Semiparametric Inference for the Functional Cox Model

Meiling Hao[a], Kin-yat Liu[b], Wei Xu[c] and Xingqiu Zhao[d]

[a]School of Statistics, University of International Business and Economics, Beijing, China

[b]Department of Mathematics and Statistics, The Hang Seng University of Hong Kong, Hong Kong

[c]Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

[d]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

This article studies penalized semiparametric maximum partial likelihood estimation and hypothesis testing for the functional Cox model in analyzing right-censored data with both functional and scalar predictors. Deriving the asymptotic joint distribution of finite-dimensional and infinite-dimensional estimators is a very challenging theoretical problem due to the complexity of semiparametric models. For the problem, we construct the Sobolev space equipped with a special inner product and discover a new joint Bahadur representation of estimators of the unknown slope function and coefficients. Using this key tool, we establish the asymptotic joint normality of the proposed estimators and the weak convergence of the estimated slope function, and then construct local and global confidence intervals for an unknown slope function. Furthermore, we study a penalized partial likelihood ratio test, show that the test statistic enjoys the Wilks phenomenon, and also verify the optimality of the test. The theoretical results are examined through simulation studies, and a right-censored data example from the Improving Care of Acute Lung Injury Patients study is provided for illustration.

KEYWORDS: Functional Cox model; Joint Bahadur representation; Partial likelihood ratio test; Penalized partial likelihood; Right-censored data.

1 Introduction

Advances in information technology enable us to collect and process densely observed data over some temporal or spatial domains. The resulting data are coined functional data in order to differentiate them from the traditional, scalar data. Examples of functional data include hippocampus radial distance data (Li and Luo, 2017), high dimensional microarray gene expression data (Chen et al., 2011), and the Sequential Organ Failure Assessment data (Gellar, et al., 2014, 2015). The explosion of functional data necessitates the development of functional data analysis. Recently, Crambes, Kneip, and Sarda (2009), Yuan and Cai (2010), Cai and Yuan (2012), Cheng and Shang (2015), and Shang and Cheng (2015), among others, proposed roughness regularization methods to control the model complexity in a continuous manner. This overcomes the imprecise control on the model complexity due to the truncation parameter in the functional principal component analysis FPCA-based approaches, as pointed out by Ramsay and Silverman (2005).

When time-to-event data are available, the proportional hazards model (Cox, 1972) is commonly used for the analysis of such data. Under the Cox model, the hazard function of a failure time for a subject takes the form:

$$h(t|Z) = h_0(t)\exp(\theta_0^\top Z),$$

where $h_0(\cdot)$ is an unspecified baseline hazard function, $Z \in \mathbb{R}^p$ is a covariate vector, and $\theta_0 \in \mathbb{R}^p$ is an unknown parameter. This model was further studied by Cox (1975), Tsiatis (1981), Andersen and Gill (1982), Johansen (1983), and Jacobsen (1984), among others. When functional covariates are involved, Chen et al. (2011) proposed the following functional Cox model:

$$h(t|Z, X(\cdot)) = h_0(t) \exp\left\{\theta_0^\top Z + \int_{\mathbb{I}} X(s)\beta_0(s)ds\right\}, \tag{1.1}$$

where $X(\cdot)$ is a functional covariate and $\beta_0(\cdot)$ is an unknown coefficient function. Clearly, this model takes into account the effect of the entire trajectory of $X(\cdot)$ on the hazard function. Note that the Cox model with a time-dependent covariate only considers the effect of $X(\cdot)$ on the hazards function at time t, where an overall effect of a functional covariate on the hazard function cannot be explained. Chen et al. (2011) applied the functional Cox model in studying the survival of diffuse large-B-cell lymphoma (DLBCL) patients in relation to the patients' high-dimensional microarray gene expression, which are expressed as functional predictors. Recently, Kong et al. (2018) established the rate of convergence of the maximum approximate partial likelihood estimator and conducted a score test for testing the nullity of the slope function related to functional predictors. Qu et al. (2016) studied the asymptotic properties of the maximum partial likelihood estimator under the framework of reproducing kernel Hilbert space and established the asymptotic normality and the efficiency of the estimator of the scalar parameter. However, the asymptotic distribution of the maximum partial likelihood estimator of the slope function has not been studied. Another important issue is to study the partial likelihood ratio test, which has not been addressed in the literature. Our goal is to address these challenging issues and to fill the gap in the study of functional Cox model.

Motivated by Cheng and Shang (2015), we explore a joint Bahadur representation to derive the asymptotic joint distribution of the maximum partial likelihood estimators of the slope function and coefficients in the functional Cox model. Different from Cheng and Shang (2015) and Shang and Cheng (2015), we deal with right-censored data. The main contributions of this paper are threefold. First, we construct the Sobolev space equipped with a special inner product and deduced the joint Bahadur representation of the maximum partial likelihood estimators of finite-dimensional and infinite-dimensional parameters. Second, we establish the joint asymptotic normality of the estimated scalar and functional coefficients and the weak convergence of the estimated functional coefficient in the Hilbert space. Third, we develop a penalized partial likelihood ratio test for testing global effects of both scalar and functional covariates on the hazard function.

The rest of this paper is organized as follows. In Section 2, we construct the Sobolev space and present a penalized estimation approach for unknown regression parameters in the functional Cox model. In Section 3, we derive a joint Bahadur representation of the maximum partial likelihood estimators of scalar and functional parameters in the space with a special inner product and establish the asymptotic properties of the proposed estimators. In Section 4, we develop a penalized likelihood ratio test for a global hypothesis. In Section 5, we present simulation results to evaluate the performance of the proposed asymptotic inference procedures. Section 6 illustrates an application of the proposed method to the data obtained from the Improving Care of Acute Lung Injury Patients (ICAP) study (Needham et al., 2006). Some concluding remarks are made in Section 7. All technical proofs are given in the Supplemental Materials.

2 Estimation Method

Denote the covariates that are incorporated in the functional Cox model (1.1) by $W = (Z^\top, X(\cdot))$. Under the right censorship, let T be the survival time, C be the censoring time, $Y = \min(T, C)$ be the observed time, and $\Delta = \mathbf{1}(T \le C)$ be the censoring indicator, where $\mathbf{1}(\cdot)$ is the indicator function. For simplicity, assume $E(\Delta Z) = 0$, and $E\{\Delta X(t)\} = 0$ for any $t \in \mathbb{I}$. Without loss of generality, we take $\mathbb{I} = [0, 1]$. As usual, assume that the survival time T and the censoring time C are conditionally independent given W. Our goal is to estimate $\alpha_0 = (\theta_0^\top, \beta_0(\cdot))$ to reveal the relationship between W and T. Suppose that $\beta_0(\cdot)$ belongs to the mth-order Sobolev space $\mathcal{H}^{(m)}(\mathbb{I})$, which is abbreviated as $\mathcal{H}^{(m)}$ for notational simplicity:

$$\mathcal{H}^{(m)}(\mathbb{I}) = \{\beta : \mathbb{I} \mapsto \mathbb{R} | \beta^{(j)} \text{ is absolutely continuous for } j = 0, 1, \ldots, m-1, \beta^{(m)} \in L_2(\mathbb{I})\},$$

where the constant $m > 1/2$ is known, $\beta^{(j)}(\cdot)$ is the jth derivative of $\beta(\cdot)$, and $L_2(\mathbb{I})$ is the L_2 space defined in $\mathbb{I}$.

