Bayesian Analysis (2024) **19**, Number 4, pp. 1225–1243

Bayesian Analysis of the Generalized Additive Proportional Hazards Model: Asymptotic Studies*

Gwangsu Kim†, Chang D Yoo‡, and Yongdai Kim§

Abstract. In this paper, we study Bayesian asymptotic properties of the proportional hazards model where the link function is modeled by the generalized additive model. As the standard generalized additive model is, the generalized additive proportional hazards model is a useful tool in finding the nonlinearity of covariate effects to survival times. We develop a data-dependent sieve prior for the generalized additive link function and use the Bayesian bootstrap prior for the baseline hazard function. We prove that the posterior contraction rate of the generalized additive link function is minimax optimal up to a $\log n$ term when the prior is carefully selected. By analyzing simulated as well as real data, we verify our theoretical results and compare with exisiting algorithms for the generalized additive proportional hazards model to illustrate that the proposed Bayesian model is a useful inference tool.

Keywords: b-spline, generalized additive model, partial likelihood, posterior contraction rate, proportional hazards model.

MSC2020 subject classifications: Primary 62F15; secondary 60G15, 62G20.

1 Introduction

For decades, the proportional hazards (PH) model (Cox, 1975) has been widely used to assess the effect of covariates on the hazard function. The typical PH model assumes that the hazard function of survival time T given a covariate vector $z \in \mathbb{R}^b$ has the form

$$\lambda(t \mid z) = \exp(z'\beta)\lambda_0(t), \ t \in [0, \infty), \tag{1}$$

where $\lambda_0(t)$ is a fully unspecified baseline hazard function. The key assumption of the PH model is that $\log \lambda(t \mid z) = z'\beta + \log \lambda_0(t)$. That is, the log hazard function is linear

*This paper was supported by research funds for newly appointed professors of Jeonbuk National University in 2023. This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.2022-0-00184, Development and Study of AI Technologies to Inexpensively Conform to Evolving Policy on Ethics] and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1A2C3A01003550).

†Jeonbuk National University, Department of Statistics (Institute of Applied Statistics), 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do, Republic of Korea, s88012@jbnu.ac.kr

‡Korea Advanced Institute of Science and Technology, School of Electrical Engineering, 291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea, cd_yoo@kaist.ac.kr

§Seoul National University, Department of Statistics, 1, Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea, ydkim0903@gmail.com

 https://doi.org/10.1214/23-BA1384

in z, which is referred to as the *proportionality*. The hazard increases as a constant rate over time when z increases in a unit value. Even though the proportionality allows interpretation of the regression coefficients to be simple, the linearity in z may not hold in many situations (Keele, 2010).

Various modifications of the PH model for nonproportionality have been proposed. Time-varying coefficients models, which allow the regression coefficient β depending on time t, have been considered (Lambert and Eilers, 2005; Tian et al., 2005; Thomas and Reyes, 2014; Kim et al., 2017), and nonparametric regression approaches, which replace $z'\beta$ in (1) by $f(z)$, have been studied (Chen et al., 2010; Hastie and Tibshirani, 1990; Huang, 1999; Shang et al., 2013; Argyropoulos and Unruh, 2015; Gaïffas and Guilloux, 2012). In this paper, we focus on the second approach – the PH model with a nonparametric link function.

Nonparametric regression approaches for the PH model assume

$$\log \lambda(t \mid z) = f(z) + \log \lambda_0(t), \tag{2}$$

where the link function f is modeled by the generalized additive model (GAM) (Hastie and Tibshirani, 1990), the partial linear model (Huang, 1999; Shang et al., 2013) or the single index model (Gaïffas and Guilloux, 2012). Among these, the GAM is known to be a useful compromise between a linear and fully nonparametric models in particular when b is large.

In this paper, we consider Bayesian analysis of the proportional hazards model with the link function f in (2) modeled by the GAM such that

$$\log \lambda(t \mid z) = \sum_{a=1}^{b} f_a(z_a) + \log \lambda_0(t), \tag{3}$$

where $f_a : \mathbb{R} \to \mathbb{R}$. We call the model (3) the Generalized Additive Proportional Hazards (GA-PH) model. Bayesian analysis of the GA-PH model has been already considered by Umlauf et al. (2018), Scheipl et al. (2012) and Yi et al. (2019).

Our main contributions are (i) to develop a prior and the corresponding posterior sampling algorithm for f and λ_0 based on a data-dependent B-spline Basis functions and Bayesian boostrap approach (Kim and Lee, 2003a) and (ii) to study asymptotic properties of the resulting posterior distribution of $f = (f_1, \ldots, f_b)$ to prove that the posterior contraction rate is minimax optimal (up to log term) and adaptive with right-censored data. Advatages of our Bayesian model are that the model is flexible and the posterior sampling is simple and efficient along with the desirable large sample properties.

Bayesian boostrap approach of (Kim and Lee, 2003a) is about the prior on λ_0. For $\lambda_0(t)$, various priors have been proposed such as gamma processes (Kalbfleisch, 1978), beta processes (Hjort, 1990), mixture of Lévy processes (Nieto-Barajas and Walker, 2004) and mixture of Weibull (Kottas, 2006). In this paper, we use the Bayesian bootstrap (BB) prior proposed by Kim and Lee (2003a). The BB prior is conceptually and computationally easy to use compared to the aforementioned full Bayesian approaches

as it does not require prior information and the marginal likelihood of f becomes the partial likelihood. In addition, Kim and Lee (2003a) proved that the BB prior could be understood as a limit of the Lévy process priors including gamma and beta processes and enjoys desirable large sample properties. Kim et al. (2011) and Kim et al. (2017) utilize the Bayesian bootstrap prior for monotone hazard models and time-dependent coefficient models.

Our study is based on Bayesian nonparametric approaches for survival analysis, which have been studied for several decasdes but still are active. Castillo and van der Pas (2021) prove the Bernstein von-Mises theorem for linear functionals of the hazard function with right censored data. Dependent priors for multiple hazard functions have been proposed by Camerlenghi et al. (2021) and Riva-Palacio et al. (2022), and Bayesian analysis for dependent censoring problems is considered by Paulon et al. (2022). A copula-based predictive update of Bayesian nonparametric survival analysis is proposed by Fong and Lehmann (2022), which does not require the specification of complex nonparametric priors.

The remainder of this paper is organized as follows. Setups of the model and prior are given in Section 2. In Section 3, the posterior contraction rate of the GA-PH, which is minimax optimal up to a $\log n$ term, is derived when the smoothness of the true link function is known. In addition, a prior on the smoothness of the link function is devised so that the posterior contraction rate is minimax optimal adaptive to the true smoothness. The results of simulation studies and real data analysis are presented in Sections 4 and 5, respectively, and concluding remarks follow in Section 6. All proofs for Lemmas and Theorems are deferred to the Supplementary Material (Kim et al., 2023).

2 Model, prior and posterior

2.1 Model and data

Let $X_1, \ldots, X_n$ be independent survival times and $z_1, \ldots, z_n \in [-B, B]^b$ be covariate vectors for $B > 0$. It is assumed that the hazard function $\lambda(t \mid z_i)$ of X_i for given z_i is given as (3). The survival times are subject to right censoring, and only $D_{(1:n)} = \{(T_1, \delta_1, z_1), \ldots, (T_n, \delta_n, z_n)\}$ are observed, where $T_i = \min\{X_i, C_i\}, \delta_i = I(X_i \leq C_i)$ and $C_1, \ldots, C_n$ are independent censoring variables having a common distribution function G. As usual, we assume that the censoring variables are independent of the X_i's. Let f^* and λ_0^* be the true link and baseline hazard function, respectively. For f^*, we assume that $\sup_{z_a \in [-B,B]} |f_a^*(z_a)| < A$ and for all $a = 1, \ldots, b$,

$$\int_{-B}^{B} f_a^*(z_a) dz_a = 0. \tag{4}$$

The constraint (4) is introduced to make f^* and λ_0^* be identifiable.

