Journal of Multivariate Analysis 135 (2015) 43–58

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A Bayesian method for analyzing combinations of continuous, ordinal, and nominal categorical data with missing values

Xiao Zhang [a,*], W. John Boscardin [b], Thomas R. Belin [c], Xiaohai Wan [d,e], Yulei He [f], Kui Zhang [e]

[a] Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, United States
[b] Department of Medicine, University of San Francisco, United States
[c] Department of Biostatistics, UCLA Jonathan and Karin Fielding School of Public Health, United States
[d] AstraZeneca Pharmaceuticals LP, United States
[e] Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, United States
[f] Office of Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention, United States

ARTICLE INFO

Article history:
Received 3 September 2013
Available online 13 December 2014

AMS subject classification:
62H99

Keywords:
Multivariate probit model
Multinomial probit model
MCMC

ABSTRACT

From a Bayesian perspective, we propose a general method for analyzing a combination of continuous, ordinal (including binary), and categorical/nominal multivariate measures with missing values. We assume multivariate normal linear regression models for multivariate continuous measures, multivariate probit models for correlated ordinal measures, and multivariate multinomial probit models for multivariate categorical/nominal measures. Then we assume a multivariate normal linear model on the continuous vector comprised of continuous variables and those underlying normal variables for ordinal variables from multivariate probit models and for categorical variables from multinomial probit models. We develop a Markov chain Monte Carlo (MCMC) algorithm to estimate unknown parameters including regression parameters, cut-points for ordinal data from the multivariate probit models, and the covariance matrix encompassing both continuous variables and the underlying normal latent variables. Combining the continuous variables and the normal latent variables allows us to model combinations of continuous, ordinal, and categorical multivariate data simultaneously. The framework incorporates flexible priors for the covariance matrix, provides a foundation for inference about the underlying covariance structure, and imputes missing data where needed. The method is illustrated through simulated examples and two real data applications.

1. Introduction

Multivariate measures and longitudinal data arise in many fields of science. There is a long history of methodological development for analyzing multivariate continuous data from both classical and Bayesian perspectives, e.g. [31,63,34,62,15,59,11].

Statistical methods for analyzing multivariate ordinal (or polytomous) data including multivariate binary (or dichotomous) data have also been established. Generalized estimating equations (GEE) methods have propelled the development

* Corresponding author.
E-mail address: xiao.zhang@cshs.org (X. Zhang).

http://dx.doi.org/10.1016/j.jmva.2014.11.007
0047-259X/

of classical methods for analyzing multivariate ordinal data, such as Zeger and Liang [66], Liang and Zeger [35], Prentice [52], Miller et al. [45], Qu et al. [53]. From a Bayesian perspective, Chib and Greenberg [9], Nandram and Chen [48], Liu [38], and Edwards and Allenby [18] analyzed multivariate ordinal data in the setting of the multivariate probit model.

Compared with multivariate continuous data and multivariate ordinal data, analyzing multivariate nominal categorical data is much less familiar. Liang et al. [36] performed multivariate regression analyses for categorical data using GEE with intensive computation. The multinomial probit model, which was developed for univariate nominal categorical data, has been generalized to the multinomial multiperiod probit model for multiple categorical data, e.g., [42,22,23,55,27]. To release the restriction on the covariance matrix of the multinomial multiperiod probit model and to generalize the multinomial probit model, Zhang et al. [69] proposed the multivariate multinomial probit model for multivariate nominal data with the multinomial probit model as a special case.

In practice, sometimes it is inevitable to analyze mixtures of multivariate continuous, ordinal and nominal measures, or various combinations of these three types of measures. This research field can be in general divided into several branches: (1) joint modeling using direct likelihood estimation and GEE methods; (2) joint modeling using general location models; (3) latent variable models (structural equation modeling, Bayesian latent variable models); (4) Other alternatives.

Joint modeling methods are to derive the joint distributions of the mixed outcomes and then to use GEE or quasi-likelihood methods to make statistical inference. Related references can be found in [6,54,24]. Modeling mixed measures using GEE include Zeger et al. [67], Legler et al. [33], and Spiess [61].

The general location model Olkin and Tate [50] has been popularized in analyzing mixed continuous and categorical data through specifying multinomial models for categorical variables and conditional multivariate normal models for continuous variables with different means across cells from those categorical variables and a common covariance matrix across cells. Using the general location model to analyze mixed measurements can be found in [37,19,59,13]. Liu and Rubin [39] extended the common covariance matrix to allow different, but proportional covariance matrices and replace the multivariate normal distribution specified for continuous variables by multivariate t distribution.

The structural equation model [30,47] has also been a popular tool to analyze mixed measures, as has path analysis [7], which refers to a similar use of linear models without latent variables. It assumes a continuous latent variable underlying several observed measures describing a common concept. The linear equation links the latent variable and the observed variables, i.e., factor analysis model is assumed for latent variables and the observed variables. Using the structural equation models to analyze mixed data types can be referred to Muthén [47], Arminger and Küsters [2], Shi and Lee [60], Lee and Zhu [32].

Using latent variable models is another active area for analyzing mixed measurement. Sammel et al. [58] proposed a latent variable mixed effects model by assuming a latent variable, linearly linked to the observed covariates, for each subject and the distribution for each type of measurement given the latent variable is from an exponential family. EM algorithm is applied to estimate the unknown parameters and the latent variables. Dunson [17] proposed Bayesian latent variable models for clustered mixed outcomes. The outcomes and the latent variables are linked through a known function and the latent variables are assumed to follow an exponential family. Adding random effect variables into the means of the specified exponential distribution accounts the correlated structure for mixed outcomes. Bayesian sampling algorithm can be derived to make statistical inference. Moustaki and Knott [46] proposed a latent trait model for various type of measurements. They assumed the distributions for mixed measurements given the latent variables are from the exponential family and estimate the unknown parameters and the latent variables using the maximum likelihood method. O'Malley et al. [51] combined the general location model and the latent trait model for mixed outcomes. Daniels and Normand [12] added latent variables to the model through the mean functions to estimate the correlations among different types of measurements. Goldstein et al. [25] proposed multilevel models for mixed data types. Weiss et al. [64] analyzed mixed outcomes through assuming various exponential distributions based on the type of the outcomes and then linearly linked the unknown mean functions with random effect variables to count for the correlated structure.

Besides the above general areas for analyzing the mixed measurements, there are other alternative methods, such as Miglioretti [44] used latent transition regression models for mixed outcomes and de Leon and Wu [14] proposed a copula-based regression models for a bivariate mixed outcome.

In this manuscript, we propose a joint modeling method using latent variables for mixed measures analyzed from a Bayesian perspective. We assume the multivariate probit model for multivariate ordinal (including binary) data. This means that there is an underlying normal latent variable for each ordinal outcome. We further assume the multivariate multinomial probit model for multivariate nominal data, where for each nominal outcome with p levels, there would be $p - 1$ normal latent variables. The multivariate multinomial probit model generalizes the multinomial probit model for the univariate nominal data. Detailed discussion about the multivariate multinomial probit model can be referred to Zhang et al. [69]. Then we combine these latent variables from multivariate ordinal and nominal measures with the multivariate continuous measures. We assume a multivariate linear model on those combined latent variables and the continuous measures and develop an MCMC algorithm to estimate the unknown parameters including the covariance matrix for the latent variables and the normal continuous variables.

Statistical methods for incomplete mixed data types are not well developed in the literature, and our paper offers expanded flexibility and generality. Although GEE methods could be considered in the present context, our proposed approach provides a foundation for drawing inferences relevant to the correlation structure of the data, and as such arguably represents an advance over GEE as well. Whereas Zhang et al. [69] describes an incomplete-data method for multivariate nominal data and Boscardin et al. [5] describes an incomplete-data method for a combination of continuous and ordinal (or binary) data, the present paper includes all of these data types (i.e., continuous, ordinal, and nominal). Our proposed method is able to extend joint modeling methods to general types of mixed measures and allows us to directly estimate the covariance matrix for latent variables. This contrasts with previously joint modeling methods targeted special types of mixed outcomes and latent variable methods for modeling the dependence structures purpose and do not allow straightforward inference for the covariance structure of the mixed measures.

Missing values in mixed measures have been considered by Little and Schluchter [37], Fitzmaurice and Laird [19], and Schafer [59]. Since missing values are ubiquitous in data sets from almost every scientific field, we also consider missing data in our model and the imputation steps for missing data are naturally built in and carried through the MCMC sampling algorithm. In this sequel, we assume the missing data mechanism is missing at random (MAR). Although one cannot regard this assumption as automatically valid, MAR is plausible in many settings and provides a foundation for general-purpose analysis procedures that can often be adapted to scenarios where a particular non-ignorable missing-data mechanism is assumed.

Our paper proceeds as follows. Section 2 gives relevant background on multivariate probit models, multinomial probit models, and multivariate multinomial probit models. In Section 3, we present a joint model that combines these elements as an associated Bayesian sampling algorithm. We use simulated examples (Section 4) and two real data examples (Sections 5 and 6) to illustrate our methodology. Section 7 includes discussion about the advantages and disadvantages for our method and proposes possible future work in this area.

2. Background review

2.1. Probit model and multivariate probit model

The idea of using a generalized linear model with a probit link function was introduced by Bliss [4] for binary or ordinal measurements. Albert and Chib [1] presented a Bayesian sampling method to estimate the probit model for independent binary and ordinal random variables. What follows is a brief review for probit model.

Suppose there are N independent binary variables $Y_1, Y_2, \ldots, Y_N$ with values 0 or 1, where each Y_i for $i = 1, \ldots, N$ has a Bernoulli distribution with the mean equal to p_i. If $p_i = \Phi(X_i\beta)$, where $\Phi(\cdot)$ is the cumulative function for standard normal distribution, X_i is a vector for observed covariates, and β is the vector for unknown regression parameters, then this model for Y_i is called a probit model. Probit model can also be described as

$$Y_i = \begin{cases} 0 & \text{if} \quad Z_i \le 0 \\ 1 & \text{if} \quad Z_i > 0 \end{cases} \tag{1}$$

where Z_i follows standard normal distribution with mean $X_i\beta$ and variance 1, denoted by $\mathrm{N}(X_i\beta, 1)$.

For N independent ordinal variables $Y_1, Y_2, \ldots, Y_N$ with possible J ordered categories, $1, \ldots, J$, let $p_{ij} = P(Y_i = j)$ and the cumulative probabilities $\eta_{ij} = \Sigma_{k=1}^{j} p_{ik}$. Then the probit model assumes $\eta_{ij} = \Phi(\gamma_j - X_i\beta)$, where γ_j for $j = 0, 1, \ldots, J$ are cut-points. Furthermore, if we assume that a latent continuous variable Z_i follows $\mathrm{N}(X_i\beta, 1)$, such that $Y_i = j$ if $\gamma_{j-1} < Z_i \le \gamma_j$ (we define that $\gamma_0 = -\infty$ and $\gamma_J = \infty$). For parameter identification, without loss of generality, we let $\gamma_1 = 0$. Then the probit model for binary data is just a special case for ordinal data with $\gamma_0 = -\infty, \gamma_1 = 0, \gamma_2 = \infty$. The unknown parameters for ordinal data include both the regression parameters β as well as the cut-points $\gamma_j, j = 2, \ldots, J-1$ for more than two categories. Since binary variables are special cases for ordinal variables with two categories, therefore, the following presentation uses the terminology for ordinal random variables.

