Statistics and Its Interface Volume 13 (2020) 49–63

Bayesian variable selection and estimation in joint confirmatory factor analysis–Cox model

Chenyi Liang and Jingheng Cai*

In this article, we propose the joint confirmatory factor analysis–Cox model to assess the effects of observed and latent risk factors on survival time. The Bayesian adaptive Lasso procedure is developed to simultaneously conduct estimation and variable selection for the proposed model. Nice features including the empirical performance of the proposed method are demonstrated by simulation studies. The proposed method is applied to analyze the bladder cancer data set obtained from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute.

AMS 2000 subject classifications: 62F15, 62N01.
Keywords and phrases: Bayesian adaptive Lasso, Proportional hazards model, Latent variables, Bladder cancer.

1. INTRODUCTION

Cox model [6] is the most famous model which reveals the interrelationships between the risk factors and the hazard rates of events concerned. Conventional Cox models are only able to deal with observed risk factors. Aside from such type of observable risk factors, latent risk factors that are not directly assessable through a single observed variable have been increasingly recognized in medical research when a medical trait is characterized by multiple observed indicators from different angles. Recently, the joint modelling approach which consists of a confirmatory factor analysis (CFA) model and the Cox model was developed to assess the effects of both observed and latent risk factors [4, 24]. More specifically, the CFA model is employed to characterize latent risk factors via related multiple observed variables, and the Cox model is used to assess the effects of the resulting latent risk factors.

Although the CFA–Cox model can help reduce the dimensionality of the risk factors efficiently, it is important that identifying prominent risk factors to obtain a more parsimonious model in statistical inference. In the Bayesian context, one of the most popular variable selection schemes is the use of model selection procedures based on model comparison statistics, such as Bayes factor [18], deviance information criterion [28], and so on. An alternative way is to employ a shrinkage method, such as Lasso [31], to perform estimation and variable selection simultaneously. Tibshirani [31] showed that the Lasso estimates for linear regression parameters via the l_1-penalized least-squares criterion can be interpreted as Bayesian posterior mode estimates when the regression parameters have independent Laplace priors. Motivated by this idea, Park and Casella [26] proposed the Bayesian framework for Lasso by imposing the double exponential prior on the regression coefficients and the gamma prior on the shrinkage parameter. Since this seminal work, a host of studies have applied the Bayesian Lasso (BLasso) approach to various models. In the Bayesian survival analysis literature, Lee et al. [19] considered a penalized semiparametric Bayesian Cox model to perform model selection by assigning shrinkage priors for the regression coefficients in the Cox model. Gu et al. [12] proposed a Bayesian two-step Lasso procedure for biomarker selection under the Cox model. However, the above proposed methods were only able to accommodate observed risk factors. In the field of latent variable modelling, Guo et al. [13] developed BLasso method for model selection in semiparametric structural equation models. Feng et al. [8] proposed BLasso method to identify the structure of semiparametric structural equation models. Despite Lasso and/or BLasso work efficiently in model selection without tedious pairwise comparisons, they have their limitations, including inconsistency in certain conditions and suffering from appreciable bias [7, 32]. To attack this problem, an adaptive Lasso [33] and Bayesian adaptive Lasso (BaLasso) procedure [1, 21] have been developed. More recently, Feng et al. [9] proposed the BaLasso procedure to conduct estimation and model selection for ordinal regression with latent variables. The aforementioned studies focused on latent variable models or survival models without latent variables separately. Thus, to the best of our knowledge, this is the first study to conduct variable selection in the CFA–Cox model via the BLasso and/or BaLasso approach.

This work is motivated by a study on mortality and its risk factors for bladder cancer patients. Bladder cancer is the fourth most common cancer and ranks eighth as a cause of death from cancer among men in the United States, with an estimation that there will be 81,190 new cases and 17,240 patients would die of bladder cancer in 2018. Moreover, for patients who suffer from bladder cancer, the 5-year survival rate is 70% for cases detected when the disease is still lo-

*Corresponding author.

calized, 35% for regional disease, and 5% for distant-stage disease.

The bladder cancer data set used in this study is obtained from Surveillance, Epidemiology, and End Results (SEER) Program [23] of the National Cancer Institute. The SEER Program is an authoritative source of information on cancer incidence and survival in the United States. SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 28 percent of the U.S. population. The SEER Program registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, the first course of treatment, and follow-up for vital status. The SEER Program is the only comprehensive source of population-based information in the United States that includes the stage of cancer at the time of diagnosis and patient survival data. There is a total of 10,050,814 records in the SEER research data for all types of cancers. The bladder cancer data set used in this study is a slice of the SEER research data, which records information of 2,095 bladder cancer patients in Connecticut diagnosed between 2010 and 2014.

There are 15 potential risk factors under consideration, including T Extension, CS Extension, N Lymph Nodes, CS Lymph Nodes, CS Met at Dx, marriage (0 = married, 1 = single), gender (0 = female, 1 = male), age at diagnosis, grade, nodes examined, tumor size, nodes removed, race (0 = non-white, 1 = white), origin (0 = non-Spanish, 1 = Spanish), and the total number of benign tumors. The descriptive statistics of the risk factors are presented in Table 1. Both T Extension and CS Extension describe the farthest documented extension of tumor away from the primary site. Moreover, N Lymph Nodes and CS Lymph Nodes simultaneously identify the lymph node chain that involved by the tumor at the time of diagnosis, while CS Met at Dx represents the distant site(s) of metastatic involvement at the time of diagnosis. Based on the medical meanings of the above risk factors, CS Extension and T Extension are grouped into a latent variable "extension (EXT)," whilst CS Lymph Nodes, CS Met at Dx, and N Lymph Nodes are grouped into the other latent variable "metastasis (META)." Given that the risk factors include both observed and latent factors, the CFA–Cox model is tailor-made for analyzing such data set. Moreover, the BaLasso procedure is employed to simultaneously identify the important observed and latent risk factors of bladder cancer and assess their effects on the survival time.

The remainder of this article is organized as follows. Section 2 introduces the joint modelling approach for CFA–Cox model. Section 3 proposes the variable selection procedure via BaLasso method. Section 4 presents the details of posterior inference, including the Gibbs sampler [11] and the Metropolis-Hastings algorithm [14, 22]. Section 5 applies the proposed model and methodology to analyze the bladder cancer data set. Section 6 conducts simulation studies to assess the empirical performance of the proposed method. Section 7 concludes the paper with a discussion.

2. JOINT MODELLING APPROACH FOR CFA–COX MODEL

Let $t_i > 0$ be the survival time associated with subject $i = 1, \ldots, n$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ir})^T$ and $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iq})^T$ are the vectors of fixed covariates and latent variables, respectively. Inspired by the proportional hazards model, the hazard rate at time t_i is related to the fixed covariates and latent variables via the following model,

$$h(t_i|\mathbf{x}_i, \boldsymbol{\xi}_i) = h_0(t_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\xi}_i^T \boldsymbol{\gamma}), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$ are vectors of regression coefficients, and $h_0(t)$ is an unspecified arbitrary baseline hazard function. Under the model defined by (1), the survival function of t_i is given by

$$S(t_i|\mathbf{x}_i, \boldsymbol{\xi}_i, H_0) = \exp\big\{ - \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\xi}_i^T \boldsymbol{\gamma}) H_0(t_i) \big\}, \tag{2}$$

where $H_0(t)$ is the cumulative baseline hazard function with $H_0(0) = 0$. One may assign the Dirichlet process to model the prior for the cumulative baseline hazard function. However, the Dirichlet process lacks a simple interpretation in terms of hazard functions. On the other hand, the gamma process is the most commonly used nonparametric prior process for modelling the cumulative baseline hazard function [5, 17]. Thus, we assign the gamma process prior for H_0, as follows:

$$H_0 \sim \mathcal{GP}(w_0 H^*, w_0), \tag{3}$$

where $\mathcal{GP}(\cdot)$ denotes the gamma process prior and $H^*(t)$ is an increasing function with $H^*(0) = 0$. Note that H^* is the mean of the process and assumed to be a known parametric function, while w_0 is a weight or confidence parameter about H^*. For example, if H^* relates to the Weibull distribution, then $H^*(t) = \zeta t^\eta$, where (ζ, η) is a specified vector of hyperparameters. In general, the statistical inference is robust to the choice of H^*. In this study, we assume that H^* associates with the exponential distribution for mathematical convenience and computational efficiency. That is, $H^*(t) = \nu t$, where ν is a hyperparameter with a given value.

With the gamma process prior, the likelihood can be only obtained for the cases when all the observed survival times are distinct [5, 17]. As grouped survival data is often encountered in practice, we will construct our model based on the grouped data likelihood [3, 16]. We first construct a finite partition of the time axis, $0 = a_0 < a_1 < a_2 < \ldots < a_K$, with $a_K > t_i$ for all $i = 1, \ldots, n$. Let $A_k = (a_{k-1}, a_k]$, we have the K disjoint intervals $A_1, \ldots, A_K$. The survival time t_i may fall in one and only one of those K disjoint intervals.

