I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2014/24

The copula-graphic estimator in censored nonparametric location-scale regression models

SUJICA, A. AND I. VAN KEILEGOM

The copula-graphic estimator in censored nonparametric location-scale regression models

Aleksandar Sujica *   Ingrid Van Keilegom *

June 4, 2014

Abstract

A common assumption when working with randomly right censored data, is the independence between the variable of interest Y (the survival time) and the censoring variable C. This assumption, which is not testable, is however unrealistic in certain situations. In this paper we assume that for a given covariate X, the dependence between the variables Y and C is described via a known copula. Additionally we assume that Y is the response variable of a heteroscedastic regression model $Y = m(X) + \sigma(X)\varepsilon$, where the error term ε is independent of the explanatory variable X, and the functions m and σ are 'smooth'. We propose an estimator of the conditional distribution of Y given X under this model, and show the asymptotic normality of this estimator. We also study the small sample performance of the estimator, and discuss the advantages/drawbacks of this estimator with respect to competing estimators.

Keywords and phrases: Asymptotic normality, asymptotic representation, copula, dependent censoring, kernel estimator, nonparametric regression, right censoring.

*Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium. E-mail address: aleksandar.sujica@uclouvain.be, ingrid.vankeilegom@uclouvain.be. This research was supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy), and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain").

1 Introduction

Consider the following nonparametric location-scale model

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

where the error ε is assumed to be independent of a one dimensional covariate X. The function $m(\cdot)$ is a conditional location functional and $\sigma(\cdot)$ is a conditional scale functional representing possible heteroscedasticity. We assume that Y is a possible (given) transformation of a survival time and is subject to random right censoring, i.e. instead of observing Y we only observe (T, Δ), where $T = \min(Y, C)$, $\Delta = I(Y \leq C)$ and C represents the censoring time. Let (T_i, X_i, Δ_i), $i = 1, \ldots, n$ be n independent vectors having the same distribution as (T, X, Δ).

The motivation for considering model (1.1) comes from the fact that the model offers important advantages with respect to the completely nonparametric model when one is interested in the estimation of the conditional distribution $F(\cdot|x) = P(Y \leq \cdot|X = x)$ of Y given $X = x$. Van Keilegom and Akritas (1999) showed how advantage can be taken from model (1.1) to estimate this conditional distribution. The advantages are especially apparent in the right tail of the distribution. In this region the completely nonparametric competitor proposed by Beran (1981) (see also Dabrowska (1989), González-Manteiga and Cadarso-Suarez (1994), Akritas (1994), Van Keilegom and Veraverbeke (1997), Du and Akritas (2002), among others) suffers from inconsistency problems especially when censoring is heavy. This phenomenon is similar to what happens in the right tail of the Kaplan-Meier estimator in the absence of covariates. Under model (1.1), Van Keilegom and Akritas (1999) showed that the right tail of the distribution $F(\cdot|x)$ can be well estimated for all values of X, provided there is a region of X where censoring is light. This is because under model (1.1) the conditional distribution $F(y|x)$ can be written as

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

where $F_e(\cdot)$ is the distribution of the error variable ε, and this error distribution is the same for all x.

Aside Van Keilegom and Akritas (1999) the nonparametric estimation of the above location-scale model with censored data has also been studied in other papers. See e.g. Lewbel and Linton (2002), Van Keilegom and Veraverbeke (2002), Chen, Dahl, and Khan (2005), Linton, Mammen, Nielsen, and Van Keilegom (2011) and Lambert (2013). Instead of studying the estimation of the model, other authors have investigated testing procedures for several aspects of the model (again in the case of censored data). We refer to Pardo-Fernández and Van Keilegom (2006) and Dette and Heuchenne (2012) for goodness-of-fit tests for the location respectively scale function, and to Pardo-Fernández, Van Keilegom, and González-Manteiga (2007) for comparing regression curves under this model. Finally, semiparametric location scale models with censored data have been studied in Heuchenne and Van Keilegom (2007a,b) among others.

Whereas all the above papers restrict attention to the case where Y and C are independent given X, we will go one step further in this paper, and consider the case where for a given value of X, the survival time and censoring time are related. The motivation for considering this situation is multifold. In many situations, the latent censoring mechanism is not of pure administrative or random nature, but is at the contrary linked (in a weak or strong way) to the survival time. This is e.g. the case when the medical condition of a patient (good or bad) makes him/her decide to leave the study or to change treatment. It is also the case in a study on the duration of unemployment, where a person might decide after long and unsuccessful attempts to find a job, to move to another region where the job market is more attractive, and hence this person will be lost to follow up. In addition, in many situations the strength of the dependence between Y and C will depend on the value of the covariate(s). Therefore, in this paper we will allow the dependence between Y and C to depend on the value of X. We will model this dependence by means of a copula function, because copulas have the attractive feature to model the dependence structure without affecting the margins (see Sklar (1959)).

In the absence of covariates and leaving the marginal distributions of Y and C completely unspecified, Zheng and Klein (1995) and Rivest and Wells (2001) supposed that the dependence structure between Y and C is known and is described by a known copula, and they showed that the marginal distributions of Y and C are identifiable under very weak conditions. They developed an estimator of the distribution of Y, which they called the copula-graphic estimator, and which reduces to the Kaplan-Meier estimator when Y and C are independent. In the presence of covariates, Braekers and Veraverbeke (2005) extended the work of Rivest and Wells (2001) to the case of a fixed design regression model, and they proposed and studied an estimator of the conditional distribution $F(\cdot|x)$ for a given covariate x without assuming any model restriction on $F(\cdot|x)$. Their estimator generalizes the Beran (1981) estimator, in the sense that it reduces to Beran's estimator when the independence copula is chosen. In the case of a random design, Sujica and Van Keilegom (2013) built further on the work of Braekers and Veraverbeke (2005). They proposed estimators of a location and scale function of Y given X and studied their asymptotic properties.

In this paper we will propose and study an estimator of the conditional distribution $F(\cdot|X)$ assuming that Y and C are copula dependent given X (as in Braekers and Veraverbeke (2005)), and that (X, Y) satisfy the nonparametric location-scale model (1.1). We will do this by first estimating the marginal error distribution $F_e(\cdot)$ taking the dependence between Y and C into account. Next, the conditional distribution $F(\cdot|x)$ will be estimated via relation (1.2), by plugging-in the obtained estimator of F_e and the estimators of $m(\cdot)$ and $\sigma(\cdot)$ studied by Sujica and Van Keilegom (2013).

Before continuing we give an overview of the paper. In the next section we introduce the precise definitions of the estimators of the error distribution and of the conditional distribution of Y given X, and we state the assumptions needed for our asymptotic results. In Section 3 we give the main results. Proofs of the main results are in Section 4. Section 5 contains results of a small sample comparison between the suggested estimator and the estimators of Van Keilegom and Akritas (1999) and Braekers and Veraverbeke (2005). Additional lemmas needed for the main results are in Appendices A and B.

2 Estimation method

We start this section with a number of definitions. Let the random vector (T, X, Δ) be as defined in Section 1 and denote $F(y|x) = P(Y \leq y|X = x)$, $G(y|x) = P(C \leq y|X = x)$, $H(y|x) = P(T \leq y|X = x)$, $H^u(y|x) = P(T \leq y, \Delta = 1|X = x)$ and $F_X(x) = P(X \leq x)$. Further, denote $F_e(y) = P(\varepsilon \leq y) = P(\frac{Y-m(X)}{\sigma(X)} \leq y)$, $G_e(y|x) = P(\frac{C-m(X)}{\sigma(X)} \leq y|X = x)$, and for $E = (T - m(X))/\sigma(X)$ denote $H_e(y) = P(E \leq y)$, $H_e^u(y) = P(E \leq y, \Delta = 1)$, $H_e(y|x) = P(E \leq y|X = x)$ and $H_e^u(y|x) = P(E \leq y, \Delta = 1|X = x)$. The probability density functions of the distribution functions defined above will be denoted by the corresponding lower case letters, and for any distribution function $F(\cdot)$, we denote the corresponding survival function by $\overline{F}(\cdot) = 1 - F(\cdot)$.

As explained in Section 1, we build further on the work of Zheng and Klein (1995), Rivest and Wells (2001), and Braekers and Veraverbeke (2005), and model the conditional dependence between Y and C via a known copula $\mathcal{C}_x$ that is allowed to depend on the value of X:

$$P(Y > y,\ C > c|X = x) = \mathcal{C}_x\big(1 - F(y|x), 1 - G(c|x)\big).$$

Moreover, we will assume that the copula belongs to the family of Archimedean copulas, which have attractive properties and which cover a broad range of different copula structures. This means that we suppose that

$$P(Y > y,\ C > c|X = x) \quad = \quad \phi_x^{-1}\Big[\phi_x\big\{1 - F(y|x)\big\} + \phi_x\big\{1 - G(c|x)\big\}\Big], \qquad (2.1)$$

for an Archimedean copula generator ϕ_x, i.e. a function from $(0, 1]$ to $I\!R^+$ that is decreasing, convex and that satisfies $\phi_x(1) = 0$.

In order to construct an estimator of the conditional distribution $F(y|x)$ given in (1.2), we start with focusing on the error distribution $F_e(y)$. The assumption of an Archimedean copula allows to write $F_e(y)$ in the following way:

$$\overline{F}_e(y) \quad = \quad \phi_{(y)}^{-1}\Big\{-\int_{B_y}\int_{-\infty}^{y}\phi'_x(\overline{H}_e(s|x))dH_e^u(s|x)dF_X(x)\Big\} \tag{2.2}$$

(see Lemma A.1 in Appendix A), where $\phi_{(y)}(u) = \int_{B_y}\phi_x(u)\,dF_X(x)$. Statement (2.2) holds for every nonempty set $B_y \subset A_y := \{x : \overline{H}_e(y|x) > \eta\}$ $(\eta > 0)$, which will be defined later. In order to derive an estimator of $F_e(\cdot)$, we will replace the distribution functions H_e, H_e^u and F_X in (2.2) by corresponding estimators $\widehat{H}_e$, $\widehat{H}_e^u$ and $\widehat{F}_X$.

We start with F_X, which we estimate by the empirical distribution function $\widehat{F}_X(\cdot) = n^{-1}\sum_{i=1}^n I(X_i \le \cdot)$. Next, to estimate H_e and H_e^u, we first need to find appropriate estimators of the functions m and σ, for which we use the following definitions:

$$m(x) = \int_0^1 F^{-1}(s|x)J(s)ds \quad \text{and} \quad \sigma^2(x) = \int_0^1 F^{-1}(s|x)^2J(s)ds - m(x)^2, \tag{2.3}$$

where $F^{-1}(s|x) = \inf\{y : F(y|x) \ge s\}$ and $J(s)$ is a given score function satisfying $\int_0^1 J(s)ds = 1$.

To estimate the functions $m(x)$ and $\sigma(x)$, we replace the conditional distribution $F(y|x)$ in (2.3) by the conditional copula-graphic estimator, introduced by Braekers and Veraverbeke (2005) and adopted to random design by Sujica and Van Keilegom (2013):

$$\overline{\widetilde{F}}(y|x) = \phi_x^{-1}\left\{-\sum_{T_i\le y,\Delta_i=1}\left[\phi_x\left(\overline{\widehat{H}}(T_i^-|x)\right) - \phi_x\left(\overline{\widehat{H}}(T_i|x)\right)\right]\right\}. \tag{2.4}$$

Here, $\widehat{H}(y|x) = \sum_{i=1}^n W_{ni}(x,h_n)I(T_i \le y)$ is the Stone (1977) estimator of the distribution of T given $X = x$, where

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

are Nadaraya-Watson weights, K is a kernel function, h_n is a bandwidth sequence, and $\overline{\widehat{H}}(y^-|x) = \lim_{t\uparrow y}\overline{\widehat{H}}(t|x)$. This leads to

$$\widehat{m}(x) = \int_0^1 \widetilde{F}^{-1}(s|x)J(s)ds \qquad \text{and} \qquad \widehat{\sigma}^2(x) = \int_0^1 \widetilde{F}^{-1}(s|x)^2J(s)ds - \widehat{m}(x)^2, \tag{2.5}$$

where the score function $J(s)$ is chosen in such a way that $\widehat{m}(x)$ and $\widehat{\sigma}(x)$ are consistent. The estimators $\widehat{m}(x)$ and $\widehat{\sigma}^2(x)$ have been the object of study in Sujica and Van Keilegom (2013), and are generalizations of the estimators proposed by Van Keilegom and Akritas (1999) to the case where Y and C are copula dependent given X.

Next, we estimate the (sub)distribution functions $H_e(y|x)$ and $H_e^u(y|x)$ by the following Stone (1977)-type estimators:

$$\widehat{H}_e(y|x) = \sum_{i=1}^n W_{ni}(x, h_n) I\big(\widehat{E}_i \le y\big) \text{ and } \widehat{H}_e^u(y|x) = \sum_{i=1}^n W_{ni}(x, h_n) I(\Delta_i = 1) I\big(\widehat{E}_i \le y\big),$$

where $\widehat{E}_i = (T_i - \widehat{m}(X_i))/\widehat{\sigma}(X_i)$, $i = 1, \dots, n$.

Plugging-in the estimators $\widehat{F}_X$, $\widehat{H}_e$ and $\widehat{H}_e^u$ in (2.2), we obtain the following estimator of the error distribution:

$$\widehat{\overline{F}}_e(y) \;=\; \widehat{\phi}_{(y)}^{-1}\Big\{ -\int_{B_y} \int_{-\infty}^y \phi_x'(\widehat{\overline{H}}_e(s|x)) d\widehat{H}_e^u(s|x) d\widehat{F}_X(x) \Big\}, \tag{2.6}$$

where $\widehat{\phi}_{(y)}(u) = \int_{B_y} \phi_x(u)\, d\widehat{F}_X(x)$. We choose the set B_y in (2.6) as a subset of the set A_y that excludes all small segments, that is $B_y = \text{argmax}_{B \in \mathcal{B}, B \subset A_y} \lambda(B)$, where $\mathcal{B} = \{B = \cup_i B_i : B \text{ is nonempty}, \; B_i \text{ is convex}, \lambda(B_i) \ge \beta\}$, λ is the Lebesgue measure and $\beta > 0$ is an arbitrary small constant. It is easy to show that $\{x \mapsto I_B(x) : B \in \mathcal{B}\}$ is a Donsker class of functions (where $I_B(x) = I(x \in B)$). We stress that the set B_y could also be estimated, but proving the asymptotic results in Section 3 would fill the paper with very technical details with no significant contribution (for more details see Remarks 3.1 and 3.2).

Finally, (2.6) together with (2.5) lead to our final estimator:

$$\widehat{F}(y|x) \;=\; \widehat{F}_e\left(\frac{y - \widehat{m}(x)}{\widehat{\sigma}(x)}\right). \tag{2.7}$$

3 Asymptotic results

3.1 Definitions and assumptions

The primary objective of this section is to study the asymptotic distribution of the estimators $\widehat{F}_e(y)$ and $\widehat{F}(y|x)$, proposed in the previous section. For establishing the asymptotic representation of these estimators we will need the following functions:

$$\begin{aligned}\xi_e(E, \Delta, y|X) = &-\int_{E\wedge y}^y \phi_X''(\overline{H}_e(s|X)) dH_e^u(s|X) + \int_{-\infty}^y \phi_X''(\overline{H}_e(s|X)) H_e(s|X) dH_e^u(s|X) \\ &+ \phi_X'(\overline{H}_e(y|X)) I(E \le y, \Delta = 1) - \int_{-\infty}^y \phi_X'(\overline{H}_e(s|X)) dH_e^u(s|X),\end{aligned}$$

$$
\begin{aligned}
\xi(T,\Delta,y|X) =& \frac{-1}{\phi'_X\left(\overline{F}(y|X)\right)}\bigg\{\int_{-\infty}^{y}\phi''_X(\overline{H}(s|X))[I(T\le s)-H(s|X)]dH^u(s|X)\\
&-\phi'_X\left(\overline{H}(y|X)\right)[I(T\le y,\Delta=1)-H^u(y|X)]\\
&-\int_{-\infty}^{y}\phi''_X(\overline{H}(s|X))[I(T\le s,\Delta=1)-H^u(s|X)]dH(s|X)\bigg\},\\
\eta\left(T,\Delta|X\right) =& \int_{-\infty}^{+\infty} J\left(F(y|X)\right)\xi\left(T,\Delta,y|X\right)dy,\\
\zeta\left(T,\Delta|X\right) =& \int_{-\infty}^{\infty} J(F(y|X))\xi\left(T,\Delta,y|X\right)\frac{y-m(X)}{\sigma(X)}dy,\\
\gamma_1(y|X) =& -\int_{-\infty}^{y}\phi''_X(\overline{H}_e(s|X))h_e(s|X)dH_e^u(s|X)+\int_{-\infty}^{y}\phi'_X(\overline{H}_e(s|X))dh_e^u(s|X),\\
\gamma_2(y|X) =& -\int_{-\infty}^{y}\phi''_X(\overline{H}_e(s|X))sh_e(s|X)dH_e^u(s|X)+\int_{-\infty}^{y}\phi'_X(\overline{H}_e(s|X))d(sh_e^u(s|X)).
\end{aligned}
$$

Finally, let $\tau_\eta = \inf\{y : \overline{H}_e(y) > \eta\}$ for some small $\eta > 0$. The following assumptions are important for proving the asymptotic results in the next section.

(A1) (*i*) The sequence h_n satisfies $nh_n^4 = o(1)$ and $nh_n^{3+2\delta}(\log h_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.

(*ii*) The support R_X of X is a bounded interval in $I\!R$.

(*iii*) The probability density function K has compact support $[-a, a]$ for some $a > 0$, $\int uK(u)du = 0$ and K is twice continuously differentiable.

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

(A2) (*i*) There exist $0 \le s_0 \le s_1 \le 1$ such that $s_1 < \inf_x F(\widetilde{T}_x|x)$, $s_0 < \inf\{s \in [0,1] : J(s) \neq 0\}$, $s_1 > \sup\{s \in [0,1] : J(s) \neq 0\}$ and $\inf_{x\in R_X}\inf_{s_0\le s\le s_1} f(F^{-1}(s|x)|x) > 0$.

(*ii*) The function J is bounded and twice continuously differentiable on the interval (s_0, s_1), $\int_0^1 J(s)ds = 1$ and $J(s) \ge 0$ for all $0 \le s \le 1$.

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

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

(A4) (*i*) The functions $\phi'_x(u) = \frac{\partial}{\partial u}\phi_x(u)$, $\phi''_x(u)$ and $\phi_x^{(3)}(u)$ exist and are continuous in $(x, u) \in R_X \times (0, 1]$.

(ii) The functions $\ddot{\phi}''_x(u) = \frac{\partial^4}{\partial x^2 \partial u^2}\phi_x(u)$, $\dot{\phi}_x^{(3)}(u)$ and $\phi_x^{(4)}(u)$ exist and are continuous in $(x,u) \in R_X \times (0,1]$.

(iii) The function ϕ_x satisfies $\phi'_x(1) < 0$.

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. (In the proofs, the function $L(y|x)$ of assumption (A5) will be either $H(y|x)$, $H_e(y|x)$, $H^u(y|x)$ or $H_e^u(y|x)$.)