2.2 Prior

We consider a sieve prior for $f = (f_1, \ldots, f_b)$ as follows. Let

$$\Theta_p = \left\{ f = (f_1, \ldots, f_b) : \|f^{(p)}\|_\infty < A, \|f\|_\infty < A, \int_{-B}^{B} f_a(z_a) dz_a = 0, a = 1, \ldots, b \right\}, \quad (5)$$

where $\|f\|_\infty = \sum_{a=1}^{b} \sup_{z_a \in [-B,B]} |f_a(z_a)|$ and $f_a^{(p)}$ is the p-th derivative of f_a. For given basis functions $B_{a_n,k}(\cdot), k = 1, \ldots, a_n$, the sieve $\mathcal{S}_n$ of Θ_p considered in this paper is given as:

$$\begin{aligned} \mathcal{S}_n \quad = \quad & \left\{ f = (f_1, \ldots, f_b) : f_a(z_a) = \sum_{k=1}^{a_n} \tau_{a,k} B_{a_n,k}(z_a), \sup_{a,k} |\tau_{a,k}| \le L \right. \\ & \left. \int_{-B}^{B} f_a(z_a) dz_a = 0, a = 1, \ldots, b \right\}. \quad (6) \end{aligned}$$

For the basis functions, we focus on B-spline basis functions of degree s having equally spaced $a_n + s + 1$ knots on $[-B, B]$. However, the proofs for the posterior contraction rate in this paper can be modified for other basis functions (e.g., Fourier basis) without much difficulty. For details of the B-spline, the definition and properties of the B-spline basis functions can be found in Györfi et al. (2002), and we summarize the formal definition of B-spline basis functions and examples in Section 1 of the Supplementary Material for reader's sake. It is also known that $\zeta \stackrel{\text{def}}{=} \sup_{f \in \mathcal{S}_n} \|f\|_\infty \le (\max\{2^s [s(s-1)]^s / s!, 1\})^{-1} L$.

For $f_a \in \mathcal{S}_n$, we can write $f_a(\cdot) = \boldsymbol{\tau}_a' \boldsymbol{B}_{a_n}(\cdot)$ for some vector $\boldsymbol{\tau}_a$, where $\boldsymbol{B}_{a_n}(\cdot)$ is the vector of B-spline basis functions

$$\boldsymbol{B}_{a_n}(\cdot) = (B_{a_n,1}(\cdot), \ldots, B_{a_n,a_n}(\cdot))'.$$

In this paper, we use f and $\boldsymbol{\tau} = (\boldsymbol{\tau}_1', \ldots, \boldsymbol{\tau}_b')'$ interchangeably unless there is any confusion. Due to the constraint (4), the free parameters are $\boldsymbol{\tau}_{a,-a_n} = (\tau_{a,1}, \ldots, \tau_{a,a_n-1})'$ in the sense that τ_{a,a_n} is defined automatically once $\boldsymbol{\tau}_{a,-a_n}$ is given. For notational simplicity, we use $\boldsymbol{\tau}_a$ and $\boldsymbol{\tau}_{a,-a_n}$ interchangeably unless otherwise stated.

The stability of the B-spline (Lyche and Mørken, 2008) implies that there exist $0 < q_1 \le q_2 < \infty$ such that

$$q_1 \|f\|^2 \quad \le \quad \boldsymbol{\tau}'\boldsymbol{\tau}/a_n \le q_2 \|f\|^2, \quad (7)$$

where $\|f\| = \left\{ \sum_{a=1}^{b} \int_{-B}^{B} |f_a(z_a)|^2 dz_a \right\}^{1/2}$. This implies the equivalence up to a constant between the l_2 norm of $\boldsymbol{\tau}$ divided by a_n and the l_2 norm of f.

The following proposition, which plays a key role in our derivation, is about the approximating property of the B-spline basis functions whose proof can be found in de Boor et al. (1976), Györfi et al. (2002) or Lyche and Mørken (2008).

Proposition 1. *Suppose that the degree s of the B-spline basis functions is greater than or equal to p. Then, for any $f^* \in \Theta_p$, there exists $f^{B,*} \in \mathcal{S}_n$ satisfying*

$$\|f^* - f^{B,*}\|_\infty = O\left(a_n^{-p}\right).$$

Remark. Here, we only consider 'the equally spaced knots' to simplify the B-spline approximation (Lyche and Mørken, 2008). In practice, B-spline basis functions with equally spaced knots approximate low curvature functions well. On the other hand, for functions having high curvature, data-dependent knots perform better (Sharef et al., 2010). For the GA-PH model, we did a small simulation study about data-dependent knots whose results are given in Section 7 of the Supplementary Material.

The proposed sieve prior for $(f_1, \cdots f_b)'$ puts prior probabilities on $\boldsymbol{\tau}$ such that

$$\begin{aligned} &\pi_n^{\mathsf{p}}(d\tau_{1,1}, \cdots, d\tau_{1,a_n-1}, \cdots, d\tau_{b,1}, \cdots, d\tau_{b,a_n-1}) \\ \propto &\prod_{a=1}^{b} \left\{ \prod_{k=1}^{a_n-1} \phi(\tau_{a,k} \mid 0, \sigma^2) \mathbb{I}(|\tau_{a,k}| \le L) d\tau_{a,1} d\tau_{a,2}, \cdots, d\tau_{a,a_n-1} \right\}, \end{aligned} \tag{8}$$

where $\phi(\cdot \mid \mu, \sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and variance σ^2, and $\mathbb{I}(\cdot)$ is the indicator function. Note that the constraint (4) implies $\int \tau_{a,a_n} B_{a_n,a_n}(z_a) dz_a = -\int \sum_{k=1}^{a_n-1} \tau_{a,k} B_{a_n,k} d(z_a) dz_a$ and thus τ_{a,a_n} is defined once $\tau_{a,1}, \ldots, \tau_{a,a_n-1}$ are given.

