Working Paper in Economics

# 202504

# Modelling the dependence structure for censored Time-to-Events: Theory and Applications


**Yassir Rabhi[1]**
**Taoufik Bouezmarini[2]**
**Abderrahim Taamouti[3]**

[1] University of Essex, yassir.rabhi@essex.ac.uk
[2] University of Sherbrooke, taoufik.bouezmarni@usherbrooke.ca
[3] University of Liverpool Management School, abderrahim.taamouti@liverpool.ac.uk



https://www.liverpool.ac.uk/economics/

# Modelling the dependence structure for censored Time-to-Events: Theory and Applications

Yassir Rabhi [1], Taoufik Bouezmarni [2], and Abderrahim Taamouti [3]

University of Essex[1], University of Sherbrooke[2] and University of Liverpool[3]

**Abstract**: Kernel estimation of copulas often requires adjustments near the borders of $[0,1]^2$ due to both practical and asymptotic considerations. In this paper, we introduce a local-polynomial estimator for copula density, designed to overcome boundary bias in the context of censored Time-to-Events data, alongside estimators for Kendall's tau and Spearman's rho. To explore the asymptotic properties of our estimators, we analyze the oscillatory behavior of a bivariate cumulative distribution function (cdf) estimator under right-censored data conditions, deriving its i.i.d. representation with an improved remainder term rate. Subsequently, we derive an i.i.d. representation for a copula cdf estimator and establish a functional Central Limit Theorem (CLT) for the copula density estimator. Additionally, we prove the weak convergence of the Kendall's tau and Spearman's rho estimators. Our results show that the local-polynomial estimator is stable and more efficient near the boundaries of $[0,1]^2$. We validate the finite sample performance of the local-polynomial estimator through an extensive Monte Carlo simulation exercise. Finally, we apply our estimators to two datasets. The first one involves insurance company indemnity claims, where we examine the dependence between indemnity payments (loss) and allocated loss adjustment expenses such as lawyers' fees and claims investigation costs. The second one analyzes the effect of the disregard rate on the duration of unemployment.

*Keywords:* Copula density, Right-censored data, Local-polynomial estimation, Measures of dependence, i.i.d. representation of copula estimator, Functional C.L.T.

[1] e-mail: `yassir.rabhi@essex.ac.uk`, School of Mathematics, Statistics and Actuarial sciences, Colchester, United Kingdom.

[2] e-mail: `taoufik.bouezmarni@usherbrooke.ca`, Department of Mathematics, Sherbrooke, Canada.

[3] e-mail: `abderrahim.taamouti@liverpool.ac.uk`, University of Liverpool Management School, Liverpool, UK.

# 1 Introduction

Copulas have become a cornerstone in econometric and statistical research for analyzing the dependence structure among variables. Unlike classical measures that capture only certain aspects of dependence, copulas offer a comprehensive view of the relationships between variables. They can identify a wide range of dependencies, including local, global, and tail dependencies. There has been a growing interest recently in modeling dependence using right-censored time-to-event data. For instance, in insurance, (7) used parametric copulas to examine the dependence between right-censored losses and the allocated loss adjustment expense (ALAE) associated with a single claim [see also (6), (11)]. In survival analysis, (21) developed semiparametric copulas for censoring settings to study the dependence between various disease events in patients with AIDS. Overall, copulas have been extensively studied using both parametric and semiparametric approaches. However, selecting the appropriate copula model is inherently complex, and these approaches are often susceptible to misspecification issues, which can lead to misleading empirical results. To address these challenges, this paper aims to develop a non-parametric approach for estimating copulas and other dependence measures. This approach is designed to be robust against boundary bias when dealing with censored Time-to-Event data.

Nonparametric estimation approaches are favored for their flexibility and reliance on minimal assumptions. A range of nonparametric methods for copula estimation have been investigated in the literature. For instance, (2) proposed an approach employing Bernstein polynomials, while (8) introduced a wavelet-based estimation technique. Additionally, (10) explored copulas based on B-splines, and both (4) and (19) presented robust kernel methods. Particularly noteworthy are the latter two works, which specifically address the issue of boundary bias in kernel estimation of the copula cumulative distribution function. (27) recently proposed an improved transformation-kernel estimator that employs a smooth tapering method to mitigate boundary biases arising from unbounded copula densities. However, the performance of their approach may depend significantly on the choice of transformation and remain vulnerable to boundary issues, particularly if the transformation does not adequately account for the boundary behavior of the data. Furthermore, as shown in this paper, transformation-kernel-based approaches result in an estimator's variance that diverges to infinity at the boundaries of $[0, 1]^2$. Moreover, none of the aforementioned approaches are applicable in the presence of censored data.

Recently, several nonparametric methods have been proposed for estimating copulas under right censoring. For example, (20) developed a kernel smoothing estimator for the conditional copula that works with right-censored data. Additionally, (9), using an estimator of the joint cumulative distribution function and a transformation of the initial variables, derived copula estimators that are valid under various bivariate censoring frameworks. Unlike the local-polynomial estimator proposed in this paper, which maintains finite variance when estimating bounded densities, the variance of the transformation-based estimator diverges to infinity at the boundaries of $[0, 1]^2$, as indicated in the previous paragraph. For unbounded densities, the local-polynomial estimator is relatively more efficient, as demonstrated in Remark 1 of Section 2.2. In many cases, the dependence structure between variables is more pronounced in the tails of their joint distribution. With its stable variance, the local-polynomial method offers a reliable and efficient approach for measuring dependence.

In this paper, we explore multivariate dependence with censored Time-to-Events data using copulas, Kendall's tau, and Spearman's rho. Specifically, we introduce a local-polynomial estimator for copulas to alleviate boundary bias within $[0, 1]^2$. Additionally, we develop nonparametric estimators for Kendall's tau and Spearman's rho. To analyze the asymptotic properties of our estimators, we initially establish an i.i.d. representation for the estimator of the bivariate cumulative distribution function $\widehat{F}$ as outlined in (23), achieving a remainder term of order $\mathcal{O}_{a.s.}\left(n^{\frac{-3}{4}}(\log n)^{\alpha_1}\right)$, an improvement over the rate reported in (23). This enhanced rate is crucial for examining the properties of our local-polynomial estimator. Subsequently, we establish an i.i.d. representation for an estimator of the copula cdf. As a by-product, we derive results on the oscillation behavior of the bivariate process $\widehat{F}$. We then investigate the weak convergence of the copula density estimator by deriving its triangular representation. Moreover, we determine the asymptotic distributions of the Kendall and Spearman estimators. Finally, we derive both local and global optimal bandwidth parameters necessary for the nonparametric estimation of copulas.

Moreover, to assess the finite sample performance of the local-polynomial estimator, we conduct a comprehensive Monte Carlo simulation study. This extensive exercise allows us to evaluate the performance of our estimator, specifically using the integrated squared error, under various scenarios. Following this validation, we demonstrate the practical utility of our estimators by applying them to two real-world datasets.

The first dataset focuses on indemnity claims from an insurance company, where we investigate the relationship between indemnity payments (losses) and the corresponding loss adjustment expenses. These expenses include legal fees, claims investigation costs, and other associated expenditures. Understanding this relationship is crucial for insurance companies as it can provide insights into cost management and risk assessment strategies.

The second dataset explores the impact of the disregard rate on the duration of unemployment. Here, we analyze how variations in the disregard rate—a measure of how certain benefits are treated in the calculation of unemployment assistance—influence the length of time individuals remain unemployed. This analysis sheds light on the effectiveness of unemployment policies and their implications for labor market dynamics.

The rest of the manuscript is structured as follows. In §2.1, we introduce a new local-polynomial estimator for the copula density of a random vector, where one of the variables is right-censored. §2.2 explores the asymptotic theory related to the estimators of the bivariate cdf, the copula function, and its density. In §3, we develop estimators for Kendall's tau and Spearman's rho and analyze their weak convergence properties. §4.1 evaluates the finite sample performance of our estimators through simulations, demonstrating that the proposed estimators generally outperform the transformation estimator. Subsequently, in §4.2 and §4.3 we apply our estimators to two empirical datasets: the first involves insurance company indemnity claims, examining the dependence between indemnity payments and allocated loss adjustment expenses; the second relates to unemployment, analyzing the impact of the disregard rate on unemployment duration. §5 concludes and the appendix contains the proofs of the theoretical results.

## 2 Copula estimation under censored Time-to-Events

In this section, we will introduce the local-polynomial estimator for copulas under censored Time-to-Events data. This estimator aims to alleviate boundary bias within the unit square $[0, 1]^2$. Following the introduction of the estimator, we will thoroughly examine its asymptotic properties, providing a detailed analysis of its behavior and performance in large samples. This includes establishing an i.i.d. representation for the estimator of the bivariate cumulative distribution function and assessing the convergence rates and asymptotic distributions of the local-polynomial estimator of the copulas.

## 2.1 A local-polynomial estimator

Consider the Time-to-Event variable $Y > 0$ and the covariate $X$ with the joint distribution $F = F_{X,Y}$, where the marginal cdfs are $F_1 = F_X$ and $F_2 = F_Y$. Suppose $Y$ is subject to right-censoring by the random variable $C$, which has distribution $G$, and that we observe $Z = \min(Y, C)$ and $\delta = \mathrm{I}(Y \leq C)$, the failure indicator. We assume that $Y$ and $C$ are independent, that $P[\delta = 1|X, Y] = P[\delta = 1|Y]$, and that $X$ is an uncensored variable. The observed i.i.d. data is of the form $(Z_i, \delta_i, X_i), , i = 1, \ldots, n$.

