Original Research Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
1–22

DOI: 10.1177/09622802241247725
journals.sagepub.com/home/smm
Sage

Bayesian analysis of joint quantile regression for multi-response longitudinal data with application to primary biliary cirrhosis sequential cohort study

Yu-Zhu Tian[1,2], Man-Lai Tang[3], Catherine Wong[4] and Mao-Zai Tian[5]

Abstract

This article proposes a Bayesian approach for jointly estimating marginal conditional quantiles of multi-response longitudinal data with multivariate mixed effects model. The multivariate asymmetric Laplace distribution is employed to construct the working likelihood of the considered model. Penalization priors on regression parameters are incorporated into the working likelihood to conduct Bayesian high-dimensional inference. Markov chain Monte Carlo algorithm is used to obtain the fully conditional posterior distributions of all parameters and latent variables. Monte Carlo simulations are conducted to evaluate the sample performance of the proposed joint quantile regression approach. Finally, we analyze a longitudinal medical dataset of the primary biliary cirrhosis sequential cohort study to illustrate the real application of the proposed modeling method.

Keywords

Joint modeling, quantile regression, multivariate longitudinal data, Markov chain Monte Carlo, sequential cohort study

1 Introduction

Longitudinal data or repeated measurement data frequently occur in studies of various disciplines such as medicine, biology, sociology, and economics. There are many tools available for modeling longitudinal data (e.g. latent growth models, cross-lagged regression models, and hierarchical linear models) that are helpful for revealing how attributes of individuals change over time. In longitudinal data modeling, mixed effects model is one of the most powerful statistical tools for depicting the relationship between the outcome variable and a group of predictors. The most appealing feature of mixed effects models is that observations among different individuals are independent, while observations (recorded over time) within the same subject are correlated. A pioneering study on mixed effects models with longitudinal data can be found by Laird and Ware.[1] For general references about longitudinal data and mixed effects models, one can refer to Diggle et al.,[2]

[1]School of Mathematics and Statistics, Northwest Normal University, LanZhou, China
[2]Gansu Provincial Research Center for Basic Disciplines of Mathematics and Statistics, Lanzhou, China
[3]Department of Physics, Astronomy and Mathematics, University of Hertfordshire, UK
[4]Digital Humanities Institut, University of Sheffield, UK
[5]Centre for Applied Statistics, School of Statistics, Renmin University of China, Beijing, China

Corresponding author:
Yu-Zhu Tian, School of Mathematics and Statistics, Northwest Normal University, 967 Anning East Road, LanZhou City, Gansu Province 730070, China; Gansu Provincial Research Center for Basic Disciplines of Mathematics and Statistics, LanZhou City, Gansu Province 730070, China.
Email: pole1999@163.com.

Hedeker and Gibbons,[3] Wu and Zhang,[4] Wu,[5] and Demidenko.[6] Among them, Wu and Zhang[4]discussed various nonparametric longitudinal data models, while Wu[5] presented a thorough discussion about linear effects models with complex data. Further, Demidenko[6] reviewed the theory and applications for mixed effects models with *R*.

Only one response is considered in most of the literatures on longitudinal data analysis. Classical linear mixed effects model is generally used to model single-response longitudinal data sets in which the response variable linearly depends on a set of covariates. In real-world applications, however, we often suffer from multiple-response longitudinal data structure in which responses on two or more characteristics are repeatedly recorded over time for an individual. We need to model the correlations among multiple or multivariate response variables via a set of common covariates with repeated measurements over time. It is noteworthy that multiple responses are not independent but statistically dependent. Therefore, separate analysis for each response would totally ignore the relationships among multiple responses and lead to unstable estimation results. Under such circumstances, a joint or simultaneous modeling approach of multi-response longitudinal data would be highly desirable. For example, in a longitudinal data analysis of the PBCseq (primary biliary cirrhosis sequential) cohort study, orthotopic liver transplantation can be treated as a potentially life-saving alternative for patients with advanced or end-stage primary biliary cirrhosis (PBC). Serum bilirubin and serum albumin are two of the primary indicators for evaluating and monitoring the absence of liver diseases. It is generally believed that there exist some relationships between serum bilirubin and serum albumin levels, and thus a joint analysis of the longitudinally collected serum bilirubin and serum albumin has received increasing attention in diagnosing liver diseases. We will analyze this multi-response longitudinal data set using the proposed joint QR (quantile regression) approach in Section 6.

For multi-response or multivariate longitudinal data, the modeling and inference methods are more complex. There have been extensive literatures about multivariate longitudinal data modeling methods. For example, Shah et al.[7] proposed a random-effects model for multiple longitudinal data with possibly missing data. Sammel et al.[8] studied multivariate linear mixed models for multiple outcomes. Lin[9] considered a mixed-effects regression model for longitudinal multivariate ordinal data. Blozis et al.[10] considered a nonlinear latent curve model for multivariate longitudinal data. Alfo and Maruotti[11] studied a hierarchical model for time dependent multivariate longitudinal data. Bandyopadhyay et al.[12] presented a review of multivariate longitudinal data analysis. Gebregziabher et al.[13] studied the joint modeling of multiple longitudinal outcomes using multivariate generalized linear mixed models. Laffont et al.[14] studied the multivariate longitudinal ordinal data with mixed-effects models. Grimm[15] applied the multivariate longitudinal data method to study the developmental relationship between depression and academic achievement. Wang et al.[16] considered an extension of the multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. Luwanda and Mwambi[17] discussed a nonlinear mixed-effects model for multivariate longitudinal data. Rajeswaran et al.[18] considered a joint modeling of multivariate longitudinal data and competing risks. Lin et al.[19] discussed the multivariate longitudinal data analysis with censored and intermittent missing responses, Hui et al.[20] studied a sparse pairwise likelihood estimation for multivariate longitudinal mixed models. Jiang et al.[21] considered an optimal design for multivariate logistic mixed models with longitudinal data. Wang[22] discussed Bayesian analysis of multivariate linear mixed models with censored and intermittent missing responses. Taavoni et al.[23] developed multivariate-t semiparametric mixed-effects models for longitudinal data with multiple characteristics. Tian and Qiu[24] studied multivariate single index modeling of longitudinal data with multiple responses.

It is noteworthy that most of the existing methods for multivariate longitudinal data are based on modeling the average effects of response variables conditionally on a set of covariates. These modeling methods only provide the mean regression analysis for multivariate longitudinal outcomes and usually require the normality assumption for the outcomes variables. In many real-world applications, we often encounter multivariate longitudinal outcomes which are non-Gaussian distributed. Traditional linear mixed models for handling multivariate longitudinal outcomes do not provide a powerful inference for such data. QR modeling, as a popular alternative to the traditional mean regression modeling, can be employed to assess the relationship between a set of predictors and a specific quantile of the response (see, Koenker[25] and Koenker et al.[26]). Quantiles generally produce a more complete picture of conditional distribution of the response than the mean, and perform more robust for non-normal data. There are many literatures in which QR approach is developed to model longitudinal data. Koenker[27] considered QR for longitudinal data analysis. Geraci and Bottai[28] studied QR for longitudinal data using the asymmetric Laplace distribution. Liu and Bottai[29] proposed the QR mixed-effects models with longitudinal data. Tian et al.[30] considered Bayesian joint QR for mixed-effects models with censoring and errors in covariates. Aghamohammadi and Mohammadi[31] considered Bayesian penalized QR for longitudinal data analysis. Alhamzawi and Ali[32] studied Bayesian QR for ordinal longitudinal data. Tian et al.[33] considered likelihood-based QR mixed-effects models for longitudinal data with multiple features via MCEM (Monte Carlo expectation-maximization) algorithm.

Although there are many literatures about QR methods for modeling longitudinal data, most of them only consider the single-response longitudinal data structure. There are shattered work on QR approach for multi-response longitudinal

data, even for multi-response regression setting with cross-sectional data. This is mainly because quantiles for multivariate outcomes are not uniquely defined. There is no unique definition of quantiles in higher dimension due to the lack of a natural ordering in Euclidean space of higher dimension framework. However, a few attempts have been made for multivariate QR analyses. Waldmann and Kneib[34] considered the Bayesian bivariate QR. For QR modeling of multi-response longitudinal data analysis, Kulkarni et al.[35] studied a joint QR model for multiple longitudinal outcomes, Ghasemzadeh et al.[36] considered a Bayesian QR for joint modeling of longitudinal mixed ordinal and continuous data, Biswas and Das[37] investigated a Bayesian QR approach for multivariate semi-continuous longitudinal data analysis. However, the aforementioned approaches only consider QR estimation at the same quantile for different responses. Recently, new research has gradually emerged on the subject of joint modeling on multivariate response QR for different quantiles. For example, Petrella and Raponi[38] proposed the joint estimation approach of conditional quantiles for multivariate linear regression models, Tian et al.[39] provided Bayesian joint inference for multivariate QR models.

In this article, we investigate a joint QR modeling approach for multi-response longitudinal data. In Section 2, we present the multi-response linear regression model and the joint QR working likelihood. Section 3 provides model specification of the proposed multi-response longitudinal mixed-effect model. In Section 4, we develop a MCMC (Markov chain Monte Carlo) algorithm for Bayesian joint QR modeling approach. Section 5 provides Monte Carlo simulations to examine the performance of the proposed estimation procedure. We illustrate our methodologies based on a real data set in Section 6. Conclusions are presented in Section 7.

For the convenience of the following description, we provide an unified statement for all formula notations in the following text. Lowercase letters stand for the scalars, boldface lowercase letters stand for vectors and boldface capital letters stand for matrices.

2 Preliminaries

Consider the following multi-response regression model

$$\boldsymbol{y}_i = \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{e}_i,\ i \in 1, \dots, N \tag{2.1}$$

where $\boldsymbol{y}_i = (y_{i1}, \dots, y_{ip})^T$ is a p-variate response vector for the i-th individual, $\boldsymbol{x}_i$ is a $k \times 1$ vector of regressors, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)^T$ is a $p \times k$ matrix of unknown parameters with $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jk})^T$, and $\boldsymbol{e}_i = (e_{i1}, \dots, e_{ip})^T$ denotes a $p \times 1$ vector of error terms.

For model (2.1), we assume that the τ_j-level quantile of the j-th component in $\boldsymbol{y}_i$ is a function of the k-dimensional covariate $\boldsymbol{x}_i$ which can be specified as $Q_{\tau_j}(y_{ij}|\boldsymbol{x}_i) = \beta_{\tau_j}^T \boldsymbol{x}_i$ for $j = 1, \dots, p$, where the $\boldsymbol{\beta}_{\tau_j}$ is the τ_j-th quantile coefficients vector corresponding to the j-th response. Denoting $\boldsymbol{\beta}_\tau = (\boldsymbol{\beta}_{\tau_1}, \dots, \boldsymbol{\beta}_{\tau_p})^T$ is a $p \times k$ matrix of parameters with $\boldsymbol{\beta}_{\tau_j} = (\beta_{j1}, \dots, \beta_{jk})^T$. Based on the common method for univariate QR model, one can estimate each $\boldsymbol{\beta}_{\tau_j}$ marginally by minimizing the objective function $\sum_{i=1}^N \rho_{\tau_j}(y_{ij} - \boldsymbol{\beta}_{\tau_j}^T \boldsymbol{x}_i)$, where $\rho_{\tau_j}(\cdot)$ denotes the τ_j-level univariate quantile check function. It can be noticed, however, that the estimators from the above objective functions totally ignore the dependence among the components in multivariate response $\boldsymbol{y}_i$. An alterative is to study the joint estimation of the p conditional quantiles by incorporating the correlations into the components of the multivariate response $\boldsymbol{y}_i$. And for that, a MAL (multivariate asymmetric Laplace) distribution can be imposed on the error term $\boldsymbol{e}_i$ of model (2.1) to specify the joint quantiles of $\boldsymbol{y}_i$ conditionally on convarate $\boldsymbol{x}_i$. The pdf (probability density function) of the p-variate random vector $\boldsymbol{Y}$ for the three-parameter MAL distribution $\text{MAL}_p(\boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\Omega})$ is given as

$$g_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\Omega}) = \frac{2\exp[(\boldsymbol{y} - \boldsymbol{\alpha})^T \boldsymbol{\Omega}^{-1} \boldsymbol{\xi}]}{(2\pi)^{p/2} |\boldsymbol{\Omega}|^{1/2}} \left(\frac{\kappa}{2 + d}\right)^{\upsilon/2} \cdot K_\upsilon\left(\sqrt{(2 + d)\kappa}\right)$$

where $d = \boldsymbol{\xi}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\xi}$, $\kappa = (\boldsymbol{y} - \boldsymbol{\alpha})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{y} - \boldsymbol{\alpha})$, $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$ are the shift and shape parameters, respectively, $\boldsymbol{\Omega}$ is the $p \times p$ positive definite matrix of scale parameters, and $K(\cdot)$ denotes the modified Bessel function of the third kind with index parameter $\nu = (2 - p)/2$. One can refer to Kotz et al.,[40] Kollo and Srivastava,[41] Visk,[42] and Hurlimann[43] for more disscussion on the MAL distribution. Letting $\boldsymbol{D} = \text{diag}(\sigma_1, \dots, \sigma_p)$ with $\sigma_j > 0$, $\boldsymbol{\xi} = \boldsymbol{D}\boldsymbol{\theta}$ and $\boldsymbol{\Omega} = \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}$, one can reparameterize the distribution $\text{MAL}_p(0, \boldsymbol{D}\boldsymbol{\theta}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})$ for $\boldsymbol{e}_i$ or the conditional distribution $\text{MAL}_p(\boldsymbol{\beta}_\tau \boldsymbol{x}_i, \boldsymbol{D}\boldsymbol{\theta}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})$ for $\boldsymbol{y}_i$ in model (2.1), where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ has generic element $\theta_j = \frac{1 - 2\tau_j}{(\tau_j(1 - \tau_j)}$, $\boldsymbol{\Sigma}$ is a $p \times p$ positive matrix such that $\boldsymbol{\Sigma} = \boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}$, and $\boldsymbol{\Psi}$ being a correlation matrix and $\boldsymbol{\nabla} = \text{diag}(\delta_1, \dots, \delta_p)$ with $\delta_j^2 = \frac{1}{\tau_j(1 - \tau_j)}$. The unknown parameters include $\boldsymbol{\beta}_\tau$, $\boldsymbol{\Psi}$, and $\boldsymbol{D}$ (i.e. $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_p)$).

