

# Bayesian analysis of nonlinear structural equation models with mixed continuous, ordered and unordered categorical, and nonignorable missing data

JING-HENG CAI, XIN-YUAN SONG AND SIK-YUM LEE

---

Structural equation models (SEMs) have been widely applied in examining inter-relationships among latent and observed variables in social, psychological, and medical research. Motivated by the fact that correlated discrete variables and missing data are frequently encountered in practical applications, a nonlinear SEM (NSEM) that accommodates covariates, mixed continuous and discrete variables, and nonignorable missing data is proposed. Bayesian methods for estimation and model comparison are discussed. One real-life data set about cardiovascular disease is used to illustrate the methodologies.

KEYWORDS AND PHRASES: Latent variables, Ordered and unordered categorical data, Nonignorable missing data, Bayesian approach.

---

## 1. INTRODUCTION

Latent variables have been widely used in assessing multidimensional concept which is related to two or more observed variables. Historically, they were analyzed by the factor analysis model, and have been widely applied to educational, social, and psychological research. Structural equation models (SEMs) [2, 15] generalize the factor analysis model by adding a structural equation to assess the effects of some independent (exogenous) latent variables to some dependent (endogenous) latent variables. As latent variables are very common in practical research, SEMs have wide applications not only in behavioral and social sciences, but also in biological, environmental, and medical sciences [see for example, 1, 17, 23, 25, among others].

The rapid growth of SEMs is due to the demand of subtle models and the related statistical methods for solving complex research problems in various fields. In this paper, the proposed model is motivated by the strong demand in analyzing genotype and phenotype latent variables and their interactions. Epidemiological and family-based analyses have indicated that diseases with complex traits have both genotype and phenotype determinants [26]. Very often, particular groups of observed genotype and phenotype

variables are highly correlated and relate to a multidimensional concept of major interest. For example, in the study of cardiovascular disease, we are interested in examining the effect of the latent variable ‘blood pressure’ (created by observed variables systolic and diastolic blood pressures) to the latent variable ‘cardiovascular disease’ (created by observed variables ischemic heart disease and cerebral vascular accident), rather than the effect of systolic blood pressure to ischemic heart disease. Moreover, the importance of genotype-genotype and genotype-phenotype interactions is now being increasingly recognized in the study of complex diseases [3]. For achieving more efficient statistical analysis and better interpretation, it is more appropriate to investigate the interactions of latent variables rather than the interactions among observed variables. Hence, there is an urgent need to develop efficient statistical methods to investigate the linear and nonlinear effects of independent phenotype and genotype latent variables on dependent latent variables. For instance, in the study of diabetic cardiovascular disease, it is advantageous to analyze the interaction of the latent variables ‘blood pressure’ and ‘lipid control’ (created by plasma triglyceride and high density lipoprotein), rather than the large number of interactions formed by the corresponding observed variables. To examine the inter-relationships among various latent variables, we propose a nonlinear SEM to fit the data which are provided by an applied genomic program in the Institute of Diabetes, The Chinese University of Hong Kong. Based on some basic knowledge of the study, we will focus on certain phenotype and genotype observed variables in this paper. These observed variables are grouped into various latent variables for formulating an SEM according to their medical meanings. The observed variables of interest and their corresponding latent variables are given in Table 1. In general, as it is important to examine the possible nonlinear effects of independent latent variables on the dependent genotype and phenotype latent variables, a model with a nonlinear structural equation will be considered. Hence, we need to develop a nonlinear SEM with mixed continuous, ordered, and unordered categorical variables. Note that the distribution of an unordered categorical variable is multinomial. Analysis

Table 1. Description of the Observed Variables and Their Associative Latent Variables in Cardiovascular Disease Data

Observed variable	Scale	Type	Associative latent variable
Ischemic heart disease (IHD)	dichotomous	phenotype	Cardiovascular disease
Cerebral vascular accident (CVA)	dichotomous	phenotype	
Systolic blood pressure (SBP)	continuous	phenotype	Blood pressure
Diastolic blood pressure (DBP)	continuous	phenotype	
High density lipoprotein cholesterol (HDL)	continuous	phenotype	Lipid control
Plasma triglyceride (TG)	continuous	phenotype	
Body mass index (BMI)	continuous	phenotype	Body shape
Waist hip ratio (WHR)	continuous	phenotype	
Paraoxonase 1 SNP1 (PON11)	unordered categorical	genotype	Gene-inflammatory
Paraoxonase 1 SNP2 (PON12)	unordered categorical	genotype	
Fibrinogen (FGB)	unordered categorical	genotype	Gene-lipid control
Endothelial leukocyte adhesion molecule-1 SNP2 (SELE2)	unordered categorical	genotype	

of correlated multinomial variables is non-trivial and has received a great deal of attention in econometrics. Recently, multinomial probit models [13, 20] have been developed to predict multinomial variables through fixed covariates with a regression model. As these models do not involve latent variables, they cannot be applied to our problem.

The data set obtained under the study of cardiovascular disease includes missing data. Often, missing data in this study is nonignorable in the sense that the reason for missingness often depends on the missing values themselves [see 19]. For example, the side effects of the treatment may make the patients worse and thereby affect patient participation. Hence, the missing data are considered as nonignorable with a nonignorable missing mechanism. Analysis of this kind of missing data has received considerable attention in statistics [see for example, 5, 12, among others]. An important issue in the analysis is in formulating a model for the missing mechanism. In this paper, motivated by Ibrahim et al. [12], a general framework based on the logistic regression model will be proposed. Under the context of the proposed non-linear SEM, various competing models under this general framework will be compared through the Bayes factor [14].

The paper is organized as follows. In Section 2, a non-linear SEM (NSEM) that accommodates fixed covariates, mixed continuous, ordered, and unordered categorical data with nonignorable missing entries is proposed together with a logistic regression model for the nonignorable missing mechanism. Bayesian methods for analyzing the proposed model, including estimation and model comparison, are presented in Section 3. Here, Markov chain Monte Carlo (MCMC) tools such as the Gibbs sampler [10] and the Metropolis-Hastings (MH) algorithm [11, 21] are applied for estimation, and a path sampling procedure [8] is used to

compute the Bayes factor for model comparison. To reveal the performance of the Bayesian methods, results obtained from a simulation study are reported in Section 4. Analysis of the real example in relation to cardiovascular disease of diabetic patients is presented in Section 5. A discussion is given in Section 6, and some technical details are given in the Appendices.

## 2. NSEM WITH MIXED CONTINUOUS AND DISCRETE VARIABLES, AND NONIGNORABLE MISSING DATA

### 2.1 NSEM with covariates, and mixed continuous and discrete data

We propose an NSEM for a  $p$  by 1 observed random vector which is composed of continuous, ordered, and unordered categorical variables. Without loss of generality, let the  $i$ th observation be  $\mathbf{d}_i = (z_{i1}, \dots, z_{i,r_1}, y_{i,r_1+1}, \dots, y_{i,r_2}, u_{i1}, \dots, u_{i_s})'$ , where  $s = p - r_2$ ,  $z_{ih}$  is ordered categorical,  $y_{ih}$  is continuous, and  $u_{ih}$  is unordered categorical. According to common practice in treating ordered categorical variables in SEMs [15, 17], an ordered categorical variable  $z_{ij}$  is defined through an underlying random variable  $y_{ij}^*$  with a continuous distribution and a threshold specification, as follows:

$$(1) \quad z_{ij} = h \quad \text{if } \alpha_{j,h} \leq y_{ij}^* < \alpha_{j,h+1}, \quad \text{for } h = 0, \dots, H_j,$$

where  $\{-\infty = \alpha_{j,0} < \alpha_{j,1} < \dots < \alpha_{j,H_j} < \alpha_{j,H_j+1} = \infty\}$  is the set of thresholds that defines  $H_j + 1$  categories. For each unordered categorical variable, we assume that it takes on one and only one of the possible values  $\{0, 1, \dots, K_j - 1\}$ . For simplicity, we assume that  $K_j = K$ ; however, this assumption can be easily relaxed. In the literature, an

unordered categorical variable, say  $u_j$ ,  $j = 1, \dots, s$ , is commonly modeled in terms of an unobserved multivariate normal random vector  $\mathbf{v}_j = (v_{j,1}, \dots, v_{j,K-1})'$  as follows [see 13, 27]: for  $j = 1, \dots, s$ ,

$$(2) \quad \begin{aligned} u_j(\mathbf{v}_j) &= 0 \text{ if } \max(\mathbf{v}_j) < 0; \text{ and} \\ u_j(\mathbf{v}_j) &= k \text{ if } \max(\mathbf{v}_j) = v_{j,k} > 0, \end{aligned}$$

where  $\max(\mathbf{v}_j)$  is the largest element in vector  $\mathbf{v}_j$ . For example, for an observed variable  $u_j$  with three outcomes, we have  $\mathbf{v}_j = (v_{j,1}, v_{j,2})'$ ;  $u_j(\mathbf{v}_j) = 0$  if both  $v_{j,1}$  and  $v_{j,2}$  are less than zero;  $u_j(\mathbf{v}_j) = 1$  if  $v_{j,1} > v_{j,2}$  and  $v_{j,1} > 0$ ; and  $u_j(\mathbf{v}_j) = 2$  if  $v_{j,2} > v_{j,1}$  and  $v_{j,2} > 0$ .

Let  $\mathbf{w}_i = (y_{i1}^*, \dots, y_{i,r_1}^*, y_{i,r_1+1}, \dots, y_{i,r_2}, \mathbf{v}'_{i1}, \dots, \mathbf{v}'_{is})'$  be the  $r_2 + s(K-1)$  by 1 latent continuous random observation which is related to the manifest random observation  $\mathbf{d}_i$  that includes the manifest ordered categorical, continuous, and unordered categorical observations. One important issue is to define a measurement equation  $\mathbf{w}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{b}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$ , in which the correlated manifest variables in  $\mathbf{w}_i$  are appropriately grouped into latent variables in  $\boldsymbol{\omega}_i$ , and the covariate  $\mathbf{b}_i$  is taken into account. Basically, we need to find an appropriate structure of  $\boldsymbol{\Lambda}$  for formulating latent variables in  $\boldsymbol{\omega}_i$  according to the meaning and properties of the manifest variables. Here the measurement equation  $\mathbf{w}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{b}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$  is defined as follows, for  $i = 1, \dots, n$ :

$$(3) \quad \begin{pmatrix} y_{i1}^* \\ \vdots \\ y_{i,r_1}^* \\ y_{i,r_1+1} \\ \vdots \\ y_{i,r_2} \\ \mathbf{v}_{i1} \\ \vdots \\ \mathbf{v}_{is} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_{r_1} \\ \mu_{r_1+1} \\ \vdots \\ \mu_{r_2} \\ \boldsymbol{\mu}_{r_2+s} \end{pmatrix} + \begin{pmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_{r_1} \\ \mathbf{A}_{r_1+1} \\ \vdots \\ \mathbf{A}_{r_2} \\ \mathbf{I}_{K-1}\mathbf{A}_{r_2+1} \\ \vdots \\ \mathbf{I}_{K-1}\mathbf{A}_{r_2+s} \end{pmatrix} \mathbf{b}_i \\ + \begin{pmatrix} \boldsymbol{\Lambda}_1 \\ \vdots \\ \boldsymbol{\Lambda}_{r_1} \\ \boldsymbol{\Lambda}_{r_1+1} \\ \vdots \\ \boldsymbol{\Lambda}_{r_2} \\ \mathbf{I}_{K-1}\boldsymbol{\Lambda}_{r_2+1} \\ \vdots \\ \mathbf{I}_{K-1}\boldsymbol{\Lambda}_{r_2+s} \end{pmatrix} \boldsymbol{\omega}_i + \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{i,r_1} \\ \epsilon_{i,r_1+1} \\ \vdots \\ \epsilon_{i,r_2} \\ \epsilon_{i,r_2+1} \\ \vdots \\ \epsilon_{i,r_2+s} \end{pmatrix},$$

where  $\boldsymbol{\mu}$  is an unknown vector of intercepts,  $\mathbf{A}_j$  ( $1 \times m_1$ ) is a row vector of unknown coefficients,  $\mathbf{b}_i$  ( $m_1 \times 1$ ) is a vector of fixed covariates,  $\boldsymbol{\Lambda}_j$  ( $1 \times q$ ) is a row vector of factor loadings,  $\mathbf{I}_{K-1}$  is a  $K-1$  by 1 vector of one's, and  $\epsilon$ 's are measurement errors. The term  $\mathbf{I}_{K-1}\boldsymbol{\Lambda}_{r_2+j}$  is used to relate the  $K-1$

by 1 vector  $\mathbf{v}_{ij}$  with the corresponding unordered categorical variable  $u_{ij}$ . Under this formulation,  $\boldsymbol{\Lambda}_{r_2+j}$  is related to  $u_{ij}$ , for  $j = 1, \dots, s$ . This relationship is useful in interpreting the specific latent variables in  $\boldsymbol{\omega}_i$  with their corresponding manifest unordered categorical variables  $u_{ij}$ , see [27]. Let  $\boldsymbol{\epsilon}_{iz} = (\epsilon_{i1}, \dots, \epsilon_{i,r_1})'$  and  $\boldsymbol{\epsilon}_{iy} = (\epsilon_{i,r_1+1}, \dots, \epsilon_{i,r_2})'$ ; we assume that  $\boldsymbol{\epsilon}_{iz} \stackrel{D}{=} N[\mathbf{0}, \boldsymbol{\Psi}_z]$ ,  $\boldsymbol{\epsilon}_{iy} \stackrel{D}{=} N[\mathbf{0}, \boldsymbol{\Psi}_y]$ , and for  $j = 1, \dots, s$ ,  $\boldsymbol{\epsilon}_{i,r_2+j} \stackrel{D}{=} N[\mathbf{0}, \boldsymbol{\Psi}_{v_j}]$ , and that  $\boldsymbol{\epsilon}_{iz}$ ,  $\boldsymbol{\epsilon}_{iy}$ , and  $\boldsymbol{\epsilon}_{i,r_2+j}$  are mutually independent. A dichotomous variable can be identified by fixing the corresponding diagonal element in  $\boldsymbol{\Psi}_z$  at 1.0; and ordered categorical variables with more than two categories can be identified by the standard methods by fixing the thresholds at both ends at preassigned values [see 15, 22]. As the multiplication of an arbitrary positive constant to each  $v_{j,k}$  does not change the value of  $u_j$  in (2), unordered categorical variables are not identified. To identify these variables, we follow the method given in the literature [6, 27] to fix  $\boldsymbol{\Psi}_{v_j} = \mathbf{I}_{K-1}$ , an identity matrix of order  $K-1$ . Finally, the NSEM can be identified by fixing appropriate elements in  $\boldsymbol{\Lambda}$  at preassigned values; see for example, an  $\boldsymbol{\Lambda}$  with the non-overlapping structure given by (14) in Section 5.

