Dependent Censoring Based On Parametric Copulas

Kaushik Raj V. Nadar

November 2024

MTH697 Project Report

Roll No.: 208160499

Under the Guidance of

Prof. Satya Prakash Singh

Department of Mathematics & Statistics,

Indian Institute of Technology Kanpur

Declaration

I hereby declare that the work presented in the project report entitled **"Dependent Censoring Based On Parametric Copulas"** is written by me in my own words and contains my own or borrowed ideas. At places, where ideas and words are borrowed from other sources, proper references and acknowledgements, as applicable, have been provided. To the best of my knowledge this work does not emanate from or resemble work created by person(s) other than those mentioned and acknowledged herein.

Name and Signature: Kaushik Raj Vaikuntaraman Nadar

Date: November 1, 2024

Abstract

We investigate the marginal distribution of a survival time T that is subject to random right censoring, where T is stochastically dependent on the censoring time C . This scenario frequently arises in practice; for instance, T may represent the time until a patient succumbs to a specific disease, while C could denote the time until the patient withdraws from the study or dies from an unrelated cause. If the reasons for study dropout are influenced by the patient's health status or if the patient dies from a condition sharing similar risk factors with the primary disease, it is likely that T and C exhibit dependence. To address this dependence, we replicate and validate a novel model[2] that utilizes a parametric copula to describe the relationship between T and C , complemented by parametric marginal distributions for both variables. Importantly, this model distinguishes itself from many existing studies by not requiring prior knowledge of the copula parameter. We establish sufficient conditions for identifying the bivariate distribution of (T, C) based on these parametric copulas and marginals, which we validate across a diverse range of common copulas and marginal distributions. Furthermore, we investigate the estimation of the model and conduct extensive simulations and analyses to evaluate its performance.

Contents

1 Introduction

1.1 Censoring in Survival Analysis

Censoring is a key challenge in survival analysis, where it refers to incomplete information about the time of an event (often failure or death). It occurs when the exact event time is unknown for certain study subjects due to various reasons, such as subjects dropping out, the study ending before all events have occurred, or subjects surviving beyond the study period.
Censoring complicates survival analysis because it biases the data, potentially leading to incorrect estimates of survival times if not properly accounted for. Standard statistical methods often assume complete data.

Understanding and addressing censoring is crucial for drawing accurate conclusions, especially in fields like medicine, engineering reliability, and social sciences where survival analysis is heavily applied.

1.1.1 Survival Time (T)

The duration of time from a defined starting point (like the beginning of a study, diagnosis, or a treatment) until the occurence of a particular event of interest (like death, relapse, failure of a device, etc.).

1.1.2 Censoring Time (C)

Censoring time is the point in time at which the observation of a subject ends, without the event of interest having occurred. For example, if the study ends before the subject shows the symptoms of a certain disease, if there is a loss of follow-up with the subject, or if the subject voluntarily drops-out of the

study.

1.2 Types of Censoring

1.2.1 Right Censoring

This is the most common type. Here, we only know that the event time is greater than some observed time. For example, if a patient is still alive at the study's end, we know only that their survival time exceeds the observed time.

Types of Right Censoring:

- **Type I censoring:** The study is set to end at a pre-defined time. If an individual has not experienced the event by this end time, their data is right-censored.
- **Type II censoring:** The study ends when a certain number of events have occurred. Those who have not experienced the event by then are right-censored.
- **Random censoring:** Censoring happens randomly due to factors like dropouts or lost-to-follow-up cases, unrelated to the study design.

1.2.2 Left Censoring

The event of interest has already occurred before the observation period begins. For instance, if we're studying the onset of a disease but only start observing individuals after they're already sick, we don't know the exact time they first contracted it.

1.2.3 Interval Censoring

The exact event time is unknown, but we know it occurred within a certain time interval. This is common in periodic follow-ups where we know the event happened between two check-ins but not the exact moment.

1.3 Existing Solutions

The existing solutions for this problem has several challenges and limitations in modeling the joint distribution of T (survival time) and C(censoring time) under right-censoring in survival analysis. Some of the issues are: Non-Identifiability in a Non-Parametric Setting[6], Dependence on Known Copulas with Fixed Association Parameters[8], Challenges with Estimating the Copula Parameter[4] and Strict Assumptions Required for Identifiability of Archimedean Copulas[3].

1.4 Objective

We demonstrate that assuming a fully specified copula is not essential for identifying the joint distribution of T and C. By modeling the marginal distributions of T and C, along with the copula function, parametrically, we prove that under certain conditions, the joint model becomes identifiable. Notably, this approach ensures the identifiability of the copula's association parameter, marking a significant advancement in employing copulas within survival analysis.

2 Proposed Model

Let T denote the survival time and C the censoring time. Due to random right censoring, we observe $Y = \min(T, C)$ and an indicator $\Delta = I(T \leq C)$. We assume potential dependence between T and C , which we model using a copula. Throughout this framework[2], we take both T and C to be non-negative, with continuous marginal distributions F_T and F_C belonging to parametric families, specified as:

$$F_T \in \{F_{T,\theta_T} : \theta_T \in \Theta_T\}, \quad F_C \in \{F_{C,\theta_C} : \theta_C \in \Theta_C\}, \tag{1}$$

where Θ_T and Θ_C denote the parameter spaces. We denote the densities of T and C as f_T and f_C , or f_{T,θ_T} and f_{C,θ_C} in the parametric context. The joint distribution $F_{T,C}$ of (T, C) is then modeled through a copula-based approach. A copula $C : [0, 1] \times [0, 1] \to [0, 1]$ with uniform margins enables us to represent the dependence structure, as established by Sklar (1959)[5]. Specifically, given the continuity of F_T and F_C , there exists a unique copula C satisfying:

$$F_{T,C}(t, c) = C(F_T(t), F_C(c)), \tag{2}$$

for any $t, c \geq 0$. We further assume that the copula itself is parametrically modeled, such that:

$$C \in \{C_\theta : \theta \in \Theta\}, \tag{3}$$

for some parameter space Θ.

Joint and Conditional Densities

Differentiating $F_{T,C}(t,c) = C(F_T(t), F_C(c))$ yields the joint density of (T, C) as:

$$f_{T,C}(t,c) = c(F_T(t), F_C(c)) f_T(t) f_C(c),$$

where c represents the copula density. Our objective is to determine the conditional distributions of T given C and C given T.

From the joint density expression, we obtain the conditional densities:

$f_{T|C}(t|c) = c(F_T(t), F_C(c)) f_T(t)$,

$f_{C|T}(c|t) = c(F_T(t), F_C(c)) f_C(c)$.

