Journal of Mathematics and Statistics

Original Research Paper

Bayesian Analysis of Longitudinal Ordinal Data Using Non-Identifiable Multivariate Probit Models

Xiao Zhang

Department of Mathematics, Michigan Technological University, United States

Article history
Received: 27-04-2022
Revised: 14-10-2022
Accepted: 15-10-2022

Email: xzhang35@mtu.edu

Abstract: Multivariate probit models have been explored for analyzing longitudinal ordinal data. However, the inherent identification issue in multivariate probit models requires the covariance matrix of the underlying latent multivariate normal variables to be a correlation matrix and thus hinders the development of efficient Bayesian sampling methods. It is known that non-identifiable models may produce Markov Chain Monte Carlo (MCMC) samplers with better convergence and mixing than identifiable models. Therefore, we were motivated to construct a non-identifiable multivariate probit model and to develop efficient MCMC sampling algorithms. In comparison with the MCMC sampling algorithm based on the identifiable multivariate probit model, which requires a Metropolis-Hastings (MH) algorithm for sampling a correlation matrix, our proposed MCMC sampling algorithms based on the non-identifiable model circumvent an MH algorithm by a Gibbs sampler for sampling a covariance matrix and thus accelerate the MCMC convergence. We illustrate our proposed methods using simulation studies and two real data applications. Both the simulation studies and the real data applications show that constructing nonidentifiable models may improve the convergence of the MCMC algorithms compared with the identifiable models. The marginalization of the redundant parameters in the non-identifiable models should be considered in developing efficient MCMC sampling algorithms. This investigation shows that construction of non-identifiable models is valuable in developing MCMC sampling methods and illustrates advantages and disadvantages of construction of non-identifiable models to improve the convergence of the MCMC sampling components.

Keywords: MCMC, Longitudinal Ordinal Data, Multivariate Probit Model, Non-Identifiable Multivariate Probit Model

Introduction

Ordinal data are ubiquitous in many scientific fields, such as medical research, behavioral research, social sciences, and customer surveys. For example, the patient pain assessment after surgery is normally measured by ordered categories, such as “no pain,” “mild pain,” “moderate pain,” and “high pain.” To evaluate the severity of illness, the outcome is often categorized as “normal,” “mildly ill,” “severely ill,” and “extremely ill.” These ordinal measures are usually collected at multiple time points.

The research field for analyzing longitudinal ordinal data can be generally divided into mixed-effects models and marginal models (Agresti, 2003; Molenberghs and Verbeke, 2005). The mixed-effects (or so-called subject-specific) models describe the mean response depending on both the covariates as fixed effects and a vector of random effects/variables varying by individuals/units (Hedeker and Gibbons, 2006; Varin and Czado, 2010; Grilli and Rampichini, 2011; Hedeker, 2015; Ursino and Gasparini, 2018). By including the random effects/variables, the mixed-effect models are flexible to incorporate the correlated data structure in the analysis. However, they are difficult to handle more than a few random effects in the maximum likelihood estimation and cannot provide a direct estimation for the correlations of multivariate data. Marginal (or so-

called population-average) models have also been popularly explored for longitudinal ordinal data and provide direct estimation for population-based means (without random effects) and correlations (Li and Schafer, 2008, Pagui *et al*., 2015; Hirk *et al*., 2019). Due to the lack of explicit likelihood functions, the maximum likelihood estimation is computationally intensive. Therefore, the Generalized Estimating Equations (GEE) approach (Liang and Zeger, 1986) provides a convenient alternative to maximum likelihood estimation, especially for longitudinal categorical and ordinal data (Lumley, 1996; Parsons *et al*., 2006; Touloumis *et al*., 2013; Nooraee *et al*., 2014; Ditlhong *et al*., 2018; da Silva *et al*., 2019). Marginalized models, integrating the random effects in the likelihood functions to get the estimation for marginal regression parameters, have also been explored for longitudinal ordinal data (Lee and Daniels, 2008; Lee *et al*., 2016; Schildcrout *et al*., 2022).

With the theoretical development of the Gibbs sampler and the maturation of Markov chains theory (Gelfand and Smith, 1990; Gilks *et al*., 1996; Gelman *et al*., 1995), the Markov chain Monte Carlo (MCMC) methods were popularized and have become a general computation tool in Bayesian inference. The MCMC methods have been developed for analyzing longitudinal ordinal data using generalized linear models (Johnson, 2003; Hadfield, 2010). Browne and Draper (2006) compared Bayesian and likelihood-based methods for fitting multilevel models and pointed out that the Bayesian methods have several advantages compared with the likelihood-based methods. Multivariate logistic and multivariate probit models have been exploited for analyzing longitudinal binary data using marginal models from a Bayesian perspective (O'Brien and Dunson, 2004, Zhang *et al*., 2006; Liu and Daniel, 2006; Zhang, 2020). However, compared with longitudinal binary data, analyzing longitudinal ordinal data using marginal models is much less familiar. In this article, we try to fill in the gaps by proposing the MCMC methods to analyze longitudinal ordinal data using the multivariate probit models.

Following Pearson (1900); Ashford and Sowden (1970), the multivariate probit models have been utilized to analyze longitudinal binary/ordinal data. The probit model assumes that there is a latent variable following a normal distribution underlying each binary/ordinal variable, whereas the multivariate probit model assumes that the latent variables underlying the longitudinal binary/ordinal variables follow a multivariate normal distribution. Due to the identification issue, the covariance matrix of these latent variables is a correlation matrix (Drasgow, 2014; Albert and Chib, 1993; Chib and Greenberg, 1998).

Tanner and Wong (1987) proposed the data augmentation algorithm by including the missing data as unknown quantities to facilitate MCMC sampling. With the probit model assumption, including the latent variables treated as missing data, data augmentation can be developed to analyze the probit model. Albert and Chib (1993) proposed data augmentation algorithms for analyzing the univariate binary, ordinal and nominal data using the probit model. Chib and Greenberg (1998) extended the data augmentation algorithm to analyze longitudinal binary data under the assumption of the multivariate probit model. However, sampling a correlation matrix, constrained by the identifiable model, brings difficulties in developing an efficient Bayesian sampling algorithm. Different MH algorithms have been proposed to sample a correlation matrix. However, using the MH algorithm to sample each element of the correlation matrix is not desirable even for moderate-dimensional data (Chib and Greenberg, 1998), while sampling the whole correlation matrix may suffer slow convergence (Zhang *et al*., 2006) or entail stringent prior distributions for the correlation matrix (Liu and Daniel, 2006). Liu (2001) used reparametrization to sample a covariance matrix instead of a correlation matrix. However, the reparametrization makes the choice of the prior distributions and the explanations inflexible.

This rigid task of sampling a correlation matrix for longitudinal binary data is inherent with longitudinal ordinal data using identifiable multivariate probit models. MacEachern (2007) pointed out that the Markov chains produced by non-identifiable models may improve the convergence rates compared with those by identifiable models, as illustrated by the Dirichlet process. The non-identifiable models can be constructed by introducing a redundant parameter or parameter vector in the identifiable models. Based on the non-identifiable models, Liu and Wu (1999) proved that the convergence of the MCMC sampler, called parameter-expanded data augmentation, is no slower than the original MCMC sampler under mild conditions and illustrated using a univariate probit model for binary data. Used a non-identifiable multivariate binomial probit model to analyze multivariate binary data. Lawrence *et al*. (2008) used a non-identifiable multivariate probit model for multivariate ordinal data, however, without considering Jacobian transformation, their algorithm is not applicable. Zhang (2020) investigated nonidentifiable multivariate probit models to analyze multivariate binary data. Other related works for using a multinomial probit model to analyze univariate categorical data can be found in McCulloch and Rossi (1994); Nobile (1998); McCulloch *et al*. (2000); Imai and van Dyk (2005).

Motivated by the above-contributed work, in this article we propose a non-identifiable multivariate probit model and develop efficient MCMC sampling methods to analyze longitudinal ordinal data. Our investigation shows that MCMC sampling algorithms based on the non-identifiable model significantly improve the convergence and mixing of the sampling components in comparison with those based on the identifiable model. The remainder of the article is organized as follows. In the Section "Multivariate Probit Model for Longitudinal Ordinal Data", we present the identifiable multivariate probit models for longitudinal ordinal data and describe the MCMC sampling algorithm based on Zhang *et al*. (2006). In the section "Non-identifiable Multivariate Probit Models and the MCMC Sampling Algorithms," we propose a nonidentifiable multivariate probit model and develop the MCMC sampling algorithms with and without marginalization of the redundant parameters. We then illustrate our proposed methods using simulations in the Section "Simulation Studies" and two real data applications in the Section "Real Data Applications". Discussion and conclusions are offered in the Section "Discussion".

