Bayesian Semi-Parametric Regression for Quantile Residual Lifetime

Taeyoung Park[1,a], Wonho Bae[a]

[a]Department of Applied Statistics, Yonsei University, Korea

Abstract

The quantile residual life function has been effectively used to interpret results from the analysis of the proportional hazards model for censored survival data; however, the quantile residual life function is not always estimable with currently available semi-parametric regression methods in the presence of heavy censoring. A parametric regression approach may circumvent the difficulty of heavy censoring, but parametric assumptions on a baseline hazard function can cause a potential bias. This article proposes a Bayesian semi-parametric regression approach for inference on an unknown baseline hazard function while adjusting for available covariates. We consider a model-based approach but the proposed method does not suffer from strong parametric assumptions, enjoying a closed-form specification of the parametric regression approach without sacrificing the flexibility of the semi-parametric regression approach. The proposed method is applied to simulated data and heavily censored survival data to estimate various quantile residual lifetimes and adjust for important prognostic factors.

Keywords: Bayesian nonparametrics, heavy censoring, median residual lifetime function, partial collapse, survival analysis.

1. Introduction

The proportional hazards model (Cox, 1972) is a powerful tool to analyze censored survival data, routinely used in diverse applications to identify risk factors or treatment effects when an outcome variable is time to event. Based on a hazard function, i.e., an instantaneous rate of occurrence of an event, the proportional hazards model is expressed as

$$h_i(t) = h_0(t) \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right), \tag{1.1}$$

where $h_i(t)$ is a hazard function for subject i at time t, $h_0(t)$ is an *unknown* baseline hazard function at time t, $\mathbf{x}_i$ is a $p \times 1$ vector of covariates for subject i, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients corresponding to $\mathbf{x}_i$. The baseline hazard function is left unspecified; however, the method of partial likelihood (Cox, 1972, 1975) allows to fit the proportional hazards model and estimate $\boldsymbol{\beta}$. The resulting estimates are not as efficient as the maximum likelihood estimates from a parametric regression approach when the baseline hazard function is correctly specified. Within a Bayesian parametric regression approach, several authors (*e.g.*, Carlin and Hodges, 1999) have proposed using an exponential or Weibull failure time distribution to model a baseline hazard function. Their method works well when the parametric assumptions are met, but estimations can be inaccurate otherwise. In the absence of covariates, a Bayesian nonparametric approach has been proposed to address the robustness issue

without sacrificing the efficiency of a Bayesian parametric approach (Kuo and Mallick, 1997; Kottas and Gelfand, 2001; Kottas, 2006; Park *et al.*, 2012a). The Bayesian nonparametric approach offers flexible modeling capabilities in a fully model-based framework. A Bayesian nonparametric method for survival analysis has two streams. The first stream develops models for log-transformed survival times under the accelerated failure time assumption (Kuo and Mallick, 1997; Kottas and Gelfand, 2001), and has been extended to Bayesian semiparametric quantile regression (Kottas *et al.*, 2007; Hjort and Petrone, 2007; Kottas and Krnjajic, 2009). However, the second stream directly models an unknown failure time distribution with a flexible Dirichlet process mixture (DPM). Following this stream, a Weibull DPM model has been recently proposed, mixing both the shape and scale parameters of the Weibull kernel (Kottas, 2006; Park *et al.*, 2012a). Kottas (2006) developed a Bayesian nonparametric method by marginalizing the mixing distribution of the Weibull DPM model, so that posterior inference on the non-linear functionals of the mixture distribution is limited. By contrast, Park *et al.* (2012a) employed a truncated stick-breaking representation (Ishwaran and James, 2001) to the Weibull DPM model, allowing direct posterior inference on the general functionals. This paper extends the work by Park *et al.* (2012a) to allow the estimation of covariate effects under the proportional hazards assumption; consequently, we consider the Bayesian semi-parametric regression method that retains proportional hazards but flexibly models an unknown baseline hazard function based on a DPM of Weibull distributions.

The interpretation of the proportional hazards model is typically based on the percentage of reduction in a hazard ratio, which is generally difficult to understand for lay audiences without solid statistical backgrounds. As an alternative summary measure of survival data to the hazard ratio, residual lifetime can then be used to better interpret the proportional hazards model (Jeong *et al.*, 2008; Park *et al.*, 2012a). Under the accelerated failure time assumption, residual lifetime has been studied using quantile regression (Gelfand and Kottas, 2003; Kim *et al.*, 2012) and the Bayesian nonparametric method (Kottas and Gelfand, 2001). This paper develops the Bayesian semi-parametric regression method to estimate the quantiles of residual lifetimes under the proportional hazards assumption. For efficient estimation, we consider the method of partial collapse proposed by van Dyk and Park (2008). The resulting partially collapsed Gibbs (PCG) sampler and its generalized variants are successfully used to efficiently fit highly structured complex models in many applications (Park *et al.*, 2008; Park and van Dyk, 2009; Park, 2011; Park *et al.*, 2012b).

This paper is organized into four additional sections. Section 2 proposes the Bayesian semi-parametric regression method for fitting the proportional hazards model with censored survival data and outlines the proposed PCG-type sampler for efficient posterior computation. In Section 3, a simulation study validates the proposed method under heavy censoring. Section 4 applies the proposed method to heavily censored survival data. Discussion follows in Section 5.

