Statistica Sinica Preprint No: SS-2017-0357	
Title	Nonparametric Inference for Right Censored Data Using Smoothing Splines
Manuscript ID	SS-2017-0357
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0357
Complete List of Authors	Meiling Hao Yuanyuan Lin and Xingqiu Zhao
Corresponding Author	Yuanyuan Lin
E-mail	ylin@sta.cuhk.edu.hk
Notice: Accepted version subject to English editing.	

Statistica Sinica

Nonparametric Inference for Right Censored Data Using Smoothing Splines

Meiling Hao

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

Yuanyuan Lin

Department of Statistics, The Chinese University of Hong Kong, Hong Kong

Xingqiu Zhao

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

Abstract: This paper introduces a penalized nonparametric maximum likelihood estimation of the log-hazard function in analyzing right censored data. The smoothing splines are employed for a smooth estimation. Our main discovery is a functional Bahadur representation, which serves as a key tool for nonparametric inference of an unknown function. Asymptotic properties of the resulting smoothing spline estimator of the unknown log-hazard function are established under regularity conditions. Moreover, the local confidence interval of the unknown log-hazard function are provided, as well as the local and global likelihood ratio tests. We also discuss issues related to the asymptotic efficiency. The theoretical results are validated by extensive simulation studies, and an application is illustrated with a real data set.

Key words and phrases: Functional Bahadur representation; Likelihood ratio test;

Nonparametric inference; Penalized likelihood; Right censored data; Smoothing splines.

1. Introduction

In survival analysis, the outcome variable of interest is the time till the occurrence of an event, such as occurrence of a disease, death, divorce, etc. The time to event or survival time is usually measured in days, weeks or years, which is typically positive. Censored observations, of which the survival time is incomplete, are collected frequently in medical studies, reliability and many other fields related to survival analysis. The most common case is right censoring. To accommodate censoring, state-of-the-art statistical models and methodologies have been developed in past decades, including parametric, semiparametric and nonparametric methods.

Parametric approaches assume that the underlying distributions of the times to event are certain known probability distributions. For example, the exponential, lognormal and Weibull distributions are among those commonly used ones. Parametric methods are appealing to practitioners owing to their convenience and ease of interpretation (Johnson and Kotz, 1970; Mann *et al.*, 1974; Lawless, 1982; Kalbfleisch and Prentice, 2011). The most extensively used semiparametric model for the analysis of survival data is Cox's proportional hazards model, assuming that the hazard function of the

survival time is multiplicatively related to an unknown baseline function and the covariates; see Cox (1972, 1975), Cox and Oake (1984), Lin and Wei (1989), Lin and Ying (1994), Chen (2004) and Chen *et al.* (2010). In contrast to parametric models, Cox's model makes no assumption on the shape of the baseline hazard function, and provides easy-to-interpret information for the relationship of the hazard function of the survival time and the covariates. The parameter regarding the covariate effect in the Cox's model is usually estimated by maximizing the partial likelihood, and its large-sample properties are beautifully justified with the martingale theory; see Anderson and Gill (1982), Kosorok (2008), Fleming and Harrington (2011). In the analysis of survival data, important alternative semiparametric models to the Cox's proportional hazards model are the accelerated failure time model (AFT) and the transformation models, which assumes the logarithm of the survival time or an unknown but monotonic transformation of the survival time is linearly related to the covariates; see Kalbfleisch and Prentice (2011), Cox and Oakes (1984), Wei (1992), Chen, Jin and Ying (2002), Zeng and Lin (2007a, 2007b). Inference methods for the AFT model and transformation models have been studied thoroughly in the literature; see Buckley and James (1979), Prentice (1978), Ritov (1990), Tsiatis (1990), Wei *et al.* (1990), Lai and Ying (1991a, b), Ying (1993), Lin and Chen

(2013), Zeng *et al.*(2009). The additive hazards model has also been examined to be useful in modeling survival data; see Breslow and Day (1987), Lin and Ying (1994), Jiang and Zhou (2007).

Parametric and semiparametric methods rely very much on the distributional or model assumptions. However, the underlying distribution or model is often unknown, and the inference based on the parametric and semiparametric models may suffer from possible mis-specification. Without making assumption about the unknown distribution or an actual model form, nonparametric inference concerned about the hazard rate, survival function or density function have been proposed in the literature. And the hazard function is closely tied to survival function or density function through a direct relationship. The celebrated Kaplan-Meier estimator (Kaplan and Meier, 1958) is the nonparametric maximum likelihood estimator of the unknown survival function, which enjoys the self-consistency and asymptotic normality; see also Efron (1967), Breslow and Crowley (1967), Lo and Singh (1986), Chen and Lo (1997), among many others. It is worth mentioning that some Bahadur-type iid representation of the product limit estimator with right censored data can be found in Lo and Singh (1986). But the discontinuous nature of the Kaplan-Meier estimator makes the inference complicated. Herein smoothed estimators of the hazard or density

function are developed. With censored data, kernel smoothing and the nearest neighbor smoothing on the time axis are popular approaches to estimate the density function or the hazard function; see Beran (1981), Tanner and Wong (1983), Dabrowska (1987), Lo, Mack and Wang (1989), Gray (1992) and Müller and Wang (1994). Penalized likelihood methods based on smoothing splines are proposed in the literature; see Anderson and Senthilselvan (1980), O'Sullivan (1988) and Rosenberg (1995). It is known that kernel estimates reflect mostly the local structure of the data, while the estimator of the density function or the hazard function based on smoothing splines with a global smoothing parameter enjoy certain global properties (O'Sullivan *et al.* 1986).

However, to the best of our knowledge, except some consistency results of the smoothing splines hazard estimate were reported (Cox and O'Sullivan, 1990), there are limited discussions on the asymptotic properties of the smoothing spline estimator of the hazard function, and the existing asymptotic representations of the product limit estimator (Lo and Singh, 1986) or the kernel smoothing estimator of the hazard function (Tanner and Wong, 1986) are not directly applicable to the smoothing spline estimator. Moreover, the nonparametric inference for the hazard function is subject to a positive constraint, which makes the computation complicated.

Recently, Shang and Cheng (2013) introduced a unified asymptotic framework for the inference of smoothing spline estimation, which is indeed novel and has broad applications for statistical inference. In this paper, similar to Kooperberg *et al.* (1995), we target at the log-hazard rate in a nonparametric framework and provide a penalized likelihood estimate based on smoothing splines. Our major contribution is to establish the asymptotic properties of the proposed log-hazard estimator with right censored data.

The rest of the paper is organized as follows. Some background and preliminaries are given in Section 2. In Section 3, we establish a new functional Bahadur representation (FBR) in the Sobolev space and study the asymptotic properties of the resulting smoothing spline estimator of the log-hazard function. We discuss the hypothesis test in Section 4 and some simulation results are presented in Section 5. The proposed method is applied to a non-Hodgkin's lymphoma dataset in Section 6. All technical proofs are deferred to the supplementary.