For the prior on λ_0, we use the BB prior of Kim and Lee (2003a). Let q_n be the number of distinct uncensored observations and $0 < t_1 < t_2 < \cdots < t_{q_n}$ be the corresponding ordered distinct uncensored observations. Let $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$. Consider the empirical likelihood $L_n(f, \Lambda_0)$ defined as:

$$\begin{aligned} &L_n(f, \Lambda_0) \\ = &\prod_{i=1}^{n} \left[\exp\left\{ \sum_{a=1}^{b} f_a(z_{ia}) \right\} \Delta\Lambda_0(T_i) \right]^{\delta_i} \exp\left[-\exp\left\{ \sum_{a=1}^{b} f_a(z_{ia}) \right\} \sum_{t_k \le T_i} \Delta\Lambda_0(t_k) \right] \end{aligned} \tag{9}$$

See Kim and Lee (2003b) for details of the empirical likelihood (9). The BB treats the empirical likelihood of (9) as the real likelihood with the parameters f and $\Delta\Lambda_0(t_1), \ldots, \Delta\Lambda_0(t_{q_n})$ and puts the prior on $\Delta\Lambda_0(t_1), \ldots, \Delta\Lambda_0(t_{q_n})$ as well as f, where

$$\pi_n^{\mathsf{p}}(\Lambda_0) \propto \prod_{k=1}^{q_n} \frac{1}{\Delta\Lambda_0(t_k)}. \tag{10}$$

2.3 BB posterior inference

The joint posterior distribution of f and Λ in the BB approach is proportional to

$$\begin{aligned} &\prod_{i=1}^{n} \left[\exp\left\{ \sum_{a=1}^{b} f_a(z_{ia}) \right\} \Delta\Lambda_0(T_i) \right]^{\delta_i} \exp\left[-\exp\left\{ \sum_{a=1}^{b} f_a(z_{ia}) \right\} \sum_{t_k \le T_i} \Delta\Lambda_0(t_k) \right] \\ \times &\prod_{k=1}^{q_n} \frac{1}{\Delta\Lambda_0(t_k)} \pi_n^{\mathsf{p}}(d\tau_{1,1}, \cdots, d\tau_{b,a_n-1}), \end{aligned} \tag{11}$$

where $f_a(\cdot) = \boldsymbol{\tau}_a' \boldsymbol{B}_{a_n}(\cdot)$. Let $D(t) = \{k : T_k = t, \delta_k = 1\}$ and $R(t) = \{j : T_j \geq t\}$. We can approximate this joint posterior distribution by the following.

By direct calculation, the marginal posterior of $\boldsymbol{\tau}$ from (11) turns out to be

$$\begin{aligned} &\pi_n^{\mathsf{p}}(d\tau_{1,1}, \cdots, d\tau_{1,a_n-1}, \cdots, d\tau_{b,1}, \cdots, d\tau_{b,a_n-1} \mid D_{(1:n)}) \\ \propto \; &\prod_{i=1}^{q_n} \frac{\sum_{k \in D(t_i)} \exp\big(\sum_{a=1}^{b} f_a(z_{ka})\big)}{\sum_{j \in R(t_i)} \exp\big(\sum_{a=1}^{b} f_a(z_{ja})\big)/n} \pi_n^{\mathsf{p}}(d\tau_{1,1}, \cdots, d\tau_{b,a_n-1}). \end{aligned} \tag{12}$$

Note the first term of (12) is the partial likelihood $\overset{\mathsf{p}}{n}(f)$ as:

$$L_n^{\mathsf{p}}(f) = \prod_{i=1}^{q_n} \frac{\sum_{k \in D(t_i)} \exp\big(\sum_{a=1}^{b} f_a(z_{ka})\big)}{\sum_{j \in R(t_i)} \exp\big(\sum_{a=1}^{b} f_a(z_{ja})\big)/n}.$$

Since the partialikelihood is log-concave, we can generate $\boldsymbol{\tau}$ from its marginal posterior distribution easily.

The conditional posterior of $\Delta\Lambda_0(t_k), k = 1, \ldots, q_n$ given f is given as:

$$\pi_n^{\mathsf{p}}(\Delta\Lambda_0(t_k) \mid f, D_{(1:n)}) = \frac{\beta_k^{\alpha^k}}{\Gamma(\alpha_k)} \Delta\Lambda_0(t_k)^{\alpha_k - 1} \exp\{-\Delta\Lambda_0(t_k)\beta_k\}, \tag{13}$$

where $\alpha_k = |D(t_k)|$ and $\beta_k = \sum_{i \in R(t_k)} \exp\left\{\sum_{a=1}^{b} f_a(z_{ia})\right\}$, which is the gamma distribution with the parameters α_k and β_k. Hence, the baseline hazard function can be generated easily from the conditional posterior distribution.

Hereafter we use the notations of $\pi_n^{\mathsf{p}}(df)$ and $\pi_n^{\mathsf{p}}(df \mid D_{(1:n)})$ instead of $\pi_n^{\mathsf{p}}(d\tau_{1,1}, \cdots, d\tau_{b,a_n-1})$ and $\pi_n^{\mathsf{p}}(d\tau_{1,1}, \cdots, d\tau_{b,a_n-1} \mid D_{(1:n)})$ for simplicity.

3 Posterior contraction rates of f

In this section we derive the posterior contraction rate of f given in (12) when the true smoothness of the true link function is known. Then, we develop a prior on the smoothness so that the resulting posterior is adaptive in the sense that the posterior contraction rate is minimax optimal (up to a $\log n$ term) regardless of the smoothness of the true link function.

We say that the posterior contraction rate is ϵ_n if for any positive diverging sequence M_n,

$$\pi_n^{\mathsf{p}}\big[\{f \in \mathcal{S}_n : \|f - f^*\| \geq \epsilon\} \mid D_{(1:n)}\big] \to 0$$

in P_n^* probability as $n \to \infty$, where P_n^* is the true probability measure of $\{(T_i, \delta_i, z_i)\}_{i=1}^n$ with the true link functions f^* and the true baseline hazard function λ_0^*.

We assume that the covariate vectors $z_1, \ldots, z_n$ are independent realizations of the random vector Z. Throughout this paper, we assume the following three regularity conditions.

(A1) The G satisfies $G(\tau) = 1, G(\tau-) < 1$ for some $\tau \in (0, \infty)$ and has only a finite number of jumps.

(A2) The $\mathrm{Cov}(Z)$ is positive definite, and Z has a probability density function h_z such that

$$h_1 \leq \inf_{w \in [-B,B]^b} h_z(w) \leq \sup_{w \in [-B,B]^b} h_z(w) \leq h_2$$

for two positive constants h_1 and h_2.

(A3) There exist positive constant a_1 and a_2 such that $a_1 \leq \lambda_0^*(t) \leq a_2$ for all $t \in [0, \tau]$.

Assumption (A1) means that the follow-up of each subject is terminated before τ, which holds in most studies. Assumption (A2) basically requires that the covariate vector is nonsigular. In this paper, we only consider continuous covariate vectors for technical simplicity but categorical covariate vectors can be easily incoporated. Assumption (A3) assumes that the distribution of survival time is absolutely continuous, which is assumed in most studies of survival analysis.

3.1 Posterior contraction rate when the smoothness of the true link function is known

The following theorem provides the posterior contraction rate of the BB posterior of f, which is the same as the minimax optimal convergence rate (Györfi et al., 2002) up to the $\log n^{p/(2p+1)}$ term.

Theorem 1. *Suppose that regularity conditions (A1)–(A3) hold and $f^* \in \Theta_p$. Then, under the B-spline prior π_n^{B} with $a_n = [(n/\log n)^{1/(2p+1)}]$, we have that for any diverging sequence M_n,*

$$\pi_n^{\mathrm{B}}\left(\|f - f^*\| \geq M_n (n/\log n)^{-p/(2p+1)} \mid D_{(1:n)}\right) \to 0$$

in probability as $n \to \infty$.

We use the BB prior on the baseline hazard function, which is not a proper prior. We may wonder what is the posterior contraction rate of a full Bayesian posterior. When the prior of the baseline hazard function is the beta process (Hjort, 1990), we can prove that the posterior contraction rate of f is the same as that in Theorem 1. The detailed proof is given in Section 2 of the Supplementary Materials.

