Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/140302

Bayesian Analysis of Longitudinal Ordinal Data with Missing Values Using Multivariate Probit Models

Xiao Zhang

Department of Mathematical Sciences, Michigan Technological University 1400 Townsend Drive, Houghton, Michigan 49931-1295, USA

Received: 12 Dec. 2024, Revised: 23 Feb. 2025, Accepted: 30 Mar. 2025
Published online: 1 May 2025

Abstract: In this paper, we propose efficient Bayesian methods to analyze longitudinal ordinal data with missing values using multivariate probit models. Longitudinal ordinal data with substantial missing values are ubiquitous in many scientific fields. Specifically, we develop the Markov chain Monte Carlo (MCMC) sampling methods based on the non-identifiable multivariate probit models and further compare their performance with the one based on the identifiable multivariate probit models. We carried out our investigation through simulation studies, which show that the proposed methods can handle substantial missing values and the method with marginalizing the redundant parameters based on the non-identifiable model outperforms the others in the mixing and convergences of the MCMC sampling components. We then present an application using data from the Russia Longitudinal Monitoring Survey-Higher School of Economics (RLMS-HSE).

Keywords: Longitudinal ordinal data; multivariate probit model; non-identifiable multivariate probit model; missing data; dropout.

1 Introduction

Longitudinal ordinal data are commonly served as primary or secondary outcome measures in many scientific fields, such as anesthesiology, psychology, and various survey studies. However, considerable missing values are inevitable. For example, in the Russia Longitudinal Monitoring Survey of the National Research University Higher School of Economics (RLMS-HSE) [1] targeting the aspects related to job satisfaction, the main outcomes are longitudinal ordinal variables with substantial dropout rates.

The random effects models have been widely employed to analyze longitudinal ordinal data, e.g., the mixed effect models [2,3,4,5] and the multilevel models [6,7]. In contrast with the random effect models, the generalized estimating equations (GEE) approach [8], being a marginal model, is generally implemented for various data structures, including longitudinal ordinal data [9,10,11,12,13,14].

With the seminal work by Gelfand and Smith [15] for Gibbs sampling and the maturation of Markov chains theory [16,17,18,19], the MCMC methods have become a general computation tool in Bayesian inference. The MCMC methods have been developed for analyzing longitudinal ordinal data using random-effects models [20,21,22].

Following Ashford and Sowden [23], the multivariate probit models have been popularly utilized to develop MCMC methods to analyze multivariate binary and ordinal data [24,25,26,27,28,29]. The multivariate probit models assume that there is a multivariate normal variable underlying each multivariate binary or ordinal variable. However, the identifiable multivariate probit models necessitate the covariance matrix for the underlying multivariate normal variable to be a correlation matrix, which is usually sampled by a rigorous Metropolis-Hastings (MH) algorithm other than a Gibbs sampler. Therefore, this restriction brings in complication to develop an efficient MCMC method.

Due to the ubiquitousness of missing data, it is a critical issue to develop a statistical method that can handle missing values. Rubin [30] introduced the concept of missing data mechanisms, which can be categorized as missing completely at

* Corresponding author e-mail: xzhang35@mtu.edu

random (MCAR), missing at random (MAR), and missing not at random (MNAR). The MCAR means the missingness is irrelevant to the variables included in the model; the MAR means the missingness is relevant to the observed values of the variables included in the model; and the MNAR means the missingness is relevant to the unobserved values of the variables included in the model. If the missingness is MCAR or MAR, and the parameters of the model and those of describing the conditional distribution of the missingness given data are distinct, i.e., they have independent prior distributions, then the missing mechanism is called ignorable. Under the assumption of ignorability, the statistical inference can be solely based on the observed quantities. However, without ignorability, the statistical inference must take account of the unobserved values, and the estimation can be sensitive to the assumption of the non-ignorable model, for which the data contains no information [31,32,33].

Both the ignorable and nonignorable missing data mechanisms have been explored in longitudinal ordinal data. Molenberghs et al. [34] investigated the nonignorable missingness using random effect models while Kaciroti et al. [35] from a Bayesian perspective. Rana et al. [36] considered nonignorability using Monte Carlo expectation–maximization (MCEM). GEE methods have been popularly explored for ignorable missingness, such as Silva et al. [37] and Ditlhong et al. [38]. Bayesian methods have also been developed for ignorable missingness. Schafer [39] proposed a multiple imputation method using expectation–maximization (EM) and data augmentation by assuming the multivariate ordinal data follows a multivariate normal distribution; however, the normality assumption may bring undesired biases for estimated quantities. Zhang et al. [40] proposed an MCMC sampling method to conduct multiple imputation using the identifiable multivariate probit models, which entails an MH algorithm for sampling a correlation matrix and causes slow convergence.

In this article, we develop MCMC methods to analyze longitudinal ordinal data with missing values based on the non-identifiable multivariate probit models, which can circumvent a rigorous MH sampling for the correlation matrix by a Gibbs sampling for the covariance matrix and therefore improve the mixing and convergence of the MCMC components. We assume the ignorable missing data mechanism, which is the basis for multiple imputation [39,41,42,43,44] and can provide useful information for possible analyses based on nonignorable missing data mechanisms [31,45].

The remainder of the article is organized as follows. In Section 2, we present the identifiable multivariate probit model and propose a non-identifiable multivariate probit model for longitudinal ordinal data. We then develop the MCMC sampling algorithms for longitudinal ordinal data with missing values in Section 3. We conduct simulation studies in Section 4 and illustrate our methods using the RLMS-HSE study in Section 5. Then we provide a discussion in Section 6.

2 Multivariate probit models for longitudinal ordinal data

Let Y_i be a univariate ordinal data with J categories for individual i for $i = 1, \cdots, n$. Then the univariate probit model assumes that there is an underlying univariate normal variable Z_i following $N(X_i\beta,\ \sigma^2)$ with X_i being the covariate matrix, β being the regression parameter vector, and σ^2 being the variance of Z_i. The relationship between Y_i and Z_i is defined by:

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

where $\gamma_j = (\gamma_0, \gamma_1, \cdots, \gamma_J)$ are the unknown cutpoints with $\gamma_0 = -\infty$ and $\gamma_J = \infty$. In other words, this means:

$$P(Y_i = l) = P(\gamma_{l-1} < Z_i \leq \gamma_l) = P\left(\frac{\gamma_{l-1} - X_i\beta}{\sqrt{\sigma^2}} < \frac{Z_i - X_i\beta}{\sqrt{\sigma^2}} \leq \frac{\gamma_l - X_i\beta}{\sqrt{\sigma^2}}\right). \tag{2}$$

It is noticeable that the model is non-identifiable due to $\gamma_j = (\gamma_0, \gamma_1, \cdots, \gamma_J)$, β, and σ^2 being unknown parameters. The identifiable probit model usually sets $\gamma_1 = 0$ and $\sigma^2 = 1$ [46].

Now we extend the univariate probit model to a multivariate probit model by assuming each individual i for $i = 1, \cdots, n$ has a $k \times 1$ ordinal outcome vector $Y_i = \{(Y_{i1}, \cdots, Y_{ik})\}^T$ and a $k \times p$ covariate matrix $X_i = \{(X_{i1}, \cdots, X_{ik})\}^T$. Each Y_{ij} has J_j ordinal categories for $i = 1, \cdots, n$ and $j = 1, \cdots, k$. And $Z_i = \{(Z_{i1}, \cdots, Z_{ik})\}^T$ is assumed to be the underlying latent multivariate normal variable with:

$$Z_i \sim N_k(X_i\beta,\ \Sigma),$$

where X_i is the covariate matrix, β is the regression parameter vector, and Σ is the covariance matrix. Then the relationship between Y_i and Z_i is defined by:

$$Y_{ij} = l \Leftrightarrow \gamma_{j,l-1} < Z_{ij} \leq \gamma_{j,l}, \quad \text{for } l = 1, \cdots, J_j, \tag{3}$$

where $\gamma_j = (\gamma_{j,0}, \gamma_{j,1}, \cdots, \gamma_{j,J_j})$ are the unknown cutpoints with $\gamma_{j,0} = -\infty$ and $\gamma_{J_j} = \infty$. Specifically,

$$P(Y_{ij} = l) = P\left(\gamma_{j,l-1} < Z_{ij} \le \gamma_{j,l}\right) = P\left(\frac{\gamma_{j,l-1} - X_{ij}^T\beta}{\sqrt{\sigma_{jj}}} < \frac{Z_{ij} - X_{ij}^T\beta}{\sqrt{\sigma_{jj}}} \le \frac{\gamma_{j,l} - X_{ij}^T\beta}{\sqrt{\sigma_{jj}}}\right). \tag{4}$$

Just as in the univariate probit model, the unknown parameters $\gamma_j = (\gamma_{j,0}, \gamma_{j,1}, \cdots, \gamma_{j,J_j})$, β, and Σ make the model unidentifiable. The identifiable model sets $\gamma_{j,1} = 0$ and $\sigma_{jj} = 1$ for $j = 1, \cdots, k$. This implies that the covariance matrix Σ is in fact a correlation matrix [24]. However, sampling a correlation matrix instead of a covariance matrix usually entails a rigorous MH algorithm, which hinders the development of an efficient MCMC sampling method to analyze the multivariate probit model [24,25,26,27].