(A5) (i) $L(y|x)$ is continuous in (x,y).

(ii) $L'(y|x)$ exists, is continuous in (x,y) and $\sup_{x,y}|yL'(y|x)| < \infty$.

(iii) $L''(y|x)$ exists, is continuous in (x,y) and $\sup_{x,y}|y^2L''(y|x)| < \infty$.

(iv) $\dot{L}(y|x)$ exists, is continuous in (x,y) and $\sup_{x,y}|y\dot{L}(y|x)| < \infty$.

(v) $\ddot{L}(y|x)$ exists, is continuous in (x,y) and $\sup_{x,y}|y^2\ddot{L}(y|x)| < \infty$.

(vi) $\dot{L}'(y|x)$ exists, is continuous in (x,y) and $\sup_{x,y}|y\dot{L}'(y|x)| < \infty$.

(vii) $\ddot{L}'(y|x)$ exists, is continuous in (x,y) and $\sup_{x,y}|y\ddot{L}'(y|x)| < \infty$.

(A6) There exist continuous and non-decreasing functions M_j with $M_j(-\infty) = 0$ and $M_j(+\infty) < \infty$ $(j = 1, ..., 4)$ such that

$$
\begin{aligned}
|L(y_2|x) - L(y_1|x)| &\le |M_1(y_2) - M_1(y_1)|, \\
|\dot{L}(y_2|x) - \dot{L}(y_1|x)| &\le |M_2(y_2) - M_2(y_1)|, \\
|L'(y_2|x) - L'(y_1|x)| &\le |M_3(y_2) - M_3(y_1)|, \\
|L'(y_2|x)y_2 - L'(y_1|x)y_1| &\le |M_4(y_2) - M_4(y_1)|,
\end{aligned}
$$

for all $x \in R_X, -\infty < y_1, y_2 < +\infty$.

Note that assumption (A6) comes from Du and Akritas (2002), and is required to prove an i.i.d. representation for our estimator $\widehat{F}(y|x)$, whose remainder term is negligible uniformly in x and y (for details see Lemma B.6).

Throughout the rest of this paper, we let C denote a generic positive constant, whose value may differ from line to line.

3.2 Asymptotic properties of the estimator $\widehat{F}_e(y)$

We will extend the result in Van Keilegom and Akritas (1999) concerning the weak convergence of the estimator of the residual distribution function under independent censoring to the case where the dependence between censoring and survival time is described via a known copula. The weak convergence of the estimator will follow from its asymptotic representation, which we give first.

Theorem 3.1. *[Asymptotic representation for $\widehat{F}_e(y)$] Assume (A1)-(A4), and assume that (A5) and (A6) hold for $H(y|x)$ and $H^u(y|x)$. Let $y \le \tau_\eta$. Then,*

$$\widehat{F}_e(y) - F_e(y) = n^{-1} \sum_{i=1}^{n} k_y(T_i, \Delta_i, X_i) + R_n(y),$$

where $\sup\{|R_n(y)| : -\infty < y \le \tau_\eta\} = o_P\left(n^{-1/2}\right)$*, and*

$$\begin{aligned} k_y(T,\Delta,X) \;=\; \frac{1}{\phi'_{(y)}(\bar{F}_e(y))} \Bigg[& I(X \in B_y)\xi_e(E,\Delta,y|X) \\ & + I(X \in B_y)\eta\,(T,\Delta|X)\,\frac{\gamma_1(y|X)}{\sigma(X)} \\ & + I(X \in B_y)\zeta\,(T,\Delta|X)\,\frac{\gamma_2(y|X)}{\sigma(X)} \\ & + \left\{ I(X \in B_y) \int_{-\infty}^{y} \phi'_X(\overline{H}_e(s|X))dH_e^u(s|X) \right. \\ & \left. \quad - \int_{B_y} \int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x))dH_e^u(s|x)dF_X(x) \right\} \\ & + \left\{ I(X \in B_y)\phi_X(\bar{F}_e(y)) - \int_{B_y} \phi_x(\bar{F}_e(y))dF_X(x) \right\} \Bigg]. \end{aligned}$$

Note that if we replace B_y by R_X, $H_e(y|x)$ by $H_e(y)$, $H_e^u(y|x)$ by $H_e^u(y)$, and set $\phi_x(u) = -\log u$, the first term above corresponds to the i.i.d. representation of the usual Kaplan-Meier estimator due to Lo and Singh (1986). Furthermore, under the same changes, the first three terms give exactly the i.i.d. representation of the estimator studied in Van Keilegom and Akritas (1999) in the case of independent censoring. The second and third term come from the fact that in the estimating procedure we replaced $(T_i - m(X_i))/\sigma(X_i)$ by $(T_i - \widehat{m}(X_i))/\widehat{\sigma}(X_i)$. Finally, the fourth and fifth terms above are caused by replacing $F_X(x)$ by $\widehat{F}_X(x)$.

Corollary 3.1. *[Weak convergence of $\widehat{F}_e(y)$] Under the assumptions of Theorem 3.1, the process $n^{1/2}(\widehat{F}_e(y) - F_e(y))$, $-\infty < y \le \tau_\eta$ converges weakly to a zero-mean Gaussian process $Z(y)$ with covariance function*

$$Cov(Z(y), Z(y')) = Cov\left(k_y(T,\Delta,X), k_{y'}(T,\Delta,X)\right).$$

3.3 Asymptotic properties of the estimator $\widehat{F}(y|x)$

Using the results from Theorem 3.1 and Corollary 3.1 we will show the asymptotic representation and the weak convergence of the estimator $\widehat{F}(y|x)$ of the conditional distribution under dependent censoring described via a known copula. This result will extend the results in Van Keilegom and Akritas (1999) which are obtained under independent censoring, to the case of dependent censoring described by a copula model.

Theorem 3.2. *[Asymptotic representation for $\widehat{F}(y|x)$] Assume (A1)-(A4), and assume that (A5) and (A6) hold for $H(y|x)$ and $H^u(y|x)$. Let $(y - m(x))/\sigma(x) \le \tau_\eta$. Then,*

$$\begin{aligned}\widehat{F}(y|x) - F(y|x) &= \widehat{F}_e\left(\frac{y-\widehat{m}(x)}{\widehat{\sigma}(x)}\right) - F_e\left(\frac{y-m(x)}{\sigma(x)}\right) \\ &= (nh_n)^{-1}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) h_y(T_i,\Delta_i|x) + R_n(y|x),\end{aligned}$$

where $\sup\{|R_n(y|x)| : (x,y)\in\Omega\} = o_P\left((nh_n)^{-1/2}\right)$, $\Omega = \{(x,y) : (y-m(x))/\sigma(x) \le \tau_\eta,\ x \in R_X\}$ *and*

$$h_y(T,\Delta|x) = \left[\eta\left(T,\Delta|x\right) + \zeta\left(T,\Delta|x\right)\frac{y-m(x)}{\sigma(x)}\right] f_e\left(\frac{y-m(x)}{\sigma(x)}\right)\sigma^{-1}(x) f_X^{-1}(x).$$

Corollary 3.2. *[Weak convergence of $\widehat{F}(y|x)$] Under the assumptions of Theorem 3.2, the process $(nh_n)^{1/2}(\widehat{F}(y|x) - F(y|x))$, $x \in R_X$ fixed, $(y-m(x))/\sigma(x) \le \tau_\eta$, converges weakly to a zero-mean Gaussian process $Z(y|x)$ with covariance function*

$$Cov(Z(y|x), Z(y'|x)) = f_X(x)\int K^2(u)du\ Cov\left(h_y(T,\Delta|x), h_{y'}(T,\Delta|x)|X = x\right).$$

Remark 3.1. *It can be shown that all the results from this section hold (with no extra terms in the i.i.d. representations in Theorems 3.1 and 3.2), if we replace B_y by an estimator $\widehat{B}_y$ that is converging a.s. to B_y in Lebesgue measure for every fixed $-\infty < y \le \tau_\eta$, and for which the set $\widehat{B}_y$ belongs a.s. to $\mathcal{B}$ for all $-\infty < y \le \tau_\eta$. The required modifications of the proofs are very technical with a small contribution of novelty, and are therefore omitted.*

Remark 3.2. *Using the notation $A_y(\eta) = \{x : \overline{H}_e(y|x) > \eta\}$ and $\widehat{A}_y(\eta) = \{x : \widehat{\overline{H}}_e(y|x) > \eta\}$, one potential estimator of B_y is $\widehat{B}_y = \text{argmax}_{B\in\mathcal{B}, B\subset\widehat{A}_y(\eta+a_n)}\lambda(B)$, where $a_n = (nh_n)^{-1/2}(\log n)^{1/2+\alpha}$ and $\alpha > 0$. This estimator satisfies the conditions of Remark 3.1 by the following reasoning. First, because of Lemma B.4, we have $P(\lim_{n\to\infty} I\{A_y(\eta+2a_n) \subset \widehat{A}_y(\eta + a_n) \subset A_y(\eta)\} = 1) = 1$. Additionally, simple calculus shows that the function*

$\eta \mapsto \lambda[A_y(\eta)]$ is a right continuous function, that is $\lambda[A_y(\eta + 2a_n)] \to \lambda[A_y(\eta)]$ a.s. Hence, we have a.s. convergence in Lebesgue measure of $\lambda[\widehat{A}_y(\eta + a_n)]$ to $\lambda[A_y(\eta)]$, which implies that $\lambda[\widehat{B}_y]$ converges to $\lambda[B_y]$. On the other hand, by definition, the set $\widehat{B}_y$ belongs to $\mathcal{B}$ for all $-\infty < y \leq \tau_\eta$.

4 Proofs of main results

In this section we give the proofs of the main results from Section 3. The proofs are based on a number of technical lemmas and propositions, which can be found in Sujica and Van Keilegom (2013) and Appendices A and B.

Proof of Theorem 3.1. [Asymptotic representation for $\widehat{F}_e(y)$]
First, we will brake down $\widehat{F}_e(y) - F_e(y)$ into several terms, in such a way that each term depends on a single plug-in estimator. This decomposition will end at (4.6). Then, in a second step we will deal with each term in this decomposition. We start by using Lemma A.1:

$$\begin{aligned}
\widehat{F}_e(y) - F_e(y) &= \phi_{(y)}^{-1}\Big\{-\int_{B_y}\int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x))dH_e^u(s|x)dF_X(x)\Big\} \\
&\quad - \widehat{\phi}_{(y)}^{-1}\Big\{-\int_{B_y}\int_{-\infty}^{y} \phi'_x(\widehat{\overline{H}}_e(s|x))d\widehat{H}_e^u(s|x)d\widehat{F}_X(x)\Big\} \\
&= \Big[\phi_{(y)}^{-1}\{U(y)\} - \phi_{(y)}^{-1}\{U_n(y)\}\Big] + \Big[\phi_{(y)}^{-1}\{U_n(y)\} - \phi_{(y)}^{-1}\{\widetilde{U}_n(y)\}\Big] \\
&\quad + \Big[\phi_{(y)}^{-1}\{\widetilde{U}_n(y)\} - \widehat{\phi}_{(y)}^{-1}\{\widetilde{U}_n(y)\}\Big] \\
&:= (\mathbf{I}) + (\mathbf{II}) + (\mathbf{III}),
\end{aligned} \tag{4.1}$$

where

$$\begin{aligned}
U(y) &= -\int_{B_y}\int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x))dH_e^u(s|x)dF_X(x), \\
U_n(y) &= -\int_{B_y}\int_{-\infty}^{y} \phi'_x(\widehat{\overline{H}}_e(s|x))d\widehat{H}_e^u(s|x)dF_X(x) \quad \text{and} \\
\widetilde{U}_n(y) &= -\int_{B_y}\int_{-\infty}^{y} \phi'_x(\widehat{\overline{H}}_e(s|x))d\widehat{H}_e^u(s|x)d\widehat{F}_X(x).
\end{aligned} \tag{4.2}$$

Next, we examine each of the three terms. Starting with the first one, we have by a second order Taylor expansion:

$$(\mathbf{I}) \;=\; \frac{1}{\phi'_{(y)}\Big(\bar{F}_e(y)\Big)}\big\{U(y) - U_n(y)\big\} + R_{n1}^{(I)}(y),$$

where

$$R_{n1}^{(I)}(y) \quad = \quad \frac{\phi''_{(y)}\left(\phi_{(y)}^{-1}(\varepsilon_1(y))\right)}{2\phi'_{(y)}\left(\phi_{(y)}^{-1}(\varepsilon_1(y))\right)^3}\{U(y)-U_n(y)\}^2,$$

with $\varepsilon_1(y)$ an intermediate value between $U_n(y)$ and $U(y)$. By adding and subtracting terms, we further have that

$$\begin{aligned}
(\mathbf{I}) = {} & \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)}\bigg\{\int_{B_y}\int_{-\infty}^{y}\left[\phi'_x(\widehat{\overline{H}}_e(s|x)) - \phi'_x(\overline{H}_e(s|x))\right]dH_e^u(s|x)dF_X(x) \\
& \quad + \int_{B_y}\int_{-\infty}^{y}\phi'_x(\overline{H}_e(s|x))d\left[\widehat{H}_e^u(s|x) - H_e^u(s|x)\right]dF_X(x)\bigg\} + R_{n1}^{(I)}(y) + R_{n2}^{(I)}(y) \\
= {} & \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)}\bigg\{-\int_{B_y}\int_{-\infty}^{y}\phi''_x(\overline{H}_e(s|x))\left[\widehat{H}_e(s|x) - H_e(s|x)\right]dH_e^u(s|x)dF_X(x) \\
& \quad + \int_{B_y}\int_{-\infty}^{y}\phi'_x(\overline{H}_e(s|x))d\left[\widehat{H}_e^u(s|x) - H_e^u(s|x)\right]dF_X(x)\bigg\} \\
& + R_{n1}^{(I)}(y) + R_{n2}^{(I)}(y) + R_{n3}^{(I)}(y), \qquad (4.3)
\end{aligned}$$

where

$$\begin{aligned}
R_{n2}^{(I)}(y) &= \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)}\int_{B_y}\int_{-\infty}^{y}\left[\phi'_x(\widehat{\overline{H}}_e(s|x)) - \phi'_x(\overline{H}_e(s|x))\right]d\left[\widehat{H}_e^u(s|x) - H_e^u(s|x)\right]dF_X(x), \\
R_{n3}^{(I)}(y) &= \frac{1}{2\phi'_{(y)}\left(\bar{F}_e(y)\right)}\int_{B_y}\int_{-\infty}^{y}\phi'''_x(\xi_1(s,x))\left[\widehat{H}_e(s|x) - H_e(s|x)\right]^2 dH_e^u(s|x)dF_X(x),
\end{aligned}$$

with $\xi_1(s,x)$ between $\widehat{\overline{H}}_e(s|x)$ and $\overline{H}_e(s|x)$.

Next, we examine $(\mathbf{II})$.

$$(\mathbf{II}) = \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)}\{U_n(y) - \widetilde{U}_n(y)\} + R_{n1}^{(II)}(y) + R_{n2}^{(II)}(y),$$

where

$$\begin{aligned}
R_{n1}^{(II)}(y) &= \frac{\phi''_{(y)}\left(\phi_{(y)}^{-1}(\varepsilon_2(y))\right)}{2\phi'_{(y)}\left(\phi_{(y)}^{-1}(\varepsilon_2(y))\right)^3}\{U_n(y) - \widetilde{U}_n(y)\}^2, \\
R_{n2}^{(II)}(y) &= \left(\frac{1}{\phi'_{(y)}\left(\phi_{(y)}^{-1}(U_n(y))\right)} - \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)}\right)\{U_n(y) - \widetilde{U}_n(y)\},
\end{aligned}$$

with $\varepsilon_2(y)$ between $\widetilde{U}_n(y)$ and $U_n(y)$. Let us further decompose (**II**):

$$\begin{aligned}(\mathbf{II}) \quad = \quad & \frac{1}{\phi'_{(y)}(\bar{F}_e(y))} \int_{B_y} \int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x)) dH_e^u(s|x) d\left[\widehat{F}_X(x) - F_X(x)\right] \\ & + R_{n1}^{(II)}(y) + R_{n2}^{(II)}(y) + R_{n3}^{(II)}(y) + R_{n4}^{(II)}(y), \end{aligned} \tag{4.4}$$

where

$$R_{n3}^{(II)}(y) = \frac{1}{\phi'_{(y)}(\bar{F}_e(y))} \int_{B_y} \int_{-\infty}^{y} \left[\phi'_x(\widehat{\overline{H}}_e(s|x)) - \phi'_x(\overline{H}_e(s|x))\right] d\widehat{H}_e^u(s|x) d\left[\widehat{F}_X(x) - F_X(x)\right],$$

$$R_{n4}^{(II)}(y) = \frac{1}{\phi'_{(y)}(\bar{F}_e(y))} \int_{B_y} \int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x)) d\left[\widehat{H}_e^u(s|x) - H_e^u(s|x)\right] d\left[\widehat{F}_X(x) - F_X(x)\right].$$

Next, we examine (**III**). By applying a second order Taylor expansion on $\widehat{\phi}_{(y)}\{\phi_{(y)}^{-1}(\widetilde{U}_n(y))\} - \widehat{\phi}_{(y)}\{\widehat{\phi}_{(y)}^{-1}(\widetilde{U}_n(y))\}$, we obtain

$$\begin{aligned}(\mathbf{III}) &= \phi_{(y)}^{-1}\left(\widetilde{U}_n(y)\right) - \widehat{\phi}_{(y)}^{-1}\left(\widetilde{U}_n(y)\right) \\ &= \frac{1}{\widehat{\phi}'_{(y)}\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right)}\left[\widehat{\phi}_{(y)}\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right) - \widetilde{U}_n(y)\right] + R_{n1}^{(III)}(y) \\ &= \frac{1}{\phi'_{(y)}\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right)}\left[\widehat{\phi}_{(y)}\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right) - \widetilde{U}_n(y)\right] + R_{n1}^{(III)}(y) + R_{n2}^{(III)}(y) \\ &= \frac{1}{\phi'_{(y)}\left(\phi_{(y)}^{-1}(U(y))\right)}\left[\widehat{\phi}_{(y)}\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right) - \widetilde{U}_n(y)\right] + R_{n1}^{(III)}(y) + R_{n2}^{(III)}(y) + R_{n3}^{(III)}(y)\end{aligned}$$

where

$$R_{n1}^{(III)}(y) = \frac{-\widehat{\phi}''_{(y)}\left(\xi_2(y)\right)}{2\widehat{\phi}'_{(y)}\left(\phi_{(y)}^{-1}\left(\widetilde{U}_n(y)\right)\right)}\left[\phi_{(y)}^{-1}\left(\widetilde{U}_n(y)\right) - \widehat{\phi}_{(y)}^{-1}\left(\widetilde{U}_n(y)\right)\right]^2,$$

$$R_{n2}^{(III)}(y) = \left[\frac{1}{\widehat{\phi}'_{(y)}\left(\phi_{(y)}^{-1}\left(\widetilde{U}_n(y)\right)\right)} - \frac{1}{\phi'_{(y)}\left(\phi_{(y)}^{-1}\left(\widetilde{U}_n(y)\right)\right)}\right]\left[\widehat{\phi}_{(y)}\left(\phi_{(y)}^{-1}\left(\widetilde{U}_n(y)\right)\right) - \widetilde{U}_n(y)\right],$$

$$R_{n3}^{(III)}(y) = \left(\frac{1}{\phi'_{(y)}\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right)} - \frac{1}{\phi'_{(y)}\left(\phi_{(y)}^{-1}(U(y))\right)}\right)\left[\widehat{\phi}_{(y)}\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right) - \widetilde{U}_n(y)\right],$$

with $\xi_2(y)$ between $\widehat{\phi}_{(y)}^{-1}(\widetilde{U}_n(y))$ and $\phi_{(y)}^{-1}(\widetilde{U}_n(y))$. Further, we continue with (**III**) by calculating

$$
\begin{aligned}
&(\mathbf{III}) - R_{n1}^{(III)}(y) - R_{n2}^{(III)}(y) - R_{n3}^{(III)}(y)\\
&= \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)} \int_{B_y} \phi_x\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right) d\left[\widehat{F}_X(x) - F_X(x)\right]\\
&= \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)} \int_{B_y} \phi_x\left(\phi_{(y)}^{-1}(U(y))\right) d\left[\widehat{F}_X(x) - F_X(x)\right] + R_{n4}^{(III)}(y)\\
&= \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)} \int_{B_y} \phi_x\left(\bar{F}_e(y)\right) d\left[\widehat{F}_X(x) - F_X(x)\right] + R_{n4}^{(III)}(y), \qquad (4.5)
\end{aligned}
$$

where

$$
R_{n4}^{(III)}(y) = \frac{1}{\phi'_{(y)}(\bar{F}_e(y))} \int_{B_y} \left[\phi_x\left(\phi_{(y)}^{-1}(\widetilde{U}_n(y))\right) - \phi_x\left(\phi_{(y)}^{-1}(U(y))\right)\right] d\left[\widehat{F}_X(x) - F_X(x)\right].
$$