Under the above setting, an estimator for the copula function $\mathbb{C}(u, v) = F\left(F_1^{-1}(u), F_2^{-1}(v)\right)$, for $u, v \in [0, 1]$, is given by

$$\widehat{\mathbb{C}}(u, v) = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{1 - \widehat{G}(Z_i^-)} \mathrm{I}\left(\widehat{F}_1(X_i) \leq u,\, \widehat{F}_2(Z_i) \leq v\right), \tag{1}$$

where $\widehat{F}_1$ is the empirical counterpart of $F_1$, and $\widehat{F}_2$ and $\widehat{G}$ are the respective Kaplan-Meier estimators of $F_2$ and $G$. For instance,

$$\widehat{F}_2(x) = 1 - \prod_{\substack{i=1 \\ z_{(i)} \leq x}}^{d} \left(1 - \frac{\sum_{j=1}^{n} \mathrm{I}(z_j = z_{(i)})}{\sum_{j=1}^{n} \mathrm{I}(z_j \geq z_{(i)})}\right),$$

where $z_{(1)}, \ldots, z_{(d)}$ are the distinct ordered Time-to-Events from the uncensored $z$'s. Notice that $\widehat{\mathbb{C}}(u, v) = \widehat{F}\left(\widehat{F}_1^{-1}(u^+)^-, \widehat{F}_2^{-1}(v^+)^-\right)$, where $\widehat{F}$ is the bivariate estimator

$$\widehat{F}(x, y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 - \widehat{G}(Z_i^-)} \mathrm{I}\big(X_i \leq x, Z_i \leq y, \delta_i = 1\big), \tag{2}$$

with $\widehat{F}(x^-, y^-) = \lim_{\substack{(u,v)\to(x,y) \\ u<x, v<y}} \widehat{F}(u, v)$ and $\widehat{F}_k^{-1}(u^+) = \lim_{\substack{t\to u \\ t>u}} \widehat{F}_k^{-1}(t)$ $(k = 1, 2)$; see (23). The estimator $\widehat{F}$ can be derived from the expression

$$F_{X,Y}(x, y) = \int_{t\leq x, s\leq y} \frac{1}{1 - G(s^-)}\, \mathrm{d}F_{X,Z,\delta}(t, s, 1),$$

by replacing $G$ by $\widehat{G}$ and $F_{X,Z,\delta}(t, s, 1) = P(X \leq t, Z \leq s, \delta = 1)$ by its empirical counterpart.

We next consider a local-polynomial approach to estimate the copula density $\mathfrak{C}$. This approach overcomes boundary bias and has the feature of having a stable variance near $0$ and $1$. Let $K$ denote a symmetric density function with support on the interval $(-1, 1)$, and let $h = h_n$ be a sequence of bandwidths that tends to zero as $n$ approaches infinity. We define three intervals as follows: $\mathcal{A}_1 = [0, h]$, $\mathcal{A}_2 = [h, 1-h]$, and $\mathcal{A}_3 = [1-h, 1]$ and

$$K_{u,h}(t) = K(t)\,\frac{a_2(u,h) - a_1(u,h)t}{a_0(u,h)a_2(u,h) - a_1^2(u,h)}\,\mathrm{I}(u \in \mathcal{A}_i),\ \ (i = 1, 2, 3), \tag{3}$$

where

$$a_\ell(u,h) = \int_{(u-1)/h}^{u/h} t^\ell K(t)\,\mathrm{d}t,\ \ (\ell = 0, 1, 2).$$

Notice that $K_{u,h} = K$ when $u \in [h, 1-h]$, $\int_{-1}^{1} K_{u,h}(t), \mathrm{d}t = 1$, and $\int_{-1}^{1} t, K_{u,h}(t), \mathrm{d}t = 0$. The function $K_{u,h}$ represents a locally linear version of $K$, originally introduced by (12) in the context of density estimation. Other variations of boundary polynomial kernels are discussed in (18). In this paper, we use $K_{u,h}$ for estimating the copula density $\mathfrak{C}$ within its compact support $[0, 1]^2$. The local polynomial estimator of $\mathfrak{C}$ is

$$\widehat{\mathfrak{C}}(u,v) = \frac{1}{nh^2}\sum_{i=1}^{n} \frac{\delta_i}{1 - \widehat{G}(Z_i^-)}\,K_{u,h}\left(\frac{u - \widehat{F}_1(X_i)}{h}\right) K_{v,h}\left(\frac{v - \widehat{F}_2(Z_i)}{h}\right), \tag{4}$$

and in the interior set $[h, 1-h]^2$ is simplified to

$$\widehat{\mathfrak{C}}_I(u,v) = \frac{1}{nh^2}\sum_{i=1}^{n} \frac{\delta_i}{1 - \widehat{G}(Z_i)}\,K\left(\frac{u - \widehat{F}_1(X_i)}{h}\right) K\left(\frac{v - \widehat{F}_2(Z_i)}{h}\right).$$

The estimator $\widehat{\mathfrak{C}}$ can be viewed as the convolution of $K_{u,h} \times K_{v,h}$ with $\widehat{\mathbb{C}}$. The rationale for using $K_{u,h}$ instead of $K$ is to mitigate boundary bias in the regions $[0, h]$ and $[1-h, 1]$, thereby improving the rate of the estimator' bias to $\mathcal{O}(h^2)$ everywhere in $[0, 1]^2$. In the context of uncensored data, (4) addressed the necessity of adjusting the standard kernel estimator of the copula function to overcome boundary bias in the intervals $[0, h)$ and $(1-h, 1]$. For regression function estimation, (18) explored various types of local polynomial kernels for estimating a compactly supported function. Specifically, when $K$ is the Epanechnikov kernel, the coefficients $a_0$, $a_1$, and $a_2$ in (3) take the following forms:

$$
\begin{aligned}
a_0(u,h) &= \left\{\frac{1}{2}+\frac{3u}{4h}-\frac{u^3}{4h^3}\right\}\mathrm{I}_{[0,h]}(u)+\left\{\frac{1}{2}+\frac{3(1-u)}{4h}-\frac{(1-u)^3}{4h^3}\right\}\mathrm{I}_{[1-h,1]}(u),\\
a_1(u,h) &= \left\{-\frac{3}{16}+\frac{3u^2}{8h^2}-\frac{3u^4}{16h^4}\right\}\mathrm{I}_{[0,h]}(u)+\left\{\frac{3}{16}-\frac{3(1-u)^2}{8h^2}+\frac{3(1-u)^4}{16h^4}\right\}\mathrm{I}_{[1-h,1]}(u),\\
a_2(u,h) &= \left\{\frac{1}{10}+\frac{u^3}{4h^3}-\frac{3u^5}{20h^5}\right\}\mathrm{I}_{[0,h]}(u)+\left\{\frac{1}{10}+\frac{(1-u)^3}{4h^3}-\frac{3(1-u)^5}{20h^5}\right\}\mathrm{I}_{[1-h,1]}(u).
\end{aligned}
$$

Now, let $L(z) = P[Z \leq z]$ and $\varrho = \sup\{z : L(z) < 1\} < \infty$. To avoid identifiability problems in $[\varrho, \infty)$ due to right censoring, we define $\widehat{F}_2$ and $\widehat{G}$ on the interval $[0, \varrho)$, and $\widehat{\mathbb{C}}$ and $\widehat{\mathfrak{C}}$ on $[0,1] \times \big[0, F_2(\varrho)\big)$. Additionally, we assume that $G(\varrho) < 1$.

## 2.2 Asymptotic Theory

In this section, we investigate an i.i.d. representation for the local-polynomial estimator $\widehat{\mathfrak{C}}$ (see Theorem 3) and its weak convergence. To achieve this, we first analyze the oscillation behavior of $\widehat{F}$ and derive a representation that features a faster remainder rate compared to similar representations in the literature; see (23). Subsequently, we establish a representation for the copula estimator $\widehat{\mathbb{C}}$.

For a given distribution $Q$, let $\overline{Q} = 1 - Q$, and define $L_0(z) = P[Z \leq z, \delta = 0]$ and

$$
\chi_i^F(x,y) = \frac{\mathrm{I}(X_i \leq x, Z_i \leq y, \delta_i = 1)}{\overline{G}(Z_i)} + \iint_{\substack{u\leq x\\ v\leq y}} \left[\frac{\mathrm{I}(Z_i \leq v, \delta_i = 0)}{\overline{L}(Z_i)} - \int_0^{v\wedge Z_i} \frac{\mathrm{d}L_0(t)}{\overline{L}^2(t)} - 1\right]\mathrm{d}F(u,v).
$$

Using the above notations, the following result establishes the i.i.d. representation of the estimator of the cumulative distribution function [see the proof of Proposition 1 in the appendix].

**Proposition 1** *Suppose that $G$ is a Lipshitz function on $[0,\varrho]$, and let $\mathcal{A} = \mathbb{R} \times [0,\varrho)$. The bivariate cdf estimator $\widehat{F}(x,y)$ admits for $(x,y) \in \mathcal{A}$ the representation*

$$
\widehat{F}(x,y) - F(x,y) = \frac{1}{n}\sum_{i=1}^{n} \chi_i^F(x,y) + r_n(x,y), \tag{5}
$$

*where $\sup_{\mathcal{A}} |r_n(x,y)| = \mathcal{O}\left(n^{-3/4}(\log n)^{\alpha_1}\right)$ a.s., with $\alpha_1 \geq 1$.*

The rate of $r_n$ in (5) is crucial for studying the asymptotic distribution of the local-polynomial estimator of the copula density. For instance, in the representation provided by (23), the remainder term is of order $o_{a.s.}(n^{-1/2})$. The covariance function of the stochastic process $\chi_1^F(x,y)$ is

$$\Sigma_{\mathcal{F}}(x,y,x_0,y_0) = \int_{\substack{u\le x\wedge x_0 \\ v\le y\wedge y_0}} \frac{\mathrm{d}F(u,v)}{\overline{G}(v)} + \int_{\substack{v\le y \\ v_0\le y_0}} \sigma_G(v,v_0)\frac{\mathrm{d}F(x,v)}{\overline{G}(v)}\frac{\mathrm{d}F(x_0,v_0)}{\overline{G}(v_0)}$$
$$-2\int_{\substack{u\le x, v\le y \\ u_0\le x_0, v_0\le y_0}} \left\{\int_0^{v\wedge v_0} \frac{\mathrm{d}L_0(t)}{\overline{L}^2(t)}\right\} \mathrm{d}F(u,v)\mathrm{d}F(u_0,v_0) - F(x,y)F(x_0,y_0),$$

where $\sigma_G(t,s)$ is the covariance function of the limiting process of $\widehat{G}$. We now have the following result on the oscillation behavior of $\widehat{F}$ [see the proof of Theorem 1 in the appendix].