Comparing the identifiable models, the non-identifiable models usually make the MCMC sampling components converge faster than the identifiable models [47]. Based on the non-identifiable models, Liu and Wu [48] proved that the convergence of the MCMC sampler, called parameter-expanded data augmentation, is no slower than the original MCMC sampler under mild conditions. Therefore, we propose a non-identifiable multivariate probit model to circumvent an MH sampling for the correlation matrix and to improve the convergence of the MCMC sampling components.

The identifiable multivariate probit model assumes the latent multivariate normal variable:

$$Z_i \sim N_k(X_i\beta,\ R),$$

where R is a correlation matrix. To make the identifiable model non-identifiable, we assume:

$$Z_i \sim N_k(D^{-1/2}X_i\beta,\ R),$$

where D is the diagonal matrix with the diagonal elements being $\sigma_{11},\ \sigma_{22},\ \cdots, \sigma_{kk}$, and $\sigma_{jj} > 0$ for $j = 1, \cdots, k$. As can be seen, D makes the model unidentifiable. Let $W_i = D^{1/2}Z_i$, and it is evident that:

$$W_i \sim N_k(X_i\beta,\ D^{1/2}RD^{1/2}),$$

i.e.,

$$W_i \sim N_k(X_i\beta,\ \Sigma),$$

with $\Sigma = D^{1/2}RD^{1/2}$. Then we can define the multivariate probit model for longitudinal ordinal data based on W_i by:

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

for $j = 1, \cdots, k$ and $l = 1, \cdots, J_j$ with $\zeta_{j,0} = -\infty$, $\zeta_{j,1} = 0$, and $\zeta_{J_j} = \infty$. Equivalently,

$$P(Y_{ij} = l) = P\left(\zeta_{j,l-1} < W_{ij} \le \zeta_{j,l}\right) = P\left(\frac{\zeta_{j,l-1} - X_{ij}^T\beta}{\sqrt{\sigma_{jj}}} < \frac{W_{ij} - X_{ij}^T\beta}{\sqrt{\sigma_{jj}}} \le \frac{\zeta_{j,l} - X_{ij}^T\beta}{\sqrt{\sigma_{jj}}}\right). \tag{5}$$

In the following section, we develop the MCMC sampling algorithms based on this non-identifiable model comprehensively.

3 MCMC sampling algorithms for longitudinal ordinal data with missing values

Continuing with the notations in Section 2, we extend the ordinal outcome $Y_i = \{(Y_{i1}, \cdots, Y_{ik})\}^T$ to have missing values and use Y_{iobs} to indicate the observed values and Y_{imiss} the missing values for individual i. Then we can rewrite:

$$Y_i = \{(Y_{i1}, \cdots, Y_{ik})\}^T = \{(Y_{i,obs}, Y_{i,mis})\}^T.$$

Correspondingly, we can write:

$$W_i = \{(W_{i1}, \cdots, W_{ik})\}^T = \{(W_{i,obs}, W_{i,mis})\}^T,$$

where $W_{i,obs}$ indicates the latent variables underlying $Y_{i,obs}$ and $W_{i,mis}$ underlying $Y_{i,mis}$. We assume that the priors of β, ζ, Σ are independent, i.e.,

$$P(\beta, \zeta, \Sigma) = P(\beta) \times P(\zeta) \times P(\Sigma).$$

Moreover, we assume the missing mechanism is ignorable, implying Bayesian inferences can be based on the observed quantities, i.e., $Y_{obs} = \{(Y_{1,obs}, \cdots, Y_{k,obs})\}^T$ [30,31]. Therefore, we derive the joint posterior density of β, ζ, Σ and W given Y_{obs} as follows:

$$P(\beta,\zeta,\Sigma,W|Y_{obs}) \propto \left(\prod_{i=1}^{n} I_i\right) \times P(\beta) \times P(\zeta) \times P(\Sigma) \times |\Sigma|^{-\frac{n}{2}} \times \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(W_i - X_i\beta)^T \Sigma^{-1}(W_i - X_i\beta)\right], \quad (6)$$

where $I_i = \prod_{j=1}^{k} I_{ij}$, and

$$I_{ij} = \sum_{t=1}^{J_j} 1_{(Y_{ij},\text{obs}=t)} 1_{(\zeta_j(t-1)<W_{ij,obs}\leq\zeta_{jt})},$$

indicating compatibility of the latent variable $W_{ij,obs} \in W_{i,obs}$ with the observed ordinal variable $Y_{ij,obs} \in Y_{i,obs}$.

MCMC Sampling Steps:

Step 3.1: Sampling $\beta|\Sigma, W, Y_{obs} \sim N_k(\hat{\beta}, V_\beta)$, where:

$$V_\beta = \left(\sum_{i=1}^{n} X_i^T \Sigma^{-1} X_i + C^{-1}\right)^{-1},$$

$$\hat{\beta} = V_\beta \left(\sum_{i=1}^{n} X_i^T \Sigma^{-1} Z_i + C^{-1} b\right),$$

assuming the prior of β follows $N_p(b, C)$.

Step 3.2: Sampling $W_{ij}|\beta, \zeta, \Sigma, Y_{obs}, W_{ik}, k \neq j$, which has two cases: 1. If $Y_{ij} \in Y_{i,obs}$, then W_{ij} is sampled from a truncated normal distribution constrained between cut-points $\zeta_{j,l-1}$ and $\zeta_{j,l}$. 2. If $Y_{ij} \in Y_{i,mis}$, then W_{ij} is sampled from an univariate normal distribution without truncation.

Step 3.3: Sampling $\zeta_{j,l}|\beta, \Sigma, W, Y_{obs}, \zeta_{j,k}, k \neq l$ from a uniform distribution:

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

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

Step 3.4: Sampling $\Sigma|\beta, \zeta, W, Y_{obs}$ from an inverse-Wishart distribution:

$$\text{Inverse-Wishart}_k\left(\sum_{i=1}^{n}(W_i - X_i\beta)(W_i - X_i\beta)^T + V, n+m+k+1\right),$$

assuming a conjugate prior $P(\Sigma) \sim \text{Inverse-Wishart}_k(V, m)$.

We term this sampling algorithm as the parameter-expanded Gibbs sampling (PX-GS) algorithm. Furthermore, we introduce a refinement, the parameter-expanded Gibbs sampling with marginalization (PX-GSM) algorithm, to improve MCMC efficiency by jointly sampling redundant parameters and latent variables [48]. Specifically, joint sampling $W_{ij}, D|\beta, \zeta, R, Y_{obs}, W_{ik}, k \neq j$, and then joint sampling $\beta, \zeta, R, D|W, Y_{obs}$. With D being in both sampling steps, the redundant parameters in D are marginalized.

4 Simulation studies

To investigate the proposed PX-GS and PX-GSM algorithms (in Section 3) based on the non-identifiable multivariate probit model (in Section 2) for longitudinal ordinal data with missing values, we conduct simulation studies based on the RLMS-HSE study [1] and compare them with the PX-MH algorithm based on the identifiable multivariate probit model, which entails a Metropolis-Hastings (MH) algorithm for sampling the correlation matrix [26,40].

In this investigation, we concentrate on the longitudinal ordinal outcome—job satisfaction with primary employment from 2010 to 2016 in the RLMS-HSE study. The job satisfaction variable is 4-categorical ordinal data with higher values indicating stronger satisfaction. The considered covariates are gender, marital status, education, and age. We choose individuals who have complete records in 2010 and 2011 for job satisfaction and the covariates, obtaining a total of 7337 individuals. The covariates X_i for each individual i are sampled from these 7337 individuals.

Based on Section 5.2, we set the regression parameters:

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T = (0.9, -0.044, 0.075, 0.252, 0.0038)^T,$$

the cut-points:

$$\boldsymbol{\zeta} = (-\infty, 0, 0.72, 2.2, \infty),$$

for each of the 7-dimensional ordinal outcomes with four categories, and the correlation matrix R as follows:

$$R = \begin{bmatrix} 1.00 & 0.47 & 0.41 & 0.34 & 0.30 & 0.39 & 0.42 \\ 0.47 & 1.00 & 0.48 & 0.39 & 0.39 & 0.39 & 0.39 \\ 0.41 & 0.48 & 1.00 & 0.46 & 0.42 & 0.39 & 0.36 \\ 0.34 & 0.39 & 0.46 & 1.00 & 0.50 & 0.43 & 0.41 \\ 0.30 & 0.39 & 0.42 & 0.50 & 1.00 & 0.53 & 0.47 \\ 0.39 & 0.39 & 0.39 & 0.43 & 0.53 & 1.00 & 0.54 \\ 0.42 & 0.39 & 0.36 & 0.41 & 0.47 & 0.54 & 1.00 \end{bmatrix}.$$

We generate the 7-dimensional latent multivariate normal variables from the multivariate normal distribution with mean vector $X_i\boldsymbol{\beta}$ and covariance matrix R, and then use the cut-points $\boldsymbol{\zeta}$ to obtain the 4-categorical longitudinal ordinal data. For each scenario, we generate 100 simulated datasets with a sample size of 500.

