Factor copula models for right-censored clustered survival data

Peer-reviewed author version

CAMPOS EUGENIO FILHO, Eleanderson; BRAEKERS, Roel; de Souza, Devanil J. & Chaves, Lucas M. (2021) Factor copula models for right-censored clustered survival data. In: LIFETIME DATA ANALYSIS,.

DOI: 10.1007/s10985-021-09525-5
Handle: http://hdl.handle.net/1942/34331

Lifetime Data Analysis manuscript No.
(will be inserted by the editor)

Factor copula models for right-censored clustered survival data

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

Received: date / Accepted: date

Abstract In this article we extend the factor copula model to deal with 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. Three estimation procedures are proposed: 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 · Factor copula models · Intracluster dependence · 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).

Despite allowing clusters to have different sizes, it is not possible to use any combination of marginal survival functions and dependence structures to build a frailty model (Hougaard, 2000). Furthermore, according to Duchateau and Janssen (2008), choosing the right frailty distribution in a frailty model is of great importance, because of its relationship (even though it is not explicit) to the shape of intracluster dependence. However, only a few distributions have been commonly used in practice (mainly the gamma and lognormal distributions). These flexibility issues have motived the alternative use of copula-based approaches to model clustered survival data.

Owing to the groundbreaking work of Sklar (1959), copula-based models can easily overcome the dependence modeling limitations 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, they impose no restrictions on the choice of the marginal survival functions and dependence structure (modeled by the copula itself). Moreover, there are several families of copulas already available in the literature, each representing a unique dependence structure. They can be easily implemented to model intracluster dependence in multivariate survival data, being readily available in several statistical software packages.

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 sizes, although important classes of copulas were still not comprehended, such as 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 with fixed-size clusters.

More recently, Barthel et al. (2018, 2019) proposed a vine copula-based approach to model clustered survival data. Within their model framework, any type of bivariate copula can be used to model the different pairwise dependencies. However, the model of Barthel et al. (2018) can only deal with clusters of fixed size. Although variable sized clusters are allowed in Barthel et al. (2019), the authors were concerned with recurrent event times, a different setting from our case. In both cases, the authors worked with small sized clusters due to the high computational cost of making inference for vine copulas in high dimensions.

Motivated by the aforementioned shortcomings, we propose a new copula-based methodology for right-censored clustered survival data modeling. The new methods impose no restrictions on the choice of copula families and allow clusters with varying sizes (possibly large). This provides the means to model multivariate 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 (allowing for covariates). Additionally, we propose a semiparametric formulation for the marginal survival function with incorporation of time dependent covariates. Our methodol-

ogy extends the one-factor copula model of Krupskii and Joe (2013) by providing new inferential tools for right-censored clustered survival data. The proposed model includes the Archimedean copula model of Prenen et al. (2017a) and the Gaussian copula model of Othus and Li (2010) as particular cases.

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 grouped 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 two propositions in Section 2.1 and 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 $\mathcal{C}_{ij}$. Considering a right censoring scheme, the observed quantities are

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

where each lifetime T_{ij} may depend on a vector $\mathbf{Z}_{ij}$ of covariates. Also, let V_i ($i = 1,...,K$) be Uniform$[0,1]$ random variables. We then assume that, within cluster i, the lifetimes are conditionally independent given V_i and the covariates. In other words, V_i behaves as a latent variable (common factor) that governs the associations between the lifetimes in cluster i. Hence, conditional on V_i and $\mathbf{Z}_{ij}$ ($j = 1,...,n_i$), we can write the joint survival function in cluster i, as

$$S(t_{i1},...,t_{in_i}|V_i,\mathbf{Z}_{i1},...,\mathbf{Z}_{in_i}) = \prod_{j=1}^{n_i} S(t_{ij}|V_i,\mathbf{Z}_{ij}),$$

where $S(t_{ij}|V_i,\mathbf{Z}_{ij})$ is the conditional survival function of $T_{ij}|\mathbf{Z}_{ij}$ given $V_i = v_i$, that is, $S(t_{ij}|V_i,\mathbf{Z}_{ij})$ is the (negative) partial derivative of the bivariate survival function $S(t_{ij},v_i|\mathbf{Z}_{ij})$ with respect to v_i:

$$S(t_{ij}|V_i,\mathbf{Z}_{ij}) = -\frac{\partial}{\partial v_i} S(t_{ij},v_i|\mathbf{Z}_{ij}).$$

By Sklar's theorem, we can rewrite the equation above in terms of a bivariate copula

$$S(t_{ij}|V_i,\mathbf{Z}_{ij}) = -\frac{\partial}{\partial v_i}\mathrm{C}_{ij}(S(t_{ij}|\mathbf{Z}_{ij}),S(v_i)),$$

where $S(t_{ij}|\mathbf{Z}_{ij})$ is the marginal survival function of $T_{ij}|\mathbf{Z}_{ij}$, $S(v_i)$ is the marginal survival function of V_i and C_{ij} is the bivariate copula that joins $T_{ij}|\mathbf{Z}_{ij}$ to V_i. Considering that $S(t_{ij}|V_i,\mathbf{Z}_{ij})$ comes from the bivariate copula C_{ij}, we shall denote the former by $\mathrm{C}_{ij|V}(S(t_{ij}|\mathbf{Z}_{ij})|v_i)$. Thus, we have that