Materials and Methods

Multivariate Probit Model for Longitudinal Ordinal Data

We begin by reviewing the univariate probit model for ordinal data. Suppose there are n individuals. Each individual has an ordinal outcome, Yi, with J ordinal categories and a $p \times 1$ covariate vector X_i for $i = 1, \ldots, n$. 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 1, denoted by $N\left(X_i^T\beta,1\right)$, where β is the $p \times 1$ regression parameter vector. The model further assumes that:

$$Y_i = l \Leftrightarrow \gamma_{l-1} < Z_i \le \gamma_l \ \ for\, l = 1,...,J$$

i.e., $P\left(Y_j \le l\right) = \Phi\left(\gamma l - X_i^T\beta\right)$ where Φ(.) is the standard normal distribution function and $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_J)$ being the unknown cut-points. It is usually defined that $\gamma_0 = -\infty$, $\gamma_1 = 0$ and $\gamma_J = \infty$ for the model identification purpose.

The multivariate probit model then assumes each individual i for $i = 1, \ldots, n$ has a $k \times 1$ longitudinal ordinal outcome vector $Y_i = (Y_{i1}, \ldots, Y_{ik})^T$ and a $k \times p$ covariate matrix $Xi = (X_{i1}, \ldots, X_{ik})^T$. Each Y_{ij} has J_j ordinal categories for $i = 1, \ldots, n$ and $j = 1, \ldots, k$. Then, the univariate probit model is still assumed for each component of Y_i, Y_{ij} by:

$$Y_{ij} = l \Leftrightarrow \gamma_{j,l-1} < Z_{ij} \le \gamma_{j,l} \ for l = 1,...,J_j \qquad (2.1)$$

where, Z_{ij} is the underlying latent variable following $N\left(X_{ij}^T\beta,1\right)$ and $\gamma_j = (\gamma_{j,0}, \gamma_{j,1}, \ldots \gamma_j, _{Jj})$ is the unknown cut-points with $\gamma j,0 = -\infty$, $\gamma_{Jj} = 0$ and $\gamma J_j = \infty$.

Since, $Y_{i1}, \ldots,$ *and* Y_{ik} are longitudinally collected from the same individual i, they are correlated to each other, and therefore, their corresponding latent variables $Z_{i1}, \ldots, Z_{ik}$ is assumed to be correlated. A univariate probit model is assumed for each Y_{ij}, thus the variance of each Z_{ij} being equal to 1 and the covariance matrix of $Z_i = (Z_{i1}, \ldots, Z_{ik})^T$ underlying each $Y_i = (Y_{i1}, \ldots, Y_{ik})^T$ is, in fact, a correlation matrix, called the polychoric correlation matrix of Y_i (Drasgow, 2014), denoted by R. It can be said that Z_i follows a multivariate normal distribution with the mean vector being $X_i\beta$ and the covariance matrix R being a correlation matrix, i.e., $Z_i \sim N_k(X_i\beta, R)$.

Assume an independent prior distribution for β, R, and γ, i.e., $P(\beta, R, \gamma) = P(\beta) \times P(R) \times P(\gamma)$. Then, we can derive the posterior joint density of β, R, γ, and Z as given Y as follows:

$$\begin{aligned} &P(\beta,R,\gamma,Z \mid Y) \propto P(\beta)P(R)\times P(Z \mid \beta,R,\gamma,Y) \\ &\propto P(\beta)\times P(R)\times P(\gamma)\times \prod_{i=1}^{n}\left[I_i \times \phi\left(Z_i; X_i\beta, R\right)\right] \end{aligned}$$

where, ϕ(.) is the standard normal density function and $I_i = \prod_{j=1}^{k} I_{ij}$, where $I_{ij} = \sum_{t=1}^{J_j} 1\left(Y_{ij} = t\right) 1_{\left(\gamma_j(t-1) < Z_{ij} \le \gamma_{jt}\right)}$, indicates the compatibility of the latent variable Z_{ij} with the ordinal variable Y_{ij} defined in (2.1). To implement the MCMC sampling, the sampling steps based on the full conditional distributions can be described in the following.

Step 2.1 $\beta \mid \gamma, R, Z, Y \sim N_k\left(\hat{\beta}, V_\beta\right)$, where, $V_\beta = \left(\sum_{i=1}^{n} X_i^T R^{-1} X_i + C^{-1}\right)^{-1}$ and $\hat{\beta} = V_\beta\left(\sum_{i=1}^{n} X_i^T R^{-1} Z_i + C^{-1} b\right)$, assuming the prior of β follows $N_p(b, C)$ with the mean vector equal to b and the covariance matrix equal to C.

Step 2.2: $Z_{ij} \mid \beta, \gamma, R, Y, Z_{ik}, k \ne j$ has interval truncated normal distribution constrained to lie between the two cut-points $\gamma_{j,\, l\text{-}1}$ and $\gamma_{j,\, l\text{-}1}$ and $\gamma_{j,\, l}$ assuming $Y_{ij} = l$.

Step 2.3: $\gamma_{j,\, l} \mid \beta, R, Z, Y, \gamma_{j,\, k}, k \ne l$ is a uniform distribution:

$$U\left(\gamma_{j,l} \mid \max\left\{\max\left[Z_{ij} : Y_{ij} = l\right], \gamma_{j,l-1}\right\}, \min\left\{\min\left[Z_{ij} : Y_{ij} = l+1\right], \gamma j, l+1\right\}\right)$$

assuming a non-informative prior for $\gamma_{j,\, l}$.

Step 2.4: The full conditional density function of R is $P(R \mid \beta, \gamma, Z, Y) \propto P(R) \times \prod_{i=1}^{n} \phi\left(Z_i; X_i\beta, R\right)$. This full conditional distribution does not belong to any standard distributions. Zhang *et al*. (2006) proposed an MH algorithm to sample R. Since it is problematic to specify a prior for R, they included artificial parameters to facilitate

the specification of the prior for *R*. Thereby, the model itself does not change and thus remains identifiable. In the following discussion, we use their notation PX-MH to denote this algorithm for the identifiable model.

Motivated by the possibility that the convergence and mixing of the MCMC sampler based on the non-identifiable models may surpass those based on the identifiable models (Liu and Wu, 1999; MacEachern, 2007), we propose a non-identifiable multivariate probit model and develop the corresponding MCMC sampling algorithms in the following section.

Non-Identifiable Multivariate Probit Models and the MCMC Sampling Algorithms

The identifiable multivariate probit model described in Section "Multivariate Probit Model for Longitudinal Ordinal Data" assumes that Z_i, the underlying multivariate normal variable, follows $N_k(X_i\beta, R)$, i.e., $Z_i \sim N_k(X_i\beta, R)$, $i = 1, \ldots, n$. To construct the non-identification in the model, we assume $Z_i \sim N_k(D^{-1/2}X_i\beta, R)$ instead of $Z_i \sim N_k(X_i\beta, R)$, where D is the diagonal matrix with the diagonal elements $d = (d_1, d_2, \ldots, d_k)$ and $d_j > 0$ for $j = 1, \ldots, k$. As noted, by including D, the identifiable model becomes non-identifiable. We then augment Z_i to be W_i by $W_i = D^{1/2}Z_i$ with $W_i \sim N_k(X_i\beta, D^{1/2}RD^{1/2})$. Therefore, the multivariate probit model can be defined by:

$$Y_{ij} = l \Leftrightarrow \sqrt{d_j}\gamma_{j,l-1} < W_{ij} \left(= \sqrt{d_j} Z_{ij}\right) \le \sqrt{d_j}\gamma_{j,l}$$

This is:

$$Y_{ij} = l \Leftrightarrow \zeta_{j,l-1} < W_{ij} \le \zeta_{j,l} \tag{3.1}$$

where, $\zeta_{j,l} = \sqrt{d_j}\gamma_{j,l}$ for $j = 1, \ldots k$ and $l = 1, \ldots, J_j$. It can be noted that $d_1, d_2, \ldots,$ *and* d_k are redundant parameters and render the identifiable model non-identifiable.

We denote $\Sigma = D^{\frac{1}{2}}RD^{\frac{1}{2}}$ and assume an independent prior distribution for β, ζ, and Σ, i.e., $P(\beta, \zeta, \Sigma) = P(\beta) \times P(\zeta) \times P(\Sigma)$ Then, the joint posterior density of β, ζ, Σ and W were given Y can be derived as follows:

$$P(\beta,\zeta,\Sigma W \mid Y) \propto \left(\prod_{i=1}^{n} I_i\right) \times P(\beta) \times P(\zeta) \times P(\Sigma) \times |\Sigma|^{-\frac{n}{2}}$$
$$\times \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(W_i - X_i\beta\right)^T \Sigma^{-1}\left(W_i - X_i\beta\right)\right]$$

where, $I_i = \prod_{j=1}^{k} I_{ij}$ and $I_{ij} = \sum_{t=1}^{J_j} 1_{(Y_{ij}=t)} 1_{\left(\zeta_j(t-1) < w_{ij} \le \zeta_{jt}\right)}$, indicating compatibility of the latent variable W_{ij} with the ordinal variable Y_{ij} defined in (3.1). Correspondingly, the MCMC sampling algorithm can be implemented by the following steps.

Step 3.1.1: Sampling $\beta \mid \Sigma, W, Y \sim N_k\left(\hat{\beta}, V_\beta\right)$, where $\hat{\beta}$ and V_β are defined as those in Step 2.1 by replacing Z with W and R with Σ and assuming the prior of β follows $N_p(b, C)$.

