DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

MODELLING AND ANALYZING BIVARIATE SURVIVAL DATA BASED ON COPULAS

by

Ece GORCEĞİZ

March, 2020

İZMİR

MODELLING AND ANALYZING BIVARIATE SURVIVAL DATA BASED ON COPULAS

A Thesis Submitted to the

Graduate School of Natural And Applied Sciences of Dokuz Eylül University

In Partial of the Requirements for the Degree of Master of Science in

Statistics, Statistics Program

by

Ece GORCEĞİZ

March, 2020

İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"MODELLING AND ANALYZING BIVARIATE SURVIVAL DATA BASED ON COPULAS"** completed by **ECE GORCEĞİZ** under supervision of **PROF. DR. BURCU HÜDAVERDİ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Burcu HÜDAVERDİ

Supervisor

Dr. Öğr. Üyesi Özge Elmastaş Gültekin

(Jury Member)

Dr. Öğr. Üyesi Senan VAHAPLAR

(Jury Member)

Prof.Dr. Kadriye ERTEKİN

Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and thanks to my advisor, Prof. Dr. Burcu HÜDAVERDİ, for her guidance, continuous support and encouragement that she gave to me so I have accomplished my thesis study. Her suggestions played a very important role for my thesis.

I am grateful to Prof. Dr. Mehmet Nurullah ORMAN, Asst. Prof. Dr. Engin YILDIZTEPE, Dr. Alper VAHAPLAR, and Dr. Selim Orhun SUSAM for their continual encouragement and support all throughout the work.

I would like to acknowledge my dissertation committee members for their constructive comments and suggestions. Their effect certainly improved my perspective, and I hope that I have carried out their very helpful suggestions in this dissertation.

Besides I want to offer my special thanks to my friends Aylin GÖÇOĞLU, Burak DİLBER, Sami AKDENİZ, Yusuf Can SEVİL and Tolga YAMUT for their support, motivations and helpful contributions.

I wish to utter my special appreciation to my parents who have unfailingly supported me through all my life. My life coach, Dr. Elif ULU ERCAN who has always listened and guided me has provided constant encouragement and positive attitudes, which I will never forget for good.

Ece GORCEĞİZ

MODELLING AND ANALYZING BIVARIATE SURVIVAL DATA BASED ON COPULAS

ABSTRACT

Modelling dependence structure of a bivariate survival data is one of the main issues in biomedical studies. Copulas are key tools to analyze the dependence structures. A bivariate survival function can be expressed as the composition of marginal survival functions and a bivariate copula. Since a survival copula is a great deal of flexibility in modelling bivariate survival data, it provides an effective approach for understanding and modelling the dependent random variables and so the dependence structure. Survival copula deals with a lifetime data and is used for modelling and understanding the distributional structure. In survival studies, the researcher can come across censored survival data. In this study, we consider modelling and analyzing the bivariate survival data in the presence of right censoring using Archimedean copula functions. We use Emura et al. (2010) goodness-of-fit testing procedure for the model selection. Throughout the model selection procedure, we obtain the goodness-of-fit statistics for Gumbel, Frank and Clayton copula models. First, we examine the heart transplant data and model the dependence structure between waiting time for transplant and post-transplant survival time to see the co-movements of these variables. Second, we examine the diabetic retinopathy data and model the dependence between the survival times of the two eyes of the same patient in case of laser photocoagulation treatment. Finally, we use the survival hazard scenario approach to evaluate the probability of exceeding some critical layers. We develop R code to implement the study.

Keywords: Copula, right censoring, bivariate survival data, Archimedean copula, survival copula, survival hazard scenario

İKİ DEĞİŞKENLİ YAŞAM VERİLERİNİN KOPULAYA DAYALI MODELLEMESİ VE ANALİZİ

ÖZ

İki değişkenli sağkalım verilerinin bağımlılık yapısının modellenmesi, biyomedikal çalışmaların temel konularından biridir. Kopulalar bağımlılık yapılarını analiz etmek için anahtar araçlardır. İki değişkenli bir sağkalım fonksiyonu, marjinal sağkalım fonksiyonlarını iki değişkenli kopulaya bağlayan fonksiyon olarak elde edilebilir. Sağkalım kopulası olarak adlandırılan bu kopula, iki değişkenli sağkalım verilerinin modellenmesinde büyük bir esneklik sağladığından, bağımlı rastsal değişkenleri ve bağımlılık yapısını anlamak ve modellemek için etkili bir yaklaşım sağlar. Sağkalım çalışmalarında, araştırmacı sansürlü hayatta kalma verilerine rastlayabilir. Bu çalışmada, Archimedean kopula fonksiyonları kullanılarak iki değişkenli sağkalım verileri sağdan sansürlü ve sansürsüz olarak modellenmiş ve analiz edilmiştir. Model seçimi için Emura ve diğer. (2010) çalışmasındaki uyum iyiliği testi kullanılmıştır. İlk olarak, kalp nakli verileri incelenmiş ve bu değişkenlerin ortak hareketlerini görmek amacıyla nakil için bekleme süresi ile nakil sonrası sağkalım süresi arasındaki bağımlılık yapısı modellenmiştir. İkinci olarak, diyabetik retinopati verileri incelenmiş ve lazer fotokogülasyon tedavisi durumunda aynı hastanın iki gözünün yaşam süreleri arasındaki bağımlılığı modellenmiştir. Son olarak bazı kritik değerleri aşma olasılıklarını hesaplamak için sağkalım tehlike senaryosu yaklaşımı kullanılmıştır. Çalışmanın uygulaması için R kodu geliştirilmiştir.

Anahtar Kelimeler: Kopula, sağdan sansürleme, iki değişkenli sağkalım verileri, Arşimet kopula, sağkalım kopula, sağkalım tehlikesi senaryosu

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Copula approach is applied for modelling the dependence structure of the random variables. Copula model is a natural and proper tool for examining dependence structure rather than a joint distribution approach or correlation-based approach. A copula that connects the marginal distributions to restore the joint distribution is a function in statistics. Also, different copulas express different dependence structures between variables. A proper copula should be chosen for representing the data (Hu, 2006).

The distribution of bivariate survival data can be defined by a bivariate survival function. According to Sklar's theorem, a bivariate survival function that has continuous marginal survival functions can be uniquely presented by its (survival) copula function commentated at its marginal survival functions (Cheng, et al. 2014). Modelling dependence structure of a bivariate survival data is one of the main issues in biomedical studies. Survival copula deals with lifetime data and is used for modelling and understanding the distributional structure.

There are several important studies about multivariate survival modelling in literature. For example, Clayton (1978) discussed the application of Cox's (1972) regression model used in epidemiological studies of chronic disease incidence for censored survival data. To analyze familial tendency in disease incidence, a related model for the relation in bivariate survival time distributions was suggested. Hougaard (1986a) suggested a new three-parameter family of distributions by using Laplace transformation in the study. He asserted that for heterogeneous populations, the distributions were concerned for applications as frailty distributions in life-table methods and he discussed properties of such distributions. Hougaard (1986b) proposed a class of continuous multivariate lifetime distributions. In his study, the dependence between individuals in a group was modeled by a group-specific

quantity. Also, it is assumed that a group-specific quantity follows positive stable distribution and it also can be interpreted as an unobserved covariate common to the individuals in the group. Thus, it could be included covariates in the model and it could be discussed whether the dependence was presented after the specific covariates were considered. In Marshall and Olkin (1988) study, they interested in families of bivariate distributions having marginals as parameters. It has been indicated that several such families can be derived from a unified method by Genest and MacKay (1986). Also, Marshall and Olkin (1988) obtained mixture models. In this article, some of the new families of bivariate distributions are extended to specific distributions by adding a parameter for making them more flexible. Shih and Louis (1995) suggested two-stage parametric and two-stage semi-parametric estimation procedures for the association parameter in copula models for censored bivariate survival data in an AIDS data set. Besides, they derived asymptotic properties of the estimators and compared their performances by simulations. Bandeen-Roche and Liang (1996) suggested a family of distributional models that accounts for multiple levels of clustering and reduces to a univariate frailty model in the case of a single clustering level. They summarized the application of our model to marginal risk regression problems. Oakes (1989) showed that, when two observed survival times depend on a proportional-hazards model on the same unobserved variable (called a frailty), this common dependence induces an association between the observed times. Oakes (1989) also dealt with the class of bivariate survival distributions that can occur in this way, extended it to allow negative association and showed that the observable bivariate distribution determines the unobserved frailty distribution up to a scale parameter. Wang and Wells (2000) proposed a model selection procedure for bivariate survival models for censored data and a goodness-of-fit-based model selection methodology in the study.

In Chapter Three, the goodness-of-fit test is proposed by Emura et al. (2010) to estimate the true relationship between lifetime and censoring variable is given. The goodness-of-fit testing procedure of the Archimedean class for bivariate survival data is investigated. After the parameter estimation process, the goodness-of-fit testing

procedure is applied to check which (survival) copula fits best to the given data. The goodness-of-fit test provides simple graphical tools and numerical techniques for selecting an appropriate model, estimating its parameters, and checking its goodness-of-fit. The goodness-of-fit tests implement that the unknown copula (C) belongs to a parametric family (C_θ) of copulas, that is, $H_0 : C \in C_\theta$ where θ is the dependence parameter versus $H_1 : C \notin C_\theta$. Graphical methods, error statistics and formal goodness-of-fit statistics are used to measure the adequacy of this hypothesis. In this study, we follow the goodness-of-fit procedure for the Archimedean copula proposed by Emura et al. (2010).

In Chapter Four, hazard scenario concept is given. Modelling the hazard scenario is used for evaluating the maximum threat expected from a studied event in a certain area and to give certain directives to authorities to prevent and reduce serious results on the environment and population infrastructures (Paulatto, Pinat & Romanelli, 2007). Moreover, it some probabilities exceeding the survival critical layer for arriving maximum threat expected in the data set are given

In Chapter Five, the application study is given. Our data are obtained from recipients of the Stanford heart transplant program (69 transplants) and diabetic retinopathy study (197 patients). Firstly, the dependence structure between waiting time for transplant and post-transplant survival time is modelled to see the co-movements of these variables. For the diabetic retinopathy study, the dependence between the survival times of the two eyes of the same patient in case of laser photocoagulation treatment is modelled. Finally, by using survival hazard scenario, the selected critical layer, critical level, observations exceeding the critical layers and probability of exceeding the critical layer are investigated.

CHAPTER TWO
COPULA

2.1 Introduction to Copula

2.1.1 Concepts Related to Copula

Copula is described as a function that joins multivariate distribution function to its one-dimensional marginal distribution functions which have uniform distribution in the interval [0,1]. Copulas are applied for analyzing multivariate results and are also helpful for creating a model in period (Kpanzou, 2008). Copulas model the dependence structure between random variables and are used for building families of multivariate distributions.

The word copula is defined as "that part of a proposition which connects the subject and predicate" (Oxford English Dictionary). Also, the copula has a meaning as "a link, bond, tie" (Cassell's Latin Dictionary). We will deal with a definition of copulas and some of their elementary properties.

Let X and Y be a pair of random variables, with distribution functions $F(x) = P[X \le x]$ and $G(y) = P[Y \le y]$ and joint distribution function $H(x, y) = P[X \le x, Y \le y]$. Each pair (X, Y) of real numbers corresponds to a point $(F(x), G(y))$ in the unit square $[0,1]\times[0,1]$ and definition interval of the joint distribution $H(x, y)$ corresponds to a number in $[0,1]$. Actually, the value of the joint distribution function to each pair of values of the marginal distribution functions appoint to a function (Nelsen, 2006). Before we examine the copula function we will define some notions.

Definition 2.1 ***(2-increasing)*****:** Let **R** indicate the ordinary real line $(-\infty,+\infty)$, and let $\bar{\mathbf{R}}$ indicate the extended real line $[-\infty,+\infty]$, and suppose that $\bar{\mathbf{R}}^2$ is the extended real plane as $\bar{\mathbf{R}}\times\bar{\mathbf{R}}$. Let also $B=[x_1,x_2]\times[y_1,y_2]$ be a rectangle in the extended real plane where vertices are $(x_1,y_1),(x_1,y_2),(x_2,y_1),(x_2,y_2)$. Besides, the unit square $\mathbf{I}^2$ is the product $\mathbf{I}\times\mathbf{I}$ where $\mathbf{I}=[0,1]$. The bivariate real function H is a function that has domain, *DomH* and has a subset of $\bar{\mathbf{R}}^2$ and whose range, *RanH*, is a subset of **R**.

Definition 2.2: Suppose S_1 and S_2 are nonempty subsets of $\bar{\mathbf{R}}$ and suppose that H is a bivariate real function such that $\mathrm{DomH}=S_1\times S_2$. Suppose $B=[x_1,x_2]\times[y_1,y_2]$ is a rectangle all and its vertices are in DomH. The *H-volume* of B is defined as

$$V_H(B)=H(x_2,y_2)-H(x_2,y_1)-H(x_1,y_2)+H(x_1,y_1) \tag{2.1}$$

Definition 2.3: If $V_H(B)\geq 0$ for all rectangles B and its vertices lie in DomH, the bivariate real function H is *2-increasing*. A 2-increasing H can be said as the H-measure of B.

2.1.2 Properties of Copula

Definition 2.4: Suppose that $\mathbf{C}'$ is a function (a two-dimensional subcopula) having the following properties;

1. S_1 and S_2 are subsets of **I** containing $[0,1]$, then $\mathrm{DomC}'=S_1\times S_2$,
2. $\mathbf{C}'$ is 2-increasing and grounded,
3. For every u and v in S_1 and S_2, respectively, $C'(u,1)=u$ and $C'(1,v)=v$.

For $\forall(u,v)$ in $\mathrm{Dom}C'$, and $0 \le C'(u,v) \le 1$, then $\mathrm{Ran}C'$ is a subset of $\mathbf{I}$.

Definition 2.5: A copula is a 2- subcopula whose domain is $\mathbf{I}^2$, and is a function $C: I^2 \to I$. The properties are,

1. For $\forall(u,v)$ in $\mathbf{I}$

$$C(u,0) = 0 = C(0,v)$$

and

$$C(u,1) = u \text{ and } C(1,v) = v;$$

2. For $\forall u_1, u_2, v_1, v_2$ in $\mathbf{I}$. If $u_1 \le u_2$ and $v_1 \le v_2$,

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \ge 0 \qquad (2.2)$$

Since $C(u,v) = V_C([0,u]\times[0,v])$, it means that $C(u,v)$ is designated as a number in $\mathbf{I}$ to the rectangle $[0,u]\times[0,v]$. The number assigned by C to all rectangles $[u_1,u_2]\times[v_1,v_2]$ in $\mathbf{I}^2$ should be nonnegative by the 2- increasing property.

Many of the properties of copulas are also the properties of the subcopulas. There is a minor distinction between a copula and a subcopula.

Theorem 2.1: Suppose that C' is a subcopula for every (u,v) in $\mathrm{Dom}C'$,

$$\max(u+v-1,0) \le C'(u,v) \le \min(u,v). \qquad (2.3)$$

Proof: Suppose that (u,v) is any point in $\mathrm{Dom}C'$. $C'(u,v) \le C'(u,1) = u$ and $C'(u,v) \le C'(1,v) = v$ and so the upper-bound is $C'(u,v) \le \min(u,v)$. Moreover,

$V_{C'}\left([u,1]\times[v,1]\right)\geq 0$ means $C'(u,v)\geq u+v-1$, and also the lower-bound is $C'(u,v)\geq \max(u+v-1,0)$.

Every copula is a subcopula and the upper-bound is denoted by $M(u,v)=\min(u,v)$ and the lower-bound is denoted by $W(u,v)=\max(u+v-1,0)$. Consequently, these bounds are themselves copulas and are showed by

$$W(u,v)\leq C(u,v)\leq M(u,v) \qquad (2.4)$$

for every copula C and $\forall(u,v)$ in $\mathbf{I}^2$. Equation (2.4) is called the *Frèchet- Hoeffding bound inequality*. Finally, a third important copula is the *product copula* which is $\Pi(u,v)=uv$ (Cherubini, et. al. 2004).

Figure 2.1 shows copulas of minimum (W), product (Π), maximum (M) in 3-dimensions (above panel) and 2-dimensions (below panel).

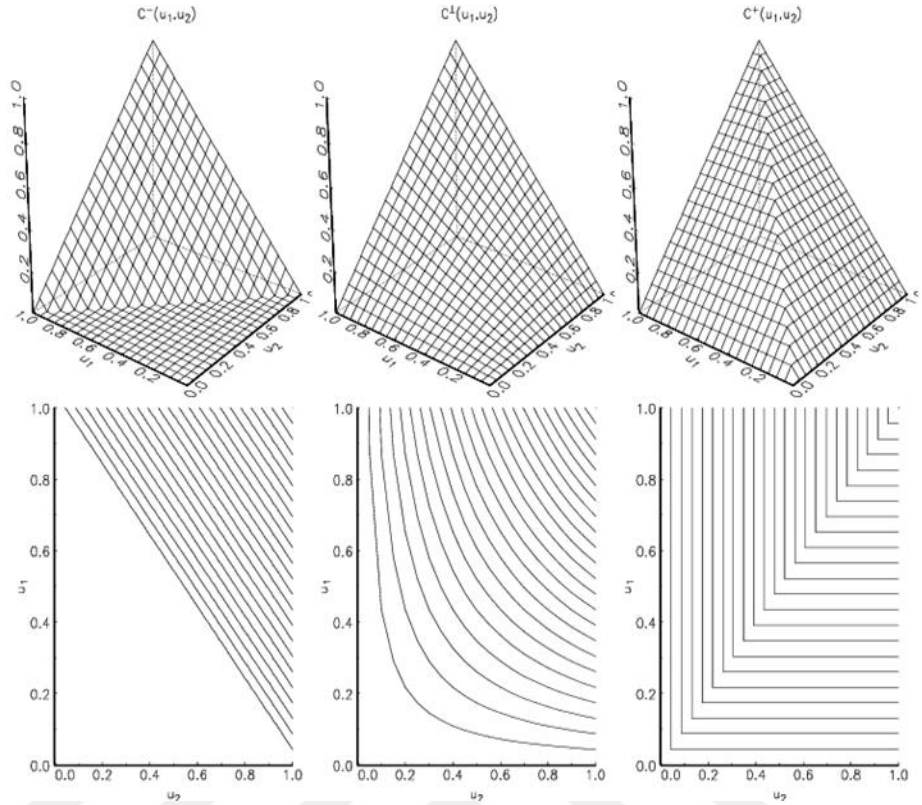

Figure 2.1 Copulas of minimum W, product Π, and maximum M copulas in 3-dimensional form (above panel) and as level curves (below panel) (Boyuè, E. et. al 2000)

Definition 2.6 (Comonotonic Copula and Countermonotonic Copula): Let $U, V \overset{i.i.d.}{\sim} U[0,1]$ for k=2 dimensions. U and V are independent if and only if C is a product copula (i.e., $C(u,v) = uv$). Besides, perfect positive dependence is known as *comonotonic* and perfect negative dependence is known as *countermonotonic*. For any (u_j, v_j) and (u_k, v_k) a *comonotonic* set is $\{u_j \le v_j, u_k \le v_k\}$ or $\{u_j \ge v_j, u_k \ge v_k\}$. Similarly, *countermonotonic* set is $\{u_j \ge v_j, u_k \le v_k\}$ or $\{u_j \le v_j, u_k \ge v_k\}$. Moreover, U is an increasing function of V if and only if $C(.) = C_U(.)$, that corresponds to *comonotonic* and perfect positive dependence. On the other hand, U is a decreasing function of V if and only if $C(.) = C_L(.)$, that corresponds to *countermonotonic* and perfect negative dependence. It means that the association is positive when the copula attains *the upper Frèchet bound* and negative when the

copula attains the *lower Frèchet bound*. Copulas have an invariance property that the dependence captured by a copula is invariant in terms of increasing and continuous transformations of the marginal distributions (Schweizer, & Sklar, 1983). For example, same copula can be used for the joint distribution of (U, V) as $(\mathrm{InU}, \mathrm{InV})$ and the marginals are stated as (U, V) or logarithmic values do not affect the copula (Trivedi, & Zimmer, 2005).

If distribution functions are continuous, the product copula (or called *independence copula*) $\prod(u, v) = uv$ characterizes the independent random variables.

In nonparametric statistics, since copulas are invariant the copula is very beneficial in strictly monotone transformations of random variables. If the distribution functions of a random variable X is continuous, and also α is a strictly monotone function whose domain contains RanX, the distribution function of the random variable $\alpha(X)$ is also continuous.

