Estimation of location and scale functionals in nonparametric regression under copula dependent censoring

February 7, 2015

Abstract

Let (X, Y) be a random vector, where Y denotes the variable of interest possibly subject to random right censoring, and X is a covariate. The variable Y is a (possible monotone transformation of a) survival time. The censoring time C and the survival time Y are allowed to be dependent, and the dependence is described via a known copula (this also includes the independent case). Under this setting we propose estimators of certain location and scale functionals of Y given X. We derive their asymptotic properties, uniformly over the support of X. In particular we derive an asymptotic representation and the uniform convergence rates for these estimators and their derivatives. We also prove asymptotic results for an estimator of the conditional distribution (the so-called conditional copula-graphic estimator), which generalizes previous results obtained by Braekers and Veraverbeke (2005). We also illustrate the results via simulations and the analysis of data on bone marrow transplantation.

Key Words: Asymptotic representation; convergence rates; copulas; dependent censoring; nonparametric regression, right censoring, survival analysis.

1 Introduction

Let (X, Y) be a random vector where Y denotes a possible transformation of the variable of interest and X is a covariate. We assume that Y 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).

In the statistical literature it is very common to assume that Y and C are independent given X. Under this assumption a lot of work has been done on the nonparametric estimation of the conditional distribution $F(\cdot|x) = P(Y \leq \cdot|X = x)$. We refer to Beran (1981), Dabrowska (1989), González-Manteiga and Cadarso-Suarez (1994), Akritas (1994), Van Keilegom and Veraverbeke (1997a,b), among many others. The nonparametric kernel estimator in this setting is often referred to as Beran's estimator, and is a generalization of the Kaplan-Meier estimator to the inclusion of covariates.

However, in various situations it is unrealistic to assume that Y and C are conditionally independent given X. Consider e.g. the situation where a patient decides to leave a medical study because he or she feels in very good shape and prefers therefore to stop the treatment. In such a case the censoring time C will likely be negatively correlated with the survival time Y, and the more the patient feels healthy the more this correlation will be negative. On the other hand we might also have patients who decide to stop a certain treatment because they are not in good health and would e.g. prefer to change treatment or hospital. These are patients for which Y and C will tend to be positively correlated. Again, the strength of the dependence will be determined by how bad the patient is feeling. More generally, in many situations the strength of the dependence between Y and C will depend on a certain number of external factors (or covariates). This motivates us to consider in this paper the situation where Y and C are dependent given X.

However, in the context without covariates, Tsiatis (1975) showed in his seminal work that the joint distribution of Y and C cannot be identified by their minimum and the censoring indicator when the dependence between Y and C is unspecified. Crowder (1991) showed that even when the marginal distributions are known, the joint distribution function is not identifiable. Tsiatis' observations have been the starting point for research on how to modify the model so as to identify the distribution of (Y, C), and a variety of modeling approaches have been studied in the past. One such approach consists in modeling the dependence structure between Y and C by means of a copula function. The advantage of using copulas is that one only models the dependence structure without affecting the margins (see Sklar (1959)). Wang (2012) showed that even if we restrict attention to Archimedean

copula functions, the model is in general not identifiable, whereas Schwarz, Jongbloed, and Van Keilegom (2013) showed that when the margins are known, the copula function can under certain conditions be identified. When the margins are unknown, Zheng and Klein (1995) and Rivest and Wells (2001) supposed that Y and C are dependent via some known copula and they showed that the marginal distribution 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. An extension of this copula-graphic estimator has been proposed by de Uña Álvarez and Veraverbeke (2013), when the full process is independently censored by some administrative censoring time.

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. We follow their approach, except that we assume that X is random, and we 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 :

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

where $1 - G(c|x) = P(C > c | X = x)$ is the conditional survival function of the censoring time C given $X = x$. Because of its nice properties and because of the broad range of different copula structures it covers, we focus attention in what follows on the subclass of Archimedean copulas, i.e. we assume that

$$P(Y > y, \ C > c | X = x) \ = \ \phi_x^{-1} \Big[\phi_x \big\{ 1 - F(y|x) \big\} + \phi_x \big\{ 1 - G(c|x) \big\} \Big], \qquad (1.1)$$

for an Archimedean copula generator ϕ_x, i .e. a function from $(0, 1]$ to $\mathbb{R}^+$ that is decreasing, convex and that satisfies $\phi_x(1) = 0$ and $\phi_x(0^+) = +\infty$. Under these conditions Braekers and Veraverbeke (2005) proposed the conditional copula-graphic estimator, which generalizes the Beran estimator to the dependent setting (1.1). Their estimator reduces to Beran's estimator when $\phi_x(\cdot) = -\log(\cdot)$.

Under this setting of dependent censoring described by an Archimedean copula, we are interested in studying location and scale functionals of Y given $X = x$, which we will denote by $m(x)$ and $\sigma^2(x)$. Because we are working in a completely nonparametric framework and because the response is subject to right censoring, we won't be able to estimate the conditional mean and the conditional variance in a consistent way. Instead we will assume that $m(x)$ and $\sigma^2(x)$ take the following form:

$$m(x) = \int_0^1 F^{-1}(s|x) J(s) ds \qquad (1.2)$$

$$\sigma^2(x) = \int_0^1 \big[F^{-1}(s|x) - m(x) \big]^2 J(s) ds = \int_0^1 F^{-1}(s|x)^2 J(s) ds - m(x)^2,$$

2

where $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ is the quantile function of Y given $X = x$, and $J(s)$ is a given positive weight function such that $\int_0^1 J(s)ds = 1$. Note that the choice $J(s) \equiv 1$ would lead to the conditional mean and variance. But because of lack of data in the right tail of $F(\cdot|x)$, we will choose $J(s)$ so that no weight is given to the regions where the conditional distribution cannot be estimated consistently. This form of location and scale functionals (known as L-functionals) is very flexible and covers a broad range of common functionals (see e.g. Serfling (1980) to learn more about L-functionals).

The objective of this paper is to propose appropriate estimators of $m(x)$ and $\sigma^2(x)$ and to study their asymptotic properties uniformly in x. These quantities have so far not been studied in the literature on nonparametric regression with copula dependent censoring. Indeed, attention has been focused on the estimation of the conditional distribution $F(y|x)$ for a fixed value of x and under a fixed design setting. See Braekers and Veraverbeke (2005), Braekers and Veraverbeke (2008) and Gaddah and Braekers (2010a,b), where the latter three papers assume that the data satisfy a conditional Koziol-Green model. As a by-product we will also study an estimator of the conditional distribution $F(y|x)$. Compared to the aforementioned papers, the main difference is that we obtain results uniformly in x and y, which are technically harder to prove.

The paper is organized as follows. In Section 2 we define the estimators of $m(x)$ and $\sigma^2(x)$, and state the assumptions that will be needed for the asymptotic results. Section 3 contains the main asymptotic results of this paper. We first study an estimator of the conditional distribution $F(y|x)$ and we next give an asymptotic representation and the uniform convergence rates for the estimators of $m(x)$ and $\sigma^2(x)$ and of their derivatives. In Section 4 we show the results of a small simulation study, and in Section 5 we illustrate our estimation method via the analysis of data on bone marrow transplantation. Finally, the appendix contains the proofs of the asymptotic results stated in Section 3.

2   Definitions and assumptions

We focus in this section on the estimation of the functions $m(x)$ and $\sigma^2(x)$ given in (1.2). These functions depend on the conditional distribution $F(\cdot|x)$, which we need to estimate first. Braekers and Veraverbeke (2005) defined the so-called conditional copula-graphic estimator of $F(\cdot|x)$:

$$\widehat{F}(y|x) = 1 - \phi_x^{-1}\Big\{ - \sum_{T_i \leq y, \Delta_i = 1} \Big[\phi_x\Big(\overline{\widehat{H}}(T_i^-|x)\Big) - \phi_x\Big(\overline{\widehat{H}}(T_i|x)\Big) \Big] \Big\}, \tag{2.1}$$

3

where

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

is the Nadaraya-Watson estimator of the conditional distribution $H(y|x) = P(T \le y|X = x)$ of T given $X = x$, $\overline{\widehat{H}}(y|x) = 1 - \widehat{H}(y|x)$, $\widehat{H}(y^-|x) = \lim_{t \uparrow y} \widehat{H}(t|x)$,

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

are Nadaraya-Watson weights, K is a probability density function (kernel), and $h \equiv h_n$ is a bandwidth sequence tending to zero when n tends to infinity. The estimator $\widehat{F}(y|x)$ is an extension of the Beran estimator in the sense that it allows for dependent censoring, and it is also an extension of the estimator proposed by Zheng and Klein (1995), since it includes covariates. Moreover, it is easily seen that $\widehat{F}(y|x)$ is a proper distribution function if $\phi_x(0) = \infty$ (i.e. the generator ϕ_x is strict) and if the last observation is uncensored (as is the case for e.g. the Kaplan-Meier or Beran estimator).

This leads to the following estimators for $m(x)$ and $\sigma^2(x)$:

$$\widehat{m}(x) = \int_0^1 \widehat{F}^{-1}(s|x) J(s) ds \quad \text{and} \quad \widehat{\sigma}^2(x) = \int_0^1 \left[\widehat{F}^{-1}(s|x) - \widehat{m}(x)\right]^2 J(s) ds. \quad (2.2)$$

Note that these estimators are in the same spirit as the ones proposed by Van Keilegom and Akritas (1999), who worked under the assumption of conditional independence between Y and C given X, and who estimated $F(\cdot|x)$ by means of the Beran estimator instead of the estimator defined in (2.1). Therefore, if we take $\phi_x(\cdot) = -\log(\cdot)$ in formula (2.1), the estimators $\widehat{m}(x)$ and $\widehat{\sigma}^2(x)$ reduce to the ones proposed in Van Keilegom and Akritas (1999).

The assumptions below are important for the establishment of the asymptotic results in Section 3. In addition to the distributions $F(\cdot|x)$, $G(\cdot|x)$ and $H(\cdot|x)$ already defined above, they concern the subdistribution $H^u(y|x) = P(T \le y, \Delta = 1|X = x)$ of the uncensored observations and the distribution $F_X(x) = P(X \le x)$ of the covariate. The probability density functions of the distribution functions defined above will be denoted by the corresponding lower case letters. Also, let $\widetilde{T}_x$ be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x \in R_X}(1 - H(\widetilde{T}_x|x)) > 0$.

For an arbitrary (sub)distribution function $L(y|x)$ we will use the notations $l(y|x) = L'(y|x) = \frac{\partial}{\partial y} L(y|x)$, $\dot{L}(y|x) = \frac{\partial}{\partial x} L(y|x)$ and similar notations will be used for higher order derivatives. (In the proofs, the function $L(y|x)$ of assumption $(A5)$ below will be either $H(y|x)$ or $H^u(y|x)$). We also use the notation $\overline{L}(y|x) = 1 - L(y|x)$ throughout the paper.

(A1) (i) The sequence h_n satisfies $nh_n^5(\log n)^{-1} = O(1)$ and $(nh_n)^{-1}\log n \to 0$.

(*ii*) The support R_X of X is a bounded interval in $\mathbb{R}$.

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

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

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

(A3) (*i*) The distribution F_X is three times continuously differentiable on the interior of R_X, 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 $\phi'_x(1) < 0$.

(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 $\sup_{x,y} |y^2\ddot{L}(y|x)| < \infty$.

(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

$$|H(y_2|x) - H(y_1|x)| \leq |M_1(y_2) - M_1(y_1)|,$$
$$|H^u(y_2|x) - H^u(y_1|x)| \leq |M_2(y_2) - M_2(y_1)|,$$
$$\left|\frac{\partial H(y_2|x)}{\partial x} - \frac{\partial H(y_1|x)}{\partial x}\right| \leq |M_3(y_2) - M_3(y_1)|,$$
$$\left|\frac{\partial H^u(y_2|x)}{\partial x} - \frac{\partial H^u(y_1|x)}{\partial x}\right| \leq |M_4(y_2) - M_4(y_1)|,$$

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

5

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.

