Biometrika (2025), **112**, 2, asaf011 https://doi.org/10.1093/biomet/asaf011
Advance Access publication 7 March 2025

Efficient estimation for functional accelerated failure time models

By CHANGYU LIU

Department of Statistics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
changyuliu@cuhk.edu.hk

WEN SU

Department of Biostatistics, City University of Hong Kong, Kowloon Tong, Hong Kong
w.su@cityu.edu.hk

KIN-YAT LIU

Department of Statistics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
kinyatliu@cuhk.edu.hk

GUOSHENG YIN

School of Computing and Data Science, The University of Hong Kong, Pokfulam Road, Hong Kong
gyin@hku.hk

and XINGQIU ZHAO

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
xingqiu.zhao@polyu.edu.hk

Summary

We propose a functional accelerated failure time model to characterize the effects of both functional and scalar covariates on the time to event of interest, and provide regularity conditions to guarantee model identifiability. For efficient estimation of model parameters, we develop a sieve maximum likelihood approach where parametric and nonparametric coefficients are bundled with an unknown baseline hazard function in the likelihood function. Not only do the bundled parameters cause immense numerical difficulties, but they also result in new challenges in theoretical development. By developing a general theoretical framework, we overcome the challenges arising from the bundled parameters and derive the convergence rate of the proposed estimator. Additionally, we prove that the finite-dimensional estimator is root-*n* consistent, asymptotically normal and achieves the semiparametric information bound. Furthermore, we demonstrate the nonparametric optimality of the functional estimator and construct the asymptotic simultaneous confidence band. The proposed inference procedures are evaluated by extensive simulation studies and illustrated with an application to the National Health and Nutrition Examination Survey data.

Some key words: Functional accelerated failure time model; Model identifiability; Right-censored data; Semiparametric information bound; Sieve maximum likelihood.

1. Introduction

Functional data are typically regarded as a realization of an underlying stochastic process; that is, $Z(\cdot)\colon \mathbb{I}_0 \to \mathbb{R}$ is a stochastic process indexed with a compact set $\mathbb{I}_0$. Technological advancement has drastically increased the capability of capturing and storing functional data, which have become increasingly important in many fields such as medicine, economics, engineering and chemometrics. With growing awareness of its importance, a vast amount of literature has been devoted to the development of functional data analysis, of which functional regression analysis has received the most attention in application and methodology development (Morris, 2015). The functional linear model was first introduced by Ramsay & Dalzell (1991), and later extended to various nonlinear functional models, including the generalized functional linear model (Marx & Eilers, 1999), the functional polynomial model (Yao & Müller, 2010) and the functional generalized additive model (McLean et al., 2014). For prediction and estimation, the approaches based on the functional principal component analysis have been popularized (Cardot et al., 1999; Müller & Stadtmüller, 2005; Yao et al., 2005; Crainiceanu et al., 2009). Moreover, other methods including different basis functions and regularization approaches have also been well developed. Additional details and insights of functional data analysis are discussed in the monographs by Ramsay & Silverman (2005) and Ferraty & Vieu (2006), as well as the reviews by Morris (2015) and Wang et al. (2016).

Recently, functional data have received a substantial amount of attention in the realm of survival analysis. Chen et al. (2011) proposed the functional Cox model and Kong et al. (2018) extended this model to the functional principal component analysis approach. Qu et al. (2016) studied the model estimation under a more general reproducing kernel Hilbert space framework, where they derived the asymptotic properties of the maximum partial likelihood estimator and established the asymptotic normality and efficiency for the finite-dimensional estimator. Furthermore, Hao et al. (2021) derived the asymptotic joint distribution of finite- and infinite-dimensional estimators. Cui et al. (2021) proposed the additive functional Cox model. Jiang et al. (2020) studied a functional censored quantile regression model to characterize the time-varying relationship between time-to-event outcomes and functional covariates. Yang et al. (2020) considered the functional linear regression for right-censored data and developed a penalized least-squares method for model estimation, while the theoretical properties of the proposed estimator have not been studied yet.

Among various survival models, the Cox proportional hazard model (Cox, 1972) has gained the most popularity in applications. However, when the proportional hazard assumption is violated, as commonly encountered in practice, the accelerated failure time (AFT) model provides a convenient and attractive alternative (Buckley & James, 1979; Miller & Halpern, 1982; Ritov, 1990; Tsiatis, 1990; Lai & Ying, 1991a,b; Ying, 1993; Jin et al., 2003, 2006; Zeng & Lin, 2007; Ding & Nan, 2011; Lin & Chen, 2013). With the transformed failure time directly regressed on the covariates, the AFT model has the advantage of straightforward interpretation inherited from the typical linear regression. To accommodate for both functional and scalar data, we consider a functional accelerated failure time (FAFT) model,

$$T = \alpha_0^{\mathrm{T}} X + \int_{\mathbb{I}_0} \beta_0(s) Z(s)\, \mathrm{d}s + \varepsilon, \tag{1}$$

where T is a failure time after a known monotone transformation, X is a p-dimensional vector of covariates, $Z(\cdot)$ is a functional covariate, α_0 is a p-dimensional parameter, $\beta_0(\cdot)$ is a functional parameter and ε is an error with an unknown distribution.

We develop a sieve maximum likelihood approach for the FAFT model with right-censored data. To investigate the asymptotic property of the sieve estimator, we need to overcome two main challenges. First, the parameters are bundled together in the loglikelihood function such that the theoretical analysis is much more difficult than the usual situations with separate parameters in objective functions. Second, the overall convergence rate of the proposed estimator is shown to be lower than the standard rate $n^{-1/2}$, which incurs considerable difficulties in deriving the asymptotic distribution of the estimator.

The main contributions are as follows.

(i) We rigorously discuss the model identifiability and provide sufficient conditions. Even for the AFT model with an unspecified error distribution, the existing statistical inference procedures are typically made by assuming the model to be identifiable.

(ii) Overcoming the challenges from bundled parameters, we establish the convergence rate of the bundled parameters, as well as separate parameters. Our theoretical development is highly nontrivial and general enough to be applicable to other bundled parameter situations.

(iii) We obtain the information bound for the finite-dimensional parameters in the semiparametric FAFT model and demonstrate the efficiency of our estimation procedure. We derive the asymptotic normality for the finite-dimensional estimator and show that it achieves the information bound asymptotically.

(iv) We establish the minimax lower bound for estimating the functional parameter and demonstrate that our functional estimator achieves it. Additionally, we introduce asymptotic simultaneous confidence bands to facilitate inference.

2. Estimation method

Let $U = \{X, Z(\cdot)\}$ denote the covariates and $\theta = \{\alpha, \beta(\cdot)\}$ denote the parameters. Define $\mu(U, \theta) = \alpha^{\mathrm{T}} X + \int_{\mathbb{I}_0} \beta(s) Z(s)\, \mathrm{d}s$. The FAFT model in (1) can be rewritten as

$$T = \mu(U, \theta_0) + \varepsilon,$$

where $\theta_0 = \{\alpha_0, \beta_0(\cdot)\}$ is the true parameter. Let R denote the censoring time after the same transformation as the failure time T. The observed survival time is $Y = \min\{T, R\}$ with censoring indicator $\Delta = I(T \leqslant R)$. Under a standard assumption that ε is independent of U and R, we subsequently have T and R being independent conditional on covariate U. Hence, the joint density function of (Y, Δ, U) is

$$f_{Y,\Delta,U}(y, \delta, u) = \lambda_0^{\delta}\{y - \mu(u, \theta_0)\} \exp[-\Lambda_0\{y - \mu(u, \theta_0)\}] H(y, \delta, u),$$

where $\lambda_0(\cdot)$ and $\Lambda_0(\cdot)$ are the hazard function and the cumulative hazard function of the error term ε, respectively, and $H(y, \delta, u)$ is a function that depends only on the distribution of U and the conditional distribution of R given U. To alleviate the positivity constraint for the hazard function, we set $g(\cdot) = \log \lambda(\cdot)$ and formulate the loglikelihood function as a function of (θ, g).

