Journal of Statistical Theory and Applications, Vol. 15, No. 4 (December 2016), 373-386

A Bayesian Joint Modeling Using Gaussian Linear Latent Variables for Mixed Correlated Outcomes with Possibility of Missing Values

Sayed Jamal Mirkamali
Shahid Beheshti University, Tehran, Iran

Mojtaba Ganjali*
Shahid Beheshti University, Tehran, Iran
m-ganjali@sbu.ac.ir

Received 12 December 2015

Accepted 5 May 2016

This paper proposes a Bayesian approach for the analysis of mixed correlated nominal, ordinal and continuous outcomes with possibility of missing values using a variation of Markov Chain Monte Carlo (MCMC) method named Parameter Expanded and Reparamerized Metropolis Hastings (PX-RPMH) method. A general latent variable model is given and a Gibbs sampler is developed to incorporate PX-RPMH and Data Augmentation (DA) steps, to estimate parameters and to impute missing values. The performance of the algorithm is evaluated by some simulation studies. An application of the model to the foreign language attitude scale dataset is also enclosed.

Keywords: Mixed Data; Correlated Outcomes; Parameter Expansion; MCMC; Data Augmentation; Longitudinal Data.

2000 Mathematics Subject Classification: 62F15, 62H86, 62H25

1. Introduction

In many longitudinal studies and surveys the multivariate measurements with different scales are of the main interest. For example in Household Budget Surveys (HBS) there are a lot of questions about possession, level of income and consumption. Most of the time, there are some correlations between the responses. So, it is important to have statistical tools for analyzing such data. Many statistical procedures are developed to analyze different types of outcomes. The standard multivariate regression and cluster analysis for continuous outcomes, logistic regression and multinomial regression for nominal outcomes and cumulative probit and ordered logistic regression for ordinal outcomes are widely used to analyze these data. Olkin and Tate [31] developed a model for categorical and continuous outcomes named the general location model (GLOM). They modeled categorical and continuous outcomes by decomposing the joint model as a marginal multinomial

*Correspondence author, Professor of Statistics.

distribution for categorical variables and a conditional multivariate normal distribution for continuous variables given the categorical variables. Some extensions of the GLOM are given in Liu and Rubin [24], Peng et al. [32] and Leon et al. [21]. A second method for joint modeling is to decompose the joint distribution as a multivariate marginal distribution for the continuous responses and a conditional distribution for categorical variables given the continuous variables. Cox and Wermuth [15] empirically examined the choice between these two methods. The third method uses simultaneous modeling of categorical and continuous variables to take into account the association between responses by the correlation between errors in the model. For more details of this approach see, for example, Heckman [18] in which a general model for simultaneously analyzing two mixed correlated responses is introduced. Catalano and Ryan [11] extended and used the model for clustered observations of discrete and continuous outcomes. The possibility of missing values for mixed discrete and continuous outcomes is studied by Fitzmaurice and Laird [17]. All the above references consider correlated nominal and continuous responses. A model for ordinal and continuous responses without considering any covariate effect is also presented by Poon and Lee [33]. Latent variable models are used by Bahrami Samani et al. [4] and the effect of missing values is studied by Bahrami Samani and Ganjali [5], Bahrami Samani et al. [6] and Bahrami Samani and Tahmasebinejad [7]. Bayesian models for analyzing mixed binary and continuous responses are presented by Liu et al. [26]. Zhang et al. [41] study a joint model for analyzing mixed nominal, ordinal and continuous data with possibility of missing values. A Bayesian joint analysis of mixed ordinal and skew continuous variables is presented by Teimourian et al. [38].

In this article, we propose a Bayesian joint model using latent variables for mixed correlated nominal, ordinal and continuous outcomes. We assume a multivariate normal distribution for latent variables and use a multivariate probit model for modeling nominal outcomes, an ordered probit model for ordinal outcomes and a multivariate regression model for continuous outcomes. To make a simultaneous analysis of these models we use a Parameter Expanded and Reparametrized Metropolis Hastings (PX-RPMH) algorithm within Gibbs sampler to get MCMC samples of parameters. In contrast to Zhang et al. [41], we use two approaches to improve the speed of the algorithm. First, we use different partitioning for cut points of ordinal latent variables. Second, we choose a jointly uniform prior for the covariance matrix and a conditionally independent sampler as our proposal in the Metropolis Hastings algorithm. We consider the possibility of missing values in each outcome and assume that the mechanism of missingness is at random (MAR). We also evaluate the sensitivity of our results when the missing mechanism is not at random (NMAR), (vide, Little and Rubin [23]).

This paper is organized as follows. In Section 2 we present the general mixed correlated model and the method of parameter estimation. In Section 3 we evaluate the model using some simulation studies. An application of our model is given in Section 4. Finally, the conclusions are presented in Section 5.

2. Bayesian Modeling

2.1. *The General Model for Analyzing Mixed Correlated Responses*

The idea of analyzing correlated binary outcomes using probit model was first proposed by Ashford and Sowden [3]. McFadden [29] generalized this idea to the Multinomial Probit (MNP) model using utility functions. A lot of work has been done about resolving the identifiability and generalizing of the MNP to the multivariate cases. See for example, McCulloch and Rossi [28], Chib and Greenberg

[13], McCulloch et al. [27], Imai and Van Dyk [19], Xu and Craig [40], and Zhang et al. [41]. Here, we present notations for the Multivariate Probit (MVP) model.