The following functions will also be needed in the sequel.

$$
\begin{aligned}
g\big(T, \Delta, y | x\big) &= \frac{-1}{\phi'_x\big(\overline{F}(y|x)\big)} \bigg\{ \int_{-\infty}^{y} \phi''_x\big(\overline{H}(s|x)\big)\big[I(T \leq s) - H(s|x)\big]dH^u(s|x) \\
&\quad -\phi'_x\big(\overline{H}(y|x)\big)\big[I(T \leq y, \Delta = 1) - H^u(y|x)\big] \\
&\quad -\int_{-\infty}^{y} \phi''_x\big(\overline{H}(s|x)\big)\big[I(T \leq s, \Delta = 1) - H^u(s|x)\big]dH(s|x) \bigg\}, \\
\eta\big(T, \Delta | x\big) &= \int_{-\infty}^{+\infty} J\big(F(y|x)\big)g\big(T, \Delta, y|x\big)dy, \\
\zeta\big(T, \Delta | x\big) &= \int_{-\infty}^{+\infty} J\big(F(y|x)\big)g\big(T, \Delta, y|x\big)\frac{y - m(x)}{\sigma(x)}dy.
\end{aligned}
$$

3 Asymptotic results

3.1 Asymptotic results for the estimator of $F(y|x)$

We start this section with some new results concerning the copula-graphic estimator $\widehat{F}(y|x)$. In particular, we will derive uniform consistency rates for $\widehat{F}(y|x)$, for its derivative $\dot{\widehat{F}}(y|x)$ and for the 'derivative' of order $1 + \delta$ defined by $(\dot{\widehat{F}}(y|x) - \dot{\widehat{F}}(y|x'))/(x - x')^{\delta}$. In addition we will also show an iid asymptotic representation for $\widehat{F}(y|x) - F(y|x)$, and establish the rate of convergence of the remainder term uniformly in x and y.

These results are useful for establishing similar results for the estimators of $m(x)$ and $\sigma^2(x)$ in the next subsection. In addition, they generalize previous results obtained by Braekers and Veraverbeke (2005), who focused attention on the estimator $\widehat{F}(y|x)$ itself and who established the rate of convergence for a fixed value of x under a fixed design setting. We refer to Subsection 3.2 for additional motivation for studying these results.

Proposition 3.1. *Assume (A1), (A3)(i), (A4)(i,iii), and H and H^u satisfy (A5)(i,iv,v). Then,*

$$
\sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} \left| \widehat{F}(y|x) - F(y|x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \quad a.s.
$$

Proposition 3.2. *Assume (A1), (A3)(i), (A4)(i,iii), $\dot{\phi}''_x(u)$ exists and is continuous in (x, u), and H and H^u satisfy (A5)(i,iv,v). Then,*

$$
\sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} \left| \dot{\widehat{F}}(y|x) - \dot{F}(y|x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \quad a.s.
$$

6

Proposition 3.3. *Assume (A1), (A3)(i), (A4), and H and H^u satisfy (A5)(i,iv,v). Then,*

$$\sup_{x,x'\in R_X} \sup_{y\leq \widetilde{T}_x \wedge \widetilde{T}_{x'}} \frac{\left| \dot{\widehat{F}}(y|x) - \dot{F}(y|x) - \dot{\widehat{F}}(y|x') + \dot{F}(y|x') \right|}{|x-x'|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log n)^{1/2}) \quad a.s.,$$

for any $0 < \delta < 1$.

Proposition 3.4. *Assume (A1), (A3)(i), (A4)(i,iii), (A6) and H and H^u satisfy (A5). Then,*

$$\widehat{F}(y|x) - F(y|x) = \frac{1}{nh_n f_X(x)} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) g(T_i, \Delta_i, y|x) + R_n(y|x),$$

where $\sup_{x\in R_X} \sup_{y\leq \widetilde{T}_x} |R_n(y|x)| = O\left((nh_n)^{-3/4}(\log n)^{3/4}\right)$ a.s.

Remark 3.1. *Without assumption (A4)(iii) all the results in this subsection remain valid if we replace $\sup_{y\leq \widetilde{T}_x}$ by $\sup_{\widetilde{t}_x \leq y \leq \widetilde{T}_x}$, where $\widetilde{t}_x$ is chosen such that $\sup_{x\in R_X}(1 - H(\widetilde{t}_x|x)) < 1$. Also note that in the next subsection we do not need to assume (A4)(iii), since by assumption (A2)(i) the score function $J(s)$ equals zero for s close to 0 or 1.*

3.2 Asymptotic results for the estimators of $m(x)$ and $\sigma^2(x)$

In this section we will derive some asymptotic properties of the location estimator $\widehat{m}(x)$ and the scale estimator $\widehat{\sigma}(x)$ defined in (2.2). In particular, we will show the uniform consistency rates of $\widehat{m}(x)$, of $\widehat{m}'(x)$ and of $(\widehat{m}'(x) - \widehat{m}'(x'))/(x-x')^\delta$, and of their analogues for $\widehat{\sigma}$. We will also prove an asymptotic representation for $\widehat{m}(x)$ and $\widehat{\sigma}(x)$. All results will be obtained uniformly in $x \in R_X$.

These results are important, since they show that with probability tending to one, the estimators $\widehat{m}$ and $\widehat{\sigma}$ belong to the class $C_K^{1+\delta}(R_X)$ of differentiable functions $f : R_X \to \mathbb{R}$ for which

$$\|f\|_{1+\delta} = \max\left\{ \sup_x |f(x)|, \sup_x |f'(x)| \right\} + \sup_{x,x'} \frac{|f'(x) - f'(x')|}{|x-x'|^\delta}$$

is bounded by $K < \infty$. This class plays a major role in empirical process theory, since its covering and bracketing number enjoy nice properties, which are needed to show that the class is Donsker (see Van der Vaart and Wellner (1996) for more details).

The results have additional value as they are necessary if one is interested in exploring the asymptotic properties of an estimator of the error distribution in a nonparametric location-scale regression model of the form $Y = m(X) + \sigma(X)\epsilon$ with ϵ and X independent. This is currently under investigation by the authors, and a research paper is in preparation that will soon be submitted for publication.

Proposition 3.5. *Assume (A1)–(A3), (A4)(i), and H and H^u satisfy (A5)(i,iv,v). Then,*

(a) $\sup_{x \in R_X} \left| \widehat{m}(x) - m(x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \quad a.s.$

(b) $\sup_{x \in R_X} \left| \widehat{\sigma}(x) - \sigma(x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \quad a.s.$

Proposition 3.6. *Assume (A1)–(A3), (A4)(i), $\dot{\phi}_x''(u)$ exists and is continuous in (x, u), and H and H^u satisfy (A5)(i,iv,v). Then,*

(a) $\sup_{x \in R_X} \left| \widehat{m}'(x) - m'(x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \quad a.s.$

(b) $\sup_{x \in R_X} \left| \widehat{\sigma}'(x) - \sigma'(x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \quad a.s.$

Proposition 3.7. *Assume (A1)–(A3), (A4)(i,ii) and H and H^u satisfy (A5)(i,iv,v). Then,*

(a) $\displaystyle\sup_{x,x' \in R_X} \frac{\left| \widehat{m}'(x) - m'(x) - \widehat{m}'(x') + m'(x') \right|}{|x - x'|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log n)^{1/2}) \quad a.s.$

(b) $\displaystyle\sup_{x,x' \in R_X} \frac{\left| \widehat{\sigma}'(x) - \sigma'(x) - \widehat{\sigma}'(x') + \sigma'(x') \right|}{|x - x'|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log n)^{1/2}) \quad a.s.,$

for any $0 < \delta < 1$.

Proposition 3.8. *Assume (A1)–(A3), (A4)(i), (A6) and H and H^u satisfy (A5). Then,*

(a)
$$\widehat{m}(x) - m(x) = \frac{-1}{nh_n f_X(x)} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \eta\left(T_i, \Delta_i | x\right) + R_{n1}(x)$$

(b)
$$\widehat{\sigma}(x) - \sigma(x) = \frac{-1}{nh_n f_X(x)} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \zeta\left(T_i, \Delta_i | x\right) + R_{n2}(x),$$

where $\sup_{x \in R_X} |R_{nj}(x)| = O\left((nh_n)^{-3/4}(\log n)^{3/4}\right)$ a.s. $(j = 1, 2)$.

8

The latter result is important for obtaining e.g. the asymptotic normality of $(nh_n)^{1/2}\big(\widehat{m}(x) - m(x)\big)$ and $(nh_n)^{1/2}\big(\widehat{\sigma}(x) - \sigma(x)\big)$ for a fixed x in R_X, from which pointwise confidence intervals for $m(x)$ and $\sigma(x)$ can be obtained, provided appropriate estimators of the variance are available. This variance basically relies on a good estimator of the function $g(T, \Delta, y|x)$, which can be obtained by replacing $H(y|x), H^u(y|x)$ and $F(y|x)$ by their corresponding estimators. An alternative could be to use an appropriate bootstrap, but that is out of the scope of this paper.

Note however that these confidence intervals will not allow us to take the uncertainty about the copula model into account, since the whole asymptotic theory relies on the assumption that the copula is known and correctly specified. On the other hand, the confidence intervals allow to have an idea of the uncertainty coming from the other components of the model.

In addition, the latter result can also be used as a first big step for constructing asymptotic confidence bands for the true unknown functions $m(\cdot)$ and $\sigma(\cdot)$, similarly as was done in Claeskens and Van Keilegom (2003) in the context of multiparameter local likelihood estimating equations, or for testing hypotheses concerning these functions.

4   Simulations

In this section we will illustrate the finite sample behavior of our estimators $\widehat{m}(x)$ and $\widehat{\sigma}(x)$ by means of Monte Carlo simulations. We will compare our estimators with the ones proposed by Van Keilegom and Akritas (1999). These estimators are based on the assumption that Y and C are conditionally independent, and are defined as follows :

$$\widetilde{m}(x) = \int_0^1 \widetilde{F}^{-1}(s|x)J(s)ds \quad \text{and} \quad \widetilde{\sigma}^2(x) = \int_0^1 \big[\widetilde{F}^{-1}(s|x) - \widetilde{m}(x)\big]^2 J(s)ds, \quad (4.1)$$

where $\widetilde{F}(\cdot|x)$ is the Beran estimator :

$$\widetilde{F}(y|x) = 1 - \prod_{T_i \le y, \Delta_i = 1} \left\{ 1 - \frac{W_{ni}(x, h_n)}{\sum_{j=1}^n I(T_j \ge T_i)W_{nj}(x, h_n)} \right\}.$$

We expect that when the conditional dependence between Y and C is strong, the estimators $\widetilde{m}(x)$ and $\widetilde{\sigma}(x)$ will behave poorly compared to the new estimators $\widehat{m}(x)$ and $\widehat{\sigma}(x)$, that take this dependence into account. The results will therefore illustrate the importance of correctly specifying the dependence between Y and C. We measure the performance of the estimators by means of their mean squared error (MSE) and by means of 90% confidence

intervals. The simulations are carried out for samples of size $n = 100$, $n = 200$ and $n = 400$, and the results are obtained by using 2000 Monte Carlo simulations.

We generate i.i.d. data from the following regression model :

$$Y = 6(X - 0.5)^2 + 0.5\varepsilon_1, \tag{4.2}$$

where X has a uniform distribution on $[0, 1]$ and the error ε_1 has a standard normal distribution and is independent of X. The censoring variable C satisfies $C = \alpha_1(X-0.5)^2 + \alpha_2 + 0.5\varepsilon_2$ for certain choices of α_1 and α_2, where ε_2 is standard normal and independent of X. 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%. We further assume that the dependence between Y and C given $X = x$ (i.e. the dependence between ε_1 and ε_2 given $X = x$) is described via a Gumbel copula :

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

where $\gamma(x) = \max(5 - 5x, 1)$. This means that the corresponding Archimedean copula generator equals $\phi_x(u) = -(\log u)^{\gamma(x)}$. With this construction the conditional dependence between Y and C given $X = x$ decreases from strong positive dependence to complete independence as x goes from 0 to 0.8, and it continues to be independent on $[0.8, 1]$. Note that Kendall tau's coefficient decreases from 0.8 to 0.

Note that, in the above setting, an equivalent way of writing the estimator $\widehat{\sigma}^2(x)$ is

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

where $\widehat{F}_b(y|x) = \min\{\widehat{F}(y|x), b\}$ and b is any value larger than $\sup\{s : J(s) \neq 0\}$, and this is close to

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

It can be easily seen that both estimators are asymptotically equivalent under certain conditions on the bandwidth. In the sequel we work with the latter estimator, since simulations showed that it outperforms somewhat the former (which is not too surprising given that $Y_i - m(X_i)$ is a proper 'error', whereas $Y_i - m(x)$ is not). The same applies to the estimator $\widetilde{\sigma}(x)$, for which we also work with the asymptotically equivalent variant.

