Statistica Sinica **16**(2006), 1117-1141

ANALYSIS OF NONLINEAR STRUCTURAL EQUATION MODELS WITH NONIGNORABLE MISSING COVARIATES AND ORDERED CATEGORICAL DATA

Sik-Yum Lee and Nian-Sheng Tang

The Chinese University of Hong Kong and Yunnan University

Abstract: The main purpose of this article is to investigate a nonlinear structural equation model with covariates and mixed continuous and ordered categorical outcomes, in the presence of missing observations and missing covariates that are missing with a nonignorable mechanism. The nonignorable missingness mechanism is specified by a logistic regression model. A Bayesian approach is proposed for obtaining the joint Bayesian estimates of structural parameters, latent variables and parameters in the logistic regression model. An algorithm that combines the Gibbs sampler and the Metropolis-Hastings algorithm is developed for sampling observations from the posterior distributions, and for obtaining the Bayesian solution. A procedure for computing the Bayes factor for model comparison is developed via path sampling. Sensitivity analyses of the results with respect to the assumed model for the missingness mechanism, the prior inputs, and the missing covariate distributions are conducted via simulation studies. An example is presented to illustrate the newly developed Bayesian methodologies.

Key words and phrases: Bayes factor, Gibbs sampler, Metropolis-Hastings algorithm, nonignorable missing data, Path sampling, sensitivity analysis.

1. Introduction

Latent variables are often encountered in behavioral, educational, medical, psychological, and social research. Investigation and analysis of latent variables for assessing the relationships of observed variables and latent variables have received a great of attention in biostatistics, psychometrics, and statistics. One approach has focused on analyzing the effects of latent variables and fixed covariates on the mean of the observed variables; see Laird and Ware (1982) and Chib and Greenberg (1998). Another approach, commonly referred to as structural equation models (SEMs, see Bollen (1989) among others), has focused on identifying the latent variables from the manifest variables, and on investigating the relationships of latent variables among themselves. On the basis of more than a dozen user-friendly software packages in the field, SEMs have been widely applied to various fields, such as organization and management (Williams, Edwards and Vandenberg (2003)), marketing (Bagozzi, Gopinath and

Nyer (1999)), quality of life (Lee, Song, Skevington and Hao (2005)), and transportation (Golob (2003)), among others.

In this paper, we first introduce nonlinear SEMs (NSEMs) that accommodate covariates, in the context of mixed continuous and ordered categorical variables. In the literature, a factor analysis model with covariates has been developed by Sammel and Ryan (1996). This model was generalized to a linear SEM by Lee and Shi (2000), and to NSEMs by Lee and Song (2003b), respectively. However, these papers did not consider missing data. Methods for the treatment of missing data that are missing at random (MAR) with an ignorable mechanism (Little and Rubin (1987)) have received much attention in SEMs without covariates; see for example Song and Lee (2002) and Lee and Song (2004a,b), among others. However, missing response data are often nonignorable in the sense that the reason for missingness depends on the missing values themselves. In light of this, the main purpose of this paper is to develop a Bayesian approach for analyzing NSEMs with covariates and mixed continuous and ordered categorical outcomes, in the presence of nonignorable missing data that can come from the observed variables as well as the covariates.

In the literature, there are important studies of maximum likelihood estimation with nonignorable missing observations for the normal random effects model (Laird and Ware (1982)), the conditional linear model (Follmann and Wu (1995)), and the generalized linear mixed model (Ibrahim, Chen and Lipsitz (2001)). Methods for treating missing covariates are given by Rathouz, Satten and Carroll (2002) for the semiparametric method in matched case-control studies, by Lipsitz, Ibrahim and Zhao (1999) and Parzen, Lipsitz, Ibrahim and Lipshultz (2002) for the weighted estimating equation method, by Ibrahim, Chen and Lipsitz (1999) for the Monte Carlo EM method in parametric regression models, and by Stubbendick and Ibrahim (2003) for the maximum likelihood method combing the Gibbs sampler and the MCEM algorithm. However, these methods cannot be applied to NSEMs with ordered categorical variables. In our development, we will employ a Bayesian approach for the following reasons: (i) it allows the use of genuine prior information for achieving better results, and (ii) as pointed out in Bayesian analyses of latent variable models and SEMs (Dunson (2000), Lee and Song (2004c) and Scheines, Hoijtink and Boomsma (1999)), sampling-based Bayesian methods do not depend on asymptotic theory, and hence give more reliable results with small sample sizes.

The rest of the paper is organized as follows. In Section 2, we formulate a NSEM with covariates, and describe the mixed continuous and ordered categorical data that can be missing with a nonignorable mechanism. The specification of the missingness mechanism is also discussed. In Section 3, we present the Bayesian approach, and derive novel conditional distributions for implementing the sampling-based methods. Numerical illustrations, which include a simulation study and an example, are presented in Section 4. Technical details are given in the appendices.

2. Model and Notation

Inspired by the LISREL model (Jöreskog and Sörbom (1996)), we propose an SEM with a measurement equation and a structural equation. The measurement equation is defined by the following factor analysis model with covariates:

$$\boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{c}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{y}_i$ is a p by 1 random vector of manifest variables, $\boldsymbol{c}_i$ is an r by 1 vector of covariates which may come from a continuous or discrete distribution, $\boldsymbol{\omega}_i$ is a q by 1 random vector of latent variables, $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ are matrices of unknown parameters, and $\boldsymbol{\epsilon}_i$ is a random vector of error measurements with distribution $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, in which $\boldsymbol{\Psi}_\epsilon$ is a diagonal matrix with diagonal elements $\psi_{\epsilon 1}, \ldots, \psi_{\epsilon p}$. It is assumed that for $i = 1, \ldots, n$, the $\boldsymbol{\omega}_i$ are independently distributed, the $\boldsymbol{\epsilon}_i$ are independently and identically distributed (i.i.d.), and the $\boldsymbol{\omega}_i$ and $\boldsymbol{\epsilon}_i$ are independent. The main purpose of this measurement equation is to identify the latent variables in $\boldsymbol{\omega}_i$ via the manifest variables (indicators) in $\boldsymbol{y}_i$. The covariates can be explanatory or other kind of variables that are helpful in achieving a better model for relating $\boldsymbol{y}_i$ with $\boldsymbol{\omega}_i$. Let $\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$ be a partition of $\boldsymbol{\omega}_i$ into endogenous latent variables in $\boldsymbol{\eta}_i$ (q_1 by 1) and exogenous latent variables in $\boldsymbol{\xi}_i$ (q_2 by 1). The following nonlinear structural equation is used to model the relationship between $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$:

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{\Gamma}\boldsymbol{H}(\boldsymbol{\xi}_i) + \boldsymbol{\zeta}_i, \tag{2}$$

where $\boldsymbol{x}_i$ is an s by 1 vector of covariates that can come from continuous or discrete distributions, $\boldsymbol{H}(\boldsymbol{\xi}_i) = (h_1(\boldsymbol{\xi}_i), \ldots, h_t(\boldsymbol{\xi}_i))^T$ is a vector-valued function containing non-zero differentiable functions $h_1, \ldots, h_t$, $t \geq q_2$, $\boldsymbol{\Pi}, \boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are matrices of unknown parameters, $\boldsymbol{\xi}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Phi}]$, $\boldsymbol{\zeta}_i$ is a vector of residuals with distribution $N[\mathbf{0}, \boldsymbol{\Psi}_\zeta]$, in which $\boldsymbol{\Psi}_\zeta$ is a diagonal matrix, and $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_i$ are independent. Similar to many SEMs, it is assumed that $\boldsymbol{\Pi}_0 = |\boldsymbol{I}_{q_1} - \boldsymbol{\Pi}|$ is nonzero and independent of any element of $\boldsymbol{\Pi}$. This condition is assumed so that the computational burden is reduced, it can be relaxed with appropriate modification. Again the covariates can be explanatory variables or other variables that are not involved in the measurement equation but have significance in explaining $\boldsymbol{\eta}$.

To account for the ordered categorical data, without loss of generality let $\boldsymbol{y} = (\boldsymbol{y}_{(1)}^T, \boldsymbol{y}_{(2)}^T)^T$, where $\boldsymbol{y}_{(1)} = (y_1, \ldots, y_k)^T$ is a subset of manifest variables that cannot be directly observed, whilst $\boldsymbol{y}_{(2)}$ can be directly observed. For $h = 1, \ldots, k$, the information of a variable y_h is given by an observable ordered categorical value z_h such that $z_h = j$ if $\alpha_{hj} < y_h \leq \alpha_{h,j+1}$, where j is in $\{0, 1, \ldots, b_h\}$. Let $\boldsymbol{\alpha}_h = (\alpha_{h1}, \ldots, \alpha_{hb_h})^T$ for $h = 1, \ldots, k$, and assume that $\alpha_{h0} = -\infty, \alpha_{h,b_h+1} = \infty$. In particular, if $b_h = 1$ and $\alpha_{h1} = 0$, the information of y_h is given by an observable dichotomous value z_h such that $z_h = 1$ if $y_h > 0$, and 0 otherwise. These ordered categorical variables can be identified by the method given by Shi and Lee (2000) and Lee and Song (2004d).

To deal with the missing data problem, let $\boldsymbol{y}_{(1)i} = \{\boldsymbol{y}_{(1)oi}, \boldsymbol{y}_{(1)mi}\}$, $\boldsymbol{y}_{(2)i} = \{\boldsymbol{y}_{(2)oi}, \boldsymbol{y}_{(2)mi}\}$, and $\boldsymbol{x}_i = \{\boldsymbol{x}_{oi}, \boldsymbol{x}_{mi}\}$, where $\boldsymbol{y}_{(2)oi}$ and $\boldsymbol{x}_{oi}$ denote the observed data, whilst $\boldsymbol{y}_{(1)mi}, \boldsymbol{y}_{(2)mi}$ and $\boldsymbol{x}_{mi}$ represent the missing data of $\boldsymbol{y}_{(1)i}, \boldsymbol{y}_{(2)i}$ and $\boldsymbol{x}_i$. For simplicity, we assume that $\boldsymbol{c}_i$ is fully observed for all $i = 1, \ldots, n$. This assumption can be relaxed with minor modification. Let $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{i,p+s})^T$ be a missing indicator vector for $\boldsymbol{v}_i = (\boldsymbol{y}_i^T, \boldsymbol{x}_i^T)^T$ such that r_{ij} equals 1 if v_{ij} is missing and 0 if v_{ij} is observed. Moreover, let $\boldsymbol{z}_{oi}$ be the observed ordered categorical data of $\boldsymbol{z}_i$ under $\boldsymbol{y}_{(1)i}$ and let $[\boldsymbol{r}_i | \boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}]$ be the conditional distribution of $\boldsymbol{r}_i$ given $\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}$, $\boldsymbol{x}_i$, $\boldsymbol{c}_i$, and $\boldsymbol{\omega}_i$, with a parameter vector $\boldsymbol{\varphi}$ and a density function $p(\boldsymbol{r}_i | \boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})$. The missing data mechanism is decided by this distribution. Let $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k\}$, and let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ in which $\boldsymbol{\theta}_1$ contains all unknown distinct parameters in $\boldsymbol{A}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon$ that are associated with the measurement equation, and $\boldsymbol{\theta}_2$ contains all unknown distinct parameters in $\boldsymbol{\Pi}, \boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\zeta$, and $\boldsymbol{\Phi}$ that are associated with the structural equation. Let $\boldsymbol{Y}_o = \{\boldsymbol{y}_{(2)oi} : i = 1, \ldots, n\}$, $\boldsymbol{Z}_o = \{\boldsymbol{z}_{oi} : i = 1, \ldots, n\}$, $\boldsymbol{Y}_m = \{(\boldsymbol{y}_{(1)mi}, \boldsymbol{y}_{(2)mi}) : i = 1, \ldots, n\}$, $\boldsymbol{Y}_{(1)o} = \{\boldsymbol{y}_{(1)oi} : i = 1, \ldots, n\}$, $\boldsymbol{Y} = \{\boldsymbol{y}_i : i = 1, \ldots, n\}$, $\boldsymbol{F} = \{\boldsymbol{\omega}_i : i = 1, \ldots, n\}$, $\boldsymbol{X}_o = \{\boldsymbol{x}_{oi} : i = 1, \ldots, n\}$, $\boldsymbol{X}_m = \{\boldsymbol{x}_{mi} : i = 1, \ldots, n\}$, $\boldsymbol{X} = \{\boldsymbol{x}_i : i = 1, \ldots, n\}$, and $\boldsymbol{C} = \{\boldsymbol{c}_i : i = 1, \ldots, n\}$. The main purpose of this paper is to develop Bayesian methods to analyze the proposed model on the basis of the missing data indicator $\boldsymbol{r} = \{\boldsymbol{r}_i : i = 1, \ldots, n\}$ and the observed data set $\{\boldsymbol{Y}_o, \boldsymbol{Z}_o, \boldsymbol{X}_o, \boldsymbol{C}\}$.