Petrella and Raponi[38] showed that the j-th marginal distribution of $\boldsymbol{y}_i$ under the assumption of MAL distribution is an univariate asymmetric Laplace distribution AL($\boldsymbol{\beta}_{\tau_j}^T \boldsymbol{x}_i, \tau_j, \sigma_j$). Besides, the correlations among the components of multivariate response $\boldsymbol{y}_i$ are included in the matrix $\boldsymbol{\Sigma}$ (or $\boldsymbol{\Psi}$). It is noticed that the density of MAL distribution is too complex for conducting statistical inference. To make statistical inference simpler, a hierarchical representation of the response $\boldsymbol{y}_i$ can be given as follows:

$$\boldsymbol{y}_i = \boldsymbol{\beta}_\tau \boldsymbol{x}_i + \boldsymbol{D}\boldsymbol{\theta} w_i + \sqrt{w_i}\boldsymbol{D}\boldsymbol{\Sigma}^{1/2}\boldsymbol{z}_i \tag{2.2}$$

where $\boldsymbol{z}_i$ follow the p-variate standard normal distribution $\mathrm{N}_p(0, \boldsymbol{I}_p)$, w_i are latent variables which follow the standard exponential distribution Exp(1), and $\boldsymbol{z}_i$ are independent of w_i. Conditioning on w_i, $\boldsymbol{y}_i$ follow the multivariate normal distribution with the mean $\boldsymbol{\beta}_\tau \boldsymbol{x}_i + \boldsymbol{D}\boldsymbol{\theta} w_i$ and variance–covariane matrix $w_i\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}$. In the following sections, we will investigate Bayesian joint QR approach for the multivariate longitudinal mixed effect model and its application.

3 Model specification

3.1 The multi-response longitudinal mixed-effect model

Suppose a data set comes from a unbalanced longitudinal study with N subjects and each subject has n_i repeated measurements over time. Each measurement has p characteristic responses for each subject. Let $\boldsymbol{y}_{it} = (y_{it}^{(1)}, \ldots, y_{it}^{(p)})^T$ be the observation vector of the p responses for the i-th individual measured at the t-th time ($i = 1, \ldots, N; t = 1, \ldots, n_i$). The multi-response mixed-effects models with a random intercept can be expressed as follows:

$$y_{it}^{(j)} = \boldsymbol{x}_{it}^T \boldsymbol{\beta}_j + b_i^{(j)} + e_{it}^{(j)}, j = 1, \ldots, p \tag{3.1}$$

where $y_{it}^{(j)}$ is the t-th observation of the i-th individual for j-th variate response, $\boldsymbol{x}_{it}$ is a $k \times 1$ vector of covariates, $\boldsymbol{\beta}_j$ is a $k \times 1$ unknown parameter vector of fixed effects for the j-th response, $b_i^{(j)}$ is the j-th random intercept term specific to subject i, $e_{it}^{(j)}$ is the error term.

The vector expression of model (3.1) is

$$\boldsymbol{y}_{it} = \boldsymbol{\beta}\boldsymbol{x}_{it} + \boldsymbol{b}_i + \boldsymbol{e}_{it} \tag{3.2}$$

where $\boldsymbol{e}_{it} = (e_{it}^{(1)}, \ldots, e_{it}^{(p)})^T$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)^T$, and $\boldsymbol{b}_i = (b_i^{(1)}, \ldots, b_i^{(p)})^T$. The random effect $\boldsymbol{b}_i$ is simply supposed to follow p-variate normal distribution $N_p(0, \boldsymbol{\Sigma}_{\boldsymbol{b}})$ for each subject, in which $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ is a $p \times p$ variance–covariance matrix. Random effects $\boldsymbol{b}_i$ in model (3.2) accounts for the longitudinal association of data from the same individual across time. The diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ quantify the variability between subjects, and the off-diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ measure the overall association between responses.

In the framework of QR modeling, we specify the τ_j-th marginal quantile of response $\boldsymbol{y}_{it}^{(j)}$ conditionally on $\boldsymbol{x}_{it}$ and $\boldsymbol{b}_i^{(j)}$ in model (3.1) as follows:

$$Q_{\tau_j}(\boldsymbol{y}_{it}^{(j)} | \boldsymbol{x}_{it}, \boldsymbol{b}_i) = \boldsymbol{x}_{it}^T \boldsymbol{\beta}_{\tau_j} + \boldsymbol{b}_i^{(j)} \tag{3.3}$$

Model (3.3) considers the marginal quantile of response $\boldsymbol{y}_{it}^{(j)}$ and totally ignores the dependences among the components of $\boldsymbol{y}_{it}$. In order to implement the joint QR analysis, based on the preliminaries in Section 2, a multivariate distribution $\mathrm{MAL}_p(0, \boldsymbol{D}\boldsymbol{\theta}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})$ is specified on the error term $\boldsymbol{e}_{it}$ of model (3.2). Similarly to model (2.5), go a step further, we obtain the hierarchical representation of model (3.2) conditionally on the random effects $\boldsymbol{b}_i$ as follows:

$$\boldsymbol{y}_{it} = \boldsymbol{\beta}_\tau \boldsymbol{x}_{it} + \boldsymbol{b}_i + \boldsymbol{D}\boldsymbol{\theta} w_{it} + \sqrt{w_{it}}\boldsymbol{D}\boldsymbol{\Sigma}^{1/2}\boldsymbol{z}_{it}, i = 1, \ldots, N, t = 1, \ldots, n_i \tag{3.4}$$

where $\boldsymbol{\beta}_\tau = (\boldsymbol{\beta}_{\tau_1}, \ldots, \boldsymbol{\beta}_{\tau_p})^T$, $\boldsymbol{z}_{it}$ follow the p-variate standard normal distribution $N_p(0, \boldsymbol{I}_p)$, w_{it} follow the standard exponential distribution Exp(1) and are independent of $\boldsymbol{z}_{it}$. For model (3.4), conditioning on latent variable w_{it} and random effect $\boldsymbol{b}_i$, $\boldsymbol{y}_{it}$ follows the multivariate normal distribution with the mean $\boldsymbol{\beta}_\tau \boldsymbol{x}_{it} + \boldsymbol{b}_i + \boldsymbol{D}\boldsymbol{\theta} w_{it}$ and variance–covariane matrix $w_{it}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}$, namely

$$\boldsymbol{y}_{it} | \boldsymbol{x}_{it}, \boldsymbol{b}_i, w_{it} \sim N(\boldsymbol{\beta}_\tau \boldsymbol{x}_{it} + \boldsymbol{b}_i + \boldsymbol{D}\boldsymbol{\theta} w_{it}, w_{it}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}) \tag{3.5}$$

3.2 The complete joint hierarchical likelihood

Denote $\boldsymbol{\Theta} = \{\boldsymbol{\beta}_\tau, \boldsymbol{\Psi}, \boldsymbol{D}, \boldsymbol{\Sigma}_{\boldsymbol{b}}\}$, $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$, $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, and $\boldsymbol{W} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N\}$. In model (3.4), the random effect $\boldsymbol{b}_i$ and w_{it} are unobserved latent variables. Hence, the observed log-likelihood function is

$$\begin{aligned} l_{\boldsymbol{\Theta}}(\boldsymbol{Y}|\boldsymbol{X}) &= \log[L_{\boldsymbol{\Theta}}(\boldsymbol{Y}|\boldsymbol{X})] = \sum_{i=1}^{N} \log L_{\boldsymbol{\Theta}}(\boldsymbol{y}_i|\boldsymbol{x}_i) \\ &= \sum_{i=1}^{N} \log \left[\int L_{\boldsymbol{\Theta}}(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{b}_i) \cdot f(\boldsymbol{b}_i) d\boldsymbol{b}_i \right] \\ &= \sum_{i=1}^{N} \log \left\{ \int \left[\int L_{\boldsymbol{\Theta}}(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{b}_i, \boldsymbol{w}_i) \cdot f(\boldsymbol{w}_i) d\boldsymbol{w}_i \right] \cdot f(\boldsymbol{b}_i) d\boldsymbol{b}_i \right\} \end{aligned} \tag{3.6}$$

where $\boldsymbol{y}_i = \{\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in_i}\}$, $\boldsymbol{x}_i = \{\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i}\}$, $\boldsymbol{w}_i = \{w_{i1}, \ldots, w_{in_i}\}$, $L_{\boldsymbol{\Theta}}(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{b}_i, \boldsymbol{w}_i)$ is the pdf of $\boldsymbol{y}_i$ conditionally on $\boldsymbol{b}_i$ and $\boldsymbol{w}_i$.

It is difficult to maximize the above marginal log-likelihood (3.6). We address this problem using the MCMC algorithm in Bayesian framework. We firstly present the joint hierarchical working likelihood of the complete data $\{\boldsymbol{Y}, \boldsymbol{b}, \boldsymbol{W}\}$ for model (3.4) as follows:

$$\begin{aligned} & L_C(\boldsymbol{Y}, \boldsymbol{b}, \boldsymbol{W}|\boldsymbol{X}, \boldsymbol{\Theta}) \\ &= \prod_{i=1}^{N} \left\{ \prod_{t=1}^{n_i} \left[f(\boldsymbol{y}_{it}|\boldsymbol{b}_i, w_{it}) \cdot f(w_{it}) \right] \cdot f(\boldsymbol{b}_i) \right\} \\ &= \prod_{i=1}^{N} \prod_{t=1}^{n_i} \left\{ \frac{(2\pi)^{-p/2}}{|w_{it}\boldsymbol{D}\boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{\Delta}\boldsymbol{D}|^{1/2}} \exp\left[-\frac{1}{2} (\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it})^T (w_{it}\boldsymbol{D}\boldsymbol{\Delta}\boldsymbol{\Psi}\boldsymbol{\Delta}\boldsymbol{D})^{-1} (\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it}) \right] \right. \\ &\quad \left. \cdot \exp(-w_{it}) \right\} \cdot \prod_{i=1}^{N} f(\boldsymbol{b}_i) \end{aligned} \tag{3.7}$$

where $\boldsymbol{\mu}_{it} = \boldsymbol{\beta}_\tau \boldsymbol{x}_{it} + \boldsymbol{b}_i + \boldsymbol{D}\boldsymbol{\theta} w_{it}$.

4 Bayesian estimation approach

4.1 Prior specifications

To select important covariates for improving prediction accuracy, various regularized penalization methods are used to conduct variable selection. Commonly used penalty functions mainly include LASSO (least absolute shrinkage and selection operator) penalty, ridge penalty, SCAD (smoothly clipped absolute deviation) penalty as well as bridge penalty. More discussion on LASSO penalty and Bayesian LASSO can be found by Tibshirani,[44] Zou,[45] Park and Casella,[46] Leng,[47] etc. This article considers Bayesian adaptive LASSO regularization of regression parameters in model (3.2). In order to do this, we impose the Laplace priors for regression parameters β_τ as follows:

$$\pi(\boldsymbol{\beta}_\tau) = \prod_{j=1}^{p} \pi(\boldsymbol{\beta}_{\tau_j}),\ \pi(\boldsymbol{\beta}_{\tau_j}) = \prod_{s=1}^{k} \pi(\beta_{js}|\lambda_{js}), \pi(\beta_{js}|\lambda_{js}) = \frac{\lambda_{js}}{2} \exp\left\{ -\lambda_{js}|\beta_{js}| \right\} \tag{4.1}$$

where $\boldsymbol{\lambda} = \{\lambda_{js}\}$, $\lambda_{js} > 0$ are tuning parameters.

However, the prior in (4.1) is analytically intractable to calculate the desirable posterior quantities. Using the mixture representation approach by Mallick and Yi (2018), we decompose the prior of $\beta_{j,s}$ as the following uniform–gamma mixture representation,

$$\pi(\beta_{js}|\lambda_{js}) = \int_0^\infty \pi(\beta_{js}|h_{js}) \cdot \pi(h_{js}|\lambda_{js}) dh_{js}$$

where $\pi(\beta_{js}|h_{js}) = \text{Uniform}(-h_{js}, h_{js})$, $\pi(h_{js}|\lambda_{js}) = \text{Gamma}(2, \lambda_{js})$.

The joint prior of $\boldsymbol{\beta}_\tau$ can be represented as

$$\pi(\boldsymbol{\beta}_\tau|\boldsymbol{H},\lambda) \propto \prod_{j=1}^{p}\prod_{s=1}^{k}[\pi(\beta_{js}|h_{js})\cdot\pi(h_{js}|\lambda_{js})]$$

where $\boldsymbol{H} = \{h_{js}\}$.

The prior of $\boldsymbol{\Sigma_b}$ is set as inverse Wishart distribution: $\pi(\boldsymbol{\Sigma_b}) \sim \text{IW}(m_b, \boldsymbol{\Phi_b})$.

