mathematics

MDPI

Article

Bayesian Joint Modeling Analysis of Longitudinal Proportional 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.); tam1973as@ynu.edu.cn (A.T.)
* Correspondence: lihuiqiong@ynu.edu.cn; Tel.: +86-871-65033700

Abstract: This paper focuses on a joint model to analyze longitudinal proportional and survival data. We utilize a logit transformation on the longitudinal proportional data and employ a partially linear mixed-effect model. With this model, we estimate the unknown function of time using the B-splines technique. Additionally, we introduce a centered Dirichlet process mixture model (CDPMM) to capture the random effects, allowing for a flexible distribution. The survival data are assumed using a Cox proportional hazard model, and the sharing random effects joint model is developed for the two types of data. We develop a Bayesian Lasso (BLasso) approach that combines the Gibbs sampler and the Metropolis–Hastings algorithm. The proposed method allows for the estimation of unknown parameters and the selection of significant covariates simultaneously. We evaluate the performance of our proposed methods through simulation studies and also provide an illustration of our methodologies using an example from the MA.5 research experiment.

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

MSC: 62N02

Citation: Liu, W.; Li, H.; Tang, A.; Cui, Z. Bayesian Joint Modeling Analysis of Longitudinal Proportional and Survival Data. *Mathematics* **2023**, *11*, 3469. https://doi.org/10.3390/math11163469

Academic Editors: Duchwan Ryu and Jiancang Zhuang

Received: 5 July 2023
Revised: 5 August 2023
Accepted: 8 August 2023
Published: 10 August 2023

1. Introduction

The joint analysis of longitudinal and survival data has gained widespread application in clinical studies on cancer and HIV/AIDS, where the primary endpoints typically involve time-to-event outcomes such as disease-free and overall survival. Notably, following the seminal work by Faucett and Thomas [1] and Wulfsohn and Tsiatis [2], the standard joint model has been extensively investigated. Researchers have extensively discussed the advantages of joint models [3–8]. However, certain patients with a compromised quality of life (QOL) may opt to discontinue their participation in the clinical trials due to disease recurrence, or they may experience mortality. In this case, the absence of QOL measures resulting from the withdrawal of patients provides informative insights into the trade-off between intensive treatment and poor QOL. To establish strong evidence, we conducted joint modeling of longitudinal life measures and survival data to investigate their relationship. For the 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 QOL 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 quality-of-life measurement data can be linearly converted into longitudinal proportional data whose value range is in the unit interval (0, 1) [12]. Song and Tan [12] emphasized that disregarding the constraint of having values between 0 and 1 could lead to erroneous interpretations. For the longitudinal component, there are two methods to deal with it. The first method applied the classic linear mixed model to the longitudinal proportional data after logit transformation [13], and the second method directly used the simplex distribution to model

the longitudinal proportional data [14,15]. The models established using 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 for the logit transformed longitudinal proportional data and use the B-splines method to model the unknown function in the model. Meanwhile, to enhance the feasibility of our proposed model, we use the CDPMM method to model random effects.

In addition, variable selection in the joint model is also considered. In traditional regression models, variable selection methods include forward selection, backward elimination, stepwise selection, and the use of information criteria such as the Akaike information criterion (AIC). However, these approaches can be computationally expensive and unstable when dealing with complex models that have a large number of covariates. To address this issue, penalized likelihood methods have been proposed, with one popular method being the Lasso of Tibshirani [16]. The Lasso estimates linear regression coefficients by applying a constraint on the L_1 norm of the least squares. Tibshirani [16] proposed that Lasso estimates can be interpreted as posterior norm estimates when the regression parameters have independent and identically Laplacian priors. Park and Casella [17] extended this idea under the Bayesian framework and introduced the Bayesian Lasso (BLasso) variable selection method. They used a double exponential prior for the regression coefficients and a gamma distribution for the shrinkage parameter. The BLasso method has been successfully applied to various models, including linear regression [18], semiparametric structural equation models [19], and joint models of longitudinal and survival data [11]. Building on this work, our paper extends the BLasso variable selection method to the joint model of longitudinal proportional data and survival data. We propose an approach called BLasso, which aims to estimate unknown parameters while also identifying the significant effects of crucial covariates.

The rest of this paper is organized as follows. In Section 2, the joint model of longitudinal proportional and survival data is introduced. In Section 3, the Bayesian estimations of the joint model are proposed. In Section 4, three numerical simulations are presented to evaluate the performance of the proposed methods. In Section 5, we utilize the proposed approach to analyze the MA.5 research experiment's data. We then provide some concluding remarks in Section 6. For more technical information, please refer to Appendix A.

2. Model and Notation

Consider a dataset consisting of n individuals. Let y_{ij} be a longitudinal proportional measurement for the i-th individual ($i = 1, 2, \ldots, n$) at observation time point t_{ij} for $j = 1, 2, \ldots, n_i$, and $y_{ij} \in (0, 1)$, where n_i represents the number of observations of individual i. We assume that y_{ij}^* is the logit transformation of y_{ij} and $y_{ij}^* = \text{logit}(y_{ij}) = \log\left(\frac{1-y_{ij}}{y_{ij}}\right)$. Furthermore, T_i^* and C_i are the true survival time and censoring time, respectively. Additionally, we have the true survival time T_i^* and the censoring time C_i for each individual i. Let $T_i = \min(T_i^*, C_i)$ denote the corresponding observed event time. Let $\delta_i = 1(T_i^* \leq C_i)$ denote the failure indicator, where $1(\cdot)$ is an indicator function.

We denote $\boldsymbol{y}^* = \{\boldsymbol{y}_1^*, \boldsymbol{y}_2^*, \ldots, \boldsymbol{y}_n^*\}$, where $\boldsymbol{y}_i^* = \left\{y_{i1}^*, y_{i2}^*, \ldots, y_{in_i}^*\right\}$. Let $\boldsymbol{T} = \{T_1, T_2, \ldots, T_n\}$ and $\boldsymbol{\Delta} = \{\delta_1, \delta_2, \ldots, \delta_n\}$. The random effects $\boldsymbol{b} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ are time-independent and underlie both the longitudinal and survival processes for the i-th individual. Given the random effects $\boldsymbol{b}_i$, we assume that y_{ij}^* follows a 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; $\boldsymbol{\beta}$ is a $p_1 \times 1$ vector of 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. Additionally, 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 [20], we specify the random effects using a 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 the known fixed effects' design matrix $\boldsymbol{W}_i$ connects the unknown $p_2 \times 1$ parameter vector $\boldsymbol{\gamma}$ to $\lambda_i(t|\boldsymbol{b}_i)$. Additionally, the unknown $q \times 1$ parameter vector $\boldsymbol{\phi}$ links $\boldsymbol{b}_i$ to $\lambda_i(t|\boldsymbol{b}_i)$. Lastly, the basic hazard function $\lambda_0(t)$ remains unknown.

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). The parameter $\boldsymbol{\phi}$ in model JMSRE reflects the correlation between transformed longitudinal proportional data and survival data, given random effects. When $\boldsymbol{\phi} = \boldsymbol{0}_{q\times 1}$, it means that the longitudinal index is not necessarily related to the event time; i.e., 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 :

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

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

We denote $\boldsymbol{\theta}_y = \{\boldsymbol{\beta}, \boldsymbol{\varphi}, \sigma^2\}$ as the unknown parameters associated with model (1) and $\boldsymbol{\theta}_T = \{\boldsymbol{\gamma}, \boldsymbol{\phi}\}$ as the unknown parameters associated with model (2). Thus, given $(\boldsymbol{\theta}_y, \boldsymbol{\theta}_T, \boldsymbol{b})$, the 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

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

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

3. Bayesian Estimation of Joint Model

3.1. Prior Specification

In order to develop Bayesian inference on the considered models, it is necessary to specify the prior distributions for $\sigma^2, \boldsymbol{\varphi}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\phi}$. For conjugation, we consider the following priors for $\sigma^2, \boldsymbol{\varphi}$:

$$\frac{1}{\sigma^2} \sim \Gamma(a_0, b_0), \quad \boldsymbol{\varphi} \sim N_L\left(\boldsymbol{0}, \boldsymbol{H}_\varphi^0\right), \tag{4}$$

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. We can set the prior distribution to a non-informative prior distribution, which just needs to have a large variance. Thus, we consider $a_0 = 1$, $b_0 = 1$ and $\boldsymbol{H}_\varphi^0 = 100\ I_4$ in the paper.

