Modelling clustered survival data through Archimedean copulas

Leen Prenen

Promotor: Prof. dr. Roel Braekers
Co-Promotor: Prof. dr. Luc Duchateau

Acknowledgements

Looking back at a seven-year chapter in my life, I would like to thank everyone who helped me in any way, big or small, during this endeavour to obtain a PhD.

First and foremost, my gratitude goes out to CenStat, for granting me the opportunity to work in such a stimulating environment. A special word of thanks goes to my supervisors Roel Braekers and Luc Duchateau, for offering me this interesting research topic. Although I had no background in biostatistics whatsoever, you were willing to launch me into the subject by sharing your expertise and insights with me. You never failed to motivate me to work hard and your guidance has brought me where I am today. Roel and Luc, thank you for your never ending positivism and enthusiasm.

To all the professors and fellow teaching assistants with whom I had the honour to collaborate during teaching assignments: it was a pleasure to work with all of you. Teaching has always been a welcome addition to my research activities and I am grateful for the many opportunities I was given to enhance my skills in education.

My colleagues at CenStat may be praised for their work ethic and superhuman intelligence, but above all they are just a bunch of nice people. This has definitely made the past seven years a lot easier. I really enjoyed the pleasant atmosphere and the enlightening conversations in the hallways and during coffee and lunch breaks. In that respect, a big shout-out to my office mates Ruth and Candida, the two most helpful and well-organized people in the world. Special thanks also to Stijn, for hooking me up with the VSC cluster and to Ewoud, for helping out with the R package.

Natuurlijk zijn er naast de werkvloer nog een heel aantal mensen die een woordje van dank verdienen. Ten eerste wil ik mijn ouders, Rik en Yvonne, in de bloemetjes zetten. Dank je voor alle kansen die jullie An en mij hebben gegeven, zonder ons ooit in een bepaalde richting te hebben willen duwen. Er is niks zo bevorderlijk om je vleugels uit te slaan, dan een veilig nest waar je altijd naar kan terugkeren. An en Xander, jullie waren een stichtend voorbeeld voor mij wat belangstelling en engagement in wetenschappelijk onderzoek betreft. Over de vraag wat er nu nuttiger is, fysica of wiskunde, zullen we nog lang mogen discussiëren, maar het is duidelijk dat - net als bij ons - de een niet zonder de ander kan.

Mijn schoonfamilie wil ik bedanken voor de gezellige momenten iedere zondag, en voor de oprechte interesse die jullie tonen in mijn werk. Suzanne, bedankt voor het lekkere eten dat je ons voorschotelt en voor de voedselpakketten die we blijven meekrijgen, ondanks de recente uitbreidingswerken in de familie. Bram, dankjewel om mijn bondgenoot te zijn als enige andere niet-leerkracht in de Govaerts-onderwijsclan.

Naast een familie vol supporters, ben ik gezegend met de liefste vrienden. Terugdenken aan de vele memorabele, hilarische en hartverwarmende momenten met jullie, bezorgt me instant happiness. Aan alle dans-, babbel-, sport-, reis-, BBQ-, etentjes-, wijntjes-, spelletjes-,... vrienden en -vriendinnen: ik ben zo blij dat jullie er zijn. Ik hoop dat we onze tradities nog heel lang mogen voortzetten.

De laatste paragraaf van dit dankwoord gaat naar mijn twee grootste schatten. Koen, bedankt om zo'n liefdevolle en geduldige echtgenoot te zijn. Je bent mijn steun en toeverlaat en een geweldige papa voor Sam. Lieve kleine Sam, al twee jaar lang help jij mij om de dingen beter in hun context te zien. Je betekent meer voor me dan ik ooit zal kunnen uitdrukken.

Diepenbeek, 8 juni 2016

Publications

This dissertation is based on the following publications and reports:

Prenen, L., Braekers, R., Duchateau, L. (2017) Extending the Archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. *Journal of the Royal Statistical Society - series B*. Epub ahead of print. (DOI: 1369–7412/17/79000)

Prenen, L., Braekers, R., Duchateau, L. (2016) Investigating the correlation structure of quadrivariate udder infection times through hierarchical Archimedean copulas. *Biometrics*. Submitted for publication.

Prenen, L., De Troyer, E., Braekers, R., Duchateau, L. (2016) Sunclarco: An R package to model multivariate survival data that are grouped in unbalanced clusters using Archimedean copulas. *Journal of Statistical Software*. To be submitted.

Other publications:

Van de Putte, G., Oben, J., **Prenen, L.**, Schobbens, J.C., Vlasselaer, J., Van Holsbeke, C., Debrock, G., Van Eycken, P., de Jonge, E., (2015) Outcome of epithelial ovarian cancer: Time for strategy trials to resolve the problem of optimal timing of surgery. *International Journal of Gynecological Cancer*,25(6):993-999.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In survival studies, focus is on time to event data. With this type of data, the variable of interest is typically positive-valued and measures the time until something happens, e.g.,

- time to death;
- time to onset or relapse of a disease;
- time to failure of electronic components;
- duration of nursing home stay.

If the event occurred in all individuals, many methods of analysis would be applicable. However, it is usual that at the end of follow-up some of the individuals have not had the event of interest yet, or have dropped out of the study. Thus, their true time to event is unknown and only a lower time bound for the event is observed. This situation is called right censoring. In this thesis, we will deal with non-informative right censoring, meaning that censoring carries no prognostic information about subsequent survival experience; in other words, those who are censored because of loss to follow-up at a given point in time should be as likely to have a subsequent event as those individuals who remain in the study.

Further, survival data are rarely normally distributed, but rather skewed and comprise typically of many early events and relatively few late ones. It is these features of the data that make the methods called survival analysis necessary. Many fine books have been written on the subject, e.g, Klein and Moeschberger (2003); Kleinbaum and Klein (2012). In Section 1.1 we will review some of the basic concepts.

1.1 Basic concepts of univariate survival analysis

Suppose we are studying the survival times of n independent individuals which are possibly right-censored. The observed time for individual $i(i = 1, \ldots, n)$ is $X_i = \min(T_i, C_i)$, where T_i is the event time and C_i is the censoring time. Event times and censoring times are assumed to be independent. In order to know whether the event actually happened, an indicator $\delta_i = I(T_i \leq C_i)$ is introduced, which equals 1 if the event happened and 0 otherwise.

Let $F(t) = P(T \leq t)$ be the cumulative distribution function of T and f the corresponding probability density function. The survival function captures the probability that the individual will survive beyond a specified time

$$S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(x)dx.$$

The hazard function (also known as the failure rate, hazard rate, or force of mortality) is the instantaneous rate of occurrence of the event

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

The numerator of this expression is the conditional probability that the event will occur in the interval $[t, t + \Delta t[$ given that it has not occurred before, and the denominator is the width of the interval. Dividing one by the other we obtain a rate of event occurrence per unit of time. Letting the interval width go to zero in the limit, we obtain an instantaneous rate of occurrence. The conditional probability in the numerator may be written as the ratio of the joint probability that T is in the interval $[t, t+\Delta t[$ and $T \geq t$ to the probability of the condition $T \geq t$. The former may be written as $f(t)\Delta t$ for small Δt, while the latter is $S(t)$ by definition. Dividing by Δt and passing to the limit gives the useful result

$$h(t) = \frac{f(t)}{S(t)},$$

from which we can deduce that

$$h(t) = -\frac{d}{dt} \log S(t).$$

This gives rise to the cumulative hazard function

$$H(t) = \int_0^t h(x)dx = -\log S(t).$$

One may think of $H(t)$ as the sum of the risks you face going from duration 0 to t.

The survival and hazard functions provide alternative but equivalent characterizations of the distribution of T. The distribution of T can be estimated either parametrically, making assumptions on the functional form of the density, or nonparametrically.

In the parametric case, the simplest possible survival distribution is obtained by assuming a constant hazard over time, corresponding to an exponential distribution of T:

$$\begin{aligned} h(t) &= \lambda \\ S(t) &= \exp(-\lambda t) \\ f(t) &= \lambda \exp(-\lambda t). \end{aligned}$$

The exponential distribution is a special case of the more flexible Weibull distribution

$$\begin{aligned} h(t) &= \rho \lambda^\rho t^{\rho-1} \\ S(t) &= \exp(-(\lambda t)^\rho) \\ f(t) &= \rho \lambda^\rho t^{\rho-1} \exp(-(\lambda t)^\rho) \end{aligned}$$

with scale parameter $\lambda > 0$ and shape parameter $\rho > 0$. If $\rho = 1$, this model reduces to the exponential. If $\rho > 1$, then the hazard increases over time. If $\rho < 1$, then the hazard decreases over time.

Another extension of the exponential model is the piecewise exponential model, where the duration is partitioned into L intervals with cutpoints $0 = \tau_0 < \tau_1 < \cdots < \tau_L = \infty$ and the hazard is assumed to be constant within each interval:

$$\begin{aligned} &h(t) = \lambda_l \text{ for } t \in [\tau_{l-1}, \tau_l[\\ &S(t) = \exp\left(-\sum_{j=1}^{l-1} \lambda_j (\tau_j - \tau_{j-1}) - \lambda_l (t - \tau_{l-1}) \right) \text{ for } t \in [\tau_{l-1}, \tau_l[\\ &f(t) = \lambda_l \exp\left(-\sum_{j=1}^{l-1} \lambda_j (\tau_j - \tau_{j-1}) - \lambda_l (t - \tau_{l-1}) \right) \text{ for } t \in [\tau_{l-1}, \tau_l[. \end{aligned}$$

In the following chapters, the Weibull distribution and piecewise exponential distribution will be used to illustrate the parametric modelling approach, because they cover a wide range of distributional shapes due to their flexibility. Other popular choices for the distribution of event times are the Gompertz, generalized gamma, log-logistic and log-normal distribution (Wienke, 2011; Klein and Moeschberger, 2003). All these distributions are equally applicable in the developed methodology.

The classical nonparametric estimator for the survival function in the presence of censoring is the Kaplan-Meier estimator (Kaplan and Meier, 1958)

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

where $t_{(i)} (i = 1, \ldots, m)$ are the ordered event times and d_i the number of events at $t_{(i)}$, while n_i is the number at risk just before $t_{(i)}$. The Kaplan-Meier estimator is a step function which jumps at the observed event times. If the largest observation happens to be a censored case, the estimator will never drop to zero. If no data are censored, the Kaplan-Meier estimator coincides with be the empirical survival function

$$\hat{S}(t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i \geq t),$$

which is the proportion alive at t.

In parametric inference, all of the evidence in a sample relevant to model parameters is contained in the likelihood function. If the lifetimes $T_1, \ldots, T_n$ are governed by a survival function $S(t)$ with associated density $f(t)$, the contribution of an uncensored subject to the likelihood is $f(t_i) = f(x_i)$, because the actual lifetime is observed. If the subject did not experience the event, all we know under non-informative right censoring is that the lifetime T_i exceeds the observed time x_i, so its contribution is $S(x_i) = P(T_i > x_i)$. The likelihood of all n subjects at once is

$$L = \prod_{i=1}^{n} f(x_i)^{\delta_i} S(x_i)^{1-\delta_i}.$$

Up to this point we have been concerned with a homogeneous population, where the lifetimes of all units are governed by the same survival function $S(t)$. The risk of failure may, however, be affected by a set of covariates $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ip})'$, which are possibly time-varying. The accelerated failure time model (AFT) models the logarithm of the (possibly unobserved) event times $T_i (i = 1, \ldots, n)$ using a conventional linear model

$$\log T_i = \boldsymbol{\beta}' \boldsymbol{Z}_i + \epsilon_i$$

where the distribution of log-survival of the ith subject is specified as a simple shift of a baseline distribution represented by the error term ϵ_i. In a model with a single binary covariate Z, this means that the survival function of subjects in group 1 is $S_1(t) = S_0(\exp(-\beta)t)$. In words, the probability that a member of group 1 has not experienced the event at time t is exactly the same as the probability that a member of group 0 will not have experienced the event at time $\exp(-\beta)t$. For $\beta = \log 2$, this would be half the time. Different kinds of parametric models are obtained by assuming different distributions for the error term. If the ϵ_i are normally distributed, the log-normal model for the T_i is obtained. Alternatively, if the ϵ_i have an extreme value distribution with probability density function $f(\epsilon) = \exp(\epsilon - \exp(\epsilon))$, the T_i are exponentially distributed.

Another, widely used approach of modelling covariate effects is the proportional hazards model that was proposed by Cox (1972). As the name suggests, the proportional hazards model focuses directly on the hazard function. The hazard at time t for a subject with covariate information $\boldsymbol{Z}_i$ is

$$h_i(t|\boldsymbol{Z}_i) = h_0(t) \exp \boldsymbol{\beta}' \boldsymbol{Z}_i.$$

In this model, $h_0(t)$ is the baseline hazard function that describes the hazard for individuals with $\boldsymbol{Z}_i = \mathbf{0}$, and $\exp \boldsymbol{\beta}' \boldsymbol{Z}_i$ is the proportionate increase or reduction in hazard, associated with the set of characteristics $\boldsymbol{Z_i}$. The survival function is

$$S(t|\boldsymbol{Z}_i) = S_0(t)^{\exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}$$

where $S_0(t) = \exp(-H_0(t))$ is the baseline survival function. Different assumptions about the baseline survival function (or hazard function) lead to different kinds of proportional hazards models. One can assume a parametric form of the baseline survival function, but Cox (1972) observed that inference about the covariate effects is also possible when there is no assumption at all on the baseline survival (or hazard) function. In Cox's proportional hazards model, estimation of covariate effects is done by maximizing the partial likelihood. In case of data without tied observations, the partial likelihood is given by

$$L(\boldsymbol{\beta}) = \prod_{i:\delta_i=1} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}{\sum_{j:X_j \geq X_i} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_j)}.$$

If event times are observed on a discrete time scale, it is possible that the event times of two or more subjects coincide. In the case of tied observations, it is recommended to use the approximation proposed by Efron (1977).

1.2 Models for multivariate survival data

Multivariate survival data consist of multiple lifetimes which are linked to each other in some sense. In clustered survival data, subjects in the same cluster are assumed to share some characteristic or environment, and are therefore expected to be more similar with respect to the hazard of the event. Examples are

- follow-up of cancer patients treated in different medical centers;
- mortality of twins;
- time to infection in the four udder parts of a dairy cow.

To analyze this type of multivariate survival data, two popular techniques exist that model the association between the individuals: frailty models and copula models.

1.2.1 Frailty models

The frailty approach aims to account for heterogeneity, caused by unmeasured covariates at the cluster level. For example, in a multi-center clinical trial, some hospitals perform better than others, although the study protocol implies the same therapy in all hospitals.

In addition to the observed covariates, a frailty model also accounts for the presence of a latent multiplicative effect on the hazard function. This effect, or frailty, is considered as a realization of a random variable U with a given frailty distribution having unit mean and finite variance:

$$E(U) = 1 \quad \text{and} \quad \mathrm{Var(U)} = \theta.$$

It is the variance of the frailty that represents the heterogeneity and which is eventually estimated. In cases where the frailty is greater than one, subjects experience an increased hazard of failure and are said to be more frail. Frailty models come in different flavours. The *shared* frailty model is used with multivariate survival data where the unobserved frailty is shared among groups of individuals. The *correlated* frailty model was originally developed for the analysis of bivariate failure time data, in which two associated random variables are used to characterize the frailty effect for each pair.

In shared frailty models, the interest lies on the hazard function of an individual, conditional on an unknown frailty term for the cluster containing this individual. In these models, we follow a conditional viewpoint and investigate the influence of different covariates on the hazard function of an individual, given the cluster. In fact, shared frailty models are extensions of the proportional hazards model. The hazard at time t for subject j in cluster i with covariate information $\boldsymbol{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijp})'$ is

$$h_{ij}(t|\boldsymbol{Z}_{ij}) = h_0(t)u_i \exp(\boldsymbol{\beta}'\boldsymbol{Z_{ij}}).$$

The frailty term $u_i (i = 1, \ldots, K)$ for each cluster expresses that we assume that different individuals in the same cluster behave in a similar but unknown manner. We consider this frailty term as a realization of a random variable U with a given frailty distribution and allow it to vary over the different clusters. These unknown frailty terms with their imposed distribution are used to express the association between the different individuals in a cluster. Due to its mathematical convenience, the one-parameter gamma distribution is a popular choice for the distribution of U. The one-parameter gamma density is

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

with $\theta > 0$. Geerdens *et al.* (2013) developed a strategy to verify the goodness of fit of the one-parameter gamma frailty distribution. Other common choices are the positive stable distribution and the inverse Gaussian distribution.

The correlated frailty model is a natural extension of the shared frailty model, allowing the frailties of individuals in a clusters to be correlated but not necessarily shared. This enables that associations are no longer the same across all pairs of individuals within a cluster. The hazard of individual j in cluster i has the form

$$h_{ij}(t|\boldsymbol{Z}_{ij}) = h_0(t)u_{ij} \exp(\boldsymbol{\beta}'\boldsymbol{Z_{ij}}).$$

Correlated frailty models are characterized by the joint distribution of a vector of frailties $(u_{i1}, \ldots, u_{in_i})$. If the frailties are independent, no clustering is present. If all frailties are equal, the shared frailty model is obtained as a special case of the correlated frailty model, with correlation one between the frailties. The flexibility of the correlated frailty model comes with a price. Extensions of the bivariate case to higher dimensions become very complex with increasing cluster size.

The frailty approach is explained in detail in Duchateau and Janssen (2008) and Wienke (2011). To estimate the different parameters in frailty models, we make use of the

conditional viewpoint of these models. Hereby we assume that different individuals within the same cluster are treated as independent of each other, conditionally on the frailty term(s). In the construction of the likelihood function of a frailty model, this assumption is utilized by first looking at the conditional contribution of an individual within a cluster to the likelihood function and afterwards integrating over the frailty distribution. In this way, the frailty model approach has the advantage that it allows that the number of individuals within a cluster may vary over the different clusters. However, the correlated frailty model becomes very complex with increasing cluster sizes. Another major disadvantage of the frailty model is that the marginal survival functions in the frailty model contain the association parameter of the frailty distribution (Goethals *et al.*, 2008). This has led to the correct observation by, e.g., Hougaard (1986, p. 676) that the association parameter in a shared frailty model can be obtained from the marginal survival functions alone. The conditional approach therefore makes it difficult to interpret the frailty parameter, because it does not simply quantify the association between event times. Additionally, overdispersion in the data, as compared to the proposed density function, is required in a frailty model in order to pick up association.

1.2.2 Copula models

Copula models, on the other hand, assume that the joint survival function of the individuals within a cluster is given by a copula function, evaluated in the marginal survival function of each individual (Sklar, 1959):

$$S(t_1, \ldots, t_n) = C(S_1(t_1), \ldots, S_n(t_n)).$$

Here, $S(t_1, \ldots, t_n) = P(T_1 > t_1, \ldots, T_n > t_n)$ is the joint survival function and $S_i(t_i) = P(T_i > t_i), i = 1, \ldots, n$ are the marginal survival functions. It is the copula function C which describes the association between the lifetimes within a cluster. An n-variate copula is a function $C : [0,1]^n \to [0,1]$ satisfying:

1. For every $(u_1, \ldots, u_n)$ in $[0,1]^n$: $C(u_1, \ldots, u_n) = 0$ if $u_i = 0$ for at least one i in $(1, \ldots, n)$.
2. For every $(u_1, \ldots, u_n)$ in $[0,1]^n$: $C(u_1, \ldots, u_n) = u_k$ if all $u_i = 1$ when $i \neq k$.
3. For every hyperrectangle $B = [u_1, v_1] \times \cdots \times [u_n, v_n]$, the C-volume of B is non-negative, i.e.,
$$\Delta_{u_n}^{v_n} \Delta_{u_{n-1}}^{v_{n-1}} \ldots \Delta_{u_2}^{v_2} \Delta_{u_1}^{v_1} C(w_1, \ldots, w_n) \geq 0$$

where, for $i = 1, \ldots, n$,

$$\begin{aligned}\Delta_{u_i}^{v_i} C(w_1, \ldots, w_n) =& C(w_1, \ldots, w_{i-1}, v_i, w_{i+1}, \ldots, w_n) \\ & - C(w_1, \ldots, w_{i-1}, u_i, w_{i+1}, \ldots, w_n).\end{aligned}$$

In this thesis, we focus on parametric copula families $C_{\boldsymbol{\theta}}$ that depend on a finite dimensional vector of parameters $\boldsymbol{\theta}$ (often just a scalar). Estimation of the different parameters in copula models can be done in two stages. In the first stage, the parameters of the marginal survival functions are estimated, and then inserted in the copula function. In the second stage, the parameter (vector) $\boldsymbol{\theta}$ of the copula function is estimated. Thus, both in the model specification and parameter estimation, the parameter(s) describing the association between event times is kept separate from the other parameters. Most reported copula models, however, only use clusters in which the cluster size is small and constant over the different clusters as it is then straightforward to define and estimate the marginal survival functions. For example, Shih and Louis (1995) introduced a copula model for multivariate survival data and provided estimation methods for the unknown parameters in a bivariate setting. Glidden (2000) and Andersen (2005) extended the approach of Shih and Louis (1995) to include covariates into the marginal survival function, but also here the clusters only had size two. Massonnet *et al.* (2009) extended these models further for clusters of size 4 to model the time until infection in the four different quarters of a cow udder. Although Glidden (2000) gives theoretical results for the Clayton copula in a balanced design with a fixed cluster size N and Othus and Li (2010) do the same in an unbalanced design for the Gaussian copula model, to our knowledge, copula models in general have not been used for clustered multivariate survival data with a cluster size of more than 4 or for a cluster size which differs over the clusters. The choice of a small and constant cluster size is a direct consequence of the difficulty to write down the likelihood function for the observed clustered survival data. For example, if the cluster size is equal to two, there are 4 different contributions to the likelihood for the observed outcomes within the cluster, depending on whether none, the first, the second or both individuals in this cluster are censored. This leads to a likelihood function consisting of 4 different terms where every term is found by taking derivatives of the joint survival function over the uncensored components in an observed couple:

$$\begin{aligned}L = \prod_{i=1}^{K} & (f(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}))^{\delta_{i1}\delta_{i2}} \left(-\frac{\partial S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1}}\right)^{\delta_{i1}(1-\delta_{i2})} \\ & \times \left(-\frac{\partial S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i2}}\right)^{(1-\delta_{i1})\delta_{i2}} (S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}))^{(1-\delta_{i1})(1-\delta_{i2})}.\end{aligned} \tag{1.1}$$

Denoting $u_{ij} = S(x_{ij}|\boldsymbol{Z}_{ij})$ and $c_\theta(u_{i1}, u_{i2}) = \dfrac{\partial^2}{\partial u_{i1}\partial u_{i2}} C_\theta(u_{i1}, u_{i2})$ we get the equivalent likelihood in terms of the copula function:

$$L = \prod_{i=1}^{K} (c_\theta(u_{i1}, u_{i2}))^{\delta_{i1}\delta_{i2}} \left(\frac{\partial C_\theta(u_{i1}, u_{i2})}{\partial u_{i1}}\right)^{\delta_{i1}(1-\delta_{i2})} \times \left(\frac{\partial C_\theta(u_{i1}, u_{i2})}{\partial u_{i2}}\right)^{(1-\delta_{i1})\delta_{i2}} (C_\theta(u_{i1}, u_{i2}))^{(1-\delta_{i1})(1-\delta_{i2})}. \quad (1.2)$$

If the cluster size is three, the number of possible combinations increases to 8, while a cluster size of 4 leads to 16 different combinations. In a general setting with a cluster size equal to n, we have 2^n possible combinations. Since a likelihood function then also contains 2^n different possible terms and each term is found by taking derivatives of the joint survival function over the uncensored components in a combination, it is a huge task if a general n-dimensional copula function is considered for the association between the different individuals within a cluster. It is in practice impossible to calculate a closed form for all the derivatives of a copula function if the order n is large.

For the class of Archimedean copula functions, we will solve this numerical problem in Chapter 2 and show that the construction of the likelihood function for this class of copula functions simplifies considerably such that we can allow the cluster size to be moderate to large and varying over the different clusters. The key to this solution is that the joint survival function of an Archimedean copula function can be rewritten as a mixture distribution of independent contributions in a similar way as in the frailty model approach. Although some of the expressions of the Archimedean copula function resemble that of the frailty model, the two models differ in an essential way due to their different inferential viewpoint, i.e., marginal versus conditional.

1.3 Data description

In this section, we will introduce two datasets that are used to exemplify the methods developed in future chapters. One dataset exhibits heavy censoring while the other includes light censoring. Both data sets are related to veterinary science, but the methodology presented in this dissertation applies to any scientific domain, as long as there are clustered survival data involved.

1.3.1 Udder infection data

Mastitis, the infection of the udder, is a frequently occurring disease in dairy cows. Mastitis control programs have been implemented due to the fact that large economic losses are associated with the disease. Control programs are focused on detection of mastitis, identification of the causative agent(s) and prevention of transmission by removing the source of the agent (milk contaminated fomites, bedding, persistently infected cows, etc.). Knowledge of mammary anatomy and physiology, mammary defense mechanism, microbial habitats, microbial virulence factors, milking machine function, and antibiotics/germicides is important in achieving effective mastitis control. In this study, 1196 cows have been followed during the lactation period, which is roughly 300-350 days but different for every cow. From each udder quarter, a milk sample is taken monthly and is screened for the presence of different bacteria. We define time to infection as the midpoint between the sampling times of the last negative result and the first positive result. To avoid convergence issues, the times to infection are multiplied by a factor $4/365.25$, so that time reflects the trimester rather than the day of the year. We model the time until infection with any bacteria, with the cow being the cluster and the quarter the experimental unit within the cluster. Observations are right-censored if no infection occurs before the end of the lactation period, or if the cow is lost to follow-up during the study. The censoring percentage is 61%.

The condition of the teat end deteriorates with an increasing number of calvings, which makes the udder more vulnerable to infections. Several studies have shown that prevalence as well as incidence of intramammary infections indeed increase with parity (Weller *et al.*, 1992). Therefore we consider the parity of the cow as a covariate. The binary parity variable is 1 for primiparous cows (cows that had one calving) and is 0 for multiparous cows (cows that had more than one calving). Logically, the parity covariate acts on the cow level.

This data set is an example of a balanced design, since all clusters have four components. We order the four parts in each cluster such that each first component corresponds to the left-front udder quarter, each second component to the left-rear udder quarter, and so on.

Because of the balanced nature and the small cluster size of this data set, it has been a very good learning example to extend copula modelling of survival data to more than two dimensions. All techniques for large and varying cluster sizes, that were

developed in Chapter 2, were double checked on this data set, because a cluster size equal to 4 is on the edge of what is possible with existing copula techniques.

In Chapter 5, different models for the association structure between the infection times of the four udder parts are compared.

cow id	location	time	status	parity
1	Left Front	0.734	1	1
1	Right Front	0.734	1	1
1	Left Rear	1.303	1	1
1	Right Rear	0.734	1	1
2	Left Front	0.728	1	0
2	Right Front	2.218	1	0
2	Left Rear	2.218	1	0
2	Right Rear	2.218	1	0
⋮	⋮	⋮	⋮	⋮
1196	Left Front	3.055	0	1
1196	Right Front	3.055	0	1
1196	Left Rear	3.055	0	1
1196	Right Rear	2.880	1	1

Table 1.1: Udder infection data. Column 1 contains the cow identification number. Column 2 denotes the location of the udder part. The observed time (column 3) is the minimum of the time to infection and the censoring time (trimester). The status is the censoring indicator: 1 if infected and 0 otherwise. The parity is 1 for a primiparous cow and 0 for a multiparous cow.

1.3.2 Insemination data

In dairy cattle, the calving interval (the time between two calvings) should be optimally between 12 and 13 months. One of the main factors determining the length of the calving interval is the time from parturition to the time of first insemination (Duchateau and Janssen, 2004). The objective of this study, amongst others, was to quantify the correlation between insemination times of cows within a herd. Insemination at a dairy farm is typically done by the farmer itself, relying on his experience. We want to get some insight into this process. The data set includes 181 clusters (farms) of different

sizes, ranging from 1 cow to 174 cows, with 10513 cows in total. As no inseminations take place in the first 29 days after calving, 29 days are subtracted from the time to first insemination since at risk time starts only then. Duchateau and Janssen (2004) suggest to transform time to months rather than days, but we did not encounter any convergence issues by not doing so, so we leave the time in days.

