UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Modelling multivariate clustered left-censored data by using copula functions.

Miguel-Angel Beynaerts
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :
Prof. dr. Roel BRAEKERS

2023
2024

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Modelling multivariate clustered left-censored data by using copula functions.

Miguel-Angel Beynaerts

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Roel BRAEKERS

Acknowledgment

I would like to extend my gratitude to Roel Braekers who proposed this topic as my thesis project. Survival analysis has always been my main focus within statistics, and I am therefore thankful to have had the opportunity to study more advanced survival topics both theoretically and in an applied context. Furthermore, I would like to thank Roel for guiding me throughout the entire project, even in busy times.

Abstract

Recent work by Prenen et al. (2017) and Othus & Li (2010) has led to a broad family of copula models for the analysis of right-censored clustered data. On the other hand, few applications for left-censored data have been considered in the literature. In this text we aim to extend the existing copula methodology to account for left-censored data. A copula model models the marginal distributions of the clustered outcomes independently from the dependency structure within the cluster. The dependency is then imposed by linking the margins through a copula function. After specification of the joint distribution function the full likelihood can then be derived taking into account the clustered structure of the data. More specifically, we will focus on the popular family of Archimedean copulas with completely monotone generator and the Gaussian copula. The latter has the advantage that the association structure is imposed by means of a distribution rather than a function which makes generalizations more straightforward. As we will focus on fully parametric models, a more flexible model alternative for the margins based on the work by Royston & Parmar (2002) will be described. Both one-stage and two-stage estimation procedures are discussed in terms of their consistency and asymptotic properties, and evaluated by an extensive simulation study. Finally, the proposed methods are demonstrated on a longitudinal study that assesses the effect of peritoneal administration of ethanol on the sleep time of mice.

Contents

1 Introduction

In survival analysis the outcome of interest generally is a time until an event of interest occurs. If the event occurred for all individuals or units within the follow-up period standard techniques would be applicable to model the outcome of interest. Many times, however, the event of interest does not occur. For example, in a clinical trial an individual might die due to causes unrelated to the event of interest or drops out from the trial. On the other hand we do know that the event of interest occurs eventually, meaning that the time to the event of interest is longer than what has been observed during the study. This mechanism, called censoring, is the foundation of time-to-event data analysis. Multivariate survival analysis generally deals with clustered data subject to a censoring mechanism. Clustering can take many forms: in a trial the event of interest is recorded multiple times for the same individual, implying that a longitudinal component is present. Similarly, in a large multi-center clinical trial, multiple individuals are clustered within centers. These examples have in common that the individuals or units within the same cluster have similar characteristics with respect to the hazard of the event of interest. This commonality is usually modeled by imposing a correlation structure between the outcomes within a cluster.

While standard techniques in survival analysis aim to deal with right censoring where it is known that the event of interest occurs at some later time point, and/or left truncation when no information whatsoever about the event of interest is known below the truncation limit. Many applications in various fields of research, however, require techniques that can deal with left censoring since the event of interest is known to occur before the start of the study. In the literature this is often referred to as data subject to a limit of detection (LOD) or a quantification limit. For example in environmental sciences and ecology, one often deals with chemicals or particles that are reported above a well-specified laboratory reporting limit (*Statistical Methods in Water Resources*, 2020). On a similar note, contamination levels in food subject to a limit of detection are recorded to assess the risk of contamination in terms of weekly food intake (Tressou, 2006). In clinical research one often deals with quantification limits, for example when using ultrasensitive assay data in measuring HIV-1 RNA levels (Jacqmin-Gadda, 2000).

Despite the increasing need for methods that can handle left-censored data, only primitive methods are readily available and popular. Helsel (2011) provides an extensive overview of the most popular methods that are being used: single imputation, also referred to as substitution or fabrication (Helsel, 2006), implies the substitution of a single measure such as for example $\text{LOD}/\sqrt{2}$ and has been shown to lead to invalid estimates (Canales et al., 2018; Hewett & Ganser, 2007). Multiple imputation on the other hand does lead to consistent estimation (Canales et al., 2018). When combining left censoring with clustered data the list of available methods is even more restrictive. Jacqmin-Gadda (2000) extend a linear mixed model to account for left censoring in a longitudinal data setting. To this end, we aim to develop a general framework that allows for easy interpretation of the within-cluster dependence structure and flexible modelling of the margins.

To model the dependency between outcomes within the same cluster, we aim to impose the dependency independent from the specification of the marginal distribution. Indeed, in a conditional model such as the frailty model, the marginal distribution depends on the association parameter which makes it difficult to interpret (Duchateau & Janssen, 2008). Alternatively the dependency structure can be imposed by coupling the margins, thus forming the joint distribution function, through a special function called the copula (Nelsen, 2006). The relationship between the multivariate distribution function and its univariate margins will play a crucial role in the modelling of the survival outcomes. In fact, Sklar's theorem (Sklar, 1959) guarantees that a unique copula $\mathcal{C}$ exists such that

$$H(x, y) = \mathcal{C}\left(G(x), G(y)\right)$$

where $H(x, y)$ is the joint cumulative distribution function of random variables X and Y with univariate margins $F(x)$ and $G(y)$. We will use Sklar's theorem extensively as the foundation of our model for multivariate left-censored data.

Firstly, we introduce the copula model in its general form. Using Sklar's theorem the full likelihood for clustered data accounting for left-censoring can be specified. At this point the general framework can be used to apply the likelihood to specific families of copulas. Furthermore we will particularly focus on how to estimate the model parameters efficiently. The asymptotic properties of the estimators will also be discussed. As parametric specification of the univariate margins is an important element in the estimation process, a flexible marginal alternative through the use of splines will be proposed. Moreover, an extensive simulation exercise will be performed to assess the performance of the estimators under various degrees of misspecification. Finally, we demonstrate the proposed methods on a longitudinal study that assesses the effect of peritoneal administration of ethanol on the sleeping time of mice by Markel et al. (1995).

2 Copula model for left-censored clustered data

In this section we extend the current copula methodology for the Archimedean family (Prenen et al., 2017) and the Gaussian family (Othus & Li, 2010) to account for left censoring. We present a general expression for the likelihood of the copula model assuming a fixed cluster size and covariate structure. Clusters of varying size and cluster-specific covariate structures will be discussed briefly. Furthermore, two parametric estimation methods that are commonly used in copula models for right-censored data will be brought forward for which we discuss some asymptotic theoretical results. Building on this theoretical foundation we introduce a new flexible parametric model based on natural cubic splines (Royston & Parmar, 2002).

2.1 Model description

To develop a copula model for clustered data subject to left censoring, we introduce the following notation used throughout the entire text. Let i be the index of a cluster such that $i = 1, \dots, K$ with K the total number of clusters in the data set. Within each cluster i let T_{ij} denote outcome j where $j = 1, \dots, n_i$ with n_i the total number of outcomes (or units) in cluster i. When developing the copula model the focus will lie in particular on the case where the cluster size is fixed $n_1 = \dots = n_K = n$. In a survival context each outcome within a cluster may or may not be observed. Therefore we further assume a random left censoring scheme denoted by the random variable C_{ij}. In practice we observe the following quantities for outcome j in cluster i

$$\begin{aligned} X_{ij} &= \max(T_{ij}, C_{ij}) \\ \delta_{ij} &= I(T_{ij} > C_{ij}). \end{aligned}$$

In general the probability distribution may depend on (possibly cluster-specific) covariates. We denote the set of covariates for outcome j in cluster i as $\boldsymbol{Z}_{ij} = (Z_{ij1}, \dots, Z_{ijp})^T$ with p the total number of covariates. This leads to the joint survival probability for cluster i

$$S(t_{i1}, \dots, t_{in_i} | \boldsymbol{Z}_{i1}, \dots, \boldsymbol{Z}_{in_i}) = P(T_{i1} > t_{i1}, \dots, T_{in_i} > t_{in_i} | \boldsymbol{Z}_{i1}, \dots, \boldsymbol{Z}_{in_i})$$

or more appropriately in the context of left censoring to the joint cumulative distribution function

$$F(t_{i1}, \dots, t_{in_i} | \boldsymbol{Z}_{i1}, \dots, \boldsymbol{Z}_{in_i}) = P(T_{i1} \leq t_{i1}, \dots, T_{in_i} \leq t_{in_i} | \boldsymbol{Z}_{i1}, \dots, \boldsymbol{Z}_{in_i}). \tag{2.1}$$

Now we are in a position to use Sklar's theorem (Sklar, 1959) to our advantage. Let $\mathcal{C}$ be the copula function with association parameter θ and $F_j(x_{ij} | \boldsymbol{Z}_{ij})$ the marginal CDF of outcome j conditional on the set of covariates $\boldsymbol{Z}_{ij}$ and evaluated in event x_{ij}, then the joint CDF can be written in function of the marginal CDF for each outcome in cluster i.

$$F(x_{i1}, \dots, x_{in_i} | \boldsymbol{Z}_{i1}, \dots, \boldsymbol{Z}_{in_i}) = \mathcal{C}(F_1(x_{ij} | \boldsymbol{Z}_{ij}), \dots, F_{n_i}(x_{in_i} | \boldsymbol{Z}_{in_i})) \tag{2.2}$$

Equation (2.2) will be the basis to derive the likelihood which will be used to estimate the association parameter θ and possibly also the marginal parameters $\boldsymbol{\beta}$. Indeed, the univariate likelihood can now

be easily extended to the multivariate clustered case since the joint density can be obtained directly from the marginal densities and the copula density (Nelsen, 2006). For a cluster k where all events are observed, we then have

$$\begin{aligned} f(x_{k1}, \dots, x_{kn_k} | \boldsymbol{Z}_{k1}, \dots, \boldsymbol{Z}_{kn_k}) &= \frac{\partial^{(n_k)} \mathcal{C}\left(F_1(x_{k1}|\boldsymbol{Z}_{k1}), \dots, F_{n_k}(x_{kn_k}|\boldsymbol{Z}_{kn_k})\right)}{\partial x_{k1} \dots \partial x_{kn_k}} \\ &= c\left(F_1(x_{k1}|\boldsymbol{Z}_{k1}), \dots, F_{n_k}(x_{kn_k}|\boldsymbol{Z}_{kn_k})\right) f_1(x_{k1}|\boldsymbol{Z}_{k1}) \dots f_{n_k}(x_{kn_k}|\boldsymbol{Z}_{kn_k}) \end{aligned}$$

where $\partial^{(n_k)}$ denotes a partial derivative of the n_k-th degree and c the copula density. On a similar note, for a cluster k where some events are observed and some are censored we have a likelihood contribution of

$$L_k = \frac{\partial^{(d_k)} \mathcal{C}\left(F_1(x_{k1}|\boldsymbol{Z}_{k1}), \dots, F_{n_k}(x_{kn_k}|\boldsymbol{Z}_{kn_k})\right)}{\partial\{\delta_{kj} = 1\}} f_1(x_{k1}|\boldsymbol{Z}_{k1})^{\delta_{k1}} \dots f_{n_k}(x_{kn_k}|\boldsymbol{Z}_{kn_k})^{\delta_{kn_k}} \tag{2.3}$$

where d_k and $\{\delta_{kj} = 1\}$ are the total number of observed events and the set of all observed events in cluster k respectively. The total likelihood is then given by

$$L = \prod_{k=1}^{K} L_k \tag{2.4}$$

In the subsequent subsections we derive the likelihood in detail for the Archimedean and Gaussian families of copulas.

2.2 Likelihood for Archimedean copulas

Consider a continuous, strictly decreasing function φ_θ which is completely monotonic and has the following properties (Nelsen, 2006):

$$\begin{cases} \varphi_\theta(0) = 1 \\ \varphi_\theta(\infty) = 0 \\ \varphi_\theta : [0, \infty[\to [0, 1]. \end{cases}$$

The function φ_θ is called the strict generator of an Archimedean copula with association parameter θ. Similarly, the inverse function φ_θ^{-1} is called the inverse generator. In practice, the generator can be obtained as a Laplace transform of a positive distribution function G_θ with $G_\theta(0) = 0$ (Nelsen, 2006)

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

Selecting the appropriate distribution G_θ, and therefore φ_θ, is important since different distributions will lead to a different tail dependency structure. Figure 2.1 shows how the association structure is realized for some popular Archimedean copulas. Using the generator φ_θ it is possible to write Equation (2.2) in a more convenient way

$$F(x_{i1},\dots,x_{in_i}|\boldsymbol{Z}_{i1},\dots,\boldsymbol{Z}_{in_i}) = \varphi_\theta\left(\varphi_\theta^{-1}\left(F_1(x_{i1}|\boldsymbol{Z}_{i1})\right) + \dots + \varphi_\theta^{-1}\left(F_{n_i}(x_{in_i}|\boldsymbol{Z}_{in_i})\right)\right). \tag{2.5}$$

Example: Consider the gamma distribution with parameters $k > 0$ and $\theta > 0$ represented by the density function

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

In the case where $k = 1/\theta$ we have a one-parameter gamma distribution with corresponding generator

$$\varphi_\theta(s) = (1+s\theta)^{1/\theta}.$$

The generator φ_θ corresponding to the one-parameter gamma distribution gives rise to the Clayton copula (Clayton, 1978), originally shown in the bivariate case,

$$\mathcal{C}(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}.$$

Note that this copula implies independence in the limiting case $\theta \to 0$.

Equation (2.5) is the general form of a n_i-dimensional Archimedean copula with association parameter θ. Using simple algebra, we will be able to derive a general likelihood that can be further developed into a concrete expression based on the choice of φ_θ.

Before deriving the likelihood for clusters of arbitrary size, let's first examine the simple yet instructive case of bivariate left-censored data. For cluster i, let F_{ij} and f_{ij} denote the marginal CDF and PDF of outcome j evaluated in x_{ij} given the set of covariates $\boldsymbol{Z}_{ij}$. Then, analogous to the case of right censoring in Andersen et al. (2005) and Prenen et al. (2017), the likelihood contribution of cluster i to the complete bivariate likelihood is given by

$$\begin{aligned} L_i =& \mathcal{C}\left(F_{i1},F_{i2}\right)^{(1-\delta_{i1})(1-\delta_{i2})} \\ & \left[\frac{\partial\mathcal{C}\left(F_{i1},F_{i2}\right)}{\partial F_{i1}}f_{i1}\right]^{\delta_{i1}(1-\delta_{i2})} \left[\frac{\partial\mathcal{C}\left(F_{i1},F_{i2}\right)}{\partial F_{i2}}f_{i2}\right]^{(1-\delta_{i1})\delta_{i2}} \\ & \left[\frac{\partial^2\mathcal{C}\left(F_{i1},F_{i2}\right)}{\partial F_{i1}\partial F_{i2}}f_{i1}f_{i2}\right]^{\delta_{i1}\delta_{i2}} \end{aligned} \tag{2.6}$$

or in terms of the Archimedean generator

$$\begin{aligned} L_i =& \varphi_\theta\left(\varphi_\theta^{-1}\left(F_{i1}\right) + \varphi_\theta^{-1}\left(F_{i2}\right)\right)^{(1-\delta_{i1})(1-\delta_{i2})} \\ & \left[\varphi_\theta'\left(\varphi_\theta^{-1}\left(F_{i1}\right) + \varphi_\theta^{-1}\left(F_{i2}\right)\right)(\varphi_\theta^{-1})'(F_{i1})f_{i1}\right]^{\delta_{i1}(1-\delta_{i2})} \\ & \left[\varphi_\theta'\left(\varphi_\theta^{-1}\left(F_{i1}\right) + \varphi_\theta^{-1}\left(F_{i2}\right)\right)(\varphi_\theta^{-1})'(F_{i2})f_{i2}\right]^{(1-\delta_{i1})\delta_{i2}} \\ & \left[\varphi_\theta^{(2)}\left(\varphi_\theta^{-1}\left(F_{i1}\right) + \varphi_\theta^{-1}\left(F_{i2}\right)\right)(\varphi_\theta^{-1})'(F_{i1})(\varphi_\theta^{-1})'(F_{i2})f_{i1}f_{i2}\right]^{\delta_{i1}\delta_{i2}}. \end{aligned} \tag{2.7}$$

The likelihood for cluster i in Equation (2.7) indeed has four terms corresponding to all possible combinations of the censoring scheme $\{\delta_{i1}, \delta_{i1}\}$ within the cluster. For a cluster of size n this leads to a likelihood contribution with 2^n terms which is infeasible analytically. Therefore, a parsimonious expression is required for easy maximization of the likelihood. Equation (2.7) generalizes quite easily since a recursive relationship based on d_i, the total number of observed events within cluster i, is present. Also note that for an Archimedean copula the existence of all derivatives is guaranteed since the generator is completely monotonic. Furthermore, the derivatives alternate in sign according to $(-1)^n \partial^{(n)} \varphi_\theta(t) \geq 0$. As a result it is easy to see that the likelihood contribution for a cluster i of size n_i can be written as

$$\begin{aligned} L_i &= \varphi_\theta^{(d_i)} \left(\varphi_\theta^{-1}(F_{i1}) + ... + \varphi_\theta^{-1}(F_{in_i}) \right) \prod_{j=1}^{n_i} \left[(\varphi_\theta^{-1})'(F_{ij}) f_{ij} \right]^{\delta_{ij}} \\ &= \varphi_\theta^{(d_i)} \left(\varphi_\theta^{-1}(F_{i1}) + ... + \varphi_\theta^{-1}(F_{in_i}) \right) \prod_{j=1}^{n_i} \left[\frac{f_{ij}}{\varphi'_\theta(\varphi_\theta^{-1}(F_{ij}))} \right]^{\delta_{ij}}. \end{aligned} \tag{2.8}$$

While this expression is easy to work with, the difficulty arises in the calculation of the higher order derivatives of φ_θ. Generally, these derivatives are not available directly and have to be calculated recursively. Hofert et al. (2012), for example, present efficient recursive techniques and implementations in the `R` programming language. As a result Equation (2.8) generally does not have a closed form and has to be approximated numerically.

Example (Cont'd): We again consider the Clayton generator and its inverse given by

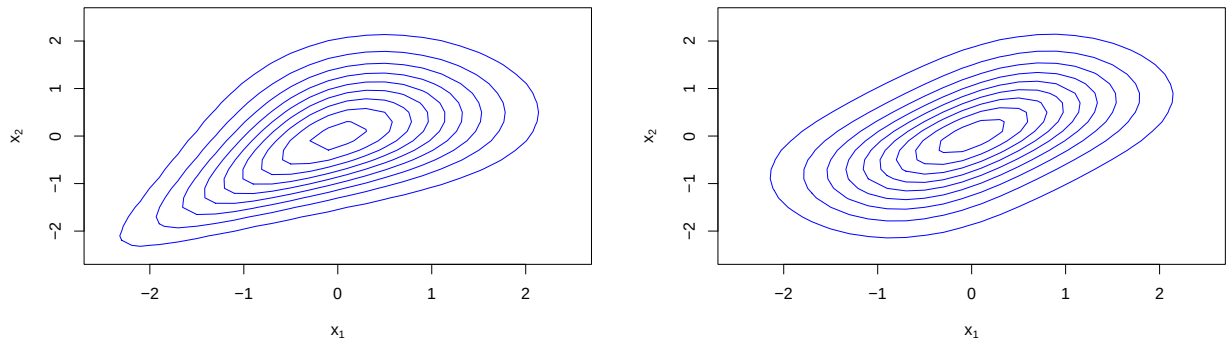
$$\begin{aligned} \varphi_\theta(s) &= (1 + s\theta)^{1/\theta} \\ \varphi_\theta^{-1}(s) &= \frac{1}{\theta}(s^{-\theta} - 1). \end{aligned}$$

This configuration is mathematically very convenient since the higher order derivatives $\varphi_\theta^{(d)}$ are analytically available by means of the Gamma function

$$\begin{aligned} \varphi_\theta^{(d)}(s) &= (-1)^d (1 + s\theta)^{-d-1/\theta} \theta^{d-1} \frac{\Gamma(d + 1/\theta)}{\Gamma(1 + 1/\theta)} \\ (\varphi_\theta^{-1})'(s) &= -s^{-\theta-1}. \end{aligned}$$

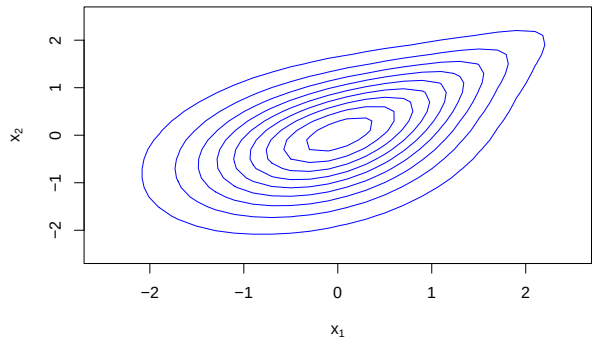
This leads to the simplified likelihood for the Clayton copula

$$L = \prod_{i=1}^{K} (-1)^{d_i} \left(F_{i1}^{-\theta} + ... + F_{in_i}^{-\theta} - n_i + 1 \right)^{-d_i - 1/\theta} \theta^{d_i - 1} \frac{\Gamma(d_i + 1/\theta)}{\Gamma(1 + 1/\theta)} \prod_{j=1}^{n_i} \left(-F_{ij}^{-\theta-1} f_{ij} \right)^{\delta_{ij}}.$$

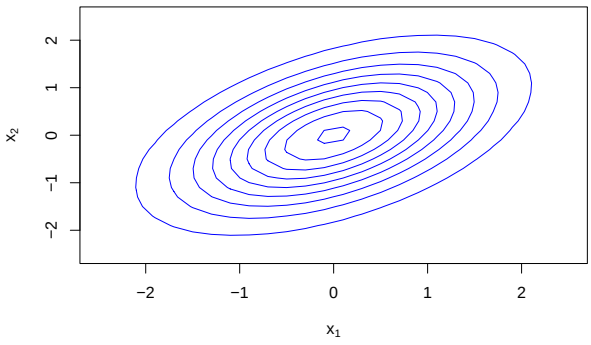


(a) Clayton

(b) Frank



(c) Gumbel-Hougaard



(d) Gaussian

Figure 2.1: Overview of the bivariate density functions of three copulas from the one-parameter Archimedean copula family and the Gaussian copula with standard normal margins (Joe (2007)). The association parameter was chosen to correspond with a Kendall's τ of 0.33. The tail dependence is clearly visible for the Clayton copula, where association is stronger in the lower tail (ie. larger survival times are correlated more strongly), while the Gumbel-Hougaard copula exhibits strong upper tail dependence (ie. smaller survival times are correlated more strongly). Both the Frank and Gaussian copula do not exhibit tail dependence.

