INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0036

DENSITY AND HAZARD ESTIMATION IN CENSORED REGRESSION MODELS

I. VAN KEILEGOM and N. VERAVERBEKE

http://www.stat.ucl.ac.be

Density and hazard estimation in censored regression models

Ingrid Van Keilegom[1]
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays 20
B-1348 Louvain-la-Neuve
Belgium

Noël Veraverbeke
Department of Mathematics
Limburgs Universitair Centrum
Universitaire Campus
B-3590 Diepenbeek
Belgium

November 21, 2000

Abstract

Let (X, Y) be a random vector, where Y denotes the variable of interest, possibly subject to random right censoring, and X is a covariate. Consider a heteroscedastic model $Y = m(X) + \sigma(X)\varepsilon$, where the error term ε is independent of X and $m(X)$ and $\sigma(X)$ are smooth but unknown functions. Further, assume that there exists a region of the covariate X where the censoring of Y is light. Under this model, we construct a nonparametric estimator for the density and hazard function of Y given X, which has a faster rate of convergence than the completely nonparametric estimator, that is constructed without making any model assumption. Moreover, the proposed estimator for the density and hazard function performs better in the right tail compared to the classical nonparametric estimator.

We prove the weak convergence of both the density and the hazard function estimator. The results are obtained by constructing asymptotic representations for the two estimators and by making use of Van Keilegom and Akritas (1999), where an estimator of the conditional distribution of Y given X is studied under the same model assumption.

KEY WORDS: Asymptotic representation; Density function; Hazard rate; Heteroscedastic regression; Right censoring; Weak convergence.

[1] This research was supported by ‘Projet d’Actions de Recherche Concertées’, No. 98/03–217, from the Belgian government.

1 Introduction

Let Y denote a possible transformation of the variable of interest and let X be a covariate. The response Y is allowed to be subject to random right censoring, while X is completely observed. It is assumed that the vector (X, Y) follows the heteroscedastic 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 function, representing possible heteroscedasticity. Under this model we estimate the density and hazard function of Y given X based on kernel methods.

Given that model (1.1) is the correct model, the estimators for the conditional density and hazard which we propose, have two important advantages over the completely nonparametric estimator (which does not make use of model (1.1)). First, there is a difference in the behavior in the right tail for the two types of estimators. The completely nonparametric kernel estimator for the conditional density or hazard at a given value x of X uses only data in a neighborhood of x and hence it behaves badly in the right tail whenever the censoring is heavy in a neighborhood of x. The proposed estimator is based on all available data points in the full range of X, not only those in a neighborhood of x. Because of this, the behavior in the right tail will be satisfactory, provided there is a region of the support of X where the censoring of Y is light.

The second advantage of these new estimators is their faster rate of convergence. The completely nonparametric kernel estimator for the conditional density or hazard requires smoothing both over the covariate space and over the time axis. As a consequence, the rate of convergence of such estimators is $O_P((na_n^2)^{-1/2})$ (for one-dimensional covariates), where a_n is the bandwidth used in the smoothing process (see e.g. McKeague and Utikal (1990) or Li and Doss (1995)). Although the proposed estimators also require smoothing in two directions, their rate of convergence is $O_P((na_n)^{-1/2})$.

Much research has been devoted to the estimation of the density and hazard function. We focus on recent papers based on kernel estimators. In the case of i.i.d. censored data (without covariates), we mention the work of Lo, Mack and Wang (1989), who estimated the density and hazard function based on the Kaplan-Meier estimator, and obtained an i.i.d. representation and the asymptotic normality of these estimators. Müller and Wang (1994) introduced boundary corrected kernels for a better estimation of the hazard function near the boundary of the covariate space. In a paper by Gefeller and Dette (1992) and Dette and Gefeller (1995), nearest neighbor weights are used for estimating the hazard function. In the context of censored response data considered in a regression model with fully observable covariates, the nonparametric estimation of the conditional hazard function has been studied by McKeague and Utikal (1990), Li and Doss (1995) and Li (1997), who use counting processes to obtain the asymptotic normality of their estimator. Gray (1996) uses binning methods to estimate the conditional hazard function.

The paper is organized as follows. In the next section, we give the definition of the estimator of the density and hazard function and we state the main assumptions under which the results will be derived. Section 3 describes the main results. Section 4 contains the proofs of the main results, while technical derivations are deferred to the Appendix.

2 Definitions and assumptions

Assume the vector (X, Y) satisfies the regression model (1.1), where the functions $m(\cdot)$ and $\sigma(\cdot)$ are respectively a location and scale functional. This means that there exist functionals S and T such that $m(x) = T(F_Y(\cdot|x))$, $\sigma(x) = S(F_Y(\cdot|x))$,

$$T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b \text{ and } S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x)),$$

for all $a \geq 0$, and $b \in I\!R$, where $F_Y(\cdot|x)$ denotes the distribution of Y given $X = x$. The response Y is allowed to be subject to random right censoring. Let C be the censoring random variable which is conditionally independent of Y given X, and suppose that the observable random vector is (X, Z, Δ), where $Z = \min(Y, C)$ and $\Delta = I(Y \leq C)$. Finally, let (X_i, Z_i, Δ_i) $(i = 1, \ldots, n)$ denote independent replications of (X, Z, Δ). We use the notation $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $H(y|x) = P(Z \leq y|x)$ and $H_1(y|x) = P(Z \leq y, \Delta = 1|x)$. Further denote $F_e(y) = P(\varepsilon \leq y)$, $G_e(y) = P(\frac{C-m(X)}{\sigma(X)} \leq y)$ and for $E = \frac{Z-m(X)}{\sigma(X)}$ we use the notation $H_e(y) = P(E \leq y)$, $H_{e1}(y) = P(E \leq y, \Delta = 1)$, $H_e(y|x) = P(E \leq y|x)$ and $H_{e1}(y|x) = P(E \leq y, \Delta = 1|x)$. The density functions of the above distribution functions will be denoted with lower case letters.

In order to estimate the conditional density function, we first need to estimate the conditional distribution function $F(y|x)$. Note that under model (1.1),

$$F(y|x) = F_e\left(\frac{y - m(x)}{\sigma(x)}\right). \tag{2.1}$$

To estimate $m(x)$ and $\sigma(x)$ we will work with the particular definitions

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

where $F^{-1}(s|x) = \inf\{t; F(t|x) \geq s\}$ 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 particular location and scale functionals then it holds for all location and scale functionals. Hence, working with the functionals $m(x)$ and $\sigma(x)$ in (2.2) is no restriction of generality. The estimation of these functions consists in replacing the unknown distribution $F(\cdot|x)$ with the following nonparametric estimator which is due to Beran (1981) :

$$\tilde{F}(y|x) = 1 - \prod_{Z_i \le y, \Delta_i = 1} \left\{ 1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \ge Z_i) W_j(x, a_n)} \right\}, \tag{2.3}$$

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

$$W_i(x, a_n) = \frac{K_1\left(\frac{x - X_i}{a_n}\right)}{\sum_{j=1}^n K_1\left(\frac{x - X_j}{a_n}\right)},$$

with K_1 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 if all weights $W_i(x, a_n)$ equal n^{-1} (i.e. if there are no covariates). This leads to

$$\hat{m}(x) = \int_0^1 \hat{F}^{-1}(s|x) J(s)\, ds, \;\; \hat{\sigma}^2(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 J(s)\, ds - \hat{m}^2(x), \tag{2.4}$$

as estimators for $m(x)$ and $\sigma(x)$. Let $\hat{E}_i = (Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$ and define

$$\hat{F}_e(y) = 1 - \prod_{\hat{E}_{(i)} \le y, \Delta_{(i)} = 1} \left(1 - \frac{1}{n - i + 1} \right), \tag{2.5}$$

where $\hat{E}_{(i)}$ is the i-th order statistic of $\hat{E}_1, \ldots, \hat{E}_n$ and $\Delta_{(i)}$ is the corresponding censoring indicator. Relation (2.1) now suggests

$$\hat{F}(y|x) = \hat{F}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)} \right) \tag{2.6}$$

as an estimator for $F(y|x)$. The estimator is an alternative for the Beran estimator $\tilde{F}(y|x)$, which often does not behave well in the right tail if the censoring is heavy. The estimator $\hat{F}(y|x)$ however behaves well in the right tail for all values of x, provided that there is a region of the covariate space where the censoring of Y is light. This is because in that case, the right tail of $F_e(\cdot)$ can be well estimated and hence, by (2.1), the same is true for the right tail of $F(\cdot|x)$ for all values of x, also for those for which the censoring of Y is heavy. The asymptotic properties of this estimator were studied in Van Keilegom and Akritas (1999). We now estimate the conditional density function $f(y|x)$ and the hazard function $\lambda(y|x) = \frac{f(y|x)}{1 - F(y|x)}$ by smoothing the estimator $\hat{F}(y|x)$ with a second kernel function K_2 (in order to make the results less technical we work with the same bandwidth as the one used for smoothing over the covariate space) :

$$\hat{f}(y|x) = a_n^{-1} \int K_2\left(\frac{y - t}{a_n} \right) d\hat{F}(t|x), \tag{2.7}$$

$$\hat{\lambda}(y|x) = \frac{\hat{f}(y|x)}{1 - \hat{F}(y|x)}. \tag{2.8}$$

The primary objective of this paper is to prove an asymptotic representation and the weak convergence of the estimators $\hat{f}(y|x)$ and $\hat{\lambda}(y|x)$.

The above estimator $\hat{f}(y|x)$ is obtained by first evaluating $F_e(y)$ in $\frac{y-m(x)}{\sigma(x)}$ (which leads to $F(y|x)$) and then smoothing $F(y|x)$. We could also start by smoothing $F_e(y)$, and then evaluating the result in $\frac{y-m(x)}{\sigma(x)}$. This leads to the following alternative estimator :

$$\hat{f}_{alt}(y|x) = \frac{1}{\hat{\sigma}(x)} \hat{f}_e \left(\frac{y - m(x)}{\sigma(x)} \right),$$

where $\hat{f}_e(y) = a_n^{-1} \int K(\frac{y-t}{a_n})\, d\hat{F}_e(t)$ is an estimator for the density $f_e(y)$ of ε. It can be seen that the estimators $\hat{f}(y|x)$ and $\hat{f}_{alt}(y|x)$ are not asymptotically equivalent, due to an extra contribution in the asymptotic representation of $\hat{f}_{alt}(y|x)$. In this paper we do not consider the estimator $\hat{f}_{alt}(y|x)$, but we note that its asymptotic properties can be obtained in a very similar way as for $\hat{f}(y|x)$.

The following functions enter in the asymptotic representation for $\hat{f}(y|x)$ and $\hat{\lambda}(y|x)$:

$$\xi_e(z, \delta, y) = (1 - F_e(y)) \left\{ - \int_{-\infty}^{y \wedge z} \frac{dH_{e1}(s)}{(1 - H_e(s))^2} + \frac{I(z \le y, \delta = 1)}{1 - H_e(z)} \right\},$$

$$\xi(z, \delta, y|x) = (1 - F(y|x)) \left\{ - \int_{-\infty}^{y \wedge z} \frac{dH_1(s|x)}{(1 - H(s|x))^2} + \frac{I(z \le y, \delta = 1)}{1 - H(z|x)} \right\},$$

$$\eta(z, \delta|x) = \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x))\, dv\, \sigma^{-1}(x),$$

$$\zeta(z, \delta|x) = \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x)) \frac{v - m(x)}{\sigma(x)}\, dv\, \sigma^{-1}(x),$$

$$\gamma_1(y|x) = \int_{-\infty}^{y} \frac{h_e(s|x)}{(1 - H_e(s))^2}\, dH_{e1}(s) + \int_{-\infty}^{y} \frac{d\, h_{e1}(s|x)}{1 - H_e(s)},$$

$$\gamma_2(y|x) = \int_{-\infty}^{y} \frac{sh_e(s|x)}{(1 - H_e(s))^2}\, dH_{e1}(s) + \int_{-\infty}^{y} \frac{d\, (sh_{e1}(s|x))}{1 - H_e(s)},$$

$$\varphi(x, z, \delta, y) = -(1 - F_e(y))\eta(z, \delta|x)\gamma_1(y|x) - (1 - F_e(y))\zeta(z, \delta|x)\gamma_2(y|x).$$

Further, we estimate $H_e(y)$ and $H_{e1}(y)$ by the empirical distribution functions

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

Let T, respectively T_x, be any value less than the upper bound of the support of $H_e(\cdot)$, respectively $H(\cdot|x)$, such that $\inf_{x \in R_X}(1 - H(T_x|x)) > 0$ and define $\Omega = \{(x, y); \frac{y-m(x)}{\sigma(x)} \le$

$T\}$. For a (sub)distribution function $L(y|x)$ we will use the notations $L'(y|x) = \frac{\partial}{\partial y} L(y|x)$, $\dot{L}(y|x) = \frac{\partial}{\partial x} L(y|x)$ and similar notations will be used for higher order derivatives. Further, let $\|K_i\|_2^2 = \int K_i^2(u)\,du$ and $\mu_2^{K_i} = \int u^2 K_i(u)\,du$ $(i = 1, 2)$.