For the weights that appear in our estimators $\widehat{m}(x)$ and $\widehat{\sigma}(x)$, and also in the estimators $\widetilde{m}(x)$ and $\widetilde{\sigma}(x)$, we use the kernel function $K(u) = (15/16)(1 - u^2)^2 I(|u| \leq 1)$, and we work with the score function $J(s) = b^{-1}I(0 \leq s \leq b)$. 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.

For each of the four estimators $\widehat{m}$, $\widehat{\sigma}$, $\widetilde{m}$ and $\widetilde{\sigma}$ we use a different bandwidth h_n. In the first step, to select the bandwidth for $\widehat{m}$, we minimize the integrated mean squared error $IMSE(\widehat{m}) = \int_{0.2}^{0.8} E[\widehat{m}(x) - m(x)]^2 dF_X(x)$ over a grid of 15 equidistant possible values of h_n between 0.05 and 0.40 (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). To calculate this $IMSE(\widehat{m})$, we use 2000 simulated data sets. For each simulated data set, we compute the integrated squared error $ISE(\widehat{m}) = \int_{0.2}^{0.8} [\widehat{m}(x) - m(x)]^2 dF_X(x)$, and we approximate $IMSE(\widehat{m})$ by taking the average over these 2000 values of $ISE(\widehat{m})$. In the second step, to select the bandwidth for $\widehat{\sigma}$ we first write

$$\sum_{i=1}^{n} \big[Y_i - \widehat{m}_1(X_i)\big]^2 \big[\widehat{F}_{1b}(Y_i|x) - \widehat{F}_{1b}(Y_i^-|x)\big] J(\widehat{F}_{2b}(Y_i|X_i)),$$

where $\widehat{m}_1$ and $\widehat{F}_{1b}$ are the estimators based on the optimal bandwidth chosen in the first step, and $\widehat{F}_{2b}$ is based on a second bandwidth. Now, we select this second bandwidth by minimizing the empirical $IMSE(\widehat{\sigma})$ (which is estimated in the same way as in the first step) over the grid $0.05, 0.1, 0.2, ..., 0.9$. The bandwidths for the estimators $\widetilde{m}$ and $\widetilde{\sigma}$ are chosen in an analogous way.

Note that we prefer to select the bandwidth for $\widehat{m}$, $\widehat{\sigma}$, $\widetilde{m}$ and $\widetilde{\sigma}$ by directly minimizing the $IMSE$ of these estimators, instead of minimizing the $IMSE$ of the estimators $\widehat{F}(\cdot|x)$ and $\widetilde{F}(\cdot|x)$. This ensures that we control the quality of these estimators instead of controlling the quality of the intermediate estimators $\widehat{F}(\cdot|x)$ and $\widetilde{F}(\cdot|x)$, which are of secondary importance in our estimation procedure.

The practical performance of the above bandwidth selection procedure is illustrated in Figures 1 and 2. The figures show the $IMSE$ as a function of the bandwidth for each considered estimator and each sample size. We see that the $IMSE$ for $\widehat{m}$ and $\widetilde{m}$ have a convexly shaped curve and that the optimal bandwidth decreases with the sample size. We also see that the bandwidths for the estimators of σ are much larger than those for the corresponding estimators of m, which is natural since σ is constant in our model. This highlights the importance of choosing different bandwidths when estimating m and σ. Finally, we notice that the $IMSE$-curves for $\widetilde{m}$ and $\widetilde{\sigma}$ are quite a bit higher than the corresponding curves for $\widehat{m}$ and $\widehat{\sigma}$, and also the minimal values of these curves differ quite a lot, which suggests that $\widehat{m}$ and $\widehat{\sigma}$ are, globally speaking, behaving better than $\widetilde{m}$ and $\widetilde{\sigma}$.

Figures 3 and 4 show the MSE of the estimators $\widehat{m}(x)$, $\widehat{\sigma}(x)$, $\widetilde{m}(x)$ and $\widetilde{\sigma}(x)$ for the bandwidths that minimize the corresponding $IMSE$ of these four estimators. The bandwidths that minimize the $IMSE$ corresponding to $n = 100, 200$ and 400 are $0.200, 0.150$ and 0.125 for $\widetilde{m}$, and $0.250, 0.225$ and 0.175 for $\widehat{m}$, respectively, while the optimal bandwidths
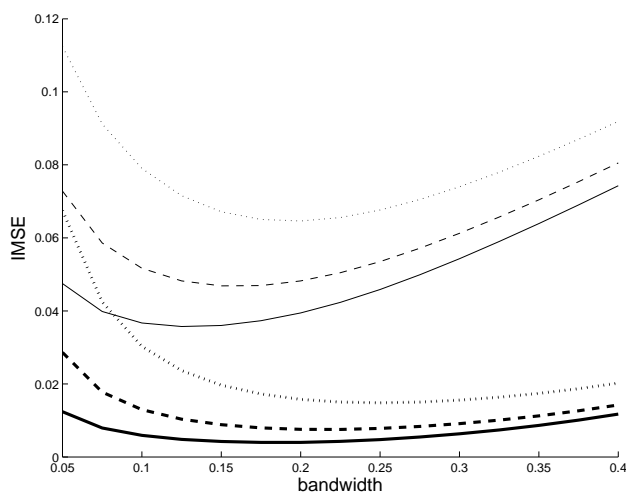
11

Figure 1: $IMSE(\widehat{m})$ (thick lines) and $IMSE(\widetilde{m})$ (thin lines) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).
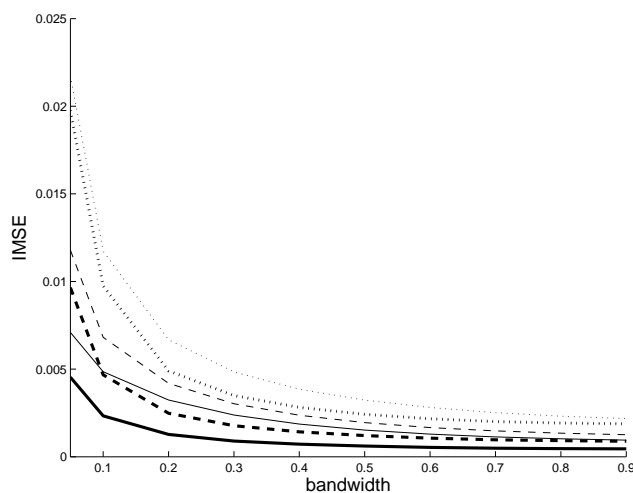

Figure 2: $IMSE(\widehat{\sigma})$ (thick lines) and $IMSE(\widetilde{\sigma})$ (thin lines) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

for both $\widetilde{\sigma}$ and $\widehat{\sigma}$ are 0.9 independently of n. As we expected, the new estimators $\widehat{m}(x)$ and $\widehat{\sigma}(x)$ outperform the estimators $\widetilde{m}(x)$ and $\widetilde{\sigma}(x)$, since the latter are incorrectly assuming that Y and C are independent given $X = x$ for all x, whereas this is only true for $0.8 \leq x \leq 1$. We see how the difference between the estimators becomes larger as x decreases, i.e. as the conditional dependence between Y and C given $X = x$ becomes stronger. The ratio of the two MSE's can be as high as 10 in case of the location function m and up to 3 in case of the scale function σ. Note that for $x = 0.8$, the two estimators are not equal in Figures 3 and 4, although $\widehat{m}(x)$ and $\widehat{\sigma}(x)$ reduce to $\widetilde{m}(x)$ and $\widetilde{\sigma}(x)$, respectively, when Y and C are
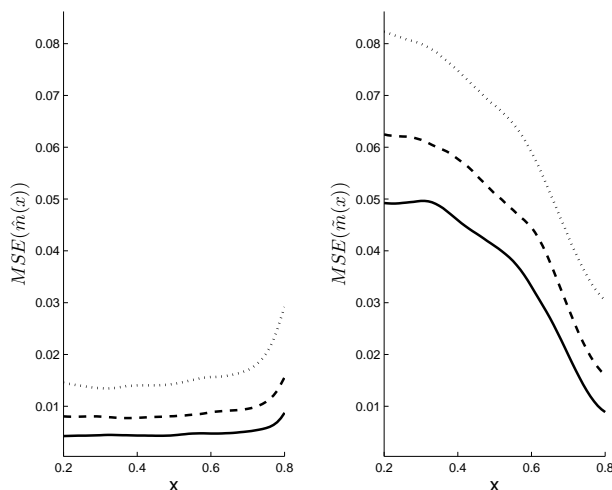
12

Figure 3: MSE of $\widehat{m}(x)$ (left panel) and $\widetilde{m}(x)$ (right panel) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).
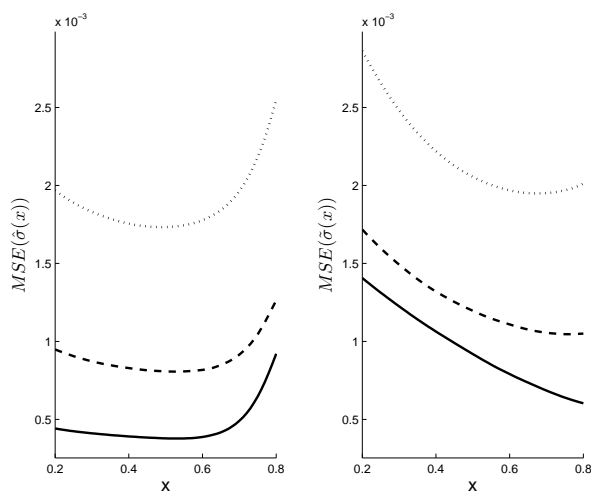

Figure 4: MSE of $\widehat{\sigma}(x)$ (left panel) and $\widetilde{\sigma}(x)$ (right panel) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

conditionally independent. This is because the bandwidths used to calculate the estimators are not the same, and are in fact determined by the behavior of the estimators for all values of the covariate.

Figure 5 shows the quantiles of order 0.05 and 0.95 of the distribution of $\widehat{m}(x) - m(x)$ and $\widetilde{m}(x) - m(x)$ respectively. The reason why we consider the 'standardized' quantity $\widehat{m}(x) - m(x)$ instead of the more natural quantity $\widehat{m}(x)$, is that in the non-standardized graph the scale of the vertical axis is much wider, and the curves are therefore very close together and hard to distinguish. We see that as we are moving to the area of stronger

dependence (small values of x) the new estimator $\widehat{m}(x)$ continues to behave well, while the estimator $\widetilde{m}(x)$ becomes increasingly biased and also slightly more variable. The same can be said for the estimators of $\sigma(x)$ shown in Figure 6 (which we do not standardize since the true σ-curve is already constant). Again, the figure shows that the new estimator is only slightly biased, whereas the estimator $\widetilde{\sigma}(x)$ shows consistent biased behavior and slightly larger variance.
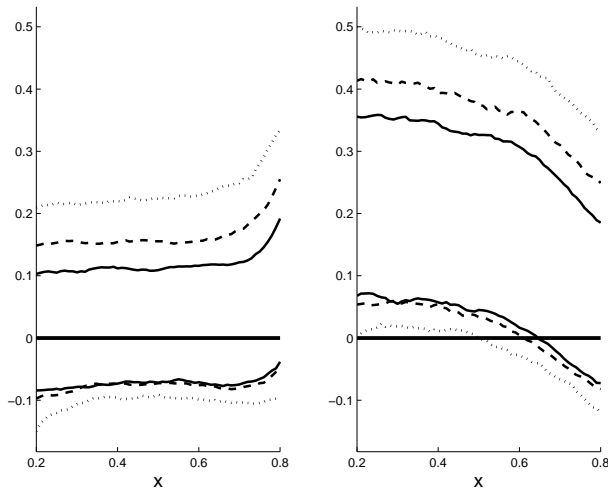

Figure 5: Quantiles of order 0.05 and 0.95 of the distribution of $\widehat{m}(x) - m(x)$ (left panel) and $\widetilde{m}(x) - m(x)$ (right panel) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