Suppose that the observations (Y_i, Δ_i, U_i), $i = 1, \ldots, n$, are independently sampled based on the FAFT model. The loglikelihood function for parameter $\xi = (\alpha, \beta, g)$ is

$$l_n(\xi) = \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i g\{Y_i - \mu(U_i, \theta)\} - \int \exp\{g(t)\} I\{Y_i - \mu(U_i, \theta) \geqslant t\}\, \mathrm{d}t\right],$$

where the parts independent of ξ are omitted. We consider estimating ξ by maximizing the loglikelihood function, for which direct estimation is infeasible. Zeng & Lin (2007) showed that the maximum of $l_n(\xi)$ does not exist even when all the covariates are scalar (i.e., no functional component). To overcome this difficulty, Zeng & Lin (2007) proposed a kernel-smoothed profile likelihood function for the estimation of regression parameters. Ding & Nan (2011) investigated the model by applying the spline method. However, neither of these approaches is applicable to the FAFT model due to the inclusion of a functional component. This inclusion introduces significant numerical challenges, particularly in the estimation of the functional parameter and its asymptotic simultaneous confidence band. Additionally, there are substantial theoretical difficulties, as establishing the convergence properties of the functional estimator and developing its asymptotic properties for inference require novel analytical approaches.

We propose an estimation approach for the FAFT model by maximizing the loglikelihood function in a sieve space. Specifically, we focus on the spline-based sieve space, where both scalar and functional parameters are estimated simultaneously as bundled together. The advantages of this spline-based sieve space are demonstrated both theoretically and numerically. The choice of sieve space is general as long as the assumptions for the theorems are satisfied.

Without loss of generality, we assume that $\mathbb{I}_0 = [0, 1]$ and that the log-hazard function g_0 is supported on $[a, b]$, as an interval of interest, where $a = \inf_{y,u}\{y - \mu(u, \theta_0)\}$ and $b = \tau < \infty$. To propose the spline-based sieve space, we first introduce some notation. For a closed interval $[c, d]$, let $\mathcal{T}_n(c, d) = \{t_i, i = 0, \ldots, m_n + 1\}$ denote a sequence of knots that partition $[c, d]$ into $m_n + 1$ subintervals, where $c \equiv t_0 < t_1 < \cdots < t_{m_n} < t_{m_n+1} \equiv d$. Let $\mathcal{S}_\ell\{\mathcal{T}_n(c, d)\}$ denote the space of splines of order $\ell \geqslant 1$ with knot sequence $\mathcal{T}_n(c, d)$, and let $q_n = m_n + \ell$. According to Corollary 4.10 of Schumaker (1981), for any function $\phi \in \mathcal{S}_\ell\{\mathcal{T}_n(c, d)\}$, there exists a q_n-dimensional vector γ such that $\phi = B_n^{\mathrm{T}}\gamma$, where $B_n = (b_1, \ldots, b_{q_n})^{\mathrm{T}}$ is a vector of B-spline basis functions. Following Shen & Wong (1994), we consider the space

$$\Phi_n(\ell, c, d) = \{B_n^{\mathrm{T}}\gamma : \|\gamma\|_\infty \leqslant c_n\},$$

where c_n grows with n slowly enough. Define $\mathcal{F}_n^\omega = \Phi_n(\lceil\omega\rceil + 1, 0, 1)$ and $\mathcal{G}_n^\kappa = \Phi_n(\lceil\kappa\rceil + 1, a, b)$, where $\lceil x\rceil$ is the ceiling function, ω and κ respectively represent the smoothness of β_0 and g_0 in Condition 4 given in the next section. The sieve space is defined as

$$\Xi_n = \mathcal{B} \times \mathcal{F}_n^\omega \times \mathcal{G}_n^\kappa = \{\xi = (\alpha, \beta, g) : \alpha \in \mathcal{B}, \beta \in \mathcal{F}_n^\omega, g \in \mathcal{G}_n^\kappa\},$$

where $\mathcal{B}$ is a known compact set. We study the sieve maximum likelihood estimator (MLE)

$$\hat{\xi}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{g}_n) = \underset{\xi \in \Xi_n}{\arg\max}\; l_n(\xi).$$

Under Condition 5 in the next section, this is equivalent to finding a $(p + m_n^{\omega} + s_n^{\kappa})$-dimensional vector $(\alpha^{\mathrm{T}}, \gamma^{\beta\mathrm{T}}, \gamma^{g\mathrm{T}})^{\mathrm{T}}$ that maximizes the loglikelihood function by taking $\beta = \gamma^{\beta\mathrm{T}} B_n^{\beta}$ and $g = \gamma^{g\mathrm{T}} B_n^{g}$, where B_n^{β} and B_n^{g} are the vectors of B-spline basis functions of $\mathcal{F}_n^{\omega}$ and $\mathcal{G}_n^{\kappa}$, respectively. Therefore, for the sieve MLE $\hat{\xi}_n$, there exists an m_n^{ω}-dimensional vector $\hat{\gamma}_n^{\beta}$ and an s_n^{κ}-dimensional vector $\hat{\gamma}_n^{g}$ such that $\hat{\beta}_n = \hat{\gamma}_n^{\beta\mathrm{T}} B_n^{\beta}$ and $\hat{g}_n = \hat{\gamma}_n^{g\mathrm{T}} B_n^{g}$. The estimate of $(\alpha^{\mathrm{T}}, \gamma^{\beta\mathrm{T}}, \gamma^{g\mathrm{T}})^{\mathrm{T}}$ is obtained by maximizing the loglikelihood function,

$$l_n(\xi) = \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i \gamma^{g\mathrm{T}} B_n^{g}\big(Y_i - \mu(U_i,\theta)\big) - \int \exp\{\gamma^{g\mathrm{T}} B_n^{g}(t)\} I\{Y_i - \mu(U_i,\theta) \geqslant t\}\, \mathrm{d}t\right],$$

where $\mu(U_i,\theta) = \alpha^{\mathrm{T}} X_i + \int_{\mathbb{I}_0} \gamma^{\beta\mathrm{T}} B_n^{\beta}(s) Z_i(s)\, \mathrm{d}s$. We apply an iterative estimation procedure that utilizes multiple randomly selected initial values to ensure the numerical convergence of the sieve MLE, addressing the local concavity of the sieve loglikelihood function. A detailed description of the numerical implementation is provided in the Supplementary Material.

3. Theoretical results

Let r_1 be a positive integer and $r_2 \in (0, 1]$ such that $r = r_1 + r_2$. Define $\mathcal{F}_r(\mathbb{I})$ as a class of functions on $\mathbb{I}$ whose r_1th derivative exists and satisfies the Lipschitz condition of order r_2:

$$\mathcal{F}_r(\mathbb{I}) = \{f : \mathbb{I} \to \mathbb{R} \mid f \text{ has bounded derivatives } f^{(j)},\ j = 1, \ldots, r_1, \\ \text{and } |f^{(r_1)}(s) - f^{(r_1)}(t)| \leqslant L|s - t|^{r_2} \text{ for } s, t \in \mathbb{I}\}$$

with L a positive constant. Define $r_\theta = Y - \mu(U, \theta)$ and $r_{\theta_0} = Y - \mu(U, \theta_0)$. To establish the asymptotic properties of the proposed estimator, we need the following conditions.

Condition 1. The true parameter α_0 belongs to the interior of a compact set $\mathcal{B} \subseteq \mathbb{R}^p$.

Condition 2.

(i) The covariate X takes values in a bounded subset $\mathcal{X} \subseteq \mathbb{R}^p$ and satisfies $E(X) = 0$, and $E(XX^{\mathrm{T}})$ is nonsingular.
(ii) The functional covariate Z takes values in the $L_2(\mathbb{I}_0)$ space. The L_2 norm of Z is bounded almost surely and $E(Z) = 0$.

