The Annals of Statistics
1994, Vol. 22, No. 3, 1299–1327

NEAREST NEIGHBOR ESTIMATION OF A BIVARIATE DISTRIBUTION UNDER RANDOM CENSORING[1]

BY MICHAEL G. AKRITAS

Pennsylvania State University

We consider the problem of estimating the bivariate distribution of the random vector (X, Y) when Y may be subject to random censoring. The censoring variable C is allowed to depend on X but it is assumed that Y and C are conditionally independent given $X = x$. The estimate of the bivariate distribution is obtained by averaging estimates of the conditional distribution of Y given $X = x$ over a range of values of x. The weak convergence of the centered estimator multiplied by $n^{1/2}$ is obtained, and a closed-form expression for the covariance function of the limiting process is given. It is shown that the proposed estimator is optimal in the Beran sense. This is similar to an optimality property the product-limit estimator enjoys. Using the proposed estimator of the bivariate distribution, an extension of the least squares estimator to censored data polynomial regression is obtained and its asymptotic normality established.

1. Introduction. Estimation of the bivariate distribution under random censoring has received considerable attention over the past 10 years [see Dabrowska (1988, 1989a), Van der Laan (1992) and references therein]. Here we consider the problem of estimating the bivariate distribution of (X, Y) when Y may be subject to random censoring but X is always uncensored. We are motivated by applications to the linear regression model where the covariate (X) is uncensored and the response variable (Y) is subject to random censoring (see Section 5 for details). In this context it is unrealistic to assume that the censoring variable C is independent from X. Therefore we work with the assumption that Y is conditionally independent of C given X, but we allow C to depend on X. This feature makes the present estimation problem different from those considered thus far, where the assumption that the censoring mechanism is independent from (X, Y) was imposed. Some identifiability issues for the bivariate distribution under conditional independence are discussed in the recent paper by Pruitt (1993).

The proposed estimator of the bivariate distribution function is an average, over covariate values, of estimates of the conditional distribution function of the response given the covariate. Although the rate of convergence of estimates of the conditional distribution function is slow, the square root n rate of convergence is recovered by the process of averaging over the covariate values. Estimation of the conditional distribution function under random censoring

---

Received July 1992; revised November 1993.

[1]Partially supported by NSF Grants DMS-90-07717 and DMS-92-08066.

AMS 1991 *subject classifications.* 62G05, 62G30, 62H12, 62J05.

Key words and phrases. Conditional empirical processes, conditional Kaplan–Meier estimator, weak convergence, Beran optimality, polynomial regression.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.*
www.jstor.org

was introduced in Beran (1981) and was further studied by Dabrowska (1987, 1989b). See also McKeague and Utikal (1990), who allow time-dependent covariates and recurrent failures, and Cheng (1989), who considered the problem of estimating the marginal distribution function of Y. The estimates of the conditional distribution functions we use are smooth nearest neighbor estimators, and thus we refer to the proposed estimator of the bivariate distribution function as the nearest neighbor estimator (NNE). In the uncensored case, nearest neighbor estimates of the conditional distribution function were studied by Stute (1984, 1986) and Horvath and Yandell (1988). To our knowledge, however, averaging of such conditional distribution functions over values of the conditioning variable was never proposed as a method for estimating the bivariate distribution function. It is worth noting that, in the uncensored case, the NNE is asymptotically equivalent to the empirical distribution function.

Clearly there is a class of such estimators, specified by the choice of the kernel function. Arguing formally it is seen that, provided the kernel function satisfies some simple conditions, all members of this class are asymptotically equivalent. Thus the covariance function of the estimator does not depend on the chosen kernel. In this paper we will work with the kernel $K(u) = 0.5I(-1 < u < 1)$. The choice of the indicator kernel function simplifies the arguments and requires fewer smoothness conditions. The closed-form expression of the variance function is slightly more complicated than the variance function of the usual Kaplan–Meier estimator on the line. A particularly important result is that the covariance function of the NNE is optimal in the Beran sense [Beran (1977)]. This optimality result is the direct analogue of a parametric optimality result established independently by Hájek (1970) and Inagaki (1970). It implies that any other regular estimator will be at least as dispersed as the NNE. That the product-limit estimator is optimal in this sense was shown by Wellner (1982).

The next section introduces the estimator of the bivariate distribution and states the assumptions. The weak convergence of the estimator of the bivariate distribution is given in Section 3. Optimality of the NNE is established in Section 4. Section 5 introduces an extension of the least squares estimator to censored data regression using the present estimator of the bivariate distribution. Finally, Section 6 discusses further research.

Throughout the paper, K in proofs stands for a generic constant whose value may differ from line to line.

2. Assumptions and definitions. In this section we state the assumptions under which the various results will be shown and present the NNE for the bivariate distribution function.

Consider a sequence of independent and identically distributed (iid) random vectors (Y_i, C_i, X_i), $i = 1, \dots, n$, such that given X_i, Y_i and C_i are independent. The observed data are of the form

$$D_i = (Z_i, \delta_i, X_i), \qquad i = 1, \dots, n, \tag{2.1}$$

where $Z_i = Y_i \wedge C_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i = Z_i)$. In what follows we let

(X, Y, Z, δ, C) be a generic random vector having the joint distribution of each $(X_i, Y_i, Z_i, \delta_i, C_i)$ and set $F(x,y) = P(X \le x,\ Y \le y)$, $S(x,y) = P(X > x,\ Y > y)$, $G(x) = F(x, \infty)$, $S(y \mid x) = P(Y > y \mid x) = 1 - F(y \mid x)$; and $F_D(D)$ will denote the joint distribution function of the data $D = (Z, \delta, X)$. The proposed estimator for the bivariate distribution of (X, Y) is based on the relation

$$S(x,y) = \int_x^\infty P(Y > y \mid X = t)\, dG(t).$$

In particular we propose the estimator

$$\widehat{S}(x,y) = n^{-1} \sum_{i=1}^{n} \widehat{S}(y \mid X_i) I(X_i > x), \tag{2.2}$$

where $\widehat{S}(y \mid x)$ is an estimate of $S(y \mid x)$. We will consider estimates $\widehat{S}(y \mid x)$ that are based on nearest neighbor estimates of the conditional subsurvival functions. Estimates of the conditional subsurvival functions are used to estimate the conditional cumulative hazard function $\Lambda(t \mid x)$, which leads to an estimate of $S(y \mid x)$ through the relation

$$S(y \mid x) = \exp\left\{-\Lambda^c(y \mid x)\right\} \prod_{s \le y} \left\{1 - \Delta\Lambda(s \mid x)\right\},$$

where $\Lambda^c(y \mid x)$ is the continuous component of $\Lambda(y \mid x)$; the product is taken over the set of discontinuities of $H_1(z \mid x)$ defined below; and $\Delta\Lambda(s \mid x) = \Lambda\ (s \mid x) - \Lambda(s-\ \mid x)$.

Note that in the univariate case (so $Z = Y \wedge C$ and there is no X) it is not possible to estimate the distribution function of Y beyond the support of the distribution of Z. It follows that in the bivariate case we can only hope to estimate $S(x,y)$ at the (x,y) values for which y is a number less than the upper bound of the support of the conditional distribution of Z given $X = x_1$, for all $x_1 > x$.

Denote the subsurvival functions by $H_\iota(z \mid x) = P(z > z,\ \delta = \iota \mid x)$, $\iota = 0, 1$, and $H(z \mid x) = H_0(z \mid x) + H_1(z \mid x)$. Thus the conditional cumulative hazard function associated with $S(y \mid x)$ is

$$\Lambda(t \mid x) = \int_0^t S(y-\ \mid x)^{-1} dF(y \mid x) = -\int_0^t H(z-\ \mid x)^{-1} dH_1(z \mid x).$$

For $\iota = 0, 1$, we consider smooth nearest neighbor estimates of the form

$$\begin{aligned} \widehat{H}(z \mid x) &= (na_n)^{-1} \sum_{j=1}^{n} I(Z_j > z,\ \delta_j = \iota) K\left(\frac{\widehat{G}(x) - \widehat{G}(X_j)}{a_n}\right), \qquad \iota = 0, 1, \\ \widehat{H}(z \mid x) &= \widehat{H}_0(z \mid x) + \widehat{H}_1(z \mid x), \end{aligned} \tag{2.3}$$

where a_n is a deterministic sequence of real numbers converging to zero, $K(u) = 0.5\, I(-1 < u < 1)$ and $\widehat{G}$ is the empirical distribution function for X_i, $i = 1, \ldots, n$.

The estimates of the conditional cumulative hazard function and the survival function of Y given $X = x$ are given by

$$\widehat{\Lambda}(t\,|\,x) = -\int_0^t \widehat{H}(z-\,|\,x)^{-1}\,d\widehat{H}_1(z\,|\,x),$$

and

$$\widehat{S}(y\,|\,x) = \prod_{Z_i \le y,\, \delta_i = 1} \left\{1 - \frac{K\Big(\big(\widehat{G}(x) - \widehat{G}(X_i)\big)/a_n\Big)}{\sum_{j=1}^n I(Z_j > Z_i-)K\Big(\big(\widehat{G}(x) - \widehat{G}(X_j)\big)/a_n\Big)}\right\}. \tag{2.4}$$

The analysis of the estimator in (2.2) will be based on the following decomposition:

$$\begin{aligned}\widehat{S}(x,y) - S(x,y) &= n^{-1}\sum_{i=1}^n \widehat{S}(y\,|\,X_i)I(X_i > x) - S(x,y)\\ &= n^{-1}\sum_{i=1}^n \Big[\widehat{S}(y\,|\,X_i) - S(y\,|\,X_i)\Big]I(X_i > x)\\ &\quad + \int S(y\,|\,t)I(t > x)\big[\widehat{G}(dt) - G(dt)\big].\end{aligned} \tag{2.5}$$

Note that the above decomposition does not display the bias term explicitly. The bias term is contained in the first term on the right-hand side of (2.5).

The assumptions are stated separately for each of the results of the following section but are collected here for convenient reference.

A1. The sequence a_n satisfies $na_n^3[\log a_n^{-1}]^{3.5} \to \infty$ and $na_n^4 \to 0$.

It should be noted that the condition $na_n^4 \to 0$ is required to make the bias term asymptotically negligible. For all other derivations $na_n^5[\log a_n^{-1}]^{-1} \to 0$ suffices.

A2. (i) The bivariate distribution function F is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$.

(ii) The joint densities $f_G(z,u)$ and $f_G(z,\delta,u)$ of $(Z, G(X))$ and $(Z,\delta,G(X))$ are twice continuously differentiable with respect to u, and

$$\int \sup\big\{|f_G''(z,u)|;\ 0 < u < 1\big\}\,dz < \infty,$$

$$\int \sup\big\{|f_G''(z,1,u)|;\ 0 < u < 1\big\}\,dz < \infty.$$

Now define the function

$$\xi_x(z,\delta,y) = -S(y\,|\,x)\left[-\int_0^y \frac{I(z > s)}{H(s\,|\,x)^2}H_1(ds\,|\,x) + \frac{I(z \le y,\ \delta = 1)}{H(z\,|\,x)}\right]. \tag{2.6}$$

The following smoothness condition is needed for showing that the bias term is asymptotically negligible.