Denote the $j-$th nominal outcome for subject i by N_{ij} for $j = 1,...,m_1$ where $N_{ij} \in \{0,1,...,k_j - 1\}$ and k_j is the number of levels for $j-$th nominal outcome. For each nominal outcome we consider $k_j - 1$ latent variables as $ZN_{ij,l}$ for $j = 1,...,m_1$ and $l = 1,...,k_j - 1$. We relate the latent variables and observations by:

$$N_{ij} = \begin{cases} r \neq 0 & ZN_{ij,r} = \max_{1 \leq l \leq k_j - 1} ZN_{ij,l} > 0; \\ 0 & \max_{1 \leq l \leq k_j - 1} ZN_{ij,l} \leq 0. \end{cases}$$

The latent variables correpond to each level of the nominal outcome are usually interpreted as the utility of that level in contrast to the utility of the first level of the nominal outcome.

Analyzing ordinal outcomes using probit model was first introduced by Aitchison and Silvey [1] and developed by Snell [35]. The identifiable multivariate ordered probit models studied by Nandram and Chen [30], Chen and Dey [12], Lawrence et al. [20], and Li and Schafer [22].

Denote the $j-$th ordinal outcome for subject i by O_{ij} for $j = 1,...,m_2$ where $O_{ij} \in \{1,2,...,c_j\}$ and c_j is the number of levels for ordinal outcome j. Let ZO_{ij} be the latent variable associated with the $j-$th ordinal response where ordinal observations and latent variables are related as follow

$$O_{ij} = r \Leftrightarrow \theta_{j,r-1} < ZO_{ij} \leq \theta_{j,r} \qquad r = 1,...,c_j,$$

where $-\infty = \theta_{j,0} \leq \theta_{j,1} \leq ... \leq \theta_{j,c_j-1} \leq \theta_{j,c_j} = \infty$ with $\theta_{j,1} = 0$ for $j = 1,...,m_2$. Let C_{ij} be $j-$th continuous response of subject i for $j = 1,...,m_3$.

Let $Y_i = (N_i^T, O_i^T, C_i^T)^T$ be $(m_1 + m_2 + m_3) \times 1$ vector of observations where $N_i^T = (N_{i1},...,N_{im_1})$, $O_i^T = (O_{i1},...,O_{im_2})$ and $C_i^T = (C_{i1},...,C_{im_3})$ for subject i, $i = 1,...,n$. Now suppose that $Z_i = (ZN_i^T, ZO_i^T, C_i^T)^T$ be $q \times 1$ vector of continuous variables where $q = \Sigma_{j=1}^{m_1} k_j - m_1 + m_2 + m_3$, $ZN_i^T = (ZN_{i1},...,ZN_{im_1})$, $ZN_{ij} = (ZN_{ij,1},...,ZN_{ij,k_j-1})$ for $j = 1,...,m_1$ and $ZO_i^T = (ZO_{i1},...,ZO_{im_2})$.

Suppose that there are some covariates available for subject i, denoted by a $q \times p$ matrix of X_i. So we can write a latent variable model as

$$Z_i = X_i\beta + \varepsilon_i, \tag{2.1}$$

where $\varepsilon_i \sim N_q(0,\Sigma)$ for $i = 1,...,n$. We take $Y = (Y_1,...,Y_n)$, $X = (X_1,...,X_n)$ and $Z = (Z_1,...,Z_n)$. Note that the design matrix X_i and β can be arranged so that the model can be used for analyzing both cross sectional and longitudinal data. For example, for analyzing longitudinal data with continuous outcomes suppose that $m_1 = m_2 = 0$. To have a model with time varying covariates for $Z_i = (Z_{i,t_1}, Z_{i,t_2},...,Z_{i,t_q})^T$ for time points $t_1,t_2,...,t_q$, we can consider $X_i = (x_{i,t_1}^T, x_{i,t_2}^T,...,x_{i,t_q}^T)^T$ and $\beta = (\beta_1,...,\beta_p)^T$. If we need a model with fixed covariates and varying coefficients we can consider $X_i = I_q \otimes X_{i,t_0}$ and $\beta = (\beta_{1,1}, \beta_{1,2},...,\beta_{q,p_0})^T$ where X_{i,t_0} is the $1 \times p_0$ vector of fixed covariates for subject i in time point t_0 and $p = q \times p_0$. Details can be found in Diggle et al. [16].

As noted by Chib and Greenberg [13] there are some restrictions on Σ in context of identifiability of the latent model. If we partition Σ according to $NO_i = (ZN_i, ZO_i)$ and C_i, we would have

$$\Sigma = \begin{bmatrix} \Sigma_{NO} & \Sigma_{NO,C} \\ \Sigma_{C,NO} & \Sigma_C \end{bmatrix}.$$

So, we need to take Σ_{NO} as a correlation matrix. Therefore, Σ is a covariance matrix with constraints that take diagonal elements corresponding to categorical outcomes to be equal to one. As a result of these restrictions, each of the off-diagonal elements of Σ_{NO} belongs to $[-1,1]$.

2.2. *Priors*

For Bayesian inference, we need to specify priors for parameters in the models described in Section 2.1. We take the priors as

$$\beta \sim N_p(\bar{\beta},\bar{\Sigma})$$

where $\bar{\beta}$ and $\bar{\Sigma}$ are hyperparameters that usually set to zero vector and $\sigma^2_{\beta}I$, respectively. We take

$$p(\Theta) \propto \Pi_{j=1}^{m_2} 1_{(0\leq\theta_{j,2}\leq\ldots\leq\theta_{j,c_j-1})}$$

where $\Theta = (\theta_{1,2},...,\theta_{1,c_1-1},\theta_{2,2},...,\theta_{m_2c_{m_2}-1})$. The method of data augmentation introduced by Tanner and Wong [37] is useful when models include latent variables. This method is used by Albert and Chib [2] and Van Dyk and Meng [39] for Bayesian analysis of binary outcomes. Hence, we can obtain the posterior distribution of parameters through

$$p(\beta,\Theta,\Sigma,Z|Y,X) \propto p(\beta)p(\Theta)p(\Sigma)\prod_{i=1}^{n} INO_i \times \phi_p(Z_i;X_i\beta,\Sigma) \tag{2.2}$$