Theoretically, any general model can be taken to specify a model for $\boldsymbol{r}_i$. However, as pointed out by Ibrahim, Chen and Lipsitz (2001), one must be careful in using a complicated or large model, because it can easily become unidentifiable. Moreover, a complex model can induce difficulty in deriving the corresponding conditional distribution of the missing manifest given the observed data, and inefficient sampling from that conditional distribution. As the covariance matrix of the error measurement, $\boldsymbol{\epsilon}_i$, is diagonal, it follows from (1) that when $\boldsymbol{\omega}_i$ is given, the components of $\boldsymbol{y}_i$ are independent. Hence, for $j \neq l \in \{1, \ldots, p\}$, it is reasonable to assume that the conditional distributions of r_{ij} and r_{il} given $\boldsymbol{\omega}_i$ are independent. Moreover, we also assume that r_{ij} and r_{il} are independent for

$j \neq l \in \{p+1, \ldots, p+s\}$. Under this assumption, we consider the following nonignorable missingness mechanism (Ibrahim, Chen and Lipsitz (2001)):

$$p(\boldsymbol{r}_i|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}) = \prod_{j=1}^{p+s} \{\mathrm{pr}(r_{ij} = 1|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\}^{r_{ij}} \times \{1 - \mathrm{pr}(r_{ij} = 1|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\}^{1-r_{ij}},$$

where $\mathrm{pr}(r_{ij} = 1|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})$ is modelled by logistic regression models

$$\mathrm{logit}\{\mathrm{pr}(r_{ij} = 1|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\} = \boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi}, \quad j = 1, \ldots, p, \tag{3}$$

$$\mathrm{logit}\{\mathrm{pr}(r_{i,p+j} = 1|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\} = \boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi}, \quad j = 1, \ldots, s, \tag{4}$$

in which $\boldsymbol{m}_{yi}$ and $\boldsymbol{m}_{xi}$ are functions of $\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i$, and $\boldsymbol{\omega}_i$, $\boldsymbol{\varphi}_y$ and $\boldsymbol{\varphi}_x$ are vectors of regression coefficients, and $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_y^T, \boldsymbol{\varphi}_x^T)^T$.

3. Bayesian Analysis of the Model

Let $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\varphi})^T$ be the vector of nuisance parameters. The Bayesian analysis is focused on $\log p(\boldsymbol{\theta}, \boldsymbol{\vartheta}|\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}) \propto \log p(\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\vartheta}) + \log p(\boldsymbol{\theta}, \boldsymbol{\vartheta})$, where $p(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ is the prior density of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$, and $p(\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r} |\boldsymbol{\theta}, \boldsymbol{\vartheta})$ is the likelihood function. In the posterior analysis, the observed data $\{\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}\}$ and $\boldsymbol{r}$ are augmented with the missing quantities $\{\boldsymbol{Y}_m, \boldsymbol{Y}_{(1)o}, \boldsymbol{X}_m, \boldsymbol{F}\}$ to produce a complete-data set $\{\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}\}$. Therefore, the joint posterior distribution of interest is $[\boldsymbol{F}, \boldsymbol{Y}_m, \boldsymbol{Y}_{(1)o}, \boldsymbol{X}_m, \boldsymbol{\theta}, \boldsymbol{\vartheta}|\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}]$.

3.1. An Algorithm for Simulating Observations

The Gibbs sampler (Geman and Geman (1984)) is used to generate a sequence of random observations from the above joint posterior distribution. In this algorithm, observations $\{\boldsymbol{F}, \boldsymbol{Y}_m, \boldsymbol{Y}_{(1)o}, \boldsymbol{X}_m, \boldsymbol{\theta}, \boldsymbol{\vartheta}\}$ are iteratively sampled from the following conditional distributions: $p(\boldsymbol{F}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_o, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\varphi})$, $p(\boldsymbol{Y}_m|\boldsymbol{X}, \boldsymbol{Y}_{(1)o}, \boldsymbol{Y}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\varphi})$, $p(\boldsymbol{X}_m|\boldsymbol{Y}, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi})$, $p(\boldsymbol{\theta} |\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{F})$, $p(\boldsymbol{\tau}|\boldsymbol{X})$, $p(\boldsymbol{\varphi}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{Z}_o, \boldsymbol{r})$ and $p(\boldsymbol{\alpha}, \boldsymbol{Y}_{(1)o}|\boldsymbol{C}, \boldsymbol{F}, \boldsymbol{Z}_o, \boldsymbol{\theta})$. As the observations are independent, we only need to derive the first three conditional distributions for each i. Note that once $\boldsymbol{y}_{(1)mi}$ is given, it is not necessary to simulate $\boldsymbol{z}_{mi}$. Thus, $\{\boldsymbol{z}_{mi} : i = 1, \ldots, n\}$ is not involved in the Gibbs sampler. The full conditional distributions that are required in the implementation of the Gibbs sampler are briefly discussed here.

For $p(\boldsymbol{F}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_o, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\varphi})$, it can be shown that

$$p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{z}_{oi}, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) \propto p(\boldsymbol{y}_i|\boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\theta}_1) p(\boldsymbol{\eta}_i|\boldsymbol{x}_i, \boldsymbol{\xi}_i, \boldsymbol{\theta}_2) p(\boldsymbol{\xi}_i|\boldsymbol{\theta}_2) p(\boldsymbol{r}_i|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}).$$

Let $\boldsymbol{\Lambda}_\omega = (\boldsymbol{\Pi}, \boldsymbol{B}, \boldsymbol{\Gamma})$ and $\boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\omega}_i) = (\boldsymbol{\eta}_i^T, \boldsymbol{x}_i^T, \boldsymbol{H}(\boldsymbol{\xi}_i)^T)^T$. Then, $p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})$ is proportional to

$$\begin{aligned}\exp\Big\{ & -\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{A}\boldsymbol{c}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\epsilon^{-1}(\boldsymbol{y}_i - \boldsymbol{A}\boldsymbol{c}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \\ & -\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\omega}_i))^T \boldsymbol{\Psi}_\zeta^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\omega}_i)) \\ & -\frac{1}{2}\boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1}\boldsymbol{\xi}_i + \sum_{j=1}^{p}(r_{ij}\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi} - \log(1 + \mathrm{e}^{\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi}})) \\ & + \sum_{j=1}^{s}(r_{i,p+j}\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi} - \log(1 + \mathrm{e}^{\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi}}))\Big\}. \end{aligned} \tag{5}$$

Consider the conditional distribution of $\boldsymbol{Y}_m$ given $(\boldsymbol{X}, \boldsymbol{Y}_{(1)o}, \boldsymbol{Y}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r})$. As $\boldsymbol{\Psi}_\epsilon$ is diagonal, $\boldsymbol{y}_{(1)mi}$ is independent of $\boldsymbol{y}_{(2)mi}$ and $\boldsymbol{y}_{(1)oi}$, and $\boldsymbol{y}_{(2)mi}$ is independent of $\boldsymbol{y}_{(2)oi}$. Hence, $p(\boldsymbol{y}_{(1)mi}, \boldsymbol{y}_{(2)mi}|\boldsymbol{y}_{(1)oi}, \boldsymbol{y}_{(2)oi}, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}) \propto p(\boldsymbol{y}_{(1)mi}|\boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\theta}_1)p(\boldsymbol{y}_{(2)mi}|\boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\theta}_1)p(\boldsymbol{r}_i \ |\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})$. According to the definition of the models for $\boldsymbol{y}_i$ and $\boldsymbol{r}_i$, we have

$$[\boldsymbol{y}_{(1)mi}|\boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\theta}_1] \stackrel{D}{=} N[\boldsymbol{A}_{(1)mi}\boldsymbol{c}_i + \boldsymbol{\Lambda}_{(1)mi}\boldsymbol{\omega}_i, \boldsymbol{\Psi}_{\epsilon(1)mi}],$$

where $\boldsymbol{A}_{(1)mi}$ and $\boldsymbol{\Lambda}_{(1)mi}$ are the submatrices of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ with rows corresponding to $\boldsymbol{y}_{(1)mi}$, respectively, and $\boldsymbol{\Psi}_{\epsilon(1)mi}$ is the submatrix of $\boldsymbol{\Psi}_\epsilon$ with rows and columns corresponding to $\boldsymbol{y}_{(1)mi}$. Moreover, $p(\boldsymbol{y}_{(2)mi}|\boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})$ is proportional to

$$\begin{aligned}\exp\{ & -\frac{1}{2}(\boldsymbol{y}_{(2)mi} - \boldsymbol{A}_{(2)mi}\boldsymbol{c}_i - \boldsymbol{\Lambda}_{(2)mi}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_{\epsilon(2)mi}^{-1}(\boldsymbol{y}_{(2)mi} - \boldsymbol{A}_{(2)mi}\boldsymbol{c}_i - \boldsymbol{\Lambda}_{(2)mi}\boldsymbol{\omega}_i) \\ & + \sum_{j=1}^{p}(r_{ij}\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi} - \log(1 + \exp(\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi}))) \\ & + \sum_{j=1}^{s}(r_{i,p+j}\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi} - \log(1 + \exp(\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi})))\}, \end{aligned} \tag{6}$$

where $\boldsymbol{A}_{(2)mi}$ and $\boldsymbol{\Lambda}_{(2)mi}$ are the submatrices of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ with rows corresponding to $\boldsymbol{y}_{(2)mi}$, and $\boldsymbol{\Psi}_{\epsilon(2)mi}$ is the submatrix of $\boldsymbol{\Psi}_\epsilon$ with rows and columns corresponding to $\boldsymbol{y}_{(2)mi}$.