Ashford and Sowden [3] extended the probit model from independent univariate Y_i to independent multivariate $Y_i = (Y_{i1}, \ldots, Y_{iK})^T$ such that $Y_{i1}, \ldots, Y_{iK}$ are correlated ordinal variables with J categories. For each $Y_{ik},\ i = 1, \ldots, N, k = 1, \ldots, K$, the probit model is assumed such that the cumulative probabilities $\eta_{ikj} = \sum_{l=1}^{j} P(Y_{ik} = l) = \Phi(\gamma_{kj} - X_{ik}\beta)$, i.e., $Y_{ik} = j$ if $\gamma_{k(j-1)} < Z_{ik} \le \gamma_{kj}$, where Z_{ik} follows $N(X_{ik}\beta, 1)$. Since $Y_{i1}, \ldots, Y_{iK}$ are correlated, it is natural to assume that $Z_{i1}, \ldots, Z_{iK}$ are correlated, too. We use Σ to denote the covariance matrix for $Z_i = (Z_{i1}, \ldots, Z_{iK})$. Since each Z_{ik} is from $N(X_{ik}\beta, 1)$, the variance of Z_{ik} is equal to 1 and therefore, Σ is a correlation matrix, sometimes called a polychoric correlation matrix Drasgow [16], instead of a covariance matrix. Further details can be referred to Chib and Greenberg [9], which is one of the precursors of Bayesian analysis of multivariate probit models for correlated binary outcomes.

2.2. Multinomial probit (MNP) model

Building on the work by McFadden [43], Hajivassiliou et al. [28], Geweke et al. [21], Albert and Chib [1], McCulloch and Rossi [42], and Nobile [49], the multinomial probit (MNP) model has generated renewed interest in the fields of both statistics and economics.

Let $i = 1, 2, \ldots, n$ index subjects and $j = 1, 2, \ldots, p$ index levels of a multinomial outcome with p levels, let $y_{ij} = 1$ if subject i has outcome j and $y_{ij} = 0$ otherwise. Let $y_i = (y_{i1}, \ldots, y_{ip})$ be a multinomial $1 \times p$ vector. More compactly, we define $d = (d_1, \ldots, d_n)^T$, where d_i contains the index of the chosen alternative, i.e., $d_i = j$ if $y_{ij} = 1$. To understand the notation, suppose we have $n = 2$ persons and each person choose one color from RED, YELLOW, and GREEN, i.e., $p = 3$. If the first person chooses RED, then $y_{11} = 1, y_{12} = 0, y_{13} = 0$, therefore have $y_1 = (y_{11}, y_{12}, y_{13}) = (1, 0, 0)$ and $d_1 = 1$. Similarly, if the second person chooses GREEN, we have $y_2 = (y_{21}, y_{22}, y_{23}) = (0, 0, 1)$ and $d_2 = 3$. Then we have $d = (d_1, d_2) = (1, 3)$.

Following the notation in economics settings where utilities underlie choices, the MNP model assumes that there is a latent $1 \times p$ vector $u_i = (u_{i1}, \ldots, u_{ip})$ underlying each multinomial vector y_i, such that the multinomial outcome is determined by the maximum u_{ij}, as would happen if the subject chooses the alternative with maximum utility score. That is

$$d_i = j \Leftrightarrow u_{ij} \geq \max_{1 \leq l \leq p} u_{il}. \tag{2}$$

The MNP model further assumes that the vector u_i follows a multivariate normal distribution with mean equal to $A_i\beta$ and covariance matrix equal to V, where A_i is a $p \times k$ covariate matrix for subject i and β is a $k \times 1$ regression parameter vector. With this notation,

$$u_i = A_i\beta + \mu_i \tag{3}$$

where $\mu_i \sim \mathrm{N}(0, V)$.

There are two identification problems in the above MNP model specification. The first identification problem is that the distribution of vector d is unchanged by adding any arbitrary constant to both sides of Eq. (3) and usually solved by subtracting the p-th row of Eq. (3) from the first $(p-1)$ rows. The model becomes

$$z_i = X_i\beta + \epsilon_i \tag{4}$$

where $\epsilon_i \sim \mathrm{N}(0, \Sigma)$ independently, $z_{ij} = u_{ij} - u_{ip}, X_{ij} = A_{ij} - A_{ip}, \epsilon_{ij} = \mu_{ij} - \mu_{ip}$ and $\Sigma = [I_{p-1}, -\mathbf{1}_{p-1}]V[I_{p-1}, -\mathbf{1}_{p-1}]^T$, with I_s denoting the $s \times s$ identity matrix and $\mathbf{1_s}$ a vector of length s comprised of 1's.

However, the model is still unidentified with multiplication of any positive constant to both sides of Eq. (4). This identification problem can be solved by restricting the first element of Σ, σ_{11}, to be equal 1. This strategy is also used by McCulloch et al. [41]. Thus, the fully identifiable MNP model can be described as follows:

$$d_i = \begin{cases} 0 & \text{if} \quad \max\limits_{1 \leq l \leq p-1} z_{il} < 0 \\ j & \text{if} \quad \max\limits_{1 \leq l \leq p-1} z_{il} = z_{ij} > 0 \end{cases} \tag{5}$$

where $z_i \sim \mathrm{N}(X_i\beta, \Sigma)$ and $\sigma_{11} = 1$.

Notice that the dimension of z_i is $p - 1$ instead of p.

2.3. Multivariate multinomial probit (MVMNP) model

The MNP model is for the univariate categorical response, i.e., each subject has one categorical outcome and this categorical outcome has p levels or categories. Extensions of MNP models to multiple categorical measures have been investigated from several perspectives. Generalizing multiple nominal measures to multiple time points has been extensively studied by McCulloch and Rossi [42], Geweke et al. [22,23], Chen and Kuo [8], and Ziegler [70]; Rendtel and Kaltenborn [55], with limited covariance matrix for the latent variable to facilitate the computation. Liang and Zeger [34], Zeger and Liang [66], Zeger [65], and Liang et al. [36] used GEE methods treating correlation matrices as nuisance parameters. Golob and Regan [26] proposed the generalized least-squares approach, forcing the magnitudes of each latent variables to be equal and this may not be appropriate without any knowledge of the latent variables. The multivariate multinomial probit model (MVMNP) proposed by Zhang et al. [69] extended the multinomial probit model and allows general covariance matrix specification. Let us introduce the MVMNP model.

Suppose for each subject i, there are g nominal measures, the first with p_1 levels, the next with p_2 levels, and so on up to the last with p_g levels. Let $d_i = (d_{i1}, \ldots, d_{ig})$ denote the index vector of the alternatives the ith subject chooses for the g measures. Assume each of these g nominal measures follows a multinomial probit model. This is, for the q-th measure, $q = 1, \ldots, g$, there is a $(p_q - 1)$-dimensional underlying utility vector z_{iq} satisfying Eq. (5) with mean equal to $X_{iq}\beta$ and covariance matrix equal to Σ_q with the first element $\sigma_q(1, 1) = 1$.

Then we stack up all the utility vectors z_{iq} for $q = 1, \ldots, g$ to be $z_i^T = (z_{i1}, \ldots, z_{ig})$ with $z_{iq} = (z_{iq1}, \ldots, z_{iq(p_q-1)})$. The MVMNP model for the g measures can be described as follows:

$$z_i = X_i\beta + \epsilon_i \tag{6}$$

where $X_i = (X_{i1}^T, \ldots, X_{ig}^T)^T$ and $\epsilon_i \sim \mathrm{N}(0, \Sigma)$ with $\sigma_{qq} = 1$, where $q = 1, p_1, \ldots, p_1 + p_2 + \cdots + p_{g-1} - g - 1$. We then specify

$$d_{iqj} = \begin{cases} 0 & \text{if} \quad \max\limits_{1 \leq l \leq p_q-1} z_{iql} < 0 \\ j & \text{if} \quad \max\limits_{1 \leq l \leq p_q-1} z_{iql} = z_{iqj} > 0 \end{cases}$$

for $i = 1, \ldots, n, q = 1, \ldots, g$, and $j = 1, \ldots, p_q - 1$.

We can see that comparing the MNP model, the MVMNP model assumes a MNP model for each nominal outcome and describes the correlated structure among multiple measurements through one covariance matrix for all the latent utility scores with restrictions on the first element of each covariance matrix from the MNP model due to the model identification.

3. Modeling and sampling mixtures of continuous, ordinal and nominal measures

3.1. Modeling mixtures of continuous, ordinal and nominal measures

Suppose $Y_i = (C_i^T, O_i^T, N_i^T)^T$ is the observed vector consisting of the continuous portion $C_i^T = (c_{i1}, \ldots, c_{im_1})$, the ordinal portion $O_i^T = (o_{i1}, \ldots, o_{im_2})$ and the nominal portion $N_i^T = (n_{i1}, \ldots, n_{im_3})$, for subject i, $i = 1, \ldots, N$.

We have three assumptions for the observed Y_i. First, we assume that C_i follows a multivariate normal distribution with mean equal to $X_{Ci}\beta$, where β is the regression parameter vector, and the variance matrix equal to Σ_C. Second, we assume that O_i follows the multivariate probit model with the continuous normal latent variable ZO_i having mean equal to $X_{Oi}\beta$ and covariance matrix equal to Σ_O. Notice that Σ_O is a correlation matrix (Section 2.1). Then we assume the MVMNP model for N_i with the continuous normal latent variable ZN_i having mean equal to $X_{Ni}\beta$ and covariance matrix equal to Σ_N, where Σ_N is a restricted covariance matrix.

Then we denote $Z_i = (C_i^T, ZO_i^T, ZN_i^T)^T$ and $X_i = (X_{Ci}^T, X_{Oi}^T, X_{Ni}^T)^T$. We further assume that Z_i follows multivariate normal distribution with mean equal to $X_i\beta$ and the covariance matrix equal to Σ with $cov(C_i) = \Sigma_C$, $cov(ZO_i) = \Sigma_O$ and $cov(ZN_i) = \Sigma_N$. Notice that Σ is a covariance matrix with some of the diagonal elements equal to 1.