The conditional distribution function of T given C = c can be derived as follows:

$$\begin{aligned} F_{T|C}(t \mid c) &= \int_0^t c\{F_T(t^*), F_C(c)\} f_T(t^*)\, dt^* \\ &= \int_0^t \frac{\partial^2}{\partial u \partial v} C(u,v)\Big|_{u=F_T(t^*), v=F_C(c)} \frac{dF_T(t)}{dt}\Big|_{t=t^*} dt^* \\ &= \frac{\partial}{\partial v} C(u,v)\Big|_{u=F_T(t), v=F_C(c)} = h_{T|C}\{F_T(t) \mid F_C(c)\}. \end{aligned}$$

where $h_{T|C}$ is defined based on the partial derivative of the copula with respect to u at $u = F_T(t)$ and $v = F_C(c)$. Similarly, we have:

$$F_{C|T}(c|t) = h_{C|T}\{F_C(c)|F_T(t)\}.$$

Marginal Distribution of $Y = \min(T, C)$

To derive the marginal distribution of Y, we note:

$$F_Y(y) = 1 - \Pr(Y > y) = 1 - \Pr(T > y, C > y).$$

Expanding this, we get:

$$F_Y(y) = F_C(y) + F_T(y) - C\{F_T(y), F_C(y)\}.$$

Joint Mixed Density of (Y, Δ)

Finally, we derive the joint density $f_{Y,\Delta}$ by observing:

$$\begin{aligned} F_{Y,\Delta}(y, 1) = \text{pr}(T \leq y, T \leq C) &= \int_0^y \text{pr}(C \geq t \mid T = t) f_T(t)\, dt \\ &= \int_0^y \left[1 - h_{C|T}\{F_C(t) \mid F_T(t)\}\right] f_T(t)\, dt, \end{aligned}$$

This results in:

$$f_{Y,\Delta}(y, 1) = \left[1 - h_{C|T}\{F_C(y)|F_T(y)\}\right] f_T(y).$$

Similarly, we have:

$$f_{Y,\Delta}(y, 0) = \left[1 - h_{T|C}\{F_T(y)|F_C(y)\}\right] f_C(y).$$

This provides the foundation for modeling dependent censoring with parametric assumptions on both the marginals and the copula structure.

3 Identifiability

We begin by establishing sufficient conditions for model identifiability. "Identifiability" means that the parameters $(\theta, \theta_T, \theta_C) \in \Theta \times \Theta_T \times \Theta_C$ uniquely determine the density of the observable variables (Y, Δ). Specifically, if $f_{Y,\Delta;\alpha_1} \equiv f_{Y,\Delta;\alpha_2}$, then $\alpha_1 = \alpha_2$, where $\alpha_j = (\theta_j, \theta_{Tj}, \theta_{Cj})^T$ for $j = 1, 2$.

Theorem 1

Suppose the following conditions hold:

1. For any $\theta_{T1}, \theta_{T2} \in \Theta_T$ and $\theta_{C1}, \theta_{C2} \in \Theta_C$, the four equivalences below must be satisfied:

$$\lim_{t\to 0} \frac{f_{T,\theta_{T1}}(t)}{f_{T,\theta_{T2}}(t)} = 1 \Rightarrow \theta_{T1} = \theta_{T2}, \quad \lim_{t\to 0} \frac{f_{C,\theta_{C1}}(t)}{f_{C,\theta_{C2}}(t)} = 1 \Rightarrow \theta_{C1} = \theta_{C2},$$

$$\lim_{t\to \infty} \frac{f_{T,\theta_{T1}}(t)}{f_{T,\theta_{T2}}(t)} = 1 \Rightarrow \theta_{T1} = \theta_{T2}, \quad \lim_{t\to \infty} \frac{f_{C,\theta_{C1}}(t)}{f_{C,\theta_{C2}}(t)} = 1 \Rightarrow \theta_{C1} = \theta_{C2}.$$

2. The parameter space $\Theta \times \Theta_T \times \Theta_C$ satisfies:

$$\lim_{t\to 0} h_{T|C,\theta}(u_t|v_t) = 0 \quad or \quad \lim_{t\to \infty} h_{T|C,\theta}(u_t|v_t) = 0, \quad \forall (\theta, \theta_T, \theta_C) \in \Theta \times \Theta_T \times \Theta_C,$$

and similarly for $h_{C|T,\theta}(v_t|u_t)$, where $u_t = F_{T,\theta_T}(t)$ and $v_t = F_{C,\theta_C}(t)$.

Under these conditions, the model specified in (1)–(3) is identifiable.

Theorem 1's first condition (i) pertains solely to the marginal distributions, whereas condition (ii) involves both the margins and the copula family. Condition (i) holds for a broad class of parametric families for the marginal densities f_T and f_C, as demonstrated in the next theorem. For other families

not explicitly covered here, condition (i) can also be verified, but we focus on the primary parametric families commonly used in survival analysis. Condition (i) requires the equivalence to hold for limits as t approaches both 0 and ∞, ensuring that when either limit is equal to 1, the parameter vectors must be identical.

Proof

We know that

$$\begin{aligned} f_{Y,\Delta}(t,1) &= \{1 - F_{C|T}(t|t)\} f_T(t) \\ &= [1 - h_{C|T}\{F_C(t)|F_T(t)\}] f_T(t). \end{aligned}$$

From condition (ii), we know $\lim_{t\to a} h_{C|T}\{F_C(t)|F_T(t)\} = 0$ for $a = 0$ or $a = \infty$.

Hence, $\lim_{t\to a} f_{Y,\Delta}(t,1) = \lim_{t\to a} f_T(t)$

Suppose now that $f_{Y,Delta,\alpha_1}(t,1) = f_{Y,Delta,\alpha_2}(t,1) \forall t$

$$\implies 1 = \lim_{t\to a} \frac{f_{Y,\Delta,\alpha_1}(t,1)}{f_{Y,\Delta,\alpha_2}(t,1)} = \lim_{t\to a} \frac{f_{T,\theta_{T_1}}(t)}{f_{T,\theta_{T_2}}(t)}$$

Using condition (i), we get, $\theta_{T_1} = \theta_{T_2}$.

Similarly, we can show for $\theta_{C_1} = \theta_{C_2}$.

Finally, to show that $\theta_1 = \theta_2$, notice

$F_{Y,\alpha_j}(t) = F_{T,\theta_{T_1}} + F_{C,\theta_{C_1}} - C_\theta\{F_{T,\theta_{T_1}}(t), F_{C,\theta_{C_1}}(t)\}$ and $F_{Y,\alpha_1}(t) = F_{Y,\alpha_2}(t)\ \forall t$.

$\therefore$ The copula is unique $\implies \theta_1 = \theta_2$.

Theorem 2

Condition (i) of Theorem 1 is satisfied for the families of lognormal, log-Student-t, Weibull, log-logistic, exponential, gamma, and truncated normal densities.

Proof

Lognormal Marginals

First, consider the lognormal density for T, which depends on $\theta_T = (\mu, \sigma)$:

$$\lim_{t\to 0} \frac{f_{T,\mu_1,\sigma_1}(t)}{f_{T,\mu_2,\sigma_2}(t)} = \lim_{t\to 0} \frac{\frac{1}{t\sigma_1}\phi\left(\frac{\log t-\mu_1}{\sigma_1}\right)}{\frac{1}{t\sigma_2}\phi\left(\frac{\log t-\mu_2}{\sigma_2}\right)} = \lim_{t'\to -\infty} \frac{\phi\left(\frac{t'-\mu_1}{\sigma_1}\right)}{\phi\left(\frac{t'-\mu_2}{\sigma_2}\right)}.$$

It is straightforward to verify that this limit is equal to 1 only when $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$. The same result holds for the limit as $t \to \infty$.

