A nonparametric conditional copula model for successive duration times, with application to insurance subscription

Yohann Le Faou[1], Olivier Lopez[2]

September 10, 2018

Abstract

We consider two dependent random times T and U, that correspond to two successive events. This setting is motivated by an application to insurance subscription, where a potential dependence exists between a time before effectiveness of the contract T, and a time U before its termination by the policyholder. The setting also extends to various types of applications involving two duration variables with some hierarchical link between the events. Indeed, since a contract can be terminated only after it becomes effective, data are subject to a particular type of censoring, where the variable U is systematically censored when the variable T is. In this framework, a nonparametric conditional copula model is considered, in the spirit of (Gijbels, Veraverbeke, & Omelka, 2011). The uniform consistency of the conditional association parameter is obtained under conditions of dependence structure and of censoring mechanism. A simulation study and a real data application show the practical behavior of the method.

Key words: Conditional copulas, censoring, insurance, kernel smoothing.

Short title: Conditional copula for successive duration times.

[1] *Forsides & Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, LPSM, 4 place Jussieu, F-75005 Paris, France, mail : yohann.lefaou@courriel.upmc.fr*

[2] *Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, LPSM, 4 place Jussieu, F-75005 Paris, France, mail: olivier.lopez@sorbonne-universite.fr*

1 Introduction

In this paper, we consider the estimation of a conditional copula function of a couple of duration variables, in a framework where the two durations are observed successively. In numerous practical situations, one can be interested in the occurrence of two events that happen in a time succession. The example we have in mind comes from the study of the termination of insurance contracts. From the subscription, the owner may have some random delay T before the contract to be effective, while the termination of the contract occurs at a time $T+U$, with $U \geq 0$. Similar situations occur in biostatistics, where T can be an infection time and U the time before the individual is cured (see for example (Wang & Wells, 1998), and (Meira-Machado, Sestelo, & Gonçalves, 2016) for a detailed review of the applications and techniques in this field). This paper aims to study the dependence structure between two such times T and U, in presence of covariates $X \in \mathbb{R}^d$ that may impact the joint distribution.

Copula theory is a quite popular way to deal with dependence, due to Sklar's Theorem (Sklar, 1959) which states that the joint distribution function $F(t,u) = \mathbb{P}(T \leq t, U \leq u)$ of a bivariate vector (T,U) can be written as

$$F(t,u) = \mathfrak{C}(F_T(t), F_U(u)),$$

where $F_T(t) = \mathbb{P}(T \leq t)$, $F_U(u) = \mathbb{P}(U \leq u)$, and $\mathfrak{C}$ is a copula function (that is a distribution function over $[0,1]^2$ with uniform margins), this decomposition being unique when the margins are continuous. Hence Sklar's Theorem ensures a separation between the marginal behaviors of T and U (defined by F_T and F_U), and the dependence structure, entirely contained in the function $\mathfrak{C}$. Conditional copulas are required when one focus on the influence of covariates X on this dependence structure (see e.g. (Gijbels et al., 2011), (Veraverbeke, Gijbels, & Omelka, 2011), (Derumigny & Fermanian, 2017)).

When dealing with duration variables, a supplementary difficulty is caused by the censoring phenomenon. In the situation we describe, a unique censoring variable C is involved, representing the time before the end of the statistical study. Indeed, if $C < T+U$, the policyholder did not stay under observation long enough to observe the whole phenomenon we are interested in. Copulas under censoring have been studied, for instance by (Lakhal-Chaieb, 2010), (Gribkova & Lopez, 2015) and (Geerdens, Acar, & Janssen, 2018). In this paper, we correct the effects of the censoring by using appropriate weights that allow our conditional copula estimator to be asymptotically consistent.

The rest of the paper is organized as follows. In Section 2, we define the observations and the methodology to estimate conditional copulas under censoring. Then, the Section 3 is devoted to the presentation of asymptotic results while we investigate the finite sample behavior of the procedure in a simulation study and a real data analysis, presented in Section 4. Technical arguments are presented in the Appendix section.

2 Observations and Methodology

2.1 Model

We consider i.i.d. replications $(T_i, U_i, X_i, C_i)_{1 \leq i \leq n}$ of a random vector (T, U, X, C) and we aim to study the dependence structure between $T \in \mathbb{R}$ and $U \in \mathbb{R}$. The random variable $X \in \mathbb{R}^d$ is a vector of covariates that may have an impact on this dependence structure (and also on the marginal distributions of T and U), and $C \in \mathbb{R}$ is a censoring variable.

The variables T and U are not always observed, due to the presence of the censoring. Instead of (T_i, U_i), one observes

$$\begin{cases} Y_i &= \min(T_i, C_i), \\ Z_i &= \min(U_i, C_i - T_i), \\ \eta_i &= \mathbf{1}_{T_i \leq C_i}, \\ \gamma_i &= \mathbf{1}_{T_i + U_i \leq C_i}. \end{cases}$$

The covariates X_i are assumed to be fully observed (not subject to censoring). For the realization of the censoring C_i, two cases exist. In the application we have in mind (see Section 2.2 and Section 4.2), C_i is known for all observations. In a more general situation, C_i may not be known. In this last case, the statistical methodology that we develop is a little bit more delicate as we will see in the following.

The following identifiability assumption is required in order to estimate the distribution of (T, U, X) from the observations.

Assumption 1 *Assume that C is independent from (T, U, X).*

Let $F(t, u|x) = \mathbb{P}(T \leq t, U \leq u|X = x)$ be the conditional distribution function of (T, U) given $X = x$, and $F_T(t|x) = \mathbb{P}(T \leq t|X = x)$ (resp. $F_U(u|x) = \mathbb{P}(U \leq u|X = x)$) be the conditional distribution function of T (resp. U), where all distribution functions are assumed to be continuous. We also define $\tau_{T+U}(x) = \inf\{z : \mathbb{P}(Y + Z \geq z|x) = 0\}$. Clearly, the distribution of (T, U, X) can not be estimated (at least nonparametrically) for values of (t, u, x) such that $t + u \geq \tau_{T+U}(x)$, since it is impossible to observe non-censored events in this part of the distribution. Sklar's Theorem ensures that $F(t, u|x) = \mathfrak{C}^{(x)}(F_T(t|x), F_U(u|x))$, where $\mathfrak{C}^{(x)}$ denotes the copula of the distribution of (T, U) conditionally on $X = x$. In the following, we assume that the copula $\mathfrak{C}^{(x)}$ stays in the same parametric copula family for all x, but with its association parameter allowed to depend on x. This is summarized in Assumption 2 below.

Assumption 2 *Let $\mathcal{C} = \{\mathfrak{C}_\theta : \theta \in \Theta\}$, with Θ a compact subset of $\mathbb{R}^k$, be a parametric family of copula functions. Assume that, for all x in the support of the random vector X, there exists $\theta(x) \in \Theta$ such that*

$$\mathfrak{C}^{(x)} = \mathfrak{C}_{\theta(x)}.$$

Our aim is to retrieve the function $\theta(x)$ from our observations.

2.2 Motivation of this model

Our method applies to a problem which arises in the field of insurance subscription. The data we consider (described in Section 4.2.1) belongs to an insurance broker who wants to have information about the quality of the underwriters who sell the insurance contracts. A first indicator would be the volume of sales per underwriter, but a crucial issue is to have insight in the quality of the contracts that have been subscribed. One element to appreciate this quality is the time the consumer keeps his contract, before terminating it and starting another contract with a different insurer. In our framework, the lifetime of the contract is the variable U, that is the difference between the date of termination of the contract and the date of effect. The date of effect of the policy is usually not the same as the date of subscription. We denote by T the time between the date of subscription and the date of effect.

It seems obvious that the two durations T and U should not be independent. The knowledge of their dependence structure is a precious indicator to develop sales strategies and to evaluate the turnover in an insurance portfolio. Additionally, many variables on the customer are usually available, and these variables may have an impact on the dependence structure. This motivates the use of conditional copulas to model the dependence between T and U.

2.3 Conditional copula estimation

Let $M(x,\theta) = E[\log c_\theta(F_T(T|X), F_U(U|X))|X = x]$, where $c_\theta(a,b) = \partial^2_{a,b}\mathfrak{C}_\theta(a,b)$ denotes the copula density associated with copula function $\mathfrak{C}_\theta$. We have, by definition of $\theta(x)$,

$$\theta(x) = \arg\max_{\theta\in\Theta} M(x,\theta).$$

To ensure identifiability of the model, we assume that for all x in the support of X, $\theta(x)$ is the unique maximum of $M(x,\theta)$. The idea of our procedure is to estimate the function M, and then to perform its maximization in order to estimate $\theta(x)$.

In an ideal situation, first consider that F_T and F_U are exactly known. If we had observed the complete data $(T_i, U_i, X_i)_{1\leq i\leq n}$, we could have estimated $M(x,\theta)$ thanks to a Nadaraya-Watson estimator ((Watson, 1964) and (Nadaraya, 1964)) such as

$$\sum_{i=1}^n w_{i,n}(x) \log c_\theta(F_T(T_i|X_i), F_U(U_i|X_i)),$$

where

$$w_{i,n}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h}\right)}, \tag{2.1}$$

and K is a kernel function (i.e. a positive and symmetric real valued function such that $\int K(u)du = 1$). However, this is impossible in our case due to the presence of censoring. If we consider nevertheless a function ϕ such that $E[|\phi(T,U,X)|] < \infty$ and $\phi(t,u,x) = 0$ for $t + u \geq \tau_{U+T}(x)$, then under Assumption 1, elementary computations show that

$$E\left[\frac{\delta\phi(Y,Z,X)}{S_C(Y+Z)}\middle| X\right] = E\left[\phi(T,U,X)|X\right], \tag{2.2}$$

with $S_C(t) = \mathbb{P}(C > t)$, and $\delta = \eta\gamma$.

As a consequence of equation (2.2), we see that if the function S_C were known, the function $M(x,\theta)$ could be estimated by the kernel estimator

$$\sum_{i=1}^n w_{i,n}(x) \frac{\delta_i \log c_\theta(F_T(Y_i|X_i), F_U(Z_i|X_i))}{S_C(Y_i + Z_i)},$$

with $w_{i,n}(x)$ as in (2.1). The kernel function K that we consider in this article is assumed to be a symmetric positive and bounded function, with $K(u) = 0$ for $\|u\| \geq 1$, $\int K(u)du = 1$ and $\int \|u\|^2 K(u)du < \infty$.

In practice, the function S_C is not known. However, it can be estimated, at least in the two situations that we mention in Section 2.1.

- **First case: C_i is observed for all individuals.** In this situation, we can estimate $S_C(t)$ with the empirical survival function, i.e.

$$\hat{S}_C^{(1)}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{C_i > t}. \tag{2.3}$$

- **Second case: Some C_i are unobserved due to the censoring.** In this second situation, the function S_C can be estimated by a Kaplan-Meier estimator, see (Kaplan & Meier, 1958). Put more precisely, consider

$$\begin{cases} Y_i' &= \min(C_i, T_i + U_i), \\ \delta_i' &= \mathbf{1}_{C_i \leq T_i + U_i}. \end{cases} \tag{2.4}$$

  In our framework, the variables (Y_i', δ_i') are observed. Moreover, C_i is independent from $T_i + U_i$ from Assumption 1. As a consequence, the survival function S_C can be consistently estimated by

$$\hat{S}_C^{(2)}(t) = \prod_{Y_i' \leq t} \left(1 - \frac{\delta_i'}{\sum_{j=1}^{n} \mathbf{1}_{Y_j' \geq Y_i'}}\right),$$

  assuming that there are no ties among the $(Y_i')_{1 \leq i \leq n}$.

Additionally, the margins F_T and F_U may not be known in practice. Several techniques may be used to estimate them, as it will be discussed in Section 2.4. To state the results in the most general form, we define $A_i = F_T(Y_i|X_i)$ and $B_i = F_U(Z_i|X_i)$, and consider that we have at our disposal pseudo-observations $(\hat{A}_i, \hat{B}_i)_{1 \leq i \leq n}$.

