The Annals of Statistics
1999, Vol. 27, No. 5, 1745–1784

TRANSFER OF TAIL INFORMATION IN CENSORED REGRESSION MODELS

By Ingrid Van Keilegom and Michael G. Akritas

Limburgs Universitair Centrum and Penn State University

Consider a heteroscedastic regression model $Y = m(X) + \sigma(X)\varepsilon$, where the functions m and σ are "smooth," and ε is independent of X. The response variable Y is subject to random censoring, but it is assumed that there exists a region of the covariate X where the censoring of Y is "light." Under this condition, it is shown that the assumed nonparametric regression model can be used to transfer tail information from regions of light censoring to regions of heavy censoring. Crucial for this transfer is the estimator of the distribution of ε based on nonparametric regression residuals, whose weak convergence is obtained. The idea of transferrring tail information is applied to the estimation of the conditional distribution of Y given $X = x$ with information on the upper tail "borrowed" from the region of light censoring, and to the estimation of the bivariate distribution $P(X \le x, Y \le y)$ with no regions of undefined mass. The weak convergence of the two estimators is obtained. By-products of this investigation include the uniform consistency of the conditional Kaplan–Meier estimator and its derivative, the location and scale estimators and the estimators of their derivatives.

1. Introduction. Let (X, Y) be a random vector where Y denotes a possible transformation of the variable of interest and X is a covariate. Let C be a "censoring" random variable which is conditionally independent of Y given X, and suppose that the observable random vector is (X, Z, Δ), where $Z = Y \wedge C$, and $\Delta = I(Y \le C)$. Finally, let (X_i, Z_i, Δ_i), $i = 1, \ldots, n$, denote independent replications of (X, Z, Δ).

In this context, nonparametric estimation of the conditional distribution $F(y \mid x)$ of Y given $X = x$ was introduced by Beran (1981), and studied by several authors [Dabrowska (1989), McKeague and Utikal (1990), Gonzalez Manteiga and Cadarso Suarez (1994), Akritas (1994), and Van Keilegom and Veraverbeke (1996, 1997a, b)]. As with the ordinary Kaplan–Meier estimator, the tails of the Beran estimator may contain little information if the censoring is "heavy." In particular, the Beran estimator cannot estimate $F(y \mid x)$ for y greater than the upper bound of the support of the conditional distribution of the censoring variable given $X = x$. This is due to the inherent lack of information and cannot be overcome in a completely general setting. Likewise all existing estimators of the bivariate distribution of (X, Y) [e.g., Dabrowska (1988), Akritas (1994), Stute (1993, 1996), van der Laan

Received April 1997; revised July 1999.
AMS 1991 *subject classifications*. Primary 62G05, 62G30, 62H12, 62J05.
Key words and phrases. Asymptotic representation, bivariate distribution, conditional distribution, nonparametric regression residuals, residual distribution, right censoring, weak convergence.

(1996)] have regions of unassigned mass when censoring in the upper tails is heavy.

In this paper we propose new estimators for $F(y \mid x)$ and for the bivariate distribution of (X, Y) which, under some additional assumptions, do not suffer from the aforementioned disadvantages. In particular, the paper proposes a method for estimating the conditional distribution $F(y \mid x)$ in such a way that tail information available in $F(y \mid x)$, with x in regions of light censoring, is transferred to the tails of $F(y \mid x)$ for x in regions of heavy censoring. The problem of transferring tail information was posed in Section 6.1 of Akritas (1994) in the context of estimating a bivariate distribution and extension of the least squares estimator to censored data; see Van Keilegom (1998) for the application to least squares estimation in the present context. To make this transfer possible, it is necessary to assume that (X, Y) follows the nonparametric regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where ε is independent of X, the function $m(\cdot)$ is the unknown regression curve and $\sigma(\cdot)$ is a conditional scale functional representing possible heteroscedasticity. This model assures that the upper tail of $F(y \mid x)$ is, up to a scale factor, the same for all x, which is the weakest possible condition under which the transfer of tail information can be accomplished. In particular, model (1.1) implies that

$$F(y \mid x) = F_e\left(\frac{y - m(x)}{\sigma(x)}\right), \tag{1.2}$$

where F_e denotes the distribution of the error variable ε. Let $\hat{m}(x)$ and $\hat{\sigma}(x)$ be nonparametric estimators for $m(x)$ and $\sigma(x)$ and $\hat{F}_e$ be the Kaplan–Meier estimator obtained from the censored nonparametric regression residuals $((Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i), \Delta_i)$, $i = 1, \ldots, n$. Then relation (1.2) suggests

$$\hat{F}(y \mid x) = \hat{F}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right) \tag{1.3}$$

as an estimator of $F(y \mid x)$. To see why this has the aforementioned advantage over the Beran estimator, suppose that for a certain region R of x-values the censoring mechanism allows the right tail of $F(y \mid x)$ to be "well estimated" for all x in R. This implies that the right tail of F_e is "well estimated" by the right tail of $\hat{F}_e$. In view of (1.3), it follows that the right tail of $F(y \mid x)$ is "well estimated" for any x.

The scenario of different degrees of censoring for different regions of the covariates arises often in practice. Consider the typical case in which the covariate is a strong prognostic variable for the event of interest. Then, in regions of the covariate indicating high risk, the amount of censoring is considerably lower than in regions indicating low risk, as here extended follow-up is needed to observe the event under study. In this situation, tail information in the high risk region can be carried over to the low risk region.

To estimate the bivariate distribution of (X, Y) without having regions of unassigned mass, we use the relation

$$F(x, y) = \int_{-\infty}^{x} F(y \mid t)\, dF_X(t), \tag{1.4}$$

where F_X is the marginal distribution function of X. Relations (1.3) and (1.4) suggest

$$\hat{F}(x, y) = \int_{-\infty}^{x} \hat{F}_e\left(\frac{y - \hat{m}(t)}{\hat{\sigma}(t)}\right) d\hat{F}_X(t) \tag{1.5}$$

as an estimator for $F(x, y)$, where $\hat{F}_X(t)$ is the empirical distribution function of the covariate X. This method for estimating the bivariate distribution was proposed in Akritas (1994), but with the Beran estimator replacing the present estimator $\hat{F}(y \mid t)$. This estimating procedure is not only reasonable, but also efficient [Akritas (1994)].

The paper is organized as follows. In the next section we give the precise definition of the estimators of the residual distribution, the conditional distribution of Y given X and the bivariate distribution of (X, Y), and we state the main assumptions under which the results will be derived. Section 3 describes the main results. A number of related results regarding the asymptotic properties of the Beran estimator, the estimators $\hat{m}(x)$, $\hat{\sigma}(x)$ and their derivatives, all of which are of independent interest and are needed for the proofs of the main results, are stated in Section 4. The main steps of the proofs are given in Section 5 while the detailed derivations are deferred to appendices. In particular, Appendix A contains proofs of results needed for the weak convergence of the Kaplan–Meier process based on nonparametric regression residuals, and results needed for the weak convergence of the bivariate distribution of (X, Y) are proved in Appendix B. The present work is part of Van Keilegom (1998) where more detailed derivations can be found.

2. Definitions and assumptions.

Consider the random vector (X, Y) satisfying the nonparametric regression model (1.1), where the smooth functions m and σ are assumed to be, respectively, a location and scale functional [as defined in, e.g., Huber (1981), pages 59, 202]. Let C, Z and Δ be as defined in Section 1 and let $F(y \mid x) = P(Y \le y \mid x)$, $G(y \mid x) = P(C \le y \mid x)$, $H(y \mid x) = P(Z \le y \mid x)$, $H_1(y \mid x) = P(Z \le y, \Delta = 1 \mid x)$ and $F_X(x) = P(X \le x)$. The assumed independence of Y and C for given X implies that $1 - H(y \mid x) = (1 - F(y \mid x))(1 - G(y \mid x))$. Further, denote $F_e(y) = P(\varepsilon \le y) = P((Y - m(X))/\sigma(X) \le y)$, $G_e(y) = P((C - m(X))/\sigma(X) \le y)$ and for $E = (Z - m(X))/\sigma(X)$ we use the notation $H_e(y) = P(E \le y)$, $H_{e1}(y) = P(E \le y, \Delta = 1)$, $H_e(y \mid x) = P(E \le y \mid x)$ and $H_{e1}(y \mid x) = P(E \le y, \Delta = 1 \mid x)$. The probability density functions of the distributions defined above will be denoted with lower case letters.

When no underlying structure is imposed on the random vector (X, Y), nonparametric estimation of the conditional distribution of Y given X is usually done by the conditional Kaplan–Meier estimator introduced by Beran (1981),

$$\tilde{F}(y \mid x) = 1 - \prod_{Z_i \le y,\, \Delta_i = 1} \left\{ 1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \ge Z_i) W_j(x, a_n)} \right\}, \tag{2.1}$$

where $W_i(x, a_n)$ are the Nadaraya–Watson weights,

$$W_i(x, a_n) = \frac{K((x - X_i)/a_n)}{\sum_{j=1}^n K((x - X_j)/a_n)},$$

with K a known probability density function (kernel) and $\{a_n\}$ a sequence of positive constants tending to zero as n tends to infinity, called a bandwidth sequence. This estimator reduces to the usual Kaplan–Meier (1958) estimator when all weights $W_i(x, a_n)$ equal n^{-1}. On the other hand, in the case of no censoring ($Z_i = Y_i, \Delta_i = 1$), the estimator equals the kernel estimator of Stone (1977). The estimator $\tilde{F}(y \mid x)$, however, shares the same drawback as the ordinary Kaplan–Meier estimator in that it often does not behave well in the right tail. By imposing a slight restriction of generality in the form of model (1.1), we propose an alternative estimator for $F(y \mid x)$ which will, in certain situations, not suffer from this drawback. This proposed estimator is based on relation (1.2) which follows from (1.1).

The first step will be to estimate the error distribution function F_e. This will be accomplished by estimating m and σ and using the Kaplan–Meier estimator on the censored residuals. To estimate m and σ we will work with the particular definitions,

$$m(x) = \int_0^1 F^{-1}(s \mid x) J(s)\, ds, \qquad \sigma^2(x) = \int_0^1 F^{-1}(s \mid x)^2 J(s)\, ds - m^2(x), \tag{2.2}$$

where $F^{-1}(s \mid x) = \inf\{t; F(t \mid x) \ge s\}$ is the quantile function of Y given x and $J(s)$ is a given score function satisfying $\int_0^1 J(s)\, ds = 1$. Note that if the assumed independence of ε and X holds for certain location and scale functionals then it holds for all location and scale functionals. Hence, working with the functionals $m(x)$ and $\sigma^2(x)$ in (2.2) is no restriction of generality. Replacing $F(s \mid x)$ with the Beran estimator $\tilde{F}(s \mid x)$ in these expressions yields

$$\hat{m}(x) = \int_0^1 \tilde{F}^{-1}(s \mid x) J(s)\, ds, \qquad \hat{\sigma}^2(x) = \int_0^1 \tilde{F}^{-1}(s \mid x)^2 J(s)\, ds - \hat{m}^2(x) \tag{2.3}$$

as estimators of $m(x)$ and $\sigma(x)$. The score function J will be chosen so that $\hat{m}(x)$ and $\hat{\sigma}(x)$ are consistent even if the tails of the Beran estimator are not consistent. Set $\hat{E}_i = (Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$ for the resulting censored residuals, and let

$$\hat{F}_e(y) = 1 - \prod_{\hat{E}_{(i)} \le y,\, \Delta_{(i)} = 1} \left(1 - \frac{1}{n - i + 1} \right), \tag{2.4}$$

where $\hat{E}_{(i)}$ is the ith order statistic of $\hat{E}_1, \ldots, \hat{E}_n$ and $\Delta_{(i)}$ is the corresponding censoring indicator, denote the proposed estimator of F_e. To our knowledge, the asymptotic properties of the Kaplan–Meier estimator based on residuals from nonparametric regression have not been considered in the literature. Relation (1.2) leads to (1.3) as an estimator of $F(y \mid x)$. As was explained in the introduction, the bivariate distribution of (X, Y) given in (1.4) is now estimated by (1.5) where $\hat{F}_X(x) = n^{-1} \sum_{i=1}^n I(X_i \le x)$. Because the asymptotic theory is based on the i.i.d. representation for $\hat{F}_e$ which is valid up to any point smaller than $\tau_{F_e} \wedge \tau_{G_e}$ (where $\tau_F = \inf\{y; F(y) = 1\}$ denotes the upper bound of the support of any distribution function F), we need to to work with a slightly modified version of (1.5). Namely the asymptotic theory developed in Section 3, pertains to

$$\hat{F}_T(x, y) = \int_{-\infty}^{x} \hat{F}_e\left(\frac{y \wedge T_t - \hat{m}(t)}{\hat{\sigma}(t)}\right) d\hat{F}_X(t), \tag{2.5}$$

where $T_t \le T\sigma(t) + m(t)$ for $t \in R_X$ and where $T < \tau_{H_e}$. This is actually an estimator of $F_T(x, y) = \int_{-\infty}^{x} F_e((y \wedge T_t - m(t))/\sigma(t))\, dF_X(t)$, which can become arbitrarily close to $F(x, y)$ if $\tau_{F_e} \le \tau_{G_e}$ and T_t, respectively, T, is chosen sufficiently close to $T\sigma(t) + m(t)$, respectively, τ_{H_e}, for all t.

The primary objective of this paper is to study the asymptotic behavior of the estimators in (2.4), (1.3) and (2.5). The following functions enter in the asymptotic representation for these estimators, which we will establish in the next section:

$$\begin{aligned}
\xi_e(z, \delta, y) &= (1 - F_e(y))\left\{-\int_{-\infty}^{y \wedge z} \frac{dH_{e1}(s)}{(1 - H_e(s))^2} + \frac{I(z \le y, \delta = 1)}{1 - H_e(z)}\right\},\\
\xi(z, \delta, y \mid x) &= (1 - F(y \mid x))\left\{-\int_{-\infty}^{y \wedge z} \frac{dH_1(s \mid x)}{(1 - H(s \mid x))^2} + \frac{I(z \le y, \delta = 1)}{1 - H(z \mid x)}\right\},\\
\eta(z, \delta \mid x) &= \int_{-\infty}^{+\infty} \xi(z, \delta, v \mid x) J(F(v \mid x))\, dv\, \sigma^{-1}(x),\\
\zeta(z, \delta \mid x) &= \int_{-\infty}^{+\infty} \xi(z, \delta, v \mid x) J(F(v \mid x)) \frac{v - m(x)}{\sigma(x)}\, dv\, \sigma^{-1}(x),\\
\gamma_1(y \mid x) &= \int_{-\infty}^{y} \frac{h_e(s \mid x)}{(1 - H_e(s))^2}\, dH_{e1}(s) + \int_{-\infty}^{y} \frac{dh_{e1}(s \mid x)}{1 - H_e(s)},\\
\gamma_2(y \mid x) &= \int_{-\infty}^{y} \frac{s h_e(s \mid x)}{(1 - H_e(s))^2}\, dH_{e1}(s) + \int_{-\infty}^{y} \frac{d(s h_{e1}(s \mid x))}{1 - H_e(s)}.
\end{aligned}$$

Finally, $H_e(y)$ and $H_{e1}(y)$ are estimated by the empirical distribution functions based on the residuals $\hat{E}_i$,

$$\hat{H}_e(y) = n^{-1} \sum_{i=1}^{n} I(\hat{E}_i \le y) \quad \text{and} \quad \hat{H}_{e1}(y) = n^{-1} \sum_{i=1}^{n} I(\hat{E}_i \le y, \Delta_i = 1),$$

$H(y \mid x)$ and $H_1(y \mid x)$ are estimated by the conditional empirical distribution functions [Stone (1977)],

$$\hat{H}(y \mid x) = \sum_{i=1}^{n} W_i(x, a_n) I(Z_i \le y)$$

and

$$\hat{H}_1(y \mid x) = \sum_{i=1}^{n} W_i(x, a_n) I(Z_i \le y, \Delta_i = 1),$$

and the density $f_X(x)$ is estimated by $\hat{f}_X(x) = (na_n)^{-1} \sum_{i=1}^n K((x - X_i)/a_n)$.

The assumptions we need for the proofs of the main results are listed below for convenient reference.

(A1) (i) The sequence a_n satisfies $na_n^5 (\log a_n^{-1})^{-1} = O(1)$ and $na_n^{3+2\delta} (\log a_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.

(ii) The sequence a_n satisfies $na_n^4 \to 0$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.

(iii) The support R_X of X is bounded, convex and its interior is not empty.

(iv) The probability density function K has compact support, $\int uK(u)\,du = 0$ and K is twice continuously differentiable.

Let $\tilde{T}_x$ be any value less than the upper bound of the support of $H(\cdot \mid x)$ such that $\inf_{x \in R_X}(1 - H(\tilde{T}_x \mid x)) > 0$.

(A2) (i) There exist $0 \le s_0 \le s_1 \le 1$ such that $s_1 \le \inf_x F(\tilde{T}_x \mid x)$, $s_0 \le \inf\{s \in [0,1]; J(s) \ne 0\}$, $s_1 \ge \sup\{s \in [0,1]; J(s) \ne 0\}$ and $\inf_{x \in R_X} \inf_{s_0 \le s \le s_1} f(F^{-1}(s \mid x) \mid x) > 0$.

(ii) J is twice continuously differentiable, $\int_0^1 J(s)\,ds = 1$ and $J(s) \ge 0$ for all $0 \le s \le 1$.

(iii) The function $x \to T_x (x \in R_X)$ is twice continuously differentiable.

(A3) (i) The distribution F_X is thrice continuously differentiable and $\inf_{x \in R_X} f_X(x) > 0$.

(ii) The functions m and σ are twice continuously differentiable and $\inf_{x \in R_X} \sigma(x) > 0$.

(iii) The error variable ε has finite expectation.

(A4) (i) The functions $\eta(z, \delta \mid x)$ and $\zeta(z, \delta \mid x)$ are twice continuously differentiable with respect to x and their first and second derivatives (with respect to x) are bounded, uniformly in $x \in R_X$, $z < \tilde{T}_x$ and δ.

(ii) The first derivatives of $\eta(z, \delta \mid x)$ and $\zeta(z, \delta \mid x)$ with respect to z are of bounded variation and the variation norms are uniformly bounded over all x.

