Statistica Sinica Preprint No: SS-2024-0180	
Title	Semiparametric Inference for Functional Survival Models
Manuscript ID	SS-2024-0180
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0180
Complete List of Authors	Hongyi Zhou, Wenqing Su, Qixian Zhong and Ying Yang
Corresponding Authors	Hongyi Zhou
E-mails	zhouhongyi99@126.com
Notice: Accepted author version.	

Statistica Sinica

Semiparametric Inference for Functional Survival Models

Hongyi Zhou, Wenqing Su, Qixian Zhong†, Ying Yang†

Tsinghua University, Xiamen Universtiy

Abstract: **In recent years, there has been growing interest among researchers in modeling right-censored survival data with functional covariates. While existing functional methods primarily focus on the Cox model, its proportional hazards assumption can be challenging to verify and may be violated in practice. To address this issue, we extend the ordinary differential equation (ODE) framework for survival data to incorporate functional covariates and develop an inference procedure for both scalar and functional parameters. Specifically, we establish asymptotic normality and semiparametric efficiency for the scalar coefficient estimators, enabling a valid inference procedure. Additionally, we derive an asymptotic simultaneous confidence band for the functional parameter. Simulations are conducted to evaluate the finite sample performance of the proposed method.**

Key words and phrases: Functional data analysis; Right censored data; Sieve maximum likelihood estimator; Semiparametric information bound; Transformation model.

† Co-corresponding authors. qxzhong@xmu.edu.cn, yangying@tsinghua.edu.cn

1. Introduction

Survival data analysis plays a crucial role in various fields, such as biomedical science and reliability engineering, where it is used to investigate the time until a specific event occurs, such as patient death in clinical applications or component failure in industrial settings. One major challenge in analyzing survival data arises from censoring, which occurs when the exact event times are not fully observed. Such censored time data are commonly encountered in longitudinal or periodic follow-up studies, such as clinical trials. To address such challenge, various statistical methods have been developed, including semiparametric models (Cox, 1972; Kalbfleisch and Prentice, 2002; Zhong et al., 2022) and nonparametric models (Kaplan and Meier, 1958; Aalen, 1978). Among these, the Cox proportional hazard model (Cox, 1972) is the most widely used and popular approach in survival analysis. For more details, we refer readers to monographs Cox and Oakes (1984); Fleming and Harrington (1991).

Recently, Tang et al. (2022) introduces an ordinary differential equation (ODE) notion for survival data which unifies many existing survival models, where the cumulative function of event time T for some subject X, denoted as $\Lambda_X(t)$, satisfies the following ordinary differential equation

$$\begin{cases} \Lambda'_X(t) = h_0(t) \exp\left(\alpha_0^T X\right) q_0(\Lambda_X(t)), \\ \Lambda_X(0) = 0, \end{cases} \tag{1.1}$$

where $\Lambda'_X(t)$ is the derivative of $\Lambda_X(t)$ with respect to argument t, α_0 is a d-dimensional parameter, $h_0(\cdot)$, $q_0(\cdot)$ are two unspecified positive functions and X is a d-dimensional covariate vector. Adopting such an ODE notion not only enables a fast and straightforward estimation and inference procedure, but also provides clear interpretation, since model (1.1) directly characterizes the evolution of hazard function over time.

In reality, covariates are not necessarily limited to scalar types. With the rapid advancement of data collection and storage techniques, functional data has achieved great attention among researchers in the field of survival analysis in recent years. Take our real data application as an example, in the Improving Care of Acute Lung Injury Patients study (Needham et al., 2006), patient data is collected at different stages during a follow-up study, beginning at baseline when patients enroll in the study. Data is then collected daily while they remain in the Intensive Care Unit until hospital discharge or death. In addition to baseline covariates such as age and gender, researchers also record each patient's overall organ failure status, measured by the sequential organ failure assessment (SOFA) score, daily during their stay in the Intensive Care Unit. The SOFA score is a type of functional data (Yao et al., 2005; Li and Hsing, 2010; Zhang and Wang, 2016; Wang et al., 2016). According to previous research (Gellar et al., 2014), these organ failure status measurements are believed to have an intrinsic impact on the patients' survival time. To accommodate such survival data with functional covariates, the Cox model has first been extended by researchers such as Chen et al. (2011) and Gellar et al. (2015). Kong et al. (2018) studied the functional Cox regression model that incorporates functional principal component analysis approach to extract functional features from surface data and derived some asymptotic properties. Qu et al. (2016) further investigated the optimality property including semiparametric efficiency under a reproducing kernel Hilbert space framework. Hao et al. (2021) derived the joint asymptotic distribution of both finite-dimensional and infinite-dimensional estimators for the functional Cox model and studied a partial likelihood ratio test. The functional Cox model is also extended to accommodate various scenarios. Shi et al. (2022) adapted the functional Cox model to interval censored data and Spreafico and Ieva (2021); Spreafico et al. (2023) considered a similar model with multiple functional predictors and recurrent events. Yang et al. (2021) studied a weighted functional Cox model under a somewhat different settings. Jiang et al. (2020) studied the relationship between functional covariates and time-to-event outcome through a quantile regression model. Other semiparametric models are also extended to analyze functional survival data. Cui et al. (2021) proposed the additive functional Cox model. Clearly, there is a growing interest in incorporating functional data into survival analysis models.

However, a significant limitation of the aforementioned methods is their heavy reliance on the proportional hazard assumption, which presumes that covariates have a parallel effect on the conditional log hazard functions. This assumption is often violated in real-world applications (Aalen and Gjessing, 2001). Jiang et al. (2020) demonstrated that when the proportional hazard assumption is violated, the functional Cox model can lead to misinterpretation of results. To alleviate these limitations, one may consider extending other semiparametric models such as the accelerated failure time model (AFT) (Ritov, 1990; Wei, 1992; Kalbfleisch and Prentice, 2002), the proportional odds model (Bennett, 1983; Murphy et al., 1997), or the linear transformation model (Cheng et al., 1995; Chen et al., 2002; Zeng and Lin, 2007a) to accommodate functional data. Among these, the AFT model has gained popularity due to its straightforward interpretation and improved stability in accounting for unobserved features. A key advantage of AFT models is that they directly assume a regression relationship between the transformed event time in relation to the covariates. However, it is important to note that the AFT model may not accurately capture the underlying distribution of survival times, posing a challenge for practitioners in selecting the most appropriate model for their specific applications.

Hence, in this paper, we aim to extend model (1.1) to accommodate functional covariates and develop appropriate estimation and inference procedure. Specifically, we assume the conditional cumulative function with a vector covariate X and a functional covariate $Z(\cdot)$, denoted as $\Lambda_{X,Z}(t)$, satisfies the following ordinary differential equation

$$\begin{cases} \Lambda'_{X,Z}(t) = h_0(t)\exp\left(\alpha_0^T X + \int_K \beta_0(s)Z(s)ds\right) q_0(\Lambda_{X,Z}(t)), \\ \Lambda_{X,Z}(0) = 0, \end{cases} \tag{1.2}$$

where $\beta_0(\cdot)$ is an unknown coefficient function and K is a compact subset of $\mathbb{R}$. Obviously, model (1.2) degenerate to model (1.1) when $\beta_0 \equiv 0$. It is also worth noting that model (1.2) is general enough to cover many existing functional model. For instance, if $q(\cdot) \equiv 1$, the model (1.2) reduces to functional Cox model (Qu et al., 2016; Kong et al., 2018; Hao et al., 2021). If $h(\cdot) \equiv 1$, the model (1.2) becomes a functional version of accelerated failure time model. Additionally, the model (1.2) includes a functional version of transformation model in Zeng and Lin (2007b), that is,

$$\varphi(T) = -\alpha_0^T X - \int_K \beta_0(s)Z(s)ds + \varepsilon, \tag{1.3}$$

where $\varphi(\cdot)$ is a monotone function and ε is an unobserved random variable independent of X and $Z(\cdot)$. The relationship between the model (1.2) and (1.3) is given by $\int_0^t h(u)du = \exp(\varphi(t))$, $\int_0^{-\ln t} 1/q(u)du = G^{-1}(t)$ and $G(t)$ is the survival function of $\exp(\varepsilon)$. It is worth-noting that the entire trajectory of Z has an overall effect on the survival time T, which is different from the transformation model with time-varying effect in Zeng and Lin (2007b), where the hazard function at time t is only affected by the current value of Z at time t.