Table 1. Summary statistics of the bladder cancer data set

Variable	Descriptive statistics		Variable	Descriptive statistics	
Marriage	Married	1892	Gender	Male	1542
	Single	203		Female	553
Age at Diagnosis	Mean	71.427	Grade	Mean	2.754
	SD	11.801		SD	1.031
Nodes Examined	Mean	1.757	Tumor Size	Mean	35.646
	SD	6.378		SD	40.913
Nodes Removed	Mean	0.559	Race	White	2018
	SD	1.565		Non-white	77
Origin	Spanish	79	Benign Tumor	Mean	0.003
	Non-Spanish	2016		SD	0.058
T Extension	Mean	83.590	CS Extension	Mean	113.786
	SD	104.667		SD	146.254
N Lymph Nodes	Mean	6.874	CS Lymph Nodes	Mean	9.690
	SD	35.244		SD	51.683
CS Met at Dx	Mean	0.926			
	SD	6.559			

Moreover, let b_k be the increment in the cumulative baseline hazard in A_k, that is

$$b_k = H_0(a_k) - H_0(a_{k-1}), \quad k = 1, 2, \ldots, K.$$

The gamma process prior in (3) implies that b_k's are independently distributed as

$$b_k \sim \text{Gamma}\big(w_0 H^*(a_k) - w_0 H^*(a_{k-1}),\ w_0\big). \tag{4}$$

Thus, the conditional probability of $t_i \in A_k$ given $\mathbf{x}_i$, $\boldsymbol{\xi}_i$, and $\mathbf{b} = (b_1, \ldots, b_K)^T$ can be specified as follows:

$$\Pr(t_i \in A_k | \mathbf{x}_i, \boldsymbol{\xi}_i, \mathbf{b}) = \exp\Big\{ -\exp(\mathbf{x}_i^T\boldsymbol{\beta} + \boldsymbol{\xi}_i^T\boldsymbol{\gamma}) \sum_{j=1}^{k-1} b_j \Big\} \times \Big[1 - \exp\big\{ -b_k \exp(\mathbf{x}_i^T\boldsymbol{\beta} + \boldsymbol{\xi}_i^T\boldsymbol{\gamma})\big\}\Big]. \tag{5}$$

The grouped data likelihood function can be obtained as follows [3, 16]:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}) \propto \prod_{k=1}^{K} L_k, \tag{6}$$

where

$$L_k = \exp\Big\{ -b_k \sum_{i \in \mathcal{R}_k - \mathcal{D}_k} \exp(\mathbf{x}_i^T\boldsymbol{\beta} + \boldsymbol{\xi}_i^T\boldsymbol{\gamma}) \Big\} \times \prod_{j \in \mathcal{D}_k} \Big[1 - \exp\big\{ -b_k \exp(\mathbf{x}_j^T\boldsymbol{\beta} + \boldsymbol{\xi}_j^T\boldsymbol{\gamma})\big\}\Big],$$

with $\mathcal{R}_k$ and $\mathcal{D}_k$ being the risk set and failure set of A_k, respectively.

The second part of the joint modelling approach is to employ the CFA model for characterizing latent variables through multiple observed variables. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^T$ $(p > q)$ be the vector of observed variables related to latent variable $\boldsymbol{\xi}_i$, the CFA model is written as follows:

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i, \tag{7}$$

where $\boldsymbol{\Lambda}$ is the factor loading matrix, $\boldsymbol{\varepsilon}_i$ is the random error term which is independent of $\boldsymbol{\xi}_i$ and follows the normal distribution $N(\mathbf{0}, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \text{diag}(\psi_1, \ldots, \psi_p)$. Finally, $\boldsymbol{\xi}_i$ is assumed to be distributed as $N(\mathbf{0}, \boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ is the unknown covariance matrix. It is well known that the CFA model is not identifiable without imposing appropriate identifiability constraints. More specifically, for an arbitrary nonsingular matrix $\mathbf{M}$, the CFA model can be rewritten as $\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i = \boldsymbol{\Lambda}\mathbf{M}\mathbf{M}^{-1}\boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i = \boldsymbol{\Lambda}^*\boldsymbol{\xi}_i^* + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{M}$ and $\boldsymbol{\xi}_i^* = \mathbf{M}^{-1}\boldsymbol{\xi}_i$. Thus, appropriate constraints should be imposed on the factor loading matrix $\boldsymbol{\Lambda}$ or the covariance matrix $\boldsymbol{\Phi}$ so that the only nonsingular matrix $\mathbf{M}$ that meets the imposing conditions is the identity matrix. There are two common ways to identify the CFA model. The first method is to take the covariance matrix $\boldsymbol{\Phi}$ as an identity matrix. The factor loading matrix $\boldsymbol{\Lambda}$ can be arbitrary in this case. The second way is specifying a non-overlapping structure of the factor loading matrix $\boldsymbol{\Lambda}$ and fixing one of the elements in each column at a nonzero constant. Following a common practice in latent variable modelling literature [2, 20, 27], we specify a non-overlapping structure of $\boldsymbol{\Lambda}$ and fix one of the elements in each column at a nonzero constant to obtain an identified model and clear interpretations of the latent variables.

From the grouped data likelihood function (6) and the CFA model (7), the complete grouped data log-likelihood

with the constant terms being disregarded is as follows:

(8)
$$\begin{aligned} L_{\text{comp}} &= \sum_{k=1}^{K} \log L_k + \sum_{i=1}^{n} \Big[\log p(\mathbf{y}_i|\boldsymbol{\xi}_i) + \log p(\boldsymbol{\xi}_i) \Big] \\ &= \sum_{k=1}^{K} \log L_k - \frac{n}{2} \Big[\log |\boldsymbol{\Psi}| + \log |\boldsymbol{\Phi}| \Big] \\ &\quad - \frac{1}{2} \sum_{i=1}^{n} \Big[(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\xi}_i)^T \boldsymbol{\Psi}^{-1} (\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\xi}_i) + \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i \Big]. \end{aligned}$$

3. VARIABLE SELECTION VIA BAYESIAN ADAPTIVE LASSO

The BLasso procedure can be employed to identify the important risk factors in $\mathbf{x}_i$ and $\boldsymbol{\xi}_i$ by assigning a Laplace prior for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, as follows [8, 13, 26]:

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{j=1}^{r} \frac{\tau}{2} \exp\big(-\tau|\beta_j|\big) \times \prod_{j=1}^{q} \frac{\tau}{2} \exp\big(-\tau|\gamma_j|\big), \tag{9}$$

where τ is the tuning parameter which serves to control the impact of the shrinkage. One advantage of BLasso is that it provides a posterior sample that can be used to summarize the entire distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The posterior mean or mode of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be regarded as their Lasso estimators. It is shown that although BLasso does not shrink the Lasso estimators exactly to zero, it does shrink them close to 0 much faster than ridge regression does [26]. However, one problem with the BLasso procedure is that the same tuning parameter τ is applied to all regression coefficients, implying that the same impact of shrinkage is introduced to different regression coefficients. This may add potential bias to the resulting estimates [7, 32]. Following the existent literature [9, 29], we employ the BaLasso procedure to address this problem. More specifically, we assign a Laplace prior with coefficient-specific tuning parameters, as follows:

(10)
$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{j=1}^{r} \frac{\tau_j}{2} \exp\big(-\tau_j|\beta_j|\big) \times \prod_{j=1}^{q} \frac{\tau_{r+j}}{2} \exp\big(-\tau_{r+j}|\gamma_j|\big),$$

where τ_j, $j = 1, \ldots, r+q$, is the coefficient-specific tuning parameter. By introducing a specific penalty to each regression coefficient, BaLasso shrinks unimportant coefficients to 0 more efficiently and produces better estimation than BLasso does. Since the Laplace distribution can be represented as a scale mixture of normals with an exponential mixing density, the hierarchical representation of the BaLasso prior (10) can be rewritten as follows:

$$\begin{aligned} &\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\beta\kappa}), \quad \boldsymbol{\Sigma}_{\beta\kappa} = \text{diag}(\kappa_1^2, \ldots, \kappa_r^2), \\ &\boldsymbol{\gamma} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma\kappa}), \quad \boldsymbol{\Sigma}_{\gamma\kappa} = \text{diag}(\kappa_{r+1}^2, \ldots, \kappa_{r+q}^2), \\ &\kappa_j^2 \sim \text{Gamma}(1, \tau_j^2/2), \quad j = 1, \ldots, r+q. \end{aligned} \tag{11}$$

Finally, we assign gamma prior for the tuning parameter, τ_j, as follows:

$$\tau_j^2 \sim \text{Gamma}(c_{\tau j}, d_{\tau j}), \quad j = 1, \ldots, r+q, \tag{12}$$

where $c_{\tau j}$ and $d_{\tau j}$ are hyperparameters whose values are predetermined. Under the prior setting, the prior variances of κ_j^2 and τ_j^2 are $4/\tau_j^4$ and $c_{\tau j}/d_{\tau j}^2$, respectively.

4. POSTERIOR INFERENCE

In a full Bayesian analysis, the first step is to specify appropriate prior distributions for the unknown parameters. The BaLasso priors for the coefficients, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the related parameters were presented in (11) and (12). Moreover, according to the common practice in Bayesian latent variable modelling [27], the following prior distributions are assigned to the parameters related in the CFA model:

$$\begin{aligned} &\boldsymbol{\Lambda}_j|\psi_j \sim N(\boldsymbol{\Lambda}_{j0}, \psi_j \boldsymbol{\Sigma}_{\lambda j0}), \quad j = 1, \ldots, p, \\ &\psi_j^{-1} \sim \text{Gamma}(c_{\psi j}, d_{\psi j}), \quad j = 1, \ldots, p, \\ &\boldsymbol{\Phi}^{-1} \sim \text{Wishart}(\mathbf{R}_0, \rho_0), \end{aligned} \tag{13}$$

where $\boldsymbol{\Lambda}_j$ is the jth row of $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}_{j0}$, $\boldsymbol{\Sigma}_{\lambda j0}$, $c_{\psi j}$, $d_{\psi j}$, $\mathbf{R}_0$, and ρ_0 are hyperparameters with predetermined values. The prior variance of ϕ_{ij} in $\boldsymbol{\Phi}$ is

$$\frac{(\rho_0 - q + 1) r_{ij}^2 + (\rho_0 - q - 1) r_{ii} r_{jj}}{(\rho_0 - q)(\rho_0 - q - 1)^2 (\rho_0 - q - 3)},$$

where r_{ij} is the ith-row and jth-column element in $\mathbf{R}_0$. Moreover, the prior covariance of ϕ_{ij} and ϕ_{kl} is

$$\frac{2 r_{ij} r_{kl} + (\rho_0 - q - 1)(r_{ik} r_{jl} + r_{il} r_{kj})}{(\rho_0 - q)(\rho_0 - q - 1)^2 (\rho_0 - q - 3)}.$$

