Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 928 – 948
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Estimation and confidence bands of a conditional survival function with censoring indicators missing at random

Qihua Wang[a, b, *], Junshan Shen[c]

[a] Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[b] Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100080, China
[c] School of Mathematical Sciences and LMAM, Peking University, Beijing 100871, China

Received 13 September 2006
Available online 12 June 2007

Abstract

The nonparametric estimator of the conditional survival function proposed by Beran is a useful tool to evaluate the effects of covariates in the presence of random right censoring. However, censoring indicators of right censored data may be missing for different reasons in many applications. We propose some estimators of the conditional cumulative hazard and survival functions which allow to handle this situation. We also construct the likelihood ratio confidence bands for them and obtain their asymptotic properties. Simulation studies are used to evaluate the performances of the estimators and their confidence bands.

AMS 1991 subject classification: primary 62N01; secondary 60G08

Keywords: Kernel regression; Confidence bands; Empirical likelihood; Missing at random; Random censorship

1. Introduction

In survival analysis and biomedical studies, the proportional hazards model is commonly used to examine the effects of covariates. The model assumes that the logarithm of the relative hazards is a linear function of covariates. However, the linearity and proportionality assumptions are often questionable in medical studies. The resulting inferences will yield biased estimators if the model is specified incorrectly. It is well known that the issue of model validity can be effectively

* Corresponding author. Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China. Fax: +852 28589041.
E-mail address: qhwang@amss.ac.cn (Q. Wang).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2007.06.002

addressed using nonparametric methods such as kernel smoothing techniques. Beran [2] proposed a kernel method to estimate the conditional cumulative hazard and survival functions. In the homogeneous case, the two kernel estimators reduce to the Aalen–Nelson and Kaplan–Meier estimates, respectively. Under the assumption that the lifetime random variable is independent of the random censoring variable given the covariables, the asymptotic properties of Beran's estimates can be found in Dabrowska [3,4], McKeague and Utikal [12], Li and Doss [9] and Li [8] among others.

Let T be a non-negative random variable and Z a covariate. Under random right censorship, rather than (T, Z), we only observe independent and identically distributed (i.i.d.) copies

$$(X_i, Z_i, \delta_i), \quad i = 1, \ldots, n$$

of the variables (X, Z, δ), where $X = \min(T, C)$, $\delta = 1_{(T \leqslant C)}$ is the censoring indicator function of $T \leqslant C$ and T and C are independent given Z.

Clearly, Beran's estimators require that the censoring indicator is always observed. In many applications, however, the censoring indicators may not be observed completely. For instance, in clinical trials, individuals may fail from one of two or multiple causes, one of which is of interest. The time to death from the cause of interest may be censored by a death from a different cause. However, cause of death may sometimes be unavailable; for example documenting whether or not death is attributable to the cause of interest may require information that is not collected to save expense or lost, or it may be difficult to determine the cause for some patients. In such cases, some censoring indicators are missing.

Let ξ_i be a missingness indicator which is 1 if δ_i is observed and is 0 otherwise. Therefore, we observe

$$O_i \equiv (X_i, Z_i, \delta_i, \xi_i = 1) \quad \text{or} \quad (X_i, Z_i, \xi_i = 0), \quad i = 1, \ldots, n. \tag{1.1}$$

In this paper, we develop approaches to estimate the conditional cumulative hazard and survival functions and construct their empirical likelihood (EL) ratio confidence bands with observations (1.1). We suppose that censoring indicators are missing at random (MAR), that is, ξ is conditionally independent of δ given X and Z:

$$P(\xi = 1|X, Z, \delta) = P(\xi = 1|X, Z).$$

In the absence of covariates, Dinse [5] obtained nonparametric maximum likelihood estimators of the survival function using the EM algorithm under the assumption that the censoring indicators are missing completely at random (MCAR), $P(\xi = 1|X, Z, \delta) = P(\xi = 1)$. However, Lo [11] proved that those estimators may be inconsistent, and he proposed two new estimates and proved their consistency. Under the MAR assumption, van der Laan and McKeague [17] proposed a sieved nonparametric maximum likelihood estimator and proved that it is asymptotically efficient.

Owen [13] introduced the EL method for construction of confidence regions. The method was studied by many authors since then. See Owen [14] for a comprehensive discussion. Because EL makes an automatic determination of the shape of confidence regions, and can incorporate side information through constraints, it is widely viewed as a desirable and natural approach to statistical inference in a variety of settings. The application of EL in survival analysis can be traced back to Thomas and Grunkemeier [16] who constructed confidence intervals for survival probabilities with censored data (see also Li [7]). EL-based confidence bands for individual survival functions have been derived by Hollander et al. [6]. Li and van Keilegom [10] obtained confidence bands for conditional survival function under random right censorship. In this paper, their method is generalized to deal with the right censored data with censoring indicators missing at random.

The paper is organized as follows. The estimators of the conditional survival function and the conditional cumulative hazard function are given in Section 2. Some asymptotic properties for the proposed estimators are given in Section 3. In Section 4, some asymptotic results of an empirical log-likelihood ratio are derived and confidence bands for the conditional survival function are obtained. A simulation study was conducted to evaluate the finite sample properties of the proposed estimators in Section 5. Proofs are postponed to the Appendices.

2. Estimation

Let $F(t|z) = P(T \leqslant t|Z = z)$ be the conditional distribution function of T given $Z = z$. Beran [2] proposed some estimators of the conditional survival function $S(t|z) = 1 - F(t|z)$ and the cumulative hazard function

$$\Lambda(t|z) = \int_0^t \frac{dF(s|z)}{S(s-|z)}.$$

To present these estimators, let $K(t)$ be a kernel function and $h_n = h_n(z)$ a smoothing bandwidth. Define Nadaraya–Watson weights

$$B_{ni}(z; h_n) = \frac{K_h(Z_i - z)}{\sum_{i=1}^n K_h(Z_i - z)},$$

where $K_h(t) = h_n^{-1} K(h_n^{-1} t)$. Let $\bar{F} = 1 - F$ for any distribution function F. Beran [2] defined the estimators of $\Lambda(t|z)$ and $S(t|z)$ by

$$\tilde{\Lambda}_n(t|z) = \sum_{i=1}^n \int_0^t \frac{\delta_i B_{ni}(z; h_n)\, dH_i(s)}{\bar{H}_n(s-|z)} \tag{2.1}$$

and

$$\tilde{S}_n(t|z) = \prod_{s \leqslant t} \{1 - \Delta\tilde{\Lambda}_n(s|z)\}, \tag{2.2}$$

respectively, where $\bar{H}_n(t-|z) = 1 - H_n(t|z)$. $H_i(t) = I(X_i \leqslant t)$ $(i = 1, \ldots, n)$,

$$H_n(t|z) = \sum_{i=1}^n I(X_i \leqslant t) B_{ni}(z; h_n)$$

is an estimator of $H(t|z) \equiv P(X \leqslant t|Z = z)$ and $\Delta\tilde{\Lambda}_n(s|z) = \tilde{\Lambda}_n(s|z) - \tilde{\Lambda}_n(s-|z)$.

Under the MAR assumption, we have

$$E\left(\tilde{\Lambda}_n(t|z)\Big|O_i,\ i = 1, \cdots n\right) = \sum_{i=1}^n \int_0^t \frac{\Big[\xi_i\delta_i + (1-\xi_i)E(\delta_i|X_i, Z_i)\Big] B_{ni}(z, h_n)\, dH_i(s)}{\bar{H}_n(s-|z)}.$$

This motivates us to define an estimator of $\Lambda(t|z)$ by replacing the censoring indicators δ_i in Beran estimator (2.1) with $E(\delta_i|X_i, Z_i)$ if $m(x, z) = E(\delta_i|X_i = x, Z_i = z)$ were a known function. In practice, however, $m(x, z)$ is unknown. We need to define an estimator of $m(x, z)$ and then replace it. Under the MAR condition, we have

$$m(x, z) = \frac{E(\xi_i \delta_i|X_i = x, Z_i = z)}{E(\xi_i|X_i = x, Z_i = z)}.$$

Therefore the kernel smoothing approach can be used to estimate $m(x, z)$ based on the observed data. Let $V(x)$ and $W(x)$ be two kernel functions. Let $a_n = a_n(z)$ and $b_n = b_n(z)$ be smoothing bandwidths. $m(x, z)$ can then be consistently estimated by the Nadaraya–Watson estimator

$$\hat{m}(x, z) = \frac{\sum_{i=1}^{n} \xi_i \delta_i V_a(X_i - x) W_b(Z_i - z)}{\sum_{i=1}^{n} \xi_i V_a(X_i - x) W_b(Z_i - z)},$$

where $V_a(\cdot) = a_n^{-1} V(a_n^{-1}(\cdot))$, $W_b(\cdot) = b_n^{-1} W(b_n^{-1}(\cdot))$. Define

$$\hat{\delta}_i = \xi_i \delta_i + (1 - \xi_i)\hat{m}(X_i, Z_i),$$

$\Lambda(t|z)$ and $S(t|z)$ can then be estimated by

$$\hat{\Lambda}_n(t|z) = \sum_{i=1}^{n} \int_0^t \frac{\hat{\delta}_i B_{ni}(z; h_n)\, dH_i(s)}{\bar{H}_n(s - |z)}$$

and

$$\hat{S}_n(t|z) = \prod_{s \leqslant t} \{1 - \Delta\hat{\Lambda}_n(s|z)\},$$

respectively. We present the uniformly strong consistency and weak convergence properties for $\hat{\Lambda}_n(t|z)$ and $\hat{S}_n(t|z)$, respectively, in the following section.