3.2 Posterior contraction rate with adaptive prior when the smoothness of the true link function is unknown

For given f^*, let p^* be the smoothness of f^* defined as $p^* = \sup_p \{p : f^* \in \Theta_p\}$. Suppose that p^* is unknown. In this case, a standard Bayesian approach is to use the hierarchical

prior where a prior is placed on p and the sieve prior conditional on p is used for f. We devise a prior on p which is adaptive to p^* in the sense that the posterior contraction rate of f is the same as that with the case of known p^*.

For an adaptive prior, we consider the following prior on p:

$$\pi_n(p) \propto \exp\left(-\eta_n n(n/\log n)^{-p/(2p+1)}\right), \tag{14}$$

where η_n is a sequence of diverging real numbers such that $\lim_{n\to\infty} \eta_n/\log n = 0$.

The prior (14), which is devised to have desirable large sample properties, is not of a standard distribution. However, there is an important implication. The prior $\pi_n(p)$ decreases in p and the decreasing rate becomes exponentially faster as n is getting larger. That is, to achieve the optimal posterior concentration rate, the prior should strongly discourage larger p. The following theorem proves that the posterior distribution of p under the prior (14) is consistent.

Theorem 2. *Suppose that regularity conditions (A1)–(A3) hold, the prior $\pi_n(p)$ satisfies (14) and the sieve prior introduced in Section 2 with $a_n = [(n/\log n)^{1/(2p+1)}]$ conditional on p is used for f. Then, for any f^* with the smoothness p^*, we have*

$$\pi_n^{\mathsf{p}}(p = p^* \mid D_{(1:n)}) \to 1$$

with probability tending to 1 as $n \to \infty$.

A direct consequence of Theorem 2 is that the hierarchical prior with the prior (14) on p and the sieve prior on f conditional on p yields the posterior distribution whose convergence rate is minimax optimal up to a $\log n$ term, which is stated in the following corollary.

Corollary 1. *Suppose that the prior $\pi_n(p)$ satisfies (14) and the sieve prior introduced in Section 2 with $a_n = [(n/\log n)^{1/(2p+1)}]$ conditional on p is used for f. If the smoothness of f^* is p^*, then we have that for any diverging sequence M_n,*

$$\pi_n^{\mathsf{p}}\left(\|f - f^*\| \geq M_n(n/\log n)^{-p^*/(2p^*+1)} \mid D_{(1:n)}\right) \to 0$$

in probability as $n \to \infty$.

4 Simulation studies

In this section, we investigate finite sample properties of the BB posterior distribution of f by analyzing simulated data. The degree of the B-spline basis functions is set to 3, for which we denote our Bayesian method GA-PH(3). The Markov chain Monte Carlo (MCMC) algorithm used in this section is described in Section 6 of the Supplementary Materials. For computational simplicity, instead of (4) we use the constraint $f_a(m_a) = 0$, where $m_a = \min(\mathrm{supp}(z_a))$. Here $\min(\mathrm{supp}(z_a))$ is the minimum value of $\mathrm{supp}(z_a)$, where $\mathrm{supp}(z_a)$ is the support of the covariate z_a.

We consider three cases for simulation. The first case considers one component GA-PH models and the second case considers a GA-PH model with two components. For these two cases, we investigate the performance of the BB posterior distribution and compare them with the two competitors: MGCV (Wood et al., 2016) and bamlss (Umlauf et al., 2018), where the first is the frequetists method based on a penalized likelihood, and the later is a Bayesian approach using the normal approximation to generate MCMC samples. Note that PH-GA(3) uses the M-H algorithm.

The third case considers a GA-PH model with two components but the true model is an one-component GA-PH model. We apply *the slab-and-spike* prior along with our sieve prior and investigate how well our Bayesian model selects the signal component. For comparison, we also apply BhGLM (Yi et al., 2019) and *relgam* (Tay and Tibshirani, 2020) that select signal components by a prior and a penalty function, respectively.

4.1 Simulation study I

We first considers the cases where the dimension of the covariate is one. We let $\lambda_0^*(t) = 1$ and the distribution of censoring time be the exponential distribution with mean $1/\eta$. The covariate is generated from the mixture of normal distributions:

$$0.5\mathcal{N}(0.6, 0.4^2) + 0.5\mathcal{N}(-0.6, 0.4^2)$$

truncated on $[-1, 1]$. For the link function f^* generating data, we consider the following four functions: $3\log(1+z)$, $2(1-z^2)$, $2\mathbb{I}(z > 0)$, and $3\sin(2\pi z)$, i.e., log, quadratic, step, and sine functions. The η is selected so that the censoring probability becomes 0.2 for the log and quadratic link functions and 0.4 and 0.5 for the step and the sine functions, respectively.

Four basis functions with equally spaced knots are assigned to the log and quadratic functions and eights and tens of basis functions are assigned to the step and sine functions, respectively. We assign more basis functions to the step and sine functions since they are more complex. For the competitors, MGCV and bamlss with sufficient numbers (no less than 10) of basis functions are assigned for fair comparison. Detailed setups for the considered algorithms including the burn-in, thinning and hyperparameter selections are given in Section 6 of the Supplementary Materials. In addition, the effect of the number of basis functions and the random knot selection instead of the equally spaced knots are discussed in Section 7 of the Supplementary Materials.

Figure 1 compares GA-PH(3) with MGCV and bamlss. For Bayesian models, i.e., GA-PH(3) and bamlss, the Bayes estimates (posterior means) of the link function and the point-wise 95% credible intervals are provided, whereas the point estimates and 95% confidence intervals are given for MGCV. We only present the results for the sample size $n = 100$ here and the results for other sample sizes are given in Section 7 of the Supplementary Materials. The interval estimates of GA-PH are much better than those of MGCV and bamlss in the sense that the intervals well include the true link functions. In addition, it is interesting that the estimates of MGCV and bamlss look almost linear for the step function while GA-PH(3) captures an abrupt change from the discontinuity point of the step function.

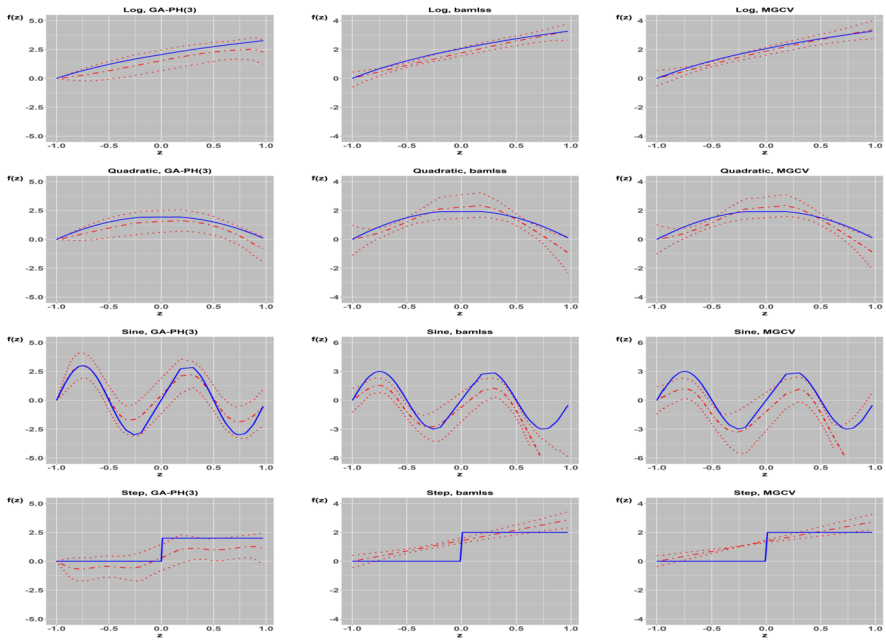

Figure 1: Bayes estimates (dot-dashed lines) and the upper and lower limits of the 95% credible intervals (dotted lines) for the four link functions when sample size is 100. Left to right: GA-PH(3), bamlss, and MGCV.