Let $\boldsymbol{\theta}$ be the vector containing all unknown parameters, $\boldsymbol{\mathcal{R}} = \{\mathcal{R}_1, \ldots, \mathcal{R}_K\}$, $\boldsymbol{\mathcal{D}} = \{\mathcal{D}_1, \ldots, \mathcal{D}_K\}$, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$, and $\mathbf{V} = \{\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{D}}, \mathbf{X}, \mathbf{Y}\}$, the Bayesian inference is based on the posterior distribution of $\boldsymbol{\theta}$ with the given observed data, $p(\boldsymbol{\theta}|\mathbf{V})$. However, owing to the existence of latent variables, this posterior distribution involves high dimensional integration and is intractable without closed form. Data augmentation [30] is employed to solve this difficulty. More specifically, the latent variables $\boldsymbol{\Xi} = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}$ are treated as hypothetical missing data and the observed data $\mathbf{V}$ are augmented with $\boldsymbol{\Xi}$ in the posterior analysis. As a result, the Bayesian analyses are based on the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\Xi}|\mathbf{V})$ rather than $p(\boldsymbol{\theta}|\mathbf{V})$. Given that $p(\boldsymbol{\theta}, \boldsymbol{\Xi}|\mathbf{V})$ does not involve integration, the subsequent computation is more straightforward. The Bayesian estimates and the standard error estimates can then be obtained through the sample means and the sample covariance matrix based on a sufficiently large number of random samples drawn from $p(\boldsymbol{\theta}, \boldsymbol{\Xi}|\mathbf{V})$. Although the joint posterior distribution does not involve high dimensional integration, drawing observations from $p(\boldsymbol{\theta}, \boldsymbol{\Xi}|\mathbf{V})$ directly is

difficult. Hence, the Gibbs sampler [11] is employed to draw a sequence of random observations from the joint posterior distribution. More specifically, the components in $\boldsymbol{\theta}$ are generated from the corresponding full conditional distributions iteratively. These full conditional distributions are provided in Appendix. Given that most full conditional distributions are non-standard, the Metropolis-Hastings algorithm [14, 22] is employed to generate samples from such non-standard distributions.

5. ANALYSIS OF BLADDER CANCER DATA SET

This section applied the proposed model with BaLasso procedure to analyze the bladder cancer data set that was introduced in Section 1. The main goal was to simultaneously identify the important observed and latent risk factors of bladder cancer and assess their associations with the hazard rate. As discussed in Section 1, the latent risk factor EXT was characterized by two observed variables, CS Extension and T Extension. The other latent risk factor META was related to three observed variables, CS Lymph Nodes, CS Met at Dx, and N Lymph Nodes. Given that no observed indicator measures more than one latent trait, the factor loading matrix $\boldsymbol{\Lambda}$ has a non-overlapping structure as follows:

$$\boldsymbol{\Lambda}^T = \begin{bmatrix} 1.0 & \lambda_{21} & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & \lambda_{42} & \lambda_{52} \end{bmatrix}, \tag{14}$$

where the ones and zeros were fixed to obtain an identified model.

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{i,11})^T = (1, \text{marriage, gender, age at diagnosis, grade, nodes examined, tumor size, nodes removed, race, origin, benign tumors})^T$, $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2})^T = (\text{EXT, META})^T$, the hazard rate in the proportional hazards model (1) is well defined with the aforementioned observed and latent risk factors. The prior inputs in (13) were assigned as follows: $\boldsymbol{\Lambda}_{j0}$ were zero vectors; $\boldsymbol{\Sigma}_{\lambda j0}$ and $\mathbf{R}_0$ were the identity matrices with appropriate dimensions; $\rho_0 = 6$, $c_{\psi j} = 1$, and $d_{\psi j} = 0.005$. Following the suggestions of the existing literatures [8, 29], we set $c_{\tau j} = 1$ and $d_{\tau j} = 0.01$ to obtain dispersed priors for τ_j. Such dispersed priors enable τ_j to be mainly determined by the data, thereby automatically imposing a large penalty on unimportant coefficients and shrinking them to 0 efficiently. There is not a general principle of choosing K. The minimum requirement is that the K disjoint intervals should be chosen so that at least one failure subject falls in each interval. In this study, we took $K = 10$ in the analysis which resulted in $(a_0, \ldots, a_{10})^T = (0, 4, 8, 12, 16, 22, 29, 36, 42, 50, 59)^T$. Moreover, w_0 in (3) represents a specification of weight or confidence attached to the initial guess H^*. Cai and Liang [4] pointed out that it is preferred to use a not very large w_0, for example, 10, to obtain reliable results under the situations that the data distribution is unknown. Finally, we set ν as its maximum likelihood estimate with $w_0 = \infty$ [3]:

$$\nu = \left(\log \left[\sum_{k=1}^{K} (n_{rk} + n_{dk}) \right] - \log \left[\sum_{k=1}^{K} n_{rk} \right] \right) / (a_1 - a_0), \tag{15}$$

where $n_{rk} = \sum_{i=1}^{n} \mathrm{I}(i \in \mathcal{R}_k - \mathcal{D}_k)$ and $n_{dk} = \sum_{i=1}^{n} \mathrm{I}(i \in \mathcal{D}_k)$ with $\mathrm{I}(\cdot)$ being an indicator function. To decide the number of burn-in iterations required for achieving convergence, we tried a few test runs with different starting values and obtained the "estimated potential scale reduction (EPSR)" statistic [10]. The EPSR values for all unknown parameters are presented in Figure 1. All EPSR values were less than 1.2 within 5,000 iterations, indicating the convergence of the algorithm. Based on this information and to be more conservative, 10,000 observations were collected after discarding the first 10,000 burn-in iterations for obtaining the Bayesian estimates. Using a single notebook computer with an Intel (R) Core (TM) i3-2350M CPU @2.30 GHz and 2.00 GB RAM, the computing time for the full Bayesian inference was about 2 hours. Our program was written in R codes which are provided on the journal website. Given that BaLasso is a sampling-based method which would not shrink the parameters exactly to 0, Hoti and Sillanpää [15] suggested setting the cutoff value c^* to identify insignificant variables. More specifically, if $|\hat{\beta}_j| \le c^*$ or $|\hat{\gamma}_j| \le c^*$ where $\hat{\beta}_j$ and $\hat{\gamma}_j$ are the means of their posterior distributions, then the corresponding observed or latent factor is classified as unimportant. The cutoff value c^* controls the size of the model, that is, more variables are identified as important with a smaller value of c^*. Following the common practice in BLasso and/or BaLasso procedure for latent variable modelling [8, 9], we set $c^* = 0.1$. The estimated factor loadings and the important coefficients along with their standard error estimates are presented in Figure 2, where the solid and dashed lines represent important and unimportant effects of risk factors, respectively. Finally, the estimates of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ in the CFA model are also presented with the arrows pointing to the corresponding observed and latent variables.

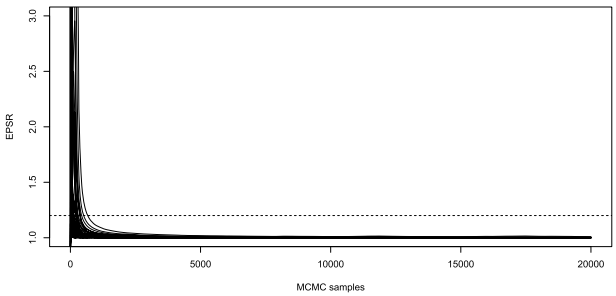

Figure 1. EPSR values in the bladder cancer study.

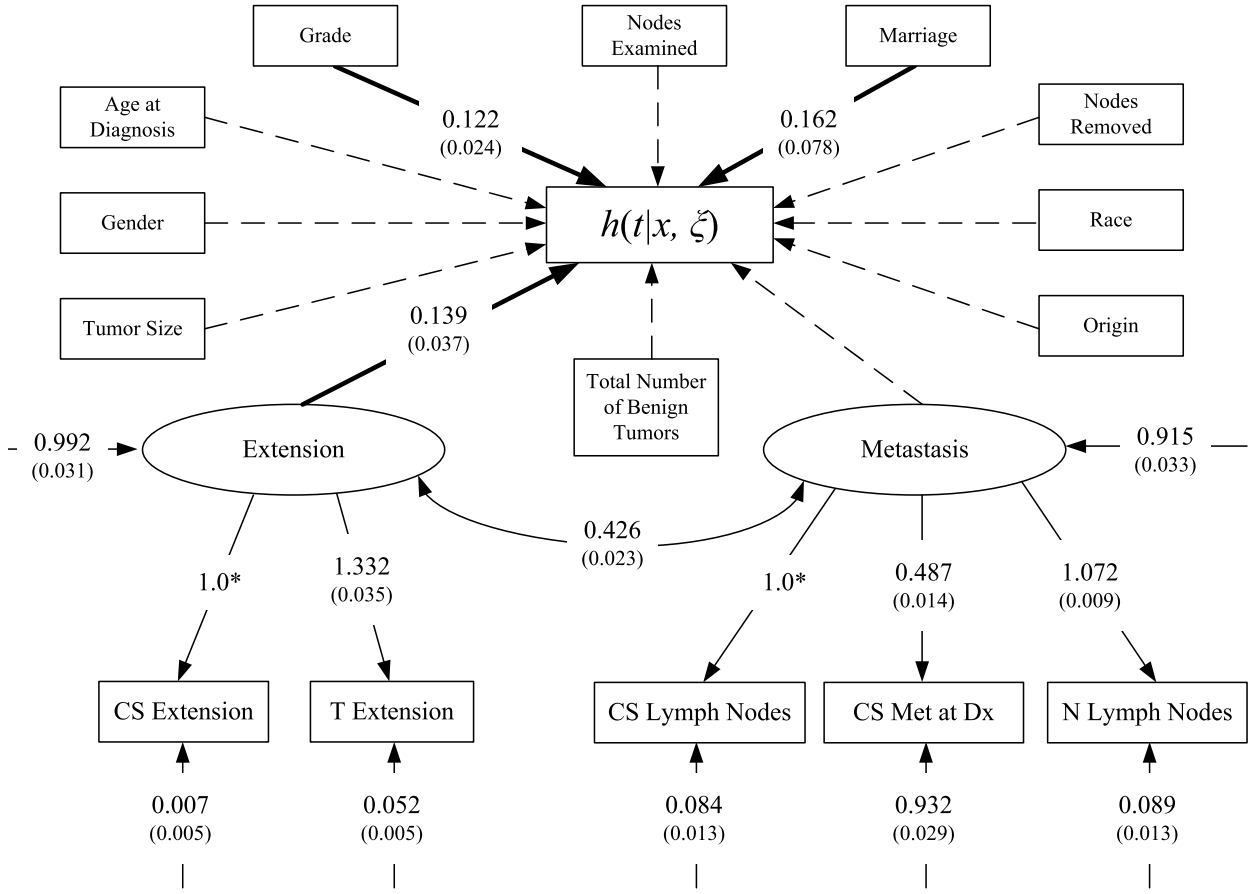

Figure 2. Path diagram in the bladder cancer study.