3. Asymptotic properties

We begin this section by giving some assumptions needed for the strong consistency. In what follows, for any cumulative distribution function F, let (a_F, b_F) be the range of F defined by

$$a_F = \inf\{x : F(x) > 0\} \quad \text{and} \quad b_F = \sup\{x : F(x) < 1\}.$$

Condition A.

(A1) Let $U(z)$ be a neighborhood of z. For $\tau_z < b_{H(\cdot|z)}$, $m(t, u)$ and $H(t|u)$ are uniformly continuous functions on $[0, \tau_z] \times U(z)$.

(A2) Let $\pi(x, z) = E(\xi|X = x, Z = z)$. $\inf_{x \leqslant \tau_z, u \in U(z)} \pi(x, u) > 0$.

(A3) $K(u)$, $W(u)$ and $V(u)$ are density functions with compact support, and symmetric around zero.

(A4) As $n \to \infty$, $h_n \to 0$, $a_n \to 0$, $b_n \to 0$, $nh_n \to \infty$, $na_n \to \infty$, $nb_n \to \infty$.

Theorem 3.1. *Suppose the condition A hold. For* $\tau_z < b_{H(\cdot|z)}$, *we have*

$$\sup_{0\leqslant t\leqslant \tau_z} \left|\hat{\Lambda}_n(t|z) - \Lambda(t|z)\right| \to 0 \quad a.s., \tag{3.1}$$

$$\sup_{0\leqslant t\leqslant \tau_z} \left|\hat{S}_n(t|z) - S(t|z)\right| \to 0 \quad a.s. \tag{3.2}$$

Condition B.

(B1) Let $f_Z(u)$ be the density function of Z. Functions $f_Z(u)$, $m(x,u)$, $\Lambda(x|u)$,$\pi(t,u)$ and $H(x|u)$ are continuously twice differentiable at u, where u belongs to a neighborhood $U(z)$ of z. Moreover, for $a_{F(\cdot|z)} \leqslant \tau_1 < \tau_2 < b_{H(\cdot|z)}$,

$$\sup_{u\in U(z),\tau_1\leqslant x\leqslant \tau_2} |H''_{uu}(x|u)| < \infty, \qquad \sup_{u\in U(z)} |f''_Z(u)| < \infty.$$

(B2) $\pi(x,u)$ satisfies $\inf_{u\in U(z),\tau_1\leqslant x\leqslant\tau_2} \pi(x,u) > 0$.

(B3) Kernel functions $K(u)$, $W(u)$ and $V(u)$ are bounded density functions with compact support, and are symmetric around zero, respectively.

(B4) As $n\to\infty, a_n\to 0, b_n\to 0, nh_n^5\to c, nh_na_n^4\to 0, nh_nb_n^4\to 0, nh_n\to\infty, na_n\to\infty, nb_n\to\infty$.

Let $\beta_K = \int x^2K(x)\,dx$, $\mu(K) = \int K^2(x)\,dx$, $H(t,z) = P(X\leqslant t|Z=z)f_Z(z)$ and

$$\mu(t|z) = \tfrac{1}{2}\beta_K \int_0^t \Lambda''_{zz}(ds|z) + \beta_K \int_0^t \frac{H'_z(s,z)\Lambda'_z(ds|z)}{H(s,z)}.$$

The following theorem states the weak convergence.

Theorem 3.2. *Suppose the condition B hold. Then for* $a_{F(\cdot|z)}\leqslant\tau_1 < \tau_2 < b_{H(\cdot|z)}$, *we have*

$$\sqrt{nh_n}\big(\hat{\Lambda}_n(t|z) - \Lambda(t|z) - \mu(t|z)h_n^2\big) \xrightarrow{\mathcal{W}} W(t|z) \quad in\ D[\tau_1,\tau_2], \tag{3.3}$$

$$\sqrt{nh_n}\big(\hat{S}_n(t|z) - S(t|z) + S(t|z)\mu(t|z)h_n^2\big) \xrightarrow{\mathcal{W}} S(t|z)W(t|z) \quad in\ D[\tau_1,\tau_2], \tag{3.4}$$

where $W(t|z)$ *is a zero mean Gaussian process with covariance function*

$$Cov\big(W(t_1|z), W(t_2|z)\big) = f_z^{-1}(z)\mu(K)\int_0^{t_1\wedge t_2}\left[\frac{d\Lambda(s|z)}{\bar{H}(s|z)} + \frac{m(s,z)(1-m(s,z))(1-\pi(s,z))\,dH(s|z)}{\pi(s,z)\bar{H}^2(s|z)}\right]. \tag{3.5}$$

Remark. Bandwidth conditions $nh_na_n^4\to 0$ and $nh_nb_n^4\to 0$ imply that $m(x,z)$ is oversmoothed and the bias of $\hat{m}(x,z)$ is asymptotically vanishing. However, under condition $nh_n^5\to c$, the bias from estimating the survival function itself still exists.

Theorem 3.2 implies that both $\hat{\Lambda}_n(t|z)$ and $\hat{S}_n(t|z)$ are asymptotically normal with asymptotic variances $\sigma^2(t|z)$ and $S^2(t|z)\sigma^2(t|z)$, respectively, where $\sigma^2(t|z) = Cov\big(W(t|z), W(t|z)\big)$. A direct method to estimate the asymptotic variance is to use the "plug in" technique by replacing $H(s|z)$, $\Lambda(t|z)$, $\pi(s,z)$, $m(s,z)$, $f_Z(z)$ and $S(t|z)$ in the asymptotic variances with their estimators. Another alternative is to use the jackknife method to estimate the asymptotic variance.

Next, we make some discussions on bandwidth selection. Suppose we define $\hat{\Lambda}_n(t|z)$ on interval $[\tau_1, \tau_2]$, where $a_{F(\cdot|z)} \leqslant \tau_1 < \tau_2 < b_{H(\cdot|z)}$. From Theorem 3.2, the asymptotic mean integrated squared error (AMISE) of $\hat{\Lambda}_n(t|z)$ on interval $[\tau_1, \tau_2]$ is

$$\mathrm{AMISE}\big(\hat{\Lambda}_n(t|z)\big) = \int_{\tau_1}^{\tau_2} \left\{\mu^2(t|z)h_n^4 + \sigma^2(t|z)(nh_n)^{-1}\right\} dt.$$

Minimizing AMISE($\hat{\Lambda}_n(t|z)$) with respect to h_n, we obtain the optimal bandwidth

$$h_{\Lambda,\mathrm{opt}} = \left[\frac{\int_{\tau_1}^{\tau_2} \sigma^2(t|z)\,dt}{4\int_{\tau_1}^{\tau_2} \mu^2(t|z)\,dt}\right]^{1/5} n^{-1/5}.$$

Similarly, the optimal bandwidth minimizing the AMISE of $\hat{S}_n(t|z)$ on interval $[\tau_1, \tau_2]$ is

$$h_{S,\mathrm{opt}} = \left[\frac{\int_{\tau_1}^{\tau_2} S^2(t|z)\sigma^2(t|z)\,dt}{4\int_{\tau_1}^{\tau_2} S^2(t|z)\mu^2(t|z)\,dt}\right]^{1/5} n^{-1/5}.$$

Since $\sigma^2(t|z)$ and $\mu^2(t|z)$ are unknown, we need to use their estimates to substitute them and obtain estimators of $h_{\Lambda,\mathrm{opt}}$ and $h_{S,\mathrm{opt}}$ in practice. However, it may need another bandwidth selection.

Another way to select a bandwidth is to use the bootstrap method. We generate repeatedly B bootstrap samples $\{O_i^*, i = 1, \ldots, n\}$ from the observed data $O_i,\ i = 1, 2, \ldots, n$ and get B bootstrapped survival functions $\hat{S}_{n,1}^*(t|z), \ldots, \hat{S}_{n,B}^*(t|z)$ for a pilot bandwidth h_n, then we choose h_n by minimizing the bootstrapped MISE

$$\frac{1}{B}\sum_{i=1}^{B}\int_{\tau_1}^{\tau_2} \big(\hat{S}_{n,i}^*(t|z) - \hat{S}_n(t|z)\big)^2\,dt. \tag{3.6}$$

Also, the AMISE given above indicates that a proper choice of a_n and b_n specified in condition (B.4) does not affect the first-order term of the mean integrated square error, though it might affect higher order terms. This also shows that the selection of a_n and b_n might not be so critical to $\hat{S}_n(t|z)$, a result which is also verified in our simulation studies.