2.3 Likelihood for Gaussian copulas

Another popular choice and alternative to Archimedean copulas are the copulas within the elliptical family. In this text we will focus in particular on the Gaussian copula. Figure 2.1 shows a distribution function from a Gaussian copula noting that, unlike some copulas from the Archimedean family, tail independence is implied. The n-dimensional Gaussian copula with parameter ρ is defined as follows (Joe, 2015; Tjøstheim et al., 2022):

$$\mathcal{C}(u_1, \dots, u_n) = \mathbf{\Phi}_{\boldsymbol{\rho}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

with $\mathbf{\Phi}_{\boldsymbol{\rho}}$ the CDF of a n-variate normal distribution with mean $\mathbf{0}$ and (exchangeable) correlation matrix $\mathbf{\Sigma}$, 1 on the diagonal and ρ elsewhere, and Φ^{-1} the quantile function of a univariate standard normal distribution. In the context of survival data with left censoring, we now rewrite Equation (2.2) to reflect a Gaussian correlation structure

$$F(x_{i1}, \dots, x_{in_i} | \boldsymbol{Z}_{i1}, \dots, \boldsymbol{Z}_{in_i}) = \mathbf{\Phi}_{\boldsymbol{\rho}} \left(\Phi^{-1}(F_1(x_{i1} | \boldsymbol{Z}_{i1})), \dots, \Phi^{-1}(F_{n_i}(x_{in_i} | \boldsymbol{Z}_{in_i})) \right). \tag{2.9}$$

We now derive an expression of the likelihood for left-censored data with clusters of varying size and covariate structure according to the methodology of Jacqmin-Gadda (2000) and Othus & Li (2010). Let X_i^o denote the d_i-vector of observed outcomes in cluster i. Similarly, let X_i^c denote the $(n_i - d_i)$-vector of censored outcomes in cluster i. We now reorder the observations within each cluster in such a way that $X_i = (X_i^o, X_i^c)$. This allows a partitioning of the covariance matrix Σ_i of the transformed observations $\Phi^{-1}(F_i(x_i))$

$$\Sigma_i = \left[\begin{array}{c|c} \Sigma_{i,o} & \Sigma_{i,oc} \\ \hline \Sigma_{i,co} & \Sigma_{i,c} \end{array} \right]$$

where $\Sigma_{i,o}$ and $\Sigma_{i,c}$ are the covariance matrices of the transformed observed and censored outcomes with dimensions $d_i \times d_i$ and $(n_i - d_i) \times (n_i - d_i)$ respectively. Consider now the likelihood contribution of the observed outcomes to cluster i

$$\begin{aligned} L_i^o &= \frac{\partial^{(d_i)} F(x_{i1}, \dots, x_{id_i})}{\partial X_i^o} \\ &= \frac{\partial^{(d_i)} \mathbf{\Phi}_{\boldsymbol{\rho}}(\Phi^{-1}(F_{i1}), \dots, \Phi^{-1}(F_{id_i}))}{\partial X_i^o} \\ &= \mathbf{\Psi}^{(\boldsymbol{d_i})} \left(\Phi^{-1}(F_{i1}), \dots, \Phi^{-1}(F_{id_i}) \right) \prod_{j=1}^{d_i} \frac{\partial \Phi^{-1}}{\partial F_{ij}} \frac{\partial F_{ij}}{\partial x_{ij}} \\ &= \mathbf{\Psi}^{(\boldsymbol{d_i})} \left(\Phi^{-1}(F_{i1}), \dots, \Phi^{-1}(F_{id_i}) \right) \prod_{j=1}^{d_i} \frac{f_{ij}}{\Psi\left(\Phi^{-1}(F_{ij})\right)} \end{aligned}$$

where $\mathbf{\Psi}^{(\boldsymbol{d_i})}$ is the d_i-variate normal distribution function with mean $\mathbf{0}$ and covariance matrix $\Sigma_{i,o}$, and Ψ the standard normal distribution function. Due to the ordering of the events, the contribution to the likelihood of the censored observation is conditional on the observed observations

$$
\begin{aligned}
L_i^c &= F(x_{i(d_i+1)}, \dots, x_{in_i} | X_i^o) \\
&= \boldsymbol{\Phi}_{\boldsymbol{c|o}}^{(\boldsymbol{n_i}-\boldsymbol{d_i})} \left(\Phi^{-1}(F_{i(d_i+1)}), \dots, \Phi^{-1}(F_{in_i}) | \Phi^{-1}(F_i(x_{ij}^o)) \right)
\end{aligned}
$$

where $\Phi_{c|o}^{(n_i-d_i)}(.|.)$ is the $(n_i - d_i)$-variate CDF of a conditional normal distribution with mean $\Sigma_{i,co}\Sigma_{i,o}^{-1}\Phi^{-1}(F_i(X_i^o))$ and covariance matrix $\Sigma_{i,c} - \Sigma_{i,co}\Sigma_{i,o}^{-1}\Sigma_{i,oc}$. We are now in a position to combine the contributions of both observed and unobserved outcomes to each cluster and give an expression for the full likelihood

$$
\begin{aligned}
L = \prod_{i=1}^{K} & \boldsymbol{\Phi}_{\boldsymbol{c|o}}^{(\boldsymbol{n_i}-\boldsymbol{d_i})} \left(\Phi^{-1}(F_{i(d_i+1)}), \dots, \Phi^{-1}(F_{in_i}) | \Phi^{-1}(F_i(x_{ij}^o)) \right) \\
& \boldsymbol{\Psi}^{(\boldsymbol{d_i})} \left(\Phi^{-1}(F_{i1}), \dots, \Phi^{-1}(F_{id_i}) \right) \prod_{j=1}^{d_i} \frac{f_{ij}}{\Psi\left(\Phi^{-1}(F_{ij})\right)}.
\end{aligned}
\tag{2.10}
$$

A practical application of Equation (2.10) for bivariate data is detailed in Appendix B. Although there we derive the likelihood term by term following the censoring scheme, it is straightforward to prove that this is equivalent to Equation (2.10).

2.4 Estimation procedures

In this part we introduce two methods to estimate the association and marginal parameters in the copula model for left-censored clustered data. Most importantly we focus on full parametric specification of the marginal baseline survival function which allows for estimation within, or close to, maximum likelihood theory. Firstly, we consider multi-parameter maximum likelihood estimation where the full set of association and marginal parameters are estimated simultaneously. While this method is well-known and straightforward to use, the complexity of the copula likelihood and the number of parameters to estimate might lead to computational difficulties or otherwise. For those situations, a two-stage method where the association parameter is estimated under a working independence assumption is proposed.

2.4.1 One-stage estimation

The following theory is based on the book by Bijma et al. (2017). Let θ denote the association parameter of copula $\mathcal{C}$ and $\boldsymbol{\beta}$ the vector of parameters attributed to the marginal distributions and covariates included in the model. The maximum likelihood estimate $(\widehat{\theta}, \widehat{\boldsymbol{\beta}})$ is obtained by maximizing the likelihood given by Equations (2.3)-(2.4)

$$
(\widehat{\theta}, \widehat{\boldsymbol{\beta}}) = \arg\max_{(\theta, \boldsymbol{\beta})} L(\theta, \boldsymbol{\beta})
$$

which implies solving simultaneously the score functions

$$\begin{cases} U_\theta(\theta, \boldsymbol{\beta}) = \frac{\partial \log L(\theta, \boldsymbol{\beta})}{\partial \theta} = 0 \\ \boldsymbol{U}_{\boldsymbol{\beta}}(\theta, \boldsymbol{\beta}) = \frac{\partial \log L(\theta, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0. \end{cases}$$

Under weak regularity conditions (Lehmann & Casella, 1998) and as $K \to \infty$, the following asymptotic result holds

$$\sqrt{K}(\widehat{\theta} - \theta, \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = MVN\left(\mathbf{0}, I = \left[-\frac{\partial^2 \log L(\theta, \boldsymbol{\beta})}{\partial \zeta_i \partial \zeta_j}\right]\right)$$

where $\zeta = (\theta, \boldsymbol{\beta})$.

In practical applications the maximum likelihood estimate $(\widehat{\theta}, \widehat{\boldsymbol{\beta}})$ can be estimated using numerical omptimization techniques (for example `optim` in `R` or `NLMIXED` in SAS). An estimate of the Fisher information I can then be directly obtained by inverting the Hessian matrix with elements $\frac{\partial^2 \log L(\widehat{\theta}, \widehat{\boldsymbol{\beta}})}{\partial \zeta_i \partial \zeta_j}$.

2.4.2 Two-stage estimation

When direct maximization of the multi-parameter likelihood is not feasible another possibility, suggested by the structure of marginal models like the copula model, is estimating the association parameter θ separately from the marginal parameters $\boldsymbol{\beta}$. Hougaard (1989) is an early example where in a bivariate setting the marginal parameters are estimated nonparametrically under the working assumption of independence and then used as plug-in estimates to estimate θ, much like the approach taken by Liang & Zeger (1986). Hougaard, however, did not investigate the inferential properties of such estimators. When the interest lies on the estimation of θ, other approaches suggest treating the baseline hazard in the Cox model (Cox, 1972) as a nuisance parameter (see for example Clayton (1978); Oakes (1982); Oakes (1986)). A general two-stage semi-parametric estimation procedure was outlined by Shih & Louis (1995) for bivariate data without covariates and by Spiekerman & Lin (1998) for clusters with fixed size, covariate structure and possibly multiple types of failure time. While a semi-parametric approach allows for flexible estimation in the margins, straightforward theory on inference is not a given in the case of left censoring since we rely on counting processes and Martingale theory. Furthermore, semi-parametric estimation is limited to fitting proportional hazards models in the margins which might not be an appropriate choice anyway. Therefore, we focus on a fully parametric model based on Shih & Louis (1995) and Spiekerman & Lin (1998) where we estimate in stage one the marginal parameters under the assumption that all observations are independent which can then be used in stage two as plug-in estimates for the estimation of the association parameter.

First we consider the case where the cluster size n is fixed for all clusters and each outcome j ($j = 1, \ldots, n$) has a fixed covariate structure independent from other outcomes $j' \neq j$ across all clusters. Let F_j denote the marginal CDF for all outcomes with position j in each cluster. We assume that the functional form of each F_j is known and has a finite number of parameters. Denote $\boldsymbol{\beta_j}$ the vector of marginal parameters related to the functional form of margin j and its covariate structure $\boldsymbol{Z}_{\cdot \boldsymbol{j}}$. Now consider the score function of margin j

$$\boldsymbol{U}^*_{\boldsymbol{\beta_j}} = \sum_{i=1}^{K} \delta_{ij} \frac{\partial \log f_j(x_{ij}|\boldsymbol{\beta_j})}{\partial \boldsymbol{\beta_j}} + (1 - \delta_{ij}) \frac{\partial \log F_j(x_{ij}|\boldsymbol{\beta_j})}{\partial \boldsymbol{\beta_j}}$$

where the summation over all clusters is possible due to the assumed between-cluster independence. In the first stage we estimate $\boldsymbol{\beta_j}$ for all j by solving $\boldsymbol{U}^*_{\boldsymbol{\beta}_j}$ separately

$$\boldsymbol{U}^*_{\boldsymbol{\beta}_j}(\overline{\boldsymbol{\beta}_j}) = 0$$

where $\overline{\boldsymbol{\beta}_j}$ is the maximum likelihood estimate of the marginal parameters of margin j under the working independence assumption. In the second stage we make use of Equations (2.3) and (2.4) to solve the score function for θ U_θ given $\overline{\boldsymbol{\beta}_j}$

$$U_\theta(\theta) = \frac{\partial \log L(\theta, \overline{\boldsymbol{\beta}_1}, \ldots, \overline{\boldsymbol{\beta}_j})}{\partial \theta} = 0. \tag{2.11}$$

Theorem 2.1. *Let $\overline{\theta}$ be the solution of $U_\theta(\theta) = 0$ and θ_0 the true value. Under regularity conditions (Cox and Hinkley (1979), p.182) and $K \to \infty$ we have*

$$\sqrt{K}(\overline{\theta} - \theta_0) \to N\left(0, \frac{1}{I_{\theta\theta}} + \frac{\boldsymbol{I}^*_{\theta\boldsymbol{\beta}}(\boldsymbol{I}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1}\boldsymbol{V}(\boldsymbol{I}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1}}{I^2_{\theta\theta}}\right)$$

We refer to Appendix C for clarification on the notation used in the theorem and its proof. Note the resemblance to the variance-covariance matrix used in Generalized Estimating Equations (Liang & Zeger, 1986) and the inclusion of the typical Huber-like covariance structure. Although it has been shown that an exact expression for the association parameter can be obtained, it becomes clear very quickly that in practical situations the use of such an expression is very cumbersome: The off-diagonal elements of $\boldsymbol{V}$, which denote the covariance between the parameters of the different margins, can only be obtained directly in a limited number of scenarios (see for example Prentice & Cai (1992)) and require numerical optimization techniques. Prenen et al. (2017) offer a valid alternative by considering all observations as independent in the first stage and thereby collapsing the data structure from multiple margins to only one margin to estimate. This approach automatically allows for clusters of varying size and covariate structure. In any case, implementation of the exact expression remains challenging. Resampling techniques have been developed (Lipsitz et al., 1994; Lipsitz & Parzen, 1996) and shown to approximate the exact variance of the association parameter very closely (Othus & Li, 2010; Prenen et al., 2017).

2.5 Flexible parametric modeling of the baseline hazard

Until now we have assumed a strong parametric form for the margins. In practice this often translates to choosing a parametric shape for the baseline hazard in an AFT model or for the baseline hazard in a Cox proportional hazards model (Cox, 1972). While this is a convenient choice in the sense that we stay well within the real of maximum likelihood theory, misspecification becomes a serious concern. Ha & Lee (2003) even state that a fully parametric model might be too strong an assumption. Other studies provide evidence that misspecification of the marginal baseline hazard negatively affects the estimation of the association parameter (Genest et al., 1995; Kim et al., 2007). Furthermore, although semiparametric estimation of the association parameter is possible in some contexts, Prenen et al. (2017) have shown that a nonparametric Breslow-type estimator for the baseline hazard in a one-stage estimation procedure is not tractable. Therefore a flexible approach to modeling the margins might offer a reliable alternative to strong parametric specification.

Recent developments by He & Lawless (2003) and Kwon et al. (2022) have seen the introduction of splines in multivariate survival data with correlated outcomes. He & Lawless (2003) introduced splines in a bivariate Cox model with Clayton copula function while Kwon et al. (2022) recently extended the work of Prenen et al. (2017) to account for splines in clustered data with varying cluster size. Both methodologies use M-spline basis functions (Ramsay, 1988) to model the baseline hazard in accordance with the existing techniques on spline modeling in the presence of dependent censoring (see for example Emura & Chen (2018) or Emura et al. (2019)). The main argument for using M-splines is the mathematical convenience that the spline always guarantees the monotonicity of the baseline hazard. On the other hand, choosing starting values for the knots is not intuitive at all and the cross-validation procedure outlined by Emura et al. (2019) is elaborate and computationally intensive. Therefore we propose a fast-fitting, flexible and intuitive spline model based on the transformation model by Royston & Parmar (2002) to be used in maximum likelihood estimation of the association and marginal parameters.

Consider a Cox model (Cox, 1972) for the margins with the usual proportional hazards assumption. Then the marginal cumulative hazard corresponding to outcome x_{ij} in cluster i is given by

$$H_{ij}(x_{ij}) = H_0(x_{ij})e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}}$$

where H_0 is the baseline cumulative hazard function. Note that, without loss of generality, H_0 is assumed to be equal for all observations regardless of cluster or position within the cluster as this allows for clusters of varying size and unequal covariate structures. In general, however, H_0 could also be estimated within each margin (denoted H_{0j}). Rather than modeling H_0 directly with splines or otherwise, consider now a log-transformation of the cumulative hazard

$$\log H_{ij}(x_{ij}) = \log H_0(x_{ij}) + \boldsymbol{\beta}^T \boldsymbol{Z}_{\boldsymbol{ij}}.$$

Since $\log H_0$ and the covariate structure are separated, unlike in the AFT framework, it is possible to apply smoothing techniques to $\log H_0$. Royston & Parmar (2002) opt for natural cubic splines with linearity constraints beyond the boundary knots as monotonic splines "make[s] the computational problem much more difficult and awkward, a cost we do not feel is in general justified". On the other hand, the authors claim that a sizable number of uncensored observations (as few as 50) in the data automatically imposes monotonicity of the fitted spline. While verifying this claim was not the purpose of this text, simulations in Section 3 do seem to indicate that a satisfactory fit of the marginal distribution is possible even in small data sets with heavy censoring. Now consider the natural cubic spline $s(\log x; \gamma)$ with boundary knots $k_{\min}$ and $k_{\max}$, m unique internal knots ($k_{\min} <$) $\quad k_1 < ... < k_m \quad (< k_{\max})$ and m basis function v_j

$$s(y; \gamma) = \gamma_0 + \gamma_1 y + \gamma_2 v_1(y) + ... + \gamma_{m+1} v_j(y) \tag{2.12}$$

where $y = \log x$ is the transformed outcome. The basis functions v_j are defined as follows

$$v_j(y) = \max(0, y - k_j)^3 - \lambda_j \max(0, y - k_{\min})^3 - (1 - \lambda_j) \max(0, y - k_{\max})^3$$

and

$$\gamma_j = \frac{k_{\max} - k_j}{k_{\max} - k_{\min}}.$$

Rather than specifying a parametric form for $\log H_0$ we are now in a position to use Equation (2.12) instead. The proportional hazards model then becomes

$$\begin{aligned}\eta = \log H_{ij}(x_{ij}) &= \log H_0(x_{ij}) + \boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{Z_{ij}} \\ &= s(y_{ij}; \gamma) + \boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{Z_{ij}}.\end{aligned} \tag{2.13}$$

A rather interesting feature of this transformation model is the transposition into a Weibull Cox model when $v_j(y) = 0$, a spline model without internal knots. The marginal distribution functions f and F are easily obtained by differentiation

$$\begin{aligned}F(x) &= 1 - S(x) \\ &= 1 - e^{-H(x)} \\ &= 1 - \exp(-e^{\eta})\end{aligned}$$

$$\begin{aligned}f(x) &= \frac{\mathrm{d}F(x)}{\mathrm{d}x} \\ &= \exp(-e^{\eta}) e^{\eta} \frac{\mathrm{d}\eta}{\mathrm{d}y} \frac{\mathrm{d}y}{\mathrm{d}x} \\ &= \frac{1}{x} \frac{\mathrm{d}s(y;\gamma)}{\mathrm{d}y} \exp(\eta - e^{\eta})\end{aligned}$$

and generally can be used in a marginal model for clustered data as obtained in Equation (2.3). In practical applications it is still necessary to find a suitable number of internal knots m, their positions and starting values for $\gamma_0, \ldots, \gamma_{m-1}$ and $\boldsymbol{\beta}$ due to the complexity of the likelihood of a copula model. Royston & Parmar (2002) recommend a maximum of 3 internal knots to preserve sufficient stability of the fitted curve. Furthermore, their position can be based on percentiles of the distribution of the uncensored observations as described in Royston & Parmar (2002) since optimizing the knots only results in marginal gains in the precision of the fit (Durrleman & Simon, 1989). In the same way, the boundary knots $k_{\min}$ and $k_{\max}$ are placed at the extremes of the uncensored observations. Starting values for γ can be obtained using regression techniques on the uncensored observations: firstly, we obtain an estimate of $\log H$ using a standard semiparametric Cox model. Having obtained an estimate for the transformed cumulative hazard, Equation (2.13) reduces to a linear regression problem with covariates v_j and $\boldsymbol{Z}$. Starting values are then obtained as the least squares estimates of the linear regression model. On a final note, it is important to mention that Equation (2.12) can be adjusted to accommodate non-proportional hazards as an alternative to for example time-varying covariates.

3 Simulation study

In this section we present an extensive simulation study to examine and verify the properties of the various estimators for the association and regression parameters. Simulations will be carried out in a bivariate setting with complete clusters and covariate structure under random left censoring of varying degree. Firstly, we consider a correct parametric specification of the copula and baseline survival function which gives insight into the bias and coverage probability for a large, but more importantly also small number of clusters. Since the one-stage and two-stage estimators are based on strong parametric model assumptions, the assessment of model robustness is crucial. Therefore, we conduct further simulations by introducing misspecification of the copula and later also misspecification of the baseline survival with various degrees of deviation from the correct specification. This study also allows a close look into whether a weak parametric estimator such as the spline model by Royston & Parmar (2002) is a viable alternative for the estimation of the within-cluster association. Since inference on the association parameter is of primary interest we include these results in this section while supplementary tables on the regression parameters are included in Appendix D.

3.1 Correct specification

We consider the case where the copula function and the baseline survival functions for both responses are correctly specified. 1000 samples are simulated for each value of the association parameter θ, with θ chosen such that Kendall's τ is comparable across different copula functions. Samples were generated using the conditional approach (Genest & Favre, 2007) and transformed to allow for a covariate structure in the AFT framework (Leemis et al., 1990). We assume a Weibull marginal survival function $S(t) = \exp(\lambda t^{\rho} \exp(\beta Z))$ with scale $\lambda = 1$ and shape $\rho = 0.5$ (see Figure 3.1) for both responses. This choice allows for comparisons to the Cox PH spline model since a Weibull AFT is equivalent to a Cox PH model with Weibull baseline hazard. Important to note is the fact that for the spline model we assume equal marginal distributions and therefore have a reduced dimensionality in the estimation process. Furthermore, a covariate structure for each response is imposed by sampling separately for each response from a standard uniform distribution and taking $\beta_1 = \beta_2 = 1$. Censoring times were generated from an exponential distribution where the scale parameter was empirically chosen to reflect the desired percentage of censoring. Starting values for the baseline spline are obtained as described in Section 2.5. Note that the number of samples for the simulations on the Gaussian copula were reduced to 500 due to the computationally intensive nature of the estimation process. Results on the association parameter are given in increasing order of θ.

Tables 3.1-3.3 show the estimated association parameter for various copulas and number of clusters using the three discussed estimation procedures. Firstly we discuss the results for the Archimedean family of copulas: in general, the estimated association parameter tends to be biased upwards. The bias is also consistently larger in magnitude when the association is relatively small and the number of clusters is low. In some cases the relative bias can reach values upwards of 30%. A notable exception is the Gumbel copula where even a low number of clusters and high censoring only leads to a maximum relative bias of less than 10% over all simulations. The two one-stage procedures markedly underperform in terms of the bias of the association parameter compared to the two-stage procedure. Since the spline model reduces to the strong parametric model with a Weibull baseline survival when $v_1(y) = 0$, it is expected that the two one-stage procedures do not produce significantly different results. On the other hand, the spline procedure does seem to reduce the bias compared to the strong parametric model. The coverage probability of the estimators is well below the nominal coverage probability of 95% in

the case of a small number of clusters and a weak association structure. Increasing the number of clusters significantly improves the coverage probability such that it stays within the acceptable range of 93 to 96%. The estimated regression parameters in Tables D.1 - D.3 are generally bias upwards, yet fairly precise in terms of bias relative to the number of clusters. The two-stage procedure produces underestimated standard errors as expected since the regression parameters are estimated under the working independence assumption. This is also reflected in the coverage probability of the two-stage estimator. Generally, both the standard and spline one-stage procedures produce acceptable coverage probabilities even in the case of a small sample size.

The Gaussian copula generally performs very well even with small sample sizes and weak association structures. the two-stage procedure, however, seems less suitable since it significantly underperforms in terms of bias compared to the one-stage procedures. Specifically for the one-stage procedures, even though the relative bias for 50 clusters is already small, the estimator for the association parameter becomes practically unbiased for a higher number of clusters regardless of the percentage of censoring. The coverage probability hovers around the nominal 95% in all other cases. Similar conclusions can be drawn for the regression parameters. The spline model also performs well but tends to have a slightly lower coverage probability overall.

As the two-stage procedure leads to inconsistent estimates of the standard errors of the regression parameters and only leads to significant improvements in the estimation of the association parameter in the case of small sample sizes, we conclude that a one-stage procedure specified in a strong or weak parametric way is overall preferable.

Table 3.1: Results for the association parameter from a simulation study with 50 clusters. The estimated value of the association parameter is given, together with the estimated standard error and coverage probability in brackets.

	θ	30% censoring			50% censoring		
		Maximum Likelihood	Two-stage	Splines (Royston & Parmar)	Maximum Likelihood	Two-stage	Splines (Royston & Parmar)
Clayton	0.2	0.272 (0.260; 82.0%)	0.252 (0.275; 89.9%)	0.284 (0.287; 84.6%)	0.387 (0.403; 78.4%)	0.319 (0.388; 84.1%)	0.355 (0.427; 77.2%)
	0.5	0.582 (0.361; 92.8%)	0.530 (0.352; 96.3%)	0.552 (0.368; 95.7%)	0.662 (0.538; 88.6%)	0.582 (0.488; 94.0%)	0.588 (0.544; 91.0%)
	1	1.118 (0.494; 94.1%)	1.028 (0.470; 94.4%)	1.112 (0.501; 93.8%)	1.239 (0.747; 93.6%)	1.083 (0.649; 93.9%)	1.173 (0.734; 95.4%)
	1.5	1.674 (0.630; 94.3%)	1.512 (0.578; 94.2%)	1.644 (0.627; 93.0%)	1.793 (0.921; 93.6%)	1.653 (0.812; 94.3%)	1.769 (0.921; 95.6%)
	2	2.233 (0.764; 95.0%)	2.036 (0.696; 94.0%)	2.137 (0.746; 92.9%)	2.332 (1.101; 93.2%)	2.124 (0.954; 93.0%)	2.356 (1.110; 94.0%)
Frank	0.82	0.763 (0.971; 94.8%)	0.823 (0.962; 94.1%)	0.782 (0.969; 94.5%)	0.870 (1.288; 95.5%)	0.847 (1.255; 97.3%)	0.831 (1.277; 96.2%)
	1.86	1.979 (1.022; 94.2%)	1.801 (0.998; 95.6%)	1.919 (1.017; 94.6%)	2.039 (1.358; 95.0%)	1.817 (1.306; 96.5%)	1.997 (1.343; 95.5%)
	3.26	3.456 (1.137; 94.1%)	3.237 (1.103; 94.0%)	3.427 (1.137; 95.5%)	3.622 (1.563; 93.3%)	3.240(1.453; 95.7%)	3.465 (1.531; 95.1%)
	4.6	4.908 (1.300; 95.0%)	4.519 (1.232; 94.4%)	4.770 (1.283; 95.6%)	5.204 (1.839; 95.3%)	4.687 (1.692; 95.3%)	5.033 (1.805; 95.7%)
	5.8	6.084 (1.457; 94.4%)	5.729 (1.386; 95.5%)	6.103 (1.460; 94.4%)	6.637 (2.156; 96.0%)	5.818 (1.932; 95.5%)	6.292 (2.033; 94.4%)
Gumbel	1.1	1.109 (0.124; 89.3%)	1.102 (0.120; 93.7%)	1.103 (0.123; 91.6%)	1.106 (0.146; 91.2%)	1.098 (0.139; 91.9%)	1.099 (0.144; 89.6%)
	1.25	1.281 (0.160; 92.9%)	1.251 (0.151; 92.7%)	1.267 (0.157; 92.2%)	1.296 (0.195; 92.1%)	1.255 (0.175; 92.0%)	1.302 (0.195; 90.6%)
	1.49	1.544 (0.221; 94.3%)	1.492 (0.205; 91.8%)	1.529 (0.218; 93.7%)	1.570 (0.277; 92.6%)	1.491 (0.242; 91.5%)	1.539 (0.266; 91.4%)
	1.75	1.831 (0.286; 93.7%)	1.731 (0.258; 90.7%)	1.790 (0.278; 94.0%)	1.891 (0.372; 94.8%)	1.735 (0.310; 92.1%)	1.851 (0.356; 94.0%)
	2	2.092 (0.344; 93.4%)	1.963 (0.308; 91.9%)	2.049 (0.337; 94.6%)	2.195 (0.462; 96.0%)	1.957 (0.377; 90.6%)	2.143 (0.445; 94.5%)
Gaussian	0.15	0.144 (0.158; 91.7%)	0.147 (0.157; 93.3%)	0.152 (0.159; 92.1%)	0.143 (0.194; 89.7%)	0.118 (0.195; 93.0%)	0.123 (0.196; 90.5%)
	0.3	0.302 (0.148; 92.6%)	0.281 (0.148; 93.9%)	0.299 (0.149; 93.6%)	0.294 (0.182; 89.2%)	0.286 (0.181; 90.3%)	0.310 (0.179; 90.7%)
	0.5	0.501 (0.123; 91.8%)	0.482 (0.125; 93.1%)	0.494 (0.125; 90.3%)	0.504 (0.153; 88.7%)	0.464 (0.156; 92.8%)	0.499 (0.154; 88.8%)
	0.65	0.655 (0.096; 90.0%)	0.625 (0.099; 93.6%)	0.641 (0.099; 94.0%)	0.663 (0.118; 86.8%)	0.626 (0.124; 92.0%)	0.638 (0.124; 90.7%)
	0.71	0.715 (0.083; 90.6%)	0.687 (0.087; 93.8%)	0.706 (0.085; 92.8%)	0.727 (0.100; 85.6%)	0.685 (0.109; 92.5%)	0.711 (0.105; 85.7%)

Table 3.2: Results for the association parameter from a simulation study with 200 clusters. The estimated value of the association parameter is given, together with the estimated standard error and coverage probability in brackets.