Step 3.1.2: Sampling $W_{ij}|\beta, \zeta, \Sigma, Y, W_{ik}, k \ne j$, from a truncated normal distribution constrained to lie between the two cut-points $\zeta_{j,l-1}$ and $\zeta_{j,l}$, assuming $Y_{ij} = l$.

Step 3.1.3: Sampling $\zeta_{j,l}|\beta, \Sigma, W, Y, \zeta_{j,k}, k \ne l$, from a uniform distribution:

$$U\left(\zeta_{j,l} \middle| \max\left\{\max\left[W_{ij} : Y_{ij} = l\right]\zeta_{j,l-1}\right\}, \min\left\{\min\left[W_{ij} : Y_{ij} = l+1\right], \zeta_{j,l+1}\right\}\right)$$

and assuming a non-informative prior for $\zeta_{j,l}$.

Step 3.1.4: Sampling Σ|β, ζ, W, Y from Inverse-Wishart$_k$ $\left(\sum_{i=1}^{n}\left(W_i - X_i\beta\right)\left(W_i - X_i\beta\right)^T + V, n+m+k+1\right)$ with a conjugate prior, $P(\Sigma) \sim$ Inv Wish$_k$ (V, m) with V being the scale matrix and m the degrees of freedom. Details can be found in Appendix A.

Then, the MCMC sampling framework can be formulated by the above four Gibbs sampling steps. In particular, the sampling of Σ (= $D^{1/2}RD^{1/2}$) in Step 3.1.4 is a Gibbs sampling instead of an MH sampling. Hence, we term this sampling algorithm as the parameter-expanded Gibbs sampling (PX-GS) algorithm.

Seeing the diagonal elements of D as the redundant parameters, we consider marginalizing D to improve the convergence and mixing of the PX-GS algorithm (Liu, 1994, Van Dyk, 2010). We propose the following sampling steps:

Step 3.2.1: W, D|β, ζ, R, Y
Step 3.2.1.1: D|β, ζ, R, Y
Step 3.2.1.2: W, D|β, ζ, R, D, Y
Step 3.2.2: β, ζ, R, D|W, Y
Step 3.2.2.1: ζ|β R, D, W, Y
Step 3.2.2.2: β|R, D, W, Y
Step 3.2.2.3: R, D|β, W, Y

Step 3.2.1, the joint sampling of W and D, can be followed by Step 3.2.1.1, D|β, ζ, R, Y and then followed by Step 3.2.1.2, W|β, ζ, R, D, Y, which is the same as Step 3.1.2. Notice that sampling D in Step 3.2.1.1 can have two circumstances. With VV being a diagonal matrix for the prior of Σ, the diagonal elements of D, $d = (d_1, d_2, \ldots, d_i)$, are independent and d_j follows an inverse-Gamma $\left(\alpha = \frac{m}{2}, \beta\frac{2}{V_j r_J}\right)$ with V_j being the j^{th} element of V and r_J being the j^{th} diagonal element of the inverse of RR for $j = 1,2, \ldots, k$. However, if V is not a diagonal matrix,

sampling D necessitates an MH algorithm. Detailed sampling schemes of D|β, ζ, R, Y for these two circumstances is given in Appendix B.

Step 3.2.2, the joint sampling of β, ζ, R, D|W, Y, can be implemented by Step 3.2.2.1, ζ|β, R, D, W, Y, which is the same as Step 3.1.3; Step 3.2.2.2, β|R, D, W, Y, which is the same as Step 3.1.1; and Step 3.2.2.3, R, D|β W, Y, which is the same as Step 3.1.4. With D marginalized in Step 3.2.1 and Step 3.2.2, we term this algorithm as a parameter-expanded Gibbs sampling with marginalization (PX-GSM) algorithm.

Software Package

We have implemented C programs for the proposed methods and applied them to simulation studies as well as two real data sets. The manual, the source C code, and the executable files for Windows operating system are available on GitHub (https://github.com/xzhang35kc/Bayesian/tree/master/Ordinal). The two real data sets, several simulated data sets, and the R program that produce the figures in the article can also be found on the same website.

Simulation Studies

We presented the PX-MH algorithm for the identifiable multivariate probit model in Section "Multivariate Probit Model for Longitudinal Ordinal Data" and developed the PX-GS and PX-GSM algorithms based on the proposed non-identifiable multivariate probit model in Section "Non-identifiable Multivariate Probit Models and the MCMC Sampling Algorithms". To investigate the performance of the PX-GS and PX-GSM algorithms and to compare them with the PX-MH algorithm, we carried out our investigation by generating 5-dimensional correlated ordinal data with the latent variables following the multivariate normal distribution with two covariates generated from the uniform distribution on the interval (-0.5, 0.5), the regression parameters $\beta^T = (\beta_0, \beta_1) = (2.0, 4.0)$ and the covariance matrix being the first-order autoregressive AR1(0.5). The ordered categories for the ordinal data were assumed to be 4, with the cut-points at 1 and 2 (the first cut-point is fixed at 0, described in Section "Multivariate Probit Model for Longitudinal Ordinal Data"). We considered sample sizes of 50 and 500 and generated 50 data sets for each investigated scenario and ran each algorithm with 10,000 iterations for each data set. The MCMC convergence diagnostics were conducted using the R package-boa by Smith (2007).

We considered two sets of priors. The first prior set, denoted by N-ID, assumes the non-informative priors for β and ζ, i.e., $P(\beta) \propto 1$ and $P(\zeta) \propto 1$ and an Inverse Wishart prior for Σ, i.e., $P(\Sigma)$~Inv Wish$_5$ ($V = I_5$, $m = 20$), where I_5 is the 5 × 5 identity matrix and assumes the correlations to be 0. The second prior set, denoted by I-CS, assumes an informative prior of β, i.e., $\beta \sim N_2(b, I_2)$, where $b = (3.0, 3.0)$ and I_2 is the 2 × 2 identity matrix, a non-informative prior for ζ, i.e., $P(\zeta) \propto 1$ and an Inverse-Wishart prior for Σ, i.e., $P(\Sigma)$~InvWish$_5$ (V = CS (0.4), m = 20), where CS (0.4) denotes the compound symmetry covariance structure, with the equal correlation being 0.4.

Table 1 presents the averaged posterior means and standard deviations for the regression parameters and the cut-points and Table 2 presents those estimated quantities for the correlation parameters based on 50 data sets under two prior scenarios. For the estimation of the regression parameters and the correlations, the I-CS prior brings an obvious effect in comparison with the N-ID prior for a sample size of 50. Specifically, the I-CS prior, assuming $\beta \sim N_2(b, I_2)$, produces the estimated values closer to the true values than the N-ID prior, which assumes $P(\beta) \propto 1$. Also, the I-CS prior specifies a CS (0.4) for the correlation matrix and thus gives larger and more precise estimated correlation parameters than the N-ID prior, which specifies the correlations to be 0. Although a noninformative prior is assumed for each cut-point, those two prior scenarios still affect the estimation of the cut-points, especially for the PX-MH and PX-GS algorithms.

As noted, the sample size of 500 gives estimated values closer to the true values and smaller standard deviations than the sample size of 50. The specification of the priors, N-ID and I-CS, does not affect much for the estimated parameters much, suggesting the sample size of 500 gives the estimated posterior quantities dominated by the data and is robust to the prior specification. It is noticeable that the PX-GSM algorithm has larger standard deviations for the estimated quantities using the I-CS prior rather than the N-ID prior. This is probably due to an MH sample algorithm required for the sampling of the redundant parameters in Step 3.2.1 (Appendix B, Circumstance 2) using the I-CS prior, which therefore produces larger variations than those using the N-ID prior.

Figure 1 shows the Auto Correlation Function (ACF) plots for selected parameters. We chose regression parameter β_1, the cut-point ζ_{11}, and the correlation parameters rr_{23} and rr_{15} for illustration. As can be seen, for a sample size of 50 (first row: N-ID and second row: ICS), both priors produce similar ACF plots; the PX-GS and PX-GSM algorithms produce almost indistinguishable ACF plots, especially for the correlations; and the PX-MH algorithm has larger ACF values than the PX-GS and PX-GSM algorithms, especially for the correlations. For a sample size of 500 (third row: N-ID and fourth row: I-CS), the ACF plots for using these three methods are distinguished. First, we can see that for the regression parameter β_1 and the cut-point ζ_{11}, the ACF plots are similar under both the N-ID and I-CS priors, while clearly, the PX-GSM algorithm outperforms the PX-MH and PX-GS algorithms with

much faster-decreased ACF values and the PX-MH algorithm outperforms the PX-GS algorithm. For the correlation parameters, the ACF plots of the PX-GS algorithm are separated from those of the PX-GSM algorithm. The N-ID prior favors the correlation rr_{15} with the true value being 0.0625, while the I-CS prior favors the correlation r_{23} with the true value being 0.5. The plots show that the PX-GS algorithm produces faster-decreased ACF values for r_{15} using the N-ID prior compared with that of using the I-CS prior, while it produces much faster-decreased ACF values for r_{23} using the I-CS prior compared with that of using the N-ID prior. In comparison with the PX-GS algorithm, the PX-GSM algorithm is robust to the prior specification. Like those for a sample size of 50, the PX-MH still has the largest ACF values among these algorithms for a sample of 500.