4. Confidence bands

In this section, we construct confidence bands for $S(t|z)$ by using different methods. We use an undersmoothing bandwidth h_n which satisfies $nh_n^5 \to 0$ so that the asymptotic bias of $\hat{S}(t|z)$ is zero. Assume τ_1, τ_2 are two numbers such that $a_{F(\cdot|z)} \leqslant \tau_1 < \tau_2 < b_{H(\cdot|z)}$. From Theorem 3.2 we get

$$\frac{\sqrt{nh_n}(\hat{S}_n(t|z) - S(t|z))}{S(t|z)} \xrightarrow{\mathcal{W}} W(t|z) \quad \text{in } D[\tau_1, \tau_2].$$

Let $B(t)$ be a Brownian bridge on [0, 1] and $u = \sigma^2(t|z)/(1 + \sigma^2(t|z))$. Since $W(t)/\sigma(t|z)$ and $B(u)/\sqrt{u(1-u)}$ have the same distribution, we have

$$\frac{\sqrt{nh_n}(\hat{S}_n(t|z) - S(t|z))}{S(t|z)\sigma(t|z)} \xrightarrow{\mathcal{W}} \frac{B(u)}{\sqrt{u(1-u)}} \quad \text{in } D[\tau_1, \tau_2].$$

Let $e_\alpha(u_1, u_2)(u_1 \leqslant u_2)$ be the α quantile of the distribution

$$\sup_{u_1 \leqslant u \leqslant u_2} \left| \frac{B(u)}{\sqrt{u(1-u)}} \right|.$$

Let

$$\hat{\pi}(t,z) = \frac{\sum_{i=1}^{n} \xi_i V_a(X_i - t) W_b(Z_i - z)}{\sum_{i=1}^{n} V_a(X_i - t) W_b(Z_i - z)},$$

$$\hat{f}_Z(z) = \frac{1}{n} \sum_{i=1}^{n} K_h(Z_i - z),$$

$$\hat{\sigma}^2(t|z) = \mu(K) \hat{f}_Z^{-1}(z) \frac{1}{n} \sum_{i=1}^{n} \left[\int_0^t \frac{\hat{\delta}_i B_{ni}(z; h_n)\, dH_i(s)}{\bar{H}_n^2(s-|z)} + \int_0^t \frac{\hat{m}(s,z)(1-\hat{m}(s,z))(1-\hat{\pi}(s,z)) B_{ni}(z;h_n)\, dH_i(s)}{\hat{\pi}(s,z)\bar{H}_n^2(s-|z)} \right]$$

and $\hat{u}_j = \hat{\sigma}^2(\tau_j|z)/(1 + \hat{\sigma}^2(\tau_j|z))$ for $j = 1, 2$. Based on the asymptotic properties of $\ln(-\ln \hat{S}_n(t|z))$, we define the following asymptotic $100(1-\alpha)\%$ transformed bands of $S(t|z)$ [1] by

$$\hat{S}_n(t|z)^{\exp\{\pm(nh_n)^{-1/2}\hat{\sigma}(t|z)e_{1-\alpha}(\hat{u}_1,\hat{u}_2)/\ln \hat{S}_n(t|z)\}}, \tag{4.1}$$

and for a fixed t, asymptotic $100(1-\alpha)\%$ transformed confidence interval of $S(t|z)$ is defined by

$$\hat{S}_n(t|z)^{\exp\{\pm(nh_n)^{-1/2}\hat{\sigma}(t|z)z_{1-\alpha/2}/\ln \hat{S}_n(t|z)\}}, \tag{4.2}$$

where z_α is the α quantile of the standard normal distribution. By using transformations, confidence interval/band (4.1) and (4.2) may avoid upper or lower limits falling outside the [0, 1] interval, and may improve the performance for small sample size.

Next we use EL method to obtain confidence interval/band of $S(t|z)$. In the absence of covariate, Lo [11] proposed the likelihood function based on data (1.1) and showed that the maximum likelihood estimators are not unique and some of them are inconsistent. Instead using the likelihood function based on data $\{O_i, i = 1, \ldots, n\}$, we define a likelihood function based on synthetic data $\{(X_i, Z_i, \hat{\delta}_i),\ i = 1, \ldots, n\}$. Without loss of generality, we suppose that $X_1 \leqslant X_2 \leqslant \cdots \leqslant X_n$ are the order statistics and $(Z_i, \hat{\delta}_i)$ are the concomitant of X_i for $i = 1, \ldots, n$. Let Γ be the space of all survival functions defined on $[0, \infty)$. For $S(t|z) \in \Gamma$, we have the following local likelihood function

$$L(S|z) = \prod_{i=1}^{n} \left\{ (S(X_i - |z) - S(X_i|z)^{\hat{\delta}_i} S(X_i|z)^{1-\hat{\delta}_i} \right\}^{B_{ni}(z;h_n)}.$$

For a fixed t, an EL ratio can then be defined by

$$\mathcal{R}(p, t|z) = \frac{\sup\{L(S|z) : S(t|z) = p, S(t|z) \in \Gamma\}}{\sup\{L(S|z) : S(t|z) \in \Gamma\}}.$$

Let $R_i = \bar{H}_n(X_i - |z)$, $B_i = \hat{\delta}_i B_{ni}(z; h_n)$. Similar to Li and van Keilegom [10], we get

$$\ln \mathcal{R}(p, t|z) = \sum_{i:X_i \leqslant t} \left\{ (R_i - B_i) \ln \left(1 + \frac{\lambda_n(t|z)}{R_i - B_i} \right) - R_i \ln \left(1 + \frac{\lambda_n(t|z)}{R_i} \right) \right\}, \tag{4.3}$$

where the lagrange multiplier $\lambda_n(t|z)$ satisfies

$$\sum_{i:X_i \leqslant t} \ln \left(1 - \frac{B_i}{R_i + \lambda_n(t|z)} \right) - \ln p = 0. \tag{4.4}$$

Define

$$\hat{\sigma}_0^2(t|z) = \mu(K) \hat{f}_Z^{-1}(z) \sum_{i:X_i \leqslant t} \frac{B_i}{R_i(R_i - B_i)}.$$

Then we have the following theorem.

Theorem 4.1. *Suppose the condition B hold and τ_1, τ_2 satisfy $a_{F(\cdot|z)} \leqslant \tau_1 < \tau_2 < b_{H(\cdot|z)}$, then*

$$-2nh_n \frac{\hat{\sigma}_0^2(t|z) \hat{f}_Z(z)}{\hat{\sigma}^2(t|z) \mu(K)} \ln \mathcal{R}(S(t|z), t|z) \xrightarrow{\mathcal{W}} \frac{B^2(u)}{u(1-u)} \quad \text{in } D[\tau_1, \tau_2],$$

where $B(t)$ is a Brownian bridge on $[0, 1]$ and $u = \sigma^2(t|z)/(1 + \sigma^2(t|z))$. Especially, for each t such that $a_{F(\cdot|z)} < t < b_{H(\cdot|z)}$, we have

$$-2nh_n \frac{\hat{\sigma}_0^2(t|z) \hat{f}_Z(z)}{\hat{\sigma}^2(t|z) \mu(K)} \ln \mathcal{R}(S(t|z), t|z) \xrightarrow{\mathcal{W}} \chi_1^2.$$

Theorem 4.1 can be used to define confidence bands for $S(t|z)$ over interval $[\tau_1, \tau_2]$ by

$$I_{n,\alpha}(\tau_1, \tau_2) = \left\{ (p, t) : -2nh_n \frac{\hat{\sigma}_0^2(t|z) \hat{f}_Z(z)}{\hat{\sigma}^2(t|z) \mu(K)} \ln \mathcal{R}(p, t, z) \leqslant e_{1-\alpha}^2(\hat{u}_1, \hat{u}_2), t \in [\tau_1, \tau_2] \right\} \tag{4.5}$$

and for a fixed t, an asymptotic $100(1-\alpha)\%$ confidence interval of $S(t|z)$ is then defined by

$$I_{n,\alpha}(t) = \left\{ p : -2nh_n \frac{\hat{\sigma}_0^2(t|z) \hat{f}_Z(z)}{\hat{\sigma}^2(t|z) \mu(K)} \ln \mathcal{R}(p, t, z) \leqslant \chi_1^2(1-\alpha) \right\}, \tag{4.6}$$

where $\chi_1^2(\alpha)$ is the α quantile of the χ_1^2 distribution.

Theoretically, the optimal bandwidth should be chosen as the value at which the coverage error of the interval/band attains the minimum. Since the coverage error cannot be observed, Li and van Keilegom [10] suggested using the bootstrap method to select the bandwidth. As an example, let us consider the bandwidth selection for constructing confidence interval (4.6). For a pilot bandwidth h_n, we generate repeatedly B bootstrap samples $\{O_i^*, i = 1, \ldots, n\}$ from the observed data O_i, $i = 1, 2, \ldots, n$ and construct B bootstrapped confidence band $I_{n,\alpha,1}^*, \ldots, I_{n,\alpha,B}^*$ using the EL method aforementioned. Then we choose the bandwidth $h = h_{\text{opt}}$ such that the bootstrapped coverage error

$$err^*(h) = \left| \frac{^{\#}\{S_n(t) \in I_{n,\alpha,i}^*,\ i = 1, \ldots, n\}}{B} - (1-\alpha) \right|$$

attains the minimum, where $^{\#}A$ denotes the number of elements of set A. Similar method can be used to construct confidence interval/band (4.1), (4.2) and (4.5).