where $\phi_p(\cdot;\mu,\Omega)$ represents the density function of $N_p(\mu,\Omega)$ distribution and INO_i is a compatibility indication which is equal to 1 if all latent variables $({ZN_i}^T,{ZO_i}^T)^T$ are compatible with their corresponding observations $({N_i}^T,{O_i}^T)$, and is equal to 0, otherwise. More precisely, $INO_i = IN_i \times IO_i$, $IN_i = \prod_{j=1}^{m_1} IN_{ij}$, $IO_i = \prod_{j=1}^{m_2} IO_{ij}$ and

$$IN_{ij} = 1_{(N_{ij}=0,\max_l ZN_{ij,l}<0)} + \sum_{r=1}^{k_j-1} 1_{(N_{ij}=r,ZN_{ij,r}=\max_l ZN_{ij,l}>0)},$$

$$IO_{ij} = \sum_{r=1}^{c_j} 1_{(O_{ij}=r,\theta_{j,r-1}<ZO_{ij}\leq\theta_{j,r})}.$$

To do Gibbs sampling we need full conditionals. It is straightforward to show that the posterior of coefficient parameters are distributed as

$$\beta|\Theta,\Sigma,Z,Y,X \sim N_p(\tilde{\beta},\tilde{\Sigma}),$$

where $\tilde{\Sigma} = \left(\sum_{i=1}^{n} {X_i}^T\Sigma^{-1}X_i + \bar{\Sigma}^{-1}\right)^{-1}$ and $\tilde{\beta} = \tilde{\Sigma}\left(\sum_{i=1}^{n} {X_i}^T\Sigma^{-1}Z_i + \bar{\Sigma}^{-1}\bar{\beta}\right)$.

For the latent variables we have

$$Z_{ij}|\beta,\Theta,\Sigma,Y,X,\{Z_{ij'};j'\neq j\} \sim TN(a_{ij},s_j,B_{ij}), \quad j=1,...,q-m_3, \tag{2.3}$$

where $a_{ij} = \mu_{i,j} - s_j\Sigma^{-1}_{j,-j}(Z_{i,-j}-\mu_{i,-j})$, $s_j = 1/\Sigma^{-1}_{jj}$ and $\mu_{i,j}$ is the $j-$th element of $X_i\beta$ and $\mu_{i,-j}$ is the vector of $X_i\beta$ without $j-$th element and Σ^{-1}_{jj} is the (j,j)-th element of Σ^{-1} and $\Sigma^{-1}_{j,-j}$ is the j-th row of Σ^{-1} without j-th element. Here, $TN(\mu,\sigma^2,B)$ is a truncated normal distribution over the interval B. The B_{ij} interval is determined according to Y_{ij}. More specifically, for ordinal outcome $O_{ij} = r$, $B_{ij} = (\theta_{j,r-1}, \theta_{j,r}]$ and for nominal outcome $N_{ij} = 0$, $B_{ij,l} = (-\infty, 0]$ for $l = 1,...,k_j - 1$ and for $N_{ij} = r \neq 0$, $B_{ij,r} = (\max\{\max_{l \neq r}\{ZN_{ij,l}\}, 0\}, \infty)$ and $B_{ij,l} = (-\infty, ZN_{ij,r})$ for $l \neq r$.

For cut points, as pointed out by Cowles [14] there is an approach to simulate the cut points by partitioning the parameters such that $\{Z, \Theta\}$ are simulated jointly, that is $p(Z, \Theta|\beta, \Sigma, Y, X) \propto p(\Theta|\beta, \Sigma, Y, X)p(Z|\beta, \Theta, \Sigma, Y, X)$. As noted by Cowles [14], this strategy increases the speed of convergence of the algorithm. So we utilize its idea and generalize it for the multivariate case to make simulations according to the following posterior density:

$$p(\Theta|\beta, \Sigma, Y, X) \propto \prod_{i=1}^{n} p(ZO_i \in B_{O_i,\Theta})$$

where $B_{O_i,\Theta} = \prod_{j=1}^{m_2} (\theta_{j,O_{ij}-1}, \theta_{j,O_{ij}}]$ and $ZO_i \sim N_{m_2}(X_{i,O}\beta, \Sigma_O)$ and $X_{i,O}$ and Σ_O are the design matrix and the covariance matrix corresponding to the ordinal latent variables, respectively. Simulation is implemented using a Metropolis-Hastings step as below:

1. Initialize $\Theta^{(0)} = (\theta_{1,2}^{(0)}, \theta_{1,3}^{(0)}, ..., \theta_{m_2,c_{m_2}-1}^{(0)})$ according to the results of ordinal probit models and set $\sigma_{MH} = 0.05/\bar{c}$ where $\bar{c} = \sum_{j=1}^{m_2} c_j$, as a rule of thumb for candidate hyperparameter to get appropriate acceptance rates.

2. For $j = 1,...,m_2$ and $r = 2,...,c_j - 1$ generate

$$\theta_{r,j}^{*} \sim TN(\theta_{r,j}^{(t-1)}, \sigma_{MH}^2, (\theta_{r-1,j}^{*}, \theta_{r+1,j}^{(t-1)}]).$$

3. Set $\Theta^{(t)} = \Theta^{(t-1)}$ with probability of $1 - R$ and set $\Theta^{(t)} = \Theta^{*}$ with probability of R as

$$R = \min\left\{1, \prod_{i=1}^{n} \frac{p(ZO_i \in B_{O_i,\Theta^*})}{p(ZO_i \in B_{O_i,\Theta^{(t-1)}})} \times \prod_{j=1}^{m_2} \prod_{r=2}^{c_j-1} \frac{\Phi((\theta_{j,r+1}^{(t-1)} - \theta_{j,r}^{(t-1)})/\sigma_{MH}) - \Phi((\theta_{j,r-1}^{*} - \theta_{j,r}^{(t-1)})/\sigma_{MH})}{\Phi((\theta_{j,r+1}^{*} - \theta_{j,r}^{*})/\sigma_{MH}) - \Phi((\theta_{j,r-1}^{(t-1)} - \theta_{j,r}^{*})/\sigma_{MH})}\right\}.$$