**Theorem 1 (Oscillation behavior of $\widehat{F}$)**
*Let $\{a_n\}$ be a sequence of positive values such that $a_n = \mathcal{O}\left(n^{-1/2}(\log n)^{\alpha_2}\right)$ ($\alpha_2 \ge 1/2$). Then with probability $1$,*

$$\sup_{(x,y)\in\mathcal{A}} \sup_{\substack{|x-x_0|\le a_n \\ |y-y_0|\le a_n}} \left|\left[\widehat{F}(x,y) - F(x,y)\right] - \left[\widehat{F}(x_0,y_0) - F(x_0,y_0)\right]\right| = \mathcal{O}\left(n^{-3/4}(\log n)^{\alpha_3}\right),$$

*where $\alpha_3 \ge 1$.*

To establish an i.i.d. representation for $\widehat{\mathbb{C}}$, we start by deriving a robust approximation for the copula estimator $\widehat{\mathbb{C}}0(u,v) = \widehat{F}\left(\widehat{F}1^{-1}(u), \widehat{F}_2^{-1}(v)\right)$. The oscillation result mentioned earlier will help us in developing this representation, as detailed in the upcoming Theorem 2 [see the proof of Theorem 2 in the appendix]. The following assumptions are required for our subsequent analysis:

**B1:** (i) The first partial derivatives of $\mathbb{C}$ are bounded.

(ii) There exist $a, b \in (0,1)$ such that $F_k$ $(k=1,2)$ is twice differentiable in $[F_k^{-1}(a) - \epsilon, F_k^{-1}(b) + \epsilon]$ for some $\epsilon > 0$.

(iii) $F_k^{(1)}$ is bounded away from zero and $F_k^{(2)}$ is bounded in absolute value $(k=1,2)$.

(iv) The second partial derivatives of $F$ are bounded.

Assumptions B1(ii)-B1(iii) are necessary to employ the i.i.d. representations of $F_1^{-1}$ and $F_2^{-1}$; see (14). Assumptions B1(i)-B1(iv) are required to derive the representation of $\widehat{\mathbb{C}}_0$ using the Taylor expansion of $F$. Let $L_1(z) = P[Z \le z, \delta = 1]$, $\eta_i(u) = u - \mathrm{I}\big(X_i \le F_1^{-1}(u)\big)$, and

$$\xi_i(v) = (1-v)\left[\frac{\mathrm{I}(Z_i \le F_2^{-1}(v), \delta_i = 1)}{\overline{L}(Z_i)} - \int_0^{F_2^{-1}(v)\wedge Z_i} \frac{\mathrm{d}L_1(t)}{\overline{L}^2(t)}\right].$$

**Theorem 2**

*Let $\mathcal{B} = (0,1) \times \big(0, F_2(\tau)\big)$, $(u_*, v_*) = \big(F_1^{-1}(u), F_2^{-1}(v)\big)$ and suppose Assumption B1 holds. The copula estimator $\widehat{\mathbb{C}}_0(u,v)$ admits for $(u,v) \in \mathcal{B}$ the representation*

$$\widehat{\mathbb{C}}_0(u,v) - \mathbb{C}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\big\{\chi_i^F(u_*, v_*) + \eta_i(u)\partial_1\mathbb{C}(u,v) + \xi_i(v)\partial_2\mathbb{C}(u,v)\big\} + r_n^1(u,v), \quad (6)$$

*where $\sup_{\mathcal{B}} |r_n^1(u,v)| = \mathcal{O}\left(n^{-3/4}(\log n)^{\alpha^*}\right)$ a.s., with $\alpha^* \ge 1$, and $\partial_1\mathbb{C}$ and $\partial_2\mathbb{C}$ denote the partial derivatives with respect to the first and second arguments of $\mathbb{C}$, respectively.*

The i.i.d. representation of $\widehat{\mathbb{C}}_0$ is derived from three key sources: the representations of $\widehat{F}$, the empirical quantile estimator $\widehat{F}1^{-1}$, and the Kaplan-Meier quantile estimator $\widehat{F}_2^{-1}$. Building on Theorem 2, we can similarly derive a representation for $\widehat{\mathbb{C}}$, given by:

$$\widehat{\mathbb{C}}(u,v) - \mathbb{C}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\big\{\chi_i^F\left(u_*^-, v_*^-\right) + \eta_i\big(u^+\big)\partial_1\mathbb{C}(u,v) - \xi_i\big(v^+\big)\partial_2\mathbb{C}(u,v)\big\} + r_n^2(u,v), \quad (7)$$

where $\sup_{\mathcal{B}} |r_n^2(u,v)| = \mathcal{O}_{a.s.}\left(n^{-3/4}(\log n)^{\alpha^*}\right)$, with $\alpha^* \ge 1$. Next, in Theorem 3 [see the proof of Theorem 3 in the appendix], we establish a triangular representation for $\widehat{\mathfrak{C}}$, which leads to a bivariate normal asymptotic distribution for this estimator. To achieve this, we require the following additional assumptions:

**K1:** (i) $\int_{\frac{u-1}{h}}^{\frac{u}{h}} K_{u,h}^2(t)\,\mathrm{d}t < \infty, \ (u = x, y)$.

(ii) $\int_{\frac{u-1}{h}}^{\frac{u}{h}} t^2\,|K_{u,h}(t)|\,\mathrm{d}t < \infty, \ (u = x, y)$.

**Theorem 3**

*Suppose Assumptions B1 and K1 hold, the second partial derivatives of $\mathfrak{C}$ are bounded, and let*

$$\chi_i^{\mathbb{C}}(u,v) = \chi_i^F\left(F_1^{-1}(u)^-, F_2^{-1}(v)^-\right) + \eta_i\big(u^+\big)\partial_1\mathbb{C}(u,v) - \xi_i\big(v^+\big)\partial_2\mathbb{C}(u,v).$$

*The copula density estimator $\widehat{\mathfrak{C}}(x,y)$ admits for $(x,y) \in \mathcal{B}$ the representation*

$$\begin{aligned}\widehat{\mathfrak{C}}(u,v) - \mathfrak{C}(u,v) =& \frac{1}{nh^2}\sum_{i=1}^{n}\int_{[-1,1]^2}\Big\{\chi_i^{\mathbb{C}}(u-th, v-sh) - \chi_i^{\mathbb{C}}(u-th, 1)\mathrm{I}(v-sh \le 1)\\ &- \chi_i^{\mathbb{C}}(1, v-sh)\mathrm{I}(u-th \le 1)\Big\}\,\mathrm{d}K_{u,h}(t)\,\mathrm{d}K_{v,h}(s) + r_n^{\mathfrak{C}}(u,v), \quad (8)\end{aligned}$$

*with $\sup_{\mathcal{B}} |r_n^{\mathfrak{C}}(u,v)| = \mathcal{O}_{a.s.}\left(n^{-3/4}h^{-2}(\log n)^{\alpha^*} + h^2\right)$.*

From the above theorem, we obtain the following Corollary.

**Corollary 1**

*Suppose Assumptions B1 and K1 hold, and* $nh^6$, $(\log n)^{4\alpha^*}/nh^4 \to 0$ *as* $n \to \infty$ *and* $h \to 0$. *Then, for* $(u,v) \in \mathcal{B}$, $n^{1/2}h\left[\widehat{\mathfrak{C}}(u,v) - \mathfrak{C}(u,v)\right]$ *converges in distribution to a zero-mean bivariate normal distribution* $N\Big(0, \sigma^2_{\mathfrak{c}}(u,v)\Big)$, *with the variance*

$$\sigma^2_{\mathfrak{c}}(u,v) = \frac{\mathfrak{C}(u,v)}{1 - G\big[F_2^{-1}(v)\big]} \left\{\int_{-1}^{1} K^2(t)\,\mathrm{d}t\right\}^2.$$

The proof of Corollary 1 follows from the triangular representation (8) by using the Lindeberg-Feller CLT theorem. Furthermore, the variance of $\widehat{\mathfrak{C}}$ satisfy

$$Var\big[\widehat{\mathfrak{C}}(u,v)\big] = \frac{1}{nh^2} \times \frac{\mathfrak{C}(u,v)}{1 - G\big[F_2^{-1}(v)\big]} \int_{\frac{u-1}{h}}^{\frac{u}{h}} K^2_{u,h}(t)\,\mathrm{d}t \times \int_{\frac{v-1}{h}}^{\frac{v}{h}} K^2_{v,h}(s)\,\mathrm{d}s + o\left(\frac{1}{nh^2}\right) \quad (9)$$

and, in the interior region $[h, 1-h]$, is simplified to

$$Var_I\big[\widehat{\mathfrak{C}}(u,v)\big] = \frac{1}{nh^2} \times \frac{\mathfrak{C}(u,v)}{1 - G\big[F_2^{-1}(v)\big]} \left\{\int_{-1}^{1} K^2(t)\,\mathrm{d}t\right\}^2 + o\left(\frac{1}{nh^2}\right).$$

**Remark 1**

*In the class of bounded copula densities, the variance of the local-polynomial estimator described in (9) remains stable along the boundary of* $\mathcal{B}$. *However, the situation differs significantly for the transformation estimator:*

$$\widehat{\mathfrak{C}}_T(u,v) = \frac{n^{-1}h^{-2}}{\phi\big[\Phi^{-1}(u)\big]\,\phi\big[\Phi^{-1}(v)\big]} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\overline{G}}(Z_i)} K_*\left(\frac{\Phi^{-1}(u) - \Phi^{-1}\big(\widehat{F}_1(X_i)\big)}{h}\right) K_*\left(\frac{\Phi^{-1}(v) - \Phi^{-1}\big(\widehat{F}_2(Z_i)\big)}{h}\right),$$

*where* $\Phi : \mathbb{R} \mapsto [0,1]$ *denotes the transformation distribution,* $\phi$ *its density, and* $K_* : \mathbb{R} \mapsto \mathbb{R}^+$ *a kernel function. The derivation of the asymptotic variance of* $\widehat{\mathfrak{C}}_T$ *leads to the following expression*

$$Var\big[\widehat{\mathfrak{C}}_T(u,v)\big] = \frac{1}{nh^2} \times \frac{1}{\phi\big[\Phi^{-1}(u)\big]\,\phi\big[\Phi^{-1}(v)\big]} \times \frac{\mathfrak{C}(u,v)}{\overline{G}\big[F_2^{-1}(v)\big]} \left\{\int_{\mathbb{R}} K^2_*(t)\,\mathrm{d}t\right\}^2 + o\left(\frac{1}{nh^2}\right).$$

