Preprints.org

Article

Bayesian Joint Modeling Analysis of Longitudinal Proportion and Survival Data

Wenting Liu , Huiqiong Li [*] , Anming Tang , Zixin Cui

Article

Bayesian Joint Modeling Analysis of Longitudinal Proportion and Survival Data

Wenting Liu, Huiqiong Li *, Anmin Tang and Zixin Cui

Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, China; liuwenting@itc.ynu.edu.cn (W.L.)

* Correspondence: lihuiqiong@ynu.edu.cn; Tel.: +86-135-29752592

Abstract: This paper focuses on a joint model to analyse longitudinal proportion and survival data. After transforming the longitudinal proportional data by a logit function, we adopt a partially linear mixed-effect model for it, in which nonlinear function of time is fitted via using B-splines technique and a centered Dirichlet Process Mixture Model (CDPMM) is specified for a general distribution of random effects. The survival data is assumed to a Cox proportional hazard model, the sharing random effects joint model is developed for the two types of data. Combining the Gibbs sampler and the Metropolis-Hastings algorithm, we propose a Bayesian Lasso (BLasso) method to simultaneously estimate unknown parameters and select important covariates. Simulation studies are conducted to investigate the finite sample performance of the proposed methods. An example from the MA.5 research experiment is used to illustrate the proposed methodologies.

Keywords: longitudinal proportional data; survival data; joint model; Bayesian variable selection; B-splines; CDPMM metnod

---

1. Introduction

Joint analysis of longitudinal and survival data have been widely applied to cancer and HIV/AIDS clinical studies where time to event outcomes, such as disease free and overall survivals, are usually the primary endpoints. In particular, after Faucett and Thomas [1] and Wulfsohn and Tsiatis [2] introduced what is the standard joint model, the joint model has been extensively explored. The advantages of joint models have been discussed by researchers [3–8]. However, in clinical trials, some patients with poor quality of life may drop out of the study due to disease recurrence or death. In this case, the absence of quality of life measures resulting from the withdrawal of these patients is informative. In order to provide strong evidence for the tradeoff between intensive treatment and the associated poor quality of life for patients, we conducted joint modeling of longitudinal life measures and survival data to investigate the relationship between them. For longitudinal quality of life and survival data, Henderson et al. [9] and Zeng and Cai [10] considered the use of shared normal distribution random effects to jointly analyze the relationship between longitudinal quality of life and survival time. Tang et al. [11] considered a novel semiparametric joint model for multivariate longitudinal and survival data to analyze data from the International Breast Cancer Study. Longitudinal life quality measurement data can be linearly converted into longitudinal proportion data whose value range is in the unit interval (0,1) [12]. For the longitudinal component, there are two methods to deal with it. The first method applied the classic linear mixed model to the longitudinal scale data after logit transformation, and the second method directly used the simplex distribution to model the longitudinal scale data. The models established by the two methods both used the EM algorithm and the Laplace approximation to estimate the unknown parameters. In order to be more flexible and practical, this paper will use a partial linear mixed effect model to the logit transformed longitudinal scale data, and use the B-splines method to model the nonlinear function in the model. Meanwhile, to promote the feasibility of our proposed model, we use CDPMM method to model random effects.

In addition, variables selection in the joint model is also considered. In traditional regression models, methods of variable selection include the forward selection method, backward elimination

method, stepwise selection method or some information criterion such as the Akaike information criterion. But these approaches are computationally expensive and unstable for the complicated models with a large number of covariates. Therefore, some penalized likelihood methods are gradually proposed, especially Lasso of Tibshirani [13] estimating linear regression coefficients by L_1-constrained least squares. Tibshirani [13] proposed that Lasso estimates can be interpreted as posterior norm estimates when the regression parameters have independent and identical Laplacian priors (ie, double exponential priors). Motivated by this connection, under the Bayesian framework, Park and Casella [14] proposed the BLasso variable selection method by imposing the double exponential prior on the regression coefficients and the gamma distribution on the shrinkage parameter. The BLasso method has been applied to linear regression model [15], semi-parametric structural equation model [16] and joint models of longitudinal and survival data [17]. Therefore, this paper extended that BLasso variable selection method to the joint model of longitudinal proportion data and survival data and a BLasso approach is developed to simultaneously estimate unknown parameters and identify the significant effect of the important covariates.

The rest of this paper is organized as follows. In Section 2, the joint model of longitudinal proportion and survival data is introduced. In Section 3, the Bayesian estimations of joint model are proposed. In Section 4, three numerical simulations are presented to evaluate the performance of the proposed methods. In Section 5, we apply the proposed approach to the MA.5 research experiment data and some concluding remarks are given in Section 6. Technical details are given in Appendix A.

2. Model and Notation

Consider a data set from n individuals. Let y_{ij} be a longitudinal proportional measurement for the i-th individual ($i = 1, 2, \cdots, n$) at observation time point t_{ij} for $j = 1, 2, \cdots, n_i$, and $y_{ij} \in (0, 1)$. We assume y_{ij}^* is the logit-transformation of y_{ij}, and $y_{ij}^* = \text{logit}(y_{ij}) = \log\left(\frac{1 - y_{ij}}{y_{ij}}\right)$. Further, T_i^* and C_i are the true survival time and censoring time, respectively. Let $T_i = \min(T_i^*, C_i)$ denotes the corresponding observed event time, and $\delta_i = \mathbf{1}(T_i^* \leq C_i)$ denotes the failure indicator, where $\mathbf{1}(\cdot)$ is a indicator function.

Denote $\boldsymbol{y}^* = \{\boldsymbol{y}_1^*, \boldsymbol{y}_2^*, \cdots, \boldsymbol{y}_n^*\}$, $\boldsymbol{y}_i^* = \left\{y_{i1}^*, y_{i2}^*, \cdots, y_{in_i}^*\right\}$, $\boldsymbol{T} = \{T_1, T_2, \cdots, T_n\}$, $\boldsymbol{\Delta} = \{\delta_1, \delta_2, \cdots, \delta_n\}$. Let $\boldsymbol{b} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n\}$ be time-independent random effects underlying both the longitudinal and survival processes for the i-th individual. Given the random effects $\boldsymbol{b}_i$, assume that y_{ij}^* follows partially linear mixed-effect model

$$y_{ij}^* | \boldsymbol{b}_i = \boldsymbol{X}_{ij}^\top \boldsymbol{\beta} + g(t_{ij}) + \boldsymbol{Z}_{ij}^\top \boldsymbol{b}_i + \varepsilon_{ij}, \tag{1}$$

where $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ represent the time-independent design vectors of fixed and random effects associated with y_{ij}^*, respectively. And $\boldsymbol{\beta}$ is a $p_1 \times 1$ fixed effects regression parameters, $\boldsymbol{b}_i$ is a $q \times 1$ random effects vector, $g(t)$ is a twice continuous differentiable unknown function, and ε_{ij} is a white noise process with variance σ^2. Also, we assume that ε_{ij}'s are independent of $\boldsymbol{b}_i$. To facilitate the feasibility of our proposed model, instead of the traditional normality assumption, which may be violated in some applications [18], it is specified the random effects by Dirichlet process (DP) mixture of normals.

For event time T_i, given random effects $\boldsymbol{b}_i$, we assume that T_i follows the hazard model

$$\lambda_i(t | \boldsymbol{b}_i) = \lambda_0(t) \exp(\boldsymbol{W}_i^\top \boldsymbol{\gamma} + \boldsymbol{\phi}^\top \boldsymbol{b}_i), \tag{2}$$

where $\boldsymbol{W}_i$ is the known fixed effects design matrix linking the unknown $p_2 \times 1$ parameter vector $\boldsymbol{\gamma}$ to $\lambda_i(t | \boldsymbol{b}_i)$, $\boldsymbol{\phi}$ is an unknown $q \times 1$ parameter vector linking $\boldsymbol{b}_i$ to $\lambda_i(t | \boldsymbol{b}_i)$, and $\lambda_0(t)$ is an unknown basic hazard function.

From the above discussion, it is suggested to link models (1) and (2) through shared random effects, called a shared random effects joint model (JMSRE). Parameter $\boldsymbol{\phi}$ in model JMSRE reflects the correlation between transformed longitudinal proportional data and survival data, given random

effects. When $\boldsymbol{\phi} = \mathbf{0}_{q \times 1}$, it means that the longitudinal index is not necessarily related to the event time, that is, longitudinal proportional data and survival data can be modeled separately. So in this case joint modeling is not necessary and longitudinal indicators can be ignored for modeling survival data.