Although to a much smaller extent than the mastitis data, this data set is also subject to right censoring. If a cow is not inseminated within 300 days after calving, or if it is culled without being inseminated, it is censored at that time. The censoring percentage is 5.5%.

The time to first insemination may be influenced by the the parity of the cow (0 if multiparous, 1 if primiparous). Due to its unbalanced cluster sizes and censoring, this data set is very well suited to illustrate the newly developed techniques in Chapter 2.

farm id	cow id	time	status	parity
1	1	82	1	0
1	2	80	1	0
⋮	⋮	⋮	⋮	⋮
1	51	219	1	1
2	52	78	1	1
2	53	116	1	0
⋮	⋮	⋮	⋮	⋮
2	98	40	1	0
⋮	⋮	⋮	⋮	⋮
181	10433	46	1	1
181	10434	251	1	0
⋮	⋮	⋮	⋮	⋮
181	10513	73	1	1

Table 1.2: Insemination data. Columns 1 and 2 contain the identification number of the farm and the cow. The observed time (column 3) is the minimum of the time to first insemination and the censoring time (days). The status is the censoring indicator: 1 if inseminated and 0 otherwise. The parity is 0 for a multiparous cow and 1 for a primiparous cow.

1.4 Outline

The main objective of the first part of this dissertation is to tackle the problem of modelling multivariate survival data that are grouped in clusters of variable size, through copulas. In Chapter 2, we show that the family of Archimedean copulas is extremely well suited for this, due to the property that the generator of an Archimedean copula can be seen as a Laplace transform. In order to improve readability and not to overcharge this chapter with hardcore mathematical technicalities, the regularity conditions, theorems and proofs are put together in Chapter 3. In 2017, a condensed version of the material in Chapters 2 and 3 will be published in the *Journal of the Royal Statistical Society - Series B*.

The existing software to fit copula models does not accommodate for multidimensional *censored* data. In order to make the extended Archimedean copula model accessible to other users that encounter clustered survival data, we took our first steps in developing an R package. The methods are implemented for a selection of copula functions and marginal survival distributions. With hopefully many more upgrades to come, an illustration of the use and output of our package `Sunclarco` can be found in Chapter 4.

In Chapter 5 we literally take the Archimedean copula model to the next level, as we allow for a hierarchy of clustering. To explore the opportunities and shortcomings of hierarchical Archimedean copula models in survival analysis, we start from the four-dimensional mastitis data set, and compare a set of models that are biologically relevant in this context. A manuscript along the lines of Chapter 5 was submitted to *Biometrics* for publication.

Each chapter concludes with a discussion of the presented material, including do's and don'ts, together with ideas for future research.

We refer the reader to Appendices A and B for supplementary mathematical details linked to Chapters 2 and 5.

Chapter 2

Extending the Archimedean copula methodology to model multivariate survival data grouped in clusters of variable size

We describe in this chapter a copula model for clustered survival data where the clusters are allowed to be moderate to large and varying in size by considering the class of Archimedean copulas with completely monotone generator. We develop both one- and two-stage estimators for the different copula parameters. Furthermore we show the consistency and asymptotic normality of these estimators. Finally, we perform a simulation study to investigate the finite sample properties of the estimators. We illustrate the method on a data set containing the time to first insemination in cows, with cows clustered in herds.

The chapter is organized as follows. In Section 2.1 we introduce a new formulation of the

Archimedean copula model by rewriting the likelihood contributions in terms of Laplace transforms. In Section 2.2 we present the theoretical results concerning estimators arising from this model, starting from parametric and semi-parametric approaches. Section 2.3 gives an overview of a large class of distributions for which the likelihood contributions are easy to generate. In Sections 2.4 and 2.5, we report simulation results along with results for a data example. The data set and our code can be found at our website (http://www.vetstat.ugent.be/research/ArchimedeanCopula/). Proofs of asymptotic results are given in Chapter 3.

2.1 Description of the model

We develop a copula model for clustered survival data in which the size of each cluster may be different. Let K be the number of clusters ($i = 1, \ldots, K$). In each cluster, we denote the lifetime for the different individuals by a positive random variable T_{ij}, $j = 1, \ldots, n_i$ where n_i is the number of individuals in cluster i. For each individual, we assume that there is an independent random censoring variable C_{ij} such that under a right censoring scheme, the observed quantities are given by

$$\begin{array}{l} X_{ij} = \min(T_{ij}, C_{ij}) \\ \delta_{ij} = I(T_{ij} \leq C_{ij}) \end{array} \qquad i = 1, \ldots, K, \quad j = 1, \ldots, n_i.$$

The risk of failure may also depend on a set of covariates $\boldsymbol{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijp})^T$. We assume that the joint survival function for the lifetime of the different individuals within cluster i is given by

$$\begin{aligned} S(t_{i1}, \ldots, t_{in_i} | \boldsymbol{Z}_{i1}, \ldots, \boldsymbol{Z}_{in_i}) &= P(T_{i1} > t_{i1}, \ldots, T_{in_i} > t_{in_i} | \boldsymbol{Z}_{i1}, \ldots, \boldsymbol{Z}_{in_i}) \\ &= \varphi_\theta \left[\varphi_\theta^{-1}(S(t_{i1} | \boldsymbol{Z}_{i1})) + \cdots + \varphi_\theta^{-1}(S(t_{in_i} | \boldsymbol{Z}_{in_i})) \right] \end{aligned}$$

where $S(t_{ij}|\boldsymbol{Z}_{ij}) = P(T_{ij} > t_{ij}|\boldsymbol{Z}_{ij})$ is a common marginal survival model for the lifetime T_{ij}, given $\boldsymbol{Z}_{ij}$. The generator $\varphi_\theta : [0, \infty[\rightarrow [0, 1]$ of a parametric Archimedean copula family is a continuous strictly decreasing function with $\varphi_\theta(0) = 1$ and $\varphi_\theta(\infty) = 0$. We denote by φ_θ^{-1} the inverse function of φ_θ. Since we want the Archimedean copula function to be correctly defined for any cluster size, we assume that this generator is completely monotonic. This means that all the derivatives exist and have alternating signs: $(-1)^m \frac{d^m}{dt^m} \varphi_\theta(t) \geq 0$, for all $t > 0$ and $m = 0, 1, 2, \ldots$ (see Nelsen (2006)). The generator φ_θ is a Laplace transformation of a positive distribution function $G_\theta(x)$ with

$\bar{G}_\theta(0) = 1$ (Joe, 1997),

$$\varphi_\theta(t) = \int_0^\infty e^{-tx} dG_\theta(x), \quad t \geq 0.$$

Hence we can rewrite the joint survival function for cluster i as

$$\begin{aligned} S(t_{i1}, \ldots, t_{in_i} | \boldsymbol{Z}_{i1}, \ldots, \boldsymbol{Z}_{in_i}) &= \varphi_\theta \left[\sum_{j=1}^{n_i} \varphi_\theta^{-1} \left(S(t_{ij} | \boldsymbol{Z}_{ij}) \right) \right] \\ &= \int_0^\infty e^{-x \sum_{j=1}^{n_i} \varphi_\theta^{-1}(S(t_{ij}|\boldsymbol{Z}_{ij}))} dG_\theta(x) \\ &= \int_0^\infty \prod_{j=1}^{n_i} e^{-x\varphi_\theta^{-1}(S(t_{ij}|\boldsymbol{Z}_{ij}))} dG_\theta(x). \end{aligned} \tag{2.1}$$

In this way, the Archimedean copula function can be seen as a mixture distribution, consisting of independent and identically distributed components which depend on a common factor that has G_θ as distribution. We use this structure to derive the likelihood function. The contribution of cluster i, with cluster size n_i, to the likelihood function corresponds to the derivative of the n_i-dimensional joint survival function over all uncensored individuals in this cluster. Since the joint survival function does not change when the individuals within the cluster are permuted, we note that only the number of uncensored individuals determines the derivative. In the bivariate case, the contribution of cluster (pair) i to the likelihood function is

- for a pair with two events $((\delta_{i1}, \delta_{i2}) = (1, 1))$:

$$f(x_{i1}, x_{i2} | \boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}) = \frac{\partial^2 S(x_{i1}, x_{i2} | \boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1} \partial x_{i2}},$$

- for a pair with one event and one censored observation $((\delta_{i1}, \delta_{i2}) = (1, 0)$ or $(\delta_{i1}, \delta_{i2}) = (0, 1))$:

$$-\frac{\partial S(x_{i1}, x_{i2} | \boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{ij}},$$

- for a pair with two censored observations $((\delta_{i1}, \delta_{i2}) = (0, 0))$:

$$S(x_{i1}, x_{i2} | \boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}).$$

In Appendix A we show that each of these contributions can be written as

$$\int_0^\infty \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij}))} \left[\frac{-xf(x_{ij}|\boldsymbol{Z}_{ij})}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})))}\right]^{\delta_{ij}} dG_\theta(x).$$

The bivariate Archimedean copula likelihood is given by the product of these contributions over all K clusters

$$L = \prod_{i=1}^{K} \int_0^\infty \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij}))} \left[\frac{-xf(x_{ij}|\boldsymbol{Z}_{ij})}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})))}\right]^{\delta_{ij}} dG_\theta(x).$$

In the case of unbalanced clusters, the contribution of cluster i to the likelihood function is

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

where $\partial\{\delta_{ij}=1\}$ is the set of uncensored individuals in cluster i and $d_i = \sum_{j=1}^{n_i} \delta_{ij}$, the size of this set. Using representation (2.1) of the joint survival function in the same way as was done in Appendix A for bivariate data, this derivative is given by

$$L_i = \int_0^\infty \prod_{j=1}^{n_i} e^{-x\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij}))} \left[\frac{-xf(x_{ij}|\boldsymbol{Z}_{ij})}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z_{ij}})))}\right]^{\delta_{ij}} dG_\theta(x).$$

Combining the contributions over the different clusters, we get the following likelihood function for survival data that are clustered in groups of unequal sizes

$$L = \prod_{i=1}^{K} \int_0^\infty \prod_{j=1}^{n_i} e^{-x\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij}))} \left[\frac{-xf(x_{ij}|\boldsymbol{Z}_{ij})}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z_{ij}})))}\right]^{\delta_{ij}} dG_\theta(x). \quad (2.2)$$

In general it is difficult to evaluate expression (2.2) except for very specific choices of the distribution G_θ. It is done for the one-parameter gamma distribution in Appendix A.2. Since the generator φ_θ is the Laplace transform of G_θ, there is an alternative expression for the likelihood function (2.2) which is found by using derivatives of this generator, i.e. $\varphi_\theta^{(m)}(t) = \int_0^\infty (-x)^m e^{-tx} dG_\theta(x)$. Hence the likelihood function can be rewritten as

$$L = \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{f(x_{ij}|\boldsymbol{Z}_{ij})}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})))}\right]^{\delta_{ij}} \right) \varphi_\theta^{(d_i)} \left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})) \right). \quad (2.3)$$

Remark. In the frailty model framework (Duchateau and Janssen, 2008, p.119), we note that we find a similar expression for the joint survival function in frailty models, with $G_\theta(x)$ as the frailty distribution of the unknown frailty term in the cluster. Starting

from the conditional viewpoint in frailty models, we find a similar expression for the joint survival function as follows. The joint conditional survival function for a cluster i is given by $S(t_{i1},\ldots,t_{in_i}|\boldsymbol{Z}_{i1},\ldots,\boldsymbol{Z}_{in_i},U_i)$ with U_i the frailty term with distribution $G_\theta(u)$ and generator $\varphi_\theta(\cdot)$. Denote the conditional cumulative hazard function for subject j from cluster i by $H(t_{ij}|\boldsymbol{Z}_{ij},U_i) = H_c(t_{ij}|\boldsymbol{Z}_{ij})U_i$. The marginal joint survival function is obtained by integrating out the frailty term:

$$\begin{aligned} S_f(t_{i1},\ldots,t_{in_i}|\boldsymbol{Z}_{i1},\ldots,\boldsymbol{Z}_{in_i}) &= \int_0^\infty S(t_{i1},\ldots,t_{in_i}|\boldsymbol{Z}_{i1},\ldots,\boldsymbol{Z}_{in_i},u_i)dG_\theta(u_i) \\ &= \int_0^\infty S(t_{i1}|\boldsymbol{Z}_{i1},u_i)\ldots S(t_{in_i}|\boldsymbol{Z}_{in_i},u_i)dG_\theta(u_i) \\ &= \int_0^\infty \exp(-u_i\sum_{j=1}^{n_i} H_c(t_{ij}|\boldsymbol{Z}_{ij}))dG_\theta(u_i) \\ &= \int_0^\infty \exp(-u_i\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_f(t_{ij}|\boldsymbol{Z}_{ij})))dG_\theta(u_i) \quad (2.4) \end{aligned}$$

due to the conditional independence assumption. The two joint survival functions (2.1) and (2.4) are indeed similar, but note that $S(t_{ij}|\boldsymbol{Z}_{ij}) \neq S_f(t_{ij}|\boldsymbol{Z}_{ij})$. More specifically $S_f(t_{ij}|\boldsymbol{Z}_{ij}) = \varphi_\theta(H_c(t_{ij}|\boldsymbol{Z}_{ij}))$ and therefore, the marginal survival function in (2.4) contains the association parameter. This an important distinction between the frailty model and the copula model.

2.2 The estimation procedures

In this section, we investigate a one- and two-stage parametric estimation method and a two-stage semi-parametric estimation method to estimate the different parameters in this model. Shih and Louis (1995) demonstrated how this can be done for a bivariate survival data set and derived asymptotic properties of the estimators. Joe (1997, 2005) discussed a general framework for studying asymptotic efficiency. We extend their results to clustered survival data with clusters of varying and possibly large size.

For equal-sized clusters with cluster size n having the same covariate structure, baseline survival functions can be estimated for each j^{th} univariate margin, $j = 1,\ldots,n$, where the j^{th} subject always has the same covariate information. Since in our application

clusters have varying size, we cannot order the components within a cluster and estimate the baseline survival of all j^{th} components. We assume that all subjects have the same baseline survival, whatever the cluster, and allow for subject specific covariate information.

2.2.1 One-stage parametric estimation

Let $\boldsymbol{\beta}$ be the q-dimensional parameter vector for the margins, containing distribution-specific parameters for the baseline survival and covariate effects. We use the likelihood function $L(\boldsymbol{\beta}, \theta)$ as derived in (2.2) and (2.3).
Write $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \theta) = \dfrac{\partial \log L(\boldsymbol{\beta}, \theta)}{\partial \boldsymbol{\beta}}, U_\theta(\boldsymbol{\beta}, \theta) = \dfrac{\partial \log L(\boldsymbol{\beta}, \theta)}{\partial \theta}$ for the score functions.
Solving

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

simultaneously, we find the maximum likelihood estimate $(\hat{\boldsymbol{\beta}}, \hat{\theta})$.
Denote the Hessian matrix as $\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\beta}}, \hat{\theta}) = \left[\dfrac{\partial^2 \log L(\hat{\boldsymbol{\beta}}, \hat{\theta})}{\partial \eta_i \partial \eta_j} \right]_{i,j=1\ldots,q+1}$ with $\boldsymbol{\eta} = (\boldsymbol{\beta}, \theta)$.
From maximum likelihood theory (Cox and Hinkley, 1974; Lehmann and Casella, 1998), we know that under regularity conditions (C.1)-(C.5) in Chapter 3, $\sqrt{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\theta} - \theta)$ converges to a multivariate normal distribution with mean vector zero and variance-covariance matrix $\mathbf{I}^{-1}$, where the Fisher information matrix $\mathbf{I} = \left[-\dfrac{\partial^2 \log L(\boldsymbol{\beta}, \theta)}{\partial \eta_i \partial \eta_j} \right]_{i,j=1\ldots,q+1}$ is partitioned into blocks:

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

Here, $K\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is the $q \times q$ variance-covariance matrix of $\mathbf{U}_{\boldsymbol{\beta}}$, $K\mathbf{I}_{\boldsymbol{\beta}\theta}$ is the $q \times 1$ covariance vector between $\mathbf{U}_{\boldsymbol{\beta}}$ and U_θ and $KI_{\theta\theta}$ is the scalar variance of U_θ. Furthermore, $\mathbf{I}_{\theta\boldsymbol{\beta}} = \mathbf{I}_{\boldsymbol{\beta}\theta}^T$.
Since the inverse of the block matrix

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

is defined by

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

we find that the lower right element of $\mathbf{I}^{-1}$ is

$$\text{Var}(\hat{\theta}) = (\text{I}_{\theta\theta} - \mathbf{I}_{\theta\boldsymbol{\beta}}\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\mathbf{I}_{\boldsymbol{\beta}\theta})^{-1}.$$

To match the appearance with the variance expressions in the sections to come, we rewrite this as

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{\mathrm{I}_{\theta\theta}} + \frac{\mathbf{I}_{\theta\boldsymbol{\beta}}(\mathbf{I}^{-1})_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{I}_{\boldsymbol{\beta}\theta}}{\mathrm{I}^2_{\theta\theta}}. \tag{2.5}$$

Note that $(\mathbf{I}^{-1})_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is the upper left $q \times q$ block of $\mathbf{I}^{-1}$, which is not the same as $\mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}$.

In practical applications, standard errors of parameter estimates can be retrieved from the diagonal elements of the inverse of the Hessian matrix $\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\beta}}, \hat{\theta})$, since we know from maximum likelihood theory that $K\mathbf{I} = E\left[-\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\beta}}, \hat{\theta})\right]$.

2.2.2 Two-stage parametric estimation

Two-stage parametric estimation, also referred to as the method of inference functions for margins (Xu, 1996), has been used mainly for multivariate models whenever a multi-parameter numerical optimization for maximum likelihood estimation is too time-consuming or infeasible. In the first stage, $\boldsymbol{\beta}$ is estimated by $\overline{\boldsymbol{\beta}}$ by considering all subjects as independent, identically distributed random variables, i.e. solving

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

Since clusters are considered as mutually independent, the $\mathbf{U}^*_{i,\boldsymbol{\beta}}(\cdot)$ are independent and identically distributed. Under regularity conditions (C.6)-(C.10) in Chapter 3, $\sqrt{K}(\overline{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges to a multivariate normal distribution with mean vector zero and variance-covariance matrix $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$, where $\mathbf{V}$ is the variance-covariance matrix of the score functions $\mathbf{U}^*_{\boldsymbol{\beta}}$;

$$\mathbf{V} = E\left[\mathbf{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta}_0)(\mathbf{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta}_0))'\right]$$

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

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

The use of the robust sandwich estimator is required since $(\mathbf{I}^*)^{-1}$ is not a consistent estimator of the asymptotic variance-covariance matrix due to the correlation between

survival times. In case of independence, we get $\mathbf{V} = \mathbf{I}^*$, and the formula $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$ reduces to the usual inverse Fisher information matrix $(\mathbf{I}^*)^{-1}$ (Pawitan, 2001).

In the second stage, the association parameter θ is estimated by plugging in the estimates for the margins into the likelihood expression (2.3), which is then maximized for the association parameter θ. The two-stage estimator for θ is the solution to

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

The asymptotic normality of the two-stage estimator is established below. The proof is deferred to Chapter 3.

Theorem 2.2.1. *Let $\overline{\theta}$ denote the solution to $U_\theta(\overline{\boldsymbol{\beta}}, \theta) = 0$ and let θ_0 be the true value of the association parameter. Under regularity conditions (C.10)-(C.14) in Chapter 3, $\sqrt{K}(\overline{\theta} - \theta_0)$ converges to a normal distribution with mean zero and variance*

$$\mathrm{Var}(\overline{\theta}) = \frac{1}{\mathrm{I}_{\theta\theta}} + \frac{\mathbf{I}_{\theta\beta}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\beta\theta}}{\mathrm{I}^2_{\theta\theta}}. \tag{2.6}$$

To estimate this quantity, we make use of $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$, the robust variance obtained in the first step; $I^{-1}_{\theta\theta}$ and $\mathbf{I}_{\boldsymbol{\beta}\theta}$ are obtained from the Hessian matrix of the one-stage procedure, which can be estimated numerically by performing one iteration of the one-stage optimization in which we evaluate the Hessian matrix in the two-stage parameter results.

2.2.3 Two-stage semi-parametric estimation

In the two-stage semi-parametric estimation procedure, the marginal survival functions are estimated using the Cox proportional hazards model (Cox, 1972). Formulas for the standard error of the estimated covariate effect $\breve{\boldsymbol{\beta}}$ and the estimated cumulative hazard $\breve{\Lambda}$ that account for clustering can be found using a sandwich formula (Spiekerman and Lin, 1998). The results in Spiekerman and Lin (1998) even accommodate for time-varying covariates.

In the second stage, $\max_\theta L(\theta; \breve{\boldsymbol{\beta}}, \breve{\Lambda})$ is solved for $\breve{\theta}$.

Theorem 2.2.2. *Under regularity conditions (C.15)-(C.21) in Chapter 3,* $(\breve{\theta}; \breve{\boldsymbol{\beta}}, \breve{\Lambda})$ *is a consistent estimator for* $(\theta_0; \boldsymbol{\beta}_0, \Lambda_0)$.

The results for $\breve{\boldsymbol{\beta}}$ and $\breve{\Lambda}$ follow from arguments along the lines of Spiekerman and Lin (1998). The proof of the consistency of $\breve{\theta}$ is quite tedious and can be found in Chapter 3. Also following Spiekerman and Lin (1998), we can show that $\sqrt{K}(\breve{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a mean zero normal distribution and that $\sqrt{K}(\breve{\Lambda} - \Lambda_0)$ converges to a mean zero Gaussian process.

Theorem 2.2.3. *Under regularity conditions (C.15)-(C.21) in Chapter 3,* $\sqrt{K}(\breve{\theta} - \theta_0)$ *converges to a normal distribution with mean zero and variance*

$$\frac{\mathrm{Var}(\Xi_1)}{W(\theta_0)^2}.$$

The proof of this theorem and the precise definition of Ξ_1 and $W(\theta_0)$, together with their estimators, can be found in Chapter 3.

2.3 Copula likelihood expression for distributions from the PVF family

The power variance function family of distributions, denoted PVF(α, δ, γ), is a large class of distributions for which Hougaard (2000) states that the Laplace transforms correspond to

$$\mathcal{L}(s) = \exp\left[-\frac{\delta}{\alpha}\left((\gamma + s)^\alpha - \gamma^\alpha\right)\right]$$

with derivatives

$$\mathcal{L}^{(k)}(s) = (-1)^k \mathcal{L}(s) \sum_{j=1}^{k} c_{k,j}(\alpha)\delta^j(\gamma + s)^{j\alpha - k},$$

where the coefficients $c_{k,j}(\alpha)$ are polynomials of order $k - j$ in α, given by the recursive formula

$$c_{k,1}(\alpha) = \frac{\Gamma(k - \alpha)}{\Gamma(1 - \alpha)}, \quad c_{k,k} = 1$$
$$c_{k,j}(\alpha) = c_{k-1,j-1}(\alpha) + c_{k-1,j}(\alpha)(k - 1 - j\alpha).$$

This allows for a closed form expression of the copula likelihood (2.3). In practical implementations, we first calculate the list of coefficients $c_{k,j}(\alpha)$, $k = 1, \ldots, \max(d_i)$,

$j = 1, \ldots, k$ from the iterative scheme and plug these into the expression for the kth derivative of the Archimedean copula generator to obtain the likelihood function (2.3). This is the modus operandi in the following Examples 2 and 3. In some rare cases, there exists an explicit expression for the kth derivative of the Laplace transform and hence the likelihood can be calculated directly. This is illustrated in Example 1.

Example 1: The one-parameter gamma distribution with density

$$g_\theta(x) = \frac{x^{1/\theta - 1} e^{-x/\theta}}{\theta^{1/\theta}\Gamma(1/\theta)}, \quad \theta > 0.$$

is found as the limiting case $\alpha = 0, \delta = \gamma = 1/\theta$. Failure times are independent when θ approaches zero. The Laplace transform is

$$\mathcal{L}(s) = \varphi_\theta(s) = (1 + \theta s)^{-1/\theta}$$

which is the generator of the Clayton copula. The first four derivatives of the Laplace transform are

$$\begin{aligned}
\varphi_\theta^{(1)}(s) &= (-1)^1 (1+\theta s)^{-1/\theta - 1} \\
\varphi_\theta^{(2)}(s) &= (-1)^2 (1+\theta)(1+\theta s)^{-1/\theta - 2} \\
\varphi_\theta^{(3)}(s) &= (-1)^3 (1+\theta)(1+2\theta)(1+\theta s)^{-1/\theta - 3} \\
\varphi_\theta^{(4)}(s) &= (-1)^4 (1+\theta)(1+2\theta)(1+3\theta)(1+\theta s)^{-1/\theta - 4}.
\end{aligned}$$

The k-th derivative of the Laplace transform is

$$\varphi_\theta^{(k)}(s) = (-1)^k (1+\theta s)^{-(k+1/\theta)} \prod_{l=0}^{k-1} (1 + l\theta).$$

This allows for a closed form expression of the copula likelihood (2.3). Denote $S_{ij} = S(x_{ij}|\boldsymbol{Z}_{ij})$ and $f_{ij} = f(x_{ij}|\boldsymbol{Z}_{ij})$.

$$\begin{aligned}
L &= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) \varphi_\theta^{(d_i)} \left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij}) \right) \\
&= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) (-1)^{d_i} \left(1 + \theta \sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij}) \right)^{-(d_i + 1/\theta)} \prod_{l=0}^{d_i - 1} (1 + l\theta) \\
&= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{-f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) \left(1 + \theta \sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij}) \right)^{-(d_i + 1/\theta)} \prod_{l=0}^{d_i - 1} (1 + l\theta).
\end{aligned}$$

Since $\varphi_\theta^{-1}(s) = (s^{-\theta} - 1)/\theta$, this equals

$$L = \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{f_{ij}}{S_{ij}^{1+\theta}} \right]^{\delta_{ij}} \right) \left(1 - n_i + \sum_{j=1}^{n_i} S_{ij}^{-\theta} \right)^{-(d_i + 1/\theta)} \prod_{l=1}^{d_i - 1} (1 + l\theta). \tag{2.7}$$

In Appendix A.2, we show that this likelihood can also be calculated directly from equation (2.2).

The corresponding log-likelihood is given by

$$\begin{aligned} \log L = \sum_{i=1}^{K} \Bigg(& \sum_{j=1}^{n_i} \delta_{ij} \left[\log f_{ij} - (1+\theta) \log S_{ij} \right] \\ & -(d_i + 1/\theta) \log \left(1 - n_i + \sum_{j=1}^{n_i} S_{ij}^{-\theta} \right) + \sum_{l=0}^{d_i - 1} \log(1 + l\theta) \Bigg). \end{aligned} \tag{2.8}$$

Just as any Archimedean copula, the n_i-variate Clayton copula is exchangeable, meaning that all bivariate copula margins are the same:

$$C(S_{ij_1}, S_{ij_2}) = C(S_{ij_2}, S_{ij_1}) = \varphi_\theta(\varphi_\theta^{-1}(S_{ij_1}) + \varphi_\theta(\varphi_\theta^{-1}(S_{ij_2})).$$

In Figure 2.1, 1000 samples from a bivariate Clayton copula are visualized. The association parameter ranges from $\theta = 0$ (independence) to $\theta = 3$. As can be seen from Figures 2.1 (b-d), the values of S_{ij_1} and S_{ij_2} are strongest correlated in the lower left corner. We say that the Clayton copula has lower tail dependence. In a survival context, where $S_{ij} = P(T_{ij} > x_{ij})$ is a monotone decreasing function, this corresponds to a stronger association between later lifetimes.