*For bounded* $\mathfrak{C}$, *the variance* $Var\big[\widehat{\mathfrak{C}}_T(u,v)\big]$ *diverges near* $u = 0, 1$ *or/and* $v = 0$. *This contrasts sharply with the behavior of the local-polynomial estimator in (9) and Corollary 1, where the variance remains finite at the boundaries of* $\mathcal{B}$. *In cases where* $\mathfrak{C}$ *exhibits an unbounded density at* $u = 0, 1$ *or/and* $v = 0, \varrho$, *the ratio* $var\big[\widehat{\mathfrak{C}}_T\big]/var\big[\widehat{\mathfrak{C}}\big] \simeq \kappa_h(u,v)/\big(\phi\big[\Phi^{-1}(u)\big], \phi\big[\Phi^{-1}(v)\big]\big) \to \infty$

*as $u \to 0, 1$ or $v \to 0$, where $\kappa_h$ is a bounded function. This observation underscores the stability and efficiency of the local-polynomial estimator $\widehat{\mathfrak{C}}$ at the boundary of $\mathcal{B}$. Accurate boundary estimation is crucial in real-world data contexts, particularly where dependencies intensify in the tails of the joint distribution. This is especially evident in financial markets, where stock returns show increased dependency during periods of market distress, marked by significant negative returns.*

**Remark 2**

*The application of a boundary kernel function establishes a revised balance between bias and variance at each $(u, v) \in \mathcal{B}$. In this case, a viable option for determining the smoothing parameter involves selecting a local bandwidth $h = h_n(u, v)$ based on the asymptotic mean square error of $\widehat{\mathfrak{C}}$;*

$$\mathrm{MSE}(\widehat{\mathfrak{C}}) = \frac{1}{nh^2}\, \sigma_h^2(u, v) + h^4\, \beta_h^2(u, v) + o\left(\frac{1}{nh^2} + h^4\right),$$

*where $\sigma_h^2/(nh^2)$ and $h^2\beta_h$ are the asymptotic variance and bias of $\widehat{\mathfrak{C}}$, respectively, with*

$$\sigma_h^2(u, v) = \frac{\mathfrak{C}(u, v)}{1 - G\big[F_2^{-1}(v)\big]} \int_{\frac{u-1}{h}}^{\frac{u}{h}} K_{u,h}^2(t)\, \mathrm{d}t \times \int_{\frac{v-1}{h}}^{\frac{v}{h}} K_{v,h}^2(t)\, \mathrm{d}t,$$

$$\beta_h(u, v) = \frac{1}{2}\left\{\frac{\partial^2 \mathfrak{C}(u, v)}{\partial u^2} \int_{\frac{u-1}{h}}^{\frac{u}{h}} t^2 K_{u,h}(t)\, \mathrm{d}t + \frac{\partial^2 \mathfrak{C}(u, v)}{\partial v^2} \int_{\frac{v-1}{h}}^{\frac{v}{h}} t^2 K_{v,h}(t)\, \mathrm{d}t\right\}.$$

*Thus, the optimal local bandwidth is defined as*

$$\widehat{h}_1(u, v) = \arg\min_h \left\{\sigma_h^2(u, v)/(nh^2) + h^4\, \beta_h^2(u, v)\right\},$$

*and, in the interior region $[h, 1-h]$, it can be simplified to achieve*

$$\widehat{h}_1(u, v) = \left\{\frac{\sigma_{\mathfrak{c}}^2(u, v)}{\beta_{\mathfrak{c}}^2(u, v)}\right\}^{\frac{1}{6}} n^{-\frac{1}{6}},$$

*where $\sigma_{\mathfrak{c}}^2$ is defined in Corollary 1 and*

$$\beta_{\mathfrak{c}}(u, v) = \frac{1}{2}\left[\frac{\partial^2 \mathfrak{C}(u, v)}{\partial u^2} + \frac{\partial^2 \mathfrak{C}(u, v)}{\partial v^2}\right] \int_{-1}^{1} t^2\, K(t)\, \mathrm{d}t.$$

*Several plug-in methods from the literature are available for determining $\widehat{h}_1(u, v)$. One approach involves using a parametric model for the copula function $\mathbb{C}$ as a benchmark. The Frank copula* *serves as a suitable choice for this reference copula. We can estimate $\sigma_h^2$ and $\beta_h$ using either a likelihood or moment approach to estimate the parameter(s) of the parametric copula. This estimation process involves substituting $F_2$ and $G$ (in $\sigma_h^2$ and/or $\sigma_{\mathfrak{C}}^2$) with their estimators $\widehat{F}_2$ and $\widehat{G}$.*

*Another approach involves selecting a global bandwidth $h$ that minimizes the integrated squared error $\mathrm{ISE}(h) = \int_{\mathcal{B}} \left[\widehat{\mathfrak{C}}(u,v;h) - \mathfrak{C}((u,v)\right]^2 \mathrm{d}u\,\mathrm{d}v$. This is equivalent to choosing $h$ that minimizes*

$$\mathrm{ISE}_*(h) = \int_{\mathcal{B}} \widehat{\mathfrak{C}}(u,v;h)^2\,\mathrm{d}u\,\mathrm{d}v - 2\int_{\mathcal{B}} \widehat{\mathfrak{C}}(u,v;h)\,\mathrm{d}\mathbb{C}(u,v).$$

*The unknown copula function $\mathbb{C}$ in the second term on the right-hand side of the latter equality can be replaced by its estimator $\widehat{\mathbb{C}}$. The data-driven bandwidth is then*

$$\widehat{h}_2 = \arg\min_h \left\{ \int_{\mathcal{B}} \widehat{\mathfrak{C}}_{-i}(u,v;h)^2\,\mathrm{d}u\,\mathrm{d}v - \frac{2}{n}\sum_{i=1}^{n} \frac{\delta_i}{1-\widehat{G}(Z_i)}\,\widehat{\mathfrak{C}}_{-i}\big(\widehat{F}_1(X_i), \widehat{F}_2(Z_i); h\big) \right\}, \tag{10}$$

*where $\widehat{\mathfrak{C}}_{-i}$ is the leave-one-out estimate of $\mathfrak{C}$, defined as*

$$\widehat{\mathfrak{C}}_{-i}(u,v;h) = \frac{1}{nh^2}\sum_{\substack{\ell=1\\ \ell\neq i}}^{n} \frac{\delta_\ell}{1-\widehat{G}(Z_\ell)} K_{u,h}\left(\frac{u-\widehat{F}_1(X_\ell)}{h}\right) K_{v,h}\left(\frac{v-\widehat{F}_2(Z_\ell)}{h}\right).$$

# 3 Estimation of Kendall's tau and Spearman's rho

Next, we introduce nonparametric estimators for two measures of association, Kendall's tau and Spearman's rho, specifically adapted for right-censored time-to-event data. We begin with Kendall's tau, which quantifies the strength and direction of the ordinal association between two variables by using the concepts of concordance and discordance. In essence, it measures the difference between the probabilities of concordance and discordance between two random variables $X$ and $Y$, and is defined as follows:

$$\tau_{X,Y} = P\Big[(X - X_0)(Y - Y_0) > 0\Big] - P\Big[(X - X_0)(Y - Y_0) < 0\Big],$$

where $(X_0, Y_0)$ and $(X, Y)$ are i.i.d. random vectors. (26) discussed the limitations of estimating Kendall's tau under right-censoring. (26) discussed the limitations of estimating Kendall's tau under right-censoring. In response, hereafter we propose an estimator that adapts to right-censoring.

Since the tail region information of the survival function of $Y$ might not be identifiable in $[\varrho, \infty)$ due to censoring, we estimate a truncated version of Kendall's tau:

$$\tau_{X,Y} = 4 \int_{\mathcal{B}} \mathbb{C}(u,v) \, \mathrm{d}\mathbb{C}(u,v) - 1,$$

defined on the set $\mathcal{B} = [0,1] \times \big[0, F_2(\varrho)\big)$. Using the above expression, we define an estimator of $\tau_{X,Y}$ for right-censored data as follows:

$$\widehat{\tau}_{X,Y} = 4 \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\delta_i \, \delta_j}{n^2 \big[1 - \widehat{G}(Z_i)\big] \big[1 - \widehat{G}(Z_j)\big]} \, \mathrm{I}\Big(X_j \leq X_i, Z_j \leq Z_i\Big) - 1. \tag{11}$$

The weight $\delta_i \delta_j / \big[n^2 \big[1 - \widehat{G}(Z_i)\big] \big[1 - \widehat{G}(Z_j)\big]$ accounts for the data censoring, replacing the uniform weight $1/n$ used in the empirical version of $\widehat{\tau}_{X,Y}$ for uncensored data. The following Theorem provides the asymptotic distribution of the Kendall's tau estimator [see the proof of Theorem 4 in the appendix].

**Theorem 4** *Suppose that the conditions in Theorems 1 hold. We have* $\sqrt{n}\big[\widehat{\tau}_{X,Y} - \tau_{X,Y}\big]$ *converges weakly to the normal variable* $Z_\tau$*, given by*

$$Z_\tau = 4 \left\{ \int_{\mathcal{B}} \mathbb{C}(u,v) \, \mathrm{d}\mathbb{C}_L(u,v) + \int_{\mathcal{B}} \mathbb{C}_L(u,v) \, \mathrm{d}\mathbb{C}(u,v) \right\},$$

*where* $\mathbb{C}_L$ *is the limiting process of* $\sqrt{n}\big[\widehat{\mathbb{C}}(u,v) - \mathbb{C}(u,v)\big]$.

We now proceed to the nonparametric estimation of Spearman's rho, an alternative measure of association between random variables. Spearman's rho evaluates the strength of monotonic relationships between two variables and is defined as follows:

$$\rho_{X,Y} = 3 \left\{ P\Big[(X - X_0)(Y - Y_1) > 0\Big] - P\Big[(X - X_0)(Y - Y_1) < 0\Big] \right\},$$

where $(X_0, Y_0)$, $(X_1, Y_1)$ and $(X, Y)$ are i.i.d. random vectors. This measure can be re-written as

$$\rho_{X,Y} = 12 \int_{\mathcal{B}} uv \, \mathrm{d}\mathbb{C}(u,v) - 3.$$

Similar to Kendall's tau, to avoid identifiability issues in $[\varrho, \infty)$ caused by right censoring, we define $\rho_{X,Y}$ on $\mathcal{B}$. Using the above expression, we define an estimator of $\rho_{X,Y}$ for right-censored data as follows:

$$\widehat{\rho}_{X,Y} = 12 \sum_{i=1}^{n} \frac{\delta_i}{n\big[1 - \widehat{G}(Z_i)\big]} \, \widehat{F}_1(X_i) \, \widehat{F}_2(Z_i) - 3. \tag{12}$$

The limit distribution of $\widehat{\rho}_{X,Y}$ is established in the following theorem [see the proof of Theorem 5 in the appendix].