As stated by Tang et al. [21], the random effects $\boldsymbol{b}_i$ can be modeled using a Dirichlet process (DP) mixture of normals. Specifically, we assume that $\boldsymbol{b}_i$ values are independently

and identically distributed according to a mixture distribution, where the mixture components are drawn from a DP with a base distribution $\mathcal{P}$ that has unknown parameters $(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g)$. To address the challenges of performing Bayesian estimation on the regression parameters $\boldsymbol{\beta}$ and the dispersion parameter σ^2 in the model (1), one common approach is to use a Dirichlet process (DP) prior to approximate the unknown form of $\mathcal{P}$. The DP prior is specified as $\mathcal{P} \sim \text{DP}(\tau F_0)$, where F_0 is a base distribution that serves as a starting point for constructing the nonparametric distribution, and τ is a weight that represents the researcher's certainty of F_0 being the distribution of $\mathcal{P}$. Sethuraman [22] demonstrated that the DP prior $\text{DP}(\tau F_0)$ can be represented using a stick-breaking prior. However, this representation has some limitations. It leads to a non-zero mean of random effects [23], which may not be desirable in certain cases. Additionally, it results in a discrete probability distribution for random effects [20], which may not accurately capture the underlying continuous distribution. The discrete Dirichlet processes proposed by Ishwaran and Zarepour [24] and Yang et al. [23] are commonly known as discrete Dirichlet processes. However, these methods may not be suitable for continuous underlying densities of random effects. Furthermore, violating the assumption of a zero mean on the random effects can lead to non-identifiability in the random effects model. Additionally, the computational complexity of the discrete DP methods with a stick-breaking prior for random effects can be high for complex models. To tackle the mentioned challenges, Tang et al. [21] proposed a truncated-approximate centered Dirichlet process mixture model (CDPMM).

In order to address these challenges, we also adopt the CDPMM approach [21] in the model (1). This method allows us to 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 $\sum_{g=1}^{\infty} \pi_g = 1$. To ease the computational intensity, we consider $G = 25$, and π_g is given by the following stick-breaking procedure, as proposed by Ishwaran and Zarepour (2000) [24]:

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

Let ϑ_g be independent and identically distributed (i.i.d.) random variables following a Beta distribution with parameters $(1, \tau)$ for $g = 1, 2, \ldots, G-1$, and let $\vartheta_G = 1$. This implies that the sum of all π_g values is equal to 1. The prior distribution for the unknown parameter τ is a Gamma distribution with hyperparameters a_1 and a_2 [25]. Here, we take the hyperparameters a_1 and a_2 to be 25 and 5, respectively.

An efficient and flexible method for solving the DP prior specified above is to represent $\boldsymbol{b}_i$ in terms of a latent variable $L_i \in \{1, 2, \ldots, G\}$, which records each $\boldsymbol{b}_i$'s cluster membership and conveys its parametric value to the distribution of $\boldsymbol{b}_i$. Let $\boldsymbol{L} = \{L_1, L_2, \ldots, L_n\}$, $\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_G\}$, $\boldsymbol{\mu}^* = \{\boldsymbol{\mu}_1^*, \boldsymbol{\mu}_2^*, \ldots, \boldsymbol{\mu}_G^*\}$ and $\boldsymbol{\Omega} = \{\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \ldots, \boldsymbol{\Omega}_G\}$, where $\boldsymbol{\Omega}_g = \text{diag}(\omega_{g1}, \omega_{g2}, \ldots, \omega_{gq})$. These variables can be reformulated as follows:

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

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

density functions of the random variables $\boldsymbol{\pi}$, $\boldsymbol{\mu}_g^*$, and ω_{g_j}. Here, $\boldsymbol{\mu}_g^*$ and ω_{g_j} can be specified by

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

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

respectively. Let $\boldsymbol{\Psi} = \text{diag}(\psi_1,\psi_2,\ldots,\psi_q)$, $\Gamma(c_1,c_2)$ denote the Gamma distribution with parameters c_1 and c_2, and $\boldsymbol{\xi}^0$, $\boldsymbol{\Psi}^0$, c_1, c_2, ω_j^a, ω_j^a and ω_j^b are prespecified hyperparameters [20]. The following values are used in the paper: $\boldsymbol{\xi}^0 = \mathbf{0}_{q\times 1}$, $\boldsymbol{\Psi}^0 = I_q$, $c_1 = 11$, $c_2 = 2.5$, $\omega_j^a = 3$, $\omega_j^a = n$, and $\omega_j^b = 10$. Given the values of L_i, $\boldsymbol{\mu}^*$, and $\boldsymbol{\Omega}$, we can sample $\boldsymbol{b}_i$ from $N_q(\boldsymbol{\mu}_{L_i},\boldsymbol{\Omega}_{L_i})$ with $\boldsymbol{\mu}_{L_i} = \boldsymbol{\mu}_{L_i}^* - \Sigma_{g=1}^G \boldsymbol{\mu}_g^*$.

Here, we will mainly introduce the variable selection principle of the BLasso method [17,19] for the proposed joint model JMSRE. We need to identify not only the important variables in models (1) and (2) but also whether the parameter $\boldsymbol{\phi}$ is $\mathbf{0}_{q\times 1}$. Our proposed BLasso method accomplishes this. In general, the prior distribution of the regression parameters is set to a multivariate normal distribution. Based on the concept of Bayesian Lasso inference [17], we adopt hierarchical priors for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\phi}$ as follows:

$$\begin{aligned}\boldsymbol{\beta}|\boldsymbol{H}_\beta &\sim N_{p_1}(0,\boldsymbol{H}_\beta),\ \text{with}\ \boldsymbol{H}_\beta = \text{diag}(h_{\beta_1}^2,h_{\beta_2}^2,\ldots,h_{\beta_{p_1}}^2),\\ f(h_{\beta_1}^2,h_{\beta_2}^2,\ldots,&h_{\beta_{p_1}}^2) = \prod_{j=1}^{p_1}\frac{\vartheta_{\beta_j}^2}{2}\exp\left(-\frac{\vartheta_{\beta_j}^2}{2}h_{\beta_j}^2\right),\end{aligned} \quad (7)$$

$$\begin{aligned}\boldsymbol{\gamma}|\boldsymbol{H}_\gamma &\sim N_{p_2}(0,\boldsymbol{H}_\gamma),\ \text{with}\ \boldsymbol{H}_\gamma = \text{diag}(h_{\gamma_1}^2,h_{\gamma_2}^2,\ldots,h_{\gamma_{p_2}}^2),\\ f(h_{\gamma_1}^2,h_{\gamma_2}^2,\ldots,&h_{\gamma_{p_2}}^2) = \prod_{j=1}^{p_2}\frac{\vartheta_{\gamma_j}^2}{2}\exp\left(-\frac{\vartheta_{\gamma_j}^2}{2}h_{\gamma_j}^2\right),\end{aligned} \quad (8)$$

$$\begin{aligned}\boldsymbol{\phi}|\boldsymbol{H}_\phi &\sim N_q(0,\boldsymbol{H}_\phi),\ \text{with}\ \boldsymbol{H}_\phi = \text{diag}(h_{\phi_1}^2,h_{\phi_2}^2,\ldots,h_{\phi_q}^2),\\ f(h_{\phi_1}^2,h_{\phi_2}^2,\ldots,&h_{\phi_q}^2) = \prod_{j=1}^{q}\frac{\vartheta_{\phi_j}^2}{2}\exp\left(-\frac{\vartheta_{\phi_j}^2}{2}h_{\phi_j}^2\right),\end{aligned} \quad (9)$$

where $\boldsymbol{\vartheta}_\beta = \left\{\vartheta_{\beta_1},\vartheta_{\beta_2},\ldots,\vartheta_{\beta_{p_1}}\right\}$, $\boldsymbol{\vartheta}_\gamma = \left\{\vartheta_{\gamma_1},\vartheta_{\gamma_2},\ldots,\vartheta_{\gamma_{p_2}}\right\}$, and $\boldsymbol{\vartheta}_\phi = \left\{\vartheta_{\phi_1},\vartheta_{\phi_2},\ldots,\vartheta_{\phi_q}\right\}$ are the regularization parameters that control the tail decay. In particular, 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 [17], we further consider the following super-priorities for these tuning parameters:

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

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

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

3.2. *Bayesian Analysis of Joint Model*

To obtain Bayesian estimates of the unknown parameters $\boldsymbol{\beta},\boldsymbol{\varphi},\sigma^2,\boldsymbol{b},\boldsymbol{\gamma}$, and $\boldsymbol{\phi}$, we use a hybrid algorithm that combines the block Gibbs sampler and the Metropolis–Hastings algorithm. This algorithm iteratively draws samples for these parameters.

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

According to Equations (3) and (7), the conditional posterior distribution $p(\boldsymbol{\beta}|\boldsymbol{\varphi},\sigma^2,\boldsymbol{b},\boldsymbol{y}^*)$ is given by

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

which yields

$$\boldsymbol{\beta}|\boldsymbol{\varphi},\sigma^2,\boldsymbol{b},\boldsymbol{y}^* \sim N_{p_1}(\mathcal{A}_\beta,\mathcal{V}_\beta), \tag{13}$$

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

(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

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

which yields

$$\boldsymbol{\varphi}|\boldsymbol{\beta},\sigma^2,\boldsymbol{b},\boldsymbol{y}^* \sim N_L(\mathcal{A}_\varphi,\mathcal{V}_\varphi), \tag{14}$$

where $\mathcal{V}_\varphi^{-1} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{\sigma^2}\boldsymbol{B}(t_{ij})\boldsymbol{B}(t_{ij})^\top + (\boldsymbol{H}_\varphi^0)^{-1}$, $\mathcal{A}_\varphi = \mathcal{V}_\varphi\left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{\sigma^2}\boldsymbol{B}(t_{ij})\left(y_{ij}^* - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} - \boldsymbol{Z}_{ij}^\top\boldsymbol{b}_i\right)\right)$.

(C) Conditional distribution of $\frac{1}{\sigma^2}$.

According to Equation (3) and the prior of σ^2 in Equation (4), the conditional distribution $p(\frac{1}{\sigma^2}|\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{b},\boldsymbol{y}^*)$ is given by

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

which yields

$$\frac{1}{\sigma^2}|\boldsymbol{\beta},\boldsymbol{\varphi},\boldsymbol{b},\boldsymbol{y}^* \sim \Gamma\left(\frac{1}{2}\sum_{i=1}^{n} n_i + a_0, \frac{1}{2}\left(y_{ij}^* - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} - \boldsymbol{B}^\top(t_{ij})\boldsymbol{\varphi} - \boldsymbol{Z}_{ij}^\top\boldsymbol{b}_i\right)^2 + b_0\right). \tag{15}$$

(D) Conditional distribution of $\boldsymbol{b_i}$.

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. [21].

(E) Conditional distribution of $\boldsymbol{\gamma}$.

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

$$\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{W}_i^\top\boldsymbol{\gamma} + \boldsymbol{\phi}^\top\boldsymbol{b}_i)}{\sum_{j\in\mathcal{R}_i}\exp(\boldsymbol{W}_j^\top\boldsymbol{\gamma} + \boldsymbol{\phi}^\top\boldsymbol{b}_j)}\right]^{\delta_i}\exp\left\{-\frac{1}{2}\boldsymbol{\gamma}^T\boldsymbol{H}_\gamma^{-1}\boldsymbol{\gamma}\right\}, \tag{16}$$

which is not a familiar distribution. Therefore, the well-known Metropolis–Hastings (MH) algorithm is adopted to simulate observations from the conditional distribution given above, which is implemented as follows. Given the current value $\boldsymbol{\gamma}^{(m)}$, new candidates $\boldsymbol{\gamma}$ are gen-

erated from $N_{p_2}(\boldsymbol{\gamma}^{(m)}, \sigma_\gamma^2 \boldsymbol{\Sigma}_\gamma)$, where $\boldsymbol{\Sigma}_\gamma = \left(-\partial^2(\ln(p(\boldsymbol{\gamma}|\boldsymbol{b}, \boldsymbol{\phi}, \boldsymbol{T}, \boldsymbol{\Delta}))/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}}\right)^{-1}$. The new $\boldsymbol{\gamma}^{(m)}$ is accepted with probability

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

where

$$\boldsymbol{\Sigma}_\gamma = \left(\sum_{i=1}^{n} \frac{\sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\boldsymbol{W}_j\boldsymbol{W}_j^\top \sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot) + \sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\boldsymbol{W}_j \sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\boldsymbol{W}_j^\top}{\left(\sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\right)^2} + \boldsymbol{H}_\gamma^{-1}\right)^{-1}. \tag{17}$$

with $\exp(\cdot) = \exp(\boldsymbol{W}_j^\top\boldsymbol{\gamma}^{(m)} + \boldsymbol{\phi}^\top\boldsymbol{b}_j)$. The variance coefficient σ_γ^2 can be adjusted to achieve an average acceptance rate of approximately 0.25 or higher.

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

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

$$\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{W}_i^\top\boldsymbol{\gamma} + \boldsymbol{\phi}^\top\boldsymbol{b}_i)}{\sum_{j\in\boldsymbol{\mathcal{R}}_i}\exp(\boldsymbol{W}_j^\top\boldsymbol{\gamma} + \boldsymbol{\phi}^\top\boldsymbol{b}_j)}\right]^{\delta_i} \exp\left\{-\frac{1}{2}\boldsymbol{\phi}^T\boldsymbol{H}_\phi^{-1}\boldsymbol{\phi}\right\}, \tag{18}$$

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 $\boldsymbol{\Sigma}_\phi = (-\partial^2(\ln(p(\boldsymbol{\phi}|\boldsymbol{\gamma}, \boldsymbol{b}, \boldsymbol{T}, \boldsymbol{\Delta}))/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top|_{\boldsymbol{\phi}=\boldsymbol{\phi}^{(m)}})^{-1}$. The new $\boldsymbol{\phi}^{(m)}$ is accepted with probability

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

where

$$\boldsymbol{\Sigma}_\phi = \left(\sum_{i=1}^{n} \frac{\sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\boldsymbol{b}_j\boldsymbol{b}_j^\top \sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot) + \sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\boldsymbol{b}_j \sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\boldsymbol{b}_j^\top}{\left(\sum_{j\in\boldsymbol{\mathcal{R}}_i} \exp(\cdot)\right)^2} + \boldsymbol{H}_\phi^{-1}\right)^{-1}, \tag{19}$$

and $\exp(\cdot) = \exp(\boldsymbol{W}_j^\top\boldsymbol{\gamma} + \boldsymbol{\phi}^{(m)^\top}\boldsymbol{b}_j)$. The variance coefficient σ_ϕ^2 can be adjusted to achieve an average acceptance rate of approximately 0.25 or higher.