We consider two missing data scenarios, generated using logistic regression for the indicators of missing values and dropouts with covariates being gender, marital status, education, and age:

– M-15-15: 15% sporadically missing values and 15% dropouts.
– M-25-50: 25% sporadically missing values and 50% dropouts (similar to the RLMS-HSE study).

Regarding prior distributions, we assume non-informative priors for the regression parameters $\boldsymbol{\beta}$ and the cut-points $\boldsymbol{\zeta}$. For the covariance matrix Σ, we assume two priors:

– **ID Prior:** $\Sigma \sim$ Inverse-Wishart$_k(I, m = 20)$ with I as the identity matrix.
– **CS Prior:** $\Sigma \sim$ Inverse-Wishart$_k(CS, m = 20)$ with CS being a compound symmetry matrix with equal correlations of 0.4.

We run each of the three algorithms (PX-GSM, PX-GS, and PX-MH) for 10,000 iterations with a burn-in of 2,000 and check their convergence using the R package BOA [50].

4.1 Results

Table 1 presents the averaged posterior means and standard deviations for the regression parameters and the cut-points under the ID prior. The PX-GS algorithm underestimates β_0 and β_3, while the PX-MH algorithm exhibits a larger standard deviation for β_0 compared to the others. Otherwise, the three algorithms provide similar estimates for the regression parameters. The PX-GSM algorithm tends to yield larger estimated means for the cut-points than the PX-GS algorithm, except for ζ_{51} and ζ_{52} in both missing data scenarios. The PX-MH algorithm produces some of the highest estimates, particularly for ζ_{62} and ζ_{72}. The PX-GS algorithm has the smallest standard deviations, whereas the PX-MH algorithm has the largest.

Table 1: Averaged posterior means and standard deviations for the regression parameters and the cut-points with the ID prior under three data scenarios: Complete data, M-15-15 and M-25-50.

Parameters	True Values	Complete			M-15-15			M-25-50		
		GSM	GS	MH	GSM	GS	MH	GSM	GS	MH
β_0	0.90	0.900	0.853	0.917	0.894	0.859	0.931	0.897	0.870	0.932
		(0.14)	(0.12)	(0.17)	(0.14)	(0.13)	(0.18)	(0.15)	(0.13)	(0.18)
β_1	-0.044	-0.042	-0.039	-0.042	-0.043	-0.041	-0.044	-0.040	-0.039	-0.041
		(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
β_2	0.075	0.077	0.077	0.079	0.078	0.077	0.081	0.079	0.078	0.083
		(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
β_3	0.252	0.251	0.242	0.257	0.250	0.241	0.259	0.246	0.240	0.255
		(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
β_4	0.0038	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
		(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
ζ_{11}	0.72	0.76	0.70	0.75	0.75	0.70	0.77	0.75	0.72	0.76
		(0.08)	(0.06)	(0.12)	(0.08)	(0.06)	(0.12)	(0.08)	(0.06)	(0.11)
ζ_{12}	2.20	2.24	2.17	2.31	2.23	2.19	2.35	2.23	2.20	2.34
		(0.18)	(0.09)	(0.32)	(0.18)	(0.09)	(0.34)	(0.18)	(0.09)	(0.32)
ζ_{21}	0.72	0.73	0.69	0.80	0.72	0.70	0.76	0.71	0.71	0.77
		(0.08)	(0.06)	(0.12)	(0.08)	(0.06)	(0.11)	(0.08)	(0.06)	(0.11)
ζ_{22}	2.20	2.21	2.16	2.44	2.20	2.18	2.33	2.20	2.19	2.34
		(0.18)	(0.09)	(0.35)	(0.18)	(0.09)	(0.33)	(0.19)	(0.09)	(0.32)
ζ_{31}	0.72	0.72	0.69	0.72	0.71	0.70	0.77	0.71	0.71	0.76
		(0.08)	(0.06)	(0.11)	(0.08)	(0.06)	(0.13)	(0.09)	(0.07)	(0.12)
ζ_{32}	2.20	2.20	2.17	2.20	2.20	2.17	2.36	2.20	2.19	2.35
		(0.19)	(0.09)	(0.29)	(0.20)	(0.10)	(0.35)	(0.23)	(0.10)	(0.33)
ζ_{41}	0.72	0.72	0.70	0.72	0.72	0.70	0.76	0.71	0.71	0.76
		(0.08)	(0.06)	(0.09)	(0.08)	(0.07)	(0.12)	(0.10)	(0.07)	(0.12)
ζ_{42}	2.20	2.23	2.18	2.27	2.21	2.19	2.34	2.22	2.21	2.35
		(0.19)	(0.09)	(0.30)	(0.20)	(0.10)	(0.35)	(0.24)	(0.11)	(0.34)
ζ_{51}	0.72	0.70	0.69	0.69	0.70	0.71	0.76	0.69	0.70	0.76
		(0.08)	(0.06)	(0.09)	(0.09)	(0.07)	(0.12)	(0.11)	(0.08)	(0.13)
ζ_{52}	2.20	2.21	2.18	2.12	2.20	2.20	2.35	2.19	2.21	2.37
		(0.19)	(0.09)	(0.28)	(0.21)	(0.10)	(0.33)	(0.27)	(0.12)	(0.34)
ζ_{61}	0.72	0.72	0.70	0.76	0.71	0.71	0.78	0.71	0.70	0.77
		(0.08)	(0.07)	(0.12)	(0.09)	(0.07)	(0.12)	(0.11)	(0.09)	(0.13)
ζ_{62}	2.20	2.20	2.17	2.28	2.19	2.18	2.36	2.19	2.19	2.39
		(0.19)	(0.10)	(0.35)	(0.21)	(0.10)	(0.32)	(0.28)	(0.13)	(0.35)
ζ_{71}	0.72	0.74	0.70	0.72	0.73	0.70	0.77	0.72	0.71	0.75
		(0.08)	(0.06)	(0.11)	(0.08)	(0.07)	(0.12)	(0.11)	(0.09)	(0.13)
ζ_{72}	2.20	2.22	2.17	2.23	2.22	2.19	2.35	2.20	2.18	2.29
		(0.18)	(0.09)	(0.30)	(0.20)	(0.10)	(0.34)	(0.27)	(0.13)	(0.33)

The posterior averaged biases and root mean squared errors (RMSE) for the regression parameters and cut-points are given in Supplementary Table 1. Figure 1 presents boxplots of the averaged biases and RMSE for β_1, β_3, ζ_{41}, and ζ_{42}. The PX-GSM and PX-GS algorithms show similar biases and RMSEs for β_1 and β_3, whereas the PX-MH algorithm indicates slightly biased estimation, particularly for β_3 under M-15-15. For ζ_{41} and ζ_{42}, the PX-GSM algorithm has the smallest biases, while the PX-MH algorithm has the largest. Regarding RMSEs, the PX-GSM algorithm yields the smallest RMSE for ζ_{41}, while the PX-MH algorithm produces the largest. The RMSEs for complete data are the smallest, while those for M-25-50 are the largest.

The averaged posterior means and standard deviations for all 21 correlation parameters under both priors are presented in Supplementary Tables 2 and 3. As an illustration, the averaged posterior means and standard deviations for the correlation parameters r_{12}, r_{46}, r_{57}, and r_{67} are presented in Table 2, and the boxplots for their biases and RMSE are drawn in Fig. 2.

It can be seen from Table 2 that for each data scenario, the averaged means for the PX-GSM algorithm are the closest to the true values, while those for the PX-MH algorithm are the opposite. The estimated standard deviations slightly increase with the missing percentages for each algorithm. We notice that the CS prior, which is close to the true values, improves the estimated correlations, especially for the PX-GS and the PX-MH algorithms. Fig. 2 illustrates that the PX-GSM algorithm produces the smallest biases and RMSE, while the PX-MH has the largest. Clearly, the CS prior improves the biases and RMSE, especially for the PX-GS and PX-MH algorithms.

This suggests that the PX-GSM algorithm is robust to the prior specification in comparison with the PX-GS and PX-MH algorithms. It is notable that the missing percentages seem to have a larger effect on the RMSE of the correlation parameters, such as r_{46}, r_{57}, and r_{67}, than those of the regression parameters and cut-points shown in Fig. 1, with the M-25-50 producing the largest RMSE and the complete data the smallest.

Table 2: Averaged means and standard deviations for the correlation parameters r_{12}, r_{46}, r_{57}, and r_{67} using the ID and CS priors under three data scenarios: Complete data, M-15-15 and M-25-50.