4. Increase $t = t + 1$ and repeat the steps (2) and (3) until an adequate number of samples are obtained.

For an unconstrained covariance Box and Tiao [9] presented the Jeffreys prior as $p(\Sigma) \propto |\Sigma|^{-(q+1)/2}$ and the inverse Wishart distribution as a conjugate prior. However, for constrained covariances there is no standard distribution available as a prior. Different priors and simulation procedures proposed by McCulloch and Rossi [28], Chib and Greenberg [13], Barnard et al. [8], Imai and Van Dyk [19], and Burgette and Nordheim [10]. Barnard et al. [8] studied marginal uniform priors and jointly uniform priors for correlation matrices. In this paper we specify a jointly uniform prior for Σ constrained to be a positive definite matrix with some diagonal elements equal to one, i.e.

$$p(\Sigma) \propto 1_{(\Sigma \text{ is positive definite and diagonals of discrete variables are equal to one})}. \tag{2.4}$$

Obviously the posterior of Σ cannot be obtained conveniently. Hence, we use the PX-RPMH algorithm within Gibbs sampler to simulate covariance matrix Σ. Details are given in the next subsection.

2.3. *The PX-RPMH Algorithm for Simulating Covariance Matrix*

To make simulations from Σ, we use the method of parameter expansion which is implemented by Liu and Daniels [25] and Liu et al. [26]. The method includes four steps of (a) parameter expansion,

(b) transformation, (c) defining candidate prior and candidate posterior and (d) simulation according to Metropolis-Hastings. Details of these steps are given in the following.

(a) In this step we need to expand Σ into a less constrained matrix $\Sigma_\circ$ which is just a positive definite matrix, and to define a reduction operator as $(\beta,\Sigma) = \text{Red}(\beta,\Sigma_\circ) = (\beta, D_\circ^{-1}\Sigma_\circ D_\circ^{-1})$ where $D_\circ$ is the matrix of square root of diagonal elements of $\Sigma_\circ$ with replacement of diagonal elements corresponding to continuous outcomes with one.

(b) This step is to transform $\{Z_i,\Sigma\}$ into $\{Z_i^*,\Sigma_\circ\}$ according to the following one-to-one mapping:

$$\begin{cases} Z_i = X_i\beta + D_\circ^{-1}Z_i^* \\ \Sigma = D_\circ^{-1}\Sigma_\circ D_\circ^{-1} \end{cases} i = 1,\ldots,n, \tag{2.5}$$

where $\sum_{i=1}^{n} Z_{ij}^{*2} = 1$ for any $j = 1,\ldots,q$. These restrictions are needed to make the transformation one-to-one mapping. Note that draws of Z_i and β implicitly correspont to draws of Z_i^* and $D_\circ$, because

$$\sum_{i=1}^{n}(Z_{ij} - x'_{ij}\beta)^2 = (D_\circ^{-1})_{jj}^2 \sum_{i=1}^{n} Z_{ij}^{*2} = (D_\circ^{-1})_{jj}^2, \qquad j = 1,\ldots,q.$$

where x'_{ij} is the j-th row of X_i and $(D_\circ^{-1})_{jj}$ is the j-th diagonal element of $D_\circ^{-1}$.

(c) Here, we find the candidate prior given by $\pi(\Sigma) \propto |\Sigma|^{-\frac{q+1}{2}}$. It can be verified by using (2.5) that

$$\pi(\Sigma_\circ|Z^*,\beta) \propto |\Sigma_\circ|^{-\frac{n+q+1}{2}} etr\left\{-\frac{1}{2}S\Sigma_\circ^{-1}\right\}, \tag{2.6}$$

where $etr\{\cdot\} = \exp\{\text{trace}(\cdot)\}$, $S = \sum_{i=1}^{n} Z_i^* Z_i^{*T}$ and $Z_i^* = D_\circ(Z_i - X_i\beta)$, that is $\Sigma_\circ|Z^*,\beta$ is distributed as inverse Wishart with n degrees of freedom and the scale matrix of S.

(d) This step is to draw $\Sigma_\circ^*$ according to the distribution of $\Sigma_\circ|Z^*,\beta$ and translating it back to $\Sigma^* = D_\circ^{*-1}\Sigma_\circ^* D_\circ^{*-1}$ and accepting the candidate Σ^* using a Metopolis-Hastings step with acceptance rate

$$\alpha = \min\left\{1, \exp\left(\frac{q+1}{2}(\log|\Sigma^*| - \log|\Sigma^{(t)}|)\right)\right\}$$

at iteration $t+1$.
More details and proofs can be found in Liu and Daniels [25] and Liu et al. [26].

2.4. *Handling Missing Values*

When there are some missing values in Y, we partition it as $Y = (Y_{obs}, Y_{mis})$ where Y_{obs} is the observed part of Y and Y_{mis} is the missing part of Y. In this article we assume the mechanism of missingness to be at random. That is $p(M_{ij}|Y) = p(M_{ij}|Y_{obs})$ where M_{ij} is equal to one if Y_{ij} is missed and is equal to zero if Y_{ij} is observed. Therefore, we can use the Data Augmentation Modeling described by distribution given in (2.3) to make draws of missing latent variables according to the model described in (2.1) for Y_{obs}. So, all of the parametric models remain valid except that $Z_{ij}|\beta,\Theta,\Sigma,Y,X,\{Z_{ij'}; j' \neq j\}$ is no longer truncated. It should be noted that when there is no missing values for continuous variables we just make draws of the latent variables of categorical variables.

However, when there are some missing values in continuous variables one should treat them as latent variables, namely ZC_{ij}s, which correspond to C_{ij}s that have missing values.