Combining (4.1), (4.3), (4.4) and (4.5), we subsequently obtain

$$
\begin{aligned}
\widehat{F}_e(y) - F_e(y) =& \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)} \Bigg\{ - \int_{B_y}\int_{-\infty}^{y} \phi''_x(\overline{H}_e(s|x))\left[\widehat{H}_e(s|x) - H_e(s|x)\right] dH_e^u(s|x)dF_X(x)\\
&+ \int_{B_y}\int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x)) d\left[\widehat{H}_e^u(s|x) - H_e^u(s|x)\right] dF_X(x)\\
&+ \int_{B_y}\int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x)) dH_e^u(s|x) d\left[\widehat{F}_X(x) - F_X(x)\right]\\
&+ \int_{B_y} \phi_x\left(\bar{F}_e(y)\right) d\left[\widehat{F}_X(x) - F_X(x)\right]\Bigg\} + R_n(y)\\
:=& \frac{1}{\phi'_{(y)}\left(\bar{F}_e(y)\right)} \left\{\widehat{\Lambda}_1(y) + \widehat{\Lambda}_2(y) + \widehat{\Lambda}_3(y) + \widehat{\Lambda}_4(y)\right\} + R_n(y). \qquad (4.6)
\end{aligned}
$$

where $R_n(y) = R_{n1}^{(I)} + R_{n2}^{(I)}(y) + R_{n3}^{(I)}(y) + R_{n1}^{(II)}(y) + R_{n2}^{(II)}(y) + R_{n3}^{(II)}(y) + R_{n4}^{(II)}(y) + R_{n1}^{(III)}(y) + R_{n2}^{(III)}(y) + R_{n3}^{(III)}(y) + R_{n4}^{(III)}(y)$. Now, by applying Lemmas A.2 and A.3 for the main terms we get after some simple algebra that

$$
\widehat{F}_e(y) - F_e(y) = \frac{1}{n}\sum_{i=1}^{n} k_y(T_i, \Delta_i, X_i) + R_n(y) + o_P(n^{-1/2}),
$$

uniformly in $-\infty < y \le \tau_\eta$.

Remainder terms $R_{n2}^{(I)}(y)$ and $R_{n3}^{(I)}(y)$ are $o(n^{-1/2})$ a.s. by Lemma B.5 and Lemma B.4, respectively. Now, by using a Taylor expansion, we can easily show that each of the other remainder terms in $R_n(y)$ can be a.s. bounded by a product consisting of some of the following terms: $U_n - U$, $\widetilde{U}_n - U$, $\widehat{\phi}_{(y)} - \phi_{(y)}$, $\int_{B_y} G_n(y,x)d[\widehat{F}_X(x) - F_X(x)]$ (where G_n is a function satisfying the assumptions of Lemma B.2). From here it is easy to verify that all the remainder terms are $o(n^{-1/2})$ a.s. by using the orders given in Lemmas B.1 to B.4. □

Proof of Corollary 3.1. We will prove the weak convergence by showing that the class of functions $\mathcal{K} = \{k_y : -\infty < y \leq \tau_\eta\}$ from Theorem 3.1 is a Donsker class. The function $x \mapsto [\phi'_{(y)}(\bar{F}_e(y))]^{-1}$ is a uniformly bounded, deterministic function because of assumptions (A3)(i), (A4)(iii) and because B_y is nonempty for all $-\infty < y \leq \tau_\eta$. The term inside the square brackets in the definition of k_y belongs to a sum of Donsker classes (see (4.6) and Lemmas A.2 and A.3), which is also a Donsker class by Example 2.10.7 in Van der Vaart and Wellner (1996). Furthermore, it is easy to show that k_y is uniformly bounded. Hence, since multiplying uniformly bounded functions preserves the Donsker property (by Example 2.10.8 in Van der Vaart and Wellner (1996)), the class of functions $\mathcal{K}$ is Donsker. □

Proof of Theorem 3.2. [Asymptotic representation for $\widehat{F}(y|x)$] Write

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

Note that $(nh_n)^{1/2}\alpha_n^1(x,y) = o_P(1)$ uniformly in $(x,y) \in \Omega$, because of the weak convergence result established in Corollary 3.1. For $\alpha_n^2(x,y)$ we have

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

for some ξ_x between $(y-m(x))/\widehat{\sigma}(x)$ and $(y-\widehat{m}(x))/\widehat{\sigma}(x)$. The second term above is of order $O\left((nh_n)^{-1}\log n\right)$ a.s. by Proposition 3.5 in Sujica and Van Keilegom (2013), together with the boundedness of f_e' (which follows from assumption (A5)(ii)). For the first term, we first replace $\widehat{\sigma}(x)$ by $\sigma(x)$ by using Proposition 3.5 and then apply Proposition 3.8 in Sujica and Van Keilegom (2013) to obtain an asymptotic representation. For $\alpha_n^3(x,y)$ we have

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

where ξ_{2x} is between $(y-m(x))/\sigma(x)$ and $(y-m(x))/\widehat{\sigma}(x)$. The second term above is $O\left((nh_n)^{-1}\log n\right)$ a.s. by Proposition 3.5 in Sujica and Van Keilegom (2013), the fact that

$\sup_y |y^2 f_e'(y)| < \infty$ and assumption (A3)(ii). After replacing $\widehat{\sigma}$ with σ by using again Proposition 3.5 in Sujica and Van Keilegom (2013), the first term above has an asymptotic representation given by Proposition 3.8 in Sujica and Van Keilegom (2013). This combined with the asymptotic representation for $\alpha_n^2(x, y)$ completes the proof. □

Proof of Corollary 3.2. The proof is similar to the proof of Corollary 3.4. in Van Keilegom and Akritas (1999). It suffices in fact to replace the function ξ defined in Van Keilegom and Akritas (1999) by the function ξ defined in Section 3.1. Apart from this, the proof is identical. □

5 Simulations

In this section we illustrate the finite sample behavior of our estimator $\widehat{F}(y|x)$ by means of Monte Carlo simulations. We compare the estimator $\widehat{F}(y|x)$ proposed in this paper with the estimator $\widetilde{F}(y|x)$ proposed by Braekers and Veraverbeke (2005) and defined in (2.4). We expect that under the assumption of the location-scale model (1.1), the estimator $\widehat{F}(y|x)$ outperforms $\widetilde{F}(y|x)$. Further, the estimator $\widehat{F}(y|x)$ is a generalization of the estimator proposed by Van Keilegom and Akritas (1999) and the two will be compared under the assumptions of the latter estimator, that is under independence between Y and C given X. We expect that they perform similarly.

To compare the performance of the estimators we use the mean squared error (MSE) and the integrated mean squared error ($IMSE$), to be defined further on. The simulations are carried out for samples of size $n = 100$, $n = 200$ and $n = 400$, and the results are obtained by using 500 Monte Carlo simulations.

In the first setting, we generate i.i.d. data from the following regression model:

$$Y = 8(X - 0.5)^2 + 0.5\varepsilon,$$

where X has a uniform distribution on $[0, 1]$ and the error ε has a standard normal distribution and is independent of X. The censoring variable C satisfies $C = \alpha_1(X-0.5)^2+\alpha_2+0.5\widetilde{\varepsilon}$, where $\widetilde{\varepsilon}$ is standard normal and independent of X, and the constants α_1 and α_2 are chosen so that the global censoring rate is 45% and the local censoring rate (for a fixed value of x) is between 42% and 48%. Finally, to model the dependence between Y and C given $X = x$ (i.e. the dependence between ε and $\widetilde{\varepsilon}$ given $X = x$) we use a Gumbel copula:

$$\mathcal{C}_x(u_1, u_2) = \exp\left\{-\big[-(\log u_1)^{\gamma(x)} - (\log u_2)^{\gamma(x)}\big]^{1/\gamma(x)}\right\}, \tag{5.1}$$

where $\gamma(x) = \max(5 - 5x, 1)$. This means that the corresponding Archimedean copula generator equals $\phi_x(u) = -(\log u)^{\gamma(x)}$. Under this setting the conditional dependence between

Y and C given $X = x$ decreases from strong positive dependence to complete independence as x goes from 0 to 1 (Kendall's tau coefficient decreases from 0.8 to 0). We work with the score function $J(s) = b^{-1}I(0 \le s \le b)$. In order to estimate the functionals $m(\cdot)$ and $\sigma(\cdot)$ consistently the constant b has to be smaller than or equal to $\inf_{x\in R_X} \widehat{F}(+\infty|x)$. Therefore, we choose $b = 0.8$ which is smaller than the average of 1000 simulated infima. Note that

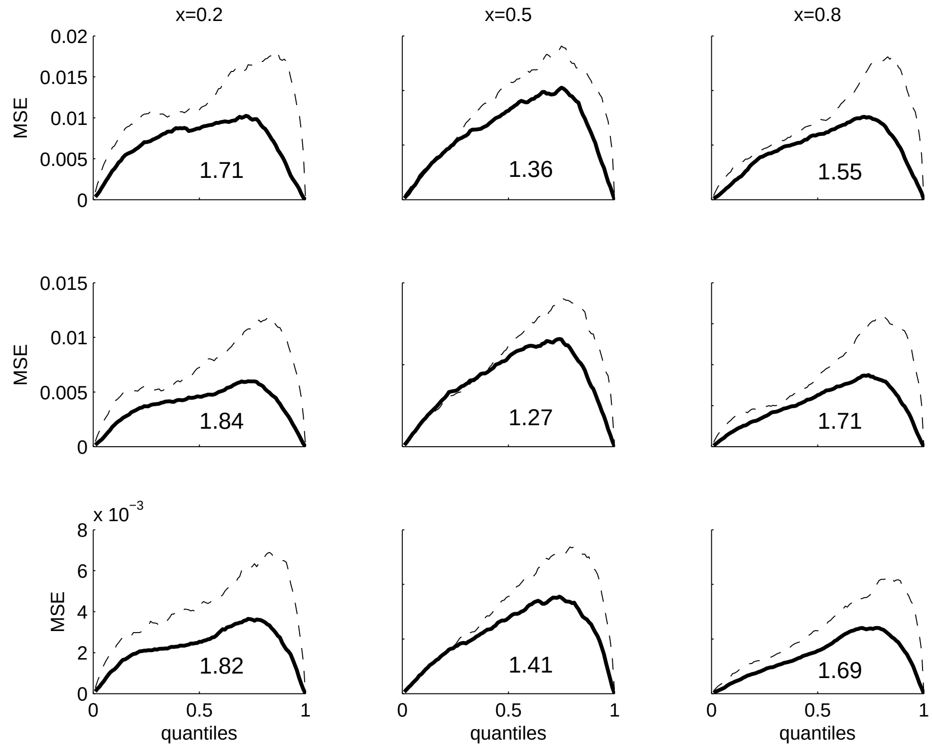

Figure 1: MSE of $\widehat{F}(y|x)$ (solid line) and $\widetilde{F}(y|x)$ (dashed line) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for covariate $x = 0.2$, $x = 0.5$ and $x = 0.8$ (column 1, column 2, column 3). The number under the curve represents the ratio of the approximated $IMSE$ of $\widetilde{F}(\cdot|x)$ and $\widehat{F}(\cdot|x)$.

an equivalent way of writing the estimator $\widehat{\sigma}^2(x)$ is

$$\widehat{\sigma}^2(x) = \sum_{i=1}^{n} \big[Y_i - \widehat{m}(x)\big]^2 \big[\widetilde{F}_b(Y_i|x) - \widetilde{F}_b(Y_i^-|x)\big] J(\widetilde{F}_b(Y_i|x)),$$

where $\widetilde{F}_b(y|x) := \min\{\widetilde{F}(y|x), b\}$, and this is close to

$$\sum_{i=1}^{n} \big[Y_i - \widehat{m}(X_i)\big]^2 \big[\widetilde{F}_b(Y_i|x) - \widetilde{F}_b(Y_i^-|x)\big] J(\widetilde{F}_b(Y_i|x)).$$

In the sequel we work with the latter estimator, since simulations showed that it outperforms the former (which is expected, since $Y_i - m(X_i)$ is a proper 'error', whereas $Y_i - m(x)$ is not).

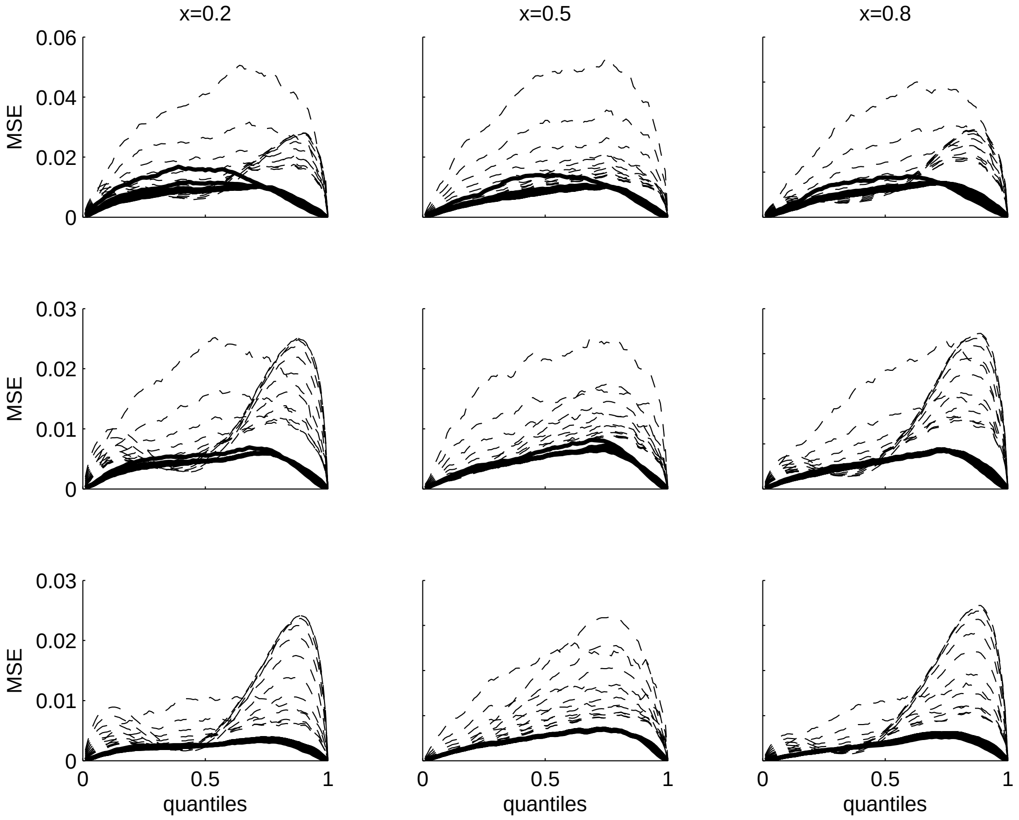

Figure 2: MSE of $\widehat{F}(y|x)$ (solid lines) and $\widetilde{F}(y|x)$ (dashed lines) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for covariate $x = 0.2$, $x = 0.5$ and $x = 0.8$ (column 1, column 2, column 3). Every subgraph contains the curves of $\widetilde{F}(\cdot|x)$ and $\widehat{F}(\cdot|x)$ for different values of the bandwidth ranging from 0.075 to 0.350 (with increments of 0.025). The bandwidths for estimating $m(\cdot)$ and $\sigma(\cdot)$ are taken equal to $h_n(\widehat{m})$ and $h_n(\widehat{\sigma})$ respectively.

Also, we can easily show that both estimators are asymptotically equivalent under certain assumptions on the bandwidth.

