Statistica Sinica **15**(2005), 645-664

A MULTIVARIATE PROBIT LATENT VARIABLE MODEL FOR ANALYZING DICHOTOMOUS RESPONSES

Xin-Yuan Song and Sik-Yum Lee

The Chinese University of Hong Kong

Abstract: We propose a multivariate probit model that is defined by a confirmatory factor analysis model with covariates for analyzing dichotomous data in medical research. Our proposal is a generalization of several useful multivariate probit models, and provides a flexible framework for practical applications. We implement a Monte Carlo EM algorithm for maximum likelihood estimation of the model, and develop a path sampling procedure to compute the observed-data log-likelihood for evaluating the Bayesian Information Criterion for model comparison. Our methodology is illustrated by analyzing two data sets in medical research.

Key words and phrases: Maximum likelihood, Monte Carlo EM algorithm, observed -data likelihood, path sampling.

1. Introduction

Biological, medical, and social studies often yield binary or dichotomous data due to the lack of adequate and direct continuous measurements. Indeed, correlated dichotomous data arise in many settings, ranging from measurements of random cross-section subjects to repeated measurements in longitudinal studies. The multivariate probit (MP) model is a popular method in biostatistics for analyzing this kind of data. This model is described in terms of a correlated multivariate normal distribution of the underlying latent variables that are manifested as discrete variables through a threshold specification, and hence allows the flexible modeling of the correlation structure and easy interpretation of the parameters.

Since the pioneer work of Ashford and Sowden (1970), numerous attempts have been made to solve the computational difficulty of evaluating the multivariate normal orthant probabilities involved. Chib and Greenberg (1998) developed a Bayesian approach and a maximum likelihood (ML) approach for a MP model with a general residual covariance structure, and applied the method to various data sets, including the canonical four-year dataset from the Six Cities Study of health effects. Their Bayesian and ML approaches require the simulation of observations from a multivariate truncated normal distribution involving an arbitrary covariance matrix. Although observations from a multivariate truncated

normal distributions can be sampled from a sequence of univariate truncated normal distributions, the computational effort is rather heavy for high dimensional problems. Other methods that use much less restrictive covariance structures have been proposed to reduce the computational burden of evaluating the probabilities: see, for example, Kolakowski and Bock (1981) and Ochi and Prentice (1984). In particular, Bock and Aitkin (1981) used the exploratory factor analysis (EFA) model for the covariance structure and applied an EM algorithm (Dempster, Laird and Rubin (1977)) to obtain the ML solution. Recently, Bock and Gibbons (1996), Gibbons and Lavigne (1998) and Gibbons and Wilcox-Gök (1998) extended the Bock and Aitkin (1981) model to an exploratory factor analysis model with fixed covariates, and provided novel applications to data on the early childhood development of psychiatric disorders (Gould, Wunsch-Hizig and Dohrenwend (1981) and Vikan (1985)), and health service utilization and insurance (Link, Long and Settle (1980) and Wolfe and Goddeeris (1991)). The approaches that were used by Bock and Aitkin (1981), Bock and Gibbons (1996), Gibbons and Lavigne (1998) and Gibbons and Wilcox-Gök (1998) applied Gauss-Hermite quadrature to approximate the integrals in relation to the marginal probabilities. Because item response models can be viewed as the factor analysis model for dichotomous variables (Takane and de Leeuw (1987)), the models that were developed by Bock and Gibbons (1996), Gibbons and Lavigne (1998) and Gibbons and Wilcox-Gök (1998) can be regarded as their generalizations. Note that the item response model has been found to be very useful for analyzing quality of life data (Donglas (1999) and Wang, Douglas and Anderson (2002)).

Meng and Schilling (1996) reanalyzed the EFA model of Bock and Aitkin (1981) and pointed out some deficiencies in using Gauss-Hermite quadrature to approximate the integrals that are associated with the marginal probabilities (Meng and Schilling (1996, p.1256)). They then recommended a better approach that is based on the Monte Carlo EM algorithm (Wei and Tanner (1990)). In contrast to Chib and Greenberg (1998), the main focus of the above work for analyzing the MP model with an EFA structure has been on estimation, and there have been few model comparison results. The main reason is probably the computational difficulty of evaluating the observed-data log-likelihood, which involves complicated integrals.

In this article, we propose a MP model that is defined with a confirmatory factor analysis model and covariates, and show that it is more general than the Gibbons and Wilcox-Gök (1998) model. Based on the recommendation of Meng and Schilling (1996), we implement a Monte Carlo EM (MCEM) algorithm for obtaining the ML estimates of the unknown parameters. Computationally, there are two key advantages of the proposed MCEM algorithm: its E-step can

be completed by observations that are directly simulated from a comparatively simple univariate truncated normal distribution; its M-step can be completed by a closed form solution that is obtained on the basis of conditional maximization. Consequently, the algorithm is rather efficient. Moreover, it can produce factor score estimates as by-products. We also show that the ML solution for the MP model that has an arbitrary covariance matrix of the residuals (see Chib and Greenberg (1998)) can be obtained with the proposed model. The important issue of model comparison is also addressed. Specifically, we utilize path sampling (Gelman and Meng (1998)) to develop a procedure for computing the observed-data log-likelihood, so that the Bayesian Information Criterion (BIC) can be evaluated for model comparison (Kass and Raftery (1995)).

In Section 2, we describe the proposed model that is based on the confirmatory factor analysis model with covariates. We consider ML estimation of the model and describe the MCEM algorithm in Section 3. Section 4 develops the path sampling procedure for computing the observed-data log-likelihood and BIC. In Section 5 we provide illustrative examples that are based on real medical datasets. Section 6 contains a discussion, and technical details are given in the Appendices.

2. The MP Model

We assume that each subject has a covariate vector that can be any mixture of discrete and continuous variables. Each subject produces J distinct quantal responses or is classified with respect to J dichotomous categories. Specifically, let $\mathbf{u}_i = (u_{i1}, \ldots, u_{iJ})'$ denote the collection of observed dichotomous 0/1 responses in J variables on the ith subject, $i = 1, \ldots, n$, $\mathbf{x}_{ij}$ be a $k_j \times 1$ vector of covariates, $k = k_1 + \cdots + k_J$, and

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i1} & 0 & \cdots & 0 \\ 0 & \mathbf{x}'_{i2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{x}'_{iJ} \end{bmatrix}$$

be a $J \times k$ matrix. The following MP model was formulated by Chib and Greenberg (1998). Let $\mathbf{z}_i = (z_{i1}, \ldots, z_{iJ})'$ denote a J-variate normal vector of "response strengths" such that

$$\mathbf{z}_i = \mathbf{X}_i\mathbf{B} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{B}' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_J)$, $\mathbf{b}_j$ is a $k_j \times 1$ unknown parameter vector, $\boldsymbol{\epsilon}_i$ is a $J \times 1$ vector of residuals that is distributed as $N[\mathbf{0}, \boldsymbol{\Sigma}]$, and

$$u_{ij} = \begin{cases} 1 & z_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, \ldots, J. \tag{2}$$

In this model, the exact measurement of "response strengths" $\mathbf{z}_i$ is not observed, and its information is given by an observed dichotomous vector $\mathbf{u}_i = (u_{i1}, \ldots, u_{iJ})'$ with u_{ij} given by (2). Here, $\mathbf{B}$ is $k \times 1$ vector of regression coefficients of $\mathbf{z}_i$ on $\mathbf{X}_i$. This MP model has seen quite wide application (see Chib and Greenberg (1998) and the references therein). Inspired by Gibbons and Wilcox-Gök (1998), Gibbons and Lavigne (1998), and the recent work in structural equation modeling (see, e.g., Lee and Song, (2003a)), we extend the above model to the following MP model with a confirmatory factor analysis model for the underlying "response strengths" $\mathbf{z}_i$:

$$\mathbf{z}_i = \mathbf{X}_i\mathbf{B} + \mathbf{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\delta}_i, \quad i = 1, \ldots, n, \tag{3}$$

where $\mathbf{\Lambda}$ is a $J \times q$ loading matrix of parameters which may be unknown or known, $\boldsymbol{\omega}_i$ is a $q \times 1$ vector of latent factors, and $\boldsymbol{\delta}_i$ is a $J \times 1$ vector of residuals. We assume that $\boldsymbol{\omega}_i$ is independently distributed as $N[\mathbf{0}, \mathbf{\Gamma}]$, $\boldsymbol{\delta}_i$ is independently distributed as $N[\mathbf{0}, \mathbf{\Psi}]$, where $\mathbf{\Gamma}$ is an arbitrary covariance or correlation matrix, $\mathbf{\Psi}$ is a diagonal covariance matrix, and $\boldsymbol{\omega}_i$ and $\boldsymbol{\delta}_i$ are uncorrelated. For brevity, we call this the MPCFA model.