Define $\eta_\alpha(W) = \theta^\top Z + \int_{\mathbb{I}} X(s)\beta(s)\, ds$, and $\mathcal{Y}(t) = \mathbf{1}(Y \ge t)$. The log partial likelihood of α under model (1.1) given the data $\{(Y_i, W_i, \Delta_i), i = 1, \ldots, n\}$ is

$$l_n(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \left(\eta_\alpha(W_i) - \log \left[\frac{1}{n} \sum_{j=1}^{n} \mathcal{Y}_j(Y_i) \exp\{\eta_\alpha(W_j)\} \right] \right).$$

To estimate α_0, we propose to use the following penalized log partial likelihood function

$$l_{n,\lambda}(\alpha) = l_n(\alpha) - \frac{\lambda}{2} J(\beta, \beta),$$

where $J(\beta_1, \beta_2) = \int_{\mathbb{I}} \beta_1^{(m)}(s)\beta_2^{(m)}(s)\, ds$ is the penalty function, and λ is the penalty parameter which controls the balance between the bias and the smoothness of the parameter. Thus, the penalized estimator of α_0 is defined by $\hat{\alpha}_{n,\lambda} = \arg\max_{\alpha \in \mathcal{H}} l_{n,\lambda}(\alpha)$, where $\mathcal{H} = \mathbb{R}^p \times \mathcal{H}^{(m)}$.

3 Asymptotic Properties

Before stating the main results, we first introduce some notation and regularity conditions. For any vector z, $z^{\otimes 2} = zz^\top$, $z^{\otimes 1} = z$, and $z^{\otimes 0} = \mathbf{1}$ with all of the elements being 1. Let $l_\lambda(\alpha)$ be limit of $l_{n,\lambda}(\alpha)$, $\mathcal{S}_\lambda(\alpha)$ be the Fréchet derivative of $l_\lambda(\alpha)$, and let D be the Fréchet derivative operator. Then

$$\mathcal{S}_\lambda(\alpha)\alpha_1 = E\left\{\int_0^\tau \left(E\eta_{\alpha_1}(W) - \frac{E\left[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}\eta_{\alpha_1}(W)\right]}{E\left[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}\right]}\right) dN(t)\right\} - \lambda J(\beta, \beta_1),$$

and

$$\begin{aligned} & D\mathcal{S}_\lambda(\alpha_0)\alpha_1\alpha_2 \\ &= -E\left\{\int_0^\tau \left(\frac{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}\eta_{\alpha_1}(W)\eta_{\alpha_2}(W)]}{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}]}\right.\right. \\ & \left.\left. - \frac{E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}\eta_{\alpha_1}(W)]E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}\eta_{\alpha_2}(W)]}{\left(E[\mathcal{Y}(t)\exp\{\eta_{\alpha_0}(W)\}]\right)^2}\right) dN(t)\right\} - \lambda J(\beta_1, \beta_2), \end{aligned}$$

where τ is the end of the study. Motivated by Cheng and Shang (2015) and Shang and Cheng (2015), we define the inner product for any $\alpha_i = (\theta_i^\top, \beta_i(\cdot)) \in \mathcal{H}$ $(i = 1, 2)$ as the negative second derivative of $l_\lambda(\alpha)$ at α_0:

$$\langle \alpha_1, \alpha_2 \rangle_\lambda = -D\mathcal{S}_\lambda(\alpha_0)\alpha_1\alpha_2.$$

The corresponding norm is denoted as $||\cdot||_\lambda$. Define

$$S_1^{(k)}(t,\alpha) = \frac{1}{n}\sum_{i=1}^n [\mathcal{Y}_i(t)\exp\{\eta_\alpha(W_i)\}Z_i^{\otimes k}],\ s_1^{(k)}(t,\alpha) = E[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}Z^{\otimes k}], k = 0, 1, 2,$$

$$S_2^{(1)}(t,s,\alpha) = \frac{1}{n}\sum_{i=1}^{n}[\mathcal{Y}_i(t)\exp\{\eta_\alpha(W_i)\}X_i(s)],\ \ s_2^{(1)}(t,s,\alpha) = E[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}X(s)],$$

$$S_2^{(2)}(t,s,v,\alpha) = \frac{1}{n}\sum_{i=1}^{n}[\mathcal{Y}_i(t)\exp\{\eta_\alpha(W_i)\}X_i(s)X_i(v)],$$

$$s_2^{(2)}(t,s,v,\alpha) = E[\mathcal{Y}(t)\exp\{\eta_\alpha(W)\}X(s)X(v)],$$

$$\Sigma = E\left\{\int_0^\tau \frac{s_1^{(2)}(t,\alpha_0)}{s_1^{(0)}(t,\alpha_0)} - \frac{s_1^{(1)}(t,\alpha_0)^{\otimes 2}}{s_1^{(0)}(t,\alpha_0)^2}\,dN(t)\right\},$$

and

$$F(s,t) = \int_0^\tau \mathrm{Cov}\{X(s),X(t)|T=v,\Delta=1\}E[\mathcal{Y}(v)\exp\{\eta_{\alpha_0}(W)\}]h_0(v)\,dv,$$

where

$$\begin{aligned}
&\mathrm{Cov}\{X(s),X(t)|T=v,\Delta=1\}\\
=&E\{X(s)X(t)|T=v,\Delta=1\} - E\{X(s)|T=v,\Delta=1\}E\{X(t)|T=v,\Delta=1\}\\
=&\frac{s_2^{(2)}(v,t,s,\alpha_0)}{s_1^{(0)}(v,\alpha_0)} - \frac{s_2^{(1)}(v,s,\alpha_0)s_2^{(1)}(v,t,\alpha_0)}{s_1^{(0)}(v,\alpha_0)^2}.
\end{aligned}$$

For any $\beta_1,\beta_2 \in \mathcal{H}^{(m)}$, define $\langle\beta_1,\beta_2\rangle_m = \int_{\mathbb{I}}\int_{\mathbb{I}} F(s,t)\beta_1(s)\beta_2(t)\,ds\,dt + \lambda J(\beta_1,\beta_2)$. Clearly, the relationship between $\langle\alpha_1,\alpha_2\rangle_\lambda$ and $\langle\beta_1,\beta_2\rangle_m$ is given by $\langle\beta_1,\beta_2\rangle_m = \langle\alpha_1,\alpha_2\rangle_\lambda$ with $\alpha_1 = (\mathbf{0}^T,\beta_1)$ and $\alpha_2 = (\mathbf{0}^T,\beta_2)$. Then $\mathcal{H}^{(m)}$ is a reproducing kernel Hilbert space (RKHS) with $\langle\cdot,\cdot\rangle_m$. For simplicity, define a linear nonnegative definite and self-adjoint operator W_λ and a bilinear operator $V(\cdot,\cdot)$ in $\mathcal{H}^{(m)}$ as $\langle W_\lambda\beta_1,\beta_2\rangle_m = \lambda J(\beta_1,\beta_2)$ and $V(\beta_1,\beta_2) = \int_{\mathbb{I}}\int_{\mathbb{I}} F(s,t)\beta_1(s)\beta_2(t)\,ds\,dt$, respectively. Then, we have $\langle\beta_1,\beta_2\rangle_m = V(\beta_1,\beta_2) + \langle W_\lambda\beta_1,\beta_2\rangle_m$. Denote the reproducing kernel in $\mathcal{H}^{(m)}$ by $K(s,t)$.

