ELEANDERSON CAMPOS

FACTOR COPULA MODELS FOR RIGHT-CENSORED CLUSTERED SURVIVAL DATA

LAVRAS – MG

2020

ELEANDERSON CAMPOS

FACTOR COPULA MODELS FOR RIGHT-CENSORED CLUSTERED SURVIVAL DATA

Thesis submitted in fulfillment of the requirements for the joint degree of Doctor in Science. Areas of concentration: Statistics and Agricultural Experimentation at UFLA, and Statistics at UHasselt.

Prof. Dr. Lucas Monteiro Chaves
Promotor (UFLA)
Prof. Dr. Roel Braekers
Promotor (UHasselt)

Prof. Dr. Devanil Jaques de Souza
Co-promotor

LAVRAS – MG
2020

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Campos, Eleanderson.
Factor copula models for right-censored clustered survival data / Eleanderson Campos. - 2020.
97 p. : il.

Orientador(a): Lucas Monteiro Chaves, Roel Braekers.
Coorientador(a): Devanil Jaques de Souza.
Tese (doutorado) - Universidade Federal de Lavras/Universiteit Hasselt, 2020.
Bibliografia.

1. Dados de sobrevivência agrupados. 2. Cópulas fatoriais. 3. Clusters de tamanhos variados. I. Monteiro Chaves, Lucas. II. Braekers, Roel. III. Jaques de Souza, Devanil. IV. Título.

Depot number - UHasselt: D/2020/2451/ 65

Universiteit Hasselt
Universidade Federal de Lavras

PhD Defence

Official report from the meeting of the examination-board on the 25th of November 2020.

The examination-board:

Prof. dr. Joao Domingos Scalon (chairman, Universidade Federal de Lavras, Brazil)
Prof. dr. Roel Braekers (promoter, Hasselt University, Belgium)
Prof. dr. Lucas Monteiro Chaves (promoter, Universidade Federal de Lavras, Brazil)
Prof. dr. Devanil Jaques de Souza (co-promoter, Universidade Federal de Lavras, Brazil)
Prof. dr. Renato Ribeiro de Lima (member of the commission, Universidade Federalde Lavras, Brazil)
Prof. dr. Daniel Furtado Ferreira (member of the commission, Universidade Federal de Lavras, Brazil)
Prof. dr. Enrico Antonio Colosimo (member of the jury, Universidade Federal de Minas Gerais, Brazil)
Prof. dr. Suely Ruiz Giolo (member of the jury, Universidade Federal do Parana, Brazil)

Has attended the public defense by

Eleanderson Campos Eugenio Filho

born in Varginha (Brazil), on the 2nd of December 1992

of his doctoral thesis :

"Factor copula models for right-censored clustered survival data"

After deliberation they have decided to grant him the diploma of

Doctor of Science: Statistics (Universiteit Hasselt)
Doctor of Science in Statistics and Agricultural Experimentation (Universidade Federal de Lavras)

Language of defence: English

25th of November 2020, Diepenbeek

The Chairman | Members of the examination board | The Secretary

1850587

À minha amada mãe Lourdes.
DEDICO

ACKNOWLEDGMENTS

Primeiramente, gostaria de expressar minha imensa gratidão e amor à minha mãe Lourdes, que sempre cuidou de mim e me incentivou a correr atrás dos meus sonhos. Muito obrigado também à minha amada companheira Giovana, pelo carinho e por ter ficado ao meu lado durante todo este período de intenso trabalho. À minha vovó Célia, tia Zilda e ao meu padrinho Elessandro, muito obrigado, eu amo vocês. Não poderia esquecer de agradecer também a todos os meus familiares e amigos de Varginha e Lavras.

I will be forever grateful to my co-supervisor and friend Prof. Devanil, for having accepted me as his pupil and for always motivating me to work hard. You are a great example for me. Many thanks also to my supervisor at UHasselt, Prof. Roel, who devoted a large part of his time to assist me in the development of this work. Your support was crucial. My sincere thanks for accepting me as your apprentice. I would also like to express my gratitude to my supervisor at UFLA, Prof. Lucas. Thank you for your support.

My unlimited gratitude goes to my special Belgian friends Jeroen and Bente. Thank you very much for everything you did for me during my stay in Hasselt, you made my days happier when I was away from my family. Many thanks also to all the friends I made at Het Eiland, I miss our relaxed and pleasant conversations. Dank u wel!

I would also like to express my gratitude to the Federal University of Lavras (UFLA) and Universiteit Hasselt (UHasselt) for giving me the opportunity to do this PhD. Many thanks also to all members and staff of I-BioStat / DSI at UHasselt and DES at UFLA, for their promptness to help and collegiality.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. I would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) for all the financial support that was essential for the accomplishment of the joint PhD between UFLA and UHasselt (PDSE - Process 88881.190074 / 2018-01).

RESUMO

Grande parte dos fenômenos aleatórios estudados em áreas aplicadas da estatística é governada por complexas estruturas de dependência. Negligenciar essas associações na análise estatística quase certamente leva a resultados enganosos. É exatamente isso que torna a análise multivariada de dados tão importante. Na análise de sobrevivência multivariada, por exemplo, particularmente quando se trata de dados agrupados, o interesse está em modelar os múltiplos tempos de vida (ou o tempo até que um evento aconteça) de indivíduos agrupados em diferentes *clusters*. Naturalmente, os tempos de vida de indivíduos dentro de um mesmo *cluster* tendem a estar associados entre si (dependência *intracluster*), uma vez que esses indivíduos estão expostos aos mesmos fatores, por exemplo. Além disso, os dados de sobrevivência são frequentemente censurados à direita, uma condição em que indivíduos sobrevivem até certo ponto no tempo, mas não se sabe exatamente quando experimentam o evento de interesse. Essas particularidades tornam o estudo de dados de sobrevivência agrupados não trivial. Em geral, existem dois tipos de modelos que são comumente utilizados para modelar esses dados: modelos de fragilidade e cópulas. Nos modelos de fragilidade se assume que os diferentes tempos de vida em um *cluster* são condicionalmente independentes dado um fator comum aleatório, o fator de fragilidade (ou variável de fragilidade). Apesar de serem amplamente utilizados para modelar dados de sobrevivência agrupados em *clusters*, os modelos de fragilidade apresentam algumas deficiências: o número de distribuições de fragilidade implementadas é limitado. Além disso, a interpretação do parâmetro de fragilidade não é direta, pois este representa a heterogeneidade entre os *clusters* e não a associação entre os tempos de vida em um *cluster*. Por outro lado, a interpretação dos parâmetros de uma cópula é mais simples, uma vez que estes modelos, por definição, conseguem separar o comportamento marginal dos tempos de vida da forma com que estão associados. Ademais, é possível encontrar um grande número de famílias de cópulas paramétricas na literatura e muitas já foram implementadas em diversos pacotes de softwares estatísticos. No entanto, até então, o potencial das cópulas não foi totalmente explorado na modelagem de dados de sobrevivência clusterizados, visto que esses modelos foram aplicados apenas em casos de *clusters* de tamanho fixo e pequeno, ou com restrição no número de famílias de cópulas utilizáveis. Visando superar essas deficiências, nós propomos neste trabalho uma nova classe de modelos baseados em cópulas fatoriais para dados censurados à direita e agrupados em *clusters* de tamanhos variados. O novo modelo permite, ainda, o uso de qualquer família de cópula para modelar a dependência *intracluster*. Além disso, nós fornecemos as rotinas computacionais em R para implementação de nossos métodos. Por meio de uma aplicação com dados reais e estudos de simulação, nós mostramos que a metodologia proposta neste trabalho possui sólidas propriedades amostrais, baixo custo computacional e fornece resultados de fácil interpretação.

Palavras-chave: Dados de sobrevivência agrupados. Cópulas fatoriais. Dados de sobrevivência multivariados. *Clusters* de tamanhos variados.

ABSTRACT

The vast majority of the random phenomena studied by applied statisticians are governed by complex dependence structures. Neglecting these associations in the statistical analysis often gives rise to misleading and biased results. This is precisely what makes multivariate data analysis so important. In multivariate survival analysis, for example, particularly when dealing with clustered survival data, the interest lies in modelling the multiple lifetimes (time until an event happens) of individuals grouped in clusters. The lifetimes of individuals inside a cluster are known to be associated to each other through a complicated dependence structure, the intra-cluster dependence. On top of this, survival data are often right-censored, a condition where a subject survives up to a certain point in time, but the exact moment of occurrence of the event of interest is not observed. These features make the study of right-censored clustered survival data non-trivial. In general, there are two types of models that are commonly used to model these forms of data: frailty models and copula models. In frailty models, we assume that the different lifetimes in a cluster are independent of each other, conditional on a common random term, the frailty term. Although frailty models are widely used to model clustered survival data, they have some deficiencies: the number of frailty distributions which are implemented is limited. Furthermore, the interpretation of the frailty parameter is not straightforward since it expresses the heterogeneity between clusters, rather than the association between lifetimes in a cluster. On the other hand, the interpretation of the parameters in copula models is easier since these models are, by their form, adapted to make a clear distinction between the marginal behaviour of a lifetime and the association between different lifetimes. Moreover, an extensive number of parametric copula families is available and already implemented in several statistical software packages. However, up to now, copula models have not been used to their full potential in clustered survival data modelling. Their usage was restricted to settings where either the size of the clusters is fixed, or the number of copula families implemented is limited. Considering these shortcomings of the current methodologies, this thesis aims to make a contribution towards the modelling of clustered survival data by copula models. In this sense, we propose a new class of models based on the flexible factor copula models that can handle right-censored clustered survival data grouped in variable sized clusters and allows the use of any copula family to model intracluster dependence. Additionally, we provide the computational routines for implementations of our methods. We show, with a real data application and simulation studies, that the newly proposed methods have solid finite sample properties, straightforward interpretation and are not computationally expensive.

Keywords: Clustered survival data. Factor copula models. Multivariate survival data. Varying cluster size.

LIST OF FIGURES

Figure 2.1 – Graphs of copulas C_U (a), C_L (b) and region that contains all copulas (c) . . 12

Figure 2.2 – Graph of the copula Π. 12

Figure 2.3 – Scatterplots from samples of size 1000 taken from a Gaussian copula with $\theta = 0.5$ (a), $\theta = 0.85$ (b), $\theta = -0.5$ (c) and $\theta = -0.85$ (d). 19

Figure 2.4 – Scatterplots from samples of size 1000 taken from a t copula ($\nu = 4$) with $\theta = 0.5$ (a), $\theta = 0.85$ (b), $\theta = -0.5$ (c) and $\theta = -0.85$ (d). 21

Figure 2.5 – Scatterplots from samples of size 1000 taken from three different Archimedean copulas under the same level of dependence ($\tau = 0.6$): Clayton copula with $\theta = 3$ (a), Frank copula with $\theta = 8$ (b) and Gumbel-Hougaard copula with $\theta = 2.5$ (c). 23

Figure 2.6 – Scatterplots from samples of size 1000 taken from the Gumbel-Hougaard copula with $\theta = 3$ (a) and the Galambos copula with $\theta = 2.3$ (b) under the same level of dependence ($\tau = 2/3$). 24

Figure 2.7 – A D-vine with 3 variables. 28

Figure 2.8 – A D-vine with 4 variables. 28

Figure 2.9 – A C-vine with 4 variables. 29

Figure 3.1 – Survival curves of Weibull model with $\lambda = 0.3$ and $\rho = 2.5$ (solid); $\lambda = 1.7$ and $\rho = 1$ (dashed); $\lambda = 3$ and $\rho = 0.4$ (dotted). 39

Figure 3.2 – Hazard rates of Weibull model with $\lambda = 0.3$ and $\rho = 2.5$ (solid); $\lambda = 1.7$ and $\rho = 1$ (dashed); $\lambda = 3$ and $\rho = 0.4$ (dotted). 41

Figure 1 – Scatterplots from samples taken from a Gaussian Factor copula with $\theta = 0$ (Independence (a)), 0.556 ($\tau = 0.2$ (b)), 0.767 ($\tau = 0.4$ (c)) and 0.899 ($\tau = 0.6$ (d)). 85

Figure 2 – Scatterplots from samples taken from a Clayton Factor copula with $\theta = 0$ (Independence (a)), 1.07 ($\tau = 0.2$ (b)), 2.383 ($\tau = 0.4$ (c)) and 4.816 ($\tau = 0.6$ (d)). 86

Figure 3 – Scatterplots from samples taken from a Galambos factor copula with $\theta = 0$ (Independence (a)), 0.866 ($\tau = 0.2$ (b)), 1.538 ($\tau = 0.4$ (c)) and 2.78 ($\tau = 0.6$ (d)). 88

LIST OF TABLES

Table 2.1 – Some important Archimedean copulas and their generators. 22

CONTENTS

FIRST PART - OVERVIEW 9
1 INTRODUCTION 9
2 COPULAS 11
2.1 Sklar's theorem: copulas in statistics 13
2.2 Some important parametric copula families 18
2.2.1 Gaussian copulas 18
2.2.2 Student's t copulas 20
2.2.3 Archimedean copulas 20
2.2.4 Extreme-value copulas 22
2.3 Multivariate copulas 24
2.4 Vine copulas 25
2.5 Factor copula models 29
2.5.1 One- and two-factor copulas 31
3 SURVIVAL ANALYSIS 38
3.1 Basic concepts and terminology 38
3.2 Maximum likelihood estimation under right censoring 41
3.2.1 Non-parametric estimation of the survival function - The Kaplan-Meier estimator 43
3.3 Regression models for survival data 43
3.3.1 The accelerated failure-time model 44
3.3.2 Cox proportional hazards model 45
3.4 Multivariate survival analysis 47
3.4.1 Frailty model 47
4 CONCLUSIONS AND PERSPECTIVES 50
REFERENCES 51
SECOND PART - PAPER 53
PAPER - Factor copula models for right-censored clustered survival data . . 53
APPENDIX A – Additional topics and proofs 81
APPENDIX B – R routines 91

FIRST PART - OVERVIEW

1 INTRODUCTION

Dependence modelling has always been an important matter in applied statistics, since in most of the cases the independence assumption is violated. Early multivariate models were mostly based on Gaussianity, therefore implying a linear dependence structure and also Gaussian marginals. Despite the increasing number of multivariate families of distributions over the course of time, it was not possible, until 1959, to model dependence structure and marginals separately, that is, the choice of a particular family of multivariate distribution implied in fixed marginals. This scenario started to change with the groundbreaking work of Abe Sklar, entitled *Fonctions de répartition à n dimensions et leurs marges* (SKLAR, 1959). In his work, Sklar introduced copula functions and announced a theorem that made it possible to split multivariate models in two parts: dependence structure (copulas) and marginal distributions. Hence, the problem of multivariate modelling became much simpler, allowing for an unconstrained choice of marginals and dependence structures (copulas), thus creating flexible models.

Since the work of Sklar, copulas have been widely used in many applied fields, specially survival analysis, economics and finance. The increasing popularity of copulas is due to a plethora of available parametric families and the possibility to combine them with any marginal distributions, making it possible to efficiently model data with complicated dependence structures, such as asymmetric and tail dependence.

Despite the large number of multivariate copula families, there were still some flexibility issues, particularly in high dimensions, because the choice of a particular copula necessarily implies a unique form of association among pairs of variables. To circumvent this problem, Bedford and Cooke (2002) introduced the vine copula models, based on the pair-copula construction (PCC) method that uses bivariate copulas as building blocks for higher dimensional distributions. However, the elevated number of parameters in such models is a major drawback for inference.

As a way to reduce the number of parameters, preserving the flexible aspect of vine copulas, Krupskii and Joe (2013) proposed the factor copula models, that, similarly to vine copulas, are multivariate models composed of bivariate copulas, but with a significant difference: they allow the reduction of the number of parameters from $O(d^2)$ to $O(d)$. With this reduction, efficient estimation methods for high dimensions can be more easily derived and the computational costs are considerably lowered.

In multivariate survival analysis, specially with clustered data, in which the different event times are linked to each other through a complicated association such as a family/sibling structure, copula models are becoming an increasingly popular alternative to the traditional frailty models. However, significant technical restrictions have been imposed in their usage. Dating from 1995, when Shih and Louis (1995) proposed a copula formulation for bivariate survival data, until today, copula models in survival analysis have their use restricted to settings where either the cluster size is fixed or the options of copula families are limited. The fine work of Prenen, Braekers and Duchateau (2017) made it possible to use Archimedean copulas to model survival data grouped in clusters of variable size. However, many important families of copulas were not comprehended by their model, e.g., those of the elliptical class (Gaussian, t), extreme-value copulas (except for the also Archimedean, Gumbel-Hougaard copula), etc. In view of this, the possibilities for dependence modelling are still limited.

Considering the shortcomings of the current methodologies for dependence modelling in survival analysis, this thesis aims to make a contribution towards the modelling of clustered survival data by copula models. For this, we propose a new class of models based on the flexible factor copulas. The challenge is to adapt the factor copula model framework to accommodate for the special features of clustered survival data, such as right-censoring and varying cluster size. In this sense, we hope to make our contribution to the field of multivariate survival analysis by developing methods that impose no restrictions for dependence modelling, allowing any copula family in their formulation and also benefiting from the reduced number of parameters of factor copulas, while being able to handle settings where the clusters have varying size.

As an additional product of this thesis, our objective is to derive computationally efficient algorithms for implementations of the proposed methods. In the attempt to do this, we provide computational routines in `R` (R Core Team, 2018) along with a detailed guideline for their usage.

This work is divided in two parts: the first consists of a technical overview on copula models and survival analysis. In the second part, we present the paper entitled "*Factor copula models for right-censored clustered survival data*", submitted for publication (in revision after peer review) in the journal *Lifetime Data Analysis*, where we tackle the main task of this thesis by developing a new methodology for clustered survival data modelling. Some additional topics and technical proofs related to the paper are given in Appendix A. The computational routines in `R` language and a step-by-step guide for their use can be found in Appendix B.

2 COPULAS

Before the advent of copulas, multivariate statistics mostly dealt with models based on Gaussianity, which in many applications is a limiting condition. Also, there were not many alternative approaches, due to the lack of flexibility of existing multivariate distribution families in representing different bivariate dependence structures. This scenario started to change in 1959, with the work of Abe Sklar, *Fonctions de répartition à n dimensions et leurs marges* (SKLAR, 1959). According to Nelsen (2007), Sklar introduced in statistics a class of functions named copulas, whose importance lies in a theorem stating that there exists, in every multivariate distribution, a function that *links* marginal distributions to their joint distribution. This theorem, known as Sklar's theorem, is the foundation for the use of copulas in statistics, since it makes possible to split a multivariate model in two parts: marginal distributions and dependence structure (copulas).

First applications of copulas were proposed in survival analysis (biostatistics, reliability, actuarial science), but eventually, all applied fields began to use copulas in dependence modelling, specially in finance and economics, where copulas became very popular due to the variety of families capable of dealing with nonlinear associations (FERMANIAN, 2017).

We now proceed by formally defining a bivariate copula and presenting some of its fundamental results. The following definitions, theorems and additional results are taken from the book "*An Introduction to copulas*" of Roger Nelsen (NELSEN, 2007).

Definition 2.1. *A bivariate copula is a function* $C : [0,1]^2 \to [0,1]$, $(u,v) \mapsto C(u,v)$, *with the following properties:*

1. *C is grounded, i.e., for every* u,v *in* $[0,1]$, $C(u,0) = 0 = C(0,v)$;

2. *For every* u,v *in* $[0,1]$, $C(u,1) = u$ *and* $C(1,v) = v$;

3. *C is* 2*-increasing, i.e., for every* u_1,u_2,v_1,v_2 *in* $[0,1]$, *such that* $u_1 \leq u_2$ *and* $v_1 \leq v_2$,

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0.$$

As for now, we have only defined a copula as a function with some particular properties, but its relationship with statistics is yet unclear. This connection will be shown later. Although bivariate copulas are defined in the unit cube, the next theorem states that there is a region where every copula is contained.

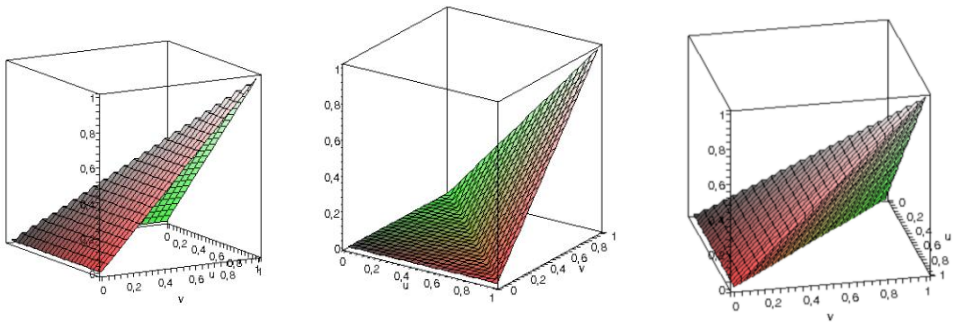
Theorem 2.1. *Let $C(u,v)$ be a copula. Then for every (u,v) in $[0,1]^2$,*

$$\max(u+v-1,0) \leq C(u,v) \leq \min(u,v). \tag{2.1}$$

The above result is the copula version of the *Fréchet-Hoeffding bounds* and establish lower and upper bounds for copulas. We will denote the *Fréchet-Hoeffding lower bound* as C_L and the *Fréchet-Hoeffding upper bound* as C_U. It can be shown that $C_L(u,v) = \max(u+v-1,0)$ and $C_U(u,v) = \min(u,v)$ are also copulas.

Owing to the Fréchet-Hoeffding bounds in (2.1), we can graphically represent the region that contains all copulas (Figure 2.1 (c)), which is the region between C_L (Figure 2.1 (b)) and C_U (Figure 2.1 (a)).

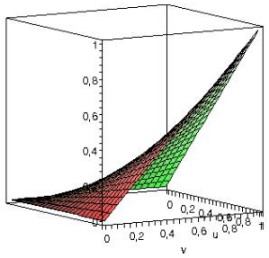
Figure 2.1 – Graphs of copulas C_U (a), C_L (b) and region that contains all copulas (c)



(a) $C_U(u,v)$ (b) $C_L(u,v)$ (c) Region between C_L and C_U

Another fundamental copula is the *product copula* $\Pi(u,v) = uv$ (Figure 2.2). As we will show further in this section, its importance lies on the fact that it characterises independence between random variables.

Figure 2.2 – Graph of the copula Π.



The next two results are of great importance, because they explore the properties of the partial derivatives of copulas. These are constantly used in copula construction methods. Besides, in order to derive the likelihood of copula based models, it is a necessary condition that $\frac{\partial^2}{\partial u \partial v} C(u,v)$ can be computed.

Theorem 2.2. *Let $C(u,v)$ be a copula. Then for every u,v in $[0,1]$, the partial derivatives $\frac{\partial}{\partial u} C(u,v)$ and $\frac{\partial}{\partial v} C(u,v)$ exists almost everywhere in $[0,1]$, also*

$$0 \leq \frac{\partial}{\partial u} C(u,v) \leq 1,$$
$$0 \leq \frac{\partial}{\partial v} C(u,v) \leq 1.$$

Theorem 2.3. *Let $C(u,v)$ be a copula. If $\frac{\partial}{\partial v} C(u,v)$ and $\frac{\partial^2}{\partial u \partial v} C(u,v)$ are continuous on $[0,1]^2$ and $\frac{\partial}{\partial u} C(u,v)$ exists for all $u \in (0,1)$ when $v = 0$, then $\frac{\partial}{\partial u} C(u,v)$ and $\frac{\partial^2}{\partial v \partial u} C(u,v)$ exist in $(0,1)^2$ and $\frac{\partial^2}{\partial u \partial v} C(u,v) = \frac{\partial^2}{\partial v \partial u} C(u,v)$.*

We are now able to explain how copula functions were introduced in statistics. The next section gives an overview on copulas as statistical models, beginning with the most important theorem that made it possible to use copula functions in a statistical context.

2.1 Sklar's theorem: copulas in statistics

Before baptised as copulas, the class of functions defined previously was already being studied from a mathematical perspective. However, it was only after the work of Sklar (1959) that these functions became popular in the scientific community (NELSEN, 2007).

Abe Sklar showed that a multivariate distribution can be written in terms of their marginal distributions and a function that joins the latter to their multivariate form. These functions were named copulas (from the Latin *copulæ*), due to their role in establishing the *link* between marginal distributions and their joint distribution, as it is presented in the theorem below, which now bears Sklar's name.

Theorem 2.4 (Sklar's Theorem)**.** *Let X and Y be continuous random variables with joint distribution $F_{X,Y}(x,y)$ and cumulated marginals $F_X(x)$ and $F_Y(x)$. Then there exists a unique copula C such that, for all $x,y \in \mathbb{R}$,*

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)). \tag{2.2}$$

Conversely, if C is a copula and $F_X(x)$, $F_Y(y)$ are continuous distribution functions, then $F_{X,Y}(x,y)$ as defined in (2.2) is a joint distribution function and it is unique.

Sklar's theorem elucidates how copulas are related to statistics/probability, by *joining* multivariate distributions to their univariate margins. In other words, Sklar's theorem makes possible to split a multivariate joint distribution in two parts: marginal distributions and a copula, the former representing individual behaviour of the random variables (RVs) and the latter as a model for the dependence structure of the RVs. Henceforth, we will often refer to the copula of X and Y as $C_{X,Y}$. Copulas can also be defined as joint distributions with standard uniform marginal distributions, that is $C(u,v) = P(U \leq u, V \leq v)$, where U and $V \sim \text{Uniform}[0,1]$.

From a practical point of view, given a set of random variables with fixed marginals, one can build a plethora of joint distributions by using different copula functions. These can be chosen from several parametric families or even constructed as a semi or nonparametric models. Conversely, a copula can be obtained if the joint distribution and its marginals are known:

Corollary 2.1. *Let X, Y, $F_{X,Y}$, F_X, F_Y and C be such as defined in Sklar's theorem. Then, for all $(u,v) \in [0,1]^2$,*

$$C_{X,Y}(u,v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)).$$

We show next how the copula $\Pi(u,v) = uv$ characterises independence between random variables.

Theorem 2.5. *Continuous random variables X and Y are independent if and only if $C_{X,Y}(u,v) = \Pi(u,v)$*

Proof. If X and Y are independent, their joint distribution is $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, then, by Corollary 2.1

$$\begin{aligned} C_{X,Y}(u,v) &= F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)) = F_X(F_X^{-1}(u))F_Y(F_Y^{-1}(v)) \\ &= uv = \Pi(u,v). \end{aligned}$$

Conversely, if $C_{X,Y}(u,v) = \Pi(u,v) = uv$, then, by Sklar's theorem, their joint distribution is

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)) = F_X(x)F_Y(y).$$

Since the joint distribution of X and Y is the product of their marginals, then, by definition, X and Y are independent. □

A useful property of a copula is its invariance under strictly increasing transformations. In general, under strictly monotone transformations, it is possible and rather simple to obtain the copula of the transformed variables if the original copula is known.

Theorem 2.6. *Let X and Y be continuous RVs with copula $C_{X,Y}$. Also, let α and β be strictly monotone transformations.*

***1**. If α and β are both strictly crescent transformations,*

$$C_{\alpha(X),\beta(Y)} = C_{X,Y}.$$

***2**. If α and β are both strictly decreasing,*

$$C_{\alpha(X),\beta(Y)}(u,v) = u+v-1+C_{X,Y}(1-u,1-v).$$

***3**. If α is strictly crescent and β is strictly decreasing,*

$$C_{\alpha(X),\beta(Y)}(u,v) = u-C_{X,Y}(u,1-v).$$

***4**. If α is strictly decreasing and β is strictly crescent,*

$$C_{\alpha(X),\beta(Y)}(u,v) = v-C_{X,Y}(1-u,v).$$

Proof. We will proof only the first two statements, since the other two cases can be similarly proved. Let $F_X, F_Y, F_{\alpha(X)}$ and $F_{\beta(Y)}$ be the distribution functions of $X, Y, \alpha(X)$ and $\beta(Y)$, respectively.

1. If $\alpha(X)$ is a strictly increasing transformation, then

$$F_{\alpha(X)}(x) = P[\alpha(X) \leq x] = P[X \leq \alpha^{-1}(x)] = F_X(\alpha^{-1}(x)).$$

The same goes for $F_{\beta(Y)} = F_Y(\beta^{-1}(y))$. This way,

$$\begin{aligned} C_{\alpha(X),\beta(Y)}(F_{\alpha(X)}(x),F_{\beta(Y)}(y)) &= F_{\alpha(X),\beta(Y)}(x,y) = P[\alpha(X) \leqslant x, \beta(Y) \leqslant y] \\ &= P[X \leqslant \alpha^{-1}(x), Y \leqslant \beta^{-1}(y))] = F_{X,Y}(\alpha^{-1}(x),\beta^{-1}(y)) \\ &= C_{X,Y}(F_X(\alpha^{-1}(x)),F_Y(\beta^{-1}(y))) \\ &= C_{X,Y}(F_{\alpha(X)}(x),F_{\beta(Y)}(y)) \\ &\Rightarrow C_{\alpha(X),\beta(Y)}(u,v) = C_{X,Y}(u,v). \end{aligned}$$

2. For $\alpha(X)$ and $\beta(Y)$ strictly decreasing,

$$\begin{aligned} C_{\alpha(X),\beta(Y)}(F_{\alpha(X)}(x),F_{\beta(Y)}(y)) &= F_{\alpha(X),\beta(Y)}(x,y) = P[\alpha(X) \leqslant x, \beta(Y) \leqslant y] \\ &= P[X > \alpha^{-1}(x), Y > \beta^{-1}(y)) = 1 - P[X \leqslant \alpha^{-1}(x)] \\ &\quad - P[Y \leqslant \beta^{-1}(y)] + P[X \leqslant \alpha^{-1}(x), Y \leqslant \beta^{-1}(y)] \\ &= 1 - F_X(\alpha^{-1}(x)) - F_Y(\beta^{-1}(y)) + F_{X,Y}(\alpha^{-1}(x),\beta^{-1}(y)) \\ &= 1 - F_{\alpha(X)}(x) - F_{\beta(Y)}(y) + C_{X,Y}(F_{\alpha(X)}(x),F_{\beta(Y)}(y)) \\ &\Rightarrow C_{\alpha(X),\beta(Y)}(u,v) = 1 - u - v + C_{X,Y}(u,v). \end{aligned}$$

□

It was already discussed why copulas are models for dependence structures of random variables, but how the intensity of the dependence can be measured? We will show next that this can be done through the parameters of the copula, by establishing a direct relationship to some measures of association, in particular, the Kendall's τ.