This leads to our final estimator of $\theta(x)$, that is

$$\hat{\theta}_h(x) = \arg\max_{\theta \in \Theta} M_{n,h}(x, \theta), \tag{2.5}$$

where

$$M_{n,h}(x,\theta) \;=\; \frac{1}{nh^d}\sum_{i=1}^{n} W_{i,n} K\left(\frac{X_i - x}{h}\right) \log c_\theta(\hat{A}_i, \hat{B}_i)\hat{w}_i(\nu_n), \tag{2.6}$$

with $W_{i,n} = \delta_i / \hat{S}_C^{(j)}(Y_i + Z_i)$ for $j = 1, 2$ depending on our ability to observe C_i or not, ν_n a sequence tending to zero which will be defined later on (see Section 3.2), and introducing a trimming function $\hat{w}_i$ defined for a sequence η_n, as

$$\hat{w}_i(\eta_n) = \mathbf{1}_{\min(\hat{A}_i, \hat{B}_i, 1-\hat{A}_i, 1-\hat{B}_i) \geq \eta_n}.$$

The presence of the trimming is required to prevent the procedure from an erratic behavior when the pseudo-observations are close to the border of the unit square. The weights $W_{i,n}$ may be seen as an approximation of $W_i = \delta_i / S_C(Y_i + Z_i)$, which, according to (2.2), would be the way we would correct the presence of the censoring if we knew exactly its distribution S_C. Let us observe that $M_{n,h}(x,\theta)$ is an estimator of $M_f(x,\theta) = M(x,\theta) f_X(x)$, where f_X denotes the density of the random vector $X \in \mathbb{R}^d$. Maximizing M or M_f with respect to θ is of course equivalent.

The practical performance of our nonparametric estimator $\hat{\theta}_h(x)$ depends on an appropriate choice of the bandwidth h. In Section 4.2.2, we present a data-driven method to select an appropriate h empirically.

2.4 Estimation of the margins

Sklar's Theorem allows to separate the marginal distributions from the dependence structure. Therefore we do not wish to specify the way the margins are estimated since various approaches may be used, possibly different from a margin to another. We will assume that this estimation has been performed separately, in a preliminary step. A parametric or semiparametric model can be put on the margins. If one only focuses on the dependence structure, a nonparametric estimation of the margins can be performed, for example using kernel estimation. In this case, one may use

$$\hat{F}_T(t|x) = \sum_{i=1}^{n} W_{i,n} \frac{K\left(\frac{X_i - x}{h'}\right)}{\sum_{j=1}^{n} K\left(\frac{X_j - x}{h'}\right)} \mathbf{1}_{Y_i \leq t}, \tag{2.7}$$

with a similar definition for $\hat{F}_U$ (replacing Y_i by Z_i). Let us note that the bandwidth h' can be different from the bandwidth used to estimate the copula parameter. Moreover, different bandwidths may be used for each of the margins. The method of equation (2.7) is the one that we use in the real data application of Section 4.2, whereas for the simulated data (Section 4.1) we use a Kaplan-Meier estimator without kernel since the marginal distributions are assumed to be independent of X.

The results that we propose in the following are valid under the condition that $\hat{A}_i = \hat{F}_T(Y_i|X_i)$ and $\hat{B}_i = \hat{F}_U(Z_i|X_i)$ are close to $A_i = F_T(Y_i|X_i)$ and $B_i = F_U(Z_i|X_i)$, but do not impose a particular method to compute the pseudo-observations. The conditions that we require (see Assumption 8) are valid for a large number of estimation techniques and at least hold for the estimator (2.7).

3 Uniform rate of convergence for $\hat{\theta}_h(x)$

As usual in nonparametric estimation, the error $\hat{\theta}_h(x)-\theta(x)$ can be decomposed into some bias term, and a stochastic term that corresponds to the fluctuations of $\hat{\theta}_h(x)$ around its central value. The most delicate convergence result to obtain is Theorem 3.2, which deals with the stochastic term, while Theorem 3.1 only studies the deterministic term, that can be handled by standard results in approximation theory.

3.1 Bias term

Let

$$\theta_h^*(x) = \arg\max_{\theta\in\Theta} \frac{1}{h^d} E\left[K\left(\frac{X-x}{h}\right)\log c_\theta(F_T(T|X), F_U(U|X))\right].$$

The difference between $\theta_h^*(x)$ and $\theta(x)$ represents the bias of the method, where the empirical mean in $M_{n,h}$ has been replaced by its limit value (we will show this convergence in the Appendix section). The aim of this section is to determine the uniform rate of convergence of this bias term on a set $\mathcal{X}$, which is assumed to be compact and strictly included in the support of the random vector X.

We use the notation $c(a,b|x) := c_{\theta(x)}(a,b)$ to denote the the conditional copula density given $X=x$. Also, let $\phi(a,b,\theta) = \log c_\theta(a,b)$, $\dot{\phi}(a,b,\theta) = \nabla_\theta \log c_\theta(a,b)$ and $\ddot{\phi}(a,b,\theta) = \nabla_\theta^2 \log c_\theta(a,b)$.

Assumptions 3 to 5 are required to obtain the convergence of the bias term. The first two can be understood as regularity assumptions on the model when x varies.

Assumption 3 *Assume that the function $(a,b,\theta)\to\phi(a,b,\theta)$ is twice continuously differentiable with respect to θ, and that for all $\theta\in\Theta, x\in\mathcal{X}$,*

$$\{|\phi(a,b,\theta)| + \|\ddot{\phi}(a,b,\theta)\|\}c(a,b|x) \le \Lambda_1(a,b), \tag{3.8}$$

with $\int \Lambda_1(a,b)dadb < \infty$.

Assumption 4 *Assume that the function $(a,b,x)\to\mathfrak{c}(a,b|x)f_X(x)$ is twice continuously differentiable with respect to x, and that for all $\theta\in\Theta, x\in\mathcal{X}$,*

$$\|\dot{\phi}(a,b,\theta)\|\cdot\|\nabla_x^2\{c(a,b|x)f_X(x)\}\| \le \Lambda_2(a,b), \tag{3.9}$$

with $\int \Lambda_2(a,b)dadb < \infty$.

The next assumption is required to ensure that the maximization problem is locally quadratically close to the true value $\theta(x)$.

Assumption 5 *Assume that there exists some $c_0 > 0$ such that, for all $x \in \mathcal{X}$, we have*

$$\forall v \in \mathbb{R}^d, \langle E[\ddot{\phi}(F_T(T|X), F_U(U|X), \theta(x))|X = x] \cdot v, v\rangle \leq -c_0\|v\|^2 \leq 0.$$

Moreover, assume that the density of X denoted by f_X is such that $f_X(x) \geq c_0$ for all $x \in \mathcal{X}$.

We now can state our result about the bias term.

Theorem 3.1 *Under Assumptions 1 to 5,*

$$\sup_{x\in\mathcal{X}} \|\theta_h^*(x) - \theta(x)\| = O(h^2).$$

This h^2 rate is classical when dealing with kernel smoothing. This rate could of course be improved by strengthening the regularity of the conditional copula function, and by considering a higher order kernel (that is a function K such that $\int u^j K(u)du = 0$ for all $j \leq k$, with k larger than 1).

The proof of Theorem 3.1 is dealt with in the Appendix section (see Section 6.1).

3.2 Stochastic term

This section presents the main theoretical result of the paper, which shows the uniform convergence of the stochastic term. The convergence rate involves two terms as given in Theorem 3.2 below; a traditional rate for kernel smoothing estimators ($[\log n]^{1/2}n^{-1/2}h^{-d/2}$), and an additional term that may become preponderant if the estimation of the margins is performed at a slow rate, or if the copula density and its derivatives behave too wildly close to the frontier of the unit square. We first begin with the assumptions required to obtain the result.

Since $\hat{\theta}_h(x)$ can be seen as a conditional version of the semiparametric copula estimator proposed by (Tsukahara, 2005) and (Genest, Ghoudi, & Rivest, 1995), the conditions required to obtain the convergence of the stochastic term are basically the same as in these two papers, with some modifications imposed by the use of smoothing and because of the censoring.

We remind that a function $r : (0,1) \to (0,\infty)$ is called $u-$shaped if r is symmetric about $1/2$ and decreasing on $(0,1/2]$. For such a $u-$shaped function r, and for $0 < \beta < 1$, define

$$r_\beta(t) = r(\beta t)\mathbf{1}_{0<t\leq 1/2} + r(1-\beta(1-t))\mathbf{1}_{1/2<t<1}.$$

A $u-$shaped function is called a reproducing $u-$shaped function if it verifies that $r_\beta \leq M_\beta r$ for all $\beta > 0$ in a neighborhood of 0, with M_β a finite constant. In the following we note $\mathcal{R}$ the set of reproducing $u-$shaped functions.

Assumptions 6 and 7 are close to assumptions A.1 to A.5 present in (Tsukahara, 2005). They ensure that the modulus of continuity of ϕ satisfies some integrability conditions, and that the derivatives of ϕ are dominated by $u-$shaped functions in order to control the explosion of these derivatives close to the border of the unit square. As shown in (Tsukahara, 2005), these conditions are satisfied by a large number of copula families.

Due to the censoring, a term $S_C(T+U)$ appears at the denominator. A similar assumption is present for example in (Gill, 1983), (Stute, 1995) or (Gribkova & Lopez, 2015). In case of heavy censoring, that is if S_C decreases too fast, the integrability conditions in Assumptions 6 and 7 may not hold. This is a classical issue in survival analysis: in such a situation, the right-tails of the distributions of T and U are rarely observed since the censoring variable tends to take small values. A solution is then to restrain the study of the distribution of T and U conditionally on $T+U \leq \tau$, where τ is a fixed bound, strictly included in the support of the variable $T+U$, though this introduces an asymptotic bias.

Assumption 6 *Assume that*

$$|\log c_\theta(a,b) - \log c_{\theta'}(a,b)| \leq R(a,b)\|\theta - \theta'\|,$$

with, for some $p > 2$ and some $\theta_0 \in \Theta$,

$$\sup_{x\in\mathcal{X}} E\left[\frac{|\log c_{\theta_0}(F_T(T|X), F_U(U|X))|^p + [R(F_T(T|X), F_U(U|X))]^p}{[S_C(T+U)]^{p-1}}\Big| X = x\right] < \infty. \quad (3.10)$$

Moreover, for a sequence η_n, let $w^*(\eta_n) = \mathbf{1}_{\min(F_T(T|X),F_U(U|X),1-F_T(T|X),1-F_U(U|X))\geq\eta_n}$ and

$$\tilde{\nu}_n = \sup_{x\in\mathcal{X}} E\left[\frac{|\log c_{\theta_0}(F_T(T|X), F_U(U|X))|^p + [R(F_T(T|X), F_U(U|X))]^p}{[S_C(T+U)]^{p-1}}(1 - w^*(2\nu_n))\Big| X = x\right], \quad (3.11)$$

the speed of convergence of the expectation on the border of the unit square.

We then define $\tilde{\tilde{\nu}}_n = \nu_n + \tilde{\nu}_n$. *Assume that*

$$(n^{1/2}h^{d/2}\log(n)^{-1/2}\tilde{\tilde{\nu}}_n)^{p-1}h^d \underset{n\to\infty}{\to} +\infty.$$

Finally, assume that, for some $p' > 1$,

$$E\left[\frac{|\log c_{\theta_0}(F_T(T|X), F_U(U|X))|^{p'} + R(F_T(T|X), F_U(U|X))^{p'}}{S_C(T+U)^{2p'-1}} w^*(\nu_n/2)\right] = O(n^{1/2}\tilde{\tilde{\nu}}_n), \tag{3.12}$$

and that $(n^{1/2}\tilde{\tilde{\nu}}_n)^{p'-1}h^d \underset{n\to\infty}{\to} +\infty$.

Let us remark that thanks to the equation (3.10), $\tilde{\nu}_n$ (and then $\tilde{\tilde{\nu}}_n$) tends to 0 from Lebesgue's dominated convergence theorem. Moreover, in most applications, the functions $\log c_{\theta_0}$ and R inside the expectation in equation (3.11) do not tend to 0 on the border of the unit square, so that $\nu_n = O(\tilde{\nu}_n)$. In these cases, the speed $\tilde{\tilde{\nu}}_n$ is typically of the same order of magnitude as $\tilde{\nu}_n$.

Assumption 7 *Assume that there exist functions* r_j *and* $\tilde{r}_j$ *in* $\mathcal{R}$, $j = 1, 2$, *such that*

$$\begin{aligned} \|\dot{\phi}(a,b,\theta)\| &\leq r_1(a)r_2(b), \\ |\partial_a\phi(a,b,\theta)| + \|\partial_a\dot{\phi}(a,b,\theta)\| &\leq \tilde{r}_1(a)r_2(b), \\ |\partial_b\phi(a,b,\theta)| + \|\partial_b\dot{\phi}(a,b,\theta)\| &\leq \tilde{r}_2(b)r_1(a). \end{aligned}$$

Considering ϵ_n *a sequence of positive numbers tending to zero (which is specified in Assumption 8 below), assume further that for some* $p'' > 1$,

$$E\left[\left(\frac{\tilde{r}_1(F_T(T|X)^{p''}r_2(F_U(U|X)^{p''} + \tilde{r}_2(F_U(U|X)^{p''}r_1(F_T(T|X)^{p''}}{S_C(T+U)^{p''-1}}\right) w^*(\nu_n)\right] = O(\epsilon_n^{-1}\tilde{\tilde{\nu}}_n) \tag{3.13}$$

and that

$$(\epsilon_n^{-1}\tilde{\tilde{\nu}}_n)^{p''-1}h^d \underset{n\to\infty}{\to} +\infty. \tag{3.14}$$

The next assumption concerns the estimation of the margins. The results we provide may hold for different strategies of estimation of the margins (nonparametric, semiparametric or parametric). We only require a rate of consistency for the margins, and a comparison of this rate with the speed $\tilde{\tilde{\nu}}_n$.