Using the above iterative process, we can obtain a series of sample $\{(\boldsymbol{\beta}^{(m)}, \boldsymbol{\varphi}^{(m)}, \sigma^{2(m)}, \boldsymbol{b}_i^{(m)}, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\phi}^{(m)}) : m = 1, 2, \ldots, M\}$. Then, Bayesian estimates of $\boldsymbol{\beta}$, $\boldsymbol{\varphi}$, σ^2, $\boldsymbol{b}_i$, $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ can be obtained using

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

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

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

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

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

4. Simulation Studies

In this section, we perform three simulation studies to examine the finite performance of the previously mentioned methods.

The model used in these studies was the one defined in models (1) and (2), involving a total of 200 individuals. The specific details of the model are as follows:

$$y_{ij}^{*}|b_i = X_{1ij}\beta_1 + X_{2ij}\beta_2 + X_{3ij}\beta_3 + X_{4ij}\beta_4 + X_{5ij}\beta_5 + X_{6ij}\beta_6 + g(t_{ij}) + b_i + \varepsilon_{ij}, \tag{20}$$

$$\lambda_i(t|b_i) = \lambda_0(t)\exp\left(W_{1i}\gamma_1 + W_{2i}\gamma_2 + W_{3i}\gamma_3 + W_{4i}\gamma_4 + \phi b_i\right). \tag{21}$$

In model (1), $\boldsymbol{Z}_{ij}$ can be either one-dimensional or multi-dimensional. However, in the following simulation study, $\boldsymbol{Z}_{ij}$ was set to be one-dimensional. In order to perform variable selection on $\boldsymbol{X}_{ij}$ and $\boldsymbol{W}_{ij}$, $\boldsymbol{X}_{ij}$ and $\boldsymbol{W}_{ij}$ were set to be multi-dimensional in the simulation study. The data were generated as follows: observation time t_{ij} was randomly generated between 0 and 3. The covariates X_{1ij} and X_{6ij} followed a Bernoulli distribution with success probabilities of 0.5 and 0.3, respectively. The covariates X_{2ij}, X_{3ij}, X_{4ij}, and X_{5ij} were generated from a multivariate normal distribution $N_4(\mathbf{0}, \boldsymbol{\Sigma})$ with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. The covariance matrix $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix with diagonal elements of 1 and all other elements of 0.5. The random error ε_{ij} was generated from a normal distribution with mean 0 and variance $\sigma^2 = 0.6^2$. We define $\boldsymbol{W}_i = (W_{i1}, W_{i2}, W_{i3}, W_{i4})^\top = (X_{3i1}, X_{4i1}, X_{5i1}, X_{6i1})^\top$. The baseline hazard function $\lambda_0(t) = 0.7$ and $\phi = 0.6$. The censoring time C_i was generated from the uniform distribution $U[0, 3]$, and T_i^* was generated from the exponential distribution with mean $1/\lambda_i(t|b_i)$, $T_i = \min(T_i^*, C_i)$. Our main objective is to utilize the proposed approaches to identify insignificant covariates and estimate non-zero coefficients. Bayesian results were obtained from 200 replications.

To demonstrate the accuracy and flexibility of our proposed method, we conducted three simulation studies. These simulations aimed to estimate parameters of interest, identify unimportant variables, and capture the features of the unknown function $g(t)$ and random effects b_i. The true values of unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ were set to be the same in Simulation I and Simulation II, and the parameter's true values included 0. The true values of unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in Simulation III are all non-zero. The settings of the unknown function $g(t)$ and random effects b_i are different between the three simulation studies. The unknown function $g(t)$ setting includes both nonlinear and linear. The random effect b_i was set to follow a mixed normal distribution with unimodal, bimodal, and trimodal distributions, respectively. By conducting these simulation studies, we can showcase the effectiveness and versatility of our method.

Simulation I

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_6)^\top = (1, 0, 0, -0.5, 0.5, -1)^\top, \quad \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_4)^\top = (0, 1, -0.5, 0)^\top,$$

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

Simulation II

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_6)^\top = (1, 0, 0, -0.5, 0.5, -1)^\top, \quad \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_4)^\top = (0, 1, -0.5, 0)^\top,$$

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

Simulation III

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

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

We utilized the proposed semiparametric Bayesian procedure to simultaneously estimate unknown parameters and identify significant covariates in each of the three simulation studies. The mean censoring rates for the survival times in these studies were 44%, 45%, and 37%. The prior hyperparameters were set as follows: $a_{\vartheta_\beta} = a_{\vartheta_\gamma} = a_{\vartheta_\phi} = 1$, $b_{\vartheta_\gamma} = b_{\vartheta_\beta} = b_{\vartheta_\phi} = 0.1$. These hyperparameters correspond to the hyperpriors for the adjustment coefficients in Equations (10)–(12). We set $a_0 = 1$, $b_0 = 1$, and $\boldsymbol{H}_\varphi^0 = 100 I_4$, which correspond to the prior parameters of σ^2 and $\boldsymbol{\varphi}$. We set the degree of B-splines $d = 3$, the number of knots $K = 4$, and $G = 25$.

To assess the convergence of the proposed algorithm, we computed the estimated potential scale reduction (EPSR) values for the parameters. Additionally, we also need to test the convergence of the unknown function fitted using the B-splines method. Figure 1 indicates that the EPSR values remained consistently below 1.2 after around 3000 iterations in all three simulation studies. Consequently, we collected 3000 observations ($M = 3000$) to calculate the Bayesian estimates of the parameters after 3000 iterations in order to produce Bayesian results for each of the 200 replications. For comparison, we also applied Gaussian priors as the prior distribution of random effects. The purpose of these simulations is to compare the semi-parametric approach based on the CDPMM prior with the parametric approach based on the Gaussian prior from a Bayesian perspective. Results obtained from three simulation studies were reported in Tables 1–3, which include five measures: "Median", "Bias", "SD", "RMS", and "F0". "Median" represents the median of the estimates from 200 replications. "Bias" indicates the difference between the true value and the mean of the estimates from 200 replications. "SD" indicates the standard deviation of the estimates from 200 replications. "RMS" is the root mean square between the estimates from 200 replications and their true values. "F0" indicates the proportion of parameters identified as zero in 200 replications, considering a parameter to be identified as zero if its 95% confidence interval contains zero.

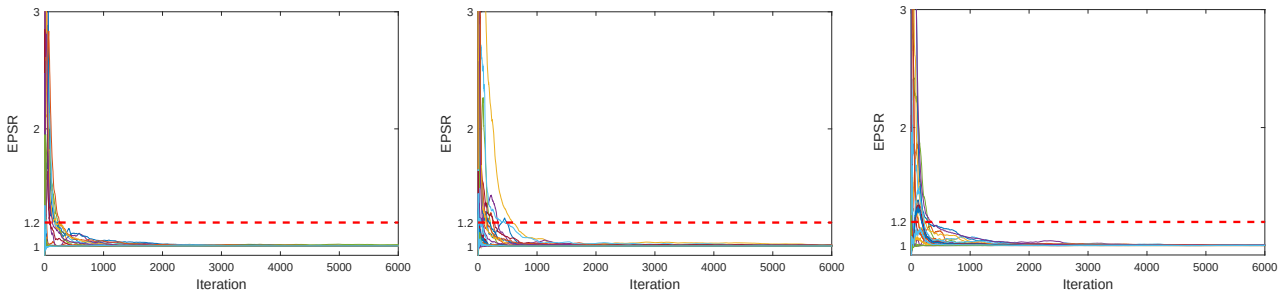

Figure 1. EPSR values of all parameters against iteration numbers for a randomly selected replication in Simulation I (**left** panel), Simulation II (**middle** panel), and Simulation III (**right** panel). The colored lines represent the EPSR values for all parameters, and the red dashed lines determine the number of iterations when all parameters converge.

Table 1. Bayesian estimates of parameters based on the CDPMM prior and Gaussian prior in Simulation I.