2. Bayesian Semi-Parametric Inference on Residual Lifetime

2.1. Bayesian semi-parametric regression model for censored survival data

Let t_i be a failure time of subject i with hazard function $h_i(t)$ in (1.1) and survival function

$$S_i(t) = S_0(t)^{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}$$
$$= \exp\left(-\mathbf{x}_i^\top \boldsymbol{\beta} \int_0^t h_0(u) du\right),$$

where $S_0(t)$ denotes a baseline survival function. Because of early termination of a trial or loss to follow-up, a failure time of subject i may not be fully observed but rather be censored at c_i that is

assumed to be independent of t_i. Then we observe $y_i = \min(t_i, c_i)$ for subject i, along with $d_i = I(t_i \leq c_i)$, where $I(\cdot)$ is an indicator function. The $100(q)^{th}$ residual life function for subject i at time t_0 is defined as

$$t^*_{i,q} = S^{-1}_{i,t_0}(1 - q),$$

where $S_{i,t_0}(t^*) = S_i(t^* + t_0)/S_i(t_0)$ is the residual survival function for $t^* > 0$ and $t_0 \geq 0$. The $100(q)^{th}$ residual life function implies that $100(q)\%$ of remaining lifetimes among subjects are below time $t^*_{i,q}$.

The semi-parametric regression method allows to flexibly model the hazard function by letting the baseline hazard function left unspecified, while adjusting for available covariates. The estimates of regression coefficients are, however, not fully efficient and, under heavy censoring, the $100(q)^{th}$ residual life function is not estimable unless the $100(q)^{th}$ percentile is reached. By contrast, the parametric regression method based on a parametric assumption on the baseline hazard function can produce fully efficient estimates for regression coefficients, but the parametric approach can suffer from the strong parametric assumption unless it is correctly specified. To flexibly model the baseline hazard function in a fully model-based framework, we propose the Bayesian semi-parametric regression method for the proportional hazards model. Let $h_0(t; \rho, \lambda) = \lambda \rho t^{\rho-1}$ and $S_0(t; \rho, \lambda) = \exp(-\lambda t^\rho)$ denote the Weibull baseline hazard and survival functions with shape parameter ρ and scale parameter λ, respectively. The proposed Weibull DPM model is expressed as

$$y_i | (\rho_i, \lambda_i, \boldsymbol{\beta}) \overset{ind}{\sim} \left\{ h_0(y_i; \rho_i, \lambda_i) \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right) \right\}^{d_i} S_0(y_i; \rho_i, \lambda_i)^{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}, \quad i = 1, \ldots, n,$$

$$(\rho_i, \lambda_i) | G \overset{iid}{\sim} G, \quad i = 1, \ldots, n,$$

$$G | (\gamma, \eta, \xi) \sim \mathrm{DP}\left(\gamma G_0(\eta, \xi)\right),$$

$$(\boldsymbol{\beta}, \gamma, \eta, \xi) \sim p(\boldsymbol{\beta})p(\gamma)p(\eta)p(\xi),$$

where the (ρ_i, λ_i)'s are independent and identically distributed according to an unknown Dirichlet process prior distribution G with base distribution G_0 on $\mathfrak{R}^+$ that has parameters η and ξ corresponding to ρ_i and λ_i, respectively. The finite stick-breaking presentation (Sethuraman, 1994) of G implies

$$G(\rho_i, \lambda_i) = \sum_{j=1}^{J} \pi_j I\left(\rho_i = \rho_j^*\right) I\left(\lambda_i = \lambda_j^*\right), \quad i = 1, \ldots, n,$$

$$\left(\rho_j^*, \lambda_j^*\right) | (\eta, \xi) \overset{iid}{\sim} G_0(\eta, \xi), \quad j = 1, \ldots, J,$$

where $\pi_j = V_j \prod_{l<j}(1 - V_l)$ is a probability weight assigned to (ρ_j^*, λ_j^*), $V_j | \gamma \overset{iid}{\sim}$ Beta$(1, \gamma)$, $j = 1, \ldots, J - 1$, and $V_J = 1$; see Ishwaran and James (2001) for theoretical arguments. For efficient and numerically stable computation, the base distribution G_0 is chosen to be conjugate with log-concavity property and a constraint is imposed on the shape parameter of the Weibull distribution such that $\rho_j^* > 1$ for $j = 1, \ldots, J$. Specifically, $G_0(\eta, \xi) = $ Trunc-Gamma$(1, \eta, 1)$Gamma$(1, \xi)$ with rate parameters η and ξ, where Trunc-Gamma(a, b, c) represents a truncated gamma distribution with shape parameter a, rate parameter b, and left truncation c, and Gamma(a, b) represents a gamma distribution with mean a/b. To facilitate the posterior computation of the Weibull DPM model, we introduce a latent variable Z_i such that $Z_i = j$ indicates subject i is allocated to cluster j. Then we can write $\rho_i = \rho_j^* = \rho_{Z_i}^*$ and $\lambda_i = \lambda_j^* = \lambda_{Z_i}^*$, and we have $\pi_j = P(Z_i = j)$.

2.2. Efficient posterior computation

For posterior computation, we impose prior distributions on parameters, $(\boldsymbol{\beta}, \gamma, \eta, \xi)$. As for the regression coefficients, the proper prior distribution is chosen as $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is positive definite. In a DPM model, γ controls the number of clusters. By assigning a conjugate gamma prior distribution, *i.e.*, $\gamma \sim \text{Gamma}(a_{\gamma}, b_{\gamma})$, the level of clustering can be a priori determined by changing the values of (a_{γ}, b_{γ}). As for the rate parameters of the base distribution G_0, we assign independent conjugate gamma distributions, *i.e.*, $(\eta, \xi) \sim \text{Gamma}(a_{\eta}, b_{\eta})\text{Gamma}(a_{\xi}, b_{\xi})$, where we choose the values of $(a_{\eta}, b_{\eta}, a_{\xi}, b_{\xi})$ to make the prior distribution of (η, ξ) diffuse.