A3. The function $\xi_{G^{-1}(u)}(z,\delta,y)$ is twice continuously differentiable with respect to u. Let ξ'_x and ξ''_x denote, respectively, the first and second derivatives of $\xi_{G^{-1}(u)}$ evaluated at $u = G(x)$. Then:

(i) $$\sup\{|\xi'_x(z,\delta,y)|;\ \text{all } (x,y) \in \Omega, z < T_x, \text{ and } \delta = 0,1\} < \infty,$$

where $\Omega = \{(x,y): x \in \mathbb{R},\ 0 < y < T_{x_1} \text{ for all } x_1 > x\}$ and T_x is any number less than the upper bound of the support of $H(\cdot\,|\,x)$.

(ii) We say that $x \in N(x_1,x_2;\varepsilon)$ if $|G(x_1)-G(x_2)| \le \varepsilon$ and $G(x)$ is between $G(x_1)$ and $G(x_2)$. There is an $\varepsilon > 0$ such that

$$\int \sup\{|\xi''_x(z_2,\delta_2,y_0)|;\ x \in N(x_1,x_2;\varepsilon) \text{ and } y_0 < T_{\tilde{x}} \text{ for all } \tilde{x} \ge x_2\} \times G(dx_1)F_D(dD_2) < \infty.$$

The last assumption is only used in the proof of Lemma 3.4, to establish a bound on certain covering numbers.

A4. For each $\varepsilon > 0$ there exists a partition of $[0,T]$ into at most $A\varepsilon^{-V}$, where A and V are constants with $V \ge 3$, intervals $[a_j, a_{j+1}]$ such that $\sup\{H_1(a_{j+1}\,|\,x) - H_2(a_j\,|\,x);\ \text{all } x \text{ and } j\} \le \varepsilon$.

REMARK 2.1. Note that the function $\xi_x(z,\delta,y)$ defined in (2.6) is a conditional version of the function used in Lo and Singh [(1986), Theorem 1]. It follows that the conditional expectation of $\xi_x(z,\delta,y)$ given x is zero for all y, and the conditional covariance of $\xi_x(z,\delta,y_1)$ and $\xi_x(z,\delta,y_2)$ given x is $S(y_1\,|\,x)S(y_2\,|\,x)C_1(y_1 \wedge y_2\,|\,x)$, where

$$C_1(t\,|\,x) = -\int_0^t H(s\,|\,x)^{-2} H_1(ds\,|\,x). \tag{2.7}$$

REMARK 2.2. The assumption that the supremum of $|\xi'_x|$ over all x is finite may not be met in certain applications. In such cases, the results of this paper are true on a restricted domain of x for which this assumption is met.

3. Weak convergence. In this section we will prove the weak convergence of the estimator in (2.2) through the decomposition in (2.5).

Let Ω be as defined in assumption A3(i).

THEOREM 3.1. *Let assumptions* A1–A4 *hold. For fixed* (x_0,y_0) *we have*

$$\begin{aligned} &n^{1/2}[\widehat{S}(x_0,y_0) - S(x_0,y_0)] \\ &\quad = \int S(y_0\,|\,x)I(x > x_0)\alpha_n(dx) \\ &\qquad + n^{1/2}\int \xi_x(z,\delta,y_0)I(x > x_0)[\widehat{F}_D(dD) - F_D(dD)] + R_n(x_0,y_0), \end{aligned} \tag{3.1}$$

where $\alpha_n(x) = n^{1/2}[\widehat{G}(x) - G(x)]$ *is the empirical process that corresponds to the covariate;* $n^{1/2}[\widehat{F}_D(D) - F_D(D)]$ *is the empirical process that corresponds to the data* $D_i = (Z_i, \delta_i, X_i)$, $i = 1, \dots, n$; *and* $\sup\{|R_n(x_0, y_0)|; (x_0, y_0) \in \Omega\} = o(1)$, *in probability.*

The proof of Theorem 3.1 follows from Lemmas 3.1–3.4 and relation (3.4) given below.

COROLLARY 3.1. *The process* $n^{1/2}[\widehat{S}(x,y) - S(x,y)]$ *converges weakly to the limiting Gaussian process* $Z(x,y)$ *with covariance function*

$$\begin{aligned} &\operatorname{Cov}\big[Z(x_1, y_1), Z(x_2, y_2)\big] \\ (3.2)\qquad &= \operatorname{Cov}\Big[S(y_1 \mid X_i) I(X_i > x_1), S(y_2 \mid X_i) I(X_i > x_2)\Big] \\ &\quad + E\Big[S(y_1 \mid X_i) S(y_2 \mid X_i) C_1(y_1 \wedge y_2 \mid X_i) I(X_i > x_1 \vee x_2)\Big], \end{aligned}$$

where $x_1 \vee x_2 = \max\{x_1, x_2\}$.

It is seen that the covariance function is slightly more complicated to compute than the covariance function of the Kaplan–Meier estimator. Alternatively, it is possible to use the bootstrap method [see Akritas (1992)].

PROOF OF COROLLARY 3.1. Using Remark 2.1 and a conditioning argument, it follows that the first two terms of the right-hand side of (3.1) are uncorrelated. It follows that the covariance of the limiting process, if weak convergence can be established, is given by (3.2). Because the two components of the process $n^{1/2}[\widehat{S}(x,y) - S(x,y)]$ are asymptotically independent it suffices to prove weak convergence for each component separately. We will use the functional central limit theorem given in Pollard [(1990), page 53]. The main concepts in this theorem are the pseudodimension of certain subsets in $\mathbb{R}^n$ and manageability with respect to the envelope function of these subjects. We will also make use of the results of Pollard [(1990), Section 5] and in particular his Lemma 5.3. Consider the first component of $n^{1/2}[\widehat{S}(x,y) - S(x,y)]$; thus, in Pollard's notation

$$f_{ni}(\omega, t) = S(t_2 \mid X_i) I(X_i > t_1)/n^{1/2}, \qquad t = (t_1, t_2).$$

To establish manageability, it suffices to consider separately the two contributions to f_{ni}. Consider the subset of $\mathbb{R}^n$

$$\mathcal{F}_n = \Big\{ \Big(S(t \mid X_1), \dots, S(t \mid X_n)\Big); t \in \mathbb{R} \Big\}.$$

Clearly it has a bounded envelope function. Next we will shown that $\mathcal{F}_n$ has pseudodimension 1. Indeed, the monotonicity of $S(t \mid x)$ as a function of t implies that $(S(t \mid X_i), S(t \mid X_j))$ cannot surround a point in $\mathbb{R}^2$ for any (i, j). The same arguments apply for the set defined by the indicator functions $\{I(X_i > t)\}$.

Manageability of the second component of $n^{1/2}[\widehat{S}(x,y) - S(x,y)]$ is established by exactly the same arguments. The remaining conditions in the functional central limit theorem are easily seen to be satisfied from the covariance calculation in (3.2) and the fact that the envelope functions are constant ($F_{ni} = K/\sqrt{n}$, for some $K > 0$). □

We now proceed with the proof of Theorem 3.1.

LEMMA 3.1. *Assume that* $na_n^3[\log a_n^{-1}]^{3.5} \to \infty$ *and* $na_n^5[\log a_n^{-1}]^{-1} \to 0$, *and let the assumptions imposed on* $f_G(z,u)$ *and* $f_G(z,1,u)$ *in Proposition* A.1 *and Remark* A.1 *hold. Let* $\xi_x(z,\delta,y)$ *be as defined in* (2.6). *Then*

$$\begin{aligned}\widehat{S}(y \mid X_i) &- S(y \mid X_i)\\ &= (na_n)^{-1}\sum_{j=1}^{n}\xi_{X_i}(Z_j,\delta_j,y)0.5I\big(|\widehat{G}(X_i) - \widehat{G}(X_j)| \le a_n\big) + r_n(y,X_i),\end{aligned}$$

where $\sup\{|r_n(y,X_i)|;\ 0 < y < T_{X_i},\ i = 1,\dots,n\} = o(n^{-1/2})$, *as* $n \to \infty$, *almost surely.*

The proof is given in Appendix B.

Relation (2.5) and Lemma 3.1 imply that

$$\begin{aligned}\widehat{S}(x,y) &- S(x,y)\\ &= \int S(y \mid t)I(t > x)\big[\widehat{G}(dt) - G(dt)\big]\\ &\quad + n^{-1}\sum_{i=1}^{n}\left[(na_n)^{-1}\sum_{j=1}^{n}\xi_{X_i}(Z_j,\delta_j,y)0.5I\big(|\widehat{G}(X_i) - \widehat{G}(X_j)| \le a_n\big)\right]\\ &\quad \times I(X_i > x) + o\big(n^{-1/2}\big)\end{aligned} \tag{3.3}$$

uniformly in $(x,y) \in \Omega$, almost surely. Let $T_n(x,y)$ denote the second term on the right-hand side of (3.3). Fix (x_0,y_0) and write $T_n = T_n(x_0,y_0)$ as

$$T_n = T_{1n} + T_{2n} + T_{3n}, \tag{3.4}$$

where

$$\begin{aligned}2T_{1n} &= a_n^{-1}\int\big[\xi_{x_1}(z_2,\delta_2,y_0) - \xi_{x_2}(z_2,\delta_2,y_0)\big]I(x_1 > x_0)\\ &\qquad \times I\big(|G(x_1) - G(x_2)| \le a_n\big)\widehat{G}(dx_1)\widehat{F}_D(dD_2)\\ 2T_{2n} &= a_n^{-1}\int \xi_{x_2}(z_2,\delta_2,y_0)I(x_1 > x_0)I\big(|G(x_1) - G(x_2)| \le a_n\big)\\ &\qquad \times \widehat{G}(dx_1)\big[\widehat{F}_D(dD_2) - F_D(dD_2)\big]\end{aligned}$$

$$2T_{3n} = a_n^{-1} \int \xi_{x_1}(z_2, \delta_2, y_0) I(x_1 > x_0) \times \Big[I\big(|\widehat{G}(x_1) - \widehat{G}(x_2)| \le a_n\big) - I\big(|G(x_1) - G(x_2)| \le a_n\big) \Big] \times \widehat{G}(dx_1)\widehat{F}_D(dD_2).$$

LEMMA 3.2. *Assume $na_n^4 \to 0$ and let assumption* A3 *hold true. Then*

$$\sup\Big\{ |n^{1/2}T_{1n}|;\ (x_0, y_0) \in \Omega \Big\} = o(1) \quad \textit{almost surely.}$$

The proof is given in Appendix B.

LEMMA 3.3. *Under the assumption $na_n^2 \to \infty$,*

$$\sup\Big\{ |n^{1/2}T_{3n}|;\ (x_0, y_0) \in \Omega \Big\} = o(1) \quad \textit{almost surely.}$$

The proof is given in Appendix B.

LEMMA 3.4. *Let assumptions* A1, A2(i) *and* A4 *hold. Let*

$$h_0(x_1, z_2, \delta_2) = \xi_{x_1}(z_2, \delta_2, y_0) I(x_1 > x_0).$$

Then

$$\sup\left\{ n^{1/2} \left| T_{2n} - \int h_0(x, z, \delta)\big[\widehat{F}_D(dD) - F_D(dD)\big] \right|;\ (x_0, y_0) \in \Omega \right\} = o(1)$$

in probability.