Pra.	True	CDPMM Prior					Gaussian Prior				
		Median	Bias	SD	RMS	F0 (%)	Median	Bias	Median	RMS	F0 (%)
β_1	1.00	0.993	−0.005	0.058	0.058	0.0	0.980	−0.016	0.124	0.125	0.0
β_2	0.00	0.001	0.002	0.036	0.036	99.0	−0.006	−0.010	0.086	0.087	95.5
β_3	0.00	−0.005	−0.002	0.033	0.033	99.0	−0.009	−0.005	0.090	0.090	95.0
β_4	−0.50	−0.493	0.005	0.041	0.041	0.0	−0.477	0.009	0.108	0.109	0.5
β_5	0.50	0.499	−0.001	0.037	0.037	0.0	0.509	0.008	0.104	0.104	0.5
β_6	−1.00	−0.985	0.013	0.064	0.066	0.0	−0.965	0.034	0.161	0.164	0.0
γ_1	0.00	0.010	0.010	0.119	0.120	95.5	−0.026	−0.028	0.122	0.125	95.5
γ_2	1.00	1.007	0.001	0.148	0.148	0.0	0.988	0.004	0.157	0.156	0.0
γ_3	−0.50	−0.478	0.018	0.137	0.138	5.0	−0.479	0.030	0.134	0.137	6.0
γ_4	0.00	0.010	−0.002	0.198	0.197	97.0	0.026	0.031	0.183	0.185	97.5
ϕ	0.60	0.610	0.006	0.106	0.106	0.0	0.598	0.032	0.121	0.125	0.0
σ^2	0.36	0.357	−0.002	0.018	0.018	–	0.363	0.004	0.021	0.022	–

Table 2. Bayesian estimates of parameters based on the CDPMM prior and Gaussian prior in Simulation II.

Pra.	True	CDPMM Prior					Gaussian Prior				
		Median	Bias	SD	RMS	F0 (%)	Median	Bias	Median	RMS	F0 (%)
β_1	1.00	0.932	−0.076	0.091	0.118	0.0	0.893	−0.113	0.142	0.182	0.0
β_2	0.00	0.003	0.002	0.047	0.047	98.5	0.000	0.003	0.098	0.098	93.0
β_3	0.00	0.004	0.000	0.049	0.049	99.0	0.000	0.003	0.109	0.109	94.5
β_4	−0.50	−0.494	0.009	0.057	0.058	0.0	−0.484	0.012	0.117	0.117	0.5
β_5	0.50	0.483	−0.013	0.056	0.058	0.0	0.487	−0.018	0.112	0.113	1.0
β_6	−1.00	−1.014	−0.007	0.101	0.101	0.0	−1.001	0.012	0.191	0.191	0.0
γ_1	0.00	0.002	0.002	0.106	0.106	98.5	−0.002	0.005	0.131	0.131	94.5
γ_2	1.00	0.964	−0.027	0.144	0.146	0.0	0.999	0.003	0.142	0.142	0.0
γ_3	−0.50	−0.457	0.031	0.128	0.131	3.5	−0.501	−0.001	0.144	0.143	6.0
γ_4	0.00	−0.006	−0.014	0.174	0.174	98.0	0.002	0.008	0.226	0.225	96.5
ϕ	0.60	0.571	−0.028	0.086	0.091	0.0	0.596	0.005	0.101	0.100	0.0
σ^2	0.36	0.361	0.000	0.020	0.020	–	0.365	0.005	0.021	0.021	–

Table 3. Bayesian estimates of parameters based on the CDPMM prior and Gaussian prior in Simulation III.

Pra.	True	CDPMM Prior					Gaussian Prior				
		Median	Bias	SD	RMS	F0 (%)	Median	Bias	Median	RMS	F0 (%)
β_1	1.00	0.990	−0.015	0.103	0.104	0.0	0.984	−0.013	0.102	0.102	0.0
β_2	0.50	0.487	−0.009	0.077	0.077	0.0	0.487	−0.013	0.083	0.084	0.0
β_3	−0.50	−0.488	0.011	0.085	0.085	0.0	−0.477	0.018	0.076	0.078	0.0
β_4	−0.50	−0.499	0.001	0.079	0.079	0.0	−0.507	−0.005	0.076	0.076	0.0
β_5	0.50	0.496	0.000	0.083	0.083	0.0	0.513	0.006	0.078	0.078	0.0
β_6	−1.00	−0.977	0.028	0.134	0.136	0.0	−0.976	0.024	0.133	0.135	0.0
γ_1	−0.50	−0.502	0.002	0.125	0.125	0.5	−0.479	0.022	0.124	0.126	2.5
γ_2	1.00	0.987	−0.008	0.141	0.141	0.0	0.983	−0.025	0.144	0.146	0.0
γ_3	−0.50	−0.492	0.008	0.127	0.127	2.5	−0.483	0.015	0.123	0.123	2.5
γ_4	1.00	0.997	−0.007	0.191	0.191	0.0	1.015	−0.001	0.205	0.205	0.0
ϕ	0.60	0.604	−0.002	0.145	0.145	1.0	0.626	0.021	0.157	0.158	0.5
σ^2	0.36	0.365	0.005	0.020	0.020	–	0.364	0.004	0.019	0.019	–

The results from Tables 1–3 suggest that the Bayesian estimates of the parameters are reasonably accurate. One can see that in all simulations, the proposed CDPMM prior performed better in both parameter estimation and inferential characteristics. This is

indicated by the fact that the bias (Bias) values of the results based on the CDPMM prior method are all less than 0.10, and the root mean square (RMS) value and standard deviation (SD) value are both less than 0.20. Furthermore, the BLasso method was able to correctly identify the important covariates in most cases, regardless of the prior inputs of parameters. This is supported by the fact that the F0 values corresponding to the important covariates were less than 10%, indicating a high level of significance. On the other hand, the F0 values corresponding to the unimportant covariates were more than 90%, indicating a lack of significance. The recovery performance of the proposed method for the unknown function $g(t)$ can be measured using the RMSE (the root mean square error), which is expressed as

$$\mathrm{RMSE}\left(g^{(r)}\right) = \sqrt{\frac{1}{300}\sum_{l=1}^{300}\left(g(u_l) - \hat{g}^{(r)}(u_l)\right)^2}, \; r = 1, 2, \ldots, 200, \tag{22}$$

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 RMSE of the unknown function, we also calculate the RMSE of the random effects. Figure 2 plots the estimated curve and estimated density of the unknown function $\hat{g}(t)$ and the random effects b_i of the replication based on different priors. The mean of the RMSE of the unknown function and the random effects is in the middle of the 200 replications and is compared against the true curves and true density in three simulation studies, respectively.

Upon inspection of Figure 2, it is evident that the Bayesian B-splines method proposed in this paper is flexible enough to accurately fit the true curve of the unknown function $g(t)$. Additionally, the CDPMM prior proposed demonstrates sufficient flexibility compared to the Gaussian prior to capture the general shapes of the three distribution assumptions considered for b_i. The results presented in Table 4, based on 200 replications in three simulation studies under the CDPMM prior and Gaussian prior, further support the robustness of the CDPMM method. The estimated means and standard deviations (SDs) of the random effects b_i closely align with their corresponding true values. Moreover, the 25%, 50%, and 75% quantiles of the RMSE of the unknown function and the random effects are sufficiently small, indicating the effectiveness of the CDPMM approach in estimating random effects.