We denote two positive sequences a_n and b_n as $a_n \asymp b_n$ if $\lim_{n\to\infty}(a_n/b_n) = c > 0$. If $c = 1$, we denote $a \sim b$. To establish the theoretical properties of the proposed estimator, we need the following regularity conditions:

(C1) (i) $0 < P(Y \geq \tau) < 1$;

(ii) There exists a constant $c_1 > 0$, for any $\alpha \in \mathcal{H}$, we have

$$E \int_0^\tau \Big[\frac{E\{\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\} \eta_\alpha(W)^2\}}{E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\}]} - \frac{(E\{\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\} \eta_\alpha(W)\})^2}{\big(E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\}]\big)^2} \Big] \, dN(t)$$

$$\geq c_1 E\{\eta_\alpha(W)\}^2.$$

(C2) There exists a sequence of functions $\{h_j\}_{j \geq 1} \subset \mathcal{H}^{(m)}$ such that $||h_j||_{L_2} \leq c_h j^a$ for each $j \geq 1$, some constants $a \geq 0$, $c_h \geq 0$, and

$$V(h_i, h_j) = \delta_{ij}, \quad J(h_i, h_j) = \rho_i \delta_{ij}, \quad \text{for any} \quad i, j \geq 1,$$

where δ_{ij} is the Kronecker's notation, and ρ_i is a nondecreasing nonnegative sequence satisfying $\rho_i \asymp i^{2k}$ for some constant $k > a + 1/2$. Furthermore, any $\beta \in \mathcal{H}^{(m)}$ admits the Fourier expansion $\beta = \sum_{i=1}^\infty V(\beta, h_i) h_i$ with the convergence in $\mathcal{H}^{(m)}$ under $\langle \cdot, \cdot \rangle_m$.

Set the projection of Z on $X(\cdot)$ as $G \equiv (G_1, G_2, \cdots, G_p)^\top$ with

$$\begin{aligned} G_k(\cdot) = &\sum_{j=1}^\infty \int_{\mathbb{I}} E\Big[\int_0^\tau \frac{E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\} Z_k X(u)]}{E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\}]} \\ &- \frac{E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\} Z_k]}{E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\}]} \frac{E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\} X(u)]}{E[\mathcal{Y}(t) \exp\{\eta_{\alpha_0}(W)\}]} \, dN(t) \Big] h_j(u) \, du h_j(\cdot) \\ \equiv &\sum_{j=1}^\infty G_{jk} h_j(\cdot). \end{aligned}$$

(C3) (i) $\Sigma - V(G, G)$ is positive definite;

(ii) There exists $b \in ((1 + 2a)/(2k), 1]$ such that $\sum_j |G_{jk}|^2 \rho_j^b < \infty$ for $k = 1, \ldots, p$.

(C4) There exist constants $s \in (0,1)$ and $M_0 > 0$ such that $E[\exp\{s(\|X\|_{L_2} + \|Z\|_2)\}] < \infty$, and $E\{|\eta_\alpha(W)|^4\} \le M_0\{E|\eta_\alpha(W)|^2\}^2$ for any $\alpha \in \mathcal{H}$.

Remark 3.1 *Condition (C1)(i) is common in survival analysis, such as Condition (D) in Andersen and Gill (1982), and Condition (C2) in Chen et al. (2010). Condition (C1)(ii) holds when $\beta = 0$ under Condition (C3).*

Remark 3.2 *Following Shang and Cheng (2015), we consider the following integro-differential equations:*

$$
\begin{aligned}
(-1)^m y_j^{(2m)}(t) &= \rho_j \int_{\mathbb{I}} F(s,t) y_j(s)\, ds, \\
y_j^{(i)}(0) &= y_j^{(i)}(1) = 0, \qquad\qquad i = m, m+1, \ldots, 2m-1.
\end{aligned}
$$

Let (ρ_j, y_j) be the corresponding eigenvalues and eigenfunctions of the above eigen-system, and let $h_j = y_j/\sqrt{V(y_j, y_j)}$. Then (ρ_j, h_j) satisfies Condition (C2) with $k = m + r + 1$ and $a = r + 1$ if one of the following additional assumptions is satisfied:

(i) $r = 0$;

(ii) $r \ge 1$, *and for any* $i = 0, 1, \ldots, r-1$, $F^{(i,0)}(0,t) = 0$ *for any* $t \in \mathbb{I}$, *where* $F^{(i,0)}(s,t)$ *is the ith order partial derivative with respect to s.*

The relationships among (h_j, ρ_j), $K(\cdot,\cdot)$ and W_λ are given as follows:

$$
K_t(\cdot) = \sum_{j=1}^{\infty} \frac{h_j(t)}{1 + \lambda\rho_j} h_j(\cdot) \qquad \text{and} \qquad (W_\lambda h_j)(\cdot) = \frac{\lambda\rho_j}{1 + \lambda\rho_j} h_j(\cdot).
$$

This can be referred to Shang and Cheng (2015).

Remark 3.3 *From the definition of G, we have $G = \mathbf{0}$ when $X(\cdot)$ and Z are independent. Furthermore, under Condition (C3)(ii), we have that $V(G, W_\lambda G) \to 0$ and $V(W_\lambda G, W_\lambda G) \to 0$ with $\lambda \to 0$. In fact, direct calculations yield*

$$\Sigma - V(G,G) = E\int_0^\tau \Bigg[\{Z - \sum_{k=1}^{\infty} G_k \int_{\mathbb{I}} X(s)h_k(s)\,ds\} - \frac{E[\mathcal{Y}(s)\exp\{\eta_{\alpha_0}(W)\}\{Z - \sum_{k=1}^{\infty} G_k \int_{\mathbb{I}} X(s)h_k(s)\,ds\}]}{E[\mathcal{Y}(s)\exp\{\eta_{\alpha_0}(W)\}]}\Bigg]^{\otimes 2} dN(s).$$

Thus Condition (C3)(i) is similar to Assumption A3 in Cheng and Shang (2015), and is required to guarantee the existence of asymptotic variance of the proposed estimator. Condition (C3)(ii) is the same as that used in Theorem 3.1 of Cheng and Shang (2015) such that $\mathcal{R}_w$, $\mathcal{P}_\lambda$ and $\tilde{\mathcal{R}}_u$ are well defined in the Supplementary Materials.

Remark 3.4 *Condition (C4) on covariates is weaker than the conditions required by Qu et al. (2016).*

In the following, we set $h = \lambda^{1/(2k)}$.

Theorem 3.1 *(Rate of Convergence) Suppose that Conditions (C1)–(C4) hold. If*

$$h = o(1) \quad \text{and} \quad n^{-1/2}h^{-(a+1)-\frac{2k-2a-1}{4m}}\{\log(n)\}^2\{\log\log(n)\}^{1/2} = o(1),$$

then $\hat{\alpha}_{n,\lambda}$ is the unique estimate for α_0 and $\|\hat{\alpha}_{n,\lambda} - \alpha_0\|_\lambda = O_p(r_n)$, where $r_n = (nh)^{-1/2} + h^k$.

This theorem shows that when we choose $\lambda = n^{-(2k)/(2k+1)}$, the estimate enjoys the same order of convergence as that in Qu et al. (2016).

For ease of interpretation, define $\mathcal{S}_n(\alpha)$ and $\mathcal{S}_{n,\lambda}(\alpha)$ be the Fréchet derivatives of $l_n(\alpha)$ and $l_{n,\lambda}(\alpha)$, respectively. A direct calculation yields that the Fréchet derivative of $l_{n,\lambda}(\alpha)$ at the direction of α_1 is

$$\mathcal{S}_{n,\lambda}(\alpha)\alpha_1 = \frac{1}{n}\sum_{i=1}^{n}\Delta_i\left[\eta_{\alpha_1}(W_i) - \frac{\sum_{j=1}^{n}\mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}\eta_{\alpha_1}(W_j)}{\sum_{j=1}^{n}\mathcal{Y}_j(Y_i)\exp\{\eta_\alpha(W_j)\}}\right] - \lambda J(\beta,\beta_1)$$

$$\equiv \mathcal{S}_n(\alpha)\alpha_1 - \lambda J(\beta,\beta_1).$$

Theorem 3.2 *(Joint Bahadur Representation) Suppose that Conditions (C1)–(C4) hold. If*

$$n^{-1/2}h^{-(a+1)-\frac{2k-2a-1}{4m}}\{\log(n)\}^2\{\log\log(n)\}^{1/2} = o(1),$$

$$nh^{2k(1+b)} = o(1), \quad \textit{and} \quad \sum_{j=1}^{\infty}V(\beta_0,h_j)^2\rho_j^2 < \infty,$$

then we have $\|\hat{\alpha}_{n,\lambda} - \alpha_0 - \mathcal{S}_{n,\lambda}(\alpha_0)\|_\lambda = O_p(a_n)$, *where*

$$a_n = n^{-1/2}h^{-(4ma+6m-1)/4m}r_n\{\log\log(n)\}^{1/2}\log(n)^2 + h^{-1/2}r_n^2, \quad \textit{and} \;\; r_n = (nh)^{-1/2} + h^k.$$

Based on the joint Bahadur representation, we can establish the asymptotic joint distribution of the proposed estimators of the slope function and the coefficients.