For $p(\boldsymbol{X}_m|\boldsymbol{Y}, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi})$, it can be shown that

$$\begin{aligned} & p(\boldsymbol{x}_{mi}|\boldsymbol{y}_i, \boldsymbol{x}_{oi}, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi}) \\ & \quad \propto p(\boldsymbol{\eta}_i|\boldsymbol{x}_i, \boldsymbol{\xi}_i, \boldsymbol{\theta}_2)p(\boldsymbol{r}_i|\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})p(\boldsymbol{x}_{mi}|\boldsymbol{\tau}). \end{aligned}$$

Then, it follows from (3) and (4) that $p(\boldsymbol{x}_{mi}|\boldsymbol{y}_i, \boldsymbol{x}_{oi}, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi})$ is proportional to

$$\begin{aligned}\exp\Big\{&-\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\xi}_i))^T \boldsymbol{\Psi}_\zeta^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\xi}_i))\\ &+\sum_{j=1}^{p}(r_{ij}\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi} - \log(1+\exp(\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi})))\\ &+\sum_{j=1}^{s}(r_{i,p+j}\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi} - \log(1+\exp(\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi})))\Big\} p(\boldsymbol{x}_{mi}|\boldsymbol{\tau}). \end{aligned} \tag{7}$$

Note that the distribution of the missing covariates is involved.

Let $p(\boldsymbol{\varphi}_y)$ be the prior density of $\boldsymbol{\varphi}_y$ with distribution $N[\boldsymbol{\varphi}_{0y}, \boldsymbol{V}_y]$, where $\boldsymbol{\varphi}_{0y}$ and $\boldsymbol{V}_y$ are the given hyper-parameters. Under this prior distribution and the fact that the distribution of r_{ij} only involves $\boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{\omega}_i$, and $\boldsymbol{\varphi}_y$ for $j \in \{1, \ldots, p\}$, it follows from the independence of r_{ij} and r_{ih}, and (3), that

$$\begin{aligned}&p(\boldsymbol{\varphi}_y|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_o, \boldsymbol{F}, \boldsymbol{r})\\ &\propto \frac{\exp\{\sum_{i=1}^{n}(\sum_{j=1}^{p} r_{ij})\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi} - \frac{1}{2}(\boldsymbol{\varphi}_y - \boldsymbol{\varphi}_{0y})^T \boldsymbol{V}_y^{-1}(\boldsymbol{\varphi}_y - \boldsymbol{\varphi}_{0y})\}}{\prod_{i=1}^{n}(1+\exp(\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi}))^p}.\end{aligned} \tag{8}$$

Similarly, let $p(\boldsymbol{\varphi}_x)$ be the prior density of $\boldsymbol{\varphi}_x$ with distribution $N[\boldsymbol{\varphi}_{0x}, \boldsymbol{V}_x]$,where $\boldsymbol{\varphi}_{0x}$ and $\boldsymbol{V}_x$ are the given hyper-parameters. Similarly, it follows from (4) that

$$\begin{aligned}&p(\boldsymbol{\varphi}_x|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_o, \boldsymbol{F}, \boldsymbol{r})\\ &\propto \frac{\exp\{\sum_{i=1}^{n}(\sum_{j=1}^{s} r_{i,p+j})\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi} - \frac{1}{2}(\boldsymbol{\varphi}_x - \boldsymbol{\varphi}_{0x})^T \boldsymbol{V}_x^{-1}(\boldsymbol{\varphi}_x - \boldsymbol{\varphi}_{0x})\}}{\prod_{i=1}^{n}(1+\exp(\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi}))^s}.\end{aligned} \tag{9}$$

Given $\boldsymbol{Y}_m$, the model defined in (1) becomes the model discussed by Lee and Zhu (2000). Hence, the conditional distribution of $(\boldsymbol{\alpha}, \boldsymbol{Y}_{(1)o})$ given $(\boldsymbol{C}, \boldsymbol{Z}_o, \boldsymbol{F}, \boldsymbol{\theta})$ can be obtained by similar derivations as given in Lee and Zhu (2000). Consider the following conjugate prior distributions for components in $\boldsymbol{\theta}$: $\psi_{\epsilon k}^{-1} \stackrel{D}{=} Gamma(\alpha_{0Ak}, \beta_{0Ak})$, $\boldsymbol{\Lambda}_{Ak} \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0Ak}, \psi_{\epsilon k}\boldsymbol{H}_{0Ak}]$, $\psi_{\zeta k}^{-1} \stackrel{D}{=} (\alpha_{0\zeta k}, \beta_{0\zeta k})$, $\boldsymbol{\Lambda}_{0\omega k} \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0\omega k}, \psi_{\zeta k}\boldsymbol{H}_{0\omega k}]$, and $\boldsymbol{\Phi} \stackrel{D}{=} IW[\boldsymbol{R}_0, \rho_0]$, where $\boldsymbol{\Lambda}_{Ak}$ is the kth row of $(\boldsymbol{A}, \boldsymbol{\Lambda})$, $\boldsymbol{\Lambda}_{\omega k}$ is the kth row of $(\boldsymbol{B}, \boldsymbol{\Gamma})$, $\psi_{\epsilon k}$ and $\psi_{\zeta k}$ are the kth diagonal elements of $\boldsymbol{\Psi}_\epsilon$ and $\boldsymbol{\Psi}_\zeta$, respectively, IW denotes the inverted Wishart distribution, and quantities with a subscript '0' are the given hyperparameter values. Under these prior distributions, the conditional distribution of $\boldsymbol{\theta}$ given $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{F})$ can also be obtained as by Lee and Zhu (2000). To save space, the related discussion is not presented. The implementation of the Metropolis-Hastings (MH) algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) and Hastings (1970)) for simulating observations from the complex conditional distributions in (5)−(9) is presented in Appendix I.

3.2. Bayesian model comparison

As pointed out by Lipsitz and Ibrahim (1996), the parametric form of the assumed missingness mechanism itself is not 'testable'. However, the Bayes factor (Berger (1985)) can be used to compare competing models M_0 and M_1, which may have different missingness mechanisms or model structures. The Bayes factor for evaluating M_1 against M_0 is defined as

$$B_{10} = \frac{p(\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}|M_1)}{p(\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}|M_0)},$$

where $p(\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}|M_k)$ is the marginal density of $(\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r})$ under M_k. As this marginal density involves an intractable multiple integral, the computation of the Bayes factor is difficult. Based on the comparative study of DiCiccio, Kass, Raftery and Wasserman (1997), bridge sampling (Meng and Wong (1996)) is an attractive method. Gelman and Meng (1998) developed path sampling, which is a generalization of bridge sampling, and argued that it has the potential to be even better. Based on the nice features presented by Gelman and Meng (1998), and inspired by many successful applications to various SEMs (see, e.g., Lee and Song (2003b)), path sampling is used to compute the Bayes factor for model comparison in the current problem. For completeness, the description of the path sampling is given in Appendix II.

4. Numerical Examples

Numerical results are presented to illustrate various aspects of the proposed Bayesian methods. In the simulation studies, two covariates are independently generated from a Binomial distribution and a normal distribution. In our example, the covariates are continuous and ordered categorical outcomes with an underlying bivariate normal distribution $N[\mathbf{0}, \boldsymbol{\Phi}_x]$, where $\boldsymbol{\Phi}_x$ is an unknown matrix. For completeness, the relating conditional distributions (see (7)) under these situations are presented in Appendix III.

4.1. Simulation Studies

The main purpose is to illustrate the influence of the missingness mechanism, and to study the sensitivity of the Bayesian estimates with respect to prior inputs and missing covariate distributions. Complete-data sets $\{\boldsymbol{y}_i, i = 1, \ldots, 500\}$

are generated from an NSEM defined in (1) and (2) with nine manifest variables that are related to three basic latent factors $\boldsymbol{\eta}_i = \eta_i$, and $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2})^T$. The covariate in the measurement equation is taken to be $\boldsymbol{c}_i = (1, \ldots, 1)^T$. For the structural equation, x_{i1} is generated from a Binomial distribution, $Bi(1, \tau_1)$, and x_{i2} is independently generated from a normal distribution $N(\tau_2, 1.0)$. Variables y_{i1} and y_{i2} are transformed to ordered categorical observations z_{i1} and z_{i2} with the same thresholds $\boldsymbol{\alpha}_k = (-1.2^*, -0.5, 0.5, 1.2^*)$ for $k = 1, 2$, where parameters with asterisks are treated as being fixed for identification. Variables y_{i3} and y_{i4} are transformed to dichotomous observations z_{i3} and z_{i4} with the fixed threshold 0.0^*, and $\psi_{\epsilon 3}$ and $\psi_{\epsilon 4}$ are fixed at 1.0^* for identification. The specifications of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ in relation to the measurement equation are: $\boldsymbol{A} = \text{diag}(0.0^*, 0.0^*, 0.0^*, 0.0^*, \mu_5, \ldots, \mu_9)$, and

$$\boldsymbol{\Lambda}^T = \begin{bmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{bmatrix}, \ \boldsymbol{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{bmatrix};$$

recall that the values 1.0^* and 0.0^* with asterisks are treated as fixed for achieving an identified model. The structural equation is defined by $\eta_i = b_1 x_{i1} + b_2 x_{i2} + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1} \xi_{i2} + \zeta_i$. True population values of the unknown parameters are given by $\mu_5 = \cdots = \mu_9 = 0.36, \lambda_{21} = \lambda_{31} = \lambda_{52} = \lambda_{62} = \lambda_{83} = \lambda_{93} = 0.36$, $\psi_{\epsilon 1} = \psi_{\epsilon 2} = \psi_{\epsilon 5} = \cdots = \psi_{\epsilon 9} = 0.36$, $b_1 = b_2 = 0.36, \gamma_1 = \gamma_2 = \gamma_3 = 0.36$, $\psi_\zeta = 0.36$, $(\phi_{11}, \phi_{12}, \phi_{22}) = (1.0, 0.36, 1.0)$, and $\tau_1 = \tau_2 = 0.5$.

Missing data of $\boldsymbol{y}_{(2)i} = (y_{i5}, \ldots, y_{i9})$ are generated from the logistic regression model,

$$\begin{aligned} &\text{logit}\{\text{pr}(r_{ij} = 1 | \boldsymbol{y}_{(2)i}, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\} \\ &\quad = \varphi_{y0} + \varphi_{y1} z_{i1} + \cdots + \varphi_{y4} z_{i4} + \varphi_{y5} y_{i5} + \cdots \\ &\qquad + \varphi_{y9} y_{i9} + \varphi_{y10} x_{i1} + \varphi_{y11} x_{i2} + \varphi_{y12} \xi_{i1} + \varphi_{y13} \xi_{i2}, \end{aligned} \tag{10}$$

with true parameters $\varphi_{y0} = -2.0$, and $\varphi_{y1} = \cdots = \varphi_{y13} = 0.1$; the missing data for (x_{i1}, x_{i2}) are generated from the logistic regression model,

$$\text{logit}\{\text{pr}(r_{i,9+j} = 1 | \boldsymbol{x}_i, \boldsymbol{\varphi})\} = \varphi_{x0} + \varphi_{x1} x_{i1} + \varphi_{x2} x_{i2}, \quad j = 1, 2, \tag{11}$$

with true parameter values $\varphi_{x0} = -2.0$ and $\varphi_{x1} = \varphi_{x2} = 0.1$. There are 50 unknown parameters in the full model. The average proportions of missing data corresponding to the y's and x's are about 0.18 and 0.13, respectively.