5   Example

In this section we will illustrate our estimation method via the analysis of the bone marrow transplantation data, which are described in Klein and Moeschberger (1997). The data are collected during a study in which 137 patients are followed in their recovery from acute leukemia after bone marrow transplantation. We are interested in the disease-free survival time Y, i.e. the time until a patient has a relapse of leukemia. However, patients can be censored by two possible events : disease-free death or disease-free and alive at the end of the study. The censoring time C is the time until the first of these two events takes place. It seems natural to believe that the time until relapse Y depends on the age X of the patient at transplantation, and on the time until disease-free death (for a given age). This indirectly implies that Y and C are dependent for a given covariate X. In Figure 7, we show a scatter plot of age versus disease-free survival time, where we distinguish between non-censored
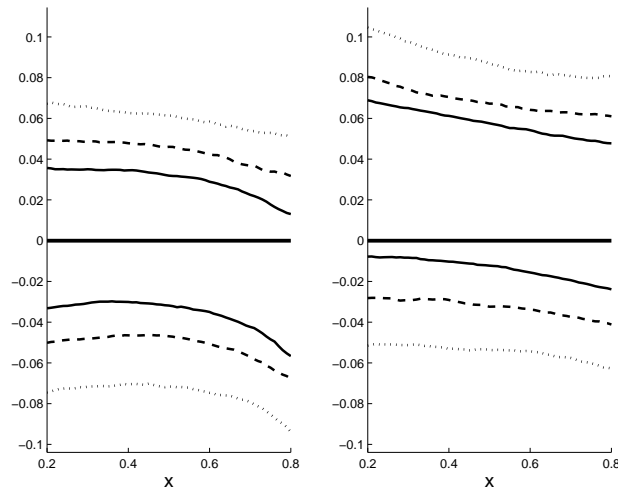
Figure 6: Quantiles of order 0.05 and 0.95 of the distribution of $\widehat{\sigma}(x)$ (left panel) and $\widetilde{\sigma}(x)$ (right panel) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

(relapsed) and censored patients. We note that the censoring rate is as high as 69%, caused by the rather short length of the study (less than 8 years).
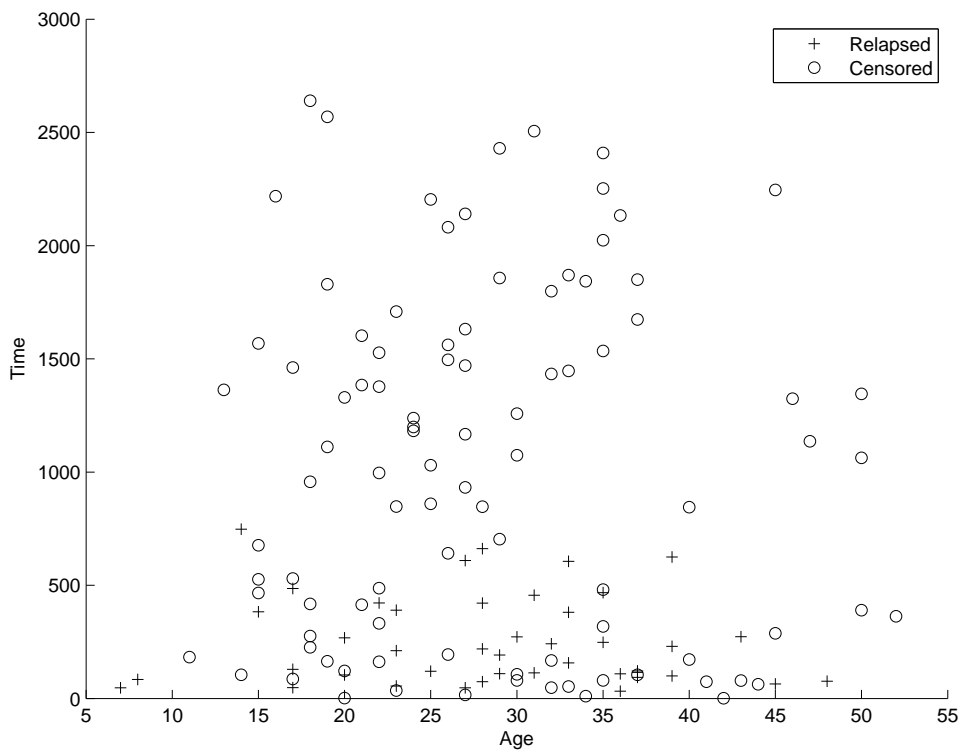

Figure 7: Scatterplot of age versus disease-free survival time.

15

First, from these data, for every fixed covariate x ranging between 20 and 40 (the area containing most of the data) we will estimate the average of the lower 30% of relapse times. Also we will estimate the standard deviation of the lower 30% of relapse times. We have to restrict to 30%, because of the rather high proportion of censoring. In fact, for some values of x, the estimator of the conditional distribution function of Y given x is only consistent up to the 0.3-th quantile. The score function corresponding to this location and scale functional is $J(s) = \frac{10}{3} I_{[0,0.3)}(s)$ (see (1.2)). To estimate these functionals we use the biweight kernel function $K(u) = (15/16)(1 - u^2)^2 I(|u| \leq 1)$.

Second, it is important to note that our estimation procedure assumes that the dependence structure between the survival time Y and the censoring time C is completely known. Therefore, in a real data application we have to select the dependence structure based on external information. For instance, we can model the dependence based on an expert's opinion or estimate it from additional data. To capture the expert's opinion on the strength and direction of the dependence between Y and C we will use Kendall's tau, which is defined as $\tau(x) = 1 + 4 \int_0^1 (\phi_x(t) \,/\, \phi_x'(t)) \, dt$ (Nelsen (2006)) and has a range from -1 to 1. The association gets stronger as $\tau(x)$ gets further away from zero, while the concordance/discordance is determined by the sign.
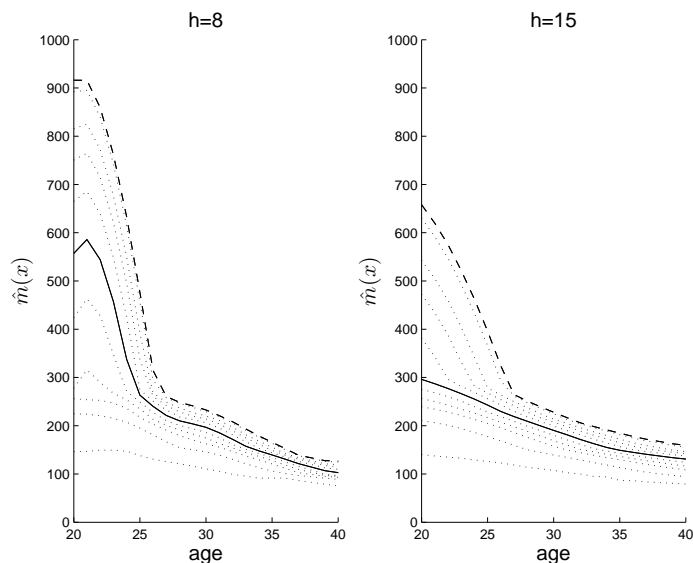

Figure 8: Estimator of the average of the lower 30% of relapse times for several choices of $\phi_x(t)$, not depending on x: independence (solid curve), Fréchet-Hoeffding lower bound (long dashed curve) and Frank family for ϕ_x corresponding to $\tau(x) = -0.5, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.5$ and 0.9 (dotted curves going from highest to lowest, respectively).

Figure 9: Estimator of the standard deviation of the lower 30% of relapse times for several choices of $\phi_x(t)$, not depending on x: independence (solid curve), Fréchet-Hoeffding lower bound (long dashed curve) and Frank family for ϕ_x corresponding to $\tau(x) = -0.5, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.5$ and 0.9 (dotted curves going from highest to lowest, respectively).
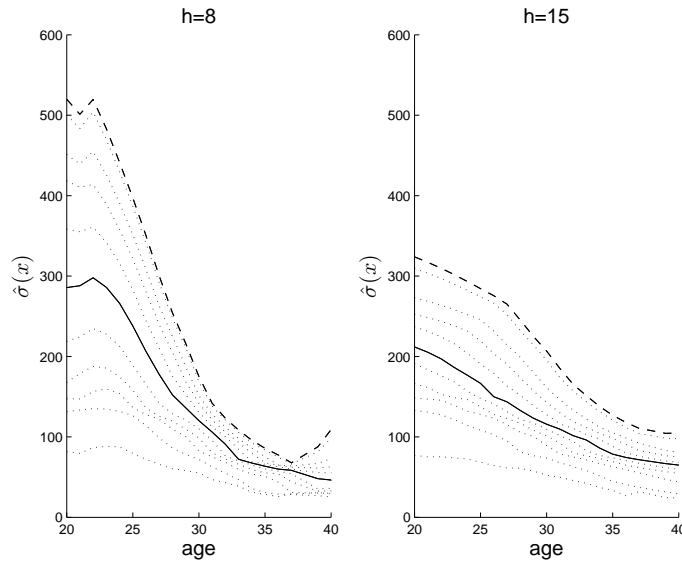
In Figure 8 we show the estimator $\widehat{m}(x)$ for $h = 8$ and $h = 15$, while Figure 9 shows the estimator $\widehat{\sigma}(x)$ for $h = 8$ and $h = 15$. Each of the plots contains estimators constructed for different choices of the copula generator $\phi_x(t)$: the generator that assumes conditional independence between Y and C ($\phi_x(t) = -\log(t)$ and $\tau(x) = 0$), the Fréchet-Hoeffding lower bound ($\phi_x(t) = 1 - t$ and $\tau(x) = -1$), which assumes that Y and C are discordant, and the generators from the Frank family ($\phi_x(t) = -\log(\exp\{-t\theta\} - 1) + \log(\exp\{-\theta\} - 1)$) with θ taking values corresponding to $\tau(x) = -0.5, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.5$ and 0.9.

From Figure 8 we can see that, for all choices of the copula and the bandwidth, the average of the lower 30% of relapse times is decreasing as age grows from 20 to 40. This means that for the 30% of worst recipients, older people have a shorter time to relapse than younger people. From Figure 9 we see that, again in all cases, the standard deviation of the lower 30% of relapse times decreases as age grows from 20 to 40. This indicates that there is more uncertainty in how younger recipients respond to transplantation.

Next, the figures also show that for the Frank copula, the value of Kendall's tau is not influencing the overall pattern of the functions $\widehat{m}$ and $\widehat{\sigma}$ as long as the dependence between Y and C does not change in an extreme way with X. Hence, the overall shape of the curves is relatively robust to the value of Kendall's tau. On the other hand, the value of the curves

17

Figure 10: Estimator of the average of the lower 30% of relapse times for several choices of $\phi_x(t)$, not depending on x: Clayton family (long dashed curves), Frank family (dotted curves) and Gumbel family (solid curves), for ϕ_x corresponding to $\tau(x) = 0.1, 0.5$ and 0.8 (curves going from highest to lowest, respectively).
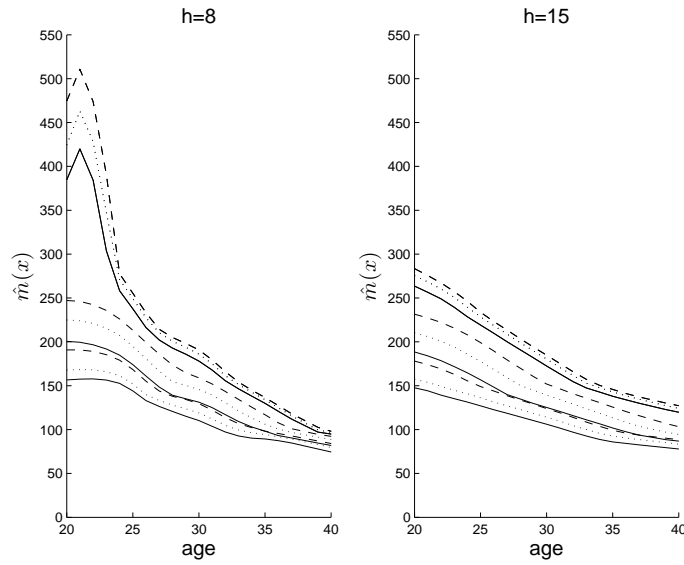

Figure 11: Estimator of the standard deviation of the lower 30% of relapse times for several choices of $\phi_x(t)$, not depending on x: Clayton family (long dashed curves), Frank family (dotted curves) and Gumbel family (solid curves), for ϕ_x corresponding to $\tau(x) = 0.1, 0.5$ and 0.8 (curves going from highest to lowest, respectively).
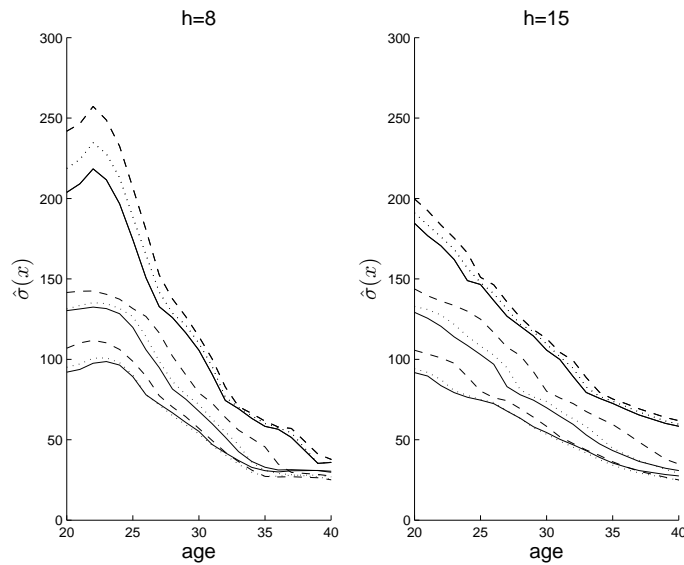
at a specific point is heavily influenced by the value of Kendall's tau, which illustrates the importance of having accurate external information regarding the strength and direction of the dependence between Y and C.

