American Journal of Theoretical and Applied Statistics
2023; 12(1): 1-12
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20231201.11
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Bayesian Analysis of Multivariate Longitudinal Ordinal Data Using Multiple Multivariate Probit Models

Xiao Zhang

Department of Mathematical Sciences, Michigan Technological University, Houghton, USA

Email address:
xzhang35@mtu.edu

To cite this article:
Xiao Zhang. Bayesian Analysis of Multivariate Longitudinal Ordinal Data Using Multiple Multivariate Probit Models. *American Journal of Theoretical and Applied Statistics*. Vol. 12, No. 1, 2023, pp. 1-12. doi: 10.11648/j.ajtas.20231201.11

Received: March 1, 2023; **Accepted**: March 17, 2023; **Published**: March 28, 2023

Abstract: Multivariate longitudinal ordinal data are often involved in longitudinal studies with each individual having more than one longitudinal ordinal measure. However, due to complicated correlation structures within each individual and no explicit likelihood functions, analyzing multivariate longitudinal ordinal data is quite challenging. In this paper, Markov chain Monte Carlo (MCMC) sampling methods are developed to analyze multivariate longitudinal ordinal data by extending multivariate probit (MVP) models for univariate longitudinal ordinal data to multiple multivariate probit models (MMVP) for multivariate longitudinal ordinal data. The identifiable MVP models require the covariance matrix of the latent multivariate normal variables underlying the longitudinal ordinal variables to be a correlation matrix, thus a Metropolis-Hastings (MH) algorithm is usually necessitated, which brings a rigorous task to develop efficient MCMC sampling methods. In contrast to the identifiable MVP models, the non-identifiable MVP models can be constructed to circumvent a MH algorithm to sample a correlation matrix by a Gibbs sampling to sample a covariance matrix, and hence improve the mixing and convergence of the MCMC components. Therefore, both the identifiable MMVP models and the non-identifiable MMVP models for multivariate longitudinal ordinal data are presented, and their corresponding MCMC sampling methods are developed. The performances of these methods are illustrated through simulation studies and an application using data from the Russia Longitudinal Monitoring Survey-Higher School of Economics (RLMS-HSE).

Keywords: Multivariate Longitudinal Ordinal Data, MCMC, Multivariate Probit Model, Multiple Multivariate Probit Model, Identification

1. Introduction

Longitudinal ordinal data appear in many scientific research fields, such as medical research and health related surveys. However, these kinds of data usually involve more than one longitudinal ordinal measure collected from the same individual. For instance, in depression research studies, each individual may be asked to fill in the depression score, an ordinal variable indicating the levels of depression during follow-up visits; the sleeping disorder score may often be collected, also an ordinal variable measuring the extent of the quality of sleeping. Since those measures are collected from the same individual, it is desirable to analyze them jointly instead of separately.

Generalized linear mixed-effects models have been widely utilized to analyze multivariate longitudinal data. Gibbons and Hedeker [8]) proposed three-level mixed-effects model for binary clustered data; Liu and Hedeker [17] extended the work of Gibbons and Hedeker [8] to longitudinal multivariate ordinal data. Cagnone et al [4] modeled multivariate longitudinal ordinal data using latent variable models by incorporating random effects to account for correlated structures. Grigorova et al [9] proposed an EM algorithm for multiple ordinal outcomes using a probit model with random effects. Laffont et al [15] proposed a multivariate probit mixed effects model for multivariate longitudinal ordinal data. Grigorova [10] proposed a random effect model for two longitudinal ordinal outcomes using EM algorithm for maximum likelihood estimation. Tran et al [23] proposed a latent linear mixed model with serial correlation structures for multivariate longitudinal ordinal data.

Multivariate marginal models have also been popularly explored for multivariate longitudinal or clustered ordinal data. Due to the lack of explicit likelihood functions, generalized estimating equations (GEE) methods are at an advantage as tools to analyze multivariate non-Gaussian data including multivariate longitudinal or clustered ordinal data. Related works include Qu et al [20], Heagerty and Zeger [12], Jiang et al [13] and Spiess [22]. Formulating the estimation equation using quasi-likelihood, Cho [6] proposed a multivariate marginal model to analyze multivariate longitudinal data. Complete reviews regarding multivariate longitudinal data can be referred to Bandyopadhyay and Ganguli [1] and Verbeke et al [24].

The MCMC methods have become general computation tools in Bayesian inference and have been generally investigated for various multivariate data. Univariate longitudinal binary data has been investigated using the MVP models [5, 18, 25, 26]. Univariate longitudinal ordinal data has been considered by Lawrence et al [16] and Zhang [27] using the non-identifiable MVP models. Dunson [7] proposed Bayesian latent variables with random effects for clustered mixed data including multivariate longitudinal ordinal data. However, multivariate longitudinal ordinal data have been seldomly inspected using marginal models. Comparing with the random-effects models, the marginal models can give direct estimation for correlations among multivariate measures. This paper is trying to fill in this gap by developing MCMC methods to analyze multivariate longitudinal ordinal data using the MMVP models. The reminder of this paper is organized as follows. Section 2 contains the identifiable MMVP model and the corresponding MCMC method for multivariate longitudinal ordinal data. The non-identifiable MMVP model is constructed in Section 3, consisting of the two proposed MCMC methods. Simulation studies regarding these proposed methods are conducted in Section 4, followed by Section 5 - an application to the RLMS-HSE data. A brief discussion is presented in Section 6.

2. Identifiable MMVP Model and Metropolis-Hastings-Within-Gibbs Sampler

2.1. Identifiable MMVP Model

The univariate probit model assumes that there is an underlying univariate normal variable corresponding to an ordinal variable. Specifically, suppose for each individual i, $i = 1, \cdots, n$, there is an ordinal outcome Y_i with J categories and a $p \times 1$ covariate vector X_i. Then the probit model assumes that there is a latent variable Z_i underlying Y_i, following a normal distribution with mean $X_i\beta$ and variance being σ^2, denoted by $N(X_i\beta, \sigma^2)$, where β is the $p \times 1$ regression parameter vector. The model further assumes that

$$Y_i = l \Leftrightarrow \gamma_{l-1} < Z_i \leq \gamma_l \text{ for } l = 1, \cdots, J,$$

i.e., $P(Y_j \leq l) = \Phi\left(\frac{\gamma_l - X_i^T\beta}{\sigma}\right)$, where $\Phi(\cdot)$ is the standard normal distribution function and $\gamma = (\gamma_0, \gamma_1, \cdots \gamma_J)$ being the unknown cut-points. The identifiable model is usually defined by $\sigma^2 = 1$ and $\gamma_0 = -\infty$, $\gamma_1 = 0$, and $\gamma_J = \infty$.

The multivariate probit (MVP) model assumes for each individual i, $Y_i = (Y_{i1}, \cdots, Y_{ik})^T$ is a k-dimensional ordinal outcome with each Y_{ij} has J_j ordinal categories for $i = 1, \cdots, n$ and $j = 1, \cdots, k$; there is an underlying multivariate normal variable $Z_i = (Z_{i1}, \cdots, Z_{ik})^T$ corresponding to $Y_i = (Y_{i1}, \cdots, Y_{ik})^T$, such that $Y_{ij} = l \Leftrightarrow \gamma_{j,l-1} < Z_{ij} \leq \gamma_{j,l}$ for $l = 1, \cdots, J_j$, $\gamma_{j,0} = -\infty$, $\gamma_{j,1} = 0$, and $\gamma_{j,J_j} = \infty$. It can be seen that the univariate probit model holds for each pair of Y_{ij} and Z_{ij}, which implies the variance of each Z_{ij} is fixed at 1 for the model to be identifiable. With $X_i = (X_{i1}, \cdots, X_{ik})^T$ being the covariate matrix for individual i, the model further assumes that $Z_i = (Z_{i1}, \cdots, Z_{ik})^T$ following a multivariate normal distribution with the mean vector being $X_i\beta$ and the covariance matrix R, which in fact is a correlation matrix, i.e., $Z_i \sim N_k(X_i\beta, R)$.

Extending the MVP model for one longitudinal ordinal variable to multiple multivariate probit (MMVP) model for multivariate longitudinal ordinal variables can be defined as follows. Suppose that there are N individuals with K longitudinal ordinal response variables, each of which is collected T_k time points. Specifically, denote $Y = (Y_1, \cdots, Y_N)^T$ with $Y_i = (Y_{i\cdot 1}, \cdots, Y_{i\cdot K})^T$ and $Y_{i\cdot k} = (Y_{i1k}, \cdots, Y_{iT_kk})^T$ for $i = 1, \cdots, N$, and $k = 1, \cdots, K$. As can be seen, for each individual i, the number of variables for the response vector Y_i is $\sum_{k=1}^K T_k$, and thus Y is a vector with $N \times \sum_{k=1}^K T_k$ elements. If each longitudinal response variable, $Y_{i\cdot k}$, is assumed to has the same number of time points, i.e., $T_1 = T_2 = \cdots = T_K$, denoted by T, then the number of elements for the response vector Y_i is $N \times T \times K$.

