KU LEUVEN
FACULTY OF ECONOMICS
AND BUSINESS

Copula based flexible modeling of associations between clustered event times

Geerdens C, Claeskens G, Janssen P.

KBI_1517

Copula based flexible modeling of associations between clustered event times

CANDIDA GEERDENS[1], **GERDA CLAESKENS**[2], **PAUL JANSSEN**[1]

[1] *Center for Statistics, I-BioStat, Universiteit Hasselt, Agoralaan 1,*
B-3590 Diepenbeek, Belgium,
[2] *ORSTAT and Leuven Statistics Research Center, KU Leuven, Naamsestraat 69,*
B-3000 Leuven, Belgium

Candida.Geerdens@uhasselt.be; Gerda.Claeskens@kuleuven.be; Paul.Janssen@uhasselt.be

Abstract

Multivariate survival data are characterized by the presence of correlation between event times within the same cluster. First, we build multi-dimensional copulas with flexible and possibly symmetric dependence structures for such data. In particular, clustered right-censored survival data are modeled using mixtures of max-infinitely divisible bivariate copulas. Second, these copulas are fit by a likelihood approach where the vast amount of copula derivatives present in the likelihood is approximated by finite differences. Third, we formulate conditions for clustered right-censored survival data under which an information criterion for model selection is either weakly consistent or consistent. Several of the familiar selection criteria are included. A set of four-dimensional data on time-to-mastitis is used to demonstrate the developed methodology.

Keywords: clustered data, copulas, model selection, multivariate data, right-censoring, survival data.

1 Introduction

Many data collections consist of observed event times that are grouped in small clusters of equal size. One example is the mastitis study of Laevens et al. (1997) from veterinary medicine, where the impact of the mastitis disease on the milk production and the milk quality of dairy cows is investigated. To this end, information on the time from parturition to infection is collected for the four udder quarters of each cow. The cow is the cluster and the event times of the four udder quarters (quadruple data) are the grouped data; see Section 2 for details. A second example is the trivariate data set on tumorigenesis in Mantel et al. (1977) and Hougaard (2000), where in a litter-matched experiment time to tumor appearance (cancer research) is registered. Modeling bivariate event times is the subject of many papers, often saying that the extension to higher dimensions is straightforward. However, it will become clear that this is an understatement if one is interested in modeling pairwise associations.

Copulas as well as frailty models are used for the statistical analysis of clustered event times. Frailty models extend Cox models and therefore provide a convenient way to model associations between subjects in the same cluster. A nice feature of frailty models is that the cluster size can vary, e.g., if the cluster is a herd, the number of animals within each herd can be different. Copula models are typically used for clusters having small and equal sizes. The data information in a cluster can then be interpreted as an observation from a multivariate random vector with survival function S_{true}.

An additional benefit of copula models is that the association between the components of the multivariate random vector is completely captured by the copula. This is not the case for frailty models, although there are similarities between frailty models and Archimedean copulas (Goethals et al., 2008). Another difference between both model types lies in the way the association between observations in the same cluster is introduced. In frailty models, it is the density of the frailty term that generates (via its Laplace transform) the association between the observations in a cluster, i.e., the association on the event times is modeled in an indirect way through the use of frailties, while in copula models the associations are modeled through the event times themselves. This implies that working with copulas becomes attractive when interest lies in a flexible modeling of the association between the components of clustered data.

Next, the main question is how to select a copula that describes the data well. For this reason we state, for multivariate right-censored data, conditions under which a model selection criterion that takes the form of a general penalized fitting criterion guarantees the selection of the model that is the best in terms of Kullback-Leibler discrepancy (Proposition 1 – weak consistency) or to select the most parsimonious model among several models achieving a similar small Kullback-Leibler discrepancy (Proposition 2 – parsimony). The familiar Akaike and Bayesian information criteria are of this form.

In Section 3 we show that the class containing the exchangeable, fully nested and partially nested Archimedean copulas (EAC, FNAC and PNAC) (Berg and Aas, 2009; Savu and Trede, 2006; Okhrin et al., 2013a,b), even though it contains many different copulas, is too restrictive to describe general association patterns and we explain why Joe-Hu copulas (Joe and Hu, 1996) allow a more flexible modeling. Following the semiparametric approach of Shih and Louis (1995), appropriate likelihood expressions for right-censored quadruple data are given in Section 4. These expressions are key ingredients of the information criteria used for model selection, properties of which are studied in Section 5. The four-variate mastitis data set is analyzed in Section 6. We give a generic R-program to fit multivariate copula models to right-censored event time data, the code is contained in the Web Appendix. The numerical performance of the developed R code is investigated by means of a small simulation study, see Section 7.

2 Motivating example: the mastitis data

The mastitis data that we consider contains information about the time to infection by certain bacteria in each of the four udder quarters of a primiparous cow, i.e., a cow with only one calving. The aim is to examine the association pattern of the infection times in order to improve the knowledge about the disease as well as to help veterinarians in taking preventive

Table 1: Censoring patterns of the mastitis data.

number of censored event times in a quadruple	0	1	2	3	4
number of cows	73	49	36	40	209

and/or curing measures. For the 407 included cows, the available data consist of the cow identification number, the minimum of the times to infection and the censoring times (both in days) for each udder quarter as well as the corresponding event indicators, e.g., for the first and last cow the data information is given by $\{1, (67, 67, 119, 67), (1, 1, 1, 1)\}$ respectively $\{407, (279, 279, 279, 263), (0, 0, 0, 1)\}$. Censoring occurs at the level of the udder quarters and it is univariate in the sense that the same censoring time applies to all udder quarters of an individual cow. Information on the censoring patterns is summarized in Table 1; censoring is present in 66.15% of the observations.

Shared frailty models (Duchateau and Janssen, 2008) and (exchangeable) Archimedean copulas (Massonnet et al., 2009) have been used to analyze the mastitis data. Both modeling strategies assume that the correlations between pairs of udder quarters are all the same (Figure 1(d)). A more general pairwise association pattern (Figure 1(f)) can be obtained via correlated frailty models (Wienke, 2011). The three different line types in Figure 1(f) represent three different dependencies; in Figure 1(d) all dependencies are the same. In Section 6 we show that an even more flexible copula modeling is needed to capture the association structure present in the mastitis data.

3 Flexible copula models

With the data example in mind, we further use quadruple data. Consider a random vector $(T_1, \ldots, T_4)$ with true joint survival function S_{true} and true marginal survival functions $S_{1,\text{true}}$, $\ldots$, $S_{4,\text{true}}$. According to Sklar (1959) there exists a four-variate distribution function C_{true} on $[0,1]^4$ with uniform marginals, called the copula corresponding to S_{true}, such that

$$S_{\text{true}}(t_1, \ldots, t_4) = C_{\text{true}}(S_{1,\text{true}}(t_1), \ldots, S_{4,\text{true}}(t_4)).$$

Throughout this paper it is assumed that the marginal distributions are continuous, which implies that C_{true} is unique and that

$$C_{\text{true}}(u_1, \ldots, u_4) = S_{\text{true}}(S_{1,\text{true}}^{-1}(u_1), \ldots, S_{4,\text{true}}^{-1}(u_4)),$$

with $S_{1,\text{true}}^{-1}$, $\ldots$, $S_{4,\text{true}}^{-1}$ the quantile functions. See Nelsen (2006) for a detailed discussion on copulas. Copulas can also be used to model binary and count data, see Nikoloulopoulos and Karlis (2008, 2010).

