are increased by 2.0, while in $\mathbf{\Gamma}_w$ remain unchanged. All these details are not presented due to space limitation, but they can be obtained upon request.

The main computation of the proposed MCEM algorithm is on generating the random observations via the Gibbs sampler in the E-step. Since the conditional distributions required for implementing the Gibbs sampler are the standard univariate truncated normal and multivariate normal, the computing burden is not heavy. Using the setting with 20 burn-in iterations and $T = 80$, it takes about 3 minutes to complete 10 MCEM iterations in a hpc4500 workstation. Our program is written in FORTRAN VI language, it is available upon request.

5. Discussion.

Polytomous data and two-level data with hierarchical structure are very common in biomedical research. In this paper, a ML approach for analyzing latent variable models with this kind of data is developed. The proposed methodology can handle binary variables which involve two categories with only one threshold. Hence, it can be viewed as an extension of Meng and Schilling (1996) in which the item factor with binary variables is investigated. Also, it is a generalization of Shi and Lee (2000) from single-level models to two-level models. Like almost all existing methods in the field, our approach requires to assume that the latent variables and error measurements are normally distributed. The assumption relating to latent variables can be assessed roughly via samples generated at the E-step. The assumption on error measurements can be assessed via estimates of residuals. Clearly, extension of results in Browne (1987) on robustness to the normality assumption for the present model represents an important open problem for further research. We also assume that covariance matrices of error measurements are diagonal. This reasonable and standard assumption in factor analysis is crucial to the efficiency of the proposed algorithm. Without this assumption, y_{gki} and y_{gkj} are not conditionally independent and the corresponding joint conditional distribution is a complicated multivariate truncated normal distribution. The computational burden for simulating observations from this distribution is more heavy.

The proposed MCEM algorithm works well for our model; and yet its efficiency can be improved. In the analysis of the real example, we use a moderate T $(= 80)$ to approximate conditional expectations at the E-step. This choice of T is ad hoc and not greatly affects the speed of the algorithm since the number of iterations to achieve convergence is small. For large models with more complicated structures, the number of iterations for convergence may be large. In this situation, a more subtle scheme (see e.g. Booth and Hobert, 1999) in selecting T may improve the efficiency of the algorithm.

For ease of explanation, the present development was based on a two-level factor analysis model. Extensions to multilevel models with more general covariance structures such as EQS (Bentler, 1992) or LISREL (Jöreskog and Sörbom, 1996) are useful for handling more complex situations. Furthermore, incorporating fixed effects with multilevel models is not trivial. Finally, theoretical and computational developments for important issues on hypothesis testing and model comparison clearly warrant future research.

ACKNOWLEDGEMENTS

The work described in this paper was fully supported by a grant from the Research Grants Council of the HKSAR (Project No: CUHK 4088/99H). The authors are thankful to the referees, the associate editor and the editor for constructive comments, and to Drs D.E.

Morisky and J.A. Stein for providing the AIDS data in the example. The assistance of Michael Leung and Esther Tam is gratefully acknowledged.

REFERENCES

Aitkin, M. and Longford, N. (1986). Statistical modeling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society, Ser A* **149**, 1-43.

Bentler, P. M. (1992). *EQS: Structural Equation Program Manual.* Los Angels: BMDP Statistical Software.

Bollen, K. A. and Long, J. S. (Eds) (1993). *Testing Structural Equation Models.* Newbury Park, CA: Sage.

Booth, J. G. and Hobert, J. P. (1999). Maximum generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **6**, 265-285.

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika* **74**, 375-384.

Catalano, P. J. and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, **87** 651-658.

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

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.

Goldstein, H. I. (1986) Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* **73**, 43-56.

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language.* Scientific Software International: Hove and London.

Legler, J. M. and Ryan, L. M. (1997). Latent variable models for multiple birth outcomes. *Journal of the American Statistical Association* **92**, 13-20.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method* (2nd ed). New York: American Elsevier.

Lee, S. Y. (1990). Multilevel analysis of structural equation models. *Biometrika* **77**, 763-772.