The MP model (Chib and Greenberg (1998)), as given at (1) and (2) with a general correlation matrix $\mathbf{\Sigma}$, can be analyzed with the MPCFA model by setting $\mathbf{\Lambda} = \mathbf{I}$, $\boldsymbol{\epsilon}_i = \boldsymbol{\omega}_i + \boldsymbol{\delta}_i$ and $\mathbf{\Sigma} = \mathbf{\Gamma} + \mathbf{\Psi}$. For any positive definite correlation matrix $\mathbf{\Sigma}$, there exists a value c such that $\mathbf{\Sigma}$ can be expressed a $\mathbf{\Gamma} + c\mathbf{I}_J$, where $\mathbf{I}_J$ is an J-dimensional identity matrix, and $\mathbf{\Gamma}$ is a positive definite matrix for defining the covariance matrix of the $\boldsymbol{\omega}_i$. For most practical problems with moderate correlations in $\mathbf{\Sigma}$, it is not necessary to choose a very small c to make $\mathbf{\Gamma}$ be positive definite; see the 'Six Cities' example in Section 5.1. For rare situations where $\mathbf{\Sigma}$ is nearly singular, c has to be small. This may induce some problems in the proposed approach for getting the ML estimates in the MP model.

If all $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iJ}$ equal a $k^* \times 1$ vector $\mathbf{x}_i$, then $\mathbf{X}_i\mathbf{B}$ in (3) can be written as $\mathbf{B}^*\mathbf{x}_i$, where $\mathbf{B}^*$ is a $J \times k^*$ matrix with rows equal to $\mathbf{b}_1', \ldots, \mathbf{b}_J'$. The form of (3) then reduces to the model given by Gibbons and Wilcox-Gök (1998). Another observation is that the covariance matrix of the latent factors in the MPCFA model is a general covariance matrix $\mathbf{\Gamma}$ rather than an identity matrix as in the Gibbons and Wilcox-Gök (1998) model. Our extension requires very little extra computing effort in the MCEM algorithm (see (A.3) in Appendix I).

Another special case of the MPCFA model is given by

$$\mathbf{z}_i = \mathbf{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\delta}_i, \quad i = 1, \ldots, n, \tag{4}$$

without any covariates. This model can be viewed as a confirmatory factor analysis model for dichotomous variables, or an item response model with correlated

factors. It is further reduced to the model of Bock and Aitkin (1981) and Meng and Schilling (1996) by restricting $\mathbf{\Gamma}$ to be an identity matrix. In psychometrics, the model that was developed in Bock and Aitkin (1981) and Meng and Schilling (1996) is called the "full-information item factor" model, and it has wide applications in educational testing and psychology.

Consider the relationship between the factor analysis model that is defined by $\boldsymbol{\omega}_i$ in (3) with the dichotomous variables in $\mathbf{u}_i$. Let $\mathbf{\Lambda}_j'$ and ψ_{jj} be the jth row of $\mathbf{\Lambda}$ and the jth diagonal element of $\mathbf{\Psi}$, respectively. It follows from equation (2) that

$$\begin{aligned}\Pr(u_{ij}=1|\boldsymbol{\omega}_i,\mathbf{b}_j,\mathbf{\Lambda}_j,\psi_{jj}) &= \Pr(z_{ij}>0|\boldsymbol{\omega}_i,\mathbf{b}_j,\mathbf{\Lambda}_j,\psi_{jj})\\ &= \Phi^*\{\mathbf{x}_{ij}'(\mathbf{b}_j/\psi_{jj}^{1/2})+(\mathbf{\Lambda}_j/\psi_{jj}^{1/2})\boldsymbol{\omega}_i\}. \qquad (5)\end{aligned}$$

Note that both $\mathbf{b}_j$, $\mathbf{\Lambda}_j$ and ψ_{jj} are not estimable, because $C\mathbf{b}_j/(C\psi_{jj})=\mathbf{b}_j/\psi_{jj}$, and $C\mathbf{\Lambda}_j/(C\psi_{jj}^{1/2})=\mathbf{\Lambda}_j/\psi_{jj}^{1/2}$ for any positive constant C. There are many ways to solve this identification problem. Meng and Schilling (1996) suggested fixing ψ_{jj} implicitly and estimating $\mathbf{b}_j$ and $\mathbf{\Lambda}_j$ instead of both $(\mathbf{b}_j,\mathbf{\Lambda}_j)$ and ψ_{jj}. We fix $\mathbf{\Psi}$ to be a diagonal matrix with preassigned diagonal elements as in the Lee and Song (2003a) model. Moreover, the factor analysis model is not identified because $\mathbf{z}_i=\mathbf{X}_i\mathbf{B}+\mathbf{\Lambda}^*\boldsymbol{\omega}_i^*$, where $\mathbf{\Lambda}^*=\mathbf{\Lambda}\mathbf{T}^{-1}$, and $\boldsymbol{\omega}_i^*=\mathbf{T}\boldsymbol{\omega}_i$ for any nonsingular matrix $\mathbf{T}$. A common method in factor analysis for solving this problem is to fix the approximate elements in $\mathbf{\Lambda}$ and/or $\mathbf{\Gamma}$ at preassigned values. In most applications of the confirmatory factor analysis model, the fixed values can be decided on the basis of substantive theory. For example, see the analysis of the multitrait-multimethod model that was given by DiMatteo and Heidi (1998). In the following ML analysis of the MPCFA model, we assume that the model is identified after fixing elements in $\mathbf{\Lambda}$ and/or $\mathbf{\Gamma}$.

3. ML Estimation of the MPCFA Model

Let $\mathbf{U}=(\mathbf{u}_1,\ldots,\mathbf{u}_n)$ be the observed data matrix of the dichotomous outcomes, $\mathbf{Z}=(\mathbf{z}_1,\ldots,\mathbf{z}_n)$ be the matrix of latent continuous measurements underlying $\mathbf{U}$, $\mathbf{\Omega}=(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_n)$ be the matrix of latent variables, and $\boldsymbol{\theta}$ be the parameter vector that includes unknown parameters in $\mathbf{B}$, $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$. Due to the discrete nature of the dichotomous variables, the observed-data likelihood function $P(\mathbf{U}|\boldsymbol{\theta})$ involves intractable multiple integrals. Obtaining ML estimates by the direct maximization of this function is very difficult. Indeed, even the special case with $\mathbf{\Gamma}=\mathbf{I}$ is difficult to handle (see Gibbons and Wilcox-Gök (1998)).