The assumptions we need to impose in the proofs of the main results are listed below :

(A1)(i) $na_n^5(\log a_n^{-1})^{-1} = O(1)$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.
(ii) The support R_X of X is bounded, convex and its interior is not empty.
(iii) For $i = 1, 2$, the probability density function K_i has compact support $[-L_i, L_i]$, $\int uK_i(u)\,du = 0$ and K_i is twice continuously differentiable.

(A2)(i) There exist $0 \leq s_0 \leq s_1 \leq 1$ such that $s_1 \leq \inf_x F(T_x|x)$, $s_0 \leq \inf\{s \in [0,1]; J(s) \neq 0\}$, $s_1 \geq \sup\{s \in [0,1]; J(s) \neq 0\}$ and $\inf_{x \in R_X} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0$.
(ii) J is twice continuously differentiable, $\int_0^1 J(s)\,ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.

(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) $E|E|^5 = E|\frac{Z - m(X)}{\sigma(X)}|^5 < \infty$.

(A4) The functions $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ are twice continuously differentiable with respect to x and their first and second derivatives (with respect to x) are bounded, uniformly in $x \in R_X, z < T_x$ and δ.

(A5) For $L(y|x) = H(y|x), H_1(y|x), H_e(y|x)$ or $H_{e1}(y|x)$: $L'(y|x)$ is continuous in (x, y) and $\sup_{x,y} |y^2 L'(y|x)| < \infty$ and the same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order four.

3 Main results

Theorem 3.1 *Assume (A1)-(A5). Then,*

$$\hat{f}(y|x) - f(y|x) = (na_n)^{-1} \sum_{i=1}^n \int \xi_e \left(E_i, \Delta_i, \frac{y - va_n - m(x)}{\sigma(x)} \right) dK_2(v)$$
$$+ (na_n)^{-1} \sum_{i=1}^n K_1 \left(\frac{x - X_i}{a_n} \right) g_{x,y}(Z_i, \Delta_i) + a_n^2 b_f(y|x) + r_n(y|x),$$

where

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

$$b_f(y|x) = \frac{1}{2}\mu_2^{K_2}\sigma^{-1}(x)\int \left[\frac{\partial}{\partial y}E\left(\varphi\left(x, Z, \Delta, \frac{y-m(x)}{\sigma(x)}\right)\middle| X = u\right) f_X(u)\right]''\Bigg|_{u=x} dx$$
$$+ \frac{1}{2}\mu_2^{K_2} f''(y|x),$$

and $\sup\{|r_n(y|x)|; (x,y) \in \Omega\} = o_P((na_n)^{-1/2}) + o_P(a_n^2)$.

Next we show the local weak convergence of this density estimator in a neighborhood $y + a_n t$ ($-\tilde{T} \leq t \leq \tilde{T}$, $\tilde{T} > 0$ arbitrary) of a fixed point y. By choosing $t = 0$, the asymptotic normality of $\hat{f}(y|x)$ follows from this result. We need to restrict attention to this local type of weak convergence, since otherwise the tightness of the process cannot be established. This is a typical feature for processes of nonparametric density, hazard or regression function estimators and can also be found in e.g. Rosenblatt (1971).

Note that Theorems 3.2 and 3.4 below are formulated both for the optimal bandwidth case ($K > 0$) and the non-optimal case ($K = 0$).

Theorem 3.2 *Assume (A1)-(A5). If $na_n^5 = K$ for some $K \geq 0$, then the process $(na_n)^{1/2}(\hat{f}(y + a_n t|x) - f(y + a_n t|x))$ ($x \in R_X$ and $y \leq T\sigma(x) + m(x)$ fixed, $-\tilde{T} \leq t \leq \tilde{T}$, $\tilde{T} > 0$ arbitrary) converges weakly to a Gaussian process $Z_f(y,t|x)$ with mean function*

$$E(Z_f(y,t|x)) = K^{1/2}b_f(y|x) + \frac{1}{2}K^{1/2}\mu_2^{K_1}[E(g_{x,y}(Z,\Delta)|X = u)f_X(u)]''|_{u=x}$$

and covariance function

$$Cov(Z_f(y,t|x), Z_f(y,t'|x)) = f_X(x)\|K_1\|_2^2 \, Var(g_{x,y}(Z,\Delta)|X = x)$$
$$+ \sigma^{-1}(x)\int K_2(w)K_2(w + t - t')\,dw \frac{h_{e1}\left(\frac{y-m(x)}{\sigma(x)}\right)}{\left(1 - G_e\left(\frac{y-m(x)}{\sigma(x)}\right)\right)^2}.$$

Theorem 3.3 *Assume (A1)-(A5). Then,*

$$\hat{\lambda}(y|x) - \lambda(y|x)$$
$$= (na_n)^{-1}(1 - F(y|x))^{-1}\sum_{i=1}^n \int \xi_e\left(E_i, \Delta_i, \frac{y - va_n - m(x)}{\sigma(x)}\right) dK_2(v)$$
$$+ (na_n)^{-1}\sum_{i=1}^n K_1\left(\frac{x - X_i}{a_n}\right) k_{x,y}(Z_i, \Delta_i) + a_n^2 b_\lambda(y|x) + r_n(y|x),$$

where

$$k_{x,y}(z,\delta) = (1-F(y|x))^{-1}g_{x,y}(z,\delta) + f(y|x)(1-F(y|x))^{-2}h_{x,y}(z,\delta),$$

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

$$b_\lambda(y|x) = (1-F(y|x))^{-1}b_f(y|x) + \frac{1}{2}\mu_2^{K_1}(1-F(y|x))^{-2}f(y|x)$$
$$\times \int \left\{ E\left[\varphi\left(x, Z, \Delta, \frac{y-m(x)}{\sigma(x)}\right)\Big| X=u\right] f_X(u)\right\}''\Big|_{u=x} dx,$$

and $\sup\{|r_n(y|x)|; (x,y)\in\Omega\} = o_P((na_n)^{-1/2}) + o_P(a_n^2)$.

We finally give the weak convergence of the hazard estimator $\hat\lambda(y|x)$. The proof is very similar to that for the density function (see Theorem 3.2) and is therefore not shown.

Theorem 3.4 *Assume (A1)-(A5). If $na_n^5 = K$ for some $K \geq 0$, then the process $(na_n)^{1/2}(\hat\lambda(y+a_nt|x) - \lambda(y+a_nt|x))$ ($x \in R_X$ and $y \leq T\sigma(x)+m(x)$ fixed, $-\tilde T \leq t \leq \tilde T$, $\tilde T > 0$ arbitrary) converges weakly to a Gaussian process $Z_\lambda(y,t|x)$ with mean function*

$$E(Z_\lambda(y,t|x)) = K^{1/2}b_\lambda(y|x) + \frac{1}{2}K^{1/2}\mu_2^{K_1}[E(k_{x,y}(Z,\Delta)|X=u)f_X(u)]''|_{u=x}$$

and covariance function

$$Cov(Z_\lambda(y,t|x), Z_\lambda(y,t'|x)) = f_X(x)\|K_1\|_2^2\, Var(k_{x,y}(Z,\Delta)|X=x)$$
$$+\sigma^{-1}(x)\int K_2(w)K_2(w+t-t')\,dw \frac{h_{e1}\left(\frac{y-m(x)}{\sigma(x)}\right)}{\left(1-H_e\left(\frac{y-m(x)}{\sigma(x)}\right)\right)^2}.$$

4 Proofs

Proof of Theorem 3.1. Write

$$\hat f(y|x) - f(y|x)$$
$$= a_n^{-1}\int K_2\left(\frac{y-t}{a_n}\right) d\hat F(t|x) - f(y|x)$$
$$= a_n^{-1}\int K_2\left(\frac{y-t}{a_n}\right) d[\hat F(t|x) - F(t|x)] + a_n^{-1}\int K_2\left(\frac{y-t}{a_n}\right) dF(t|x) - f(y|x)$$
$$= -a_n^{-1}\int [\hat F(t|x) - F(t|x)]\, dK_2\left(\frac{y-t}{a_n}\right) + a_n^{-1}\int K_2\left(\frac{y-t}{a_n}\right) dF(t|x) - f(y|x)$$
$$= a_n^{-1}\int [\hat F(y-va_n|x) - F(y-va_n|x)]\, dK_2(v) + a_n^{-1}\int K_2\left(\frac{y-t}{a_n}\right) dF(t|x) - f(y|x).$$

To the first term we apply Proposition A.8. The last term equals

$$
\begin{aligned}
& a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right)[f(t|x)-f(y|x)]\,dt \\
& = f'(y|x) a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right)(t-y)\,dt \\
& \quad + \frac{1}{2} f''(y|x) a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right)(t-y)^2\,dt + a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) o(|t-y|^2)\,dt \\
& = \frac{1}{2} a_n^2 \mu_2^{K_2} f''(y|x) + o(a_n^2)
\end{aligned}
$$

from which the result follows.

Proof of Theorem 3.2. For showing the weak convergence of the main term in the representation, use will be made of Theorem 2.11.9 in van der Vaart and Wellner (1996). We start with calculating the covariance of the process. That

$$
\begin{aligned}
& \operatorname{Cov}\left\{K_1\left(\frac{x-X}{a_n}\right) g_{x,y+a_n t}(Z,\Delta), K_1\left(\frac{x-X}{a_n}\right) g_{x,y+a_n t'}(Z,\Delta)\right\} \\
& = a_n f_X(x) \|K_1\|_2^2 \operatorname{Var}\{g_{x,y}(Z,\Delta)|X=x\} + O(a_n^2)
\end{aligned}
$$

follows easily from a Taylor expansion. Next, let $e_v(t) = (y + (t-v)a_n - m(x))/\sigma(x)$ and $e = (y - m(x))/\sigma(x)$. Using the expression of the covariance of the function ξ_e given in Lo and Singh (1986), we have

$$
\begin{aligned}
& \operatorname{Cov}\left[\int \xi_e(E,\Delta,e_v(t))\,dK_2(v), \int \xi_e(E,\Delta,e_v(t'))\,dK_2(v)\right] \\
& = \int\int E\{\xi_e(E,\Delta,e_v(t))\xi_e(E,\Delta,e_w(t'))\}\,dK_2(v)\,dK_2(w) \\
& = \int\int (1-F_e(e_v(t)))(1-F_e(e_w(t'))) \int_{-\infty}^{e_v(t)\wedge e_w(t')} \frac{dH_{e1}(s)}{(1-H_e(s))^2}\,dK_2(v)\,dK_2(w) \\
& = \int \int_{-\infty}^{w+t-t'} (1-F_e(e_v(t)))\,dK_2(v)(1-F_e(e_w(t'))) \int_{-\infty}^{e_w(t')} \frac{dH_{e1}(s)}{(1-H_e(s))^2}\,dK_2(w) \\
& \quad + \int \int_{-\infty}^{v-t+t'} (1-F_e(e_w(t')))\,dK_2(w)(1-F_e(e_v(t))) \int_{-\infty}^{e_v(t)} \frac{dH_{e1}(s)}{(1-H_e(s))^2}\,dK_2(v). \quad (4.1)
\end{aligned}
$$

The first term of (4.1) equals

$$
\int K_2(w+t-t')(1-F_e(e_w(t')))^2 \int_{-\infty}^{e_w(t')} \frac{dH_{e1}(s)}{(1-H_e(s))^2}\,dK_2(w)
$$

$$-\frac{a_n}{\sigma(x)}\int\int_{-\infty}^{w+t-t'} K_2(v)f_e(e_v(t))\,dv(1-F_e(e_w(t')))\int_{-\infty}^{e_w(t')}\frac{dH_{e1}(s)}{(1-H_e(s))^2}\,dK_2(w).$$