2.5. *Initializations and Tunings*

At the first step of MCMC we need some initial values for parameters. In this study we set them randomly. However, we recommend the use of the output of an ordered logistic regression model for initial values of cut points, pairwise sample association coefficients for initial values of correlation matrix and the output of a linear regression model for initial values of coefficients.

3. Simulation Study

3.1. *Data Generation*

We use the following simulated example to illustrate the Bayesian model and the MCMC algorithm proposed in section 2 for analyzing mixed nominal, ordinal and continuous correlated outcomes. For $n = 300$ individuals we take a nominal outcome N with three levels, an ordinal outcome O with four levels and a continuous outcome Z. We also generated three covariates from $U(-0.5, 0.5)$ and set the first column of X_i to one, which means $p = 4$. So the latent variables model can be expressed as

$$Z_i = X_i\beta + \varepsilon_i \qquad i = 1, ..., n,$$

where $Z_i = (ZN_{i1}, ZN_{i2}, ZO_i, C_i)$ and $\varepsilon_i \sim N_4(0, \Sigma)$. We take $\beta = (1, -2, 0, 1)^T$ and $(\theta_2, \theta_3) = (0.5, 1)$. According to restrictions on the covariance matrix, we have

$$\Sigma = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & 1 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & 1 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix}.$$

Here, we take $\sigma_{44} = 2$ and $\sigma_{ij} = 0.5$ for $i \neq j$.

3.1.1. *Evaluation of Parameter Estimates*

Suppose that the parameter space is $\Xi = \{\beta, \Theta, \Sigma\}$ and $\xi \in \Xi$ is any parameter of the model. To make an estimation of ξ we iterated the MCMC algorithm for $n_{iter} = 75$ times with $n_{gibbs} = 2000$ Gibbs samples and $n_{burn} = 500$ burning samples and the thinning number $n_{thin} = 3$ to get $n_{mcmc} = 500$ samples of posterior densities denoted by $\{\xi_{i,j}\}$ for $i = 1, ..., n_{mcmc}$; $j = 1, ..., n_{iter}$. We estimated the parameter ξ and its standard error by $\hat{\mu}_\xi$ and $\hat{\sigma}_\xi$, respectively as $\hat{\mu}_\xi = \sum_{j=1}^{n_{iter}} \bar{\xi}_j / n_{iter}$; and $\hat{\sigma}_\xi = \sqrt{\sum_{j=1}^{n_{iter}} (\bar{\xi}_j - \bar{\bar{\xi}}_j)/(n_{iter} - 1)}$; where $\bar{\xi}_j = \sum_{i=1}^{n_{mcmc}} \xi_{i,j} / n_{mcmc}$ for $j = 1, ..., n_{iter}$.

The results are given in Table 1. In this table we see that MCMC sampler is successful to estimate $\beta_0, \beta_1, \beta_2, \beta_3$ and the cut points. However, there is an underestimation regarding the covariances by a factor of 10% for sample size $n = 300$. The standard error of parameters are relatively low, with respect to the magnitude of true values, that show a measure of goodness of convergence. The standard errors reduce as the sample size increases.

Table 1. Summary of 75 iterations of MCMC

Parameters	True Values	$n = 300$		$n = 1000$	
		Est.	S.E.	Est.	S.E.
β_0	1.0	1.007	0.072	0.997	0.030
β_1	-2.0	-1.998	0.145	-2.002	0.074
β_2	0.0	-0.023	0.110	-0.005	0.053
β_3	1.0	1.009	0.125	1.006	0.078
θ_2	0.5	0.503	0.070	0.498	0.034
θ_3	1.0	1.010	0.091	0.991	0.040
σ_{12}	0.5	0.458	0.112	0.483	0.059
σ_{13}	0.5	0.461	0.106	0.492	0.071
σ_{14}	0.5	0.464	0.166	0.496	0.092
σ_{23}	0.5	0.473	0.117	0.480	0.069
σ_{24}	0.5	0.472	0.163	0.499	0.082
σ_{34}	0.5	0.505	0.103	0.501	0.051
σ_{44}	2.0	2.074	0.175	2.020	0.085

3.2. *Prediction and Validation*

We used a 5-fold cross validation for a sample of size $n = 300$ that is we partitioned the sample into 5 subsamples and used four subsamples to estimate the parameters and one subsample to predict the observations jointly and to find the accuracy of predictions. We measured the accuracy of nominal and ordinal variables by the percentage of correct predictions and the accuracy of continuous variable by mean square prediction error (MSPE) defined as $MSPE = \frac{1}{m}\sum_{i=1}^{m}(\hat{W}_i - W_i)^2$, where $\hat{W}_i$ is the predicted value for observation W_i and $m = n/5$, the number of predictions. According to Table 2, on average 67% of nominal outcomes predicted exactly. This index for ordinal outcome is about 50% while this outcome has 4 levels and the correct predictions of a random guess would be only 25%. The positive correlations between predictions and the true values show that the predictions are in the same direction of the true values. The MSPE of continuous variable is near the true variance showing that the model does not inflate the variations. It should be noted that the accuracy of the model highly depends on the variations of predictors and the variations of errors, that is the accuracy of cross validation will increase when variation of predictors increase and it will decrease as the variations of errors increase.