**Theorem 5** *Under the assumptions of Theorems 1, $\sqrt{n}\big[\widehat{\rho}_{X,Y} - \rho_{X,Y}\big]$ converges in distribution to the Gaussian variable*

$$Z_\rho = 12 \int_{\mathcal{B}} u\, v \, \mathrm{d}\mathbb{C}_L(u, v).$$

# 4 Simulation and data analysis

In this section, we provide a comprehensive examination of our theoretical results through simulation studies and data analysis. The first subsection focuses on simulations, where we evaluate the performance of our proposed local polynomial estimators for copula density under various Data Generating Processes (DGPs) that reflect practical scenarios. This analysis offers valuable insights into the bias and efficiency of the estimators, as measured by the integrated squared error. In the second subsection, we apply these estimators to real-world data, conducting rigorous analyses to validate our theoretical findings and extract meaningful economic conclusions.

## 4.1 Simulation study

We perform a simulation study to assess the finite-sample performance of our estimator, $\widehat{\mathfrak{C}}$ in (4), in comparison to the transformation estimator, $\widehat{\mathfrak{C}}_T$, as described in Remark 1. The performance is evaluated using the integrated squared error (ISE) for both estimators:

$$\mathrm{ISE}(\widehat{\mathfrak{C}}) = \int_{\mathcal{B}_*} \Big[\widehat{\mathfrak{C}}(u, v) - \mathfrak{C}(u, v)\Big]^2 \mathrm{d}u \, \mathrm{d}v,$$

where $\mathcal{B}_* = [0.005, 0.995] \times [0.005, \varrho)$.

For our simulation settings, we consider five copula models: Frank, Gumbel, Clayton, Joe, and Gaussian. We evaluate and compare the performance of the local-polynomial and transformation-kernel-based estimators by computing their integrated squared error, $\mathrm{ISE}(\widehat{\mathfrak{C}})$, across various scenarios. Specifically, we examine three levels of dependence ($\tau = 0.25, , 0.5, , 0.75$), two censoring levels ($20\%,\ 40\%$), and three sample sizes ($n = 100,\ 200,\ 400$). The kernel function used is $K(x) = 0.75(1 - x^2), \mathrm{I}_{[-1,1]}(x)$, and the bandwidth $h_n$ is selected to minimize the integrated squared error.

**Table 1.** Integrated Squared Error (ISE) under various copula models with Kendall' tau $\tau = 0.25$, based on the average of 1000 simulations of ISE ($\times 10^{-2}$).

| | | $n = 100$ | | $n = 200$ | | $n = 400$ | |
|---|---|---|---|---|---|---|---|
| Copula | censoring | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ |
| Frank | $20\%$ | 10.83 | 30.47 | 5.28 | 16.74 | 2.63 | 8.82 |
| | $40\%$ | 21.88 | 41.95 | 11.02 | 25.93 | 5.29 | 14.36 |
| Clayton | $20\%$ | 13.57 | 31.04 | 8.48 | 17.57 | 5.73 | 9.02 |
| | $40\%$ | 24.52 | 42.45 | 13.62 | 25.10 | 8.17 | 13.90 |
| Gaussian | $20\%$ | 11.57 | 31.87 | 5.94 | 17.90 | 3.04 | 9.03 |
| | $40\%$ | 23.02 | 43.19 | 11.88 | 26.80 | 5.86 | 14.76 |
| Gumbel | $20\%$ | 14.07 | 34.00 | 7.86 | 19.30 | 4.91 | 10.20 |
| | $40\%$ | 27.61 | 47.19 | 14.29 | 29.86 | 8.18 | 17.27 |
| Joe | $20\%$ | 16.91 | 36.46 | 10.94 | 20.70 | 7.58 | 10.93 |
| | $40\%$ | 31.66 | 51.85 | 17.46 | 33.67 | 11.15 | 20.06 |

Tables 1, 2, and 3 present the average integrated squared error (ISE) for both local-polynomial and transformation-kernel-based estimators, based on 1000 replications. The results demonstrate that our estimator, $\widehat{\mathfrak{C}}$, generally outperforms the transformation-kernel-based estimator $\widehat{\mathfrak{C}}_T$ across most copula models. An exception occurs with the Clayton model, where $\widehat{\mathfrak{C}}_T$ performs well for $\tau = 0.5$ and $\tau = 0.75$; however, our estimator still shows superior performance for $\tau = 0.25$. Furthermore, for both estimators, find that the average ISE decreases as the sample size increases and the level of censoring decreases.

## 4.2 Indemnity payments and allocated loss adjustment expenses

We illustrate the methodology described in Sections 2 and 3 by analyzing a dataset of censored insurance indemnity claims. The dataset consists of $n = 1500$ liability claims selected based on late settlement lags from an insurance company. Each claim includes the indemnity payment (the loss) $Y$, the failure indicator $\delta$ (with $\delta = 1$ if $Y$ is observed and $\delta = 0$ if censored), and the allocated loss adjustment expense (ALAE) $X$. In this context, the loss variable $Y$ is subject to random right-censoring, while the ALAE variable $X$ is uncensored. Specifically, each claim also contains information on the policy limit, which represents the maximum allowable claim amount. The policy limit induces censoring, as losses cannot exceed this specified maximum.

**Table 2.** Integrated Squared Error (ISE) under various copula models with Kendall' tau $\tau = 0.5$, based on the average of 1000 simulations of ISE ($\times 10^{-2}$).

| | | $n = 100$ | | $n = 200$ | | $n = 400$ | |
|---|---|---|---|---|---|---|---|
| Copula | censoring | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ |
| Frank | 20% | 10.4 | 31.1 | 5.83 | 17.81 | 3.45 | 9.76 |
| | 40% | 22.6 | 44.4 | 11.29 | 27.48 | 6.18 | 15.83 |
| Clayton | 20% | 52.0 | 40.8 | 42.10 | 26.51 | 34.50 | 19.30 |
| | 40% | 64.2 | 52.5 | 51.20 | 35.10 | 41.96 | 25.28 |
| Gaussian | 20% | 13.7 | 33.7 | 8.33 | 19.17 | 5.83 | 10.33 |
| | 40% | 24.8 | 48.2 | 14.53 | 30.15 | 8.97 | 17.72 |
| Gumbel | 20% | 24.5 | 43.1 | 18.63 | 25.15 | 14.30 | 14.74 |
| | 40% | 39.2 | 62.3 | 27.01 | 42.24 | 18.90 | 25.70 |
| Joe | 20% | 50.3 | 60.6 | 40.06 | 42.10 | 27.13 | 32.58 |
| | 40% | 68.1 | 89.9 | 52.27 | 63.21 | 40.35 | 48.23 |

**Table 3.** Integrated Squared Error (ISE) under various copula models with Kendall' tau $\tau = 0.75$, based on the average of 1000 simulations of ISE.

| | | $n = 100$ | | $n = 200$ | | $n = 400$ | |
|---|---|---|---|---|---|---|---|
| Copula | censoring | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ | $\widehat{\mathfrak{C}}$ | $\widehat{\mathfrak{C}}_T$ |
| Frank | 20% | 0.28 | 0.48 | 0.18 | 0.35 | 0.11 | 0.25 |
| | 40% | 0.46 | 0.66 | 0.29 | 0.47 | 0.19 | 0.35 |
| Clayton | 20% | 3.71 | 2.90 | 3.34 | 1.82 | 2.93 | 1.51 |
| | 40% | 4.07 | 3.13 | 3.65 | 2.21 | 3.23 | 1.87 |
| Gaussian | 20% | 0.50 | 0.56 | 0.35 | 0.38 | 0.24 | 0.29 |
| | 40% | 0.74 | 0.80 | 0.48 | 0.56 | 0.34 | 0.41 |
| Gumbel | 20% | 1.12 | 1.22 | 0.91 | 0.95 | 0.65 | 0.78 |
| | 40% | 1.45 | 1.63 | 1.15 | 1.28 | 0.94 | 0.97 |
| Joe | 20% | 3.00 | 3.13 | 2.69 | 2.55 | 2.34 | 1.94 |
| | 40% | 3.46 | 3.83 | 3.04 | 3.27 | 2.60 | 2.74 |

For claims where the policy limit was not available, we assumed there was no policy limit.

The presence of censoring complicates the estimation of the dependence structure and joint distribution of losses and expenses; see (28) and (29). The latter paper focused on modeling the relationship between indemnity payments and ALAE using parametric copulas, highlighting the sensitivity of reinsurance premium calculations to copula model misspecification.

In this section, our primary objective is to apply our local polynomial estimator for cop-

**Table 4.** Kendall and Spearman estimates for Loss $\&$ ALAE insurance data, with corresponding Bootstrap 95$\%$ confidence interval (CI).

| Variables | $\widehat{\tau}_{X,Y}$ | CI of $\tau$ | $\widehat{\rho}_{X,Y}$ | CI of $\rho$ |
|---|---|---|---|---|
| ALAE $\&$ Loss | 0.333 | [0.297, 0.364] | 0.512 | [0.444, 0.560] |

ula density to nonparametrically estimate and examine the dependence between losses and expenses. This approach aims to address challenges posed by censoring and potential model misspecification.

Using the theoretical results in section 3, Table 4 presents the estimated values of Kendall's tau and Spearman's rho, along with their corresponding $95\%$ confidence intervals. These intervals are constructed using the bootstrap methods discussed in (5; 20; 27), among others. From the original sample, we draw simple random vectors $\{(Y_i, \delta_i, X_i), i = 1, \ldots, n\}$ with replacement. We perform $B = 1000$ bootstrap replications, and for each $b^{th}$ bootstrap sample, we compute the estimators of Kendall's tau $\hat{\tau}^{(b)}$ and Spearman's rho $\hat{\rho}^{(b)}$, for $b = 1, \ldots, B$. The $95\%$ bootstrap confidence intervals for $\tau$ and $\rho$ are then constructed using the $2.5^{th}$ and $97.5^{th}$ percentiles of the bootstrap estimates $\hat{\tau}^{(b)}$ and $\hat{\rho}^{(b)}$, respectively.