Several findings are obtained from the results. First, the prominent observed and latent risk factors associated with the hazard of bladder cancer are marriage, grade, and EXT. Second, both marriage and grade have positive effects on the hazard rate of mortality, implying that the single (marriage = 1) and poorer cell differentiated patients have a higher risk of death suffered from bladder cancer. Third, latent variable EXT has positive effects on the hazard rate of death, indicating that a more serious extension of bladder cancer leads to higher mortality. Fourth, other observed and latent risk factors, including gender, age at diagnosis, nodes examined, tumor size, nodes removed, race, origin, the total number of benign tumors, and META have little effects on the survival time in this study. The findings are useful for understanding the relationships between the mortality and the related observed and latent risk factors. Finally, we conducted analyses based on $K = 5$ and $K = 15$. We obtained similar results which indicated that the inferences were not sensitive to the choice of K.

For comparison, we reanalyzed the same data set based on a standard analysis, in which all the observed variables were treated as independent covariates. The estimated coefficients and the standard error estimates in both joint and standard analyses are presented in Table 2. The variables with an asterisk are the important ones identified by the BaLasso procedure. Despite the results are similar, the standard analysis is not able to address the following major question in the current study: How do the latent risk factors, EXT and META, affect the hazard rate of the bladder cancer overall?

We also conducted the analysis with H^* corresponding to the Weibull distribution, that is, $H^*(t) = \zeta t^\eta$ for the gamma process prior. We set $\zeta = \frac{1}{3}$ and $\eta = \frac{1}{2}$ to make H^* approximate to its maximum likelihood estimate. The estimated coefficients and the standard error estimates with $H^*(t) = \nu t$ and $H^*(t) = \zeta t^\eta$ are presented in Table 3. The variables with an asterisk are the important variables identified by the BaLasso procedure. The choice of H^* does not affect the statistical inference. Table 3 also shows how the size of the model varies according to the cutoff value c^*. If we took a smaller cutoff value, for example, $c^* = 0.05$, four extra variables including gender, age at diagnosis, race, and META were identified as important ones. On the other hand, if we took a larger cutoff value, $c^* = 0.15$, only one observed risk factor, marriage, was chosen.

Table 2. Comparison of the joint and standard analyses in the bladder cancer study

	Variable	Est	SE
Joint Analysis	Marriage*	0.162	0.078
	Gender	−0.098	0.053
	Age at Diagnosis	0.082	0.024
	Grade*	0.122	0.024
	Nodes Examined	−0.041	0.035
	Tumor Size	0.011	0.018
	Nodes Removed	−0.035	0.025
	Race	−0.066	0.086
	Origin	0.012	0.079
	Benign Tumors	0.015	0.160
	EXT*	0.139	0.037
	META	0.093	0.028
Standard Analysis	Marriage*	0.164	0.076
	Gender*	−0.109	0.051
	Age at Diagnosis	0.088	0.024
	Grade*	0.105	0.022
	Nodes Examined	−0.030	0.034
	Tumor Size	0.011	0.018
	Nodes Removed	−0.036	0.023
	Race	−0.078	0.087
	Origin	0.019	0.078
	Benign Tumors	0.025	0.165
	CS Extension	0.003	0.067
	T Extension*	0.141	0.074
	CS Lymph Nodes	0.016	0.039
	CS Met at Dx	0.095	0.024
	N Lymph Nodes	0.048	0.041

Note: The variables with an asterisk are the important variables identified by the BaLasso procedure with $c^* = 0.1$.

Table 3. Comparison of the analyses between $H^*(t) = \nu t$ and $H^*(t) = \zeta t^\eta$ in the bladder cancer study

	$H^*(t) = \nu t$		$H^*(t) = \zeta t^\eta$	
Variable	Est	SE	Est	SE
Marriage*	0.162	0.078	0.160	0.079
Gender	−0.098	0.053	−0.099	0.054
Age at Diagnosis	0.082	0.024	0.083	0.025
Grade*	0.122	0.024	0.121	0.026
Nodes Examined	−0.041	0.035	−0.045	0.036
Tumor Size	0.011	0.018	0.012	0.017
Nodes Removed	−0.035	0.025	−0.032	0.024
Race	−0.066	0.086	−0.050	0.091
Origin	0.012	0.079	0.012	0.080
Benign Tumors	0.015	0.160	0.024	0.161
EXT*	0.139	0.037	0.141	0.035
META	0.093	0.028	0.089	0.028

Note: The variables with an asterisk are the important variables identified by the BaLasso procedure with $c^* = 0.1$.

Finally, we conducted sensitivity analyses by perturbing the hyperparameters in prior inputs, the results of variable selection and Bayesian estimates were robust to such perturbations and not presented to save space.

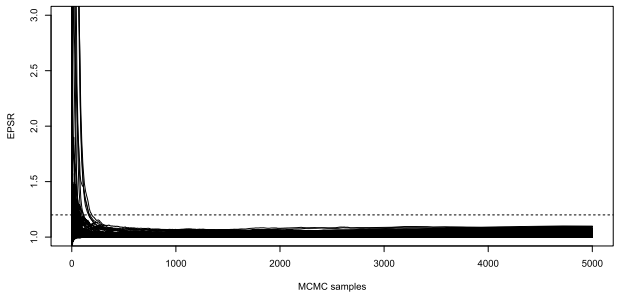

Figure 3. EPSR values in the simulation study.

6. SIMULATION STUDY

In this section, we conducted simulation studies to examine the performance of the proposed method. We set $p = 15$, $q = 5$ and a non-overlapping structure for

$$\mathbf{\Lambda} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.9 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.9 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.9 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.9 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.8 \end{bmatrix}$$

where the ones and zeros were fixed to obtain an identified model. $\mathbf{\Phi}$ was taken to be the correlation matrix with correlation coefficients being 0.5. Finally, $\mathbf{\Psi} = \text{diag}(0.3, \ldots, 0.3)$.

In the proportional hazards model defined by (1), a total of 21 covariates were set in order to assess the accuracy of the variable selection procedure, $\mathbf{x}_i = (x_{i1}, \ldots, x_{i,21})^T$ where $x_{i1} = 1$ and $x_{i2}, x_{i3}, \ldots, x_{i,21}$ were independently generated from $N(0, 1)$. Besides, we set $\beta_1 = 1$, $\beta_2 = \ldots = \beta_{11} = 0.5$, and $\beta_{12} = \ldots = \beta_{21} = 0$. Note that β_1 played a role as the intercept. The first eleven variables, $x_{i1}, x_{i2}, \ldots, x_{i,11}$ were prominent covariates with the corresponding regression coefficients being nonzero, while the last ten covariates did not affect the survival times. Moreover, we let $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)^T = (0.5, 0.5, 0.5, 0, 0)^T$, which indicates that the first three of the latent factors are prominent while the rest two are unimportant to the hazard rate. We considered two distributions to generate the survival time t_i, Exponential(ϑ_i) and Weibull$(2, \vartheta_i)$ where $\vartheta_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\xi}_i^T \boldsymbol{\gamma})$. The simulation was conducted based on 500 replications with sample sizes $n = 200$ and 500 with censoring rates (CR) being 20% and 40%.

Table 4. Number of correct specifications in the simulation study under exponential distribution

		$n = 200$				$n = 500$			
		CR = 20%		CR = 40%		CR = 20%		CR = 40%	
Par	Truth	BLasso	BaLasso	BLasso	BaLasso	BLasso	BaLasso	BLasso	BaLasso
β_1	1	500	500	500	500	500	500	500	500
β_2	0.5	500	500	500	495	500	500	500	500
β_3	0.5	500	500	500	495	500	500	500	500
β_4	0.5	500	500	495	500	500	500	500	500
β_5	0.5	500	500	500	500	500	500	500	500
β_6	0.5	500	500	500	495	500	500	500	500
β_7	0.5	500	500	500	495	500	500	500	500
β_8	0.5	500	500	500	500	500	500	500	500
β_9	0.5	500	495	500	471	500	500	500	500
β_{10}	0.5	500	500	500	500	500	500	500	500
β_{11}	0.5	500	500	495	500	500	500	500	500
β_{12}	0	406	495	409	457	481	495	454	475
β_{13}	0	401	442	389	476	495	490	475	500
β_{14}	0	415	457	409	471	476	486	479	496
β_{15}	0	401	457	394	466	495	490	454	492
β_{16}	0	406	462	370	438	486	495	454	479
β_{17}	0	377	471	385	442	486	495	450	483
β_{18}	0	382	466	361	457	495	490	454	500
β_{19}	0	434	476	389	476	500	495	433	488
β_{20}	0	406	447	409	481	481	490	471	479
β_{21}	0	420	466	361	471	495	500	475	496
γ_1	0.5	500	486	500	481	500	500	500	500
γ_2	0.5	500	486	500	486	500	500	500	500
γ_3	0.5	500	495	500	481	500	500	500	500
γ_4	0	354	423	317	394	428	462	442	458
γ_5	0	335	418	317	404	423	486	440	458

In the simulation study, the prior inputs in (13) were assigned as follows: $\mathbf{\Lambda}_{j0}$ were zero vectors; $\mathbf{\Sigma}_{\lambda j0}$ and $\mathbf{R}_0$ were the identity matrices with appropriate dimensions; $\rho_0 = 6$, $c_{\psi j} = 1$, $d_{\psi j} = 0.005$, $c_{\tau j} = 1$, and $d_{\tau j} = 0.01$. We chose $w_0 = 10$ and $K = 10$. Moreover, the value of ν was obtained by (15).

