Ann. Inst. Statist. Math.
Vol. 53, No. 4, 708–729 (2001)

A BOOTSTRAP APPROACH TO NONPARAMETRIC REGRESSION FOR RIGHT CENSORED DATA

GANG LI[1] AND SOMNATH DATTA[2]

[1]*Department of Biostatistics, School of Public Health, University of California, Los Angeles, CA 90095-1772, U.S.A.*
[2]*Department of Statistics, Franklin College of Arts and Sciences, University of Georgia, 204 Statistics Building, Athens, GA 30602-1952, U.S.A.*, e-mail: datta@stat.uga.edu

(Received May 13, 1999; revised April 4, 2000)

Abstract. In this paper a two-stage bootstrap method is proposed for nonparametric regression with right censored data. The method is applied to construct confidence intervals and bands for a conditional survival function. Its asymptotic validity is established using counting process techniques and martingale central limit theory. The performance of the bootstrap method is investigated in a Monte Carlo study. An illustration is given using a real data.

Key words and phrases: Bootstrap, Beran's estimate, censoring, confidence bands, confidence intervals, Kaplan-Meier estimate, nonparametric regression, quantile regression.

1. Introduction and a review

In lifetime data analysis, nonparametrically estimated conditional survival curves (such as the conditional Kaplan-Meier estimate) are useful for assessing the influence of risk factors, predicting survival probabilities, and checking goodness-of-fit of various survival regression models. However, it has not been an easy task to assess the variability of the estimated conditional survival curves. Consider the right censored survival regression data consisting of n i.i.d. triples $(X_1, \delta_1, Z_1), \dots, (X_n, \delta_n, Z_n)$, where $X_i = \min(T_i, C_i)$, $\delta_i = I(T_i \le C_i)$, and $T_i \ge 0$, $C_i \ge 0$, and Z_i represent the survival time, the censoring time, and the covariate, respectively, for the i-th subject under study, $i = 1, \dots, n$. To ensure the identifiability of the model, we assume that for each i, T_i and C_i are conditionally independent given Z_i. Let $S(t \mid z) = P(T_i > t \mid Z_i = z)$ and $\Lambda(t \mid z) = -\int_0^t S(ds \mid z)/S(s- \mid z)$ denote the conditional survival function and the conditional cumulative hazard function of T_i given $Z_i = z$, respectively. We study the problem of constructing nonparametric confidence bands and intervals for $S(t \mid z)$ and $\Lambda(t \mid z)$ using the optimal rate conditional Kaplan-Meier estimate of Beran (1981). Such confidence bands and intervals can be used to assess the variability of the estimated conditional survival probabilities and provide a useful scale against which unusual features of the estimated conditional survival curve may be evaluated.

Nonparametric estimation of the conditional survival function and its related functions was initiated by Beran (1981) and has been further studied by Dabrowska (1987, 1989, 1992), Li (1997), Li and Doss (1995), McKeague and Utikal (1990), and others. For the convenience of discussion, let us consider a simple version of Beran's (1981) estimate $\hat{S}_h(t \mid z)$ of $S(t \mid z)$ that is defined as the Kaplan-Meier estimate constructed using only those observations whose covariate Z fall inside a neighborhood $(z - h_n, z + h_n)$ of z, where h_n is called the bandwidth (the general definition of Beran's estimate is given in Section 2). Because only a portion of the data are used, the rate of convergence of $\hat{S}_h(t \mid z)$ to $S(t \mid z)$ is typically slower than the $n^{-1/2}$ rate for the ordinary Kaplan-Meier estimate. The actual rate of convergence depends on how fast h_n goes to 0 as well as on the smoothness of S. A decrease in the bandwidth would reduce the bias but, on the other hand, increase the variance of the estimate, and vice versa. Under certain smoothness conditions, it has been shown that (cf. Dabrowska (1987) and Li (1997)) if h_n is of order $n^{-1/5}$, then $\hat{S}_h(t \mid z)$ converges to $S(t \mid z)$ at an optimal rate. Moreover,

$$(1.1) \qquad (nh_n)^{1/2}(\hat{S}_h(t \mid z) - S(t \mid z)) \xrightarrow{d} U(t \mid z),$$

where for fixed z, $U(t \mid z)$ is a continuous Gaussian martingale process with a *nonzero* mean (see (2.6) and (2.7) below for explicit expressions of the mean and variance function of U). In particular, the optimal rate of convergence for $\hat{S}_h(t \mid z)$ is $(nh_n)^{-1/2} = O(n^{-2/5})$.

It, however, remains an open problem as to how to construct confidence bands and intervals for $S(t \mid z)$ using the optimal rate weak convergence result (1.1). The major hurdles are that U has an unknown nonzero mean and that the distribution of $\sup_t |U(t \mid z)|$ is intractable. One possible solution is to "undersmooth" the Beran estimate: if $nh_n^5 \to 0$, then the limiting process U will have a zero mean and a Hall-Wellner (1980) type confidence band for $S(t \mid z)$ can be constructed (cf. Li and Doss (1995)). The drawback of this approach is that undersmoothing slows down the rate of convergence, which is not desired.

The main purpose of this paper is to study a bootstrap method for censored nonparametric regression and use it to construct confidence bands and intervals for $S(t \mid z)$ from the optimal rate Beran estimate. The basic idea of bootstrap is to first resample from the observed data, then reconstruct the estimate of interest, say Beran's estimate $\hat{S}_h^*(t \mid z)$, from the resampled data, and finally approximate the distribution of $(nh_n)^{1/2}(\hat{S}_h(t \mid z) - S(t \mid z))$ by that of its bootstrap version which can be obtained via computer simulation. An advantage of the bootstrap approach is that with an appropriately designed resampling method, the bootstrap will correctly account for the bias of the estimated survival function $\hat{S}_h(t \mid z)$. Therefore it does not require additional bias estimation or the use of suboptimal rate estimate (undersmoothing) for constructing confidence bands or intervals. It also automatically adapts to different variances of the estimated survival function at different covariate locations. However, it is not obvious what resampling scheme should be used for censored nonparametric regression as discussed below.

The idea of bootstrap was first introduced by Efron (1979) for homogeneous i.i.d. complete data setup in which the bootstrap is carried out by "resampling with replacement" from the sample data. This approach has since become a powerful tool in many statistical applications. See, for example, Efron and Tibshirani (1993) for the bootstrap method and its applications. Bootstrap for right censored data with no covariate was first studied by Efron (1981) who proposed two equivalent versions of bootstrap: a "simple" version and an "obvious" version. The "simple" bootstrap of Efron (1981) "resamples with replacement" from the observed pairs $\{(X_i, \delta_i), i = 1, \ldots, n\}$. The validity of this method was established by Akritas (1986) and Lo and Singh (1986). For the nonparametric right censored regression setup considered in this paper, it seemed at first sight that a natural extension of Efron's (1981) "simple" bootstrap would be to resample with replacement from the observed triples $\{(X_i, \delta_i, Z_i), i = 1, \ldots, n\}$. This resampling scheme, however, has a serious flaw that it fails to adequately take account of the type of bias of the optimal rate Beran estimate $\hat{S}_h$. Consequently, it will not give asymptotically correct results. This is very similar to the inappropriateness of the "usual" bootstrap in nonparametric density estimation and in ordinary nonparametric regression with complete data (cf. Hall (1992), Sections 4.4.2 and 4.5). Bootstrap in nonparametric regression with no censoring has been investigated by Hall ((1992), Section 4.5), Härdle and Bowman (1988) and Härdle and Marron (1991), among others. An ordinary nonparametric regression model for an observed data set (T_i, Z_i), $i = 1, \ldots, n$, is $T_i = m(Z_i) + \epsilon_i$, $i =, \ldots, n$, where $m(z) = E(T \mid Z = z)$ is the regression mean and the errors ϵ_i are independent and identically distributed with zero mean. In this case, the essential idea used in the aforementioned works is to resample from the estimated residuals $\hat{\epsilon}_i = T_i - \hat{m}_h(Z_i)$, $1 \le i \le n$ where $\hat{m}_h(z)$ is a pilot nonparametric estimate (such as the kernel estimate) of the conditional mean $m(z)$. Some of the resampling techniques for ϵ_i^* proposed in the literature include resampling from a set residuals determined by a window function (cf. Härdle and Bowman (1988)) and resampling each residual from a two-point distribution (*wild bootstrapping*) (cf. Härdle and Marron (1991)). After resampling, the bootstrap sample are formed as $\{(T_i^*, Z_i)\}_{i=1}^n$ where $T_i^* = \hat{m}_g(Z_i) + \epsilon_i^*$, and g is taken to be larger than h. Obviously, it is not easy to extend these methods to the nonparametric censored regression setup because the estimated residual is not available when the survival time T is censored.