In practice, it might not be an easy task to select the best copula family based on external information. Therefore, we want to investigate the robustness of the estimators $\widehat{m}$ and $\widehat{\sigma}$ to the choice of the copula family under the same value of Kendall's tau. In Figures 10 and 11 we show again the estimators $\widehat{m}(x)$ and $\widehat{\sigma}(x)$, respectively, for $h = 8$ and $h = 15$ and several choices of $\phi_x(t)$ (not depending on x) : the Clayton family ($\phi_x(t) = \frac{1}{\theta}(t^{-\theta} - 1)$), the Frank family and the Gumbel family ($\phi_x(t) = -(\log t)^\theta$), for θ corresponding to $\tau = 0.1, 0.5$ and 0.8 (representing small, significant and very strong positive dependence, respectively). In this investigation we restrict ourselves to positive values of τ, since the Gumbel family is only able to produce positive correlation. Note that the dependence structure between Y and C differs a lot from one copula family to another : the Clayton family gives lots of weight to the left lower corner (or left tail) of the unit square, the Gumbel family to the left and the right tail, and the Frank family is distributed more homogeneously along all points close to the bisector.

Both Figures 10 and 11 show that, for a fixed τ, the difference between the estimators caused by different choices of the copula family is not exceeding 25%. All figures strongly indicate the general tendency of the curves, independently of the choice of the copula family, as long as the dependence between Y and C does not change in an extreme way with X. (Note that even though the copula generators ϕ_x used in the figures do not depend on x, one can easily see how the estimator would behave if ϕ_x would change with x.) Hence, the figures show that for a fixed value of Kendall's tau, our estimators are quite robust to misspecification of the copula family.

To conclude, the choice of the copula family (Gumbel, Clayton, ...) is less important than the correct specification of the strength of the dependence between Y and C (via e.g. Kendall's tau), and the way this dependence depends on X. In many situations in practice, we do have an idea of the strength of this dependence, e.g. we know whether the dependence is negative or positive and whether it is strong or rather weak. This can help in choosing an appropriate copula function.

A   Appendix : Proofs

Proof of Proposition 3.1. Let x be an arbitrary value in R_X and let $y \leq \widetilde{T}_x$. Throughout this proof we will use that any random process $\alpha_n(x, y)$ that lies between $\overline{H}(y|x)$ and $\widehat{\overline{H}}(y|x)$

for n large enough, can be a.s. bounded from below for n large enough :

$$\alpha_n(x,y) > \gamma \text{ on } \{(x,y) : x \in R_X, \ y \le \widetilde{T}_x\} \tag{A.1}$$

for some $\gamma > 0$. This follows from the uniform consistency of $\widehat{H}(y|x)$ (see Proposition 1 in Akritas and Van Keilegom (2001)) and from the definition of $\widetilde{T}_x$.

Using similar calculations as in the proof of Theorem 1 in Braekers and Veraverbeke (2005), we can write

$$
\begin{aligned}
&\widehat{F}(y|x) - F(y|x) \\
&= \Big[-\phi_x^{-1} \Big\{ - \sum_{T_i \le y, \Delta_i = 1} \Big[\phi_x\big(\widehat{\overline{H}}(T_i^-|x)\big) - \phi_x\big(\widehat{\overline{H}}(T_i^-|x) - W_{ni}(x,h_n)\big) \Big] \Big\} \\
&\quad + \phi_x^{-1} \Big\{ - \sum_{T_i \le y, \Delta_i = 1} \phi_x'\big(\widehat{\overline{H}}(T_i^-|x)\big) W_{ni}(x,h_n) \Big\} \Big] \\
&\quad - \Big[\phi_x^{-1}\Big\{ -\int_{-\infty}^{y} \phi_x'\big(\widehat{\overline{H}}(s^-|x)\big) d\widehat{H}^u(s|x) \Big\} - \phi_x^{-1}\Big\{ -\int_{-\infty}^{y} \phi_x'\big(\overline{H}(s|x)\big) dH^u(s|x) \Big\} \Big] \\
&:= Q_1(x,y) + Q_2(x,y).
\end{aligned}
$$

Next, we will calculate the order of convergence of each of these terms. Starting with the second term, we apply the mean value theorem and obtain

$$
\begin{aligned}
&Q_2(x,y) \tag{A.2} \\
&= \frac{-1}{\phi_x'\big(\phi_x^{-1}(\alpha_1(x,y))\big)} \Big[-\int_{-\infty}^{y} \phi_x'\big(\widehat{\overline{H}}(s^-|x)\big) d\widehat{H}^u(s|x) + \int_{-\infty}^{y} \phi_x'\big(\overline{H}(s|x)\big) dH^u(s|x) \Big],
\end{aligned}
$$

where $\alpha_1(x,y)$ lies between $-\int_{-\infty}^{y} \phi_x'\big(\widehat{\overline{H}}(s^-|x)\big) d\widehat{H}^u(s|x)$ and $-\int_{-\infty}^{y} \phi_x'\big(\overline{H}(s|x)\big) dH^u(s|x)$. From (A4)(iii) we know that the first factor on the right hand side of (A.2) is uniformly bounded. Now, by adding and subtracting terms, we further obtain

$$Q_2(x,y) = O(1)\Big[Q_2^{(1)}(x,y) + Q_2^{(2)}(x,y) \Big],$$

where

$$
\begin{aligned}
Q_2^{(1)}(x,y) &= -\int_{-\infty}^{y} \Big[\phi_x'\big(\widehat{\overline{H}}(s^-|x)\big) - \phi_x'\big(\overline{H}(s|x)\big) \Big] d\widehat{H}^u(s|x), \\
Q_2^{(2)}(x,y) &= -\int_{-\infty}^{y} \phi_x'\big(\overline{H}(s|x)\big) d\Big[\widehat{H}^u(s|x) - H^u(s|x) \Big].
\end{aligned}
$$

Applying the mean value theorem on the term $Q_2^{(1)}(x,y)$ we get:

$$Q_2^{(1)}(x,y) = -\int_{-\infty}^{y} \phi_x''\big(\alpha_2(x,s)\big) \Big[\widehat{\overline{H}}(s^-|x) - \overline{H}(s|x) \Big] d\widehat{H}^u(s|x),$$

20

where $\alpha_2(x,s)$ lies between $\widehat{\overline{H}}(s^-|x)$ and $\overline{H}(s|x)$. This gives us that

$$\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|Q_2^{(1)}(x,y)\right| \leq \sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|\phi_x''(\alpha_2(x,y))\right|\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|\widehat{H}(y^-|x)-H(y|x)\right|. \quad\text{(A.3)}$$

From Proposition 1 in Akritas and Van Keilegom (2001), we have that

$$\sup_{x,y}\left|\widehat{H}(y^-|x)-H(y|x)\right| = O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \text{ a.s.}$$

Hence, applying (A.1) on the first supremum at the right hand side of (A.3) yields that

$$\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|Q_2^{(1)}(x,y)\right| = O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \text{ a.s.} \quad\text{(A.4)}$$

For $Q_2^{(2)}(x,y)$ we integrate by parts, and using similar calculations as for $Q_2^{(1)}(x,y)$, we easily obtain that

$$
\begin{aligned}
\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|Q_2^{(2)}(x,y)\right| &\leq \sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|\phi_x'(\overline{H}(y|x))\right|\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|\widehat{H}^u(y|x)-H^u(y|x)\right| \\
&\quad + \sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|\phi_x''(\overline{H}(y|x))\right|\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|\widehat{H}^u(y|x)-H^u(y|x)\right| \\
&= O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \text{ a.s.} \quad\text{(A.5)}
\end{aligned}
$$

Now, combining results (A.4) and (A.5), yields

$$\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|Q_2(x,y)\right| = O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \text{ a.s.} \quad\text{(A.6)}$$

For $Q_1(x,y)$, we repeatedly apply the mean value theorem to get

$$
\begin{aligned}
Q_1(x,y) &= \frac{-1}{\phi_x'\big(\phi_x^{-1}(\alpha_3(x,y))\big)}\left\{-\sum_{T_i\leq y,\Delta_i=1}\phi_x\big(\widehat{\overline{H}}(T_i^-|x)\big)-\phi_x\big(\widehat{\overline{H}}(T_i^-|x)-W_{ni}(x,h_n)\big)\right. \\
&\quad \left.+\sum_{T_i\leq y,\Delta_i=1}\phi_x'\big(\widehat{\overline{H}}(T_i^-|x)\big)W_{ni}(x,h_n)\right\} \\
&= \frac{1}{2\phi_x'\big(\phi_x^{-1}(\alpha_3(x,y))\big)}\sum_{T_i\leq y,\Delta_i=1}\phi_x''\big(\alpha_4(x,i)\big)W_{ni}^2(x,h_n),
\end{aligned}
$$

where $\alpha_3(x,y)$ lies between $-\sum_{T_i\leq y,\Delta_i=1}\left[\phi_x(\widehat{\overline{H}}(T_i^-|x))-\phi_x(\widehat{\overline{H}}(T_i^-|x)-W_{ni}(x,h_n))\right]$ and $-\sum_{T_i\leq y,\Delta_i=1}\phi_x'\big(\widehat{\overline{H}}(T_i^-|x)\big)W_{ni}(x,h_n)$, and $\alpha_4(x,i)$ lies between $\widehat{\overline{H}}(T_i^-|x)$ and $\widehat{\overline{H}}(T_i^-|x)-W_{ni}(x,h_n)$. This leads to

$$
\begin{aligned}
&\sup_{x\in R_X}\sup_{y\leq\widetilde{T}_x}\left|Q_1(x,y)\right| \\
&\leq \frac{1}{2}\left[\inf_{x\in R_X}\inf_{y\leq\widetilde{T}_x}\left|\phi_x'\big(\phi_x^{-1}(\alpha_3(x,y))\big)\right|\right]^{-1}\sup_{x\in R_X,i=1,\dots n}\left|\phi_x''(\alpha_4(x,i))\right|\sum_{i=1}^n W_{ni}^2(x,h_n).
\end{aligned}
$$

21

The infimum is bounded from below because of (A4)(iii) and the supremum is bounded thanks to (A4) and equation (A.1). On the other hand, $\sum_{i=1}^{n} W_{ni}^2(x, h_n)$ is of order $O((nh_n)^{-1})$ a.s., since we know from standard kernel smoothing theory that $\sup_x \max_{i=1,\dots n} W_{ni}(x, h_n) = O((nh_n)^{-1})$ a.s. It now follows that

$$\sup_{x \in R_X} \sup_{y \leq \tilde{T}_x} \left| Q_1(x, y) \right| = O\left((nh_n)^{-1/2} (\log n)^{1/2} \right) \quad a.s. \tag{A.7}$$