Log-Student-t Marginals

Similarly, for the log-Student-t density, we have

$$\lim_{t\to 0,\infty} \frac{f_{T,\nu_1,\mu_1,\sigma_1}(t)}{f_{T,\nu_2,\mu_2,\sigma_2}(t)} = \frac{c_{\nu_1,\sigma_1}}{c_{\nu_2,\sigma_2}} \lim_{t\to 0,\infty} \frac{\left(1+\frac{1}{\nu_1}\left(\frac{\log t-\mu_1}{\sigma_1}\right)^2\right)^{-(\nu_1+1)/2}}{\left(1+\frac{1}{\nu_2}\left(\frac{\log t-\mu_2}{\sigma_2}\right)^2\right)^{-(\nu_2+1)/2}},$$

where $c_{\nu,\sigma} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}$, with Γ denoting the gamma function. It is easy to see that this limit is equal to 1 if and only if $(\nu_1, \mu_1, \sigma_1) = (\nu_2, \mu_2, \sigma_2)$.

Weibull Marginals

The same conclusion holds for the Weibull density, where

$$\lim_{t\to 0,\infty} \frac{f_{T,\lambda_1,\rho_1}(t)}{f_{T,\lambda_2,\rho_2}(t)} = \frac{\lambda_1\rho_1}{\lambda_2\rho_2} \lim_{t\to 0,\infty} \frac{t^{\rho_1-1}\exp(-\lambda_1 t^{\rho_1})}{t^{\rho_2-1}\exp(-\lambda_2 t^{\rho_2})},$$

which is equal to 1 only if $\rho_1 = \rho_2$ and $\lambda_1 = \lambda_2$.

Log-logistic Marginals

For the log-logistic density, we have the limit

$$\lim_{t\to 0,\infty} \frac{f_{T,\lambda_1,\kappa_1}(t)}{f_{T,\lambda_2,\kappa_2}(t)} = \frac{\kappa_1\lambda_1^{\kappa_1}}{\kappa_2\lambda_2^{\kappa_2}} \lim_{t\to 0,\infty} \frac{t^{\kappa_1-1}\left(1+(\lambda_2 t)^{\kappa_2}\right)^2}{t^{\kappa_2-1}\left(1+(\lambda_1 t)^{\kappa_1}\right)^2}.$$

Again, this limit equals 1 if and only if $\kappa_1 = \kappa_2$ and $\lambda_1 = \lambda_2$.

Exponential Marginals

First, we consider the exponential density for T , which depends on the parameter $\theta_T = \lambda$:

$$f_T(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

The limit of the ratio of the exponential densities for two parameters λ_1 and λ_2 as $t \to 0$ and $t \to \infty$ is given by:

$$\lim_{t\to 0} \frac{f_{T,\lambda_1}(t)}{f_{T,\lambda_2}(t)} = \lim_{t\to 0} \frac{\lambda_1 e^{-\lambda_1 t}}{\lambda_2 e^{-\lambda_2 t}} = \frac{\lambda_1}{\lambda_2}$$

$$= 1 \iff \lambda_1 = \lambda_2$$

Similarly, as $t \to \infty$:

$$\lim_{t\to \infty} \frac{f_{T,\lambda_1}(t)}{f_{T,\lambda_2}(t)} = \lim_{t\to \infty} \frac{\lambda_1 e^{-\lambda_1 t}}{\lambda_2 e^{-\lambda_2 t}} = 0 \text{ or } \infty$$

depending on λ_1 and λ_2.

These limits indicate that for the exponential density, the ratio is equal to 1 only if $\lambda_1 = \lambda_2$.

Similarly, we can prove that ratio is 1 iff parameters are equal in the C case.

Gamma Marginals

First, we consider the Gamma density for T , which depends on the parameters $\theta_T = (\alpha, \beta)$:

$$f_T(t; \alpha, \beta) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}, \quad t \geq 0,$$

where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the rate parameter, and $\Gamma(\alpha)$ is the gamma function.

The limit of the ratio of the Gamma densities for two sets of parameters (α_1, β_1) and (α_2, β_2) as $t \to 0$ and $t \to \infty$ is given by:

$$\lim_{t\to 0} \frac{f_{T,\alpha_1,\beta_1}(t)}{f_{T,\alpha_2,\beta_2}(t)} = \lim_{t\to 0} \frac{\frac{\beta_1^{\alpha_1} t^{\alpha_1-1} e^{-\beta_1 t}}{\Gamma(\alpha_1)}}{\frac{\beta_2^{\alpha_2} t^{\alpha_2-1} e^{-\beta_2 t}}{\Gamma(\alpha_2)}} = \frac{\beta_1^{\alpha_1}\Gamma(\alpha_2)}{\beta_2^{\alpha_2}\Gamma(\alpha_1)} \lim_{t\to 0} \frac{t^{\alpha_1-1}}{t^{\alpha_2-1}}.$$

As $t \to 0$:

- If $\alpha_1 = 1$ and $\alpha_2 > 1$, the limit tends to 0.
- If $\alpha_1 > 1$ and $\alpha_2 = 1$, the limit tends to ∞.
- If $\alpha_1 = \alpha_2$, the limit approaches $\frac{\beta_1^{\alpha_1}\Gamma(\alpha_2)}{\beta_2^{\alpha_2}\Gamma(\alpha_1)}$.

Next, as $t \to \infty$:

$$\lim_{t\to \infty} \frac{f_{T,\alpha_1,\beta_1}(t)}{f_{T,\alpha_2,\beta_2}(t)} = \lim_{t\to \infty} \frac{\frac{\beta_1^{\alpha_1} t^{\alpha_1-1} e^{-\beta_1 t}}{\Gamma(\alpha_1)}}{\frac{\beta_2^{\alpha_2} t^{\alpha_2-1} e^{-\beta_2 t}}{\Gamma(\alpha_2)}} = 0.$$

Thus, we find that these limits equal to 1 only if $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. We will reach the same conclusion with the C marginal case.

Truncated Normal Marginals

Consider the truncated normal density for T , which depends on the parameters $\theta_T = (\mu, \sigma, a, b)$:

$$f_T(t; \mu, \sigma, a, b) = \frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{\sigma\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right]}, \quad a < t < b,$$

where ϕ is the probability density function (pdf) of the standard normal distribution, and Φ is the cumulative distribution function (cdf) of the standard normal distribution.