Data size	Algorithm	Log	Quadratic	Sine	Step	Average
		ℓ_2 (CP)	ℓ_2 (CP)	ℓ_2 (CP)	ℓ_2 (CP)	ℓ_2 (CP)
$n = 100$	GA-PH(3)	0.13 (0.95)	0.15 (0.95)	0.45 (0.97)	0.22 (0.97)	0.23 (0.96)
	bamlss	0.09 (1.00)	0.27 (1.00)	1.70 (0.79)	0.30 (0.61)	0.59 (0.85)
	MGCV	0.10 (1.00)	0.27 (0.97)	1.53 (0.97)	0.33 (0.57)	0.56 (0.88)
$n = 300$	GA-PH(3)	0.16 (1.00)	0.13 (1.00)	0.33 (0.97)	0.14 (1.00)	0.19 (0.99)
	bamlss	0.14 (0.83)	0.11 (1.00)	0.32 (0.97)	0.16 (1.00)	0.18 (0.95)
	MGCV	0.13 (0.19)	0.12 (0.16)	0.35 (0.96)	0.18 (0.99)	0.20 (0.99)
$n = 500$	GA-PH(3)	0.08 (0.99)	0.11 (0.99)	0.21 (0.96)	0.13 (1.00)	0.13 (0.98)
	bamlss	0.08 (1.00)	0.10 (0.99)	0.26 (1.00)	0.11 (0.99)	0.14 (1.00)
	MGCV	0.09 (1.00)	0.11 (1.00)	0.23 (1.00)	0.13 (0.97)	0.14 (0.99)

Table 1: ℓ_2 is defined as $\ell_2^2 = \min_c \sum_{i=1}^n (\hat{f}(z_i) - f(z_i) - c)^2/n$, where $\hat{f}(\cdot)$ and $f(\cdot)$ are the Bayes estimate and true function respectively, and CP is the proportion of z_is where the point-wise credible/confidence intervals of f at each z_i include the value of the true link function.

Table 1 compares the ℓ_2 distances of the point estimates from the true link function defined as $\ell_2^2 = \min_c \sum_{i=1}^n (\hat{f}(z_i) - f(z_i) - c)^2/n$, and the coverage proportions (CPs) defined as the proportion of z_is where the point-wise credible/confidence intervals of f at each z_i include the value of the true link function. In Table 1, we can clearly see

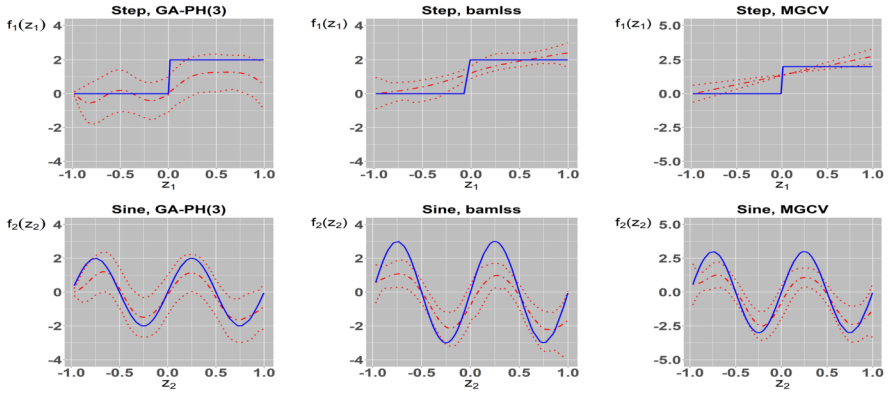

Figure 2: Bayes estimates (red dot-dashed lines) and the upper and lower limits of the 95% credible/confidence intervals (dotted lines) for two component functions when the sample size is 100. Left to right: GA-PH(3), bamlss, and MGCV. The censoring rate is set to be 0.5.

Functions	Data size	GA-PH(3)	bamlss	MGCV
		ℓ_2 (CP)	ℓ_2 (CP)	ℓ_2 (CP)
Step (1st)	$n = 100$	0.48 (0.93)	0.43 (0.82)	0.50 (0.56)
	$n = 300$	0.32 (0.93)	0.25 (0.94)	0.25 (0.94)
	$n = 500$	0.31 (0.92)	0.23 (0.92)	0.24 (0.91)
Sine (2nd)	$n = 100$	0.53 (1.00)	1.05 (0.46)	0.92 (0.57)
	$n = 300$	0.26 (0.99)	0.90 (0.29)	0.90 (0.34)
	$n = 500$	0.20 (0.97)	0.65 (0.37)	0.69 (0.36)
Average of two components	$n = 100$	0.50 (0.97)	0.74 (0.64)	0.71 (0.56)
	$n = 300$	0.29 (0.96)	0.58 (0.62)	0.58 (0.64)
	$n = 500$	0.26 (0.95)	0.44 (0.64)	0.47 (0.64)

Table 2: ℓ_2 is defined as $\ell_2^2 = \min_c \sum_{i=1}^n (\hat{f}(z_i) - f(z_i) - c)^2/n$, where $\hat{f}(\cdot)$ and $f(\cdot)$ are the Bayes estimate and true function respectively, and CP is the proportion of z_is where the point-wisecredible/confidence intervals of f at each z_i include the value of the true link function.

that GA-PH(3) performs much better in terms of ℓ_2 (i.e. the accuracy of the point estimate), especially when either the sample sizes are small or the true link function is the sine function. Moreover, the CPs of MGCV and bamlss become worse when data size increases while those of GA-PH(3) remain stable with respect to the sample size.

4.2 Simulation study II: Multivariate case

We consider a two component GA-PH model, where the true link function for data generation is given as $f^*(z) = f_1(z_1) + f_2(z_2)$ where $z \in [-1, 1]^2$,

$$f_1^*(z_1) = 2\mathbb{I}(z_1 > 0), \text{ and } f_2^*(z_2) = 2\sin(2\pi z_2).$$

For simulation, we generate z_1 and z_2 independently from the uniform distribution on $[-1, 1]$. For the data-dependent prior, the numbers of basis functions are eight and ten for f_1 and f_2, respectively. The censoring rate is set to be 0.5 and the other setups are the same as those in Section 4.1.

Figure 2 presents the point estimates and credible/confidence intervals for the two components when $n = 100$. The figures for other sample sizes are presented in Section 7 of the Supplementary Material. It is observed that the true link function is almost encapsulated by the 95% credible intervals of the GA-PH(3). In contrast, the 95% intervals of bamlss and MGCV do not include the sine function properly. In addition, the point estimates of *bamlss* and *MGCV* for the step function look linear.

Table 2 presents the ℓ_2s and CPs of the two components. For the point estimation, GA-PH(3) is inferior for the step function but superior for the sine function. It is observed that GA-PH(3) measures the uncertainties much better than bamlss and MGCV do, especially when the sample size is large. The CPs of the sine function for bamlss and MGCV becomes worse as the sample size increases, which indicates that bamlss and MGCV would not quantify the uncertainties properly.