3.2. Bayesian sampling for mixtures of continuous, ordinal and nominal measures without missing values

To use the MCMC scheme, we need to write down the posterior density for β, Γ (cutpoints for the ordinal portion), Σ, and $Z = (Z_1, \ldots, Z_N)$ given the observed Y_i, which is

$$p(\beta, \Gamma, \Sigma, Z|Y) \propto p(\beta) \times p(\Gamma) \times p(\Sigma) \times \prod_{i=1}^{n}[I_i \times \phi(Z_i; X_i\beta, \Sigma)]$$

where ϕ is the standard normal density function, and I_i is the indicator function indicating the compatibility of Z_i and Y_i, combining the compatibility from Sections 2.1 and 2.2. This is, $I_i = IO_i \times IN_i$, where IO_i is the indicator function for the multivariate probit model for ordinal measures and IN_i is the indicator function for the MVMNP model for nominal measures. We have

$$IO_i = \prod_{j=1}^{m_2} I_{ij}, \quad \text{and} \quad I_{ij} = \sum_{l=1}^{J} 1_{(O_{ij}=l)} 1_{(\gamma_{jl} < ZO_{ij} < \gamma_{j(l+1)})}.IN_i = \prod_{q=1}^{m_3} I_{iq},$$

and

$$I_{iq} = \mathbf{1}_{(d_{iqj}=0, ZN_{iqj}<0, j=1,\ldots,p_q-1)} + \sum_{k=1}^{p_q-1} \mathbf{1}_{(d_{iqj}=k, ZN_{iqk}=\max_{1\le l\le p_q-1}(ZN_{iql},0))}.$$

To implement our MCMC algorithms, we build on the following:

- Assuming $\beta \sim N(b, C)$ as a prior distribution for β and using standard Bayesian linear model results, $\beta|\Gamma, \Sigma, Z, Y$ has a multivariate normal distribution:

  $$\beta|\Gamma, \Sigma, Z, Y \sim \mathrm{N}(\hat{\beta}, V_\beta),$$

  where $V_\beta = \left(\sum_{i=1}^{N} X_i^T \Sigma^{-1} X_i + C^{-1}\right)^{-1}$ and $\hat{\beta} = V_\beta \left(\sum_{i=1}^{N} X_i^T \Sigma^{-1} Z_i + C^{-1} b\right)$.
- The latent variable $Z_{ij}|\beta, \Gamma, \Sigma, Z_{ij'}, j' \neq j, Y_i$ has a truncated normal distribution that can be represented:

  $$\begin{aligned} p(Z_{ij}|\beta, \Gamma, \Sigma, Y_{ij}) &\propto I_i \times p(Z_{ij}|\beta, \Gamma, \Sigma, Z_{ij'}, j' \neq j) \\ &= I_i \times \phi(Z_{ij}; \mu_{ij}, \Sigma_{ij}) \end{aligned}$$

  where: μ_{ij} and Σ_{ij} are the conditional mean and variance of Z_{ij} given $Z_{ij'}, j' \neq j$. The truncation is based on the comparability of Z_{ij} and its corresponding Y_{ij}.
- Considering the full conditional distribution foreach γ_{mn}, $m = \{1, \ldots, p_o, n = 2, \ldots, J_j - 1\}$, where p_o is the number of ordinal measures, we have

  $$\begin{aligned} p(\gamma_{mn}|\beta, \Sigma, Z, Y, \gamma_{mj}, j \neq n) &= p(\gamma_{mn}|ZO_{im}, C_{im}, i = 1, \ldots, N, \gamma_{mj}, j \neq n) \\ &\propto \prod_{i=1}^{N}[I_{(O_{im}=n)} I_{(\gamma_{m(n-1)} < ZO_{im} < \gamma_{mn})} + I_{(O_{im}=n+1)} I_{(\gamma_{mn} < ZO_{im} < \gamma_{m(n+1)})}] \\ &= U(\gamma_{mn}; \max\{\max\{ZO_{im} : O_{im} = n\}, \gamma_{m(n-1)}\}, \\ &\quad \min\{\min\{ZO_{im} : O_{ij} = n + 1\}, \gamma_{m(n+1)}\}). \end{aligned}$$
- Assuming Σ has prior density $p(\Sigma)$, we have $p(\Sigma|\beta, \Gamma, Z, Y)$ is proportional to $p(\Sigma) \times \prod_{i=1}^{n} \phi(Z_i; X_i\beta, \Sigma)$. It is not easy to directly draw simulations from the posterior distribution of the covariance matrix Σ with some diagonal elements equal to 1. In the next section, we elaborate in detail the steps involved in drawing $p(\Sigma|\beta, \Gamma, Z, d)$ using the parameter-extended Metropolis–Hastings (PX-MH) algorithm.

3.3. PX-MH algorithm

In this section we review the PX-MH algorithm developed by Zhang et al. [68]. To sample a correlation matrix, R, in a multivariate probit model, Zhang et al. sample a covariance matrix, W, using the decomposition $W = D^{1/2}RD^{1/2}$ where D is a diagonal matrix of artificial variance components governed by a joint prior distribution, $p(R, D)$, for the correlation matrix R and D. We present the PX-MH algorithm as follows.

Set initial value of $(R^{(0)}, D^{(0)})$ through setting $W^{(0)} = D^{(0)\frac{1}{2}}R^{(0)}D^{(0)\frac{1}{2}}$ to an initial covariance matrix.
Then, at iteration $(t + 1)$

1. Generate (R^*, D^*) by generating $W^* = D^{*\frac{1}{2}}R^*D^{*\frac{1}{2}}$ from $Wishart_p(m, W^{(t)})$.
2. Take

$$(R^{(t+1)}, D^{(t+1)}) = \begin{cases} (R^*, D^*) & \text{with probability } \alpha \\ (R^{(t)}, D^{(t)}) & \text{otherwise} \end{cases}$$

where $\alpha = \min\left\{\frac{p(R^*,D^*|\beta,Z,Y)}{p(R^{(t)},D^{(t)}|\beta,Z,Y)}\frac{q(W^{(t)}|W^*)}{q(W^*|W^{(t)})}, 1\right\}$. Here, $p(R, D|\beta, Z, Y)$ is the joint posterior density of (R, D) and $q(.|W^{(t)})$, the proposal density, is equal to the product of the Jacobian term for the transformation $(W \to R, D)$ and the Wishart density with m degrees of freedom and scale matrix equal to $W^{(t)}$.

For the multivariate probit model, Σ_O is a correlation matrix. So the diagonal elements of Σ corresponding to Σ_O are all equal to 1. For the MVMNP model, Σ_N has g diagonal elements equal to 1, where g is the number of nominal measures. The covariance matrix Σ_C of the continuous measures is a covariance matrix without any restrictions. We decompose $\Sigma = D^0RD^0$ where R is the correlation matrix of Σ and D^0 is the diagonal standard deviation matrix with some of the elements equal to 1. Then we consider a diagonal matrix D replacing those elements of D^0 equal to 1 with unknown parameters. Therefore, the matrix $W = DRD$ is a covariance matrix without restrictions on the diagonal elements. We use the above PX-MH algorithm to sample W, thereby obtaining a draw of Σ. A slight distinction between sampling Σ in the mixture model and sampling R in the multivariate probit model is that some of the diagonal elements of D are identified parameters in the mixture model, while for the multivariate probit model, all the diagonal elements of D are artificial; this distinction does not alter the character of the algorithm, however.

For the prior distribution of Σ, we use a PXW prior proposed by Zhang et al. [68], with density given by the product of the Jacobian term for the transformation $(W \to R, D)$ and the Wishart density with m_0 degrees of freedom and scale matrix equal to Λ. The scale matrix Λ reflects the prior guess for the covariance matrix Σ with higher values of m_0 representing greater prior precision.

Including the artificial parameters from the ordinal and nominal measures, the joint posterior density of β, Γ, R, D, Z given Y is

$$p(\beta, \Gamma, R, D, Z|Y) \propto p(\beta) \times p(\Gamma) \times p(R, D) \times \prod_{i=1}^{N}[I_i \times \phi(Z_i; X_i\beta, \Sigma)].$$

The conditional distributions for β, Γ and Z_i given other parameters are the same as described in Section 3.2. Through this joint posterior density, we have $p(R, D|\beta, \Gamma, R, D, Z, Y)$ is proportional to $p(R, D) \times \prod_{i=1}^{N} \phi(Z_i; X_i\beta, \Sigma)$. As suggested above, the prior density $p(R, D)$ can be specified by letting the joint prior distribution of (R, D) be from the $PXW(m_0, \Lambda)$ family of distributions. Therefore, one cycle of the algorithm consists of Gibbs steps to sample β, each component of the latent variable z_i, each components of the cutpoint vector Γ, and a Metropolis–Hastings step for sampling (R, D), with Σ generated as a byproduct of the PX-MH step.

3.4. Missing data investigation for mixtures of continuous, ordinal and nominal measures

In this sequel, we assume that there are no missing values for adjusted covariates and concentrate on the missing values for mixed data types. In Section 3.2, we present the Bayesian sampling algorithm for mixed data without missing values. However, in reality, missing values happen quite often. This means the mixed Y is comprised of Y_{obs} denoting the observed portion and Y_{mis} denoting the missing portion. Specifically, we further distinguish the missing values from the continuous variables denoted by $Y_{mis,c}$ and those from the discrete (ordinal and categorical) variables by $Y_{mis,d}$. Since the continuous variables are part of the latent variables Z, therefore, we incorporating this information to the joint posterior distribution, we have

$$p(\beta, \Gamma, \Sigma, Z, Y_{mis,d}|Y_{obs}) \propto p(\beta) \times p(\Gamma) \times p(\Sigma) \times \prod_{i=1}^{n}[I_i \times \phi(Z_i; X_i\beta, \Sigma)].$$

We notice that this joint posterior distribution has the same formulation as that without missing values in Section 3.2. However, $Y_{mis,c}$ is included through Z and $Y_{mis,d}$ is in indicator function I_i.

Without missing values for continuous variables, the sampled steps for latent variables are only those corresponding to the ordinal and categorical data. With missing values $Y_{mis,c}$, the latent variable corresponding to each component $Y_{mis,c}$ is sampled, too. Notice that this sampling step is a conditional univariate normal distribution without truncation. However, the full conditional distribution for each component of $Y_{mis,d}$ is determined through the indicator function I_i. Due to this determinate relationship, each component $Y_{mis,d}$ has no opportunity to make a transition to any other values once given the initial value from its corresponding latent component. Hence, the produced Markov chain is reducible and the samplings from the full conditional distributions do not converge.

To solve the reducibility issue, we integrate $Y_{mis,d}$ from the joint posterior distribution and consider $p(\beta, \Gamma, \Sigma, Z|Y_{obs})$, which satisfies the positive condition implying the convergence of the Markov chain Robert and Casella [56]. We notice that the sampling for each latent component corresponding to the missed component from $Y_{mis,d}$ is a univariate normal distribution without truncation. The other samplings remain the same formulation as those without missing values, including the PX-MH sampling step for covariance matrix Σ. Then based on their determinate relationship, we can impute the missing values $Y_{mis,d}$ through their sampled corresponding latent variables. For the purpose of multiple imputation, after diagnosing the convergence of the Markov chain, we can thin the chain to get 5–10 approximately independent samples, or obtain the samples by running 5–10 parallel chains.

4. Simulated examples

We use the following simulated examples to illustrate the Bayesian model and the MCMC algorithm proposed in Section 3 for analyzing combinations of continuous, ordinal and nominal measures.