Since T and C are measurements of time, their support $= [0, \infty)$. Hence, we can consider a=0 and b=∞ in the truncated normal marginals. Thus, the density depends only on μ and σ in this case.

$$f_T(t; \mu, \sigma) = \frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{\sigma\left[1 - \Phi\left(\frac{-\mu}{\sigma}\right)\right]}, \quad 0 < t < \infty,$$

The limit of the ratio of the Truncated-normal densities for the two sets of parameters (μ_1, σ_1) and (μ_2, σ_2) as $t \to 0$ and $t \to \infty$ is given by:

$$\begin{aligned}
\lim_{t\to 0} \frac{f_{T,\mu_1,\sigma_1}(t)}{f_{T,\mu_2,\sigma_2}(t)} &= \lim_{t\to 0} \frac{\sigma_2}{\sigma_1} \frac{\Phi(\frac{\mu_1}{\sigma_1})}{\Phi(\frac{\mu_2}{\sigma_2})} \frac{\phi(\frac{t-\mu_1}{\sigma_1})}{\phi(\frac{t-\mu_2}{\sigma_2})} \\
&= \frac{\sigma_2}{\sigma_1} \frac{\Phi(\frac{\mu_1}{\sigma_1})}{\Phi(\frac{\mu_2}{\sigma_2})} \exp\left(\frac{-1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2}\right)\right) \\
&= 1 \iff \frac{\mu_1}{\sigma_1} = \frac{\mu_2}{\sigma_2} \text{ and } \sigma_1 = \sigma_2 \\
&\iff \mu_1 = \mu_2
\end{aligned}$$

Similarly, we can prove that ratio is 1 if parameters are equal in the C case. Therefore, condition (i) is satisfied for each of these densities.

Classes of Copula

Archimedean Copula

$$C(u, v) = \psi^{[-1]}\{\psi(u) + \psi(v)\} \tag{4}$$

where ψ is a generator, that is, $\psi : [0, 1] \to [0, \infty)$ is a continuous, strictly decreasing and convex function such that $\psi(1) = 0$. Here, $\psi^{[-1]}$ is the pseudo-inverse of ψ, i.e., $\psi^{[-1]}(t) = \psi^{-1}(t)$ if $0 \leq t \leq 1$ and $\psi^{[-1]}(t) = 0$ if $t \geq \psi(0)$. Following are some important families of archimedean copula:

- Frank Family: $\psi_\theta(u) = -\log\left[\frac{e^{-\theta u}-1}{e^{-\theta}-1}\right]$, $\theta \in \mathbb{R}\backslash\{0\}$
- Clayton Family: $\psi_\theta(u) = \frac{u^{-\theta}-1}{\theta}$ with $\theta \in [-1, \infty)\backslash\{0\}$
- Gumbel Family: $\psi_\theta(u) = \{-\log(u)\}^\theta$ with $\theta \in [1, \infty)$

Differentiation of (4) gives

$$h_{T|C}(u|v) = \frac{\psi'(v)}{\psi'\left[\psi^{-1}\{\psi(u) + \psi(v)\}\right]} \quad \text{and} \quad h_{C|T}(v|u) = \frac{\psi'(u)}{\psi'\left[\psi^{-1}\{\psi(u) + \psi(v)\}\right]}$$

for $0 < u, v < 1$, provided the derivatives and inverses in the formula exist.

Lemma 1

Suppose the generator ψ is differentiable on $(0, 1)$. If $\lim_{v\to 1} \psi'(v) \in (-\infty, 0)$, then $\lim_{t\to\infty} h_{T|C,\theta}\{F_{T,\theta T}(t)|F_{C,\theta C}(t)\} = 1$.

Proof of Lemma 1:

Let $u_t = F_{T,\theta_T}(t)$ and $v_t = F_{C,\theta_C}(t)$. Note that

$$\psi(1) = 0, \lim_{t\to 0} \psi^{-1}(t) = 1 \text{ and } \lim_{u\to 1} \psi'(u) = c \in (-\infty, 0)$$
$$\implies \lim_{t\to\infty} h_{T|C,\theta}(u_t|v_t) = \lim_{t\to\infty} \frac{\psi'(v_t)}{\psi'\left[\psi^{-1}\{\psi(u_t) + \psi(v_t)\}\right]} = \frac{c}{c} = 1$$

Gaussian Copula

$$C_\theta(u, v) = \Phi_\theta\{\Phi^{-1}(u), \Phi^{-1}(v)\}$$

where,

Φ: CDF of Standard Normal distribution

Φ_θ: CDF of Bivariate Standard Normal distribution with correlation θ

Theorem 3

Condition (ii) of Theorem 1 is satisfied by the following cases:

1. The Frank copula, regardless of the marginal distributions and the parameter space.

2. The Gumbel copula, if

$$\lim_{t\to 0} \frac{\log F_{T,\theta_T}(t)}{\log F_{C,\theta_C}(t)} \in (0, \infty) \quad \text{for all} \quad (\theta_T, \theta_C) \in \Theta_T \times \Theta_C.$$

3. The Gaussian copula, if either:

$$\lim_{t\to 0} A_{\theta,F_{T,\theta_T},F_{C,\theta_C}}(t) = -\infty \quad \forall(\theta, \theta_T, \theta_C) \in \Theta \times \Theta_T \times \Theta_C,$$

or

$$\lim_{t\to\infty} A_{\theta,F_{T,\theta_T},F_{C,\theta_C}}(t) = -\infty \quad \forall(\theta,\theta_T,\theta_C) \in \Theta \times \Theta_T \times \Theta_C.$$

The same conditions apply for $A_{\theta,F_{C,\theta_C},F_{T,\theta_T}}$, where
$A_{\theta,F_1,F_2}(t) = \Phi^{-1}\{F_1(t)\} - \theta\Phi^{-1}\{F_2(t)\}$.

Proof

We begin by examining the Frank copula. Simple calculations reveal that

$$\lim_{u\to 1} \psi'(u) = \frac{\theta e^{-\theta}}{e^{-\theta}-1} < 0 \quad \text{for} \quad \theta \neq 0.$$

Therefore, by applying Lemma 1, we conclude that

$$\lim_{t\to\infty} h_{T|C}\{F_T(t)|F_C(t)\} = 1.$$

This indicates that the first part of condition (ii) can only be fulfilled if

$$\lim_{t\to 0} h_{T|C}\{F_T(t)|F_C(t)\} = 0.$$

Some calculations yield:

$$\begin{aligned}\lim_{t\to 0} h_{T|C}\{F_T(t)|F_C(t)\} &= \lim_{t\to 0} \frac{\psi'(F_C(t))}{\{\psi^{-1}\{\psi(F_T(t)) + \psi(F_C(t))\}\}} \\ &= \frac{\exp\{-\theta F_C(t)\}\left[\exp\{-\theta F_T(t)\}-1\right]}{\left[\exp\{-\theta F_C(t)\}-1\right]\cdot\left[\exp\{-\theta F_T(t)\}-1\right]+e^{-\theta}-1} = 0\end{aligned}$$

for $\quad \theta \neq 0.$

Thus, the first part of condition (ii) is satisfied. A similar analysis can demonstrate that the second part is also satisfied.