5. Simulation results

We conducted Monte Carlo simulations to evaluate the finite sample performances of the estimator $\hat{S}_n(t|z)$. In our simulation, the conditional distribution of T given $Z = z$ is exponential with mean $a_0+a_1z+a_2z^2$, the censoring time C has an exponential distribution with mean function $b_0 + b_1z + b_2z^2$ and the covariate Z is uniform distribution on the interval [0, 1]. The parameters were adjusted to produce censoring rates of 33% ($a_0 = 0, a_1 = a_2 = 1, b_0 = 0, b_1 = b_2 = 2$) and 50% ($a_0 = b_0 = 0, a_1 = a_2 = b_1 = b_2 = 1$) in different simulations. The selected probability function was taken to follow the logistic model

$$P(\xi_i = 1|X_i, Z_i) = \frac{1}{1 + \exp(c_1X_i + c_2Z_i)}.$$

The parameters c_1 and c_2 were adjusted to produce different missing rates with $c_1 = c_2$. Kernel functions K, V and W were selected as the Epanechnikov kernel and the bandwidths were taken to be $a_n = an^{-1/4}$, $b_n = bn^{-1/4}$ and $h_n = hn^{-1/5}$. At first we consider the sensitivities of those bandwidths. Let $z = 0.5, a_0 = 0, a_1 = 1, a_2 = 1, b_0 = 0, b_1 = 1, b_2 = 1, c_1 = c_2 = 1.25$ and the sample size $n = 80$. For 1000 duplications, we draw the average MISE of $\hat{S}_n(t|z)$ ($t \in [0, 3]$) as function of h, a and b over interval [0.3, 3], respectively. The results are summarized in Fig. 1.

From Fig. 1 we see that MISE of $\hat{S}_n(t|z)$ are sensitive to bandwidth h_n, but not sensitive to bandwidths a_n and b_n. In what follows, we let $a_n = n^{-1/4}$, $b_n = n^{-1/4}$ and $h_n = hn^{-1/5}$. The bandwidth parameter h was selected at interval [0.3, 3] such that the bootstrapped MISE (3.6) attains the minimum. For $z = 0.5$, Table 1 presents the simulation results for MISE of $\hat{S}_n(t|z)$ over the interval [0, 3] with different missing rates, including that of Beran estimator $\tilde{S}_n(t|z)$ without

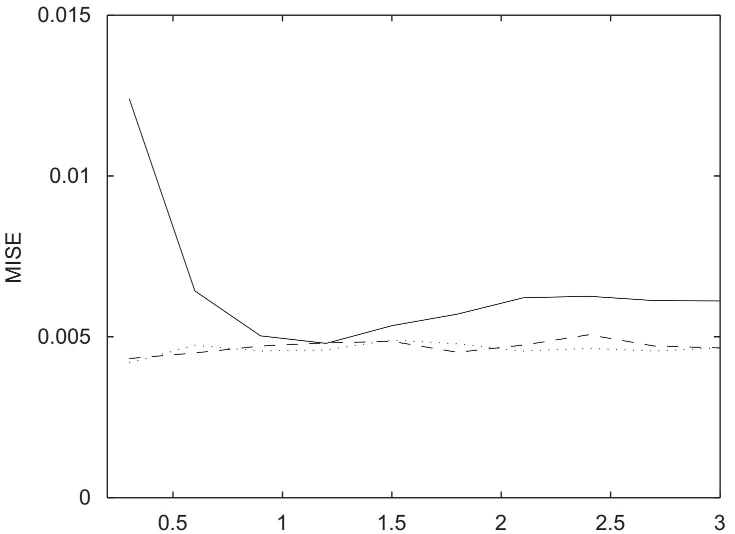

Fig. 1. Bandwidths sensitivities. The smooth solid line is MISE of $\hat{S}_n(t|z)$ as a function of h ($a = 1, b = 1$), while the dotted line is MISE of $\hat{S}_n(t|z)$ as a function of a ($h = 1.2, b = 1$), the dashed line is MISE of $\hat{S}_n(t|z)$ as a function of b ($h = 1.2, a = 1$).

Table 1
MISE for $\hat{S}_n(t|z)$ at $z = 0.5$

Censoring rate	n	No missing	$E\xi = 0.75$	$E\xi = 0.45$	$E\xi = 0.35$
0.33	40	0.0089	0.0092	0.0099	0.0110
	80	0.0049	0.0051	0.0053	0.0060
	120	0.0034	0.0034	0.0040	0.0043
0.50	40	0.0111	0.0117	0.0121	0.0131
	80	0.0063	0.0066	0.0071	0.0083
	120	0.0045	0.0046	0.0054	0.0059

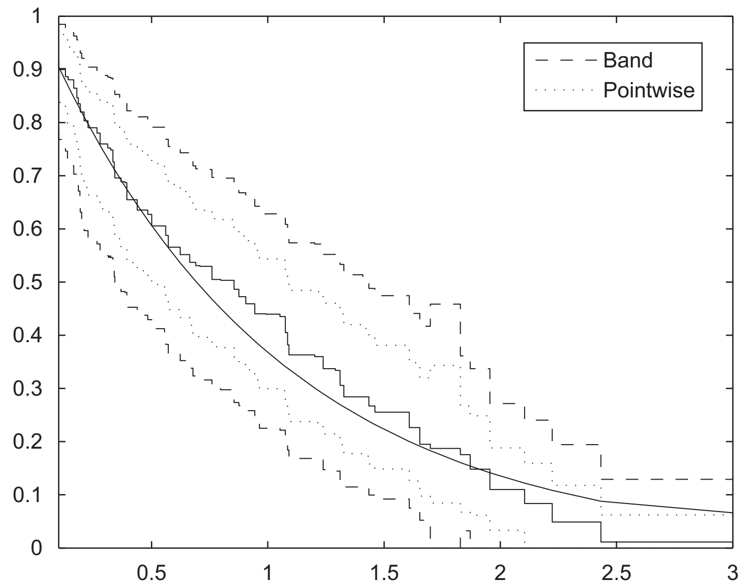

Fig. 2. Estimated conditional survival curve and confidence bands based on empirical likelihood. The smooth solid line is $S(t|z)$, while the step solid line is $\hat{S}_n(t|z)$, the dashed lines are confidence bands (4.5) and the dotted lines are pointwise confidence bands (4.6). The parameters were selected as $z = 0.5, a_0 = 0.5, a_1 = 1, a_2 = 0, b_0 = 2, b_1 = 1, b_2 = 0, c_1 = c_2 = 1.25$ and the sample size $n = 80$.

missing data. The Beran estimator can serve as a gold standard, even though it is practically unachievable because of the missingness of censoring indicators. For each sample size n (n=40, 80, 120), 1000 duplications were calculated in simulations.

From Table 1 we see that the MISE of $\hat{S}_n(t|z)$ is very small and is close to that of the Beran estimator, the gold standard, especially when the missing rate is small. The MISE of $\hat{S}_n(t|z)$ decreases as sample size increases and increases as missing rate or censoring rate increases. As an example, we plotted the curves of $S(t|z)$ and $\hat{S}_n(t|z)$ for $z = 0.5$ in Fig. 1 (also in Fig. 2) for $n = 80$. It can be seen that the curve of the estimator is close to the true conditional survival function. All in all, our estimator $\hat{S}_n(t|z)$ performs well in terms of the MISE.

Next, we used Monte Carlo simulation to compare the EL and the normal type confidence intervals of $S(t|z)$ in terms of coverage accuracy. The bandwidths $a_n = n^{-1/3}$, $b_n = n^{-1/3}$

Table 2
Confidence intervals of $S(0.5|z = 0.5)$

$E\xi$	n	EL		NA	
		Accuracy	Length	Accuracy	Length
$\alpha = 0.90$					
0.75	40	0.887	0.396	0.920	0.435
	80	0.892	0.320	0.915	0.340
	120	0.897	0.295	0.902	0.310
0.45	40	0.885	0.402	0.886	0.451
	80	0.891	0.346	0.894	0.371
	120	0.903	0.307	0.905	0.324
0.30	40	0.882	0.413	0.884	0.462
	80	0.894	0.360	0.892	0.378
	120	0.895	0.311	0.895	0.327
$\alpha = 0.95$					
0.75	40	0.943	0.427	0.963	0.521
	80	0.949	0.393	0.956	0.415
	120	0.949	0.346	0.954	0.359
0.45	40	0.943	0.444	0.945	0.545
	80	0.944	0.410	0.947	0.442
	120	0.952	0.361	0.953	0.386
0.30	40	0.925	0.465	0.927	0.560
	80	0.945	0.407	0.941	0.432
	120	0.947	0.371	0.953	0.396

and $h_n = an^{-1/3}$ and we selected a at interval $[0.5, 2]$ such that the bootstrapped coverage error attains the minimum. The censoring rate is 50% ($a_0 = a_1 = a_2 = b_0 = b_1 = b_2 = 1$) and $E\xi$ were selected as 0.75, 0.45 and 0.30. For $t = 0.5$ and $z = 0.5$, simulation results for two type confidence intervals of $S(t|z)$ with asymptotic coverage probability α are summarized in Table 2, where "EL" denotes confidence interval (4.6) and "NA" denotes confidence interval (4.2).