We generated two simulated data sets, one with a sample size of 200 and the other with a sample size of 2000. Each subject i was assumed to have two nominal outcomes (y_{i1} with three categorical levels and y_{i2} with four categorical levels), two ordinal outcomes (y_{i3} and y_{i4}, each having four levels from 0 up to 3) and two continuous outcomes (y_{i5} and y_{i6}). Based on Section 2.2, we use z_{i1} and z_{i2} to denote the two latent variables corresponding to y_{i1}, use z_{i3}, z_{i4}, and z_{i5} to denote the latent variable corresponding to y_{i2}, and use z_{i6} and z_{i7} to denote the latent variables for ordinal outcomes y_{i3} and y_{i4}, respectively. There are two unknown cutpoints for an ordinal outcome having four levels based on Section 2.1 and we set them to be 0.5 and 1.0 for both ordinal outcomes, y_{i3} and y_{i4}. The covariate matrix X_i was generated from i.i.d. uniform $(-0.5, 0.5)$. We set the regression parameter β equal to 2.0. We generate the multivariate normal vector $z_i = (z_{i1}, \ldots, z_{i7}, y_{i5}, y_{i6})$ with mean equal to $X_i\beta$ and covariance matrix equal to Σ with the variance components for $(z_{i2}, z_{i4}, z_{i5}, y_{i5}, y_{i6})$ being 1, the correlations between z_{i1} and z_{i2}, among z_{i3}, z_{i4} and z_{i5}, between z_{i6} and z_{i7}, and between y_{i5} and y_{i6} being 0.3, and other correlations being 0.2. Based on the assumptions for the multivariate multinomial probit model for multivariate nominal outcomes and the multivariate probit model for multivariate ordinal outcomes in Section 2, the data $y_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, y_{i6})$ with y_{i1} and y_{i2} being the categorical variables, y_{i3} and y_{i4} being the ordinal variables, and y_{i5} and y_{i6} being the continuous variables is generated according to the its correspondence to z_i.

To perform statistical inference using the combined model in Section 3.1, we consider two alternative prior formulations for β and Σ. First, we take the prior distribution for β to be $N(0, 100)$, which is very weakly informative, and we assume Σ has a $PXW(m_0 = 10, I)$ distribution, i.e. a prior guess that the covariance matrix is equal to the identity matrix with ten degrees of freedom. A proper prior distribution for the 9 by 9 covariance matrix Σ requires m_0 to be greater than or equal to 9, and thus the prior distribution reflects a weakly informative belief in a scenario where the levels of the nominal variables, the latent variables of the ordinal outcomes and the normal continuous variables have no association with one another. We also examined a strongly informative prior scenario with a $N(0, 1)$ prior distribution for β and a $PXW(m_0 = 50, CS(0.4))$ prior distribution for Σ, where $CS(0.4)$ indicates a compound symmetry structure with equal correlation 0.4. The degrees-of-freedom parameter m_0 in this case reflects a strong prior belief that the covariance matrix has a CS(0.4) structure. We label the first approach PXW_I_weak and the second approach PXW_CS_strong. We set the non-informative prior for four unknown cutpoints denoted by $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ with γ_1 and γ_2 for y_{i3}, and γ_3 and γ_4 for y_{i4} in both PXW_I_weak and PXW_CS_strong scenarios, i.e., $p(\gamma_1, \gamma_2, \gamma_3, \gamma_4) \propto 1$. These two prior scenarios are illustrated in Fig. 1, showing that the PXW_CS_strong scenario gives much tighter information for all of the parameters than the PXW_I_weak scenario does.

We ran the MCMC algorithm for 201,000 iterations, discarding the first 1000 iterations as a burn-in period for each of these two prior distribution scenarios. Acceptance rates for the proposed draws in the PX-MH step were roughly 15% under both prior scenarios for both simulated data. The posterior means and posterior standard deviations for the regression parameter β, the variance parameters $(\sigma_{22}, \sigma_{44}, \sigma_{55}, \sigma_{88}, \sigma_{99})$ and four cutpoints $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are presented in Table 1. Table 2 contains the posterior mean and standard deviation for each correlation. Through Table 1, we see the estimated values for the regression parameters, the variance parameters and the cutpoints are quite similar under each prior scenario for both simulated data. However, the estimated correlation parameters appear to depend somewhat on the specification of prior distributions for the data with sample size equal to 200. Not surprisingly, the posterior means of the correlations under the PXW_CS_strong scenario are pulled toward the assumed value of 0.4. In comparison with the data with sample size equal to 200, the data with sample size equal to 2000 gives much more similar estimated correlation parameters under these two prior scenarios. Also, the posterior standard deviations for all parameters under the PXW_CS_strong scenario are uniformly smaller than those under the PXW_I_weak scenario. The coverage of true values appears satisfactory in both scenarios with different sample size, partly because substantial posterior uncertainty remains. In stands to reason that better prior specification may give better estimated values, but, in general, inference appears to be fairly robust to the choice between these two groups of priors, presumably because the sample sizes of 200 and 2000 are sufficient to dominate the prior in either scenario. Unsurprisingly, the data with larger sample size gives better estimated values and is more robust to

Table 1
Posterior means (posterior standard deviations) for the regression parameter (β), five variance parameters ($\sigma_{22}, \sigma_{44}, \sigma_{55}, \sigma_{88}, \sigma_{99}$) and four cutpoints ($\gamma_1, \gamma_2, \gamma_3, \gamma_4$) under PXW_I_weak and PXW_CS_strong scenarios. This table shows that the posterior means for all parameters are similar under these two groups of priors and the posterior standard deviations under the PXW_CS prior are uniformly smaller than those under the PXW_I prior. Under both PXW_I_weak and PXW_CS_strong scenarios, the posterior standard deviations for the data with the sample size equal to 2000 are uniformly smaller than those for the data with the sample size equal to 200.

Parameters	True	$n = 200$		$n = 2000$	
		PXW_I_weak	PXW_CS_strong	PXW_I_weak	PXW_CS_strong
β	2.0	2.00(0.11)	1.94(0.10)	2.04(0.04)	2.02(0.04)
σ_{22}	1.00	1.11(0.31)	0.95(0.15)	0.98(0.10)	0.92(0.08)
σ_{44}	1.00	1.04(0.21)	0.91(0.14)	1.07(0.14)	0.96(0.10)
σ_{55}	1.00	0.92(0.26)	0.88(0.15)	1.01(0.14)	0.92(0.10)
σ_{88}	1.00	1.14(0.11)	1.08(0.10)	0.97(0.03)	0.97(0.03)
σ_{99}	1.00	1.16(0.11)	1.12(0.10)	1.02(0.03)	1.02(0.03)
γ_1	0.50	0.51(0.07)	0.51(0.07)	0.47(0.02)	0.47(0.02)
γ_2	1.00	0.98(0.10)	0.97(0.10)	0.96(0.03)	0.96(0.03)
γ_3	0.50	0.52(0.08)	0.52(0.07)	0.47(0.02)	0.46(0.02)
γ_4	1.00	1.05(0.11)	1.04(0.10)	1.02(0.03)	1.02(0.03)

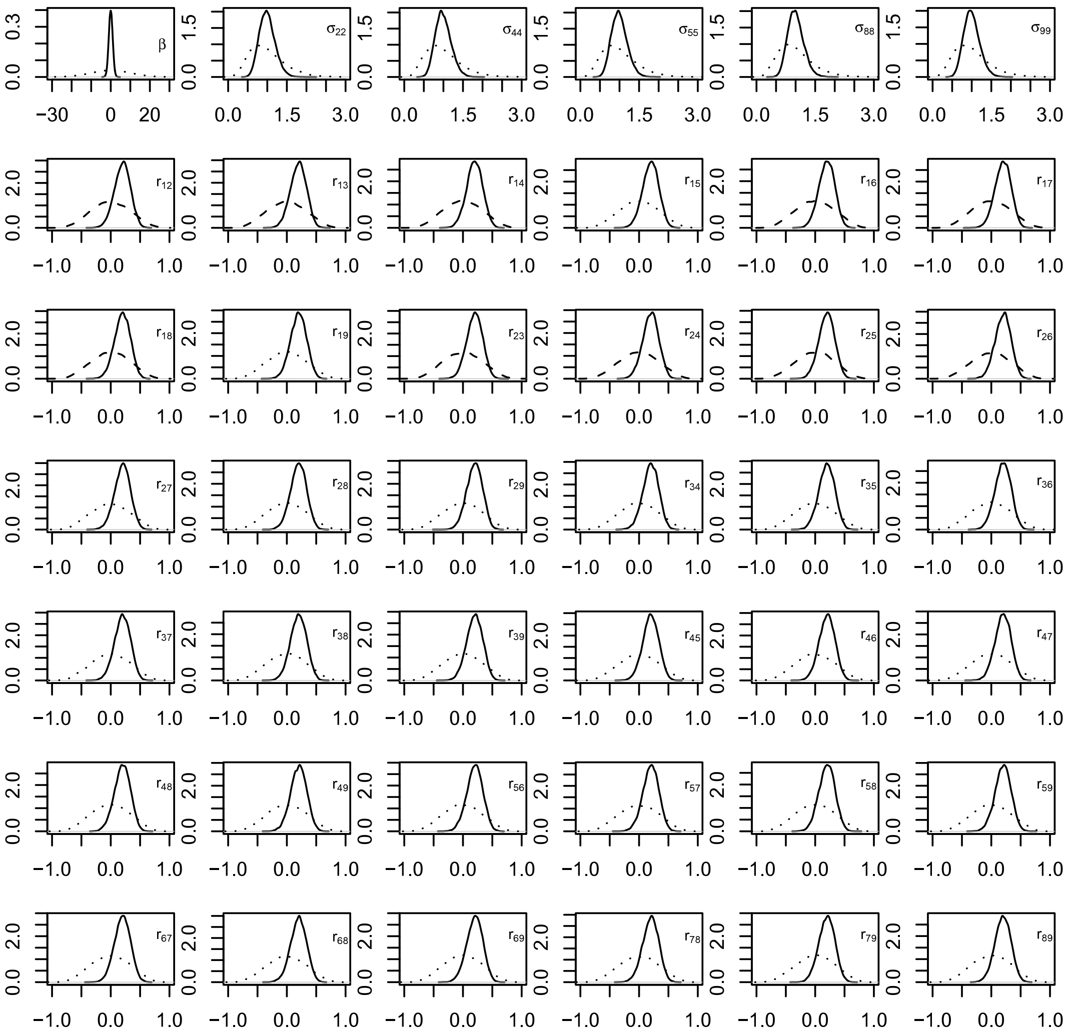

Fig. 1. Prior density plots for the regression parameter (β), variance parameters ($\sigma_{22}, \sigma_{44}, \sigma_{55}, \sigma_{88}, \sigma_{99}$), and 36 correlations ($r_{ij}$). The dotted lines are for the PXW_I_weak scenario, the solid lines are for the PXW_CS_strong scenario.

the prior choices in comparison with the one with smaller sample size. Further illustration for marginal posterior densities for each simulated data scenario are presented in Figs. 2 and 3, respectively.