Further, to make Bayesian inference on $\boldsymbol{\beta}$ based on model (1), we approximate $g(t)$ through a B-splines method

$$g(t) \approx B_1(t)\varphi_1 + B_2(t)\varphi_2 + \cdots + B_L(t)\varphi_L = \boldsymbol{B}^\top(t)\boldsymbol{\varphi},$$

where $L = d + K + 1$, d is the degree of B-splines, K is the number of knots, $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \cdots, \varphi_L)^\top$ is a $L \times 1$ unknown coefficient vector, and $\boldsymbol{B}(t) = (B_1(t), B_2(t), \cdots, B_L(t))^\top$.

Denote $\boldsymbol{\theta}_y = \{\boldsymbol{\beta}, \boldsymbol{\varphi}, \sigma^2\}$ be unknown parameters associated with model (1), and $\boldsymbol{\theta}_T = \{\boldsymbol{\gamma}, \boldsymbol{\phi}\}$ be unknown parameters associated with model (2). Thus, given $(\boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{b})$, a joint likelihood function of $(\boldsymbol{y}^*, \boldsymbol{T}, \boldsymbol{\Delta})$ can be written as

$$p(\boldsymbol{y}^*, \boldsymbol{T}, \boldsymbol{\Delta} | \boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{b}) = \prod_{i=1}^n p(\boldsymbol{y}_i^* | \boldsymbol{b}_i; \boldsymbol{\theta}_y) \, p(\boldsymbol{T}, \boldsymbol{\Delta} | \boldsymbol{b}_i; \boldsymbol{\theta}_T), \tag{3}$$

where

$$p(\boldsymbol{y}_i^* | \boldsymbol{b}_i; \boldsymbol{\theta}_y) = \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\left(y_{ij}^* - \boldsymbol{X}_{ij}^\top \boldsymbol{\beta} - \boldsymbol{B}^\top(t_{ij})\boldsymbol{\varphi} - \boldsymbol{Z}_{ij}^\top \boldsymbol{b}_i\right)^2}{2\sigma^2} \right\},$$

$$p(\boldsymbol{T}, \boldsymbol{\Delta} | \boldsymbol{b}_i; \boldsymbol{\theta}_T) = \prod_{i=1}^n \left[\frac{\exp(\boldsymbol{W}_i^\top \boldsymbol{\gamma} + \boldsymbol{\phi}^\top \boldsymbol{b}_i)}{\sum_{j \in \mathcal{R}_i} \exp(\boldsymbol{W}_j^\top \boldsymbol{\gamma} + \boldsymbol{\phi}^\top \boldsymbol{b}_j)} \right]^{\delta_i}, \quad \mathcal{R}_i = \{j : T_j \geq T_i\}.$$

3. Bayesian estimation of joint model

3.1. Prior specification

To develop Bayesian inference on the considered models, we need specifying the prior distributions for $\sigma^2, \boldsymbol{\varphi}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\phi}$. For conjugation, we consider the following priors for $\sigma^2, \boldsymbol{\varphi}$:

$$\sigma^{-2} \sim \Gamma(a_0, b_0), \quad \boldsymbol{\varphi} \sim N_L\left(\mathbf{0}, \boldsymbol{H}_\varphi^0\right), \tag{4}$$

where a_0, b_0, and $\boldsymbol{H}_\varphi^0$ are pre-given hyperparameters, $\Gamma(a_0, b_0)$ denotes the Gamma distribution with parameter a_0 and the shape parameter b_0.

Following Tang et al. [17], we specify the DP mixture of normals for the random effects $\boldsymbol{b}_i$, that is, $\boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^\infty \pi_g N_q(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g)$ with $(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g) \sim \mathcal{P}$, where $\mathcal{P}$ is a random probability with an unknown form. one common approach is to specify DP prior to approximate $\mathcal{P}$, i.e., $\mathcal{P} \sim \mathrm{DP}(\tau F_0)$ in which F_0 is a base distribution that serves as a starting-point for constructing the nonparametric distribution, and τ is a weight that indicates the researcher's certainty of F_0 as the distribution of $\mathcal{P}$. To gain the form of $\mathcal{F}$, motivated by Ohlssen et al. [18] and Yang et al. [19], we consider the following truncated approximate centered Dirichlet Process Mixture Model (CDPMM), which can specify the prior distribution of $\boldsymbol{b}_i$ as follows:

$$\boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^G \pi_g N_q(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g) \text{ with } \boldsymbol{\mu}_g = \boldsymbol{\mu}_g^* - \sum_{g=1}^G \pi_g \boldsymbol{\mu}_g^* \text{ and } (\boldsymbol{\mu}_g^*, \boldsymbol{\Omega}_g) \overset{\text{i.i.d.}}{\sim} F_0,$$

where $1 \leq G < \infty$, and π_g is a random probability weight chosen to be independent of $(\boldsymbol{\mu}_g^*, \boldsymbol{\Omega}_g)$ such that $0 \leq \pi_g \leq 1$ and $\sum_{g=1}^\infty \pi_g = 1$. In order to ease the computational intensity, according to take

moderate value proposed by Ishwaran and Zarepour (2000) [20], we consider $G = 25$, and π_g is given by the following stick-breaking procedure:

$$\pi_1 = \vartheta_1 \ \text{ and } \ \pi_g = \vartheta_g \prod_{\iota=1}^{g-1} (1 - \vartheta_\iota) \ \text{ for } g = 2, \ldots, G, \tag{5}$$

where $\vartheta_g \overset{\text{i.i.d.}}{\sim} \text{Beta}(1, \tau)$ for $g = 1, 2, \ldots, G - 1$, and $\vartheta_G = 1$ so that $\sum_{g=1}^{G} \pi_g = 1$. The prior distribution for the unknown parameter τ is given by $\tau \sim \Gamma(a_1, a_2)$ with prespecified hyperparameters a_1 and a_2 [21]. Here, we took the hyperparameters a_1 and a_2 to be 25 and 5, respectively.

Sampling from posterior distributions of b_i with the above specified DP prior is quite difficult and inefficient. An efficient and flexible method for solving the above question is to represent b_i in terms of a latent variable $L_i \in \{1, 2, \ldots, G\}$, which records each b_i's cluster membership and conveys its parametric value to the distribution of b_i. Let $L = \{L_1, L_n, \ldots, L_n\}$, $\pi = \{\pi_1, \pi_2, \ldots, \pi_G\}$, $\mu^* = \{\mu_1^*, \mu_2^*, \ldots, \mu_G^*\}$ and $\Omega = \{\Omega_1, \Omega_2, \ldots, \Omega_G\}$ in which $\Omega_g = \text{diag}(\omega_{g1}, \omega_{g2}, \ldots, \omega_{gq})$, and they can be reformulated by

$$L_i | \pi \overset{\text{i.i.d}}{\sim} \sum_{g=1}^{G} \pi_g \delta_g(\cdot) \ \text{ and } \ (\pi, \mu^*, \Omega) \sim f_1(\pi) f_2(\mu^*) f_3(\Omega),$$

where $\delta_g(\cdot)$ denotes a discrete probability measure concentrated at g, $f_1(\pi)$ is specified by the stick-breaking prior as given in Equation (5), $f_2(\mu^*) = \prod_{g=1}^{G} f_2(\mu_g^*)$ and $f_3(\Omega) = \prod_{g=1}^{G} \prod_{j=1}^{q} f_3(\omega_{g_j})$. Here, μ_g^* and ω_{g_j} can be specified by

$$\mu_g^* | \xi, \Psi \overset{\text{i.i.d}}{\sim} N_q(\xi, \Psi), \ \xi | \xi^0, \Psi^0 \sim N_q(\xi^0, \Psi^0), \ \psi_j^{-1} | c_1, c_2 \sim \Gamma(c_1, c_2) \ \text{ for } j = 1, 2, \ldots, q, \tag{6}$$

$$\omega_{g_j}^{-1} | \varpi_j^a, \varpi_j \sim \Gamma(\varpi_j^a, \varpi_j) \ \text{ and } \ \varpi_j | \varpi_j^a, \varpi_j^b \sim \Gamma(\varpi_j^a, \varpi_j^b),$$

respectively, where $\Psi = \text{diag}(\psi_1, \psi_2, \ldots, \psi_q)$, $\Gamma(c_1, c_2)$ denotes the Gamma distribution with parameters c_1 and c_2, and $\xi^0, \Psi^0, c_1, c_2, \varpi_j^a, \varpi_j^a$ and ϖ_j^b are prespecified hyperparameters [18]. We consider $\xi^0 = 0_{q \times 1}$, $\Psi^0 = I_q$, $c_1 = 11$, $c_2 = 2.5$, $\varpi_j^a = 3$, $\varpi_j^a = n$ and $\varpi_j^b = 10$ in the paper. Given the values of L_i, μ^* and Ω, we can sample b_i from $N_q(\mu_{L_i}, \Omega_{L_i})$ with $\mu_{L_i} = \mu_{L_i}^* - \Sigma_{g=1}^{G} \mu_g^*$.