After applying integration by parts to both terms above, we get

$$-\int K_2(w)K_2'(w+t-t')(1-F_e(e_w(t')))^2\int_{-\infty}^{e_w(t')}\frac{dH_{e1}(s)}{(1-H_e(s))^2}\,dw$$
$$-\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,dw\,(1-F_e(e))f_e(e)\int_{-\infty}^{e}\frac{dH_{e1}(s)}{(1-H_e(s))^2}$$
$$+\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,dw\,(1-F_e(e))^2\frac{h_{e1}(e)}{(1-H_e(e))^2}+O(a_n^2).$$

Using the same techniques, we find that the second term of (4.1) equals

$$\int K_2(w)K_2'(w+t-t')(1-F_e(e_w(t')))^2\int_{-\infty}^{e_w(t')}\frac{dH_{e1}(s)}{(1-H_e(s))^2}\,dw$$
$$+\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,dw\,(1-F_e(e))f_e(e)\int_{-\infty}^{e}\frac{dH_{e1}(s)}{(1-H_e(s))^2}+O(a_n^2).$$

Hence, (4.1) is equal to

$$\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,dw\frac{h_{e1}\left(\frac{y-m(x)}{\sigma(x)}\right)}{\left(1-G_e\left(\frac{y-m(x)}{\sigma(x)}\right)\right)^2}+O(a_n^2).$$

Next we need to calculate $\mathrm{Cov}(K_1(\frac{x-X}{a_n})g_{x,y+a_nt}(Z,\Delta),\int\xi_e(E,\Delta,e_v(t'))\,dK_2(v))$. Since the calculations are similar to the ones above, we restrict to giving an outline. Since the function $g_{x,y}(z,\delta)$ is formed of the functions $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$, which contain the function $\xi(z,\delta,s|x)$, we need to calculate the covariance of $\xi(Z,\Delta,s|x)$ and $\xi_e(E,\Delta,y)$, given X. Next, integration by parts over the v variable shows that the order of the covariance is $O(a_n^2)$.
Let us now calculate the bias of the second term of the representation in Theorem 3.1 (the first one is unbiased). This follows from a Taylor expansion :

$$a_n^{-1}\int K_1\left(\frac{x-u}{a_n}\right)E(g_{x,y+a_nt}(Z,\Delta)|X=u)\,dF_X(u)$$
$$=\frac{1}{2}a_n^2\mu_2^{K_1}[E(g_{x,y}(Z,\Delta)|X=u)f_X(u)]''|_{u=x}+o(a_n^2),$$

since $E(g_{x,y}(Z,\Delta)|X=x)=0$. Next, we show the convergence of the finite dimensional distributions. By the Cramér-Wold device, we need to show the convergence of any linear combination of the functions $(na_n)^{1/2}(\hat f(y+a_nt_j|x)-f(y+a_nt_j|x))$ $(-\tilde T\le t_1,\ldots,t_k\le\tilde T$

arbitrary, k arbitrary). We do this by verifying Liapunov's condition. Since the functions $g_{x,y}$ and ξ_e are bounded, the Liapunov ratio is easily seen to be $O((na_n)^{-1/2}) = o(1)$. It remains to verify the three displayed conditions in Theorem 2.11.9 in van der Vaart and Wellner (1996). The first one is obviously satisfied, since the functions $g_{x,y}(z,\delta)$ and $\xi_e(z,\delta,y)$ are bounded. We will next show that

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

where $N_{[]}$ is the bracketing number, $\mathcal{F} = \{W_n(t); -\tilde{T} \le t \le \tilde{T}\}$, where

$$\begin{aligned} W_n(t) &= \int \xi_e\left(E, \Delta, \frac{y + (t-v)a_n - m(x)}{\sigma(x)}\right) dK_2(v) + K_1\left(\frac{x - X}{a_n}\right) g_{x,y+a_n t}(Z, \Delta) \\ &= W_{n1}(t) + W_{n2}(t), \end{aligned}$$

P is the probability measure corresponding to the joint distribution of (Z, Δ, X), and $L_2(P)$ is the L_2-norm. Partition $[-\tilde{T}, \tilde{T}]$ into $O(\varepsilon^{-1})$ subintervals $[t_j, t_{j+1}]$ of length at most $K\varepsilon$ for some $K > 0$. We will show that

$$a_n^{-1} E \sup_{t_j \le t, t' \le t_{j+1}} |W_n(t) - W_n(t')|^2 \le \varepsilon^2. \tag{4.2}$$

This does not only imply the third condition in van der Vaart and Wellner (1996), but also the second, since the partitions are independent of n. It is easily seen that $|W_{n2}(t) - W_{n2}(t')| \le K\varepsilon a_n$ whenever $|t - t'| \le K\varepsilon$ and hence (4.2) is satisfied for W_n replaced by W_{n2}. Since the function $\xi_e(z,\delta,y)$ consists of three terms, also $W_{n1}(t)$ can be decomposed into three terms. The most difficult term to deal with is the second one, which equals $Z_n(t) - E[Z_n(t)]$, where

$$Z_n(t) = \int \frac{I(E \le e_v(t), \Delta = 1)}{1 - G_e(e_v(t))}\, dK_2(v).$$

We will prove that (4.2) is satisfied if $W_n(t)$ is replaced by $Z_n(t)$ (the derivation for $E[Z_n(t)]$ follows immediately by integration by parts). Consider

$$\begin{aligned} &Z_n(t) - Z_n(t') \\ &= \int \frac{I(E \le e_v(t), \Delta = 1) - I(E \le e_v(t'), \Delta = 1)}{1 - G_e(e_v(t))}\, dK_2(v) \\ &\quad + \int I(E \le e_v(t'), \Delta = 1)\left[\frac{1}{1 - G_e(e_v(t))} - \frac{1}{1 - G_e(e_v(t'))}\right] dK_2(v). \end{aligned}$$

We concentrate on the first term which equals, using the notation $E_n(x,y) = (E\sigma(x) - y + m(x))/a_n$ and choosing $t' \le t$,

$$
\begin{aligned}
& I(\Delta = 1) \int_{t'-E_n(x,y)}^{t-E_n(x,y)} \frac{1}{1 - G_e(e_v(t))} \, dK_2(v) \\
&= I(\Delta = 1) \left\{ \frac{K_2(t - E_n(x,y))}{1 - G_e(E)} - \frac{K_2(t' - E_n(x,y))}{1 - G_e(E + \sigma^{-1}(x)(t - t')a_n)} \right\} \\
&\quad + \frac{a_n}{\sigma(x)} I(\Delta = 1) \int_{t'-E_n(x,y)}^{t-E_n(x,y)} \frac{K_2(v)}{(1 - G_e(e_v(t)))^2} g_e(e_v(t)) \, dv.
\end{aligned}
$$

The last term above is not more than $K\varepsilon a_n$, which is sufficiently small for the required condition to be satisfied. For the first term above it suffices to consider

$$
\begin{aligned}
& a_n^{-1} E \sup_{t_j \le t, t' \le t_{j+1}} |K_2(t - E_n(x,y)) - K_2(t' - E_n(x,y))|^2 \\
&= a_n^{-1} E \Bigg[\sup_{t_j \le t, t' \le t_{j+1}} |K_2(t - E_n(x,y)) - K_2(t' - E_n(x,y))|^2 \\
&\qquad\qquad \times I(t_j - L_2 \le E_n(x,y) \le t_{j+1} + L_2) \Bigg] \\
&\le K\varepsilon^2,
\end{aligned}
$$

where $[-L_2, L_2]$ is the support of the kernel K_2. This finishes the proof.

Proof of Theorem 3.3. Write

$$
\begin{aligned}
& \hat{\lambda}(y|x) - \lambda(y|x) = \frac{\hat{f}(y|x)}{1 - \hat{F}(y|x)} - \frac{f(y|x)}{1 - F(y|x)} \\
&= \frac{\hat{f}(y|x) - f(y|x)}{1 - F(y|x)} + \frac{f(y|x)[\hat{F}(y|x) - F(y|x)]}{(1 - F(y|x))^2} \\
&\quad + \frac{(\hat{f}(y|x) - f(y|x))(\hat{F}(y|x) - F(y|x))}{1 - \hat{F}(y|x)} + \frac{f(y|x)[\hat{F}(y|x) - F(y|x)]^2}{(1 - \hat{F}(y|x))(1 - F(y|x))^2}.
\end{aligned}
$$

The representation for the first term above follows from Theorem 3.1. For the second term we use a slight generalization of the representation for $\hat{F}(y|x)$ given in Theorem 3.3 in Van Keilegom and Akritas (1999). The latter theorem assumes that $na_n^5 \to 0$, which implies that all bias terms are $o((na_n)^{-1/2})$. Here we assume weaker conditions on the bandwidth, which makes that a bias term has to be added in the representation. In order to show that the third and fourth term are of lower order, we will prove that

$$\sup_{(x,y)\in\Omega} |\hat{F}(y|x) - F(y|x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \quad \text{a.s.}$$

$$\sup_{(x,y)\in\Omega} |\hat{f}(y|x) - f(y|x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \quad \text{a.s.}$$

We will show the latter result, the former can be obtained in a similar way. The representation for $\hat{f}(y|x)$ established in Theorem 3.1 has two main terms, call them $A_1(y|x)$ and $A_2(y|x)$. The term $A_1(y|x)$ contains the function $g_{x,y}(z,\delta)$, which is formed from the functions $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$. These functions are integrals of the function $\xi(z,\delta,y)$. Now, since

$$\begin{aligned}
&(na_n)^{-1}\sum_{i=1}^n K_1\left(\frac{x-X_i}{a_n}\right)\xi(Z_i,\Delta_i,y|x) \qquad (4.3)\\
&= (na_n)^{-1}\sum_{i=1}^n K_1\left(\frac{x-X_i}{a_n}\right)\left[\int_{-\infty}^{y}\frac{I(Z_i\le s)-H(s|x)}{(1-H(s|x))^2}dH_1(s|x)\right.\\
&\quad\left. + \frac{I(Z_i\le y,\Delta_i=1)-H_1(y|x)}{1-H(y|x)} - \int_{-\infty}^{y}\frac{I(Z_i\le s,\Delta_i=1)-H_1(s|x)}{(1-H(s|x))^2}dH(s|x)\right]
\end{aligned}$$

is $O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ a.s. uniformly in x and $y \le T_x$ by Lemma A.1 in Van Keilegom and Akritas (1999), it follows that $A_1(y|x)$ is uniformly of the right order. For $A_2(y|x)$ we first note that $n^{-1}\sum_{i=1}^n \xi_e(E_i,\Delta_i,y)$ can be decomposed in three terms in a similar way as was done in (4.3) for the function $\xi(z,\delta,y|x)$. We consider the second term, which is the most difficult one. Denote $e_v(y) = \frac{y-va_n-m(x)}{\sigma(x)}$ and $e(y) = \frac{y-m(x)}{\sigma(x)}$. Then,

$$\begin{aligned}
&(na_n)^{-1}\sum_{i=1}^n \int \frac{I(E_i\le e_v(y),\Delta_i=1)-H_{e1}(e_v(y))}{1-H_e(e_v(y))}\,dK_2(v)\\
&= (na_n)^{-1}\sum_{i=1}^n \int (1-H_e(e_v(y)))^{-1}[I(E_i\le e_v(y),\Delta_i=1)-H_{e1}(e_v(y))\\
&\quad - I(E_i\le e(y),\Delta_i=1)+H_{e1}(e(y))]\,dK_2(v)\\
&= O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \quad \text{a.s.}
\end{aligned}$$

by Lemma 2.4 in Stute (1982). This finishes the proof.

Appendix

This appendix contains a number of technical results, needed in the proofs of the main results.