Additionally, we propose a sieve maximum likelihood estimator that operates within a spline-based sieve space and provide theoretical justification for the large sample properties of our estimators. The main challenge involved in deriving the estimators and establishing their asymptotic properties is the fact that the functional parameter poses theoretical challenges due to the different convergence rates of the estimators for the scalar coefficient α_0 and functional coefficient $\beta_0(\cdot)$. The rate of convergence for the functional parameter not only depends on the sample size, but also on the choice of B-spline basis. Consequently, the method by Ding and Nan (2011) and Tang et al. (2022) can not be applied for deriving the asymptotic distribution of functional parameters. We further establish the asymptotic distribution of the functional parameter and construct both pointwise confidence intervals and simultaneous confidence bands, which have not yet been addressed by Tang et al. (2022). These contributions represent significant advancements of this paper compared to their work.

Overcoming the aforementioned challenges, we proved the convergence rate of both the finite-dimensional and infinite-dimensional parameters within the framework of ordinary differential equations. This theoretical analysis provides valuable insights into the performance of our estimators. We also derive the information bound for the finite-dimensional parameter and demonstrate that our proposed estimators asymptotically achieve this information bound and thus are semiparametric efficient. Furthermore, we derive the asymptotic distribution for the infinite-dimensional parameter, facilitating the construction of both pointwise and simultaneous confidence bands.

The rest of the paper is organized as follows. In Section 2, we discuss the model identifiability and introduce the estimation approach. Section 3 states the regularity conditions and develop the asymptotic properties, including consistency and semiparametric efficiency, of the estimators. Section 4 presents simulation results to evaluate finite sample properties of the estimators. Section 5 illustrates an application of proposed method to a sequential organ failure assessment data. In Section 6, we make some concluding remarks and discuss several topics for future research. Technical proofs are relegated to the Supplementary material.

2. Methodology

2.1 Identifiability

We first address the issue of model identification. Note that the equation (1.3) still holds if φ, ε are replaced by $\varphi + c_1$ and $\varepsilon + c_1$ for some constant c_1, or $\varphi, \alpha, \beta, \varepsilon$ are replaced by $c_2\varphi, c_2\alpha, c_2\beta, c_2\varepsilon$ for some constant c_2. Therefore, a location and scale normalization is essential for identifiability. Follow the discussion on page 169 of Horowitz (1996), the model parameters are identifiable up to a location and scale normalization if X has at least one continuously distributed component with a non-zero coefficient. Correspondingly, we provide the sufficient conditions for identifiablity of the model (1.2) as follows:

Proposition 1. *Suppose there exists at least one covariate in X with a nonzero α coefficient. If $(q(\cdot), \alpha, \beta(\cdot), h(\cdot))$ specify the same survival function of T as $\left(\tilde{q}(\cdot), \tilde{\alpha}, \tilde{\beta}(\cdot), \tilde{h}(\cdot)\right)$, then there exists constants c_1 and c_2 such that $\tilde{\alpha} = c_1\alpha, \tilde{\beta} = c_1\beta, \int_0^t \tilde{h}(s)ds = c_2\left(\int_0^t h(s)ds\right)^{c_1}$ and $\int_0^t \tilde{q}^{-1}(s)ds = c_2\left(\int_0^t q^{-1}(s)ds\right)^{c_1}$.*

Remark 1. If one of $h(\cdot)$ or $q(\cdot)$ is specified, then scalar parameter α and functional parameter $\beta(\cdot)$ are identifiable. When $h(\cdot)$ and $q(\cdot)$ are both unspecified, Proposition 1 shows that $(\alpha, \beta, \log\ h(\cdot), \log\ q(\cdot))$ is identifiable up to a location and scale normalization. Therefore, we can set $\log h(t^*) = c$ for some fixed t^* and c and set the first element of α to be 1 to guarantee the identifiability of the model. To interpret the coefficient which is set to be 1, we need to confirm that the first variable has a significant impact on the survival time T with prior or domain knowledge. The remaining coefficients are then interpreted as the relative impact of their corresponding variables on T, compared to the first variable.

2.2 Estimation

We set $\gamma(t) = \log(h(t)), g(t) = \log(q(t))$ to overcome the nonnegative constraint on $h(\cdot)$ and $q(\cdot)$. Without loss of generality, we suppose the support of $\beta(\cdot)$ is $K = [0, 1]$. Then the model (1.2) becomes

$$\begin{cases} \Lambda'_{X,Z}(t) = \exp\left(\alpha^T X + \int_0^1 \beta(s)Z(s)ds + \gamma(t) + g(\Lambda_{X,Z}(t))\right), \\ \Lambda_{X,Z}(0) = 0. \end{cases} \tag{2.4}$$

We consider right-censored data. Denote the event time as T and the censoring time as C. Let $Y = \min\{T, C\}$ being the observed time and $\Delta = I(T \leq C)$ be the censoring indicator with $\Delta = 1$ if the survival time is uncensored and $\Delta = 0$ otherwise. We denote covariate $U = (X, Z(s), s \in K)$. Suppose the observations $W_i = (Y_i, \Delta_i, U_i), i = 1, \ldots, n$, are independent and identically distributed copies of (Y, Δ, U) from the model (2.4). Throughout this paper, we assume that the survival time T and censored time C are independent conditional on the covariates U. Then the log-likelihood of the parameters $\theta = (\alpha, \beta(\cdot), \gamma(\cdot), g(\cdot))$ based on observations $\{W_i, i = 1, \ldots, n\}$ under model (2.4) is given by

$$\begin{aligned} l_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\Bigg\{\Delta_i\left[\alpha^T X_i + \int_0^1 \beta(s) Z_i(s) ds + \gamma(Y_i) + g(\Lambda_{U_i}(Y_i, \theta))\right] \\ &\quad -\Lambda_{U_i}(Y_i, \theta)\Bigg\}, \end{aligned} \tag{2.5}$$

where $\Lambda_{U_i}(t, \theta)$ denotes the solution of the ODE (2.4) parameterized by $\theta = (\alpha, \beta(\cdot), \gamma(\cdot), g(\cdot))$.

We estimate function $\beta(\cdot), \gamma(\cdot), g(\cdot)$ by using B-spline functions. Let $\Pi_n^1 = \{t_1, \ldots, t_{K_{n,1}}\}$ with $0 = t_0 < t_1 < \ldots < t_{K_{n,1}} = 1$ be a sequence of knot which partition $[0, 1]$ into $K_{n,1}$ subintervals. Let $\mathcal{S}_n(p_1, \Pi_n^1)$ with $p_1 \geq 1$ denote the space of splines of order $\lceil p_1 \rceil$ with knot sequence Π_n^1, where $\lceil \cdot \rceil$ represents the ceiling function. Similarly, let Π_n^2 and Π_n^3 be two set of partition points of $[0, \tau]$ and $[0, \mu]$ respectively, where τ and μ will be specified in Section 3. Let $\mathcal{S}_n(p_2, \Pi_n^2)$ and $\mathcal{S}_n(p_3, \Pi_n^3)$ with $p_2 \geq 1, p_3 \geq 1$ denote the space of splines of order $\lceil p_2 \rceil, \lceil p_3 \rceil$ with knot sequence Π_2 and