To investigate the sensitivity of Bayesian estimates to prior inputs, the following hyper-parameters are considered. Type I: the hyper-parameters for $\boldsymbol{\Lambda}_{0Ak} = (\boldsymbol{A}_{0k}^T, \boldsymbol{\Lambda}_{0k}^T)^T, \boldsymbol{\Lambda}_{0\omega k} = (\boldsymbol{B}_{0k}^T, \boldsymbol{\Gamma}_{0k}^T)^T$, $\boldsymbol{\varphi}_{0y}$ and $\boldsymbol{\varphi}_{0x}$ are taken to be their corresponding true values, $\alpha_{0Ak} = \alpha_{0\zeta k} = 10$, $\beta_{0Ak} = \beta_{0\zeta k} = 4$ and $\boldsymbol{H}_{0Ak}$ and

$\boldsymbol{H}_{0\omega k}$ are diagonal matrices with diagonal elements 0.25, $\rho_0 = 8$ and $\boldsymbol{R}_0 = 5\boldsymbol{\Phi}_0$, where $\boldsymbol{\Phi}_0$ is the true value of $\boldsymbol{\Phi}$. This can be regarded as a situation with good prior information. Type II: non-informative priors. The results of all of the simulation studies are based on 100 replications. In Tables 1 to 3, 'Bias' denotes the difference between the true value and the mean of the estimates, and 'RMS' is the root mean square between the estimates and its true value. The results obtained from this simulation are reported in Table 1. It can be seen that the Bayesian estimates obtained are reasonably accurate under different prior inputs, and not sensitive to prior inputs.

Table 1. Performance of the Bayesian estimates in the simulation study.

	Type I		Type II			Type I		Type II	
Par.	Bias	RMS	Bias	RMS	Par.	Bias	RMS	Bias	RMS
u_5	0.003	0.030	-0.004	0.032	α_{12}	-0.020	0.056	-0.006	0.057
u_6	-0.001	0.035	-0.003	0.031	α_{13}	-0.017	0.050	-0.010	0.055
u_7	-0.018	0.056	-0.013	0.058	α_{22}	0.003	0.047	-0.001	0.044
u_8	-0.003	0.035	-0.004	0.033	α_{23}	-0.007	0.036	-0.003	0.041
u_9	-0.007	0.035	-0.008	0.036	φ_{y0}	-0.097	0.173	-0.112	0.273
λ_{21}	-0.004	0.039	0.004	0.055	φ_{y1}	0.022	0.060	0.022	0.074
λ_{31}	0.019	0.072	0.007	0.093	φ_{y2}	0.003	0.070	0.024	0.091
λ_{52}	0.037	0.072	0.012	0.077	φ_{y3}	-0.036	0.058	-0.030	0.052
λ_{62}	0.059	0.093	0.013	0.078	φ_{y4}	-0.023	0.054	-0.029	0.069
λ_{83}	0.015	0.046	-0.010	0.060	φ_{y5}	0.044	0.150	0.027	0.152
λ_{93}	0.016	0.052	-0.010	0.062	φ_{y6}	0.039	0.116	0.046	0.184
ψ_1	-0.004	0.051	-0.003	0.140	φ_{y7}	0.032	0.140	0.029	0.366
ψ_2	-0.002	0.030	-0.001	0.035	φ_{y8}	0.025	0.116	0.019	0.153
ψ_5	0.002	0.030	-0.002	0.036	φ_{y9}	0.012	0.111	0.035	0.160
ψ_6	-0.012	0.032	-0.003	0.031	φ_{y10}	0.010	0.109	-0.002	0.124
ψ_7	0.035	0.066	-0.056	0.171	φ_{y11}	-0.023	0.074	-0.018	0.070
ψ_8	0.000	0.028	0.007	0.033	φ_{y12}	-0.021	0.202	-0.024	0.288
ψ_9	-0.005	0.029	0.001	0.032	φ_{y13}	-0.035	0.207	-0.069	0.497
b_1	0.020	0.070	0.034	0.088	φ_{x0}	-0.009	0.122	-0.026	0.151
b_2	-0.084	0.095	-0.089	0.100	φ_{x1}	0.008	0.216	0.031	0.301
γ_1	0.027	0.079	0.007	0.091	φ_{x2}	-0.095	0.212	-0.111	0.269
γ_2	0.009	0.062	-0.014	0.075					
γ_3	0.056	0.105	0.024	0.108					
ψ_ζ	0.040	0.068	0.046	0.142					
ϕ_{11}	-0.063	0.215	0.059	0.350					
ϕ_{12}	0.000	0.077	-0.005	0.090					
ϕ_{22}	-0.038	0.122	-0.084	0.204					
τ_1	0.002	0.024	0.000	0.027					
τ_2	-0.027	0.071	-0.028	0.084					
ΣRMS		1.819		2.452	ΣRMS		2.379		3.471

The following simulation study is used to illustrate the sensitivity of Bayesian estimates to the missingness mechanism. Complete data sets are simulated on the basis of the NSEM with the above settings. Missing data are created as follows.

Type A. Nonignorable missingness mechanisms that are different from (10) and (11):

$$\begin{aligned}&\text{logit}\{\text{pr}(r_{ij}=1|\boldsymbol{y}_{(2)i},\boldsymbol{x}_i,\boldsymbol{z}_{oi},\boldsymbol{c}_i,\boldsymbol{\omega}_i,\boldsymbol{\varphi})\}\\&\quad=\varphi_{y0}+\varphi_{y1}z_{i1}+\cdots+\varphi_{y4}z_{i4}+\varphi_{y5}y_{i5}+\cdots+\varphi_{y9}y_{i9}\\&\qquad+\varphi_{y10}x_{i1}+\varphi_{y11}x_{i2}+\varphi_{y12}\xi_{i1}+\varphi_{y13}\xi_{i2}+\varphi_{y14}x_{i1}^2+\varphi_{y15}x_{i2}^2+\varphi_{y16}x_{i1}x_{i2},\\&\text{logit}\{\text{pr}(r_{i,9+j}=1|\boldsymbol{x}_i,\boldsymbol{\varphi})\}\\&\quad=\varphi_{x0}+\varphi_{x1}x_{i1}+\varphi_{x2}x_{i2}+\varphi_{x3}x_{i1}^2+\varphi_{x4}x_{i2}^2+\varphi_{x5}x_{i1}x_{i2},\end{aligned}$$

with $\varphi_{y0}=\varphi_{x0}=-2.0$, $\varphi_{y1}=\cdots=\varphi_{y16}=\varphi_{x1}=\cdots=\varphi_{x5}=0.1$.

Type B. Logistic regression models that are different from (10) and (11):

$$\begin{aligned}&\text{logit}\{\text{pr}(r_{ij}=1|\boldsymbol{y}_{(2)i},\boldsymbol{x}_i,\boldsymbol{z}_{oi},\boldsymbol{c}_i,\boldsymbol{\omega}_i,\boldsymbol{\varphi})\}\\&\quad=\varphi_{y0}+\varphi_{y1}z_{i1}+\cdots+\varphi_{y4}z_{i4}+\varphi_{y5}y_{i5}+\cdots+\varphi_{y9}y_{i9}+\varphi_{y10}x_{i1}+\varphi_{y11}x_{i2}\\&\qquad+\varphi_{y12}\xi_{i1}\xi_{i2}\\&\text{logit}\{\text{pr}(r_{i,9+j}=1|\boldsymbol{x}_i,\boldsymbol{\varphi})\}\\&\quad=\varphi_{x0}+\varphi_{x1}x_{i1}+\varphi_{x2}x_{i2}+\varphi_{x3}y_{i6}+\varphi_{x4}y_{i7}y_{i8}+\varphi_{x5}\xi_{i1}+\varphi_{x6}\xi_{i2}\end{aligned}$$

with $\varphi_{y0}=\varphi_{x0}=-2.0,\varphi_{y1}=\cdots=\varphi_{y12}=\varphi_{x1}=\cdots=\varphi_{x6}=0.1$.

Type C. MAR missingness mechanism.

Type D. Logistic regression models given in (10) and (11).

All estimates are obtained with Type I prior inputs. For the first three missingness mechanisms, estimates are computed via the incorrect models (10) and (11). For Type D, estimates are obtained under the incorrect MAR assumption. Results are reported in Table 2. From the columns under Types A, B and C, we observe that even when the true missingness mechanism models are more complicated or the true missing data are MAR, the estimates obtained using the models defined by (10) and (11) are quite accurate. In contrast, it can be seen from the column under Type D that the estimates obtained under the incorrect MAR assumption are inaccurate. Hence, it seems that the results obtained by the proposed logistic regression model are robust to the different choices of the missingness mechanism, but it is important to take the nonignorable missingness mechanism into account.

Table 2. Bayesian estimates under different missingness mechanisms.

		Type A		Type B		Type C		Type D	
Par.	true	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
u_5	0.36	-0.003	0.035	-0.006	0.036	0.001	0.026	-0.245	0.264
u_6	0.36	-0.006	0.034	-0.002	0.036	0.002	0.029	-0.263	0.279
u_7	0.36	-0.016	0.061	-0.017	0.057	0.009	0.050	-0.660	0.669
u_8	0.36	-0.010	0.036	-0.007	0.040	0.005	0.037	-0.288	0.294
u_9	0.36	-0.005	0.039	-0.002	0.033	0.002	0.047	-0.292	0.303
λ_{21}	0.36	0.007	0.040	0.003	0.044	0.002	0.039	-0.044	0.055
λ_{31}	0.36	0.008	0.070	0.010	0.076	0.002	0.074	-0.034	0.075
λ_{52}	0.36	0.017	0.067	-0.007	0.060	0.039	0.072	-0.036	0.182
λ_{62}	0.36	0.018	0.067	-0.004	0.068	0.035	0.067	-0.040	0.155
λ_{83}	0.36	0.007	0.040	0.006	0.044	0.008	0.044	-0.029	0.111
λ_{93}	0.36	0.003	0.039	0.011	0.049	0.010	0.049	-0.013	0.107
ψ_1	0.36	0.009	0.054	0.003	0.053	0.067	0.094	-0.097	0.106
ψ_2	0.36	-0.002	0.027	0.003	0.032	0.003	0.033	0.007	0.037
ψ_5	0.36	-0.006	0.033	0.000	0.027	-0.006	0.030	-0.018	0.058
ψ_6	0.36	-0.003	0.034	-0.004	0.032	-0.001	0.029	-0.009	0.049
ψ_7	0.36	0.015	0.057	0.023	0.060	0.018	0.070	-0.037	0.078
ψ_8	0.36	-0.001	0.026	-0.004	0.032	-0.004	0.030	-0.009	0.047
ψ_9	0.36	-0.008	0.026	0.002	0.028	-0.002	0.033	-0.017	0.046
b_1	0.36	0.069	0.102	0.054	0.089	-0.042	0.075	-0.137	0.143
b_2	0.36	-0.119	0.127	-0.088	0.099	0.084	0.095	0.454	0.457
γ_1	0.36	-0.009	0.079	-0.008	0.085	0.066	0.106	0.047	0.113
γ_2	0.36	-0.005	0.059	0.005	0.064	0.021	0.084	-0.088	0.138
γ_3	0.36	0.014	0.081	0.014	0.088	0.070	0.119	0.271	0.288
ψ_ζ	0.36	0.056	0.078	0.053	0.075	-0.054	0.158	-0.256	0.256
ϕ_{11}	1.00	0.026	0.218	0.069	0.291	-0.120	0.219	-0.170	0.259
ϕ_{12}	0.36	0.006	0.088	-0.007	0.072	-0.031	0.082	-0.016	0.136
ϕ_{22}	1.00	0.007	0.105	-0.035	0.113	-0.021	0.113	-0.188	0.246
α_{12}	-0.50	-0.014	0.063	-0.011	0.056	0.010	0.061	-0.062	0.084
α_{13}	0.50	-0.006	0.044	-0.011	0.052	0.001	0.048	-0.048	0.071
α_{22}	-0.50	-0.010	0.048	-0.007	0.046	0.001	0.044	-0.023	0.049
α_{23}	0.50	-0.003	0.041	-0.005	0.038	-0.000	0.038	-0.024	0.046
τ_1	0.50	0.003	0.029	-0.002	0.024	-0.005	0.027	-0.146	0.156
τ_2	0.50	-0.060	0.086	-0.029	0.070	0.003	0.074	-0.302	0.314
ΣRMS			2.069		1.966		2.196		5.701