Denote the covariate matrix of individual i for the kth longitudinal ordinal variable $Y_{i\cdot k}$ as $X_{i\cdot k} = \left(X_{i1k}, \cdots, X_{iT_kk}\right)^T_{T_k \times p_k}$, and the regression parameter vector $\beta_k = (\beta_{k1}, \cdots, \beta_{kp_k})^T$. Then the covariate matrix for individual i can be written as $X_i = \text{block-diagonal}\left(\mathrm{I}_{T_1 \times T_1}, \cdots, \mathrm{I}_{T_K \times T_K}\right) \times (X_{i\cdot 1}, \cdots, X_{i\cdot K})^T$ with $\mathrm{I}_{T_k \times T_k}$ being the identity matrix for $k = 1, \cdots, K$, and the regression parameter vector β as $\beta = (\beta_1^T, \cdots, \beta_K^T)^T$ with $\beta_k = (\beta_{k1}, \cdots, \beta_{kp_k})^T$. Assuming that $X_{i\cdot 1} = \cdots = X_{i\cdot K} = X_i$, i.e., each longitudinal ordinal variable $Y_{i\cdot k}$ has the same covariate matrix, then $X_i = \mathrm{I}_K \otimes X_i$, where I_K is a $K \times K$ identity matrix and $\otimes$ is the left Kronecker product.

The MVP model is assumed for each longitudinal ordinal response variable $Y_{i\cdot k} = (Y_{i1k}, \cdots, Y_{iT_kk})^T$, there is a latent multivariate normal variable $Z_{i\cdot k} = (Z_{i1k}, \cdots, Z_{iT_kk})^T$ follows a multivariate normal distribution with mean vector being $X_{i\cdot k}\beta_k$ and covariance matrix R_k, which is a correlation matrix for the identifiable model. Assume that Y_{itk} has c_{tk} categories, then the cut-points for the latent variable Z_{itk} can be defined as follows: $Y_{itk} = c \Leftrightarrow \gamma_{t,k,c-1} < Z_{itk} \leq \gamma_{t,k,c}$, for $c = 1, \cdots, c_{tk}$, $\gamma_{t,k,0} = -\infty$, $\gamma_{t,k,1} = 0$, and $\gamma_{t,k,c_{tk}} = \infty$; $\gamma = (\gamma_1, \cdots, \gamma_K)^T$ with $\gamma_k = (\gamma_{1k}, \cdots, \gamma_{T_kk})$ and $\gamma_{tk} = (\gamma_{t,k,1}, \cdots, \gamma_{t,k,c_{tk}})$ for $k = 1, \cdots, K$ and $t = 1, \cdots, T_k$. Denote $Z = (Z_1, \cdots, Z_N)^T$ and $Z_i = (Z_{i\cdot 1}, \cdots, Z_{i\cdot K})^T$ for $i = 1, \cdots, N$. It can be shown that $Z_i \sim N_{\sum_{k=1}^K T_k}(X_i\beta, R)$, and R is a correlation matrix with the block diagonal matrices being $R_1, \cdots, R_K$.

2.2. Metropolis-Hastings-Within-Gibbs Sampler (MH-GS)

The priors for β, γ and R are assumed to be independent, i.e., $P(\beta,\gamma,R) = P(\beta)\times P(\gamma)\times P(R)$. Then the posterior joint density of $\beta,\gamma,R,$ and Z given the observed multivariate longitudinal ordinal outcome $Y=(Y_1,\cdots,Y_N)^T$ can be derived in the following:

$$P(\beta,\gamma,R,Z\,\big|\,Y) \propto P(\beta)\times P(\gamma)\times P(R)\times P(Z|\beta,\gamma,R,Y)$$

$$\propto P(\beta)\times P(\gamma)\times P(R)\times \prod_{i=1}^{N}[\mathrm{I}_i\times\phi(Z_i;X_i\beta,R)]$$

where $\phi(\cdot)$ is the standard normal density function, and $\mathrm{I}_i=\prod_{k=1}^{K}\mathrm{I}_{ik}$, where $\mathrm{I}_{ik}=\sum_{t=1}^{T_k}1_{(Y_{itk}=l)}\,1_{(\gamma_{t,k,l-1}<Z_{itk}\le\gamma_{t,k,l})}$, indicating compatibility of the latent variable Z_i with the ordinal variable Y_i, $i=1,\cdots,N$. To implement the MCMC sampling, each full conditional distribution is derived as follows:

$\beta|\,\gamma,R,Z,Y\sim N_{\sum_{k=1}^{K}p_k}(\hat{\beta},V_\beta)$ where $V_\beta=(\sum_{i=1}^{N}X_i^TR^{-1}X_i+C^{-1})^{-1}$ and $\hat{\beta}=V_\beta(\sum_{i=1}^{N}X_i^TR^{-1}Z_i+C^{-1}b)$, assuming the prior of β follows $N_{\sum_{k=1}^{K}p_k}(b,\mathrm{C})$ with the mean vector equal to b and the covariance matrix equal to C.

$Z_{itk}|\beta_k,\gamma_k,R,Y_{itk},Z_{ijk},j\neq t$ is the interval truncated normal distribution constrained to lie between two cut-points $\gamma_{t,k,c-1}$ and $\gamma_{t,k,c}$, assuming $Y_{itk}=c$.

$\gamma_{t,k,c}|\,\beta,R,Z,Y,\gamma_{t,k,l}\ c\neq l$ is a uniform distribution, $U(\gamma_{t,k,c}|\max\{\max[Z_{itk}\!:Y_{itk}=c],\gamma_{t,k,c-1}\},\min\{\min[Z_{itk}\!:Y_{itk}=c+1],\gamma_{t,k,c+1}\})$, assuming a non-informative prior for $\gamma_{t,k,c}$, i.e., $P(\gamma_{t,k,c})\propto 1$.

The full conditional density function of R can be derived as $P(R|\beta,\gamma,Z,Y)\propto P(R)\times\prod_{i=1}^{N}\phi(Z_i;X_i\beta,R)$, which does not belong to any standard density function due to R being a correlation matrix instead of a covariance matrix. Zhang et al [26] proposed a MH algorithm to sample R for univariate longitudinal binary data and used Wishart distributions to derive the prior and proposal distributions for R. For MMVP model, the proposed method for sampling R is extended to allow more flexible priors and proposal distributions including those based on Wishart distributions.

Several priors for a correlation matrix have been employed, such as the Jeffreys' prior ($R_{p\times p}\propto|R|^{-\frac{p+1}{2}}$) [3], truncated multivariate normal prior [5], marginally uniform prior ($P(r_{ij})\propto 1$ [2], and joint uniform prior ($R\propto 1$)[2, 18]. It can be shown that if $\Sigma\left(=D^{\frac{1}{2}}RD^{\frac{1}{2}}\right)\sim\mathrm{Wishart}_p(m,V)$ with m degrees of freedom and V being a diagonal matrix, R and D (a diagonal matrix with diagonal elements being the variance parameters) are independent and $P(R_{p\times p})\propto|R|^{\frac{m-p-1}{2}}$; if V is a non-diagonal matrix, R and D are not independent. If $\Sigma\left(=D^{\frac{1}{2}}RD^{\frac{1}{2}}\right)\sim\text{Inverse-Wishart}_p(m,V)$, R and D are not independent for any scale matrix V being diagonal or non-diagonal matrices.

Although there is not an explicit formula for the marginal density of R derived from a Wishart distribution with a non-diagonal scale matrix and an inverse-Wishart distribution, the joint density of R and D can be derive as $P(R,D)=P(\Sigma)\times|J_{\Sigma\to R,D}|$ with Jacobian matrix $|J_{\Sigma\to R,D}|=|D|^{\frac{p-1}{2}}$ and $P(\Sigma)$ being the density function of Wishart or inverse-Wishart distribution. The weak point to use $P(R,D)$ as the prior of R is that it includes redundant D with diagonal elements serving as variance parameters; however, the advantage is that it allows more flexible prior specification for R through the specification of the scale matrix V other than a diagonal matrix. If $P(R,D)$ is used for the prior of R, then the full conditional density function of R and D is $P(R,D|\beta,\gamma,Z,Y)\propto P(R,D)\times\prod_{i=1}^{N}\phi(Z_i;X_i\beta,R)$. As can be seen, D is involved only through the prior of $P(R,D)$, the model itself does not change and still remains identifiable. The MH algorithm is proposed to sample R and D as follows:

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

Generate (R^*,D^*) by generating $\Sigma^*=D^{*\frac{1}{2}}R^*D^{*\frac{1}{2}}$ from $\mathrm{Wishart}_{\sum_{k=1}^{K}T_k}\left(m_p,\Sigma^{(t)}/m_p\right)$ or Inverse-$\mathrm{Wishart}_{\sum_{k=1}^{K}T_k}\left(m_p,m_p\Sigma^{(t)}\right)$ with m_p degrees of freedom.