2. Preliminaries

2.1 Notation and Methodology

We introduce some notation that will be used throughout this paper. Let T be the survival time, let C be the censoring time and let τ be the

end of the study. We define the observed time $Y = \min(T, C)$ and the censoring indicator $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Moreover, denote $\lambda(t)$ as the hazard rate function of the survival time and $g_0(t) = \log(\lambda(t))$. The hazard function $\lambda(t) : [0, \tau] \mapsto R$ is bounded away from 0 and infinity. Without loss of generality, we consider $\mathbb{I} \doteq [0, \tau] = [0, 1]$. Suppose that the observed data $(Y_i, \Delta_i), i = 1, \ldots, n$, are independent and identically distributed (i.i.d) copies of (Y, Δ). Then, the log-likelihood of g is

$$l_n(g) = -\int_{\mathbb{I}} \exp\{g(t)\} S_n(t)\, dt + \frac{1}{n} \sum_{i=1}^{n} \Delta_i g(Y_i),$$

where $S_n(\cdot) = n^{-1} \sum_{i=1}^{n} I(Y_i \geq t)$ is the empirical survival function of Y; see O'Sullivan (1988). Let $l(g) \equiv E\{l_n(g)\}$. A direct calculation yields that

$$l(g) = -\int_{\mathbb{I}} \exp\{g(t)\} S(t)\, dt + \int_{\mathbb{I}} \exp\{g_0(t)\} g(t) S(t)\, dt,$$

where $S(t) = Pr(Y \geq t)$. Throughout this paper, we suppose the true target function $g_0(t)$ belongs to the mth-order Sobolev space $\mathcal{H}^m(\mathbb{I})$ shorten as $\mathcal{H}^m$:

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

where the constant $m > 1/2$ and is assumed to be known, $g^{(j)}$ is the jth derivative of g and $L_2(\mathbb{I})$ is the L_2 space defined in $\mathbb{I}$. Define $J(g, \tilde{g}) =$

$\int_{\mathbb{I}} g^{(m)}(t)\tilde{g}^{(m)}(t)\,dt$. The penalized likelihood of $g(\cdot)$ is defined as:

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

where $J(g,g)$ is the roughness penalty and λ is the smoothing parameter, which converges to 0 as $n \to \infty$.

For the inference of $g_0(t)$, we propose to use B-splines to approximate g in $l_{n,\lambda}(g)$. For the finite closed interval $\mathbb{I}$, we define a partition of $\mathbb{I}$:

$$0 = t_1 = \ldots = t_m < t_{m+1} < \ldots < t_{m_n+m} < t_{m_n+m+1} = \ldots = t_{m_n+2m} = 1,$$

with which the interval $[0,1]$ is partitioned into $m_n + 1$ subintervals with knots set $\mathcal{I} \equiv \{t_i\}^{m_n+2m}$, and $m_n = o(n^v)$ for $0 < v < 1/2$. Let $\{B_{i,m}, 1 \leq i \leq q_n\}$ denote the B-spline basis functions with $q_n = m_n + m$. Let $\Psi_{m,\mathcal{I}}$ (with order m and knots $\mathcal{I}$) be the linear space spanned by the B-spline functions, that is

$$\Psi_{m,\mathcal{I}} = \left\{ \sum_{i=1}^{q_n} \theta_i B_{i,m} : \theta_i \in R, i = 1, \ldots, q_n \right\}.$$

It follows from Schumaker (1981) that there exists a smoothing spline $g_n(t) \in \Psi_{m,\mathcal{I}}$ such that $\|g_n(t) - g_0(t)\|_\infty = O(n^{-vm})$ and $\|g(t)\|_\infty \equiv \sup_{t\in\mathbb{I}} |g(t)|$. Hence, we define

$$\begin{aligned} \hat{g}_{n,\lambda} &\equiv \arg\max_{g\in\Psi_{m,\mathcal{I}}} l_{n,\lambda}(g) \\ &= \arg\max_{g\in\Psi_{m,\mathcal{I}}} \left\{ l_n(g) - \frac{\lambda}{2}J(g,g) \right\} \end{aligned}$$

as the estimator of $g_0(t)$. The numerical implementation of solving the above objective function is available in O'Sullivan (1988) with a fast computation algorithm. Moreover, a data-driven method based on AIC criterion was suggested to select the smoothing parameter λ.

2.2 Reproducing Kernel Hilbert Space

We now present some useful properties about the reproducing kernel Hilbert space (RKHS); see Shang and Cheng (2013). Under conditions (C1) and (C3) in the Appendix, $\mathcal{H}^m$ is a RKHS with the inner product

$$< g, \tilde{g} >_\lambda = \int_{\mathbb{I}} g(t)\tilde{g}(t) \exp\{g_0(t)\} S(t)\, dt + \lambda J(g, \tilde{g})$$

and the norm $\|g\|_\lambda^2 =< g, g >_\lambda$ when $m > 1/2$. Furthermore, there exists a positive definite self-adjoint operator $W_\lambda : \mathcal{H}^m \mapsto \mathcal{H}^m$, which satisfies: $< W_\lambda g, \tilde{g} >_\lambda = \lambda J(g, \tilde{g})$ for any $g, \tilde{g} \in \mathcal{H}^m$. Denote $V(g, \tilde{g}) = \int_{\mathbb{I}} g(t)\tilde{g}(t) \exp\{g_0(t)\} S(t)\, dt$. Then, it follows directly that

$$< g, \tilde{g} >_\lambda = V(g, \tilde{g}) + < W_\lambda g, \tilde{g} >_\lambda .$$

Let $K(\cdot, \cdot)$ be the reproducing kernel of $\mathcal{H}^m$ defined on $\mathbb{I} \times \mathbb{I}$, which is known to possess the following properties:

(P_1) $K_t(\cdot) = K(t, \cdot)$ and $< K_t, g >_\lambda = g(t)$ for any g in $\mathcal{H}^m$ and any t in $\mathbb{I}$.

(P_2) There exists a constant c_m depending only on m, such that $\|K_t\|_\lambda \leq c_m h^{-1/2}$ for any $t \in \mathbb{I}$ and $h = \lambda^{1/(2m)}$. Hence, we have $\|g(t)\|_\infty \leq$

$c_m h^{-1/2}\|g\|_\lambda$ for any $g \in \mathcal{H}^m$.

We denote two positive sequences a_n and b_n as $a_n \asymp b_n$ if $\lim_{n\to\infty}(a_n/b_n) = c > 0$. There exists a sequence of eigenfunctions $h_j \in \mathcal{H}^m$ and eigenvalues γ_j satisfying the following properties:

(P_3) $\sup_{j\in\mathcal{N}} \|h_j\|_\infty < \infty$, $\gamma_j \asymp j^{2m}$, where $\mathcal{N} = \{0, 1, \ldots\}$.

(P_4) $V(h_i, h_j) = \delta_{ij}$, $J(h_i, h_j) = r_j\delta_{ij}$, where δ_{ij} is a Kronecker delta, that is $\delta_{ij} = 1$ when $i = j$; $\delta_{ij} = 0$, otherwise.

(P_5) For any $g \in \mathcal{H}^m$, we have $g = \sum_{j=0}^\infty V(g, h_j)h_j$ with convergence in the $\|\cdot\|_\lambda$-norm.

(P_6) For any $g \in \mathcal{H}^m$ and $t \in \mathbb{I}$, we have $\|g\|_\lambda^2 = \sum_{j=0}^\infty V(g, h_j)^2(1 + \lambda\gamma_j)$, $K_t(\cdot) = \sum_{j=0}^\infty h_j(t)h_j(\cdot)/(1+\lambda\gamma_j)$ and $(W_\lambda h_j)(\cdot) = (\lambda\gamma_j)/(1+\lambda\gamma_j)h_j(\cdot)$.