Hereafter we mainly introduce the variable selection principle of BLasso method [14,16] for the proposed joint model JMSRE. We need to identify not only the important variables in models (1) and (2). In particular, we also need to identify whether the parameter ϕ is $0_{q \times 1}$, and our proposed BLasso method does just that. In general, the prior distribution of the regression parameters is set to a multivariate normal distribution. Following the idea of Bayesian Lasso inference [14], we consider the following hierarchical priors for β, γ and ϕ:

$$\beta | H_\beta \sim N_{p_1}(0, H_\beta), \ \text{ with } \ H_\beta = \text{diag}(h_{\beta_1}^2, h_{\beta_2}^2, \cdots, h_{\beta_{p_1}}^2),$$
$$f(h_{\beta_1}^2, h_{\beta_2}^2, \cdots, h_{\beta_{p_1}}^2) = \prod_{j=1}^{p_1} \frac{\vartheta_{\beta_j}^2}{2} \exp\left(-\frac{\vartheta_{\beta_j}^2}{2} h_{\beta_j}^2\right), \tag{7}$$

$$\gamma | H_\gamma \sim N_{p_2}(0, H_\gamma), \ \text{ with } \ H_\gamma = \text{diag}(h_{\gamma_1}^2, h_{\gamma_2}^2, \cdots, h_{\gamma_{p_2}}^2),$$
$$f(h_{\gamma_1}^2, h_{\gamma_2}^2, \cdots, h_{\gamma_{p_2}}^2) = \prod_{j=1}^{p_2} \frac{\vartheta_{\gamma_j}^2}{2} \exp\left(-\frac{\vartheta_{\gamma_j}^2}{2} h_{\gamma_j}^2\right), \tag{8}$$

$$\boldsymbol{\phi}|\boldsymbol{H}_\phi \sim N_q(0, \boldsymbol{H}_\phi), \text{ with } \boldsymbol{H}_\phi = \text{diag}(h_{\phi_1}^2, h_{\phi_2}^2, \cdots, h_{\phi_q}^2),$$

$$f(h_{\phi_1}^2, h_{\phi_2}^2, \cdots, h_{\phi_q}^2) = \prod_{j=1}^{q} \frac{\vartheta_{\phi_j}^2}{2} \exp\left(-\frac{\vartheta_{\phi_j}^2}{2} h_{\phi_j}^2\right), \tag{9}$$

where $\boldsymbol{\vartheta}_\beta = \left\{\vartheta_{\beta_1}, \vartheta_{\beta_2}, \cdots, \vartheta_{\beta_{p_1}}\right\}$, $\boldsymbol{\vartheta}_\gamma = \left\{\vartheta_{\gamma_1}, \vartheta_{\gamma_2}, \cdots, \vartheta_{\gamma_{p_2}}\right\}$ and $\boldsymbol{\vartheta}_\phi = \left\{\vartheta_{\phi_1}, \vartheta_{\phi_2}, \cdots, \vartheta_{\phi_q}\right\}$ are the regularization parameters that control the tail decay. In particular, in order to better control the effect of tail decay, this paper sets different regularization parameters for different components of the same parameter. Inspired by Park and Casella [14], we further consider the following super-priories for these tuning parameters:

$$\vartheta_{\beta_j}^2 \sim \Gamma\left(a_{\vartheta_\beta}, b_{\vartheta_\beta}\right), \quad j = 1, 2, \cdots, p_1, \tag{10}$$

$$\vartheta_{\gamma_j}^2 \sim \Gamma\left(a_{\vartheta_\gamma}, b_{\vartheta_\gamma}\right), \quad j = 1, 2, \cdots, p_2, \tag{11}$$

$$\vartheta_{\phi_j}^2 \sim \Gamma\left(a_{\vartheta_\phi}, b_{\vartheta_\phi}\right), \quad j = 1, 2, \cdots, q. \tag{12}$$

3.2. Bayesian analysis of joint model

To obtain Bayesian estimates of unknown parameters $\boldsymbol{\beta}, \boldsymbol{\varphi}, \sigma^2, \boldsymbol{b}, \boldsymbol{\gamma}$ and $\boldsymbol{\phi}$, a hybrid algorithm combining the block Gibbs sampler and the Metropolis-Hastings algorithm is employed to iteratively draw $\{\boldsymbol{\beta}, \boldsymbol{\varphi}, \sigma^2, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{\phi}\}$ as follows.

(A) Conditional distribution of $\boldsymbol{\beta}$

It follows from Equation (3) and Equation (7) that the conditional posterior distribution $p(\boldsymbol{\beta}|\boldsymbol{\varphi}, \sigma^2, \boldsymbol{b}, \boldsymbol{y}^*)$ is expressed as

$$p(\boldsymbol{\beta}|\boldsymbol{\varphi}, \sigma^2, \boldsymbol{b}, \boldsymbol{y}^*) \propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{\sigma^2}\left(y_{ij}^* - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} - \boldsymbol{B}^\top(t_{ij})\boldsymbol{\varphi} - \boldsymbol{Z}_{ij}^\top\boldsymbol{b}_i\right) + \boldsymbol{\beta}^\top\boldsymbol{H}_\beta^{-1}\boldsymbol{\beta}\right]\right\},$$

which yields

$$\boldsymbol{\beta}|\boldsymbol{\varphi}, \sigma^2, \boldsymbol{b}, \boldsymbol{y}^* \sim N_{p_1}\left(\mathcal{A}_\beta, \mathcal{V}_\beta\right),$$

where $\mathcal{V}_\beta = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\sigma^{-2}\boldsymbol{X}_{ij}^\top\boldsymbol{X}_{ij} + \boldsymbol{\Sigma}_\beta^{-1}$, $\mathcal{A}_\beta = \mathcal{V}_\beta\left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\boldsymbol{X}_{ij}\left(y_{ij}^* - \boldsymbol{Z}_{ij}^\top\boldsymbol{b}_i - \boldsymbol{B}^\top(t_{ij})\boldsymbol{\varphi}\right)\sigma^{-2}\right)$.

(B) Conditional distribution of $\boldsymbol{\varphi}$ According to Equation (3) and the prior of $\boldsymbol{\varphi}$ in Equation (4), the conditional distribution $p(\boldsymbol{\varphi}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{b}, \boldsymbol{y}^*)$ is given by

$$p(\boldsymbol{\varphi}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{b}, \boldsymbol{y}^*) \propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{\sigma^2}\left(y_{ij}^* - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} - \boldsymbol{B}^\top(t_{ij})\boldsymbol{\varphi} - \boldsymbol{Z}_{ij}^\top\boldsymbol{b}_i\right) + \boldsymbol{\varphi}^\top(\boldsymbol{H}_\varphi^0)^{-1}\boldsymbol{\varphi}\right]\right\},$$

which yields

$$\boldsymbol{\varphi}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{b}, \boldsymbol{y}^* \sim N_L\left(\mathcal{A}_\varphi, \mathcal{V}\varphi\right),$$

where $\mathcal{V}\varphi = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\sigma^{-2}\boldsymbol{B}(t_{ij})^\top\boldsymbol{B}(t_{ij}) + \boldsymbol{\Sigma}_\varphi^{-1}$, $\mathcal{A}\varphi = \mathcal{V}\varphi\left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\boldsymbol{B}(t_{ij})\left(y_{ij}^* - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} - \boldsymbol{Z}_{ij}^\top\boldsymbol{b}_i\right)\sigma^{-2}\right)$.