Take

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

where $\alpha=\min\left\{\frac{P(R^*,D^*|\beta,\gamma,Z,Y)}{P(R^{(t)},D^{(t)}|\beta,\gamma,Z,Y)}\frac{f(\Sigma^{(t)}|\Sigma^*)}{f(\Sigma^*|\Sigma^{(t)})},1\right\}$, and the proposal density $f(\Sigma^*|\Sigma^{(t)})$ is equal to $|J_{\Sigma^*\to R^*,D^*}|\times\mathrm{Wishart}_{\sum_{k=1}^{K}T_k}\left(m_p,\Sigma^{(t)}/m_p\right)$ or $|J_{\Sigma^*\to R^*,D^*}|\times\text{Inverse-Wishart}_{\sum_{k=1}^{K}T_k}\left(m_p,m_p\Sigma^{(t)}\right)$. Notice that if R and D are prior independent, then $P(R,D|\beta,\gamma,Z,Y)=P(R|\beta,\gamma,Z,Y)$.

Accordingly, the MCMC sampling framework can be implemented through three Gibbs sampling steps for β, Z and γ and one MH sampling step for R, and this sampling method is denoted as Metropolis-Hastings-within-Gibbs Sampler (MH-GS).

3. Non-Identifiable MMVP Model and Parameter-Expanded Data Augmentation

3.1. Non-Identifiable MMVP Model

The identifiable MVP model assumes that the latent variable $Z_i\sim N_k(X_i\beta,R)$ with R being a correlation matrix. The non-identifiable MVP model can be constructed by assuming $Z_i\sim N_k(D^{-1/2}X_i\beta,R)$ or $D^{\frac{1}{2}}Z_i\sim N_k(X_i\beta,D^{\frac{1}{2}}RD^{\frac{1}{2}})$, with D being a diagonal matrix with diagonal elements $(\sigma_{11},\sigma_{22},\cdots,\sigma_{kk})$ serving as redundant variance parameters. Define $W_i=D^{\frac{1}{2}}Z_i$ and $\Sigma=D^{\frac{1}{2}}RD^{\frac{1}{2}}$, then $W_i\sim N_k(X_i\beta,\Sigma)$

with Σ being a covariance matrix without restrictions in the diagonal elements [27]. Then the non-identifiable MVP model can be defined as follows:

$$Y_{ij} = l \Leftrightarrow \zeta_{j,l-1} < W_{ij} \le \zeta_{j,l},$$

with $\zeta_j = (\zeta_{j,0}, \zeta_{j,1}, \cdots \zeta_{j,J_j})$ being the unknown cut-points with $\zeta_{j,0} = -\infty$, $\zeta_{j,1} = 0$, and $\zeta_{J_j} = \infty$, for $j = 1, \cdots, k$.

The non-identifiable MVP model for univariate longitudinal ordinal data can be extended to the MMVP model for multivariate longitudinal ordinal data. The notations for the response variable $Y = (Y_1, \cdots, Y_N)^T$, the covariate matrix X_i for individual i, and the regression parameter vector β remain the same as those defined in Section 2.1. But the non-identifiable MVP model is assumed for each longitudinal ordinal response variable $Y_{i\cdot k} = (Y_{i1k}, \cdots, Y_{iT_k k})^T$, there is a latent multivariate normal variable $W_{i\cdot k} = (W_{i1k}, \cdots, W_{iT_k k})^T$ following a multivariate normal distribution with mean vector being $X_{i\cdot k}\beta_k$ and covariance matrix Σ_k, i.e., $W_{i\cdot k}\left(= D_k^{1/2} Z_{i\cdot k}\right) \sim N_{T_k}(X_{i\cdot k}\beta, \Sigma_k)$ with D_k being a diagonal matrix with diagonal elements $\left(\sigma_{T_k 1}, \sigma_{T_k 2}, \cdots, \sigma_{T_k T_k}\right)$. Denote $W = (W_1, \cdots, W_N)^T$ and $W_i = (W_{i\cdot 1}, \cdots, W_{i\cdot K})^T$ for $i = 1, \cdots, N$. It can be shown that $W_i \sim N_{\sum_{k=1}^K T_k}(X_i\beta, \Sigma)$, and Σ is a covariance matrix with the block diagonal matrices being $\Sigma_1, \cdots, \Sigma_K$. Assume that Y_{itk} has c_{tk} categories, then the cut-points for the latent variable W_{itk} can be defined as follows: $Y_{itk} = c \Leftrightarrow \zeta_{t,k,c-1} < W_{itk} \le \zeta_{t,k,c}$, for $c = 1, \cdots, c_{tk}$, $\zeta_{t,k,0} = -\infty$, $\zeta_{t,k,1} = 0$, and $\zeta_{t,k,c_{tk}} = \infty$; $\zeta = (\zeta_1, \cdots, \zeta_K)^T$ with $\zeta_k = (\zeta_{1k}, \cdots, \zeta_{T_k k})$ and $\zeta_{tk} = (\zeta_{t,k,1}, \cdots, \zeta_{t,k,c_{tk}})$ for $k = 1, \cdots, K$ and $t = 1, \cdots, T_k$. The identifiable cut-points γ_k can be obtained by $D_k^{-1/2}\zeta_k$, and thus $\gamma = D^{-\frac{1}{2}}\zeta$, with D being a diagonal matrix with the block diagonal matrices being $D_1, \cdots, D_K$. There are two circumstances to get the identifiable β: if X_i is the identity matrix, then $D^{-\frac{1}{2}}\beta$ is the identifiable values for β; otherwise use $\frac{1}{T_k}\sum_{t=1}^{T_k} \sigma_{T_k t}^{-1/2} \beta_{kp}$ as the identifiable value of β_{kp}, for $k = 1, \cdots, K$, $t = 1, \cdots, T_k$, and $p = 1, \cdots, p_k$.

3.2. Parameter-Expanded Data Augmentation Sampling Algorithms

Assume the priors for β, ζ and Σ are independent, i.e., $P(\beta, \zeta, \Sigma) = P(\beta) \times P(\zeta) \times P(\Sigma)$. Then the posterior joint density of β, ζ, Σ, and W given the observed multivariate longitudinal ordinal outcome $Y = (Y_1, \cdots, Y_N)^T$ can be derived in the following:

$$P(\beta, \zeta, \Sigma, W \mid Y) \propto P(\beta) \times P(\zeta) \times P(\Sigma) \times P(W|\beta, \zeta, \Sigma, Y)$$

$$\propto P(\beta) \times P(\zeta) \times P(\Sigma) \times \prod_{i=1}^N [\mathrm{I}_i \times \phi(W_i; X_i\beta, \Sigma)]$$

where $\phi(\cdot)$ is the standard normal density function, and $\mathrm{I}_i = \prod_{k=1}^K \mathrm{I}_{ik}$, where $\mathrm{I}_{ik} = \sum_{t=1}^{T_k} 1_{(Y_{itk}=l)}\, 1_{(\zeta_{t,k,l-1} < W_{itk} \le \zeta_{t,k,l})}$, indicating compatibility of the latent variable W_i with the ordinal variable Y_i, $i = 1, \cdots, N$. Then the MCMC sampling algorithm can be derived as follows:

$\beta \mid \zeta, \Sigma, W, Y \sim N_{\sum_{k=1}^K p_k}(\hat{\beta}, V_\beta)$ where $V_\beta = (\sum_{i=1}^N X_i^T \Sigma^{-1} X_i + C^{-1})^{-1}$ and $\hat{\beta} = V_\beta(\sum_{i=1}^N X_i^T \Sigma^{-1} W_i + C^{-1} b)$, assuming the prior of β follows $N_{\sum_{k=1}^K p_k}(b, \mathrm{C})$ with the mean vector equal to b and the covariance matrix equal to C.

$W_{itk} | \beta_k, \zeta_k, \Sigma, Y_{itk}, W_{ijk}, j \ne t$ is the interval truncated normal distribution constrained to lie between two cut-points $\zeta_{t,k,c-1}$ and $\zeta_{t,k,c}$, assuming $Y_{itk} = c$.

$\zeta_{t,k,c} | \beta, \Sigma, W, Y, \zeta_{t,k,l}\ c \ne l$ is a uniform distribution, $U(\zeta_{t,k,c} | \max\{\max[W_{itk}: Y_{itk} = c], \zeta_{t,k,c-1}\}, \min\{\min[W_{itk}: Y_{itk} = c+1], \zeta_{t,k,c+1}\})$, assuming a non-informative prior for $\zeta_{t,k,c}$, i.e., $P(\zeta_{t,k,c}) \propto 1$.

$\Sigma | \beta, \zeta, W, Y \sim$ Inverse-Wishart$_{\sum_{k=1}^K T_k}(\sum_{i=1}^N (W_i - X_i\beta)(W_i - X_i\beta)^T + V, N + \sum_{k=1}^K T_k + m + 1)$, assuming $P(\Sigma) =$ Inverse-Wishart$_{\sum_{k=1}^K T_k}(m, V)$, a conjugate prior for Σ.