Π_3 respectively. Let $q_{n,i} = K_{n,i} + \lceil p_i \rceil + 1$ for $i = 1, 2, 3$, then Corollary 4.10 of Schumaker (2007) shows that there exists three sets of B-spline basis $\{B_j^{\beta}, 1 \le j \le q_{n,1}\}$, $\{B_j^{\gamma}, 1 \le j \le q_{n,2}\}$ and $\{B_j^{g}, 1 \le j \le q_{n,3}\}$ such that for any $\hat{\beta}(s) \in \mathcal{S}_n(p_1, \Pi_n^1)$, $\hat{\gamma} \in \mathcal{S}_n(p_2, \Pi_n^2)$ and $\hat{g} \in \mathcal{S}_n(p_3, \Pi_n^3)$, we may write $\hat{\beta}(t) = \sum_{j=1}^{q_{n,1}} a_j B_j^{\beta}(t)$, $\hat{\gamma}(t) = \sum_{j=1}^{q_{n,2}} b_j B_j^{\gamma}(t)$ and $\hat{g}(t) = \sum_{j=1}^{q_{n,3}} c_j B_j^{g}(t)$. We consider the following spaces as in Shen and Wong (1994): $\mathcal{F}_n^{p_1} = \left\{\sum_{j=1}^{q_{n,1}} a_j B_j^{\beta}(t) : \sum_{j=1}^{q_{n,1}} a_j^2 \le l_n\right\}$, $\Gamma_n^{p_2} = \left\{\sum_{j=1}^{q_{n,2}} b_j B_j^{\gamma}(t) : \sum_{j=1}^{q_{n,2}} b_j^2 \le l_n\right\}$ and $\mathcal{G}_n^{p_3} = \left\{\sum_{j=1}^{q_{n,3}} c_j B_j^{g}(t) : \sum_{j=1}^{q_{n,3}} c_j^2 \le l_n\right\}$. where l_n grows with n slowly enough, and $l_n \to \infty$ as $n \to \infty$. Denote the sieve space by

$$\Theta_n = \mathcal{B} \times \mathcal{F}_n^{p_1} \times \Gamma_n^{p_2} \times \mathcal{G}_n^{p_3} = \left\{\theta = (\alpha, \beta, \gamma, g) : \alpha \in \mathcal{B}, \beta \in \mathcal{F}_n^{p_1}, \gamma \in \Gamma_n^{p_2}, g \in \mathcal{G}_n^{p_3}\right\},$$

where $\mathcal{B}$ is a known compact set of $\mathbb{R}^d$. Our estimator is obtained by maximizing the likelihood in the sieve space, that is, $\hat{\theta}_n = \left(\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n\right) := \arg\max_{\theta \in \Theta_n} l_n(\theta)$, which is equivalent to find a $(d + q_{n,1} + q_{n,2} + q_{n,3})$-dimensional vector $\theta = (\alpha, a, b, c)^T$ that maximizes the following log-likelihood function

$$\begin{aligned} l_n(\theta) &= \frac{1}{n} \sum_{i=1}^{n} \Bigg\{ \Delta_i \Bigg[\alpha^T X_i + \sum_{j=1}^{q_{n,1}} a_j \int_0^1 B_j^{\beta}(s) Z_i(s) ds + \sum_{j=1}^{q_{n,2}} b_j B_j^{\gamma}(t) \\ &\quad + \sum_{j=1}^{q_{n,3}} c_j B_j^{g}(\Lambda_{U_i}(Y_i, \theta)) \Bigg] - \Lambda_{U_i}(Y_i, \theta) \Bigg\}, \end{aligned} \tag{2.6}$$

where $\Lambda_{U_i}(t, \theta)$ is the solution of the following ODE equation with respect to $\Lambda(\cdot)$

$$\begin{cases} \Lambda'(t) = \exp\left(\alpha^T X_i + \sum_{j=1}^{q_{n,1}} a_j \int_0^1 B_j^{\beta}(s) Z_i(s) ds + \sum_{j=1}^{q_{n,2}} b_j B_j^{\gamma}(t) + \sum_{j=1}^{q_{n,3}} c_j B_j^{g}\left(\Lambda(t)\right)\right), \\ \Lambda(0) = 0. \end{cases} \tag{2.7}$$

It should be noted that the objective function (2.6) involves the solution of a parameterized ordinary differential equation for which there is no closed-form solution. This distinguishes it from most conventional optimization problems. We use a gradient-based optimization algorithm to optimize (2.5), which is also adopted by Tang et al. (2022) for survival data with scalar covariates. It's worth mentioning that in our proposed estimation method, the log-likelihood of each individual is solely determined by its own observations. This allows for the simultaneous evaluation of independent data points. Consequently, the proposed method is more computationally efficient than the partial likelihood-based estimation method because the partial likelihood of each individual involves computing the at-risk set.

3. Theoretical Results

To ensure identifiability, we suppose the first coordinate of X is continuous and constrain the first element of α be 1. We denote $\alpha = (1, \tilde{\alpha})$ and $X = (X_1, X_{-1})$. For simplicity of notation, we replace $\tilde{\alpha}$ by α and X_{-1} by

X. To study the large-sample properties of the parameter estimators, we require the following regularity conditions on the true parameters, covarites and censoring mechanism. Let $\mathcal{F}_p([a,b])$ be the class of functions f on $[a,b]$ with bounded derivatives $f^{(j)}, j = 1, \ldots, k$ and the $k-$th derivative satisfies the m-Hölder continuity condition:

$$|f^{(k)}(s) - f^{(k)}(t)| \leq M|s-t|^m, \quad \forall s, t \in [a,b],$$

where M is a positive constant, k is a positive integer and $m \in (0,1]$ with $p = k + m$. We also assume the following regularity conditions:

(A1) The event time T and censoring time C are conditionally independent given U.

(A2) Study ends at time $\tau < \infty$ and there exists a positive constant δ_0, such that $P(Y > \tau|U) \geq \delta_0$ almost surely with respect to the probability measure F_U, where F_U represents the joint distribution of $U = (X, Z(\cdot))$.

(A3) (i) The covariate X takes values in a bounded subset $\mathcal{X} \subseteq \mathbb{R}^d$ and satisfies $E(X) = 0$ and $E(XX^T)$ is nonsingular.

(ii) The functional covariate Z takes value in a compact subset $\mathcal{Z} \subseteq L_2(K)$ space, the L_2-norm of Z is bounded almost surely and satisfies $E(Z(s)) = 0$ for all $s \in K$, where K is a compact subset of $\mathbb{R}$.

(A4) The true parameter α_0 belongs to the interior of a compact set $\mathcal{B} \subseteq \mathbb{R}^d$.

(A5) The true functional parameter β_0 belongs to $\mathcal{F}^{p_1}([0,1])$, where $p_1 \geq 2$. The true functions γ_0 and g_0 belongs to $\Gamma^{p_2} := \{\gamma \in \mathcal{F}^{p_2}([0,\tau]) : \gamma(t^*) = 0\}$ and $\mathcal{G}^{p_3} := \mathcal{F}^{p_3}([0,\mu])$ with $p_2 \geq 2, p_3 \geq 3$, respectively.

(A6) (i) For $\mathcal{F}_n^{p_1}$, Let $\Pi(0,1) = \{t_i, i = 0, \ldots, K_n^1 + 1\}$ denote the corresponding knot sequence satisfies $\max_{0 \leq i \leq K_n^1} |t_{i+1} - t_i| = O(n^{-\nu_1})$ and $K_n^1 = O(n^{\nu_1})$ for $\nu_1 \in (0, 0.5)$.

(ii) For $\Gamma_n^{p_2}$ Let $\Pi(0,\tau) = \{t_i, i = 0, \ldots, K_n^2 + 1\}$ denote the corresponding knot sequence satisfies $\max_{0 \leq i \leq K_n^2} |t_{i+1} - t_i| = O(n^{-\nu_2})$ and $K_n^2 = O(n^{\nu_2})$ for $\nu_2 \in (0, 0.5)$.