Table 2. Cross validation summary

Variable	Accuracy (%)	Correlation	MSPE
Nominal	67.41	–	–
Ordinal	49.72	0.48	–
Continuous	–	0.41	2.03

3.3. *The Effect of Missing at Random Assumption*

To make an evaluation of the effect of missingness we generated a sample of size $n = 300$, and then removed some of the outcomes as missing values according to three different scenarios:

Table 3. Summary of 75 iterations of MCMC with MAR values

Parameters	True Values	M1		M2		M3	
		Est.	S.E.	Est.	S.E.	Est.	S.E
β_0	1.0	1.002	0.062	0.997	0.073	0.990	0.060
β_1	-2.0	-2.032	0.151	-2.008	0.175	-2.044	0.150
β_2	0.0	-0.018	0.121	0.017	0.142	-0.024	0.152
β_3	1.0	1.011	0.149	1.041	0.131	1.055	0.156
θ_2	0.5	0.506	0.066	0.490	0.068	0.488	0.070
θ_3	1.0	1.009	0.089	0.999	0.082	0.979	0.091
σ_{12}	0.5	0.429	0.118	0.451	0.121	0.423	0.144
σ_{13}	0.5	0.462	0.106	0.452	0.173	0.486	0.170
σ_{14}	0.5	0.465	0.165	0.464	0.145	0.470	0.160
σ_{23}	0.5	0.463	0.107	0.444	0.181	0.480	0.160
σ_{24}	0.5	0.451	0.154	0.469	0.138	0.460	0.142
σ_{34}	0.5	0.470	0.106	0.506	0.103	0.482	0.086
σ_{44}	2.0	2.065	0.199	2.059	0.159	2.072	0.143

M1. Missing values only correspond to the continuous variable with

$$p(MC_{ij} = 1|O_{ij} = o_{ij}) = \begin{cases} 0.4 & o_{ij} = 4 \\ 0.1 & o_{ij} \neq 4 \end{cases}$$

where MC_{ij} indicates missing status of C_{ij}.

M2. Missing values only correspond to the ordinal variable with

$$p(MO_{ij} = 1|N_{ij} = n_{ij}) = \begin{cases} 0.5 & n_{ij} = 0 \\ 0.1 & n_{ij} \neq 0 \end{cases}$$

where MO_{ij} indicates missing status of O_{ij}.

M3. Missing values only correspond to the nominal variable with

$$p(MN_{ij} = 1|C_{ij} = c_{ij}) = \begin{cases} 0.1 & c_{ij} \leq 0 \\ 0.3 & c_{ij} > 0 \end{cases}$$

where MN_{ij} indicates missing status of N_{ij}.

In all three scenarios we have one variable with missing values and other variables are observed for all individuals. In each scenario, the probabilities of missingness are chosen such that the rate of missingness is about 25%. The estimated parameters and standard errors reported in Table 3 show that the algorithm is successful even if there are MAR values in the dataset.

3.4. *Sensitivity to Non-Ignorable Missing Values*

To make an evaluation of the effect of missing not at random mechanism we generated a sample of size $n = 300$, and then replaced some of the responses of continuous variable with missing values according to the following two different scenarios:

M4. Missing values correspond to the continuous variable with

$$p(MC_{ij} = 1|C_{ij} = c_{ij}) = \begin{cases} 0.4 & c_{ij} \leq 0 \\ 0.1 & c_{ij} > 0 \end{cases}$$

Table 4. Summary of 75 iterations of MCMC for different patterns of missing values

Parameters	True Values	No Missing		M1 (MAR)		M4 (NMAR)		M5 (NMAR)	
		Est.	S.E.	Est.	S.E.	Est.	S.E	Est.	S.E.
β_0	1.0	1.007	0.072	1.007	0.068	0.928	0.065	1.247	0.076
β_1	-2.0	-1.998	0.145	-2.036	0.168	-2.061	0.178	-1.839	0.285
β_2	0.0	-0.023	0.110	-0.020	0.128	-0.009	0.118	0.001	0.121
β_3	1.0	1.009	0.125	1.008	0.147	1.026	0.143	0.906	0.194
θ_2	0.5	0.503	0.070	0.502	0.065	0.457	0.066	0.612	0.082
θ_3	1.0	1.010	0.091	1.008	0.085	0.936	0.079	1.151	0.101
σ_{12}	0.5	0.458	0.112	0.433	0.117	0.352	0.142	0.478	0.292
σ_{13}	0.5	0.461	0.106	0.455	0.107	0.431	0.117	0.433	0.301
σ_{14}	0.5	0.464	0.166	0.459	0.167	0.457	0.193	0.217	0.202
σ_{23}	0.5	0.473	0.117	0.462	0.109	0.429	0.131	0.423	0.314
σ_{24}	0.5	0.472	0.163	0.444	0.154	0.481	0.169	0.212	0.197
σ_{34}	0.5	0.505	0.103	0.470	0.107	0.508	0.114	0.254	0.110
σ_{44}	2.0	2.074	0.175	2.053	0.199	2.195	0.197	1.504	2.664

M5. Missing values correspond to the continuous variable with

$$p(MC_{ij}=1|C_{ij}=c_{ij}) = \begin{cases} 1 & c_{ij} \le 0 \\ 0 & c_{ij} > 0 \end{cases}$$

The estimated parameters in Table 4 show that the true values of parameters are within two standard deviation intervals. However, the biases of parameters raise in non-ignorable missing patterns especially for those of the covariance parameters which are directly connected to variables containing missing values. The analysis of standard errors shows that there is an increase in error of estimations, especially for the covariance parameters, when the probability of non-ignorable missingness increases.

4. Application

As an application of our model we used the Foreign Language Attitude Scale (FLAS) data which was analyzed by Schafer [34] using GLOM. It contains $n = 279$ samples of students who enrolled in foreign language courses at the Pennsylvania State university. We investigated the association between the kind of foreign language studied (LAN), Modern Language Aptitude Test (MLAT), and the age of students (AGE) with the FLAS variable as a predictor. Table 5 presents some descriptive statistics of the dataset.