For the Gumbel family, we observe that

$$\lim_{u\to 1}\psi'(u) = 0,$$

indicating that Lemma 1 does not apply. Therefore, we analyze the limits as t approaches 0 and ∞:

$$\begin{aligned}\lim_{t\to 0,\infty} h_{T|C}\{F_T(t)|F_C(t)\} = &\lim_{t\to 0,\infty}\left[1+(-\log F_C(t))^{-\theta}(-\log F_T(t))^{\theta}\right]^{-1+\frac{1}{\theta}} \\ &\times \lim_{t\to 0,\infty}\exp\left\{-\left[(-\log F_T(t))^{\theta}+(-\log F_C(t))^{\theta}\right]^{\frac{1}{\theta}}-\log F_C(t)\right\}.\end{aligned}$$

Refer to Aas et al.(2009)[1] for the formula of $h_{T|C}\{F_T(t)|F_C(t)\}$ for the Gumbel family. Assuming that $\frac{\log F_T(t)}{\log F_C(t)} \to c$ for some $0 < c < \infty$ as $t \to 0$, the exponential term tends to 0 as t approaches 0 and tends to 1 as t approaches infinity. The factor in front of this exponential term converges to a constant within the interval $[0, 1]$, depending on the limit of $\frac{\log F_T(t)}{\log F_C(t)}$ as t tends to 0 or infinity. This indicates that the product of the two limits equals 0 as t approaches 0, while the limit can be zero or strictly positive as t approaches infinity, depending on the limit of $\frac{\log F_T(t)}{\log F_C(t)}$ as t increases without bound.

Now, we turn to the Gaussian copula. Note that

$$\begin{aligned}Pr\left[\Phi^{-1}\{F_T(T)\} \le t, \Phi^{-1}\{F_C(C)\} \le c\right] &= \Pr\left(T \le F_T^{-1}(\Phi(t)), C \le F_C^{-1}(\Phi(c))\right) \\ &= \Phi_\theta(t, c)\end{aligned}$$

, where Φ_θ is the bivariate normal distribution with correlation parameter θ. Consequently,

$$\Phi^{-1}\{F_T(T)\}|\Phi^{-1}\{F_C(C)\} \sim N\left(\theta\Phi^{-1}\{F_C(C)\}, 1-\theta^2\right).$$

This allows us to write, omitting the parameters θ, θ_T, and θ_C for simplicity,

$$\begin{aligned} h_{T|C}\{F_T(t)|F_C(t)\} = F_{T|C}(t|t) &= Pr\left[\Phi^{-1}\{F_T(T)\} \leq \Phi^{-1}\{F_T(t)\}|\Phi^{-1}\{F_C(C)\} = \Phi^{-1}\{F_C(t)\}\right] \\ &= \Phi\left[\frac{\Phi^{-1}\{F_T(t)\} - \theta\Phi^{-1}\{F_C(t)\}}{\sqrt{1-\theta^2}}\right] = \frac{A_{\theta,F_T,F_C}(t)}{\sqrt{1-\theta^2}}, \end{aligned}$$

where

$$A_{\theta,F_T,F_C}(t) = \Phi^{-1}\{F_T(t)\} - \theta\Phi^{-1}\{F_C(t)\}.$$

Since $A_{\theta,F_T,F_C}(t) \to -\infty$ as $t \to 0$ or $t \to \infty$, it follows that $h_{T|C}\{F_T(t)|F_C(t)\} \to 0$ as t approaches either 0 or infinity. This verifies the first part of condition (ii). A similar approach can be used to prove the second part.

Theorem 4

Assume that condition (i) of Theorem 1 is met, and that the parameter spaces $\Theta_T \times \Theta_C$ satisfy $\lim_{t\to 0} \frac{F_{T,\theta_T}(t)}{F_{C,\theta_C}(t)}$ is either 0 or $+\infty$ for all $\theta_T \in \Theta_T$ and $\theta_C \in \Theta_C$. Additionally, suppose the copula C_θ is a Clayton copula with $\theta > 0$. Then, the model defined in equations (1)–(3) is identifiable.

Proof

Assume that $\lim_{t\to 0} \frac{F_{T,\theta_T}(t)}{F_{C,\theta_C}(t)} = \infty$; a similar approach can be taken if the limit equals zero. Then, from equation (6), we see that $\lim_{t\to 0} h_{C|T,\theta}\{F_{C,\theta_C}(t)|F_{T,\theta_T}(t)\} =$

0. Following arguments similar to those in the proof of Theorem 1, condition (i) implies that θ_T is identifiable. Using the form of $f_{Y,\theta}(\cdot, 1)$ in equation (4), it also follows that the function $t \to h_{C|T,\theta}\{F_{C,\theta_C}(t)|F_{T,\theta_T}(t)\}$ is identifiable.

Now, for sufficiently large t, $F_T(t)^\theta / F_{C,\theta_C}(t)^\theta - F_T(t)^\theta$ is close to zero (where θ_T has been omitted as it is identifiable). Thus, we can apply a Taylor expansion to approximate:

$$\begin{aligned}\log h_{C|T,\theta}\{F_{C,\theta_C}(t)|F_T(t)\} &= -\frac{\theta+1}{\theta}\log\left(1 + \frac{F_T(t)^\theta}{F_{C,\theta_C}(t)^\theta} - F_T(t)^\theta\right) \\ &\approx -\frac{\theta+1}{\theta}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\left(F_T(t)^\theta \cdot \{F_{C,\theta_C}(t)^\theta - 1\}\right)^k\end{aligned}$$

for large t, which yields a polynomial in $u = F_T(t)$. Therefore, for two parameter sets (θ, θ_C) and (θ^*, θ_C^*), we have:

$$\frac{\theta+1}{\theta}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\left(F_{C,\theta_C}(t)^{-\theta} - 1\right)^k u_t^{\theta k} = \frac{\theta^*+1}{\theta^*}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\left(F_{C,\theta_C^*}(t)^{-\theta^*} - 1\right)^k u_t^{\theta^* k}$$

for large t. This equality holds only if $\theta = \theta^*$ and $\theta_C = \theta_C^*$.

3.1 Examples of some models

From Theorem 3, it is clear that we can create an identifiable model with the Frank Copula and any of the marginal distribution satisfying condition (i) of Theorem 1.

Let's look at Gumbel Copula. The condition $\lim_{t\to 0}\frac{F_{T,\theta_T}(t)}{F_{C,\theta_C}(t)} = 0$ or $+\infty$ is often satisfied in parametric families. For instance, in the lognormal family, where $\log T \sim N(\mu_T, \sigma_T^2)$ and $\log C \sim N(\mu_C, \sigma_C^2)$, we have:

$$\lim_{t\to 0}\frac{F_T(t)}{F_C(t)} = \lim_{t\to 0}\exp\left(-\frac{(\log t-\mu_T)^2}{2\sigma_T^2}+\frac{(\log t-\mu_C)^2}{2\sigma_C^2}\right).$$

This limit approaches zero if $\sigma_T < \sigma_C$ or if $\sigma_T = \sigma_C$ and $\mu_T > \mu_C$. Conversely, it approaches infinity if $\sigma_T > \sigma_C$ or if $\sigma_T = \sigma_C$ and $\mu_T < \mu_C$. Thus, local identifiability is achieved near (μ_T, σ_T) and (μ_C, σ_C).