We considered both BLasso and BaLasso methods for comparison. The EPSR values for all unknown parameters are presented in Figure 3. All EPSR values were less than 1.2 within 1,000 iterations, indicating the convergence of the algorithm. We collected 3,000 posterior samples after a burn-in phase of 2,000 iterations for Bayesian inference. The cutoff value c^* was set as 0.1 to identify important variables. The number of correct specifications for all regression coefficients in the proportional hazards models based on 500 replications are shown in Tables 4 and 5. Given that the Bayesian estimates obtained from BLasso and BaLasso procedures are similar, we only present the summaries of the Bayesian estimates obtained from BaLasso procedure to save space. The averages of the bias (BIAS), the root mean square error (RMS), the averages of the posterior standard deviations (SD), and the coverage probabilities of 95% highest posterior density (HPD) intervals of the unknown parameters (CP) in the proportional hazards models are presented in Tables 6–9. Moreover, as the Bayesian estimates of unknown parameters in the CFA models are robust to the sample sizes and censoring rates, we only present the summaries of the unknown parameters in the CFA models under the cases with $n = 200$ and CR = 40% in Tables 10 and 11.

The main findings in the simulation study are as follows. First, Tables 4 and 5 show that both BLasso and BaLasso methods can identify the prominent variables in most replications. However, the number of correct specifications obtained by BaLasso are slightly larger than those obtained by BLasso, indicating BaLasso works better in variable selection than BLasso. This result is expected because BaLasso enables the data to determine the coefficient-specific penalty, which adaptively penalizes the coefficients of unimportant variables and shrinks them to 0 faster. Second, Tables 6–11 show that the proposed methodologies provide reliable results in the CFA–Cox model in the sense that the BIAS and RMS of the estimates are small except β_1. However, β_1 can be treated as the adjustment to the baseline hazard function $h_0(t)$. It is not the major parameter and it neither affects the statistical inference about other parameters. Third, owing to the shrinkage effects from BaLasso priors, the estimates of nonzero coefficients, $\beta_2, \ldots, \beta_{11}, \gamma_1, \gamma_2, \gamma_3$, are shrunk to zero, which leads to large negative biases. Consequently, the corresponding

Table 5. Number of correct specifications in the simulation study under Weibull distribution

		$n = 200$				$n = 500$			
		CR = 20%		CR = 40%		CR = 20%		CR = 40%	
Par	Truth	BLasso	BaLasso	BLasso	BaLasso	BLasso	BaLasso	BLasso	BaLasso
β_1	1	462	438	447	421	495	500	500	496
β_2	0.5	500	490	495	500	500	500	500	500
β_3	0.5	495	500	500	500	500	500	500	500
β_4	0.5	500	500	490	500	500	500	500	500
β_5	0.5	500	500	500	495	500	500	500	500
β_6	0.5	495	495	500	500	500	500	500	500
β_7	0.5	500	500	500	500	500	500	500	500
β_8	0.5	490	500	500	495	500	500	500	500
β_9	0.5	500	495	500	490	500	500	500	500
β_{10}	0.5	500	500	495	485	500	500	500	500
β_{11}	0.5	500	500	500	480	500	500	500	500
β_{12}	0	442	481	394	460	490	500	475	496
β_{13}	0	486	466	413	465	490	500	500	488
β_{14}	0	481	486	433	470	495	495	479	492
β_{15}	0	442	476	418	485	486	500	467	492
β_{16}	0	462	490	399	470	486	500	475	496
β_{17}	0	471	471	413	470	500	500	492	488
β_{18}	0	471	476	466	455	495	495	479	492
β_{19}	0	462	481	457	450	500	500	462	488
β_{20}	0	447	476	409	480	486	500	475	492
β_{21}	0	466	490	466	470	495	495	496	488
γ_1	0.5	495	495	495	480	500	500	500	500
γ_2	0.5	500	490	486	490	500	500	500	500
γ_3	0.5	500	490	495	475	500	500	500	500
γ_4	0	413	423	346	396	471	486	442	467
γ_5	0	365	466	341	431	443	476	433	446

CPs are not close to the nominal level. On the other hand, the biases of $\beta_{12}, \ldots, \beta_{21}, \gamma_4, \gamma_5$ are close to zero, which means that the RMS values are approximately equal to the standard deviations of the estimates. However, Tables 6–9 show that the SD values are slightly larger than the RMS values, indicating the posterior standard deviations are inflated. Therefore, the CPs are greater than the nominal level.

7. DISCUSSION

In this article, we considered BaLasso procedure to simultaneously conduct estimation and variable selection for the CFA–Cox model. The related MCMC algorithms were developed. Simulation studies showed that the proposed method was able to shrink the coefficients of unimportant variables towards 0 efficiently to identify the prominent observed and latent risk factors. An application to the bladder cancer data set was also presented.

There are several limitations in this study. First, we characterized latent variables through multiple observed variables via a CFA model. In many substantive studies, the number of latent variables and the structure of the factor loading matrix can be obtained by subject knowledge and the meanings of the observed variables. Moreover, exploratory factor analysis is useful in cross-validating the structure. More recently, penalty methods have been employed to obtain the structure of a matrix [25]. Such methods are potentially generalized to determine the number of latent variables and identify the factor loading matrix. Second, latent variables and random errors were assumed normally distributed in the CFA model. This normality assumption may be violated in practice. Employing more sophisticated techniques to relax the normality assumption of random errors and/or latent variables is of our further research interest. Third, the proposed method is performed based on right-censored data. It is worthy to generalize the proposed method to handle left-censored and/or interval-censored data. Fourth, the cutoff value c^* in this study was chosen based on the existent literature, it is essential to develop more rigorous methods to choose the appropriate cutoff value. Fifth, we did not provide any theories given that the Bayesian consistency in the latent variable modelling is an open question. It is an interesting topic in future studies. Finally, the coverage probabilities of conventional HPD intervals of the parameters in the proportional hazards models are not acceptable. It is necessary to develop adjusted credible intervals that have satisfactory coverage probabilities.

Table 6. Bayesian estimates of the parameters in the survival model in the simulation study under exponential distribution with $n = 200$

	CR = 20%				CR = 40%			
Par	BIAS	RMS	SD	CP (%)	BIAS	RMS	SD	CP (%)
β_1	−3.170	3.178	0.116	0.0	−2.940	2.950	0.154	0.0
β_2	−0.076	0.134	0.099	84.6	−0.119	0.168	0.114	79.8
β_3	−0.098	0.145	0.098	80.8	−0.156	0.193	0.114	68.3
β_4	−0.105	0.149	0.099	83.7	−0.123	0.167	0.113	79.8
β_5	−0.108	0.152	0.099	78.8	−0.134	0.181	0.115	76.9
β_6	−0.079	0.132	0.098	80.8	−0.143	0.179	0.116	78.8
β_7	−0.106	0.160	0.099	77.9	−0.128	0.178	0.115	76.9
β_8	−0.082	0.140	0.099	80.8	−0.128	0.172	0.115	77.9
β_9	−0.108	0.147	0.099	79.8	−0.149	0.196	0.113	76.0
β_{10}	−0.103	0.152	0.098	77.9	−0.122	0.169	0.114	75.0
β_{11}	−0.088	0.137	0.099	80.8	−0.139	0.180	0.114	77.9
β_{12}	0.002	0.048	0.071	99.0	0.001	0.054	0.079	100.0
β_{13}	0.000	0.061	0.074	100.0	−0.001	0.054	0.080	99.0
β_{14}	0.005	0.056	0.072	99.0	0.000	0.049	0.080	100.0
β_{15}	−0.001	0.059	0.073	99.0	0.000	0.056	0.080	100.0
β_{16}	−0.005	0.058	0.072	99.0	−0.004	0.062	0.081	100.0
β_{17}	−0.005	0.052	0.072	100.0	0.002	0.059	0.081	99.0
β_{18}	0.006	0.055	0.072	99.0	0.000	0.061	0.081	100.0
β_{19}	0.004	0.049	0.072	98.1	0.005	0.055	0.080	100.0
β_{20}	−0.002	0.056	0.073	99.0	−0.007	0.051	0.079	100.0
β_{21}	0.001	0.054	0.072	100.0	−0.002	0.054	0.080	100.0
γ_1	−0.095	0.179	0.142	87.5	−0.109	0.186	0.160	89.4
γ_2	−0.094	0.180	0.144	83.7	−0.107	0.203	0.161	82.7
γ_3	−0.081	0.169	0.146	89.4	−0.163	0.217	0.159	80.8
γ_4	0.036	0.073	0.096	100.0	0.041	0.083	0.106	100.0
γ_5	0.035	0.084	0.098	98.1	0.042	0.081	0.108	100.0

APPENDIX A. FULL CONDITIONAL DISTRIBUTIONS AND COMPUTATIONAL SCHEMES

(A1) Full conditional distribution of $\boldsymbol{\xi}_i$

The full conditional distribution of $\boldsymbol{\xi}_i$ is as follows:

(16)

$$
\begin{aligned}
p(\boldsymbol{\xi}_i|\cdot) \propto & \Big[\prod_{k=1}^{K} L_k^{(i)}\Big] \times p(\mathbf{y}_i|\boldsymbol{\xi}_i, \boldsymbol{\theta}) \times p(\boldsymbol{\xi}_i|\boldsymbol{\theta}) \\
\propto & \Big[\prod_{k=1}^{K} L_k^{(i)}\Big] \times \exp\Big\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\xi}_i)^T \boldsymbol{\Psi}^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\xi}_i) \\
& - \frac{1}{2}\boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1}\boldsymbol{\xi}_i\Big\},
\end{aligned}
$$

where $p(\boldsymbol{\xi}_i|\cdot)$ denotes full conditional distribution of $\boldsymbol{\xi}_i$ given all other quantities, and

$$
\begin{aligned}
L_k^{(i)} = \exp\Big\{ & - b_k \exp(\mathbf{x}_i^T\boldsymbol{\beta} + \boldsymbol{\xi}_i^T\boldsymbol{\gamma})\mathrm{I}(i \in \mathcal{R}_k - \mathcal{D}_k) + \\
& \log\Big[1 - \exp\big\{ - b_k \exp(\mathbf{x}_i^T\boldsymbol{\beta} + \boldsymbol{\xi}_i^T\boldsymbol{\gamma})\big\}\Big]\mathrm{I}(i \in \mathcal{D}_k)\Big\}.
\end{aligned}
$$

The distribution in (16) is non-standard and the Metropolis-Hastings (MH) algorithm [14, 22] is employed to sample from the non-standard distribution.

Given the current value $\boldsymbol{\xi}_i^{(l)}$, we simulate a new candidate $\boldsymbol{\xi}_i^*$ from proposal distribution $N(\boldsymbol{\xi}_i^{(l)}, \sigma_\xi^2 \boldsymbol{\Sigma}_{\xi i}^*)$, where

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\xi i}^{*-1} = & \boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} + \boldsymbol{\Phi}^{-1} + \sum_{k=1}^{K} b_k^{(i)}\mathrm{I}(i \in \mathcal{R}_k - \mathcal{D}_k)\boldsymbol{\gamma}\boldsymbol{\gamma}^T \\
& + \sum_{k=1}^{K} \frac{b_k^{(i)*}}{\big[1 - \exp\{-b_k^{(i)}\}\big]^2}\mathrm{I}(i \in \mathcal{D}_k)\boldsymbol{\gamma}\boldsymbol{\gamma}^T
\end{aligned}
$$

with $b_k^{(i)} = b_k \exp(\mathbf{x}_i^T\boldsymbol{\beta})$ and $b_k^{(i)*} = \big[b_k^{(i)2} - b_k^{(i)}\big]\exp\{-b_k^{(i)}\} \times \big[1 - \exp\{-b_k^{(i)}\}\big] + b_k^{(i)2}\exp\{-2b_k^{(i)}\}$. $\boldsymbol{\xi}_i^*$ is then accepted as new observation $\boldsymbol{\xi}_i^{(l+1)}$ with the following probability:

$$
\min\left\{1, \quad \frac{p(\boldsymbol{\xi}_i^*|\cdot)}{p(\boldsymbol{\xi}_i^{(l)}|\cdot)}\right\}.
$$

Tuning parameter σ_ξ^2 is selected such that the average acceptance rate is 0.25 or more.

Table 7. Bayesian estimates of the parameters in the survival model in the simulation study under exponential distribution with $n = 500$

	CR = 20%				CR = 40%			
Par	BIAS	RMS	SD	CP (%)	BIAS	RMS	SD	CP (%)
β_1	−3.423	3.427	0.080	0.0	−3.288	3.294	0.108	0.0
β_2	−0.084	0.099	0.057	69.5	−0.093	0.114	0.067	72.5
β_3	−0.082	0.104	0.057	68.6	−0.082	0.107	0.068	74.2
β_4	−0.096	0.111	0.057	61.0	−0.086	0.107	0.068	76.7
β_5	−0.092	0.110	0.057	62.9	−0.094	0.112	0.068	71.7
β_6	−0.079	0.100	0.057	65.7	−0.088	0.110	0.067	75.8
β_7	−0.082	0.097	0.058	71.4	−0.086	0.108	0.068	80.8
β_8	−0.075	0.100	0.057	70.5	−0.085	0.109	0.068	76.7
β_9	−0.087	0.103	0.058	65.7	−0.081	0.107	0.067	75.8
β_{10}	−0.094	0.107	0.057	62.9	−0.087	0.109	0.067	74.2
β_{11}	−0.076	0.097	0.057	70.5	−0.090	0.112	0.067	73.3
β_{12}	−0.002	0.035	0.046	100.0	−0.001	0.044	0.054	95.0
β_{13}	0.008	0.037	0.047	98.1	−0.004	0.036	0.054	100.0
β_{14}	−0.005	0.041	0.047	98.1	0.002	0.038	0.054	99.2
β_{15}	−0.002	0.036	0.047	99.0	0.005	0.040	0.054	100.0
β_{16}	−0.005	0.035	0.047	99.0	−0.001	0.048	0.055	96.7
β_{17}	0.008	0.032	0.046	99.0	0.001	0.042	0.055	98.3
β_{18}	0.003	0.038	0.047	99.0	−0.002	0.036	0.055	98.3
β_{19}	0.001	0.040	0.047	99.0	−0.005	0.038	0.054	98.3
β_{20}	−0.003	0.038	0.047	99.0	−0.002	0.042	0.054	95.8
β_{21}	0.002	0.033	0.046	100.0	−0.002	0.039	0.054	100.0
γ_1	−0.095	0.121	0.081	77.1	−0.100	0.134	0.095	84.2
γ_2	−0.084	0.108	0.082	87.6	−0.086	0.124	0.095	87.5
γ_3	−0.087	0.124	0.082	80.0	−0.087	0.132	0.095	90.0
γ_4	0.026	0.054	0.064	98.1	0.020	0.055	0.073	96.7
γ_5	0.022	0.049	0.063	100.0	0.031	0.061	0.075	98.3

(A2) Full conditional distribution of $(\boldsymbol{\Lambda}_j, \psi_j)$

Considering that $\boldsymbol{\Lambda}_j$ and ψ_j are only involved in the CFA model defined in (7), the full conditional distribution of $(\boldsymbol{\Lambda}_j, \psi_j)$ only depends on $\mathbf{Y}$ and $\boldsymbol{\Xi}$. Therefore, with the conjugate prior distributions in (13), the full conditional distribution can be easily obtained, as follows [27]: For $j = 1, \ldots, p$,

$$(\boldsymbol{\Lambda}_j|\cdot) \sim N(\boldsymbol{\Lambda}_j^*, \psi_j \boldsymbol{\Sigma}_{\lambda j}^*), \quad (\psi_j^{-1}|\cdot) \sim \text{Gamma}(c_{\psi j}^*, d_{\psi j}^*), \tag{17}$$

where $\boldsymbol{\Sigma}_{\lambda j}^{*-1} = \sum_{i=1}^n \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T + \boldsymbol{\Sigma}_{\lambda j0}^{-1}$, $\boldsymbol{\Lambda}_j^* = \boldsymbol{\Sigma}_{\lambda j}^* \big[\sum_{i=1}^n \boldsymbol{\xi}_i y_{ij} + \boldsymbol{\Sigma}_{\lambda j0}^{-1} \boldsymbol{\Lambda}_{j0}\big]$, $c_{\psi j}^* = c_{\psi j} + n/2$, $d_{\psi j}^* = d_{\psi j} + \frac{1}{2}\big[\sum_{i=1}^n y_{ij}^2 - \boldsymbol{\Lambda}_j^{*T} \boldsymbol{\Sigma}_{\lambda j}^{*-1} \boldsymbol{\Lambda}_j^* + \boldsymbol{\Lambda}_{j0}^T \boldsymbol{\Sigma}_{\lambda j0}^{-1} \boldsymbol{\Lambda}_{j0}\big]$.

(A3) Full conditional distribution of $\boldsymbol{\Phi}$

It can be easily derived that the full conditional distribution of $\boldsymbol{\Phi}$ only depends on $\boldsymbol{\Xi}$. More specifically,

$$(\boldsymbol{\Phi}^{-1}|\cdot) \sim \text{Wishart}(\mathbf{R}^*, \rho^*), \tag{18}$$

where $\mathbf{R}^* = (\mathbf{R}_0^{-1} + \sum_{i=1}^n \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T)^{-1}$, and $\rho^* = \rho_0 + n$.