Definition 2.2. *Let (X_1,Y_1) and (X_2,Y_2) be independent copies of a random vector (X,Y) whose joint cumulative distribution function is $F_{X,Y}$. The Kendall's τ of the pair (X,Y) is defined as*

$$\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

The Kendall's τ is a measure of association based on the concepts of concordance and discordance of pairs. For example, two pairs (x_i,y_i) and (x_j,y_j) are said to be concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$. Note that, indeed, the Kendall's τ in the above definition measures the probability of concordance, $\mathtt{P}_C = P[(X_1 - X_2)(Y_1 - Y_2) > 0]$, minus the probability of discordance, $\mathtt{P}_D = P[(X_1 - X_2)(Y_1 - Y_2) < 0]$. In this sense, a random

pair (X,Y) has a high value of Kendall's τ if it is also high the probability of concordance, $\mathtt{P}_C = P[X_2 > X_1, Y_2 > Y_1] + P[X_2 < X_1, Y_2 < Y_1]$. In other words, the Kendall's $\tau_{X,Y}$ is a value between -1 (perfect discordance) and $+1$ (perfect concordance) that measures the strength of monotonic association between X and Y. It is important to note that, if X and Y are independent, then $\tau_{X,Y} = 0$, but the the converse is not necessarily true.

Theorem 2.7. *Let X and Y be continuous RVs with copula C. Then, the Kendall's tau of X and Y ($\tau_{X,Y}$) can be expressed as*

$$\tau_{X,Y} = 4\int_0^1 \int_0^1 C(u,v)dC(u,v) - 1. \tag{2.3}$$

Proof. From Definition 2.2, we have that

$$\tau_{X,Y} = \mathtt{P}_C - \mathtt{P}_D = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0],$$

where (X_1, Y_1) and (X_2, Y_2) are independent copies of (X,Y). Also, the probability of discordance can be expressed as $\mathtt{P}_D = 1 - \mathtt{P}_C$. Then

$$\tau_{X,Y} = 2\mathtt{P}_C - 1.$$

The probability of concordance

$$\mathtt{P}_C = P[(X_1 - X_2)(Y_1 - Y_2) > 0] = P[X_1 < X_2, Y_1 < Y_2] + P[X_1 > X_2, Y_1 > Y_2]$$

can be rewritten as

$$\mathtt{P}_C = 2P[X_1 < X_2, Y_1 < Y_2],$$

since $P[X_1 > X_2, Y_1 > Y_2] = P[X_2 > X_1, Y_2 > Y_1]$. Therefore,

$$\tau_{X,Y} = 4P[X_1 < X_2, Y_1 < Y_2] - 1.$$

Furthermore, by expressing $P[X_1 < X_2, Y_1 < Y_2]$ in terms of the copula of X and Y, we get

$$\begin{aligned}
P[X_1 < X_2, Y_1 < Y_2] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{x_2}\int_{-\infty}^{y_2} f_{X_1,Y_1,X_2,Y_2}(x_1,y_1,x_2,y_2)dx_1dy_1dx_2dy_2 \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{x_2}\int_{-\infty}^{y_2} f_{X_1,Y_1}(x_1,y_1)f_{X_2,Y_2}(x_2,y_2)dx_2dy_2dx_1dy_1 \\
&= \iint_{\mathbb{R}^2} F_{X,Y}(x,y)f_{X,Y}(x,y)dxdy \\
&= \iint_{\mathbb{R}^2} F_{X,Y}(x,y)dF_{X,Y}(x,y) \\
&= \iint_{\mathbb{R}^2} C(F_X(x),F_Y(y))dC(F_X(x),F_Y(y)) \\
&= \int_0^1\int_0^1 C(u,v)dC(u,v)
\end{aligned}$$

Hence $\tau_{X,Y} = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1$. □

The above result states that copulas, besides representing different types of dependence structures, also contains the information of the strength of association between random variables (Kendall's tau). Moreover, both the dependence structure and the strength of association are represented by the copula alone.

2.2 Some important parametric copula families

In this section we give a brief overview on some of the most popular bivariate copula families in the literature, namely: Gaussian copulas, Student's t copulas, Archimedean copulas and extreme-value copulas. These models have been extensively used in practical applications due to their good properties and simple formulations.

2.2.1 Gaussian copulas

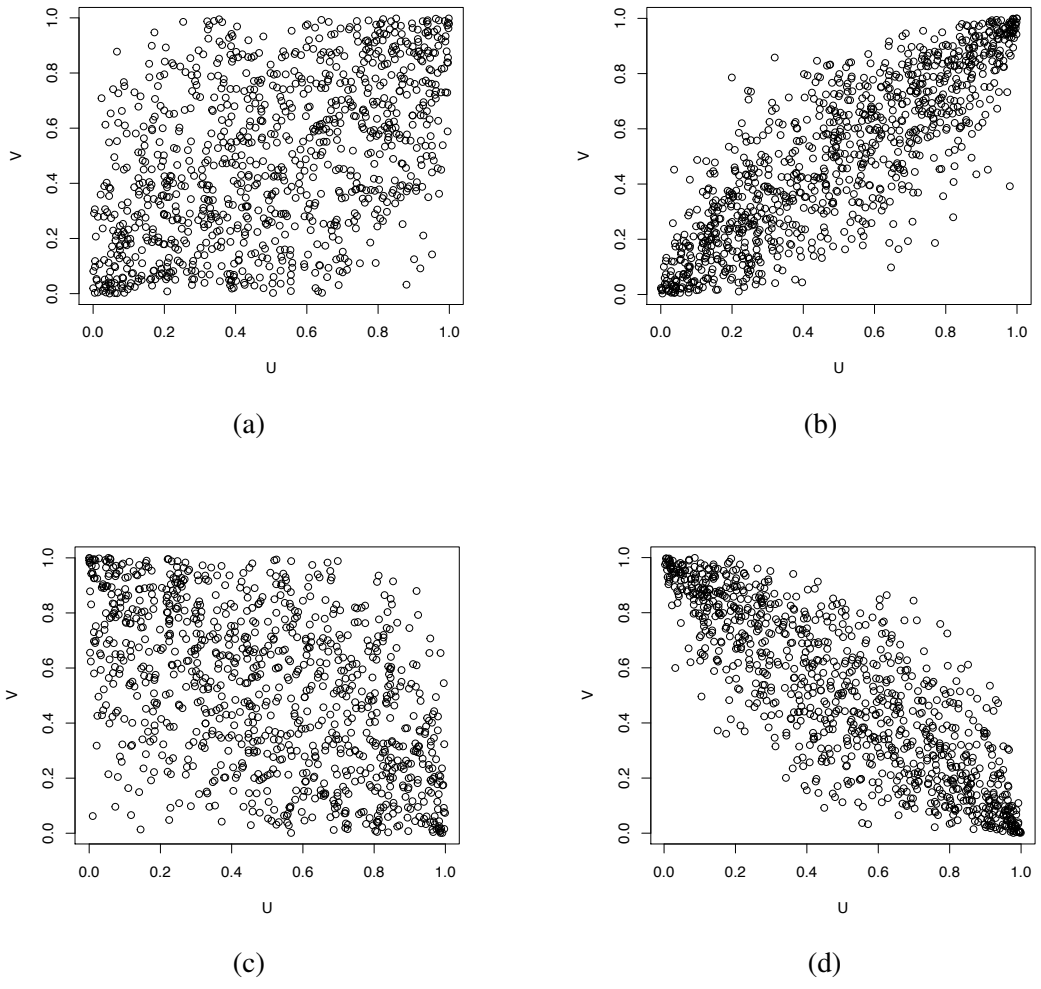
The Gaussian copula, or normal copula, is a model for linear dependence between random variables. It can be obtained by using the result in Corollary 2.1 on a bivariate normal

distribution, i.e.,

$$C_{Ga}(u,v) = \Phi_2(\Phi^{-1}(u),\Phi^{-1}(v);\theta)$$
$$= \frac{1}{2\pi\sqrt{1-\theta^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left(-\frac{s^2-2\theta st+t^2}{2(1-\theta^2)}\right) dsdt, \quad -1 \leq \theta \leq +1.$$

It is easy to see that the Gaussian copula's parameter θ is, in fact, the Pearson's correlation coefficient ρ. Hence, it represents the strength of linear dependence between the random variables U and V. In Figure 2.3, we illustrate the shape of the dependence imposed by the Gaussian copula under different parametric settings.

Figure 2.3 – Scatterplots from samples of size 1000 taken from a Gaussian copula with $\theta = 0.5$ (a), $\theta = 0.85$ (b), $\theta = -0.5$ (c) and $\theta = -0.85$ (d).



(a) (b)

(c) (d)

By using Sklar's theorem with the Gaussian copula it is possible to derive the standard bivariate normal distribution. This can be done by setting $u = \Phi(x)$ and $v = \Phi(y)$ in $C_{Ga}(u,v)$. Moreover, a different choice of copula with the same standard normal marginals would build a different joint distribution, that is, we can create several joint distributions with standard normal marginals that are not the bivariate standard normal distribution. For $C = \Pi$ we also end up with a bivariate standard normal distribution but with independent margins:

$$\begin{aligned}\Pi(\Phi(x),\Phi(y)) &= \Phi(x)\Phi(y)\\ &= \frac{1}{2\pi}e^{-\frac{(x^2+y^2)}{2}}.\end{aligned}$$

2.2.2 Student's t copulas

This family of copulas can be obtained through a procedure similar to the Gaussian's. Let $t_{\nu,\theta}$ be the bivariate Student's t distribution with central Student's t margins, t_ν. Then, by using Corollary 2.1, the Student's t copula is given by

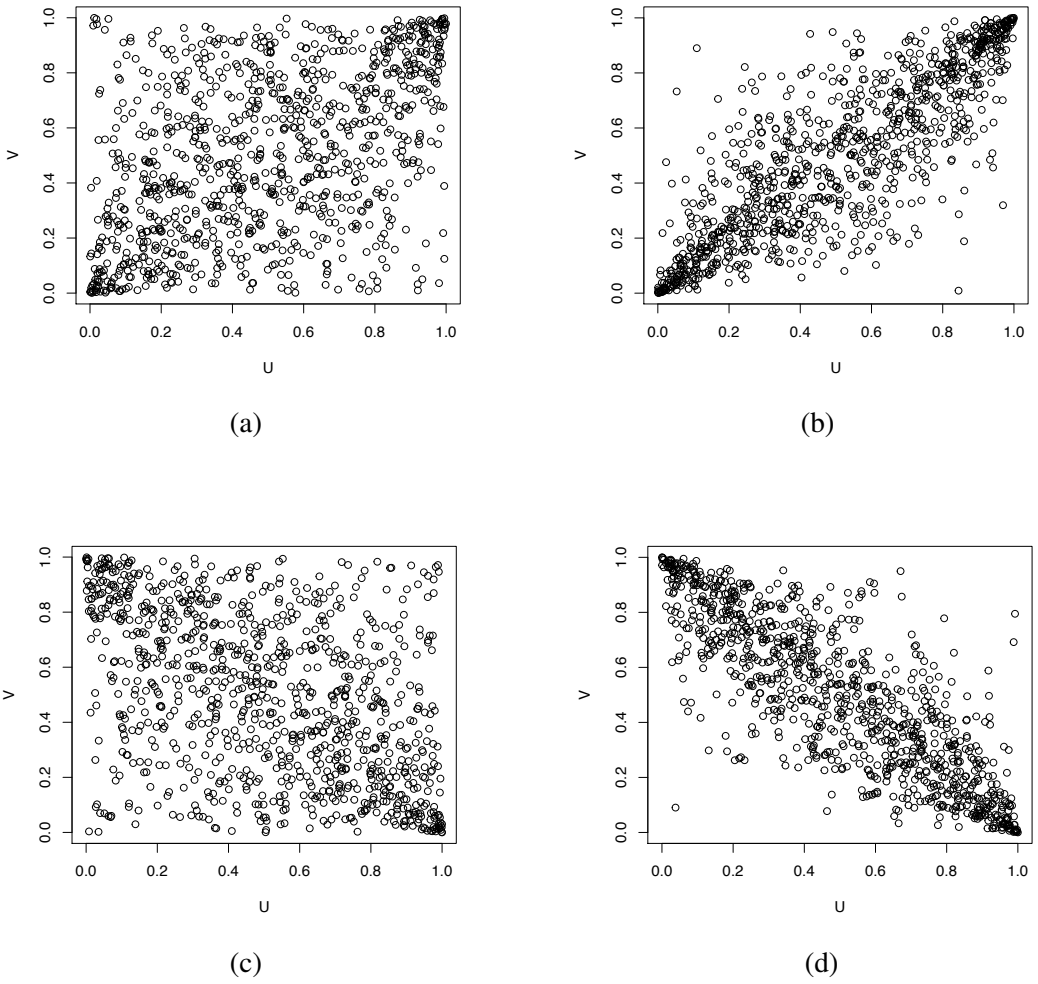
$$\begin{aligned}C_t(u,v;\nu,\theta) &= t_{\nu,\theta}(t_\nu^{-1}(u),t_\nu^{-1}(v))\\ &= \frac{1}{2\pi\sqrt{1-\theta^2}}\int_{-\infty}^{t_\nu^{-1}(u)}\int_{-\infty}^{t_\nu^{-1}(v)}\left(1+\frac{s^2-2\theta st+t^2}{\nu(1-\theta^2)}\right)^{-\frac{\nu+2}{2}}dsdt, \quad -1\le\theta\le+1, \nu>0.\end{aligned}$$

The dependence structure imposed by the Student's t copula, or simply t copula, is shown in Figure 2.4 under different values of θ. This copula differs from the Gaussian due to an increased association in the tails, that is, it represents a dependence structure in which the probability of joint occurrence of large/small values of the RVs is higher than in the Gaussian copula case. This can be noticed by comparing the concentration of data points in the tails in Figure 2.4 and Figure 2.3.

2.2.3 Archimedean copulas

Archimedean copulas are a general class that comprehends several copula families. These families are constructed via a generator ϕ, a continuous and decreasing convex function $\phi : [0,1] \to [0,+\infty]$, satisfying $\phi(1) = 0$. If $\phi(0) = +\infty$, ϕ is a strict generator and has

Figure 2.4 – Scatterplots from samples of size 1000 taken from a t copula ($\nu = 4$) with $\theta = 0.5$ (a), $\theta = 0.85$ (b), $\theta = -0.5$ (c) and $\theta = -0.85$ (d).



(a) (b)

(c) (d)

inverse ϕ^{-1}. For $\phi(0) < +\infty$, we have to define a pseudo-inverse (NELSEN, 2007)

$$\phi^{[-1]} = \begin{cases} \phi^{-1} & , 0 \leq t \leq \phi(0) \\ 0 & , \phi(0) \leq t \leq +\infty \end{cases}, \tag{2.4}$$

such that $\phi^{[-1]}(\phi(t)) = t$, $\forall t \in [0,1]$.

Definition 2.3. *Let ϕ be a generator function and $\theta \in \mathbb{R}$. An Archimedean copula is defined as*

$$C_A(u,v;\theta) = \phi_\theta^{[-1]}(\phi_\theta(u) + \phi_\theta(v)). \tag{2.5}$$

The use of the generator function in the construction of Archimedean copulas makes possible to create more simple relations to some quantities, such as the Kendall's tau. Indeed, Genest and MacKay (1986) showed that the expression in 2.3 can be written in the form

$$\tau_{C_A} = 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt + 1.$$

Some of the most popular Archimedean copula families are given in Table 2.1. For more details on Archimedean copulas and an extensive list of families of such class, the reader is referred to Nelsen (2007).

Table 2.1 – Some important Archimedean copulas and their generators.

Family	$\phi_\theta(t)$	$\theta \in$	$C_\theta(u,v)$
Gumbel-Hougaard	$(-\ln t)^\theta$	$[1,+\infty)$	$\exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}}\}$
Clayton	$\frac{1}{\theta}(t^{-\theta}-1)$	$[-1,+\infty)\backslash\{0\}$	$[\max(u^{-\theta}+v^{-\theta}-1,0)]^{-\frac{1}{\theta}}$
Ali-Mikhail-Haq	$\ln\frac{1-\theta(1-t)}{t}$	$[-1,1)$	$\frac{uv}{1-\theta(1-u)(1-v)}$
Frank	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$(-\infty,\infty)\backslash\{0\}$	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$

It is possible to find a reasonable number of families in the class of Archimedean copulas with distinct dependence structures. In Figure 2.5, we illustrate this feature by providing the scatterplots from samples of different Archimedean copulas. One can see (Figure 2.5), for example, that the Gumbel-Hougaard copula induces a more accentuated association at the upper tail in comparison to the Frank and Clayton. Indeed, as we show in the next section, the Gumbel-Hougaard copula also belongs to the family of extreme-value copulas, which characterises its particular feature. The Clayton copula on the other hand imposes a stronger association at the lower tail.

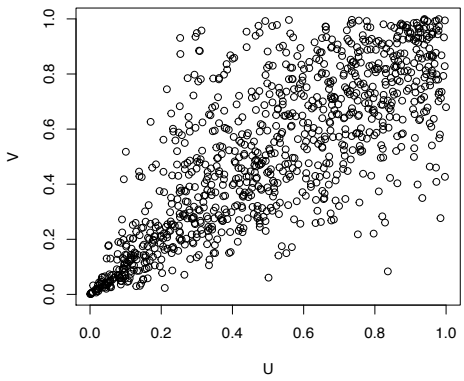
2.2.4 Extreme-value copulas

Extreme-value copulas are one of the most popular in the fields of finance and actuarial risk analysis. They arise naturally as the limits of copulas of componentwise maxima in independent random samples and represent dependence structures for rare events (GUDENDORF; SEGERS, 2010).

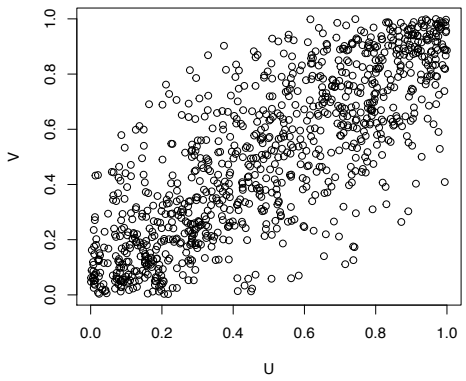
Definition 2.4. *A copula C is an extreme-value copula if there exists a copula C_* such that*

$$C(u,v) = \lim_{n\to+\infty}\left[C_*(u^{1/n}, v^{1/n})\right]^n, \quad \forall (u,v)\in[0,1]^2. \tag{2.6}$$

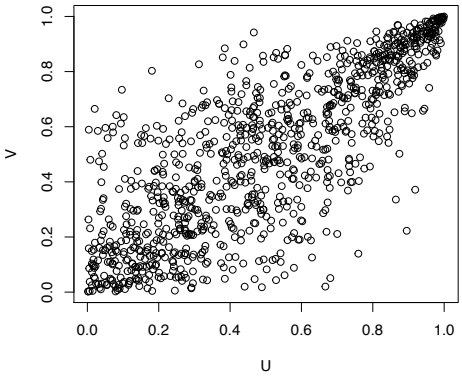
Figure 2.5 – Scatterplots from samples of size 1000 taken from three different Archimedean copulas under the same level of dependence ($\tau = 0.6$): Clayton copula with $\theta = 3$ (a), Frank copula with $\theta = 8$ (b) and Gumbel-Hougaard copula with $\theta = 2.5$ (c).



(a)



(b)



(c)

It is said that C_* is in the *domain of attraction* of C.

The importance of extreme-value copulas lies in their capability of modelling rare (extreme) events that are dependent on each other (higher probability of joint occurrence of unusually large values). The use of correct models for these type of events in finance and actuarial risk management, for example, is extremely important, since their effects could be catastrophic, e.g., large financial losses.

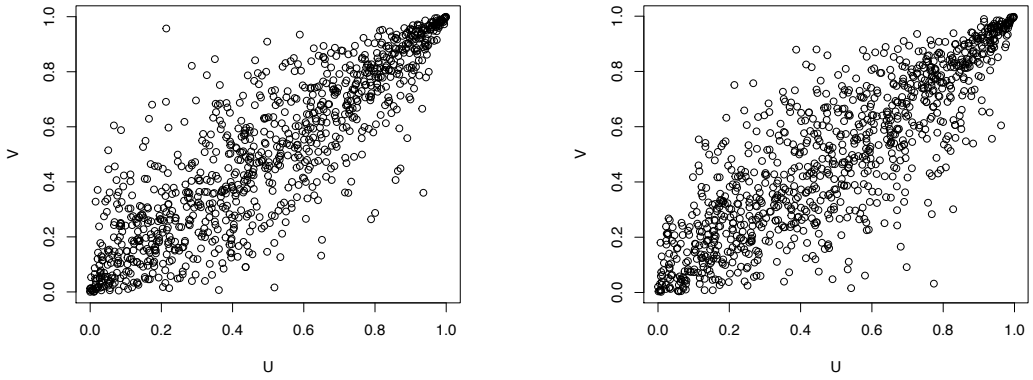
Two examples of extreme-value copulas are the Gumbel-Hougaard and the Galambos copulas, with the former also belonging to the class of Archimedean copulas (see Table 2.1).

The Galambos copula is defined as (JOE, 2014)

$$C(u,v;\theta) = uv\exp\left\{\left[(-\log u)^{-\theta} + (-\log v)^{-\theta}\right]^{-1/\theta}\right\}, \quad 0 \leq \theta < +\infty.$$

For the sake of illustration, we give in Figure 2.6 the scatterplots of samples taken from the Gumbel-Hougaard and the Galambos copulas. As is the characteristic of extreme-value copulas, one can note that the sample points are more concentrated in the upper tails.

Figure 2.6 – Scatterplots from samples of size 1000 taken from the Gumbel-Hougaard copula with $\theta = 3$ (a) and the Galambos copula with $\theta = 2.3$ (b) under the same level of dependence ($\tau = 2/3$).



(a) (b)

2.3 Multivariate copulas

So far, we have defined copulas as two-dimensional functions, but it is possible to naturally extend the definition to d dimensions Nelsen (2007):

Definition 2.5. *A multivariate copula is a function* $C : [0,1]^d \to [0,1]$, $(u_1,...,u_d) \mapsto C(u_1,...,u_d)$, *with the following properties:*

1. *C is grounded, i.e., if at least one* $u_j = 0$ $j = 1,2,...,d$, *then* $C(u_1,...,u_{j-1},0,u_{j+1},...,u_d) = 0$;

2. *For every* $u_1,...,u_d$ *in* $[0,1]$, $C(1,...,u_j,...,1) = u_j$;

3. *C is d-increasing, i.e., for every hyperrectangle* $[u_1,v_1] \times [u_2,v_2] \times ... \times [u_d,v_d] \in [0,1]^d$, *such that* $u_j \leq u_{j-1}$ *and* $v_j \leq v_{j-1}$ $j = 2,3,...,d$,

$$\Delta_{u_d}^{v_d}\Delta_{u_{d-1}}^{v_{d-1}}...\Delta_{u_1}^{v_1}C(t_1,...,t_d) \geqslant 0,$$

where $\Delta_{u_j}^{v_j} C(t_1,...,t_d) = C(t_1,...,v_j,...,t_d) - C(t_1,...,u_j,...,t_d)$.

Sklar's theorem is also valid for multivariate copulas. Let $X_1,...,X_d$ be continuous random variables with cumulative distribution functions $F_{X_1},...,F_{X_d}$, respectively, and joint distribution $F_{X_1,...,X_d}$. There is a unique copula C such that $F_{X_1,...,X_d}(x_1,...,x_d) = C(F_{X_1}(x_1),...,F_{X_d}(x_d))$.

Besides Sklar's theorem, many other theorems and definitions have analogous multivariate versions, although, not all do. For a detailed study on multivariate copulas, Nelsen (2007) and Joe (2014) can be consulted.

The reason why we have focused more on bivariate copulas, instead of their d-dimensional formulation, is due to the lack of parametric families of multivariate copulas with flexible dependence structures. While there is a plethora of parametric bivariate copula families, the options of parametric multivariate copulas are limited (COOKE; JOE; AAS, 2011). Moreover, multivariate copulas impose the dependence structure for pairs in a random vector. For this reason, they are not widely used for dependence modelling. Instead, construction methods based on bivariate copulas are preferred for multivariate dependence modelling, as we show in the next sections.

2.4 Vine copulas

Vine copulas are models based on pair-copula construction (PCC) methods that allow high-dimensional models to be built using a cascade of bivariate copulas. This is done by decomposing a joint probability density function in a product of univariate marginals and bivariate copulas (pair-copulas). Hence, the flexibility of such models comes from the possibility of choosing any bivariate copula family to model the pairwise dependencies in a multivariate context. The number of combinations unfolded by the pair-copula construction method is abundant, e.g., it is possible to create a model by combining extreme-value copulas, elliptical copulas (Gaussian and t), Archimedean copulas and so on (COOKE; JOE; AAS, 2011).

Before defining vine copula, we will present some basic definitions and results that will be needed henceforth.

Definition 2.6. *Let* $C_{12}(u_1,u_2)$ *be a copula. Its corresponding density is given by*

$$c_{12}(u_1,u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C_{12}(u_1,u_2).$$

Proposition 2.1. *Let $X_1 \sim F_1$ and $X_2 \sim F_2$ be RVs with joint cumulative distribution function (CDF) $F_{12}(x_1,x_2)$ and copula C_{12}. Then*
(1) The joint probability density function (pdf) of X_1, X_2 is given by

$$f_{12}(x_1,x_2) = c_{12}(F_1(x_1),F_2(x_2))f_1(x_1)f_2(x_2).$$

(2) The conditional pdf of X_2 given X_1 is

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1),F_2(x_2))f_2(x_2).$$

Proof. (1) By Sklar's theorem, $F_{12}(x_1,x_2) = C_{12}(F_1(x_1),F_2(x_2))$, thus

$$\begin{aligned} f_{12}(x_1,x_2) &= \frac{\partial^2 F_{12}}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial F_{12}}{\partial x_2}\right) = \frac{\partial}{\partial x_1}\left(\frac{\partial C_{12}}{\partial F_2}\frac{dF_2}{dx_2}\right) = \frac{dF_1}{dx_1}\frac{dF_2}{dx_2}\frac{\partial^2 C_{12}}{\partial F_1 \partial F_2} \\ &= f_1(x_1)f_2(x_2)\frac{\partial^2 C_{12}}{\partial F_1 \partial F_2} = c_{12}(F_1(x_1),F_2(x_2))f_1(x_1)f_2(x_2). \end{aligned}$$

(2) It follows from (1) and the definition of conditional pdf that

$$f_{2|1}(x_2|x_1) = \frac{f_{12}(x_1,x_2)}{f_1(x_1)} = c_{12}(F_1(x_1),F_2(x_2))f_2(x_2). \qquad \square$$

We now proceed by giving an example of how the pair-copula construction method (vine copulas) works. Let (X_1,X_2,X_3) be a random vector with joint pdf $f_{1,2,3}(x_1,x_2,x_3)$. The idea is to decompose the joint pdf $f_{1,2,3}$ into a product of univariate marginals and bivariate copulas. First, note that

$$f_{123}(x_1,x_2,x_3) = f_{3|12}(x_1|x_2,x_3)f_{2|1}(x_2|x_1)f_1(x_1).$$

Using part (2) of Proposition 2.1:

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1),F_2(x_2))f_2(x_2).$$

The density $f_{3|12}(x_1|x_2,x_3)$ can also be expressed as

$$f_{3|12}(x_3|x_1,x_2) = \frac{f_{123}(x_1,x_2,x_3)}{f_{12}(x_1,x_2)} = \frac{f_{13|2}(x_1,x_3|x_2)f_2(x_2)}{f_{1|2}(x_1,x_2)f_2(x_2)} = \frac{f_{13|2}(x_1,x_3|x_2)}{f_{1|2}(x_1,x_2)}.$$

Before proceeding with the demonstration, we will need some additional notation for the copula that *joins* conditional CDFs, usually called *conditional copula*. For example, in $F_{13|2}(x_1,x_3|x_2) = C(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2))$, the copula C is joining $F_{1|2}(x_1|x_2)$ and $F_{3|2}(x_3|x_2)$ to $F_{13|2}$, thus we will denote it as $C_{13|2}$. This way,

$$f_{13|2}(x_1,x_3|x_2) = c_{13|2}(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2))f_{1|2}(x_1|x_2)f_{3|2}(x_3|x_2).$$

Hence

$$f_{3|12}(x_3|x_1,x_2) = \frac{f_{13|2}(x_1,x_3|x_2)}{f_{1|2}(x_1,x_2)} = c_{13|2}(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2))f_{3|2}(x_3|x_2),$$

but, since $f_{3|2}(x_3|x_2) = c_{23}(F_2(x_2),F_3(x_3))f_3(x_3)$, it follows that

$$\begin{aligned} f_{123}(x_1,x_2,x_3) = {} & c_{13|2}(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2)) \\ & \times c_{23}(F_2(x_2),F_3(x_3))f_3(x_3)c_{12}(F_1(x_1),F_2(x_2)) \\ & \times f_3(x_3)f_2(x_2)f_1(x_1). \end{aligned} \tag{2.7}$$

This factorisation of the joint density f_{123} is called a pair-copula construction and represents a vine copula model, that is, an expression (not unique!) of a multivariate model as products of marginal densities and bivariate conditional and unconditional copulas. The choices for these copulas are unconstrained, giving high flexibility in the dependence modelling.

Two other possible decompositions for f_{123} would be using $C_{12|3}$ or $C_{23|1}$ instead of $C_{13|2}$, resulting in a total of 3 different vine copulas. For $d = 4$ variables, the number of possible decompositions is 24 and, for d variables, the number is $\frac{d!}{2} \times 2^{(n-2)(n-3)/2}$ (MORALES-NAPOLES, 2011). As the number of vines grows large for higher dimensions, in order to help organizing them, Bedford and Cooke (2002) introduced a graphical model called *regular vine*. It allows to represent every pair-copula decomposition by a nested set of connected trees (acyclic connected simple graphs).

Definition 2.7. *A regular vine or R-vine V on d elements is a graph organised as a nested set of $d-1$ trees such that*

1. T_j ($j = 1,...,d-1$) is a connected tree;

2. T_1 has nodes $N_1 = \{1,...,d\}$, and edge set E_1. The nodes of T_j ($j = 2,...,d-1$) are the edges of tree T_{j-1}, i.e., $N_j = E_{j-1}$;

3. Two nodes in T_j ($j = 2,...,d-1$) are connected if the corresponding edges share a node in T_{j-1}.

In Figure 2.7 we give an example of a regular vine representing the decomposition in 2.7 for 3 variables. This particular form of vine is called D-vine (drawable vine), where every node is connected to no more than 2 edges. For the case of 4 variables, two possible vine copulas are illustrated: in Figure 2.8, a D-vine and in Figure 2.9, a C-vine (canonical vine), which is another type of regular vine that, in each tree T_j, a unique node (root of the vine) is connected to exactly $d-j$ edges. This type of vine copula is related to cases where a particular variable is known to govern interactions in the dataset (KJERSTI et al., 2009).