The prior of $\boldsymbol{\Psi}$ is set as inverse Wishart distribution: $\pi(\boldsymbol{\Psi}) \sim \text{IW}(m_0, \boldsymbol{\Phi_0})$.

The prior of $\boldsymbol{D}$ is assumed to be the following informative prior

$$\pi(\boldsymbol{D}) = \pi(\sigma_1,\ldots,\sigma_p) \propto \prod_{j=1}^{p}\frac{1}{\sigma_j}$$

The prior of $\boldsymbol{\lambda}$ is set as $\pi(\lambda) = \prod_{j=1}^{p}\prod_{s=1}^{k}\pi(\lambda_{js})$, where $\lambda_{js} \sim \text{Gamma}(c_{js}, d_{js})$.

Hence, the joint hierarchical prior of all parameters is

$$\pi(\boldsymbol{\Theta},\lambda) \propto \pi(\boldsymbol{\beta}_\tau|\boldsymbol{H},\lambda)\pi(\boldsymbol{\Sigma_b})\pi(\boldsymbol{\Psi})\pi(\boldsymbol{D})\pi(\lambda) \tag{4.2}$$

4.2 The fully conditional posteriors and Gibbs sampling algorithm

Incorporating joint prior (4.2) into the joint working likelihood (3.7) results in the joint posterior density of all parameters as follows:

$$\pi(\boldsymbol{\Theta},\lambda,\boldsymbol{b},\boldsymbol{W},\boldsymbol{Y}|\boldsymbol{X}) \propto L_C(\boldsymbol{Y},\boldsymbol{b},\boldsymbol{W}|\boldsymbol{X},\boldsymbol{\Theta})\cdot\pi(\boldsymbol{\Theta},\lambda) \tag{4.3}$$

Gibbs sampler procedures are employed to carry MCMC algorithm. The hierarchical expression of posterior distributions of all unknown parameters and latent variables can be presented as follows:

$$\begin{cases} \boldsymbol{Y}|\boldsymbol{W},\boldsymbol{b} \sim \prod_{i=1}^{N}\prod_{t=1}^{n_i} N_p(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it} + \boldsymbol{b}_i + \boldsymbol{D\theta}w_{it}, w_{it}\boldsymbol{D\Sigma D}) \\ \boldsymbol{W} \sim \prod_{i=1}^{N}\prod_{t=1}^{n_i}\text{Exp}(1) \\ \boldsymbol{D} \sim \prod_{j=1}^{p}\frac{1}{\sigma_j} \\ \boldsymbol{b} \sim \prod_{i=1}^{N} N_p(0,\boldsymbol{\Sigma_b}) \\ \boldsymbol{\beta}_\tau|\boldsymbol{H} \sim \prod_{j=1}^{p}\prod_{s=1}^{k}\text{Uniform}(-h_{js}, h_{js}) \\ \boldsymbol{H}|\lambda \sim \prod_{j=1}^{p}\prod_{s=1}^{k}\text{Gamma}(2,\lambda_{js}) \\ \boldsymbol{\Psi} \sim IW(m_0,\boldsymbol{\Phi_0}) \\ \boldsymbol{\Sigma_b} \sim IW(m_b,\boldsymbol{\Phi_b}) \\ \boldsymbol{\lambda} \sim \prod_{j=1}^{p}\prod_{s=1}^{k}\text{Gamma}(c_{js},d_{js}) \end{cases} \tag{4.4}$$

Let $\boldsymbol{\Theta}\_$ denote the remaining parameters subset apart from the present sample parameter. The full conditional posterior distributions of unknown parameters and latent variables can be presented as follows, respectively.

• Sample $\boldsymbol{\beta}_\tau$ from the truncated matrix normal distribution

$$N_{p\times k}(\boldsymbol{M},\boldsymbol{\Phi}\otimes\boldsymbol{V})\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\text{I}(|\beta_{js}| < h_{js})$$

In terms of the properties of matrix normal distribution, we have

$$\text{Vec}(\boldsymbol{\beta}_\tau) \sim N_{pk}(\text{Vec}(\boldsymbol{M}),\boldsymbol{\Phi}\otimes\boldsymbol{V})\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\text{I}(|\beta_{js}| < h_{js})$$

Specially, we can marginally sample the component β_{js} of $\text{Vec}(\boldsymbol{\beta}_\tau)$ via the following truncated normal distribution:

$$N(\text{Vec}(\boldsymbol{M})_{(s-1)p+j},(\boldsymbol{\Phi}\otimes\boldsymbol{V})_{(s-1)p+j,(s-1)p+j})\cdot\text{I}(|\beta_{js}| < h_{js}), j = 1,\ldots,p, s = 1,\ldots,k$$

The theoretical posterior distribution of $\boldsymbol{\beta}_\tau$ is derived as follows:

$$
\begin{aligned}
&\pi(\boldsymbol{\beta}_\tau|\boldsymbol{\Theta}\_) \propto L_C(\boldsymbol{Y},\boldsymbol{W}|\boldsymbol{X},\boldsymbol{\beta}_\tau,\boldsymbol{D},\boldsymbol{\Psi})\cdot\pi(\boldsymbol{\beta}_\tau|\boldsymbol{H})\\
&\quad\propto \exp\Big\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{n_i}(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it}-\boldsymbol{\eta}_{it})^T(w_{it}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it}-\boldsymbol{\eta}_{it})\Big\}\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\mathrm{I}(|\beta_{js}|<h_{js})\\
&\quad\propto \exp\Big\{-\frac{1}{2}tr\Big(\sum_{i=1}^{N}\sum_{t=1}^{n_i}(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it}-\boldsymbol{\eta}_{it})^T(w_{it}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it}-\boldsymbol{\eta}_{it})\Big)\Big\}\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\mathrm{I}(|\beta_{js}|<h_{js})\\
&\quad\propto \exp\Big\{-\frac{1}{2}tr\Big(\sum_{i=1}^{N}\sum_{t=1}^{n_i}\Big[(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it})^T(w_{it}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it})-2\boldsymbol{\eta}_{it}^T(w_{it}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}(\boldsymbol{\beta}_\tau\boldsymbol{x}_{it})\Big]\Big)\Big\}\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\mathrm{I}(|\beta_{js}|<h_{js})\\
&\quad\propto \exp\Big\{-\frac{1}{2}tr\Big[\Big(\sum_{i=1}^{N}\sum_{t=1}^{n_i}w_{it}^{-1}\boldsymbol{x}_{it}\boldsymbol{x}_{it}^T\Big)\cdot(\boldsymbol{\beta}_\tau-\boldsymbol{M})^T(\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}(\boldsymbol{\beta}_\tau-\boldsymbol{M})\Big]\Big\}\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\mathrm{I}(|\beta_{js}|<h_{js})\\
&\quad\propto \exp\Big\{-\frac{1}{2}tr\Big(\Phi^{-1}\cdot(\boldsymbol{\beta}_\tau-\boldsymbol{M})^T\boldsymbol{V}^{-1}(\boldsymbol{\beta}_\tau-\boldsymbol{M})\Big)\Big\}\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\mathrm{I}(|\beta_{js}|<h_{js})\\
&\quad\sim N_{p\times k}(\boldsymbol{M},\boldsymbol{\Phi}\otimes\boldsymbol{V})\cdot\prod_{j=1}^{p}\prod_{s=1}^{k}\mathrm{I}(|\beta_{js}|<h_{js})
\end{aligned}
$$

where $N_{p\times k}(\boldsymbol{M},\boldsymbol{\Phi}\otimes\boldsymbol{V})$ denotes a $p\times k$ matrix normal distribution with parameters $\boldsymbol{M}$, $\boldsymbol{\Phi}$, $\boldsymbol{V}$, and $\boldsymbol{\Phi}=(\sum_{i=1}^{N}\sum_{t=1}^{n_i}w_{it}^{-1}\boldsymbol{x}_{it}\boldsymbol{x}_{it}^T)^{-1}$, $\boldsymbol{M}=\sum_{i=1}^{N}\sum_{t=1}^{n_i}(w_{it}^{-1}\boldsymbol{\eta}_{it}\boldsymbol{x}_{it}^T)\cdot\boldsymbol{\Phi}$, $\boldsymbol{V}=\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}$, and $\boldsymbol{\eta}_{it}=\boldsymbol{y}_{it}-\boldsymbol{b}_i-\boldsymbol{D}\boldsymbol{\theta}w_{it}$.

• Sample σ_j from the following posterior distribution:

$$
\begin{aligned}
&\pi(\sigma_j|\Theta\_)\\
&\quad\propto|\boldsymbol{D}|^{-(\sum_{i=1}^{N}n_i+1)}\exp\Big\{-\frac{1}{2}tr\Big[\sum_{i=1}^{N}\sum_{t=1}^{n_i}(\boldsymbol{e}_{it}-\boldsymbol{D}\boldsymbol{\theta}w_{it})^T(w_{it}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}(\boldsymbol{e}_{it}-\boldsymbol{D}\boldsymbol{\theta}w_{it})\Big]\Big\}\\
&\quad\propto\sigma_j^{-(\sum_{i=1}^{N}n_i+1)}\cdot\exp\Big\{-\frac{1}{2}tr\Big[\sum_{i=1}^{N}\sum_{t=1}^{n_i}(w_{it}^{-1}\boldsymbol{e}_{it}\boldsymbol{e}_{it}^T)\cdot(\boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}^{-1})-\boldsymbol{\theta}(\sum_{i=1}^{N}\sum_{t=1}^{n_i}\boldsymbol{e}_{it}^T)\cdot\boldsymbol{\Sigma}^{-1}\boldsymbol{D}^{-1}\Big]\Big\}
\end{aligned}
$$

where $\boldsymbol{e}_{it}=\boldsymbol{y}_{it}-\boldsymbol{\beta}_\tau\boldsymbol{x}_{it}-\boldsymbol{b}_i$.

Noting that $\pi(\sigma_j|\boldsymbol{\Theta}\_)$ is not a standard distribution, we can sample σ_j using the MH (Metropolis-Hastings) algorithm by the following steps:

(I) Generate a random number $\sigma_j^{candidate}$ from the truncated normal distribution $\mathrm{TN}(0,\gamma^2,0,\infty)$ which is specified as the proposal distribution in the MH algorithm, where γ^2 denotes the hyperparameter of variance and $(0,\infty)$ is the sampling interval of the truncated normal distribution. The pdf of the proposal distribution is denoted as $q(\cdot|\boldsymbol{\Theta}\_)$.

(II) Generate a random number u from the standard uniform distribution $U[0,1]$.

(III) Compute the acceptance probability $r=\min\{1,\frac{q(\sigma_j^{old}|\boldsymbol{\Theta}\_)}{q(\sigma_j^{candidate}|\boldsymbol{\Theta}\_)}\cdot\frac{\pi(\sigma_j^{candidate}|\boldsymbol{\Theta}\_)}{\pi(\sigma_j^{old}|\boldsymbol{\Theta}\_)}\}$

(IV) If $u<r$, then accept the proposal $\sigma_j\leftarrow\sigma_j^{candidate}$, otherwise reject the proposal and let $\sigma_j\leftarrow\sigma_j^{old}$.

• Sample $\boldsymbol{\Psi}$ from the following inverse Wishart distribution:

$$
IW\Big(\sum_{i=1}^{N}n_i+m_0,\quad(\boldsymbol{\nabla}\boldsymbol{D})^{-1}\cdot\sum_{i=1}^{N}\sum_{t=1}^{n_i}(w_{it}^{-1}\boldsymbol{\alpha}_{it}\boldsymbol{\alpha}_{it}^T)\cdot(\boldsymbol{D}\boldsymbol{\nabla})^{-1}+\boldsymbol{\Phi}_{\mathbf{0}}\Big)
$$

The theoretical posterior distribution of $\boldsymbol{\Psi}$ is as follows:

$$\begin{aligned}
\pi(\boldsymbol{\Psi}|\boldsymbol{\Theta}_{-}) &\propto L_C(\boldsymbol{Y},\boldsymbol{b},\boldsymbol{W}|\boldsymbol{X},\boldsymbol{\beta}_{\tau},\boldsymbol{D},\boldsymbol{\Psi},\boldsymbol{\Sigma}_{\boldsymbol{b}})\cdot\pi(\boldsymbol{\Psi})\\
&\propto \prod_{i=1}^{N}\prod_{t=1}^{n_i}\Big[\frac{(2\pi)^{-p/2}}{|w_{it}\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D}|^{1/2}}\exp\Big\{-\frac{1}{2}\boldsymbol{\alpha}_{it}^{T}(w_{it}\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D})^{-1}\boldsymbol{\alpha}_{it}\Big\}\Big]\frac{|\boldsymbol{\Psi}|^{-\frac{m_0+p+1}{2}}\exp\{-\frac{1}{2}tr(\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}_{\mathbf{0}})\}}{2^{\frac{m_0p}{2}}|\boldsymbol{\Phi}_{\mathbf{0}}|^{-\frac{m_0}{2}}\cdot\Gamma_p(\frac{m_0}{2})}\\
&\propto |\boldsymbol{\Psi}|^{-\frac{\prod_{i=1}^{N}n_i+m_0+p+1}{2}}\exp\Big\{-\frac{1}{2}\Big[\sum_{i=1}^{N}\sum_{t=1}^{n_i}tr\Big(\boldsymbol{\alpha}_{it}^{T}(w_{it}\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D})^{-1}\boldsymbol{\alpha}_{it}\Big)+tr(\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}_{\mathbf{0}})\Big]\Big\}\\
&\propto |\boldsymbol{\Psi}|^{-\frac{\prod_{i=1}^{N}n_i+m_0+p+1}{2}}\exp\Big\{-\frac{1}{2}tr\Big[\boldsymbol{\Psi}^{-1}\Big((\boldsymbol{\nabla}\boldsymbol{D})^{-1}\cdot\sum_{i=1}^{N}\sum_{t=1}^{n_i}(w_{it}^{-1}\boldsymbol{\alpha}_{it}\boldsymbol{\alpha}_{it}^{T})\cdot(\boldsymbol{D}\boldsymbol{\nabla})^{-1}+\boldsymbol{\Phi}_{\mathbf{0}}\Big)\Big]\Big\}\\
&\sim IW\Big(\prod_{i=1}^{N}n_i+m_0,(\nabla D)^{-1}\cdot\sum_{i=1}^{N}\sum_{t=1}^{n_i}(w_{it}^{-1}\boldsymbol{\alpha}_{it}\boldsymbol{\alpha}_{it}^{T})\cdot(\boldsymbol{D}\boldsymbol{\nabla})^{-1}+\boldsymbol{\Phi}_{\mathbf{0}}\Big)
\end{aligned}$$

where $\boldsymbol{\alpha}_{it}=\boldsymbol{y}_{it}-\boldsymbol{\beta}_{\tau}\boldsymbol{x}_{it}-\boldsymbol{b}_i-\boldsymbol{D}\boldsymbol{\theta}w_{it}$.