In this paper, we treat the joint survival function as a semiparametric copula-based multivariate model, i.e., a model with two sets of parameters: the unknown marginal survival functions

$S_{1,\text{true}}, \ldots, S_{4,\text{true}}$ (the nonparametric part) and the copula C_{true} with true unknown parameter ζ_{true} (the parametric part). In practice we estimate the margins by their Kaplan-Meier counterpart and select a copula from a set of parametric copula families.

Commonly used copulas are the exchangeable and the nested Archimedean ones (Savu and Trede, 2006; Berg and Aas, 2009; Okhrin et al., 2013a,b), see Section 3.1. We show that these copulas are too restrictive to model general association patterns and that the Joe-Hu copula family (Joe and Hu, 1996), see Section 3.2, provides the needed flexibility.

3.1 Exchangeable and nested Archimedean copulas

With ψ a Laplace transform, a four-variate exchangeable Archimedean copula (EAC) is defined as

$$C(u_1, \ldots, u_4) = \psi\left(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \psi^{-1}(u_3) + \psi^{-1}(u_4)\right),$$

and is therefore completely determined by the choice of ψ. The latter implies a restrictive dependence structure; indeed since only ψ can be specified, all bivariate marginal copulas are exactly the same.

Fully and partially nested Archimedean copulas (FNAC, resp. PNAC) are extensions of EAC's allowing more flexible association patterns. For four-dimensional data, the association patterns are given in Figure 1. The corresponding copula expressions are

$$C(u_1, \ldots, u_4) = \psi_3\left(\psi_3^{-1}(u_4) + \psi_3^{-1}\left[\psi_2\left(\psi_2^{-1}(u_3) + \psi_2^{-1}\left[\psi_1\left(\psi_1^{-1}(u_2) + \psi_1^{-1}(u_1)\right)\right]\right)\right]\right), \quad (1)$$

respectively,

$$C(u_1, \ldots, u_4) = \psi_3\left(\psi_3^{-1}\left[\psi_1\left(\psi_1^{-1}(u_1) + \psi_1^{-1}(u_2)\right)\right] + \psi_3^{-1}\left[\psi_2\left(\psi_2^{-1}(u_3) + \psi_2^{-1}(u_4)\right)\right]\right) \quad (2)$$

where ψ_i $(i = 1, 2, 3)$ are Laplace transforms. From the formulas it follows that FNAC and PNAC allow the free specification of *only* three out of the six bivariate margins. The specification of the three remaining bivariate margins is implied by the chosen copula structure. Note that FNAC is not based on the geometry (front versus rear, or left versus right) of the data. Further, to be a valid copula, each combination of Laplace transforms within a NAC, as given by $\psi_i^{-1} \circ \psi_j, i \neq j \in \{1, 2, 3\}$, needs to satisfy the complete monotonicity condition (Nelsen, 2006). For two Laplace transforms stemming from the same family (e.g. Clayton) the latter is equivalent to claiming that the degree of dependence, as expressed by the bivariate copula parameters, decreases with the level of nesting. However, the mixing of diverse Laplace transforms needs to be handled with more care (e.g. the combination of a Clayton with a Gumbel Laplace transform always leads to an improper copula) (Joe, 1997; Hofert, 2008).

3.2 Joe-Hu copulas

A copula family that is much more flexible than the one of exchangeable and nested Archimedean copulas is the Joe-Hu family (Joe and Hu, 1996), which is constructed as follows. Let K_{ij},

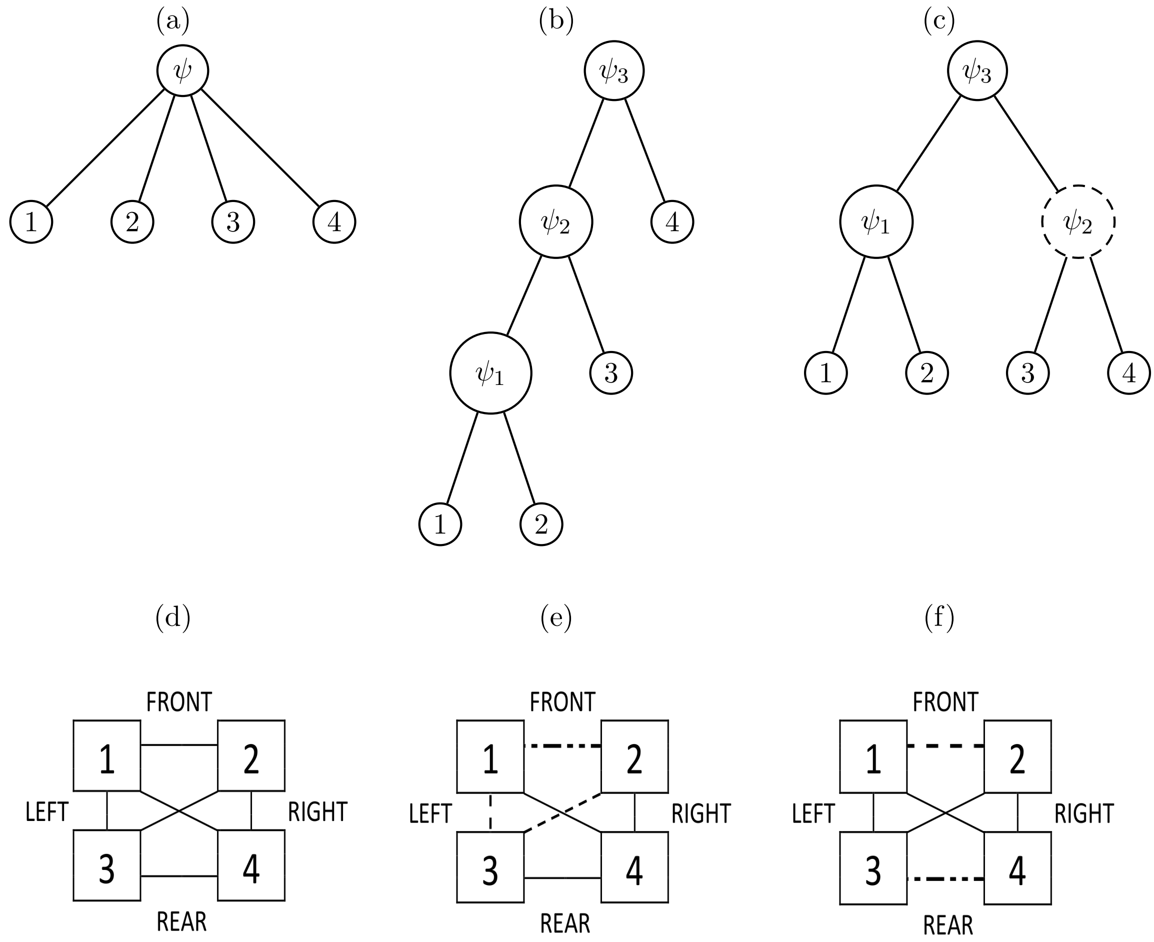

Figure 1: With θ the Laplace parameter of ψ and θ_j the Laplace parameter of ψ_j, $j = 1, 2, 3$, EAC, FNAC and PNAC generate the following association patterns with corresponding tree structures (a larger node for ψ_j corresponds to a larger value of the parameter θ_j): (a) and (d) EAC: all pairs have the same dependence parameter θ; (b) and (e) FNAC: pair (1,2) has association parameter θ_1, pairs (1,3) and (2,3) have association parameter θ_2 and all other pairs have association parameter θ_3 (with $\theta_1 \geq \theta_2 \geq \theta_3$); (c) and (f) PNAC: pair (1,2), resp. pair (3,4), has association parameter θ_1, resp. θ_2, and all other pairs have association parameter θ_3 (with $\theta_1 \geq \theta_3, \theta_2 \geq \theta_3$).