Table 2
The posterior means and standard deviations for the correlations of the vector $(z_{i1}, z_{i2}, z_{i3}, z_{i4}, z_{i5}, z_{i6}, z_{i7}, y_{i5}, y_{i6})$. Above diagonal: the posterior correlation means (posterior standard deviations) for the data with 2000 sample size under PXW_CS_strong prior and PXW_CS_weak prior. Below diagonal: the posterior correlation means (posterior standard deviations) for the data with 200 sample size under PXW_CS_strong prior and PXW_CS_weak prior.

True	0.30	0.20	0.20	0.20	0.20	0.20	0.20	0.20
PXW_CS	0.27	0.26	0.22	0.23	0.21	0.22	0.26	0.21
	(0.23)	(0.22)	(0.22)	(0.23)	(0.19)	(0.19)	(0.18)	(0.18)
PXW_I	0.26	0.24	0.19	0.21	0.19	0.21	0.26	0.20
	(0.23)	(0.23)	(0.23)	(0.23)	(0.19)	(0.20)	(0.18)	(0.18)
0.30	True	0.20	0.20	0.20	0.20	0.20	0.20	0.20
0.30	PXW_CS	0.17	0.13	0.24	0.23	0.19	0.23	0.16
(0.31)		(0.23)	(0.22)	(0.23)	(0.19)	(0.19)	(0.18)	(0.18)
0.24	PXW_I	0.14	0.09	0.22	0.21	0.17	0.21	0.14
(0.40)		(0.24)	(0.24)	(0.23)	(0.19)	(0.20)	(0.18)	(0.18)
0.2	0.2	True	0.3	0.3	0.2	0.2	0.2	0.2
0.27	0.26	PXW_CS	0.23	0.27	0.27	0.25	0.20	0.24
(0.32)	(0.31)		(0.26)	(0.25)	(0.20)	(0.20)	(0.19)	(0.18)
0.20	0.10	PXW_I	0.21	0.24	0.26	0.24	0.19	0.23
(0.39)	(0.38)		(0.29)	(0.29)	(0.20)	(0.21)	(0.19)	(0.19)
0.2	0.2	0.3	True	0.3	0.2	0.2	0.2	0.2
0.23	0.34	0.25	PXW_CS	0.28	0.25	0.29	0.18	0.16
(0.32)	(0.31)	(0.33)		(0.26)	(0.20)	(0.20)	(0.19)	(0.19)
0.13	0.26	0.06	PXW_I	0.31	0.24	0.28	0.16	0.14
(0.39)	(0.37)	(0.45)		(0.28)	(0.20)	(0.21)	(0.19)	(0.19)
0.2	0.2	0.3	0.3	True	0.2	0.2	0.2	0.2
0.30	0.29	0.28	0.29	PXW_CS	0.19	0.19	0.17	0.22
(0.31)	(0.30)	(0.33)	(0.32)		(0.20)	(0.20)	(0.18)	(0.19)
0.22	0.09	0.14	0.17	PXW_I	0.17	0.18	0.15	0.20
(0.39)	(0.40)	(0.44)	(0.42)		(0.21)	(0.21)	(0.19)	(0.19)
0.2	0.2	0.2	0.2	0.2	True	0.3	0.2	0.2
0.09	0.18	0.27	0.21	0.23	PXW_CS	0.26	0.19	0.22
(0.30)	(0.29)	(0.30)	(0.30)	(0.30)		(0.17)	(0.16)	(0.16)
−0.13	0.01	0.16	0.04	0.08	PXW_I	0.25	0.18	0.21
(0.34)	(0.33)	(0.34)	(0.35)	(0.35)		(0.17)	(0.16)	(0.16)
0.2	0.2	0.2	0.2	0.2	0.3	True	0.2	0.2
0.30	0.35	0.25	0.28	0.36	0.36	PXW_CS	0.24	0.23
(0.29)	(0.29)	(0.30)	(0.30)	(0.29)	(0.27)		(0.16)	(0.16)
0.22	0.31	0.12	0.20	0.32	0.33	PXW_I	0.23	0.23
(0.34)	(0.34)	(0.36)	(0.36)	(0.36)	(0.30)		(0.16)	(0.16)
0.2	0.2	0.2	0.2	0.2	0.2	0.2	True	0.3
0.18	0.25	0.23	0.27	0.31	0.25	0.19	PXW_CS	0.27
(0.28)	(0.28)	(0.28)	(0.28)	(0.28)	(0.26)	(0.27)		(0.14)
0.06	0.16	0.12	0.19	0.24	0.19	0.09	PXW_I	0.27
(0.31)	(0.31)	(0.32)	(0.32)	(0.33)	(0.28)	(0.29)		(0.14)
0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	True
0.21	0.36	0.29	0.34	0.41	0.27	0.25	0.28	PXW_CS
(0.28)	(0.27)	(0.28)	(0.28)	(0.27)	(0.26)	(0.26)	(0.23)	
0.11	0.31	0.20	0.28	0.39	0.22	0.16	0.24	PXW_I
(0.31)	(0.31)	(0.32)	(0.32)	(0.32)	(0.28)	(0.28)	(0.25)	

The convergence of the MCMC algorithm was assessed by several procedures recommended by Cowles and Carlin [10]. We calculated Gelman and Rubin's potential scale reduction factor, $\sqrt{\hat{R}}$ for five dispersed chains with the first 1000 iterations discarded as burn-in [20]. The jumping distribution degrees-of-freedom parameter of $m = 1200$ for the data with sample size equal to 200 and of $m = 10\,000$ for the sample size equal to 2000 gave an acceptance rate of about 15% for the PX-MH step of the algorithm. Although this is below the value of 25% recommended by Roberts and Sahu [57], we find in practice that higher values for m substantially increased autocorrelations. For the single regression parameter β, 36 correlation parameters r_{ij}, five variance parameters (σ_{22}, σ_{44}, σ_{55}, σ_{88}, and σ_{99}), and four cutpoints (g_1, g_2, g_3, g_4), the values of $\sqrt{\hat{R}}$ for these two simulated data were all below 1.25 after 40,000 iterations and were all below 1.04 after further 60,000 iterations. The multivariate potential scale reduction factor for these 46 parameters was 1.21 after 40,000 iterations, improving to 1.04 at 100,000 iterations for each simulated data.

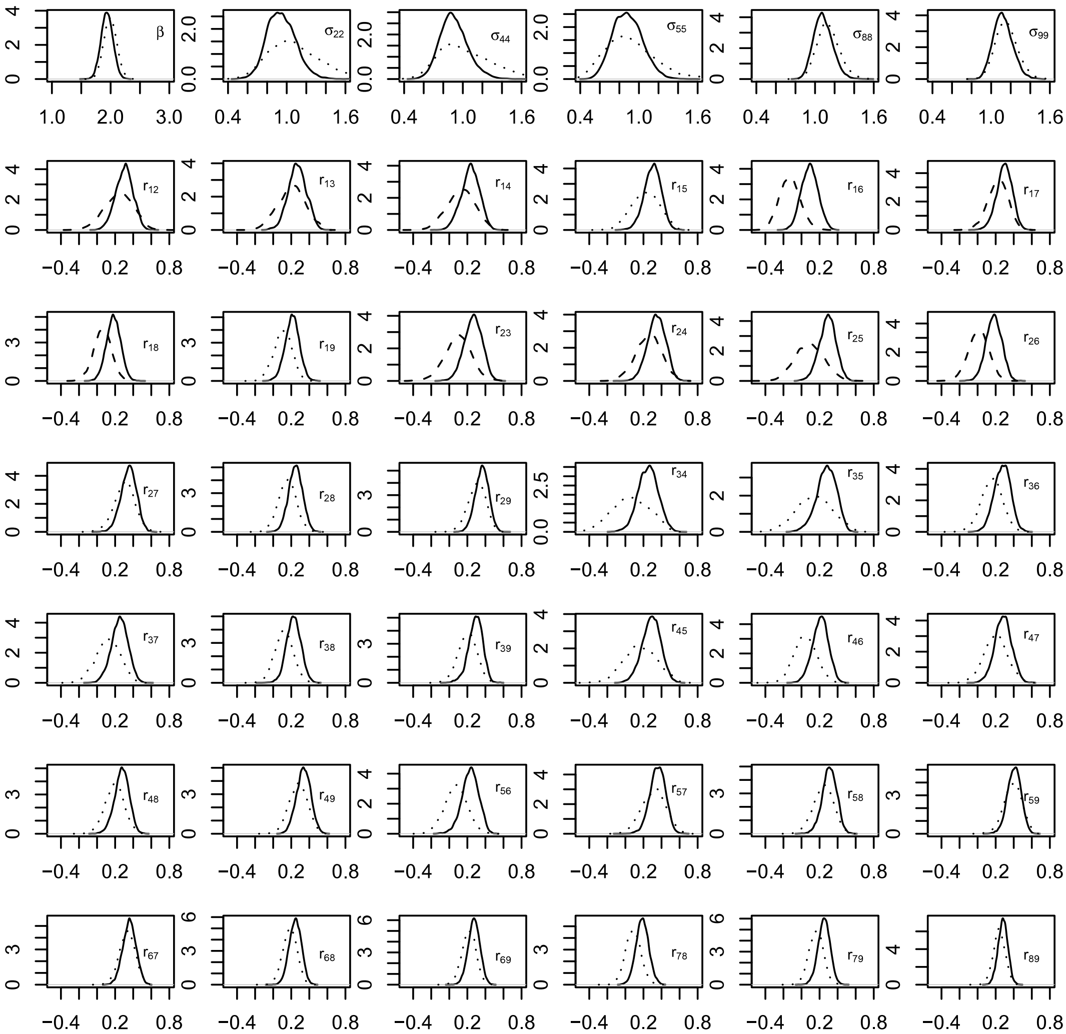

Fig. 2. Posterior density plots for the regression parameter (β), variance parameters (σ_{22}, σ_{44}, σ_{55}, σ_{88}, σ_{99}), and 36 correlations (r_{ij}) for the data set with 200 sample size. Dotted lines correspond to the PXW_I_weak scenario, and solid lines correspond to the PXW_CS_strong scenario. The plots show that the posterior distributions are robust to the prior specification for the data with 200 sample size.

5. Application to the St Louis Risk Research Project data

Table 3 presents the descriptive information for the St Louis Risk Research Project data used by Little and Schluchter [37], Schafer [59], and Liu and Rubin [39]. The data set was collected from 69 families and contains one categorical variable G with three categories indicating the parental risk levels, two binary variables D_1 and D_2 indicating the levels of psychiatric symptoms for first and second child, and four continuous variables V_1, V_2, R_1, and R_2 being the verbal and reading comprehension scores for first and second child, respectively. From Table 3, we can see that except for categorical variable G with no missing values, the other variables' missing percentages are not trivial. The last column in Table 3 contains the two latent variables $z_{i,1}$ and $z_{i,2}$ from the categorical variable G, two latent variables $z_{i,3}$ and $z_{i,4}$ from binary D_1 and D_2, respectively, and four continuous variables $y_{i,5}$, $y_{i,6}$, $y_{i,7}$ and $y_{i,8}$ for the verbal and reading comprehension scores. To illustrate our methodology using this example, we apply our proposed method through modeling latent variables $z_{i,1}$, $z_{i,2}$, $z_{i,3}$, $z_{i,4}$ and the continuous variables $y_{i,5}$, $y_{i,6}$, $y_{i,7}$, $y_{i,8}$ simultaneously. We assume non-informative or uniform priors for all unknown parameters and the covariance matrix. We ran the MCMC sampling algorithm elaborated in Section 3 and checked the convergence after running 2,000,000 iterations, yielding a 14.5% acceptance rate. The MCMC samplers pass the criterion proposed by Heidelberger and Welch [29] with Monte Carlo errors less than 0.01 using R package BOA.