Lee, S. Y. and Poon, W. Y. (1998). Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica* **8** 749-766.

Lee, S. Y., Poon, W. Y. and Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology* **48**, 339-358.

Louis, T. A. (1982). Finding the observed information matrix when using EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226-233.

Longford, N. T. and Muthen, B. O. (1992). Factor analysis for clustered observations. *Psychometrika* **57**, 581-597.

McDonald, R. P. and Goldstein, H. I. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology* **42**, 215-232.

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

Meng, X. L. and Wong, W. H. (1996). Simulation ratios of normalizing constants via simple identity, a theoretical exploration. *Statist. Sin.* **6**, 831-860.

Morisky, D. E., Tiglao, T. V., Sneed, C. D., Tempongko, S. B., Baltazar, J. C., Detels, R. and Stein, J. A. (1998). The effects of establishment practices, knowledge and attitudes on condom use among Filipina sex workers. *AIDS Care* **10**, 213-220.

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

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

Roberts, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing* **5**, 121-125.

Rubin, D. B. (1991) EM and beyond. *Psychometrika* **56**, 241-254.

Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 220-243.

Sammel, M. D. Ryan, L. M. and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society*, it Ser B **59**, 667-678.

Shi, J. Q. and Lee, S. Y. (1998). Bayesian sampling-based approach for factor analysis models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **51**, 233-252.

Shi, J. Q. and Lee, S. Y. (2000). Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society, Ser B* **62**, 77-87.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association* **85**, 699-704.

APPENDIX

The Conditional distributions are derived as follows. (i) $[\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{F},\boldsymbol{\alpha}]$: From (1), $\boldsymbol{u}_{gi}$ has the normal distribution $N(\boldsymbol{v}_g+\boldsymbol{\Lambda}_{wg}\boldsymbol{\xi}_{gi},\ \boldsymbol{\Psi}_{wg})$ and the $\boldsymbol{u}_{gi}$ are mutually independent for different g and i given $\boldsymbol{v}_g$, $\boldsymbol{\xi}_{gi}$, $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}_k$. Since $\boldsymbol{y}_{gi}$ is a subset of $\boldsymbol{u}_{gi}$, conditional distributions of $\boldsymbol{y}_{gi}$ are normally distributed and mutually independent for different g and i. Moreover, since $\boldsymbol{\Psi}_{wg}$ is a diagonal matrix, $\boldsymbol{y}_k=\{y_{kgi}, g=1,\cdots,G, i=1,\cdots,N_g\}$ are mutually and conditionally independent for different k, $k=1,\cdots s$, the conditional distribution of y_{kgi} is the univariate truncated normal distribution $N(v_{kg}+\boldsymbol{\Lambda}_{wkg}\boldsymbol{\xi}_{gi},\ \psi_{wkg})I(y_{kgi}\in(\alpha_{k,z_{kgi}},\alpha_{k,z_{kgi}+1}])$. A random variate from an univariate truncated normal can be generated via the algorithm proposed by Roberts (1995). (ii) $[\boldsymbol{F}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{Y}]$: Let $\boldsymbol{U}=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n)$, since

$$\begin{aligned} p(\boldsymbol{F}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{Y}) &= \prod_{g=1}^{G} p(\boldsymbol{v}_g,\boldsymbol{\xi}_g,\boldsymbol{\xi}_{gi},\ i=1,\cdots,N_g|\boldsymbol{U},\boldsymbol{\theta}) \\ &= \prod_{g=1}^{G} p(\boldsymbol{v}_g|\boldsymbol{U},\boldsymbol{\theta})\ p(\boldsymbol{\xi}_g,\boldsymbol{\xi}_{gi},\ i=1,\cdots,N_g|\boldsymbol{v}_g,\boldsymbol{U},\boldsymbol{\theta}), \end{aligned}$$

the conditional probability density function $p(\boldsymbol{v}_g|\boldsymbol{U},\boldsymbol{\theta})$ is proportional to