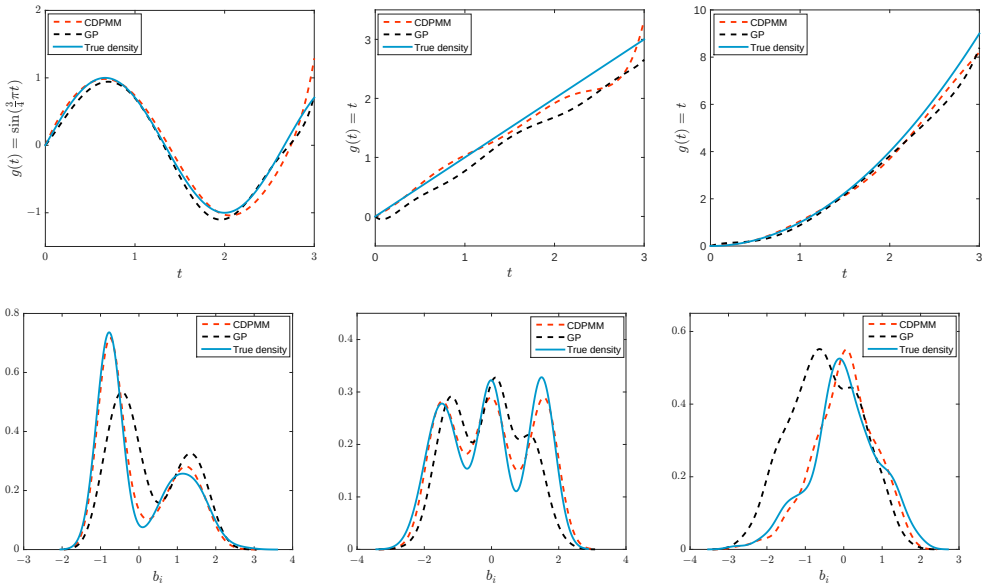

Figure 2. Estimation versus true values of unknown function $g(t)$ (**upper** panels) and estimated versus true densities for random effects b_i (**lower** panels) based on the CDPMM prior (CDPMM) and Gaussian prior (GP) method in Simulation I (**left** panels), Simulation II (**middle** panels), and Simulation III (**right** panels).

All these findings show that, compared with the Gaussian prior method, our CDPMM prior method makes the Bayesian B-spline curve flexible enough to accurately fit the real curve of nonlinear data. Additionally, the Bayesian procedure effectively captures the true information of b_i, regardless of their true distributions and forms. Furthermore, BLasso has a high probability of correctly identifying the true model.

Table 4. Estimated mean and standard deviation for random effects and quantiles of RMSE for unknown functions and random effects based on the CDPMM prior (CDPMM) and Gaussian prior (GP) method in three simulation studies.

	Method	Est of Random Effects				Quantile of RMSE		
		Mean	Est Mean	SD	Est SD	25%	50%	75%
Simulation I	CDPMM	−0.011	−0.007	1.004	0.961	0.091	0.115	0.138
	GP	0.004	−0.036	1.052	0.903	0.130	0.159	0.198
Simulation II	CDPMM	−0.040	0.009	1.251	1.236	0.086	0.112	0.139
	GP	−0.151	−0.040	1.224	1.135	0.088	0.111	0.140
Simulation III	CDPMM	−0.001	−0.001	0.879	0.797	0.109	0.152	0.227
	GP	0.018	0.004	0.739	0.663	0.102	0.144	0.209

"Mean" denotes true empirical mean of the distribution; "Est mean" denotes mean of the posterior samples. "SD" denotes true empirical standard deviation of the distribution; "Est SD" denotes standard deviation of the posterior samples.

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 pertain to 716 women with early-stage breast cancer before menopause. A total of 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, during each of the six treatment cycles, and every three months after treatment. At each visit, medical history and physical examination were conducted, and the Breast Cancer Questionnaire (BCQ) is used to assess the patient's QOL. The dataset consists of a total of 7807 observations. By the end of the study, 366 patients had died, resulting in a censoring rate of approximately 49%. For a detailed study of these data, please refer to Song et al. [26] and Levine et al. [27]. We linearly convert the evaluated BCQ score into a unit interval $(0, 1)$, and the longitudinal data constrained to the interval $(0, 1)$ are the longitudinal proportional data of interest. The trial focuses on the recurrence-free survival time (RFS), which is the duration between randomization and disease recurrence. Different treatment options, age, and the number of tumor-positive lymph nodes may directly affect RFS and the patient's QOL. We fitted the MA.5 research experiment dataset to the following model:

$$y_{ij}^{*}|b_i = \beta_1 \mathrm{EM}_i + \beta_2 \mathrm{NODE\_POS}_i + \beta_3 \mathrm{AGE}_i + g(t_{ij}) + b_i + \varepsilon_{ij}, \tag{23}$$

$$\lambda_i(t|b_i) = \lambda_0(t) \exp\left(\gamma_1 \mathrm{EM}_i + \gamma_2 \mathrm{NODE\_POS}_i + \gamma_3 \mathrm{AGE}_i + \phi b_i\right), \tag{24}$$

where variable y_{ij}^{*} represents the BCQ score after applying the logit function transformation. EM_i is a two-class treatment index, where $\mathrm{EM}_i = 1$ indicates that the i-th patient underwent CEF treatment, and $\mathrm{EM}_i = 0$ indicates that the i-th patient underwent CMF treatment. Age and the number of lymph node metastases are binary variables. Patients who are 40 years old or younger are classified as belonging to the younger group, denoted as AGE = 1. Patients who are older than 40 years old belong to the elderly group, denoted as AGE = 0. When the number of lymph node metastases is 0–3, NODE_POS = 0; otherwise, it is 1.

The term $g(t)$ in Equation (23) represents an unknown function related to the observation time t.

The unknown function $g(t)$ is estimated using a cubic B-spline function, and the domain of the cubic B-spline function is $[\min(t_{ij}), \max(t_{ij})]$. The prior distributions and 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. The results indicate that after approximately 3000 iterations, all EPSR values are less than 1.2. Therefore, we use the 3000 iterations after the 3000th iteration to calculate the Bayesian estimation. The results of the example analysis are shown in Table 5 based on two different prior methods.

Table 5. Bayesian estimations of parameters based on the CDPMM prior and Gaussian prior in the MA.5 experimental research study.

Pra.	CDPMM Prior			Gaussian Prior		
	Est	**SD**	**IC**	**Est**	**SD**	**IC**
β_1	0.239	0.035	(0.176, 0.312)	0.242	0.045	(0.163, 0.331)
β_2	0.304	0.041	(0.219, 0.377)	0.296	0.043	(0.220, 0.379)
β_3	0.269	0.049	(0.180, 0.375)	0.275	0.059	(0.166, 0.387)
γ_1	−0.341	0.150	(−0.625, −0.033)	−0.316	0.152	(−0.636, −0.048)
γ_2	0.745	0.133	(0.480, 1.013)	0.747	0.141	(0.472, 1.017)
γ_3	0.628	0.136	(0.355, 0.902)	0.611	0.154	(0.310, 0.934)
ϕ	0.269	0.126	(0.036, 0.519)	0.292	0.133	(0.020, 0.551)
σ^2	0.180	0.003	(0.174, 0.186)	0.180	0.003	(0.174, 0.186)

From Table 5, the following observations can be made. (i) The parameter estimation based on the CDPMM prior proposed in this paper has a smaller standard deviation (SD) and a shorter confidence interval than that based on the Gaussian prior. This suggests that the approach proposed in this paper is more effective. (ii) Under the CDPMM prior, the risk ratio of randomly receiving CEF and CMF treatment is HR $= \exp(\gamma_1) = 71.106\%$, implying that patients who randomly receive CEF chemotherapy have a lower risk. (iii) The credible interval $(0.176, 0.312)$ for β_1 does not include 0, indicating that different adjuvant chemotherapy regimens have a significant impact on patients' QOL. Additionally, it suggests that CEF chemotherapy is more toxic than CMF chemotherapy. (iv) The risk ratio for the number of lymph node metastases being greater than or equal to four compared to less than four is calculated as HR $= \exp(\gamma_2) = 210.644\%$. This implies that patients with a higher number of lymph node metastases have a greater risk of breast cancer recurrence and a shorter RFS. (v) The regression coefficient β_2 for lymph node metastasis numbers greater than or equal to 4 is 0.304, and its credible interval does not include 0, indicating high significance. This suggests that patients with a higher number of lymph nodes experience a lower QOL, which aligns with clinical experience; (vi) The risk ratio between the young group and the old group is HR $= \exp(\gamma_3) = 187.386\%$, implying that the risk of breast cancer recurrence is higher and the RFS is shorter in the young group; (vii) The credible interval $(0.180, 0.375)$ for $\beta_3 = 0.269$ does not contain 0, indicating that age has a significant impact on where variable y_{ij}^* represents the BCQ score after applying the logit function transformation. This suggests that the quality of life for the elderly group is better than that of the young group. (viii) The value of ϕ is 0.269, and the credible interval for ϕ is $(0.036, 0.519)$, which does not include 0. This indicates that ϕ is significantly different from 0, suggesting a significant correlation between the longitudinal proportional data and survival data. Therefore, the JMSRE model proposed in this paper is applicable and reasonable for analyzing the MA.5 research experiment's data.