(iii) For $\mathcal{G}_n^{p_3}$ Let $\Pi(0,\mu) = \{t_i, i = 0, \ldots, K_n^3 + 1\}$ denote the corresponding knot sequence satisfies $\max_{0 \leq i \leq K_n^3} |t_{i+1} - t_i| = O(n^{-\nu_3})$ and $K_n^1 = O(n^{\nu_3})$ for $\nu_3 \in (0, 0.5)$.

(A7) Define $L(t) = \int_0^t \exp(\gamma_0(s))ds, V = \alpha_0^T X + \int_0^1 \beta_0(s)Z(s)ds$ and $R = e^V L(Y)$. There exists a constant $\eta_1 \in (0,1)$ independent of R and V, such that

$$\begin{aligned}\operatorname{Var}&\left(\alpha^T X + \int_0^1 \beta(s)Z(s)ds \middle| R, V\right) \\ &\geq \eta_1 E\left[\left\{\alpha^T X + \int_0^1 \beta(s)Z(s)ds\right\}^2 \middle| R, V\right] \qquad (3.8)\end{aligned}$$

holds almost surely for any $\alpha \in \mathcal{B}$ and $\beta \in \mathcal{F}^{p_1}$.

(A8) Let $\psi(t,x,z,\alpha,\beta,\gamma,g) = \alpha^T X + \int_0^1 \beta(s)Z(s)ds + \gamma(t) + g(\Lambda(t,x,z,\alpha,\beta,\gamma,g))$. Denote the functional derivatives with respect to $\gamma(\cdot)$ and $g(\cdot)$ along the direction $v(\cdot)$, $w(\cdot)$ at the true parameter by $\psi'_{0\gamma}(t,u)[v]$ and $\psi'_{0g}(t,u)[w]$, respectively. Then there exists a constant $\eta_2 \in (0,1)$ such that

$$
\begin{aligned}
& \left(E\left\{\psi'_{0\gamma}(Y,U)[v]\psi'_{0g}(Y,U)[w]|\Delta=1\right\}\right)^2 \\
\leq\;& \eta_2 E\left\{\left(\psi'_{0\gamma}(Y,U)[v]\right)^2|\Delta=1\right\} E\left\{\left(\psi'_{0g}(Y,U)[w]\right)^2|\Delta=1\right\}.
\end{aligned}
$$

Condition (A1) is a common assumption when analyzing right-censored data, which ensures the censoring mechanism does not bring extra information. Condition (A2) means that the study is conducted in a time period $[0,\tau]$ and $P(Y>\tau|U) \geq \delta_0$ almost surely implies that there are some subject still alive at the end of the study, which implies $\Lambda_0(\tau) \leq -\log\delta_0 =: \mu$. Condition (A3) is the same as that in Qu et al. (2016), which requires the boundedness of covariates. Condition (A4) and (A5) impose constraints on the parameter spaces, where the latter requires smoothness of the functional parameters in order to control the error rate of spline approximation. Similar assumptions is often adopted in spline estimation, see Huang (1999); Ding and Nan (2011); Tang et al. (2022). Condition (A7) is assumed to guarantee the identifiability of α_0 and $\beta_0(\cdot)$, similar assumptions is also used by Wellner and Zhang (2007); Ding and Nan (2011). Condition (A8) is essential for the identifiability of γ and g when they are both unspecified.

Denote the counting process martingale associated with the process $\left\{I_{\{R\le t\}}, t\ge 0\right\}$ with R defined in Assumption (A7), as $M(t) = \Delta I\{(R \le t)\} - \int_0^t I\{R\ge s\} d\tilde{\Lambda}_0(s)$, where $\tilde{\Lambda}_0(t)$ is the solution of $\tilde{\Lambda}_0'(t) = \exp\left(g_0(\tilde{\Lambda}_0)\right)$ with $\tilde{\Lambda}_0(0) = 0$. We first derive the efficient score function and the information bound for the estimation of α_0.

Theorem 1. *Under Conditions (A1)-(A5) and (A7), the efficient score function $\dot{l}^*_{\alpha_0}$ for the estimation α_0 is*

$$\int \Bigg\{ \left(g_0'(\tilde{\Lambda}_0(t)) \exp\left(g_0(\tilde{\Lambda}_0(t))\right) t + 1\right) \left(X - \int_0^1 \boldsymbol{h}_1^*(s) Z(s) ds\right)$$
$$- g_0'(\tilde{\Lambda}_0(t)) \exp\left(g_0(\tilde{\Lambda}_0(t))\right) \int_0^t \boldsymbol{h}_2^*(L^{-1}(se^{-V})) ds + \boldsymbol{h}_2^*(L^{-1}(se^{-V}))$$
$$- g_0'(\tilde{\Lambda}_0(t)) \exp\left(g_0(\tilde{\Lambda}_0(t))\right) \int_0^{\tilde{\Lambda}_0(t)} \exp(-g_0(s)) \boldsymbol{h}_3^*(s) ds + \boldsymbol{h}_3^*(\tilde{\Lambda}_0(t)) \Bigg\} dM(t),$$

where $(\boldsymbol{h}_1^*, \boldsymbol{h}_2^*, \boldsymbol{h}_3^*)$ is defined by

$$\underset{(\boldsymbol{h}_1,\boldsymbol{h}_2,\boldsymbol{h}_3)\in\mathbb{T}}{\arg\min}\ E\Bigg[\Delta\Bigg\| \left(g_0'(\tilde{\Lambda}_0(R)) \exp\left(g_0(\tilde{\Lambda}_0(R))\right) R + 1\right) \left(X - \int_0^1 \boldsymbol{h}_1(s) Z(s) ds\right)$$
$$- g_0'(\tilde{\Lambda}_0(R)) \exp\left(g_0(\tilde{\Lambda}_0(R))\right) \int_0^R \boldsymbol{h}_2(L^{-1}(se^{-V})) ds + \boldsymbol{h}_2(L^{-1}(Re^{-V}))$$
$$- g_0'(\tilde{\Lambda}_0(R)) \exp\left(g_0(\tilde{\Lambda}_0(R))\right) \int_0^{\tilde{\Lambda}_0(R)} \exp(-g_0(s)) \boldsymbol{h}_3(s) ds + \boldsymbol{h}_3(\tilde{\Lambda}_0(R)) \Bigg\|^2\Bigg]$$

and the domain $\mathbb{T}$ will be defined in Supplementary materials. The infor-

mation bound for α_0 *is*

$$I(\alpha_0) = E\left[l_{\alpha_0}^{*\otimes 2}\right], \tag{3.9}$$

where $x^{\otimes 2} = xx^T$ *for any vector* $x \in \mathbb{R}^p$.

Remark 2. It is challenging to obtain an explicit solution for the efficient direction $(\mathbf{h}_1^*, \mathbf{h}_2^*, \mathbf{h}_3^*)$ due to the integrations; however, we can find their solutions numerically. For example, when $g_0 \equiv 0$, it reduces to a functional Cox model, which is the most widely used model in practice, and in this case, the direction is $\mathbf{h}_3^* = 0$. Then the direction $\mathbf{h}_1^*, \mathbf{h}_2^*$ in Theorem 2 minimizes

$$E\left\{\Delta\left\|X - \int \mathbf{h}_1(s)Z(s)ds + \mathbf{h}_2(Y)\right\|^2\right\},$$

which is equivalent to the directions given by Qu et al. (2016) in functional Cox model whose explicit form of $\mathbf{h}_1^*, \mathbf{h}_2^*$ are challenging to derive. However, as suggested by Qu et al. (2016), their numerical solutions can be obtained using the population version of the ACE algorithm. (Wang and Murphy, 2004)

Next, we give the convergence rate of the estimator. We measure the accuracy of $\hat{\alpha}_n$ by the usual Euclidean norm and measure the accuracy of $\hat{\beta}_n$ by the weighted L_2-norm $\|\hat{\beta}_n - \beta_0\|_C$, where $\|\beta\|_C^2 = \int\int \beta(s)C(s,t)\beta(t)dsdt$, $C(s,t) = E\left(Z(s)Z(t)\right)$. The accuracy of $\gamma(\cdot)$ is measured by the L_2-norm