(C) Conditional distribution of σ^{-2}

It follows from Equation (3) and the prior of σ^2 in Equation (4) that the conditional distribution $p(\sigma^{-2}|\boldsymbol{\beta}, \boldsymbol{\varphi}, \boldsymbol{b}, \boldsymbol{y}^*)$ is given by

$$p(\sigma^{-2}|\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{b},\boldsymbol{y}^*) \propto \exp\left\{-\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(y_{ij}^* - \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} - \boldsymbol{B}^{\top}(t_{ij})\boldsymbol{\varphi} - \boldsymbol{Z}_{ij}^{\top}\boldsymbol{b}_i\right) + b_0\right)\sigma^{-2}\right\}$$
$$\times (\sigma^{-2})^{\frac{1}{2}\sum_{i=1}^{n}n_i + a_0 - 1},$$

which yields

$$\sigma^{-2}|\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{b},\boldsymbol{y}^* \sim \Gamma\left(\frac{1}{2}\sum_{i=1}^{n}n_i + a_0, \frac{1}{2}\left(y_{ij}^* - \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} - \boldsymbol{B}^{\top}(t_{ij})\boldsymbol{\varphi} - \boldsymbol{Z}_{ij}^{\top}\boldsymbol{b}_i\right) + b_0\right).$$

(D) Conditional distribution of b_i

For reasons of space, the sampling of $\boldsymbol{b}_i, i = 1, 2, \ldots, n$ follows the steps in Appendix A, which can also be seen in Tang et al. [24]

(E) Conditional distribution of γ

It follows from Equation (3) and Equation (8) that the conditional distribution $p(\gamma|\boldsymbol{b},\boldsymbol{\phi},\boldsymbol{T},\boldsymbol{\Delta})$ is proportional to

$$\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{W}_i^{\top}\gamma + \boldsymbol{\phi}^{\top}\boldsymbol{b}_i)}{\sum_{j\in\mathcal{R}_i}\exp(\boldsymbol{W}_j^{\top}\gamma + \boldsymbol{\phi}^{\top}\boldsymbol{b}_j)}\right]^{\delta_i}\exp\left\{-\frac{1}{2}\gamma^T\boldsymbol{H}_{\gamma}^{-1}\gamma\right\},$$

which is not a familiar distribution. Therefore, the well-known Metropolis-Hastings (MH) algorithm is adopted to simulate observations from the above given conditional distribution, which is implemented as follows. Given the current value $\gamma^{(m)}$, new candidates γ are generated from $N_{p_2}(\gamma^{(m)}, \sigma_{\gamma}^2\boldsymbol{\Sigma}_{\gamma})$, where $\boldsymbol{\Sigma}_{\gamma} = \left(-\partial^2(\ln(p(\gamma|\boldsymbol{b},\boldsymbol{\phi},\boldsymbol{T},\boldsymbol{\Delta}))/\partial\gamma\partial\gamma^{\top}|_{\gamma=\gamma^{(m)}}\right)^{-1}$. The new $\gamma^{(m)}$ is accepted with probability

$$\min\left\{1, \frac{p(\gamma|\boldsymbol{b},\boldsymbol{\phi},\boldsymbol{T},\boldsymbol{\Delta})}{p(\gamma^{(m)}|\boldsymbol{b},\boldsymbol{\phi},\boldsymbol{T},\boldsymbol{\Delta})}\right\},$$

where

$$\boldsymbol{\Sigma}_{\gamma} = \left(\sum_{i=1}^{n}\frac{\sum_{j\in\mathcal{R}_i}\exp(\cdot)\boldsymbol{W}_j\boldsymbol{W}_j^{\top}\sum_{j\in\mathcal{R}_i}\exp(\cdot) + \sum_{j\in\mathcal{R}_i}\exp(\cdot)\boldsymbol{W}_j\sum_{j\in\mathcal{R}_i}\exp(\cdot)\boldsymbol{W}_j^{\top}}{\left(\sum_{j\in\mathcal{R}_i}\exp(\cdot)\right)^2} + \boldsymbol{H}_{\gamma}^{-1}\right)^{-1}.$$

with $\exp(\cdot) = \exp(\boldsymbol{W}_j^{\top}\gamma^{(m)} + \boldsymbol{\phi}^{\top}\boldsymbol{b}_j)$. The variance coefficient σ_{γ}^2 can be chosen such that the average acceptance rate are approximately 0.25 or more.

(F) Conditional distribution of $\boldsymbol{\phi}$

From Equation (3) and Equation (9), the conditional distribution $p(\boldsymbol{\phi}|\gamma,\boldsymbol{b},\boldsymbol{T},\boldsymbol{\Delta})$ is proportional to

$$\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{W}_i^{\top}\gamma + \boldsymbol{\phi}^{\top}\boldsymbol{b}_i)}{\sum_{j\in\mathcal{R}_i}\exp(\boldsymbol{W}_j^{\top}\gamma + \boldsymbol{\phi}^{\top}\boldsymbol{b}_j)}\right]^{\delta_i}\exp\left\{-\frac{1}{2}\gamma^T\boldsymbol{H}_{\phi}^{-1}\gamma\right\},$$

which is not a familiar distribution. Similar to above (E), given the current value $\boldsymbol{\phi}^{(m)}$, new candidates $\boldsymbol{\phi}$ are generated from $N_q(\boldsymbol{\phi}^{(m)}, \sigma_{\phi}^2\boldsymbol{\Sigma}_{\phi})$, where $\boldsymbol{\Sigma}_{\phi} = \left(-\partial^2(\ln(p(\boldsymbol{\phi}|\boldsymbol{b},\gamma,\boldsymbol{T},\boldsymbol{\Delta}))/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^{\top}|_{\boldsymbol{\phi}=\boldsymbol{\phi}^{(m)}}\right)^{-1}$. The new $\boldsymbol{\phi}^{(m)}$ is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\phi}|\gamma,\boldsymbol{b},\boldsymbol{T},\boldsymbol{\Delta})}{p(\boldsymbol{\phi}^{(m)}|\gamma,\boldsymbol{b},\boldsymbol{T},\boldsymbol{\Delta})}\right\},$$

where

$$\boldsymbol{\Sigma}_{\phi} = \left(\sum_{i=1}^{n}\frac{\sum_{j\in\mathcal{R}_i}\exp(\cdot)\boldsymbol{b}_j\boldsymbol{b}_j^{\top}\sum_{j\in\mathcal{R}_i}\exp(\cdot) + \sum_{j\in\mathcal{R}_i}\exp(\cdot)\boldsymbol{b}_j\sum_{j\in\mathcal{R}_i}\exp(\cdot)\boldsymbol{b}_j^{\top}}{\left(\sum_{j\in\mathcal{R}_i}\exp(\cdot)\right)^2} + \boldsymbol{H}_{\phi}^{-1}\right)^{-1},$$

and $\exp(\cdot) = \exp(W_j^\top \gamma + \phi^{(m)^\top} b_j)$. The variance coefficient σ_ϕ^2 can be chosen such that the average acceptance rate are approximately 0.25 or more.

By the above iterative process, we can obtain a series of sample $\{(\beta^{(m)}, \varphi^{(m)}, \sigma^{2(m)}, b_i^{(m)}, \gamma^{(m)}, \phi^{(m)}) : m = 1, 2, \ldots, M\}$. Then, Bayesian estimates of $\beta, \varphi, \sigma^2, b_i, \gamma$ and ϕ can be obtained by

$$\hat{\beta} = \frac{1}{M} \sum_{m=1}^{M} \beta^{(m)}, \ \hat{\varphi} = \frac{1}{M} \sum_{m=1}^{M} \varphi^{(m)}, \ \hat{\sigma^2} = \frac{1}{M} \sum_{m=1}^{M} \sigma^{2(m)},$$

$$\hat{b}_i = \frac{1}{M} \sum_{m=1}^{M} b_i^{(m)} \ \hat{\gamma} = \frac{1}{M} \sum_{m=1}^{M} \gamma^{(m)}, \ \hat{\phi} = \frac{1}{M} \sum_{m=1}^{M} \phi^{(m)}.$$

Similarly, the consistent estimates of the posterior covariance matrices of $\mathbf{var}(\beta|y^*, X, Z)$, $\mathbf{var}(\varphi|y^*, X, Z)$, $\mathbf{var}(\sigma^2|y^*, X, Z)$, $\mathbf{var}(\gamma|W, T, \Delta)$, and $\mathbf{var}(\phi|y^*, X, Z, W, T, \Delta)$ can be obtained via the sample covariance matrices. For example

$$\hat{\mathbf{var}}(\beta|y^*, X, Z) = \frac{1}{M-1} \sum_{m=1}^{M} (\beta^{(m)} - \hat{\beta})(\beta^{(m)} - \hat{\beta})^\top.$$

Therefore, the variance of the corresponding parameter can be obtained through the diagonal elements of the sample covariance matrix of the random sample sequence.

4. Simulation Studies

In this section, we conducted three simulation studies to investigate the finite performance of the above proposed methods.