Inspired by the useful strategy suggested by Rubin (1991), Meng and Schilling (1996) and Lee and Shi (2001), we solve this problem by treating $\mathbf{Z}$ and $\mathbf{\Omega}$ as hypothetical missing quantities, reformulating the problem as a missing data

problem that can be solved with the well-known EM algorithm (Dempster, Laird and Rubin (1977)). Augmenting the observed dichotomous data $\mathbf{U}$ with $(\mathbf{Z}, \boldsymbol{\Omega})$, the complete-data set is $(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega})$, and the complete-data likelihood is given by

$$\begin{aligned} &P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}) = P(\mathbf{U}|\mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\theta})P(\mathbf{Z}|\boldsymbol{\Omega}, \boldsymbol{\theta})P(\boldsymbol{\Omega}|\boldsymbol{\theta}) \\ &= (2\pi)^{-np/2} \prod_{i=1}^{n} \exp\{-\frac{1}{2}(\mathbf{z}_i - \mathbf{X}_i\mathbf{B} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)'\boldsymbol{\Psi}^{-1}(\mathbf{z}_i - \mathbf{X}_i\mathbf{B} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)\} I(\mathbf{z}_i \in A_i) \\ &\quad \times (2\pi)^{-nq/2}|\boldsymbol{\Gamma}|^{-n/2} \exp(-\frac{1}{2}\sum_{i=1}^{n} \boldsymbol{\omega}_i'\boldsymbol{\Gamma}^{-1}\boldsymbol{\omega}_i), \end{aligned} \tag{6}$$

where $I(\cdot)$ is an indicator function which takes value 1 if $\mathbf{z}_i \in A_i$ and 0 otherwise, and A_i is an appropriate J-dimensional cell corresponding to $\mathbf{u}_i$ with its jth side of the form $(-\infty, 0)$ or $[0, \infty)$, for $j = 1, \ldots, J$. Note that for each $\mathbf{z}_i$, there is only one A_i such that $\mathbf{z}_i \in A_i$ and the corresponding value of the density function is nonzero.

Note that $P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta})$ involves no integrals and is much simpler than the observed-data likelihood $P(\mathbf{U}|\boldsymbol{\theta})$. Let $L_c(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}) = \log P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta})$ be the complete-data log-likelihood. The E-step at the rth iteration of the EM algorithm with a current value $\boldsymbol{\theta}^{(r)}$ is to evaluate $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = E\{L_c(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta})|\mathbf{U}, \boldsymbol{\theta}^{(r)}\}$, where the expectation is taken with respect to the joint conditional distribution of $\mathbf{Z}$ and $\boldsymbol{\Omega}$ given $\mathbf{U}$ and $\boldsymbol{\theta}^{(r)}$. The M-step is to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ with respect to $\boldsymbol{\theta}$.

Due to the complexities of the MPCFA model and the dichotomous data, the direct evaluation of the conditional expectations in the E-step is tedious. To ease this problem, we use the idea of the MCEM algorithm (Wei and Tanner (1990)) to approximate these conditional expectations via sample means of a sequence of observations of $(\mathbf{Z}, \boldsymbol{\Omega})$ that is generated from the conditional distribution $[\mathbf{Z}, \boldsymbol{\Omega}|\mathbf{U}, \boldsymbol{\theta}]$. Simulation of these observations is done by using a MCMC method, namely the Gibbs sampler (Geman and Geman (1984)), which iteratively simulates $\mathbf{Z}$ from $[\mathbf{Z}|\boldsymbol{\Omega}, \mathbf{U}, \boldsymbol{\theta}]$ and $\boldsymbol{\Omega}$ from $[\boldsymbol{\Omega}|\mathbf{Z}, \mathbf{U}, \boldsymbol{\theta}]$. Note that an estimate of any $\boldsymbol{\omega}_i$ in $\boldsymbol{\Omega}$ can be obtained via the sample mean of the simulated observations of $\boldsymbol{\omega}_i$ at the last iteration. Some details of the implementation of this procedure are outlined in Appendix I.

Several approaches can be applied to complete the M-step. For example, as proposed by Bock and Gibbons (1996) and Gibbons and Wilcox-Gök (1998), classical iterative procedures such as the Newton-Raphson algorithm and the scoring algorithm can be considered for maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$. We use the method of conditional maximization (Meng and Rubin (1993)), which gives a closed-form solution and demands less computational effort. This method performs well in

analyzing complex latent variable models in psychometrics (Lee and Zhu (2002)). Brief details of its implementation are given in Appendix II.

As suggested by Meng and Schilling (1996) and Lee and Shi (2001), the convergence of the MCEM algorithm is monitored by bridge sampling (Meng and Wong (1996)), and standard error estimates are computed via the identity that was given by Louis (1982). To save space, details are not presented.

4. Model Comparison

We now consider the problem of comparing alternative MPCFA models. Typically, competing models arise from restrictions on the covariates and/or different forms of the covariance structure $\mathbf{\Sigma}$. One example of the restriction on the covariates is that $\mathbf{b}_j = \mathbf{b}$ across the J responses. Examples of the different forms of $\mathbf{\Sigma}$ are (i) $\mathbf{\Sigma}$ is an identity matrix; (ii) $\mathbf{\Sigma}$ is in the equi-correlated form $(1-\rho)\mathbf{I}_J + \rho \mathbf{1}_J \mathbf{1}_J'$, where $|\rho| < 1$ and $\mathbf{1}_J$ is a $J \times 1$ vector of 1's; (iii) $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$, the covariance structure of an exploratory factor analysis model with a specific number of q uncorrelated factors with variance 1; (iv) $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Gamma}\mathbf{\Lambda}' + \mathbf{\Psi}$, in relation to a confirmatory factor analysis with a specific number of q correlated factors; and (v) $\mathbf{\Sigma}$ is an arbitrary correlation matrix.

The Bayesian Information Criterion (BIC) is used to compare two competitive models, M_1 and M_2:

$$\mathrm{BIC}_{12} = -2[\log P(\mathbf{U}|\hat{\boldsymbol{\theta}}_1, M_1) - \log P(\mathbf{U}|\hat{\boldsymbol{\theta}}_2, M_2)] + (d_1 - d_2)\log n, \qquad (7)$$

where $\hat{\boldsymbol{\theta}}_a$ is the ML estimate of the parameter vector $\boldsymbol{\theta}_a$ under model M_a, and d_a is the dimension of $\boldsymbol{\theta}_a$. See Raftery (1993) for some advantages of the BIC over the likelihood ratio test. The interpretation of the BIC for model comparison was given by Kass and Raftery (1995) in terms of Bayes factors. Negative values of BIC_{12} provide evidence for M_1, while positive values of BIC_{12} provide evidence for M_2. According to Kass and Raftery (1995), a value of BIC_{12} between 0 to 2 is weak, between 2 to 6 is substantial, between 6 to 10 is strong, and greater than 10 is very strong. It can be seen from (7) that the evaluation of BIC_{12} involves the computation of the observed-data log-likelihood functions, $\log P(\mathbf{U}|\hat{\boldsymbol{\theta}}_1, M_1)$ and $\log P(\mathbf{U}|\hat{\boldsymbol{\theta}}_2, M_2)$. As we mentioned before, their computation involves intractable multiple integrals. A procedure that is based on path sampling (Gelman and Meng (1998)) is described below to solve this problem.

Path sampling is a powerful method for computing (ratios of) normalizing constants of statistical models. We now apply it to computing $\log p(\mathbf{U}|\boldsymbol{\theta}, M_1)$ under model M_1 with any realization of its parameter vector $\boldsymbol{\theta}$, including $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_1$. First augment the observed data $\mathbf{U}$ with $(\mathbf{Z}, \mathbf{\Omega})$, as in the estimation. We

then select an auxiliary model M_0, such that $M_0 \subset M_1$ and $P(\mathbf{U}|\boldsymbol{\theta}, M_0)$ can be computed easily. Consider a continuous path that is defined for t in $[0, 1]$ by

$$v(t) = \int P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t) d\mathbf{Z} d\boldsymbol{\Omega}, \tag{8}$$

where $P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t)$ is a density function for each t, $P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}, a) = P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}, M_a)$, with $P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}, M_a)$ denoting the complete-data likelihood under model M_a, $a = 0, 1$. Hence,

$$v(a) = \int P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}, M_a) d\mathbf{Z} d\boldsymbol{\Omega} = P(\mathbf{U}|\boldsymbol{\theta}, M_a), \quad a = 0, 1 \tag{9}$$

which is the observed-data likelihood under M_a. The problem is to evaluate the integral in (9) to obtain $v(1)$. Let $V(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t) = d \log P(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t)/dt$, and let $\{t_{(s)}\}_{s=1}^{S}$ be fixed with $t_{(0)} = 0 < t_{(1)} < \cdots < t_{(S+1)} = 1$. It can be shown, by reasoning similar to that of Gelman and Meng (1998), that (see Appendix III)