The proof is given in Appendix B.

4. Asymptotic optimality. In this section we will show that the limiting distribution of any regular estimator (as defined prior to Theorem 4.1) of the bivariate distribution function that is a function of the data (Z_i, δ_i, X_i), $i = 1, \dots, n$, can be represented as a convolution of the distribution that corresponds to the limiting process of the proposed NNE with another distribution on $C(\Omega)$. This will imply that any other regular estimator is at least as dispersed as the NNE. See Beran (1977) for a nice discussion of this point. For notational simplicity, in this section we will assume that $T_x = T$, constant for all x. With this assumption $\Omega = \mathbb{R} \times [0, T]$. See also Remark 4.1 at the end of this section.

Such representation of regular estimators was first established in the parametric case independently by Hájek (1970) and Inagaki (1970). For an excellent textbook presentation of this theorem see Roussas (1972). Beran (1977) established a nonparametric version of this theorem and showed that the empirical distribution function is optimal. Wellner (1982) used the same techniques to show that, when the data are subject to random censoring, the product limit estimator is also optimal in this sense.

Let ν_1 be a measure on the real line $\mathbb{R}$ with respect to which the conditional distributions $F(\cdot\,|\,x)$ and $C(\cdot\,|\,x)$ have densities $f(\cdot\,|\,x)$ and $c(\cdot\,|\,x)$. The measure ν_1 can depend on x, but this will not be made explicit in the notation. Let ν_2 be a measure on $\mathbb{R}$ with respect to which G has a density g. It then follows that the observed random vector $D = (Z, \delta, X)$ has density

$$f_D(z,\delta,x) = f_D(z,\delta\,|\,x)g(x) \tag{4.1}$$

with respect to the measure $\mu = \nu_1 \times \nu_2 \times \{\text{counting}\}$ on $S_1 = \mathbb{R}^2 \times \{0,1\}$, where

$$f_D(z,\delta\,|\,x) = \Big\{C(z\,|\,x)f(z\,|\,x)\Big\}^{\delta}\Big\{S(z\,|\,x)c(z\,|\,x)\Big\}^{1-\delta} \tag{4.2}$$

is the conditional, given x, density of (Z,δ) with respect to the measure $\mu_1 = \nu_1 \times \{\text{counting}\}$ on $S_2 = \mathbb{R} \times \{0,1\}$.

Note that f without the subscript D denotes the density of F, the bivariate distribution that is to be estimated. It is possible to recover F or, equivalently, S from f_D through the relation

$$\begin{aligned} S(y\,|\,x) &= \exp\left(-\int_0^y \frac{1}{H(s\,|\,x)}\right) f_D(z,1\,|\,x)\,d\nu_1(z), \\ H(s\,|\,x) &= \sum_{\delta=0}^{1}\int_s^{\infty} f_D(z,\delta\,|\,x)\,d\nu_1(z), \\ S(x,y) &= \int_x^{\infty} S(y\,|\,t)g(t)\,d\nu_2(t). \end{aligned} \tag{4.3}$$

Let $\|\cdot\|_{\mu}, \|\cdot\|_{\mu_1}, \|\cdot\|_{\nu_2}, \langle\cdot,\cdot\rangle_{\mu}, \langle\cdot,\cdot\rangle_{\mu_1}, \langle\cdot,\cdot\rangle_{\nu_2}$, denote the usual norm and inner products on $L^2(S_1,\mu), L^2(S_2,\mu_1)$ and $L^2(\mathbb{R},\nu_2)$, respectively. Let $\mathcal{F}(\mu)$ denote the set of all densities on S_1 with respect to μ; let $\mathcal{F}_1(\mu_1)$ denote the set of all densities on S_2 with respect to μ_1; and let $\mathcal{F}_2(\nu_2)$ denote the set of all the densities on $\mathbb{R}$ with respect to ν_2. Finally, let $\mathcal{C}(f_D,\beta)$, $\mathcal{C}_1(f_D(\cdot,\cdot\,|\,x),\beta_1(\cdot,\cdot\,|\,x))$ and $\mathcal{C}_2(g,\beta_2)$ denote the set of all sequences of densities $\{f_{Dm} \in \mathcal{F}(\mu)\}$, $\{f_{Dm}(\cdot,\cdot\,|\,x) \in \mathcal{F}_1(\mu_1)\}$ and $\{g_m \in \mathcal{F}_2(\nu_2)\}$, respectively, such that

$$\begin{aligned} \lim_{m\to\infty} \left\|m^{1/2}\left(f_{Dm}^{1/2} - f_D^{1/2}\right) - \beta\right\|_{\mu} &= 0, \\ \lim_{m\to\infty} \left\|m^{1/2}\left(f_{Dm}^{1/2}(\cdot,\cdot\,|\,x) - f_D^{1/2}(\cdot,\cdot\,|\,x)\right) - \beta_1(\cdot,\cdot\,|\,x)\right\|_{\mu_1} &= 0, \\ \lim_{m\to\infty} \left\|m^{1/2}\left(g_m^{1/2} - g^{1/2}\right) - \beta_2\right\|_{\nu_2} &= 0. \end{aligned} \tag{4.4}$$

It is easy to see that relations (4.4) imply

$$\left\langle\beta, f_D^{1/2}\right\rangle_{\mu} = 0, \qquad \left\langle\beta_1(\cdot,\cdot\,|\,x), f_D^{1/2}(\cdot,\cdot\,|\,x)\right\rangle_{\mu_1} = 0 \quad \text{and} \quad \left\langle\beta_2, g^{1/2}\right\rangle_{\nu_2} = 0.$$

This is most easily seen by writing

$$\int \left[m^{1/2}(f_{Dm}^{1/2} - f_D^{1/2}) - \beta\right]^2 d\mu$$
$$= -2m \int \left[f_{Dm}^{1/2} - f_D^{1/2} - m^{1/2}\beta\right] f_d^{1/2}\, d\mu - 2m^{1/2} \int (f_{Dm}^{1/2} - f_D^{1/2})\beta\, d\mu$$
$$+ \int \beta^2\, d\mu - 2m^{1/2} \int \beta f_D^{1/2}\, d\mu,$$

and similarly for the other implications.

LEMMA 4.1. *Let*

$$\{g_m\} \in \mathcal{C}_2(g, \beta_2) \quad \textit{and} \quad \left\{f_{Dm}(\cdot,\cdot\,|\,x)\right\} \in \mathcal{C}_1\Big(f_D(\cdot,\cdot\,|\,x), \beta_1(\cdot,\cdot\,|\,x)\Big),$$

for almost surely $[\nu_2]$ *all* x, *and let*

$$\begin{aligned} f_D(z,\delta,x) &= f_D(z,\delta\,|\,x)g(x), \\ f_{Dm}(z,\delta,x) &= f_{Dm}(z,\delta\,|\,x)g_m(x), \\ \beta(z,\delta,x) &= \beta_1(z,\delta\,|\,x)g^{1/2}(x) + \beta_2(x)f_D^{1/2}(z,\delta\,|\,x). \end{aligned} \tag{4.5}$$

Then $\{f_{Dm}\} \in \mathcal{C}(f_D, \beta)$. *Conversely, each sequence* $\{f_{Dm}\} \in \mathcal{C}(f_D, \beta)$ *specifies sequences* $\{g_m\} \in \mathcal{C}(g, \beta_2)$ *and* $\{f_{Dm}(\cdot,\cdot\,|\,x)\} \in \mathcal{C}_1(f_D(\cdot,\cdot\,|\,x), \beta(\cdot,\cdot\,|\,x))$ *such that* $f_D(z,\delta,x)$, $f_D(z,\delta\,|\,x)$ *and* $g(x)$ *are related through the first relation in* (4.5), *and* $\beta_2(x) = \int \beta(z,\delta,x)f_D^{1/2}(z,\delta\,|\,x)\,d\mu_1(z,\delta)$ *and* $\beta_1(z,\delta\,|\,x)$ *is given in terms of* β *and* β_2 *through the third relation in* (4.5). *In addition,*

$$\|\beta\|_\mu^2 = \int \left\|\beta_1(\cdot,\cdot\,|\,x)\right\|_{\mu_1}^2 g(x)\,d\nu_2(x) + \|\beta_2\|^2. \tag{4.6}$$

The proof of Lemma 4.1 follows by straightforward L^2 calculations.

Let now $\mathcal{C}(f_D)$ denote the union of all sets $\{\mathcal{C}(f_\nu, \beta)\colon \beta \in L^2(S_1, \mu),\ \beta \perp f_D^{1/2}\}$. Let $\{f_{Dm}\} \in \mathcal{C}(f_D)$, and let $\{S_m(x,y)\}$ be the sequence of bivariate survival functions obtained from (4.3) with f_{Dm} substituted for f_D. Consider the corresponding sequence of experiments where in the nth experiment we observe n independent random vectors $(Z_{ni}, \delta_{ni}, X_{ni})$, $i = 1, \dots, n$, with joint density $\Pi_{i=1}^n f_{Dn}(Z_{ni}, \delta_{ni}, X_{ni})$ on S_1^n. Let $\{\widetilde{S}_n\}$ be any sequence of $C(\Omega)$-valued estimators of S, where $\widetilde{S}_n$ is a function of $(Z_{ni}, \delta_{ni}, X_{ni})$, $i = 1, \dots, n$. We say that an estimating sequence, or estimator $\{\widetilde{S}_n\}$, is regular at f_D if the distributions $\mathcal{L}\{n^{1/2}(\widetilde{S}_n - S_n)\}$ on $C(\Omega)$ converge weakly to the same distribution $\mathcal{D} = \mathcal{D}_{f_D}$ depending only on f_D for all sequences $\{f_{Dm}\} \in \mathcal{C}(f_D)$. Of course $\mathcal{D}$ may also depend on the estimator $\{\widetilde{S}_n\}$. Exclusion of estimators that are not regular avoids the complicating issue of the existence of estimators that are superefficient at a specified distribution [see Beran (1977) for a discussion of this issue].

Let $\mathcal{D}_Z$ denote the law on $C(\Omega)$ corresponding to the process Z defined in Corollary 3.1.

THEOREM 4.1. *For any regular estimator $\tilde{S}_n$ of S that is based on (Z_i, δ_i, X_i), $i = 1, \dots, n$, the limiting law $\mathcal{D}$ on $C(\Omega)$ may be represented as $\mathcal{D}_Z * \mathcal{D}_W$, where $\mathcal{D}_W$ is the distribution of some process W that is independent from Z. Alternatively, if $\tilde{Z}$ denotes the limiting process of $n^{1/2}[\tilde{S}_n - S]$ under f_D, $\tilde{Z} = Z + W$.*

The proof of the theorem proceeds by a number of lemmas. Let

$$L_n = 2\log \prod_{i=1}^{n} \left\{ \frac{f_{Dn}(Z_i, \delta_i, X_i)}{f_D(Z_i, \delta_i, X_i)} \right\}, \tag{4.7}$$

for $\{f_{Dn}\} \in \mathcal{C}(f_D, \beta)$, $\beta \in L^2(S_1, \mu)$. The first lemma is a straightforward analogue of corresponding lemmas in Beran (1977) and Wellner (1982) and can be deduced easily from LeCam's second lemma.