Following the arguments in Shang and Cheng (2013, page 2613), the eigenvalues and eigenfunctions can be solved through the ordinary differential equations (ODE):

$$\begin{aligned} &(-1)^m h_j^{(2m)}(\cdot) = \gamma_j \exp(g_0(\cdot))S(\cdot)h_j(\cdot), \\ &\qquad h_j^{(k)}(0) = h_j^{(k)}(1) = 0, \qquad\qquad k = m, m+1, \cdots, 2m-1. \end{aligned} \tag{2.1}$$

For ease of presentation, we introduce more notations related to the Fréchet derivatives. Let $\mathcal{S}_n(g)$ and $\mathcal{S}_{n,\lambda}(g)$ be the Fréchet derivatives of

$l_n(g)$ and $l_{n,\lambda}(g)$, respectively. Similarly, let $\mathcal{S}(g)$ and $\mathcal{S}_\lambda(g)$ be the Fréchet derivatives of $l(g)$ and $l_\lambda(g)$, respectively. Let D be the Fréchet derivative operator and $g_1, g_2, g_3 \in \mathcal{H}^m$ be any direction. Then, we have

$$\begin{aligned} Dl_{n,\lambda}(g)g_1 &= -\int_{\mathbb{I}} \exp\{g(t)\} g_1(t) S_n(t)\, dt + \frac{1}{n}\sum_{i=1}^n \Delta_i g_1(Y_i) - < W_\lambda g, g_1 >_\lambda \\ &= < \mathcal{S}_n(g), g_1 >_\lambda - < W_\lambda g, g_1 >_\lambda \\ &= < \mathcal{S}_{n,\lambda}(g), g_1 >_\lambda, \end{aligned}$$

where $\mathcal{S}_n(g) = -\int_{\mathbb{I}} \exp\{g(t)\} K_t S_n(t)\, dt + n^{-1}\sum_{i=1}^n \Delta_i K_{Y_i}$ and $\mathcal{S}_{n,\lambda}(g) = \mathcal{S}_n(g) - W_\lambda g$. Moreover,

$$D^2 l_{n,\lambda}(g) g_1 g_2 = -\int_{\mathbb{I}} \exp\{g(t)\} g_1(t) g_2(t) S_n(t)\, dt - < W_\lambda g_1, g_2 >_\lambda,$$

$$D^3 l_{n,\lambda}(g) g_1 g_2 g_3 \;=\; -\int_{\mathbb{I}} \exp\{g(t)\} g_1(t) g_2(t) g_3(t) S_n(t)\, dt.$$

Further, by a direct calculation, we can express

$$\mathcal{S}(g) = Dl(g) = -\int_{\mathbb{I}} \exp\{g(t)\} K_t S(t)\, dt + \int_{\mathbb{I}} \exp\{g_0(t)\} K_t S(t)\, dt = E\{\mathcal{S}_n(g)\},$$

as well as $\mathcal{S}_\lambda(g) = \mathcal{S}(g) - W_\lambda g$. Besides,

$$D\{\mathcal{S}(g) g_1\} g_2 = D^2 l(g) g_1 g_2 = -\int_{\mathbb{I}} \exp\{g(t)\} g_1(t) g_2(t) S(t)\, dt.$$

Hence, we got the following result:

$$\begin{aligned}
< D\mathcal{S}_\lambda(g_0)f, g >_\lambda &= < D\{\mathcal{S}(g_0) - W_\lambda g_0\}f, g >_\lambda \\
&= < D\mathcal{S}(g_0)f, g >_\lambda - < W_\lambda f, g >_\lambda \\
&= < -\int_{\mathbb{I}} \exp\{g_0(t)\}f(t)K_t S(t)\, dt, g >_\lambda - < W_\lambda f, g >_\lambda \\
&= -\int_{\mathbb{I}} g(t)f(t)\exp\{g_0(t)\}S(t)\, dt - \lambda J(g, f) \\
&= - < f, g >_\lambda .
\end{aligned}$$

Proposition 1. *$D\mathcal{S}_\lambda(g_0) = -id$, where id is the identity operator in $\mathcal{H}^m$.*

This proposition plays an important role in deriving a functional Bahadur representation (FBR) of the proposed estimator.

3. Functional Bahadur Representation

In this section, we derive and present the key technical tool: functional Bahadur representation (FBR), which laid down a theoretical foundation for statistical inference procedures in later sections. With the help of the FBR, we establish the asymptotic normality of the proposed smoothing spline estimate. Likelihood ratio test procedure is also justified rigorously. To begin with, we present a lemma regarding the consistency of the proposed estimate for obtaining the FBR. All theoretical conditions and proofs are deferred to Appendix.

Lemma 1. *(Consistency). Suppose conditions (C1)-(C3) given in Appendix hold. If* $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ *as* $n \to \infty$ *for* $0 < v < 1/2$*, then for* n *large enough,*

$$\|\hat{g}_{n,\lambda} - g_0\|_\infty = o_p(1),$$

$$J(\hat{g}_{n,\lambda} - g_0, \hat{g}_{n,\lambda} - g_0) < \tilde{C},$$

where $\tilde{C}$ *is a constant larger than* 1.

In fact, the consistency of the estimator with the infinity norm can be derived along the lines of Cox and O'Sullivan (1990). But the second result in Lemma 1 is given by us.

To obtain the rate of convergence of the proposed estimator, we next drive a concentration inequality of certain empirical process. Define $\mathcal{G} = \{g \in \mathcal{H}^m : \|g\|_\infty \leq 1, J(g,g) \leq \tilde{C}\}$, with $\tilde{C}$ specified in Lemma 1. We next define

$$\mathcal{Z}_n(g) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\varphi_n(Y_i, g) - E\{\varphi_n(Y_i, g)\}],$$

where $\varphi_n(Y_i, g)$ is a real-valued function in $\mathcal{H}^m$.

Lemma 2. *Suppose that* $\varphi_n(Y, g)$ *satisfies the following condition:*

$$\|\varphi_n(Y, f) - \varphi_n(Y, g)\|_\lambda \leq \|f - g\|_\infty \qquad \textit{for any} \qquad f, g \in \mathcal{G}. \tag{3.1}$$

Then,

$$\lim_{n\to\infty} P\left[\sup_{g\in\mathcal{G}} \frac{\|\mathcal{Z}_n(g)\|_\lambda}{\|g\|_\infty^{1-1/(2m)} + n^{-1/2}} \le \{5\log\log(n)\}^{1/2}\right] = 1.$$

By Lemmas 1 and 2, we obtain the convergence rate of our estimate, which is presented in the following theorem:

Theorem 1. *(Convergence Rate). Assume that conditions (C1)-(C3) given in Appendix are satisfied. Then, when* $\log\{\log(n)\}/(nh^2) \to 0$ *and* $\lambda\{n^{(1-v)/2} + n^{vm}\} \to 0$ *as* $n \to \infty$*,*

$$\|\hat{g}_{n,\lambda} - g_0\|_\lambda = O_p\big((nh)^{-1/2} + h^m\big).$$

Remark 1. When $h \asymp n^{-1/(2m+1)}$, Theorem 1 suggests that $\hat{g}_{n,\lambda}$ achieves the optimal rate of convergence when we estimate $g_0 \in \mathcal{H}^m$, that is $O_p(n^{-\frac{m}{2m+1}})$. This result is in accordance to that in Gu (1991).