The following simulation study investigates the sensitivity of the results with respect to the choice of the missing covariate distributions. The true distributions of the covariates are given as follows. Type E: $x_{i1} \sim Bi(1, 0.5)$ and $x_{i2}|x_{i1} \sim N(0.5 + x_{i1}, 1)$. Type F: $x_{i2} \sim N(0.5, 1)$ and $x_{i1}|x_{i2} \sim Bi(1, p_x)$, where

$\log(p_x/(1-p_x)) = 0.5 - x_{i2}$. Missing data are generated via (10) and (11). Bayesian estimates are obtained via the correct missingness mechanism models (10) and (11), but under the incorrect distributional assumption of the covariates as stated at the beginning of this section. The results are reported in Table 3. It seems that the Bayesian estimates are not sensitive to the mis-specification of the missing covariate distributions.

Table 3. Bayesian estimates obtained under different missing covariate distributions.

		Type E		Type F	
Par.	true	Bias	RMS	Bias	RMS
u_5	0.36	-0.001	0.037	-0.004	0.034
u_6	0.36	-0.005	0.031	0.000	0.030
u_7	0.36	-0.005	0.051	-0.020	0.058
u_8	0.36	-0.000	0.033	-0.007	0.033
u_9	0.36	-0.003	0.033	-0.006	0.033
λ_{21}	0.36	0.004	0.037	0.006	0.040
λ_{31}	0.36	0.009	0.056	0.011	0.071
λ_{52}	0.36	0.026	0.071	0.043	0.073
λ_{62}	0.36	0.029	0.075	0.036	0.074
λ_{83}	0.36	0.014	0.045	0.022	0.051
λ_{93}	0.36	0.011	0.048	0.007	0.045
ψ_1	0.36	-0.011	0.048	0.009	0.049
ψ_2	0.36	-0.003	0.033	-0.003	0.030
ψ_5	0.36	0.003	0.029	-0.002	0.029
ψ_6	0.36	-0.005	0.031	-0.002	0.032
ψ_7	0.36	0.028	0.066	0.035	0.072
ψ_8	0.36	-0.011	0.031	-0.004	0.031
ψ_9	0.36	-0.002	0.027	-0.007	0.024
b_1	0.36	0.102	0.131	-0.031	0.068
b_2	0.36	-0.073	0.082	-0.087	0.096
γ_1	0.36	0.020	0.094	0.037	0.089
γ_2	0.36	-0.001	0.076	0.003	0.064
γ_3	0.36	0.028	0.090	0.064	0.110
ψ_ζ	0.36	0.042	0.063	0.037	0.060
ϕ_{11}	1.00	-0.042	0.241	-0.131	0.264
ϕ_{12}	0.36	-0.011	0.075	-0.022	0.081
ϕ_{22}	1.00	-0.029	0.115	-0.037	0.114
α_{12}	-0.50	-0.012	0.062	-0.023	0.062
α_{13}	0.50	-0.011	0.049	-0.017	0.052
α_{22}	-0.50	-0.002	0.051	-0.001	0.045
α_{23}	0.50	-0.005	0.039	-0.005	0.038
ΣRMS			1.950		1.952

To illustrate the path sampling procedure for comparing various logistic regression models in relation to missing data y_{ij}, we consider the following different forms of logistic models:

$$M_{y0}: m_{y0}(\boldsymbol{y}_{(2)i}, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}_y) = \varphi_{y0} + \varphi_{y1} z_{i1} + \cdots + \varphi_{y4} z_{i4} + \varphi_{y5} y_{i5} + \cdots + \varphi_{y9} y_{i9},$$

$$M_{y1}: m_{y1}(\boldsymbol{y}_{(2)i}, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}_y) = \varphi_{y0} + \varphi_{y1} z_{i1} + \cdots + \varphi_{y4} z_{i4} + \varphi_{y5} y_{i5} + \cdots + \varphi_{y9} y_{i9} + \varphi_{y10} x_{i1} + \varphi_{y11} x_{i2},$$

$$M_{y2}: m_{y2}(\boldsymbol{y}_{(2)i}, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}_y) = \varphi_{y0} + \varphi_{y1} z_{i1} + \cdots + \varphi_{y4} z_{i4} + \varphi_{y5} y_{i5} + \cdots + \varphi_{y9} y_{i9} + \varphi_{y10} x_{i1} + \varphi_{y11} x_{i2} + \varphi_{y12} \xi_{i1} + \varphi_{y13} \xi_{i2}.$$

We simulate the complete data sets on the basis of the same NSEM with the specifications stated previously, and take M_{y0} and (11) as the true missingness mechanism model for creating nonignorable missing data. Defining a path $t \in [0, 1]$ to link any two of the above models (see Appendix II) is straightforward. For example, M_{y0} and M_{y1} can be linked by

$$M_{ty01}: m_{y01}(\boldsymbol{y}_{(2)i}, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}_y) = \varphi_{y0} + \varphi_{y1} z_{i1} + \cdots + \varphi_{y4} z_{i4} + \varphi_{y5} y_{i5} + \cdots + \varphi_{y9} y_{i9} + t\varphi_{y10} x_{i1} + t\varphi_{y11} x_{i2}.$$

Clearly, M_{ty01} is equal to M_{y0} or M_{y1} when $t = 0$ or 1. The logarithm Bayes factors computed by the path sampling procedure via different hyperparameters as in Type I and Type II, are $\widehat{\log B_{y10}} = -1.998$ and -5.975, and $\widehat{\log B_{y20}} = -4.371$ and -12.01, respectively. Based on the criterion given by Kass and Raftery (1995), the true model M_{y0} is selected.

To illustrate the model comparison of various SEMs, we compare models that have the same measurement equation but have the following different structural equations:

$$M_{\eta 0}: \eta_i = b_1 x_{i1} + b_2 x_{i2} + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1} \xi_{i2} + \zeta_i,$$

$$M_{\eta 1}: \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1} \xi_{i2} + \zeta_i.$$

Complete data sets are simulated on the basis of $M_{\eta 0}$, and missing data are created via (10) and (11). The $\widehat{\log B_{\eta 10}}$ obtained with Type I and Type II priors are -146.62 and -41.5881, respectively. This gives strong evidence that we should select the true model $M_{\eta 0}$. The above results illustrate the reliability of the path sampling procedure.

4.2. An example

We analyze a portion of the data set that was obtained from a study of the relationship between acquired immune deficiency syndrome (AIDS) and the use of condoms (Morisky, Tiglao, Sneed, Tempongko, Baltazar, Detels and Stein (1998)). The data were obtained from female sex workers in Philippine cities, on items about knowledge of AIDS and attitude towards AIDS, belief, self efficiency of condom use, etc. Items 33, 32, 31, 43, 72, 74, 27h, 27e and 27i in the questionnaire (see Appendix IV) are taken as manifest variables in $\boldsymbol{y} = (y_1, \ldots, y_9)$. The first and last three items are ordered categorical variables with 5-point scales, and the remaining items are continuous. Moreover, a continuous item x_1 (item 37, see Appendix IV) and an ordered categorical item x_2 (item 21) are taken as covariates. Similarly, we consider x_2 as the observed ordered categorical value of an underlying continuous variable v, and assume that the joint bivariate distribution of $(x_1, v)'$ is $N[\mathbf{0}, \boldsymbol{\Phi}_x]$. There are 1116 random observations with missing data; a rough picture of the missing patterns is displayed in Table 4. To unify the scales of the continuous variables, the raw continuous data are standardized.

Based on the meanings of the questions that correspond to the selected manifest variables, we propose an NSEM with the following specifications. In the measurement equation, we consider $\boldsymbol{A} = \mathrm{diag}(0.0^*, 0.0^*, 0.0^*, \mu_4, \mu_5, \mu_6, 0.0^*, 0.0^*, 0.0^*)$ with fixed 0.0^* values for identification, and $\boldsymbol{c}_i = (1.0, \ldots, 1.0)^T$. From the meaning of the items (see Appendix IV), the first three items (y_1, y_2, y_3) are related to a latent variable, η, which can be roughly interpreted as 'threat of AIDS', while the next three items (y_4, y_5, y_6) and the last three items (y_7, y_8, y_9) are respectively related to latent variables ξ_1 and ξ_2, which can be interpreted as 'aggressiveness of the sex worker', and 'worry of contracting AIDS'. Hence, we consider the following common non-overlapping structure of $\boldsymbol{\Lambda}$ with some fixed zeros entries at the appropriate positions:

$$\boldsymbol{\Lambda}^T = \begin{bmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{bmatrix}.$$

Following a common practice in factor analysis and SEM, the 1.0^* is fixed to identify the scale of the latent factor. The following nonlinear structural equation is considered: $\eta = b_1x_1 + b_2x_2 + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta$. To identify the parameters that are related to the ordered categorical variables in $\boldsymbol{y}$, we follow the suggestion of Shi and Lee (2000) to set $\alpha_{jh} = \Phi^{*-1}(f_{jh})$ for $h = 1, 4$, and $j = 1, 2, 3, 7, 8, 9$, where f_{jh} are the observed cumulative marginal proportions of the categories with $z_j < h$, and Φ^* is the distributional function of $N(0, 1)$.

Table 4. Missing patterns and their sample sizes: AIDS data, "x" and "o" indicate missing and observed entries, respectively.