In this article, we study a different resampling approach for the nonparametric censored regression setup. The proposed resampling is carried out in two steps. In the first stage, we resample with replacement from the set $\{Z_1, \ldots, Z_n\}$ to obtain the bootstrap sample for the covariate $\{Z_1^*, \ldots, Z_n^*\}$. Then, in the second stage, we generate a pair (X_i^*, δ_i^*) for each Z_i^* using ideas similar to that of Efron (1981). Let us use a special case to explain how it works. For each Z_i^* $(1 \le i \le n)$, we obtain (X_i^*, δ_i^*) by randomly selecting a pair from $\mathcal{A}_i = \{(X_j, \delta_j) : Z_i^* - g_n \le Z_j \le Z_i^* + g_n\}$, the set of observations for those cases whose covariate fall inside a neighborhood of Z_i^*, where $g_n > 0$ determines the size of the neighborhood. It can be shown (cf. Efron (1981)) that the resampling methods used in the second stage is equivalent to generating $T_i^* \sim \hat{S}_g(t \mid Z_i^*)$, $C_i^* \sim \hat{G}_g(t \mid Z_i^*)$ and letting $X_i^* = \min\{T_i^*, C_i^*\}$, $\delta_i^* = I(T_i^* \le C_i^*)$, where $\hat{S}_g(t \mid Z_i^*)$ and $\hat{G}_g(t \mid Z_i^*)$ are the Kaplan-Meier estimates of the survival distributions of T and C, respectively, constructed from the data $\mathcal{A}_i$. It is interesting to mention that if $g_n = 0$, then this two-step procedure is equivalent to resampling with replacement from the set of triples $\{(X_i, \delta_i, Z_i),\ i = 1, \ldots n\}$. This provides an intuitive explanation why "resample with replacement" from the sample triples would not be appropriate for bootstrapping the estimated conditional survival function $\hat{S}_h(t \mid z)$. In order to properly account for the bias of $\hat{S}_h(t \mid z)$, g_n has to be larger than h_n. Recall that h_n is the bandwidth used in the construction of Beran's estimate whose distribution needs to be bootstrapped. Although the technical reasons will be discussed in later sections, we point out that intuitively, a larger g_n enables one to catch more bias. The conditions imposed on g_n will be given in the next section. More discussion on this point can be found in Remarks 2.1 and 4.1.

Earlier, Van Keilegom and Veraverbeke (1997) studied bootstrap for censored nonparametric regression under fixed design where one only needs to resample the survival times from the conditional distribution. For random designs, one has to decide how to deal with the random covariate in the resampling process. Some discussion of the difference in bootstrapping fixed and random design linear regression models for complete data can be found in Freedman (1981). In this paper we propose a two-step bootstrap for random design that involves resampling the covariates in the first stage and the survival times in the second stage as discussed earlier. This does not resemble resampling residuals as what has been done for nonparametric regression with complete data. To the best of our knowledge, the two-step procedure is new in the literature even for complete data. The additional phase of resampling of the covariates also introduces more technical challenge for establishing the asymptotic validity of the bootstrap. Our justification of the two-stage bootstrap for random design is different from that of Van Keilegom and Veraverbeke (1997) for fixed design. Our approach may also be adopted to study bootstrap in nonparametric regression for other important incomplete survival data such as left-truncated data and both left-truncated and right-censored data, which are well known special cases of the counting process model. We finally note that bootstrap methods have been studied for some semiparametric regression models such as the Cox proportional hazards model; see, e.g., Hjort (1985) and Burr (1994), among others.

As mentioned earlier, a major use of bootstrap in this paper is to solve the problem of bandwidth selection in censored regression. We will present some numerical results on the performance of the proposed bootstrap method and illustrate its use through a real data example. It is worth noting that Li and Doss (1995) proposed a different class of estimators for the conditional survival function using local linear hazard smoothing. Their estimators have less bias near the boundary of the covariate space than Beran's estimators. Although nontrivial modifications may be needed, bootstrapping the estimators of Li and Doss (1995) could be studied along similar lines. Finally, we note that random design is more common than fixed design in clinical trials and epidemiological studies where the covariate is usually observational.

In Section 2, we give two equivalent resampling algorithms for bootstrapping the optimal rate Beran estimate of the conditional survival function. The algorithms actually use weighted resampling which include the local resampling discussed earlier as a special case. We state conditions on g_n under which the proposed method is asymptotically valid. We describe how to construct confidence intervals and equal precision bands using the proposed bootstrap method. We also discuss the use of bootstrap for data-driven bandwidth selection in censored nonparametric regression. Further extensions to pointwise and simultaneous confidence intervals for a quantile regression function at different covariate points are also discussed. In Section 3, we illustrate our method on a real data set and present numerical results about the performance of the proposed bootstrap method for finite sample sizes. The proofs are collected in Section 4.

2. Main results

2.1 *Notations and assumptions*

Recall that $A(t \mid z)$ and $S(t \mid z)$ denote the conditional cumulative hazard function and conditional survival function, respectively, of the survival time T given $Z = z$. Let $G(t \mid z) = P(C_i > t \mid Z_i = z)$ denote the conditional survival function of the censoring time C given $Z = z$. Assume that Z has a density $f(z)$ and let $\alpha(t \mid z) = \partial A(t \mid z)/\partial t$ denote the conditional hazard function of T given $Z = z$.

Let

$$N_i(t) = I(X_i \le t, \delta_i = 1) \quad \text{and} \quad Y_i(t) = I(X_i \ge t), \ i = 1, \ldots, n.$$

In addition, define

$$
\begin{aligned}
&H_1(t \mid z) = E(N_i(t) \mid Z_i = z) = P(X_i \le t, \delta_i = 1 \mid Z_i = z),\\
&H_2(t \mid z) = E(Y_i(t) \mid Z_i = z) = P(X_i \ge t \mid Z_i = z), \qquad (2.1)\\
&H_k(t, z) = H_k(t \mid z) f(z), \qquad k = 1, 2.
\end{aligned}
$$

The following assumptions will come into play in what follows.

ASSUMPTION A. Let $\omega(\cdot)$ be a density function vanishing outside $[-1, 1]$.
(A.1) $\int u\omega(u) = 0$ and $\int u^2\omega(u)du < \infty$.
(A.2) $\omega(\cdot)$ is of bounded variation.
(A.3) The derivative function $\omega'(\cdot)$ is of bounded variation.

Let $\tau > 0$ be a real number such that $S(\tau \mid z)G(\tau \mid z) > 0$, $I = [a, b]$ $(-\infty < a < b < \infty)$ be a fixed interval, and $I_\delta = (a - \delta, b + \delta)$ $(\delta > 0)$ be a δ-neighborhood of I. For any real function $f(t, z)$ and any interval J, denote $\|f\|_J^\tau = \sup\{|f(t, z)| : 0 \le t \le \tau, z \in J\}$.

ASSUMPTION B.
(B.1) $\inf\{H_2(\tau \mid z) : z \in I_\delta\} > 0$ and $\inf\{f(z) : z \in I_\delta\} > 0$ for some $\delta > 0$.
(B.2) $\|\partial^i H_k / \partial z^i\|_{I_\delta}^\tau < \infty$ for $i = 0, 1, 2, 3$ and $k = 1, 2$.
(B.3) For each $s \in [0, \tau]$, $\alpha(s \mid z)$ is twice differentiable with respect to z and satisfies

$$\alpha(s \mid w) = \alpha(s \mid z) + \alpha_z'(s \mid z)(w - z) + \alpha_{zz}''(s \mid z)(w - z)^2 + \gamma(s, z)(w - z)^3$$

where there exists a constant $K_1 > 0$ independent of s and z such that $\alpha(s \mid z)$, $\alpha_z'(s \mid z)$, $\alpha_{zz}''(s \mid z)$, and $\gamma(s, z)$ are all bounded by K_1 for all $s \in [0, \tau]$ and $z \in I$. Here α_z' and α_{zz}'' denote the first and second order partial derivatives of α with respect to z.

2.2 *The Beran estimate*

For readers' convenience, we review some results about the Beran-type estimates. The Beran (1981) kernel estimates of the conditional cumulative hazard function $A(t \mid z)$ and the conditional survival function $S(t \mid z)$ are defined by

$$\hat{A}_h(t \mid z) = \int_0^t \frac{\sum_{i=1}^n W_{hi}(z) dN_i(s)}{\sum_{i=1}^n W_{hi}(z) Y_i(s)}, \qquad \left(\frac{0}{0} \equiv 0\right), \qquad (2.2)$$

and $\hat{S}_h(t \mid z) = \prod_{s \le t}(1 - \Delta\hat{A}_h(s \mid z))$, where $\Delta\hat{A}_h(s \mid z) \equiv \hat{A}_h(s \mid z) - \hat{A}_h(s- \mid z)$ and for each i, $W_{hi}(z)$ is a kernel weight function given by

$$W_{hi}(z) = w\left(\frac{Z_i - z}{h_n}\right) \Big/ \sum_{j=1}^n w\left(\frac{Z_j - z}{h_n}\right), \qquad 1 \le i \le n, \qquad (2.3)$$

with $w(\cdot)$ being a density function in $\mathbf{R}$ and $h_n > 0$ is the bandwidth parameter. The Beran estimates can be considered as weighted average estimates. For example, the jump size $\Delta\hat{A}_h(t \mid z)$ of $\hat{A}_h(t \mid z)$ at time t is a weighted average of the jumps of the N_i's at time t for those subjects who are at risk. It is easy to see that if $W_{hi} \equiv 1$ for all i, then $\hat{A}_h$ and $\hat{S}_h$ reduce to the ordinary Nelson-Aalen estimate and Kaplan-Meier estimate for homogeneous data. If $w(\cdot)$ is taken to be the uniform density on $(-1, 1)$, then $\hat{A}_h$ is a local Nelson-Aalen estimator based on only those observations whose covariate Z fall inside the neighborhood $(z - h_n, z + h_n)$.