Table 4 contains the posterior means and standarddeviations for all eight regression parameters (identity matrix is assumed for design matrix), one variance from categorical data G, four variances for all the continuous score variables. Table 5 presents the posterior means (above the diagonal) and the posterior standard deviations (below the diagonal) for the 8 by 8 correlation matrix. We can see that the parental risk levels (detailed definition can be referred to Little and Schluchter [37]) are significantly correlated with both children's psychiatric symptoms; there are positive correlations among the children's

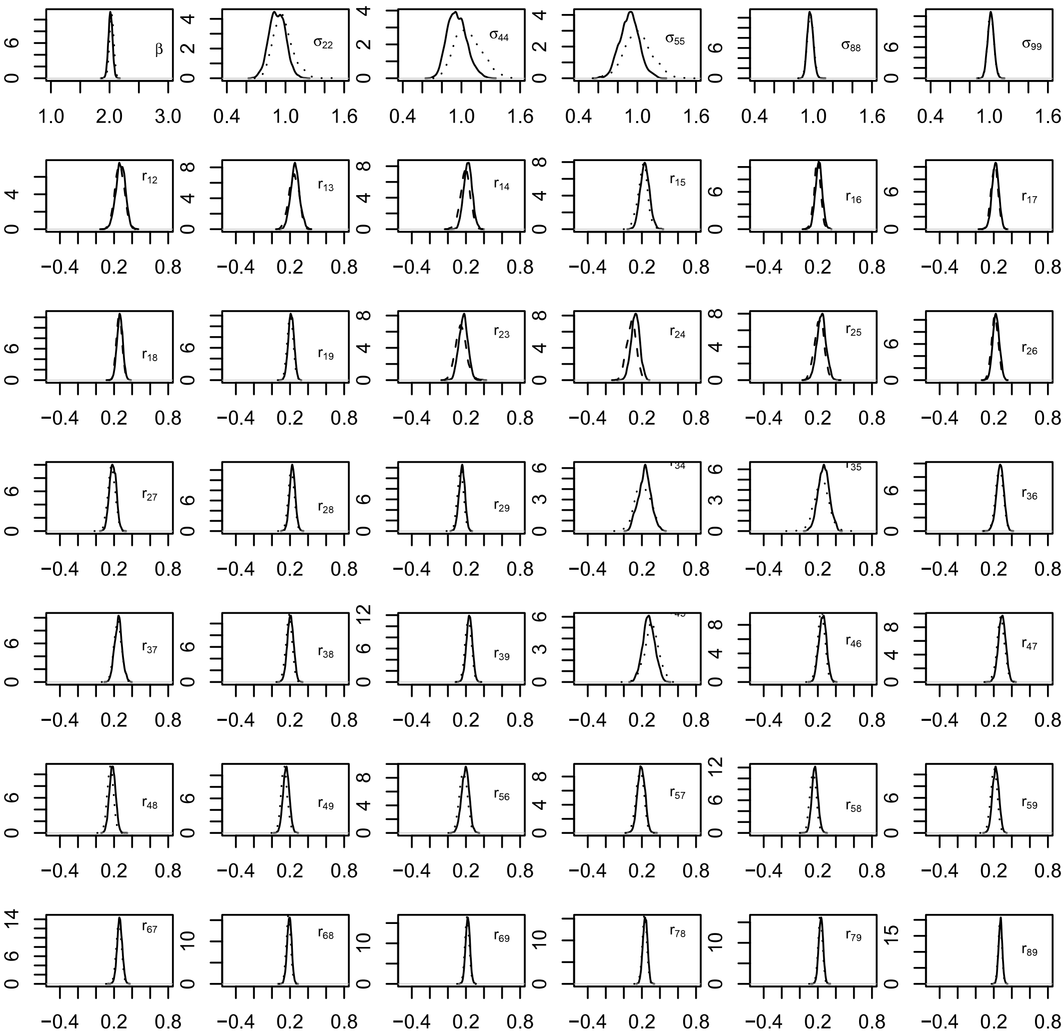

Fig. 3. Posterior density plots for the regression parameter (β), variance parameters (σ_{22}, σ_{44}, σ_{55}, σ_{88}, σ_{99}), and 36 correlations (r_{ij}) for the data set with 2000 sample size. Dotted lines correspond to the PXW_I_weak scenario, and solid lines correspond to the PXW_CS_strong scenario. The plots show that the posterior distributions are quite robust to the prior specification for the data with 2000 sample size.

Table 3
Description of variables in St Louis Risk Research Project data.

Variable	Description	Categories	Sample size	Missing (%)	Latent variables
G	Parental risk group	1. Normal	27	0.0	(Reference)
		2. Moderate	24		$z_{i,1}$
		3. High	18		$z_{i,2}$
D_1	Symptoms of first child	1. Low	21	40.6	(Reference)
		2. High	20		$z_{i,3}$
D_2	Symptoms of second child	1. Low	13	40.6	(Reference)
		2. High	28		$z_{i,4}$
V_1	Standard verbal comprehension score for first child			43.5	$y_{i,5}$
V_2	Standard verbal comprehension score for second child			24.6	$y_{i,6}$
R_1	Standard reading comprehension score for first child			30.4	$y_{i,7}$
R_2	Standard reading comprehension score for second child			23.2	$y_{i,8}$

testing scores and the parental risk levels and the children's psychiatric symptoms; and significant correlations exist among children's verbal and reading scores. It seems that the parental risk levels affect both children's psychiatric symptoms, but do not have a significant negative effect on children's verbal and reading scores.

Table 4
Posterior means and posterior standard deviations for the regression parameter (β), one variance parameters (σ_{22}) from the categorical variable G, and four variances $\sigma_{55}, \sigma_{66}, \sigma_{77}, \sigma_{88}$ for four continuous variables V_1, V_2, R_1, R_2.

Parameters	Posterior means	Posterior standard deviations
β_1	−0.16	0.185
β_2	−0.30	0.201
β_3	−0.10	0.199
β_4	0.61	0.223
β_5	122.8	0.895
β_6	116.0	0.822
β_7	106.9	0.599
β_8	103.8	0.567
σ_{22}	0.97	0.541
σ_{55}	37.43	1.07
σ_{66}	37.86	1.08
σ_{77}	18.57	0.90
σ_{88}	17.88	0.89

Table 5
The posterior means and standard deviations for the correlations of the vector ($z_{i1}, z_{i2}, z_{i3}, z_{i4}, y_{i5}, y_{i6}, y_{i7}, y_{i8}$). Above diagonal: the posterior correlation means. Below diagonal: the posterior standard deviations.

1.00	0.30	0.46*	0.41*	−0.03	0.01	0.08	0.16*
0.45	1.00	0.35*	0.49*	−0.01	0.05	0.10	0.05
0.38	0.40	1.00	0.42*	−0.03	0.22*	0.12	0.22*
0.38	0.36	0.37	1.00	0.11*	0.16*	0.28*	0.46*
0.21	0.22	0.22	0.22	1.00	0.51*	0.61*	0.36*
0.22	0.21	0.25	0.24	0.14	1.00	0.43*	0.52*
0.25	0.25	0.25	0.26	0.14	0.16	1.00	0.33*
0.25	0.25	0.28	0.28	0.16	0.15	0.18	1.00

* Indicates those 95% credible intervals excluding 0.

Table 6
Expected frequencies for each cell and estimated cell means. The first row is listed by 'C' for cell, 'F' for frequency, 'V_1' for verbal score for first child, 'V_2' for verbal score for second child, 'R_1' for reading score for first child, 'R_2' for reading score for second child. Column denoted by A and column denoted by B are for two maximum likelihood methods (without and with cell probability restrictions) from Little and Schluchter [37] and column denoted by M is for using our proposed MCMC algorithm.

C			F			V_1			V_2			R_1			R_2		
G	D_1	D_2	A	B	M	A	B	M	A	B	M	A	B	M	A	B	M
1	1	1	10.2	4.8	10.2	133.7	140.9	130.1	119.4	129.5	118.3	110.2	113.6	109.1	99.8	103.0	101.3
1	1	2	9.0	8.8	9.8	161.1	160.1	146.8	132.1	131.0	127.1	123.4	122.8	118.1	116.0	115.4.0	113.4
1	2	1	3.6	3.7	3.4	147.7	136.9	135.6	126.9	111.6	135.4	111.2	105.3	111.8	110.0	101.7	103.5
1	2	2	4.2	9.7	3.6	123.9	120.8	138.5	151.4	148.0	148.1	118.0	114.5	116.0	111.9	111.1	120.0
2	1	1	2.2	4.3	2.6	81.1	81.7	98.6	103.3	104.2	88.9	87.6	88.4	94.2	101.1	101.5	89.1
2	1	2	7.2	7.8	6.2	134.6	134.8	124.5	109.6	109.0	110.0	104.3	104.4	104.0	109.4	109.6	108.6
2	2	1	2.3	3.3	2.8	122.6	122.0	106.7	146.1	145.3	99.6	96.4	96.1	97.0	134.5	134.3	90.9
2	2	2	12.3	8.6	12.4	104.7	104.5	113.6	102.4	102.3	113.7	106.7	106.6	105.2	97.0	96.8	106.4
3	1	1	2.1	3.2	3.4	137.7	137.5	128.4	96.3	96.0	98.4	115.8	115.7	106.8	82.9	82.8	85.3
3	1	2	7.8	5.9	5.2	127.9	119.4	110.5	128.3	117.1	103.7	105.7	100.7	103.4	100.8	96.1	98.4
3	2	1	1.0	2.5	1.2	58.3	90.4	74.4	105.4	148.6	120.0	56.2	76.2	67.9	88.2	108.3	95.0
3	2	2	7.1	6.4	8.2	107.2	107.2	122.1	104.8	104.8	124.1	107.3	107.4	110.5	107.0	107.3	108.0

Little and Schluchter [37] applied the general location model and considered the model for each categorical cell with and without cell probability restrictions. They gave maximum likelihood estimates using EM algorithm for expected frequencies of all cells comprised by categorical variable G and two binary variables D_1 and D_2 and expected means for all continuous score variables given each cell. To make a comparison with their estimation, we calculate the averaged frequencies of all cells and the averaged means for all continuous score variables for each cell based on the MCMC sampling for imputing the missing values at 120,000th, 140,000th, 160,000th, 180,000th, and 200,000th iterations. We listed the results from Little and Schluchter [37] and ours in Table 6. The rows are for each cell, column F is for estimated frequency, and columns V_1, V_2, R_1 and R_2 are for estimated scores. In each column, A column and B column are for two maximum likelihood methods (without and with cell probability restrictions) from Little and Schluchter [37] and M column is for using our proposed MCMC algorithm. Through Table 6, we can see that our estimated cell frequencies are similar to either one of those from Little and Schluchter [37]. We also notice that the big differences found between Little and Schluchter's and ours are the estimated mean verbal and reading score for second children for cell ($G = 2, D_1 = 2, D_2 = 1$). Investigating these differences due to substantial missing percentages or modeling issues is worth further attention.