We consider a partition of  $\boldsymbol{\omega}_i$  ( $q \times 1$ ) into subvectors  $\boldsymbol{\eta}_i$  ( $q_1 \times 1$ ) and  $\boldsymbol{\xi}_i$  ( $q_2 \times 1$ ) that respectively contain endogenous and exogenous latent variables. In general,  $\boldsymbol{\eta}_i$  may contain phenotype and genotype latent variables that are respectively related with manifest variables in  $\{z_{i1}, \dots, z_{i,r_1}, y_{i,r_1+1}, \dots, y_{i,r_2}\}$  and  $\{u_{i1}, \dots, u_{is}\}$ . The structural equation for assessing the interrelationships among  $\boldsymbol{\eta}_i$  and  $\boldsymbol{\xi}_i$  together with some fixed covariates is defined, as follows:

$$(4) \quad \boldsymbol{\eta}_i = \mathbf{A}^* \mathbf{c}_i + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i, \quad i = 1, \dots, n,$$

where  $\mathbf{c}_i$  ( $m_2 \times 1$ ) is a vector of fixed covariates,  $\mathbf{F}(\boldsymbol{\xi}_i) = (f_1(\boldsymbol{\xi}_i), \dots, f_r(\boldsymbol{\xi}_i))'$  is a vector of differentiable functions  $f_1, \dots, f_r$  that are linearly independent,  $\mathbf{A}^*$  ( $q_1 \times m_2$ ),  $\mathbf{B}$  ( $q_1 \times q_1$ ), and  $\boldsymbol{\Gamma}$  ( $q_1 \times r$ ) are unknown parameter matrices, and  $\boldsymbol{\xi}_i$  and  $\boldsymbol{\delta}_i$  are independently distributed as  $N[\mathbf{0}, \boldsymbol{\Phi}]$  and  $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$ , respectively, where  $\boldsymbol{\Psi}_\delta$  is diagonal. Covariates in  $\mathbf{b}_i$  and  $\mathbf{c}_i$  can be different or equal. For convenience, we assume that  $\mathbf{I} - \mathbf{B}$  is nonsingular and its determinant is independent of elements in  $\mathbf{B}$ . Although our illustrative application focuses on analyzing interaction terms of latent variables in  $\boldsymbol{\xi}_i$ , the proposed framework can be used to analyze general differentiable functions  $f_1, \dots, f_r$ . The fixed covariates  $\mathbf{c}_i$  (and/or  $\mathbf{b}_i$ ) could come from arbitrary continuous or discrete distributions. Let  $\boldsymbol{\Pi} = (\mathbf{A}^*, \mathbf{B}, \boldsymbol{\Gamma})$ , and  $\mathbf{G}(\boldsymbol{\omega}_i) = (\mathbf{c}'_i, \boldsymbol{\eta}'_i, \mathbf{F}(\boldsymbol{\xi}_i))'$ , (4) can be rewritten as

$$(5) \quad \boldsymbol{\eta}_i = \boldsymbol{\Pi}\mathbf{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i, \quad i = 1, \dots, n,$$

The model defined by (3) and (4) is confirmatory. The subject matter knowledge about the manifest variables is very important in model building; for instance, in deciding  $q$ ,  $q_1$ , and  $q_2$ , and the positions and preassigned values of some elements in  $\boldsymbol{\Lambda}$ ; see the illustrative application in Section 5.

## 2.2 Missingness mechanism

To deal with missing entries in  $\mathbf{d}_i = (d_{i1}, \dots, d_{ip})' = (z_{i1}, \dots, z_{i,r_1}, y_{i,r_1+1}, \dots, y_{i,r_2}, u_{i1}, \dots, u_{is})'$ , we define a missing indicator  $\mathbf{r}_i = (r_{i1}, \dots, r_{ip})'$  such that  $r_{ij} = 1$  if  $d_{ij}$  is missing, and  $r_{ij} = 0$  if  $d_{ij}$  is observed. Let  $\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_n)$  and  $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_n)$ ; and let  $\mathbf{D}_{\text{mis}}$  and  $\mathbf{D}_{\text{obs}}$  be the missing and observed data in  $\mathbf{D}$ , respectively. If the distribution of  $\mathbf{r}$  is independent of  $\mathbf{D}_{\text{mis}}$ , the missing mechanism is defined to be missing at random (MAR); otherwise, the missing mechanism is nonignorable [19]. In analyzing a nonignorable missing mechanism, an important issue is to define an appropriate model for  $p(\mathbf{r}|\mathbf{D}, \boldsymbol{\varphi})$  with an unknown parameter vector  $\boldsymbol{\varphi}$ . We may take the matrix of latent variables  $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_n)$  into account, and consider the conditional distribution of  $\mathbf{r}$  given  $(\mathbf{D}, \boldsymbol{\Omega}, \boldsymbol{\varphi})$ . However, we have to be careful not to use a complicated model as it can easily become unidentifiable. Moreover, a too complex model will also lead to inefficient sampling from the related conditional distributions. Based on the above understanding, and the fact that the latent variables in the proposed NSEM are closely related with their corresponding manifest variables (indicators), we propose the following model for the nonignorable missingness mechanism that is independent of  $\boldsymbol{\Omega}$ :

$$(6) \quad p(\mathbf{r}|\mathbf{D}, \boldsymbol{\varphi}) = \prod_{i=1}^n \prod_{j=1}^p \{p(r_{ij} = 1|\mathbf{d}_i, \boldsymbol{\varphi})\}^{r_{ij}} \times \{1 - p(r_{ij} = 1|\mathbf{d}_i, \boldsymbol{\varphi})\}^{1-r_{ij}}.$$

Ibrahim et al. [12] pointed out that since  $r_{ij}$  is binary, one can use a sequence of logistic regressions for modeling  $p(r_{ij} = 1|\mathbf{d}_i, \boldsymbol{\varphi})$  in (6). Hence, the following logistic regression model is used:

$$(7) \quad \begin{aligned} & \text{logit}\{p(r_{ij} = 1|\mathbf{d}_i, \boldsymbol{\varphi})\} \\ &= \varphi_0 + \varphi_1 z_{i1} + \dots + \varphi_r z_{i,r_1} + \varphi_{r_1+1} y_{i,r_1+1} + \dots \\ & \quad + \varphi_{r_2} y_{i,r_2} + \varphi_{r_2+1} u_1 + \dots + \varphi_{r_2+s} u_s \\ &= \boldsymbol{\varphi}' \mathbf{d}_i^* = \varphi_0 + \boldsymbol{\varphi}'_I \mathbf{d}_i, \end{aligned}$$

where  $\mathbf{d}_i^* = (1, \mathbf{d}_i')'$ ,  $\boldsymbol{\varphi} = (\varphi_0, \boldsymbol{\varphi}'_I)'$ , and  $\boldsymbol{\varphi}_I = (\varphi_1, \dots, \varphi_p)'$ . Based on the similar rationale as given by Ibrahim et al. [12] in omitting the latent random effects in a normal random effect model, the latent variables are not used in modeling our missing mechanism. However, all the indicators for endogenous and exogenous latent variables in  $\boldsymbol{\eta}$  and  $\boldsymbol{\xi}$  are included in this missing mechanism. In the Appendices, we derive the corresponding conditional distributions required in the estimation, and describe the implementation of the MH algorithm [11, 21] in generating observations from these conditional distributions.

## 3. BAYESIAN ANALYSIS OF THE MODEL

### 3.1 Estimation with MCMC methods

Bayesian estimation of the unknown parameters in the proposed model defined by (1), (2), (3), (4), (6), and (7) is

discussed in this section. The thresholds in  $\boldsymbol{\alpha}$  are regarded as nuisance parameters in the model because they are not usually involved in competing models for comparison. Following the suggestion of Lee and Zhu [18], we use the following non-informative prior distribution for  $\boldsymbol{\alpha}_j$ ,  $j = 1, \dots, r_1$ :

$$p(\boldsymbol{\alpha}_j) = p(\alpha_{j,2}, \dots, \alpha_{j,H_j-1}) \propto \text{constant}, \\ \text{for } \alpha_{j,2} < \dots < \alpha_{j,H_j-1}.$$

For the other unknown parameters, the commonly used conjugate distributions [see for example, 15] are used. These distributions are proper and induce no problem in using the Bayes factor for model comparison. More specifically, the following independent conjugate prior distributions are used:

$$(8) \quad \begin{aligned} p(\boldsymbol{\mu}) &\stackrel{D}{=} N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \\ p(\boldsymbol{\Lambda}_j|\psi_{\epsilon_j}) &\stackrel{D}{=} N(\boldsymbol{\Lambda}_{0j}, \psi_{\epsilon_j} \mathbf{H}_{0j}), \quad j = 1, \dots, r_2 \\ p(\psi_{\epsilon_j}^{-1}) &\stackrel{D}{=} \text{Gamma}(\alpha_{0j}, \beta_{0j}), \quad j = 1, \dots, r_2, \\ p(\boldsymbol{\Lambda}_j) &\stackrel{D}{=} N(\boldsymbol{\Lambda}_{0j}, \mathbf{H}_{0j}), \quad j = r_2 + 1, \dots, p, \\ p(\boldsymbol{\Pi}_j|\psi_{\delta_j}) &\stackrel{D}{=} N(\boldsymbol{\Pi}_{0j}, \psi_{\delta_j} \mathbf{H}_{0\delta_j}), \quad j = 1, \dots, q_1 \\ p(\psi_{\delta_j}^{-1}) &\stackrel{D}{=} \text{Gamma}(\alpha_{0\delta_j}, \beta_{0\delta_j}), \quad j = 1, \dots, q_1, \\ p(\boldsymbol{\Phi}) &\stackrel{D}{=} \text{IW}_{q_2}(\mathbf{R}_0, \rho_0), \quad p(\boldsymbol{\varphi}) \stackrel{D}{=} N(\mathbf{0}, \mathbf{C}), \end{aligned}$$

where ' $p(\cdot) \stackrel{D}{=}$ ' is defined as 'the distribution of  $p(\cdot)$  is equal to',  $\text{IW}_{q_2}$  denotes a  $q_2$  dimensional inverted Wishart distribution,  $\psi_{\epsilon_j}$ ,  $j = 1, \dots, r_2$  is the  $j$ th diagonal element of  $\boldsymbol{\Psi}_z$  and  $\boldsymbol{\Psi}_y$ ,  $\boldsymbol{\Pi}_j$  and  $\psi_{\delta_j}$ ,  $j = 1, \dots, q_1$  are the  $j$ th row of  $\boldsymbol{\Pi}$  and the  $j$ th diagonal element of  $\boldsymbol{\Psi}_\delta$ , respectively; and  $\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_{0j}, \boldsymbol{\Pi}_{0j}, \alpha_{0j}, \beta_{0j}, \alpha_{0\delta_j}, \beta_{0\delta_j}, \rho_0$ , and positive definite matrices  $\boldsymbol{\Sigma}_0, \mathbf{H}_{0j}, \mathbf{H}_{0\delta_j}, \mathbf{R}_0$ , and  $\mathbf{C}$  are hyperparameters whose values are assumed to be given by the prior information.

Let  $\mathbf{y}_i^*$  and  $\mathbf{v}_i = (\mathbf{v}'_{i1}, \dots, \mathbf{v}'_{is})'$  be the continuous unobserved random vectors corresponding to  $\mathbf{z}_i$  and  $\mathbf{u}_i$ , respectively. Let  $\mathbf{Y}^* = \{\mathbf{y}_1^*, \dots, \mathbf{y}_n^*\}$  and  $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  be the unobserved random data corresponding to  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  and  $\mathbf{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ , respectively. Let  $\{\mathbf{Z}_{\text{obs}}, \mathbf{Y}_{\text{obs}}, \mathbf{U}_{\text{obs}}\}$  and  $\{\mathbf{Z}_{\text{mis}}, \mathbf{Y}_{\text{mis}}, \mathbf{U}_{\text{mis}}\}$  be the observed and missing ordered categorical, continuous, and unordered categorical data, respectively. Furthermore, let  $\mathbf{Y}_{\text{obs}}^*$  and  $\mathbf{V}_{\text{obs}}$  be the unobserved continuous variables corresponding to  $\mathbf{Z}_{\text{obs}}$  and  $\mathbf{U}_{\text{obs}}$ , respectively; and let  $\mathbf{Y}_{\text{mis}}^*$  and  $\mathbf{V}_{\text{mis}}$  be the unobserved continuous variables corresponding to  $\mathbf{Z}_{\text{mis}}$  and  $\mathbf{U}_{\text{mis}}$ , respectively. Finally, let  $\boldsymbol{\theta}$  be the vector that contains all other unknown parameters in the measurement and structural equations of the proposed model,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{r_1})$ , and let  $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi})$  and  $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}|\mathbf{D}_{\text{obs}}, \mathbf{r})$  be the prior distribution and posterior distribution of  $(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi})$ , respectively. The Bayesian estimate of  $(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi})$  will be obtained through the sample mean of a sufficiently large sample of observations generated from  $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}|\mathbf{D}_{\text{obs}}, \mathbf{r})$ . As this posterior distribution is rather complicated, we use the

idea of data augmentation [28] to augment the observed data  $\{\mathbf{D}_{\text{obs}}, \mathbf{r}\}$  with  $\{\mathbf{Y}_{\text{obs}}^*, \mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}_{\text{mis}}, \mathbf{V}_{\text{obs}}, \boldsymbol{\Omega}\}$  in the posterior analysis. A sequence of random observations will be generated from the joint posterior distribution  $p(\boldsymbol{\alpha}, \mathbf{Y}_{\text{obs}}^*, \mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}_{\text{mis}}, \mathbf{V}_{\text{obs}}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi} | \mathbf{D}_{\text{obs}}, \mathbf{r})$  by the Gibbs sampler [10] coupled with the MH algorithm [11, 21]. More specifically,  $(\boldsymbol{\alpha}, \mathbf{Y}_{\text{obs}}^*, \mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}_{\text{mis}}, \mathbf{V}_{\text{obs}}, \boldsymbol{\Omega}, \boldsymbol{\theta},$  and  $\boldsymbol{\varphi}$  are iteratively generated from the corresponding full conditional distributions. Under the above selected prior distributions, these full conditional distributions are derived in Appendix A. We find that the full conditional distributions corresponding to  $(\boldsymbol{\alpha}, \mathbf{Y}_{\text{obs}}^*), \mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}_{\text{mis}}, \boldsymbol{\Omega},$  and  $\boldsymbol{\varphi}$  are nonstandard. Brief details on the application of the MH algorithm in generating observations from these nonstandard distributions are presented in Appendix B.