The bandwidth h_n controls the tradeoff between the bias and variance of the estimators: the smaller the h_n, the smaller the bias and the larger the variance. If $S(t \mid z)$ is twice differentiable with respect to z, then the optimal rate of convergence for $\hat{A}_h$ and $\hat{S}_h$ is of order $n^{-2/5}$ with h_n tending to zero at rate $n^{-1/5}$ (cf. Dabrowska (1992), Theorem 1 and Li (1997), Corollary 1). In particular, if $nh_n^5 \to c > 0$, then, under regularity conditions

$$\sqrt{nh_n}(\hat{A}_h(\cdot \mid z) - A(\cdot \mid z)) \xrightarrow{d} U(\cdot \mid z), \tag{2.4}$$

$$\sqrt{nh_n}(\hat{S}_h(\cdot \mid z) - S(\cdot \mid z)) \xrightarrow{d} S(\cdot \mid z)U(\cdot \mid z), \tag{2.5}$$

in $D[0,\tau]$ for any τ such that $S(\tau \mid z)G(\tau \mid z) > 0$, where $D[0,\tau]$ is the standard Skorohod space on $[0,\tau]$ and $U(\cdot \mid z)$ is a continuous Gaussian martingale with mean

$$\mu(t \mid z) = \sqrt{c}\left[\int u^2 w(u)du\right]\left[\int_0^t \alpha''_{zz}(s \mid z)ds\right] + \sqrt{c}\left[\int_0^t \frac{\alpha'_z(s \mid z)H'_{2z}(s,z)}{H_2(s,z)}ds\right]\left[\int u^2 w(u)du\right] \tag{2.6}$$

and variance function

$$\sigma^2(t \mid z) = \left[\int w^2(u)du\right]\left[\int_0^t \frac{\alpha(s \mid z)}{H_2(s,z)}ds\right]. \tag{2.7}$$

2.3 *Bootstrapping the Beran estimate*

We first give two equivalent algorithms for bootstrapping the Beran estimate of the conditional survival function. Then we state the main results that provide a theoretical justification of the proposed bootstrap method.

The Simple Weighted Bootstrap. Generate $Z_1^*, \ldots, Z_n^*$ i.i.d. from the empirical distribution of $\{Z_1, \ldots, Z_n\}$. For each i, generate a pair (X_i^*, δ_i^*) from the weighted empirical distribution $\hat{F}_g(\cdot,\cdot \mid Z_i^*)$ of $\{(X_1,\delta_1),\ldots,(X_n,\delta_n)\}$, where

$$\hat{F}_g(u,v \mid z) = \sum_{i=1}^n W_{gi}(z)I(X_i \le u, \delta_i \le v) \tag{2.8}$$

and $W_{gi}(z)$ is defined by (2.3) with a bandwidth g_n. Our bootstrap sample is formed as $\{(X_1^*, \delta_1^*, Z_1^*), \ldots, (X_n^*, \delta_n^*, Z_n^*)\}$.

The obvious bootstrap. Let $\hat{S}_g(t \mid z)$ and $\hat{G}_g(t \mid z)$ be the Beran estimates of $S(t \mid z)$ and $G(t \mid z)$, respectively, using the same weight function $W_{gi}(z)$. Here we force $\hat{S}_g(t \mid Z_i^*)$ and $\hat{G}_g(t \mid Z_i^*)$ to 0 beyond the larger of the last jump points of the two step functions to make both proper survival functions in order to sample the failure and censoring times described below. Generate $Z_1^*, \ldots, Z_n^*$ i.i.d. from the empirical distribution of $\{Z_1, \ldots, Z_n\}$. For each i, generate T_i^* from $\hat{S}_g(t \mid Z_i^*)$ and C_i^* from $\hat{G}_g(t \mid Z_i^*)$ independently, and define

$$X_i^* = \min(T_i^*, C_i^*), \qquad \delta_i^* = I(T_i^* \le C_i^*). \tag{2.9}$$

Our bootstrap sample is $\{(X_1^*, \delta_1^*, Z_1^*), \ldots, (X_n^*, \delta_n^*, Z_n^*)\}$.

PROPOSITION 2.1. *Suppose that there is no ties in the sample values $X_1, \dots, X_n$. The simple weighted bootstrap and the obvious bootstrap are equivalent. More precisely, if (X_i^*, δ_i^*) are defined by (2.9), then the conditional distribution of (X_i^*, δ_i^*) given Z_i^* is $\hat{F}_g(\cdot, \cdot \mid Z_i^*)$, where $\hat{F}_g(\cdot, \cdot \mid Z_i^*)$ is defined by (2.8).*

The equivalence of the two resampling algorithms is parallel to that of Efron's (1981) resampling methods for homogeneous data. We note that the obvious bootstrap algorithm has a clear motivation on its own. On the other hand, the simple weighted bootstrap algorithm is more convenient for computer implementation. In our simulations with a uniform kernel function, the "simple" bootstrap was much faster than the "obvious" version.

Let $\hat{S}_g^*(t \mid z)$ and $\hat{A}_g^*(t \mid z)$ be the Beran estimates of $S(t \mid z)$ and $A(t \mid z)$, respectively, from the bootstrap sample using the bandwidth g_n. We propose to estimate the distributions of $\sqrt{nh_n}(\hat{A}_h(\cdot \mid z) - A(\cdot \mid z))$ and $\sqrt{nh_n}(\hat{S}_h(\cdot \mid z) - S(\cdot \mid z))$ by their bootstrap counterparts $\sqrt{nh_n}(\hat{A}_h^*(\cdot \mid z) - \hat{A}_g(\cdot \mid z))$ and $\sqrt{nh_n}(\hat{S}_h^*(\cdot \mid z) - \hat{S}_g(\cdot \mid z))$, respectively. The distributions of the bootstrap processes can then be approximated by generating a large number of bootstrap samples.

In order to properly account for the bias, g_n needs to be slightly larger than h_n. In the following theorem we give conditions under which the proposed bootstrap method "works".

THEOREM 1. *Let $z \in (a, b)$ be fixed. Assume* (A.1)-(A.3) *and* (B.1)-(B.3). *If $h_n = cn^{-1/5}$ $(0 < c < \infty)$ and g_n satisfies the following conditions*

$$\frac{g_n^r}{h_n} = O(1) \quad for\ some \quad 0 < r < 2, \ and \ [-\log g_n]\left(\frac{h_n}{g_n}\right)^3 \to 0, \tag{2.10}$$

then, conditional on $(X_1, \delta_1, Z_1), \dots, (X_n, \delta_n, Z_n)$,

$$\sqrt{nh_n}(\hat{A}_h^*(t \mid z) - \hat{A}_g(t \mid z)) \xrightarrow{d} U(\cdot \mid z), \tag{2.11}$$

$$\sqrt{nh_n}(\hat{S}_h^*(t \mid z) - \hat{S}_g(t \mid z)) \xrightarrow{d} S(\cdot \mid z)U(\cdot \mid z), \tag{2.12}$$

in $D[0, \tau]$ for almost all sample sequences $(X_1, \delta_1, Z_1), (X_2, \delta_2, Z_2), \dots$, where the distributions of the limiting processes $U(\cdot \mid z)$ and $S(\cdot \mid z)U(\cdot \mid z)$ are the same as the limiting processes of $\sqrt{nh_n}(\hat{A}_h(\cdot \mid z) - A(\cdot \mid z))$ and $\sqrt{nh_n}(\hat{S}_h(\cdot \mid z) - S(\cdot \mid z))$ given in (2.4) *and* (2.5), *respectively.*

Remark 2.1. If $g_n \sim n^{-\beta}$, then it is easy to see that (2.8) is satisfied if and only if $\frac{1}{10} < \beta < \frac{1}{5}$. Thus, g_n should go to zero at a slower rate than h_n. It would be of interest to know what value of β should be recommended in practice. A general answer is beyond the scope of this paper and a thorough investigation is needed in future research. We observed that the value $\beta = 0.11$ has demonstrated satisfactory performance in term of coverage probabilities in our limited simulation studies. It is interesting that this value is very close to the theoretical value $\beta = 1/9$ obtained by Cao and González-Manteiga (1993) for uncensored regression data. We also observed in our simulations that the coverage probability were not very sensitive to small changes of β. One possible explanation is that the ratio $g_n/h_n \sim n^{.20-\beta}$ is not very sensitive to β $(0.10 < \beta < 0.20)$ when the sample size n is not extremely large.

2.4 *Bootstrap confidence intervals and bands for* $S(t \mid z)$

Confidence intervals for $S(t \mid z)$ at a fixed time t can be easily obtained by approximating the distribution of $\sqrt{nh_n}(\hat{S}_h(t \mid z) - S(t \mid z))$ by its bootstrap counterpart $\sqrt{nh_n}(\hat{S}_h^*(t \mid z) - \hat{S}_g(t \mid z))$.

Below we describe how to construct equal precision type confidence bands for the conditional survival function $S(t \mid z)$.

Define

$$\hat{\sigma}^2(t \mid z) = nh_n \int_0^t \sum_{i=1}^n c_i^2(s, z) d\hat{A}_h(s \mid Z_i),$$

where for $1 \leq i \leq n$,

$$c_i(t, z) = W_{hi}(z)Y_i(t) \Big/ \sum_{j=1}^n W_{hj}(z)Y_j(t).$$

Then, similar to (4.32) in Section 4, it can be shown that $\hat{\sigma}^2(t \mid z)$ converges almost surely to $\sigma^2(t \mid z)$ uniformly in $t \in [0, \tau]$. This, together with Theorem 1 and the strong uniform consistency of $S(t \mid z)$ (cf. Dabrowska (1989)), implies that

$$\frac{\sqrt{nh_n}(\hat{S}_h^*(t \mid z) - \hat{S}_g(t \mid z))}{\hat{S}_h(\cdot \mid z)\hat{\sigma}(\cdot \mid z)} \xrightarrow{d} \frac{U(\cdot \mid z)}{\sigma(\cdot \mid z)},$$

where the limiting distribution is the same as that of $\frac{\sqrt{nh_n}(\hat{S}_h(t|z) - S(t|z))}{\hat{S}_h(\cdot|z)\hat{\sigma}(\cdot|z)}$. Hence, we have the following result.