For the Clayton copula, if $\mu_T = \mu_C$ and $\sigma_T = \sigma_C$, we find that $h_{T|C}\{F_T(t)|F_C(t)\} = (2 - F_C(t)^\theta)^{-(\theta+1)/\theta}$, yielding $\lim_{t\to 0} h_{T|C}\{F_T(t)|F_C(t)\} = 2^{-(\theta+1)/\theta}$.

For the lognormal density, $\lim_{t\to 0}\frac{f_{C,\mu_{C1},\sigma_{C1}}(t)}{f_{C,\mu_{C2},\sigma_{C2}}(t)}$ can only be 0, 1, or ∞, ensuring that $(\mu_{C1}, \sigma_{C1}) = (\mu_{C2}, \sigma_{C2})$, confirming model identifiability.

Similarly, if $T \sim \text{Wei}(\lambda_T, \rho_T)$ and $C \sim \text{Wei}(\lambda_C, \rho_C)$, then $\lim_{t\to 0}\frac{F_T(t)}{F_C(t)} = 0$ if $\rho_T > \rho_C$, ∞ if $\rho_T < \rho_C$, and $\frac{\lambda_T}{\lambda_C}$ if $\rho_T = \rho_C$. Thus, satisfying the conditions for Theorem 4.

4 Estimation

In this section, we discuss parameter estimation for the joint parametric model of the survival time T and censoring time C, as specified in equations (1)–(3). We assume an independent and identically distributed sample $D = \{(y_i, \delta_i), i = 1, \ldots, n\}$ is available. The joint log-likelihood function for the parameter vector $\alpha = (\theta, \theta_T, \theta_C)^T$ is given by

$$\ell(\alpha; D) = \sum_{i=1}^{n} \left[\delta_i \log f_{T,\theta_T}(y_i)\big(1 - h_{C|T,\theta}\{F_{C,\theta_C}(y_i)|F_{T,\theta_T}(y_i)\}\big)\right] + \sum_{i=1}^{n} \left[(1-\delta_i) \log f_{C,\theta_C}(y_i)\big(1 - h_{T|C,\theta}\{F_{T,\theta_T}(y_i)|F_{C,\theta_C}(y_i)\}\big)\right].$$

To estimate parameters, we use a maximum likelihood approach by maximizing this log-likelihood:

$$\hat{\alpha} = (\hat{\theta}, \hat{\theta}_T, \hat{\theta}_C)^T = \arg\max_{\alpha \in \mathcal{A}} \ell(\alpha; D),$$

where $\mathcal{A} = \Theta \times \Theta_T \Theta_\times C$. For example, with lognormal margins for T and C and single-parameter copula families, this involves optimizing over five parameters.

To establish the asymptotic normality of $(\theta, \theta_T, \theta_C)$, we utilize White's (1982)[7] results, which provide sufficient conditions for the asymptotic normality of an estimator obtained by maximizing a criterion function, even if the model is misspecified. Define the parameter vector $\alpha^* = (\theta^*, \theta_T^*, \theta_C^*)^\top$ as the one that minimizes the Kullback–Leibler information criterion $E\{\log f_{Y,\delta}(Y, \delta) - \log f_{Y,\delta,\alpha}(Y, \delta)\}$. Let $d = \dim(\theta) + \dim(\theta_T) + \dim(\theta_C)$. The following result,

based on White (1982), holds. White's regularity conditions (A1)–(A6) are assumptions about the true density $f_{Y,\delta}(y,\delta)$, the assumed density $f_{Y,\delta,\alpha}(y,\delta)$, and the density's derivatives with respect to α and y.

Theorem 5

(i) Under the regularity conditions (A1)–(A3) in White (1982)[7],

$$(\hat{\theta}, \hat{\theta}_T, \hat{\theta}_C) \to (\theta^*, \theta_T^*, \theta_C^*) \quad \text{in probability as } n \to \infty.$$

(ii) Under the regularity conditions (A1)–(A6) in White (1982)[7],

$$\sqrt{n}\left((\hat{\theta}, \hat{\theta}_T, \hat{\theta}_C) - (\theta^*, \theta_T^*, \theta_C^*)\right) \to N(0, V) \quad \text{in distribution as } n \to \infty,$$

where $V = A(\alpha^*)^{-1}B(\alpha^*)A(\alpha^*)^{-1}$, with

$$A(\alpha) = \left[E\left(\frac{\partial^2}{\partial\alpha_j \partial\alpha_k} \log f_{Y,\delta,\alpha}(Y,\delta)\right)\right]_{j,k=1}^{d},$$

and

$$B(\alpha) = \left[E\left(\frac{\partial}{\partial\alpha_j} \log f_{Y,\delta,\alpha}(Y,\delta) \frac{\partial}{\partial\alpha_k} \log f_{Y,\delta,\alpha}(Y,\delta)\right)\right]_{j,k=1}^{d}.$$

If the model is correctly specified, V simplifies to $A(\alpha)^{-1}$, the inverse of the Fisher information matrix.

5 Simulation

We analyze the performance of the maximum likelihood estimators for θ, θ_T, and θ_C across four parametric copula families: Frank, Clayton, Gumbel, and Gaussian, using lognormal margins for T and C. In the lognormal model, the parameters μ and σ represent the mean and standard deviation of $\log(X)$, respectively, for a lognormal random variable X. Two simulation scenarios are considered, with parameter settings detailed in Table 1. Visualizations of the theoretical density, survival, and hazard functions of Y for each scenario are provided in Figures 1 and 2.

Table 1: Parameter specifications for the simulation scenarios with lognormal margins, with mean parameters μ_T and μ_C and standard deviation parameters σ_T and σ_C for T and C, respectively, and dependency measured by Kendall's τ

Scenario	μ_T	σ_T	μ_C	σ_C	τ	θ_{Frank}	θ_{Clayton}	θ_{Gumbel}	θ_{Gauss}
1	2.2	1.0	2.0	0.25	0.2	1.86	0.50	1.25	0.31
					0.5	5.74	2.00	2.00	0.71
					0.7	11.74	4.67	3.33	0.89
2	2.5	1.0	2.0	0.50	0.2	1.86	0.50	1.25	0.31
					0.5	5.74	2.00	2.00	0.71
					0.7	11.74	4.67	3.33	0.89

In Scenario 1, the marginal density of the observable random variable is nonstandard, which is anticipated because it results from the sum of two subdensities. We observe that the strength of dependence between and impacts the skewness of Y.