$1 \leq i < j \leq 4$, be bivariate copulas that are max-id, i.e., K_{ij}^{γ} is a distribution function for all $\gamma > 0$. Further, let $H_1,\ldots,H_4$ be univariate cdf's on $[0, 1]$ and let M be the distribution of a positive random variable α whose Laplace transform is defined by ψ. Joe and Hu (1996) define a four-variate copula by considering the following mixture:

$$\begin{aligned}
&\int_0^\infty \prod_{1\le i<j\le 4} K_{ij}^{\alpha}(H_i(u_i),H_j(u_j)) \prod_{i=1}^{4} H_i^{\nu_i\alpha}(u_i)\ dM(\alpha)\\
&= \psi\big(-\log\big\{\prod_{1\le i<j\le 4} K_{ij}(H_i(u_i),H_j(u_j)) \prod_{i=1}^{4} H_i^{\nu_i}(u_i)\big\}\big)\\
&= \psi\big(-\sum_{1\le i<j\le 4} \log K_{ij}(H_i(u_i),H_j(u_j)) - \sum_{i=1}^{4} \nu_i \log H_i(u_i)\big),
\end{aligned}$$

where the ν_i's are chosen fixed constants satisfying $\nu_i > -3$. The ν_i's are usually nonnegative, but they can be negative if some of the K_{ij} correspond to an independence copula.

To be a copula, the margins need to be uniform on $[0,1]$. This is achieved by choosing $H_i(u_i)$ to be $\exp\{-p_i\psi^{-1}(u_i)\}$ with $p_i = (\nu_i+3)^{-1}$, $i = 1,\ldots,4$. One then obtains the copula

$$\begin{aligned}
&C(u_1,\ldots,u_4)\\
&= \psi\big(-\sum_{1\le i<j\le 4} \log K_{ij}(\exp\{-p_i\psi^{-1}(u_i)\},\exp\{-p_j\psi^{-1}(u_j)\}) + \sum_{i=1}^{4} \nu_i p_i \psi^{-1}(u_i)\big). \quad (3)
\end{aligned}$$

The presence of the ν_i's ensures that the above family of multivariate copulas is closed under margins and that the (i,j)-bivariate marginal copula is given by

$$\begin{aligned}
C(u_i,u_j) = \psi\big(&-\log K_{ij}(\exp\{-p_i\psi^{-1}(u_i)\},\exp\{-p_j\psi^{-1}(u_j)\})\\
&+(\nu_i+2)p_i\psi^{-1}(u_i) + (\nu_j+2)p_j\psi^{-1}(u_j)\big). \quad (4)
\end{aligned}$$

See Joe and Hu (1996) for technical details where it is further shown that $C(u_i,u_j)$ is more concordant (or more positive quadrant dependent) than its Archimedean counterpart, see the section on concordance in Joe (1993).

The dependence structure in (3) is completely determined by the choice of the K_{ij}, $1 \le i < j \le 4$, as well as by the Laplace transform ψ. The Laplace transform determines a minimal level of overall dependence, while the copulas K_{ij} allow a fine-tuning of the dependence for each of the six bivariate margins. Examples of popular Archimedean copulas that are max-id and therefore can be used as building blocks K_{ij} are listed in Table 2. These copulas have different dependence properties, e.g., a Clayton copula is lower tail dependent, a Gumbel and a Joe copula have upper tail dependence while a Frank copula exhibits no tail dependence. Using (a combination of) them in (3) thus allows the construction of copulas with flexible dependence patterns. Here, a copula is used to model the joint survival function of event times. Therefore, a copula with lower tail dependence models the association between late event times, whereas a copula with upper tail dependence captures the association between events that occur early in time. We refer to the Web Appendix for both a description and a visualization of the dependence properties of some bivariate Joe-Hu copulas as well as an illustration of the impact of the choice of ν_i on the modeled association. Conditions to check the max-id assumption can be found in Section 4 of Joe and Hu (1996).

Table 2: Choices of bivariate Archimedean max-id copulas and their Laplace transforms.

Family	$K(u,v)$	$\psi(s)$	$\eta \in$
Clayton	$(u^{-\eta}+v^{-\eta}-1)^{-1/\eta}$	$(1+\eta s)^{-1/\eta}$	$]0,\infty[$
Gumbel	$e^{-\{(-\ln u)^{\eta}+(-\ln v)^{\eta}\}^{1/\eta}}$	$e^{-s^{1/\eta}}$	$[1,\infty[$
Frank	$-\frac{1}{\eta}\ln\left\{1+\frac{(e^{-\eta u}-1)(e^{-\eta v}-1)}{e^{-\eta}-1}\right\}$	$-\frac{1}{\eta}\ln\{1-(1-e^{-\eta})e^{-s}\}$	$]0,\infty[$
Joe	$1-\{(1-u)^{\eta}+(1-v)^{\eta}-(1-u)^{\eta}(1-v)^{\eta}\}^{1/\eta}$	$1-(1-e^{-s})^{1/\eta}$	$[1,\infty[$

By taking $K_{ij}(u_i,u_j)=u_iu_j$, i.e., the independence copula, for all pairs (i,j), $i\neq j\in\{1,\ldots,4\}$, the copula in (3) is Archimedean with Laplace transform ψ. Furthermore, by appropriate choices of the K_{ij}, $1\leq i<j\leq 4$ with $K_{ij}(u_i,u_j)\neq u_iu_j$ for some pairs (i,j) it is possible to create a dependence structure that is the same as the one generated by a nested Archimedean copula and this without the required modeling restrictions of the latter e.g. the association structure in Figure 1(e) can be obtained by taking $K_{12}\neq K_{13}=K_{23}\neq K_{14}=K_{24}=K_{34}$, while the correlation pattern in Figure 1(f) can be constructed by setting $K_{12}\neq K_{34}\neq K_{13}=K_{23}=K_{14}=K_{24}$. Exchangeable and nested Archimedean copulas are thus, in that sense, a subclass of the Joe-Hu family.

4 Maximum quasi-likelihood estimation

Let Y_{sr} denote the observed time for observation r ($r=1,\ldots,4$) in cluster s ($s=1,\ldots,S$). That is, $Y_{sr}=\min(T_{sr},C_{sr})$ with T_{sr} the event time and C_{sr} the censoring time. We assume that $T_s=(T_{s1},\ldots,T_{s4})$ and $C_s=(C_{s1},\ldots,C_{s4})$ are independent random vectors ($s=1,\ldots,S$). The actual data are denoted as $y_{sr}=\min(t_{sr},c_{sr})$ and $\delta_{sr}=I(t_{sr}\leq c_{sr})$ is the value of the corresponding event indicator. Let u_{sr} denote the value of the true r-th marginal survival function evaluated at y_{sr}, i.e., $u_{sr}=S_{r,\text{true}}(y_{sr})$.