THEOREM 2. *Assume that the conditions of Theorem* 1 *hold. Let* $0 < \alpha < 1$ *be a fixed constant. Choose* γ_n *from the bootstrap distribution so that*

$$P\left\{ \sup_{t \in [0,\tau]} \frac{\sqrt{nh_n}|\hat{S}_h^*(t \mid z) - \hat{S}_g(t \mid z)|}{\hat{S}_h(t \mid z)\hat{\sigma}(t \mid z)} \leq \gamma_n \mid (X_i, \delta_i, Z_i), i = 1, \ldots, n \right\} = 1 - \alpha. \tag{2.13}$$

Then

$$P\left\{ \hat{S}_h(t \mid z) - \gamma_n \frac{\hat{S}_h(t \mid z)\hat{\sigma}(t \mid z)}{\sqrt{nh_n}} \leq S(t \mid z) \leq \hat{S}_h(t \mid z) + \gamma_n \frac{\hat{S}_h(t \mid z)\hat{\sigma}(t \mid z)}{\sqrt{nh_n}} \right.$$
$$\left. \text{for all} \quad t \in [0, \tau] \right\} \to 1 - \alpha. \tag{2.14}$$

It is not hard to show that if $\hat{S}_h(t \mid z)\hat{\sigma}(t \mid z)$ in (2.11) is replaced by its bootstrapped version, then the conclusion in (2.12) still holds.

Bootstrap confidence bands for the conditional cumulative function can be constructed similarly.

3. Application and simulation study

In this section we illustrate the proposed bootstrap method on a data set involving survival of patients diagnosed of primary biliary cirrhosis of the liver (PBC). We also carry out a simulation study to assess the performance of the bootstrap confidence bands and intervals.

The PBC data come from a randomized clinical trial study conducted by the Mayo Clinic between January 1974 and May 1984, comparing the drug D-penicillamine (DPCA) with a placebo. Among the 424 PBC patients who met eligibility criteria, 312 cases participated the randomized trial and contain largely complete data. For each patient, the date of randomization, the disease and survival status as of July 1986, and a large number of risk factors such as age and serum libirubin were recorded. As of July 1986, 124 were observed to die from the disease and the remaining were censored observations. Of the additional 112 cases who did not participate in the randomized trial, but consented to have basic measurements recorded and to be followed for survival, the data are available on only 106 cases because six of those cases lost follow-up shortly after diagnosis. The data are given in Fleming and Harrington ((1991), Appendix D.1).

In their analysis, Fleming and Harrington (1991) showed that there are no detectable differences between the distributions of survival times for the DPCA and the placebo groups. Thus we combined the two groups in the randomized trial to study the association between the survival time and risk factors. Fleming and Harrington (1991) also showed that the variable Z="*serum bilirubin*", (in mg/dl) is the strongest univariate predictor of survival for PBC patients. Although a complete analysis requires the use of other risk factors, for illustration purpose, here we focus only on the influence of *serum bilirubin* on survival probabilities. Specifically, we construct bootstrap confidence bands and intervals for the conditional survival curve $S(t \mid z)$ for given z using the 312 cases participated the randomized trial. 1000 bootstrap samples were used in the studies.

Figure 1 gives the 90% bootstrap pointwise confidence intervals for the conditional survival functions $S(t \mid Z = 0.5)$ and $S(t \mid Z = 2.35)$ at times $t = i \times 365/4, i = 1, \ldots, 20$. Figure 2 gives the 90% bootstrap simultaneous confidence bands for $S(t \mid Z = 0.5)$ and $S(t \mid Z = 2.35)$ on the interval $[0, 1825]$ (in days), or over the first 5 years. For $Z = 0.5$, we used $c = 1.7$ which gives $h_n = cn^{-1/5} = 0.54$ and $g_n = cn^{-0.11} = 0.904$. For $Z = 2.35$, we used $c = 6.6$ which corresponds to $h_n = 2.09$ and $g_n = 3.51$. Here c was chosen to minimize the bootstrapped integrated mean square error. Details of a bootstrap bandwidth selection procedure can be found in Li and Datta (1999).

It is seen from Fig. 1 that a patient with $Z = 0.5$ has a very high chance of survival for more than 5 years. For instance, the 90% confidence interval for $S(1825 \mid Z = 0.5)$ is $[0.91, 0.97]$. That is, with 90% confidence, the 5-year survival probability of a patient with $Z = 0.5$ is at least .91 and it can be as high as .97. On the other hand, the 90% confidence interval for $S(1825 \mid Z = 2.35)$, the 5-year survival probability of a patient with $Z = 2.35$, is $[0.77, 0.86]$ which is noticeably lower than that for $Z = 0.5$. Furthermore, both the confidence intervals and bands revealed that the differences between the estimated survival curves $\hat{S}(t \mid 0.5)$ and $\hat{S}(t \mid 2.35)$ are not just caused by random variation. Instead, they reflect the significant drop in survival probability when Z is increased from 0.5 to 2.35. Finally, we point out that the data are too sparse to produce useful confidence intervals or bands for survival probabilities when Z is large (e.g., $Z = 15$), unless one use a very large bandwidth which in turn introduces severe bias. Detailed analysis of the PBC data using various parsimonious models can be found in Fleming and Harrington (1991).

We finally mention that we also carried out a similar analysis by including the additional 106 cases. As expected, we observed very little differences from the results presented above.

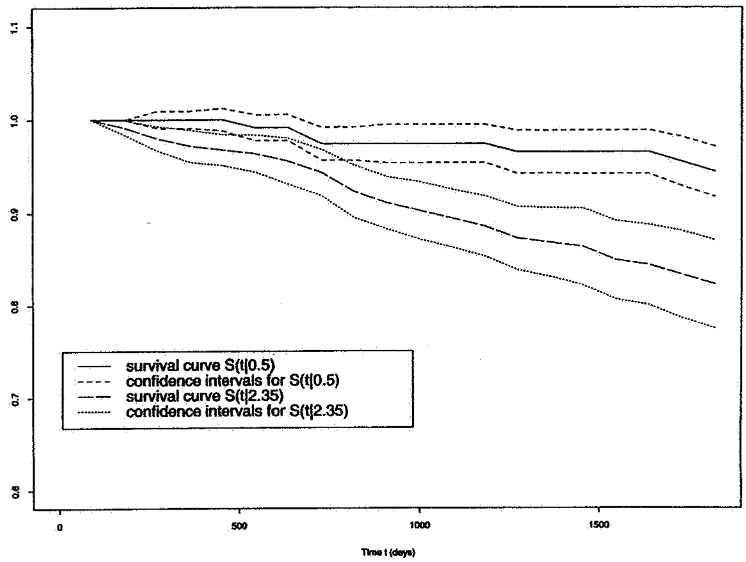

Fig. 1. 90% confidence intervals for $S(t \mid 0.5)$ and $S(t \mid 2.35)$.

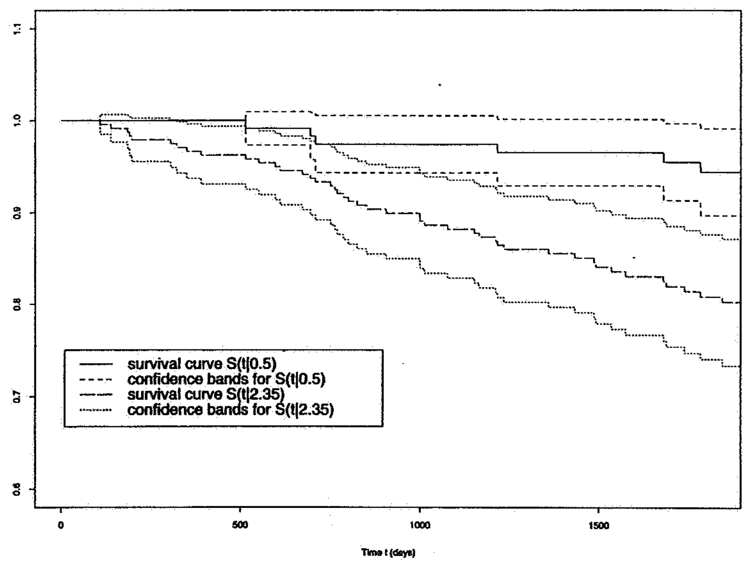

Fig. 2. 90% simultaneous confidence bands for $S(t \mid 0.5)$ and $S(t \mid 2.35)$ over 5 years.

In Table 1 we report the results of simulations to estimate the coverage probabilities of the bootstrap confidence band for the conditional survival function $S(t \mid 0.5)$ on the interval $[0, 2]$. Here the covariate Z has a uniform distribution on (0,1) and the conditional survival and censoring distributions (given $Z = z$) are $F(t \mid z) = 1 - \exp(-z^2 t)$ and $G(t \mid z) = 1 - \exp(-bz^2 t)$. The parameter b is adjusted to give the prescribed censoring rate. Note that for this example, $P(T > C \mid Z = z) = b/(1 + b)$ which does not involve z. In the simulation, we used $h_n = cn^{-1/5}$ and $g_n = cn^{-0.11}$ with

Table 1. Observed coverage probabilities of bootstrap bands for $S(t \mid z)$ over the interval $[0, 2]$.

Sample size	Nominal Level=.90				Nominal Level=.95			
	Censoring rate				Censoring rate			
n	5%	20%	40%	70%	5%	20%	40%	70%
50	.91	.92	.90	.90	.95	.95	.93	.95
100	.94	.92	.92	.81	.96	.96	.95	.88
200	.88	.88	.87	.80	.91	.91	.89	.87
500	.91	.90	.90	.84	.97	.98	.97	.95
1000	.90	.90	.91	.85	.95	.95	.95	.94

Table 2. Observed coverage probabilities of bootstrap confidence intervals for $S(t \mid z)$ at $t = 2$.