We also compare the behaviors of the likelihood ratio confidence band (4.5) and the normal approximation based band (4.1) over the interval $[u_1, u_2] = [0.1, 0.9]$. We let $a_0 = 0.5, b_0 = 2, a_1 = b_1 = 1$ and $a_2 = b_2 = 0$. Censoring rate was taken to be 72%. The missing indicators ξ follow the same logistic model as before, and the missing rates $E\xi$ were selected as 0.5 and 0.8. For each sample size n ($n = 40$, 80, 120), 1000 duplications were calculated in simulations. For $z = 0.5$, Table 3 gives the simulation results, where "EL" denotes confidence band (4.5) and "NA" denotes confidence band (4.1).

Comparing the coverage accuracy of the EL method and normal approximation method from Tables 2 and 3, we can see both methods perform quite well in terms of coverage accuracy, but confidence interval/band produced by EL method has shorter length/smaller width.

In Figs. 2 and 3, we plotted the pointwise confidence bands and uniform confidence bands over interval $[0.1, 3]$ with coverage level $\alpha = 0.90$ based on EL and normal approximation, respectively. From those figures we see both methods produce bands falling inside the [0, 1] interval automatically.

Table 3
Confidence bands of $S(t|z = 0.5)$ over $u \in [0.1, 0.9]$

$E\xi$	n	EL		NA	
		Accuracy	Width	Accuracy	Width
$\alpha = 0.90$					
0.8	40	0.906	0.420	0.912	0.433
	80	0.908	0.329	0.913	0.337
	120	0.897	0.294	0.906	0.302
0.5	40	0.898	0.435	0.912	0.447
	80	0.904	0.354	0.909	0.360
	120	0.894	0.305	0.907	0.312
$\alpha = 0.95$					
0.8	40	0.952	0.457	0.964	0.471
	80	0.952	0.362	0.955	0.371
	120	0.951	0.320	0.947	0.329
0.5	40	0.951	0.479	0.961	0.497
	80	0.946	0.375	0.955	0.388
	120	0.949	0.337	0.951	0.346

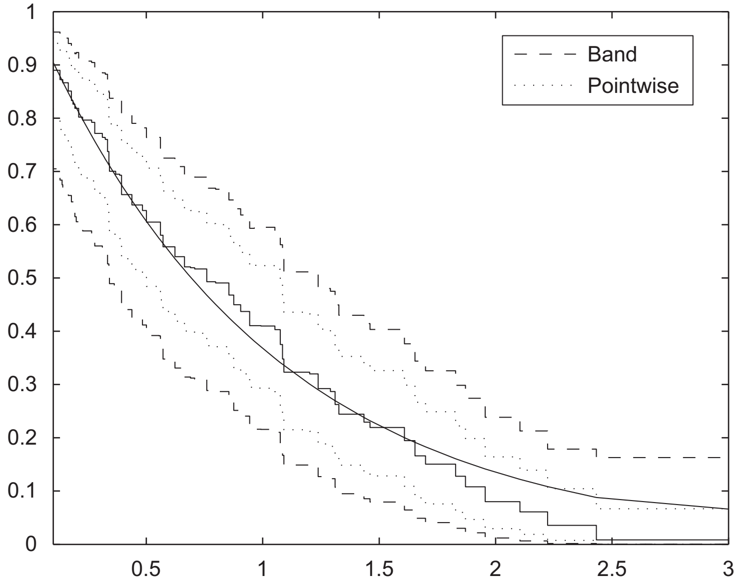

Fig. 3. Estimated conditional survival curve and confidence bands based on normal approximation. The smooth solid line is $S(t|z)$, while the step solid line is $\hat{S}_n(t|z)$, the dashed lines are confidence bands (4.1) and the dotted lines are pointwise confidence bands (4.2).The parameters were selected as $z = 0.5, a_0 = 0.5, a_1 = 1, a_2 = 0, b_0 = 2, b_1 = 1, b_2 = 0, c_1 = c_2 = 1.25$ and the sample size $n = 80$.

Acknowledgments

We are grateful to the referees and the associate editor for their helpful comments. The research was supported by the National Science Fund for Distinguished Young Scholars, the National

Natural Science Foundation of China (Key grant: 10231030; General grant: 10671198) and a grant from the research Grants Council of the Hong Kong, China (HKU 7050/06P).

Appendix A

Proof of Theorem 3.1. Since

$$\hat{\Lambda}_n(t|z) - \Lambda(t|z) = \big(\hat{\Lambda}_n(t|z) - \Lambda(t|z)\big) + J_{n1}(t) + J_{n2}(t), \tag{A.1}$$

where

$$J_{n1}(t) \equiv \sum_{i=1}^{n} \int_0^t \frac{(1-\xi_i)\big(m(s, Z_i) - \delta_i\big) B_{ni}(z; h_n)\, dH_i(s)}{\bar{H}_n(s-|z)} - \Lambda(t|z),$$

$$J_{n2}(t) \equiv \sum_{i=1}^{n} \int_0^t \frac{(1-\xi_i)(\hat{m}(s, Z_i) - m(s, Z_i)) B_{ni}(z; h_n)\, dH_i(s)}{\bar{H}_n(s-|z)}.$$

Since $H(\tau_z|z) > 0$, we have $\sup_{0 \leqslant t \leqslant \tau_z} \left|\bar{H}_n^{-1}(t|z) - \bar{H}^{-1}(t|z)\right| \overset{a.s.}{\to} 0$. This together with the fact $n^{-1}\sum_{i=1}^n K_h(Z_i - z) \overset{a.s.}{\to} f_Z(z)$ proves

$$J_{n1}(t) = \frac{1}{n f_Z(z)} \sum_{i=1}^{n} \frac{(1-\xi_i)\big(m(X_i, Z_i) - \delta_i\big) K_h(Z_i - z)}{\bar{H}(X_i|z)} I(X_i \leqslant t) + o(1) \quad a.s.$$

uniformly for $t \in [0, \tau_z]$. Under MAR assumption, since $E|(1-\xi_i)\bar{H}^{-1}(X_i|z)\big(m(X_i, Z_i) - \delta_i\big) K_h(Z_i - z) I(X_i \leqslant t)| < \infty$ and

$$E\, \frac{(1-\xi_i)\big(m(X_i, Z_i) - \delta_i\big) K_h(Z_i - z)}{\bar{H}(X_i|z)} I(X_i \leqslant t) = 0,$$

by the strong law of large number, we get

$$J_{n1}(t) \to 0 \quad a.s. \tag{A.2}$$

On the other hand, by $\left|\hat{m}(t, z) - m(t, z)\right| \to 0\ a.s.$ and $\bar{H}_n(t-|z) \to \bar{H}(t|z)\ a.s.$, we get

$$\left|J_{n2}(t)\right| \leqslant |\hat{m}(t, z) - m(t, z)| \bar{H}_n^{-1}(t|z) \to 0 \quad a.s. \tag{A.3}$$

By (A.1)–(A.3) and the strong consistency of $\hat{\Lambda}_n(t|z)$ due to Dabrowska [3], it follows that $\hat{\Lambda}_n(t|z)$ is pointwise consistent. Since $\hat{\Lambda}_n(t|z)$ and $\Lambda(t|z)$ are bounded monotone functions on interval $[0, \tau_z]$ and $\Lambda(t|z)$ is continuous, (3.1) is then proved.

Since

$$\max_{1 \leqslant i \leqslant n} \left|B_{ni}(z; h_n)\right| = \max_{1 \leqslant i \leqslant n} \left|\frac{n^{-1} K_h(Z_i - z)}{\hat{f}_Z(z)}\right| = O((nh)^{-1}) \quad a.s.,$$

by inequality $|\ln(1-x)+x| \leqslant x^2$, we get

$$\sup_{0\leqslant t\leqslant \tau_z} \left|\ln \hat{S}_n(t|z) + \hat{\Lambda}_n(t|z)\right| \leqslant \sup_{0\leqslant t\leqslant \tau_z} \sum_{i=1}^{n} \int_0^t \frac{\hat{\delta}_i^{\,2} B_{ni}^2(z;h_n)\, dH_i(s)}{\bar{H}_n^2(s-|z)}$$
$$\leqslant \max_{1\leqslant i\leqslant n} \left|B_{ni}(z;h_n)\right| \bar{H}_n^{-2}(\tau_z - |z) = O((nh)^{-1}) \quad a.s.$$

It follows by Taylor expansion that

$$\begin{aligned}\hat{S}_n(t|z) - S(t|z) &= \Big\{ \exp\big(\ln \hat{S}_n(t|z) + \hat{\Lambda}_n(t|z)\big) \exp\big(-\hat{\Lambda}_n(t|z) + \Lambda(t|z)\big) - 1 \Big\} S(t|z) \\ &= \Big\{ \exp\big(-\hat{\Lambda}_n(t|z) + \Lambda(t|z)\big) - 1 \Big\} S(t|z) + o(1) \\ &= \big(-\hat{\Lambda}_n(t|z) + \Lambda(t|z)\big) S(t|z) + o(1) \quad a.s. \end{aligned} \tag{A.4}$$

uniformly for $t \in [0, \tau_z]$. We get (3.2) by (3.1) and (A.4). □

Appendix B

The weak convergence results need the following asymptotic representation of $\hat{\Lambda}_n(t|z)$ and $\hat{S}_n(t|z)$.