Example 2: The choice $\alpha = \theta, \delta = \theta, \gamma = 0$ leads to the positive stable distribution with density

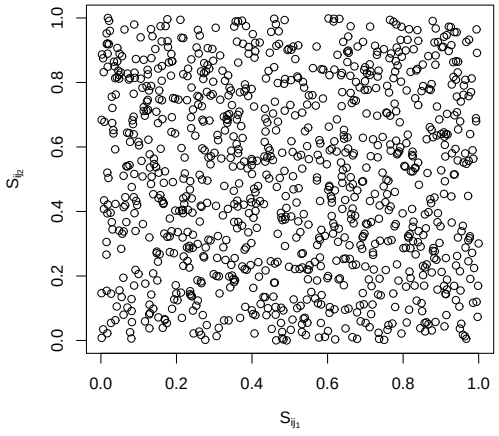
$$g_\theta(x) = -\frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{\Gamma(k\theta + 1)}{k!} (-x^{-\theta})^k \sin(\theta k \pi)$$

with $0 < \theta < 1$. Feller (1971) shows that this density function can be found by Fourier inversion of the Laplace transform

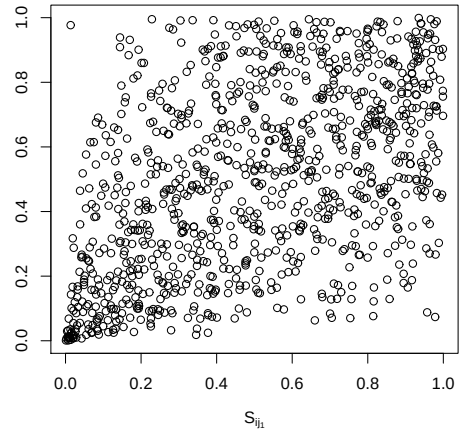
$$\mathcal{L}(s) = \varphi_\theta(s) = e^{-s^\theta}$$

which is the generator of the Gumbel-Hougaard copula. Small values of θ provide large correlation and survival times are independent as θ approaches 1. The Gumbel-Hougaard copula has upper tail dependence, as can be seen from Figure 2.2, implying a stronger correlation between the lower survival times.

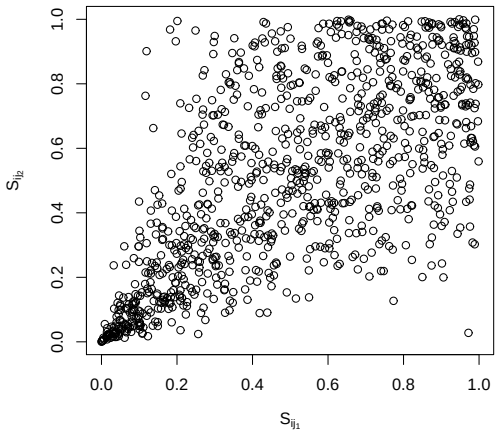
(a) Clayton copula, $\theta = 0$ (independence)



(b) Clayton copula, $\theta = 1$



(c) Clayton copula, $\theta = 2$



(d) Clayton copula, $\theta = 3$

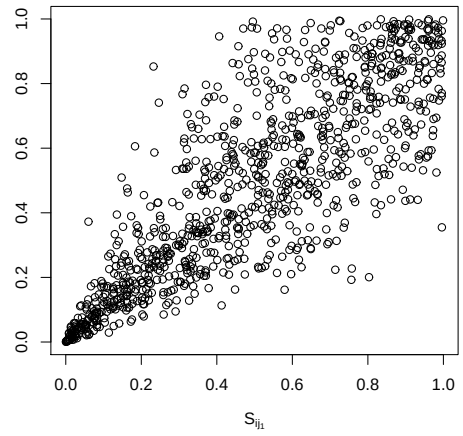
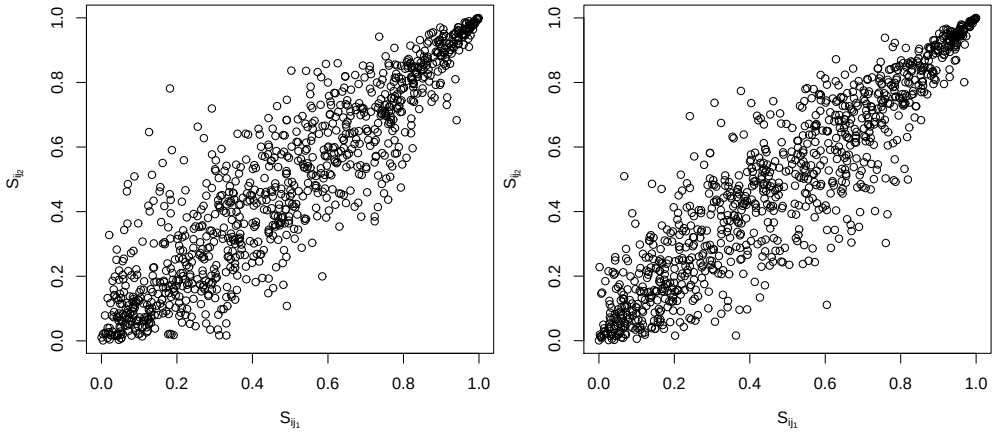


Figure 2.1: Bivariate Clayton copula with $\theta = 0, 1, 2, 3$.

(a) Gumbel copula, $\theta = 0.25$

(b) Gumbel copula, $\theta = 0.33$



(c) Gumbel copula, $\theta = 0.5$

(d) Gumbel copula, $\theta = 1$ (independence)

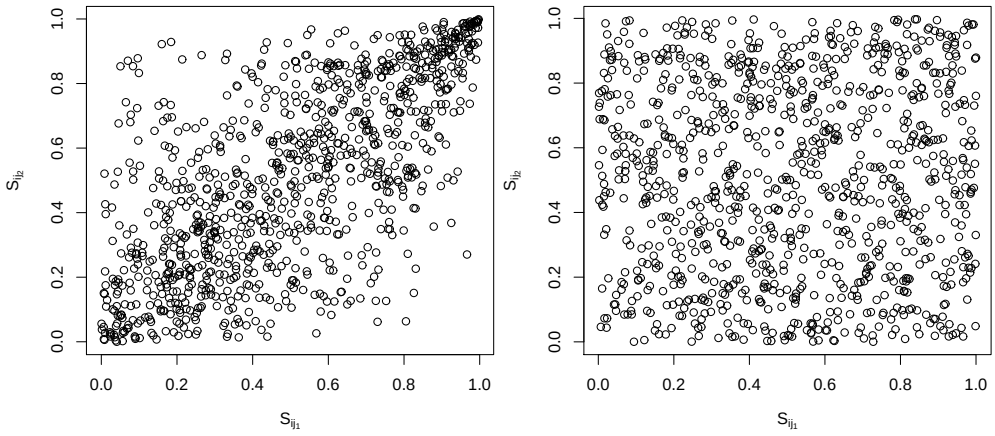


Figure 2.2: Bivariate Gumbel copula with $\theta = 0.25, 0.33, 0.5, 1$.

Example 3: Another PVF distribution is obtained by choosing $\alpha = 1/2, \delta = (2\theta)^{-1/2}, \gamma = (2\theta)^{-1}$. This is the inverse Gaussian distribution with variance θ. The density is defined by

$$f_\theta(x) = \sqrt{\frac{1}{2\pi\theta}} x^{-3/2} \exp\left(\frac{-1}{2x\theta}(x-1)^2\right)$$

with $\theta > 0$. The Laplace transform is

$$\mathcal{L}(s) = \varphi_\theta(s) = \exp\left(\frac{1}{\theta} - \left(\frac{1}{\theta^2} + 2\frac{s}{\theta}\right)^{1/2}\right).$$

The tail dependence of the inverse Gaussian copula is in between upper and lower tail dependence. The distributions in these three examples therefore cover a wide range of association structures in the data.

2.4 Simulation study

We generate 1000 data sets with 50, 200 or 500 clusters of size varying uniformly between 2 and 50. Survival times are simulated from respectively a Clayton copula with $\theta_0 = 0.2, 0.5, 1.0, 1.5$ or from a Gumbel-Hougaard copula with $\theta_0 = 0.2, 0.5, 0.65, 0.8$, and with, in both settings, Weibull marginal survival functions $S(t) = \exp(\lambda t^\rho \exp(\beta' Z))$, choosing $\rho = 1.5$, $\lambda = 0.0316$ and Z a dichotomous covariate with effect $\beta = 3$. The values of the association parameter θ for both copula models are chosen such that the according values of Kendall's tau are comparable. Data are generated using the sampling algorithm of Marshall and Olkin (1988). The censoring distribution is also Weibull, with parameters $(\lambda_C = 0.0274, \rho_C = 1.5)$ and $(\lambda_C = 0.1464, \rho_C = 1.5)$ yielding censoring percentages of 25% and 50%, respectively. The performances of one-stage parametric estimation, two-stage parametric estimation and two-stage semi-parametric estimation are summarized in Tables 2.2, 2.3 and 2.4. Since the focus is on the association parameter, the results for the marginal parameters λ and ρ and the covariate effect β are not reported. For each copula, simulation results are listed in increasing order of association. For the Clayton copula, higher values of θ correspond to a higher degree of association via $\tau = \frac{\theta}{\theta+2}$ whereas the inverse link holds for the Gumbel-Hougaard copula $(\tau = 1 - \theta)$. For each degree of association, we report the mean estimated values of $\hat{\theta}$, $\overline{\theta}$ and $\breve{\theta}$ in the first row. Mean standard errors together with the coverage are reported in the second row. Standard errors of one-stage parametric estimators are calculated from the inverse Hessian matrix. In the two-stage parametric approach, standard errors are

found via formula (2.6). In the two-stage semi-parametric case, we used the grouped jackknife to obtain standard errors (Lipsitz *et al.*, 1994; Lipsitz and Parzen, 1996). As in the work of Othus and Li (2010) we noted that the variance expression in the two-stage semi-parametric estimation method is rather complicated to implement. We assessed the performance of the jackknife procedure in the two-stage parametric model by comparing the standard error through the theoretical expression with a jackknife alternative. Since the results were virtually the same, we only show the standard error calculated from the theoretical expression.

Note that, as the number of clusters increases from $K = 50$ (Table 2.2) to $K = 200$ (Table 2.3), standard errors are halved since they are proportional to $1/\sqrt{K}$. For the Gumbel-Hougaard copula, the bias of the estimates are not noticeably affected by an increasing percentage of censoring, Only when we go from the one-stage parametric estimation method to the two-stage estimation methods we have an increase in the bias. However the standard errors become a bit larger when more censoring is present. For the Clayton copula, we observe that the bias of the estimators increases more when the percentage of censoring increases than in the case of the Gumbel-Hougaard copula. For the standard errors, we see in the Clayton copula similar results as for the Gumbel-Hougaard copula. The combined effect of the increased bias and slightly different standard errors for the Clayton model in comparison of the Gumbel-Hougaard model explain why the coverages are smaller in the Clayton model than in the Gumbel-Hougaard model. A general observation is that biases and standard errors tend to shrink as θ_0 approaches independence. In each of Tables 2.2, 2.3 and 2.4, the largest biases are found in the semi-parametric cases where θ_0 has moved far away from independence. The transition from $K = 50$ to $K = 200$ and $K = 500$ leads to a reduction of the bias, which also follows from the asymptotic proofs in Chapter 3. However, when the number of clusters is small and the variability of cluster sizes is large, the two-stage parametric and semi-parametric procedures are not recommended. Although computationally more intensive, the one-stage parametric procedure yields the best results in every setting.

2.5 Modelling time to first insemination in cows clustered in herds

In this data example, the insemination data set, introduced in Section 1.3.2, is analysed. In the parametric approach, we first assume a Weibull distribution for the times to first

insemination

$$S(t) = \exp(-\lambda \exp(\beta' Z) t^{\rho})$$

and model the association structure by a Clayton copula and a Gumbel-Hougaard copula. In Table 2.1, the results are listed for the parity effect and association parameter, using the one-stage parametric, two-stage parametric and two-stage semi-parametric estimation procedures. In addition, a model with piecewise constant baseline hazard was also fitted, because it has the advantage of a flexible baseline hazard - making it a good alternative for the semi-parametric model - but is also parametric , and thus the one-stage estimation procedure can be used. Hereby cutpoints are chosen such that each time interval contains 5% of the events. A piecewise constant hazard corresponds to a piecewise exponential distribution of event times and will be abbreviated by PWE.

In both copula models, the results for the parity effect are similar for all estimation approaches (see Table 2.1). A primiparous cow has a significantly lower hazard of being inseminated than a multiparous cow. The hazard ratio in the one-stage Weibull-Clayton model equals 0.92 (95% CI: $[0.89, 0.95]$), and is 0.95 (95% CI: $[0.92, 0.97]$) for the Weibull-Gumbel-Hougaard model. Both the parametric Weibull and semi-parametric two-stage approaches lead to a hazard ratio of 0.94 (95% CI: $[0.90, 0.98]$). For the PWE-Clayton and PWE-Gumbel-Hougaard models, hazard ratios are 0.93 (95% CI: $[0.90, 0.96]$) and 0.94 (95% CI: $[0.92, 0.97]$), respectively. Within each copula model, the parameter estimates for θ vary over the different estimation techniques. The lowest values of θ are observed for the one-stage Weibull models and the highest for the two-stage semi-parametric models. Regarding the simulation results in Section 2.4, we emphasize that for relatively small sample sizes, the one-stage parametric procedure is most reliable. If the Weibull assumption is questionable, a piecewise exponential model for the hazard function is recommended.

	Clayton copula				Gumbel-Hougaard copula			
	Weibull one-stage	Weibull two-stage	PWE one-stage	Semi-par. two-stage	Weibull one-stage	Weibull two-stage	PWE one-stage	Semi-par. two-stage
β	-0.082	-0.066	-0.070	-0.060	-0.055	-0.066	-0.058	-0.060
	(0.017)	(0.022)	(0.016)	(0.021)	(0.013)	(0.022)	(0.014)	(0.021)
θ	0.212	0.324	0.352	0.447	0.624	0.766	0.661	0.790
	(0.015)	(0.050)	(0.034)	(0.063)	(0.016)	(0.018)	(0.013)	(0.016)

Table 2.1: Estimation results for time to first insemination data

A visual check of the estimated marginal survival curves (see Figure 2.3) reveals why

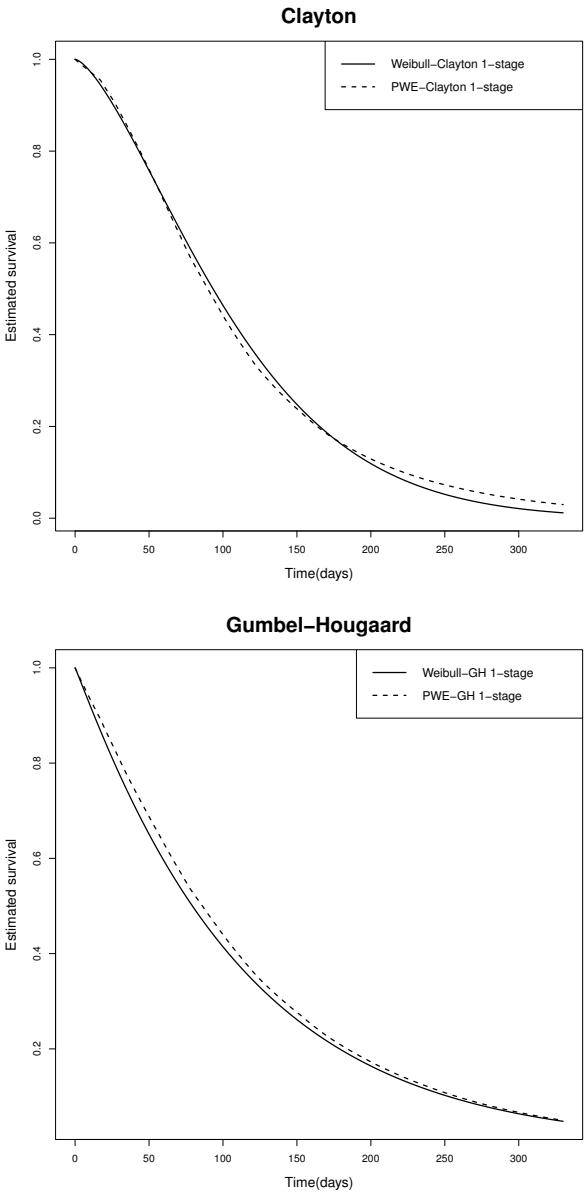


Figure 2.3: Estimated survival curves for multiparous cows

the difference between the estimated association parameter θ in the one-stage Weibull-Clayton and PWE-Clayton is so large (0.212 versus 0.352). The difference between the estimated marginal survival functions is largest for later times, which are the times when the Clayton copula imposes a higher dependency. If the Weibull assumption is incorrect, the estimated association parameter will also lack accuracy. In this example, we used both a Clayton and a Gumbel-Hougaard copula to illustrate our techniques. At this moment, we did not focus on a goodness-of-fit test for the selection of the copula function. This will be done in the future and will be added to our R package `Sunclarco`.

2.6 Discussion

The current copula methodology only allows the modelling of multivariate survival data that are grouped in clusters of small and equal size. A new formulation for the likelihood of Archimedean copula models for survival data is developed, that allows for clusters of large and variable size. The failure times within a cluster are assumed to be exchangeable and the whole data set is used to estimate a common marginal baseline survival. The survival functions of subjects differ through the incorporation of covariates (possibly time-dependent). For copula members of the PVF family, a closed form expression of the likelihood exists, whereas other choices require numerical integration. We investigated the parametric one-stage and two-stage approach as well as the semi-parametric two-stage approach and derived asymptotic results for the estimators under a reasonable set of conditions. Simulation results show that all three methods work well for cluster sizes ranging from 2 to 50. Even larger clusters can be attained, at the cost of larger computing time. For samples with less than 100 clusters, the two-stage estimation approaches are not recommended since they lead to larger bias and less coverage. As an alternative to the flexible semi-parametric model, a piecewise constant hazard (or, by extension, e.g. splines) can be used while modelling the marginal survival function. This chapter is an extension of the work of Shih and Louis (1995), who derived founding results for bivariate data, and the work of Glidden (2000), who investigated the two-stage semi-parametric model for the Clayton copula, as it describes the use of copula functions for clusters with large and varying cluster size.

Copula model	τ	θ_0	0% censoring Parametric one-stage	0% censoring Parametric two-stage	0% censoring Semi-parametric two-stage	25% censoring Parametric one-stage	25% censoring Parametric two-stage	25% censoring Semi-parametric two-stage	50% censoring Parametric one-stage	50% censoring Parametric two-stage	50% censoring Semi-parametric two-stage
Clayton	0.09	0.2	0.197	0.193	0.191	0.196	0.194	0.194	0.197	0.195	0.195
			(0.043; 93.1%)	(0.042; 89.7%)	(0.045; 85.9%)	(0.047; 92.0%)	(0.047; 90.7%)	(0.049; 88.9%)	(0.055; 92.1%)	(0.055; 91.3%)	(0.056, 88.9%)
	0.2	0.5	0.498	0.486	0.463	0.496	0.489	0.479	0.495	0.491	0.485
			(0.084; 93.2%)	(0.091; 84.3%)	(0.010; 76.8%)	(0.091; 92.9%)	(0.097; 88.1%)	(0.105; 85.3%)	(0.101; 92.7%)	(0.106; 89.8%)	(0.113; 87.8%)
	0.33	1.0	0.997	0.973	0.875	0.996	0.981	0.938	0.997	0.990	0.959
			(0.160; 93.5%)	(0.176; 81.9%)	(0.174; 71.8%)	(0.166; 92.9%)	(0.182; 86.9%)	(0.195; 81.7%)	(0.178; 92.3%)	(0.194; 88.9%)	(0.205; 85.7%)
	0.43	1.5	1.479	1.436	1.226	1.478	1.451	1.365	1.476	1.469	1.402
			(0.234; 92.1%)	(0.253; 83.9%)	(0.229; 63.0%)	(0.240; 92.6%)	(0.262; 87.5%)	(0.273; 81.3%)	(0.252; 91.7%)	(0.278; 88.5%)	(0.287; 84.9%)
G-H	0.2	0.8	0.803	0.801	0.803	0.804	0.802	0.803	0.804	0.802	0.804
			(0.034; 93.6%)	(0.042; 88.6%)	(0.041; 89.0%)	(0.036; 94.3%)	(0.045; 89.0%)	(0.044; 87.7%)	(0.039; 94.9%)	(0.048; 87.8%)	(0.048; 86.0%)
	0.35	0.65	0.656	0.655	0.661	0.656	0.656	0.662	0.656	0.656	0.664
			(0.040; 93.5%)	(0.048; 89.4%)	(0.049; 89.5%)	(0.042; 93.3%)	(0.051; 89.2%)	(0.052; 88.9%)	(0.045; 94.6%)	(0.055; 88.0%)	(0.056; 86.4%)
	0.5	0.5	0.507	0.507	0.521	0.508	0.508	0.522	0.507	0.509	0.525
			(0.040; 93.3%)	(0.046; 91.2%)	(0.047; 90.4%)	(0.041; 93.6%)	(0.048; 90.5%)	(0.050; 90.2%)	(0.043; 94.3%)	(0.051; 88.4%)	(0.054; 86.9%)
	0.8	0.2	0.205	0.208	0.247	0.205	0.209	0.250	0.205	0.211	0.258
			(0.022; 94.7%)	(0.023; 92.3%)	(0.030; 68.5%)	(0.022; 94.2%)	(0.025; 92.6%)	(0.032; 67.2%)	(0.023; 95.1%)	(0.026; 89.7%)	(0.035; 60.9%)

Table 2.2: Simulation results for 50 clusters of varying sizes ranging from 2 to 50

Copula model	τ	θ_0	0% censoring Parametric one-stage	0% censoring Parametric two-stage	0% censoring Semi-parametric two-stage	25% censoring Parametric one-stage	25% censoring Parametric two-stage	25% censoring Semi-parametric two-stage	50% censoring Parametric one-stage	50% censoring Parametric two-stage	50% censoring Semi-parametric two-stage
Clayton	0.09	0.2	0.199	0.198	0.197	0.199	0.198	0.198	0.199	0.199	0.199
			(0.021; 94.4%)	(0.022; 93.3%)	(0.025; 90.8%)	(0.024; 94.4%)	(0.024; 94.0%)	(0.025; 92.6%)	(0.027; 95.3%)	(0.028; 94.8%)	(0.029; 94.0%)
	0.2	0.5	0.498	0.498	0.489	0.498	0.498	0.495	0.498	0.499	0.497
			(0.042; 94.3%)	(0.052; 90.8%)	(0.059; 88.8%)	(0.045; 94.6%)	(0.052; 92.6%)	(0.057; 92.0%)	(0.050; 93.3%)	(0.055; 93.4%)	(0.060; 92.5%)
	0.33	1.0	0.994	0.990	0.953	0.993	0.990	0.978	0.994	0.992	0.984
			(0.079; 95.3%)	(0.101; 90.8%)	(0.110; 86.8%)	(0.083; 94.3%)	(0.099; 92.6%)	(0.108; 90.7%)	(0.088; 95.1%)	(0.102; 93.4%)	(0.109; 92.0%)
	0.43	1.5	1.494	1.484	1.401	1.494	1.488	1.463	1.496	1.491	1.472
			(0.118; 94.2%)	(0.147; 90.1%)	(0.152; 82.2%)	(0.121; 94.4%)	(0.145; 91.2%)	(0.155; 90.4%)	(0.127; 94.8%)	(0.148; 91.8%)	(0.157; 91.1%)
G-H	0.2	0.8	0.802	0.801	0.802	0.801	0.801	0.801	0.801	0.800	0.801
			(0.017; 95.8%)	(0.022; 93.2%)	(0.022; 92.4%)	(0.018; 94.6%)	(0.024; 92.4%)	(0.024; 91.6%)	(0.020; 95.5%)	(0.026; 92.2%)	(0.026; 91.6%)
	0.35	0.65	0.652	0.652	0.654	0.652	0.652	0.654	0.652	0.652	0.655
			(0.020; 95.2%)	(0.025; 93.9%)	(0.026; 93.0%)	(0.021; 95.0%)	(0.027; 93.3%)	(0.028; 93.4%)	(0.022; 95.4%)	(0.030; 92.8%)	(0.030; 93.0%)
	0.5	0.5	0.503	0.503	0.507	0.502	0.503	0.508	0.502	0.503	0.509
			(0.020; 94.8%)	(0.024; 93.8%)	(0.024; 93.4%)	(0.020; 94.7%)	(0.025; 93.7%)	(0.026; 93.3%)	(0.021; 95.1%)	(0.028; 93.2%)	(0.029; 93.0%)
	0.8	0.2	0.201	0.202	0.215	0.201	0.203	0.217	0.201	0.203	0.220
			(0.011; 95.3%)	(0.012; 94.7%)	(0.014; 81.4%)	(0.011; 94.0%)	(0.013; 94.5%)	(0.015; 81.2%)	(0.011; 94.7%)	(0.013; 93.9%)	(0.016; 76.9%)

Table 2.3: Simulation results for 200 clusters of varying sizes ranging from 2 to 50

Copula model	τ	θ_0	0% censoring			25% censoring			50% censoring		
			Parametric one-stage	Parametric two-stage	Semi-parametric two-stage	Parametric one-stage	Parametric two-stage	Semi-parametric two-stage	Parametric one-stage	Parametric two-stage	Semi-parametric two-stage
Clayton	0.09	0.2	0.200	0.200	0.199	0.200	0.200	0.200	0.200	0.199	0.200
			(0.013; 95.4%)	(0.014; 94.2%)	(0.016; 92.6%)	(0.015; 95.3%)	(0.016; 94.1%)	(0.017; 93.7%)	(0.017; 94.9%)	(0.018; 95.2%)	(0.019; 94.8%)
	0.2	0.5	0.501	0.499	0.493	0.501	0.499	0.498	0.501	0.500	0.499
			(0.027; 95.4%)	(0.033; 92.2%)	(0.039; 89.9%)	(0.029; 94.9%)	(0.033; 92.8%)	(0.037; 91.7%)	(0.032; 94.6%)	(0.035; 93.4%)	(0.038; 93.2%)
	0.33	1.0	0.999	0.994	0.973	1.000	0.996	0.990	0.999	0.997	0.992
			(0.050; 94.8%)	(0.065; 91.8%)	(0.072; 89.1%)	(0.053; 94.1%)	(0.064; 92.9%)	(0.070; 92.1%)	(0.056; 94.2%)	(0.065; 93.0%)	(0.070; 93.2%)
	0.43	1.5	1.498	1.496	1.453	1.498	1.497	1.485	1.497	1.498	1.490
			(0.075; 93.8%)	(0.098; 93.4%)	(0.104; 88.8%)	(0.077; 94.3%)	(0.095; 93.7%)	(0.101; 93.4%)	(0.081; 93.5%)	(0.095; 93.4%)	(0.102; 93.9%)
G-H	0.2	0.8	0.800	0.801	0.801	0.801	0.801	0.801	0.801	0.801	0.802
			(0.011; 93.6%)	(0.014; 94.3%)	(0.014; 92.9%)	(0.011; 95.5%)	(0.015; 93.3%)	(0.015; 93.0%)	(0.013; 95.4%)	(0.017; 93.6%)	(0.017; 93.0%)
	0.35	0.65	0.651	0.652	0.653	0.651	0.652	0.653	0.652	0.652	0.654
			(0.013; 95.4%)	(0.016; 95.3%)	(0.017; 95.1%)	(0.013; 95.9%)	(0.017; 93.7%)	(0.018; 93.6%)	(0.014; 94.6%)	(0.019; 93.8%)	(0.020; 92.9%)
	0.5	0.5	0.501	0.502	0.504	0.501	0.502	0.505	0.502	0.502	0.505
			(0.013; 96.9%)	(0.015; 95.0%)	(0.016; 94.8%)	(0.013; 94.9%)	(0.016; 93.8%)	(0.017; 93.4%)	(0.014; 95.8%)	(0.018; 93.9%)	(0.018; 93.6%)
	0.8	0.2	0.201	0.201	0.208	0.201	0.201	0.209	0.201	0.202	0.211
			(0.007; 95.7%)	(0.008; 95.4%)	(0.009; 86.7%)	(0.007; 95.9%)	(0.008; 95.4%)	(0.009; 86.1%)	(0.007; 95.0%)	(0.009; 94.6%)	(0.010; 83.2%)

Table 2.4: Simulation results for 500 clusters of varying sizes ranging from 2 to 50

Chapter 3

Conditions, theorems and proofs in Chapter 2

We state regularity conditions as well as proofs for the consistency and asymptotic normality of the estimators developed in Chapter 2 . For the parametric estimation procedures, the regularity conditions are adapted from Cox and Hinkley (1974) and Lehmann and Casella (1998). To prove consistency and asymptotic normality in the semi-parametric setting, an approach based on empirical processes is required. Therefore the set of conditions is adapted from Spiekerman and Lin (1998) and Othus and Li (2010).