	θ	30% censoring			50% censoring		
		Maximum Likelihood	Two-stage	Splines (Royston & Parmar)	Maximum Likelihood	Two-stage	Splines (Royston & Parmar)
Clayton	0.2	0.211 (0.135; 93.0%)	0.207 (0.135; 95.5%)	0.219 (0.138; 95.8%)	0.241 (0.195; 88.3%)	0.228 (0.189; 94.1%)	0.232 (0.201; 91.2%)
	0.5	0.512 (0.173; 94.7%)	0.506 (0.173; 96.4%)	0.508 (0.174; 93.9%)	0.532 (0.253; 95.1%)	0.512 (0.236; 96.2%)	0.534 (0.255; 95.4%)
	1	1.029 (0.232; 94.2%)	1.008 (0.227; 94.8%)	1.022 (0.234; 95.0%)	1.035 (0.327; 94.3%)	1.032 (0.325; 96.0%)	1.024 (0.308; 94.4%)
	1.5	1.521 (0.288; 95.0%)	1.508 (0.280; 95.2%)	1.524 (0.294; 93.5%)	1.565 (0.406; 94.9%)	1.512 (0.374; 93.4%)	1.555 (0.407; 95.3%)
	2	2.023 (0.346; 94.5%)	1.998 (0.334; 93.8%)	2.049 (0.357; 94.1%)	2.087 (0.482; 95.4%)	2.031 (0.447; 95.6%)	2.055 (0.481; 93.8%)
Frank	0.82	0.832 (0.471; 95.3%)	0.803 (0.471; 94.8%)	0.839 (0.472; 94.3%)	0.813 (0.599; 94.7%)	0.810 (0.599; 95.6%)	0.817 (0.601; 95.5%)
	1.86	1.893 (0.492; 95.5%)	1.885 (0.491; 95.8%)	1.898 (0.493; 94.0%)	1.873 (0.630; 95.2%)	1.864 (0.627; 95.3%)	1.917 (0.630; 95.5%)
	3.26	3.294 (0.541; 96.7%)	3.256 (0.540; 94.7%)	3.282 (0.542; 95.2%)	3.322 (0.703; 93.8%)	3.271 (0.696; 95.1%)	3.277 (0.701; 94.7%)
	4.6	4.677 (0.612; 94.8%)	4.574 (0.604; 96.3%)	4.674 (0.617; 95.1%)	4.706 (0.810; 94.9%)	4.550 (0.792; 94.8%)	4.653 (0.807; 94.3%)
	5.8	5.880 (0.686; 95.3%)	5.780 (0.676; 95.2%)	5.862 (0.692; 94.1%)	5.923 (0.925; 95.4%)	5.828 (0.912; 95.6%)	5.908 (0.927; 95.7%)
Gumbel	1.1	1.105 (0.056; 94.0%)	1.103 (0.056; 94.1%)	1.105 (0.056; 94.1%)	1.104 (0.062; 93.6%)	1.101 (0.062; 93.2%)	1.101 (0.062; 92.8%)
	1.25	1.257 (0.076; 94.8%)	1.248 (0.074; 94.0%)	1.255 (0.076; 94.2%)	1.260 (0.087; 95.1%)	1.256 (0.086; 93.8%)	1.257 (0.087; 94.2%)
	1.49	1.501 (0.105; 95.0%)	1.490 (0.102; 93.9%)	1.496 (0.104; 92.7%)	1.506 (0.125; 94.1%)	1.485 (0.119; 92.9%)	1.501 (0.125; 95.0%)
	1.75	1.768 (0.135; 95.3%)	1.752 (0.129; 92.0%)	1.766 (0.135; 95.6%)	1.734 (0.158; 94.3%)	1.738 (0.153; 92.5%)	1.768 (0.163; 94.8%)
	2	2.030 (0.165; 94.7%)	1.979 (0.153; 91.4%)	2.008 (0.163; 95.1%)	2.030 (0.201; 95.1%)	1.981 (0.187; 92.1%)	2.040 (0.204; 95.0%)
Gaussian	0.15	0.150 (0.079; 95.3%)	0.143 (0.079; 94.6%)	0.147 (0.079; 93.6%)	0.149 (0.097; 95.6%)	0.144 (0.097; 93.0%)	0.142 (0.194; 90.3%)
	0.3	0.300 (0.074; 94.1%)	0.295 (0.074; 95.1%)	0.300 (0.074; 95.4%)	0.299 (0.091; 93.1%)	0.300 (0.090; 95.9%)	0.282 (0.184; 89.7%)
	0.5	0.502 (0.062; 94.2%)	0.495 (0.062; 93.6%)	0.500 (0.062; 94.2%)	0.506 (0.075; 94.1%)	0.491 (0.076; 94.0%)	0.489 (0.155; 91.5%)
	0.65	0.652 (0.048; 94.9%)	0.644 (0.048; 93.5%)	0.648 (0.049; 93.2%)	0.652 (0.060; 92.8%)	0.643 (0.060; 94.4%)	0.657 (0.119; 89.1%)
	0.71	0.710 (0.042; 94.3%)	0.704 (0.041; 95.4%)	0.706 (0.043; 95.0%)	0.711 (0.052; 91.9%)	0.702 (0.052; 94.2%)	0.701 (0.108; 89.5%)

Table 3.3: Results for the association parameter from a simulation study with 500 clusters. The estimated value of the association parameter is given, together with the estimated standard error and coverage probability in brackets.

	θ	30% censoring			50% censoring		
		Maximum Likelihood	Two-stage	Splines (Royston & Parmar)	Maximum Likelihood	Two-stage	Splines (Royston & Parmar)
Clayton	0.2	0.202 (0.087; 95.4%)	0.202 (0.087; 95.8%)	0.201 (0.087; 95.7%)	0.210 (0.125; 93.4%)	0.209 (0.119; 95.6%)	0.213 (0.128; 96.4%)
	0.5	0.503 (0.108; 94.2%)	0.502 (0.108; 95.1%)	0.494 (0.108; 95.0%)	0.509 (0.157; 94.7%)	0.500 (0.147; 95.3%)	0.504 (0.157; 94.8%)
	1	1.012 (0.145; 94.8%)	1.010 (0.143; 94.9%)	1.004 (0.145; 94.0%)	1.017 (0.203; 94.4%)	1.006 (0.191; 95.7%)	1.012 (0.205; 94.8%)
	1.5	1.521 (0.181; 94.9%)	1.507 (0.177; 94.6%)	1.510 (0.184; 94.0%)	1.535 (0.251; 94.2%)	1.517 (0.235; 95.6%)	1.514 (0.252; 93.8%)
	2	2.007 (0.216; 95.5%)	1.998 (0.21; 94.8%)	2.011 (0.221; 95.3%)	2.033 (0.297; 95.3%)	2.005 (0.277; 94.7%)	2.022 (0.301; 94.1%)
Frank	0.82	0.829 (0.297; 94.3%)	0.815 (0.297; 93.9%)	0.825 (0.297; 96.1%)	0.830 (0.376; 94.8%)	0.829 (0.375; 94.6%)	0.818 (0.376; 95.5%)
	1.86	1.862 (0.309; 95.4%)	1.843 (0.309; 95.8%)	1.853 (0.310; 94.6%)	1.862 (0.393; 95.7%)	1.870 (0.392; 94.2%)	1.838 (0.392; 94.9%)
	3.26	3.281 (0.340; 95.1%)	3.254 (0.339; 95.2%)	3.254 (0.341; 94.4%)	3.272 (0.437; 95.2%)	3.275 (0.438; 94.8%)	3.257 (0.439; 94.5%)
	4.6	4.630 (0.383; 95.2%)	4.608 (0.381; 96.4%)	4.619 (0.388; 94.9%)	4.624 (0.502; 95.5%)	4.593 (0.498; 95.9%)	4.609 (0.502; 95.7%)
	5.8	5.854 (0.431; 95.5%)	5.794 (0.426; 95.2%)	5.822 (0.433; 94.5%)	5.871 (0.574; 94.1%)	5.803 (0.569; 95.6%)	5.844 (0.575; 95.8%)
Gumbel	1.1	1.100 (0.035; 94.5%)	1.099 (0.035; 95.0%)	1.103 (0.035; 94.4%)	1.104 (0.039; 95.2%)	1.100 (0.039; 93.1%)	1.100 (0.039; 94.0%)
	1.25	1.256 (0.048; 94.4%)	1.251 (0.047; 94.0%)	1.250 (0.047; 94.3%)	1.254 (0.055; 95.0%)	1.249 (0.054; 93.0%)	1.254 (0.055; 95.0%)
	1.49	1.494 (0.066; 95.7%)	1.488 (0.064; 93.0%)	1.491 (0.066; 94.9%)	1.494 (0.078; 96.0%)	1.489 (0.075; 95.3%)	1.493 (0.078; 95.0%)
	1.75	1.756 (0.085; 94.7%)	1.745 (0.081; 93.4%)	1.755 (0.085; 94.9%)	1.720 (0.099; 94.6%)	1.737 (0.097; 92.7%)	1.751 (0.102; 94.3%)
	2	2.008 (0.103; 94.8%)	1.994 (0.097; 93.7%)	2.005 (0.102; 94.8%)	2.021 (0.126; 94.9%)	1.992 (0.118; 92.7%)	2.006 (0.125; 93.9%)

3.2 Misspecification of the copula function

We investigate the performance of the estimation procedures under misspecification of the assumed copula function using a small simulation study. The same marginal survival functions, covariate structure and censoring distributions are retained from the study on correct specification. Now we specify the Gumbel-Hougaard copula (Hougaard, 1986a, 1986b) as the true copula function and generate 1000 (500 for the Gaussian copula) samples using the conditional approach (Genest & Favre, 2007). Furthermore, the scope of the numerical study is limited to a moderate association corresponding to a kendall's τ of 0.333. Note that the association parameter θ will be estimated on the scale of the fitted copula function and transformed to Kendall's τ to facilitate comparisons with other copulas.

In general, all estimates for the parameters of interest in Tables 3.4 - 3.6 are biased upwards with larger relative bias for smaller numbers of clusters. No significant difference between the one-stage procedures is found, while the two-stage procedure results in estimates with marginally less bias compared to the other procedures. As previously explained, the standard errors of the regression parameter estimates are not consistent and should be treated with caution. The coverage probability of the association parameter is clearly decreased compared to a correctly specified model. This is particularly evident when the Gaussian copula is fitted. The regression parameter estimates do not seem to be sensitive to misspecification of the copula function in terms of bias and coverage probability contrary to the findings of Kwon et al. (2022). In general, the amount of censoring in the data is shown to be crucial in the estimation of the association parameter. Clearly, severe censoring in the data leads to a highly reduced coverage probability and large increases in the bias for the estimates of the association parameter. This effect seems to be exacerbated when the number of clusters is increased.

Table 3.4: Overview of the parameter estimates for $K = 50$ under misspecification of the copula function. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring			50% censoring		
		One-stage	Two-stage	Spline	One-stage	Two-stage	Spline
Clayton	θ	1.142 (0.507; 92.6%)	1.070 (0.494; 93.4%)	1.206 (0.550; 92.5%)	2.004 (1.011; 94.1%)	1.719 (1.892; 99.2%)	1.920 (0.996; 93.6%)
	τ	0.363 (0.197)	0.349 (0.195)	0.376 (0.211)	0.500 (0.337)	0.462 (0.662)	0.490 (0.336)
	β_1	1.056 (0.498; 92.3%)	1.063 (0.395; 93.7%)	1.064 (0.383; 95.5%)	1.084 (0.526; 93.8%)	1.079 (0.412; 92.3%)	1.079 (0.401; 93.6%)
	β_2	1.072 (0.499; 95.3%)	1.069 (0.396; 93.0%)	1.076 (0.385; 95.5%)	1.075 (0.525; 93.8%)	1.082 (0.410; 92.5%)	1.079 (0.402; 94.4%)
Frank	θ	3.660 (1.167; 92.6%)	3.397 (1.129; 94.7%)	3.637 (1.177; 90.9%)	4.365 (1.683; 92.6%)	4.038 (1.596; 95.9%)	4.424 (1.734; 93.2%)
	τ	0.362 (0.237)	0.341 (0.265)	0.360 (0.242)	0.415 (0.246)	0.391 (0.269)	0.419 (0.247)
	β_1	1.056 (0.481; 93.7%)	1.049 (0.394; 94.1%)	1.083 (0.370; 93.6%)	1.066 (0.513; 94.1%)	1.062 (0.411; 93.2%)	1.094 (0.394; 94.0%)
	β_2	1.064 (0.481; 93.8%)	1.051 (0.394; 94.0%)	1.086 (0.369; 94.5%)	1.085 (0.512; 94.1%)	1.079 (0.411; 92.4%)	1.087 (0.393; 94.1%)
Gaussian	θ	0.549 (0.114; 81.0%)	0.524 (0.118; 86.9%)	0.539 (0.117; 82.8%)	0.586 (0.134; 77.4%)	0.554 (0.140; 83.8%)	0.557 (0.140; 80.4%)
	τ	0.370 (0.087)	0.351 (0.088)	0.362 (0.088)	0.399 (0.105)	0.374 (0.107)	0.376 (0.107)
	β_1	1.067 (0.463; 94.3%)	1.064 (0.394; 93.9%)	1.079 (0.358; 94.1%)	1.070 (0.491; 94.2%)	1.081 (0.411; 93.3%)	1.119 (0.387; 92.3%)
	β_2	1.056 (0.462; 94.0%)	1.077 (0.393; 93.9%)	1.083 (0.360; 94.2%)	1.082 (0.489; 94.5%)	1.084 (0.412; 91.9%)	1.114 (0.384; 94.5%)

Table 3.5: Overview of the parameter estimates for $K = 200$ under misspecification of the copula function. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring			50% censoring		
		One-stage	Two-stage	Spline	One-stage	Two-stage	Spline
Clayton	θ	1.054 (0.239; 92.0%)	1.031 (0.236; 91.6%)	1.087 (0.254; 92.0%)	1.638 (0.422; 73.9%)	1.595 (0.415; 74.9%)	1.689 (0.444; 71.5%)
	τ	0.345 (0.095)	0.340 (0.051)	0.352 (0.100)	0.450 (0.064)	0.444 (0.064)	0.458 (0.156)
	β_1	1.013 (0.242; 95.1%)	1.008 (0.198; 93.9%)	1.027 (0.188; 95.8%)	1.024 (0.251; 93.6%)	1.011 (0.207; 95.5%)	1.029 (0.195; 93.7%)
	β_2	1.027 (0.242; 94.4%)	1.012 (0.198; 95.4%)	1.028 (0.188; 95.6%)	1.012 (0.251; 93.9%)	1.017 (0.207; 94.6%)	1.032 (0.196; 94.4%)
Frank	θ	3.516 (0.555; 90.2%)	3.417 (0.553; 92.8%)	3.545 (0.567; 91.1%)	4.017 (0.755; 85.5%)	3.914 (0.753; 87.0%)	4.074 (0.766; 83.1%)
	τ	0.351 (0.122)	0.343 (0.128)	0.353 (0.123)	0.389 (0.129)	0.382 (0.135)	0.394 (0.127)
	β_1	1.033 (0.233; 95.4%)	1.001 (0.198; 94.0%)	1.031 (0.180; 94.0%)	1.029 (0.246; 93.1%)	1.010 (0.207; 95.6%)	1.045 (0.191; 92.9%)
	β_2	1.025 (0.232; 95.7%)	0.999 (0.198; 94.8%)	1.033 (0.179; 94.1%)	1.033 (0.246; 94.3%)	1.004 (0.207; 95.9%)	1.035 (0.190; 94.2%)
Gaussian	θ	0.539 (0.058; 81.4%)	0.539 (0.058; 82.4%)	1.083 (0.360; 94.2%)	0.580 (0.067; 69.2%)	0.580 (0.067; 70.0%)	0.587 (0.066; 65.9%)
	τ	0.363 (0.044)	0.362 (0.044)	0.367 (0.044)	0.394 (0.052)	0.394 (0.052)	0.400 (0.052)
	β_1	1.023 (0.225; 95.9%)	1.010 (0.198; 95.7%)	1.045 (0.176; 95.8%)	1.028 (0.235; 95.1%)	1.015 (0.206; 94.4%)	1.038 (0.182; 96.2%)
	β_2	1.024 (0.225; 94.7%)	1.000 (0.199; 95.9%)	1.043 (0.175; 95.0%)	1.036 (0.236; 94.6%)	1.015 (0.208; 94.4%)	1.041 (0.182; 94.4%)

Table 3.6: Overview of the parameter estimates for $K = 500$ under misspecification of the copula function. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring			50% censoring		
		One-stage	Two-stage	Spline	One-stage	Two-stage	Spline
Clayton	θ	1.038 (0.149; 92.3%)	1.034 (0.149; 91.7%)	1.074 (0.160; 90.1%)	1.591 (0.261; 39.3%)	1.573 (0.259; 42.1%)	1.648 (0.278; 37.4%)
	τ	0.342 (0.059)	0.341 (0.032)	0.349 (0.063)	0.443 (0.093)	0.440 (0.041)	0.452 (0.098)
	β_1	1.011 (0.152; 95.9%)	1.005 (0.126; 95.3%)	1.022 (0.119; 93.6%)	1.009 (0.158; 94.9%)	1.003 (0.132; 94.7%)	1.020 (0.123; 95.6%)
	β_2	1.003 (0.152; 95.6%)	1.006 (0.126; 95.8%)	1.021 (0.118; 94.3%)	1.013 (0.158; 93.9%)	1.006 (0.131; 95.8%)	1.019 (0.123; 95.3%)
Frank	θ	3.485 (0.348; 88.4%)	3.446 (0.349; 92.8%)	3.493 (0.358; 88.5%)	3.964 (0.470; 68.7%)	3.931 (0.471; 69.9%)	3.997 (0.477; 66.3%)
	τ	0.348 (0.078)	0.345 (0.080)	0.349 (0.080)	0.385 (0.082)	0.383 (0.084)	0.388 (0.082)
	β_1	1.019 (0.146; 94.7%)	1.005 (0.126; 95.7%)	1.018 (0.113; 94.5%)	1.016 (0.154; 94.7%)	1.000 (0.132; 95.9%)	1.023 (0.120; 95.7%)
	β_2	1.017 (0.146; 94.6%)	1.004 (0.126; 95.4%)	1.019 (0.113; 94.8%)	1.017 (0.154; 95.2%)	1.002 (0.132; 95.5%)	1.026 (0.120; 95.0%)
Gaussian	θ	0.538 (0.037; 77.5%)	0.542 (0.037; 72.7%)	0.543 (0.037; 75%)	0.580 (0.042; 49.1%)	0.583 (0.043; 50.6%)	0.584 (0.042; 47%)
	τ	0.361 (0.028)	0.365 (0.028)	0.366 (0.028)	0.394 (0.033)	0.396 (0.034)	0.397 (0.033)
	β_1	1.019 (0.141; 96.0%)	1.002 (0.126; 94.9%)	1.033 (0.11; 93.9%)	1.022 (0.148; 96.3%)	1.004 (0.131; 96.1%)	1.041 (0.114; 94.4%)
	β_2	1.016 (0.141; 96.5%)	1.003 (0.127; 95.0%)	1.033 (0.11; 94.1%)	1.027 (0.148; 95.0%)	1.004 (0.132; 95.9%)	1.045 (0.115; 93.4%)

3.3 Misspecification of the baseline survival function

We now investigate the influence of the baseline survival function on the estimation of the association and regression parameters using the standard one-stage proceedure. Rather than the Weibull baseline survival we have considered until now, we specify a Gompertz (with fixed shape $\lambda = 1$ and variable rate α) and lognormal baseline survival in a model with functional form $h(t) = h_0(t)e^{\beta^T Z}$ (Cox, 1972). Note that for a lognormal baseline survival this model is not closed under the proportional hazards assumption which implies that the fitted baseline survival does not need to be monotonic. Therefore, any result with the lognormal baseline survival will be considered supplemental to the main analysis and is provided in Appendix D. A Weibull baseline survival function was fitted in the case of the strong parametric one-stage procedure. Starting values for the spline method with a total of 3 knots were obtained using the procedure described in Section 2.5. Again, we draw 1000 (500 for the Gaussian copula) samples using the conditional approach (Genest & Favre, 2007). The association parameter will be kept constant across the copulas to simulate a moderate assocation corresponding to a Kendall's τ of 0.333. Furthermore, the covariate structure and censoring distributions will be retained from previous simulations. As it is desirable to investigate the estimation procedures under various degrees of misspecification, three configurations of the Gompertz and lognormal baseline survival function were considered. Figure 3.1 shows the severity of misspecification compared to a Weibull baseline survival.

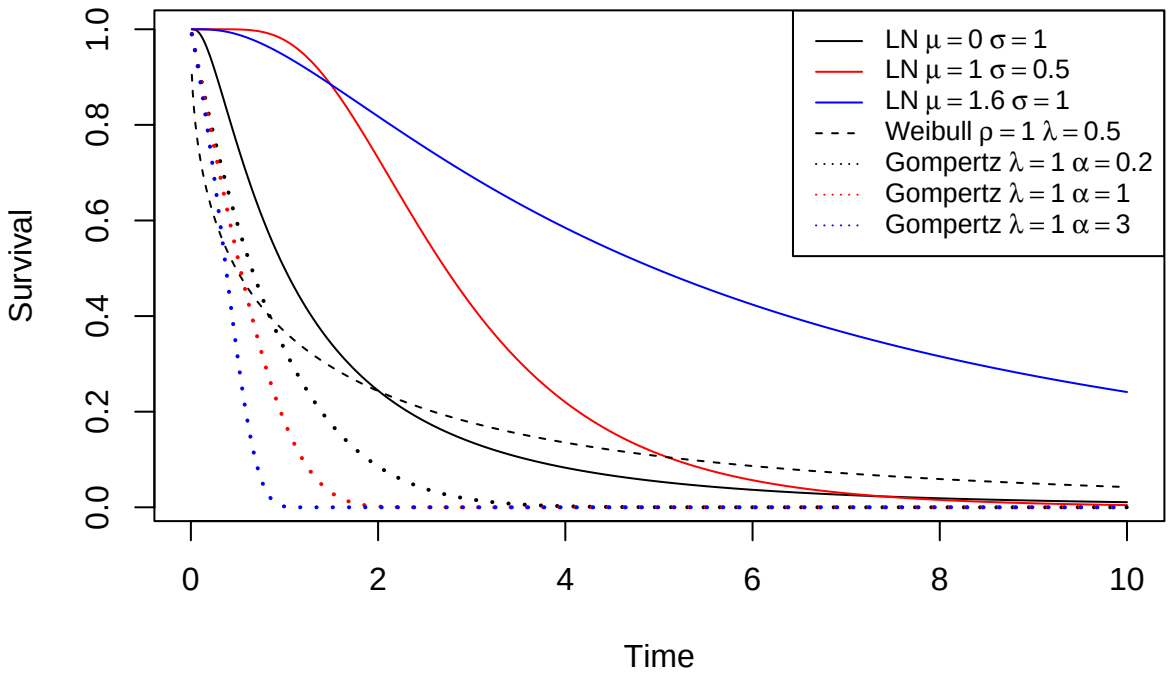


Figure 3.1: Overview of the functions used to specify the baseline hazards in the simulation excercise.