Lemma B.1. *Suppose condition B hold, then for $\tau_z < b_{H(\cdot|z)}$, we have*

$$\hat{\Lambda}_n(t|z) - \Lambda(t|z) - \mu(t|z)h_n^2 = \frac{1}{n}\sum_{i=1}^{n} IC_{ni}(t|z) + r_n(t|z), \tag{B.1}$$

$$\hat{S}_n(t|z) - S(t|z) + S(t|z)\mu(t|z)h_n^2 = \frac{1}{n}\sum_{i=1}^{n} S(t|z) IC_{ni}(t|z) + r_n'(t|z), \tag{B.2}$$

where

$$IC_{ni}(t|z) \equiv \frac{K_h(Z_i - z)}{\hat{f}_Z(z)} \int_0^t \frac{dM_i(s)}{\bar{H}_n(s-|z)} + \frac{K_h(Z_i - z)}{f_Z(z)} \frac{(\xi_i - \pi(X_i, Z_i))\big(\delta_i - m(X_i, Z_i)\big) I(X_i \leqslant t)}{\pi(X_i, Z_i)\bar{H}(X_i|z)}$$

with

$$M_i(t) \equiv \delta_i H_i(t) - \int_0^t I(X_i \geqslant s)\, d\Lambda(s|Z_i)$$

and

$$\sup_{t\in[0,\tau_z]} |r_n(t|z)| = o_p((nh_n)^{-1/2}), \qquad \sup_{t\in[0,\tau_z]} |r_n'(t|z)| = o_p((nh_n)^{-1/2}).$$

Proof. At first we represent $\hat{\Lambda}_n(t|z) - \Lambda(t|z)$ as

$$\hat{\Lambda}_n(t|z) - \Lambda(t|z) = I_{n1}(t) + I_{n2}(t) + I_{n3}(t), \tag{B.3}$$

where

$$I_{n1}(t) \equiv \tilde{\Lambda}_n(t|z) - \Lambda(t|z),$$

$$I_{n2}(t) \equiv \sum_{i=1}^{n} \int_0^t \frac{(1-\xi_i)(m(s, Z_i) - \delta_i) B_{ni}(z; h_n)\, dH_i(s)}{\bar{H}_n(s-|z)},$$

$$I_{n3}(t) \equiv \sum_{i=1}^{n} \int_0^t \frac{(1-\xi_i)(\hat{m}(s, Z_i) - m(s, Z_i)) B_{ni}(z; h_n)\, dH_i(s)}{\bar{H}_n(s-|z)}.$$

Since

$$I_{n1}(t) - \mu(t|z)h_n^2 = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{K_h(Z_i - z)\, dM_i(s)}{\hat{f}_Z(z)\bar{H}_n(s-|z)} + \int_0^t \frac{dU_n(s|z)}{\hat{f}_Z(z)\bar{H}_n(s-|z)} \tag{B.4}$$

where

$$M_i(t) = \delta_i H_i(t) - \int_0^t I(X_i \geqslant s)\, d\Lambda(s|Z_i)$$

is a local martingale with respect to the filtration generated by

$$\left\{ I(X_i \leqslant s, \delta_i = 1), I(X_i \leqslant s, \delta_i = 0), Z_i, i = 1, \ldots, n, s \leqslant t \right\}$$

and

$$U_n(t|z) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t K_h(Z_i - z) I(X_i \geqslant s)\, d[\Lambda(s|Z_i) - \Lambda(s|z)] \\ - \int_0^t \hat{f}_Z(z)\bar{H}_n(s-|z)\mu(ds|z)h_n^2.$$

Since $\hat{f}_Z(z)\bar{H}_n(t-|z) \to^p f_Z(z)\bar{H}(t|z)$ uniformly in $[0, \tau_z]$, Lemma 2.2 of Dabrowska's [4] shows $(nh_n)^{1/2}U_n(t|z) \to^p 0$ uniformly in $t \in [0, \tau_z]$. Following Dabrowska's [4] proof of Theorem 2.1, we have

$$\sup_{t\in[0,\tau_z]} \left| \int_0^t \frac{dU_n(s|z)}{\hat{f}_Z(z)\bar{H}_n(s|z)} \right| = o_p((nh_n)^{-1/2}). \tag{B.5}$$

Combining (B.4) and (B.5), we get

$$I_{n1}(t) - \mu(t|z)h_n^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(Z_i - z)}{\hat{f}_Z(z)} \int_0^t \frac{dM_i(s)}{\bar{H}_n(s-|z)} + o_p((nh_n)^{-1/2}) \tag{B.6}$$

uniformly for $t \in [0, \tau_z]$. Since

$$\sup_{t\in[0,\tau_z]} \left(H_n(t-|z)\hat{f}_Z(z) - H(t|z) f_Z(z) \right)^2 = o_p((nh_n)^{-1/2}),$$

uniformly for $t \in [0, \tau_z]$, the second term in (B.3) can be represented as

$$\begin{aligned}
I_{n2}(t) &= \sum_{i=1}^{n} \int_0^t \frac{(1-\xi_i)(m(s,Z_i)-\delta_i)B_{ni}(z;h_n)dH_i(s)}{\bar{H}_n(s-|z)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_0^t \frac{(1-\xi_i)(m(s,Z_i)-\delta_i)K_h(Z_i-z)dH_i(s)}{\bar{H}(s-|z)f_Z(z)} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \int_0^t \frac{(1-\xi_i)\Big(m(s,Z_i)-\delta_i\Big)\Big(H_n(s-|z)\hat{f}_Z(z)-H(s|z)f_Z(z)\Big)K_h(Z_i-z)\,dH_i(s)}{\bar{H}^2(s|z)f_Z^2(z)} \\
&\quad + o_p((nh_n)^{-1/2}) \\
&\equiv I_{n21}(t) + I_{n22}(t) + o_p((nh_n)^{-1/2}).
\end{aligned}$$

Notice

$$\begin{aligned}
I_{n22}(t) = \frac{1}{n(n-1)} \sum_{i \neq j} \Bigg\{ &\frac{(1-\xi_i)\big(m(X_i,Z_i)-\delta_i\big)K_h(Z_i-z)}{\bar{H}^2(X_i|z)f_Z^2(z)} \\
&\times \big(I(X_j \leqslant X_i)K_h(Z_j-z) - H(X_i|z)f_Z(z)\big)\Bigg\} I(X_i \leqslant t) + o_p((nh_n)^{-1/2}),
\end{aligned}$$

by condition (B4) and properties of U-statistics processes [15, Theorem 1.2], uniformly for $t \in [0, \tau_z]$,

$$\begin{aligned}
I_{n22}(t) &= \frac{1}{n}\sum_{i=1}^{n} \Bigg\{ \frac{(1-\xi_i)\big(m(X_i,Z_i)-\delta_i\big)K_h(Z_i-z)}{\bar{H}^2(X_i|z)f_Z^2(z)} \\
&\qquad \times \Big(E\big(I(X \leqslant X_i)K_h(Z-z)|X_i\big) - H(X_i|z)f_Z(z)\Big)\Bigg\} I(X_i \leqslant t) + o_p((nh_n)^{-1/2}) \\
&= o_p((nh_n)^{-1/2}). \qquad \text{(B.7)}
\end{aligned}$$

Thus

$$I_{n2}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{(1-\xi_i)(m(X_i,Z_i)-\delta_i)K_h(Z_i-z)I(X_i \leqslant t)}{\bar{H}(X_i|z)f_Z(z)} + o_p((nh_n)^{-1/2}) \qquad \text{(B.8)}$$

uniformly for $t \in [0, \tau_z]$. Define

$$\hat{f}_1(t,z) \equiv \frac{1}{n}\sum_{i=1}^{n} \xi_i\delta_i V_a(X_i-t)W_b(Z_i-z),$$

$$\hat{f}_2(t,z) \equiv \frac{1}{n}\sum_{i=1}^{n} \xi_i V_a(X_i-t)W_b(Z_i-z),$$

then $\hat{m}(t,z) = \hat{f}_1(t,z)/\hat{f}_2(t,z)$. Let $f(t,z)$ be the density function of $P(X \leqslant t, Z \leqslant z)$ and define $f_1(t,z) \equiv \pi(t,z)m(t,z)f(t,z)$, $f_2(t,z) \equiv \pi(t,z)f(t,z)$. By MAR assumption and conditions B, we get

$$\sup_{t\in[0,\tau_z]} |\hat{f}_1(t,z) - f_1(t,z)| = O_p((nh_n)^{-1/2}),$$

$$\sup_{t\in[0,\tau_z]} |\hat{f}_2(t,z) - f_2(t,z)| = O_p((nh_n)^{-1/2}).$$

So uniformly for $t \in [0, \tau_z]$, the function $I_{n3}(t)$ can be decomposed as

$$I_{n3}(t) = I_{n31}(t) - I_{n32}(t) + o_p((nh_n)^{-1/2}), \tag{B.9}$$

where

$$I_{n31}(t) \equiv \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{(1-\xi_i)\big(\hat{f}_1(s,Z_i) - \hat{f}_2(s,Z_i)m(s,Z_i)\big)K_h(Z_i - z)\,dH_i(s)}{f_2(s,Z_i)\bar{H}(s|z)f_Z(z)}$$

and

$$I_{n32}(t) \equiv \frac{1}{n}\sum_{i=1}^{n}\int_0^t \Bigg[\frac{(1-\xi_i)\big(\hat{f}_1(s,Z_i) - \hat{f}_2(s,Z_i)m(s,Z_i)\big)K_h(Z_i - z)}{f_2^2(s,Z_i)\bar{H}(s|z)f_Z(z)} \times\big(\hat{f}_2(s,Z_i) - f_2(s,Z_i)\big)\Bigg]\,dH_i(s).$$