Proposition A.1 *Assume (A1)-(A5). Then,*

$$\sup_{-\infty<y<+\infty} |(na_n)^{-1} \sum_{i=1}^{n} \int \{I(\hat{E}_i \le y - va_n) - I(E_i \le y - va_n) - P(\hat{E} \le y - va_n|\mathcal{X}_n)$$
$$+ P(E \le y - va_n)\} \, dK_2(v)| = o_P((na_n)^{-1/2}),$$

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

Proof. We use the notation $y_v = y - va_n$ throughout the proof. The expression between absolute values equals

$$\begin{aligned}
&(na_n)^{-1} \sum_{i=1}^{n} \int \{I(E_i \le d_{n1}(X_i) + y_v(1 + d_{n2}(X_i))) - P(E \le d_{n1}(X) + y_v(1 + d_{n2}(X))) \\
&\quad - J_i(y_v, d_{n1}(X_i), d_{n2}(X_i)) + E[J(y_v, d_{n1}(X), d_{n2}(X))] \\
&\quad -I(E_i \le y_v) + P(E \le y_v) + J_i(y_v, 0, 0) - E[J(y_v, 0, 0)]\} \, dK_2(v) \\
&+ (na_n)^{-1} \sum_{i=1}^{n} \int \{J_i(y_v, d_{n1}(X_i), d_{n2}(X_i)) - E[J(y_v, d_{n1}(X), d_{n2}(X))] \\
&\quad - J_i(y_v, 0, 0) + E[J(y_v, 0, 0)]\} \, dK_2(v) \\
&= A_n(y) + B_n(y),
\end{aligned}$$

where

$$\begin{aligned}
&J_i(y, a, b) \\
&= \begin{cases} 0 & a + y(1 + b) \le E_i - c_n(1 + b) \\ \dfrac{a + y(1 + b) - E_i + c_n(1 + b)}{2c_n(1 + b)} & E_i - c_n(1 + b) \le a + y(1 + b) \le E_i + c_n(1 + b) \\ 1 & a + y(1 + b) \ge E_i + c_n(1 + b), \end{cases}
\end{aligned}$$

$c_n = a_n^{1/2}$, $d_{n1}(x) = (\hat{m}(x) - m(x))/\sigma(x)$ and $d_{n2}(x) = (\hat{\sigma}(x) - \sigma(x))/\sigma(x)$. We start with $B_n(y)$. Using integration by parts, we can write :

$$\begin{aligned}
B_n(y) = (2nc_n)^{-1} \sum_{i=1}^{n} \int K_2(v)\{&I(E_i \le d_{n1}(X_i) + (y_v + c_n)(1 + d_{n2}(X_i))) \\
&- P(E \le d_{n1}(X) + (y_v + c_n)(1 + d_{n2}(X))) \\
&- I(E_i \le y_v + c_n) + P(E \le y_v + c_n)\} \, dv \\
- (2nc_n)^{-1} \sum_{i=1}^{n} \int K_2(v)\{&I(E_i \le d_{n1}(X_i) + (y_v - c_n)(1 + d_{n2}(X_i)))
\end{aligned}$$

$$
\begin{aligned}
&- P(E \le d_{n1}(X) + (y_v - c_n)(1 + d_{n2}(X))) \\
&- I(E_i \le y_v - c_n) + P(E \le y_v - c_n)\} \, dv \\
= {}& (2nc_n)^{-1} \sum_{i=1}^n \int K_2(v) \{ I(\hat{E}_i \le y_v + c_n) - P(\hat{E} \le y_v + c_n) \\
&- I(E_i \le y_v + c_n) + P(E \le y_v + c_n)\} \, dv \\
&- (2nc_n)^{-1} \sum_{i=1}^n \int K_2(v) \{ I(\hat{E}_i \le y_v - c_n) - P(\hat{E} \le y_v - c_n) \\
&- I(E_i \le y_v - c_n) + P(E \le y_v - c_n)\} \, dv.
\end{aligned}
$$

Both terms above are bounded by

$$
(2nc_n)^{-1} \sup_y \Big| \sum_{i=1}^n \{ I(\hat{E}_i \le y) - P(\hat{E} \le y) - I(E_i \le y) + P(E \le y)\} \Big|
$$

and this is $o_P(n^{-1/2} c_n^{-1}) = o_P((na_n)^{-1/2})$, by Lemma B.1 in Van Keilegom and Akritas (1999). The proof for $A_n(y)$ is based on results in van der Vaart and Wellner (1996). Let

$$
\begin{aligned}
Z_{ni}(y, d_1, d_2) = n^{-1/2} a_n^{-1/2-\delta} \int & \{ I(E_i \le d_1(X_i) + y_v(1 + d_2(X_i))) \\
& - J_i(y_v, d_1(X_i), d_2(X_i)) - I(E_i \le y_v) + J_i(y_v, 0, 0)\} \, dK_2(v)
\end{aligned}
$$

$(i = 1, \ldots, n)$. We consider Z_{ni} as a process over the class

$$
\mathcal{F} = \{ (y, d_1, d_2); -\infty < y < +\infty, d_i(X) = e_n \tilde{d}_i(X) \text{ and } \tilde{d}_i(X) \in C_1^\delta(R_X) \quad (i = 1, 2) \},
$$

where $e_n = (na_n^{1+\delta})^{-1/2} (\log a_n^{-(1+\delta)})^{1/2}$ ($\delta > 0$ sufficiently small) and where $C_1^\delta(R_X)$ is the class of all differentiable functions d defined on the domain R_X of X such that

$$
\|d\|_\delta = \sup_x |d(x)| + \sup_{x,x'} \frac{|d(x) - d(x')|}{|x - x'|^\delta} \le 1.
$$

Note that by Propositions 4.3 and (a slight modification of) 4.5 in Van Keilegom and Akritas (1999), we have that $P(e_n^{-1} d_{ni} \in C_1^\delta(R_X) \quad (i = 1, 2)) \to 1$ as $n \to \infty$. We will show that $\sum_{i=1}^n (Z_{ni} - EZ_{ni})$ converges weakly to a Gaussian process. From this it follows that $\sup_y A_n(y) = O_P(n^{-1/2} a_n^{-1/2+\delta})$ and hence $\sup_y A_n(y) = o_P((na_n)^{-1/2})$. To show the weak convergence of the given process, we will verify the conditions of Theorem 2.11.9 in van der Vaart and Wellner (1996). We start with calculating the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, L_2^n)$, which is the minimal number of sets N_ε in a partition of $\mathcal{F}$ into sets $\mathcal{F}_{\varepsilon j}^n$ such that, for every partitioning set $\mathcal{F}_{\varepsilon j}^n$,

$$
\sum_{i=1}^n E\{ \sup |Z_{ni}(y, d_1, d_2) - Z_{ni}(y', d'_1, d'_2)|^2 \} \le \varepsilon^2, \tag{A.1}
$$

where the supremum is taken over all $(y,d_1,d_2),(y',d_1',d_2') \in \mathcal{F}^n_{\varepsilon j}$. Partition the line into $m_1 = O(\varepsilon^{-17/3})$ subintervals y_i $(i=1,\ldots,m_1)$: let $y_0 = -\infty, y_1 = -\varepsilon^{-2/3}, y_{m_1-1} = \varepsilon^{-2/3}, y_{m_1} = +\infty$ and divide the interval $[y_1, y_{m_1-1}]$ into $m_1 - 2$ equally spaced subintervals of length $O(\varepsilon^5)$. Next, note that in Corollary 2.7.2 of the aforementioned book, it is stated that $m_2 = N_{[]}(\varepsilon, C_1^\delta(R_X), L_2(P)) \leq \exp(K\varepsilon^{-\frac{1}{\delta}})$ and $m_3 = N_{[]}(\varepsilon^{5/3}, C_1^\delta(R_X), L_2(P)) \leq \exp(K\varepsilon^{-\frac{5}{3\delta}})$. Let $d_{j1}^L \leq d_{j1}^U$ $(j=1,\ldots,m_2)$ be the functions defining the m_2 ε-brackets for $C_1^\delta(R_X)$ and let $d_{k2}^L \leq d_{k2}^U$ $(k=1,\ldots,m_3)$ be the functions defining the m_3 $\varepsilon^{5/3}$-brackets for $C_1^\delta(R_X)$.
We will show that the brackets for the considered process are given by $[y_i, y_{i+1}] \times [d_{j1}^L, d_{j1}^U] \times [d_{k2}^L, d_{k2}^U]$ $(i=1,\ldots,m_1, j=1,\ldots,m_2, k=1,\ldots,m_3)$. We start with $i=0$ and with arbitrary j and k. We will prove that

$$
\begin{aligned}
& a_n^{-1-2\delta} E \left\{ \sup{}^* \left| \int [I(E \leq d_1(X) + y_v(1+d_2(X))) \right.\right. \\
& \left.\left. - J(y_v, d_1(X), d_2(X)) - I(E \leq y_v) + J(y_v,0,0)]\, dK_2(v) \right| \right\}^2 \leq K\varepsilon^2, \qquad \text{(A.2)}
\end{aligned}
$$

where the supremum is taken over all $(y,d_1,d_2) \in (y_0,y_1] \times [d_{j1}^L, d_{j1}^U] \times [d_{k2}^L, d_{k2}^U]$. First note that, for a fixed value of E, the above supremum can only be different from zero if $E \leq -\varepsilon^{-2/3} + L_2 a_n + e_n + e_n\varepsilon^{-2/3} + c_n(1+e_n) \leq -\frac{1}{2}\varepsilon^{-2/3}$ (since without loss of generality we can assume that $\varepsilon \leq 1$). Denote the integrand in (A.2) by $I(y,d_1,d_2,v)$. Then, the left hand side of (A.2) is bounded by

$$
\begin{aligned}
& a_n^{-1-2\delta} E \left\{ \left[\sup{}^* \int^* |I(y,d_1,d_2,v)| dK_2(v) + \sup{}^* \int^{**} |I(y,d_1,d_2,v)| dK_2(v) \right] \right. \\
& \left. \times I\left(E \leq -\frac{1}{2}\varepsilon^{-2/3}\right) \right\}^2, \qquad \text{(A.3)}
\end{aligned}
$$

where $\int^*$ denotes the integral over all v for which y_v and $d_1(X) + y_v(1+d_2(X))$ belong to the same side of E and $\int^{**}$ denotes the integral over the remaining v's. The first integral is bounded by $K|d_1(X) + Ed_2(X)|/(2c_n(1+d_2(X)))$ and hence,

$$
\sup{}^* \int^* |I(y,d_1,d_2,v)| dK_2(v) \leq \frac{K|d_1(X) + Ed_2(X)|}{2c_n(1+d_2(X))} \leq K\frac{e_n}{c_n}|E|.
$$

The length of the interval over which is integrated in the second integral in (A.3) is $O(e_n/a_n)|E|$ if y_v or $d_1(X) + y_v(1+d_2(X))$ belong to $[E-c_n, E+c_n]$ (which implies that $|y_v| \leq K(|E| + c_n)$ for some $K > 0$). If none of the two points belongs to the interval, then the integrand will be zero. Hence, this integral is bounded above by

$$
\sup{}^* \int^{**} |I(y,d_1,d_2,v)| dK_2(v) \leq K\frac{e_n}{a_n}|E|.
$$

It now follows that (A.3) is not larger than

$$Ka_n^{-3-2\delta}e_n^2 \int_{-\infty}^{-\frac{1}{2}\varepsilon^{-2/3}} z^2\,dH_e(z) \le (na_n^{4+3\delta})^{-1}\log a_n^{-(1+\delta)}\varepsilon^2,$$

for ε small enough, since

$$\int_{-\infty}^{-\frac{1}{2}\varepsilon^{-2/3}} z^2\,dH_e(z) \le 8\varepsilon^2 \int_{-\infty}^{-\frac{1}{2}\varepsilon^{-2/3}} |z|^5\,dH_e(z) \le \varepsilon^2$$

for ε small enough. This shows that all intervals of the form $(y_0, y_1] \times [d_{j1}^L, d_{j1}^U] \times [d_{k2}^L, d_{k2}^U]$ $(j = 1, \ldots, m_2, k = 1, \ldots, m_3)$ satisfy the bracketing condition (A.1). In a completely similar way we can show that all three dimensional intervals containing the right tail $[y_{m_1-1}, y_{m_1})$ of the line satisfy this condition. It remains to consider the case where the y-interval is one of the finite intervals $[y_i, y_{i+1}]$ $(i = 1, \ldots, m_1 - 1)$. We consider two cases depending on whether ε is smaller or larger than $n^{-1/15}$. First assume that $\varepsilon \le n^{-1/15}$. In this case we will prove that

$$\begin{aligned} &a_n^{-1-2\delta}E\Big\{\sup{}^*\Big|\int[I(E \le d_1(X) + y_v(1 + d_2(X))) - J(y_v, d_1(X), d_2(X)) \\ &- I(E \le d_1'(X) + y_v'(1 + d_2'(X))) + J(y_v', d_1'(X), d_2'(X))]\,dK_2(v)\Big|\Big\}^2 \le K\varepsilon^2, \quad \text{(A.4)} \end{aligned}$$