We considered the model defined in models (1) and (2) with 200 individuals. Here, the specific model is as follows:

$$y_{ij}^*|b_i = X_{1ij}\beta_1 + X_{2ij}\beta_2 + X_{3ij}\beta_3 + X_{4ij}\beta_4 + X_{5ij}\beta_5 + X_{6ij}\beta_6 + \sin\left(2\pi t_{ij}\right) + b_i + \varepsilon_{ij},$$

$$\lambda_i\left(t|b_i\right) = \lambda_0\left(t\right) \exp\left(W_{1i}\gamma_1 + W_{2i}\gamma_2 + W_{3i}\gamma_3 + W_{4i}\gamma_4 + \phi b_i\right).$$

The data were generated as follows: observation time t_{ij} was generated randomly between 0 and 3, the covariates X_{1ij} and X_{6ij} followed the Bernoulli distribution with the success probability 0.5 and 0.3, respectively. The covariates $X_{2ij}, X_{3ij}, X_{4ij}$ and X_{5ij} were generated form the multivariate normal distribution $N_4\left(\mathbf{0}, \Sigma\right)$ with mean vector $\mathbf{0}$ and covariance matrix Σ. And the covariance matrix Σ is a symmetric positive definite matrix whose diagonal elements are 1 and the rest are 0.5. The random error ε_{ij} is generated from the normal distribution with mean 0 and variance $\sigma^2 = 0.6^2$. And $W_i = (W_{i1}, W_{i2}, W_{i3}, W_{i4})^\top = (X_{3i1}, X_{4i1}, X_{5i1}, X_{6i1})^\top$. The baseline hazard function $\lambda_0\left(t\right) = 0.7$ and $\phi = 0.6$. The censoring time C_i was generated from the Uniform distribution $U\left[0, 3\right]$, and T_i^* was generated from the Exponential distribution with mean $1/\lambda_i\left(t|b_i\right)$, $T_i = \min(T_i^*, C_i)$. Our main purpose is to use the proposed approaches to identify the unimportant covariates and estimate nonzero coefficients. Bayesian results were obtained from 200 replications.

To show that our proposed method above can both accurately estimate parameters of interest and identify unimportant variables, but also flexibly capture the feature of nonlinear function $g\left(t\right)$ and random effects b_i, we consider three different simulation studies as follows.

Simulation I.

$$\beta = (\beta_1, \cdots, \beta_6)^\top = (1, 0, 0, -0.5, 0.5, -1)^\top, \quad \gamma = (\gamma_1, \cdots, \gamma_4)^\top = (0, 1, -0.5, 0)^\top,$$

$$g(t) = \sin\left(\frac{3}{4}\pi t\right), \quad b_i \overset{\text{i.i.d}}{\sim} 0.6N(-0.8, 0.1^2) + 0.4N(1.2, 0.5^2),$$

Simulation II.

$$\beta = (\beta_1, \cdots, \beta_6)^\top = (1, 0, 0, -0.5, 0.5, -1)^\top, \quad \gamma = (\gamma_1, \cdots, \gamma_4)^\top = (0, 1, -0.5, 0)^\top,$$

$$g(t) = t, \quad b_i \overset{\text{i.i.d}}{\sim} 0.4N(0, 0.3^2) + 0.3N(-1.5, 0.1^2) + 0.3N(1.5, 0.1^2),$$

Simulation III.

$$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_6)^\top = (1, 0.5, -0.5, -0.5, 0.5, -1)^\top, \quad \boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_4)^\top = (-0.5, 1, -0.5, 1)^\top,$$

$$g(t) = t^2, \quad b_i \overset{\text{i.i.d}}{\sim} N(0, 0.8^2).$$

which the mean censoring rates of the survival times in the three simulation studies are 44%, 45%, and 37%, respectively.

For each of the above generated data sets, the proposed semiparametric Bayesian procedure was used to simultaneously evaluate Bayesian estimates of unknown parameters and select the important covariates. The prior hyperparameters $a_{\vartheta_\beta} = a_{\vartheta_\gamma} = a_{\vartheta_\phi} = 1, b_{\vartheta_\gamma} = b_{\vartheta_\beta} = b_{\vartheta_\phi} = 0.1$, corresponding to the hyperparameters of the hyper priors for the adjustment coefficients in Equations (10)-(12). $a_0 = 1, b_0 = 1$ and $\boldsymbol{H}_\phi^0 = 100I_4$, corresponding to the prior parameters of $\sigma^2, \boldsymbol{\varphi}$. We set the degree of B-splines $d = 3$, the number of konts $K = 4$, and $G = 25$. To investigate the convergence of the proposed algorithm, we calculated the estimated potential scale reduction (EPSR) values of parameters. In addition, we also need to test whether the nonlinear function fitted by the B-splines method is convergent. It can be seen from Figure 1 that the EPSR values were less than 1.2 after about 3000 iterations in three simulation studies. Thus, 3000 ($M = 3000$) observations were collected to calculated Bayesian estimates of parameters after 3000 iterations in producing Bayesian results for each of 200 replications. Results obtained under three simulation studies were reported in Table 1, where "Bias" is the difference between the true value and the mean of the estimates based on 200 replications, "RMS" is the root mean square between the estimates based on 200 replications and its true value, and "F0" is the proportion that parameter is identified to be zero in 200 replications in terms of the criterion that a parameter is identified to be 0 if its 95% confidence interval contains 0.
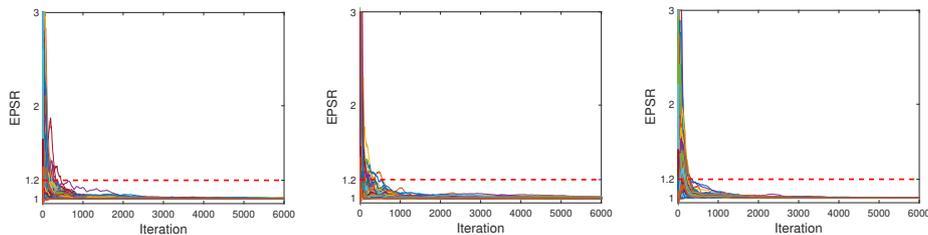

Figure 1. EPSR values of all parameters against iteration numbers for a randomly selected replication in Simulation I (left panel), Simulation II (middel panel) and Simulation III (right panel).

Table 1. Bayesian estimates of parameters in the simulation studies.

Pra.	True	Simulation I			Simulation II			Simulation III			
		Bias	RMS	F0(%)	Bias	RMS	F0(%)	True	Bias	RMS	F0(%)
β_1	1.00	-0.005	0.058	0.0	-0.080	0.112	0.0	1.00	-0.015	0.104	0.0
β_2	0.00	0.002	0.036	99.0	0.001	0.049	98.0	0.50	-0.009	0.077	0.0
β_3	0.00	-0.002	0.033	99.0	0.005	0.042	99.5	-0.50	0.011	0.085	0.0
β_4	-0.50	0.005	0.041	0.0	0.003	0.056	0.0	-0.50	0.001	0.079	0.0
β_5	0.50	-0.001	0.037	0.0	-0.009	0.063	0.0	0.50	0.000	0.083	0.0
β_6	-1.00	0.013	0.066	0.0	0.006	0.091	0.0	-1.00	0.028	0.136	0.0
γ_1	0.00	0.010	0.120	95.5	-0.004	0.105	98.0	-0.50	0.002	0.125	0.5
γ_2	1.00	0.001	0.148	0.0	-0.052	0.144	0.0	1.00	-0.008	0.141	0.0
γ_3	-0.50	0.018	0.138	5.0	0.039	0.136	6.5	-0.50	0.008	0.127	2.5
γ_4	0.00	-0.002	0.197	97.0	0.010	0.178	97.5	1.00	-0.007	0.191	0.0
ϕ	0.60	0.006	0.106	0.0	-0.019	0.088	0.0	0.60	-0.002	0.145	1.0
σ^2	0.36	-0.002	0.018	–	-0.001	0.019	–	0.36	0.005	0.020	–