LEMMA 4.2. *Let L_n be given by* (4.7). *For every $\varepsilon > 0$,*

$$\lim_{n\to\infty} P_{f_D}\left\{ \left| L_n - 2n^{-1/2}\sum_{i=1}^{n} \beta(Z_i, \delta_i, X_i) f_D^{-1/2}(Z_i, \delta_i, X_i) + 2\|\beta\|_\mu^2 \right| > \varepsilon \right\} = 0.$$

For each $(x, y) \in \mathbb{R}^2$ define the function

$$\gamma_y(z, \delta \mid x) = I(z < y)\left[C_1(y \mid x) - C_1(z \mid x) + \frac{I(\delta = 1)}{H(z \mid x)} \right] f_D^{1/2}(z, \delta \mid x). \tag{4.8}$$

Note that this corresponds to the function in relation (3.2) of Wellner (1982).

LEMMA 4.3. *Let $\{f_{Dn}\} \in \mathcal{C}(f_D, \beta)$, and S_n is the corresponding sequence of bivariate survival functions obtained from* (4.3) *with f_{Dn} substituting f_D. Let $\beta_1(z, \delta \mid x)$ and $\beta_2(x)$ be the functions related to β through Lemma* 4.1. *Then*

$$\sup_{0<y<T} \left| n^{1/2}\left[S_n(y \mid x) - S(y \mid x) \right] + 2S(y \mid x)\left\langle \beta_1(\cdot, \cdot \mid x), \gamma_y(\cdot, \cdot \mid x) \right\rangle_{\mu_1} \right| \to 0,$$

as $n \to \infty$, for almost surely $[\nu_2]$ all x, and

$$\sup_{(x,y)\in\Omega} \left| n^{1/2}[S_n(x,y) - S(x,y)] + 2\int_x^\infty S(y \mid x_1)\left\langle \beta_1(\cdot, \cdot \mid x_1), \gamma_y(\cdot, \cdot \mid x_1) \right\rangle_{\mu_1} g(x_1)\, d\nu_2(x_1) - 2\int_x^\infty S(y \mid x_1)\beta_2(x_1) g^{1/2}(x_1)\, d\nu_2(x_1) \right| \to 0.$$

The first relation in Lemma 4.3 follows directly from Wellner [(1982), Lemma 2], while the second follows by writing

$$\begin{aligned} n^{1/2}&\left[S_n(x,y) - S(x,y)\right] \\ &= n^{1/2}\int_x^\infty \left[S_n(y\mid x_1) - S(y\mid x_1)\right]g(x_1)\,d\nu_2(x_1) \\ &\quad + n^{1/2}\int_x^\infty S_n(y\mid x_1)\left[g_n(x_1) - g(x_1)\right]d\nu_2(x_1) \end{aligned}$$

and some easy L^2 calculations.

Let v_1 and v_2 be functions of bounded variation on $\mathbb{R}$ and $[0,T]$, respectively. Set

$$\begin{aligned} V(z\mid x) &= \int_z^T S(y\mid x)\,dv_2(y), \\ U(z\mid x) &= \int_z^T S(y\mid x)C_1(y\mid x)\,dv_2(y), \\ \eta(z,\delta\mid x) &= \left[\frac{V(z\mid x)I(\delta=1)}{H(z\mid x)} + U(z\mid x) - C_1(z\mid x)V(z\mid x)\right] \\ &\quad \times f_D^{1/2}(z,\delta\mid x)I(0<z<T), \end{aligned} \tag{4.9}$$

and

$$\sigma_2^2(v_2\mid x) = \int_0^T \left[V(z\mid x)\right]^2 dC_1(z\mid x). \tag{4.10}$$

Note that the functions in relation (4.9) are analogues to the functions defined in Wellner [(1982), (3.3) and (3.4)] while the variance in relation (4.10) is the analogue of relation (3.6) of the last reference.

LEMMA 4.4. *Let $\{f_{Dn}\}$ and S_n be as in Lemma* 4.3, *and let v_1 and v_2 be as specified above. Then*

$$\begin{aligned} \lim_{n\to\infty} &\int_\Omega n^{1/2}\left[S_n(x,y) - S(x,y)\right]dv_1(x)\,dv_2(y) \\ &= -2\int \left[v_1(x_1) - v_1(0)\right]\left\langle \beta_1(\cdot,\cdot\mid x_1), \eta(\cdot,\cdot\mid x_1)\right\rangle_{\mu_1} g(x_1)\,d\nu_2(x_1) \\ &\quad + 2\int \left[v_1(x_1) - v_1(0)\right]\int_0^T S(y\mid x_1)\,dv_2(y)\beta_2(x_1)g^{1/2}(x_1)\,d\nu_2(x_1). \end{aligned}$$

The proof of Lemma 4.4 follows easily from Lemma 4.3 and the fact that

$$\int_0^T S(y\mid x_1)\left\langle \beta_1(\cdot,\cdot\mid x_1), \gamma_y(\cdot,\cdot\mid x_1)\right\rangle_{\mu_1} dv_2(y) = \left\langle \beta_1(\cdot,\cdot\mid x_1), \eta(\cdot,\cdot\mid x_1)\right\rangle_{\mu_1}$$

[see also Wellner (1982), Lemma 3].

LEMMA 4.5. *The characteristic functional of the Gaussian process Z on Ω with covariance function* (3.2) *is*

$$E\exp\left[i\int_{\Omega} Z(x,y)\,dv_1(x)\,dv_2(y)\right] = \exp\left[\frac{-(\sigma_1^2+\sigma_2^2)}{2}\right], \tag{4.11}$$

where $\sigma_1^2 = \sigma_1^2(v_1, v_2)$ and $\sigma_2^2 = \sigma_2^2(v_1, v_2)$ are given by

$$\begin{aligned}\sigma_1^2 &= \int\left[(v_1(x_1)-v_1(0))\int_0^T S(y\,|\,x_1)\,dv_2(y)\right]^2 g(x_1)\,d\nu_2(x_1)\\ &\quad-\left[\int (v_1(x_1)-v_1(0))\int_0^T S(y\,|\,x_1)\,dv_2(y)g(x_1)\,d\nu_2(x_1)\right]^2,\\ \sigma_2^2 &= \int [v_1(x_1)-v_1(0)]^2\sigma_n^2(v_2\,|\,x_1)g(x_1)\,d\nu_2(x_1),\end{aligned} \tag{4.12}$$

with $\sigma_2^2(v_1\,|\,x)$ given in (4.10).

This follows easily by noting that the two components of the Z-process are independent and that the function $\xi_x(z,\delta,y)$ appears in the iid representation of the product limit estimator [Lo and Singh (1986)].

Now choose

$$\begin{aligned}\beta_2^*(x) &= \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}g^{1/2}(x)\\ &\quad\times\left\{[v_1(x)-v_1(0)]\int_0^T S(y\,|\,x)\,dv_2(y)\right.\\ &\qquad\left.-\int [v_1(x_1)-v_1(0)]\int_0^T S(y\,|\,x_1)\,dv_2(y)g(x_1)\,d\nu_2(x_1)\right\},\\ \beta_1^*(z,\delta\,|\,x) &= \frac{-1}{\sqrt{\sigma_1^2+\sigma_2^2}}[v_1(x)-v_1(0)]\eta_0(z,\delta\,|\,x),\end{aligned} \tag{4.13}$$

where $\eta_0(z,\delta\,|\,x) = \eta(z,\delta\,|\,x) - U(0\,|\,x)f_D^{1/2}(z,\delta\,|\,x)$. It is easy to see that

$$\begin{aligned}&\langle\beta_2^*, g^{1/2}\rangle_{\nu_2} = 0, \qquad \langle\beta_2^*,\beta_2^*\rangle_{\nu_2} = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2},\\ &\int [v_1(x)-v_1(0)]\int_0^T S(y\,|\,x)\,dv_2(y)g^{1/2}(x)\beta_2^*(x)\,d\nu_2(x) = \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}.\end{aligned} \tag{4.14}$$

Next, from Wellner [(1982), Lemma 5], it follows that

$$\|\eta(\cdot,\cdot\,|\,x)\|^2 = \sigma_2^2(v_2\,|\,x) + U(0\,|\,x)^2$$

and

$$\left\langle\eta(\cdot,\cdot\,|\,x), f_D^{1/2}(\cdot,\cdot\,|\,x)\right\rangle_{\mu_1} = U(0\,|\,x),$$

so that

$$\left\langle \eta_0(\cdot,\cdot\,|\,x), f_D^{1/2}(\cdot,\cdot\,|\,x)\right\rangle_{\mu_1} = 0$$

and

$$\|\eta_0(\cdot,\cdot\,|\,x)\|_{\mu_1}^2 = \left\langle \eta_0(\cdot,\cdot\,|\,x), \eta(\cdot,\cdot\,|\,x)\right\rangle_{\mu_1} = \sigma_2^2(v_2\,|\,x).$$

Therefore,

$$\left\langle \beta_1^*(\cdot,\cdot\,|\,x), f_D^{1/2}(\cdot,\cdot\,|\,x)\right\rangle_{\mu_1} = 0,$$

$$\left\langle \beta_1^*(\cdot,\cdot\,|\,x), \beta_1^*(\cdot,\cdot\,|\,x)\right\rangle_{\mu_1} = \frac{\left(v_1(x) - v_1(0)\right)^2}{\sigma_1^2 + \sigma_2^2}\sigma_2^2(v_2\,|\,x), \tag{4.15}$$

$$\int [v_1(x) - v_1(0)]\left\langle \beta_1^*(\cdot,\cdot\,|\,x), \eta(\cdot,\cdot\,|\,x)\right\rangle_{\mu_1} g(x)\,d\nu_2(x) = \frac{-\sigma_2^2(v_1, v_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

To prove Theorem 5.1, let $\{\widetilde{S}_n\}$ be a regular estimator of S. The characteristic functional of $n^{1/2}(\widetilde{S}_n - S_n)$ under f_{Dn} is

$$\begin{aligned}
E_{f_{Dn}} &\exp\left\{ i \int_\Omega n^{1/2}\left[\widetilde{S}_n(x,y) - S_n(x,y)\right] dv_1(x)\,dv_2(y)\right\} \\
&= E_{fD} \exp\bigg\{ i \int_\Omega n^{1/2}\left[\widetilde{S}_n(x,y) - S(x,y)\right] dv_1(x)\,dv_2(y) + L_n \\
&\qquad + 2i \int [v_1(x_1) - v_1(0)]\left\langle \beta_1(\cdot,\cdot\,|\,x_1), \eta(\cdot,\cdot\,|\,x_1)\right\rangle g(x_1)\,d\nu_2(x_1) \\
&\qquad - 2i \int [v_1(x_1) - v_1(0)] \int_0^T S(y\,|\,x_1) \\
&\qquad\qquad \times dv_2(y)\beta_2(x_1) g^{1/2}(x_1)\,d\nu_2(x_1)\bigg\} + o_p(1),
\end{aligned} \tag{4.16}$$

by Lemma 4.4. By regularity, (4.16) converges to

$$E \exp\left(i \int_\Omega \widetilde{Z}(x,y)\,dv_1(x)\,dv_2(y)\right).$$

The rest of the proof proceeds as in Beran (1977) except that we choose $\beta = h\beta^*$, where β^* is specified by β_1^* and β_2^* defined in (4.13) through Lemma 4.1. Note that by Lemma 4.1 and relations (4.14) and (4.15) we have $\beta \perp f_D^{1/2}$ and $\|\beta\|^2 = h^2$. The final result is

$$\phi(v_1, v_2, 0) = \phi(v_1, v_2, -\sigma)\exp\left(-\sigma^2/2\right),$$

where $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$, $\phi(v_1, v_2, 0)$ is the characteristic functional of the process $\widetilde{Z}$, $\exp(-\sigma^2/2)$ is the characteristic functional of Z by Lemma 4.5 and $\phi(v_1, v_2, -\sigma)$ is the characteristic functional of the process W.