The target posterior distribution is then given by $p(\mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \gamma, \eta, \xi | \mathbf{Y})$, where $\mathbf{Z} = \{Z_i\}_{i=1}^n$ denotes a collection of latent indicator variables, $\mathbf{V} = \{V_j\}_{j=1}^{J-1}$ denotes a set of random variables from a truncated stick-breaking process, $\boldsymbol{\rho}^* = \{\rho_j^*\}_{j=1}^J$ and $\boldsymbol{\lambda}^* = \{\lambda_j^*\}_{j=1}^J$ denote a set of atoms for the Weibull shape and scale parameters, respectively, and $\mathbf{Y} = \{(y_i, d_i)\}_{i=1}^n$ denotes the observed data. For efficient posterior computation, we consider a certain component to be partially collapsed out of the model (van Dyk and Park, 2008; Park and van Dyk, 2009). In particular, $\boldsymbol{\lambda}^*$ is partially collapsed out of the sampling steps for $\boldsymbol{\beta}$ and $\boldsymbol{\rho}^*$ for better convergence characteristics. The resulting PCG-type sampler based on the method of partial collapse iterates among the following steps. First, Z_i is drawn from its complete conditional distribution that is multinomial with probabilities

$$p\left(Z_i = j | \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \gamma, \eta, \xi, \mathbf{Y}\right) = \frac{V_j \prod_{l<j}(1-V_l) \left\{h_0\left(y_i; \rho_j^*, \lambda_j^*\right) \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right\}^{d_i} S_0\left(y_i; \rho_j^*, \lambda_j^*\right)^{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}}{\sum_{k=1}^J V_k \prod_{l<k}(1-V_l) \left\{h_0\left(y_i; \rho_k^*, \lambda_k^*\right) \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right\}^{d_i} S_0\left(y_i; \rho_k^*, \lambda_k^*\right)^{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}},$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Second, V_j is drawn from its complete conditional distribution, *i.e.*, a beta distribution,

$$V_j | (\mathbf{Z}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \gamma, \eta, \xi, \mathbf{Y}) \overset{ind}{\sim} \text{Beta}\left(1 + \sum_{i=1}^n I(Z_i = j), \gamma + \sum_{i=1}^n I(Z_i > j)\right),$$

for $j = 1, \ldots, J - 1$. Third, the reduced conditional distribution of $\boldsymbol{\beta}$ is computed after partially collapsing $\boldsymbol{\lambda}^*$ out of the model and is proportional to

$$\exp\left\{\left(\boldsymbol{\mu}_{\boldsymbol{\beta}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} + \sum_{i=1}^n d_i \mathbf{x}_i^\top\right)\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta} - \sum_{j=1}^J \left(1 + \sum_{i:Z_i=j} d_i\right)\log\left(\xi + \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right)\right\}.$$

Fourth, $(\boldsymbol{\rho}^*, \boldsymbol{\lambda}^*)$ are jointly drawn from their complete conditional distributions,

$$p\left(\boldsymbol{\rho}^*, \boldsymbol{\lambda}^* | \mathbf{Z}, \mathbf{V}, \boldsymbol{\beta}, \gamma, \eta, \xi, \mathbf{Y}\right) = \prod_{j=1}^J p\left(\rho_j^* | \mathbf{Z}, \mathbf{V}, \boldsymbol{\beta}, \gamma, \eta, \xi, \mathbf{Y}\right) p\left(\lambda_j^* | \mathbf{Z}, \mathbf{V}, \rho_j^*, \boldsymbol{\beta}, \gamma, \eta, \xi, \mathbf{Y}\right),$$

where the reduced conditional distribution of ρ_j^* is obtained by partially collapsing $\boldsymbol{\lambda}^*$ out of the model and is proportional to

$$\left(\rho_j^*\right)^{\sum_{i:Z_i=j} d_i} \exp\left\{-\eta\rho_j^* + \rho_j^* \sum_{i:Z_i=j} d_i \log y_i - \left(1 + \sum_{i:Z_i=j} d_i\right)\log\left(\xi + \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right)\right\},$$

Table 1: Summary statistics of coefficients under 30% and 50% censoring

Coefficient	True value	30% censoring		50% censoring	
		Mean	95% interval	Mean	95% interval
β_1	1	1.133	(0.930, 1.369)	0.969	(0.805, 1.132)
β_2	−1	−1.084	(−1.375, −0.800)	−0.808	(−1.129, −0.501)

for $\rho_j^* > 1$, and the complete conditional distribution of λ_j^* is gamma,

$$\lambda_j^* | \left(\mathbf{Z}, \mathbf{V}, \rho_j^*, \boldsymbol{\beta}, \gamma, \eta, \xi, \mathbf{Y} \right) \overset{ind}{\sim} \text{Gamma}\left(1 + \sum_{i:Z_i=j} d_i, \xi + \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta} \right) \right),$$

for $j = 1, \ldots, J$. Last, (γ, η, ξ) are jointly drawn from their complete conditional distributions,

$$p\left(\gamma, \eta, \xi | \mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \mathbf{Y} \right) = p\left(\gamma | \mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \mathbf{Y} \right) p\left(\eta | \mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \mathbf{Y} \right) p\left(\xi | \mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \mathbf{Y} \right),$$

where

$$\gamma | \left(\mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \mathbf{Y} \right) \sim \text{Gamma}\left(a_\gamma + J, b_\gamma - \sum_{j=1}^{J-1} \log(1 - V_j) \right),$$

$$\eta | \left(\mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \mathbf{Y} \right) \sim \text{Gamma}\left(a_\eta + J, b_\eta + \sum_{j=1}^{J} \rho_j^* - J \right),$$

$$\xi | \left(\mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}, \mathbf{Y} \right) \sim \text{Gamma}\left(a_\xi + J, b_\xi + \sum_{j=1}^{J} \rho_j^* \right).$$