Figure 2.7 – A D-vine with 3 variables.

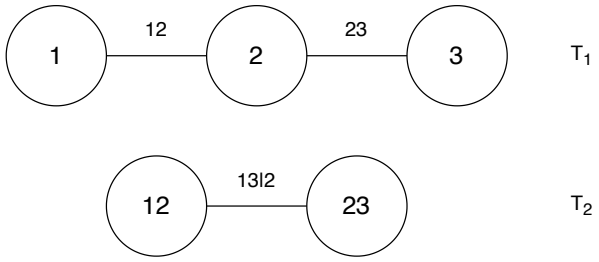
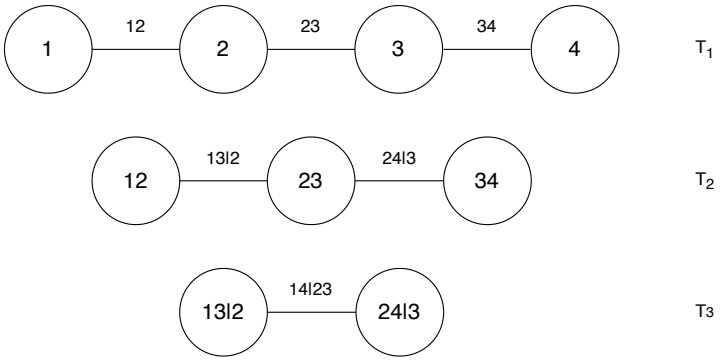
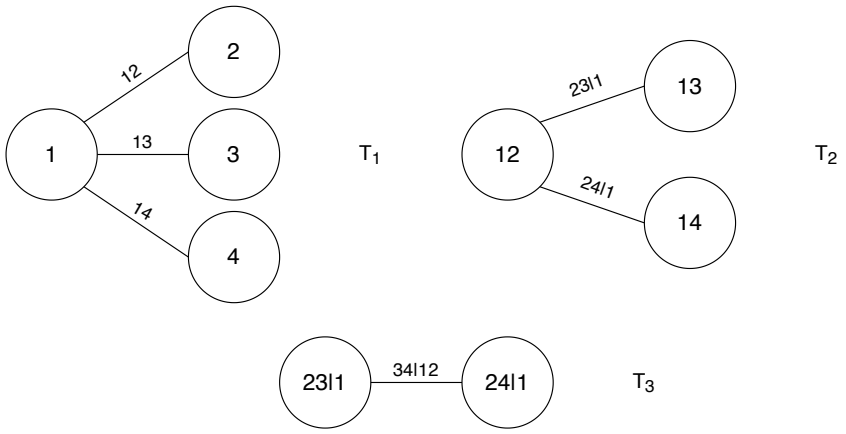


Figure 2.8 – A D-vine with 4 variables.



One major drawback of vine copulas is in the full estimation of the model, consisting of three parts: the first is the specification of the general structure of dependence, or the form of the regular vine, which, according to Kjersti et al. (2009) is, in practice, infeasible for higher dimensions; the second consists of selecting each of the $d(d-1)/2$ (number of edges) bivariate copulas from a plethora of families; the third being the parameter estimation of all $d(d-1)/2$ copulas, another problem due to the elevated number of parameters (of order $O(d^2)$). An alternative approach is to use factor copula models, which are closely related to vine copulas but considerably simpler in terms of estimation.

Figure 2.9 – A C-vine with 4 variables.



2.5 Factor copula models

In multivariate analysis it is usual to build models under the assumption of normality, i.e., correlation-based models. Despite being relatively easy to deal with, those models have the shortcoming of neglecting nonlinear dependencies, such as tail dependencies (OH; PATTON, 2017). Copulas can easily overcome these drawbacks, since they are very general dependence models that embrace both linear (e.g., Gaussian copula) and nonlinear cases.

In many applications, one has to deal with high-dimensional data. In these situations, multivariate copula models may be used to model the dependence structure. However, the number of parametric multivariate copulas is limited, and one major drawback is that they can become inflexible in high dimensions, due to not allowing the specification of pairwise dependence.

Vine copulas are flexible models for high-dimensional data mainly because they allow pairwise modelling of a d-variate dependence structure via $d(d-1)/2$ bivariate copulas. One shortcoming of these models is the high number of dependence parameters (of order $O(d^2)$), which considerably hinders estimations.

Truncated vine copulas or t-truncated vine copulas are an alternative to reduce the number of parameters, since the pair-copulas starting from tree T_{t+1} are all considered to be Π (product copula). Therefore diminishing the number of parametric copulas (JOE, 2014). Another way is to uncover p latent variables $V_1, V_2, ..., V_p$, with p considerably lower than d, such that the d observable variables $U_1, U_2, ..., U_d$ are independent given $V_1, V_2, ..., V_p$, that is,

$$F_{U_1,...,U_d|\mathbf{v}} = \prod_{j=1}^{d} F_{U_j|\mathbf{v}}.$$

Hence, the number of parameters in the multivariate model reduces from $O(d^2)$ to $O(d)$ as we detail next.

Let $U_1, U_2, ..., U_d$ be dependent Uniform$[0,1]$ random variables and $V_1, V_2, ..., V_p$ independent and identically distributed (i.i.d.) Uniform$[0,1]$ latent variables. We assume that the observable random variables $U_1, U_2, ..., U_d$ are conditionally independent given the latent variables $V_1, V_2, ..., V_p$. Let $c_{\mathbf{U},\mathbf{V}}$ be the joint density of $U_1, U_2, ..., U_d$ and $V_1, V_2, ..., V_p$. Then,

$$
\begin{aligned}
c_{\mathbf{U}} &= \int_{[0,1]^p} c_{\mathbf{U},\mathbf{V}}\, dv_1 ... dv_p \\
\frac{\partial^d}{\partial u_1 ... \partial u_d} C_{\mathbf{U}} &= \int_{[0,1]^p} \frac{\partial^{d+p}}{\partial u_1 ... \partial u_d \partial v_1 ... \partial v_p} C_{\mathbf{U},\mathbf{V}}\, dv_1 ... dv_p \\
\int_{[0,u_j]^d} \frac{\partial^d}{\partial u_1 ... \partial u_d} C_{\mathbf{U}}\, du_1 ... du_d &= \int_{[0,u_j]^d} \int_{[0,1]^p} \frac{\partial^{d+p}}{\partial u_1 ... \partial u_d \partial v_1 ... \partial v_p} C_{\mathbf{U},\mathbf{V}}\, dv_1 ... dv_p\, du_1 ... du_d \\
C_{\mathbf{U}}(\mathbf{u}) &= \int_{[0,u_j]^d} \frac{\partial^d}{\partial u_1 ... \partial u_d} \left[\int_{[0,1]^p} \frac{\partial^p}{\partial v_1 ... \partial v_p} C_{\mathbf{U},\mathbf{V}}\, dv_1 ... dv_p \right] du_1 ... du_d \\
&= \int_{[0,1]^p} \frac{\partial^p}{\partial v_1 ... \partial v_p} C_{\mathbf{U},\mathbf{V}}\, dv_1 ... dv_p \\
&= \int_{[0,1]^p} C_{\mathbf{U}|\mathbf{V}}\, dv_1 ... dv_p \\
C_{\mathbf{U}}(\mathbf{u}) &= \int_{[0,1]^p} \prod_{j=1}^{d} C_{U_j|\mathbf{V}}\, dv_1 ... dv_p. \qquad (2.8)
\end{aligned}
$$

The joint CDF of $U_1, U_2, ..., U_d$, i.e., the copula of $U_1, U_2, ..., U_d$, as represented in (2.8) was denominated *factor copula model* by Krupskii and Joe (2013).

Factor copulas are conditional independence models, since $U_1, U_2, ..., U_d$ are conditionally independent given $V_1, V_2, ..., V_p$. Considering this, any conditional independence model based on $d + p$ variables can be converted into a factor copula after transformations to Uniform$[0,1]$ random variables (JOE, 2014).

The dependence structure of $\mathbf{U}$, consisting of d conditional CDFs of the form $C_{U_j|V_1,...,V_p} = C_{j|\mathbf{V}}$, can be expressed in terms of a sequence of bivariate copulas linking the observable variables U_j ($j = 1, ..., d$) to the latent variables V_k ($k = 1, ..., p$). This way, as we will show in the next section for $p = 1$ and $p = 2$ latent variables, the resulting factor copula model will be composed of pd bivariate copulas. Hence, for one-parameter copulas, the factor copula of $\mathbf{U}$ will have pd parameters.

2.5.1 One- and two-factor copulas

For the case of $p = 1$ latent variable, V_1, let C_{j,V_1} and c_{j,V_1} be, respectively, the joint CDF and density of (U_j, V_1), for $j = 1, 2, ..., d$, that is, all copulas are absolutely continuous. Then, Eq. (2.8) becomes

$$C_{\mathbf{U}}(\mathbf{u}) = \int_0^1 \prod_{j=1}^{d} C_{j|V_1}(u_j|v_1)\, dv_1. \tag{2.9}$$

Since $\frac{\partial}{\partial v_1} C_{j,V_1}(u_j, v_1) = C_{j|V_1}(u_j|v_1)$, under regularity conditions (differentiation under the integral sign) we have that

$$\begin{aligned} c(\mathbf{u}) &= \frac{\partial^d}{\partial u_1 ... \partial u_d} C(\mathbf{u}) = \frac{\partial^d}{\partial u_1 ... \partial u_d} \int_0^1 \prod_{j=1}^{d} \frac{\partial}{\partial v_1} C_{j,V_1}(u_j, v_1)\, dv_1 \\ &= \int_0^1 \prod_{j=1}^{d} \frac{\partial^2}{\partial u_j \partial v_1} C_{j,V_1}(u_j, v_1)\, dv_1 \\ &= \int_0^1 \prod_{j=1}^{d} c_{j,V_1}(u_j, v_1)\, dv_1. \end{aligned}$$

Therefore, the dependence structure of $\mathbf{U}$ is modelled by d bivariate copulas linking the observable variables U_j to the latent variable V_1. The model expressed in Eq. (2.9) is called one-factor copula (JOE, 2014).

As stated by Krupskii and Joe (2013), factor copulas can be nicely interpreted. An example in finance, for one-factor copula, refers to the prices of stocks (observable random variables) in a common sector that are affected by the state (latent variable) of this sector. The sector index (measurable) on the other hand, might not entirely explain the dependence in the stock prices, justifying the use of a factor copula model.

A special case of one-factor copula model is derived when C_{j,V_1} is a Gaussian copula for all $j = 1, 2, ..., d$. Then, (2.9) becomes the multivariate Gaussian distribution with a one-factor correlation matrix, as we show next.

Let C_{j,V_1} be a Gaussian copula with correlation θ_{j1},

$$C_{j,V_1}(u_j, v_1) = \Phi_2(\Phi^{-1}(u_j), \Phi^{-1}(v_1); \theta_{j1}). \tag{2.10}$$

It is known that, if Z and W are standard normal random variables with correlation θ, the conditional distribution of Z given $W = w$ is also normal, with mean $\mu = \theta w$ and variance

$\sigma^2 = (1-\theta^2)$. This way, by setting $u_j = \Phi(z_j)$ and $v_1 = \Phi(w_1)$ in (2.10), we obtain

$$F_{Z_j,W_1}(z_j,w_1) := C_{j,V_1}(\Phi(z_j),\Phi(w_1)) = \Phi_2(z_j,w_1;\theta_{j1}).$$

Hence, $F_{Z_j|W_1}(z_j|w_1) = C_{j|V_1}(\Phi(z_j)|\Phi(w_1))$ is normal with mean $\theta_{j1}w_1$ and variance $(1-\theta_{j1}^2)$ and

$$C_{j|V_1}(\Phi(z_j)|\Phi(w_1)) = \Phi\left(\frac{z_j - \theta_{j1}w_1}{\sqrt{1-\theta_{j1}^2}}\right). \tag{2.11}$$

Combining (2.11) with (2.9) and remembering that $v_1 = \Phi(w_1)$ so $dv_1 = \phi(w_1)dw_1$, results in

$$F(z_1,z_2,...,z_d) := C(\Phi(z_1),\Phi(z_2),...,\Phi(z_d)) = \int_{-\infty}^{+\infty}\left[\prod_{j=1}^{d}\Phi\left(\frac{z_j-\theta_{j1}w_1}{\sqrt{1-\theta_{j1}^2}}\right)\right]\phi(w_1)\,dw_1. \tag{2.12}$$

Model (2.12) can also be derived from the classical factor model representation

$$Z_j - \mu = \lambda_{j1}W + \epsilon_j, \quad j = 1,...,d,$$

where W is a factor, or latent variable, $\epsilon_1,...,\epsilon_d$ are error terms and $W,\epsilon_1,...,\epsilon_d$ are i.i.d. standard normal random variables. Let $\mu = 0$ and express $\mathbf{Z}$ as

$$Z_j = \theta_{j1}W + \sqrt{1-\theta_{j1}}\epsilon_j, \quad j = 1,...,d, \tag{2.13}$$

so that $E(Z_j) = 0$, $Var(Z_j) = 1$ and $Cov(Z_j,Z_i) = \theta_{j1}\theta_{i1}$, $i \neq j$.

Proposition 2.1. *The joint pdf of $Z_1,...,Z_d$ may be written in the form (2.12).*

Proof. Let $Z_1,Z_2,...,Z_d$ be as in (2.13) and $Z_0 = W$. Since the variables $W,\epsilon_1,\epsilon_2,...,\epsilon_d$ are i.i.d. standard normal RVs, their joint density is $f_{W,\epsilon_1,...,\epsilon_d} = \phi(w)\phi(\epsilon_1)...\phi(\epsilon_d)$. Given the set of transformation functions $g_0,g_1,g_2,...,g_d$

$$\begin{aligned}
Z_0 &= g_0(W,\epsilon_1,...,\epsilon_d) = W\\
Z_1 &= g_1(W,\epsilon_1,...,\epsilon_d) = \theta_{11}W + \sqrt{1-\theta_{11}^2}\epsilon_1\\
Z_2 &= g_2(W,\epsilon_1,...,\epsilon_d) = \theta_{21}W + \sqrt{1-\theta_{21}^2}\epsilon_2\\
&\vdots\\
Z_d &= g_d(W,\epsilon_1,...,\epsilon_d) = \theta_{d1}W + \sqrt{1-\theta_{d1}^2}\epsilon_d,
\end{aligned}$$

take the set of inverse transformation functions $g_0^{-1}, g_1^{-1}, ..., g_d^{-1}$

$$
\begin{aligned}
W &= g_0^{-1}(Z_0, Z_1, ..., Z_d) = Z_0 \\
\epsilon_1 &= g_1^{-1}(Z_0, Z_1, ..., Z_d) = \frac{Z_1 - \theta_{11} Z_0}{\sqrt{1-\theta_{11}^2}} \\
\epsilon_2 &= g_2^{-1}(Z_0, Z_1, ..., Z_d) = \frac{Z_2 - \theta_{11} Z_0}{\sqrt{1-\theta_{21}^2}} \\
&\vdots \\
\epsilon_d &= g_d^{-1}(Z_0, Z_1, ..., Z_d) = \frac{Z_d - \theta_{11} Z_0}{\sqrt{1-\theta_{d1}^2}}.
\end{aligned}
$$

The Jacobian of the transformation can be computed by taking the determinant of the matrix of partial derivatives

$$
D_Z = \begin{pmatrix}
\frac{\partial W}{\partial Z_0} & \frac{\partial W}{\partial Z_1} & \cdots & \frac{\partial W}{\partial Z_d} \\
\frac{\partial \epsilon_1}{\partial Z_0} & \frac{\partial \epsilon_1}{\partial Z_1} & \cdots & \frac{\partial \epsilon_1}{\partial Z_d} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial \epsilon_d}{\partial Z_0} & \frac{\partial \epsilon_d}{\partial Z_1} & \cdots & \frac{\partial \epsilon_d}{\partial Z_d}
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
-\theta_{11}/\sqrt{1-\theta_{11}^2} & 1/\sqrt{1-\theta_{11}^2} & 0 & \cdots & 0 \\
-\theta_{21}/\sqrt{1-\theta_{11}^2} & 0 & 1/\sqrt{1-\theta_{21}^2} & & \vdots \\
\vdots & \vdots & & \ddots & \\
-\theta_{d1}/\sqrt{1-\theta_{d1}^2} & 0 & 0 & \cdots & 1/\sqrt{1-\theta_{d1}^2}
\end{pmatrix}
$$

$$
J = |D_Z| = \prod_{j=1}^{d} \frac{1}{\sqrt{1-\theta_{j1}^2}}.
$$

Hence the joint density of $Z_0, Z_1, ..., Z_d$ is

$$
f_{Z_0, Z_1, ..., Z_d} = \phi(z_0) \left[\prod_{j=1}^{d} \phi \left(\frac{z_1 - \theta_{j1} z_0}{\sqrt{1-\theta_{j1}^2}} \right) \right] \left[\prod_{i=1}^{d} \frac{1}{\sqrt{1-\theta_{i1}^2}} \right].
$$

Integrating with respect to z_0

$$
f_{Z_1, ..., Z_d} = \int_{-\infty}^{+\infty} \phi(z_0) \left[\prod_{j=1}^{d} \phi \left(\frac{z_1 - \theta_{j1} z_0}{\sqrt{1-\theta_{j1}^2}} \right) \right] \left[\prod_{i=1}^{d} \frac{1}{\sqrt{1-\theta_{i1}^2}} \right] dz_0
$$

$$
F_{Z_1, ..., Z_d} = \int_{-\infty}^{+\infty} \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_d} \left[\prod_{j=1}^{d} \phi \left(\frac{s_j - \theta_{j1} z_0}{\sqrt{1-\theta_{j1}^2}} \right) \right] \left[\prod_{i=1}^{d} \frac{1}{\sqrt{1-\theta_{i1}^2}} \right] \phi(z_0) \, ds_1 ... ds_d \, dz_0.
$$

Therefore, by making $x_j = (s_j - \theta_{j1} z_0)/(\sqrt{1-\theta_{j1}^2})$, $ds_j = \sqrt{1-\theta_{j1}^2}\, dx_j$ and

$$F_{Z_1,\ldots,Z_d} = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{\frac{z_1-\theta_{11}z_0}{\sqrt{1-\theta_{11}^2}}} \cdots \int_{-\infty}^{\frac{z_d-\theta_{d1}z_0}{\sqrt{1-\theta_{d1}^2}}} \left[\prod_{j=1}^{d} \phi\left(x_j\right) \right] dx_1 \ldots dx_d \right\} \phi(z_0) dz_0.$$

Finally, by setting $z_0 = w_1$ we get

$$F_{Z_1,\ldots,Z_d} = \int_{-\infty}^{+\infty} \left[\prod_{j=1}^{d} \Phi\left(\frac{z_j - \theta_{j1} w_1}{\sqrt{1-\theta_{j1}^2}} \right) \right] \phi(w_1) dw_1. \qquad \square$$

If another factor is introduced in the model, i.e., for the case of $p = 2$ factors, the resulting two-factor copula model is of the form

$$C(\mathbf{u}) = \int_0^1 \int_0^1 \prod_{j=1}^{d} C_{j|V_1,V_2}(u_j|v_1,v_2)\, dv_1 dv_2. \tag{2.14}$$

The conditional distribution $C_{j|V_1,V_2}(u_j|v_1,v_2)$ can be written as

$$\begin{aligned} C_{j|V_1,V_2}(u_j|v_1,v_2) &= P[U_j \leqslant u_j | V_1 = v_1, V_2 = v_2] = \frac{\partial}{\partial v_2} P[U_j \leqslant u_j, V_2 \leqslant v_2 | V_1 = v_1] \\ &= \frac{\partial}{\partial v_2} C_{j,V_2|V_1}(C_{j|V_1}(u_j|v_1), C_{V_2|V_1}(v_2|v_1)), \end{aligned}$$

where $C_{j,V_2|V_1}$ is the conditional copula for margins $C_{j|V_1}$ and $C_{V_2|V_1}$ and has density $c_{j,V_2|V_1}$. Since the factors V_1 and V_2 are independent Uniform$[0,1]$ random variables, $C_{V_2|V_1}(v_2|v_1) = C_{V_2}(v_2) = v_2$. Therefore

$$\begin{aligned} C_{j|V_1,V_2}(u_j|v_1,v_2) &= \frac{\partial}{\partial v_2} C_{j,V_2|V_1}(C_{j|V_1}(u_j|v_1), C_{V_2|V_1}(v_2|v_1)) = \frac{\partial}{\partial v_2} C_{j,V_2|V_1}(C_{j|V_1}(u_j|v_1), v_2) \\ &= C_{j|V_1\big|V_2}(C_{j|V_1}(u_j|v_1)|v_2). \end{aligned} \tag{2.15}$$

Following the simplifying assumption on vine copulas (HAFF; AAS; FRIGESSI, 2010) it is assumed here that the copula $C_{j,V_2|V_1}$ does not depend on v_1, that is

$$C_{j,V_2|V_1}(C_{j|V_1}(u_j|v_1), v_2; v_1) = C_{j,V_2|V_1}(C_{j|V_1}(u_j|v_1), v_2),$$

in other words, $C_{j,V_2|V_1}(C_{j|V_1}(u_j|v_1), v_2)$ depends on v_1 only through its marginal $C_{j|V_1}(u_j|v_1)$. According to Joe (2014), this assumption is not strong in the case of the factor copula model,

due to the nature of the factors (latent variables). Combining (2.15) with (2.14) we have

$$C(\mathbf{u}) = \int_0^1 \int_0^1 \prod_{j=1}^{d} C_{j|V_1}\big|_{V_2}(C_{j|V_1}(u_j|v_1)|v_2)\, dv_1 dv_2 \tag{2.16}$$

and by differentiating with respect to $u_1, ..., u_d$,

$$c(\mathbf{u}) = \int_0^1 \int_0^1 \prod_{j=1}^{d} c_{j,V_2|V_1}(C_{j|V_1}(u_j|v_1), v_2) c_{j,V_1}(u_j, v_1)\, dv_1 dv_2,$$

the density of the two-factor copula. This way, the dependence structure of $\mathbf{U}$ is defined by $2d$ bivariate copulas: d linking V_1 to each U_j and d linking V_2 to U_j, conditioned on V_1, i.e., $C_{U_1,V_1}, C_{U_2,V_1}, ..., C_{U_d,V_1}$ and $C_{U_1,V_2|V_1}, C_{U_2,V_2|V_1}, ..., C_{U_d,V_2|V_1}$, respectively.

The multivariate Gaussian distribution with a two-factor correlation matrix can be derived as a special case of a two-factor copula model, when the $2d$ bivariate copulas C_{U_j,V_1}, $C_{U_j,V_2|V_1}$ $(j = 1, ..., d)$ are all Gaussian with parameters $\theta_{j1} = \alpha_{j1}$ and $\theta_{j2} = \alpha_{j2} \Big/ \sqrt{1-\alpha_{j1}^2}$, respectively. Also, $U_1, ..., U_d$ need to be transformed to standard normal random variables and V_1 and V_2 to independent standard normal RVs, say $Z_1, ..., Z_d$ and W_1 and W_2, respectively. In this case, α_{j1} is the correlation between Z_j and W_1, α_{j2} is the correlation between Z_j and W_2 and θ_{j2} the partial correlation between Z_j and W_2 given $W_1 = w_1$ (JOE, 2014).

If C_{j,V_1} is a Gaussian copula, then $C_{j,V_1}(u_j, v_1) = \Phi_2(\Phi^{-1}(u_j), \Phi^{-1}(v_1); \theta_{j1})$. Setting $u_j = \Phi(z_j)$, $w_1 = \Phi(v_1)$ and $w_2 = \Phi(v_2)$, by Sklar's theorem,

$$\begin{aligned} C_{j,V_1}(\Phi(z_j), \Phi(w_1)) &= C_{j,W_1}(\Phi(z_j), \Phi(w_1)) \\ &= \Phi_2(z_j, w_1; \theta_{j1}). \end{aligned}$$

We now need to obtain $C_{j|V_1}$, or $C_{j|W_1}$, that is, the conditional distribution of Z_j given W_1, for all $j = 1, ..., d$. Since Z_j and W_1 are both standard normal RVs with correlation θ_{j1}, it is known that the conditional CDF of Z_j given W_1 is also normal with mean $\theta_{j1} w_1$ and variance $(1 - \theta_{j1}^2)$. Therefore,

$$C_{j|W_1}(\Phi(z_j)|w_1) = \Phi\left(\frac{z_j - \theta_{j1} w_1}{\sqrt{1-\theta_{j1}^2}}\right).$$

Also, if $C_{j,V_2|V_1}$ is a Gaussian copula,

$$C_{j,V_2|V_1}(u_j, v_2) = \Phi_2(\Phi^{-1}(u_j), \Phi^{-1}(v_2); \theta_{j2}).$$

We can now retrieve $C_{j|W_1}\big|_{W_2}(C_{j|W_1}(\Phi(z_j)|w_1)|w_2)$. In view of $Z_j|W_1 \sim C_{j|W_1}(\Phi(z_j)|w_1)$ being normally distributed with mean $\theta_{j1}w_1$ and variance $(1-\theta_{j1}^2)$, or equivalently,

$$\frac{z_j - \theta_{j1}w_1}{\sqrt{1-\theta_{j1}^2}} \sim \text{Normal}(0,1),$$

it follows that the conditional CDF of $Z_j|W_1$ given $W_2 \sim \text{Normal}(0,1)$, $C_{j|W_1}\big|_{W_2}(C_{j|W_1}(\Phi(z_j)|w_1)|w_2)$, is also normal with mean $\theta_{j1}w_1 + \theta_{j2}w_2\sqrt{1-\theta_{j1}^2}$ and variance $(1-\theta_{j1}^2)(1-\theta_{j2}^2)$, that is,

$$C_{j|W_1}\big|_{W_2}(C_{j|W_1}(\Phi(z_j)|w_1)|w_2) = \Phi\left(\frac{z_j - \theta_{j1}w_1 - \sqrt{1-\theta_{j1}^2}\,\theta_{j2}w_2}{\sqrt{(1-\theta_{j1}^2)(1-\theta_{j2}^2)}}\right). \tag{2.17}$$

Therefore, by plugging (2.17) in (2.16), remembering that $dv_1dv_2 = \phi(w_1)\phi(w_2)dw_1dw_2$, we get

$$\begin{aligned} &C(\Phi(z_1),...,\Phi(z_d)) := F_{Z_1,...,Z_d}(z_1,...,z_d) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\prod_{j=1}^{d}\Phi\left(\frac{z_j - \theta_{j1}w_1 - \sqrt{1-\theta_{j1}^2}\,\theta_{j2}w_2}{\sqrt{(1-\theta_{j1}^2)(1-\theta_{j2}^2)}}\right)\phi(w_1)\phi(w_2)dw_1dw_2, \end{aligned}$$

the multivariate Gaussian model with a two-factor correlation structure, that, as we have shown for $p=1$, can also be derived from the classical factor model representation

$$Z_j = \alpha_{j1}W_1 + \alpha_{j2}W_2 + \sqrt{\left(1-\alpha_{j1}^2\right)\left(1-\frac{\alpha_{j2}}{1-\alpha_{j1}^2}\right)}\,\epsilon_j, \quad j = 1,...,d,$$

where W_1, W_2 are factors, or latent variables, $\epsilon_1,...,\epsilon_d$ are error terms and $W_1, W_2, \epsilon_1,...,\epsilon_d$ are i.i.d. standard normal. Also, $E(Z_j) = 0$, $Var(Z_j) = 1$ and $Cov(Z_j, Z_i) = \alpha_{j1}\alpha_{i1} + \alpha_{j2}\alpha_{i2}, \quad i \neq j$. In fact, the multivariate Gaussian model with a p-factor correlation structure given by $\Sigma = \mathbf{A}\mathbf{A}^T + \mathbf{\Psi}^2$, where $\mathbf{A}_{d\times p}$ is a matrix of loadings and $\mathbf{\Psi}^2$ is a diagonal matrix of residual variances, can be seen as a special case of factor copula model (JOE, 2014).

The multivariate Gaussian model with a p-factor correlation structure is just one example of many other possible models that can be achieved with factor copulas. Consider the case of d-dimensional one-factor copulas, for instance. If each of the d bivariate copulas in the model is chosen from a list with r different families (any bivariate copula can be used!), then

the number of possible constructions is r^d, each with only d parameters to be estimated (assuming one-parameter families). That is why factor copulas are adequate models for dependence modelling when the number of variables is large.

3 SURVIVAL ANALYSIS

It is often the case in statistical modelling that data are specified as the time until an event happens. This event could be death, recovery from some disease, failure of an electronic component, etc. There are some particular features that make the analysis of such data non-trivial, the most important and common being right censoring. This kind of censoring arises when the event of interest does not occur during the observation time, i.e., we only know that the individual survived up to a certain point in time, but we have no information about the occurrence of the event of interest. There are several reasons that lead to right censoring, and, in most cases, these are not under the control of the researcher, e.g., the study ends before the event occurs, the subject leaves the study or any kind of lost to follow-up happens. Standard statistical techniques are not suitable to analyse such types of data, also known as time-to-event data or survival data, hence a collection of techniques was developed to this end, giving rise to a branch of statistics called survival analysis.

In this chapter, we present some basic concepts and terminology of survival analysis, along with some fundamental estimation techniques for right-censored time-to-event data. The main reference used in this chapter is Klein and Moeschberger (2006). Another fine works on survival analysis are David and Kleinbaum (2016) and Hougaard (2012).

3.1 Basic concepts and terminology

Our interest hereon is to give a statistical modelling perspective on time-to-event phenomena so, firstly, let T be a nonnegative random variable describing the time until an event happens. As discussed earlier, this event could be anything from death, to recovery from a particular disease, or it may be related to the failure of a mechanical or electronic system/component. It could be the case that right censoring is present in the data. In this case, let C denote the censoring time variable (independent of T) and $\delta = I(T \leq C)$ a dichotomous variable assuming 0 or 1, indicating if the individual is right-censored or not, respectively. In practice, what we observe, under a right censoring scheme, is a realisation of the random variable $X = \min(T, C)$.