$$\alpha = \log \left[\frac{v(1)}{v(0)}\right] \doteq \frac{1}{2} \sum_{s=1}^{S} (t_{(s+1)} - t_{(s)})(\bar{V}_{(s+1)} + \bar{V}_{(s)}), \tag{10}$$

where

$$\bar{V}_{(s)} = L^{-1} \sum_{l=1}^{L} V(\mathbf{U}, \mathbf{Z}^{(l)}, \boldsymbol{\Omega}^{(l)}, \boldsymbol{\theta}, t_{(s)}), \tag{11}$$

with $\{(\mathbf{Z}^{(l)}, \boldsymbol{\Omega}^{(l)}), l = 1, \ldots, L\}$ being the simulated observations drawn from $P(\mathbf{Z}, \boldsymbol{\Omega}|\mathbf{U}, \boldsymbol{\theta}, t_{(s)})$. It follows from (10) that

$$\log v(1) = \log v(0) + \alpha \doteq \log v(0) + \frac{1}{2} \sum_{s=0}^{S} (t_{(s+1)} - t_{(s)})(\bar{V}_{(s+1)} + \bar{V}_{(s)}). \tag{12}$$

As a program for simulating observations from $p(\mathbf{Z}, \boldsymbol{\Omega}|\mathbf{U}, \boldsymbol{\theta})$ has been constructed in the ML estimation, the implementation of this procedure is simple. Another advantage is that the summand in equation (12) is on a log scale, which is generally more stable.

Finding an appropriate auxiliary model M_0 and a good path $v(t)$ to link M_1 and M_0 is important in applying the path sampling procedure. In practice, there is a natural choice of M_0 and t, as we illustrate below. Consider the general MPCFA model

$$M_1: \qquad \mathbf{z}_i = \mathbf{X}_i \mathbf{B}_1 + \boldsymbol{\Lambda}_1 \boldsymbol{\omega}_i + \boldsymbol{\delta}_i, \quad i = 1, \ldots, n, \tag{13}$$

with parameter matrices $\mathbf{B}_1$, $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Gamma}_1$, and $\boldsymbol{\Psi}_1$. Let the auxiliary model be

$$M_0: \qquad \mathbf{z}_i = \mathbf{X}_i \mathbf{B}_1 + \boldsymbol{\delta}_i, \quad i = 1, \ldots, n. \tag{14}$$

Note that because $\mathbf{\Psi}$ is a diagonal matrix, the observed-data likelihood $v(0)$ can be computed easily. Models M_1 and M_0 can be linked via t in [0,1] as follows:

$$M_{t10}: \qquad \mathbf{z}_i = \mathbf{X}_i\mathbf{B}_1 + t\mathbf{\Lambda}_1\boldsymbol{\omega}_i + \boldsymbol{\delta}_i, \quad i = 1, \ldots, n. \tag{15}$$

Consequently, we can obtain α via (10) and (11), and finally the observed-data log-likelihood $\log v(1)$ via (12).

5. Illustrative Examples

5.1. Six cities study

The first example is based on a well-known data set from the Six Cities study, a longitudinal study of the health effects of air pollution. This data set was analyzed by Glonek and McCullagh (1995) and others with a multivariate logit model, and by Chib and Greenberg (1998) with a multivariate probit model. The data that were presented by Chib and Greenberg (1998) contain repeated dichotomous measures of the wheezing status (1 = yes, 0 = no) of 537 children from Stuebenvile, Ohio, at ages 7, 8, 9 and 10 years. The objective of this longitudinal study is to model the probability of wheeze status over time as a function of a dichotomous indicator variable that represents the mother's smoking habit during the first year of the study and the age of the child. Interpreting age as category j, we fit the following three MP models (see (1)) to the data set as did by Chib and Greenberg (1998): the full multivariate probit model where $\mathbf{\Sigma}$ is an arbitrary correlation matrix, M_1; the equi-correlated model where the correlations are equal, M_2; and the independent probit model where $\mathbf{\Sigma}$ is an identity matrix, M_3. In each MP model, the "response strengths" are specified as

$$z_{ij} = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \epsilon_i, \tag{16}$$

where x_{i1} is the age of the child, centered at 9 years, x_{i2} is a binary indicator that represents the mother's smoking habit (1 = yes, 0 = no), and x_{i3} is an interaction between smoking habit and age. Note that the regression parameter is constrained to be constant across j. To avoid sampling from a multivariate truncated normal distribution, we treat (16) as a special case of the MPCFA model defined in (3), with $\mathbf{z}_i = (z_{i1}, \ldots, z_{i4})'$, $\mathbf{X}_i = (1, x_{i1}, x_{i2}, x_{i3})$, $\mathbf{B} = (b_0, b_1, b_2, b_3)'$, $\mathbf{\Lambda} = \mathbf{I}_4$, a 4×4 identity matrix, and $\mathbf{\Psi}$ is fixed at $0.2\mathbf{I}_4$ to identify the model. For model M_1, $\mathbf{\Gamma} = (\gamma_{ij})$ is taken to be a symmetric matrix with unknown off-diagonal elements, but the diagonal elements are all equal to 0.8 so that $\mathbf{\Sigma} = \mathbf{\Gamma} + \mathbf{\Psi}$ is a correlation matrix. For M_2, $\mathbf{\Gamma} = (0.8 - \rho)\mathbf{I}_4 + \rho\mathbf{1}_4\mathbf{1}_4'$, so that $\mathbf{\Sigma}$ is again a correlation matrix. For M_3, $\mathbf{\Gamma} = 0.8\mathbf{I}_4$, so that $\mathbf{\Sigma}$ is an identity matrix.

In the ML estimation, conditional expectations at the E-step of the MCEM algorithm are approximated by 20 observations which are generated from the

conditional distributions for the first 30 MCEM steps and 200 observations for the later MCEM steps. The algorithm converges after about 50 iterations. To be conservative, we take the parameters' values at the 60th iteration as their ML estimates. The results are reported in Table 1. Note that the ML estimates and standard error estimates are close to those which were reported by Chib and Greenberg (1998).

Table 1. ML estimates and their standard errors for M_1, M_2 and M_3: Six Cities data.

	M_1		M_2		M_3	
	MLE	SE	MLE	SE	MLE	SE
b_0	-1.118	0.066	-1.118	0.056	-1.122	0.046
b_1	-0.077	0.031	-0.078	0.032	-0.074	0.034
b_2	0.151	0.107	0.167	0.090	0.165	0.073
b_3	0.037	0.052	0.038	0.052	0.035	0.056
γ_{21}	0.599	0.076	0.601	0.012	–	
γ_{31}	0.592	0.053	–		–	
γ_{41}	0.475	0.084	–		–	
γ_{23}	0.687	0.056	–		–	
γ_{24}	0.610	0.034	–		–	
γ_{34}	0.653	0.082	–		–	

In computing $p(\mathbf{U}|\hat{\boldsymbol{\theta}}_a, M_a)$, $a = 1, 2, 3$, we use $M_0 : \mathbf{z}_i = \mathbf{X}_i\mathbf{B} + \boldsymbol{\delta}_i$ as the auxiliary model. Models M_a and M_0 are linked by a model M_t via path $t \in [0, 1]$ as follows:

$$M_t : \quad \mathbf{z}_i = \mathbf{X}_i\mathbf{B} + t\boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\delta}_i, \quad i = 1, \ldots, n,$$

where $\boldsymbol{\omega}_i$ is the latent factor defined according to M_a. Clearly, $M_t = M_a$ when $t = 1$ and $M_t = M_0$ when $t = 0$, and $v(0)$ can be easily evaluated in closed form. The observed-data log-likelihood $\log P(\mathbf{U}|\hat{\boldsymbol{\theta}}_a, M_a)$, $a = 1, 2, 3$, as estimated by the path sampling procedure, are -968.156, -976.842, and -990.471, respectively. Hence, $\mathrm{BIC}_{12} = 14.058$ and $\mathrm{BIC}_{23} = -20.972$. These results give decisive evidence against M_1 and M_3, and in favor of the equi-correlated model M_2. These conclusions also agree with those of Chib and Greenberg (1998). The estimated equation in relation to the "response strengths" under the selected model M_2 is given by (see Table 1) $z_{ij} = -1.118 - 0.078x_{i1} + 0.167x_{i2} + 0.038x_{i3}$, and the wheezing status across the ages of 7, 8, 9 and 10 is equi-correlated with a correlation 0.601. Similar results are obtained with $\boldsymbol{\Psi}$ fixed at $0.1\mathbf{I}_4$ to identify the model, and with the diagonal elements of $\boldsymbol{\Gamma}$ adjusted accordingly.