As for the sampling steps that are not in closed form, the corresponding conditional distributions can be shown to be log-concave; refer to Appendix A. We thus employ adaptive rejection sampling (Gilks and Wild, 1992) to efficiently implement these sampling steps.

3. Simulation Study

3.1. Under heavy censoring

A simulation study is conducted to validate the proposed method under heavy censoring. Covariates $\mathbf{x}_i = (x_{i1}, x_{i2})^\top$ are generated as $x_{i1} \overset{iid}{\sim} \text{Bernoulli}(0.5)$ and $x_{i2} \overset{iid}{\sim} \text{Uniform}(0, 1)$ for $i = 1, \ldots, n$. The corresponding coefficients are set to $\boldsymbol{\beta} = (\beta_1, \beta_2) = (1, -1)^\top$. Under the baseline Weibull hazard function with shape parameter $\rho = 2$ and scale parameter $\lambda = 0.01$, we generate $n = 1000$ failure times from the proportional hazards model. Independent of the failure times, censoring times are generated from a uniform distribution between 1 and c^*, where c^* is chosen to adjust an expected censoring proportion to 30% and 50%. We use $(a_\gamma, b_\gamma, a_\eta, b_\eta, a_\xi, b_\xi) = (2, 0.1, 2, 0.1, 2, 0.1)$ for the prior distributions of (γ, η, ξ), and $\boldsymbol{\mu_\beta} = \mathbf{0}$ and $\boldsymbol{\Sigma_\beta} = \mathbf{I}$ are used for the prior distribution of $\boldsymbol{\beta}$. The stick-breaking process is truncated at $J = 100$.

Based on the test data, the proposed PCG-type sampler developed in Section 2 is run with two separate chains of 10000 iterations with different starting values. The convergence of the sampler is checked with the $\hat{R}^{1/2}$ statistic (Gelman and Rubin, 1992) being less than 1.2, and the 10000 draws collected from the second halves of the two chains are used for posterior inference. Table 1 shows the summary statistics of regression coefficients under heavy censoring of 30% and 50%. In Table 1, the true values of regression coefficients are closely estimated by posterior means and well covered by the corresponding 95% posterior intervals.
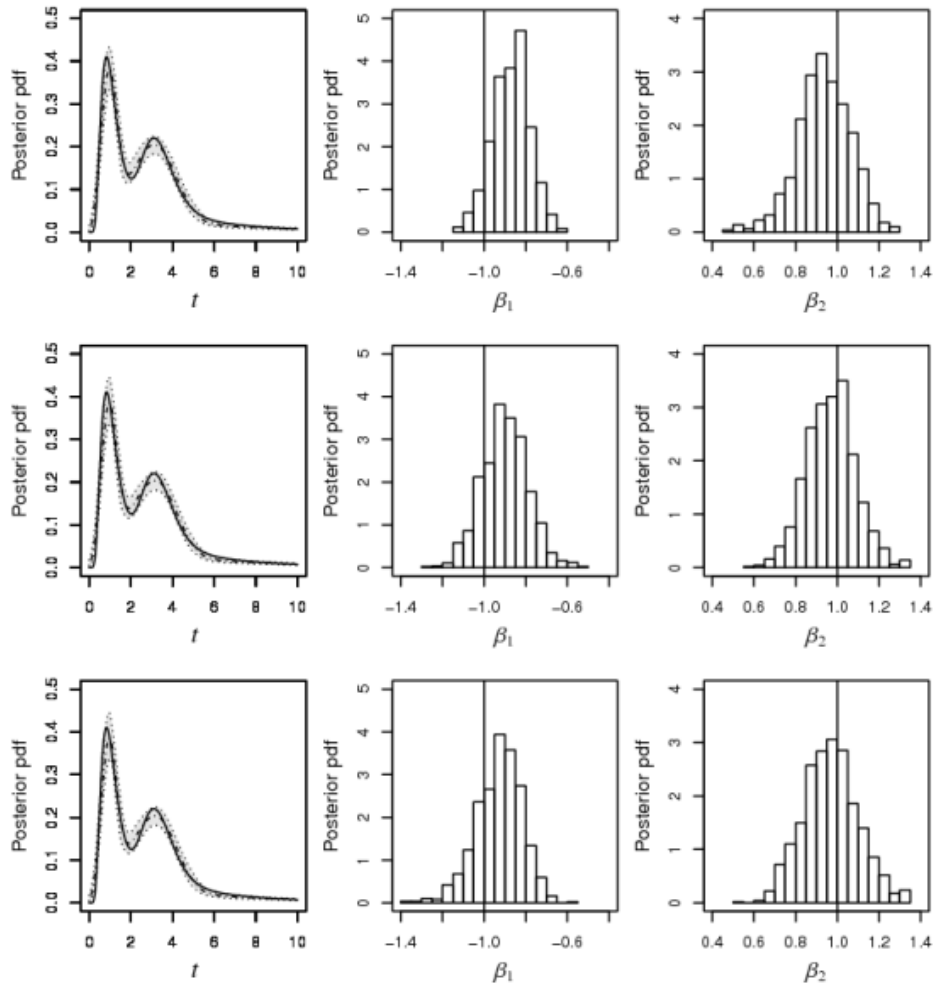
Figure 1: *Posterior failure time distribution and posterior distributions of regression coefficients under three different prior distributions for γ*