where the supremum is taken over all $(y, d_1, d_2), (y', d_1', d_2') \in [y_i, y_{i+1}] \times [d_{j1}^L, d_{j1}^U] \times [d_{k2}^L, d_{k2}^U]$. This together with the analogue of (A.4) for y_v and y_v' implies (A.1). If $d_1(X) + y_v(1 + d_2(X))$ and $d_1'(X) + y_v'(1 + d_2'(X))$ are both smaller or larger than E, than the integrand on the left hand side of (A.4) is $O((\varepsilon^5 + e_n\varepsilon)/c_n)$, while the probability that for a given value of E, the integrand in (A.4) is different from zero is $O(c_n)$. Hence, the left hand side of (A.4) is bounded by $Ka_n^{-\frac{3}{2}-2\delta}(\varepsilon^{10} + e_n^2\varepsilon^2) \le \varepsilon^2$. If $d_1(X) + y_v(1 + d_2(X))$ and $d_1'(X) + y_v'(1 + d_2'(X))$ each belong to a different side of E, then the length of the interval over which is integrated in (A.4) is $O((\varepsilon^5 + e_n\varepsilon)/a_n)$ and the probability that for a given value of E, the integrand is different from zero is $O(\varepsilon^5 + a_n + e_n\varepsilon) = O(\varepsilon^5 + a_n)$. This gives that the integral in (A.4) is not larger than $Ka_n^{-3-2\delta}(\varepsilon^{15} + a_n\varepsilon^{10} + e_n^2\varepsilon^7 + a_ne_n^2\varepsilon^2) \le \varepsilon^2$. This finishes the case $\varepsilon \le n^{-1/15}$. Finally, assume that $\varepsilon \ge n^{-1/15}$. In this case, we need to consider the four points together (this in opposite to the previous case where we proved the bracketing condition (A.1) for two sets of two points). There are several cases we need to consider. The one for which the bound is the sharpest is when three (out of the four) points belong to one side of E and one point to the other side. We restrict attention to this case (the derivation for the other cases is similar). In that case, the length of the interval over which we integrate is $O(\min(\varepsilon^5 + e_n\varepsilon, e_n\varepsilon^{-2/3})/a_n)$, while the integrand can be as big as 1. The probability that for a given value of E, the integrand is different from zero is $O(\varepsilon^5 + a_n + e_n\varepsilon^{-2/3})$. Combining these bounds yields that the left hand side of (A.1) is at most ε^2.

We have now shown that the bracketing number for the considered process is $O(\varepsilon^{-17/3}\exp(K\varepsilon^{-5/(3\delta)}))$ and hence

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2^n)}\, d\varepsilon \to 0$$

for all $\delta_n \downarrow 0$.

We next prove the convergence of the marginals of the process $\sum_{i=1}^n (Z_{ni} - EZ_{ni})$. This will be achieved by proving that Liapunov's ratio

$$\frac{\sum_{i=1}^n E|Z_{ni} - E[Z_{ni}]|^3}{[\sum_{i=1}^n \mathrm{Var}(Z_{ni})]^{3/2}}$$

tends to zero as n tends to infinity. We start with calculating the second moment of $Z_n(y, d_1, d_2)$.

$$\begin{aligned}
&E^2[Z_n(y,d_1,d_2)] \\
&= 2n^{-1}a_n^{-1-2\delta} \int \int_{-L_2}^{v} E\{[L(y_v,d_1,d_2) - L(y_v,0,0)] \\
&\quad \times [L(y_w,d_1,d_2) - L(y_w,0,0)]\}\, dK_2(w)\, dK_2(v) \\
&= 2n^{-1}a_n^{-1-2\delta} \int \int_{-L_2}^{v} \{E[L(y_v,d_1,d_2)L(y_w,d_1,d_2)] - E[L(y_v,d_1,d_2)L(y_w,0,0)] \\
&\quad - E[L(y_v,0,0)L(y_w,d_1,d_2)] + E[L(y_v,0,0)L(y_w,0,0)]\}\, dK_2(w)\, dK_2(v), \qquad \text{(A.5)}
\end{aligned}$$

where $[-L_2, L_2]$ is the support of the kernel K_2 and

$$\begin{aligned}
L(y,a,b) &= I(E \le a + y(1+b)) - J(y,a,b) \\
&= \begin{cases} 0 & a + y(1+b) \le E - c_n(1+b) \\ -\dfrac{a + y(1+b) - E + c_n(1+b)}{2c_n(1+b)} & E - c_n(1+b) \le a + y(1+b) \le E \\ -\dfrac{a + y(1+b) - E - c_n(1+b)}{2c_n(1+b)} & E \le a + y(1+b) \le E + c_n(1+b) \\ 1 & a + y(1+b) \ge E + c_n(1+b). \end{cases}
\end{aligned}$$

We restrict attention for a while to the integrand of the last term of (A.5) for fixed values of $v > w$. Since $L_2 a_n < c_n = a_n^{1/2}$, we have that $y_w - c_n < y_v$ and hence the points

$-\infty < y_v - c_n < y_w - c_n < y_v < y_w < y_v + c_n < y_w + c_n < +\infty$ partition the line into seven regions, say $R_j((y_v,0,0),(y_w,0,0))$ $(j=1,\ldots,7)$. Hence, the integrand of the last term of (A.5) equals

$$\sum_{i=1}^{7} E[L(y_v,0,0)L(y_w,0,0)I\{E \in R_j((y_v,0,0),(y_w,0,0))\}].$$

In a similar way, we can write the other three terms of (A.5) as a sum of seven terms. In this way, we obtain seven terms, each containing four terms. Denote these terms by $A_{n1},\ldots,A_{n7}$. The sixth and seventh term equal zero, since for these terms one (or both) of the functions $L(y_v,\cdot,\cdot)$ and $L(y_w,\cdot,\cdot)$ are zero. For the remaining five terms, we restrict the calculations for simplicity to the homoscedastic case where $d_2 \equiv 0$. The general case where d_2 can be any function of the class $\mathcal{F}$ that we consider, can be dealt with in a similar way. Some calculations show that

$$\begin{aligned}
A_{n1} &= -2n^{-1}a_n^{-1-2\delta}E\left\{\int \left[K(v) - K\left(v - \frac{d_1(X)}{a_n}\right)\right]\left[K'(v) - K'\left(v - \frac{d_1(X)}{a_n}\right)\right]\right. \\
&\qquad \left. \times H_e(d_1(X) + y_v - c_n|X)\right\} dv \\
&\quad + 2n^{-1}a_n^{-1-2\delta}E\left\{\int \left[K(v) - K\left(v - \frac{d_1(X)}{a_n}\right)\right]^2 h_e(d_1(X) + y_v - c_n|X)\, dv\right\} \\
&= O(n^{-1}e_n^2 a_n^{-3-2\delta}) \\
A_{n2} &= n^{-1}a_n^{1-2\delta}c_n^{-1}E\left\{\int K(v) \int_{v+\frac{d_1(X)}{a_n}}^{v} \left[K(w) - K\left(w - \frac{d_1(X)}{a_n}\right)\right]\right. \\
&\qquad \left. \times h_e(d_1(X) + y_w - c_n|X)\, dw\, dv\right\} \\
&= O(n^{-1}a_n^{-3/2-2\delta}e_n^2).
\end{aligned}$$

The calculations for the terms A_{n3}, A_{n4} and A_{n5} are very long but straightforward and will not be given. The asymptotic order of these terms is smaller than the order of the term A_{n1} and so we can conclude that

$$E^2(Z_n(y,d_1,d_2)) = O(n^{-1}e_n^2 a_n^{-3-2\delta}) = O(n^{-2}a_n^{-4-3\delta}\log a_n^{-(1+\delta)}).$$

In a similar, but much easier way, we can show that

$$E(Z_n(y,d_1,d_2)) = O(n^{-\frac{1}{2}}e_n a_n^{-\delta}) = O(n^{-1}a_n^{-\frac{1}{2}-\frac{3\delta}{2}}(\log a_n^{-(1+\delta)})^{1/2})$$

and so $(E[Z_n(y,d_1,d_2)])^2$ is negligible with respect to $E^2[Z_n(y,d_1,d_2)]$. This means that the Liapunov ratio equals

$$O\left(\frac{a_n^{\frac{3+\delta}{2}}}{(\log a_n^{-(1+\delta)})^{\frac{1}{2}}}\right) \to 0,$$

uniformly in y, d_1 and d_2, since $\sup_{y,d_1,d_2} |Z_n(y,d_1,d_2)| \le Kn^{-1/2}a_n^{-1/2-\delta}$. This shows the convergence of the marginals of the process $\sum_{i=1}^n (Z_{ni} - E[Z_{ni}])$. The remaining conditions in Theorem 2.11.9 of van der Vaart and Wellner (1996) are easily seen to be satisfied and hence the result follows.

Proposition A.2 *Assume (A1)-(A5). Then,*

$$\begin{aligned}
&\hat{H}_e(y) - H_e(y) \\
&= n^{-1}\sum_{i=1}^n [-\{\eta(Z_i,\Delta_i|X_i) + \zeta(Z_i,\Delta_i|X_i)y\}h_e(y|X_i) + I(E_i \le y) - H_e(y)] \\
&\quad - \frac{1}{2}a_n^2\mu_2^{K_1}\int\{[E(\eta(Z,\Delta|x)|X=u)f_X(u)]''|_{u=x} \\
&\qquad + [E(\zeta(Z,\Delta|x)|X=u)f_X(u)]''|_{u=x}y\}h_e(y|x)\,dx + R_n(y),
\end{aligned}$$

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

We omit the proof of this result, since it can be obtained from the proof of the next result, by omitting the integration over the variable v, or from Proposition B.2 in Van Keilegom and Akritas (1999), where the above result was shown under the extra condition that $na_n^4 \to 0$ (which implies that the bias term in the above representation is negligible).

Proposition A.3 *Assume (A1)-(A5). Then,*

$$\begin{aligned}
&a_n^{-1}\int \frac{\hat{H}_e(y - va_n) - H_e(y - va_n)}{1 - G_e(y - va_n)}\,dK_2(v) \\
&= (na_n)^{-1}\sum_{i=1}^n \int \frac{I(E_i \le y - va_n) - H_e(y - va_n)}{1 - G_e(y - va_n)}\,dK_2(v) \\
&\quad - \frac{1}{2}a_n^2\mu_2^{K_1}\int\left\{[E(\eta(Z,\Delta|x)|X=u)f_X(u)]''|_{u=x}\frac{\partial}{\partial y}\left[\frac{h_e(y|x)}{1 - G_e(y)}\right]\right. \\
&\qquad \left. + [E(\zeta(Z,\Delta|x)|X=u)f_X(u)]''|_{u=x}\frac{\partial}{\partial y}\left[\frac{yh_e(y|x)}{1 - G_e(y)}\right]\right\}\,dx + R_n(y),
\end{aligned}$$

where $\sup\{|R_n(y)|; -\infty < y \le T\} = o_P((na_n)^{-1/2}) + o_P(a_n^2)$*.*

Proof. It follows from Proposition A.1 that

$$
\begin{aligned}
& a_n^{-1} \int (1 - G_e(y - va_n))^{-1} (\hat{H}_e(y - va_n) - H_e(y - va_n))\, dK_2(v) \\
& = a_n^{-1} \int \int (1 - G_e(y - va_n))^{-1} \left\{ H_e \left(\frac{(y - va_n)\hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)} \middle| x \right) \right. \\
& \qquad \left. - H_e(y - va_n|x) \right\} dF_X(x) dK_2(v) \\
& \quad + (na_n)^{-1} \sum_{i=1}^{n} \int (1 - G_e(y - va_n))^{-1} \{ I(E_i \leq y - va_n) - H_e(y - va_n) \}\, dK_2(v) \\
& \quad + (na_n)^{-1} \sum_{i=1}^{n} \int (1 - G_e(y - va_n))^{-1} \{ I(\hat{E}_i \leq y - va_n) - I(E_i \leq y - va_n) \\
& \qquad - P(\hat{E} \leq y - va_n | \mathcal{X}_n) + P(E \leq y - va_n) \}\, dK_2(v) \\
& = a_n^{-1} \int \int \frac{h_e(y - va_n|x)}{1 - G_e(y - va_n)} \frac{(y - va_n)(\hat{\sigma}(x) - \sigma(x)) + \hat{m}(x) - m(x)}{\sigma(x)}\, dF_X(x) dK_2(v) \\
& \quad + \frac{1}{2} a_n^{-1} \iint \frac{h_e'(y_{vx}|x)}{1 - G_e(y - va_n)} \left(\frac{(y - va_n)(\hat{\sigma}(x) - \sigma(x)) + \hat{m}(x) - m(x)}{\sigma(x)} \right)^2 dF_X(x) dK_2(v) \\
& \quad + (na_n)^{-1} \sum_{i=1}^{n} \int (1 - G_e(y - va_n))^{-1} \{ I(E_i \leq y - va_n) - H_e(y - va_n) \}\, dK_2(v) \\
& \quad + o_P((na_n)^{-1/2}),
\end{aligned}
\tag{A.6}
$$