We start by a saturated model with separate coefficients for each latent variable. Due to different sample sizes for levels of the nominal outcome we used a mildly informative prior regarding to coefficients of nominal latent variables as $(\beta_{0,ZN_1}, \beta_{1,ZN_1}, ..., \beta_{1,ZN_3}) \sim N(0,I)$ and non-informative priors for other coefficients. The results of the MCMC algorithm are presented in Table 6. The outputs of the saturated model show that the FLAS score has no effect on the latent variables of Spanish and German languages. However, there is a negative effect on the latent variable of Russian language. It means that those who get more FLAS scores are less likely to choose Russian language. Obviously, the posterior distributions of the latent variables of Spanish and German languages are very similar. Also the negative coefficient corresponding to ordinal latent variable means that those who get more FLAS score are less likely to be older. The estimated coefficients for continuous outcomes are similar to the standard regression model $MLAT_i = 18 + 0.08FLAS_i$. Because there are

Table 5. Summary of variables in foreign language attitude scale analysis

Variable	Description	Categories	Percent	Missing	Latents
LAN	Foreign language studied	French	24.0 %	0.0 %	–
		Spanish	28.0 %		ZN_1
		German	40.9 %		ZN_2
		Russian	7.1 %		ZN_3
AGE	Age group	< 20	46.3 %	3.9 %	ZO
		20-21	43.3 %		
		≥ 22	10.4 %		
MLAT	Modern Language Aptitude Test			17.6 %	ZC
FLAS	Foreign Language Attitude Scale			0.0 %	–

many non-significant coefficients in the saturated model, we consider a reduced model by assuming $\beta_{1,ZN_1} = \beta_{1,ZN_2} = 0$. The results of the reduced model show a substantial reduction in standard error of parameters. For comparison of the full model and the reduced model we use the DIC index (Spiegelhalter et al. [36]). The difference between DICs is quite large showing that the fit of the reduced model is better than that of the full model. The positive correlation between latent variables of Spanish and German languages ($\sigma_{ZN_1,ZN_2} = 0.811$) shows that the utilities of these choices are in the same direction. The negative correlations between the latent variables of German and Russian languages with AGE group show that younger students are expected to choose German and Russian courses. The positive covariance of $\sigma_{ZN_3,ZC}$ and the negative covariance of $\sigma_{ZO,ZC}$ suggest that the students who choose Russian language and younger students are expected to get more MLAT scores.

5. Conclusions

We have developed a joint model for correlated mixed outcomes using MCMC approach. The results of simulations show that our method is successful in estimation of parameters of the model. A cross validation analysis of simulated data with MAR values shows that the method is reliable regarding parameter estimation and imputations. Further investigations about NMAR mechanism show that our method is not successful about parameter estimation and this type of missingness will raise standard errors slightly. As an application of our method we studied the foreign language aptitude scale dataset. Also, the thetracoric correlations for this dataset were investigated.

It should be noted that the MVP model is sensitive to zero counts in the cross tabulation of nominal outcomes and non-informative priors for coefficients may result in divegence of MCMC algorithm. So, we recommend to put informative priors for those parameters.

Our extensive studies suggest that our Bayesian methodology can be used effectively in the studies where joint analysis of a combination of correlated nominal, ordinal and continuous outcomes is of the main interests. It is obvious that our method can be used for longitudinal studies with arbitrary correlation matrices.

Published by Atlantis Press

Table 6. Estimated parameters of the latent variables for FLAS dataset

Parameters	Saturated Model		Reduced Model	
	Estimates	S.E.	Estimates	S.E.
β_{0,ZN_1}	0.181	0.355	0.137*	0.058
β_{1,ZN_1}	0.001	0.006	–	–
β_{0,ZN_2}	0.019	0.378	0.213*	0.059
β_{1,ZN_2}	0.003	0.004	–	–
β_{0,ZN_3}	-1.278*	0.510	-0.390*	0.073
β_{1,ZN_3}	-0.004	0.005	-0.013*	0.001
$\beta_{0,ZO}$	1.339*	0.410	0.999*	0.407
$\beta_{1,ZO}$	-0.015*	0.006	-0.010*	0.005
$\beta_{0,ZC}$	16.643*	2.409	17.220*	2.293
$\beta_{1,ZC}$	0.092*	0.028	0.086*	0.027
θ_2	1.507*	0.144	1.431*	0.093
σ_{ZN_1,ZN_2}	0.811*	0.302	0.822*	0.160
σ_{ZN_1,ZN_3}	-0.698*	0.224	-0.496	0.246
$\sigma_{ZN_1,ZO}$	-0.077	0.218	0.050	0.112
$\sigma_{ZN_1,ZC}$	0.459	0.781	-0.195	0.511
σ_{ZN_2,ZN_3}	-0.754*	0.154	0.007	0.071
$\sigma_{ZN_2,ZO}$	-0.167	0.150	-0.113*	0.076
$\sigma_{ZN_2,ZC}$	1.131*	0.400	0.427	0.408
$\sigma_{ZN_3,ZO}$	-0.221	0.116	-0.295*	0.069
$\sigma_{ZN_3,ZC}$	-0.781	0.642	1.117*	0.439
$\sigma_{ZO,ZC}$	-1.155*	0.478	-1.189*	0.512
σ^2_{ZC}	41.771*	9.991	40.137*	3.543
DIC	2979.33		2904.89	

*: Significant at the 5% level

References

[1] Aitchison, J. and Silvey, S. D., The generalization of probit analysis to the case of multiple responses. *Biometrika*, **44**(1-2)(1957) 131-140.

[2] Albert, J. H. and Chib, S., Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, **88**(422)(1993), 669-679.

[3] Ashford, J. R. and Sowden, R. R., Multi-variate probit analysis. *Biometrics*(1970) 535-546.

[4] Bahrami Samani, E., Ganjali, M. and Khodaddadi, A., A latent variable model for mixed continuous and ordinal responses. *Journal of Statistical Theory and Applications*, **7**(3)(2008) 337-349.

[5] Bahrami Samani, E. and Ganjali, M., A Multivariate Latent Variable Model for Mixed Data from Continuous and Ordinal Responses with Possibility of Missing Responses. *Appl. Appl. Math.: Int. J*, **5**(10)(2010) 1564-1584.