For the weights that appear in our estimators $\widetilde{F}(y|x)$ and $\widehat{F}(y|x)$ we use the kernel function $K(u) = (15/16)(1-u^2)^2 I(|u| \leq 1)$. For a fixed x, to select a bandwidth for $\widetilde{F}(\cdot|x)$, we minimize the integrated mean squared error $IMSE(\widetilde{F}(\cdot|x)) := E[\int\{\widetilde{F}(y|x) - F(y|x)\}^2 dF(y|x)]$ over a grid of 12 equidistant possible values of h_n between 0.075 and 0.350. The so-obtained estimator is denoted by $h_n(\widetilde{F}(\cdot|x))$. To calculate $IMSE(\widetilde{F}(\cdot|x))$, we use 500 simulated data sets. For each simulated data set, we compute the integrated squared error $\int\{\widetilde{F}(y|x) - F(y|x)\}^2 dF(y|x)$, and we approximate $IMSE(\widetilde{F}(\cdot|x))$ by taking the average over these 500 values. On the other hand, to estimate $\widehat{F}(\cdot|x)$ we need to choose 4 bandwidths corresponding to $\widehat{m}(\cdot)$, $\widehat{\sigma}(\cdot)$, $\widehat{H}_e(\cdot|\cdot)$ and $\widehat{H}_e^u(\cdot|\cdot)$. In the first step we choose bandwidths $h_n(\widehat{m})$ and $h_n(\widehat{\sigma})$ for $\widehat{m}(\cdot)$ and $\widehat{\sigma}(\cdot)$, so that they minimize the approximated

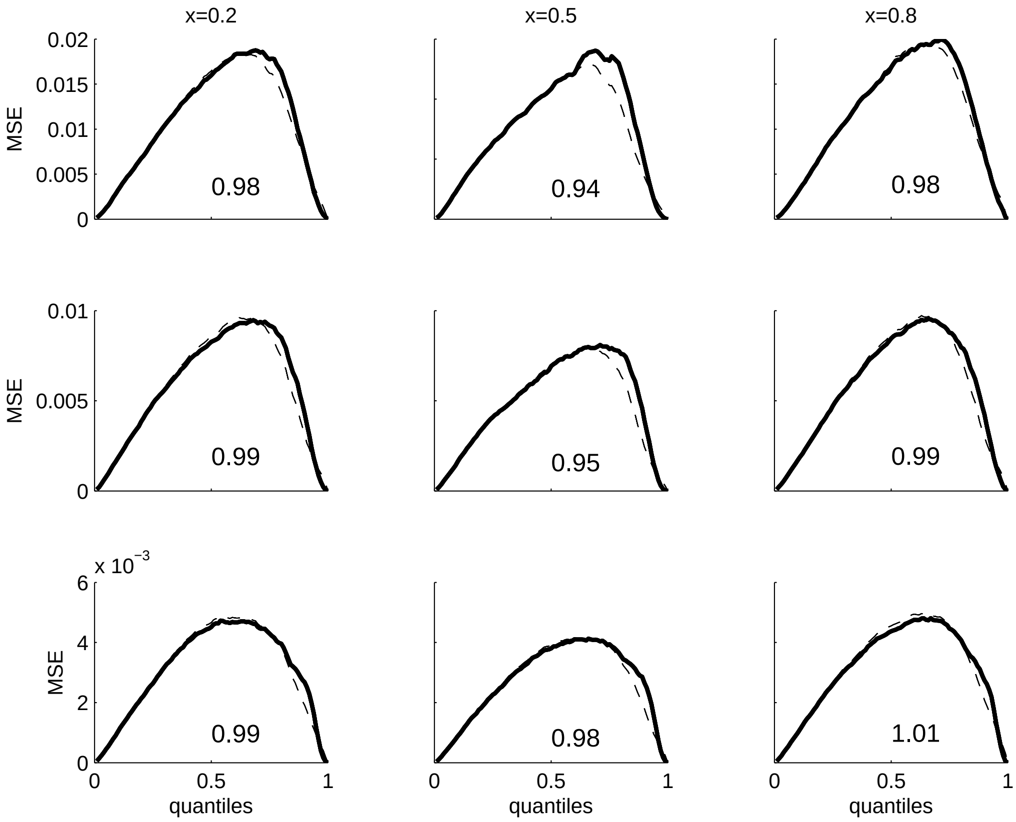

Figure 3: MSE of $\check{F}(y|x)$ (dashed line) and $\widehat{F}(y|x)$ (solid line) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for covariate $x = 0.2$, $x = 0.5$ and $x = 0.8$ (column 1, column 2, column 3). The number under the curve represents the ratio of the approximated $IMSE$ of $\check{F}(\cdot|x)$ and $\widehat{F}(\cdot|x)$.

$IMSE(\widehat{m}) := E[\int_{0.2}^{0.8}\{\widehat{m}(x) - m(x)\}^2 dF_X(x)]$ and $IMSE(\widehat{\sigma})$, respectively, where, by abuse of notation, $IMSE$ is used to denote both the integrated mean squared error of $\widehat{m}$, $\widehat{\sigma}$, $\widetilde{F}(\cdot|x)$ and $\widehat{F}(\cdot|x)$. Note that we do not take into account values of x close to the boundary of the support of X to avoid boundary effects of the Nadaraya-Watson weights. We calculate the approximated $IMSE(\widehat{m})$ and $IMSE(\widehat{\sigma})$ by using the same approach as for calculating the approximated $IMSE(\widetilde{F}(\cdot|x))$. In the next step we choose the same bandwidth for both $\widehat{H}_e(\cdot|x)$ and $\widehat{H}_e^u(\cdot|x)$ so that it minimizes the corresponding approximated $IMSE(\widehat{F}(\cdot|x))$ (where $\widehat{F}(\cdot|x)$ uses the bandwidths $h_n(\widehat{m})$ and $h_n(\widehat{\sigma})$ for $\widehat{m}$ and $\widehat{\sigma}$, obtained in the previous step). This bandwidth is denoted by $h_n(\widehat{F}(\cdot|x))$.

Figure 1 shows the MSE of $\widetilde{F}(y|x)$ and $\widehat{F}(y|x)$ for bandwidths chosen by the above procedure. Each subgraph contains the ratio of the approximated $IMSE$ of $\widetilde{F}(\cdot|x)$ and $\widehat{F}(\cdot|x)$, which shows that $\widehat{F}(\cdot|x)$ outperforms $\widetilde{F}(\cdot|x)$ for all sample sizes and all values of the covariate. We believe that this is a consequence of the fact that $\widehat{F}(y|x)$ uses the extra information given by the location-scale regression model (1.1). One could argue that in this comparison, having two optimization steps for choosing the bandwidths for $\widehat{F}(\cdot|x)$ gives an unfair advantage over a one step optimization for choosing the bandwidth for $\widetilde{F}(\cdot|x)$. To address this issue, in Figure 2, we plot the MSE of $\widetilde{F}(y|x)$ and of $\widehat{F}(y|x)$ for several bandwidths ranging from 0.075 to 0.350. To estimate m and σ we use the optimal bandwidths $h_n(\widehat{m})$ and $h_n(\widehat{\sigma})$ for all curves. Figure 2 shows that the second optimization step has little influence on the performance of the estimator $\widehat{F}(y|x)$, which once again shows the advantage of our estimator $\widehat{F}(y|x)$ over $\widetilde{F}(y|x)$.

Figure 3 shows the MSE of the estimator $\check{F}(y|x)$ and the estimator $\widehat{F}(y|x)$ under the assumption that Y and C are independent given X (i.e. in (5.1) we use the independent copula). The estimator $\check{F}(y|x)$ is defined by relation (2.7), when $\phi_x(t) = -\log t$ (which is used in estimating $\widehat{m}$ and $\widehat{\sigma}$, that is in estimating $\widetilde{F}$), and $\widehat{F}_e$ is replaced by

$$\check{F}_e(y) := 1 - \prod_{\widehat{E}_i \le y, \Delta_i = 1} \left(1 - \frac{1}{n-i+1}\right).$$

Note that we use the same $\widehat{m}$ and $\widehat{\sigma}$ as plug-in estimators for $\check{F}(\cdot|x)$ as we do for $\widehat{F}(\cdot|x)$. The only bandwidth selection for $\check{F}(\cdot|x)$ is the one used for $\widehat{m}$ and $\widehat{\sigma}$, and is obtained as above. The second-step bandwidth for $\widehat{F}(\cdot|x)$ is chosen from the grid [0.075,0.350] as before, so that it minimizes $IMSE(\widehat{F}(\cdot|x))$. As expected, we see that the MSE of the estimators $\widehat{F}(y|x)$ and $\check{F}(y|x)$ are virtually identical, with the ratio of the approximated $IMSE$ of $\widehat{F}(\cdot|x)$ and $\check{F}(\cdot|x)$ approaching 1, as the sample size increases.

A Appendix A

Appendices A and B contain the results needed for the proof of Theorem 3.1. Lemma A.1 is a small technical result that helps to break down $\widehat{F}_e(y) - F_e(y)$ into a sum of terms each depending only on one plug-in estimator, and remainder terms (see (4.6)). The two main lemmas in this section are Lemmas A.2 and A.3, which derive asymptotic representations for the main terms in the representation of $\widehat{F}_e(y) - F_e(y)$, while Appendix B contains the results needed for showing the negligibility of the remainder terms.

In the proofs of Lemmas A.2 and A.3 we use empirical process theory. The result that we use most frequently is Lemma A.5, which is an adaptation of a result in Giné and Nickle (2008) to our framework. This result is allowing us to asymptotically approximate sums of the form $\sum_{i=1}^n \int \widehat{g}_y(x, V_i) W_{ni}(x, h_n) dF_X(x)$ by $\frac{1}{n}\sum_{i=1}^n \widehat{g}_y(X_i, V_i) + o_P(n^{-1/2})$, where $V_i = (X_i, T_i)$. Lemma A.4 is used to show the Donsker property of classes of functions appearing in the proofs of Lemmas A.2 and A.3.

Before proceeding, we define a certain set of functions that will be used throughout Appendix A:

$$C_M^\alpha(R_X) := \{f : R_X \mapsto R : \ \|f\|_\alpha < M\},$$
$$\widetilde{C}_M^\alpha(R_X) := \{f \in C_M^\alpha(R_X) : \ \inf_{x\in R_X} f(x) > b_{\inf}\}, \tag{A.1}$$

where $0 < M < \infty$, $b_{\inf} = \inf_{x\in R_X} \sigma(x)/2$ and

$$\|f\|_\alpha := \max_{k\le\lfloor\alpha\rfloor} \sup_{x\in R_X} |f^{(k)}(x)| + \sup_{x_1,x_2\in R_X} \left|f^{\lfloor\alpha\rfloor}(x_1) - f^{\lfloor\alpha\rfloor}(x_2)\right| |x_1 - x_2|^{-(\alpha-\lfloor\alpha\rfloor)},$$

where $\lfloor\alpha\rfloor$ is the greatest integer smaller then α.

The following lemma gives a specific form for the error distribution F_e, which we use to construct the estimator $\widehat{F}_e$.

Lemma A.1. *Assume that $H(y|x)$ and $H^u(y|x)$ satisfy (A5)(ii) and let $\phi_x'(u)$ exist for $(x,u) \in R_X \times (0,1]$. Then for every $y \le \tau_\eta$,*

$$\overline{F}_e(y) \ = \ \phi_{(y)}^{-1}\Big\{ - \int_{B_y} \int_{-\infty}^y \phi_x'(\overline{H}_e(s|x))dH_e^u(s|x)dF_X(x)\Big\}.$$

Proof. Define

$$\overline{H}_e(y_1, y_2|x) = P\left(\frac{Y - m(X)}{\sigma(X)} > y_1, \frac{C - m(X)}{\sigma(X)} > y_2 \middle| X = x\right).$$

Then, using relation (2.1), we can easily show that $\overline{H}_e(y_1, y_2|x) = \phi_x^{-1}\{\phi_x(\overline{F}_e(y_1))+\phi_x(\overline{G}_e(y_2|x)\}$. Using Tsiatis (1975), for all $y \le \tau_\eta$ we get

$$H_e^{u'}(y|x) = -\frac{d}{dy_1}\overline{H}_e(y_1, y_2|x)\Big|_{y_1=y_2=y} = -\frac{\phi_x'\left(\overline{F}_e(y)\right)\overline{F}_e'(y)}{\phi_x'\left(\overline{H}_e(y|x)\right)}.$$

For every x such that $\overline{H}_e(y|x) > \eta$ this leads to

$$-\int_{-\infty}^y \phi_x'(\overline{H}_e(s|x))dH_e^u(s|x) = \int_{-\infty}^y \phi_x'(\overline{F}_e(s))\overline{F}_e'(s)ds$$
$$= \int_1^{\overline{F}_e(y)} \phi_x'(w)dw = \phi_x\left(\overline{F}_e(y)\right).$$

Now, by integrating over the set B_y with respect to $dF_X(x)$, and solving for $F_e(y)$ we finish the proof. □

In the next two lemmas we establish i.i.d. representations for the terms $\widehat{\Lambda}_1(y)$, $\widehat{\Lambda}_2(y)$, $\widehat{\Lambda}_3(y)$ and $\widehat{\Lambda}_4(y)$ given in (4.6). These representations are needed for the i.i.d. representation of $\widehat{F}_e(y) - F_e(y)$ given in Theorem 3.1.

Lemma A.2. *[Asymptotic representation for $\widehat{\Lambda}_1(y)$]*
Assume (A1)-(A4), and assume that (A5) holds for $H(y|x)$ and $H^u(y|x)$. Then for $y \le \tau_\eta$,

$$\widehat{\Lambda}_1(y) \;=\; \frac{1}{n}\sum_{i=1}^n g_y^{(1)}(X_i,T_i,\Delta_i) + R_n(y),$$

where $g_y^{(1)}(X,T,\Delta) = I(X \in B_y)\int_{-\infty}^y \phi_X''(\overline{H}_e(s|X))[-\frac{h_e(s|X)}{\sigma(X)}\{s\zeta(T,\Delta|X)+\eta(T,\Delta|X)\} - I(E \le s) + H_e(s|X)]dH_e^u(s|X)$, $\sup_{y\le\tau_\eta}|R_n(y)| = o_P(n^{-1/2})$ and $\widehat{\Lambda}_1(y)$ is defined in (4.6). Furthermore, the class of functions $\mathcal{G}^{(1)} = \{(x,t,\delta) \mapsto g_y^{(1)}(x,t,\delta) : -\infty < y \le \tau_\eta\}$ is a Donsker class of functions.

***Proof*.** By using the notation

$$g^{[1]}_{y,\widehat{m},\widehat{\sigma}}(x_1,x_2,t) = I(x_1\in B_y)\int_{-\infty}^y \phi_{x_1}''\left(\overline{H}_e(s|x_1)\right)\left[H_e(s|x_1) - I\left(\frac{t-\widehat{m}(x_2)}{\widehat{\sigma}(x_2)} \le s\right)\right]h_e^u(s|x_1)ds,$$

we can write

$$\begin{aligned}\widehat{\Lambda}_1(y) &= \sum_{i=1}^n \int_{B_y}\int_{-\infty}^y \phi_x''(\overline{H}_e(s|x))\left[H_e(s|x) - I(\widehat{E}_i \le s)\right]h_e^u(s|x)dsW_{ni}(x,h_n)f_X(x)dx \\ &= \sum_{i=1}^n \int g^{[1]}_{y,\widehat{m},\widehat{\sigma}}(x,X_i,T_i)W_{ni}(x,h_n)f_X(x)dx.\end{aligned}$$

Now, thanks to Lemma A.5 we have

$$\begin{aligned}\widehat{\Lambda}_1(y) &= \frac{1}{n}\sum_{i=1}^n g^{[1]}_{y,\widehat{m},\widehat{\sigma}}(X_i,X_i,T_i) + o_P(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^n I(X_i\in B_y)\int_{-\infty}^y \phi_{X_i}''(\overline{H}_e(s|X_i))\left[H_e(s|X_i) - I(\widehat{E}_i\le s)\right]h_e^u(s|X_i)ds + o_P(n^{-1/2}).\end{aligned}$$

By adding and subtracting terms we get,

$$\begin{aligned}\widehat{\Lambda}_1(y) &= \frac{1}{n}\sum_{i=1}^n I(X_i\in B_y)\int_{-\infty}^y \phi_{X_i}''(\overline{H}_e(s|X_i))\left[I(E_i\le s) - I(\widehat{E}_i\le s)\right]h_e^u(s|X_i)ds \\ &\quad + \frac{1}{n}\sum_{i=1}^n I(X_i\in B_y)\int_{-\infty}^y \phi_{X_i}''(\overline{H}_e(s|X_i))\left[H_e(s|X_i) - I(E_i\le s)\right]h_e^u(s|X_i)ds + o_P(n^{-1/2}) \\ &= \widehat{\Lambda}_{11}(y) + \widehat{\Lambda}_{12}(y) + o_P(n^{-1/2}). \qquad \text{(A.2)}\end{aligned}$$

The second term on the right hand side is of the desired form, and the corresponding class of functions $\{(x,t,\delta)\mapsto I(x\in B_y)\int_{-\infty}^y \phi_x''(\overline{H}_e(s|x))[H_e(s|x) - I(\frac{t-m(x)}{\sigma x} \le s)]h_e^u(s|x)ds : -\infty < y \le \tau_\eta\}$ is a Donsker class by Lemma A.4. To deal with the term $\widehat{\Lambda}_{11}(y)$, we define the class of functions

$$\mathcal{G}^{[2]} := \left\{(x,t)\mapsto g^{[2]}_{y,m_1,\sigma_1}(x,t) :\; y\le\tau_\eta,\; m_1\in C_M^{1+\delta},\; \sigma_1\in\widetilde{C}_M^{1+\delta}\right\},$$

where $g^{[2]}_{y,m_1,\sigma_1}(x,t) = I(x \in B_y) \int_{-\infty}^{y} \phi''_x(\overline{H}_e(s|x))[I(\frac{t-m(x)}{\sigma(x)} \leq s) - I(\frac{t-m_1(x)}{\sigma_1(x)} \leq s)]$ $\times h^u_e(s|x)ds$. The class of functions $\mathcal{G}^{[2]}$ is Donsker by Lemma A.4. Therefore, by Corollary 2.3.12 in the book of Van der Vaart and Wellner (1996) we have

$$\lim_{\alpha \downarrow 0} \limsup_{n\to\infty} P\left(\sup_{g^{[2]}\in\mathcal{G}^{[2]},\ Var(g^{[2]})<\alpha} n^{-1/2} \left| \sum_{i=1}^{n} g^{[2]}(X_i,T_i) - E[g^{[2]}(X,T)] \right| > \varepsilon \right) = 0,$$

for every $\varepsilon > 0$. Since it follows from Propositions 3.5, 3.6 and 3.7 in Sujica and Van Keilegom (2013) that $\lim_{n\to\infty} P(g^{[2]}_{y,\widehat{m},\widehat{\sigma}} \in \mathcal{G}^{[2]}) = 1$, and since it can be easily verified that $Var(g^{[2]}_{y,\widehat{m},\widehat{\sigma}}(X,T)) = o(1)$ a.s., we can approximate $g^{[2]}_{y,\widehat{m},\widehat{\sigma}}$ in $\widehat{\Lambda}_{11}(y)$ by its expectation:

$$\widehat{\Lambda}_{11}(y) = E\left\{ I(X \in B_y) \int_{-\infty}^{y} \phi''_X(\overline{H}_e(s|X)) \left[I(E \leq s) - I(\widehat{E} \leq s) \right] h^u_e(s|X)ds \right\} + o_P\left(n^{-1/2}\right).$$