Condition 3. There is a truncation time $\tau < \infty$ such that, for some constant δ, $P(r_{\theta_0} > \tau \mid U) \geqslant \delta > 0$ almost surely with respect to the probability measure of U. This implies that $\Lambda_0(\tau) \leqslant -\log\delta < \infty$.

Condition 4. The true functional parameter β_0 belongs to $\mathcal{F}^{\omega} \equiv \mathcal{F}_{\omega}([0, 1])$, where $\omega \geqslant 1$, and the closed support set of β_0 belongs to $(0, 1)$. The true log-hazard function g_0 belongs to $\mathcal{G}^{\kappa} \equiv \mathcal{F}_{\kappa}([a, b])$, where $\kappa \geqslant 3$, and g_0 is a nonconstant and nonperiodic function.

Condition 5.

(i) For $\mathcal{F}_n^{\omega}$, let $\mathcal{T}_n(0, 1) = \{t_i,\ i = 0, \ldots, m_n + 1\}$ denote the corresponding knot sequence. The maximum spacing of the knots satisfies $\max_{1\leqslant i\leqslant m_n+1} |t_i - t_{i-1}| = O(n^{-\nu})$ and $m_n = O(n^{\nu})$ for $\nu \in (0, 0.5)$. Define $m_n^{\omega} = m_n + \lceil\omega\rceil + 1$.

(ii) For $\mathcal{G}_n^\kappa$, let $\mathcal{T}_n(a,b) = \{t_i,\ i = 0,\dots,s_n+1\}$ denote the corresponding knot sequence. The maximum spacing of the knots satisfies $\max_{1\leqslant i\leqslant s_n+1} |t_i - t_{i-1}| = O(n^{-q})$ and $s_n = O(n^q)$ for $q \in (0, 0.5)$. Define $s_n^\kappa = s_n + \lceil\kappa\rceil + 1$.

Condition 6. For some $\eta \in (0,1)$, $\text{var}\{\mu(U,\theta) \mid r_{\theta_0}\} \geqslant \eta E\{\mu(U,\theta)^2 \mid r_{\theta_0}\}$ holds almost surely for any $\theta \in \mathcal{B} \times \mathcal{F}^\omega$.

Condition 7. Conditional on $\Delta = 1$ and $\varepsilon = t$, the conditional densities of X and $Z(s)$ have bounded jth derivatives with respect to t for $j = 1,\dots,\lfloor\kappa\rfloor$.

Condition 8. There exists a nonnegative integer ϱ and a signed measure ζ with bounded variation such that, for any ϕ with a continuous $(\varrho+1)$th derivative and $\phi^{(j)}(1) = \phi^{(j)}(0) = 0$ for $j = 0,\dots,\varrho$, it holds that $\int_0^1\int_0^1 \phi^{(\varrho+1)}(s)\phi^{(\varrho+1)}(t)C(s,t)\,\mathrm{d}s\,\mathrm{d}t = \int_0^1\int_0^1 \phi(s)\phi(t)\,\mathrm{d}\zeta(s,t)$, where $C(s,t) = E\{Z(s)Z(t)\}$.

Conditions 1 and 4 place restrictions on the parameter space, which require α_0 not to be on the boundary of the parameter space as well as g_0 and β_0 satisfying certain smoothness conditions. Such a smoothness assumption is often adopted in nonparametric estimation and can be easily satisfied. Similar regularity conditions are commonly imposed in the literature (Huang, 1999; Zeng & Lin, 2007; Ding & Nan, 2011). Condition 2 places a boundedness restriction on the covariates, which is also assumed by Qu et al. (2016). Condition 3 is the same as that in Ding & Nan (2011). Condition 5 is a regularity condition about the spline-based sieve space. Condition 6 guarantees that the convergence rate of each parameter can be derived from the result of the bundled parameter $g(r_\theta)$. Furthermore, it is feasible to impose this condition within a small neighbourhood of θ_0, while maintaining the validity of the main results. Condition 7 is required to show that the score functions in the least favourable direction are nearly zero, which is a key step in the derivation of the asymptotic normality of the scalar estimator. Condition 8 is essential in establishing the convergence rate of the functional estimator. Let $C^{(k,l)}(s,t) = \partial^{k+l}C(s,t)/(\partial s^k \partial t^l)$. Condition 8 requires a level of smoothness in $C(s,t)$, which is satisfied when $C^{(\varrho+1,\varrho+1)}(s,t)$ exists. Importantly, Condition 8 is less restrictive than differentiability, encompassing cases where $C(s,t)$ can be a generalized function. For more information on generalized functions and signed measures, we refer the reader to Stein & Shakarchi (2011) and Folland (1999, Ch. 3), respectively. Furthermore, it is worth noting that Condition 8 is less stringent than the Sacks–Ylvisaker condition of order ϱ, which is commonly assumed in the functional regression literature (Yuan & Cai, 2010). The Sacks–Ylvisaker condition is stated in the Supplementary Material for completeness.

Define the parameter space as

$$\Xi = \mathcal{B} \times \mathcal{F}^\omega \times \mathcal{G}^\kappa = \{\xi = (\alpha,\beta,g) \colon \alpha \in \mathcal{B},\ \beta \in \mathcal{F}^\omega,\ g \in \mathcal{G}^\kappa\}.$$

The sequence of spaces $\{\Xi_n\}_{n\geqslant 1}$ approximates to Ξ and is called a sieve. For notational simplicity, we also denote $\xi = (\alpha,\beta,g)$ by $\xi = (\theta,g)$ with $\theta = (\alpha,\beta)$. For the parameter space Ξ, we define the pseudometric $d(\cdot,\cdot)$ as

$$d(\xi_1,\xi_2) = P\{\mu(U,\theta_1-\theta_2)^2\}^{1/2} + \|g_1 - g_2\|_{\mathcal{G}},$$

where P denotes the probability measure with $Pf = \int f\,\mathrm{d}P$ and $\|g\|_{\mathcal{G}}^2 = P(\Delta g\{r_{\theta_0}\}^2)$. A pseudometric is a distance function that is weaker than a metric and may assign a value of zero to nonidentical points. As shown in Theorem 2 below, based on this pseudometric, the accuracy of $\hat{\alpha}_n$ can be measured by the Euclidean norm $|\hat{\alpha}_n - \alpha_0|$, and the accuracy of $\hat{\beta}_n$ can be measured by $\|\hat{\beta}_n - \beta_0\|_C$, where

$$\|\beta\|_C^2 = \iint \beta(s)C(s,t)\beta(t)\,\mathrm{d}s\,\mathrm{d}t \quad \text{with} \quad C(s,t) = E\{Z(s)Z(t)\}.$$

This measurement has been widely used for functional linear models (Cai & Yuan, 2012) and has also been investigated in the functional Cox model (Qu et al., 2016).

Direct investigation of the estimator under $d(\cdot,\cdot)$ is challenging because parameters g and θ are bundled together, which makes the information of separate parameters difficult to derive. To overcome this difficulty, we propose first investigating the space for bundled parameters, and then applying the results to study the parameters separately. Furthermore, Proposition 2 below provides sufficient conditions to ensure parameter identifiability under $d(\cdot,\cdot)$. This pseudometric-based analysis method is specifically tailored for the FAFT model, effectively addressing the complexities associated with bundled parameters and the inclusion of a functional parameter. Specifically, we define the space of bundled parameters as follows. For any given θ, let $r_\theta(\cdot)$ be a mapping from $\mathbb{R} \times \mathcal{X} \times L_2(\mathbb{I}_0)$ to $\mathbb{R}$ defined by $r_\theta(y,u) = y - \mu(u,\theta)$ for any $y \in \mathbb{R}$ and $u \in \mathcal{X} \times L_2(\mathbb{I}_0)$. The space of bundled parameters is defined by

$$\mathcal{A} = \{g[r_\theta(\cdot)] \colon \mathbb{R} \times \mathcal{X} \times L_2(\mathbb{I}_0) \mapsto \mathbb{R} \mid (\theta, g) \in \Xi\}.$$

For notational simplicity, we denote $g\{r_\theta(\cdot)\}$ by $g(r_\theta)$. As $g(r_\theta)$ can also represent a random variable, the meaning of $g(r_\theta)$ is according to the context. To measure the difference between any two elements in $\mathcal{A}$, we consider the pseudometric

$$\|g_1(r_{\theta_1}) - g_2(r_{\theta_2})\|_{\mathcal{A}} = P[\Delta\{g_1(r_{\theta_1}) - g_2(r_{\theta_2})\}^2]^{1/2}.$$

We first derive the efficient score function and the information bound.