Assumption 8 *Assume that*

$$\sup_{1\leq i\leq n} |\hat{A}_i - A_i| + |\hat{B}_i - B_i| = O_P(\varepsilon_n),$$

with $\varepsilon_n = o(\nu_n)$.

We now can state the main result of this section, which is proven in Section 6.7.

Theorem 3.2 *Assume that*

$$nh^d \log(n)^{-1} \underset{n\to\infty}{\to} \infty. \tag{3.15}$$

Then, under Assumptions 1 to 8,

$$\sup_{x\in\mathcal{X}} \|\hat{\theta}_h(x) - \theta^*_h(x)\| = O_P(\tilde{\tilde{\nu}}_n + [\log n]^{1/2} n^{-1/2} h^{-d/2}).$$

In this result, we can see that the rate of convergence of the stochastic term can be decomposed into the classical convergence rate $[\log n]^{1/2}n^{-1/2}h^{-d/2}$ (that is the standard convergence rate one would obtain if the true margins were known), and a second term $\tilde{\tilde{\nu}}_n$ which combines the convergence rate of the margins and the difficulty to replace the pseudo-observations by their limit values near the frontier of the unit square. Let us recall that this term $\tilde{\tilde{\nu}}_n$ can be close to the rate ε_n of estimation of the margins, still with $\varepsilon_n = o(\tilde{\tilde{\nu}}_n)$ (in fact, $\varepsilon_n = o(\nu_n)$ thanks to Assumption 8, and $\nu_n \le \tilde{\tilde{\nu}}_n$) by definition of $\tilde{\tilde{\nu}}_n$). Moreover we can chose $\tilde{\tilde{\nu}}_n$ since it depends on the trimming function we use.

To clarify the situation, let us discuss the conditions we need in order to achieve the ideal rate $\tilde{\tilde{\nu}}_n = [\log n]^{1/2}n^{-1/2}h^{-d/2}$. Achieving this rate essentially depends on ε_n and on the parameter p'' involved in (3.13) and condition (3.14).

First assume that the margins have been estimated with the rate $\varepsilon_n = n^{-1/2}$ (e.g. a parametric model is put on the margins). Take $\tilde{\tilde{\nu}}_n = [\log n]^{1/2}n^{-1/2}h^{-d/2}$. In this case, the condition (3.14) in Assumption 7 holds if $p'' \ge 3$. A condition on p'' corresponds to a constraint on the functions $\tilde{r}_j$ and r_j ($j = 1, 2$), that is a constraint on the domination of the derivatives of the log-likelihood near the frontier of the unit square (and on the strength of the censoring due to the presence of S_C at the denominator). If the explosion of the functions $\tilde{r}_j$ and r_j ($j = 1, 2$) is too strong, that is if (3.13) only holds for a $p'' < 3$, then $\tilde{\tilde{\nu}}_n$ should be increased in order to achieve condition (3.14). In this latter case the term $\tilde{\tilde{\nu}}_n$ then becomes preponderant in the convergence rate of Theorem 3.2.

On the other hand, if ε_n is slower than $n^{-1/2}$ (e.g. the margins are estimated non-parametrically), then $p'' = 3$ is not sufficient if we want condition (3.14) to be satisfied by $\tilde{\tilde{\nu}}_n = [\log n]^{1/2}n^{-1/2}h^{-d/2}$. An higher value of p'' is required. The slower ε_n converges (without being slower than $[\log n]^{1/2}n^{-1/2}h^{-d/2}$), the higher p'' should be so that condition (3.14) holds. In an extremely favorable situation, that is if $\tilde{r}_j$ and r_j are bounded near the border of the unit square, and if the censoring is not too heavy (i.e. S_C decreases slowly), then p'' can be taken as $+\infty$ and $\tilde{\tilde{\nu}}_n$ may be chosen close to ε_n.

To summarize, in order to achieve the optimal convergence rate for the stochastic term, the rate of convergence for the estimation of the margins should be fast enough compared to the explosion of the log-likelihood and its derivatives near the boundaries of the unit square. If the required conditions are not satisfied, then a deterioration of the rate of convergence appears.

4 Experiments of the method using data

In the following part, the method developed in Section 2 to estimate the conditional dependence parameter of a copula is illustrated numerically.

Four families of parametric copulas (Gaussian, Clayton, Gumbel and Frank) are tested to model the dependence between T and U :

- the Gaussian copula family $\mathcal{C}^1 = \{\mathfrak{C}^1_\theta : \theta \in [-1;1]\}$, where $\mathfrak{C}^1_\theta(a,b) = g_\theta\left(g^{-1}(a) + g^{-1}(b)\right)$, with g_θ is the cumulative distribution function of a bivariate Gaussian vector (V_1, V_2) – with mean $E(V_1, V_2) = (0,0)$, marginal variances $\mathrm{Var}(V_1) = \mathrm{Var}(V_2) = 1$ and covariance $\mathrm{Cov}(V_1, V_2) = \theta$ – and g^{-1} is the inverse cumulative distribution function of a standard normal random variable.
- the Clayton copula family $\mathcal{C}^2 = \{\mathfrak{C}^2_\theta : \theta > 0\}$, with $\mathfrak{C}^2_\theta(a,b) = \left(a^{-\theta} + b^{-\theta} - 1\right)^{-1/\theta}$
- the Gumbel copula family $\mathcal{C}^3 = \{\mathfrak{C}^3_\theta : \theta \geq 1\}$, with

$$\mathfrak{C}^3_\theta(a,b) = \exp\left[-\left((-\log(a))^\theta + (-\log(b))^\theta\right)^{1/\theta}\right]$$

- the Frank copula family $\mathcal{C}^4 = \{\mathfrak{C}^4_\theta : \theta \in \mathbb{R}\backslash\{0\}\}$ with

$$\mathfrak{C}^4_\theta(a,b) = -\frac{1}{\theta}\log\left[1 + \frac{(\exp(-\theta a) - 1)(\exp(-\theta b) - 1)}{\exp(-\theta) - 1}\right]$$

4.1 Simulated data

4.1.1 Data setting

We first consider simulated data for the model testing. The covariate vector X is taken as one dimensional, with uniform law on $[0,1]$. On the other hand, the successive times T and U are independent of X and follow log-normal distributions : $\log(T) \sim \mathcal{N}(0,1)$ (resp. $\log(U) \sim \mathcal{N}(0,1)$).

Given a sample $(T_i, U_i)_{1\leq i\leq n}$, let $n_c = \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{T_i<T_j, U_i<U_j}$ be the number of concordant pairs in the sample, and $n_d = \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{T_i<T_j, U_i>U_j}$ be the number of discordant pairs. Then the Kendall tau (Kendall, 1938) between T and U expresses as

$$\tau_n = \frac{n_c - n_d}{n_c + n_d}. \tag{4.16}$$

Its expected value is given by $\tau = 2E\left[\mathbf{1}_{(\tilde{T}_1-\tilde{T}_2)(\tilde{U}_1-\tilde{U}_2)>0}\right] - 1$ where $(\tilde{T}_1, \tilde{U}_1)$ (resp. $(\tilde{T}_2, \tilde{U}_2)$) follows the same law as (T, U) and $(\tilde{T}_1, \tilde{U}_1)$ is independent of $(\tilde{T}_2, \tilde{U}_2)$.

The dependence between T and U in the simulations is set using the Kendall τ through the relation :

$$\log(\tau/1 - \tau) = a + b \cdot X \tag{4.17}$$

with $a = -3$ and $b = 4$. Indeed, for any of the four copula families we consider, there is a bijection between the copula parameter θ and the expected value of the Kendall tau (Genest & MacKay, 1986). Then, making the assumption that the dependence between T and U belongs to some copula family, it is enough to set a value for τ to specify the conditional copula between T and U (see Section 4.1.2).

Let $q = P(T + U > C)$ the censoring rate of the variable $T + U$. The influence of q on the results is studied in the experiments. To do so, the distribution of the censoring variable C, whose log is assumed to follow an exponential distribution, is adjusted so that the desired censoring rate is achieved among the simulated data ($q = 0.3$ or $q = 0.5$).

4.1.2 Description of the experiments

Each simulated dataset consists of $n = 1000$ observations. For each copula family $(\mathcal{C}^l)_{l=1,\ldots,4}$, we estimate the copula parameter θ at five different x values, which correspond to five distinct values $\tau(x)$ and $\theta^l(x)$ that are summarized in Tab. 1. The marginal laws of T and U are estimated with the Kaplan-Meier estimator, as well as the survival function of the censoring S_C, used to compute the weights $W_{i,n}$ (see equation (2.6)). Also, we use a quadratic kernel : $K(u) = 15/16 \cdot (1 - u^2)^2 \cdot \mathbf{1}_{|u|\leq 1}$ to localize the estimation in x neighborhoods, and different candidate values for the bandwidth parameter h (see Fig. 1). For each bandwidth h, this results in estimators $(\hat{\theta}^l_h(x))_{l=1,..4}$ of the conditional copula parameters at the point x, for the copula families $(\mathcal{C}^l)_{l=1,..,4}$. We can compare them with the exact parameters $(\theta^l(x))_{l=1,..4}$ by computing the quadratic errors at each x value : $\forall\ l = 1, \ldots, 4,\ \epsilon^l_h(x) = (\hat{\theta}^l_h(x) - \theta^l(x))^2$. Then the error of a copula model $\mathcal{C}^l$ with bandwidth h is taken as the average of $\epsilon^l_h(x)$ over the five x values: $\forall\ l = 1, \ldots, 4,\ \epsilon^l_h = 1/5 \sum_{x\in\{x\ values\}} \epsilon^l_h(x)$.

4.1.3 Results

The results of the simulations are shown in Fig. 1. We represent the mean values of ϵ_h^l, computed over 100 i.i.d. replications of the above procedure of data simulation and copula fitting, as a function of h. The error is split into a bias part and a variance part, which are known to form an additive decomposition of the total error, and that we also represent in Fig. 1.

All the six graphics present the same pattern in a u-shape for the total error. For small h, the bias is low but the high variance of the estimator leads to a bad precision of the estimator overall. The situation is reversed when h takes big values, with low variance and high bias for the estimator. The optimal value for the bandwidth then has to be taken among middle values of h.

As expected we also remark that the increase of the rate of censoring deteriorates the precision of the estimation.

Although we have shown in Section 3 that the estimation procedure we propose is asymptotically consistent, it was important to verify that the method behaves well with finite samples. In this regard, the results of Fig. 1 show that, for the four copula families, reasonable errors could be achieved with our method. In the four cases, a value of h equal to 0.15 is a good compromise between the bias and the variance, so that the total error is close from its minimum.

x	0.10	0.30	0.50	0.70	0.90
Kendall tau : $\tau(x)$	0.07	0.14	0.27	0.45	0.65
Gaussian : $\theta^1(x)$	0.11	0.22	0.41	0.65	0.85
Clayton : $\theta^2(x)$	0.15	0.33	0.74	1.64	3.64
Gumbel : $\theta^3(x)$	1.07	1.17	1.37	1.82	2.82
Frank : $\theta^4(x)$	0.62	1.30	2.57	4.90	9.29

Tab. 1: Values of the Kendall tau and of the exact copula parameters at the five different x points