A mixed effect model is defined by $\mathbf{z}_i = \mathbf{X}_i\mathbf{B} + \boldsymbol{\Lambda}_i^*\boldsymbol{\omega}_i + \boldsymbol{\delta}_i$, where $\boldsymbol{\Lambda}_i^*$ is a known design matrix rather than a matrix of parameters, and $\boldsymbol{\omega}_i$ and $\boldsymbol{\delta}_i$ are

independently distributed as $N[\mathbf{0}, \mathbf{\Gamma}]$ and $N[\mathbf{0}, \sigma^2\mathbf{I}]$, respectively. As pointed out by a reviewer, the data set in this example can be analyzed as a special case of a mixed effect model in which $\mathbf{\Lambda}_i^*$ is fixed at the identity matrix such that $\mathbf{z}_i \stackrel{D}{=} N[\mathbf{X}_i\mathbf{B}, \mathbf{\Gamma} + \sigma^2\mathbf{I}]$. The main distinctions between a mixed effect model and the MPCFA model are on the different natures and properties between $\mathbf{\Lambda}_i^*$ and $\mathbf{\Lambda}$, and the different dimensions of $\boldsymbol{\omega}_i$, while a minor distinction is the covariance matrix of $\boldsymbol{\delta}_i$.

5.2. A compliance study of patients

The purpose of this example is to illustrate the use of the MPCFA model as a latent variable model with covariates. It has recently been pointed out that patient adherence to prescribed medication is crucial to the success of medical treatment (Czajkowski, Margaret and Ashley (1998)), and that nonadherence leads to misjudgment of the effectiveness of medication (Rand and Kathleen (1998)). In the promotion of adherence, it is desirable to establish a statistical model to study the correlation between nonadherence and its core factors such as health condition, patient knowledge of medication, attitudes and beliefs concerning medication (Andre and Lynda (1991)), and so on. To enrich existing knowledge about patient nonaderence, the Department of Medicine and Therapeutics, Community and Family Medicine, and Pharmacy at the Chinese University of Hong Kong conducted a survey of ethnic Chinese patients who had been diagnosed as suffering from hypertension (Czajkowski et al. (1998)). One objective was to measure and examine correlations among latent variables such as physician advice and concern, patient knowledge and belief, social cognition, and social influence, and the subsequent study reported nonadherence with reference to a factor analysis model or a more general structural equation model. Because the study involved many dichotomous variables and the manifest indicators for the factors are influenced by covariates, the MPCFA model is useful.

To demonstrate the methodology, suppose we are interested in establishing a MPCFA model with two fixed covariates about patient education (coded by 0, 1, 2, 3) and the existence of "side-effects" (coded by 0 and 1), as well as latent factors of patient "nonadherence", "knowledge of medication", and "health condition", by analyzing the related portion of the whole data set. Nine dichotomous manifest variables are selected as indicators of the latent variables mentioned above. Translations of the corresponding questions from Chinese into English are listed in Table 2, together with their frequencies. For brevity, we omit a small number of observations with missing entries, and the remaining sample size is 837.

The resultant data set is analyzed using a MPCFA model (3). Although other structures for the loading matrix can be considered, for clear interpretation we

Table 2. Questions associated with the manifest variables. Frequencies of (Yes '1'/No '0') are in parentheses.

u_1:	Did you have any surplus in the previous prescribed drugs? (175/662)
u_2:	Did you stop/reduce/increase the dosage? (69/768)
u_3:	Did you forget to take medications? (391/446)
u_4:	Do you feel you have hypertension? (363/474)
u_5:	Do you know the reasons for taking drugs? (650/187)
u_6:	Do you know the reasons for taking drugs for a long term? (605/232)
u_7:	In the past two weeks, did you have emotional problems? (387/450)
u_8:	In the past two weeks, did your health cause any difficulties in daily activities? (181/656)
u_9:	In the past two weeks, did your health cause any difficulties in social activities? (177/660)

choose the structure that gives nonoverlapping latent factors. Hence, the following specification of the loading matrix $\mathbf{\Lambda}$ is used:

$$\mathbf{\Lambda}' = \begin{bmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{73} & \lambda_{83} & \lambda_{93} \end{bmatrix},$$

where the λ_{ij}s are the unknown factor loading parameters, while the 0's are fixed in the estimation for achieving an identified model. From the meanings of the questions (see Table 2), it is clear that this structure gives three nonoverlapping factors (latent variables), which can be interpreted as the "*nonadherence*, ω_1", "*knowledge of medication*, ω_2" and "*health condition*, ω_3" of the patients. To identify the model, we also fix $\mathbf{\Psi}$ to be an identity matrix and $\mathbf{\Gamma} = (\gamma_{ij})$ to be a correlation matrix.

We compare the following three models.

M_1: An MPCFA model that involves covariates and correlated latent factors with the above specification for the loading matrix $\mathbf{\Lambda}$.

M_2: An MPCFA model as in M_1, but with uncorrelated latent factors that each have variance 1, that is, $\mathbf{\Gamma}$ is an identity matrix.

M_3: A MP model with diagonal $\mathbf{\Sigma}$ but without latent variables, so that $\boldsymbol{\omega}_i = \mathbf{0}$.

In the ML estimation, conditional expectations at the E-step of the MCEM algorithm are approximated by 20 observations generated from the conditional distributions for the first 80 MCEM steps and 200 observations for the later MCEM steps. The algorithm converges at the 100th iteration. The ML estimates for models M_1, M_2 and M_3 are reported in Table 3, together with their standard error estimates. The observed log-likelihoods that correspond to M_1,

Table 3. ML estimates and their standard errors for M_1, M_2 and M_3 in the Compliance Study.

	M_1		M_2		M_3	
	MLE	SE	MLE	SE	MLE	SE
λ_{11}	1.091	0.070	1.791	0.161	–	
λ_{21}	1.418	0.291	0.845	0.206	–	
λ_{31}	0.311	0.044	0.450	0.096	–	
λ_{42}	0.099	0.052	0.217	0.069	–	
λ_{52}	1.370	0.145	1.367	0.133	–	
λ_{62}	1.471	0.201	1.395	0.165	–	
λ_{73}	0.658	0.067	0.609	0.072	–	
λ_{83}	2.271	0.136	2.242	0.192	–	
λ_{93}	2.244	0.207	2.285	0.144	–	
b_{11}	-0.722	0.026	-1.006	0.054	-0.506	0.043
b_{21}	-1.207	0.187	-0.935	0.083	-0.733	0.063
b_{31}	-0.022	0.033	-0.023	0.034	-0.022	0.048
b_{41}	-0.121	0.032	-0.123	0.035	-0.123	0.067
b_{51}	0.876	0.068	0.891	0.076	0.532	0.051
b_{61}	0.729	0.070	0.710	0.084	0.418	0.065
b_{71}	-0.081	0.032	-0.085	0.039	-0.075	0.037
b_{81}	-1.179	0.024	-1.162	0.034	-0.488	0.046
b_{91}	-1.163	0.106	-1.180	0.145	-0.486	0.121
b_{12}	0.097	0.137	0.090	0.210	0.024	0.176
b_{22}	0.150	0.087	0.068	0.156	0.029	0.364
b_{32}	0.099	0.088	0.117	0.122	0.102	0.166
b_{42}	0.314	0.108	0.322	0.115	0.310	0.179
b_{52}	0.267	0.155	0.328	0.185	0.237	0.139
b_{62}	-0.076	0.211	-0.045	0.156	-0.016	0.186
b_{72}	0.262	0.113	0.270	0.123	0.224	0.136
b_{82}	0.545	0.195	0.504	0.144	0.198	0.132
b_{92}	0.518	0.199	0.502	0.243	0.203	0.143
γ_{21}	-0.425	0.072	–		–	
γ_{31}	0.492	0.044	–		–	
γ_{23}	-0.479	0.041	–		–	