The basic quantity used to describe the probability that an individual survives up to a certain period of time T is the survival function. Let $F(t) = P[T \leq t]$ be the cumulative probability function of T with density $f(t)$. Then its survival function is given by

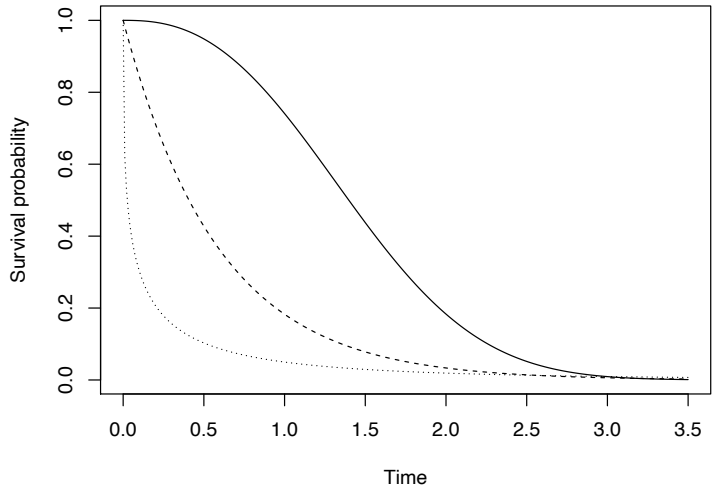
$$S(t) = P[T > t] = 1 - F(t) = \int_t^{+\infty} f(u)du,$$

such that

$$f(t) = -\frac{dS}{dt}.$$

One common parametric survival model is the Weibull, with survival function $S(t) = \exp(-\lambda t^{\rho})$, where $\rho \in (0, +\infty)$ is the shape parameter and $\lambda \in (0, +\infty)$ the scale parameter. The well known exponential model is a particular case of the Weibull when the shape parameter is equal to 1, i.e., the Weibull is a very reasonable generalisation of the exponential model that can assume a variety of forms due to its shape parameter $\rho \in (0, +\infty)$. We exemplify this characteristic of the Weibull model by giving the graph of its survival functions under different parametric settings in Figure 3.1.

Figure 3.1 – Survival curves of Weibull model with $\lambda = 0.3$ and $\rho = 2.5$ (solid); $\lambda = 1.7$ and $\rho = 1$ (dashed); $\lambda = 3$ and $\rho = 0.4$ (dotted).



As opposed to the cumulative probability function, the survival function is monotone decreasing, meaning that for any two $t_1 \leq t_2$, the survival probabilities are non-increasing, $S(t_1) \leq S(t_2)$. Although being an important quantity to compare the survival probabilities of individuals, there are other insightful ways of doing so rather than using only the survival function. One classical example of this is the hazard function, $h(t)$, defined as the risk of experiencing the event immediately after time t given survival up to t. The hazard function is mathematically defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t + \Delta t | T \geq t]}{\Delta t}.$$

By further working on the above expression, one gets

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t + \Delta t]}{P[T \geq t]\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t + \Delta t]}{S(t)\Delta t} = \frac{f(t)}{S(t)}, \end{aligned}$$

such that the hazard function is written in terms of the survival function and its density. The cumulative version of the hazard function is simply defined as

$$H(t) = \int_0^t h(u)du,$$

analogously to the cumulative distribution function and its density.

It is also possible to derive some useful identities involving the hazard function (cumulative hazard function) and the survival function. As a matter of fact, one is fully determined by the other, i.e., the hazard function can be obtained from the survival function alone, and vice versa, e.g.,

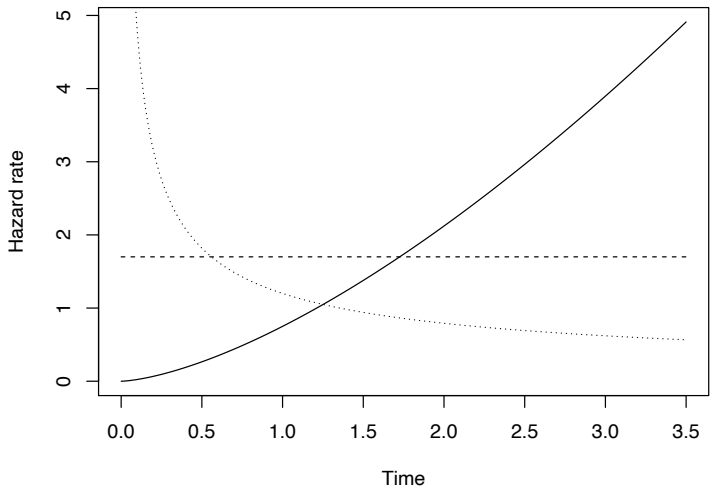
$$\begin{aligned} S(t) &= \exp\left\{-\int_0^t h(u)du\right\} = \exp\{-H(t)\}, \\ h(t) &= -\frac{dS(t)/dt}{S(t)} = -\frac{d\log S(t)}{dt}, \\ H(t) &= -\log S(t). \end{aligned}$$

The hazard function carries useful information about how an individual experiences the event of interest over time. It can assume a varied range of shapes, with the only restriction of being nonnegative. Take for example the Weibull model, its hazard function is given by

$$h(t) = \rho\lambda t^{\rho-1}.$$

For $\rho = 1$ (exponential model), the Weibull hazard function is constant over time, $h(t) = \lambda$, for $\rho > 1$ it is an increasing function of t and for $0 < \rho < 1$, it is a decreasing function. This feature is better observed in Figure 3.2, under the same settings used in Figure 3.1.

Figure 3.2 – Hazard rates of Weibull model with $\lambda = 0.3$ and $\rho = 2.5$ (solid); $\lambda = 1.7$ and $\rho = 1$ (dashed); $\lambda = 3$ and $\rho = 0.4$ (dotted).



3.2 Maximum likelihood estimation under right censoring

When right censoring is present in the data, the likelihood function assumes a different form from the classical approach, this is because we do not observe the actual realisations of the random variable T, but rather from $X = \min(T,C)$ and $\delta = I(T \leq C)$. This way, each individual contributes to the likelihood in a different way. If it is not censored, the procedure is as usual, i.e., its contribution is the density evaluated at the realised value of T. For a right-censored observation, the contribution is the survival function evaluated at the observed value of X. The reason for this is detailed below.

Let $T_1, ..., T_n$ be a random sample from a population with survival function $S_T(t;\theta) = 1 - F_T(t;\theta)$ and density $f_T(t;\theta)$. Due to a potential right censoring, we may not observe the true realisations of $T_1, ..., T_n$, instead, we observe the realisations of the random pairs $(X_1, \delta_1), ..., (X_n, \delta_n)$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$, for $i = 1, ..., n$, with $C_1, ..., C_n$ being independent and identically distributed censoring times coming from an unknown cumulative distribution function $F_C(c;\phi)$. We are now interested in deriving the joint density of (X, δ) in order to determine the contributions to the likelihood. For this, we will split the process in two situations, the first being the case when the true survival time is known, that is, when $\delta = 1$, and the other when the individual is right-censored, i.e., $\delta = 0$.

For the case of $\delta = 1$, the joint density of (X, δ), namely $g(x, \delta)$, can be regarded as $g(x, \delta = 1)$. Hence, by assuming C independent of T, the joint density can be rewritten as (DUCHATEAU; JANSSEN, 2008)

$$\begin{aligned} g(x, \delta = 1) &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} P(x - \epsilon \leq X \leq x + \epsilon, \delta = 1) \\ &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} P(x - \epsilon \leq T \leq x + \epsilon, T \leq C) \\ &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} \int_{t}^{+\infty} dF_C(c)\, dF_T(t) \\ &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} S_C(t)\, dF_T(t) \\ &= S_C(x; \phi) f_T(x; \theta), \end{aligned} \tag{3.1}$$

where $S_C(x; \phi) = 1 - F_C(x; \phi)$.

When $\delta = 0$, i.e., the individual is right-censored, the joint density of (X, δ), assuming C independent of T, is given by

$$\begin{aligned} g(x, \delta = 0) &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} P(x - \epsilon \leq X \leq x + \epsilon, \delta = 0) \\ &= \lim_{\epsilon \to 0} \frac{1}{2\epsilon} P(x - \epsilon \leq C \leq x + \epsilon, T > C) \\ &= S_T(x; \theta) f_C(x; \phi), \end{aligned} \tag{3.2}$$

where $f_C(x; \phi)$ is the probability density function of C.

The joint density of (X, δ) can now be obtained by joining parts (3.1) and (3.2) together

$$h(x, \delta; \theta, \phi) = [f_T(x; \theta) S_C(x; \phi)]^{\delta} [f_C(x; \phi) S_T(x; \theta)]^{1-\delta},$$

such that the likelihood for (θ, ϕ) takes the form

$$L(\theta, \phi | (x_1, \delta_1), ..., (x_n, \delta_n)) = \prod_{i=1}^{n} [f_T(x_i; \theta) S_C(x_i; \phi)]^{\delta_i} [f_C(x_i; \phi) S_T(x_i; \theta)]^{1-\delta_i}.$$

Since the parameters θ and ϕ are not related to each other (T independent of C), the quantities $S_C(x_i; \phi)$ and $f_C(x_i; \phi)$ $(i = 1, ..., n)$ can be regarded as constants and ruled out of the likelihood

function. Hence

$$L(\theta) = \prod_{i=1}^{n} [f_T(x_i;\theta)]^{\delta_i} [S_T(x_i;\theta)]^{1-\delta_i} \tag{3.3}$$
$$= \prod_{i=1}^{n} [h_T(x_i;\theta)]^{\delta_i} S_T(x_i;\theta),$$

where h_T is the hazard function of T. We are now able to make inference only on θ, our main objective at first.

3.2.1 Non-parametric estimation of the survival function - The Kaplan-Meier estimator

Using a non-parametric estimator can be a viable option when one does not want to make any assumptions on the form of the survival function. One classical example of non-parametric estimator of the survival function that takes censoring into consideration is the Kaplan-Meier estimator (KAPLAN; MEIER, 1958), also called product limit estimator.

In order to define the Kaplan-Meier estimator, first let $t_1 < ... < t_k$ be all the k distinct observed survival times from a random sample $T_1, ..., T_n$, with $k \leq n$. At time t_i $(i = 1, ..., n)$ there are Y_i individuals who are said to be at risk, that is, they have not experienced the event of interest prior to t_i, and d_i will experience the event at t_i. Also, let $t_0 = 0$ and $d_0 = 0$. Hence, the Kaplan-Meier estimator is defined as (KOROSTELEVA, 2008):

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{Y_i}\right).$$

The estimated survival curve produced by the Kaplan-Meier estimator is a step function, with jumps on the observed survival times (right-continuous). One advantage of this estimator is precisely its flexibility on not having to assume any particular parametric model as correct, on top of allowing for censored data. On the other hand, a major drawback is that the hazard function cannot be estimated and the estimated survival curve is not well defined beyond t_k if there are censored observations beyond that time (HOUGAARD, 2012).

3.3 Regression models for survival data

In many applications, it is common to use explanatory variables when dealing with time-to-event data. This has not been the case so far, since we dealt with data coming from homogeneous populations, that is, we were only able to assess time-to-event data with no explanatory

variables. We will present in this section two common regression methods to make inference on time-to-event data coming from a random variable T that depends on a vector of covariates, or explanatory variables, $\mathbf{Z} = (Z_1, ..., Z_p)$. This set of covariates can be quantitative (blood pressure, body fat percentage, doses of a drug, temperature, age, height, weight, etc.), qualitative (gender, treatment, marital status, etc.), or even time-varying $\mathbf{Z}(\mathbf{t}) = (Z_1(t), ..., Z_p(t))$. These time dependent covariates can arise, for example, in a situation where a particular quantity is measured several times over time in the experiment (KLEIN; MOESCHBERGER, 2006).

3.3.1 The accelerated failure-time model

One of the most common parametric models for regression analysis with survival data is the accelerated failure time-model (AFT). In this model, the natural logarithm of the survival time is modelled as a function of the covariates, such that we have a transformation of T to the real line:

$$\log T = \beta_0 + \beta_1 Z_1 + ... + \beta_p Z_p + \sigma W, \tag{3.4}$$

where $(\beta_0, \beta_1, ..., \beta_p)$ are the parameters for the vector of covariates $\mathbf{Z} = (Z_1, ..., Z_p)$, with β_0 being the intercept; σ is a real constant and W is the random error.

The two most classical distributions for the survival time T are the Weibull and exponential, the last being a particular case of the first. These distributions of T arise when the error term W follows the extreme-value distribution with density

$$f_W(w) = \exp\{w - e^w\}, \quad w \in \mathbb{R}. \tag{3.5}$$

If it is the case that $\sigma = 1$ and W follows an extreme-value distribution with the above density, then T will follow an exponential distribution. For $\sigma \neq 1$, T will follow a Weibull distribution (KOROSTELEVA, 2008).

The reason why this model is called the accelerated failure-time model (AFT) can be understood by doing the follow: let $S_0(t)$ denote the survival function of T when all the covariates are set to zero, that is, $S_0(t)$ is the baseline survival function of $\exp(\beta_0 + \sigma W)$. Then, also

note that

$$\begin{aligned}
P[T > t|\mathbf{Z}] &= P[\log T > \log t|\mathbf{Z}] \\
&= P[\beta_0 + \beta_1 Z_1 + ... + \beta_p Z_p + \sigma W > \log t|\mathbf{Z}] \\
&= P[\beta_0 + \sigma W > \log t - (\beta_1 Z_1 + ... + \beta_p Z_p)|\mathbf{Z}] \\
&= P[\exp\{\beta_0 + \sigma W\} > \exp\left\{\log t - (\beta_1 Z_1 + ... + \beta_p Z_p)\right\}|\mathbf{Z}] \\
&= P[\exp\{\beta_0 + \sigma W\} > t\exp\left\{-(\beta_1 Z_1 + ... + \beta_p Z_p)\right\}|\mathbf{Z}] \\
&= S_0(t\exp\left\{-(\beta_1 Z_1 + ... + \beta_p Z_p)\right\}).
\end{aligned}$$

In other words, depending on the sign of $(\beta_1 Z_1 + ... + \beta_p Z_p)$ the effect of the covariates on the baseline survival time (original time scale) will be to either shift it up (accelerate) or down (KLEIN; MOESCHBERGER, 2006).

As previously discussed, when the error term W follows an extreme-value distribution with density defined as in (3.5), the distribution of the survival time T will be Weibull. The survival function $S_T(t)$ in this case will have a modified scale (parameter λ) structure. Its form is given by

$$S_T(t) = \exp\left\{-t^{\rho}\exp[-(\beta_0 + \beta_1 Z_1 + ... + \beta_p Z_p)/\sigma]\right\},$$

where $\sigma = 1/\rho$ comes from the definition of the accelerated failure time model in (3.4). Estimation of the parameters in the accelerated failure time model can be carried out via maximum likelihood estimation, and, for example, by making use of the likelihood formula given in (3.3), right-censored observations can also be admitted in the estimation process.

3.3.2 Cox proportional hazards model

Another widely used regression model for time-to-event data is the semiparametric Cox proportional hazards model, proposed by Cox (1972). This procedure is focused on modelling the hazard function adjusted for a set of covariates, such that the effects of those explanatory variables can be compared in terms of their hazard.

Let $\mathbf{Z} = (Z_1, ..., Z_p)$ be a vector of covariates (possibly time dependent) which may affect the survival time T. The Cox proportional hazards model is defined as

$$h(t|\mathbf{Z}) = h_0(t)\exp(\beta_1 Z_1 + ... + \beta_p Z_p), \tag{3.6}$$

where $h_0(t)$ is the baseline hazard function, i.e., the hazard function of an hypothetical individual whose covariate values are all equal to zero, and $(\beta_1, ..., \beta_p)$ are the effects (parameters) of the covariate vector.

The reason why the Cox proportional hazards is a semiparametric model is that the baseline hazard function $h_0(t)$ is estimated non-parametrically, whereas the covariate effects are treated parametrically in the form $\exp(\beta_1 Z_1 + ... + \beta_p Z_p)$.

When comparing the ratio between the hazard rates of two individuals with covariate values $\mathbf{Z}^*$ and $\mathbf{Z}$, we get

$$\begin{aligned} \frac{h(t|\mathbf{Z}^*)}{h(t|\mathbf{Z})} &= \frac{h_0(t)\exp(\beta^T\mathbf{Z}^*)}{h_0(t)\exp(\beta^T\mathbf{Z})} \\ &= \frac{\exp(\beta^T\mathbf{Z}^*)}{\exp(\beta^T\mathbf{Z})} = \exp(\beta^T(\mathbf{Z}^* - \mathbf{Z})), \end{aligned} \tag{3.7}$$

that is, since the ratio of their hazards is constant over time, they are proportional. Hence the proportional hazards in the name of the model. The hazard ratio as in (3.7) quantifies the relative risk of an individual with covariate value $\mathbf{Z}^*$ experience the event of interest as compared to another individual with covariate value $\mathbf{Z}$. This is the main interest when using the Cox proportional hazards model (KLEIN; MOESCHBERGER, 2006). For example, suppose that we are interested in assessing the effect of a particular drug on the time to recovery of a disease. An experiment is conducted with two groups of subjects afflicted by the hypothetical disease, such that one group receives the drug ($Z = 1$) and the other, a placebo ($Z = 0$). A Cox proportional hazards model is adjusted in order to quantify the relative risk of recovery (drug vs placebo) and the estimated (assuming it is significant) β was 1.2. This means that the relative risk is equal to $\exp(1.2) \approx 3.32$, indicating that the drug is effective on treating the hypothetical disease.

Due to its semiparametric nature, usual maximum likelihood estimation cannot be used in Cox's model. Instead, a different approach based on the partial likelihood is taken. Let $(X_j, \delta_j, \mathbf{Z}_j)$, $j = 1, ..., n$, be the observed quantities in a sample of size n, with censoring times independent of the survival times. Also, assuming that there are no ties in the data, let $t_1 < ... < t_k$ denote the ordered survival times ($k \leq n$) and $\mathbf{Z} = (Z_1, ..., Z_p)$ be the vector of covariates that may explain the survival time, with associated parametric vector $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$. Let $R(t_i)$ denote the set of individuals who are at risk at time t_i, that is, the collection of individuals who did not experience the event just prior to t_i. The partial likelihood of the parametric vector $\boldsymbol{\beta}$ is

defined as (KLEIN; MOESCHBERGER, 2006)

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{Z}_j)},$$

where the parametric vector is estimated by finding the maxima of $L(\boldsymbol{\beta})$.

In the presence of ties in the data, estimation in the Cox model can be carried out by using the methods proposed by Breslow (1974) or Efron (1977).

3.4 Multivariate survival analysis

There are several circumstances in which the hypothesis of independence between the survival times of individuals may be misleading. Take for example time-to-event data observed from twins, couples, livestock herds, groups of patients from different hospitals, or even from a single subject that suffers from a chronic condition that has subsequent periods of remission and relapse. These are typical situation where the survival times may be dependent. Multivariate survival analysis arises from settings like as the ones aforementioned, where an additional information has to be taken into account: the dependence between the survival times (KLEIN; MOESCHBERGER, 2006).

Multivariate survival data come in a variety of settings, which are generated by different dependence mechanisms. Here the focus will be on clustered time-to-event data. A more detailed study on multivariate survival data can be found in Hougaard (2012).

Clustered survival data arise, for example, when individuals in a group share a common characteristic, and/or are affected by the same environment. In these scenarios, there may be a common risk factor, usually unknown or unmeasurable, that is shared among the subjects in a same group/cluster, hence making their hazard rates similar in some sense. Two popular models that deals with clustered time-to-event data are the frailty and the copula models. We give an overview on frailty models next. Copula models are discussed in the first chapter, from a technical perspective and a new copula model for clustered survival data is presented in the second part of this thesis.

3.4.1 Frailty model

A frailty model is a conditional hazard model in which the information about the heterogeneity between clusters is incorporated in its formulation as a multiplicative term coming from

a given distribution with unit mean and unknown variance. This multiplicative term is called the frailty factor and receives such name because it can affect the hazard rate of individuals within a cluster in two different ways. If the frailty factor is greater than 1 in a given cluster, it will increase the hazard, i.e., individuals belonging to that cluster will have a greater risk of experiencing the event if compared to those with a frailty factor equal to 1. For this reason, these individuals are considered more frail. On the other hand, a frailty factor between 0 and 1 decreases the hazard rate, and individuals will have a lesser risk of experiencing the event (DUCHATEAU; JANSSEN, 2008; KLEIN; MOESCHBERGER, 2006).

The most simple and common frailty model is the shared frailty model, which can be seen as an extension of the Cox proportional hazards model. It is defined as follows:

$$h_{ij}(t) = h_0(t)u_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}), \tag{3.8}$$

where $h_{ij}(t)$ is the conditional hazard rate of individual j ($j = 1,...,n_i$) within cluster i ($i = 1,...,K$) given the frailty factor u_i. $h_0(t)$ is an arbitrary baseline hazard function and $\mathbf{Z}_{ij} = (Z_{ij1},...,Z_{ijp})$ is the vector of covariates with associated parametric vector $\boldsymbol{\beta}$ of dimension p. This frailty model is called shared frailty because the frailty factor is common to all individuals belonging to the same cluster.

One feature of frailty models is that the frailty factor U is treated as a random variable following a particular distribution (the frailty distribution) with mean $E[U] = 1$ and variance $\text{var}[\text{U}] = \theta$. It is precisely the variance of the frailty distribution that is estimated from the data and carries the information of the heterogeneity between the clusters. There are several models that can be assumed for the frailty distribution, the most populars being the gamma and the lognormal (DUCHATEAU; JANSSEN, 2008).

Two possible ways of modelling the baseline hazard $h_0(t)$ in (3.8) are either nonparametrically or parametrically, with the latter being much easier to be handled in the estimation process. A formulation of the shared frailty model with a one-parameter gamma frailty factor and parametric baseline hazard, for example, simplifies the estimation by making it possible to integrate out the frailties from the conditional likelihoods of each cluster

$$L_i(\boldsymbol{\xi}, \boldsymbol{\beta}|u_i) = \prod_{j=1}^{n_i} \left[h_0(x_{ij})u_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\right]^{\delta_{ij}} \exp\left[-H_0(x_{ij})u_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\right],$$

where δ_{ij} is the event indicator, $\boldsymbol{\xi}$ contains the parameters of the baseline hazard and x_{ij} is the observed event (or censoring) time from the j^{th} individual of the i^{th} cluster. Just like in the specification of the likelihood function in (3.3), we assume that the censoring times are independent of the survival times. This results in a simple marginal likelihood:

$$\begin{aligned} L_{\text{marg},i}(\boldsymbol{\zeta}) &= \int_0^{+\infty} \prod_{j=1}^{n_i} \left[h_0(x_{ij})u_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\right]^{\delta_{ij}} \exp\left[-H_0(x_{ij})u_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\right] f_U(u_i)du_i \\ &= \frac{\Gamma(d_i + \theta^{-1}) \prod_{j=1}^{n_i} \left[h_0(x_{ij})u_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\right]^{\delta_{ij}}}{\left[\theta^{-1} + \sum_{j=1}^{n_i} H_0(x_{ij})u_i \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\right]^{1/\theta + d_i} \theta^{1/\theta}\Gamma(1/\theta)}, \end{aligned} \tag{3.9}$$

where $\boldsymbol{\zeta} = (\theta, \boldsymbol{\xi}, \boldsymbol{\beta})$, $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ and

$$f_U(u) = \frac{u^{1/\theta - 1} \exp(-u/\theta)}{\theta^{1/\theta}\Gamma(1/\theta)}$$

is the probability density function of the one-parameter gamma distribution, with Γ being the gamma function. Note that $E[U] = 1$ and var[U] $= \theta$. The cluster contributions to the likelihood in (3.9) can then be combined into the general marginal likelihood by taking their products:

$$L_{\text{marg}}(\boldsymbol{\zeta}) = \prod_{i=1}^{K} L_{\text{marg},i}(\boldsymbol{\zeta}). \tag{3.10}$$

Finally, because of the parametric baseline hazard and the gamma frailty factor, the marginal likelihood in (3.10) can be maximised with respect to $\boldsymbol{\zeta}$.

Although coming with some drawbacks, which we discuss in the second part of this thesis, frailty models are able to handle a great variety of settings of multivariate time-to-event data, and the different techniques and extensions of such models are manifold. For a more thorough study on frailty models, the reader is referred to the works of Duchateau and Janssen (2008) and Wienke (2011).

4 CONCLUSIONS AND PERSPECTIVES

As we have shown in the first chapter, copula models are great tools for dependence modelling and they come in many shapes and forms. A reasonable number of copula constructions methods and copula-based models have been proposed to deal with high-dimensional data, offering unique possibilities to be used in several fields of application of statistics. In multivariate survival analysis, particularly in clustered survival data modelling, the challenge is to use copula models to their full potential, accommodating for the special features of such data, e.g., censoring, clusters with variable size (possibly large). The factor copula models stands out as a significant candidate for this task, although, from a computational perspective, the high complexity of such models and the possible large clusters impose a formidable problem.

REFERENCES

BEDFORD, T.; COOKE, R. M. Vines: A new graphical model for dependent random variables. **Annals of Statistics**, v. 30, n. 4, p. 1031–1068, 2002.

BRESLOW, N. Covariance analysis of censored survival data. **Biometrics**, JSTOR, p. 89–99, 1974.

COOKE, R. M.; JOE, H.; AAS, K. Vines arise. In: ____. **Dependence modeling: vine copula handbook**. [S.l.]: World Scientific, 2011. p. 37–71.

COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 34, n. 2, p. 187–202, 1972.

DAVID, G.; KLEINBAUM, K. **Survival analysis: a self-learning text**. [S.l.]: Springer-Verlag New York, 2016.

DUCHATEAU, L.; JANSSEN, P. **The Frailty Model**. [S.l.]: Springer, Berlin, 2008.

EFRON, B. The efficiency of cox's likelihood function for censored data. **Journal of the American statistical Association**, Taylor & Francis, v. 72, n. 359, p. 557–565, 1977.

FERMANIAN, J.-D. Recent developments in copula models. **Econometrics**, v. 5, n. 3, p. 1–3, 2017.

GENEST, C.; MACKAY, J. The joy of copulas: Bivariate distributions with uniform marginals. **The American Statistician**, v. 40, n. 4, p. 280–283, 1986.

GUDENDORF, G.; SEGERS, J. Extreme-value copulas. In: ____. **Copula theory and its applications**. Berlin: Springer, 2010. p. 127–145.

HAFF, I. H.; AAS, K.; FRIGESSI, A. On the simplified pair-copula construction — simply useful or too simplistic? **Journal of Multivariate Analysis**, v. 101, n. 5, p. 1296–1310, 2010.

HOUGAARD, P. **Analysis of multivariate survival data**. [S.l.]: Springer Science & Business Media, 2012.

JOE, H. **Dependence Modeling with Copulas**. [S.l.]: Chapman & Hall/CRC, 2014.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958.

KJERSTI, A. et al. Pair-copula constructions of multiple dependence. **Insurance: Mathematics and economics**, v. 44, n. 2, p. 182–198, 2009.

KLEIN, J. P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. [S.l.]: Springer Science & Business Media, 2006.

KOROSTELEVA, O. **Clinical statistics: introducing clinical trials, survival analysis, and longitudinal data analysis**. [S.l.]: Jones & Bartlett Publishers, 2008.

KRUPSKII, P.; JOE, H. Factor copula models for multivariate data. **Journal of Multivariate Analysis**, v. 120, p. 85–101, 2013.

MORALES-NAPOLES, O. Counting vines. In: ____. **Dependence modeling: vine copula handbook**. [S.l.]: World Scientific, 2011. p. 189–218.

NELSEN, R. B. **An introduction to copulas**. [S.l.]: Springer Science & Business Media, 2007.

OH, D. H.; PATTON, A. J. Modeling dependence in high dimensions with factor copulas. **Journal of Business & Economic Statistics**, v. 35, n. 1, p. 139–154, 2017.

PRENEN, L.; BRAEKERS, R.; DUCHATEAU, L. Extending the archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 79, n. 2, p. 483–505, 2017.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018.

SHIH, J. H.; LOUIS, T. A. Inferences on the association parameter in copula models for bivariate survival data. **Biometrics**, JSTOR, v. 51, p. 1384–1399, 1995.

SKLAR, A. Fonctions de répartition á n dimensions et leurs marges. **Publ. Inst. Statist. Univ. Paris**, v. 8, p. 229–231, 1959.

WIENKE, A. **Frailty models in survival analysis**. [S.l.]: Boca Raton: Chapman and Hall, 2011.

SECOND PART - PAPER

PAPER - Factor copula models for right-censored clustered survival data

According to the guidelines of the journal *Lifetime Data Analysis* (currently under revision after peer review).

Factor copula models for right censored clustered survival data

Eleanderson Campos · Roel Braekers · Devanil J. de Souza · Lucas M. Chaves

Abstract In this article we propose a general class of copula-based models for right censored event time data grouped in clusters. The new methodology allows for clusters to have variable sizes ranging from small to large and intracluster dependence to be flexibly modeled by any parametric family of bivariate copulas, thus encompassing a wide range of dependence structures. Incorporation of covariates (possibly time dependent) in the margins is also supported. Additionally, the joint behavior of subjects within a cluster is assumed to be governed by an unknown common factor, leading to conditional independence of survival times at cluster level. Three estimation procedures are investigated: both one- and two-stage parametric and a two-stage semiparametric method where marginal survival functions are estimated by using a Cox proportional hazards model. We prove that the estimators are consistent and asymptotically normally distributed, and assess their finite sample behavior with simulation studies. Furthermore, we illustrate the proposed methods on a data set containing the time to first insemination after calving in dairy cattle clustered in herds of different sizes.

Keywords Clustered survival data · Conditional independence · Factor copula models · Multivariate survival data · Varying cluster size

Eleanderson Campos
Department of Statistics, Federal University of Lavras, Lavras, Minas Gerais, Brazil.
Data Science Institute, Interuniversity Institute for Biostatistics and statistical Bioinformatics - I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium.
E-mail: eleandersoncampos@estudante.ufla.br

Roel Braekers
Data Science Institute, Interuniversity Institute for Biostatistics and statistical Bioinformatics - I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium.
I-BioStat, KU Leuven, Leuven, Belgium.

Devanil J. de Souza
Department of Statistics, Federal University of Lavras, Lavras, Minas Gerais, Brazil.

Lucas M. Chaves
Department of Statistics, Federal University of Lavras, Lavras, Minas Gerais, Brazil.

1 Introduction

In many applications involving survival data analysis, there is a concomitant interest in assessing both covariate effect and the relationship between failure times. Multivariate survival analysis arises from this class of problems. In these settings, the independence assumption is often misleading, since failure times can be governed by an unknown dependence structure. Moreover, when subjects are allocated in clusters, we expect dependence to be more prominent within clusters rather than among clusters. This is an intuitive assumption, because subjects in a same cluster are affected by the environment in a similar fashion and tend to share some characteristics. Clustered survival data analysis is crucial in many areas. For example, in biomedical studies, when dealing with multicenter clinical trials, where patients are clustered according to their treatment center. In agriculture, when analyzing infectious disease data of livestock grouped in herds. In finance, when assessing time to default over different portfolios. Copula models and frailty models are two techniques commonly used to analyze these types of clustered event time data.