PROPOSITION 1. *Under Conditions 1–4 and 6, the efficient score function for estimating α_0 in the FAFT model is*

$$\dot{l}^\star_{\alpha_0} = \int \left\{ -\dot{g}_0(t)X + \dot{g}_0(t)\int_0^1 b^\star(s)Z(s)\,\mathrm{d}s - \phi^\star(t) \right\} \mathrm{d}M(t),$$

where $M(t) = \Delta I(r_{\theta_0} \leqslant t) - \int_{-\infty}^t I(r_{\theta_0} \geqslant u)\lambda_0(u)\,\mathrm{d}u$, $\dot{g}_0$ denotes the first derivative of g_0 and $(b^\star, \phi^\star)$ is a solution that minimizes

$$E\left[\Delta\left| -\dot{g}_0(r_{\theta_0})X + \dot{g}_0(r_{\theta_0})\int_0^1 b(s)Z(s)\,\mathrm{d}s - \phi(r_{\theta_0})\right|^2\right].$$

The information bound for estimation of α_0 is

$$I(\alpha_0) = E[\dot{l}^{\star\otimes 2}_{\alpha_0}] = E\left[\Delta\left\{ -\dot{g}_0(r_{\theta_0})X + \dot{g}_0(r_{\theta_0})\int_0^1 b^\star(s)Z(s)\,\mathrm{d}s - \phi^\star(r_{\theta_0})\right\}^{\otimes 2}\right],$$

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

Remark 1. Detailed discussions on the explicit forms of $b^\star$ and $\phi^\star$ in the efficient score function can be found in § 1.3 of the Supplementary Material.

PROPOSITION 2. *For any* $\xi^\star = (\theta^\star, g^\star)$ *that maximizes* $El_n(\xi)$*, it holds that* $\|g^\star(r_{\theta^\star}) - g_0(r_{\theta_0})\|_{\mathcal{A}} = 0$. *Under Conditions 1–4, if* $I(\alpha_0)$ *is nonsingular, we further establish* $g^\star = g_0$, $\alpha^\star = \alpha_0$ *and* $\|\beta^\star - \beta_0\|_C = 0$. *Moreover,* $\beta^\star = \beta_0$ *holds when the covariance function* $C(s,t)$ *satisfies the Sacks–Ylvisaker conditions of order k, with integer* $1 \leqslant k < \lceil \omega \rceil - 1$, $C^{(0,i)}(\cdot, 0) = 0$ *for* $i = 0, 1, \ldots, k-1$ *and the jth derivative of* $\beta^\star$ *equals* 0 *at points* 0 *and* 1 *for* $j = 0, \ldots, \lceil \omega \rceil - 1$.

Remark 2. The result of Proposition 2 provides sufficient conditions to guarantee the identifiability of model (1). These conditions entail specific smoothness requirements for $C(s,t)$ and β_0 and require that g_0 is nonconstant and nonperiodic. Such identifiability is the key to statistical inference. However, in the AFT model, the accelerated hazard regression model and the longitudinal data model, statistical inference is often based on a direct assumption of model identifiability (Zeng & Lin, 2007; Zhao et al., 2017; Kong et al., 2018).

We next give the convergence rate of the bundled estimator $\hat{g}_n(r_{\hat{\theta}_n})$.

THEOREM 1. *Under Conditions 1–5 and 8, we have*

$$\|\hat{g}_n(r_{\hat{\theta}_n}) - g_0(r_{\theta_0})\|_{\mathcal{A}} = O_p(n^{-c}),$$

where $c = \min[\nu(\omega + \varrho + 1), \kappa q, (1 - \max\{\nu, q\})/2]$.

Next, the consistency of each estimator is derived separately. For $\hat{g}_n$, we first show that the sequence $\{\hat{g}_n\}_{n \geqslant 1}$ is precompact and then apply the Arzelà–Ascoli theorem. Let $\dot{\hat{g}}_n$ denote the first derivative of $\hat{g}_n$. The result indicates that both $\hat{g}_n$ and $\dot{\hat{g}}_n$ converge in probability under the supremum norm $\|\cdot\|_\infty$. Similar approaches were studied by Murphy et al. (1999) and Kuchibhotla & Patra (2020); however, they were rarely applied to survival analysis. Next, to derive the consistency of $\hat{\beta}_n$, we define an integral operator of $C(s,t)$ and derive the consistency based on the compactness of the operator. When $I(\alpha_0)$ is nonsingular, the accuracy of $\hat{\alpha}_n$ and $\hat{\beta}_n$ can be measured by $|\cdot|$ and $\|\cdot\|_C$, respectively.

THEOREM 2. *Suppose that Conditions 1–6 and 8 hold. If* $I(\alpha_0)$ *is nonsingular then the following statements hold.*

(i) *(Consistency.) We have* $\|\hat{g}_n - g_0\|_\infty + \|\dot{\hat{g}}_n - \dot{g}_0\|_\infty = o_p(1)$, $|\hat{\alpha}_n - \alpha_0| = o_p(1)$ *and* $\|\hat{\beta}_n - \beta_0\|_C = o_p(1)$.

(ii) *(Convergence rate.) Let* $c = \min[\nu(\omega + \varrho + 1), \kappa q, (1 - \max\{\nu, q\})/2]$. *We have*

$$d(\hat{\xi}_n, \xi_0) = O_p(n^{-c}).$$

(iii) *We have* $|\hat{\alpha}_n - \alpha_0| + \|\hat{\beta}_n - \beta_0\|_C + \|\hat{g}_n - g_0\|_{\mathcal{G}} = O_p(n^{-c})$.

When $\nu = 1/(2\omega + 2\varrho + 3)$ and $q = 1/(1 + 2\kappa)$, Theorem 2(ii) implies that the convergence rate of the sieve estimator $\hat{\xi}_n$ could reach the slower rate between $n^{-(\omega+\varrho+1)/(2\omega+2\varrho+3)}$ and $n^{-\kappa/(1+2\kappa)}$. Combining the derivation with Theorem 2(iii), it can be shown that, when g_0 has a weaker smoothness property, $\hat{g}_n$ could reach the optimal rate in nonparametric regression, as given by Stone (1982), and when β_0 and $C(s,t)$ have weaker smoothness properties, $\hat{\beta}_n$ could reach the optimal rate, as shown in Theorem 4 below. Next, we derive the convergence rate of scalar estimator $\hat{\alpha}_n$ and show that it could reach $n^{-1/2}$.

THEOREM 3. *Suppose that Conditions 1–8 hold and that the information bound $I(\alpha_0)$ is nonsingular. Let $c_1 = \min[\nu\omega, \kappa q, (1-\max\{\nu, q\})/2]$. When $c_1 - q > 1/4$, we have*

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\text{D}} N(0, \Sigma),$$

where $\Sigma = I(\alpha_0)^{-1}$ and '$\xrightarrow{\text{D}}$' denotes convergence in distribution.

The above result shows that $\hat{\alpha}_n$ achieves the information bound. Therefore, it is asymptotically efficient among all the regular estimators. In Theorem 3, the condition on ν and q is relatively mild and can be conveniently satisfied in most cases. For example, when $\nu = q$ and $\omega = \kappa$, the condition is satisfied if $1/(2(1+\kappa)) < q < 1/(2\kappa)$ for $\kappa \geqslant 3$. We also establish the minimax lower bound of the convergence rate for estimating β_0 and show that it reaches $n^{-(\omega+\varrho+1)/(2\omega+2\varrho+3)}$.