Using Theorem 1, we are ready to present the key technical tool of this paper, a new version of *functional Bahadur representation* compared with that of Shang and Cheng (2013). Define $M_i(t) \equiv N_i(t) - \int_0^t I(Y_i \ge s)\exp\{g_0(s)\}\,ds$, which is a martingale.

Theorem 2. *(Functional Bahadur Reprensentation). Assume that conditions (C1)-(C3) hold. Then, if* $\log\{\log(n)\}/(nh^2) \to 0$*,* $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ *as* $n \to \infty$*,*

$$\|\hat{g}_{n,\lambda} - g_0 - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda = O_p(\alpha_n),$$

where

$$\mathcal{S}_{n,\lambda}(g_0) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} K_t\, dM_i(t) - W_\lambda g_0$$

and

$$\alpha_n = n^{-1/2-vm} + n^{-vm}\{(nh)^{-1/2} + h^m\} + h^{-1/2}\{(nh)^{-1} + h^{2m}\}$$
$$+ h^{-(6m-1)/(4m)} n^{-1/2}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}.$$

In fact, Proposition 1 is crucial to derive the FBR in Theorem 2; see the Appendix for the proofs of Theorem 2. Moreover, Theorem 2 reveals that the "bias" of our estimate $\hat{g}_{n,\lambda}$ can be approximated by $\mathcal{S}_{n,\lambda}(g_0)$, a sum of martingale integral. Applying this result, we immediately obtain the asymptotic normality:

Theorem 3. *Assume conditions (C1)-(C3) hold. For $m > 3/4 + \sqrt{5}/4$ and $1/(4m) \le v \le 1/(2m)$, suppose $nh^{4m-1} \to 0$ and $nh^3 \to \infty$ as $n \to \infty$. Then, for any $t_0 \in \mathbb{I}$,*

$$\sqrt{nh}\{\hat{g}_{n,\lambda}(t_0) - g(t_0) + (W_\lambda g_0)(t_0)\} \xrightarrow{d} N(0, \sigma_{t_0}^2),$$

where $\sigma_{t_0}^2 \equiv \lim_{h\to 0} h \sum_{j=0}^{\infty} h_j^2(t_0)/(1+\lambda\gamma_j)^2$ and $\xrightarrow{d}$ means convergence in distribution.

Corollary 1. *Assume conditions (C1)-(C3) hold. For $m > 3/2$ and $1/(4m) \le v \le 1/(2m)$, suppose $nh^{2m} \to 0$ and $nh^3 \to \infty$ as $n \to \infty$. Then, for any*

$t_0 \in \mathbb{I}$,

$$\sqrt{nh}\{\hat{g}_{n,\lambda}(t_0) - g_0(t_0)\} \xrightarrow{d} N(0, \sigma_{t_0}^2),$$

where $\sigma_{t_0}^2$ *is defined the same as in Theorem 3.*

Remark 2. Corollary 1 implies that, under certain under-smoothing conditions, the asymptotic bias for the estimation of $g_0(t_0)$ vanishes. Hence, Corollary 1 together with the so-called Delta-method immediately gives the pointwise confidence interval (CI) for some real-valued smooth function of $g_0(t)$ at any fixed point $t_0 \in \mathbb{I}$, denoted by $\rho\big(g_0(t_0)\big)$. Let $\dot{\rho}(\cdot)$ be the first derivative of $\rho(\cdot)$. By Corollary 1, for any fixed point $t_0 \in \mathbb{I}$ and $\dot{\rho}\big(g_0(t_0)\big) \neq 0$, we have

$$P\left(\rho\big(g_0(t_0)\big) \in \left[\rho\big(\hat{g}_{n,\lambda}(t_0)\big) \pm \Phi_{\frac{\alpha}{2}} \frac{\dot{\rho}\big(g_0(t_0)\big)\sigma_{t_0}}{\sqrt{nh}}\right]\right) \to 1 - \alpha$$

as $n \to \infty$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and Φ_α is the lower α-th quantitle of $\Phi(\cdot)$, that is $\Phi(\Phi_\alpha) = \alpha$.

4. Likelihood Ratio Test

With the help of the FBR, we consider further inference of $g_0(\cdot)$ by testing local and global hypothesis. In this section, we focus on likelihood ratio tests for testing $g_0(\cdot)$.

4.1 Local Likelihood Ratio Test

We consider the following hypothesis for some pre-specified (t_0, ω_0):

$$H_0 : g(t_0) = \omega_0 \qquad \text{versus} \qquad H_1 : g(t_0) \neq \omega_0.$$

The penalized log-likelihood under H_0, or the "constrained" penalized log-likelihood by Shang and Cheng (2013), is defined as:

$$L_{n,\lambda}(g) = -\int_{\mathbb{I}} \exp\{g(t) + \omega_0\} S_n(t)\, dt + \frac{1}{n}\sum_{i=1}^{n} \Delta_i \{g(Y_i) + \omega_0\} - \frac{\lambda}{2} J(g, g),$$

where $g \in \mathcal{H}_0 = \{g \in \mathcal{H}^m : g(t_0) = 0\}$. We consider the following likelihood ratio test (LRT) statistic:

$$\text{LRT}_{n,\lambda} = L_{n,\lambda}(\omega_0 + \hat{g}^0_{n,\lambda}) - L_{n,\lambda}(\hat{g}_{n,\lambda}),$$

where $\hat{g}^0_{n,\lambda} \equiv \arg\max_{g \in \Psi^0_{m,\mathcal{I}}} L_{n,\lambda}(g)$ is the MLE of g in

$$\Psi^0_{m,\mathcal{I}} = \{\sum_{i=1}^{q_n} \theta_i B_{i,m}, \sum_{i=1}^{q_n} \theta_i B_{i,m}(t_0) = 0\}.$$

Clearly, $\mathcal{H}_0$ is a closed subset in $\mathcal{H}^m$, and hence it is a Hilbert space endowed with the norm $\|\cdot\|_\lambda$.

The following proposition states the reproducing kernel and penalty operator of $\mathcal{H}_0$ inherited from $\mathcal{H}^m$ without proofs.

Proposition 2. *The reproducing kernel and penalty operator of $\mathcal{H}_0$ inherited from $\mathcal{H}^m$ satisfy the following properties:*

(a) Recall that $K(t_1, t_2)$ is the reproducing kernel for $\mathcal{H}^m$ under $< \cdot, \cdot >_\lambda$.

Then, the bivariate function

$$K^*(t_1, t_2) = K(t_1, t_2) - K(t_0, t_1)K(t_0, t_2)/K(t_0, t_0)$$

is a reproducing kernel for $(\mathcal{H}_0, < \cdot, \cdot >_\lambda)$. *That is, for any* $t \in \mathbb{I}$ *and* $g \in \mathcal{H}_0$, *we have* $K_t^* \equiv K^*(t, \cdot) \in \mathcal{H}_0$ *and* $< K_t^*, g >_\lambda = g(t)$. *Moreover, we have* $\|K^*\|_\lambda \leq \sqrt{2} c_m h^{-1/2}$, *where* c_m *is the same as in* P_2.