### 3.2 Model comparison using Bayes factor

An important issue in analyzing SEMs is model comparison. In this paper, we propose Bayes factor as a statistic for model comparison. Suppose that the observed data  $\mathbf{D}_{\text{obs}}$  arise from one of the two competing models,  $M_0$  and  $M_1$ . Let  $p(\mathbf{D}_{\text{obs}}, \mathbf{r} | M_1)$  and  $p(\mathbf{D}_{\text{obs}}, \mathbf{r} | M_0)$  be the probability densities of  $(\mathbf{D}_{\text{obs}}, \mathbf{r})$  given  $M_1$  and  $M_0$ , respectively. The Bayes factor is defined by:

$$(9) \quad B_{10} = \frac{p(\mathbf{D}_{\text{obs}}, \mathbf{r} | M_1)}{p(\mathbf{D}_{\text{obs}}, \mathbf{r} | M_0)}.$$

Criterion for interpreting  $B_{10}$  is given in Kass and Raftery [14]. Very often, it is difficult to obtain  $B_{10}$  analytically. Inspired by nice features presented in Gelman and Meng [8], and its applications to other SEMs [see 15], a path sampling procedure is proposed to compute the logarithm  $B_{10}$  for model comparison. Similar to the Bayesian estimation, the observed data  $\{\mathbf{D}_{\text{obs}}, \mathbf{r}\}$  are augmented with the latent data  $\{\mathbf{Y}_{\text{obs}}^*, \mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}_{\text{mis}}, \mathbf{V}_{\text{obs}}, \boldsymbol{\Omega}\}$  in the computation. Suppose  $M_t$  is a linked model that links  $M_1$  and  $M_0$ . Let  $\mathbf{Y}^* = \{\mathbf{Y}_{\text{obs}}^*, \mathbf{Y}_{\text{mis}}^*\}, \mathbf{V} = \{\mathbf{V}_{\text{obs}}, \mathbf{V}_{\text{mis}}\}, p(\mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \mathbf{D}_{\text{obs}}, \mathbf{r} | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, t)$  be the complete-data likelihood function under the linked model  $M_t$ , and

$$(10) \quad \begin{aligned} & U(\boldsymbol{\alpha}, \mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, t) \\ &= \frac{d}{dt} \log \{p(\mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \mathbf{D}_{\text{obs}}, \mathbf{r} | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, t)\}. \end{aligned}$$

Moreover, let  $\{t_{(l)}, l = 0, \dots, L\}$  be grids in  $[0, 1]$  such that  $0 = t_{(0)} < t_{(1)} < \dots < t_{(L)} < t_{(L+1)} = 1$ . The logarithm Bayes factor is computed as

$$(11) \quad \log B_{10} = \frac{1}{2} \sum_{l=0}^L (t_{(l+1)} - t_{(l)}) (\bar{U}_{(l+1)} + \bar{U}_{(l)}),$$

where  
(12)

$$\bar{U}_{(l)} = J^{-1} \sum_{j=1}^J U(\boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{*(j)}, \mathbf{Y}_{\text{mis}}^{(j)}, \mathbf{V}^{(j)}, \boldsymbol{\Omega}^{(j)}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\varphi}^{(j)}, t_{(l)}),$$

in which  $\{(\boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{*(j)}, \mathbf{Y}_{\text{mis}}^{(j)}, \mathbf{V}^{(j)}, \boldsymbol{\Omega}^{(j)}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\varphi}^{(j)}), j = 1, \dots, J\}$  are simulated observations from the posterior distribution  $p(\boldsymbol{\alpha}, \mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi} | \mathbf{D}_{\text{obs}}, \mathbf{r}, t_{(l)})$ . These observations are drawn via the MCMC methods in the estimation described in Section 3.1. Note that the missing mechanism is involved in the conditional distributions of  $\mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}$ , and  $\boldsymbol{\varphi}$  in simulating observations  $\mathbf{Y}^{*(j)}, \mathbf{Y}_{\text{mis}}^{(j)}, \mathbf{V}^{(j)}$ , and  $\boldsymbol{\varphi}^{(j)}$  for getting  $\bar{U}_{(l)}$ , see (10) and (12). Hence, different missing mechanisms give distinct values of  $\bar{U}_{(l)}$  and hence different Bayes factor values.

## 4. A SIMULATION STUDY

The objective of this section is to present results of a simulation study to reveal the empirical performances of the Bayesian estimates and the Bayes factor for model comparison. A complete data set  $\{\mathbf{d}_i, i = 1, \dots, n\}$  is generated from an NSEM defined in (1), (2), (3), and (4) with 15 manifest variables that are related with six basic latent variables  $\boldsymbol{\eta}_i = \eta_i$ , and  $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{i5})'$ . Among these manifest variables in the measurement equation, the first three are dichotomous, the next six are continuous, and the last six are unordered categorical. The true values of  $\boldsymbol{\mu}, \mathbf{A}$ , and  $\boldsymbol{\Lambda}$  in the measurement equation are given by  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_9, \boldsymbol{\mu}'_{10}, \dots, \boldsymbol{\mu}'_{15})' = (0, \dots, 0, \mathbf{0}', \dots, \mathbf{0}')'$ ,  $\mathbf{A} = (A_1, \dots, A_9, A_{10}, \dots, A_{15})' = (-0.3, \dots, -0.3, 0.3, \dots, 0.3)'$ ,  $\boldsymbol{\Psi}_\epsilon = \text{diag}(1^*, 1^*, 1^*, 0.3, 0.3, 0.4, 0.4, 0.5, 0.5, \mathbf{I}^*, \dots, \mathbf{I}^*)$ , and  $\boldsymbol{\Lambda}^* = (\boldsymbol{\Lambda}'_1, \dots, \boldsymbol{\Lambda}'_{15})'$ , where

$$\boldsymbol{\Lambda}^* = \begin{pmatrix} 1.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ \lambda_{21} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 1.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & \lambda_{52} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 1.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & \lambda_{73} & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & \lambda_{94} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & \lambda_{11,5} & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & \lambda_{12,5} & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & \lambda_{14,6} & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & \lambda_{15,6} & 0.0^* \end{pmatrix}.$$

Here, the parameters with an asterisk are treated as fixed. The true values of  $\lambda$ 's are given by:  $\lambda_{21} = \lambda_{31} = \lambda_{94} = \lambda_{11,5} = \lambda_{14,6} = 0.8$ ,  $\lambda_{52} = \lambda_{63} = \lambda_{12,5} = \lambda_{15,6} = 0.7$ . The fixed covariate  $\mathbf{b}_i = b_i$  in the measurement equation is sampled from  $N(0, 1)$ . The structural equation is defined by

$$\begin{aligned} \eta_i &= a^* c_i + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \gamma_4 \xi_{i4} + \gamma_5 \xi_{i5} + \gamma_6 \xi_{i2} \xi_{i4} \\ &\quad + \gamma_7 \xi_{i3} \xi_{i5} + \delta_i, \end{aligned}$$

where  $\delta_i \stackrel{D}{=} N(0, 0.36)$ , and  $\boldsymbol{\xi}_i \stackrel{D}{=} N(\mathbf{0}, \boldsymbol{\Phi})$  with  $\phi_{ii} = 1$  and  $\phi_{ij} = 0.2, i \neq j$ , and the fixed covariate  $c_i$  is sampled from

Table 2. Performance of the Bayesian Estimates of Structural Parameters in the Simulation Study with Type I and Type II Prior Inputs, Under the Nonignorable Missing Mechanism and Setting A

Par.	Type I		Type II		Par.	Type I		Type II	
	Bias	RMS	Bias	RMS		Bias	RMS	Bias	RMS
$\mu_1$	0.037	0.118	0.027	0.115	$\lambda_{52}$	0.036	0.050	0.031	0.051
$\mu_2$	-0.024	0.091	0.028	0.084	$\lambda_{73}$	0.036	0.060	0.033	0.059
$\mu_3$	-0.020	0.039	0.035	0.089	$\lambda_{94}$	0.038	0.079	0.022	0.065
$\mu_4$	-0.009	0.082	0.003	0.042	$\lambda_{11,5}$	0.121	0.187	0.078	0.190
$\mu_5$	-0.006	0.085	0.002	0.031	$\lambda_{12,5}$	0.098	0.184	0.049	0.192
$\mu_6$	-0.010	0.071	0.001	0.038	$\lambda_{14,6}$	0.109	0.194	0.045	0.185
$\mu_7$	-0.005	0.074	0.003	0.030	$\lambda_{15,6}$	0.100	0.174	0.029	0.154
$\mu_8$	-0.003	0.068	0.007	0.043	$\psi_{\epsilon 4}$	0.047	0.059	0.064	0.069
$\mu_9$	-0.002	0.032	0.006	0.041	$\psi_{\epsilon 5}$	-0.016	0.029	-0.001	0.020
$\mu_{10,1}$	-0.033	0.082	0.013	0.094	$\psi_{\epsilon 6}$	0.043	0.070	0.064	0.077
$\mu_{10,2}$	-0.038	0.085	0.008	0.089	$\psi_{\epsilon 7}$	-0.018	0.038	-0.003	0.027
$\mu_{11,1}$	0.013	0.071	0.029	0.075	$\psi_{\epsilon 8}$	0.030	0.082	0.037	0.065
$\mu_{11,2}$	0.005	0.074	0.021	0.075	$\psi_{\epsilon 9}$	-0.030	0.060	-0.002	0.038
$\mu_{12,1}$	0.006	0.062	0.022	0.068	$a^*$	0.026	0.107	0.030	0.125
$\mu_{12,2}$	0.006	0.068	0.021	0.073	$\gamma_1$	0.018	0.063	0.053	0.090
$\mu_{13,1}$	0.032	0.081	0.029	0.100	$\gamma_2$	0.037	0.085	0.085	0.118
$\mu_{13,2}$	-0.034	0.079	0.026	0.094	$\gamma_3$	-0.021	0.079	-0.042	0.104
$\mu_{14,1}$	0.011	0.073	0.026	0.073	$\gamma_4$	0.089	0.155	0.100	0.171
$\mu_{14,2}$	0.007	0.075	0.022	0.075	$\gamma_5$	0.104	0.190	0.107	0.223
$\mu_{15,1}$	0.006	0.073	0.023	0.079	$\gamma_6$	-0.037	0.118	-0.040	0.121
$\mu_{15,2}$	-0.001	0.076	0.016	0.079	$\gamma_7$	0.086	0.159	0.067	0.152
$a_1$	-0.016	0.073	-0.045	0.090	$\phi_{11}$	-0.063	0.089	-0.068	0.088
$a_2$	-0.011	0.054	-0.008	0.053	$\phi_{21}$	-0.008	0.035	-0.003	0.036
$a_3$	-0.008	0.066	-0.005	0.066	$\phi_{22}$	-0.038	0.097	-0.057	0.095
$a_4$	-0.006	0.036	-0.004	0.035	$\phi_{31}$	-0.005	0.037	-0.004	0.038
$a_5$	-0.004	0.028	-0.002	0.028	$\phi_{32}$	0.003	0.041	0.000	0.042
$a_6$	-0.005	0.041	-0.002	0.040	$\phi_{33}$	-0.027	0.105	-0.022	0.088
$a_7$	0.000	0.029	0.003	0.029	$\phi_{41}$	-0.019	0.055	-0.008	0.053
$a_8$	0.004	0.040	0.006	0.040	$\phi_{42}$	-0.015	0.053	-0.005	0.051
$a_9$	0.001	0.034	0.003	0.033	$\phi_{43}$	-0.013	0.053	-0.003	0.052
$a_{10}$	-0.014	0.056	-0.004	0.059	$\phi_{44}$	-0.130	0.205	-0.031	0.221
$a_{11}$	0.007	0.051	0.004	0.050	$\phi_{51}$	-0.018	0.047	0.004	0.053
$a_{12}$	0.002	0.046	0.000	0.045	$\phi_{52}$	-0.016	0.054	0.003	0.058
$a_{13}$	-0.016	0.058	0.002	0.067	$\phi_{53}$	-0.017	0.058	-0.003	0.057
$a_{14}$	0.014	0.061	0.010	0.058	$\phi_{54}$	-0.024	0.054	0.011	0.069
$a_{15}$	0.006	0.049	0.004	0.050	$\phi_{55}$	-0.126	0.232	0.021	0.241
$\lambda_{21}$	-0.021	0.111	-0.126	0.163	$\psi_{\delta}$	0.007	0.084	0.108	0.142
$\lambda_{31}$	-0.016	0.119	-0.118	0.160					

Bernoulli(0.7). The true values of the parameters in this equation are given by:  $a^* = 0.3$ ,  $\gamma_1 = 0.2$ ,  $\gamma_2 = 0.3$ ,  $\gamma_3 = -0.3$ ,  $\gamma_4 = 0.5$ ,  $\gamma_5 = 0.7$ ,  $\gamma_6 = -0.2$ , and  $\gamma_7 = 0.4$ .

We first generate the complete data set  $\{d_{ij} : i = 1, \dots, n, j = 1, \dots, p\}$  via equations (1), (2), (3), and (4) with  $n = 1400$  and  $p = 15$ . Then the missing data are then generated from the missingness mechanism defined by (6) and (7) with the following two settings. Setting A: From the last six hundred observations with  $i = 801, \dots, 1400$ , determining whether  $d_{ij}$  is missing or not via the missingness mechanism in (6) and (7) with the true values of  $\varphi_0 = -1.0$ ,  $\varphi_1 = \dots = \varphi_{15} = 0.3$ . The missing proportion is about 0.3. Setting B: For all  $\mathbf{d}_i$ ,  $i = 1, \dots, 1400$ , determining whether  $d_{ij}$  is missing or not via the missingness mech-

anism in (6) and (7) with the true values of  $\varphi_0 = -2.3$ ,  $\varphi_1 = \dots = \varphi_{15} = 0.3$ . The missing proportion is slightly larger than 0.5.