M_2 and M_3 computed via the path sampling procedure are -5619.6, -5864.6 and -5898.7, respectively. Consequently, BIC_{12}=-469.7 and $\text{BIC}_{13} = -470.7$. These results give decisive evidence in favor of the MPCFA model M_1. As both M_2 and M_3 are nested in M_1, we can consider the following likelihood ratio test on the basis of $LR_{1k} = -2[\log P(\mathbf{U}|\hat{\boldsymbol{\theta}}_1, M_1) - \log P(\mathbf{U}|\hat{\boldsymbol{\theta}}_k, M_k)]$, $k = 2, 3$. As $LR_{12} = 490.0$ and $LR_{13} = 558.2$, models M_2 and M_3 are clearly rejected at type I error 0.1, based on χ^2-distributions with degrees of freedom 3 and 12,

respectively. This is the conclusion obtained from the BIC.

On the basis of model M_1 and the meaning of the questions (Table 2), the most important interpretations of the ML estimates are as follows: (i) From the estimates $\hat{b}_{11}, \ldots, \hat{b}_{91}$ that correspond to the effects of the first covariate about patient education, we see that higher education has (a) a negative effect on the indicators for "nonadherence"; (b) a positive effect on the more relevant indicators (u_4 and u_5) for "knowledge of medication"; and (c) a negative effect on the indicators for weaker or worse "health condition". (ii) From the estimates $\hat{b}_{21}, \ldots, \hat{b}_{29}$ that correspond to the effects of the second covariate about side-effects, we see that the presence of the side-effects has (a) a positive effect on the indicators for "nonadherence"; (b) a positive effect on the first two indicators for "knowledge of medication"; and (c) a positive effect on the indicators for weaker or worse "health condition". (iii) As $\hat{\gamma}_{21} = -0.425$, we know that "nonadherence" is negatively correlated with "knowledge of medication". (iv) As $\hat{\gamma}_{31} = 0.492$, we know that "nonadherence" is positively correlated with weaker or worse "health condition". (v) As expected, "knowledge of medication" is negatively correlated with weaker and worse "health condition", as $\hat{\gamma}_{23} = -0.479$ indicates. Based on the above results, we arrive at a conclusion that it is desirable to better educate patients about their illness and encourage them to pay more attention to their health.

6. Discussion

We propose a model for analyzing multivariate dichotomous data that combines the MP model in biostatistics and the confirmatory factor analysis model in psychometrics. Methods for ML estimation and model comparison are developed with such powerful tools of statistical computing as the MCEM algorithm, the Gibbs sampler, and path sampling. It may be possible to fit the proposed model with dichotomous data by using some software in psychometrics, for example Mplus (Muthén and Muthén (2001)). However, one obtains neither the ML estimates nor the value of the observed data log-likelihood in this way. Hence estimates they give statistically less optimal and one cannot be used to do model comparison.

The factor analysis model is a special case of the more general structural equation models (SEMs) (Bentler (1992), Bollen (1989), Everitt (1984) and Jöreskog and Sörbom(1996)). SEMs, which are sometimes called latent variable models, have been extensively applied to behavioral, psychological, and social research for assessing the latent traits of manifest variables. Recently, SEMs and their submodels have also received much attention in biostatistics, and have been widely applied to medical research (Douglas (1999), Lee and Song (2003a), Bentler and Stein (1992), Beacon and Thompson (1998), Palta (1999) and Chan

et al. (1996)). By integrating these psychometric models into biostatistics and statistics computing methods, our development can be generalized to handle complex models such as nonlinear SEMs with covariates (Lee and Song (2003b)) and two-level SEMs (Lee and Shi (2001) and Ansari and Jedidi (2000)), and such other complex data structure as mixed continuous and polytomous data (Sammel, Ryan and Legler (1997) and Shi and Lee (2000)) and missing data (Song and Lee (2002)). Furthermore, the sensitivity of the ML results in relation to the model and data inputs can be analyzed via the local and global influence approaches of Zhu and Lee (2001) and Zhu, Lee, Wei and Zhou (2001), respectively.

If useful prior information about the parameters in the MPCFA model is available, then the Monte Carlo simulation in the E-step of the proposed MCEM algorithm can be extended to a full Bayesian analysis to achieve more accurate results. The Bayes factor (see Berger and Perrichi (2001) and Kass and Raftery (1995)) is an important statistic for Bayesian model comparison. It is well-known that BIC is an asymptotic approximation to the Bayes factor. Hence, BIC should be used with confidence only in situations where there is a large sample size, relative to the number of parameters. Usually the number of unknown parameters in MPCFA model is not small. However, as the sample sizes for most studies in behavioral and social sciences are quite large, this problem is not serious. For medical research that is not related to some rare disease, for example the compliance study of patients in Section 5.2, it is also likely to have large sample sizes. While BIC is an approximation of the Bayes factor with the relative error O(1), see Kass and Raftery (1995), a more accurate approximation has given by Berger and Perrichi (2001). Still, the model that is selected by the BIC may not be the true model. Hence it is always desirable to examine residuals (see (3)): $\hat{\boldsymbol{\delta}}_i = \hat{\mathbf{z}}_i - \mathbf{X}_i\hat{\mathbf{B}} - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\omega}}_i, i = 1, \ldots, n$, where $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Lambda}}$ are the ML estimates, $\hat{\mathbf{z}}_i$ and $\hat{\boldsymbol{\omega}}_i$ are the estimates of $\mathbf{z}_i$ and $\boldsymbol{\omega}_i$ that can be obtained from the simulated observations in the E-step of the last iteration of the MCEM algorithm.

Based on similar reasoning as given in Appendix III, see also Lee and Song (2003b, c), path sampling can be applied for computing Bayes factor B_{12} for comparing MPCFA models M_1 and M_2 in a Bayesian context. Let $\boldsymbol{\theta}$ be the parameter vector that contains common and distinct parameters in M_1 and M_2, and let $V^*(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t) = dP(\mathbf{U}, \mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\theta}|t)/dt$. It can be shown that

$$\log B_{12} \doteq \frac{1}{2}\sum_{s=1}^{S}(t_{(s+1)} - t_{(s)})(\bar{V}^*_{(s+1)} + \bar{V}^*_{(s)}),$$

where $\bar{V}^* = L^{-1}\sum_{l=1}^{L} V^*(\mathbf{U}, \mathbf{Z}^{(l)}, \boldsymbol{\Omega}^{(l)}, \boldsymbol{\theta}^{(l)}, t_{(s)})$, with $\{(\mathbf{Z}^{(l)}, \boldsymbol{\Omega}^{(l)}, \boldsymbol{\theta}^{(l)}),\ l = 1, \ldots, L\}$ being the simulated observations from the joint posterior distribution

$P(\mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\theta}|\mathbf{U}, t_{(s)})$ via some Markov chain Monte Carlo method, such as the Gibbs sampler. Note that the above expression for $\log B_{12}$ is similar to (10). The main difference is that $\boldsymbol{\theta}$ is not fixed in computing $\bar{V}^*$, but a sequence of $\boldsymbol{\theta}^{(l)}$ that is simulated from the joint posterior distribution $P(\mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\theta}|\mathbf{U}, t_{(s)})$ is used.