(b) *The operator* W_λ^*, *defined by* $W_\lambda^* g \equiv W_\lambda g - W_\lambda g(t_0) K_{t_0}/K(t_0, t_0)$, *is bounded linear from* $\mathcal{H}_0$ *to* $\mathcal{H}_0$ *and satisfies* $< W_\lambda^* g, \tilde{g} >= \lambda J(g, \tilde{g})$, *for any* $g, \tilde{g} \in \mathcal{H}_0$.

Based on Proposition 2, we are in the position to derive the functional Bahadur representation (FBR) for $\hat{g}_{n,\lambda}^0$ under null hypothesis, or the so-called "restricted" FBR for $\hat{g}_{n,\lambda}^0$, which will be used to obtain the limiting distribution under the null. A direct calculation yields the Fréchet derivatives of $L_{n,\lambda}$ (along directions in $\mathcal{H}_0$). Consider $g_1, g_2, g_3 \in \mathcal{H}_0$. The first-order Fréchet derivative of $L_{n,\lambda}$, denoted by $\mathcal{S}_{n,\lambda}^0$, can be calculated as

follows:

$$
\begin{aligned}
&DL_{n,\lambda}(g)g_1 \\
&= -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)g_1(t)\,dt + \frac{1}{n}\sum_{i=1}^n \Delta_i g_1(Y_i) - < W_\lambda^* g, g_1 >_\lambda \\
&= -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t) < K_t^*, g_1 >_\lambda \, dt + \frac{1}{n}\sum_{i=1}^n \Delta_i < K_{Y_i}^*, g_1 >_\lambda \\
&\qquad\qquad - < W_\lambda^* g, g_1 >_\lambda \\
&=< -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)K_t^*\,dt, g_1 >_\lambda + \frac{1}{n}\sum_{i=1}^n \Delta_i < K_{Y_i}^*, g_1 >_\lambda \\
&\qquad\qquad - < W_\lambda^* g, g_1 >_\lambda \\
&=< \mathcal{S}_n^0(g), g_1 >_\lambda - < W_\lambda^* g, g_1 >_\lambda \\
&=< \mathcal{S}_{n,\lambda}^0(g), g_1 >_\lambda,
\end{aligned}
$$

where $\mathcal{S}_n^0(g) = -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)K_t^*\,dt + n^{-1}\sum_{i=1}^n \Delta_i K_{Y_i}^*$ and $\mathcal{S}_{n,\lambda}^0(g) = \mathcal{S}_n^0(g) - W_\lambda^* g$. Define $\mathcal{S}^0(g) \equiv E\{\mathcal{S}_n^0(g)\}$ and $\mathcal{S}_\lambda^0(g) \equiv \mathcal{S}^0(g) - W_\lambda^* g$. Next, we denote the second-order and the third-order Fréchet derivatives of $L_{n,\lambda}(g)$ as $D^2L_{n,\lambda}(g)g_1g_2$ and $D^3L_{n,\lambda}(g)g_1g_2g_3$ respectively. Further calculation yields that

$$
D^2L_{n,\lambda}(g)g_1g_2 = -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)g_1(t)g_2(t)\,dt - < W_\lambda^* g_2, g_1 >_\lambda,
$$

and

$$
D^3L_{n,\lambda}(g)g_1g_2g_3 = -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)g_1(t)g_2(t)g_3(t)\,dt.
$$

We consider the derivative of $\mathcal{S}_\lambda^0(g)$ and obtain

$$D\mathcal{S}_\lambda^0(g)g_1g_2 = -\int_0^1 \exp\{g(t)+\omega_0\}S(t)g_1(t)g_2(t)\,dt - < W_\lambda^* g_2, g_1 >_\lambda .$$

Then, by defining $g_0^0(t) = g_0(t) - \omega_0$, we get the following important equation:

$$\begin{aligned} < D\mathcal{S}_\lambda^0(g_0^0)f, g >_\lambda &=< D\{\mathcal{S}^0(g_0^0)\}f, g >_\lambda - < W_\lambda^* f, g > \\ &= -\int_0^1 \exp\{g_0^0(t)+\omega_0\}S(t)f(t)g(t)\,dt - < W_\lambda^* f, g >_\lambda \\ &= - < f, g > . \end{aligned}$$

We state this result as the next proposition.

Proposition 3. *$D\mathcal{S}_\lambda^0(g_0^0) = -id$, where id is the identity operator.*

Similar to Theorem 1 in Section 3, we need to prove the rate of convergence of the resulting estimator so as to obtain the FBR.

Proposition 4. *(Convergence Rate). Assume conditions (C1)-(C3) hold. Under H_0, if $\log\{\log(n)\}/(nh^2) \to 0$ and $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ as $n \to \infty$, we have*

$$\|\hat{g}_{n,\lambda}^0 - g_0^0\|_\lambda = O_p\big((nh)^{-1/2} + h^m\big).$$

The proof of Proposition 4 is similar to that of Theorem 1 and it is omitted. The next theorem follows directly from Propositions 2-4.

Theorem 4. *(Restricted Functional Bahadur Representation). Suppose that conditions (C1)-(C3) are satisfied. Also, we assume that under* H_0, $\log\{\log(n)\}/(nh^2) \to 0$ *and* $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ *as* $n \to \infty$. *Then,*

$$\|\hat{g}^0_{n,\lambda} - g^0_0 - \mathcal{S}^0_{n,\lambda}(g^0_0)\|_\lambda = O_p(\alpha_n),$$

where α_n *is defined in Theorem 2.*

Our main result on the local likelihood ratio test follows immediately from Theorem 4 and is presented below.

Theorem 5. *(Local Likelihood Ratio Test). Assume conditions (C1)-(C3) hold. For* $m > (5+\sqrt{21})/4$ *and* $1/(4m) \le v \le 1/(2m)$, *suppose that* $nh^{2m} \to 0$ *and* $nh^4 \to \infty$ *as* $n \to \infty$. *Furthermore, for any* $t_0 \in \mathbb{I}$, *if* $\sigma_{t_0} \neq 0$, *let* $c_{t_0} = \lim_{h\to 0} V(K_{t_0}, K_{t_0})/\|K_{t_0}\|^2_\lambda \in (0,1]$. *Then, under* H_0,

(i) $\|\hat{g}_{n,\lambda} - \hat{g}^0_{n,\lambda} - \omega_0\|_\lambda = O_p(n^{-1/2})$;

(ii) $-2nLRT_{n,\lambda} = n\|\hat{g}_{n,\lambda} - \hat{g}^0_{n,\lambda} - \omega_0\|^2_\lambda + o_p(1)$;

(iii) $-2nLRT_{n,\lambda} \xrightarrow{d} c_{t_0}\chi^2_1$.

Remark 3. The central Chi-square limiting distribution in part (iii) of the theorem is established under those under-smoothing assumptions in Theorem 5. One may also relax those conditions for h at the price of obtaining a noncentral Chi-square limiting distribution. We also note

that the convergence rate stated in theorem 5 is reasonable under local restriction.