Consider a collection $\mathcal{M}$ of D copula families M_d ($d=1,\ldots,D$) consisting of EAC, FNAC and PNAC as well as Joe-Hu families:

$$\mathcal{M}=\bigcup_{d=1}^{D}M_d$$

with $M_d=\{C_d(u_1,\ldots,u_4;\zeta_d):\zeta_d\in A_d\subset\mathbb{R}^{p_d}\}$. Here, A_d is the parameter space, a subset of the p_d-dimensional Euclidean space. The parameter ζ_d contains, for each single family, all specific parametric characteristics: the Laplace transform ψ_d (EAC, Joe-Hu), the Laplace transforms $\psi_{d,i}$ (FNAC, PNAC) ($i=1,2,3$) and the bivariate max-id copulas $K_{d,ij}$ (Joe-Hu) ($1\leq i<j\leq 4$). To keep the notation simple, we write C_d to represent a general member of the copula family M_d.

Further, with:

$$\Delta_s = \prod_{r=1}^{4}(1-\delta_{sr}); \quad \Delta_s(p) = \delta_{sp}\prod_{r=1;r\neq p}^{4}(1-\delta_{sr}); \quad \Delta_s(1,2,3,4) \;=\; \prod_{r=1}^{4}\delta_{sr}$$

$$\Delta_s(p,q) = \delta_{sp}\delta_{sq}\prod_{r=1;r\neq p,q}^{4}(1-\delta_{sr}) \quad \text{for } p\neq q$$

$$\Delta_s(p,q,v) = \delta_{sp}\delta_{sq}\delta_{sv}(1-\delta_{sw}) \quad \text{for } w\neq p,q,v \text{ and } p\neq q\neq v,$$

and

$$\begin{aligned} & l_{s,d}(u_{s1},\ldots,u_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta_d) \\ = \;& \Delta_s \log\{C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)\} \\ & +\sum_{p=1}^{4}\left[\Delta_s(p)\log\left\{\frac{\partial C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{sp}}\right\}\right] \\ & +\sum_{p\neq q}\left[\Delta_s(p,q)\log\left\{\frac{\partial^2 C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{sp}\partial u_{sq}}\right\}\right] \\ & +\sum_{p\neq q\neq v}\left[\Delta_s(p,q,v)\log\left\{\frac{\partial^3 C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{sp}\partial u_{sq}\partial u_{sv}}\right\}\right] \\ & +\Delta_s(1,2,3,4)\log\left\{\frac{\partial^4 C_d(u_{s1},u_{s2},u_{s3},u_{s4};\zeta_d)}{\partial u_{s1}\partial u_{s2}\partial u_{s3}\partial u_{s4}}\right\}, \end{aligned} \tag{5}$$

the loglikelihood of copula C_d for right-censored data is given by (Shih and Louis, 1995; Massonnet et al., 2009)

$$\log L_{S,d}(\zeta_d) = \sum_{s=1}^{S} l_{s,d}(u_{s1},\ldots,u_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta_d).$$

Note that the presence of censoring leads to a non-classical form of the likelihood. From this loglikelihood ζ_d can be estimated if $S_{r,\text{true}}$, $r=1,\ldots,4$ are known. However, in practice the latter are typically unknown and the two-step estimator proposed by Shih and Louis (1995) is applied. First, we construct the Kaplan-Meier estimators $\widetilde{S}_{r,\text{true}}$ of the unknown marginal survival functions $S_{r,\text{true}}$ and obtain $\tilde{u}_{sr} = \widetilde{S}_{r,\text{true}}(y_{sr})$ ($r=1,\ldots,4$, $s=1,\ldots,S$). Then we maximize the quasi-loglikelihood

$$\log \widetilde{L}_{S,d}(\zeta_d) = \sum_{s=1}^{S} l_{s,d}(\tilde{u}_{s1},\ldots,\tilde{u}_{s4},\delta_{s1},\ldots,\delta_{s4};\zeta_d),$$

to obtain the maximum quasi-likelihood estimator

$$\hat{\zeta}_{S,d} = \arg\max_{\zeta_d\in A_d}\log\tilde{L}_{S,d}(\zeta_d).$$

We implemented the optimization of the quasi-loglikelihood in a generic R-program which is discussed in Section 7.

5 Model selection

The collection of copulas from which we select one is $\mathcal{M} = \bigcup_{d=1}^{D} M_d$. To derive a model selection criterion, we work under a general misspecification setting, i.e., the true copula C_{true} does not need to be in $\mathcal{M}$. Consequently, we need a selection criterion that gives us the 'best possible' copula in $\mathcal{M}$ or the copula in $\mathcal{M}$ that is the 'closest' we can get to the true copula C_{true} (Sin and White, 1996; Claeskens and Hjort, 2008).

The information criterion we use for model selection is the penalized quasi-loglikelihood

$$\text{IC}_S(M_d) = -2 \log \widetilde{L}_{S,d}(\widehat{\zeta}_{S,d}) + \text{pen}(S, d). \tag{6}$$

Common examples of the penalty term are $\text{pen}(S, d) = 2p_d$, in which case $\text{IC}_S(M_d)$ is the Akaike information criterion, AIC (Akaike, 1973) and $\text{pen}(S, d) = p_d \log(S)$, resulting in a version of the Bayesian information criterion, BIC (Schwarz, 1978). The smaller the value of $\text{IC}_S(M_d)$, the better the model M_d is in terms of Kullback-Leibler discrepancy

$$\frac{1}{S} E_{\text{true}} \left[\log L_{\text{true}}(\zeta_{\text{true}}) - \log L_{S,d}(\zeta_d) \right],$$

with E_{true} the expectation with respect to the true unknown copula C_{true} and $\log L_{\text{true}}$ the version of $\log L_{S,d}$ using C_{true} instead of C_d.

Propositions 1 and 2 show that, based on the above selection criterion, the best model is chosen in a consistent way. Since we work under a general misspecification setting, the pseudo-true parameter values $\zeta^*_{S,d}$ $(d = 1, \ldots, D)$ are needed where

$$\zeta^*_{S,d} = \arg \max_{\zeta_d \in A_d} S^{-1} \sum_{s=1}^{S} E_{\text{true}} \left[l_{s,d}(U_1, \ldots, U_4, \delta_1, \ldots, \delta_4; \zeta_d) \right].$$

The pseudo-true value is the value that minimizes the average Kullback-Leibler information criterion (KLIC), i.e., it is the value for which the loglikelihood using the dth model is as close as possible to the true (unknown) loglikelihood. For the considered data generating process (right censored data with T_s and C_s independent random vectors) we have that

$$\zeta^*_{S,d} = \zeta^*_d = \arg \max_{\zeta_d \in A_d} E_{\text{true}} \left[l_d(U_1, \ldots, U_4, \delta_1, \ldots, \delta_4; \zeta_d) \right].$$

For the remainder of this section we prefer to present the model selection procedure in its general form.

First, suppose that there is exactly one model M_{d_0} in the set of considered models $\mathcal{M}$, which reaches the smallest Kullback-Leibler discrepancy. Further, assume that the conditions C1–C5 of Chen et al. (2010) needed to ensure the convergence of the estimator $\widehat{\zeta}_{S,d}$ to the pseudo-true value $\zeta^*_{S,d}$ hold. The next proposition states the necessary condition on the penalty such that information criterion (6) is weakly consistent, i.e. such that it selects model M_{d_0} with probability tending to one. We follow the terminology as introduced by Sin and White (1996) where weak consistency makes use of a weak law of large numbers, while strong consistency uses a strong law of large numbers.