(A5) The function $y \to P(m(X) + e\sigma(X) \le y)$ $(y \in \mathbb{R})$ is differentiable for all $e \in \mathbb{R}$ and the derivative is uniformly bounded over all $e \in \mathbb{R}$.

For a (sub)distribution function $L(y \mid x)$ we will use the notations $l(y \mid x) = L'(y \mid x) = (\partial/\partial y) L(y \mid x)$, $\dot{L}(y \mid x) = (\partial/\partial x) L(y \mid x)$ and similar notations will be used for higher order derivatives. [In the proofs, the function $L(y \mid x)$ of assumption (A6) will be either $H(y \mid x)$, $H_1(y \mid x)$, $H_e(y \mid x)$ or $H_{e1}(y \mid x)$.]

(A6) (i) $L(y \mid x)$ is continuous.
(ii) $L'(y \mid x) = l(y \mid x)$ exists, is continuous in (x, y) and $\sup_{x,y} |yL'(y \mid x)| < \infty$.
(iii) $L''(y \mid x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y^2 L''(y \mid x)| < \infty$.
(iv) $\dot{L}(y \mid x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y\dot{L}(y \mid x)| < \infty$.
(v) $\ddot{L}(y \mid x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y^2 \ddot{L}(y \mid x)| < \infty$.
(vi) $\dot{L}'(y \mid x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y\dot{L}'(y \mid x)| < \infty$.
(vii) $\ddot{L}'(y \mid x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y\ddot{L}'(y \mid x)| < \infty$.
(viii) $L'''(y \mid x)$ exists, is continuous in (x, y) and $\sup_{x,y} |y^3 L'''(y \mid x)| < \infty$.

3. Main Results. The results for $\hat{F}_e(y)$, $\hat{F}(y \mid x)$ and $\hat{F}(x, y)$ are given separately.

3.1. *Asymptotics for the Kaplan–Meier estimator based on nonparametric regression residuals.* The results of this section extend the classical results of Durbin (1973) and Loynes (1980) concerning the weak convergence of the empirical distribution function when parameters are estimated, to the present nonparametric regression setting with censored data. The weak convergence of the Kaplan–Meier process $n^{1/2}(\hat{F}_e(y) - F_e(y))$ follows from a Lo and Singh (1986)-type asymptotic representation, which we give first.

THEOREM 3.1. *Assume* (A1), (A2) (i), (ii), (A3) (i), (ii), (A4) (i), $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i)–(vi), *and* $H_e(y \mid x)$ *and* $H_{e1}(y \mid x)$ *satisfy* (A6) (ii), (iii), (vi), (vii). *Then,*

$$\hat{F}_e(y) - F_e(y) = n^{-1} \sum_{i=1}^{n} \varphi(X_i, Z_i, \Delta_i, y) + R_n(y),$$

where $\sup\{|R_n(y)|; -\infty < y \le T\} = o_P(n^{-1/2})$ *and, with* $S_e = 1 - F_e$,

$$\varphi(x, z, \delta, y) = \xi_e\left(\frac{z - m(x)}{\sigma(x)}, \delta, y\right) - S_e(y)\eta(z, \delta \mid x)\gamma_1(y \mid x) - S_e(y)\zeta(z, \delta \mid x)\gamma_2(y \mid x).$$

Note that the i.i.d representation of the usual Kaplan–Meier estimator due to Lo and Singh (1986) contains only the term ξ_e. The extra term in the representation above is caused by the fact that we replaced $(Z_i - m(X_i))/\sigma(X_i)$ with $(Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$. We continue with the statement of the weak convergence result for the process $n^{1/2}(\hat{F}_e(\cdot) - F_e(\cdot))$.

COROLLARY 3.2. *Under the assumptions of Theorem* 3.1, *the process* $n^{1/2}(\hat{F}_e(y) - F_e(y))$, $-\infty < y \le T$ *converges weakly to a zero-mean Gaussian process* $Z(y)$ *with covariance function*

$$\mathrm{Cov}(Z(y), Z(y')) = E(\varphi(X, Z, \Delta, y)\varphi(X, Z, \Delta, y')).$$

3.2. *Asymptotics for the estimator* $\hat{F}(y \mid x)$. The results of Theorem 3.1 and Corollary 3.2 enable us to show an asymptotic representation and the weak convergence of $\hat{F}(y \mid x)$.

THEOREM 3.3. *Assume* (A1), (A2) (i), (ii), (A3) (i) (ii), (A4) (i), $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i)–(vi), *and* $H_e(y \mid x)$ *and* $H_{e1}(y \mid x)$ *satisfy* (A6) (ii), (iii), (vi), (vii). *Then,*

$$\begin{aligned}\hat{F}(y \mid x) - F(y \mid x) &= \hat{F}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - m(x)}{\sigma(x)}\right)\\ &= (na_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{a_n}\right) h_{x,y}(Z_i, \Delta_i) + R_n(x, y),\end{aligned}$$

where $\sup\{|R_n(x, y)|; (x, y) \in \Omega\} = o_P((na_n)^{-1/2})$, $\Omega = \{(x, y);\ x \in R_X,\ (y - m(x))/\sigma(x) \le T\}$ *and*

$$h_{x,y}(z, \delta) = \left[\eta(z, \delta \mid x) + \zeta(z, \delta \mid x)\frac{y - m(x)}{\sigma(x)}\right] f_e\left(\frac{y - m(x)}{\sigma(x)}\right) f_X^{-1}(x).$$

COROLLARY 3.4. *Under the assumption of Theorem* 3.3, *the process* $(na_n)^{1/2}$ $(\hat{F}(y \mid x) - F(y \mid x))$, $x \in R_X$ *fixed*, $(y - m(x))/\sigma(x) \le T$ *converges weakly to a zero-mean Gaussian process* $Z(y \mid x)$ *with covariance function*

$$\begin{aligned}&\operatorname{Cov}\big(Z(y \mid x), Z(y' \mid x)\big)\\ &\quad = f_X(x) \int K^2(u)\, du \operatorname{Cov}\big(h_{x,y}(Z, \Delta), h_{x,y'}(Z, \Delta) \mid X = x\big).\end{aligned}$$

REMARK 3.1. Condition (A1) (ii) is needed for the proof of Theorem 3.1 which is used in the proof of Theorem 3.3. However, the proof of Theorem 3.3 requires only that the remainder term in Theorem 3.1 satisfies $\sup_{y \le T} |R_n(y)| = o_p((na_n)^{-1/2})$. This rate can be obtained by using $na_n^5 \to 0$ instead of $na_n^4 \to 0$ of assumption (A1) (ii).

REMARK 3.2. It is easily seen that in the uncensored case and when $J(s) \equiv 1$ [that is, $m(x) = E(Y \mid x)$ and $\sigma^2(x) = \operatorname{Var}(Y \mid X)$], the asymptotic variance function of the process $(na_n)^{1/2}(\hat{F}(\cdot \mid x) - F(\cdot \mid x))$ reduces to

$$\begin{aligned}&f_X^{-1}(x) \int K^2(u)\, du\, f_e^2\left(\frac{y - m(x)}{\sigma(x)}\right)\\ &\qquad \times \left[\frac{1}{4}\operatorname{Var}(\varepsilon^2)\left(\frac{y - m(x)}{\sigma(x)}\right)^2 + E(\varepsilon^3)\frac{y - m(x)}{\sigma(x)} + 1\right],\end{aligned}$$

while the asymptotic variance of the process $(na_n)^{1/2} \sum_{i=1}^{n} W_i(x, a_n)(I(Y_i \le y) - F(y \mid x))$, which corresponds to the usual kernel estimator $\tilde{F}(y \mid x)$, is $f_X^{-1}(x) \int K^2(u)\, du\, F(y \mid x)(1 - F(y \mid x))$. Figure 1 illustrates that when the error distribution is normal and there is no censoring, the present estimator

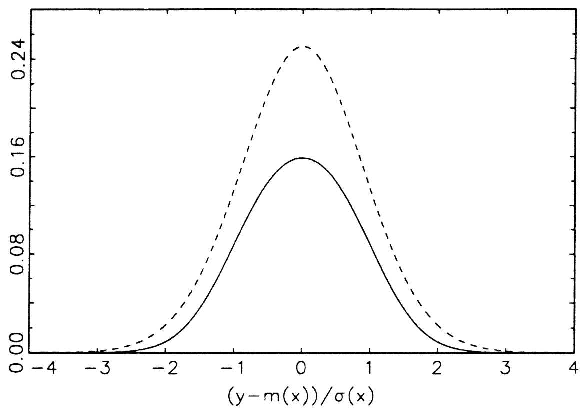

FIG. 1. *Graph of the variance of function of* $\hat{F}(y\,|\,x)$ *(full line) and* $\tilde{F}(y\,|\,x)$ *(dashed line) divided by their common factor* $f_X^{-1}(x)\int K^2(u)\,du$ *for normal survival times and in case of no censoring.*

is more efficient and should be used when model (1.1) can be assumed. [The plot does not include the common component of the variance functions; i.e., they correspond to the variance functions divided by $f_X^{-1}(x)\int K^2(u)\,du$.]

However, even though the estimator $\hat{F}(y\,|\,x)$ is constructed under model (1.1) whereas the estimator $\tilde{F}(y\,|\,x)$ does not require any model assumptions, the variance of $\hat{F}(y\,|\,x)$ is not always smaller than that of $\tilde{F}(y\,|\,x)$. This is because $\hat{F}(y\,|\,x)$ involves the estimation of $m(x)$ and $\sigma^2(x)$, which introduces extra variability in the estimator. We note that the primary consideration in choosing the score function J is to achieve uniform consistency of $\hat{m}(x)$ and $\hat{\sigma}^2(x)$, and that this objective is incompatible with efficient estimation of $m(x)$ and $\sigma^2(x)$ when the degree of censoring varies with x.

To illustrate that the variance of $\hat{F}(y\,|\,x)$ is not always smaller than that of $\tilde{F}(y\,|\,x)$ we consider the following distributions for the survival and censoring times:

$$(Y\,|\,X=x)\sim \mathrm{Exp}(\lambda_x),\qquad (C\,|\,X=x)\sim \mathrm{Exp}(\mu_x),$$

for certain choices of the parameters λ_x and μ_x. We use $J(s)=b^{-1}I(0\le s\le b)$, where $b\in(0,1)$ is chosen in such a way that $m(x)$ and $\sigma^2(x)$ can be estimated consistently (see below for the actual choices). [We do realize that this indicator function does not satisfy the smoothness conditions imposed in assumption (A2) (i), however (a) the indicator score function greatly simplifies the (long) calculations, and (b) it can be approximated arbitrarily well by a score function that does satisfy the conditions, meaning that similar results can be obtained from calculations with an appropriate score function]. In Figure 2 we show the variance of the Beran estimator $\tilde{F}(y\,|\,x)$ and the new estimator $\hat{F}(y\,|\,x)$ for $\lambda_x=1$ and for μ_x determined in such a way that the probability of censoring (given by $\mu_x/(\lambda_x+\mu_x)$) is, respectively, 0.2, 0.4, 0.6 and 0.8. For determining the endpoint b of the score function J, we first note that for a given sample, the value of b should be smaller than (or equal to) $\max_y \hat{F}(y \mid x)$ [otherwise $\hat{m}(x)$ is not defined]. Therefore, we calculated this maximum for 1000 independent samples of size 50 and defined b as the average of these maxima. In this way we obtain $b = 0.98, 0.93, 0.81$ and 0.52 for the situation where the probability of censoring is, respectively, 0.2, 0.4, 0.6 and 0.8. Figure 2 shows the variance of $\tilde{F}(y \mid x)$ and $\hat{F}(y \mid x)$ for $y \in [0, F^{-1}(0.99 \mid x)]$ (except for Figure 2d where we restricted to the interval $[0, F^{-1}(0.85 \mid x)]$ since, in that case, the variance of $\tilde{F}(y \mid x)$ increases extremely fast as y tends to infinity). From the figures it follows that the variance of the new estimator $\hat{F}(y \mid x)$ is not always smaller than that of the Beran estimator $\tilde{F}(y \mid x)$. However, we can still say that $\hat{F}(y \mid x)$ behaves better than $\tilde{F}(y \mid x)$ on the average. This is especially so for cases with heavy censoring (see Figure 2c and 2d) since in these cases the variance of the Beran estimator tends to infinity (this is so for all cases where the probability of censoring is larger than 0.5).

Finally we note that simulation studies comparing the finite sample performance of $\tilde{F}(y \mid x)$ and $\hat{F}(y \mid x)$ have been carried out in Van Keilegom, Akritas

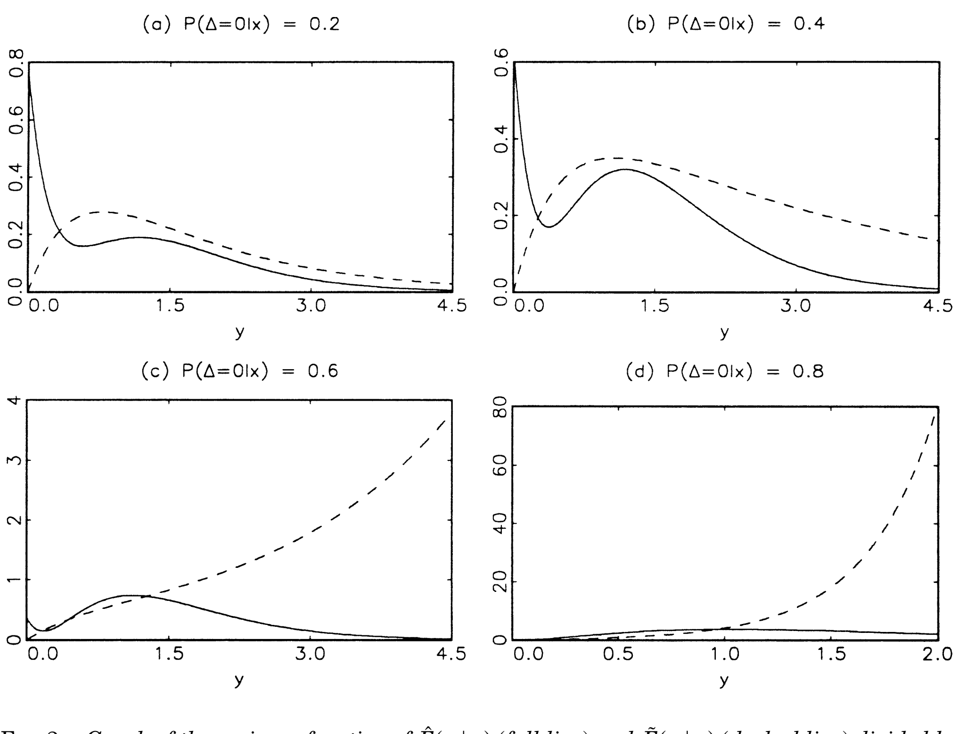

FIG. 2. *Graph of the variance function of* $\hat{F}(y \mid x)$ *(full line) and* $\tilde{F}(y \mid x)$ *(dashed line) divided by their common factor* $f_X^{-1}(x) \int K^2(u)\, du$ *for exponential survival and censoring times.*

and Veraverbeke (1998). As these simulations indicate, the estimator $\hat{F}(y \mid x)$ outperforms (on average) $\tilde{F}(y \mid x)$ also for finite sample sizes.

3.3. *Asymptotics for the estimator* $\hat{F}(y, x)$.

THEOREM 3.5. *Assume* (A1)–(A5), $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i)–(vi), *and* $H_e(y \mid x)$ *and* $H_{e1}(y \mid x)$ *satisfy* (A6) (ii), (iii), (vi), (vii), (viii). *Then,*

$$\hat{F}_T(x, y) - F_T(x, y) = n^{-1} \sum_{i=1}^{n} g_{x,y}(X_i, Z_i, \Delta_i) + R_n(x, y),$$

where $\sup\{|R_n(x, y)|;\ x \in R_X,\ y \in \mathbb{R}\} = o_P(n^{-1/2})$, *and* $g_{x,y}(t, z, \delta) = \sum_{i=1}^{3} g_{x,y}^i(t, z, \delta)$, *where*

$$g_{x,y}^1(t, z, \delta) = E\left\{\varphi\left(t, z, \delta, \frac{y \wedge T_X - m(X)}{\sigma(X)}\right) I(X \le x)\right\},$$

$$g_{x,y}^2(t, z, \delta) = \left[\eta(z, \delta \mid t) + \zeta(z, \delta \mid t) \frac{y \wedge T_t - m(t)}{\sigma(t)}\right] \times f_e\left(\frac{y \wedge T_t - m(t)}{\sigma(t)}\right) I(t \le x),$$

$$g_{x,y}^3(t, z, \delta) = F_e\left(\frac{y \wedge T_t - m(t)}{\sigma(t)}\right) I(t \le x) - E\left[F_e\left(\frac{y \wedge T_X - m(X)}{\sigma(X)}\right) I(X \le x)\right].$$

COROLLARY 3.6. *Under the assumptions of Theorem* 3.5, *the process* $n^{1/2} (\hat{F}_T(x, y) - F_T(x, y))$, $x \in R_X$, $y \in \mathbb{R}$, *converges weakly to a zero-mean Gaussian process* $Z(x, y)$ *with covariance function*

$$\begin{aligned} &\operatorname{Cov}(Z(x, y), Z(x', y')) \\ &\quad = E\{[g_{x,y}^1(X, Z, \Delta) + g_{x,y}^2(X, Z, \Delta)][g_{x',y'}^1(X, Z, \Delta) + g_{x',y'}^2(X, Z, \Delta)]\} \\ &\qquad + \operatorname{Cov}\left[F_e\left(\frac{y \wedge T_X - m(X)}{\sigma(X)}\right) I(X \le x), F_e\left(\frac{y' \wedge T_X - m(X)}{\sigma(X)}\right) I(X \le x')\right]. \end{aligned}$$

4. Some auxiliary results. The asymptotic representation and weak convergence results stated in the previous section require some results concerning the location estimator $\hat{m}(x)$ and the scale estimator $\hat{\sigma}(x)$ which are stated below as propositions. In particular, the uniform consistency of $\hat{m}(x)$, $\hat{m}'(x)$ and $(\hat{m}'(x) - \hat{m}'(x'))/(x - x')^{\delta}$ and their analogues for $\hat{\sigma}(x)$ will be established, as well as asymptotic representations for $\hat{m}(x)$ and $\hat{\sigma}(x)$. In turn, these results require the uniform (both in y and x) consistency of the Beran estimator and its derivative. The proofs of these results can be found in Van Keilegom and Akritas (1998) and Van Keilegom (1998).