REMARK 4.1. Although the assumption $\Omega = \mathbb{R} \times [0, T]$ was imposed in this section for notational convenience, it should be noted that it is not very restrictive. Indeed, the real line can be partitioned into a finite number of intervals I_k. For each interval I_k take $T_k = \inf\{T_x;\, x \in I_k\}$ and apply the above arguments for the process that restricts x on I_k.

5. Application to least squares estimation. As was explained in the Introduction, this research was motivated by applications to linear regression. This section proposes an extension of the least squares estimator (LSE) to polynomial regression with censored data. An asymptotic representation of the proposed LSE is obtained from which its asymptotic normality follows. Applications to general multiple regression are discussed in Remark 5.3 and in the next section.

Suppose that some strictly monotonic transformation of the response variable, such as $\log Y$, satisfies a polynomial regression model. Since we will only refer to this transformed variable in this section we denote it also by Y. (Similarly we will keep the same notation for the bivariate and conditional distributions and all other quantities.) Thus we assume $E(\mathbf{Y} \mid \mathcal{X}) = \mathcal{X}\beta$, where $\mathbf{Y}$ is the $n \times 1$ vector of the transformed response, and $\mathcal{X}$ denotes the $n \times (p+1)$ design matrix whose kth column has elements X_i^k, $k = 0, \dots, p$. With uncensored data the LSE is

$$P^{-1}\mathcal{X}'\mathbf{Y} = P^{-1}\begin{pmatrix} \sum X_i^0 Y_i \\ \vdots \\ \sum X_i^p Y_i \end{pmatrix} = nP^{-1}\begin{pmatrix} \int x^0 y\, d\widehat{F}(x,y) \\ \vdots \\ \int x^p y\, d\widehat{F}(x,y) \end{pmatrix},$$

where $P = \mathcal{X}'\mathcal{X}$ and $\widehat{F}$ is the bivariate empirical distribution function from $(X_1, Y_1), \dots, (X_n, Y_n)$. The proposed extension consists in replacing the uncensored-data empirical distribution function by the present estimate of the bivariate distribution. It is worth noting that the proposed estimator estimates the same functional as the uncensored data LSE. Thus it is suitable even in cases where the errors are asymmetric and/or heteroscedastic. In fact the asymptotics developed below do not assume symmetry or homoscedasticity.

The idea of replacing the uncensored-data empirical distribution with a censored-data version was also used in Miller's (1976) extension of the LSE. The difference is that Miller used it on the quadratic expression to be minimized whereas we use it directly on the expression for the LSE. Most other approaches for extending the LSE use a transformation of the data and apply ordinary least squares to the transformed data, without regard to censoring.

The first such method was proposed by Buckley and James (1979), but their transformation is in terms of the error distribution function so it requires iteration. Lai and Ying (1991) established asymptotic properties of a modified Buckley–James estimator. Koul, Susarla and Van Ryzin (1981) and Leurgans (1987) have proposed different data transformations that make possible the direct application of the least squares principle. Zhou (1992) establishes the asymptotic normality of Leurgan's estimator and claims his result also holds for heteroscedastic regression models. His derivations, however, assume that the censoring distribution does not depend on the covariate. The same assumption is imposed in Koul, Susarla and Van Ryzin (1981). On a different note, Stute (1992) proposed a weighted LSE which is based on an estimate of the multivariate distribution of the response variable and the covariates. Under the assumption that the censoring variable is independent of Y, Stute establishes strong consistency of his estimator.

The following asymptotic theory for the proposed LSE requires three additional assumptions.

LS1. The upper bound for the support of the conditional distribution of Y given $X = x$ is less than or equal to the upper bound of the support of the conditional distribution of C given $X = x$ for all x.

LS2. The distribution of X has bounded support.

LS3. The p-variate distribution of $(X, X^2, \dots, X^p)$ is nonsingular.

The first assumption can be unrealistic in certain real-life applications. A method that could weaken this assumption is described in the next section.

Write $\beta = [E(n^{-1}P)]^{-1}(\int x^0 y\, dF(x,y), \dots, \int x^p y\, dF(x,y))'$. Since the bivariate distribution F is estimated only in a restricted domain, what is estimated is

$$\beta_T = \left[E\left(n^{-1}P\right)\right]^{-1}\left(\int x^0 \int_0^T y\, dF(y\,|\,x)\, dG(x), \dots, \int x^p \int_0^T y\, dF(y\,|\,x)\, dG(x)\right)'.$$

By assumption LS1 the bias $\beta_T - \beta$ can be made arbitrarily small by choosing T large enough.

REMARK 5.1. One of the referees conjectures that the assumption

$$n^{1/2} \int x^k \int_{Y_{\max}}^{r(x)} yF(dy\,|\,x)G(dx) \to_P 0,$$

for $k = 0, \dots, p$, would suffice for consistency and asymptotic normality of the LSE. The above condition is the regression analogue of condition (3.3) of Gill (1983). The proof of such a result requires extending the results of Gill (1983) to estimates of the conditional survival function as well as to estimates of the bivariate distribution. It should be noted that the iid representation of the estimate of the bivariate distribution, as developed in this paper, holds for a restricted domain largely due to the use of results similar in nature of those in Lo and Singh (1986).

In what follows we write β instead of β_T. Then

$$\widehat{\beta} - \beta = nP^{-1}\begin{pmatrix} \int x^0 y\, d\big[\widehat{F}(x,y) - F(x,y)\big] \\ \vdots \\ \int x^p y\, d\big[\widehat{F}(x,y) - F(x,y)\big] \end{pmatrix} + n\Big(P^{-1} - \big[E(P)\big]^{-1}\Big)\begin{pmatrix} \int x^0 y\, dF(x,y) \\ \vdots \\ \int x^p y\, dF(x,y) \end{pmatrix}. \tag{5.1}$$

Using the fact that the present estimate of the bivariate distribution is obtained by averaging conditional distribution functions, it follows that, for $k = 0, \ldots, p$,

$$\begin{aligned} &\int x^k y\, d\big[\widehat{F}(x,y) - F(x,y)\big] \\ &\quad = \int x^k \int y\, dF(y\,|\,x)\, d\big[\widehat{G}(x) - G(x)\big] \\ &\quad + \int x^k \int y\, d\Big[\widehat{F}(y\,|\,x) - F(y\,|\,x)\Big]\, d\widehat{G}(x), \end{aligned} \tag{5.2}$$

where the insides of the double integrals go from $-\infty$ to T. Integrating by parts and using Lemma 3.1, the second term on the right-hand side of (5.2) is seen to be asymptotically equivalent to

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n}(na_n)^{-1}\sum_{j=1}^{n} X_i^k \left[\int_{-\infty}^{T} \xi_{X_i}(Z_j, \delta_j, y)\, dy - T\xi_{X_i}(Z_j, \delta_j, T)\right] \\ &\quad \times 0.5 I\big(|\widehat{G}(X_i) - G(X_j)| \le a_n\big). \end{aligned} \tag{5.3}$$

Note that this expression is similar to the second term on the right-hand side of (3.3). Using an expansion similar to (3.4) it is easily seen, using assumption LS2, that arguments similar to those in Lemmas 3.2–3.4 go through. Thus (5.3) is asymptotically equivalent to

$$n^{-1}\sum_{j=1}^{n} X_j^k\left(\int_{-\infty}^{T} \xi_{X_j}(Z_j, \delta_j, y)\, dy - T\xi_{X_j}(Z_j, \delta_j, T)\right) = n^{-1}\sum_{j=1}^{n}\Psi_{kj}, \tag{5.4}$$

say, which is a sum of iid random variables with zero mean and finite variance. From that it is easily seen that the second term on the right-hand side of (5.2) is asymptotically independent from the first as well as from the second term on the right-hand side of (5.1). Relations (5.2)–(5.4) establish a convenient asymptotic representation of the first term of the right-hand side of (5.1). To complete the asymptotic theory of the proposed LSE, we will establish an asymptotic representation of the second term of the right-hand side of (5.1). Write $n^{-1}P = A_n$, $E(n^{-1}P) = A$ and

$$n\left(P^{-1} - \left[E(P)\right]^{-1}\right) = A_n^{-1} - A^{-1} = \left(g_{lk}(A_n) - g_{lk}(A)\right)_{l,k}$$

where $g_{lk}(A_n)$ is the (l,k)th element of A_n^{-1}, and similarly for $g_{lk}(A)$. Note that assumption LS3 implies that A is of full rank; also, the same assumption and arguments similar to those in Arnold [(1981), Theorem 17.8, page 316] imply that A_n is of full rank with probability 1. By an application of the δ-method to each element $g_{lk}(A_n) - g_{lk}(A)$, it follows that an asymptotically equivalent expression for $A_n^{-1} - A^{-1}$ is

$$\sum_{r=1}^{p}\sum_{s=1}^{p} \widetilde{G}_{rs}(A)(\widehat{a}_{rs} - a_{rs}), \tag{5.5}$$

where a_{rs} and $\widehat{a}_{rs}$ are the (r,s)th element of A and A_n, respectively, and $\widetilde{G}_{rs}(A)$ is the $(p+1)\times(p+1)$ matrix with (l,k)th element $g_{lk}^{rs}(A)$, the partial derivative of $g_{lk}(A)$ with respect to a_{rs}. Combining (5.1)–(5.5), we have the following asymptotic representation for $\widehat{\beta} - \beta$ whose covariance matrix is straightforward to compute.

THEOREM 5.1. *Under assumptions* A1–A4 *and* LS1, LS2 *and* LS3, *we have that the right-hand side of* (5.1) *is asymptotically equivalent to*

$$nP^{-1}\begin{pmatrix} n^{-1}\sum_j \Psi_{0j} \\ \vdots \\ n^{-1}\sum_j \Psi_{pj} \end{pmatrix} + nP^{-1}\begin{pmatrix} \int x^0 \int y\,dF(y\,|\,x)d\left[\widehat{G}(x) - G(x)\right] \\ \vdots \\ \int x^p \int y\,dF(y\,|\,x)d\left[\widehat{G}(x) - G(x)\right] \end{pmatrix}$$

$$+\sum_{r=1}^{p+1}\sum_{s=1}^{p+1} \widetilde{G}_{rs}(A)\begin{pmatrix} \int x^0 y\,dF(x,y) \\ \vdots \\ \int x^p y\,dF(x,y) \end{pmatrix}(\widehat{a}_{rs} - a_{rs}).$$

REMARK 5.2. In the uncensored case, one usually obtains the conditional distribution of the LSE given $\mathcal{X}$. In the present formulation, it is more convenient to obtain the unconditional distribution.