Definition 2.7 (Concordant, Discordant): Suppose that (x_i, y_i) and (x_j, y_j) are observations of the two pairs. When $y_i < y_j$ and $x_i < x_j$ when $x_i > x_j$ and $y_i > y_j$, so observations (x_i, y_i) and (x_j, y_j) are *concordant*. Otherwise, (x_i, y_i) and (x_j, y_j) are *discordant* when $x_i < x_j$ and $y_i > y_j$ or when $x_i > x_j$ and $y_i < y_j$. It is given as concordant formulation,

$$(x_i - x_j)(y_i - y_j) > 0, \tag{2.5}$$

and as discordant formulation,

$$(x_i - x_j)(y_i - y_j) < 0. \tag{2.6}$$

Definition 2.8 (Spearman's Rank Correlation): Suppose that X and Y are random variables having distribution functions F_1 and F_2. For sample, the estimator of *Spearman's rank correlation* $\hat{\rho}_S(X,Y)$ is given as,

$$\hat{\rho}(X,Y)=\frac{12}{n\left(n^2-1\right)}\sum_{i=1}^{n}\left(\text{rank}\left(X_i\right)-\frac{n+1}{2}\right)\left(\text{rank}\left(Y_i\right)-\frac{n+1}{2}\right) \qquad (2.7)$$

and for population, *Spearman's rank correlation* $\rho_S(X,Y)$ is,

$$\rho_S(X,Y)=\rho\left(F_1(X),F_2(Y)\right), \qquad (2.8)$$

where ρ is the usual linear correlation.

Definition 2.9 (Kendall's Tau (τ)): *Kendall's tau (τ)* is known as rank correlation coefficient. The rank correlation examines the correlation between the rankings of X and Y. They are unaffected by any increasing transformation of X and Y.

Kendall's tau is defined as the probability of concordance minus the probability of discordance. *Kendall's τ coefficient* is shown as

$$\tau=P\left[\left(X-X'\right)\left(Y-Y'\right)\geq 0\right]-P\left[\left(X-X'\right)\left(Y-Y'\right)\leq 0\right], \qquad (2.9)$$

where (X,Y) and (X',Y') are two independent pairs of random variables from a common cumulative distribution function F.

For the sample, Kendall's τ statistic is defined as $\hat{\tau}=\frac{c-d}{c+d}=(c-d)\Big/\binom{n}{2}$ where c is the number of concordant pairs and d is the number of discordant pairs from a sample of n observations from (X,Y).

Theorem 2.2: Suppose that (X, Y) and (X', Y') are independent vectors of continuous random variables having joint distribution functions H_1 and H_2 respectively. (X, X') and (Y, Y') have marginal distribution functions F and G. Suppose that C_1 and C_2 show the copulas of (X, Y) and (X', Y'), respectively, so that $H_1(x, y) = C_1(F(x), G(y))$ and $H_2(x, y) = C_2(F(x), G(y))$. Using Kendall's τ coefficient equation (2.9),

$$\tau = \tau(C_1, C_2) = 4\int\int_{I^2} C_2(u, v)\, dC_1(u, v) - 1. \tag{2.10}$$

Proof. See, Nelsen, 2006.

This equation can be interpreted as the expected value of the function $C(U, V)$ of uniform $(0,1)$ random variables U and V that their joint distribution is C,

$$\tau_C = 4E(C(U, V)) - 1. \tag{2.11}$$

Corollary: Suppose that X and Y are random variables having an Archimedean copula C generated by φ in Ω. For the population version, Kendall's tau is

$$\tau_C = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt. \tag{2.12}$$

Proof*:* Suppose that U and V are uniform $(0,1)$ random variables with joint distribution function C, and suppose that K_C is the distribution function of $C(U, V)$. Then

$$\tau_C = 4E(C(U, V)) - 1 = 4\int_0^1 t K_C(t) - 1 \tag{2.13}$$

which, upon integration by parts, yields

$$\tau_C = 3 - 4\int_0^1 K_C(t)dt . \tag{2.14}$$

The function K_C of $C(U,V)$ is

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)} \tag{2.15}$$

so

$$\tau_C = 3 - 4\int_0^1 \left[t - \frac{\varphi(t)}{\varphi'(t^+)} \right] dt = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \tag{2.16}$$

Since the concave functions are differentiable almost everywhere it is put away $\varphi'(t^+)$ by $\varphi'(t)$ in the denominator of the integrand (Nelsen, 2006; Lai, & Xie, 2006; Chen, 2012).

The advantages of rank correlation are that

1. The rank correlation is quite robust against outliers
2. The invariance in monotonic transformations and perfect dependence corresponds to correlations of 1 and -1.

However, disadvantage of rank correlation is that rank correlations do not give themselves to same elegant variance-covariance manipulations, in other words they are not moment-based correlations. For calculations, there are cases that rank correlations are easier to calculate.

Definition 2.10 (Tail Dependence): *The Kendall's tau* and Spearman rank correlation are two global measures of association. By *tail dependence* concept, the dependence can be characterized more locally. The notion of bivariate tail dependence depends on the tails and relates to the dependence on extreme values. Suppose that X and Y are random variables having distribution F_1 and F_2. The *coefficient of upper tail dependence* of X and Y is defined as

$$\lim_{\alpha \to 1^-} P\left[Y > F_2^{-1}(\alpha) \middle| X > F_1^{-1}(\alpha) \right] = \lambda_{u,} \tag{2.17}$$

for a limit $\lambda_u \in [0,1]$. Here,

$$F^{-1}(\alpha) = \inf \left\{ x \middle| F(x) \geq \alpha \right\}, \text{ for } \alpha \in (0,1). \tag{2.18}$$

When $\lambda \in (0,1]$, X and Y are said to be *asymptotically dependent* in the upper tail. When $\lambda = 0$, X and Y are said to be *asymptotically independent* in the upper tail. The *coefficient of lower tail dependence* is given as

$$\lim_{\alpha \to 0^-} P\left[Y \leq F_2^{-1}(\alpha) \middle| X \leq F_1^{-1}(\alpha) \right] = \lambda_{\ell}, \tag{2.19}$$

for a limit $\lambda_{\ell} \in [0,1]$.

When $\lambda \in (0,1]$, X and Y are said to be *asymptotically dependent* in the lower tail. When $\lambda = 0$, X and Y are said to be *asymptotically independent* in the lower tail.

Linear correlation causes problems when we work with heavy-tailed distributions. The copulas are proper functions for modelling tail dependence because the copulas are invariant under strictly increasing transformations of the margins. When F_1 and F_2 are continuous, it can be verified, under assumption that the limit exists, that

$$\begin{aligned}\lambda_u &= \lim_{\alpha\to 1^-} P\left[Y > F_2^{-1}(\alpha)\middle| X > F_1^{-1}(\alpha)\right] \\ &= \lim_{\alpha\to 1^-} \frac{P\left[Y > F_2^{-1}(\alpha), X > F_1^{-1}(\alpha)\right]}{P\left[X > F_1^{-1}(\alpha)\right]} \\ &= \lim_{\alpha\to 1^-} \frac{\bar{C}(\alpha,\alpha)}{1-\alpha}\end{aligned} \tag{2.20}$$

where $\bar{C}(u,u) = 1 - 2u + C(u,u)$ and C is the unique copula related to $(X,Y)^t$,

$$\begin{aligned}\lambda_\ell &= \lim_{\alpha\to 0^+} P\left[Y \le F_2^{-1}(\alpha)\middle| X \le F_1^{-1}(\alpha)\right] \\ &= \lim_{\alpha\to 0^+} \frac{P\left[Y \le F_2^{-1}(\alpha), X \le F_1^{-1}(\alpha)\right]}{P\left[X \le F_1^{-1}(\alpha)\right]} \\ &= \lim_{\alpha\to 0^+} \frac{C(\alpha,\alpha)}{\alpha}\end{aligned} \tag{2.21}$$

In equation (2.20) and (2.21), tail dependence is best understood as an asymptotic property of the copula. When the copula has a simple closed-form, λ is calculated easily.

Theorem 2.3: Suppose that C is a copula and its partial derivative according to u $\partial C(u,v)/\partial u$ is

$$0 \le \frac{\partial}{\partial u} C(u,v) \le 1, \tag{2.22}$$

for any v in I and nearly all u, for such v and u. Otherwise, the partial derivative according to v, $\partial C(u,v)/\partial v$ is

$$0 \le \frac{\partial}{\partial v} C(u,v) \le 1, \tag{2.23}$$

for almost v and any u in I, for such u, v.

Moreover, the functions $u \mapsto \partial C(u,v)/\partial v$ and $v \mapsto \partial C(u,v)/\partial u$ are *nondecreasing* almost everywhere on **I** .

Theorem 2.4 (Probability- Integral Transforms): Suppose that X is a random variable on probability space (Ω, F, P) having distribution function is given by $F_X(x)$. The continuous random variable $Y = F_X(X)$ is distributed uniform distribution $F_X(X) \sim U[0,1]$, namely, $P(Y \le y) = y$, $0 < y < 1$ (Ostwald, 2019).

2.1.3 Sklar's Theorem

Sklar's Theorem is vital for the theory of copula and is the basis of the applications of theory in Statistics. It can indicate the relationship of copula to the multivariate distributions and their marginal distributions. Clearly, it states that the multivariate distribution functions are represented as copula and univariate margins. We deal with distribution functions in this section.

Theorem 2.5 (Sklar's theorem): Suppose that a random vector (X, Y) is given on a probability space (Ω, F, P) and suppose that $H(x,y) := P(X \le x, Y \le y)$ is the distribution function and $F_1(x)$ and $F_2(y)$ are the marginal distributions. There is a dimensional copula

$$H(x,y) = C(F_1(x), F_2(y)), \tag{2.24}$$

When the marginals F_1, F_2 are continuous, the copula C is uniquely defined.

Remark 2.1: By calculating, if H in equation (2.24) is an absolutely continuous distribution, the density of H is shown as

$$c(x,y) = c\left(f_1(x), f_2(y)\right) \tag{2.25}$$

where F_1, F_2 are the univariate margins having densities f_1, f_2 and c is the *density of the copula* C.

When the marginals F_1 and F_2 of distribution function are continuous, (unique) copula by Sklar's theorem can be proved.

Lemma 2.1: Considering Sklar's Theorem, if F_1, F_2 are continuous, there is a unique copula C. For $\forall u \in I^2$, it is defined,

$$C(u) = H\left(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2)\right), \text{ for } j \in \{1,2\} \tag{2.26}$$

where $F_j^{(-1)}$ is the quasi-inverse of F_j.

Proof: Considering the univariate probability integral transform, when F_j is continuous, the distribution function of $(F_1 \circ X, F_2 \circ Y)$ has uniform univariate marginals and it is a copula. For $\forall t \in R^2$,

$$\begin{aligned} H(t) &= P(X \le x, Y \le y) \\ &= P\left(F_1(X) \le F_1(x), F_2(Y) \le F_2(y)\right) \\ &= C\left(F_1(x), F_2(y)\right) \end{aligned} \tag{2.27}$$

Remark 2.2: A copula is understood as transforming the continuous random variables X, Y into the other random variables $F_1(X), F_2(Y)$ having all uniform distribution on I.

Remark 2.3: Lemma 2.1 obtains visualizing the dependence structure (namely copula) of a random sample from an unknown distribution function H. Let d be 2-dimensional. Suppose that a random sample of (X_i, Y_i) is from a random pair

$(X, Y) \sim H$ for $i = 1, ..., n$. To visualize the dependence of the data, the marginal distributions F and G of X and Y respectively are estimated and are plotted $\left(F(X_i), G(Y_i)\right)$ for $i = 1, ..., n$. If marginal distribution functions are unknown as in several cases, the marginal distribution functions may be estimated using empirical distribution functions as,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le x\}} \quad \text{and} \quad G_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i \le y\}}. \qquad (2.28)$$

Remark 2.4: According to the previous Lemma 2.1, if U vector is a random variable distributed as the copula C and F_1, F_2 are univariate distribution functions, the random vector $\left(F_1^{-1}(U_1), F_2^{-1}(U_2)\right)$ is distributed as $H = C(F_1, F_2)$. In order to simulate a random sample from H, the following is done;

1. Simulate (u_1, u_2) from the copula C.
2. Return $\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right)$

When the inverse of the univariate margin exists, a copula may be simulated, and also any multivariate distribution function may be simulated (Durante, & Sempi, 2015).

2.2 Survival Copula

In this section, we focus on the notion of survival copula. It deals with lifetime data and is used for modelling and understanding the distributional structure. The random variables can be the lifetimes of the individuals or objects for some population in many studies. Firstly we need to know the notion of the survival function and its relation with some notions for understanding the survival copula.

Definition 2.11: Let T_1 be the probability of an individual living or surviving time. The survival function is $\bar{F}(t_1) = P[T_1 > t_1] = 1 - F(t_1)$, where F denotes the distribution function of T_1 for $[0, \infty)$.

Definition 2.12: Let T_1 and T_2 be two possibly dependent lifetimes. The joint distribution function of survival times determined as the probability of the joint event $P(T_1 \le t_1, T_2 \le t_2)$ for $n = 2$ dimension. It is indicated as:

$$\begin{aligned} F(t_1, t_2) &= P[T_1 \le t_1, T_2 \le t_2], \\ &= 1 - P[T_1 > t_1] - P[T_2 > t_2] + P[T_1 > t_1, T_2 > t_2]. \end{aligned} \quad (2.29)$$

The joint survival probability $P(T_1 > t_1, T_2 > t_2)$ is

$$\begin{aligned} S(t_1, t_2) &= P[T_1 > t_1, T_2 > t_2], \\ &= 1 - F(t_1) - F(t_2) + F(t_1, t_2), \\ &= S_1(t_1) + S_2(t_2) - 1 + \bar{C}(1 - S_1(t_1), 1 - S_2(t_2)), \end{aligned} \quad (2.30)$$

where $\bar{C}(u, v)$ is the *survival copula*, $S_1(t_1)$ and $S_2(t_2)$ are the marginal survival distributions and $S(t_1, t_2)$ is the joint survival function (Trivedi, & Zimmer, 2005).

Definition 2.13: The survival copula is related to the copula C, as

$$\bar{C}(u, v) = u + v - 1 + C(1 - u, 1 - v). \quad (2.31)$$

It represents the probability that two standard uniform variable having copula C be greater than u and v since

$$\begin{aligned} \bar{C}(1 - u, 1 - v) &= 1 - u + 1 - v - 1 + C(u, v) \\ &= 1 - P(U \le u) + 1 - P(V \le v) - 1 + P(U \le u, V \le v) \\ &= P(U > u) + P(V > v) - 1 + P(U \le u, V \le v) \\ &= P(U > u, V > v) \end{aligned} \quad (2.32)$$

(Cherubini, Luciano, & Vecchiato, 2004).

2.2.1 Properties of Survival Copula

Theorem 2.6: 1. The margins of $\bar{C}$ are *uniform*:

$$\bar{C}(u,1) = u + C(1-u,0) = u$$

and similarly

$$\bar{C}(1,v) = v$$

2. $\bar{C}$ is a *grounded function*:

$$\begin{aligned}\bar{C}(u,0) &= \bar{C}(0,u)\\ &= u-1+C(1,1-u)\\ &= u-1+1-u\\ &= 0\end{aligned}$$

3. $\bar{C}$ is *2- increasing*:

$$V_{\bar{C}}\left([u_1,v_1]\times[u_2,v_2]\right) = \bar{C}(v_1,v_2) - \bar{C}(v_1,u_2) - \bar{C}(u_1,v_2) + \bar{C}(u_1,u_2) \geq 0 \tag{2.33}$$

for $(u_1,u_2)\in[0,1]^2$, $(v_1,v_2)\in[0,1]^2$ in $0\leq u_1\leq v_1\leq 1$ and $0\leq u_2\leq v_2\leq 1$. Hence we calculate

$$V_{\bar{C}}\left([u_1,v_1]\times[u_2,v_2]\right) = C(1-v_1,1-v_2) - C(1-v_1,1-u_2) - C(1-u_1,1-v_2) + C(1-u_1,1-u_2) \tag{2.34}$$

Applying the notations $\dot{u}_i = 1-u_i$ and $\dot{v}_i = 1-v_i$, then

$$V_{\bar{C}}\left(\left[u_1, v_1\right]\times\left[u_2, v_2\right]\right) = V_C\left(\left[\dot{v}_1, \dot{u}_1\right]\times\left[\dot{v}_2, \dot{u}_2\right]\right)$$
$$\geq 0$$

where $0 \leq \dot{v}_1 \leq \dot{u}_1 \leq 1$ and $0 \leq \dot{v}_2 \leq \dot{u}_2 \leq 1$ and $\bar{C}$ is 2-increasing.

Theorem 2.7: For 2-dimensional case, the relationship between the survival function of the copula $\breve{C}$ and the survival copula $\bar{C}$ is given by

$$\bar{C}\left(u_1, u_2\right) = \breve{C}\left(1-u_1, 1-u_2\right)$$

with

$$\breve{C}\left(u_1, u_2\right) = \sum_{n=0}^{2}\left[\left(-1\right)^n \sum_{v\left(u_1,u_2\right)\in Z\left(2-n,2,1\right)} C\left(v_1, v_2\right)\right] \quad (2.35)$$

where the set $\left\{v\in\left[0,1\right]^2 \middle| v_n \in \left\{u_n, \varepsilon\right\}, \sum_{n=1}^{2} X_{\left\{\varepsilon\right\}}\left(v_n\right) = M\right\}$.

Because survival copula $\left(\bar{C}\right)$ is a copula function, conclusions of copula to the survival copula can be applied. For instance,

$$C^- \prec \bar{C} \prec C^+ \quad (2.36)$$

Property 2.1: The copula is *radially symmetric* if and only if $\bar{C} = C$.

Proof. A random vector $\left(T_1, T_2\right)$ is called to be radially symmetric about $\left(t_1^*, t_2^*\right)$ if and only if (for detail, see Nelsen, 1999)

$$F\left(t_1^* + t_1, t_2^* + t_2\right) = S\left(t_1^* - t_1, t_2^* - t_2\right) \quad \text{for all } t \in \mathbf{R}^2 \tag{2.37}$$

When it is dealt with the previous equation

$$\begin{aligned} & C\left(F_1\left(t_1^* + t_1\right), F_2\left(t_2^* + t_2\right)\right) = \bar{C}\left(S_1\left(t_1^* - t_1\right), S_2\left(t_2^* - t_2\right)\right) \quad \text{for all } t \in \mathbf{R}^2 \\ \Leftrightarrow\ & C\left(F_1\left(t_1^* + t_1\right), F_2\left(t_2^* + t_2\right)\right) = \bar{C}\left(F_1\left(t_1^* + t_1\right), F_2\left(t_2^* + t_2\right)\right) \text{ for all } t \in \mathbf{R}^2 \\ \Leftrightarrow\ & C\left(u_1, u_2\right) = \bar{C}\left(u_1, u_2\right) \text{ for all } u \in [0,1]^2 \end{aligned} \tag{2.38}$$

Because

$$\begin{aligned} F_n\left(t_n^* + t_n\right) &= F\left(+\infty, ..., +\infty, t_n^* + t_n, +\infty, ..., +\infty\right) \\ &= S\left(-\infty, ..., -\infty, t_n^* - t_n, -\infty, ..., -\infty\right) \\ &= S_n\left(t_n^* - t_n\right). \end{aligned} \tag{2.39}$$

So it has the similar properties with a copula

- The product copula C' for $N = 2$

$$\begin{aligned} \bar{C}'(u, v) &= u + v - 1 + (1-u)(1-v) \\ &= uv \\ &= C'(u, v). \end{aligned} \tag{2.40}$$

- The bivariate upper Frèchet copula C^+

$$\begin{aligned} \bar{C}^+(u, v) &= u + v - 1 + \min(1-u, 1-v) \\ &= u + v - \max(u, v) \\ &= \min(u, v) \\ &= C^+(u, v). \end{aligned} \tag{2.41}$$

- The bivariate lower Frèchet copula C^-

$$\begin{aligned}\bar{C}^-(u,v) &= u+v-1+\max(1-u-v,0)\\ &= -\min(1-u-v,0)\\ &= \max(u+v-1,0)\\ &= C^-(u,v).\end{aligned} \tag{2.42}$$

Property 2.2: If $C_1 \succ C_2$ then $\bar{C}_1 \succ \bar{C}_2$ for bivariate case.

Proof. : If $C_1 \succ C_2$,

$$\begin{aligned}&C_1(1-u,1-v) \geq C_2(1-u,1-v)\\ \Leftrightarrow\ &u+v-1+C_1(1-u,1-v) \geq u+v-1+C_2(1-u,1-v)\\ \Leftrightarrow\ &\bar{C}_1(u,v) \geq \bar{C}_2(u,v).\end{aligned} \tag{2.43}$$

Property 2.3: Suppose that X_1, X_2 are random variables with continuous distributions F_1, F_2 and a copula C. Suppose that G_1, G_2 are continuous distributions with 2-dimensions and it is showed $T_n = G_n^{(-1)}\left(1-F_n(X_n)\right)$. It indicates that the margins and copula of the random vector (T_1, T_2) are G_1, G_2 and the survival copula $\bar{C}$.