Real Data Applications

The Pain Score Data

Our first application is the study of abdominal suction to reduce shoulder tip pain after laparoscopic surgery (Jorgensen, 1995; Lumley, 1996). There was a total of 41 patients randomized to suction or no suction on the drain. For each patient, the shoulder tip pain score, an ordinal measurement with 5 categories from 1 to 5, was longitudinally collected after surgery from mornings and afternoons for three days. The age and gender of each patient were also collected. We used our proposed methods to analyze this data with the outcome being a 6-dimensional ordinal variable with 4 categories by combining the last two categories and the covariates being the visit time, gender (1: Male; 0: Female), treatment (1: Treatment; 0: Control) and age with the intercept term.

We investigated two prior scenarios: N-ID and ICS. The N-ID assumes $P(\beta) \propto 1$, $P(\zeta) \propto 1$ and $P(\Sigma)\sim\text{InvWish}_5(V = I_6, m = 30)$; the I-CS assumes $\beta\sim N_5(b, I_5)$, where $b = (0, 0, 0, -1, 0)$, which is close to the estimated quantities, $P(\zeta) \propto 1$ and $P(\Sigma)\sim\text{InvWish}_5(V = \text{CS}(0.5), mm = 30)$. We ran each algorithm with 10,000 iterations.

Table 3 presents the 95% credible intervals for the regression parameters and the posterior means and standard deviations for each cut-point as well. As shown, the treatment effect is significant for all three methods under both priors, suggesting that abdominal suction helps to reduce shoulder tip pain after laparoscopic surgery. Noticeably, the age effect is significant for all three methods under the N-ID prior, but not with the I-CS prior. Also, the I-CS prior produces smaller estimated cut-points with smaller standard deviations than the N-ID prior does.

Then we investigated two additional models: One with gender, treatment, and age by excluding visit time and the other with treatment and age by excluding both visit time and gender. Then we calculated Bayesian Information Criterion (BIC) (Grigorova and Gueorguieva, 2016) for these three models and the results are presented in Table 4. As can be seen, for all three models, the BIC values for the PX-GS and PXGSM algorithms are similar and both have smaller values than those for the PX-MH algorithm, suggesting the PX-MH algorithm is inferior to the PX-GS and PXGSM algorithms. By excluding visit time and/or gender (Model 1, Model 2), all BIC values for each algorithm are improved, suggesting visit time and gender can be dropped from the modeling.

Table 5 contains the estimated posterior means and standard deviations for all the correlations under both priors. We can see that the I-CS prior gives a larger estimated value with smaller standard deviations in comparison with the N-ID prior. This implies that with a sample size of 41, the prior specification plays a part in the posterior estimation of unknown parameters, as addressed in the Section "Simulation Studies" for simulation studies with a sample size of 50.

The Schizophrenia Study

We then applied the proposed methods to the National Institute of Mental Health Schizophrenia Collaborative Study (Hedeker and Gibbons, 1994ab). There were 437 patients randomized to receive either a placebo or one of three different anti-psychotic drugs (chlorpromazine, fluphenazine, or thioridazine). Each patient was assessed at weeks 0, 1, 3, and 6. The original primary outcome was item 79 on the Inpatient Multidimensional Psychiatric Scale, which indicates the severity of illness using 7 ordinal scales. Archer *et al.* (2015) used a mixed-effects logistic regression model to analyze this data by combining the 7 ordinal scales to be 4 ordinal scales to ensure enough observations for each category and choosing the treatment group, square root of the time, and the interaction of the treatment group and the square root of the time as the covariates.

With the same outcome and covariates as Archer *et al.* (2015), we conducted our analysis using the PX-MH, PXGS, and PX-GSM algorithms. Since the sample size was 437, based on the simulation studies for a sample size of 500 in Section "Simulation Studies", the prior specification may not affect much for the estimated posterior means and standard deviations of the unknown parameters. Therefore, we chose the non-informative priors for the regression parameters and the cut-points and an Inverse-Wishart prior for the correlation matrix with the scale matrix being the identity matrix and 20 degrees of freedom. We ran each algorithm with 20,000 iterations with the first 5,000 as the burn-in. Table 6 presents the 95% credible intervals for the regression parameters and the posterior estimated

means and standard deviations for the cut-points and correlations. It can be seen that all three methods illustrate the significant effect of the interaction of the treatment and the square root of the time; both the PX-MH and PX-GSM algorithms show that the time effect is significantly negative, while the PX-GS algorithm does not. The estimated standard deviations of the cut points using the PX-GSM algorithm are larger than those of the other two algorithms. This may be caused by the slow convergences and mixings for the PXMH and PX-GS algorithms in comparison with the PXGSM algorithm for the cut points (Fig. 1 for a sample size of 500).

Table 1: Averaged posterior means and standard deviations for the regression parameters and cut-points based on 50 simulated datasets under two prior scenarios

		PX-MH				PX-GS				PX-GSM			
		n = 50		n = 500		n = 50		n = 500		n = 50		n = 500	
Parameters	True	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS
β_{11}	2.0	2.15 (0.30)	2.12 (0.28)	2.02 (0.09)	2.02 (0.09)	2.19 (0.41)	2.10 (0.37)	2.01 (0.17)	1.99 (0.17)	2.19 (0.39)	2.08 (0.47)	2.09 (0.24)	2.03 (0.33)
β_{22}	4.0	4.28 (0.37)	4.09 (0.34)	4.06 (0.11)	4.02 (0.11)	4.34 (0.64)	4.00 (0.57)	4.04 (0.32)	3.99 (0.32)	4.34 (0.59)	3.98 (0.79)	4.19 (0.47)	4.03 (0.64)
ζ_{11}	1.0	1.25 (0.24)	1.24 (0.24)	1.19 (0.07)	1.18 (0.07)	1.43 (0.38)	1.28 (0.32)	1.15 (0.19)	1.15 (0.19)	1.34 (0.31)	1.39 (0.39)	1.16 (0.19)	1.15 (0.23)
ζ_{12}	2.0	2.02 (0.30)	1.99 (0.29)	2.20 (0.10)	2.19 (0.10)	2.37 (0.54)	2.10 (0.43)	2.12 (0.28)	2.12 (0.28)	2.23 (0.43)	2.24 (0.55)	2.18 (0.31)	2.14 (0.40)
ζ_{21}	1.0	0.96 (0.19)	0.94 (0.20)	1.12 (0.09)	1.11 (0.09)	1.16 (0.33)	1.01 (0.27)	1.09 (0.17)	1.09 (0.17)	1.04 (0.26)	1.02 (0.29)	1.04 (0.18)	1.07 (0.21)
ζ_{22}	2.0	1.94 (0.27)	1.91 (0.26)	2.21 (0.11)	2.20 (0.11)	2.39 (0.57)	2.09 (0.44)	2.18 (0.28)	2.18 (0.28)	2.12 (0.41)	2.18 (0.53)	2.11 (0.29)	2.13 (0.37)
ζ_{31}	1.0	1.22 (0.22)	1.17 (0.21)	1.05 (0.08)	1.04 (0.07)	1.63 (0.44)	1.39 (0.34)	0.99 (0.17)	0.99 (0.16)	1.36 (0.30)	1.39 (0.37)	0.98 (0.17)	0.96 (0.19)
ζ_{32}	2.0	2.22 (0.31)	2.14 (0.29)	2.13 (0.11)	2.11 (0.11)	2.91 (0.70)	2.48 (0.53)	2.01 (0.27)	2.01 (0.26)	2.47 (0.48)	2.51 (0.62)	2.01 (0.29)	2.00 (0.34)
ζ_{41}	1.0	1.24 (0.24)	1.21 (0.22)	1.05 (0.09)	1.05 (0.09)	1.38 (0.38)	1.21 (0.30)	1.00 (0.16)	1.00 (0.16)	1.28 (0.31)	1.37 (0.38)	0.96 (0.16)	0.98 (0.18)
ζ_{42}	2.0	2.38 (0.31)	2.28 (0.29)	2.08 (0.10)	2.07 (0.10)	2.61 (0.59)	2.30 (0.46)	2.01 (0.27)	2.01 (0.26)	2.49 (0.47)	2.62 (0.62)	1.95 (0.28)	1.99 (0.33)
ζ_{51}	1.0	0.99 (0.21)	0.93 (0.20)	1.00 (0.07)	0.99 (0.07)	1.12 (0.31)	0.96 (0.25)	0.93 (0.15)	0.93 (0.15)	1.06 (0.26)	1.05 (0.30)	0.97 (0.16)	0.96 (0.19)
ζ_{52}	2.0	1.87 (0.26)	1.79 (0.26)	2.04 (0.10)	2.02 (0.10)	2.10 (0.48)	1.83 (0.38)	1.91 (0.26)	1.91 (0.26)	2.02 (0.38)	2.02 (0.48)	2.02 (0.28)	1.98 (0.35)