$\|\hat{\gamma}-\gamma\|_2 := \left(\int_0^\tau (\gamma(t))^2 dt\right)^{1/2}$. Note that the parameters g and α, β, γ are bundled together, which makes the information of separate parameters difficult to derive. Therefore, for any given (α, β, γ), we investigate the composite function $g\left(\Lambda(\cdot,\cdot,\cdot,\alpha,\beta,\gamma,g)\right)$ directly as a function from $\mathbb{R} \times \mathcal{X} \times L_2([0,1])$. To be specific, define the collection of functions

$$\begin{aligned} \mathcal{H}^{p_3} &= \Big\{\zeta(\cdot,\alpha,\beta,\gamma) : \zeta(t,u,\alpha,\beta,\gamma) = g\left(\Lambda(t,u,\alpha,\beta,\gamma,g)\right), t \in [0,\tau], \\ &\qquad u \in \mathcal{X} \times L_2([0,1]), \alpha \in \mathbb{R}^d, \beta \in \mathcal{F}^{p_1}, \gamma \in \Gamma^{p_2}, g \in \mathcal{G}^{p_3}\Big\} \end{aligned}$$

for any $\zeta(\cdot,\alpha,\beta,\gamma) \in \mathcal{H}^{p_3}$, define its norm as:

$$\|\zeta(\cdot,\alpha,\beta,\gamma)\|_2 = \left[\int_{\mathcal{X}\times L_2([0,1])} \int_0^\tau |\zeta(t,u,\alpha,\beta,\gamma)|^2 d\Lambda_0(t,u) dF_U(u)\right]^{1/2}.$$

Theorem 2 (Rate of convergence)**.** *Suppose conditions (A1)-(A8) hold and the information bound $I(\alpha_0)$ defined in Theorem 1 is nonsingular, then we have*

$$\begin{aligned} \|\hat{\beta}-\beta_0\|_C = O_p(n^{-c}), \|\hat{\gamma}-\gamma_0\|_2 = O_p(n^{-c}), \\ \left\|\hat{\zeta}(\cdot,\hat{\alpha},\hat{\beta},\hat{\gamma}) - \zeta_0(\cdot,\alpha_0,\beta_0,\gamma_0)\right\|_2 = O_p(n^{-c}), \end{aligned} \tag{3.10}$$

where $c = \min\{p_1\nu_1, p_2\nu_2, p_3\nu_3, \frac{1-\max\{\nu_1,\nu_2,\nu_3\}}{2}\}$.

Theorem 2 gives the convergence rate of sieve MLE. When $\nu_1 = \nu_2 = \nu_3 = \nu$, $p_1 = p_2 = p_3 = p$, the convergence rate becomes $O_p(n^{-\min\{p\nu, \frac{1-\nu}{2}\}})$, which is the same convergence rate as the case in (Ding and Nan, 2011) when there is no functional parameter. If we further assume $\nu = \frac{1}{2p+1}$, then the convergence rate becomes $O_p(n^{-\frac{p}{2p+1}})$, which is the optimal convergence rate in nonparametric regression (Stone, 1985).

The next theorem will show that the sieve MLE of the scalar parameter remains asymptotically normal and reach $n^{1/2}$ convergence rate despite the slower convergence rate than $O(n^{1/2})$ of the nonparametric part.

Theorem 3 (Semi-parametric efficiency)**.** *Suppose Conditions (A1)-(A8) hold and the information bound $I(\alpha_0)$ defined in Theorem 1 is nonsingular. If ν_1, ν_2, ν_3 satisfies $\frac{1}{2(p_1+2)} < \nu_1 < \frac{1}{2p_1}$, $\frac{1}{2(p_2+2)} < \nu_2 < \frac{1}{2p_2}$, $\frac{1}{2(p_3+1)} < \nu_3 < \frac{1}{2p_3}$, $\nu_3 > \frac{1}{2(p_3-1)} - \frac{2\min\{\nu_1,\nu_2\}}{p_3-1}$, $\nu_3 > \frac{1}{2} - \max\{p_1\nu_1, p_2\nu_2\}$ and $\max\{\nu_1, \nu_2, \nu_3\} < 2\min\{\nu_1, \nu_2, \nu_3\}$, then we have*

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{D}} N(0, I(\alpha_0)^{-1}), \tag{3.11}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

Theorem 3 shows that $\hat{\alpha}_n$ achieves the information bound displayed in Theorem 1. Therefore, it is asymptotically efficient among all the regular estimators. The restrictions on ν and p is relatively mild and can be satisfied when ν_1, ν_2 and ν_3 are not far away from each other. For example, if $\nu_1 = \nu_2 = \nu_3 = \nu$, the restrictions hold when $\frac{1}{2p+1} < \nu < \frac{1}{2p}$.

To derive the asymptotic simultaneous confidence band for the estimate of β_0, we need the following additional regularity condition. For any $s, t \in [0,1]$, let $K(s) = E[Z(s)|X], s \in [0,1]$ and

$$K(s,t) = E\left[(Z(s) - K(s))\,(Z(t) - K(t))\right].$$

In what follows, we further assume

(A9) $K(s,t)$ satisfies the Sack-Ylvisaker conditions with boundary condition $K(\cdot, 0) = 0$.

The Sack-Ylvisaker conditions (Sacks and Ylvisaker, 1966) imposes smoothness restrictions on the covariance kernel functions, which is a common assumption in the literature of functional linear regression (Yuan and Cai, 2010; Ritter et al., 1995). Ritter et al. (1995) reveals that under Sack-Ylvisaker conditons, the reproducing kernel Hilbert space induced by kernel function K, denoted as $\mathcal{H}(K)$, only differs from Sobolev space $\mathcal{W}_2^1[0,1]$ by a finite dimensional space. This relationship between $\mathcal{H}(K)$ and $\mathcal{W}_2^1[0,1]$ is essential for our theoretical development. The rigorous definition of Sack-Ylvisaker conditions are deferred to Appendix.

Theorem 4 (Asymptotic simultaneous confidence band)**.** *Assume the conditions in Theorem 3 hold, we further assume* $\frac{1}{2p_1+2} < \nu_1 < \max\{-\frac{1}{2} + 2c - \nu_3, \frac{1}{6}\}$*, the true function* β_0 *vanishes at the boundary of* $[0,1]$*. Then there exists a Gaussian process* $\{G_n(s), s \in (0,1)\}$ *with* $E[G_n(s)] = 0$*,*

$Var[G_n(s)] = 1$ and $E[G_n(s)G_n(t)] = G_n(s,t)$ such that

$$\sup_{s\in(0,1)}\left|\frac{\sqrt{n}\left(\hat{\beta}_n(s)-\beta_0(s)\right)}{\sqrt{[\mathbf{B}_n^{\beta}(s)]^T J_n^{-1}\mathbf{B}_n^{\beta}(s)}} - G_n(s)\right| \xrightarrow{P} 0,$$

where $\mathbf{B}_n^{\beta}(s) = \left(B_1^{\beta}(s),\ldots,B_{q_{n,1}}^{\beta}(s)\right)^T$ is a B-spline basis for estimating β_0 and J_n is a $q_{n,1}\times q_{n,1}$ matrix. The exact form of $G_n(s,t)$ and J_n will be given in Supplementary materials.