[6] Bahrami Samani, E., Ganjali, M. and Eftekhari, S., A latent variable model for mixed continuous and ordinal responses with nonignorable missing responses: Assessing the local influence via covariance structure. *Sankhya B*, **72**(1)(2010) 38-57.

[7] Bahrami Samani, E. and Tahmasebinejad, Z., Joint Modelling of Mixed Correlated Nominal, Ordinal and Continuous Responses. *Journal of Statistical Research*, **45**(1)(2011) 37-47.

[8] Barnard, J., McCulloch, R. and Meng, X. L., Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, **10**(4)(2000) 1281-1312.

[9] Box, G. E. and Tiao, G. C., *Bayesian inference in statistical analysis.* (Addison-Wesely publishing company,1973).

[10] Burgette, L. F. and Nordheim, E. V., The trace restriction: An alternative identification strategy for the Bayesian multinomial probit model. *Journal of Business & Economic Statistics*, **30**(3)(2012) 404-410.

[11] Catalano, P. J. and Ryan, L. M., Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, **87**(419)(1992) 651-658.

[12] Chen, M. H. and Dey, D. K., Bayesian analysis for correlated ordinal data models. *Biostatistics Basel*, **5**(2000) 133-158.

[13] Chib, S. and Greenberg, E., Analysis of multivariate probit models. *Biometrika*, **85**(2)(1998) 347-361.

[14] Cowles, M. K., Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, **6**(2)(1996) 101-111.

[15] Cox, D. R. and Wermuth, N., Response models for mixed binary and quantitative variables. *Biometrika*, **79**(3)(1992) 441-461.

[16] Diggle, P., Heagerty, P., Liang, K. Y. and Zeger, S., *Analysis of longitudinal data.* (Oxford University Press, New York, 2002).

[17] Fitzmaurice, G. M. and Laird, N. M., Regression models for mixed discrete and continuous responses with potentially missing values. *Biometrics*(1997) 110-122.

[18] Heckman, J. J., Dummy endogenous variables in a simultaneous equation system. *Econometrica*, **46**(4)(1978) 931-959.

[19] Imai, K. and Van Dyk, D. A., A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of econometrics*, **124**(2)(2005) 311-334.

[20] Lawrence, E., Bingham, D., Liu, C. and Nair, V. N., Bayesian inference for multivariate ordinal data using parameter expansion. *Technometrics*, **50**(2)(2008) 182-191.

[21] Leon, D., Alexander, R. and Carrigre, K. C., General mixed data model: Extension of general location and grouped continuous models. *Canadian Journal of Statistics*, **35**(4)(2007) 533-548.

[22] Li, Y. and Schafer, D. W., Likelihood analysis of the multivariate ordinal probit regression model for repeated ordinal responses. *Computational Statistics & Data Analysis*, **52**(7)(2008) 3474-3492.

[23] Little, R. J. and Rubin, D. B., *Statistical analysis with missing data.* (John Wiley and Sons, New York, 2002).

[24] Liu, C. and Rubin, D. B., Ellipsoidally symmetric extensions of the general location model for mixed categorical and continuous data. *Biometrika*, **85**(3)(1998) 673-688.

[25] Liu, X. and Daniels, M. J., A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *Journal of Computational and Graphical Statistics*, **15**(4)(2006) 897-914.

[26] Liu, X., Daniels, M. J. and Marcus, B., Joint Models for the Association of Longitudinal Binary and Continuous Processes With Application to a Smoking Cessation Trial, *Journal of the American Statistical Association*, **104**(2009) 429-438.

[27] McCulloch, R. E., Polson, N. G. and Rossi, P. E., A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, **99**(1)(2000) 173-193.

[28] McCulloch, R. and Rossi, P., An exact likelihood approach to analysis of the MNP model. *Journal of Econometrics*, **64**(1-2)(1994) 207-240.

[29] McFadden, D., Conditional logit analysis of qualitative choice behavior. In Frontiers of Econometrics, ed. by P. Zarembka. *Academic Press*(1973) 105-142.

[30] Nandram, B. and Chen, M. H., Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation*, **54**(1-3)(1996) 129-144.

[31] Olkin, I. and Tate, R. F., Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics*(1961) 448-465.

[32] Peng, Y., Little, R. J. and Raghunathan, T. E., An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics*, **60**(3)(2004) 598-607.

[33] Poon, W. Y. and Lee, S. Y., Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika*, **52**(3)(1987) 409-430.

[34] Schafer, J. L., *Analysis of incomplete multivariate data.* (CRC press, London, 1997).

[35] Snell, E. J., A scaling procedure for ordered categorical data. *Biometrics*, **20**(1964) 592-607.

[36] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A., Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4)(2002) 583-639.

Published by Atlantis Press

[37] Tanner, M. A. and Wong, W. H., The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, **82**(398)(1987) 528-540.
[38] Teimourian, M., Baghfalaki, T., Ganjali, M. and Berridge, D., Joint modeling of mixed skewed continuous and ordinal longitudinal responses: a Bayesian approach. *Journal of Applied Statistics*, **42**(10)(2015) 1-24.
[39] Van Dyk, D. A. and Meng, X. L., The art of data augmentation. *Journal of Computational and Graphical Statistics*, **10**(1)(2001) 1-50.
[40] Xu, H. and Craig, B. A., A probit latent class model with general correlation structures for evaluating accuracy of diagnostic tests. *Biometrics*, **65**(4)(2009) 1145-1155.
[41] Zhang, X., Boscardin, W. J., Belin, T. R., Wan, X., He, Y. and Zhang, K., A Bayesian method for analyzing combinations of continuous, ordinal, and nominal categorical data with missing values. *Journal of Multivariate Analysis*, **135**(2015) 43-58.