$$\begin{aligned} & p(\boldsymbol{v}_g)\prod_{i=1}^{N_g} p(\boldsymbol{u}_{gi}|\boldsymbol{v}_g,\boldsymbol{\theta}) \propto \exp\{-\frac{1}{2}\boldsymbol{v}_g{}'\boldsymbol{\Sigma}_b^{-1}\boldsymbol{v}_g\}\prod_{i=1}^{N_g}\exp\{-\frac{1}{2}(\boldsymbol{u}_{gi}-\boldsymbol{v}_g)'\boldsymbol{\Sigma}_{wg}^{-1}(\boldsymbol{u}_{gi}-\boldsymbol{v}_g)\} \\ \propto\ & \exp\{-\frac{1}{2}(\boldsymbol{v}_g-\boldsymbol{\mu}_g)'\boldsymbol{\Omega}_g^{-1}(\boldsymbol{v}_g-\boldsymbol{\mu}_g)\}, \end{aligned}$$

where $\boldsymbol{\mu}_g=N_g\boldsymbol{\Omega}_g\boldsymbol{\Sigma}_{wg}^{-1}\overline{\boldsymbol{u}}_g$ and $\boldsymbol{\Omega}_g=(\boldsymbol{\Sigma}_b^{-1}+N_g\boldsymbol{\Sigma}_{wg}^{-1})^{-1}$, with $\overline{\boldsymbol{u}}_g$ be the sample mean of $\boldsymbol{u}_{gi}$ for $i=1,\cdots,N_g$. So, $p(\boldsymbol{v}_g|\boldsymbol{U},\boldsymbol{\theta})$ is proportional to $N(\boldsymbol{\mu}_g,\boldsymbol{\Omega}_g)$. Moreover, $\boldsymbol{\xi}_g$ and $\boldsymbol{\xi}_{gi}$ are independent; and $\boldsymbol{\xi}_{gi}$ and $\boldsymbol{\xi}_{gj}$ are independent for different i and j. Hence, it suffices to obtain conditional distributions $[\boldsymbol{\xi}_g|\boldsymbol{v}_g,\boldsymbol{U},\boldsymbol{\theta}]$ and $[\boldsymbol{\xi}_{gi}|\boldsymbol{v}_g,\boldsymbol{U},\boldsymbol{\theta}]$. Since $p(\boldsymbol{\xi}_g|\boldsymbol{v}_g,\boldsymbol{U},\boldsymbol{\theta}) = p(\boldsymbol{\xi}_g|\boldsymbol{v}_g,\boldsymbol{\theta}) \propto p(\boldsymbol{\xi}_g|\boldsymbol{\theta})p(\boldsymbol{v}_g|\boldsymbol{\xi}_g,\boldsymbol{\theta})$, where $p(\boldsymbol{\xi}_g|\boldsymbol{\theta})$ and $p(\boldsymbol{v}_g|\boldsymbol{\xi}_g,\boldsymbol{\theta})$ are density functions of $N(\boldsymbol{0},\boldsymbol{\Gamma}_b)$ and $N(\boldsymbol{\Lambda}_b\boldsymbol{\xi}_g,\boldsymbol{\Psi}_b)$ respectively, the conditional distribution of $\boldsymbol{\xi}_g$ given $(\boldsymbol{v}_g,\boldsymbol{\theta})$ is a normal distribution $N(\boldsymbol{\Omega}_b\boldsymbol{\Lambda}_b{}'\boldsymbol{\Psi}_b^{-1}\boldsymbol{v}_g,\ \boldsymbol{\Omega}_b)$, where $\boldsymbol{\Omega}_b=(\boldsymbol{\Gamma}_b^{-1}+\boldsymbol{\Lambda}_b{}'\boldsymbol{\Psi}_b^{-1}\boldsymbol{\Lambda}_b)^{-1}$. Similarly, the conditional distribution of $\boldsymbol{\xi}_{gi}$ given $(\boldsymbol{v}_g,\boldsymbol{u}_{gi},\boldsymbol{\theta})$ is $N(\boldsymbol{\Omega}_{wg}\boldsymbol{\Lambda}_{wg}{}'\boldsymbol{\Psi}_{wg}^{-1}(\boldsymbol{u}_{gi}-\boldsymbol{v}_g),\ \boldsymbol{\Omega}_{wg})$, where $\boldsymbol{\Omega}_{wg}=(\boldsymbol{\Gamma}_{wg}^{-1}+\boldsymbol{\Lambda}_{wg}{}'\boldsymbol{\Psi}_{wg}^{-1}\boldsymbol{\Lambda}_{wg})^{-1}$.