Tables 3.7 - 3.9 show the estimates of the association and regression parameters for different copula functions under various degrees of misspecification compared to the true Gompertz baseline survival function. Focusing first on the association parameter, note that the bias generally increases with in-

creasing sample size. Moreover, the bias is shown to be affected significantly by the amount of censoring present for both estimation procedures. For example, when fitting the Clayton copula, the bias even changes direction from downward to upward when the percentage of censoring is increased leading to a highly overestimated estimate of the association parameter. Increasing the parameter space to a total of 5 knots and repeating the analysis did improve the bias (results not shown in this text). Note that the Clayton and Gumbel copula models tend to prefer the standard parametric one-stage procedure in terms of bias of the association parameter. The Frank and Gaussian copula models on the other hand produce less, albeit sometimes only slightly less, biased estimates using the spline method. This clear division between copulas can potentially be attributed to the fact that the Clayton and Gumbel copulas have a pronounced tail dependency property while the Gaussian and Frank copulas do not have this specific feature. This indicates that the spline method struggles to fit the baseline hazard well in the tails where the Clayton and Gumbel copulas happen to impose a stronger dependency. A possible solution could be to move the single internal towards the tails to allow the spline to capture a larger proportion of the variability in the tails, although this option was not investigated any further. Even though the difference in bias between the two procedures is sometimes large, the spline method does seem to preserve the coverage probability significantly better compared the standard parametric method, especially when the estimate is highly biased. In terms of the regression parameter estimates both estimation procedures seem to perform similarly in terms of bias and coverage probability. The spline method, however, does offer increased precision of the regression estimates compared to the standard parametric method.

Tables D.4 - D.6 give a similar overview of the estimates of interest but now with different variations of the lognormal baseline survival. The lognormal is a special case in the sense that it does not lead to a proportional hazards model due to the non-monotonicity of the baseline hazard. Therefore we expect the spline method to have a significant advantage since it relies heavily on the data to impose monotonicity rather than having it imposed by design as is the case when fitting a strong parametric baseline hazard. This is also evident from the simulation study: estimates of both the regression and association parameters are highly biased in the standard parametric model. The association parameter in particular tends to be highly biased upwards and has a significantly reduced coverage probability. The spline method leads to estimates that are close to unbiased and constrains the coverage probability within an acceptable range of the nominal coverage probability. Similarly for the regression parameters, the spline method maintains a large quantitative advantage compared to the standard parametric method.

Table 3.7: Overview of the parameter estimates for $K = 50$ using the strong parametric one-stage and spline procedures for various configurations of the Gompertz baseline survival function. The copula association parameter is estimated on its original scale and transformed to Kendall's τ to allow for easy comparisons. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
	α	0.2	1	3	0.2	1	3
Clayton	θ	0.670 (0.308; 67.8%)	0.559 (0.276; 57.5%)	0.524 (0.265; 50.5%)	0.836 (0.405; 84.1%)	0.845 (0.397; 83.7%)	0.919 (0.427; 87.9%)
	τ	0.251 (0.132)	0.218 (0.121)	0.207 (0.117)	0.295 (0.167)	0.297 (0.164)	0.315 (0.174)
	β_1	1.005 (0.502; 95.0%)	0.937 (0.504; 95.8%)	0.962 (0.512; 95.8%)	1.046 (0.388; 94.6%)	1.043 (0.387; 96.4%)	1.101 (0.386; 93.3%)
	β_2	1.011 (0.506; 94.8%)	0.966 (0.507; 96.2%)	0.955 (0.512; 96.8%)	1.042 (0.385; 93.7%)	1.02 (0.388; 95%)	1.108 (0.386; 91.2%)
Frank	θ	3.325 (1.002; 93.7%)	3.255 (0.986; 95%)	3.267 (1.014; 93.3%)	3.407 (1.017; 95.4%)	3.355 (1.035; 91.4%)	3.338 (1.048; 93.3%)
	τ	0.335 (0.245)	0.329 (0.252)	0.330 (0.257)	0.342 (0.237)	0.337 (0.249)	0.336 (0.255)
	β_1	1.030 (0.469; 94.3%)	1.030 (0.470; 95.2%)	0.946 (0.471; 94.9%)	1.030 (0.360; 94.1%)	1.062 (0.378; 95.2%)	1.049 (0.368; 92.6%)
	β_2	1.016 (0.469; 94.7%)	0.988 (0.471; 94.9%)	0.952 (0.474; 95.8%)	1.038 (0.362; 93.7%)	1.064 (0.381; 94.1%)	1.069 (0.370; 94.7%)
Gumbel	θ	1.467 (0.199; 92.2%)	1.490 (0.200; 90.9%)	1.501 (0.203; 93%)	1.438 (0.196; 88.2%)	1.438 (0.195; 90.6%)	1.426 (0.197; 87.6%)
	τ	0.318 (0.092)	0.329 (0.090)	0.334 (0.090)	0.305 (0.095)	0.305 (0.094)	0.299 (0.097)
	β_1	0.962 (0.446; 93.7%)	0.930 (0.437; 91.6%)	0.882 (0.437; 95.2%)	1.017 (0.350; 92.2%)	0.998 (0.350; 91.8%)	0.990 (0.356; 92.5%)
	β_2	0.979 (0.442; 93.0%)	0.943 (0.438; 93.0%)	0.910 (0.437; 92.8%)	1.010 (0.349; 93.1%)	0.984 (0.348; 90.3%)	0.976 (0.353; 93.3%)
Gaussian	θ	0.509 (0.108; 90.9%)	0.487 (0.111; 90.3%)	0.471 (0.114; 93.3%)	0.500 (0.111; 90.3%)	0.490 (0.114; 92.6%)	0.493 (0.116; 93.0%)
	τ	0.340 (0.080)	0.324 (0.081)	0.312 (0.082)	0.333 (0.082)	0.326 (0.083)	0.328 (0.085)
	β_1	1.061 (0.463; 94.7%)	0.972 (0.466; 96.6%)	0.952 (0.468; 95.8%)	1.064 (0.357; 93.7%)	1.053 (0.362; 92.0%)	1.090 (0.364; 93.3%)
	β_2	1.018 (0.459; 91.2%)	0.990 (0.467; 94.1%)	0.922 (0.470; 95.8%)	1.063 (0.359; 93.3%)	1.037 (0.358; 91.9%)	1.049 (0.362; 94.3%)
		50% censoring					
	α	0.2	1	3	0.2	1	3
Clayton	θ	1.319 (0.632; 91.3%)	1.285 (0.673; 87.2%)	1.511 (0.955; 87.3%)	1.539 (0.733; 92.5%)	1.586 (0.824; 92.9%)	1.817 (1.089; 89.1%)
	τ	0.397 (0.238)	0.391 (0.255)	0.43 (0.347)	0.435 (0.265)	0.442 (0.295)	0.476 (0.374)
	β_1	1.057 (0.526; 94.7%)	1.032 (0.557; 92.7%)	1.019 (0.626; 94.4%)	1.064 (0.397; 93.5%)	1.078 (0.422; 95.0%)	1.096 (0.468; 89.5%)
	β_2	1.060 (0.522; 95.0%)	1.023 (0.550; 92.0%)	1.009 (0.619; 93.4%)	1.072 (0.399; 94.1%)	1.075 (0.419; 93.3%)	1.099 (0.469; 91.3%)
Frank	θ	3.514 (1.359; 94.1%)	3.466 (1.496; 94.3%)	3.860 (1.918; 95.6%)	3.403 (1.351; 93.9%)	3.481 (1.492; 95.8%)	3.643 (1.941; 94%)
	τ	0.350 (0.299)	0.347 (0.338)	0.378 (0.352)	0.341 (0.316)	0.348 (0.334)	0.361 (0.398)
	β_1	1.060 (0.518; 94.1%)	1.021 (0.541; 94.1%)	0.974 (0.594; 92.2%)	1.093 (0.402; 94.7%)	1.088 (0.417; 94.3%)	1.127 (0.476; 93.3%)
	β_2	1.028 (0.516; 95.4%)	1.006 (0.541; 93.9%)	0.994 (0.601; 92.7%)	1.119 (0.404; 93.7%)	1.107 (0.423; 94.1%)	1.133 (0.476; 93.7%)
Gumbel	θ	1.373 (0.215; 81.7%)	1.425 (0.247; 85.7%)	1.455 (0.317; 85.3%)	1.356 (0.213; 81%)	1.365 (0.234; 82.4%)	1.395 (0.299; 81.4%)
	τ	0.272 (0.114)	0.298 (0.122)	0.313 (0.15)	0.263 (0.116)	0.267 (0.126)	0.283 (0.154)
	β_1	1.011 (0.498; 93.0%)	0.952 (0.511; 91.8%)	0.923 (0.583; 91.6%)	1.053 (0.389; 92.0%)	1.057 (0.41; 91.6%)	1.083 (0.465; 91.6%)
	β_2	1.016 (0.498; 88.8%)	0.939 (0.507; 92.2%)	0.905 (0.566; 93.3%)	1.045 (0.388; 90.7%)	1.054 (0.41; 92.9%)	1.072 (0.461; 91.4%)
Gaussian	θ	0.494 (0.142; 90.7%)	0.499 (0.151; 88.4%)	0.510 (0.177; 82.4%)	0.493 (0.141; 89.5%)	0.491 (0.153; 90.1%)	0.488 (0.187; 88.8%)
	τ	0.329 (0.104)	0.332 (0.111)	0.341 (0.131)	0.328 (0.103)	0.327 (0.112)	0.324 (0.136)
	β_1	1.035 (0.506; 95.2%)	0.974 (0.528; 94.9%)	0.970 (0.582; 94.1%)	1.092 (0.393; 90.5%)	1.084 (0.413; 95.6%)	1.060 (0.464; 92.4%)
	β_2	1.034 (0.512; 93.0%)	0.983 (0.528; 91.2%)	1.004 (0.592; 92.9%)	1.088 (0.391; 90.7%)	1.066 (0.41; 93.8%)	1.074 (0.462; 91.8%)

Table 3.8: Overview of the parameter estimates for $K = 200$ using the strong parametric one-stage and spline procedures for various configurations of the Gompertz baseline survival function. The copula association parameter is estimated on its original scale and transformed to Kendall's τ to allow for easy comparisons. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
	α	0.2	1	3	0.2	1	3
Clayton	θ	0.621 (0.148; 31.2%)	0.526 (0.131; 9.3%)	0.473 (0.124; 5.5%)	0.764 (0.189; 68.1%)	0.800 (0.194; 75.5%)	0.812 (0.200; 77.5%)
	τ	0.237 (0.064)	0.208 (0.058)	0.191 (0.055)	0.276 (0.079)	0.286 (0.081)	0.289 (0.083)
	β_1	0.982 (0.245; 94.9%)	0.946 (0.247; 95.2%)	0.921 (0.25; 96.4%)	1.019 (0.189; 96.2%)	1.010 (0.188; 94.7%)	1.024 (0.190; 94.9%)
	β_2	1.012 (0.246; 95.0%)	0.952 (0.248; 97.1%)	0.914 (0.25; 94.7%)	1.015 (0.189; 96.0%)	1.007 (0.188; 96.2%)	1.024 (0.191; 96.6%)
Frank	θ	3.246 (0.486; 96.0%)	3.159 (0.477; 96.0%)	3.135 (0.485; 93.3%)	3.259 (0.493; 96.0%)	3.252 (0.495; 93.9%)	3.252 (0.509; 92.8%)
	τ	0.328 (0.125)	0.321 (0.129)	0.319 (0.134)	0.329 (0.126)	0.329 (0.127)	0.329 (0.13)
	β_1	1.000 (0.228; 96.0%)	0.932 (0.229; 93.5%)	0.911 (0.230; 94.3%)	1.008 (0.177; 94.9%)	1.014 (0.178; 93.3%)	0.993 (0.179; 93.9%)
	β_2	0.994 (0.228; 95.0%)	0.950 (0.229; 94.5%)	0.921 (0.230; 93.9%)	1.016 (0.177; 96.6%)	1.012 (0.177; 94.5%)	0.996 (0.179; 95.2%)
Gumbel	θ	1.430 (0.096; 86.5%)	1.453 (0.096; 91.0%)	1.471 (0.098; 92.8%)	1.424 (0.095; 86.3%)	1.421 (0.096; 87.6%)	1.409 (0.096; 80.8%)
	τ	0.301 (0.047)	0.312 (0.045)	0.320 (0.045)	0.298 (0.047)	0.296 (0.048)	0.290 (0.048)
	β_1	0.952 (0.217; 91.6%)	0.879 (0.212; 88.6%)	0.841 (0.210; 89.0%)	0.945 (0.169; 92.6%)	0.941 (0.170; 92.2%)	0.935 (0.172; 90.1%)
	β_2	0.943 (0.216; 92.2%)	0.881 (0.212; 90.5%)	0.848 (0.210; 89.0%)	0.947 (0.169; 92.8%)	0.945 (0.170; 91.2%)	0.942 (0.172; 91.0%)
Gaussian	θ	0.494 (0.055; 95.6%)	0.484 (0.055; 93.7%)	0.477 (0.056; 93.5%)	0.501 (0.056; 95.6%)	0.499 (0.056; 93.7%)	0.502 (0.057; 95.4%)
	τ	0.329 (0.04)	0.321 (0.04)	0.316 (0.041)	0.334 (0.041)	0.333 (0.041)	0.335 (0.042)
	β_1	0.975 (0.225; 93.9%)	0.967 (0.227; 96.4%)	0.923 (0.229; 94.5%)	1.012 (0.174; 95.6%)	1.003 (0.175; 94.3%)	1.001 (0.176; 95.4%)
	β_2	0.985 (0.226; 95.4%)	0.941 (0.227; 94.7%)	0.897 (0.228; 94.9%)	1.013 (0.175; 95.8%)	1.006 (0.175; 93.3%)	0.995 (0.176; 96.0%)
		50% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
	α	0.2	1	3	0.2	1	3
Clayton	θ	1.055 (0.268; 92.8%)	0.96 (0.272; 83.6%)	0.832 (0.297; 75.8%)	1.298 (0.336; 87.7%)	1.291 (0.353; 88.9%)	1.262 (0.417; 90.5%)
	τ	0.345 (0.106)	0.324 (0.11)	0.294 (0.123)	0.393 (0.127)	0.392 (0.133)	0.387 (0.158)
	β_1	0.999 (0.255; 93.7%)	0.945 (0.266; 94.9%)	0.949 (0.295; 97.1%)	1.036 (0.196; 94.7%)	1.021 (0.206; 95.8%)	1.028 (0.229; 95.2%)
	β_2	1.001 (0.255; 95.6%)	0.947 (0.266; 93.5%)	0.934 (0.296; 95.4%)	1.039 (0.196; 96.4%)	1.021 (0.205; 97.1%)	1.026 (0.230; 96.0%)
Frank	θ	3.257 (0.626; 94.9%)	3.188 (0.672; 96.0%)	3.215 (0.821; 94.3%)	3.252 (0.631; 94.3%)	3.304 (0.698; 94.7%)	3.382 (0.854; 94.3%)
	τ	0.329 (0.16)	0.323 (0.179)	0.326 (0.215)	0.329 (0.161)	0.333 (0.173)	0.340 (0.202)
	β_1	1.002 (0.247; 95.0%)	0.968 (0.260; 95.0%)	0.939 (0.286; 94.9%)	1.020 (0.193; 94.3%)	1.013 (0.202; 95.0%)	1.041 (0.225; 92.8%)
	β_2	1.012 (0.248; 96.8%)	0.955 (0.258; 96.4%)	0.926 (0.286; 95.0%)	1.024 (0.194; 94.5%)	1.010 (0.201; 95.4%)	1.036 (0.225; 94.7%)
Gumbel	θ	1.338 (0.100; 65.7%)	1.366 (0.111; 73.9%)	1.394 (0.137; 82.5%)	1.338 (0.101; 60.8%)	1.328 (0.107; 60%)	1.326 (0.127; 68.3%)
	τ	0.253 (0.056)	0.268 (0.059)	0.283 (0.071)	0.252 (0.056)	0.247 (0.061)	0.246 (0.072)
	β_1	0.946 (0.238; 93.3%)	0.914 (0.245; 93.7%)	0.870 (0.270; 92.0%)	0.966 (0.186; 91.6%)	0.974 (0.196; 93.5%)	0.972 (0.221; 92.9%)
	β_2	0.951 (0.239; 93.1%)	0.912 (0.243; 92.6%)	0.901 (0.270; 92.2%)	0.973 (0.186; 93.9%)	0.967 (0.196; 92.8%)	0.984 (0.221; 92.7%)
Gaussian	θ	0.495 (0.070; 95.0%)	0.499 (0.075; 93.1%)	0.497 (0.089; 93.9%)	0.495 (0.070; 96%)	0.501 (0.075; 91.9%)	0.495 (0.093; 93.1%)
	τ	0.330 (0.051)	0.333 (0.055)	0.331 (0.065)	0.330 (0.051)	0.334 (0.055)	0.330 (0.068)
	β_1	1.005 (0.244; 95.2%)	0.945 (0.252; 94.3%)	0.927 (0.280; 95.2%)	1.012 (0.189; 93.3%)	1.007 (0.198; 95%)	1.024 (0.222; 94.7%)
	β_2	1.009 (0.243; 95.4%)	0.947 (0.253; 94.9%)	0.926 (0.281; 96.0%)	1.012 (0.189; 94.9%)	1.012 (0.198; 96.6%)	1.021 (0.222; 93.9%)

Table 3.9: Overview of the parameter estimates for $K = 500$ using the strong parametric one-stage and spline procedures for various configurations of the Gompertz baseline survival function. The copula association parameter is estimated on its original scale and transformed to Kendall's τ to allow for easy comparisons. Note that the Gaussian copula is excluded due to computational needs of the estimation procedures under repeated sampling. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
	α	0.2	1	3	0.2	1	3
Clayton	θ	0.603 (0.092; 3.8%)	0.515 (0.082; 0.2%)	0.460 (0.077; 0.0%)	0.757 (0.118; 45.5%)	0.770 (0.120; 50.1%)	0.783 (0.123; 55.0%)
	τ	0.232 (0.040)	0.205 (0.036)	0.187 (0.035)	0.275 (0.050)	0.278 (0.050)	0.281 (0.051)
	β_1	0.991 (0.155; 95.6%)	0.939 (0.155; 95.2%)	0.903 (0.157; 90.5%)	1.021 (0.119; 94.1%)	1.010 (0.119; 94.7%)	1.013 (0.12; 93.9%)
	β_2	0.988 (0.155; 96.2%)	0.944 (0.156; 95.4%)	0.896 (0.157; 93.5%)	1.019 (0.119; 95.4%)	1.006 (0.119; 95.4%)	1.015 (0.12; 94.3%)
Frank	θ	3.199 (0.305; 95.8%)	3.132 (0.300; 91.6%)	3.084 (0.304; 92.8%)	3.268 (0.311; 95.2%)	3.235 (0.312; 93.1%)	3.271 (0.320; 89.7%)
	τ	0.324 (0.081)	0.319 (0.083)	0.315 (0.086)	0.330 (0.079)	0.327 (0.081)	0.330 (0.081)
	β_1	0.978 (0.144; 93.7%)	0.952 (0.144; 94.7%)	0.912 (0.145; 90.7%)	1.006 (0.112; 94.1%)	1.000 (0.112; 96.0%)	1.001 (0.113; 94.9%)
	β_2	0.988 (0.144; 94.7%)	0.954 (0.144; 94.1%)	0.895 (0.145; 92.4%)	1.012 (0.112; 95.0%)	1.005 (0.112; 95.0%)	0.998 (0.113; 93.9%)
Gumbel	θ	1.430 (0.060; 77.3%)	1.448 (0.060; 86.3%)	1.467 (0.062; 90.9%)	1.425 (0.061; 79.2%)	1.419 (0.060; 78.9%)	1.412 (0.061; 73.3%)
	τ	0.301 (0.029)	0.309 (0.029)	0.318 (0.029)	0.298 (0.030)	0.295 (0.030)	0.292 (0.031)
	β_1	0.941 (0.136; 90.5%)	0.875 (0.133; 85.7%)	0.832 (0.132; 76.0%)	0.942 (0.106; 89.1%)	0.933 (0.107; 87.4%)	0.932 (0.108; 88.2%)
	β_2	0.934 (0.135; 89.0%)	0.877 (0.133; 83.0%)	0.842 (0.133; 77.1%)	0.941 (0.106; 92.0%)	0.929 (0.106; 86.9%)	0.931 (0.107; 88.0%)
		50% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
	α	0.2	1	3	0.2	1	3
Clayton	θ	1.044 (0.168; 89.5%)	0.906 (0.164; 79.6%)	0.784 (0.178; 65.5%)	1.229 (0.201; 80.4%)	1.240 (0.218; 79.6%)	1.212 (0.257; 85.9%)
	τ	0.343 (0.067)	0.312 (0.067)	0.282 (0.074)	0.381 (0.077)	0.383 (0.083)	0.377 (0.099)
	β_1	0.989 (0.160; 95.0%)	0.946 (0.168; 94.7%)	0.925 (0.185; 93%)	1.017 (0.124; 95.8%)	1.014 (0.129; 94.3%)	1.023 (0.144; 94.7%)
	β_2	0.987 (0.160; 95.0%)	0.950 (0.168; 95.4%)	0.924 (0.185; 95%)	1.014 (0.124; 95.0%)	1.014 (0.129; 95.2%)	1.024 (0.144; 96.0%)
Frank	θ	3.229 (0.391; 94.7%)	3.168 (0.419; 94.1%)	3.139 (0.505; 92.8%)	3.264 (0.397; 95.6%)	3.255 (0.433; 94.5%)	3.288 (0.528; 93.3%)
	τ	0.327 (0.101)	0.322 (0.113)	0.319 (0.139)	0.33 (0.101)	0.329 (0.111)	0.332 (0.132)
	β_1	0.990 (0.156; 94.5%)	0.958 (0.162; 95%)	0.918 (0.179; 93.0%)	1.006 (0.121; 93.9%)	1.002 (0.127; 93.3%)	0.997 (0.141; 94.9%)
	β_2	0.989 (0.156; 96.0%)	0.946 (0.163; 95%)	0.916 (0.179; 94.7%)	1.007 (0.121; 93.7%)	1.000 (0.127; 92.0%)	1.005 (0.141; 95.0%)
Gumbel	θ	1.332 (0.063; 30.9%)	1.348 (0.068; 44.2%)	1.368 (0.083; 64.8%)	1.327 (0.063; 28.2%)	1.321 (0.067; 29.1%)	1.317 (0.079; 41.3%)
	τ	0.249 (0.036)	0.258 (0.037)	0.269 (0.044)	0.247 (0.036)	0.243 (0.038)	0.240 (0.046)
	β_1	0.953 (0.15; 92.2%)	0.901 (0.154; 88.8%)	0.859 (0.169; 86.3%)	0.952 (0.117; 91.6%)	0.959 (0.124; 94.5%)	0.968 (0.138; 92.8%)
	β_2	0.941 (0.15; 92.8%)	0.892 (0.154; 88.8%)	0.873 (0.169; 89.0%)	0.948 (0.117; 90.5%)	0.952 (0.123; 93.3%)	0.971 (0.138; 92.6%)

4 Ethanol-induced sleep time data

To demonstrate the practical use of the discussed copula methods we now turn to the analysis of ethanol-induced sleep time in genetically-selected strains of mice. Markel et al. (1995) originally designed a study where the aim is to investigate the genetic influence on sleep time. The mice included in the study were selected from two inbred strains, the isogenic F_1 population and the segregating F_2 population induced by crossbreeding from the F_1 population. The original study was designed to allow for a repeated measurements design where, after intraperitoneal injection of a 4.1 g/kg dose of ethanol, the length of the induced sleep time was measured twice with a prespecified interval between measurements. Since some mice might not fall asleep or sleep only a very short time a detection limit of 1 minute was imposed. In practice, this results in the application of left censoring for mice that did not sleep or slept less than 1 minute.