REMARK 5.3. Similar formulas for the LSE hold for general multiple regression, provided the censoring variable is independent from all covariates.

The case that the censoring variable depends on the covariates but, given the covariates, it is conditionally independent from the response variable is briefly addressed in the next section.

6. Further research.

1. The method of estimating the bivariate distribution by averaging conditional distributions over a range of values of the conditioning variable has desirable properties for applications to regression analysis. Indeed, when linear regression seems a reasonable model for the relation between X and Y, the present method allows one to use information about the tails of the residuals from a region with light censoring in order to achieve better estimation of the tails in regions of heavy censoring. This can be used to decrease the bias of the proposed LSE.
2. For applications to general multiple regression with covariates $X_1, \dots, X_p$, the present estimate of the bivariate distribution will not suffice. Indeed, under the assumption that Y and C are conditionally independent given $(X_1, \dots, X_p)$, the present estimate of the bivariate distribution of (X_1, Y), say, will not have the stated properties. To circumvent this difficulty, one needs to estimate the multivariate distribution of $(X_1, \dots, X_p, Y)$ when only Y is subject to censoring. This can be achieved by extending the present method of averaging the estimated conditional distributions of Y given X to multiple covariates. This estimate of the multivariate distribution will yield estimates of the bivariate marginal distributions of (X_k, Y), $k = 1, \dots, p$, and the marginal distribution of Y which can be used in the form of the LSE as described in Section 5.
3. The NNE can be extended to the case that the covariate X is also subject to random censoring. An obvious way to achieve this extension is to use the product-limit estimator for G to determine the "nearest neighbors" among the uncensored x-values. This estimate of the bivariate distribution can then be used to extend the proposed LSE in this context. The LSE when both x and y are subject to censoring has applications in astronomy.

APPENDIX A

Results needed for the main proofs. This appendix establishes a rate of convergence and an oscillation result for the conditional process

$$\beta_n(z \mid X_i) = (na_n)^{-1/2} \sum_{j=1}^{n} \left[I(Z_j > z) - H(z \mid X_i) \right] I\left(|\widehat{G}(X_j) - \widehat{G}(X_i)| \le a_n \right)$$

uniformly in i and, as a corollary, obtains the rate for a certain integral uniformly in i. These results are used in the proofs of Appendix B.

PROPOSITION A.1. *Assume that the joint density $f_G(z, u)$ of $(Z, G(X))$ is twice continuously differentiable with respect to u. Let f_G'' denote the second derivative* *of f_G with respect to u, and assume that $\int \sup\{|f_G''(z,u)|;\ 0 < u < 1\}\,dz < \infty$. Assume also that $na_n^5[\log a_n^{-1}]^{-1} \to 0$. Then*

$$\sup\left\{\left[\log a_n^{-1}\right]^{-1/2}\left|\beta_n(z \mid X_i)\right|;\ 0 < z < T_{X_i},\ i = 1,\dots,n\right\} = O(1),$$

almost surely, where T_x denotes a number less than the upper bound of the support of the conditional distribution of Z given x.

PROOF. Write

$$\beta_n(z \mid X_i) = \beta_n^1(z \mid X_i) + \beta_n^2(z \mid X_i) + \beta_n^3(z \mid X_i), \tag{A.1}$$

where

$$\begin{aligned}
\beta_n^1(z \mid x) &= (na_n)^{-1/2}\sum_{j=1}^{n}\left[I(Z_j > z) - H_n(z \mid x)\right] \\
&\qquad \times I\left(|G(X_j) - G(x)| \le a_n\right), \\
\beta_n^2(z \mid x) &= (na_n)^{-1/2}\sum_{j=1}^{n}\left[H_n(z \mid x) - H(z \mid x)\right]I\left(|\widehat{G}(X_j) - \widehat{G}(x)| \le a_n\right), \\
\beta_n^3(z \mid x) &= (na_n)^{-1/2}\sum_{j=1}^{n}\left[I(Z_j > z) - H_n(z \mid x)\right] \\
&\qquad \times \left[I\left(|\widehat{G}(X_j) - \widehat{G}(x)| \le a_n\right) - I\left(|G(X_j) - G(x)| \le a_n\right)\right],
\end{aligned}$$

with $H_n(z \mid x) = P(Z > z \mid |G(X) - G(x)| \le a_n)$. Using the assumptions made on $f_G(z,u)$, it is easy to verify that $H_n(z \mid x) - H(z \mid x) = O(a_n^2)$ uniformly in x and z. It follows that $\beta_n^2(z \mid X_i)$ multiplied by $[\log a_n^{-1}]^{-1/2}$ is

$$O\left(\left(na_n^5\left[\log a_n^{-1}\right]^{-1}\right)^{1/2}\right) \to 0,$$

almost surely uniformly in i, z. For $\beta_n^3(z \mid X_i)$ we note that $|G(X_j) - G(x)| \le a_n$ implies that

$$\left|\left[\widehat{G}(X_j) - \widehat{G}(x)\right] - \left[G(X_j) - G(x)\right]\right| \le Kn^{-1/2}\left[a_n \log a_n^{-1}\right]^{1/2}$$

almost surely for n sufficiently large [Stute (1982)]. It follows that, for n large enough,

$$I\left(|\widehat{G}(X_j) - \widehat{G}(x)| \le a_n\right) - I\left(|G(X_j) - G(x)| \le a_n\right) \ne 0$$

only if $[\widehat{G}(X_j) - \widehat{G}(x)]/a_n$ is in an interval around 1 (or -1) of length

$$Kn^{-1/2}a_n^{-1/2}\left[\log a_n^{-1}\right]^{1/2}.$$

Thus, for each i,

$$\text{(A.2)}\qquad \begin{aligned}&\sum_j \left|I\left(|\widehat{G}(X_j) - \widehat{G}(X_i)| \le a_n\right) - I\left(|G(X_j) - G(X_i) \le a_n\right)\right| \\ &\qquad < K n^{1/2} a_n^{1/2} \left[\log a_n^{-1}\right]^{1/2},\end{aligned}$$

for n sufficiently large almost surely. Consequently, for n large enough, $\beta_n^3(z \mid X_i)$ multiplied by $[\log a_n^{-1}]^{-1/2}$ remains bounded uniformly in i and z. Finally, by the Dvoretzky–Kiefer–Wolfowitz inequality [Dvoretzky, Kiefer and Wolfowitz (1956)],

$$\begin{aligned}&P\left(\sup\left\{\left[\log a_n^{-1}\right]^{-1/2} \left|\beta_n^1(z \mid X_i)\right|;\ 0 < z < T_{X_i}\right\} > d\right) \\ &\qquad < K \exp\left[-2d^2 \log a_n^{-1}\right].\end{aligned}$$

Selecting $d = [l/2]^{1/2}$, where l satisfies $\sum_{n=1}^{\infty} n a_n^l < \infty$, implies the statement of the proposition with β_n^1 substituted for β_n. This completes the proof of the proposition. □

PROPOSITION A.2. *Let* $J_c(x) = \{(z_1, z_2): |H(z_1 \mid x) - H(z_2 \mid x)| \le c\}$, *and set*

$$\text{(A.3)}\qquad \omega_n(c \mid x) = \sup\left\{\left|\beta_n(z_1 \mid x) - \beta_n(z_2 \mid x)\right|;\ (z_1, z_2) \in J_c(x)\right\}.$$

Consider the assumptions on the joint density $f_G(z, u)$ *imposed in Proposition* A.1. *Let* $\bar{a}_n$ *be any sequence of positive numbers tending to zero that satisfies*

$$\bar{a}_n a_n^{-2} \to \infty, \qquad n a_n^5 \bar{a}_n^{-1} \left[\log a_n^{-1}\right]^{-1} \to 0, \qquad n a_n \bar{a}_n^2 \left[\log a_n^{-1}\right]^{-1} \to \infty,$$

and there exists an $l > 0$ *such that* $\sum_{n=1}^{\infty} n \bar{a}_n^l < \infty$. *Then*

$$\sup\left\{\frac{\omega_n(\bar{a}_n \mid X_i)}{\left[\bar{a}_n \log \bar{a}_n^{-1}\right]^{1/2}};\ i = 1, \ldots, n\right\} = O(1),$$

almost surely.

PROOF. Consider the decomposition of β_n given in (A.1) and let $\omega_n^k(c \mid x)$ be defined as in (A.3) but with $\beta_n^k(z \mid x)$ substituted for $\beta_n(z \mid x)$, $k = 1, 2, 3$. The assumption that $\bar{a}_n a_n^{-2} \to \infty$ and the fact that

$$\left|H_n(z_1 \mid x) - H_n(z_2 \mid x)\right| < \left|H(z_1 \mid x) - H(z_2 \mid x)\right| + O\left(a_n^2\right)$$

implies that $|H_n(z_1 \mid x) - H_n(z_2 \mid x)| \le K\bar{a}_n$ whenever $|H(z_1 \mid x) - H(z_2 \mid x)| \le \bar{a}_n$. Thus, Lemma 2.4 of Stute (1982) with $s = [2(1 + l)(1 - \delta_1)^{-5} \log a_n^{-1}]^{1/2}$, some $\delta_1 \in (0, 1)$ as in the aforementioned lemma and any l that satisfies $\sum_{n=1}^{\infty} n \bar{a}_n^l < \infty$ implies that the statement of the proposition remains true when ω_n^1 is substituted for ω_n. Using again the fact that $H_n(z \mid x) - H(z \mid x) = O(a_n^2)$, uniformly in

x and z, it is seen that the assumption $na_n^5\bar{a}_n^{-1}[\log a_n^{-1}]^{-1} \to 0$ implies that the statement of the proposition is true when ω_n^2 is substituted for ω_n. Finally, to show the same statement for ω_n^3, note that, by (A.2) and for $z_1 < z_2$,

$$
\begin{aligned}
&\left|\beta_n^3(z_1 \mid X_i) - \beta_n^3(z_2 \mid X_i)\right| \\
&\quad \le \left|H_n(z_1 \mid X_i) - H_n(z_2 \mid X_i)\right| K\left[\log a_n^{-1}\right]^{1/2} \\
&\quad + \left|(na_n)^{-1/2}\sum_{j=1}^{n}\left[I(z_1 < Z_j < z_2) - H_n^*(z_1 \mid X_i) + H_n^*(z_2 \mid X_i)\right] I\left(A_n(X_j, X_i)\right)\right| \\
&\quad + \left|H_n^*(z_1 \mid X_i) - H_n^*(z_2 \mid X_i)\right| K\left[\log a_n^{-1}\right]^{1/2},
\end{aligned}
\tag{A.4}
$$

where

$$
\begin{aligned}
A_n(X_j, x) = &\left[a_n \le |G(x) - G(X_j)| \le a_n + Kn^{-1/2}\left[a_n \log a_n^{-1}\right]^{1/2}\right] \\
&\cup \left[a_n - Kn^{-1/2}\left[a_n \log a_n^{-1}\right]^{-1} \le |G(x) - G(X_j)| \le a_n\right]
\end{aligned}
$$

and $H_n^*(z \mid x) = P(z_j > z \mid A_n(x_j, x))$. Using the assumptions made on f_G, it is easy to verify that

$$
\left|H_n^*(z_1 \mid x) - H_n^*(z_2 \mid x)\right| \le \left|H(z_1 \mid x) - H(z_2 \mid x)\right| + O(a_n^2) = O(\bar{a}_n)
$$

uniformly on x and $(z_1, z_2) \in J_{\bar{a}_n}(x)$; as before, the same in true for $H_n(z_1 \mid x) - H_n(z_2 \mid x)$. It follows that the supremum of the first and third terms on the right-hand side of (A.4) over all $(z_1, z_2) \in J_{\bar{a}_n}(X_i)$ and over all i, divided by $[\bar{a} \log a_n^{-1}]^{1/2}$ tends to zero. Application of Lemma 2.4 of Stute (1982) with the same choice of s specified above gives that the supremum of the second term in (A.4) over all $(z_1, z_2) \in J_{\bar{a}_n}(X_i)$ and over all i, divided by $[\bar{a}_n \log a_n^{-1}]^{1/2}$ remains bounded almost surely. Note that the assumption $na_n\bar{a}_n^2[\log a_n^{-1}]^{-1} \to \infty$ guarantees that condition (iv) of Lemma 2.4 of Stute (1982) holds (with n replaced by $Kn^{1/2}a_n^{1/2}[\log a_n^{-1}]^{1/2}$) for n large enough. This concludes the proof of the proposition. □