- Sample w_{it} from the generalized inverse Gaussian distribution

$$\mathrm{GIG}\Big(1-\frac{p}{2},\boldsymbol{e}_{it}^{T}(\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}\boldsymbol{e}_{it},\boldsymbol{\theta}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}+2\Big)$$

The theoretical posterior distribution of w_{it} is derived as follows:

$$\begin{aligned}
&\pi(w_{it}|\boldsymbol{\Theta}_{-})\\
&\quad\propto\frac{1}{|w_{it}\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D}|^{1/2}}\cdot\exp\Big\{-\frac{1}{2}(\boldsymbol{e}_{it}-\boldsymbol{D}\boldsymbol{\theta}w_{it})^{T}(w_{it}\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D})^{-1}(\boldsymbol{e}_{it}-\boldsymbol{D}\boldsymbol{\theta}w_{it})\Big\}\cdot\exp\{-w_{it}\}\\
&\quad\propto|w_{it}|^{-\frac{p}{2}}\cdot\exp\Big\{-\frac{1}{2}(\boldsymbol{e}_{it}-\boldsymbol{D}\boldsymbol{\theta}w_{it})^{T}(w_{it}\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D})^{-1}(\boldsymbol{e}_{it}-\boldsymbol{D}\boldsymbol{\theta}w_{it})\Big\}\cdot\exp\{-w_{it}\}\\
&\quad\propto|w_{it}|^{-\frac{p}{2}}\cdot\exp\Big\{-\frac{1}{2}\Big[\boldsymbol{e}_{it}^{T}(\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D})^{-1}\boldsymbol{e}_{it}\cdot w_{it}^{-1}+\big(\boldsymbol{\theta}^{T}(\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla})^{-1}\boldsymbol{\theta}+2\big)\cdot w_{it}\Big]\Big\}\\
&\quad\sim\mathrm{GIG}\Big(1-\frac{p}{2},\boldsymbol{e}_{it}^{T}(\boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla}\boldsymbol{D})^{-1}\boldsymbol{e}_{it},\boldsymbol{\theta}^{T}(\boldsymbol{\nabla}\boldsymbol{\Psi}\boldsymbol{\nabla})^{-1}\boldsymbol{\theta}+2\Big)
\end{aligned}$$

- Sample λ_{js} from the Gamma distribution: Gamma$(2+c_{js},d_{js}+h_{js})$.
- Sample h_{js} from the left-truncated exponential distribution $\mathrm{Exp}(\lambda_{js})I\{h_{js}>|\beta_{js}|\}$, using the inversion method, which can be conducted by the following two steps:
  (I) Generate $h_{js}^{*}\sim\mathrm{Exp}(\lambda_{js})$.
  (II) Generate $h_{js}=h_{js}^{*}+|\beta_{js}|$.
- Sample $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ from the following inverse Wishart distribution

$$IW\Big(N+m_b,\ \sum_{i=1}^{N}\boldsymbol{b}_i\boldsymbol{b}_i^{T}+\boldsymbol{\Phi}_{\boldsymbol{b}}\Big)$$

The theoretical derivation of the posterior distribution of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ is as follows:

$$\begin{aligned}
\pi(\boldsymbol{\Sigma}_{\boldsymbol{b}}|\boldsymbol{\Theta}_{-})&\propto\prod_{i=1}^{N}\Big[\frac{(2\pi)^{-p/2}}{|\boldsymbol{\Sigma}_{\boldsymbol{b}}|^{1/2}}\exp\Big\{-\frac{1}{2}\boldsymbol{b}_i^{T}{\boldsymbol{\Sigma}_{\boldsymbol{b}}}^{-1}\boldsymbol{b}_i\Big\}\Big]\cdot\frac{|\boldsymbol{\Sigma}_{\boldsymbol{b}}|^{-\frac{m_b+p+1}{2}}\exp\{-\frac{1}{2}tr({\boldsymbol{\Sigma}_{\boldsymbol{b}}}^{-1}\boldsymbol{\Phi}_{\boldsymbol{b}})\}}{2^{\frac{m_bp}{2}}|\boldsymbol{\Phi}_{\boldsymbol{b}}|^{-\frac{m_b}{2}}\cdot\Gamma_p(\frac{m_b}{2})}\\
&\propto|\boldsymbol{\Sigma}_{\boldsymbol{b}}|^{-\frac{N+m_b+p+1}{2}}\exp\Big\{-\frac{1}{2}\Big[\sum_{i=1}^{N}tr\Big({\boldsymbol{b}_i}^{T}{\boldsymbol{\Sigma}_{\boldsymbol{b}}}^{-1}\boldsymbol{b}_i\Big)+tr({\boldsymbol{\Sigma}_{\boldsymbol{b}}}^{-1}\boldsymbol{\Phi}_{\boldsymbol{b}})\Big]\Big\}
\end{aligned}$$

$$\propto |\boldsymbol{\Sigma_b}|^{-\frac{N+m_b+p+1}{2}} \exp\Big\{ -\frac{1}{2} tr\Big[\boldsymbol{\Sigma_b}^{-1}\Big(\sum_{i=1}^{N} \boldsymbol{b}_i\boldsymbol{b}_i^T + \boldsymbol{\Phi_b}\Big)\Big]\Big\}$$

$$\sim IW\Big(N + m_b, \sum_{i=1}^{N} \boldsymbol{b}_i\boldsymbol{b}_i^T + \boldsymbol{\Phi_b}\Big)$$

- Sample $\boldsymbol{b}_i$ from normal distribution $N(\boldsymbol{\mu}^*_{\boldsymbol{b}_i}, \boldsymbol{\Sigma}^*_{\boldsymbol{b}_i})$, where $\boldsymbol{\varsigma}_{it} = \boldsymbol{y}_{it} - \boldsymbol{\beta}_\tau \boldsymbol{x}_{it} - \boldsymbol{D\theta} w_{ij}$, and $\boldsymbol{\mu}^*_{\boldsymbol{b}_i} = \boldsymbol{\Sigma}^*_{\boldsymbol{b}_i} \cdot (\boldsymbol{D\Sigma D})^{-1} \sum_{j=1}^{n_i} w_{it}^{-1} \boldsymbol{\varsigma}_{it}$, $\boldsymbol{\Sigma}^*_{\boldsymbol{b}_i} = ((\boldsymbol{D\Sigma D})^{-1} \sum_{j=1}^{n_i} w_{it}^{-1} + \boldsymbol{\Sigma_b}^{-1})^{-1}$.

For carrying out a Bayesian analysis, an efficient MCMC algorithm is used to sample $\boldsymbol{\beta}_\tau, \boldsymbol{\Psi}, \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{b}, \boldsymbol{\Sigma_b}$, and $\boldsymbol{\lambda}$ from the above conditional posterior distributions. In particular, for correlation matrix $\boldsymbol{\Psi}$, we draw its sample from the posterior inverse Wishart distribution, and then standardize it as the correlation coefficient matrix.

5 Simulation studies

In this section, we investigate the performance of the proposed Bayesian procedures by conducting some Monte Carlo simulations. We repeatedly generate 50 data sets from the following three-variate longitudinal mixed model

$$\boldsymbol{y}_{it} = \boldsymbol{\beta x}_{it} + \boldsymbol{b}_i + \boldsymbol{e}_{it}, i = 1, \ldots, N, t = 1, \ldots, m \tag{5.1}$$

where $N = 20$ and equal cluster sizes $m = 5$ are considered to illustrate the finite-sample performance. In model (5.1), the elements of $\boldsymbol{x}_{it}$ are independently generated from the standard normal distribution, random effects $\boldsymbol{b}_i$ is drawn from three-variate normal distribution with zero mean and covariance matrix $\boldsymbol{\Sigma_b}$ of dimension 3×3. $\boldsymbol{\Sigma_b}$ is constructed by simulating from a Wishart distribution with four degrees of freedom and a diagonal scale matrix with element 0.1.

In model (5.1), the true parameter matrix $\boldsymbol{\beta}$ is set as the following cases:

Model 1: Dense case:

$$\boldsymbol{\beta}_{3\times3} = \begin{pmatrix} -0.382 & -0.372 & 0.715 \\ 1.993 & 0.650 & 0.764 \\ 0.670 & 1.079 & 0.584 \end{pmatrix}$$

Model 2: Sparse case:

$$\boldsymbol{\beta}_{3\times3} = \begin{pmatrix} -0.382 & 0 & 0.715 \\ 0 & 0.650 & 0 \\ 0.670 & 0 & 0 \end{pmatrix}$$

Model 3: Extremely sparse case:

$$\boldsymbol{\beta}_{3\times10} = \begin{pmatrix} -0.382 & 0 & 0.715 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.650 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.670 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We set elements of matrices $\boldsymbol{D}$ and $\boldsymbol{\Psi}$ as $\sigma_1 = 0.13, \sigma_2 = 0.30, \sigma_3 = 0.23$, and $\rho_{12} = 0.5, \rho_{13} = 0.3, \rho_{23} = 0.4$. Three cases of quantile levels, that is, $\tau = (0.5, 0.5, 0.5)$, $\tau = (0.25, 0.5, 0.75)$, and $\tau = (0.75, 0.5, 0.25)$, are considered for the three models. Two different distributions for the error term $\boldsymbol{e}_{it}$ in model (5.1) are considered as follows. Case I: A multivariate normal distribution (MN) with zero mean and a variance–covariance matrix equal to $\boldsymbol{D\theta\theta}^T\boldsymbol{D} + \boldsymbol{D\Sigma D}$; Case II: A multivariate student t distribution with three degrees of freedom (Mt_3), non-centrality parameter $\boldsymbol{D\theta}$ and scale parameter equal to $\boldsymbol{D\Sigma D}$. An illustration of convergence diagnosis of the MCMC algorithm is shown in Section 5.1. The substantive calculations of the considered three models are implemented in Section 5.2. Section 5.3 provides some additional simulations.

Simulation studies in Section 5 and real-world data analysis in Section 6 are conducted on a Dell desktop [OptiPlex 7050, Intel(R) Core(TM) i7-7700U CPU] via statistical software R3.5.2. All codes of simulations and computations in this article can be requested on the first author. In addition, about the computation time, we conduct a test by taking the setting

of $\boldsymbol{e}_{it} \sim \mathrm{Mt}_3$, $N = 20, m = 5$, and $\tau = (0.25, 0.5, 0.75)$ as an example. The computing times of the proposed joint QR approach for accomplishing one replication are 2.85 min for Model 1, 2.84 min for Model 2, and 2.89 min for Model 3, respectively.

5.1 Convergence diagnosis

To guide the MCMC convergence, we carry out a few test runs under different initial values using the joint QR approach based on the settings of Model 2, and $\boldsymbol{e}_{it} \sim \mathrm{Mt}_3$ (Case II) and $\tau = (0.25, 0.5, 0.75)$. The hyperparameters of the priors discussed in Section 4 are set as follows: $c = d = 0.1, m_b = 6, m_0 = 4, \boldsymbol{\Phi_0} = \mathrm{diag}(1, \ldots, 1)$, and $\boldsymbol{\Phi_b} = \mathrm{diag}(0.1, \ldots, 0.1)$. We consider three groups of initial values for parameters $\boldsymbol{\beta}$, $\sigma_j, j = 1, 2, 3$ and $\boldsymbol{\Psi}$ as follows:

$$\textbf{Initial values 1: } \boldsymbol{\beta}_{3\times3}^{(0)} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{\Psi}^{(0)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \sigma_j^{(0)} = 0.5, j = 1, 2, 3;$$

$$\textbf{Initial values 2: } \boldsymbol{\beta}_{3\times3}^{(0)} = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}, \quad \boldsymbol{\Psi}^{(0)} = \begin{pmatrix} 1 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{pmatrix}, \quad \sigma_j^{(0)} = 5, j = 1, 2, 3;$$

$$\textbf{Initial values 3: } \boldsymbol{\beta}_{3\times3}^{(0)} = \begin{pmatrix} -2 & -2 & -2 \\ -2 & -2 & -2 \\ -2 & -2 & -2 \end{pmatrix}, \quad \boldsymbol{\Psi}^{(0)} = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}, \quad \sigma_j^{(0)} = 10, j = 1, 2, 3.$$

Initial values of other parameters were simply set as $\boldsymbol{b}_i^{(0)} \sim N_p(0, \boldsymbol{I}_p)$, $\boldsymbol{\Sigma_b}^{(0)} = \boldsymbol{I}_p$, $\boldsymbol{\Psi^{(0)}} = I_p$, $\lambda_{js}^{(0)} \sim Exp(1)$, $h_{js}^{(0)} \sim Exp(1)$, and $w_{it}^{(0)} \sim Exp(1)$.