As the study of Markel et al. (1995) has been investigated several times (see for example Braekers & Markel (2006) or Fogap (2007)) we will focus mainly on the application of the proposed methods. Furthermore we limit ourselves to extending the univariate regression analysis of Braekers & Grouwels (2016) where the influence of sex, albinism, its interaction and the weight of the mouse on the sleep times of the F_2 population is investigated. Now we will also be able to model the association structure between response times for the same mouse through the use of copulas. Firstly, we introduce a Cox model for the margins with the required covariate structure:

$$H_j(y_{ij}) = H_{0j}(y_{ij}) \exp\Big(\beta_1 + \beta_2 \text{Sex}_i + \beta_3 \text{Alb}_i + \beta_4 \text{Sex:Alb}_i + \beta_{5j} \text{Weight}_{ij}\Big)$$

where y_{ij} is the sleep time for mouse i in trial j (j=1,2), H_j the cumulative hazard for trial j and H_{0j} the baseline cumulative hazard for trial j. Firstly note that, unlike other covariates, the weight of the mouse is recorded once for each trial. Therefore the weight is a margin-specific covariate whereas the other covariates have a $\log HR$ that is constrained to be the same between the two trials. Furthermore, since the covariate structure is equal for all mice within the same trial a trial-specific baseline hazard can be specified. Lastly, note that the model is identified by setting a male mouse with no albinism as the reference level.

Estimation of the model will be done in different ways: firstly the model is estimated under the working independence assumption. We will assume either a Weibull or a spline baseline hazard to investigate the difference in model fit. The next step will be estimating the association parameter θ as described in Section 2. The independence model naturally leads to the two-stage estimation procedure. Furthermore, we extend the analysis to full maximum likelihood estimation both in a strong and weak parametric framework. Table 4.1 shows the estimated independence model when fitting a Weibull or a spline with 3 knots as the baseline hazard. In both the Weibull and spline model we find a significant effect of sex on the induced sleep time meaning that female mice generally have a shorter sleep time. Furthermore, an increased weight at the start of trial 2 significantly increases the hazard indicating a shorter sleep time. Comparing our bivariate analysis with the univariate analysis of trial 1 by Braekers & Grouwels (2016), we observe a much larger effect of sex when taking into account the repeated measurements design of the study. Moreover, when comparing to the proposed semi-parametric model by Braekers & Grouwels (2016), the spline model seems to be similar in fit. Even though a formal likelihood ratio test between the two non-nested independence models is not possible, a comparison of the log-likelihood values also indicates that flexible parametric modeling of the baseline hazard in this case allows for a superior fit ($-2ll = 21548$ for the spline model against 22233 for the Weibull model).

Now we extend the independence model by including the association parameter θ that will be estimated

Table 4.1: Parameter estimates of the bivariate independence model where the baseline hazard is assumed to be either a Weibull hazard or flexible spline. Standard errors are given in brackets. * indicates a significant p-value at the 5% significance level.

		Trial 1	Trial 2
Weibull	Sex	0.162 (0.062)*	
	Albinism	0.103 (0.071)	
	Sex:Albinism	-0.048 (0.102)	
	Weight	-0.003 (0.011)	0.027 (0.012)*
Spline	Sex	0.193 (0.062)*	
	Albinism	0.062 (0.071)	
	Sex:Albinism	0.020 (0.102)	
	Weight	-0.004 (0.011)	0.039 (0.012)*

by means of a copula model. Since the outcome structure is bivariate the likelihood expressions in Appendices A and B are valid. Furthermore, the functional form of the baseline hazard is retained from the independence model. The association parameter and/or regression parameters will be estimated by means of the parametric two-stage and one-stage procedures. Starting values for the regression parameters in the Weibull one-stage procedure are obtained from the two-stage procedure whereas starting values and knot positions for the spline baseline hazard are obtained by means of the method described in Section 2.5. Table 4.2 shows the parameter estimates for all copula models. Note that the regression estimates of the two-stage procedure are equal to the independence model as expected. Moreover, it is important to note that estimation with the Clayton copula was attempted but resulted in unstable numerical optimization due to the low amount of censoring ($\sim$ 5%) in the data. To this end the analysis was limited to the Gumbel-Hougaard, Frank and Gaussian copulas.

Examining the estimates in Table 4.2 it is clear that imposing an association structure on the bivariate response structure has an impact on the magnitude of the covariate effects. Comparing the regression estimates of the standard and spline one-stage procedures with the two-stage regression estimates obtained under the working independence assumption, a clear decrease in the effect of albinism and sex on the sleep time is noted. For example, mice with albinism now undergo a decrease in hazard which implies a longer sleep time, while under the implausible independence assumption one would on average expect shorter sleep times. The interaction term generally seems to have a larger effect in the spline models compared to the standard one-stage procedure. In terms of the association parameter, all models indicate a low to moderate association between the two trials. Although it is important to note that our simulations showed that both one-stage procedures are susceptible to upward bias of the association parameter when the copula is misspecified, it is reasonable to assume that in this case the bias will be rather minimal since it has been shown to decrease with a decreasing percentage of censoring. Testing for independence can easily be done by means of a likelihood ratio test. The test statistic under the null hypothesis $H_0 : \theta = \theta_0$ against a two-sided alternative hypothesis where θ_0 is the value of the association parameter that imposes independence generally follows a χ^2-distribution with 1 degree of freedom. For the Frank and Gaussian copula we have $\theta_0 = 0$ and for the Gumbel-Hougaard copula $\theta_0 = 1$. Note that θ_0 lies on the boundary of the parameter space in the case of the Gumbel-Hougaard copula. This implies that a standard likelihood ratio test is not valid and should be adjusted to reflect the boundary problem. In this case the null distribution of the test statistic is a mixture χ^2-distribution with 0 and 1 degrees of freedom and equal weighting probabilities. A likelihood ratio test therefore implies rejecting the null hypothesis that the two trials are independent at the 5% significance level for all

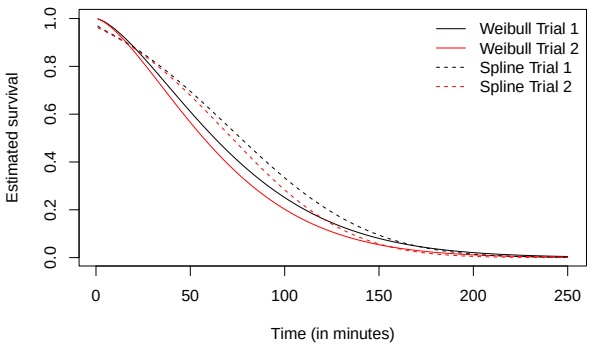
Table 4.2: Parameter estimates of the copula models. Note that the regression estimates from the Weibull independence model are displayed for the two-stage procedure. Standard errors are given in brackets. * indicates a significant p-value at the 5% significance level.

		Two-stage	One-stage Weibull	Spline 3 knots
Gumbel	θ	1.508 (0.041)	1.565 (0.045)	1.471 (0.047)
	Sex	0.162 (0.062)*	0.086 (0.071)	0.139 (0.070)*
	Albinism	0.103 (0.071)	-0.048 (0.083)	-0.068 (0.081)
	Sex:Albinism	-0.048 (0.102)	0.063 (0.119)	0.113 (0.116)
	Weight Trial 1	-0.003 (0.011)	-0.022 (0.011)*	-0.019 (0.011)
	Weight Trial 2	0.027 (0.012)*	-0.005 (0.012)	0.009 (0.012)
Frank	θ	4.061 (0.254)	4.230 (0.248)	4.253 (0.241)
	Sex	0.162 (0.062)*	0.057 (0.072)	0.084 (0.072)
	Albinism	0.103 (0.071)	-0.001 (0.089)	-0.024 (0.087)
	Sex:Albinism	-0.048 (0.102)	0.077 (0.123)	0.149 (0.121)
	Weight Trial 1	-0.003 (0.011)	-0.019 (0.011)	-0.018 (0.011)
	Weight Trial 2	0.027 (0.012)*	-0.004 (0.012)	0.004 (0.012)
Gaussian	θ	0.329 (0.031)	0.331 (0.023)	0.466 (0.026)
	Sex	0.162 (0.062)*	0.125 (0.069)	0.134 (0.072)
	Albinism	0.103 (0.071)	-0.005 (0.081)	-0.075 (0.084)
	Sex:Albinism	-0.048 (0.102)	0.049 (0.115)	0.151 (0.119)
	Weight Trial 1	-0.003 (0.011)	-0.010 (0.011)	-0.015 (0.011)
	Weight Trial 2	0.027 (0.012)*	0.014 (0.012)	0.015 (0.012)

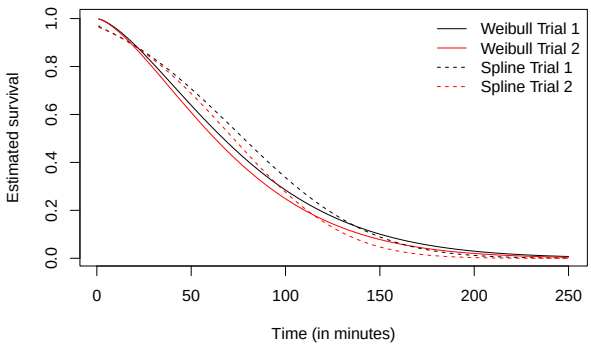
proposed models. Since the estimates of the association parameter differ to some extent, especially for the Gaussian copula, we further investigate model fit. Table 4.3 shows the AIC values for all one-stage models considered in this analysis. Clearly, a flexible spline model offers significant gains in term of model fit relative to the Weibull models. Between the spline models the relative difference in fit is small but the Frank model is preferred. Visual differences in model fit are also evident. Figure 4.1 shows the estimated survival curves for a male mouse with no albinism and average weight. Firstly, the difference in the association parameter between the Gumbel-Hougaard models can be explained by the difference in fit in the upper tails of the survival curves as the copula imposes a stronger dependency for shorter sleep times. Similarly, the estimated survival curves of the spline Gaussian model seem to differ quite significantly in shape between the Weibull models, potentially explaining the large differences in the association parameter. While the Frank models also show a difference in fit, the spline model produces a fairly similar shape to the Weibull model. Furthermore, the Frank copula does not have a tail dependence property which results in a combination that possibly explains the smaller difference in the association parameter.

Table 4.3: AIC values of all one-stage models.

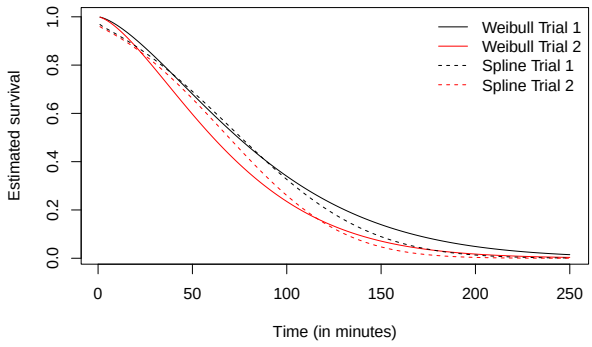
	Weibull	Spline
Gumbel	21961.58	21359.09
Frank	21937.45	21241.59
Gaussian	22085.35	21341.16



(a) Gumbel-Hougaard



(b) Frank



(c) Gaussian

Figure 4.1: Estimated survival curves for a male mouse with no albinism and average weight using different copulas in the two proposed one-stage procedures.

5 Conclusion

Based on existing methodology on the modelling of clustered data subject to right censoring, we propose a parametric copula model that allows for marginal modelling of clustered data subject to left censoring. In the context of the Archimedean family of copulas, the proposed model is an extension of the copula model by Prenen et al. (2017) which allows for clusters of varying size and subject-specific covariate structure. Furthermore an alternative model using the Gaussian copula based on techniques by Othus & Li (2010) is proposed. More specifically, the full likelihood for bivariate data subject to left censoring is derived for various copulas configurations. Since full parametric specification of the baseline survival might be too strong to assume, a flexible parametric alternative based on the work of Royston & Parmar (2002) can be specified in combination with the proposed copula model.

Estimation can be performed in several ways. Firstly, the one-stage procedure is a multivariate maximum likelihood procedure where the association parameter of the copula and the marginal parameters are estimated simultaneously. Secondly, when one-stage estimation is not feasible computationally or otherwise, the two-stage procedure can be used. In the first stage the marginal and regression parameters are estimated under the working independence assumption. Using the estimates from the first stage as plug-in estimates in the full likelihood, a univariate maximum likelihood procedure is used to estimate the association parameter. Since the marginal parameters are estimated without taking into account the true association structure the regression estimates are not suitable for inference. Furthermore, the variance of the association parameter requires a Huber-like correction to impose consistency of the estimator. This is often cumbersome from a practical point of view and therefore resampling methods to approximate the variance are preferred. Extensive simulations show that the one-stage procedure is preferable under correct specification of the copula and baseline survival. On the other hand, the one-stage estimator of the association parameter is rather sensitive to misspecification of the copula and shows significant upward bias under heavy censoring. When misspecifying the baseline survival, we furthermore find that using a flexible baseline survival in the one-stage procedure increases precision and reduces bias in the regression estimates.

Until now we have assumed that the censoring mechanism is completely independent from the outcome. This assumption is strong and generally not realistic in many settings, leading to biased model estimates. Copula models have shown to be useful in the modelling of the dependence structure between the censoring and survival times (Emura & Chen, 2018). An extension of the models we have considered in this text to allow for dependent censoring therefore would seem natural. A recent example of such an extension for bivariate data is provided by Deresa et al. (2022). On a similar note, recent developments in the field of dependent censoring (Deresa & Van Keilegom, 2020) show that parametric transformation models, not unlike the model by Royston & Parmar (2002), can be used to effectively model dependent censoring. Indeed, the dependence structure imposed by a bivariate normal distribution might be replaced by a more flexible copula model combined with splines as proposed by this text. In any case it is clear that there is still ample room for copula models to grow and develop into more flexible and complex models.

6 References

Andersen, P. K., Ekstrøm, C. T., Klein, J. P., Shu, Y., & Zhang, M.-J. (2005). A Class of Goodness of Fit Tests for a Copula Based on Bivariate Right-Censored Data. *Biometrical Journal*, *47*(6), 815–824. https://doi.org/10.1002/bimj.200410163

Bijma, F., Jonker, M., & Vaart, A. W. van der. (2017). *An introduction to mathematical statistics*. Amsterdam University Press.

Braekers, R., & Grouwels, Y. (2016). A semi-parametric cox's regression model for zero-inflated left-censored time to event data. *Communications in Statistics - Theory and Methods*, *45*(7), 1969–1988. https://doi.org/10.1080/03610926.2013.870207

Braekers, R., & Markel, P. (2006). *Modelling ethanol-induced sleeping time in inbred strains of mice : A repeated measures design with left-censored data*.

Canales, R. A., Wilson, A. M., Pearce-Walker, J. I., Verhougstraete, M. P., & Reynolds, K. A. (2018). Methods for Handling Left-Censored Data in Quantitative Microbial Risk Assessment. *Applied and Environmental Microbiology*, *84*(20), e01203–18. https://doi.org/10.1128/AEM.01203-18

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*(1), 141–151. https://doi.org/10.1093/biomet/65.1.141

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Deresa, N. W., & Van Keilegom, I. (2020). Flexible parametric model for survival data subject to dependent censoring. *Biometrical Journal*, *62*(1), 136–156. https://doi.org/10.1002/bimj.201800375

Deresa, N. W., Van Keilegom, I., & Antonio, K. (2022). Copula-based inference for bivariate survival data with left truncation and dependent censoring. *Insurance: Mathematics and Economics*, *107*, 1–21. https://doi.org/10.1016/j.insmatheco.2022.07.011

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

Durrleman, S., & Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, *8*(5), 551–561. https://doi.org/10.1002/sim.4780080504

Emura, T., & Chen, Y.-H. (2018). *Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches*. Springer Singapore. https://doi.org/10.1007/978-981-10-7164-5

Emura, T., Matsui, S., & Rondeau, V. (2019). *Survival Analysis with Correlated Endpoints: Joint Frailty-Copula Models*. Springer Singapore. https://doi.org/10.1007/978-981-13-3516-7

Fogap, N. T. (2007). *Modelling the ethanol-induced sleeping time in mice through a zero inflated model* [Master's thesis].

Genest, C., & Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, *12*(4), 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Genest, C., Ghoudi, K., & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, *82*(3), 543–552. https://doi.org/10.1093/biomet/82.3.543

Ha, I. D., & Lee, Y. (2003). Estimating Frailty Models via Poisson Hierarchical Generalized Linear Models. *Journal of Computational and Graphical Statistics*, *12*(3), 663–681. https://doi.org/10.1198/1061860032256

He, W., & Lawless, J. F. (2003). Flexible Maximum Likelihood Methods for Bivariate Proportional Hazards Models. *Biometrics*, *59*(4), 837–848. https://doi.org/10.1111/j.0006-341X.2003.00098.x

Helsel, D. R. (2006). Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere*, *65*(11), 2434–2439. https://doi.org/10.1016/j.chemosphere.2006.04.051

Helsel, D. R. (2011). *Statistics for Censored Environmental Data* Using Minitab® and R*: Helsel/Statistics for Environmental Data 2E*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118162729

Hewett, P., & Ganser, G. (2007). A Comparison of Several Methods for Analyzing Censored Data. *The Annals of Occupational Hygiene*. https://doi.org/10.1093/annhyg/mem045

Hofert, M., Mächler, M., & McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, *110*, 133–150. https://doi.org/10.1016/j.jmva.2012.02.019

Hougaard, P. (1986a). A Class of Multivariate Failure Time Distributions. *Biometrika*, *73*(3), 671. https://doi.org/10.2307/2336531

Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, *73*(2), 387–396. https://doi.org/10.1093/biomet/73.2.387

Hougaard, P. (1989). Fitting a multivariate failure time distribution. *IEEE Transactions on Reliability*, *38*(4), 444–448. https://doi.org/10.1109/24.46460

Jacqmin-Gadda, H. (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics*, *1*(4), 355–368. https://doi.org/10.1093/biostatistics/1.4.355

Joe, H. (2015). *Dependence modeling with copulas*. CRC Press, Taylor & Francis Group.

Kim, G., Silvapulle, M. J., & Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, *51*(6), 2836–2850. https://doi.org/10.1016/j.csda.2006.10.009

Kwon, S., Ha, I. D., Shih, J.-H., & Emura, T. (2022). Flexible parametric copula modeling approaches for clustered survival data. *Pharmaceutical Statistics*, *21*(1), 69–88. https://doi.org/10.1002/pst.2153

Leemis, L. M., Shih, L.-H., & Reynertson, K. (1990). Variate generation for accelerated life and proportional hazards models with time dependent covariates. *Statistics & Probability Letters*, *10*(4), 335–339. https://doi.org/10.1016/0167-7152(90)90052-9

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation* (2nd ed). Springer.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lipsitz, S. R., Dear, K. B. G., & Zhao, L. (1994). Jackknife Estimators of Variance for Parameter Estimates from Estimating Equations with Applications to Clustered Survival Data. *Biometrics*, *50*(3), 842. https://doi.org/10.2307/2532797

Lipsitz, S. R., & Parzen, M. (1996). A Jackknife Estimator of Variance for Cox Regression for Correlated Survival Data. *Biometrics*, *52*(1), 291. https://doi.org/10.2307/2533164

Markel, P. D., DeFries, J. C., & Johnson, T. E. (1995). Use of Repeated Measures in an Analysis of Ethanol-Induced Loss of Righting Reflex in Inbred Long-Sleep and Short-Sleep Mice. *Alcoholism: Clinical and Experimental Research*, *19*(2), 299–304. https://doi.org/10.1111/j.1530-0277.1995.tb01506.x

Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed). Springer.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *44*(3), 414–422. https://doi.org/10.1111/j.2517-6161.1982.tb01222.x

Oakes, D. (1986). Semiparametric Inference in a Model for Association in Bivariate Survival Data. *Biometrika*, *73*(2), 353. https://doi.org/10.2307/2336211

Othus, M., & Li, Y. (2010). A Gaussian Copula Model for Multivariate Survival Data. *Statistics in Biosciences*, *2*(2), 154–179. https://doi.org/10.1007/s12561-010-9026-x

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: Statistical Methodology*, *79*(2), 483–505. https://doi.org/10.1111/rssb.12174

Prentice, R. L., & Cai, J. (1992). Covariance and survivor function estimation using censored multivari-

ate failure time data. *Biometrika*, *79*(3), 495–512. https://doi.org/10.1093/biomet/79.3.495

Ramsay, J. O. (1988). Monotone Regression Splines in Action. *Statistical Science*, *3*(4). https://doi.org/10.1214/ss/1177012761

Royston, P., & Parmar, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, *21*(15), 2175–2197. https://doi.org/10.1002/sim.1203

Shih, J. H., & Louis, T. A. (1995). Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics*, *51*(4), 1384. https://doi.org/10.2307/2533269

Sklar, M. (1959). *Fonctions de Répartition À N Dimensions Et Leurs Marges*. Université Paris 8.

Spiekerman, C. F., & Lin, D. Y. (1998). Marginal Regression Models for Multivariate Failure Time Data. *Journal of the American Statistical Association*, *93*(443), 1164–1175. https://doi.org/10.1080/01621459.1998.10473777

Statistical Methods in Water Resources (Techniques and Methods). (2020). [Techniques and {{Methods}}].

Tjøstheim, D., Otneim, H., & Støve, B. (2022). Local Gaussian correlation and the copula. In *Statistical Modeling Using Local Gaussian Approximation* (pp. 135–159). Elsevier. https://doi.org/10.1016/B978-0-12-815861-6.00012-2

Tressou, J. (2006). Nonparametric Modeling of the Left Censorship of Analytical Data in Food Risk Assessment. *Journal of the American Statistical Association*, *101*(476), 1377–1386. https://doi.org/10.1198/016214506000000573

A Derivation of the bivariate Archimedean likelihood

A.1 Derivation of the bivariate Archimedean likelihood

In Section 2 we have outlined a copula model for clusters of varying size under left-censoring. A general expression for the likelihood contribution of a given cluster is given in Equation (2.6) for the simplified example of bivariate clustering. We now validate the claim that Equation (2.6) reduces to Equation (2.7) for the family of Archimedean copulas specifically.

To model the full likelihood in terms of the marginal CDF, we define the following Archimedean copula for cluster i with generator φ_θ:

$$F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}) = \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}), F_2(x_{i2}|\boldsymbol{Z}_{i2})) = \varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big). \tag{A.1}$$

We now derive the likelihood contribution of cluster i for each permutation of the pair $\{\delta_{i1}, \delta_{i2}\}$ separately as shown in Equation (2.6)

$$\begin{aligned}
\{\delta_{i1} = 0, \delta_{i2} = 0\}: \quad L_i &= F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}) \\
&= \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}), F_2(x_{i2}|\boldsymbol{Z}_{i2})) \\
&= \varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big).
\end{aligned}$$

$$\begin{aligned}
\{\delta_{i1} = 1, \delta_{i2} = 0\}: \quad L_i &= \frac{\partial F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1}} \\
&= \frac{\partial \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}), F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial x_{i1}} \\
&= \frac{\partial \varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big)}{\partial x_{i1}} \\
&= \frac{\partial \varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big)}{\partial \varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))} \frac{\partial \varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))}{\partial F_1(x_{i1}|\boldsymbol{Z}_{i1})} \frac{\partial F_1(x_{i1}|\boldsymbol{Z}_{i1})}{\partial x_{i1}} \\
&= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) \frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})))}.
\end{aligned}$$

$$
\begin{aligned}
\{\delta_{i1}=0,\delta_{i2}=1\}: \quad L_i &= \frac{\partial F(x_{i1},x_{i2}|\boldsymbol{Z}_{i1},\boldsymbol{Z}_{i2})}{\partial x_{i2}} \\
&= \frac{\partial \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}),F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial x_{i2}} \\
&= \frac{\partial \varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))+\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big)}{\partial x_{i2}} \\
&= \frac{\partial \varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))+\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big)}{\partial \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}\frac{\partial \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial F_2(x_{i2}|\boldsymbol{Z}_{i2})}\frac{\partial F_2(x_{i2}|\boldsymbol{Z}_{i2})}{\partial x_{i2}} \\
&= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))+\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{\varphi'_\theta(\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}.
\end{aligned}
$$

Therefore the full likelihood for K clusters of size 2 can be written as

$$
\begin{aligned}
L = \prod_{i=1}^{K} & \varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))+\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big)^{(1-\delta_{i1})(1-\delta_{i2})} \\
& \left(\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))+\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))\frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})))}\right)^{\delta_{i1}(1-\delta_{i2})} \\
& \left(\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))+\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{\varphi'_\theta(\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}\right)^{(1-\delta_{i1})\delta_{i2}} \\
& \left(\varphi''_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))+\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))\frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})))}\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{\varphi'_\theta(\varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}\right)^{\delta_{i1}\delta_{i2}}
\end{aligned}
$$

which is completely equivalent to Equation (2.7).