In frailty models, it is assumed that failure times within a cluster are conditionally independent given the frailty. This frailty is incorporated in the model as a multiplicative term represented by u_i, the frailty effect for the i^{th} cluster, and is an actual realization of a latent variable U following a particular distribution (the frailty distribution) with unit mean and finite variance. In this sense, individuals with $u_i > 1$ are considered frail, due to an increased hazard, i.e., higher risk of failure, whereas individuals with $u_i < 1$ have a lower risk of failure. In these models, the interest lies in obtaining the hazard function of an individual given the frailty effect. A thorough study of frailty models can be found in Duchateau and Janssen (2008) and Wienke (2011).

Frailty models have the advantage of allowing clusters to have different sizes. However, a major drawback is in the fact that association is not captured directly, since the frailty parameter (variance of the frailty distribution) only represents the heterogeneity between clusters and not how the individuals are linked to each other within a cluster. Also, it can be shown that this parameter is exclusively contained in the marginal survival functions (Goethals et al., 2008). In other words, it is not possible to explicitly model the intracluster dependence, but rather show how different are the clusters. Moreover, the estimation of the frailty parameter is only reliable when there is overdispersion in the data (Prenen et al., 2017a).

Owing to the groundbreaking work of Sklar (1959), copula-based models can easily overcome the fundamental issues in frailty models. This follows from the role that copulas play in multivariate models by working (in a survival analysis context) as a link between the marginal survival functions and the joint survival function for subjects. That is, copulas can be seen as dependence models separated from the margins. In view of this, it is possible to explicitly estimate the association between subjects, in our case, intracluster dependence.

Parametric estimation in copula models is usually done by taking either a frequentist or Bayesian approach. Frequentist approaches typically rely on maximum likelihood estimation methods, e.g., one-stage and two-stage. The first is the classical maximum likelihood, where the parameters from the copula and marginal survival

functions are simultaneously estimated; while in the second method, commonly referred to as inference functions for margins (Xu, 1996), estimation is done in two stages: parameters from the margins are estimated first, by assuming independence, and, in a second stage, the association parameter of the copula is obtained by maximizing the likelihood function with the parameters from the marginal survival functions fixed at the estimates from the first stage. Although less reliable, the two-stage method is preferred over the one-stage when estimation is computationally expensive.

One of the pioneer works in multivariate survival data modeling with copulas is due to Shih and Louis (1995). The authors provided estimation methods and derived asymptotic results for the estimators under a bivariate setting. Following their work, Andersen (2005) incorporated covariates in the model for bivariate data. Slightly improving the cluster size issue, Massonnet et al. (2009) proposed a quadrivariate copula model to study the time until infection in the four quarters of a cow udder. Despite the fact that clusters with a fixed size K were admissible in the model of Glidden (2000) and with varying size in Othus and Li (2010), the choice of copulas was restricted to the Clayton and Gaussian families, respectively.

A more flexible class of models, built with Archimedean copulas based on Laplace transforms, was proposed by Prenen et al. (2017a). In their case, clusters were allowed to have variable size, although important classes of copulas were still not comprehended. For example, elliptical (Gaussian, t) and extreme-value copulas (except for the also Archimedean, Gumbel-Hougaard copula). In a similar way, Romeo et al. (2018) proposed a model based on the two-parameter Archimedean family of Power Variance Function (PVF) copulas, but differently from the frequentist methods employed by Prenen et al. (2017a), estimation was performed by taking a one-stage Bayesian approach.

In this work, we extend the current techniques on clustered survival data modeling to a new level by developing a comprehensive class of copula-based models that imposes no restriction on the choice of copula families. This provides the means to model survival data that exhibit different types of dependence behaviors, ranging from symmetric positive to tail dependence with possible tail asymmetries. Marginal survival functions can also be flexibly modeled using different parametric families and covariate effects. A semiparametric formulation of the marginal survival function is also possible, with incorporation of time dependent covariates. Furthermore, there are no restrictions on cluster sizes, which may also be variable. The new class of models is based on the factor copula proposed by Krupskii and Joe (2013) and broadly extends the works of Prenen et al. (2017a) and Othus and Li (2010).

We provide three estimation methods: both one- and two-stage parametric, and a two-stage semiparametric method with marginal survival functions estimated by using a Cox proportional hazards model (Cox, 1972). Estimators derived under all three estimation procedures are shown to be consistent and asymptotically normally distributed. Their finite sample behavior is investigated in simulation studies. All estimation methods were implemented in `R` (R Core Team, 2018), and, in general, numerical computations are reasonably fast. We also provide an analytical alternative to the jackknife method employed by Othus and Li (2010) and Prenen et al. (2017a), which drastically reduces the computational cost for estimations under the semiparametric procedure.

Our paper is arranged as follows: a detailed description of the model is given in Section 2. The estimation procedures are explored in Section 3 and a simulation study is detailed in Section 4. In Section 5 we illustrate the methodology with a real data example. Proofs for the different asymptotic results stated in Section 3.3 can be found in the Appendix.

2 Description of the model

We consider the case of clusters with variable sizes, but settings with fixed cluster size are also supported by our methodology. Denote the number of clusters by K and the lifetime of individuals by a positive random variable T_{ij}, with $j = 1,...,n_i$ representing the j^{th} individual within cluster i ($i = 1,...,K$), and n_i the size of the i^{th} cluster. For every individual we assume an independent random censoring variable C_{ij}. Considering a right censoring scheme, the observed quantities are

$$\begin{aligned} &X_{ij} = \min(T_{ij}, C_{ij}) \quad \text{and} \\ &\delta_{ij} = I(T_{ij} \leqslant C_{ij}) \quad i = 1,...,K, \;\; j = 1,...,n_i. \end{aligned}$$

Assuming that each lifetime T_{ij} depends on a vector $\mathbf{Z}_{ij}$ of covariates (possibly time dependent), the joint survival function for cluster i is given by

$$\begin{aligned} S(t_{i1},...,t_{in_i}|\mathbf{Z}_{i1},....,\mathbf{Z}_{in_i}) &= P[T_{i1} > t_{i1},...,T_{in_i} > t_{in_i}|\mathbf{Z}_{i1},....,\mathbf{Z}_{in_i}] \\ &= \mathrm{C}(S(t_{i1}|\mathbf{Z}_{i1}),...,S(t_{in_i}|\mathbf{Z}_{in_i})), \end{aligned} \tag{1}$$

where $S(t_{ij}|\mathbf{Z}_{ij})$ is the marginal survival function of individual j within cluster i and C is the copula that joins the marginal survival functions to the joint survival function of the individuals in cluster i. Even though (1) being a straightforward representation of the joint survival function, we shall not use it as is. The main reason is that multivariate copulas can become inflexible for large clusters. Instead, we adopt conditional independence as a construction method, such that the more flexible bivariate copulas can be used as building blocks for the joint survival function in cluster i.

Let V_i ($i = 1,...,K$) be Uniform$[0,1]$ random variables. We assume that, within cluster i, the lifetimes are conditionally independent given V_i. In other words, V_i behaves as a latent variable that explains the dependence in cluster i. Hence, conditional on V_i, we can redefine (1), the joint survival function in cluster i, as

$$\begin{aligned} S(t_{i1},...,t_{in_i}|V_i,\mathbf{Z}_{i1},...,\mathbf{Z}_{in_i}) &= \mathrm{C}(S_{\cdot|V}(t_{i1}|V_i,\mathbf{Z}_{i1}),...,S_{\cdot|V}(t_{in_i}|V_i,\mathbf{Z}_{in_i})) \\ &= \prod_{j=1}^{n_i} S_{\cdot|V}(t_{ij}|V_i,\mathbf{Z}_{ij}), \end{aligned}$$

where $S_{\cdot|V}(t_{ij}|V_i,\mathbf{Z}_{ij})$ is the baseline conditional survival function of $T_{ij}|\mathbf{Z}_{ij}$ given $V_i = v_i$, that is, $S_{\cdot|V}(t_{ij}|V_i,\mathbf{Z}_{ij})$ is the partial derivative of the bivariate copula $\mathrm{C}_{\cdot V}(u_{ij},v_i;\theta)$ with respect to v_i for $u_{ij} = S(t_{ij}|\mathbf{Z}_{ij})$, the baseline marginal survival function of $T_{ij}|\mathbf{Z}_{ij}$

$$S_{\cdot|V}(t_{ij}|V_i,\mathbf{Z}_{ij}) = \frac{\partial}{\partial v_i}\mathrm{C}_{\cdot V}(S(t_{ij}|\mathbf{Z}_{ij}),v_i;\theta).$$

Due to the direct relationship of the baseline conditional survival function $S_{\cdot|V}(t_{ij}|V_i,\mathbf{Z}_{ij})$ with the bivariate copula $\mathrm{C}_{\cdot V}(u_{ij},v_i;\theta)$, we shall denote the former by $\mathrm{C}_{\cdot|V}(S(t_{ij}|\mathbf{Z}_{ij})|v_i)$. Thus, we have that

$$S(t_{i1},...,t_{in_i}|V_i,\mathbf{Z}_{i1},...,\mathbf{Z}_{in_i})=\prod_{j=1}^{n_i}\mathrm{C}_{\cdot|V}(S(t_{ij}|\mathbf{Z}_{ij})|v_i). \tag{2}$$

Now we can retrieve the unconditional joint survival function of cluster i by integrating V_i out of (2)

$$S(t_{i1},...,t_{in_i}|\mathbf{Z}_{i1},...,\mathbf{Z}_{in_i})=\int_0^1\prod_{j=1}^{n_i}\mathrm{C}_{\cdot|V}(S(t_{ij}|\mathbf{Z}_{ij})|v_i)dv_i. \tag{3}$$

Following Krupskii and Joe (2013), we will call (3) a one-factor copula model.

In the presence of right censoring, the contribution of cluster i to the likelihood is obtained by taking derivatives over the uncensored observations in cluster i

$$L_i=(-1)^{d_i}\frac{\partial^{d_i}}{(\partial x_{i1})^{\delta_{i1}}...\,(\partial x_{in_i})^{\delta_{in_i}}}S(x_{i1},...,x_{in_i}|\mathbf{Z}_{i1},....,\mathbf{Z}_{in_i}),$$

where $d_i=\sum_{j=1}^{n_i}\delta_{ij}$. Using representation (3) and assuming that differentiation and integration are interchangeable, the contribution to the likelihood can be expressed as

$$\begin{aligned}L_i&=(-1)^{d_i}\frac{\partial^{d_i}}{(\partial x_{i1})^{\delta_{i1}}...\,(\partial x_{in_i})^{\delta_{in_i}}}\int_0^1\prod_{j=1}^{n_i}\mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)\,dv_i\\&=(-1)^{d_i}\int_0^1\frac{\partial^{d_i}}{(\partial x_{i1})^{\delta_{i1}}...\,(\partial x_{in_i})^{\delta_{in_i}}}\prod_{j=1}^{n_i}\mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)\,dv_i\\&=(-1)^{d_i}\int_0^1\prod_{j=1}^{n_i}\left\{\frac{\partial}{\partial x_{ij}}\mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)\right\}^{\delta_{ij}}\times\mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)^{1-\delta_{ij}}dv_i.\end{aligned}$$

We also assume that the bivariate copula $\mathrm{C}_{\cdot V}$ is absolutely continuous, such that its density $\mathrm{c}_{\cdot V}(u_{ij},v_i)=\frac{\partial}{\partial u_{ij}}\mathrm{C}_{\cdot|V}(u_{ij}|v_i)=\frac{\partial^2}{\partial u_{ij}\partial v_i}\mathrm{C}_{\cdot V}(u_{ij},v_i)$ exists. Then

$$\begin{aligned}L_i&=(-1)^{d_i}\int_0^1\prod_{j=1}^{n_i}\left\{\mathrm{c}_{\cdot V}(S(x_{ij}|\mathbf{Z}_{ij}),v_i)(-f(x_{ij}|\mathbf{Z}_{ij}))\right\}^{\delta_{ij}}\times\mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)^{1-\delta_{ij}}dv_i\\&=\int_0^1\prod_{j=1}^{n_i}\left\{\mathrm{c}_{\cdot V}(S(x_{ij}|\mathbf{Z}_{ij}),v_i)f(x_{ij}|\mathbf{Z}_{ij})\right\}^{\delta_{ij}}\times\mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)^{1-\delta_{ij}}dv_i,\end{aligned}$$

where $f(x_{ij}|\mathbf{Z}_{ij})=-dS/dx_{ij}$ is the conditional density of the lifetime X_{ij}. Therefore, by taking the product $\prod_{i=1}^{K}L_i$, that is, combining the contribution of all clusters, we have the likelihood function

$$L=\prod_{i=1}^{K}\int_0^1\prod_{j=1}^{n_i}\left\{\mathrm{c}_{\cdot V}(S(x_{ij}|\mathbf{Z}_{ij}),v_i)f(x_{ij}|\mathbf{Z}_{ij})\right\}^{\delta_{ij}}\times\mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)^{1-\delta_{ij}}dv_i. \tag{4}$$

One advantage of the proposed model is that the likelihood function is only determined by the number of uncensored observations in each cluster. This follows from the joint survival functions of the clusters having exchangeable margins. Therefore, it is possible for clusters to have different sizes. On the other hand, a direct consequence of these unbalanced settings is that the integrals in (4) are, in practice, infeasible. This is because every configuration leads to a different and complicated integral. For this reason, numerical integration methods are required to evaluate the likelihood function. One avenue is to use Gauss-Legendre quadrature, as suggested by Krupskii and Joe (2013). In this case, the expression of the likelihood becomes

$$L \approx \prod_{i=1}^{K} \sum_{k=1}^{n_q} w_k \prod_{j=1}^{n_i} \left\{ \mathrm{c}_{\cdot V}(S(x_{ij}|\mathbf{Z}_{ij}), y_k^*) f(x_{ij}|\mathbf{Z}_{ij}) \right\}^{\delta_{ij}} \times \mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|y_k^*)^{1-\delta_{ij}},$$

where w_k and $y_k^* = 0.5y_k + 0.5$ are the weights and nodes of the quadrature, respectively. Krupskii and Joe (2013) also pointed out that a reasonable choice for the number of points of the quadrature, n_q, is around 21-25 for a one-factor copula model. However, we find that in our case estimation results are only reliable for $n_q \geq 50$. As an alternative, we also use the adaptive quadrature method of Gauss-Kronrod for numerical integration. It can be the case that, when an elevated number of quadrature points is needed in the Gauss-Legendre quadrature, the adaptive method tends to be computationally more efficient. Additional details about the computational aspects are given in Section 4.

3 Estimation

Our estimation procedures are based in two common frequentist techniques for copula models, the one- and two-stage methods. The former estimates the association and the marginals parameters simultaneously, whereas the latter splits the estimation procedure in two parts, first estimating the parameters of the marginal survival functions and then, conditional on these estimates, the association parameter is estimated in a second step. We investigate these two methods and, in addition, a two-stage semi-parametric approach, where the marginal survival functions are estimated by using a Cox proportional hazards model. Under Archimedean copula models, Prenen et al. (2017a) studied the same estimation procedures and derived asymptotic results. We extend their work by considering a more general factor copula model.

In a balanced design, with all clusters having a fixed size n, it is possible to order the components within the clusters, therefore allowing the estimation of a different baseline survival function for each element in the cluster, whilst having the same covariate information for every subject. In our case, the clusters have different sizes, thus making it impossible to assume a different survival function for each individual. For this reason, we proceed by defining a unique baseline survival function for all individuals, allowing for subject-specific covariate information.

3.1 One-stage procedure

The one-stage procedure is the classical maximum likelihood approach, where the association and the marginal survival function's parameters are simultaneously estimated by finding the maxima of the likelihood function. Let $\boldsymbol{\beta}$ be the p-dimensional parametric vector for the baseline survival function S, containing distribution and covariate information. Also, let θ be the association parameter for individuals within every cluster, i.e., the parameter of the underlying copula C. Let $L(\boldsymbol{\beta},\theta)$ be the likelihood function as derived in (4). The maximum likelihood estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\theta}$ are yielded by solving the score equations

$$\begin{cases} \mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\theta) = \mathbf{0} \\ U_{\theta}(\boldsymbol{\beta},\theta) = 0, \end{cases}$$

where $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\theta) = \frac{\partial}{\partial \boldsymbol{\beta}} \log L(\boldsymbol{\beta},\theta)$ and $U_{\theta}(\boldsymbol{\beta},\theta) = \frac{\partial}{\partial \theta} \log L(\boldsymbol{\beta},\theta)$. It is known from maximum likelihood theory (Cox and Hinkley, 1974; Lehmann and Casella, 1998) that, under customary regularity conditions, $\sqrt{K}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widehat{\theta} - \theta\right)$ converges to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}^{-1}$. We partition the Fisher information matrix $\mathbf{I} = E_{\boldsymbol{\eta}}\left[-\nabla^2 \log L(\boldsymbol{\eta})\right]$, $\boldsymbol{\eta} = (\boldsymbol{\beta},\theta)$ with size $(p+1)$ as follows

$$\mathbf{I} = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\beta}\theta} \\ \mathbf{I}_{\theta\boldsymbol{\beta}} & I_{\theta\theta} \end{pmatrix}.$$

Using that the inverse of a block matrix (Henderson and Searle, 1981)

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

is

$$\mathbf{A}^{-1} = \begin{pmatrix} \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & -\mathbf{A}_{11}\mathbf{A}_{12}\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \end{pmatrix},$$

we get the expression for the asymptotic variance of $\widehat{\theta}$, the lower right element of $\mathbf{I}^{-1}$,

$$\operatorname{var}(\widehat{\theta}) = \left(I_{\theta\theta} - \mathbf{I}_{\theta\boldsymbol{\beta}}\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\mathbf{I}_{\boldsymbol{\beta}\theta}\right)^{-1},$$

which, by the identity

$$(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} = \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\left(\mathbf{A}_{22}^{-1} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$

can be rewritten as

$$\operatorname{var}(\widehat{\theta}) = \frac{1}{I_{\theta\theta}} + \frac{\mathbf{I}_{\theta\boldsymbol{\beta}}\left(\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \frac{\mathbf{I}_{\boldsymbol{\beta}\theta}\mathbf{I}_{\boldsymbol{\beta}\theta}^{T}}{I_{\theta\theta}}\right)^{-1}\mathbf{I}_{\boldsymbol{\beta}\theta}}{I_{\theta\theta}^2}.$$

Considering that the information matrix $\mathbf{I}$ depends on the unknown quantities $\boldsymbol{\beta}$ and θ, in practical applications standard errors for parameter estimates are given by the square root of the diagonal of the inverse of the Hessian matrix evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\theta = \widehat{\theta}$.

3.2 Two-stage parametric estimation

In the two-stage procedure, we use the method of inference functions for margins (IFM), proposed by Xu (1996). Differently from the one-stage procedure, estimation now is carried out in two steps. The first stage consists in estimating the marginal survival function's parameters alone, not taking into account the intracluster dependence. In the second stage we estimate the copula's association parameter whilst fixing the likelihood for the estimates of the first stage. This method is preferred over the one-stage procedure when full likelihood estimation is computationally expensive.

Formally: Let $\boldsymbol{\beta}$ and θ be defined as in Section 3.1. In the first stage, $\boldsymbol{\beta}$ is estimated considering the lifetimes T_{ij} as independent and identically distributed random variables, i.e., by solving

$$\begin{aligned}\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta}) &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\delta_{ij}\frac{\partial \log f(x_{ij}|\mathbf{Z}_{ij})}{\partial \boldsymbol{\beta}} + (1-\delta_{ij})\frac{\partial \log S(x_{ij}|\mathbf{Z}_{ij})}{\partial \boldsymbol{\beta}}\right)\\ &= \sum_{i=1}^{K}\mathbf{U}^*_{i,\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\mathbf{U}^*_{i,j,\boldsymbol{\beta}}(\boldsymbol{\beta}) = \mathbf{0}. \qquad (5)\end{aligned}$$

Let $\bar{\boldsymbol{\beta}}$ be the estimator obtained from (5). Under regularity conditions, $\sqrt{K}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$, where $\mathbf{V}$ is the variance-covariance matrix of the score functions $\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta})$;

$$\mathbf{V} = E[\mathbf{U}^*_{i,\boldsymbol{\beta}}(\boldsymbol{\beta}_0)\mathbf{U}^*_{i,\boldsymbol{\beta}}(\boldsymbol{\beta}_0)^T]$$

and $\mathbf{I}^*$ is the Fisher information matrix of $\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta})$;

$$\mathbf{I}^* = E\left[-\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}^*_{i,\boldsymbol{\beta}}(\boldsymbol{\beta}_0)\right].$$

$\boldsymbol{\beta}_0$ is the true parametric vector.

Due to misspecification of the model, i.e., assuming independence between the random variables T_{ij} when they are actually dependent, the usual inverse of the Fisher information, $(\mathbf{I}^*)^{-1}$, is not a consistent estimator of the asymptotic variance-covariance matrix. Hence, we use the robust sandwich estimator $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$.

In the second stage, the association parameter θ is estimated by plugging $\bar{\boldsymbol{\beta}}$, obtained in the first stage, in the full likelihood (4) and solving

$$U_\theta(\bar{\boldsymbol{\beta}}, \theta) = \frac{\partial}{\partial \theta}\log L(\bar{\boldsymbol{\beta}}, \theta) = 0$$

for θ. Thus, obtaining the two-stage estimator for θ.

Theorem 1 *Let $\bar{\theta}$ be the two-stage estimator for θ, obtained from $U_\theta(\bar{\boldsymbol{\beta}}, \theta) = 0$. Under regularity conditions (see Xu, 1996), $\sqrt{K}(\bar{\theta} - \theta)$ converges to a normal distribution with mean 0 and variance*

$$\operatorname{var}(\bar{\theta}) = \frac{1}{I_{\theta\theta}} + \frac{\mathbf{I}_{\theta\boldsymbol{\beta}}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\boldsymbol{\beta}\theta}}{I^2_{\theta\theta}}. \qquad (6)$$

A proof of this theorem is given in Prenen et al. (2017a) for Archimedean copulas, but it holds all the same in our context.

We estimate the variance term in (6) by replacing $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$ with $\left(\widehat{\mathbf{I}}^*\right)^{-1}\widehat{\mathbf{V}}\left(\widehat{\mathbf{I}}^*\right)^{-1}$, where

$$\widehat{\mathbf{I}}^* = \sum_{i=1}^{K}\sum_{j=1}^{n_i} -\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}^*_{i,j,\boldsymbol{\beta}}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}$$

and

$$\widehat{\mathbf{V}} = \sum_{i=1}^{K}\left(\sum_{j=1}^{n_i}\mathbf{U}^*_{i,j,\boldsymbol{\beta}}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}\right)\left(\sum_{j=1}^{n_i}\mathbf{U}^*_{i,j,\boldsymbol{\beta}}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}\right)^T.$$

The quantities $I_{\theta\theta}$, $\mathbf{I}_{\theta\boldsymbol{\beta}}$ and $\mathbf{I}_{\boldsymbol{\beta}\theta}$ are obtained from the Hessian matrix by performing one iteration of the one-stage procedure with $\boldsymbol{\beta}$ fixed at $\bar{\boldsymbol{\beta}}$ and θ at $\bar{\theta}$.

3.3 Two-stage semiparametric estimation

If a more flexible setting for the margins is desired, rather than using fully parametric models, it is possible to estimate the margins by taking a semiparametric approach. In this case, we use the Cox proportional hazards model (Cox, 1972). Estimation now consists in obtaining, for the first stage, $\check{\boldsymbol{\beta}}$ and $\check{\Lambda}$, the estimated covariate effects and cumulative hazard function, respectively. As in the two-stage parametric method, it is assumed that the subjects are independent in the first stage, the so-called independence working assumption. Also, a common baseline hazard function is assumed for all individuals, but allowing for subject-specific covariate information, which can also depend on time. Estimators for $\boldsymbol{\beta}$ and Λ along with formulas for their standard errors can be found in Spiekerman and Lin (1998).

In the second stage, the estimate $\check{\theta}$ of the copula's association parameter is retrieved by maximizing the likelihood for θ whilst fixing for the first stage estimates, i.e., by solving $\max_\theta\left\{L\left(\theta,\check{\boldsymbol{\beta}},\check{\Lambda}\right)\right\}$.

Theorem 2 *Under regularity conditions 1-8 in the Appendix,* $\left(\check{\theta},\check{\boldsymbol{\beta}},\check{\Lambda}\right)$ *are consistent estimators for* $(\theta_0,\boldsymbol{\beta}_0,\Lambda_0)$*, the true parameters.*

The proof for the consistency of $\left(\check{\boldsymbol{\beta}},\check{\Lambda}\right)$ can be found in Spiekerman and Lin (1998) whereas the consistency of $\check{\theta}$ is proved in the Appendix, following ideas from Prenen et al. (2017a) and Othus and Li (2010).

Theorem 3 *Under regularity conditions 1-8 in the Appendix,* $\sqrt{K}\left(\check{\theta}-\theta_0\right)$ *converges to a normal distribution with mean 0 and variance equal to*

$$\mathrm{var}\left(\Xi\right)/W(\theta_0)^2. \tag{7}$$

A proof for Theorem 3 and the formal definitions of Ξ and $W(\theta_0)$ are presented in the Appendix. We derive this proof by extending the results in Prenen et al. (2017a) from Archimedean copulas to the more general factor copula models. We also provide the formulae to compute estimates for the standard error of $\breve{\theta}$. This drastically reduces the computational cost when compared to the jackknife alternative used by Prenen et al. (2017a).

4 Simulation study

In order to assess the finite sample behavior of the estimators, we simulate 1000 data sets under different settings. For the number of clusters, we use $K = 50$, 200 and 500, with each cluster having size varying uniformly from 2 to 50. We use the Clayton ($\theta = 1.07$, 2.383, 4.816), Gaussian ($\theta = 0.556$, 0.767, 0.899) and Galambos ($\theta = 0.866$, 1.538, 2.78) copulas to simulate intracluster dependence, such that we have representatives from different classes (Archimedean, elliptical and extreme-value copulas, respectively) and three levels of dependence for each case (Kendall's $\tau \approx 0.35$, 0.55, 0.7). Individual lifetimes are generated from a Weibull distribution, with survival function given by $S(t|Z) = \exp\{-\lambda \exp(\beta z) t^{\rho}\}$ and choosing $\lambda = 0.5$, $\rho = 1.6$ and Z a dichotomous covariate with effect $\beta = 2$. Data are generated using the sampling algorithm proposed by Joe (2014). We consider three different censoring scenarios: 25%, 50% and no censoring. Censoring times are obtained from a Weibull distribution with parameters $\lambda_C = 0.425$, $\rho_C = 1.6$, for 25% of censoring and $\lambda_C = 2.241$, $\rho_C = 1.6$, for 50%.