Proposition 1 (Weak consistency). *Suppose that there is a unique model M_{d_0} in the set of considered models $\mathcal{M}$ which reaches the smallest Kullback-Leibler discrepancy, thus,*

$$\liminf_{S\to\infty} \min_{d\neq d_0} \frac{1}{S} E_{\text{true}}[\log L_{S,d_0}(\zeta^*_{S,d_0}) - \log L_{S,d}(\zeta^*_{S,d})] > 0.$$

Define $\Delta IC_S(d_0, d) = IC_S(M_{d_0}) - IC_S(M_d)$. If for all $d = 1, \ldots, D$, pen(S, d_0) − pen$(S, d) = o_p(S)$, and conditions C1–C5 of Chen et al. (2010) are satisfied, then weak consistency holds,

$$\lim_{S\to\infty} P(\max_{d\neq d_0} \Delta IC_S(d_0, d) < 0) = 1.$$

Proof. Under conditions C1–C5 of Chen et al. (2010), we get from their proposition 3.1(b), for $d = 1, \ldots, D$

$$\frac{1}{S} \log \widetilde{L}_{S,d}(\hat{\zeta}_{S,d}) = \frac{1}{S} E_{\text{true}}[\log L_{S,d}(\zeta^*_{S,d})] + o_P(1).$$

Therefore,

$$\Delta \text{IC}_S(d_0, d) = -2(E_{\text{true}}[\log L_{S,d_0}(\zeta^*_{S,d_0})] - E_{\text{true}}[\log L_{S,d}(\zeta^*_{S,d})]) + o_P(S) + \text{pen}(S, d_0) - \text{pen}(S, d).$$

The statement follows from the assumption on the penalty term as in the proof of Proposition 4.2(a) of Sin and White (1996), though now for the censored multivariate survival models. □

Clearly the condition on the penalty for weak consistency is satisfied by both AIC and BIC-type penalties.

Next, suppose that there are two or more models that achieve nearly the same small Kullback-Leibler discrepancy value. To this end, define $\mathcal{J}$ and $\widetilde{\mathcal{J}}$ to be sets of such 'good' models and state that M_d, $M_{d'} \in \mathcal{J}$ if and only if

$$\limsup_{S\to\infty} \frac{1}{\sqrt{S}} \big(E_{\text{true}}[\log L_{S,d'}(\zeta^*_{S,d'})] - E_{\text{true}}[\log L_{S,d}(\zeta^*_{S,d})]\big) < \infty, \tag{7}$$

and M_d, $M_{d'} \in \widetilde{\mathcal{J}}$ if and only if

$$\log \widetilde{L}_{S,d'}(\zeta^*_{S,d'}) - \log \widetilde{L}_{S,d}(\zeta^*_{S,d}) = O_P(1). \tag{8}$$

We investigate under which assumptions the information criterion is able to select from the set of 'good' models $\mathcal{J}$ or $\widetilde{\mathcal{J}}$ (assumed to be non-empty) a model which is the most parsimonious, in the sense of having the smallest number of parameters. Note that there might be more than one parsimonious model. Define $\mathcal{J}_0 \subset \mathcal{J}$ the subset of $\mathcal{J}$ with the most parsimonious models, i.e., $\mathcal{J}_0 = \{M_{d_0} \in \mathcal{J} : p_{d_0} = \min\{p_d : M_d \in \mathcal{J}\}\}$ and $\widetilde{\mathcal{J}}_0 \subset \widetilde{\mathcal{J}}$ the subset of $\widetilde{\mathcal{J}}$ with the most parsimonious models, i.e., $\widetilde{\mathcal{J}}_0 = \{M_{d_0} \in \widetilde{\mathcal{J}} : p_{d_0} = \min\{p_d : M_d \in \widetilde{\mathcal{J}}\}\}$.

Since the copula models are not necessarily nested, we use the full set of assumptions as in Proposition 4.2 of Chen et al. (2010), that is, their assumptions C1–C5, A1–A4 hold and in addition a Lindeberg condition holds on the likelihood components minus their expected values at the pseudo-true values. Conditions A1–A4 are needed to ensure the asymptotic normality of the estimator $\widehat{\zeta}_{S,d}$.

Proposition 2 (Parsimony)**.** *Assume the condition as in Proposition* 4.2 *of Chen et al. (2010).*

(a) With $\mathcal{J}$ defined in (7), assume that for all $M_{d_0} \in \mathcal{J}_0$, for all $M_d \in \mathcal{J}\backslash\mathcal{J}_0$, the penalty is such that $P(\{pen(S,d) - pen(S,d_0)\}/\sqrt{S} \to \infty) = 1$. Then, with probability tending to one a most parsimonious model will be selected, that is,

$$\lim_{S\to\infty} P(\max_{M_d \notin \mathcal{J}_0} \Delta IC_S(d_0,d) < 0) = 1.$$

(b) With $\widetilde{\mathcal{J}}$ defined in (8), assume that for all $M_{d_0} \in \widetilde{\mathcal{J}}_0$, for all $M_d \in \widetilde{\mathcal{J}}\backslash\widetilde{\mathcal{J}}_0$: $P(pen(S,d) - pen(S,d_0) \to \infty) = 1$. Then, with probability tending to one a most parsimonious model will be selected, that is,

$$\lim_{S\to\infty} P(\max_{M_d \notin \widetilde{\mathcal{J}}_0} \Delta IC_S(d_0,d) < 0) = 1.$$

Proof. Using the set of assumptions (a) we start from Proposition 4.2 of Chen et al. (2010) to obtain that

$$\Delta \mathrm{IC}_S(d_0,d) = -2(E_{\mathrm{true}}[\log L_{S,d_0}(\zeta^*_{S,d_0})] - E_{\mathrm{true}}[\log L_{S,d}(\zeta^*_{S,d})]) + O_P(\sqrt{S}) + \mathrm{pen}(S,d_0) - \mathrm{pen}(S,d).$$

Using the set of assumptions (b) and the proof of Proposition 4.2 of Chen et al. (2010) we arrive at

$$\Delta \mathrm{IC}_S(d_0,d) = -2(\log \tilde{L}_{S,d_0}(\zeta^*_{S,d_0}) - \log \tilde{L}_{S,d}(\zeta^*_{S,d})) + O_P(1) + \mathrm{pen}(S,d_0) - \mathrm{pen}(S,d).$$

The proof now proceeds as in Sin and White (1996), Proposition 4.2 (b) and (c). ☐

Proposition 2 yields the well-known difference in asymptotic behavior for criteria of the type AIC and BIC. For a BIC-type penalty, this form of parsimony in selection holds true, provided the other conditions of the proposition hold. One drawback might be that for sample size dependent penalties it holds that the larger the sample size, the larger the penalty, and as a consequence that simpler models tends to be selected for larger samples, which is often counter-intuitive. Since AIC's penalty does not depend on the sample size, it does not satisfy either penalty condition for consistency. One implication of this is that if there indeed are several models with similar small Kullback-Leibler discrepancy, AIC does not necessarily select amongst these models the one with the smallest number of parameters. This overfitting property of AIC makes statements regarding efficiency of the selection possible. Consistency and efficiency cannot be true together (Yang, 2005), implying that consistent criteria cannot be efficient and vice versa.