6. Concluding Remarks

In this paper, a semiparametric joint model is proposed for longitudinal proportional data and survival data. The model does not assume the normality of random effects and does not require the specification of an unknown function influencing longitudinal

responses. The proposed model offers several advantages. Firstly, it improves the flexibility of jointly modeling longitudinal proportional data and survival data. Secondly, the proposed B-splines method effectively captures different unknown functions in a flexible manner. Thirdly, compared to a Gaussian prior, the proposed CDPMM method accurately captures the unimodal, bimodal, and multimodal features of random effects. Lastly, the computational burden is not heavy, with the replication in the simulation study taking approximately 4 min and the breast cancer dataset taking about 78 min to run.

Our simulation studies and example analysis demonstrate that the Bayesian estimation approach proposed based on the joint model is accurate and robust. The use of Bayesian B-splines allows for a more flexible estimation of the unknown function curve, enabling it to capture the true characteristics of the unknown function more effectively. Additionally, compared with the Gaussian prior method, the CDPMM method effectively captures the true information of $\boldsymbol{b}_i$. Furthermore, the BLasso method has a high probability of correctly identifying the true model. In comparison to the method proposed by Song et al. [26] for jointly modeling longitudinal proportional data and survival data, the joint model proposed in this paper offers greater flexibility.

The joint model of longitudinal proportional data and survival data proposed in this paper still has many unsolved problems, and we need to address the following issues in the future: (i) It does not impose any constraints on the form of the basic hazard function. (ii) We should consider more complex spline models, such as automatically selecting nodes to enhance the performance of the proposed model. (iii) We should also explore a joint model for the variable longitudinal proportional outcome and the multivariate survival outcome.

Author Contributions: Methodology, W.L., A.T. and Z.C.; Writing—review & editing, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was partially supported by grants from the Natural Science Foundation of China [Grant Number 12261102], a grant from key project of the Yunnan Province Foundation, China [Grant Number 202001BB050049, 202201BF070001-004, 202301AS070044].

Data Availability Statement: The real data that are used to illustrate the proposed methods may be available from the corresponding author upon considerable request. The data are not publicly available due to ethical restrictions and privacy.

Acknowledgments: The authors wish to thank the Editor-in-Chief, the Associate Editor and two reviewers for their many helpful and insightful comments and suggestions that greatly improved the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Conditional Distribution of $\boldsymbol{b}_i$

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

Step (a) The conditional distribution of $\boldsymbol{\xi}$ given $(\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b})$ is a normal distribution given by

$$\boldsymbol{\xi}|\boldsymbol{\mu}^*, \boldsymbol{\Psi}, \boldsymbol{b} \sim N_q(\mathbb{A}, \mathbb{B}), \tag{A1}$$

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

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

$$\psi_j^{-1}|\boldsymbol{\mu}^*, \boldsymbol{\xi} \sim \Gamma(c_1 + \frac{G}{2}, c_2 + \frac{1}{2}\sum_{g=1}^{G}(\mu_{g_j}^* - \xi_j)^2), \tag{A2}$$

where $\mu_{g_j}^*$ is the jth element of $\boldsymbol{\mu}_g^*$ and ξ_j is the jth element of $\boldsymbol{\xi}$.

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

$$\omega_j|\boldsymbol{\Omega} \sim \Gamma(\omega_j^a, \omega_j^b + \sum_{g=1}^{G} \omega_{g_j}^{-1}), \tag{A3}$$

where ω_{g_j} is the jth diagonal element of $\boldsymbol{\Omega}_g$.

Step (d) According to Ishwaran and Zarepour [24], the conditional distribution of τ can be determined based on the given $\boldsymbol{\pi}$:

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

where ν_g^* represents a randomly sampled weight from the beta distribution.

Step (e) Given $\boldsymbol{L}$ and τ, the conditional distribution of $\boldsymbol{\pi}$ can be obtained by following generalized Dirichlet distribution:

$$\boldsymbol{\pi}|\boldsymbol{L}, \tau \sim \mathrm{Dir}(a_1^*, b_1^*, a_2^*, b_2^*, \ldots, a_{G-1}^*, b_{G-1}^*). \tag{A4}$$

where $a_g^* = 1 + d_g$ and $b_g^* = \tau + \sum_{\iota=g+1}^{G} d_\iota$ for $g = 1, 2, \ldots, G-1$. Here, d_g represents the number of $L_i's$ values that are equal to g, and ν_g^* is generated autonomously from a Beta distribution characterized by the parameters (a_g^*, b_g^*). Then, the values $\pi_1, \pi_2, \ldots, \pi_G$ are derived using the following formula:

$$\pi_1 = \nu_1^*, \;\; \pi_G = 1 - \sum_{g=1}^{G-1} \pi_g, \;\; \text{and} \;\; \pi_g = \prod_{\iota=1}^{g-1}(1 - \nu_\iota^*)\nu_g^*, \;\; \text{for } g \neq 1 \;\text{ or }\; G. \tag{A5}$$

Step (f) Let $L_1^*, L_2^*, \ldots, L_d^*$ represent the d distinct values of $\{L_1, L_2, \ldots, L_n\}$ (i.e., the unique number of "clusters"). For $g = 1, 2 \ldots, G$, the conditional distribution of $\boldsymbol{\mu}_g^*$ is as follows:

$$\boldsymbol{\mu}_g^*|\boldsymbol{\xi}, \boldsymbol{\Psi} \sim N_q(\boldsymbol{\xi}, \boldsymbol{\Psi}) \;\text{ for }\; g \notin \{L_1^*, L_2^*, \ldots, L_d^*\}, \tag{A6}$$

$$\boldsymbol{\mu}_g^*|\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{L}, \boldsymbol{b} \sim N_q(\mathbb{E}_g, \mathbb{F}_g) \;\text{ for }\; g \in \{L_1^*, L_2^*, \ldots, L_d^*\}, \tag{A7}$$

where $\mathbb{F}_g$ is defined as $(\boldsymbol{\Psi}^{-1} + \Sigma_{\{i:L_i=g\}}\boldsymbol{\Omega}_i^{-1})^{-1}$, and $\mathbb{E}_g$ is defined as $\mathbb{F}_g(\boldsymbol{\Psi}^{-1}\boldsymbol{\xi} + \Sigma_{\{i:L_i=g\}}\boldsymbol{\Omega}_i^{-1}\boldsymbol{b}_i)$ for $g \in \{L_1^*, L_2^*, \ldots, L_d^*\}$. Given $\boldsymbol{\mu}_g^*$, $\boldsymbol{\mu}_g = \boldsymbol{\mu}_g^* - \Sigma_{g=1}^{G}\pi_g\boldsymbol{\mu}_g^*$, $\boldsymbol{\mu}^* = \{\boldsymbol{\mu}_1^*, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_G^*\}$, and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_G\}$.