Table 7
Description of variables used in foreign language data analysis.

Variable	Description	Categories	Percentage	Missing (%)	Latent variables
LAN	Foreign language studied	French	24.0	0.0	$z_{i,1}$
		Spanish	28.0		$z_{i,2}$
		German	40.9		$z_{i,3}$
		Russian	7.1		(Reference)
AGE	Age group	Less than 20	46.3	3.9	$z_{i,4}$
		20–21	43.3		$z_{i,5}$
		22+	10.4		(Reference)
SEX		Female	45.7	0.4	$z_{i,6}$
		Male	54.3		(Reference)
PRI	Number of prior Foreign language courses	None	26.5	3.9	$z_{i,7}$
		1–2	27.2		
		3+	46.3		
GRD	Final grade in Foreign language course	C, D, F	17.2	16.8	$z_{i,8}$
		B	28.4		
		A	54.4		(Reference)
FLAS	Score on foreign language attitude scale			0.0	$y_{i,9}$
MLAT	Modern Language Aptitude Test			17.6	$y_{i,10}$
SATV	Scholastic Aptitude Test, verbal score			12.2	$y_{i,11}$
SATM	Scholastic Aptitude Test, math score			12.2	$y_{i,12}$
ENG	Score on Penn State English placement exam			13.3	$y_{i,13}$
HGPA	High school grade point average			0.4	$y_{i,14}$
CGPA	Current college grade point average			12.2	$y_{i,15}$

6. Application to a foreign language study

The foreign language study aimed to investigate the usefulness of a newly developed instrument, the Foreign Language Attitude Scale (FLAS), for predicting success in the study of foreign language, in comparison with the other established instruments, such as the Modern Language Aptitude Test (MLAT), the Scholastic Aptitude Test (SATV) and other test scores. The data were collected on a sample of 279 students who enrolled in foreign language courses at The Pennsylvania State University in the early 1980s. This data set was analyzed by Schafer [59] and de Leon and Carrière [13] using general location models.

Here we investigate the between-measure and within-measure associations across the levels of two nominal variables (*LAN* with 4 levels and *AGE* with 3 levels), one binary variable (*SEX*), two ordinal variables (*PRI* and *GRD* with 3 levels both) and 7 continuous variables (*FLAS*, *MLAT*, *SATV*, *SATM*, *ENG*, *HGPA*, *CGPA*). Descriptive information about these 12 variables is presented in Table 7.

We used our proposed model and MCMC algorithm considering missing values and chose an independent $N(0, 100)$ prior for each regression parameter and a $PXW(m_0 = 20, I)$ prior for the covariance matrix for vector z_i. We ran 401,000 iterations with 1000 burn-in iterations, yielding a roughly 15% acceptance rate for the proposed draws in the PX-MH step. We assessed the convergence of the MCMC algorithm by Cowles and Carlin [10] and values of univariate potential scale reduction factors were all well below 1.11 and the multivariate scale reduction factor was 1.21.

In Table 8, we present the significant estimated correlations with 95% credible intervals. We can see that the FLAS score is positively related to GRD (the final grade in foreign language course), MALT (Modern Language Aptitude Test) and ENG (score on Penn State English placement exam). Not surprisingly, GRD (the final grade in foreign language course) is positively correlated with MLAT, SATV, SATM, ENG, HGPA, CGPA which have higher positive correlations among each other. There are other interesting correlations, such as female students tended to get higher scores on the language tests while get lower scores on math tests, and the students who studied Spanish tended to get higher language scores while those who studied French did not.

Schafer [59] used the general location model for conducting missing data imputation on the FLAS data. However, there are total 1000 cells comprised of all the ordered and nominal categorical variables including the main variable GRD and it is not feasible to get the estimated correlation between continuous FLAS and ordinal GRD. Therefore, Schafer [59] considered a logistic or proportional odds model for GRD while adjusting FLAS and other variables. de Leon and Carrière [13] generalized the usual general location model by using grouped continuous model to estimate the correlation structure for both continuous and ordinal variables given each cell obtained by combining all categorical variables. Their method can give the estimated correlation among all the continuous and ordinal variables, and among each combined category and those continuous and ordinal variables as well. However, all these estimated correlations are separate analyses based on each combined category. In comparison with Schafer [59] and de Leon and Carrière [13], our method can perform the missing data imputation and analysis simultaneously and jointly estimate the correlations among all variables.

Table 8
Posterior means for the correlation parameters with the 95% credible intervals excluding 0.

French	FLAS	MLAT	SATV	SATM	ENG	
Mean	−0.16	−0.21	−0.20	−0.20	−0.24	
95% CI	(−0.30, −0.01)	(−0.38, −0.03)	(−0.36, −0.03)	(−0.36, −0.03)	(−0.40, −0.06)	
Spanish	GRG	MLAT	SATV	ENG		
Mean	0.24	0.22	0.16	0.20		
95% CI	(0.07, 0.41)	(0.06, 0.37)	(0.00, 0.31)	(0.04, 0.34)		
German	Age<20	PRI	SATM			
Mean	−0.26	0.26	−0.22			
95% CI	(−0.50, −0.01)	(0.02, 0.48)	(−0.42, −0.01)			
Age<20	Female	GRD	FLAS	MLAT	ENG	
Mean	−0.21	0.19	−0.17	−0.17	−0.21	
95% CI	(−0.38, −0.04)	(0.02, 0.36)	(−0.30, −0.02)	(−0.32, −0.02)	(0.35, −0.07)	
Age 20–21	Female	PRI	CGPA			
Mean	−0.39	−0.22	−0.28			
95% CI	(−0.58, −0.17)	(−0.41, −0.01)	(−0.47, −0.09)			
Female	GRD	FLAS	MLAT	SATM	CGPA	
Mean	0.20	0.29	0.18	−0.25	0.25	
95% CI	(0.05, 0.36)	(0.16, 0.42)	(0.03, 0.32)	(−0.39, −0.11)	(0.10, 0.38)	
PRI	GRD					
Mean	0.18					
95% CI	(0.02, 0.32)					
GRD	FLAS	MLAT	SATM	ENG	HGPA	CGPA
Mean	0.25	0.44	0.22	0.25	0.60	0.44
95% CI	(0.13, 0.38)	(0.32, 0.56)	(0.08, 0.35)	(0.12, 0.38)	(0.50, 0.69)	(0.32,0.56)
FLAS	MLAT	ENG				
Mean	0.12	0.14				
95% CI	(0.01, 0.24)	(0.03, 0.25)				
MLAT	SATV	SATM	ENG	HGPA	CGPA	
Mean	0.31	0.39	0.49	0.35	0.44	
95% CI	(0.19, 0.42)	(0.28, 0.49)	(0.39, 0.58)	(0.24, 0.46)	(0.33, 0.54)	
SATV	SATM	ENG	HGPA	CGPA		
Mean	0.34	0.68	0.24	0.22		
95% CI	(0.23, 0.44)	(0.62, 0.74)	(0.12, 0.35)	(0.10, 0.33)		
SATM	ENG	HGPA	CGPA			
Mean	0.37	0.26	0.32			
95% CI	(0.27, 0.47)	(0.15, 0.37)	(0.21, 0.43)			
ENG	HGPA	CGPA				
Mean	0.27	0.28				
95% CI	(0.17, 0.38)	(0.17, 0.39)				
HGPA	CGPA					
Mean	0.45					
95% CI	(0.34, 0.54)					

7. Conclusions

In this manuscript, we proposed a general model for analyzing combinations of continuous, ordinal and categorical measures with missing values, using the MCMC algorithm in a Bayesian framework.

The proposed approach for handling combinations of various types of multivariate data has several advantages. First, this model allows us to analyze continuous, ordinal and categorical multivariate data, or any subsets of these three types of data simultaneously by combining the continuous variables and latent variables from the multivariate probit model and the multivariate multinomial probit model. Second, the PX-MH algorithm for sampling the restricted covariance matrix gives the covariance structure of continuous variables and the associated latent measures and embeds a flexible prior distribution on the covariance matrix. Third, the sampling algorithm can naturally handle missing values in the multivariate measures.

Our method can be extended to handle mixed data type involving repeated continuous or ordinal or nominal measures. This may induce variance component model or autoregressive process for the covariance matrix of repeated measures and necessitates a structured instead of an unstructured covariance matrix for the mixed data to reflect the patterned covariance structure. The covariance structure for repeated continuous and ordinal measures is straightforward. The specification of the covariance matrix for repeated nominal measures can be referred to McCulloch and Rossi [42].

Due to the model identification issue, we impose restrictions on the diagonal elements of the covariance matrix for the continuous measures and the latent variables from the multivariate probit model and the multivariate multinomial probit model and apply the PX-MH algorithm to sample the covariance matrix with restrictions for the identified model. Based on the discussion by MacEachern [40], the Markov chain produced by the non-identifiable model/algorithm has better convergence rate than those from identifiable models/algorithms illustrated through an example of three-state Dirichlet process of mixture model. MacEachern [40] shows that the non-identifiable model has the smallest second eigenvalue compared with the other identified models and proves that for countable state space, one step of the chain based on larger conditioning sets (i.e., the sampler based on the non-identifiable model) is preferable to one step of the chain based on the smaller conditioning sets. Based on MacEachern [40], particular consideration should be given to inducing non-identification by adding symmetries such as relabeling of the clusters in the simple three-state Dirichlet process. Although the covariance matrix with and without restrictions is sampled from a continuous state Markov chain, the non-identifiable model may still have better convergence rate than the identifiable model. McCulloch and Rossi [42] elaborate the convergence issues for the non-identifiable multinomial probit model without imposing the restriction on the covariance matrix. Nobile [49] modifies the algorithm by McCulloch and Rossi [42] through adding one Metropolis step along a direction of constant likelihood induced by the non-identifiable model. To consider a modified algorithm based on the covariance matrix without restrictions may accelerate the MCMC convergence and is worth our attention for analyzing the mixed data types.

Further research on the use of the algorithm for multiple imputation is also worth pursuing. Another concern for using the proposed method is the normality assumption. The effect of using this approach with non-normal data, diagnosing the effect of non-normality, and proposing a more general method would be worthy areas for future investigation.