Table 2: Averaged posterior means and standard deviations for the correlation parameters based on 50 simulated datasets under two prior scenarios

		PX-MH				PX-GS				PX-GSM			
		n = 50		n = 500		n = 50		n = 500		n = 50		n = 500	
Parameters	True	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS
r_{12}	0.5000	0.32 (0.17)	0.42 (.16)	0.47 (0.06)	0.49 (0.06)	0.30 (0.16)	0.43 (0.14)	0.45 (0.07)	0.48 (0.06)	0.23 (0.14)	0.40 (0.16)	0.47 (0.06)	0.49 (0.07)
r_{13}	0.2500	0.16 (0.17)	0.27 (.17)	0.23 (0.06)	0.24 (0.07)	0.18 (0.17)	0.32 (0.15)	0.23 (0.07)	0.26 (0.06)	0.14 (0.15)	0.26 (0.17)	0.24 (0.07)	0.26 (0.07)
r_{14}	0.1250	0.08 (0.17)	0.20 (.18)	0.10 (0.07)	0.11 (0.07)	0.11 (0.17)	0.25 (0.15)	0.11 (0.07)	0.13 (0.07)	0.08 (0.15)	0.18 (0.18)	0.11 (0.07)	0.12 (0.07)
r_{15}	0.0625	0.00 (0.18)	0.11 (.19)	0.05 (0.07)	0.06 (0.07)	0.02 (0.17)	0.18 (0.16)	0.06 (0.07)	0.09 (0.07)	0.02 (0.15)	0.09 (0.18)	0.06 (0.07)	0.07 (0.07)
r_{23}	0.5000	0.38 (0.16)	0.48 (.15)	0.46 (0.06)	0.48 (0.06)	0.35 (0.15)	0.46 (0.13)	0.44 (0.07)	0.47 (0.06)	0.27 (0.14)	0.45 (0.15)	0.45 (0.06)	0.48 (0.07)
r_{24}	0.2500	0.17 (0.17)	0.30 (.17)	0.23 (0.06)	0.25 (0.07)	0.21 (0.16)	0.33 (0.15)	0.24 (0.06)	0.26 (0.06)	0.16 (0.15)	0.30 (0.17)	0.24 (0.07)	0.26 (0.07)
r_{25}	0.1250	0.08 (0.17)	0.20 (.18)	0.12 (0.07)	0.14 (0.07)	0.11 (0.17)	0.25 (0.16)	0.13 (0.07)	0.16 (0.07)	0.08 (0.15)	0.18 (0.18)	0.13 (0.07)	0.15 (0.07)
r_{34}	0.5000	0.36 (0.16)	0.47 (.15)	0.48 (0.06)	0.50 (0.06)	0.34 (0.15)	0.45 (0.13)	0.46 (0.07)	0.49 (0.06)	0.26 (0.14)	0.45 (0.15)	0.47 (0.06)	0.50 (0.07)
r_{35}	0.2500	0.13 (0.17)	0.25 (.17)	0.24 (0.06)	0.26 (0.07)	0.16 (0.17)	0.30 (0.15)	0.24 (0.06)	0.27 (0.06)	0.12 (0.15)	0.25 (0.17)	0.25 (0.07)	0.27 (0.07)
r_{45}	0.5000	0.32 (0.17)	0.41 (.16)	0.45 (0.06)	0.47 (0.06)	0.29 (0.16)	0.42 (0.14)	0.43 (0.07)	0.46 (0.06)	0.22 (0.14)	0.41 (0.16)	0.45 (0.06)	0.48 (0.07)

Table 3: Posterior 95% credible intervals of the regression parameters and posterior means and standard deviations for each cut-point using the pain score data

	PX-MH		PX-GS		PX-GSM	
Parameters	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS
Visit time	(-0.18, 0.02)	(-0.17, 0.02)	(-0.19, 0.01)	(-0.17, 0.02)	(-0.21, 0.01)	(-0.17, 0.03)
Gender	(-0.46, 0.53)	(-0.41, 0.63)	(-0.46, 0.53)	(-0.33, 0.65)	(-0.48, 0.60)	(-0.35, 0.69)
Treatment	(-1.53, -0.58)	(-1.44, -0.44)	(-1.49, -0.51)	(-1.40, -0.41)	(-1.53, -0.43)	(-1.45, -0.35)
Age	(-0.04, -0.01)	(-0.03, 0.00)	(-0.04, -0.003)	(-0.03, 0.004)	(-0.04, -0.01)	(-0.03, 0.003)
ζ_{11}	0.57 (0.18)	0.54 (0.17)	0.57 (0.18)	0.52 (0.16)	0.54 (0.17)	0.49 (0.16)
ζ_{12}	1.32 (0.26)	1.26 (0.25)	1.34 (0.25)	1.25 (0.24)	1.27 (0.26)	1.19 (0.24)
ζ_{21}	0.66 (0.17)	0.62 (0.16)	0.64 (0.17)	0.60 (0.16)	0.63 (0.18)	0.58 (0.16)
ζ_{22}	1.26 (0.22)	1.18 (0.21)	1.24 (0.23)	1.16 (0.21)	1.23 (0.24)	1.15 (0.22)
ζ_{31}	0.67 (0.18)	0.62 (0.17)	0.60 (0.19)	0.57 (0.18)	0.59 (0.19)	0.57 (0.18)
ζ_{32}	1.33 (0.21)	1.22 (0.20)	1.31 (0.24)	1.23 (0.22)	1.28 (0.25)	1.22 (0.23)
ζ_{41}	0.51 (0.15)	0.46 (0.14)	0.52 (0.17)	0.48 (0.15)	0.51 (0.17)	0.47 (0.15)
ζ_{42}	1.06 (0.20)	0.96 (0.18)	1.09 (0.22)	1.00 (0.21)	1.06 (0.23)	0.99 (0.21)
ζ_{51}	0.73 (0.21)	0.68 (0.19)	0.73 (0.25)	0.68 (0.21)	0.70 (0.22)	0.67 (0.21)
ζ_{52}	1.59 (0.29)	1.48 (0.27)	1.59 (0.32)	1.48 (0.29)	1.54 (0.33)	1.47 (0.30)
ζ_{61}	1.08 (0.26)	1.02 (0.25)	1.05 (0.26)	1.01 (0.25)	1.01 (0.26)	0.99 (0.25)
ζ_{62}	2.66 (0.54)	2.47 (0.49)	2.57 (0.52)	2.43 (0.48)	2.51 (0.53)	2.40 (0.48)

Table 4: Posterior means and standard deviations of correlation parameters using the pain score data

	PX-MH		PX-GS		PX-GSM	
Parameters	N-ID	I-CS	N-ID	I-CS	N-ID	I-CS
r_{12}	0.40 (0.13)	0.49 (0.11)	0.47 (0.13)	0.55 (0.11)	0.42 (0.13)	0.53 (0.11)
r_{13}	0.36 (0.15)	0.49 (0.13)	0.45 (0.13)	0.54 (0.11)	0.40 (0.13)	0.52 (0.11)
r_{14}	0.27 (0.15)	0.44 (0.13)	0.40 (0.14)	0.50 (0.12)	0.34 (0.14)	0.47 (0.12)
r_{15}	0.06 (0.15)	0.26 (0.15)	0.23 (0.15)	0.36 (0.13)	0.19 (0.15)	0.35 (0.14)
r_{16}	0.24 (0.15)	0.38 (0.15)	0.31 (0.15)	0.42 (0.13)	0.26 (0.15)	0.41 (0.13)
r_{23}	0.72 (0.11)	0.78 (0.10)	0.65 (0.10)	0.71 (0.08)	0.65 (0.10)	0.71 (0.08)
r_{24}	0.57 (0.12)	0.68 (0.12)	0.57 (0.11)	0.64 (0.10)	0.57 (0.12)	0.64 (0.10)
r_{25}	0.23 (0.15)	0.43 (0.13)	0.34 (0.15)	0.45 (0.13)	0.35 (0.15)	0.46 (0.13)
r_{26}	0.08 (0.15)	0.32 (0.14)	0.28 (0.15)	0.39 (0.13)	0.28 (0.15)	0.40 (0.13)
r_{34}	0.66 (0.14)	0.74 (0.12)	0.58 (0.11)	0.66 (0.09)	0.60 (0.11)	0.67 (0.09)
r_{35}	0.32 (0.14)	0.51 (0.13)	0.38 (0.14)	0.50 (0.12)	0.41 (0.14)	0.52 (0.12)
r_{36}	0.20 (0.16)	0.43 (0.13)	0.34 (0.14)	0.46 (0.12)	0.35 (0.15)	0.47 (0.13)
r_{45}	0.48 (0.14)	0.63 (0.12)	0.45 (0.13)	0.57 (0.11)	0.48 (0.13)	0.59 (0.11)
r_{46}	0.45 (0.13)	0.59 (0.12)	0.45 (0.14)	0.57 (0.11)	0.45 (0.13)	0.57 (0.11)
r_{56}	0.74 (0.11)	0.78 (0.10)	0.59 (0.11)	0.70 (0.09)	0.58 (0.11)	0.70 (0.09)