Let $\tilde{f}_1(t,z) = E\hat{f}_1(t,z)$ and $\tilde{f}_2(t,z) = E\hat{f}_2(t,z)$. Since

$$\begin{aligned}I_{n31}(t) &= \frac{1}{n(n-1)}\sum_{i\neq j} \\ &\quad\times\left[\frac{\xi_j(1-\xi_i)(\delta_j - m(X_i,Z_i))V_a(X_j - X_i)W_b(Z_j - Z_i)K_h(Z_i - z)}{f_2(X_i,Z_i)\bar{H}(X_i|z)f_Z(z)}\right] \\ &\quad\times I(X_i \leqslant t) + o_p((nh_n)^{-1/2}),\end{aligned}$$

uniformly for $t \in [0, \tau_z]$, we have

$$\begin{aligned}I_{n31}(t) &= \frac{1}{n}\sum_{k=1}^{n}\int K_h(u-z)\,dF_Z(u) \\ &\quad\times\int_0^t \frac{\xi_k\big(1-\pi(s,u)\big)\big(\delta_k - m(s,u)\big)V_a(X_k - s)W_b(Z_k - u)\,dH(s|u)}{f_2(s,u)\bar{H}(s|z)f_Z(z)} \\ &\quad+\frac{1}{n}\sum_{k=1}^{n}\int_0^t \frac{(1-\xi_k)\big(\tilde{f}_1(s,Z_k) - \tilde{f}_2(s,Z_k)m(s,Z_k)\big)K_h(Z_k - z)\,dH_k(s)}{f_2(X_k,Z_k)\bar{H}(X_k|z)f_Z(z)} \\ &\quad+o_p((nh_n)^{-1/2}).\end{aligned} \tag{B.10}$$

uniformly for $t \in [0, \tau_z]$, by properties of U-statistics processes [15, Theorem 1.2]. Taylor expansion of $m(x, z)$, $\bar{H}(x|z)$, $f_Z(x)$ shows

$$\frac{1}{n}\sum_{k=1}^{n}\int K_h(u-z)\,dF_Z(u) \times \int_0^t \frac{\xi_k\big(1-\pi(s,u)\big)\big(\delta_k - m(s,u)\big)V_a(X_k - s)W_b(Z_k-u)\,dH(s|u)}{f_2(s,u)\bar{H}(s|z)f_Z(z)}$$
$$= \frac{1}{n}\sum_{k=1}^{n}\frac{\xi_k\big(1-\pi(X_k,Z_k)\big)\big(\delta_k - m(X_k,Z_k)\big)K_h(Z_k-z)f(X_k,Z_k)I(X_k \leqslant t)}{f_2(X_k,Z_k)\bar{H}(s|z)f_Z(z)} + o_p((nh_n)^{-1/2}). \tag{B.11}$$

By condition (B4), we get

$$\sup_{t\in[0,\tau_z]} |\tilde{f}_1(t,z) - f_2(t,z)| = o_p((nh_n)^{-1/2}),$$
$$\sup_{t\in[0,\tau_z]} |\tilde{f}_2(t,z) - f_2(t,z)| = o_p((nh_n)^{-1/2}).$$

Thus

$$\frac{1}{n}\sum_{k=1}^{n}\int_0^t \frac{(1-\xi_k)\big(\tilde{f}_1(s,Z_k) - \tilde{f}_2(s,Z_k)m(s,Z_k)\big)K_h(Z_k - z)\,dH_k(s)}{f_2(X_k,Z_k)\bar{H}(X_k|z)f_Z(z)} = o_p((nh_n)^{-1/2}), \tag{B.12}$$

Combining (B.10)–(B.12), we have

$$I_{n31}(t) = \frac{1}{n}\sum_{k=1}^{n}\frac{\xi_k\big(1-\pi(X_k,Z_k)\big)\big(\delta_k - m(X_k,Z_k)\big)K_h(Z_k-z)f(X_k,Z_k)I(X_k \leqslant t)}{f_2(X_k,Z_k)\bar{H}(s|z)f_Z(z)} + o_p((nh_n)^{-1/2}) \tag{B.13}$$

uniformly for $t \in [0, \tau_z]$. Since $I_{n32}(t)$ can be represented as

$$I_{n32}(t) = \frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k} \times \Bigg[\frac{\xi_j(1-\xi_i)(\delta_j - m(X_i,Z_i))V_a(X_j - X_i)W_b(Z_j - Z_i)K_h(Z_i - z)}{f_2^2(X_i,Z_i)\bar{H}(X_i|z)f_Z(z)} \times\Big(\xi_k V_a(X_k - X_i)W_b(Z_k - Z_i) - f_2(X_i,Z_i)\Big)\Bigg] I(X_i \leqslant t) + o_p((nh_n)^{-1/2}),$$

similar argument as used in the proof of (B.13) shows

$$\sup_{t\in[0,\tau_z]} I_{n32}(t) = o_p((nh_n)^{-1/2}). \tag{B.14}$$

Notice $\pi(X, Z) = f_2(X, Z)/f(X, Z)$, by (B.9)–(B.14), we get

$$I_{n3}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i\big(1-\pi(X_i, Z_i)\big)\big(\delta_i - m(X_i, Z_i)\big)K_h(Z_j - z)I(X_i \leqslant t)}{\pi(X_i, Z_i)\bar{H}(X_i|z)f_Z(z)} + o_p((nh_n)^{-1/2}). \tag{B.15}$$

uniformly for $t \in [0, \tau_z]$. Combining (B.3), (B.6), (B.8) and (B.15), we complete the proof of (B.1). Under condition B, similar to (A.4), it can be proved

$$\hat{S}_n(t|z) - S(t|z) = \big(-\hat{\Lambda}_n(t|z) + \Lambda(t|z)\big)S(t|z) + o_p((nh_n)^{-1/2}) \tag{B.16}$$

uniformly for $t \in [0, \tau_z]$. (B.2) is then proved. □

Proof of Theorem 3.2. We only analyze $\hat{\Lambda}_n(t|z)$. The properties of $\hat{S}_n(t|z)$ can be obtained by (B.16). Using the fact $EIC_{ni}(t|z) = 0$, $\langle M_i(t), M_i(t)\rangle = \int_0^t I(X_i \geqslant s)\, d\Lambda(s|Z_i)$ and

$$E\left\{\frac{K_h^2(Z_i - z)}{\hat{f}_Z^2(z)}\,\frac{(\xi_i - \pi(X_i, Z_i))\big(\delta_i - m(X_i, Z_i)\big)I(X_i \leqslant t)}{\pi(X_i, Z_i)\bar{H}(X_i|z)}\int_0^t \frac{dM_i(s)}{\bar{H}_n(s-|z)}\right\} = 0,$$

we have

$$\begin{aligned}
&EIC_{ni}(t_1|z)IC_{ni}(t_2|z)\\
&= E\left\{\frac{K_h^2(Z_i - z)}{\hat{f}^2(z)}\int_0^{t_1\wedge t_2}\frac{I(X_i \geqslant s)\, d\Lambda(s|Z_i)}{\bar{H}_n^2(s-|z)}\right\}\\
&\quad + E\left\{\frac{K_h(Z_i - z)}{f(z)}\,\frac{(\xi_i - \pi(X_i, Z_i))\big(\delta_i - m(X_i, Z_i)\big)I(X_i \leqslant t_1\wedge t_2)}{\pi(X_i, Z_i)\bar{H}(X_i|z)}\right\}^2\\
&= f_z^{-1}(z)\mu(K)\int_0^{t_1\wedge t_2}\frac{d\Lambda(s|z)}{\bar{H}(s|z)}\\
&\quad + f_z^{-1}(z)\mu(K)\int_0^{t_1\wedge t_2}\frac{m(s, z)(1-m(s, z))(1-\pi(s, z))\, dH(s|z)}{\pi(s, z)\bar{H}^2(s|z)} + o(1)\\
&\equiv Cov(W(t_1|z), W(t_2|z)) + o(1).
\end{aligned}$$