In the uncensored case, the consistency of derivatives of a regression function has been studied by Schuster and Yakowitz (1979), Gasser and Müller (1984), Härdle and Gasser (1985), Mack and Müller (1989), Rychlik (1990), among others. The uniform consistency of a regression function was the object of study in Devroye (1978),Schuster and Yakowitz (1979), Müller and Stadtmüller (1987) and Härdle, Janssen and Serfling (1988). In the latter paper strong uniform consistency rates are established for kernel type estimators of functionals of the conditional distribution function, under general conditions.

LEMMA 4.1. *Assume* (A1) (i), (iii), (iv), F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y\,|\,x)$ *satisfies* (A6) (i), (iv), (v). *Then*

$$\sup_{x\in R_X}\sup_{-\infty<y<\infty}|\hat H(y\,|\,x)-H(y\,|\,x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\quad a.s.,$$

$$\sup_{x\in R_X}\sup_{-\infty<y<\infty}|\dot{\hat H}(y\,|\,x)-\dot H(y\,|\,x)| = O((na_n^3)^{-1/2}(\log a_n^{-1})^{1/2})\quad a.s.$$

LEMMA 4.2. *Assume* (A1) (i), (iii), (iv), F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y\,|\,x)$ *satisfies* (A6) (i), (iv), (v). *Then*,

$$\sup_{x,\,x',\,y}\frac{\big|\dot{\hat H}(y\,|\,x)-\dot H(y\,|\,x)-\dot{\hat H}(y\,|\,x')+\dot H(y\,|\,x')\big|}{|x-x'|^\delta}$$
$$= O\big((na_n^{3+2\delta})^{-1/2}(\log a_n^{-1})^{1/2}\big)\quad a.s.,$$

where $\delta>0$ *is as in assumption* (A1).

PROPOSITION 4.3. *Assume* (A1) (i), (iii), (iv), F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y\,|\,x)$ *and* $H_1(y\,|\,x)$ *satisfy* (A6) (i), (iv), (v). *Then*

$$\sup_x\sup_{y\le\tilde T_x}|\tilde F(y\,|\,x)-F(y\,|\,x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\quad a.s.,$$

$$\sup_x\sup_{y\le\tilde T_x}|\dot{\tilde F}(y\,|\,x)-\dot F(y\,|\,x)| = O((na_n^3)^{-1/2}(\log a_n^{-1})^{1/2})\quad a.s.$$

PROPOSITION 4.4. *Assume* (A1) (i), (iii), (iv), F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y\,|\,x)$ *and* $H_1(y\,|\,x)$ *satisfy* (A6) (i), (iv), (v). *Then*,

$$\sup_{x,\,x'}\sup_{y\le\tilde T_x\wedge\tilde T_{x'}}\frac{|\dot{\tilde F}(y\,|\,x)-\dot F(y\,|\,x)-\dot{\tilde F}(y\,|\,x')+\dot F(y\,|\,x')|}{|x-x'|^\delta}$$
$$= O\big((na_n^{3+2\delta})^{-1/2}(\log a_n^{-1})^{1/2}\big)\quad a.s.,$$

where $\delta>0$ *is as in assumption* (A1).

PROPOSITION 4.5. *Assume* (A1) (i), (iii), (iv), (A2) (i), J *is continuous*, $\int_0^1 J(s)\,ds = 1$, $J(s) \geq 0$ *for all* $0 \leq s \leq 1$, F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i), (iv), (v). *Then*,

$$\sup_x |\hat{m}(x) - m(x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \quad a.s.$$

If in addition $\inf_x \sigma(x) > 0$, *then*,

$$\sup_x |\hat{\sigma}(x) - \sigma(x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \quad a.s.$$

PROPOSITION 4.6. *Assume* (A1) (i), (iii), (iv), (A2) (i), J *is continuously differentiable*, $\int_0^1 J(s)\,ds = 1$, $J(s) \geq 0$ *for all* $0 \leq s \leq 1$, F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i), (iv), (v). *Then*,

$$\sup_x |\hat{m}'(x) - m'(x)| = O((na_n^3)^{-1/2}(\log a_n^{-1})^{1/2}) \quad a.s.$$

If in addition m *and* σ *are continuously differentiable and* $\inf_{x\in R_X} \sigma(x) > 0$, *then*,

$$\sup_x |\hat{\sigma}'(x) - \sigma'(x)| = O((na_n^3)^{-1/2}(\log a_n^{-1})^{1/2}) \quad a.s.$$

PROPOSITION 4.7. *Assume* (A1) (i), (iii), (iv), (A2) (i), (ii), F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i), (iv), (v). *Then*,

$$\sup_{x,\,x'} \frac{|\hat{m}'(x) - m'(x) - \hat{m}'(x') + m'(x')|}{|x - x'|^\delta} = O((na_n^{3+2\delta})^{-1/2}(\log a_n^{-1})^{1/2}) \quad a.s.$$

If in addition (A3) (ii) *holds, then*,

$$\sup_{x,\,x'} \frac{|\hat{\sigma}'(x) - \sigma'(x) - \hat{\sigma}'(x') + \sigma'(x')|}{|x - x'|^\delta} = O((na_n^{3+2\delta})^{-1/2}(\log a_n^{-1})^{1/2}) \quad a.s.,$$

where $\delta > 0$ *is as in assumption* (A1).

PROPOSITION 4.8. *Assume* (A1) (i), (iii), (iv), (A2) (i), J *is continuously differentiable*, $\int_0^1 J(s)\,ds = 1$, $J(s) \geq 0$ *for all* $0 \leq s \leq 1$, F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x) > 0$, *and* $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i)–(vi). *Then*,

$$\hat{m}(x) - m(x) = -(na_n)^{-1} f_X^{-1}(x)\sigma(x) \sum_{i=1}^n K\left(\frac{x - X_i}{a_n}\right)\eta(Z_i, \Delta_i \mid x) + R_n(x),$$

where $\sup\{|R_n(x)|; x \in R_X\} = O((na_n)^{-3/4}(\log n)^{3/4})$ *a.s.*

PROPOSITION 4.9. *Assume* (A1) (i), (iii), (iv), (A2) (i), J *is continuously differentiable*, $\int_0^1 J(s)\,ds = 1$, $J(s) \geq 0$ *for all* $0 \leq s \leq 1$, F_X *is twice continuously differentiable*, $\inf_{x \in R_X} f_X(x) > 0$, $\inf_{x \in R_X} \sigma(x) > 0$ *and* $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i)–(vi). *Then*,

$$\hat{\sigma}(x) - \sigma(x) = -(na_n)^{-1} f_X^{-1}(x)\sigma(x) \sum_{i=1}^{n} K\left(\frac{x - X_i}{a_n}\right)\zeta(Z_i, \Delta_i \mid x) + \tilde{R}_n(x),$$

where $\sup\{|\tilde{R}_n(x)|; x \in R_X\} = O((na_n)^{-3/4}(\log n)^{3/4})$ *a.s.*

5. Proofs of main results.

This section contains the proofs of the results stated in Section 3. Some technical results needed in the proof of Theorem 3.1 are deferred to Appendix A, while results needed for showing Theorem 3.5 are proved in Appendix B.

PROOF OF THEOREM 3.1. We start with

$$\begin{aligned}
&\int_{-\infty}^{y} \frac{d\hat{H}_{e1}(s)}{1 - \hat{H}_e(s)} - \int_{-\infty}^{y} \frac{dH_{e1}(s)}{1 - H_e(s)} \\
&\quad = \int_{-\infty}^{y} \left[\frac{1}{1 - \hat{H}_e(s)} - \frac{1}{1 - H_e(s)}\right] dH_{e1}(s) \\
&\qquad + \int_{-\infty}^{y} \frac{1}{1 - H_e(s)}\, d(\hat{H}_{e1}(s) - H_{e1}(s)) \\
&\qquad + \int_{-\infty}^{y} \left[\frac{1}{1 - \hat{H}_e(s)} - \frac{1}{1 - H_e(s)}\right] d(\hat{H}_{e1}(s) - H_{e1}(s)).
\end{aligned}$$

The last term on the right-hand side is $o_P(n^{-1/2})$ by Corollary A.5. Using Proposition A.3, the sum of the first and second term can be written as

$$\begin{aligned}
&\int_{-\infty}^{y} \frac{\hat{H}_e(s) - H_e(s)}{(1 - H_e(s))^2}\, dH_{e1}(s) + \int_{-\infty}^{y} \frac{1}{1 - H_e(s)}\, d(\hat{H}_{e1}(s) - H_{e1}(s)) + o_P(n^{-1/2}) \\
&\quad = n^{-1} \sum_{i=1}^{n} \varphi(X_i, Z_i, \Delta_i, y)(1 - F_e(y))^{-1} + o_P(n^{-1/2}),
\end{aligned}$$

where the equality follows by Propositions A.2 and its analogue for $\hat{H}_{e1}$. Now, write (using a Taylor expansion)

$$\begin{aligned}
&\log(1 - \hat{F}_e(y)) + \int_{-\infty}^{y} \frac{d\hat{H}_{e1}(s)}{1 - \hat{H}_e(s-)} \\
&\quad = -\frac{1}{2} \sum_{i=1}^{n} \frac{I(\hat{E}_{(i)} \leq y, \Delta_{(i)} = 1)}{(n - i + 1)^2} \frac{1}{(1 - R_i)^2} = O(n^{-1})
\end{aligned}$$

a.s., uniformly in y, where R_i is between 0 and $1/(n - i + 1)$. The result now follows by noting that $\hat{F}_e(y) - F_e(y) = -(1 - F_e(y))(\log(1 - \hat{F}_e(y)) -$

$\log(1 - F_e(y))) + o_P(n^{-1/2})$, and that

$$\log(1 - F_e(y)) = -\int_{-\infty}^{y} (1 - H_e(s))^{-1} \, dH_{e1}(s).$$

PROOF OF COROLLARY 3.2. We will make use of Theorem 2.5.6 in van der Vaart and Wellner (1996), that is, we will show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, L_2(P))} \, d\varepsilon < \infty, \tag{5.1}$$

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of (X, Z, Δ), $L_2(P)$ is the L_2-norm and $\mathscr{F} = \{\varphi(X, Z, \Delta, y); -\infty < y \le T\}$. Proving this entails that the class $\mathscr{F}$ is Donsker and hence the weak convergence of the given process follows from pages 81 and 82 in van der Vaart and Wellner's book. Since the functions $x \to S_e(y)\gamma_i(y \,|\, x)$ $(i = 1, 2)$ are bounded (uniformly in y), as well as their first derivatives, their bracketing number is $O(\exp(K\varepsilon^{-1}))$ by Corollary 2.7.2 of the aforementioned book. Hence, since $\eta(z, \delta \,|\, x)$ and $\zeta(z, \delta \,|\, x)$ are uniformly bounded, the bracketing number of the second and third terms of $\varphi(x, z, \delta, y)$ is $O(\exp(K\varepsilon^{-1}))$. For the first term, we note that the first term of $\xi_e((z - m(x))/\sigma(x), \delta, y)$ is decreasing in $(z - m(x))/\sigma(x)$. Hence, its bracketing number is $m = O(\exp(K\varepsilon^{-1}))$ by Theorem 2.7.5 in van der Vaart and Wellner (1996). Also the class of functions of the form $(z - m(x))/\sigma(x) \to (1 - F_e(y))I((z - m(x))/\sigma(x) \le y)$ needs m brackets by Theorem 2.7.5. Finally, we note that since $I(\delta = 1)(1 - H_e((z - m(x))/\sigma(x)))^{-1}$ is bounded (for $z \le T$) and independent of y, the second term of $\xi_e((z - m(x))/\sigma(x), \delta, y)$ has bracketing number m. This concludes the proof, since the integration in (5.1) can be restricted to the interval $[0, 2M]$, if $|\varphi(x, z, \delta, y)| \le M$ for all x, z, δ and y (for $\varepsilon > 2M$ we take $N_{[]}(\varepsilon, \mathscr{F}, L_2(P)) = 1$). $\square$

PROOF OF THEOREM 3.3. Write

$$\begin{aligned}
&\hat{F}(y \,|\, x) - F(y \,|\, x) \\
&\quad = \left[\hat{F}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right)\right] \\
&\qquad + \left[F_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - m(x)}{\hat{\sigma}(x)}\right)\right] \\
&\qquad + \left[F_e\left(\frac{y - m(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - m(x)}{\sigma(x)}\right)\right] \\
&\quad = \alpha_n^1(x, y) + \alpha_n^2(x, y) + \alpha_n^3(x, y).
\end{aligned}$$

We start with $\alpha_n^2(x, y)$.

$$\alpha_n^2(x, y) = -\frac{\hat{m}(x) - m(x)}{\hat{\sigma}(x)} f_e\left(\frac{y - m(x)}{\hat{\sigma}(x)}\right) + \frac{1}{2}\left(\frac{\hat{m}(x) - m(x)}{\hat{\sigma}(x)}\right)^2 f_e'(A_x),$$

for some A_x between $(y - m(x))/\hat{\sigma}(x)$ and $(y - \hat{m}(x))/\hat{\sigma}(x)$. The second term of these two terms is $O((na_n)^{-1} \log a_n^{-1})$ a.s. by Proposition 4.5. For the first term, we first replace $\hat{\sigma}(x)$ by $\sigma(x)$ (using Proposition 4.5) and then apply Proposition 4.8. For $\alpha_n^3(x, y)$ we have

$$\alpha_n^3(x, y) = -\frac{\hat{\sigma}(x)-\sigma(x)}{\hat{\sigma}(x)} \frac{y-m(x)}{\sigma(x)} f_e\left(\frac{y-m(x)}{\sigma(x)}\right) + \frac{1}{2}\left(\frac{\hat{\sigma}(x)-\sigma(x)}{\hat{\sigma}(x)}\right)^2 \left(\frac{y-m(x)}{\sigma(x)}\right)^2 f_e'(B_x),$$

where B_x is between $(y - m(x))/\sigma(x)$ and $(y - m(x))/\hat{\sigma}(x)$. The second term above is $O((na_n)^{-1} \log a_n^{-1})$ a.s. by Proposition 4.5 and the fact that $\sup_y |y^2 f_e'(y)| < \infty$, and the first term has an asymptotic representation given by Proposition 4.9. The above show that

$$\begin{aligned} &\alpha_n^2(x, y) + \alpha_n^3(x, y) \\ &\quad = (na_n)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{a_n}\right) h_{x,y}(Z_i, \Delta_i) \\ &\qquad + O((na_n)^{-3/4} (\log a_n^{-1})^{3/4}) \quad \text{a.s.} \end{aligned} \tag{5.2}$$

Next note that $(na_n)^{1/2}\alpha_n^1(x, y) = o_P(1)$, uniformly in $(x, y) \in \Omega$, follows from the order of the remainder term in Theorem 3.1, and the weak convergence established in Corollary 3.2. This and (5.2) complete the proof of the theorem. □

PROOF OF COROLLARY 3.4. Let $Z_{ni}(y \mid x) = (na_n)^{-1/2} K((x - X_i)/a_n) h_{x,y}(Z_i, \Delta_i)$, $i = 1, \ldots, n$, be a triangular array of random processes, with $y \in \mathscr{F} = \{y; (y - m(x))/\sigma(x) \leq T\}$ and x fixed. We will prove that the conditions displayed in Theorem 2.11.9 in van der Vaart and Wellner (1996) are satisfied for $\sum_{i=1}^n Z_{ni}$. Endow $\mathscr{F}$ with the semimetric ρ defined by

$$\rho(y, y') = \max\left\{\left|f_e\left(\frac{y'-m(x)}{\sigma(x)}\right) - f_e\left(\frac{y-m(x)}{\sigma(x)}\right)\right|, \left|\frac{y'-m(x)}{\sigma(x)} f_e\left(\frac{y'-m(x)}{\sigma(x)}\right) - \frac{y-m(x)}{\sigma(x)} f_e\left(\frac{y-m(x)}{\sigma(x)}\right)\right|\right\}.$$

Since $|f_e(z)|$ and $|zf_e(z)|$ are bounded for $z \in \mathbb{R}$, we can divide $\mathscr{F}$ for every $\varepsilon > 0$ into $N_\varepsilon = O(\varepsilon^{-1})$ subintervals $\mathscr{F}_{\varepsilon j}$, $j = 1, \ldots, N_\varepsilon$, such that $\rho(y, y') \leq C\varepsilon$, $C > 0$ for all $y, y' \in \mathscr{F}_{\varepsilon j}$ and hence $\sum_{i=1}^n \sup_{y, y' \in \mathscr{F}_{\varepsilon j}} |Z_{ni}(y' \mid x) - Z_{ni}(y \mid x)|^2 \leq \varepsilon^2$ by proper choice of C, since $\eta(z, \delta \mid x)$ and $\zeta(z, \delta \mid x)$ are bounded as well. This shows that the bracketing number is $O(\varepsilon^{-1})$ and hence the third displayed condition is satisfied. The other two conditions are easily seen to hold.