The results in Table 4 clearly indicate a positive and a statistically significant relationship between the censored Loss variable and ALAE. Furthermore, Figure 1(a) displays the nonparametric estimate of the copula density, revealing prominent peaks near the points $(1, 1)$ and $(0, 0)$ that correspond to the tails of the joint distribution of losses and ALAE. This suggests that large losses are significantly associated with high values of ALAE, while smaller losses are associated to lower ALAE values. Figure 1(b) shows the copula density based on the parametric Gumbel model, with an estimated parameter value of $\widehat{\theta} = 1.45$, as used by (29). Notably, both the parametric Gumbel model and our nonparametric estimator from (4) exhibit similar patterns in the copula density, with peaks at $(1, 1)$ and $(0, 0)$. This is not surprising as (29) used the procedure developed by (30) to identify the appropriate copula model for the variables losses and ALAE. For our nonparametric estimation, we used the kernel function $K(x) = 0.75(1 - x^2), \mathrm{I}_{[-1,1]}(x)$, and the bandwidth $h$ was selected using the formula in (10) as detailed in Remark 2.

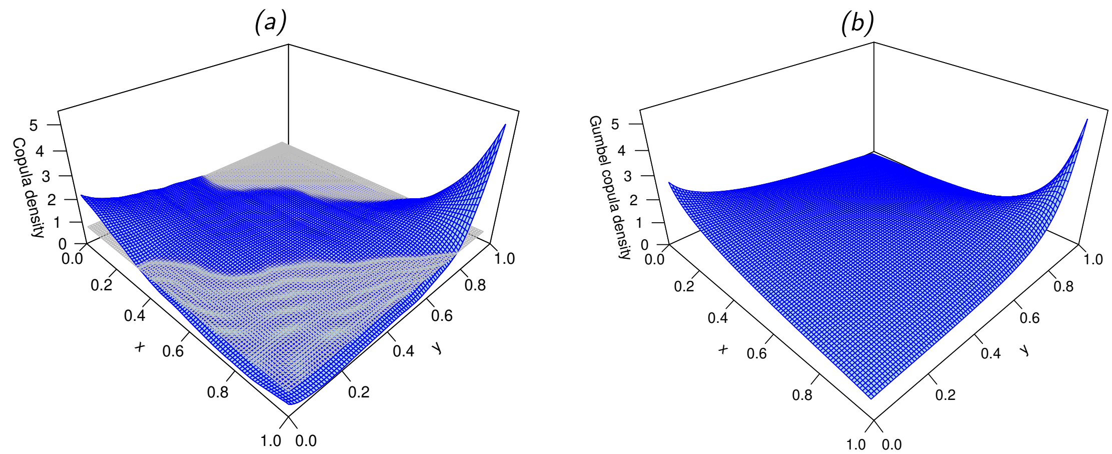


*Figure 1: (a) Local-polynomial estimate of copula density (blue) and plane $z = 1$ (grey), and (b) Gumbel parametric model ($\widehat{theta} = 1.45$) of copula density, for Loss and ALAE.*

## 4.3 Disregard rate on the duration of unemployment

In this second empirical application, we investigate the dependence structure between unemployment duration and the disregard rate using employment-censored data, as discussed in (31). The dataset comprises unemployment durations for $n = 3241$ individuals, of whom 1986 found employment, while 1255 were right-censored during the follow-up period. In this analysis, unemployment duration $Y$ represents the time elapsed between an individual's last and new job, with the censoring indicator $\delta$ defined as $1$ if the individual found a job and $0$ if they remained unemployed. The disregard rate is denoted by $X$. In a previous study using the same dataset, (31) employed a structural parametric econometric model to analyze how unemployment insurance (UI) policies influence workers' decisions to accept part-time work or remain unemployed while searching for full-time positions. Their findings suggest that more generous UI benefits are associated with longer durations of unemployment. Additionally, their research indicates that UI rules allowing for partial benefit collection while working part-time can incentivize part-time employment. The primary objective of this empirical study is to apply our local polynomial estimator for copula density to reassess the relationship between the disregard rate and the duration of unemployment in a model-free framework.

The results are provided in Figure 2 and Table 5. Figure 2 illustrates the local-polynomial estimate of the copula density for unemployment duration and disregard rate, along with the

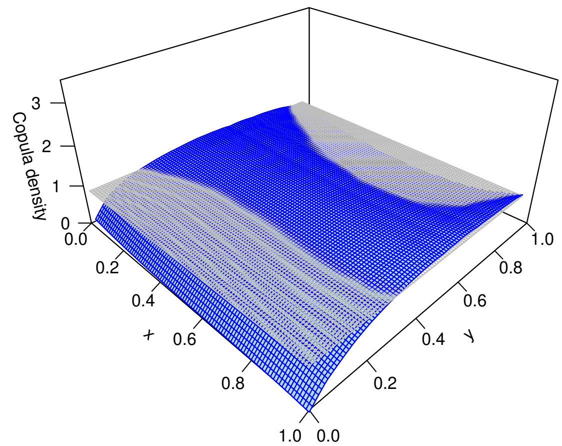


*Figure 2: Local-polynomial estimate of copula density (blue) and plane $z = 1$ (grey) for unemployment duration and disregard rate.*

**Table 5.** Kendall and Spearman estimates for Unemployement data, with corresponding Bootstrap 95$\%$ confidence interval (CI).

| Variables | $\widehat{\tau}_{X,Y}$ | CI of $\tau$ | $\widehat{\rho}_{X,Y}$ | CI of $\rho$ |
|---|---|---|---|---|
| Disreg-Rate $\&$ Unemp-Dur | 0.0172 | [-0.0811,0.1250] | -0.0303 | [-0.3357,0.3038] |

plane $z = 1$ shown in grey color. The copula density curve appears nearly flat at the level of the $z = 1$ plane, indicating weak relationship between unemployment duration and disregard rate. This result is confirmed by Table 5 that shows the estimated values of Kendall's and Spearman's rank correlation coefficients between the two variables of interest, along with their respective $95\%$ confidence intervals. This suggests that the disregard rate has little to no significant impact on an individual's likelihood of searching for a new job after losing one.

# 5 Conclusion

In this paper, we have developed a local-polynomial estimator for copula density, specifically designed to address boundary bias issues in the context of censored Time-to-Events data. Our approach also includes robust estimators for Kendall's tau and Spearman's rho. Through a rigorous analysis of the asymptotic properties of these estimators, we derived an i.i.d. representation with an improved remainder term rate for a bivariate cdf estimator under right-censored data conditions. This served as a foundation for further deriving an i.i.d. representation for a copula cdf estimator and establishing a functional Central Limit Theorem (CLT) for the copula density estimator. Additionally, the weak convergence of the Kendall's tau and Spearman's rho estimators has been proven. Moreover, we derive both local and global optimal bandwidth parameters necessary for the nonparametric estimation of copulas.

Our findings indicate that the local-polynomial estimator exhibits stability and greater efficiency near the boundaries of $[0, 1]^2$, addressing a crucial challenge in kernel estimation of copulas. The practical implications of our work were validated through an extensive Monte Carlo simulation exercise, which showcased the finite sample performance of the local-polynomial estimator.

To illustrate the practical usefulness of our estimators, we applied them to two real-world datasets. The first dataset, from the insurance industry, allowed us to explore the dependence between indemnity payments and allocated loss adjustment expenses. The second dataset examined the impact of the disregard rate on the duration of unemployment. These applications underscore the broad applicability of our estimators across different datasets.

In conclusion, our research makes a significant contribution to the literature on dependence estimation by offering a more efficient and reliable approach for copula density estimation in the presence of boundary bias and censored data. The local-polynomial estimator of copula, along with the nonparametric estimators for Kendall's tau and Spearman's rho, presents a valuable tool for both theoretical analysis and practical applications.

**Acknowledgements**.

This work has been supported by the Canadian Statistical Sciences Institute and Statlab-CRM (Canada), the Natural Sciences and Engineering Research Council of Canada (NSERC) and Le Fond de recherche du Québec (Nature et technologies).

# A Appendix: Proofs of main results

**Proof of Proposition 1.**

Let $H(u,v,0) = P[X \leq u, Z \leq v, Y - C \leq 0] = F_{X,Z,\delta}(u,v,1)$ and $\widehat{H}$ the empirical counterpart of $H$. First, Notice that

$$\widehat{F}(x,y) = \iint_{u \leq x, v \leq y} \frac{1}{\widehat{\overline{G}}(v)} \, \mathrm{d}\widehat{H}(u,v,0),$$

and the difference $\widehat{F}(x,y) - F(x,y)$ can be written as

$$\begin{aligned}\widehat{F}(x,y) - F(x,y) = &\int_{v \leq y} \left[\widehat{\overline{G}}^{-1}(v) - \overline{G}^{-1}(v)\right] \mathrm{d}\big[\widehat{H}(x,v,0) - H(x,v,0)\big] - F(x,y) \\ &+ \int_{v \leq y} \overline{G}^{-1}(v) \, \mathrm{d}\widehat{H}(x,v,0) + \int_{v \leq y} \left[\widehat{\overline{G}}^{-1}(v) - \overline{G}^{-1}(v)\right] \mathrm{d}H(x,v,0).\end{aligned}$$

Using the uniform convergence result of the Kaplan-Meier estimator $\widehat{G}$, this difference is equal to

$$\begin{aligned}\widehat{F}(x,y) - F(x,y) = &\int_{v \leq y} \frac{\widehat{G}(v) - G(v)}{\overline{G}^2(v)} \, \mathrm{d}\big[\widehat{H}(x,v,0) - H(x,v,0)\big] - F(x,y) \\ &+ \int_{v \leq y} \overline{G}^{-1}(v) \, \mathrm{d}\widehat{H}(x,v,0) + \int_{v \leq y} \frac{\widehat{G}(v) - G(v)}{\overline{G}^2(v)} \, \mathrm{d}H(x,v,0) + r_{1,n}(x,y),\end{aligned} \tag{13}$$

where $\sup_{x,y} |r_{1,n}(x,y)| = \mathcal{O}_{a.s.}\left(n^{-1} \log\log n\right)$. Let

$$I_n(x,y) = \int_{v \leq y} \frac{\widehat{G}(v) - G(v)}{\overline{G}^2(v)} \, \mathrm{d}\big[\widehat{H}(x,v,0) - H(x,v,0)\big].$$