Theorem 4 enables us to construct both asymptotic simultaneous confidence band and pointwise confidence intervals for β_0. Let $z_{\alpha,n}$ be the α-th percentile of $\sup_{s\in(0,1)} G_n(s)$, then an asymptotic $100(1-\alpha)\%$ simultaneous confidence band for β_0 would be

$$\Bigg[\hat{\beta}_n(s) - \frac{1}{\sqrt{n}} z_{\alpha/2,n}\sqrt{[\mathbf{B}_n^{\beta}(s)]^T J_n^{-1}\mathbf{B}_n^{\beta}(s)},\\ \hat{\beta}_n(s) + \frac{1}{\sqrt{n}} z_{\alpha/2,n}\sqrt{[\mathbf{B}_n^{\beta}(s)]^T J_n^{-1}\mathbf{B}_n^{\beta}(s)}\Bigg]$$

and the asymptotic 100(1-α)% pointwise confidence interval can be constructed as

$$\Bigg[\hat{\beta}_n(s) - \frac{1}{\sqrt{n}} z'_{1-\alpha/2,n}\sqrt{[\mathbf{B}_n^{\beta}(s)]^T J_n^{-1}\mathbf{B}_n^{\beta}(s)},\\ \hat{\beta}_n(s) - \frac{1}{\sqrt{n}} z'_{\alpha/2,n}\sqrt{[\mathbf{B}_n^{\beta}(s)]^T J_n^{-1}\mathbf{B}_n^{\beta}(s)}\Bigg],$$

where $z'_{\alpha,n}$ is the α-th percentile of standard normal distribution. It is easy to know from the definition that $z_{\alpha,n} > z'_{\alpha,n}$. Therefore, the simultaneous confidence band is usually strictly larger wider than the confidence interval. The implementary detail of constructing confidence interval and confidence band are deferred to Supplementary materials.

Remark 3. In practice, one can use the empirical distribution of Gaussian process with zero mean, unit variance and covariance function $G_n(s,t)$ to approximate the distribution of $G_n(s)$. Subsequently, the α-th percentile of $\sup_{s\in(0,1)} G_n(s)$ can be approximated by the empirical percentile of its numerous realizations. It is important to note that G_n is only related to n through the B-spline basis $\mathbf{B}_n^{\beta}$. This observation suggests that a different sample size with the same B-spline basis may result in the same distribution of G_n.

4. Simulation

In this section, we conduct simulations under different settings to evaluate the finite sample performance of the proposed method. We also provide numerical comparisons with the Functional Cox model (FCox) proposed by Qu et al. (2016). The estimation procedure is implemented by Python with some existing packages. Specifically, we construct the B-spline functions using the "*Bspline*" function of the "*scipy.interpolate*" package and solve ordinary differential equations using "*solve_ivp*" function of "*scipy.integrate*" package.

For the functional covariates, we employ a design similar to that used by Qu et al. (2016), where the functional covariates $Z(\cdot)$ is generated using a set of cosine basis functions, that is, $Z(s) = \sum_{k=1}^{50} \xi_k U_k \phi_k(s)$, 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 $\phi_{k+1} = \sqrt{2}\cos(k\pi s)$ for $k \geq 1$. The coefficient function $\beta_0(s)$ is set as $\beta_0(s) = \sum_{i=1}^{50}(-1)^k k^{-3/2}\phi_k(s)$. The scalar covariates X follows standard normal distribution truncated at ± 2. The event time T is generated based on model:

$$\Lambda_{X,Z}(t) = h_0(t) \exp\left(\alpha_0^T X + \int_0^1 \beta_0(s)Z(s)ds\right) q_0\left(\Lambda_{X,Z}(t)\right).$$

We consider the following settings for $\alpha_0, h_0(\cdot)$ and $q_0(\cdot)$:

Setting 1: $\alpha_0 = 1, q_0(t) = 1$ and $h_0(t) = 1 + t^3$;

Setting 2: $\alpha_0 = (1, 1), q_0(t) = 1$ and $h_0(t) = 1 + t^3$;

Setting 3: $\alpha_0 = 1, h_0(t) = 1$ and $q_0(t) = \exp(2/(1 + t))$;

Setting 4: $\alpha_0 = (1, 1), h_0(t) = 1$ and $q_0(t) = \exp(2/(1 + t))$;

Setting 5: $\alpha_0 = (1, 1), h_0(t) = 1+\log(1+2t^3)$ and $q_0(t) = 1+\log(1+t)$.

In setting 1 the functional Cox model is specified, we set $q(t) = 1$ as fixed and leave $h(t)$ unspecified. In setting 3 the functional AFT model is specified, we set $h(t)=1$ as fixed and leave $q(t)$ unspecified. In setting 2, 4 and 5, we set $\alpha_2=1$ in order to make the model identifiable and leave both $h(t)$ and $q(t)$ as unspecified. In each setting, the censoring time C is generated from an independent uniform distribution $U(0,c)$, where the value of c is selected to achieve censoring rates ranging from approximately 15% to 30%. The sample size varies from 200, 400, 600 and 800. We estimate β_0 using a cubic B-spline with $\lceil n^{1/5}\rceil$ interior nodes that are equally spaced at the interval $[0,1]$. We fit $\log\ h(t)$ and $\log\ q(t)$ by cubic B-spline with $\lceil n^{1/7}\rceil$ interior nodes that are equally spaced at the interval $[0,\tau]$ and $[0,\mu]$ respectively, where μ is chosen large enough to cover the value of estimated cumulative hazard at all observed event time.

The simulation results are based on 1000 replications. Table 1 and Table 2 compare the performance of the proposed estimators for the scalar and functional parameter with MPLE-based estimators under setting 1 and 2. BIAS is calculated as the difference between the mean of the estimates and the true value. SE represents the standard error of parameter estimators, SEE is the mean of standard error estimator obtained by inverting the estimated information matrix. CP represents the corresponding coverage proportion of the 95% confidence interval. The results shows that under setting 1 and 2 (when the proportional harzard assumption is satisfied), the

Table 1: Simulation results for scalar parameter α and comparison with MPLE based method under setting 1 and 2.

			Censoring rate $\approx$ 15%				Censoring rate $\approx$ 30%			
Setting	n	Method	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP
1	200	FunODE	0.045	0.102	0.112	0.958	0.037	0.112	0.123	0.961
		MPLE	0.010	0.100	0.100	0.942	0.024	0.114	0.103	0.925
	400	FunODE	0.021	0.071	0.074	0.958	0.011	0.077	0.081	0.972
		MPLE	-0.005	0.066	0.069	0.957	0.004	0.080	0.076	0.943
	600	FunODE	0.021	0.059	0.059	0.941	0.009	0.062	0.065	0.957
		MPLE	0.001	0.062	0.056	0.933	0.001	0.063	0.066	0.952
	800	FunODE	0.009	0.048	0.050	0.960	0.005	0.053	0.055	0.951
		MPLE	0.000	0.051	0.048	0.936	0.000	0.055	0.058	0.938
2	200	FunODE	0.023	0.152	0.146	0.922	0.055	0.154	0.168	0.941
		MPLE	0.004	0.100	0.099	0.951	0.019	0.113	0.103	0.920
	400	FunODE	0.009	0.096	0.101	0.966	0.005	0.119	0.110	0.921
		MPLE	-0.010	0.074	0.069	0.941	-0.005	0.076	0.074	0.954
	600	FunODE	0.012	0.066	0.073	0.960	-0.008	0.073	0.083	0.963
		MPLE	-0.007	0.063	0.056	0.935	-0.004	0.061	0.058	0.942
	800	FunODE	-0.004	0.058	0.063	0.957	-0.006	0.062	0.071	0.955
		MPLE	-0.002	0.053	0.048	0.940	-0.002	0.054	0.050	0.923

Table 2: Simulation results for functional parameter β and comparison with MPLE based method under setting 1 and 2.

			Censoring rate $\approx$ 15%		Censoring rate $\approx$ 30%	
Setting	n	Method	IMSE	RIMSE	IMSE	RIMSE
1	200	FunODE	0.037	0.052	0.052	0.061
		MPLE	0.027	0.039	0.030	0.042
	400	FunODE	0.017	0.024	0.020	0.028
		MPLE	0.018	0.025	0.019	0.027
	600	FunODE	0.015	0.021	0.016	0.022
		MPLE	0.013	0.019	0.015	0.021
	800	FunODE	0.010	0.014	0.012	0.017
		MPLE	0.011	0.015	0.013	0.018
2	200	FunODE	0.052	0.073	0.057	0.080
		MPLE	0.028	0.040	0.031	0.043
	400	FunODE	0.024	0.034	0.025	0.035
		MPLE	0.027	0.038	0.023	0.032
	600	FunODE	0.018	0.025	0.021	0.029
		MPLE	0.021	0.030	0.018	0.026
	800	FunODE	0.013	0.018	0.016	0.022
		MPLE	0.012	0.017	0.017	0.024

Table 3: Simulation results for scalar parameter α under setting 3-5.