Parameters	True Values	Complete			M-15-15			M-25-50		
		GSM	GS	MH	GSM	GS	MH	GSM	GS	MH
$\mathbf{r}_{12}$ (ID)	0.47	0.46 (0.04)	0.45 (0.04)	0.43 (0.04)	0.46 (0.04)	0.45 (0.04)	0.44 (0.04)	0.46 (0.04)	0.45 (0.04)	0.44 (0.04)
$\mathbf{r}_{12}$ (CS)		0.47 (0.04)	0.46 (0.04)	0.46 (0.04)	0.47 (0.04)	0.47 (0.04)	0.46 (0.04)	0.47 (0.04)	0.47 (0.04)	0.46 (0.04)
$\mathbf{r}_{46}$ (ID)	0.43	0.41 (0.05)	0.40 (0.04)	0.37 (0.04)	0.41 (0.05)	0.40 (0.05)	0.38 (0.05)	0.41 (0.07)	0.39 (0.07)	0.35 (0.06)
$\mathbf{r}_{46}$ (CS)		0.42 (0.05)	0.42 (0.04)	0.41 (0.04)	0.43 (0.05)	0.42 (0.05)	0.42 (0.05)	0.43 (0.07)	0.41 (0.06)	0.41 (0.06)
$\mathbf{r}_{57}$ (ID)	0.47	0.47 (0.04)	0.45 (0.04)	0.44 (0.04)	0.46 (0.05)	0.45 (0.05)	0.43 (0.05)	0.47 (0.06)	0.45 (0.07)	0.42 (0.06)
$\mathbf{r}_{57}$ (CS)		0.48 (0.04)	0.47 (0.04)	0.47 (0.04)	0.47 (0.05)	0.47 (0.05)	0.46 (0.05)	0.49 (0.06)	0.47 (0.06)	0.47 (0.06)
$\mathbf{r}_{67}$ (ID)	0.54	0.52 (0.04)	0.50 (0.04)	0.49 (0.04)	0.52 (0.05)	0.50 (0.04)	0.49 (0.05)	0.51 (0.06)	0.49 (0.06)	0.46 (0.06)
$\mathbf{r}_{67}$ (CS)		0.53 (0.04)	0.52 (0.04)	0.52 (0.04)	0.53 (0.04)	0.52 (0.04)	0.52 (0.04)	0.53 (0.06)	0.51 (0.06)	0.51 (0.06)

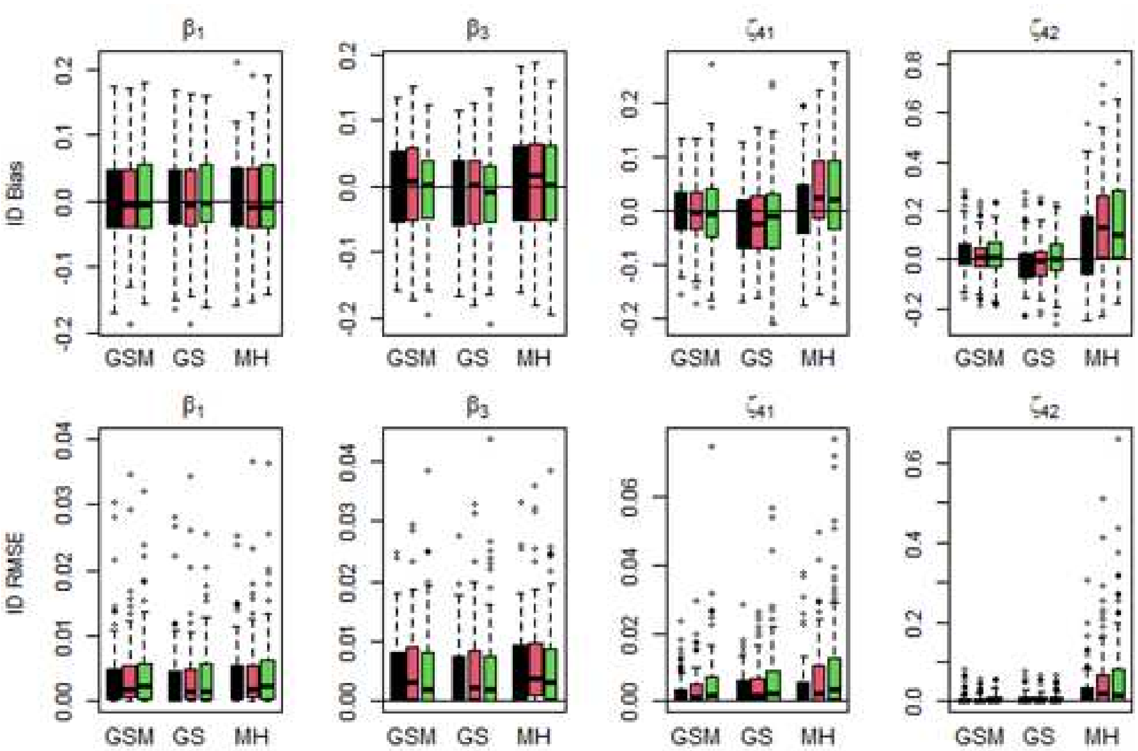

Fig. 1: Averaged biases (first row) and RMSE (second row) for the regression parameters β_1 and β_3 and the cut-points ζ_{41} and ζ_{42} using the ID prior for complete data (black), M-15-15 (red), and M-25-50 (green).

Table 3: Missing values and dropout rates for the RLMS-HSE study.

Years	Total Missing Numbers (%)	Dropout Numbers (%)
2012	1430 (19.49)	922 (12.57)
2013	2141 (29.18)	729 (9.94)
2014	3383 (46.11)	1174 (16.00)
2015	3694 (50.35)	504 (6.87)
2016	3852 (52.50)	523 (7.13)

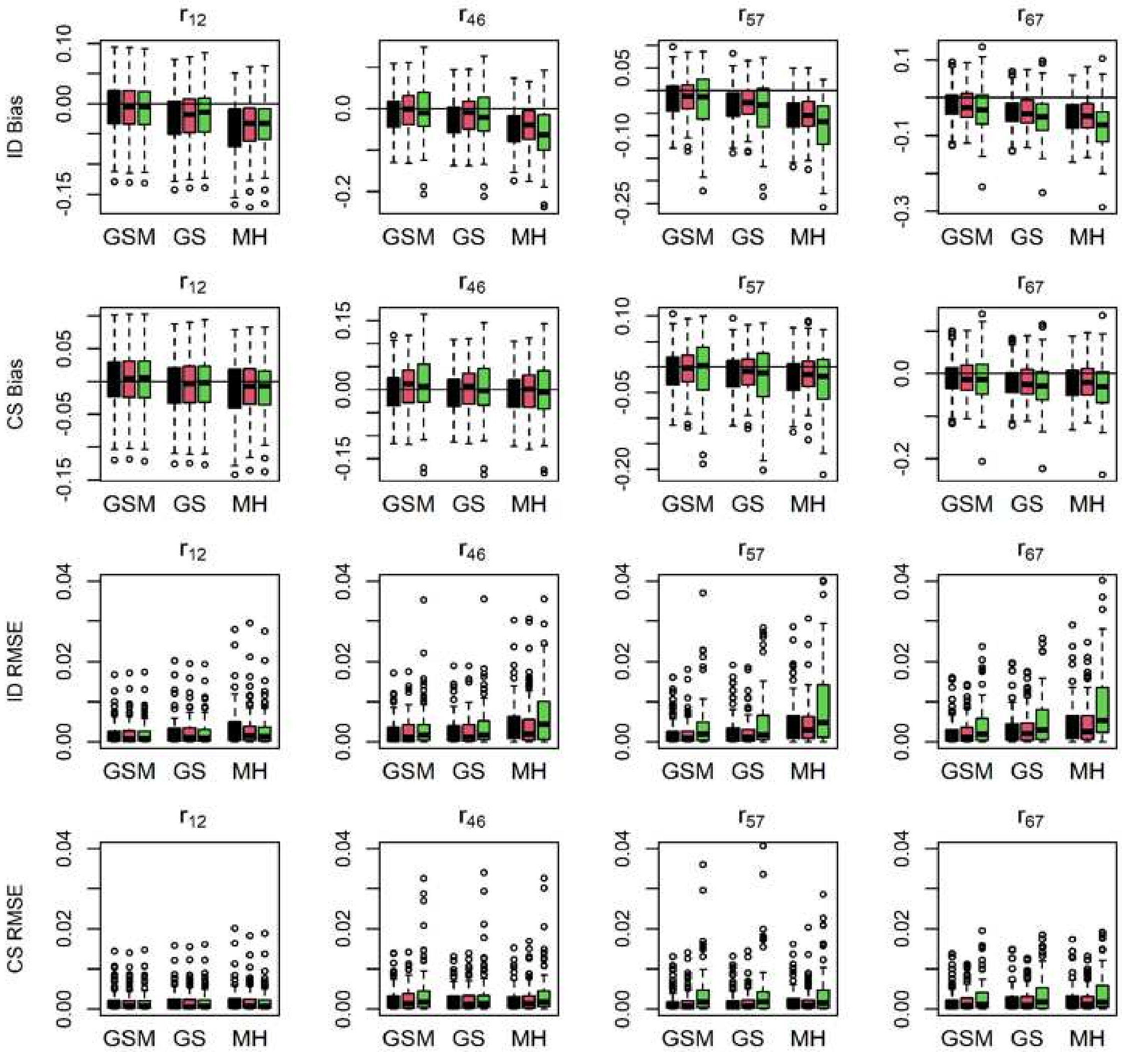

Fig. 2: Bias (the first row for the ID prior and the second row for the CS prior) and RMSE (the third row for the ID prior and the fourth row for the CS prior) for selected correlation parameters r_{12}, r_{46}, r_{57}, and r_{67} using the ID and CS priors for complete data (black), M-15-15 (red), and M-25-50 (green).