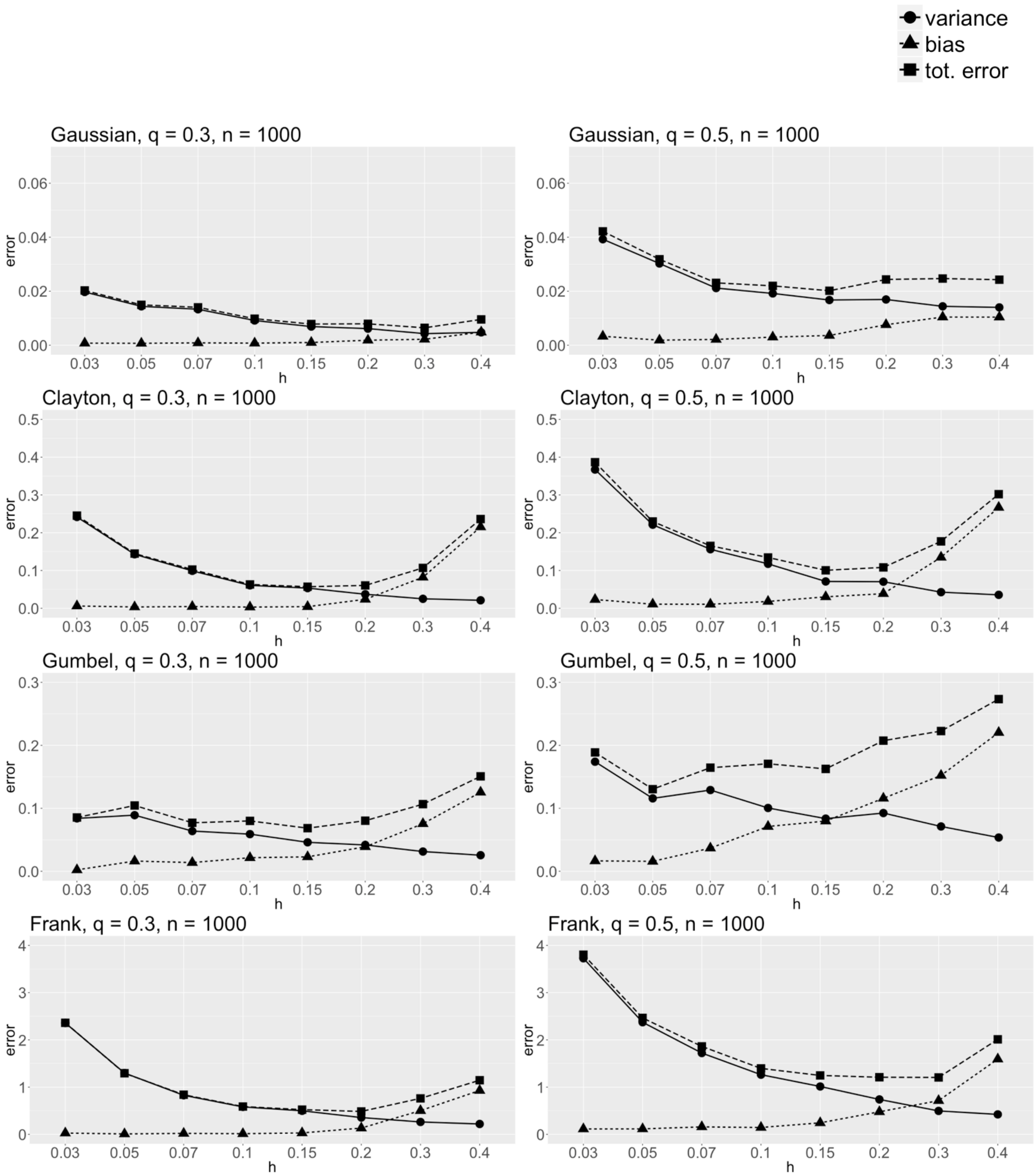

Fig. 1: For each copula family $\mathcal{C}^l$ and each rate of censoring $q \in \{0.3, 0.5\}$, mean values of ϵ_h^l (100 repetitions) as a function of h, and their decompositions into bias and variance terms. Size of the simulated datasets: n = 1000.

4.2 Application on real data

4.2.1 Description of the data

We have applied our estimation methodology to data provided by a broker of health insurance contracts. In the context of the study, the time T corresponds to the effective time of a contract (i.e. the duration between the date of subscription and the date of effect of the contract), whereas U is the termination time of the contract (i.e. the duration between the date of effect and the date of termination of the contract). As the churn of a contract holder impacts the commission received by a broker for this contract, it is important from the broker's point of view to understand the dependence between the successive times T and U, and especially to measure it given some characteristics X of the contract holder. Indeed, evaluating such dependence allows to fairly compare two underwriters that have been selling insurance products to customers with different characteristics, by weighting the performances of the underwriters with a value factor (that is a factor which gives the expected value of a given customer). This study of the dependence should take into account the censoring that is present in the data, which is due to the fact that any contract may stop to be under observation before T, or U, occurs (e.g. due to the end of the study, or the end of the observation period). In the following, we tackle the problem of the estimation of the dependence between T and U, given the age of the contract holder at the subscription.

The dataset that we study has 224897 entries, recorded from 1^{st} October 2009 to 31^{th} July 2016. We observe that the dataset can be split into two parts according to whether a contract date of effect is on January 1, or not. Indeed, contracts coming into effect on January 1 represent more than one third of the database, and are generally associated with longer delays before the date of effect. For those contracts, the dependence between T and U is stronger. We focus on this part of the database in the following application.

4.2.2 Methodology in the experiments

For the modelling of the dependence between T and U conditionally on X, which is in our situation the age of the contract holder at the subscription, we analyze the results given by the Gaussian, Clayton, Gumbel and Frank parametric copula families (see Section 4).

The estimation is made according to the equation (2.5). We use a quadratic kernel as in the simulated data experiments. The survival function of the censoring is estimated thanks to the estimator of the equation (2.3). Indeed, in this application the censoring variable is the age of the contract (from its subscription to the end of the observation period), hence it is always observed since the perimeter of the study corresponds to contracts that have been subscribed. The pseudo observations $\hat{A}_i$ and $\hat{B}_i$ are derived using local Kaplan-Meier estimators of the marginal distributions of T and U, whose formula are given at equation (2.7).

Assuming no prior on the direction of the dependence between T and U, our fitting procedure for copula needs to be adapted to cases where the sign of the dependence changes as the age of the policyholder varies. This requirement is not problematic for the Gaussian copula and the Frank copula, which may model positive and negative dependences, but the Clayton copula and the Gumbel copula can only model positive dependences. Moreover, the Gaussian copula and the Frank copula are symmetric (i.e. the copula densities satisfy $c_\theta(a,b) = c_\theta(1-b, 1-a)$), whereas the Clayton copula and the Gumbel copula are not. Hence, for the latter two copulas, we need to fit the copula four times to cover all possible dependence relations between T and U: we successively fit the copula on the pseudo-observations $(\hat{A}_i, \hat{B}_i)$, $(1-\hat{A}_i, \hat{B}_i)$, $(\hat{A}_i, 1-\hat{B}_i)$ and $(1-\hat{A}_i, 1-\hat{B}_i)$ where $\hat{A}_i$ and $\hat{B}_i$ are defined in Section 2.4. This gives four candidate maximums of the criteria $M_{n,h}$ (equation (2.6)), from which we can select the highest maximum. The dependence between T and U is then positive if the maximum corresponds to the case $(\hat{A}_i, \hat{B}_i)$ or $(1-\hat{A}_i, 1-\hat{B}_i)$, and negative otherwise.

In the following numerical experiments, we use a train-test approach with 100 repetitions. For each of the 100 iterations, two non-overlapping subsamples $\mathcal{D}_{tr}$ (train) and $\mathcal{D}_{te}$ (test), of size 10000 ($n_{tr} = n_{te} = 10000$), are drawn from the initial dataset. On the training set, and for each age x from 20 to 80, we apply our method to compute $(\hat{\theta}^l_{h,tr}(x))_{l=1,..4}$, the estimates of the conditional copula parameters corresponding to the copula families $(\mathcal{C}^l)_{l=1,..,4}$ and to the values $h = 1, 5, 10, 20, 40$. On the test set, we estimate the conditional Kendall tau between T and U, at each age x. This is done using the following kernel estimator of the Kendall tau : for all test observations i, let $W_{i,n}(x) = W_{i,n}K((X_i - x)/h_1)$, where $W_{i,n}$ is defined at equation (2.6), $(X_i)_{i=1,..,n_{te}}$ denotes the age values and $h_1 = 1$. Then let $n^w_c = \sum_{i=1}^{n_{te}} \sum_{j=1}^{n_{te}} W_{i,n}(x) \mathbf{1}_{T_i<T_j, U_i<U_j}$ and $n^w_d = \sum_{i=1}^{n_{te}} \sum_{j=1}^{n_{te}} W_{i,n}(x) \mathbf{1}_{T_i<T_j, U_i>U_j}$; we can estimate the conditional Kendall tau at the age x by $\hat{\tau}^w_{h_1,te}(x) = (n^w_c - n^w_d)/(n^w_c + n^w_d)$. The conditional Kendall tau is also estimated on the training set ($\hat{\tau}^w_{h_1,tr}(x)$), using the same method.

To compare the results of the different copula families and to identify the optimal value for the bandwidth parameter h, we compute train and test errors based on Kendall tau estimates. Thanks to the one to one relations between the parameter θ and the Kendall tau for the copula families $(\mathcal{C}^l)_{l=1,..,4}$ (see Section 4.1.1), we deduce from the estimators $(\hat{\theta}^l_{h,tr}(x))_{l=1,..4}$ train estimates of the conditional Kendall tau $(\hat{\tau}^l_{h,tr}(x))_{l=1,..4}$. Then we define the test (resp. train) error at a point x as $\epsilon^l_{h,te}(x) = (\hat{\tau}^l_{h,tr}(x) - \hat{\tau}^w_{h_1,te}(x))^2$ (resp. $\epsilon^l_{h,tr}(x) = (\hat{\tau}^l_{h,tr}(x) - \hat{\tau}^w_{h_1,tr}(x))^2$), and thereafter the test (resp. train) error of the copula model as the aggregated error over all x values : $\epsilon^l_{h,te} = \sum_{x=20}^{80} \bar{w}_{x,te}\epsilon^l_{h,te}(x)$ (resp. $\epsilon^l_{h,tr} = \sum_{x=20}^{80} \bar{w}_{x,tr}\epsilon^l_{h,tr}(x)$), with $\bar{w}_{x,te} = w_{x,te}/\sum_{x=20}^{80} w_{x,te}$ and $w_{x,te} = \sum_{i\in\mathcal{D}_{te}} K((X_i - x)/h_1)$ (resp. $\bar{w}_{x,tr} = w_{x,tr}/\sum_{x=20}^{80} w_{x,tr}$ and $w_{x,tr} = \sum_{i\in\mathcal{D}_{tr}} K((X_i - x)/h_1)$).

4.2.3 Results

We represent in Fig. 2 box plots of the train and test square root errors $((\epsilon^{\,l}_{h,tr})^{1/2})_{l=1,..,4}$ and $((\epsilon^{\,l}_{h,te})^{1/2})_{l=1,..,4}$ measured over the 100 iterations, for each copula family and each bandwidth value h. The results show that the Frank copula achieves the lowest error (both train and test) on the data, and should be privileged to model the dependence between the two durations. Also, we observe that the test error is generally minimal for $h = 20$, which indicates that 20 years is the appropriate time scale to observe trends in the conditional dependence.

On Fig. 3, we show the average values of the conditional copula parameters for the fitted copulas, as well as 95% confidence intervals for the exact parameters. The graphic for the Frank copula indicates that the strength of the dependence between T and U decreases between the ages 20 and 40, and then increases from 40 to 65 (i.e. until the age of retirement), before it starts to decrease again. This means that young adults (20-30) and seniors (55-75) are more likely than other age categories to have their decision to terminate their contract impacted by the effective time of the contract, in the sens that a long effective time causes an higher probability of rapid termination of the contract.

The Fig. 4 shows that a different evolution of the dependence is observed for the contracts associated with higher product's level (i.e. better guarantees). These contracts exhibit a strong negative dependence overall, which becomes even stronger for the people aged over 60.