The proof is finished by combining (A.7) and (A.6). $\qquad\square$

Proof of Proposition 3.2. Using the notation $U(x, y) := \phi_x(\overline{F}(y|x))$ and $U_n(x, y) := \phi_x(\widehat{\overline{F}}(y|x))$, we calculate

$$
\begin{aligned}
\frac{\partial}{\partial x}\left[U(x, y) - U_n(x, y) \right] &= -\phi_x'(\overline{F}(y|x))\dot{F}(y|x) + \dot{\phi}_x(\overline{F}(y|x)) \\
&\quad + \phi_x'(\widehat{\overline{F}}(y|x))\dot{\widehat{F}}(y|x) - \dot{\phi}_x(\widehat{\overline{F}}(y|x)) \\
&= \phi_x'(\overline{F}(y|x))\left[\dot{\widehat{F}}(y|x) - \dot{F}(y|x) \right] + \left[\phi_x'(\widehat{\overline{F}}(y|x)) - \phi_x'(\overline{F}(y|x)) \right]\dot{\widehat{F}}(y|x) \\
&\quad + \dot{\phi}_x'(\alpha(x, y))\left[\widehat{F}(y|x) - F(y|x) \right],
\end{aligned}
$$

where $\alpha(x, y)$ is between $\widehat{\overline{F}}(y|x)$ and $\overline{F}(y|x)$. From here we can write

$$
\begin{aligned}
\dot{\widehat{F}}(y|x) - \dot{F}(y|x) &= \frac{1}{\phi_x'(\overline{F}(y|x))}\Bigg\{ -\left[\phi_x'(\widehat{\overline{F}}(y|x)) - \phi_x'(\overline{F}(y|x)) \right]\dot{\widehat{F}}(y|x) \\
&\quad - \dot{\phi}_x'(\alpha(x, y))\left[\widehat{F}(y|x) - F(y|x) \right] + \frac{\partial}{\partial x}\left[U(x, y) - U_n(x, y) \right] \Bigg\}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\left[\dot{\widehat{F}}(y|x) - \dot{F}(y|x) \right]&\left[1 + \frac{\phi_x'(\widehat{\overline{F}}(y|x)) - \phi_x'(\overline{F}(y|x))}{\phi_x'(\overline{F}(y|x))} \right] \tag{A.8} \\
&= \frac{1}{\phi_x'(\overline{F}(y|x))}\Bigg\{ -\left[\phi_x'(\widehat{\overline{F}}(y|x)) - \phi_x'(\overline{F}(y|x)) \right]\dot{F}(y|x) \\
&\quad - \dot{\phi}_x'(\alpha(x, y))\left[\widehat{F}(y|x) - F(y|x) \right] + \frac{\partial}{\partial x}\left[U(x, y) - U_n(x, y) \right] \Bigg\}.
\end{aligned}
$$

Hence, it follows from Proposition 3.1 and the conditions on ϕ_x that the statement of the proposition follows provided we can show that

$$\sup_{x \in R_X} \sup_{y \leq \tilde{T}_x} \left| \frac{\partial}{\partial x}[U_n(x, y) - U(x, y)] \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \quad a.s.$$

22

Consider

$$
\begin{aligned}
\frac{\partial}{\partial x}\big[U(x,y)-U_n(x,y)\big] &= \frac{\partial}{\partial x}\Bigg[-\sum_{T_i\le y,\Delta_i=1}\Big\{\phi_x\big(\widehat{\overline{H}}(T_i^-|x)\big)-\phi_x\big(\widehat{\overline{H}}(T_i^-|x)-W_{ni}(x,h_n)\big)\Big\} \\
&\qquad + \sum_{T_i\le y,\Delta_i=1}\phi_x'\big(\widehat{\overline{H}}(T_i|x)\big)W_{ni}(x,h_n)\Bigg] \\
&\quad -\frac{\partial}{\partial x}\Bigg[\int_{-\infty}^{y}\phi_x'\big(\widehat{\overline{H}}(s|x)\big)d\widehat{H}^u(s|x)-\int_{-\infty}^{y}\phi_x'\big(\overline{H}(s|x)\big)dH^u(s|x)\Bigg] \\
&:= Q_1(x,y)+Q_2(x,y).
\end{aligned}
$$

The second term $Q_2(x,y)$ can be further decomposed in the sum of three terms :

$$
\begin{aligned}
Q_2^{(1)}(x,y) &= -\frac{\partial}{\partial x}\int_{-\infty}^{y}\Big[\phi_x'\big(\widehat{\overline{H}}(s|x)\big)-\phi_x'\big(\overline{H}(s|x)\big)\Big]dH^u(s|x), \\
Q_2^{(2)}(x,y) &= \frac{\partial}{\partial x}\int_{-\infty}^{y}\phi_x'\big(\overline{H}(s|x)\big)d\Big[H^u(s|x)-\widehat{H}^u(s|x)\Big], \\
Q_2^{(3)}(x,y) &= -\frac{\partial}{\partial x}\int_{-\infty}^{y}\Big[\phi_x'\big(\widehat{\overline{H}}(s|x)\big)-\phi_x'\big(\overline{H}(s|x)\big)\Big]d\Big[\widehat{H}^u(s|x)-H^u(s|x)\Big].
\end{aligned}
$$

We will show that $Q_2^{(1)}(x,y)$ is of the desired order. The terms $Q_2^{(2)}(x,y)$ and $Q_2^{(3)}(x,y)$ can be dealt with in a similar way with additional use of integration by parts. Write

$$
\begin{aligned}
&Q_2^{(1)}(x,y) \\
&= -\int_{-\infty}^{y}\Big[\phi_x''\big(\widehat{\overline{H}}(s|x)\big)\dot{\widehat{\overline{H}}}(s|x)+\dot{\phi}_x'\big(\widehat{\overline{H}}(s|x)\big)-\phi_x''\big(\overline{H}(s|x)\big)\dot{\overline{H}}(s|x)-\dot{\phi}_x'\big(\overline{H}(s|x)\big)\Big]h^u(s|x)ds \\
&\quad -\int_{-\infty}^{y}\Big[\phi_x'\big(\widehat{\overline{H}}(s|x)\big)-\phi_x'\big(\overline{H}(s|x)\big)\Big]\frac{\partial}{\partial x}h^u(s|x)ds \\
&= -\int_{-\infty}^{y}\Big\{\phi_x''\big(\widehat{\overline{H}}(s|x)\big)\Big[\dot{\widehat{\overline{H}}}(s|x)-\dot{\overline{H}}(s|x)\Big]+\phi_x^{(3)}\big(\alpha_1(x,s)\big)\Big[\widehat{\overline{H}}(s|x)-\overline{H}(s|x)\Big]\dot{\overline{H}}(s|x) \\
&\qquad\quad +\dot{\phi}_x''\big(\alpha_2(x,s)\big)\Big[\widehat{\overline{H}}(s|x)-\overline{H}(s|x)\Big]\Big\}h^u(s|x)ds \\
&\quad -\int_{-\infty}^{y}\phi_x''\big(\alpha_3(x,s)\big)\Big[\widehat{\overline{H}}(s|x)-\overline{H}(s|x)\Big]\frac{\partial}{\partial x}h^u(s|x)ds,
\end{aligned}
$$

where $\alpha_j(x,s)$ is between $\overline{H}(s|x)$ and $\widehat{\overline{H}}(s|x)$ ($j=1,2,3$). From assumption (A4) we know that $\sup_{x\in R_X}\sup_{y\le\widetilde{T}_x}|\phi_x''(\overline{H}(y|x))|<\infty$ and similarly for $\phi_x^{(3)}$ and $\dot{\phi}_x''$. Hence, it follows from Proposition 1 in Akritas and Van Keilegom (2001) that $\sup_{x\in R_X}\sup_{y\le\widetilde{T}_x}|Q_2^{(1)}(x,y)|=O((nh_n^3)^{-1/2}(\log n)^{1/2})$ a.s.

To finish the proof it remains to show that $Q_1(x,y)$ is of the desired order. We use the abbreviated notation $\widehat{\overline{H}}_{ix}=\widehat{\overline{H}}(T_i|x)$, $\widehat{\overline{H}}_{i^-x}=\widehat{\overline{H}}(T_i^-|x)$ and similarly for $\dot{\widehat{\overline{H}}}_{ix}$ and $\dot{\widehat{\overline{H}}}_{i^-x}$.

23

Write

$$Q_1(x,y)$$
$$= - \sum_{T_i \leq y, \Delta_i = 1} \left\{ \dot{\phi}_x(\widehat{\overline{H}}_{i^- x}) + \phi'_x(\widehat{\overline{H}}_{i^- x})\dot{\widehat{\overline{H}}}_{i^- x} - \dot{\phi}_x(\widehat{\overline{H}}_{ix}) - \phi'_x(\widehat{\overline{H}}_{ix})\dot{\widehat{\overline{H}}}_{ix} \right\}$$
$$+ \sum_{T_i \leq y, \Delta_i = 1} \left\{ \dot{\phi}'_x(\widehat{\overline{H}}_{ix})W_{ni}(x,h_n) + \phi'_x(\widehat{\overline{H}}_{ix})\dot{W}_{ni}(x,h_n) + \phi''_x(\widehat{\overline{H}}_{ix})W_{ni}(x,h_n)\dot{\widehat{\overline{H}}}_{ix} \right\}$$
$$= - \sum_{T_i \leq y, \Delta_i = 1} \left\{ \dot{\phi}'_x(\widehat{\overline{H}}_{ix})W_{ni}(x,h_n) + \frac{1}{2}\dot{\phi}''_x(\alpha_1(x,i))W_{ni}^2(x,h_n) \right.$$
$$\left. + \phi'_x(\widehat{\overline{H}}_{i^- x})\dot{W}_{ni}(x,h_n) + \phi''_x(\widehat{\overline{H}}_{ix})W_{ni}(x,h_n)\dot{\widehat{\overline{H}}}_{ix} + \frac{1}{2}\phi_x^{(3)}(\alpha_2(x,i))W_{ni}^2(x,h_n)\dot{\widehat{\overline{H}}}_{ix} \right\}$$
$$+ \sum_{T_i \leq y, \Delta_i = 1} \left\{ \dot{\phi}'_x(\widehat{\overline{H}}_{ix})W_{ni}(x,h_n) + \phi'_x(\widehat{\overline{H}}_{ix})\dot{W}_{ni}(x,h_n) + \phi''_x(\widehat{\overline{H}}_{ix})W_{ni}(x,h_n)\dot{\widehat{\overline{H}}}_{ix} \right\}$$
$$= - \sum_{T_i \leq y, \Delta_i = 1} \left\{ \frac{1}{2}\dot{\phi}''_x(\alpha_1(x,i))W_{ni}^2(x,h_n) + \phi''_x(\alpha_3(x,i))\dot{W}_{ni}(x,h_n)W_{ni}(x,h_n) \right.$$
$$\left. + \frac{1}{2}\phi_x^{(3)}(\alpha_2(x,i))W_{ni}^2(x,h_n)\dot{\widehat{\overline{H}}}_{ix} \right\}, \tag{A.9}$$

where $\alpha_j(x,i)$ is between $\widehat{\overline{H}}_{ix}$ and $\widehat{\overline{H}}_{i^- x}$ $(j = 1, 2, 3)$. From Proposition 1 in Akritas and Van Keilegom (2001) it follows that $\inf_{x \in R_X} \min_{i:T_i \leq \widetilde{T}_x} \alpha_j(x,i) > 0$ for $j = 1, 2, 3$ and for n large enough. This together with assumption (A4) implies that

$$\sup_{x \in R_X, y \leq \widetilde{T}_x} |Q_1(x,y)| \leq C_1 \sum_{i=1}^{n} W_{ni}^2(x,h_n) + C_2 \sum_{i=1}^{n} |\dot{W}_{ni}(x,h_n)|W_{ni}(x,h_n)$$

for some $0 < C_1, C_2 < \infty$. The first term is of the order $O((nh_n)^{-1})$ a.s. because $\max_i \sup_{x \in R_X} W_{ni}(x,h_n) = O((nh_n)^{-1})$ a.s., whereas the second term is $O((nh_n^2)^{-1}) = O((nh_n^3)^{-1/2}(\log n)^{1/2})$ a.s., since $\max_i \sup_{x \in R_X} |\dot{W}_{ni}(x,h_n)| = O((nh_n^2)^{-1})$ a.s. $\qquad \square$