Simulation results for the three estimation methods are summarized in Tables 1, 2 and 3 for $K = 50, 200$ and 500 clusters, respectively. In all three scenarios, we provide, for the Clayton, Gaussian and Galambos copulas, the mean estimated values of $\hat{\theta}$, $\bar{\theta}$ and $\breve{\theta}$ in the first rows, along with their mean estimated standard errors and coverage of 95% confidence intervals in the second rows. In the parametric one-stage method, standard errors are retrieved from the inverse of the Hessian matrix, whereas in the parametric and semiparametric two-stage, we obtain the estimates of the standard errors via formulas (6) and (7), respectively. Moreover, by using the plug-in estimator of the standard error in the semiparametric two-stage method, we noticeably reduce the computing time if compared to the grouped jackknife alternative employed by Prenen et al. (2017a) and Othus and Li (2010). We deal with the infeasible integrals in the likelihood expressions by using a Gauss-Legendre quadrature rule with $n_q = 50$ points, resulting in reasonable accuracy at a small computational cost for the parametric and semiparametric two-stage methods. In contrast, the parametric one-stage estimator is highly sensible (specially for high values of Kendall's τ) to the number of quadrature points, thus making necessary to use at least $n_q = 200$ points when $K = 50$. This effect is magnified for larger values of K, therefore making the parametric one-stage computationally expensive. However, by using the adaptive quadrature of Gauss-Kronrod we were able to mitigate this issue in the one-stage method. Nevertheless, the two-stage methods are still the better option regarding computational time. As evidenced in Tables 2 and 3, the parametric and semiparametric two-stage

Table 1 Simulation results for 50 clusters of varying sizes ranging from 2 to 50. Mean estimated values of $\hat{\theta}$, $\bar{\theta}$ and $\breve{\theta}$ are in the first rows, along with their mean estimated standard errors and coverage of 95% confidence intervals in the second rows

Copula model	τ	θ_0	0% censoring			25% censoring			50% censoring		
			Parametric one-stage	Parametric two-stage	Semiparametric two-stage	Parametric one-stage	Parametric two-stage	Semiparametric two-stage	Parametric one-stage	Parametric two-stage	Semiparametric two-stage
Clayton	0.35	1.07	1.069	1.051	1.018	1.065	1.060	1.049	1.053	1.059	1.047
			(0.176; 92.6%)	(0.189; 85.3%)	(0.212; 83.1%)	(0.186; 94.7%)	(0.206; 87.4%)	(0.224; 89.1%)	(0.199; 93.8%)	(0.213; 90.9%)	(0.231; 92.1%)
	0.55	2.383	2.362	2.325	2.161	2.378	2.352	2.284	2.373	2.363	2.316
			(0.352; 90.1%)	(0.385; 84.1%)	(0.435; 83.4%)	(0.378; 93.5%)	(0.427; 90.2%)	(0.483; 90.3%)	(0.402; 93.1%)	(0.453; 90.7%)	(0.495; 93.2%)
	0.7	4.816	4.696	4.665	4.159	4.738	4.686	4.443	4.837	4.785	4.591
			(0.664; 83.0%)	(0.723; 79.2%)	(0.825; 79.5%)	(0.680; 89.7%)	(0.801; 87.1%)	(0.934; 90.5%)	(0.780; 94.5%)	(0.903; 93.0%)	(1.014; 92.7%)
Gaussian	0.35	0.556	0.549	0.544	0.547	0.548	0.546	0.550	0.549	0.545	0.547
			(0.041; 93.2%)	(0.042; 92.1%)	(0.045; 91.7%)	(0.043; 93.7%)	(0.044; 92.0%)	(0.045; 92.5%)	(0.045; 93.3%)	(0.046; 93.9%)	(0.050; 92.1%)
	0.55	0.767	0.762	0.757	0.755	0.762	0.758	0.755	0.760	0.757	0.758
			(0.027; 95.1%)	(0.033; 91.3%)	(0.037; 94.8%)	(0.028; 94.7%)	(0.034; 91.7%)	(0.038; 93.7%)	(0.031; 94.3%)	(0.036; 93.4%)	(0.041; 91.4%)
	0.7	0.899	0.897	0.892	0.855	0.897	0.894	0.889	0.895	0.892	0.887
			(0.012; 93.8%)	(0.018; 91.0%)	(0.022; 94.0%)	(0.013; 94.6%)	(0.019; 94.0%)	(0.021; 94.5%)	(0.015; 94.4%)	(0.021; 93.4%)	(0.027; 93.4%)
Galambos	0.35	0.866	0.858	0.854	0.849	0.859	0.858	0.851	0.860	0.856	0.850
			(0.069; 94.3%)	(0.085; 89.8%)	(0.091; 84.4%)	(0.076; 93.4%)	(0.093; 89.3%)	(0.092; 83.5%)	(0.087; 93.4%)	(0.104; 87.6%)	(0.099; 81.9%)
	0.55	1.538	1.535	1.519	1.495	1.529	1.524	1.479	1.527	1.506	1.475
			(0.126; 93.1%)	(0.168; 89.9%)	(0.163; 82.2%)	(0.139; 92.6%)	(0.185; 88.6%)	(0.161; 79.7%)	(0.159; 92.4%)	(0.197; 85.5%)	(0.169; 75.6%)
	0.7	2.78	2.705	2.718	2.573	2.703	2.724	2.566	2.715	2.710	2.529
			(0.225; 88.4%)	(0.317; 88.0%)	(0.262; 76.1%)	(0.252; 88.1%)	(0.342; 87.3%)	(0.268; 70.8%)	(0.273; 87.6%)	(0.375; 84.2%)	(0.271; 64.6%)

Table 2 Simulation results for 200 clusters of varying sizes ranging from 2 to 50. Mean estimated values of $\hat{\theta}$, $\bar{\theta}$ and $\breve{\theta}$ are in the first rows, along with their mean estimated standard errors and coverage of 95% confidence intervals in the second rows

Copula model	τ	θ_0	0% censoring			25% censoring			50% censoring		
			Parametric one-stage	Parametric two-stage	Semiparametric two-stage	Parametric one-stage	Parametric two-stage	Semiparametric two-stage	Parametric one-stage	Parametric two-stage	Semiparametric two-stage
Clayton	0.35	1.07	1.067	1.067	1.053	1.068	1.067	1.062	1.068	1.060	1.060
			(0.088; 94.3%)	(0.106; 89.5%)	(0.115; 87.9%)	(0.092; 93.1%)	(0.108; 93.6%)	(0.110; 92.2%)	(0.101; 94.0%)	(0.110; 94.6%)	(0.110; 93.2%)
	0.55	2.383	2.364	2.362	2.305	2.372	2.373	2.339	2.392	2.379	2.372
			(0.184; 93.5%)	(0.222; 90.7%)	(0.230; 89.7%)	(0.190; 94.1%)	(0.229; 94.2%)	(0.233; 92.6%)	(0.201; 94.4%)	(0.231; 93.1%)	(0.234; 93.7%)
	0.7	4.816	4.769	4.715	4.561	4.831	4.787	4.686	4.837	4.822	4.755
			(0.346; 90.6%)	(0.419; 88.7%)	(0.412; 84.7%)	(0.369; 94.2%)	(0.431; 92.1%)	(0.441; 92.2%)	(0.386; 93.2%)	(0.460; 93.6%)	(0.468; 93.3%)
Gaussian	0.35	0.556	0.554	0.555	0.553	0.555	0.554	0.553	0.554	0.553	0.555
			(0.020; 94.0%)	(0.022; 93.9%)	(0.021; 91.4%)	(0.021; 93.6%)	(0.022; 94.3%)	(0.021; 94.4%)	(0.023; 93.2%)	(0.024; 94.3%)	(0.022; 92.7%)
	0.55	0.767	0.765	0.765	0.764	0.766	0.765	0.765	0.766	0.764	0.764
			(0.013; 95.6%)	(0.017; 94.2%)	(0.016; 92.2%)	(0.014; 95.0%)	(0.018; 94.6%)	(0.017; 92.3%)	(0.015; 96.2%)	(0.019; 94.9%)	(0.018; 90.0%)
	0.7	0.899	0.899	0.897	0.895	0.898	0.897	0.896	0.898	0.897	0.896
			(0.006; 95.3%)	(0.009; 92.8%)	(0.009; 92.4%)	(0.006; 93.9%)	(0.010; 94.3%)	(0.009; 95.5%)	(0.007; 94.7%)	(0.010; 94.0%)	(0.010; 90.3%)
Galambos	0.35	0.866	0.863	0.865	0.862	0.864	0.867	0.860	0.863	0.863	0.857
			(0.035; 94.1%)	(0.045; 93.8%)	(0.051; 89.0%)	(0.038; 94.5%)	(0.050; 93.8%)	(0.053; 86.4%)	(0.044; 91.9%)	(0.056; 91.4%)	(0.057; 83.7%)
	0.55	1.538	1.537	1.532	1.524	1.538	1.533	1.520	1.531	1.536	1.519
			(0.065; 95.1%)	(0.090; 93.0%)	(0.087; 86.5%)	(0.073; 94.7%)	(0.098; 91.6%)	(0.094; 86.5%)	(0.083; 94.7%)	(0.111; 91.4%)	(0.102; 84.4%)
	0.7	2.78	2.778	2.757	2.709	2.754	2.754	2.711	2.766	2.747	2.693
			(0.128; 92.6%)	(0.171; 92.0%)	(0.146; 82.7%)	(0.140; 91.6%)	(0.185; 91.8%)	(0.153; 80.6%)	(0.158; 93.3%)	(0.208; 90.9%)	(0.167; 75.3%)

Table 3 Simulation results for 500 clusters of varying sizes ranging from 2 to 50. Mean estimated values of $\hat{\theta}$, $\bar{\theta}$ and $\breve{\theta}$ are in the first rows, along with their mean estimated standard errors and coverage of 95% confidence intervals in the second rows

Copula model	τ	θ_0	0% censoring			25% censoring			50% censoring		
			Parametric one-stage	Parametric two-stage	Semiparametric two-stage	Parametric one-stage	Parametric two-stage	Semiparametric two-stage	Parametric one-stage	Parametric two-stage	Semiparametric two-stage
Clayton	0.35	1.07	1.07	1.073	1.061	1.071	1.066	1.066	1.07	1.074	1.068
			(0.056; 94.2%)	(0.074; 93.0%)	(0.073; 93.3%)	(0.057; 92.6%)	(0.068; 94.4%)	(0.069; 92.4%)	(0.063; 92.3%)	(0.070; 94.9%)	(0.069; 93.3%)
	0.55	2.383	2.377	2.371	2.345	2.376	2.373	2.362	2.395	2.382	2.377
			(0.117; 92.0%)	(0.149; 93.4%)	(0.148; 90.3%)	(0.120; 93.3%)	(0.150; 93.5%)	(0.145; 93.5%)	(0.125; 94.0%)	(0.146; 94.8%)	(0.145; 94.2%)
	0.7	4.816	4.761	4.753	4.658	4.807	4.807	4.741	4.839	4.811	4.790
			(0.229; 94.0%)	(0.283; 93.3%)	(0.268; 86.2%)	(0.232; 95.7%)	(0.277; 93.4%)	(0.277; 92.4%)	(0.248; 93.8%)	(0.293; 94.0%)	(0.291; 94.3%)
Gaussian	0.35	0.556	0.556	0.555	0.556	0.556	0.555	0.556	0.555	0.555	0.556
			(0.013; 94.0%)	(0.014; 94.3%)	(0.013; 94.7%)	(0.013; 93.4%)	(0.014; 94.5%)	(0.013; 93.0%)	(0.014; 91.2%)	(0.015; 96.3%)	(0.014; 93.8%)
	0.55	0.767	0.766	0.766	0.766	0.766	0.766	0.766	0.767	0.767	0.766
			(0.008; 95.4%)	(0.011; 92.5%)	(0.010; 94.0%)	(0.009; 94.0%)	(0.011; 95.1%)	(0.010; 95.0%)	(0.010; 94.1%)	(0.012; 93.7%)	(0.011; 96.1%)
	0.7	0.899	0.899	0.898	0.898	0.899	0.899	0.898	0.899	0.898	0.898
			(0.004; 94.0%)	(0.006; 93.4%)	(0.005; 93.5%)	(0.004; 94.1%)	(0.006; 93.9%)	(0.006; 93.8%)	(0.005; 94.8%)	(0.007; 94.7%)	(0.006; 92.2%)
Galambos	0.35	0.866	0.866	0.863	0.862	0.865	0.865	0.862	0.864	0.864	0.866
			(0.022; 93.3%)	(0.029; 93.0%)	(0.033; 89.3%)	(0.024; 93.0%)	(0.032; 93.7%)	(0.034; 89.7%)	(0.028; 94.8%)	(0.037; 93.8%)	(0.038; 88.9%)
	0.55	1.538	1.538	1.533	1.525	1.536	1.533	1.527	1.538	1.527	1.526
			(0.041; 93.8%)	(0.058; 93.7%)	(0.057; 87.7%)	(0.046; 92.4%)	(0.063; 93.0%)	(0.060; 87.8%)	(0.055; 94.7%)	(0.072; 93.0%)	(0.066; 85.4%)
	0.7	2.78	2.769	2.769	2.739	2.758	2.770	2.743	2.752	2.758	2.728
			(0.073; 94.3%)	(0.111; 96.1%)	(0.095; 84.9%)	(0.090; 93.8%)	(0.121; 94.3%)	(0.101; 83.8%)	(0.103; 93.0%)	(0.137; 92.3%)	(0.110; 80.6%)

methods perform well when $K \geq 200$, yielding small biases and appropriate coverage probabilities at a much lower computational cost.

Censoring affects the performance of the estimators in different ways for the Clayton, Gaussian and Galambos copulas, although standard errors systematically increase as the censoring percentage increases (as expected). Due to the opposite nature of the Clayton and Galambos copulas (lower tail dependence versus upper tail dependence), we notice that, as the censoring percentage increases, coverage probabilities also increase for the Clayton copula, while the opposite happens for the Galambos copula. This can be seen for $K = 50$, 200, and, to a lesser extent, for $K = 500$ clusters. Coverage probabilities for the Gaussian copula are not significantly affected by censoring percentage, owing to its symmetrical dependence structure. As can be seen in Tables 1, 2 and 3, the strength of association, represented by the three values of Kendall's τ, has an intuitive impact on the estimators, i.e., higher values of Kendall's τ impose inferior results, while the results tend to be better for smaller values of Kendall's τ. This is specially perceivable for $K = 50$, and for the semiparametric two-stage method when $K \leq 200$. It is important to note that for samples with a small number of clusters (K around 50 clusters), the two-stage methods are not much reliable (Table 1). The parametric one-stage is recommended in these scenarios, as it gives better results in terms of bias and coverage probability. Fortunately, for $K \geq 200$, the two-stage methods have a good performance and a low computational cost.

5 Real data example - Insemination dataset

One possible application of the proposed methods is to model the time to first insemination after calving in dairy cattle clustered in herds. For this, we use the insemination dataset, available in the `R` package `Sunclarco` (Prenen et al., 2017b). This dataset consists of 181 clusters (farms) of different sizes, containing 10513 cows in total. The cluster sizes range from 1 to 174 cows and the times to first insemination are subject to right censoring, which makes this dataset suited for our purposes. Despite representing only 5.5% of the data, right censoring is still present, making it necessary to be considered in the modeling. This right censoring is due to no insemination of a cow within 330 days or if it is culled before insemination. The insemination dataset also contains covariate information, represented by the dichotomous covariate parity, which is 0 for multiparous cows and 1 for primiparous cows.

According to Duchateau and Janssen (2004), the time from parturition until first insemination is one of the main factors that determines the calving interval, which should be optimally between 12 and 13 months in order to maximize milk production. Usually, insemination is done by the farmer, relying only on his experience. By modeling the association between insemination times, we can get more insight into this process.

We use three different factor copulas to model the association between times to first insemination: the first built with a bivariate Clayton copula; the second using a bivariate Gaussian copula; and the third built with a bivariate Galambos copula. In all three settings, we use a baseline Weibull survival function (one- and two-stage para-

metric estimation) to model the times to first insemination and a Cox proportional hazards model for the two-stage semiparametric method, allowing for covariate information (parity of the cow). The expression for the Weibull survival function used is

$$S(t|Z) = \exp\{-\lambda \exp(\beta z) t^{\rho}\},$$

where β is the parity effect for the dichotomous covariate Z.

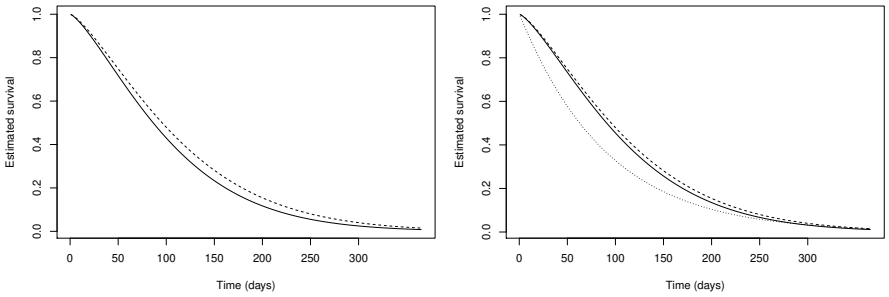
Using the one- and two-stage procedures for estimation, we provide results for the parity effect and association parameters for all three factor copula settings considered, along with their Akaike information criterion (AIC) (see Table 4). It is important to note that, since the parametric and semiparametric models are different in nature, one must be careful not to use the AIC to compare them.

Table 4 Estimation results for the insemination dataset.

	Results for Clayton copula			Results for Gaussian copula			Results for Galambos copula		
	Weibull one stage	Weibull two stage	Semiparametric two-stage	Weibull one stage	Weibull two stage	Semiparametric two-stage	Weibull one stage	Weibull two stage	Semiparametric two-stage
β	−0.138 (0.016)	−0.066 (0.022)	−0.060 (0.021)	−0.135 (0.019)	−0.066 (0.022)	−0.060 (0.021)	−0.100 (0.015)	−0.066 (0.022)	−0.060 (0.021)
θ	0.779 (0.023)	0.829 (0.126)	0.985 (0.098)	0.624 (0.029)	0.575 (0.034)	0.510 (0.021)	1.302 (0.048)	0.916 (0.038)	0.753 (0.035)
AIC	−705.0	−700.8	154.2	−703.8	−700.4	520.7	−690.3	−673.2	1072.4

The parity of the cow has a similar and coherent effect for all settings, with primiparous cows having a significantly lower hazard of experiencing the event (insemination). Indeed, for the one-stage method, hazard ratios are 0.87 (95% confidence interval (CI) $[0.84, 0.91]$), 0.87 (95% confidence interval (CI) $[0.84, 0.90]$) and 0.90 (95% confidence interval (CI) $[0.88, 0.93]$) for the Gaussian, Clayton and Galambos factor copulas, respectively. For the two-stage parametric method, all models lead to a hazard ratio of 0.94 (95% confidence interval (CI) $[0.89, 0.98]$). The hazard ratio for the semiparametric method is 0.94 (95% confidence interval (CI) $[0.90, 0.98]$) under all three copula models. According to the AIC, the Clayton factor copula presented the best fit among the three models for every estimation procedure. Considering that the Clayton copula has lower tail dependence, it can be inferred, in this context, that later times of insemination have a stronger association and lower values are weakly correlated. Moreover, under the one-stage method, the estimation results show that the times until insemination are significantly affected by the farm (aggregate of many exogenous variables). The strength of this association can be measured by the estimate of the association parameter (0.779 with 95% confidence interval (CI) $[0.734, 0.824]$), which is equivalent to a Kendall's τ of 0.28. This means that the new methodology is not only capable of controlling for cluster effect, but to assess the shape of intracluster dependence (any copula family can be used) and its strength.

As can be seen in Figure 1, for the Weibull-Clayton model, the estimated survival curve for primiparous cows is greater than the one from multiparous cows, meaning



(a) **(b)**

Fig. 1 (a) Estimated survival curves for multiparous (continuous) and primiparous (dashed) cows (Clayton model under one-stage estimation) and (b) estimated unconditional survival curves for the Clayton (continuous), Gaussian (dashed) and Galambos (dotted) factor copula models under one-stage estimation.

that multiparous cows are inseminated earlier than primiparous cows, with approximately 50% of the multiparous cows being inseminated before 86 days, while for primiparous cows, the estimated median is 96 days. This difference in the median time is more accentuated when comparing the estimated marginal survival curves for the three models (see Figure 1). Due to the upper tail dependence feature of the Galambos copula (stronger association for lower times of insemination), the estimated survival curve for the Weibull-Galambos model is notably less than the other two for lower values of the variable time until insemination. Indeed, estimated median survival times for the Gaussian and Clayton factor copulas are approximately 91 days, while for the Galambos it is 63 days. It is also possible to verify that the Clayton factor copula is more suited to the data by checking the sample median time until first insemination, which is 90 days for multiparous cows. This value is closest to the estimated median survival time in the Weibull-Clayton model.

6 Discussion

Current methodologies restrict clustered survival data modeling to settings where either cluster sizes are small and fixed or the number of copula families implemented is limited. This work aims to overcome these limitations. By using factor copulas, we developed a comprehensive class of models that allow for clusters to be large and with variable size, altogether with the flexibility of supporting any copula family in its structure. Owing to clusters having different sizes, we assume exchangeability between lifetimes within a cluster and proceed by estimating a common baseline survival function using the whole data set. Nonetheless, subject-specific covariate information is introduced (possibly time dependent). One drawback of the proposed models is the lack of analytical expressions for the likelihood, a consequence of the

infeasible integral in its definition. However, we can still obtain reliable results by using Gauss-Legendre integration with an appropriate number of quadrature points, or even adaptive methods, such as the Gauss-Kronrod quadrature. Three estimation methods were investigated: parametric one-stage and two-stage along with a semi-parametric two-stage approach. Additionally, we derived estimators and proved their consistency and asymptotic normality for all the three methods. Simulation results showed that the three methods behave reasonably well under different settings, with the one-stage procedure being, in general, more reliable for samples with a small number of clusters. On the other hand, the one-stage method is computationally demanding for a large number of clusters ($K \geq 200$). This is not an issue for both the parametric and semiparametric two-stage methods, since they yielded up to standard results for settings with a large number of clusters ($K \geq 200$). Moreover, the computational cost in the two-stage procedures is substantially reduced. This paper is an extension of the works of Prenen et al. (2017a), who investigated similar estimation methods under Archimedean copula based models, and Othus and Li (2010), who explored a semiparametric two-stage approach using Gaussian copulas. We also mention the foundational work of Shih and Louis (1995), who derived essential results for bivariate data.

Acknowledgements This research was supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES - PDSE; Process 88881.190074/2018-01). We would also like to express our gratitude to our colleagues from the Department of Statistics at the Federal University of Lavras (DES-UFLA) and I-BioStat at Universiteit Hasselt, who greatly assisted this work.

Appendix

For convenience, we first introduce some notations and definitions adapted from Prenen et al. (2017a) and Othus and Li (2010).

$$Y_{ij}(t) = I_{\left\{X_{ij} \geq t\right\}}$$

$$\check{\Lambda}(t) = \int_0^t \frac{d\sum_{i=1}^{K}\sum_{j=1}^{n_i} \delta_{ij} I_{\left\{X_{ij} \leq u\right\}}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} Y_{ij}(u)\exp\left\{\check{\boldsymbol{\beta}}'\mathbf{Z}_{ij}(u)\right\}} = \sum_{i=1}^{K}\sum_{j=1}^{n_i} \frac{\delta_{ij} I_{\left\{X_{ij} \leq t\right\}}}{\sum_{k=1}^{K}\sum_{l=1}^{n_k} I_{\left\{X_{kl} \leq X_{ij}\right\}}\exp\left\{\check{\boldsymbol{\beta}}'\mathbf{Z}_{kl}(X_{ij})\right\}},$$

$$H_{ij} = \exp\left[-\int_0^\tau Y_{ij}(u)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\right\} d\Lambda(u)\right],$$

$$H_{ij}^0 = \exp\left[-\int_0^\tau Y_{ij}(u)\exp\left\{\boldsymbol{\beta}_0'\mathbf{Z}_{ij}(u)\right\} d\Lambda_0(u)\right],$$

$$\check{H}_{ij} = \exp\left[-\int_0^\tau Y_{ij}(u)\exp\left\{\check{\boldsymbol{\beta}}'\mathbf{Z}_{ij}(u)\right\} d\check{\Lambda}(u)\right],$$

$$H_{ij}(t) = \exp\left[-\int_0^\tau Y_{ij}(u)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\right\} d\left\{\Lambda + t\left(\Gamma - \Lambda\right)\right\}(u)\right].$$

Note that $H_{ij} = H_{ij}(0)$.

$$\begin{aligned} L(\theta,\boldsymbol{\beta},\Lambda) &= \prod_{i=1}^{K} L_i(\theta,\boldsymbol{\beta},\Lambda) \\ &= \prod_{i=1}^{K} \int_0^1 \prod_{j=1}^{n_i} \mathrm{c}_{\cdot V}(H_{ij},v_i)^{\delta_{ij}} \times \mathrm{C}_{\cdot|V}(H_{ij}|v_i)^{1-\delta_{ij}} dv_i \\ &= \prod_{i=1}^{K} \int_0^1 \exp\left\{ \sum_{j=1}^{n_i} \log\left(\mathrm{c}_{\cdot V}(H_{ij},v_i)^{\delta_{ij}} \times \mathrm{C}_{\cdot|V}(H_{ij}|v_i)^{1-\delta_{ij}} \right) \right\} dv_i \\ &= \prod_{i=1}^{K} \int_0^1 \exp\left\{ \sum_{j=1}^{n_i} \log \mathbf{C}(H_{ij},v_i;\theta) \right\} dv_i, \end{aligned}$$

where $\mathbf{C}(H_{ij},v_i,\theta) = \mathrm{c}_{\cdot V}(H_{ij},v_i;\theta)^{\delta_{ij}} \times \mathrm{C}_{\cdot|V}(H_{ij}|v_i;\theta)^{1-\delta_{ij}}$.

$$\begin{aligned} l_K(\theta) &= K^{-1} \log\{L(\theta,\boldsymbol{\beta},\Lambda)\}, \\ l_{K0}(\theta) &= K^{-1} \log\{L(\theta,\boldsymbol{\beta}_0,\Lambda_0)\}, \\ \check{l}_K(\theta) &= K^{-1} \log\left\{ L\left(\theta,\check{\boldsymbol{\beta}},\check{\Lambda}\right) \right\}, \end{aligned}$$

$$\begin{aligned} U_K(\theta) = \frac{\partial l_K(\theta)}{\partial\theta} &= K^{-1} \frac{\partial \log\{L(\theta,\boldsymbol{\beta},\Lambda)\}}{\partial\theta} \\ &= K^{-1} \sum_{i=1}^{K} \frac{\displaystyle\int_0^1 \exp\left\{ \sum_{j=1}^{n_i} \log \mathbf{C}(H_{ij},v_i,\theta) \right\} \left\{ \sum_{j=1}^{n_i} \frac{\partial}{\partial\theta} \log \mathbf{C}(H_{ij},v_i,\theta) \right\} dv_i}{\displaystyle\int_0^1 \exp\left\{ \sum_{j=1}^{n_i} \log \mathbf{C}(H_{ij},v_i,\theta) \right\} dv_i}, \end{aligned}$$

$$\begin{aligned} U_{K0}(\theta) &= \frac{\partial l_{K0}(\theta)}{\partial\theta} = K^{-1} \frac{\partial \log\{L(\theta,\boldsymbol{\beta}_0,\Lambda_0)\}}{\partial\theta}, \\ \check{U}_K(\theta) &= \frac{\partial \check{l}_K(\theta)}{\partial\theta} = K^{-1} \frac{\partial \log\left\{ L\left(\theta,\check{\boldsymbol{\beta}},\check{\Lambda}\right) \right\}}{\partial\theta}. \end{aligned}$$

The following notation is copied from Spiekerman and Lin (1998). Let $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}'\mathbf{a}$ and $r = 0,1,2$:

$$\mathbf{S}^{(r)}(\boldsymbol{\beta},t) = K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i} Y_{ij}(t)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(t)\right\}\mathbf{Z}_{ij}(t)^{\otimes r}, \qquad \mathbf{s}^{(r)} = E\left[\mathbf{S}^{(r)}(\boldsymbol{\beta},t)\right],$$

$$\mathbf{E}(\boldsymbol{\beta},t) = \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)},$$

$$\mathbf{e}(\boldsymbol{\beta},t) = \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)},$$

$$\mathbf{V}(\boldsymbol{\beta},t) = \frac{\mathbf{S}^{(2)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)} - \mathbf{E}(\boldsymbol{\beta},t)^{\otimes 2},$$

$$\mathbf{v}(\boldsymbol{\beta},t) = \frac{\mathbf{s}^{(2)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)} - \mathbf{e}(\boldsymbol{\beta},t)^{\otimes 2}.$$

Assume the following regularity conditions, where $\tau > 0$ is a constant denoting the last survival time of the uncensored subjects:

Condition 1 $\boldsymbol{\beta}$ *is in a compact subset of* $\mathbb{R}^p$.

Condition 2 $\Lambda(t) < \infty$.

Condition 3 $\theta \in \nu$, *where* ν *is a compact subset of* Θ.

Condition 4 $P(C_{ij} \geq t, \forall t \in [0,\tau]) > \delta_c > 0$ *for* $i = 1,...,K$ *and* $j = 1,...,n_i$.

Condition 5 *Let* $\mathbf{Z}_{ij}(t) = \left\{Z_{ij1}(t),...,Z_{ijp}(t)\right\}$. *For* $i = 1,...,K$, $j = 1,...,n_i$ *and* $k = 1,...,p$,

$$\left|Z_{ijk}(0)\right| + \int_0^\tau \left|dZ_{ijk}(t)\right| \leq B_Z < \infty \qquad \textit{almost surely for some constant } B_z.$$

Condition 6 $E[\log\{L_i(\theta_1,\boldsymbol{\beta},\Lambda)/L_i(\theta_2,\boldsymbol{\beta},\Lambda)\}$ *exists for all* $\theta_1,\theta_2 \in \Theta$, $i = 1,...,K$.

Condition 7 $\mathbf{A} = \int_0^\tau \mathbf{v}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)d\Lambda_0(u)$ *is positive definite.*

Condition 8 *The bivariate copula* $\mathrm{C}_{\cdot V}(u_{ij},v_i;\theta)$ *is absolutely continuous.*

Proof of Theorem 2

Since the consistency of $\breve{\boldsymbol{\beta}}$ and $\breve{\Lambda}$ was already proved in Spiekerman and Lin (1998), we only show the consistency of $\breve{\theta}$. This is done by extending the results in Prenen et al. (2017a) and Othus and Li (2010).

Accounting for the fact that we use plug-in estimators for $\boldsymbol{\beta}$ and Λ, we proceed by taking a Taylor series expansion of the log-likelihood of θ in the neighbourhood of $\boldsymbol{\beta}$ and Λ. In view of Λ being an unspecified function, we need to include a functional expansion term. The concept of Hadamard differentiability is suitable in this case. In order to use this approach, we must first verify that the log-likelihood $l(\theta)$ is Hadamard differentiable with respect to Λ: By condition 5, the

term $\int_0^\tau Y_{ij}(u)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\right\}d\Lambda(u)$ in H_{ij} is Hadamard differentiable. Furthermore, by the chain rule for Hadamard derivatives (Van der Vaart, 2000), we conclude that $l(\theta)$ is Hadamard differentiable with respect to Λ.

Let $BV[0,\tau]$ denote the class of functions with bounded total variation on $[0,\tau]$. The Hadamard derivative of $l(\theta)$ with respect to Λ at $\Gamma-\Lambda\in BV[0,\tau]$ can be obtained by taking the derivative of $K^{-1}\log\left[L\left\{\theta,\boldsymbol{\beta},\Lambda+t\left(\Gamma-\Lambda\right)\right\}\right]$ with respect to t and then making $t=0$:

$$\frac{d}{dt}\left(K^{-1}\log\left[L\left\{\theta,\boldsymbol{\beta},\Lambda+t\left(\Gamma-\Lambda\right)\right\}\right]\right)\Bigg|_{t=0}=\int_0^\tau\zeta_K(\theta,\Lambda)(u)d(\Gamma-\Lambda)(u),$$

where $\zeta_K\left(\theta,\Lambda\right)(u)$ is equal to

$$\begin{aligned}
&K^{-1}\sum_{i=1}^{K}\frac{\displaystyle\int_0^1\exp\left\{\sum_{j=1}^{n_i}\log\mathbf{C}(H_{ij},v_i,\theta)\right\}\left[\sum_{j=1}^{n_i}\left\{\left(\frac{\partial}{\partial H_{ij}}\log\mathbf{C}(H_{ij},v_i,\theta)\right)D_{ij}^{\Lambda}\right\}\right]dv_i}{\displaystyle\int_0^1\exp\left\{\sum_{j=1}^{n_i}\log\mathbf{C}(H_{ij},v_i,\theta)\right\}dv_i}\\
&=K^{-1}\sum_{i=1}^{K}\int_0^1 P(v_i|H_{i\cdot},\theta)\left[\sum_{j=1}^{n_i}\left\{\left(\frac{\partial}{\partial H_{ij}}\log\mathbf{C}(H_{ij},v_i,\theta)\right)D_{ij}^{\Lambda}\right\}\right]dv_i\\
&=K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i}D_{ij}^{\Lambda}E\left[\frac{\partial}{\partial H_{ij}}\log\mathbf{C}(H_{ij},v_i,\theta)\right],
\end{aligned}$$

$$D_{ij}^{\Lambda}=(-H_{ij})Y_{ij}(u)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\right\},$$

and

$$P(v_i|H_{i\cdot},\theta)=\frac{\exp\left\{\sum_{j=1}^{n_i}\log\mathbf{C}(H_{ij},v_i,\theta)\right\}}{\displaystyle\int_0^1\exp\left\{\sum_{j=1}^{n_i}\log\mathbf{C}(H_{ij},v_i,\theta)\right\}dv_i}\tag{8}$$

is a probability density function of a random variable V_i assuming values in $[0,1]$. Similarly, the derivative of $l(\theta)$ with respect to $\boldsymbol{\beta}$ is

$$\begin{aligned}
\zeta_K(\theta,\boldsymbol{\beta})&=K^{-1}\sum_{i=1}^{K}\int_0^1 P(v_i|H_{i\cdot},\theta)\left[\sum_{j=1}^{n_i}\left\{\left(\frac{\partial}{\partial H_{ij}}\log\mathbf{C}(H_{ij},v_i,\theta)\right)D_{ij}^{\boldsymbol{\beta}}\right\}\right]dv_i\\
&=K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i}D_{ij}^{\boldsymbol{\beta}}E\left[\frac{\partial}{\partial H_{ij}}\log\mathbf{C}(H_{ij},v_i,\theta)\right],
\end{aligned}$$

where

$$D_{ij}^{\boldsymbol{\beta}}=(-H_{ij})\int_0^\tau Y_{ij}(u)\mathbf{Z}_{ij}(u)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\right\}d\Lambda(u).$$

Let $\|\cdot\|$ denote the Euclidean norm and let $\|\cdot\|_\infty$ denote the supremum norm on $[0,\tau]$. To prove the consistency of $\breve{\theta}$, we need $\|\zeta_K(\theta,\Lambda)\|_\infty$ and $\|\zeta_K(\theta,\boldsymbol{\beta})\|$ to be

bounded. Note that, by the definition of H_{ij} and conditions 2 and 5, the terms $\left\| D_{ij}^{\Lambda} \right\|_{\infty}$ and $\left\| D_{ij}^{\boldsymbol{\beta}} \right\|$ are bounded. Therefore, in order to satisfy the boundedness condition of $\|\zeta_K(\theta,\Lambda)\|_{\infty}$ and $\|\zeta_K(\theta,\boldsymbol{\beta})\|$, we shall require the expectations in their formulae to be finite.