Sample size	Nominal Level=.90				Nominal Level=.95			
	Censoring rate				Censoring rate			
n	5%	20%	40%	70%	5%	20%	40%	70%
50	.91	.90	.87	.80	.96	.94	.91	.85
100	.92	.90	.88	.71	.97	.97	.95	.83
200	.91	.93	.94	.73	.97	.99	.95	.78
500	.97	.95	.92	.83	.98	.97	.99	.89
1000	.89	.90	.88	.77	.95	.94	.93	.90

$c = 0.70$ which is close to the optimal value of c for a global bandwidth. For a given sample, 1,000 bootstrap samples were used to construct confidence bands and intervals. Each entry in the table was based on 100 Monte Carlo samples. (It would be ideal to use a larger number of Monte Carlo samples. However, this would require enormous computation time for the simulation when the sample size n exceeds 500. We did run the simulation for $n = 50$ and 100 using 1,000 Monte Carlo samples and the results were consistent with those reported here.)

Table 2 was similarly constructed except that it reports estimated coverage probabilities of the bootstrap confidence intervals for a single conditional survival probability $S(2 \mid 0.5)$.

It is seen from both tables that with the exception of heavy censoring, the coverage probabilities are observed to be close to their nominal values. In the case of heavy censoring, a much larger sample size would be needed.

4. Proofs

PROOF OF PROPOSITION 2.1. The proof is similar to that of Efron (1981) and is omitted. A detailed proof can be obtained from the author upon request. □

Next we prove Theorem 1.

Because (2.12) is a consequence of (2.11) and the functional delta method, we only prove (2.11). Our proof of (2.11) is carried out by verifying the conditions of the martingale central limit theorem for $\sqrt{nh_n}(\hat{A}_h^*(t \mid z) - \hat{A}_g(t \mid z))$, conditional on the data. It is important to note that the techniques involved in the proof are substantially different from those used in Li (1997) for establishing weak convergence of $\sqrt{nh_n}(\hat{A}_h(t \mid z) - A(t \mid z))$. In particular, we need to derive the rate of uniform (in z) almost sure convergence for the Beran estimate $\hat{A}$ and its partial derivative $\partial\hat{A}/\partial z$ and for some nonparametric mean regression estimators. In contrast, only weak consistency results with fixed z were needed in Li (1997).

The following lemma gives the rates of uniform strong convergence for $\hat{A}$ and $\partial\hat{A}/\partial z$, which will play a crucial role in the proof of Theorem 1.

LEMMA 1. (Rate of uniform strong convergence for $\hat{A}_g$ and its first order derivative) *Assume that* (A.1), (B.1), *and* (B.2) *hold.*

(a) *If* $\alpha_n = (\frac{\log g_n^{-1}}{ng_n})^{1/2} \to 0$ *and* $\sum_{n=1}^{\infty} g_n^{\rho} < \infty$ *for some* $\rho > 0$, *then, with probability* 1,

$$\|\hat{A}_g - A\|_I^{\tau} = O(\max(\alpha_n, g_n^2)) \quad as \quad n \to \infty, \tag{4.1}$$

where $\|f\|_I^{\tau} = \sup\{|f(t,z)| : 0 \le t \le \tau, z \in I\}$ *for any real function* $f(t,z)$.

(b) *If* $\beta_n = (\frac{\log g_n^{-1}}{ng_n^3})^{1/2} \to 0$ *and* $\sum_{n=1}^{\infty} g_n^{\rho} < \infty$ *for some* $\rho > 0$, *then, with probability* 1,

$$\left\|\frac{\partial\hat{A}_g}{\partial z} - \frac{\partial A}{\partial z}\right\|_I^{\tau} = O(\max(\beta_n, g_n^2)) \quad as \quad n \to \infty. \tag{4.2}$$

PROOF. (a) Let

$$\hat{H}_{g1}(t,z) = \frac{1}{ng_n}\sum_{i=1}\omega\left(\frac{Z_i - z}{g_n}\right)N_i(t) \quad \text{and} \quad \hat{H}_{g2}(t,z) = \frac{1}{ng_n}\sum_{i=1}\omega\left(\frac{Z_i - z}{g_n}\right)Y_i(t).$$

It follows from Dabrowska ((1989), p. 1165, lines 5 and 7) that

$$P(\|\hat{H}_{gk} - E\hat{H}_{gk}\|_I^{\tau} > \epsilon) \le c_{0k}g_n^{-1}\exp(-c_{1k}\epsilon^2 ng_n), \qquad k = 1,2, \quad \text{for all } \epsilon > 0, \tag{4.3}$$

where c_{01}, c_{02}, c_{11}, and c_{12} are some universal positive constants. Letting $\epsilon = \alpha_n\{(1+\rho)/c_{1k}\}^{1/2}$, the above inequalities reduces to

$$P(\|\hat{H}_{gk} - E\hat{H}_{gk}\|_I^{\tau} > \alpha_n\{(1+\rho)/c_{1k}\}^{1/2}) \le c_{0k}g_n^{\rho}, \quad k = 1,2.$$

This, combined with the Borel-Cantelli lemma, implies that with probability 1,

$$\|\hat{H}_{gk} - E\hat{H}_{gk}\|_I^{\tau} = O(\alpha_n), \qquad k = 1,2. \tag{4.4}$$

Note that $E\hat{H}_{gk}(t,z) = \int \omega(u)H_k(t, z + g_n u)du$ for $k = 1,2$. An application of the Taylor expansion and (A.1) leads to

$$\|E\hat{H}_{gk}(t,z) - H_k(t,z)\|_I^{\tau} \le g_n^2\left[\int u^2\omega(u)du\right]\left\|\frac{\partial^2 H_k}{\partial z^2}\right\|_{I_6}^{\tau} = O(g_n^2), \qquad k = 1,2, \tag{4.5}$$

for sufficiently large n.

It follows from (4.4) and (4.5) that

$$\|\hat{H}_{gk} - H_k\|_I^{\tau} \xrightarrow{a.s.} 0 \qquad k = 1,2. \tag{4.6}$$

Write

$$\hat{A}_g(t \mid z) - A(t \mid z) = \int_0^t \frac{1}{E\hat{H}_{g2}(s,z)} d[\hat{H}_{g1}(s,z) - E\hat{H}_{g1}(s,z)]$$
$$+ \int_0^t \left[\frac{1}{\hat{H}_{g2}(s,z)} - \frac{1}{E\hat{H}_{g2}(s,z)}\right] d\hat{H}_{g1}(s,z)$$
$$+ \int_0^t \frac{1}{H_2(s,z)} d[E\hat{H}_{g1}(s,z) - H_1(s,z)]$$
$$+ \int_0^t \left[\frac{1}{E\hat{H}_{g2}(s,z)} - \frac{1}{H_2(s,z)}\right] dE\hat{H}_{g1}(s,z)$$

and denote by I_l the l-th term on the right hand side of the equality, $l = 1, \ldots, 4$. It can be shown that there exist some positive constants K_1, K_2, K_3 and K_4 such that for sufficiently large n

$$\|I_1\|_I^\tau \le K_1 \|\hat{H}_{g1} - E\hat{H}_{g1}\|_I^\tau, \qquad \|I_2\|_I^\tau \le K_2 \|\hat{H}_{g2} - E\hat{H}_{g2}\|_I^\tau,$$
$$\|I_3\|_I^\tau \le K_3 \|E\hat{H}_{g1} - H_1\|_I^\tau, \qquad \text{and} \qquad \|I_4\|_I^\tau \le K_4 \|E\hat{H}_{g2} - H_2\|_I^\tau.$$

Therefore, for sufficiently large n

$$\|\hat{A}_g - A\|_I^\tau \le \|I_1\|_I^\tau + \|I_2\|_I^\tau + \|I_3\|_I^\tau + \|I_4\|_I^\tau \le K \sum_{k=1}^{2} (\|\hat{H}_{gk} - E\hat{H}_{gk}\|_I^\tau + \|E\hat{H}_{gk} - H_k\|_I^\tau) \tag{4.7}$$

with probability 1, for some positive constant K. This, combined with (4.4) and (4.5), implies (4.1).

(b) Now we prove (4.2). Similar to (4.3), it can be shown that

$$P\left(\left\|\frac{\partial \hat{H}_{gk}}{\partial z} - E\frac{\partial \hat{H}_{gk}}{\partial z}\right\|_I^\tau > \epsilon\right) \le b_{0k} g_n^{-1} \exp(-b_{1k}\epsilon^2 n g_n^3), \qquad k = 1, 2,$$
$$\text{for all} \quad \epsilon > 0,$$

for some positive constants b_{01}, b_{02}, b_{11} and b_{12}. This, together with the argument leading to (4.4), implies that with probability 1,

$$\left\|\frac{\partial \hat{H}_{gk}}{\partial z} - E\frac{\partial \hat{H}_{gk}}{\partial z}\right\|_I^\tau = O(\beta_n), \qquad k = 1, 2. \tag{4.8}$$

Similar to (4.5), the Taylor expansion and (A.1) leads to

$$\left\|E\left[\frac{\partial \hat{H}_{gk}}{\partial z}\right] - \frac{\partial H_k}{\partial z}\right\|_I^\tau \le g_n^2 \left[\int u^2 \omega(u) du\right] \left\|\frac{\partial^3 H_k}{\partial z^3}\right\|_{I_\delta}^\tau = O(g_n^2), \qquad k = 1, 2. \tag{4.9}$$

Finally,

$$\frac{\partial \hat{A}_g(t \mid z)}{\partial z} = \int_0^t \frac{d\left[\frac{\partial \hat{H}_{g1}(s,z)}{\partial z}\right]}{\hat{H}_{g2}(s,z)} - \int_0^t \frac{\left[\frac{\partial \hat{H}_{g2}(s,z)}{\partial z}\right]}{[\hat{H}_{g2}(s,z)]^2} d\hat{H}_{g1}(s,z)$$

and

$$\frac{\partial A(t \mid z)}{\partial z} = \int_0^t \frac{d\left[\frac{\partial H_1(s,z)}{\partial z}\right]}{H_2(s,z)} - \int_0^t \frac{\left[\frac{\partial H_2(s,z)}{\partial z}\right]}{[H_2(s,z)]^2} dH_1(s,z).$$