(A4) Full conditional distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$

Let $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{r+q})^T = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ and $\mathbf{x}_{\xi i} = (\mathbf{x}_i^T, \boldsymbol{\xi}_i^T)^T$. Since $\boldsymbol{\varphi}$ is independent of other quantities with given $\mathbf{b}$ and $\boldsymbol{\Xi}$, the full conditional distribution is as follows:

$$\begin{aligned} p(\boldsymbol{\varphi}|\cdot) &= p(\boldsymbol{\varphi}|\mathbf{b}, \boldsymbol{\Xi}) \propto p(\mathbf{b}|\boldsymbol{\varphi}, \boldsymbol{\Xi}) \times p(\boldsymbol{\varphi}) \\ &= \prod_{k=1}^K L_k \times p(\boldsymbol{\beta}) \times p(\boldsymbol{\gamma}) \\ &\propto \prod_{k=1}^K \exp\Big\{ - b_k \sum_{i \in \mathcal{R}_k - \mathcal{D}_k} \exp(\mathbf{x}_{\xi i}^T \boldsymbol{\varphi}) \Big\} \times \\ &\quad \prod_{j \in \mathcal{D}_k} \Big[1 - \exp\big\{ - b_k \exp(\mathbf{x}_{\xi j}^T \boldsymbol{\varphi})\big\}\Big] \times \\ &\quad \exp\left\{ -\frac{1}{2} \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\beta\kappa}^{-1} \boldsymbol{\beta} - \frac{1}{2} \boldsymbol{\gamma}^T \boldsymbol{\Sigma}_{\gamma\kappa}^{-1} \boldsymbol{\gamma} \right\}. \end{aligned} \tag{19}$$

MH algorithm can be also employed to efficiently sample $\boldsymbol{\varphi}$. Given the current value $\boldsymbol{\varphi}^{(l)}$, we simulate a new candidate $\boldsymbol{\varphi}^*$ from proposal distribution $N(\boldsymbol{\varphi}^{(l)}, \sigma_\varphi^2 \boldsymbol{\Sigma}_\varphi^*)$, where

$$\begin{aligned} \boldsymbol{\Sigma}_\varphi^{*-1} = \sum_{k=1}^K \Big[& b_k \sum_{i \in \mathcal{R}_k - \mathcal{D}_k} \mathbf{x}_{\xi i} \mathbf{x}_{\xi i}^T \\ & + \sum_{j \in \mathcal{D}_k} \frac{b_k^*}{[1 - \exp\{-b_k\}]^2} \mathbf{x}_{\xi j} \mathbf{x}_{\xi j}^T \Big] \end{aligned}$$

Table 8. Bayesian estimates of the parameters in the survival model in the simulation study under Weibull distribution with $n = 200$

	CR = 20%				CR = 40%			
Par	BIAS	RMS	SD	CP (%)	BIAS	RMS	SD	CP (%)
β_1	−1.350	1.396	0.173	1.0	−1.479	1.524	0.208	1.0
β_2	−0.193	0.212	0.096	51.0	−0.166	0.201	0.112	65.3
β_3	−0.176	0.193	0.097	54.8	−0.155	0.195	0.112	70.3
β_4	−0.179	0.198	0.096	56.7	−0.175	0.201	0.112	67.3
β_5	−0.182	0.202	0.096	49.0	−0.185	0.211	0.110	63.4
β_6	−0.186	0.206	0.097	49.0	−0.161	0.191	0.111	72.3
β_7	−0.181	0.198	0.095	52.9	−0.174	0.201	0.112	66.3
β_8	−0.192	0.209	0.095	45.2	−0.192	0.212	0.112	60.4
β_9	−0.187	0.202	0.098	48.1	−0.183	0.214	0.112	64.4
β_{10}	−0.183	0.202	0.096	50.0	−0.175	0.208	0.111	65.3
β_{11}	−0.169	0.187	0.096	61.5	−0.184	0.219	0.111	58.4
β_{12}	−0.002	0.035	0.046	100.0	0.003	0.058	0.079	100.0
β_{13}	−0.006	0.049	0.067	100.0	−0.004	0.062	0.079	99.0
β_{14}	0.000	0.042	0.072	100.0	−0.004	0.055	0.079	100.0
β_{15}	0.004	0.050	0.070	100.0	−0.004	0.050	0.078	100.0
β_{16}	−0.006	0.048	0.069	99.0	−0.006	0.055	0.077	99.0
β_{17}	−0.009	0.048	0.069	100.0	0.000	0.047	0.076	100.0
β_{18}	0.000	0.049	0.071	100.0	−0.003	0.054	0.078	100.0
β_{19}	−0.006	0.045	0.069	100.0	−0.002	0.056	0.079	100.0
β_{20}	0.010	0.044	0.071	100.0	−0.002	0.047	0.078	100.0
β_{21}	0.003	0.043	0.070	100.0	−0.002	0.053	0.078	100.0
γ_1	−0.185	0.221	0.132	67.3	−0.147	0.212	0.154	82.2
γ_2	−0.180	0.213	0.130	71.2	−0.182	0.231	0.152	75.2
γ_3	−0.181	0.220	0.134	72.1	−0.189	0.236	0.152	70.3
γ_4	0.041	0.074	0.090	99.0	0.051	0.088	0.104	100.0
γ_5	0.033	0.058	0.089	100.0	0.044	0.073	0.105	100.0

$$+\begin{pmatrix}\boldsymbol{\Sigma}_{\beta\kappa}^{-1} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{\gamma\kappa}^{-1}\end{pmatrix},$$

with $b_k^* = (b_k^2 - b_k)\exp\{-b_k\}[1-\exp\{-b_k\}] + b_k^2\exp\{-2b_k\}$. $\boldsymbol{\varphi}^*$ is then accepted as new observation $\boldsymbol{\varphi}^{(l+1)}$ with the following probability:

$$\min\left\{1,\ \frac{p(\boldsymbol{\varphi}^*|\cdot)}{p(\boldsymbol{\varphi}^{(l)}|\cdot)}\right\}.$$

Tuning parameter σ_φ^2 is selected such that the average acceptance rate is 0.25 or more.

(A5) Full conditional distribution of b_k

The full conditional distribution of b_k only involves $\boldsymbol{\Xi}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, as follows:

$$\begin{aligned} p(b_k|\cdot) &= p(b_k)\times p(b_k|\boldsymbol{\Xi},\boldsymbol{\beta},\boldsymbol{\gamma})\\ &\propto \Big[b_k^{\alpha_{0k}-\alpha_{0,k-1}-1}\exp(-w_0 b_k)\Big]\times\\ &\quad \exp\Big\{-b_k\sum_{i\in\mathcal{R}_k-\mathcal{D}_k}\exp(\mathbf{x}_i^T\boldsymbol{\beta}+\boldsymbol{\xi}_i^T\boldsymbol{\gamma})\Big\}\times\\ &\quad \prod_{j\in\mathcal{D}_k}\Big[1-\exp\big\{-b_k\exp(\mathbf{x}_j^T\boldsymbol{\beta}+\boldsymbol{\xi}_j^T\boldsymbol{\gamma})\big\}\Big], \end{aligned} \tag{20}$$

where $\alpha_{0k} = w_0 H^*(a_k)$. Although the above distribution is non-standard, it can be well approximated by the following Gamma distribution [16]:

$$\begin{aligned} (b_k|\cdot) \sim \text{Gamma}\Big(&\alpha_{0k}-\alpha_{0,k-1}+n_{dk},\\ &w_0+\sum_{i\in\mathcal{R}_k-\mathcal{D}_k}\exp(\mathbf{x}_i^T\boldsymbol{\beta}+\boldsymbol{\xi}_i^T\boldsymbol{\gamma})\Big). \end{aligned} \tag{21}$$

(A6) Full conditional distribution of κ_j^2

Since κ_j^2 is only related to φ_j and τ_j^2, the full conditional distribution of κ_j^2 is given by:

$$p(\kappa_j^2|\cdot) = p(\varphi_j|\kappa_j^2)\times p(\kappa_j^2|\tau_j^2) \propto \kappa_j^{-1}\times\exp\left(-\frac{\varphi_j^2}{2\kappa_j^2}-\frac{\tau_j^2\kappa_j^2}{2}\right). \tag{22}$$

Such density turns out to be the following Inverse Gaussian distribution:

$$({\kappa_j^2}^{-1}|\cdot) \sim \text{Inverse Gaussian}\left(\frac{\tau_j}{|\varphi_j|},\tau_j^2\right). \tag{23}$$

Table 9. Bayesian estimates of the parameters in the survival model in the simulation study under Weibull distribution with $n = 500$

	CR = 20%				CR = 40%			
Par	BIAS	RMS	SD	CP (%)	BIAS	RMS	SD	CP (%)
β_1	−1.929	1.949	0.122	0.0	−2.134	2.165	0.149	0.0
β_2	−0.138	0.147	0.057	29.8	−0.115	0.131	0.067	57.5
β_3	−0.151	0.158	0.057	16.3	−0.120	0.133	0.067	56.7
β_4	−0.139	0.148	0.057	26.0	−0.116	0.132	0.066	60.8
β_5	−0.136	0.145	0.057	26.9	−0.110	0.130	0.066	57.5
β_6	−0.149	0.157	0.057	24.0	−0.103	0.121	0.066	63.3
β_7	−0.139	0.148	0.057	34.6	−0.115	0.128	0.066	59.2
β_8	−0.139	0.148	0.057	29.8	−0.110	0.127	0.067	63.3
β_9	−0.148	0.155	0.057	25.0	−0.115	0.136	0.067	55.8
β_{10}	−0.155	0.163	0.057	26.9	−0.123	0.136	0.067	55.0
β_{11}	−0.146	0.153	0.056	22.1	−0.115	0.131	0.067	60.8
β_{12}	0.004	0.032	0.046	100.0	0.003	0.037	0.052	100.0
β_{13}	−0.003	0.032	0.046	100.0	−0.003	0.040	0.053	99.2
β_{14}	−0.002	0.032	0.046	99.0	0.007	0.037	0.052	99.2
β_{15}	0.004	0.030	0.045	100.0	−0.001	0.039	0.053	100.0
β_{16}	−0.001	0.032	0.046	100.0	0.003	0.036	0.052	100.0
β_{17}	0.000	0.033	0.046	100.0	0.000	0.042	0.053	100.0
β_{18}	0.000	0.033	0.046	99.0	0.001	0.035	0.052	100.0
β_{19}	−0.004	0.032	0.046	100.0	0.001	0.042	0.052	99.2
β_{20}	0.000	0.031	0.046	100.0	−0.009	0.038	0.052	99.2
β_{21}	−0.004	0.031	0.046	99.0	−0.001	0.043	0.053	99.2
γ_1	−0.143	0.162	0.078	54.8	−0.115	0.138	0.093	80.0
γ_2	−0.160	0.177	0.078	48.1	−0.120	0.148	0.093	75.8
γ_3	−0.163	0.175	0.079	42.3	−0.133	0.158	0.094	69.2
γ_4	0.023	0.045	0.060	100.0	0.026	0.057	0.069	99.2
γ_5	0.024	0.049	0.061	100.0	0.029	0.060	0.070	98.3