A.2 Clayton copula

We consider the Clayton copula with generator $\varphi_\theta(s) = (1+\theta s)^{-1/\theta}$ and its inverse $\varphi_\theta^{-1}(s) = (s^{-\theta}-1)/\theta$. Note that we have the following derivatives of φ_θ and φ_θ^{-1}:

$\varphi'_\theta(s) = -(1+\theta s)^{-1/\theta-1}$
$\varphi''_\theta(s) = (1+\theta)(1+\theta s)^{-1/\theta-2}$
$(\varphi_\theta^{-1})'(s) = -s^{-\theta-1}$
$\varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big) = (F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta}$
$\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = -(F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta-1}$
$\varphi''_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = (1+\theta)(F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta-2}$
$(\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1})) = -F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta-1}$
$(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = -F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta-1}$

The likelihood can now be calculated term by term:

$$\{\delta_{i1}=0,\delta_{i2}=0\}: \quad L_i = (F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta}$$

$$\begin{aligned}\{\delta_{i1}=1,\delta_{i2}=0\}: \quad L_i &= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))(\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1}))f_1(x_{i1}|\boldsymbol{Z}_{i2})\\ &= -(F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta-1}\left(-F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta-1}\right)f_1(x_{i1}|\boldsymbol{Z}_{i2})\\ &= (F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta-1}F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta-1}f_1(x_{i1}|\boldsymbol{Z}_{i2})\end{aligned}$$

$$\begin{aligned}\{\delta_{i1}=0,\delta_{i2}=1\}: \quad L_i &= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2}))f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= -(F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta-1}\left(-F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta-1}\right)f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= (F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta-1}F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta-1}f_2(x_{i2}|\boldsymbol{Z}_{i2})\end{aligned}$$

$$\begin{aligned}\{\delta_{i1}=1,\delta_{i2}=1\}: \quad L_i &= \varphi''_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\\ &\quad (\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1}))(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2}))f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= (1+\theta)(F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1)^{-1/\theta-2}\\ &\quad \left(-F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta-1}\right)\left(-F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta-1}\right)f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})\end{aligned}$$

Setting $\kappa = F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta} + F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta} - 1$, the full likelihood can be expressed as

$$\begin{aligned}L = \prod_{i=1}^{K} &\left(\kappa^{-1/\theta}\right)^{(1-\delta_{i1})(1-\delta_{i2})}\\ &\left(\kappa^{-1/\theta-1}f_1(x_{i1}|\boldsymbol{Z}_{i2})F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta-1}\right)^{\delta_{i1}(1-\delta_{i2})}\\ &\left(\kappa^{-1/\theta-1}f_2(x_{i2}|\boldsymbol{Z}_{i2})F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta-1}\right)^{(1-\delta_{i1})\delta_{i1}}\\ &\left((1+\theta)\kappa^{-1/\theta-2}f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})F_1(x_{i1}|\boldsymbol{Z}_{i1})^{-\theta-1}F_2(x_{i2}|\boldsymbol{Z}_{i2})^{-\theta-1}\right)^{\delta_{i1}\delta_{i2}}\end{aligned}$$

A.3 Frank copula

We consider the Frank copula with generator $\varphi_\theta(s) = \frac{-1}{\theta}\log\big(1+e^{-s}(e^{-\theta}-1)\big)$ and its inverse $\varphi_\theta^{-1}(s) = -\log\left(\frac{e^{-s\theta}-1}{e^{-\theta}-1}\right)$ for $\theta > 0$. We have the following derivatives of φ_θ and φ_θ^{-1}, where we use $\alpha_1 = e^{-\theta F_1(x_{i1}|\boldsymbol{Z}_{i1})}-1$, $\alpha_2 = e^{-\theta F_2(x_{i2}|\boldsymbol{Z}_{i2})}-1$, $\beta = e^{-\theta}-1$ and $\gamma = \frac{\alpha_1\alpha_2}{\beta}$ to simplify the expressions:

$\varphi'_\theta(s) = \frac{1}{\theta}\frac{e^{-s}\beta}{1+e^{-s}\beta}$
$\varphi''_\theta(s) = -\frac{1}{\theta}\frac{e^{-s}\beta}{(1+e^{-s}\beta)^2}$
$(\varphi_\theta^{-1})'(s) = \theta\frac{e^{-\theta s}}{e^{-\theta s}-1}$
$\varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big) = -\frac{1}{\theta}\log(1+\gamma)$
$\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = \frac{1}{\theta}\frac{\gamma}{1+\gamma}$
$\varphi''_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = -\frac{1}{\theta}\frac{\gamma}{(1+\gamma)^2}$
$(\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1})) = \theta\frac{\alpha_1+1}{\alpha_1}$
$(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = \theta\frac{\alpha_2+1}{\alpha_2}$

The likelihood can now be calculated term by term:

$$\{\delta_{i1}=0,\delta_{i2}=0\}: \quad L_i = -\frac{1}{\theta}\log(1+\gamma).$$

$$\begin{aligned}\{\delta_{i1}=1,\delta_{i2}=0\}: \quad L_i &= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))(\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1}))f_1(x_{i1}|\boldsymbol{Z}_{i2})\\ &= \frac{1}{\theta}\frac{\gamma}{1+\gamma}\theta\frac{\alpha_1+1}{\alpha_1}f_1(x_{i1}|\boldsymbol{Z}_{i2})\\ &= \frac{\alpha_2}{\beta(1+\gamma)}(\alpha_1+1)f_1(x_{i1}|\boldsymbol{Z}_{i2}).\end{aligned}$$

$$\begin{aligned}\{\delta_{i1}=0,\delta_{i2}=1\}: \quad L_i &= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2}))f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= \frac{1}{\theta}\frac{\gamma}{1+\gamma}\theta\frac{\alpha_2+1}{\alpha_2}f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= \frac{\alpha_1}{\beta(1+\gamma)}(\alpha_2+1)f_2(x_{i2}|\boldsymbol{Z}_{i2}).\end{aligned}$$

$$\begin{aligned}\{\delta_{i1}=1,\delta_{i2}=1\}: \quad L_i &= \varphi''_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\\ &\qquad (\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1}))(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2}))f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= -\frac{1}{\theta}\frac{\gamma}{(1+\gamma)^2}\theta\frac{\alpha_1+1}{\alpha_1}\theta\frac{\alpha_2+1}{\alpha_2}f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= -\frac{\theta}{\beta(1+\gamma)^2}(\alpha_1+1)(\alpha_2+1)f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})\end{aligned}$$

The full likelihood can be expressed as

$$
\begin{aligned}
L = \prod_{i=1}^{K} & \left(-\frac{1}{\theta}\log(1+\gamma)\right)^{(1-\delta_{i1})(1-\delta_{i2})} \\
& \left(\frac{\alpha_2}{\beta(1+\gamma)}(\alpha_1+1)f_1(x_{i1}|\boldsymbol{Z}_{i2})\right)^{\delta_{i1}(1-\delta_{i2})} \\
& \left(\frac{\alpha_1}{\beta(1+\gamma)}(\alpha_2+1)f_2(x_{i2}|\boldsymbol{Z}_{i2})\right)^{(1-\delta_{i1})\delta_{i2}} \\
& \left(-\frac{\theta}{\beta(1+\gamma)^2}(\alpha_1+1)(\alpha_2+1)f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})\right)^{\delta_{i1}\delta_{i2}}.
\end{aligned}
$$

A.4 Gumbel-Hougaard copula

The Gumbel-Hougaard copula is associated with the positive stable distribution and has generator $\varphi_\theta(s) = \exp(-s^{1/\theta})$ with inverse $\varphi_\theta^{-1}(s) = (-\log s)^\theta$ for $\theta > 1$. To derive the bivariate likelihood we consider the following derivatives of φ_θ and φ_θ^{-1} with $\mu = (-\log F_1(x_{i1}|\boldsymbol{Z}_{i1}))^\theta + (-\log F_2(x_{i2}|\boldsymbol{Z}_{i2}))^\theta$ to simplify

$$\varphi'_\theta(s) = -\varphi_\theta(s)\frac{s^{1/\theta-1}}{\theta}$$
$$\varphi''_\theta(s) = \varphi_\theta(s)\left(s^{2/\theta-2} + (\theta-1)\,s^{1/\theta-2}\right)/\theta^2$$
$$(\varphi_\theta^{-1})'(s) = \frac{-\theta}{s}(-\log s)^{\theta-1}$$
$$\varphi_\theta\Big(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))\Big) = \exp\left(-\mu^{1/\theta}\right)$$
$$\varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = -\exp\left(-\mu^{1/\theta}\right)\frac{\mu^{1/\theta-1}}{\theta}$$
$$\varphi''_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) = -\frac{\exp(-\mu^{1/\theta})}{\theta^2}\left(\mu^{2/\theta-2} + (1-\theta)\mu^{1/\theta-2}\right)$$
$$(\varphi_\theta^{-1})'(s) = \frac{-\theta}{s}(-\log s)^{\theta-1}$$

The likelihood can now be calculated term by term:

$$\{\delta_{i1}=0,\delta_{i2}=0\}: \quad L_i = \exp\left(-\mu^{1/\theta}\right).$$

$$\begin{aligned}\{\delta_{i1}=1,\delta_{i2}=0\}: \quad L_i &= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))(\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1}))f_1(x_{i1}|\boldsymbol{Z}_{i2})\\ &= -\exp\left(-\mu^{1/\theta}\right)\frac{\mu^{1/\theta-1}}{\theta}\frac{-\theta}{F_1(x_{i1}|\boldsymbol{Z}_{i1})}(-\log F_1(x_{i1}|\boldsymbol{Z}_{i1}))^{\theta-1}f_1(x_{i1}|\boldsymbol{Z}_{i2})\\ &= \exp\left(-\mu^{1/\theta}\right)\mu^{1/\theta-1}\frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{F_1(x_{i1}|\boldsymbol{Z}_{i1})}(-\log F_1(x_{i1}|\boldsymbol{Z}_{i1}))^{\theta-1}\end{aligned}$$

$$\begin{aligned}\{\delta_{i1}=0,\delta_{i2}=1\}: \quad L_i &= \varphi'_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2}))f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= -\exp\left(-\mu^{1/\theta}\right)\frac{\mu^{1/\theta-1}}{\theta}\frac{-\theta}{F_2(x_{i2}|\boldsymbol{Z}_{i2})}(-\log F_2(x_{i2}|\boldsymbol{Z}_{i2}))^{\theta-1}f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= \exp\left(-\mu^{1/\theta}\right)\mu^{1/\theta-1}\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{F_2(x_{i2}|\boldsymbol{Z}_{i2})}(-\log F_2(x_{i2}|\boldsymbol{Z}_{i2}))^{\theta-1}\end{aligned}$$

$$\begin{aligned}\{\delta_{i1}=1,\delta_{i2}=1\}: \quad L_i &= \varphi''_\theta(\varphi_\theta^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})) + \varphi_\theta^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))\\ &\quad (\varphi_\theta^{-1})'(F_1(x_{i1}|\boldsymbol{Z}_{i1}))(\varphi_\theta^{-1})'(F_2(x_{i2}|\boldsymbol{Z}_{i2}))f_1(x_{i1}|\boldsymbol{Z}_{i2})f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= \frac{\exp\left(-\mu^{1/\theta}\right)}{\theta^2}\left(\mu^{2/\theta-2} + (\theta-1)\mu^{1/\theta-2}\right)\\ &\quad \frac{-\theta}{F_1(x_{i1}|\boldsymbol{Z}_{i1})}(-\log F_1(x_{i1}|\boldsymbol{Z}_{i1}))^{\theta-1}f_1(x_{i1}|\boldsymbol{Z}_{i2})\frac{-\theta}{F_2(x_{i2}|\boldsymbol{Z}_{i2})}(-\log F_2(x_{i2}|\boldsymbol{Z}_{i2}))^{\theta-1}f_2(x_{i2}|\boldsymbol{Z}_{i2})\\ &= \exp\left(-\mu^{1/\theta}\right)\left(\mu^{2/\theta-2} + (\theta-1)\mu^{1/\theta-2}\right)\\ &\quad \frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{F_1(x_{i1}|\boldsymbol{Z}_{i1})}\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{F_2(x_{i2}|\boldsymbol{Z}_{i2})}(-\log F_1(x_{i1}|\boldsymbol{Z}_{i1}))^{\theta-1}(-\log F_2(x_{i2}|\boldsymbol{Z}_{i2}))^{\theta-1}\end{aligned}$$

The full likelihood can now be written as

$$L = \prod_{i=1}^{K} \exp\left(-\mu^{1/\theta}\right)$$
$$\left(\mu^{1/\theta-1}\frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{F_1(x_{i1}|\boldsymbol{Z}_{i1})}(-\log F_1(x_{i1}|\boldsymbol{Z}_{i1}))^{\theta-1}\right)^{\delta_{i1}}\left(\mu^{1/\theta-1}\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{F_2(x_{i2}|\boldsymbol{Z}_{i2})}(-\log F_2(x_{i2}|\boldsymbol{Z}_{i2}))^{\theta-1}\right)^{\delta_{i2}}$$
$$\left((\theta-1)\mu^{-1/\theta}+1\right)^{\delta_{i1}\delta_{i2}}$$

B Derivation of the bivariate Gaussian likelihood

The full likelihood contribution of K clusters of cluster size 2 under left-censoring is given by

$$L = \prod_{i=1}^{K} F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})^{(1-\delta_{i1})(1-\delta_{i2})} \left(\frac{\partial F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1}}\right)^{\delta_{i1}(1-\delta_{i2})} \left(\frac{\partial F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i2}}\right)^{\delta_{i2}(1-\delta_{i1})} \left(\frac{\partial^2 F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1}\partial x_{i2}}\right)^{\delta_{i1}\delta_{i2}}$$

To model the full likelihood in terms of the marginal CDF, we use the Gaussian copula for cluster i as defined in Equation (2.9) where $n_i = 2$ for all i. We now derive the likelihood contribution of cluster i for each permutation of the pair $\{\delta_{i1}, \delta_{i2}\}$ separately using the same notation as in Section 2.3.

$$\begin{aligned}
\{\delta_{i1} = 0, \delta_{i2} = 0\}: \quad L_i &= F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}) \\
&= \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}), F_2(x_{i2}|\boldsymbol{Z}_{i2})) \\
&= \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})), \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))) \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))} \int_{-\infty}^{\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))} \exp\left(-\frac{s^2 + t^2 - 2\rho st}{2(1-\rho^2)}\right) dsdt.
\end{aligned}$$

$$\begin{aligned}
\{\delta_{i1} = 1, \delta_{i2} = 0\}: \quad L_i &= \frac{\partial F(x_{i1}, x_{i2}|\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})}{\partial x_{i1}} \\
&= \frac{\partial \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}), F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial x_{i1}} \\
&= \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})), \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial x_{i1}} \\
&= \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})), \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial \varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))} \frac{\partial \varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))}{\partial x_{i1}} \\
&= \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})), \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial \varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))} \frac{f_1(x_{i1})}{\varphi'(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})))} \\
&= \Psi(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})))\varphi\left(\frac{\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) - \rho\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))}{\sqrt{1-\rho^2}}\right) \frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{\varphi'(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})))} \\
&= \varphi\left(\frac{\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})) - \rho\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))}{\sqrt{1-\rho^2}}\right) f_1(x_{i1}|\boldsymbol{Z}_{i2}).
\end{aligned}$$

$$
\begin{aligned}
\{\delta_{i1}=0,\delta_{i2}=1\}: \quad L_i &= \frac{\partial F(x_{i1},x_{i2}|\boldsymbol{Z}_{i1},\boldsymbol{Z}_{i2})}{\partial x_{i2}} \\
&= \frac{\partial \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}),F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial x_{i2}} \\
&= \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial x_{i2}} \\
&= \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}\frac{\partial \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial x_{i2}} \\
&= \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{\varphi'(\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))} \\
&= \Psi(\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))\varphi\left(\frac{\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))-\rho\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\sqrt{1-\rho^2}}\right)\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{\varphi'(\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))} \\
&= \varphi\left(\frac{\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))-\rho\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\sqrt{1-\rho^2}}\right)f_2(x_{i2}|\boldsymbol{Z}_{i2}).
\end{aligned}
$$

$$
\begin{aligned}
\{\delta_{i1}=1,\delta_{i2}=1\}: \quad L_i &= \frac{\partial^2 F(x_{i1},x_{i2}|\boldsymbol{Z}_{i1},\boldsymbol{Z}_{i2})}{\partial x_{i1}\partial x_{i2}} \\
&= \frac{\partial^2 \mathcal{C}(F_1(x_{i1}|\boldsymbol{Z}_{i1}),F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial x_{i1}\partial x_{i2}} \\
&= \frac{\partial^2 \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial x_{i1}\partial x_{i2}} \\
&= \frac{\partial^2 \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial \varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))\partial \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}\frac{\partial \varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))}{\partial x_{i1}}\frac{\partial \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\partial x_{i2}} \\
&= \frac{\partial^2 \boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{\partial \varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))\partial \varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}\frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{\varphi'(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})))}\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{\varphi'(\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))} \\
&= \boldsymbol{\Psi}^{(\mathbf{2})}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))\frac{f_1(x_{i1}|\boldsymbol{Z}_{i2})}{F_1(x_{i1}|\boldsymbol{Z}_{i1})}\frac{f_2(x_{i2}|\boldsymbol{Z}_{i2})}{F_2(x_{i2}|\boldsymbol{Z}_{i2})}.
\end{aligned}
$$

Therefore the full likelihood for K clusters of size 2 can be written as

$$
\begin{aligned}
L = \prod_{i=1}^{K} &\boldsymbol{\Phi}_{\boldsymbol{\rho}}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))^{(1-\delta_{i1})(1-\delta_{i2})} \\
&\varphi\left(\frac{\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))-\rho\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))}{\sqrt{1-\rho^2}}\right)^{\delta_{i1}(1-\delta_{i2})}\varphi\left(\frac{\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1}))-\rho\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2}))}{\sqrt{1-\rho^2}}\right)^{(1-\delta_{i1})\delta_{i2}} \\
&\left(\frac{\boldsymbol{\Psi}^{(\mathbf{2})}(\varphi^{-1}(F_1(x_{i1}|\boldsymbol{Z}_{i1})),\varphi^{-1}(F_2(x_{i2}|\boldsymbol{Z}_{i2})))}{F_1(x_{i1}|\boldsymbol{Z}_{i1})F_2(x_{i2}|\boldsymbol{Z}_{i2})}\right)^{\delta_{i1}\delta_{i2}} f_1(x_{i1}|\boldsymbol{Z}_{i2})^{\delta_{i1}}f_2(x_{i2}|\boldsymbol{Z}_{i2})^{\delta_{i2}}.
\end{aligned}
$$

C Proof of Theorem 2.1

Let $\boldsymbol{\beta_{j0}}$ denote the vector of the true value of the marginal parameters $\boldsymbol{\beta_j}$ for margin j with $j = 1, \ldots, n$ and $\overline{\boldsymbol{\beta_j}}$ the estimated values of the marginal parameters under the working independence assumption. We now expand the score functions $\boldsymbol{U}^*_{\boldsymbol{\beta}_j}$ of the first step in the two-stage procedure around the true vector $\boldsymbol{\beta_{j0}}$ using a first-order Taylor series approximation and evaluate it in the estimate $\overline{\boldsymbol{\beta_j}}$

$$\boldsymbol{U}^*_{\boldsymbol{\beta}_j}(\overline{\boldsymbol{\beta_j}}) = \boldsymbol{U}^*_{\boldsymbol{\beta}_j}(\boldsymbol{\beta_{j0}}) - \frac{-\partial \boldsymbol{U}^*_{\boldsymbol{\beta}_j}(\boldsymbol{\beta_{j0}})}{\partial \boldsymbol{\beta_j}}(\overline{\boldsymbol{\beta_j}} - \boldsymbol{\beta_{j0}}) + \mathcal{O}(\sqrt{K}) = 0$$

where $\frac{-\partial \boldsymbol{U}^*_{\boldsymbol{\beta}_j}(\boldsymbol{\beta_{j0}})}{\partial \boldsymbol{\beta_j}}$ is a partial derivative of the score function evaluated in $\boldsymbol{\beta_{j0}}$. Similarly for step two (estimation of the association parameter θ) we expand the score function U_θ of the second stage around the true value θ_0 and evaluate it in $\overline{\theta}$

$$U_\theta(\overline{\theta}) = U_\theta(\theta_0) - \frac{-\partial U_\theta(\theta_0)}{\partial \boldsymbol{\beta_1}}(\overline{\boldsymbol{\beta_1}} - \boldsymbol{\beta_{10}}) - \ldots - \frac{-\partial U_\theta(\theta_0)}{\partial \boldsymbol{\beta_n}}(\overline{\boldsymbol{\beta_n}} - \boldsymbol{\beta_{n0}}) - \frac{-\partial U_\theta(\theta_0)}{\partial \theta}(\overline{\theta} - \theta_0) + \mathcal{O}(\sqrt{K}) = 0.$$

Using the law of large numbers we have, as $K \to \infty$,

$$\begin{aligned}
\frac{-\partial \boldsymbol{U}^*_{\boldsymbol{\beta}_j}(\boldsymbol{\beta_{j0}})}{\partial \boldsymbol{\beta_j}} &\to K\boldsymbol{I}^*_{\boldsymbol{\beta}_j\boldsymbol{\beta}_j} \\
\frac{-\partial U_\theta(\theta_0)}{\partial \boldsymbol{\beta_j}} &\to K\boldsymbol{I}^*_{\theta\boldsymbol{\beta}_j} \\
\frac{-\partial U_\theta(\theta_0)}{\partial \theta} &\to KI_{\theta\theta}
\end{aligned}$$

which implies

$$\frac{1}{\sqrt{K}}\begin{pmatrix} \boldsymbol{U}^*_{\boldsymbol{\beta}_1}(\boldsymbol{\beta_{10}}) \\ \vdots \\ \boldsymbol{U}^*_{\boldsymbol{\beta}_n}(\boldsymbol{\beta_{n0}}) \\ U_\theta(\theta_0) \end{pmatrix} = \sqrt{K}\begin{pmatrix} \boldsymbol{I}^*_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}^*_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{I}^*_{\theta\boldsymbol{\beta}_1} & \boldsymbol{I}^*_{\theta\boldsymbol{\beta}_2} & \cdots & I_{\theta\theta} \end{pmatrix}\begin{pmatrix} \overline{\boldsymbol{\beta_1}} - \boldsymbol{\beta_{10}} \\ \vdots \\ \overline{\boldsymbol{\beta_n}} - \boldsymbol{\beta_{n0}} \\ \overline{\theta} - \theta_0 \end{pmatrix} \tag{C.1}$$

On the other hand we have by the Central Limit Theorem that the left hand side of Equation (C.1) converges to a multivariate normal distribution with mean $\boldsymbol{0}$ and variance-covariance matrix

$$\begin{pmatrix} \boldsymbol{I}^*_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} & \boldsymbol{I}^*_{\boldsymbol{\beta}_1\boldsymbol{\beta}_2} & \cdots & \boldsymbol{I}^*_{\boldsymbol{\beta}_1\boldsymbol{\beta}_n} & \boldsymbol{0} \\ \boldsymbol{I}^*_{\boldsymbol{\beta}_2\boldsymbol{\beta}_1} & \boldsymbol{I}^*_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2} & \cdots & \boldsymbol{I}^*_{\boldsymbol{\beta}_2\boldsymbol{\beta}_n} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{I}^*_{\boldsymbol{\beta}_n\boldsymbol{\beta}_1} & \boldsymbol{I}^*_{\boldsymbol{\beta}_n\boldsymbol{\beta}_2} & \cdots & \boldsymbol{I}^*_{\boldsymbol{\beta}_n\boldsymbol{\beta}_n} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & I_{\theta\theta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V} & \boldsymbol{0} \\ \boldsymbol{0} & I_{\theta\theta} \end{pmatrix}.$$

Therefore we have that $\sqrt{K}(\overline{\boldsymbol{\beta_1}} - \boldsymbol{\beta_{10}}, \ldots, \overline{\boldsymbol{\beta_n}} - \boldsymbol{\beta_{n0}}, \overline{\theta} - \theta_0)^T$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix

$$\begin{pmatrix} \boldsymbol{I}^*_{\boldsymbol{\beta\beta}} & \mathbf{0} \\ \boldsymbol{I}^*_{\boldsymbol{\theta\beta}} & I_{\theta\theta} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{V} & \mathbf{0} \\ \mathbf{0} & I_{\theta\theta} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}^*_{\boldsymbol{\beta\beta}} & \mathbf{0} \\ \boldsymbol{I}^*_{\boldsymbol{\theta\beta}} & I_{\theta\theta} \end{pmatrix}^{-1^T}.$$

Performing the matrix multplication and taking the element in the lower right corner of the resulting variance-covariance matrix will give the desired result.