THEOREM 4. *Under the conditions of Theorem 2, there is a positive number a such that*

$$\lim_{n\to\infty} \inf_{\hat{\beta}} \sup_{\xi\in\Xi} P_\xi\{\|\hat{\beta} - \beta\|_C \geqslant a n^{-(\omega+\varrho+1)/(2\omega+2\varrho+3)}\} = 1,$$

where the infimum is taken over all possible estimators $\hat{\beta}$ based on the observed data.

We then develop asymptotic simultaneous confidence bands (ASCBs) for $\hat{\beta}_n$. Let $\xi_n(s)$ be a Gaussian process with $E[\xi_n(s)] \equiv 0, \text{var}[\xi_n(s)] \equiv 1$ and covariance matrix

$$\text{cov}\{\xi_n(s), \xi_n(s')\} = \frac{B_n(s)^{\mathrm{T}} D_n^{-1} B_n(s')}{\{B_n(s)^{\mathrm{T}} D_n^{-1} B_n(s)\}^{1/2}\{B_n(s')^{\mathrm{T}} D_n^{-1} B_n(s')\}^{1/2}} \quad \text{for any } s, s' \in [0,1].$$

Here $B_n = (b_1, \ldots, b_{m_n^\omega})^{\mathrm{T}}$ is the vector of B-spline basis functions for $\mathcal{F}_n^\omega$ and $D_n = (\bar{d}_{ij})$ is an $m_n^\omega \times m_n^\omega$ matrix. Specifically, $\bar{d}_{ij}$ takes the value of the expected negative second derivative of the loglikelihood function at directions $(\tilde{h}_{1i}, b_i, \tilde{h}_{3i})$ and $(\tilde{h}_{1j}, b_j, \tilde{h}_{3j})$. Here $\tilde{h}_{1i}$ and $\tilde{h}_{3i}$ for $i = 1, \ldots, m_n^\omega$ are obtained through the minimization problem

$$\min_{h_1\in\mathcal{B},\, h_3\in\mathcal{G}_n^\kappa} P\left[\Delta\left| -\dot{g}_0(r_{\theta_0})\int_0^1 b_i(s)Z(s)\,\mathrm{d}s + \dot{g}_0(r_{\theta_0})X^{\mathrm{T}}h_1 - h_3(r_{\theta_0})\right|^2\right].$$

We define the $100(1-\alpha)$th percentile of the absolute maxima distribution of $\xi_n(s)$ as $Q_n(\alpha)$, which satisfies

$$P\left\{\sup_{s\in[0,1]} |\xi_n(s)| \leqslant Q_n(\alpha)\right\} = 1 - \alpha.$$

We now present the theorem that establishes the ASCBs. Let $K(s) = P\{Z(s) \mid X\}$ and $K(s,t) = P[\{Z(s) - K(s)\}\{Z(t) - K(t)\}]$. Define $K^{(k,l)}(s,t) = \partial^{k+l}K(s,t)/(\partial s^k \partial t^l)$.

THEOREM 5. *Assume that $K(s,t)$ satisfies the Sacks–Ylvisaker conditions of order k, with the integer $1 \leqslant k < \lceil\omega\rceil - 1$, and $K^{(0,i)}(\cdot, 0) = 0$ for $i = 0, 1, \ldots, k-1$. Under Conditions 1–8, $1/2 + \nu - \min\{\nu(\omega+\varrho+1), \kappa q\} < 0$, $0 < \nu < 1/6$ and $c > \nu + q + 1/4$, we have*

$$\lim_{n\to\infty} P\left\{\sup_{s\in[0,1]}\left|\frac{\sqrt{n}\{\hat{\beta}_n(s) - \beta_0(s)\}}{\{B_n(s)^{\mathrm{T}} D_n^{-1} B_n(s)\}^{1/2}}\right| \leqslant Q_n(\alpha)\right\} = 1 - \alpha.$$

4. Simulation studies

We conducted simulation studies to evaluate the finite-sample performance of the proposed method. For generating the functional covariate, we considered a similar set-up as Qu et al. (2016), defining $Z(\cdot)$ as $Z(s) = \sum_{k=1}^{50} \xi_k U_k \phi_k(s)$, where the U_k are independently sampled from Un$[-1, 1]$, $\xi_k = (-1)^{k+1} k^{-1/2}$, $\phi_1 \equiv 1$ and $\phi_{k+1}(s) = \sqrt{2}\cos(k\pi s)$ for $k \geqslant 1$. Then, the covariance function is $C(s, t) = \sum_{k=1}^{50} \phi_k(s)\phi_k(t)/(3k)$. The functional parameter β_0 is defined as $\beta_0(s) = \sum_{k=1}^{50}(-1)^k k^{-3/2}\phi_k(s)$. The scalar covariates X_1 followed Ber(0.5) and X_2 followed $N(0, 0.5)$ truncated at ± 2. The transformed failure time T was generated from the FAFT model

$$T = X_1 + X_2 + \int_{\mathbb{I}_0} \beta_0(s) Z(s)\, \mathrm{d}s + \varepsilon.$$

We considered three error term cases: (a) $\exp(\varepsilon) \sim \text{Ex}(0.6)$; (b) $\varepsilon \sim 0.8N(0, 1) + 0.2N(0, 9)$ and (c) the extreme-value distribution with location and scale parameters equal to 0 and 2, respectively. We generated censoring time R from Un$[0, \tau]$, where τ was chosen to produce desired censoring rates. The transformed observation time was $Y = \min\{T, \log(R)\}$. We considered censoring rates of 25%, 40% and 75% and sample sizes $n = 400, 600$ and 800.

To estimate the functional parameters $\beta_0(\cdot)$ and $g_0(\cdot)$, we used B-spline functions with equally spaced interior knots and $q_n = \lfloor n^{1/4} \rfloor$, resulting in four basis functions for $n = 400$ and 600 and five basis functions for $n = 800$. Let $\{\psi_k(\cdot),\ k = 1, \ldots, q_n\}$ and $\{\eta_k(\cdot),\ k = 1, \ldots, q_n\}$ be the spline basis functions with support on $\mathbb{I}_0 = [0, 1]$ and $[a, b]$, respectively. The functional parameters $\beta_0(\cdot)$ and $g_0(\cdot)$ were approximated by $\beta(\cdot) = \sum_{k=1}^{q_n} \beta_k \psi_k(\cdot)$ and $g(\cdot) = \sum_{k=1}^{q_n} g_k \eta_k(\cdot)$, respectively. Parameter $\xi = (\alpha_1, \alpha_2, \beta, g)$ was estimated based on the loglikelihood function

$$l_n(\xi) = \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i \sum_{k=1}^{q_n} g_k \eta_k\{Y_i - \mu(U_i, \theta)\} - \int \exp\left\{\sum_{k=1}^{q_n} g_k \eta_k(t)\right\} I\{Y_i - \mu(U_i, \theta) \geqslant t\}\, \mathrm{d}t\right],$$

where $\mu(U, \theta) = \alpha_1 X_1 + \alpha_2 X_2 + \int_{\mathbb{I}_0} \sum_{k=1}^{q_n} \beta_k \psi_k(s) Z(s)\, \mathrm{d}s$. The chosen support $[a, b]$ was wide enough such that it covered all residual terms, $Y_i - \mu(U_i, \theta)$ for $i = 1, \ldots, n$. Denote the estimator by $\hat{\xi} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}, \hat{g})$. The standard errors of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ were obtained from the first two diagonal entries of $(H^{-1}/n)^{1/2}$, where H was the Hessian matrix of l_n. The estimation of D_n^{-1} for the ASCBs followed a similar method. The details and validation of this variance estimation method are provided in the Supplementary Material. For each combination of error distribution, censoring rate and sample size, the simulation was repeated 1000 times.