Sin and White (1996) also obtain conditions under which they prove strong consistency of this type of information criterion, where the selection takes place almost surely. Those statements hold by using the law of the iterated logarithm. In order for similar statements to be phrased for the multivariate semiparametric copula models, one would need to study the application

Table 3: Joe-Hu dependence structures fitted to the mastitis data

pattern	pairs (i,j) for which $K_{ij}(u,v) \neq uv$	number of parameters
a	(1,2)	2
b	(1,2), (3,4) with $K_{12} \neq K_{34}$	3
c	(1,2), (3,4), (1,3), (2,4) with $K_{12} \neq K_{34} \neq K_{13} = K_{24}$	4
d	(1,2), (3,4), (1,3), (2,4), (2,3) with $K_{12} \neq K_{34} \neq K_{13} = K_{24} \neq K_{23}$	5

of the law of the iterated logarithm under model misspecification as in the framework of Chen et al. (2010). We do not pursue that in this paper.

This study of the asymptotic behavior of an information criterion IC in the style of the AIC is different in spirit from that by Grønneberg and Hjort (2014). Those authors aim to construct a first order unbiased estimator of the Kullback-Leibler discrepancy between a model and the true copula, and name it the copula information criterion, CIC. The penalty used in their CIC versions adds to the AIC-type penalty of $2p_d$ one or two additional terms. Note that no censoring is allowed in that paper. Since the criterion CIC is of the same type as the general criterion IC in (6), its (weak) consistency could be investigated as in Propositions 1 and 2.

6 Mastitis data revisited

In this section we investigate the association pattern of the infection times in the udder quarters of a cow by fitting a variety of copulas to the data at hand. Based on the bivariate copulas and their corresponding Laplace transforms in Table 2, we construct 4 exchangeable, 4 fully and 4 partially nested Archimedean copulas with an implied dependence structure as visualized in Figure 1 as well as 64 Joe-Hu copulas obtained by combining diverse elements from Table 2. Here we set $\nu_i = 0$, $i = 1, \ldots, 4$. Further, within a specific dependency pattern all $K_{ij}(u,v)$ not equal to an independence copula are chosen to stem from the same copula family. By doing so, the diverse four-dimensional Joe-Hu models are nested within each other in the sense that the four Joe-Hu dependence structures depicted in Figure 2 exhibit an increasing flexibility in the way the association between pairs is modeled. It is possible to combine $K_{ij}(u,v)$'s from different copula families, but since this would lead to a large number of copulas to be considered we do not apply this option here. To obtain the patterns in Figure 2, we need to specify for which pairs of udder quarters an extra dependence is added on top of the dependence captured by the Laplace transform ψ, i.e., we need to specify for which pairs (i,j) we have $K_{ij}(u,v) \neq I(u,v) = uv$ (I is the independence copula). The concrete situation is summarized in Table 3. $K_{14}(u,v) = I(u,v)$ in every pattern, this choice is justified by the fit of more general four-variate Joe-Hu models leading to parameter values for $K_{14}(u,v)$ near independence. Note that the dependence pattern in Figure 2(b) is the same as the one of PNAC in Figure 1(c). However, the copula underlying the pattern in Figure 2(b)

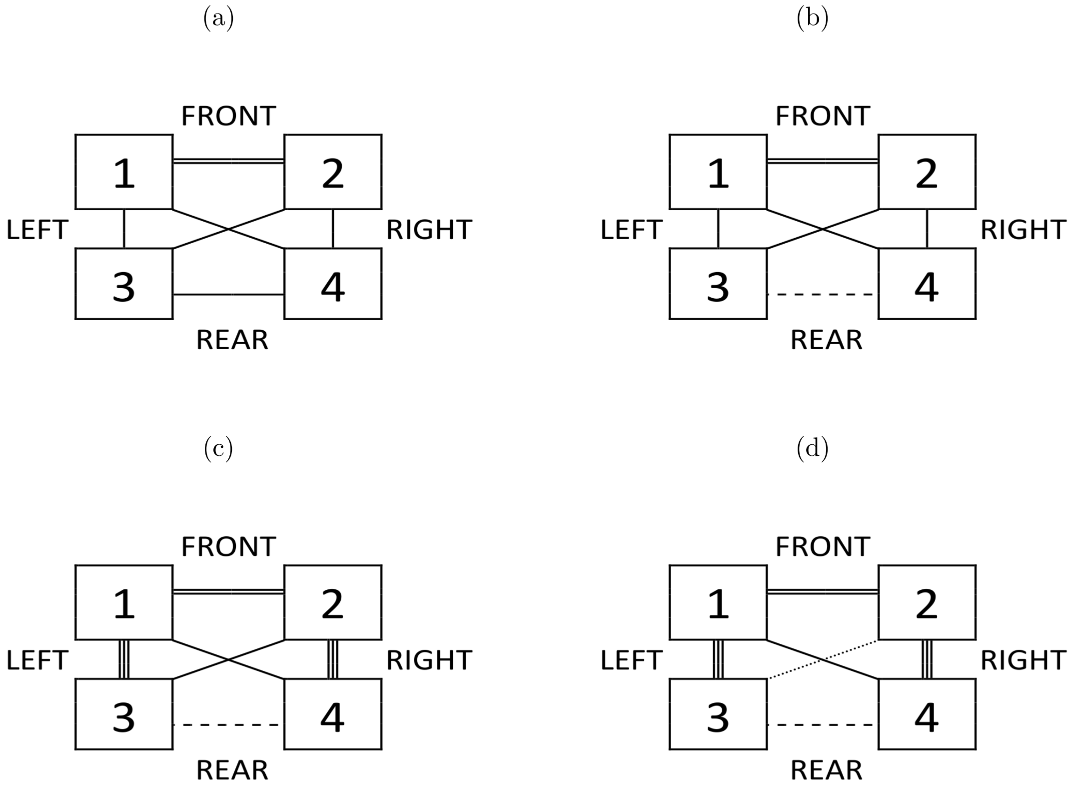

Figure 2: Considered Joe-Hu structures for the mastitis data - different line types represent different dependencies.

allows a more flexible mixing of different types of copulas (see also Section 3.1).

The obtained AIC and BIC values are listed in Tables 4–5, with the three best models marked in bold. From these tables it can be observed that for the mastitis data the simpler models, i.e., EAC, FNAC and PNAC clearly are insufficient. It is the most elaborate dependence structure, pattern d, that outperforms all other models. The estimates of the Laplace and copula parameters in the best model are listed in the left part of Table 6. The corresponding values of the lower and upper tail dependencies are displayed in the right part of Table 6. Note that the tail dependencies for each udder pair (i, j) are calculated using both the estimated parameter of the Laplace transform (θ^C) as well as the estimated value of the parameter θ^J_{ij} of K_{ij} ($i < j \in \{1, 2, 3, 4\}$) (see the Web Appendix for a detailed discussion). To obtain standard errors 1000 bootstrap samples are used; the resampling algorithms (parametric bootstrap) are given in the Web Appendix. From Table 6 it follows that the infection times are substantially correlated, i.e., the lower tail dependence equals 0.74 for all udder quarter pairs and the upper tail dependence of the udder quarter pairs ranges from 0 to 0.31. Late event times thus follow a similar association pattern, while for early event times the association changes from pair to pair. No specific symmetries are present.

Table 4: AIC/BIC-values for EAC, FNAC and PNAC.