In the following, we show that the rate of $\sup_{x,y} |I_n(x,y)|$ is of order $\mathcal{O}_{a.s.}\left(n^{-3/4}(\log n)^{\alpha_1}\right)$ ($\alpha_1 \geq 1$). We have

$$|I(x,y)| \leq \|\widehat{G} - G\|.\|\overline{G}^{-2}\|. \int_{v=0}^{v=y} \left|\mathrm{d}\big[\widehat{H}(x,v,0) - H(x,v,0)\big]\right|. \tag{14}$$

Divide $[0,y]$ into $m$ sub-intervals $[0,y_1]$, $[y_1,y_2], \ldots, [y_{m-1},y_m]$ of equal length $\ell = a_0 n^{-1/2}(\log n)^q$ ($q \geq 1/2$ and $a_0 > 0$ is some constant), so $m$ is of order $\mathcal{O}\left(n^{1/2}(\log n)^{-q}\right)$. Then,

$$\begin{aligned}|I(x,y)| &\leq \frac{\|\widehat{G} - G\|}{\|\overline{G}^2\|}. \sum_{i=0}^{m-1} \int_{v=y_i}^{v=y_{i+1}} \left|\mathrm{d}\big[\widehat{H}(x,v,0) - H(x,v,0)\big]\right| \\ &\leq \frac{\|\widehat{G} - G\|}{\|\overline{G}^2\|}. \sum_{i=0}^{m-1} \sup_{u,v \in [y_i, y_{i+1}]} \left|\big[\widehat{H}(x,v,0) - H(x,v,0)\big] - \big[\widehat{H}(x,v,0) - H(x,u,0)\big]\right|.\end{aligned} \tag{15}$$

The sup-norm term, inside the summation, on the R.H.S. of (A) is of order $\mathcal{O}_{a.s.}\left(n^{-3/4}(\log n)^{\frac{1+q}{2}}\right)$, as $n \to \infty$, by the oscillation result in (22) (see theorem 2.3). Since $\|\widehat{G} - G\|$ and $m$ are of order $\mathcal{O}_{a.s.}\left(n^{-1/2}\big(\log\log n\big)^{1/2}\right)$ and $\mathcal{O}\left(n^{-1/2}(\log n)^{-q}\right)$, respectively, the term on the R.H.S. of (A) is of order $\mathcal{O}_{a.s.}\left(n^{-3/4}(\log n)^{\alpha_1}\right)$ $(\alpha_1 \geq 1)$. Hence,

$$\sup_{x,y} |I_n(x,y)| = \mathcal{O}_{a.s.}\left(n^{-3/4}(\log n)^{\alpha_1}\right),$$

and therefore,

$$\widehat{F}(x,y) - F(x,y) = \int_{v\leq y} \left[\widehat{G}(v) - G(v)\right] \frac{\mathrm{d}H(x,v,0)}{\overline{G}^2(v)} + \int_{v\leq y} \overline{G}^{-1}(v)\,\mathrm{d}\widehat{H}(x,v,0) - F(x,y) + r_{2,n}(x,y), \tag{16}$$

where $\sup_{x,y}|r_{2,n}(x,y)| = \mathcal{O}_{a.s.}\left(n^{-3/4}(\log n)^{\alpha_1}\right)$. By using the i.i.d. representation of $\widehat{G}(v) - G(v)$ in (13), we complete the proof. ■

**Proof of Theorem 1.**

Denote $\widehat{\mathbb{F}}(x,y) = \widehat{F}(x,y) - F(x,y)$, $\widehat{\mathbb{G}}(y) = \widehat{G}(y) - G(y)$ and $\widehat{\mathbb{H}}(x,y) = \widehat{H}(x,y) - H(x,y)$, where $H(x,y) = H(x,y,0)$ is defined in Lemma 1' proof and $\widehat{H}$ is its empirical counterpart. Let $x_0$ and $y_0$ be two positive values such that $|x - x_0|, |y - y_0| \leq a_n$, and denote $\underline{x} = (x, x_0)$ and $\underline{y} = (y, y_0)$. We have

$$\begin{aligned}
\widehat{\mathbb{F}}(x,y) - \widehat{\mathbb{F}}(x_0,y_0) &= \int_0^{y_0} \left[\widehat{\overline{G}}^{-1}(v) - \overline{G}^{-1}(v)\right] \mathrm{d}\big[H(x,v) - H(x_0,v)\big] + \int_{y_0}^{y} \left[\widehat{\overline{G}}^{-1}(v) - \overline{G}^{-1}(v)\right] \mathrm{d}H(x,v) \\
&\quad + \int_0^{y_0} \widehat{\overline{G}}^{-1}(v)\,\mathrm{d}\big[\widehat{\mathbb{H}}(x,v) - \widehat{\mathbb{H}}(x_0,v)\big] + \int_{y_0}^{y} \widehat{\overline{G}}^{-1}(v)\,\mathrm{d}\widehat{\mathbb{H}}(x,v) \\
&= \int_0^{y_0} \frac{\widehat{\mathbb{G}}(v)}{\overline{G}^2(v)}\,\mathrm{d}\big[H(x,v) - H(x_0,v)\big] + \int_{y_0}^{y} \frac{\widehat{\mathbb{G}}(v)}{\overline{G}^2(v)}\,\mathrm{d}H(x,v) \\
&\quad + \int_0^{y_0} \widehat{\overline{G}}^{-1}(v)\,\mathrm{d}\big[\widehat{\mathbb{H}}(x,v) - \widehat{\mathbb{H}}(x_0,v)\big] + \int_{y_0}^{y} \widehat{\overline{G}}^{-1}(v)\,\mathrm{d}\widehat{\mathbb{H}}(x,v) + r_n'(\underline{x},\underline{y}),
\end{aligned} \tag{17}$$

where $\|r_n'\| = \mathcal{O}_{a.s.}\left(n^{-1}\log\log n\right)$, by using the uniform convergence of $\widehat{G}$. Let $I_n^1(\underline{x},\underline{y})$, $I_n^2(\underline{x},\underline{y})$ and $I_n^3(\underline{x},\underline{y})$ be, respectively, the sum of the first two terms, the third term and the fourth term in (17). We want to find the rates of the sup-norm of $I_n^k(\underline{x},\underline{y})$, for $k = 1,2,3$. First, we have

$$|I_n^1(\underline{x},\underline{y})| \leq \|\overline{G}^{-2}\|.\|\widehat{G} - G\|.\left(\int_0^{y_0} \left|\frac{\partial H}{\partial v}(x,v) - \frac{\partial H}{\partial v}(x_0,v)\right| \mathrm{d}v + |H(x,y) - H(x,y_0)|\right),$$

and by using Taylor expansion of first order for $|x - x_0|, |y - y_0| \le a_n$, under bounded first and second partial derivatives of $H$, and the uniform convergence of $\widehat{G}$,

$$\sup_{\substack{|x-x_0|\le a_n\\ |y-y_0|\le a_n}} |I_n^1(\underline{x}, \underline{y})| = \mathcal{O}_{a.s.}\left(a_n n^{-1/2}\big(\log\log n\big)^{1/2}\right).$$

For the rates of $I_n^2(\underline{x}, \underline{y})$ and $I_n^3(\underline{x}, \underline{y})$, notice that by using partial integration

$$|I_n^2(\underline{x}, \underline{y})| \le \|\widehat{\overline{G}}^{-1}\|.\|\widehat{\overline{G}}^{-1} - 1\|. \sup_{\substack{|x-x_0|\le a_n\\ v\le y_0}} \left|\widehat{\mathbb{H}}(x, v) - \widehat{\mathbb{H}}(x_0, v)\right|,$$

and $I_n^3(\underline{x}, \underline{y})$ can be written as

$$\begin{aligned} I_n^3(\underline{x}, \underline{y}) = &\left[\widehat{\mathbb{H}}(x, y) - \widehat{\mathbb{H}}(x_0, y_0)\right]\widehat{\overline{G}}^{-1}(y) + \left[\widehat{\mathbb{H}}(x_0, y_0) - \widehat{\mathbb{H}}(x, y_0)\right]\widehat{\overline{G}}^{-1}(y_0) \\ &+ \int_{y_0}^{y}\left[\widehat{\mathbb{H}}(x_0, y_0) - \widehat{\mathbb{H}}(x, v)\right]\mathrm{d}\widehat{\overline{G}}^{-1}(v). \end{aligned}$$

Hence, by using theorem 2.3 in (22)

$$\sup_{\substack{|x-x_0|\le a_n\\ |y-y_0|\le a_n}} |I_n^k(\underline{x}, \underline{y})| = \mathcal{O}_{a.s.}\left(n^{-3/4}\big(\log n\big)^{1/2}\big(\log\log n\big)^{1/4}\right),$$

for $k = 2, 3$, and the result follows. ■

**Proof of Theorem 2.**

Using the oscillation result in Proposition 1 and Taylor expansion, the representation of $\widehat{\mathbb{C}}_0(u, v)$ follows from the i.i.d. representations of $\widehat{F}(x, y)$, in Proposition 1, and that of $\widehat{F}_1^{-1}(x)$ and $\widehat{F}_2^{-1}(x)$ in (1) and (14), respectively. ■