Proof. We know that $G_n^{(-1)}$ and $1-F_n$ are strictly increasing and strictly decreasing functions respectively. The distribution function of T_n is

$$\begin{aligned}P\{T_n \leq t_n\} &= P\left\{G_n^{(-1)}\left(1-F_n(X_n)\right) \leq t_n\right\}\\ &= 1-P\{F_n(X_n) \leq 1-G_n(t_n)\}\\ &= G_n(t_n).\end{aligned} \tag{2.44}$$

Property 2.4: Let the survival copula $\bar{C}$ be the absolutely continuous component at the points (u_1, u_2) and let the copula C be the absolutely continuous component at

the points $(1-u_1,1-u_2)$. Thus survival copula $\bar{C}$ is equal to the copula C. Besides, it is symmetric about point $\left(\frac{1}{2},\frac{1}{2}\right)$.

Proof. The margins of $\bar{C}$ are uniform and $\bar{C}$ is a grounded function. Suppose that C is absolutely continuous, then

$$\partial_{1,2}\bar{C}(u_1,u_2)=(-1)^2\,\partial_{1,2}C(1-u_1,1-u_2)\times\underbrace{(-1)\times(-1)}_{\substack{2\\ \text{times}}} \qquad (2.45)$$

It is known that $\bar{C}$ is a positive measure and also $\bar{C}$ is a copula function. If C is not absolutely continuous, it is assumed that the N-increasing property holds (Georges, et al. 2001).

Theorem 2.8: (The Inversion Method) Suppose that H is a joint distribution function and $\bar{H}$ is the survival function. The bivariate distribution function H has continuous marginals F and G. Then the survival copula is

$$\hat{C}(u,v)=\bar{H}\left(\bar{F}^{-1}(u),\bar{G}^{-1}(v)\right) \qquad (2.46)$$

where $\bar{F}$ and $\bar{G}$ are univariate survival functions (Kpanzou, 2008).

2.3 Archimedean Copulas

In this section, we present some basic definitions and properties for the class of Archimedean copulas.

One of the class of copulas is an *Archimedean copula*. There are some reasons that Archimedean copulas find a wide area of application in practice;

- Archimedean copulas are very easy to build,
- Many of the parametric families of copulas belong to this class,
- The Archimedean copulas have various types of dependence structure,
- The Archimedean copulas have simply-closed form expressions.
- The Archimedean representation allows reducing a multivariate copula to a single univariate function.
- The Archimedean copulas are useful in empirical modelling.

2.3.1 Definitions and Some Properties

Definition 2.14: Suppose that Φ is a set of functions $\varphi:[0,1]\to[0,\infty]$ which are continuous, *strictly decreasing*, *convex* and $\varphi(0)=\infty$, $\varphi(1)=0$. $\forall\varphi\in\Phi$ has inverse $\varphi^{-1}:[0,\infty]\to[0,1]$ which are continuous, strictly decreasing, convex and $\varphi^{-1}(0)=1$, $\varphi^{-1}(\infty)=0$.

A bivariate distribution function with margins uniform on $[0,1]$ given by

$$C(u,v)=\varphi^{-1}\left(\varphi(u)+\varphi(v)\right) \qquad 0\le u,v\le 1. \tag{2.47}$$

where each member of Φ generates a copula C. This equation is called Archimedean copula. The copula is defined by a function φ as an element of the class Φ.

It is not essential that $\varphi(0)$ is infinite in order that φ generates a copula. If $\varphi(0)$ is finite, Archimedean copula generated by φ is given by a *pseudo-inverse* of φ.

$$\varphi^{[-1]}=\begin{cases}\varphi^{-1}, 0\le t\le\varphi(0)\\ 0, \varphi(0)\le t\le\infty\end{cases}$$

The generators for $\varphi(0)=\infty$ are examined because it is essential that φ has an inverse on $(0,\infty)$.

If we build Archimedean copulas for different copula functions, we need to know the functions that work as generators and describe the corresponding copula.

Theorem 2.9: Suppose that C is an Archimedean copula with the generator φ. Then,

- C is symmetric; $C(u,v)=C(v,u)$ for all $u,v\in[0,1]$
- C is associative; $C(C(u,v),w)=C(u,C(v,w))$ for all $u,v,w\in[0,1]$
- If $c>0$ is any constant then $c\varphi$ is a generator of C.
- The marginal distributions of U and V are uniform on the interval $(0,1)$, so $C(u,1)=u$ for all $0\le u\le 1$ and $C(1,v)=v$ for all $0\le v\le 1$.

Proof. Suppose that $u,v,w\in[0,1]$, so

- $C(u,v)=\varphi^{-1}(\varphi(u)+\varphi(v))=\varphi^{-1}(\varphi(v)+\varphi(u))=C(v,u)$
- $$\begin{aligned} C(C(u,v),w) &= \varphi^{-1}\left(\varphi\left[\varphi^{-1}(\varphi(u)+\varphi(v))\right]+\varphi(w)\right) \\ &= \varphi^{-1}(\varphi(u)+\varphi(v)+\varphi(w)) \\ &= \varphi^{-1}\left(\varphi(u)+\varphi\left[\varphi^{-1}(\varphi(v)+\varphi(w))\right]\right) \\ &= C(u,C(v,w)) \end{aligned}$$

- The inverse of $y=c\varphi$ is $\varphi^{-1}\left(\frac{1}{c}y\right)$

$$\begin{aligned} C_{c\varphi}(u,v) &= \varphi^{-1}\left(\frac{1}{c}\left[c\varphi(u)+c\varphi(v)\right]\right) \\ &= \varphi^{-1}\left(\frac{1}{c}\left[c\{\varphi(u)+\varphi(v)\}\right]\right) \end{aligned}$$

$$= \varphi^{-1}\left(\varphi(u)+\varphi(v)\right)$$

- $C(u,1) = \varphi^{-1}\left(\varphi(u)+\varphi(1)\right) = \varphi^{-1}\left(\varphi(u)+0\right) = u$.

Theorem 2.10: Suppose that C is an Archimedean copula generated by a generator φ and $K_C(t)$ show the C- measure of the set $\left\{(u,v)\in I^2 \middle| C(u,v)\le t\right\}$, or equivalently, of the set $\left\{(u,v)\in I^2 \middle| \varphi(u)+\varphi(v)\ge \varphi(t)\right\}$.Then

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)} \quad \text{for any } t\in I. \tag{2.48}$$

Proof: See, De Matteıs, 2001

Proposition 2.1: Suppose that U and V are uniform random variables and their joint distribution function is the Archimedean copula C generated by $\varphi\in\Omega$.. The function $K_C(T\le t)$ is the distribution function of the random variable $C(U,V)$ (De Matteis, 2001).

Theorem 2.11: "The *density* of Archimedean copula is "

$$c(u,v) = \frac{-\varphi''(C(u,v))\varphi'(u)\varphi'(v)}{\left(\varphi'(C(u,v))\right)^3} \tag{2.49}$$

(Cherubini, Luciano & Vecchiato, 2004, 122).

Now, we present several commonly used Archimedean copulas. Table 2.1 gives dependence parameter space, generator and bivariate copula for Clayton, Gumbel, Frank and independence copula as a whole.

2.3.2 Clayton Copula

The Clayton copula is given as:

$$C_\theta(u,v)=\left[\max\left(u^{-\theta}+v^{-\theta}-1\right),0\right]^{-1/\theta} \tag{2.50}$$

where the generator function is $\varphi_\theta(t)=\left(t^{-\theta}-1\right)/\theta$ with $\theta\in[-1,\infty)-\{0\}$. The copula is strict and it is given as

$$C_\theta(u,v)=\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta} \text{ for } \theta>0 \tag{2.51}$$

If θ approaches zero, the Clayton copula become independence copula, and if θ approaches infinity, the Clayton copula become the two-dimensional comonotonic copula namely the Clayton copula reaches the Frèchet upper bound, but it reaches the Frèchet lower bound for no value. The Clayton copula is not explained with the negative dependence. It has been used to examine correlated risks since it presents *strong left tail dependence* and relatively *weak right tail dependence*. When the correlation between two events is strongest in the left tail of the joint distribution, Clayton copula is an appropriate in model selection.

2.3.3 Frank Copula

The Frank copula is given as:

$$C_\theta(u,v)=-\frac{1}{\theta}\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{\left(e^{-\theta}-1\right)}\right) \tag{2.52}$$

where the generator function is $\varphi_\theta(t)=\log\left(e^{-\theta}-1\right)-\log\left(e^{-\theta t}-1\right)$ with $\theta\in(-\infty,\infty)-\{0\}$. Frank copulas are strict Archimedean copulas. If θ approaches

zero, the Frank copula approaches the independence copula; if θ approaches infinity, the Frank copula becomes comonotonicity copula (or become Frèchet upper bound); similarly, if θ approaches negative infinity, the Frank copula becomes the counter-monotonicity copula (or Frèchet lower bound), which equal to the perfectly negative dependence between two variables. Frank copula is important for several causes:

- The dependence is symmetric in both tails,
- The Frank copula allows negative dependence between the marginals,
- The Frank copula interpolates between perfectly negative dependence and perfectly positive dependence.
- It is "comprehensive" in that Frèchet lower bound and Frèchet upper bound consist of the range of permissible dependence.

The dependence in the tails of the Frank copula tends to be relatively weak in regard to the Gaussian copula and the strongest dependence is centered in the middle of the distribution. Thus it indicates that Frank copula is most appropriate for the data that exhibits weak tail dependence. Moreover, Frank copula has been commonly used in empirical applications.

2.3.4 Gumbel Copula

The Gumbel copula is given as

$$C_\theta(u,v) = \exp\left\{-\left[\left(-\log(u)\right)^\theta + \left(-\log(v)\right)^\theta\right]^{\frac{1}{\theta}}\right\}, \qquad (2.53)$$

where the generator function is $\varphi_\theta(t) = \left(-\log(t)\right)^\theta$ with $\theta \in [1,\infty)$, and also φ_θ is a strict generator. Moreover, $\varphi(t)$ is a continuous function of t and $\varphi(1) = 0$.

If θ is equal to 1, the Gumbel copula is an independent copula. If θ approaches infinity, the limit of Gumbel copula is a comonotonic of the copula (or Frèchet upper

bound). Actually, this condition corresponds to the perfectly positive dependence between two variables. The Gumbel copula allows perfectly positive dependence but does not allow negative dependence. The parameter θ shows the strength of the dependence. The Gumbel copula shows strong right tail dependence and relatively weak left tail dependence. The Gumbel copula is an appropriate choice when outcomes are known to be strongly correlated at high values but less correlated at low values (Mei, 2016; Trivedi, & Zimmer, 2005).

Figure 2.2 gives contour plots of three Archimedean copulas (Gumbel, Frank and Clayton copula), thus it can be shown that three Archimedean copulas are different amongst themselves.

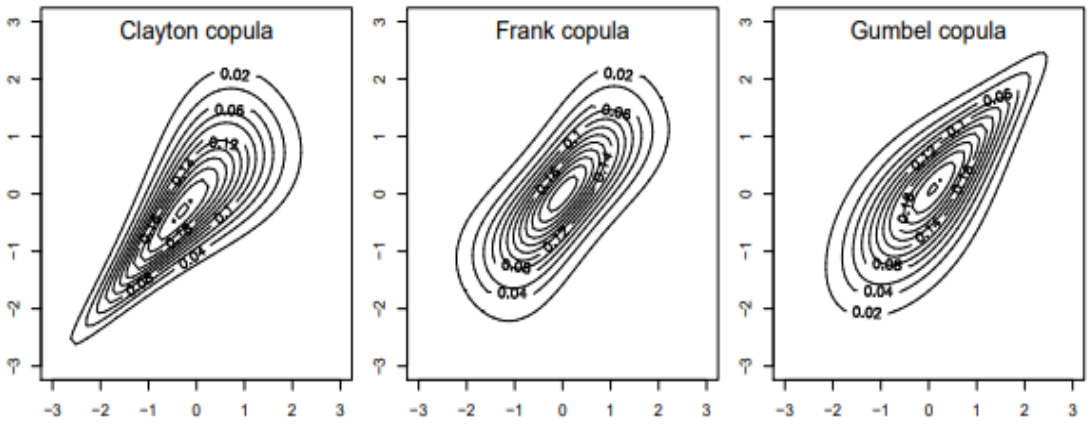

Figure 2.2 Contour plots of three Archimedean copulas

Table 2.1 Archimedean copulas and their generators

	Dependence Parameter (θ) Space	Generator φ(t)	Bivariate Copula $C_{\theta(u,v)}$
Clayton	$\theta > 0$	$\left(t^{-\theta}-1\right)/\theta$	$C_\theta(u,v)=\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$
Gumbel	$\theta \geq 0$	$\{-\log(t)\}^{\theta+1}$	$C_\theta(u,v)=\exp\left\{-\left[\left(-\log(u)\right)^\theta+\left(-\log(v)\right)^\theta\right]^{\frac{1}{\theta}}\right\}$
Frank	$\theta \neq 0$	$-\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$C_\theta(u,v)=-\frac{1}{\theta}\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{\left(e^{-\theta}-1\right)}\right)$
Independence	Not applicable	$-\mathrm{In}(t)$	uv

2.4 Kendall Distribution

Genest and Rivest (1993) first examined *Kendall distribution function* in two dimensions and Barbe et al. (1996) studied it in more general cases. The Kendall distribution function is derived that the random variable $V=S(T_1,T_2)$ has the cumulative distribution function. When the copula model is considered, let $T=(T_1,T_2)\sim C_\rho$, where C_ρ has a 2-dimensional copula, so the Kendall distribution function is defined as

$$K_\rho(v)=P\left(C_\rho \leq v\right) \tag{2.54}$$

It is showed by Genest and Rivest (1993) that the Kendall distribution function of V is $K_{\rho}(v) = v - \frac{\varphi(v)}{\varphi'(v)}$ and also its density function is $k(v) = \frac{\varphi(v)\varphi''(v)}{\varphi'(v)^2}$ for Archimedean copula models Chen (2012).

CHAPTER THREE
PARAMETER ESTIMATION AND GOODNESS-OF-FIT TESTING

3.1 Parameter Estimation

The *maximum likelihood estimator* (MLE) method is applied for estimating parameters; because the importance of the maximum likelihood estimator method is that the maximum likelihood estimaters have many 'good' properties (*consistency, asymptotic normality and asymptotic efficiency*). In this thesis, we will handle MLE based estimation methods (Exact Maximum Likelihood, Inference Functions for Margins, Canonical Maximum Likelihood) for copula estimation. Firstly, *Exact Maximum Likelihood* (EML) is known as 'classical' MLE procedure. Thus, it estimates jointly the parameter for the margins and parameter(s) defining the dependence structure, the copula parameter(s). When applying EML, it is not needed that margins are separated from the dependence structure. Using the EML method has two disadvantages. One of them is that this method will become computationally intensive for higher-dimensional distributions (Bouyè al. 2000). The second disadvantage is that the copula parameters are estimated under distributional assumptions for the margins.

A second approach is *Inference Functions for Margins* (IFM) method. When applying IFM, it is required to know the copula theory of separation of the margins and the dependence structure. It divides the parameters into certain parameters for the marginal distributions and common parameters for the dependence structure (the copula parameters). The estimation of the parameters of the univariate marginal distributions is the first step. In the second step, we maximize the likelihood function to find the copula parameter estimates when using the results of the parameter estimation. Quality of the resulting copula estimates depend upon the quality of the assumptions that were made about the margins.

The third approach is the *Canonical Maximum Likelihood* (CML). This method also uses a separation of margins and dependence structure in copula. IFM and EML

have the disadvantage that we estimate the parameters of the copula under distributional assumptions for the margins. The estimates for the copula parameters are margin-dependent. It is better to use the observed margins in the estimation procedure. Thus, it can be done by transforming the data into uniform variates using the empirical distributions for all margins (Genest et al. 1995). The misspecification of the margins causes to obtain biased estimates for the EML and IFM method.

3.1.1 Exact Maximum Likelihood (EML)

Suppose that a sample is $\left\{x_1^t,...,x_n^t\right\}_{t=1}^{T}$ having n-dimension over T periods and also the joint distributions of these variables is defined as the copula $C\left(F_1\left(x_1^t\right),...,F_n\left(x_n^t\right)\right)$ where $F_1,...,F_n$ are the marginal distributions. The expression for the log-likelihood function is

$$\ell(\theta)=\sum_{t=1}^{T}\mathrm{Inc}\left(F_1\left(x_1^t\right),...,F_n\left(x_n^t\right)\right)+\sum_{t=1}^{T}\sum_{i=1}^{n}\mathrm{Inf}_i\left(x_i^t\right) \qquad (3.1)$$

with θ a $k\times 1$ vector of parameters, consisting of both the margin and the copula parameters. By maximizing the log-likelihood according to θ, we obtain the ML estimator ($\hat{\theta}_{ML}$). For this estimator it holds that:

$$\sqrt{T}\left(\hat{\theta}_{ML}-\theta_0\right)-N\left(0,J^{-1}\left(\theta_0\right)\right) \qquad (3.2)$$

with $J(\theta_0)$ the Fisher information matrix.

3.1.2 Inference Functions for Margins (IFM)

The IFM method uses idea of separation of margins from the dependence structure. It is obtained as the log-likelihood as:

$$\ell(\theta) = \sum_{t=1}^{T} \mathrm{Inc}\left(\mathrm{F}_1\left(\mathrm{x}_1^t;\theta_1\right),...,\mathrm{F}_n\left(\mathrm{x}_n^t;\theta_n\right);\alpha\right) + \sum_{t=1}^{T}\sum_{i=1}^{n} \mathrm{Inf}_i\left(\alpha_i^t;\theta_i\right) \qquad (3.3)$$

with $\theta = (\theta_1,...,\theta_n,\alpha)$. The estimation of θ has two steps:

- First step; the estimation of the parameters in the univariate margins:

$$\hat{\theta}_i = \arg\max_{\theta_i} \ell^i(\theta_i) := \arg\max_{\theta_i} \sum_{t=1}^{T} \mathrm{Inf}_i\left(\mathrm{x}_i^t;\theta_i\right) \quad \text{for} \quad i = 1,...,n. \qquad (3.4)$$

- Second step; the estimation of the copula parameter, using the estimates obtained for the margins:

$$\hat{\alpha} = \arg\max_{\alpha} \ell^C(\theta) := \arg\max_{\alpha} \sum_{t=1}^{T} \mathrm{Inc}\left(\mathrm{F}_1\left(\mathrm{x}_1^t;\hat{\theta}_1\right),...,\mathrm{F}_n\left(\mathrm{x}_n^t;\hat{\theta}_n\right);\alpha\right) \qquad (3.5)$$

The IFM estimator is defined as:

$$\hat{\theta}_{\mathrm{IFM}} = \left(\hat{\theta}_1,...,\hat{\theta}_n,\hat{\alpha}\right) \qquad (3.6)$$

The IFM estimator verifies the property of asymptotic normality. So we can show as

$$\sqrt{T}\left(\hat{\theta}_{\mathrm{IFM}} - \theta_0\right) \to \mathrm{N}\left(0, \mathrm{V}^{-1}(\theta_0)\right) \qquad (3.7)$$

where $\mathrm{V}(\theta_0) = \left(\mathrm{E}\left[\frac{\partial}{\partial\theta}\mathrm{g}(\theta)^{\mathrm{T}}\right]\right)^{-1}\left(\mathrm{E}\left[\mathrm{g}(\theta)^{\mathrm{T}}\,\mathrm{g}(\theta)\right]\right)\left(\left(\mathrm{E}\left[\frac{\partial}{\partial\theta}\mathrm{g}(\theta)^{\mathrm{T}}\right]\right)^{-1}\right)^{\mathrm{T}}$

3.1.3 Canonical Maximum Likelihood (CML)

In the IFM method, the estimation θ depends on the distributional assumptions that we apply about the margins. First step; we transform the data $\left(x_1^t,...,x_n^t\right)$ into the uniform variables $\left(\hat{u}_1^t,...,\hat{u}_n^t\right)$ using the empirical distribution functions of the univariate margins. Second step; we estimate the copula parameters α by maximum likelihood estimation

$$\hat{\alpha} = \arg\max \sum_{t=1}^{T} \mathrm{Inc}\left(\hat{u}_1^t,...,\hat{u}_n^t;\alpha\right) \tag{3.8}$$

Using the transformation of the empirical distribution function, data is transformed into the uniform variables. $F_1^*,...,F_n^*$ is defined as the empirical distribution function:

$$F_j^*\left(x\right) = \frac{1}{T}\sum_{t=1}^{T} I\left(X_j^t \le x\right) \tag{3.9}$$

where $I(.)$ is the indicator function. In the estimation procedure, when some of u_i are equal to 1, we use $\frac{T}{T+1}F_j^*$, to prevent possible problems with unboundedness of the copula density. Also, the transformation of the data into the standard uniform random variables is implemented by:

$$\hat{u}_j^t = \frac{T}{T+1}F_j^T\left(x_j^t\right) \quad \text{for} \quad j=1,...,n. \tag{3.10}$$

The resulting estimator α is also called a *Pseudo-Likelihood Estimator* (Genest et al. 1995).