Similar to (4.7), it can be shown that there exists a positive constant K such that for sufficiently large n,

$$\left\| \frac{\partial \hat{A}_g}{\partial z} - \frac{\partial A}{\partial z} \right\|_I^{\mathcal{T}} \le K \sum_{k=1}^{2} \left(\|\hat{H}_{gk} - E\hat{H}_{gk}\|_I^{\mathcal{T}} + \|E\hat{H}_{gk} - H_k\|_I^{\mathcal{T}} \right.$$
$$\left. \cdot \left\| \frac{\partial \hat{H}_{gk}}{\partial z} - E\frac{\partial \hat{H}_{gk}}{\partial z} \right\|_I^{\mathcal{T}} + \left\| E\frac{\partial \hat{H}_{gk}}{\partial z} - \frac{\partial H_k}{\partial z} \right\|_I^{\mathcal{T}} \right)$$

with probability 1. This, combined with (4.4), (4.5), (4.8), and (4.9), proves (4.2). □

LEMMA 2. *Assume that* $(\xi_1, Z_1), \ldots, (\xi_n, Z_n)$ *are iid random vectors taking values in* $[0,1] \times \mathbf{R}$ *and that the density function* $f(z)$ *of* Z_i *satisfies* $\inf_{a \le z \le b} f(z) > 0$ *for some* $-\infty < a < b < \infty$. *Define*

$$m(z) = E(\xi_i \mid Z_i = z) \quad \text{and} \quad m_n(z) = \sum_{i=1}^{n} W_{hi}(z)\xi_i,$$

where $W_{hi}(z)$ *is defined by* (2.3). *In addition, for any function* $g(z)$, *define*

$$\psi(z) = \frac{d[f(z)g(z)]}{dz} \int u^2 \omega(u)du \quad \text{and} \quad \psi_n(z) = \frac{1}{nh_n^3} \sum_{i=1}^{n} \omega\left(\frac{Z_i - z}{h_n}\right)(Z_i - z)g(Z_i).$$

Assume that (A.1) *and* (A.2) *hold and that* $g(z)$ *is continuously differentiable and* $m(z)$ *and* $f(z)$ *are twice continuously differentiable on* $(a-\epsilon, b+\epsilon)$ *for some* $\epsilon > 0$.

(a) *If* $h_n \to 0$ *and for some fixed* $0 < r < 2$,

$$\sum_{n=1} \exp(-\rho n h_n^{2+2r}) < \infty, \quad \text{for all} \quad \rho > 0, \tag{4.10}$$

then, as $n \to \infty$,

$$\sup_{a \le z \le b} h_n^{-r} |m_n(z) - m(z)| \xrightarrow{a.s.} 0. \tag{4.11}$$

(b) *If* (4.10) *holds for* $r = 1$, *then*

$$\sup_{a \le z \le b} |\psi_n(z) - \psi(z)| \xrightarrow{a.s.} 0. \tag{4.12}$$

PROOF. (a) Let $\phi(z) = m(z)f(z)$,

$$\phi_n(z) = (nh_n)^{-1} \sum_{i=1}^{n} \omega\left(\frac{Z_i - z}{h_n}\right)\xi_i \quad \text{and} \quad f_n(z) = (nh_n)^{-1} \sum_{i=1}^{n} \omega\left(\frac{Z_i - z}{h_n}\right). \tag{4.13}$$

Then, $m_n(z) = \phi_n(z)/f_n(z)$. Thus it suffices to show that

$$\sup_{a \le z \le b} h_n^{-r} |\phi_n(z) - \phi(z)| \xrightarrow{a.s.} 0, \quad \text{as} \quad n \to \infty, \tag{4.14}$$

and that

$$\sup_{a\leq z\leq b} h_n^{-r}|f_n(z)-f(z)| \xrightarrow{a.s.} 0, \quad \text{as} \quad n\to\infty. \tag{4.15}$$

By (A.1) and the Taylor series expansion, we have

$$E\phi_n(z)-\phi(z) = h_n^2\int u^2\omega(u)\phi''(z+\lambda h_n u)du, \qquad (|\lambda|\leq 1.)$$

Hence, as $n\to\infty$,

$$\begin{aligned}\sup_{a\leq z\leq b} h_n^{-r}|E\phi_n(z)-\phi(z)| \\ \leq h_n^{2-r}\left[\sup_{a\leq z\leq b}\sup_{|\delta|\leq h_n}|\phi''(z+\delta)-\phi''(z)|+\phi''(z)\right]\int u^2\omega(u)du \to 0.\end{aligned} \tag{4.16}$$

On the other hand, Lemma 2 of Nadaraya (1970) implies that

$$\begin{aligned}P\left(\sup_{a\leq z\leq b} h_n^{-r}|\phi_n(z)-E\phi_n(z)|>\rho\right) &\leq c_1\exp(-\alpha_1\rho^2 nh_n^{2+2r}) \\ &+c_2\exp(-\alpha_2\rho^2 nh_n^{2+2r}), \qquad \rho>0,\end{aligned}$$

for some universal positive constants c_1, c_2, α_1, and α_2. This, combined with the Borel-Cantelli lemma and (4.16), implies that (4.14) holds. We can prove (4.15) along the same lines. Therefore (4.11) is proved.

(b) From the Taylor expansion and (A.1), we have

$$E\psi_n(z) = \int u^2 w(u)\frac{d[f(y)g(y)]}{dy}\bigg|_{y=z+\lambda h_n u} du, \qquad |\lambda|\leq 1.$$

Thus, similar to (4.16), we have

$$\sup_{a\leq z\leq b}|E\psi_n(z)-\psi(z)|\to 0. \tag{4.17}$$

Let $F^Z(z)$ denote the cumulative distribution function of Z_1 and let $F_n^Z(z)$ be the empirical distribution function of $Z_1,\ldots,Z_n$. By integrating by parts, it is shown that for $z\in[a,b]$ and sufficiently large n,

$$|\psi_n(z)-E\psi_n(z)|\leq K_0 h_n^{-2}\sup|F_n^Z(u)-F^Z(u)|,$$

for some constant $K_0>0$, where the second equality is obtained by integrating by parts. This, combined with the following inequality (cf. Nadaraya (1970), (6))

$$P\left(\sup|F_n^Z(u)-F^Z(u)|>\frac{\lambda}{\sqrt{n}}\right)\leq c_0\exp(-2\lambda^2), \qquad (c_0>0 \text{ is a universal constant})$$

implies that

$$P\left(\sup_{a\leq z\leq b}|\psi_n(z)-E\psi_n(z)|>\rho\right)\leq c_0\exp(-\alpha_0\rho^2 nh_n^4),$$

where $\alpha_0 = 2/K_0^2$. Applying the Borel-Cantelli lemma, we have

$$\sup_{a\le z\le b} |\psi_n(z) - E\psi_n(z)| \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty. \tag{4.18}$$

Therefore, (4.12) follows from (4.17) and (4.18). □

LEMMA 3. *Let $Z_1^*, \ldots, Z_n^*$ be defined as in Subsection 2.3. Define $f_n^*(z) = \frac{1}{nh_n} \sum_{i=1}^n \omega(\frac{Z_i^*-z}{h_n})$ to be the bootstrap counterpart of $f_n(z)$ in (4.13). Assume that $h_n \to 0$ and $\sum_{n=1}^\infty \exp(-\rho n h_n) < \infty$ for all $\rho > 0$. Then, for almost all sequences $Z_1, Z_2, \ldots$, conditional on $(Z_1, \ldots, Z_n)$, $f_n^*(z) \to^P f(z)$ as $n \to \infty$ at every continuity point z of f.*

PROOF. Let z be a continuity point of f. By Rao ((1983), Theorem 3.1.5), $f_n(z) \longrightarrow^{\text{a.s.}} f(z)$ as $n \to \infty$. The conclusion follows from the facts that for almost all sequences $Z_1, Z_2, \ldots, E(f_n^*(z) \mid Z_1, \ldots, Z_n) = f_n(z) \to f(z)$, and $\text{Var}(f_n^*(z) \mid Z_1, \ldots, Z_n) = \frac{1}{nh_n} \cdot \frac{1}{nh_n} \sum_{i=1}^n \omega^2(\frac{Z_i - z}{h_n}) - \frac{1}{n} f_n^2(z) \to 0$. □

The following lemma gives a list of results that will be repeatedly used in the proof of Theorem 1.