	Sample	Manifest Variables										Sample	Manifest Variables								
Pattern	Size	1	2	3	4	5	6	7	8	9	Pattern	Size	1	2	3	4	5	6	7	8	9
1	755	o	o	o	o	o	o	o	o	o	27	1	x	o	x	x	o	o	o	o	o
2	40	x	o	o	o	o	o	o	o	o	28	1	x	o	x	o	o	o	x	o	o
3	77	o	x	o	o	o	o	o	o	o	29	1	x	o	x	o	o	o	o	x	o
4	31	o	o	x	o	o	o	o	o	o	30	3	x	o	x	o	o	o	o	o	x
5	5	o	o	o	x	o	o	o	o	o	31	2	o	x	x	x	o	o	o	o	o
6	12	o	o	o	o	o	o	x	o	o	32	1	o	x	x	o	o	o	x	o	o
7	15	o	o	o	o	o	o	o	x	o	33	2	o	x	x	o	o	o	o	x	o
8	32	o	o	o	o	o	o	o	o	x	34	2	o	x	x	o	o	o	o	o	x
9	9	x	x	o	o	o	o	o	o	o	35	1	o	x	o	o	o	o	x	o	x
10	14	x	o	x	o	o	o	o	o	o	36	3	o	x	o	o	o	o	o	x	x
11	1	x	o	o	o	o	o	x	o	o	37	2	o	o	x	o	o	o	x	x	o
12	6	x	o	o	o	o	o	o	x	o	38	1	o	o	x	o	o	o	x	o	x
13	1	x	o	o	o	o	o	o	o	x	39	19	o	o	o	o	o	o	x	x	x
14	14	o	x	x	o	o	o	o	o	o	40	2	x	x	x	o	o	o	o	o	x
15	2	o	x	o	o	o	o	x	o	o	41	1	x	x	o	o	o	o	x	x	o
16	2	o	x	o	o	o	o	o	x	o	42	1	x	o	x	o	o	o	x	x	o
17	7	o	x	o	o	o	o	o	o	x	43	1	x	o	x	o	o	o	x	o	x
18	5	o	o	x	o	o	o	x	o	o	44	1	x	o	x	o	o	o	o	x	x
19	4	o	o	x	o	o	o	o	x	o	45	2	x	o	o	o	o	o	x	x	x
20	6	o	o	x	o	o	o	o	o	x	46	1	o	x	x	o	o	o	x	x	o
21	1	o	o	o	x	o	o	o	o	x	47	3	o	x	x	o	o	o	x	o	x
22	1	o	o	o	o	o	o	x	x	o	48	1	o	x	o	o	o	o	x	x	x
23	8	o	o	o	o	o	o	x	o	x	49	1	o	o	x	o	o	o	x	x	x
24	6	o	o	o	o	o	o	o	x	x	50	1	x	o	x	o	o	o	x	x	x
25	4	x	x	x	o	o	o	o	o	o	51	1	o	o	x	x	o	o	x	x	x
26	2	x	x	o	o	o	o	o	o	x											

The path sampling procedure is used to compute the logarithm Bayes factor for comparing the model under the MAR assumption (M_1) with the model with the following nonignorable missingness mechanism:

$$
\begin{aligned}
M_0 : &\operatorname{logit}\{\operatorname{pr}(r_{ij} = 1 | \boldsymbol{y}_{(2)i}, \boldsymbol{z}_{oi}, \boldsymbol{x}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\} \\
&\quad = \varphi_{y0} + \varphi_{y1} z_{i1} + \cdots + \varphi_{y3} z_{i3} + \varphi_{y4} y_{i4} + \cdots + \varphi_{y6} y_{i6} + \varphi_{y7} z_{i7} + \cdots + \varphi_{y9} z_{i9}, \\
&\operatorname{logit}\{\operatorname{pr}(r_{i,9+j} = 1 | \boldsymbol{x}_i, \boldsymbol{\varphi})\} = \varphi_{x0} + \varphi_{x1} x_{i1} + \varphi_{x2} x_{i2}.
\end{aligned}
\tag{12}
$$

We have to assign values to the hyperparameters in the conjugate prior distributions. To provide an illustration, we consider a situation in which we have no external prior information. Among many other alternatives in selecting the hyperparameter values, we randomly select 316 observations to conduct an auxiliary Bayesian estimation with non-informative prior distributions for getting the prior inputs. The estimates $\tilde{\mathbf{\Lambda}}_{Ak}$, $\tilde{\mathbf{\Lambda}}_{\omega k}$, $\tilde{\boldsymbol{\varphi}}_y$, $\tilde{\boldsymbol{\varphi}}_x$, and $\tilde{\mathbf{\Phi}}$ of $\mathbf{\Lambda}_{Ak}, \mathbf{\Lambda}_{\omega k}, \boldsymbol{\varphi}_y, \boldsymbol{\varphi}_x$, and $\mathbf{\Phi}$ obtained from the auxiliary estimations are respectively used as prior inputs for $\mathbf{\Lambda}_{0Ak}, \mathbf{\Lambda}_{0\omega k}, \boldsymbol{\varphi}_{0y}, \boldsymbol{\varphi}_{0x}$ and $\boldsymbol{R}_0(= 5\tilde{\mathbf{\Phi}})$ with $\rho_0 = 8$, $\boldsymbol{H}_{0Ak}$ and $\boldsymbol{H}_{0\omega k}$ are taken as $0.5\boldsymbol{I}$, $\boldsymbol{V}_y$ and $\boldsymbol{V}_x$ are taken as $\boldsymbol{I}$, and $\boldsymbol{R}_{0x} = 5\boldsymbol{I}$ with $\rho_{0x} = 8$. The hyperparameter values associated with $\psi_{\epsilon k}$ and $\psi_{\zeta k}$ are taken as $\alpha_{0Ak} = \alpha_{0\zeta k} = 10$, and $\beta_{0Ak} = \beta_{0\zeta k} = 4$ to allow comparatively large variances in the prior distributions for flexibility. The Bayesian results are then obtained on the basis of the remaining 800 observations. The number of grids is taken to be 10, and for each t_l, 4,000 simulated observations are used to compute $\bar{H}_l$ after 3,000 burn-in iterations. To reveal the convergence, plots of the EPSR values for all the unknown parameters against the iteration numbers are presented in Figure 1. The logarithm Bayes factor computed via the path sampling procedure is equal to -160.63. According to the criterion given by Kass and Raftery (1995), M_0 is significantly better than M_1. That is, the nonignorable missingness mechanism defined in M_0 is better than MAR. The PPP p-value (see Bayarri and Berger (2000)) corresponding to the missingness mechanism model M_0 is equal to 0.405. This indicates that the proposed NSEM and the selected missingness mechanism model fit the data.

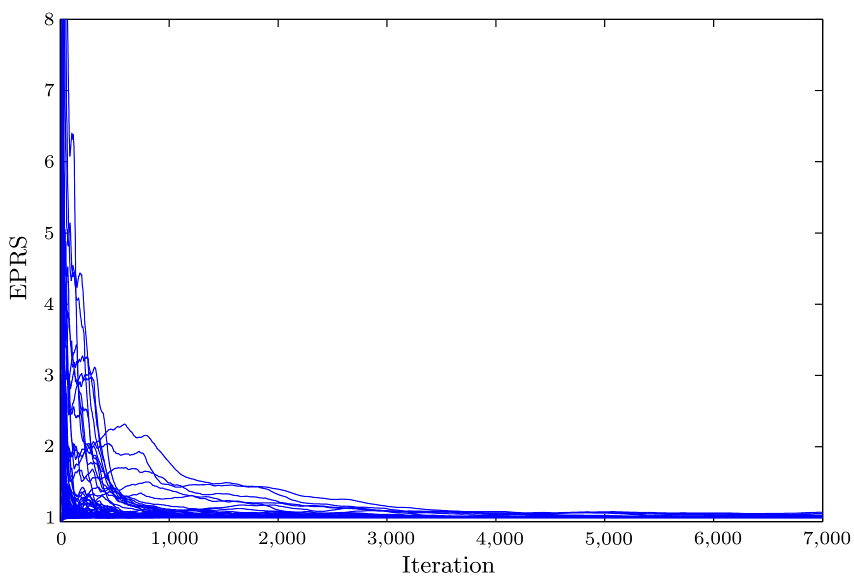

Figure 1. EPSR values of all parameters against iteration numbers: AIDS.

The Bayesian estimates and their standard error estimates of the unknown parameters in the selected model are presented in Table 5. It can be seen that the factor loading estimates are quite large, which indicates a strong association of the latent variables and their corresponding indicators. Before providing interpretations of the estimated nonlinear structural equation, we note from the scale of the ordered categorical variables $(y_1, y_2, y_3, y_7, y_8, y_9)$ that comparatively large (positive) values of η and ξ_2 imply that an individual feels a high threat from AIDS and is more worried about contracting AIDS. With this understanding, the estimated nonlinear structural equation $\eta = -0.08x_1 + 0.096x_2 - 0.357\xi_1 + 0.308\xi_2 - 0.151\xi_1\xi_2$ has the following interpretations. (i) From $\hat{b}_1 = -0.08$, the longer sex workers

Table 5. Bayesian estimates and their standard errors: AIDS.

Para.	EST	SD	Para.	EST	SD
u_4	-0.011	0.021	α_{12}	-0.992	0.024
u_5	-0.001	0.020	α_{13}	-0.624	0.024
u_6	0.004	0.017	α_{22}	0.057	0.019
λ_{21}	0.356	0.076	α_{23}	0.365	0.021
λ_{31}	0.813	0.105	α_{32}	-1.141	0.034
λ_{52}	2.023	0.206	α_{33}	-0.742	0.028
λ_{62}	1.399	0.110	α_{72}	-1.579	0.037
λ_{83}	0.598	0.092	α_{73}	-0.705	0.016
λ_{93}	1.056	0.156	α_{82}	-0.389	0.019
$\psi_{\epsilon 1}$	0.627	0.051	α_{83}	0.243	0.018
$\psi_{\epsilon 2}$	0.994	0.053	α_{92}	-1.012	0.025
$\psi_{\epsilon 3}$	0.739	0.053	α_{93}	-0.113	0.014
$\psi_{\epsilon 4}$	0.968	0.031	ϕ_{x11}	1.055	0.031
$\psi_{\epsilon 5}$	0.572	0.046	ϕ_{x12}	0.078	0.021
$\psi_{\epsilon 6}$	0.520	0.027	ϕ_{x22}	1.001	0.029
$\psi_{\epsilon 7}$	0.738	0.060	φ_{x0}	-8.940	0.377
$\psi_{\epsilon 8}$	0.938	0.049	φ_{x1}	-6.141	0.438
$\psi_{\epsilon 9}$	0.726	0.058	φ_{x2}	-0.063	0.323
b_1	-0.080	0.019	φ_{y0}	0.176	0.085
b_2	0.096	0.019	φ_{y1}	-0.230	0.020
γ_1	-0.357	0.089	φ_{y2}	-0.251	0.023
γ_2	0.308	0.069	φ_{y3}	-0.244	0.019
γ_3	-0.151	0.151	φ_{y4}	-0.162	0.058
ψ_ζ	0.301	0.037	φ_{y5}	-0.205	0.045
ϕ_{11}	0.116	0.018	φ_{y6}	-0.099	0.049
ϕ_{12}	-0.018	0.008	φ_{y7}	-0.297	0.021
ϕ_{22}	0.253	0.044	φ_{y8}	-0.122	0.020
			φ_{y9}	-0.244	0.019

are in their jobs, the less threat they feel from AIDS; and from $\hat{b}_2 = 0.096$, the more that they think they know about AIDS, the more threat they feel from AIDS. (ii) From $\hat{\gamma}_1 = -0.357$, more aggressive sex workers seem to feel less threat from AIDS, and from $\hat{\gamma}_2 = 0.308$, sex workers who are more worried about contracting AIDS feel more of a threat from AIDS. (iii) From $\hat{\gamma}_3 = -0.151$, ξ_1 and ξ_2 have an interaction effect on 'threat of AIDS, η'. The basic interpretation is that the 'additive' effect of 'aggressiveness of the sex worker' and 'worry about contracting AIDS' is inadequate to account for their relationship with 'threat of AIDS', and an interaction effect has to be added. In different situations, this interaction term (with a negative sign) has different effects. For example, a less aggressive sex worker (with a relatively negative ξ_1) who is more worried about contracting AIDS (with a positive ξ_2) would feel an increased threat from AIDS (-0.151 $\xi_1\xi_2$ would be positive). From $\hat{\phi}_{11}$, $\hat{\phi}_{12}$, and $\hat{\phi}_{22}$, the estimated correlation between ξ_1 and ξ_2 is -0.105. Hence, 'aggressiveness' and 'worry' are negatively correlated. From the estimates of $\varphi_{x0}, \varphi_{x1}$ and $\varphi_{y1}, \ldots, \varphi_{y9}$ and their standard error estimates, we see that a nonignorable missingness mechanism for modelling the missing data is necessary. This result is consistent with the conclusion that was obtained by model comparison.