To be conservative, for each simulation, we run the Gibbs sampling algorithm 8000 iterations to assess the convergence of the MCMC algorithm. The MCMC trace plots under three initial values are displayed in Figures 1 and 2. It can be seen in Figure 1, three MCMC chains of regression coefficients starting from the above three initial values are mixed rapidly which shows a sufficient convergence of the algorithm. Figure 2 presents the trace plots of all 8000 posterior iterations of one MCMC chain for all regression coefficients under the setting of **Initial values 1**, which reconfirm that the MCMC chains rapidly converge to their stationary distributions. We also depict the autocorrelation function (ACF) plots in Figure 3 by discarding the first 2000 burn-in iterations to check the autocorrelation between stationary posterior samples under the setting of **Initial values 1**.

5.2 Substantive simulations

Parameters estimation and variable selection using the proposed Bayesian joint QR approach are conducted for dense Model 1 and sparse Models 2 and 3. The initial values of all parameters are set as the case of **Initial values 1**. The hyperparameters of priors are taken as the same values in Section 5.1. To illustrate the superiority of the joint QR approach, we also provide the estimation results of the single QR approach for each response for the aim of comparison. Fifty repeated simulations are conducted by running the Gibbs sampling algorithm for each model and each quantile combination. For each case, we run 8000 iterations of Gibbs sampling algorithm for all parameters and latent variables, the first 2000 burn-in iterations are discarded and the remaining 6000 stationary iterations are retained to conduct posterior inference. Based on 50 repeated simulations, the averaged estimation biases (Bias) and root mean square error (RMSE) of regression parameters for different settings are reported in Tables 1 to 3. A total of 95% credible intervals for the regression coefficients are omitted here. For sparse models, we select the important covariates based on the sizes of estimated values of regression coefficients compared with a predetermined threshold value. Covariates with the absolute values of coefficients are greater than the threshold value are specified as important or "significant" predictors. The threshold value in simulations is consistently taken as 0.1 for all cases. Variable selection results based on various settings are reported in Tables 4 to 6, where"NC" denotes the average correctly identified number of important covariates, "NIC" denotes the average wrongly identified number of unimportant

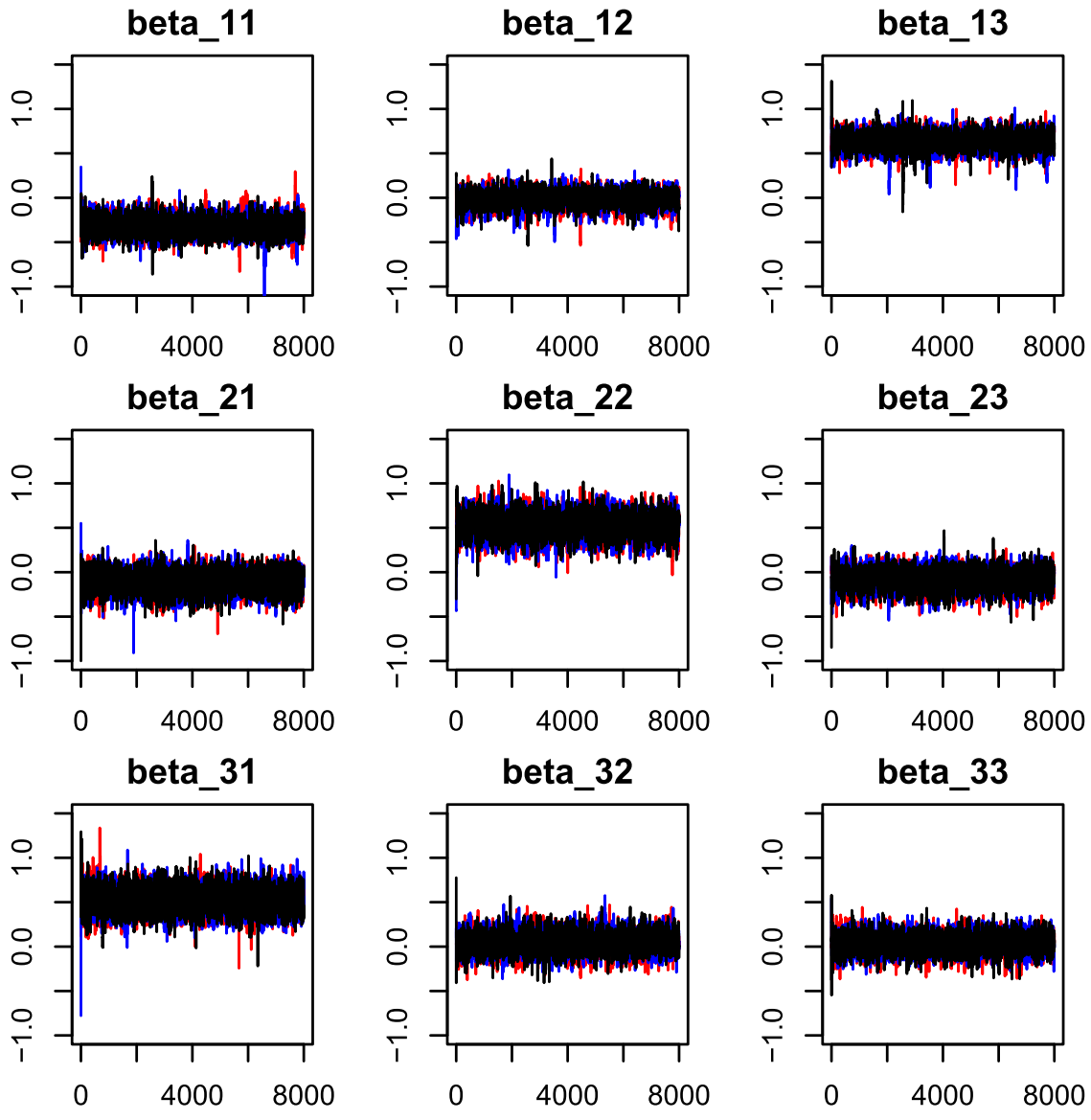

Figure 1. MCMC chains starting from different initial values of the proposed joint QR. *Note:* The red line denotes the first setting of initial values; the blue line denotes the second setting; and the black line denotes the third setting. MCMC: Markov chain Monte Carlo; QR: quantile regression.

covariates. The averaged posterior mean square error (APMSE) of the identified model for 50 simulations is given by

$$\text{APMSE} = \frac{1}{50}\sum_{h=1}^{50} \mathbf{tr}[(\hat{\boldsymbol{\beta}}_{\tau}^{(h)} - \boldsymbol{\beta}_{\tau}^{(true)})(\hat{\boldsymbol{\beta}}_{\tau}^{(h)} - \boldsymbol{\beta}_{\tau}^{(true)})^{T}] \tag{5.2}$$

where $\hat{\boldsymbol{\beta}}^{(h)}$ is the h-th estimated value.

From Tables 1 to 3, we find that the joint QR modeling approach totally performs superior to the single QR approach. Although both two approaches yield unbiased estimates, the former apparently produces more accurate estimates with smaller RMSEs for all considered settings. For variable selection, we observe that both joint and single QR approaches yield the same NC values for the dense Model 1 (see Table 4) except that our approach produces smaller APMSE values. For Models 2 and 3, the joint QR approach obviously gives significantly smaller APMSE and NIC values than those of the single QR approach for the settings under consideration (see Tables 5 and 6). Simulation results show that the joint

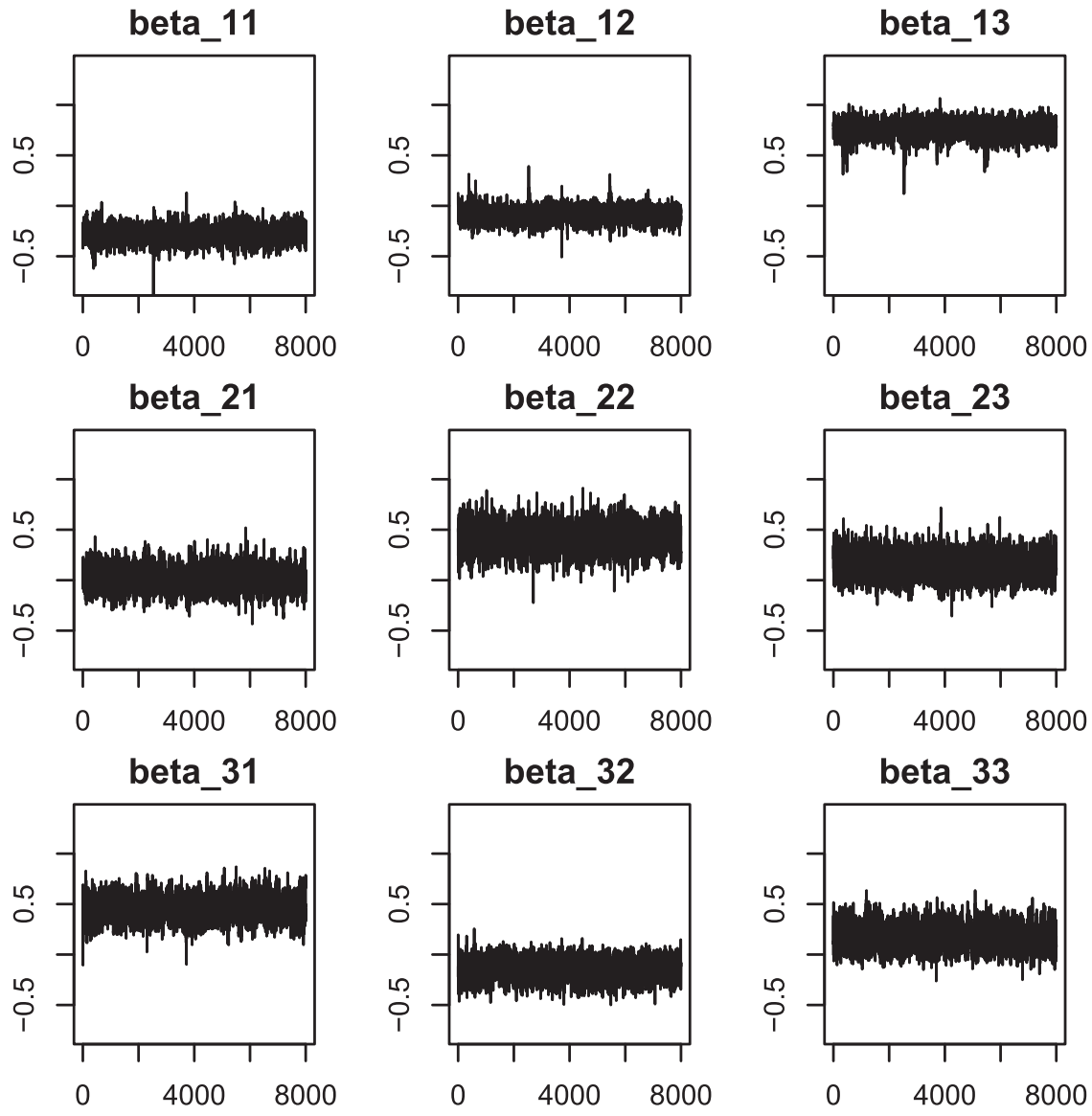

Figure 2. Markov chain Monte Carlo (MCMC) trace plots.

QR approach is more efficient than the single QR approach for both dense models and sparse models for different quantile combinations and error distributions.

5.3 Additional simulations

We implement additional simulations to assess the performances of other settings in this subsection. In the following, the specifics of the settings are indicated. All other conditions are identical to Section 5.2.

Model 4 (high correlation case): $p = 3, K = 3$, $\tau = (0.25, 0.5, 0.75)$, $\boldsymbol{e}_{it} \sim$ MN, and

$$\boldsymbol{\beta}_{3\times3} = \boldsymbol{\beta}^{*}_{3\times3} = \begin{pmatrix} -0.382 & 0 & 0.715 \\ 0 & 0.650 & 0 \\ 0.670 & 0 & 0 \end{pmatrix}$$

where $\boldsymbol{x}_{it} \sim N(0, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is a variance-covariance matrix with off-diagonal elements 0.5.

Model 5 (high-dimensional case): $p = 3, K = 50$, $\tau = (0.25, 0.5, 0.75)$, $\boldsymbol{e}_{it} \sim$ MN, the submatrix consisting of the first three columns of $\boldsymbol{\beta}$ is $\boldsymbol{\beta}^{*}_{3\times3}$ in Model 4, other elements are zero. $\boldsymbol{x}_{it} \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an identity matrix.