Proof of Proposition 3.3. When $|x - x'| \geq h_n$, then Proposition 3.3 is trivially true. In what follows we consider the case when $|x - x'| < h_n$. By using decomposition (A.8) we can easily see that

$$\left| \dot{\widehat{F}}(y|x) - \dot{F}(y|x) - \dot{\widehat{F}}(y|x') + \dot{F}(y|x') \right| |x - x'|^{-\delta}$$
$$\leq \left| Q_1(x,y) - Q_1(x',y) \right| |x - x'|^{-\delta} + \left| Q_2(x,y) - Q_2(x',y) \right| |x - x'|^{-\delta} \tag{A.10}$$
$$+ O((nh_n^{1+2\delta})^{-1/2}(\log n)^{1/2}) \quad a.s.,$$

where $Q_1(x,y)$ and $Q_2(x,y)$ are defined in the proof of the previous proposition. For the first term on the right hand side of (A.10) we use (A.9), which leads to

$$
\begin{aligned}
&Q_1(x,y) - Q_1(x',y) \\
&= -\sum_{T_i \leq y, \Delta_i = 1} \left\{ \frac{1}{2} \dot{\phi}_x''(\alpha_1(x,i)) W_{ni}^2(x,h_n) - \frac{1}{2} \dot{\phi}_{x'}''(\alpha_1(x',i)) W_{ni}^2(x',h_n) \right. \tag{A.11} \\
&\qquad + \phi_x''(\alpha_3(x,i)) \dot{W}_{ni}(x,h_n) W_{ni}(x,h_n) - \phi_{x'}''(\alpha_3(x',i)) \dot{W}_{ni}(x',h_n) W_{ni}(x',h_n) \\
&\qquad \left. + \frac{1}{2} \phi_x^{(3)}(\alpha_2(x,i)) W_{ni}^2(x,h_n) \dot{\overline{\widehat{H}}}(T_i|x) - \frac{1}{2} \phi_{x'}^{(3)}(\alpha_2(x',i)) W_{ni}^2(x',h_n) \dot{\overline{\widehat{H}}}(T_i|x') \right\}.
\end{aligned}
$$

By adding and subtracting terms we can see that the sum of the first two terms of (A.11) equals

$$
\begin{aligned}
&-\frac{1}{2} \sum_{T_i \leq y, \Delta_i = 1} \left[\dot{\phi}_x''(\alpha_1(x,i)) - \dot{\phi}_{x'}''(\alpha_1(x',i)) \right] W_{ni}^2(x,h_n) \\
&-\frac{1}{2} \sum_{T_i \leq y, \Delta_i = 1} \dot{\phi}_{x'}''(\alpha_1(x',i)) [W_{ni}^2(x,h_n) - W_{ni}^2(x',h_n)],
\end{aligned}
$$

which under assumption (A4) when multiplied with $|x-x'|^{-\delta}$ is of the order $O((nh_n^{1+\delta})^{-1})$ a.s. In a similar way we can show the order of the other terms of (A.11).

By straightforward algebra and assumption (A4) the second term on the right hand side of (A.10) can be written in terms of $\left| \dot{\hat{H}}(y|x) - \dot{H}(y|x) - \dot{\hat{H}}(y|x') + \dot{H}(y|x') \right| |x-x'|^{-\delta}$, which by Lemma 4.2 in Van Keilegom and Akritas (1999) is of the desired order. $\qquad \square$

Lemma A.1. *Under the assumptions of Proposition 3.4, we have*

$$
\sup_{x \in R_X} \sup_{y \leq \tilde{T}_x} \left| \int_{-\infty}^{y} \left[\phi_x'(\overline{\widehat{H}}(t|x)) - \phi_x'(\overline{H}(t|x)) \right] d\left[\widehat{H}^u(t|x) - H^u(t|x) \right] \right| = O\left((nh_n)^{-3/4} (\log n)^{3/4} \right) a.s.
$$

Proof. First note that it follows from Lemma 4.4 in Du and Akritas (2002) that

$$
\begin{aligned}
\sup_{x \in R_X} \sup_{|M(y_2)-M(y_1)| \leq a_n} &\left| \overline{\widehat{H}}(y_2|x) - \overline{H}(y_2|x) - \overline{\widehat{H}}(y_1|x) + \overline{H}(y_1|x) \right| \\
&= O\left((nh_n)^{-3/4} (\log n)^{3/4} \right) \ a.s., \tag{A.12}
\end{aligned}
$$

where $a_n = O\left((nh_n)^{-1/2} (\log n)^{1/2} \right)$ and $M = M_1 + M_3$, and where the functions M_1 and M_3 come from assumption (A6). Equality (A.12) also holds for H^u with the function $M = M_2 + M_4$. Note that the result in Du and Akritas (2002) is in fact slightly less general, since they impose that y_1 and y_2 are bounded by some finite T, but it is easy to see that this is by no means necessary. Further note that after redefining the function M as $M := M_1 + ... + M_4$, equality (A.12) holds for both H and H^u.

25

Now, because of assumption (A6) we can partition $\mathbb{R}$ into $m = Ca_n^{-1}$ subintervals $[t_i, t_{i+1}]$ such that $|M(t_{i+1}) - M(t_i)| \leq a_n$. Consequently, we have

$$\sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} \left| \int_{-\infty}^{y} \left[\phi'_x(\widehat{\overline{H}}(t|x)) - \phi'_x(\overline{H}(t|x)) \right] d\left[\widehat{H}^u(t|x) - H^u(t|x) \right] \right|$$

$$\leq \sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} \left| \sum_{i=2}^{k(y)} \int_{t_{i-1}}^{t_i} \left[\phi'_x(\widehat{\overline{H}}(t|x)) - \phi'_x(\overline{H}(t|x))) \right] d\left[\widehat{H}^u(t|x) - H^u(t|x) \right] \right.$$

$$\left. + \int_{t_{k(y)}}^{y} \left[\phi'_x(\widehat{\overline{H}}(t|x)) - \phi'_x(\overline{H}(t|x)) \right] d\left[\widehat{H}^u(t|x) - H^u(t|x) \right] \right|$$

$$:= \sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} \left| \alpha_{n1}(x,y) + \alpha_{n2}(x,y) \right|,$$

where $t_{k(y)} \leq y < t_{k(y)+1}$. The term $\alpha_{n2}(x,y)$ can be treated in the same way as $\alpha_{n1}(x,y)$, so we will restrict ourselves to showing that $\sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} |\alpha_{n1}(x,y)| = O\left((nh_n)^{-3/4}(\log n)^{3/4}\right)$ a.s. Write

$$\sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} |\alpha_{n1}(x,y)|$$

$$\leq \sup_{x \in R_X} \sum_{i=2}^{m(x)} \left| \int_{t_{i-1}}^{t_i} \left[\phi'_x(\widehat{\overline{H}}(t|x)) - \phi'_x(\overline{H}(t|x)) - \phi'_x(\widehat{\overline{H}}(t_i|x)) + \phi'_x(\overline{H}(t_i|x)) \right] \right.$$

$$\left. d\left[\widehat{H}^u(t|x) - H^u(t|x) \right] \right|$$

$$+ \sup_{x \in R_X} \sum_{i=2}^{m(x)} \left| \left[\phi'_x(\widehat{\overline{H}}(t_i|x)) - \phi'_x(\overline{H}(t_i|x)) \right] \int_{t_{i-1}}^{t_i} d\left[\widehat{H}^u(t|x) - H^u(t|x) \right] \right|,$$

where $m(x)$ is such that $t_{m(x)} \leq \widetilde{T}_x < t_{m(x)+1}$. Using the notation $J_{a_n}^x = \{(t_1, t_2) : M(t_2) - M(t_1) < a_n, \ t_1, t_2 \leq \widetilde{T}_x\}$, we have

$$\sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} |\alpha_{n1}(x,y)|$$

$$\leq 2 \sup_{x \in R_X} \sup_{(t_1, t_2) \in J_{a_n}^x} \left| \phi'_x(\widehat{\overline{H}}(t_1|x)) - \phi'_x(\overline{H}(t_1|x)) - \phi'_x(\widehat{\overline{H}}(t_2|x)) + \phi'_x(\overline{H}(t_2|x)) \right|$$

$$+ Ca_n^{-1} \sup_{x \in R_X} \sup_{t \in J_{a_n}^x} \left| \phi'_x(\widehat{\overline{H}}(t|x)) - \phi'_x(\overline{H}(t|x)) \right|$$

$$\times \sup_{x \in R_X} \sup_{|M(y_2) - M(y_1)| \leq a_n} \left| \widehat{H}^u(t_1|x) - H^u(t_1|x) - \widehat{H}^u(t_2|x) + H^u(t_2|x) \right|$$

$$:= \alpha_{n11} + \alpha_{n12}.$$

After applying a Taylor expansion we can bound the second term on the right hand side :

$$|\alpha_{n12}| \leq Ca_n^{-1} \sup_{x \in R_X} \sup_{t \leq \widetilde{T}_x} \left| \phi''_x(\xi_{t,x}) \right| \sup_{x \in R_X} \sup_{t \in \mathbb{R}} \left| \widehat{\overline{H}}(t|x) - \overline{H}(t|x) \right|$$

$$\times \sup_{x \in R_X} \sup_{|M(y_2) - M(y_1)| \leq a_n} \left| \widehat{H}^u(t_1|x) - H^u(t_1|x) - \widehat{H}^u(t_2|x) + H^u(t_2|x) \right|,$$

26

where $\xi_{t,x}$ is between $\overline{H}(t|x)$ and $\widehat{\overline{H}}(t|x)$. By Proposition 1 in Akritas and Van Keilegom (2001) we know that for n large enough there is $\gamma > 0$ such that $\inf_{x \in R_X, t \leq \widetilde{T}_x} |\xi_{t,x}| > \gamma$ a.s., from where we have that $\sup_{x \in R_X, t \leq \widetilde{T}_x} |\phi_x''(\xi_{t,x})|$ is a.s. finite. Now, by using Proposition 1 in Akritas and Van Keilegom (2001) and (A.12) to bound the second and third factor respectively, we get that α_{n12} is of the desired order $O\big((nh_n)^{-3/4}(\log n)^{3/4}\big)$ a.s. By again applying a Taylor expansion, we can bound α_{n11} :

$$
\begin{aligned}
|\alpha_{n11}| \quad \leq \quad & 2 \sup_{x \in R_X} \sup_{(t_1, t_2) \in J_{a_n}^x} \left| \phi_x''\big(\overline{H}(t_1|x)\big)\big[\widehat{\overline{H}}(t_1|x) - \overline{H}(t_1|x)\big] \right. \\
& \qquad\qquad\qquad -\phi_x''\big(\overline{H}(t_2|x)\big)\big[\widehat{\overline{H}}(t_2|x) - \overline{H}(t_2|x)\big] \Big| \\
& + \sup_{x \in R_X} \sup_{(t_1, t_2) \in J_{a_n}^x} \left| \phi_x^{(3)}\big(\xi_{t_1,x}\big)\big[\widehat{\overline{H}}(t_1|x) - \overline{H}(t_1|x)\big]^2 \right. \\
& \qquad\qquad\qquad -\phi_x^{(3)}\big(\xi_{t_2,x}\big)\big[\widehat{\overline{H}}(t_2|x) - \overline{H}(t_2|x)\big]^2 \Big|, \qquad (A.13)
\end{aligned}
$$

where $\xi_{t,x}$ is between $\widehat{\overline{H}}(t|x)$ and $\overline{H}(t|x)$. Following the same argument as for α_{n12}, we have that $\sup_{x \in R_X, t \leq \widetilde{T}_x} |\phi_x^{(3)}(\xi_{t,x})|$ is a.s. finite. Now, we can conclude that the second term on the right hand side of (A.13) is of the order $O\big((nh_n)^{-1} \log n\big)$ a.s. by Proposition 1 in Akritas and Van Keilegom (2001). For the first term on the right hand side, we add and subtract terms followed by a Taylor expansion, and obtain the following bound :

$$
\begin{aligned}
& 2 \sup_{x \in R_X} \sup_{(t_1, t_2) \in J_{a_n}^x} \left| \phi_x''\big(\overline{H}(t_1|x)\big)\big[\widehat{\overline{H}}(t_1|x) - \overline{H}(t_1|x) - \widehat{\overline{H}}(t_2|x) + \overline{H}(t_2|x)\big] \right| \\
& + 2 \sup_{x \in R_X} \sup_{(t_1, t_2) \in J_{a_n}^x} \left| \phi_x^{(3)}\big(\xi_{t_1,t_2,x}\big)\big[\overline{H}(t_2|x) - \overline{H}(t_1|x)\big]\big[\widehat{\overline{H}}(t_2|x) - \overline{H}(t_2|x)\big] \right|,
\end{aligned}
$$

where $\xi_{t_1,t_2,x}$ is between $\overline{H}(t_1|x)$ and $\overline{H}(t_2|x)$. Following an analogous reasoning as before we conclude that $\sup_{x \in R_X, t \leq \widetilde{T}_x} |\phi_x''(\overline{H}(t|x))|$ and $\sup_{x \in R_X} \sup_{(t_1, t_2) \in J_{a_n}^x} |\phi_x^{(3)}(\xi_{t_1,t_2,x})|$ are a.s. finite. The second term above is of the order $O\big((nh_n)^{-1} \log n\big)$ a.s. by the definition of $J_{a_n}^x$ and Proposition 1 in Akritas and Van Keilegom (2001). The first term is of the desired order $O\big((nh_n)^{-3/4}(\log n)^{3/4}\big)$ a.s. thanks to (A.12). $\qquad\square$

Proof of Proposition 3.4. Using a similar decomposition as in the proof of Theorem 1 in Braekers and Veraverbeke (2005), we can write

$$
\begin{aligned}
& \widehat{F}(y|x) - F(y|x) \\
& = \sum_{i=1}^n W_{ni}(x, h_n) g(T_i, \Delta_i, y|x) + R_{n1}(y|x) + R_{n2}(y|x) + R_{n3}(y|x) + R_{n4}(y|x),
\end{aligned}
$$

where we refer to the aforementioned paper for the precise definition of the remainder terms $R_{nj}(y|x)$ ($j = 1, 2, 3, 4$). We can easily show that the terms $R_{n1}(y|x)$, $R_{n2}(y|x)$ and $R_{n3}(y|x)$

27

are uniformly of the order $O\big((nh_n)^{-1}\log n\big)$ a.s. by following the same arguments as in the proof of Theorem 1 in Braekers and Veraverbeke (2005). In fact, the most important difference with their proof is that we use Proposition 1 in Akritas and Van Keilegom (2001) (which gives the rate of convergence of $\widehat{H}(y|x) - H(y|x)$ uniformly in x and y) instead of Lemma A.4 in Van Keilegom and Veraverbeke (1997b) (which gives the same result but for Gasser-Müller weights and for a fixed value of x). The order of the remainder term $R_{n4}(y|x)$ is given by $O\big((nh_n)^{-3/4}(\log n)^{3/4}\big)$ a.s. uniformly in x and $y \leq \widetilde{T}_x$ by Lemma A.1.