Further, we calculate

$$\begin{aligned}
&\widehat{\Lambda}_{11}(y) - o_P\left(n^{-1/2}\right) \\
&= \int_{-\infty}^{y} \int_{B_y} \phi''_x(\overline{H}_e(s|x)) h^u_e(s|x) \left[P\left(E \leq s|X=x\right) - P\left(\widehat{E} \leq s|\mathcal{X}_n, X=x\right) \right] dF_X(x)ds \\
&= \int_{-\infty}^{y} \int_{B_y} \phi''_x(\overline{H}_e(s|x)) h^u_e(s|x) \left[P\left(E \leq s|\, X=x\right) - P\left(\frac{T-\widehat{m}(x)}{\widehat{\sigma}(x)} \leq s \middle| X=x \right) \right] dF_X(x)ds \\
&= \int_{-\infty}^{y} \int_{B_y} \phi''_x(\overline{H}_e(s|x)) h^u_e(s|x) \left[H_e\left(s|\, x\right) - H_e\left(\frac{s\widehat{\sigma}(x) + \widehat{m}(x) - m(x)}{\sigma(x)} \middle| x \right) \right] dF_X(x)ds \\
&= -\int_{-\infty}^{y} \int_{B_y} \phi''_x(\overline{H}_e(s|x)) h^u_e(s|x) \frac{h_e(s|x)}{\sigma(x)} \left\{ s[\widehat{\sigma}(x) - \sigma(x)] + \widehat{m}(x) - m(x) \right\} dF_X(x)ds \\
&\quad - \frac{1}{2} \int_{-\infty}^{y} \int_{B_y} \phi''_x(\overline{H}_e(s|x)) h^u_e(s|x) \frac{h'_e(s_1|x)}{\sigma(x)^2} \left\{ s[\widehat{\sigma}(x) - \sigma(x)] + \widehat{m}(x) - m(x) \right\}^2 dF_X(x)ds,
\end{aligned}$$

where s_1 is between s and $\sigma(x)^{-1}[s\widehat{\sigma}(x) + \widehat{m}(x) - m(x)]$, and $\mathcal{X}_n = \{(T_i, \Delta_i, X_i) :\ i = 1, ..., n\}$. Because of assumption (A5)(iii) we have that $\sup_{y,x} |y^2 h'_e(y|x)| < \infty$, so the second term above is of order $O((nh_n)^{-1} \log n)$ a.s. by Proposition 3.5 in Sujica and Van Keilegom (2013). By using the asymptotic representation for $\widehat{m}$ and $\widehat{\sigma}$ given in Proposition 3.8 in Sujica and Van Keilegom (2013), we get

$$\begin{aligned}
&\widehat{\Lambda}_{11}(y) - o_P\left(n^{-1/2}\right) \\
&= -\frac{1}{n} \sum_{i=1}^{n} \int_{B_y} \int_{-\infty}^{y} \phi''_x(\overline{H}_e(s|x)) \frac{h_e(s|x)}{\sigma(x)} \left[s\zeta(T_i, \Delta_i|x) + \eta(T_i, \Delta_i|x) \right] \frac{K\left(\frac{x-X_i}{h_n}\right)}{h_n} dH^u_e(s|x)dx \\
&:= \frac{1}{n} \sum_{i=1}^{n} \int g^{[3]}_y(X_i + uh_n, T_i, \Delta_i) K\left(u\right) du.
\end{aligned}$$

We can write

$$g_y^{[3]}(x,t,\delta) = -I(x \in B_y) \int_{-\infty}^{y} q_s(x,t,\delta) dH_e^u(s|x),$$

where $q_s(x,t,\delta) = \phi_x''(\overline{H}_e(s|x)) \frac{h_e(s|x)}{\sigma(x)} [s\zeta(t,\delta|x) + \eta(t,\delta|x)]$. By assumptions (A4) and (A5)(ii,iii) we have that for every $s \leq y \leq \tau_\eta$ the function $x \mapsto \phi_x''(\overline{H}_e(s|x)) \frac{h_e(s|x)}{\sigma(x)} s$ is uniformly bounded (by a constant not depending on s and y) in $x \in B_y$, as well as the corresponding first derivative. Therefore, using the notation $\mathcal{Q} = \{x \mapsto \phi_x''(\overline{H}_e(s|x)) \frac{h_e(s|x)}{\sigma(x)} s : s \leq \tau_\eta\}$, we have that the bracketing number $N_{[\,]}(\varepsilon, \mathcal{Q}, L_2(P))$ equals $\exp(-C\varepsilon)$, for some constant C (see Corollary 2.7.2 in Van der Vaart and Wellner (1996)). By a similar reasoning we have that the bracketing number of the class $\mathcal{Q}' = \{x \mapsto \phi_x''(\overline{H}_e(s|x)) \frac{h_e(s|x)}{\sigma(x)} : s \leq \tau_\eta\}$ equals $\exp(-C\varepsilon)$. Now, $\mathcal{Q}$ and $\mathcal{Q}'$ are Donsker by Theorem 2.5.6 in Van der Vaart and Wellner (1996). Hence, Lemma A.4 entails that $\mathcal{G}^{[3]} = \{(x,t,\delta) \mapsto g_y^{[3]}(x,t,\delta) : -\infty < y \leq \tau_\eta\}$ is a Donsker class of functions, while Lemma A.5 entails that

$$\widehat{\Lambda}_{11}(y) = \frac{1}{n} \sum_{i=1}^{n} g_y^{[3]}(X_i, T_i, \Delta_i) + o_P\left(n^{-1/2}\right).$$

Now, by using (A.2) we get $\widehat{\Lambda}_1(y) = \frac{1}{n}\sum_{i=1}^n g_y^{(1)}(X_i,T_i,\Delta_i) + o_P\left(n^{-1/2}\right)$. The class of functions $\mathcal{G}^{(1)}$ is a Donsker class as it is a sum of two Donsker classes (see Example 2.10.7 in Van der Vaart and Wellner (1996)), which completes the proof. □

Lemma A.3. *[Asymptotic representation for $\widehat{\Lambda}_2(y)$, $\widehat{\Lambda}_3(y)$ and $\widehat{\Lambda}_4(y)$]*
Assume (A1)-(A4), and assume that (A5) holds for $H(y|x)$ and $H^u(y|x)$. Then for $y \leq \tau_\eta$,

$$\begin{aligned}
\widehat{\Lambda}_2(y) &= \frac{1}{n} \sum_{i=1}^{n} g_y^{(2)}(X_i, T_i, \Delta_i) + R_n(y), \\
\widehat{\Lambda}_3(y) &= \frac{1}{n} \sum_{i=1}^{n} g_y^{(3)}(X_i, T_i, \Delta_i), \\
\widehat{\Lambda}_4(y) &= \frac{1}{n} \sum_{i=1}^{n} g_y^{(4)}(X_i, T_i, \Delta_i),
\end{aligned}$$

where

$$\begin{aligned}
g_y^{(2)}(X,T,\Delta) &= I(X \in B_y) \int_{-\infty}^{y} \phi_X'(\overline{H}_e(s|X)) d\Bigg[\frac{h_e^u(s|X)}{\sigma(X)} [s\zeta(T,\Delta|X) + \eta(T,\Delta|X)] \\
&\qquad + I(E \leq s, \Delta = 1) - H_e^u(s|X)\Bigg], \\
g_y^{(3)}(X,T,\Delta) &= I(X \in B_y) \int_{-\infty}^{y} \phi_X'(\overline{H}_e(s|X)) dH_e^u(s|X) \\
&\qquad - \int_{B_y} \int_{-\infty}^{y} \phi_x'(\overline{H}_e(s|x)) dH_e^u(s|x) dF_X(x),
\end{aligned}$$

$$g_y^{(4)}(X,T,\Delta) \; = I(X \in B_y)\phi_X(\bar{F}_e(y)) - \int_{B_y} \phi_x(\bar{F}_e(y))dF_X(x),$$

$\sup_{y\le\tau_\eta} |R_n(y)| = o_P\left(n^{-1/2}\right)$, *and the terms* $\widehat{\Lambda}_2(y)$, $\widehat{\Lambda}_3(y)$ *and* $\widehat{\Lambda}_4(y)$ *are defined in (4.6). Furthermore, the class of functions* $\mathcal{G}^{(i)} = \{(x,t,\delta) \mapsto g_y^{(i)}(x,t,\delta) : -\infty < y \le \tau_\eta\}$ *is a Donsker class of functions for every* $i = 2,3,4$.

Proof. The expressions of $\widehat{\Lambda}_3(y)$ and $\widehat{\Lambda}_4(y)$ are obtained from simple algebra. Showing the Donsker property of $\mathcal{G}^{(3)}$ and $\mathcal{G}^{(4)}$ is very similar, and therefore we will only show it for $\mathcal{G}^{(3)}$. The first factor of the first term in $g_y^{(3)}(x,t,\delta)$ belongs to a class of functions that is Donsker by construction. The second factor of the first term in $g_y^{(3)}(x,t,\delta)$ is bounded and monotone in the parameter y, uniformly in $y \le \tau_\eta$. Therefore, its ε-bracketing number is bounded by $C\varepsilon^{-1}$, which by Theorem 2.5.6 in Van der Vaart and Wellner (1996) implies the Donsker property. The second term of $g_y^{(3)}(x,t,\delta)$ is a deterministic, uniformly bounded function in $y \le \tau_\eta$ by (A4), and therefore trivially Donsker. Since multiplying and adding uniformly bounded functions preserves the Donsker property (by Examples 2.10.7 and 2.10.8 in Van der Vaart and Wellner (1996)), the class of functions $\mathcal{G}^{(3)}$ is also Donsker. To deal with the expression of $\widehat{\Lambda}_2(y)$ we start by using integration by parts:

$$\begin{aligned}\widehat{\Lambda}_2(y) \;&=\; \int_{B_y} \phi_x'(\overline{H}_e(y|x)) \left[\widehat{H}_e^u(y|x) - H_e^u(y|x)\right] dF_X(x) \\ &\quad + \int_{B_y}\int_{-\infty}^{y} \phi_x''(\overline{H}_e(s|x))h_e(s|x) \left[\widehat{H}_e^u(s|x) - H_e^u(s|x)\right] ds dF_X(x).\end{aligned}$$

From here on the proof is very similar to the proof of Lemma A.2 and is therefore omitted. □

The following lemma shows the Donsker property for the classes of functions that are showing up in the proofs of Lemmas A.2, A.3 and A.5.

Lemma A.4. *Assume (A1)-(A4), and assume that (A5) holds for* $H(y|x)$ *and* $H^u(y|x)$. *Then, the following classes of functions are Donsker:*

$$\mathcal{G}_1 = \left\{(x_1,x_2,t) \mapsto g_{y,m_1,\sigma_1,1}(x_1,x_2,t) : y \le \tau_\eta,\; m_1 \in C_M^{1+\delta}(R_X),\; \sigma_1 \in \widetilde{C}_M^{1+\delta}(R_X)\right\},$$

$$\mathcal{G}_2 = \left\{(x_1,x_2,t) \mapsto g_{y,m_1,\sigma_1,2}(x_1,x_2,t) : y \le \tau_\eta,\; m_1 \in C_M^{1+\delta}(R_X),\; \sigma_1 \in \widetilde{C}_M^{1+\delta}(R_X)\right\},$$

$$\mathcal{G}_3 = \left\{(x,t,\delta) \mapsto g_{y,\psi,3}(x,t,\delta) : y \le \tau_\eta, \psi \in \{\zeta,\eta\}\right\},$$

$$\mathcal{G}_4 = \left\{(x,t,\delta) \mapsto g_{y,\psi,4}(x,t,\delta) : y \le \tau_\eta, \psi \in \{\zeta,\eta\}\right\},$$

where

$$g_{y,m_1,\sigma_1,1}(x_1,x_2,t) = I(x_1 \in B_y)\int_{-\infty}^{y} q_s(x_1)\left[I\left(\frac{t-m_1(x_2)}{\sigma_1(x_2)} \le s\right) - H_e(s|x_1)\right]h_e^u(s|x_1)ds,$$

$$g_{y,m_1,\sigma_1,2}(x_1,x_2,t) = I(x_1 \in B_y)q_y(x_1)\left[I\left(\frac{t-m_1(x_2)}{\sigma_1(x_2)} \le y\right) - H_e(y|x_1)\right]h_e^u(y|x_1),$$

$$g_{y,\psi,3}(x,t,\delta) = I(x \in B_y)\int_{-\infty}^{y} q_s(x)\psi(t,\delta|x)h_e^u(s|x)ds,$$

$$g_{y,\psi,4}(x,t,\delta) = I(x \in B_y)q_y(x)\psi(t,\delta|x)h_e^u(y|x),$$

the sets $C_M^{1+\delta}(R_X)$ *and* $\widetilde{C}_M^{1+\delta}(R_X)$ *are defined in (A.1), and the class* $\{x \mapsto q_y(x) : y \le \tau_\eta\}$ *is a Donsker class of non-negative (non-positive) functions, such that* $\sup_{y\le\tau_\eta}\sup_{x\in B_y}\{|q_y(x)|, |\frac{\partial}{\partial x}q_y(x)|\} < \infty$.

Proof. First, we will prove that the class of functions $\mathcal{G}_2$ is Donsker. Note that the class of functions $\{(x_1,x_2,t) \mapsto [I(\frac{t-m_1(x_2)}{\sigma_1(x_2)} \le y) - H_e(y|x_1)] : \ y \le \tau_\eta, \ \ m_1 \in C_M^{1+\delta}(R_X), \ \sigma_1 \in \widetilde{C}_M^{1+\delta}(R_X)\}$ is Donsker by calculations in Lemma A.1 in Van Keilegom and Akritas (1999). The class of functions $\{x \mapsto I_B(x) : B \in \mathcal{B}\}$ is Donsker (see Section 2). Hence, the class of functions $\mathcal{G}_2$ is Donsker since it is a product of uniformly bounded, Donsker classes of functions (see Example 2.10.8 in the book of Van der Vaart and Wellner (1996)). Since it is easy to see that the functions ζ and η are uniformly bounded, $\mathcal{G}_4$ is a product of uniformly bounded, Donsker classes of functions, and therefore Donsker itself (see Example 2.10.8 in Van der Vaart and Wellner (1996)).

Proving that $\mathcal{G}_1$ and $\mathcal{G}_3$ are Donsker is similar, therefore we will only prove that $\mathcal{G}_1$ is Donsker, which is the hardest of the two. We will use results from the book of Van der Vaart and Wellner (1996). By Theorem 2.5.6 in their book it is sufficient to show that

$$\int_0^\infty \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}_1, \|\cdot\|_\infty)}d\varepsilon < \infty. \tag{A.3}$$

We will restrict ourselves to showing (A.3) for $\mathcal{G}_1^1 = \{(x_1,x_2,t) \mapsto g^1_{y,m_1,\sigma_1}(x_1,x_2,t) : \ y \le \tau_\eta, \ m_1 \in C_M^{1+\delta}(R_X), \ \sigma_1 \in \widetilde{C}_M^{1+\delta}(R_X)\}$, where $g^1_{y,m_1,\sigma_1}(x_1,x_2,t) = \int_{-\infty}^{y} q_s(x_1)I(\frac{t-m_1(x_2)}{\sigma_1(x_2)} \le s)h_e^u(s|x_1)ds$. By Theorem 2.7.1 in the aforementioned book we know that $C_M^{1+\delta}(R_X)$ and $\widetilde{C}_M^{1+\delta}(R_X)$ can be covered by $M_1 = \exp(C_1\varepsilon^{-1/(1+\delta)})$ and $M_2 = \exp(C_2\varepsilon^{-1/(1+\delta)})$ ε-brackets with respect to the supremum norm, respectively. Let $\{[m_j^l, m_j^r] : j = 1,...,M_1\}$ and $\{[\sigma_k^l, \sigma_k^r] : k = 1,...,M_2\}$ be those ε-brackets. Let $x_1,...,x_{M_3}$ be a grid of R_X such that $x_{r+1} - x_r \le \varepsilon$, $r = 1,...,M_3 - 1$ and let $\{y_{ri} : r = 1,...,M_3, i = 1,...,M_4\}$ be such that $H_e^u(y_{ri+1}|x_j) - H_e^u(y_{ri}|x_j) \le \varepsilon$. Let $\{y_i : i = 1,...,M_5\}$ be the union of all y_{ri} in ascending order. There are $M_5 = C\varepsilon^{-2}$ of them. Now, we define brackets $\{[g_{y_i,m_j^l,\sigma_k^l}, g_{y_{i+1},m_j^r,\sigma_k^r}]\}_{i,j,k}$. There are at most $C\varepsilon^{-2}\exp\{C\varepsilon^{-1/(1+\delta)}\}$ of them. Hence, they satisfy condition (A.3) and they cover $\mathcal{G}_1^1$, because $g^1_{t,m,\sigma}(\cdot,\cdot)$ is a monotone function of its parameters. Now, to show that $\mathcal{G}_1^1$ is Donsker, we only need to show that the defined brackets are ε-brackets:

$$\begin{aligned}
&\|g_{y_i,m_j^l,\sigma_k^l} - g_{y_{i+1},m_j^r,\sigma_k^r}\|_\infty \\
&\le \|g_{y_i,m_j^l,\sigma_k^l} - g_{y_i,m_j^l,\sigma_k^r}\|_\infty + \|g_{y_i,m_j^l,\sigma_k^r} - g_{y_i,m_j^r,\sigma_k^r}\|_\infty + \|g_{y_i,m_j^r,\sigma_k^r} - g_{y_{i+1},m_j^r,\sigma_k^r}\|_\infty.
\end{aligned} \tag{A.4}$$

We start with the second term:

$$\begin{aligned}
&\|g_{y_i,m_j^l,\sigma_k^r} - g_{y_i,m_j^r,\sigma_k^r}\|_\infty \\
&\le C \sup_{t\le\tau_\eta,x_1,x_2\in R_X} \int_{-\infty}^{\infty} \left| I\left(\frac{t-m_j^l(x_2)}{\sigma_k^r(x_2)} \le s\right) - I\left(\frac{t-m_j^r(x_2)}{\sigma_k^r(x_2)} \le s\right)\right| h_e^u(s|x_1)ds \\
&\le C \sup_{t\le\tau_\eta,x_1,x_2\in R_X} \left| H_e^u\left(\frac{t-m_j^r(x_2)}{\sigma_k^r(x_2)}\middle| x_1\right) - H_e^u\left(\frac{t-m_j^l(x_2)}{\sigma_k^r(x_2)}\middle| x_1\right)\right| \\
&\le \|h_e^u\|_\infty \sup_{x_2\in R_X} \left|\frac{m_j^l(x_2)-m_j^r(x_2)}{\sigma_k^r(x_2)}\right| \\
&\le C\ \|m_j^l - m_j^r\|_\infty \le C\varepsilon.
\end{aligned}$$

The fourth inequality follows from (A5)(ii) for H_e^u, and the fact that σ_k is bounded away from zero as a function belonging to $\widetilde{C}_M^{1+\delta}(R_X)$. Similarly, we can bound the first term:

$$\begin{aligned}
\|g_{y_i,m_j^l,\sigma_k^l} - g_{y_i,m_j^l,\sigma_k^r}\|_\infty &\le \|h_e^u\|_\infty \sup_{y\le\tau_\eta,x_2\in R_X} \left|\frac{(y-m_j^l(x_2))(\sigma_k^l(x_2)-\sigma_k^r(x_2))}{\sigma_k^l(x_2)\sigma_k^r(x_2)}\right| \\
&\le C\ \|\sigma_k^l - \sigma_k^r\|_\infty \le C\varepsilon.
\end{aligned}$$

The second inequality follows from (A5)(ii) for H_e^u, and the fact that σ_k and σ_l are bounded away from zero as functions belonging to $\widetilde{C}_M^{1+\delta}(R_X)$. For the third term we calculate

$$\begin{aligned}
\|g_{y_i,m_j^r,\sigma_k^r} - g_{y_{i+1},m_j^r,\sigma_k^r}\|_\infty &\le C \sup_{x\in R_X} \int_{y_i}^{y_{i+1}} h_e^u(s|x)ds \\
&\le C \sup_{x\in R_X} |H_e^u(y_{i+1}|x) - H_e^u(y_i|x)| \\
&\le C \max_j |H_e^u(y_{i+1}|x_j) - H_e^u(y_i|x_j)| + 2\sup_{y,x}|h_e^u(y|x)|\varepsilon \\
&\le C\varepsilon.
\end{aligned}$$

To get the last inequality we used the uniform boundedness of h_e^u. Now, we have that expression (A.4) is bounded by $C\varepsilon$. Because C is independent of the bracket selection (i.e. of ε), we have that the brackets $\{[g_{y_i,m_j^l,\sigma_k^l}, g_{y_{i+1},m_j^r,\sigma_k^r}]\}_{i,j,k}$ are ε-brackets, therefore by the previous reasoning $\mathcal{G}_1^1$ (and hence $\mathcal{G}_1$) is Donsker. □

The next lemma is an adaptation of the result on convergence of smoothed processes in Giné and Nickle (2008) to our framework. We will use this result in the proofs of Lemmas A.2 and A.3 to asymptotically replace expressions of the form $\sum_{i=1}^n \int \widehat{g}_y(x, V_i) W_{ni}(x, h_n) dF_X(x)$ by the simpler expressions $\frac{1}{n}\sum_{i=1}^n \widehat{g}_y(X_i, V_i)$, where $V_i = (X_i, T_i)$.