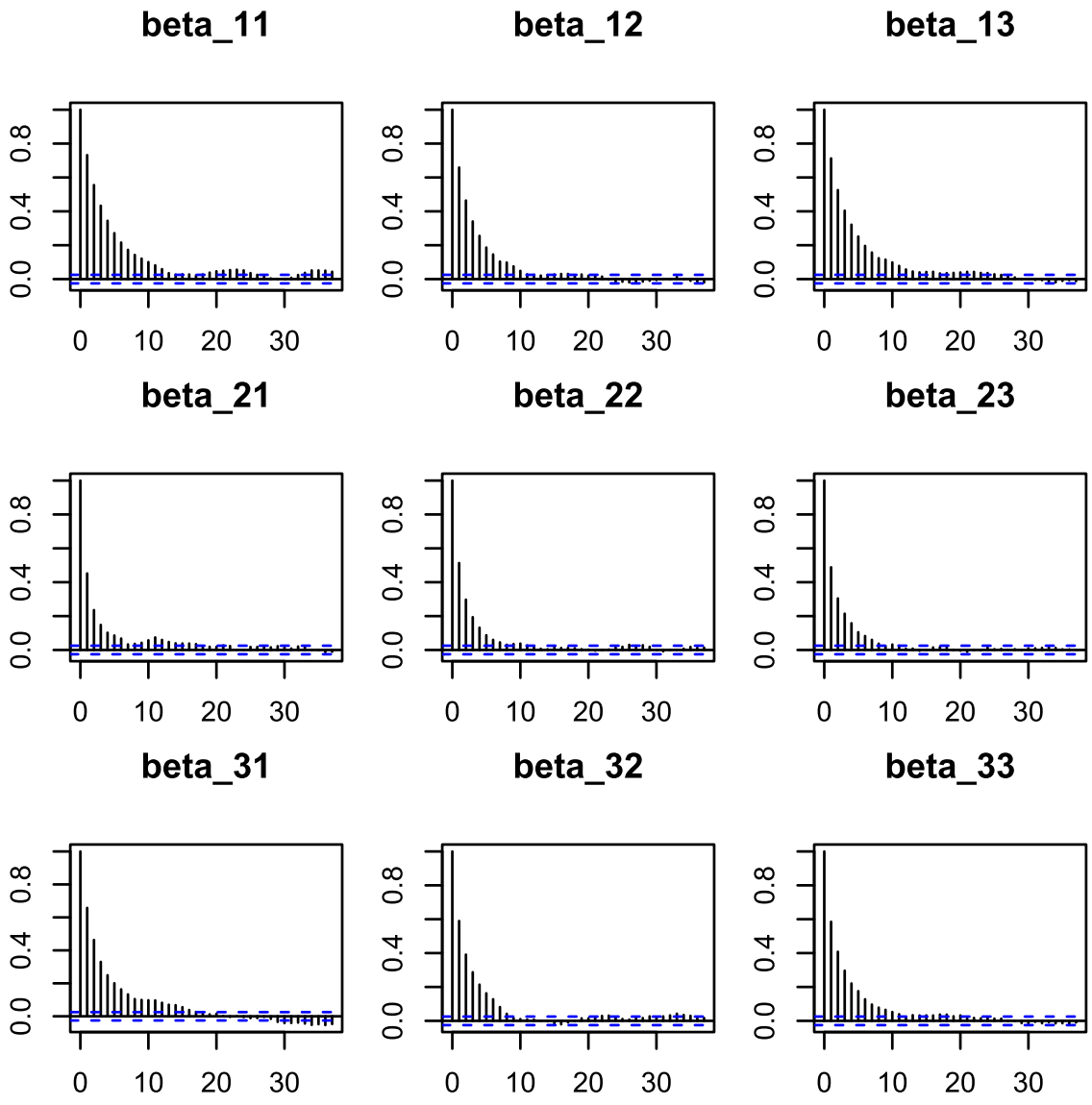

Figure 3. Autocorrelation function (ACF) plots.

Model 6 (ultra high-dimensional case): $p = 3, K = 200$, $\tau = (0.25, 0.5, 0.75)$, $\boldsymbol{e}_{it} \sim$ MN, the submatrix consisting of the first three columns of $\boldsymbol{\beta}$ is $\boldsymbol{\beta}^*_{3\times3}$ in Model 4, other elements are zero. $\boldsymbol{x}_{it} \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an identity matrix.

Under the same initial values, priors and Gibbs sampling algorithm in Section 5.2, we implement the simulation tests for Models 4 to 6 for the joint QR approach. Compared with Models 2 and 3, the performance of the proposed approach is equally good in Model 4, unsatisfactory in Model 5, and breaking down in Model 6. The computation results for Models 4 to 6 are omitted here. For high-dimensional and ultra-high-dimensional cases of Models 5 and 6, we can incorporate Bayesian feature screening algorithm into the joint QR approach to improve the unsatisfactory endings. However, we do not discuss this issue here.

6 Multivariate longitudinal analysis of PBCseq cohort study

In this section, we analyze a subset of longitudinal data on PBCseq cohort study using the proposed annroach. A total of 312 patients were recruited from the Mayo Clinic between January 1974 and May 1984, and participated in either of two double-blind, placebo-controlled, randomized trials with D-penicillamine for treating primary biliary cirrhosis until April 1988. A clinical laboratory database which comprised ID number, time-dependent variables (age and total number of follow-up days), categorical variables (sex, drug, and status), and two continuous measurement variables (natural logarithm scale of

Table 1. Estimation results of Model 1.

Error	Methods	Evaluation	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{33}
		τ = (0.5, 0.5, 0.5)									
MN	Joint QR	Bias	0.002	0.005	0.008	0.029	−0.046	−0.011	0.002	−0.030	−0.010
		RMSE	0.047	0.044	0.048	0.113	0.112	0.111	0.084	0.084	0.080
	Single QR	Bias	−0.030	−0.040	−0.102	−0.873	−0.137	−0.130	−0.326	−0.288	−0.105
		RMSE	0.086	0.092	0.128	0.876	0.158	0.160	0.345	0.300	0.151
Mt_3	Joint QR	Bias	0.005	−0.003	−0.011	−0.009	−0.041	−0.020	−0.021	−0.014	−0.024
		RMSE	0.049	0.053	0.053	0.118	0.134	0.109	0.083	0.090	0.120
	Single QR	Bias	−0.027	−0.035	−0.141	−0.879	−0.142	−0.129	−0.353	−0.286	−0.128
		RMSE	0.075	0.076	0.163	0.883	0.195	0.181	0.376	0.300	0.202
		τ = (0.25, 0.5, 0.75)									
MN	Joint QR	Bias	0.014	0.054	−0.019	0.004	−0.012	−0.005	−0.015	−0.037	−0.031
		RMSE	0.080	0.118	0.115	0.111	0.096	0.105	0.136	0.164	0.129
	Single QR	Bias	0.004	0.026	−0.201	−0.876	−0.151	−0.191	−0.282	−0.335	−0.157
		RMSE	0.090	0.124	0.221	0.880	0.170	0.210	0.315	0.355	0.195
Mt_3	Joint QR	Bias	0.021	0.045	−0.014	−0.035	−0.008	−0.003	−0.061	−0.055	−0.020
		RMSE	0.083	0.094	0.099	0.108	0.137	0.118	0.142	0.145	0.121
	Single QR	Bias	0.050	0.029	−0.203	−0.913	−0.158	−0.193	−0.319	−0.373	−0.158
		RMSE	0.092	0.106	0.229	0.916	0.207	0.231	0.348	0.395	0.195
		τ = (0.75, 0.5, 0.25)									
MN	Joint QR	Bias	0.049	0.032	−0.029	−0.008	−0.030	−0.032	−0.077	−0.030	−0.054
		RMSE	0.099	0.099	0.087	0.096	0.107	0.136	0.151	0.179	0.156
	Single QR	Bias	0.039	0.012	−0.206	−0.916	−0.162	−0.207	−0.341	−0.316	−0.173
		RMSE	0.099	0.110	0.218	0.918	0.183	0.239	0.364	0.342	0.209
Mt_3	Joint QR	Bias	0.019	0.011	−0.026	−0.013	−0.036	0.006	0.008	−0.032	−0.010
		RMSE	0.080	0.080	0.110	0.102	0.146	0.117	0.113	0.119	0.115
	Single QR	Bias	0.027	0.021	−0.221	−0.889	−0.178	−0.210	−0.312	−0.386	−0.193
		RMSE	0.106	0.101	0.247	0.891	0.207	0.243	0.350	0.401	0.226

QR: quantile regression; MN: multivariate normal distribution; RMSE: root mean square error; Bias: biases.

Table 2. Estimation results of sparse Model 2.

Error	Methods	Evaluation	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{33}
		$\tau = (0.5, 0.5, 0.5)$									
MN	Joint QR	Bias	0.006	−0.001	0.002	−0.023	0.008	0.001	−0.024	0.004	0.008
		RMSE	0.052	0.030	0.049	0.095	0.102	0.074	0.091	0.078	0.059
	Single QR	Bias	0.028	−0.034	−0.270	−0.032	−0.256	−0.072	−0.208	−0.056	−0.016
		RMSE	0.071	0.079	0.274	0.129	0.261	0.121	0.218	0.091	0.087
Mt_3	Joint QR	Bias	0.003	0.000	−0.017	−0.006	−0.011	−0.003	−0.024	−0.005	−0.001
		RMSE	0.059	0.049	0.060	0.086	0.115	0.089	0.086	0.060	0.070
	Single QR	Bias	0.030	−0.022	−0.254	−0.030	−0.250	−0.094	−0.239	−0.059	−0.015
		RMSE	0.075	0.078	0.262	0.122	0.261	0.144	0.252	0.094	0.087
		$\tau = (0.25, 0.5, 0.75)$									
MN	Joint QR	Bias	0.060	−0.003	−0.043	0.022	−0.033	−0.006	−0.076	−0.020	−0.003
		RMSE	0.136	0.053	0.119	0.078	0.114	0.077	0.205	0.098	0.111
	Single QR	Bias	0.093	−0.021	−0.335	0.042	−0.250	−0.058	−0.257	−0.051	−0.025
		RMSE	0.150	0.076	0.343	0.091	0.258	0.109	0.284	0.124	0.117
Mt_3	Joint QR	Bias	0.013	−0.004	−0.033	−0.035	−0.047	0.007	−0.070	0.001	−0.004
		RMSE	0.079	0.064	0.103	0.106	0.126	0.086	0.150	0.089	0.078
	Single QR	Bias	0.070	−0.006	−0.308	−0.030	−0.265	−0.062	−0.246	−0.049	−0.015
		RMSE	0.109	0.084	0.316	0.116	0.275	0.121	0.260	0.104	0.083
		$\tau = (0.75, 0.5, 0.25)$									
MN	Joint QR	Bias	0.037	−0.001	−0.031	−0.001	−0.012	−0.006	−0.043	−0.001	0.014
		RMSE	0.107	0.069	0.109	0.071	0.117	0.103	0.157	0.112	0.086
	Single QR	Bias	0.077	−0.007	−0.319	−0.008	−0.243	−0.056	−0.236	−0.039	−0.007
		RMSE	0.121	0.079	0.330	0.092	0.252	0.120	0.252	0.115	0.097
Mt_3	Joint QR	Bias	0.010	0.004	−0.033	−0.012	−0.007	−0.002	−0.010	−0.004	0.010
		RMSE	0.099	0.055	0.079	0.090	0.133	0.081	0.126	0.095	0.088
	Single QR	Bias	0.091	−0.010	−0.321	0.000	−0.238	−0.062	−0.259	−0.051	−0.019
		RMSE	0.134	0.089	0.329	0.102	0.254	0.117	0.272	0.112	0.101

QR: quantile regression; MN: multivariate normal distribution; RMSE: root mean square error; Bias: biases.

Table 3. Estimation results of very sparse Model 3.

Error	Methods	Evaluation	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{33}
		$\tau = (0.5, 0.5, 0.5)$									
MN	Joint QR	Bias	0.015	0.003	0.009	0.022	−0.013	0.015	−0.009	−0.008	0.007
		RMSE	0.051	0.037	0.046	0.101	0.127	0.086	0.082	0.061	0.069
	Single QR	Bias	0.026	−0.007	−0.292	0.024	−0.282	−0.053	−0.213	−0.051	−0.020
		RMSE	0.074	0.053	0.295	0.116	0.288	0.119	0.221	0.089	0.076
Mt_3	Joint QR	Bias	0.013	0.007	−0.011	−0.001	−0.044	−0.002	−0.050	0.003	−0.002
		RMSE	0.065	0.051	0.068	0.100	0.150	0.099	0.104	0.068	0.073
	Single QR	Bias	0.049	−0.005	−0.281	−0.003	−0.281	−0.075	−0.247	−0.041	−0.021
		RMSE	0.086	0.081	0.285	0.144	0.294	0.131	0.256	0.106	0.095
		$\tau = (0.25, 0.5, 0.75)$									
MN	Joint QR	Bias	0.033	0.004	−0.034	−0.017	0.008	−0.005	−0.043	0.006	0.009
		RMSE	0.116	0.074	0.102	0.085	0.095	0.086	0.182	0.096	0.073
	Single QR	Bias	0.112	0.023	−0.352	−0.014	−0.257	−0.054	−0.271	−0.029	−0.001
		RMSE	0.158	0.109	0.357	0.101	0.265	0.110	0.289	0.141	0.093
Mt_3	Joint QR	Bias	0.060	−0.001	−0.035	0.017	0.001	−0.009	−0.020	0.011	0.016
		RMSE	0.102	0.062	0.090	0.103	0.131	0.101	0.153	0.105	0.090
	Single QR	Bias	0.127	−0.023	−0.340	0.026	−0.243	−0.054	−0.279	−0.013	−0.005
		RMSE	0.154	0.089	0.345	0.101	0.254	0.118	0.302	0.146	0.110
		$\tau = (0.75, 0.5, 0.25)$									
MN	Joint QR	Bias	0.082	0.004	−0.045	0.012	−0.016	−0.011	−0.078	0.001	−0.013
		RMSE	0.132	0.051	0.115	0.081	0.117	0.083	0.170	0.084	0.098
	Single QR	Bias	0.120	−0.010	−0.349	0.023	−0.260	−0.070	−0.279	−0.037	−0.024
		RMSE	0.160	0.087	0.357	0.108	0.267	0.101	0.295	0.119	0.123
Mt_3	Joint QR	Bias	0.049	0.012	−0.049	−0.013	−0.012	−0.012	−0.056	−0.006	−0.003
		RMSE	0.121	0.073	0.097	0.079	0.143	0.099	0.137	0.086	0.087
	Single QR	Bias	0.103	0.018	−0.344	0.002	−0.269	−0.069	−0.272	−0.052	−0.009
		RMSE	0.145	0.085	0.351	0.106	0.279	0.117	0.291	0.121	0.096

QR: quantile regression; MN: multivariate normal distribution; RMSE: root mean square error; Bias: biases.