3.1 Conditions for one-stage parametric estimation

The regularity conditions for one-stage parametric estimation are:

(C.1) The parameter space $\mathbf{\Omega}$ has finite dimension, is closed and compact, and the true parameter vector $\boldsymbol{\eta} = (\boldsymbol{\beta}, \theta)$ lies in the interior of $\mathbf{\Omega}$.

(C.2) $E\left[\mathbf{U}_{\beta_j}(\boldsymbol{\beta_0}, \theta_0)\right] = \mathbf{0}$ for $j = 1, \ldots q$ and $U_\theta(\boldsymbol{\beta_0}, \theta_0) = 0$.

(C.3) $E\left[\frac{\partial \log L(\boldsymbol{\beta}_0,\theta_0)}{\partial \eta_i}\cdot\frac{\partial \log L(\boldsymbol{\beta}_0,\theta_0)}{\partial \eta_j}\right]=E\left[-\frac{\partial^2 \log L(\boldsymbol{\beta},\theta)}{\partial \eta_i \partial \eta_j}\right]$ for $i,j=1,\ldots q+1$.

(C.4) The Fisher information matrix $\mathbf{I}=\left[-\frac{\partial^2 \log L(\boldsymbol{\beta},\theta)}{\partial \eta_i \partial \eta_j}\right]_{i,j=1\ldots,q+1}$ is positive definite for all $\boldsymbol{\eta}=(\boldsymbol{\beta},\theta)\in\boldsymbol{\Omega}$.

(C.5) Third order partial derivatives of $\log L(\boldsymbol{\beta},\theta)$ are bounded integrable, i.e., $\left|\frac{\partial^3 \log L(\boldsymbol{\beta},\theta)}{\partial \eta_i \eta_j \eta_k}\right|<M_{ijk}$ for all $\boldsymbol{\eta}\in\Omega$, where $E[M_{ijk}]<\infty$ for $i,j,k=1,\ldots,q+1$.

3.2 Conditions, theorems and proofs for two-stage parametric estimation

The regularity conditions for the first stage of the two-stage parametric estimation procedure, i.e., estimation of the marginal parameters $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_q)'$, are given by (C.6)-(C.9). Denote

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

(C.6) The parameter space $\boldsymbol{\Omega}_1$ has finite dimension, is closed and compact, and the true parameter vector β lies in the interior of $\boldsymbol{\Omega}_1$.

(C.7) $E\left[\mathbf{U}^*_{1,\beta_j}(\boldsymbol{\beta}_0)\right]=\mathbf{0}$ for $j=1,\ldots,q$.

(C.8) The Fisher information matrix

$$\mathbf{I}^*=E\left[-\frac{\partial}{\partial \beta_j}\mathbf{U}^*_{1,\beta_i}(\boldsymbol{\beta})\right]_{i,j=1\ldots,q}\equiv E\left[-\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta})\right]$$

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

(C.9) Second order partial derivatives of $\mathbf{U}^*_{\beta_i}(\boldsymbol{\beta})$ are bounded integrable, i.e. $\left|\frac{\partial^2 \mathbf{U}^*_{\beta_i}(\boldsymbol{\beta})}{\partial \beta_j \beta_k}\right|<M_{ijk}$ for all $\boldsymbol{\beta}\in\boldsymbol{\Omega}_1$, where $E[M_{ijk}]<\infty$ for $i,j,k=1,\ldots,q$.

The regularity conditions for the second stage of the two-stage parametric estimation procedure, i.e., estimation of the association parameter θ, are

(C.10) The parameter space Ω_2 is closed and compact, and the true parameter θ lies in the interior of Θ.

(C.11) $E[U_\theta(\boldsymbol{\beta}_0, \theta_0)] = 0$.

(C.12) $E\left[U_\theta^2(\boldsymbol{\beta}_0, \theta_0)\right] = E\left[-\dfrac{\partial U_\theta(\boldsymbol{\beta}_0, \theta_0)}{\partial \theta}\right]$

(C.13) The Fisher information matrix $I_{\theta\theta} = E\left[-\dfrac{\partial U_\theta(\boldsymbol{\beta}, \theta)}{\partial \theta}\right]$ is positive definite for all θ in Θ.

(C.14) Second order partial derivatives of $U_\theta(\boldsymbol{\beta}, \theta)$ are bounded integrable, i.e. $\left|\dfrac{\partial^2 U_\theta(\boldsymbol{\beta}, \theta)}{\partial \theta^2}\right| < M$ for all $\theta \in \Theta$, where $E[M] < \infty$.

Theorem (2.2.1). *Let $\overline{\theta}$ denote the solution to $U_\theta(\overline{\boldsymbol{\beta}}, \theta) = 0$ and let θ_0 be the true value of the association parameter. Under regularity conditions (C.10)-(C.14) as stated above, $\sqrt{K}(\overline{\theta} - \theta_0)$ converges to a normal distribution with mean zero and variance*

$$\mathrm{Var}(\overline{\theta}) = \frac{1}{\mathrm{I}_{\theta\theta}} + \frac{\mathbf{I}_{\theta\beta}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\beta\theta}}{\mathrm{I}_{\theta\theta}^2}. \tag{2.6}$$

Proof. Let $\boldsymbol{\beta}_0$ denote the true parameter vector for the margins. Expanding the score function $\mathbf{U}_{\boldsymbol{\beta}}^*$ in a Taylor series around $\boldsymbol{\beta}_0$ and evaluating it at $\boldsymbol{\beta} = \overline{\boldsymbol{\beta}}$, we get under regularity conditions of maximum likelihood theory

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

The o_p-notation stands for convergence in probability, i.e., $X_K = o_p(\sqrt{K})$ is defined as $\lim_{K\to\infty} P(|X_K/\sqrt{K}| \geq \varepsilon) = 0$ for every positive ε.

Similarly,

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

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

$$
\begin{aligned}
-\frac{1}{K}\left.\frac{\partial \mathbf{U}^*_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} &= \frac{1}{K}\sum_{i=1}^{K}\left[-\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{U}^*_{i,\boldsymbol{\beta}}(\boldsymbol{\beta}_0)\right] \to \mathbf{I}^* \\
-\frac{1}{K}\left.\frac{\partial U_\theta}{\partial \boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},\theta)=(\boldsymbol{\beta}_0,\theta_0)} &= \frac{1}{K}\sum_{i=1}^{K}\left[-\frac{\partial}{\partial \boldsymbol{\beta}}U_{i,\theta}(\boldsymbol{\beta}_0,\theta_0)\right] \to \mathbf{I}_{\theta\boldsymbol{\beta}} \\
-\frac{1}{K}\left.\frac{\partial U_\theta}{\partial \theta}\right|_{(\boldsymbol{\beta},\theta)=(\boldsymbol{\beta}_0,\theta_0)} &= \frac{1}{K}\sum_{i=1}^{K}\left[-\frac{\partial}{\partial \theta}U_{i,\theta}(\boldsymbol{\beta}_0,\theta_0)\right] \to I_{\theta\theta}.
\end{aligned}
$$

Hence

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

By the central limit theorem, $\frac{1}{\sqrt{K}}\begin{pmatrix} \mathbf{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) \\ U_\theta(\boldsymbol{\beta}_0,\theta_0) \end{pmatrix}$ converges to multivariate normal with mean $\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}$ and variance-covariance matrix $\begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & I_{\theta\theta} \end{pmatrix}$ with $\mathbf{V} = \mathrm{Var}\left(\mathbf{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta}_0)\right) = \mathrm{E}\left[\mathbf{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta}_0)(\mathbf{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta}_0))'\right]$. Thus, $\sqrt{K}\begin{pmatrix} \overline{\boldsymbol{\beta}}-\boldsymbol{\beta}_0 \\ \overline{\theta}-\theta_0 \end{pmatrix}$ converges to multivariate normal with mean vector zero and variance-covariance matrix

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

(Note that $(\mathbf{I}^*)^{-1^T} = (\mathbf{I}^*)^{-1}$ since $(A^{-1})^T = (A^T)^{-1}$ and $(\mathbf{I}^*)^T = \mathbf{I}^*$.)

The lower right element of this matrix is the asymptotic variance of $\sqrt{K}(\overline{\theta}-\theta_0)$ and we denote this by σ^2.

$$
\sigma^2 = \frac{1}{I_{\theta\theta}} + \frac{\mathbf{I}_{\theta\boldsymbol{\beta}}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\boldsymbol{\beta}\theta}}{I^2_{\theta\theta}}.
$$

□

3.3 Conditions, theorems and proofs for two-stage semi-parametric estimation

Before we prove Theorems 2.2.2 and 2.2.3, we first introduce some notation.

$$
\begin{aligned}
Y_{ij}(t) &= I_{\{X_{ij}\geq t\}} \\
\check{\Lambda}(t) &= \int_0^t \frac{d\sum_{i=1}^K\sum_{j=1}^{n_i}\delta_{ij}I_{\{X_{ij}\leq u\}}}{\sum_{i=1}^K\sum_{j=1}^{n_i}Y_{ij}(u)\exp[\check{\boldsymbol{\beta}}'\boldsymbol{Z}_{ij}(u)]} = \sum_{i=1}^K\sum_{j=1}^{n_i}\frac{\delta_{ij}I_{\{X_{ij}\leq t\}}}{\sum_{k=1}^K\sum_{l=1}^{n_k}I_{\{X_{kl}\leq X_{ij}\}}\exp[\check{\boldsymbol{\beta}}'\boldsymbol{Z}_{kl}(X_{ij})]} \\
H_{ij} &= \exp\left(-\int_0^\tau Y_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}(u)]d\Lambda(u)\right) \\
H_{ij}^0 &= \exp\left(-\int_0^\tau Y_{ij}(u)\exp[\boldsymbol{\beta}_0'\boldsymbol{Z}_{ij}(u)]d\Lambda_0(u)\right) \\
\check{H}_{ij} &= \exp\left(-\int_0^\tau Y_{ij}(u)\exp[\check{\boldsymbol{\beta}}'\boldsymbol{Z}_{ij}(u)]d\check{\Lambda}(u)\right) \\
H_{ij}(t) &= \exp\left(-\int_0^\tau Y_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}(u)]d(\Lambda+t(\Gamma-\Lambda))(u)\right)
\end{aligned}
$$

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

$$
\begin{aligned}
L(\theta;\boldsymbol{\beta},\Lambda) &= \prod_{i=1}^K L_i(\theta;\boldsymbol{\beta},\Lambda) \\
&= \prod_{i=1}^K\left(\prod_{j=1}^{n_i}\left[\frac{1}{\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)}\right]^{\delta_{ij}}\right)\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right) \\
l_K(\theta) &= K^{-1}\log L(\theta;\boldsymbol{\beta},\Lambda) \\
&= K^{-1}\sum_{i=1}^K\left\{\sum_{j=1}^{n_i}\delta_{ij}\log\left[\frac{1}{\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)}\right] + \log\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)\right\} \\
l_{K0}(\theta) &= K^{-1}\log L(\theta;\boldsymbol{\beta}_0,\Lambda_0) \\
\check{l}_K(\theta) &= K^{-1}\log L(\theta;\check{\boldsymbol{\beta}},\check{\Lambda}) \\
U_K(\theta) &= \frac{\partial l_K(\theta)}{\partial\theta} = K^{-1}\frac{\partial\log L(\theta;\boldsymbol{\beta},\Lambda)}{\partial\theta} \\
&= K^{-1}\sum_{i=1}^K\left\{\sum_{j=1}^{n_i}\delta_{ij}\left[\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)\right]\frac{\partial}{\partial\theta}\left[\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)\right]^{-1}\right. \\
&\qquad \left. + \left[\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)\right]^{-1}\frac{\partial}{\partial\theta}\left[\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)\right]\right\}
\end{aligned}
$$

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

We copy the following notation from Spiekerman and Lin (1998) where $\boldsymbol{a}^{\otimes 0} = 1, \boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$ and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}'\boldsymbol{a}$:

$$\boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t) = K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i} Y_{ij}(t)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}(t)]\boldsymbol{Z}_{ij}(t)^{\otimes r}, \qquad \boldsymbol{s}^{(r)} = E\left[\boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t)\right] \quad (r = 0, 1, 2)$$
$$\boldsymbol{E}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)}, \qquad \boldsymbol{e}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)}$$
$$\boldsymbol{V}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}^{(2)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - \boldsymbol{E}(\boldsymbol{\beta}, t)^{\otimes 2}, \qquad \boldsymbol{v}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} - \boldsymbol{e}(\boldsymbol{\beta}, t)^{\otimes 2}$$

Assume the following regularity conditions where $\tau > 0$ is a constant (e.g. end of study time).

(C.15) $\boldsymbol{\beta}$ is in a compact subset of $\mathbb{R}^p$.

(C.16) $\Lambda(\tau) < \infty$.

(C.17) $\theta \in \nu$, where ν is a compact subset of Θ.

(C.18) $P(C_{ij} \geq t \quad \forall t \in [0, \tau]) > \delta_c > 0$ for $i = 1, \ldots, K$ and $j = 1, \ldots, n_i$.

(C.19) Write $\boldsymbol{Z}_{ij}(t) = \{Z_{ij1}(t), \ldots, Z_{ijp}(t)\}$. For $i = 1, \ldots, K, j = 1 \ldots, n_i, k = 1, \ldots, p$

$$|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(t)| \leq B_Z < \infty \quad \text{a.s. for some constant } B_Z.$$

(C.20) $E\left[\log \frac{L_i(\theta_1; \boldsymbol{\beta}, \Lambda)}{L_i(\theta_2; \boldsymbol{\beta}, \Lambda)}\right]$ exists for all $\theta_1, \theta_2 \in \Theta, i = 1, \ldots, K$.

(C.21) $\boldsymbol{A} = \int_0^\tau \boldsymbol{v}(\boldsymbol{\beta}_0, u)s^{(0)}(\boldsymbol{\beta}_0, u)d\Lambda_0(u)$ is positive definite.

Theorem (2.2.2). *Under regularity conditions (C.15)-(C.21), the two-stage semi-parametric estimator $(\check{\theta}; \check{\boldsymbol{\beta}}, \check{\Lambda})$ is a consistent estimator for $(\theta_0; \boldsymbol{\beta}_0, \Lambda_0)$.*

Proof. The results for $\check{\boldsymbol{\beta}}$ and $\check{\Lambda}$ follow from arguments along the lines of Spiekerman and Lin (1998). We will now show the consistency of $\check{\theta}$ using ideas of Othus and Li (2010).

To account for the fact that plug-in estimates of $\boldsymbol{\beta}$ and Λ are used in the likelihood for θ, we will need to take a Taylor series expansion of the likelihood of θ around $\boldsymbol{\beta}_0$ and Λ_0. Since Λ_0 is an unspecified function, this expansion will need to include a functional expansion term. An expansion using Hadamard derivatives is appropriate for this situation. Hereto, we must verify that the log-likelihood $l_K(\theta)$ is Hadamard differentiable with respect to Λ.

We find the Hadamard derivative of l_K w.r.t. Λ at $\Gamma - \Lambda \in BV[0,\tau]$ by taking the derivative of $K^{-1}\log L(\theta;\boldsymbol{\beta},\Lambda + t(\Gamma-\Lambda))$ with respect to t en then putting $t=0$:

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

where

$$\zeta_K(\theta;\Lambda)(u) = K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i} D_{ij}^l Y_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}(u)]$$

and

$$D_{ij}^l = \left\{\delta_{ij}\frac{-\varphi_\theta''(\varphi_\theta^{-1}(H_{ij}))}{\varphi_\theta'(\varphi_\theta^{-1}(H_{ij}))} + \frac{\varphi_\theta^{(d_i+1)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)}{\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)}\right\}\frac{-H_{ij}}{\varphi_\theta'(\varphi_\theta^{-1}(H_{ij}))}.$$

The derivative of $l_K(\theta)$ w.r.t. $\boldsymbol{\beta}$ is

$$\zeta_K(\theta;\boldsymbol{\beta}) = K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i} D_{ij}^l\left(\int_0^\tau Y_{ij}(u)\boldsymbol{Z}_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}(u)]d\Lambda(u)\right).$$

To prove consistency for $\check{\theta}$, we will require $||\zeta_K(\theta;\Lambda)||_\infty$ and $||\zeta_K(\theta;\boldsymbol{\beta})||$ to be bounded. This can be obtained when the common factor $||D_{ij}^l||_\infty$ is bounded and also the terms unique to $\zeta_K(\theta;\boldsymbol{\beta})$ and $\zeta_K(\theta;\Lambda)$ have to be bounded. This requirement is not too restrictive, e.g. for the Clayton copula we have

$$||D_{ij}^l||_\infty = \left|\left|\delta_{ij}(1+\theta) - \frac{(1+d_i\theta)H_{ij}^{-\theta}}{\left(-n_i+1+\sum_{j=1}^{n_i}H_{ij}^{-\theta}\right)}\right|\right|_\infty.$$

Due to the definition of H_{ij} and condition (C.16), this expression is bounded. By condition (C.19),

$||Y_{ij}\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}]||_\infty$ and $\left|\left|\int_0^\tau Y_{ij}(u)\boldsymbol{Z}_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij(u)}d\Lambda(u)]\right|\right|$ are bounded.

An expansion of $\check{l}_K(\theta)$ around $\boldsymbol{\beta}_0$ and Λ_0 can be written as

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

Another (intuitive) notation is:

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

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

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

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

In order to prove $\check{\theta}$ is consistent we will 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 $l_{K0}(\theta)$:

$$\sup_{\theta\in\nu} |\check{l}_K(\theta) - E[l_{K0}(\theta)]| = o_p(1). \tag{3.1}$$

This can be shown as follows:

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

Due to the law of large numbers, for fixed θ,

$$l_{K0}(\theta) - E[l_{K0}(\theta)] \xrightarrow{p} 0. \tag{3.2}$$

Since $||\zeta_K(\theta;\boldsymbol{\beta})||$ is bounded, say $||\zeta_K(\theta;\boldsymbol{\beta})|| \le M_1$, we have

$$\sup_{\theta\in\nu} \left|\zeta_K(\theta;\boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right| \le M_1||\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||. \tag{3.3}$$

Since $||\zeta_K(\theta;\Lambda)(u)||_\infty$ is bounded, say $||\zeta_K(\theta;\Lambda)(u)||_\infty \le M_2$, we have

$$\sup_{\theta\in\nu} \left|\int_0^\tau \zeta_K(\theta;\Lambda)(t)d(\check{\Lambda} - \Lambda_0)(t)\right| \le M_2||\check{\Lambda} - \Lambda_0||_\infty. \tag{3.4}$$

Therefore

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

Using (3.2), the consistency of $\check{\boldsymbol{\beta}}$, the uniform consistency of $\check{\Lambda}$ and the fact that $R = o_p(1)$, we get

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

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

$$E[l_{K0}(\theta)] - E[l_{K0}(\theta_0)] < 0. \tag{3.5}$$

Due to independence between clusters and the fact that all lower dimensional copulas can be regarded as margins of the highest dimensional copula, the log-likelihood $l_K(\theta)$ can be written as a sum of i.i.d. random variables

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

with

$$\begin{aligned} L_i &= (-1)^{d_i}\frac{\partial^{d_i}}{\partial\{\delta_{ij}=1\}}S(x_{i1},\ldots,x_{in_i}) \\ &= \left(\prod_{j=1}^{n_i}\left[\frac{1}{\varphi'_\theta\left(\varphi_\theta^{-1}\left(e^{-\Lambda(x_{ij})}\right)\right)}\right]^{\delta_{ij}}\right)\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}\left(e^{-\Lambda(x_{ij})}\right)\right) \end{aligned}$$

where $\partial\{\delta_{ij}=1\}$ is the set of uncensored individuals in cluster i.

Take $\theta \neq \theta_0$. The law of large numbers, Jensen's inequality and condition (C.20) imply that

$$\begin{aligned} \lim_{K\to\infty} l_{K0}(\theta) - l_{K0}(\theta_0) &= E[l_{K0}(\theta)] - E[l_{K0}(\theta_0)] \\ &= E\left[K^{-1}\sum_{i=1}^{K}\log L_i(\theta;\boldsymbol{\beta}_0,\Lambda_0)\right] - E\left[K^{-1}\sum_{i=1}^{K}\log L_i(\theta_0;\boldsymbol{\beta}_0,\Lambda_0)\right] \\ &= E\left[\log L_1(\theta;\boldsymbol{\beta}_0,\Lambda_0) - \log L_1(\theta_0;\boldsymbol{\beta}_0,\Lambda_0)\right] \\ &= E\left[\log\frac{L_1(\theta;\boldsymbol{\beta}_0,\Lambda_0)}{L_1(\theta_0;\boldsymbol{\beta}_0,\Lambda_0)}\right] \\ &\leq \log E\left[\frac{L_1(\theta;\boldsymbol{\beta}_0,\Lambda_0)}{L_1(\theta_0;\boldsymbol{\beta}_0,\Lambda_0)}\right] \\ &= \log 1 \\ &= 0. \end{aligned}$$

The before last equality results from $L_1(\theta;\boldsymbol{\beta}_0,\Lambda_0)$ being the contribution of cluster 1 to the likelihood $L(\theta;\boldsymbol{\beta}_0,\Lambda_0)$, which is the joint density function of

$(x_{11}, \ldots, x_{1n_1}; \delta_{11}, \ldots, \delta_{1n_1})$.
Since $\check{\theta}$ maximizes $\check{l}_K(\theta)$, (3.1) implies that

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

$$\Downarrow$$

$$E[l_{K0}(\theta_0)] \leq \check{l}_K(\check{\theta}) + o_p(1).$$

Subtract $E[l_{K0}(\check{\theta})]$ from each side of the inequality to write

$$E[l_{K0}(\theta_0)] - E[l_{K0}(\check{\theta})] \leq \check{l}_K(\check{\theta}) - E[l_{K0}(\check{\theta})] + o_p(1) \leq \sup_{\theta \in \Theta} |\check{l}_K(\theta) - E[l_{K0}(\theta)]| + o_p(1) = o_p(1). \tag{3.6}$$

Now take θ such that $|\theta - \theta_0| \geq \varepsilon$ for any fixed $\varepsilon > 0$. By (3.5) there must exist some $\gamma_\varepsilon > 0$ such that

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

It follows that

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

Equation (3.6) implies that

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

Therefore

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

which proves the consistency of $\check{\theta}$. □

Theorem (2.2.3). *Under regularity conditions (C.15)-(C.21), $\sqrt{K}(\check{\theta} - \theta_0)$ converges to a normal distribution with mean zero and variance*

$$\frac{\mathrm{Var}(\Xi_1)}{W(\theta_0)^2}.$$

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

$$\hat{U}_K(\hat{\theta}) = \hat{U}_K(\theta_0) + (\hat{\theta} - \theta_0) \left. \frac{\partial \hat{U}_K}{\partial \theta} \right|_{\theta = \theta^*} \tag{3.7}$$

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

$$\sqrt{K}(\hat{\theta} - \theta_0) = \frac{\sqrt{K}\hat{U}_K(\theta_0)}{-\left.\frac{\partial \hat{U}_K}{\partial \theta}\right|_{\theta=\theta^*}}. \tag{3.8}$$

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

$$\left.\frac{\partial \hat{U}_K}{\partial \theta}\right|_{\theta=\theta^*} \xrightarrow{P} W(\theta_0) = \lim_{K\to\infty} \left.\frac{\partial U_K}{\partial \theta}\right|_{\theta=\theta_0} \quad \text{(Fisher information)}.$$

We will show that the score equation $\hat{U}_K(\theta_0)$ in the numerator of (3.8) follows a normal distribution. Hereto we need a Taylor series expansion of $\hat{U}_K(\theta_0)$ around $\boldsymbol{\beta}_0$ and Λ_0. Because Λ_0 is an unspecified function, we will use the Hadamard derivative of $U_K(\theta_0)$ w.r.t. Λ at $\Gamma - \Lambda \in BV[0, \tau]$.

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

where

$$\xi_K(\theta; \Lambda)(u) = K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i} D_{ij}^U Y_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}(u)]$$

and

$$\begin{aligned} D_{ij}^U = \Bigg\{ & \delta_{ij}\frac{\varphi_\theta''\left(\varphi_\theta^{-1}(H_{ij})\right)}{\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)}\frac{\partial}{\partial\theta}\left[\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)\right]^{-1} \\ & + \delta_{ij}\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)\frac{\partial}{\partial\theta}\left[-\frac{\varphi_\theta''\left(\varphi_\theta^{-1}(H_{ij})\right)}{\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)^3}\right] \\ & - \frac{\varphi_\theta^{(d_i+1)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)}{\left[\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)\right]^2}\frac{1}{\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)}\frac{\partial}{\partial\theta}\left[\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)\right] \\ & + \frac{1}{\varphi_\theta^{(d_i)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)}\frac{\partial}{\partial\theta}\left[\frac{\varphi_\theta^{(d_i+1)}\left(\sum_{j=1}^{n_i}\varphi_\theta^{-1}(H_{ij})\right)}{\varphi_\theta'\left(\varphi_\theta^{-1}(H_{ij})\right)}\right]\Bigg\}(-H_{ij}). \end{aligned}$$

The derivative of $U_K(\theta)$ w.r.t. $\boldsymbol{\beta}$ is given by

$$\xi_K(\theta; \boldsymbol{\beta}) = K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i} D_{ij}^U \int_0^\tau Y_{ij}(u)\boldsymbol{Z}_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}(u)]d\Lambda(u).$$

We require $||\xi_K(\theta;\Lambda)||_\infty$ and $||\xi_K(\theta;\boldsymbol{\beta})||$ to be bounded. By condition (C.19), the terms unique to $\xi_K(\theta;\Lambda)$ and $\xi_K(\theta;\boldsymbol{\beta})$, i.e.

$$||Y_{ij}\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij}]||_\infty \quad \text{and} \quad \left|\left|\int_0^\tau Y_{ij}(u)\boldsymbol{Z}_{ij}(u)\exp[\boldsymbol{\beta}'\boldsymbol{Z}_{ij(u)}d\Lambda(u)]\right|\right|$$

are bounded. The common term $||D_{ij}^U||_\infty$ is also bounded.

A Taylor series expansion of $\hat{U}_K(\theta_0)$ around $\boldsymbol{\beta}_0$ and Λ_0 gives

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

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 too. Therefore

$$\sqrt{K}\hat{U}_K(\theta_0) = \sqrt{K}\left(U_{K0}(\theta_0) + \xi(\theta_0;\boldsymbol{\beta}_0)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \int_0^\tau \xi(\theta_0;\Lambda_0)(t)d[\check{\Lambda}(t) - \Lambda_0(t)]\right) + o_p(1). \tag{3.9}$$

By Spiekerman and Lin (1998)

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

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

$$\boldsymbol{w}_{i\cdot} = \sum_{j=1}^{n_i}\int_0^\tau \{\boldsymbol{Z}_{ij}(u) - E(\boldsymbol{\beta}_0,u)\}dM_{ij}(u)$$

with

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

They also showed that

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

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

$$E\left[\Psi_1(t)^2\right]$$

with

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

and

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

That's why

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

The central limit theorem implies that $\sqrt{K}\hat{U}_K(\theta_0)$ converges to a normally distributed random variable with mean zero and variance equal to the variance of Ξ_1.

Thus we have

$$\sqrt{K}(\hat{\theta} - \theta_0) = \frac{\sqrt{K}\hat{U}_K(\theta_0)}{-\left.\frac{\partial \hat{U}_K}{\partial\theta}\right|_{\theta=\theta^*}} \tag{3.10}$$

where

$$\sqrt{K}\hat{U}_K(\theta_0) \xrightarrow{D} N(0, \mathrm{Var}(\Xi_1))$$

and

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

By Slutsky's theorem, $\sqrt{K}(\hat{\theta} - \theta_0)$ converges to a normal distribution with mean zero and variance equal to

$$\frac{\mathrm{Var}(\Xi_1)}{W(\theta_0)^2}.$$

The variance of Ξ_1 (note that $\mathrm{Var}(\Xi_1) = \mathrm{E}[\Xi_1^2]$) can be estimated by $K^{-1}\sum_{i=1}^K \hat{\Xi}_i^2$ where $\hat{\Xi}_i$ is obtained from Ξ_i replacing parameter values by their estimators.

$W(\theta_0)$ can be estimated by the (minus) derivative of the pseudo score function $\hat{U}_K(\theta)$, evaluated in $\hat{\theta}$. □

Chapter 4

R package Sunclarco

The methods for survival data that are grouped in unbalanced clusters using Archimedean copulas, that were developed and proven in Chapters 2 and 3, are implemented in the R package **Sunclarco**. The package covers one-stage and two-stage estimation procedures for a selection of copula functions and marginal survival distributions.