Lemma A.5. *Assume the conditions and notations of Lemma A.4. Then, using the notation $\widehat{g}_{y,r} := g_{y,\widehat{m},\widehat{\sigma},r}$, for r=1,2, and $\widehat{g}_{y,r} := g_{y,\psi,r}$, for $r = 3, 4$, we have*

$$\sup_{y\le\tau_\eta}\left|\frac{1}{n}\sum_{i=1}^n\left\{\int \widehat{g}_{y,r}(X_i+uh_n, V_i)K(u)\,du - \widehat{g}_{y,r}(X_i, V_i)\right\}\right| = o_P\left(n^{-1/2}\right), \tag{A.5}$$

$$\sup_{y\le\tau_\eta}\left|\sum_{i=1}^n\left\{\int \widehat{g}_{y,r}(t, V_i)W_{ni}(t, h_n)f_X(t)dt - \frac{1}{n}\widehat{g}_{y,r}(X_i, V_i)\right\}\right| = o_P\left(n^{-1/2}\right). \tag{A.6}$$

***Proof*.** Before we start, we give the results that will be used throughout the proof and which will be shown at the end of this proof. For $r = 1, 2, 3, 4$ we have

$$\sup_{g\in\mathcal{G}_r}\|g\|_\infty < \infty, \tag{A.7}$$

$$\lim_{n\to\infty} P\left(\{\widehat{g}_{y,r} : -\infty < y \le \tau_\eta\} \subset \mathcal{G}_r\right) = 1, \tag{A.8}$$

$$\sup_{g\in\mathcal{G}_r} E\left[\int g(X+uh_n, V)K(u)\,du - g(X, V)\right]^2 = o(1), \tag{A.9}$$

$$\sup_{y\le\tau_\eta}\left|E\left[\int \widehat{g}_{y,r}(X+uh_n, V)K(u)\,du - \widehat{g}_{y,r}(X, V)\middle|\mathcal{X}_n\right]\right| = o(n^{-1/2})\ a.s., \tag{A.10}$$

where $V = (X, T)$ and $\mathcal{X}_n = \{(X_i, T_i, \Delta_i) : i = 1, ..., n\}$. The aforementioned results also hold when we replace $\widehat{g}_{y,r}$ and $\mathcal{G}_r$ by $\widehat{g}'_{y,r} = \widehat{g}_{y,r}\frac{f_X - \widehat{f}_X}{\widehat{f}_X}$ and $\mathcal{G}'_r := \mathcal{G}_r \times C_M^{1+\delta}(R_X)$, respectively, where $\widehat{f}_X(x) = \frac{1}{nh_n}\sum_{i=1}^n K(\frac{X_i - x}{h_n})$.

In the following calculations we will use the notation $\widehat{g}_y$ to represent $\widehat{g}_{y,r}$, for all $r = 1, 2, 3, 4$, since the proof is the same in all cases. Similarly, for all $r = 1, 2, 3, 4.$, we will use the notation g for $g_r \in \mathcal{G}'_r$, and $\mathcal{G}'$ for $\mathcal{G}'_r$. Conditions (A.7) to (A.10) allow us to apply Theorem 2 (a) in Giné and Nickle (2008) to the term in (A.5), which gives

$$\begin{aligned}&\sup_{y\le\tau_\eta}\left|\frac{1}{n}\sum_{i=1}^n\left\{\int \widehat{g}_y(t, V_i)\frac{1}{h_n}K\left(\frac{X_i - t}{h_n}\right)dt - \widehat{g}_y(X_i, V_i)\right\}\right| \\ &= \sup_{y\le\tau_\eta}\left|E\left[\int \widehat{g}_y(t, V)\frac{1}{h_n}K\left(\frac{X - t}{h_n}\right)dt - \widehat{g}_y(X, V)\middle|\mathcal{X}_n\right]\right| + o_P\left(n^{-1/2}\right).\end{aligned}$$

The first term on the right hand side is $o(n^{-1/2})$ a.s. by (A.10), which implies (A.5). Next, to prove statement (A.6), we write

$$
\begin{aligned}
Q(\widehat{g}_y) &:= \sum_{i=1}^n \int \widehat{g}_y(t,V_i) W_{ni}(t,h_n) f_X(t) dt \\
&= \frac{1}{n}\sum_{i=1}^n \int \widehat{g}_y(t,V_i) \frac{f_X(t)}{\widehat{f}_X(t)} \frac{1}{h_n} K\left(\frac{X_i-t}{h_n}\right) dt \\
&= \frac{1}{n}\sum_{i=1}^n \int \widehat{g}_y(t,V_i) \frac{1}{h_n} K\left(\frac{X_i-t}{h_n}\right) dt \\
&+ \frac{1}{n}\sum_{i=1}^n \int \widehat{g}_y(t,V_i) \frac{f_X(t)-\widehat{f}_X(t)}{\widehat{f}_X(t)} \frac{1}{h_n} K\left(\frac{X_i-t}{h_n}\right) dt + o(n^{-1/2}) \ a.s. \\
&:= Q_1(\widehat{g}_y) + Q_2(\widehat{g}_y) + o(n^{-1/2}) \ a.s.
\end{aligned}
$$

Because of conditions (A.7)-(A.10) for $\widehat{g}'_y$ and $\mathcal{G}'$, we can use (A.5) to write $Q_2(\widehat{g}_y)$ as

$$
Q_2(\widehat{g}_y) \ = \ \frac{1}{n}\sum_{i=1}^n \widehat{g}_y(X_i,V_i)\frac{f_X(X_i)-\widehat{f}_X(X_i)}{\widehat{f}_X(X_i)} + o_P\left(n^{-1/2}\right).
$$

By Corollary 2.3.12 in the book of Van der Vaart and Wellner (1996) we have

$$
\lim_{\alpha\downarrow 0}\limsup_{n\to\infty} P\left(\sup_{g'\in\mathcal{G}',\ Var(g')<\alpha} n^{-1/2}\left|\sum_{i=1}^n g'(X_i,V_i) - E\left(g'(X,V)\right)\right| > \varepsilon\right) = 0, \qquad \text{(A.11)}
$$

for every $\varepsilon > 0$. Since, $\lim_{n\to\infty} P(\widehat{g}_y \frac{f_X-\widehat{f}_X}{\widehat{f}_X} \in \mathcal{G}') = 1$ and $Var(\widehat{g}_y \frac{f_X-\widehat{f}_X}{\widehat{f}_X}) = o(1)$, we can use (A.11) to approximate $Q_2(\widehat{g}_y)$ with the corresponding expectation:

$$
\begin{aligned}
Q_2(\widehat{g}_y) \ &= \ E\left[\widehat{g}_y(X,V)\frac{f_X(X)-\widehat{f}_X(X)}{\widehat{f}_X(X)}\middle|\mathcal{X}_n\right] + o_P\left(n^{-1/2}\right) \\
&= \ E\left[E[\widehat{g}_y(X,V)|X,\mathcal{X}_n]\frac{f_X(X)-\widehat{f}_X(X)}{\widehat{f}_X(X)}\middle|\mathcal{X}_n\right] + o_P\left(n^{-1/2}\right) \\
&= \ o_P\left(n^{-1/2}\right).
\end{aligned}
$$

To obtain the last equality we used that $\|f_X - \widehat{f}_X\|_\infty = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. and $\sup_{x\in R_X}|E[\widehat{g}_y(X,V)|X=x,\mathcal{X}_n]| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. (see Proposition 3.5 in Sujica and Van Keilegom (2013)). Now, applying (A.5) on $Q_1(\widehat{g}_y)$ we get

$$
Q(\widehat{g}_y) = \frac{1}{n}\sum_{i=1}^n \widehat{g}_y(X_i,V_i) + o_P\left(n^{-1/2}\right),
$$

which concludes the proof.

Now, it remains to show that conditions (A.7), (A.8), (A.9) and (A.10) are satisfied (proving the corresponding conditions for $\widehat{g}'_y$ and $\mathcal{G}'$ will be omitted since it uses the same techniques and reasoning). Condition (A.7) is easily verified, while condition (A.8) follows from Propositions 3.5, 3.6 and 3.7 in Sujica and Van Keilegom (2013). Because of condition (A.7) we can bound the expectation in condition (A.9) by

$$C \sup_{u\in[-a,a]} E|g(X+uh_n,V)-g(X,V)|.$$

Now, we decompose the function g as $g(X,V) = I(X \in B_y)z(X,V)$ and write

$$\begin{aligned} Q_g(uh_n) &= E\big|E[g(X+uh_n,V)|X] - E[g(X,V)|X]\big| \\ &= E\big|I(X+uh_n \in B_y)E[z(X+uh_n,V)|X] - I(X \in B_y)E[z(X,V)|X]\big|. \end{aligned}$$

By using assumptions (A4) and (A5)(iv,vi), it is easy to see that the function $(x,e) \mapsto E[z(x+e,V)|X=x]$ has uniformly bounded partial derivative over e, for e small enough. This allows us to use a Taylor expansion and to obtain

$$\begin{aligned} Q_g(uh_n) &= E\left\{|I(X+uh_n \in B_y) - I(X \in B_y)|\,|E[z(X,V)|X]|\right\} + O(uh_n) \\ &\le C \int_{R_X} |I(x+uh_n \in B_y) - I(x \in B_y)|dx + O(uh_n) \\ &= C \int_{R_X} I\left(x \in B_y \Delta \{B_y - uh_n\}\right) dx + O(uh_n) \\ &\le C \sum_{i=1}^{k} \int_{\sup B_{yi}-uh_n}^{\sup B_{yi}+uh_n} dx + C \sum_{i=1}^{k} \int_{\inf B_{yi}-uh_n}^{\inf B_{yi}+uh_n} dx + O(uh_n) \\ &= O(uh_n), \end{aligned}$$

where Δ is the symmetric difference, $B_y = \bigcup_{i=1}^{k} B_{yi}$, the sets B_{yi} are convex and $k \le \lambda(R_X)/\beta < \infty$ (for the definition of β see Section 2). This implies condition (A.9). To prove condition (A.10), we use the notation $E[\,\cdot\,] = E[\,\cdot\,|\mathcal{X}_n]$ and write the expectation in condition (A.10) as

$$\int E\Big[\widehat{g}_y(X+uh_n,V) - \widehat{g}_y(X,V)\Big] K(u)\, du.$$

Using the decomposition $\widehat{g}_y(x,v) = I(x \in B_y)\widehat{z}_y(x,v)$, we can write the expectation in the integral above as

$$\begin{aligned} Q_{\widehat{g}_y}(uh_n) =& E\left\{I(X+uh_n \in B_y)\Big[E[\widehat{z}_y(X+uh_n,V)|X] - E[\widehat{z}_y(X,V)|X]\Big]\right\} \\ &+ E\left\{\Big[I(X+uh_n \in B_y) - I(X \in B_y)\Big]E[\widehat{z}_y(X,V)|X]\right\} \\ :=& Q_{\widehat{g}_y,1}(uh_n) + Q_{\widehat{g}_y,2}(uh_n). \end{aligned}$$

In order to bound the second term on the right hand side, we use similar calculations as when dealing with $Q_g(uh_n)$ to get

$$\begin{aligned}\left|Q_{\widehat{g}_y,2}(uh_n)\right| \leq & C\sum_{i=1}^{k}\int_{\sup B_{yi}-uh_n}^{\sup B_{yi}+uh_n} |E[\widehat{z}_y(x,V)|X=x]|dx \\ & + C\sum_{i=1}^{k}\int_{\inf B_{yi}-uh_n}^{\inf B_{yi}+uh_n} |E[\widehat{z}_y(x,V)|X=x]|dx \\ \leq & 2Ck \sup_{y\leq\tau_\eta}\sup_{d(x,B_y)\leq uh_n} |E[\widehat{z}_y(x,V)|X=x]|uh_n \\ = & O\left((nh_n)^{-1/2}(\log n)^{1/2}h_n\right) = o(n^{-1/2}) \ a.s.,\end{aligned}$$

uniformly in $y \leq \tau_\delta$ and $u \in [-a,a]$, where $d(x,B_y) = \inf\{|x-x_1| : x_1 \in B_y\}$. To get the first equality we used Proposition 3.5 in Sujica and Van Keilegom (2013), and assumptions (A4) and (A5)(iv,vi). The term $Q_{\widehat{g}_y,1}(uh_n)$ can be bounded as follows:

$$\begin{aligned}\left|Q_{\widehat{g}_y,1}(uh_n)\right| & \leq \sup_{y\leq\tau_\eta}\sup_{d(x,B_y)\leq uh_n}\left|E\left[\frac{\partial}{\partial x}E\left[\widehat{z}_y(x,V)|X=x\right]\right]uh_n\right| \\ & = O\left((nh_n)^{-1/2}(\log n)^{1/2}h_n\right) = o(n^{-1/2}) \ a.s.\end{aligned}$$

uniformly in $y \leq \tau_\delta$ and $u \in [-a,a]$. To get the first equality we used again Proposition 3.5 in Sujica and Van Keilegom (2013), and assumptions (A4) and (A5)(iv,vi). This implies (A.10). □

B Appendix B

This section contains results on uniform convergence rates of quantities needed for showing that the remainder terms in (4.6) of Theorem 3.1 are $o(n^{-1/2})$ a.s.

Lemma B.1. *Assume (A1)(i,ii) and (A4). Then, for $\alpha \in (0,1]$ and $0 < M < \infty$,*

$$\sup_{v\in[\alpha,1],\ y\leq\tau_\eta} |\widehat{\phi}_{(y)}(v) - \phi_{(y)}(v)| = O(n^{-1/3}(\log n)^{1/2}) \ a.s., \tag{B.1}$$

$$\sup_{v\in[\alpha,1],\ y\leq\tau_\eta} |\widehat{\phi}'_{(y)}(v) - \phi'_{(y)}(v)| = O(n^{-1/3}(\log n)^{1/2}) \ a.s., \tag{B.2}$$

$$\sup_{v\in[\alpha,1],\ y\leq\tau_\eta} |\widehat{\phi}''_{(y)}(v) - \phi''_{(y)}(v)| = O(n^{-1/3}(\log n)^{1/2}) \ a.s., \tag{B.3}$$

$$\sup_{u\in[0,M],\ y\leq\tau_\eta} |\widehat{\phi}^{-1}_{(y)}(u) - \phi^{-1}_{(y)}(u)| = O(n^{-1/3}(\log n)^{1/2}) \ a.s. \tag{B.4}$$

Proof. Statements (B.1), (B.2) and (B.3) can be proven in an analogous way. Therefore, we will prove only statement (B.3). We can write

$$\widehat{\phi}''_{(y)}(v) - \phi''_{(y)}(v) = \int_{B_y} \phi''_x(v) d[\widehat{F}_X(x) - F_X(x)].$$

Now, since $\phi''_x(v)$ and $\phi'''_x(v)$ are uniformly bounded in $(v, x) \in [\alpha, 1] \times R_X$, we can easily show by following calculations done for proving (B.5) in Lemma B.2 below, that statement (B.3) is true. To prove statement (B.4) we use a first order Taylor expansion to get

$$\widehat{\phi}^{-1}_{(y)}(u) - \phi^{-1}_{(y)}(u) \quad = \quad \frac{-1}{\widehat{\phi}'_{(y)}(\xi(u,y))} \left[\widehat{\phi}_{(y)}\left(\phi^{-1}_{(y)}(u)\right) - \widehat{\phi}_{(y)}\left(\widehat{\phi}^{-1}_{(y)}(u)\right) \right],$$

where $\xi(u, y)$ is between $\widehat{\phi}^{-1}_{(y)}(u)$ and $\phi^{-1}_{(y)}(u)$. The second factor on the right hand side is $O(n^{-1/3}(\log n)^{1/2})$ *a.s.* by (B.1). The first factor on the right hand side is a.s. uniformly bounded (in $u \in [0, M]$ and $y \leq \tau_\eta$) because of assumption (A4)(iii). □

When showing the negligibility of the remainder terms in the proof of Theorem 3.1 via Lemma B.2 below, we need to verify certain assumptions regarding the rate of $G_n(y, x)$ and $\frac{\partial}{\partial x} G_n(y, x)$, where $G_n(y, x)$ is a stochastic process. This will either be trivial or will reduce to verifying that $\sup_{y,x} |\widehat{L}(y|x) - L(y|x)| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s., $\sup_{y,x} |\dot{\widehat{L}}(y|x) - \dot{L}(y|x)| = O(1)$ a.s. or $\sup_{y,x} |\dot{\widehat{L}}(y|x) - \dot{L}(y|x)| = O((nh_n^3)^{-1/2}(\log n)^{1/2})$ a.s., for $L \in \{H_e, H_e^u\}$, which is true by Lemma B.4.