**Proof of Theorem 3.**

The proof is given for representation (8) when $u, v \in \mathcal{A}_3 = [1 - h, 1]$. The proof is similar for the other cases of $u$ and $v$. Given $K_{u,h}$ is continuously differentiable, by partial integration

$$\begin{aligned} \widehat{\mathfrak{C}}(u, v) = h^{-2}\Big\{ &\int_{v-h}^{1}\int_{u-h}^{1} K_{u,h}^{(1)}\left(\frac{u-t}{h}\right)K_{v,h}^{(1)}\left(\frac{v-s}{h}\right)\frac{\mathrm{d}t}{h}\frac{\mathrm{d}s}{h} \\ &- \int_{v-h}^{1}\int_{u-h}^{1} \widehat{\mathbb{C}}(1, s)K_{u,h}^{(1)}\left(\frac{u-t}{h}\right)K_{v,h}^{(1)}\left(\frac{v-s}{h}\right)\frac{\mathrm{d}t}{h}\frac{\mathrm{d}s}{h} \\ &- \int_{v-h}^{1}\int_{u-h}^{1} \widehat{\mathbb{C}}(t, 1)K_{u,h}^{(1)}\left(\frac{u-t}{h}\right)K_{v,h}^{(1)}\left(\frac{v-s}{h}\right)\frac{\mathrm{d}t}{h}\frac{\mathrm{d}s}{h} \\ &+ \int_{v-h}^{1}\int_{u-h}^{1} \widehat{\mathbb{C}}(t, s)K_{u,h}^{(1)}\left(\frac{u-t}{h}\right)K_{v,h}^{(1)}\left(\frac{v-s}{h}\right)\frac{\mathrm{d}t}{h}\frac{\mathrm{d}s}{h}\Big\}, \end{aligned}$$

and by using the substitutions $t^* = (u-t)/h$ and $s^* = (v-s)/h$,

$$\widehat{\mathfrak{C}}(u,v) = h^{-2} \iint_{[-1,1]^2} \Big[\widehat{\mathbb{C}}(u-th, v-sh) - \widehat{\mathbb{C}}(u-th,1)\mathrm{I}(v-sh \le 1) - \widehat{\mathbb{C}}(1, v-sh)\mathrm{I}(u-th \le 1) + \mathrm{I}(u-th \le 1, v-sh \le 1)\Big] \mathrm{d}K_{u,h}(t)\mathrm{d}K_{v,h}(s).$$

The difference $\widehat{\mathfrak{C}} - \mathfrak{C}$ can be written as

$$\begin{aligned}\widehat{\mathfrak{C}}(u,v) - \mathfrak{C}(u,v) = h^{-2} \iint_{[-1,1]^2} \Big\{ &\Big[\widehat{\mathbb{C}}(u-th, v-sh) - \mathbb{C}(u-th, v-sh)\Big] \\ &- \Big[\widehat{\mathbb{C}}(u-th,1) - \mathbb{C}(u-th,1)\Big]\mathrm{I}(v-sh \le 1) \\ &- \Big[\widehat{\mathbb{C}}(1,v-sh) - \mathbb{C}(1,v-sh)\Big]\mathrm{I}(u-th \le 1)\Big\} \mathrm{d}K_{u,h}(t)\mathrm{d}K_{v,h}(s) \\ + \iint_{[-1,1]^2} &\Big[\mathfrak{C}(u-th, v-sh) - \mathfrak{C}(u,v)\Big] K_{u,h}(t)\, K_{v,h}(s)\, \mathrm{d}t\, \mathrm{d}s.\end{aligned}$$

By employing the i.i.d. representation of $\widehat{\mathbb{C}}$ in (7) and Taylor expansion of second order, the result follows by using the fact that $\int_{-1}^{1} u\, K_{u,h}(t)\, \mathrm{d}t = \int_{-1}^{1} s\, K_{v,h}(s)\, \mathrm{d}s = 0$. ■

**Proof of Theorem 4.**

1. First, notice that the estimator $\widehat{F}$ depends on the pair $(\widehat{F}_{X,Z,\delta}, \widehat{G})$ through the map

$$(f,g) \mapsto \int_{[0,x]\times[0,y]} \frac{1}{1-g}\, \mathrm{d}f.$$

Using similar arguments to that of (25) (Lemma 1) and (24) (Lemma 3.9.17), the composition map is Hadamard differentiable on a domain of the type $\{(f,g) : \int |df| \le M, g \ge \epsilon\}$, for $M, \epsilon > 0$, at every point $(f,g)$ such that $1/g$ is of bounded variation. Now, using the empirical central limit theorem and the fact that $\sqrt{n}\big[\widehat{G} - G\big]$ converges to a tight univariate gaussian process $\mathbb{G}$, the process $\sqrt{n}\big(\widehat{F}_{X,Z,\delta} - F_{X,Z,\delta}, \widehat{\overline{G}} - \overline{G}\big]$ converges to a tight zero-mean gaussian process $\big(\mathbb{F}_{X,Z,\delta}, \overline{\mathbb{G}}\big)$. Hence, by the delta method, the process $\sqrt{n}\big[\widehat{F} - F\big]$ converges to the tight process

$$\int_{[0,x]\times[0,y]} \frac{1}{\overline{G}}\, \mathrm{d}\mathbb{F}_{X,Z,\delta} - \int_{[0,x]\times[0,y]} \frac{\overline{\mathbb{G}}}{\overline{G}^2}\, \mathrm{d}F_{X,Z,\delta}.$$

By Lemma 3.9.28 in (24), the map $\varphi$ defined by $\varphi(F)(u,v) = \mathbb{C}(u,v)$ is Hadamard differentiable at $F$ tangentially to the set of continuous functions on $\overline{\mathbb{R}}^2$. Therefore, using similar arguments to that of (24) (example 3.9.29), $\sqrt{n}\big[\widehat{\mathbb{C}} - \mathbb{C}\big]$ converges in distribution to a tight process $\mathbb{C}_L$.

2. Now, note that by lemma 1 in (25) the map $\phi_1 : \mathbb{C} \to 4\int_{\mathcal{B}} \mathbb{C}\, \mathrm{d}\mathbb{C} - 1$ is Hadamard-differentiable at $\mathbb{C}$ tangentially to the set of continuous functions on $\mathcal{B}$, with derivative

$$\phi'_{1,\mathbb{C}}(\xi) = 4\left\{\int \mathbb{C}\, \mathrm{d}\xi + \int \xi\, \mathrm{d}\mathbb{C}\right\}.$$

Thus, by the functional delta method

$$\sqrt{n}\big[\widehat{\tau}_{X,Y} - \tau_{X,Y}\big] = 4\sqrt{n}\big[\phi_1(\widehat{\mathbb{C}}) - \phi_1(\mathbb{C})\big] \xrightarrow{d} \phi'_{1,\mathbb{C}}(\mathbb{C}_L).$$

■

**Proof of Theorem 5.**

Analogously to the proof of lemma 1 in (25), the map $\phi_2 : \mathbb{C} \to 12\int_{\mathcal{B}} u\, v\, \mathrm{d}\mathbb{C} - 3$ is Hadamard-differentiable at $\mathbb{C}$ tangentially to the set of continuous functions on $\mathcal{B}$, with derivative

$$\phi'_{2,\mathbb{C}}(\xi) = 12 \int u\, v\, \mathrm{d}\xi.$$

Thus, by the functional delta method

$$\sqrt{n}\big[\widehat{\rho}_{X,Y} - \rho_{X,Y}\big] = 12\sqrt{n}\big[\phi_2(\widehat{\mathbb{C}}) - \phi_2(\mathbb{C})\big] \xrightarrow{d} \phi'_{2,\mathbb{C}}(\mathbb{C}_L).$$

■

# References


[1] Bahadur, R.R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3), 577–580.

[2] Bouezmarni, T., El-Ghouch, A., and Taamouti, A. (2013). Bernstein estimator for unbounded copula densities. *Statistics and Risk Modeling* 30(4), 343–360.

[3] Bouezmarni, T., Lemyre, F.C., and El-Ghouch, A. (2019). Estimation of a bivariate conditional copula when a variable is subject to random right censoring. *Electronic J.Statistics* 13, 5044–5087.

[4] Chen, S.X. and Huang, T. (2007). Nonparametric estimation of copula functions for dependent modeling. *Canadian J. Statistics*, 35, 265–282.

[5] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Bullettin de l'académie Royal de Belgique, Classe des Sciences* 65, 274–29.

[6] Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. New York: Wiley.

[7] Frees, E. and Valdez, E. (1998). Annuity Valuation with Dependant Mortality. *North American Actuarial Journal* 2, 1–25.

[8] Genest, C., Masiello, E., and Tribouley, K. (2009). Estimating copula densities through wavelets. *Insurance: Mathematics and Economics* 44, 170–181.

[9] Gribkova, S. and Lopez, O. (2015). Non-parametric copula estimation under bivariate censoring. *Scandinavian Journal of Statistics* 42(4), 925–946.

[10] Kauermann, G., Schellhase, C., and Ruppert, D. (2013). Flexible Copula Density Estimation with Penalized Hierarchical B-splines. *Scandinavian J. statistics* 40(4), 685–705.

[11] Klugman, S.A. and Parsa, A. (1999). Fitting Bivariate Distributions with Copulas. *Insurance: Mathematics and Economics* 24, 139-148.

[12] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis* 14(4), 457–471.

[13] Lo, S., Mack, Y., and Wang, J. (1989). Density and hazard rate estimation for censored data via strong reptresentation of the kaplan-meier estimator. *Probability Theory and Relatd Fields* 80, 473–473.

[14] Lo, S. and Singh, K. (1985). The Product-Limit Estimator and the Bootstrap: Some Asymptotic Representations. *Probability Theory and Related Fields* 71, 455–465.

[15] Loprinzi, C.L., et al. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. *Journal of Clinical Oncology* 12, 601–607.

[16] Miller, R. and Halpern, J. (1982). Regression with censored Data. *Biometrika* 69(3), 521–531.

[17] Meng, X.-L., Bassiakos, Y. and Lo, S.-H. (1991). Large-sample properties for a general estimator of the treatment effect in the two-sample problem with right censoring. *The Annals of Statistics* 19(4), 1786–1812.

[18] Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* 78(3), 521–530.

[19] Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics* 37(5B), 3023–3058.

[20] Rabhi, Y. and Bouezmarni, T. (2020). Nonparametric inference for copulas and measures of dependence under length-biased sampling and informative censoring. *Journal of the American Statistical Association*, 115 (531), 1268–1278.

[21] Shih, J. and Louis, T. (1995). Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics* 51, 1384–1399.

[22] Stute, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *The Annals of Probability* 12(2), 361–379.

[23] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian J. Statistics* 23(4), 461–471.

[24] Van Der-Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer.

[25] Veraverbeke, N., Omelka, M. and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian J. Statistics* 38(4), 766–780.

[26] Wang, W. and Wells, M.T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association* 95(449), 62–72.

[27] Wen, K., and Wu, X. (2020). Transformation-kernel estimation of copula densities. *Journal of Business  Economic Statistics* 38 (1), 148–164.

[28] Hogg, R.V. and Klugman, S.A. (1984). Loss Distributions, New York: Wiley

[29] Frees, E.W. and Valdez, E.A. (1998). Understanding relationships using copulas. *North American Actuarial Journal* 2 (1), 1–25.

[30] Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 88, 1034–1043.

[31] McCall, B.P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometrica: Journal of the Econometric Society* 647-682.