The total number of unknown parameters is 90. Bayesian estimates in 100 replications are obtained with the following two different prior inputs in the conjugate prior distributions. Type I: The hyper-parameter values of  $\boldsymbol{\mu}_0$ ,  $\boldsymbol{\Lambda}_{0j}$ ,  $\boldsymbol{\Pi}_{0j}$ , and  $\boldsymbol{\varphi}_0$  are taken to be their corresponding true values;  $\boldsymbol{\Sigma}_0$ ,  $\mathbf{H}_{0j}$ ,  $\mathbf{H}_{0\delta j}$ ,  $\mathbf{C}$  are identity matrices with appropriate dimensions,  $\alpha_{0j} = \alpha_{0\delta j} = 9$ ,  $\beta_{0j} = \beta_{0\delta j} = 4$ ,  $\rho_0 = 12$ , and  $\mathbf{R}_0^{-1} = 6\boldsymbol{\Phi}$ ; Type II:  $\boldsymbol{\mu}_0 = \mathbf{1}$ ,  $\boldsymbol{\Lambda}_{0j}$ ,  $\boldsymbol{\Pi}_{0j}$ ,  $\boldsymbol{\varphi}_0$ , are fixed at vector  $\mathbf{0}$ ;  $\boldsymbol{\Sigma}_0$ ,  $\mathbf{H}_{0j}$ ,  $\mathbf{H}_{0\delta j}$ ,  $\mathbf{C}$  are two times identity matrices with appropriate dimensions,  $\alpha_{0j} = 5$ ,  $\beta_{0j} = 6$ ,  $\alpha_{0\delta j} = 10$ ,  $\beta_{0\delta j} = 5$ ,  $\rho_0 = 10$ , and  $\mathbf{R}_0^{-1} = 4\mathbf{I}$ .

Table 3. Performance of the Bayesian Estimates of Structural Parameters in the Simulation Study with Type I and Type II Prior Inputs, Under the Nonignorable Missing Mechanism and Setting B

Par.	Type I		Type II		Par.	Type I		Type II	
	Bias	RMS	Bias	RMS		Bias	RMS	Bias	RMS
$\mu_1$	-0.012	0.118	0.077	0.158	$\lambda_{52}$	0.046	0.064	0.032	0.047
$\mu_2$	0.010	0.099	0.059	0.110	$\lambda_{73}$	0.045	0.080	0.033	0.054
$\mu_3$	-0.009	0.104	0.041	0.101	$\lambda_{94}$	0.061	0.101	0.028	0.064
$\mu_4$	-0.019	0.051	-0.006	0.048	$\lambda_{11,5}$	0.147	0.228	0.120	0.277
$\mu_5$	-0.002	0.031	0.004	0.032	$\lambda_{12,5}$	0.157	0.234	0.112	0.227
$\mu_6$	-0.009	0.051	0.005	0.049	$\lambda_{14,6}$	0.157	0.243	0.111	0.270
$\mu_7$	-0.004	0.037	0.004	0.037	$\lambda_{15,6}$	0.160	0.232	0.093	0.250
$\mu_8$	-0.013	0.054	-0.001	0.053	$\psi_{\epsilon 4}$	0.067	0.082	0.091	0.097
$\mu_9$	-0.003	0.046	0.002	0.047	$\psi_{\epsilon 5}$	-0.014	0.030	0.019	0.029
$\mu_{10,1}$	-0.051	0.093	0.020	0.107	$\psi_{\epsilon 6}$	0.061	0.090	0.092	0.105
$\mu_{10,2}$	-0.044	0.089	0.026	0.114	$\psi_{\epsilon 7}$	-0.018	0.048	0.015	0.039
$\mu_{11,1}$	0.034	0.104	0.057	0.119	$\psi_{\epsilon 8}$	0.044	0.091	0.064	0.083
$\mu_{11,2}$	0.033	0.108	0.056	0.126	$\psi_{\epsilon 9}$	-0.029	0.070	-0.003	0.121
$\mu_{12,1}$	0.027	0.095	0.051	0.110	$a^*$	-0.001	0.127	0.070	0.128
$\mu_{12,2}$	0.025	0.092	0.049	0.108	$\gamma_1$	0.008	0.089	0.086	0.140
$\mu_{13,1}$	-0.059	0.102	0.019	0.102	$\gamma_2$	0.028	0.103	0.085	0.118
$\mu_{13,2}$	-0.061	0.105	0.016	0.096	$\gamma_3$	-0.028	0.106	-0.065	0.133
$\mu_{14,1}$	0.033	0.103	0.055	0.125	$\gamma_4$	0.102	0.183	0.106	0.201
$\mu_{14,2}$	0.039	0.110	0.059	0.127	$\gamma_5$	0.113	0.213	0.072	0.206
$\mu_{15,1}$	0.029	0.099	0.048	0.114	$\gamma_6$	-0.029	0.138	-0.030	0.162
$\mu_{15,2}$	0.020	0.092	0.035	0.108	$\gamma_7$	0.102	0.187	0.061	0.171
$a_1$	-0.003	0.074	-0.052	0.110	$\phi_{11}$	-0.085	0.111	-0.090	0.115
$a_2$	-0.019	0.075	-0.010	0.071	$\phi_{21}$	-0.016	0.053	-0.007	0.049
$a_3$	-0.007	0.077	0.001	0.079	$\phi_{22}$	-0.066	0.119	-0.073	0.116
$a_4$	0.007	0.036	0.008	0.035	$\phi_{31}$	-0.017	0.047	-0.008	0.047
$a_5$	-0.001	0.031	0.001	0.031	$\phi_{32}$	-0.010	0.057	0.001	0.055
$a_6$	-0.002	0.042	0.001	0.041	$\phi_{33}$	-0.065	0.134	-0.059	0.115
$a_7$	-0.004	0.033	-0.002	0.033	$\phi_{41}$	-0.023	0.064	-0.004	0.065
$a_8$	-0.008	0.043	-0.006	0.044	$\phi_{42}$	-0.018	0.063	0.012	0.068
$a_9$	0.002	0.036	0.004	0.036	$\phi_{43}$	-0.025	0.065	-0.004	0.069
$a_{10}$	-0.017	0.059	-0.002	0.067	$\phi_{44}$	-0.171	0.258	-0.034	0.289
$a_{11}$	0.019	0.064	0.014	0.061	$\phi_{51}$	-0.020	0.056	0.009	0.060
$a_{12}$	0.015	0.055	0.011	0.054	$\phi_{52}$	-0.026	0.058	0.006	0.060
$a_{13}$	-0.017	0.061	0.002	0.069	$\phi_{53}$	-0.023	0.058	-0.005	0.067
$a_{14}$	0.017	0.066	0.014	0.064	$\phi_{54}$	-0.030	0.064	0.010	0.076
$a_{15}$	0.013	0.054	0.013	0.056	$\phi_{55}$	-0.186	0.275	-0.004	0.295
$\lambda_{21}$	0.050	0.202	-0.138	0.223	$\psi_{\delta}$	0.009	0.071	0.131	0.153
$\lambda_{31}$	0.003	0.152	-0.166	0.217					

Based on some pilot test runs in getting some idea about the required number of iterations for convergence, we took  $T = 20,000$  simulated observations after 20,000 burn-ins to get the estimates in each replication. Under setting A, the bias (Bias) and root mean squares (RMS) values corresponding to the parameters in the NSEM are presented in Table 2. We observed that most of these values are reasonably small. We also observed the Bias and RMS corresponding to the estimates of the nuisance parameters in  $\varphi$  are small; details are not presented to save space. As expected, the Bias and RMS values of some parameters that are associated with the ordered and unordered categorical variables, for example,  $\lambda_{11,5}$ ,  $\lambda_{12,5}$ ,  $\lambda_{14,6}$ ,  $\lambda_{15,6}$ ,  $\gamma_4$ ,  $\gamma_5$ ,  $\phi_{44}$ , and  $\phi_{55}$ , are larger. This phenomenon may be due to the

discrete variables provide less information than the continuous variables. Moreover, we learned that under the selected sample size the different prior inputs have minor effects on the parameter estimates. Results obtained with more severe misspecification of the prior inputs are expected to be worse. However, as Bayesian estimation is not very sensitive to prior inputs for reasonably large sample sizes [14], we expect that the problem is not serious. For setting B, the overall performance is similar to that obtained with setting A; although as expected, it is slightly worse since the missing proportion is larger. For completeness, Bias and RMS values corresponding to the estimates of the parameters in the NSEM are presented in Table 3.

Table 4. Performance of the Bayesian Estimates of Structural Parameters in the Simulation Study with Type I and Type II Prior Inputs, Under the Ignorable Missing Mechanism and Setting A

Par.	Type I		Type II		Par.	Type I		Type II	
	Bias	RMS	Bias	RMS		Bias	RMS	Bias	RMS
$\mu_1$	-0.187	0.221	-0.211	0.233	$\lambda_{52}$	0.035	0.049	0.035	0.048
$\mu_2$	-0.144	0.169	-0.165	0.193	$\lambda_{73}$	0.035	0.058	0.046	0.057
$\mu_3$	-0.140	0.170	-0.175	0.209	$\lambda_{94}$	0.030	0.077	0.024	0.051
$\mu_4$	-0.098	0.106	-0.142	0.152	$\lambda_{11,5}$	0.165	0.245	0.112	0.166
$\mu_5$	-0.072	0.079	-0.102	0.118	$\lambda_{12,5}$	0.127	0.200	0.091	0.191
$\mu_6$	-0.104	0.110	-0.131	0.140	$\lambda_{14,6}$	0.136	0.236	0.116	0.206
$\mu_7$	-0.077	0.083	-0.113	0.119	$\lambda_{15,6}$	0.102	0.185	0.089	0.190
$\mu_8$	-0.092	0.100	-0.126	0.132	$\psi_{\epsilon 4}$	0.046	0.058	0.064	0.071
$\mu_9$	-0.077	0.086	-0.118	0.126	$\psi_{\epsilon 5}$	-0.016	0.030	-0.004	0.021
$\mu_{10,1}$	-0.121	0.137	-0.144	0.164	$\psi_{\epsilon 6}$	0.038	0.065	0.063	0.075
$\mu_{10,2}$	-0.144	0.157	-0.155	0.173	$\psi_{\epsilon 7}$	-0.017	0.039	-0.009	0.027
$\mu_{11,1}$	-0.065	0.091	-0.108	0.129	$\psi_{\epsilon 8}$	0.019	0.079	0.036	0.064
$\mu_{11,2}$	-0.091	0.112	-0.140	0.161	$\psi_{\epsilon 9}$	-0.024	0.059	-0.008	0.032
$\mu_{12,1}$	-0.058	0.081	-0.099	0.129	$a^*$	0.029	0.116	0.040	0.136
$\mu_{12,2}$	-0.077	0.099	-0.105	0.135	$\gamma_1$	0.015	0.063	0.049	0.083
$\mu_{13,1}$	-0.115	0.135	-0.150	0.172	$\gamma_2$	0.059	0.100	0.074	0.101
$\mu_{13,2}$	-0.136	0.150	-0.178	0.192	$\gamma_3$	-0.069	0.113	-0.059	0.107
$\mu_{14,1}$	-0.068	0.095	-0.105	0.131	$\gamma_4$	0.134	0.192	0.139	0.208
$\mu_{14,2}$	-0.089	0.113	-0.115	0.143	$\gamma_5$	0.092	0.176	0.081	0.178
$\mu_{15,1}$	-0.062	0.092	-0.093	0.130	$\gamma_6$	-0.025	0.111	-0.028	0.093
$\mu_{15,2}$	-0.086	0.114	-0.128	0.154	$\gamma_7$	0.106	0.170	0.116	0.187
$a_1$	-0.006	0.071	-0.004	0.079	$\phi_{11}$	-0.054	0.084	-0.069	0.096
$a_2$	-0.002	0.052	0.016	0.054	$\phi_{21}$	0.004	0.037	-0.004	0.032
$a_3$	0.002	0.066	0.010	0.063	$\phi_{22}$	-0.034	0.093	-0.059	0.098
$a_4$	-0.001	0.038	-0.002	0.041	$\phi_{31}$	0.006	0.039	0.009	0.041
$a_5$	0.000	0.028	-0.004	0.029	$\phi_{32}$	0.013	0.044	0.007	0.046
$a_6$	0.001	0.042	0.009	0.043	$\phi_{33}$	-0.012	0.100	-0.020	0.088
$a_7$	0.005	0.030	0.008	0.035	$\phi_{41}$	-0.014	0.047	-0.014	0.052
$a_8$	0.007	0.041	0.004	0.045	$\phi_{42}$	-0.008	0.048	-0.012	0.054
$a_9$	0.003	0.036	0.002	0.039	$\phi_{43}$	-0.002	0.051	0.000	0.048
$a_{10}$	-0.010	0.054	0.006	0.065	$\phi_{44}$	-0.149	0.255	-0.102	0.259
$a_{11}$	0.015	0.053	0.009	0.053	$\phi_{51}$	-0.002	0.049	0.000	0.057
$a_{12}$	0.009	0.046	0.018	0.059	$\phi_{52}$	-0.002	0.055	0.003	0.062
$a_{13}$	-0.007	0.054	0.001	0.055	$\phi_{53}$	-0.005	0.057	-0.015	0.062
$a_{14}$	0.020	0.063	0.023	0.049	$\phi_{54}$	-0.013	0.058	-0.009	0.059
$a_{15}$	0.012	0.051	0.012	0.043	$\phi_{55}$	-0.112	0.249	-0.062	0.270
$\lambda_{21}$	-0.025	0.109	-0.057	0.126	$\psi_{\delta}$	-0.003	0.069	0.071	0.097
$\lambda_{31}$	-0.015	0.114	-0.066	0.116					

To study whether the missing data generated from non-ignorable missing mechanism can be treated as MAR, the same simulated data sets are reanalyzed under the MAR assumption. The corresponding simulation results are presented in Tables 4 and 5, respectively. From these tables, we observe that the Bias and RMS values of the parameter estimates (particular those in  $\mu$ ) are generally larger than those in Tables 2 and 3. These results indicate that the performance under the MAR assumption is inferior.