Lemma B.2. *Assume (A1)(i,ii). Let* $G_n(y, x)$ *be a stochastic process that is satisfying* $\sup_{y \leq \tau_\eta} \sup_{x \in B_y} \{|G_n(y, x)|, |\frac{\partial}{\partial x} G_n(y, x)|\} = O(1)$ *a.s. Then,*

$$\sup_{y \leq \tau_\eta} \left| \int_{B_y} G_n(y, x) d\left[\widehat{F}_X(x) - F_X(x)\right] \right| = O(n^{-1/3}(\log n)^{1/2}) \; a.s. \tag{B.5}$$

If in addition we assume that $\sup_{y \leq \tau_\eta} \sup_{x \in B_y} |G_n(y, x)| = O((nh_n)^{-1/2}(\log n)^{1/2})$ *a.s. and* $\sup_{y \leq \tau_\eta} \sup_{x \in B_y} |\frac{\partial}{\partial x} G_n(y, x)| = O((nh_n^3)^{-1/2}(\log n)^{1/2})$ *a.s., then*

$$\sup_{y \leq \tau_\eta} \left| \int_{B_y} G_n(y, x) d\left[\widehat{F}_X(x) - F_X(x)\right] \right| = o(n^{-1/2}) \; a.s. \tag{B.6}$$

Proof. We start by partitioning $R_X = [a, b]$ using a grid $a = x_1 < x_2 < ... < x_m = b$ such that $x_{i+1} - x_i < a_n$, $i = 1, ..., m-1$, where $m = C\, a_n^{-1}$ and a_n is a sequence of constants to be specified further on. We can then write

$$
\begin{aligned}
&\Big| \int_{B_y} G_n(y,x) d\left[\widehat{F}_X(x) - F_X(x)\right] \Big| \\
&= \Big| \sum_{\substack{i=1 \\ x_i \in B_y}}^{m-1} G_n(y,x_i) \int_{B_y \cap [x_i, x_{i+1}]} d\left[\widehat{F}_X(x) - F_X(x)\right] \\
&\quad + \sum_{\substack{i=1 \\ x_i \in B_y}}^{m-1} \int_{B_y \cap [x_i, x_{i+1}]} [G_n(y,x) - G_n(y,x_i)] d\left[\widehat{F}_X(x) - F_X(x)\right] \Big| \\
&\leq C\ a_n^{-1} \sup_{y \leq \tau_\eta} \sup_{x \in B_y} |G_n(y,x)| \times \sup_{|x - x'| < a_n} |\widehat{F}_X(x') - F_X(x') - \widehat{F}_X(x) + F_X(x)| \\
&\quad + 2k \sup_{y \leq \tau_\eta} \sup_{x \in B_y} \left| \frac{\partial}{\partial x} G_n(y,x) \right| a_n,
\end{aligned}
$$

where the last equality holds for n big enough, and $k \leq \lfloor \lambda(R_X)/\beta \rfloor$. Under the assumption that $\frac{\partial}{\partial x} G_n(y,x)$ is a.s. uniformly bounded, by defining $a_n = n^{-1/3}$, the second term on the right hand side is of the order $O(a_n) = O(n^{-1/3})$ a.s. By using Theorem 0.2 in Stute (1982) we can bound the first term by $a_n^{-1} O(n^{-1/2} a_n^{1/2} (\log n)^{1/2}) = O(n^{-1/3} (\log n)^{1/2})$ a.s. This proves (B.5).

By taking now $a_n = n^{-1/2}$, statement (B.6) is true, because under the additional assumptions for (B.6), the second term above is $a_n O(n^{-1/2} h_n^{-3/2} (\log n)^{1/2}) = o(n^{-1/2})$ a.s. Again by using Theorem 0.2 in Stute (1982) we can bound the first term above by $a_n^{-1} O((nh_n)^{-1/2} (\log n)^{1/2}) \times O(n^{-1/2} a_n^{1/2} (\log n)^{1/2}) = o(n^{-1/2})$ a.s. □

Lemma B.3. *Under the conditions of Theorem 3.1 we have*

$$\sup_{y \leq \tau_\eta} \left| U_n(y) - U(y) \right| = O\left((nh_n)^{-1/2} (\log n)^{1/2} \right) \ a.s., \tag{B.7}$$

$$\sup_{y \leq \tau_\eta} \left| \widetilde{U}_n(y) - U(y) \right| = O\left(n^{-1/3} (\log n)^{1/2} \right) \ a.s., \tag{B.8}$$

where $U(y)$, $U_n(y)$ and $\widetilde{U}_n(y)$ are defined in (4.2)

Proof. To deal with (B.7) we calculate

$$
\begin{aligned}
U_n(y) - U(y) = &- \int_{B_y} \int_{-\infty}^{y} \phi_x' \left(\widehat{\overline{H}}_e(s|x) \right) d\left[\widehat{H}_e^u(s|x) - H_e^u(s|x) \right] dF_X(x) \\
&+ \int_{B_y} \int_{-\infty}^{y} \left[\phi_x' \left(\overline{H}_e(s|x) \right) - \phi_x' \left(\widehat{\overline{H}}_e(s|x) \right) \right] dH_e^u(s|x) dF_X(x).
\end{aligned}
$$

Now, by using integration by parts for the first term, and a first order Taylor expansion for the second term we get

$$\begin{aligned} U_n(y) - U(y) = & - \int_{B_y} \phi_x' \left(\widehat{\overline{H}}_e(y|x) \right) \left[\widehat{H}_e^u(y|x) - H_e^u(y|x) \right] dF_X(x) \\ & + \int_{B_y} \int_{-\infty}^{y} \phi_x'' \left(\widehat{\overline{H}}_e(s|x) \right) \left[\widehat{H}_e^u(s|x) - H_e^u(s|x) \right] d\widehat{\overline{H}}_e(s|x) dF_X(x) \\ & - \int_{B_y} \int_{-\infty}^{y} \phi_x'' \left(\overline{H}_e(s|x) \right) \left[\widehat{\overline{H}}_e(s|x) - \overline{H}_e(s|x) \right] dH_e^u(s|x) dF_X(x) \\ & + \frac{1}{2} \int_{B_y} \int_{-\infty}^{y} \phi_x''' \left(\xi(s,x) \right) \left[\widehat{\overline{H}}_e(s|x) - \overline{H}_e(s|x) \right]^2 dH_e^u(s|x) dF_X(x), \end{aligned}$$

where $\xi(s,x)$ is between $\overline{H}_e(s|x)$ and $\widehat{\overline{H}}_e(s|x)$. Now, by using assumption (A4) and Lemma B.4 below we get the desired order.

Statement (B.8) can be bounded by $\sup_{y \le \tau_\eta} |\widetilde{U}_n(y) - U_n(y)| + \sup_{y \le \tau_\eta} |U_n(y) - U(y)|$. The first term is of the order $O\left(n^{-1/3}(\log n)^{1/2}\right)$ a.s. by Lemma B.2, while the second term is $O\left((nh_n)^{-1/2}(\log n)^{1/2}\right)$ a.s. by (B.7). This concludes the proof. □

The following lemma is a generalization of Lemma 4.1 in Van Keilegom and Akritas (1999) regarding the uniform (in $y \in I\!R$ and $x \in R_X$) rate of convergence of the difference $\widehat{H}(y|x) - H(y|x)$ to the uniform rate of convergence of $\widehat{H}_e(y|x) - H_e(y|x)$. The former difference is a sum of i.i.d. random variables, while the latter difference is a sum of non-independent random variables.

Lemma B.4. *Assume (A1)-(A3), and assume that (A5)(i,ii,iv,v) holds for* $H_e(y|x)$ *and* $H_e^u(y|x)$. *Then,*

(i) $$\sup_{x \in R_X} \sup_{y \in I\!R} \left| \widehat{H}_e(y|x) - H_e(y|x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \ a.s.$$

(ii) $$\sup_{x \in R_X} \sup_{y \in I\!R} \left| \dot{\widehat{H}}_e(y|x) - \dot{H}_e(y|x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \ a.s.$$

(iii) $$\sup_{x \in R_X} \sup_{y \in I\!R} \left| \widehat{H}_e^u(y|x) - H_e^u(y|x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \ a.s.$$

(iv) $$\sup_{x \in R_X} \sup_{y \in I\!R} \left| \dot{\widehat{H}}_e^u(y|x) - \dot{H}_e^u(y|x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \ a.s.$$

Proof. The proofs of statements (i), (ii), (iii) and (iv) use the same idea so we will only show (i). For that we define a new estimator

$$\widehat{H}_e^*(y|x) = \sum_{i=1}^{n} W_{ni}(x,h_n) I(E_i \le y),$$

for which statement (i) is true by Lemma 4.1 in Van Keilegom and Akritas (1999). To finish the proof we will show that the difference between $\widehat{H}_e$ and $\widehat{H}_e^*$ is uniformly of the order $O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. Consider

$$
\begin{aligned}
\left|\widehat{H}_e(y|x) - \widehat{H}_e^*(y|x)\right| &= \left|\sum_{i=1}^n W_{ni}(x,h_n)\left[I(\widehat{E}_i \le y) - I(E_i \le y)\right]\right| \\
&= \left|\sum_{i=1}^n W_{ni}(x,h_n)\left[I\left(E_i \le y\frac{\widehat{\sigma}(X)}{\sigma(X)} + \frac{\widehat{m}(X)-m(X)}{\sigma(X)}\right) - I(E_i \le y)\right]\right| \\
&\le \sum_{i=1}^n W_{ni}(x,h_n)\left[I(E_i \le y + |y|\beta_n + \alpha_n) - I(E_i \le y - |y|\beta_n - \alpha_n)\right] \\
&= \widehat{H}_e^*(y + |y|\beta_n + \alpha_n|x) - \widehat{H}_e^*(y - |y|\beta_n - \alpha_n|x),
\end{aligned}
$$

where $\alpha_n = \sup_{x\in R_X} |(\widehat{m}(x) - m(x))/\sigma(x)|$ and $\beta_n = \sup_{x\in R_X} |(\widehat{\sigma}(x) - \sigma(x))/\sigma(x)|$. Now, by adding and subtracting $H_e(y + |y|\beta_n + \alpha_n|x) - H_e(y - |y|\beta_n - \alpha_n|x)$, we get

$$
\begin{aligned}
\sup_{y\in I\!R,\ x\in R_X}\left|\widehat{H}_e(y|x) - \widehat{H}_e^*(y|x)\right| &\le \sup_{y\in I\!R,\ x\in R_X}\left|\widehat{H}_e^*(y + |y|\beta_n + \alpha_n|x) - H_e(y + |y|\beta_n + \alpha_n|x)\right| \\
&\quad + \sup_{y\in I\!R,\ x\in R_X}\left|\widehat{H}_e^*(y - |y|\beta_n - \alpha_n|x) - H_e(y - |y|\beta_n - \alpha_n|x)\right| \\
&\quad + \sup_{y\in I\!R,\ x\in R_X}\left|H_e(y + |y|\beta_n + \alpha_n|x) - H_e(y - |y|\beta_n - \alpha_n|x)\right| \\
&= 2\sup_{y\in I\!R,\ x\in R_X}\left|\widehat{H}_e^*(y|x) - H_e(y|x)\right| \\
&\quad + \sup_{y\in I\!R,\ x\in R_X}\left|h_e(\xi_{x,y}|x)\left[2\beta_n|y| + 2\alpha_n\right]\right|,
\end{aligned}
$$

where $\xi_{x,y}$ is between $y - |y|\beta_n - \alpha_n$ and $y + |y|\beta_n + \alpha_n$. As explained in the beginning the first term on the right hand side is of the order $O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. The second term is of the same order because $\sup_{y\in I\!R,\ x\in R_X} |h_e(y|x)y| < \infty$ by assumption (A5)(ii) and because α_n and β_n are $O((nh_n)^{-1/2}(\log n)^{1/2})$ *a.s.* by Proposition 3.5 in Sujica and Van Keilegom (2013). □

Lemma B.5. *Under the assumptions of Theorem 3.1 we have*

$$
\sup_{y\le\tau_\eta}\sup_{x\in B_y}\int_{-\infty}^{y}\left[\phi_x'(\widehat{\overline{H}}_e(s|x)) - \phi_x'\left(\overline{H}_e(s|x)\right)\right] d\left[\widehat{H}_e^u(s|x) - H_e^u(s|x)\right]
$$
$$
= O((nh_n)^{-3/4}(\log n)^{3/4+\alpha})\ a.s., \qquad \text{(B.9)}
$$

where $\alpha > 0$ is an arbitrarily small constant.

***Proof*.** The proof is very analogous to the proof of Lemma A.1 in Sujica and Van Keilegom (2013). The only difference is that we use Lemma B.6 below for the uniform rate of convergence of the modulus of continuity $\widehat{H}_e(y_1|x) - \widehat{H}_e(y_2|x) - H_e(y_1|x) + H_e(y_2|x)$, whereas they use Lemma 4.4 in Du and Akritas (2002) for $\widehat{H}(y_1|x) - \widehat{H}(y_2|x) - H(y_1|x) + H(y_2|x)$. Details of the proof are omitted and can be found in Lemma A.1 in Sujica and Van Keilegom (2013). □

The following lemma is a uniform modulus of continuity result for the Nadaraya-Watson type estimator $\widehat{H}_e(y|x)$. It is an adaptation of Lemma 4.4 in Du and Akritas (2002). A major difficulty in this adaptation is that $\widehat{H}_e(y|x)$ is not a sum of independent random variables.

Lemma B.6. *Assume (A1)-(A4), and assume that (A5) and (A6) hold for $H_e(y|x)$. Let $a_n = O((nh_n)^{-1/2}(\log n)^{1/2})$. Then,*

$$\sup_{x\in R_X}\sup_{(y_1,y_2)\in J_{a_n}} |\ \widehat{H}_e(y_1|x) - \widehat{H}_e(y_2|x) - H_e(y_1|x) + H_e(y_2|x)\ |$$
$$= O\left(a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2+\alpha} + a_n^2 + a_nh_n\right) = o\left(n^{-1/2}\right)\ a.s., \quad \text{(B.10)}$$

for an arbitrarily small constant $\alpha > 0$, $J_{a_n} = \{(y_1, y_2) : \ |M(y_1) - M(y_2)| \le a_n\}$ and $M(y) = \sum_{i=1}^4 M_i(y)$ (see (A6) for $L = H_e$).

Proof. In this proof we will use an index $(-r)$ to denote estimators that leave out the random variables E_r and Δ_r:

$$\widehat{H}_{(-r)}(y|x) := \sum_{i=1,i\neq r}^{n} W_{ni}(x,h_n) I\left(T_i \le y\right),$$
$$\widetilde{\widetilde{F}}_{(-r)}(y|x) := \phi_x^{-1}\left\{-\sum_{T_i\le y,\Delta_i=1,i\neq r}\left[\phi_x\left(\widehat{\widehat{H}}_{(-r)}(T_i^-|x)\right) - \phi_x\left(\widehat{\widehat{H}}_{(-r)}(T_i|x)\right)\right]\right\},$$
$$\widehat{m}_{(-r)}(x) := \sum_{i=1,i\neq r}^{n} \Delta\widetilde{F}_{(-r)}(T_i|x)T_iJ(\widetilde{F}_{(-r)}(T_i|x)),$$
$$\widehat{\sigma}^2_{(-r)}(x) := \sum_{i=1,i\neq r}^{n} \Delta\widetilde{F}_{(-r)}(T_i|x)T_i^2J(\widetilde{F}_{(-r)}(T_i|x)) - \widehat{m}^2_{(-r)}(x).$$

We will also use the notation $\Delta(y,x) = y\Delta_2(x) + \Delta_1(x)$, where $\Delta_2(x) = \frac{\widehat{\sigma}(x)-\sigma(x)}{\sigma(x)}$ and $\Delta_1(x) = \frac{\widehat{m}(x)-m(x)}{\sigma(x)}$. Further we will denote $\Delta_{(-r)}$, $\Delta_{1(-r)}$ and $\Delta_{2(-r)}$ when replacing $\widehat{m}$ and $\widehat{\sigma}$ by $\widehat{m}_{(-r)}$ and $\widehat{\sigma}_{(-r)}$ in the functions Δ, Δ_1 and Δ_2, respectively.

We will prove the main statement of the Lemma by showing the following two statements:

$$\sup_{x\in R_X}\sup_{(y_1,y_2)\in J_{a_n}} |\ \widehat{H}_e(y_1|x) - \widehat{H}_e(y_2|x) - \widetilde{H}_e(y_1|x) + \widetilde{H}_e(y_2|x)\ |$$
$$= O(a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2+\alpha})\ a.s., \quad \text{(B.11)}$$

$$\sup_{x\in R_X}\sup_{(y_1,y_2)\in J_{a_n}} |\ \widetilde{H}_e(y_1|x) - \widetilde{H}_e(y_2|x) - H_e(y_1|x) + H_e(y_2|x)\ |$$
$$= O(a_n^2 + a_nh_n)\ a.s., \quad \text{(B.12)}$$

where $\widetilde{H}_e(y|x) = \sum_{r=1}^n W_{nr}(x, h_n) H_e(y + \Delta_{(-r)}(y, X_r)|X_r)$. To show (B.12) we calculate, for all y_1 and y_2 such that $|(M_2 + M_3 + M_4)(y_1) - (M_2 + M_3 + M_4)(y_2)| < a_n$, the following:

$$
\begin{aligned}
&\widetilde{H}_e(y_1|x) - \widetilde{H}_e(y_2|x) - H_e(y_1|x) + H_e(y_2|x) \\
&= \sum_{r=1}^n W_{nr}(x, h_n) \Big\{ H_e(y_1 + \Delta_{(-r)}(y_1, X_r)|X_r) - H_e(y_2 + \Delta_{(-r)}(y_2, X_r)|X_r) \\
&\qquad\qquad - H_e(y_1|x) + H_e(y_2|x) \Big\} \\
&= \sum_{r=1}^n W_{nr}(x, h_n) \Big\{ H'_e(y_1|X_r)\Delta_{(-r)}(y_1, X_r) - H'_e(y_2|X_r)\Delta_{(-r)}(y_2, X_r) \\
&\qquad\qquad + \frac{1}{2} H''_e(\xi_{1r}|X_r)\Delta^2_{(-r)}(y_1, X_r) - \frac{1}{2} H''_e(\xi_{2r}|X_r)\Delta^2_{(-r)}(y_2, X_r) \\
&\qquad\qquad + [\dot{H}_e(y_1|x'_r) - \dot{H}_e(y_2|x'_r)](X_r - x) \Big\} \\
&= O(a_n^2 + a_n^2 + a_n h_n) \ a.s.,
\end{aligned}
$$

where x'_r is between x and X_r, ξ_{ir} is between y_i and $y_i + \Delta_{(-r)}(y_i, X_r)$ for $i = 1, 2$ and $r = 1, ..., n$. In the last equality we used assumptions (A5)(ii,iii), Proposition 3.5 in Sujica and Van Keilegom (2013), and the fact that $\sup_{x \in R_X} \max_{r=1,\dots,n} |W_{nr}(x, h_n)| = O((nh_n)^{-1})$ a.s., which is easy to show.