4.2 Global Likelihood Ratio Test

It is of paramount importance to study the global behavior of a smooth function. In this section, we consider the following "global" hypothesis:

$$H_0^{global} : g = g_0 \qquad \text{versus} \qquad H_1 : g \neq g_0,$$

where $g_0 \in \mathcal{H}^m$ can be either known or unknown. The penalized likelihood ratio rest (PLRT) statistic is defined as

$$\text{PLRT}_{n,\lambda} \equiv l_{n,\lambda}(g_0) - l_{n,\lambda}(\hat{g}_{n,\lambda}).$$

We next derive the null limiting distribution of $\text{PLRT}_{n,\lambda}$.

Theorem 6. *Assume conditions (C1)-(C3) hold. For $m > (3+\sqrt{5})/4$ and $1/(4m) \leq v \leq 1/(2m)$, suppose that $nh^{2m+1} = O(1)$ and $nh^3 \to \infty$ as $n \to \infty$. Define $\sigma_\lambda^2 \equiv \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)$, $\rho_\lambda^2 \equiv \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)^2$, $\gamma_\lambda \equiv \sigma_\lambda^2/\rho_\lambda^2$, $\nu_\lambda \equiv h^{-1}\sigma_\lambda^4/\rho_\lambda^2$. Then, under H_0^{global},*

$$(2\nu_\lambda)^{-1/2}(-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda \|(W_\lambda g_0)(t)\|_\lambda^2 - \nu_\lambda) \xrightarrow{d} N(0,1).$$

It is worth noting that the null limiting distribution above remains unchanged even when g_0 in the null hypothesis is unknown. Moreover, it can be easily verified that $h \asymp n^{-d}$ with $1/(2m+1) \leq d < 1/3$ satisfies those

conditions in Theorem 6. We can also show that $n\|W_\lambda g_0\|^2 = o(h^{-1}) = o(\nu_\lambda)$. Thus, $-2n\gamma_\lambda \mathrm{PLRT}_{n,\lambda}$ is asymptotical $N(\nu_\lambda, 2\nu_\lambda)$, which approaches $\chi^2_{\nu_\lambda}$ as n goes to infinity. In other words, we have

$$-2n\gamma_\lambda \mathrm{PLRT}_{n,\lambda} \xrightarrow{d} \chi^2_{\nu_\lambda}$$

suggesting the Wilks phenomenon holds for the PLRT.

Lastly, to conclude this section, we are going to show that the PLRT achieves the optimal minimax rate of testing given by Ingster (1993) based on the uniform version of the FBR. To this end, we consider the alternative hypothesis $H_{1n} : g = g_{n_0}$, where $g_{n_0} = g_0 + g_n$, $g_0 \in \mathcal{H}^m$ and g_n belongs to the alternatives value set $\mathcal{A} = \{g \in \mathcal{H}^m, \exp\{g_n(t)\} \le \zeta, J(g,g) \le \zeta\}$ for some constant $\zeta > 0$.

Theorem 7. *Assume that conditions (C1)-(C3) are satisfied. For $m > (3+\sqrt{5})/4$ and $1/(4m) \le v \le 1/(2m)$, suppose that $h \asymp n^{-d}$ for $1/(2m+1) \le d < 1/3$ and uniformly over $g_n \in \mathcal{A}$, $\|\hat{g}_{n,\lambda} - g_{n_0}\|_\lambda = O_p\big((nh)^{-1/2} + h^m\big)$ holds under $H_{1n} : g = g_{n_0}$. Then, for any $\delta \in (0,1)$, there always exist positive constants b_0 and N such that*

$$\inf_{n \ge N} \inf_{g_n \in \mathcal{A}, \|g_n\|_\lambda \ge b_0 \eta_n} P(\textit{reject} \quad H_0^{global} | H_{1n} \textit{ is } \textit{ true}) \ge 1 - \delta,$$

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

Importantly, when $h = h^{**} = n^{-2/(4m+1)}$, Theorem 7 suggests that the PLRT can detect any local alternatives with separation rate no faster than $n^{-2m/(4m+1)}$, which is exactly the minimax rate of hypothesis testing in the sense of Ingster (1993).

5. Simulation studies

To evaluate the theoretical results, we present the simulation results in this section. In the simulation studies, we set $v = 1/5$ and the number of knots is $[3 \times n^{1/5}]$ for spline approximation, where $[x]$ is the integer part of x. Fore ease of presentation, more notation is needed. We define

$$H \equiv \int_0^1 \exp\{g(t)\} B(s) B(s)^\top S_n(s)\, ds,$$
$$\Omega_{lk} \equiv \int_0^1 \ddot{B}_{k,m}(s) \ddot{B}_{l,m}(s)\, ds, \qquad k, l = 1, 2, \ldots, q_n,$$

$\Omega \equiv (\Omega_{lk})$ is a matrix with the (l, k) element being Ω_{lk}, and $\ddot{B}_{l,m}(s)$ is the second derivative of $B_{l,m}(s)$. The following AIC criterion proposed by O'Sullivan (1988) is used to select the smoothing parameter λ:

$$AIC(\lambda) = -l_n(\hat{g}_{n,\lambda}) + \frac{trace[(\hat{H} + \lambda\Omega)^{-1}\hat{H}]}{n}.$$

In linear algebra, the trace of an n-by-n square matrix A is defined to be the sum of the elements on the main diagonal (the diagonal from the upper left to the lower right) of A.

To examine the performance of the pointwise confidence interval given

in Section 3, we compare our method with the average length of Bayesian confidence interval proposed by Wahba (1983), denoted by LBCI. And its corresponding coverage probability is denoted by BCP. We refer the average length of our proposed pointwise (local) confidence interval and its coverage probability as LLCI and LCP.

To generate data, we suppose the failure time follows the truncated Weibull distribution on $[1, \infty]$ with density function

$$f(t) \propto (t/\tau)^{k-1} \exp\{-(t/\tau)^k\}, \qquad t \in [1, \infty],$$

with $k = 1.5$ and $\tau = 1.2$. We generate the censoring time from the truncated Weibull distribution on $[1, 2]$ with $\tau = 3$ and k is chosen to yield 30% and 40% censoring rate. For the estimate of g_0, we compared our estimator and the kernel-smoothed Nelson estimator (Müller and Wang, 1994) denoted as Smoothed NE, which shows that the spline estimate is more stable than the kernel method, especially at the boundary region. Here, we set $m = 2$. Similar to Shang and Cheng (2013), we get the eigenvalues and eigenfunctions through the ODE function in (2.1) and then plug in the formula of the definition of σ_{t_0}. The simulation results are presented in Figures 1-2. We observe that the average length of our proposed local confidence interval (LLCI) is shorter than that of Wahba (1983). The LCP is close to 95% for $t \in [1.2, 1.7]$ and $t \in [1.2, 1.6]$ with censoring rate being 30% and

40%, respectively, while the BCP is almost 1 due to over-estimation in the variance.