We now continue with the proof by taking an expansion of $\check{l}_K(\theta)$ around $\boldsymbol{\beta}_0$ and Λ_0, given by

$$\check{l}_K(\theta) = l_{K0}(\theta) + \zeta_K(\theta,\boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \int_0^{\tau} \zeta_K(\theta,\Lambda_0)(t)d(\check{\Lambda} - \Lambda_0)(t) + R.$$

Another (intuitive) notation is

$$l_{K,\theta}(\check{\boldsymbol{\beta}},\check{\Lambda}) = l_{K,\theta}(\boldsymbol{\beta}_0,\Lambda_0) + \frac{\partial}{\partial \boldsymbol{\beta}} l_{K,\theta}(\boldsymbol{\beta}_0,\Lambda_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{\partial}{\partial \Lambda} l_{K,\theta}(\boldsymbol{\beta}_0,\Lambda_0)(\check{\Lambda} - \Lambda_0) + R.$$

The remainder term R is of order $o_p\left(\max\left\{\left\|\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|, \left\|\check{\Lambda} - \Lambda_0\right\|_{\infty}\right\}\right)$. This can be seen from the definition of Hadamard differentiability, since

$$\left\| \frac{l_{K,\theta}\left\{\boldsymbol{\beta},\Lambda_0 + t(\check{\Lambda} - \Lambda_0)\right\} - l_{K,\theta}(\boldsymbol{\beta},\check{\Lambda})}{t} - \frac{\partial}{\partial \Lambda} l_{K,\theta}(\boldsymbol{\beta},\Lambda_0)(\check{\Lambda} - \Lambda_0) \right\|_{\infty} \to 0, \quad \text{as } t \downarrow 0$$

uniformly in $\check{\Lambda} - \Lambda_0$ in all compact subsets of $\mathbb{D}$, the space of cumulative hazard functions. Since $\check{\boldsymbol{\beta}}$ is consistent and $\check{\Lambda}$ is uniformly consistent (Spiekerman and Lin, 1998), $R = o_p(1)$. To prove that $\check{\theta}$ is consistent we need to verify the uniform convergence of the log-likelihood with the plug-in estimate of Λ to the expected value of the log-likelihood evaluated at the true value of Λ, denoted $l_{K0}(\theta)$:

$$\sup_{\theta \in \nu} \left| \check{l}_K(\theta) - E[l_{K0}(\theta)] \right| = o_P(1). \tag{9}$$

This can be shown as follows:

$$\begin{aligned} \check{l}_K(\theta) - E[l_{K0}(\theta)] = {} & l_{K0}(\theta) - E[l_{K0}(\theta)] + \zeta_K(\theta,\boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ & + \int_0^{\tau} \zeta_K(\theta,\Lambda_0)(t)d(\check{\Lambda} - \Lambda_0)(t) + R. \end{aligned}$$

By the law of large numbers, for fixed θ

$$l_{K0}(\theta) - E[l_{K0}(\theta)] \xrightarrow{p} 0. \tag{10}$$

Since $\|\zeta_K(\theta,\boldsymbol{\beta})\|$ and $\|\zeta_K(\theta,\Lambda)(u)\|_{\infty}$ are bounded, say $\|\zeta_K(\theta,\boldsymbol{\beta})\| \leq M_1$ and $\|\zeta_K(\theta,\Lambda)(u)\|_{\infty} \leq M_2$, we have

$$\begin{aligned} \sup_{\theta \in \nu} \left| \zeta_K(\theta,\boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \right| &\leq M_1 \left\| \check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|, \\ \sup_{\theta \in \nu} \left| \int_0^{\tau} \zeta_K(\theta,\Lambda_0)(t)d(\check{\Lambda} - \Lambda_0)(t) \right| &\leq M_2 \left\| \check{\Lambda} - \Lambda_0 \right\|_{\infty}. \end{aligned}$$

For this reason,

$$\sup_{\theta\in\nu}\left|\check{l}_K(\theta)-E[l_{K0}(\theta)]\right| \leq \sup_{\theta\in\nu}|l_{K0}(\theta)-E[l_{K0}(\theta)]|+M_1\left\|\check{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\| + M_2\left\|\check{\Lambda}-\Lambda_0\right\|_\infty + R.$$

Using result (10), the consistency of $\check{\boldsymbol{\beta}}$ and the uniform consistency of $\check{\Lambda}$ and the fact that $R = o_p(1)$, we obtain

$$\sup_{\theta\in\nu}\left|\check{l}_K(\theta)-E[l_{K0}(\theta)]\right| = o_p(1).$$

Finally, to verify that $\check{\theta}$ is consistent, we need to show that the expected log-likelihood is maximized at the true value:

$$E[l_{K0}(\theta)]-E[l_{K0}(\theta_0)]<0. \tag{11}$$

Since clusters are independent, the log-likelihood $l_K(\theta)$ can be written as a sum of independent and identically distributed random variables:

$$K^{-1}\sum_{i=1}^{K}\log\{L_i(\theta,\boldsymbol{\beta},\Lambda)\},$$

with

$$\begin{aligned}L_i &= (-1)^{d_i}\frac{\partial^{d_i}}{(\partial x_{i1})^{\delta_{i1}}\ldots(\partial x_{in_i})^{\delta_{in_i}}}S(x_{i1},\ldots,x_{in_i})\\ &= \prod_{i=1}^{K}\int_0^1\exp\left\{\sum_{j=1}^{n_i}\log\left(\mathrm{c}_{\cdot V}(\exp\{-\Lambda(x_{ij})\},v_i)^{\delta_{ij}}\right.\right.\\ &\quad\left.\left.\times\mathrm{C}_{\cdot|V}(\exp\{-\Lambda(x_{ij})\}|v_i)^{1-\delta_{ij}}\right)\right\}dv_i.\end{aligned}$$

Take $\theta\neq\theta_0$. The law of large numbers, Jensen's inequality and condition 6 imply that

$$\begin{aligned}\lim_{K\to\infty} l_{K0}(\theta)-l_{K0}(\theta_0) &= E[l_{K0}(\theta)]-E[l_{K0}(\theta_0)]\\ &= E\left[K^{-1}\sum_{i=1}^{K}\log\{L_i(\theta,\boldsymbol{\beta}_0,\Lambda_0)\}\right]-E\left[K^{-1}\sum_{i=1}^{K}\log\{L_i(\theta_0,\boldsymbol{\beta}_0,\Lambda_0)\}\right]\\ &= E\left[\log\{L_i(\theta,\boldsymbol{\beta}_0,\Lambda_0)\}-\log\{L_i(\theta_0,\boldsymbol{\beta}_0,\Lambda_0)\}\right]\\ &= E\left[\log\{L_i(\theta,\boldsymbol{\beta}_0,\Lambda_0)/L_i(\theta_0,\boldsymbol{\beta}_0,\Lambda_0)\}\right]\\ &\leq \log\left(E\left[\log\{L_i(\theta,\boldsymbol{\beta}_0,\Lambda_0)/L_i(\theta_0,\boldsymbol{\beta}_0,\Lambda_0)\}\right]\right)\\ &\leq \log\left(E\left[\log\{L_i(\theta,\boldsymbol{\beta}_0,\Lambda_0)/L_i(\theta_0,\boldsymbol{\beta}_0,\Lambda_0)\}\right]\right)\\ &= \log(1)=0.\end{aligned}$$

Since $\check{\theta}$ maximizes $\check{l}(\theta)$, equation (9) implies that

$$0 \leq \check{l}_K\left(\check{\theta}\right) - \check{l}_K\left(\theta_0\right) + E\left[l_{K0}\left(\theta_0\right)\right] - E\left[l_{K0}\left(\theta_0\right)\right] = \check{l}_K\left(\check{\theta}\right) - E\left[l_{K0}\left(\theta_0\right)\right] + o_p(1)$$
$$\Longrightarrow E\left[l_{K0}\left(\theta_0\right)\right] \leq \check{l}_K\left(\check{\theta}\right) + o_p(1).$$

Subtracting $E\left[l_{K0}\left(\check{\theta}\right)\right]$ from each side of the inequality we get

$$\begin{aligned} &E\left[l_{K0}\left(\theta_0\right)\right] - E\left[l_{K0}\left(\check{\theta}\right)\right] \leq \check{l}_K\left(\check{\theta}\right) - E\left[l_{K0}\left(\check{\theta}\right)\right] + o_p(1) \\ &\leq \sup_{\theta\in\Theta}\left|\check{l}_K\left(\theta\right) - E\left[l_{K0}\left(\theta\right)\right]\right| + o_p(1) = o_p(1). \end{aligned} \tag{12}$$

Now take θ such that $|\theta - \theta_0| \geq \varepsilon$ for any fixed $\varepsilon > 0$. By inequality (11), there must be some $\gamma_\varepsilon > 0$ such that

$$E\left[l_{K0}\left(\check{\theta}\right)\right] + \gamma_\varepsilon < E\left[l_{K0}\left(\theta_0\right)\right].$$

It follows that

$$P(|\check{\theta} - \theta_0| \geq \varepsilon) \leq P\left\{E\left[l_{K0}\left(\check{\theta}\right)\right] + \gamma_\varepsilon < E\left[l_{K0}\left(\theta_0\right)\right]\right\}.$$

Equation (12) implies that

$$P\left\{E\left[l_{K0}\left(\check{\theta}\right)\right] + \gamma_\varepsilon < E\left[l_{K0}\left(\theta_0\right)\right]\right\} \to 0 \qquad \text{as } K \to \infty.$$

Therefore

$$P(|\check{\theta} - \theta_0| \geq \varepsilon) \to 0 \qquad \text{as } K \to \infty.$$

Proof of Theorem 3

Take a first order Taylor series expansion of $\check{U}_K(\check{\theta})$ around θ_0:

$$\check{U}_K(\check{\theta}) = \check{U}_K(\theta_0) + \left(\check{\theta} - \theta_0\right)\left.\frac{\partial \check{U}_K}{\partial \theta}\right|_{\theta=\theta^*},$$

where θ^* is between $\check{\theta}$ and θ_0. It must be that $\check{U}_K(\check{\theta}) = 0$ since $\check{\theta}$ was taken to be the maximum of $L(\theta, \check{\boldsymbol{\beta}}, \check{\Lambda})$. For this reason

$$\sqrt{K}\left(\check{\theta} - \theta_0\right) = \frac{\sqrt{K}\check{U}_K\left(\theta_0\right)}{-\partial \check{U}_K/\partial\theta\big|_{\theta=\theta^*}}. \tag{13}$$

We already showed that $\check{\theta}$ consistently estimates θ_0, so the law of large numbers implies that

$$\left.\frac{\partial \check{U}_K}{\partial \theta}\right|_{\theta=\theta^*} \to W\left(\theta_0\right) = \lim_{K\to\infty}\left.\frac{\partial \check{U}_K}{\partial \theta}\right|_{\theta=\theta_0}.$$

We shall show that the score equation $\check{U}_K(\theta_0)$ in the numerator of equation (13) follows a normal distribution. Hereto, we need a Taylor series expansion of $\check{U}_K(\theta_0)$ in

the neighbourhood of $\boldsymbol{\beta}_0$ and Λ_0. Because Λ_0 is an unspecified function, we shall use the Hadamard derivative of $U_K(\theta_0)$ with respect to Λ at $\Gamma - \Lambda \in BV[0,\tau]$:

$$\frac{d}{dt}\left(K^{-1}\frac{\partial \log\left[L\{\theta,\boldsymbol{\beta},\Lambda + t(\Gamma-\Lambda)\}\right]}{\partial\theta}\right)\Bigg|_{t=0} = \int_0^\tau \xi_K(\theta,\Lambda)(u)d(\Gamma-\Lambda)(u),$$

where $\xi_K(\theta,\Lambda)(u)$ is equal to

$$\begin{aligned}
&K^{-1}\sum_{i=1}^{K}\left[\int_0^1 P(v_i|H_{i\cdot},\theta)\left\{\sum_{j=1}^{n_i}\frac{\partial^2 \log\mathbf{C}(H_{ij},v_i,\theta)}{\partial\theta\partial H_{ij}}D_{ij}^{\Lambda} + \sum_{j=1}^{n_i}\frac{\partial \log\mathbf{C}(H_{ij},v_i,\theta)}{\partial\theta}\right.\right.\\
&\quad\left.\times\sum_{j=1}^{n_i}\frac{\partial \log\mathbf{C}(H_{ij},v_i,\theta)}{\partial H_{ij}}D_{ij}^{\Lambda}\right\}dv_i - \int_0^1 P(v_i|H_{i\cdot},\theta)\left\{\sum_{j=1}^{n_i}\frac{\partial \log\mathbf{C}(H_{ij},v_i,\theta)}{\partial\theta}\right\}dv_i\\
&\quad\left.\times\int_0^1 P(v_i|H_{i\cdot},\theta)\left\{\sum_{j=1}^{n_i}\frac{\partial \log\mathbf{C}(H_{ij},v_i,\theta)}{\partial H_{ij}}D_{ij}^{\Lambda}\right\}dv_i\right]\\
&= K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i}D_{ij}^{\Lambda}\left\{E\left[\frac{\partial^2 \log\mathbf{C}(H_{ij},v_i,\theta)}{\partial\theta\partial H_{ij}}\right]\right.\\
&\quad\left.+\sum_{k=1}^{n_i}\mathrm{Cov}\left[\frac{\partial \log\mathbf{C}(H_{ij},v_i,\theta)}{\partial\theta},\frac{\partial \log\mathbf{C}(H_{ik},v_i,\theta)}{\partial H_{ik}}\right]\right\},
\end{aligned}$$

$$D_{ij}^{\Lambda} = (-H_{ij})Y_{ij}(u)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\right\},$$

and $P(v_i|H_{i\cdot},\theta)$ has the same definition as in expression (8). The derivative of $U_K(\theta_0)$ with respect to $\boldsymbol{\beta}$ is given by the same expression as $\xi_K(\theta,\Lambda)(u)$, replacing D_{ij}^{Λ} for

$$D_{ij}^{\boldsymbol{\beta}} = (-H_{ij})\int_0^\tau Y_{ij}(u)\mathbf{Z}_{ij}(u)\exp\left\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\right\}d\Lambda(u).$$

By the same arguments used to show the consistency of $\check{\theta}$, we also need $\|\xi_K(\theta,\Lambda)\|_\infty$ and $\|\xi_K(\theta,\boldsymbol{\beta})\|$ to be bounded. For this reason, we shall require the expectation and covariance in their formulae to be finite. Hence, we proceed by taking a Taylor series expansion of $\check{U}_K(\theta_0)$ in the neighbourhood of $\boldsymbol{\beta}_0$ and Λ_0 which gives

$$\check{U}_K(\theta_0) = U_{K0}(\theta_0) + \xi_K(\theta_0,\boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \int_0^\tau \xi_K(\theta_0,\Lambda_0)(t)d\left\{\check{\Lambda}(t)-\Lambda_0(t)\right\} + G_K,$$

where G_K is the remainder term for the Taylor series. Since $\check{\Lambda}$ is $\sqrt{K}$ consistent, it can be shown that $G_K = o_p(K^{-1/2})$. Define the pointwise limit of $\xi_K(\theta,\Lambda)(t)$ as $\xi(\theta,\Lambda)(t)$ and denote $\xi(\theta,\boldsymbol{\beta}) = E[\xi_K(\theta,\boldsymbol{\beta})]$. Since $\|\xi_K(\theta,\Lambda)\|_\infty$ and $\|\xi_K(\theta,\boldsymbol{\beta})\|$ are bounded, $\|\xi(\theta,\Lambda)\|_\infty$ and $\|\xi(\theta,\boldsymbol{\beta})\|$ are also. Therefore

$$\begin{aligned}
\sqrt{K}\check{U}_K(\theta_0) = \sqrt{K}\Big[&U_{K0}(\theta_0) + \xi_K(\theta_0,\boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&+\int_0^\tau \xi_K(\theta_0,\Lambda_0)(t)d\left\{\check{\Lambda}(t)-\Lambda_0(t)\right\}\Big] + o_p(1).
\end{aligned}$$

By Spiekerman and Lin (1998)

$$\sqrt{K}(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to \mathbf{A}^{-1} \sum_{i=1}^{K} \mathbf{w}_{i.},$$

where $\mathbf{w}_{i.}$ is the ith component of the score function for $\boldsymbol{\beta}$ under the independence working assumption, evaluated at $\boldsymbol{\beta}_0$:

$$\mathbf{w}_{i.} = \sum_{j=1}^{n_i} \int_0^{\tau} \left\{ \mathbf{Z}_{ij}(u) - E\left[\boldsymbol{\beta}_0, u\right] \right\} dM_{ij}(u),$$

with

$$M_{ij}(t) = \delta_{ij} Y_{ij}(t) - \int_0^t Y_{ij}(u) \exp\left\{ \boldsymbol{\beta}'_0 \mathbf{Z}_{ij}(u) \right\} d\Lambda_0(u).$$

They also showed that

$$\sqrt{K} \left\{ \check{\Lambda}_0(t, \check{\boldsymbol{\beta}}) - \Lambda_0(t) \right\} \to \mathcal{W}(t) = K^{-1/2} \sum_{i=1}^{K} \Psi_i(t),$$

where $\mathcal{W}(t)$ is a zero mean Gaussian process with variance function

$$E\left[\Psi_i(t)^2 \right],$$

with

$$\Psi_i(t) = \int_0^t \frac{dM_{i.}(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} + \mathbf{h}^T(t) \mathbf{A}^{-1} \mathbf{w}_{i.}$$

and

$$\mathbf{h}(t) = -\int_0^t \mathbf{e}(\boldsymbol{\beta}_0, u)\, d\Lambda_0(u).$$

That is why

$$\begin{aligned}
&\sqrt{K} \left[U_{K0}(\theta_0) + \xi_K(\theta_0, \boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \int_0^{\tau} \xi_K(\theta_0, \Lambda_0)(t) d\left\{ \check{\Lambda}(t) - \Lambda_0(t) \right\} \right] \\
&= \sqrt{K} \left[K^{-1} \sum_{i=1}^{K} \phi_i(\theta_0) + \xi_K(\theta_0, \boldsymbol{\beta}_0) K^{-1/2} \mathbf{A}^{-1} \sum_{i=1}^{K} \mathbf{w}_{i.} \right. \\
&\quad \left. + K^{-1/2} \int_0^{\tau} \xi_K(\theta_0, \Lambda_0)(t) d\left\{ K^{-1/2} \sum_{i=1}^{K} \Psi_i(t) \right\} \right] \\
&= \sqrt{K} \sum_{i=1}^{K} \left[K^{-1} \phi_i(\theta_0) + \xi_K(\theta_0, \boldsymbol{\beta}_0) K^{-1/2} \mathbf{A}^{-1} \mathbf{w}_{i.} + K^{-1} \int_0^{\tau} \xi_K(\theta_0, \Lambda_0)(t) d\Psi_i(t) \right] \\
&= K^{-1/2} \sum_{i=1}^{K} \left[\phi_i(\theta_0) + \xi_K(\theta_0, \boldsymbol{\beta}_0) \sqrt{K} \mathbf{A}^{-1} \mathbf{w}_{i.} + \int_0^{\tau} \xi_K(\theta_0, \Lambda_0)(t) d\Psi_i(t) \right] \\
&= K^{-1/2} \sum_{i=1}^{K} \Xi_i
\end{aligned}$$

The central limit theorem implies that $\sqrt{K}\check{U}_K(\theta_0)$ converges to a normally distributed random variable with mean 0 and variance equal to the variance of Ξ. Thus we have

$$\sqrt{K}\left(\check{\theta}-\theta_0\right)=\frac{\sqrt{K}\check{U}_K(\theta_0)}{-\partial\check{U}_K/\partial\theta\big|_{\theta=\theta^*}},$$

where

$$\sqrt{K}\check{U}_K(\theta_0)\xrightarrow{D}N\left\{0,\mathrm{var}\left(\Xi\right)\right\}$$

and

$$\partial\check{U}_K/\partial\theta\big|_{\theta=\theta^*}\xrightarrow{P}W\left(\theta_0\right).$$

By Slutsky's theorem, $\sqrt{K}\left(\check{\theta}-\theta_0\right)$ converges to a normal distribution with mean 0 and variance

$$\mathrm{var}\left(\Xi\right)/W(\theta_0)^2.$$

The variance of Ξ (note that $\mathrm{var}[\Xi]=E[\Xi^2]$) can be estimated by $K^{-1}\sum_{i=1}^{K}\check{\Xi}^2$, where $\check{\Xi}$ is obtained from Ξ replacing parameter values by their estimates. $W(\theta_0)$ can be estimated by the (minus) derivative of the pseudoscore function $\check{U}_K(\theta)$ evaluated at $\check{\theta}$.

References

Andersen EW (2005) Two-stage estimation in copula models used in family studies. Lifetime Data Analysis 11(3):333–350

Cox DR (1972) Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) 34(2):187–202

Cox DR, Hinkley D (1974) Theoretical Statistics. Chapman and Hall, London

Duchateau L, Janssen P (2004) Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. Biometrics 60(3):608–614

Duchateau L, Janssen P (2008) The Frailty Model. Springer, Berlin

Glidden DV (2000) A two-stage estimator of the dependence parameter for the clayton-oakes model. Lifetime Data Analysis 6(2):141–156

Goethals K, Janssen P, Duchateau L (2008) Frailty models and copulas: similarities and differences. Journal of Applied Statistics 35(9):1071–1079

Henderson HV, Searle SR (1981) On deriving the inverse of a sum of matrices. Siam Review 23(1):53–60

Joe H (2014) Dependence Modeling with Copulas. Chapman & Hall/CRC

Krupskii P, Joe H (2013) Factor copula models for multivariate data. Journal of Multivariate Analysis 120:85–101

Lehmann EL, Casella G (1998) Theory of Point Estimation. Springer

Massonnet G, Janssen P, Duchateau L (2009) Modelling udder infection data using copula models for quadruples. Journal of Statistical Planning and Inference 139(11):3865–3877

Othus M, Li Y (2010) A gaussian copula model for multivariate survival data. Statistics in Biosciences 2(2):154–179

Prenen L, Braekers R, Duchateau L (2017a) Extending the archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79(2):483–505

Prenen L, Braekers R, Duchateau L, De Troyer E (2017b) Sunclarco: Survival Analysis using Copulas. R package version 1.0.0

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Romeo JS, Meyer R, Gallardo DI (2018) Bayesian bivariate survival analysis using the power variance function copula. Lifetime data analysis 24(2):355–383

Shih JH, Louis TA (1995) Inferences on the association parameter in copula models for bivariate survival data. Biometrics 51:1384–1399

Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. Publ Inst Statist Univ Par 8:229–231

Spiekerman CF, Lin D (1998) Marginal regression models for multivariate failure time data. Journal of the American Statistical Association 93(443):1164–1175

Van der Vaart AW (2000) Asymptotic statistics, vol 3. Cambridge university press

Wienke A (2011) Frailty models in survival analysis. Boca Raton: Chapman and Hall

Xu JJ (1996) Statistical modelling and inference for multivariate and longitudinal discrete response data. PhD thesis, University of British Columbia

APPENDIX A – Additional topics and proofs

Some details on the properties of the model

If compared to other models for clustered survival data, the proposed factor copula based model has the advantage of allowing flexible dependence structures. This is due to the many options of bivariate copulas that can be used in its formulation. These are not constrained to a specific family or class, as opposed to the models of Prenen et al. (2017), Romeo et al. (2017) and Othus and Li (2010). This feature of factor copulas makes possible to achieve a wide range of dependence structures (see Section 3 of Krupskii and Joe (2013)). The Archimedean copula model of Prenen et al. (2017), for example, has the limitation of only supporting a few positive dependence structures. This is due to a very limited number of existing generators from the class $\varphi \in \mathscr{L}_\infty$, used to yield the bivariate Archimedean copulas in their model.

As we show next, our factor copula based model can be seen as a generalisation of the Archimedean copula model of Prenen et al. (2017).

Proposition. *Let $\varphi_\theta \in \mathscr{L}_\infty$ be a generator function from the class of Laplace transforms of non-negative random variables with no mass at 0 (see Chapter 4 of Nelsen (2007)). Then, there exists $C_{\cdot|V}(u_j|v)$, such that for all $u_j \in [0,1]$ $(j = 1,...,n)$,*

$$\int_0^1 \prod_{j=1}^{n} C_{\cdot|V}(u_j|v)\,dv = \int_0^{+\infty} \prod_{j=1}^{n} \exp\left\{-x\varphi_\theta^{-1}(u_j)\right\} dG_\theta(x). \tag{1}$$

That is, the model of Prenen et al. (2017) (right-hand side of (1)) is a subclass of the herein proposed model (left-hand side of (1)).

Proof. Assuming that $G_\theta(x)$ is differentiable, such that $g_\theta(x) = \frac{d}{dx}G_\theta(x)$, we can make $x = G_\theta^{-1}(v)$ with $dx = \frac{dv}{g_\theta(G_\theta^{-1}(v))} = \frac{dv}{g_\theta(x)}$. This way, we can rewrite equation (1) as

$$\int_0^1 \prod_{j=1}^{n} C_{\cdot|V}(u_j|v)\,dv = \int_0^1 \prod_{j=1}^{n} \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\} dv, \tag{2}$$

where $\varphi_\theta : [0,+\infty) \to [0,1]$, the generator of an Archimedean copula, is a continuous strictly decreasing function with $\varphi_\theta(0) = 1$, $\varphi_\theta(+\infty) = 0$ and inverse φ_θ^{-1}. The function φ_θ is also the Laplace transform of a distribution function $G_\theta(x)$ with inverse $G_\theta^{-1}(x)$ and $G_\theta(0) = 0$.

By subtracting $\int_0^1 \prod_{j=1}^{n} \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\} dv$ from both sides of Equation (2), we get

$$\int_0^1 \prod_{j=1}^{n} C_{\cdot|V}(u_j|v)\,dv - \int_0^1 \prod_{j=1}^{n} \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\} dv = 0$$
$$\iff \int_0^1 \left(\prod_{j=1}^{n} C_{\cdot|V}(u_j|v) - \prod_{j=1}^{n} \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\} \right) dv = 0.$$

Hence, for the above condition to hold, it is sufficient that

$$\prod_{j=1}^{n} C_{\cdot|V}(u_j|v) - \prod_{j=1}^{n} \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\} = 0,$$

or, equivalently,

$$C_{\cdot|V}(u_j|v) = \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\},$$

which is the conditional distribution derived from

$$\begin{aligned} C_{\cdot V}(u_j, v) &= \int_0^v \exp\left\{-G_\theta^{-1}(t)\varphi_\theta^{-1}(u_j)\right\} dt \\ &= \int_0^{G_\theta^{-1}(v)} \exp\left\{-s\varphi_\theta^{-1}(u_j)\right\} dG_\theta(s), \end{aligned}$$

a bivariate function with the following properties:

1) $C_{\cdot V}(u_j, v)$ is grounded

$$C_{\cdot V}(0, v) = \int_0^v \exp\left\{-G_\theta^{-1}(t)\varphi_\theta^{-1}(0)\right\} dt = \int_0^v 0\,dt = 0$$
$$C_{\cdot V}(u_j, 0) = \int_0^0 \exp\left\{-G_\theta^{-1}(t)\varphi_\theta^{-1}(u_j)\right\} dt = 0.$$

2) $C_{\cdot V}(u_j, v)$ has margins u_j and v

$$\begin{aligned} C_{\cdot V}(1, v) &= \int_0^v \exp\left\{-G_\theta^{-1}(t)\varphi_\theta^{-1}(1)\right\} dt = \int_0^v dt = v \\ C_{\cdot V}(u_j, 1) &= \int_0^{G_\theta^{-1}(1)} \exp\left\{-s\varphi_\theta^{-1}(u_j)\right\} dG_\theta(s) = \int_0^{+\infty} \exp\left\{-s\varphi_\theta^{-1}(u_j)\right\} dG_\theta(s) \\ &= \varphi_\theta\left(\varphi_\theta^{-1}(u_j)\right) = u_j. \end{aligned}$$

3) $C_{\cdot V}(u_j, v)$ is 2-increasing, i.e., $\forall u_j, u_j^*, v, v^* \in [0,1]$ with $u_j \leq u_j^*$, $v \leq v^*$, it follows that $C_{\cdot V}(u_j^*, v^*) - C_{\cdot V}(u_j, v^*) - C_{\cdot V}(u_j^*, v) + C_{\cdot V}(u_j, v) \geqslant 0$

$$\begin{aligned}
&\int_0^{v^*} \left[\exp\left\{-G_\theta^{-1}(t)\varphi_\theta(u_j^*)\right\} - \exp\left\{-G_\theta^{-1}(t)\varphi_\theta(u_j)\right\}\right] dt \\
&- \int_0^{v} \left[\exp\left\{-G_\theta^{-1}(t)\varphi_\theta(u_j^*)\right\} - \exp\left\{-G_\theta^{-1}(t)\varphi_\theta(u_j)\right\}\right] dt \geqslant 0 \\
&\iff \int_v^{v^*} \left[\exp\left\{-G_\theta^{-1}(t)\varphi_\theta(u_j^*)\right\} - \exp\left\{-G_\theta^{-1}(t)\varphi_\theta(u_j)\right\}\right] dt \geqslant 0.
\end{aligned}$$

Therefore, $C_{\cdot V}(u_j, v)$ is a bivariate copula. □

The Kendall's τ in a one-factor copula model

It is important to note that, while dependence is explicitly determined between two random variables in a classic bivariate copula, within our one-factor copula model framework, intracluster relationships are shaped implicitly through a latent variable V (the common factor) in an exchangeable fashion. Therefore, in order to compute the Kendall's tau for any given pair of individuals (free of V) inside a cluster, the following must be done: let U_1 and U_2 denote the lifetimes of two individuals belonging to the same cluster, such that they are conditionally independent given a latent variable V. Then, without loss of generality, we have that the factor copula of (U_1, U_2) for a cluster with size two is given by

$$C_{U_1,U_2}(u_1, u_2; \theta) = \int_0^1 C_{\cdot|V}(u_1|v;\theta) C_{\cdot|V}(u_2|v;\theta) dv. \tag{3}$$

Following a well known result (see page 164 of Nelsen (2007)), we can write the Kendall's tau of U_1 and U_2 as

$$\tau = 1 - 4\int_0^1 \int_0^1 C_{2|1}(u_2|u_1) C_{1|2}(u_1|u_2) du_1 du_2.$$

Hence, we must obtain $C_{2|1}(u_2|u_1)$ and $C_{1|2}(u_1|u_2)$ from (3):

$$\begin{aligned}
C_{2|1}(u_2|u_1) &= \frac{\partial}{\partial u_1} \int_0^1 C_{\cdot|V}(u_1|v) C_{\cdot|V}(u_2|v) dv \\
&= \int_0^1 c_{\cdot V}(u_1, v) C_{\cdot|V}(u_2|v) dv.
\end{aligned}$$

Similarly,

$$C_{1|2}(u_1|u_2) = \int_0^1 C_{\cdot|V}(u_1|v) c_{\cdot V}(u_2, v) dv.$$

Therefore, the Kendall's tau for the pair (U_1, U_2) in a one-factor copula model is given by

$$\tau_{1,2} = 1 - 4\int_0^1 \int_0^1 \left(\int_0^1 c_{\cdot V}(u_1, v) C_{\cdot|V}(u_2|v) dv \right) \left(\int_0^1 C_{\cdot|V}(u_1|v) c_{\cdot V}(u_2, v) dv \right) du_1 du_2. \quad (4)$$

This expression cannot be evaluated analytically, but it can be easily computed with numerical integral methods for any given expression of $C_{\cdot|V}$ together with the value of its parameter θ.