Theorem 3.3 *(Asymptotic Joint Distribution) Suppose that the conditions of Theorem 3.2 hold. Furthermore, assume that* $\sup_{j\geq 1}\|h_j\|_\infty \leq c_h j^a$, $n^{1/2}a_n h^{-(a+1/2)} = o(1)$, $n^{1/2}h^{k(1+b)} = o(1)$, $\sum_{j=1}^{\infty}V(\beta_0,h_j)^2\rho_j^2 < \infty$, *and* $h^{(2a+1)}\sum_{j=1}^{\infty}\frac{\|h_j(t)\|_\infty^2}{(1+\lambda\rho_j)^2} \asymp \sigma_t^2 > 0$. *Then we have*

$$\begin{bmatrix} \sqrt{n}(\hat{\theta}_{n,\lambda} - \theta_0) \\ \sqrt{nh}h^a\{\hat{\beta}_{n,\lambda}(t) - \beta_0(t)\} \end{bmatrix} \to N(0,\Phi),$$

where

$$\Phi = \begin{bmatrix} \{\Sigma - V(G,G)\}^{-1} & 0 \\ 0 & \sigma_t^2 \end{bmatrix}.$$

Here, Σ can be consistently estimated by

$$\begin{aligned}\hat{\Sigma} =& \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \hat{\text{Var}}(Z|T=t,\Delta=1)\mathcal{Y}_i(t)\exp\{\eta_{\hat{\alpha}}(W_i)\}\,d\hat{\Lambda}_0(t)\right] \\ =& \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \Big[\frac{S_1^{(2)}(t,\hat{\alpha})}{S_1^{(0)}(t,\hat{\alpha})} - \frac{\{S_1^{(1)}(t,\hat{\alpha})\}^{\otimes 2}}{\{S_1^{(0)}(t,\hat{\alpha})\}^2}\Big]\mathcal{Y}_i(t)\exp\{\eta_{\hat{\alpha}}(W_i)\}\,d\hat{\Lambda}_0(t)\right],\end{aligned}$$

where

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{k=1}^n dN_k(s)}{\sum_{j=1}^n \mathcal{Y}_j(s)\exp\{\eta_{\hat{\alpha}}(W_j)\}}.$$

Theorem 3.3 implies that, with certain under-smoothing conditions, the asymptotic bias for the estimation of $\beta_0(t_0)$ vanishes. Hence, applying Theorem 3.3 together with the Delta-method immediately yields the pointwise confidence interval (CI) for some real-valued smooth function of $\beta_0(t)$ at any fixed point $t_0 \in \mathbb{I}$, denoted by $\rho\big\{\beta_0(t_0)\big\}$. Let $\dot{\rho}(\cdot)$ be the first derivative of $\rho(\cdot)$. By Theorem 3.3, for any fixed point $t_0 \in \mathbb{I}$ and $\dot{\rho}\big\{\beta_0(t_0)\big\} \neq 0$, we have

$$P\left(\rho\big\{\beta_0(t_0)\big\} \in \left[\rho\big\{\hat{\beta}_{n,\lambda}(t_0)\big\} \pm z_{\xi/2}\frac{\dot{\rho}\big\{\hat{\beta}_{n,\lambda}(t_0)\big\}\sqrt{\sum_{j=1}^\infty(|h_j(t)|^2/(1+\lambda\rho_j)^2)}}{\sqrt{n}}\right]\right) \to 1-\xi$$

as $n \to \infty$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and z_ξ is the lower ξ-th quantile of $\Phi(\cdot)$, that is $\Phi(z_\xi) = 1-\xi$.

Theorem 3.4 *(Weak Convergence of $\hat{\beta}_{n,\lambda}(\cdot)$) Assume that the conditions in Theorem 3.3 hold. Then $\sqrt{nh}h^a\{\hat{\beta}_{n,\lambda}(s) - \beta_0(s)\}$ converges to a mean zero Gaussian process $\mathcal{G}(s)$ in the Hilbert space $\mathcal{H}^{(m)}$ with the inner product $V(\cdot,\cdot)$, where the covariance for $\mathcal{G}(s)$ at s_1 and s_2 is given by*

$$\Gamma(s_1, s_2) = h^{1+2a} \sum_{j=1}^{\infty} \frac{h_j(s_1)h_j(s_2)}{(1+\lambda\rho_j)^2}.$$

To construct a simultaneous confidence band for $\beta_0(s)$ over a closed subinterval $[\zeta, 1-\zeta] \subseteq \mathbb{I}$, we can employ the resampling method of Lin *et al.* (1993) for the distributional approximation. For illustration, let $(\epsilon_1, \ldots, \epsilon_n)$ be independent standard normal random variables, independent of the data $(Y_i, \Delta_i, W_i), i = 1, \ldots, n$. It can be shown that the distribution of the limiting process $\mathcal{G}(s)$ can be approximated by

$$\hat{\mathcal{G}}(s) \equiv \frac{1}{\sqrt{n}h^{-a-1/2}} \sum_{i=1}^{n} \int_{\mathbb{I}} K_t(s)\, d\tilde{W}_i(t)\epsilon_i,$$

where

$$\tilde{W}_i(s) = \int_0^\tau \left[X_i(s) - \frac{\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}X_j(s)}{\sum_{j=1}^n \mathcal{Y}_j(t)\exp\{\eta_{\alpha_0}(W_j)\}} \right] dM_i(t).$$

In view of this fact, we obtain a large number of realizations of $\hat{\mathcal{G}}(s)$ by repeatedly generating the standard normal random samples $(\epsilon_1, \ldots, \epsilon_n)$ while fixing the data. Thus, one may use the empirical distribution of these random samples to approximate the distribution of $\mathcal{G}(s)$. In particular, the αth-percentile of $\sup_{\zeta \le s \le 1-\zeta} |\mathcal{G}(s)|$ can be approximated by the empirical percentile of a large number of realizations of $\sup_{\zeta \le s \le 1-\zeta} |\hat{\mathcal{G}}(s)|$, denoted by $\hat{\mathcal{G}}_\alpha$. As a result, we can construct the global confidence band for $\beta_0(s)$ with $s \in [\zeta, 1-\zeta]$ as follows:

$$\left(\hat{\beta}_{n,\lambda}(s) - \frac{1}{\sqrt{nh}h^a}\hat{\mathcal{G}}_\alpha, \quad \hat{\beta}_{n,\lambda}(s) + \frac{1}{\sqrt{nh}h^a}\hat{\mathcal{G}}_\alpha \right).$$

4 Penalized Partial Likelihood Ratio Test

In this section, we consider testing the following "global" hypothesis:

$$H_0 : \alpha = \alpha_0 \qquad \text{versus} \qquad H_1 : \alpha \neq \alpha_0,$$

where $\alpha_0 \in \mathcal{H}$. The penalized partial likelihood ratio rest (PLRT) statistic is defined by

$$\text{PLRT}_{n,\lambda} \equiv l_{n,\lambda}(\alpha_0) - l_{n,\lambda}(\hat{\alpha}_{n,\lambda}).$$

Next, we derive the asymptotic null distribution of $\text{PLRT}_{n,\lambda}$.