To assess the performance of the global likelihood ratio test given in Section 4, we compute the sizes and the powers of the test based on simulated data for different situations. For this purpose, we consider the null hypothesis $H_0^{global} : g = g_0$ against $H_1 : g \neq g_0$, where $g_0(t) = \log(k) + (k-1)\log(t) - k\log(\tau)$ with $k = 1.5$ and $\tau = 1.2$. Take $g_1(t) = \log(k) + (k-1)\log(t) - k\log(\tau) + c\{\log(t) - \log(\tau)\} + \log(1 + c/k)$ with $c = 0, 0.5, 1, 1.5$. To perform the test, we generate the failure time from the truncated Weibull distribution on $[1, \infty]$ with log-hazard g_1 and the censoring time from Weibull distribution on $[1, 2]$ with $\lambda = 5$ and k chosen to yield 30% and 40% censoring rate, respectively. Again by solving the ODE in (2.1), we can get the eigenvalues of g_0 are $\gamma_j \approx (\alpha j)^{2m}$, with $\alpha = 1.8852$ or 1.8920 for the case of 30% censoring rate with $n = 250$ or 500, respectively. For the case of 40% censoring rate, we get $\alpha = 1.9944$ or 2.0126 with $n = 250$ or 500, respectively. Plugging in the value of α and by Lemma 6.1 in Shang and Cheng (2013), we can get $\gamma_\lambda = 1.333$, $h\nu_\lambda = 0.7856$ or 0.7828 for the case of 30% censoring rate with $n = 250$ or 500, respectively, and $h\nu_\lambda = 0.7426$ or 0.7359 for the case of 40% censoring rate with $n = 250$ or 500, respectively. The results of the global likelihood

ratio tests are reported in Table 1. The estimated size is around 5% for $c = 0$, while the estimated power approaches to 1 with the increasing of the sample size or c, which shows that the test has a good power.

6. Application

For illustration, we apply the proposed methods to analyze the study of non-Hodgkin's lymphoma (Dave *et al.*, 2004). The goal of the experiment is to detect the effect of follicular lymphoma on the patients' survival time. The data were obtained from seven institutions from 1974 to 2001. The samples are from 191 patients with untreated follicular lymphoma, who are diagnosed at the ages from 23 to 81 years (median 51). The follow-up times are ranging from 1.0 to 28.2 years (median 6.6). After removing 4 samples with missing censoring indicator and observation time, we have $n = 187$ samples and around 50% censoring rate. As suggested by Iglewicz and Hoaglin (1973), we also calculate an outlier statistic: $Z_i = 0.6745|Y_i - \text{median}(Y)|/\text{mad}(Y)$, where i refers to the ith subject, median(Y) and mad(Y) are the median and median absolute deviation of the 187 observation times, respectively. According to Iglewicz and Hoaglin (1973), an observation is an outlier if $Z > 3.5$. In this analysis, we observe that the 170th subject is the outlier. Then we clean it out and use the left samples to do the data analysis. We standardize the survival times to range

from 0 to 1. The results are summarized in Figure 3.

For comparison, we also compute the Kaplan-Meier estimate (KME), the smoothed Kaplan-Meier estimate (Smoothed KME) and our proposed estimate of the cumulative hazard function $\Lambda(t)$. The results are presented in the left panel of Figure 3. Our approach provides similar estimates to that of other methods. The right panel in Figure 3 presents estimation results of the log-hazard of our proposed method, the LCI as well as BCI. It can be seen that the pointwise interval of our proposal is shorter than that of Wahba (1983), which is accordance to the simulation results.

7. Conclusion

This paper focuses on the nonparametric inference of the log-hazard function for right censored data. The major advantage of doing so is that there is no constraint on the target function, and hence it simplifies the computation. It is well known that the penalized nonparametric maximum likelihood estimation is useful in balancing the smoothness and goodness-of-fit of the resulting estimator. We adopt the approach to estimate the log-hazard function in the presence of right censoring. On the other hand, the idea of smoothing B-spline can also be found in the literature for a smooth estimation, for example, Schumaker (1981). The main discovery of the article is a functional Bahadur representation established in the Sobolev space

$\mathcal{H}^m$ with a proper inner product, which serves as a key tool for nonparametric inference of the unknown parameter/function. Asymptotic properties of the resulting estimate of the unknown log-hazard function are justified rigorously. The local confidence interval of the unknown log-hazard function is provided, as well as the local and global likelihood ratio tests. We need to emphasize that, the penalized global likelihood ratio test is able to detect any local alternatives with minimax separation rate in the sense of Ingster (1993), which is closely related to the asymptotic efficiency. As suggested by one anonymous reviewer, we can extend our method to make inference of the survival function of the form $S_T(t) = \exp\{-\int_0^t \lambda(s)ds\}$. The inference procedures described in sections 3-4 can be modified accordingly.

We admit that the penalization on the coefficient of the B-spline function is not considered in the present paper, hence we cannot provide a constant estimate even when the true function takes constant value. The proposed inference approach can be also extended to handle other complicated data, for example, interval censored data. Although this extension seems to be conceptually straightforward, much more effort is needed to establish the theoretical properties of the estimators. Especially, it is a nontrivial task to develop an appropriate inner product for the Sobolev space. This problem is under investigation and is beyond the scope of the

current paper.

Supplementary Materials

Supplementary materials include all the technical proofs.

Acknowledgment

The authors are indebted to the Editor, the Associate Editor and two anonymous reviewers for their professional review and insightful comments that lead to substantial improvements in the paper. Yuanyuan Lin's research is supported in part by the Hong Kong Research Grants Council (Grants No. 509413 and 14311916) and Direct Grants for Research, The Chinese University of Hong Kong. Xingqiu Zhao's research is supported in part by the National Natural Science Foundation of China (No. 11771366) and the Hong Kong Polytechnic University.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Anderson, J. and Senthilselvan, A. (1980). Smooth estimates for the hazard function. *J. R. Statist. Soc. B* **42**, 322-327.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. *Technical*

Report. Univ. California, Berkeley.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2**, 437-453.

Breslow, N. E. and Day, N. E. (1987).*Statistical Models in Cancer Research 2: The Design and Analysis of Cohort Studies*. International Agency for Research on Cancer, Lyon.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Chen, K. (2004). Statistical estimation in the proportional hazards model with risk set sampling. *Ann. Statist.* **32**, 1513-1532.

Chen, K., Guo, S., Sun, L. and Wang, J. L. (2010). Global partial likelihood for nonparametric proportional hazards models. *J. Amer. Statist. Assoc.* **105**, 750-760.

Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659-668.

Chen, K. and Lo, S. H. (1997). On the rate of uniform convergence of the product-limit estimator: strong and weak laws. *Ann. Statist.* **25**, 1050-1087.

Cox, D. D. and OSullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Ann. Statist.* **18**, 1676-1695.

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

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

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* London: Chapman and Hall.

Dave, S. S., Wright, G., Tan, B., Rosenwald, A., Gascoyne, R. D., Chan, W. C., Fisher, R. I., Braziel, R. M., Rimsza, L. M., Grogan, T. M., Miller, T. P., Leblanc, M., Greiner, T. C., Weisenburger, D. D., Lynch, J. C., Vose, J., Armitage, J. O., Smeland, E. B., Kvaloy, S., Holte, H., Delabie, J., Connors, J. M., Lansdorp, P. M., Ouyang, Q., Lister, T. A., Davies, A. J., Norton, A. J., MÜller-Hermelink, H. K., Ott, G., Campo, E., Montserrat, E., Wilson, W. H., Jaffe, E. S., Simon, R., Yang, L., Powell, J., Zhao, H., Goldschmidt, N., Chiorazzi, M. and Staudt, L. M. (2004). Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N. Engl. J. Med.* **351**, 2159-2169.