D Supplementary tables

Table D.1: Overview of the estimates for the regression parameters for $K = 50$ under different copula models. All estimation procedures discussed are used for comparison. The standard error and coverage probability are given in brackets.

		30% censoring					
		β_1			β_2		
	θ	Maximum Likelihood	Two-stage	Spline	Maximum Likelihood	Two-stage	Spline
Clayton	0.2	1.023 (0.528; 94.5%)	1.037 (0.396; 93.0%)	1.067 (0.411; 93.0%)	1.043 (0.527; 94.1%)	1.032 (0.393; 91.9%)	1.060 (0.412; 94.4%)
	0.5	1.058 (0.521; 92.5%)	1.068 (0.395; 92.5%)	1.030 (0.403; 94.3%)	1.055 (0.520; 95.3%)	1.057 (0.394; 93.0%)	1.031 (0.404; 93.9%)
	1	1.069 (0.498; 94.6%)	1.047 (0.395; 94.0%)	1.064 (0.387; 95.3%)	1.064 (0.500; 93.1%)	1.044 (0.394; 93.0%)	1.072 (0.388; 94.7%)
	1.5	1.040 (0.471; 94.0%)	1.044 (0.393; 92.6%)	1.058 (0.368; 94.7%)	1.055 (0.473; 94.6%)	1.043 (0.395; 93.8%)	1.054 (0.368; 94.5%)
	2	1.061 (0.446; 95.3%)	1.042 (0.394; 93.1%)	1.038 (0.347; 94.0%)	1.064 (0.445; 93.2%)	1.048 (0.394; 92.6%)	1.041 (0.348; 94.1%)
Frank	0.82	1.046 (0.528; 95.6%)	1.055 (0.394; 93.6%)	1.052 (0.410; 93.8%)	1.045 (0.528; 93.8%)	1.063 (0.393; 93.3%)	1.052 (0.411; 94.3%)
	1.86	1.046 (0.515; 94.8%)	1.045 (0.394; 93.8%)	1.060 (0.398; 94.6%)	1.039 (0.516; 93.8%)	1.069 (0.393; 93.6%)	1.054 (0.399; 94.5%)
	3.26	1.032 (0.482; 94.0%)	1.043 (0.396; 93.3%)	1.059 (0.372; 93.4%)	1.080 (0.485; 92.6%)	1.060 (0.395; 94.0%)	1.064 (0.371; 93.7%)
	4.6	1.033 (0.445; 93.4%)	1.050 (0.393; 93.4%)	1.055 (0.342; 94.8%)	1.062 (0.447; 94.6%)	1.054 (0.392; 94.3%)	1.039 (0.344; 93.6%)
	5.8	1.044 (0.420; 93.6%)	1.030 (0.392; 94.6%)	1.026 (0.318; 94.0%)	1.055 (0.417; 93.8%)	1.025 (0.393; 94.9%)	1.040 (0.316; 94.3%)
Gumbel	1.1	1.036 (0.513; 92.3%)	1.049 (0.392; 91.9%)	1.063 (0.404; 92.1%)	1.067 (0.516; 93.6%)	1.038 (0.394; 92.0%)	1.067 (0.403; 92.3%)
	1.25	1.025 (0.483; 93.8%)	1.065 (0.393; 92.6%)	1.062 (0.373; 95.0%)	1.060 (0.483; 94.1%)	1.056 (0.394; 93.7%)	1.081 (0.375; 94.6%)
	1.49	1.041 (0.431; 93.2%)	1.042 (0.393; 93.9%)	1.061 (0.332; 93.4%)	1.048 (0.430; 94.4%)	1.067 (0.394; 94.7%)	1.066 (0.333; 94.2%)
	1.75	1.054 (0.381; 93.8%)	1.060 (0.396; 93.2%)	1.057 (0.296; 94.1%)	1.041 (0.381; 92.6%)	1.049 (0.393; 92.4%)	1.058 (0.297; 93.5%)
	2	1.049 (0.346; 93.8%)	1.047 (0.390; 94.8%)	1.049 (0.269; 93.5%)	1.077 (0.345; 92.0%)	1.036 (0.392; 93.8%)	1.043 (0.268; 94.5%)
Gaussian	0.15	1.041 (0.525; 94.3%)	1.073 (0.393; 94.0%)	1.041 (0.411; 93.8%)	1.081 (0.527; 93.1%)	1.062 (0.394; 92.9%)	1.046 (0.409; 94.8%)
	0.3	1.053 (0.512; 93.6%)	1.055 (0.393; 92.1%)	1.048 (0.396; 93.4%)	1.050 (0.511; 94.9%)	1.039 (0.391; 93.7%)	1.050 (0.396; 95.6%)
	0.5	1.044 (0.474; 93.5%)	1.051 (0.393; 92.8%)	1.075 (0.364; 94.5%)	1.038 (0.473; 94.0%)	1.045 (0.393; 92.2%)	1.062 (0.364; 94.9%)
	0.65	1.035 (0.428; 94.0%)	1.047 (0.393; 93.9%)	1.055 (0.334; 94.1%)	1.061 (0.430; 94.7%)	1.050 (0.395; 93.0%)	1.048 (0.333; 93.3%)
	0.71	1.040 (0.404; 94.2%)	1.037 (0.392; 94.0%)	1.064 (0.315; 94.7%)	1.052 (0.405; 93.6%)	1.038 (0.395; 94.9%)	1.049 (0.315; 93.7%)
		50% censoring					
		β_1			β_2		
	θ	Maximum Likelihood	Two-stage	Spline	Maximum Likelihood	Two-stage	Spline
Clayton	0.2	1.070 (0.562; 94.0%)	1.068 (0.410; 92.5%)	1.048 (0.444; 94.4%)	1.097 (0.560; 93.6%)	1.062 (0.408; 92.8%)	1.049 (0.444; 94.1%)
	0.5	1.100 (0.559; 93.3%)	1.065 (0.407; 92.2%)	1.057 (0.427; 93.7%)	1.068 (0.557; 92.4%)	1.063 (0.410; 92.9%)	1.066 (0.430; 92.5%)
	1	1.066 (0.541; 94.0%)	1.037 (0.409; 92.9%)	1.063 (0.428; 93.8%)	1.095 (0.541; 92.7%)	1.045 (0.405; 91.8%)	1.064 (0.419; 93.8%)
	1.5	1.083 (0.526; 93.3%)	1.038 (0.406; 92.7%)	1.072 (0.406; 94.4%)	1.051 (0.524; 92.8%)	1.043 (0.404; 93.1%)	1.086 (0.407; 94.0%)
	2	1.072 (0.509; 93.9%)	1.040 (0.408; 92.5%)	1.032 (0.391; 93.4%)	1.069 (0.510; 93.9%)	1.057 (0.407; 93.3%)	1.048 (0.390; 93.5%)
Frank	0.82	1.058 (0.558; 93.2%)	1.065 (0.409; 92.2%)	1.064 (0.435; 94.8%)	1.043 (0.554; 94.7%)	1.075 (0.410; 91.5%)	1.072 (0.436; 93.7%)
	1.86	1.043 (0.548; 93.3%)	1.070 (0.411; 92.9%)	1.068 (0.425; 94.3%)	1.056 (0.549; 93.5%)	1.082 (0.409; 92.4%)	1.088 (0.425; 94.1%)
	3.26	1.074 (0.523; 94.7%)	1.072 (0.411; 92.9%)	1.087 (0.406; 92.9%)	1.070 (0.524; 93.7%)	1.077 (0.412; 92.3%)	1.088 (0.406; 93.4%)
	4.6	1.080 (0.494; 93.3%)	1.083 (0.412; 93.2%)	1.061 (0.381; 94.7%)	1.068 (0.497; 94.6%)	1.059 (0.409; 93.7%)	1.047 (0.382; 94.6%)
	5.8	1.078 (0.464; 95.3%)	1.071 (0.412; 93.8%)	1.063 (0.363; 94.7%)	1.056 (0.467; 92.6%)	1.064 (0.409; 92.6%)	1.047 (0.362; 94.8%)
Gumbel	1.1	1.059 (0.546; 92.3%)	1.051 (0.410; 91.5%)	1.051 (0.427; 92.9%)	1.042 (0.548; 91.5%)	1.061 (0.413; 92.6%)	1.054 (0.427; 93.7%)
	1.25	1.097 (0.519; 94.0%)	1.062 (0.410; 93.3%)	1.063 (0.393; 92.6%)	1.072 (0.515; 94.6%)	1.068 (0.412; 92.5%)	1.078 (0.396; 92.9%)
	1.49	1.109 (0.467; 92.8%)	1.063 (0.408; 94.3%)	1.089 (0.360; 92.7%)	1.085 (0.466; 91.3%)	1.079 (0.412; 94.2%)	1.089 (0.358; 92.0%)
	1.75	1.076 (0.414; 93.7%)	1.058 (0.411; 91.4%)	1.096 (0.323; 91.3%)	1.046 (0.413; 92.3%)	1.062 (0.411; 91.5%)	1.102 (0.323; 91.2%)
	2	1.069 (0.374; 93.3%)	1.067 (0.409; 93.6%)	1.097 (0.297; 93.2%)	1.076 (0.373; 93.0%)	1.078 (0.411; 94.1%)	1.100 (0.301; 92.6%)
Gaussian	0.15	1.034 (0.551; 93.2%)	1.062 (0.411; 92.5%)	1.064 (0.432; 94.5%)	1.046 (0.559; 94.2%)	1.067 (0.411; 93.0%)	1.088 (0.435; 94.7%)
	0.3	1.049 (0.542; 93.8%)	1.068 (0.414; 92.8%)	1.051 (0.418; 92.3%)	1.062 (0.547; 92.2%)	1.048 (0.411; 93.3%)	1.065 (0.420; 92.7%)
	0.5	1.076 (0.514; 93.7%)	1.059 (0.414; 91.7%)	1.075 (0.399; 90.4%)	1.046 (0.511; 94.0%)	1.070 (0.411; 91.5%)	1.069 (0.397; 91.3%)
	0.65	1.062 (0.465; 93.5%)	1.066 (0.412; 93.7%)	1.058 (0.366; 95.4%)	1.073 (0.469; 93.3%)	1.067 (0.412; 93.3%)	1.054 (0.367; 94.8%)
	0.71	1.042 (0.443; 93.2%)	1.078 (0.412; 92.6%)	1.071 (0.344; 94.0%)	1.053 (0.444; 92.0%)	1.079 (0.412; 93.8%)	1.071 (0.347; 92.9%)

Table D.2: Overview of the estimates for the regression parameters for $K = 200$ under different copula models. All estimation procedures discussed are used for comparison. The standard error and coverage probability are given in brackets.

		30% censoring					
		β_1			β_2		
	θ	Maximum Likelihood	Two-stage	Spline	Maximum Likelihood	Two-stage	Spline
Clayton	0.2	1.006 (0.256; 94.4%)	1.010 (0.199; 94.9%)	1.012 (0.203; 95.2%)	1.010 (0.256; 96.1%)	1.013 (0.199; 94.9%)	1.001 (0.201; 94.1%)
	0.5	1.001 (0.251; 94.6%)	1.012 (0.200; 93.0%)	1.011 (0.197; 94.7%)	1.012 (0.252; 96.4%)	1.006 (0.199; 94.4%)	1.011 (0.197; 94.3%)
	1	1.014 (0.241; 94.9%)	1.006 (0.198; 94.5%)	1.023 (0.188; 93.7%)	1.013 (0.241; 94.6%)	0.997 (0.199; 94.5%)	1.020 (0.189; 94.5%)
	1.5	1.011 (0.230; 94.6%)	1.009 (0.199; 95.6%)	1.008 (0.179; 95.1%)	1.012 (0.230; 96.4%)	1.010 (0.199; 94.5%)	1.012 (0.179; 94.6%)
	2	1.012 (0.217; 93.9%)	1.017 (0.200; 95.1%)	0.997 (0.169; 95.2%)	1.015 (0.218; 95.3%)	1.018 (0.199; 94.8%)	1.002 (0.169; 94.4%)
Frank	0.82	1.010 (0.256; 95.1%)	1.002 (0.198; 95.9%)	1.009 (0.200; 94.9%)	1.014 (0.255; 94.3%)	1.005 (0.198; 94.4%)	1.010 (0.200; 94.3%)
	1.86	1.013 (0.248; 95.7%)	1.007 (0.199; 94.4%)	1.021 (0.193; 95.2%)	1.015 (0.249; 95.5%)	1.009 (0.199; 94.6%)	1.019 (0.193; 94.7%)
	3.26	1.012 (0.233; 94.9%)	1.014 (0.199; 95.6%)	1.003 (0.180; 93.8%)	1.010 (0.233; 95.1%)	1.022 (0.199; 94.7%)	1.014 (0.181; 94.5%)
	4.6	1.013 (0.216; 94.0%)	1.015 (0.199; 95.1%)	1.011 (0.167; 94.9%)	1.020 (0.216; 94.8%)	1.015 (0.199; 95.3%)	1.013 (0.167; 94.6%)
	5.8	1.005 (0.200; 93.8%)	1.015 (0.199; 94.2%)	1.008 (0.155; 95.0%)	1.008 (0.200; 94.9%)	1.011 (0.199; 94.9%)	1.013 (0.155; 95.7%)
Gumbel	1.1	1.010 (0.249; 94.8%)	1.008 (0.200; 94.4%)	1.020 (0.195; 94.5%)	1.012 (0.249; 93.6%)	1.007 (0.199; 94.8%)	1.011 (0.195; 94.7%)
	1.25	1.016 (0.233; 94.6%)	1.006 (0.199; 96.1%)	1.019 (0.181; 95.0%)	1.002 (0.231; 93.6%)	1.003 (0.199; 95.5%)	1.018 (0.181; 94.6%)
	1.49	1.014 (0.204; 93.2%)	1.014 (0.200; 95.0%)	1.012 (0.161; 94.9%)	1.011 (0.204; 95.2%)	1.015 (0.199; 95.2%)	1.014 (0.160; 93.9%)
	1.75	1.019 (0.180; 95.7%)	1.008 (0.199; 95.2%)	1.017 (0.142; 93.5%)	1.014 (0.180; 94.0%)	1.012 (0.199; 95.4%)	1.008 (0.142; 92.9%)
	2	1.007 (0.161; 94.4%)	1.009 (0.199; 95.5%)	1.016 (0.129; 94.1%)	1.017 (0.161; 94.7%)	1.010 (0.198; 95.7%)	1.018 (0.129; 92.9%)
Gaussian	0.15	1.012 (0.255; 94.5%)	1.015 (0.199; 94.8%)	1.006 (0.199; 95.2%)	1.006 (0.254; 95.1%)	1.008 (0.199; 94.5%)	1.009 (0.200; 94.0%)
	0.3	1.017 (0.248; 95.4%)	1.021 (0.199; 95.5%)	1.008 (0.193; 94.2%)	1.012 (0.247; 95.6%)	1.024 (0.200; 95.8%)	1.011 (0.193; 94.6%)
	0.5	1.005 (0.230; 92.9%)	1.004 (0.199; 95.7%)	1.008 (0.179; 96.2%)	1.022 (0.230; 94.0%)	1.010 (0.199; 94.3%)	1.008 (0.178; 95.4%)
	0.65	1.012 (0.207; 94.3%)	1.015 (0.200; 94.5%)	1.013 (0.161; 94.8%)	1.011 (0.208; 95.4%)	1.020 (0.199; 94.9%)	1.010 (0.161; 95.2%)
	0.71	1.011 (0.196; 94.4%)	1.021 (0.199; 93.9%)	1.024 (0.152; 94.2%)	1.009 (0.195; 95.0%)	1.017 (0.198; 95.9%)	1.019 (0.153; 95.2%)
		50% censoring					
		β_1			β_2		
	θ	Maximum Likelihood	Two-stage	Spline	Maximum Likelihood	Two-stage	Spline
Clayton	0.2	1.015 (0.268; 95.2%)	1.013 (0.206; 93.8%)	1.024 (0.208; 94.3%)	1.015 (0.268; 94.9%)	1.007 (0.206; 94.6%)	1.021 (0.209; 94.4%)
	0.5	1.013 (0.264; 96.0%)	1.006 (0.206; 95.2%)	1.010 (0.209; 94.5%)	1.028 (0.265; 93.2%)	1.015 (0.205; 95.0%)	1.008 (0.209; 93.9%)
	1	1.009 (0.259; 94.7%)	1.012 (0.205; 94.3%)	1.022 (0.203; 94.1%)	1.014 (0.259; 95.8%)	1.011 (0.205; 93.9%)	1.024 (0.203; 94.8%)
	1.5	1.009 (0.251; 94.4%)	1.014 (0.206; 94.4%)	1.017 (0.197; 95.0%)	0.999 (0.251; 95.4%)	1.012 (0.206; 95.0%)	1.014 (0.197; 95.0%)
	2	1.013 (0.243; 94.7%)	1.009 (0.206; 94.3%)	1.022 (0.190; 94.9%)	1.007 (0.243; 94.3%)	1.013 (0.206; 94.5%)	1.025 (0.190; 94.7%)
Frank	0.82	1.017 (0.267; 94.9%)	1.021 (0.207; 94.4%)	1.022 (0.211; 94.3%)	1.026 (0.268; 94.7%)	1.024 (0.208; 93.3%)	1.014 (0.210; 95.0%)
	1.86	1.018 (0.263; 93.5%)	1.024 (0.208; 95.1%)	1.028 (0.205; 95.2%)	1.024 (0.262; 94.7%)	1.029 (0.208; 94.5%)	1.024 (0.205; 95.2%)
	3.26	1.005 (0.251; 95.7%)	1.002 (0.208; 94.6%)	1.006 (0.195; 95.2%)	1.011 (0.251; 94.8%)	1.001 (0.207; 94.5%)	1.013 (0.195; 95.5%)
	4.6	1.007 (0.238; 95.0%)	1.011 (0.207; 94.8%)	1.018 (0.185; 94.3%)	1.022 (0.237; 93.3%)	1.009 (0.207; 94.2%)	1.011 (0.184; 95.8%)
	5.8	1.016 (0.225; 95.2%)	1.021 (0.208; 95.9%)	1.004 (0.174; 94.4%)	1.008 (0.225; 94.9%)	1.023 (0.208; 94.7%)	1.004 (0.174; 94.9%)
Gumbel	1.1	1.023 (0.260; 95.4%)	1.015 (0.207; 94.0%)	1.027 (0.205; 94.4%)	1.012 (0.260; 95.7%)	1.013 (0.207; 94.9%)	1.027 (0.204; 93.8%)
	1.25	1.006 (0.243; 95.3%)	1.014 (0.207; 94.9%)	1.022 (0.190; 93.9%)	1.009 (0.243; 94.9%)	1.009 (0.207; 95.1%)	1.023 (0.190; 94.2%)
	1.49	1.005 (0.216; 94.3%)	1.011 (0.208; 95.5%)	1.017 (0.170; 94.3%)	1.011 (0.216; 94.5%)	1.019 (0.208; 95.0%)	1.021 (0.170; 94.9%)
	1.75	1.012 (0.196; 94.2%)	1.020 (0.207; 94.8%)	1.018 (0.152; 94.1%)	1.007 (0.196; 94.3%)	1.025 (0.207; 93.0%)	1.017 (0.152; 95.2%)
	2	1.018 (0.174; 94.7%)	1.017 (0.207; 95.0%)	1.023 (0.139; 93.1%)	1.015 (0.174; 94.5%)	1.027 (0.207; 95.4%)	1.020 (0.139; 94.2%)
Gaussian	0.15	1.018 (0.267; 95.1%)	1.013 (0.208; 94.5%)	1.074 (0.429; 94.3%)	1.011 (0.266; 93.6%)	1.019 (0.208; 94.3%)	1.074 (0.429; 93.1%)
	0.3	1.015 (0.261; 95.0%)	1.006 (0.207; 94.1%)	1.076 (0.423; 95.0%)	1.016 (0.261; 94.9%)	1.006 (0.207; 94.1%)	1.094 (0.425; 95.2%)
	0.5	1.022 (0.245; 93.6%)	1.022 (0.208; 94.1%)	1.085 (0.398; 92.2%)	0.999 (0.245; 94.2%)	1.008 (0.207; 96.1%)	1.079 (0.396; 94.2%)
	0.65	1.005 (0.224; 94.9%)	1.013 (0.207; 94.3%)	1.035 (0.359; 93.7%)	1.007 (0.225; 93.9%)	1.017 (0.207; 94.7%)	1.042 (0.360; 95.2%)
	0.71	1.022 (0.213; 95.5%)	1.021 (0.207; 95.1%)	1.055 (0.347; 93.9%)	1.009 (0.214; 95.1%)	1.012 (0.208; 96.4%)	1.051 (0.348; 93.8%)

Table D.3: Overview of the estimates for the regression parameters for $K = 500$ under different copula models. All estimation procedures discussed are used for comparison. Note that a Gaussion copula model is not considered due to the computation requirements involved in the one-stage procedure under repeated sampling. The standard error and coverage probability are given in brackets.