Table 1

Questions Associated with the Manifest Variables.

$u_g(1)$:	How often did you perform vaginal sex in the last 7 days?
$u_g(2)$:	How often did you perform manual sex in the last 7 days?
$u_g(3)$:	How often did you perform oral sex in the last 7 days?
$u_g(4)$:	How much of a threat do you think AIDS is to the health of people?
$u_g(5)$:	What are the chances that you yourself might get AIDS?
$u_g(6)$:	How worried are you about getting AIDS?
	How great is the risk of getting AIDS or the AIDS virus from sexual intercourse with someone:
$u_g(7)$:	Who has the AIDS virus using a condom?
$u_g(8)$:	Whom you don't know very well without using a condom?
$u_g(9)$:	Who injects drugs?

Table 2

ML Estimates and Standrd Errors (in brackets) of the Parameters

Thresholds		Among-gp Model		Within-gp Model	
Parameter	ML Est.	Parameters	ML Est.	Parameter	ML Est.
α_{12}	-1.095	λ_{b11}	.194 (.054)	λ_{w11}	.224 (.055)
α_{13}	-.607	λ_{b21}	.322 (.069)	λ_{w21}	.513 (.059)
α_{22}	-.112	λ_{b31}	.138 (.038)	λ_{w31}	.394 (.055)
α_{23}	.384	λ_{b42}	-.099 (.058)	λ_{w42}	.568 (.075)
α_{32}	-.900	λ_{b52}	.649 (.071)	λ_{w52}	.237 (.049)
α_{33}	-.388	λ_{b62}	.142 (.063)	λ_{w62}	.403 (.058)
α_{42}	-.473	λ_{b73}	.316 (.035)	λ_{w73}	.244 (.060)
α_{43}	.262	λ_{b83}	.099 (.072)	λ_{w83}	.592 (.075)
α_{52}	-1.647	λ_{b93}	.195 (.067)	λ_{w93}	.507 (.062)
α_{53}	-.718				
α_{62}	-1.090	ψ_{b1}	.032 (.018)	ψ_{w1}	.883 (.050)
α_{63}	-.098	ψ_{b2}	.097 (.035)	ψ_{w2}	.500 (.057)
		ψ_{b3}	.013 (.008)	ψ_{w3}	.808 (.054)
		ψ_{b4}	.064 (.029)	ψ_{w4}	.564 (.081)
		ψ_{b5}	.060 (.038)	ψ_{w5}	.569 (.049)
		ψ_{b6}	.127 (.037)	ψ_{w6}	.616 (.057)
		ψ_{b7}	.012 (.004)	ψ_{w7}	.881 (.075)
		ψ_{b8}	.093 (.035)	ψ_{w8}	.588 (.082)
		ψ_{b9}	.102 (.041)	ψ_{w9}	.676 (.071)
		γ_{b12}	.594 (.150)	γ_{w12}	-.218 (.082)
		γ_{b13}	.025 (.205)	γ_{w13}	-.221 (.087)
		γ_{b23}	-.680 (.133)	γ_{w23}	.310 (.081)

Figure captions:

Figure 1. AIDS data: values of likelihood ratios in the bridge sampling.

Figure 2. AIDS data: values of estimates of (i) $\boldsymbol{\alpha}$ and (ii) $\boldsymbol{\theta}$ at iterations of the EM algorithm.