4.3 Spike-and-slab prior for GA-PH

In this section, we consider the component selection in the GA-PH model. For this purpose, the *spike-and-slab* prior is used with the GA-PH model, which is formulated as:

$$\begin{aligned} \log \lambda(t \mid z_1, \ldots, z_b, \gamma_1, \ldots \gamma_b) &= \sum_{a=1}^{b} \gamma_a f_a(z_a) + \log \lambda_0(t), \\ \pi(\gamma_a = 1) &= v \ (0 < v < 1), \ a = 1, \ldots, b. \end{aligned} \tag{15}$$

See O'Hara et al. (2009) for related literature about Bayesian model selection methods including the *spike-and-slab* prior.

For the *spike-and-slab* prior, a standard MCMC algorithm could suffer from slow mixing in particular for γ_a. For generating γ_a from its conditional posterior, a Metropolis-Hastings algorithm is used. The proposal distribution for $\boldsymbol{\tau}_a$ should be carefully designed to improve the acceptance rate. In our MCMC algorithm, the normal distribution with the mean vector $\hat{\boldsymbol{\tau}}_a$ and the covariance matrix of $H_a(\hat{\boldsymbol{\tau}}_a)^{-1}$ is used for the proposal distribution of $\boldsymbol{\tau}_a$ when $\gamma_a = 0$, where $H_a(\hat{\boldsymbol{\tau}}_a)$ is the negative Hessian matrix of the log-partial likelihood with respect to $\boldsymbol{\tau}_a$ evaluated at $\hat{\boldsymbol{\tau}}_a$. Details are explained in Section 6 of the Supplementary Material.

For simulation, we let $b = 2$ and let $f_1^*(z_1) = 0$ and $f_2^*(z_2) = \mathbb{I}(z_2 > 0)$. The covariate vectors are generated as:

$$z_1 = \sqrt{w} U_1 + \sqrt{1 - w} U_2, \ z_2 = U_2,$$

where $U_i \overset{i.i.d.}{\sim} Unif[-1, 1]$, in which z_1 ad z_2 are correlated with the correlation $\sqrt{1 - w}$. We investigate how the correlation affects the posterior distribution. Usually, correlations between covariates affect the efficiency of the estimator negatively. For example,

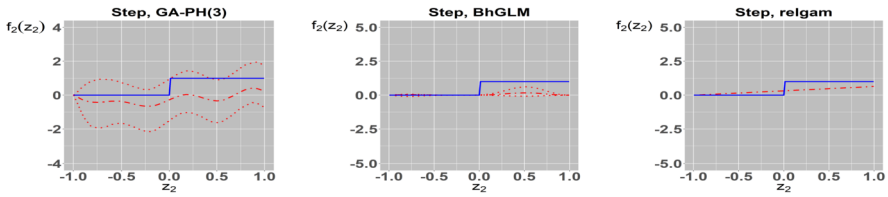

Figure 3: Tree figures (GA-PH(3), BhGLM, and relgam) are Bayes estimates (dot-dashed lines) for the second component and the upper and lower limits of the 95% credible/confidence intervals (dotted lines), when the sample size is 100 and $r = 0.4$.

Function (r)	Data size	GA-PH(3)	BhGLM	relgam
		ℓ_2 (CP)	ℓ_2 (CP)	ℓ_2 (CP)
Null (0.00)	$n = 100$	0.1332 (1.0000)	0.0002 (1.0000)	0.1714 (NA)
	$n = 300$	0.0068 (1.0000)	< 0.0000 (1.0000)	0.0330 (NA)
	$n = 500$	0.0071 (1.0000)	0.0050 (1.0000)	0.0781 (NA)
Null (0.40)	$n = 100$	0.1239 (1.0000)	0.0001 (1.0000)	0.2636 (NA)
	$n = 300$	0.0234 (1.0000)	0.0095 (1.0000)	0.1050 (NA)
	$n = 500$	0.0062 (1.0000)	0.0136 (1.0000)	0.0119 (NA)
Null (0.80)	$n = 100$	0.0566 (1.0000)	0.0244 (1.0000)	< 0.0000 (NA)
	$n = 300$	0.0050 (1.0000)	0.0122 (1.0000)	< 0.0000 (NA)
	$n = 500$	0.0707 (1.0000)	0.0767 (1.0000)	0.0939 (NA)
Step (0.00)	$n = 100$	0.2817 (0.8800)	0.5186 (0.5200)	0.2241 (NA)
	$n = 300$	0.1782 (1.0000)	0.4279 (0.6400)	0.2040 (NA)
	$n = 500$	0.1261 (0.9940)	0.4116 (0.8920)	0.1735 (NA)
Step (0.40)	$n = 100$	0.3140 (0.9800)	0.4725 (0.5500)	0.3366 (NA)
	$n = 300$	0.2142 (0.9533)	0.4506 (0.6333)	0.2306 (NA)
	$n = 500$	0.1531 (0.9620)	0.4406 (0.8720)	0.1778 (NA)
Step (0.80)	$n = 100$	0.2396 (0.9900)	0.4937 (0.4900)	0.3788 (NA)
	$n = 300$	0.2488 (1.0000)	0.4926 (0.1233)	0.2480 (NA)
	$n = 500$	0.1702 (0.8840)	0.5323 (0.2060)	0.2160 (NA)

Table 3: ℓ_2 is defined as $\ell_2^2 = \min_c \sum_{i=1}^n (\hat{f}(z_i) - f(z_i) - c)^2/n$, where $\hat{f}(\cdot)$ and $f(\cdot)$ are the Bayes estimate and true function respectively, and CP is the proportion of z_is where the point-wise credible/confidence intervals of f at each z_i include the value of the true link function. Here, ρ in the parenthesis is the correlation coefficient of two covariates.

Huang (1999) observed this phenomenon in the partially linear proportional hazards model.

For simulation studies, we let the censoring variable be the exponential distribution with mean r and choose the r so that the censoring rates for all setups are between 0.20∼0.25. We set v in (15) to be 1/11 which results in $\pi_n(\gamma_a = 1)/\pi_n(\gamma_a = 0) = 0.1$. The other setups are the same as those in Sections 4.1–2.

For comparison, we consider the algorithms of BhGLM (Nengjun et al., 2018) and

	Null			Step		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
$r = 0.0$	0.5233	0.1120	0.1280	0.9300	1.0000	1.0000
$r = 0.4$	0.4420	0.1933	0.1166	0.5800	1.0000	1.0000
$r = 0.8$	0.2380	0.3561	0.4213	0.8806	1.0000	1.0000

Table 4: Posterior mean of γ_a.

relgam (Tay and Tibshirani, 2020). BhGLM uses a double-exponential prior for each basis function. BhGLM is a recently developed Bayesian method which enables variable selection in the GA-PH model. The hyper-parameters used in BhGLM are given in Section 7 of the Supplementary Material. In contrast, relgam, which selects components in the GA-PH model by use of a sparse penalty, provides only a point estimate but does not yield a confidence interval. Thus we obtain the point estimates and credible intervals for GA-PH (3) and BhGLM while we only obtain the point estimates for relgam.

The Bayes estimates and pointwise probability intervals of the signal components are drawn in Figure 3 (the results for $n = 100$ are presented and the results for other sample sizes are given in Section 7 of the Supplementary Material) and the ℓ_2 and CPs are reported in Table 3. The results indicate that our Bayesian model estimates the signal component much better than BhGLM. In particular, the CPs of BhGLM of the signal component are much less than the nominal level 0.95. GA-PH(3) is competitive to relgam in point estimate, but uncertainty quantification is not available for relgam.