The difference between IFM and CML is that

- IFM estimates the marginal distributions under distributional assumptions and also transform the data into uniform variables.
- CML directly transforms the data into uniform variables using the empirical distribution of the data.

It can be shown that (under certain regularity conditions):

$$\sqrt{\mathrm{T}}\left(\hat{\theta}_{\mathrm{CML}}-\theta_0\right)\rightarrow \mathrm{N}\left(0,\sigma^2\right) \tag{3.11}$$

where θ_0 is the true copula parameter (under the assumption that this specific copula conducts the data generating process) and σ is not dependent upon T (Wulp, 2003).

3.2 Goodness-of-Fit Test

When a certain model has been determined and estimated, it has to be checked whether the certain model assumption is realistic. Thus researchers are faced with the notion of the goodness-of-fit problem (Fermanian, 2013). Some authors have examined the goodness-of-fit tests for copulas, but there is not overwhelmingly recommended method for checking whether a parametric family of copulas can appropriately describe the dependence structure for the data. Some researchers applied tests based on the probability integral transform presented by Rosenblatt (1952). Fermanian (2005) and Fermanian and Scaillet (2005) proposed goodness-of-fit tests based on the empirical copula process. Scailet (2005) proposed a non-parametric goodness-of-fit test based on a kernel smoothed estimator for the copula density.

Survival times are important in many study fields but such data are complex for modelling and analyzing. Because the problem of dependent censoring occurs and

subsequent survival times when the duration of follow-up is limited and the times for a certain individual are not independent. Moreover, for second and later times, the marginal survival distributions are non-identified, since they are observable only if previous survival times for an individual are uncensored. Also, in some studies, a significant proportion of individuals may never have the first event. As summary, survival times have more than one failure time that is observed in paralleled or simultaneously and do not satisfy any order restrictions. In the past years, Marshall and Olkin (1967) presented the common shock model in which the occurrence of a shock or disaster causes dependencies between independent lifetimes. For using copulas, many of the empirical applications were in the survival analysis field. One of the examples was Clayton (1978) who, in his study of joint lifetimes of fathers and sons, derived the joint survival function with joint frailty defined by the gamma distribution. Dependence of lifetimes on pairs of individuals (for instance, father and son, husband and wife, twins) is known. Oakes (1982) suggested similar results in a random-effects setting, and so joint survival models of this form were called as Clayton-Oakes models. Clayton and Cuzick (1985) suggested an EM algorithm to estimate these models. In a related study, Hougaard (1986b) modeled joint survival times of groups of rats inflicted with tumors. In the recent research, frailty models examined joint survival times of family members, especially interested devoted in to annuity valuation. Frees et al. (1996) deal with a joint and last-survivor annuity contracts using Frank's copula. Wang (2003) searched survival times of bone marrow and heart transplant patients. Hougaard et al. (1992) obtained data from Danish twins born between 1881 and 1930. The researchers analyzed joint survival data assuming Weibull conditionals and the assumption that frailty follows a positive stable distribution.

The methods for fitting copula models to data are developed, but when the lifetimes are put to censoring, there are a few applications on tests of fit for survival copulas. Especially, most of the proposed procedures do not examine the censored data and those that do suffer from limitations.

We examine the Archimedean method proposed for evaluating the fit of a copula in this section. We will see that the Archimedean method deals with censored bivariate lifetime data for the special case of the Clayton, Gumbel and Frank copula.

Definition 3.1 (Cross-Ratio Function): A *cross-ratio function* is important in terms of the characterization results and inferential procedures and diagnostic plots that are defined and shown by real examples since a cross-ratio function estimated from censored data. For the cross-ratio function, we benefit from measure *local dependence*. The *local odds ratio* function was proposed by Oakes (1989) and its definition;

$$\theta^{*}(x,y)=\frac{\partial^{2}P(X>x,Y>y)/\partial x\partial y P(X>x,Y>Y)}{\partial P(X>x,Y>y)/\partial x\partial P(X>x,Y>y)/\partial y}$$
$$=\frac{P(\Delta_{ij}=1\,|\,\tilde{X}_{ij}>x,\tilde{Y}_{ij}>y)}{P(\Delta_{ij}=0\,|\,\tilde{X}_{ij}>x,\tilde{Y}_{ij}>y)} \qquad (3.12)$$

where $\tilde{X}_{ij}=X_i\wedge X_j$ and $\tilde{Y}_{ij}=Y_i\wedge Y_j$, and $a\wedge b=\min(a,b)$ (Emura, et. al. 2011). If the pair (X,Y) is independent at (x,y) then $\theta^{*}(x,y)$ is a bivariate function measuring local dependence and equal 1. And also it is indicated that $\theta^{*}(x,y)=\theta_{\alpha}\{S(x,y)\}$ for the Archimedean copula class. $\theta_{\alpha}\{S(x,y)\}$ is the univariate function and it can be obtained as

$$\theta_{\alpha}(v)=-v\left\{\frac{\left[\phi_{\alpha}^{''}(v)\right]}{\left[\phi_{\alpha}^{'}(v)\right]}\right\}, \qquad (3.13)$$

Genest and Rivest (1993) indicated that $K(\nu)$ is related to $\phi_\alpha(\nu)$ through the differential equation

$$\lambda(v) = \nu - K(\nu) = \frac{\phi_\alpha(\nu)}{\phi'_\alpha(\nu)} \tag{3.14}$$

where $\phi'_\alpha(v) = \partial\phi_\alpha(v)/\partial v$ and $\nu = S(X, Y)$ (Wang and Wels, 2000). And also it is showed that the function $\phi_\alpha(v)$ can be obtained by the univariate function $K(v) \equiv P\{F(X, Y) \le v\}$ where $K(v)$ and $\lambda(v)$ are related to the generator function $\phi_\alpha(v)$ and they determine the dependence structure of the Archimedean copula class.

In the equation (3.12), $\Delta_{ij} = 1$ if (X_i, X_j) and (Y_i, Y_j) are concordant and $\Delta_{ij} = 0$ if (X_i, X_j) and (Y_i, Y_j) are discordant.

For a selected Archimedean copula model, semiparametric inference of the association parameter α has been an important research issue in the literature. Although the marginal distribution is not determined, the form of $\phi(.)$ is assumed. So, how to select an appropriate $\phi(.)$ with the data at hand is an important topic for applications. Oakes (1989) derived the analytic properties of Archimedean copula models that are useful for model selection. It is shown that

$$E\left[\Delta_{ij} \middle| \tilde{X}_{ij} = x, \tilde{Y}_{ij} = y\right] = \frac{\theta_\alpha\{S(x, y)\}}{\theta_\alpha\{S(x, y)\} + 1}, \tag{3.15}$$

where $\tilde{X}_{ij} = \min\left(X_i, X_j\right)$ and $\tilde{Y}_{ij} = \min\left(Y_i, Y_j\right)$. Moreover, Oakes (1989) found that a conditional version of Kendall's tau

$$\tau\left(t_1, t_2\right) = P\left(\left(T_{1i} - T_{1j}\right)\left(T_{2i} - T_{2j}\right) > 0 \middle| \min\left(T_{1i}, T_{1j}\right) = t_1, \min\left(T_{2i}, T_{2j}\right) = t_2\right)$$
$$-P\left(\left(T_{1i} - T_{1j}\right)\left(T_{2i} - T_{2j}\right) < 0 \middle| \min\left(T_{1i}, T_{1j}\right) = t_1, \min\left(T_{2i}, T_{2j}\right) = t_2\right) \quad (3.16)$$

is a function of $\theta\left(S\left(t_1, t_2\right)\right)$ such that

$$\tau\left(t_1, t_2\right) = \frac{\theta\left(S\left(t_1, t_2\right)\right) - 1}{\theta\left(S\left(t_1, t_2\right)\right) + 1} \quad (3.17)$$

where

$$\theta\left(S\left(t_1, t_2\right)\right) = \frac{P\left(\left(T_{1i} - T_{1j}\right)\left(T_{2i} - T_{2j}\right) > 0 \middle| \min\left(T_{1i}, T_{1j}\right) = t_1, \min\left(T_{2i}, T_{2j}\right) = t_2\right)}{P\left(\left(T_{1i} - T_{1j}\right)\left(T_{2i} - T_{2j}\right) < 0 \middle| \min\left(T_{1i}, T_{1j}\right) = t_1, \min\left(T_{2i}, T_{2j}\right) = t_2\right)} \quad (3.18)$$

When the bivariate survival function is unknown, Oakes (1989) conditioned on the size R_{ij} of the corresponding bivariate risk set $\left\{k : T_{1k} \geq \min\left(T_{1i}, T_{1j}\right), T_{2k} \geq \min\left(T_{2i}, T_{2j}\right)\right\}$ in equation (3.18) and obtained

$$\gamma\left(r\right) = \frac{P\left(\left(T_{1i} - T_{1j}\right)\left(T_{2i} - T_{2j}\right) > 0 \middle| R_{ij} = r\right)}{P\left(\left(T_{1i} - T_{1j}\right)\left(T_{2i} - T_{2j}\right) < 0 \middle| R_{ij} = r\right)} \quad (3.19)$$

for $r = 2, ..., n$. Oakes indicated that equation 3.19 is closed to equation 3.18. So, Oakes presented a graphical Archimedean copula selection procedure by plotting

$-1/\log(r/n)$ versus $\gamma(r)$ that is achieved by inserting the estimate of the unknown parameter found by on Kendall's tau (Yilmaz, 2009).

3.2.1 The Goodness-of-Fit Test Procedure for the Survival Copula

In this section, we deal with *uncensored data* firstly, and then we modify the model for *the right censored data* Emura et al. (2010). The main hypothesis is given by

$$H_0 : C(u,v) = \phi_\alpha^{-1}\left[\phi_\alpha(u) + \phi_\alpha(v)\right] \text{ for some } \alpha \in \mathbf{R},$$

where the alternative hypothesis is any other copula.

3.2.2 The GOF procedure for the uncensored data

Let $\{(X_i, Y_i); (i = 1,.....,n)\}$ be the uncensored data, and Δ_{ij} be the *concordance indicator*, then

$$U_k(\alpha) = \sum_{i<j} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij})\left[\Delta_{ij} - E\left[\Delta_{ij} \middle| \tilde{X}_{ij} = x, \tilde{Y}_{ij} = y\right]\right] \tag{3.20}$$

$$U_k(\alpha) = \sum_{i<j} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij})\left[\Delta_{ij} - \frac{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1}\right] \tag{3.21}$$

where $W_k(\tilde{x}_{ij}, \tilde{y}_{ij})$ is *weight function* and $\hat{S}(x,y)$ is an estimator of $S(x,y)$. $\hat{S}(x,y)$ is called empirical survival function and it can be expressed as $\hat{S}(x,y) = \frac{1}{n}\sum_i I(X_i \geq x, Y_i \geq y)$. In this GoF procedure, two weight functions are used

to obtain $\hat{\alpha}_k$ solve $U_k(\alpha)=0$ for $k=1,2$. When H_0 is true, $n^{1/2}\{\hat{\alpha}_1-\hat{\alpha}_2\}$ converges to a mean zero normal distribution. The power of the test depends to choose the weight functions (Emura et al., 2010).

3.2.3 The weight functions

Shih (1998) compared *the unweighted* and *weighted* concordance estimators of the association parameter and indicated that if the proposed model fits, the difference of these two estimates converges to zero. Using the idea of the likelihood approach of Clayton (1978), the estimating function can be written in terms of $U_k(\alpha)$. We can define the set of grid points as follows,

$$\psi=\left\{(x,y)\left|\sum_{i=1}^{n} I(X_i=x,Y_i\geq y)=1,\sum_{i=1}^{n} I(X_i\geq x,Yi=y)=1\right.\right\} \qquad (3.22)$$

Also let $D(x,y)=\sum_i I(X_i=x,Y_i=y)$ be the number of observed failures at (x,y), and $R(x,y)=r=\sum_{i=1} I(X_i\geq x,Y_i\geq y)$ measures the number of risk at $(x,y)\in\psi$. Then, $D(x,y)$ is distributed as *Bernoulli* with the success probability

$$P\left\{D(x,y)=1\middle|R(x,y)=r,(x,y)\in\psi\right\}=\frac{\theta_\alpha\{S(x,y)\}}{r-1+\theta_\alpha\{S(x,y)\}} \qquad (3.23)$$

By using conditional probability, the likelihood function can be obtained as

$$L(\alpha)=\prod_{(x,y)\in\psi}\left[\frac{\theta_\alpha\{S(x,y)\}}{R(x,y)-1+\theta_\alpha\{S(x,y)\}}\right]^{D(x,y)}\times\left[\frac{R(x,y)-1}{R(x,y)-1+\theta_\alpha\{S(x,y)\}}\right]^{1-D(x,y)}$$

(3.24)

.

with all points in ψ under the independence assumption among the grids.

Then, *the estimating function* is

$$U_1(\alpha)=\sum_{i<j}\frac{\dot\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}\left[\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}+1\right]}{\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}\left[R_{ij}-1+\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}\right]}\left[\Delta_{ij}-\frac{\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}}{\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}+1}\right] \quad (3.25)$$

$$=\sum_{i<j}W_1(\tilde X_{ij},\tilde Y_{ij})\left[\Delta_{ij}-\frac{\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}}{\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}+1}\right], \quad (3.26)$$

where $R_{ij}=R(\tilde X_{ij},\tilde Y_{ij})=n\hat S(\tilde X_{ij},\tilde Y_{ij})$ and $\dot\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}=\partial\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}/\partial\alpha$. We show that *the second estimating function* is

$$U_2(\alpha)=\sum_{i<j}\left[\Delta_{ij}-\frac{\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}}{\theta_\alpha\{\hat S(\tilde X_{ij},\tilde Y_{ij})\}+1}\right]. \quad (3.27)$$

By solving $U_1(\alpha)=0$ and $U_2(\alpha)=0$, we find $\hat\alpha_k$ for $k=1,2$. Frank copula is defined as $\phi_\alpha(v)=\log\{(1-\alpha^{-1})/(1-\alpha^{-v})\}$ and with $\theta_\alpha\{S(x,y)\}=S(x,y)\log(\alpha)/\left(1-e^{-S(x,y)\log(\alpha)}\right)$, then

$$W_1(\tilde{X}_{ij},\tilde{Y}_{ij})=\frac{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})e^{\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}(e^{\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}-\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha-1)/(e^{\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}-1)^2)(((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))+1)}{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})} \tag{3.28}$$

Therefore,

$$U_1(\alpha)=\sum_{i<j}\frac{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})e^{\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}(e^{\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}-\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha-1)/(e^{\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}-1)^2)(((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))+1)}{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})}$$

$$\times\left[\Delta_{ij}-\frac{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))}{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\tilde{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))+1}\right] \tag{3.29}$$

and,

$$U_2(\alpha)=\sum_{i<j}\left[\Delta_{ij}-\frac{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))}{((\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-e^{-\hat{\tilde{S}}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha}))+1}\right] \tag{3.30}$$

3.2.4 Asymptotic distribution theory

Many important test statistics can be written with *U-statistics*. So, in the asymptotic distribution theory, the concordance probability estimate is derived by using the U-statistic theory. Firstly, U-statistics for independent identically distributed random variables comes from Hoeffding (1948). Then, asymptotics for U-statistics are produced from the dependent sample. So, there are many asymptotic results for U-statistics based on weakly dependent random processes (Klicnarová, 2017).

The derivative of *the log pseudolikeihood* proposed a possible proof of the asymptotic normality of $U(\alpha)$. If it is indicated that $U(\alpha)-U'(\alpha)$ (where $U'(\alpha)$

is derivative of $U(\alpha)$) is asymptotically trivial, $U'(\alpha)$ has finite variance, then the asymptotic normality of $U(\alpha)$ would be observed immediately (Oakes, 1986).

We obtain the weighted and unweighted function approaches to the true parameter value of Frank copula as

$$U_1(\alpha)=\sum_{i<j}\frac{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})/(1-\exp(-\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha))-\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})^2\alpha/(1-\exp(-\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha))^2\exp(-\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})a)\left[(\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-\exp(-\alpha\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})))+1\right]}{\left[(\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-\exp(-\alpha\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})))\right]\left[R_{ij}-1+\left[(\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-\exp(-\alpha\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})))\right]\right]}$$
$$\times\left[\Delta_{ij}-\frac{\left[(\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-\exp(-\alpha\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})))\right]}{\left[(\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-\exp(-\alpha\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})))+1\right]}\right] \tag{3.31}$$

$$U_2(\alpha)=\sum_{i<j}\left[\Delta_{ij}-\frac{\left[(\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-\exp(-\alpha\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})))\right]}{\left[(\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\alpha)/(1-\exp(-\alpha\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})))+1\right]}\right] \tag{3.32}$$

We obtain the weighted and unweighted function approaches to the true parameter value of Gumbel copula as

$$U_1(\alpha)=\sum_{i<j}\frac{2\log\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-\alpha}{\left\{\log\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-\alpha\right\}\left\{\alpha-R_{ij}\log\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}}\left[\Delta_{ij}-\frac{\log\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-\alpha}{2\log\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-\alpha}\right] \tag{3.33}$$

$$U_2(\alpha)=\sum_{i<j}\left[\Delta_{ij}-\frac{\log\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-\alpha}{2\log\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-\alpha}\right] \tag{3.34}$$

We obtain the weighted and unweighted function approaches to the true parameter value of Clayton copula as

$$U_1(\alpha)=\sum_{i<j}\frac{\left[(\alpha+1)+1\right]}{(\alpha+1)\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})}\left[\Delta_{ij}-\frac{(\alpha+1)}{\left[(\alpha+1)+1\right]}\right] \tag{3.35}$$

$$U_2(\alpha)=\sum_{i<j}\left[\Delta_{ij}-\frac{(\alpha+1)}{\left[(\alpha+1)+1\right]}\right] \tag{3.36}$$

For, Frank copula, the asymptotic distribution of $n^{1/2}(\hat{\alpha}_1-\hat{\alpha}_2)$ and $n^{1/2}(\log\hat{\alpha}_1-\log\hat{\alpha}_2)$ approximate to a normal distribution with mean zero and variance $\sigma^2=4E\left[h\{(X_1,Y_1),(X_2,Y_2)\}h\{(X_1,Y_1),(X_3,Y_3)\}\right]$. And also $(\log\hat{\alpha}_1-\log\hat{\alpha}_2)$ converges zero in probability, so that $U_1(\alpha)$ and $U_2(\alpha)$ involve the estimator. However, this causes difficulty in technical. We hold U –statistics as

$$\tilde{U}_1(\alpha)=\sum_{i<j}\frac{\dot{\theta}_\alpha\left\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}\left[\theta_\alpha\left\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}+1\right]}{\theta_\alpha\left\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}S(\tilde{X}_{ij},\tilde{Y}_{ij})}\left[\Delta_{ij-}\frac{\theta_\alpha\left\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}}{\theta_\alpha\left\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}+1}\right], \tag{3.37}$$

and

$$\tilde{U}_2(\alpha)=\sum_{i<j}\Delta_{ij}-\frac{\theta_\alpha\left\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}}{\theta_\alpha\left\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}+1}. \tag{3.38}$$

Lemma 3.1: Under the correct model and regularity conditions

$$n\binom{n}{2}^{-1} U_1(\alpha) = \binom{n}{2}^{-1} \tilde{U}_1(\alpha) + O_p(1), \qquad \binom{n}{2}^{-1} U_2(\alpha) = \binom{n}{2}^{-1} \tilde{U}_2(\alpha) + O_p(1),$$

where $O_p(1)$ is uniform in α.

If the parameter α is positive, natural logarithm transformation can recuperate to normal approximation (Emura et al., 2010).