As illustrated, the above four sampling steps are Gibbs sampling, and especially the posterior sample for Σ, a covariance matrix, can be directly sampled from an inverse-Wishart distribution, instead of a MH step to sample R, a correlation matrix, for the identifiable model in Section 2.2. This algorithm is termed as the parameter-expanded data augmentation (PX-DA) algorithm. However, including the redundant parameters in the diagonal of D, the model becomes non-identifiable and the sampled values of β and ζ are not identifiable. Therefore, marginalizing D can be considered by jointly sampling $W, D|\beta, \zeta, R, Y$ (sampling $D|\beta, \zeta, R, Y$ followed by sampling $W|\beta, \zeta, R, D, Y$) and then jointly sampling $\beta, \zeta, R, D|W, Y$ (sampling $\beta|R, D, W, Y$, $R, D|\beta, W, Y$, and then $\zeta|\beta, R, D, W, Y$). It is noticeable that $W|\beta, \zeta, R, D, Y$, $\beta|R, D, W, Y$, $R, D|\beta, W, Y$, and $\zeta|\beta, R, D, W, Y$ are the same as those in the PX-DA algorithm by replacing R and D by $\Sigma\left(= D^{\frac{1}{2}} R D^{\frac{1}{2}}\right)$. And the sampling of $D|\beta, \zeta, R, Y$ is actually sampling $D|R$, the conditional prior of D given R. With $P(\Sigma) =$ Inverse-Wishart$_{\sum_{k=1}^K T_k}(m, V)$ and V being a diagonal matrix, the diagonal elements of D are independent given R, and each follows an inverse-Gamma ($\alpha = \frac{m}{2}, \beta = \frac{2}{V_{jj}\widetilde{r_{JJ}}}$) with V_{jj} being the jth element of V and $\widetilde{r_{JJ}}$ being the j th diagonal element of inverse of R for $j = 1, 2, \cdots, \sum_{k=1}^K T_k$ [11]. However, if V is not a diagonal matrix, sampling D necessitates a MH algorithm. Replacing $P(R, D|\beta, \gamma, Z, Y)$ or $P(R|\beta, \gamma, Z, Y)$ by $P(D|R)$, the MH sampling method in section 2.2 can be used to sample D given R. This algorithm is termed as the parameter-expanded data augmentation with marginalization (PX-DAM) algorithm.

4. Simulation Studies

The MH-GS algorithm for the identifiable MMVP model is proposed in Section 2 and the PX-DA and PX-DAM algorithms based on the non-identifiable MMVP model are developed in Section 3. To investigate these algorithms, simulation studies are conducted by generating two 5-dimensional longitudinal ordinal response variables, each with 4 categories. The 5×2 covariate matrix for each longitudinal ordinal variable was generated from the uniform distribution on the interval (-0.5, 0.5), and the regression parameters for the first longitudinal ordinal variable is $\beta_1^T = (\beta_{11}, \beta_{12}) = (1.0, 3.0)$ and those for the second longitudinal ordinal variable is $\beta_2^T = (\beta_{21}, \beta_{22}) = (2.0, 5.0)$. The correlation matrix for the first longitudinal ordinal variable is the first-order autoregressive AR1(0.5), that for the second longitudinal ordinal variable is AR1(0.7), and the correlations between the first and the second longitudinal ordinal variables are set at 0.2. The cut-points for the latent multivariate normal variables are 1 and 2 (the first cut-point is fixed at 0). The non-informative priors for regression parameters β and the cut-points for the identifiable MMVP model γ and for the non-identifiable model ζ are chosen; an Inverse-Wishart prior for Σ with degrees of freedom $m = 20$ and scale matrix V being an identity matrix. Sample sizes of 500, 1000 and 2000 are investigated and 100 data sets are generated for each investigated scenario. Each algorithm runs 10,000 iterations with 2,000 burn-in. The MCMC convergence diagnostics were conducted using the R package-coda by Plummer et al [19] and R package-boa by Smith [21].

The averaged posterior means and standard deviations with 95% credible interval coverage probabilities for the regression parameters and cut-points are presented in Tables 1 and 2 for sample sizes being 500 and 2,000 (the results for sample size being 1,000 is presented in Appendix Table A1). As can be seen, these three methods produce similar estimated values for sample sizes being 500 and 1,000; for sample size being 2,000, the PX-DA algorithm produces biased estimation with lower 95% credible interval coverage probabilities, such as β_{22} with estimated value being 4.873 and 77% coverage probability, $\gamma_{1,1,1}$ with estimated value being 0.93 and 61% coverage probability and all the rest cut-points. Overall, it seems that the MH-GS algorithm gives the most precise estimated values for the cut-points; the PX-DA algorithm shows biased estimation for data with large sample sizes such as 2,000; while the PX-DAM algorithm has the largest standard deviations and thus maximum 95% interval credible coverage probabilities.

Table 1. *Averaged posterior means (Mean), standard deviations (SD) and 95% credible interval coverage probabilities (CP%) for regression parameters and cut-points with sample size 500 based on 100 generated datasets.*

	True Values	MH-GS			PX-DA			PX-DAM		
		Mean	**SD**	**CP%**	**Mean**	**SD**	**CP%**	**Mean**	**SD**	**CP%**
β_{11}	1.0	1.018	0.08	93	1.014	0.08	93	1.017	0.08	95
β_{12}	3.0	3.022	0.09	96	3.000	0.09	96	3.013	0.12	100
β_{21}	2.0	2.007	0.09	91	2.003	0.09	91	2.016	0.10	95
β_{22}	5.0	5.052	0.14	92	5.034	0.14	92	5.068	0.19	100
$\gamma_{1,1,1}$	1.0	1.01	0.07	96	1.01	0.07	93	1.00	0.12	100
$\gamma_{1,1,2}$	2.0	2.04	0.10	95	2.03	0.11	94	2.04	0.22	100
$\gamma_{2,1,1}$	1.0	1.01	0.06	90	0.99	0.07	90	0.98	0.11	100
$\gamma_{2,1,2}$	2.0	2.01	0.10	97	1.99	0.10	90	1.98	0.21	100
$\gamma_{3,1,1}$l	1.0	1.00	0.07	95	0.99	0.07	96	1.00	0.11	100
$\gamma_{3,1,2}$	2.0	2.03	0.10	96	2.00	0.10	96	2.00	0.21	100
$\gamma_{4,1,1}$	1.0	1.00	0.06	94	0.99	0.07	93	0.98	0.11	100
$\gamma_{4,1,2}$	2.0	2.03	0.10	97	2.01	0.10	96	2.00	0.21	100
$\gamma_{5,1,1}$	1.0	1.01	0.06	95	0.97	0.07	93	0.99	0.11	99
$\gamma_{5,1,2}$	2.0	2.01	0.10	93	2.00	0.11	91	2.02	0.21	100
$\gamma_{1,2,1}$	1.0	1.00	0.07	91	1.00	0.08	94	1.04	0.13	100
$\gamma_{1,2,2}$	2.0	2.01	0.09	87	2.01	0.11	93	2.09	0.25	100
$\gamma_{2,2,1}$	1.0	1.03	0.07	92	1.01	0.08	97	0.99	0.13	100
$\gamma_{2,2,2}$	2.0	2.02	0.09	93	2.00	0.11	95	1.97	0.23	100
$\gamma_{3,2,1}$	1.0	1.02	0.07	95	1.01	0.08	96	1.00	0.13	100
$\gamma_{3,2,2}$	2.0	2.03	0.09	88	2.02	0.21	94	2.00	0.23	100
$\gamma_{4,2,1}$	1.0	1.00	0.07	97	1.00	0.08	95	0.99	0.13	100
$\gamma_{4,2,2}$	2.0	2.02	0.09	91	2.01	0.11	94	2.00	0.23	100
$\gamma_{5,2,1}$	1.0	1.01	0.07	95	1.00	0.08	95	1.06	0.12	99
$\gamma_{5,2,2}$	2.0	2.02	0.09	94	2.03	0.11	93	2.12	0.22	100

Table 2. *Averaged posterior means (Mean), standard deviations (SD) and 95% credible interval coverage probabilities (CP%) for regression parameters and cut-points with sample size 2000 based on 100 generated datasets.*

	True Values	MH-GS			PX-DA			PX-DAM		
		Mean	**SD**	**CP%**	**Mean**	**SD**	**CP%**	**Mean**	**SD**	**CP%**
β_{11}	1.0	1.002	0.04	95	0.978	0.04	97	1.000	0.05	98
β_{12}	3.0	3.003	0.04	93	2.928	0.07	83	3.000	0.09	100
β_{21}l	2.0	2.014	0.04	94	1.956	0.05	90	2.008	0.07	100
β_{22}	5.0	5.023	0.07	89	4.873	1.11	77	5.004	0.14	100
$\gamma_{1,1,1}$	1.0	0.99	0.03	94	0.93	0.05	61	0.99	0.10	100
$\gamma_{1,1,2}$	2.0	2.00	0.04	92	1.91	0.07	71	2.00	0.19	100
$\gamma_{2,1,1}$	1.0	1.00	0.03	84	0.93	0.05	78	0.98	0.10	100