3.2. Under a non-standard baseline hazard function

This section illustrates the proposed method under a non-standard baseline hazard function and assesses prior sensitivity. We consider the same covariates and regression coefficients as in Section 3.1, but a baseline hazard function is specified by the failure time distribution that is a mixture of three lognormal distributions as in Park *et al.* (2012a),

$$0.4 \times \text{LN}(0, 0.2) + 0.3 \times \text{LN}(1.2, 0.05) + 0.3 \times \text{LN}(1.4, 0.5), \tag{3.1}$$

where $\text{LN}(\mu, \sigma^2)$ represents a lognormal distribution with mean $\exp(\mu + \sigma^2/2)$. The failure distribution in (3.1) has a bimodal shape and is heavy-tailed. With the non-standard baseline hazard function, we generate $n = 1000$ failure times from the proportional hazards model with no censoring. While the prior distributions for (η, ξ) and $\boldsymbol{\beta}$ remain the same as in Section 3.1, we consider three different prior distributions for γ. The precision parameter γ plays a key role to controll the number of clusters in

Table 2: Summary statistics of coefficients under different prior distributions for γ

Coefficient	True value	$\gamma \sim$ Gamma$(1, 1)$		$\gamma \sim$ Gamma$(2, 0.5)$		$\gamma \sim$ Gamma$(3, 0.1)$	
		Mean	95% interval	Mean	95% interval	Mean	95% interval
β_1	1	0.940	(0.603, 1.312)	0.985	(0.651, 1.313)	0.914	(0.609, 1.238)
β_2	−1	−1.081	(−1.326, −0.871)	−1.091	(−1.302, −0.906)	−1.004	(−1.210, −0.813)

a DPM model and may affect performance in the Bayesian semi-parametric inference for the proportional hazards model. Our choices are Gamma$(1, 1)$, Gamma$(2, 0.5)$, and Gamma$(3, 0.1)$ in the order of increasing the expected number of clusters in the Weibull DPM model.

The proposed PCG-type sampler is implemented on the test data, and two separate chains with different starting values run for 10000 iterations. Upon convergence based on the criterion $\hat{R}^{1/2} < 1.2$, the second halves of the two chains are collected and used to compute posterior summary statistics. Figure 1 displays the posterior distribution of failure times along with the posterior distributions of regression coefficients, β_1 and β_2: The top, middle, and bottom rows correspond to $\gamma \sim$ Gamma$(1, 1)$, $\gamma \sim$ Gamma$(2, 0.5)$, and $\gamma \sim$ Gamma$(3, 0.1)$, respectively. The left column of Figure 1 shows the posterior failure time distribution, where the solid lines represent the true curve for a failure time distribution and the dashed lines represent pointwise posterior medians and 95% posterior intervals for a failure time distribution. Figure 1 shows that the non-standard failure time distribution is well estimated by pointwise posterior medians and 95% posterior intervals. The middle and right columns of Figure 1 present the posterior distributions of regression coefficients, where the vertical solid lines represent the true values for regression coefficients. Under all different prior distributions for γ, the true values for the regression coefficients are well covered and estimated by posterior distributions. Table 2 presents the posterior mean and 95% posterior interval for regression coefficients using the three prior distributions for γ. Regardless of the choice of a prior distribution for γ, Table 2 shows that posterior estimates for regression coefficients are similar and illustrates that prior knowledge about γ does not have a significant impact on posterior inference about β.

4. Application to Heavily Censored Survival Data

We illustrate the proposed Bayesian semi-parametric regression method using data from a breast cancer clinical trial that assessed the effect of a hormonal therapy using tamoxifen in female patients with lymph node-negative and estrogen receptor-positive breast cancer (Park *et al.*, 2012a). The study randomized 1413 patients to the placebo group and 1404 patients to the tamoxifen-treated group, and the mean follow-up time was 13 years. The observed covariates included clinical tumor size and age, where the mean (95% confidence interval) of log tumor size is 2.95 (1.95, 3.91) in the placebo group and 2.96 (1.95, 4.09) in the tamoxifen-treated group, and the mean (95% confidence interval) of age is 54.67 (34, 69) in the placebo group and 54.80 (35, 69) in the tamoxifen-treated group. The endpoint of overall survival is time to any death. The median residual life function was not estimable using the semi-parametric regression method due to the heavy censoring (66%) of the censored survival data. In this paper, we use the Bayesian semi-parametric regression method to compute the 25th percentile and median residual life functions, and compare them between the tamoxifen-treated and placebo groups, while adjusting for available covariates.