It remains to calculate the covariance of the limiting process. Let $\mathbf{X} = (X_1, \ldots, X_n)'$. Since $\sup_{x,y} |E(\hat{H}(y \mid x) \mid \mathbf{X}) - H(y \mid x)|$ is easily seen to be $O(a_n^2)$, it follows that $\sum_{i=1}^n W_i(x, a_n)\ E(\xi(Z_i,\ \Delta_i,\ y \mid x) \mid \mathbf{X}) = O(a_n^2)$ uniformly in x and y. Noting that $h_{x,y}$ is defined in terms of the functions η

and ζ which include the function ξ, it can be seen that the above implies $\sum_{i=1}^n W_i(x, a_n)E(h_{x,y}(Z_i, \Delta_i) \mid \mathbf{X}) = O(a_n^2)$. We will show that

$$(na_n)^{-1} \sum_{i=1}^n E\left\{K^2\left(\frac{x - X_i}{a_n}\right)\gamma(X_i, x, y, y')\right\} \tag{5.3}$$

tends to the displayed expression, where

$$\begin{aligned}\gamma(X_i, x, y, y') = E\{&[h_{x,y}(Z_i, \Delta_i) - E(h_{x,y}(Z_i, \Delta_i) \mid X_i)] \\ &\times [h_{x,y'}(Z_i, \Delta_i) - E(h_{x,y'}(Z_i, \Delta_i) \mid X_i)] \mid \mathbf{X}\}.\end{aligned}$$

Using the fact that for all $s \le t$,

$$\begin{aligned}&E\{[I(Z_i \le s) - H(s \mid X_i)][I(Z_i \le t) - H(t \mid X_i)] \mid \mathbf{X}\} \\ &\quad = H(s \mid x) - H(s \mid x)H(t \mid x) + O(a_n),\end{aligned}$$

holds whenever $K((X_i - x)/a_n) \ne 0$, it follows after some simple algebra that $|\gamma(X_i, x, y, y') - \gamma(x, x, y, y')| = O(a_n)$ for $K((X_i - x)/a_n) \ne 0$. Hence, (5.3) equals

$$\begin{aligned}&f_X(x) \int K^2(u)\, du\, \gamma(x, x, y, y') + o(1) \\ &\quad = f_X(x) \int K^2(u)\, du\, \operatorname{Cov}(h_{x,y}(Z, \Delta), h_{x,y'}(Z, \Delta) \mid X = x) + o(1).\end{aligned}$$

PROOF OF THEOREM 3.5. Write

$$\begin{aligned}&\hat{F}_T(x, y) - F_T(x, y) \\ &\quad = \int_{-\infty}^x \left[\hat{F}_e\left(\frac{y \wedge T_t - \hat{m}(t)}{\hat{\sigma}(t)}\right) - F_e\left(\frac{y \wedge T_t - m(t)}{\sigma(t)}\right)\right] d\hat{F}_X(t) \\ &\qquad + \int_{-\infty}^x F_e\left(\frac{y \wedge T_t - m(t)}{\sigma(t)}\right) d\big(\hat{F}_X(t) - F_X(t)\big).\end{aligned} \tag{5.4}$$

For the first term on the right-hand side of (5.4) we will make use of Theorem 3.3. However, the remainder term in that representation is not $o_P(n^{-1/2})$, as is required. Let $\alpha_n^i(x, y)$, $i = 1, 2, 3$, be as in the proof of that representation. Then, the first term on the right-hand side of (5.4) equals

$$\int_{-\infty}^x \left[\alpha_n^1(t, y \wedge T_t) + \alpha_n^2(t, y \wedge T_t) + \alpha_n^3(t, y \wedge T_t)\right] d\hat{F}_X(t). \tag{5.5}$$

Using equation (5.2), $\alpha_n^2(t, y \wedge T_t) + \alpha_n^3(t, y \wedge T_t)$ can be expressed as the sum of a leading term and a remainder term which is $o_P(n^{-1/2})$ uniformly in (t, y). Thus only the term $\alpha_n^1(t, y \wedge T_t)$ needs further attention. However, using Theorem 3.1 together with Lemma B.1, it is easily seen that this term can also be written as a leading term and a remainder of order $o_P(n^{-1/2})$. Combining the two leading terms, it follows that (5.5) equals

$$\begin{aligned}
&n^{-2}\sum_{i=1}^n\sum_{j=1}^n \varphi\biggl(X_i, Z_i, \Delta_i, \frac{y\wedge T_{X_j}-m(X_j)}{\sigma(X_j)}\biggr) I(X_j\le x)\\
&\quad + n^{-2}a_n^{-1}\sum_{i=1}^n\sum_{j=1}^n K\biggl(\frac{X_j-X_i}{a_n}\biggr) f_X^{-1}(X_j)\\
&\qquad\times\biggl[\eta(Z_i,\Delta_i\,|\,X_j)+\zeta(Z_i,\Delta_i\,|\,X_j)\frac{y\wedge T_{X_j}-m(X_j)}{\sigma(X_j)}\biggr]\\
&\qquad\times f_e\biggl(\frac{y\wedge T_{X_j}-m(X_j)}{\sigma(X_j)}\biggr) I(X_j\le x) + o_P(n^{-1/2}).
\end{aligned}\tag{5.6}$$

The result now follows from Propositions B.5, its analogue for ζ and B.6. □

PROOF OF COROLLARY 3.6. To prove the weak convergence of the given process, we will make use of results in van der Vaart and Wellner (1996). Let $\mathscr{F} = \{(t,z,\delta)\to g_{x,y}(t,z,\delta)\colon x\in R_X,\ y\in\mathbb{R}\}$. We will show that the class $\mathscr{F}$ is Donsker by showing that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon,\mathscr{F},L_2(P))}\,d\varepsilon<\infty \tag{5.7}$$

[see page 130 in van der Vaart and Wellner (1996)] where $N_{[]}$ is the bracketing number. Set $e_T(X)=(y\wedge T_X-m(X))/\sigma(X)$ and write

$$\begin{aligned}
g^1_{x,y}(t,z,\delta) &= -\int_{-\infty}^x\int_{-\infty}^{(z-m(t))/\sigma(t)} \frac{1-F_e(e_T(u))}{(1-H_e(s))^2} I(s\le e_T(u))\,dH_{e1}(s)\,dF_X(u)\\
&\quad+\frac{I(\delta=1)}{1-H_e((z-m(t))/\sigma(t))}\\
&\quad\times\int_{-\infty}^x (1-F_e(e_T(u))) I\biggl(\frac{z-m(t)}{\sigma(t)}\le e_T(u)\biggr)\,dF_X(u)\\
&\quad-\eta(z,\delta\,|\,t)\int_{-\infty}^x S_e(e_T(u))\gamma_1(e_T(u)\,|\,t)\,dF_X(u)\\
&\quad-\zeta(z,\delta\,|\,t)\int_{-\infty}^x S_e(e_T(u))\gamma_2(e_T(u)\,|\,t)\,dF_X(u)\\
&=\sum_{i=1}^4 g^{1i}_{x,y}(t,z,\delta).
\end{aligned}$$

The term $g^{11}_{x,y}(t,z,\delta)$ is decreasing as a function of $(z-m(t))/\sigma(t)$ and bounded and hence $O(\exp(K\varepsilon^{-1}))$ brackets are required by Theorem 2.7.5 in van der Vaart and Wellner (1996). Similarly, the bracketing number for the integral in $g^{12}_{x,y}(t,z,\delta)$ is at most $O(\exp(K\varepsilon^{-1}))$ and hence the same holds for $g^{12}_{x,y}(t,z,\delta)$ itself since the expression in front of the integral is bounded and independent of x and y. Since the integrals in $g^{13}_{x,y}(t,z,\delta)$ and $g^{14}_{x,y}(t,z,\delta)$ are bounded (uniformly in t), as well as their first derivatives (with respect to t), their bracketing number is $O(\exp(K\varepsilon^{-1}))$ by Corollary 2.7.2. Therefore, the terms $g^{13}_{x,y}(t,z,\delta)$ and $g^{14}_{x,y}(t,z,\delta)$ themselves also need $O(\exp(K\varepsilon^{-1}))$ brackets, because $\eta(z,\delta\,|\,t)$ and $\zeta(z,\delta\,|\,t)$ are bounded and independent of x and y. All together, we have $O(\exp(K\varepsilon^{-1}))$ brackets for $g^{1}_{x,y}(t,z,\delta)$.

For $g^{2}_{x,y}(t,z,\delta)$ we first note that the bracketing number for the class of functions of the form $t\to I(t\le x)$ is $O(\exp(K\varepsilon^{-1}))$ by Theorem 2.7.5 in the aforementioned book. For the functions $t\to f_e((y\wedge T_t-m(t))/\sigma(t))$, we consider three cases. If $y\le \min_t T_t$, then Corollary 2.7.2 in van der Vaart and Wellner's book entails that $O(\exp(K\varepsilon^{-1}))$, brackets are needed. If $y\ge \max_t T_t$, then $f_e((y\wedge T_t-m(t))/\sigma(t))$ does not depend on y and hence one bracket suffices. For the intermediate case, that is, for $\min_t T_t\le y\le \max_t T_t$, we apply Theorem 2.11.9 in van der Vaart and Wellner (1996) on $\sum_{i=1}^n Z_{ni}(y)=n^{-1/2}\sum_{i=1}^n f_e((y\wedge T_{X_i}-m(X_i))/\sigma(X_i))$. For each $\varepsilon>0$, divide the interval $[\min_t T_t,\max_t T_t]$ into $N_\varepsilon=O(\varepsilon^{-1})$ subintervals $\mathscr{F}^n_{\varepsilon j}$ of length not more than $K\varepsilon$ for some $K>0$. Then, for each $j=1,\dots,N_\varepsilon$, $\sum_{i=1}^n \sup_{y,y'\in\mathscr{F}^n_{\varepsilon j}}|Z_{ni}(y)-Z_{ni}(y')|^2\le\varepsilon^2$, by proper choice of $K>0$. This shows that the bracketing number is $O(\varepsilon^{-1})$ in this case. In total, we have $O(\exp(K\varepsilon^{-1}))$ brackets for the class $\{t\to f_e((y\wedge T_t-m(t))/\sigma(t));\,y\in\mathbb{R}\}$. Similarly, the class $\{t\to(y\wedge T_t-m(t))/\sigma(t)f_e((y\wedge T_t-m(t))/\sigma(t))\}$ also requires $O(\exp(K\varepsilon^{-1}))$ brackets. This shows that the bracketing number for $g^{2}_{x,y}(t,z,\delta)$ is $O(\exp(K\varepsilon^{-1}))$. Finally, for the term $g^{3}_{x,y}(t,z,\delta)$, analogous arguments as before show that also $O(\exp(K\varepsilon^{-1}))$ brackets are needed for this term.

This shows that the integral in (5.7) is bounded by $K\int_o^{2M} d\varepsilon/\varepsilon^{1/2}<\infty$ where M is an upper bound for $|g_{x,y}(t,z,\delta)|$ (because for $\varepsilon>2M$, one bracket suffices to cover $\mathscr{F}$). This shows that the class $\mathscr{F}$ is Donsker. The result now follows from pages 81 and 82 in van der Vaart and Wellner (1996).

APPENDIX A

A result needed for Theorem 3.1.

LEMMA A.1. *Assume* (A1) (i), (iii), (iv), (A2) (i), (ii), (A3) (ii), F_X *is twice continuously differentiable*, $\inf_{x\in R_X} f_X(x)>0$, $H(y\,|\,x)$ *and* $H_1(y\,|\,x)$ *satisfy* (A6) (i), (iv), (v), *and* $H_e(y\,|\,x)$ *satisfies* (A6) (ii). *Then*,

$$\sup_{-\infty<y<+\infty}\Big|n^{-1}\sum_{i=1}^n\{I(\hat E_i\le y)-I(E_i\le y)-P(\hat E\le y\,|\,\mathscr{X}_n)+P(E\le y)\}\Big|=o_P(n^{-1/2}),$$

where $P(\hat E\le y\,|\,\mathscr{X}_n)$ *is the distribution of* $\hat E=(Z-\hat m(X))/\hat\sigma(X)$ *conditioning on* (X_j,Z_j,Δ_j), $j=1,\dots,n$.

PROOF. The expression between braces equals $I(E_i \le yd_{n2}(X_i) + d_{n1}(X_i)) - I(E_i \le y) - P(E \le yd_{n2}(X) + d_{n1}(X)) + P(E \le y)$, where $d_{n1}(x) = (\hat{m}(x) - m(x))/\sigma(x)$ and $d_{n2}(x) = \hat{\sigma}(x)/\sigma(x)$. The proof is based on results in van der Vaart and Wellner (1996). Let

$$\begin{aligned}\mathscr{F} = \{&I(E \le yd_2(X) + d_1(X)) - I(E \le y)\\ &-P(E \le yd_2(X) + d_1(X)) + P(E \le y);\\ &-\infty < y < +\infty, d_1 \in C_1^{1+\delta}(R_X) \text{ and } d_2 \in \tilde{C}_2^{1+\delta}(R_X)\},\end{aligned}$$

where $C_1^{1+\delta}(R_X)$ is the class of all differentiable functions d defined on the domain R_X of X such that $\|d\|_{1+\delta} \le 1$, $\tilde{C}_2^{1+\delta}(R_X)$ is the class of all differentiable functions d defined on R_X such that $\|d\|_{1+\delta} \le 2$, and $\inf_x\{d(x)\} \ge \frac{1}{2}$ and $\|d\|_{1+\delta} = \max\{\sup_x |d(x)|, \sup_x |d'(x)|\} + \sup_{x,x'} |d'(x) - d'(x')|/|x - x'|^\delta$. Note that by Propositions 4.5, 4.6 and 4.7, we have that $P(d_{n1} \in C_1^{1+\delta}(R_X)$ and $d_{n2} \in \tilde{C}_2^{1+\delta}(R_X)) \to 1$ as $n \to \infty$. In a first step we will show that the class $\mathscr{F}$ is Donsker. From Theorem 2.5.6 in van der Vaart and Wellner (1996), it follows that it suffices to show that

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, L_2(P))}\, d\varepsilon < \infty, \tag{A.1}$$

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of (E, X), $L_2(P)$ is the L_2-norm. We will restrict ourselves to showing (A.1) for the class $\mathscr{F}_1 = \{I(E \le yd_2(X) + d_1(X));\ -\infty < y < +\infty, d_1 \in C_1^{1+\delta}(R_X)$ and $d_2 \in \tilde{C}_2^{1+\delta}(R_X)\}$, since the other terms are similar, but much easier. In Corollary 2.7.2 of the aforementioned book it is stated that $m_1 = N_{[]}(\varepsilon^2, C_1^{1+\delta}(R_X), L_2(P)) \le \exp(K\varepsilon^{-2/(1+\delta)})$ and $m_2 = N_{[]}(\varepsilon^2, \tilde{C}_2^{1+\delta}(R_X), L_2(P)) \le \exp(K\varepsilon^{-2/(1+\delta)})$. Let $d_1^L \le d_1^U, \dots, d_{m_1}^L \le d_{m_1}^U$ be the functions defining the m_1 brackets for $C_1^{1+\delta}(R_X)$ and let $\tilde{d}_1^L \le \tilde{d}_1^U, \dots, \tilde{d}_{m_2}^L \le \tilde{d}_{m_2}^U$ be the functions defining the m_2 brackets for $\tilde{C}_2^{1+\delta}(R_X)$. Thus, for each d_1 and d_2 and each fixed y,

$$\begin{aligned}I(E \le y\tilde{d}_j^L(X) + d_i^L(X)) &\le I(E \le yd_2(X) + d_1(X))\\ &\le I(E \le y\tilde{d}_j^U(X) + d_i^U(X)).\end{aligned}$$

Define $F_{ij}^L(y) = P(E \le y\tilde{d}_j^L(X) + d_i^L(X))$ and let y_{ijk}^L, $k = 1, \dots, O(\varepsilon^{-2})$, partition the line in segments having F_{ij}^L-probability less than or equal to a fraction of ε^2. Similarly, define $F_{ij}^U(y) = P(E \le y\tilde{d}_j^U(X) + d_i^U(X))$ and let y_{ijk}^U, $k = 1, \dots, O(\varepsilon^{-2})$, partition the line in segments having F_{ij}^U-probability less than or equal to a fraction of ε^2. Now, define the following bracket for y: $y_{ijk_1}^L \le y \le y_{ijk_2}^U$, where $y_{ijk_1}^L$ is the largest of the y_{ijk}^L with the property of being less than or equal to y and $y_{ijk_2}^U$ is the smallest of the y_{ijk}^U with the property of being greater than or equal to y. We will now show that the brackets for our function are given by

$$I\big(E \le y^L_{ijk_1}\tilde d^L_j(X) + d^L_i(X)\big) \le I\big(E \le y d_2(X) + d_1(X)\big) \le I\big(E \le y^U_{ijk_2}\tilde d^U_j(X) + d^U_i(X)\big).$$

Let us calculate

$$\Big\| I\big(E \le y^U_{ijk_2}\tilde d^U_j(X) + d^U_i(X)\big) - I\big(E \le y^L_{ijk_1}\tilde d^L_j(X) + d^L_i(X)\big)\Big\|_2^2 = F^U_{ij}(y^U_{ijk_2}) - F^L_{ij}(y^L_{ijk_1}) = F^U_{ij}(y) - F^L_{ij}(y) + K\varepsilon^2.$$

Applying a Taylor expansion to the function H_e, yields

$$\begin{aligned}
&F^U_{ij}(y) - F^L_{ij}(y)\\
&\quad= \int \big[H_e(y\tilde d^U_j(x) + d^U_i(x)\,|\,x) - H_e(y\tilde d^L_j(x) + d^L_i(x)\,|\,x)\big]\,dF_X(x)\\
&\quad= \int h_e\big(y\tilde\xi_j(x) + \xi_i(x)\,|\,x\big)\\
&\qquad\quad \times \big[y(\tilde d^U_j(x) - \tilde d^L_j(x)) + (d^U_i(x) - d^L_i(x))\big]\,dF_X(x)\\
&\quad= \int h_e\big(y\tilde\xi_j(x) + \xi_i(x)\,|\,x\big)\big(y\tilde\xi_j(x) + \xi_i(x)\big)\big(\tilde\xi_j(x)\big)^{-1}\\
&\qquad\qquad \times \big(\tilde d^U_j(x) - \tilde d^L_j(x)\big)\,dF_X(x)\\
&\qquad - \int h_e\big(y\tilde\xi_j(x) + \xi_i(x)\,|\,x\big)\xi_i(x)(\tilde\xi_j(x))^{-1}\\
&\qquad\qquad \times (\tilde d^U_j(x) - \tilde d^L_j(x))\,dF_X(x)\\
&\qquad + \int h_e\big(y\tilde\xi_j(x) + \xi_i(x)\,|\,x\big)(d^U_i(x) - d^L_i(x))\,dF_X(x).
\end{aligned}\tag{A.2}$$

Here, $\xi_i(x)$ is between $d^L_i(x)$ and $d^U_i(x)$ and $\tilde\xi_j(x)$ is between $\tilde d^L_j(x)$ and $\tilde d^U_j(x)$. Since we can choose the brackets d_i and $\tilde d_j$ such that $\sup_x |d^U_i(x)| \le 1$ and $\inf_x |\tilde d^L_j(x)| \ge \frac{1}{2}$ (for all i and j) and since $\sup_{x,\,y} |yh_e(y\,|\,x)| < \infty$, (A.2) is bounded in absolute value by

$$K_1\|\tilde d^U_j - \tilde d^L_j\|_{P,1} + K_2\|d^U_i - d^L_i\|_{P,1} \le (K_1 + K_2)\varepsilon^2$$

(since $\|d\|_{P,1} \le \|d\|_{P,2}$ for any function d, where $\|d\|_{P,1}$, respectively, $\|d\|_{P,2}$ is the $L_1(P)$-norm, respectively, $L_2(P)$-norm of d). Hence, for the class $\mathscr{F}_1$ and for each $\varepsilon > 0$, we have at most $O(\varepsilon^{-2}\exp(K\varepsilon^{-2/(1+\delta)}))$ brackets in total. However, for $\varepsilon > 1$, one bracket suffices. So we have $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathscr{F}_1, L_2(P))}\,d\varepsilon < \infty$. This shows that the class $\mathscr{F}_1$ (and hence $\mathscr{F}$) is Donsker.