	True Values	MH-GS			PX-DA			PX-DAM		
		Mean	SD	CP%	Mean	SD	CP%	Mean	SD	CP%
$\gamma_{2,1,2}$	2.0	2.00	0.04	92	1.91	0.07	73	1.98	0.19	100
$\gamma_{3,1,1}$	1.0	1.01	0.03	88	0.93	0.05	70	0.97	0.10	100
$\gamma_{3,1,2}$	2.0	2.01	0.05	92	1.93	0.07	87	1.99	0.19	100
$\gamma_{4,1,1}$	1.0	1.00	0.03	85	0.93	0.05	70	0.98	0.10	100
$\gamma_{4,1,2}$	2.0	2.01	0.05	88	1.92	0.07	80	1.99	0.19	100
$\gamma_{5,1,1}$	1.0	0.99	0.03	91	0.93	0.05	69	1.00	0.10	100
$\gamma_{5,1,2}$	2.0	2.00	0.05	96	1.94	0.07	87	2.01	0.18	100
$\gamma_{1,2,1}$	1.0	1.00	0.03	89	0.92	0.06	65	1.02	0.12	100
$\gamma_{1,2,2}$l	2.0	2.00	0.04	89	1.89	0.08	70	2.05	0.23	100
$\gamma_{2,2,1}$	1.0	1.01	0.03	95	0.93	0.06	73	0.95	0.10	100
$\gamma_{2,2,2}$	2.0	2.01	0.04	96	1.90	0.08	80	1.93	0.20	100
$\gamma_{3,2,1}$	1.0	1.00	0.03	91	0.93	0.06	71	0.97	0.11	100
$\gamma_{3,2,2}$	2.0	2.02	0.04	88	1.91	0.08	71	1.98	0.21	100
$\gamma_{4,2,1}$	1.0	1.00	0.03	88	0.92	0.06	80	0.96	0.11	100
$\gamma_{4,2,2}$	2.0	2.01	0.04	94	1.90	0.08	75	1.94	0.21	100
$\gamma_{5,2,1}$	1.0	1.00	0.03	93	0.92	0.06	76	1.02	0.10	100
$\gamma_{5,2,2}$	2.0	2.01	0.04	91	1.90	0.08	75	2.07	0.20	100

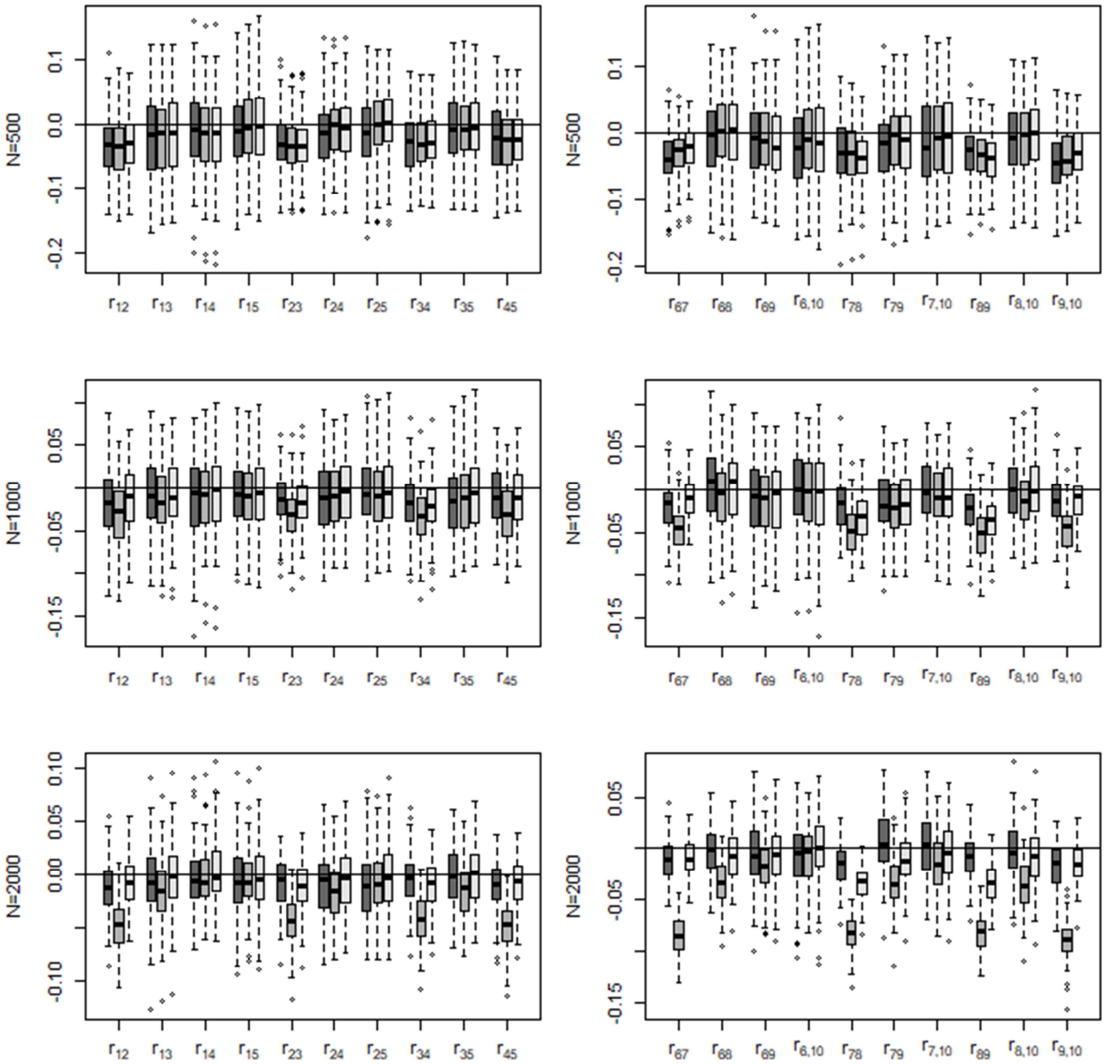

Figure 1. *Boxplots of the absolute biases of correlations for sample sizes being 500, 1000, and 2000 based on 100 generated datasets (those with dark gray for the MH-GS algorithm, those with medium gray for the PX-DA algorithm and those with light gray for PX-DAM algorithm).*

Figure 1 illustrates the boxplots of the absolute biases for correlations of the latent multivariate normal variables from the first longitudinal ordinal response (the first column) and the second longitudinal ordinal response (the second column). It shows that with sample size being 500 (the first row), these three methods produce similar absolute biases for those correlations; with sample size being 1,000 (the second row), the MH-GS and PX-DAM algorithms give similar absolute biases while the PX-DA produces biased estimated correlations, especially for the second longitudinal response; with sample size being 2,000 (the third row), the MH-GS and PX-DAM algorithms still give similar absolute biases while the PX-DA produces serious biased estimated correlations for both longitudinal responses. The boxplots of the absolute biases for the correlations between those two MVP models are presented in Appendix Figure A1. It shows the similarity of these three algorithms for sample sizes being 500 and 1,000 while the PX-DA algorithm has serious biases for sample size being 2,000.

The autocorrelation function (ACF) plots for selected parameters are shown in Figure 2. It is evident that in general the MH-GS algorithm has the slowest decreased ACF values while the PX-DAM algorithm shows the fastest. The ACF values for the PX-DA algorithm display that with large sample size (such as 2,000) the ACF values of the regression parameter β_{22} decreased slower even than that for the MH-GS algorithm; the ACF values of the cut-points significantly decrease slower with larger sample size; the ACF values for the correlations are affected not only by sample sizes but also by the values themselves, i.e., $r_{15} = 0.0625$ without being affected while $r_{67} = 0.7$ being significantly affected by increasing sample sizes.