The proposed PCG-type sampler is used to fit the proportional hazards model given by

$$h_i(t) = h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}),$$

and run with two separate chains of 10000 iterations, where x_{i1} is an indicator for the tamoxifen-treated group, x_{i2} is the log of tumor size for patient i, and x_{i3} is the age of patient i. After convergence
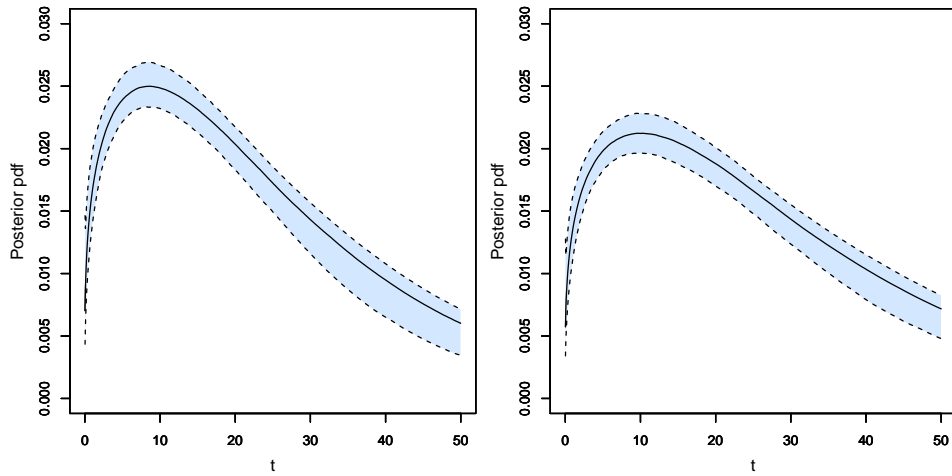
Figure 2: *Posterior failure time distributions for placebo and tamoxifen-treated groups*
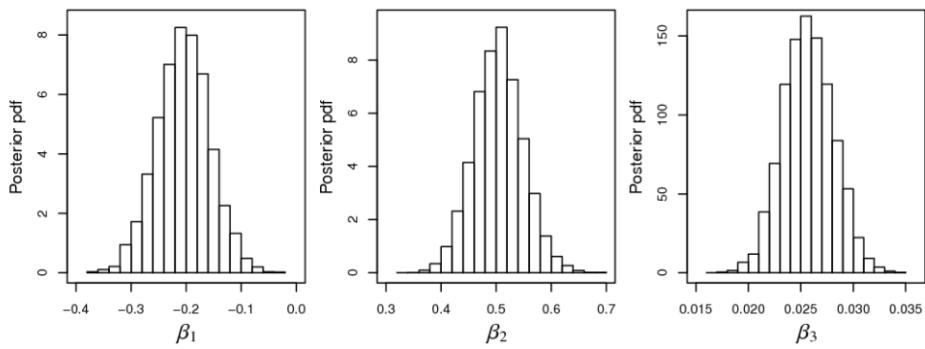

Figure 3: *Posterior distributions for regression coefficients to the tamoxifen treatment indicator, log tumor size, and age*

Table 3: Summary statistics of coefficients for the analysis of heavily censored survival data

Coefficient	Mean	95% interval
β_1	−0.205	(−0.301, −0.109)
β_2	0.505	(0.419, 0.597)
β_3	0.026	(0.021, 0.030)

is assessed by the criterion $\hat{R}^{1/2} < 1.2$, the second halves of the two chains are collected to compute posterior summary statistics. Figure 2 shows the posterior failure time distributions for the placebo group on the left panel and the tamoxifen-treated group on the right panel, where the values of log tumor size and age are fixed at their means, *i.e.*, $x_{i2} = 3$ and $x_{i3} = 55$. Figure 2 illustrates that the posterior failure time distribution for the tamoxifen-treated group has a fatter and longer tail than the placebo group, which implies better survival. Figure 3 presents the posterior distributions of regression coefficients and Table 3 provides the posterior summary statistics. The summary statistics of regression coefficients illustrate that all covariates were statistically significant because 95% credible intervals do not contain zero. The estimated hazard ratio in the tamoxifen-treated group when compared with the placebo group was $\exp(\hat{\beta}_1) = 0.82$ with a 95% credible interval of (0.74, 0.90). Because

Table 4: Summary statistics of quantiles for residual lifetime

log(tumor size)	age	t_0	q	Placebo Mean	Placebo 95% interval	Tamoxifen Mean	Tamoxifen 95% interval
2	40	0	0.25	21.93	(20.37, 23.93)	25.76	(23.53, 28.50)
			0.50	43.54	(39.87, 48.93)	51.16	(45.88, 58.24)
		1	0.25	21.27	(19.71, 23.26)	25.09	(22.86, 27.88)
			0.50	42.83	(39.13, 48.28)	50.43	(45.12, 57.61)
		2	0.25	20.74	(19.10, 22.75)	24.54	(22.29, 27.44)
			0.50	42.22	(38.44, 47.77)	49.80	(44.48, 57.08)
		3	0.25	20.28	(18.64, 22.33)	24.06	(21.79, 26.96)
			0.50	41.66	(37.86, 47.32)	49.23	(43.90, 56.60)
4	40	0	0.25	10.25	(9.12, 11.51)	12.04	(10.79, 13.40)
			0.50	20.36	(18.22, 22.74)	23.91	(21.35, 26.77)
		1	0.25	9.67	(8.57, 10.89)	11.44	(10.24, 12.77)
			0.50	19.70	(17.59, 48.28)	23.24	(20.61, 26.10)
		2	0.25	9.24	(8.21, 10.43)	10.99	(9.78, 12.30)
			0.50	19.18	(17.05, 21.55)	22.70	(20.09, 25.57)
		3	0.25	8.89	(7.86, 10.07)	10.62	(9.42, 11.97)
			0.50	18.73	(16.58, 21.11)	22.22	(19.60, 25.17)
2	60	0	0.25	14.84	(13.57, 16.14)	17.44	(15.87, 19.05)
			0.50	29.48	(26.75, 32.49)	34.63	(31.16, 38.48)
		1	0.25	14.23	(12.98, 15.52)	16.81	(15.24, 18.47)
			0.50	28.79	(26.05, 48.28)	33.93	(30.41, 37.85)
		2	0.25	13.75	(12.50, 15.05)	16.30	(14.74, 17.99)
			0.50	28.22	(25.44, 31.41)	33.35	(29.76, 37.33)
		3	0.25	13.34	(12.10, 14.67)	15.87	(14.31, 17.53)
			0.50	27.72	(24.93, 30.88)	32.82	(29.17, 36.92)
4	60	0	0.25	6.93	(6.23, 7.55)	8.14	(7.43, 8.85)
			0.50	13.77	(12.70, 14.89)	16.17	(14.85, 17.49)
		1	0.25	6.40	(5.77, 6.99)	7.59	(6.91, 8.26)
			0.50	13.15	(12.10, 14.26)	15.54	(14.25, 16.92)
		2	0.25	6.02	(5.50, 6.57)	7.19	(6.54, 7.82)
			0.50	12.68	(11.62, 13.80)	15.05	(13.73, 16.46)
		3	0.25	5.73	(5.23, 6.26)	6.88	(6.24, 7.50)
			0.50	12.28	(11.21, 13.39)	14.63	(13.31, 16.08)