Next, let us calculate

$$\operatorname{Var}\big(I(E \le y d_{n2}(X) + d_{n1}(X)) - I(E \le y) - P\big(E \le y d_{n2}(X) + d_{n1}(X)\big) + P(E \le y)\big)$$

$$
\begin{aligned}
&= \operatorname{Var}\big(I(E \le y d_{n2}(X) + d_{n1}(X)) - I(E \le y)\big) \\
&\le E\big[E(\{I(E \le y d_{n2}(X) + d_{n1}(X)) - I(E \le y)\}^2 \mid X)\big] \qquad \text{(A.3)} \\
&= E\big[H_e(y d_{n2}(X) + d_{n1}(X) \mid X) \\
&\quad - H_e(\min(y, y d_{n2}(X) + d_{n1}(X)) \mid X)\big] \\
&\quad + E\big[H_e(y \mid X) - H_e(\min(y, y d_{n2}(X) + d_{n1}(X)) \mid X)\big] \\
&= E\big[h_e(y a_{n2}(X) + a_{n1}(X) \mid X)\big|y(d_{n2}(X) - 1) + d_{n1}(X)\big|\big],
\end{aligned}
$$

for some $a_{n1}(X)$ between 0 and $d_{n1}(X)$ and some $a_{n2}(X)$ between 1 and $d_{n2}(X)$. Since

$$
\begin{aligned}
&\sup_x \big|y h_e(y a_{n2}(x) + a_{n1}(x)|x)\big| \\
&\quad \le \sup_x \Big|\big\{\big|y a_{n2}(x) + a_{n1}(x)\big| a_{n2}(x)^{-1} + \big|a_{n1}(x)\big| a_{n2}(x)^{-1}\big\} \\
&\qquad\qquad \times h_e(y a_{n2}(x) + a_{n1}(x) \mid x)\Big| \\
&\quad \le K_1 \quad \text{(say)}
\end{aligned}
$$

[by Proposition 4.5, because $\sup_{x,y} h_e(y \mid x) < \infty$, $\sup_{x,y} |y h_e(y \mid x)| < \infty$ and $\inf_x \sigma(x) > 0$], (A.3) is bounded by

$$
\begin{aligned}
&K_1 E|d_{n2}(X) - 1| + \sup_{x,y} h_e(y \mid x) E|d_{n1}(X)| \\
&\quad \le K_1 \sup_x \left|\frac{\hat\sigma(x)}{\sigma(x)} - 1\right| + K_2 \sup_x |\hat m(x) - m(x)| \to 0 \quad \text{a.s.},
\end{aligned}
$$

again by Proposition 4.5. Since the class $\mathscr{F}$ is Donsker, it follows from Corollary 2.3.12 in van der Vaart and Wellner (1996) that

$$
\lim_{\alpha \downarrow 0} \limsup_{n \to \infty} P\left(\sup_{f \in \mathscr{F}, \operatorname{Var}(f) < \alpha} n^{-1/2} \left| \sum_{i=1}^n f(X_i) \right| > \varepsilon \right) = 0,
$$

for each $\varepsilon > 0$. By restricting the supremum inside this probability to the elements in $\mathscr{F}$ corresponding to $d_1(X) = d_{n1}(X)$ and $d_2(X) = d_{n2}(X)$ as defined above, the result follows.

PROPOSITION A.2. *Assume* (A1), (A2) (i), (ii), (A3) (i), (ii), (A4) (i), $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i)–(vi), *and* $H_e(y \mid x)$ *satisfies* (A6) (ii), (iii), (vi), (vii). *Then*,

$$
\begin{aligned}
&\hat H_e(y) - H_e(y) \\
&\quad = n^{-1} \sum_{i=1}^n \big[-\{\eta(Z_i, \Delta_i \mid X_i) + \zeta(Z_i, \Delta_i \mid X_i) y\} h_e(y \mid X_i) \\
&\qquad\qquad + I(E_i \le y) - H_e(y)\big] + R_n(y),
\end{aligned}
$$

where $\sup\{|R_n(y)|; -\infty < y < +\infty\} = o_P(n^{-1/2})$.

PROOF. Using Lemma A.1 and the notation in the statement of that lemma,

$$
\begin{aligned}
&\hat{H}_e(y) - H_e(y) \\
&\quad = \int \left\{ H_e\left(\frac{y\hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)} \middle| x \right) - H_e(y \mid x) \right\} dF_X(x) \\
&\qquad + n^{-1} \sum_{i=1}^{n} \{ I(E_i \le y) - H_e(y) \} + o_P(n^{-1/2}) \\
&\quad = \int h_e(y \mid x) \frac{y(\hat{\sigma}(x) - \sigma(x)) + (\hat{m}(x) - m(x))}{\sigma(x)} \, dF_X(x) \qquad \text{(A.4)} \\
&\qquad + \frac{1}{2} \int h_e'(y_i \mid x) \left(\frac{y(\hat{\sigma}(x) - \sigma(x)) + (\hat{m}(x) - m(x))}{\sigma(x)} \right)^2 dF_X(x) \\
&\qquad + n^{-1} \sum_{i=1}^{n} \{ I(E_i \le y) - H_e(y) \} + o_P(n^{-1/2}),
\end{aligned}
$$

where y_i is between y and $(y\hat{\sigma}(X_i) + \hat{m}(X_i) - m(X_i))/\sigma(X_i)$. The second term above is $o(n^{-1/2})$ a.s. by Proposition 4.5 and since $\sup_{x,y} |y^2 h_e'(y \mid x)| < \infty$. The first term on the right-hand side of (A.4) splits naturally in two parts. We will deal with the second one only. Using Proposition 4.8 it follows that

$$
\begin{aligned}
&\int h_e(y \mid x) \frac{\hat{m}(x) - m(x)}{\sigma(x)} \, dF_X(x) \\
&\quad = -(na_n)^{-1} \sum_{i=1}^{n} \int f_X^{-1}(x) h_e(y \mid x) K\left(\frac{x - X_i}{a_n} \right) \eta(Z_i, \Delta_i \mid x) \, dF_X(x) \\
&\qquad + o(n^{-1/2}) \qquad \text{(A.5)} \\
&\quad = -(na_n)^{-1} \sum_{i=1}^{n} \int K\left(\frac{x - X_i}{a_n} \right) \beta(x, Z_i, \Delta_i, y) \, dF_X(x) \\
&\qquad + o(n^{-1/2}),
\end{aligned}
$$

a.s., uniformly in y, where $\beta(x, z, \delta, y) = f_X^{-1}(x) h_e(y \mid x) \eta(z, \delta \mid x)$. Using a two-term Taylor expansion of $\beta(x, Z_i, \Delta_i, y)$ around X_i, (A.5) can be written as

$$
\begin{aligned}
&-(na_n)^{-1} \sum_{i=1}^{n} \beta(X_i, Z_i, \Delta_i, y) \int K\left(\frac{x - X_i}{a_n} \right) dF_X(x) \\
&-(na_n)^{-1} \sum_{i=1}^{n} \beta'(X_i, Z_i, \Delta_i, y) \int K\left(\frac{x - X_i}{a_n} \right) (x - X_i) \, dF_X(x) \\
&-\frac{1}{2}(na_n)^{-1} \sum_{i=1}^{n} \int K\left(\frac{x - X_i}{a_n} \right) (x - X_i)^2 \beta''(\xi_i, Z_i, \Delta_i, y) \, dF_X(x) + o(n^{-1/2})
\end{aligned}
$$

$$= -n^{-1} \sum_{i=1}^{n} \beta(X_i, Z_i, \Delta_i, y) f_X(X_i) + O(a_n^2) + o(n^{-1/2})$$

$$= -n^{-1} \sum_{i=1}^{n} h_e(y \mid X_i) \eta(Z_i, \Delta_i \mid X_i) + o(n^{-1/2}),$$

where ξ_i is between x and X_i. This completes the proof. □

PROPOSITION A.3. *Assume* (A1) (i), (iii), (iv), (A2) (i), (ii), (A3) (ii), F_X *is twice continuously differentiable*, $\inf_{x \in R_X} f_X(x) > 0$, $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i), (iv), (v), *and* $H_e(y \mid x)$ *satisfies* (A6) (ii). *Then*,

$$\sup_{-\infty < y < +\infty} |\hat{H}_e(y) - H_e(y)| = O((na_n)^{-1/2} (\log a_n^{-1})^{1/2}) \quad a.s.$$

PROOF. Applying Lemma A.1 and a Taylor expansion yields that

$$\hat{H}_e(y) - H_e(y) = \int h_e(y_x \mid x) \frac{y(\hat{\sigma}(x) - \sigma(x)) + (\hat{m}(x) - m(x))}{\sigma(x)} \, dF_X(x)$$
$$+ n^{-1} \sum_{i=1}^{n} \{I(E_i \le y) - H_e(y)\} + o_P(n^{-1/2}).$$

where y_x is between y and $((y\hat{\sigma}(x) + \hat{m}(x) - m(x)))/\sigma(x)$. The first term is $O((na_n)^{-1/2} (\log a_n^{-1})^{1/2})$ a.s. by Proposition 4.5. For the second one, the Dvoretzky–Kiefer–Wolfowitz (1956) inequality yields $O(n^{-1/2} (\log n)^{1/2})$ a.s. □

REMARK A.1. The representation in Proposition A.2 can also be used to obtain a bound of $O_P(n^{-1/2})$, instead of $O((na_n)^{-1/2} (\log a_n^{-1})^{1/2})$ a.s. as established above. Note that the nature of such results is entirely different from the case where Euclidean parameters for location and scale are estimated. The rate of convergence of the present nonparametric estimators is slower than $n^{-1/2}$ and recovery of the root n convergence is due to the averaging over the covariate values. Similar results can be found in Cristóbal Cristóbal, Foraldo Roca and González Manteiga (1987) and in Akritas (1994, 1996).

PROPOSITION A.4. *Assume* (A1) (i), (iii), (iv), (A2) (i), (ii), (A3) (ii), F_X *is twice continuously differentiable*, $\inf_{x \in R_X} f_X(x) > 0$, $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i), (iv), (v), *and* $H_e(y \mid x)$ *satisfies* (A6) (ii), (iii), (vi). *Let* $J_c = \{(y_1, y_2); |H_e(y_2) - H_e(y_1)| \le c\}$. *Then*,

$$\sup\{|\hat{H}_e(y_2) - \hat{H}_e(y_1) - H_e(y_2) + H_e(y_1)|; (y_1, y_2) \in J_{\bar{a}_n}\} = o_P(n^{-1/2}),$$

where $\bar{a}_n$ *is any sequence of positive numbers tending to zero as* n *tends to infinity that satisfies* $\bar{a}_n a_n^{-1} \log a_n^{-1} \to 0$.

PROOF. Consider the decomposition

$$
\begin{aligned}
&\hat{H}_e(y_2) - \hat{H}_e(y_1) - H_e(y_2) + H_e(y_1) \\
&\quad = \int \biggl\{ H_e\biggl(\frac{y_2 \hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)} \biggm| x \biggr) - H_e(y_2 \mid x) \\
&\qquad - H_e\biggl(\frac{y_1 \hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)} \biggm| x \biggr) + H_e(y_1 \mid x) \biggr\} \, dF_X(x) \qquad \text{(A.6)} \\
&\qquad + n^{-1} \sum_{i=1}^{n} \{ I(E_i \le y_2) - H_e(y_2) - I(E_i \le y_1) + H_e(y_1) \} + o_P(n^{-1/2}),
\end{aligned}
$$

uniformly in y_1 and y_2 by Lemma A.1. The second term on the right-hand side of (A.6) is $O(\bar{a}_n^{1/2} n^{-1/2} (\log n)^{1/2})$ a.s. by Lemma 2.4 in Stute (1982). The first term equals

$$
\begin{aligned}
&\int \biggl\{ \frac{\hat{m}(x) - m(x)}{\sigma(x)} (h_e(y_2 \mid x) - h_e(y_1 \mid x)) \\
&\qquad + \frac{\hat{\sigma}(x) - \sigma(x)}{\sigma(x)} (y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x)) \biggr\} \, dF_X(x) \\
&\qquad + O\bigl((n a_n)^{-1} \log a_n^{-1}\bigr) \\
&\quad \le \sup_x |\hat{m}(x) - m(x)| \Bigl(\inf_x \sigma(x) \Bigr)^{-1} \qquad \text{(A.7)} \\
&\qquad\quad \times \int \bigl| h_e(y_2 \mid x) - h_e(y_1 \mid x) \bigr| \, dF_X(x) \\
&\qquad + \sup_x \bigl| \hat{\sigma}(x) - \sigma(x) \bigr| \Bigl(\inf_x \sigma(x) \Bigr)^{-1} \\
&\qquad\quad \times \int \bigl| y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x) \bigr| \, dF_X(x) \\
&\qquad + O((n a_n)^{-1} \log a_n^{-1}).
\end{aligned}
$$

We will show that for all $x \in R_X$,

$$
|y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x)| \le \frac{K}{\bar{a}_n^{1/2}} \bigl| H_e(y_2 \mid x) - H_e(y_1 \mid x) \bigr| + 2 \bar{a}_n^{1/2}. \qquad \text{(A.8)}
$$

Showing this implies that for $(y_1, y_2) \in J_{\bar{a}_n}$,

$$
\begin{aligned}
&\int \bigl| y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x) \bigr| \, dF_X(x) \\
&\qquad \le \frac{K}{\bar{a}_n^{1/2}} \int \bigl| H_e(y_2 \mid x) - H_e(y_1 \mid x) \bigr| \, dF_X(x) + 2 \bar{a}_n^{1/2} \\
&\qquad = \frac{K}{\bar{a}_n^{1/2}} \bigl| H_e(y_2) - H_e(y_1) \bigr| + 2 \bar{a}_n^{1/2} \le (K + 2) \bar{a}_n^{1/2}.
\end{aligned}
$$

This together with Proposition 4.5 and the assumption on $\bar{a}_n$ implies $o_P(n^{-1/2})$ for the second term of (A.7). The derivation for the first term is similar, but easier. For the proof of (A.8), define $A_x = \{y; |yh_e(y \mid x)| \geq \bar{a}_n^{1/2}\}$. Clearly, in the case where $y_1 \notin A_x$ and $y_2 \notin A_x$ there is nothing to show. Consider next the case where $y_1 \notin A_x$ and $y_2 \in A_x$. Then there exists a y between y_1 and y_2 such that $|yh_e(y \mid x)| = \bar{a}_n^{1/2}$ and $u \in A_x$ for all u between y and y_2. Using this y, write

$$\text{(A.9)} \qquad \begin{aligned} \big|y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x)\big| &\leq \big|y_2 h_e(y_2 \mid x) - y h_e(y \mid x)\big| \\ &\quad + \big|y h_e(y \mid x) - y_1 h_e(y_1 \mid x)\big|. \end{aligned}$$

The second term on the right-hand side of (A.9) is clearly less than $2\bar{a}_n^{1/2}$. To deal with the first term define $v_e(y \mid x) = yh_e(y \mid x)$ and write

$$\text{(A.10)} \qquad \begin{aligned} &\big|v_e(y_2 \mid x) - v_e(y \mid x)\big| \\ &\quad = \big|\big(v_e(\cdot \mid x) \circ H_e^{-1}(\cdot \mid x)\big)'(H_e(u \mid x))\big|(H_e(y_2 \mid x) - H_e(y \mid x)) \\ &\quad = \frac{|v_e'(u \mid x)|}{h_e(u \mid x)}\big(H_e(y_2 \mid x) - H_e(y \mid x)\big) \\ &\quad = \frac{|h_e(u \mid x) + uh_e'(u \mid x)|}{h_e(u \mid x)}\big(H_e(y_2 \mid x) - H_e(y \mid x)\big) \\ &\quad \leq \left(1 + \frac{\sup_u |u^2 h_e'(u \mid x)|}{\inf_{u \in A_x} |uh_e(u \mid x)|}\right)\big(H_e(y_2 \mid x) - H_e(y \mid x)\big) \\ &\quad \leq K\bar{a}_n^{-1/2}\big(H_e(y_2 \mid x) - H_e(y \mid x)\big), \end{aligned}$$

where the u in the first equality is between y and y_2. Finally, consider the case where both $y_1, y_2 \in A_x$. If we have that $y \in A_x$ for all y between y_1 and y_2 then (A.10) shows that $|y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x)| \leq K\bar{a}_n^{-1/2}(H_e(y_2 \mid x) - H_e(y_1 \mid x))$. It remains to consider the possibility that there exists a y between y_1 and y_2 with $y \notin A_x$. Let $u_1 (u_2)$ be the smallest (largest) number between y_1 and y_2 such that $u_1 h_e(u_1 \mid x) = \bar{a}_n^{1/2}$ (and similarly for u_2). Also assume without loss of generality that $y_1 < y_2$. Then $|y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x)| \leq K\bar{a}_n^{1/2}$ follows from the decomposition $|y_2 h_e(y_2 \mid x) - y_1 h_e(y_1 \mid x)| \leq |y_2 h_e(y_2 \mid x) - u_2 h_e(u_2 \mid x)| + |u_2 h_e(u_2 \mid x) - u_1 h_e(u_1 \mid x)| + |u_1 h_e(u_1 \mid x) - y_1 h_e(y_1 \mid x)|$ and (A.10). □