Acknowledgement

This research was fully supported by a grant (CUHK 4243/03H) from the Research Grants Council of the Hong Kong Special Administration Region. We are thankful to the Editor, an associate editor, and a referee for valuable comments that improved the paper, and to D. E. Morisky and J. A. Stein for the use of their AIDS data set.

Appendix I. Implementation of the MH Algorithms

To simulate observations from conditional distribution in relation to (5), let $\mathbf{\Pi}_0 = \boldsymbol{I} - \mathbf{\Pi}$, $\mathbf{\Delta}_H = \partial \boldsymbol{H}(\boldsymbol{\xi}_i)/\partial \boldsymbol{\xi}_i^T|_{\boldsymbol{\xi}_i=\mathbf{0}}$, and

$$\mathbf{\Omega}_\omega^{-1} = \mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} + \begin{pmatrix} \mathbf{\Pi}_0^T \mathbf{\Psi}_\zeta^{-1} \mathbf{\Pi}_0 & -\mathbf{\Pi}_0^T \mathbf{\Psi}_\zeta^{-1} \mathbf{\Gamma} \mathbf{\Delta}_H \\ -\mathbf{\Delta}_H^T \mathbf{\Gamma}^T \mathbf{\Psi}_\zeta^{-1} \mathbf{\Pi}_0 & \mathbf{\Phi}^{-1} + \mathbf{\Delta}_H^T \mathbf{\Gamma}^T \mathbf{\Psi}_\zeta^{-1} \mathbf{\Gamma} \mathbf{\Delta}_H \end{pmatrix}$$
$$+ a_y \boldsymbol{\varphi}_{yII} \boldsymbol{\varphi}_{yII}^T + a_x \boldsymbol{\varphi}_{xII} \boldsymbol{\varphi}_{xII}^T,$$

where $a_y = p \exp(\varphi_{y0} + \boldsymbol{\varphi}_{yI}^T \boldsymbol{m}_{yIi})/(1 + \exp(\varphi_{y0} + \boldsymbol{\varphi}_{yI}^T \boldsymbol{m}_{yIi}))^2$, $a_x = s \exp(\varphi_{x0} + \boldsymbol{\varphi}_{xI}^T \boldsymbol{m}_{xIi})/(1 + \exp(\varphi_{x0} + \boldsymbol{\varphi}_{yI}^T \boldsymbol{m}_{xIi}))^2$, $\boldsymbol{\varphi}_y = (\boldsymbol{\varphi}_{yI}, \boldsymbol{\varphi}_{yII})$, where $\boldsymbol{\varphi}_{yII}$ is the subvector of $\boldsymbol{\varphi}_y$ in relation to $\boldsymbol{\omega}_i$, $\boldsymbol{\varphi}_x = (\boldsymbol{\varphi}_{xI}, \boldsymbol{\varphi}_{xII})$, where $\boldsymbol{\varphi}_{xII}$ is the subvector of $\boldsymbol{\varphi}_x$ in relation to $\boldsymbol{\omega}_i$, $\boldsymbol{m}_{yIi}$ is the subvector of $\boldsymbol{m}_{yij}$ that corresponds to $\boldsymbol{\varphi}_{yI}$, and $\boldsymbol{m}_{xIi}$ is the subvector of $\boldsymbol{m}_{xij}$ that corresponds to $\boldsymbol{\varphi}_{xI}$. The MH algorithm is implemented as follows. At the $(j+1)$st iteration with a $\boldsymbol{\omega}_i^{(j)}$, a new candidate $\boldsymbol{\omega}_i$ is generated from $N(\boldsymbol{\omega}_i^{(j)}, \sigma_\omega^2 \boldsymbol{\Omega}_\omega)$. It is accepted with probability $\min\{1, p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})/p(\boldsymbol{\omega}_i^{(j)}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})\}$. The variance σ_ω^2 is chosen such that the average acceptance rate is about 0.25 or more (see Gelman, Roberts and Gilks (1995)).

Similarly, the MH algorithm for sampling $\boldsymbol{y}_{(2)mi}$ from $p(\boldsymbol{y}_{(2)mi}|\boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})$ in (6) is implemented as follows. At the $(j+1)$st iteration with a current value $\boldsymbol{y}_{(2)mi}^{(j)}$, a new candidate $\boldsymbol{y}_{(2)mi}$ is generated from the proposal distribution $N(\boldsymbol{y}_{(2)mi}^{(j)}, \sigma_y^2 \boldsymbol{\Omega}_{ym})$, where $\boldsymbol{\Omega}_{ym}^{-1} = \boldsymbol{\Psi}_{\epsilon(2)mi}^{-1} + c_y \boldsymbol{\varphi}_{my} \boldsymbol{\varphi}_{my}^T + c_x \boldsymbol{\varphi}_{mx} \boldsymbol{\varphi}_{mx}^T$, with $c_y = p \exp(\varphi_{y0} + \sum_{l \in \bar{D}_y} \varphi_{yl} m_{yil})/(1 + \exp(\varphi_{y0} + \sum_{l \in \bar{D}_y} \varphi_{yl} m_{yil}))^2$, $c_x = s \exp(\varphi_{x0} + \sum_{l \in \bar{D}_x} \varphi_{xl} m_{xil})/(1 + \exp(\varphi_{y0} + \sum_{l \in \bar{D}_x} \varphi_{yl} m_{yil}))^2$, $\boldsymbol{\varphi}_{my}$ and $\boldsymbol{\varphi}_{mx}$ are subvectors of $\boldsymbol{\varphi}_y$ and $\boldsymbol{\varphi}_x$ corresponding to $\boldsymbol{y}_{(2)mi}$, respectively, $\boldsymbol{\Psi}_{\epsilon(2)mi}$ is a submatrix of $\boldsymbol{\Psi}_\epsilon$ that corresponds to $\boldsymbol{y}_{(2)mi}$, $\bar{D}_y$ is the set of the indexing numbers that corresponds to the elements in $\boldsymbol{\varphi}_y$ but not in $\boldsymbol{\varphi}_{my}$, $\bar{D}_x$ is similarly defined, and σ_y^2 is chosen as before. The acceptance probability is $\min\{1, p(\boldsymbol{y}_{(2)mi}|\boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})/ p(\boldsymbol{y}_{(2)mi}^{(t)}|\boldsymbol{x}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\varphi})\}$.

To sample $\boldsymbol{\varphi}_y$ from $p(\boldsymbol{\varphi}_y|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_0, \boldsymbol{F}, \boldsymbol{r})$ as given in (8), let $\boldsymbol{\Omega}_{\varphi y}^{-1} = (p/4) \sum_{i=1}^n \boldsymbol{m}_{yi} \boldsymbol{m}_{yi}^T + \boldsymbol{V}_y^{-1}$. The MH algorithm is implemented as follows. At the $(j+1)$st iteration with a current value $\boldsymbol{\varphi}_y^{(j)}$, a new candidate $\boldsymbol{\varphi}_y$ is generated from $N[\boldsymbol{\varphi}_y^{(j)}, \sigma_{\varphi y}^2 \boldsymbol{\Omega}_{\varphi y}]$. It is accepted with the probability $\min\{1, p(\boldsymbol{\varphi}_y|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_0, \boldsymbol{F}, \boldsymbol{r})/p(\boldsymbol{\varphi}_y^{(t)}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_0, \boldsymbol{F}, \boldsymbol{r})\}$.

Sampling $\boldsymbol{\varphi}_x$ from $p(\boldsymbol{\varphi}_x|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_0, \boldsymbol{F}, \boldsymbol{r})$ as given in (9) can be implemented as follows. At the $(j+1)$st iteration with a current value $\boldsymbol{\varphi}_x^{(j)}$, a new candidate $\boldsymbol{\varphi}_x$ is generated from $N[\boldsymbol{\varphi}_x^{(j)}, \sigma_{\varphi x}^2 \boldsymbol{\Omega}_{\varphi x}]$, where $\boldsymbol{\Omega}_{\varphi x}^{-1} = (s/4) \sum_{i=1}^n \boldsymbol{m}_{xi} \boldsymbol{m}_{xi}^T + \boldsymbol{V}_x^{-1}$, and s is the dimension of $\boldsymbol{x}$. It is accepted with the probability $\min\{1, p(\boldsymbol{\varphi}_x|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_0, \boldsymbol{F}, \boldsymbol{r})/p(\boldsymbol{\varphi}_x^{(t)}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{C}, \boldsymbol{Z}_0, \boldsymbol{F}, \boldsymbol{r})\}$. The implementation of the MH algorithm for sampling from (7) is similar.

Appendix II. Computation of the Bayes Factor

In the development of the path sampling procedure, $[\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}]$ is augmented with $[\boldsymbol{Y}_m, \boldsymbol{Y}_{(1)o}, \boldsymbol{X}_m, \boldsymbol{F}]$. Consider the following class of densities,

$$z(t) = \int p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}; \boldsymbol{\theta}, \boldsymbol{\vartheta}, t) d\boldsymbol{Y}_m d\boldsymbol{Y}_{(1)o} d\boldsymbol{X}_m d\boldsymbol{F} d\boldsymbol{\theta} d\boldsymbol{\vartheta},$$

where t is a continuous parameter that belongs to interval $[0, 1]$, and $p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}; \boldsymbol{\theta}, \boldsymbol{\vartheta}, t)$ is the density of model M_t that links M_0 and M_1 with the continuous parameter t, such that $M_t = M_0$ if $t = 0$, and $M_t = M_1$ if $t = 1$. By reasoning similar to that of Gelman and Meng (1998), it can be shown that

$$\log B_{10} = \log \frac{z(1)}{z(0)} = \int_0^1 E^*[H(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, t)]dt,$$

where $H(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, t) = d \log p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}, t|\boldsymbol{\theta}, \boldsymbol{\vartheta})/dt$, and E^* is the expectation with respect to density $p(\boldsymbol{Y}_m, \boldsymbol{Y}_{(1)o}, \boldsymbol{X}_m, \boldsymbol{F}, \boldsymbol{\theta}, \boldsymbol{\vartheta}|\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}, t)$. Let $0 = t_0 < t_1 < t_2 < \cdots < t_I < t_{I+1} = 1$. Then $\log B_{10}$ can be estimated by $\widehat{\log B_{10}} = (1/2)\sum_{l=0}^{I}(t_{l+1} - t_l)(\bar{H}_{l+1} + \bar{H}_l)$, where $\bar{H}_l = J^{-1}\sum_{j=1}^{J} H(\boldsymbol{Y}_o, \boldsymbol{Y}_m^{(j)}, \boldsymbol{Y}_{(1)o}^{(j)}, \boldsymbol{X}_o, \boldsymbol{X}_m^{(j)}, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}^{(j)}, \boldsymbol{r}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\vartheta}^{(j)}, t_l)$, and $\{(\boldsymbol{Y}_m^{(j)}, \boldsymbol{Y}_{(1)o}^{(j)}, \boldsymbol{X}_m^{(j)}, \boldsymbol{F}^{(j)}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\vartheta}^{(j)}) : j = 1, \ldots, J\}$ are observations that are simulated from $p(\boldsymbol{Y}_m, \boldsymbol{Y}_{(1)o}, \boldsymbol{X}_m, \boldsymbol{F}, \boldsymbol{\theta}, \boldsymbol{\vartheta}|\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{r}, t_l)$. In the simulation study, we take $I = 10$ and $J = 2,000$ after a burn-in of $3,000$ iterations.