To illustrate the application of the Bayes factor, we consider the comparison of the following models for the missing mechanism:

$$M_0: \text{logit}\{p(r_{ij} = 1|\mathbf{d}_i, \varphi)\} = \varphi_0 + \varphi_1 z_{i1} + \cdots + \varphi_3 z_{i3} + \varphi_4 y_{i4} + \cdots + \varphi_9 y_{i9} + \varphi_{10} u_{i1} + \cdots + \varphi_{15} u_{i6},$$

$$M_1: \text{logit}\{p(r_{ij} = 1|\mathbf{d}_i, \varphi)\} = \varphi_0 + \varphi_1 z_{i1} + \cdots + \varphi_3 z_{i3} + \varphi_4 y_{i4} + \cdots + \varphi_9 y_{i9},$$

$$M_2: \text{logit}\{p(r_{ij} = 1|\mathbf{d}_i, \varphi)\} = \varphi_0 + \varphi_{10} u_{i1} + \cdots + \varphi_{15} u_{i6},$$

$$M_3: \text{logit}\{p(r_{ij} = 1|\mathbf{d}_i, \varphi)\} = \varphi_0,$$

where  $M_0$  is the model that involves all the manifest variables,  $M_1$  only involves the ordered categorical and continuous variables,  $M_2$  only involves the unordered categorical variables, and  $M_3$  is MAR. The rationale for considering these missing mechanisms is to examine the impact of the phenotype and genotype variables. For example, the rationale for considering  $M_1$  is to examine whether the missingness depends on the genotype variables or not. In addition to comparing the models with different missing mechanisms, we also consider a competing model with missing mechanism



Table 5. Performance of the Bayesian Estimates of Structural Parameters in the Simulation Study with Type I and Type II Prior Inputs, Under the Ignorable Missing Mechanism and Setting B

Par.	Type I		Type II		Par.	Type I		Type II	
	Bias	RMS	Bias	RMS		Bias	RMS	Bias	RMS
$\mu_1$	-0.341	0.368	-0.298	0.329	$\lambda_{52}$	0.050	0.065	0.031	0.045
$\mu_2$	-0.256	0.275	-0.193	0.213	$\lambda_{73}$	0.055	0.090	0.034	0.056
$\mu_3$	-0.265	0.281	-0.204	0.223	$\lambda_{94}$	0.063	0.113	0.029	0.064
$\mu_4$	-0.271	0.275	-0.265	0.268	$\lambda_{11,5}$	0.300	0.411	0.132	0.283
$\mu_5$	-0.185	0.187	-0.182	0.184	$\lambda_{12,5}$	0.252	0.353	0.128	0.242
$\mu_6$	-0.282	0.285	-0.274	0.277	$\lambda_{14,6}$	0.296	0.437	0.125	0.308
$\mu_7$	-0.201	0.204	-0.197	0.200	$\lambda_{15,6}$	0.242	0.370	0.099	0.251
$\mu_8$	-0.291	0.294	-0.001	0.053	$\psi_{\epsilon 4}$	0.070	0.079	0.094	0.100
$\mu_9$	-0.235	0.238	0.002	0.047	$\psi_{\epsilon 5}$	-0.013	0.029	0.019	0.028
$\mu_{10,1}$	-0.207	0.223	0.020	0.107	$\psi_{\epsilon 6}$	0.068	0.098	0.097	0.108
$\mu_{10,2}$	-0.285	0.295	0.026	0.114	$\psi_{\epsilon 7}$	-0.021	0.054	0.015	0.039
$\mu_{11,1}$	-0.143	0.163	0.057	0.119	$\psi_{\epsilon 8}$	0.049	0.097	0.067	0.085
$\mu_{11,2}$	-0.228	0.242	0.056	0.126	$\psi_{\epsilon 9}$	-0.033	0.074	0.017	0.047
$\mu_{12,1}$	-0.124	0.146	0.051	0.110	$a^*$	0.012	0.133	0.006	0.150
$\mu_{12,2}$	-0.210	0.224	0.049	0.108	$\gamma_1$	0.015	0.093	0.053	0.122
$\mu_{13,1}$	-0.216	0.231	0.019	0.102	$\gamma_2$	0.080	0.135	0.140	0.198
$\mu_{13,2}$	-0.300	0.312	0.016	0.096	$\gamma_3$	-0.150	0.192	-0.199	0.240
$\mu_{14,1}$	-0.153	0.172	0.055	0.125	$\gamma_4$	0.215	0.288	0.200	0.288
$\mu_{14,2}$	-0.228	0.242	0.059	0.127	$\gamma_5$	0.091	0.247	0.032	0.236
$\mu_{15,1}$	-0.135	0.157	0.048	0.114	$\gamma_6$	-0.070	0.221	-0.053	0.197
$\mu_{15,2}$	-0.227	0.239	0.035	0.108	$\gamma_7$	0.172	0.300	0.101	0.260
$a_1$	0.054	0.091	0.022	0.091	$\phi_{11}$	-0.194	0.204	-0.204	0.212
$a_2$	0.036	0.079	0.042	0.081	$\phi_{21}$	-0.145	0.152	-0.143	0.149
$a_3$	0.043	0.087	0.047	0.088	$\phi_{22}$	-0.191	0.212	-0.204	0.216
$a_4$	0.045	0.057	0.045	0.057	$\phi_{31}$	-0.138	0.144	-0.139	0.145
$a_5$	0.028	0.042	0.028	0.043	$\phi_{32}$	-0.141	0.150	-0.143	0.152
$a_6$	0.040	0.059	0.040	0.058	$\phi_{33}$	-0.174	0.205	-0.177	0.194
$a_7$	0.028	0.043	0.028	0.043	$\phi_{41}$	-0.152	0.158	-0.142	0.152
$a_8$	0.031	0.056	0.032	0.057	$\phi_{42}$	-0.153	0.162	-0.136	0.149
$a_9$	0.035	0.052	0.035	0.052	$\phi_{43}$	-0.154	0.160	-0.148	0.157
$a_{10}$	0.020	0.063	0.037	0.084	$\phi_{44}$	-0.352	0.408	-0.158	0.378
$a_{11}$	0.070	0.100	0.056	0.088	$\phi_{51}$	-0.147	0.153	-0.134	0.144
$a_{12}$	0.057	0.082	0.050	0.075	$\phi_{52}$	-0.162	0.169	-0.153	0.163
$a_{13}$	0.024	0.063	0.045	0.088	$\phi_{53}$	-0.158	0.165	-0.159	0.168
$a_{14}$	0.064	0.096	0.053	0.087	$\phi_{54}$	-0.162	0.169	-0.139	0.152
$a_{15}$	0.056	0.083	0.049	0.076	$\phi_{55}$	-0.341	0.396	-0.106	0.418
$\lambda_{21}$	0.007	0.204	-0.151	0.219	$\psi_{\delta}$	0.010	0.073	0.110	0.146
$\lambda_{31}$	-0.026	0.179	-0.165	0.222					

$M_0$  and with the following structural equation:

$$M_4: \eta_i = a^* c_i + \gamma_1 \xi_{i1} + \dots + \gamma_5 \xi_{i5} + \delta_i.$$

In the path sampling procedure for computing the Bayes factors, we took  $L = 20$ , and  $J = 4,000$  collected after 1,000 burn-ins. In setting A, we observe that the Bayes factor chooses the true model  $M_0$  in all the replications. In setting B, the Bayes factor chooses the true model  $M_0$  when comparing with  $M_1$ ,  $M_2$ , and  $M_3$ . Hence, it always selects the true missing mechanism. When comparing  $M_0$  with  $M_4$ , the Bayes factor chooses the false model  $M_4$  six and three times out of 100 replications with Type I and Type II prior inputs, respectively. This phenomenon may be due to the large amount of missing data in setting B, which gives in-

correct results in selecting the true structural equation in a small number of replications. The summary statistic of the computed logarithm Bayes factors for comparing the above models with Type I and Type II prior inputs of two settings are presented in Table 6. It seems that the Bayes factor is not sensitive to the selected prior inputs under the selected sample sizes.

## 5. ANALYSIS OF THE CARDIOVASCULAR DISEASE DATA

The proposed model and the Bayesian methods are illustrated by an application to examine the inter-relationships of phenotype and genotype variables in relation to cardio-

Table 6. Summary of the Estimated Twice Logarithm Bayes Factors in the Simulation Study with Type I and Type II Prior Inputs, Under Settings A and B

	Setting A				Setting B			
	Type I		Type II		Type I		Type II	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$2 \log B_{01}$	563.06	129.91	417.94	97.96	1382.38	234.41	679.54	138.39
$2 \log B_{02}$	420.45	65.97	350.31	80.61	798.24	94.58	674.53	77.36
$2 \log B_{03}$	1246.67	133.32	1049.45	109.92	2781.07	967.36	1730.48	130.12
$2 \log B_{04}$	13.89	6.93	13.36	5.65	7.54	5.39	5.92	3.39

vascular disease of diabetes patients as mentioned in the ‘Introduction’. Data are obtained from patients who participated in an applied genomic program conducted by the Institute of Diabetes at the Chinese University of Hong Kong. They underwent a comprehensive assessment of complications and risk factors based on the European DiabCare protocol. The whole data set involves measurements of many phenotype variables, such as those related to age at diagnosis, duration of disease, SBP, DBP, HDL, TG, and existence of symptoms, etc; as well as a panel of 65 biallelic single nucleotide polymorphisms (SNPs) in 36 genes from biochemical pathways implicated in the development of cardiovascular disease. We are interested in incorporating multiple genetic and phenotypic pathways to understand more about the cardiovascular disease. The endogenous variable of cardiovascular disease was treated as a latent variable reflected by two manifest dichotomous variables which are related to the existence of the following symptoms: IHD and CVA. Based on some preliminary data analysis and medical knowledge, several continuous phenotype variables and several unordered categorical genotype variables are used to formulate phenotype and genotype exogenous latent variables. The phenotype variables are: WHR, BMI, SBP, DBP, HDL, and ln TG (the logarithm transformation is used for achieving a distribution that is close to normal) while the genotype variables are PON11, PON12, FGB, and SELE2. Note that regressing IHD (or CVA) on the above 10 manifest variables with a single multiple regression model would result in a large number of linear and interaction terms that induce difficulties in formulating the model and achieving meaningful medical interpretations. Moreover, we are interested in the latent variable about ‘cardiovascular disease’ which should be measured by IHD and CVA, rather than by a single manifest variable. Hence, an NSEM with mixed continuous, ordered categorical (dichotomous), and unordered categorical variables is needed.

In this application, we consider manifest random observations which are related to eight phenotype variables and four genotype variables, namely:  $\mathbf{d} = (\text{IHD}, \text{CVA}, \text{WHR}, \text{BMI}, \text{SBP}, \text{DBP}, \text{ln TG}, \text{HDL}, \text{PON11}, \text{PON12}, \text{SELE2}, \text{FGB})'$ . Here,  $r_1 = 2$ ,  $r_2 = r_1 + 6 = 8$ , and  $s = 4$ . The sample size is 1385. We encountered a number of missing entries (given in parentheses) in WHR(6), BMI (6), ln TG (10), HDL (21), PON11 (20), PON12 (37), SELE2 (12), and

FGB (25). As we expect that the reason for missingness depends on the missing values themselves, we treat the missing data as nonignorable. The frequencies of ‘Yes=1/No=0’ corresponding to the dichotomous variables IHD and CVA are {111,1274} and {117,1268}, respectively. The sample mean and standard deviation (computed based on complete data) of {WHR, BMI, SBP, DBP, ln TG, HDL} are {0.89, 24.77, 138.16, 78.91, 0.37, 1.27} and {0.07, 3.72, 22.79, 11.40, 0.64, 0.36}, respectively. In the analysis, continuous data were standardized to unify the scale. The frequencies of {AA, Aa, aa} corresponding to PON11, PON12, SELE2, and FGB are {0, 70, 1295}, {167, 608, 573}, {1267, 103, 3}, and {724, 522, 114}, respectively.

The proposed NSEM and the following nonignorable missing mechanism was used to analyze this data set with nonignorable missing data:

$$(13) \quad \begin{aligned} & \text{logit}\{p(r_{ij} = 1 | \mathbf{d}_i, \boldsymbol{\varphi})\} \\ & = \varphi_0 + \varphi_1 z_{i1} + \varphi_2 z_{i2} + \varphi_3 y_{i3} + \cdots + \varphi_8 y_{i8} + \varphi_9 u_{i1} \\ & \quad + \cdots + \varphi_{12} u_{i4}. \end{aligned}$$

The measurement equation of the proposed NSEM is given by (3), where the the fixed covariates  $\mathbf{b}_i = b_i$  is the ‘duration of disease’, and  $\boldsymbol{\Lambda}^* = (\boldsymbol{\Lambda}'_1 \cdots \boldsymbol{\Lambda}'_{12})'$  are defined by

$$(14) \quad \boldsymbol{\Lambda}^* = \begin{pmatrix} 1.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ \lambda_{21} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 1.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & \lambda_{42} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 1.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & \lambda_{63} & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & \lambda_{84} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & \lambda_{10,5} & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & \lambda_{12,6} \end{pmatrix}.$$

The covariate  $b_i$  is included in examining the effects of the duration of disease to the phenotype and genotype observed variables. Here, the parameters with an asterisk are treated as fixed. Based on the medical meaning of the manifest phenotype and genotype variables, this non-overlapping structure of  $\boldsymbol{\Lambda}^*$  is proposed for achieving clear interpretation

of latent variables. The structure of  $\mathbf{\Lambda}$  (corresponding to  $\mathbf{\Lambda}^*$ ) provides four phenotype latent variables: ‘cardiovascular disease,  $\eta$ ’, ‘body shape,  $\xi_1$ ’, ‘blood pressure,  $\xi_2$ ’, and ‘lipid control,  $\xi_3$ ’; and two genotype latent variables ‘gene-inflammatory,  $\xi_4$ ’ and ‘gene-lipid control,  $\xi_5$ ’. While the  $\mathbf{\Lambda}^*$  (equivalently  $\mathbf{\Lambda}$ ) defined by (14) provides a natural and reasonable formulation of the measurement equation, there are many choices of the structural equation that relates  $\eta$  with various linear and nonlinear terms of  $\xi_1, \dots, \xi_5$ , as well as some covariates. We have compared a number of competing models through the Bayes factors that were computed by the path sampling; under the missing mechanism as defined by (13), and the following hyper-parameter values  $\alpha_{0j} = \alpha_{0\delta j} = 5$ ,  $\beta_{0j} = \beta_{0\delta j} = 6$ ,  $\rho_0 = 8$ ,  $\mathbf{R}_0^{-1} = 2\mathbf{I}$ ,  $\mathbf{\Sigma}_0$ ,  $\mathbf{H}_{0j}$ ,  $\mathbf{H}_{0\delta j}$ , and  $\mathbf{C}$  are fixed at two times identity matrices with appropriate dimensions,  $\boldsymbol{\mu}_0$ ,  $\mathbf{\Lambda}_{0j}$ ,  $\mathbf{\Pi}_{0j}$ , and  $\boldsymbol{\varphi}_0$  are fixed at vector  $\mathbf{0}$ . Among the competing models, we finally selected the model with the following structural equation (as the comparison is rather standard, we ignore the details):

$$(15) \quad \eta_i = a^* c_i + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \gamma_4 \xi_{i4} + \gamma_5 \xi_{i5} \\ + \gamma_6 \xi_{i2} \xi_{i4} + \gamma_7 \xi_{i3} \xi_{i5} + \delta_i,$$

where  $c_i$  is a fixed covariate about the smoking status (Yes=1/No=0). In addition to the linear effects of the phenotype and genotype latent variables, this structural equation involves two pheno-genotype latent variable interactions  $\{\xi_2 \xi_4, \xi_3 \xi_5\}$ . To illustrate the use of Bayes factor in comparing various missing mechanisms, we consider the following models under the selected NSEM with structural equation (15):