COROLLARY A.5. *Assume* (A1) (i), (iii), (iv), (A2) (i), (ii), (A3) (ii), F_X *is twice continuously differentiable* $\inf_{x \in R_X} f_X(x) > 0$, $H(y \mid x)$ *and* $H_1(y \mid x)$ *satisfy* (A6) (i), (iv), (v) *and* $H_e(y \mid x)$ *and* $H_{e1}(y \mid x)$ *satisfy* (A6) (ii), (iii), (vi). *Then*,

$$\text{(A.11)} \quad \sup_{-\infty < y \leq T}\left|\int_{-\infty}^{y}\left[\frac{1}{1 - \hat{H}_e(s)} - \frac{1}{1 - H_e(s)}\right] d(\hat{H}_{e1}(s) - H_{e1}(s))\right| = o_P(n^{-1/2}).$$

PROOF. Partitioning the interval $(-\infty, T]$ into k_n subintervals $[y_i, y_{i+1}]$ such that $H_e(y_{i+1}) - H_e(y_i) \leq (na_n)^{-1/2}(\log a_n^{-1})^{1/2} = \bar{a}_n$, where $k_n =$

$O((na_n)^{1/2}(\log a_n^{-1})^{-1/2})$, we have that the integral in (A.11) is bounded by

$$k_n \sup\left\{\left|\frac{1}{1-\hat{H}_e(y)} - \frac{1}{1-H_e(y)}\right|; -\infty < y \le T\right\}$$
$$\times \sup\left\{|\hat{H}_{e1}(y_2) - \hat{H}_{e1}(y_1) - H_{e1}(y_2) + H_{e1}(y_1)|; (y_1, y_2) \in J_{\bar{a}_n}\right\}$$
$$+2\sup\left\{\left|\frac{1}{1-\hat{H}_e(y_2)} - \frac{1}{1-\hat{H}_e(y_1)} - \frac{1}{1-H_e(y_2)} + \frac{1}{1-H_e(y_1)}\right|; (y_1, y_2) \in J_{\bar{a}_n} \cap (-\infty, T]^2\right\},$$

where $J_{\bar{a}_n} = \{(y_1, y_2); |H_e(y_2) - H_e(y_1)| \le \bar{a}_n\}$. Applying Proposition A.3 and an analogue of Proposition A.4 for the distribution H_{e1}, the first term above is easily seen to be $o_P(n^{-1/2})$. Again using Proposition A.3, the expression inside the supremum of the second term can be written as

$$\begin{aligned}
&\left|\big(1 - H_e(y_1)\big)^{-2}\big(\hat{H}_e(y_1) - H_e(y_1)\big) - \big(1 - H_e(y_2)\big)^{-2}\big(\hat{H}_e(y_2) - H_e(y_2)\big)\right| \\
&\qquad + O\big((na_n)^{-1}\log a_n^{-1}\big) \\
&\quad = \big(1 - H_e(y_1)\big)^{-2}\big|\hat{H}_e(y_2) - \hat{H}_e(y_1) - H_e(y_2) + H_e(y_1)\big| \\
&\qquad + O((na_n)^{-1}\log a_n^{-1}) \quad \text{a.s.,}
\end{aligned}$$

uniformly on $(-\infty, T]$. This is $o_P(n^{-1/2})$ by Proposition A.4. □

APPENDIX B

Results needed for Theorem 3.5.

LEMMA B.1. *Let the assumptions imposed in Theorem* 3.5 *hold. Then*,

$$n^{-1/2} \sup_{\substack{x \in R_X \\ y \in \mathbb{R}}} \left|\sum_{i=1}^{n}\left\{\varphi\left(X_i, Z_i, \Delta_i, \frac{y \wedge T_x - \hat{m}(x)}{\hat{\sigma}(x)}\right) - \varphi\left(X_i, Z_i, \Delta_i, \frac{y \wedge T_x - m(x)}{\sigma(x)}\right)\right\}\right| \to_P 0.$$

PROOF. Consider the classes $\mathscr{F}_1 = \{\varphi_{1y}(X, Z, \Delta); -\infty < y \le \bar{T}\}$ and $\mathscr{F}_2 = \{\varphi_{2y}(X, Z, \Delta); -\infty < y \le \bar{T}\}$, where $\varphi_{1y}(x, z, \delta) = \xi_e((z - m(x))/\sigma(x), \delta, y)$, $\varphi_{2y}(x, z, \delta) = -S_e(y)\eta(z, \delta\,|\,x)\gamma_1(y\,|\,x) - S_e(y)\zeta(z, \delta\,|\,x)\gamma_2(y\,|\,x)$ and $T < \bar{T} < \tau_{H_e}$. [Note that $\varphi(x, z, \delta, y) = \varphi_{1y}(x, z, \delta) + \varphi_{2y}(x, z, \delta)$.] In the proof of Corollary 3.2 it was shown that the classes $\mathscr{F}_1$ and $\mathscr{F}_2$ are Donsker.

Hence, it follows from Theorem 2.3.12 in van der Vaart and Wellner (1996) that for every decreasing sequence $\delta_n \downarrow 0$,

$$\text{(B.1)}\quad n^{-1/2}\sup{}^*\left|\sum_{i=1}^n \{\varphi_{ky_2}(X_i, Z_i, \Delta_i) - \varphi_{ky_1}(X_i, Z_i, \Delta_i)\}\right| \to_P 0, \qquad k = 1, 2,$$

where $\sup^*$ is the supremum over all (y_1, y_2) such that $\mathrm{Var}(\varphi_{ky_2} - \varphi_{ky_1}) \le \delta_n$. We will show that for n large, this condition is satisfied for all pairs of the form $(e, \hat{e}) = ((y \wedge T_x - m(x))/\sigma(x), (y \wedge T_x - \hat{m}(x))/\hat{\sigma}(x))$. Let $\delta_n = (na_n)^{-1/2}(\log a_n^{-1})^{1/2}$. First note that, by Proposition 4.5, for n large, $|F_e(\hat{e}) - F_e(e)| \le \delta_n(K_1 \sup_y f_e(y) + K_2 \sup_y |yf_e(y)|)$ and similarly for H_{e1}, where $K_1, K_2 > 0$ do not depend on x or y. Using this, and Proposition 4.5, it is easy to see that $\sup_{t,z,\delta} |\varphi_{2\hat{e}}(t, z, \delta) - \varphi_{2e}(t, z, \delta)| \le C_2\delta_n$ for some $C_2 > 0$, which implies that $\mathrm{Var}(\varphi_{2\hat{e}} - \varphi_{2e}) \le C_2^2\delta_n^2$. This argument cannot be used for $(\varphi_{1\hat{e}} - \varphi_{1e})(t, z, \delta)$, because φ_{1y} is not continuous in y. Note, however, that since the function ξ_e is the function ξ in the notation of Lo and Singh (1986), it follows that for any y_1, y_2, $\mathrm{Cov}(\varphi_{1y_1}, \varphi_{1y_2}) = (1 - F_e(y_1))(1 - F_e(y_2)) \int_{-\infty}^{y_1 \wedge y_2} dH_{e1}(s)/(1 - H_e(s))^2$, and hence $\mathrm{Var}(\varphi_{1\hat{e}} - \varphi_{1e}) = \mathrm{Var}(\varphi_{1e}) + \mathrm{Var}(\varphi_{1\hat{e}}) - 2\,\mathrm{Cov}(\varphi_{1e}, \varphi_{1\hat{e}}) \le C_1\delta_n$ for some $C_1 > 0$. The result now follows from (B.1). □

LEMMA B.2. *Let the assumptions imposed in Theorem* 3.5 *hold. Then*

$$\begin{aligned} n^{-3/2}a_n^{-1} \sup_{\substack{x \in R_X \\ y \in \mathbb{R}}} \Bigg| \sum_{i=1}^n \sum_{j=1}^n K\left(\frac{X_j - X_i}{a_n}\right) & \\ \times \Bigg[f_X^{-1}(X_j)\eta(Z_i, \Delta_i \mid X_j) f_e\left(\frac{y \wedge T_{X_j} - m(X_j)}{\sigma(X_j)}\right) & \\ - f_X^{-1}(X_i)\eta(Z_i, \Delta_i \mid X_i) f_e\left(\frac{y \wedge T_{X_i} - m(X_i)}{\sigma(X_i)}\right)\Bigg] I(X_j \le x) \Bigg| & \\ \to 0 \quad a.s. & \end{aligned}$$

PROOF. Define $h(x, z, \delta, y) = f_X^{-1}(x)\eta(z, \delta \mid x) f_e((y - m(x))/\sigma(x))$. Then,

$$\begin{aligned} h(X_j, Z_i, \Delta_i, y \wedge T_{X_j}) &= h(X_j, Z_i, \Delta_i, y) I(y \le T_{X_j}) + h(X_j, Z_i, \Delta_i, T_{X_j}) I(y > T_{X_j}) \\ &= h(X_j, Z_i, \Delta_i, y) I(y \le T_{X_i}) + h(X_j, Z_i, \Delta_i, T_{X_j}) I(y > T_{X_i}) \\ &\quad + [h(X_j, Z_i, \Delta_i, y) - h(X_j, Z_i, \Delta_i, T_{X_j})] \\ &\qquad \times [I(y \le T_{X_j}) - I(y \le T_{X_i})] \\ &= T_{ij}^{(1)}(y) + T_{ij}^{(2)}(y). \end{aligned}$$

The term $T_{ij}^{(2)}(y)$ differs from zero only if y is between T_{X_i} and T_{X_j}. Since $|T_{X_j} - T_{X_i}| = O(a_n)$ if $K((X_j - X_i)/a_n) \ne 0$, it follows that $|T_{X_j} - y| = O(a_n)$

in that case. Hence, using a one-term Taylor expansion,

$$n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{X_j-X_i}{a_n}\right)T_{ij}^{(2)}(y)I(X_j\le x)=O(a_n^2)$$

uniformly in x and y. It now follows that

$$\begin{aligned}
&n^{-2}a_n^{-1}\sum_{i,\,j=1}^{n}K\left(\frac{X_j-X_i}{a_n}\right)\\
&\qquad\times\left[h(X_j,Z_i,\Delta_i,y\wedge T_{X_j})-h(X_i,Z_i,\Delta_i,y\wedge T_{X_i})\right]I(X_j\le x)\\
&\quad=n^{-2}a_n^{-1}\sum_{i,\,j=1}^{n}K\left(\frac{X_j-X_i}{a_n}\right)\\
&\qquad\qquad\times\left[h(X_j,Z_i,\Delta_i,y)-h(X_i,Z_i,\Delta_i,y)\right]I(y\le T_{X_i})I(X_j\le x)\\
&\qquad+n^{-2}a_n^{-1}\sum_{i,\,j=1}^{n}K\left(\frac{X_j-X_i}{a_n}\right)\\
&\qquad\qquad\times\left[h(X_j,Z_i,\Delta_i,T_{X_j})-h(X_i,Z_i,\Delta_i,T_{X_i})\right]\\
&\qquad\qquad\times I(y>T_{X_i})I(X_j\le x).
\end{aligned}$$

We will show that the second term above is $o(n^{-1/2})$ a.s. The proof for the first term is completely analogous. Let $g(x,z,\delta)=h(x,z,\delta,T_x)$. Writing $g(X_j,Z_i,\Delta_i)-g(X_i,Z_i,\Delta_i)=(X_j-X_i)g'(X_i,Z_i,\Delta_i)+\frac{1}{2}(X_j-X_i)^2$ $g''(\xi_{ij},Z_i,\Delta_i)$ for some ξ_{ij} between X_i and X_j (and where $g'$and g'' denote, respectively, the first and second derivative of $g(x,z,\delta)$ with respect to x), it is clear that it suffices to consider

$$\begin{aligned}
&n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{X_j-X_i}{a_n}\right)(X_j-X_i)g'(X_i,Z_i,\Delta_i)\\
&\qquad\times I(y>T_{X_i})I(X_j\le x)\\
&\quad=n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{X_j-X_i}{a_n}\right)(X_j-X_i)g'(X_i,Z_i,\Delta_i)\\
&\qquad\qquad\times I(y>T_{X_i})I(X_i\le x)\\
&\qquad+n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{X_j-X_i}{a_n}\right)(X_j-X_i)g'\big(X_i,Z_i,\Delta_i\big)I(y>T_{X_i})\\
&\qquad\qquad\times\left[I(X_j\le x)-I(X_i\le x)\right].
\end{aligned}$$

The second term above is $O(a_n^2)$, since there are only $O(na_n)$ i's and $O(na_n)$ j's involved. The first one equals

$$
\begin{aligned}
&(na_n)^{-1}\sum_{i=1}^{n}\int K\left(\frac{u-X_i}{a_n}\right)(u-X_i)\,d(\hat F_X(u)-F_X(u))g'(X_i,Z_i,\Delta_i)\\
&\qquad\times I(y>T_{X_i})I(X_i\le x)\\
&\quad +O(a_n^2) \qquad\qquad \text{(B.2)}\\
&= n^{-3/2}a_n^{-1}\sum_{i=1}^{n}\int K\left(\frac{u-X_i}{a_n}\right)(u-X_i)\,d(\alpha_n(u)-\alpha_n(x_i^*))\\
&\qquad\times g'(X_i,Z_i,\Delta_i)I(y>T_{X_i})I(X_i\le x)+O(a_n^2),
\end{aligned}
$$

where $\alpha_n(u) = n^{1/2}(\hat F_X(u) - F_X(u))$ and $x_i^* \in R_X$ is such that $K((x_i^* - X_i)/a_n) \neq 0$. Making the substitution $v = (u - X_i)/a_n$, using integration by parts and the fact that $\sup_{|x_2-x_1|\le a_n} |\alpha_n(x_2) - \alpha_n(x_1)| = O(a_n^{1/2}(\log a_n^{-1})^{1/2})$ a.s. [see Theorem 0.2 in Stute (1982)], (B.2) is easily seen to be $o(n^{-1/2})$ a.s. This completes the proof of Lemma B.2. □

LEMMA B.3. *Let the assumptions imposed in Theorem* 3.5 *hold. Then*,

$$
\begin{aligned}
&\sup\left|\int_{-\infty}^{z}\int_{-\infty}^{x}(\hat F_X(x_1-va_n)-F_X(x_1-va_n))\right.\\
&\qquad\left.\times\, d(\hat F_D(x_1,z_1,\delta)-F_D(x_1,z_1,\delta))\right| = O_P(n^{-1}), \qquad \text{(B.3)}
\end{aligned}
$$

where the supremum is taken over all v belonging to the support of the kernel K, all $x \in R_X$ and all $-\infty < z < +\infty$, where $\hat F_D(x,z,\delta) - F_D(x,z,\delta) = n^{-1}\sum_{i=1}^n I(X_i \le x, Z_i \le z, \Delta_i = \delta) - P(X \le x, Z \le z, \Delta = \delta)$ and where $\delta = 0$ or 1 is fixed.

PROOF. We will apply Theorem 7 in Nolan and Pollard (1988) on the degenerate class of functions

$$
\begin{aligned}
\mathscr{F} = \Big\{&(v,x,z)\to I(X_1\le X_2-va_n, X_2\le x, Z_2\le z, \Delta_2=\delta)\\
&- F_X(X_2-va_n)I(X_2\le x, Z_2\le z,\Delta_2=\delta)\\
&- \int_{-\infty}^{x}\int_{-\infty}^{z} I(X_1\le x_1-va_n)\,dF_D(x_1,z_1,\delta)\\
&\quad + \int_{-\infty}^{x}\int_{-\infty}^{z} F_X(x_1-va_n)\,dF_D(x_1,z_1,\delta)\Big\}.
\end{aligned}
$$

Showing that the three displayed conditions in the aforementioned theorem are satisfied will imply the weak convergence of the process stated between absolute values in (B.3) and hence the result will follow. In what follows we will show that these conditions are satisfied for the class $\mathscr{F}_1 = \{(v,x,z) \to I(X_1 \le X_2 - va_n, X_2 \le x, Z_2 \le z, \Delta_2 = \delta)\}$. The other terms are dealt with in an easier way. First note that the envelope function F_1 equals 1. Let $\varepsilon > 0$ be fixed. For the first condition, we need to find a subset $\mathscr{F}_1^*$ of $\mathscr{F}_1$ (allowed to depend on the given sample), such that for each f in $\mathscr{F}_1$, there exists a f^* in $\mathscr{F}_1^*$, such that $E_{T_n} |f - f^*|^2 \le \varepsilon^2 n^2$, where the measure T_n is as defined on page 1293 in Nolan and Pollard (1988). If f (respectively, f^*) corresponds to the triplet (v, x, z) [respectively, (v^*, x^*, z^*)], this means that

$$\begin{aligned} E_{T_n}\big|&I(X_1 \le X_2 - va_n, X_2 \le x, Z_2 \le z, \Delta_2 = \delta) \\ &- I(X_1 \le X_2 - v^* a_n, X_2 \le x^*, Z_2 \le z^*, \Delta_2 = \delta)\big| \le \varepsilon^2 n^2. \end{aligned} \tag{B.4}$$

Partition the domain of $X_2 - X_1$ into $O(\varepsilon^{-2})$ subintervals $[v_j, v_{j+1}]$ such that the number of $X_{i_2} - X_{i_1}$'s between each $v_j a_n$ and $v_{j+1} a_n$ is less than $K_1 \varepsilon^2 n^2$ for some constant $K_1 > 0$ to be specified later. Divide in a similar way R_X into $O(\varepsilon^{-2})$ intervals $[x_k, x_{k+1}]$ such that $\hat{F}_X(x_{k+1}) - \hat{F}_X(x_k) \le K_2 \varepsilon^2$. Finally, divide the real line into $O(\varepsilon^{-2})$ subintervals $[z_l, z_{l+1}]$, such that the number of Z_i's between z_l and z_{l+1} is never more than $K_3 \varepsilon^2 n^2$. It is easily seen that, by proper choice of K_1, K_2 and K_3, for any (v, x, z) there exist j, k and l such that (B.4) is satisfied for $(v^*, x^*, z^*) = (v_j, x_k, z_l)$. This means that the covering number for the class $\mathscr{F}_1$ is $O(\varepsilon^{-6})$ and hence the covering integral $J(1, T_n, \mathscr{F}_1, F_1)$ is 6, uniformly over all samples and over all n. The second condition in Theorem 7 in Nolan and Pollard (1988) is satisfied by noting that for any $\gamma > 0$, $J(\gamma, T_n, \mathscr{F}_1, F_1) = 6\gamma(1 - \log \gamma)$, which equals 0 for $\gamma = e$. Finally, for the third, condition, we have to calculate both $N(\varepsilon, P \times P_n, \mathscr{F}_1, F_1)$ and $N(\varepsilon, P_n \times P, \mathscr{F}_1, F_1)$, since the functions of the class $\mathscr{F}_1$ are not symmetric in (X_1, Z_1) and (X_2, Z_2). For $N(\varepsilon, P \times P_n, \mathscr{F}_1, F_1)$, we can use the same partitions for x and z as before and for v we choose $O(\varepsilon^{-2})$ points v_j such that $|v_{j+1} - v_j| \le K_1 \varepsilon^2$ for some $K_1 > 0$. In a similar way as before, the class of functions corresponding to all triplets of the form (v_j, x_k, z_l) can be used to show that the covering number $N(\varepsilon, P \times P_n, \mathscr{F}_1, F_1)$ is $O(\varepsilon^{-6})$ uniformly over all samples and all n. Selecting the points v_j, x_k and z_l $(j, k, l = 1, \dots, O(\varepsilon^{-2}))$ satisfying $|v_{j+1} - v_j| \le K_1 \varepsilon^2$, $|x_{k+1} - x_k| \le K_2 \varepsilon^2$ and $|H(z_{l+1}) - H(z_l)| \le K_3 \varepsilon^2$ (where H is the distribution of Z), it is easy to show that also $N(\varepsilon, P_n \times P, \mathscr{F}_1, F_1)$ is $O(\varepsilon^{-6})$. This shows that also the third condition of Theorem 7 in Nolan and Pollard (1988) is satisfied and hence the result follows. □

The next result will be used in Proposition B.6 for establishing uniform convergence to zero of a U-process.