Table 4. Variable selection results of dense Model 1.

Quantiles τ	Error	Methods	APMSE	NC	NIC
(0.50, 0.50, 0.50)	MN	Joint QR	0.063 (0.035)	9	0
		Single QR	1.081 (0.196)	9	0
	Mt_3	Joint QR	0.079 (0.051)	9	0
		Single QR	1.158 (0.231)	8.94	0
(0.25, 0.50, 0.75)	MN	Joint QR	0.125 (0.062)	9	0
		Single QR	1.182 (0.240)	8.96	0
	Mt_3	Joint QR	0.123 (0.073)	8.98	0
		Single QR	1.320 (0.266)	8.96	0
(0.75, 0.50, 0.25)	MN	Joint QR	0.143 (0.084)	9	0
		Single QR	1.294 (0.243)	8.96	0
	Mt_3	Joint QR	0.108 (0.058)	9	0
		Single QR	1.310 (0.244)	8.96	0

QR: quantile regression; MN: multivariate normal distribution; APMSE: averaged posterior mean square error; NC: average correctly identified number of important covariates; NIC: average wrongly identified number of unimportant covariates.

Table 5. Variable selection results of sparse Model 2.

Quantiles τ	Error	Methods	APMSE	NC	NIC
(0.50, 0.50, 0.50)	MN	Joint QR	0.048 (0.035)	4	0.64
		Single QR	0.248 (0.048)	4	1.52
	Mt_3	Joint QR	0.053 (0.038)	4	0.76
		Single QR	0.263 (0.068)	4	1.64
(0.25, 0.50, 0.75)	MN	Joint QR	0.123 (0.088)	3.96	1.12
		Single QR	0.341 (0.121)	3.90	1.78
	Mt_3	Joint QR	0.090 (0.058)	4	1.04
		Single QR	0.306 (0.066)	4	1.58
(0.75, 0.50, 0.25)	MN	Joint QR	0.100 (0.062)	4	1.16
		Single QR	0.301 (0.092)	4	1.56
	Mt_3	Joint QR	0.082 (0.060)	4	0.70
		Single QR	0.318 (0.062)	3.98	1.62

QR: quantile regression; MN: multivariate normal distribution; APMSE: averaged posterior mean square error; NC: average correctly identified number of important covariates; NIC: average wrongly identified number of unimportant covariates.

Table 6. Variable selection results of very sparse Model 3.

Quantiles τ	Error	Methods	APMSE	NC	NIC
(0.50, 0.50, 0.50)	MN	Joint QR	0.137 (0.061)	4	3
	Single QR	0.367 (0.064)	4	4.26	
	Mt_3	Joint QR	0.175 (0.073)	4	3.62
	Single QR	0.432 (0.094)	4	5.62	
(0.25, 0.50, 0.75)	MN	Joint QR	0.238 (0.092)	4	4.66
		Single QR	0.524 (0.134)	3.96	6.54
	Mt_3	Joint QR	0.234 (0.094)	4	5.14
		Single QR	0.539 (0.118)	3.94	7.26
(0.75, 0.50, 0.25)	MN	Joint QR	0.231 (0.090)	3.98	4.70
		Single QR	0.522 (0.149)	3.92	6.34
	Mt_3	Joint QR	0.249 (0.106)	3.98	5.42
		Single QR	0.546 (0.140)	3.92	7.30

QR: quantile regression; MN: multivariate normal distribution; APMSE: averaged posterior mean square error; NC: average correctly identified number of important covariates; NIC: average wrongly identified number of unimportant covariates.

Table 7. Summary of parameter estimates along with standard errors of fixed effects (in parentheses) for the PBCseq data based on joint QR approach.

τ	Variables	Constant	X_1	X_2	X_3	X_4	X_5
(0.25, 0.25)	$Y^{(1)}$ (lbili)	1.366	−0.543	0.015	−0.009	0.038	−0.001
		(0.732)	(0.330)	(0.124)	(0.009)	(0.050)	(0.004)
	$Y^{(2)}$ (lalbumin)	1.283	−0.010	0.010	−0.001	−0.015	0.000
		(0.167)	(0.065)	(0.043)	(0.002)	(0.021)	(0.002)
(0.25, 0.50)	$Y^{(1)}$ (lbili)	1.044	−0.628	0.044	−0.009	0.026	0.000
		(0.519)	(0.269)	(0.105)	(0.007)	(0.049)	(0.004)
	$Y^{(2)}$ (lalbumin)	1.335	−0.005	0.010	−0.001	−0.013	0.000
		(0.138)	(0.057)	(0.040)	(0.002)	(0.021)	(0.002)
(0.25, 0.75)	$Y^{(1)}$ (lbili)	1.048	−0.642	0.009	−0.010	0.039	0.000
		(0.818)	(0.416)	(0.127)	(0.011)	(0.054)	(0.004)
	$Y^{(2)}$ (lalbumin)	1.354	0.000	0.008	−0.001	−0.012	0.000
		(0.163)	(0.063)	(0.052)	(0.003)	(0.025)	(0.002)
(0.50, 0.25)	$Y^{(1)}$ (lbili)	1.158	−0.575	0.024	−0.009	0.035	0.000
		(0.474)	(0.251)	(0.119)	(0.006)	(0.046)	(0.003)
	$Y^{(2)}$ (lalbumin)	1.306	−0.012	0.007	−0.001	−0.012	0.000
		(0.159)	(0.069)	(0.044)	(0.003)	(0.025)	(0.002)
(0.50, 0.50)	$Y^{(1)}$ (lbili)	1.249	−0.606	0.035	−0.009	0.030	0.000
		(0.439)	(0.254)	(0.114)	(0.006)	(0.044)	(0.003)
	$Y^{(2)}$ (lalbumin)	1.330	−0.006	0.006	−0.001	−0.015	0.000
		(0.148)	(0.058)	(0.044)	(0.002)	(0.021)	(0.002)
(0.50, 0.75)	$Y^{(1)}$ (lbili)	1.163	−0.553	0.014	−0.008	0.040	0.000
		(0.506)	(0.251)	(0.111)	(0.006)	(0.047)	(0.004)
	$Y^{(2)}$ (lalbumin)	1.331	−0.005	0.010	−0.001	−0.012	0.000
		(0.161)	(0.066)	(0.046)	(0.003)	(0.022)	(0.002)
(0.75, 0.25)	$Y^{(1)}$ (lbili)	0.956	−0.452	0.022	−0.010	0.036	0.000
		(0.748)	(0.325)	(0.149)	(0.010)	(0.057)	(0.004)
	$Y^{(2)}$ (lalbumin)	1.311	−0.003	0.012	−0.001	−0.011	0.000
		(0.170)	(0.067)	(0.049)	(0.003)	(0.026)	(0.003)
(0.75, 0.50)	$Y^{(1)}$ (lbili)	1.176	−0.597	0.028	−0.010	0.025	0.001
		(0.527)	(0.286)	(0.112)	(0.007)	(0.047)	(0.004)
	$Y^{(2)}$ (lalbumin)	1.334	0.001	0.008	−0.001	−0.013	0.000
		(0.143)	(0.057)	(0.040)	(0.002)	(0.020)	(0.002)
(0.75, 0.75)	$Y^{(1)}$ (lbili)	1.439	−0.526	0.039	−0.008	0.037	0.000
		(0.773)	(0.289)	(0.133)	(0.009)	(0.053)	(0.004)
	$Y^{(2)}$ (lalbumin)	1.306	−0.012	0.005	−0.002	−0.016	0.000
		(0.241)	(0.071)	(0.048)	(0.003)	(0.025)	(0.002)

PBCseq: primary biliary cirrhosis sequential; QR: quantile regression; lbili: logarithm of serum bilirubin; lalbumin: logarithm of serum albumin.

bili and albumin), was established on each patient who was collected repeatedly and prospectively at yearly intervals under standardized forms, definitions, and study protocols. In the second paragraph of Section 1, we have alluded that serum bilirubin and serum albumin are two of the primary indicators to help evaluate and track the absence of liver diseases. An extremely higher or lower level than the standards that bilirubin is excreted in bile and urine can indicate certain diseases. Also, extreme higher or lower circulating serum albumin levels are harmful to human body. Additionally, there exist some relationship between serum bilirubin and serum albumin levels. Fukui et al.[48] showed that the serum bilirubin level is associated with microalbuminuria and subclinical atherosclerosis in patients with type 2 diabetes. A separate analysis for those two markers may lose important information about evolutional relationships among multiple responses. Thus, a joint analysis of the longitudinally collected serum bilirubin and serum albumin may be more appropriate in diagnosing liver diseases. The PBCseq data set is available from "mixAK" package of R[49] and has been analyzed by Wang[50] and Taavoni et al.,[23] etc. Wang[50] analyzed this data set using a mixture of multivariate *t* linear mixed models with heterogeneity, Taavoni et al.[23] analyzed the data set using multivariate *t* semiparametric mixed-effects model with multiple characteristics.

We concentrate on modeling the dependence of the longitudinal profiles of two markers with the natural logarithm of serum bilirubin (lbili) and the natural logarithm of serum albumin (lalbumin), on time (visited years) and other covariates of interest (e.g. sex, drug, and age). We conduct the joint QR analysis for the longitudinal PBCseq data set on responses

Table 8. Summary of parameter estimates along with standard errors of fixed effects (in parentheses) for the PBCseq data based on single QR approach.

τ	Variables	Constant	X_1	X_2	X_3	X_4	X_5
(0.25, 0.25)	$Y^{(1)}$ (lbili)	0.142	−0.187	−0.014	−0.007	0.038	−0.010
		(1.255)	(0.634)	(0.445)	(0.034)	(0.249)	(0.028)
	$Y^{(2)}$ (lalbumin)	0.117	0.070	0.002	−0.001	0.021	−0.008
		(1.016)	(0.571)	(0.377)	(0.027)	(0.205)	(0.025)
(0.25, 0.50)	$Y^{(1)}$ (lbili)	0.416	−0.189	−0.033	−0.013	0.065	−0.012
		(1.124)	(0.505)	(0.425)	(0.031)	(0.226)	(0.028)
	$Y^{(2)}$ (lalbumin)	0.581	0.065	−0.008	−0.001	−0.003	0.000
		(0.963)	(0.492)	(0.332)	(0.022)	(0.165)	(0.020)
(0.25, 0.75)	$Y^{(1)}$ (lbili)	0.534	−0.224	−0.011	−0.012	0.041	−0.012
		(1.397)	(0.646)	(0.460)	(0.031)	(0.246)	(0.034)
	$Y^{(2)}$ (lalbumin)	0.969	0.027	0.010	0.001	−0.051	0.011
		(1.328)	(0.666)	(0.426)	(0.029)	(0.191)	(0.029)
(0.50, 0.25)	$Y^{(1)}$ (lbili)	0.562	−0.261	−0.019	−0.007	0.010	0.000
		(1.230)	(0.587)	(0.475)	(0.030)	(0.228)	(0.029)
	$Y^{(2)}$ (lalbumin)	0.157	0.057	0.011	−0.003	0.021	−0.008
		(1.134)	(0.632)	(0.470)	(0.030)	(0.227)	(0.031)
(0.50, 0.50)	$Y^{(1)}$ (lbili)	0.680	−0.232	−0.041	−0.007	0.006	−0.002
		(1.266)	(0.545)	(0.439)	(0.030)	(0.239)	(0.031)
	$Y^{(2)}$ (lalbumin)	0.626	0.049	−0.010	0.000	−0.016	0.001
		(1.111)	(0.475)	(0.334)	(0.025)	(0.190)	(0.023)
(0.50, 0.75)	$Y^{(1)}$ (lbili)	0.843	−0.266	0.009	−0.008	0.033	−0.003
		(1.308)	(0.588)	(0.458)	(0.029)	(0.221)	(0.029)
	$Y^{(2)}$ (lalbumin)	1.089	−0.059	0.017	0.002	−0.030	0.009
		(1.495)	(0.627)	(0.507)	(0.031)	(0.230)	(0.030)
(0.75, 0.25)	$Y^{(1)}$ (lbili)	0.798	−0.245	0.012	−0.004	−0.003	0.007
		(1.382)	(0.684)	(0.480)	(0.034)	(0.249)	(0.034)
	$Y^{(2)}$ (lalbumin)	0.374	0.050	0.004	−0.004	0.013	−0.010
		(1.281)	(0.634)	(0.450)	(0.030)	(0.224)	(0.029)
(0.75, 0.50)	$Y^{(1)}$ (lbili)	0.925	−0.251	0.040	−0.004	0.004	0.007
		(1.547)	(0.622)	(0.463)	(0.032)	(0.237)	(0.031)
	$Y^{(2)}$ (lalbumin)	0.746	0.036	0.022	−0.002	−0.004	0.001
		(1.236)	(0.452)	(0.345)	(0.024)	(0.179)	(0.023)
(0.75, 0.75)	$Y^{(1)}$ (lbili)	1.167	−0.304	0.007	−0.008	0.004	0.009
		(1.514)	(0.617)	(0.487)	(0.032)	(0.248)	(0.031)
	$Y^{(2)}$ (lalbumin)	1.069	−0.059	0.007	0.000	−0.044	0.009
		(1.392)	(0.561)	(0.432)	(0.027)	(0.204)	(0.025)