Results of Table 1 indicated that (i) Bayesian estimates of parameters are reasonably accurate because their absolute biases are less than 0.10 and the root mean square of their RMS values are less than 0.20; (ii) BLasso could identify the correct models in most cases regardless of prior inputs of parameters because the F0 values

corresponding to the important covariates were less than 10%, but the F0 values corresponding to unimportant covariates were more than 90%. The performance of the proposed approach to recover the nonlinear function $g(t)$ can be measured by the root mean square error

$$\text{RASE}\left(g^{(r)}\right) = \sqrt{\frac{1}{300}\sum_{l=1}^{300}\left(g(u_l) - \hat{g}^{(r)}(u_l)\right)}, r = 1, 2, \cdots, 200,$$

where $\hat{g}^{(r)}(t) = \boldsymbol{B}^{\top}(t)\hat{\boldsymbol{\varphi}}^{(r)}$, $\hat{\boldsymbol{\varphi}}^{(r)}$ represents the Bayesian estimated value of the parameter vector $\boldsymbol{\varphi}$ in the r-th replication. Similar to the root mean square error of nonlinear function, we also calculate the root mean square error of random effects. Figure 2 plots the estimated curve and estimated density of the nonlinear function $\hat{g}(t)$ and the random effects b_i of the replication in which the mean of the RASE of the nonlinear function and the random effects is in the middle of the 200 replications, and against true curves and true density in three simulation studies, respectively. Inspection of Figure 2 showed that (i) the Bayesian B-splines method proposed in this paper can be flexible enough fit the true curve of the nonlinear function $g(t)$; (ii) the CDPMM method proposed is flexible enough to capture the general shapes of our considered three distribution assumptions for b_i; (iii) the estimated 95% confidence region of the nonlinear function can cover its true curve with a fairly narrow area. From Table 2, based on the results of 200 replications in three simulation studies, the estimated means and standard deviation (SD) of random effects b_i are enough near to their corresponding true values, and 25%, 50% and 75% quantile of the mean of the RASE of the nonlinear function and the random effects are enough small and close to illustrate that it is robust to apply CDPMM to estimate random effects. All these findings indicated that (i) our proposed Bayesian B-splines can be flexible enough fit the true curve of the nonlinear; (ii) Bayesian procedure could well capture the true information of b_i regardless of their true distributions and forms; (iii) BLasso could identify the true model with a high probability.
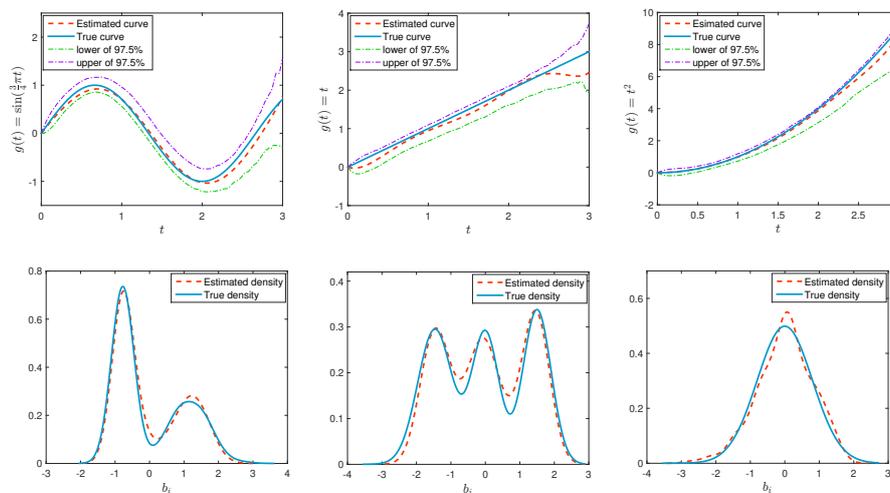

Figure 2. Estimated versus trues of nonlinear function $g(t)$ (upper panels,) and estimated versus true denisties for random effects b_i (lower panels) in Simulation I (left panel), Simulation II (middel panel) and Simulation III (right panel).

Table 2. Estimated mean and standard deviation for random effects, and quantiles of mean RASE for nonlinear functions and random effects.

	Est of random effects				Quantile of RASE		
	Mean	Est mean	SD	Est SD	25%	50%	75%
Simulation I	-0.011	-0.007	1.004	0.961	0.091	0.115	0.138
Simulation II	-0.023	0.014	1.268	1.202	0.135	0.181	0.239
Simulation III	-0.001	-0.001	0.879	0.797	0.109	0.152	0.227

5. An Example

In this section, we apply the method proposed in the previous sections to the MA.5 research experiment conducted by the Clinical Trial Group of the National Cancer Institute of Canada. The data contains 716 women with early-stage breast cancer before menopause. 356 patients were randomly selected to receive cyclophosphamide, epirubicin, and fluorouracil (CEF) adjuvant chemotherapy as the experimental group. The remaining 360 patients received cyclophosphamide, methotrexate, and fluorouracil (CMF) adjuvant chemotherapy as the control group of the trial. In clinical trials, visits were made before the start of treatment, each of the 6 treatment cycles, and every three months after treatment. At each visit, medical history and physical examination are accepted, and BCQ (Breast Cancer Questionnaire) is used to assess the patient's quality of life. The dataset has a total of 7807 observations. By the end of the study, a total of 366 patients had died, and the censorship rate was about 49%. For a detailed study of this data, please refer to the literature of Song et al. [22] and Levine et al. [24]. We linearly convert the evaluated BCQ score into a unit interval $(0, 1)$, and the longitudinal data constrained on the interval $(0, 1)$ is the longitudinal proportion data that we are interested in. The survival time of interest in the trial is the recurrence-free survival time (RFS), which is defined as the time from randomization to disease recurrence. Different treatment options, age, and the number of tumor-positive lymph nodes may directly affect RFS and the quality of life of patients. Similar to Tang et al. [17], we fitted the MA.5 research experiment data set to the following model:

$$y_{ij}^* | b_i = \beta_1 \text{EM}_i + \beta_2 \text{NODE\_POS}_i + \beta_3 \text{AGE}_i + g\left(t_{ij}\right) + b_i + \varepsilon_{ij},$$

$$\lambda_i\left(t | b_i\right) = \lambda_0\left(t\right) \exp\left(\gamma_1 \text{EM}_i + \gamma_2 \text{NODE\_POS}_i + \gamma_3 \text{AGE}_i + \phi b_i\right),$$

where y_{ij}^* represents the BCQ score after logit function transformation, EM_i is a two-class treatment index, $\text{EM}_i = 1$ indicates that the i-th patient underwent CEF treatment, $\text{EM}_i = 0$ indicates that the *i*-th patient underwent CMF treatment. Age and the number of lymph node metastases are binary variables. Patients who are less than or equal to 40 years old belong to the younger group, that is, $\text{AGE} = 1$. Patients who are older than 40 years old belong to the elderly group, that is, $\text{AGE} = 0$. When the number of lymph node metastases is 0-3, $\text{NODE\_POS} = 0$, otherwise it is 1. The nonlinear $g(t)$ is estimated by using cubic B-splines function, and the domain of cubic B-splines function is $[\min(t_{ij}), \max(t_{ij})]$. The prior distributions and the values of all hyperparameters in the case study are the same as those set in the simulation study above. Based on the above settings, we calculated EPSR values for all parameters and are shown in the left panel in Figure 3, which shows that after about 3000 iterations, all EPSR values are less than 1.2. Therefore, we use the 3000 iterations after the 3000th time to calculate the Bayesian estimation. The example analysis results are shown in Table 3.
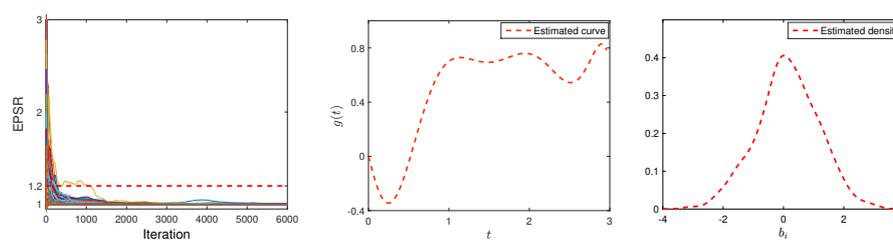

Figure 3. EPSR values of all parameters (left panel), estimated curve for nonlinear function $g(t)$ (middle panel) and estimated density for random effects b_i (right panel) in the the MA.5 research experiment study.

Inspection of Figure 3 showed that (i) the estimated density of random effects is approximately normally distributed. The distribution settings for random effects in our Simulation III correspond to. This shows that our CDPMM method is effective and reasonable; (ii) the estimated nonlinear function first decreases, then increases and finally stabilizes with time t. Although there are some fluctuations in the tail, this is reasonable and can be ignored. From Table 3, we can see that (i) the risk ratio of randomly receiving CEF and CMF treatment is $\text{HR} = \exp\left(\gamma_1\right) = 71.605\%$, which means patients who randomly receive CEF chemotherapy have a lower risk; (ii) the credible interval $(0.171, 0.321)$ of β_1 does not contain 0, indicating that different adjuvant chemotherapy regimens have a significant impact on the QOL of patients and CEF chemotherapy is more toxic than CMF

chemotherapy; (iii) the risk ratio of the number of lymph node metastases greater than or equal to 4 to less than 4 is $HR = \exp(\gamma_2) = 205.238\%$, which means that the number of lymph node metastases is greater than or equal to 4. The risk of breast cancer recurrence is greater in patients, and the RFS is shorter; (iv)the regression coefficient β_2 of the patient's lymph node metastasis number greater than or equal to 4 is 0.309, and its credible interval does not contain 0, which is highly significant. This suggests that patients with more lymph nodes have a lower quality of life, which is also consistent with clinical experience; (v) the risk ratio between the young group and the old group is $HR = \exp(\gamma_3) = 182.212\%$, which means the risk of breast cancer recurrence is higher and the RFS is shorter in the young group; (vii) the credible interval $(0.168, 0.344)$ of $\beta_3 = 0.263$ does not contain 0, indicating that different ages have a significant impact on the QOL of patients. It indicates that the quality of life of the olderly group is better than that of the young group. (viii) $\phi = 0.259$ and the credible interval $(0.003, 0.542)$ of ϕ does not contain 0, which means that ϕ is significantly different from 0. This shows that there is a significant correlation between the longitudinal proportional data and survival data, and the JMSRE model proposed in this paper is applicable and reasonable for the analysis of the MA.5 research experiment data.