Table 1 reports the performance of the proposed estimates of α_1 and α_2, including the average bias, the sample standard error, the estimated standard error and the coverage probability. Evidently, both the sample standard error and estimated standard error decrease with larger sample sizes and lower censoring rates. Moreover, the bias is small and the coverage probability approximates the theoretical level of 95% across all simulation scenarios. Table 2 shows the performances of $\hat{\beta}(\cdot)$ and the bundled estimator $\hat{g}(r_{\hat{\theta}})$. The performance is reasonable for all simulation scenarios, which obviously improves as the sample size increases. Figure 1 shows the pointwise averages of $\hat{\beta}(\cdot)$, where the estimates are within close proximity of the true values for all simulation scenarios. Table 3 reports the performance of the ASCBs for $\hat{\beta}(\cdot)$ with a confidence level of 95% and $s \in [0.2, 0.8]$. The coverage

Table 1. *The performance of the proposed estimates of α_1 and α_2 with different error distributions: (a) exponential, (b) Gaussian mixture, (c) extreme value*

Error distribution	n	Censoring rate (%)	$\hat{\alpha}_1$ BIAS	$\hat{\alpha}_1$ SSE	$\hat{\alpha}_1$ ESE	$\hat{\alpha}_1$ CP	$\hat{\alpha}_2$ BIAS	$\hat{\alpha}_2$ SSE	$\hat{\alpha}_2$ ESE	$\hat{\alpha}_2$ CP
(a)	400	25	−0.012	0.210	0.198	0.957	−0.009	0.206	0.200	0.966
		40	−0.010	0.194	0.215	0.955	−0.006	0.187	0.217	0.963
		75	−0.016	0.338	0.351	0.953	0.009	0.342	0.352	0.954
	600	25	−0.002	0.175	0.161	0.939	0.002	0.176	0.162	0.946
		40	−0.003	0.163	0.175	0.943	0.003	0.163	0.176	0.951
		75	−0.012	0.281	0.283	0.940	0.001	0.283	0.283	0.950
	800	25	0.004	0.152	0.140	0.954	0.001	0.154	0.141	0.951
		40	0.009	0.139	0.153	0.956	0.008	0.141	0.154	0.947
		75	−0.001	0.258	0.242	0.901	−0.007	0.247	0.243	0.912
(b)	400	25	0.007	0.165	0.154	0.958	0.010	0.163	0.155	0.953
		40	0.007	0.149	0.164	0.947	0.014	0.151	0.165	0.948
		75	0.012	0.245	0.246	0.950	0.025	0.237	0.245	0.948
	600	25	−0.004	0.115	0.119	0.963	0.007	0.115	0.119	0.965
		40	0.003	0.124	0.126	0.956	0.013	0.124	0.127	0.961
		75	0.019	0.196	0.198	0.954	0.019	0.197	0.198	0.948
	800	25	0.008	0.103	0.105	0.967	0.005	0.114	0.106	0.959
		40	0.012	0.098	0.112	0.971	0.008	0.104	0.113	0.948
		75	0.019	0.161	0.173	0.923	0.025	0.167	0.173	0.902
(c)	400	25	−0.020	0.220	0.226	0.949	−0.020	0.216	0.228	0.956
		40	−0.014	0.218	0.231	0.953	−0.021	0.219	0.233	0.964
		75	0.001	0.268	0.288	0.957	−0.008	0.280	0.291	0.959
	600	25	−0.031	0.182	0.184	0.948	−0.029	0.186	0.185	0.946
		40	−0.020	0.185	0.188	0.938	−0.022	0.188	0.189	0.946
		75	0.001	0.232	0.236	0.948	0.001	0.228	0.237	0.965
	800	25	−0.017	0.154	0.154	0.946	−0.015	0.159	0.155	0.943
		40	−0.016	0.157	0.159	0.956	−0.015	0.161	0.160	0.947
		75	−0.001	0.193	0.203	0.937	−0.002	0.196	0.204	0.937

BIAS, bias; SSE, sample standard error; ESE, estimated standard error; CP, coverage probability.

Table 2. *The performance of the proposed estimates of $\beta_0(\cdot)$ and $g_0(r_{\theta_0})$ with different error distributions: (a) exponential, (b) Gaussian mixture, (c) extreme value*

Error distribution	n	Censoring rate = 25% $\|\hat{\beta}-\beta_0\|_C$	Censoring rate = 25% $\|\hat{g}(r_{\hat{\theta}})-g_0(r_{\theta_0})\|_{\mathcal{A}}$	Censoring rate = 40% $\|\hat{\beta}-\beta_0\|_C$	Censoring rate = 40% $\|\hat{g}(r_{\hat{\theta}})-g_0(r_{\theta_0})\|_{\mathcal{A}}$	Censoring rate = 75% $\|\hat{\beta}-\beta_0\|_C$	Censoring rate = 75% $\|\hat{g}(r_{\hat{\theta}})-g_0(r_{\theta_0})\|_{\mathcal{A}}$
(a)	400	0.204	0.159	0.222	0.158	0.378	0.158
	600	0.176	0.135	0.191	0.134	0.305	0.129
	800	0.166	0.139	0.179	0.137	0.295	0.133
(b)	400	0.166	0.318	0.183	0.300	0.296	0.210
	600	0.128	0.250	0.137	0.236	0.255	0.193
	800	0.120	0.246	0.130	0.233	0.234	0.173
(c)	400	0.222	0.218	0.226	0.194	0.330	0.169
	600	0.185	0.203	0.187	0.177	0.283	0.149
	800	0.173	0.163	0.177	0.152	0.248	0.133

probabilities under all simulation scenarios approach the target confidence level. Overall, simulation results validate that both the scalar and functional parameter estimators are consistent, and the proposed variance estimation procedure provides reasonable estimates. Furthermore, the empirical coverage probabilities are close to the theoretical level of