5 Application

The RLMS-HSE study [1] targets aspects related to satisfaction with primary and secondary employment. One of the main outcomes is job satisfaction with primary employment, a 5-categorical ordinal variable with:

–1: Absolutely satisfied
–2: Mostly satisfied
–3: Neutral
–4: Not very satisfied
–5: Absolutely unsatisfied

For our investigation, we focus on the longitudinal data from the years 2010 to 2016 for individuals. We include gender (1 being male and 0 female), marital status (1 being in a registered marriage and 0 otherwise), education (1 having a higher education diploma or more and 0 otherwise), and age as the covariates. We then choose those who have complete records in the years 2010 and 2011 for job satisfaction as well as the covariates and obtain a total of 7337 individuals.

The missing values and dropout information for job satisfaction are presented in Table 3. As can be seen, the dropout percentages are substantial and accumulate year by year. We then investigate the possible missing data mechanisms by conducting logistic regression (or chi-squared test) for both the missing and dropout indicators on each covariate and job satisfaction in the years 2010 and 2011. We find that the missing and dropout indicators are significantly related to each covariate but are not significantly related to job satisfaction in years 2010 and 2011. This investigation serves as the foundation for us to focus on the MAR missing mechanism and the design of simulation studies in Section 4.

We recode the 5-categorical job satisfaction variable as a 4-categorical ordinal variable with:

–0: Absolutely unsatisfied or Not very satisfied
–1: Neutral
–2: Mostly satisfied
–3: Absolutely satisfied

We then conduct a longitudinal data analysis for job satisfaction with the 7337 individuals using the PX-GSM, PX-GS, and PX-MH algorithms by running 40,000 iterations with 20,000 burn-in. With a large sample size of 7337, we use non-informative priors for the regression parameters, the cut-points, and the correlations. The results are presented in Table 4.

Table 4: Posterior means and standard deviations for the regression parameters, cut-points and correlations for the RLMS-HSE study with 95% credible intervals for the regression parameters.

Parameters	GSM	GS	MH
β_0	0.889 (0.0694)	0.796 (0.0343)	0.904 (0.0331)
β_1	-0.036 (0.0195)	-0.032 (0.0168)	-0.036 (0.0181)
95% CI	(-0.08, 0.001)	(-0.07, 0.001)	(-0.07, -0.001)
β_2	0.070 (0.019)	0.066 (0.0173)	0.070 (0.0186)
95% CI	(0.033, 0.107)	(0.032, 0.100)	(0.033, 0.106)
β_3	0.255 (0.0258)	0.236 (0.0185)	0.256 (0.0194)
95% CI	(0.204, 0.305)	(0.200, 0.272)	(0.217, 0.294)
β_4	0.004 (0.0008)	0.003 (0.0007)	0.003 (0.0007)
95% CI	(0.002, 0.005)	(0.002, 0.005)	(0.002, 0.005)
ζ_{11}	0.74 (0.06)	0.61 (0.02)	0.73 (0.01)
ζ_{12}	2.08 (0.16)	1.93 (0.03)	2.12 (0.02)
ζ_{21}	0.72 (0.05)	0.65 (0.02)	0.74 (0.01)
ζ_{22}	2.28 (0.17)	2.19 (0.03)	2.33 (0.03)
ζ_{31}	0.73 (0.07)	0.68 (0.03)	0.73 (0.01)
ζ_{32}	2.21 (0.21)	2.14 (0.04)	2.21 (0.02)
ζ_{41}	0.79 (0.08)	0.75 (0.03)	0.79 (0.02)
ζ_{42}	2.25 (0.23)	2.19 (0.04)	2.22 (0.03)
ζ_{51}	0.72 (0.09)	0.68 (0.03)	0.74 (0.02)
ζ_{52}	2.27 (0.27)	2.23 (0.04)	2.29 (0.03)
ζ_{61}	0.76 (0.09)	0.72 (0.04)	0.79 (0.02)
ζ_{62}	2.28 (0.27)	2.22 (0.05)	2.30 (0.03)
ζ_{71}	0.77 (0.09)	0.69 (0.02)	0.74 (0.02)
ζ_{72}	2.33 (0.26)	2.24 (0.05)	2.29 (0.03)

Parameters	GSM	GS	MH
r_{12}	0.47 (0.02)	0.43 (0.01)	0.44 (0.01)
r_{13}	0.41 (0.02)	0.37 (0.01)	0.39 (0.01)
r_{14}	0.34 (0.02)	0.31 (0.02)	0.33 (0.01)
r_{15}	0.30 (0.02)	0.27 (0.02)	0.29 (0.02)
r_{16}	0.29 (0.02)	0.26 (0.02)	0.28 (0.02)
r_{17}	0.27 (0.02)	0.24 (0.02)	0.26 (0.02)
r_{23}	0.47 (0.02)	0.43 (0.01)	0.46 (0.01)
r_{24}	0.39 (0.02)	0.36 (0.02)	0.39 (0.01)
r_{25}	0.39 (0.02)	0.34 (0.02)	0.39 (0.02)
r_{26}	0.35 (0.02)	0.31 (0.02)	0.35 (0.02)
r_{27}	0.34 (0.02)	0.30 (0.02)	0.34 (0.02)
r_{34}	0.46 (0.02)	0.41 (0.02)	0.46 (0.01)
r_{35}	0.42 (0.02)	0.37 (0.02)	0.42 (0.02)
r_{36}	0.39 (0.02)	0.34 (0.02)	0.39 (0.02)
r_{37}	0.36 (0.02)	0.31 (0.02)	0.37 (0.02)
r_{45}	0.50 (0.02)	0.43 (0.02)	0.51 (0.02)
r_{46}	0.43 (0.02)	0.38 (0.02)	0.44 (0.02)
r_{47}	0.41 (0.02)	0.35 (0.02)	0.42 (0.02)
r_{56}	0.53 (0.02)	0.46 (0.02)	0.54 (0.02)
r_{57}	0.46 (0.02)	0.40 (0.02)	0.42 (0.02)
r_{67}	0.53 (0.02)	0.47 (0.02)	0.48 (0.01)

As can be seen, the 95% credible intervals for Gender (β_1) include 0 for the PX-GSM and PX-GS algorithms, while the PX-MH algorithm indicates that females tend to have higher satisfaction than males. This suggests that the PX-MH algorithm based on the identifiable model may have an advantage over those based on non-identifiable models regarding inferences for the regression parameters.

These three algorithms have consistent 95% credible intervals for the other covariates:

–Marital status with a registered marriage produces higher satisfaction than other statuses.
–Higher education also has a positive effect on job satisfaction.
–Older individuals have higher satisfaction than younger individuals.