$\exp(\hat{\beta}_2) = 1.66$, there was a 66% increase in the expected hazard relative to a one-unit increase in log tumor size; in addition, $\exp(\hat{\beta}_3) = 1.03$ implies that there was a 3% increase in the expected hazard relative to a one-year increase in age.

For better interpretation of the proportional hazards model, Table 4 presents summary statistics of quantiles for residual lifetime given certain values of the covariates. Because the log tumor size and age are continuous variables, we consider four cases where the covariates are given as $(x_{i2}, x_{i3}) = (2, 40), (4, 40), (2, 60),$ and $(4, 60)$. The quantile residual life functions are estimated for the tamoxifen-treated ($x_{i1} = 1$) and placebo ($x_{i1} = 0$) groups with $q = 0.25$ and 0.5 at $t_0 = 0, 1, 2,$ and 3. Table 4 shows that patients in the tamoxifen-treated group are expected to consistently have longer residual lifetime than those in the placebo group. To examine a treatment effect for longer residual lifetime, Table 5 provides the comparison of quantile residual lifetimes between the tamoxifen-treated and placebo groups, where $t^*_{i,q,1}$ and $t^*_{i,q,2}$ denote the $100(q)^{th}$ residual life function for subject i in the tamoxifen-treated and placebo groups, respectively, at time t_0. The posterior mean of difference in quantile residual lifetime between the two groups was large relative to its posterior standard deviation, so that the posterior probability of $t^*_{i,q,1} > t^*_{i,q,2}$ was one for all combinations of covariates. After adjusting for covariates, the data analysis illustrates that patients in the tamoxifen-treated group have a significantly longer residual lifetime than those in the placebo group.

Table 5: Comparison of quantile residual lifetimes between the tamoxifen-treated and placebo groups

| Covariate | | t_0 | q | $t^*_{i,q,1} - t^*_{i,q,2}$ | |
log(tumor size)	age			Mean	Standard deviation
2	40	0	0.25	3.83	0.93
			0.50	7.62	1.89
		1	0.25	3.82	0.93
			0.50	7.60	1.89
		2	0.25	3.80	0.93
			0.50	7.59	1.89
		3	0.25	3.78	0.92
			0.50	7.57	1.88
4	40	0	0.25	1.79	0.43
			0.50	3.55	0.86
		1	0.25	1.77	0.42
			0.50	3.54	0.85
		2	0.25	1.75	0.42
			0.50	3.52	0.85
		3	0.25	1.73	0.41
			0.50	3.50	0.84
2	60	0	0.25	2.60	0.62
			0.50	5.15	1.26
		1	0.25	2.58	0.62
			0.50	5.14	1.26
		2	0.25	2.56	0.62
			0.50	5.12	1.25
		3	0.25	2.54	0.61
			0.50	5.10	1.24
4	60	0	0.25	1.21	0.28
			0.50	2.40	0.56
		1	0.25	1.19	0.28
			0.50	2.39	0.56
		2	0.25	1.17	0.27
			0.50	2.37	0.55
		3	0.25	1.14	0.27
			0.50	2.34	0.55

5. Discussion

This paper develops the Bayesian semi-parametric regression method to fit the proportional hazards model in the presence of covariates. The existing semi-parametric regression method cannot deal with a heavy censoring problem to estimate quantile residual life functions until the targeted quantiles are reached. The existing parametric regression method can estimate any quantile residual life function regardless of the degree of censoring, but may suffer from strong parametric assumptions. Thus, this paper proposes the Bayesian semi-parametric regression method that takes advantages from both approaches. Under the proposed Bayesian semi-parametric approach, any quantile residual life function is estimable in a fully model-based framework without sacrificing the flexibility of modeling the unknown baseline hazard function.