Theorem 4.1 *(Partial Likelihood Ratio Test) Suppose that Conditions (C1)–(C4) hold. Assume that*

$$nh^{2k(1+b)} = O(1), \quad nh^2 \to \infty, \quad n^{1/2}a_n = o(1), \quad nr_n^3 = o(1), \quad \sum_{j=1}^{\infty} V(\beta_0, h_j)^2 \rho_j^2 < \infty,$$

$$n^{1/2}h^{-\{a+1/2+(2k-2a-1)/(4m)\}} r_n^2 \{\log(n)\}^2 \{\log\log(n)\}^{1/2} = o(1),$$

and

$$n^{1/2}h^{-\{2a+1+(2k-2a-1)/(4m)\}} r_n^3 \{\log(n)\}^3 \{\log\log(n)\}^{1/2} = o(1).$$

Then under H_0, we have

$$(2\nu_\lambda)^{-1/2}(-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda \|W_\lambda \beta_0\|_m^2 - \nu_\lambda) \xrightarrow{d} N(0,1),$$

where $\sigma_\lambda^2 \equiv \sum_{j=1}^{\infty} h/(1+\lambda\rho_j)$, $\rho_\lambda^2 \equiv \sum_{j=1}^{\infty} h/(1+\lambda\rho_j)^2$, $\gamma_\lambda \equiv \sigma_\lambda^2/\rho_\lambda^2$, and $\nu_\lambda \equiv h^{-1}\sigma_\lambda^4/\rho_\lambda^2$.

It follows from Theorem 3.3 that $n\|W_\lambda \beta_0\|_m^2 = o(n\lambda) = o(\nu_\lambda)$. Hence, we have $-2n\gamma_\lambda PLRT_{n,\lambda} \sim N(\nu_\lambda, 2\nu_\lambda)$, which is nearly $\chi^2_{\nu_\lambda}$ as $n \to \infty$. This shows that PLRT enjoys the Wilks phenomenon.

Remark 4.1 *(Composite Hypothesis) Similar to Remark 5.1 of Shang and Cheng (2015), we can deal with some composite hypothesis testing. By examining the proof of Theorem 4.1, we can find that the asymptotic null distribution derived therein remains the same even when the hypothesized value* α_0 *is unknown. An important consequence is that the proposed likelihood ratio approach can also be used to test a composite hypothesis such as*

$$H_0 : \theta = \theta_0 \text{ and } \beta \in \mathcal{P}_d,$$

where $\mathcal{P}_d = \{\beta(t) : \beta(t) = \sum_{j=0}^{d} t^j b_j\}$*. Using the similar arguments used in Remark 5.1 of Shang and Chang (2015), we can conclude that the asymptotic null distribution for testing such composite hypothesis follows* $\chi^2_{\nu_\lambda}$*, which is the same as that given in Theorem 4.1.*

To conclude this section, we show that the PLRT achieves the optimal minimax rate of testing given by Ingster (1993) based on the uniform version of the joint Bahadur representation. To this end, we consider the alternative hypothesis $H_{1n} : \alpha = \alpha_{n_0}$, where $\alpha_{n_0} = \alpha_0 + \alpha_n$, $\alpha_0 \in \mathcal{H}$ and α_n belongs to the alternative value set $\mathcal{A} = \{\alpha \in \mathcal{H}, \|\theta\|_2 \leq \zeta, \|\beta\|_{L^2} \leq \zeta, J(\beta,\beta) \leq \zeta\}$ for some constant $\zeta > 0$.

Theorem 4.2 *Suppose that the conditions of Theorem 4.1 hold, and under* $H_{1n} : \alpha = \alpha_{n_0}$, $\|\hat{\alpha}_{n,\lambda} - \alpha_{n_0}\|_\lambda = O_p\{(nh)^{-1/2} + h^k\}$ *holds uniformly over* $\alpha_{n_0} \in \mathcal{A}$*. If* $nh^{3/2+a/2} \to \infty$ *as* $n \to \infty$*, then, for any* $\delta \in (0,1)$*, there always exist positive constants* b_0 *and* N *such that*

$$\inf_{n \geq N} \inf_{\alpha_n \in \mathcal{A}, \|\alpha_n\|_\lambda \geq b_0 \eta_n} P(\text{reject} \quad H_0 | H_{1n} \text{ is } \text{ true}) \geq 1 - \delta,$$

where $\eta_n \geq \sqrt{h^{2k} + (nh^{1/2})^{-1}}$*. Moreover, the minimal lower bound of* η_n *is* $n^{-2k/(4k+1)}$*, which can be achieved when* $h = h^{**} = n^{-2/(4k+1)}$*.*

5 Simulation Studies

In this section, we conduct simulation studies to assess the finite-sample performance of the estimated confidence interval given in Section 3 and the PLRT developed in Section 4.

We use a setup similar to that in Qu et al. (2016). The functional covariate $X(\cdot)$ is defined as

$$X(s) = \sum_{k=1}^{50} \xi_k U_k \phi_k(s),$$

where U_k are independently sampled from the uniform distribution on $[-3, 3]$, $\xi_k = (-1)^{k+1} k^{-1/2}$, $\phi_1 = 1$, and $\phi_{k+1}(s) = \sqrt{2}\cos(k\pi s)$ for $k \geq 1$.

We set $\beta_0(t) = 9/(50 - 45t) - 0.9$, which is from a Sobolov space $\mathcal{H}^{(2)}(\mathbb{I})$. The penalty function is $J(\beta, \beta) = \int_{\mathbb{I}} \{\beta^{(2)}(t)\}^2 \, dt$. The scalar covariate Z is set to be univariate with distribution $N(0, 1)$ and the corresponding coefficient θ_0 to be 1. The failure time T is generated from the functional Cox model:

$$h(t|W) = h_0(t) \exp\left\{\theta_0^\top Z + \int_0^1 X(s)\beta_0(s)ds\right\},$$

where $h_0(t) = t^2$. Given W, the failure time T follows a Weibull distribution. The censoring time C is generated independently, following an exponential distribution with parameter γ which controls the censoring rate. Here, $\gamma = 15$ and 3.9 result in censoring rates around 12% and 33%, respectively. We consider the sample sizes $n = 200$ and 400. We adopt the cubic spline functions for the estimation of the functional coefficient. The number of knots is at the order of $q_n = [2n^{1/5}]$, and the knots are equally spaced. The order m of Sobolev space is 2.

The proposed estimation and testing procedures are implemented in MATLAB programming language. In particular, the eigenvalues and eigenfunctions are solved using `eigs` function from `Chebfun` package (version 5), an open source software package. For details, one can refer to Driscoll, Bornemann and Trefethen (2008). For each combination of censoring rate and n, the simulation is repeated 1000 times.

For the determination of the optimal tuning parameter, we use the cross-validated log partial likelihood method (CVL) (Verweij and Houwelingen, 1993). Let $\hat{\alpha}^{\lambda}_{(-i)}$ be the value of α that maximizes $l_{\lambda,(-i)}$, the penalized log partial likelihood when observation i is omitted. Given a value of λ, the CVL is given by $CVL_{\lambda} = \sum_{i=1}^{n} l_{\lambda,i}(\alpha^{\lambda}_{(-i)})$, where $l_{\lambda,i}(\cdot) = l_{\lambda}(\cdot) - l_{\lambda,(-i)}(\cdot)$ is the contribution of observation i to the penalized log partial likelihood. Using the CVL, we find that the optimal tuning parameter is about 10^{-6} based on 10 Monte Carlo trials. The tuning parameter, therefore, is chosen to be $\lambda = 10^{-6}$ to reduce the computation time in our simulations. One may also consider less computationally intensive methods such as AIC (Gellar et al., 2015) and GCV (Qu et al., 2016).