where y_{vx} is between $y - va_n$ and $((y - va_n)\hat{\sigma}(x) + \hat{m}(x) - m(x))/\sigma(x)$. That the second term on the right hand side of (A.6) is $O((na_n^2)^{-1} \log a_n^{-1}) = o((na_n)^{-1/2})$ a.s. follows from the consistency of the estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$ given by Proposition 4.3 in Van Keilegom and Akritas (1999). The first term naturally splits in two parts, whose derivations are similar. We restrict the proof to the second part. Applying integration by parts, this term can be written as

$$
\int \left[\frac{h_e(y|x)}{1 - G_e(y)} \right]' \frac{\hat{m}(x) - m(x)}{\sigma(x)}\, dF_X(x) + o_P((na_n)^{-1/2}),
$$

where the order of the remainder term follows from the fact that $\sup_x |\hat{m}(x) - m(x)| = O((na_n)^{-1/2} (\log a_n^{-1})^{1/2})$ a.s. (see Proposition 4.3 in Van Keilegom and Akritas (1999)). Using again this representation for $\hat{m}(x)$, it follows that

$$
a_n^{-1} \int \int \frac{h_e(y - va_n|x)}{1 - G_e(y - va_n)} \frac{\hat{m}(x) - m(x)}{\sigma(x)}\, dF_X(x) dK_2(v)
$$

$$
\begin{aligned}
&= -(na_n)^{-1} \sum_{i=1}^n \int f_X^{-1}(x) \left[\frac{h_e(y|x)}{1-G_e(y)}\right]' K_1\left(\frac{x-X_i}{a_n}\right) \eta(Z_i,\Delta_i|x)\, dF_X(x) + o_P((na_n)^{-1/2}) \\
&= -(na_n)^{-1} \sum_{i=1}^n \int \left[\frac{h_e(y|x)}{1-G_e(y)}\right]' K_1\left(\frac{x-X_i}{a_n}\right) \{\eta(Z_i,\Delta_i|x) - [\eta(Z_i,\Delta_i|x)|X_i]\}\, dx \\
&\quad - \int \left[\frac{h_e(y|x)}{1-G_e(y)}\right]' \left[(na_n)^{-1} \sum_{i=1}^n K_1\left(\frac{x-X_i}{a_n}\right) E[\eta(Z_i,\Delta_i|x)|X_i]\right. \\
&\qquad \left. - a_n^{-1} E\left\{K_1\left(\frac{x-X}{a_n}\right)\eta(Z,\Delta|x)\right\}\right] dx \\
&\quad - a_n^{-1} \int \left[\frac{h_e(y|x)}{1-G_e(y)}\right]' a_n^{-2} E\left\{K_1\left(\frac{x-X}{a_n}\right)\eta(Z,\Delta|x)\right\} dx + o_P((na_n)^{-1/2}) \\
&= \alpha_{n1}(y) + \alpha_{n2}(y) + \alpha_{n3}(y) + o_P((na_n)^{-1/2}).
\end{aligned}
$$

Let

$$
Q_i(y|x) = \left[\frac{h_e(y|x)}{1-G_e(y)}\right]' \{\eta(Z_i,\Delta_i|x) - E[\eta(Z_i,\Delta_i|x)|X_i]\}.
$$

Then, using a Taylor expansion,

$$
\begin{aligned}
\alpha_{n1}(y) &= -(na_n)^{-1} \sum_{i=1}^n \int K_1\left(\frac{x-X_i}{a_n}\right) Q_i(y|x)\, dx \\
&= -n^{-1} \sum_{i=1}^n \left[\frac{h_e(y|X_i)}{1-G_e(y)}\right]' \eta(Z_i,\Delta_i|X_i) - \frac{1}{2} n^{-1} a_n^2 \mu_2^{K_1} \sum_{i=1}^n \ddot{Q}_i(y|X_i) + o(a_n^2), \quad \text{(A.7)}
\end{aligned}
$$

where $\dot{Q}_i(y|x)$, respectively $\ddot{Q}_i(y|x)$, denotes the first, respectively second, partial derivative of $Q_i(y|x)$ with respect to x. We will show that both terms in (A.7) are asymptotically negligible. We start with proving that the first term, considered as a process in y, and multiplied with $n^{1/2}$, converges to a zero mean Gaussian process. This implies that this term is $O_P(n^{-1/2}) = o_P((na_n)^{-1/2})$, uniformly in y. We will make use of Theorem 2.5.6 in van der Vaart and Wellner (1996), i.e. we will show that

$$
\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty, \quad \text{(A.8)}
$$

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

$$
\mathcal{F} = \left\{ \left[\frac{h_e'(y|X)}{1-G_e(y)} + \frac{h_e(y|X)}{(1-G_e(y))^2} g_e(y)\right] \eta(Z,\Delta|X); -\infty < y \le T \right\}.
$$

Proving this entails that the class $\mathcal{F}$ is Donsker and hence the weak convergence of the given process follows from p. 81-82 in van der Vaart and Wellner (1996). Since the functions $x \to \frac{h_e'(y|x)}{1-G_e(y)} + \frac{h_e(y|x)}{(1-G_e(y))^2} g_e(y)$ 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 $\eta(z,\delta|x)$ is uniformly bounded, the bracketing number for the class $\mathcal{F}$ is $O(\exp(K\varepsilon^{-1}))$. This shows that the integral in (A.8) is finite, since the integration can be restricted to the interval $[0, 2M]$, if the functions of the class $\mathcal{F}$ are bounded by M (for $\varepsilon > 2M$, we take $N_{[]}(\varepsilon, \mathcal{F}, L_2(P)) = 1$). It now follows that the first term of (A.7) is $O_P(n^{-1/2})$, uniformly in y. In a completely similar way, we can show that the second term is $O_P(n^{-1/2}a_n^2)$, uniformly in y.
Using the notation $P(y|x) = \left[\frac{h_e(y|x)}{1-G_e(y)}\right]'$, the first term of $\alpha_{n2}(y)$ equals

$$
\begin{aligned}
&-(na_n)^{-1}\sum_{i=1}^n \int K_1\left(\frac{x-X_i}{a_n}\right) P(y|x)E[\eta(Z_i,\Delta_i|x)|X_i]\,dx \\
&= -(na_n)^{-1}\sum_{i=1}^n \int K_1\left(\frac{x-X_i}{a_n}\right) dx\, P(y|X_i)E[\eta(Z_i,\Delta_i|X_i)|X_i] \\
&\quad -(na_n)^{-1}\sum_{i=1}^n \int K_1\left(\frac{x-X_i}{a_n}\right)(x-X_i)\,dx\,\frac{\partial}{\partial x}\left\{P(y|x)E[\eta(Z_i,\Delta_i|x)|X_i]\right\}|_{x=X_i} \\
&\quad -\frac{1}{2}(na_n)^{-1}\sum_{i=1}^n \int K_1\left(\frac{x-X_i}{a_n}\right)(x-X_i)^2\,dx\,\frac{\partial^2}{\partial x^2}\left\{P(y|x)E[\eta(Z_i,\Delta_i|x)|X_i]\right\}|_{x=X_i} \\
&\quad + o(a_n^2) \\
&= -\frac{1}{2}n^{-1}a_n^2\mu_2^{K_1}\sum_{i=1}^n \frac{\partial^2}{\partial x^2}\left\{\left[\frac{h_e'(y|x)}{1-G_e(y)} + \frac{h_e(y|x)}{(1-G_e(y))^2}g_e(y)\right]E[\eta(Z_i,\Delta_i|x)|X_i]\right\}\Bigg|_{x=X_i} \\
&\quad + o(a_n^2).
\end{aligned}
$$

Using a similar Taylor expansion, the second term of $\alpha_{n2}(y)$ can be written as

$$
-\frac{1}{2}a_n^2\mu_2^{K_1}\frac{\partial^2}{\partial x^2}\left\{\left[\frac{h_e'(y|x)}{1-G_e(y)} + \frac{h_e(y|x)}{(1-G_e(y))^2}g_e(y)\right]E[\eta(Z,\Delta|x)]\right\} + o(a_n^2).
$$

In an analogous way as for the first term of (A.7), one can show that $\alpha_{n2}(y)$ is $O_P(n^{-1/2}a_n^2)$, uniformly in y.
Finally, using again a Taylor expansion, we can write

$$
\alpha_{n3}(y) = -\frac{1}{2}a_n^2\mu_2^{K_1}\int\left[\frac{h_e(y|x)}{1-G_e(y)}\right]'[E(\eta(Z,\Delta|x)|X=u)f_X(u)]''|_{u=x}\,dx + o(a_n^2).
$$

This finishes the proof.

We next give a slight generalization of Proposition B.5 in Van Keilegom and Akritas (1999), in the sense that in the result below the conditions on the sequence $\bar{a}_n$ are less restrictive. The proof is completely similar.

Proposition A.4 *Assume (A1)-(A5). Then,*

$$\sup |\hat{H}_e(y_2) - \hat{H}_e(y_1) - H_e(y_2) + H_e(y_1)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}\bar{a}_n^{1/2}) + o_P(n^{-1/2}),$$

where the supremum is taken over all (y_1, y_2) *satisfying* $|H_e(y_2) - H_e(y_1)| \le \bar{a}_n$ *and where* $\bar{a}_n$ *is any sequence of positive constants satisfying* $\bar{a}_n \to 0$ *and* $(n\bar{a}_n)^{-1}\log a_n^{-1} \to 0$.

The proof of the next result is inspired by the proof of the above result. Since some parts have changed substantially, we give a complete proof.

Proposition A.5 *Assume (A1)-(A5). Then,*

$$\sup \left| a_n^{-1} \int \left\{ \hat{H}_e(y_2 - va_n) - \hat{H}_e(y_1 - va_n) - H_e(y_2 - va_n) + H_e(y_1 - va_n) \right\} dK_2(v) \right|$$
$$= O(n^{-1/2}a_n^{-1}(\log a_n^{-1})^{1/2}\bar{a}_n^{1/2}) + o_P((na_n)^{-1/2}),$$

where the supremum is taken over all (y_1, y_2) *satisfying* $|H_e(y_2) - H_e(y_1)| \le \bar{a}_n$ *and where* $\bar{a}_n$ *is any sequence of positive constants satisfying* $\bar{a}_n \to 0$ *and* $(n\bar{a}_n)^{-1}\log a_n^{-1} \to 0$.