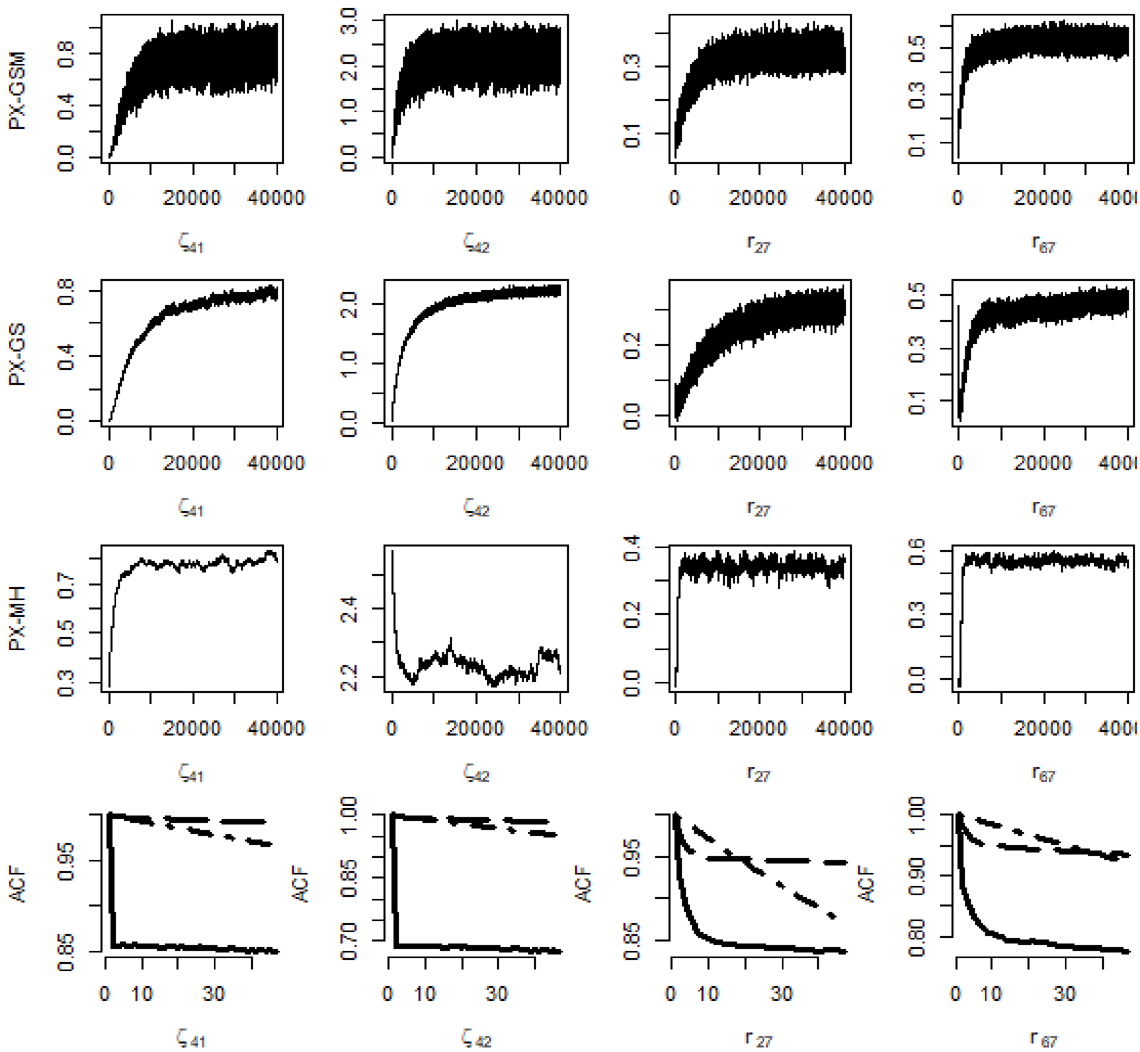

Fig. 3: Trace plots (the first three rows) for the cut-points (ζ_{41} and ζ_{42}) and the correlations (r_{27} and r_{67}) and the ACF plots (the last row), with the solid lines representing the PX-GSM algorithm, the long dashes representing the PX-GS algorithm, and the dot-dash lines representing the PX-MH algorithm.

Regarding the cut-points, Table 4 shows that job satisfaction has roughly the same cut-points for each year. The PX-GS gives the smallest estimated values, while the PX-GSM algorithm has the largest standard deviations among the three algorithms. The first two columns of Fig. 3 show the trace plots and the autocorrelation function (ACF) plots for the two cut-points ζ_{41} and ζ_{42} for illustration.

As can be seen from the trace plots (the first three rows), the PX-GSM algorithm stabilizes around 5,000 iterations, while the PX-GS algorithm stabilizes around 15,000 and the PX-MH algorithm around 10,000. The ACF plots (the last row) show that the ACF values of the PX-GSM algorithm decrease much faster than those of the other two algorithms. It is apparent that the PX-GSM algorithm has much better mixing than the PX-GS and PX-MH algorithms, and their ACF plots demonstrate the superiority of the PX-GSM algorithm. This explains why the much smaller standard deviations of the PX-GS and PX-MH algorithms are due to slow mixing and convergence.

From Table 4, it may be seen that the correlation structure of the underlying multivariate normal variables for the 7-dimensional ordinal data is between an autoregressive AR(1) and compound symmetry. The PX-GS algorithm produces the smallest estimated values for the correlation parameters among the three algorithms, but all three algorithms have similar standard deviations.

The third and fourth columns in Fig. 3 show the trace plots and the ACF plots for the two correlation parameters r_{27} and r_{67} as an example. The trace plots (the first three rows) illustrate that the PX-GSM and PX-MH algorithms stabilize much faster than the PX-GS algorithm, and the PX-GSM algorithm shows better mixing than the PX-MH algorithm, whereas the PX-GS algorithm does not appear to stabilize even after 30,000 iterations for r_{27}. The ACF plots (the last row) indicate that the PX-GSM algorithm has the fastest decreasing ACF values, whereas the PX-GS algorithm has the slowest. Overall, the PX-GSM algorithm demonstrates superior performance among the three algorithms.

6 Discussion

In this manuscript, we develop the PX-GSM and PX-GS algorithms for analyzing longitudinal ordinal data with missing values using a non-identifiable multivariate probit model and compare their performances with the PX-MH algorithm based on the identifiable model. The PX-GSM algorithm involves marginalization for the redundant parameters, while the PX-GS algorithm does not include marginalization.

The simulation studies show that the PX-GSM algorithm outperforms the PX-GS and PX-MH algorithms overall. Specifically, the PX-GSM algorithm produces less bias and smaller RMSE for the estimation of regression parameters, cut-points, and correlations than the PX-GS and PX-MH algorithms. The PX-GSM algorithm also shows better mixing and convergence than the other two algorithms. Good priors for correlation parameters significantly improve performance, especially for the PX-GS and PX-MH algorithms, suggesting that the PX-GSM algorithm is robust to prior specification in comparison with the other two algorithms. The effect of missing percentages within each algorithm for the correlation parameters appears prominent (bias and RMSE) compared with that for the regression parameters and cut-points.

Furthermore, the application to the RLMS-HSE study shows that the PX-MH algorithm, based on the identifiable model, produces a significant gender effect, whereas the PX-GSM and PX-GS algorithms, based on the non-identifiable model, fail to do so. The PX-GSM algorithm outperforms the PX-GS and PX-MH algorithms in terms of mixing and convergence for the cut-points and correlation parameters. Although the PX-GS algorithm based on the non-identifiable model circumvents an MH algorithm for sampling the correlation matrix, the convergence for the correlations may still be underperformed.

In summary, our investigation shows that the PX-GSM algorithm, based on the non-identifiable model with marginalization of redundant parameters, outperforms the PX-GS algorithm (which is also based on the non-identifiable model but without marginalization) and the PX-MH algorithm based on the identifiable model. This illustrates that non-identifiable models with marginalization of redundant parameters should be considered in developing efficient MCMC sampling methods.

We assume ignorable missing data mechanisms in both the simulation studies and the real application. A future research direction may be to extend the non-identifiable multivariate probit model to multi-level models, which could provide a possibility to consider non-ignorable missing data mechanisms.

Conflicts of Interest Statement

The authors declare that they have no conflict of interest.

Acknowledgments

This work is supported by NIH Grant R15GM151700 to Xiao Zhang.

References

[1] Kozyreva, P., Kosolapov, M., Popkin, B.M. (2016). 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*, 45(2): 395–401.

[2] Liu, C., Hedeker, D.A. (2006). A Mixed-Effects Regression Model for Longitudinal Multivariate Ordinal Data. *Biometrics*, 62(1): 261-268.

[3] Varin, C.A. (2010). Mixed Autoregressive Probit Model for Ordinal Longitudinal Data. *Biostatistics*, 11(1): 127–138.

[4] Laffont, C.M, Vandemeulebroecke, M., Concordet, D. (2014). Multivariate Analysis of Longitudinal Ordinal Data with Mixed Effects Models, with Application to Clinical Outcomes in Osteoarthritis. *Journal of the American Statistical Association*, 109(507): 955-966.

[5] Archer, K.J., Hedeker, D., Nordgren, R., Gibbons, R.D. (2015). mixor: An R Package for longitudinal and clustered ordinal response modeling. Available at: https://rdrr.io/cran/mixor/f/inst/doc/mixor.pdf.

[6] Deleeuw, J., Meijer, E. (eds). (2008). *Handbook of Multilevel Analysis*. Springer.

[7] Grilli, L., Rampichini, C. (2012). *Multilevel Models for Ordinal Data in Modern Analysis of Customer Surveys: with Applications using R*. Wiley.

[8] Liang, K. Y., and Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73(1): 13-22.

[9] Lumley, T. (1996). Generalized Estimating Equations for Ordinal Data: a Note on Working Correlation Structures. *Biometrics*, 52: 354-361.

[10] Parsons, N., Edmondson, R., Gilmour, S. (2006). A Generalized Estimating Equation Method for Fitting Autocorrelated Ordinal Score Data with an Application in Horticultural Research. *Journal of the Royal Statistical Society C*, 55: 507-524.

[11] Touloumis, A., Agresti, A., Kateri, M. (2013). GEE for Multinomial Responses Using a Local Odds Ratios Parameterization. *Biometrics*, 69(3): 633-640.

[12] Nooraee, N., Molenberghs, G., van den, H.E.R. (2014). GEE for Longitudinal Ordinal Data: Comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. *Computational Statistics and Data Analysis*, 77: 70-83.