LEMMA B.4. *Let D, $\tilde{D}$ be random vectors of dimension r, $\tilde{r}$, respectively, and let $(D_i, \tilde{D}_i)$, $i = 1, \dots, n$, be a random sample drawn from the joint distribution of $(D, \tilde{D})$. Let $V(d_1, \tilde{d}_1, d_2, v)$ be a real-valued function defined on $\mathbb{R}^r \times \mathbb{R}^{\tilde{r}} \times \mathbb{R}^r \times T$, where $T = \mathbb{R}^r \times \mathbb{R}^s$ (where s can be zero) and where $v = (v_1, v_2)$ with $v_1 \in \mathbb{R}^r$ and $v_2 \in \mathbb{R}^s$. Assume that the function V satisfies $\sup_{d_1, \tilde{d}_1, d_2, v} |V(d_1, \tilde{d}_1, d_2, v)| < \infty$, $E_{D_1, \tilde{D}_1}(V(D_1, \tilde{D}_1, d_2, v)) = 0$ for all*

$d_2 \in \mathbb{R}^r$ and $v \in T$, that as a process in v_1, $\mathscr{B}_n(d, \tilde{d}, v_1, v_2) = n^{-1/2} \sum_{j=1}^n V(d, \tilde{d}, D_j, v_1, v_2)$ is a pure jump process with jumps at $v_1 = D_j$, and that $\mathscr{B}_n(d, \tilde{d}, v)$ converges weakly to a zero-mean Gaussian process $\mathscr{B}(d, \tilde{d}, v)$, which is continuous in all the indices d, $\tilde{d}$ and v. Then,

$$n^{-3/2} \sup_{v \in T} \left| \sum_{i=1}^n \sum_{j=1}^n V(D_i, \tilde{D}_i, D_j, v) \right| \to_P 0.$$

PROOF. The continuity of the limiting Gaussian process together with the Skorohod–Dudley–Wichura theorem [cf. Shorack and Wellner (1986), page 47], implies the existence of versions of the processes $\mathscr{B}_n(d, \tilde{d}, v)$ $(n \geq 1)$, and $\mathscr{B}(d, \tilde{d}, v)$ which are defined on the same probability space and such that $\sup_{d, \tilde{d}, v} |\mathscr{B}_n(d, \tilde{d}, v) - \mathscr{B}_n(d, \tilde{d}, v)| \to 0$, almost surely. [For simplicity, we denote the a.s. convergent versions of the processes by the same symbols as the original processes. We will do the same for the versions of the random vectors $(D_j, \tilde{D}_j)$, $j = 1, \ldots, n$ to be defined in the new space.] Note that in this new space $\mathscr{B}_n(d, \tilde{d}, v_1, v_2)$ is also a pure jump process in v_1 and the points where it jumps define the realization of the random vectors D_j. Thus the representation $\mathscr{B}_n(d, \tilde{d}, v_1, v_2) = n^{-1/2} \sum_{j=1}^n V(d, \tilde{d}, D_j, v_1, v_2)$ holds also in the new space. Let $\tilde{D}_j$ be a random variable generated according to the conditional distribution of $\tilde{D}$ given $D = D_j$. Assumption $E_{D_1, \tilde{D}_1}(V(D_1, \tilde{D}_1, d_2, v)) = 0$ implies

$$E_{D_1,\tilde{D}_1}\Big[\mathscr{B}_n(D_1, \tilde{D}_1, v)|\omega\Big] = \int \mathscr{B}_n(\omega, d_1, \tilde{d}_1, v)\, dF_{D_1, \tilde{D}_1}(d_1, \tilde{d}_1) = 0, \tag{B.5}$$

where conditioning on ω means conditioning on each sample path of the process $\mathscr{B}_n(d, \tilde{d}, v)$, and $\mathscr{B}_n(\omega, d, \tilde{d}, v)$ denotes the sample path. Consider now the process $n^{-1} \sum_{i=1}^n \mathscr{B}(D_i, \tilde{D}_i, v)$ as a process in v and write it as $n^{-1} \sum_{i=1}^n \mathscr{B}_v(\omega, D_i, \tilde{D}_i)$ to stress the fact that D_i and $\tilde{D}_i$, $i = 1, \ldots, n$, are not the only sources of randomness in this process. We will show the weak convergence of this process to the zero process. To show the convergence of the finite dimensional distributions, it suffices to show that for a fixed v, $n^{-1} \sum_{i=1}^n \mathscr{B}_v(\omega, D_i, \tilde{D}_i)$ as $P \to 0$. Write

$$\begin{aligned} E\Big[n^{-1} \sum_{i=1}^n \mathscr{B}_v(\omega, D_i, \tilde{D}_i)\Big]^2 &= n^{-2} \sum_{i_1 \neq i_2} E\Big[E\{\mathscr{B}_v(\omega, D_{i_1}, \tilde{D}_{i_1}) \,|\, \omega\} E\{\mathscr{B}_v(\omega, D_{i_2}, \tilde{D}_{i_2}) \,|\, \omega\}\Big] \\ &\quad + n^{-2} \sum_{i=1}^n E\big[\mathscr{B}_v^2(\omega, D_i, \tilde{D}_i)\big]. \end{aligned} \tag{B.6}$$

By adding and subtracting the zero term in (B.5), it can be seen that $|E(\mathscr{B}_v(\omega, D_i, \tilde{D}_i) \,|\, \omega)| \leq \sup_{d, \tilde{d}, v} |\mathscr{B}_n(\omega, d, \tilde{d}, v) - \mathscr{B}(\omega, d, \tilde{d}, v)| \to 0$, for almost all ω, which implies that $E(\mathscr{B}_v(\omega, D_i, \tilde{D}_i) \,|\, \omega) = 0$ a.s. Thus, the first term on the right-hand side of (B.6) equals zero. The second term of (B.6) is $O(n^{-1})$, since V is uniformly bounded over all variables. This shows that $n^{-1}\sum_{i=1}^{n}\mathscr{B}_v(\omega, D_i, \tilde{D}_i) = o_P(1)$ by Chebyshev's inequality. Finally, we consider

$$P\biggl(\sup_{\rho(v,v')<\delta}\biggl|n^{-1}\sum_{i=1}^{n}\bigl[\mathscr{B}_{v'}(\omega, D_i, \tilde{D}_i) - \mathscr{B}_v(\omega, D_i, \tilde{D}_i)\bigr]\biggr| > \varepsilon\biggr)$$
$$\leq \biggl(\sup_{\rho(v,v')<\delta}\sup_{d,\tilde{d}}\bigl|\mathscr{B}(d, \tilde{d}, v') - \mathscr{B}(d, \tilde{d}, v)\bigr| > \varepsilon\biggr).$$

Since $\mathscr{B}$ is a tight process, this shows the tightness of the process $n^{-1}\cdot \sum_{i=1}^{n}\mathscr{B}_v(\omega, D_i, \tilde{D}_i)$ and completes the proof. □

PROPOSITION B.5. *Let the assumptions imposed in Theorem* 3.5 *hold. Then,*

$$\begin{aligned} n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} & K\biggl(\frac{X_j - X_i}{a_n}\biggr) f_X^{-1}(X_j)\eta(Z_i, \Delta_i \mid X_j) \\ & \times f_e\biggl(\frac{y\wedge T_{X_j} - m(X_j)}{\sigma(X_j)}\biggr) I(X_j \leq x) \\ = n^{-1}\sum_{i=1}^{n} & \eta(Z_i, \Delta_i \mid X_i) f_e\biggl(\frac{y\wedge T_{X_i} - m(X_i)}{\sigma(X_i)}\biggr) I(X_i \leq x) + R_{n1}(x, y), \end{aligned}$$

where $\sup\{|R_{n1}(x,y)|; x \in R_X, y \in \mathbb{R}\} = o_P(n^{-1/2})$.

PROOF. Using Lemma B.2, it is clear that it suffices to show that

$$\begin{aligned} & n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K\biggl(\frac{X_j - X_i}{a_n}\biggr) f_X^{-1}(X_i)\eta(Z_i, \Delta_i \mid X_i) \\ & \qquad \times f_e\biggl(\frac{y\wedge T_{X_i} - m(X_i)}{\sigma(X_i)}\biggr) I(X_j \leq x) \\ & \quad - n^{-1}\sum_{i=1}^{n}\eta(Z_i, \Delta_i \mid X_i) f_e\biggl(\frac{y\wedge T_{X_i} - m(X_i)}{\sigma(X_i)}\biggr) I(X_i \leq x) \\ & = n^{-1}\sum_{i=1}^{n}\bigl[\hat{f}_X(X_i) - f_X(X_i)\bigr] f_X^{-1}(X_i)\eta(Z_i, \Delta_i \mid X_i) \\ & \qquad \times f_e\biggl(\frac{y\wedge T_{X_i} - m(X_i)}{\sigma(X_i)}\biggr) I(X_i \leq x) \\ & \quad + n^{-2}a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K\biggl(\frac{X_j - X_i}{a_n}\biggr) f_X^{-1}(X_i)\eta(Z_i, \Delta_i \mid X_i) \\ & \qquad \times f_e\biggl(\frac{y\wedge T_{X_i} - m(X_i)}{\sigma(X_i)}\biggr) \\ & \qquad \times \bigl[I(X_j \leq x) - I(X_i \leq x)\bigr] \end{aligned} \tag{B.7}$$

is $o_P(n^{-1/2})$ uniformly in (x, y), where $\hat f_X(x) = (na_n)^{-1}\sum_{i=1}^n K((x - X_i)/a_n)$ is an estimator for the density $f_X(x)$. We will prove that the first term on the right-hand side of (B.7) is $o_P(n^{-1/2})$. The proof for the second term is similar and will be left to the reader. Writing

$$\begin{aligned}\hat f_X(x) - f_X(x) &= a_n^{-1}\int K\left(\frac{x-u}{a_n}\right) d(\hat F_X(u) - F_X(u)) + O(a_n^2)\\ &= a_n^{-1}\int(\hat F_X(x - va_n) - F_X(x - va_n))K'(v)\,dv + O(a_n^2),\end{aligned}$$

and introducing the notation $\hat F_D(D) - F_D(D)$ for the empirical process that corresponds to the data $D_i = (X_i, Z_i, \Delta_i)$, $i = 1, \dots, n$, and the notation

$$h_y(x_1, z_1, \delta_1) = f_X^{-1}(x_1)\eta(z_1, \delta_1 \mid x_1) f_e\left(\frac{y \wedge T_{x_1} - m(x_1)}{\sigma(x_1)}\right),$$

the first term on the right-hand side of (B.7) can be written as

$$\begin{aligned}&a_n^{-1}\int(\hat F_X(x_1 - va_n) - F_X(x_1 - va_n))h_y(x_1, z_1, \delta_1)\\ &\qquad \times I(x_1 \le x)\,d(\hat F_D(D_1) - F_D(D_1))K'(v)\,dv.\end{aligned}$$

Further, let

$$\begin{aligned}A_v(x, z) = \int_{-\infty}^{z}\int_{-\infty}^{x} &n^{1/2}\big(\hat F_X(x_1 - va_n) - F_X(x_1 - va_n)\big)\,dn^{1/2}\\ &\times \big(\hat F_D(x_1, z_1, \delta) - F_D(x_1, z_1, \delta)\big)\end{aligned}$$

for fixed $\delta = 0, 1$. Then, it can be easily seen that it suffices to consider

$$\begin{aligned}&(na_n)^{-1}\int h_y(x_1, z_1, \delta)I(x_1 \le x)\,dA_v(x_1, z_1)K'(v)\,dv\\ &\quad = (na_n)^{-1}\int\int_{-\infty}^{z_1}\int_{x_L}^{x_1} dh_y(x_2, z_2, \delta)I(x_1 \le x)\,dA_v(x_1, z_1)K'(v)\,dv\\ &\qquad + (na_n)^{-1}\int h_y(x_L, z_1, \delta)I(x_1 \le x)\,dA_v(x_1, z_1)K'(v)\,dv\\ &\qquad + (na_n)^{-1}\int\int_{x_L}^{x_1} dh_y(x_2, -\infty, \delta)I(x_1 \le x)\,dA_v(x_1, z_1)K'(v)\,dv,\end{aligned}\tag{B.8}$$

where x_L is the left endpoint of the support of R_X. The first term on the right-hand side of (B.8) can be written as

$$\begin{aligned}&(na_n)^{-1}\int[A_v(x, +\infty) - A_v(x, z_2) - A_v(x_2, +\infty) + A_v(x_2, z_2)]\\ &\qquad \times I(x_2 > x_L)\,dh'_y(x_2, z_2, \delta)\,dx_2 K'(v)\,dv\end{aligned}$$

[where $h'_y(x_2, z_2, \delta)$ denotes the partial derivative of $h_y(x_2, z_2, \delta)$ with respect to x_2] and this is bounded by [see, for example, Lemma B, page 254 in Serfling (1980)] $K(na_n)^{-1}\sup_{x,z,v}|A_v(x, z)|\sup_{y,x_2}\|h'_y(x_2, z_2, \delta)\|_V$ for some $K > 0$

[where $\|h'_y(x_2, z_2, \delta)\|_V$ is the variation norm of the function $h'_y(x_2, z_2, \delta)$ considered as a function in z_2], which is $O_P((na_n)^{-1})$ by Lemma B.3 and assumption (A4) (ii). The derivation for the second and third term of (B.8) is similar, but easier. This completes the proof. $\square$

PROPOSITION B.6. *Let the assumptions imposed in Theorem* 3.5 *hold*. *Then*,

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\varphi\left(X_i, Z_i, \Delta_i, \frac{y\wedge T_{X_j}-m(X_j)}{\sigma(X_j)}\right)I(X_j\le x)$$
$$= n^{-1}\sum_{i=1}^{n}E\left\{\varphi\left(X_i, Z_i, \Delta_i, \frac{y\wedge T_X-m(X)}{\sigma(X)}\right)I(X\le x)\middle| X_i, Z_i, \Delta_i\right\}$$
$$+ R_{n2}(x, y),$$

where $\sup\{|R_{n2}(x, y)|; x\in R_X, y\in\mathbb{R}\} = o_P(n^{-1/2})$.