To prove (B.11) we start by partitioning $I\!R$ into $m_n = \lfloor M(+\infty)/a_n \rfloor$ subintervals $-\infty = y_0 < y_1 < ... < y_{m_n} = \infty$, such that $M(y_{i+1}) - M(y_i) = \bar{a}_n := M(+\infty)/m_n$. For each $i = 1, ..., m_n - 1$ define $I_{ni} = [y_{i-1}, y_{i+1}]$. Further partition each interval I_{ni} into $2b_n$ smaller intervals $[y_{ij}, y_{i,j+1}]$ for $j = -b_n, -b_n + 1, ..., b_n$, where $b_n = O(a_n^{1/2}(nh_n)^{1/2}(\log n)^{-1/2})$, such that $M(y_{i,j+1}) - M(y_{ij}) = \frac{\bar{a}_n}{b_n}$. It can be easily verified that $a_n \le \bar{a}_n \le 2a_n$ for n large enough, and for any $y_1, y_2 \in I\!R$ with $|M(y_1) - M(y_2)| < a_n$, there exists an interval I_{ni} such that $y_1, y_2 \in I_{ni}$. Hence, by using the monotonicity of $\widehat{H}_e(\cdot|x)$, it can be seen that (B.11) is bounded by

$$\sup_{x \in R_X} \max_{1 \le i \le m_n - 1} \max_{-b_n \le j,k \le b_n} | \widehat{H}_e(y_{ik}|x) - \widehat{H}_e(y_{ij}|x) - \widetilde{H}_e(y_{ik}|x) + \widetilde{H}_e(y_{ij}|x) | \quad \text{(B.13)}$$

$$+2 \sup_{x \in R_X} \sup\{| \widetilde{H}_e(y_1|x) - \widetilde{H}_e(y_2|x) | : |M(y_1) - M(y_2)| \le \bar{a}_n/b_n\}. \quad \text{(B.14)}$$

We can write the term between absolute values in (B.14) as

$$\sum_{r=1}^n W_{nr}(x, h_n) \Big\{ H_e(y_1 + \Delta_{(-r)}(y_1, X_r)|X_r) - H_e(y_2 + \Delta_{(-r)}(y_2, X_r)|X_r) \Big\}$$

$$= \sum_{r=1}^{n} W_{nr}(x, h_n) \Big\{ H_e(y_1|X_r) - H_e(y_2|X_r)$$
$$+ H_e'(y_1|X_r)\Delta_{(-r)}(y_1, X_r) - H_e'(y_2|X_r)\Delta_{(-r)}(y_2, X_r)$$
$$+ \frac{1}{2} H_e''(\xi_{1r}|X_r)\Delta^2_{(-r)}(y_1, X_r) - \frac{1}{2} H_e''(\xi_{2r}|X_r)\Delta^2_{(-r)}(y_2, X_r) \Big\},$$

where ξ_{ir} is between y_i and $y_i + \Delta_{(-r)}(y_i, X_r)$ for $i = 1, 2$. The first difference on the right hand side is of the order $\frac{a_n}{b_n} = O(a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2})$. The third difference is of the order $O((nh_n)^{-1}\log n)$ a.s., because of assumption (A5)(iii), and because the terms $\Delta_{1(-r)}(y_i, X_r)$ and $\Delta_{2(-r)}(y_i, X_r)$ are uniformly $O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. (by Proposition 3.5 in Sujica and Van Keilegom (2013), and the relation $\sup_{x\in R_X} \max_{r=1,\ldots,n} |W_{nr}(x, h_n)| = O((nh_n)^{-1})$ a.s.). To show that the second difference is negligible we calculate

$$\big| \big[H_e'(y_1|X_r)y_1 - H_e'(y_2|X_r)y_2\big]\Delta_{2(-r)}(X_r) + \big[H_e'(y_1|X_r) - H_e'(y_2|X_r)\big]\Delta_{1(-r)}X_r)\big|$$
$$\le \big|M(y_1) - M(y_2)\big| \, \big|\Delta_{2(-r)}(X_r)\big| + \big|M(y_1) - M(y_2)\big| \, \big|\Delta_{1(-r)}(X_r)\big|$$
$$= O\left(\frac{a_n}{b_n}(nh_n)^{-1/2}(\log n)^{1/2}\right) \ a.s.$$

Hence, we showed that (B.14) is $O(a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2})$ a.s.

To deal with (B.13) we define the grid $a = x_0 < x_1 < \ldots < x_{k_n} = b$ of $R_X = [a, b]$ such that $x_i - x_{i-1} \le a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2}h_n^3$, $i = 1, .., k_n$. By similar calculations as in Lemma 4.2 in Du and Akritas (2002), we have uniformly, up to a remainder term of order $O(a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2})$ a.s., that (B.13) is bounded by

$$A_n := \max_{l=1,\ldots,k_n} \max_{1\le i\le m_n-1} \max_{-b_n\le j,k\le b_n} \left| \widehat{H}_e(y_{ik}|x_l) - \widehat{H}_e(y_{ij}|x_l) - \widetilde{H}_e(y_{ik}|x_l) + \widetilde{H}_e(y_{ij}|x_l) \right|.$$

Before continuing we define D_n as a set where for a fixed constant $0 < C' < \infty$ the following conditions are satisfied:

$$\sup_{x\in R_X} \max_{r=1,\ldots,n} |W_{nr}(x, h_n)| \le C'(nh_n)^{-1}(\log n)^{\alpha}, \tag{B.15}$$

$$\sup_{x\in R_X} \max_{r=1,\ldots,n} \max_{i=1,2} |\Delta_{i(-r)}(x)| \le C'(nh_n)^{-1/2}(\log n)^{1/2+\alpha}, \tag{B.16}$$

$$\sup_{x\in R_X} \max_{r=1,\ldots,n} \max_{i=1,2} |\Delta_{i(-r)}(x) - \Delta_i(x)| \le C'(nh_n)^{-1}(\log n)^{\alpha}. \tag{B.17}$$

We can show that $P(\cup_{m=1}^{\infty} \cap_{n=m}^{\infty} D_n) = 1$. Indeed, as mentioned before we have that $\sup_{x\in R_X} \max_{r=1,\ldots,n} |W_{nr}(x, h_n)| = O((nh_n)^{-1})$ a.s. By additionally using Proposition 3.5 in Sujica and Van Keilegom (2013), we have that the term on the left hand side in (B.16) is $O((nh_n)^{-1})$ a.s. Proving that the term on the left hand side in (B.17) is $O((nh_n)^{-1}\log n)$

a.s. is a straightforward but tedious calculation and is omitted. On the set D_n, we can write $\widehat{H}_e(y_{ik}|x) - \widehat{H}_e(y_{ij}|x) - \widetilde{H}_e(y_{ik}|x_l) + \widetilde{H}_e(y_{ij}|x_l) = \sum_{r=1}^n X_{rijkl}$, where

$$X_{rijkl} = \widetilde{\widetilde{W}}_{nr}(x_l, h_n)\Big\{ I(E_r \le y_{ik} + \widetilde{\Delta}_{(-r)}(y_{ik}, X_r)) - I(E_r \le y_{ij} + \widetilde{\Delta}_{(-r)}(y_{ij}, X_r)) + \widetilde{Z}_{rijkl} \\ - H_e(y_{ik} + \widetilde{\Delta}_{(-r)}(y_{ik}, X_r)|X_r) + H_e(y_{ij} + \widetilde{\Delta}_{(-r)}(y_{ij}, X_r)|X_r) \Big\},$$

$Z_r(y) := I(E_r \le y + \Delta(y, X_r)) - I(E_r \le y + \Delta_{(-r)}(y, X_r))$, $\widetilde{Z}_{rijk} = [Z_r(y_{ik}) - Z_r(y_{ij})]I_{D_n}$, $\widetilde{\widetilde{W}}_{nr}(x, h_n) = W_{nr}(x, h_n)I_{D_n}$, $\widetilde{\Delta}_{(-r)}(y, x) = y\widetilde{\Delta}_{2(-r)}(x) + \widetilde{\Delta}_{1(-r)}(x)$ and $\widetilde{\Delta}_{i(-r)}(x) = \Delta_{i(-r)}(x)$ $\times I_{D_n}$, $i = 1, 2$. We define a (centered) version of the random variable X_{rijkl} by $\widetilde{X}_{rijkl} = X_{rijkl} - E[\widetilde{Z}_{rijk}|\mathfrak{X}_{(-r)}, X_r]$, where $\mathfrak{X}_{(-r)} = \{X_i, T_i, \Delta_i\}_{i \ne r}$. Now in order to prove (B.13) we will use a modification of Bernstein's inequality (see Theorem 1.2A in de la Peña (1999)). Using the notation $\widetilde{\mathfrak{X}}_{r-1,ijkl} = \{\widetilde{X}_{1,ijkl}, ..., \widetilde{X}_{r-1,ijkl}\}$, if the following is satisfied

$$E[\widetilde{X}_{rijkl}|\widetilde{\mathfrak{X}}_{r-1,ijkl}] = 0, \tag{B.18}$$

$$\sum_{r=1}^n E[\widetilde{X}_{rijkl}^2|\widetilde{\mathfrak{X}}_{r-1,ijkl}] \le v_n. \tag{B.19}$$

$$E[\widetilde{X}_{rijkl}^p|\widetilde{\mathfrak{X}}_{r-1,ijkl}] \le \frac{1}{2}E[\widetilde{X}_{rijkl}^2|\widetilde{\mathfrak{X}}_{r-1,ijkl}]L_n^{p-2}p!, \tag{B.20}$$

where $L_n > 0$ and $v_n > 0$ are constants, we have

$$P\left(\sum_{r=1}^n \widetilde{X}_{rijkl} > \lambda_n\right) \le \exp\left\{-\lambda_n^2/2\left[v_n + L_n\lambda_n\right]\right\}.$$

We will use now, and show later that the conditions (B.18), (B.19) and (B.20) can be verified for constants $v_n = Ca_n(nh_n)^{-1}(\log n)^\alpha$ and $L_n = C(nh_n)^{-1}(\log n)^\alpha$. Now, for $\lambda_n = c_1 a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2+\alpha}$, where c_1 is a positive constant to be specified further on, we have

$$\begin{aligned} P\left(\{A_n > 2\lambda_n\} \cap D_n\right) &\le P\left(\max_{l=1,...,k_n}\max_{1\le i\le m_n-1}\max_{-b_n\le j,k\le b_n}\left|\sum_{r=1}^n X_{rijkl}\right| > 2\lambda_n\right) \\ &\le P\left(\max_{l=1,...,k_n}\max_{1\le i\le m_n-1}\max_{-b_n\le j,k\le b_n}\left|\sum_{r=1}^n \widetilde{X}_{rijkl}\right| > \lambda_n\right) \\ &\le 2k_n(m_n-1)(2b_n+1)^2\exp\left\{-C\ \lambda_n^2/\left(\frac{a_n(\log n)^\alpha}{nh_n} + \frac{\lambda_n(\log n)^\alpha}{nh_n}\right)\right\} \\ &\le 2k_n(m_n-1)(2b_n+1)^2 n^{-c_1 C}, \end{aligned}$$

for some constant $C > 0$. For the second inequality we used that $\lambda_n > \sum_{r=1}^n E[\widetilde{Z}_{rijk}|\mathfrak{X}_{(-r)}, X_r]$ by (B.23) below, while for the third inequality we used the modified Bernstein's inequality.

Since, by proper choice of c_1, this can be made summable, using the Borel-Cantelli lemma we get $P(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\{A_n > 2\lambda_n\} \cap D_n\}) = 0$. From here we can calculate

$$P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{A_n > 2\lambda_n\}\right) \leq P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\left[\{\{A_n > 2\lambda_n\} \cap D_n\} \cup D_n^C\right]\right)$$
$$\leq P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\{A_n > 2\lambda_n\} \cap D_n\}\right) + P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} D_n^C\right).$$

As we have just shown, the first term on the right hand side is 0 by the Borel-Cantelli lemma, while the second term is 0, since $P(\cup_{m=1}^{\infty}\cap_{n=m}^{\infty} D_n) = 1$. Therefore, this proves (B.13) and subsequently (B.11). Finally, to show the conditions that allowed us to use the modified Bernstein's inequality, we start with condition (B.18), that is satisfied since by definition $E[\widetilde{X}_{rijkl}|\mathfrak{X}_{(-r)}, X_r] = 0$. Condition (B.20) follows from

$$E[\widetilde{X}_{rijkl}^p|\widetilde{\mathfrak{X}}_{r-1,ijkl}] \leq E[\widetilde{X}_{rijkl}^2\widetilde{W}_{nr}^{p-2}(x_l, h_n)|\widetilde{\mathfrak{X}}_{r-1,ijkl}]$$
$$\leq E[\widetilde{X}_{rijkl}^2|\widetilde{\mathfrak{X}}_{r-1,ijkl}](C'(nh_n)^{-1}(\log n)^{\alpha})^{(p-2)},$$

where the last inequality is uniform in r, i, j, k and l. Before continuing we calculate:

$$\begin{aligned}E[\widetilde{X}_{rijkl}^2|\mathfrak{X}_{(-r)}, X_r] =&\widetilde{W}_{nr}^2(x_l, h_n)\Big\{ \Big|H_e(y_{ik} + \widetilde{\Delta}_{(-r)}(y_{ik}, X_r)|X_r) - H_e(y_{ij} + \widetilde{\Delta}_{(-r)}(y_{ij}, X_r)|X_r)\Big| \\ &- \Big[H_e(y_{ik} + \widetilde{\Delta}_{(-r)}(y_{ik}, X_r)|X_r) - H_e(y_{ij} + \widetilde{\Delta}_{(-r)}(y_{ij}, X_r)|X_r)\Big]^2 \\ &+ E\Big[\{\widetilde{Z}_{rijk} - E[\widetilde{Z}_{rijk}|\mathfrak{X}_{(-r)}, X_r]\}^2|\mathfrak{X}_{(-r)}, X_r\Big]\Big\} \qquad \text{(B.21)} \\ =&\widetilde{W}_{nr}^2(x_l, h_n)\,[a_n + R_n]\,,\end{aligned}$$

where $|R_n| \leq C(a_n^2(\log n)^{2\alpha} + (nh_n)^{-1}(\log n)^{\alpha})$ uniformly in r, i, j, k and l. To bound the first and the second difference above we use the following calculation:

$$\begin{aligned}&H_e(y_{ik} + \widetilde{\Delta}_{(-r)}(y_{ik}, X_r)|X_r) - H_e(y_{ij} + \widetilde{\Delta}_{(-r)}(y_{ij}, X_r)|X_r) \\ &= [H_e(y_{ik}|X_r) - H_e(y_{ij}|X_r)] \\ &\quad + \Big[H_e'(y_{ik}|x)\widetilde{\Delta}_{(-r)}(y_{ik}, X_r) - H_e'(y_{ij}|x)]\widetilde{\Delta}_{(-r)}(y_{ij}, X_r)\Big] \\ &\quad + \Big[\frac{1}{2}H_e''(\xi_{ik}|x)\widetilde{\Delta}_{(-r)}^2(y_{ik}, X_r) - \frac{1}{2}H_e''(\xi_{ij}|x)]\widetilde{\Delta}_{(-r)}^2(y_{ij}, X_r)\Big] \qquad \text{(B.22)} \\ &= O(a_n) + R_n',\end{aligned}$$

uniformly in r, i, j and k, where $|R_n'| \leq C(a_n^2(\log n)^{\alpha} + a_n^2(\log n)^{2\alpha})$ and ξ_{it} is between y_{it} and $y_{it} + \widetilde{\Delta}_{(-r)}(y_{it}, X_r)$, $t = k, j$. For the second and the third term of (B.22) we used (B.16) and assumptions (A5)(ii) and (A5)(iii). To bound the third term in (B.21) we use the following reasoning. We first define $c_n = C'(nh_n)^{-1}(\log n)^{\alpha}$. Now, on the set D_n we can conclude that the variable $I(E_r \le s + \Delta(s, X_r))$ lies in between $I(E_r \le s + \Delta_{(-r)}(s, X_r) - sc_n - c_n)$ and $I(E_r \le s + \Delta_{(-r)}(s, X_r) + sc_n + c_n)$. Therefore, we have that

$$\begin{aligned}
\big|E[Z_r(s) I_{D_n} | \mathcal{X}_{(-r)}, X_r]\big| &\le E\Big[I(E_r \le s + \widetilde{\Delta}_{(-r)}(s, X_r) + sc_n + c_n) \\
&\qquad - I(E_r \le s + \widetilde{\Delta}_{(-r)}(s, X_r) - sc_n - c_n) \Big| \mathcal{X}_{(-r)}, X_r \Big] \\
&\le H_e\left(s + \widetilde{\Delta}_{(-r)}(s, X_r) + sc_n + c_n | X_r\right) \\
&\qquad - H_e\left(s + \widetilde{\Delta}_{(-r)}(s, X_r) - sc_n - c_n | X_r\right) \\
&\le 2 \sup_{s,x} \big| H_e'(s|x)[|s|c_n + c_n] \big| \\
&\le C\ (nh_n)^{-1}(\log n)^{\alpha}, \qquad \text{(B.23)}
\end{aligned}$$

for some constant $C > 0$, uniformly in $-\infty < s < \infty$ and for all $r \le n$. The last inequality follows from assumption (A5)(ii). This proves (B.21). Condition (B.19) is now satisfied, since $\sum_{r=1}^{n} E[\widetilde{X}_{rijkl}^2 | \widetilde{\mathcal{X}}_{r-1,ijkl}] \le\ Ca_n(nh_n)^{-1}(\log n)^{\alpha}$ uniformly, because of (B.21). This completes the proof. □

References

Akritas, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics 22*, 1299–1327.

Beran, R. (1981). *Nonparametric Regression with Randomly Censored Survival Data.* University of California, Berkeley: Technical Report.

Braekers, R. and N. Veraverbeke (2005). A copula-graphic estimator for the conditional survival function under dependent censoring. *Canadian Journal of Statistics 33*, 429–447.

Chen, S., G. B. Dahl, and S. Khan (2005). Nonparametric identification and estimation of a censored location-scale regression model. *The Journal of the American Statistical Association 100*, 212–221.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Annals of Statistics 17*, 1157–1167.

de la Peña, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *The Annals of Probability 27*, 537–564.

Dette, H. and C. Heuchenne (2012). Scale checks in censored regression. *Scandinavian Journal of Statistics 39*, 323–339.

Du, Y. and M. G. Akritas (2002). Uniform strong representation of the conditional Kaplan-Meier process. *Mathematical Methods of Statistics 11*, 152–182.

Giné, E. and R. Nickle (2008). Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields 141*, 333–387.

González-Manteiga, W. and C. Cadarso-Suarez (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics 4*, 65–78.

Heuchenne, C. and I. Van Keilegom (2007a). Nonlinear regression with censored data. *Technometrics 49*, 34–44.

Heuchenne, C. and I. Van Keilegom (2007b). Polynomial regression with censored data based on preliminary nonparametric estimation. *Annals of the Institute of Statistical Mathematics 59*, 273–297.

Lambert, P. (2013). Nonparametric additive location-scale models for interval censored data. *Statistics and Computing 23*, 75–90.

Lewbel, A. and O. Linton (2002). Nonparametric censored and truncated regression. *Econometrica 70*, 765–779.

Linton, O., E. Mammen, J. P. Nielsen, and I. Van Keilegom (2011). Nonparametric regression with filtered data. *Bernoulli 17*, 60–87.

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

Pardo-Fernández, J. C. and I. Van Keilegom (2006). Comparison of regression curves with censored responses. *Scandinavian Journal of Statistics 33*, 409–434.

Pardo-Fernández, J. C., I. Van Keilegom, and W. González-Manteiga (2007). Goodness-of-fit tests for parametric models in censored regression. *Canadian Journal of Statistics 35*, 249–264.

Rivest, L. and M. T. Wells (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *Journal of Multivariate Analysis 79*, 138–155.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris 8*, 229–231.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics 5*, 595–645.

Stute, W. (1982). The oscillation behavior of empirical processes. *Annals of Probability 10*, 86–107.

Sujica, A. and I. Van Keilegom (2013). Estimation of location and scale functionals in nonparametric regression under copula dependent censoring. *Canadian Journal of Statistics (under revision).*

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America 72*, 20–22.

Van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes.* New York: Springer.

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

Van Keilegom, I. and N. Veraverbeke (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals of the Institute of Statistical Mathematics 49*, 467–491.

Van Keilegom, I. and N. Veraverbeke (2002). Density and hazard estimation in censored regression models. *Bernoulli 8*, 607–625.

Zheng, M. and J. Klein (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika 82*, 127–138.