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

It is reasonable to assume that every subject in cluster i is affected by V_i in a similar fashion (exchangeability). In this case, it follows that

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

Now we can retrieve the unconditional joint survival function of cluster i by integrating V_i out of (1)

$$S(t_{i1},...,t_{in_i}|\mathbf{Z}_{i1},...,\mathbf{Z}_{in_i}) = \int_0^1\prod_{j=1}^{n_i}\mathrm{C}_{\cdot|V}(S(t_{ij}|\mathbf{Z}_{ij})|v_i)dv_i. \tag{2}$$

Following Krupskii and Joe (2013), we will call (2) 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 (2) 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

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

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

where $f(x_{ij}|\mathbf{Z}_{ij}) = -dS/dx_{ij}$ is the 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

$$L = \prod_{i=1}^{K} \int_0^1 \prod_{j=1}^{n_i} \left\{ \mathrm{c}_{\cdot V}(S(x_{ij}|\mathbf{Z}_{ij}), v_i) f(x_{ij}|\mathbf{Z}_{ij}) \right\}^{\delta_{ij}} \times \mathrm{C}_{\cdot|V}(S(x_{ij}|\mathbf{Z}_{ij})|v_i)^{1-\delta_{ij}} dv_i. \quad (3)$$

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 (3) do not have analytical solutions. 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 $y_k^* = 0.5 y_k + 0.5$ 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.

2.1 Frailty models *versus* copula-based models

Frailty models have been traditionally used in the literature to model survival data grouped in clusters of variable size and, more recently, copula-based models became a very popular alternative. Although dealing with the same data settings, these two approaches are quite different. As discussed by Goethals et al. (2008), a notable distinction between these models is in the way that the survival function of an individual

is modeled, for example. In copula-based models, the survival function is treated separately, from a *marginal* perspective (population-average), while in frailty models, the survival function of a subject is modeled from a *conditional* (to the cluster) point of view. In addition, there is an important difference regarding the interpretation of the dependence parameter in the two models. Since copulas are able to exclusively model the dependence between random variables, they allow a direct interpretation of their parameters, whereas in frailty models the interpretation of the parameter is different (heterogeneity between clusters). Curiously, as observed by Goethals et al. (2008), there are a few structural similarities between frailty and copula models, but these are limited to the Archimedean world. We are, indeed, speaking about two distinct classes of models with their own specificities.

The increasing popularity of the copula-based approaches to clustered survival data modeling is due to their flexibility when dealing with dependence modeling: there are numerous families of copulas available, each representing a different dependence structure, and they are fairly simple to implement from a computational viewpoint. The R package *copula*, for example, offers many possibilities (Yan et al., 2007). In practice, the number of different copula families is enormous, since there are even methods for creating new copulas, such as the methods based on copula mixtures and rotations. In this sense, copula-based models provide almost unlimited options to model intracluster dependence. For example, within our model framework, by choosing

$$\mathrm{C}_{\cdot|V}(u_j|v_i) = \Phi\left(\frac{\Phi^{-1}(u_j) - \theta\Phi^{-1}(v)}{(1-\theta^2)^{1/2}}\right), \quad j = 1,...,n_i \tag{4}$$

in (2), where the expression in the RHS is the partial derivative with respect to v of a bivariate Gaussian copula ($\theta > 0$), we can model a positive (symmetric) relationship with weak tail association between subjects in a cluster, such as in the model of Othus and Li (2010). Within the same class of elliptical copulas, stronger tail associations can be captured by the t-copula if we set $\mathrm{C}_{\cdot|V}(u_j|v_i)$ to be the partial derivative (with respect to v) of a bivariate t-copula. Alternatively, asymmetrical tail dependence structures (association earlier *versus* later in time) can be modeled by using the Archimedean Clayton (stronger association between late lifetimes) copula, such as in the model of Prenen et al. (2017a), or the extreme-value Galambos (stronger association between early lifetimes) copula. The flexibility here lies precisely on the extensive list of bivariate copula families that can be used within our model framework to achieve, virtually, any type of intracluster dependence. In fact, we show that the Archimedean copula model of Prenen et al. (2017a) and the Gaussian copula model of Othus and Li (2010) (for $\sigma > 0$) are subclasses of our model.

Proposition 1 *Let $\varphi_\theta \in \mathcal{L}_\infty$ be a generator function from the class of Laplace transforms of non-negative random variables with no mass at 0, i.e., $\varphi_\theta : [0,+\infty) \to [0,1]$, 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}. Also, φ_θ is the Laplace transform of a distribution function $G_\theta(x)$ with inverse $G_\theta^{-1}(x)$ and $G_\theta(0) = 0$. Then, for*

$$\mathrm{C}_{\cdot|V}(u_j|v) = \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\}, \tag{5}$$

it follows that, for all $u_j \in [0,1]$ (j=1,...,n),

$$\int_0^1 \prod_{j=1}^{n} \mathrm{C}_{\cdot|V}(u_j|v)\,dv = \int_0^{+\infty} \prod_{j=1}^{n} \exp\left\{-x\varphi_\theta^{-1}(u_j)\right\} dG_\theta(x), \tag{6}$$

where the expression on the right-hand side is the Archimedean copula model of Prenen et al. (2017a).