Appendix A: Log-Concavity of Non-Standard Conditional Distributions

The sampling steps for β and ρ^*_j are not standard, so that we have to resort in an indirect sampling method like adaptive rejection sampling. To do so, we show that the conditional distributions of β and

ρ_j^* are log-concave. First, the Hessian matrix of the log conditional distribution of $\boldsymbol{\beta}$ can be written as

$$\nabla^2 \log p\left(\boldsymbol{\beta}|\mathbf{Z}, \mathbf{V}, \boldsymbol{\rho}^*, \gamma, \eta, \xi, \mathbf{Y}\right) = -\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} - \sum_{j=1}^{J} \frac{\left(1 + \sum_{i:Z_i=j} d_i\right)\left(\xi \sum_{i:Z_i=j} \mathbf{x}_i y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right)\mathbf{x}_i^{\top} + \Xi\right)}{\left(\xi + \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right)\right)^2},$$

where

$$\Xi = \sum_{i:Z_i=j} \mathbf{x}_i y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right)\mathbf{x}_i^{\top} \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right) - \sum_{i:Z_i=j} \mathbf{x}_i y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right) \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right)\mathbf{x}_i^{\top}.$$

Because of the Cauchy-Schwarz inequality, we have $\text{tr}(\Xi) \geq 0$ for all $\boldsymbol{\beta}$ and thus Ξ is positive semi-definite. Combined with the fact that $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}$ is positive definite, this implies that the Hessian matrix of the log conditional distribution of $\boldsymbol{\beta}$ is negative definite, so that the conditional distribution of $\boldsymbol{\beta}$ is shown to be log-concave. Second, the second derivative of the log conditional distribution of ρ_j^* is given by

$$\frac{d^2 \log p(\rho_j^*|\mathbf{Z}, \mathbf{V}, \boldsymbol{\beta}, \gamma, \eta, \xi, \mathbf{Y})}{d(\rho_j^*)^2} = -\frac{\sum_{i:Z_i=j} d_i}{(\rho_j^*)^2} - \frac{\left(1 + \sum_{i:Z_i=j} d_i\right)\left(\xi \sum_{i:Z_i=j}(\log y_i)^2 y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right) + \psi\right)}{\left(\xi + \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right)\right)^2},$$

where

$$\psi = \sum_{i:Z_i=j}(\log y_i)^2 y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right) \sum_{i:Z_i=j} y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right) - \sum_{i:Z_i=j}(\log y_i) y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right) \sum_{i:Z_i=j}(\log y_i) y_i^{\rho_j^*} \exp\left(\mathbf{x}_i^{\top}\boldsymbol{\beta}\right)$$

that is non-negative based on the Cauchy-Schwarz inequality. As long as at least one subject belongs to cluster j, the second derivative of the log conditional distribution of ρ_j^* is negative for all ρ_j^*, so that the conditional distribution of ρ_j^* is shown to be log-concave .

References

Carlin, B. P. and Hodges, J. S. (1999). Hierarchical proportional hazards models for highly stratified data, *Biometrics*, **55**, 1162–1170.

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

Cox, D. R. (1975). Partial likelihood, *Biometrika*, **62**, 269–276.

Gelfand, A. E. and Kottas, A. (2003). Bayesian semiparametric regression for median residual life, *Scandinavian Journal of Statistics*, **30**, 651–665.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulations using multiple sequences (with discussion), *Statistical Science*, **7**, 457–472.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, **41**, 337–348.

Hjort, N. L. and Petrone, S. (2007). Nonparametric quantile inference using Dirichlet processes. In Nair, V., editor, *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A. Doksum*, 463–492. World Scientific Publishing Company, Singapore.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors, *Journal of the American Statistical Association*, **96**, 161–173.

Jeong, J. H., Jung, S. H. and Costantino, J. P. (2008). Nonparametric inference on median residual life function, *Biometrics*, **64**, 157–163.

Kim, M.-O., Zhou, M., and Jeong, J.-H. (2012). Censored quantile regression for residual lifetimes, *Lifetime Data Analysis*, **18**, 177–194.

Kottas, A. (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions, *Journal of Statistical Planning and Inference*, **136**, 578–596.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling, *Journal of the American Statistical Association*, **96**, 1458–1468.

Kottas, A. and Krnjajic, M. (2009). Bayesian semiparametric modelling in quantile regression, *Scandinavian Journal of Statistics*, **36**, 297–319.

Kottas, A., Krnjajic, M. and Taddy, M. (2007). Model-based approaches to nonparametric Bayesian quantile regression, In *Proceedings of the 2007 Joint Statistical Meetings*, 1137–1148.

Kuo, L. and Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model, *Canadian Journal of Statistics*, **25**, 457–472.

Park, T. (2011). Bayesian analysis of individual choice behavior with aggregate data, *Journal of Computational and Graphical Statistics*, **20**, 158–173.

Park, T., Jeong, J. and Lee, J. (2012a). Bayesian nonparametric inference on quantile residual life function: Application to breast cancer data, *Statistics In Medicine*, **31**, 1972–1985.

Park, T., Krafty, R. T. and Sanchez, A. I. (2012b). Bayesian semi-parametric analysis of Poisson change-point regression models: Application to policy making in Cali, Colombia, *Journal of Applied Statistics*, **39**, 2285–2298.

Park, T. and van Dyk, D. A. (2009). Partially collapsed Gibbs samplers: Illustrations and applications, *Journal of Computational and Graphical Statistics*, **18**, 283–305.

Park, T., van Dyk, D. A. and Siemiginowska, A. (2008). Searching for narrow emission lines in X-ray spectra: Computation and methods, *The Astrophysical Journal*, **688**, 807–825.

Sethuraman, J. A. (1994). A constructive definition of Dirichlet priors, *Statistica Sinica*, **4**, 639–650.

van Dyk, D. A. and Park, T. (2008). Partially collapsed Gibbs samplers: Theory and methods, *Journal of the American Statistical Association*, **103**, 790–796.