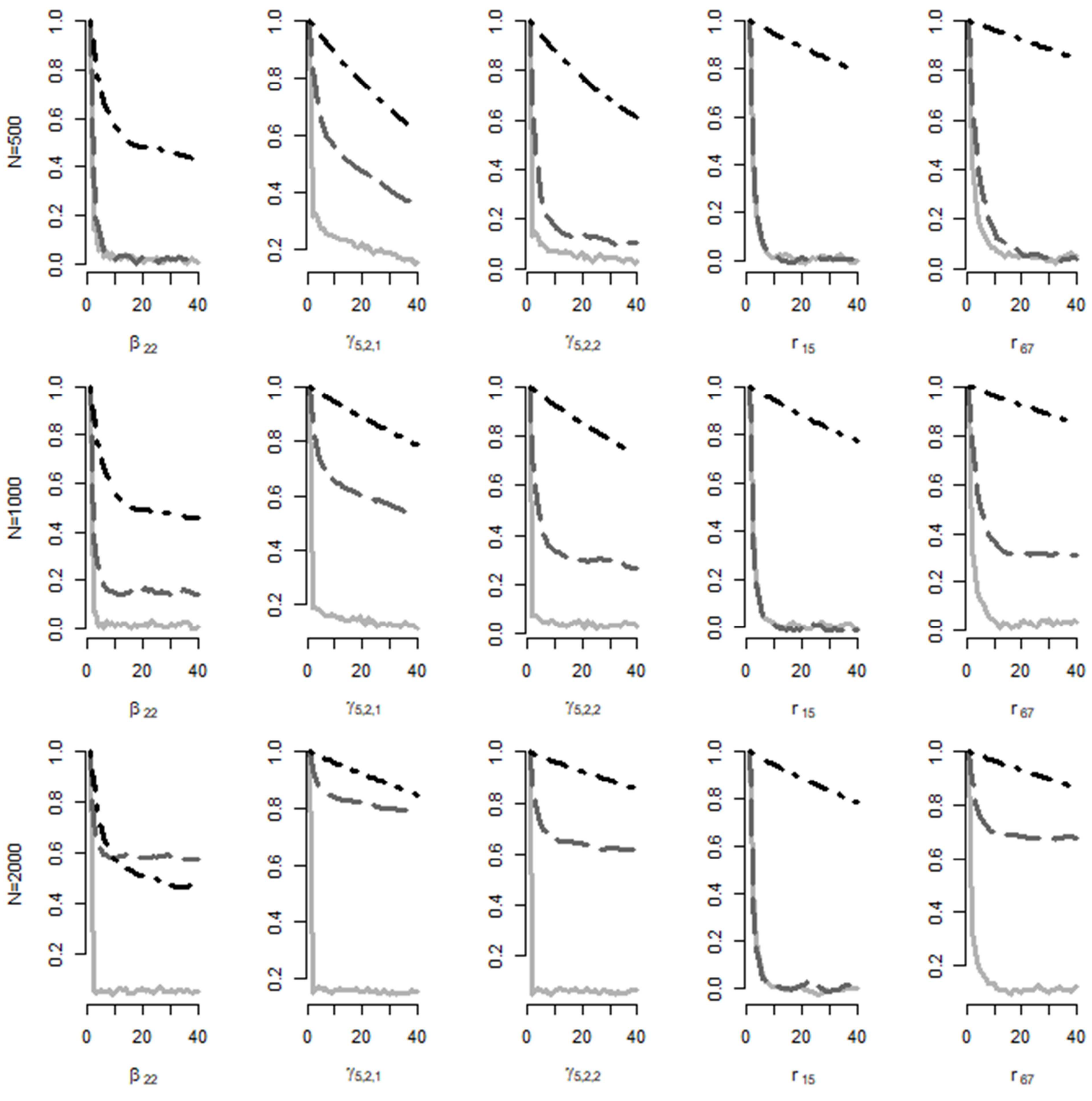

Figure 2. *The ACF plots for selected parameters with sample sizes being 500, 1,000 and 2,000 based on 100 generated datasets. The black dot-dash lines indicate the MH-GS algorithm, the medium gray long-dashed lines indicate the PX-DA algorithm and the light gray solid lines indicate the PX-DAM algorithm.*

5. Application to the RLMS-HSE

The Russia Longitudinal Monitoring Survey-Higher Schools of Economics (RLMS-HSE) was to collect data to study the effects of Russian reforms on the health and economic welfare of households and individuals [14]. One of the main aspects focuses on the job satisfaction with the primary and secondary employments, which is a 5-categorical ordinal variable defined as “absolutely satisfied”, “mostly satisfied”, “neutral”, “not very satisfied” and “absolutely unsatisfied”. To study the relationship between the job satisfaction and other satisfaction, such as satisfaction of education, family, environment, living conditions, etc, is also another focus of the survey. This application is targeted to investigate the relationship between the job satisfaction and the satisfaction with life at present, which is another longitudinal ordinal variable with five categories. The data is chosen from 2013 to 2019 and include gender (1 being male and 0 female), marital status (1 being in a registered marriage and 0 otherwise), education (1 having university diploma and 0 otherwise), and age as the covariates. There are a total of 2704 individuals after excluding those with missing values. Then categories of “not very satisfied” and “absolutely unsatisfied” for job satisfaction are combined and categories “fully satisfied” and “rather satisfies” are combined for the satisfaction with life at present due to sparse sizes in those categories.

The model selection is conducted for job satisfaction using the MVP model: first considering the main effect model including gender, marital status, education and age (all the effects are significant); then - constructing two-way interaction model by including all two-way interaction besides the main effects (only the interaction of marital status and gender is significant); finally, considering the model with all main effects and the interaction of marital status and gender. The log likelihoods, BIC and AIC and selected the model with main effect without any interactions are calculated. The model selection for the satisfaction with life at present is also conducted with the similar conclusion as those for job satisfaction. Therefore, in the following multivariate longitudinal analysis, the main effects as the covariates for both the job satisfaction and the satisfaction with life at present are considered.

The MH-GS, PX-DA and PX-DAM algorithms are run with 20,000 iterations with 10,000 burn-in using the same prior specifications as those in Section 4. Table 3 presents the log-likelihood, the estimated posterior means with standard deviations of the regression parameters for job satisfaction and satisfaction with life at present. It shows that the MH-GS algorithm has the largest log-likelihood values, while the PX-DA algorithm has the smallest among those three algorithms. All the covariates are significant (using 95% credible intervals Specifically, female tends to have a higher job satisfaction than male, while male tend to have a high satisfaction of life at present than female; older people tend to have a high job satisfaction while have a lower satisfaction of life at present; people with married status and diplomas (institute, university, academy) tend to have a higher job satisfaction and satisfaction of life in present as well.

Table 3. *The log-likelihood, estimated regression parameters with standard deviations for job satisfaction and satisfaction of life at present.*

Log-likelihood	**Job Satisfaction**				**Satisfaction of Life at Present**			
	Gender	**Marital**	**Educ**	**Age**	**Gender**	**Marital**	**Educ**	**Age**
MH-GS	-0.0908	0.0835	0.2202	0.0048	0.1655	0.2934	0.2815	-0.0049
(-35653.3)	(.0298)	(.0308)	(.0305)	(.0012)	(.031)	(.0322)	(.0323)	(.0013)
PX-DA	-0.0876	0.0791	0.2057	0.0045	0.1552	0.2749	0.2626	-0.0043
(-35847.2)	(.0294)	(.0303)	(.0299)	(.0013)	(.0311)	(.0318)	(.0323)	(.0013)
PX-DAM	-0.0901	0.0794	0.2179	0.006	0.158	0.2797	0.2681	-0.0033
(-35664.6)	(.0289)	(.0303)	(.0307)	(.0012)	(.0317)	(.0329)	(.0327)	(.0015)

Table 4. *The estimated correlation matrix of the underlying latent multivariate normal variables corresponding to job satisfaction and satisfaction of life from year 2013 to year 2019 using MH-GS.*

Job Satisfaction							**Satisfaction of Life at Present**						
2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1	*0.46*	*0.42*	*0.42*	*0.41*	*0.36*	*0.33*	0.3	0.27	0.23	0.16	0.22	0.2	0.2
0.46	*1*	*0.53*	*0.44*	*0.42*	*0.37*	*0.36*	0.26	0.36	0.27	0.17	0.22	0.24	0.23
0.42	*0.53*	*1*	*0.5*	*0.49*	*0.43*	*0.41*	0.25	0.29	0.32	0.2	0.23	0.25	0.28
0.42	*0.44*	*0.5*	*1*	*0.58*	*0.54*	*0.51*	0.22	0.26	0.27	0.28	0.3	0.27	0.29
0.41	*0.42*	*0.49*	*0.58*	*1*	*0.56*	*0.52*	0.26	0.31	0.31	0.26	0.37	0.33	0.32
0.36	*0.37*	*0.43*	*0.54*	*0.56*	*1*	*0.58*	0.2	0.22	0.24	0.21	0.26	0.37	0.32
0.33	*0.36*	*0.41*	*0.51*	*0.52*	*0.58*	*1*	0.22	0.27	0.24	0.2	0.25	0.33	0.4
0.3	0.26	0.25	0.22	0.26	0.2	0.22	*1*	*0.51*	*0.47*	*0.42*	*0.43*	*0.41*	*0.35*
0.27	0.36	0.29	0.26	0.31	0.22	0.27	*0.51*	*1*	*0.53*	*0.48*	*0.44*	*0.42*	*0.4*
0.23	0.27	0.32	0.27	0.31	0.24	0.24	*0.47*	*0.53*	*1*	*0.55*	*0.5*	*0.45*	*0.44*
0.16	0.17	0.2	0.28	0.26	0.21	0.2	*0.42*	*0.48*	*0.55*	*1*	*0.55*	*0.49*	*0.45*
0.22	0.22	0.23	0.3	0.37	0.26	0.25	*0.43*	*0.44*	*0.5*	*0.55*	*1*	*0.57*	*0.49*
0.2	0.24	0.25	0.27	0.33	0.37	0.33	*0.41*	*0.42*	*0.45*	*0.49*	*0.57*	*1*	*0.58*
0.2	0.23	0.28	0.29	0.32	0.32	0.4	*0.35*	*0.4*	*0.44*	*0.45*	*0.49*	*0.58*	*1*

Table 5. *The estimated correlation matrix of the underlying latent multivariate normal variables corresponding to job satisfaction and satisfaction of life from year 2013 to year 2019 using PX-DA.*