Proposition 2 *Let $\tilde{T}_{ij} = \sqrt{\sigma}b_i + \varepsilon_{ij}$ be the transformed lifetime of the j^{th} ($j = 1,...,n_i$) subject in cluster i ($i = 1,...,K$) as defined in the marginalized frailty model extension of Othus and Li (2010), where $\sigma > 0$, $b_i \sim N(0,1)$ is a cluster-level frailty and $\varepsilon_{ij} \sim N(0,1-\sigma)$ is an error term. Assume also that $b_i, \varepsilon_{i1}, ..., \varepsilon_{in_i}$ are mutually independent. Then, the joint cumulative distribution function of $\tilde{T}_{i1}, ..., \tilde{T}_{in}$ can be written as a one-factor copula model.*

The proof of the propositions above can be found in the Appendix. For a more complete list of bivariate copula families with their dependence properties, the reader is referred to the books of Nelsen (2007) and Joe (2014).

If clusters have a fixed size n, it is also possible to adapt our methodology to the new settings. This can be done by allowing $\mathrm{C}_{\cdot|V}(u_j|v)$ in (2) to be different for every j, resulting in

$$S(t_{i1},...,t_{in}|\mathbf{Z}_{i1},...,\mathbf{Z}_{in}) = \int_0^1 \prod_{j=1}^{n} \mathrm{C}_{j|V}(S(t_{ij}|\mathbf{Z}_{ij})|v_i)dv_i.$$

By doing this, we can achieve different pairwise dependencies for the lifetimes in a cluster, similarly as done by Barthel et al. (2018) using vine copulas. In this sense, it is even possible to use a different survival function for each component of the cluster.

2.2 Measuring intracluster dependence: the role of the Kendall's tau

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 T_j and T_k denote the lifetimes of two arbitrary individuals belonging to a cluster with joint survival function given by expression 2. Owing to the factor copula's framework, the relationship between the lifetimes T_j and T_k is uniquely determined by a latent variable V through

$$\mathrm{C}_{jk}(S(t_j|\mathbf{Z}_j), S(t_k|\mathbf{Z}_k);\theta) = \int_0^1 \mathrm{C}_{\cdot|V}(S(t_j|\mathbf{Z}_j)|v;\theta)\mathrm{C}_{\cdot|V}(S(t_k|\mathbf{Z}_k)|v;\theta)dv,$$

which is the same as looking at the relationship between U_j and U_k in the underlying copula

$$\mathrm{C}_{jk}(u_j,u_k;\theta) = \int_0^1 \mathrm{C}_{\cdot|V}(u_j|v;\theta)\mathrm{C}_{\cdot|V}(u_k|v;\theta)dv. \tag{7}$$

Following a well known result (see Nelsen, 2007, p. 164), we can write the Kendall's tau of U_j and U_k as

$$\tau_{jk} = 1 - 4\int_0^1 \int_0^1 \mathrm{C}_{j|k}(u_j|u_k)\mathrm{C}_{k|j}(u_k|u_j)du_j du_k.$$

Hence, we must obtain the conditional cumulative distribution functions $\mathrm{C}_{j|k}(u_j|u_k)$ and $\mathrm{C}_{k|j}(u_k|u_j)$ from (7):

$$\begin{aligned}\mathrm{C}_{k|j}(u_k|u_j) &= \frac{\partial}{\partial u_j}\int_0^1 \mathrm{C}_{\cdot|V}(u_j|v)\mathrm{C}_{\cdot|V}(u_k|v)dv \\ &= \int_0^1 c_{\cdot V}(u_j,v)\mathrm{C}_{\cdot|V}(u_k|v)dv.\end{aligned}$$

Similarly,

$$\mathrm{C}_{j|k}(u_j|u_k) = \int_0^1 \mathrm{C}_{\cdot|V}(u_j|v)c_{\cdot V}(u_k,v)dv.$$

Therefore, the Kendall's tau for the pair (U_j,U_k) - or, equivalently, (T_j,T_k) - in a one-factor copula model is given by

$$\tau_{jk} = 1-4\int_0^1 \int_0^1 \left(\int_0^1 c_{\cdot V}(u_j,v)\mathrm{C}_{\cdot|V}(u_k|v)dv\right)\left(\int_0^1 \mathrm{C}_{\cdot|V}(u_j|v)c_{\cdot V}(u_k,v)dv\right)du_j du_k. \tag{8}$$

This expression cannot be evaluated analytically, except for some specific choices of the bivariate copula $\mathrm{C}_{\cdot V}$ (e.g., Gaussian copula and FGM copula), but it can be easily computed with numerical integration methods for any given expression of $\mathrm{C}_{\cdot|V}$ together with the value of its parameter θ.