It is interesting to observe the behavior of γ_a in the GA-PH(3), summarized in Table 4. The posterior mean of γ_1 for the null function seems to go to the zero when the data size increases, and r is not large. However, the posterior mean of γ_1 is relatively large for $r = 0.80$, which is partly due to the correlation between z_1 and z_2. On the other hand, GA-PH(3) identifies the non-zero function quickly as the sample size increases.

Remark. Though the ℓ_2 of GA-PH(3) for the null function behaves slightly irregularly with respect to data size. This is because the posterior mean is not exactly zero for the null function. In contrast, if we use the posterior median as the Bayes estimate, all estimates of the null function become 0, whose results are given in Section 7 of the Supplementary Material. As shown in the Supplementary Material, using the median as the Bayes estimates is preferable when we are interested in component selection.

5 Data analysis

We analyze the Chemotherapy Data for Stage B/C colon cancer (CDS) archived in R survival package (Moertel et al., 1990). In CDS, there are eight covariates consisting of six binary variables and two continuous variables. The six binary variables are 1) the treatment Lev(amisole), 2) Lev(amisole)+5-FU, 3) sex, 4) obstruction of colon by tumor (Obstruction), 5) perforation of colon (Perforation), 6) adherence of nearby organs (Adherence), and the two continuous covariates are age (Age) and the number of lymph nodes detectable cancer (Nodes). We consider the following partially linear

Variable	BE	90% CI	95% CI
Lev(amisole)	-0.0398	$(-0.2145, 0.1456)$	$(-0.2550, 0.1900)$
Lev(amisole)+5-FU	-0.5269	$(-0.7158, -0.3338)$	$(-0.7592, -0.2930)$
Sex	-0.1242	$(-0.2810, 0.0239)$	$(-0.3113, 0.0505)$
Obstruction	0.2106	$(0.0212, 0.3967)$	$(-0.0092, 0.4347)$
Perforation	0.1565	$(-0.2638, 0.5596)$	$(-0.3647, 0.6542)$
Adherence	0.2525	$(0.0486, 0.4485)$	$(0.0066, 0.4874)$

Table 5: Bayes estimates (BEs) and 90% and 95% credible intervals (CIs) of the regression coefficients of the binary covariates.

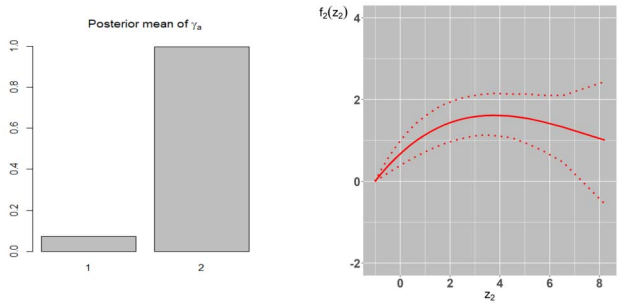

Figure 4: Left panel: the posterior probabilities of $\gamma_a = 1$, Right panel: Bayes estimate and point-wise 95% credible interval for the component of Nodes.

GA-PH model with the spike-and-slab prior:

$$\log \lambda(t \mid x_1, \ldots, x_6, z_1, z_2) = \sum_{k=1}^{6} \beta_k x_k + \sum_{a=1}^{2} \gamma_a f_a(z_a) + \log \lambda_0(t),$$

where $\gamma_a \in \{0, 1\}$ and $\pi(\gamma_a = 1) = v$ for $a = 1, 2$. Here, x_ks are binary covariates, and z_1 and z_2 are Age and Nodes, respectively. The four B-spline basis functions of order 3 are used for the sieve prior.

Table 5 summarizes the Bayes estimates and 90% and 95% credible intervals of the six binary covariates, and Figure 4 presents the posterior probabilities of $\gamma_a = 1$ and the Bayes estimates of the second component of nonlinear term with the 95% point-wise credible intervals. Note that the posterior probability of γ_1 is close 0 and thus we eliminate the first component of nonlinear term for further analysis. It is interesting to see that the effect of Nodes on the hazard is quite nonlinear. The hazard increases when $z_2 \leq 4$ but stops increasing after that.

6 Concluding remarks

We only considered the GAM in the proportional hazards model. Our results could be extended for other nonparametric models such as the single index model (Shang et al.,

2013). Moreover, similar results could be obtained for other semi-parametric regression models for survival data, such as the generalized transformation model (Doksum, 1987) even though computation would be demanding compared to GA-PH.

We have used a sieve prior. However, our techniques to derive the posterior contraction rate cannot be directly applicable to other nonparametric priors such as Gaussian process priors (van der Vaart and van Zanten, 2008). It would be interesting to drive posterior contraction rates for other nonparametric priors.

The MCMC algorithms used for our numerical studies and data analysis are not the fully efficient. More efficient MCMC algorithms could be developed using the log concavity of $G_n(f)$ as is done by Görür and Teh (2011) and Dwivedi et al. (2018).

Supplementary Material

Supplementary Material of Bayesian analysis of the generalized additive proportional hazards model: Asymptotic studies (DOI: 10.1214/23-BA1384SUPP; .pdf). Proofs, details of numerical studies and MCMC algorithm, and additional numerial studies.

References

Argyropoulos, C. and Unruh, M. (2015). "Analysis of time to event outcomes in randomized controlled trials by generalized additive models." *PloS one*, 10(4): e0123784. 1226

Camerlenghi, F., Lijoi, A., and PrüNnster, I. (2021). "Survival analysis via hierarchically dependent mixture hazards." *The Annals of Statistics*, 49(2): 863–884. MR4255111. doi: https://doi.org/10.1214/20-aos1982. 1227

Castillo, I. and van der Pas, S. (2021). "Multiscale Bayesian survival analysis." *The Annals of Statistics*, 49(6): 3559–3582. MR4352541. doi: https://doi.org/10.1214/21-aos2097. 1227

Chen, K., Guo, S., Sun, L., and Wang, J.-L. (2010). "Global partial likelihood for nonparametric proportional hazards models." *Journal of the American Statistical Association*, 105(490): 750–760. MR2724858. doi: https://doi.org/10.1198/jasa.2010.tm08636. 1226

Cox, D. (1975). "Partial likelihood." *Biometrika*, 62(2): 269–276. MR0400509. doi: https://doi.org/10.1093/biomet/62.2.269. 1225

de Boor, C., Lyche, T., and Schumaker, L. L. (1976). "On calculating with B-splines II. Integration." In *Numerische Methoden der Approximationstheorie/Numerical Methods of Approximation Theory*, 123–146. Springer. MR0510689. 1228

Doksum, K. (1987). "An extension of partial likelihood methods for proportional hazard models to general transformation models." *The Annals of Statistics*, 325–345. MR0885740. doi: https://doi.org/10.1214/aos/1176350269. 1240

Dwivedi, R., Chen, Y., Wainwright, M. J., and Yu, B. (2018). "Log-concave sampling: Metropolis-Hastings algorithms are fast!" *arXiv preprint arXiv:1801.02309*. MR4048994. 1240

Fong, E. and Lehmann, B. (2022). "A Predictive Approach to Bayesian Nonparametric Survival Analysis." In *International Conference on Artificial Intelligence and Statistics*, 6990–7013. PMLR. MR0395032. 1227

Gaïffas, S. and Guilloux, A. (2012). "High-dimensional additive hazards models and the Lasso." *Electronic Journal of Statistics*, 6: 522–546. MR2988418. doi: https://doi.org/10.1214/12-EJS681. 1226