Step (g) Given a value g, for $j = 1, 2, \ldots, q$, the conditional distribution of the diagonal elements of $\boldsymbol{\Omega}_g$ is as follows:

$$\omega_{g_j} \sim \Gamma(\omega_j^a, \omega_j) \;\text{ for }\; g \notin \{L_1^*, L_2^*, \ldots, L_d^*\}, \tag{A8}$$

$$\omega_{g_j} \sim \Gamma(\frac{d_g}{2} + \omega_j^a, \omega_j + \sum_{\{i:L_i=g\}} \frac{1}{2}(b_{i_j} - \mu_{g_j})^2) \;\text{ for }\; g \in \{L_1^*, L_2^*, \ldots, L_d^*\}, \tag{A9}$$

where b_{i_j} represents the jth element of vector $\boldsymbol{b}_i$, while μ_{g_j} denotes the jth element of vector $\boldsymbol{\mu}_g$. Additionally, given the value of ω_{g_j}, we can construct the diagonal matrix $\boldsymbol{\Omega}_g = \mathrm{diag}(\omega_{g_1}, \omega_{g_2}, \ldots, \omega_{g_q})$. Finally, the set $\boldsymbol{\Omega}$ consists of matrices $\{\Omega_1, \Omega_2, \ldots, \Omega_G\}$.

Step (h) Given $\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{b}$, the conditional distribution of L_i is obtained by

$$L_i|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{b} \overset{\text{i.i.d}}{\sim} \text{Multinomial}(\pi_{ig}^*, g = 1, 2, \ldots, G), \tag{A10}$$

the value of π_{ig}^* is directly proportional to $\pi_g p(\boldsymbol{b}_i|\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g)$, where $\boldsymbol{b}_i|\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g \sim N_q(\boldsymbol{\mu}_g, \boldsymbol{\Omega}_g)$. The values of π_g ($g = 1, 2, \ldots, G$) are randomly selected from step (e). When given L_i, $\boldsymbol{\mu}$, and $\boldsymbol{\Omega}$, the prior distribution of $\boldsymbol{b}_i$ follows a normal distribution $N_q(\boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i})$, where $\boldsymbol{\mu}_{L_i}$ and $\boldsymbol{\Omega}_{L_i}$ represent the L_i elements of the sets $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$, respectively.

Step (i) The conditional distribution $p(\boldsymbol{b}_i|\boldsymbol{\beta}, \boldsymbol{\varphi}, \sigma^2, \gamma, \boldsymbol{\phi}, \boldsymbol{y}^*, \boldsymbol{T}, \boldsymbol{\Delta})$ cannot be directly derived using Gibbs sampling for $i = 1, 2, \ldots, n$ as it is non-standard. Specifically, it can be expressed as follows:

$$p(\boldsymbol{b}_i|\boldsymbol{\beta}, \boldsymbol{\varphi}, \sigma^2, \gamma, \boldsymbol{\phi}, \boldsymbol{y}^*, \boldsymbol{T}, \boldsymbol{\Delta}) \propto p(\boldsymbol{b}_i|\boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i}) p\left(\boldsymbol{y}_i^*|\boldsymbol{b}_i; \boldsymbol{\theta}_y\right) p(\boldsymbol{T}, \boldsymbol{\Delta}|\boldsymbol{b}; \boldsymbol{\theta}_T). \tag{A11}$$

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

$$\min\left\{1, \frac{p(\boldsymbol{b}_i|\boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i}) p\left(\boldsymbol{y}_i^*|\boldsymbol{b}_i; \boldsymbol{\theta}_y\right) p(\boldsymbol{T}, \boldsymbol{\Delta}|\boldsymbol{b}; \boldsymbol{\theta}_T)}{p(\boldsymbol{b}_i^{(m)}|\boldsymbol{\mu}_{L_i}, \boldsymbol{\Omega}_{L_i}) p\left(\boldsymbol{y}_i^*|\boldsymbol{b}_i^{(m)}; \boldsymbol{\theta}_y\right) p\left(\boldsymbol{T}, \boldsymbol{\Delta}|\boldsymbol{b}_i^{(m)}, \boldsymbol{b}_{-i}; \boldsymbol{\theta}_T\right)}\right\}, \tag{A12}$$

The remaining random effects, denoted as $\boldsymbol{b}_{-i}$, represent the random effects of all individuals except the ith individual. The value of the variance σ_b^2 can be adjusted to ensure that the average acceptance rate is about 0.25 or higher.

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. [CrossRef]
2. Wulfsohn, M.S.; Tsiatis, A.A. A Joint model for survival and longitudinal data measured with error. *Biometrics* **1997**, *53*, 330–339. [CrossRef] [PubMed]
3. Schluchter, M.D. Methods for the analysis of informatively censored longitudinal data. *Stat. Med.* **1992**, *11*, 1861–1870. [CrossRef] [PubMed]
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. [CrossRef]
5. Tsiatis, A.A.; Davidian, M. Joint modeling of longitudinal and time-to-event data: An overview. *Stat. Sin.* **2004**, *14*, 809–834. [CrossRef]
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. [CrossRef]
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. [CrossRef] [PubMed]
8. Lawrence Gould, A.; Boye, M.E.; Crowther, M.J.; Ibrahim, J.G.; Quartey, G.; Micallef, S.; Bois, F.Y. Joint modeling of survival and longitudinal non-survival data: Current methods and issues. *Stat. Med.* **2015**, *34*, 2181–2195. [CrossRef]
9. Henderson, R.; Diggle, P.; Dobson, A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* **2000**, *1*, 465–480. [CrossRef]
10. Zeng, D.; Cai, J. Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures. *Lifetime Data Anal.* **2005**, *11*, 151–174. [CrossRef]
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. [CrossRef]
12. Song P.X.; Tan, M. Marginal models for longitudinal continuous proportional data. *Biometrics* **2000**, *56*, 496–502. [CrossRef] [PubMed]
13. Lesaffre, E.; Rizopoulos, D.; Tsonaka, R. The Logistic Transform for Bounded Outcome Scores. *Biostatistics* **2007**, *8*, 72–85. [CrossRef]
14. Barndorff-Nielsen, O.E.; Jørgensen, B. Some Parametric Models on the Simplex. *Multivar. Anal.* **1991**, *39*, 106–116. [CrossRef]
15. Qiu, Z.; Song, P.X.K. Simplex Mixed-Effects Models for Longitudinal Proportional Data. *Scand. J. Stat.* **2008**, *35*, 577–596. [CrossRef]
16. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B.* **1996**, *58*, 267–288. [CrossRef]
17. Park T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [CrossRef]
18. Hans, C. Bayesian lasso regression. *Biometrika* **2009**, *96*, 835–845. [CrossRef]

19. Guo, R.; Zhu, H.; Chow, S.M.; Ibrahim, J.G. Bayesian Lasso for semiparametric sructural equation models. *Biometrics* **2012**, *68*, 567–577. [CrossRef]
20. 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. [CrossRef]
21. 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. [CrossRef]
22. Sethuraman, J. A Constructive Definition of Dirichlet Priors. *Stat. Sin.* **1994**, *4*, 639–650.
23. Yang, M.; Dunson, D.B.; Baird, D. Semiparametric bayes hierarchical models with mean and variance constraints. *Comput. Stat. Data. Anal.* **2010**, *54*, 2172–2186. [CrossRef] [PubMed]
24. Ishwaran, H.; Zarepour, M. Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. *Biometrika* **2000**, *87*, 371–390. [CrossRef]
25. 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. [CrossRef] [PubMed]
26. 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 Anal.* **2017**, *23*, 183–206. [CrossRef] [PubMed]
27. Levine, M.N.; Bramwell, V.H.; Pritchard, K.I.; Norris, B.D.; Shepherd, L.E.; Abu-Zahra, H.; Findlay, B.; Warr, D.; Bowman, D.; Myles, J.; 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. Clin. Oncol.* **1998**, *16*, 2651–2658. [CrossRef]