Table 5: BIC for pain score data for model selection

Models	PX-MH	PX-GS	PX-GSM
Model 1: Treatment, Age	518.61	515.35	515.69
Model 2: Gender, Treatment, Age	524.57	518.68	518.34
Model 3: Visit time, Gender, Treatment, Age	526.85	520.01	521.16

Table 6: Posterior 95% credible intervals of the regression parameters and posterior means and standard deviations of the correlation parameters and cut-points using the schizophrenia study

Parameters	PX-MH	PX-GS	PX-GSM
Treatment	(-0.27, 0.23)	(-0.31, 0.31)	(-0.31, 0.32)
$\sqrt{Time}$	(-0.54, -0.24)	(-0.50, 0.02)	(-0.87, -0.45)
Treatment × $\sqrt{Time}$	(-0.59, -0.31)	(-0.89, -0.40)	(-0.75, -0.31)
r_{12}	0.55(0.04)	0.54(0.05)	0.55(0.05)
r_{13}	0.39(0.06)	0.40(0.06)	0.40(0.06)
r_{14}	0.18(0.07)	0.21(0.07)	0.22(0.07)
r_{23}	0.65(0.04)	0.65(0.04)	0.67(0.04)
r_{24}	0.52(0.05)	0.53(0.05)	0.55(0.05)
r_{34}	0.66(0.04)	0.67(0.04)	0.68(0.04)
ζ_{11}	1.23(0.10)	1.48(0.28)	2.05(0.28)
ζ_{12}	2.12(0.10)	2.37(0.27)	2.95(0.38)
ζ_{21}	1.28(0.07)	1.22(0.08)	1.33(0.16)
ζ_{22}	2.07(0.07)	2.00(0.11)	2.11(0.23)
ζ_{31}	1.07(0.06)	1.02(0.07)	1.03(0.13)
ζ_{32}	2.01(0.07)	1.95(0.10)	1.93(0.22)
ζ_{41}	1.20(0.07)	1.19(0.08)	1.13(0.14)
ζ_{42}	1.80(0.09)	1.8(0.10)	1.74(0.20)

Figure 2 contains the trace plots for selected parameters: The regression parameter, β_3, the square root of the time, the correlation r_{14} being the smallest correlation parameter, the correlation.

r_{34} being the largest correlation parameter and two cut-points ζ_1 and ζ_{42}. As can be seen from the trace plots of β_3, the PX-MH algorithm reaches stabilization faster than the PX-GS and PX-GSM algorithms; the PX-GSM algorithm reaches stabilization after 5,000 iterations, while the PX-GS seems not to do so. This may cause the PX-GS algorithm to fail to show the significant time effect in Table 5. The trace plots of the correlations rr_{14} and rr_{34} for the PX-GS and PX-GSM algorithms are similar and show better mixing than the PX-MH algorithm. The trace plots for the cut-point ζ_1 show that the PX-GSM algorithm reaches stabilization after 5,000 iterations, while the PX-MH and PX-GS algorithms do not. For the cut-point ζ_{42}, the PX-GS and PX-GSM algorithms exceed the PX-MH algorithm, with the PX-GSM algorithm still having the best mixing sampler among the three algorithms.

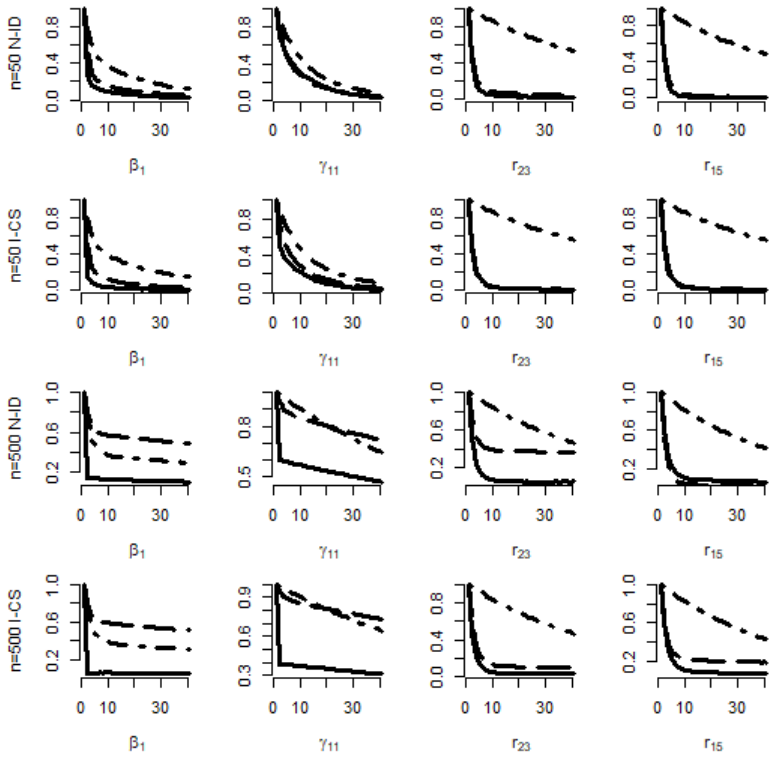

Fig. 1: ACF plots of selected parameters for simulation studies with sample sizes being 50 and 500 under the N-ID and I-CS priors. The solid line: is PX-GSM; the long-dashed: is PX-GS; the dot-dashed: PX-MH

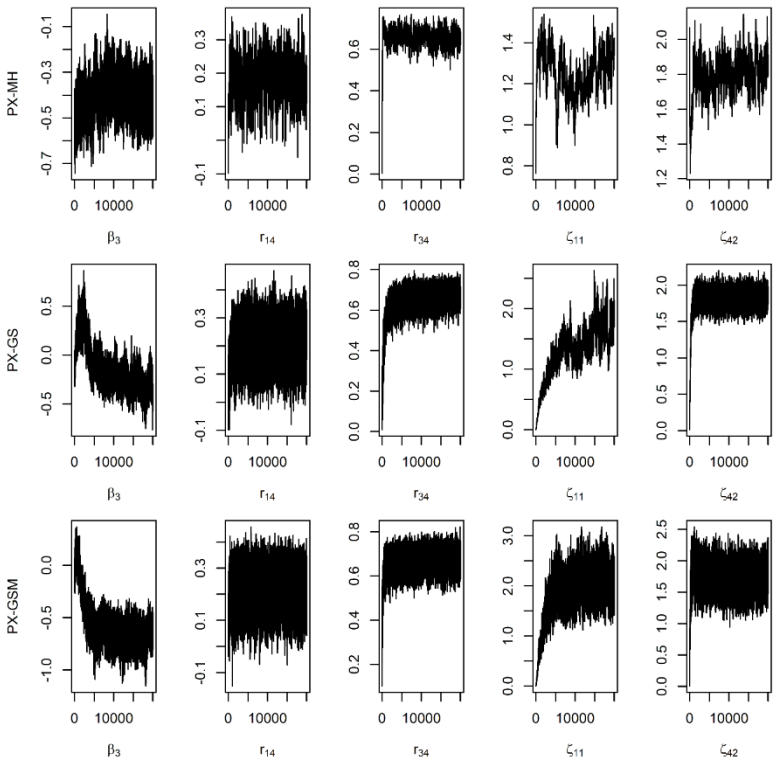

Fig. 2: Trace plots of selected parameters for the schizophrenia study using PX-MH, PX-GS, and PX-GSM algorithms

Results and Discussion

In this article, we proposed a non-identifiable multivariate probit model to analyze longitudinal ordinal data and developed the PX-GS and PX-GSM algorithms, with the PX-GSM algorithm marginalizing the redundant parameters. We conducted simulation studies and two real data applications to investigate and compare the PX-GS and PX-GSM algorithms with the PX-MH algorithm, which is based on the identifiable multivariate probit model. The PX-GSM algorithm outperforms the PX-GS and PX-MH algorithms in the convergence and mixing of the regression parameters, the cut-points, and correlations.

For data with small sample sizes, such as 50, the ACF plots for the PX-GS and PX-GSM algorithms are almost indistinguishable and have much faster-decreased ACF values than the PX-MH algorithm does, especially for correlations. This indicates that the PX-GS and PX-GSM algorithms based on a non-identifiable model outperform the PX-MH algorithm based on the identifiable model. For data with a large sample size, such as 500, the PX-GS algorithm is affected by the prior specification, while the PX-GSM algorithm is robust to the priors for correlations.

It is also shown that the ACF values (β_1, γ_{11}, r_{23}) of the PX-GSM and PX-MH algorithms decrease faster than the PX-GS algorithm. In the schizophrenia study with 437 patients, the trace plots of β_3 and γ_{11} illustrate that the PX-GS algorithm does not converge after 20,000 iterations while the PX-MH and PX-GSM algorithms do. This explains the PX-MH and PX-GSM algorithms show a significant time effect, while the PX-GS fails. This suggests one possibility that the redundant parameters of the PX-GS algorithm may converge slower for data with a large sample size than for the data with small sample size, thus leading to the slow convergence of the identifiable parameters, such as the regression parameters, cut-points, and correlations. In this circumstance, the PX-GS algorithm may be inferior to the PX-MH algorithm and this issue is worth further investigation.