Acknowledgements

The work that is described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. CUHK 4243/02H). The authors are greatly indebted to an associate editor and two anonymous reviewers for valuable comments which improve the paper substantially. They are also grateful to Juliana C. N. Chan, Associate Professor, Department of Medicine and Therapentics, CUHK, and Grace Chan, Pharmacist, Prince Wales Hospital, Hong Kong for introducing them to the CUHK study on hypertensive patients, and for providing the data in the example.

Appendix I. Implementation of the Gibbs Sampler

To implement the Gibbs sampler for simulating observations in the E-step, we start with initial values $(\mathbf{Z}^{(0)}, \boldsymbol{\Omega}^{(0)})$, simulate $(\mathbf{Z}^{(1)}, \boldsymbol{\Omega}^{(1)})$, and continue as follows. At the mth iteration with current $(\mathbf{Z}^{(m)}, \boldsymbol{\Omega}^{(m)})$:

$$\begin{aligned} &\text{(a) Generate } \mathbf{Z}^{(m+1)} \text{ from } [\mathbf{Z}|\boldsymbol{\Omega}^{(m)}, \mathbf{U}, \boldsymbol{\theta}], \\ &\text{(b) Generate } \boldsymbol{\Omega}^{(m+1)} \text{ from } [\boldsymbol{\Omega}|\mathbf{Z}^{(m+1)}, \mathbf{U}, \boldsymbol{\theta}]. \end{aligned} \tag{A.1}$$

It has been shown (Geman and Geman (1984) and Geyer (1992)) that under mild conditions and after a sufficiently large number of iterations, the joint distribution of $(\mathbf{Z}^{(m)}, \boldsymbol{\Omega}^{(m)})$ converges at an exponential rate to the desired posterior distribution $[\mathbf{Z}, \boldsymbol{\Omega}|\mathbf{U}, \boldsymbol{\theta}]$. The required conditional distributions that are involved in (A.1) are briefly derived below.

$[\mathbf{Z}|\boldsymbol{\Omega}, \mathbf{U}, \boldsymbol{\theta}]$: Let $\boldsymbol{\Lambda}_j$ be the j-row of $\boldsymbol{\Lambda}$. As the $\mathbf{z}_i$ are mutually independent, it follows from (3) that

$$p(\mathbf{Z}|\boldsymbol{\Omega}, \mathbf{U}, \boldsymbol{\theta}) = \prod_{i=1}^{n} P(\mathbf{z}_i|\boldsymbol{\omega}_i, \mathbf{u}_i, \boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{J} P(z_{ij}|\boldsymbol{\omega}_i, u_{ij}, \boldsymbol{\theta}), \tag{A.2}$$

where

$$P(z_{ij}|\boldsymbol{\omega}_i, u_{ij}, \boldsymbol{\theta}) \sim \begin{cases} N[\mathbf{x}'_{ij}\mathbf{b}_j + \boldsymbol{\Lambda}_j\boldsymbol{\omega}_i, \psi_{jj}]I_{(-\infty,0)}(z_{ij}), \text{ if } u_{ij} = 0 \\ N[\mathbf{x}'_{ij}\mathbf{b}_j + \boldsymbol{\Lambda}_j\boldsymbol{\omega}_i, \psi_{jj}]I_{(0,\infty)}(z_{ij}), \text{ if } u_{ij} = 1, \end{cases}$$

Note that (A.2) involves univariate rather than multivariate truncated normal distributions. The commonly used inverse distribution method, as given by Devroye (1985), can be employed to simulate observations from this relatively simple distribution.

$[\boldsymbol{\Omega}|\mathbf{Z},\mathbf{U},\boldsymbol{\theta}]$: As $\boldsymbol{\Omega}$ is independent of $\mathbf{U}$ with $\mathbf{Z}$ given, and $\boldsymbol{\omega}_i$, $i = 1,\ldots,n$ are mutually independent,

$$P(\boldsymbol{\Omega}|\mathbf{Z},\mathbf{U},\boldsymbol{\theta}) = P(\boldsymbol{\Omega}|\mathbf{Z},\boldsymbol{\theta}) = \prod_{i=1}^{n} P(\boldsymbol{\omega}_i|\mathbf{z}_i,\boldsymbol{\theta}_i), \tag{A.3}$$

where $[\boldsymbol{\omega}_i|\mathbf{z}_i,\boldsymbol{\theta}] \overset{D}{=} N[\boldsymbol{\Sigma}^*\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}(\mathbf{z}_i - \mathbf{X}_i\mathbf{B}), \boldsymbol{\Sigma}^*]$, with $\boldsymbol{\Sigma}^* = (\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}$. The simulation of observations from the standard normal distribution is fast and straightforward. We emphasize here that an identity or an arbitrary correlation matrix $\boldsymbol{\Gamma}$ requires the same computational effort in simulating $\boldsymbol{\omega}_i$.

Appendix II. Conditional Maximization

At the M-step, we solve the system of equations

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})}{\partial \boldsymbol{\theta}} = E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} L_c(\mathbf{U},\mathbf{Z},\boldsymbol{\Omega}|\boldsymbol{\theta}) \middle| \mathbf{U}, \boldsymbol{\theta}^{(r)} \right\} = 0 \tag{A.4}$$

by means of three conditional maximizations (see Meng and Rubin (1993)). The solutions for updating $\mathbf{B}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are:

$$\begin{aligned}
\hat{\mathbf{B}} &= (\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i')^{-1} \sum_{i=1}^{n} \mathbf{X}_i' \mathrm{E}[(\mathbf{z}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)|\mathbf{U},\boldsymbol{\theta}^{(r)}], \\
\hat{\boldsymbol{\Lambda}}_j &= \left(\sum_{i=1}^{n} \mathrm{E}[\boldsymbol{\omega}_i\boldsymbol{\omega}_i'|\mathbf{U},\boldsymbol{\theta}] \right)^{-1} \sum_{i=1}^{n} \mathrm{E}[\boldsymbol{\omega}_i(z_{ij} - \mathbf{x}_{ij}'\hat{\mathbf{b}}_j)|\mathbf{U},\boldsymbol{\theta}^{(r)}], \\
\hat{\boldsymbol{\Gamma}} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(\boldsymbol{\omega}_i\boldsymbol{\omega}_i'|\mathbf{U},\boldsymbol{\theta}^{(r)}).
\end{aligned} \tag{A.5}$$

The conditional expectations that are involved in (A.5) are approximated by the simulated observations that are obtained with the Gibbs sampler at the E-step.

Appendix III. Proof of (10) in the Path Sampling Procedure

Equation (10) is proved on the basis of the definition of $v(t)$ that is given in (8) and reasoning which is similar to that of Gelman and Meng (1998). Assuming the legitimacy of interchange of integration with differentiation, it follows from (8) that

$$\begin{aligned}
\frac{d\log v(t)}{dt} &= \int P(\mathbf{Z},\boldsymbol{\Omega}|\mathbf{U},\boldsymbol{\theta},t) \frac{d\log P(\mathbf{U},\mathbf{Z},\boldsymbol{\Omega}|\boldsymbol{\theta},t)}{dt} d\mathbf{Z}d\boldsymbol{\Omega} \\
&= E_{Z,\Omega}[V(\mathbf{U},\mathbf{Z},\boldsymbol{\Omega},\boldsymbol{\theta},t)],
\end{aligned} \tag{A.6}$$

where $E_{Z,\Omega}$ denotes the expectation with respect to the sampling distribution of $P(\mathbf{Z}, \mathbf{\Omega}|\ \mathbf{U}, \boldsymbol{\theta}, t)$, and $V(\mathbf{U}, \mathbf{Z}, \mathbf{\Omega}, \boldsymbol{\theta}, t) = d \log P(\mathbf{U}, \mathbf{Z}, \mathbf{\Omega}|\boldsymbol{\theta}, t)/dt$. Integrating (A.6) from 0 to 1, we have

$$\alpha = \log v(1) - \log v(0) = \int_0^1 E_{Z,\Omega}[V(\mathbf{U}, \mathbf{Z}, \mathbf{\Omega}, \boldsymbol{\theta}, t)]dt. \tag{A.7}$$

Let $\{t_{(s)}\}_{s=1}^S$ be such that $t_{(0)} = 0 < t_{(1)} < \cdots < t_{(S+1)} = 1$. The integration of the right hand side of (A.7) is estimated by $(1/2)\sum_{s=0}^S (t_{(s+1)} - t_{(s)})(\bar{V}_{(s+1)} + \bar{V}_{(s)})$, where $\bar{V}_{(s)} = L^{-1}\sum_{l=1}^L V(\mathbf{U}, \mathbf{Z}^{(l)}, \mathbf{\Omega}^{(l)}, \boldsymbol{\theta}, t_{(s)})$, with $\{(\mathbf{Z}^{(l)}, \mathbf{\Omega}^{(l)}), l = 1, \ldots, L\}$ being the simulated observations that are drawn from $P(\mathbf{Z}, \mathbf{\Omega}|\ \mathbf{U}, \boldsymbol{\theta}, t_{(s)})$.