LEMMA 4. *Assume that $h_n \to 0$ and $g_n \to 0$. Assume further that*

$$\sum_{n=1} \exp(-\rho n h_n^4) < \infty, \quad \textit{for all} \quad \rho > 0, \tag{4.19}$$

and that for some constant $0 < r < 2$,

$$\frac{g_n^r}{h_n} = O(1), \tag{4.20}$$

and

$$\sum_{n=1} \exp(-\rho n g_n^{2+2r}) < \infty, \quad \textit{for all} \quad \rho > 0. \tag{4.21}$$

In addition, assume that the density $f(z)$ of Z satisfies the conditions of Lemma 2. Let

$$\hat{H}_{h2}^*(s,z) = \frac{1}{nh_n} \sum_{i=1}^n \omega\Big(\frac{Z_i^* - z}{h_n}\Big) Y_i^*(s),$$

and, for $1 \le i \le n$, let

$$c_i^*(t,z) = W_{hi}^*(z) Y_i^*(t) \Big/ \sum_{j=1}^n W_{hj}^*(z) Y_j^*(t), \tag{4.22}$$

where $N_i^*(t)$, $Y_i^*(t)$, and $W_{hi}^*(z)$ are the bootstrap counterparts of $N_i(t)$, $Y_i(t)$, and $W_{hi}(z)$, respectively. Let $a < z < b$ be fixed. Then, conditional on $(X_1, \delta_1, Z_1), \ldots, (X_n, \delta_n, Z_n)$,

$$\sup_{s\in[0,\tau]} |\hat{H}_{h2}^*(s,z) - H_2(s,z)| \xrightarrow{P} 0, \tag{4.23}$$

$$\text{(4.24)} \quad \sup_{s\in[0,\tau]} \left| (nh_n)^\delta \sum_{i=1}^n c_i^{*1+\delta}(s,z) - [H_2(s,z)]^{-\delta} \int \omega^{1+\delta}(u)du \right| \xrightarrow{P} 0, \quad \delta > 0,$$

$$\text{(4.25)} \quad \sup_{s\in[0,\tau]} \left| h_n^{-2} \sum_{i=1}^n c_i^*(s,z)(Z_i^* - z) - H_2^{-1}(s,z) \left[\frac{\partial H_2(s,z)}{\partial z}\right] \int u^2 w(u)du \right| \xrightarrow{P} 0,$$

$$\text{(4.26)} \quad \sup_{s\in[0,\tau]} \left| h_n^{-2} \sum_{i=1}^n c_i^*(s,z)(Z_i^* - z)^2 - \int u^2 w(u)du \right| \xrightarrow{P} 0,$$

$$\text{(4.27)} \quad \sum_{i=1}^n \int_0^t |c_i^{*p}(ds,z)| = \sum_{i=1}^n c_i^{*p}(t,z) + O_p((nh_n)^{1-p}), \quad \textit{for any} \quad p > 0,$$

for almost all sequences $(X_1, \delta_1, Z_1), \ldots$, *where* $H_2(s,z)$ *is defined in* (2.1).

PROOF. Throughout the proof expectations are taken conditional on $(X_1, \delta_1, Z_1), \ldots, (X_n, \delta_n, Z_n)$.

(a) Let $0 \le s \le \tau$ be fixed and define

$$\text{(4.28)} \qquad \hat{H}_{g2}(s \mid z) = \sum_{i=1}^n W_{gi}(z) Y_i(s).$$

It can be shown that

$$E\hat{H}_{h2}^*(s,z) = \frac{1}{nh_n}\sum_{i=1}^n \omega\left(\frac{Z_i - z}{h_n}\right)[\hat{H}_{g2}(s \mid Z_i) - H_2(s \mid Z_i)] + \frac{1}{nh_n}\sum_{i=1}^n \omega\left(\frac{Z_i - z}{h_n}\right) H_2(s \mid Z_i).$$

Denote by I_1 and I_2 the first and second term on the right hand side of the last equality. Then, as $n \to \infty$,

$$|I_1| \le \left[\frac{1}{nh_n}\sum_{i=1}^n \omega\left(\frac{Z_i - z}{h_n}\right)\right] \sup_{a\le z\le b} |\hat{H}_{g2}(s \mid z) - H_2(s \mid z)| \xrightarrow{a.s.} f(z)\cdot 0 = 0,$$

where the almost sure convergence follows from (4.15) and (4.11). Moreover, by (4.14), $I_2 \longrightarrow^{\text{a.s.}} H_2(s \mid z) f(z) = H_2(s,z)$, as $n \to \infty$. Thus, $E\hat{H}_{h2}^*(s,z) \longrightarrow^{\text{a.s.}} H_2(s,z)$.

Similarly, it can be shown that as $n \to \infty$

$$\text{Var}(\hat{H}_{h2}^*(s,z)) = \frac{1}{nh_n}\cdot\frac{1}{nh_n}\sum_{i=1}^n \omega^2\left(\frac{Z_i - z}{h_n}\right)\hat{H}_{g2}(s \mid Z_i) - \frac{1}{n}[E\hat{H}_{h2}^*(s,z)]^2 \xrightarrow{a.s.} 0.$$

Therefore, for any s, conditional on $(X_1, \delta_1, Z_1), \ldots, (X_n, \delta_n, Z_n)$, $\hat{H}_{h2}^*(s,z) \to^P H_2(s,z)$ along almost all sequences $(X_1, \delta_1, Z_1), \ldots$. One can also show that for any s, conditional on $(X_1, \delta_1, Z_1), \ldots, (X_n, \delta_n, Z_n)$, $\hat{H}_{h2}^*(s+,z) \to^P H_2(s+,z)$ along almost all sample sequences $\{(X_i, \delta_i, Z_i), i = 1, 2, \ldots\}$. Using the standard argument similar to those in the proof of Theorem 5.5.1 of Chung (1974) and the fact that $H_2(s,z)$ is left-continuous and nonincreasing, we conclude (4.23).

Parts (c)-(d) can be proved by applying (a) and Lemmas 2 and 3. Here we omit the details. □

PROOF OF THEOREM 1. It is not difficult to verify that the conditions of g_n in Lemma 1 and the conditions of h_n and g_n in Lemma 4 are satisfied if $nh_n^5 \to c > 0$ and (2.10) holds. So, all results in Lemmas 1 and 4 apply here. Throughout the proof all convergence are conditional on the observed sample data $(X_1, \delta_1, Z_1), \ldots, (X_n, \delta_n, Z_n)$.

Now let's prove (2.11). Note that

$$
\begin{aligned}
&(\hat{A}_h^*(t \mid z) - \hat{A}_g(t \mid z)) \\
&= \left(\int_0^t \sum_{i=1}^n c_i^*(s,z) dN_i^*(s) - \hat{A}_g(t \mid z) \right) \\
&= \int_0^t \sum_{i=1}^n c_i^*(s,z) dM_i^*(s) + \int_0^t \sum_{i=1}^n c_i^*(s,z) d(\hat{A}_g(s \mid Z_i^*) - \hat{A}_g(s \mid z)) \\
&= W_{1n}^*(t \mid z) + W_{2n}^*(t \mid z), \qquad (4.29)
\end{aligned}
$$

where $c_i^*(s,z)$ and $N_i^*(s)$ $(1 \le i \le n)$ are defined in Lemma 4, $M_i^*(t) = N_i^*(t) - \int_0^t Y_i^*(s) d\hat{A}_g(s \mid Z_i^*)$, $t \in [0, \tau]$, $i = 1, \ldots, n$, are orthogonal locally square integral martingales (cf. Gill (1980), pp. 26–31 and Fleming and Harrington (1991), Section 2.6), and $W_{1n}^*(t \mid z)$ and $W_{2n}^*(t \mid z)$ denote the first and second term on the right hand side of the second equality. Thus, to prove (2.9), it suffices to show that conditional on $(X_1, \delta_1, Z_1), \ldots, (X_n, \delta_n, Z_n)$,

$$\sqrt{nh_n} W_{1n}^*(\cdot \mid z) \xrightarrow{d} V(\cdot \mid z), \qquad (4.30)$$

$$\sup_{t \in [0,\tau]} |\sqrt{nh_n} W_{2n}^*(t \mid z) - \sqrt{c}\mu(t \mid z)| \xrightarrow{P} 0, \qquad (4.31)$$

along almost all sequences (X_1, δ_1, Z_1), $(X_2, \delta_2, Z_2), \ldots$, where $\mu(t \mid z)$ is defined in (2.6) and $V(\cdot \mid z)$ is a continuous Gaussian martingale with mean 0 and variance function $\sigma^2(t \mid z)$ given by (2.7).

We first prove (4.30). Writing $I_1(t) = nh_n \int_0^t \sum_{i=1}^n c_i^{*2}(s,z) \alpha(s \mid Z_i^*) ds$ and $I_2(t) = nh_n \int_0^t \sum_{i=1}^n c_i^{*2}(s,z) d(\hat{A}_g(s \mid Z_i^*) - A(s \mid Z_i^*))$, we have

$$\langle \sqrt{nh_n} W_{1n}^*, \sqrt{nh_n} W_{1n}^* \rangle(t) = nh_n \int_0^t \sum_{i=1}^n c_i^{*2}(s,z) d\hat{A}_g(s \mid Z_i^*) = I_1(t) + I_2(t).$$

From Assumption (B.3), $\alpha(s \mid w) = \alpha(s \mid z) + \eta_1(s,w)(w - z)$, where $\eta_1(s,w)$ is bounded. Thus, with probability 1,

$$
\begin{aligned}
I_1(t) = &\int_0^t \left[\alpha(s \mid z) nh_n \sum_{i=1}^n c_i^{*2}(s,z) + nh_n \sum_{i=1}^n c_i^{*2}(s,z) \eta_1(s, Z_i^*)(Z_i^* - z) \right] ds \\
\xrightarrow{P} &\left[\int_0^t \alpha(s \mid z) H_2^{-1}(s,z) ds \right] \left[\int \omega^2(u) du \right] = \sigma^2(t \mid z),
\end{aligned}
$$

uniformly in $t \in [0, \tau]$, where the convergence follows from (4.24) and the simple fact that

$$\left| nh_n \sum_{i=1}^n c_i^{*2}(s,z) \eta_1(s, Z_i^*)(Z_i^* - z) \right| \le Ch_n \cdot (nh_n) \sum_{i=1}^n c_i^{*2}(s,z),$$

for some constant $C > 0$. Moreover, integrating by parts, we have

$$\begin{aligned}
|I_2(t)| &= \Big| nh_n \sum_{i=1}^{n} c_i^{*2}(t,z)(\hat{A}_g(t \mid Z_i^*) - A(t \mid Z_i^*)) \\
&\quad - nh_n \sum_{i=1}^{n} \int_0^t (\hat{A}_g(s \mid Z_i^*) - A(s \mid Z_i^*)) dc_i^{*2}(s,z) \Big| \\
&\le \|\hat{A}_g - A\|_I^\tau \cdot nh_n \sum_{i=1}^{n} c_i^{*2}(t,z) + \|\hat{A}_g - A\|_I^\tau \cdot nh_n \sum_{i=1}^{n} \int_0^t |\, dc_i^{*2}(s,z)| \\
&= \|\hat{A}_g - A\|_I^\tau \left(2nh_n \sum_{i=1}^{n} c_i^{*2}(t,z) + (nh_n) \cdot O_p\left((nh_n)^{-1}\right) \right) \\
&\xrightarrow{P} 0 \qquad \text{uniformly in} \quad t \in [0,\tau],
\end{aligned}$$

with probability 1, where in the third step we have used (4.27) and the last step follows from (4.1) and (4.24). Hence, with probability 1,

$$(4.32)\quad \langle \sqrt{nh_n} W_{1n}^*, \sqrt{nh_n} W_{1n}^* \rangle(t) = I_1(t) + I_2(t) \xrightarrow{P} \sigma^2(t \mid z) \qquad \text{uniformly in} \quad t \in [0,\tau].$$