We also noticed that for the PX-GS algorithm in the schizophrenia study, the mixing of the trace plot for the cut-point $\zeta\zeta_{11}$ was not as good as that for the PX-GSM algorithm, while the plot for the cut-point $\zeta\zeta_{42}$ seemed similar to that of the PX-GSM algorithm. In this article, we specified only the non-informative priors for the cutpoints. This suggests that further investigation regarding the convergence of the cut-points for the PX-GS algorithm and possible informative priors may be considered in our future research work.

Conclusion

Our investigation illustrates that constructing nonidentifiable models may improve the convergence of the MCMC sampling components compared with the identifiable models. The marginalization of the redundant parameters in the non-identifiable models should be considered in developing efficient MCMC sampling algorithms, especially for data with large sample sizes. Due to the improved convergence of the correlation parameters, applying the PX-GSM algorithm to large and high-dimensional ordinal data may become one of our future investigations.

Acknowledgment

I would like to thank the Editors and reviewers for their valuable comments and suggestions that improve the quality of the paper significantly.

Ethics

This article is original and contains unpublished materials. The corresponding author has read and approved the manuscript and no ethical issues are involved.

References

Agresti, A. (2003). *Categorical data analysis*. John Wiley & Sons. https://doi.org/10.1002/0471249688

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669-679. https://doi.org/10.1080/01621459.1993.10476321

Archer, K. J., Hedeker, D., Nordgren, R., & Gibbons, R. D. (2015). mixor: An R package for longitudinal and clustered ordinal response modeling. https://cran.microsoft.com/snapshot/2015-09-16/web/packages/mixor/vignettes/mixor.pdf

Ashford, J. R., & Sowden, R. R. (1970). Multi-variate probit analysis. *Biometrics*, 535-546. https://doi.org/10.2307/2529107

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian analysis*, *1*(3), 473-514. https://doi.org/10.1214/06-BA117

Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*(2), 347-361. https://doi.org/10.1093/biomet/85.2.347

da Silva, J. L., Colosimo, E. A., & Demarqui, F. N. (2019). A general GEE framework for the analysis of longitudinal ordinal missing data and related issues. *Statistical Modelling*, *19*(2), 174-193. https://doi.org/10.1177/1471082X17752753

Ditlhong, K. E., Ngesa, O. O., & Kombo, A. Y. (2018). A comparative analysis of generalized estimating equations methods for incomplete longitudinal ordinal data with ignorable dropouts. *Open Journal of Statistics*, *8*(05), 770. https://doi.org/10.4236/ojs.2018.85051

Drasgow, F. (2014). Polychoric and polyserial correlations. *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.stat02493

Edwards, Y. D., & Allenby, G. M. (2003). Multivariate analysis of multiple response data. *Journal of Marketing research*, *40*(3), 321-334. https://doi.org/10.1509/jmkr.40.3.321.19233

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398-409. https://doi.org/10.1080/01621459.1990.10476213

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429258411

Gilks, W. R., Richardson, S., & Spieglhalter, D. J. (1996). Markov Chain Monte Carlo in Practice, 520 pp. https://doi.org/10.1201/b14835

Grigorova, D., & Gueorguieva, R. (2016). Correlated probit analysis of repeatedly measured ordinal and continuous outcomes with application to the health and retirement study. *Statistics in Medicine*, *35*(23), 4202-4225. https://doi.org/10.1002/sim.6982

Grilli, L., & Rampichini, C. (2011). Multilevel models for ordinal data. *Modern analysis of customer surveys: With applications using R*, 391-411. https://doi.org/10.1002/9781119961154.ch19

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, *33*, 1-22. https://doi.org/10.18637/jss.v033.i02

Hedeker, D. (2015). Methods for multilevel ordinal data in prevention research. *Prevention Science*, *16*(7), 997-1006. https://doi.org/10.1007/s11121-014-0495-x

Hedeker, D., & Gibbons, R. D. (1994b). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 933-944. https://doi.org/10.2307/2533433

Hedeker, D. & Gibbons, R. D. (1994a). Application of random-effects probit regression models. *Journal of consulting and clinical psychology*, *62*(2), 285. https://doi.org/10.1037/0022-006X.62.2.285

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Wiley-Interscience.

Hirk, R., Hornik, K., & Vana, L. (2019). Multivariate ordinal regression models: An analysis of corporate credit ratings. *Statistical Methods & Applications*, *28*(3), 507-539. https://doi.org/10.1007/s10260-018-00437-7

Imai, K., & Van Dyk, D. A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, *124*(2), 311-334. https://doi.org/10.1016/j.jeconom.2004.02.002

Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika*, *68*(4), 563-583. https://doi.org/10.1007/BF02295612

Jorgensen, J. O., Gillies, R. B., Hunt, D. R., Caplehorn, J. R., & Lumley, T. (1995). A simple and effective way to reduce postoperative pain after laparoscopic cholecystectomy. *Australian and New Zealand Journal of surgery*, *65*(7), 466-469. https://doi.org/10.1111/j.1445-2197.1995.tb01787.x

Lawrence, E., Bingham, D., Liu, C., & Nair, V. N. (2008). Bayesian inference for multivariate ordinal data using parameter expansion. *Technometrics*, *50*(2), 182-191. https://doi.org/10.1198/004017008000000064

Lee, K., & Daniels, M. J. (2008). Marginalized models for longitudinal ordinal data with application to quality of life studies. *Statistics in Medicine*, *27*(21), 4359-4380. https://doi.org/10.1002/sim.3352

Lee, K., Sohn, I., & Kim, D. (2016). Analysis of long series of longitudinal ordinal data using marginalized models. *Computational Statistics & Data Analysis*, *94*, 363-371. https://doi.org/10.1016/j.csda.2015.07.010

Li, Y., & Schafer, D. W. (2008). Likelihood analysis of the multivariate ordinal probit regression model for repeated ordinal responses. *Computational Statistics & Data Analysis*, *52*(7), 3474-3492. https://doi.org/10.1016/j.csda.2007.10.025

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22. https://doi.org/10.1093/biomet/73.1.13

Liu, C. (2001). [The Art of Data Augmentation]: Discussion. *Journal of Computational and Graphical Statistics*, *10*(1), 75-81. https://doi.org/10.1198/10618600152418746

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, *89*(427), 958-966. https://doi.org/10.1080/01621459.1994.10476829

Liu, J. S., & Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, *94*(448), 1264-1274. https://doi.org/10.1080/01621459.1999.10473879

Liu, X., & Daniels, M. J. (2006). A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *Journal of Computational and Graphical Statistics*, *15*(4), 897-914. https://doi.org/10.1198/106186006X160681

Lumley, T. (1996). Generalized estimating equations for ordinal data: A note on working correlation structures. *Biometrics*, 354-361. https://doi.org/10.2307/2533173

MacEachern, S. N. (2007). Comment on the article by Jain and Neal. *Bayesian Analysis*, *2*(3), 483-494. https://doi.org/10.1214/07-BA219C

McCulloch, R. E., Polson, N. G., & Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, *99*(1), 173-193. https://doi.org/10.1016/S0304-4076(00)00034-8

McCulloch, R., & Rossi, P. E. (1994). Exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*(1-2), 207-240. https://doi.org/10.1016/0304-4076(94)90064-7

Molenberghs, G., & Verbeke, G. (2005). Models for discrete longitudinal data. Spring Science+ Business Media. *Inc., New York*.

Nobile, A. (1998). A hybrid Markov chain for the Bayesian analysis of the multinomial probit model. *Statistics and Computing*, *8*(3), 229-242. https://doi.org/10.1023/A:1008905311214

Nooraee, N., Molenberghs, G., & van den Heuvel, E. R. (2014). GEE for longitudinal ordinal data: Comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. *Computational Statistics & Data Analysis*, *77*, 70-83. https://doi.org/10.1016/j.csda.2014.03.009

O'brien, S. M., & Dunson, D. B. (2004). Bayesian multivariate logistic regression. *Biometrics*, *60*(3), 739-746. https://doi.org/10.1111/j.0006-341X.2004.00224.x

Pagui, K., Clovis, E., & Canale, A., (2015). Pairwise likelihood inference for multivariate ordinal responses with applications to customer satisfaction. *Applied Stochastic Models in Business and Industry*, 32 (2), 273-282. https://doi.org/10.1002/asmb.2147

Parsons, N. R., Edmondson, R. N., & Gilmour, S. G. (2006). A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *55*(4), 507-524. https://doi.org/10.1111/j.1467-9876.2006.00550.x

Pearson, K. (1900). I. Mathematical contributions to the theory of evolution.—VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *195*(262-273), 1-47. https://doi.org/10.1098/rsta.1900.0022

Schildcrout, J. S., Harrell Jr, F. E., Heagerty, P. J., Haneuse, S., Di Gravio, C., Garbett, S. P., ... & Shepherd, B. E. (2022). Model-assisted analyses of longitudinal, ordinal outcomes with absorbing states. *Statistics in Medicine*. https://doi.org/10.1002/sim.9366