Appendix III: Conditional distributions $p(\boldsymbol{X}_m|\boldsymbol{Y}, \boldsymbol{X}_o, \boldsymbol{Z}_o, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi})$ and $p(\boldsymbol{\tau}|\boldsymbol{X}_m)$

In the simulation study, for independent x_{i1} and x_{i2} such that $x_{i1} \stackrel{D}{=} Bi(1, \tau_1)$, $x_{i2} \stackrel{D}{=} N[\tau_2, 1.0]$, it follows from (7) that

$$\begin{aligned}
&p(x_{i1}|\boldsymbol{y}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi}) \\
&\quad \propto \exp\Big\{ -\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\xi}_i))^T \boldsymbol{\Psi}_\zeta^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\xi}_i)) \\
&\qquad\qquad + \sum_{j=1}^{p}(r_{ij}\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi} - \log(1 + \exp(\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi}))) \\
&\qquad\qquad + \sum_{j=1}^{s}(r_{i,j+p}\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi} - \log(1 + \exp(\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi})))\Big\}\tau_1^{x_{i1}}(1 - \tau_1)^{1 - x_{i1}},
\end{aligned}$$

$$\begin{aligned}
&p(x_{i2}|\boldsymbol{y}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi}) \\
&\quad \propto \exp\Big\{ -\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\xi}_i))^T \boldsymbol{\Psi}_\zeta^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{x}_i, \boldsymbol{\xi}_i)) \\
&\qquad\qquad + \sum_{j=1}^{p}(r_{ij}\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi} - \log(1 + \exp(\boldsymbol{\varphi}_y^T \boldsymbol{m}_{yi}))) \\
&\qquad\qquad + \sum_{j=1}^{s}(r_{i,j+p}\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi} - \log(1 + \exp(\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi}))) - \frac{1}{2}(x_{i2} - \tau_2)^2\Big\}.
\end{aligned}$$

Observation x_{i1} can be sampled from $[x_{i1}|\boldsymbol{y}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi}]$ as follows: randomly generate a u from uniform distribution $U[0, 1]$, then $x_{i1} = 0$ if $u \leq \bar{p}_0/(\bar{p}_0 + \bar{p}_1)$ and 1 otherwise, where $\bar{p}_k = p(x_{i1} = k|\boldsymbol{y}_i, \boldsymbol{z}_{oi}, \boldsymbol{c}_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\varphi})$ for $k = 0, 1$.

The following conjugate prior distributions (Lindley and Smith (1972)) of τ_1 and τ_2 are used, $\tau_1 \stackrel{D}{=} Beta(\nu_1, \nu_2)$, $\tau_2 \stackrel{D}{=} N[\tau_2^0, \nu_3]$, where ν_1, ν_2, ν_3, and τ_2^0 are the given hyper-parameters. Moreover, it is assumed that τ_1 and τ_2 are independent. It can be shown that

$$[\tau_1|\boldsymbol{X}] \stackrel{D}{=} Beta[\nu_1 + \sum_{i=1}^n x_{i1}, \nu_2 + n - \sum_{i=1}^n x_{i1}], \quad [\tau_2|\boldsymbol{X}] \stackrel{D}{=} N\left[\frac{\sum\limits_{i=1}^n x_{i2} + \frac{\tau_2^0}{\nu_3}}{n + \frac{1}{\nu_3}}, \frac{1.0}{n + \frac{1}{\nu_3}}\right].$$

In our example, $\boldsymbol{\alpha}_v$ is the vector of unknown thresholds, and let $\boldsymbol{v} = (v_1, \ldots, v_n)$, $\boldsymbol{x}_i = (x_{i1}, x_{i2})^T$, $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ and $\boldsymbol{x}_i^+ = (x_{i1}, v_i)^T$. Then it follows from (7) and (12) that

$$\begin{aligned} &p(x_{i1}|x_{i2}, v_i, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\Phi}_x, \boldsymbol{\varphi}) \\ &\quad \propto \exp\Big\{ -\frac{1}{2}(\eta_i - \boldsymbol{B}\boldsymbol{x}_i - \boldsymbol{\Gamma}\boldsymbol{H}(\boldsymbol{\xi}_i))^T \boldsymbol{\Psi}_\zeta^{-1}(\eta_i - \boldsymbol{B}\boldsymbol{x}_i - \boldsymbol{\Gamma}\boldsymbol{H}(\boldsymbol{\xi}_i)) \\ &\qquad\qquad + \sum_{j=1}^s (r_{i,j+p}\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi} - \log(1 + \exp(\boldsymbol{\varphi}_x^T \boldsymbol{m}_{xi}))) - \frac{1}{2}\boldsymbol{x}_i^{+T} \boldsymbol{\Phi}_x^{-1} \boldsymbol{x}_i^+ \Big\}, \end{aligned}$$

$$p(\boldsymbol{\alpha}_v, \boldsymbol{v}|\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{r}_i, \boldsymbol{\theta}, \boldsymbol{\Phi}_x, \boldsymbol{\varphi}) \propto \prod_{i=1}^n \phi(\sigma_{\mu x}^{1/2}[v_i - \mu_{xi}]) I_{(\alpha_{x_{i2}}, \alpha_{x_{i2}+1})}(v_i),$$

where $\sigma_{\mu x} = b_2^2/\psi_\zeta + \phi_x^{22}$, $\mu_{xi} = [b_2(\eta_i - b_1 x_{i1} - \boldsymbol{\Gamma}\boldsymbol{H}(\boldsymbol{\xi}_i))/\psi_\zeta - \phi_x^{12} x_{i1}]/\sigma_{\mu x}$, and ϕ_x^{11}, ϕ_x^{12} and ϕ_x^{22} are the elements of $\boldsymbol{\Phi}_x^{-1}$. The prior distribution of $\boldsymbol{\Phi}_x$ is taken as $p(\boldsymbol{\Phi}_x) \sim IW[\boldsymbol{R}_{0x}, \rho_{0x}, s]$, where $\boldsymbol{R}_{0x}$ are the given hyperparameters. It can be shown that $p(\boldsymbol{\Phi}_x|\boldsymbol{X}) \sim IW[\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T + \boldsymbol{R}_{0x}, n + \rho_{0x}, s]$.

Appendix IV. Selected Items in the AIDS Data

The number of the variables in the questionnaire is given in parentheses.

y_1: How worried are you about getting AIDS? (33)
not at all worried 1/2/3/4/5 extremely worried.
y_2: What are the chances that you yourself might get AIDS? (32)
none 1/2/3/4/5 very great.
y_3: How much of a threat do you think AIDS is to the health of people? (31)
no threat at all 1/2/3/4/5 very great.
y_4: How many times did you have vaginal sex in the last 7 days? (43).
y_5: How many 'hand jobs' did you give in the last 7 days? (72)
y_6: How many 'blow jobs' did you give in the last 7 days? (74)
How great is the risk of getting AIDS from the following activities.

y_7: Sexual intercourse with someone you don't know very well without using a condom (27h).

y_8: Sexual intercourse with someone who has the AIDS virus using a condom? (27e).

y_9: Sexual intercourse with someone who injects drugs? (27i).
The scale for y_7, y_8 and y_9 is: no risk 1/2/3/4/5 great risk.

x_1: How long (in months) have you been working at a job where people pay to have sex with you? (37).

x_2: How much do you think you know about the disease called AIDS? (21).
nothing 1/2/3/4/5 a great deal.

References

Bagozzi, R. P., Gopinath, M. and Nyer, P. U. (1999). The role of emotions in marketing. *J. Academy of Marketing Science* **27**, 184-206.

Bayarri, M. J. and Berger, J. O. (2000). P values for composite null models. *J. Amer. Statist. Assoc.* **95**, 1127-1142.

Berger, J. O. (1985). *Statistical Decison Theory and Bayesian Analysis.* Springer-Verlag, New York.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociol. Methods Res.* **17**, 303-316.

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

DiCiccio, T. J., Kass, R. E., Raftery, A. and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *J. Amer. Statist. Assoc.* **92**, 903-915.

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *J. Roy. Statist. Soc. Ser. B* **62**, 355-366.

Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151-168.

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

Gelman, A., Roberts, G. O. and Gilks, W. R. (1995). Efficient Metropolis jumping rules. In *Bayesian Statistics* **5** (Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), 599-607. Oxford University Press, Oxford.

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

Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B* **37**, 1-25.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika* **57**, 97-109.

Ibrahim, J. G., Chen, M. H. and Lipsitz, S. R. (1999). Monte Carlo EM for missing covariates in parametric regression models. *Biometrics* **55**, 591-596.

Ibrahim, J. G., Chen, M. H. and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorbale. *Biometrika* **88**, 551-564.

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modelling 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.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

Lee, S. Y. and Shi, J. Q. (2000). Bayesian analysis of structural equation model with fixed covariates. *Struct. Equ. Model.* **7**, 411-430.

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. (2004a). Maximum likelihood estimation and model comparison for mixtures of structural equation models with ignorable missing data. *J. Classification* **20**, 221-225.

Lee, S. Y. and Song, X. Y. (2004b). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British J. Math. Statist. Psych.* **57**, 131-150.

Lee, S. Y. and Song, X. Y. (2004c). Evaluation of Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research* **39**, 653-686.

Lee, S. Y. and Song, X. Y. (2004d). Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics* **60**, 624-636.

Lee, S. Y., Song, X. Y., Skevington, S. and Hao, Y. T. (2005). Application of structural equation models to quality of life. *Structural Equation Modeling- A Multidisciplinary Journal* **12**, 435-453.

Lee, S. Y. and Zhu, H. T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British J. Math. Statist. Psych.* **53**, 209-232.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 1-42.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

Lipsitz, S. R., Ibrahim, J. G. and Zhao, L. P. (1999). A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *J. Amer. Statist. Assoc.* **94**, 1147-1160.

Lipsitz, S. R. and Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika* **83**, 916-922.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machine. *J. Chemical Physics* **21**, 1087-1091.

Meng, X. L. and Wong, Y. X. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statist. Sinica* **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**, 213-320.

Parzen, M., Lipsitz, S. R., Ibrahim, J. G. and Lipshultz, S. (2002). A weighted estimating equation for linear regression with missing covariate data. *Statist. Medicine* **21**, 2421-2436.

Rathouz, P. J., Satten, G. A. and Carroll, R. J. (2002). Semiparametric inference in matched case-control studies with missing covariate data. *Biometrika* **89**, 905-916.

Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 650-663.

Scheines, R., Hoijtink, H. and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika* **64**, 37-52.

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.

Stubbendick, A. L. and Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing response and covariates in random effects models. *Biometrics* **59**, 1140-1150.

Williams, L. J., Edwards, J. R. and Vandenberg, R. J. (2003). Recent advances in causal modeling methods for organizational and management research. *J. Management* **29**, 903-936.

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

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

Department of Statistics, Yunnan University, Kunming 650091, China.

E-mail: nstang@ynu.edu.ch

(Received September 2004; accepted May 2005)