Lemma 3.2: Under the correct model and regularity conditions. When $\gamma = \log\alpha$ and $\hat{\gamma}_k = \log\hat{\alpha}_k$, then

$$n^{1/2}(\hat{\gamma}_1 - \hat{\gamma}_2) = n^{1/2}\binom{n}{2}^{-1} \sum_{i<j} h\left\{(X_i, Y_i), (X_j, Y_j)\right\} + O_p(1), \tag{3.39}$$

where the function h is symmetric and

$$h\left\{(X_i, Y_i), (X_j, Y_j)\right\} \equiv \frac{1}{\alpha}\left(\frac{\dot{\theta}_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\}\left[\theta_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\} + 1\right]}{A_L\theta_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\}S(\tilde{X}_{ij}, \tilde{Y}_{ij})} - \frac{1}{A}\right)\left[\Delta_{ij} - \frac{\theta_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\}}{\theta_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\} + 1}\right] \tag{3.40}$$

where $A \equiv E\left(\frac{\dot{\theta}\left\{S(\tilde{X}_{12}, \tilde{Y}_{12})\right\}}{\left[\theta_\alpha\left\{S(\tilde{X}_{12}, \tilde{Y}_{12})\right\}+1\right]^2} \right)$ and

$$A_L \equiv E\left(\frac{\dot{\theta}_\alpha\left[\left\{S(\tilde{X}_{12}, \tilde{Y}_{12})\right\}\right]^2}{\theta_\alpha\left\{S(\tilde{X}_{12}, \tilde{Y}_{12})\right\}\left[\theta_\alpha\left\{S(\tilde{X}_{12}, \tilde{Y}_{12})\right\}+1\right]} \right)$$

3.2.5 The testing procedure

When $|\hat{\gamma}_1 - \hat{\gamma}_2|/\hat{\sigma}$ is greater than $Z_{1-\alpha/2}$ for α- level test and $Z_{1-\alpha/2}$ is the p^{th} percentile of the standard normal distribution. In Archimedean copula models, derivation of the variance estimator is difficult and also this formula becomes complex for right-censored data. Therefore, Emura et al. (2010) proposed the *jackknife method* for the variance estimation. The jackknife method is defined as

$$\sigma^2_{jack} = \frac{n-1}{n}\sum_{i=1}^{n}\left\{\hat{\gamma}_1^{(-i)} - \hat{\gamma}_2^{(-i)} - (\hat{\gamma}_1^{(.)} - \hat{\gamma}_2^{(.)})\right\}^2, \qquad (3.41)$$

where $\hat{\gamma}^{(-i)}$ is the estimator after deleting the ith observation and $\hat{\gamma}_1^{(.)} - \hat{\gamma}_2^{(.)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_1^{(-i)} - \hat{\gamma}_2^{(-i)}$

3.2.6 In the case of right-censoring

Let (A_i, B_i) be the independently identically distributed bivariate censoring variables. (A_i, B_i) and (X_i, Y_i) are independent of each other. The goodness-of-fit test for right-censored bivariate data can be set up as follows.

1. Estimation step: Brown, Holander and Korwar (1974) proposed a non-parametric estimator for censored data. It can be applied to this approach.

Also, Shih and Louis (1995) proposed a semiparametric estimator for θ. Similarly, it can be estimated with this approach. It can be estimated the marginal survivor functions by Kaplan-Meier estimates and the joint survivor function by the Dabrowska estimator Dabrowska (1988). As basic, it is used $\left(\hat{K}_i, \hat{L}_i\right)$ to denote the imputed or estimated bivariate random vector corresponding to $\left(K_i, L_i\right)$

2. Data imputation step: For imputation, it is replaced the unknown parameter and survivor functions with the corresponding estimators obtained from the estimation step.

   a) If $\left(X_i, Y_i\right)$ is double censored, namely $X_i > a_i$, $Y_i > b_i$ for some observed pair $\left(A_i, B_i\right) = \left(a_i, b_i\right)$, we generate $\hat{L}_i = l$ from the distribution of $\left(L \middle| X > a_i, Y > b_i\right)$ using the inverse CDF method. We generate $\hat{K}_i$ from the distribution of $\left(K \middle| L = l, X > a_i, Y > b_i\right)$. $\left(K \middle| L = l, X > a_i, Y > b_i\right)$ is uniform distribution on $[0,1]$ and independent of L. It can be seen that we can generate a random pair $\left(\hat{K}_i, \hat{L}_i\right)$ that follows $\left\{(K, L) \middle| X > a_i, Y > b_i\right\}$ distribution asymptotically.

   b) If $\left(X_i, Y_i\right)$ is single censored, i.e., $X_i > a_i$, $Y_i = b_i$, we generate $\hat{L}_i$ from the distribution of $\left(L \middle| X > a_i, Y = b_i\right)$ to obtain $\hat{L}_i = l$ and estimate K_i by

$$\hat{K}_i = \frac{\varphi_{\hat{\theta}}(l) - \varphi_{\hat{\theta}}\left\{\hat{S}_2\left(y_i\right)\right\}}{\varphi_{\hat{\theta}}(l)} \tag{3.42}$$

c) If (X_i, Y_i) is single censored and $X_i = x_i$, $Y_i > b_i$, we generate $\hat{L}_i$ from the distribution of $\left(L \middle| X = x_i, Y > b_i\right)$ to obtain $\hat{L}_i = 1$ and estimate K_i by

$$\hat{K}_i = \frac{\varphi_{\hat{\theta}}\left\{\hat{S}_1(x_i)\right\}}{\varphi_{\hat{\theta}}(1)} \tag{3.43}$$

d) If (X_i, Y_i) is uncensored, i.e., $X_i = x_i$, $Y_i = y_i$, we estimate K_i and L_i by

$$\hat{K}_i = \frac{\varphi_{\hat{\theta}}\left(\hat{S}_1(x_i)\right)}{\varphi_{\hat{\theta}}\left\{\hat{S}(x_i, y_i)\right\}} \quad \text{and} \quad \hat{L}_i = \hat{S}(x_i, y_i), \tag{3.44}$$

respectively.

It is repeated the above imputation procedures to generate m imputed data sets of size n.

3. Test step: For each complete data sample obtained from the previous data imputation step, we conduct the Z-test proposed in the uncensored case and combine the test results for $m(m \geq 3)$ imputed samples as described in Rubin (1987). The detailed procedure is from m complete data sets and m values for $\hat{Z}$, $\hat{Z}_1, ..., \hat{Z}_m$, calculate the combined test statistic

$$T_m = \frac{\bar{Z}_m}{\sqrt{L_m}} \tag{3.45}$$

where $\bar{Z}_m = \sum_{l=1}^{m} \hat{Z}_l / m$ and $L_m = (1 + 1/m) E_m + 1/n$, with $E_m = \sum_{l=1}^{m} \left(\hat{Z}_l - \bar{Z}_m\right)^2 \Big/ (m-1)$. The p-value related to T_m is

$P_m = 2\left[1 - D_d\left(|T_m|\right)\right]$, where D_d is the distribution function of a t-distributed random variable having $d = (m-1)\left[L_m \Big/ \left(1 + m^{-1}\right) E_m\right]^2$ degrees of freedom (Wang, 2010).

Through the right censoring process, it is applied that, $\breve{Y}_i = \min(Y_i, B_i)$, $\delta_i^x = I(X_i \le A_i)$ and $\delta_i^y = I(Y_i \le B_i)$. The order of X_i and X_j is known if and only if $\tilde{X}_{ij} \le \tilde{A}_{ij}$, where $\tilde{X}_{ij} = X_i \Lambda X_j$ and $\tilde{A}_{ij} = A_i \Lambda A_j$. Similarly, the order of Y_i and Y_j is iknown if and only if $\tilde{Y}_{ij} \le \tilde{B}_{ij}$. Let Z_{ij} indicates whether the ordering relationship is certain or not. Then, then U- statistic is

$$U_k(\alpha) = \sum_{i<j} Z_{ij} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, \widehat{S}) \left[\Delta_{ij} - \frac{\theta_\alpha\left\{\widehat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\}}{\theta_\alpha\left\{\widehat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\} + 1} \right] \quad k = (1,2) \ , \quad (3.46)$$

where $\widehat{S}(x, y)$ is an estimator of $S(x, y)$ and $\Delta_{ij} = I\left\{(\breve{X}_i - \breve{X}_j)(\breve{Y}_i - \breve{Y}_j) > 0\right\}$. The weight function formula is

$$W_1(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, S) = \frac{\dot{\theta}_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\}\left[\theta_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\} + 1\right]}{\theta_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\}\left[R_{ij} - 1 + \theta_\alpha\left\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\right\}\right]} \quad (3.47)$$

where $W_2(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, S) = 1$ and $R_{ij} = \sum_{i=1}^{n} I(\breve{X}_I \ge \tilde{X}_{ij}, \breve{Y}_I \ge \tilde{Y}_{ij})$.

To solve $\hat{\alpha}_k$, the following equation is used;

$$\sum_{i<j} Z_{ij} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, \tilde{S}^{(l-1)}) \left[\Delta_{ij} - \frac{\theta_\alpha \left\{ \tilde{S}^{(l-1)}(\tilde{X}_{ij}, \tilde{Y}_{ij}) \right\}}{\theta_\alpha \left\{ \tilde{S}^{(l-1)}(\tilde{X}_{ij}, \tilde{Y}_{ij}) \right\} + 1} \right] = 0 \qquad l = 1, 2, \ldots \quad (3.48)$$

where $\tilde{S}^{(l)}$ shows the estimated value in the l-step and also defined as $\tilde{S}_\alpha(x, y)$.

As in the censored case, the results of asymptotic normality are valid and

$$\tilde{U}_2(\alpha) = \sum_{i<j} Z_{ij} \left[\Delta_{ij} - \frac{\theta_\alpha \left\{ S(\tilde{X}_{ij}, \tilde{Y}_{ij}) \right\}}{\theta_\alpha \left\{ S(\tilde{X}_{ij}, \tilde{Y}_{ij}) \right\} + 1} \right], \quad (3.49)$$

U-statistic approximate to $U_k(\alpha)$ as in the above $U_2(\alpha)$ function.

$$\tilde{U}_2(\alpha) = \sum_{i<j} Z_{ij} \left[\Delta_{ij} - \frac{\theta_\alpha \left\{ S(\tilde{X}_{ij}, \tilde{Y}_{ij}) \right\}}{\theta_\alpha \left\{ S(\tilde{X}_{ij}, \tilde{Y}_{ij}) \right\} + 1} \right]. \quad (3.50)$$

CHAPTER FOUR
HAZARD SCENARIOS

4.1 Introduction to Hazard Scenarios

A "*Hazard*" is known as a situation or property with the potential to construct harm and also denotes the potential for an accident with undesired results in process safety. Some examples for hazard are toxic materials, high pressure, high- voltage power supply, etc. (Baybutt, 2003). The main aim of modelling the *hazard scenario* is to evaluate the maximum threat expected from a studied event in a certain area and to give some ideas about preventing and reducing the serious results on the environment and population, infrastructures (Paulatto, Pinat & Romanelli, 2007).

Hazard scenarios may be multivariate and dependent. In hazard scenarios, several scenarios are tendered and comparisons are applied. Firstly we will deal with some notation in order to explain the hazard scenario.

Definition 4.1 (Critical Layer): If an n-dimensional continuous distribution function is $F = C(F_1, ..., F_d)$ and F has strictly increased margins, *the critical layer* L_t^F of the level $t \in (0,1)$ is defined as;

$$L_t^F = \{x \in R^d : F(x) = t\}. \tag{4.1}$$

L_t^F is the *iso-hyper-surface* (with n-1 dimension) and F equals the constant value t: so, L_t^F is a *(iso)line* for bivariate distributions, a *(iso)surface* for trivariate ones, and so on. For any certain $x \in R^d$, there exists a unique critical layer L_t^F supporting x: namely, the one identified by the level $t = F(x)$. Note that, by Probability Integral Transformation, there exists a one-to-one correspondence between the two iso-hyper-surfaces $L_t^C = \{u \in I^d : C(u) = t\}$ (pertaining to C in I^d) and L_t^F (pertaining to F in R^d) (Salvadori, De Michele & Durante, 2011).

Definition 4.2 (Upper Set): $S \subseteq R^d$ is an *Upper Set* if, and only if, $x \in S$ and $y \geq x$ component-wise refer $y \in S$. The concept of Upper Set tackles with the intuitive (and practical) reasoning that, if an event is risky, "larger" realizations may be threatening.

Definition 4.3 (Hazard Scenario): Suppose that **X** is a random vector defining the event of interest. A hazard scenario of a level $\alpha \in (0,1)$ is any Upper Set that $S = S_\alpha \subseteq R^d$ relation

$$P(X \in S) = \alpha \tag{4.2}$$

holds. The dependence upon α will be eliminated when no complexity may occur to hold the formula simple. According to the proper criterion, a hazard scenario can be planned as a set of containing all the events x's reputed to be "dangerous". If $x \in S$ and $y \geq x$ component-wise, y could be thought of as a dangerous event.

Since a hazard scenario is an Upper Set, it is tackled with states that large values of the variables of interest are related to the dangerous terms. If small values of variables of interest (some variables X_j's) may be dangerous, it is changed the sign of the variables of interest or achieving other proper transformations as in AghaKouchak et al. (2014).

Given a realization $x \in R^d$, there are several ways related to x with a proper hazard scenario, denoted by $S_x \subseteq R^d$ for formula facility and clearness. By the Probability Integral Transformation, there is a unique realization $u = T_F(x)$ corresponding to x, and the unique region $S_u \subseteq I^d$ corresponding to S_x in I^d. Copula can help to calculate the level of S_u and S_x.

The type of events considered to be menacing, and their probabilities of occurrence, is given below

1. "OR" scenario S^{v}: For the realization of the event $\left\{X \in S_x^v\right\}$, it is sufficient one of the variables X_j's, with $i = 1,...,d,$ exceeds the corresponding threshold x_j. The figure of a bivariate *OR Hazard Scenario* $S_{u,v}^{v}$ is considered the pair $(U,V) \in I^2$.

2. "AND" scenario S^: For the realization of the event $\left\{X \in S_x^{\wedge}\right\}$ it is necessary all the variables X_j's, with $i = 1,...,d,$ exceed the corresponding thresholds x_i's. The figure of a bivariate *AND Hazard Scenario* $S_{u,v}^{\wedge}$, considered the pair $(U,V) \in I^2$.

3. "Kendall" scenario S^{K}: Suppose that $x \in R^d$ is a given event, with $t = F(x)$, and suppose that L_t is the level set crossing x. A bivariate *Kendall Hazard Scenario* is considered the pair $(U,V) \in I^2$. S_t^K is the region "exceeding" the critical layer L_t and L_t may represent a critical multivariate threshold.

4. "Survival Kendall" scenario $S^{\breve{K}}$: Suppose that $x \in R^d$ is a given event, with $t = \bar{F}(x)$, and suppose that $\bar{L}_t$ is the survival level set crossing x. A bivariate *Survival Kendall Hazard* Scenario is considered the pair $(U,V) \in I^2$. $S_t^{\breve{K}}$ is the region "exceeding" the *survival critical layer* $\bar{L}_t$ and $\bar{L}_t$ may represent a critical multivariate threshold.

5. **Structural scenario** S^{Ψ}. A *structure function* Ψ is determined interactions between the physical structure and loads acting in the set of dangerous events (and its shape). Also, Ψ is continuous for $\Psi : R^d \to R$ and $y \geq x$ component-wise refer $\Psi(y) \geq \Psi(x)$. If Ψ is a (highly) nonlinear function, the calculation is useless as analytical. So *Monte Carlo approach* is applied for the proper solutions (Straub 2014).

Table 4.1 gives the shapes (second column) and probabilities (third column) formulation for the basic hazard scenarios (OR, AND, Kendall and Survival Kendall Scenarios).

Table 4.1 Shapes and probabilities formulation for the basic Hazard Scenarios

Scenarios	**Shapes**	**Probabilities**
OR	$S_x^{v} = \bigcup_{j=1}^{d} \left(R \times ... \times \left(x_j, +\infty \right) \times ... \times R \right)$	$\alpha_x^{v} = P\left(X \in S_x^{v} \right) = 1 - C\left(F_1(x_1), ..., F_d(x_d) \right)$
AND	$S_x^{\wedge} = \bigcap_{j=1}^{d} \left(R \times ... \times \left(x_j + \infty \right) \times ... \times R \right)$	$\alpha_x^{\wedge} = P\left(X \in S_x^{\wedge} \right) = \hat{C}\left(\bar{F}_1(x_1), ..., \bar{F}_d(x_d) \right)$
Kendall	$S_t^{K} = \left\{ y \in R^d : F(y) > t \right\} = \left\{ y \in R^d : C\left(F_1(y_1), ..., F_d(y_d) \right) > t \right\}$	$\alpha_u^{K} = \alpha_x^{K} = \alpha_t^{K} = P\left(X \in S_t^{K} \right) = 1 - K(t)$
Survival Kendall	$S_t^{\breve{K}} = \left\{ y \in R^d : \bar{F}(y) < t \right\} = \left\{ y \in R^d : \hat{C}\left(\bar{F}_1(y_1), ..., \bar{F}_d(y_d) \right) < t \right\}$	$\alpha_u^{\breve{K}} = \alpha_x^{\breve{K}} = \alpha_t^{\breve{K}} = P\left(X \in S_t^{\breve{K}} \right) = 1 - \breve{K}(t) = \hat{K}(t)$
Structural	$S_z^{\Psi} = \left\{ x \in R^d : \psi(x) > z \right\}$	$\alpha_z^{\psi} = P\left(X \in S_z^{\psi} \right) = P\left(\Psi(X) > z \right)$

Figure 4.1 presents a bivariate illustration of the hazard scenarios (shaded areas) as different approaches, in the copula domain.

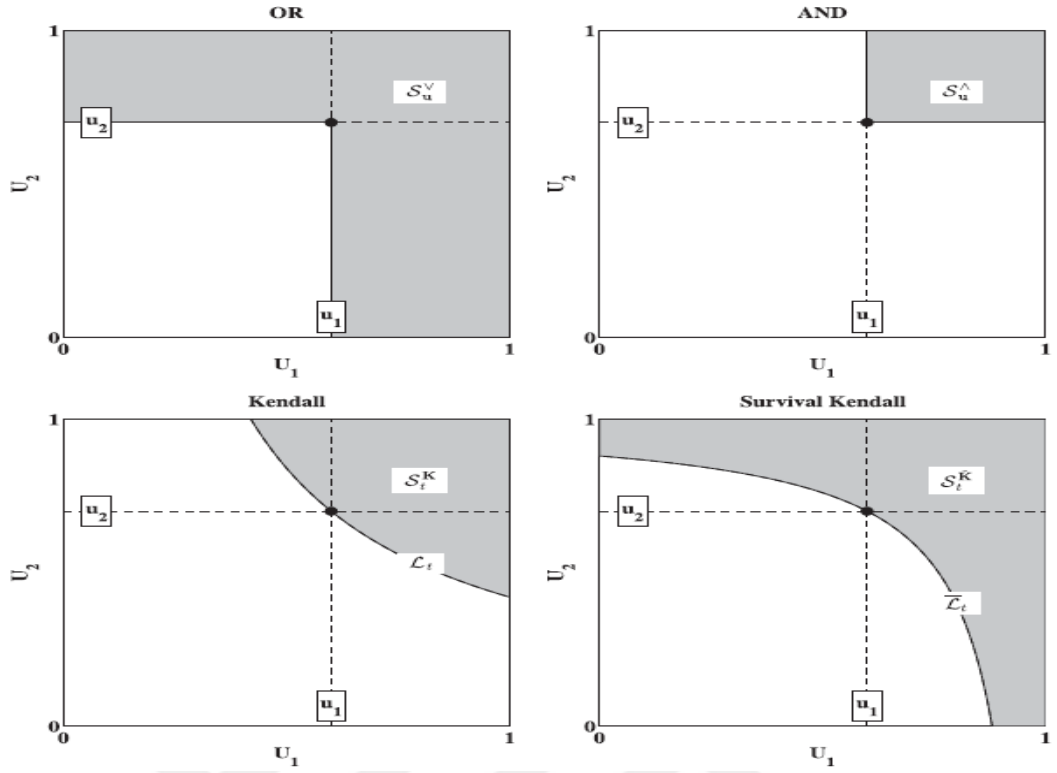

Figure 4.1 Bivariate illustration of the Hazard Scenarios examined (shaded areas) as different approaches, in the copula domain- the (top-left) OR, (top-right) AND, (bottom-left) Kendall, and (bottom-right) the Survival Kendall. The lower bound of the hazard region corresponds to the critical isoline L_t in the Kendall case and the lower bound of the hazard region corresponds to the critical isoline $\overline{L}_t$ in the Survival Kendall case (Salvadori, et al., 2016)

S_t^K and $S_t^{\check{K}}$ may overlap over a set with positive probability, i.e, $P\left(X \in S_t^K \bigcap S_t^{\check{K}}\right) > 0$ for certain $t \in I$, so the Kendall and Survival Kendall cases will be examined separately.

1. **The Kendall case** Suppose that $u = F(x_1, ..., x_d) \in I^d$ is fixed and suppose that L_t is the critical layer crossing u, where $t = C(u)$.

$$S_u^\wedge \subseteq S_t^K \subseteq S_u^v \tag{4.3}$$

and

$$\alpha_u^\wedge \leq \alpha_t^K \leq \alpha_u^v \tag{4.4}$$

α_u^v is constant as u crosses over L_t, as well as α_t^K, but $\alpha_u^\wedge$ usually changes. S_t^K is unique for all events having L_t, the hazard scenario S_u^v's restate and partially overlap.