Job Satisfaction							**Satisfaction of Life at Present**						
2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1	0.43	0.4	0.39	0.39	0.34	0.31	0.3	0.27	0.24	0.16	0.22	0.2	0.2
0.43	1	0.5	0.42	0.4	0.35	0.35	0.26	0.34	0.28	0.19	0.22	0.23	0.22
0.4	0.5	1	0.48	0.47	0.41	0.39	0.25	0.29	0.32	0.21	0.24	0.24	0.27
0.39	0.42	0.48	1	0.54	0.5	0.47	0.23	0.26	0.27	0.28	0.31	0.27	0.3
0.39	0.4	0.47	0.54	1	0.52	0.48	0.26	0.3	0.31	0.26	0.36	0.32	0.31
0.34	0.35	0.41	0.5	0.52	1	0.53	0.22	0.23	0.26	0.22	0.26	0.35	0.31
0.31	0.35	0.39	0.47	0.48	0.53	1	0.22	0.27	0.25	0.21	0.25	0.31	0.37
0.3	0.26	0.25	0.23	0.26	0.22	0.22	1	0.49	0.47	0.43	0.44	0.41	0.37
0.27	0.34	0.29	0.26	0.3	0.23	0.27	0.49	1	0.53	0.48	0.44	0.41	0.4
0.24	0.28	0.32	0.27	0.31	0.26	0.25	0.47	0.53	1	0.54	0.5	0.45	0.45
0.16	0.19	0.21	0.28	0.26	0.22	0.21	0.43	0.48	0.54	1	0.54	0.49	0.46
0.22	0.22	0.24	0.31	0.36	0.26	0.25	0.44	0.44	0.5	0.54	1	0.55	0.49
0.2	0.23	0.24	0.27	0.32	0.35	0.31	0.41	0.41	0.45	0.49	0.55	1	0.57
0.2	0.22	0.27	0.3	0.31	0.31	0.37	0.37	0.4	0.45	0.46	0.49	0.57	1

Table 6. *The estimated correlation matrix of the underlying latent multivariate normal variables corresponding to job satisfaction and satisfaction of life from year 2013 to year 2019 using PX-DAM.*

Job Satisfaction							**Satisfaction of Life at Present**						
2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1	0.47	0.42	0.42	0.41	0.36	0.33	0.33	0.29	0.25	0.17	0.23	0.21	0.21
0.47	1	0.54	0.45	0.43	0.37	0.37	0.28	0.37	0.3	0.2	0.23	0.24	0.23
0.42	0.54	1	0.51	0.5	0.44	0.42	0.27	0.31	0.35	0.22	0.25	0.26	0.29
0.42	0.45	0.51	1	0.58	0.54	0.51	0.24	0.28	0.29	0.3	0.33	0.28	0.31
0.41	0.43	0.5	0.58	1	0.56	0.51	0.28	0.33	0.33	0.27	0.39	0.34	0.33
0.36	0.37	0.44	0.54	0.56	1	0.57	0.23	0.24	0.27	0.23	0.28	0.38	0.33
0.33	0.37	0.42	0.51	0.51	0.57	1	0.24	0.29	0.27	0.22	0.26	0.33	0.4
0.33	0.28	0.27	0.24	0.28	0.23	0.24	1	0.54	0.51	0.45	0.47	0.44	0.39
0.29	0.37	0.31	0.28	0.33	0.24	0.29	0.54	1	0.57	0.51	0.47	0.44	0.43
0.25	0.3	0.35	0.29	0.33	0.27	0.27	0.51	0.57	1	0.58	0.54	0.48	0.48
0.17	0.2	0.22	0.3	0.27	0.23	0.22	0.45	0.51	0.58	1	0.58	0.52	0.49
0.23	0.23	0.25	0.33	0.39	0.28	0.26	0.47	0.47	0.54	0.58	1	0.59	0.52
0.21	0.24	0.26	0.28	0.34	0.38	0.33	0.44	0.44	0.48	0.52	0.59	1	0.6
0.21	0.23	0.29	0.31	0.33	0.33	0.4	0.39	0.43	0.48	0.49	0.52	0.6	1

Tables 4, 5 and 6 present the estimated correlations of the underlying latent multivariate normal variables corresponding to job satisfaction and satisfaction of life at present (the estimated standard deviations vary from 0.02 to 0.03 for each method). It indicates that the correlations of the latent variables for satisfaction of life at present varying from 0.39 to 0.60 are a little higher than those for job satisfaction varying from 0.33 to 0.58; the correlations between those two measures vary from 0.17 to 0.40, suggesting that those two measures are modestly correlated. As can be seen, the estimated correlations using the PX-DA algorithm are unanimously smaller than the MH-GS and PX-DAM algorithms; for instance, r_{12} (year 2013 vs year 2014 for job satisfaction) is estimated to be 0.43 for the PX-DA algorithm while be 0.46 for the MH-GS algorithm and 0.47 for the PX-DAM algorithm, and r_{89} (year 2013 vs year 2014 for life satisfaction) is estimated to be 0.49 for the PX-DA algorithm while 0.51 for the MH-GS algorithm and 0.54 for the PX-DAM algorithm. This phenomenon is consistent with the findings in simulation studies in section 5, i.e., the PX-DA algorithm may produce biased or underestimated correlations in comparison with the MH-GS and PX-DAM algorithms.

6. Discussion

In this manuscript the MH-GS algorithm based on the identifiable MMVP model and the PX-DA and PX-DAM algorithms based on the non-identifiable MMVP model are proposed for analyzing multivariate longitudinal ordinal data.

Simulation studies show that those proposed algorithms produce similar estimated regression parameters, cut-points and correlations for data with sample size being 500, while the PX-GS algorithm tends to give biased estimation for data with large sample sizes, such as 1,000 and 2,000. However, the PX-DA and PX-DAM algorithms have faster convergences for almost all parameters than the MH-GS algorithm does, especially for sample size being 500 and 1,000. This suggests that the sampling methods based on the non-identifiable models improve the convergences of the MCMC components compared with those based on the identifiable models. However, for data with large sample sizes, marginalizing the redundant parameters in the non-identifiable models should be considered, otherwise it may produce biased estimated values.

Real data application illustrates that the MH-GS algorithm has the largest log-likelihood values among those three algorithms, suggesting that the algorithm based on the identifiable model produces more precise estimated values in comparison with those based on the non-identifiable models. The observation regarding correlations estimation, i.e., the PX-DA algorithm produces biased or underestimated correlations in comparison with the MH-GS and PX-DAM algorithms, is consistent with the findings of simulation studies.

7. Conclusion

The MCMC methods proposed in this manuscript can provide a joint analysis of multivariate longitudinal ordinal data and provide a direct estimation for the correlations among the underlying multivariate normal variables for multivariate longitudinal ordinal data. However, both mixed-effects models and separate analysis cannot provide a direct estimation for correlations of those multivariate measurements.

The investigation further demonstrates that MCMC sampling methods based on non-identifiable models may improve the convergences in comparison with those based on the identifiable models, while those based on the identifiable models may produce model with larger log-likelihood values than those based on the non-identifiable models. However, non-identifiable models may produce serious biased estimation without marginalization of the redundant parameters for data with large sample sizes.

Further investigation of those proposed algorithms in other multivariate longitudinal data structures, such as mixed longitudinal ordinal and continuous data, will be one of our future research focuses.

Appendix

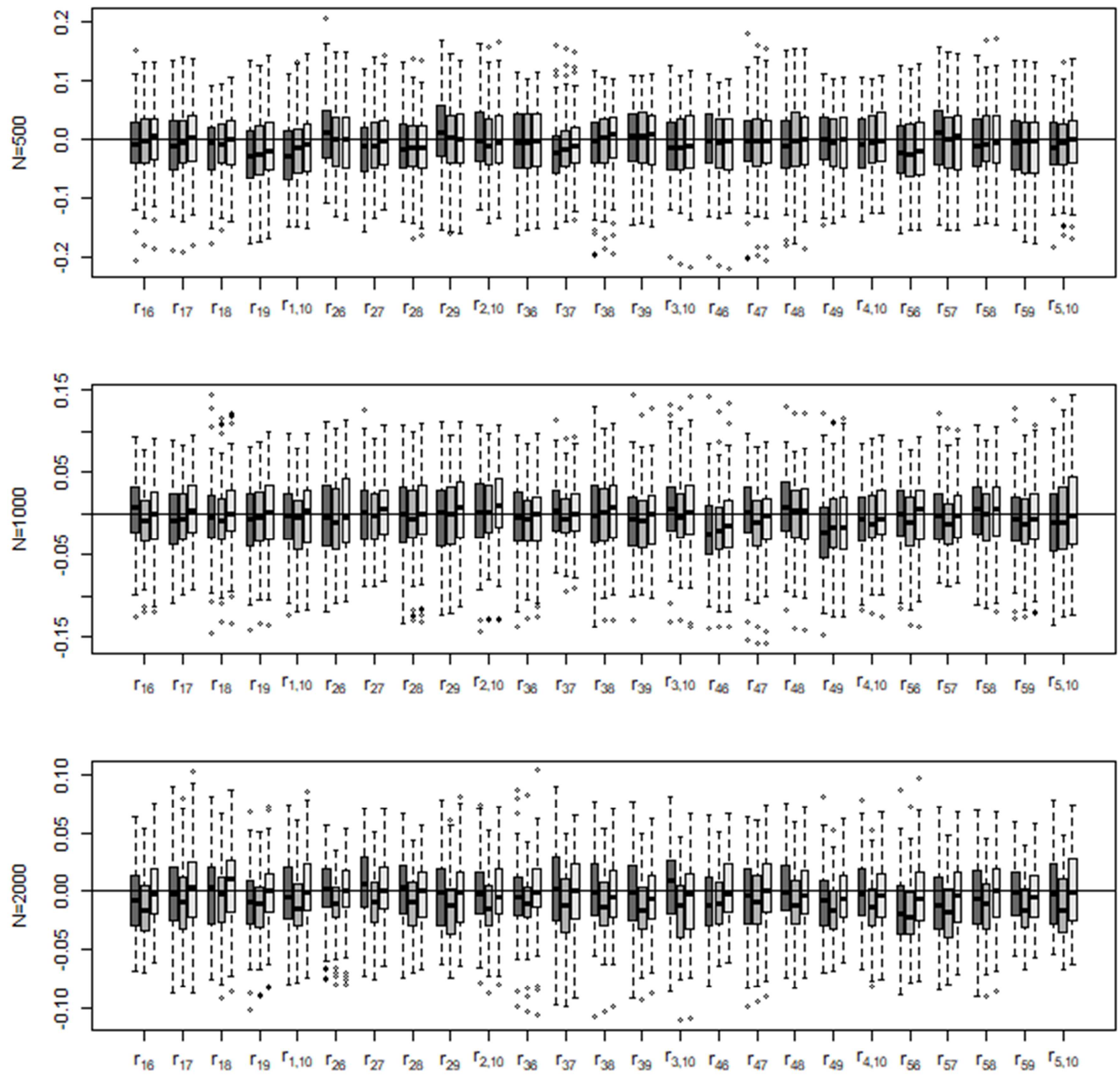

Figure A1. *Boxplots of the absolute biases of correlations for sample sizes being 500, 1000, and 2000 based on 100 generated datasets (those with dark gray for the MH-GS algorithm, those with medium gray for the PX-DA algorithm and those with light gray for PX-DAM algorithm).*