5 Conclusion

In this paper, we proposed a methodology for estimating a conditional copula function under random censoring, when the two variables linked through the copula are successive

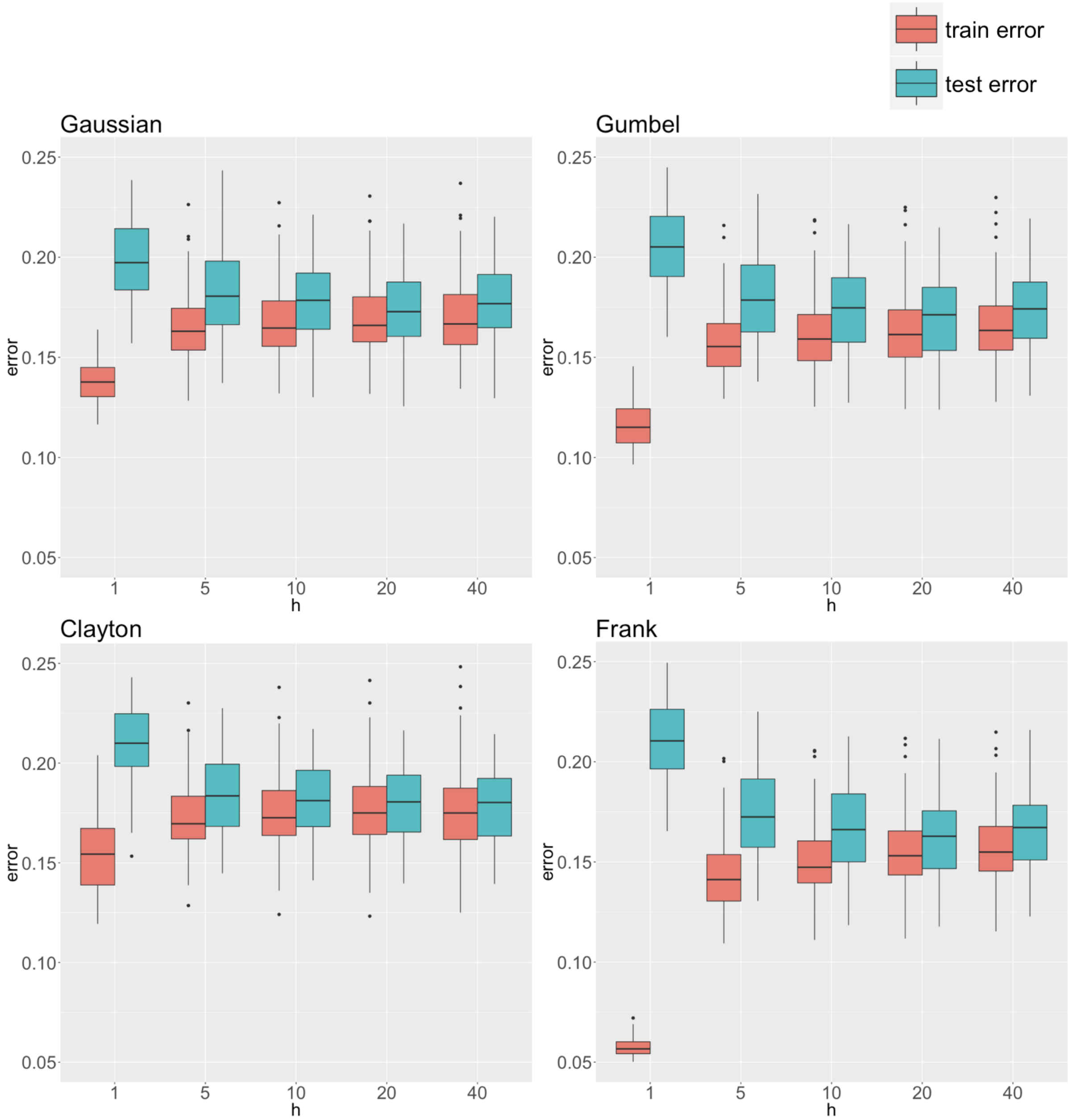

Fig. 2: For each copula family and each bandwidth value h, box plot of the train and test square root errors $((\epsilon^l_{h,tr})^{1/2})_{l=1,..,4}$ and $((\epsilon^l_{h,te})^{1/2})_{l=1,..,4}$ ($n = 10000$, 100 repetitions).

times. The model is semiparametric, since we assume that the conditional copula does not leave a parametric family, but with a nonparametric assumption on the dependence of the association parameter on the covariates. From a numerical point of view, the procedure is simple, since it relies on a weighted log-likelihood approach. The kernel smoothing approach can be extended to local linear modeling, as in (Gijbels et al., 2011) in presence of complete data. Let us mention that our results hold with only standard conditions

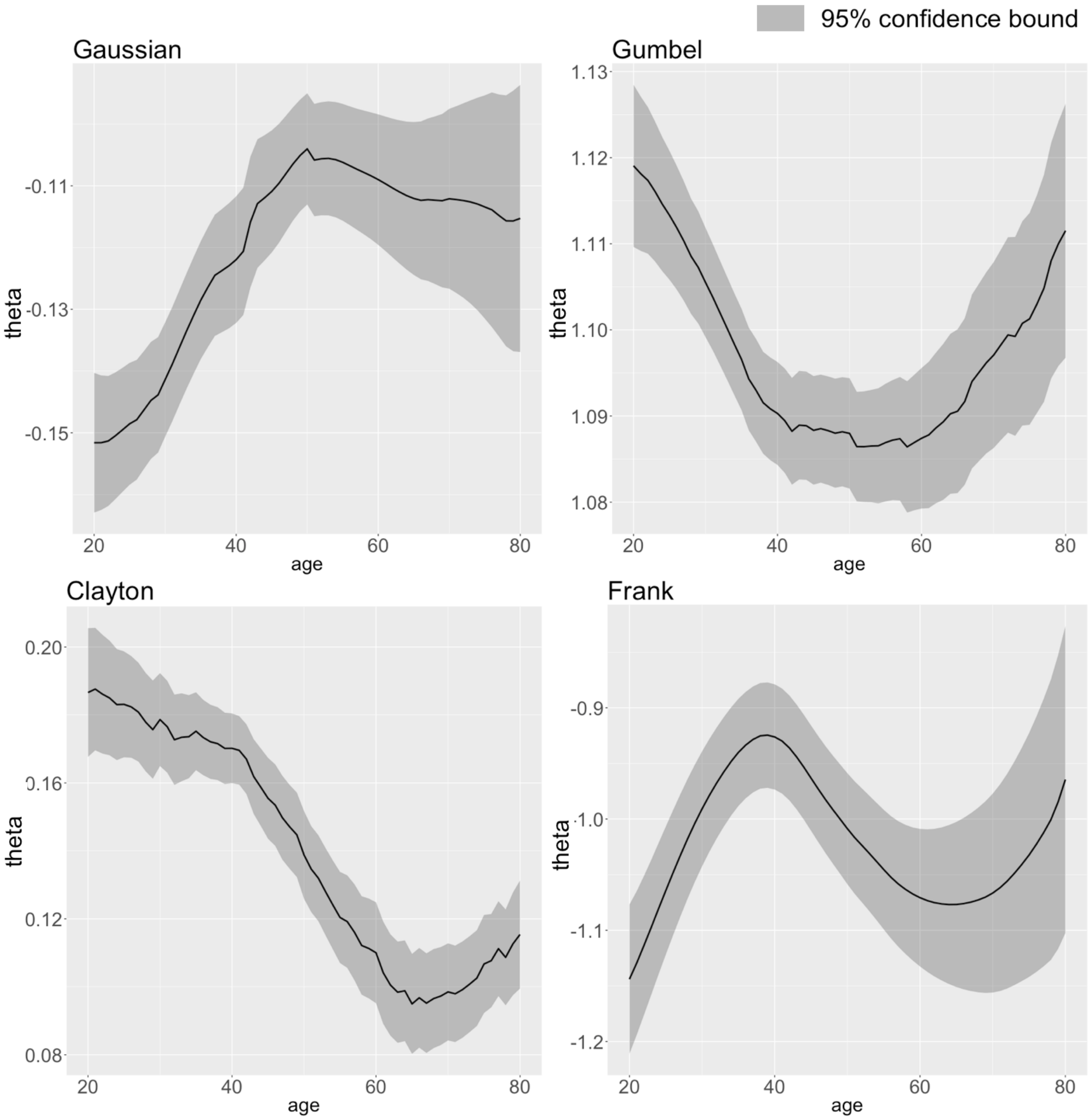

Fig. 3: For each copula family, mean value of the conditional copula parameter as a function of the age x ($h = 20$, $n = 10000$, 100 repetitions). As we notice in Section 4.2.2, the Gumbel copula and the Clayton copula don't vary in the same direction as the Gaussian copula and the Frank copula.

on the estimation of the margins, giving a relative freedom to practitioners on how they want to perform this estimation. Moreover, we provide conditions on the censoring which allow to understand the behavior of the method even in the tail of the distribution (that is near the right and upper corner of the unit square when looking at the copula). This

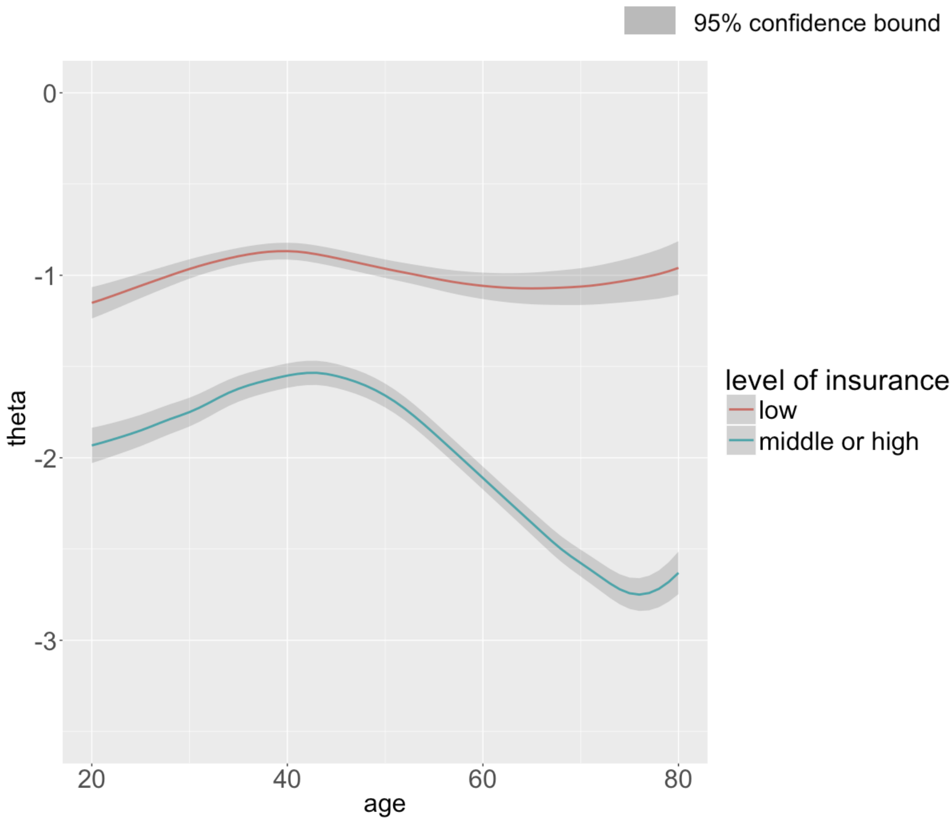

Fig. 4: Impact of the variable *level of insurance* on the conditional dependence between T and U, given the age of the prospect (Frank copula, $h = 20$, 100 repetition).

indication is precious since under random censoring an important question is to control the behavior of the method near the right tail.

Code

The code used for producing the results is available at the address :
`github.com/YohannLeFaou/copula-successive-duration-times`.

Acknowledgements

We would like to thank the company Forsides that supports us in this project, and the company Santiane who provided the data that served for the experiments.

6 Appendix: technical results

This appendix section gathers the proofs of Theorems 3.1 (Section 6.1) and 3.2 (Sections 6.2 to 6.7). The auxiliary results required to prove Theorem 3.2 consist of replacing the weights $W_{i,n}$ by W_i in the criterion $M_{n,h}$ (Section 6.3), dealing with the trimming function (Section 6.4), replacing the pseudo-observations by their limit values (Section 6.5), and providing a control of the stochastic term uniformly in x (Section 6.6).

The uniform consistency result is then obtained by applying the results of (Einmahl & Mason, 2005) on kernel smoothing (Section 6.9).