From the perspective of our model, the Kendall's tau as computed by (8) 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 ($\tau_{\cdot V}$) and, as a consequence, the same measure of association between each other, i.e., τ_{jk} is the same for every $j,k = 1,...,n$ with $j \neq k$ and n being the cluster size.

Although $\tau_{\cdot V}$ is allowed to assume any value in $[-1,1]$, the negative range is not covered by τ_{jk} when the clusters have variable sizes in a one-factor copula model. This can be demonstrated by showing that $\tau_{jk} = 1$ when $\mathrm{C}_{\cdot V}$ is equal to either W or M, the Fréchet-Hoeffding lower bound (perfect negative dependence) and upper bound (perfect positive dependence), respectively. Since $\tau_{jk} = 0$ when $\mathrm{C}_{\cdot V} = \Pi$ (independence/product copula) and knowing that every copula with a negative dependence structure lies between W and Π, it follows that $0 \leq \tau_{jk} \leq 1$ for any given pair (T_j,T_k) of lifetimes in our model. Furthermore, it can also be shown that $\tau_{jk} \leq \tau_{\cdot V}$, i.e., the strength of association between T_j and T_k is weaker than that between T_j (or T_k) and V, unless $\tau_{\cdot V} = 1$, in which case τ_{jk} will also be equal to 1.

3 Estimation

Our estimation procedures are based on 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 (3). 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 $\mathbf{I}^{-1}$, the inverse of the Fisher information matrix. 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

$$\begin{aligned}\mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta}) &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\delta_{ij}\frac{\partial \log f(x_{ij}|\mathbf{Z}_{ij})}{\partial \boldsymbol{\beta}} + (1-\delta_{ij})\frac{\partial \log S(x_{ij}|\mathbf{Z}_{ij})}{\partial \boldsymbol{\beta}}\right)\\ &= \sum_{i=1}^{K}\mathbf{U}^*_{i,\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\mathbf{U}^*_{i,j,\boldsymbol{\beta}}(\boldsymbol{\beta}) = \mathbf{0}. \qquad (9)\end{aligned}$$

Let $\bar{\boldsymbol{\beta}}$ be the estimator obtained from (9). 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 (3) 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*

$$\mathrm{var}(\bar{\theta}) = \frac{1}{I_{\theta\theta}} + \frac{\mathbf{I}_{\theta\boldsymbol{\beta}}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\boldsymbol{\beta}\theta}}{I^2_{\theta\theta}}, \qquad (10)$$

where the Fisher information matrix $\mathbf{I} = E_{\boldsymbol{\eta}}\left[-\nabla^2 \log L(\boldsymbol{\eta})\right]$, $\boldsymbol{\eta} = (\boldsymbol{\beta}, \theta)$ *with size* $(p+1)$*, is partitioned as follows*

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

A proof of this theorem is given in Prenen et al. (2017a).

Therefore, estimates for the standard errors of $\bar{\boldsymbol{\beta}}$ and $\bar{\theta}$ can be obtained, respectively, by the square root of the diagonal of $\left(\widehat{\mathbf{I}}^*\right)^{-1}\widehat{\mathbf{V}}\left(\widehat{\mathbf{I}}^*\right)^{-1}$, where

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

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

and by the square root of expression (10) after replacing the quantities $I_{\theta\theta}$, $\mathbf{I}_{\theta\boldsymbol{\beta}}$ and $\mathbf{I}_{\boldsymbol{\beta}\theta}$ for their empirical counterparts 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 $\bar{\theta}$. The matrix $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$ should also be replaced by $\left(\widehat{\mathbf{I}}^*\right)^{-1}\widehat{\mathbf{V}}\left(\widehat{\mathbf{I}}^*\right)^{-1}$.

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, $\breve{\boldsymbol{\beta}}$ and $\breve{\Lambda}$, 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 practice, the `R` package `survival` (Therneau, 2015) provides the necessary functions to retrieve $\breve{\boldsymbol{\beta}}$, $\breve{\Lambda}$ and their standard errors.

In the second stage, the estimate $\breve{\theta}$ 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 $\max_\theta \left\{L\left(\theta, \breve{\boldsymbol{\beta}}, \breve{\Lambda}\right)\right\}$.

Theorem 2 *Under regularity conditions 1-8 in the Appendix,* $\left(\breve{\theta}, \breve{\boldsymbol{\beta}}, \breve{\Lambda}\right)$ *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*

$$\operatorname{var}(\Xi) / W(\theta_0)^2. \tag{11}$$

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 copulas. An analytical formula to compute the standard error of $\check{\theta}$ is also provided in the Appendix and it will be soon implemented in the `R` package `Sunclarco`. With this formula, we were able to drastically reduce the computing time of the two-stage semiparametric procedure compared to the grouped jackknife alternative employed by Prenen et al. (2017a) and Othus and Li (2010).