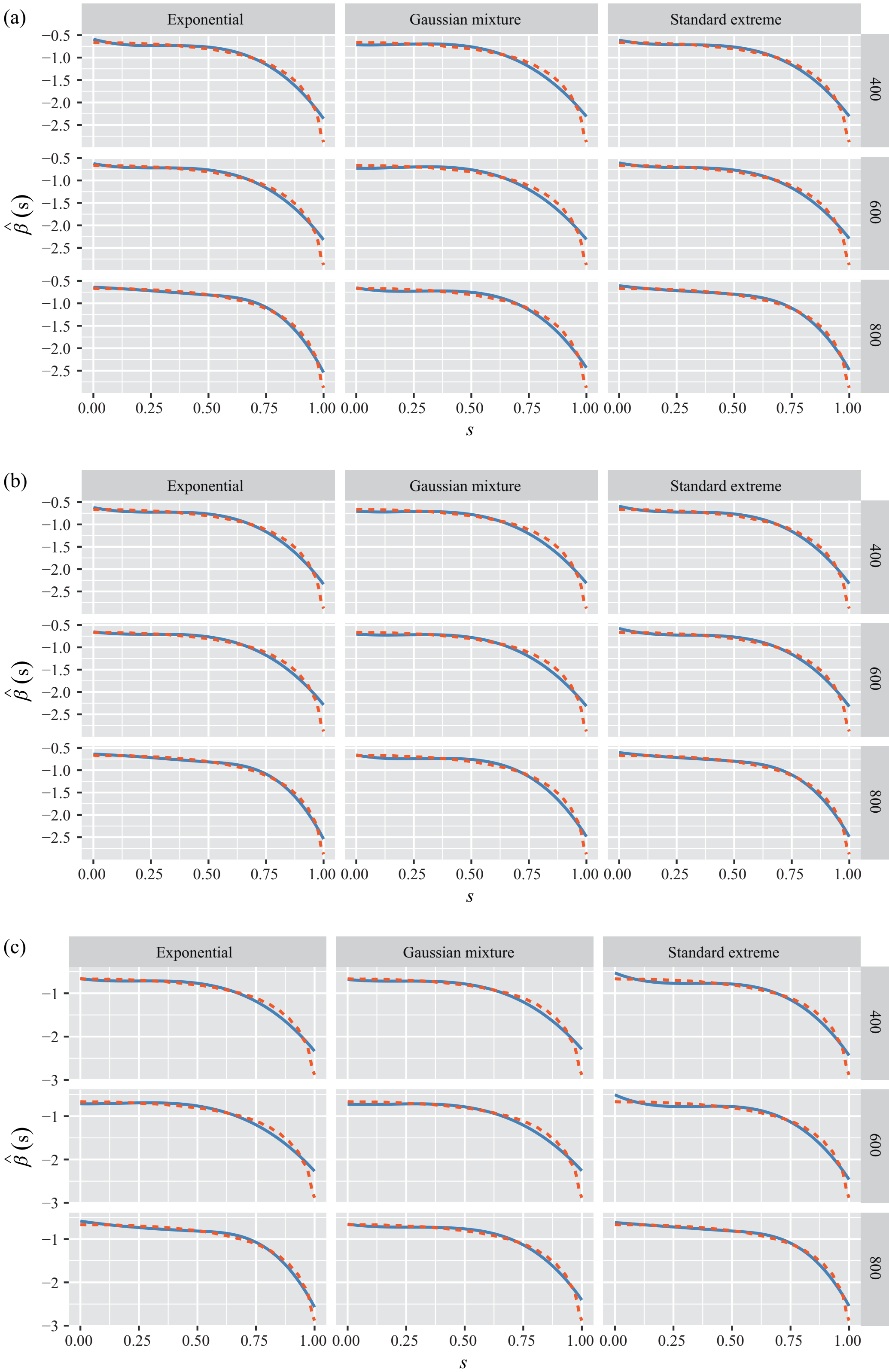

Fig. 1. Graphical displays of the pointwise averages $\hat{\beta}(\cdot)$ for censoring rates of (a) 25%, (b) 40% and (c) 75%. The dashed lines represent $\beta_0(\cdot)$, whereas the solid lines represent the pointwise averages of $\hat{\beta}(\cdot)$.

Table 3. *The performance of the asymptotic simultaneous confidence bands for $\hat{\beta}(\cdot)$ at a confidence level of 95% and $s \in [0.2, 0.8]$ with different error distributions: (a) exponential, (b) Gaussian mixture; (c) extreme value*

Error distribution	n	Censoring rate = 25%	Censoring rate = 40%	Censoring rate = 75%
(a)	400	0.959	0.957	0.966
	600	0.930	0.937	0.970
	800	0.962	0.965	0.971
(b)	400	0.919	0.907	0.956
	600	0.932	0.928	0.933
	800	0.960	0.950	0.964
(c)	400	0.963	0.967	0.927
	600	0.965	0.972	0.909
	800	0.955	0.963	0.954

95%, verifying the validity of normal approximation and the established ASCBs. Additional numerical comparisons between the FAFT and functional Cox models are presented in the Supplementary Material.

5. Application

As an illustration, we applied the proposed FAFT model to analyse data from the National Health and Nutrition Examination Survey (NHANES) (Mirel et al., 2013). NHANES was a study conducted by the Centers for Disease Control to assess the health and nutritional status of individuals in the United States, and the dataset is available in the R package `rnhanesdata` (Leroux et al., 2018; R Development Core Team, 2025). A distinct feature of the dataset was the inclusion of high-resolution physical activity and time-to-death data. The physical activity was measured using hip-worn accelerometers by participants for seven consecutive days. The recorded data were represented as minute-level log-transformed activity counts (LACs), a measure commonly adopted in the physical activity research literature (Varma et al., 2017, 2018).

The NHANES accelerometry data were collected from a total of 14 631 study participants. We excluded data that had missing mortality information, resulting in a dataset of 9590 participants, with a censoring rate of 84.8%. A common approach to include the physical activity in classical survival models was to calculate a daily average of LACs and treat the mean of these averages as a scalar covariate. However, much information would be lost during such an aggregation process. The proposed FAFT model provided a more effective alternative by averaging the LACs at each time-point over available days, smoothing the data using the procedure described by Cui et al. (2021) and treating the resulting smoothed, averaged LACs as a functional covariate. The transformed event time was the natural logarithm of the number of months until death since the day the accelerators were worn. We incorporated all available scalar covariates in the model: age, body mass index, gender, various health conditions (mobility problems, coronary heart disease, congestive heart failure, stroke, cancer, diabetes), self-reported overall health (poor or not poor), smoking status (never, former or current smoker), alcohol consumption (heavy drinker or not heavy drinker), employment status, educational attainment (less than high school, high school or more than high school), poverty-income ratio (< 1 or $\geqslant 1$) and race (White, Black, Mexican American, other Hispanic or other). Among the 9590 participants, 2507 had covariates with missing values. We imputed these missing values using the modes for qualitative variables and the means for quantitative variables, where the qualitative variables are binary, taking the value 1 for yes and 0 for no. In the analysis, we standardized

Table 4. *Estimation results of regression coefficients for scalar covariates in the NHANES data analysis, where participants with missing mortality information were excluded and missing covariates were imputed ($n = 9590$, censoring rate $= 84.8\%$)*

Covariates	$\hat{\alpha}$	SE	t-statistic	p-value
Age	−0.432	0.030	−14.502	< 0.001
Body mass index	0.050	0.035	1.421	0.155
Gender (female = 1, male = 0)	0.390	0.049	7.911	< 0.001
Mobility problem	−0.568	0.050	−11.362	< 0.001
Diabetes	−0.156	0.060	−2.587	0.010
Coronary heart disease	−0.075	0.080	−0.934	0.350
Congestive heart failure	−0.581	0.088	−6.604	< 0.001
Stroke	−0.347	0.080	−4.327	< 0.001
Cancer	−0.482	0.062	−7.783	< 0.001
Overall health is poor	−0.220	0.089	−2.468	0.014
Former/current smoker	−0.083	0.046	−1.783	0.075
Heavy drinker	−0.120	0.094	−1.272	0.204
Employed	0.655	0.062	10.563	< 0.001
High school	0.162	0.058	2.787	0.005
More than high school	0.302	0.055	5.444	< 0.001
Poverty-income ratio ⩾ 1	−0.066	0.049	−1.343	0.179
White	−0.217	0.137	−1.587	0.112
Black	0.058	0.143	0.407	0.684
Mexican American	−0.069	0.144	−0.474	0.635
Other Hispanic	0.155	0.220	0.706	0.480

continuous variables for numerical stability and utilized absolute values of the standardized body mass index, as deviations from the average body mass index may influence mortality. We centred the functional covariate by subtracting the LACs from the pointwise averages. We adopted cubic spline functions to estimate the functional coefficient with $q_n = \lfloor n^{1/4} \rfloor = 7$ basis functions using equally spaced knots.

Table 4 presents a summary of the estimated regression coefficients for the scalar covariates, where the t-statistic is defined as the ratio of the estimate to the corresponding estimated standard error. It reveals that patients' existing health status (such as mobility problems, diabetes, congestive heart failure, stroke and cancer) was negatively associated with mortality, while educational attainment was positively associated with mortality. Figure 2 shows the estimated functional coefficient $\hat{\beta}(\cdot)$ and the corresponding 95% asymptotic simultaneous confidence band, which indicates a functional association between physical activity and patient mortality. In particular, regular physical activity from 11:30 am to 9 pm was associated with a lower risk of death.

We also experimented with two additional data-processing methods: excluding participants with either missing mortality information or missing scalar covariates ($n = 7083$, censoring rate $= 85.9\%$), and excluding those with either missing information or fewer than seven days of accelerometer wear time ($n = 979$, censoring rate $= 73.4\%$). Additionally, we performed a goodness-of-fit evaluation to validate the suitability of the FAFT model for the data application under all three data-processing methods. The results of these analyses are provided in the Supplementary Material.

Acknowledgement

The authors would like to thank the editor, the associate editor and the three reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. This research was supported in part by the Research Grant Council of Hong Kong and The Hong Kong Polytechnic University.