Table 10. Bayesian estimates of the parameters in the CFA model in the simulation study under exponential distribution with $n = 200$, CR = 40%

Par	BIAS	RMS	SD	CP(%)	Par	BIAS	RMS	SD	CP(%)
λ_{21}	−0.002	0.054	0.059	96.2	ψ_{11}	0.005	0.042	0.043	99.0
λ_{31}	−0.012	0.055	0.059	94.2	ψ_{12}	0.005	0.042	0.044	96.2
λ_{52}	0.000	0.061	0.057	96.2	ψ_{13}	−0.010	0.055	0.054	92.3
λ_{62}	−0.003	0.061	0.057	95.2	ψ_{14}	0.013	0.048	0.043	93.3
λ_{83}	0.002	0.051	0.056	98.1	ψ_{15}	0.007	0.040	0.042	96.2
λ_{93}	0.003	0.052	0.057	97.1	ϕ_{11}	0.018	0.128	0.133	96.2
$\lambda_{11,4}$	−0.013	0.055	0.058	93.3	ϕ_{21}	0.009	0.086	0.091	95.2
$\lambda_{12,4}$	−0.011	0.055	0.058	94.2	ϕ_{22}	0.015	0.139	0.134	96.2
$\lambda_{14,5}$	−0.008	0.053	0.057	96.2	ϕ_{31}	0.010	0.084	0.092	96.2
$\lambda_{15,5}$	−0.001	0.053	0.057	95.2	ϕ_{32}	0.012	0.096	0.092	98.1
ψ_1	−0.004	0.051	0.050	94.2	ϕ_{33}	0.011	0.136	0.137	97.1
ψ_2	0.006	0.044	0.044	94.2	ϕ_{41}	0.008	0.084	0.092	95.2
ψ_3	0.017	0.044	0.044	97.1	ϕ_{42}	0.008	0.090	0.092	93.3
ψ_4	0.004	0.061	0.054	92.3	ϕ_{43}	0.010	0.085	0.093	97.1
ψ_5	0.002	0.049	0.042	89.4	ϕ_{44}	0.032	0.128	0.134	98.1
ψ_6	0.004	0.042	0.042	90.4	ϕ_{51}	0.022	0.092	0.092	96.2
ψ_7	−0.002	0.057	0.060	95.2	ϕ_{52}	0.007	0.092	0.092	97.1
ψ_8	−0.004	0.037	0.039	96.2	ϕ_{53}	0.004	0.080	0.092	96.2
ψ_9	0.002	0.040	0.040	95.2	ϕ_{54}	0.008	0.087	0.092	99.0
ψ_{10}	−0.004	0.047	0.050	94.2	ϕ_{55}	0.027	0.132	0.136	96.2

Table 11. Bayesian estimates of the parameters in the CFA model in the simulation study under Weibull distribution with $n = 200$, CR = 40%

Par	BIAS	RMS	SD	CP(%)	Par	BIAS	RMS	SD	CP(%)
λ_{21}	−0.005	0.064	0.058	90.1	ψ_{11}	0.006	0.046	0.044	95.0
λ_{31}	0.003	0.062	0.058	92.1	ψ_{12}	0.007	0.040	0.044	96.0
λ_{52}	−0.004	0.056	0.057	96.0	ψ_{13}	−0.004	0.061	0.054	87.1
λ_{62}	−0.001	0.058	0.057	93.1	ψ_{14}	0.008	0.045	0.042	96.0
λ_{83}	−0.004	0.050	0.056	99.0	ψ_{15}	0.000	0.044	0.042	94.1
λ_{93}	0.004	0.050	0.056	96.0	ϕ_{11}	0.016	0.130	0.132	95.0
$\lambda_{11,4}$	−0.004	0.062	0.059	92.1	ϕ_{21}	0.017	0.084	0.092	97.0
$\lambda_{12,4}$	−0.009	0.055	0.058	96.0	ϕ_{22}	0.013	0.118	0.134	97.0
$\lambda_{14,5}$	−0.008	0.058	0.057	92.1	ϕ_{31}	0.027	0.098	0.093	92.1
$\lambda_{15,5}$	−0.005	0.051	0.057	98.0	ϕ_{32}	0.027	0.098	0.093	98.0
ψ_1	−0.002	0.053	0.049	96.0	ϕ_{33}	0.035	0.140	0.139	96.0
ψ_2	0.002	0.042	0.043	93.1	ϕ_{41}	0.016	0.094	0.092	96.0
ψ_3	0.007	0.044	0.044	95.0	ϕ_{42}	0.019	0.099	0.092	95.0
ψ_4	−0.003	0.057	0.053	95.0	ϕ_{43}	0.021	0.108	0.094	95.0
ψ_5	0.002	0.043	0.041	94.1	ϕ_{44}	0.045	0.148	0.136	95.0
ψ_6	0.000	0.043	0.042	92.1	ϕ_{51}	0.030	0.094	0.093	94.1
ψ_7	−0.009	0.065	0.060	94.1	ϕ_{52}	0.026	0.087	0.093	96.0
ψ_8	0.003	0.038	0.040	94.1	ϕ_{53}	0.032	0.096	0.094	96.0
ψ_9	−0.003	0.042	0.040	93.1	ϕ_{54}	0.026	0.099	0.093	98.0
ψ_{10}	0.005	0.050	0.051	95.0	ϕ_{55}	0.033	0.142	0.135	96.0

(A7) Full conditional distributions of τ_j^2

The posterior distribution of τ_j^2 can be derived from:

(24)
$$p(\tau_j^2|\cdot) = p(\kappa_j^2|\tau_j^2) \times p(\tau_j^2) \propto (\tau_j^2)^{c_{\tau j}} \times \exp\left\{-(d_{\tau j} + \frac{\kappa_j^2}{2})\tau_j^2\right\},$$

which is the density of the following Gamma distribution:

(25)
$$(\tau_j^2|\cdot) \sim \text{Gamma}\Big(c_{\tau j} + 1, d_{\tau j} + \frac{\kappa_j^2}{2}\Big).$$

ACKNOWLEDGEMENTS

The authors thank the Editor, the Associate Editor, and three reviewers for their constructive comments that have improved the initial version of this paper.

Received 9 July 2018

REFERENCES

[1] Alhamzawi, R., Yu, K. and Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling: An International Journal* **12** 279–297. MR3179503

[2] Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley, New York.

[3] Burridge, J. (1981). Empirical Bayes analysis of survival time data. *Journal of the Royal Statistical Society, Series B* **43** 65–75. MR0610379

[4] Cai, J. H. and Liang, C. Y. (2019). Bayesian analysis of semiparametric Cox models with latent variables. *Statistical Methods in Medical Research* **28** 2150–2164. MR3977098

[5] Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47** 467–485.

[6] Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34** 187–220. MR0341758

[7] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* **96** 1348–1360. MR1946581

[8] Feng, X. N., Wang, G. C., Wang, Y. F. and Song, X. Y. (2015). Structure detection of semiparametric structural equation models with Bayesian adaptive group lasso. *Statistics in Medicine* **34** 1527–1547. MR3334674

[9] Feng, X. N., Wu, H. T. and Song, X. Y. (2017). Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods and Research* **46** 926–953. MR3715833

[10] Gelman, A. (1996). Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds.) pages 131–144. Chapman and Hall, London. MR1397966

[11] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.

[12] Gu, X., Yin, G. and Lee, J. J. (2013). Bayesian two-step Lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints. *Contemporary Clinical Trials* **36** 642–650.

[13] Guo, R. X., Zhu, H. T., Chow, S. M. and Ibrahim, J. G. (2012). Bayesian Lasso for semiparametric structural equation models. *Biometrics* **68** 567–577. MR2959624

[14] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437

[15] Hoti, F. and Sillanpää, M. J. (2006). Bayesian mapping of genotype × expression interactions in quantitative and qualitative traits. *Heredity* **97** 4–18.

[16] Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York. MR1876598

[17] Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B* **40** 214–221. MR0517442

[18] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90** 773–795. MR3363402
[19] Lee, H. K., Chakraborty, S. and Sun, J. (2011). Bayesian variable selection in semiparametric proportional hazards model for high dimensional survival data. *The International Journal of Biostatistics* **7** 1–32. MR2787411
[20] Lee, S. Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. Wiley, Chichester. MR2290938
[21] Leng, C. L., Tran, M. N. and Nott, D. (2013). Bayesian adaptive Lasso. *Annals of the Institute of Statistical Mathematics* **66** 221–244. MR3171404
[22] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machine. *The Journal of Chemical Physics* **21** 1087–1091.
[23] National Cancer Institute, DCCPS, Surveillance Research Program (2018). *Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973–2015)*. Released April 2018, based on the November 2017 submission.
[24] Pan, D., Kang, K., Wang, C. J. and Song, X. Y. (2019). Bayesian proportional hazards model with latent variables. *Statistical Methods in Medical Research* **28** 986–1002. MR3934631
[25] Pan, J. H., Ip, E. H. and Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian Lasso. *Psychological Methods* **22** 687–704.
[26] Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103** 681–686. MR2524001
[27] Song, X. Y. and Lee, S. Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. John Wiley & Sons, Ltd, Hoboken.
[28] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **64** 583–639. MR1979380
[29] Tang, A. M., Zhao, X. Q. and Tang, N. S. (2017). Bayesian variable selection and estimation in semiparametric joint models of multivariate longitudinal and survival data. *Biometrical Journal* **59** 57–78. MR3593721
[30] Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82** 528–550. MR0898357
[31] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58** 267–288. MR1379242
[32] Wang, H., Li, G. D. and Tsai, C. L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **69** 63–78. MR2301500
[33] Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* **101** 1418–1429. MR2279469

Chenyi Liang
Department of Statistics
Sun Yat-sen University
Guangzhou
China
E-mail address: liangchy5@mail2.sysu.edu.cn
Jingheng Cai
Department of Statistics
Sun Yat-sen University
Guangzhou
China
E-mail address: caijheng@mail.sysu.edu.cn