Dabrowska, D. M. (1987). Non-parametric Regression with Censored Survival Time Data. *Scand. J. Statist.* **14**, 181-197.

Efron, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.* **4**, 831-853.

Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis.* John Wiley and Sons.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *J. Amer. Statist. Assoc.* **87**, 942-951.

Gu,C. (1991). Penalized Likelihood Hazard Estimation. *Technical Report.* Purdue Univeristy.

Hao, M., Song, X. and Sun, L. (2014). Reweighting estimators for the additive hazards model with missing covariates. *Can. J. Statist.* **42**(2), 285-307.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

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

Iglewicz, B. and Hoaglin, D. C. (1993). *How to Detect and Handle Outliers.* Asq Press.

Jiang, J. and Zhou, H. (2007). Additive hazard regression with auxiliary covariates. *Biometrika* **94** (2), 359-369.

Jong, D. P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75**(2), 261-277.

Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions.* Boston: Houghton Mifflin.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data.* John Wiley and Sons.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Kooperberg, C., Stone, C. J. and Truong, Y. K. (1995). Hazard regression. *J. Amer. Statist. Assoc.* **90**, 78-94.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer Science and Business Media.

Lai, T. L. and Ying, Z. (1991a). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370-1402.

Lai, T. L. and Ying, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531-556.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

Lin, D. Y. and Wei, L. J.(1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84** , 1074-1078.

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

Lin, Y. and Chen, K. (2013). Efficient estimation of the censored linear regression model. *Biometrika* **100**, 525-530.

Lo, S. H., Mack, Y. P., and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Prob. Theory and Related Fields* **30**(3), 461-473.

Lo, S. H. and Singh, K. (1986). The product-limit estimator and the bootstrap: some asymp-

totic representations. *Prob. Theory and Related Fields* **71**, 455-465.

Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974). *Methods for statistical Analysis of Reliability and Lifetime Data.* New York: Wiley.

Müller, H. G., and Wang, J. L.(1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61-76.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167-179.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303-328.

Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics* **51**, 874-887.

O'Sullivan, F. (1988). Fast computation of fully automated log-density and log- hazard estimators. *SIAM J. Sci. Statist. Comp.* **9**, 363- 379.

O'Sullivan, F., Yandell, B. S. and Raynor, W. J. (1986). Automatic Smoothing of Regression Functions in Generalized Linear Models. *J. Amer. Statist. Assoc.* **81**, 96-103.

Shang, Z. and Cheng, G. (2013). Local and global asymptotic inference in smoothing spline models. *Ann. Statist.* **41**(5), 2608-2638.

Tanner, M. A. and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* **11**, 989-993.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-372.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence*. Springer New York.

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline, *J. R. Statist. Soc. B* **45**,133-150.

Wang, H., Li, R. and Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94**, 553-568.

Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statist. Med.* **11**, 1871-1879.

Wei, L., Ying, Z. and Lin, D. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-851.

Zeng, D., Chen, Q. and Ibrahim, J. G. (2009). Gamma frailty transformation models for multivariate survival times. *Biometrika* **96**, 277-291.

Zeng, D. and Lin, D. Y. (2007a). Efficient estimation for the accelerated failure time model. *J. Am. Statist. Assoc.* **102**, 1387-96.

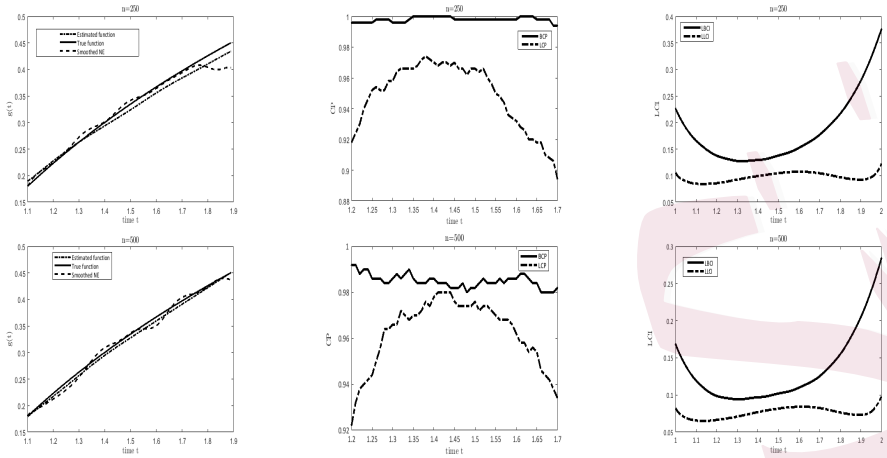
Zeng, D. and Lin, D. Y. (2007b). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Statist. Soc. B.* **69**, 507-564.

Meiling Hao, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

E-mail: meiling.hao@connect.polyu.hk

Yuanyuan Lin, Department of Statistics, The Chinese University of Hong Kong, Hong Kong

Figure 1. *Simulation results with around 30% censoring rate. CP: coverage probablity; LCI: average length of confidence interval.*



E-mail: ylin@sta.cuhk.edu.hk

Xingqiu Zhao, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

E-mail: xingqiu.zhao@polyu.edu.hk

Table 1. *The estimated size and power of the PLRT.*

Censoring rate	n	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$
30%	250	0.056	0.984	1.000	1.000
	500	0.048	1.000	1.000	1.000
40%	250	0.068	0.962	1.000	1.000
	500	0.052	0.998	1.000	1.000

Figure 2. *Simulation results with around 40% censoring rate. CP: coverage probablity; LCI: average length of confidence interval.*

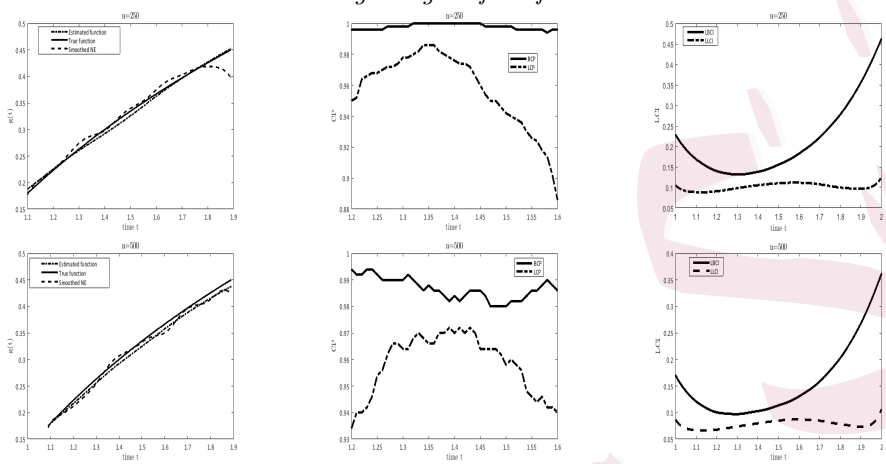


Figure 3. *Analysis results of the real data. The left panel displays the cumulative hazard estimation and the right panel presents the log-hazard estimation and its confidence interval using different methods.*