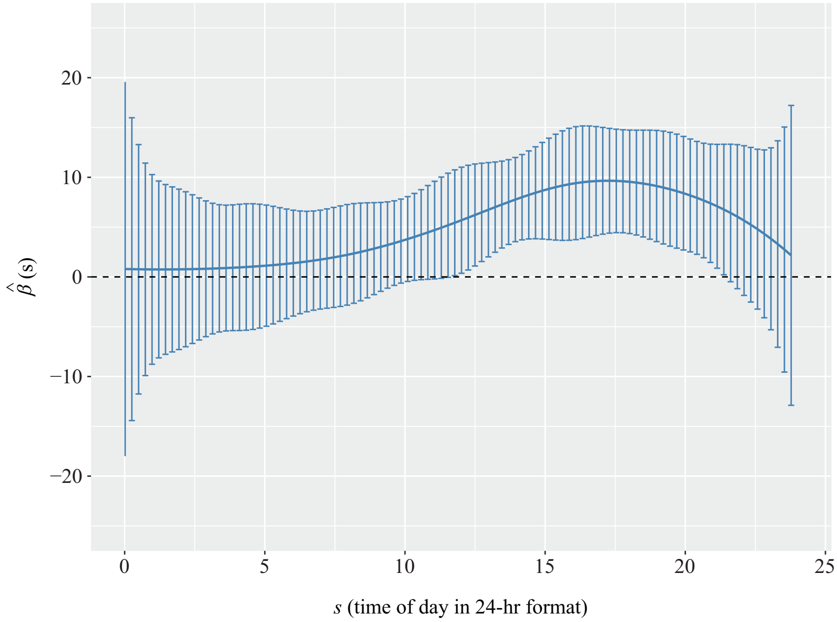

Fig. 2. Estimated functional coefficient $\hat{\beta}(\cdot)$ for the NHANES data, where participants with missing mortality information were excluded and missing covariates were imputed ($n = 9590$, censoring rate $= 84.8\%$).

Supplementary material

The Supplementary Material provides all technical details and proofs, as well as a detailed description of the numerical implementation procedure. It also includes additional numerical simulations and results from the data application.

References

Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–36.

Cai, T. T. & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Am. Statist. Assoc.* **107**, 1201–16.

Cardot, H., Ferraty, F. & Sarda, P. (1999). Functional linear model. *Statist. Prob. Lett.* **45**, 11–22.

Chen, K., Chen, K., Müller, H. G. & Wand, J. L. (2011). Stringing high-dimensional data for functional analysis. *J. Am. Statist. Assoc.* **106**, 275–84.

Cox, D. R. (1972). Regression models and life-tables. *J. R. Statist. Soc.* B. **34**, 187–220.

Crainiceanu, C. M., Staicu, A. M. & Di, C. Z. (2009). Generalized multilevel functional regression. *J. Am. Statist. Assoc.* **104**, 1550–61.

Cui, E., Crainiceanu, C. M. & Leroux, A. (2021). Additive functional Cox model. *J. Comp. Graph. Statist.* **30**, 780–93.

Ding, Y. & 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. *Ann. Statist.* **39**, 3032–61.

Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer.

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York: John Wiley and Sons.

Hao, M., Liu, K. Y., Xu, W. & Zhao, X. (2021). Semiparametric inference for the functional Cox model. *J. Am. Statist. Assoc.* **116**, 1319–29.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27**, 1536–63.

Jiang, F., Cheng, Q., Yin, G. & Shen, H. (2020). Functional censored quantile regression. *J. Am. Statist. Assoc.* **115**, 931–44.

Jin, Z., Lin, D. Y., Wei, L. J. & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–53.

Jin, Z., Lin, D. Y. & Ying, Z. (2006). On least-squares regression with censored data. *Biometrika* **93**, 147–61.
Kong, D., Ibrahim, J. G., Lee, E. & Zhu, H. (2018). FLCRM: functional linear Cox regression model. *Biometrics* **74**, 109–17.
Kong, S., Nan, B., Kalbfleisch, J. D., Saran, R. & Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *J. Am. Statist. Assoc.* **113**, 357–68.
Kuchibhotla, A. K. & Patra, R. K. (2020). Efficient estimation in single index models through smoothing splines. *Bernoulli* **26**, 1587–618.
Lai, T. L. & Ying, Z. (1991a). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531–56.
Lai, T. L. & Ying, Z. (1991b). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370–402.
Leroux, A., Crainiceanu, C., Smirnova, E. & Cao, Q. (2018). rnhanesdata: NHANES accelerometry data pipeline. R package version 1.
Lin, Y. & Chen, K. (2013). Efficient estimation of the censored linear regression model. *Biometrika* **100**, 525–30.
Marx, B. D. & Eilers, P. H. (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics* **41**, 1–13.
McLean, M. W., Hooker, G., Staicu, A. M., Scheipl, F. & Ruppert, D. (2014). Functional generalized additive models. *J. Comp. Graph. Statist.* **23**, 249–69.
Miller, R. & Halpern, J. (1982). Regression with censored data. *Biometrika* **69**, 521–31.
Mirel, L. B., Mohadjer, L. K., Dohrmann, S. M., Clark, J., Burt, V. L., Johnson, C. L. & Curtin, L. R. (2013). National Health and Nutrition Examination Survey: estimation procedures, 2007–2010. *Vital Health Statist.* **159**, 1–17.
Morris, J. S. (2015). Functional regression. *Ann. Rev. Statist. Appl.* **2**, 321–59.
Müller, H. G. & Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774–805.
Murphy, S. A., van der Vaart, A. W. & Wellner, J. A. (1999). Current status regression. *Math. Meth. Statist.* **8**, 407–25.
Qu, S., Wang, J. L. & Wang, X. (2016). Optimal estimation for the functional Cox model. *Ann. Statist.* **44**, 1708–38.
R Development Core Team (2025). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.
Ramsay, J. O. & Dalzell, C. J. (1991). Some tools for functional data analysis. *J. R. Statist. Soc.* B. **53**, 539–72.
Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*. New York: Springer.
Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303–28.
Schumaker, L. L. (1981). *Spline Functions: Basic Theory*. New York: John Wiley and Sons.
Shen, X. & Wong, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22**, 580–615.
Stein, E. M. & Shakarchi, R. (2011). *Functional Analysis: Introduction to Further Topics in Analysis*. Princeton, NJ: Princeton University Press.
Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040–53.
Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–72.
Varma, V. R., Dey, D., Leroux, A., Di, J., Urbanek, J., Xiao, L. & Zipunnikov, V. (2017). Re-evaluating the effect of age on physical activity over the lifespan. *Prev. Med.* **101**, 102–8.
Varma, V. R., Dey, D., Leroux, A., Di, J., Urbanek, J., Xiao, L. & Zipunnikov, V. (2018). Total volume of physical activity: TAC, TLAC or TAC(λ). *Prev. Med.* **106**, 233–5.
Wang, J. L., Chiou, J. M. & Müller, H. G. (2016). Functional data analysis. *Ann. Rev. Statist. Appl.* **3**, 257–95.
Yang, S. J., Shin, H., Lee, S. H., Lee, S. & Alzheimer's Disease Neuroimaging Initiative. (2020). Functional linear regression model with randomly censored data: predicting conversion time to Alzheimer's disease. *Comp. Statist. Data Anal.* **150**, 107009.
Yao, F. & Müller, H. G. (2010). Functional quadratic regression. *Biometrika* **97**, 49–64.
Yao, F., Müller, H. G. & Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873–903.
Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76–99.
Yuan, M. & Cai, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38**, 3412–44.
Zeng, D. & Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *J. Am. Statist. Assoc.* **102**, 1387–96.
Zhao, X., Wu, Y. & Yin, G. (2017). Sieve maximum likelihood estimation for a general class of accelerated hazards models with bundled parameters. *Bernoulli* **23**, 3385–411.

[*Received on* 21 *September* 2022. *Editorial decision on* 9 *January* 2025]