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 degrees of association for each case (Kendall's $\tau \approx 0.2$, 0.4, 0.6 computed with formula (8)). 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 $\beta = 3$. 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 $\lambda_{\mathcal{C}} = 0.425$, $\rho_{\mathcal{C}} = 1.6$, for 25% of censoring and $\lambda_{\mathcal{C}} = 2.241$, $\rho_{\mathcal{C}} = 1.6$, 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 $\check{\theta}$ 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 (10) and (11), 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

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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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	0.2	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%)
	0.4	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%)
	0.6	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%)

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 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. Although this behavior appears to be counterintuitive, similar results were obtained in the simulation studies of Prenen et al. (2017a). One possible explanation is that the upper tail of extreme-value copulas is directly affected by the right censoring, i.e., the higher the percentage of censoring, the more information in the upper tail of an EV copula is lost. Therefore, results are expected to be worse. On the other hand, since the Clayton copula has lower tail dependence, increasing the percentage of right censoring would give more emphasis on the lower tail. 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 $K \leq 200$. This effect was studied by Joe (2005) for copula models under two-stage estimation (IFM), who showed that the efficiency of the IFM estimators tend to be affected when there is strong dependence (high values of θ). 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.

We also conduct an additional simulation study to assess the use of the AIC on the discrimination between different factor copula models under the two-stage inference method. In this case, we simulate 500 data sets coming from different copula models under a Kendall's $\tau = 0.35$: Clayton PBD from the Archimedean model of Prenen et al. (2017a) ($\theta = 1.107$), Factor Clayton ($\theta = 2$), Gaussian OL model of Othus and Li (2010) ($\theta = 0.723$) and the Factor Galambos ($\theta = 1.339$). We then

use the two-stage parametric method to fit, to each generated data set (coming from a particular copula, the true model), the four copula models and calculate the delta AIC, $\Delta = \text{AIC}_{\text{Fitted}} - \text{AIC}_{\text{True}}$. The results can be seen in Table 4 for a scenario with $K = 50$ clusters with size varying uniformly from 2 to 50 and for 25% of right censoring. Our criterion here is that two models are indistinguishable if their delta AIC is less than two. As can be seen in Table 4, the high values of mean delta AIC (and low standard deviation), indicate that the AIC was able to easily identify the best model in almost every situation. Another evidence is the small number of delta AICs less than two. Therefore, we conclude that, at least for applications where the global scenario is close to the one we have considered, our inference method is robust to detect deviations from the true copula model.

Table 4 Comparison of different copula models based on the delta AIC values ($\Delta = \text{AIC}_{\text{Fitted}} - \text{AIC}_{\text{True}}$). Values on the first rows represent the mean delta AIC calculated from 500 simulations, while the second rows contain the standard deviation of Δ and the number of times $\Delta \leq 2$.

True model	Fitted model			
	Factor Clayton	Clayton PBD	Gaussian OL	Factor Galambos
Factor Clayton	0	53.19 (17.77; 0)	96.61 (28.57; 0)	213.62 (47.85; 0)
Clayton PBD	47.83 (16.02; 0)	0	120.49 (27.28; 0)	221.67 (41.53; 0)
Gaussian OL	133.84 (41.69; 0)	112.28 (32.45; 0)	0	41.17 (19.00; 7)
Factor Galambos	297.42 (88.48; 0)	251.93 (81.23; 0)	44.41 (15.24; 0)	0

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 four different factor copulas to model the association between times to first insemination: the first built with a bivariate Clayton copula (Factor Clayton); the second using a bivariate Gaussian copula (Gaussian OL or Factor Gaussian), which leads to the model of Othus and Li (2010); the third built with a bivariate Galambos copula (Factor Galambos) and the fourth being the Clayton copula model (Clayton PBD) of Prenen et al. (2017a). Being particular cases of the one-factor copula, the Gaussian OL and Clayton PBD models can be obtained by setting $C_{\cdot|V}(u_j|v)$ as in expressions (4) and (5) (using the generator of the Clayton copula and the gamma distribution with parameters shape θ and scale $1/\theta$), respectively. In all four settings, we use a baseline Weibull survival function (one- and two-stage parametric 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

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

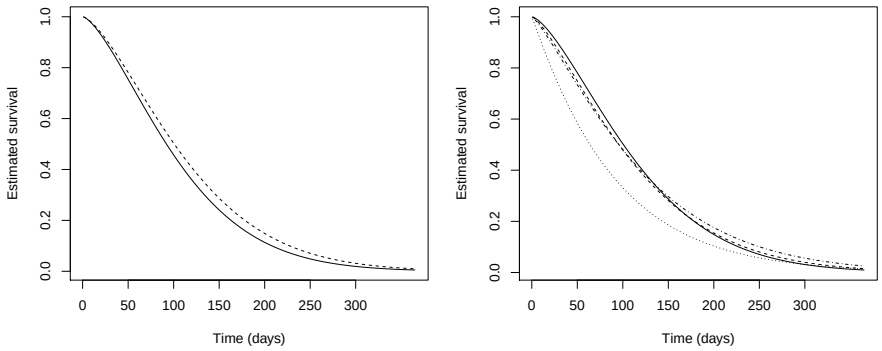
where β is the parity effect for the dichotomous covariate Z.

Using the one- and two-stage procedures for estimation, we provide results for the parity effect and association parameters for all four factor copula settings considered, along with their AIC (see Table 5). 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 5 Estimation results for the insemination dataset.