Görür, D. and Teh, Y. W. (2011). "Concave-convex adaptive rejection sampling." *Journal of Computational and Graphical Statistics*, 20(3): 670–691. MR2878996. doi: https://doi.org/10.1198/jcgs.2011.09058. 1240

Györfi, L., Kohler, M., Krzyźak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer. MR1920390. doi: https://doi.org/10.1007/b97848. 1228, 1231

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*, volume 43. CRC press. MR1082147. 1226

Hjort, N. (1990). "Nonparametric Bayesian estimators based on beta processes in models for lie history data." *The Annals of Statistics*, 18(3): 1259–1294. MR1062708. doi: https://doi.org/10.1214/aos/1176347749. 1226, 1231

Huang, J. (1999). "Efficient estimation of the partly linear additive Cox model." *The Annals of Statistics*, 27(5): 1536–1563. MR1742499. doi: https://doi.org/10.1214/aos/1017939141. 1226, 1237

Kalbfleisch, J. (1978). "Non-parametric Bayesian analysis of survival time data." *Journal of the Royal Statistical Society. series B (Methodological)*, 32(4): 214–221. MR0517442. 1226

Keele, L. (2010). "Proportionally difficult: testing for nonproportional hazards in Cox models." *Political Analysis*, 18(2): 189–205. 1226

Kim, G., Kim, Y., and Choi, T. (2017). "Bayesian Analysis of the Proportional Hazards Model with Time-Varying Coefficients." *Scandinavian Journal of Statistics*, 44(2): 524–544. MR3658525. 1226, 1227

Kim, Y., Kim, J., and Kim, G. (2011). "Bayesian analysis for monotone hazard ratio." *Lifetime Data Analysis*, 17(2): 302–320. MR2777122. doi: https://doi.org/10.1007/s10985-010-9181-x. 1227

Kim, Y. and Lee, J. (2003a). "Bayesian analysis of proportional hazard models." *The Annals of Statistics*, 31(2): 483–511. MR1983539. doi: https://doi.org/10.1214/aos/1051027878. 1226, 1227, 1229

Kim, Y. and Lee, J. (2003b). "Bayesian bootstrap for proportional hazards models." *The Annals of Statistics*, 31(6): 1905–1922. MR2036394. doi: https://doi.org/10.1214/aos/1074290331. 1229

Kim, G., Yoo, C. D., and Kim, Y. (2023). "Supplementary Material for "Bayesian Analysis of the Generalized Additive Proportional Hazards Model: Asymptotic Studies"." *Bayesian Analysis*. doi: https://doi.org/10.1214/23-BA1384SUPPA. 1227

Kottas, A. (2006). "Nonparametric Bayesian survival analysis using mixtures of Weibull distributions." *Journal of Statistical Planning and Inference*, 136(3): 578–596. MR2181970. doi: https://doi.org/10.1016/j.jspi.2004.08.009. 1226

Lambert, P. and Eilers, P. H. (2005). "Bayesian proportional hazards model with time-varying regression coefficients: a penalized Poisson regression approach." *Statistics in Medicine*, 24(24): 3977–3989. MR2221979. doi: https://doi.org/10.1002/sim.2396. 1226

Lyche, T. and Mørken, K. (2008). *Spline Methods (draft)*. Department of Informatics, Centre of Mathematics for Applications, University of Oslo. 1228, 1229

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., et al. (1990). "Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma." *New England Journal of Medicine*, 322(6): 352–358. 1238

Nengjun, Y., Zaixiang, T., Xinyan, Z., and Boyi, G. (2018). "BhGLM: Bayesian hierarchical GLMs and survival models, with applications to genomics and epidemiology." *Bioinformatics*, 35(8): 1419–1421. 1237

Nieto-Barajas, L. and Walker, S. (2004). "Bayesian nonparametric survival analysis via Lévy driven Markov processes." *Statistica Sinica*, 1127–1146. MR2126344. 1226

O'Hara, R. B., Sillanpää, M. J., et al. (2009). "A review of Bayesian variable selection methods: what, how and which." *Bayesian analysis*, 4(1): 85–117. MR2486240. doi: https://doi.org/10.1214/09-BA403. 1236

Paulon, G., Müller, P., and y Rosas, V. G. S. (2022). "Bayesian nonparametric bivariate survival regression for current status data." *Bayesian Analysis*, 1(1): 1–27. 1227

Riva-Palacio, A., Leisen, F., and Griffin, J. (2022). "Survival regression models with dependent Bayesian nonparametric priors." *Journal of the American Statistical Association*, 117(539): 1530–1539. MR4480729. doi: https://doi.org/10.1080/01621459.2020.1864381. 1227

Scheipl, F., Fahrmeir, L., and Kneib, T. (2012). "Spike-and-slab priors for function selection in structured additive regression models." *Journal of the American Statistical Association*, 107(500): 1518–1532. MR3036413. doi: https://doi.org/10.1080/01621459.2012.737742. 1226

Shang, S., Liu, M., Zeleniuch-Jacquotte, A., Clendenen, T., Krogh, V., Hallmans, G., and Lu, W. (2013). "Partially linear single index Cox regression model in nested case-control studies." *Computational Statistics & Data Analysis*, 67: 199–212. MR3079597. doi: https://doi.org/10.1016/j.csda.2013.05.011. 1226, 1239

Sharef, E., Strawderman, R. L., Ruppert, D., Cowen, M., and Halasyamani, L. (2010). "Bayesian adaptive B-spline estimation in proportional hazards frailty models." *Elec-*

tronic Journal of Statistics, 4: 606–642. MR2660535. doi: https://doi.org/10.1214/10-EJS566. 1229

Tay, J. K. and Tibshirani, R. (2020). "Reluctant Generalised Additive Modelling." *International Statistical Review*, 88(S1): S205–S224. 1233, 1238

Thomas, L. and Reyes, E. M. (2014). "Tutorial: survival estimation for Cox regression models with time-varying coefficients using SAS and R." *J. Stat. Softw.*, 61(c1): 1–23. 1226

Tian, L., Zucker, D., and Wei, L. (2005). "On the Cox model with time-varying regression coefficients." *Journal of the American Statistical Association*, 100(469): 172–183. MR2156827. doi: https://doi.org/10.1198/016214504000000845. 1226

Umlauf, N., Klein, N., and Zeileis, A. (2018). "BAMLSS: Bayesian Additive Models for Location, Scale, and Shape (and Beyond)." *Journal of Computational and Graphical Statistics*, 27(3): 612–627. MR3863762. doi: https://doi.org/10.1080/10618600.2017.1407325. 1226, 1233

van der Vaart, A. and van Zanten, J. (2008). "Rates of contraction of posterior distributions based on Gaussian process priors." *The Annals of Statistics*, 36(3): 1435–1463. MR2418663. doi: https://doi.org/10.1214/009053607000000613. 1240

Wood, S. N., Pya, N., and Säfken, B. (2016). "Smoothing Parameter and Model Selection for General Smooth Models." *Journal of the American Statistical Association*, 111(516): 1548–1563. MR3601714. doi: https://doi.org/10.1080/01621459.2016.1180986. 1233

Yi, N., Tang, Z., Zhang, X., and Guo, B. (2019). "BhGLM: Bayesian hierarchical GLMs and survival models, with applications to genomics and epidemiology." *Bioinformatics*, 35(8): 1419–1421. 1226, 1233