Smith, B. J. (2007). boa: An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*, *21*, 1-37. https://doi.org/10.18637/jss.v021.i11

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528-540. https://doi.org/10.1080/01621459.1987.10478458

Touloumis, A., Agresti, A., & Kateri, M. (2013). GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, *69*(3), 633-640. https://doi.org/10.1111/biom.12054

Ursino, M., & Gasparini, M. (2018). A new parsimonious model for ordinal longitudinal data with application to subjective evaluations of gastrointestinal disease. *Statistical Methods in Medical Research*, *27*(5), 1376-1393. https://doi.org/10.1177/0962280216661370

Varin, C., & Czado, C. (2010). A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, *11*(1), 127-138. https://doi.org/10.1093/biostatistics/kxp042

Zhang, X. (2020). Parameter-expanded data augmentation for analyzing correlated binary data using multivariate probit models. *Statistics in Medicine*, *39*(25), 3637-3652. https://doi.org/10.1002/sim.8685

Zhang, X., Boscardin, W. J., & Belin, T. R. (2006). Sampling correlation matrices in Bayesian models with correlated latent variables. *Journal of Computational and Graphical Statistics*, *15*(4), 880-896. https://doi.org/10.1198/106186006X160050

Appendix A. Details of Sampling Σ|β, ζ, W, Y

In this Appendix, we give details of sampling $\Sigma|\ \beta, \zeta, W, Y$ in step3.1.4 for the PX-GS algorithm.

We assume $P(\Sigma)\sim$ Inv Wish (m, V), i.e.:

$$P(\Sigma)=\frac{1}{2^{mk/2}}\frac{1}{\Gamma_p\left(\frac{m}{2}\right)}|V|^{m/2}|\Sigma|-\frac{m+k+1}{2}\exp\left[-\frac{1}{2}tr\left(V\sum{}^{-1}\right)\right]$$

Then, $P(\Sigma|\beta, \zeta, W, Y) \propto P(\beta, \zeta, W, |Y)$:

$$\propto P(\Sigma)\times|\Sigma|^{-\frac{n}{2}}\times\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(W_i-X_i\beta\right)^T\sum{}^{-1}\left(W_i-X_i\beta\right)\right]$$

$$\propto|V|^{\frac{m}{2}}\times|\Sigma|^{-\frac{m+k+1}{2}}\exp\left[-\frac{1}{2}tr\left(V\sum{}^{-1}\right)\right]\times|\Sigma|^{-\frac{n}{2}}$$

$$\times\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(W_i-X_i\beta\right)^T\sum{}^{-1}\left(W_i-X_iX_i\beta\right)\right]$$

$$\propto|\Sigma|^{-\frac{n+m+k+1}{2}}\exp\left\{-\frac{1}{2}tr\left[\left(\sum_{i=1}^{n}\left(W_i-X_i\beta\right)\left(W_i-X_i\beta\right)^T+V\right)\sum{}^{-1}\right]\right\}$$

So:

$$\Sigma\mid\beta,W,Y\sim\ln vWish_q\left(\sum_{i=1}^{n}\left(W_i-X_i\beta\right)\left(W_i-X_i\beta\right)^T+V,n+m+k+1\right)$$

In this Appendix, we give details of sampling $D|\ \beta, \zeta, W, Y$ in Step 3.2.1.1 for the PX-GSM algorithm.

First, we realize that $D|\ \beta, \zeta, W, Y$ is the condition before D given R, denoted by $D|R$.

We assume $P(\Sigma)\sim$Inv Wish$_k$ (m, V). Then

$$P(\Sigma)=\frac{1}{2^{mk/2}}\frac{1}{\Gamma_p\left(\frac{m}{2}\right)}|V|^{m/2}|\Sigma|^{-\frac{m+k+1}{2}}\exp\left[-\frac{1}{2}tr\left(V\sum{}^{-1}\right)\right]$$

The Jacobian of the transformation from Σ to (*R*, *D*) is denoted by $|J_{(\Sigma)}(R, D)|$ which can be calculated to be equal to $|D|^{\frac{k-1}{2}}$. Then we have $P(R,D) = P(\Sigma) \times \left|J_{(\Sigma)}(R,D)\right| = P(\Sigma) \times |D|^{\frac{k-1}{2}}$.

Therefore:

$$
\begin{aligned}
&P(D \mid R) \propto P(\Sigma) \times \left|J_{(\Sigma)}(R,D)\right| \\
&\propto P(\Sigma) \times |D|^{\frac{k-1}{2}} \\
&\propto \frac{1}{2^{\frac{mk}{2}}} \frac{1}{\Gamma_p\left(\frac{m}{2}\right)} |V|^{\frac{m}{2}} |\Sigma|^{-\frac{m+k+1}{2}} \exp\left[-\frac{1}{2} tr\left(V \sum^{-1}\right)\right] \times |D|^{\frac{k-1}{2}} \\
&\propto |\Sigma|^{-\frac{m+k+1}{2}} \exp\left[-\frac{1}{2} tr\left(V \sum^{-1}\right)\right] \times |D|^{\frac{k-1}{2}} \\
&\propto |\Sigma|^{-\frac{m+k+1}{2}} \exp\left[-\frac{1}{2} tr\left(VD^{-1/2} R^{-1} D^{-1/2}\right)\right] \times |D|^{\frac{k-1}{2}} \\
&\propto |R|^{-\frac{m+k+1}{2}} \times |D|^{-\frac{m+k+1}{2}} \exp\left[-\frac{1}{2} tr\left(VD^{-1/2} R^{-1} D^{-1/2}\right)\right] \times |D|^{\frac{k-1}{2}} \\
&\propto |D|^{-\frac{m}{2}-1} \exp\left[-\frac{1}{2} tr\left(VD^{-1/2} R^{-1} D^{-1/2}\right)\right]
\end{aligned}
$$

Circumstance 1: $V = diag(v_{11}, v_{22}, \ldots, v_{kk})$:

$$
\begin{aligned}
&P(D \mid R) \propto |D|^{-\frac{m}{2}-} \exp\left[-\frac{1}{2} tr\left(VD^{-1} R^{-1}\right)\right] \\
&\propto |D|^{-\frac{m}{2}-} \exp\left[-\frac{1}{2}\left(v_{11} d_1^{-1} r_{11} + v_{22} d_2^{-1} r_{22} + \cdots + v_{kk} d_k^{-1} r_{kk}\right)\right] \\
&\propto \left(d_1 d_2 \ldots d_k\right)^{-\frac{m}{2}-1} \exp\left[-\frac{1}{2}\left(v_{11} d_1^{-1} r_{11} + v_{22} d_2^{-1} r_{22} + \cdots + v_{kk} d_k^{-1} r_{kk}\right)\right]
\end{aligned}
$$

Then, $P(D|R) = P(d_1|\mathrm{R}) \times P(d_2|\mathrm{R}) \times \ldots \times P(d_k|R)$ and:

$$P\left(d_1 \mid R\right) = \left(d_1\right)^{-\frac{m}{2}-1} \exp\left[-\frac{1}{2}\left(v_{11} d_1^{-1} r_{11}\right)\right] \sim IG\left(\frac{m}{2}, \frac{v_{11} r_{11}}{2}\right)$$

$$P\left(d_2 \mid R\right) = \left(d_2\right)^{-\frac{m}{2}-1} \exp\left[-\frac{1}{2}\left(v_{22} d_2^{-1} r_{22}\right)\right] \sim IG\left(\frac{m}{2}, \frac{v_{22} r_{22}}{2}\right)$$

:

$$P\left(d_k \mid R\right) = \left(d_k\right)^{-\frac{m}{2}-1} \exp\left[-\frac{1}{2}\left(v_{kk} d_k^{-1} r_{kk}\right)\right] \sim IG\left(\frac{m}{2}, \frac{v_{kk} r_{kk}}{2}\right)$$

Then, $P\left(d_i^{-1} \mid R\right) \sim G\left(\frac{m}{2}, \frac{2}{v_{ii} r_{11}}\right)$, for $i = 1, \ldots, k$.

The, r_{11} is the *ith* diagonal element for R^{-1}, for $i = 1, \ldots, k$.

Circumstance 2: If *VV* is not a diagonal matrix, then we use the MH algorithm to sample *D* given *R*; this is to sample $P(D|R) \propto P(\Sigma) \times |J_{(\Sigma)}(R, D)|$.

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

1. Generate D^* by generating $\sum^* = D^{*\frac{1}{2}} R^* D^{*\frac{1}{2}}$ from Inverse-Wishart$_k$ $(m_p, m_p \times \Sigma^{(t)})$
2. Take

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

where $\alpha \min_{\alpha = \min}\left\{\frac{P\left(D^* \mid R\right)}{P\left(D^{(t)} \mid R\right)} \frac{f\left(\Sigma^{(t)} \mid \Sigma^*\right)}{f\left(\Sigma^* \mid \Sigma^{(t)}\right)}, 1\right\}$ and the proposal density $f(\Sigma^*|\Sigma^{(t)})$ is equal to $J_{\Sigma^* \to R^*, D^*} \times$ Inverse-Wishart$_k$ $(m_p, m_p \times \Sigma^{(t)})$.