	Factor Clayton			Gaussian OL			Factor Galambos			Clayton PBD		
	Weibull one stage	Weibull two stage	Cox two stage	Weibull one stage	Weibull two stage	Cox two stage	Weibull one stage	Weibull two stage	Cox two stage	Weibull one stage	Weibull two stage	Cox two stage
β	−0.138	−0.066	−0.060	−0.132	−0.066	−0.060	−0.100	−0.066	−0.060	−0.132	−0.066	−0.060
	(0.016)	(0.022)	(0.021)	(0.017)	(0.022)	(0.021)	(0.015)	(0.022)	(0.021)	(0.018)	(0.022)	(0.021)
θ	0.779	0.829	0.995	0.623	0.575	0.520	1.236	0.916	0.768	0.209	0.325	0.448
	(0.023)	(0.143)	(0.098)	(0.020)	(0.034)	(0.021)	(0.042)	(0.038)	(0.035)	(0.014)	(0.043)	(0.062)
τ	0.132	0.143	0.177	0.254	0.214	0.174	0.321	0.218	0.164	0.095	0.140	0.183
AIC	−7144.8	−7008.0	149.1	−7038.4	−7004.2	511.3	−6941.7	−6731.9	1032.2	−7445.3	−7355.2	−71.59

The parity of the cow had 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 0.88 (95% confidence interval (CI) $[0.84, 0.91]$), 0.87 (95% confidence interval (CI) $[0.84, 0.90]$), 0.90 (95% confidence interval (CI) $[0.88, 0.93]$) and 0.88 (95% confidence interval (CI) $[0.84, 0.91]$) for the Gaussian OL, Factor Clayton, Factor Galambos and Clayton PBD models, 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 four copula models. According to the AIC, the Clayton PBD copula presented the best fit among the four models for every estimation procedure. Considering that the Clayton copula has lower tail dependence, it can be inferred, in this context, that later



(a) **(b)**

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

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 (0.209 with 95% confidence interval (CI) $[0.182, 0.236]$), which is equivalent to a Kendall's τ of 0.095. This means that the new methodology is not only capable of controlling for cluster effect, but to assess the shape of intra-cluster dependence (allowing the use of any copula family) and its strength. This can be crucial when Archimedean or Gaussian copulas are not suited to the data.

As can be seen in Figure 1, for the Weibull-Clayton PBD model, the estimated survival curve for primiparous cows is greater than the one from multiparous cows, meaning that multiparous cows are inseminated earlier than primiparous cows, with approximately 50% of the multiparous cows being inseminated before 92 days, while for primiparous cows, the estimated median is 101 days. This difference in the median time is more accentuated when comparing the estimated marginal survival curves for the four 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 three for lower values of the variable time until insemination. Indeed, estimated median survival times for the Gaussian OL, Clayton PBD and Clayton factor copulas are, respectively, 91, 101 and 91 days for multiparous cows, while for the Galambos it is 64 days.

By deriving the expression for the survival function of a subject in a given cluster (using the fitted factor copula model and the observed data in the cluster),

$$S(t^*|X_{i1} = x_{i1}, ..., X_{in_i} = x_{in_i}), \tag{12}$$

we make it possible to rank the clusters in terms of any quantiles, or even the mean or median survival times. This had not been done so far for copula-based models as pointed out by Schneider et al. (2020). We illustrate this concept in Figure 2 for the insemination dataset, where we plot the median times until first insemination of primiparous and multiparous cows for each cluster under the Weibull-Clayton PBD model (one-stage method).

Based on an optimal calving interval between 12 and 13 months and an average gestation period of 282 days, we calculate the ideal times until first insemination of 83 and 113 days, which are displayed in Figure 2. In this way, it is possible to verify which farms are performing better. In other words, we expect that the median times until first insemination should be within the interval of 83 and 113 days or close.

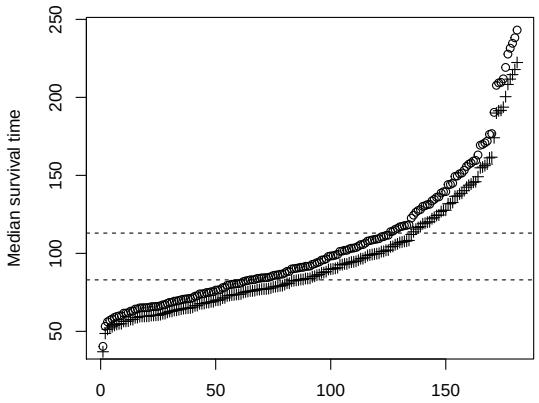


Fig. 2 Median times until first insemination (by cluster) for primiparous (circle) and multiparous (cross) cows obtained from the fitted Weibull-Clayton PBD factor copula model under one-stage estimation. Dashed lines represent optimal times of insemination of 83 and 113 days.

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 methodology that allows for clusters to be large and with variable size, altogether with the flexibility of supporting any copula family. 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 semiparametric 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.