[13] Jiang, Z., Liu, Y., Wahed, A.S., Molenberghs, G. (2018). Joint Modeling of Multiple Ordinal Adherence Outcomes via Generalized Estimating Equations with Flexible Correlation Structure. *Statistics in Medicine*, 37: 983–995.

[14] Schildcrout, J. S., Harrell Jr, F. E., Heagerty, P. J., Haneuse, S., Di Gravio, C., Garbett, S. P., Rathouz, P. J., Shepherd, B. E. (2022). Model-assisted analyses of longitudinal, ordinal outcomes with absorbing states. *Statistics in Medicine*, 41(14): 2497-2512.

[15] Gelfand, A.E., Smith, A.F.M. (1990). Sampling-based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85: 398-409.

[16] Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 4: 1701-1962.

[17] Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

[18] Meyn, S., Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*. Second Edition. Cambridge University Press.

[19] Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC. Third Edition.

[20] Johnson, T.R. (2003). On the Use of Heterogeneous Thresholds Ordinal Regression Models to Account for Individual Differences in Response Style. *Psychometrika*, 68: 563–583.

[21] Hadfield, J.D. (2010). MCMC Methods for Multi-response Generalized Linear Mixed Models: the MCMCglmm R Package. *Journal of Statistical Software*, 33: 1–22.

[22] Bürkner, P. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10: 395-411.

[23] Ashford, T., Sowden, R.R. (1970). Multivariate probit analysis. *Biometrics*, 26: 535-546.

[24] Chib, S., Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, 85: 347-361.

[25] Liu, C. (2001). Bayesian Analysis of Multivariate Probit Models – Discussion on 'The art of data augmentation' by van Dyk and Meng. *Journal of Computational and Graphical Statistics*, 10: 75-81.

[26] Zhang, X., Boscardin, W.J., Belin, T. (2006). Sampling Correlation Matrices in Bayesian Models with Correlated Latent Variables. *Journal of Computational and Graphical Statistics*, 15: 880-896.

[27] 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: 897-914.

[28] Lawrence, E., Liu, C., Bingham, D., Nair, V.N. (2008). Bayesian Inference for Multivariate Ordinal Data Using Parameter Expansion. *Technometrics*, 50: 182-191.

[29] Zhang, X. (2020). Parameter-expanded Data Augmentation for Analyzing Correlated Binary Data Using Multivariate Probit Models. *Statistics in Medicine*, 39(25): 3637-3652.

[30] Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63: 581–592.

[31] Harel, O., Zhou, X. (2007). Multiple Imputation: Review of Theory, Implementation and Software. *Statistics in Medicine*, 26: 3057–3077.

[32] Little, R. J. A., Rubin, D. B. (2019). *Statistical Analysis with Missing Data*. Third Edition. Wiley.

[33] Carpenter, J. R., Smuk, M. (2021). Missing Data: A Statistical Framework for Practice. *Biometrical Journal*, 63(5): 915-947.

[34] Molenberghs, G., Kenward, M.G., Lesaffre, E. (1997). The Analysis of Longitudinal Ordinal Data with Nonrandom Drop-out. *Biometrika*, 84: 33-44.

[35] Kaciroti, A.K., Raghunathan, T.E., Schork, A., Clark, N.M., Gong, M. (2006). A Bayesian Approach for Clustered Longitudinal Ordinal Outcome with Nonignorable Missing Data: Evaluation of an Asthma Education Program. *Journal of the American Statistical Association*, 101: 435-446.

[36] Rana, S., Roy, S., Das, K. (2018). Analysis of Ordinal Longitudinal Data under Nonignorable Missingness and Misreporting: An Application to Alzheimer's Disease Study. *Journal of Multivariate Analysis*, 166: 62–77.
[37] Silva, J.L., Colosimo, E.A., Demarqui, F.N. (2015). Modeling the Association Structure in Doubly Robust GEE for Longitudinal Ordinal Missing Data. *arXiv:1506.04452v1*.
[38] 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: 770-792.
[39] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. CRC Press, Boca Raton.
[40] Zhang, X., Li, Q., Cropsey, K., Yang, X., Zhang, K., Belin, T. (2017). A Multiple Imputation Method for Incomplete Correlated Ordinal Data Using Multivariate Probit Models. *Communications in Statistics - Simulation and Computation*, 46: 2360-2375.
[41] Yucel, R.M. (2011). State of the Multiple Imputation Software. *Journal of Statistical Software*, 45 (1).
[42] van Buuren, S., Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3).
[43] van Buuren, S. (2018). *Flexible Imputation of Missing Data*. Second Edition. Chapman & Hall/CRC.
[44] Wijesuriya, R., Moreno-Betancur, M., Carlin, J.B., White, I.R., Quartagno, M., Lee, K.J. (2025). Multiple Imputation for Longitudinal Data: A Tutorial. *Statistics in Medicine*, 44.
[45] Hammon, A., Zinn, S. (2020). Multiple Imputation of Binary Multilevel Missing not at Random Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 69(3): 547–564.
[46] Albert, J., Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, 88: 669-679.
[47] MacEachern, S.N. (2007). Comment on Article by Jain and Neal. *Bayesian Analysis*, 2: 483-494.
[48] Liu, J., Wu, Y. (1999). Parameter Expansion for Data Augmentation. *Journal of the American Statistical Association*, 94: 1264-1274.
[49] Gupta, A.K., Nagar, D.K. (2000). *Matrix Variate Distribution*. Chapman & Hall.
[50] Smith, B.J. (2007). boa: An R package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software*, 21(11).

Table 5: Supplementary Table 1: Averaged biases and root mean square errors for the regression parameters and cut-points using the ID prior based on 100 simulated datasets.

Parameters (True Values)	Complete			M-15-15			M-25-50		
	GSM	GS	MH	GSM	GS	MH	GSM	GS	MH
β_0 (0.90) (mean)	0.000	-0.047	0.017	-0.006	-0.041	0.031	-0.003	-0.030	0.032
(RMSE)	0.014	0.015	0.019	0.014	0.014	0.023	0.014	0.015	0.018
β_1 (-0.044) (mean)	0.002	0.005	0.002	0.001	0.003	0.000	0.004	0.005	0.003
(RMSE)	0.004	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.005
β_2 (0.075) (Bias)	0.002	0.002	0.003	0.003	0.002	0.006	0.003	0.003	0.008
(RMSE)	0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.005	0.005
β_3 (0.252) (Bias)	-0.001	-0.010	0.005	0.002	-0.009	0.007	-0.006	-0.012	0.003
(RMSE)	0.005	0.005	0.006	0.005	0.005	0.006	0.005	0.005	0.006
β_4 (0.0038) (Bias)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
(RMSE)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ζ_{11} (0.72) (Bias)	0.043	-0.020	0.029	0.033	-0.015	0.046	0.032	-0.004	0.045
(RMSE)	0.005	0.004	0.006	0.005	0.004	0.010	0.004	0.004	0.007
ζ_{12} (2.2) (Bias)	0.044	-0.027	0.113	0.034	-0.014	0.145	0.032	-0.001	0.139
(RMSE)	0.008	0.008	0.047	0.007	0.008	0.068	0.007	0.008	0.053
ζ_{21} (0.72) (Bias)	0.006	-0.025	0.077	-0.002	-0.021	0.042	-0.005	-0.013	0.047
(RMSE)	0.003	0.004	0.014	0.003	0.004	0.008	0.003	0.003	0.010
ζ_{22} (2.2) (Bias)	0.008	-0.037	0.242	-0.000	-0.024	0.128	-0.004	-0.012	0.140
(RMSE)	0.005	0.008	0.113	0.005	0.007	0.057	0.006	0.007	0.056
ζ_{31} (0.72) (Bias)	-0.003	-0.028	-0.002	-0.005	-0.022	0.050	-0.014	-0.017	0.044
(RMSE)	0.004	0.005	0.006	0.004	0.005	0.011	0.005	0.006	0.011
ζ_{32} (2.2) (Bias)	0.001	-0.031	0.004	-0.003	-0.025	0.157	-0.002	-0.006	0.149
(RMSE)	0.008	0.010	0.027	0.008	0.010	0.080	0.009	0.012	0.076
ζ_{41} (0.72) (Bias)	0.002	-0.023	0.004	-0.003	-0.018	0.035	-0.005	-0.014	0.035
(RMSE)	0.003	0.004	0.005	0.003	0.005	0.007	0.005	0.007	0.010
ζ_{42} (2.2) (Bias)	0.027	-0.018	0.065	0.019	-0.008	0.141	0.015	0.009	0.154
(RMSE)	0.008	0.009	0.030	0.007	0.009	0.055	0.008	0.010	0.063
ζ_{51} (0.72) (Bias)	-0.021	-0.030	-0.034	-0.021	-0.015	0.043	-0.027	-0.015	0.045
(RMSE)	0.004	0.004	0.005	0.004	0.004	0.008	0.008	0.007	0.013
ζ_{52} (2.2) (Bias)	0.007	-0.018	-0.077	0.002	-0.003	0.146	-0.011	0.007	0.171
(RMSE)	0.006	0.007	0.025	0.007	0.008	0.052	0.012	0.014	0.081
ζ (0.72) (Bias)	0.001	-0.019	0.040	-0.005	-0.012	0.056	-0.015	-0.020	0.055
(RMSE)	0.003	0.003	0.007	0.003	0.004	0.011	0.005	0.007	0.013
ζ_{62} (2.2) (Bias)	0.004	-0.027	0.080	-0.009	-0.021	0.155	-0.008	-0.012	0.188
(RMSE)	0.004	0.005	0.034	0.004	0.005	0.067	0.008	0.012	0.081
ζ_{71} (0.72) (Bias)	0.022	-0.023	0.004	0.009	-0.017	0.047	0.005	-0.013	0.034
(RMSE)	0.003	0.004	0.005	0.003	0.005	0.011	0.006	0.008	0.011
ζ_{72} (2.2) (Bias)	0.024	-0.025	0.033	0.019	-0.012	0.155	-0.003	-0.022	0.093
(RMSE)	0.006	0.007	0.026	0.006	0.008	0.082	0.009	0.011	0.050

Table 6: Supplementary Table 2: Averaged means and standard deviations for the correlation parameters using the ID prior based on 100 simulated datasets.