Table A1. *Averaged posterior means (Mean), standard deviations (SD) and 95% credible interval coverage probabilities (CP%) for regression parameters and cut-points with sample size 1000 based on 100 generated datasets.*

	True Values	MH-GS			PX-DA			PX-DAM		
		Mean	SD	CP%	Mean	SD	CP%	Mean	SD	CP%
β_{11}	1.0	1.008	0.05	98	0.995	0.05	98	1.006	0.06	98
β_{12}	3.0	3.008	0.07	97	2.971	0.07	93	3.000	0.10	100
β_{21}	2.0	2.001	0.06	92	1.978	0.06	93	2.001	0.08	97
β_{22}	5.0	5.022	0.09	95	4.969	0.11	94	5.033	0.16	99
$\gamma_{1,1,1}$	1.0	1.00	0.04	90	0.98	0.05	94	1.00	0.10	100
$\gamma_{1,1,2}$	2.0	2.01	0.07	95	1.97	0.08	93	2.01	0.20	100
$\gamma_{2,1,1}$	1.0	1.00	0.04	94	0.96	0.05	92	0.99	0.10	100
$\gamma_{2,1,2}$	2.0	2.01	0.07	93	1.97	0.08	93	1.99	0.20	100
$\gamma_{3,1,1}$	1.0	1.00	0.04	91	0.96	0.05	86	0.99	0.10	100
$\gamma_{3,1,2}$	2.0	2.01	0.07	93	1.97	0.08	93	1.99	0.20	100
$\gamma_{4,1,1}$	1.0	1.00	0.04	94	0.97	0.05	94	0.98	0.10	100
$\gamma_{4,1,2}$	2.0	2.02	0.07	94	1.98	0.08	94	1.99	0.20	100
$\gamma_{5,1,1}$	1.0	1.01	0.04	90	0.98	0.05	94	1.01	0.10	100
$\gamma_{5,1,2}$	2.0	2.00	0.07	93	1.96	0.08	92	2.01	0.19	100
$\gamma_{1,2,1}$	1.0	1.00	0.05	95	0.97	0.06	98	1.04	0.13	100
$\gamma_{1,2,2}$	2.0	2.00	0.06	91	1.97	0.08	95	2.07	0.23	100
$\gamma_{2,2,1}$	1.0	1.00	0.04	92	0.97	0.05	95	0.97	0.11	100
$\gamma_{2,2,2}$	2.0	2.00	0.06	91	1.96	0.08	94	1.95	0.21	100
$\gamma_{3,2,1}$	1.0	1.00	0.05	96	0.97	0.06	96	0.99	0.12	100
$\gamma_{3,2,2}$	2.0	2.01	0.06	97	1.95	0.08	88	1.97	0.22	100
$\gamma_{4,2,1}$	1.0	1.00	0.05	87	0.96	0.06	87	0.97	0.11	100
$\gamma_{4,2,2}$	2.0	2.00	0.06	92	1.97	0.08	91	1.96	0.22	100
$\gamma_{5,2,1}$	1.0	1.00	0.05	96	0.97	0.06	92	1.05	0.11	100
$\gamma_{5,2,2}$	2.0	2.00	0.06	89	1.96	0.08	94	2.07	0.20	100

References

[1] Bandyopadhyay S and Ganguli B. A review of multivariate longitudinal data analysis. Statistical Methods in Medical Research. 2011; 20: 299–330.

[2] Barnard, J., McCulloch, R. and Meng, X. L. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica, 2000, 1281-1311.

[3] Box, G. E. P. and Tiao, G. (1973) Bayesian inference in statistical analysis. Reading Addison-Wesley, New York.

[4] Cagnone S, Moustaki I and Vasdekis V. Latent variable models for multivariate longitudinal ordinal responses. British Journal of Mathematical and Statistical Psychology 2009; 62: 401–415.

[5] Chib S, Greenberg E. Analysis of multivariate probit models. *Biometrika*, 1998; 85: 347-361.

[6] Cho, H. The analysis of multivariate longitudinal data using multivariate marginal models. Journal of Multivariate Analysis. 2016; 143: 481-491.

[7] Dunson D. B. Bayesian Latent Variable Models for Clustered Mixed Outcomes. Journal of the Royal Statistical Society. Series B. 2000; 62: 355-366.

[8] Gibbons, R. and Hedeker, D. Random effects probit and logistic regression models for three-level data. Biometrics 1997; 53: 1527–1537.

[9] Grigorova D, Encheva E, Gueorguieva R. EM algorithm for mle of a probit model for multiple ordinal outcomes. Serdica J. Computing, 2013; 3, 227–244.

[10] Grigorova D. EM algorithm for maximum likelihood estimation of correlated probit model for two longitudinal ordinal outcomes. Ann. Sofia Univ., Fac. Math and Inf., 2017; 104: 217–232.

[11] Gupta, A. K., and Nagar, D. K. (2000), Matrix Variate Distributions, Boca Raton: Chapman & Hall/CRC.

[12] Heagerty P. J. and Zeger S. L. Marginal regression models for clustered ordinal measurements. Journal of the American Statistical Association. 1996; 435: 1024-1036.

[13] Jiang Z, Liu Y, Wahed AS, Molenberghs G. Joint modeling of multiple ordinal adherence outcomes via generalized estimating equations with flexible correlation structure. Statistics in Medicine 2018; 37: 983-995.

[14] Kozyreva, P., Kosolapov, M., and Popkin, B. M. Data resource profile: The Russia Longitudinal Monitoring Survey—Higher School of Economics (RLMS-HSE) Phase II: Monitoring the Economic and Health Situation in Russia, 1994–2013. International Journal of Epidemiology, 2016; 395–401.

[15] Laffont C. M., Vandemeulebroecke M, Concordet D. Multivariate analysis of longitudinal ordinal data with mixed effects models, with application to clinical outcomes in osteoarthritis. Journal of the American Statistical Association 2014; 109: 955-966.

[16] Lawrence E, Liu C, Bingham D, and Nair VN. Bayesian inference for multivariate ordinal data using parameter expansion. Technometrics. 2008; 50: 182-191.

[17] Liu L. C. and Hedeker D. A Mixed-Effects Regression Model for Longitudinal Multivariate Ordinal Data. Biometrics 2006; 62: 261–268.

[18] Liu X, Daniels MJ. A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *Journal of Computational and Graphical Statistics*, 2006; 15: 897-914.

[19] Plummer, M., Best, N., Cowles, K. and Vines, K., 2015. Package 'coda'. URL http://cran. r-project. org/web/packages/coda/coda. Pdf.

[20] Qu, Y, Piedmonte, M. R. and Medendorp, S. V. Latent variable models for clustered ordinal data. Biometrics. 1995; 51: 268-275.

[21] Smith, B. J. 2016. Package 'boa'. URL http://cran.r-project.org/web/packages/boa/boa. pdf, accessed June, 23, p. 2016.

[22] Spiess, M, Fernández, D, Nguyen T, Liu I. Generalized estimating equations to estimate the ordered stereotype logit model for panel data. Statistics in Medicine. 2020; 39: 1919-1940.

[23] Tran T. D, Lesaffre M, Verbeke G, Molenberghs G. Serial correlation structures in latent linear mixed models for analysis of multivariate longitudinal ordinal responses. Statistics in Medicine. 2021; 40: 578–592.

[24] Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: A review. Stat Methods Med Res. 2014; 23: 42–59.

[25] Zhang X, Boscardin WJ, Belin T. Sampling correlation matrices in Bayesian models with correlated latent variables. *Journal of Computational and Graphical Statistics*, 2006; 15: 880-896.

[26] Zhang X. Parameter-expanded data augmentation for analyzing correlated binary data using multivariate probit models. Statistics in Medicine. 2020; 39: 3637-3652.

[27] Zhang, X.: Bayesian analysis of longitudinal ordinal data using non-identifiable multivariate probit models. Journal of Mathematics and Statistics. 2022; 18: 163-175.