PROOF. We start with the second term of φ on which we will apply Lemma B.4 with $D = X$, $\tilde{D} = (Z, \Delta)$, $v = (x, y)$ and

$$V(d_1, \tilde{d}_1, d_2, u) = V(x_1, z_1, \delta_1, x_2, x, y)$$
$$= S_e\left(\frac{y\wedge T_{x_2}-m(x_2)}{\sigma(x_2)}\right)\eta(z_1, \delta_1\,|\,x_1)\gamma_1\left(\frac{y\wedge T_{x_2}-m(x_2)}{\sigma(x_2)}\middle| x_1\right)I(x_2\le x)$$
$$-E\left[S_e\left(\frac{y\wedge T_{X_2}-m(X_2)}{\sigma(X_2)}\right)\eta(z_1, \delta_1\,|\,x_1)\gamma_1\right.$$
$$\left.\times\left(\frac{y\wedge T_{X_2}-m(X_2)}{\sigma(X_2)}\middle| x_1\right)I(X_2\le x)\right].$$

Since $E[\xi(Z, \Delta, y\,|\,X)\,|\,X] = 0$, we also have that $E[\eta(Z, \Delta\,|\,X)\,|\,X] = 0$ and hence it suffices for the second term of φ to show the weak convergence of the process $n^{-1/2}\sum_{j=1}^{n} V(x_1, z_1, \delta_1, X_j, x, y)$ to a zero-mean Gaussian process. This will be done by showing that the class $\mathscr{F} = \{x_2\to V^*(x_1, z_1, \delta_1, x_2, x, y)\}$ is Donsker, where V^* equals the first term of V [see van der Vaart and Wellner (1996), page 81]. For this we need to show that $\int_0^\infty\sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, L_2(P))}\,d\varepsilon<\infty$, where $N_{[]}$ denotes the bracketing number (see Theorem 2.5.6 in the same book). Since the function $x_2\to S_e((y-m(x_2))/\sigma(x_2))$ is bounded and continuously differentiable, it follows from Corollary 2.7.2 in the aforementioned book that $m = O(\exp(K\varepsilon^{-1}))$ brackets are required for the class $\{x_2\to S_e((y-m(x_2))/\sigma(x_2))\}$. Hence, by truncating these brackets at $S_e((T_{x_2}-m(x_2))/\sigma(x_2))$, the same number is needed for the class $\{x_2\to S_e((y\wedge T_{x_2}-m(x_2))/\sigma(x_2))\}$, since $S_e((y\wedge T_{x_2}-m(x_2))/\sigma(x_2)) = S_e((y-m(x_2))/\sigma(x_2))\vee S_e((T_{x_2}-m(x_2))/\sigma(x_2))$. For γ_1, we first divide R_X into $O(\varepsilon^{-1})$ subintervals $[t_i, t_{i+1}]$ such that $|t_{i+1}-t_i|\le K\varepsilon$ for some $K>0$. For each fixed i, there exist m brackets that cover the class $\{x_2\to\gamma_1((y-m(x_2))/\sigma(x_2)\,|\,t_i); y\in\mathbb{R}\}$. Truncating these brackets at $\gamma_1((T_{x_2}-m(x_2))/\sigma(x_2)\,|\,t_i)$ shows that the bracketing number for the truncated class $\{x_2 \to \gamma_1((y \wedge T_{x_2} - m(x_2))/\sigma(x_2) \mid t_i); y \in \mathbb{R}\}$ is also m. Using the (more general) definition of bracketing number stated on page 211 in van der Vaart and Wellner (1996), it is clear that the class $\{x_2 \to \gamma_1((y \wedge T_{x_2} - m(x_2))/\sigma(x_2) \mid x_1); x_1 \in R_X, y \in \mathbb{R}\}$ can be covered by using all the $O(\varepsilon^{-1}\exp(K\varepsilon^{-1}))$ brackets corresponding to all the points t_i ($i = 1, \ldots, O(\varepsilon^{-1})$), since for any $t_i \le x_1, x_1' \le t_{i+1}$ and for any $y, y' \in \mathbb{R}$,

$$\begin{aligned}
&\gamma_1\biggl(\frac{y' \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| x_1'\biggr) - \gamma_1\biggl(\frac{y \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| x_1\biggr)\\
&\quad \le \gamma_1\biggl(\frac{y' \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| x_1'\biggr) - \gamma_1\biggl(\frac{y' \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| t_i\biggr)\\
&\qquad + \gamma_1\biggl(\frac{y' \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| t_i\biggr) - \gamma_1\biggl(\frac{y \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| t_i\biggr)\\
&\qquad + \gamma_1\biggl(\frac{y \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| t_i\biggr) - \gamma_1\biggl(\frac{y \wedge T_{x_2} - m(x_2)}{\sigma(x_2)}\bigg| x_1\biggr).
\end{aligned}$$

The L_2-norm of each of the three terms above is less than (a constant times) ε, provided y and y' belong to the same bracket from the class of m brackets determined by t_i. Finally, since $\eta(z_1, \delta_1 \mid x_1)I(x_2 \le x)$ is increasing in x_2 and bounded, it requires $O(\exp(K\varepsilon^{-1}))$ brackets by Theorem 2.7.5 in van der Vaart and Wellner (1996). This shows the weak convergence of the process $n^{-1/2}\sum_j V(x_1, z_1, \delta_1, X_j, x, y)$. The third term of φ is dealt with in a similar way. For the first term, we use the following decomposition of the function ξ_e:

$$\begin{aligned}
\xi_e(E, \Delta, y) &= (1 - F_e(y))\biggl\{-\int_{-\infty}^{E\wedge y} \frac{dH_{e1}(s)}{(1 - H_e(s))^2} + \int_{-\infty}^{y} \frac{dH_{e1}(s)}{1 - H_e(s)}\biggr\}\\
&\quad + (1 - F_e(y))\biggl\{\frac{I(E \le y, \Delta = 1)}{1 - H_e(E)} - \int_{-\infty}^{y} \frac{dH_{e1}(s)}{1 - H_e(s)}\biggr\}\\
&= \xi_{e1}(E, \Delta, y) + \xi_{e2}(E, \Delta, y).
\end{aligned}$$

For $n^{-2}\sum_{i,j}\xi_{e1}(E_i, \Delta_i, (y \wedge T_{X_j} - m(X_j))/\sigma(X_j))I(X_j \le x)$, the same arguments as for the second term of φ show that this term is asymptotically equivalent to $n^{-1}\sum_{i=1}^n E\{\xi_{e1}(E_i, \Delta_i, (y \wedge T_X - m(X))/\sigma(X))I(X \le x) \mid E_i, \Delta_i\}$. To show that the class

$$\biggl\{t \to \biggl(1 - F_e\biggl(\frac{y \wedge T_t - m(t)}{\sigma(t)}\biggr)\biggr)\int_{-\infty}^{(y\wedge T_t - m(t))/\sigma(t)} \frac{I(s \le E)}{(1 - H_e(s))^2}\, dH_{e1}(s) I(t \le x)\biggr\}$$

is Donsker, use similar arguments as above for $F_e((y \wedge T_t - m(t))/\sigma(t))$ and for $I(t \le x)$, and apply Corollary 2.7.2 in van der Vaart and Wellner (1996) on the integral. On the term ξ_{e2} we will apply Theorem 7 in Nolan and Pollard

(1988). Let

$$\mathscr{F} = \bigg\{(x, y) \to \xi_{e2}\bigg(\frac{Z_1 - m(X_1)}{\sigma(X_1)}, \Delta_1, \frac{y \wedge T_{X_2} - m(X_2)}{\sigma(X_2)}\bigg) I(X_2 \le x) - E\bigg[\xi_{e2}\bigg(\frac{Z_1 - m(X_1)}{\sigma(X_1)}, \Delta_1, \frac{y \wedge T_{X_2} - m(X_2)}{\sigma(X_2)}\bigg) I(X_2 \le x)\bigg| X_1, Z_1, \Delta_1\bigg]; x \in R_X, y \in \mathbb{R}\bigg\}.$$

Note that this is a degenerate class of functions. We will restrict ourselves to verifying the conditions in that theorem for the class:

$$\mathscr{F}_1 = \bigg\{(x, y) \to \bigg(1 - F_e\bigg(\frac{y \wedge T_{X_2} - m(X_2)}{\sigma(X_2)}\bigg)\bigg)\bigg(1 - H_e\bigg(\frac{Z_1 - m(X_1)}{\sigma(X_1)}\bigg)\bigg)^{-1} \times I\bigg(\frac{Z_1 - m(X_1)}{\sigma(X_1)} \le \frac{y \wedge T_{X_2} - m(X_2)}{\sigma(X_2)}, \Delta_1 = 1\bigg) I(X_2 \le x); x \in R_X, y \in \mathbb{R}\bigg\}.$$

The verification of the conditions in Theorem 7 in Nolan and Pollard (1988) for the other terms is similar, but easier. Let $F_1 \equiv (1 - H_e(T))^{-1}$ denote the envelope function for this class and let $\varepsilon > 0$. We start with showing that there exists a partition $[y_i, y_{i+1}]$ of the real line such that for all i, all $y_i \le y \le y_{i+1}$ and all $x \in R_X$,

$$F_e\bigg(\frac{y \wedge T_x - m(x)}{\sigma(x)}\bigg) - F_e\bigg(\frac{y_i \wedge T_x - m(x)}{\sigma(x)}\bigg) \le \varepsilon. \tag{B.9}$$

To see this, first note that for ε small enough $F_e^{-1}(\varepsilon) \ge \varepsilon^{-1}$ and $F_e^{-1}(1-\varepsilon) \le \varepsilon^{-1}$. Hence, all $y \le -\varepsilon^{-1} \sup_x \sigma(x) + \inf_x m(x) = a_\varepsilon$ satisfy $F_e((y - m(x))/\sigma(x)) \le \varepsilon$ (for all x) and all $y \ge \varepsilon^{-1} \sup_x \sigma(x) + \sup_x m(x) = b_\varepsilon$ satisfy $F_e((y - m(x))/\sigma(x)) \ge 1 - \varepsilon$ (for all x). Now divide the interval $[a_\varepsilon, b_\varepsilon]$ into $K = O(\varepsilon^{-2})$ subintervals $[y_i, y_{i+1}]$ ($y_1 = a_\varepsilon$, $y_K = b_\varepsilon$) such that $|y_{i+1} - y_i| \le \varepsilon(\sup_z f_e(z))^{-1} \inf_x \sigma(x)$. Then, if $y_0 = -\infty$ and $y_{K+1} = +\infty$, it is readily shown that (B.9) is satisfied. Next, let us divide the line into $O(\varepsilon^{-2})$ subintervals $[\tilde{y}_k, \tilde{y}_{k+1}]$ such that the number of couples (i, j) for which $\tilde{y}_k \le ((Z_i - m(X_i))/\sigma(X_i))\sigma(X_j) + m(X_j) \le \tilde{y}_{k+1}$ is less than $\varepsilon^2 n^2$. Finally, partition R_X into $O(\varepsilon^{-2})$ subintervals $[x_i, x_{i+1}]$ such that $\hat{F}_X(x_{i+1}) - \hat{F}_X(x_i) \le \varepsilon^2$. Now let $\mathscr{F}_1^*$ be the subclass of $\mathscr{F}_1$ consisting of all functions for which the corresponding couple (x, y) equals (x_i, y_j) or $(x_i, \tilde{y}_k)$ ($i, j, k = 1, \ldots, O(\varepsilon^{-2})$). It is readily seen that for any f in $\mathscr{F}_1$, there is a f^* in $\mathscr{F}_1^*$ for which $E_{T_n}|f - f^*|^2 \le 3(1 - H_e(T))^{-2}\varepsilon^2 n^2$, where T_n is the measure defined on page 1293 in Nolan and Pollard (1988). This shows that the covering number $N(\varepsilon, T_n, \mathscr{F}_1, F_1)$ is $O(\varepsilon^{-4})$ uniformly over all samples and over all n and hence the covering integral $J(1, T_n, \mathscr{F}_1, F_1)$ is 4. Hence, the first condition in Theorem 7 in Nolan and Pollard (1988) is satisfied. The second one is easily verified by an application of Chebyshev's inequality and by noting that the covering integral $J(\gamma, T_n, \mathscr{F}_1, F_1) = 4\gamma(1 - \log \gamma)$ equals zero for $\gamma = e$. Finally, for the third condition, we calculate first the covering number $N(\varepsilon, P \times P_n, \mathscr{F}_1, F_1)$ (note that since the functions in the class $\mathscr{F}_1$ are not symmetric, we have to consider both $P_n \times P$ and $P \times P_n$). For this, the partitions x_i ($i = 1, \dots, O(\varepsilon^{-2})$) and y_j ($j = 1, \dots, O(\varepsilon^{-2})$) constructed above can be used here as well to take care of, respectively, the first and the last factor of the functions in $\mathscr{F}_1$. For the indicator $I((Z_1 - m(X_1))/\sigma(X_1) \le (y \wedge T_{X_2} - m(X_2))/\sigma(X_2), \Delta_1 = 1)$, we partition the real line into $O(\varepsilon^{-2})$ subintervals $[\bar{y}_k, \bar{y}_{k+1}]$ such that $H_{e1}((y \wedge T_x - m(x))/\sigma(x)) - H_{e1}((\bar{y}_k \wedge T_x - m(x))/\sigma(x)) \le \varepsilon$ for all $\bar{y}_k \le y \le \bar{y}_{k+1}$ and for all $x \in R_X$ [this can be done in a similar way as for equation (B.9)]. The subclass of $\mathscr{F}_1$ corresponding to all couples (x_i, y_j) and $(x_i, \bar{y}_k)$ can be used to show that the covering number is not more than $O(\varepsilon^{-4})$. For the covering number with respect to the measure $P_n \times P$, we again use the partition y_j ($j = 1, \dots, O(\varepsilon^{-2})$) for the factors $1 - F_e((y \wedge T_{X_2} - m(X_2))/\sigma(X_2))$. For $I(X_2 \le x)$ we divide R_X into $O(\varepsilon^{-2})$ intervals $[\tilde{x}_i, \tilde{x}_{i+1}]$ such that $F_X(\tilde{x}_{i+1}) - F_X(\tilde{x}_i) \le \varepsilon^2$. A more complicated partition is needed for the remaining factor, for which subintervals (say $[y_k^*, y_{k+1}^*]$) need to be constructed satisfying

$$\begin{aligned}\sum_{i=1}^n \bigg\{ & P\bigg(\frac{Z_i - m(X_i)}{\sigma(X_i)} \sigma(X) + m(X) \le y_{k+1}^* \bigg| X_i, Z_i \bigg) \\ & - P\bigg(\frac{Z_i - m(X_i)}{\sigma(X_i)} \sigma(X) + m(X) \le y_k^* \bigg| X_i, Z_i \bigg) \bigg\} \le \varepsilon^2 n. \end{aligned} \tag{B.10}$$

Let V be the number of $(Z_i - m(X_i))/\sigma(X_i)$ that are less than $H_e^{-1}(\varepsilon^2/2)$ or greater than $H_e^{-1}(1 - \varepsilon^2/2)$. Then $V \sim \mathrm{Bin}(n, \varepsilon^2)$ and hence, for n large, $V - n\varepsilon^2 \le 2(\varepsilon^2(1 - \varepsilon^2) n \log\log n)^{1/2}$ a.s. [see, e.g., Serfling (1980), page 35]. So, we only need to consider the terms in (B.10) for which $H_e^{-1}(\varepsilon^2/2) \le (Z_i - m(X_i))/\sigma(X_i) \le H_e^{-1}(1 - \varepsilon^2/2)$. As before, we have that $[H_e^{-1}(\varepsilon^2/2), H_e^{-1}(1-\varepsilon^2/2)] \subseteq [-\varepsilon^{-2}/2, \varepsilon^{-2}/2]$ for ε small enough. Hence, it suffices to consider values of y for which $y_L = -(\varepsilon^{-2}/2)\sup_x \sigma(x) + \inf_x m(x) \le y \le (\varepsilon^{-2}/2)\sup_x \sigma(x) + \sup_x m(x) = y_U$. Using assumption (A5), (B.10) can be achieved for these values of y by using at most $O(\varepsilon^{-4})$ subintervals $[y_k^*, y_{k+1}^*]$. This shows that also the last condition of Theorem 7 in Nolan and Pollard (1988) is satisfied and hence

$$\begin{aligned} n^{-1} \sup_{\substack{x \in R_X \\ y \in \mathbb{R}}} \bigg| \sum_{i=1}^n \sum_{j=1}^n \bigg\{ & \xi_{e2}\bigg(\frac{Z_i - m(X_i)}{\sigma(X_i)}, \Delta_i, \frac{y \wedge T_{X_j} - m(X_j)}{\sigma(X_j)} \bigg) I(X_j \le x) \\ & - E\bigg[\xi_{e2}\bigg(\frac{Z_i - m(X_i)}{\sigma(X_i)}, \Delta_i, \frac{y \wedge T_{X_j} - m(X_j)}{\sigma(X_j)} \bigg) I(X_j \le x) \bigg| X_i, Z_i, \Delta_i \bigg] \bigg\} \bigg| \\ = O_P(1). & \end{aligned}$$

This completes the proof. □

Acknowledgments. The authors thank Professor Noël Veraverbeke and the Limburgs Universitair Centrum for setting the stage for the collaboration that led to this work and for discussions related to this manuscript. A travel grant from the Fund for Scientific Research–Flanders (Belgium) is also gratefully acknowledged.

REFERENCES

AKRITAS, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.* **22** 1299–1327.

AKRITAS, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics* **52** 1342–1362.

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

CRISTÓBAL CRISTÓBAL, J. A., FARALDO ROCA, P. and GONZÁLEZ MANTEIGA, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.* **15** 603–609.

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

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

DEVROYE, L. P. (1978). The uniform convergence of the Nadaraya–Watson regression function estimate. *Canad. J. Statist.* **6** 179–191.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

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.

GASSER, T. and MÜLLER, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 171–185.

GONZÁLEZ MANTEIGA, W. and CADARSO SUAREZ, C. (1994). Asymptotic properties of a generalized Kaplan–Meier estimator with some applications. *J. Nonparametr. Statist.* **4** 65–78.

HÄRDLE, W. and GASSER, T. (1985). On robust kernel estimation of derivatives of regression functions. *Scand. J. Statist.* **12** 233–240.

HÄRDLE, W., JANSSEN, P. and SERFLING, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16** 1428–1449.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

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

LOYNES, R. M. (1980). The empirical sample distribution function of residuals from generalized regression. *Ann. Statist.* **8** 285–298.

MACK, Y. P. and MÜLLER, H.-G. (1989). Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā Ser. A* **51** 59–72.

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

MÜLLER, H. G. and STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625.

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

RYCHLIK, T. (1990). Unbiased nonparametric estimation of the derivative of the mean. *Statist. Probab. Lett.* **10** 329–333.

SCHUSTER, E. and YAKOWITZ, S. (1979). Contribution to the theory of nonparametric regression, with application to system identification. *Ann. Statist.* **7** 139–149.

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

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

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

STUTE, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45** 89–103.

STUTE, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23** 461–471.

VAN DER LAAN, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.* **24** 596–627.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes*. Springer, New York.

VAN KEILEGOM, I. (1998). *Nonparametric estimation of the conditional distribution in regression with censored data*. Ph.D. thesis, Limburgs Univ. Centrum.

VAN KEILEGOM, I. and AKRITAS, M. G. (1998). Transfer of tail information in censored regression models. Technical report, Limburgs Univ. Centrum.

VAN KEILEGOM, I., AKRITAS, M. G. and VERAVERBEKE, N. (1998). Estimation of the conditional distribution in regression with censored data: a comparative study. *Comput. Statist. Data Anal.* (in press).

VAN KEILEGOM, I. and VERAVERBEKE, N. (1996). Uniform strong convergence results for the conditional Kaplan–Meier estimator and its quantiles. *Comm. Statist. Theory Methods* **25** 2251–2265.

VAN KEILEGOM, I. and VERAVERBEKE, N. (1997a). Weak Convergence of the bootstrapped conditional Kaplan–Meier process and its quantile process. *Comm. Statist. Theory Methods* **26** 853–869.

VAN KEILEGOM, I. and VERAVERBEKE, N. (1997b). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.* **49** 467–491.

DEPARTMENT OF MATHEMATICS
EINDHOVEN UNIVERSITY OF TECHNOLOGY
P.O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
E-MAIL: i.v.keilegom@tue.nl

DEPARTMENT OF STATISTICS
PENN STATE UNIVERSITY
326 THOMAS BUILDING
UNIVERSITY PARK, PENNSYLVANIA 16802
E-MAIL: mga@stat.psu.edu