Our perspectives for future works include the extension of our methods to also allow for dependent censoring and semicompeting risks, following ideas from Schneider et al. (2020) and Emura et al. (2017). We also plan on broadening the current factor copula methodology to accommodate for nested structures with more latent variables ($p > 1$). The computational routines that we have developed will soon be included in the `R` package `Sunclarco` as well.

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 deep appreciation to the editor, the associate editor and the reviewers for their comments on our manuscript. They have allowed us to majorly improve our work.

Appendix

Proof of Proposition 1

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 (6) as

$$\int_0^1 \prod_{j=1}^{n} \mathrm{C}_{\cdot|V}(u_j|v)\,dv = \int_0^1 \prod_{j=1}^{n} \exp\left\{-G_\theta^{-1}(v)\varphi_\theta^{-1}(u_j)\right\} dv. \tag{13}$$

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 (13), we get

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

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

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

or, equivalently,

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

which is the conditional distribution derived from

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

a bivariate function with the following properties:

1) $\mathrm{C}_{\cdot V}(u_j,v)$ is grounded

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

2) $\mathrm{C}_{\cdot V}(u_j,v)$ has margins u_j and v

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

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

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

Therefore, $\mathrm{C}_{\cdot V}(u_j, v)$ is a bivariate copula, proving that the Archimedean copula model of Prenen et al. (2017a) is a subclass of our model.

Proof of Proposition 2

Let $\tilde{T}_{i0} = b_i$ and $\varepsilon_{ij} \sim N(0,1)$, such that $\tilde{T}_{ij} = \sqrt{\sigma} b_i + \sqrt{1-\sigma}\varepsilon_{ij}$. Since $b_i, \varepsilon_{i1}, ..., \varepsilon_{in_i}$ are independent standard normal random variables, their joint density is $f_{b_i,\varepsilon_{i1},...,\varepsilon_{in_i}} = \phi(b_i)\phi(\varepsilon_{i1})...\phi(\varepsilon_{in_i})$. Given the set of transformation functions

$$\begin{aligned}
\tilde{T}_{i0} &= g_0(b_i, \varepsilon_{i1}, ..., \varepsilon_{in_i}) = b_i \\
\tilde{T}_{i1} &= g_1(b_i, \varepsilon_{i1}, ..., \varepsilon_{in_i}) = \sqrt{\sigma} b_i + \sqrt{1-\sigma}\varepsilon_{i1} \\
&\vdots \\
\tilde{T}_{in_i} &= g_{n_i}(b_i, \varepsilon_{i1}, ..., \varepsilon_{in_i}) = \sqrt{\sigma} b_i + \sqrt{1-\sigma}\varepsilon_{in_i},
\end{aligned}$$

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

$$\begin{aligned}
b_i &= g_0^{-1}(\tilde{T}_{i0}, \tilde{T}_{i1}, ..., \tilde{T}_{in_i}) = \tilde{T}_{i0} \\
\varepsilon_{i1} &= g_1^{-1}(\tilde{T}_{i0}, \tilde{T}_{i1}, ..., \tilde{T}_{in_i}) = \frac{\tilde{T}_{i1} - \sqrt{\sigma}\tilde{T}_{i0}}{\sqrt{1-\sigma}} \\
&\vdots \\
\varepsilon_{in_i} &= g_{n_i}^{-1}(\tilde{T}_{i0}, \tilde{T}_{i1}, ..., \tilde{T}_{in_i}) = \frac{\tilde{T}_{in_i} - \sqrt{\sigma}\tilde{T}_{i0}}{\sqrt{1-\sigma}}.
\end{aligned}$$

It can be shown that the Jacobian of the transformations above is given by

$$J = \prod_{j=1}^{n_i} \frac{1}{\sqrt{1-\sigma}}.$$

Hence, the joint cumulative distribution function of $\tilde{T}_{i1}, ..., \tilde{T}_{in_i}$ after integrating out $\tilde{T}_{i0}$ and making appropriate variable changes is given by

$$\begin{aligned}
F_{\tilde{T}_{i1},...,\tilde{T}_{in_i}} &:= \mathrm{C}(\Phi(\tilde{t}_{i1}), ..., \Phi(\tilde{t}_{in_i})) = \int_{-\infty}^{+\infty} \left[\prod_{j=1}^{n_i} \Phi\left(\frac{\tilde{t}_{ij} - \sqrt{\sigma} b_i}{\sqrt{1-\sigma}} \right) \right] \phi(b_i) db_i \\
&= \int_0^1 \prod_{j=1}^{n_i} \Phi\left(\frac{\tilde{t}_{ij} - \sqrt{\sigma}\Phi^{-1}(v_i)}{\sqrt{1-\sigma}} \right) dv_i,
\end{aligned}$$

a one-factor copula model where the bivariate linking copulas are all Gaussian with correlation parameter $\theta = \sqrt{\sigma}$.