6.1 Proof of Theorem 3.1 (Bias term)

Let

$$M_f^{(c)}(x,\theta) = \frac{1}{h^d}E\left[K\left(\frac{X-x}{h}\right)\log c_\theta(F_T(T|X),F_U(U|X))\right].$$

First observe that

$$\int K(v)\phi(F_T(t|x+hv),F_U(u|x+hv),\theta)dF(t,u|x+hv)f_X(x+hv)dv =$$
$$\int K(v)\phi(a,b,\theta)\mathfrak{c}(a,b|x+hv)f_X(x+hv)dvdadb.$$

The right-hand side converges towards

$$\int \phi(a,b,\theta)\mathfrak{c}(a,b|x)dadbf_X(x) = M(x,\theta)f_X(x) \tag{6.18}$$

as $h\to 0$, uniformly in θ and x, from Lebesgue's dominated convergence theorem and Assumption 3. Thus,

$$\sup_{x\in\mathcal{X}}\|\theta_h^*(x)-\theta(x)\| \underset{h\to 0}{=} o(1). \tag{6.19}$$

Next, we use a Taylor expansion to show the speed of convergence of $\sup_x\|\theta_h^*(x)-\theta(x)\|$. For all $j=1,\ldots,k$, let $\nabla_{\theta_j}M_f^{(c)}(x,\theta)$ the j^{th} component of the gradient $\nabla_\theta M_f^{(c)}(x,\theta)$. Since $\nabla_\theta M_f^{(c)}(x,\theta_h^*(x))=0$, we have

$$\forall j=1,\ldots,k,\ \langle\nabla_\theta\nabla_{\theta_j}M_f^{(c)}(x,\tilde\theta_h^j(x)),\theta_h^*(x)-\theta(x)\rangle = -\nabla_{\theta_j}M_f^{(c)}(x,\theta(x)), \tag{6.20}$$

with $\forall j,\ \tilde\theta_h^j(x)\in[\theta_h^*(x);\theta(x)]$. For the left hand side of (6.20) we have,

$$\|\nabla_\theta\nabla_{\theta_j}M_f^{(c)}(x,\tilde\theta_h^j(x))-\nabla_\theta\nabla_{\theta_j}M(x,\theta(x))f_X(x)\| \le \|\nabla_\theta\nabla_{\theta_j}M_f^{(c)}(x,\tilde\theta_h^j(x))-\nabla_\theta\nabla_{\theta_j}M_f^{(c)}(x,\theta(x))\|$$
$$+\|\nabla_\theta\nabla_{\theta_j}M_f^{(c)}(x,\theta(x))-\nabla_\theta\nabla_{\theta_j}M(x,\theta(x))f_X(x)\|.$$

Clearly $\sup_x \|\nabla_\theta\nabla_{\theta_j} M_f^{(c)}(x, \tilde{\theta}_h^j(x)) - \nabla_\theta\nabla_{\theta_j} M_f^{(c)}(x, \theta(x))\| = o(1)$ by (6.19) and the smoothness condition of ϕ in Assumption 3.
Moreover, $\sup_x \|\nabla_\theta\nabla_{\theta_j} M_f^{(c)}(x, \theta(x)) - \nabla_\theta\nabla_{\theta_j} M(x, \theta(x)) f_X(x)\| = o(1)$ using the same kind of development as in (6.18) (applied to $\nabla_\theta\nabla_{\theta_j}\phi(a, b, \theta)$ instead of $\phi(a, b, \theta)$), and Lebesgue's theorem.
Hence, one gets

$$\sup_x \|\nabla_\theta\nabla_{\theta_j} M_f^{(c)}(x, \tilde{\theta}_h^j(x)) - \nabla_\theta\nabla_{\theta_j} M(x, \theta(x)) f_X(x)\| \underset{h\to 0}{=} o(1),$$

so that using Assumption 5 and combining the k equations of (6.20), we have for h sufficiently small,

$$\forall x \in \mathcal{X}, \ \sum_{j=1}^{k} \left|\langle \nabla_\theta\nabla_{\theta_j} M_f^{(c)}(x, \tilde{\theta}_h^j(x)), \theta_h^*(x) - \theta(x)\rangle\right| \geq c_0' \|\theta_h^*(x) - \theta(x)\|, \tag{6.21}$$

with $c_0' > 0$ a given constant.

Moreover, we have for the right hand side of (6.20),

$$\forall j, \ \sup_{x\in\mathcal{X}} \left|\nabla_{\theta_j} M_f^{(c)}(x, \theta(x))\right| = O(h^2). \tag{6.22}$$

Indeed, a second order Taylor expansion leads to

$$\nabla_\theta M_f^{(c)}(x, \theta(x)) = \frac{h^2}{2} \int K(v)\dot{\phi}(a, b, \theta(x))\langle \nabla_x^2\{\mathfrak{c}(a, b|\tilde{x}) f_X(\tilde{x})\}\cdot v, v\rangle dv da db,$$

for some $\tilde{x}$ between x and $x + hv$, and the right-hand side is $O(h^2)$ uniformly in x thanks to Assumption 4.
Combining equations (6.21) and (6.22) then leads to

$$\sup_x \|\theta_h^*(x) - \theta(x)\| \underset{h\to 0}{=} O(h^2).$$

6.2 Consistency of the Stochastic term

Before showing the convergence rate of $\hat{\theta}_h(x)$, we first show its uniform consistency in Proposition 6.1, by looking at its difference with the bias term $\theta_h^*(x)$.

Proposition 6.1 *Under the assumptions of Theorem 3.2,*

$$\sup_x \|\hat{\theta}_h(x) - \theta_h^*(x)\| \underset{n\to\infty}{=} o_P(1). \tag{6.23}$$

Proof. To show (6.23), first decompose

$$|M_{n,h}(x,\theta) - M_f^{(c)}(x,\theta)| \leq |M_{n,h}(x,\theta) - M_{n,h}^*(x,\theta)| + |M_{n,h}^*(x,\theta) - M_f^{(c)}(x,\theta)|,$$

where

$$M_{n,h}^*(x,\theta) = \frac{1}{nh^d}\sum_{i=1}^n W_{i,n} K\left(\frac{X_i - x}{h}\right) \log c_\theta(A_i, B_i)\hat{\omega}_i(\nu_n). \tag{6.24}$$

From Lemma 6.4, we get $\sup_{x\in\mathcal{X},\theta\in\Theta} |M_{n,h}(x,\theta) - M_{n,h}^*(x,\theta)| = o_P(1)$. Next, from Lemma 6.2 and Lemma 6.3,

$$M_{n,h}^*(x,\theta) = \frac{1}{nh^d}\sum_{i=1}^n W_i K\left(\frac{X_i - x}{h}\right) \log c_\theta(A_i, B_i) + o_P(1),$$

uniformly in x and θ, using Assumption 6 and equation (3.15).

Then, Theorem 4 of (Einmahl & Mason, 2005) applies using Assumption 6 to show that $\sup_{x\in\mathcal{X},\theta\in\Theta} |\frac{1}{nh^d}\sum_{i=1}^n W_i K\left(\frac{X_i-x}{h}\right)\log c_\theta(A_i,B_i) - M_f^{(c)}(x,\theta)| = o_P(1)$, which concludes the proof. ■

6.3 Estimation of S_C

Lemma 6.2 below shows that, provided that an integrability condition holds, the weights $W_{i,n}$ (relying on the estimation $\hat{S}_C$ of the distribution of the censoring) are asymptotically equivalent to the weights W_i.

Lemma 6.2 *Let $\mathcal{F}$ denote a class of function ψ such that, $\forall\psi \in \mathcal{F}$, $|\psi(y,z,x)| \leq \Psi(y,z,x)$. Assume equation (3.15) holds, and that, for some $p' > 1$,*

$$E\left[\frac{\Psi(T,U,X)^{p'}}{S_C(T+U)^{2p'-1}} w^*(\nu_n/2)\right] = O(n^{1/2}\tilde{\tilde{\nu}}_n), \tag{6.25}$$

with $(n^{1/2}\tilde{\tilde{\nu}}_n)^{p'-1}h^d \underset{n\to\infty}{\to} +\infty$.
Then,

$$\sup_{x\in\mathcal{X},\psi\in\mathcal{F}} \left|\frac{1}{nh^d}\sum_{i=1}^n (W_{i,n} - W_i)\psi(Y_i,Z_i,X_i)K\left(\frac{X_i-x}{h}\right)\hat{\omega}_i(\nu_n)\right| = O_P(\tilde{\tilde{\nu}}_n). \tag{6.26}$$

Proof. Let $\mathfrak{S}_{(1)} \leq \mathfrak{S}_{(2)} \leq ... \leq \mathfrak{S}_{(n)}$ denote the order statistics of $\mathfrak{S}_i = Y_i + Z_i$. Observe that

$$\left|\frac{1}{nh^d}\sum_{i=1}^n (W_{i,n} - W_i)\psi(Y_i,Z_i,X_i)K\left(\frac{X_i-x}{h}\right)\hat{\omega}_i(\nu_n)\right| \leq \sup_{t\leq\mathfrak{S}_{(n)}}\left|\hat{S}_C(t) - S_C(t)\right| \sup_{t\leq\mathfrak{S}_{(n)}}\left|\frac{S_C(t)}{\hat{S}_C(t)}\right|$$
$$\times \frac{1}{nh^d}\sum_{i=1}^n \frac{W_i|\Psi(Y_i,Z_i,X_i)|K\left(\frac{X_i-x}{h}\right)}{S_C(Y_i+Z_i)}\hat{\omega}_i(\nu_n).$$

First notice that $\sup_{t\leq\mathfrak{S}_{(n)}}|\hat{S}_C(t)-S_C(t)|=O_P(n^{-1/2})$ and $\sup_{t\leq\mathfrak{S}_{(n)}}S_C(t)\hat{S}_C(t)^{-1}=O_P(1)$ (see (Shorack & Wellner, 2009) when $\hat{S}_C$ is the empirical survival function (2.3), and (Gill, 1983) when $\hat{S}_C$ is the Kaplan-Meier estimator (2.4)).
Let $E_n(M)=\{\sup_{1\leq i\leq n}|\hat{A}_i-A_i|+|\hat{B}_i-B_i|\leq M\varepsilon_n\}$. On $E_n(M)$ and for n large enough, we have $\hat{\omega}_i(\nu_n)\leq w_i(\nu_n/2)$, with

$$w_i(\nu_n/2)=\mathbf{1}_{\min(A_i,B_i,1-A_i,1-B_i)\geq\nu_n/2}. \tag{6.27}$$

Hence, since

$$\frac{1}{h^d}E\left[\frac{W_i^{p'}\Psi(Y_i,Z_i,X_i)^{p'}}{S_C(Y_i+Z_i)^{p'}}\omega_i(\nu_n/2)\right]=\frac{1}{h^d}E\left[\frac{\Psi(T_i,U_i,X_i)^{p'}}{S_C(T_i+U_i)^{2p'-1}}\omega_i^*(\nu_n/2)\right]$$

(noting $w_i^*(\nu_n/2)=\mathbf{1}_{\min(F_T(T_i|X_i),F_U(U_i|X_i),1-F_T(T_i|X_i),1-F_U(U_i|X_i))\geq\nu_n/2}$), we get from Lemma 6.5*(i)* that on $E_n(M)$,

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{nh^d}\sum_{i=1}^n\frac{W_i\Psi(Y_i,Z_i,X_i)K\left(\frac{X_i-x}{h}\right)}{S_C(Y_i+Z_i)}\hat{\omega}_i(\nu_n)\right|=O_P(n^{1/2}\tilde{\tilde{\nu}}_n).$$

This concludes the proof given that $\overline{\lim}_{M\to\infty}\lim_{n\to\infty}\mathbb{P}(E_n(M))=1$ from Assumption 8.
■

6.4 Trimming function

To control the potential erratic behavior of $\hat{A}_i$ and $\hat{B}_i$ close to the border of the unit square, we introduced some trimming $\hat{\omega}_i(\nu_n)$. Lemma 6.3 then shows the consistency of this trimming approach.

Lemma 6.3 *Let $\mathcal{F}$ denote a class of functions ψ such that $\forall\psi\in\mathcal{F}$, $|\psi(y,z,x)|\leq\Psi(y,z,x)$. Assume equation (3.15) holds, and that, for some $p>2$,*

$$\sup_{x\in\mathcal{X}}E\left[\frac{\Psi(T,U,X)^p}{S_C(T+U)^{p-1}}(1-w^*(2\nu_n))\right]=O(\tilde{\nu}_n)), \tag{6.28}$$

with $(n^{1/2}h^{d/2}\log(n)^{-1/2}\tilde{\tilde{\nu}}_n)^{p-1}h^d\underset{n\to\infty}{\to}+\infty$.
Then,

$$\sup_{x\in\mathcal{X},\psi\in\mathcal{F}}\left|\frac{1}{nh^d}\sum_{i=1}^n W_iK\left(\frac{X_i-x}{h}\right)\psi(Y_i,Z_i,X_i)(1-\hat{w}_i(\nu_n)))\right|=O_P(\tilde{\tilde{\nu}}_n).$$

Proof. As in the proof of Lemma 6.2, we consider the event $E_n(M)$. The result is then a direct application of Lemma 6.5*(ii)*. ■

6.5 Pseudo-observations

The aim of this section is to show that the pseudo-observations $\hat{A}_i$ and $\hat{B}_i$ can be asymptotically replaced by A_i and B_i.

Lemma 6.4 *Let $M^*_{n,h}(x,\theta)$ as defined in (6.24). Under Assumptions 7, 8, and assuming (3.15) holds,*

$$\sup_{x\in\mathcal{X},\theta\in\Theta} |M_{n,h}(x,\theta) - M^*_{n,h}(x,\theta)| = o_P(1).$$

Proof. We have, from a Taylor expansion and Assumption 7,

$$\begin{aligned}
|M_{n,h}(x,\theta) - M^*_{n,h}(x,\theta)| &= \frac{1}{nh^d}\sum_{i=1}^n W_{i,n} K\left(\frac{X_i - x}{h}\right)(\log c_\theta(\hat{A}_i,\hat{B}_i) - \log c_\theta(A_i,B_i))\hat{w}_i(\nu_n) \\
&\leq \frac{1}{nh^d}\sum_{i=1}^n W_{i,n} K\left(\frac{X_i - x}{h}\right)\tilde{r}_1(\tilde{A}_i) r_2(\tilde{B}_i)|\hat{A}_i - A_i|\hat{\omega}_i(\nu_n) \\
&\quad + \frac{1}{nh^d}\sum_{i=1}^n W_{i,n} K\left(\frac{X_i - x}{h}\right) r_1(\tilde{A}_i)\tilde{r}_2(\tilde{B}_i)|\hat{B}_i - B_i|\hat{\omega}_i(\nu_n),
\end{aligned}$$

where for all $i = 1,\ldots,n$, $\tilde{A}_i \in [A_i,\hat{A}_i]$ (resp. $\tilde{B}_i \in [B_i,\hat{B}_i]$).