		Censoring rate $\approx$ 15%				Censoring rate $\approx$ 30%			
Setting	n	BIAS	SE	SEE	CP	BIAS	SE	SEE	CP
3	200	0.023	0.129	0.127	0.898	0.012	0.134	0.144	0.926
	400	0.018	0.090	0.086	0.951	-0.001	0.090	0.098	0.950
	600	0.012	0.066	0.073	0.960	-0.008	0.073	0.083	0.963
	800	-0.004	0.058	0.063	0.957	-0.006	0.062	0.071	0.955
4	200	0.023	0.152	0.146	0.922	0.055	0.154	0.168	0.941
	400	0.009	0.096	0.101	0.966	0.005	0.119	0.110	0.921
	600	0.008	0.077	0.082	0.963	0.009	0.086	0.083	0.939
	800	0.002	0.067	0.067	0.943	$<$0.001	0.073	0.073	0.949
5	200	0.041	0.101	0.111	0.925	-0.009	0.105	0.126	0.942
	400	0.031	0.068	0.074	0.941	0.030	0.071	0.084	0.954
	600	0.028	0.053	0.059	0.946	0.023	0.057	0.066	0.960
	800	0.018	0.045	0.052	0.944	0.009	0.051	0.057	0.947

proposed estimators are strongly competitive with MPLE-based estimators. Table 3 summarizes the performance of estimator for the scalar parameter in general settings. As shown in Table 1 and 3, in all the five settings, the mean of estimators is close to the true value and both the standard error (SE) and the mean of standard error estimator (SEE) decreases as the sample size n increases with censoring rate fixed, thereby confirming the consistency of our proposed estimator. Furthermore, the SE exhibit lower values at the 15% censoring rate compared to the values observed at the 30% censoring rate. This observation aligns with the expected outcome, as a lower censoring rate typically leads to more accurate estimates.

Besides, we calculated information matrix through inverting the empirical Hessian matrix of log-likelihood (2.6) and then construct confidence interval based on Theorem 3. When the sample size exceeds 400, the corresponding coverage probability closely aligns with the theoretical level of 95%, indicating that a normal approximation is suitable.

Table 2 shows the IMSE and relative integrated mean square error (RIMSE) for the functional parameter estimator $\hat{\beta}(\cdot)$, which are defined as follows: $\text{IMSE}(\hat{\beta}) = \int_0^1 (\hat{\beta}(s) - \beta_0(s))^2 ds$ and $\text{RIMSE}(\hat{\beta}) = \text{IMSE}(\hat{\beta}) / \|\beta_0\|_2^2$. As shown in Table 2, the IMSE exhibits an obvious decreasing trend as the sample size n increases in all five scenarios. The empirical coverage proba-

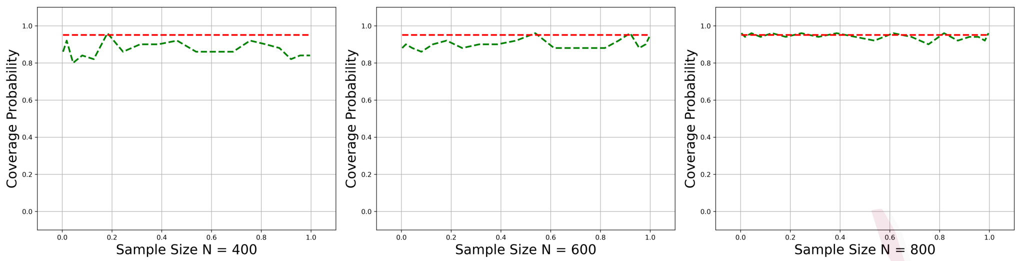

Figure 1: Pointwise coverage probability with sample size $n = 400, 600, 800$. The dashed red line represents the theoretical value of 0.95, while the dashed green line represents the empirical coverage probability of the pointwise confidence interval.

bilities of simultaneous confidence band for β are deferred to supplementary materials.

Figure 1 displays the empirical pointwise coverage probability of the pointwise confidence interval under setting 5 for sample size $n = 400$ and 800. The dash green line represents empirical coverage probability of pointwise confidence interval for each point on $[0, 1]$ and the dash red line represents the theoretical value 0.95. The results demonstrate that the empirical coverage probability closely aligns with its theoretical value as sample size increases.

5. Real Data Example

This section presents an application of the proposed functional transformation model to the Sequential Organ Failure Assessment (SOFA) data acquired from the Improving Care of Acute Lung Injury Patients (ICAP) study (Needham et al., 2006; Gellar et al., 2014). The ICAP study aims to investigate the long-term complications of patients who suffer from acute lung injury/acute respiratory distress syndrome (ALI/ARDS). A total of 520 subjects were involved in the study, with 237 (46%) of them passing away in the intensive care unit (ICU). Our analysis excludes 107 individuals (31.0%) who died within the first five days in ICU. The number of days in ICU till death are regarded as the event time.

During the ICAP study, patient data were collected upon admission to the ICU and then daily throughout their hospitalization. One of the measurements recorded daily was the Sequential Organ Failure Assessment (SOFA) score, which provides an assessment of a patient's overall organ function status. The SOFA score includes six components: respiratory, cardiovascular, coagulation, liver, renal, and neurological, with scores ranging from 0 to 4. Higher scores indicate poorer organ function. The SOFA score is calculated as the sum of these six component scores and ranges from 0 to 24.

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

	$\hat{\alpha}$	S.E.	p-value
Age	0.204	0.065	<0.001
Gender(male=1)	-0.023	0.100	0.822

To account for the evolution of each subject's organ function, we consider their history of SOFA scores during the first five days as a functional covariate denoted as $Z(s)$, where s represents the number of days since admission to the ICU. The model also include age, gender and Charlson co-morbidity index as three scalar covariates. Both scalar and functional covariates are centralized in order to satisfy Condition (A3). We adopted cubic spline functions with $\lceil n^{1/5} \rceil$ equally spaced interior nodes to estimate the functional coefficient and with $\lceil n^{1/7} \rceil$ equally spaced interior nodes to estimate the nuisance parameter γ and g. The estimated functional coefficient $\hat{\beta}(s)$ is shown in Figure 2. we can see that $\hat{\beta}(s)$ shows an increasing trend and the 95% confidence band does not cover the horizontal line when $s \in [0.6, 1]$, suggesting that higher SOFA score in the fourth and fifth day may lead to higher mortality rate.

The estimation of the regression coefficients of the scalar covariates is summarized in Table 4. Alongside the functional covariate, the analysis reveals that patients' age has a positive impact on the hazard, whereas gender does not show a significant association with the hazard of death. This agrees with the recent study by Gellar et al. (2015). Further discussions about our real data example is deferred to supplementary materials.

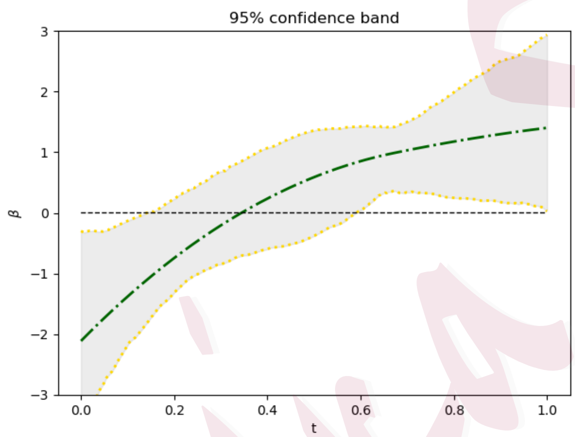

Figure 2: The estimated functional coefficient $\hat{\beta}(\cdot)$ and the pointwise 95% confidence interval.