References

[1] J.H. Albert, S. Chib, Bayesian analysis of binary and polychotomous response data, J. Amer. Statist. Assoc. 88 (1993) 669–679.
[2] G. Arminger, U. Küsters, Latent trait models with indicators of mixed measurement level, in: Latent Trait and Latent Class Models, Plenum, New York, 1988.
[3] J.R. Ashford, R.R. Sowden, Multivariate probit analysis, Biometrics 26 (1970) 535–546.
[4] C.I. Bliss, The calculation of the dosage-mortality curve, Ann. Appl. Biol. 22 (1935) 134–167.
[5] J. Boscardin, X. Zhang, T. Belin, Modeling a mixture of ordinal and continuous repeated outcomes, J. Stat. Comput. Simul. 78 (2008) 873–886.
[6] P.J. Catalano, L.M. Ryan, Bivariate latent variable models for clustered discrete and continuous outcomes, J. Amer. Statist. Assoc. 87 (1992) 651–658.
[7] M. Chavance, S. Escolano, M. Romon, A. Basdevant, B. de Lauzon-Guillain, M.A. Charles, Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology, BMC Med. Res. Methodol. 10 (2010) 37.
[8] Z. Chen, L. Kuo, Discrete choice models based on the scale mixture of multivariate normal distributions, Indian J. Statist. 64 (2002) 192–213.
[9] S. Chib, E. Greenberg, Analysis of multivariate probit models, Biometrika 85 (1998) 347–361.
[10] M.K. Cowles, B.P. Carlin, Markov chain Monte Carlo convergence diagnostics: a comparative review, J. Amer. Statist. Assoc. 91 (1996) 883–904.
[11] M.J. Daniels, R.E. Kass, Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models, J. Amer. Statist. Assoc. 94 (1999) 1254–1263.
[12] M.J. Daniels, S.-L.T. Normand, Longitudinal profiling of health care units based on continuous and discrete patient outcomes, Biostatistics 7 (2006) 1–15.
[13] A.R. de Leon, K.C. Carrière, General mixed-data model: extension of general location and grouped continuous models, Canad. J. Statist. 35 (2007) 533–548.
[14] A.R. de Leon, B. Wu, Copula-based regression models for a bivariate mixed discrete and continuous outcomes, Stat. Med. 30 (2011) 175–185.
[15] P.J. Diggle, K.-Y. Liang, S.L. Zeger, Analysis of Longitudinal Data, Oxford Science Publications, 1994.
[16] F. Drasgow, Polychoric and polyserial correlations, in: Encyclopedia of Statistical Sciences, John Wiley & Sons, 1986.
[17] D.B. Dunson, Bayesian latent variable models for clustered mixed outcomes, J. R. Stat. Soc. Ser. B (2000) 355–366.
[18] Y.D. Edwards, G.M. Allenby, Multivariate analysis of multiple response data, J. Mark. Res. XL (2003) 321–334.
[19] G.M. Fitzmaurice, N.M. Laird, Regression models for mixed discrete and continuous responses with potentially missing values, Biometrics 53 (1997) 110–122.
[20] A. Gelman, D.B. Rubin, Inference from iterative simulation using multiple sequences, Statist. Sci. 7 (1992) 457–511. (with discussion).
[21] J. Geweke, M. Keane, D. Runkle, Alternative computational approaches to inference in the multinomial probit model, Rev. Econ. Stat. 76 (1994) 609–632.
[22] J. Geweke, M. Keane, D. Runkle, Recursively simulating multinomial multiperiod probit probabilities, Tech. Rep., Proceedings of the American Statistical Association, Business and Economic Statistics Section, 1994.
[23] J. Geweke, M. Keane, D. Runkle, Statistical inference in the multinomial multiperiod probit model, J. Econometrics 80 (1997) 125–165.
[24] H. Geys, M.M. Regan, P.J. Catalano, G. Molenberghs, Two latent variable risk assessment approaches for mixed continuous and discrete outcomes from developmental toxicity data, J. Agric. Biol. Environ. Stat. 6 (2001) 340–355.
[25] H. Goldstein, J. Carpenter, M.G. Kenward, K.A. Levin, Multilevel models with multivariate mixed response types, Statist. Model. 9 (2009) 173–197.
[26] T. Golob, A. Regan, Trucking industry adoption of information technology: a multivariate discrete choice model, Transp. Res. C 10 (2002) 205–228.
[27] W. Greene, Convenient estimators for the panel probit model: further results, Empir. Econom. 29 (2004) 21–47.
[28] V.A. Hajivassiliou, D.L. McFadden, P.A. Ruud, Simulation of multivariate normal orthan probabilities: methods and programs. MIT Mimeo, 1990.
[29] P. Heidelberger, P.D. Welch, Simulation run length control in the presence of an initial transient, Comput. Oper. Res. 31 (1983) 1109–1144.
[30] K.G. Jöreskog, D. Sörbom, LISREL VI-Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables and Least Squares Methods, Scientific Software, Mooresville, Indiana, 1984.
[31] N.M. Laird, J.H. Ware, Random-effects models for longitudinal data, Biometrics 38 (1982) 963–974.
[32] S.-Y. Lee, H.-T. Zhu, Statistical analysis of nonlinear structural equation models with continuous and polytomous data, British J. Math. Statist. Psych. 53 (2000) 209–232.
[33] J.M. Legler, M. Lefkopoulou, L.M. Ryan, Efficiency and power of tests for multiple binary outcomes, J. Amer. Statist. Assoc. 90 (1995) 680–693.
[34] K.-Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.
[35] K.-Y. Liang, S.L. Zeger, A class of logistic regression models for multivariate binary time series, J. Amer. Statist. Assoc. 84 (1989) 447–451.
[36] K.-Y. Liang, S.L. Zeger, B. Qaqish, Multivariate regression analyses for categorical data, J. R. Stat. Soc. Ser. B 54 (1992) 3–40.
[37] R.J.A. Little, M.D. Schluchter, Maximum likelihood estimation for mixed continuous and categorical data with missing values, Biometrika 72 (1985) 497–512.

[38] C. Liu, Bayesian analysis of multivariate probit model: discussion of ''the art of data augmentation'' by Van Dyk and Meng, J. Comput. Graph. Statist. 10 (2001) 75–81.
[39] C. Liu, D.B. Rubin, Ellipsoidally symmetric extensions of the general location model for mixed categorical and continuous data, Biometrika 3 (1998) 673–688.
[40] S.N. MacEachern, Comment on article by Jain and Neal, Bayesian Anal. 2 (2007) 483–494.
[41] R.E. McCulloch, N.G. Polson, P.E. Rossi, A Bayesian analysis of the multinomial probit model with fully identified parameters, J. Econometrics 99 (2000) 173–193.
[42] R. McCulloch, P.E. Rossi, An exact likelihod analysis analysis of the multinomial probit model, J. Econometrics 64 (1994) 207–240.
[43] D. McFadden, A method of simulated moments for estimation of discrete response models without numerical integration, Econometrica 57 (1989) 995–1026.
[44] D.L. Miglioretti, Latent transition regression for mixed outcomes, Biometrics 59 (2003) 710–720.
[45] M.E. Miller, C.S. Davis, J.R. Landis, The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares, Biometrics 49 (1993) 1033–1044.
[46] I. Moustaki, M. Knott, Generalized latent trait models, Psychometrika 65 (2000) 391–411.
[47] B. Muthén, A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators, Psychometrika 1 (1984) 115–132.
[48] B. Nandram, M.-H. Chen, Accelerating Gibbs sampler convergence in the generalized linear models via a reparameterization, J. Stat. Comput. Simul. 81 (1994) 27–40.
[49] A. Nobile, A hybrid Markov chain for the Bayesian analysis of the multinomial probit model, Stat. Comput. 8 (1998) 229–242.
[50] I. Olkin, R. Tate, Multivariate correlation models with discrete and continuous variables, Ann. Math. Statist. 32 (1961) 448–465.
[51] A.J. O'Malley, S. Lise Normand, R.E. Kuntz, Application of models for multivariate mixed outcomes to medical device trials: coronary artery stenting, Stat. Med. 22 (2003) 313–336.
[52] R.L. Prentice, Correlated binary regression with covariates specific to each binary observation, Biometrics 44 (1988) 1033–1048.
[53] Y. Qu, M. Piedmonte, S.V. Medendorp, Latent variable models for clustered ordinal data, Biometrics 51 (1995) 268–275.
[54] M.M. Regan, P.J. Catalano, Regression models and risk estimation for mixed discrete and continuous outcomes in developmental toxicology, Risk Anal. 20 (2000) 363–376.
[55] U. Rendtel, U. Kaltenborn, Diskussionsbeiträge des Fachbereichs Wirtschaftswissenschaft der FU Berlin, Tech. Rep., Volkswirtschaftliche Reihe, 2004.
[56] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, 2000.
[57] G. Roberts, S. Sahu, Updating schemes, correlation structure, blocking and parameterisation for the Gibbs sampler, J. R. Stat. Soc. Ser. B 59 (1997) 291–317.
[58] M.D. Sammel, L.M. Ryan, J.M. Legler, Latent variable models for mixed discrete and continuous outcomes, J. R. Stat. Soc. Ser. B 59 (1997) 667–678.
[59] J.L. Schafer, Analysis of Incomplete Multivariate Data, Chapman & Hall/CRC, 1997.
[60] J.-Q. Shi, S.-Y. Lee, Latent variable models with mixed continuous and polytomous data, J. R. Stat. Soc. 62 (2000) 77–87.
[61] M. Spiess, Estimation of a two-equation panel model with mixed continuous and ordered categorical outcomes and missing data, Appl. Stat. 55 (2006) 525–538.
[62] M.A. Tanner, W.H. Wong, The calculation of posterior distributions by data augmentation, J. Amer. Statist. Assoc. 82 (1987) 528–540.
[63] J.H. Ware, Linear models for the analysis of longitudinal studies, J. Amer. Statist. Assoc. 39 (1985) 95–101.
[64] R.E. Weiss, J. Jia, M.A. Suchard, A Bayesian model for the common effects of multiple predictors on mixed outcomes, Inferface Focus 1 (2011) 886–894.
[65] S.L. Zeger, The analysis of discrete longitudinal data: commentary, Stat. Med. 7 (1987) 161–168.
[66] S.L. Zeger, K.-Y. Liang, Longitudinal data analysis for discrete and continuous outcomes, Biometrics 42 (1986) 121–130.
[67] S.L. Zeger, K.-Y. Liang, P.S. Albert, Models for longitudinal data: a generalized estimating equation approach, Biometrics 44 (1988) 1049–1060.
[68] X. Zhang, W.J. Boscardin, T.R. Belin, Sampling correlation matrices in Bayesian models with correlated latent variables, J. Comput. Graph. Statist. 15 (2006) 880–896.
[69] X. Zhang, W.J. Boscardin, T.R. Belin, Bayesian analysis of multivariate nominal measures using multivariate multinomial probit models, Comput. Statist. Data Anal. 52 (2008) 3697–3708.
[70] A. Ziegler, Simulated classical tests in the multiperiod multinomial probit model, Tech. Rep., Center for European Economic Research, Mannheim, 2002.