2. The Survival Kendall case Suppose that $u = \bar{F}(x_1,...,x_d) \in I^d$ is fixed and suppose that $\bar{L}_t$ is the survival critical layer crossing u, where $t = \hat{C}(1-u)$.

$$S_u^\wedge \subseteq S_t^{\check{K}} \subseteq S_u^v \tag{4.5}$$

and

$$\alpha_u^\wedge \le \alpha_t^{\check{K}} \le \alpha_u^v \tag{4.6}$$

$\alpha_u^\wedge$ is constant as u crosses over $\bar{L}_t$, as well as $\alpha_t^{\check{K}}$, but α_u^v usually changes. $S_t^{\check{K}}$ is unique for all events having $\bar{L}_t$, the hazard scenario $S_u^\wedge$'s restate and partially overlap.

4.1.1 The Usage of Hazard Scenarios

The use of a certain hazard scenarios should be conducted as a certain combination of the most proper variables that cause accruing of the dangerous events. Firstly the shape of a hazard scenario has been decided, and then it is related to probability (viz., the level) that sets the degree of threatening. The use of AND or OR Scenarios is depended on what case will devastate the structure. When the OR approach can be applied when flood peak or flood volume exceeding a certain magnitude will cause damage. AND approach is applied, if the flood volume and flood peak must exceed a certain magnitude that will cause damage.

When we tackle OR and AND scenarios' shape equation in Table 4.1, it is determined a certain event x. Otherwise, Kendall and Survival Kendall Hazard Scenarios are defined with a critical layer of level t (namely, an infinity of events x's). For the OR case,

$$S_t^K = \bigcap_{u \in L_t} S_u^{\vee}, \tag{4.7}$$

it can be seen in Figure 4.1c, and for the AND case

$$S_t^{\check{K}} = \bigcup_{u \in \bar{L}_t} S_u^{\wedge}, \tag{4.8}$$

It can be seen in Figure 4.1d. So, Kendall and Survival Kendall Hazard Scenarios present to the intersection and the union of infinite OR and AND Hazard Scenarios. Kendall and Survival Kendall Hazard Scenarios are related to any of the events x's in exceed of the critical layer of level t defining the dangerous area.

For a certain level t, S_t^K and $S_t^{\check{K}}$ are shown all the threatening realizations characterizing the OR and AND cases. In the OR and AND cases, the shapes of the hazard regions are defined a priori and it is determined methods of the distribution function of X and its level curves by dangerous region. Moreover, Kendall and Survival Kendall Hazard Scenarios do not have a comment as structural and physical notions directly. Besides, Survival Kendall and Kendall Hazard Scenarios do not determine a priori which variables may present to the main resources of risk.

The Survival Kendall case indicates regions of bounded safe events, so ensuring that nondangerous events take on bounded values of all variables.

When we tackle the relation between Kendall Scenario and Survival Kendall Scenario for probabilities, we can see that any probability can be rewritten and calculate by And Scenario probability ($\alpha_u^{\wedge}$):

Firstly, if we hold AND Scenario probabilities:

$$\begin{aligned}\alpha^{\vee}_{u_1,u_2} &= 1-C(u_1,u_2)\\ &= 1-\left[u_1+u_2-1+\hat{C}(1-u_1,1-u_2)\right]\\ &= 2-\left(u_1+u_2+\hat{\alpha}_{u_1,u_2}\right)\end{aligned} \tag{4.9}$$

For Kendall Scenario probabilities

$$\begin{aligned}\alpha^{K}_{u_1,u_2} &= 1-K(t)\\ &= 1-K\left(C(u_1,u_2)\right)\\ &= 1-K\left(u_1+u_2-1+\alpha^{\wedge}_{u_1,u_2}\right)\end{aligned} \tag{4.10}$$

For Survival Scenario probabilities

$$\begin{aligned}\alpha^{\hat{K}}_{u_1,u_2} &= 1-\breve{K}(t)\\ &= 1-\breve{K}\left(\hat{C}(1-u_1,1-u_2)\right)\\ &= 1-\breve{K}\left(\alpha^{\wedge}_{u_1,u_2}\right)\end{aligned} \tag{4.11}$$

where $K(t)$ is Kendall distribution function (calculated as $K(t)=P\left(F(X_1,...,X_d)\le t\right)=P\left(C\left(F_1(X_1),...,F_d(X_d)\right)\le t\right)$) and also $\hat{K}(t)$ is *upper-orthant Kendall distribution function* (calculated as $\hat{K}(t)=P\left(\bar{F}(X_1,...,X_d)\le t\right)=P\left(\hat{C}\left(\bar{F}_1(X_1),...,\bar{F}_d(X_d)\right)\le t\right)$), $\breve{K}(t)$ is Survival Kendall distribution function (calculated as $\breve{K}(t)=1-\hat{K}(t)$).

Definition 4.4 (The Failure Probability Approach): This notion is provided on how to calculate the probability of observing a Hazard Scenario at least once in a certain design lifetime to us.

Suppose that $T>0$ is an arbitrary design lifetime for a given structure and also suppose that $X_1,...,X_T$ is the dimensional variate random vectors defining the event in an investigation at times $1,...,T$. Let $(S_1,...,S_T)$ be the hazard scenario vector and suppose that S_i^c are hazard scenario's complements namely their complements S_i^c's

could be labeled as “safe”. For *The Failure Probability Approach*, it can be denoted as a general formula;

$$\rho_T = 1 - P\left(X_1 \in S_1^c, ..., X_T \in S_T^c\right). \tag{4.12}$$

For example, considering hazard scenario is fixed and variables are independent identical distributed, then the formula is;

$$\rho_T = 1 - \prod_{j=1}^{T} P\left(X \in S^c\right) = 1 - \left(1-\alpha\right)^T \tag{4.13}$$

If X_i’s are independent and are not identically distributed and also hazard scenarios are changing in time then,

$$\rho_T = 1 - \prod_{i=1}^{T} P\left(X_i \in S_i^C\right) = 1 - \prod_{i=1}^{T} \left(1-\alpha_i\right), \tag{4.14}$$

where under assuming, ρ_T is calculated that the estimated joint distribution at the time i is given by a proper function F_i.

While basic hazard scenarios’ calculate the Failure Probabilities (Table 4.2), it is considered that for events sharing a joint multivariate critical threshold $\tilde{x}$, and independent and identically distributed X_i’s.

Table 4.2 The Failure probabilities for basic Hazard Scenarios

Hazard Scenarios	**The Failure Probabilities**
OR Scenario	$\rho_T^{\vee} = 1 - \left(C\left(F_1\left(\tilde{x}_1\right), ..., F_d\left(\tilde{x}_d\right)\right)\right)^T$
AND Scenario	$\rho_T^{\wedge} = 1 - \left(1 - \hat{C}\left(\bar{F}_1\left(\tilde{x}_1\right), ..., \bar{F}_d\left(\tilde{x}_d\right)\right)\right)^T$
Kendall Scenario	$\rho_T^{K} = 1 - K\left(F\left(\tilde{x}\right)\right)^T$
Survival Kendall Scenario	$\rho_T^{\breve{K}} = 1 - \breve{K}\left(1 - F\left(\tilde{x}\right)\right)^T$

For the Failure Probabilities related to different bivariate scenarios formulas:

$$\begin{aligned}\rho_{T,u}^{\vee} &= 1 - \left(-1 + u_1 + u_2 + \alpha_u^{\wedge}\right)^T \\ &= 1 - \left(u_1 + u_2 - \left(1 - \rho_{T,u}^{\wedge}\right)^{1/T}\right)^T \end{aligned} \tag{4.15}$$

$$\begin{aligned}\rho_{T,u}^{K} &= 1 - K\left(-1 + u_1 + u_2 + \alpha_u^{\wedge}\right)^T \\ &= 1 - K\left(u_1 + u_2 - \left(1 - \rho_{T,u}^{\wedge}\right)^{1/T}\right)^T \end{aligned} \tag{4.16}$$

$$\begin{aligned}\rho_{T,u}^{\breve{K}} &= 1 - \breve{K}\left(\alpha_u^{\wedge}\right)^T \\ &= 1 - \breve{K}\left(1 - \left(1 - \rho_{T,u}^{\wedge}\right)^{1/T}\right)^T \end{aligned} \tag{4.17}$$

Any Failure Probability can be written in terms of $\rho_{T,u}^{\wedge}$ and $\rho_{T,u}^{\wedge}$ can be calculated as a function of the other Failure Probabilities.

CHAPTER FIVE
APPLICATION

5.1 Heart Transplant Data

Our data is obtained from heart transplant recipients from the Stanford heart transplant program (Crowley and Hu, 1977). We tackle 69 transplant data from 103 transplants and also our data is used for the uncensored case. We model the dependence structure between waiting time for transplant and post-transplant survival time to see the co-movements of these variables and how they affect each other. Table 5.1 shows the goodness-of-fit results of the copula models. For the Frank copula model, we obtain $\hat{\alpha}_1 = 2.1930472$ and $\hat{\alpha}_2 = 2.134277$. Also, we calculate τ-values based on $\hat{\alpha}$ values as $\hat{\tau}_1 = 0.2328296$ and $\hat{\tau}_2 = 0.2271088$. As seen in Table 5.1, the Frank copula has the best fit (p-value=0.3594) for heart transplant data. The Gumbel copula model cannot be rejected at the 5% significance level also, but its GoF statistic is higher than the Frank copula model. The model has symmetric dependence structure and transplant and post-transplant survival times have symmetric co-movements. Figure 5.1, Figure 5.2 and Figure 5.3 are constructed using λ-functions of Clayton, Gumbel and Frank copula model, respectively. In Figures, the left graph is for the empirical λ-function of the data, the middle one is for the theoretical λ-function and the right one is for both empirical and theoretical λ-functions. When Figure 5.1, Figure 5.2 and Figure 5.3 are examined, it can be concluded that the Frank copula model has again best fit to the data.

Table 5.1 The goodness-of-fit test results for three Archimedean copula models based on the Stanford heart transplant program (Crowley & Hu, 1977)

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$(\hat{\gamma}_1 - \hat{\gamma}_2)/\hat{\sigma}_{jack}$	*p value*
Clayton	0.1692321	0.6449004	-6.437196	0
Frank	2 .1930472	2.134277	0.3676287	0.3594
Gumbel	1.2789026	1.2598375	0.9582767	0.1711

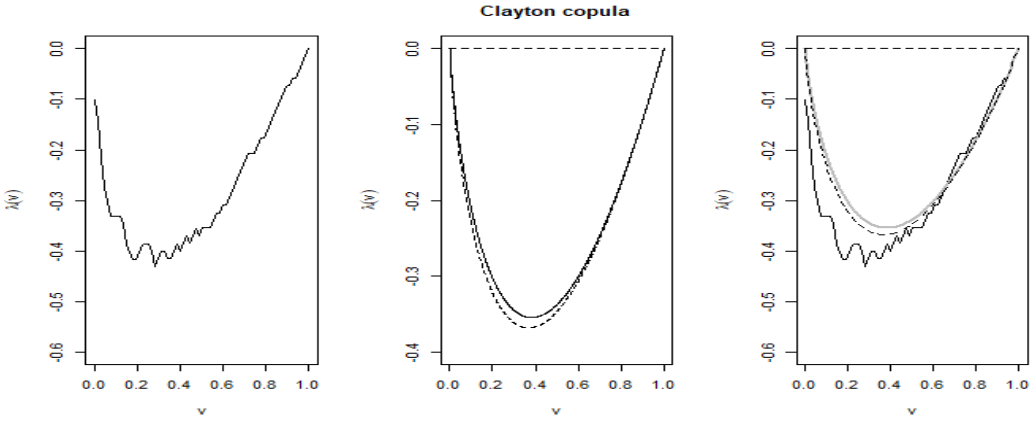

Figure 5.1 λ-functions based on the Stanford heart transplant program for Clayton copula model

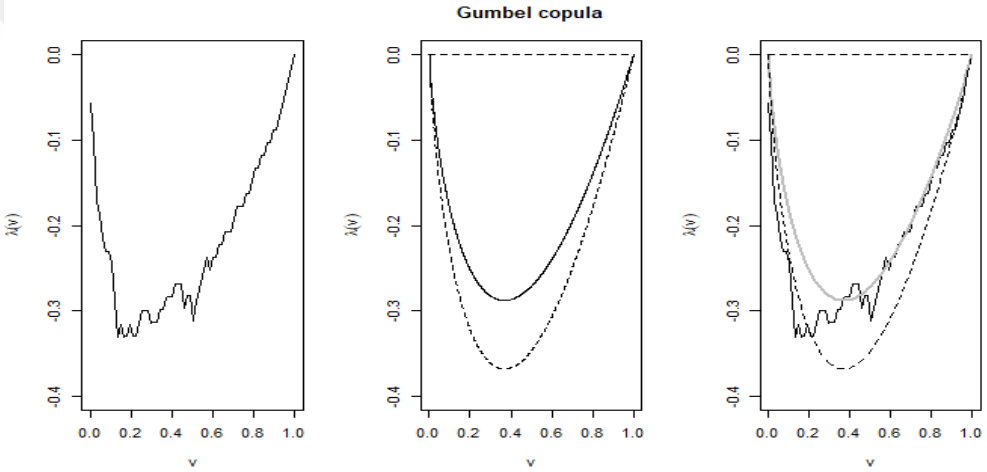

Figure 5.2 λ-functions based on the Stanford heart transplant program for Gumbel copula model

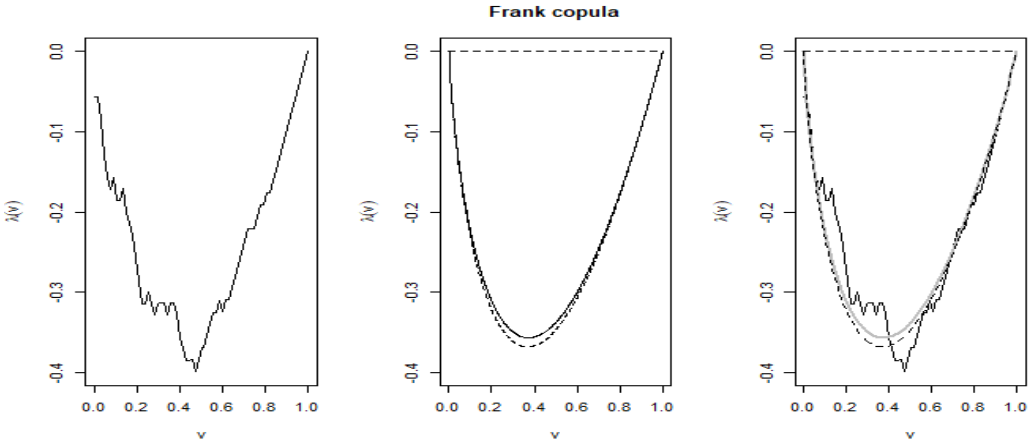

Figure 5.3 λ-functions based on the Stanford heart transplant program for Frank copula model

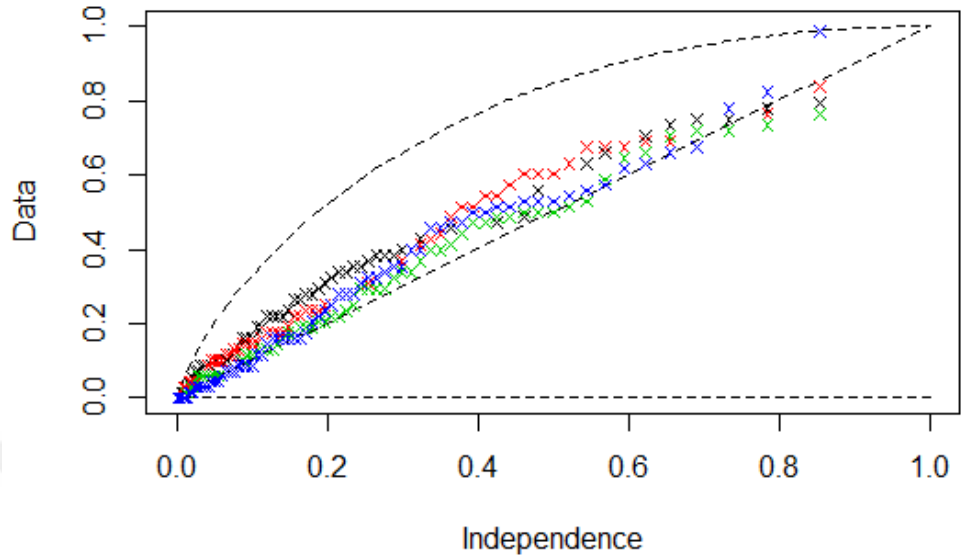

Figure 5.4 Kendall plot based on the Stanford heart transplant program

NOTE: Green (Frank copula), red (Clayton copula), blue (Gumbel copula)

Kendall distribution of the selected copula model for the data set is visualized in Figure 5.4. It can also be concluded that the Frank copula model has the best fit to the Stanford heart transplant data.

Frank copula allows negative dependence between the marginals and also the dependence is symmetric in tails. Frank copula is "comprehensive" in the sense that Fréchet lower bound and Fréchet upper bound are included in the range of dependence. In theory, Frank copula can be applied to the model outcomes with the strong positive or negative dependence (Trivedi, & Zimmer, 2005).

The generator, λ-function and Kendall's tau (τ) formulation based on joint survival distribution for families of Archimedean copulas are given in Table 5.2

Table 5.2 Examples of families of Archimedean copulas

	ϕ	λ	τ
Clayton	$\left(S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)^{-\alpha}-1\right)\Big/\alpha$	$-S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)\left(1-S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)^{\alpha}\right)\Big/\alpha$	$\alpha/(\alpha+2)$
Frank	$\log\left(\frac{1-\exp(-\alpha)}{1-\exp\left(-\alpha S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)\right)}\right)$	$-\frac{1-\exp\left(-\alpha S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)\right)}{\alpha\exp\left(-\alpha S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)\right)}\log\left(\frac{1-\exp(-\alpha)}{1-\exp\left(-\alpha S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)\right)}\right)$	$1+4\left\{D_1(\alpha)-1\right\}/\alpha$
Gumbel	$\left\{-\log\left(S(\tilde{X}_{ij},\tilde{Y}_{ij})\right\}^{\alpha+1}\right.$	$S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)\log\left(S\left(\tilde{X}_{ij},\tilde{Y}_{ij}\right)\right)\Big/(\alpha+1)$	$\alpha/(\alpha+1)$

NOTE: The three families are indexed by a single real parameter α

* D_1 is the Debye function of order 1, $D_1(\alpha)=\int_0^{\alpha}\left\{t/\alpha\left(e^t-1\right)\right\}dt$ (Genest and Rivest, 1993).

Table 5.3 Some probabilities exceeding the critical layer for selected heart transplant data

Survival time	**Waiting time**	$\mathrm{t}=\mathrm{H}(\mathrm{x},\mathrm{y})$	$\alpha_{\mathrm{t}}^{\check{\mathrm{K}}}$
5	5	0.1527	0.3913
16	2	0.5326	0.9130
16	1	0.7826	0.9710
17	5	0.3804	0.7246
30	5	0.4750	0.8405
39	38	0.1156	0.3043
39	36	0.1386	0.3478
43	20	0.3190	0.6521
45	1	0.8382	0.9855
51	12	0.4527	0.7971

In Table 5.3, Survival Hazard Scenario is presented for heart transplant study. In Table 5.3, the survival time and the waiting time and its Survival Hazard scenario levels $\alpha_{\mathrm{t}}^{\check{\mathrm{K}}}$ are given. The third column in Table 5.3 is bivariate critical threshold for the Frank copula for the given x, y. The fourth column $\alpha_{\mathrm{t}}^{\check{\mathrm{K}}}$ shows the probability of exceeding the critical layer. For example, the probability of exceeding in the critical layer 0.1527 is 0.3913 for the value of $(\mathrm{x},\mathrm{y})=(5,5)$ in Table 5.3 (Susam & Ucer, 2019).

Figure 5.5 shows rank-plot of fitted survival copula (regular lines) based on heart transplant study (waiting time data are y-axis and survival time data are x-axis) for Frank copula.