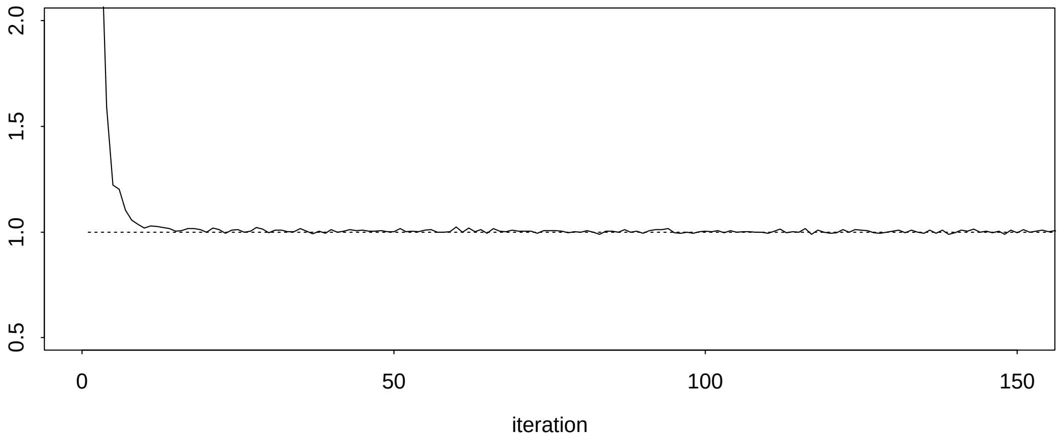

Figure 1.

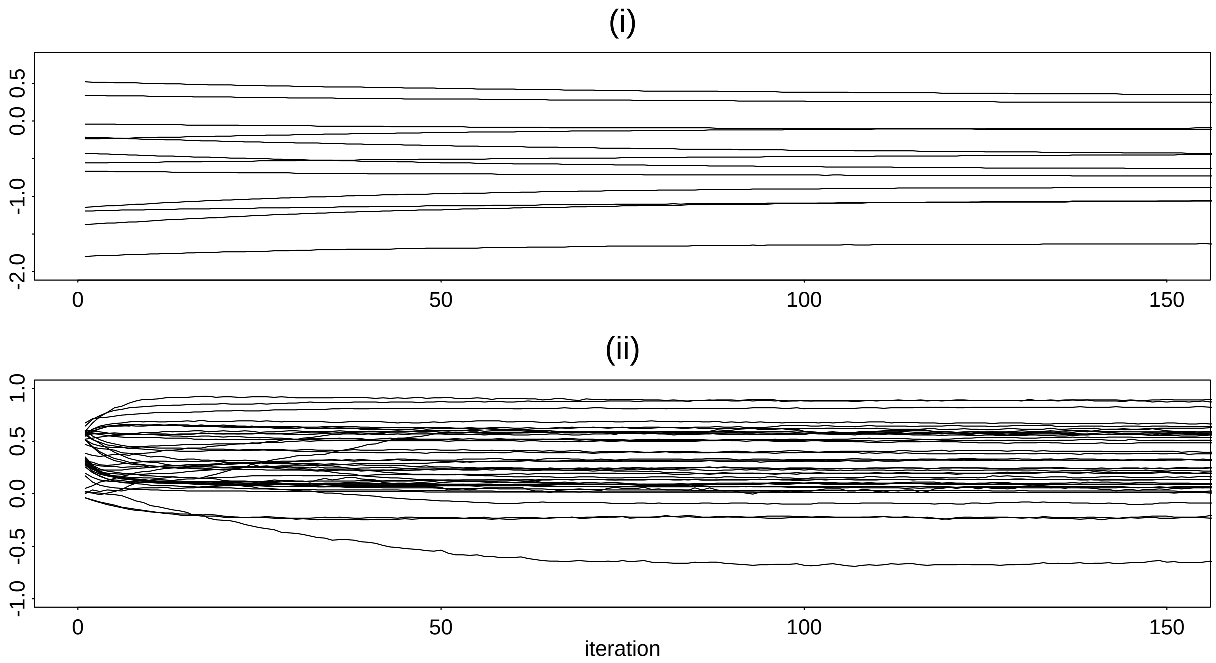

Figure 2.