ψ	EAC	FNAC	PNAC
Clayton	307.38 / 311.38	308.72 / 320.75	305.85 / 317.88
Gumbel	386.75 / 390.76	372.05 / 384.07	358.40 / 370.43
Frank	315.23 / 319.23	312.66 / 324.69	307.72 / 319.74
Joe	440.61 / 444.62	421.09 / 433.11	405.67 / 417.70

We compare the results of the four-variate analysis, listed in Table 6, with those of the bivariate analysis of each udder pair separately. To this end, we apply model (4) to the six possible pairs of udder quarters. We take ψ a Clayton Laplace transform, K_{ij} a bivariate Joe copula and $\nu_i = 0$ ($i < j \in \{1,2,3,4\}$); this setting corresponds to a of the preferred four-variate Joe-Hu copula (except for the (1,4)-margin which is a Clayton copula). The findings are summarized in Table 7. For most udder quarter pairs the results of the bivariate and the four-variate analysis are similar (e.g., the estimated tail dependencies), however, for the crossed udder quarter pairs (1,4) and (2,3) different estimates are obtained. By reducing the four-variate data to bivariate data the overall association between the four udder quarters is not properly taken into account and hence somewhat misleading results are obtained. These findings are in line with the discussion in Hougaard (2000), where it is indicated that bivariate analyzes do not capture all the information available in multivariate data. A further reference illustrating the advantage of using all available data to estimate a quantity of interest is Akritas and Van Keilegom (2003). The bivariate analysis might serve as an initial search for plausible combinations of ψ and K_{ij}, and the estimated copula parameter values can be used as starting values for the optimization of the four-variate copula loglikelihood, but should not be considered as giving the final result. Further, even though a Joe-Hu model can be applied to bivariate data, it should be done with some care. If ψ and K_{ij} are from the same family (e.g., if ψ and K_{ij} are both taken to be Gumbel), then a Joe-Hu copula may not really make sense: why add an extra association via K_{ij} if ψ already models the desired type of dependence? Further, identifiability issues may arise in dimension two, i.e., it may not be possible to retrieve the parameters of ψ and K_{ij} in a unique way (e.g., if ψ and K_{ij} are both set to be Joe and if in (4) $\nu_i = \nu_j = -2$, then the resulting bivariate copula is a Joe copula with parameter equal to the product of the parameters of ψ and K_{ij}). These problems do not arise in dimension at least three.

7 Generic program and simulation study

7.1 R-program

To allow flexible choices for ψ and the K_{ij}'s in the Joe-Hu family, we developed a generic R-program (see the Web Appendix) based on the general copula formula (3). Due to the generality, some numerical care is needed, e.g., evaluation of the exponent of the negative inverse Laplace transform might be tedious due to limited precision. We addressed this issue

Table 5: AIC/BIC-values for Joe-Hu copulas.

ψ	K_{ij}	Pattern a	Pattern b	Pattern c	Pattern d
Clayton	Clayton	283.46 / 291.48	252.54 / 264.56	232.89 / 248.92	228.80 / 248.84
	Gumbel	279.44 / 287.46	254.75 / 266.78	214.56 / 230.59	**199.13 / 219.17**
	Frank	267.85 / 275.87	217.24 / 229.26	202.44 / 218.48	**198.16 / 218.21**
	Joe	280.43 / 288.44	256.34 / 268.36	214.02 / 230.06	**197.99 / 218.04**
Gumbel	Clayton	352.45 / 360.47	311.70 / 323.72	272.61 / 288.64	262.12 / 282.16
	Gumbel	355.27 / 363.29	326.87 / 338.89	271.07 / 287.10	247.56 / 267.60
	Frank	354.63 / 362.65	327.53 / 339.56	289.13 / 305.17	267.22 / 287.26
	Joe	355.85 / 363.86	327.96 / 339.99	271.86 / 287.89	248.54 / 268.58
Frank	Clayton	291.69 / 299.71	261.45 / 273.48	238.21 / 254.25	232.33 / 252.37
	Gumbel	287.43 / 295.45	262.29 / 274.32	221.71 / 237.74	205.12 / 225.17
	Frank	289.48 / 297.50	281.57 / 293.60	272.15 / 288.18	273.58 / 293.62
	Joe	287.48 / 295.50	262.36 / 274.38	221.68 / 237.72	205.03 / 225.07
Joe	Clayton	397.31 / 405.33	349.35 / 361.38	298.41 / 314.45	284.41 / 304.46
	Gumbel	402.36 / 410.37	369.28 / 381.31	301.67 / 317.71	275.06 / 295.10
	Frank	398.69 / 406.71	378.79 / 390.81	329.38 / 345.42	293.10 / 313.14
	Joe	402.89 / 410.91	370.26 / 382.28	302.59 / 318.62	276.21 / 296.26

Table 6: Estimates of the Clayton Laplace parameter (θ^C) and of the bivariate Joe copula parameters (θ^J_{ij}) in the four-variate Joe-Hu copula as well as the corresponding estimated lower and upper tail dependencies for each udder pair; the estimated standard errors se_1 and se_2 are obtained using the bootstrap Algorithms 1 and 2 (Web Appendix).

parameter	estimated parameter (se_1, se_2)	pair of udder quarters	estimated lower tail dependence (se_1, se_2)	estimated upper tail dependence (se_1, se_2)
θ^C	2.31 (0.30, 0.29)	1-4	0.74 (0.03, 0.03)	0
θ^J_{12}	2.62 (0.73, 0.69)	1-2	0.74 (0.03, 0.03)	0.23 (0.03, 0.03)
θ^J_{23}	1.38 (0.23, 0.23)	2-3	0.74 (0.03, 0.03)	0.12 (0.04, 0.04)
θ^J_{34}	4.44 (1.03, 1.05)	3-4	0.74 (0.03, 0.03)	0.28 (0.02, 0.02)
$\theta^J_{13} = \theta^J_{24}$	9.65 (1.35, 1.37)	1-3 and 2-4	0.74 (0.03, 0.03)	0.31 (0.01, 0.01)

by high precision calculations, i.e., instead of using the double precision numbers in R, we use multiple precision floating point numbers (Maechler, 2014). Further, note that the terms in the likelihood expressions in (5) often contain (higher order) partial derivatives. The exact expressions are typically quite cumbersome to obtain, in the program we therefore used finite forward differences as approximations. The simulation study below shows that the generic program works well.

Table 7: Estimates of the Clayton Laplace parameter (θ^C) and of the bivariate Joe copula parameter (θ^J_{ij}) in the bivariate Joe-Hu copula as well as the corresponding estimated lower and upper tail dependencies for each udder pair; the estimated standard errors se_1 and se_2 are obtained using the bootstrap Algorithms 1 and 2 (Web Appendix).

pair of udder quarters	$\widehat{\theta}^C$ (se_1, se_2)	$\widehat{\theta}^J_{ij}$ (se_1, se_2)	estimated lower tail dependence (se_1, se_2)	estimated upper tail dependence (se_1, se_2)
1-2	2.42 (0.45, 0.44)	4.17 (1.14, 1.13)	0.75 (0.05, 0.05)	0.27 (0.03, 0.03)
1-3	2.25 (0.45, 0.45)	9.85 (2.16, 2.31)	0.74 (0.05, 0.05)	0.31 (0.02, 0.02)
1-4	1.50 (0.39, 0.38)	3.36 (0.93, 0.98)	0.63 (0.07, 0.07)	0.26 (0.04, 0.04)
2-3	2.03 (0.45, 0.46)	4.78 (1.31, 1.25)	0.71 (0.06, 0.06)	0.28 (0.03, 0.03)
2-4	2.26 (0.44, 0.45)	8.56 (1.90, 2.07)	0.74 (0.05, 0.05)	0.31 (0.02, 0.02)
3-4	2.15 (0.46, 0.47)	4.19 (1.17, 1.15)	0.72 (0.06, 0.06)	0.27 (0.03, 0.03)