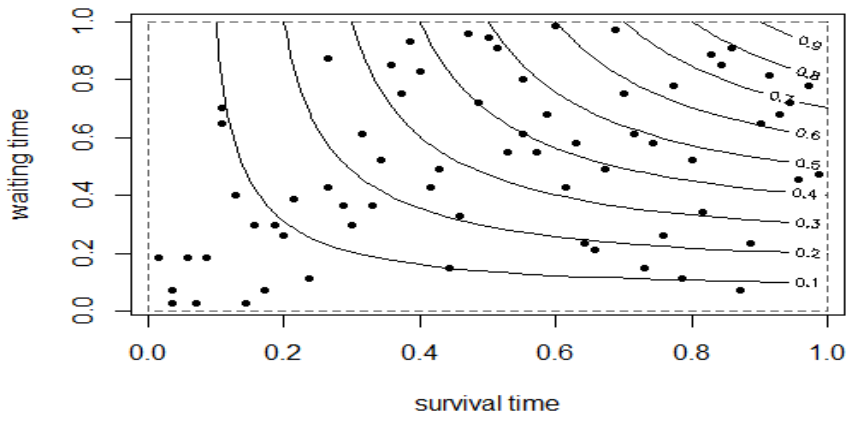

Figure 5.5 Rank-plot of fitted survival copula (regular lines) based on heart transplant study for Frank Copula (The isolines have levels 0.1: (0.1): 0.9 from left to right)

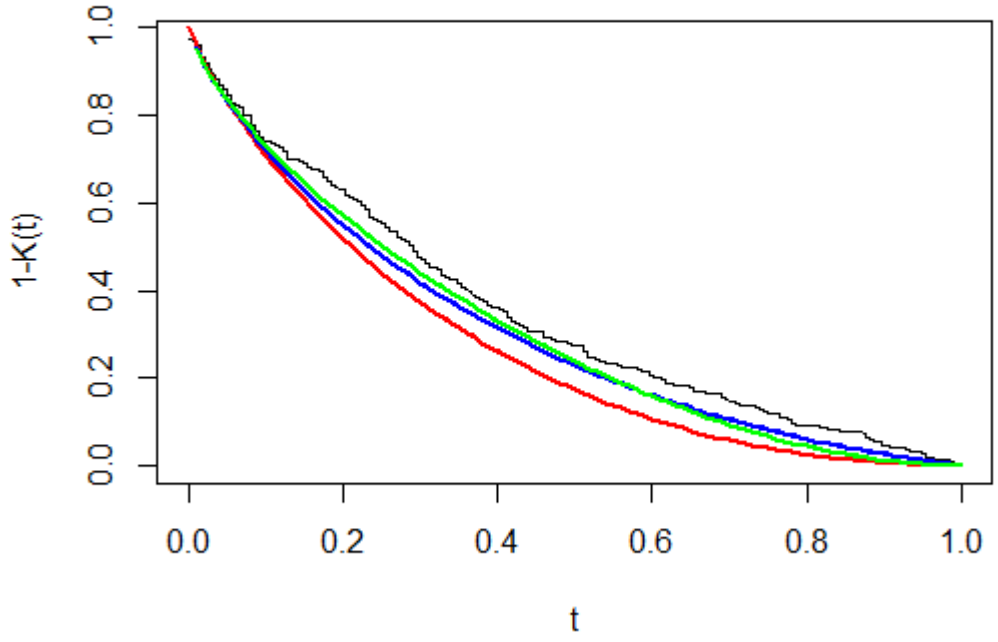

Figure 5.6 Survival Kendall plot based on heart transplant program

NOTE: Green (Frank copula), red (Clayton copula), blue (Gumbel copula)

We calculate Survival Kendall function $\bar{K}$. Figure 5.6 shows the empirical estimate of $\bar{K}$ and the function associated with the survival copula fitted to the data.

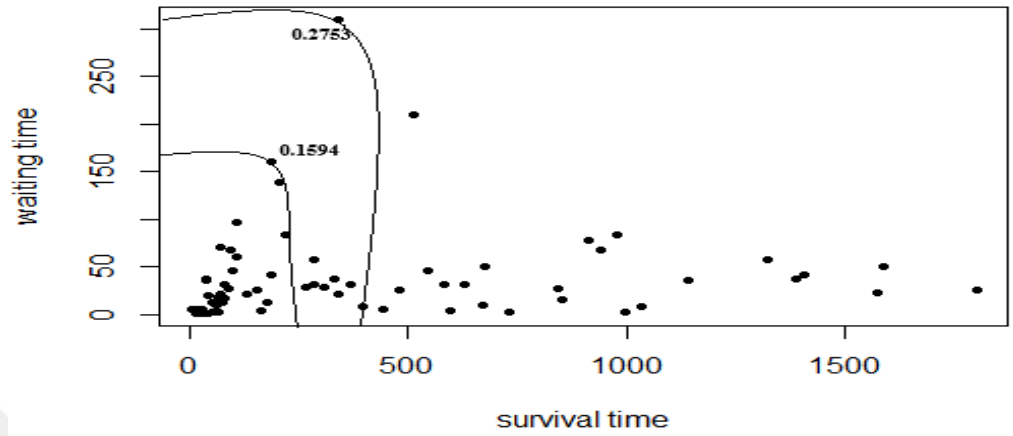

Figure 5.7 Observed pairs of heart transplant data, selected critical layer $L_t^{\check{K}}$ and level $\alpha_t^{\check{K}}$

In Figure 5.7 we can see the observed pairs of the data, the selected critical layer $L_t^{\check{K}}$ and the level $\alpha_t^{\check{K}}$ and also we determine the observations exceeding the critical layers. For example when the survival time is 340 days and the waiting time is 310 days, it is shown risk at level %27.5 as heart transplant study.

5.2 Diabetic Retinopathy Data

Our other data is obtained from diabetic retinopathy study (Manatunga, & Oakes, 1999). We tackle 197 patients for our censored study. One eye of patients was randomly selected for photocoagulation treatment and an experimental treatment was applied to one eye, a standard treatment to the other eye, of each patient enrolled. Manatunga and Oakes (1999) concentrated the effect of laser- photocoagulation therapy, and whether this effect differs for adult onset and juvenile onset patients and association between the survival times of the two eyes of the same patient for their analyses in their study. We examine that of the 197 patients, 38 experienced failure in both eyes, 79 experienced failure in one eye and 80 experienced no failure. Considering censored status, we use the variables as treated and untreated eyes and

also we deal with time to loss of vision or last follow-up of the patients as survival time.

In this data, we model the dependence structure between the survival times of the two eyes of the same patient after laser- photocoagulation therapy to see how they affect each other. In these two studies, we deal with Frank copula, Gumbel copula and Clayton copula models. The GoF results are given in Table 5.4. When we examine Table 5.4, the Frank copula model has best fit to our data (p-value=0.435 (Frank)). Also, for the Frank copula model, we obtain $\hat{\alpha}_1 = 1.5647446$ and $\hat{\alpha}_2 = 1.6743183$. The visual comparisons are also given in Figure 5.8, Figure 5.9 and Figure 5.10 via λ-functions. In Figures, the left graph is for the empirical λ-function of the data, the middle one is for the theoretical λ-function and the right one is for both empirical and theoretical λ-functions. When Figure 5.8, Figure 5.9 and Figure 5.10 are examined, it can be concluded that the Frank copula model has again best fit to the data.

Table 5.4 The goodness-of-fit test results for three Archimedean copula models based on the diabetic retinopathy study (Manatunga & Oakes, 1999)

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$(\hat{\gamma}_1 - \hat{\gamma}_2)/\hat{\sigma}_{jack}$	*p value*
Clayton	1.164139	1.246918	-1.00857	0.1587
Frank	3.897294	3.944172	-0.1755508	0.435
Gumbel	1.5647446	1.6743183	-2.603582	0.0047

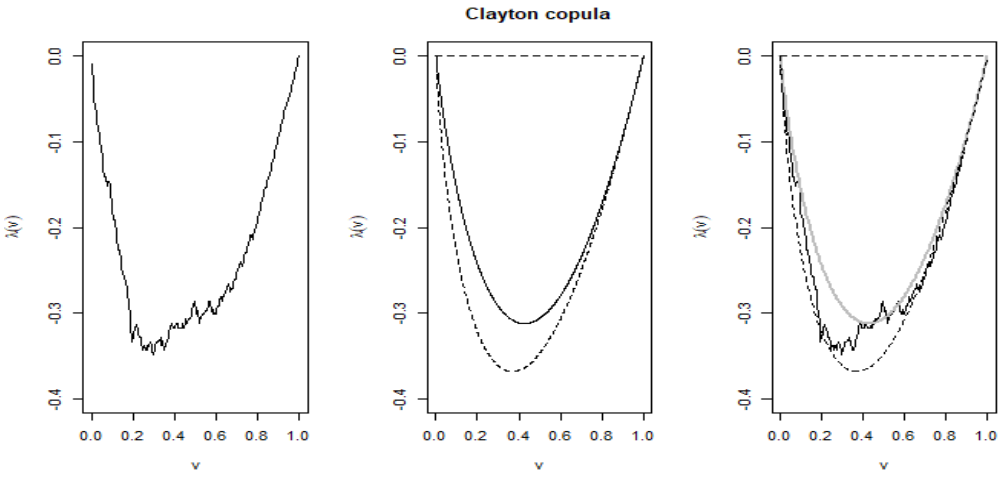

Figure 5.8 λ-functions based on the diabetic retinopathy study for Clayton copula model

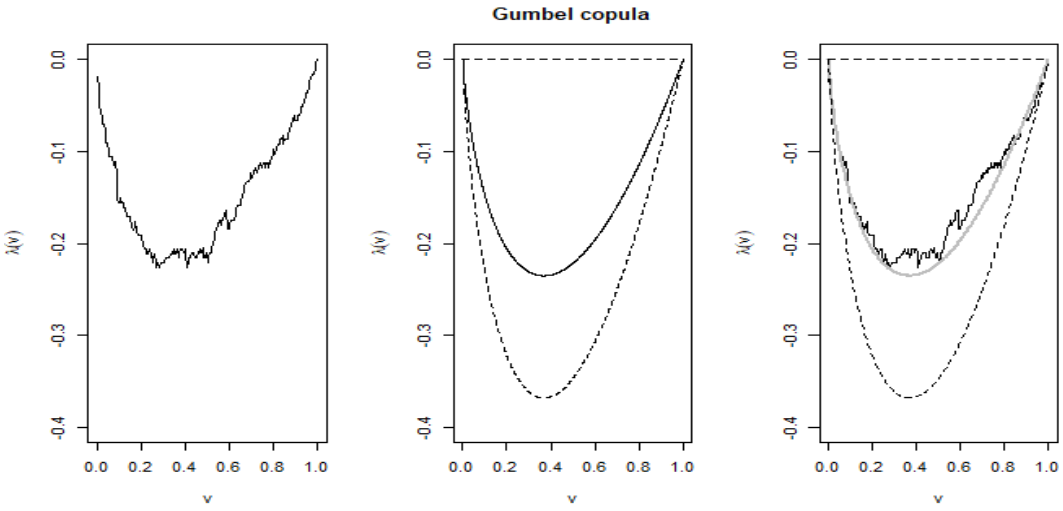

Figure 5.9 λ-functions based on the diabetic retinopathy study for Gumbel copula model

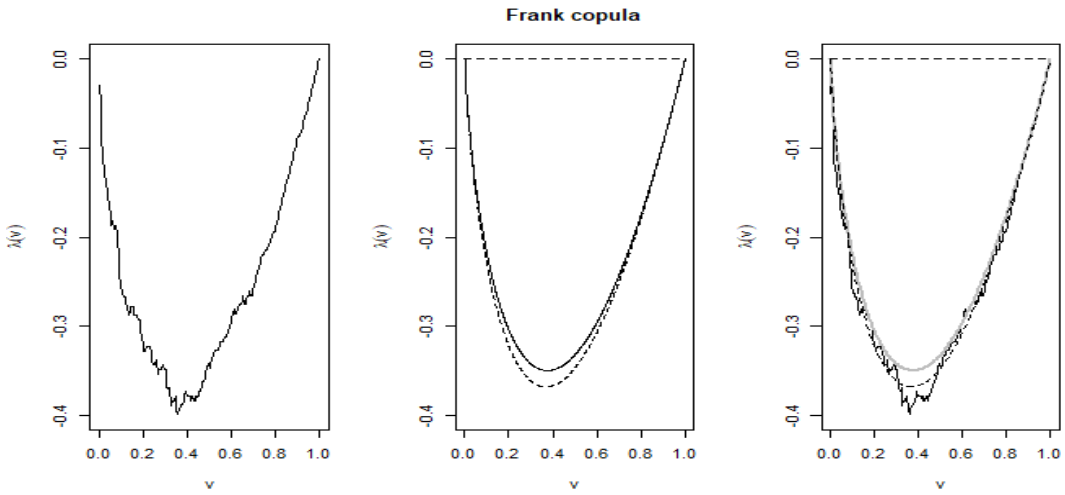

Figure 5.10 λ-functions based on the diabetic retinopathy study for Frank copula model

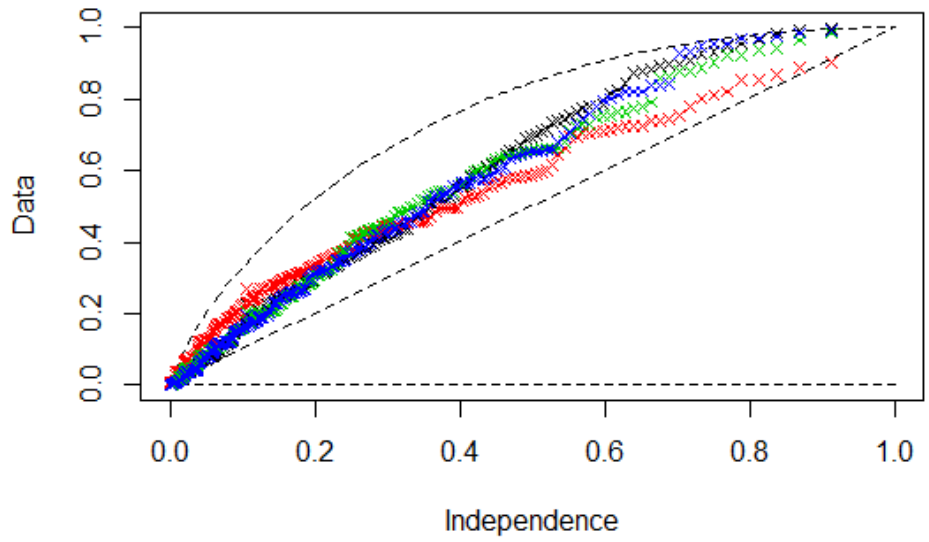

Figure 5.11 Kendall plot based on the diabetic retinopathy study

NOTE: Green (Frank copula), red (Clayton copula), blue (Gumbel copula)

Kendall distribution of the selected copula model for the data set is visualized in Figure 5.11. It can also be concluded that the Frank copula model has the best fit for diabetic retinopathy data.

Table 5.5 Some probabilities exceeding the critical layer for selected diabetic retinopathy data

Experimental treatment time	**Photocoagulation treatment time**	$t = H(x, y)$	$\alpha_t^{\breve{K}}$
46.23	46.23	0.2330	0.4771
31.3	42.5	0.2769	0.5431
42.27	42.27	0.2849	0.5532
20.6	20.6	0.4995	0.8172
0.3	38.77	0.3001	0.5736
54.27	65.23	0.0294	0.0152
10.8	63.5	0.0487	0.0913
23.17	23.17	0.4792	0.7868
13.83	58.07	0.0993	0.2385
48.53	46.43	0.2123	0.4365

In Table 5.5, Survival Hazard Scenario is presented for diabetic retinopathy study. In Table 5.5, for one of eyes nontreatment time and treatment time and its Survival Hazard Scenario levels $\alpha_t^{\breve{K}}$ are given. The third column in Table 5.5 is bivariate critical threshold for the Frank copula for the given x, y. The fourth column $\alpha_t^{\breve{K}}$ shows the probability of exceeding the critical layer.

Figure 5.12 is shows the rank-plot of fitted survival copula (regular lines) based on heart transplant study (experiment treatment (time) data is on the y-axis and photocoagulation treatment (time) data is on the x-axis) for Frank copula.

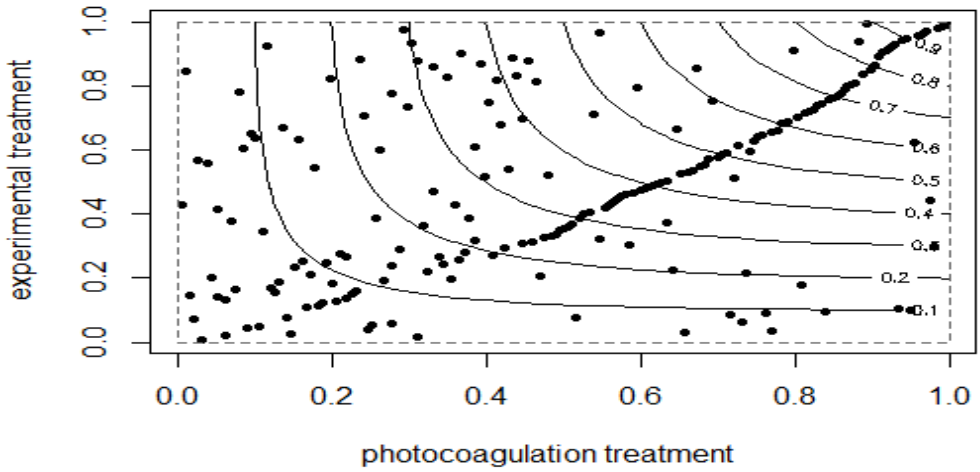

Figure 5.12 Rank-plot of fitted survival copula (regular lines) based on diabetic retinopathy study for Frank Copula (The isolines have levels 0.1: (0.1): 0.9 from left to right)

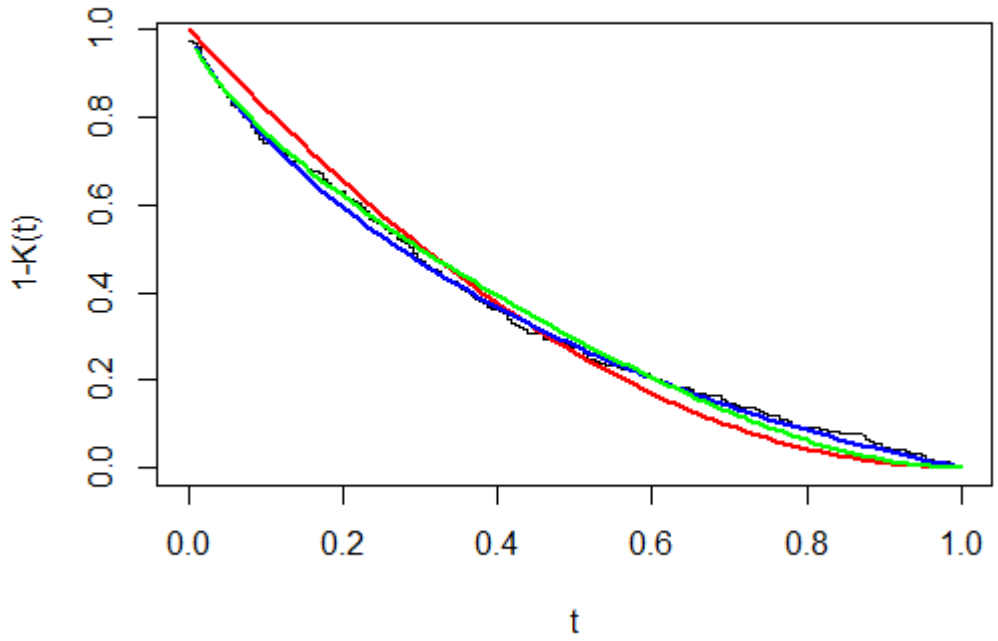

Figure 5.13 Survival Kendall plot based on diabetic study

NOTE: Green (Frank copula), red (Clayton copula), blue (Gumbel copula)

We calculate Survival Kendall function $\bar{K}$ and so it is spotted the survival critical layers. Figure 5.13 shows the empirical estimate of $\bar{K}$ and the function associated with the survival copula fitted on the data.

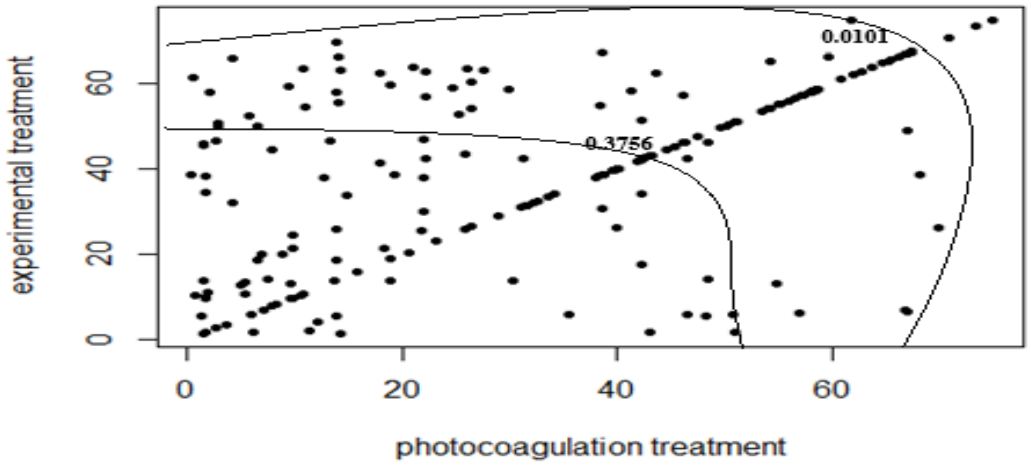

Figure 5.14 Observed pairs of diabetic retinopathy data, selected critical layer $L_t^{\check{K}}$ and $\alpha_t^{\check{K}}$

In Figure 5.14 we can see the observed pairs, the selected critical layer $L_t^{\check{K}}$ and the level $\alpha_t^{\check{K}}$ and also we can determine observations exceeding the critical layers. For example, in diabetic retinopathy study, when nontreatment time is 66.2 months and treatment time is 66.2 months it is shown risk at level % 1.