We analyze two sample sizes, and , repeating each simulation setting 100 times. To calculate the asymptotic standard errors, we numerically determine the required Hessian matrix. However, numerical evaluation of the Hessian

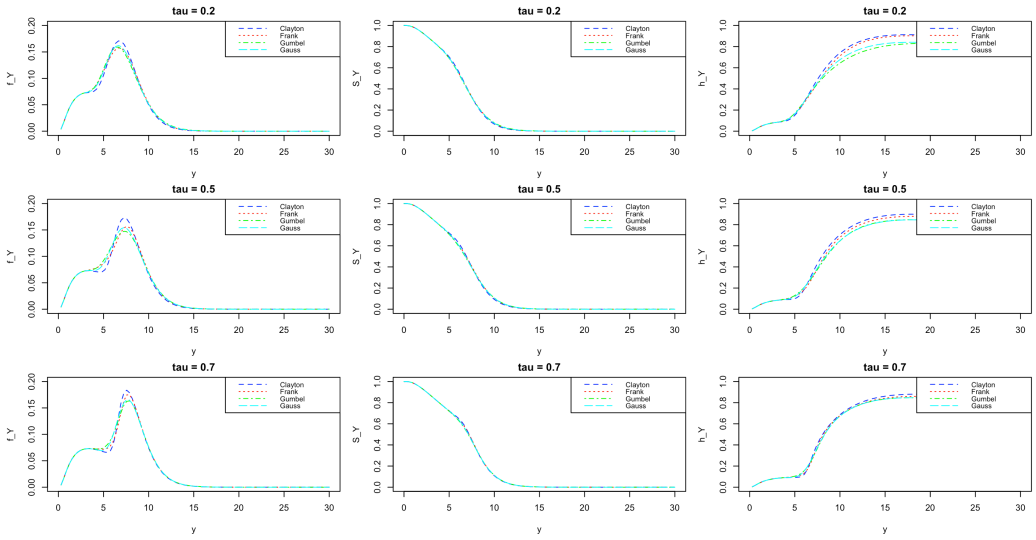

Figure 1: Theoretical density (left column), survival function (middle column) and hazard (right row) of Y for four copula families and three values under Scenario 1.

becomes unstable with extremely large copula parameter values. Therefore, we limit the copula parameter to ensure that the resulting Kendall's does not exceed 0.8. This constraint is not overly restrictive, as it corresponds to a correlation of 0.95 in the Gaussian copula. L-BFGS was used as the optimisation algorithm since it is suitable to run on limited amount of computer memory.

For our simulation experiments, we report the average estimate, the empirical standard error of this estimate, the average asymptotic standard error estimates for the parameter estimators, and the empirical root mean squared error based on 100 replications. The results for Scenario 1 are presented in Tables 2, 3, 4, and 5 for the Frank, Clayton, Gumbel, and Gaussian copulas, respectively, demonstrating satisfactory performance of the estimation procedure. As anticipated, the average root mean squared error decreases

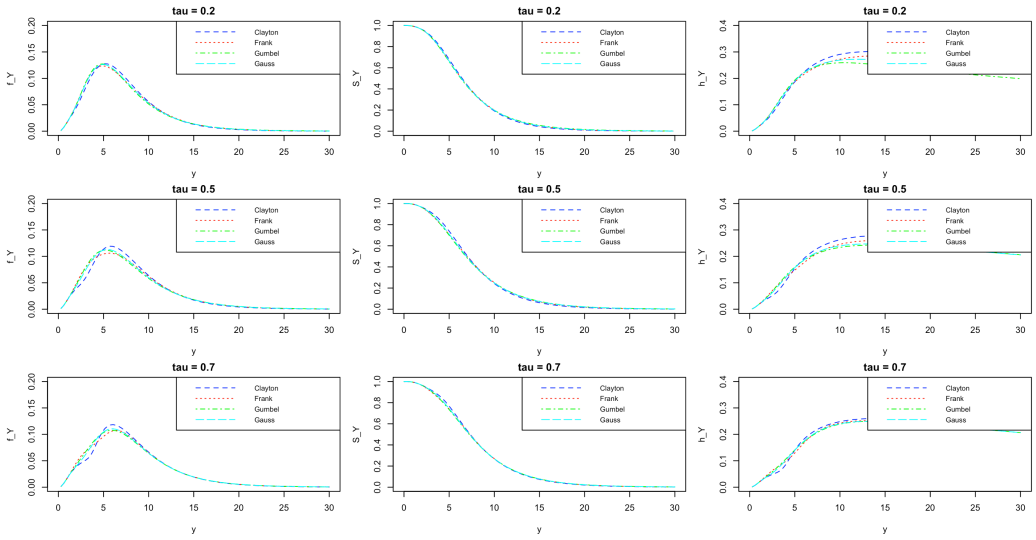

Figure 2: Theoretical density (left column), survival function (middle column) and hazard (right column) of Y for four copula families and three values under Scenario 2.

with an increase in sample size. The marginal parameters are accurately estimated in all cases, and the asymptotic standard error estimates align well with the empirical standard errors.

Moreover, the marginal parameters exhibit nearly perfect unbiasedness even at , leading to an almost exact agreement between the root mean squared error and the empirical standard errors. While 100 Monte Carlo replications may seem limited, the results indicate that this number is adequate.

Table 2: Simulation results for the Frank copula

	τ	μ_T	σ_T	μ_C	σ_C	Estimated τ
n=200						
aver.est	0.7	2.22	1.03	2.00	0.25	0.70
sd.aver.est	0.7	0.11	0.10	0.03	0.02	0.04
aver.asderr	0.7	0.10	0.10	0.03	0.02	0.06
RMSE	0.7	0.11	0.11	0.03	0.02	0.04
aver.est	0.5	2.22	1.02	2.01	0.24	0.44
sd.aver.est	0.5	0.13	0.12	0.05	0.02	0.18
aver.asderr	0.5	0.11	0.10	0.04	0.02	0.14
RMSE	0.5	0.13	0.12	0.05	0.02	0.18
aver.est	0.2	2.26	1.00	2.00	0.25	0.22
sd.aver.est	0.2	0.14	0.09	0.02	0.02	0.12
aver.asderr	0.2	0.12	0.11	0.04	0.02	0.18
RMSE	0.2	0.15	0.08	0.02	0.02	0.12
n=500						
aver.est	0.7	2.19	0.98	2.00	0.25	0.71
sd.aver.est	0.7	0.06	0.06	0.02	0.01	0.05
aver.asderr	0.7	0.06	0.06	0.02	0.01	0.04
RMSE	0.7	0.06	0.07	0.01	0.01	0.04
aver.est	0.5	2.19	0.99	2.00	0.25	0.47
sd.aver.est	0.5	0.08	0.07	0.03	0.02	0.08
aver.asderr	0.5	0.07	0.06	0.02	0.01	0.08
RMSE	0.5	0.08	0.07	0.03	0.02	0.08
aver.est	0.2	2.18	0.96	1.99	0.25	0.25
sd.aver.est	0.2	0.05	0.04	0.02	0.01	0.09
aver.asderr	0.2	0.07	0.06	0.03	0.01	0.11
RMSE	0.2	0.06	0.06	0.03	0.01	0.10