By the multivariate central limit theorem and Lemma B.1, the finite-dimensional distribution of $n^{-1/2}\sum_{i=1}^n IC_{ni}(t|z)$ converge to the finite-dimensional distribution of a mean zero Gaussian process $W(t|z)$ with covariance function $Cov(W(t_1|z), W(t_2|z))$. On the other hand, notice that

$$\begin{aligned}
&\frac{(\xi_i - \pi(X_i, Z_i))\big(\delta_i - m(X_i, Z_i)\big)I(X_i \leqslant t)}{\pi(X_i, Z_i)\bar{H}(X_i|z)}\\
&= \frac{\big(\xi_i\delta_i - \xi_i m(X_i, Z_i) - \delta_i\pi(X_i, Z_i) + \pi(X_i, Z_i)m(X_i, Z_i)\big)I(X_i \leqslant t)}{\pi(X_i, Z_i)\bar{H}(X_i|z)},
\end{aligned}$$

thus $n^{-1/2}\sum_{i=1}^n IC_{ni}(t|z)$ can be represent as a sum of a martingale integral and some monotone processes on $[\tau_1, \tau_2]$. Thus the tightness of $n^{-1/2}\sum_{i=1}^n IC_{ni}(t|z)$ follows from properties of martingale integral and Example 2.11.16 of van der Vaart and Wellner [18]. Eq. (3.3) is then proved. □

Appendix C

Proof of Theorem 4.1. Note that

$$\lambda_n(t|z)\left[\ln\left(1-\frac{B_i}{R_i+\lambda_n(t|z)}\right)-\ln\left(1-\frac{B_i}{R_i}\right)\right]\geqslant 0.$$

By (4.3), (4.4) and inequality $|\ln(1-x)-\ln(1-y)|\geqslant|x-y|$ for $x, y\in(0,1)$, we get

$$\begin{aligned}
&\lambda_n(t|z)\Big[\ln(S(t|z))-\ln(\hat{S}_n(t|z))\Big]\\
&\quad=\sum_{i:X_i\leqslant t}\lambda_n(t|z)\ln\left(1-\frac{B_i}{R_i+\lambda_n(t|z)}\right)-\ln\left(1-\frac{B_i}{R_i}\right)\\
&\quad=\sum_{i:X_i\leqslant t}|\lambda_n(t|z)|\left|\ln\left(1-\frac{B_i}{R_i+\lambda_n(t|z)}\right)-\ln\left(1-\frac{B_i}{R_i}\right)\right|\\
&\quad\geqslant\sum_{i:X_i\leqslant t}|\lambda_n(t|z)|\left|\frac{B_i}{R_i+\lambda_n(t|z)}-\frac{B_i}{R_i}\right|\\
&\quad\geqslant\frac{\lambda_n(t|z)^2}{1+|\lambda_n(t|z)|/\min_{i:X_i\leqslant t}R_i}\sum_{i:X_i\leqslant t}\frac{B_i}{R_i^2}.
\end{aligned}\tag{C.1}$$

Using similar arguments used in the proof of Theorem 3.1, we get

$$\sum_{i:X_i\leqslant t}\frac{B_i}{R_i^2}=\sum_{i=1}^{n}\int_0^t\frac{\delta_i B_{ni}(z;h_n)\,dH_i(s)}{\bar{H}_n^2(s-|z)}=\int_0^t\frac{d\Lambda(s|z)}{\bar{H}(s|z)}+o(1)\quad a.s.$$

uniformly for $t\in[\tau_1,\tau_2]$. Almost surely for large n, for $t\in[\tau_1,\tau_2]$,

$$\min_{i:X_i\leqslant t}R_i\geqslant\bar{H}_n(t-|z)\geqslant\bar{H}(\tau_2|z)/2.$$

Thus together with the fact that $\sup_{t\in[\tau_1,\tau_2]}|\ln\hat{S}_n(t|z)-\ln S(t|z)|=O_p((nh_n)^{-1/2})$ proves

$$\sup_{t\in[\tau_1,\tau_2]}\lambda_n(t|z)=O_p((nh_n)^{-1/2}).$$

Repeating the proof of Lemma A.3 and Lemma A.4 of Li and van Keilegom [10], we get

$$\lambda_n(t|z)=\frac{\mu(K)\big(\ln S(t|z)-\ln\hat{S}_n(t|z)\big)}{\hat{f}_Z(z)\hat{\sigma}_0^2(t|z)}+O_p((nh_n)^{-1})\tag{C.2}$$

uniformly for $t\in[\tau_1,\tau_2]$, where

$$\hat{\sigma}_0^2(t|z)=\frac{\int K^2(u)\,du}{\hat{f}_Z(z)}\sum_{i:X_i\leqslant t}\frac{B_i}{R_i(R_i-B_i)}$$

and

$$-2\ln\mathcal{R}(S(t|z),t|z)=\lambda_n^2(t|z)\sum_{i=1}^{n}\frac{B_i}{R_i(R_i-B_i)}+o_p((nh_n)^{-1/2})$$

uniformly for $t \in [\tau_1, \tau_2]$. Thus

$$-2nh_n \frac{\hat{\sigma}_0^2(t|z)\hat{f}_Z(z)}{\hat{\sigma}^2(t|z)\mu(K)} \ln \mathcal{R}(S(t|z), t|z) = \frac{nh_n(\ln S(t|z) - \ln \hat{S}_n(t|z))^2}{\hat{\sigma}^2(t|z)} + o_p((nh_n)^{-1/2}).$$

Notice that

$$\hat{\sigma}^2(t|z) = \mu(K)\hat{f}_Z^{-1}(z) \int_0^t \frac{d\hat{\Lambda}_n(s|z)}{\bar{H}_n(s-|z)} + \int_0^t \frac{\hat{m}(s,z)(1-\hat{m}(s,z))(1-\hat{\pi}(s,z))\, dH_n(s|z)}{\hat{\pi}(s,z)\bar{H}_n^2(s-|z)}.$$

By the uniform consistencies of $\hat{\Lambda}_n(t|z)$, $H_n(t|z)$, $\hat{m}(t,z)$, $\hat{\pi}(t,z)$ for $t \in [\tau_1, \tau_2]$, with the similar argument to that of Theorem 3.1, we get $\sup_{t\in[\tau_1,\tau_2]} |\hat{\sigma}^2(t|z) - \sigma^2(t|z)| \xrightarrow{P} 0$. Thus

$$-2nh_n \frac{\hat{\sigma}_0^2(t|z)\hat{f}_Z(z)}{\hat{\sigma}^2(t|z)\mu(K)} \ln \mathcal{R}(S(t|z), t|z) \xrightarrow{\mathcal{W}} \frac{B^2(u)}{u(1-u)} \quad \text{in } D[\tau_1, \tau_2]. \qquad \square$$

References

[1] P.K. Andersen, O. Borgan, R.D. Gill, N. Keiding, Statistical Models Based on Counting Processes, Springer, New York, 1993.

[2] R. Beran, Nonparametric regression with randomly censored survival data, Technical Report, Dept. Statistics, Univ. California, Berkeley, 1981.

[3] D.M. Dabrowska, Nonparametric regression with censored survival time data, Scand. J. Statist. 14 (1987) 181–197.

[4] D.M. Dabrowska, Variable bandwidth conditional Kaplan–Meier estimate, Scand. J. Statist. 19 (1992) 351–361.

[5] G.E. Dinse, Nonparametric estimation for partially-complete time and type of failure data, Biometrics 38 (1982) 417–431.

[6] M. Hollander, I.W. McKeague, J. Yang, Likelihood ratio-based confidence bands for survival functions, J. Amer. Statist. Assoc. 92 (1997) 215–227.

[7] G. Li, On nonparametric likelihood ratio estimation of survival probabilities for censored data, Statist. Probab. Lett. 25 (1995) 95–104.

[8] G. Li, Optimal rate local smoothing in a multiplicative intensity counting process model, Math. Methods Statist. 6 (1997) 224–244.

[9] G. Li, H. Doss, An approach to nonparametric regression for life history data using local linear fitting, Ann. Statist. 23 (1995) 787–823.

[10] G. Li, I. van Keilegom, Likelihood ratio confidence bands in nonparametric regression with censored data, Scand. J. Statist. 29 (2002) 547–562.

[11] S.-H. Lo, Estimating a survival function with incomplete cause-of-death data, J. Multivariate Anal. 39 (1991) 217–235.

[12] I.W. McKeague, K.J. Utikal, Inference for a nonlinear counting process regression model, Ann. Statist. 18 (1990) 1172–1187.

[13] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.

[14] A.B. Owen, Empirical Likelihood, Chapman & Hall, London, 2001.

[15] W. Stute, *U*-statistics processes: a martingale approach, Ann. Probab. 22 (1994) 1725–1744.

[16] D.R. Thomas, G.L. Grunkemeier, Confidence interval estimation of survival probabilities for censored data, J. Amer. Statist. Assoc. 70 (1975) 865–871.

[17] M.J. van der Laan, I.W. McKeague, Efficient estimation from right-censored data when failure indicators are missing at random, Ann. Statist. 26 (1998) 164–182.

[18] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.