We illustrate the **Sunclarco** package with the insemination data set which contains the time to first insemination of cows clustered in herds (see Section 1.3.2). We notify the reader that the exemplifying code in this chapter is based on the development version of the package, which is available on R-Forge (https://r-forge.r-project.org/), and thus subject to change.

```
R> R.Version()[["version.string"]]
[1] "R version 3.2.5 (2016-04-14)"

R> library(Sunclarco)
```

The insemination dataset is available in the package via the command **data("insem")** and its first 6 lines are given by

```
R> head(insem)
```

```
  Cowid Time Status Herd Parity
1     1   82      1    1      0
2     2   80      1    1      0
3     3  152      1    1      0
4     4  287      1    1      1
5     5  236      1    1      0
6     6   83      1    1      0
```

The fifth line reads as: the cow with cow identification number 5 was inseminated at time 236. It is a multiparous cow that belongs to herd (farm) 1.

The **CopulaModel_1stage** command fits an Archimedean copula model with parametric baseline to the data, using the one-stage estimation procedure. It requires the following input:

data: name of the dataset

time: name of the column including the survival times

cluster: the cluster variable's name

status: name of the column including the censoring status

covariates: a vector containing all covariates' names

marginal: the marginal baseline distribution of the survival times (can be either **"Weibull"** or **"PiecewiseExp"**)

copula: either **"Clayton"** or **"GH"**

Initial values for optimization can be passed via **init.values**. A default is provided, using the results the independence working model for the marginal parameters and an intermediate value for the association parameter.

To obtain the results from Table 2.1, we will first fit the model with a Weibull baseline distribution and a Clayton copula.

```
mod_weibCL <- CopulaModel_1stage(data=insem,time="Time",
+   status="Status",cluster="Herd",covariates="Heifer",
+   marginal="Weibull",copula="Clayton")
```

This command results in the log-likelihood of the model, parameter estimates and standard errors, an estimate of Kendall's tau and the variance-covariance matrix of all parameters.

```
R> str(mod_weibCL)

List of 5
 $ Parameters    :'data.frame': 4 obs. of  2 variables:
  ..$ Estimates     : num [1:4] 0.000881 1.470265 0.212645 -0.082131
  ..$ StandardErrors: num [1:4] 6.82e-05 1.41e-02 1.50e-02 1.73e-02
 $ Kendall_Tau   : Named num [1:2] 0.0961 0.00612
  ..- attr(*, "names")= chr [1:2] "Estimates" "StandardErrors"
 $ ParametersCov : num [1:4, 1:4] 4.65e-09 -8.59e-07 2.14e-07
                                  -6.43e-08 -8.59e-07 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:4] "lambda" "rho" "theta" "beta_Heifer"
  .. ..$ : chr [1:4] "lambda" "rho" "theta" "beta_Heifer"
 $ logllh        : num -54930
 $ parameter.call:List of 8
  ..$ time       : chr "Time"
  ..$ status     : chr "Status"
  ..$ clusters   : chr "Herd"
  ..$ covariates : chr "Heifer"
  ..$ init.values: Named num [1:4] 0.0015 1.3439 0.5000 -0.0657
  .. ..- attr(*, "names")= chr [1:4] "lambda" "rho" "theta" "beta_Heifer"
  ..$ marginal   : chr "Weibull"
  ..$ copula     : chr "Clayton"
  ..$ n.piecewise: num 20
 - attr(*, "class")= chr "Sunclarco"
```

The most important information is summarized by the **summary** command:

```
R> summary(mod_weibCL)

Execution Time: 1.859906 mins
Copula: Clayton
Marginal survival distribution: Weibull
```

```
Loglikelihood: -54929.69

                  ESTIMATE   SE
lambda         0.000880872   6.820698e-05
rho            1.470335455   1.412170e-02
beta_Heifer   -0.082447868   1.730574e-02
theta          0.212409846   1.496303e-02

Kendall's Tau  0.096008362    0.006113899
```

For piecewise exponential margins, the range of the time variable is split up into 20 intervals by default, but this can be changed through the **n.pieces** option. The larger the number of cutpoints, the longer it will take to compute all estimates and standard errors.

```
R> mod_pweCL <- CopulaModel_1stage(data=insem,time="Time",status="Status",
+   cluster="Herd",covariates="Heifer",
+   marginal="PiecewiseExp",copula="Clayton")

R> summary(mod_pweCL)

Execution Time: 11.54703 mins
Copula: Clayton
Marginal survival distribution: Piecewise exponential (20 intervals)
Loglikelihood: -54829.0

                ESTIMATE   SE
lambda1       0.002694161   0.0001725595
lambda2       0.005415180   0.0003344959
lambda3       0.006911665   0.0004277045
lambda4       0.007529730   0.0004598037
lambda5       0.008336969   0.0005187667
lambda6       0.009163350   0.0005622532
lambda7       0.010229630   0.0006182957
lambda8       0.010758852   0.0006527896
lambda9       0.011540465   0.0007007856
lambda10      0.010582665   0.0006485056
```

```
lambda11      0.011615798   0.0006915840
lambda12      0.012473235   0.0007760987
lambda13      0.012451307   0.0007577423
lambda14      0.013168029   0.0008240393
lambda15      0.011932527   0.0007523800
lambda16      0.011947366   0.0007586338
lambda17      0.012984590   0.0008396998
lambda18      0.011850726   0.0007995940
lambda19      0.011569655   0.0008124611
lambda20      0.011210448   0.0008354135
beta_Heifer  -0.069862056   0.0158143980
theta         0.351739438   0.0343196550

Kendall's Tau 0.14956565     0.01241065
```

In Section 2.4, we showed that the one-stage parametric procedure is preferred over two-stage parametric estimation when the number of clusters is relatively small, i.e., about 50 or lower. When the number of clusters increases, the two-stage procedure is less time-consuming and yields nearly the same results as the one-stage procedure if the parametric assumption for the baseline holds. In the two-stage estimation procedure, the additional **"Cox"** option for the **marginal** input variable is available. This option must be called if the user does not want to specify the baseline distribution. The user must however be aware that even though semi-parametric estimation offers more flexibility, the parametric estimation will be more powerful if the form of the baseline hazard is somehow known in advance. Also, the estimation technique is more simple. To compute standard errors for a semi-parametric model, the grouped jackknife procedure is used, which takes a lot of time when the number of clusters is large. In the current version of **Sunclarco**, the grouped jackknife procedure is also used for the estimation of standard errors in the two-stage parametric procedure with piecewise exponential margins. We are planning to speed up the two-stage parametric procedure with piecewise exponential margins by making use of the variance expression (2.6), in order to have a flexible and fast alternative to semi-parametric model fitting.

As an example of the two-stage estimation, we fit the semi-parametric model with the Gumbel-Hougaard copula. The reported log-likelihood is the log-likelihood from the second stage of the estimation procedure.

```
R> mod_coxGH <- CopulaModel_2stage(data=insem,time="Time",status="Status",
+   cluster="Herd",covariates="Heifer",
+   marginal="Cox",copula="GH")


R> summary(mod_coxGH)

Execution Time: 15.36016 mins
Copula: Gumbel-Hougaard
Marginal survival distribution: Cox model
Loglikelihood: -271.3222

                ESTIMATE     SE
beta_Heifer  -0.06034839     0.02096202
theta          0.7904656     0.01621859

Kendall's Tau  0.2095344     0.02096202
```

In upcoming releases, we will gradually extend the possible choices for the marginal survival function and copula function. We also plan to further equip the `Sunclarco` package with p-values of estimated covariate effects, predictions and tools for plotting and model selection.

Chapter 5

Investigating the correlation structure of quadrivariate udder infection times through Archimedean copulas

The correlation structure imposed on multivariate time to event data is often of a simple nature, such as in the shared frailty model where pairwise correlations between event times in a cluster are all the same. In modelling the infection times of the four udder quarters clustered within the cow, more complex correlation structures are possibly required, and if so, such more complex correlation structures give more insight in the infection process.

In this chapter, we will choose a marginal approach to study more complex correlation structures, therefore leaving the modelling of marginal distributions unaffected by the association parameters. The dependency of failure times will be induced through copula functions. The methods are shown for (mixtures of) the Clayton copula, but can be generalized to mixtures of Archimedean copulas for which the nesting conditions are met (McNeil, 2008; Hofert, 2011).

5.1 Introduction

Time to event data are often clustered and different techniques have been developed to cope with the clustering in the data. Most commonly used approaches only accommodate a simple association structure between the event times in a cluster. For instance, the underlying assumption of the shared frailty model is that the correlation between any two event times is the same (Duchateau and Janssen, 2008). The correlated frailty model allows more complex structures, but has mostly been used to model bivariate survival data (Wienke, 2011), and the extension of the correlated frailty model based on the gamma density function imposes quite a few restrictions on the correlation structure.

An alternative modelling technique is based on copula functions. Copulas have also been mostly used for bivariate survival data, and in the case clusters were larger, the development was most often also restricted to simple association structures.

The data set studied here warrants the development of more complex association structures. We investigate the appropriateness of different association structures for the quadrivariate udder quarter infection times clustered in the cow. It was shown in previous analyses using frailty models with a simple association structure that strong correlation exists between the infection times within an udder (Goethals *et al.*, 2009; Ampe *et al.*, 2012). The udder quarters, however, can be ranked in space, and special correlation structures can therefore be proposed. For instance, it is biologically plausible that infection times of left and right udder quarters in front are more correlated than the right front and rear udder quarters. Because the distance between two front (or rear) parts is smaller than the distance between one front and one rear part, it is likely that bacteria will spread more easily from left to right, than from front to rear (or vice versa). The method proposed here provides the tools to test such biologically plausible hypotheses. The findings have large impact on the prevention of infections in udder quarters, as large correlations could signify that bacteria are spread from one udder quarter to the next, which could be prevented with proper hygienic measures.

In this chapter, we will use hierarchical Archimedean copula models as it allows us to impose the correlation structures with biological relevance to the quadrivariate udder quarter infection times. One challenge in using these types of models in multivariate survival data typically is that one needs to calculate all possible first and higher order partial derivatives of the joint survival function, which can get complicated if you allow for

hierarchical structures. In Section 5.2, we discuss the infection time data in cow udders and the general construction of the likelihood function. In Section 5.3, we introduce models with different correlation structures and also discuss the choice of the baseline hazard function and the one- and two stage approach to model fitting. In Section 5.4, we describe the results of the different models and compare them with each other. There is a higher level of association within the two front and the two rear udder parts, than between pairs where one part is located front and one is located rear. There is no difference in association between infection times in multiparous and primiparous cows. Size and power calculations are performed in Section 5.5.

5.2 The general likelihood function for a cow udder

We investigate the correlation structure between the times to infection of the four udder quarters nested in a cow. The data set was described in Section 1.3.1. We use the schematic representation described by the diagram in Figure 5.1.

Let K be the number of cows ($i = 1, \ldots, K$). In each udder, we denote the lifetime

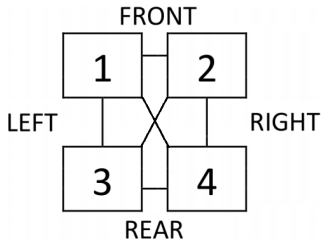


Figure 5.1: Schematic representation of a cow udder.

for the different parts by a positive random variable T_{ij}, $j = 1, \ldots, 4$. For each cow, we assume that there is an independent random censoring variable C_i such that under a right censoring scheme, the observed quantities are given by

$$\begin{matrix} X_{ij} = \min(T_{ij}, C_i) \\ \delta_{ij} = I(T_{ij} \leq C_i) \end{matrix} \quad , i = 1, \ldots, K, \quad j = 1, \ldots, 4.$$

The risk of infection may also depend on a set of covariates $\boldsymbol{Z}_{ij}$, which are possibly time-varying. As denoted in the previous section, we will consider the parity as a covariate:

$$Z_i = \begin{cases} 1 & \text{for primiparous cows} \\ 0 & \text{for multiparous cows.} \end{cases}$$

We denote the (possibly unobserved) time until infection for the different udder quarters of cow i by $(t_{i1}, t_{i2}, t_{i3}, t_{i4})$. The likelihood contribution of a cluster of size 4 is one term out of 16 possibilities. If cow i has four infected udder parts, its contribution to the likelihood is the joint density function of the infection times $f(t_{i2}, t_{i2}, t_{i3}, t_{i4}|Z_i)$. If the cow has only one infected udder part, we need to take the derivative of the joint survival function with respect to that event time, and so on. A general expression for the full likelihood is given by (5.1).

$$\begin{aligned}
&\prod_{i=1}^{K} (f(x_{i2}, x_{i2}, x_{i3}, x_{i4}|Z_i))^{\delta_{i1}\delta_{i2}\delta_{i3}\delta_{i4}} \\
&\quad\times \left(-\frac{\partial S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}}\right)^{\delta_{i1}(1-\delta_{i2})(1-\delta_{i3})(1-\delta_{i4})} \left(-\frac{\partial S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i2}}\right)^{(1-\delta_{i1})\delta_{i2}(1-\delta_{i3})(1-\delta_{i4})} \\
&\quad\times \left(-\frac{\partial S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i3}}\right)^{(1-\delta_{i1})(1-\delta_{i2})\delta_{i3}(1-\delta_{i4})} \left(-\frac{\partial S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i4}}\right)^{(1-\delta_{i1})(1-\delta_{i2})(1-\delta_{i3})\delta_{i4}} \\
&\quad\times \left(\frac{\partial^2 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}\partial x_{i2}}\right)^{\delta_{i1}\delta_{i2}(1-\delta_{i3})(1-\delta_{i4})} \left(\frac{\partial^2 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i2}\partial x_{i3}}\right)^{(1-\delta_{i1})\delta_{i2}\delta_{i3}(1-\delta_{i4})} \\
&\quad\times \left(\frac{\partial^2 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}\partial x_{i3}}\right)^{\delta_{i1}(1-\delta_{i2})\delta_{i3}(1-\delta_{i4})} \left(\frac{\partial^2 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}\partial x_{i4}}\right)^{\delta_{i1}(1-\delta_{i2})(1-\delta_{i3})\delta_{i4}} \\
&\quad\times \left(\frac{\partial^2 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i2}\partial x_{i4}}\right)^{(1-\delta_{i1})\delta_{i2}(1-\delta_{i3})\delta_{i4}} \left(\frac{\partial^2 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i3}\partial x_{i4}}\right)^{(1-\delta_{i1})(1-\delta_{i2})\delta_{i3}\delta_{i4}} \\
&\quad\times \left(-\frac{\partial^3 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}\partial x_{i2}\partial x_{i3}}\right)^{\delta_{i1}\delta_{i2}\delta_{i3}(1-\delta_{i4})} \left(-\frac{\partial^3 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i2}\partial x_{i3}\partial x_{i4}}\right)^{(1-\delta_{i1})\delta_{i2}\delta_{i3}\delta_{i4}} \\
&\quad\times \left(-\frac{\partial^3 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}\partial x_{i3}\partial x_{i4}}\right)^{\delta_{i1}(1-\delta_{i2})\delta_{i3}\delta_{i4}} \left(-\frac{\partial^3 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}\partial x_{i2}\partial x_{i4}}\right)^{\delta_{i1}\delta_{i2}(1-\delta_{i3})\delta_{i4}} \\
&\quad\times f(S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i))^{(1-\delta_{i1})(1-\delta_{i2})(1-\delta_{i3})(1-\delta_{i4})}
\end{aligned} \tag{5.1}$$

5.3 Different models for the association structure

In this section, we go from models with simple correlation structure to models with more complex correlation structure. As the models are nested, likelihood ratio testing can be used to test whether such more complex correlation structures are required. Association structures are expressed in terms of the copula function, i.e., correlation is introduced through the copula function that links the marginal survival functions into the joint survival function. We will use the shorthand notation S_j for the marginal survival functions $S_j(x_j|Z)$, $j = 1, 2, 3, 4$.

5.3.1 No clustering (Model 0)

In the case of independence between the udder quarter infection times in a cow, the joint survival function is given by

$$S(x_1, x_2, x_3, x_4|Z) = S_1 S_2 S_3 S_4$$

where $S(x_1, x_2, x_3, x_4|Z)$ is the joint survival function and $S_1, \ldots, S_4$ are the marginal survival functions for the left front, right front, left rear and right rear udder quarters respectively.

For this simplest model and all models following, both a parametric and semiparametric approach will be considered. In the parametric approach, the baseline hazard is assumed to be Weibull. The general likelihood function (5.1) can then be maximized in one stage, i.e., maximizing the likelihood jointly for the parameter(s) of the marginal survival functions and the parameter in the copula function, or in two stages, first estimating the parameter(s) of the marginal survival functions, plugging those in (5.1) and then maximizing only for the parameter in the copula function. In the semiparametric approach, the baseline hazard is unspecified and only a two stage approach is feasible: partial likelihood maximization is used to estimate the marginal survival functions, plugging those in (5.1) and then maximizing only for the parameter in the copula function. The two-stage approach is straightforward as the first stage, estimating the marginal survival functions, is based on basic survival models without clustering, and only one parameter remains in (5.1) for which it needs to be maximized.

5.3.2 One level of clustering (Model 1)

If we assume that the association between each two different udder quarters is the same, we model the joint survival function by a four-dimensional Archimedean copula function with generator φ. This situation is depicted in Figure 5.2. The joint survival function is

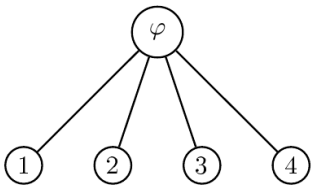


Figure 5.2: Four-dimensional Archimedean copula with generator φ.

represented as

$$S(x_1, x_2, x_3, x_4|Z) = C_{\theta_0}(S_1, S_2, S_3, S_4),$$

or equivalently,

$$S(x_1, x_2, x_3, x_4|Z) = \varphi\left[\varphi^{-1}(S_1) + \varphi^{-1}(S_2) + \varphi^{-1}(S_3) + \varphi^{-1}(S_4)\right]$$

where $\varphi : [0, \infty[\rightarrow [0, 1]$ is a continuous, strictly decreasing function which is completely monotonic and has $\varphi(0) = 1$ and $\varphi(\infty) = 0$ (Nelsen, 2006). The generator φ depends on the association parameter θ_0. As an example, the association structure is induced here through a Clayton copula with generator $\varphi(t) = (1 + \theta_0 t)^{-1/\theta_0}$ with $\theta_0 > 0$. Infection times are independent when θ_0 approaches zero.

For the case of a parametric baseline hazard and using the one-stage procedure, the contributions to the likelihood expression (5.1) are derived in Appendix B. In Chapter 2 we showed that maximizing the likelihood expression is equivalent to solving $\dfrac{d \log L}{d\boldsymbol{\eta}} = \mathbf{0}$ with

$$\log L(\boldsymbol{\eta}) = \sum_{i=1}^{K} \left(\sum_{j=1}^{n_i} \delta_{ij} \log \left[\frac{-f_{ij}}{\varphi'(\varphi^{-1}(S_{ij}))} \right] + \sum_{l=0}^{d_i - 1} \log(1 + l\theta_0) \right.$$
$$\left. -(d_i + 1/\theta_0) \log \left(1 + \theta_0 \sum_{j=1}^{n_i} \varphi^{-1}(S_{ij}) \right) \right), \qquad (5.2)$$

where $n_i = 4$, $d_i = \sum_{j=1}^{n_i} \delta_{ij}$, $S_{ij} = S_j(x_{ij}|Z_i)$ and $f_{ij} = f(x_{ij}|Z_i) = -\frac{dS_{ij}}{dx_{ij}}$, $i = 1, \ldots, K$, $j = 1, \ldots, 4$. The parameter vector $\boldsymbol{\eta}$ contains the baseline parameters from the four margins, the parity effect β and the association parameter θ_0.

In the model above, it was not only assumed that all cows can be described by the same correlation structure, but also that the correlations themselves are the same. As primiparous and multiparous cows react quite differently with respect to udder quarter infections, it is worthwhile to test whether primiparous and multiparous cows share the same values for the correlations within the same correlation structure.

To test whether the association between infection times depends on the parity of the cow, we use the following copula function $\varphi(t) = (1 + \theta_p t)^{-1/\theta_p}$ for primiparous cows and $\varphi(t) = (1 + \theta_m t)^{-1/\theta_m}$ for multiparous cows and test the hypothesis

$$H_0 : \theta_m = \theta_p \quad \text{versus} \quad H_1 : \theta_m \neq \theta_p$$

which can then be tested through the likelihood ratio test.

5.3.3 Multilevel clustering: parent copula with two identical child copulas (Model 2)

We assume that the front udder quarters have the same association as the rear udder quarters, which is indicated by the solid lines in the left panel of Figure 5.3. Another type of association occurs between the front and rear udder quarters, indicated by the dashed lines. This type of association structure is captured by a partially nested Archimedean copula function where the parent copula C_{θ_0} has two identical child copulas C_{θ_1} and C_{θ_1}:

$$C_{\theta_0}\left[C_{\theta_1}(S_1, S_2), C_{\theta_1}(S_3, S_4)\right],$$

or equivalently,

$$\begin{aligned} S(x_1, x_2, x_3, x_4|Z) = \varphi_0 \big[&\varphi_0^{-1} \circ \varphi_1 \left\{\varphi_1^{-1}(S_1) + \varphi_1^{-1}(S_2)\right\} \\ &+\varphi_0^{-1} \circ \varphi_1 \left\{\varphi_1^{-1}(S_3) + \varphi_1^{-1}(S_4)\right\}\big]. \end{aligned}$$

The generator φ_0 describes the association between front and rear udder quarters, while φ_1 describes the association within front udder quarters and within rear udder quarters. The right panel of Figure 5.3 is a schematic representation of the partially nested Archimedean

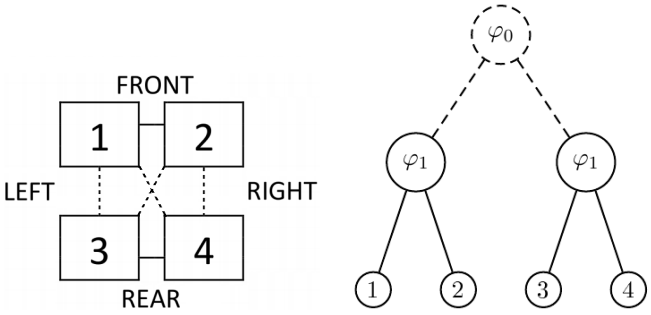


Figure 5.3: Parent copula with two identical child copulas.

copula. According to McNeil (2008), for a general nested Archimedean structure to be a proper copula, it is sufficient that all appearing nodes of the form $\varphi_k^{-1} \circ \varphi_l$ have completely monotone derivatives. The sufficient nesting condition is often easily verified if all generators appearing in the nested structure come from the same parametric family. For the Archimedean families of Ali-Mikhail-Haq, Clayton, Frank, Gumbel and Joe, two generators φ_k and φ_l of the same family with corresponding parameters θ_k and θ_l fulfill the sufficient nesting condition if $\theta_k \leq \theta_l$ (Hofert, 2011).

Since φ_0 is a completely monotonic copula generator, we can look at φ_0 as a Laplace transform of a positive distribution function $G_0(x)$,

$$\varphi_0(t) = \int_0^\infty e^{-tx} dG_0(x).$$

This means we can write the joint survival function as

$$\begin{aligned} S(x_1, x_2, x_3, x_4|Z) &= \int_0^\infty e^{-x\left[\varphi_0^{-1}\circ\varphi_1\left(\varphi_1^{-1}(S_1)+\varphi_1^{-1}(S_2)\right)+\varphi_0^{-1}\circ\varphi_1\left(\varphi_1^{-1}(S_3)+\varphi_1^{-1}(S_4)\right)\right]} dG_0(x) \\ &= \int_0^\infty e^{-x\varphi_0^{-1}\circ\varphi_1\left(\varphi_1^{-1}(S_1)+\varphi_1^{-1}(S_2)\right)} \times e^{-x\varphi_0^{-1}\circ\varphi_1\left(\varphi_1^{-1}(S_3)+\varphi_1^{-1}(S_4)\right)} dG_0(x). \end{aligned}$$

Defining in this expression $\phi_{1x}(\cdot) = e^{-x\varphi_0^{-1}\circ\varphi_1(\cdot)}$, we can rewrite

$$\begin{aligned} e^{-x\varphi_0^{-1}\circ\varphi_1\left(\varphi_1^{-1}(S_1)+\varphi_1^{-1}(S_2)\right)} &= \phi_{1x}\left(\varphi_1^{-1}(S_1) + \varphi_1^{-1}(S_2)\right) \\ &= \phi_{1x}\left(\phi_{1x}^{-1}\left(e^{-x\varphi_0^{-1}(S_1)}\right) + \phi_{1x}^{-1}\left(e^{-x\varphi_0^{-1}(S_2)}\right)\right). \end{aligned}$$

Since $\varphi_0^{-1} \circ \varphi_1$ has a completely monotonic derivative, we note that ϕ_{1x} is a Laplace transform, and therefore is a generator of an Archimedean copula. Hence there exists a

distribution G_{1x} such that

$$\phi_{1x}(t) = e^{-x\varphi_0^{-1}\circ\varphi_1(t)} = \int_0^\infty e^{-x_1 t} dG_{1x}(x_1).$$

This leads to

$$\begin{aligned} S(x_1,x_2,x_3,x_4) &= \int_0^\infty \phi_{1x}\left(\varphi_1^{-1}(S_1)+\varphi_1^{-1}(S_2)\right) \times \phi_{1x}\left(\varphi_1^{-1}(S_3)+\varphi_1^{-1}(S_4)\right) dG_0(x) \\ &= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_1)+\varphi_1^{-1}(S_2)\right)} dG_{1x}(x_1) \\ &\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_3)+\varphi_1^{-1}(S_4)\right)} dG_{1x}(x_1) dG_0(x). \end{aligned} \tag{5.3}$$

In Appendix B we show that the likelihood for K clusters can be written as

$$\begin{aligned} L = \prod_{i=1}^K \int_0^\infty \int_0^\infty \prod_{j=1}^2 e^{-x_1\left(\varphi_1^{-1}(S_{ij})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right)^{\delta_{ij}} dG_{1x}(x_1) \\ \times \int_0^\infty \prod_{j=3}^4 e^{-x_1\left(\varphi_1^{-1}(S_{ij})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right)^{\delta_{ij}} dG_{1x}(x_1) dG_0(x), \end{aligned} \tag{5.4}$$

with $S_{ij} = S_j(x_{ij}|Z_i)$, $f_{ij} = \frac{dS_{ij}}{dx_{ij}}$, $i = 1,\ldots,K$, $j = 1,\ldots,4$. We notice that the likelihood expression (5.4) is a modified version of (2.2).

Choosing parent copula C_{θ_0} and child copulas C_{θ_1} to be Clayton copulas with generators $\varphi_0(t) = (1+\theta_0 t)^{-1/\theta_0}$ and $\varphi_1(t) = (1+\theta_1 t)^{-1/\theta_1}$, the distribution G_{1x} is a special case of the exponentially tilted stable distribution (Hofert, 2011). Unfortunately, for this choice, the inner integral in the likelihood has no closed form. We will therefore not use the nested integral structure (5.4) for computations, but instead we will calculate the joint survival function and its mixed partial derivatives up to fourth order. The joint survival function is

$$S(x_1,x_2,x_3,x_4) = \left[-1+\left(-1+S_1^{-\theta_1}+S_2^{-\theta_1}\right)^{\theta_0/\theta_1}+\left(-1+S_3^{-\theta_1}+S_4^{-\theta_1}\right)^{\theta_0/\theta_1}\right]^{-1/\theta_0}.$$

The nesting condition for this setting is $\theta_0 \leq \theta_1$. This means that there must be a stronger association of infection times within front and within rear udder parts, than there is between front and rear udder parts.

The contributions to the likelihood expression (5.1) for the one-stage parametric approach are given in Appendix B.

5.3.4 Multilevel clustering: parent copula with two different child copulas (Model 3)

We will now assume that the association within the front udder quarters is different from the association within the rear udder quarters. A third type of association occurs between front and rear udder quarters.