References

Andre, T. and Lynda, B. (1991). Knowledge of acquired immune deficiency syndrome and sexual responsibility among high school students. *Youth Soc.* **22**, 339-361.

Ansari, A. and Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika* **65**, 475-498.

Ashford J. R. and Sowden R. R. (1970). Multivariate probit analysis. *Biometrics* **26**, 535-546.

Beacon, H. J. and Thompson, S. G. (1998). The analysis of complex patterns of longitudinal binary response: an example of transient dysphagia following radiotherapy. *Statist. Medicine* **17**, 2551-2561.

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

Bentler, P. M. and Stein, J. A. (1992). Structural equation models in medical research. *Statist. Meth. Medical Res.* **1**, 159-181.

Berger, J. O. and Pericchi, L. R. (2001). Objective Bayesian methods for model selection: introduction and comparison. In *Model Selection.* (Edited by P. Lahiri). Beachwood, Ohio.

Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika* **46**, 443-445.

Bock, R. D. and Gibbons, R. D. (1996). High dimensional multivariate probit analysis. *Biometrics* **52**, 1183-1194.

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

Chan, C. N., Chan, Y. W., Cheung, C. K., Swaminathan, R., Lan, M. C., Cockrama, S. and Wooa, J. (1996). The metabolic syndrome in Hong Kong Chinese: the interrelationship among its components analyzed by structural equation modeling. *Diabetes Care* **19**, 953-959.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika* **85**, 347-361.

Czajkowski, S. M., Margaret, A. C. and Ashley, W. S. (1998). Adherence and the placebo effect, In *The Handbook of Health Behavior Change*. 2nd edition. (Edited by A. S. Sally, B. S. Eleanor, K. O. Judith and L. M. Wendy), 513-534. Springer, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Devroye, L. (1985). *Non-Uniform Random Variate Generation*. Springer Verlag, New York.

DiMatteo, M. R. and Heidi, S. L. (1998). Promoting Adherence to courses of treatment: mutual collaboration in the physician-patient relationship, In *Health Communication Research: A*

Guide to Developments and Directions (Edited by D. J. Lorraine, K. D. Bernard), 71-98. Greenwood, Westport, Connecticut.

Douglas, J. A. (1999). Item response models for longitudinal quality of life data in clinical trials. *Statist. Medicine* **18**, 2917-2931.

Everitt, B. S. (1984). *An Introduction to Latent Variable Models*. Chapman and Hall, London.

Gelman, A. and Meng, X. L. (1998). Simulating normalizing constant: From importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13**, 163-185.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721-741.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.* **7**, 473-511.

Gibbons, R. D. and Lavigne, J. V. (1998). Emergence of childhood psychiatric disorders: a multivariate probit analysis. *Statist. Medicine* **17**, 2487-2499.

Gibbons, R. D. and Wilcox-Gök, V. (1998). Health service utilization and insurance coverage: a multivariate probit model. *J. Amer. Statist. Assoc.* **93**, 63-72.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *J. Roy. Statist. Soc. Ser. B* **57**, 533-546.

Gould, M. S. and Wunsch-Hizig, R. and Dohrenwend, B. (1981). Estimating the prevalence of childhood psychopathology. *J. Amer. Acad. Child and Adolescent Psychiatry* **20**, 462-476.

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

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90**, 773-795.

Kolakowski, D. and Bock, R. D. (1981). A multivariate generalization of probit analysis. *Biometrics* **37**, 541-551.

Lee, S. Y. and Shi, J. Q. (2001). Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics* **57**, 787-794.

Lee, S. Y. and Song, X. Y. (2003a). Bayesian analysis of structural equation models with dichotomous variables. *Statist. Medicine* **22**, 3073-3088.

Lee, S. Y. and Song, X. Y. (2003b). Model Comparison of nonlinear structural equation models with fixed covariates. *Psychometrika* **68**, 27-47.

Lee, S. Y. and Song, X. Y. (2003c). Bayesian model selection for mixtures of structural equation models with an unknown number of components. *British J. Math. Statist. Psych.* **56**, 145-165.

Lee, S. Y. and Zhu, H. T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika* **67**, 189-210.

Link, C. R., Long, S. H. and Settle, R. (1980). Cost sharing, supplementary insurance, and health services utilization among the medicare elderly. *Health Care Financing Rev.* **2**, 25-31.

Louis, T. A. (1982). Finding the observed information matrix when using EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44**, 226-233.

Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267-278.

Meng, X. L. and Schilling, S. (1996). Fitting full-information item factor models and an empirical investigation of bridge sampling. *J. Amer. Statist. Assoc.* **91**, 1254-1267.

Meng, X. L. and Wong, H. W. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statist. Sinica* **6**, 831-360.

Muthén, L. K. and Muthén, B. O. (2001). *Mplus User's Guide*. Muthén and Muthén, Los Angeles.

Ochi, Y. and Prentice, R. L. (1984). Likelihood inference in a correlated probit regression model. *Biometrika* **71**, 531-543.

Palta, M. (1999). Latent variables, measurement error and methods for analyzing longitudinal binary and ordinal data. *Statist. Medicine* **18**, 385-396.

Raftery, A. E. (1993). Bayesian model selection in structural equation models. In *Testing Structural Equation Models* (Edited by K. A. Bollen and J. S. Long), 163-180. Sage, Beverly Hills, California.

Rand, C. S. and Kathleen, W. (1998). Measuring adherence with medication regiments in clinical care and research, In *The Handbook of Health Behavior Change*. 2nd edition. (Edited by A. S. Sally, B. S. Eleanor, K. O. Judith, L. M. Wendy), 71-103. Springer, New York.

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

Sammel, M. D., Ryan, L. M. and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *J. Roy. Statist. Soc. Ser. B* **59**, 667-678.

Shi, J. Q. and Lee, S. Y. (2000). Latent variable models with mixed continuous and polytomous data. *J. Roy. Statist. Soc. Ser. B* **62**, 77-87.

Song, X. Y. and Lee, S. Y. (2002). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika* **67**, 261-288.

Takane, Y. and de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* **52**, 393-408.

Vikan, A. (1985). Psychiatric epidemiology in a sample of 1510 ten-year-old children. *J. Child Psych. Psychiatry* **26**, 55-76.

Wang, C., Douglas, J. and Anderson, S. (2002). Item response models for joint analysis of quality of life and survival. *Statist. Medicine* **21**, 129-142.

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. *J. Amer. Statist. Assoc.* **85**, 699-704.

Wolfe, J. R. and Goddeeris, J. H. (1991). Adverse selection, moral hazard, and wealth effects in the medigap insurance market. *J. Health Economics* **10**, 433-459.

Zhu, H. T. and Lee, S. Y. (2001). Local influence for incomplete-data models. *J. Roy. Statist. Soc. Ser. B* **63**, 111-126.

Zhu, H. T., Lee, S. Y., Wei, B. C. and Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika* **88**, 727-737.

Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

E-mail: xysong@sta.cuhk.edu.hk

Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

E-mail: sylee@sparc2.sta.cuhk.edu.hk

(Received December 2003; accepted June 2004)