$M_0$ : the missing mechanism defined by (13),

$$M_1: \text{logit}\{p(r_{ij} = 1 | \mathbf{d}_i, \boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 z_{i1} + \varphi_2 z_{i2} + \varphi_3 y_{i3} \\ + \dots + \varphi_8 y_{i8},$$

$$M_2: \text{logit}\{p(r_{ij} = 1 | \mathbf{d}_i, \boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 z_{i1} + \varphi_2 z_{i2},$$

$$M_3: \text{logit}\{p(r_{ij} = 1 | \mathbf{d}_i, \boldsymbol{\varphi})\} = \varphi_0.$$

In the path sampling procedure for computing the Bayes factor, we took  $L = 10$ , and  $J = 10,000$  collected after 4,000 burn-ins. The computed logarithm Bayes factors are  $2\log \widehat{B}_{01} = 16.48$ ,  $2\log \widehat{B}_{02} = 14.15$ ,  $2\log \widehat{B}_{03} = 13.91$ . Hence  $M_0$  is selected.

To give some idea about the convergence of the Gibbs sampler in the analysis, the ‘estimated potential scale reduction (EPSR)’ values [see 7] of the parameter sequences obtained under  $M_0$  using different starting values are displayed in Figure 1. We observe from Figure 1 that the Gibbs sampler converges after 15,000 iterations. Bayesian estimates and their standard error estimates were obtained through 20,000 simulated observations after 15,000 burn-ins.

The path diagram of the selected model  $M_0$  is presented in Figure 2, together with some Bayesian estimates and their standard error estimates. We note that the standard

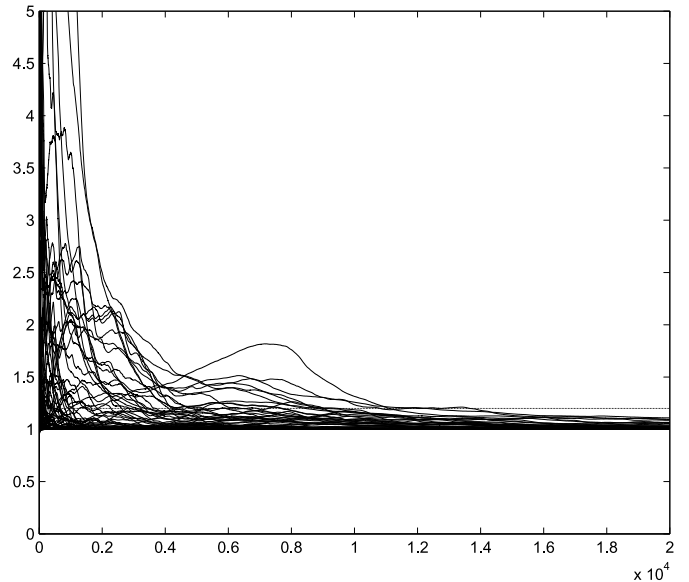


Figure 1. EPSR Values Against the Iterations.

error estimates of parameters that involve unordered categorical variables are substantially larger than the others. This phenomenon may be due to the unordered discrete nature of the genotype variables. The large magnitudes of the  $\lambda$ 's estimates indicate strong associations between the latent variables and their corresponding indicators. The estimated structural equation is equal to

$$\eta = 0.153c^* + 0.237\xi_1 + 0.465\xi_2 + 0.283\xi_3 + 0.505\xi_4 \\ (0.136) \quad (0.155) \quad (0.102) \quad (0.156) \quad (0.481) \\ - 0.300\xi_5 - 0.265\xi_2\xi_4 - 0.312\xi_3\xi_5 + \delta, \\ (0.345) \quad (0.314) \quad (0.385)$$

where the values in parentheses are standard error estimates. Although some of the parameter estimates are not statistically significant, we interpret this equation based on the magnitude of the estimates. The interpretation of this equation is similar to the regression model by considering  $\eta$  as a dependent variable, and  $\xi_1, \dots, \xi_5$  as the independent variables. From the magnitudes of  $\hat{\gamma}_1 = 0.237$ ,  $\hat{\gamma}_2 = 0.465$ , and  $\hat{\gamma}_3 = 0.283$ , it seems that the phenotype latent variable ‘body shape,  $\xi_1$ ’, ‘blood pressure,  $\xi_2$ ’, and ‘lipid control,  $\xi_3$ ’ have substantial direct effects on ‘cardiovascular disease,  $\eta$ ’. From the magnitudes of  $\hat{\gamma}_4 = 0.505$  and  $\hat{\gamma}_5 = -0.300$ , it seems that the genotype variables ‘gene-inflammatory,  $\xi_4$ ’ and ‘gene-lipid control,  $\xi_5$ ’ also have substantial direct effects on ‘cardiovascular disease,  $\eta$ ’. From  $\hat{\gamma}_6 = -0.265$ ,  $\hat{\gamma}_7 = -0.312$ , ‘blood pressure,  $\xi_2$ ’ and ‘gene-inflammatory,  $\xi_4$ ’, as well as ‘lipid control,  $\xi_3$ ’ and ‘gene-lipid control,  $\xi_5$ ’ have negative interactions ( $\xi_2 \xi_4$  and  $\xi_3 \xi_5$ , respectively) on ‘cardiovascular disease,  $\eta$ ’. The basic interpretation is that the additive linear direct effects of the latent variables  $\xi_1, \xi_2, \xi_3, \xi_4$ , and  $\xi_5$  in the structural equation are inadequate to account for their relationships with the dependent latent variable  $\eta$ , and

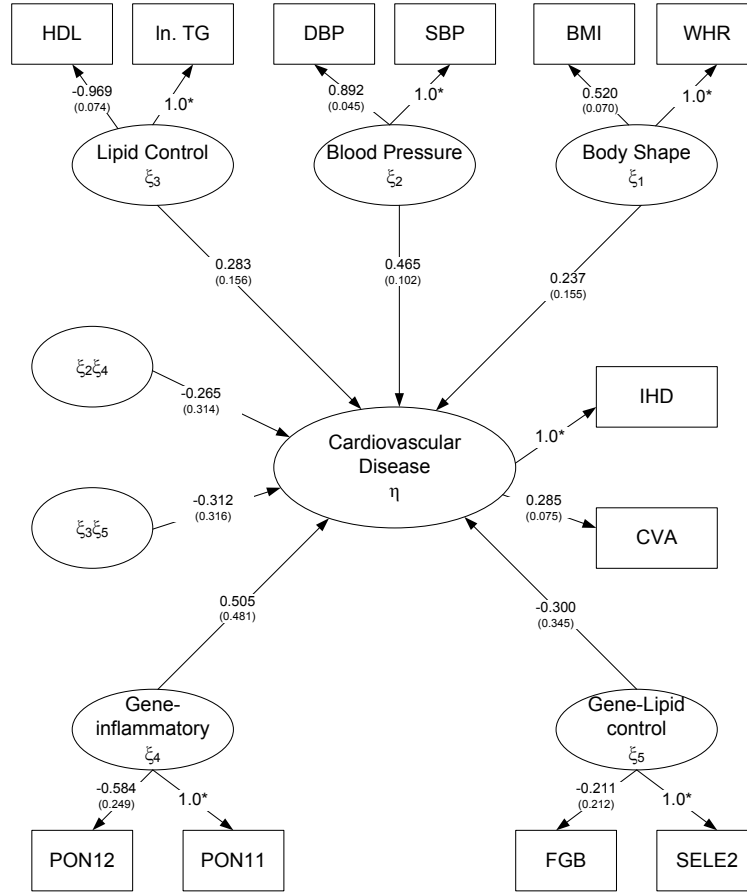


Figure 2. Path Diagram of the Model  $M_0$ .

Table 7. Bayesian Estimates of Other Parameters in Cardiovascular Disease Data, Under  $M_0$  with Nonignorable Missing Mechanism

Par.	Est	SE	Par.	Est	SE	Par.	Est	SE	Par.	Est	SE
$\mu_1$	-2.225	0.143	$a_1$	0.190	0.029	$\psi_{\epsilon 7}$	0.468	0.043	$\phi_{55}$	0.199	0.039
$\mu_2$	-1.473	0.066	$a_2$	0.117	0.050	$\psi_{\epsilon 8}$	0.529	0.042	$\psi_{\delta}$	0.999	0.207
$\mu_3$	-0.004	0.028	$a_3$	-0.004	0.026	$\phi_{11}$	0.454	0.060	$\varphi_0$	-2.217	0.329
$\mu_4$	-0.002	0.027	$a_4$	-0.111	0.027	$\phi_{21}$	0.124	0.027	$\varphi_1$	-0.263	0.324
$\mu_5$	-0.008	0.025	$a_5$	0.108	0.027	$\phi_{22}$	0.693	0.048	$\varphi_2$	-0.422	0.316
$\mu_6$	-0.007	0.025	$a_6$	-0.200	0.027	$\phi_{31}$	0.276	0.026	$\varphi_3$	-0.009	0.110
$\mu_7$	-0.003	0.027	$a_7$	-0.069	0.029	$\phi_{32}$	0.140	0.026	$\varphi_4$	-0.180	0.147
$\mu_8$	-0.014	0.027	$a_8$	0.078	0.028	$\phi_{33}$	0.535	0.048	$\varphi_5$	0.156	0.108
$\mu_{9,1}$	0.394	0.193	$a_9$	0.170	0.165	$\phi_{41}$	0.078	0.038	$\varphi_6$	-0.172	0.111
$\mu_{9,2}$	2.638	0.184	$a_{10}$	0.028	0.035	$\phi_{42}$	-0.008	0.044	$\varphi_7$	-0.203	0.138
$\mu_{10,1}$	0.446	0.053	$a_{11}$	-0.049	0.052	$\phi_{43}$	0.116	0.030	$\varphi_8$	-0.504	0.167
$\mu_{10,2}$	0.407	0.052	$a_{12}$	0.019	0.029	$\phi_{44}$	0.176	0.053	$\varphi_9$	-1.940	0.194
$\mu_{11,1}$	-1.568	0.059	$\psi_{\epsilon 3}$	0.546	0.060	$\phi_{51}$	0.035	0.044	$\varphi_{10}$	0.160	0.208
$\mu_{11,2}$	-2.951	0.161	$\psi_{\epsilon 4}$	0.872	0.038	$\phi_{52}$	0.029	0.038	$\varphi_{11}$	-0.680	0.420
$\mu_{12,1}$	-0.254	0.035	$\psi_{\epsilon 5}$	0.300	0.032	$\phi_{53}$	0.047	0.044	$\varphi_{12}$	0.159	0.181
$\mu_{12,2}$	-1.217	0.051	$\psi_{\epsilon 6}$	0.413	0.030	$\phi_{54}$	-0.012	0.027			

two negative interaction terms ( $-0.265\xi_2\xi_4$  and  $-0.312\xi_3\xi_5$ ) have to be added. Because of the pheno-genotype interactions, the corresponding phenotype and genotype effects should be interpreted together with the interaction effects.

The remaining Bayesian estimates of  $\mu$ ,  $\mathbf{A}$ ,  $\Psi_{\epsilon y}$ ,  $\psi_{\delta}$ ,  $\Phi$ , and  $\varphi$  and their standard error estimates are presented in Table 7. From  $\hat{\phi}_{ij}$ , we observe that the correlations of the phenotype and genotype latent variables are small. From the

estimates of parameters in  $\varphi$ , we observe that it is necessary to incorporate the missing mechanism in analyzing the missing data. This reconfirms the above model comparison results. To save space, straightforward interpretation of the less important parameter estimates is not discussed.

We also analyzed the data set under the MAR assumption ( $M_3$ ). As expected, we observe that the estimates of some parameters are quite different from those in  $M_0$ . For example, the estimates of  $\mu_{9,1}, \lambda_{10,4}, \gamma_4, \gamma_5, \gamma_6, \gamma_7$ , and  $\phi_{43}$  are equal to 0.720, 0.406,  $-0.362$ ,  $-0.515$ , 0.115,  $-0.656$ , and  $-0.138$ , which are very different from the corresponding estimates obtained through the nonignorable missing mechanism. Hence, the conclusion obtained with the incorrect MAR assumption is quite different. For instance, the estimates of the path coefficient corresponding to  $\xi_2\xi_4$  obtained with the nonignorable missing mechanism and the MAR assumption are equal to  $-0.265$  and 0.115, respectively. These different estimates give quite a different conclusion on the effect of phenotype-genotype interaction to cardiovascular disease.

Using a SUN Enterprise 4500 server, the computer time for obtaining the Bayesian estimates of parameters in  $M_0$  in this application is about 120 minutes. Our program is written in C language and is available upon request.

## 6. DISCUSSION

We have proposed a nonlinear structural equation model for analyzing linear and nonlinear relationships of latent variables that are identified from mixed continuous, ordered, and unordered categorical manifest variables. The proposed model also accommodates fixed covariates and nonignorable missing data. Bayesian methods for estimation and model comparison have been presented and applied to study the linear and interaction effects of phenotype and genotype latent variables to the cardiovascular disease of diabetes patients. As the importance of genotype-genotype and genotype-phenotype interactions is being increasingly recognized in the study of pathogenesis of many complex diseases [3], we expect the methodologies presented here will have wide applications in biomedical science. Moreover, as ordered and unordered categorical variables are very common in geosciences, social, and behavioral sciences, these developed methodologies are also useful in these fields.