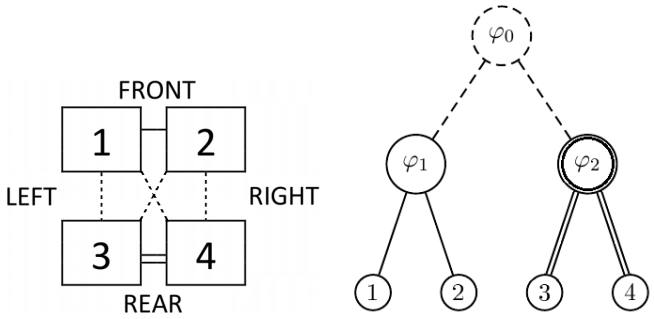


Figure 5.4: Parent copula with two different child copulas.

The hierarchical Archimedean copula function that represents this situation is

$$C_{\theta_0}\left[C_{\theta_1}(S_1, S_2), C_{\theta_2}(S_3, S_4)\right],$$

or equivalently,

$$\begin{aligned} S(x_1, x_2, x_3, x_4) = \varphi_0 \big[&\varphi_0^{-1} \circ \varphi_1 \left\{\varphi_1^{-1}(S_1) + \varphi_1^{-1}(S_2)\right\} \\ &+\varphi_0^{-1} \circ \varphi_2 \left\{\varphi_2^{-1}(S_3) + \varphi_2^{-1}(S_4)\right\}\big]. \end{aligned}$$

The generator φ_0 describes the association between the front and rear udder quarters, while generators φ_1 and φ_2 describe the association within the front udder quarters and within the rear udder quarters, respectively.

We will choose φ_0, φ_1 and φ_2 to be generators of Clayton copulas with association parameters θ_0, θ_1 and θ_2. In that case, the joint survival function is given by

$$S(x_1, x_2, x_3, x_4) = \left[-1 + \left(-1 + S_1^{-\theta_1} + S_2^{-\theta_1}\right)^{\theta_0/\theta_1} + \left(-1 + S_3^{-\theta_2} + S_4^{-\theta_2}\right)^{\theta_0/\theta_2}\right]^{-1/\theta_0}.$$

The nesting conditions are $\theta_0 \leq \theta_1$ and $\theta_0 \leq \theta_2$. The contributions to the likelihood function are given in Appendix B.

5.4 Results

5.4.1 The marginal survival functions

When assuming a parametric form of the marginal survival functions, the Weibull distribution is a popular choice. Under the Weibull assumption, the marginal survival functions are

$$S_{ij}(t|Z_i) = \exp(-\lambda_j t^{\rho_j} \exp(\beta Z_i)), \quad j = 1, \ldots, 4.$$

The parity Z_i is cow-specific and therefore we assume that the parity effect is the same in each of the four quarters ($\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$).

In a model without clustering (Model 0), standard survival methods yield the parameter estimates in Table 5.1.

Weibull baseline parameters				parity
Left Front	Right Front	Left Rear	Right Rear	effect β
$\lambda_1 = 0.168(0.011)$	$\lambda_2 = 0.178(0.012)$	$\lambda_3 = 0.145(0.011)$	$\lambda_4 = 0.157(0.011)$	$-0.418(0.051)$
$\rho_1 = 1.321(0.052)$	$\rho_2 = 1.270(0.050)$	$\rho_3 = 1.325(0.055)$	$\rho_4 = 1.280(0.053)$	

Table 5.1: Parameter estimates for marginal Weibull distributions and parity effect.

When no parametric baseline is assumed, a stratified Cox proportional hazards model (Cox, 1972) can be used, where the baseline hazard $h_{0j}(\cdot)$ is left unspecified:

$$h_{ij}(t|Z_i) = h_{0j}(t) \exp(\beta Z_i).$$

In this (semi-parametric) model, the estimated parity effect is $-0.407(0.051)$.

Following either the parametric or semi-parametric approach, these parameter estimates are consistent and are used in the second stage of the (semi-)parametric two-stage estimation procedure of models 1, 2 and 3.

5.4.2 Fitting a hierarchy of association structures

Models 0, 1, 2 and 3 are fitted and the parameter estimates are reported in Table 5.2. Corresponding standard errors are in brackets. In the one-stage procedure, these were

retrieved from the inverse Hessian matrix. In the two-stage procedure, standard errors of the association parameters were determined using the grouped jackknife method (Lipsitz *et al.*, 1994; Lipsitz and Parzen, 1996).

Implementation of the models was done in R, based on the likelihood contributions that were calculated in Appendix B. As an example, the R code that was used to fit Model 3 using the one-stage parametric procedure is given in Appendix B.4.1.

As pointed out in the previous section, in two-stage estimation, the estimates of the baseline and the parity effect are equal to the estimates arising from the independence model.

	1-stage parametric estimation						2-stage estimation	
							parametric	semiparametric
	Weibull baseline parameters				parity	copula	copula	copula
	$i=1$	$i=2$	$i=3$	$i=4$	effect β	parameter(s)	parameter(s)	parameter(s)
Model 0	λ_i 0.168(0.011)	0.178(0.012)	0.145(0.011)	0.157(0.011)	-0.418(0.051)	-	-	-
	ρ_i 1.321(0.052)	1.270(0.050)	1.325(0.055)	1.280(0.053)				
Model 1	λ_i 0.156(0.011)	0.166(0.011)	0.134(0.010)	0.141(0.010)	$-0.344(0.068)$	$\theta_0 = 3.184(0.182)$	$\theta_0 = 2.938(0.201)$	$\theta_0 = 3.227(0.216)$
	ρ_i $1.297(0.048)$	$1.262(0.047)$	$1.310(0.052)$	$1.266(0.050)$				
Model 2	λ_i 0.157(0.011)	0.166(0.011)	0.134(0.010)	0.142(0.010)	-0.340(0.068)	$\theta_0 = 3.050(0.184)$	$\theta_0 = 2.825(0.186)$	$\theta_0 = 3.110(0.215)$
	ρ_i 1.299(0.048)	1.264(0.047)	1.310(0.052)	1.269(0.050)		$\theta_1 = 3.552(0.221)$	$\theta_1 = 3.299(0.232)$	$\theta_1 = 3.626(0.263)$
Model 3	λ_i 0.157(0.011)	0.166(0.011)	0.134(0.010)	0.142(0.010)	-0.340(0.068)	$\theta_0 = 3.048(0.185)$	$\theta_0 = 2.821(0.187)$	$\theta_0 = 3.107(0.215)$
	ρ_i 1.299(0.048)	1.264(0.047)	1.311(0.052)	1.270(0.050)		$\theta_1 = 3.589(0.271)$	$\theta_1 = 3.363(0.282)$	$\theta_1 = 3.674(0.323)$
						$\theta_2 = 3.513(0.274)$	$\theta_2 = 3.231(0.278)$	$\theta_2 = 3.575(0.309)$

Table 5.2: Parameter estimates from copula models 0, 1, 2 and 3.

To investigate which association structure is most appropriate, we test the hypotheses

$$\begin{array}{llll} H_0^A : \theta_0 = 0 & \text{versus} & H_1^A : \theta_0 \neq 0 & \text{in Model 1} \\ H_0^B : \theta_1 = \theta_0 & \text{versus} & H_1^B : \theta_1 \neq \theta_0 & \text{in Model 2} \\ H_0^C : \theta_2 = \theta_1 & \text{versus} & H_1^C : \theta_2 \neq \theta_1 & \text{in Model 3} \end{array}$$

In words, test A is used to detect the presence of clustering in the data. With test B we determine whether it is necessary to account for front and rear subclusters. Test C is used to detect a different level of association in the front and rear subclusters. For testing $H_0^A : \theta = 0$, we use a likelihood ratio test with a mixed chi-squared distribution, since the null hypothesis lies at the boundary of the parameter space (Duchateau *et al.*, 2002). In Section 5.5.1, we take a closer look at this distribution. To test hypotheses H_0^B and H_0^C, the likelihood ratio statistic follows a $\chi^2(1)$ distribution. The likelihood ratio tests are performed for both one-stage and two-stage estimation procedures. The resulting p-values are given Table 5.3. From Table 5.3 we conclude that there is in fact clustering of infection times, and the front and rear subclusters are detected, but there is no need to set up a model which includes a different level of association within each subcluster. The most appropriate model for the udder quarter infection times, is therefore Model 2.

Test	Estimation procedure	LRT	p-value
H_0^A vs H_1^A	one-stage parametric	1938.344	< 0.0001
	two-stage parametric	1931.996	< 0.0001
	two-stage semi-parametric	1949.055	< 0.0001
H_0^B vs H_1^B	one-stage parametric	14.156	0.00017
	two-stage parametric	14.703	0.00013
	two-stage semi-parametric	15.397	< 0.0001
H_0^C vs H_1^C	one-stage parametric	0.056	0.812
	two-stage parametric	0.228	0.633
	two-stage semi-parametric	0.111	0.739

Table 5.3: Likelihood ratio tests for the association structure.

5.4.3 The association structure as a function of the parity covariate

As mentioned at the end of Section 5.3.2, it is worthwhile to test the hypothesis

$$H_0^Z : \theta_m = \theta_p \quad \text{versus} \quad H_1^Z : \theta_m \neq \theta_p$$

in Model 1. The p-values of the likelihood ratio tests for the different estimation approaches are listed in Table 5.4. None of the estimation procedures lead to a significant

Test	Estimation procedure	LRT	p-value	$\hat{\theta}_m$	$\hat{\theta}_p$
H_0^Z vs H_1^Z	one-stage parametric	1.436	0.231	3.055	3.515
	two-stage parametric	2.732	0.098	2.829	3.302
	two-stage semi-parametric	0.780	0.377	3.160	3.430

Table 5.4: Likelihood ratio test for $H_0^Z : \theta_m = \theta_p$ versus $H_1^Z : \theta_m \neq \theta_p$ in Model 1.

difference between θ_m and θ_p. Consequently, there is no need to model the association structure as a function of the parity covariate.

5.5 Size and power analysis

We simulate survival data that resemble the udder infection data. All subjects are sampled from the same marginal distribution, i.e. a Weibull distribution with parameters comparable to the estimated parameters of the udder infection data set: $\lambda = 0.11, \rho = 1.3, \beta = 0.4$. The censoring variable is Weibull distributed with $\rho_C = 1.3$ and $\lambda_C = 0.21$, yielding a censoring percentage around 61%. The aim is to assess the size and the power of the likelihood ratio tests when comparing the different association structures. We only investigate the performance of the two-stage parametric estimation procedure. In the first simulation setting, we simulate four-dimensional survival data sets with one level of clustering, and calculate the size and the power to detect departures from the independence model. In the second simulation setting, we simulate data from a two-level hierarchical copula model with two identical child copulas, and compute the size and power to detect the subclusters. In the third simulation setting, data were simulated from a two-level hierarchical Archimedean copula model with two different child copulas, and the size and the power to detect the difference between the two subclusters were determined. In the

last setting, we study the size and the power to detect a covariate effect on the association parameter in the model with one level of clustering.

5.5.1 Testing for independence versus one-level clustering

Let the true value of θ range from 0 to 0.5 by steps of size 0.05. We simulate 1000 data sets with 200 clusters of size 4 from a Clayton copula for each specific value of the association parameter θ. Our aim is to pick up deviations from independence. The power of the likelihood ratio test for independence is plotted versus the value of θ. At the boundary of the parameter space $(\theta = 0)$, the likelihood ratio statistic follows a mixed chi-squared distribution

$$2 \log \frac{\text{likelihood alternative model}}{\text{likelihood null model}} \sim 0.5\chi^2(0) + 0.5\chi^2(1).$$

Figure 5.5 shows that a value of $\theta = 0.25$, corresponding to a Kendall's tau of 0.11, is detected with a probability over 80%. In the model with one level of clustering, deviations from independence are hence quickly detected. From $\theta = 0.35$ onwards, the power level approaches 1. At $\theta = 0$, we approximately attain the size of the test by a value of 0.043.

5.5.2 Testing for one-level clustering versus two-level clustering with one parent and two identical child copulas

We let the true values of θ_0 and θ_1 range from 0.02 to 1 by steps of length 0.02, only considering those combinations of (θ_0, θ_1) for which the nesting condition $\theta_1 \geq \theta_0$ is met. We simulate 1000 data sets with 200 clusters of size 4 from a hierarchical Clayton copula with parent copula C_{θ_0} and 2 identical child copulas C_{θ_1} for each eligible pair (θ_0, θ_1) and calculate the probability to detect the subclusters. In order to make use of the $\chi^2(1)$ distribution, however, values of θ_0 and θ_1 close to the boundary, i.e., $\theta_0 = \theta_1 = 0$, should be excluded, as demonstrated in the discussion section. Therefore, first, the independence hypothesis is tested, and if not rejected, the simulated data set is discarded and not used in the future testing of $H_0^B : \theta_1 = \theta_0$. Therefore, in simulation settings with θ_1 and θ_0 close to zero, a substantial number of the 1000 simulations might be discarded. If less than 20% of the 1000 data sets remain, the symbol ▲ is used for plotting in the top panel

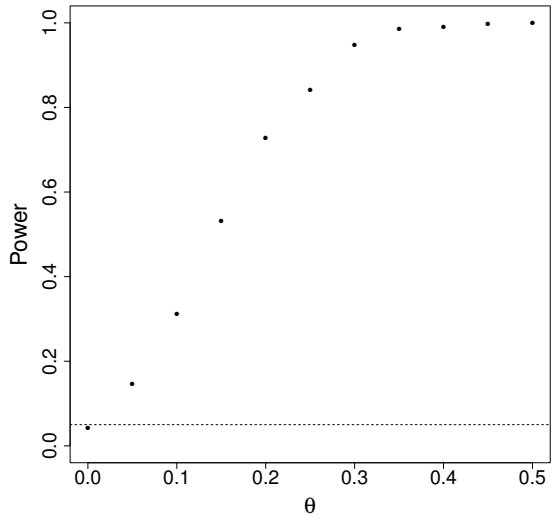


Figure 5.5: The power of the likelihood ratio test for $H_0^A : \theta = 0$ versus $H_1^A : \theta \neq 0$.

of Figure 5.6. As demonstrated in Section 5.5.1 and Figure 5.5 in particular, this occurs only for very small values of the association parameter ($\theta_0 = \theta_1 \leq 0.06$). The size of the test is shown in the upper panel of Figure 5.6; most values are below 0.065 and thus quite acceptable. In the lower panel of Figure 5.6, the line $\theta_0 = \theta_1$ indicates the null model, i.e., no subclusters. To obtain a power of 80%, values must differ quite substantially, e.g., $(0.2, 0.75)$ or $(0.3, 0.9)$.

5.5.3 Testing for two-level clustering with one parent and two identical child copulas versus two-level clustering with one parent and two different child copulas

We fix the value of θ_0 at 0.5 and let θ_1 and θ_2 range from 0.5 to 2.0 by steps of length 0.1. We simulate 1000 data sets with 200 clusters of size 4 from a hierarchical Clayton copula with parent copula C_{θ_0} and 2 child copulas C_{θ_1} and C_{θ_2}. For each combination of θ_1 and θ_2, we calculate the probability to detect the different levels of association in the subclusters. In Figure 5.7, the line $\theta_1 = \theta_2$ indicates the model with two identical child copulas. On this line, we achieve the size of the test.

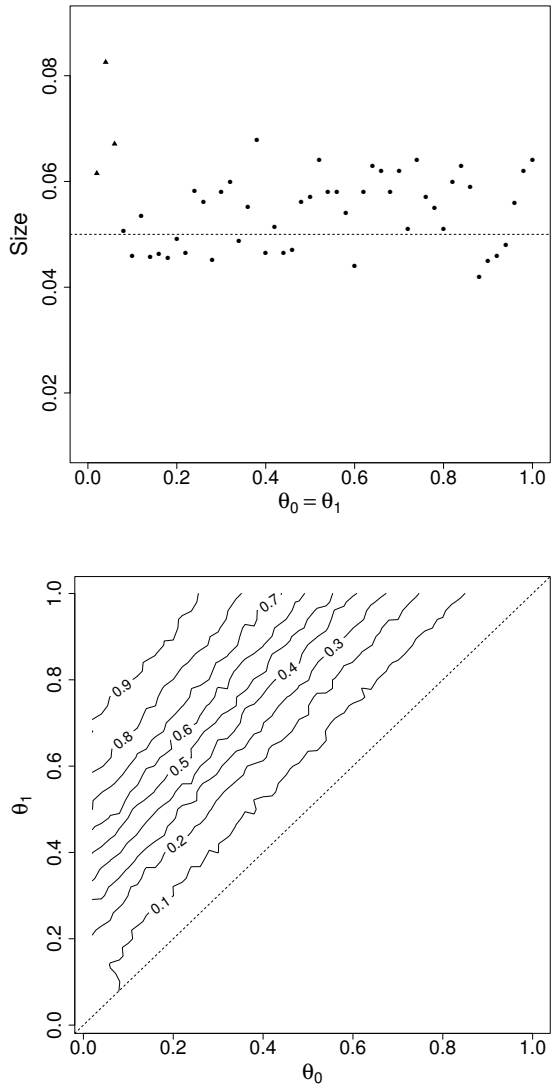


Figure 5.6: Size (top) and power (bottom) of the likelihood ratio test for $H_0^B : \theta_0 = \theta_1$ versus $H_1^B : \theta_0 \neq \theta_1$.

5.5.4 Testing for differing association structures as a function of a covariate

We fix the true value of θ_p at 1.5 and let θ_m range from 0 to 3.5 by steps of length 0.1. We simulate 1000 data sets with 200 clusters of size 4 from a Clayton copula C_{θ_m} for a

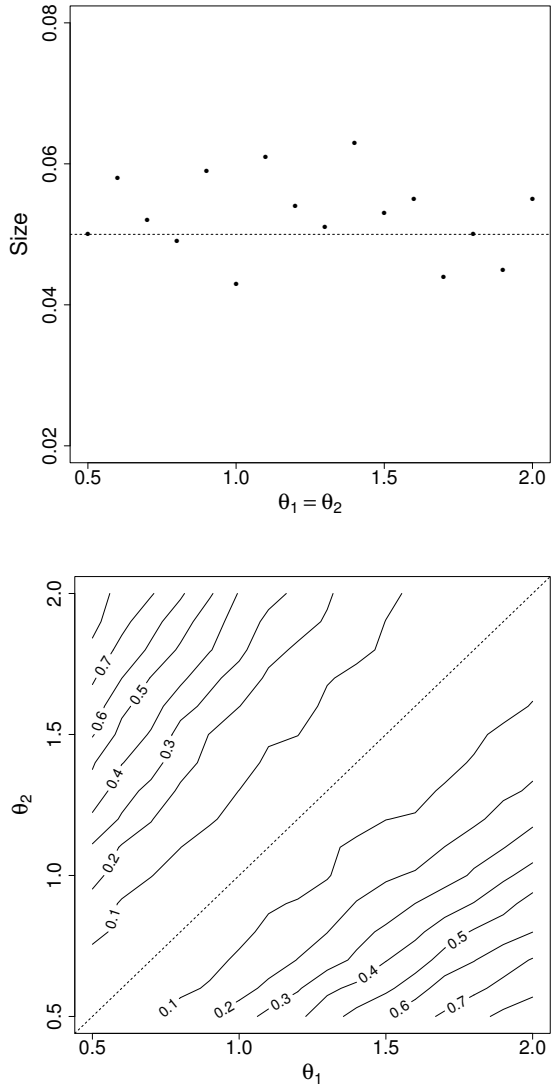


Figure 5.7: Size (top) and power (bottom) of the likelihood ratio test for $H_0^C : \theta_1 = \theta_2$ versus $H_1^C : \theta_1 \neq \theta_2$.

multiparous cow and from a Clayton copula C_{θ_p} for a primiparous cow.

We determine how many times the difference between the association parameters θ_m and θ_p is picked up. In Figure 5.8, when $\Delta\theta = \theta_m - \theta_p$ approaches -1.5, i.e., when C_{θ_m}

approaches the independence copula, the power increases quickly. On the right hand side of Figure 5.8, where $\Delta\theta$ is positive, the power to detect a covariate effect on the association parameter increases more gradually.

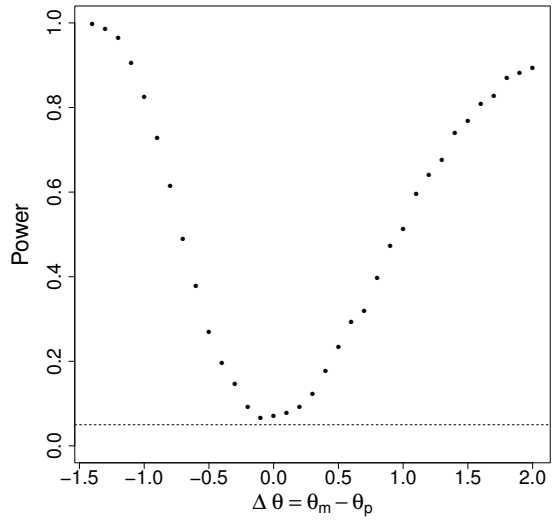


Figure 5.8: The power of the likelihood ratio test for $H_0^Z : \theta_m = \theta_p$ versus $H_1^Z : \theta_m \neq \theta_p$.

5.6 Discussion

We compared different hierarchical Archimedean copula models for the association between infection times of the four udder parts in dairy cows. The most adequate model for the quadrivariate udder infection data is the nested copula model where the association between front and rear udder quarters is smaller than the association between two front, resp. rear, udder quarters. The within-front association is not significantly different from the within-rear association. According to the best fitting copula, i.e., Model 2, the association parameter between two quarters either on the rear or on the front side corresponds to 3.552, or a Kendall's tau equal to 0.64. As expected the association parameter between two quarters not on the same rear or front side is smaller and equal to 3.050 with a corresponding Kendall's tau equal to 0.60. Although these two association parameters differ significantly from each other, it is not important from a practical point of view as both are large. It is important to know for a dairy holder that noninfected udder quarters

are highly at risk whenever one of the udder quarters of a cow is infected.

The reason to choose all apparent copulas to be Clayton copulas is threefold. From a computational point of view, Clayton copulas are convenient to work with, since there exists a closed form expression for the derivatives of the copula generator φ. For a hierarchical Clayton copula to be well-defined, a simple nesting condition has to be met, i.e., $\theta_k \leq \theta_l$ for all appearing nodes of the form $\varphi_k^{-1} \circ \varphi_l$. Additionally, the Clayton copula has lower tail dependence. In a survival context, this translates to a stronger association later in time. It's therefore important to extend these copulas to other members of the Archimedean copula family to investigate when the correlation in time is strongest.

In Section 5.5.1, the power of testing for the presence of simple clustering, i.e., the same pairwise correlation between all udder quarter, in four-dimensional data was assessed using a likelihood ratio test with a mixed chi-squared distribution, since the null hypothesis of no association lies on the boundary of the parameter space. The top panel of Figure 5.9 illustrates that the empirical cumulative distribution function of the likelihood ratio statistic, calculated for 1000 simulated datasets without clustering, agrees with the $0.5\chi^2(0) + 0.5\chi^2(1)$ distribution function. In Section 5.5.2, we assessed the power of testing for the presence of subclusters in four-dimensional data using a likelihood ratio test with a $\chi^2(1)$ distribution. The null hypothesis $H_0^B : \theta_0 = \theta_1$ lies on the boundary of the nesting condition $\theta_0 \leq \theta_1$, however, since the nesting condition only is sufficient and not necessary, the mixed chi-squared distribution does not apply unless $\theta_0 = \theta_1 = 0$. Before testing for multiple levels of clustering, it is therefore necessary to test first for the presence of simple clustering. Omitting this preliminary test can lead to test sizes much larger than the nominal significance level. This is illustrated in Figure 5.10, where the size of the likelihood ratio test for $H_0^B : \theta_0 = \theta_1$ versus $H_1^B : \theta_0 \neq \theta_1$ was determined for 1000 data sets with 200 clusters of size 4 that were simulated from a Clayton copula with association parameter ranging from 0.02 to 1. For small values of θ, the size clearly deviates from the desired 0.05 level. In the upper panel of Figure 5.6 in Section 5.5.2, we remedy this problem by only looking at those data sets in which the preliminary test detected the presence of simple clustering. In the lower panel of Figure 5.9, the likelihood ratio test for $H_0^B : \theta_0 = \theta_1$ versus $H_1^B : \theta_0 \neq \theta_1$ was performed on 1000 data sets that were simulated from a unilevel Clayton copula model with association parameter equal to 1.2. As can be seen from this figure, the empirical cumulative distribution function of the likelihood ratio statistic coincides with the $\chi^2(1)$ distribution.

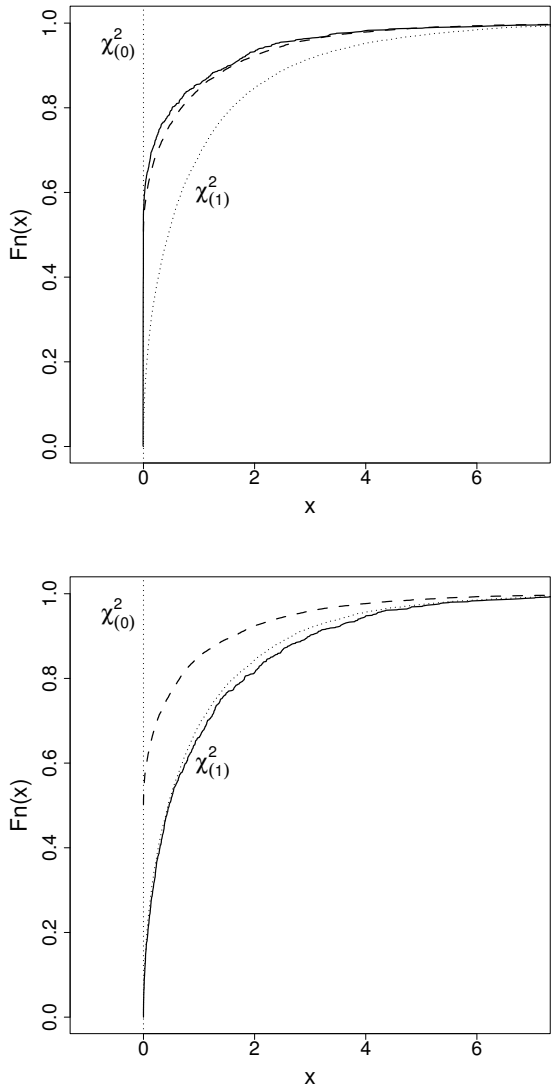


Figure 5.9: Top: the empirical cumulative distribution function (ECDF) under $H_0^A : \theta = 0$ (solid line), ECDF of $\chi^2(0), \chi^2(1)$ (dotted lines) and $0.5\chi^2(0)+0.5\chi^2(1)$ (dashed line). Bottom: ECDF under $H_0^B : \theta_0 = \theta_1 = 1.2$ (solid line), ECDF of $\chi^2(0), \chi^2(1)$ (dotted lines) and $0.5\chi^2(0)+0.5\chi^2(1)$ (dashed line).

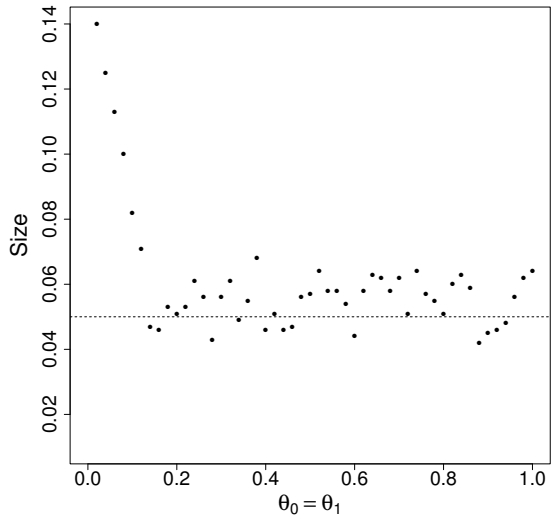


Figure 5.10: The size of the likelihood ratio test for $H_0^B : \theta_0 = \theta_1$ versus $H_1^B : \theta_0 \neq \theta_1$, including also the data sets for which no simple clustering was detected.

Bibliography

Ampe, B., Goethals, K., Laevens, H. and Duchateau, L. (2012) Investigating clustering in interval-censored udder quarter infection times in dairy cows using a gamma frailty model. *Preventive Veterinary Medicine*, **106**, 251–257.

Andersen, E. W. (2005) Two-stage estimation in copula models used in family studies. *Lifetime Data Analysis*, **11**, 333–350.