REMARK A.1. Replacing the condition on $f_G(z, u)$ by a similar condition on $f_G(z, 1, u)$, the joint density of $(Z, \delta, G(X))$ at $\delta = 1$, it is easily seen that the result of Proposition A.2 is true for oscillations of the process $(na_n)^{1/2}[\widehat{H}_1(z \mid X_i) - H_1(z \mid X_i)]$.

COROLLARY A.1. *Assume that* $na_n^3[\log a_n^{-1}]^{3.5} \to \infty$ *and* $na_n^5[\log a_n^{-1}]^{-1} \to 0$, *and let the assumptions imposed on* $f_G(z, u)$ *and* $f_G(z, 1, u)$ *in Proposition* A.1 *and Remark* A.1 *hold. Then*

$$
\sup\left\{\left|\int_0^y \left[\widehat{H}(z \mid X_i)^{-1} - \widehat{H}(z \mid X_i)^{-1}\right] d\left(H_1(z \mid X_i) - H_1(z \mid X_i)\right)\right|; \; 0 < y < T_{X_i},\, i = 1, \dots, n\right\} = o(n^{-1/2}),
\tag{A.5}
$$

almost surely.

PROOF. For each y, the interval $[0, T_{X_i}]$ can be partitioned into subintervals $[z_j, z_{j+1}]$, $j = 1, \dots, k_n$, $0 = z_1 < z_2 < \cdots < z_{k_n+1} = T_{X_i}$, where $k_n = O((na_n)^{1/2} [\log a_n^{-1}]^{-3/4})$, such that

$$H(z_j \mid X_i) - H(z_{j+1} \mid X_i) < K(na_n)^{-1/2}\big[\log a_n^{-1}\big]^{3/4} = \bar{a}_n,$$

and the integral in (A.5) is bounded by

$$\begin{aligned}
&k_n \sup\Big\{\big|\widehat{H}(z \mid X_i)^{-1} - H(z \mid X_i)^{-1}\big|;\ 0 < z < T_{X_i}\Big\}\\
&\quad \times \sup\Big\{\big|\widehat{H}_1(z_1 \mid X_i) - \widehat{H}_1(z_2 \mid X_i) - H_1(z_1 \mid X_i) + H_1(z_2 \mid X_i)\big|; (z_1, z_2) \in J_{\bar{a}_n}(X_i)\Big\}\\
&\quad + 2\sup\Big\{\big|\widehat{H}(z_1 \mid X_i)^{-1} - \widehat{H}(z_2 \mid X_i)^{-1} - H(z_1 \mid X_i)^{-1} + H(z_2 \mid X_i)^{-1}\big|;\\
&\qquad\qquad (z_1, z_2) \in J_{\bar{a}_n}(X_i)\Big\}.
\end{aligned}$$

Note that the sequence $\bar{a}_n$ that bounds $H(z_j \mid x) - H(z_{j+1} \mid x)$ satisfies the conditions of Proposition A.2. Use now Proposition A.1 and Remark A.1 to conclude that the supremum over all $y \in [0, T_{X_i}]$ and over all $i = 1, \dots, n$ of the first term above is $O(n^{-3/4} a_n^{-3/4} [\log a_n^{-1}]^{5/8})$. Using Proposition A.1, the expression inside the supremum of the second term can be written as

$$\begin{aligned}
&\Big|H(z_1 \mid X_i)^{-2}\Big[H(z_1 \mid X_i) - \widehat{H}(z_1 \mid X_i)\Big] - H(z_2 \mid X_i)^{-2}\Big[H(z_2 \mid X_i) - \widehat{H}(z_2 \mid X_i)\Big]\Big|\\
&\qquad + O\big((na_n)^{-1} \log a_n^{-1}\big)\\
&\quad = \Big|H(z_1 \mid X_i)^{-2}\Big[H(z_1 \mid X_i) - \widehat{H}(z_1 \mid X_i) - H(z_2 \mid X_i) + \widehat{H}(z_2 \mid X_i)\Big]\Big|\\
&\qquad + O\Big((na_n)^{-1}\big[\log a_n^{-1}\big]^{5/4}\Big),
\end{aligned}$$

almost surely, where the remainder term is uniform in $z \in [0, T_{x_i}]$ and in i. Using Proposition A. 2, this is seen to be $O(n^{-3/4} a_n^{-3/4} [\log a_n^{-1}]^{7/8})$ almost surely and uniformly in $z \in [0, T_{x_i}]$ and in i. This completes the proof of the corollary. □

APPENDIX B

Proofs.

PROOF OF LEMMA 3.1. Use relation (2.4) and a two-term Taylor expansion to write

$$\begin{aligned}
&\log \widehat{S}(y \mid X_i) - \int_0^y \widehat{H}(z- \mid X_i)^{-1} \widehat{H}_1(dz \mid X_i)\\
&\quad = \frac{1}{2} \sum_{j=1}^n \frac{I(Z_j \le y,\ \delta_j = 1) I\big(|\widehat{G}(X_i) - \widehat{G}(X_j)| \le a_n\big)}{\Big[\sum_{k=1}^n I(Z_k > Z_j -) I\big(|\widehat{G}(X_i) - \widehat{G}(X_k)| \le a_n\big)\Big]^2} \frac{1}{(1 - R_{ij})^2},
\end{aligned}$$

where R_{ij} is between zero and

$$\frac{I(|\widehat{G}(X_i) - \widehat{G}(X_j)| \le a_n)}{\sum_{k=1}^{n} I(Z_k > Z_{j-}) I(|\widehat{G}(X_i) - \widehat{G}(X_k)| \le a_n)}.$$

Proposition A.1 implies that, for n large enough,

$$H(z \mid X_i) - K\left[\log a_n^{-1}\right]^{1/2} (na_n)^{-1/2} < \widehat{H}(z \mid X_i)$$

uniformly in $0 < z < T_{X_i}$ and in i. Thus, for n large enough, $\widehat{H}(z \mid X_i)$ is bounded away from zero uniformly in $z \in [0, T_{X_i}]$ and in i. This implies that

$$\begin{aligned} &\sup\left\{\left|\log \widehat{S}(y \mid X_i) - \int_0^y \widehat{H}(z- \mid X_i)^{-1} \widehat{H}_1(dz \mid X_i)\right|;\right. \\ &\qquad\qquad\qquad \left. 0 < y < T_{X_i},\ i = 1, \dots, n\right\} \\ &\quad = O\left((na_n)^{-1}\right), \end{aligned} \tag{B.1}$$

for n large enough almost surely. Next write

$$\begin{aligned} &\int_0^y \widehat{H}(z \mid X_i)^{-1} \widehat{H}_1(dz \mid X_i) - \int_0^y H(z \mid X_i)^{-1} H_1(dz \mid X_i) \\ &\quad = -\int_0^y H(z \mid X_i)^{-2} \widehat{H}(z \mid X_i) H_1(dz \mid X_i) \\ &\qquad + \int_0^y H(z \mid X_i)^{-1} \widehat{H}_1(dz \mid X_i) \\ &\qquad + \int_0^y \frac{\left[H(z \mid X_i) - \widehat{H}(z \mid X_i)\right]^2}{H(z \mid X_i)^2 \widehat{H}(z \mid X_i)} H_1(dz \mid X_i) \\ &\qquad + \int_0^y \left[\widehat{H}(z \mid X_i)^{-1} - H(z \mid X_i)^{-1}\right]\left[\widehat{H}_1(dz \mid X_i) - H_1(dz \mid X_i)\right] \\ &\quad = (na_n)^{-1} \sum_{j=1}^{n} \xi_{X_i}(Z_j, \delta_j, y) \frac{0.5 I(|\widehat{G}(X_i) - \widehat{G}(X_j)| \le a_n)}{S(y \mid X_i)} + o(n^{-1/2}) \end{aligned} \tag{B.2}$$

uniformly in $y \in [0, T_{X_i}]$ and $i = 1, \dots, n$, by Proposition A.1 and Corollary A.1. The result of the lemma now follows from (B.1), (B.2) and a Taylor expansion. □

PROOF OF LEMMA 3.2. Using a two-term Taylor expansion, we have (see assumption A3 for notation)

$$\begin{aligned}T_{1n} &= n^{1/2}a_n^{-1}\int [G(x_1)-G(x_2)]\xi'_{x_2}(z_2,\delta_2,y_0)I(x_1>x_0)\\ &\quad\times I\big(|G(x_1)-G(x_2)|\le a_n\big)\widehat{G}(dx_1)\widehat{F}_D(dD_2)\\ &\quad+\frac{n^{1/2}a_n^{-1}}{2}\int [G(x_1)-G(x_2)]^2\xi''_x(z_2,\delta_2,y_0)\\ &\qquad\qquad\times I(x_1>x_0)I\big(|G(x_1)-G(x_2)|\le a_n\big)\widehat{G}(dx_1)\widehat{F}_D(dD_2)\\ &= I_1+I_2,\quad \text{say,}\end{aligned}$$

where x is such that $G(x)$ is between $G(x_1)$ and $G(x_2)$. By the assumption $na_n^4\to 0$ and assumption A3, it is easily seen that I_2 goes to zero almost surely and uniformly in x_0, y_0.