To control the two terms in this last expression, the problems arise when $\tilde{A}_i$ and/or $\tilde{B}_i$ are close to 1 or 0. We explain how to study the case $\tilde{A}_i$ close to 0 since the other ones are similar. Therefore, we consider the case where both $\hat{A}_i$ and A_i are less than 1/2.

If $\hat{A}_i \geq A_i$, then $\tilde{r}_1(\tilde{A}_i) \leq \tilde{r}_1(A_i)$ and $r_1(\tilde{A}_i) \leq r_1(A_i)$ by Assumption 7. To treat the case $\hat{A}_i \leq A_i$, consider that we are on the event $E_n(M) = \{\sup_{1\leq i\leq n} |\hat{A}_i - A_i| + |\hat{B}_i - B_i| \leq M\varepsilon_n\}$. Then, when $\hat{\omega}_i(\nu_n) = 1$ and for n large enough, $\hat{A}_i \geq A_i/2$ (indeed, note that $A_i \leq \nu_n + M\varepsilon_n$ when $\hat{\omega}_i(\nu_n) = 1$). Hence, from $\hat{A}_i \geq A_i/2$, $\tilde{r}_1(\tilde{A}_i) \leq C\tilde{r}_1(A_i)$ and $r_1(\tilde{A}_i) \leq Cr_1(A_i)$ for some constant C using the reproducibility property of the u-shaped functions in Assumption 7.

Noting also that $\hat{\omega}_i(\nu_n) \leq \omega_i(\nu_n - M\epsilon_n)$ on $E_n(M)$ (with w_i as defined in 6.27), we have on $E_n(M)$ and for n large enought that $|M_{n,h}(x,\theta) - M^*_{n,h}(x,\theta)|$ is bounded by

$$\mathcal{T}_1 := \frac{C}{nh^d}\sum_{i=1}^n W_{i,n} K\left(\frac{X_i - x}{h}\right)\left(\tilde{r}_1(A_i) r_2(B_i)|\hat{A}_i - A_i| + r_1(A_i)\tilde{r}_2(B_i)|\hat{B}_i - B_i|\right)\omega_i(\nu_n - M\epsilon_n).$$

Moreover we have on $E_n(M)$,

$$\mathcal{T}_1 \leq \frac{CM\epsilon_n}{nh^d}\cdot \sup_{t\leq\mathfrak{S}_{(n)}}\left|\frac{S_C(t)}{\hat{S}_C(t)}\right|\cdot \sum_{i=1}^n W_i K\left(\frac{X_i - x}{h}\right)(\tilde{r}_1(A_i) r_2(B_i) + r_1(A_i)\tilde{r}_2(B_i))\,\omega_i(\nu_n - M\epsilon_n),$$

with $\sup_{t\leq \mathfrak{S}_{(n)}} |S_C(t)/\hat{S}_C(t)| = O_P(1)$.

By Lemma 6.5*(i)*, and given the Assumption 7 and the equation (3.15), we have

$$\frac{1}{nh^d}\sup_{x\in\mathcal{X}}\left|\sum_{i=1}^n W_i K\left(\frac{X_i-x}{h}\right)(\tilde{r}_1(A_i)r_2(B_i)+r_1(A_i)\tilde{r}_2(B_i))\,\omega_i(\nu_n-M\epsilon_n)\right| = O_P(\epsilon_n^{-1}\tilde{\tilde{\nu}}_n).$$

Note that to obtain this result, we replaced $w^*(\nu_n)$ by $w^*(\nu_n - M\epsilon_n)$ in Assumption 7 thanks to the reproducibility of the functions $\tilde{r}_1, r_1, \tilde{r}_2, r_2$.

This concludes the proof since we have $\overline{\lim}_{M\to\infty}\lim_{n\to\infty}\mathbb{P}(E_n(M)^c) = 0$ from Assumption 8 (where $E_n(M)^c$ denotes the complementary set of $E_n(M)$).

■

6.6 Uniform rate of convergence of the Stochastic term

In this section, we use a result from (Einmahl & Mason, 2005) to obtain uniform rates of convergence for our estimator. Lemma 6.6 below is a direct consequence of Inequality p. 1390 and Proposition 1 in (Einmahl & Mason, 2005). These results are the key arguments in the proof of Lemma 6.5 which provides the conditions for the uniform convergence. In particular, a moment condition of order $p > 1$ is required. Case *(i)* gives the result under a condition of moment of order p and a condition of speed of convergence, whereas Case *(ii)* provides the same result as Case *(i)* under a weaker assumption of speed of convergence and a stronger moment condition.

Lemma 6.5 *Let*

$$\mathfrak{K}_n(x) = \sum_{i=1}^n V_{i,n} K\left(\frac{X_i - x}{h}\right),$$

for $(V_{i,n})_{i=1,\dots,n}$ a sequence of i.i.d. random variables having the same distribution as a variable V_n.

We define

$$\|K_n\|_{\mathcal{X}} = \sup_{x\in\mathcal{X}} |\mathfrak{K}_n(x)|,$$

Assume that equation (3.15) holds, and that $E[|V_n|^p] = O(\eta_n)$ for some $p > 1$. If one of the following conditions is satisfied :

(i) $\eta_n^{p-1}\ h^d \underset{n\to\infty}{\to} +\infty$,

(ii) $\sup_{x\in\mathcal{X}} E[V_n|X=x] = O(\eta_n)$ *and* $(n^{1/2}h^{d/2}log(n)^{-1/2}\eta_n)^{p-1}\ h^d \underset{n\to\infty}{\to} +\infty$.

Then, $\|K_n\|_{\mathcal{X}} = O_P(nh^d\eta_n)$.

Proof. Let $M \geq 0$, and define

$$\begin{aligned}
\mathfrak{K}_n^M(x) &= \sum_{i=1}^n V_{i,n}\mathbf{1}_{|V_{i,n}|\leq M} K\left(\frac{X_i - x}{h}\right), \\
\mathfrak{K}_n^{\bar{M}}(x) &= \sum_{i=1}^n V_{i,n}\mathbf{1}_{|V_{i,n}|> M} K\left(\frac{X_i - x}{h}\right),
\end{aligned}$$

$\|K_n^M\|_{\mathcal{X}} = \sup_{x\in\mathcal{X}} |\mathfrak{K}_n^M(x)|$, and $\|K_n^{\bar{M}}\|_{\mathcal{X}} = \sup_{x\in\mathcal{X}} |\mathfrak{K}_n^{\bar{M}}(x)|$.

For Case *(i)*, take $t = nh^d\eta_n$ and $M = \eta_n$. By Lemma 6.6, let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be two constants such that

$$P(\|K_n^M\|_{\mathcal{X}} \geq 2nh^d\eta_n + \mathfrak{A}_1(n^{1/2}h^{d/2}\eta_n log(h^{-1})^{1/2}) \leq 4\exp\left(-\mathfrak{A}_2 nh^d\right).$$

Since $n^{1/2}h^{d/2}\eta_n log(h^{-1})^{1/2} = o(nh^d\eta_n)$ and $nh^d log(n)^{-1} \underset{n\to\infty}{\to} +\infty$, we get $\|K_n^M\|_{\mathcal{X}} = O_P(nh^d\eta_n)$.

On the other hand, we have

$$\begin{aligned}
P(\|K_n^{\bar{M}}\|_{\mathcal{X}} \geq t) &\leq E\left[\sup_{x\in\mathcal{X}}\left|\sum\nolimits_{i=1}^n V_{i,n}\mathbf{1}_{|V_{i,n}|>M}K((X_i - x)/h)\right|\right] \Big/ t \\
&\leq n\|K\|_\infty E\Big[|V_{i,n}|\mathbf{1}_{|V_{i,n}|>M}\Big]/t
\end{aligned}$$

by Markov's inequality. Using Hölder's inequality,

$$E[|V_{i,n}|\mathbf{1}_{|V_{i,n}|\geq M}] \leq E[|V_{i,n}|^p]^{1/p} P(|V_{i,n}| \geq M)^{1-1/p}$$

with $P(|V_{i,n}| \geq M) \leq E[|V_{i,n}|^p]/M^p$.

Finally, using the fact that $E[|V_n|^p] = O(\eta_n)$, we get

$$P(\|K_n^{\bar{M}}\|_{\mathcal{X}} \geq t) = O(1/(\eta_n^{p-1}h^d)),$$

and $\|K_n^{\bar{M}}\|_{\mathcal{X}} = O_P(nh^d\eta_n)$, which concludes the proof of Case *(i)*.

The proof of the Case *(ii)* is similar to the Case *(i)*, taking $t = nh^d\eta_n$ and $M = n^{1/2}h^{d/2}log(h^{-1})^{-1/2}\eta_n$. ∎

Lemma 6.6 *Let $M > 0$ and consider $\|K_n\|_{\mathcal{X}}$ and Cases (i) and (ii) as defined in Proposition 6.5. Suppose that $|V_{i,n}| \leq M$. Then, there exists constants $\mathfrak{A}_1$ and $\mathfrak{A}_2$ such that*

$$P\left(\|K_n\|_{\mathcal{X}} \geq nh^d C + \mathfrak{A}_1(n^{1/2}h^{d/2}Mlog(h^{-1})^{1/2} + t\right) \leq 2\left(\exp\left(-\frac{\mathfrak{A}_2 t^2}{nM^2h^d}\right) + \exp\left(-\frac{\mathfrak{A}_2 t}{M}\right)\right), \tag{6.29}$$

with $C = M$ for Case (i) and $C = \sup_{x\in\mathcal{X}} E[V_n|X = x]$ for Case (ii).

Proof. This result is the consequence of the application of Inequality p. 1390 and Proposition 1 in (Einmahl & Mason, 2005).

We apply Proposition 1 to the random vector (V_n, X) and to the class of functions $\mathcal{G} = \{(v, u) \to v\mathbf{1}_{|v|\leq M}K((u - x)/h), x \in \mathcal{X}\}$. In particular, we use the Lemma 22 in the article of (Nolan & Pollard, 1987) to verify the assumption on the covering number of the class $\mathcal{G}$. ■

6.7 Proof of Theorem 3.2 (Stochastic term)

For the sake of simplicity, we assume in this section that $\theta \in \mathbb{R}$. The multidimensional case can be studied similarly, component by component, as we did in Section 6.1.

By Proposition 6.1, we already have that $\hat{\theta}_h(x) - \theta_h^*(x)$ tends uniformly to zero. To obtain the convergence rate, the key result consists in controlling the deviations of the process

$$\mathcal{Z}_h(x, \theta) = \frac{M_{n,h}(x, \theta) - M_{n,h}(x, \theta_h^*(x)) - M_f^{(c)}(x, \theta) + M_f^{(c)}(x, \theta_h^*(x))}{|\theta - \theta_h^*(x)|}.$$

Indeed, by definition of $\hat{\theta}_h(x)$ and $\theta_h^*(x)$,

$$M_{n,h}(x, \hat{\theta}_h(x)) - M_{n,h}(x, \theta_h^*(x)) \geq 0,$$

and

$$M_f^{(c)}(x, \theta_h^*(x)) - M_f^{(c)}(x, \hat{\theta}_h(x)) \geq 0.$$

Therefore,

$$0 \leq \frac{M_f^{(c)}(x, \theta_h^*(x)) - M_f^{(c)}(x, \hat{\theta}_h(x))}{|\hat{\theta}_h(x) - \theta_h^*(x)|} \leq \mathcal{Z}_h(x, \hat{\theta}_h(x)).$$

Moreover, by a second order Taylor expansion,

$$M_f^{(c)}(x, \hat{\theta}_h(x)) - M_f^{(c)}(x, \theta_h^*(x)) = \frac{(\hat{\theta}_h(x) - \theta_h^*(x))^2}{2}\nabla_\theta^2 M_f^{(c)}(x, \tilde{\theta}_h(x)),$$

where $\tilde{\theta}_h(x)$ belongs to the interval $[\hat{\theta}_h(x), \theta^*_h(x)]$. Due to Assumption 5 and to the consistency of $\hat{\theta}_h(x)$ shown in Proposition 6.1, $|\nabla^2_\theta M_f^{(c)}(x, \tilde{\theta}_h(x))| \geq c'_0 > 0$ for h small enough and n sufficiently large. The result of Theorem 3.2 then follows from Proposition 6.7 below.