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

Cox, D. R. and Hinkley, D. (1974) *Theoretical Statistics*. Chapman and Hall.

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

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R. and Sylvester, R. (2002) The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics and Data Analysis*, **40**, 603–620.

Efron, B. (1977) The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72(359)**, 557–565.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*. Wiley.

Geerdens, C., Claeskens, G. and Janssen, P. (2013) Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics*, **14**, 433–446.

Glidden, D. V. (2000) A two-stage estimator of the dependence parameter for the clayton-oakes model. *Lifetime Data Analysis*, **6**, 141–156.

Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and Duchateau, L. (2009) Modeling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural Biological and Environmental Statistics*, **14**, 1–14.

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

Hofert, M. (2011) Efficiently sampling nested archimedean copulas. *Computational Statistics and Data Analysis*, **55**, 57–70.

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

Hougaard, P. (2000) *Analysis of Multivariate Survival Data*. Springer.

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

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

Kaplan, E. and Meier, P. (1958) Nonparametric estimation of incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Klein, J. and Moeschberger, M. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*. Springer.

Kleinbaum, D. and Klein, J. (2012) *Survival Analysis: A Self-Learning Text*. Springer.

Lehmann, E. L. and Casella, G. (1998) *Theory of Point Estimation*. Springer.

Lipsitz, S. R., Dear, K. B. and Zhao, L. (1994) Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics*, **50**, 842–846.

Lipsitz, S. R. and Parzen, M. (1996) A jackknife estimator of variance for cox regression for correlated survival data. *Biometrics*, **52**, 291–298.

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

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

McNeil, A. J. (2008) Sampling nested archimedean copulas. *Journal of Statistical Computation and Simulation*, **6**, 567–581.

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

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

Pawitan, Y. (2001) *In All Likelihood. Statistical Modelling and Inference Using Likelihood*. Oxford Science Publications.

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

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

Spiekerman, C. F. and Lin, D. Y. (1998) Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, **93**, 1164–1175.

Weller, J., Saran, A. and Zeliger, Y. (1992) Genetic and environmental relationships among somatic cell count, bacterial infection, and clinical mastitis. *Journal of Dairy Science*, **75**, 2535–2540.

Wienke, A. (2011) *Frailty Models in Survival Analysis*. Chapman & Hall.

Xu, J. (1996) Statistical modelling and inference for multivariate and longitudinal discrete response data. *Ph.D. Thesis*.

Appendix A

Validation of the Archimedean copula likelihood expression

A.1 Validation of the bivariate copula likelihood for a general Archimedean copula

Lemma A.1.1. *For a sample of bivariate survival data* $\{(x_{i1}, \delta_{i1}), (x_{i2}, \delta_{i2})\}, i = 1, \dots K$ *the contribution of pair* i *to the Archimedean copula likelihood can be expressed as*

$$\int_0^\infty \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S_{ij})} \left[\frac{-x f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})))} \right]^{\delta_{ij}} dG_\theta(x).$$

Proof. We calculate the bivariate copula likelihood for a general Archimedean copula, starting from formula (1.1) on page 9. We invoke some useful properties of the Archimedean copula function.

The generator $\varphi_\theta : [0, \infty[\to [0, 1]$ of an Archimedean copula function $C : [0, 1]^n \to [0, 1]$, can be written as a Laplace transform

$$\varphi_\theta(s) = \int_0^\infty e^{-sx} dG_\theta(x), \quad s \geq 0.$$

The first and second order derivatives of the generator are $\varphi'_\theta(s) = \int_0^\infty -xe^{-sx}dG_\theta(x)$ and $\varphi''_\theta(s) = \int_0^\infty x^2e^{-sx}dG_\theta(x)$. To ease notation, we write $S_{ij} = S(x_{ij}|\boldsymbol{Z}_{ij})$ and $f_{ij} = f(x_{ij}|\boldsymbol{Z}_{ij})$. Since f is the density function and S the survival function, they are related via $f_{ij} = -\frac{dS_{ij}}{dx_{ij}}$. In a sample of K pairs of time-to-event data $\{(x_{11},\delta_{11}),(x_{12},\delta_{12})\},\ldots,\{(x_{K1},\delta_{K1}),(x_{K2},\delta_{K2})\}$, we distinguish four possible situations for pair i:

- Two censored observations ($\delta_{i1}=0,\delta_{i2}=0$). The contribution to the likelihood is the joint survival function

$$\begin{aligned} L_{i,(0,0)} &= S(x_{i1},x_{i2}|\boldsymbol{Z}_{i1},\boldsymbol{Z}_{i1}) \\ &= C(S_{i1},S_{i2}) = \varphi_\theta(\varphi_\theta^{-1}(S_{i1})+\varphi_\theta^{-1}(S_{i2})) \\ &= \int_0^\infty e^{-x(\varphi_\theta^{-1}(S_{i1})+\varphi_\theta^{-1}(S_{i2}))}dG_\theta(x). \end{aligned}$$

- An event for the first subject and a censored observation for the second subject ($\delta_{i1}=1,\delta_{i2}=0$):

$$\begin{aligned} L_{i,(1,0)} &= -\frac{\partial S(x_{i1},x_{i2}|\boldsymbol{Z}_{i1},\boldsymbol{Z}_{i2})}{\partial x_{i1}} \\ &= -\varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1})+\varphi_\theta^{-1}(S_{i2})\right)\frac{d}{dx_{i1}}\left[\varphi_\theta^{-1}(S_{i1})\right] \\ &= \varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1})+\varphi_\theta^{-1}(S_{i2})\right)\frac{f_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))} \\ &= \int_0^\infty -xe^{-x\left(\varphi_\theta^{-1}(S_{i1})+\varphi_\theta^{-1}(S_{i2})\right)}dG_\theta(x)\frac{f_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))} \\ &= \int_0^\infty e^{-x\left(\varphi_\theta^{-1}(S_{i1})+\varphi_\theta^{-1}(S_{i2})\right)}\left[\frac{-xf_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))}\right]dG_\theta(x). \end{aligned}$$

- A censored observation for the first subject and an event for the second subject

$(\delta_{i1} = 0, \delta_{i2} = 1)$:

$$\begin{aligned}
L_{i,(0,1)} &= -\frac{\partial S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i2}} \\
&= -\varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) \frac{d}{dx_{i2}}\left[\varphi_\theta^{-1}(S_{i2})\right] \\
&= \varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) \frac{f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))} \\
&= \int_0^\infty -x e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x) \frac{f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))} \\
&= \int_0^\infty e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} \left[\frac{-x f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))}\right] dG_\theta(x).
\end{aligned}$$

- Two events $(\delta_{i1} = 1, \delta_{i2} = 1)$:

$$\begin{aligned}
L_{i,(1,1)} &= f(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}) = \frac{\partial^2 S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1} \partial x_{i2}} \\
&= \varphi''_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) \frac{d}{dx_{i1}}\left[\varphi_\theta^{-1}(S_{i1})\right] \frac{d}{dx_{i2}}\left[\varphi_\theta^{-1}(S_{i2})\right] \\
&= \varphi''_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) \frac{f_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))} \frac{f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))} \\
&= \int_0^\infty x^2 e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x) \frac{f_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))} \frac{f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))} \\
&= \int_0^\infty e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} \left[\frac{-x f_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))}\right] \left[\frac{-x f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))}\right] dG_\theta(x).
\end{aligned}$$

In all four cases, the contribution to the bivariate copula likelihood can be written as

$$\int_0^\infty \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S_{ij})} \left[\frac{-x f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})))}\right]^{\delta_{ij}} dG_\theta(x).$$

□

Example: Bivariate Clayton copula likelihood

We calculate the bivariate copula likelihood for the Clayton copula starting from formula (1.1) on page 9 and show that the resulting expression is the same when starting from formula (2.2) on page 18.

The generator of the Clayton copula is $\varphi_\theta(s) = (1 + \theta s)^{1-/\theta}$ and its inverse is $\varphi_\theta^{-1}(s) = (s^{-\theta} - 1)/\theta$. In a sample of K pairs of time-to-event data

$\{(x_{11}, \delta_{11}), (x_{12}, \delta_{12})\}, \ldots, \{(x_{K1}, \delta_{K1}), (x_{K2}, \delta_{K2})\}$, we distinguish four possible situations for pair i:

- Two censored observations ($\delta_{i1} = 0, \delta_{i2} = 0$). The contribution to the likelihood is the joint survival function

$$\begin{aligned} L_{i,(0,0)} &= S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i1}) = C(S_{i1}, S_{i2}) \quad = \varphi_\theta(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})) \\ &= (S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta}. \end{aligned}$$

- An event for the first subject and a censored observation for the second subject ($\delta_{i1} = 1, \delta_{i2} = 0$):

$$\begin{aligned} L_{i,(1,0)} &= -\frac{\partial S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1}} \\ &= (S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta - 1} S_{i1}^{-\theta - 1} f_{i1}. \end{aligned}$$

- A censored observation for the first subject and an event for the second subject ($\delta_{i1} = 0, \delta_{i2} = 1$):

$$\begin{aligned} L_{i,(0,1)} &= -\frac{\partial S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i2}} \\ &= (S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta - 1} S_{i2}^{-\theta - 1} f_{i2}. \end{aligned}$$

- Two events ($\delta_{i1} = 1, \delta_{i2} = 1$):

$$\begin{aligned} L_{i,(1,1)} &= f(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}) \\ &= \frac{\partial^2 S(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1} \partial x_{i2}} \\ &= (1 + \theta)(S_{i1}^{-\theta} + S_i 2^{-\theta} - 1)^{-1/\theta - 2} S_{i1}^{-\theta - 1} S_{i2}^{-\theta - 1} f_{i1} f_{i2}. \end{aligned}$$

Combing the contributions of all K pairs according to (1.1), the bivariate Clayton copula likelihood is thus given by

$$\begin{aligned} L = \prod_{i=1}^{K} & \left[(1 + \theta)(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta - 2} S_{i1}^{-\theta - 1} S_{i2}^{-\theta - 1} f_{i1} f_{i2} \right]^{\delta_{i1}\delta_{i2}} \\ & \left[(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta - 1} S_{i1}^{-\theta - 1} f_{i1} \right]^{\delta_{i1}(1 - \delta_{i2})} \\ & \left[(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta - 1} S_{i2}^{-\theta - 1} f_{i2} \right]^{(1 - \delta_{i1})\delta_{i2}} \\ & \left[(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta} \right]^{(1 - \delta_{i1})(1 - \delta_{i2})}. \end{aligned} \tag{A.1}$$

We show that this very result is obtained when starting from formula (2.2):

$$L = \prod_{i=1}^{K} \int_0^\infty \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S_{ij})} \left[\frac{-x f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} dG_\theta(x). \tag{2.2}$$

We decompose the likelihood into the four possible contributions of a pair $\{(x_{i1}, \delta_{i1}), (x_{i2}, \delta_{i2})\}$

$$\begin{aligned} L = \prod_{i=1}^{K} & \left[\int_0^\infty x^2 \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S_{ij})} dG_\theta(x) \frac{f_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))} \frac{f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))} \right]^{\delta_{i1}\delta_{i2}} \\ & \left[\int_0^\infty -x \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S_{ij})} dG_\theta(x) \frac{f_{i1}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i1}))} \right]^{\delta_{i1}(1-\delta_{i2})} \\ & \left[\int_0^\infty -x \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S_{ij})} dG_\theta(x) \frac{f_{i2}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{i2}))} \right]^{(1-\delta_{i1})\delta_{i2}} \\ & \left[\int_0^\infty \prod_{j=1}^{2} e^{-x\varphi_\theta^{-1}(S_{ij}))} dG_\theta(x) \right]^{(1-\delta_{i1})(1-\delta_{i2})} . \end{aligned}$$

We use that $\frac{f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} = -\frac{d}{dx_{ij}} \left(\varphi_\theta^{-1}(S_{ij})\right) = [\varphi_\theta^{-1}(S_{ij})]' f_{ij}$ with $[\varphi_\theta^{-1}(S_{ij})]' \equiv \frac{d}{ds}[\varphi_\theta^{-1}(s)]\Big|_{s=S_{ij}}$ and write the product of exponential functions as an exponential function of a sum.

$$\begin{aligned} L = \prod_{i=1}^{K} & \left[\int_0^\infty x^2 e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x) [\varphi_\theta^{-1}(S_{i1})]' [\varphi_\theta^{-1}(S_{i2})]' f_{i1} f_{i2} \right]^{\delta_{i1}\delta_{i2}} \\ & \left[\int_0^\infty -x e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x) [\varphi_\theta^{-1}(S_{i1})]' f_{i1} \right]^{\delta_{i1}(1-\delta_{i2})} \\ & \left[\int_0^\infty -x e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x) [\varphi_\theta^{-1}(S_{i2})]' f_{i2} \right]^{(1-\delta_{i1})\delta_{i2}} \\ & \left[\int_0^\infty e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x) \right]^{(1-\delta_{i1})(1-\delta_{i2})} . \end{aligned}$$

Since we can write $\varphi_\theta(s)$ as a Laplace transform, $\varphi_\theta(s) = \int_0^\infty e^{-sx} dG_\theta(x)$, its first and second order derivatives are $\varphi'_\theta(s) = \int_0^\infty -x e^{-sx} dG_\theta(x)$ and $\varphi''_\theta(s) =$

$\int_0^\infty x^2 e^{-sx} dG_\theta(x)$, yielding

$$\varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) = \int_0^\infty -x e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x)$$
$$\varphi''_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) = \int_0^\infty x^2 e^{-x\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)} dG_\theta(x).$$

Substituting this into the likelihood expression, we get

$$\begin{aligned} L = \prod_{i=1}^{K} & \left[\varphi''_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)\left[\varphi_\theta^{-1}(S_{i1})\right]'\left[\varphi_\theta^{-1}(S_{i2})\right]' f_{i1} f_{i2}\right]^{\delta_{i1}\delta_{i2}} \\ & \left[\varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)\left[\varphi_\theta^{-1}(S_{i1})\right]' f_{i1}\right]^{\delta_{i1}(1-\delta_{i2})} \\ & \left[\varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)\left[\varphi_\theta^{-1}(S_{i2})\right]' f_{i2}\right]^{(1-\delta_{i1})\delta_{i2}} \\ & \left[\varphi_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right)\right]^{(1-\delta_{i1})(1-\delta_{i2})}. \end{aligned}$$

Using the explicit representation of the Clayton copula generator φ_θ and its inverse φ_θ^{-1} and all their requisite derivatives,

$$\begin{array}{lll} \varphi_\theta(s) = (1+\theta s)^{-1/\theta} & \varphi'_\theta(s) = -(1+\theta s)^{-1/\theta-1} & \varphi''_\theta(s) = (1+\theta s)^{-1/\theta-2} \\ \varphi_\theta^{-1}(s) = (s^{-\theta}-1)/\theta & \left[\varphi_\theta^{-1}(s)\right]' = -s^{-\theta-1} & \end{array}$$

$$\begin{aligned} \varphi_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) &= \left(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1\right)^{-1/\theta} \\ \varphi'_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) &= -\left(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1\right)^{-1/\theta-1} \\ \varphi''_\theta\left(\varphi_\theta^{-1}(S_{i1}) + \varphi_\theta^{-1}(S_{i2})\right) &= (1+\theta)\left(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1\right)^{-1/\theta-2} \end{aligned}$$

we get to the bivariate Clayton copula likelihood

$$\begin{aligned} L = \prod_{i=1}^{K} & \left[(1+\theta)(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta-2} S_{i1}^{-\theta-1} S_{i2}^{-\theta-1} f_{i1} f_{i2}\right]^{\delta_{i1}\delta_{i2}} \\ & \left[(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta-1} S_{i1}^{-\theta-1} f_{i1}\right]^{\delta_{i1}(1-\delta_{i2})} \\ & \left[(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta-1} S_{i2}^{-\theta-1} f_{i2}\right]^{(1-\delta_{i1})\delta_{i2}} \\ & \left[(S_{i1}^{-\theta} + S_{i2}^{-\theta} - 1)^{-1/\theta}\right]^{(1-\delta_{i1})(1-\delta_{i2})}. \end{aligned}$$

This is exactly the same as (A.1).

A.2 Validation of Clayton copula likelihood

In Section 2.3, the Clayton copula likelihood (2.7) on page 25 was derived from (2.3) using the closed form expression of the derivatives of its generator $\varphi_\theta(s) = (1+\theta s)^{-1/\theta}$

$$L = \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{f(x_{ij}|\boldsymbol{Z}_{ij})}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})))} \right]^{\delta_{ij}} \right) \varphi_\theta^{(d_i)} \left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij})) \right). \tag{2.3}$$

We show the equivalence of representation (2.2) and (2.3) for the Clayton copula, by calculating the Clayton copula likelihood (2.7) starting from (2.2)

$$L = \prod_{i=1}^{K} \int_0^{+\infty} \prod_{j=1}^{n_i} e^{-x\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z}_{ij}))} \left[\frac{-x f(x_{ij}|\boldsymbol{Z}_{ij})}{\varphi'_\theta(\varphi_\theta^{-1}(S(x_{ij}|\boldsymbol{Z_{ij}})))} \right]^{\delta_{ij}} dG_\theta(x). \tag{2.2}$$

The generator of the Clayton copula is the Laplace transform of the one-parameter gamma distribution, so

$$dG_\theta(x) = \frac{x^{1/\theta-1}\exp(-x/\theta)}{\theta^{1/\theta}\Gamma(1/\theta)} dx.$$

Using the shorthand notation $S_{ij} = S(x_{ij}|\boldsymbol{Z}_{ij})$ and $f_{ij} = f(x_{ij}|\boldsymbol{Z}_{ij})$, we rewrite the likelihood

$$\begin{aligned}
L &= \prod_{i=1}^{K} \int_0^{\infty} \prod_{j=1}^{n_i} e^{-x\varphi_\theta^{-1}(S_{ij})} \left[\frac{-x f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} dG_\theta(x) \\
&= \prod_{i=1}^{K} \int_0^{\infty} \prod_{j=1}^{n_i} e^{-x\varphi_\theta^{-1}(S_{ij})} \left[\frac{-x f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \frac{x^{1/\theta-1}e^{-x/\theta}}{\theta^{1/\theta}\Gamma(1/\theta)} dx \\
&= \prod_{i=1}^{K} \int_0^{\infty} e^{-x\left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij})+1/\theta\right)} x^{d_i} \left(\prod_{j=1}^{n_i} \left[\frac{-f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) \frac{x^{1/\theta-1}e^{-x/\theta}}{\theta^{1/\theta}\Gamma(1/\theta)} dx \\
&= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{-f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) \frac{1}{\theta^{1/\theta}\Gamma(1/\theta)} \int_0^{\infty} e^{-x\left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij})+1/\theta\right)} x^{d_i+1/\theta-1} dx \\
&= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{-f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) \frac{1}{\theta^{1/\theta}\Gamma(1/\theta)} \left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij}) + 1/\theta \right)^{-(d_i+1/\theta-1)} \\
&\qquad \int_0^{\infty} e^{-x\left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij})+1/\theta\right)} \left(x \left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij}) + 1/\theta \right) \right)^{d_i+1/\theta-1} dx.
\end{aligned}$$

We now use the definition of the gamma function $\Gamma(z) = \int_0^\infty e^{-t} z^{t-1} dt$:

$$L = \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{-f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) \frac{1}{\theta^{1/\theta}\Gamma(1/\theta)} \left(\sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij}) + 1/\theta \right)^{-(d_i+1/\theta)} \Gamma(d_i + 1/\theta).$$

At this point, a useful feature of the gamma function comes into play. From the property $\Gamma(z+1) = z\Gamma(z)$ we derive

$$\begin{aligned}
\Gamma(z+k) &= (z+k-1)\Gamma(z+k-1) \\
&= (z+k-1)(z+k-2)\Gamma(z+k-2) \\
&= (z+k-1)(z+k-2)\dots(z-1)z\Gamma(z) \\
&= \Gamma(z) \prod_{l=0}^{k-1} (z+l).
\end{aligned}$$

So

$$\begin{aligned}
\frac{\Gamma(d_i + 1/\theta)}{\Gamma(1/\theta)} &= \prod_{l=0}^{d_i-1} (1/\theta + l) \\
&= \prod_{l=0}^{d_i-1} (1/\theta(1 + l\theta)) \\
&= \frac{1}{\theta}(1+0\theta) \times \frac{1}{\theta}(1+\theta)\dots\frac{1}{\theta}(1+(d_i-1)\theta) \\
&= \theta^{-d_i} \prod_{l=0}^{d_i-1} (1 + l\theta).
\end{aligned}$$

Plugging this into the likelihood, we get exactly equation (2.7) on page 25:

$$\begin{aligned}
L &= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{-f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(S_{ij}))} \right]^{\delta_{ij}} \right) \left(1 + \theta \sum_{j=1}^{n_i} \varphi_\theta^{-1}(S_{ij}) \right)^{-(d_i+1/\theta)} \prod_{l=1}^{d_i-1} (1 + l\theta) \\
&= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{-f_{ij}}{-S_{ij}^{1+\theta}} \right]^{\delta_{ij}} \right) \left(1 + \theta \sum_{j=1}^{n_i} \frac{S_{ij}^{-\theta} - 1}{\theta} \right)^{-(d_i+1/\theta)} \prod_{l=1}^{d_i-1} (1 + l\theta) \\
&= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \left[\frac{f_{ij}}{S_{ij}^{1+\theta}} \right]^{\delta_{ij}} \right) \left(1 - n_i + \sum_{j=1}^{n_i} S_{ij}^{-\theta} \right)^{-(d_i+1/\theta)} \prod_{l=1}^{d_i-1} (1 + l\theta).
\end{aligned}$$

Appendix B

Likelihood contributions for quadrivariate survival data

The contributions to the likelihood (5.1) for the different association structures presented in Chapter 5 are given in this Appendix. For a sample of quadrivariate survival data

$$\{(x_{i1}, \delta_{i1}), (x_{i2}, \delta_{i2}), (x_{i3}, \delta_{i3}), (x_{i4}, \delta_{i4})\} \quad i = 1, \ldots, K,$$

the contribution L_i of quadruple i to the likelihood depends on the censoring status of the event times:

- For a cluster with no events: $L_i = S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)$.
- For a cluster with one event: $L_i = \dfrac{\partial S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{ij}}$.
- For a cluster with two events: $L_i = \dfrac{\partial^2 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{ij}\partial x_{ik}}$.
- For a cluster with three events: $L_i = \dfrac{\partial^3 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{ij}\partial x_{ik}\partial x_{il}}$.
- For a cluster with four events: $L_i = \dfrac{\partial^4 S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)}{\partial x_{i1}\partial x_{i2}\partial x_{i3}\partial x_{i4}}$.

According to the association structure that is specific to Model 0, 1, 2 and 3, the joint survival function $S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)$ takes on a different form.

We denote the joint survival function $S = S(x_{i1}, x_{i2}, x_{i3}, x_{i4}|Z_i)$ and the marginal survival functions $S_j = S_j(x_{ij}|Z_i)$, $j = 1, 2, 3, 4$. Furthermore, $S'_j = \dfrac{dS_j}{dx_{ij}}$.

B.1 Contributions to the likelihood of Model 0

In the independence model, the joint survival function of the quadrivariate lifetimes is

$$S = S_1 S_2 S_3 S_4$$

with derivatives

$$\begin{aligned}
\frac{\partial S}{\partial x_{ij}} &= S'_j S_k S_l S_m \text{ for } \{j,k,l,m\} = \{1,2,3,4\} \\
\frac{\partial^2 S}{\partial x_{ij}\partial x_{ik}} &= S'_j S'_k S_l S_m \\
\frac{\partial^3 S}{\partial x_{ij}\partial x_{ik}\partial x_{il}} &= S'_j S'_k S'_l S_m \\
\frac{\partial^4 S}{\partial x_{i1}\partial x_{i2}\partial x_{i3}\partial x_{i4}} &= S'_1 S'_2 S'_3 S'_4.
\end{aligned}$$

B.2 Contributions to the likelihood of Model 1

In the model with one level of clustering, the joint survival function of the quadrivariate lifetimes is

$$S = \varphi\left[\varphi^{-1}(S_1) + \varphi^{-1}(S_2) + \varphi^{-1}(S_3) + \varphi^{-1}(S_4)\right].$$

For the Clayton copula, $\varphi(t) = (1+\theta t)^{-1/\theta}$ and $\varphi^{-1}(t) = \frac{t^{-\theta}-1}{\theta}$, yielding

$$S = \left[S_1^{-\theta} + S_2^{-\theta} + S_3^{-\theta} + S_4^{-\theta} - 3\right]^{-1/\theta}.$$

Now put

$$\begin{aligned}
A &= \left[S_1^{-\theta} + S_2^{-\theta} + S_3^{-\theta} + S_4^{-\theta} - 3\right] \\
C_j &= S_j^{-\theta-1} S'_j \quad j = 1, 2, 3, 4,
\end{aligned}$$

then

$$\begin{aligned}
\frac{\partial S}{\partial x_{ij}} &= A^{-1/\theta-1}C_j \\
\frac{\partial^2 S}{\partial x_{ij}\partial x_{ik}} &= (1+\theta)A^{-1/\theta-2}C_jC_k \\
\frac{\partial^3 S}{\partial x_{ij}\partial x_{ik}\partial x_{il}} &= (1+\theta)(1+2\theta)A^{-1/\theta-3}C_jC_kC_l \\
\frac{\partial^4 S}{\partial x_{i1}\partial x_{i2}\partial x_{i3}\partial x_{i4}} &= (1+\theta)(1+2\theta)(1+3\theta)A^{-1/\theta-4}C_1C_2C_3C_4.
\end{aligned}$$

B.3 Contributions to the likelihood of Model 2

Lemma B.3.1. *For a sample of quadrivariate survival data*

$$\{(x_{i1},\delta_{i1}),(x_{i2},\delta_{i2}),(x_{i3},\delta_{i3}),(x_{i4},\delta_{i4})\} \qquad i=1,\ldots,K$$

the contribution of quadruple i to the nested Archimedean copula likelihood of Model 2 can be expressed as

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \prod_{j=1}^{2} e^{-x_1\left(\varphi_1^{-1}(S_{ij})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right)^{\delta_{ij}} dG_{1x}(x_1) \\
&\quad\times \int_0^\infty \prod_{j=3}^{4} e^{-x_1\left(\varphi_1^{-1}(S_{ij})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right)^{\delta_{ij}} dG_{1x}(x_1)dG_0(x).
\end{aligned}$$

Proof. In Section 5.3.3 we demonstrated that in the model with a parent copula with two identical child copulas, the joint survival function of the quadrivariate lifetimes is

$$\begin{aligned}
S &= \varphi_0\left[\varphi_0^{-1}\circ\varphi_1\left\{\varphi_1^{-1}(S_1)+\varphi_1^{-1}(S_2)\right\}+\varphi_0^{-1}\circ\varphi_1\left\{\varphi_1^{-1}(S_3)+\varphi_1^{-1}(S_4)\right\}\right] \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_1)+\varphi_1^{-1}(S_2)\right)} dG_{1x}(x_1) \\
&\quad\times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_3)+\varphi_1^{-1}(S_4)\right)} dG_{1x}(x_1)dG_0(x).
\end{aligned}$$

The contributions to the likelihood are the derivatives of the joint survival function w.r.t. the uncensored observations. We distinguish the following cases:

- One event

$$
\begin{aligned}
L_i &= \frac{\partial S}{\partial x_{ij}} \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right) dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} dG_{1x}(x_1) dG_0(x) \qquad \text{if } j = 1,2 \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right) dG_{1x}(x_1) dG_0(x) \\
&\qquad\qquad \text{if } j = 3,4.
\end{aligned}
$$

- Two events in one subcluster

$$
\begin{aligned}
L_i &= \frac{\partial^2 S}{\partial x_{ij} \partial x_{ik}} \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} \left(\frac{-x_1 f_{i1}}{\varphi_1'(\varphi_1^{-1}(S_{i1}))}\right) \left(\frac{-x_1 f_{i2}}{\varphi_1'(\varphi_1^{-1}(S_{i2}))}\right) dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} dG_{1x}(x_1) dG_0(x) \qquad \text{if } (j,k) = (1,2) \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} \left(\frac{-x_1 f_{i3}}{\varphi_1'(\varphi_1^{-1}(S_{i3}))}\right) \left(\frac{-x_1 f_{i4}}{\varphi_1'(\varphi_1^{-1}(S_{i4}))}\right) dG_{1x}(x_1) dG_0(x) \\
&\qquad\qquad \text{if } (j,k) = (3,4).
\end{aligned}
$$