Proof. Using the same decomposition as in Proposition A.3, we can write

$$a_n^{-1} \int \left\{ \hat{H}_e(y_2 - va_n) - \hat{H}_e(y_1 - va_n) - H_e(y_2 - va_n) + H_e(y_1 - va_n) \right\} dK_2(v)$$
$$= a_n^{-1} \int\int \left\{ \frac{\hat{m}(x) - m(x)}{\sigma(x)} \{h_e(y_2 - va_n|x) - h_e(y_1 - va_n|x)\} \right.$$
$$\left. + \frac{\hat{\sigma}(x) - \sigma(x)}{\sigma(x)} \{(y_2 - va_n)h_e(y_2 - va_n|x) - (y_1 - va_n)h_e(y_1 - va_n|x)\} \right\}$$
$$dF_X(x)\, dK_2(v)$$
$$+ (na_n)^{-1} \sum_{i=1}^n \int \{I(E_i \le y_2 - va_n) - I(E_i \le y_1 - va_n) - H_e(y_2 - va_n)$$
$$+ H_e(y_1 - va_n)\}\, dK_2(v) + o_P((na_n)^{-1/2}).$$

The second term above is $O(n^{-1/2}a_n^{-1}(\log n)^{1/2}\bar{a}_n^{1/2})$ a.s. by Lemma 2.4 in Stute (1982). The first term consists of two similar parts. The second part is bounded by

$$a_n^{-1} \frac{\sup_x |\hat{\sigma}(x) - \sigma(x)|}{\inf_x \sigma(x)} \left| \int\int [(y_2 - va_n)h_e(y_2 - va_n|x) \right.$$
$$\left. - (y_1 - va_n)h_e(y_1 - va_n|x)]\, dF_X(x)\, dK_2(v) \right|. \quad \text{(A.9)}$$

The integral between absolute values equals

$$a_n \int\int K_2(v)[(y_2 - va_n)h_e'(y_2 - va_n|x) - (y_1 - va_n)h_e'(y_1 - va_n|x)]\, dF_X(x)\, dv$$
$$+ a_n \int\int K_2(v)[h_e(y_2 - va_n|x) - h_e(y_1 - va_n|x)]\, dF_X(x)\, dv.$$

The first integral above can also be written as (the derivation for the second term is similar)

$$a_n \int [y_2 h_e'(y_2|x) - y_1 h_e'(y_1|x)]\, dF_X(x) + O(a_n^3).$$

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

$$|y_2 h_e'(y_2|x) - y_1 h_e'(y_1|x)| \le \frac{K}{\bar{a}_n^{1/2}} |H_e(y_2|x) - H_e(y_1|x)| + 2\bar{a}_n^{1/2}. \tag{A.10}$$

Showing this implies that for all (y_1, y_2) satisfying $|H_e(y_2) - H_e(y_1)| \le \bar{a}_n$,

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

This together with Proposition 4.3 in Van Keilegom and Akritas (1999) and the assumption on $\bar{a}_n$ implies the order $O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}\bar{a}_n^{1/2})$ a.s. for (A.9). For the proof of (A.10) define $A_x = \{y; |yh_e'(y|x)| \ge \bar{a}_n^{1/2}\}$. 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 $|yh_e'(y|x)| = \bar{a}_n^{1/2}$ and $u \in A_x$ for all u between y and y_2. Using this y write

$$|y_2 h_e'(y_2|x) - y_1 h_e'(y_1|x)| \le |y_2 h_e'(y_2|x) - yh_e'(y|x)| + |yh_e'(y|x) - y_1 h_e'(y_1|x)|. \tag{A.11}$$

The second term on the right hand side of (A.11) is clearly less than $2\bar{a}_n^{1/2}$. To deal with the first term define $v_e(y|x) = yh_e'(y|x)$ and write

$$\begin{aligned}|v_e(y_2|x) - v_e(y|x)| &= |(v_e(\cdot|x) \circ H_e^{-1}(\cdot|x))'(H_e(u|x))|(H_e(y_2|x) - H_e(y|x)) \\ &= \frac{|v_e'(u|x)|}{h_e(u|x)}(H_e(y_2|x) - H_e(y|x)) \\ &= \frac{|h_e'(u|x) + uh_e''(u|x)|}{h_e(u|x)}(H_e(y_2|x) - H_e(y|x)) \\ &\le \left(\frac{\sup_u |uh_e'(u|x)| + \sup_u |u^2 h_e''(u|x)|}{\inf_{u \in A_x} |uh_e(u|x)|} \right) (H_e(y_2|x) - H_e(y|x)) \\ &\le K\bar{a}_n^{-1/2}(H_e(y_2|x) - H_e(y|x)), \qquad (A.12)\end{aligned}$$

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.12) shows that $|y_2 h_e'(y_2|x) - y_1 h_e'(y_1|x)| \leq K\bar{a}_n^{-1/2}(H_e(y_2|x) - H_e(y_1|x))$. 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 $u_1 h_e(u_1|x) = \bar{a}_n^{1/2}$ and $u_2 h_e(u_2|x) = \bar{a}_n^{1/2}$. Also assume without loss of generality that $y_1 < y_2$. Then $|y_2 h_e'(y_2|x) - y_1 h_e'(y_1|x)| \leq K\bar{a}_n^{1/2}$ follows from the decomposition $|y_2 h_e'(y_2|x) - y_1 h_e'(y_1|x)| \leq | y_2 h_e'(y_2|x) - u_2 h_e'(u_2|x)| + |u_2 h_e'(u_2|x) - u_1 h_e'(u_1|x)| + |u_1 h_e'(u_1|x) - y_1 h_e'(y_1|x)|$ and (A.12).

Proposition A.6 *Assume (A1)-(A5). Then,*

$$\sup_{-\infty<y\leq T}\left|a_n^{-1}\int(1-F_e(y-va_n))\int_{-\infty}^{y-va_n}\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}\right]d(\hat{H}_{e1}(s)-H_{e1}(s))\,dK_2(v)\right|$$
$$= o_P((na_n)^{-1/2}). \tag{A.13}$$

Proof. The expression on the left hand side in (A.13) is bounded by

$$\sup_y\left|a_n^{-1}(1-F_e(y))\int\int_{-\infty}^{y-va_n}\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}\right]d(\hat{H}_{e1}(s)-H_{e1}(s))\,dK_2(v)\right|$$
$$+K\sup_y\left|\int_{-\infty}^{y}\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}\right]d(\hat{H}_{e1}(s)-H_{e1}(s))\right|.$$

The second term above is $o_P(n^{-1/2})$ by Corollary B.6 in Van Keilegom and Akritas (1999). Partition the interval $(-\infty, T]$ into $O(a_n^{-1})$ subintervals $[y_i, y_{i+1}]$ satisfying $H_e(y_{i+1}) - H_e(y_i) \leq a_n$. Then,

$$a_n^{-1}\int\int_{-\infty}^{y-va_n}\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}\right]d(\hat{H}_{e1}(s)-H_{e1}(s))\,dK_2(v)$$
$$=a_n^{-1}\int K_2\left(\frac{y-s}{a_n}\right)\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}\right]d(\hat{H}_{e1}(s)-H_{e1}(s))$$
$$=a_n^{-1}\int K_2\left(\frac{y-s}{a_n}\right)\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}-\frac{1}{1-\hat{H}_e(y_i)}+\frac{1}{1-H_e(y_i)}\right]$$
$$\times d(\hat{H}_{e1}(s)-H_{e1}(s))$$
$$+a_n^{-1}\left[\frac{1}{1-\hat{H}_e(y_i)}-\frac{1}{1-H_e(y_i)}\right]\int K_2\left(\frac{y-s}{a_n}\right)d(\hat{H}_{e1}(s)-H_{e1}(s))$$
$$=\alpha_{n1}(y)+\alpha_{n2}(y),$$

where $y_i \le y \le y_{i+1}$. Then, if $K_2(\frac{y-s}{a_n}) \neq 0$, it follows that $|H_e(y_i) - H_e(s)| \le K a_n$. We start with $\alpha_{n1}(y)$.

$$
\begin{aligned}
&|\alpha_{n1}(y)| \\
&\le K a_n^{-1} \sup{}^* \left| \frac{1}{1-\hat{H}_e(t)} - \frac{1}{1-H_e(t)} - \frac{1}{1-\hat{H}_e(s)} + \frac{1}{1-H_e(s)} \right| \\
&\quad \times \{\hat{H}_{e1}(y+L_2 a_n) - \hat{H}_{e1}(y-L_2 a_n) + H_{e1}(y+L_2 a_n) - H_{e1}(y-L_2 a_n)\} \\
&\le K a_n^{-1} \left\{ \sup{}^* \left| \frac{\hat{H}_e(t) - H_e(t)}{(1-H_e(t))^2} - \frac{\hat{H}_e(s) - H_e(s)}{(1-H_e(s))^2} \right| + O((na_n)^{-1} \log a_n^{-1}) \right\} \\
&\quad \times \{\sup{}^* |\hat{H}_{e1}(t) - H_{e1}(t) - \hat{H}_{e1}(s) + H_{e1}(s)| + 2(H_{e1}(y+L_2 a_n) - H_{e1}(y-L_2 a_n))\} \\
&\le K \left\{ \sup{}^* \frac{|\hat{H}_e(t) - H_e(t) - \hat{H}_e(s) + H_e(s)|}{(1-H_e(t))^2} + O((na_n)^{-1/2} (\log a_n^{-1})^{1/2} a_n) \right\} \\
&= O((na_n)^{-1/2} (\log a_n^{-1})^{1/2} a_n^{1/2}),
\end{aligned}
$$

using Proposition A.4 and its analogue for H_{e1}, where $\sup^*$ denotes the supremum over all (s,t) satisfying $|H_e(t) - H_e(s)| \le K a_n$ and where the second inequality follows from the fact that $\sup_y |\hat{H}_e(y) - H_e(y)| = O((na_n)^{-1/2} (\log a_n^{-1})^{1/2})$ a.s. (see Proposition B.4 in Van Keilegom and Akritas (1999)). For the term $\alpha_{n2}(y)$ we have :

$$
\begin{aligned}
&\alpha_{n2}(y) \\
&= a_n^{-1} (\hat{H}_e(y_i) - H_e(y_i)) \{(1-H_e(y_i))^{-2} + o(1)\} \int (\hat{H}_{e1}(y - va_n) - H_{e1}(y - va_n))\, dK_2(v) \\
&= O((na_n^2)^{-1} \log a_n^{-1}),
\end{aligned}
$$

uniformly over all y.

Proposition A.7 *Assume (A1)-(A5). Then,*

$$
\begin{aligned}
&a_n^{-1} \int \{\hat{F}_e(y - va_n) - F_e(y - va_n)\}\, dK_2(v) \\
&= (na_n)^{-1} \sum_{i=1}^n \int \xi_e(E_i, \Delta_i, y - va_n)\, dK_2(v) + \beta(y) + R_n(y)
\end{aligned}
$$

where

$$
\beta(y) = \frac{1}{2} a_n^2 \mu_2^{K_2} \int \left[\frac{\partial}{\partial y} E(\varphi(x, Z, \Delta, y) | X = u) f_X(u) \right]'' \Bigg|_{u=x} dx
$$

and $\sup\{|R_n(y)|; -\infty < y \le T\} = o_P((na_n)^{-1/2}) + o_P(a_n^2)$.

Proof. We start with

$$a_n^{-1}\int(1-F_e(y-va_n))\left\{\int\limits_{-\infty}^{y-va_n}\frac{d\hat{H}_{e1}(s)}{1-\hat{H}_e(s)}-\int\limits_{-\infty}^{y-va_n}\frac{dH_{e1}(s)}{1-H_e(s)}\right\}dK_2(v)$$

$$\begin{aligned}
&= a_n^{-1}\int(1-F_e(y-va_n))\left\{\int\limits_{-\infty}^{y-va_n}\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}\right]dH_{e1}(s)\right.\\
&\quad+\int\limits_{-\infty}^{y-va_n}\frac{1}{1-H_e(s)}d(\hat{H}_{e1}(s)-H_{e1}(s))\\
&\quad\left.+\int\limits_{-\infty}^{y-va_n}\left[\frac{1}{1-\hat{H}_e(s)}-\frac{1}{1-H_e(s)}\right]d(\hat{H}_{e1}(s)-H_{e1}(s))\right\}dK_2(v).
\end{aligned}$$

The last term on the right hand side is $o_P((na_n)^{-1/2})$ by Proposition A.6. From the fact that $\sup_y|\hat{H}_e(y)-H_e(y)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ a.s. (see Proposition B.4 in Van Keilegom and Akritas (1999)), it follows that the sum of the first and second term can be written as

$$\begin{aligned}
&a_n^{-1}\int(1-F_e(y-va_n))\left\{\int\limits_{-\infty}^{y-va_n}\frac{\hat{H}_e(s)-H_e(s)}{(1-H_e(s))^2}dH_{e1}(s)+\frac{\hat{H}_{e1}(y-va_n)-H_{e1}(y-va_n)}{1-H_e(y-va_n)}\right.\\
&\left.-\int\limits_{-\infty}^{y-va_n}\frac{\hat{H}_{e1}(s)-H_{e1}(s)}{(1-H_e(s))^2}\,dH_e(s)\right\}dK_2(v)+o_P((na_n)^{-1/2}) \qquad \text{(A.14)}\\
&=\alpha_{n1}(y)+\alpha_{n2}(y)+\alpha_{n3}(y)+o_P((na_n)^{-1/2}).
\end{aligned}$$