Figure 1 displays an instance of estimated $\beta_0(\cdot)$ and the pointwise 95% confidence intervals among 1000 simulations. The pointwise average of the estimated $\beta_0(\cdot)$ and the empirical coverage probability of the 95% pointwise confidence interval based on 1000 simulations are shown in Figures 2 and 3, respectively. Table 1 reports the bias (BIAS), the sample standard error of the estimates (SSE), the average of the estimated standard errors (ESE), and the empirical coverage probability (CP) at $t = 0.1, 0.5, 0.9$. The simulation results are consistent with Theorem 3.3. It is apparent that when n increases from 200 to 400 with the censoring rate fixed, the average bias and the standard error decrease steadily. In particular, these results suggest that the estimator of $\beta_0(\cdot)$ is consistent. Furthermore, the empirical coverage probability also approaches to the nominal level 95%. The average ESE at 12% censoring rate is lower in comparison to that at 33% censoring rate. This is in line with the expectation that the lower the censoring rate is, the more accurate the estimate becomes.

For the regression coefficient of the scalar covariate, the BIAS, SSE, ESE, and CP of the estimated $\hat{\theta}_{n,\lambda}$ are given in Table 2 for each setting of censoring rate and sample size over 1000 repetitions. As the sample size increases, the average of $\hat{\theta}_{n,\lambda}$ approaches to the true value, the standard deviation reduces, and the coverage probability approaches to 95% given a fixed censoring rate. Similarly, we observe these trends as the censoring rate reduces for a given sample size.

In summary, the simulation results in Tables 1 and 2 indicate that the estimates of both scalar and functional parameters are consistent and the proposed variance estimation procedure provides reasonable estimation of variances. Also the results on the empirical coverage probability suggest that the normal approximation is appropriate.

To study the performance of the penalized partial likelihood ratio test, we calculate the estimated sizes and powers of the PLRT under $H_0 : \alpha = (\theta_0, \beta_0(\cdot))$, that is, the percentage of rejecting H_0. We consider α under different signal strengths. Specifically, $\alpha = (\theta_0 + c, \beta_0(\cdot) + c)$, where $c = 0.0, 0.1, 0.3, 0.5$. Table 3 summarizes the percentages of rejecting H_0 over 1,000 simulations. These results demonstrate the good performance of the PLRT. The power of the test increases as sample size n increases, and the power slightly decreases as the censoring rate increases.

6 An Application

In this section, we apply the proposed method to the Sequential Organ Failure Assessment (SOFA) data obtained from the Improving Care of Acute Lung Injury Patients (ICAP) study (Needham et al., 2006; Gellar et al., 2014, 2015). The primary goal of this prospective cohort study is to investigate the long-term complications of patients who suffer from acute lung injury/acute respiratory distress syndrome (ALI/ARDS).

The ICAP study involves 520 subjects, with 237 (46%) dying in the intensive care unit (ICU). Out of the 520 subjects, 161 subjects (31.0%) died within the first week in ICU, and they are excluded from the analysis. Therefore, the proposed method is applied to the remaining 359 subjects. In the ICAP study, data were recorded once the patients were admitted in the ICU, and then daily during hospitalization. The SOFA score is one of the measurements recorded daily. SOFA is a measure of the overall organ function status of a patient. It is composed of respiratory, cardiovascular, coagulation, liver, renal, and neurological components. The score of each component ranges from 0 to 4, with higher scores suggesting inferior organ function. The SOFA score, ranging from 0 to 24, is then the sum of these six scores. We treat the history of each subject's SOFA scores, in the first week, as a functional covariate, $X(s)$, where s is the number of days since the admission to the ICU. Trajectories of the SOFA score of subjects who died after the first week of ICU hospitalization and those who survived are depicted in Figure 4. It is apparent that among patients who manage to survive, the pointwise averages of SOFA scores are declining, whereas among patients who died after the first week of ICU hospitalization, the averages are relatively stable. Our model includes three scalar covariates as controls of a subject's baseline risk. They are age, gender, and Charlson co-morbidity index (Charlson et al., 1987).

Our goal is to estimate the association between the trajectories of SOFA score and mortality among subjects who were hospitalized in ICU for more than a week. We adopt the cubic spline functions for the estimation of the functional parameter. The number of knots is at the order of $q_n = [2n^{1/5}] = 7$, and the knots are equally spaced. The $\lambda = 10^{-3}$ leads to the optimal penalty according to CVL.

We plot the estimated coefficient function $\hat{\beta}_{n,\lambda}(\cdot)$ in Figure 5. The result suggests that there is a functional association between time to death during the ICU stay and the SOFA score function for $t \in [0.75, 1]$, which corresponds to the sixth and the seventh day of ICU stay. This implies that the SOFA score in last two days in the first week of ICU stay may be used as an indicator of the one's hazard.

Table 4 summarizes the estimation of the regression coefficients of the scalar covariates. In addition to the functional covariate, patients' age and Charlson co-morbidity index seem to have positive effects on the hazard. On the other hand, the gender shows no significant association with the hazard of death.

7 Concluding Remarks

This article focuses on the development of semiparametric inference for the functional Cox model with right-censored data. We have proposed a penalized partial likelihood approach for the estimation of model parameters and established the asymptotic properties including the consistency, the convergence rate, and the limiting distribution of the proposed estimators. In particular, since the overall convergence rate of the proposed estimators cannot achieve the standard rate $n^{-1/2}$, deriving the asymptotic joint distribution of the functional and scalar estimators becomes more difficult. To overcome the difficulty, we have investigated the joint Bahadur representation of finite-dimensional and infinite-dimensional estimators in the Sobolev space equipped with a proper inner product. There are two significant contributions made to the study of the functional Cox model. One is that the asymptotic joint normality of the estimators of the slope function and coefficients has been obtained; another is that the partial likelihood ratio test with the Wilks phenomenon and the optimality has been developed. These two important issues have not been addressed in the previous research. Our new results will provide more insights and deeper understanding about effects of functional predictors on the hazard function of failure time. Our simulation studies demonstrate that the proposed estimation approach performs well and the penalized partial likelihood ratio test has a good power.

Note that the definition of the inner product plays a key role for deriving the theoretical properties. Based on the specific inner product, we can establish Lemma B.1 and then derive the joint Bahadur representation of the proposed estimators. Clearly, the asymptotic joint distribution of the proposed estimator in Theorem 3.3 and the asymptotic null distribution of the proposed partial likelihood ratio test statistic in Theorem 4.1 as well as the optimality of the test in Theorem 4.2 do not reply on the definition of the inner product. Some other inner products satisfying Lemma B.1 may be used to derive the joint Bahadur representation in Theorem 3.2. This needs to be further explored.

The proposed approach can be extended to making inference for the following nonparametric Cox's proportional hazards models in Chen et al. (2010):

$$h(t|X) = \exp\{g_0(X)\}h_0(t),$$

where $h(t|X)$ is the hazard function, $h_0(t)$ is the baseline hazard function, and $g_0(\cdot)$ is an unknown function. Denote $l_n(g)$ and $l_{n,\lambda}(g)$ as the partial likelihood and the penalized partial likelihood, respectively. Then we have

$$\begin{aligned} l_{n,\lambda}(g) \equiv & l_n(g) - \frac{\lambda}{2} J(g,g) \\ = & \frac{1}{n}\sum_{i=1}^{n} \Delta_i \left[g(X_i) - \log\left\{ \frac{1}{n}\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp(g(X_j)) \right\} \right] - \frac{\lambda}{2} J(g,g). \end{aligned}$$