Similarly, one can use Lemma 1 and Lemma 4 to verify the Lindeberg condition that for any $\epsilon > 0$,

$$\begin{aligned}
&\int_0^\tau \sum_{i=1}^{n} (\sqrt{nh_n} c_i^*(s,z))^2 Y_i^*(s) I(\sqrt{nh_n} c_i(s,z_0) > \epsilon) d\hat{A}_g(s \mid Z_i^*) \\
&\qquad \le \epsilon^{-1} \int_0^\tau (nh_n)^{3/2} \sum_{i=1}^{n} c_i^{*3}(s,z) d\hat{A}_g(s \mid Z_i^*) \\
(4.33)\qquad &\qquad \xrightarrow{P} 0,
\end{aligned}$$

with probability 1. Therefore, (4.30) follows from (4.32), (4.33), and Rebolledo's martingale central limit theorem (cf. Andersen and Gill (1982), Theorem I.2).

Now, we verify (4.31). Write

$$\begin{aligned}
(4.34)\qquad \sqrt{nh_n} W_{2n}^*(t \mid z) &= \sqrt{nh_n} \int_0^t \sum_{i=1}^{*} c_i^*(s,z)(\alpha(s \mid Z_i^*) - \alpha(s|z)) ds \\
&\quad + \sqrt{nh_n} \int_0^t \sum_{i=1}^{*} c_i^*(s,z) \\
&\quad \cdot d([\hat{A}_g(s \mid Z_i^*) - A(s \mid Z_i^*)] - [\hat{A}_g(s \mid z) - A(s \mid z)]) \\
&= \sqrt{nh_n} \int_0^t \sum_{i=1}^{*} c_i^*(s,z)(\alpha(s \mid Z_i^*) - \alpha(s \mid z)) ds \\
&\quad + \sqrt{nh_n} \int_0^t \sum_{i=1}^{n} c_i^*(s,z) \\
&\quad \cdot d\left[\frac{\partial}{\partial z} (\hat{A}_g(s \mid \tilde{z}_i(s)) - A(s \mid \tilde{z}_i(s))) \right] (Z_i^* - z) \\
&\equiv J_1(t) + J_2(t),
\end{aligned}$$

where $\tilde{z}_i(s)$ is between Z_i^* and z, and $J_1(t)$ and $J_2(t)$ denote the first and second term on the right hand side of the second equation. Then,

$$
\begin{aligned}
J_1(t) &= \int_0^t \alpha_z'(s \mid z)\left[\sqrt{nh_n}\sum_{i=1}^n c_i^*(s,z)(Z_i^*-z)\right]ds \\
&\quad+\int_0^t \alpha_{zz}''(s,z)\left[\sqrt{nh_n}\sum_{i=1}^n c_i^*(s,z)(Z_i^*-z)^2\right]ds \\
&\quad+\int_0^t \left[\sqrt{nh_n}\sum_{i=1}^n c_i(s,z)\gamma(s,Z_i^*)(Z_i^*-z)^3\right]ds \\
&\xrightarrow{P} \sqrt{c}\mu(t \mid z) \qquad (4.35)
\end{aligned}
$$

uniformly in $t \in [0,\tau]$, where the first equality is from Assumption (B.3) and the last step follows from (4.25), (4.26), the assumption that $nh_n^5 \to c$, and the fact that

$$
\left|\int_0^t \left[\sqrt{nh_n}\sum_{i=1}^n c_i(s,z)\gamma(s,Z_i^*)(Z_i^*-z)^3\right]ds\right| \le K_1 h_n \int_0^t \sqrt{nh_n}\sum_{i=1}^n c_i(s,z)(Z_i^*-z)^2 ds.
$$

Moreover, integrating by parts, we have

$$
\begin{aligned}
|J_2(t)| &\le \left|\sqrt{nh_n}\sum_{i=1}^n c_i^*(t,z)(Z_i^*-z)\left(\frac{\partial \hat{A}_g(t \mid \tilde{z}_i(t))}{\partial z}-\frac{\partial A(t \mid \tilde{z}_i(t))}{\partial z}\right)\right| \\
&\quad+\sqrt{nh_n}\int_0^t \sum_{i=1}^n |(Z_i^*-z)|\cdot\left|\frac{\partial \hat{A}_g(s \mid \tilde{z}_i(s))}{\partial z}-\frac{\partial A(s \mid \tilde{z}_i(s))}{\partial z}\right|\cdot|dc_i^*(s,z)| \\
&\le \sqrt{nh_n^3}\left\|\frac{\partial \hat{A}_g}{\partial z}-\frac{\partial A}{\partial z}\right\|_I^\tau\left(\sum_{i=1}^n c_i^*(t,z)+\int_0^t |dc_i^*(s,z)|\right) \\
&= \sqrt{nh_n^3}\left\|\frac{\partial \hat{A}_g}{\partial z}-\frac{\partial A}{\partial z}\right\|_I^\tau\left(2\sum_{i=1}^n c_i^*(t,z)+O_p(1)\right) \\
&= O\left(\max\left\{\sqrt{\frac{h_n^3\log g_n^{-1}}{g_n^3}},\sqrt{nh_n^3 g_n^4}\right\}\right)(2+O_p(1)) \\
&= o(1), \qquad (4.36)
\end{aligned}
$$

where the third step uses (4.27), the fourth step follows from (4.2), and the last step follows from (2.10). Combining (4.34), (4.35), and (4.36), we prove (4.31). □

Remark 4.1. It is seen from the above proof that in order for the bootstrap to pick up the correct amount of bias (i.e. for (4.31) to hold), $\frac{\partial \hat{A}_g}{\partial z}$ has to be uniformly strong consistent at appropriate speed (see the last two steps in (4.36)). This, in turn, it requires that g_n satisfies condition (2.10). As mentioned in Remark 2.1, this means that g_n has to go to 0 at a slower rate than h_n.

Acknowledgements

The authors are grateful to a referee to providing constructive comments. The research of Gang Li was partially supported by a National Institute of Health grant and a National Science Foundation grant. The research of Somnath Datta was partially supported by a grant from the National Security Agency.

References

Akritas, M. (1986). Bootstrapping the Kaplan-Meier estimator, *J. Amer. Statist. Assoc.*, **81**, 1032–1039.

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

Beran, R. (1981). Nonparametric regression with randomly censored survival data, Tech. Report, Department of Statistics, University of California, Berkeley.

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model, *J. Amer. Statist. Assoc.*, **89**, 1290–1302.

Cao, R. and González-Manteiga, W. (1993). Bootstrap methods in regression smoothing, *J. Nonparametr. Statist.*, **2**, 379–388.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed., Academic Press, New York.

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

Dabrowska, D. M. (1989). Uniform consistency of the Kernel conditional Kaplan-Meier estimate, *Ann. Statist.*, **7**, 1–26.

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

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Ann. Statist.*, **7**, 1–26.

Efron, B. (1981). Censored data and the bootstrap, *J. Amer. Statist. Assoc.*, **76**, 312–319.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Monogr. Statist. Appl. Probab., No. 57, Chapman & Hall, London.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*, Wiley, New York.

Freedman, D. A. (1981). Bootstrapping regression models, *Ann. Statist.*, **6**, 1218–1228.

Gill, R. D. (1980). Censoring and stochastic integrals, *Mathematical Centre Tracts*, **124**, Mathematical Centre, Amsterdam.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data, *Biometrika*, **67**, 133–143.

Härdle, W. and Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands, *J. Amer. Statist. Assoc.*, **83**, 102–110.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *Ann. Statist.*, **19**, 778–796.

Hjort, N. L. (1985). Bootstrapping Cox's regression model, Tech. Report, No. 241, Department of Statistics, Stanford University, California.

Li, G. (1997). Optimal rate local smoothing in a multiplicative intensity counting process model, *Journal of Mathematical Methods of Statistics*, **6**, 224–244.

Li, G. and Datta, S. (1999). A bootstrap approach to nonparametric regression for right censored data, Tech. Report, Sta 99-8, Department of Statistics, University of Georgia, Athens.

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

Lo, S. H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations, *Probab. Theory Related Fields*, **71**, 455–465.

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

Nadaraya, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves, *Theory Probab. Appl.*, **15**, 134–137.

Rao, B. L. S. Prakasa (1983). *Nonparametric Functional Estimation*, Academic Press, New York.

Van Keilegom, I. and Veraverbeke, N. (1997). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process, *Comm. Statist. Theory Methods*, **26**, 853–869.