We abbreviate the representation for $\hat{H}_e(y)$ given in Proposition A.2 by $\hat{H}_e(y)-H_e(y) = n^{-1}\sum_{i=1}^n A_{i\eta}(y)+n^{-1}\sum_{i=1}^n A_{i\zeta}(y)+n^{-1}\sum_{i=1}^n B_i(y)-\frac{1}{2}a_n^2\mu_2^{K_1}\int b_\eta(x)h_e(y|x)\,dx-\frac{1}{2}a_n^2\mu_2^{K_1}$ $\int b_\zeta(x)yh_e(y|x)\,dx+R_n(y)$. Then, $\alpha_{n1}(y)$ naturally splits into six terms. In a completely similar way as for the first term on the right hand side of (A.7), we can show that the first of the six terms of $\alpha_{n1}(y)$ equals

$$\begin{aligned}
&-n^{-1}\sum_{i=1}^n\int K_2(v)\left[-f_e(y-va_n)\int\limits_{-\infty}^{y-va_n}\frac{h_e(s|x)}{(1-H_e(s))^2}\,dH_{e1}(s)\right.\\
&\left.+(1-F_e(y-va_n))\frac{h_e(y-va_n|x)}{(1-H_e(y-va_n))^2}\,h_{e1}(y-va_n)\right]\,dv\,\eta(Z_i,\Delta_i|X_i)+O_P(n^{-1/2}a_n^2)\\
&=O_P(n^{-1/2})=o_P((na_n)^{-1/2}).
\end{aligned}$$

In a similar way, the second one is $o_P((na_n)^{-1/2})$. Using integration by parts, the fourth term of $\alpha_{n1}(y)$ equals

$$-\frac{1}{2}a_n^2\mu_2^{K_2}\int\int K_2(v)\left[(1-F_e(y-va_n))\int_{-\infty}^{y-va_n}\frac{h_e(s|x)}{(1-H_e(s))^2}dH_{e1}(s)\right]' dv\, b_\eta(x)\, dx$$

$$= -\frac{1}{2}a_n^2\mu_2^{K_2}\int\left[(1-F_e(y))\int_{-\infty}^{y}\frac{h_e(s|x)}{(1-H_e(s))^2}dH_{e1}(s)\right]' b_\eta(x)\, dx + o(a_n^2).$$

A similar expression can be given for the fifth term. Using again integration by parts, it is easy to show that the sixth term is $o_P(n^{-1/2}) + o_P(a_n^2)$ uniformly in y. The representation for $\alpha_{n2}(y)$ is given in Proposition A.3, while the derivation for $\alpha_{n3}(y)$ parallels completely that of $\alpha_{n1}(y)$. All together we have that (A.14) equals

$$(na_n)^{-1}\sum_{i=1}^n\int\xi_e(E_i,\Delta_i,y-va_n)dK_2(v)+\beta(y)+o_P((na_n)^{-1/2})+o_P(a_n^2).$$

Now, write (using a Taylor expansion)

$$\log(1-\hat{F}_e(y)) + \int_{-\infty}^{y}\frac{d\hat{H}_{e1}(s)}{1-\hat{H}_e(s-)}$$

$$= -\frac{1}{2}\sum_{i=1}^n\frac{I(\hat{E}_{(i)}\le y,\Delta_{(i)}=1)}{(n-i+1)^2}\frac{1}{(1-R_i)^2} = O(n^{-1}) \qquad \text{a.s.,}$$

uniformly in y, where R_i is between 0 and $1/(n-i+1)$. The result now follows by noting that $a_n^{-1}\int\{\hat{F}_e(y-va_n)-F_e(y-va_n)\}\,dK_2(v) = -a_n^{-1}\int(1-F_e(y-va_n))[\log(1-\hat{F}_e(y-va_n))-\log(1-F_e(y-va_n))]\,dK_2(v)+o_P(n^{-1/2})$, and that $\log(1-F_e(y)) = -\int_{-\infty}^{y}(1-H_e(s))^{-1}dH_{e1}(s)$, where the order of the remainder term follows from the fact that

$$\sup_y|\hat{F}_e(y)-F_e(y)| = O_P(n^{-1/2}+a_n^2). \qquad \text{(A.15)}$$

This can be derived from a slight generalization of the asymptotic representation for $\hat{F}_e(y)-F_e(y)$ given by Theorem 3.1 in Van Keilegom and Akritas (1999).

Proposition A.8 *Assume (A1)-(A5). Then,*

$$a_n^{-1}\int\{\hat{F}(y-va_n|x)-F(y-va_n|x)\}\,dK_2(v)$$

$$= (na_n)^{-1}\sum_{i=1}^n\int\xi_e\left(E_i,\Delta_i,\frac{y-va_n-m(x)}{\sigma(x)}\right)dK_2(v)$$

$$+ (na_n)^{-1}\sum_{i=1}^n K_1\left(\frac{x-X_i}{a_n}\right)g_{x,y}(Z_i,\Delta_i)+\frac{1}{\sigma(x)}\beta\left(\frac{y-m(x)}{\sigma(x)}\right)+R_n(y|x),$$

where $\sup\{|R_n(y|x)|; (x,y) \in \Omega\} = o_P((na_n)^{-1/2}) + o_P(a_n^2)$.

Proof. Write

$$\begin{aligned}
&a_n^{-1} \int \{\hat{F}(y - va_n|x) - F(y - va_n|x)\}\, dK_2(v) \\
&= a_n^{-1} \int \left\{ \left[\hat{F}_e\left(\frac{y - va_n - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - va_n - \hat{m}(x)}{\hat{\sigma}(x)}\right) \right] \right. \\
&\quad + \left[F_e\left(\frac{y - va_n - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - va_n - m(x)}{\hat{\sigma}(x)}\right) \right] \\
&\quad \left. + \left[F_e\left(\frac{y - va_n - m(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - va_n - m(x)}{\sigma(x)}\right) \right] \right\} dK_2(v) \\
&= \alpha_{n1}(x,y) + \alpha_{n2}(x,y) + \alpha_{n3}(x,y).
\end{aligned}$$

We start with $\alpha_{n2}(x,y)$.

$$\begin{aligned}
&\alpha_{n2}(x,y) \\
&= a_n^{-1} \int \left\{ -\frac{\hat{m}(x) - m(x)}{\hat{\sigma}(x)} f_e\left(\frac{y - va_n - m(x)}{\hat{\sigma}(x)}\right) + \frac{1}{2}\left(\frac{\hat{m}(x) - m(x)}{\hat{\sigma}(x)}\right)^2 f_e'(A_{xv}) \right\} dK_2(v),
\end{aligned}$$

for some A_{xv} between $\frac{y - va_n - m(x)}{\hat{\sigma}(x)}$ and $\frac{y - va_n - \hat{m}(x)}{\hat{\sigma}(x)}$. The second term of these two terms is $O(a_n^{-1}(na_n)^{-1} \log a_n^{-1})$ a.s., which follows from the consistency of $\hat{m}(x)$ given by Proposition 4.3 in Van Keilegom and Akritas (1999). For the first term, we first replace $\hat{\sigma}(x)$ by $\sigma(x)$ (using the same proposition) and write

$$a_n^{-1} \int f_e\left(\frac{y - va_n - m(x)}{\sigma(x)}\right) dK_2(v) = \sigma^{-1}(x) f_e'\left(\frac{y - m(x)}{\sigma(x)}\right) + o(a_n).$$

Using again the consistency of $\hat{m}(x)$ and applying the representation for $\hat{m}(x)$ given by Proposition 4.6 in the same paper, it follows that

$$\begin{aligned}
\alpha_{n2}(x,y) &= (na_n)^{-1} f_X^{-1}(x) \sigma^{-1}(x) \sum_{i=1}^n K_1\left(\frac{x - X_i}{a_n}\right) \eta(Z_i, \Delta_i|x) f_e'\left(\frac{y - m(x)}{\sigma(x)}\right) \\
&\quad + o_P((na_n)^{-1/2}).
\end{aligned}$$

In a similar way we can show that

$$\begin{aligned}
\alpha_{n3}(x,y) &= (na_n)^{-1} f_X^{-1}(x) \sigma^{-1}(x) \sum_{i=1}^n K_1\left(\frac{x - X_i}{a_n}\right) \zeta(Z_i, \Delta_i|x) \\
&\quad \times \left\{ \frac{y - m(x)}{\sigma(x)} f_e'\left(\frac{y - m(x)}{\sigma(x)}\right) + f_e\left(\frac{y - m(x)}{\sigma(x)}\right) \right\} + o_P((na_n)^{-1/2}).
\end{aligned}$$

The above show that

$$\alpha_{n2}(x,y)+\alpha_{n3}(x,y)=(na_n)^{-1}\sum_{i=1}^{n}K_1\left(\frac{x-X_i}{a_n}\right)g_{x,y}(Z_i,\Delta_i)+o_P((na_n)^{-1/2}).$$

On the term $\alpha_{n1}(x,y)$ we will apply Proposition A.7. Note however that the argument of the function F_e in Proposition A.7 is $y-va_n$, while in $\alpha_{n1}(x,y)$ the argument is $\frac{y-va_n-\hat{m}(x)}{\hat{\sigma}(x)}$. It can be shown that the following slight variant of Proposition A.7 is valid :

$$\alpha_{n1}(x,y)=(na_n)^{-1}\sum_{i=1}^{n}\int\xi_e\left(E_i,\Delta_i,\frac{y-va_n-\hat{m}(x)}{\hat{\sigma}(x)}\right)dK_2(v)$$
$$+\frac{1}{\hat{\sigma}(x)}\beta\left(\frac{y-\hat{m}(x)}{\hat{\sigma}(x)}\right)+o_P((na_n)^{-1/2})+o_P(a_n^2).$$

It is easy to see that

$$\frac{1}{\hat{\sigma}(x)}\beta\left(\frac{y-\hat{m}(x)}{\hat{\sigma}(x)}\right)-\frac{1}{\sigma(x)}\beta\left(\frac{y-m(x)}{\sigma(x)}\right)=o(a_n^2).$$

Hence, using the notation $e_v(y)=(y-va_n-m(x))/\sigma(x)$ and $\hat{e}_v(y)=(y-va_n-\hat{m}(x))/\hat{\sigma}(x)$, it remains to show that

$$(na_n)^{-1}\sum_{i=1}^{n}\int[\xi_e(E_i,\Delta_i,\hat{e}_v(y))-\xi_e(E_i,\Delta_i,e_v(y))]\,dK_2(v)=o_P((na_n)^{-1/2}).$$

Note that

$$(na_n)^{-1}\sum_{i=1}^{n}\xi_e(E_i,\Delta_i,\hat{e}_v(y))$$
$$=a_n^{-1}(1-F_e(\hat{e}_v(y)))\left\{\int_{-\infty}^{\hat{e}_v(y)}\frac{\hat{H}_e(s)-H_e(s)}{(1-H_e(s))^2}\,dH_{e1}(s)+\frac{\hat{H}_{e1}(\hat{e}_v(y))-H_{e1}(\hat{e}_v(y))}{1-H_e(\hat{e}_v(y))}\right.$$
$$\left.-\int_{-\infty}^{\hat{e}_v(y)}\frac{\hat{H}_{e1}(s)-H_{e1}(s)}{(1-H_e(s))^2}\,dH_e(s)\right\}.$$

It is easy to show that replacing $1-F_e(\hat{e}_v(y))$ in the above expression by $1-F_e(e_v(y))$ yields a remainder term of order $O((na_n^2)^{-1}\log a_n^{-1})$ a.s., using the rate of convergence of $\hat{H}_e(y)$ and $\hat{m}(x)$ (see Propositions B.4, respectively 4.3 in Van Keilegom and Akritas (1999)). The result now follows from Proposition A.5 and (again) the rate of convergence of $\hat{H}_e(y)$ and $\hat{m}(x)$.

References

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

Dette, H. and Gefeller, O. (1995). The impact of different definitions of nearest neighbour distances for censored data on nearest neighbour kernel estimators of the hazard rate. *J. Nonpar. Statist.*, **4**, 271-282.

Gefeller, O. and Dette, H. (1992). Nearest neighbour kernel estimation of the hazard function from censored data. *J. Statist. Comput. Simul.*, **43**, 93-101.

Gray, R.J. (1996). Hazard rate regression using ordinary nonparametric regression smoothers. *J. Comput. Graph. Statist.*, **5**, 190-207.

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

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

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

Lo, S.-H., Mack, Y.P. and Wang, J.-L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Th. Rel. Fields*, **80**, 461-473.

Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. *Probab. Th. Rel. Fields*, **71**, 455-465.

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

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

Rosenblatt, M. (1971). Curve estimates. *Ann. Math. Statist.*, **42**, 1815-1842.

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

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York.

Van Keilegom, I. and Akritas, M.G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.*, **27**, 1745-1784.