Para	True	Complete			M-15-15			M25-50		
		GSM	GS	MH	GSM	GS	MH	GSM	GS	MH
r_{12}	0.47	0.46	0.45	0.43	0.46	0.45	0.44	0.46	0.45	0.44
		0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
r_{13}	0.41	0.41	0.39	0.38	0.41	0.39	0.37	0.40	0.39	0.37
		0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05
r_{14}	0.34	0.34	0.32	0.31	0.33	0.32	0.30	0.33	0.32	0.30
		0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.05
r_{15}	0.30	0.30	0.29	0.26	0.30	0.29	0.26	0.30	0.29	0.25
		0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06
r_{15}	0.30	0.30	0.29	0.26	0.30	0.29	0.26	0.30	0.29	0.25
		0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06
r_{16}	0.29	0.29	0.28	0.25	0.30	0.29	0.26	0.29	0.28	0.24
		0.05	0.05	0.05	0.05	0.05	0.05	0.07	0.07	0.07
r_{17}	0.27	0.27	0.26	0.24	0.27	0.26	0.24	0.28	0.26	0.23
		0.05	0.05	0.05	0.05	0.05	0.05	0.07	0.07	0.07
r_{23}	0.48	0.48	0.46	0.45	0.48	0.47	0.45	0.48	0.47	0.45
		0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.04
r_{24}	0.39	0.39	0.37	0.35	0.39	0.37	0.35	0.39	0.38	0.35
		0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05
r_{25}	0.39	0.39	0.37	0.35	0.39	0.38	0.35	0.38	0.37	0.33
		0.05	0.05	0.04	0.05	0.05	0.05	0.06	0.06	0.05
r_{26}	0.35	0.34	0.33	0.30	0.35	0.34	0.31	0.34	0.33	0.29
		0.05	0.05	0.04	0.05	0.05	0.05	0.07	0.06	0.06
r_{27}	0.34	0.34	0.32	0.30	0.33	0.32	0.30	0.34	0.33	0.29
		0.05	0.05	0.04	0.05	0.05	0.05	0.07	0.06	0.06
r_{34}	0.46	0.45	0.44	0.42	0.45	0.44	0.42	0.45	0.43	0.41
		0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05
r_{35}	0.42	0.42	0.40	0.37	0.41	0.40	0.38	0.42	0.40	0.36
		0.05	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.06
r_{36}	0.39	0.38	0.37	0.34	0.39	0.38	0.35	0.38	0.37	0.33
		0.05	0.05	0.04	0.05	0.05	0.05	0.06	0.07	0.07
r_{37}	0.36	0.36	0.35	0.32	0.35	0.34	0.32	0.36	0.35	0.31
		0.05	0.05	0.04	0.05	0.05	0.05	0.07	0.07	0.07
r_{45}	0.50	0.49	0.47	0.45	0.49	0.47	0.46	0.48	0.46	0.43
		0.04	0.04	0.04	0.05	0.05	0.04	0.06	0.06	0.06
r_{46}	0.43	0.41	0.40	0.37	0.41	0.40	0.38	0.41	0.39	0.35
		0.05	0.04	0.04	0.05	0.05	0.05	0.07	0.07	0.06
r_{47}	0.41	0.40	0.38	0.36	0.40	0.38	0.36	0.40	0.38	0.34
		0.05	0.04	0.04	0.05	0.05	0.05	0.07	0.07	0.06
r_{56}	0.53	0.51	0.49	0.48	0.51	0.49	0.48	0.50	0.48	0.45
		0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.07	0.06
r_{57}	0.47	0.47	0.45	0.44	0.46	0.45	0.43	0.47	0.45	0.42
		0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.07	0.06
r_{67}	0.54	0.52	0.50	0.49	0.52	0.50	0.49	0.51	0.49	0.46
		0.04	0.04	0.04	0.05	0.04	0.05	0.06	0.06	0.06

Table 7: Supplementary Table 3: Averaged means and standard deviations for the correlation parameters using the CS prior based on 100 simulated datasets.

Para	True	Complete			M-15-15			M25-50		
		GSM	GS	MH	GSM	GS	MH	GSM	GS	MH
r_{12}	0.47	0.47	0.46	0.46	0.47	0.47	0.46	0.47	0.47	0.46
		0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
r_{13}	0.41	0.42	0.41	0.41	0.42	0.41	0.40	0.42	0.41	0.40
		0.04	0.04	0.04	0.05	0.05	0.04	0.05	0.05	0.05
r_{14}	0.34	0.35	0.34	0.34	0.35	0.34	0.34	0.35	0.34	0.34
		0.05	0.05	0.04	0.05	0.05	0.05	0.06	0.05	0.05
r_{15}	0.30	0.31	0.31	0.30	0.31	0.31	0.30	0.31	0.31	0.30
		0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06
r_{16}	0.29	0.30	0.30	0.29	0.31	0.31	0.30	0.31	0.30	0.30
		0.05	0.05	0.05	0.05	0.05	0.05	0.07	0.06	0.06
r_{17}	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.28
		0.05	0.05	0.04	0.05	0.05	0.05	0.07	0.06	0.06
r_{23}	0.48	0.49	0.48	0.48	0.49	0.48	0.48	0.49	0.48	0.48
		0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.04
r_{24}	0.39	0.40	0.39	0.39	0.40	0.39	0.39	0.40	0.39	0.39
		0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05
r_{25}	0.39	0.40	0.39	0.39	0.40	0.39	0.39	0.40	0.39	0.38
		0.05	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.06
r_{26}	0.35	0.35	0.35	0.34	0.36	0.35	0.35	0.36	0.35	0.35
		0.05	0.05	0.04	0.05	0.05	0.05	0.06	0.06	0.06
r_{27}	0.34	0.35	0.34	0.34	0.35	0.34	0.34	0.36	0.35	0.35
		0.05	0.05	0.04	0.05	0.05	0.05	0.06	0.06	0.06
r_{34}	0.46	0.46	0.45	0.45	0.46	0.45	0.45	0.46	0.45	0.45
		0.04	0.04	0.04	0.05	0.04	0.04	0.05	0.05	0.05
r_{35}	0.42	0.43	0.42	0.41	0.43	0.42	0.41	0.43	0.42	0.42
		0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.06
r_{36}	0.39	0.39	0.39	0.38	0.40	0.39	0.39	0.40	0.39	0.39
		0.05	0.04	0.04	0.05	0.05	0.05	0.07	0.06	0.06
r_{37}	0.36	0.37	0.36	0.36	0.37	0.36	0.36	0.38	0.37	0.37
		0.05	0.05	0.04	0.05	0.05	0.05	0.07	0.07	0.06
r_{45}	0.50	0.49	0.49	0.49	0.50	0.49	0.49	0.49	0.48	0.47
		0.04	0.04	0.04	0.05	0.04	0.04	0.06	0.06	0.06
r_{46}	0.43	0.42	0.42	0.41	0.43	0.42	0.42	0.43	0.41	0.41
		0.05	0.04	0.04	0.05	0.05	0.05	0.07	0.06	0.06
r_{47}	0.41	0.42	0.40	0.40	0.41	0.40	0.40	0.42	0.40	0.40
		0.05	0.04	0.04	0.05	0.05	0.05	0.07	0.06	0.06
r_{56}	0.53	0.52	0.51	0.51	0.52	0.51	0.51	0.52	0.50	0.49
		0.04	0.04	0.04	0.05	0.04	0.04	0.06	0.06	0.06
r_{57}	0.47	0.48	0.47	0.47	0.47	0.47	0.46	0.49	0.47	0.47
		0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.06
r_{67}	0.54	0.53	0.52	0.52	0.53	0.52	0.52	0.53	0.51	0.51
		0.04	0.04	0.04	0.04	0.04	0.04	0.06	0.06	0.06