On the original time scale:

$$S(t_{i1}, ..., t_{in_i} | \mathbf{Z}_{i1}, ..., \mathbf{Z}_{in_i}) = \int_0^1 \prod_{j=1}^{n_i} \Phi\left(\frac{\Phi^{-1}(S(t_{ij}|\mathbf{Z}_{ij})) - \sqrt{\sigma}\Phi^{-1}(v_i)}{\sqrt{1-\sigma}} \right) dv_i,$$

where $S(t_{ij}|\mathbf{Z}_{ij})$ is the marginal survival function associated with a Cox proportional hazards model. □

For convenience, we first introduce some notations and definitions adapted from Prenen et al. (2017a) and Othus and Li (2010) before we continue with the other proofs.

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

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

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

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

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

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

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

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

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

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

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* $k = 1,...,p$,

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

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 $\check{\boldsymbol{\beta}}$ and $\check{\Lambda}$ was already proved in Spiekerman and Lin (1998), we only show the consistency of $\check{\theta}$. 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

$$P(v_i|H_{i\cdot},\theta)=\frac{\exp\left\{\sum_{j=1}^{n_i}\log\mathbf{C}(H_{ij},v_i,\theta)\right\}}{\displaystyle\int_0^1\exp\left\{\sum_{j=1}^{n_i}\log\mathbf{C}(H_{ij},v_i,\theta)\right\}dv_i} \tag{14}$$

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 $\breve{l}_K(\theta)$ around $\boldsymbol{\beta}_0$ and Λ_0, given by

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

Another (intuitive) notation is

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

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

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

uniformly in $\breve{\Lambda}-\Lambda_0$ in all compact subsets of $\mathbb{D}$, the space of cumulative hazard functions. Since $\breve{\boldsymbol{\beta}}$ is consistent and $\breve{\Lambda}$ is uniformly consistent (Spiekerman and Lin, 1998), $R = o_p(1)$. To prove that $\breve{\theta}$ 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 $l_{K0}(\theta)$:

$$\sup_{\theta\in\nu}\left|\breve{l}_K(\theta) - E[l_{K0}(\theta)]\right| = o_P(1). \tag{15}$$

This can be shown as follows:

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

By the law of large numbers, for fixed θ

$$l_{K0}(\theta) - E[l_{K0}(\theta)] \xrightarrow{p} 0. \tag{16}$$

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

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

For this reason,

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

Using result (16), the consistency of $\check{\boldsymbol{\beta}}$ and the uniform consistency of $\check{\Lambda}$ and the fact that $R = o_p(1)$, we obtain

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

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

$$E[l_{K0}(\theta)] - E[l_{K0}(\theta_0)] < 0. \tag{17}$$

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 (15) implies that

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

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

$$\begin{aligned}
&E\left[l_{K0}\left(\theta_0\right)\right] - E\left[l_{K0}\left(\check{\theta}\right)\right] \leq \check{l}_K\left(\check{\theta}\right) - E\left[l_{K0}\left(\check{\theta}\right)\right] + o_p(1) \\
&\leq \sup_{\theta\in\Theta}\left|\check{l}_K\left(\theta\right) - E\left[l_{K0}\left(\theta\right)\right]\right| + o_p(1) = o_p(1). \qquad (18)
\end{aligned}$$

Now take θ such that $|\theta - \theta_0| \geq \varepsilon$ for any fixed $\varepsilon > 0$. By inequality (17), 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 (18) 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

$$\sqrt{K}\left(\check{\theta} - \theta_0\right) = \frac{\sqrt{K}\check{U}_K(\theta_0)}{-\partial \check{U}_K/\partial\theta\big|_{\theta=\theta^*}}. \tag{19}$$

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

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

We shall show that the score equation $\check{U}_K(\theta_0)$ in the numerator of Equation (19) 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

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

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

and $P(v_i|H_{i\cdot},\theta)$ has the same definition as in expression (14). 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

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

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 $\check{\theta}$. Therefore, the standard error of $\check{\theta}$ is given by the square root of

$$\mathrm{var}(\check{\theta})=\frac{K^{-2}\sum_{i=1}^K\check{\Xi}^2}{\left(\partial\check{U}_K/\partial\theta\big|_{\theta=\check{\theta}}\right)^2}.$$

References

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

Barthel N, Geerdens C, Killiches M, Janssen P, Czado C (2018) Vine copula based likelihood estimation of dependence patterns in multivariate event time data. Computational Statistics & Data Analysis 117:109–127

Barthel N, Geerdens C, Czado C, Janssen P (2019) Dependence modeling for recurrent event times subject to right-censoring with d-vine copulas. Biometrics 75(2):439–451

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

Emura T, Nakatochi M, Murotani K, Rondeau V (2017) A joint frailty-copula model between tumour progression and death for meta-analysis. Statistical methods in medical research 26(6):2649–2666

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

Hougaard P (2000) Analysis of multivariate survival data. Springer-Verlag

Joe H (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of multivariate Analysis 94(2):401–419

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

Nelsen RB (2007) An introduction to copulas. Springer Science & Business Media

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

Schneider S, Demarqui FN, Colosimo EA, Mayrink VD (2020) An approach to model clustered survival data with dependent censoring. Biometrical Journal 62(1):157–174

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

Therneau TM (2015) A Package for Survival Analysis in S. R package version 2.38

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

Yan J, et al. (2007) Enjoy the joy of copulas: with a package copula. Journal of Statistical Software 21(4):1–21