Next write $I_1 = I_3 + I_4$, where

$$\begin{aligned}I_3 &= n^{1/2}a_n^{-1}\int [G(x_1)-G(x_2)]\xi'_{x_2}(z_2,\delta_2,y_0)I(x_2>x_0)\\ &\qquad\times I\big(|G(x_1)-G(x_2)|\le a_n\big)\widehat{G}(dx_1)\widehat{F}_D(dD_2)\\ I_4 &= n^{1/2}a_n^{-1}\int [G(x_1)-G(x_2)]\xi'_{x_2}(z_2,\delta_2,y_0)\big[I(x_1>x_0)-I(x_2>x_0)\big]\\ &\qquad\times I\big(|G(x_1)-G(x_2)|\le a_n\big)\widehat{G}(dx_1)\widehat{F}_D(dD_2).\end{aligned}$$

By the assumption $na_n^4\to 0$ and assumption A3, it is easily seen that I_4 goes to zero almost surely and uniformly in x_0, y_0. Finally, using assumption A3,

$$\begin{aligned}|I_3| \le \sup_i \Big|&\xi'_{X_i}(Z_i,\delta_i,y_0)I(X_i>x_0)a_n^{-1}\\ &\times\int [G(x_1)-G(X_i)]I\big(|G(x_1)-G(X_i)|\le a_n\big)\alpha_n(dx_1)\Big|\end{aligned}$$

almost surely and uniformly in i since, integrating by parts, we have

$$\begin{aligned}a_n^{-1}&\int [G(x_1)-G(X_i)]I\big(|G(x_1)-G(X_i)|\le a_n\big)\alpha_n(dx_1)\\ &= \Big[\alpha_n\Big(G^{-1}\big(G(X_i)+a_n\big)\Big)-\Gamma_i\Big]-\Big[\alpha_n\Big(G^{-1}\big(G(X_i)-a_n\big)\Big)-\Gamma_i\Big]\\ &\quad\times a_n^{-1}\int \big[\alpha_n(x_1)-\Gamma_i\big]I\big(|G(x_1)-G(X_i)|\le a_n\big)\,dG(x_1);\end{aligned}$$

taking $\Gamma_i = \alpha_n(x^*)$, for some x^* such that $|G(x^*)-G(X_i)|\le a_n$, the above is easily seen to converge to zero almost surely and uniformly in i. This completes the proof of Lemma 3.2. □

PROOF OF LEMMA 3.3. Consider a sequence of kernel functions $K_n(u)$ such that $K_n(u) = I(-1+n^{-2}\le u\le 1-n^{-2})$ and $K_n(u)$ vanishes outside the interval

$(-1-n^{-2}, 1+n^{-2})$ and is continuously differentiable. It is not hard to verify that

$$n^{1/2}|2T_{3n} - A_n| \to 0 \quad \text{almost surely},$$

uniformly in $(z_0, y_0) \in \Omega$, where

$$n^{1/2}A_n = n^{1/2}a_n^{-1} \int h_0(x_1, z_2, \delta_2) \times \left[K_n\left(\frac{G_n(x_1) - G_n(x_2)}{a_n}\right) - K_n\left(\frac{G(x_1) - G(x_2)}{a_n}\right)\right]\widehat{G}(dx_1)\widehat{F}_D(dD_2),$$

and $h_0(x_1, z_2, \delta_2)$ is defined in Lemma 3.4. Using a one-term Taylor expansion and (A.2), it is seen that

$$|n^{1/2}A_n| \le K\left[n^{1/2}a_n^{-2}n^{-1/2}(a_n \log a_n^{-1})^{1/2}n^{1.5}(a_n \log a_n^{-1})^{1/2}\right]\Big/n^2$$

almost surely for all n large enough, uniformly in $(x_0, y_0) \in \Omega$. This completes the proof of the lemma. □

PROOF OF LEMMA 3.4. Write

$$\begin{aligned} n^{1/2}&\left[T_{2n} - \int h_0(x, z, \delta)[\widehat{F}_D(dD) - F_D(dD)]\right] \\ &= n^{1/2}\int h_0(x, z, \delta)Q_n(x)[\widehat{F}_D(dD) - F_D(dD)] \\ &\quad + n^{1/2}a_n^{-1}\int \xi_{x_2}(z_2, \delta_2, y_0)[I(x_1 > x_0) - I(x_2 > x_0)] \\ &\qquad \times 0.5I(|G(x_1) - G(x_2)| \le a_n)\widehat{G}(dx_1)[\widehat{F}_D(dD_2) - F_D(dD_2)] \\ &= I_5 + I_6, \end{aligned}$$

say, where

$$Q(x) = (na_n)^{-1}\sum_{i=1}^{n}\left[0.5I(|G(X_i) - G(x)| \le a_n) - a_n\right].$$

Define the process

$$W_n(t) = n^{1/2}\int h_0(x, z, \delta)I(x \le t)[\widehat{F}_D(dD) - F_D(dD)],$$

indexed by t, x_0 and y_0. Next let

$$Q_{1n}(x) = (2na_n)^{-1}\sum_{i=1}^{n} I(G(x) < G(X_i) + a_n) - \frac{1 - G(x) + a_n}{2a_n},$$

$$Q_{2n}(x) = (2na_n)^{-1}\sum_{i=1}^{n} I(G(x) < G(X_i) - a_n) - \frac{1 - G(x) - a_n}{2a_n},$$

and note that

$$|Q_n(\ x\ ) - Q_{1n}(x) + Q_{2n}(x)| \le (2na_n)^{-1},$$
$$Q_{1n}(x) = -0.5n^{-1/2}a_n^{-1}\alpha_n\Big(G^{-1}\big(G(x) - a_n\big)\Big),$$
$$Q_{2n}(x) = -0.5n^{-1/2}a_n^{-1}\alpha_n\Big(G^{-1}\big(G(x) - a_n\big)\Big),$$

for each fixed x and for n large enough. In this notation,

$$
\begin{aligned}
I_5 &= \int Q_n(x)\,dW_n(x) \\
&= \int \big[Q_{1n}(x) + Q_{2n}(x)\big]\,dW_n(x) + o_p(1) \qquad \text{(B.3)}\\
&= -\int W_n(x-)\,dQ_{1n}(x) - \int W_n(x-)\,dQ_{2n}(x) + o_p(1).
\end{aligned}
$$

Note that the remainder term in (B.3) converges to zero uniformly in (x_0, y_0). It is easy to verify that

$$
\begin{aligned}
&\int W_n(x-)\,dQ_{1n}(x) \\
&\quad = n^{-1/2}a_n^{-1}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_0(X_j, Z_j, \delta_j)\big[G(X_j) - I(X_i \le X_j)\big]
\end{aligned}
$$

and that

$$V_n = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_0(X_j, Z_j, \delta_j)\big[G(X_j) - I(X_i \le X_j)\big] \qquad \text{(B.4)}$$

remains bounded in probability. [Take the expected value of V_n^2 or see Serfling (1980), page 223.] Thus the first term on the right-hand side of (B.3) converges to zero in probability. That the second term on the right-hand side of (B.3) converges to zero is shown similarly. The same argument can also be used to show that I_6 converges to zero in probability. The proof of the lemma will now be completed by showing the stochastic equicontinuity [Pollard (1990), page 51] of I_5 and I_6 in (x_0, y_0). We will do this by verifying the conditions of Nolan and Pollard [(1988), Theorem 7]. Set

$$
\begin{aligned}
f(\xi_i, \xi_j) = 0.5\Big\{& h_0(X_j, Z_j, \delta_j)\big[G(X_j) - I(X_i \le X_j)\big] \\
&+ h_0(X_i, Z_i, \delta_i)\big[G(X_i) - I(X_j \le X_i)\big]\Big\}.
\end{aligned}
$$

Thus V_n in (B.4) is $V_n = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} f(\xi_i, \xi_j)$. Note that the dependence of $f(\xi_i, \xi_j)$ on (x_0, y_0) is not made explicit. Different values of (x_0, y_0) yield different functions $f(\cdot, \cdot)$, which generate the class $\mathcal{F}$. Now let

$$U_n = n^{-1} \sum_{1 \le i \ne j \le n} f(\xi_i, \xi_j). \qquad \text{(B.5)}$$

Note first that the class $\mathcal{F}$ has a constant envelope function F. Let T_n be the measure described in Nolan and Pollard [(1988), page 1293]. Using assumption A4, it can be seen that the covering number $N(\varepsilon, T_n, \mathcal{F}, F)$, which is the smallest cardinality of a subclass $\mathcal{F}^*$ of $\mathcal{F}$ such that $\min\{T_n \mid f - f^*|^2; f^* \in \mathcal{F}^*\} \leq \varepsilon^2 T_n F^2$ for each f in $\mathcal{F}$, is bounded by $A\varepsilon^{-V}$. This implies that conditions (i) and (ii) of Nolan and Pollard [(1988), Theorem 7] are satisfied. Using assumption A4 again it is seen that condition (iii) of the above theorem is also satisfied. Next,

$$V_n - U_n = n^{-1} \sum_{i=1}^{n} f(\xi_i, \xi_i),$$

which can be seen to converge to zero uniformly in (x_0, y_0) by an application of Theorem 8.2 of Pollard (1990). This and (B.3) show that I_5 converges to zero in probability uniformly in (x_0, y_0). Similar arguments show that I_6 converges to zero in probability uniformly in (x_0, y_0). This completes the proof of the lemma. □

Acknowledgments. This research was carried out while the author was a Visiting Fellow in the Centre for Mathematics and Its Applications at the Australian National University. The author is grateful to Peter Hall for encouragement and for a number of helpful discussions, and to the referees and an Associate Editor for their helpful comments.

REFERENCES

AKRITAS, M. G. (1992). Bootstrapping the nearest neighbor estimator of a bivariate distribution function. Unpublished manuscript.

ARNOLD, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

BERAN, R. (1977). Estimating a distribution function. *Ann. Statist.* **5** 400–404.

BERAN, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

CHENG, P. E. (1989). Nonparametric estimation of survival curve under dependent censoring. *J. Statist. Plann. Inference* **23** 181–192.

DABROWSKA, D. (1987). Non-parametric regression with censored survival time data. *Scand. J. Statist.* **14** 181–197.

DABROWSKA, D. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489.

DABROWSKA, D. (1989a). Kaplan–Meier estimate on the plane: weak convergence, LIL, and the bootstrap. *J. Multivariate Anal.* **29** 308–325.

DABROWSKA, D. (1989b). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *Ann. Statist.* **17** 1157–1167.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

GILL, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–339.

HORVATH, L. and YANDELL, B. S. (1988). Asymptotics of conditional empirical processes. *J. Multivariate Anal.* **26** 184–206.

INAGAKI, N. (1970). On the limitating distribution of a sequence of estimators with uniformity property. *Ann. Inst. Statist. Math.* **22** 1–13.

KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9** 1276–1288.

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

LEURGANS, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74** 301–309.

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.

MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.

NOLAN, D. and POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** 1291–1298.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* IMS, Hayward, California.

PRUITT, R. C. (1993). Identifiability of bivariate survival curves from censored data. *J. Amer. Statist. Assoc.* **88** 573–579.

ROUSSAS, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge Univ. Press.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistic.* Wiley, New York.

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

STUTE, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *Ann. Statist.* **12** 917–926.

STUTE, W. (1986). Conditional empirical processes. *Ann. Statist.* **14** 638–647.

STUTE, W. (1992). Consistent estimation under random censorship when covariables are present. Unpublished manuscript.

VAN DER LAAN, M. J. (1992). Efficient estimator of the bivariate survival function for right censored data. Preprint, Dept. Mathematics, Univ. Utrecht.

WELLNER, J. A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist.* **10** 595–602.

ZHOU, M. (1992). Asymptotic normality of the 'synthetic data' regression estimator for censored data. *Ann. Statist.* **20** 1002–1021.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
326 CLASSROOM BUILDING
UNIVERSITY PARK, PENNSYLVANIA 16802