6. Concluding Remarks

In this paper, we have proposed a general class of survival model for analyzing right-censored survival data, which encompasses the functional Cox model and functional accelerated failure time model as special cases. Within the ODE framework, we developed a sieve maximum likelihood estimator. Our rigorous theoretical analysis has revealed the large sample properties of the estimators, including their consistency and semiparametric efficiency. Furthermore, we have derived an asymptotic simultaneous confidence band for the functional parameter, ensuring the reliability of inferences.

Our proposed method can be readily extended to handle scenarios with high dimensional scalar covariates. However, in high-dimensional settings, interpretability becomes a major concern and the classical large sample theories may lead to invalid inference, as the Fisher information matrix is singular when the number of scalar parameters $d > n$. Therefore, detecting and analyzing sparsity in survival models with functional covariate in high dimensional settings would also be an intriguing avenue for future research.

7. Acknowledgements

The authors would like to thank the editor, the associate editor and the two reviewers for their constructive and insightful comments that greatly improved this article. Zhong's research was supported by National Key R&D Program of China (1003800, 1015600), the National Natural Science Foundation of China grant (12201527). Yang's research was supported by the National Natural Science Foundation of China grant (12271286, 11931001).

References

Aalen, O. (1978). Nonparametric Inference for a Family of Counting Processes. *The Annals of Statistics*, 6:701–726.

Aalen, O. O. and Gjessing, H. K. (2001). Understanding the Shape of the Hazard Rate: A Process Point of View (With comments and a rejoinder by the authors). *Statistical Science*, 16:1–22.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277.

Chen, K., Chen, K., Müller, H.-G., and Wang, J.-L. (2011). Stringing High-Dimensional Data for Functional Analysis. *Journal of the American Statistical Association*, 106:275–284.

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric Analysis of Transformation Models with Censored Data. *Biometrika*, 89:659–668.

Cheng, S. C., Wei, L. J., and Ying, Z. (1995). Analysis of Transformation Models with Censored Data. *Biometrika*, 82:835–845.

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society: Series B*, 34:187–202.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Monographs on Statistics and Applied Probability. Chapman and Hall, New York, NY.

Cui, E., Crainiceanu, C. M., and Leroux, A. (2021). Additive Functional Cox Model. *Journal* *of Computational and Graphical Statistics*, 30:780–793.

Ding, Y. and Nan, B. (2011). A Sieve M-theorem for Bundled parameters in Semiparametric Models, with Application to the Efficient Estimation in a Linear model for Censored Data. *The Annals of Statistics*, 39:3032–3061.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Wiley, New York.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2014). Variable-Domain Functional Regression for Modeling ICU Data. *Journal of the American Statistical Association*, 109:1425–1439.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2015). Cox Regression Models with Functional Covariates for Survival Data. *Statistical Modelling*, 15:256–278.

Hao, M., Liu, K., Xu, W., and Zhao, X. (2021). Semiparametric Inference for the Functional Cox Model. *Journal of the American Statistical Association*, 116:1319–1329.

Horowitz, J. L. (1996). Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable. *Econometrica*, 64:103–117.

Huang, J. (1999). Efficient Estimation of the Partly Linear Additive Cox Model. *The Annals of Statistics*, 27:1536–1563.

Jiang, F., Cheng, Q., Yin, G., and Shen, H. (2020). Functional Censored Quantile Regression.

Journal of the American Statistical Association, 115:931–944.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data (2nd Edition)*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53:457–481.

Kong, D., Ibrahim, J. G., Lee, E., and Zhu, H. (2018). FLCRM: Functional Linear Cox Regression Model. *Biometrics*, 74:109–117.

Li, Y. and Hsing, T. (2010). Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data. *The Annals of Statistics*, 38:3321–3351.

Murphy, S. A., Rossini, A. J., and van der Vaart, A. W. (1997). Maximum Likelihood Estimation in the Proportional Odds Model. *Journal of the American Statistical Association*, 92:968–976.

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.

Ritov, Y. (1990). Estimation in a Linear Regression Model with Censored Data. *The Annals of Statistics*, 18:303–328.

Ritter, K., Wasilkowski, G. W., and Wozniakowski, H. (1995). Multivariate Integration and Approximation for Random Fields Satisfying Sacks-Ylvisaker Conditions. *The Annals of Applied Probability*, 5:518–540.

Sacks, J. and Ylvisaker, D. (1966). Designs for Regression Problems With Correlated Errors. *The Annals of Mathematical Statistics*, 37:66–89.

Schumaker, L. (2007). *Spline Functions: Basic Theory (3rd Edition)*. Cambridge University Press, Cambridge Mathematical Library.

Shen, X. and Wong, W. H. (1994). Convergence Rate of Sieve Estimates. *The Annals of Statistics*, 22:580–615.

Shi, H., Ma, D., Beg, M. F., and Cao, J. (2022). A Functional Proportional Hazard Cure Rate Model for Interval-censored Data. *Statistical Methods in Medical Research*, 31:154–168.

Spreafico, M. and Ieva, F. (2021). Functional Modeling of Recurrent Events on Time-to-event Processes. *Biometrical Journal*, 63:948–967.

Spreafico, M., Ieva, F., and Fiocco, M. (2023). Modelling Time-varying Covariates Effect on Survival via Functional Data Analysis: Application to the MRC BO06 Trial in Osteosarcoma. *Statistical Methods & Applications*, 32:271–298.

Stone, C. J. (1985). Additive Regression and Other Nonparametric Models. *The Annals of*

Statistics, 13:689–705.

Tang, W., He, K., Xu, G., and Zhu, J. (2022). Survival Analysis via Ordinary Differential Equations. *Journal of the American Statistical Association*, 118:2406–2421.

Wang, D. L. and Murphy, M. (2004). Estimating optimal transformations for multiple regression using the ace algorithm. *Journal of Data ence*, 2(4):329–346.

Wang, J. L., Chiou, J.-M., and Müller, H. G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

Wei, L. J. (1992). The Accelerated Failure Time Model: A useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*, 11:1871–1879.

Wellner, J. A. and Zhang, Y. (2007). Two Likelihood-based Semiparametric Estimation Methods for Panel Count Data with Covariates. *The Annals of Statistics*, 35:2106–2142.

Yang, H., Zhu, H., Ahn, M., and Ibrahim, J. G. (2021). Weighted Functional Linear Cox Regression Model. *Statistical Methods in Medical Research*, 30:1917–1931.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional Linear Regression Analysis for Longitudinal Data. *The Annals of Statistics*, 33:2873–2903.

Yuan, M. and Cai, T. T. (2010). A Reproducing Kernel Hilbert Space Approach to Functional Linear Regression. *The Annals of Statistics*, 38:3412–3444.

Zeng, D. and Lin, D. Y. (2007a). Efficient Estimation for the Accelerated Failure Time Model. *Journal of the American Statistical Association*, 102:1387–1396.

Zeng, D. and Lin, D. Y. (2007b). Maximum Likelihood Estimation in Semiparametric Regression Models with Censored Data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69:507–564.

Zhang, X. and Wang, J.-L. (2016). From Sparse to Dense Functional Data and Beyond. *The Annals of Statistics*, 44:2281–2321.

Zhong, Q., Mueller, J., and Wang, J.-L. (2022). Deep Learning for the Partially Linear Cox Model. *The Annals of Statistics*, 50:1348–1375.

Department of Mathematical Science, Tsinghua University, Beijing, China

E-mail: zhou-hy21@mails.tsinghua.edu.cn

Department of Mathematical Science, Tsinghua University, Beijing, China

E-mail: swq21@mails.tsinghua.edu.cn

Department of Statistics and Data Science, Xiamen University, Fujian, China

E-mail: qxzhong@xmu.edu.cn

Department of Mathematical Science, Tsinghua University, Beijing, China

E-mail: yangying@tsinghua.edu.cn