PBCseq: primary biliary cirrhosis sequential; QR: quantile regression; lbili: logarithm of serum bilirubin; lalbumin: logarithm of serum albumin.

of lbili and lalbumin by establishing model (3.1). Denote $\boldsymbol{y}_{it} = (y_{it}^{(1)}, y_{it}^{(2)})^T, \boldsymbol{x}_{it} = (1, x_{it1}, x_{it2}, x_{it3}, x_{it4}, x_{it5})^T, i \in 1, \ldots, 312$, where $\boldsymbol{y}_{it}$ is the bivariate response for the i-th patient, $y_{it}^{(1)}$ and $y_{it}^{(2)}$ represent lbili and lalbumin levels, and $\boldsymbol{x}_{it}$ is a 6×1 vector of regressors, x_{it1}(gender) denotes the gender indicator (0 = male and 1 = female), x_{it2}(drug) denotes the drug treatment indicator (0 = patient treated with placebo and 1 = patient treated with D-penicillamine), x_{it3}(age) denotes the age at entry in years, x_{it4}(month/12) denotes time (visited years), and $x_{it5} = x_{it4}^2$. Thus, the parameter matrix of fixed effects is denoted as

$$\boldsymbol{\beta}_{2\times 5} = \begin{pmatrix} \beta_{10} & \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} \\ \beta_{20} & \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} & \beta_{25} \end{pmatrix}$$

Nine quantile combinations, that is, $\tau = (0.25, 0.25)$, $\tau = (0.25, 0.55)$, $\tau = (0.25, 0.75)$, $\tau = (0.50, 0.25)$, $\tau = (0.50, 0.50)$, $\tau = (0.50, 0.75)$, $\tau = (0.75, 0.25)$, $\tau = (0.75, 0.50)$, and $\tau = (0.75, 0.75)$, are considered for the response $(y_{it}^{(1)}, y_{it}^{(2)})^T$. The priors and initial values are set as the same setting in Section 5. We run 8000 iterations of Gibbs sampling algorithm for each quantile combination, the first 2000 burn-in iterations are removed and the remaining 6000 stationary iterations are retained to conduct posterior inference. Estimation results including average estimation values and standard errors of the

considered nine quantile combinations are reported in Table 7. For the aim of comparison, we also provide the estimation results of single QR approach in Table 8.

Compared Table 7 with Table 8, we again conclude that the proposed joint QR approach has better estimation results with apparently smaller standard errors for almost all parameters. In addition, from Table 7, we find that five covaraites have not exactly the same impacts on two responses for each quantile combination. The impacts of five covariates on two responses also totally vary with the quantile combinations. Further, at all quantile levels under consideration, covariates X_1 (gender) and X_3 (age) have simultaneous negative effects on responses Y_1 (lbili) and Y_2 (lalbumin), and covariates X_2 (drug) have simultaneous positive effects on two responses. Yet, X_1 has a clearly bigger impact on response Y_1 than Y_2 while X_2 has a slightly bigger impact on response Y_1 than Y_2. Whereas, covariates X_4 (visited years) has positive effects on response Y_1 and negative effects on response Y_2 for all quantiles. Specially, we find X_5 has almost no impact on responses Y_1 and Y_2 for all considered quantiles since the coefficients are approximately estimated as zeros. Some useful conclusions can be drawn based on the above quantitative analysis. lbili and lalbumin becomes slightly smaller as the ages of patients increase. Female patients ($X_1 = 1$) have faster change for lbili and lalbumin than male patients ($X_1 = 0$). Patients treated with D-penicillamine ($X_2 = 1$) have bigger change for lbili and lalbumin than patients treated with placebo ($X_2 = 0$). Generally, in the diagnosis of clinical liver disease patients, the increase of serum bilirubin and the decrease of serum albumin indicate liver cell damage. Hence, conclusions suggest elderly female patients with liver disease more should receive medication intervention of D-penicillamine to more effectively reduce the risk of liver disease development.

7 Concluding remarks

We investigated the Bayesian joint QR modeling of multi-response mixed-effects model with longitudinal data in this paper. A MAL distribution was imposed on the errors of the considered model to build the working likelihood of Bayesian joint inference. For implementing the efficient Bayesian inference, the location-scale mixture reparameterization of the working likelihood and LASSO-type penalization priors of regression coefficients were employed to construct the Bayesian joint hierarchical QR model. The conditional posterior distributions of parameters and latent variables based on MCMC algorithm were derived. Monte Carlo simulation examples were presented to illustrate the proposed joint QR approach. Simulation results showed that the proposed joint QR approach has more satisfactory performance than the single QR approach. Finally, we analyzed a real-world data set of a longitudinal PBCseq cohort study using the proposed joint modeling approach. The proposed joint QR approach can be extended to more complex longitudinal data models and applications.

Acknowledgements

The authors thank editors and two referees for their constructive comments and suggestions which have greatly improved the paper. The work of Tian Yu-Zhu was jointly supported by grants from the National Natural Science Foundation of China (Grant 12061065) and Funds for Innovative Fundamental Research Group Project of Gansu Province of China (Grant 23JRRA684). The work of Tang Man-Lai is partially supported by the Research Matching Grant (700006) from the Research Grants Council of the Hong Kong Special Administration Region and FDS Grant (UGC/FDS14/P05/20) from the Big Data Intelligence Center in The Hang Seng University of Hong Kong.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iD

Catherine Wong https://orcid.org/0000-0002-5421-1458

Supplemental material

Supplemental material for this article is available online.

References

1. Laird NM and Ware H. Random-effect models for longitudinal data. *Biometrics* 1982; **38**: 963–974.
2. Diggle P, Heagerty P, Liang kY, et al. *Analysis of longitudinal data*. USA: Oxford University Press, 2002.
3. Hedeker D and Gibbons RD. *Longitudinal data analysis*. Wiley, New Jersey: Blackwell Publishing Ltd, 2006.
4. Wu H and Zhang JT. *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. New York: Wiley, 2006.
5. Wu L. *Mixed effects models for complex data*. Boca Raton: Chapman & Hall/CRC Press, 2010.

6. Demidenko E. *Mixed Models: Theory and Applications with R*. (2nd ed). Hoboken Wiley, 2013.
7. Shah A, Laird N and Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. *J Am Stat Assoc* 1997; **438**: 775–779.
8. Sammel M, Lin X and Ryan L. Multivariate linear mixed models for multiple outcomes. *Stat Med* 1999; **18**: 2479–2492.
9. Lin TI. A mixed-effects regression model for longitudinal multivariate ordinal data. *Int Biometric Soc* 2006; **62**: 261–268.
10. Blozis SA, Conger KJ and Harring JR. Nonlinear latent curve models for multivariate longitudinal data. *Int J Behav Dev* 2007; **31**: 340–346.
11. Alfo M and Maruotti A. *A hierarchical model for time dependent multivariate longitudinal data*. Berlin, Heidelberg: Springer, 2010.
12. Bandyopadhyay S, Ganguli B and Chatterjee A. A review of multivariate longitudinal data analysis. *Stat Methods Med Res* 2011; **20**: 299–330.
13. Gebregziabher M, Zhao Y and Dismuke CE, etc. Joint modeling of multiple longitudinal cost outcomes using multivariate generalized linear mixed models. *Health Serv Outc Res Methodol* 2013; **13**: 39–57.
14. Laffont CM, Vandemeulebroecke M and Concordet D. Multivariate analysis of longitudinal ordinal data with mixed effects models, with application to clinical outcomes in osteoarthritis. *J Am Stat Assoc* 2014; **109**: 955–966.
15. Grimm KJ. Multivariate longitudinal methods for studying developmental relationships between depression and academic achievement. *Int J Behav Dev* 2015; **31**: 328–339.
16. Wang WL, Lin TI and Lachos VH. Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Stat Methods Med Res* 2015; **27**: 48–64.
17. Luwanda AG and Mwambi HG. A nonlinear mixed-effects model for multivariate longitudinal data with partially observed outcomes with application to HIV disease dynamics. *J Appl Stat* 2017; **44**: 441–456.
18. Rajeswaran J, Blackstone EH and Barnard J. Joint modeling of multivariate longitudinal data and competing risks using multiphase sub-models. *Stat Biosci* 2018; **10**: 651–685.
19. Lin TI, Lachos VH and Wang WL. Multivariate longitudinal data analysis with censored and intermittent missing responses. *Stat Med* 2018; **37**: 2822–2835.
20. Hui FKC, Mueller S and Welsh AH. Sparse pairwise likelihood estimation for multivariate longitudinal mixed models. *J Am Stat Assoc* 2018; **113**: 1759–1769.
21. Jiang HY, Yue RX and Zhou XD. Optimal designs for multivariate logistic mixed models with longitudinal data. *Commun Stat-Theory Method* 2019; **48**: 850–864.
22. Wang WL. Bayesian analysis of multivariate linear mixed models with censored and intermittent missing responses. *Stat Med* 2020; **39**: 2518–2535.
23. Taavoni M, Arashi M, Wang WL, et al. Multivariate *t* semiparametric mixed-effects model for longitudinal data with multiple characteristics. *J Stat Comput Simul* 2021; **91**: 260–281.
24. Tian Z and Qiu P. Multivariate single index modeling of longitudinal data with multiple responses. *Stat Med* 2023; **42**: 2982–2998.
25. Koenker R. *Quantile regression*. Cambridge: Cambridge University Press, 2005.
26. Koenker R, Chernozhukov V, He X, et al. *Handbook of Quantile Regression*. Florida CRC Press, 2017.
27. Koenker R. Quantile regression for longitudinal data. *J Multivar Anal* 2004; **91**: 74–89.
28. Geraci M and Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 2007; **8**: 140–154.
29. Liu Y and Bottai M. Mixed-effects models for conditional quantiles with longitudinal data. *Int J Biostat* 2009; **5**: 1–24.
30. Tian YZ, Li EQ and Tian MZ. Bayesian joint quantile regression for mixed effects models with censoring and errors in covariates. *Comp Stat* 2016; **31**: 1–27.
31. Aghamohammadi A and Mohammadi S. Bayesian analysis of penalized quantile regression for longitudinal data. *Stat Pap* 2017; **58**: 1035–1053.
32. Alhamzawi R and Ali HTM. Bayesian quantile regression for ordinal longitudinal data. *J Appl Stat* 2018; **45**: 815–828.
33. Tian YZ, Wang LY, Tang ML, et al. Likelihood-based quantile mixed effects models for longitudinal data with multiple features via MCEM algorithm. *Commun Stat-Simul Comput* 2020; **49**: 317–334.
34. Waldmann E and Kneib T. Bayesian bivariate quantile regression. *Stat Model* 2014; **15**: 326–344.
35. Kulkarni H, Biswas J and Das K. A joint quantile regression model for multiple longitudinal outcomes. *AStA-Adv Stat Anal* 2019; **103**: 453–473.
36. Ghasemzadeh S, Ganjali M and Baghfalaki T. Bayesian quantile regression for joint modeling of longitudinal mixed ordinal and continuous data. *Commun Stat Simul Comput* 2020; **49**: 375–395.
37. Biswas J and Das K. A Bayesian quantile regression approach to multivariate semi-continuous longitudinal data. *Comput Stat* 2021; **36**: 241–260.
38. Petrella L and Raponi V. Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. *J Multivar Anal* 2019; **173**: 70–84.
39. Tian YZ, Tang ML and Tian MZ. Bayesian joint inference for multivariate quantile regression model with $L_{1/2}$ penalty. *Comput Stat* 2021; **36**: 2967–2994.
40. Kotz S and Kozubowski TJ. *Symmetric multivariate Laplace distribution. Asymmetric multivariate Laplace distribution*. Boston, MA: Springer, 2001.

41. Kollo T and Srivastava MS. Estimation and testing of parameters in multivariate Laplace distribution. *Commun Stat-Theory Method* 2005; **33**: 2363–2387.
42. Visk H. On the parameter estimation of the asymmetric multivariate Laplace distribution. *Commun Stat-Theory Method* 2009; **38**: 461–470.
43. Hurlimann W. A moment method for the multivariate asymmetric Laplace distribution. *Stat Probab Lett* 2013; **83**: 1247–1253.
44. Tibshirani RJ. Regression shrinkage and selection via the LASSO. *J R Stat Soc (Ser B)* 1996; **73**: 273–282.
45. Zou H. The adaptive LASSO and its oracle properties. *J Am Stat Assoc* 2006; **101**: 1418–1429.
46. Park T and Casella G. The Bayesian LASSO. *J Am Stat Assoc* 2008; **103**: 681–686.
47. Leng C, Tran MN and Nott D. Bayesian adaptive LASSO. *Ann Inst Stat Math* 2014; **66**: 221–244.
48. Fukui M, Tanaka M, Shiraishi E, et al. Relationship between serum bilirubin and albuminuria in patients with type 2 diabetes. *Kidney Int* 2008; **74**: 1197–1201.
49. Komarek A and Komarkova L. Capabilities of R package mixAK for clustering based on multivariate continuous and discrete longitudinal data. *J Stat Softw* 2014; **59**: 1–38.
50. Wang WL. Mixture of multivariate *t* linear mixed models for multi-outcome longitudinal data with heterogeneity. *Stat Sin* 2017; **27**: 733–760.