Table 3. Bayesian estimatations of parameters in the MA.5 research experiment study.

Par.	Est	SD	CI
β_1	0.241	0.038	(0.171, 0.321)
β_2	0.309	0.050	(0.214, 0.414)
β_3	0.263	0.046	(0.168, 0.344)
γ_1	-0.334	0.142	(-0.608,-0.057)
γ_2	0.719	0.143	(0.452, 0.999)
γ_3	0.600	0.145	(0.315, 0.880)
ϕ	0.259	0.135	(0.003, 0.542)
σ_2	0.180	0.003	(0.174, 0.186)

6. Concluding Remarks

In this paper, a semiparametric joint model for longitudinal proportional data and survival data is proposed by relaxing the normality assumption of random effects and without specifying a nonlinear function that affects longitudinal responses. The advantages of the proposed model include: (i) the proposed method improves the flexibility of jointly modeling longitudinal proportional data and survival data; (ii) the proposed B-splines method can flexibly capture different nonlinear functions; (iii) the proposed CDPMM method can flexibly capture the unimodal, bimodal and multimodal features of random effects; (iv) the computational burden is not heavy, e.g. it takes about 4 minutes to run the replication in the above simulation study and about 2 hours to run the breast cancer dataset.

Our simulation studies and example analysis show that the Bayesian estimation proposed based on the joint model are quite accurate and relatively robust. The nonlinear function curve estimated by Bayesian B-splines is more flexible and can capture the true characteristics of the nonlinear function. And CDPMM method could also well capture the true information of $\boldsymbol{b}_i$. BLasso method could identify the true model with a high probability. Compared with the method proposed by Song et al. [22] to jointly model longitudinal proportional data and survival data, the joint model proposed in this paper is more flexible.

The joint model of longitudinal proportional data and survival data proposed in this paper still has many unsolved problems and we need to solve the following problems in the future: (i) does not make any constraints on the form of the basic hazard function; (ii) the distribution of random effects does not depend on the normality assumption and still has robust inferences; (iii) consider more complex spline models, for example nodes can be automatically selected to improve the performance of the proposed model; (iv) consider a joint model for the variable longitudinal proportional outcome and the multivariate survival outcome.

Appendix A

Conditional distribution of $\boldsymbol{b}_i$

Let $\boldsymbol{\theta}_{b_i}$ denotes all unknown parameters associated with distribution of $\boldsymbol{b}_i$, for $i = 1, 2, \ldots, n$. $\boldsymbol{\theta}_{b_i}$ can be iteratively draw by using the following steps:

Step (a) Conditional distribution of $\boldsymbol{\zeta}$ given $(\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b})$ is given by

$$\boldsymbol{\zeta}|\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b} \sim N_q(\mathbb{A}, \mathbb{B}),$$

where $\mathbb{B} = (G\Psi^{-1} + (\Psi^0)^{-1})^{-1}$ and $\mathbb{A} = \mathbb{B}((\Psi^0)^{-1}\xi^0 + \Psi^{-1}\sum_{g=1}^G \mu_g^*)$.

Step (b) For $j = 1, 2, \ldots, q$, the diagonal elements of Ψ is conditionally distributed as

$$\psi_j^{-1}|\mu^*, \xi \sim \Gamma(c_1 + \frac{G}{2}, c_2 + \frac{1}{2}\sum_{g=1}^G (\mu_{g_j}^* - \xi_j)^2),$$

where $\mu_{g_j}^*$ is the jth element of μ_g^* and ξ_j is the jth element of ξ.

Step (c) For $j = 1, 2, \ldots, q$, $\varpi_j|\Omega$ is conditionally distributed as

$$\varpi_j|\Omega \sim \Gamma(\varpi_j^a, \varpi_j^b + \sum_{g=1}^G \omega_{g_j}^{-1}),$$

where ω_{g_j} is the j th diagonal element of Ω_g.

Step (d) Following Ishwaran and Zarepour [20], given π, the conditional distribution of τ can be given by

$$\tau|\pi \sim \Gamma(a_1 + G - 1, a_2 - \sum_{g=1}^{G-1} \log(1 - v_g^*)),$$

where v_g^* is a random weight sampled from the the following beta distribution.

Step (e) Given L and τ, the conditional distribution of π can be obtained by following generalized Dirichlet distribution:

$$\pi|L, \tau \sim \text{Dir}(a_1^*, b_1^*, a_2^*, b_2^*, \ldots, a_{G-1}^*, b_{G-1}^*).$$

where $a_g^* = 1 + d_g, b_g^* = \tau + \sum_{l=g+1}^G d_l$ for $g = 1, 2, \ldots, G - 1$, and d_g is the number of $L_i's$ whose value equals to g. v_g^* is independently generated from a Beta distribution (a_g^*, b_g^*). Then, $\pi_1, \pi_2, \ldots, \pi_G$ are obtained from the following formula:

$$\pi_1 = v_1^*, \quad \pi_G = 1 - \sum_{g=1}^{G-1} \pi_g, \quad \text{and} \quad \pi_g = \prod_{l=1}^{g-1}(1 - v_l^*)v_g^*, \quad \text{for } g \neq 1 \text{ or } G.$$

Step (f) Let $L_1^*, L_2^*, \ldots, L_d^*$ be the d unique values of $\{L_1, L_2, \ldots, L_n\}$ (i.e., unique number of "clusters"), for $g = 1, 2 \ldots, G, \mu_g^*$ is conditionally distributed as follows:

$$\mu_g^*|\xi, \Psi \sim N_q(\xi, \Psi) \text{ for } g \notin \{L_1^*, L_2^*, \ldots, L_d^*\},$$

$$\mu_g^*|\xi, \Psi, \Omega, L, b \sim N_q(\mathbb{E}_g, \mathbb{F}_g) \text{ for } g \in \{L_1^*, L_2^*, \ldots, L_d^*\},$$

where $\mathbb{F}_g = (\Psi^{-1} + \Sigma_{\{i:L_i=g\}}\Omega_i^{-1})^{-1}$ and $\mathbb{E}_g = \mathbb{F}_g(\Psi^{-1}\xi + \Sigma_{\{i:L_i=g\}}\Omega_i^{-1}b_i)$ for $g \in \{L_1^*, L_2^*, \ldots, L_d^*\}$. Given μ_g^*, $\mu_g = \mu_g^* - \Sigma_{g=1}^G \pi_g\mu_g^*, \mu^* = \{\mu_1^*, \mu_2, \ldots, \mu_G^*\}$ and $\mu = \{\mu_1, \mu_2, \ldots, \mu_G\}$.

Step (g) Given g, for $j = 1, 2, \ldots, q$, the jth diagonal element of Ω_g is conditionally distributed as

$$\omega_{g_j} \sim \Gamma(\varpi_j^a, \varpi_j) \text{ for } g \notin \{L_1^*, L_2^*, \ldots, L_d^*\},$$

$$\omega_{g_j} \sim \Gamma(\frac{d_g}{2} + \varpi_j^a, \varpi_j + \sum_{\{i:L_i=g\}} \frac{1}{2}(b_{i_j} - \mu_{g_j})^2) \text{ for } g \in \{L_1^*, L_2^*, \ldots, L_d^*\},$$

where b_{i_j} is the jth element of b_i and μ_{g_j} is the jth element of μ_g. Given ω_{g_j}, $\Omega_g = \text{diag}(\omega_{g_1}, \omega_{g_2}, \ldots, \omega_{g_q})$ and $\Omega = \{\Omega_1, \Omega_2, \ldots, \Omega_G\}$.

Step (h) Given π, μ, Ω, b, the conditional distribution of L_i is obtained by

$$L_i|\pi, \mu, \Omega, b \overset{\text{i.i.d}}{\sim} \text{Multinomial}(\pi_{ig}^*, g = 1, 2, \ldots, G),$$

where π_{ig}^* is proportional to $(\pi_g p(b_i|\mu_g, \Omega_g)$ with $b_i|\mu_g, \Omega_g \sim N_q(\mu_g, \Omega_g)$, and $\pi_g(g = 1, 2, \ldots, G)$ are sampled from step (e). Given L_i, μ and Ω, the prior of b_i is distributed as $N_q(\mu_{L_i}, \Omega_{L_i})$ with μ_{L_i} and Ω_{L_i} being the L_i elements of sets μ and Ω, respectively.