From the perspective of our model, the Kendall's tau as computed by (4) can be regarded as an exchangeable measure for the intracluster associations, that is, every subject in a cluster is equally affected by the common factor V, so they all share the same Kendall's tau with respect to V ($\tau_{U_j,V}$) and, as a consequence, the same measure of association between each other, i.e., τ_{U_j,U_k} is the same for every $j,k = 1,...,n$ with $j \neq k$ and n being the cluster size.

We now give three examples to illustrate the flexible nature of the factor copula model.

Example 1. In a factor copula model, when $C_{j|V_1}$ comes from a Gaussian copula with parameter $\theta \in [-1,1]$ for all $j = 1,2,...,d$, then the resulting factor copula model is of the form

$$C(u_1,...,u_d) = \int_0^1 \prod_{j=1}^{d} \Phi\left(\frac{\Phi^{-1}(u_j) - \theta\Phi^{-1}(v)}{(1-\theta^2)^{1/2}} \right) dv.$$

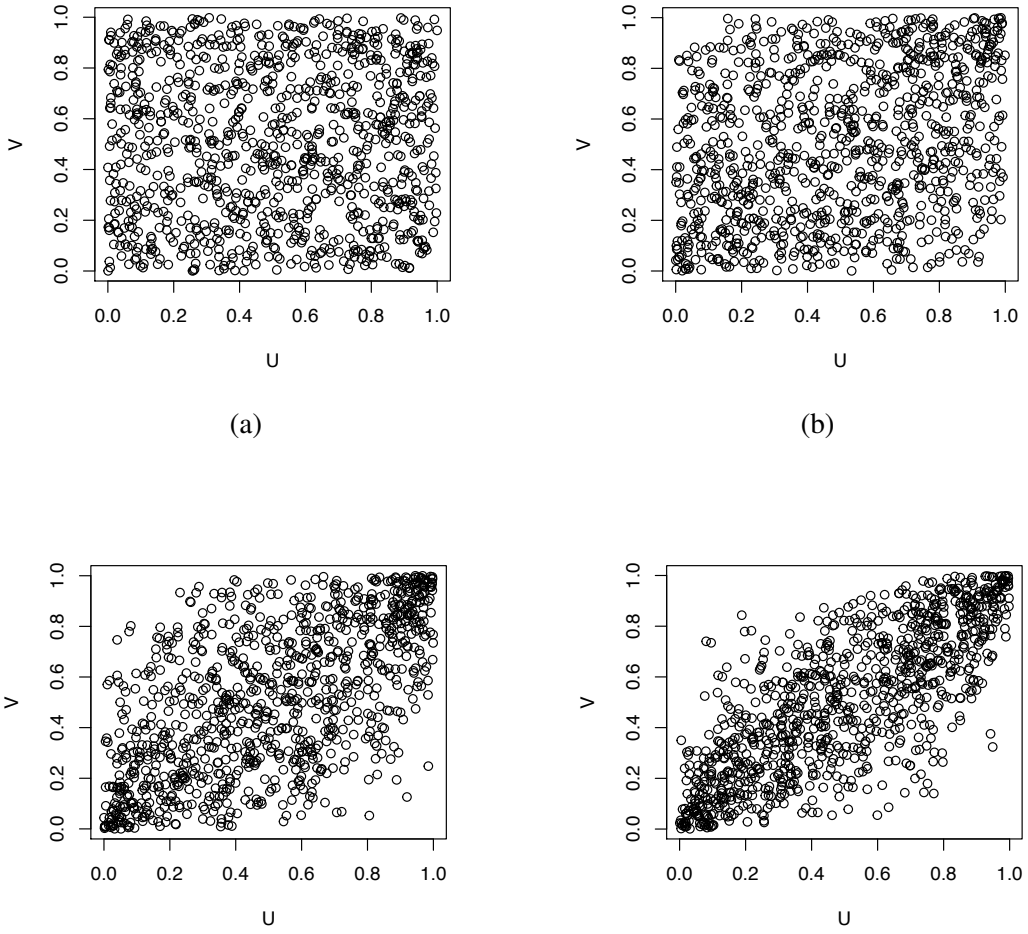
Under a right censoring scheme, the likelihood expression for the Gaussian factor copula model is

$$L = \prod_{i=1}^{K} \int_0^1 \prod_{j=1}^{n_i} \left[\frac{\phi_2(\Phi^{-1}(u_{ij}), \Phi^{-1}(v_i); \theta) f(x_{ij}|\mathbf{Z}_{ij})}{\phi(\Phi^{-1}(u_{ij}))\phi(\Phi^{-1}(v_i))} \right]^{\delta_{ij}} \times \left[\Phi\left(\frac{\Phi^{-1}(u_{ij}) - \theta\Phi^{-1}(v_i)}{(1-\theta^2)^{1/2}} \right) \right]^{1-\delta_{ij}} dv_i,$$

where $u_{ij} = S(x_{ij}|\mathbf{Z}_{ij})$.

The factor copula model built with Gaussian bivariate copulas has the properties of reflection symmetry and weak/intermediate tail dependence, as can be seen in Figure 1 for the bivariate case.

Figure 1 – Scatterplots from samples taken from a Gaussian Factor copula with $\theta = 0$ (Independence (a)), 0.556 ($\tau = 0.2$ (b)), 0.767 ($\tau = 0.4$ (c)) and 0.899 ($\tau = 0.6$ (d)).



(a) (b)

(c) (d)

Example 2. When all bivariate copulas in a factor copula model are Clayton with parameter $\theta \in [0, +\infty)$, the resulting factor copula model is of the form

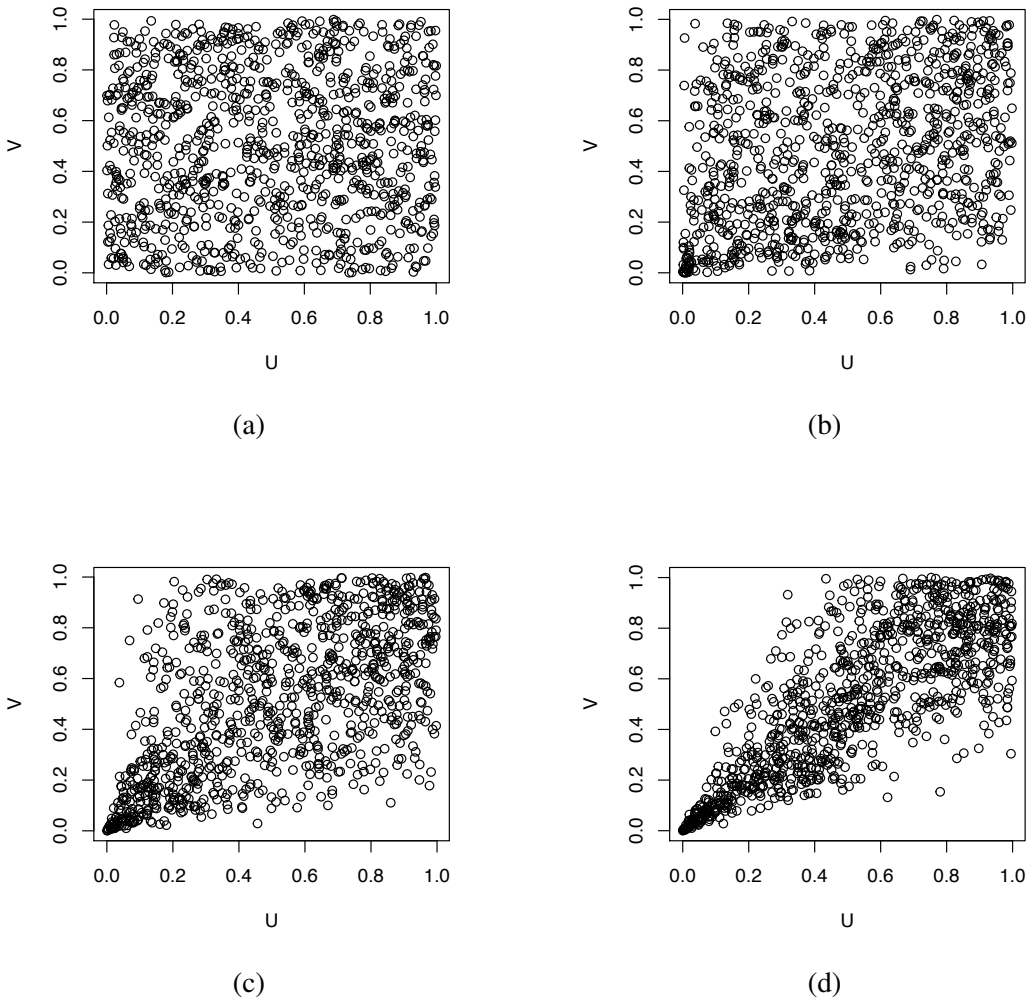
$$C(u_1, ..., u_d) = \int_0^1 \prod_{j=1}^{d} \left[1 + v^{\theta}(u_j^{-\theta} - 1)\right]^{-1-1/\theta} dv.$$

Under a right censoring scheme, the likelihood expression is given by

$$L = \prod_{i=1}^{K} \int_0^1 \prod_{j=1}^{n_i} \left[(1+\theta)(u_{ij}v_i)^{-\theta-1}(u_{ij}^{-\theta} + v_i^{-\theta} - 1)^{-2-1/\theta} f(x_{ij}|\mathbf{Z}_{ij})\right]^{\delta_{ij}}$$
$$\times \left[\left(1 + v^{\theta}(u^{-\theta} - 1)\right)^{-2-1/\theta}\right]^{1-\delta_{ij}} dv_i.$$

The Clayton factor copula model has strong lower tail dependence and weak upper tail dependence, as can be seen in Figure 2 for the bivariate case.

Figure 2 – Scatterplots from samples taken from a Clayton Factor copula with $\theta = 0$ (Independence (a)), 1.07 ($\tau = 0.2$ (b)), 2.383 ($\tau = 0.4$ (c)) and 4.816 ($\tau = 0.6$ (d)).



Example 3. When using bivariate Galambos copulas ($\theta \in [0, +\infty)$) as building blocks for a factor copula model, the resulting distribution is of the form

$$C(u_1, ..., u_d) = \int_0^1 \prod_{j=1}^{d} \left\{ u - u\left[1 + (\log v / \log u_j)^{\theta}\right]^{-1-1/\theta} \right\} \times \exp\left\{ \left[(-\log u_j)^{-\theta} + (-\log v)^{-\theta} \right]^{-1/\theta} \right\} dv.$$

Under a right censoring scheme, the formula for the likelihood is as follows

$$L=\prod_{i=1}^{K}\int_0^1\prod_{j=1}^{n_i}\left[A(u_{ij},v_i)f(x_{ij}|\mathbf{Z}_{ij})\right]^{\delta_{ij}}\times\left[B(u_{ij},v_i)\right]^{1-\delta_{ij}}dv_i,$$

where

$$\begin{aligned}A(s_{ij},t_i)=&\left[1-\left[s_{ij}^{-\theta}+t_i^{-\theta}\right]^{-1-1/\theta}\times\left[s_{ij}^{-\theta-1}+t_i^{-\theta-1}\right]+\left[s_{ij}^{-\theta}+t_i^{-\theta}\right]^{-2-1/\theta}\right.\\&\left.\times(st)^{-2-1/\theta}\left\{1+\theta+\left[s_{ij}^{-\theta}+t_i^{-\theta}\right]^{-1/\theta}\right\}\right]\exp\left\{\left[s_{ij}^{-\theta}+t_i^{-\theta}\right]^{-1/\theta}\right\},\end{aligned}$$

$$\begin{aligned}B(u_{ij},v_i)=&\left\{u_{ij}-u_{ij}\left[1+(\log v_i/\log u_{ji})^{\theta}\right]^{-1-1/\theta}\right\}\\&\times\exp\left\{\left[(-\log u_{ij})^{-\theta}+(-\log v_i)^{-\theta}\right]^{-1/\theta}\right\},\end{aligned}$$

and $s_{ij}=-\log S(x_{ij}|\mathbf{Z}_{ij})$, $t_i=-\log v_i$.

Owing to its extreme-value feature, the Galambos factor copula model has a positive dependence structure with strong upper tail dependence, as can be seen in Figure 3.

Regularity conditions for the two-stage parametric procedure and proof of Theorem 1 (Paper)

The regularity conditions for the two-stage parametric procedure are adapted from Cox and Hinkley (1974), Lehmann and Casella (1998) and Xu (1996), and are as follows:

Condition 1. *The parameter space* $\boldsymbol{\Omega}_1$ *of the marginal parameters* $\boldsymbol{\beta}$ *has finite dimension, is closed and compact, and* $\boldsymbol{\beta}_0$*, the true parameter vector, lies in the interior of* $\boldsymbol{\Omega}_1$.

Condition 2. $E\left[\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0)\right]=\mathbf{0}$*, where* $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_p)'$.

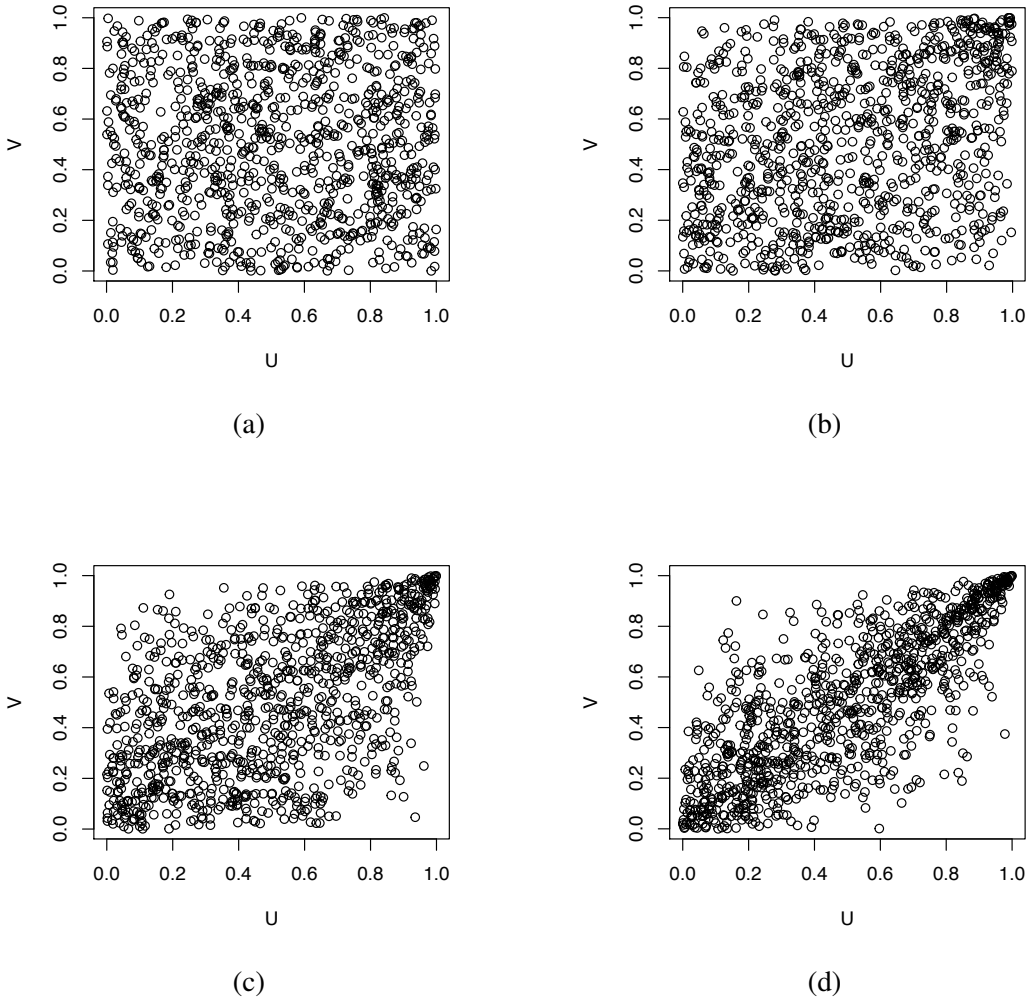
Condition 3. *The Fisher information matrix*

$$\mathbf{I}^*=E\left[-\frac{\partial}{\partial\boldsymbol{\beta}}\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta})\right]$$

is positive definite for all $\boldsymbol{\beta}\in\boldsymbol{\Omega}_1$.

Condition 4. *Second order partial derivatives of* $\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ *are bounded integrable, i.e.,* $\left|\frac{\partial^2\mathbf{U}^*_{\boldsymbol{\beta}_i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k}\right|<M_{ijk}$ *for all* $\boldsymbol{\beta}\in\boldsymbol{\Omega}_1$*, where* $E\left[M_{ijk}\right]<\infty$ *for* $i,\ j,\ k=1,...,p$.

Figure 3 – Scatterplots from samples taken from a Galambos factor copula with $\theta = 0$ (Independence (a)), 0.866 ($\tau = 0.2$ (b)), 1.538 ($\tau = 0.4$ (c)) and 2.78 ($\tau = 0.6$ (d)).



(a) (b)

(c) (d)

Condition 5. *The parameter space Ω_2 of the copula's parameter θ is closed and compact, and the true value θ_0 lies in the interior of Ω_2.*

Condition 6. $E\left[U_\theta(\boldsymbol{\beta}_0, \theta_0)\right] = 0$.

Condition 7. $E\left[U_\theta^2(\boldsymbol{\beta}_0, \theta_0)\right] = E\left[-\frac{\partial U_\theta(\boldsymbol{\beta}_0, \theta_0)}{\partial \theta}\right]$.

Condition 8. *The Fisher information*

$$I_{\theta\theta} = E\left[-\frac{\partial U_\theta(\boldsymbol{\beta}_0, \theta_0)}{\partial \theta}\right]$$

is greater than zero for all $\theta \in \Omega_2$.

Condition 9. *Second order partial derivatives of* $U_\theta(\boldsymbol{\beta},\theta)$ *are bounded integrable, i.e.,* $\left|\frac{\partial^2 U_\theta(\boldsymbol{\beta},\theta)}{\partial\theta^2}\right| < M$ *for all* $\theta \in \Omega_2$*, where* $E[M] < \infty$*.*

Condition 10. *The support of* T_{ij} $(i = 1,...,K$ *and* $j = 1,...,n_i)$ *does not depend on any* $(\boldsymbol{\beta},\theta) \in \boldsymbol{\Omega}_1 \times \Omega_2$*.*

Proof of Theorem 1

Let $\boldsymbol{\beta}_0$ denote the true parameter vector for the margins. Expanding the score function $\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ in a Taylor series around $\boldsymbol{\beta}_0$ and evaluating it at $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, we obtain, under regularity conditions 1-10,

$$\mathbf{U}^*_{\boldsymbol{\beta}}\left(\bar{\boldsymbol{\beta}}\right) = \mathbf{0} = \mathbf{U}^*_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0\right) + \left.\frac{\partial \mathbf{U}^*_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\left(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + o_p\left(\sqrt{K}\right).$$

Similarly,

$$U_\theta\left(\bar{\boldsymbol{\beta}},\bar{\theta}\right) = 0 = U_\theta\left(\boldsymbol{\beta}_0,\theta_0\right) + \left.\frac{\partial U_\theta}{\partial \boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},\theta)=(\boldsymbol{\beta}_0,\theta_0)}\left(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \left.\frac{\partial U_\theta}{\partial \theta}\right|_{(\boldsymbol{\beta},\theta)=(\boldsymbol{\beta}_0,\theta_0)}\left(\bar{\theta} - \theta_0\right) + o_p\left(\sqrt{K}\right).$$

By the law of large numbers, as $K \to \infty$,

$$-K^{-1}\left.\frac{\partial \mathbf{U}^*_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = K^{-1}\sum_{i=1}^{K} -\frac{\partial \mathbf{U}^*_{i,\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}} \to \mathbf{I}^* = E\left[-\frac{\partial \mathbf{U}^*_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}}\right],$$

$$-K^{-1}\left.\frac{\partial U_\theta}{\partial \boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},\theta)=(\boldsymbol{\beta}_0,\theta_0)} = K^{-1}\sum_{i=1}^{K} -\frac{\partial U_{i,\theta}\left(\boldsymbol{\beta}_0,\theta_0\right)}{\partial \boldsymbol{\beta}} \to \mathbf{I}_{\theta\boldsymbol{\beta}}$$

$$-K^{-1}\left.\frac{\partial U_\theta}{\partial \theta}\right|_{(\boldsymbol{\beta},\theta)=(\boldsymbol{\beta}_0,\theta_0)} = K^{-1}\sum_{i=1}^{K} -\frac{\partial U_{i,\theta}\left(\boldsymbol{\beta}_0,\theta_0\right)}{\partial \theta} \to I_{\theta\theta}.$$

Hence

$$K^{-1/2}\begin{pmatrix} \mathbf{U}^*_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0\right) \\ U_\theta\left(\boldsymbol{\beta}_0,\theta_0\right) \end{pmatrix} \to \sqrt{K}\begin{pmatrix} \mathbf{I}^* & \mathbf{0} \\ \mathbf{I}_{\theta\boldsymbol{\beta}} & I_{\theta\theta} \end{pmatrix}\begin{pmatrix} \overline{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \overline{\theta} - \theta_0 \end{pmatrix}.$$

By the central limit theorem,

$$K^{-1/2}\begin{pmatrix} \mathbf{U}^*_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0\right) \\ U_\theta\left(\boldsymbol{\beta}_0,\theta_0\right) \end{pmatrix}$$

converges to a multivariate normal distribution with mean $\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}$ and variance-covariance matrix

$$\begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & I_{\theta\theta} \end{pmatrix}$$

with $\mathbf{V} = \text{var}\left[\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0)\right] = E\left[\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0)^2\right]$. Thus,

$$\sqrt{K}\begin{pmatrix} \overline{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \overline{\theta} - \theta_0 \end{pmatrix}$$

converges to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix

$$\begin{pmatrix} \mathbf{I}^* & \mathbf{0} \\ \mathbf{I}_{\theta\boldsymbol{\beta}} & I_{\theta\theta} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & I_{\theta\theta} \end{pmatrix} \begin{pmatrix} \mathbf{I}^* & \mathbf{0} \\ \mathbf{I}_{\theta\boldsymbol{\beta}} & I_{\theta\theta} \end{pmatrix}^{-1^T} = \begin{pmatrix} (\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T} & \frac{(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T}\mathbf{I}_{\boldsymbol{\beta}\theta}}{I_{\theta\theta}} \\ \frac{-\mathbf{I}_{\theta\boldsymbol{\beta}}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T}}{I_{\theta\theta}} & \frac{1}{I_{\theta\theta}} + \frac{\mathbf{I}_{\theta\boldsymbol{\beta}}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T}\mathbf{I}_{\theta\boldsymbol{\beta}}}{I^2_{\theta\theta}} \end{pmatrix}.$$

The lower right-hand element of this matrix is the asymptotic variance of $\sqrt{K}\left(\overline{\theta} - \theta_0\right)$.

The references for this section can be found in the reference list of the paper (Second part of the thesis).

APPENDIX B – R routines

In this section we describe how to implement the proposed methods in the `R` language. The source code, with all necessary functions, will be incorporated in the `R` package `Sunclarco` in the near future. Until there, they can be requested via e-mail to the author (*eleandersoncampos@estudante.ufla.br*). Required additional packages are reported in the header of the source code. The usage of the functions is straightforward, as we detail next with an example of estimation using the three procedures.

For the sake of illustration, we simulate the data to be modeled. This can be done by using the function `Data`, available in the source code. The call for this command is as follows:

```
set.seed(1)
sim_data <- Data(nclusters = 20, copula = "clayton", theta = 1.5, rho = 1.4,
                 lambda = .4, beta = 2, percens = .25)
```

where the argument `nclusters` determines the number of clusters (K), each with size varying uniformly from 2 to 50. `copula` indicates the bivariate copula family ("gaussian", "clayton"or "galambos") to be used as intracluster dependence structure, with parameter `theta`. We stress that any parametric bivariate copula family can be used in our model, but we have implemented so far the Clayton, Gaussian and Galambos families, representing the class of Archimedean, elliptical and extreme-value copulas. The arguments `rho`, `lambda` and `beta` represents the parameters of the Weibull baseline marginal survival function, $S(t|Z) = \exp\{-\lambda \exp(\beta z) t^{\rho}\}$, where Z is a dichotomous covariate. The censoring distribution is also Weibull, with the argument `percens` $\in [0,1]$ indicating the approximate percentage of right censoring in the data. The data frame `sim_data` will have the following form:

```
   Cluster Cov         X delta
1        1   1 0.2599620     0
2        1   0 0.5813684     1
3        1   1 0.4204703     1
⋮
15       1   0 0.5567494     1
16       2   0 0.7760581     1
17       2   0 0.6878404     0
```

```
⋮

574      20   0 0.8925011     1
575      20   0 2.2655296     1
576      20   1 0.1338729     1
```

where `Cluster` denotes the number i of the cluster ($i = 1,...,K$) which contains subject j ($j = 1,...,n_i$, with n_i being the size of cluster i). The values of the dichotomous covariate Z_{ij}, $X_{ij} = \min(T_{ij}, C_{ij})$ and δ_{ij} are indicated by `Cov`, `X` and `delta`, respectively, such that in line 16, for example, $Z_{2,1} = 0$, $X_{2,1} = 0.776$ and $\delta_{2,1} = 1$ ($\delta_{ij} = 0$ if the survival time is censored).

We now proceed by using the function `twostageP` to fit a Weibull-Clayton model using the two-stage parametric procedure:

```
TSP_results <- twostageP(data = sim_data, copula = "clayton",
                         start = c(0,1,1,1), nq = 50)
```

where the argument `data` receives the data frame containing the data in the same format as generated by the function `Data`. Intracluster dependence structure is determined by the argument `copula`, with available options “gaussian", “clayton"or “galambos"as the bivariate copula models. Starting values for minimization of the likelihood are supplied as a vector $(\gamma, \rho, \lambda, \beta)$ in the argument `start`. If there is no prior knowledge for starting values, a reasonable guess would be `start = c(0,1,1,1)`. We start with $\gamma = 0$ because we optimize the likelihood either for $\gamma = \log(\theta)$ (Clayton and Galambos copulas) or $\gamma = -\log(1/\theta - 1)$ (Gaussian copula). Furthermore, the number of Gauss-Legendre quadrature points is defined by the argument `nq`. We recommend using 50 quadrature points for sufficient accuracy. The output of `TSP_results` gives us

```
$estimates
[1] 1.6269760 1.4417907 0.3506424 1.9389392
$stderr_robust
[1] 0.47634720 0.11841220 0.06516016 0.23204626
$stderr_naive
[1] 0.19279228 0.05466236 0.02823527 0.12206452
$z_values
[1]  3.415525 12.176032  5.381239  8.355831
```

```
$AIC
[1] 600.1981
```

where `$estimates` represents the estimated values of $(\theta, \rho, \lambda, \beta)$ and `$stderr_robust` their standard errors derived from the sandwich estimator. Naive standard errors are also provided for illustrational purposes and are represented by `$stderr_naive`. These are obtained from the diagonal of the inverse of the Hessian matrix. Z values and the AIC are also provided. If we use, for example, `copula = "galambos"` or `"gaussian"` as arguments of `twostageP` for the same data set, we would get an AIC of 688 and 654, respectively. These values are coherent, because the data was simulated using a Clayton copula, meaning that the AIC could be a reasonable tool for comparing the models in this context. The average computing time for two-stage parametric estimation under the settings described above is 0.83 seconds using a 2.5 GHz Intel(R) Core(TM) i5 computer running macOS Catalina. For $K = 200$ clusters, the average time is 9.1 seconds.

We now can use the estimated values from the two-stage procedure as starting values for one-stage estimation by running

```
OSP_results <- onestageP(data = sim_data, copula = "clayton",
                         start = TSP_results$estimates)
```

The arguments of `onestageP` are the same as in `twostageP`, except for `nq`, since we have implemented adaptive Gauss-Kronrod integration instead of Gauss-Legendre. The output is

```
$estimates
[1] 1.5213726 1.4515814 0.3687886 1.9295434
$stderr
[1] 0.36822983 0.07191544 0.06018928 0.11222566
$z_values
[1]  4.131584 20.184557  6.127146 17.193425
$AIC
[1] 599.993
```

and the average running time is 10.6 seconds. Although the computing time is somewhat higher, results yielded by this method are more reliable (specially for a small number of clusters).

Two-stage semiparametric estimation can be performed with the following command

```
TSS_results <- twostageS(data = sim_data, copula = "clayton",
                         start = TSP_results$estimates[1], nq = 50)
```

where the argument `start` only receives a starting value of γ. The average computing time is 1.28 seconds. Reported results are

```
$theta_estimate
[1] 1.537098
$theta_stderr
[1] 0.5957812
$AIC
[1] -1.89
$coxPH
Call:
coxph(formula = Surv(X, delta) ~ Cov + cluster(Cluster), data = samp,
    method = "efron")
      coef exp(coef) se(coef) robust se     z        p
Cov 2.0054    7.4288   0.1324    0.2494 8.042 8.84e-16
Likelihood ratio test=242.2  on 1 df, p=< 2.2e-16
n= 576, number of events= 411
```

This method makes use of the function `coxph` (from the `R` package `survival`) to fit a Cox proportional hazards model to the marginal survival functions. In view of this, the output `$coxPH` is a `coxph.object`.