Like many other statistical models, the proposed NSEM is developed under the assumption that the distributions of the latent variables and the measurement errors are normal. In practice, as the normality assumption for the latent variables is difficult to check, and this assumption may not be valid in some substantive research, it is important to develop some robust methods that are less reliant on this assumption. Recently, a Bayesian semi-parametric approach that utilizes the approximate Dirichlet process has been applied to some special cases of the proposed NSEM [see for example, 16]. Clearly, extension of this robust approach to

the complex SEMs presented here represents an interesting topic of research. Moreover, to enhance our understanding of complex diseases, developing second-order factor models, multiple-group models, and longitudinal models is useful.

## APPENDIX A. FULL CONDITIONAL DISTRIBUTIONS

(a)  $p(\Omega|\mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{D}_{\text{obs}}, \mathbf{r}) = \prod_{i=1}^n p(\omega_i|\mathbf{w}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i)$ . Since  $\omega_i$  is independent with  $\{\boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i\}$  when given  $\{\mathbf{w}_i, \boldsymbol{\theta}\}$ , so  $p(\omega_i|\mathbf{w}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i) = p(\omega_i|\mathbf{w}_i, \boldsymbol{\theta})$ , and

$$(16) \quad \begin{aligned} & p(\omega_i|\mathbf{w}_i, \boldsymbol{\theta}) \\ & \propto p(\mathbf{w}_i|\omega_i, \boldsymbol{\theta})p(\eta_i|\xi_i, \boldsymbol{\theta})p(\xi_i|\boldsymbol{\theta}) \\ & \propto \exp\left\{-\frac{1}{2}(\mathbf{w}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{b}_i - \boldsymbol{\Lambda}\omega_i)' \boldsymbol{\Psi}_\epsilon^{-1}(\mathbf{w}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{b}_i - \boldsymbol{\Lambda}\omega_i) \right. \\ & \quad \left. - \frac{1}{2}(\eta_i - \boldsymbol{\Pi}\mathbf{G}(\omega_i))' \boldsymbol{\Psi}_\delta^{-1}(\eta_i - \boldsymbol{\Pi}\mathbf{G}(\omega_i)) - \frac{1}{2}\xi_i' \boldsymbol{\Phi}^{-1}\xi_i\right\}, \end{aligned}$$

where  $\boldsymbol{\Psi}_\epsilon = \text{diag}(\psi_{\epsilon 1}, \dots, \psi_{\epsilon r_2}, \mathbf{I}, \dots, \mathbf{I})$ .

(b)  $p(\mathbf{Y}_{\text{mis}}^*|\mathbf{Y}_{\text{obs}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{D}_{\text{obs}}, \mathbf{r}) = \prod_{i=1}^n p(\mathbf{y}_{i,\text{mis}}^*|\mathbf{w}_i \setminus \mathbf{y}_{i,\text{mis}}^*, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i)$ , where  $\mathbf{w}_i \setminus \mathbf{y}_{i,\text{mis}}^*$  is the  $\mathbf{w}_i$  with  $\mathbf{y}_{i,\text{mis}}^*$  deleted, and

$$(17) \quad \begin{aligned} & p(\mathbf{y}_{i,\text{mis}}^*|\mathbf{w}_i \setminus \mathbf{y}_{i,\text{mis}}^*, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i) \\ & \propto p(\mathbf{r}_i|\mathbf{d}_i, \boldsymbol{\alpha}, \boldsymbol{\varphi})p(\mathbf{y}_{i,\text{mis}}^*|\boldsymbol{\omega}_i, \boldsymbol{\theta}) \\ & \propto \exp\left\{-\frac{1}{2}(\mathbf{y}_{i,\text{mis}}^* - \tilde{\boldsymbol{\mu}}_{i,\text{mis}}^*)' \boldsymbol{\Psi}_{z,\text{mis}}^{-1}(\mathbf{y}_{i,\text{mis}}^* - \tilde{\boldsymbol{\mu}}_{i,\text{mis}}^*) \right. \\ & \quad \left. + \left(\sum_{j=1}^p r_{ij}\right) (\varphi_0 + \boldsymbol{\varphi}'_I \mathbf{d}_i) - p \log(1 + \exp(\varphi_0 + \boldsymbol{\varphi}'_I \mathbf{d}_i))\right\}, \end{aligned}$$

where  $\tilde{\boldsymbol{\mu}}_{i,\text{mis}}^* = \boldsymbol{\mu}_{i,\text{mis}}^* + \mathbf{A}_{i,\text{mis}}^* \mathbf{b}_i + \boldsymbol{\Lambda}_{i,\text{mis}}^* \boldsymbol{\omega}_i$ ,  $\boldsymbol{\mu}_{i,\text{mis}}^*$  is the subvector of  $\boldsymbol{\mu}$ ,  $\mathbf{A}_{i,\text{mis}}^*$ ,  $\boldsymbol{\Lambda}_{i,\text{mis}}^*$ , and  $\boldsymbol{\Psi}_{z,\text{mis}}$  are the submatrices of  $\mathbf{A}$ ,  $\boldsymbol{\Lambda}$ , and  $\boldsymbol{\Psi}_z$  corresponding to  $\mathbf{y}_{i,\text{mis}}^*$ , respectively. Finally,  $\mathbf{d}_i = \{\mathbf{d}_{i,\text{obs}}, \mathbf{d}_{i,\text{mis}}\}$ , and the values of  $\mathbf{d}_{i,\text{mis}}$  are determined from the values  $\mathbf{w}_i$  and  $\boldsymbol{\alpha}$  according to (1) and (2).

(c)  $p(\mathbf{Y}_{\text{mis}}|\mathbf{Y}^*, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{D}_{\text{obs}}, \mathbf{r}) = \prod_{i=1}^n p(\mathbf{y}_{i,\text{mis}}|\mathbf{w}_i \setminus \mathbf{y}_{i,\text{mis}}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i)$ , where  $\mathbf{w}_i \setminus \mathbf{y}_{i,\text{mis}}$  is  $\mathbf{w}_i$  with  $\mathbf{y}_{i,\text{mis}}$  deleted, and

$$(18) \quad \begin{aligned} & p(\mathbf{y}_{i,\text{mis}}|\mathbf{w}_i \setminus \mathbf{y}_{i,\text{mis}}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i) \\ & \propto p(\mathbf{r}_i|\mathbf{d}_i, \boldsymbol{\alpha}, \boldsymbol{\varphi})p(\mathbf{y}_{i,\text{mis}}|\boldsymbol{\omega}_i, \boldsymbol{\theta}) \\ & \propto \exp\left\{-\frac{1}{2}(\mathbf{y}_{i,\text{mis}} - \tilde{\boldsymbol{\mu}}_{i,\text{mis}})' \boldsymbol{\Psi}_{y,\text{mis}}^{-1}(\mathbf{y}_{i,\text{mis}} - \tilde{\boldsymbol{\mu}}_{i,\text{mis}}) \right. \\ & \quad \left. + \left(\sum_{j=1}^p r_{ij}\right) (\varphi_0 + \boldsymbol{\varphi}'_I \mathbf{d}_i) - p \log(1 + \exp(\varphi_0 + \boldsymbol{\varphi}'_I \mathbf{d}_i))\right\}, \end{aligned}$$

where  $\tilde{\boldsymbol{\mu}}_{i,\text{mis}} = \boldsymbol{\mu}_{i,\text{mis}} + \mathbf{A}_{i,\text{mis}}\mathbf{b}_i + \boldsymbol{\Lambda}_{i,\text{mis}}\boldsymbol{\omega}_i$ ,  $\boldsymbol{\mu}_{i,\text{mis}}$  is the sub-vector of  $\boldsymbol{\mu}$ ,  $\mathbf{A}_{i,\text{mis}}$ ,  $\boldsymbol{\Lambda}_{i,\text{mis}}$ , and  $\boldsymbol{\Psi}_{y,\text{mis}}$  are the submatrices of  $\mathbf{A}$ ,  $\boldsymbol{\Lambda}$ , and  $\boldsymbol{\Psi}_y$  corresponding to  $\mathbf{y}_{i,\text{mis}}$ , respectively.

(d)  $p(\mathbf{V}_{\text{mis}}|\mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}_{\text{obs}}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{D}_{\text{obs}}, \mathbf{r}) = \prod_{i=1}^n \times \prod_{j,\text{mis}} p(\mathbf{v}_{ij}|\mathbf{w}_i \setminus \mathbf{v}_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i)$ , where  $\mathbf{w}_i \setminus \mathbf{v}_{ij}$  is  $\mathbf{w}_i$  with  $\mathbf{v}_{ij}$  deleted, and

$$(19) \quad \begin{aligned} & p(\mathbf{v}_{ij}|\mathbf{w}_i \setminus \mathbf{v}_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\varphi}, \mathbf{z}_{i,\text{obs}}, \mathbf{u}_{i,\text{obs}}, \mathbf{r}_i) \\ & \propto p(\mathbf{r}_i|\mathbf{d}_i, \boldsymbol{\varphi}, \boldsymbol{\alpha})p(\mathbf{v}_{ij}|\boldsymbol{\omega}_i, \boldsymbol{\theta}) \\ & \propto \exp\left\{-\frac{1}{2}(\mathbf{v}_{ij} - \boldsymbol{\mu}_j - \mathbf{1}\mathbf{A}_j\mathbf{b}_i - \mathbf{1}\boldsymbol{\Lambda}_j\boldsymbol{\omega}_i)' \right. \\ & \quad \times (\mathbf{v}_{ij} - \boldsymbol{\mu}_j - \mathbf{1}\mathbf{A}_j\mathbf{b}_i - \mathbf{1}\boldsymbol{\Lambda}_j\boldsymbol{\omega}_i) \\ & \quad + \left(\sum_{j=1}^p r_{ij}\right)(\varphi_0 + \boldsymbol{\varphi}'_I\mathbf{d}_i) \\ & \quad \left. - p\log(1 + \exp(\varphi_0 + \boldsymbol{\varphi}'_I\mathbf{d}_i))\right\}. \end{aligned}$$

(e) The full conditional distribution of  $\boldsymbol{\varphi}$  is

$$(20) \quad \begin{aligned} & p(\boldsymbol{\varphi}|\mathbf{Y}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{D}_{\text{obs}}, \mathbf{r}) \propto p(\mathbf{r}|\mathbf{D}, \boldsymbol{\alpha}, \boldsymbol{\varphi})p(\boldsymbol{\varphi}) \\ & \propto \exp\left\{\sum_{i=1}^n \left[\left(\sum_{j=1}^p r_{ij}\right)\boldsymbol{\varphi}'\mathbf{d}_i^* - \log(1 + \exp(\boldsymbol{\varphi}'\mathbf{d}_i^*))\right] \right. \\ & \quad \left. - \frac{1}{2}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)' \mathbf{C}^{-1}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_0)\right\}. \end{aligned}$$

(f) For  $j = 1, \dots, r_1$ , let  $\boldsymbol{\alpha}_j, \mathbf{y}_{j,\text{obs}}^*$  be the  $j$ th rows of  $\boldsymbol{\alpha}, \mathbf{Y}_{\text{obs}}^*$ , respectively. Due to the nonignorable term,  $\{(\boldsymbol{\alpha}_j, \mathbf{y}_{j,\text{obs}}^*, j = 1, \dots, r_1)\}$  are not independent, it is difficult to generate  $p(\boldsymbol{\alpha}, \mathbf{Y}_{\text{obs}}^*|\mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{D}_{\text{obs}}, \mathbf{r})$  directly, so we use the Gibbs sampler [10], simulating  $(\boldsymbol{\alpha}_j, \mathbf{y}_{j,\text{obs}}^*)$  from  $p(\boldsymbol{\alpha}_j, \mathbf{y}_{j,\text{obs}}^*|\mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{D}_{\text{obs}}, \mathbf{r}, \boldsymbol{\alpha}_{-j}, \mathbf{Y}_{-j,\text{obs}}^*)$  iteratively, where  $\boldsymbol{\alpha}_{-j}$  and  $\mathbf{Y}_{-j,\text{obs}}^*$  are  $\boldsymbol{\alpha}$  and  $\mathbf{Y}_{\text{obs}}^*$  with  $\boldsymbol{\alpha}_j$  and  $\mathbf{y}_{j,\text{obs}}^*$  deleted. Let  $\mathbf{z}_{j,\text{obs}}$  be the  $j$ th row of  $\mathbf{Z}_{\text{obs}}$ , it follows from (3) and (6) that

$$(21) \quad \begin{aligned} & p(\boldsymbol{\alpha}_j, \mathbf{y}_{j,\text{obs}}^*|\mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{D}_{\text{obs}}, \mathbf{r}, \boldsymbol{\alpha}_{-j}, \mathbf{Y}_{-j,\text{obs}}^*) \\ & \propto p(\mathbf{r}|\mathbf{D}, \boldsymbol{\alpha}, \boldsymbol{\varphi})p(\mathbf{y}_{j,\text{obs}}^*|\boldsymbol{\alpha}_j, \mathbf{z}_{j,\text{obs}}, \boldsymbol{\theta}, \boldsymbol{\Omega})p(\boldsymbol{\alpha}_j|\mathbf{z}_{j,\text{obs}}, \boldsymbol{\theta}, \boldsymbol{\Omega}). \end{aligned}$$

(g) Other full conditional distributions are the same as those given in Song et al. [27].