- One event in the first subcluster and one event in the second subcluster

$$
\begin{aligned}
L_i &= \frac{\partial^2 S}{\partial x_{ij} \partial x_{ik}} \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right) dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} \left(\frac{-x_1 f_{ik}}{\varphi_1'(\varphi_1^{-1}(S_{ik}))}\right) dG_{1x}(x_1) dG_0(x) \\
&\qquad\qquad \text{if } (j,k) = (1,3),(1,4),(2,3),(2,4).
\end{aligned}
$$

- Three events

$$
\begin{aligned}
L_i &= \frac{\partial^3 S}{\partial x_{ij}\partial x_{ik}\partial x_{il}} \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} \left(\frac{-x_1 f_{i1}}{\varphi_1'(\varphi_1^{-1}(S_{i1}))}\right)\left(\frac{-x_1 f_{i2}}{\varphi_1'(\varphi_1^{-1}(S_{i2}))}\right) dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} \left(\frac{-x_1 f_{il}}{\varphi_1'(\varphi_1^{-1}(S_{il}))}\right) dG_{1x}(x_1)dG_0(x) \\
&\qquad\qquad \text{if } (j,k,l) = (1,2,3),(1,2,4) \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right) dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} \left(\frac{-x_1 f_{i3}}{\varphi_1'(\varphi_1^{-1}(S_{i3}))}\right)\left(\frac{-x_1 f_{i4}}{\varphi_1'(\varphi_1^{-1}(S_{i4}))}\right) dG_{1x}(x_1)dG_0(x) \\
&\qquad\qquad \text{if } (j,k,l) = (1,3,4),(2,3,4).
\end{aligned}
$$

- Four events

$$
\begin{aligned}
L_i &= \frac{\partial^4 S}{\partial x_{i1}\partial x_{i2}\partial x_{i3}\partial x_{i4}} \\
&= \int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} \left(\frac{-x_1 f_{i1}}{\varphi_1'(\varphi_1^{-1}(S_{i1}))}\right)\left(\frac{-x_1 f_{i2}}{\varphi_1'(\varphi_1^{-1}(S_{i2}))}\right) dG_{1x}(x_1) \\
&\quad \times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} \left(\frac{-x_1 f_{i3}}{\varphi_1'(\varphi_1^{-1}(S_{i3}))}\right)\left(\frac{-x_1 f_{i4}}{\varphi_1'(\varphi_1^{-1}(S_{i4}))}\right) dG_{1x}(x_1)dG_0(x).
\end{aligned}
$$

In all cases, the contribution to the likelihood of Model 2 for cluster i is given by

$$
\begin{aligned}
L_i = &\int_0^\infty \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i1})+\varphi_1^{-1}(S_{i2})\right)} \left(\frac{-x_1 f_{i1}}{\varphi_1'(\varphi_1^{-1}(S_{i1}))}\right)^{\delta_{i1}}\left(\frac{-x_1 f_{i2}}{\varphi_1'(\varphi_1^{-1}(S_{i2}))}\right)^{\delta_{i2}} dG_{1x}(x_1) \\
&\times \int_0^\infty e^{-x_1\left(\varphi_1^{-1}(S_{i3})+\varphi_1^{-1}(S_{i4})\right)} \left(\frac{-x_1 f_{i3}}{\varphi_1'(\varphi_1^{-1}(S_{i3}))}\right)^{\delta_{i3}}\left(\frac{-x_1 f_{i4}}{\varphi_1'(\varphi_1^{-1}(S_{i4}))}\right)^{\delta_{i4}} dG_{1x}(x_1)dG_0(x).
\end{aligned}
$$

. This can be rewritten as

$$
\begin{aligned}
L_i = &\int_0^\infty \int_0^\infty \prod_{j=1}^{2} e^{-x_1\left(\varphi_1^{-1}(S_{ij})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right)^{\delta_{ij}} dG_{1x}(x_1) \\
&\times \int_0^\infty \prod_{j=3}^{4} e^{-x_1\left(\varphi_1^{-1}(S_{ij})\right)} \left(\frac{-x_1 f_{ij}}{\varphi_1'(\varphi_1^{-1}(S_{ij}))}\right)^{\delta_{ij}} dG_{1x}(x_1)dG_0(x).
\end{aligned}
$$

□

As explained in Section 5.3.3, Lemma B.3 is more of theoretical rather than computational importance. Weighing the cost of using elaborate techniques for evaluating the inner

integrals, we choose to calculate the partial derivatives of the joint survival function to obtain the contributions to the likelihood of Model 2.

Choosing parent copula C_{θ_0} and child copulas C_{θ_1} to be Clayton copulas with generators $\varphi_0(t) = (1+\theta_0 t)^{-1/\theta_0}$ and $\varphi_1(t) = (1+\theta_1 t)^{-1/\theta_1}$, the joint survival function becomes

$$S = \left[-1 + \left(-1 + S_1^{-\theta_1} + S_2^{-\theta_1}\right)^{\theta_0/\theta_1} + \left(-1 + S_3^{-\theta_1} + S_4^{-\theta_1}\right)^{\theta_0/\theta_1}\right]^{-1/\theta_0}.$$

Now put

$$A = \left[-1 + \left(-1 + S_1^{-\theta_1} + S_2^{-\theta_1}\right)^{\theta_0/\theta_1} + \left(-1 + S_3^{-\theta_1} + S_4^{-\theta_1}\right)^{\theta_0/\theta_1}\right]$$

$$B_{12} = \left(-1 + S_1^{-\theta_1} + S_2^{-\theta_1}\right)$$

$$B_{34} = \left(-1 + S_3^{-\theta_1} + S_4^{-\theta_1}\right)$$

$$C_j = S_j^{-\theta_1 - 1} S_j' \quad j = 1,2,3,4$$

then

$$\frac{\partial S}{\partial x_{ij}} = \begin{cases} A^{-1/\theta_0 - 1} B_{12}^{\theta_0/\theta_1 - 1} C_j & \text{if } j = 1,2 \\ A^{-1/\theta_0 - 1} B_{34}^{\theta_0/\theta_1 - 1} C_j & \text{if } j = 3,4 \end{cases}$$

$$\frac{\partial^2 S}{\partial x_{ij} \partial x_{ik}} = \begin{cases} A^{-1/\theta_0 - 2} B_{ij}^{\theta_0/\theta_1 - 2} C_j C_k \left[(1+\theta_0) B_{jk}^{\theta_0/\theta_1} + (-\theta_0 + \theta_1) A\right] & \text{if } (j,k) = (1,2),(3,4) \\ (1+\theta_0) A^{-1/\theta_0 - 2} B_{12}^{\theta_0/\theta_1 - 1} B_{34}^{\theta_0/\theta_1 - 1} C_j C_k & \text{else} \end{cases}$$

$$\frac{\partial^3 S}{\partial x_{ij} \partial x_{ik} \partial x_{il}} = \begin{cases} (1+\theta_0) A^{-1/\theta_0 - 3} B_{12}^{\theta_0/\theta_1 - 2} B_{34}^{\theta_0/\theta_1 - 1} C_j C_k C_l \left[(1+2\theta_0) B_{12}^{\theta_0/\theta_1} + (-\theta_0 + \theta_1) A\right] & \\ \qquad \text{if } (j,k,l) = (1,2,3),(1,2,4) & \\ (1+\theta_0) A^{-1/\theta_0 - 3} B_{12}^{\theta_0/\theta_1 - 1} B_{34}^{\theta_0/\theta_1 - 2} C_j C_k C_l \left[(1+2\theta_0) B_{34}^{\theta_0/\theta_1} + (-\theta_0 + \theta_1) A\right] & \\ \qquad \text{if } (j,k,l) = (1,3,4),(2,3,4) & \end{cases}$$

$$\begin{aligned} \frac{\partial^4 S}{\partial x_{i1} \partial x_{i2} \partial x_{i3} \partial x_{i4}} &= (1+\theta_0) A^{-1/\theta_0 - 4} B_{12}^{\theta_0/\theta_1 - 2} B_{34}^{\theta_0/\theta_1 - 2} C_1 C_2 C_3 C_4 \\ &\quad \left[(1+2\theta_0)(1+3\theta_0) B_{12}^{\theta_0/\theta_1} B_{34}^{\theta_0/\theta_1} + (1+2\theta_0)(-\theta_0 + \theta_1) A \left(B_{12}^{\theta_0/\theta_1} + B_{34}^{\theta_0/\theta_1}\right)\right. \\ &\quad \left. + (-\theta_0 + \theta_1)^2 A^2\right]. \end{aligned}$$

B.4 Contributions to the likelihood of Model 3

In the model with a parent copula with two different child copulas, the joint survival function of the quadrivariate lifetimes is, for the Clayton copula

$$S = \left[-1 + \left(-1 + S_1^{-\theta_1} + S_2^{-\theta_1}\right)^{\theta_0/\theta_1} + \left(-1 + S_3^{-\theta_2} + S_4^{-\theta_2}\right)^{\theta_0/\theta_2}\right]^{-1/\theta_0}.$$

Now put

$$A = \left[-1 + \left(-1 + S_1^{-\theta_1} + S_2^{-\theta_1}\right)^{\theta_0/\theta_1} + \left(-1 + S_3^{-\theta_2} + S_4^{-\theta_2}\right)^{\theta_0/\theta_2}\right]$$

$$B_{12} = \left(-1 + S_1^{-\theta_1} + S_2^{-\theta_1}\right)$$

$$B_{34} = \left(-1 + S_3^{-\theta_2} + S_4^{-\theta_2}\right)$$

$$C_{j1} = S_j^{-\theta_1 - 1} S_j' \quad j = 1, 2$$

$$C_{j2} = S_j^{-\theta_2 - 1} S_j' \quad j = 3, 4$$

then

$$\frac{\partial S}{\partial x_{ij}} = \begin{cases} A^{-1/\theta_0 - 1} B_{12}^{\theta_0/\theta_1 - 1} C_{j1} & \text{if } j = 1, 2 \\ A^{-1/\theta_0 - 1} B_{34}^{\theta_0/\theta_2 - 1} C_{j2} & \text{if } j = 3, 4 \end{cases}$$

$$\frac{\partial^2 S}{\partial x_{ij} \partial x_{ik}} = \begin{cases} A^{-1/\theta_0 - 2} B_{12}^{\theta_0/\theta_1 - 2} C_{11} C_{21} \left[(1+\theta_0) B_{12}^{\theta_0/\theta_1} + (-\theta_0 + \theta_1) A\right] & \text{if } (j,k) = (1,2) \\ A^{-1/\theta_0 - 2} B_{34}^{\theta_0/\theta_2 - 2} C_{32} C_{42} \left[(1+\theta_0) B_{34}^{\theta_0/\theta_2} + (-\theta_0 + \theta_2) A\right] & \text{if } (j,k) = (3,4) \\ (1+\theta_0) A^{-1/\theta_0 - 2} B_{12}^{\theta_0/\theta_1 - 1} B_{34}^{\theta_0/\theta_2 - 1} C_{j1} C_{k2} & \text{if } j \in \{1,2\} \text{ and } k \in \{3,4\} \end{cases}$$

$$\frac{\partial^3 S}{\partial x_{ij} \partial x_{ik} \partial x_{il}} = \begin{cases} (1+\theta_0) A^{-1/\theta_0 - 3} B_{12}^{\theta_0/\theta_1 - 2} B_{34}^{\theta_0/\theta_2 - 1} C_{j1} C_{k1} C_{l2} \left[(1+2\theta_0) B_{12}^{\theta_0/\theta_1} + (-\theta_0 + \theta_1) A\right] \\ \qquad\qquad \text{if } (j,k,l) = (1,2,3), (1,2,4) \\ (1+\theta_0) A^{-1/\theta_0 - 3} B_{12}^{\theta_0/\theta_1 - 1} B_{34}^{\theta_0/\theta_2 - 2} C_{j1} C_{k2} C_{l2} \left[(1+2\theta_0) B_{34}^{\theta_0/\theta_2} + (-\theta_0 + \theta_2) A\right] \\ \qquad\qquad \text{if } (j,k,l) = (1,3,4), (2,3,4) \end{cases}$$

$$\begin{aligned} \frac{\partial^4 S}{\partial x_{i1} \partial x_{i2} \partial x_{i3} \partial x_{i4}} &= (1+\theta_0) A^{-1/\theta_0 - 4} B_{12}^{\theta_0/\theta_1 - 2} B_{34}^{\theta_0/\theta_2 - 2} C_{11} C_{21} C_{32} C_{42} \\ &\quad \Big[(1+2\theta_0)(1+3\theta_0) B_{12}^{\theta_0/\theta_1} B_{34}^{\theta_0/\theta_2} + (1+2\theta_0) A((-\theta_0 + \theta_2) B_{12}^{\theta_0/\theta_1} + (-\theta_0 + \theta_1) B_{34}^{\theta_0/\theta_2}) \\ &\quad + (-\theta_0 + \theta_1)(-\theta_0 + \theta_2) A^2\Big]. \end{aligned}$$

B.4.1 R code for one-stage estimation in Model 3

```
R> head(udder)

  cowid location      time status parity
1     1       LF 0.7337440      1      1
2     1       RF 0.7337440      1      1
3     1       LR 1.3032170      1      1
4     1       RR 0.7337440      1      1
5     2       LF 0.7282683      1      0
6     2       RF 2.2176591      1      0
```

```
Q1 <-udder[udder$location=="LF",]
Q2 <-udder[udder$location=="RF",]
Q3 <-udder[udder$location=="LR",]
Q4 <-udder[udder$location=="RR",]

t1 <- Q1$time; c1 <- Q1$status; par1 <- Q1$parity
t2 <- Q2$time; c2 <- Q2$status; par2 <- Q2$parity
t3 <- Q3$time; c3 <- Q3$status; par3 <- Q3$parity
t4 <- Q4$time; c4 <- Q4$status; par4 <- Q4$parity

loglik <- function(p){ #loglikelihood for one-stage estimation of Model 3

    lambda1 <- exp(p[1]) #exp transform to make sure that
    rho1    <- exp(p[2]) #estimated parameters are positive
    lambda2 <- exp(p[3])
    rho2    <- exp(p[4])
    lambda3 <- exp(p[5])
    rho3    <- exp(p[6])
    lambda4 <- exp(p[7])
    rho4    <- exp(p[8])
    beta    <- p[9]      #covariate effect can be negative
    theta1  <- exp(p[10])
    theta2  <- exp(p[11])
    theta3  <- exp(p[12])

    S1 <- exp(-lambda1*t1^rho1*exp(beta*par1));
    S2 <- exp(-lambda2*t2^rho2*exp(beta*par2));
    S3 <- exp(-lambda3*t3^rho3*exp(beta*par3));
    S4 <- exp(-lambda4*t4^rho4*exp(beta*par4));

    DS1 <- -lambda1*rho1*t1^(rho1-1)*exp(beta*par1)*S1;
    DS2 <- -lambda2*rho2*t2^(rho2-1)*exp(beta*par2)*S2;
    DS3 <- -lambda3*rho3*t3^(rho3-1)*exp(beta*par3)*S3;
```

```
DS4 <- -lambda4*rho4*t4^(rho4-1)*exp(beta*par4)*S4;

A <- -1+(-1+S1^(-theta2)+S2^(-theta2))^(theta1/theta2)+(-1+S3^(-theta3)+
     S4^(-theta3))^(theta1/theta3)
B12 <- -1+S1^(-theta2)+S2^(-theta2)
B34 <- -1+S3^(-theta3)+S4^(-theta3)
C1 <- S1^(-theta2-1)*DS1
C2 <- S2^(-theta2-1)*DS2
C3 <- S3^(-theta3-1)*DS3
C4 <- S4^(-theta3-1)*DS4

#joint survival function
S <- A^(-1/theta1)

#first order partial derivatives
dS1 <- A^(-1/theta1-1)*B12^(theta1/theta2-1)*C1
dS2 <- A^(-1/theta1-1)*B12^(theta1/theta2-1)*C2
dS3 <- A^(-1/theta1-1)*B34^(theta1/theta3-1)*C3
dS4 <- A^(-1/theta1-1)*B34^(theta1/theta3-1)*C4

#second order partial derivatives
d2S12 <- A^(-1/theta1-2)*B12^(theta1/theta2-2)*C1*C2*((1+theta1)*B12^(theta1/theta2)+
         (-theta1+theta2)*A)
d2S13 <- (1+theta1)*A^(-1/theta1-2)*B12^(theta1/theta2-1)*B34^(theta1/theta3-1)*C1*C3
d2S14 <- (1+theta1)*A^(-1/theta1-2)*B12^(theta1/theta2-1)*B34^(theta1/theta3-1)*C1*C4
d2S23 <- (1+theta1)*A^(-1/theta1-2)*B12^(theta1/theta2-1)*B34^(theta1/theta3-1)*C2*C3
d2S24 <- (1+theta1)*A^(-1/theta1-2)*B12^(theta1/theta2-1)*B34^(theta1/theta3-1)*C2*C4
d2S34 <- A^(-1/theta1-2)*B34^(theta1/theta3-2)*C3*C4*((1+theta1)*B34^(theta1/theta3)+
         (-theta1+theta3)*A)

#third order partial derivatives
d3S123 <- (1+theta1)*A^(-1/theta1-3)*B12^(theta1/theta2-2)*B34^(theta1/theta3-1)*
          C1*C2*C3*((1+2*theta1)*B12^(theta1/theta2)+(-theta1+theta2)*A)
d3S124 <- (1+theta1)*A^(-1/theta1-3)*B12^(theta1/theta2-2)*B34^(theta1/theta3-1)*
          C1*C2*C4*((1+2*theta1)*B12^(theta1/theta2)+(-theta1+theta2)*A)
d3S134 <- (1+theta1)*A^(-1/theta1-3)*B12^(theta1/theta2-1)*B34^(theta1/theta3-2)*
          C1*C3*C4*((1+2*theta1)*B34^(theta1/theta3)+(-theta1+theta3)*A)
d3S234 <- (1+theta1)*A^(-1/theta1-3)*B12^(theta1/theta2-1)*B34^(theta1/theta3-2)*
          C2*C3*C4*((1+2*theta1)*B34^(theta1/theta3)+(-theta1+theta3)*A)

#fourth order partial derivatives
d4S1234 <- (1+theta1)*A^(-1/theta1-4)*B12^(theta1/theta2-2)*B34^(theta1/theta3-2)*
           C1*C2*C3*C4*((1+2*theta1)*(1+3*theta1)*B12^(theta1/theta2)*B34^(theta1/theta3)+
           (1+2*theta1)*A*((-theta1+theta3)*B12^(theta1/theta2)+(-theta1+theta2)*
           B34^(theta1/theta3))+(-theta1+theta2)*(-theta1+theta3)*A^2)

terms <- log(S^((1-c1)*(1-c2)*(1-c3)*(1-c4))*
```

```
                    (-dS1)^(c1*(1-c2)*(1-c3)*(1-c4))*
                    (-dS2)^((1-c1)*c2*(1-c3)*(1-c4))*
                    (-dS3)^((1-c1)*(1-c2)*c3*(1-c4))*
                    (-dS4)^((1-c1)*(1-c2)*(1-c3)*c4)*
                    d2S12^(c1*c2*(1-c3)*(1-c4))*
                    d2S13^(c1*(1-c2)*c3*(1-c4))*
                    d2S14^(c1*(1-c2)*(1-c3)*c4)*
                    d2S23^((1-c1)*c2*c3*(1-c4))*
                    d2S24^((1-c1)*c2*(1-c3)*c4)*
                    d2S34^((1-c1)*(1-c2)*c3*c4)*
                    (-d3S123)^(c1*c2*c3*(1-c4))*
                    (-d3S124)^(c1*c2*(1-c3)*c4)*
                    (-d3S134)^(c1*(1-c2)*c3*c4)*
                    (-d3S234)^((1-c1)*c2*c3*c4)*
                    d4S1234^(c1*c2*c3*c4))
    -sum(terms)}

init.vals <- c(rep(log(0.5),8),0.5,rep(log(0.5),3))
res <- nlm(loglik,init.vals,hessian=TRUE)

lambda1 <- exp(res$estimate[1])
lambda2 <- exp(res$estimate[3])
lambda3 <- exp(res$estimate[5])
lambda4 <- exp(res$estimate[7])
rho1 <- exp(res$estimate[2])
rho2 <- exp(res$estimate[4])
rho3 <- exp(res$estimate[6])
rho4 <- exp(res$estimate[8])
beta <- res$estimate[9]
theta1 <- exp(res$estimate[10])
theta2 <- exp(res$estimate[11])
theta3 <- exp(res$estimate[12])

#Calculate standard errors using delta method
stderr <- sqrt(diag(solve(res$hessian)))
se <- c(lambda1,rho1,lambda2,rho2,lambda3,rho3,lambda4,rho4,1,theta1,theta2,theta3)*stderr

lambda1; se[1] #0.156654(0.01055878)
rho1; se[2]    #1.298542(0.04833772)
lambda2; se[3] #0.1657836(0.01089846)
rho2; se[4]    #1.263716(0.04677189)
lambda3; se[5] #0.1338474(0.009678422)
rho3; se[6]    #1.311082(0.05189555)
lambda4; se[7] #0.1419053(0.009947461)
rho4; se[8]    #1.269656(0.04981163)
beta; se[9]    #-0.3400679(0.06777242)
theta1; se[10] #3.047683(0.184657)
```

```
theta2; se[11] #3.588606(0.2711444)
theta3; se[12] #3.512632(0.2742904)
```

Samenvatting

In overlevingsstudies analyseert men de tijd totdat een bepaalde gebeurtenis zich voordoet. Naast de levensduur van personen, zijn andere voorbeelden van overlevingstijden: de tijd tot hervallen van een ziekte, de levensduur van elektronische componenten, of de tijd dat een persoon in een bejaardentehuis verblijft. Als de gebeurtenis zich voordoet bij alle personen of studieobjecten, zijn er vele technieken toepasbaar. Maar in vele gevallen hebben we geen exacte informatie over de overlevingstijd, bijvoorbeeld als de patiënt nog leeft of als de elektronische component het nog niet begeven heeft aan het einde van de studie. In zo'n gevallen spreken we van censurering en komen er specifieke statistische technieken aan te pas. We willen immers zoveel mogelijk informatie meenemen in de analyse, ook als dit slechts een ondergrens is voor de ware overlevingstijd.

Vaak zijn overlevingstijden ook nog eens gegroepeerd in clusters, zoals bij het opvolgen van kankerpatiënten in verschillende ziekenhuizen, de levensduur van tweelingen of de tijd tot infectie van de 4 uierkwartieren bij een melkkoe. De overlevingstijden binnen een cluster vertonen doorgaans een zekere associatie. Om met geclusterde overlevingsdata te werken, zijn er verschillende technieken voorhanden: frailty modellen en copula modellen. Beide modellen hebben hun voordelen en beperkingen. Bij het frailty model is de grootste beperking dat de functies die de marginale overlevingskansen beschrijven, afhangen van de associatieparameter die de gezamenlijke overlevingsfunctie bepaalt. Bij het copula model is dit niet zo en staan de marginale overlevingsfuncties los van de associatiestructuur. Het copula model is dan weer minder geschikt wanneer de clusters groot zijn, of in grootte verschillen.

In Hoofdstuk 2 proberen we om de methodologie van copula modellen meer flexibel te maken zodat deze toch met variabele clustergroottes om kan. We doen dit voor de belangrijke klasse van de Archimedische copula's. Deze copula's hebben een breed toepassingsdomein omwille van verscheidene redenen: ze zijn gemakkelijk te construeren, er is een grote variëteit aan mogelijke associatiestructuren, en ze hebben mooie wiskundige eigenschappen. Door deze eigenschappen slim te benutten, kunnen we het Archimedisch copula model op zo'n manier herschrijven dat de grootte van de clusters geen enkel probleem meer vormt. Si-

mulatiestudies tonen aan dat de schatters van de associatieparameter die uit dit vernieuwde model voortkomen, "goede" schatters zijn (in wiskundetaal: consistent en asympotisch normaal verdeeld). We bewijzen dit formeel in Hoofdstuk 3 en geven eveneens uitdrukkingen voor de variantie van de schatters.

Natuurlijk bestaat er naast de theoretische fraaiheid van het vernieuwde Archimedische copula model, ook nog het praktische aspect. We illustreren hoe dit model kan gebruikt worden bij een dataset die stamt uit de diergeneeskunde. Bij melkkoeien ligt de tijd tussen twee worpen optimaal tussen 12 en 13 maanden. Een belangrijke factor die de lengte van dit interval beïnvloedt, is de tijd tussen het werpen en de eerstvolgende inseminatie. In een melkveebedrijf worden de koeien doorgaans door de boer zelf geïnsemineerd, afhankelijk van zijn ervaring. Om meer inzicht te krijgen in het inseminatieproces, fitten we ons model op een dataset die gegevens bevat uit 181 melkveebedrijven van uiteenlopende grootte. Omdat het aantal worpen van een koe een invloed kan hebben op de tijd tot de eerstvolgende inseminatie, nemen we deze informatie mee op in het model. We vergelijken verschillende schattingsprocedures voor een greep uit het assortiment aan Archimedische copula functies en voor verschillende keuzes van de baseline overlevingsfuncties. Om ons model toegankelijk te maken voor andere gebruikers, hebben we een softwarepakket ontwikkeld voor de statistische softwareomgeving R. In Hoofdstuk 4 geven we hiervan een korte gebruiksaanwijzing.

In de praktijk komen we regelmatig datasets tegen waarin overlevingstijden meerdere niveaus van clustering vertonen. Denk bijvoorbeeld aan de levensduur van leden van Afrikaanse families, die gegroepeerd zijn in afgelegen dorpen. Om de mogelijkheden en tekortkomingen van het Archimedische copula model in zo'n setting te onderzoeken, starten we in Hoofdstuk 5 vanuit een dataset met dimensie 4. Deze dataset bevat de tijden tot infectie van de 4 uierdelen van 1196 melkkoeien. Een uierinfectie is nefast voor de melkproductie en de kwaliteit van de melk, daarom is het belangrijk om inzicht te verwerven in de manier waarop een infectie zich over de uier verspreidt. We vertalen dit probleem naar copula-taal door de associatie tussen de infectietijden van de vier uierdelen te onderzoeken. Het aantal keren dat een koe gekalfd heeft, kan invloed hebben op de staat van de uierspenen, en dus ook op de vatbaarheid voor infecties. De pariteit, die we dichotomiseren als primipaar (1 keer gekalfd) of multipaar (meer dan eens gekalfd), wordt als verklarende variabele opgenomen in het model. We vergelijken modellen die verschillende associatiestructuren handhaven, waarbij we rekening houden met de biologische relevantie van deze modellen. Als je een koe van dichtij bekijkt, zie je dat de voorste twee uierspenen dicht bij elkaar liggen en de achterste twee ook. Er is meer ruimte tussen de voorste en achterste uierdelen. De voorste en achterste paren kunnen dus als subclusters van de uier worden beschouwd. Zoals verwacht komt het geneste Archimedische copula model met 2 evenwaardige subclusters als beste model uit onze vergelijking. De correlatie tussen infectietijden van de voorste uierdelen is gelijk aan de correlatie tussen de achterste uierdelen, en is groter dan de correlatie tussen een speen vooraan en een speen achteraan. Maar doordat de correlatie globaal gezien erg groot is, heeft dit resultaat weinig praktisch nut voor de melkveehouder. Zodra er een speen geïnfecteerd is, moeten er preventieve zorgen geboden worden aan de andere spenen omdat de infectie zich snel verspreidt over de gehele uier.

In de ontwikkeling van hiërarchische copula modellen is het echter wel van belang om te kunnen zien of er onderscheid kan gemaakt worden in de associatie van hoofd-en subclusters. Daarom hebben we een aantal simulatiestudies uitgevoerd om o.a. de gevoeligheid te bepalen waarmee het onderscheid tussen hoofd-en subclusters gedetecteerd wordt. Het vierdimensionale geval wordt in Hoofdstuk 5 als uitgangspunt gebruikt om de modellen te kunnen veralgemenen naar hogere dimensies. Ondanks de mooie eigenschappen van de klasse van Archimedische copula's, blijft het een uitdaging om de berekening van de bouwstenen van het hiërarchische model te generaliseren. Dit is iets dat in de toekomst zeker nog de nodige aandacht verdient.