This together with the uniform rate of convergence of $\widehat{F}(y|x) - F(y|x)$ given in Proposition 3.1, entails that

$$\sup_{x \in R_X} \sup_{y \leq \widetilde{T}_x} \Big| \sum_{i=1}^n W_{ni}(x, h_n) g(T_i, \Delta_i, y|x) \Big| = O\big((nh_n)^{-1/2}(\log n)^{1/2}\big) \ a.s.,$$

and hence

$$\widehat{F}(y|x) - F(y|x)$$
$$= \frac{1}{nh_n f_X(x)} \sum_{i=1}^n K\Big(\frac{x - X_i}{h_n}\Big) g(T_i, \Delta_i, y|x) + O\big((nh_n)^{-3/4}(\log n)^{3/4}\big),$$

since $\sup_x |\widehat{f}_X(x) - f_X(x)| = O\big((nh_n)^{-1/2}(\log n)^{1/2}\big)$ a.s. This finishes the proof. $\qquad\square$

Proof of Proposition 3.5. For every $x \in R_X$ we can write

$$
\begin{aligned}
\widehat{m}(x) - m(x) &= \int_0^1 \int_{-\infty}^{+\infty} I(F^{-1}(s|x) \leq t \leq \widehat{F}^{-1}(s|x)) I(\widehat{F}^{-1}(s|x) > F^{-1}(s|x)) \, dt \, J(s) \, ds \\
&\quad - \int_0^1 \int_{-\infty}^{+\infty} I(\widehat{F}^{-1}(s|x) \leq t \leq F^{-1}(s|x)) I(\widehat{F}^{-1}(s|x) \leq F^{-1}(s|x)) \, dt \, J(s) \, ds \\
&= \int_{-\infty}^{+\infty} \int_{\widehat{F}(t|x)}^{F(t|x)} I(\widehat{F}^{-1}(s|x) > F^{-1}(s|x)) J(s) \, ds \, dt \\
&\quad - \int_{-\infty}^{+\infty} \int_{F(t|x)}^{\widehat{F}(t|x)} I(\widehat{F}^{-1}(s|x) \leq F^{-1}(s|x)) J(s) \, ds \, dt \\
&= \int_{-\infty}^{+\infty} \int_{\widehat{F}(t|x)}^{F(t|x)} J(s) \, ds \, dt \\
&= \int_{-\infty}^{+\infty} \Big[L(F(t|x)) - L(\widehat{F}(t|x)) \Big] dt,
\end{aligned}
$$

where $L(u) = \int_0^u J(s)ds$ for all $0 \leq u \leq 1$. Using the substitution $t = F^{-1}(s|x)$ we get

$$
\begin{aligned}
\widehat{m}(x) - m(x) &= \int_0^1 \Big[L(s) - L(\widehat{F}(F^{-1}(s|x)|x)) \Big] dF^{-1}(s|x) \\
&= \int_0^1 \Big[L(s) - L(\widehat{F}(F^{-1}(s|x)|x)) \Big] \frac{1}{f\big(F^{-1}(s|x)|x\big)} ds.
\end{aligned}
$$

Finally, by using a Taylor expansion, it follows that

$$\sup_{x \in R_x} |\widehat{m}(x) - m(x)|$$

$$\leq \left[\inf_{x \in R_X} \inf_{s_0 \leq s \leq s_1} f\big(F^{-1}(s|x)|x\big)\right]^{-1} \sup_{s_0 \leq s \leq s_1} J(s) \sup_{x \in R_x} \sup_{F^{-1}(s_0|x) \leq y \leq F^{-1}(s_1|x)} \big|\widehat{F}(y|x) - F(y|x)\big|.$$

Note that for all x, $[F^{-1}(s_0|x), F^{-1}(s_1|x)] \subset (-\infty, \widetilde{T}_x]$, since $s_1 \leq \inf_x F(\widetilde{T}_x|x)$ by assumption (A2). The result now follows by using again assumption (A2) together with Proposition 3.1. Part (b) can be shown in a similar way. $\qquad\square$

Proof of Proposition 3.6. The proof is very analogous to the proof of Proposition 4.6 in Van Keilegom and Akritas (1999). The only difference is that we use our Proposition 3.2 for the uniform rate of convergence of $\dot{\widehat{F}}(y|x) - \dot{F}(y|x)$, whereas they use the second statement of their Proposition 4.3. $\qquad\square$

Proof of Proposition 3.7. We follow exactly the same steps as in the proof of Proposition 4.7 in Van Keilegom and Akritas (1999), with the only exception that instead of using their Proposition 4.4, we use our Proposition 3.3. $\qquad\square$

Proof of Proposition 3.8. Using the notation $L(u) = \int_0^u J(s)ds$ we write

$$\widehat{m}(x) - m(x)$$
$$= -\int_{-\infty}^{+\infty} \left[L\big(\widehat{F}(y|x)\big) - L\big(F(y|x)\big)\right] dy$$
$$= -\int_{-\infty}^{+\infty} J\big(F(y|x)\big)\left[\widehat{F}(y|x) - F(y|x)\right] dy + \frac{1}{2}\int_{-\infty}^{+\infty} J'\big(\beta(x,y)\big)\left[\widehat{F}(y|x) - F(y|x)\right]^2 dy,$$

with $\beta(x,y)$ between $\widehat{F}(y|x)$ and $F(y|x)$. By Proposition 3.1, the second term above is $O\big((nh_n)^{-1}\log n\big)$ a.s. Using Proposition 3.4, the first term can be written as

$$-\frac{1}{nh_n f_X(x)}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{h_n}\Big)\int_{-\infty}^{+\infty} J\big(F(y|x)\big)g(T_i, \Delta_i, y|x)dy$$
$$+\int_{-\infty}^{+\infty} J\big(F(y|x)\big)R_n(y|x)dy \quad a.s.$$

The rate of convergence of the last term above is $O\big((nh_n)^{-3/4}(\log n)^{3/4}\big)$ a.s., which completes the proof for part (a).

To prove the second assertion in Proposition 3.8, we mimic Van Keilegom and Akritas (1999) and write

$$\widehat{\sigma}(x) - \sigma(x) = \frac{\widehat{\sigma}^2(x) - \sigma^2(x)}{2\sigma(x)} - \frac{\left[\widehat{\sigma}(x) - \sigma(x)\right]^2}{2\sigma(x)}.$$

It follows from Proposition 3.5 that the second term above is $O\big((nh_n)^{-1}\log n\big)$ a.s. uniformly in $x \in R_X$. For the first term, we have similarly as in the proof of part (a),

$$\widehat{\sigma}^2(x) - \sigma^2(x) = \int_0^1 \left[\widehat{F}^{-1}(s|x)^2 - F^{-1}(s|x)^2\right]J(s)ds - \left[\widehat{m}^2(x) - m^2(x)\right]$$

$$= -2\int_{-\infty}^{+\infty} \left[L\big(\widehat{F}(y|x)\big) - L\big(F(y|x)\big)\right]ydy - \left[\widehat{m}^2(x) - m^2(x)\right]. \quad \text{(A.14)}$$

Now, by a Taylor expansion we have

$$\int_{-\infty}^{+\infty} \left[L\big(\widehat{F}(y|x)\big) - L\big(F(y|x)\big)\right]ydy$$

$$= \int_{-\infty}^{+\infty} J\big(F(y|x)\big)\left[\widehat{F}(y|x) - F(y|x)\right]ydy$$

$$+ \frac{1}{2}\int_{-\infty}^{+\infty} J\big(\gamma(x,y)\big)\left[L\big(\widehat{F}(y|x)\big) - L\big(F(y|x)\big)\right]^2 ydy$$

$$= -\frac{1}{nh_n f_X(x)}\sum_{i=1}^n \int_{-\infty}^{+\infty} J\big(F(y|x)\big)K\Big(\frac{x - X_i}{h_n}\Big)g\big(T_i, \Delta_i, y|x\big)ydy$$

$$+ O\big((nh_n)^{-3/4}(\log n)^{3/4}\big) \ a.s., \quad \text{(A.15)}$$

where $\gamma(x,y)$ is between $F(y|x)$ and $\widehat{F}(y|x)$. In the above, the second equality follows from Proposition 3.1 and by using the same reasoning as at the end of part (a) of this proof. Next, we write

$$\widehat{m}^2(x) - m^2(x) \quad \text{(A.16)}$$

$$= 2m(x)\left[\widehat{m}(x) - m(x)\right] + \left[\widehat{m}(x) - m(x)\right]^2$$

$$= -\frac{2m(x)}{nh_n f_X(x)}\sum_{i=1}^n K\Big(\frac{x - X_i}{h_n}\Big)\int_{-\infty}^{+\infty} J\big(F(y|x)\big)g\big(T_i, \Delta_i, y|x\big)dy$$

$$+ O\big((nh_n)^{-3/4}(\log n)^{3/4}\big) \ a.s.,$$

with the last equality following from part (a) and from Proposition 3.5(a). Combining (A.14), (A.15) and (A.16), we get the result. $\qquad\square$

References

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

Akritas, M. G. and I. Van Keilegom (2001). Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics 28*, 549–568.

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.

Braekers, R. and N. Veraverbeke (2008). The conditional Koziol-Green model under dependent censoring. *Statistics and Probability Letters 78*, 927–937.

Claeskens, G. and I. Van Keilegom (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics 31*, 1852–1884.

Crowder, M. (1991). On the identifiability crisis in competing risks analysis. *Scandinavian Journal of Statistics 18*, 223–233.

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

de Uña Álvarez, J. and N. Veraverbeke (2013). Generalized copula-graphic estimator. *TEST 22*, 343–360.

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

Gaddah, A. and R. Braekers (2010a). Testing under the extended Koziol-Green model. In Durante, F., Härdle, W., Jaworski, P., Rychlik, T. (Eds). *Workshop on Copula Theory and its Applications, Lecture Notes in Statistics - Proceedings. Springer, Berlin/Heidelberg 98*, 279–288.

Gaddah, A. and R. Braekers (2010b). Weak convergence for the conditional distribution function in a Koziol-Green model under dependent censoring. *Journal of Statistical Planning and Inference 139*, 930–943.

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.

Klein, J. P. and M. L. Moeschberger (1997). *Survival Analysis, Techniques for Censored and Truncated Data*. New York: Springer.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer-Verlag.

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.

Schwarz, M., G. Jongbloed, and I. Van Keilegom (2013). On the identifiability of copulas in bivariate competing risks models. *Canadian Journal of Statistics 41*, 291–303.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

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.

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 (1997a). 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 (1997b). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantiles process. *Communications in Statistics: Theory and Methods 26*, 853–869.

Wang, A. (2012). On the nonidentifiability property of Archimedean copula models under dependent censoring. *Statistics & Probability Letters 82*, 621–625.

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