Table 3: Simulation results for the Clayton copula

	τ	μ_T	σ_T	μ_C	σ_C	Estimated τ
n=200						
aver.est	0.7	2.22	1.03	2.00	0.25	0.71
sd.aver.est	0.7	0.08	0.09	0.02	0.02	0.04
aver.asderr	0.7	0.10	0.09	0.03	0.02	0.07
RMSE	0.7	0.08	0.10	0.02	0.02	0.04
aver.est	0.5	2.20	1.02	2.01	0.24	0.45
sd.aver.est	0.5	0.10	0.08	0.05	0.03	0.18
aver.asderr	0.5	0.10	0.09	0.04	0.03	0.17
RMSE	0.5	0.10	0.08	0.05	0.03	0.18
aver.est	0.2	2.16	0.96	1.99	0.26	0.25
sd.aver.est	0.2	0.10	0.07	0.05	0.03	0.19
aver.asderr	0.2	0.11	0.09	0.07	0.03	0.25
RMSE	0.2	0.10	0.08	0.05	0.03	0.20
n=500						
aver.est	0.7	2.22	1.00	2.00	0.25	0.69
sd.aver.est	0.7	0.07	0.07	0.02	0.01	0.04
aver.asderr	0.7	0.06	0.06	0.02	0.01	0.05
RMSE	0.7	0.07	0.07	0.02	0.01	0.04
aver.est	0.5	2.19	1.00	2.00	0.25	0.52
sd.aver.est	0.5	0.07	0.07	0.02	0.02	0.07
aver.asderr	0.5	0.06	0.06	0.02	0.02	0.09
RMSE	0.5	0.07	0.07	0.02	0.02	0.07
aver.est	0.2	2.18	0.98	2.01	0.25	0.19
sd.aver.est	0.2	0.06	0.05	0.04	0.02	0.13
aver.asderr	0.2	0.07	0.06	0.05	0.02	0.19
RMSE	0.2	0.06	0.06	0.04	0.02	0.12

Table 4: Simulation results for the Gumbel copula

	τ	μ_T	σ_T	μ_C	σ_C	Estimated τ
n=200						
aver.est	0.2	2.21	1.00	1.99	0.25	0.24
sd.aver.est	0.2	0.11	0.10	0.04	0.02	0.13
aver.asderr	0.2	0.12	0.10	0.04	0.02	0.16
RMSE	0.2	0.10	0.10	0.04	0.02	0.13
aver.est	0.5	2.19	0.99	1.99	0.25	0.50
sd.aver.est	0.5	0.11	0.09	0.03	0.02	0.13
aver.asderr	0.5	0.11	0.10	0.03	0.02	0.13
RMSE	0.5	0.11	0.09	0.03	0.02	0.12
aver.est	0.7	2.24	1.04	2.01	0.24	0.67
sd.aver.est	0.7	0.11	0.07	0.02	0.02	0.05
aver.asderr	0.7	0.11	0.10	0.03	0.02	0.08
RMSE	0.7	0.12	0.08	0.03	0.02	0.06
n=500						
aver.est	0.2	2.19	0.98	2.00	0.25	0.22
sd.aver.est	0.2	0.05	0.05	0.03	0.01	0.12
aver.asderr	0.2	0.07	0.06	0.02	0.01	0.10
RMSE	0.2	0.06	0.05	0.03	0.01	0.12
aver.est	0.5	2.19	1.00	2.00	0.25	0.48
sd.aver.est	0.5	0.09	0.07	0.03	0.01	0.09
aver.asderr	0.5	0.07	0.06	0.02	0.01	0.09
RMSE	0.5	0.08	0.07	0.03	0.01	0.09
aver.est	0.7	2.18	0.99	2.00	0.25	0.70
sd.aver.est	0.7	0.07	0.05	0.02	0.01	0.04
aver.asderr	0.7	0.06	0.06	0.02	0.02	0.05
RMSE	0.7	0.07	0.05	0.02	0.01	0.04

Table 5: Simulation results for the Gauss copula

	τ	μ_T	σ_T	μ_C	σ_C	Estimated τ
n=500						
aver.est	0.7	2.19	1.00	2.00	0.25	0.71
sd.aver.est	0.7	0.05	0.07	0.02	0.01	0.04
aver.asderr	0.7	0.06	0.06	0.02	0.01	0.04
RMSE	0.7	0.05	0.07	0.02	0.01	0.04
aver.est	0.5	2.20	1.00	2.00	0.25	0.48
sd.aver.est	0.5	0.07	0.05	0.02	0.01	0.08
aver.asderr	0.5	0.07	0.06	0.02	0.02	0.09
RMSE	0.5	0.07	0.05	0.02	0.01	0.08
aver.est	0.2	2.18	0.99	1.99	0.25	0.23
sd.aver.est	0.2	0.05	0.05	0.03	0.02	0.13
aver.asderr	0.2	0.07	0.06	0.03	0.01	0.13
RMSE	0.2	0.05	0.05	0.03	0.02	0.13
n=200						
aver.est	0.7	2.22	1.00	2.00	0.24	0.68
sd.aver.est	0.7	0.09	0.08	0.02	0.02	0.08
aver.asderr	0.7	0.11	0.09	0.03	0.02	0.08
RMSE	0.7	0.09	0.08	0.02	0.02	0.08
aver.est	0.5	2.23	1.03	2.00	0.25	0.46
sd.aver.est	0.5	0.11	0.09	0.03	0.02	0.14
aver.asderr	0.5	0.11	0.10	0.04	0.02	0.15
RMSE	0.5	0.11	0.09	0.03	0.02	0.14
aver.est	0.2	2.19	1.00	1.99	0.25	0.21
sd.aver.est	0.2	0.10	0.11	0.04	0.02	0.16
aver.asderr	0.2	0.12	0.10	0.05	0.02	0.20
RMSE	0.2	0.09	0.11	0.04	0.02	0.16

6 Conclusion

The replication of the referenced paper[2] has been successfully achieved. In addition, we have established the identifiability of a copula model under dependent censoring, without presuming knowledge of the copula's association parameter. Our work also expands upon the original author's findings by demonstrating identifiability for additional marginal distributions, specifically Exponential, Gamma, and truncated Normal distributions.

Furthermore, we can conduct additional analyses to assess the model's sensitivity to misspecification. To enhance flexibility, we can explore semiparametric or nonparametric margins, raising the question of whether the identifiability of the model can still be assured. Additionally, we can incorporate covariates into the model. There is also considerable potential in utilizing more generalized survival models, such as competing risks models, cure models, and administrative censoring and truncation, in conjunction with dependent censoring.

References

[1] Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

[2] C Czado and I Van Keilegom. Dependent censoring based on parametric copulas. *Biometrika*, 110(3):721–738, 12 2022.

[3] Jongbloed G. Schwarz, M. and I. Van Keilegom. On the identifiability of copulas in bivariate competing risks models. *Canadian Journal of Statistics*, 41(2):291–303, 2013.

[4] Lee W. Sun L. Shih, J. and T. Emura. Fitting competing risks data to bivariate pareto models. *Communication in Statistics- Theory and Methods*, 2018.

[5] Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*.

[6] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 1975.

[7] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.

[8] M. Zheng and J. P. Klein. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 1995.