Note that the first and second Fréchet derivatives of $l_{n,\lambda}(g)$ at the direction of g_1 and g_2 are given by

$$\mathcal{S}_{n,\lambda}(g)g_1 = \frac{1}{n}\sum_{i=1}^{n} \Delta_i \left[g_1(X_i) - \frac{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{g(X_j)\}g_1(X_j)}{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{g(X_j)\}} \right] - \lambda J(g_1, g),$$

$$\begin{aligned} D\mathcal{S}_{n,\lambda}(g)g_1g_2 = & -\frac{1}{n}\sum_{i=1}^{n} \Delta_i \left[\frac{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{g(X_j)\}g_1(X_j)g_2(X_j)}{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{g(X_j)\}} \right. \\ & \left. - \frac{\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{g(X_j)\}g_1(X_j)\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{g(X_j)\}g_2(X_j)}{(\sum_{j=1}^{n} \mathcal{Y}_j(Y_i)\exp\{g(X_j)\})^2} \right] - \lambda J(g_1, g_2). \end{aligned}$$

Assume that $g_0(\cdot)$ belongs to the mth-order Sobolev space $\mathcal{H}^{(m)}$. Define an inner product in the space as

$$\begin{aligned}\langle g_1, g_2\rangle_\lambda = & E\int_0^\tau \left[\frac{E[\mathcal{Y}(t)\exp\{g_0(X)\}g_1(X)g_2(X)]}{E[\mathcal{Y}(t)\exp\{g_0(X)\}]}\right. \\ & \left. - \frac{E[\mathcal{Y}(t)\exp\{g_0(X)\}g_1(X)]E[\mathcal{Y}(t)\exp\{g_0(X)\}g_2(X)]}{(E\mathcal{Y}(t)\exp\{g_0(X)\})^2}\right] dN(t) + \lambda J(g_1, g_2).\end{aligned}$$

Thus $\mathcal{H}^{(m)}$ is a reproducing kernel Hilbert space. Following this step, our method can be used to handle the model proposed in Chen *et al.* (2010).

A further interesting research is to explore other useful functional models such as functional accelerated failure time models and functional additive hazards models with right-censored data, where a partial likelihood is unavailable.

Supplementary Materials

The Supplementary Materials include the proofs of lemmas and theorems.

Acknowledgement

The authors would like to thank the Editor, the Associate Editor and the two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. Hao's research is partly supported by the Fundamental Research Funds for the Central Universities (No. CXTD10-09) and the National Natural Science Foundation of China (No. 11901087). Xu's research is partly supported by the Canadian Institutes of Health Research (CIHR, Grant No. 145546). Zhao's research is partly supported by the Research Grant

Council of Hong Kong (15301218), the National Natural Science Foundation of China (No. 11771366), and The Hong Kong Polytechnic University.

References

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* **10**, 1100-1120.

Cai, T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association* **107**, 1201-1216.

Charlson, M. E., Pompei, P., Ales, K. L. and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases and Management* **40**, 373-383.

Chen, K., Chen, K., Müller, H. and Wang, J. L. (2011). Stringing high-dimensional data for functional analysis. *Journal of the American Statistical Association* **106**, 275-284.

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**, 750-760.

Cheng, G. and Shang, Z. (2015). Joint asymptotics for semi-nonparametric regression models with partially linear structure. *The Annals of Statistics* **43**, 1351-1390.

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

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

Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics* **37**, 35-72.

Driscoll, T., Bornemann, F. and Trefethen, L. (2008). The chebop system for automatic solution of differential equations. *BIT Numerical Mathematics* **48**, 701-723.

Gellar, J., Colantuoni, E., Needham, D. and Crainiceanu, C. (2014). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association* **109**, 1425-1439.

Gellar, J., Colantuoni, E., Needham, D. and Crainiceanu, C. (2015). Cox regression models with functional covariates for survival data. *Statistical Modelling* **15**, 256-278.

Ingster, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives $I-III$. *The journal Mathematical Methods of Statistics* **2**, 85-114; **3**, 171-189; **4**, 249-268.

Jacobsen, M. (1984). Maximum likelihood estimation in the multiplicative intensity model: a survey. *International Statistical Review* **52**, 193-207.

Johansen, S. (1983). An extension of Cox's regression model. *International Statistical Review* **51**, 165-174.

Kong, D., Ibrahim, J. G., Lee, E. and Zhu, H. (2018). FLCRM: functional linear cox regression model. *Biometrics* **74**, 109-117.

Li, K. and Luo, S. (2017). Functional joint model for longitudinal and time-to-event data: an application to alzheimer's disease. *Statistics in Medicine* **36**, 3560-3572.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-biased residuals. *Biometrika* **85**, 605–619.

Needham, D. M., Dennison, C. R., Dowdy, D. W., Mendez-Tellez, P. a., Ciesla, N., Desai, S. V., Sevransky, J., Shanholtz, C., Scharfstein, D., Herridge, M. S. and Pronovost, P. J. (2006). Study Protocol: The Improving Care of Acute Lung Injury Patients (ICAP) study. *Critical Care* **10**, R9.

Qu, S., Wang, J. L. and Wang, X. (2016). Optimal estimation for the functional Cox model. *The Annals of Statistics* **44**, 1708-1738.

Ramsay, S. and Silverman, B. W. (2005). Functional data analysis (2nd ed., Springer series in statistics). New York ; Berlin: Springer.

Shang, Z. and Cheng, G. (2015). Nonparametric inference in generalized functional linear models. *The Annals of Statistics* **43**, 1742-1773.

Tsiatis, A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics* **9**, 93-108.

Verweij, P. and Van Houwelingen, H. (1993). Cross-validation in survival analysis. *Statistics in Medicine* **12**, 2305-2314.

Yuan, M. T. and Cai, T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics* **38**, 3412-3444.

Table 1: Simulation results for the proposed estimate of $\beta_0(t)$ at $t = 0.1, 0.5, 0.9$.

	$n = 200$			$n = 400$		
	0.1	0.5	0.9	0.1	0.5	0.9
12% BIAS	-0.0504	-0.0431	-0.0747	-0.0189	-0.0218	-0.0400
SSE	0.1518	0.1372	0.1751	0.1042	0.1088	0.1223
ESE	0.1927	0.1602	0.2156	0.1343	0.1117	0.1501
CP	0.9750	0.9680	0.9740	0.9840	0.9510	0.9820
33% BIAS	-0.0539	-0.0514	-0.0914	-0.0241	-0.0270	-0.0531
SSE	0.1704	0.1578	0.1919	0.1245	0.1269	0.1419
ESE	0.1999	0.1658	0.2238	0.1391	0.1158	0.1558
CP	0.9750	0.9510	0.9560	0.9690	0.9200	0.9570

Table 2: Simulation results for the proposed estimate of θ_0.

		$n = 200$	$n = 400$
12%	BIAS	0.0339	0.0154
	SSE	0.1070	0.0717
	ESE	0.1170	0.0811
	CP	0.9520	0.9690
33%	BIAS	0.0392	0.0212
	SSE	0.1261	0.0811
	ESE	0.1224	0.0849
	CP	0.9240	0.9500

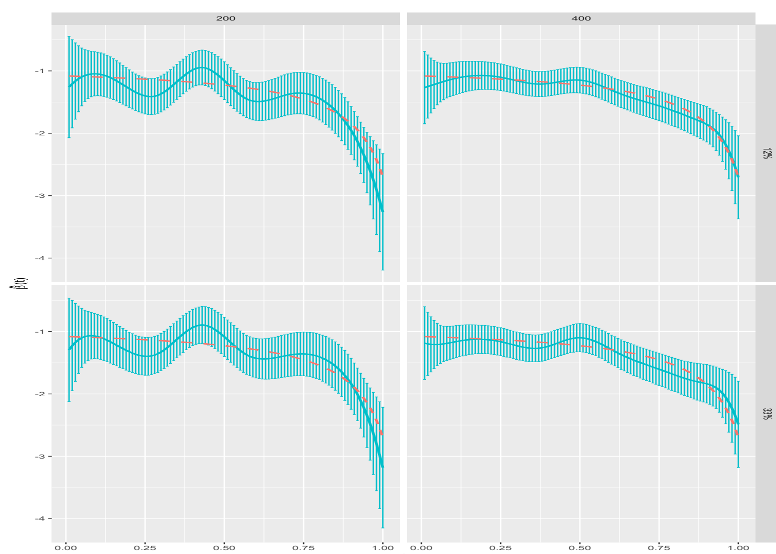

Figure 1: Graphical displays of $\hat{\beta}_{n,\lambda}(\cdot)$ and the pointwise 95% confidence intervals of $\beta_0(t)$. The dashed lines represent $\beta_0(\cdot)$ whereas the solid lines represent $\hat{\beta}_{n,\lambda}(\cdot)$.