## APPENDIX B. IMPLEMENTATION OF THE MH ALGORITHM

The general procedure of the MH algorithm [11, 21] with target distribution  $p(\cdot)$  is implemented as follows. At the  $m$ th iteration of the MH algorithm with a current  $X_m$ , the

next  $X_{m+1}$  is chosen by first sampling from a candidate point  $Y$  from a proposal distribution  $q(\cdot|X_m)$  which is easy to sample. This candidate point  $Y$  is accepted as  $X_{m+1}$  with probability

$$\min\left(1, \frac{p(Y)q(X_m|Y)}{p(X_m)q(Y|X_m)}\right).$$

If the candidate point  $Y$  is rejected, then  $X_{m+1} = X_m$  and the chain does not move. The proposal distributions for generating observations from our non-standard full conditional distributions are presented as follows:

(a) For the full conditional distribution of  $\boldsymbol{\omega}_i$  as given in (16), we choose  $N(\cdot, \sigma_\omega^2 \boldsymbol{\Omega}_\omega)$  as the proposal distribution, where  $\boldsymbol{\Omega}_\omega^{-1} = \boldsymbol{\Sigma}_\omega + \boldsymbol{\Lambda}'\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda}$ ,

$$\boldsymbol{\Sigma}_\omega = \begin{pmatrix} \boldsymbol{\Pi}'_0\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Phi}_0 & -\boldsymbol{\Pi}'_0\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta} \\ -\boldsymbol{\Delta}'\boldsymbol{\Gamma}'\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Pi}_0 & \boldsymbol{\Phi}^{-1} + \boldsymbol{\Delta}'\boldsymbol{\Gamma}'\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta} \end{pmatrix},$$

with  $\boldsymbol{\Pi}_0 = \mathbf{I} - \mathbf{B}$ ,  $\boldsymbol{\Delta} = (\partial\mathbf{F}(\boldsymbol{\xi}_i)/\partial\boldsymbol{\xi}_i)'|_{\boldsymbol{\xi}_i=\mathbf{0}}$ , and  $\sigma_\omega^2$  is chosen such that the acceptance rate is approximately 0.25 or more [see 9].

(b) For the full conditional distribution of  $\mathbf{y}_{i,\text{mis}}^*$  as given in (17), we choose  $N(\cdot, \sigma_{y^*}^2 \boldsymbol{\Psi}_{z,\text{mis}})$  as the proposal distribution, where  $\sigma_{y^*}^2$  is similarly chosen as before.

(c) For the full conditional distribution of  $\mathbf{y}_{i,\text{mis}}$  as given in (18), we choose  $N(\cdot, \sigma_y^2 \boldsymbol{\Omega}_y)$  as the proposal distribution, where

$$\boldsymbol{\Omega}_y^{-1} = \boldsymbol{\Psi}_{y,\text{mis}}^{-1} + \frac{p \exp(\varphi_0 + \sum_{l \in \bar{E}_i} \boldsymbol{\varphi}'_l \mathbf{w}_{il})}{[1 + \exp(\varphi_0 + \sum_{l \in \bar{E}_i} \boldsymbol{\varphi}'_l \mathbf{w}_{il})]^2} \boldsymbol{\varphi}_{im} \boldsymbol{\varphi}'_{im},$$

in which  $\boldsymbol{\varphi}_{im}$  is a subvector of  $\boldsymbol{\varphi}_I$  corresponding to  $\mathbf{y}_{i,\text{mis}}$ ,  $\bar{E}_i$  is the set of indexes corresponding to  $\mathbf{w}_i \setminus \mathbf{y}_{i,\text{mis}}$ , and  $\sigma_y^2$  is similarly chosen as before.

(d) For the full conditional distribution of  $\mathbf{v}_{ij}$  as given in (19), we choose  $N(\cdot, \sigma_v^2 \mathbf{I})$  as the proposal distribution, and  $\sigma_v^2$  is similarly chosen as before.

(e) For the full conditional distribution of  $\boldsymbol{\varphi}$  as given in (20), we choose  $N(\cdot, \sigma_\varphi^2 \boldsymbol{\Omega}_\varphi)$  as the proposal distribution where

$$\boldsymbol{\Omega}_\varphi^{-1} = \frac{p}{4} \sum_{i=1}^n \mathbf{d}_i^* \mathbf{d}_i^{*'} + \mathbf{C}^{-1},$$

and  $\sigma_\varphi^2$  is similarly chosen as before.

(f) At  $(m+1)$ th the iteration, we generate a vector of thresholds  $(\alpha_{j,2}^{(m+1)}, \dots, \alpha_{j,H_j-1}^{(m+1)})$  from the following univariate truncated normal distribution:

$$(22) \quad \alpha_{j,k} \stackrel{D}{=} N(\alpha_{j,k}^{(m)}, \sigma_{\alpha_j}^2) \mathbf{I}_{(\alpha_{j,k-1}^{(m+1)}, \alpha_{j,k+1}^{(m)})}^*(\alpha_{j,k}),$$

for  $k = 2, \dots, H_j - 1$ ,

where  $\alpha_{j,k}^{(m)}$  and  $\alpha_{j,k+1}^{(m)}$  are the thresholds at the  $m$ th iteration,  $\mathbf{I}^*(\cdot)$  is the indicator function which takes value 1 if  $\alpha_{j,k} \in (\alpha_{j,k-1}^{(m+1)}, \alpha_{j,k+1}^{(m)})$ , and 0 otherwise, and  $\sigma_{\alpha_j}^2$  is an appropriate preassigned constant such that the acceptance rate is 0.25 or more. And given  $\alpha_j^{(m+1)}$ ,  $\theta$ ,  $\Omega$ ,  $\mathbf{Z}_{\text{obs}}$ , we generate  $\mathbf{y}_{j,\text{obs}}^{*(m+1)}$  from the following truncated normal distribution:

$$(23) \quad \mathbf{y}_{ij,\text{obs}}^* \stackrel{D}{=} N(\mu_j + \mathbf{A}_j \mathbf{b}_i + \mathbf{A}_j \boldsymbol{\omega}_i, \psi_{\epsilon_j}) \mathbf{I}_{(\alpha_{j,z_{ij}}^{(m+1)}, \alpha_{j,z_{ij}+1}^{(m+1)})}^*(\mathbf{y}_{ij,\text{obs}}^*).$$

The acceptance probability for  $(\alpha_j^{(m+1)}, \mathbf{y}_{j,\text{obs}}^{*(m+1)})$  as a new observation is  $\min\{1, R_j\}$ , where

$$R_j = \frac{p(\alpha_j^{(m+1)}, \mathbf{y}_{j,\text{obs}}^{*(m+1)} | \mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \Omega, \theta, \varphi, \mathbf{D}_{\text{obs}}, \mathbf{r}, \alpha_{-j}, \mathbf{Y}_{-j,\text{obs}}^*)}{p(\alpha_j^{(m)}, \mathbf{y}_{j,\text{obs}}^{*(m)} | \mathbf{Y}_{\text{mis}}^*, \mathbf{Y}_{\text{mis}}, \mathbf{V}, \Omega, \theta, \varphi, \mathbf{D}_{\text{obs}}, \mathbf{r}, \alpha_{-j}, \mathbf{Y}_{-j,\text{obs}}^*)} \times \frac{p(\alpha_j^{(m)}, \mathbf{y}_{j,\text{obs}}^{*(m)} | \alpha_j^{(m+1)}, \mathbf{y}_{j,\text{obs}}^{*(m+1)}, \Omega, \theta, \mathbf{Z}_{\text{obs}})}{p(\alpha_j^{(m+1)}, \mathbf{y}_{j,\text{obs}}^{*(m+1)} | \alpha_j^{(m)}, \mathbf{y}_{j,\text{obs}}^{*(m)}, \Omega, \theta, \mathbf{Z}_{\text{obs}})}.$$

It can be shown from (21), (22), and (23) that

$$(24) \quad R_j = \prod_{k=2}^{H_j-1} \frac{\Phi^*\{(\alpha_{j,k+1}^{(m)} - \alpha_{j,k}^{(m)})/\sigma_{\alpha_j}\} - \Phi^*\{(\alpha_{j,k-1}^{(m+1)} - \alpha_{j,k}^{(m)})/\sigma_{\alpha_j}\}}{\Phi^*\{(\alpha_{j,k+1}^{(m+1)} - \alpha_{j,k}^{(m+1)})/\sigma_{\alpha_j}\} - \Phi^*\{(\alpha_{j,k-1}^{(m)} - \alpha_{j,k}^{(m)})/\sigma_{\alpha_j}\}} \times \prod_{i,\text{obs}} \frac{\Phi^*\{\psi_{\epsilon_j}^{-1/2}(\alpha_{j,z_{ij}+1}^{(m+1)} - \mu_{ij}^*)\} - \Phi^*\{\psi_{\epsilon_j}^{-1/2}(\alpha_{j,z_{ij}}^{(m+1)} - \mu_{ij}^*)\}}{\Phi^*\{\psi_{\epsilon_j}^{-1/2}(\alpha_{j,z_{ij}+1}^{(m)} - \mu_{ij}^*)\} - \Phi^*\{\psi_{\epsilon_j}^{-1/2}(\alpha_{j,z_{ij}}^{(m)} - \mu_{ij}^*)\}} \times \prod_{i=1}^n \exp \left\{ \left[ \sum_{j=1}^p (r_{ij})(\varphi' \mathbf{d}_i^{*(m+1)}) - p \log(1 + \exp(\varphi' \mathbf{d}_i^{*(m+1)})) \right] \right. \\ \left. \left[ \sum_{j=1}^p (r_{ij})(\varphi' \mathbf{d}_i^{*(m)}) - p \log(1 + \exp(\varphi' \mathbf{d}_i^{*(m)})) \right] \right\},$$

where  $\mu_{ij}^* = \mu_j + \mathbf{A}_j \mathbf{b}_i + \mathbf{A}_j \boldsymbol{\omega}_i$ , and  $\mathbf{d}_i^{*(m)}$  and  $\mathbf{d}_i^{*(m+1)}$  are the different values of  $\mathbf{d}_i$  determined from  $\alpha_j^{(m)}$  and  $\alpha_j^{(m+1)}$  according to (1), respectively. According to Cowles [4], when there are no missing data,  $R_j$  only depends on the old and new values of  $\alpha_j$  and not on the  $\mathbf{y}_{j,\text{obs}}^*$ , so it does not require to generate a new  $\mathbf{y}_{j,\text{obs}}^*$  in any iteration in which the new value of  $\alpha_j$  is not accepted. This property still holds from (24). Random observations from the univariate truncated normal in (22) and (23) are simulated via the algorithm of Robert [24].

## ACKNOWLEDGEMENTS

This research was fully supported by grants (CUHK 404507 and 450607) from the Research Grant Council of the Hong Kong Special Administrative Region.

Received 10 January 2008

- [1] BENTLER, P. M. and STEIN, J. A. (1992). Structural equation models in medical research. *Statistical Methods in Medical Research* **1** 159–181.
- [2] BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. Wiley, New York. [MR0996025](#)
- [3] CHEN, C. C., LU, R. B., CHAN, M. F., CHANG, Y. C., and LI, Y. K. (1999). Interaction between the functional polymorphisms of the alcohol-etalobism genes in protection against alcoholism. *The American Journal of Human Genetics* **65** 795–807.
- [4] COWLES, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6** 101–111.
- [5] DIGGLE, P. and KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* **43** 49–93.
- [6] DUNSON, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* **62** 355–366. [MR1749544](#)
- [7] GELMAN, A. (1996). Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. Chapman and Hall, London, 131–144. [MR1397966](#)
- [8] GELMAN, A. and MENG, X. L. (1998). Simulating normalizing constant: From importance sampling to bridge sampling to path sampling. *Statistical Science* **13** 163–185. [MR1647507](#)
- [9] GELMAN, A., ROBERTS, G. O., and GILKS, W. R. (1995). Efficient Metropolis jumping rules. In *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds. Oxford University Press, 599–607. [MR1425429](#)
- [10] GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.
- [11] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika* **57** 97–109.
- [12] IBRAHIM, J. G., CHEN, M. H., and LIPSITZ, S. R. (2001). Missing responses in generalised linear mixed models when missing data mechanism is nonignorable. *Biometrika* **88** 551–564. [MR1844851](#)
- [13] IMAI, K. and VAN DYK, D. A. (2005). Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics* **124** 311–334. [MR2125369](#)
- [14] KASS, R. E. and RAFTERY, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90** 773–795.
- [15] LEE, S. Y. (2007). *Structural Equation Modelling: A Bayesian Approach*. John Wiley & Sons. [MR2290938](#)
- [16] LEE, S. Y., LU, B., and SONG, X. Y. (2008). Semiparametric Bayesian analysis of structural equation models with fixed covariates. *Statistics in Medicine*. In press.
- [17] LEE, S. Y. and SONG, X. Y. (2004). Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics* **69** 624–636. [MR2089437](#)
- [18] LEE, S. Y. and ZHU, H. T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **53** 209–232.
- [19] LITTLE, R. J. A. and RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York. [MR0890519](#)
- [20] MCCULLOCH, R. and ROSSI, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* **64** 207–240. [MR1310524](#)
- [21] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., and TELLER, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics* **21** 1087–1091.
- [22] MOUSTAKI, I. (2003). A general class of latent variable methods for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology* **56** 337–357. [MR2101773](#)

- [23] PUGESEK, B. H., TOMER, A., and VON EYE, A. (2003). *Structural Equation Modeling Applications in Ecological and Evolutionary Biology*. Cambridge University Press, New York.
- [24] ROBERT, C. P. (1995). Simulation of truncated normal variable. *Statistics and Computing* **5** 121–125.
- [25] SANCHEZ, B. N., BUDTZ-JORGENDER, E., RYAN, L. M., and HU, H. (2005). Structural equation models: A review with applications to environmental epidemiology. *Journal of the American Statistical Association* **100** 1443–1455. [MR2236454](#)
- [26] SEAQUIST, E. R., GOETZ, F. C., RICH, S., and BARBOSA, J. (1989). Familiar clustering of diabetic kidney disease: Evidence for genetic susceptibility to diabetic nephropathy. *The New England Journal of Medicine* **320** 1161–1165.
- [27] SONG, X. Y., LEE, S. Y., NG, M. C. Y., SO, W. Y., and CHAN, J. C. N. (2007). Bayesian analysis of structural equation models with multinomial variables and an application to type 2 diabetic nephropathy. *Statistics in Medicine* **26** 2348–2369. [MR2368420](#)
- [28] TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82** 528–550. [MR0898357](#)

Jing-Heng Cai  
 Dept. of Statistics  
 The Chinese University of Hong Kong  
 Hong Kong, China  
 E-mail address: [s050389@sta.cuhk.edu.hk](mailto:s050389@sta.cuhk.edu.hk)

Xin-Yuan Song  
 Dept. of Statistics  
 The Chinese University of Hong Kong  
 Hong Kong, China  
 E-mail address: [xysong@sta.cuhk.edu.hk](mailto:xysong@sta.cuhk.edu.hk)

Sik-Yum Lee  
 Dept. of Statistics  
 The Chinese University of Hong Kong  
 Hong Kong, China  
 E-mail address: [sylee@sta.cuhk.edu.hk](mailto:sylee@sta.cuhk.edu.hk)