7.2 Simulation study

To evaluate the numerical performance of the generic R-program, we set up a small simulation study. Based on the conditional inversion algorithm described in Embrechts et al. (2003) (see the Web Appendix) we generate 600 datasets, each containing 500 clusters of size 4 from either the Clayton copula with $\theta = 3.19$, the Frank copula with $\theta = 5.76$ or the Joe-Hu copula

$$C(u_1, \dots, u_4) = \left(1 + \left\{\left(u_1^{-\theta} - 1\right)^{\alpha} + \left(u_2^{-\theta} - 1\right)^{\alpha}\right\}^{\frac{1}{\alpha}} + \left\{\left(u_3^{-\theta} - 1\right)^{\alpha} + \left(u_4^{-\theta} - 1\right)^{\alpha}\right\}^{\frac{1}{\alpha}}\right)^{-\frac{1}{\theta}},$$

with $\theta = 2.91$ and $\alpha = 1.17$. This Joe-Hu copula can be constructed by taking $\nu_1 = \ldots = \nu_4 = -2$, ψ a Clayton Laplace transform with parameter θ, $K_{12} = K_{34}$ a bivariate Gumbel copula with parameter α and all other K_{ij} independence copulas ($1 \leq i < j \leq 4$).

To obtain event times t_{sr} ($s = 1, \dots, 500$, $r = 1, \dots, 4$) we take four Weibull margins with scale $\lambda = 0.5$ and shape $\rho = 1.5$. The censoring mechanism is assumed to be univariate, i.e. $c_{sr} = c_s$ ($s = 1, \dots, 500$, $r = 1, \dots, 4$), Weibull with scale and shape given by $\lambda = 0.15, \rho = 1.5$, resp. $\lambda = 0.85, \rho = 1.5$, leading to approximately 23%, resp. 63% censoring. The latter corresponds to the censoring present in the mastitis data. The observed data can then be calculated as $y_{sr} = \min(t_{sr}, c_{sr})$ and $\delta_{sr} = I(t_{sr} \leq c_{sr})$ ($s = 1, \dots, 500$, $r = 1, \dots, 4$).

The results, obtained by applying the generic R-program with a stepsize of either 0.001 or 0.0005 for the finite forward differences, are summarized in Table 8. From this it can be seen that, on average and taking the empirical standard deviation into account, the estimation is on target. The performance is somewhat more accurate for the Frank copula and slightly better for light censored data.

Table 8: Simulation results - mean and empirical standard deviation

copula	censoring	stepsize	$\widehat{\theta}$ (se)	$\widehat{\alpha}$ (se)
Clayton $\theta = 3.19$	23%	0.001	3.094 (0.226)	
		0.0005	3.089 (0.230)	
	63%	0.001	3.041 (0.328)	
		0.0005	3.041 (0.330)	
Frank $\theta = 5.76$	23%	0.001	5.737 (0.266)	
		0.0005	5.727 (0.263)	
	63%	0.001	5.722 (0.376)	
		0.0005	5.723 (0.374)	
Joe-Hu $\theta = 2.91$, $\alpha = 1.17$	23%	0.001	2.813 (0.226)	1.178 (0.033)
		0.0005	2.809 (0.225)	1.180 (0.032)
	63%	0.001	2.770 (0.326)	1.180 (0.043)
		0.0005	2.776 (0.324)	1.182 (0.042)

8 Discussion

In this paper we show that, compared to exchangeable and nested Archimedean copulas (EAC, FNAC, PNAC), the Joe-Hu family allows a flexible way of modeling the association patterns. Given the variety of possible copula constructions we used a model selection criterion for which methodological support is given in Propositions 1 and 2. For right-censored quadruple data, a generic R-program is given to fit the model as well as to perform the model selection.

Vines (Berg and Aas, 2009; Aas et al., 2009) also provide flexible modeling of association patterns. The basic idea is a decomposition of the copula density into a cascade of bivariate conditional densities which then allows a copula specification for each pair of variables. So far, vine theory is only developed for complete data. The extension of the existing vine methodology to right-censored data is a challenging research problem that we are currently investigating.

Acknowledgements

The authors wish to thank Dr. H. Laevens (Catholic University College Sint-Lieven, Sint-Niklaas, Belgium), for permission to use the mastitis data for this research. This work was supported by the IAP Research Network P6/03 of the Belgian State (Belgian Research Policy), by the Fund for Scientific Research Flanders and by KU Leuven grant GOA/12/14. For the simulations we used the infrastructure of the VSC - Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government – department EWI. The authors thank all the reviewers of this paper for the constructive remarks and questions that have led to a clearer presentation.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, 44:182–198.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, pages 267–281. Akadémiai Kiadó, Budapest.

Akritas, M. G. and Van Keilegom, I. (2003). Estimation of bivariate and marginal distributions with censored data. *Journal of the Royal Statistical Society: Series B*, 65:457–471.

Berg, D. and Aas, K. (2009). Models for construction of multivariate dependence: A comparison study. *The European Journal of Finance*, 15:639–659.

Chen, X., Fan, Y., Pouzo, D., and Ying, Z. (2010). Estimation and model selection of semiparametric multivariate survival functions under general censorship. *J. Econometrics*, 157:129–142.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.

Duchateau, L. and Janssen, P. (2008). *The frailty model*. Springer.

Embrechts, P., Lindskog, F., and McNeil, A. J. (2003). Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. Elsevier.

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

Grønneberg, S. and Hjort, N. L. (2014). The copula information criteria. *Scandinavian Journal of Statistics*, 41:436–459.

Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics and Data Analysis*, 52:5163–5174.

Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer.

Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 46:262–282.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall.

Joe, H. and Hu, T. (1996). Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, 57:240–265.

Laevens, H., Deluyker, H., Schukken, Y. H., De Meulemeester, L., Vandermeersch, R., De Meulenaere, E., and De Kruif, A. (1997). Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. *Journal of Dairy Science*, 80:3219–3226.

Maechler, M. (2014). *Rmpfr: R MPFR - Multiple Precision Floating-Point Reliable.* R package version 0.5-6.

Mantel, N., Bohidar, N., and Ciminera, J. (1977). Mantel-Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Research*, 37:3863–3868.

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

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, 27:6393–6406.

Nikoloulopoulos, A. K. and Karlis, D. (2010). Modeling multivariate count data using copulas. *Communications in Statistics - Simulation and Computation*, 39:172–187.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013a). Determining the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, 173:189–204.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013b). Properties of hierarchical Archimedean copulas. *Statistics and Risk Modeling*, 30:21–53.

Savu, C. and Trede, M. (2006). Hierarchical Archimedean copulas. *International Conference on High Frequency Finance, Konstanz, Germany.*

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:1384–1399.

Sin, C. and White, H. (1996). Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, 71:207–225.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l' Intitut de Statistique de l' Université de Paris*, 8:229–231.

Wienke, A. (2011). *Frailty models in survival analysis.* Chapman-Hall.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? *Biometrika*, 92:937–950.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be