Proposition 6.7 *Under Assumptions 6 to 8 and equation (3.15),*

$$\sup_{x,\theta} |\mathcal{Z}_h(x,\theta)| = O_P(\tilde{\tilde{\nu}}_n + [\log n]^{1/2} n^{-1/2} h^{-d/2}).$$

Proof of Proposition 6.7. With $M^*_{n,h}$ defined in (6.24), decompose

$$\mathcal{Z}_h(x,\theta) = \mathcal{Z}^*_h(x,\theta) + \mathcal{Z}^{(c)}_h(x,\theta),$$

where

$$\begin{aligned}
\mathcal{Z}^*_h(x,\theta) &= \frac{M_{n,h}(x,\theta) - M_{n,h}(x,\theta^*_h(x)) - M^*_{n,h}(x,\theta) + M^*_{n,h}(x,\theta^*_h(x))}{|\theta - \theta^*_h(x)|}, \\
\mathcal{Z}^{(c)}_h(x,\theta) &= \frac{M^*_{n,h}(x,\theta) - M^*_{n,h}(x,\theta^*_h(x)) - M_f^{(c)}(x,\theta) + M_f^{(c)}(x,\theta^*_h(x))}{|\theta - \theta^*_h(x)|}.
\end{aligned}$$

$\mathcal{Z}^*_h$ corresponds to the replacement of (A_i, B_i) by pseudo-observations $(\hat{A}_i, \hat{B}_i)$, while $\mathcal{Z}^{(c)}_h$ comes from the difference between the criterion when the margins are known and its expectation. These two terms are studied separately in Lemma 6.8 and Lemma 6.9. ■

6.7.1 Auxiliary Lemmas

Lemma 6.8 *Under Assumptions 7 and 8, and equation (3.15),*

$$\sup_{x,\theta} |\mathcal{Z}^*_h(x,\theta)| = O_P(\tilde{\tilde{\nu}}_n).$$

Proof. From a first order Taylor expansion, we get

$$|\phi(a,b,\theta) - \phi(a,b,\theta^*_h(x)) - \phi(\hat{a},\hat{b},\theta) + \phi(\hat{a},\hat{b},\theta^*_h(x))| \leq |\dot{\phi}(a,b,\tilde{\theta}_h(x)) - \dot{\phi}(\hat{a},\hat{b},\tilde{\theta}_h(x))| \cdot |\theta - \theta^*_h(x)|,$$

for some $\tilde{\theta}_h(x)$ between θ and $\theta^*_h(x)$.

Next, from another Taylor expansion we have :

$$|\dot{\phi}(a,b,\tilde{\theta}_h(x)) - \dot{\phi}(\hat{a},\hat{b},\tilde{\theta}_h(x))| \leq |\partial_a \dot{\phi}(\tilde{a},\tilde{b},\tilde{\theta}_h(x))| \cdot |\hat{a} - a| + |\partial_b \dot{\phi}(\tilde{a},\tilde{b},\tilde{\theta}_h(x))| \cdot |\hat{b} - b|,$$

with, $\tilde{a} \in [a, \hat{a}]$ and $\tilde{b} \in [b, \hat{b}]$. Hence, we use Assumption 7 to shaw that

$$|\mathcal{Z}^*_h(x,\theta)| \leq \frac{1}{nh^d} \sum_{i=1}^n W_{i,n} K\left(\frac{X_i - x}{h}\right) \left\{ \tilde{r}_1(\tilde{A}_i) r_2(\tilde{B}_i) |\hat{A}_i - A_i| + r_1(\tilde{A}_i) \tilde{r}_2(\tilde{B}_i) |\hat{B}_i - B_i| \right\} \hat{\omega}_i(\nu_n),$$

with $\tilde{A}_i \in [A_i, \hat{A}_i]$ (resp. $\tilde{B}_i \in [B_i, \hat{B}_i]$).

In order to obtain the desired result, we need to control terms of the same form as in the proof of Lemma 6.4. Using the same arguments, we can show that on the set $E_n(M) = \{\sup_{1\leq i\leq n} |\hat{A}_i - A_i| + |\hat{B}_i - B_i| \leq M\varepsilon_n\}$, we have for n large enough,

$$|\mathcal{Z}_h^*(x,\theta)| \leq CM\varepsilon_n \sup_{t\leq\mathfrak{S}_{(n)}} \left|\frac{S_C(t)}{\hat{S}_C(t)}\right| \cdot \frac{1}{nh^d}\sum_{i=1}^n W_i K\left(\frac{X_i - x}{h}\right)\{\tilde{r}_1(A_i)r_2(B_i) + r_1(A_i)\tilde{r}_2(B_i)\}\,\omega_i(\nu_n - M\epsilon_n),$$

for some constants C.

This allows us to conclude the proof noting that $\sup_{t\leq\mathfrak{S}_{(n)}} |S_C(t)/\hat{S}_C(t)| = O_P(1)$ and

$$\frac{1}{nh^d}\sum_{i=1}^n W_i K\left(\frac{X_i - x}{h}\right)\{\tilde{r}_1(A_i)r_2(B_i) + r_1(A_i)\tilde{r}_2(B_i)\}\,\omega_i(\nu_n - M\epsilon) = O_P(\tilde{\tilde{\nu}}_n/\epsilon_n)$$

from Assumption 7, equation (3.15), and Lemma 6.5. ■

Lemma 6.9 *Under Assumptions 6, 8, and equation (3.15),*

$$\sup_{x,\theta} |\mathcal{Z}_h^{(c)}(x,\theta)| = O_P(\tilde{\tilde{\nu}}_n + [\log n]^{1/2} n^{-1/2} h^{-d/2}).$$

Proof. Let

$$\phi_{\theta,\theta'}(y,z,x) = \frac{\log c_\theta(F_T(y|x), F_U(z|x)) - \log c_{\theta'}(F_T(y|x), F_U(z|x))}{(\theta - \theta')}.$$

Let $\mathcal{A}$ denote the class which contains all the functions $\phi_{\theta,\theta'}$. We have

$$\phi_{\theta,\theta'}(y,z,x) \leq \Psi(y,z,x) := R(F_T(y|x), F_U(z|x)),$$

where we used Assumption 6.

It follows from Lemma 6.2 and Lemma 6.3 that

$$\begin{aligned}
\frac{M_{n,h}^*(x,\theta) - M_{n,h}^*(x,\theta_h^*(x))}{(\theta - \theta_h^*(x))} &= \frac{1}{nh^d}\sum_{i=1}^n W_i K\left(\frac{X_i - x}{h}\right)\phi_{\theta,\theta_h^*(x)}(Y_i, Z_i, X_i) + O_P(\tilde{\tilde{\nu}}_n), \\
&=: L_n(x,h,\theta) + O_P(\tilde{\tilde{\nu}}_n),
\end{aligned}$$

where the O_P-rate is uniform in x and θ.

On the other hand,

$$\sup_{x,\theta}\left| L_n(x,h,\theta) - \frac{M_f^{(c)}(x,\theta) - M_f^{(c)}(x,\theta_h^*(x))}{\theta - \theta_h^*(x)}\right| = O_P([\log n]^{1/2} n^{-1/2} h^{-d/2}).$$

This result is obtained using Theorem 4 in (Einmahl & Mason, 2005). Let $\mathcal{A}^\delta = \{(y,z,c,x) \to \mathbf{1}_{y+z\leq c}\ a(y,z,x)S_C(y+z)^{-1} : a \in \mathcal{A}\}$, and $\Psi^\delta(T,U,C,X) = \mathbf{1}_{T+U\leq C}\Psi(T,U,X)S_C(T+U)^{-1}$. The conditions in Theorem 4 in (Einmahl & Mason, 2005) hold if we check that

$$N(\varepsilon, \mathcal{A}^\delta, \Psi^\delta) \leq \Delta\varepsilon^{-\alpha}, \tag{6.30}$$

for some $\alpha > 0$ and $\Delta > 0$, and if

$$E\left[\left\{\frac{\delta\Psi(Y,Z,X)}{S_C(Y+Z)}\right\}^p\right] < \infty, \tag{6.31}$$

for some $p > 2$. Condition (6.31) is easy to check, since

$$E\left[\left\{\frac{\delta\Psi(Y,Z,X)}{S_C(Y+Z)}\right\}^p\right] = E\left[\frac{\Psi(T,U,X)^p}{S_C(T+U)^{p-1}}\right]$$

which is finite from (3.10) in Assumption 6.

To check (6.30), observe that

$$\phi_{\theta_1,\theta_2}(y,z,x) - \phi_{\theta_3,\theta_4}(y,z,x) = \dot{\phi}(F_T(y|x), F_U(z|x), \tilde{\theta}) - \dot{\phi}(F_T(y|x), F_U(z|x), \bar{\theta}),$$

for $\tilde{\theta}$ between θ_1 and θ_2, and $\bar{\theta}$ between θ_3 and θ_4. Then we get, from a Taylor expansion,

$$\phi_{\theta_1,\theta_2}(y,z,x) - \phi_{\theta_3,\theta_4}(y,z,x) = \ddot{\phi}(F_T(y|x), F_U(z|x), \theta^*)(\tilde{\theta} - \bar{\theta}),$$

for θ^* between $\tilde{\theta}$ and $\bar{\theta}$. From Assumption 5, we deduce that the class $\mathcal{A}$ satisfies (6.30) thanks to Lemma 19.31 of (van der Vaart, 1998). ■

References

Derumigny, A., & Fermanian, J.-D. (2017). About tests of the simplifying assumption for conditional copulas. *Dependence modeling*(5), 154 - 197.

Einmahl, U., & Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, *33*(3), 1380–1403. Retrieved from `http://dx.doi.org/10.1214/009053605000000129` doi: 10.1214/009053605000000129

Geerdens, C., Acar, E. F., & Janssen, P. (2018). Conditional copula models for right-censored clustered event time data. *Biostatistics*, *19*(2), 247-262. Retrieved from `http://dx.doi.org/10.1093/biostatistics/kxx034` doi: 10.1093/biostatistics/kxx034

Genest, C., Ghoudi, K., & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, *82*(3), 543–552. Retrieved from `http://www.jstor.org/stable/2337532`

Genest, C., & MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, *40*(4), 280–283.

Gijbels, I., Veraverbeke, N., & Omelka, M. (2011). Conditional copulas, association measures and their applications. *Computational Statistics & Data Analysis*, *55*(5), 1919 - 1932. Retrieved from `http://www.sciencedirect.com/science/article/pii/S016794731000438X` doi: https://doi.org/10.1016/j.csda.2010.11.010

Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.*, *11*(1), 49–58. Retrieved from `http://dx.doi.org/10.1214/aos/1176346055` doi: 10.1214/aos/1176346055

Gribkova, S., & Lopez, O. (2015). Non-parametric copula estimation under bivariate censoring. *Scandinavian Journal of Statistics*, *42*(4), 925–946. Retrieved from `http://dx.doi.org/10.1111/sjos.12144` (10.1111/sjos.12144) doi: 10.1111/sjos.12144

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

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81–93.

Lakhal-Chaieb, M. (2010). Copula inference under censoring. *Biometrika*, *97*(2), 505–512. Retrieved from `http://www.jstor.org/stable/25734101`

Meira-Machado, L., Sestelo, M., & Gonçalves, A. (2016). Nonparametric estimation of the survival function for ordered multivariate failure time data: A comparative study. *Biometrical Journal*, *58*(3), 623–634.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, *9*(1), 141–142.

Nolan, D., & Pollard, D. (1987). U-processes: rates of convergence. *The Annals of Statistics*, 780–799.

Shorack, G. R., & Wellner, J. A. (2009). *Empirical processes with applications to statistics* (Vol. 59). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Retrieved from `https://doi.org/10.1137/1.9780898719017.ch1` (Reprint of the 1986 original [MR0838963]) doi: 10.1137/1.9780898719017.ch1

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst.*

Statist. Univ. Paris, *8*, 229–231.

Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.*, *23*(2), 422–439. Retrieved from http://dx.doi.org/10.1214/aos/1176324528 doi: 10.1214/aos/1176324528

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics*, *33*(3), 357–375. Retrieved from http://dx.doi.org/10.1002/cjs.5540330304 doi: 10.1002/cjs.5540330304

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press.

Veraverbeke, N., Gijbels, I., & Omelka, M. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, *38*(4), 766-780. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2011.00744.x doi: 10.1111/j.1467-9469.2011.00744.x

Wang, W., & Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika*, *85*(3), 561–572. Retrieved from http://www.jstor.org/stable/2337386

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359–372.