CHAPTER SIX
CONCLUSION

In this study, the adequacy of a copula-based bivariate survival model is investigated. We deal with modelling and analyzing bivariate survival uncensored and right-censored data for three Archimedean copula models (Clayton, Frank and Gumbel copula). Some problems occur in parameter estimation process when the survival times are dependent. This causes dependent censoring and non-identifiability of the marginal distributions for survival times. It is hard to find the best fit model especially under the right-censored case with unknown parameters. Thus, we apply the goodness-of-fit method proposed by Emura et al. (2010) to select the best appropriate Archimedean copula model for our data. Different from the conventional methods in survival analysis, this method yields a formal goodness-of-fit test using the weight function based on conditional likelihood. In the presence of right-censoring for the data, deleting non-orderable pairs from the equation is used. This strategy was also used by Oakes (1982). We use Kendall distribution function and λ-functions to choose the best fit visually. We adapt U-statistics to right-censored data. Since the variance estimator is complicated for the right-censored data, the jack-knife estimator is obtained. For both of the heart transplant data and diabetic retinopathy data, the Frank copula model is selected. The data sets have a symmetric dependence structure.

Finally, a probabilistically consistent framework for dealing with bivariate Survival Hazard Scenario is used. It provides valuable resulting for assessing the probability of threatening events. The hazard scenario is defined by means of the concept of Upper Set and determined in terms of a proper probability level. Besides, the hazard scenario is examined menacing events coherently and also in order to model various types of events; it examines a wide variety of setting. The hazard scenario framework outlined in this work deals with the concept of Failure Probability for design and risk assessment. We calculate the probability that patients of the survival time and waiting time jointly exceed the critical layer in heart

transplant program and also those patients of the treatment time and nontreatment time jointly exceed the critical layer in diabetic retinopathy program for Frank copula model. We deal with the Survival Hazard scenario to determine the suitable probability levels of extreme events for both of the data. The hazard scenario approach can be provided for the evaluation of the joint risk for the variables.

REFERENCES

AghaKouchak, A., Cheng, L., Mazdiyasni, O., & Farahmand, A. (2014). Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought. *Geophysical Research Letters, 41,* 8847–8852.

Barbe, P., Genest, C., Ghoudi, K., & Rèmiiard, B. (1996), On Kendall's process. *Journal of Multivariate Analysis, 58,*197-229.

Bandeen-Roche, K. J., & Liang, K. Y. (1996), Modelling failure-time associations in data with multiple levels of clustering. *Biometrika, 83* (1), 29-39.

Baybutt, P. (2003). Major hazards analysis – an improved process hazard analysis method. *Process Safety Progress, 22* (1), 21–26.

Brown, B. W. M., Hollander, M., & Korwar, R. M. (1974). Nonparametric tests of independence for censored data with applications to heart transplant studies. *Reliability & Biometry: Statistical Analysis of Life Length,* 327-354.

Bouyè, E., Durrleman, V., Nikeghbali, A., Riboulet, G., & Roncalli, T. (2000). *Copulas for finance - a reading guide and some applications.* France: Groupe de Recherche Op´erationelle, Cr´edit Lyonnais.

Chen, Z. (2012). *A flexible copula model for bivariate survival data*. Phd Thesis, University of Rochester, New York.

Cheng G., Zhou L., Chen, X., & Huang, J. Z. (2014). Efficient estimation of semiparametric copula models for bivariate survival data. *Journal of Multivariate Analysis, 123,* 330-344.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula methods in finance*. New York: John Wiley.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application to epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65* (1), 141-151.

Clayton, D., & Cuzick, J. (1985). Multivariate generalisations of the proportional hazards model. *Journal of the Royal Statistical Society, 148*, 82-117.

Crowley, J., & Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistic Association, 72*, 27-36.

Cox, D.R. (1972), Regression models and life-tables. *Journal of the Royal Statistical Society, 34*, 187-220.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *The Annals of Statistics, 16* (4), 1475-1489.

De Matteis, R. (2001). *Fitting copulas to data*. Phd thesis, University of Zurich, Zurich.

Durante, F., & Sempi, C. (2015). *Principles of copula theory*, London: CRC/Chapman & Hall.

Emura, T., Lin, C. W., & Wang, W. (2010). A goodness-of-fit test for Archimedean copula models in the presence of right censoring. *Computational Statistics & Data Analysis, 54* (12), 3033–3043.

Emura, T., Wang, W., & Hung, H. N. (2011). Semi-parametric inference for copula models for truncated data. *Statistica Sinica, 21*, 349–367.

Fermanian, J. D. (2005). Goodness-of-fit tests for copulas. *Journal Multivariate Analysis, 95*, 119-152.

Fermanian, J. D., & Scaillet, O. (2005). Some statistical pitfalls in copula modelling for financial applications. *In Capital formation, governance and banking* (57–72). New York: Nova Science Publishing.

Fermanian, J. D. (2013). An overview of the goodness-of-fit test problem for copulas. *Copulae in mathematical and quantitative finance*. 61–89. Berlin Heidelberg: Springer.

Frees, E. W., Carriere, J. F., & Valdez, E. A. (1996). Annuity valuation with dependence mortality. *Journal of Risk and Insurance, 63* (2), 229-261.

Genest, C., & MacKay, R. J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *American Statistics, 40*, 280-283.

Genest, C., & Rivest, L. P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association, 88*, 1034-1043.

Genest, C., Ghoudi, K., & Rivest, L. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika, 82*, 543-552.

Georges, P., Lamy, A. G., Nicolas, A., Quibel, G., & Roncalli, T. (2001). *Multivariate survival modelling: a unified approach wiith copulas*, France: Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Hu, L. (2006). Dependence patterns across financial markets: a mixed copula approach. *Applied Financial Economics, 16* (10), 717–729.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical. Statistics, 19,* 293-325.

Hougaard, P., Harvald B., & Holm, N. V. (1992). Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930. *Journal of the American Statistical Association, 87*, 17–24.

Hougaard, P. (1986a), Survival models for heterogeneous populations derived from stable distributions. *Biometrika, 73* (2), 387-396.

Hougaard, P. (1986b), A class of multivariate failure time distributions. *Biometrika, 73* (3), 671-678.

Klicnarová, J. (2017). On weighted U-statistics for stationary random fields. *Kybernetika, 53* (2), 220–230.

Kpanzou, T. A. (2008). *Aspects of copula and goodness of fit*. Master Thesis, Stellenbosch University, Stellenbosch.

Lai, C. D., & Xie, M. (2006). Concepts and Measures of Dependence in Reliability *Stochastic ageing and dependence for reliability, (1st ed.),* New York: Springer.

Manatunga, A. K., & Oakes, D. (1999). Parametric analysis of matched pair survival data. *Lifetime Data Analysis, 5*, 371 – 387.

Marshall, A. W., & Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association, 62*, 30-44.

Marshall, A.W., & Olkin I. (1988), Families of multivariate distributions. *Journal of the American Statistical Association, 83*, 834-841.

Mei, M. (2016). *A goodness-of-fit test for semi-parametric copula models of right censored bivariate survival times*. Master's Thesis, Simon Fraser University, Canada.

Nelsen, R. (1999). *An introduction to copulas*. New York: Springer.

Nelsen, R. (2006). *An introduction to copulas*. New York: Springer.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, 44,* 414-422.

Oakes, D. (1986). Semiparametric inference in a model for association in bivariate survival data. *Biometrika,* 73, 353-361.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association, 84,* 487493.

Ostwald, D. (2019). *Statistics for Data Science*. Retrieved December 28, 2019, from https://www.ewi-psy.fu-berlin.de

Paulatto, M., Pinat, T., & Romanelli, F. (2007). Tsunami hazard scenarios in the Adriatic Sea domain. *Natural Hazards and Earth System Sciences, 7,* 309–325.

Rosenblatt, M. (1952). Remarks on a multivariate transform. *Annals of Mathematical Statistics, 23,* 470–472.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley.

Salvadori, G., De Michele, C., & Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences, 15,* 3293-3305.

Salvadori, G., Durante, F., De Michele, C., Bernardi, M., & Petrella, L. (2016). A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. *Water Resources Research, 52,* 3701–3721.

Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Insurance and Risk Management Journal, 74* (1), 639-660.

Schweizer, B., & Sklar, A. (1983), *Probabilistic metric spaces.* New York: North Holland.

Shih, J., & Louis, T. A. (1995). Inference on association parameter in copula models for bi variate survival data. *Biometrics, 26*, 183.

Shih, J., H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika, 85,* 189- 200.

Straub, D. (Ed.). (2014). *Engineering risk assessment. In risk - A multidisciplinary introduction*, (333–362). New York: Springer.

Susam, S. O., & Ucer, B. (2019). Modelling the dependence structure of CO2 emissions and energy consumption based on the Archimedean copula approach: the case of the United States. *Energy Sources, Part B: Economics, Planning, and Policy*, *14*, 274-289.

Trivedi, P. K., & Zimmer, D. M. (2005). Copula modelling: an introduction for practitioners. *Foundations and Trends in Econometrics, 1* (1), 1-111.

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

Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society, 65*, 257–273.

Wang, A. (2010). Goodness-of-Fit tests for Archimedean copula models. *Statistica Sinica, 20*, 441-453.

Wulp, G. (2003). *Using copula in risk management*. Master's Thesis, Tilburg University, Holland.

Yilmaz, Y. E. (2009). *Estimation and goodness of fit for multivariate survival models based on copulas*. Phd Thesis, University of Waterloo, Canada.

APPENDICES

```
#R codes for the Frank copula based on diabetic retinopathy (with right-censored)

#U-statistics calculations


library("copula")

n=197 # row

c=2 # column

cdiab<-cbind(diabet1$time...3,diabet1$time...9)


#(Xi,Yi) are subject to right censoring by (Ai, Bi)

ab1<-matrix(nrow = n, ncol = 2)

ı=matrix(nrow=n, ncol=1)

r=matrix(nrow=n, ncol=1)

for (i in 1:n) {

  ı[i]<-min(cdiab[i,1],cdiab[i,1])

  r[i]<-min(cdiab[i,2],cdiab[i,2])

}

ab1[,1]<-ı[,1]

ab1[,2]<-r[,1]
```

```
#minimum value of Ai,Aj and minimum value of Bi,Bj
ıı=matrix(nrow=n, ncol=n)
rr=matrix(nrow=n, ncol=n)
for (a in 1:196) {
  for (j in (a+1):197) {
    ıı[a,j]<-min(cdiab[a,1],cdiab[j,1])
    rr[a,j]<-min(cdiab[a,2],cdiab[j,2])
  }
}

#Rxy is measures the number at risk at (x,y)
Rxy<- function (ab1,a,b)
{
  return(sum(as.numeric(ab1[,1]>=a & ab1[,2]>=b)))
}
y=matrix(nrow=n, ncol=n)
y1=matrix(nrow=n, ncol=n)
Rif=matrix(nrow = n, ncol = n)
for (a in 1:(n-1)) {
  for (j in (a+1):n) {
```

```
 #minimum values of X's and Y's
    y[a,j]<-min(cdiab[a,1],cdiab[j,1])
    y1[a,j]<-min(cdiab[a,2],cdiab[j,2])
    Rif[a,j]<-Rxy(ab1, y[a,j], y1[a,j])
  }
}

n=197
m<-1:197
fZi<-(Rif[m,m]/n) #emprical survival distribution

#Zij indicates whether the ordering relationship is certain or not
Zi=((y[m,m]<=ıı[m,m])&(y1[m,m]<=rr[m,m]))
Zi=Zi*1

#fdelta is concordance indicator for observable
fdelta=matrix(nrow=n, ncol=n)
for (a in 1:(n-1)){
  for (j in (a+1):n) {
    if ((ab1[a,1]-ab1[j,1])*(ab1[a,2]-ab1[j,2])>0) {
      fdelta[a,j]=1
```

```
   } else {
     fdelta[a,j]=0
   }
  }
}

print(fdelta)

#θ(S(x,y))<- ((fZ*alpha)/(1-exp(-alpha*fZ)))
#(cross-ratio function for survival function)

# first U-statistics
fU1<- function(alpha){
  e=matrix(0L, nrow = n, ncol=n)
  for (v in 1:(n-1)) {
    for(vv in (v+1):(n)){

      e[v,vv]<-     Zi[v,vv]*((fZi[v,vv]/(1-exp(-fZi[v,vv]*alpha))-fZi[v,vv]^2*alpha/(1-
exp(-fZi[v,vv]*alpha))^2*exp(-fZi[v,vv]*alpha))*(fZi[v,vv]*alpha/(1-exp(-
fZi[v,vv]*alpha))+1)/fZi[v,vv]/alpha*(1-exp(-fZi[v,vv]*alpha))/(Rif[v,vv]-
1+fZi[v,vv]*alpha/(1-exp(-fZi[v,vv]*alpha)))*(fdelta[v,vv]-fZi[v,vv]*alpha/(1-exp(-
fZi[v,vv]*alpha))/(fZi[v,vv]*alpha/(1-exp(-fZi[v,vv]*alpha))+1)))
}
```

```
  }
  sum(e)
}
b<-try(uniroot(fU1, lower = 0.00001, upper=4))

#second U-statistics
fU2<-function(alpha){
  aa=matrix(0L, nrow = n, ncol=n)

  for (v in 1:(n-1)) {
    for(vv in (v+1):(n)){
      aa[v,vv]<-Zi[v,vv]*(fdelta[v,vv]-fZi[v,vv]*alpha/(1-exp(-
fZi[v,vv]*alpha))/(fZi[v,vv]*alpha/(1-exp(-fZi[v,vv]*alpha))+1))
}
  }
  sum(aa)
}
bb<-try(uniroot(fU2, lower = 0.000001, upper = 4))

library(boot)
library(bootstrap)
y<-cdiab[,1]
```

```
z<-cdiab[,2]

nn<-nrow(cdiab)


#Jackknife Estimation

theta <- function(x,cdiab){ cor(cdiab[x,1],cdiab[x,2]) }

results <- jackknife(1:n,theta,cdiab)

stat<-(log(b$root)-log(bb$root))/results$jack.se


stat



#R codes for the Frank copula based on diabetic retinopathy (with right-censored)

#Survival Hazard Scenario calculations


library("VineCopula")

diabetic<-cbind(diabet1[,3],diabet1[,9])

d<-pobs(diabetic)


#pit(probability integral transformation)

alfa= 3.897294

RVM<-RVineStructureSelect(d)

pit<-RVinePIT(d, RVM)
```

```
a<-pit[,1]

b<-pit[,2]

aa<-1-pit[,1]

bb<-1-pit[,2]


#the Frank copula and survival copula

nn=197

frank=(-1/alfa)*log(1+(((exp(-alfa*a)-1)*(exp(-alfa*b)-1))/(exp(-alfa)-1)))

surcop=aa+bb-1+frank #S(x,y)=v t=H(x,y)



ii=rep(1:nn, each=nn)

jj=rep(1:nn,nn)

t=min(surcop) #Lt (survival critical layer)


tt=(surcop[ii]<surcop[jj])

tt1=tapply(tt, ii, sum)

tt2=tapply(tt, ii, sum)/nn #subcritical layer


ttt=(surcop[ii]>surcop[jj])

ttt1=tapply(ttt,ii,sum)

ttt2=tapply(ttt, ii, sum)/nn #supcritical layer
```

```
plot(d,xlab="photocoagulation treatment",ylab="experimental treatment",pch = 20, cex = 1)


#Survival Kendall Hazard Scenario

surken=surcop-(log((exp(-alfa*surcop)-1)/(exp(-alfa)-1))/alfa)*(exp(-alfa*surcop)-1)

sur=(surken[ii]>=surken[jj])

sur1=tapply(sur,ii,sum)

surkendall=tapply(sur,ii,sum)/nn #alpha Kt



#R codes for the Frank copula based on Stanford heart transplant data

#(with non-censored)

#U-statistics calculations

library("copula")

n=69

c=2


#fdelta is concordance indicator for observable

fdelta=matrix(nrow=n, ncol=n)

for (i in 1:(n-1)){

  for (j in (i+1):n) {
```

```
    if ((heart[i,1]-heart[j,1])*(heart[i,2]-heart[j,2])>0) {
      fdelta[i,j]=1
    } else {
      fdelta[i,j]=0
    }
  }
}
print(fdelta)


#Rxy is measures the number at risk at (x,y)
Rxy<- function (heart,a,b)
{
  return(sum(as.numeric(heart[,1]>=a & heart[,2]>=b)))
}


#minimum value of (Xi,Yi)'s
y=matrix(nrow=n, ncol=n)
yy=matrix(nrow=n, ncol=n)
fR=matrix(nrow = n, ncol = n)
for (a in 1:(n-1)) {
  for (j in (a+1):n) {
```

```
   #(X,Y'lerin minimum değerleri)
   y[a,j]<-min(heart[a,1],heart[j,1])
   yy[a,j]<-min(heart[a,2],heart[j,2])
   fR[a,j]<-Rxy(heart, y[a,j], yy[a,j])
  }
}

m<-1:69
fZ<-(fR[m,m]/n)  #emprical survival function

#θ(S(x,y))<- ((fZ*alpha)/(1-exp(-alpha*fZ)))
library("pracma")

#first U-statistics
fU1<-function(alpha) {
  u=matrix(0L, nrow = n, ncol=n)
  for (v in 1:(n-1)) {
    for(vv in (v+1):(n)){
      u[v,vv]<-(fZ[v,vv]/(1-exp(-fZ[v,vv]*alpha))-fZ[v,vv]^2*alpha/(1-exp(-
fZ[v,vv]*alpha))^2*exp(-fZ[v,vv]*alpha))*(fZ[v,vv]*alpha/(1-exp(-
fZ[v,vv]*alpha))+1)/fZ[v,vv]^2/alpha*(1-exp(-fZ[v,vv]*alpha))*(fdelta[v,vv]-
```

```
fZ[v,vv]*alpha/(1-exp(-fZ[v,vv]*alpha))/(fZ[v,vv]*alpha/(1-exp(-
fZ[v,vv]*alpha))+1))
    }
  }
  sum(u)
}

b<-try(uniroot(fU1, lower = 1, upper = 4))

#second U-statistics
fU2<-function(alpha){
  uu=matrix(0L, nrow = n, ncol=n)
  for (v in 1:(n-1)) {
    for(vv in (v+1):(n)){
      uu[v,vv]<-fdelta[v,vv]-fZ[v,vv]*alpha/(1-exp(-
fZ[v,vv]*alpha))/(fZ[v,vv]*alpha/(1-exp(-fZ[v,vv]*alpha))+1)
    }
  }
  sum(uu)
}

bb<-try(uniroot(fU2, lower = 1, upper = 3))
```

```
library(boot)

library(bootstrap)

y<-heart$`heart[,1]`

z<-heart$`heart[,2]`

nn<-nrow(heart)


#Jackknife Estimation

theta <- function(x,heart){ cor(heart[x,1],heart[x,2]) }

results <- jackknife(1:n,theta,heart)


stat<-(log(b$root)-log(bb$root))/results$jack.se

stat




#R codes for the Frank copula based on diabetic retinopathy (with right-censored)

#Survival Hazard Scenario calculations


library("VineCopula")

heart<-cbind(heart[,1],heart[,2])

h<-pobs(heart)
```

```
#pit(probability integral transformation)

alfa=2.1930472

RVM<-RVineStructureSelect(h)

pit<-RVinePIT(h, RVM)

a<-pit[,1]

b<-pit[,2]

aa<-1-pit[,1]

bb<-1-pit[,2]


#the Frank copula and survival copula

nn=69

frank=(-1/alfa)*log(1+(((exp(-alfa*a)-1)*(exp(-alfa*b)-1))/(exp(-alfa)-1)))

surcop=aa+bb-1+frank #S(x,y)=v t=H(x,y)


ii=rep(1:nn, each=nn)

jj=rep(1:nn,nn)

t=min(surcop) #Lt (survival critical layer)


tt=(surcop[ii]<surcop[jj])

tt1=tapply(tt, ii, sum)

tt2=tapply(tt, ii, sum)/nn #subcritical layer
```

```
ttt=(surcop[ii]>surcop[jj])

ttt1=tapply(ttt,ii,sum)

ttt2=tapply(ttt, ii, sum)/nn #supcritical layer


plot(h,xlab="waiting time",ylab="survival time",pch = 20, cex = 1)


#Survival Kendall Hazard Scenario

surken=surcop-(log((exp(-alfa*surcop)-1)/(exp(-alfa)-1))/alfa)*(exp(-alfa*surcop)-1)

sur=(surken[ii]>=surken[jj])

sur1=tapply(sur,ii,sum)

surkendall=tapply(sur,ii,sum)/nn #alpha Kt


######
```