Step (i) The conditional distribution $p(\boldsymbol{b}_i|\boldsymbol{\beta},\boldsymbol{\varphi},\sigma^2,\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{Y}^*,\boldsymbol{T},\boldsymbol{\Delta},\boldsymbol{W})$ is non-standard and cannot be derived directly via Gibbs sampling for $i=1,2,\ldots,n$. Specially,

$$p(\boldsymbol{b}_i|\boldsymbol{\beta},\boldsymbol{\varphi},\sigma^2,\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{Y}^*,\boldsymbol{T},\boldsymbol{\Delta}) \propto p(\boldsymbol{b}_i|\boldsymbol{\mu}_{L_i},\boldsymbol{\Omega}_{L_i})p\left(\boldsymbol{Y}_i^*|\boldsymbol{b}_i;\boldsymbol{\theta}_Y\right)p\left(\boldsymbol{T},\boldsymbol{\Delta}|\boldsymbol{b};\boldsymbol{\theta}_T\right).$$

The Metropolis-Hastings algorithm used to sample $\boldsymbol{b}_i$ is implemented as follows. At the mth iteration with a current value $\boldsymbol{b}_i^{(m)}$, a new candidate $\boldsymbol{b}_i$ is drawn from the normal distribution $N_q(\boldsymbol{b}_i^{(m)},\sigma_b^2\boldsymbol{\Sigma}_{b_i})$, where $\boldsymbol{\Sigma}_{b_i}=(\boldsymbol{\Omega}_{L_i}^{-1}+\boldsymbol{\Xi}_i)^{-1}$ and $\boldsymbol{\Xi}_i=-\partial^2(\ln(p\left(\boldsymbol{Y}_i^*|\boldsymbol{b}_i;\boldsymbol{\theta}_Y\right)p\left(\boldsymbol{T},\boldsymbol{\Delta}|\boldsymbol{b};\boldsymbol{\theta}_T\right))/\partial\boldsymbol{b}_i\partial\boldsymbol{b}_i^\top|_{\boldsymbol{b}_i=\boldsymbol{b}_i^{(m)}}$. The new $\boldsymbol{b}_i$ is accepted with probability

$$\min\left\{1,\frac{p(\boldsymbol{b}_i|\boldsymbol{\mu}_{L_i},\boldsymbol{\Omega}_{L_i})p\left(\boldsymbol{Y}_i^*|\boldsymbol{b}_i;\boldsymbol{\theta}_Y\right)p\left(\boldsymbol{T},\boldsymbol{\Delta}|\boldsymbol{b};\boldsymbol{\theta}_T\right)}{p(\boldsymbol{b}_i^{(m)}|\boldsymbol{\mu}_{L_i},\boldsymbol{\Omega}_{L_i})p\left(\boldsymbol{Y}_i^*|\boldsymbol{b}_i^{(m)};\boldsymbol{\theta}_Y\right)p\left(\boldsymbol{T},\boldsymbol{\Delta}|\boldsymbol{b}_i^{(m)},\boldsymbol{b}_{-i};\boldsymbol{\theta}_T\right)}\right\},$$

where $\boldsymbol{b}_{-i}$ represents the remaining random effects except the random effects of the ith individual. The variance σ_b^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

References

1. Faucett, C.L.; Thomas, D.C. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Stat. Med.* **1996**, *15*, 1663–1685. https://doi.org/10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1.

2. Wulfsohn, M.S.; Tsiatis, A.A. A Joint model for survival and longitudinal data measured with error. *Biometrics* **1997**, *53*, 330–339. https://doi.org/10.2307/2533118.

3. Schluchter, M.D. Methods for the analysis of informatively censored longitudinal data. *Stat. Med.* **1992**, *11*, 1861–1870. https://doi.org/10.1002/sim.4780111408.

4. Tsiatis, A.A.; Degruttola, V.; Wulfsohn, M.S. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *J. Am. Stat. Assoc.* **1995**, *90*, 27–37. https://doi.org/10.1080/01621459.1995.10476485.

5. Tsiatis, A.A.; Davidian, M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat. Sin.* **2004**, *14*, 809–834. https://doi.org/doi:10.1007/s00440-004-0348-9.

6. Yu, M.; Law, N.J.; Taylor, J.M.G.; Sandler, H.M. Joint longitudinal-survival-cure models and their application to prostate cancer. *Stat. Sin.* **2004**, *14*, 835–862. https://doi.org/10.1007/s00440-004-0348-9.

7. Diggle, P.J.; Sousa, I; Chetwynd, A.G. Joint modelling of repeated measurements and time-to-event outcomes: the fourth armitage lecture. *Stat. Med.* **2008**, *27*, 2981–2998. https://doi.org/10.1002/sim.3131.

8. Lawrence, A.; Gould, et al. Joint modeling of survival and longitudinal non-survival data: current methods and issues. *Stat. Med.* **2015**, *34*, 2181–2195. https://doi.org/10.1002/sim.6141.

9. Henderson, R.; Diggle, P.; Dobson, A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* **2000**, *1*, 465-480. https://doi.org/10.1093/biostatistics/1.4.465.

10. Zeng, D.; Cai, J. Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures. *Lifetime Data Analysis* **2005**, *11*, 151-174. https://doi.org/10.1007/s10985-004-0381-0.

11. Tang, A.; Zhao, X.; Tang, N. Bayesian variable selection and estimation in semiparametric joint models of multivariate longitudinal and survival data. *Biom. J.* **2017**, *59*, 57–78. https://doi.org/10.1002/bimj.201500070.

12. Song P.X.; Tan, M. Marginal models for longitudinal continuous proportional data. *Biometrics* **2000**, *56*, 496–502 https://doi.org/10.1111/j.0006-341X.2000.00496.x.

13. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B.* **1996**, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

14. Park T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681-686. https://doi.org/10.1198/016214508000000337.

15. Hans, C. Bayesian lasso regression. *Biometrika* **2009**, *96*, 835–845. https://doi.org/10.1093/biomet/asp047.

16. Guo, R.; Zhu, H.; Chow, S.M.; Ibrahim, J.G. Bayesian Lasso for semiparametric sructural equation models. *Biometrics* **2012**, *68*, 567–577. https://doi.org/10.1111/j.1541-0420.2012.01751.x.

17. Tang, A.; Zhao, X.; Tang, N. Bayesian variable selection and estimation in semiparametric joint models of multivariate longitudinal and survival data. *Biom. J.* **2017**, *59*, 57–78. https://doi.org/10.1002/bimj.201500070.

18. Ohlssen, D.I.; Sharples, L.D.; Spiegelhalter, D.J. Flexible random effects models using Bayesian semiparametric models: applications to institutional comparisons. *Stat. Med.* **2007**, *26*, 2088–2112. https://doi.org/10.1002/sim.2666.

19. Yang, M.; Dunson, D.B.; Baird, D. Semiparametric bayes hierarchical models with mean and variance constraints. *Comput. Stat. Data. Anal.* **2010**, *54*, 2172-2186. https://doi.org/10.1016/j.csda.2010.03.025.

20. Ishwaran, H.; Zarepour, M. Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. *Biometrika* **2000**, *87*, 371–390. https://doi.org/10.1093/biomet/87.2.371.

21. Chow, S.; Tang, N.; Yuan, Y.; Song, X.; Zhu, H. Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using the Dirichlet process prior. *Br. J. Math. Stat. Psychol.* **2011**, *64*, 69–106. https://doi.org/10.1348/000711010X497262.

22. Song, H.; Peng, Y; Tu, D. Jointly modeling longitudinal proportional data and survival times with an application to the quality of life data in a breast cancer trial. *Lifetime Data Analysis* **2017**, *23*, 183–206. https://doi.org/10.1007/s10985-015-9346-8.

23. Levine, M., Bramwell, V., Pritchard, K., et al. Randomized trial of intensive cyclophosphamide, epirubicin, and fluorouracil chemotherapy compared with cyclophosphamide, methotrexate, and fluorouracil in premenopausal women with node-positive breast cancer. *J. Clini. Oncol.* **1998**, *16*, 2651-2658. https://doi.org/10.1200/JCO.1998.16.8.2651.

24. Tang, A.; Duan, X.; Zhao, Y. Bayesian variable selection and estimation in semiparametric simplex mixed-effects models with longitudinal proportional data. *Entropy* **2022**, *24*, 1466. https://doi.org/10.3390/e24101466.