Table 3: The simulated sizes and powers of the likelihood ratio test for $H_0 : \alpha = (\theta_0^\top, \beta_0(\cdot))$.

	c	200	400
12%	0.0	0.0510	0.0410
	0.1	0.2320	0.5680
	0.3	1.0000	1.0000
	0.5	1.0000	1.0000
33%	0.0	0.0510	0.0490
	0.1	0.1610	0.4570
	0.3	0.9950	1.0000
	0.5	1.0000	1.0000

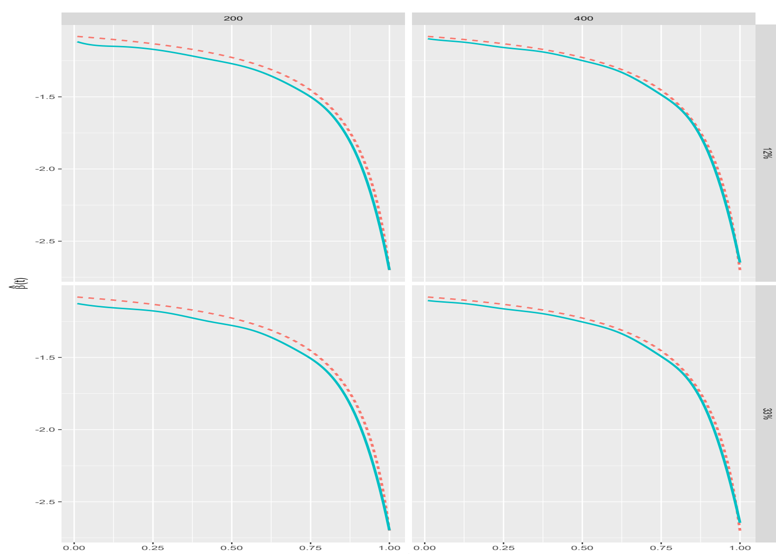

Figure 2: Graphical displays of the pointwise averages $\hat{\beta}_{n,\lambda}(\cdot)$. The dashed lines represent $\beta_0(\cdot)$ whereas the solid lines represent the pointwise averages of $\hat{\beta}_{n,\lambda}(\cdot)$.

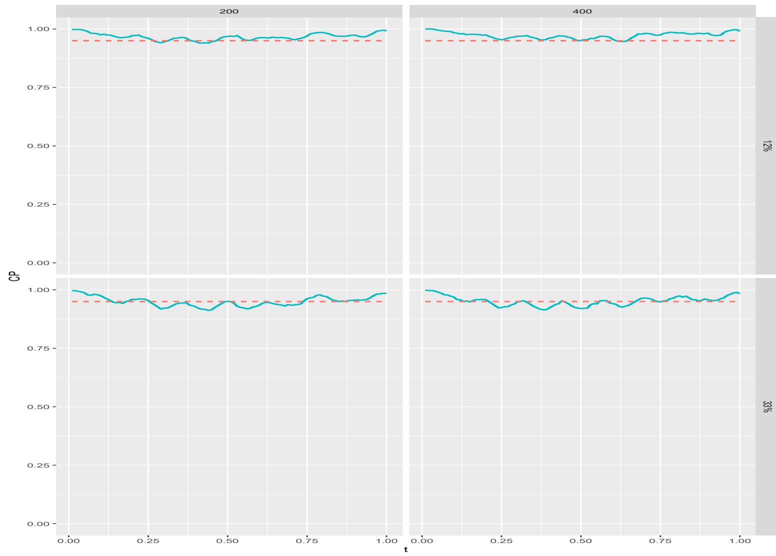

Figure 3: Graphical displays of the pointwise coverage probabilities (CP). The dashed lines represent 95% whereas the solid lines represent the pointwise CP of $\beta_0(\cdot)$.

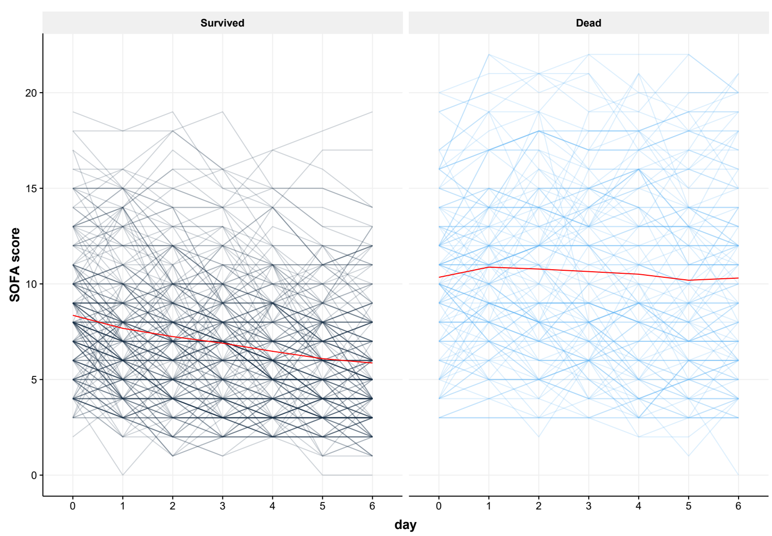

Figure 4: Trajectories of the SOFA score of subjects who died after the first week of the ICU hospitalization and those who survived. The red lines are the pointwise average of the SOFA score.

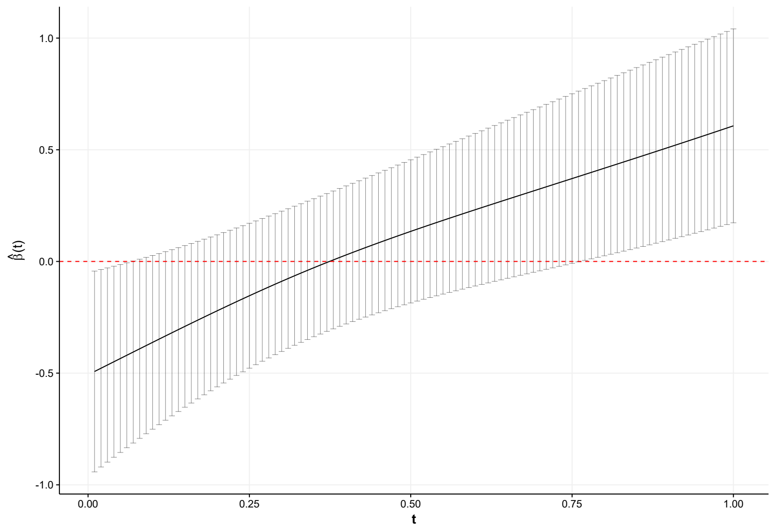

Figure 5: The estimated coefficient function $\hat{\beta}_{n,\lambda}(\cdot)$ and the pointwise 95% confidence interval for the SOFA data analysis.

Table 4: Estimation results of regression coefficients of scalar covariates for the SOFA data analysis

	$\hat{\theta}_{n,\lambda}$	$S.E.$	t-value
Age	0.0151	0.0015	10.0667
Gender (male=1)	0.1640	0.1331	1.2322
Charlson Index	-0.0348	0.0034	-10.2353