		30% censoring					
		β_1			β_2		
	θ	Maximum Likelihood	Two-stage	Spline	Maximum Likelihood	Two-stage	Spline
Clayton	0.2	0.991 (0.160; 93.7%)	1.007 (0.126; 94.5%)	1.004 (0.128; 93.0%)	0.996 (0.161; 95.9%)	1.011 (0.126; 95.0%)	1.004 (0.128; 94.3%)
	0.5	1.010 (0.158; 94.1%)	1.005 (0.127; 95.4%)	1.010 (0.124; 93.9%)	0.999 (0.158; 95.7%)	1.006 (0.126; 94.6%)	1.010 (0.124; 94.2%)
	1	1.003 (0.152; 93.9%)	1.000 (0.126; 95.9%)	0.999 (0.118; 94.7%)	1.002 (0.152; 95.6%)	0.999 (0.126; 94.9%)	1.003 (0.118; 95.3%)
	1.5	1.004 (0.144; 94.9%)	1.012 (0.126; 93.7%)	0.986 (0.112; 93.0%)	1.004 (0.144; 94.9%)	1.010 (0.126; 95.3%)	0.989 (0.112; 94.5%)
	2	1.004 (0.137; 95.4%)	1.005 (0.126; 95.8%)	0.998 (0.106; 94.4%)	1.002 (0.137; 94.3%)	1.006 (0.126; 96.1%)	1.001 (0.106; 94.4%)
Frank	0.82	1.013 (0.160; 96.0%)	1.004 (0.126; 93.9%)	1.000 (0.126; 93.5%)	1.002 (0.161; 93.6%)	1.001 (0.126; 95.1%)	0.998 (0.126; 92.9%)
	1.86	1.005 (0.156; 94.5%)	1.007 (0.126; 94.2%)	1.003 (0.122; 93.9%)	1.005 (0.156; 95.0%)	1.006 (0.126; 95.5%)	1.004 (0.122; 92.8%)
	3.26	1.009 (0.147; 94.7%)	1.010 (0.126; 94.7%)	0.997 (0.114; 93.6%)	1.002 (0.147; 95.4%)	1.007 (0.126; 94.2%)	0.999 (0.114; 92.9%)
	4.6	0.999 (0.136; 95.3%)	1.004 (0.126; 95.4%)	1.003 (0.105; 94.1%)	1.006 (0.136; 95.3%)	1.006 (0.126; 95.1%)	1.006 (0.105; 94.9%)
	5.8	1.001 (0.126; 94.9%)	1.005 (0.126; 96.5%)	0.998 (0.097; 94.5%)	1.002 (0.126; 94.5%)	1.006 (0.126; 96%)	1.004 (0.097; 95.5%)
Gumbel	1.1	1.011 (0.157; 94.8%)	1.005 (0.126; 94.1%)	0.999 (0.123; 93.9%)	1.008 (0.157; 95.1%)	1.005 (0.126; 94.4%)	1.002 (0.123; 94.3%)
	1.25	1.002 (0.145; 94.3%)	1.007 (0.126; 95.0%)	1.008 (0.113; 93.6%)	1.012 (0.145; 93.8%)	1.003 (0.126; 93.9%)	1.005 (0.113; 93.7%)
	1.49	1.002 (0.128; 94.3%)	1.004 (0.127; 95.5%)	1.003 (0.101; 95.0%)	1.007 (0.128; 96.6%)	1.007 (0.126; 95%)	1.002 (0.101; 93.3%)
	1.75	1.005 (0.112; 96.0%)	1.010 (0.126; 95.0%)	1.004 (0.089; 95.0%)	1.004 (0.113; 95.0%)	1.008 (0.126; 95.0%)	1.004 (0.089; 95.6%)
	2	1.000 (0.101; 96.7%)	1.006 (0.126; 94.4%)	1.009 (0.080; 94.6%)	1.011 (0.101; 95.0%)	1.006 (0.126; 95.3%)	1.008 (0.080; 94.7%)
		50% censoring					
		β_1			β_2		
	θ	Maximum Likelihood	Two-stage	Spline	Maximum Likelihood	Two-stage	Spline
Clayton	0.2	0.994 (0.168; 94.7%)	1.001 (0.130; 94.3%)	1.010 (0.133; 94.4%)	1.007 (0.168; 95.9%)	1.001 (0.130; 94.9%)	1.014 (0.133; 94.1%)
	0.5	1.001 (0.166; 94.9%)	1.009 (0.130; 95.8%)	1.014 (0.131; 94.6%)	1.008 (0.167; 95.5%)	1.008 (0.131; 95.3%)	1.009 (0.131; 94.5%)
	1	1.002 (0.162; 95.3%)	1.006 (0.131; 94.9%)	1.009 (0.128; 94.5%)	1.007 (0.162; 95.6%)	1.003 (0.130; 94.2%)	1.008 (0.128; 94.8%)
	1.5	1.003 (0.158; 94.7%)	1.005 (0.130; 94.9%)	1.015 (0.124; 92.9%)	1.009 (0.157; 95.2%)	1.009 (0.130; 95.2%)	1.012 (0.124; 93.6%)
	2	1.009 (0.153; 94.8%)	1.014 (0.130; 95.7%)	1.007 (0.119; 94.8%)	1.005 (0.153; 95.3%)	1.014 (0.131; 95.8%)	1.006 (0.119; 94.4%)
Frank	0.82	1.006 (0.167; 96.0%)	1.005 (0.132; 94.5%)	1.012 (0.132; 94.1%)	1.011 (0.167; 95.0%)	1.003 (0.131; 95.6%)	1.006 (0.132; 93.9%)
	1.86	1.008 (0.164; 95.0%)	1.000 (0.131; 94.6%)	1.006 (0.129; 94.5%)	1.008 (0.164; 95.2%)	1.002 (0.132; 95.2%)	1.001 (0.129; 96.0%)
	3.26	1.008 (0.157; 94.1%)	1.009 (0.132; 94.0%)	1.002 (0.123; 96.0%)	1.006 (0.158; 94.6%)	1.011 (0.132; 94.8%)	1.003 (0.123; 93.7%)
	4.6	1.000 (0.149; 95.5%)	1.010 (0.131; 95.6%)	1.008 (0.116; 94.7%)	1.009 (0.149; 95.2%)	1.006 (0.131; 95.4%)	1.007 (0.116; 94.6%)
	5.8	1.006 (0.141; 94.3%)	0.997 (0.131; 95.4%)	1.000 (0.109; 94.7%)	1.001 (0.140; 94.1%)	1.000 (0.131; 96.4%)	0.998 (0.109; 95.0%)
Gumbel	1.1	1.005 (0.163; 95.5%)	1.007 (0.132; 94.2%)	1.003 (0.128; 94.5%)	1.008 (0.163; 94.7%)	1.009 (0.132; 94.0%)	1.003 (0.128; 94.5%)
	1.25	1.008 (0.152; 95.0%)	1.006 (0.132; 94.7%)	1.008 (0.119; 94.5%)	1.002 (0.152; 93.3%)	1.003 (0.132; 94.5%)	1.009 (0.119; 95.6%)
	1.49	1.010 (0.135; 95.9%)	1.000 (0.131; 95.4%)	1.002 (0.106; 95.8%)	1.012 (0.135; 95.3%)	1.003 (0.132; 95.2%)	1.002 (0.107; 96.2%)
	1.75	1.013 (0.122; 95.0%)	1.003 (0.132; 94.2%)	1.010 (0.096; 94.3%)	1.013 (0.122; 94.6%)	1.003 (0.132; 95.8%)	1.012 (0.096; 94.5%)
	2	1.011 (0.108; 94.2%)	1.016 (0.132; 93.5%)	1.008 (0.088; 93.5%)	1.006 (0.108; 93.5%)	1.013 (0.132; 94.5%)	1.008 (0.088; 92.9%)

Table D.4: Overview of the parameter estimates for $K = 50$ using the strong parametric one-stage and spline procedures for various configurations of the lognormal baseline survival function. The copula association parameter is estimated on its original scale and transformed to Kendall's τ to allow for easy comparisons. It is also important to mention that the non-monotonicity of the lognormal baseline survival leads to a violation of the proportional hazards assumption. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
		$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$	$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$
Clayton	θ	1.555 (0.482; 86.8%)	1.645 (0.478; 80.7%)	1.652 (0.482; 79.0%)	1.054 (0.370; 93.5%)	1.029 (0.358; 93.3%)	1.025 (0.357; 93.2%)
	τ	0.437 (0.174)	0.451 (0.169)	0.452 (0.171)	0.345 (0.146)	0.340 (0.142)	0.339 (0.142)
	β_1	1.066 (0.448; 93.0%)	1.037 (0.436; 92.9%)	0.988 (0.433; 92.2%)	1.054 (0.366; 92.2%)	1.053 (0.365; 93.6%)	1.032 (0.365; 94.5%)
	β_2	1.063 (0.448; 91.6%)	1.047 (0.437; 91.4%)	1.038 (0.434; 92.5%)	1.050 (0.364; 94.5%)	1.044 (0.365; 94.1%)	1.035 (0.364; 94.6%)
Frank	θ	3.778 (1.121; 90.8%)	3.794 (1.116; 92.5%)	3.862 (1.125; 91.2%)	3.348 (0.992; 93.7%)	3.324 (0.982; 94.6%)	3.364 (0.989; 94.2%)
	τ	0.371 (0.214)	0.373 (0.212)	0.378 (0.206)	0.337 (0.240)	0.335 (0.241)	0.338 (0.237)
	β_1	1.137 (0.472; 90.8%)	1.154 (0.475; 91.3%)	1.156 (0.474; 89.8%)	1.066 (0.361; 94.9%)	1.045 (0.361; 93.8%)	1.042 (0.36; 93.0%)
	β_2	1.162 (0.471; 92.0%)	1.146 (0.474; 90.4%)	1.128 (0.473; 89.9%)	1.049 (0.360; 93.6%)	1.035 (0.360; 92.8%)	1.045 (0.36; 92.8%)
Gumbel	θ	1.543 (0.22; 91.1%)	1.543 (0.222; 90.8%)	1.544 (0.221; 89.9%)	1.515 (0.195; 94.4%)	1.512 (0.192; 92.3%)	1.520 (0.195; 93.4%)
	τ	0.352 (0.092)	0.352 (0.093)	0.352 (0.093)	0.340 (0.085)	0.339 (0.084)	0.342 (0.084)
	β_1	1.247 (0.439; 86.6%)	1.267 (0.441; 84.9%)	1.240 (0.439; 84.4%)	1.061 (0.332; 93.9%)	1.076 (0.327; 94.4%)	1.074 (0.327; 92.8%)
	β_2	1.253 (0.438; 85.0%)	1.221 (0.438; 85.8%)	1.208 (0.440; 86.6%)	1.052 (0.328; 93.5%)	1.086 (0.327; 92.5%)	1.055 (0.325; 92.8%)
Gaussian	θ	0.533 (0.11; 86.1%)	0.529 (0.111; 85.1%)	0.524 (0.112; 88.1%)	0.495 (0.108; 92.8%)	0.487 (0.107; 91.1%)	0.496 (0.106; 92.5%)
	τ	0.358 (0.083)	0.355 (0.083)	0.351 (0.084)	0.33 (0.079)	0.324 (0.078)	0.33 (0.078)
	β_1	1.153 (0.466; 88.5%)	1.153 (0.465; 89.5%)	1.124 (0.465; 89.7%)	1.068 (0.357; 94.0%)	1.050 (0.357; 95.4%)	1.036 (0.356; 93.5%)
	β_2	1.189 (0.466; 90.1%)	1.183 (0.466; 89.3%)	1.200 (0.467; 88.5%)	1.054 (0.358; 94.6%)	1.050 (0.359; 95.4%)	1.049 (0.358; 93.9%)
		50% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
		$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$	$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$
Clayton	θ	1.472 (0.591; 92.8%)	1.595 (0.483; 84.4%)	1.584 (0.484; 83.3%)	1.089 (0.464; 93.9%)	1.044 (0.361; 93.8%)	1.037 (0.363; 93.6%)
	τ	0.424 (0.216)	0.444 (0.173)	0.442 (0.173)	0.352 (0.182)	0.343 (0.143)	0.341 (0.144)
	β_1	1.088 (0.490; 91.5%)	1.043 (0.445; 91.5%)	1.072 (0.446; 91.9%)	1.056 (0.387; 92.9%)	1.031 (0.364; 94.6%)	1.032 (0.366; 93.5%)
	β_2	1.054 (0.488; 91.2%)	1.041 (0.445; 92.4%)	1.060 (0.446; 93.3%)	1.059 (0.389; 94.4%)	1.024 (0.365; 95.0%)	1.046 (0.367; 94.1%)
Frank	θ	3.779 (1.268; 93.2%)	3.788 (1.121; 91.9%)	3.808 (1.119; 93.2%)	3.417 (1.148; 94.1%)	3.261 (0.983; 95%)	3.348 (0.986; 94.1%)
	τ	0.371 (0.242)	0.372 (0.213)	0.374 (0.211)	0.343 (0.266)	0.330 (0.250)	0.337 (0.238)
	β_1	1.131 (0.492; 90.5%)	1.128 (0.476; 89.7%)	1.142 (0.472; 90.5%)	1.013 (0.375; 94.7%)	1.046 (0.363; 92.5%)	1.030 (0.359; 94.8%)
	β_2	1.147 (0.493; 91.3%)	1.156 (0.471; 91.0%)	1.135 (0.471; 89.8%)	1.018 (0.376; 94.4%)	1.043 (0.364; 93.7%)	1.020 (0.359; 93.4%)
Gumbel	θ	1.533 (0.234; 91.7%)	1.544 (0.222; 91.2%)	1.528 (0.218; 91.0%)	1.520 (0.218; 93.9%)	1.518 (0.196; 94.2%)	1.508 (0.193; 93.4%)
	τ	0.348 (0.100)	0.352 (0.093)	0.346 (0.093)	0.342 (0.094)	0.341 (0.085)	0.337 (0.085)
	β_1	1.227 (0.455; 89.8%)	1.284 (0.443; 86.7%)	1.268 (0.442; 85.1%)	1.057 (0.340; 94.3%)	1.066 (0.328; 91.4%)	1.054 (0.329; 94.1%)
	β_2	1.219 (0.456; 89.1%)	1.278 (0.439; 85.6%)	1.258 (0.441; 86.2%)	1.068 (0.338; 94.5%)	1.065 (0.326; 92.7%)	1.044 (0.329; 93.8%)
Gaussian	θ	0.509 (0.126; 88.1%)	0.529 (0.111; 85.1%)	0.534 (0.110; 87.7%)	0.491 (0.123; 92.2%)	0.492 (0.108; 94.4%)	0.486 (0.108; 92.8%)
	τ	0.340 (0.093)	0.355 (0.083)	0.358 (0.083)	0.326 (0.090)	0.328 (0.079)	0.323 (0.079)
	β_1	1.133 (0.488; 91.1%)	1.169 (0.467; 90.7%)	1.158 (0.463; 88.3%)	1.088 (0.374; 94.6%)	1.076 (0.361; 95.2%)	1.033 (0.359; 94.0%)
	β_2	1.147 (0.487; 92.1%)	1.180 (0.469; 88.9%)	1.216 (0.465; 89.7%)	1.046 (0.373; 94.2%)	1.091 (0.360; 93.5%)	1.041 (0.358; 92.4%)

Table D.5: Overview of the parameter estimates for $K = 200$ using the strong parametric one-stage and spline procedures for various configurations of the lognormal baseline survival function. The copula association parameter is estimated on its original scale and transformed to Kendall's τ to allow for easy comparisons. It is also important to mention that the non-monotonicity of the lognormal baseline survival leads to a violation of the proportional hazards assumption. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
		$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$	$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$
Clayton	θ	1.586 (0.240; 31.2%)	1.658 (0.237; 17.5%)	1.663 (0.238; 17.6%)	1.015 (0.178; 93.4%)	1.009 (0.177; 93.6%)	1.016 (0.174; 94.0%)
	τ	0.442 (0.086)	0.453 (0.084)	0.454 (0.084)	0.337 (0.071)	0.335 (0.071)	0.337 (0.069)
	β_1	0.993 (0.214; 92.0%)	0.973 (0.209; 91.6%)	0.984 (0.209; 92.8%)	1.015 (0.179; 94.2%)	1.007 (0.178; 94.5%)	1.006 (0.178; 94.9%)
	β_2	1.000 (0.214; 93.0%)	0.971 (0.209; 92.8%)	0.986 (0.209; 91.1%)	1.018 (0.179; 93.8%)	1.013 (0.178; 94.4%)	1.006 (0.178; 93.9%)
Frank	θ	3.826 (0.554; 83.0%)	3.824 (0.553; 81.5%)	3.860 (0.555; 80.6%)	3.252 (0.483; 93.8%)	3.301 (0.483; 95.1%)	3.276 (0.482; 94.1%)
	τ	0.375 (0.103)	0.375 (0.103)	0.378 (0.102)	0.329 (0.124)	0.333 (0.12)	0.331 (0.122)
	β_1	1.098 (0.226; 89.7%)	1.124 (0.226; 88.0%)	1.119 (0.226; 88.0%)	1.006 (0.176; 94.3%)	1.019 (0.176; 95%)	1.015 (0.176; 92.4%)
	β_2	1.106 (0.226; 88.9%)	1.131 (0.227; 87.8%)	1.128 (0.226; 89.2%)	1.012 (0.177; 94.5%)	1.018 (0.176; 95%)	1.015 (0.176; 93.6%)
Gumbel	θ	1.488 (0.108; 91.6%)	1.499 (0.110; 91.9%)	1.497 (0.109; 93.4%)	1.496 (0.095; 95.1%)	1.493 (0.094; 94.3%)	1.497 (0.095; 94.2%)
	τ	0.328 (0.049)	0.333 (0.049)	0.332 (0.049)	0.331 (0.042)	0.330 (0.042)	0.332 (0.042)
	β_1	1.229 (0.211; 76.9%)	1.241 (0.211; 74.6%)	1.241 (0.211; 75.6%)	1.021 (0.158; 95.0%)	1.020 (0.158; 95.0%)	1.024 (0.158; 93.2%)
	β_2	1.237 (0.213; 76.9%)	1.234 (0.211; 77.2%)	1.242 (0.212; 74.2%)	1.021 (0.159; 94.4%)	1.025 (0.158; 95.1%)	1.024 (0.158; 93.8%)
Gaussian	θ	0.532 (0.056; 84.6%)	0.538 (0.056; 85.3%)	0.542 (0.055; 82.0%)	0.498 (0.054; 93.8%)	0.498 (0.053; 92.6%)	0.494 (0.053; 94.7%)
	τ	0.357 (0.042)	0.361 (0.042)	0.365 (0.042)	0.332 (0.040)	0.332 (0.039)	0.329 (0.039)
	β_1	1.144 (0.224; 87.7%)	1.167 (0.223; 83.0%)	1.144 (0.223; 85.7%)	1.011 (0.174; 93.2%)	1.003 (0.174; 94.2%)	0.997 (0.174; 93.9%)
	β_2	1.160 (0.224; 86.3%)	1.152 (0.224; 84.6%)	1.170 (0.223; 83.2%)	1.015 (0.174; 93.8%)	1.009 (0.174; 94.2%)	1.004 (0.175; 92.7%)
		50% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
		$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$	$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$
Clayton	θ	1.391 (0.283; 76.1%)	1.617 (0.239; 24.5%)	1.604 (0.24; 26.2%)	1.017 (0.222; 93.7%)	1.017 (0.178; 94.2%)	1.017 (0.177; 94.4%)
	τ	0.410 (0.105)	0.447 (0.085)	0.445 (0.086)	0.337 (0.088)	0.337 (0.071)	0.337 (0.071)
	β_1	1.054 (0.235; 92.4%)	0.988 (0.212; 92.4%)	1.001 (0.213; 93.2%)	1.018 (0.190; 94.5%)	1.013 (0.179; 95.1%)	1.007 (0.178; 95.2%)
	β_2	1.057 (0.236; 92.1%)	0.993 (0.212; 91.9%)	1.000 (0.212; 92.3%)	1.007 (0.190; 94.1%)	1.014 (0.178; 95.3%)	1.012 (0.178; 95.0%)
Frank	θ	3.725 (0.615; 88.4%)	3.850 (0.556; 79.4%)	3.859 (0.555; 80.8%)	3.236 (0.551; 93.9%)	3.288 (0.486; 94.6%)	3.240 (0.483; 95.1%)
	τ	0.367 (0.121)	0.377 (0.103)	0.378 (0.102)	0.327 (0.142)	0.332 (0.122)	0.328 (0.125)
	β_1	1.086 (0.235; 91.5%)	1.115 (0.225; 88.0%)	1.106 (0.226; 88.8%)	1.010 (0.184; 94.4%)	1.009 (0.176; 95.2%)	1.014 (0.177; 95.3%)
	β_2	1.085 (0.234; 91.5%)	1.120 (0.226; 89.8%)	1.119 (0.226; 85.9%)	1.013 (0.184; 95.1%)	1.017 (0.177; 94.7%)	1.018 (0.177; 94.8%)
Gumbel	θ	1.484 (0.113; 91.9%)	1.496 (0.109; 92.7%)	1.499 (0.109; 91.3%)	1.500 (0.106; 94.7%)	1.495 (0.095; 93.0%)	1.500 (0.096; 95.1%)
	τ	0.326 (0.051)	0.331 (0.049)	0.333 (0.049)	0.333 (0.047)	0.331 (0.043)	0.333 (0.043)
	β_1	1.156 (0.216; 84.3%)	1.222 (0.211; 77.3%)	1.234 (0.211; 76.1%)	1.012 (0.162; 94.3%)	1.015 (0.159; 93.2%)	1.023 (0.158; 95.2%)
	β_2	1.186 (0.217; 83.5%)	1.241 (0.211; 73.9%)	1.231 (0.211; 75.7%)	1.012 (0.163; 94.6%)	1.013 (0.159; 92.2%)	1.027 (0.158; 94.6%)
Gaussian	θ	0.511 (0.063; 92.3%)	0.534 (0.056; 84.6%)	0.540 (0.055; 80.8%)	0.498 (0.062; 95.2%)	0.500 (0.053; 93.6%)	0.500 (0.053; 94.6%)
	τ	0.341 (0.047)	0.359 (0.042)	0.363 (0.042)	0.332 (0.046)	0.334 (0.039)	0.334 (0.039)
	β_1	1.096 (0.234; 89.3%)	1.138 (0.224; 85.1%)	1.156 (0.224; 84.8%)	1.008 (0.181; 94.6%)	1.021 (0.174; 95.4%)	1.019 (0.175; 94.0%)
	β_2	1.092 (0.233; 92.1%)	1.138 (0.224; 86.3%)	1.146 (0.223; 85.0%)	1.000 (0.181; 95.4%)	1.017 (0.174; 95.0%)	1.021 (0.174; 95.0%)

Table D.6: Overview of the parameter estimates for $K = 500$ using the strong parametric one-stage and spline procedures for various configurations of the lognormal baseline survival function. The copula association parameter is estimated on its original scale and transformed to Kendall's τ to allow for easy comparisons. Note that the Gaussian copula is excluded due to computational needs of the estimation procedures under repeated sampling. It is also important to mention that the non-monotonicity of the lognormal baseline survival leads to a violation of the proportional hazards assumption. The standard error and coverage probability are given in brackets, except τ for which only the standard error is reported.

		30% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
		$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$	$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$
Clayton	θ	1.587 (0.151; 2.1%)	1.672 (0.150; 0.5%)	1.661 (0.149; 0.9%)	1.006 (0.112; 94.4%)	1.007 (0.111; 95.1%)	1.012 (0.11; 94.4%)
	τ	0.442 (0.054)	0.455 (0.053)	0.454 (0.053)	0.335 (0.045)	0.335 (0.044)	0.336 (0.044)
	β_1	0.998 (0.134; 89.8%)	0.967 (0.131; 90.0%)	0.981 (0.131; 91.2%)	1.006 (0.113; 92.8%)	1.000 (0.112; 94.4%)	0.999 (0.112; 94.1%)
	β_2	0.996 (0.134; 90.4%)	0.973 (0.132; 92.8%)	0.974 (0.130; 90.3%)	1.007 (0.113; 93.6%)	0.998 (0.112; 94.1%)	0.993 (0.112; 94.5%)
Frank	θ	3.799 (0.348; 66.0%)	3.831 (0.35; 62.9%)	3.837 (0.349; 60.5%)	3.241 (0.305; 94.2%)	3.228 (0.303; 94.9%)	3.258 (0.303; 95%)
	τ	0.373 (0.066)	0.375 (0.065)	0.376 (0.065)	0.328 (0.079)	0.327 (0.079)	0.329 (0.077)
	β_1	1.112 (0.142; 84.1%)	1.104 (0.141; 84.9%)	1.111 (0.142; 85.3%)	1.008 (0.111; 94.8%)	1.008 (0.111; 93%)	1.010 (0.111; 95.2%)
	β_2	1.099 (0.141; 85.8%)	1.112 (0.141; 85.3%)	1.120 (0.142; 83.6%)	1.004 (0.111; 94.9%)	1.006 (0.111; 95%)	1.014 (0.111; 95.5%)
Gumbel	θ	1.492 (0.069; 92.6%)	1.488 (0.069; 90.5%)	1.491 (0.069; 91.6%)	1.493 (0.060; 94.1%)	1.489 (0.060; 95.4%)	1.489 (0.060; 94.4%)
	τ	0.33 (0.031)	0.328 (0.031)	0.329 (0.031)	0.330 (0.027)	0.329 (0.027)	0.328 (0.027)
	β_1	1.218 (0.132; 61.6%)	1.232 (0.134; 54.9%)	1.243 (0.132; 55.5%)	1.010 (0.100; 94.9%)	1.009 (0.099; 95.0%)	1.005 (0.099; 96.7%)
	β_2	1.220 (0.132; 59.7%)	1.236 (0.132; 56.4%)	1.244 (0.132; 55.2%)	1.010 (0.099; 94.6%)	1.012 (0.100; 94.4%)	1.007 (0.100; 94.9%)
		50% censoring					
		Maximum Likelihood			Splines Royston & Parmar (2002)		
		$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$	$(\mu = 0, \sigma = 1)$	$(\mu = 1, \sigma = 0.5)$	$(\mu = 1.6, \sigma = 1)$
Clayton	θ	1.393 (0.179; 41.0%)	1.643 (0.152; 0.9%)	1.602 (0.151; 1.5%)	1.002 (0.139; 93.9%)	1.008 (0.112; 94.2%)	1.006 (0.112; 95.4%)
	τ	0.411 (0.066)	0.451 (0.054)	0.445 (0.054)	0.334 (0.056)	0.335 (0.045)	0.335 (0.045)
	β_1	1.047 (0.147; 92.8%)	0.990 (0.132; 90.8%)	0.988 (0.133; 91.0%)	1.001 (0.119; 93.6%)	1.006 (0.112; 95.3%)	1.007 (0.112; 94.6%)
	β_2	1.052 (0.147; 91.0%)	0.980 (0.132; 91.3%)	1.000 (0.133; 91.7%)	1.008 (0.119; 93.3%)	1.001 (0.112; 94.6%)	1.006 (0.113; 95.1%)
Frank	θ	3.708 (0.388; 75.4%)	3.822 (0.351; 63.7%)	3.831 (0.349; 62.0%)	3.241 (0.348; 94.7%)	3.240 (0.305; 95.0%)	3.255 (0.304; 94.7%)
	τ	0.366 (0.077)	0.375 (0.066)	0.375 (0.065)	0.328 (0.090)	0.328 (0.079)	0.329 (0.078)
	β_1	1.072 (0.148; 90.5%)	1.113 (0.142; 84.4%)	1.108 (0.141; 85.2%)	1.008 (0.116; 95.4%)	1.007 (0.111; 94.4%)	1.005 (0.111; 93.9%)
	β_2	1.087 (0.148; 89.5%)	1.104 (0.142; 84.9%)	1.108 (0.141; 85.4%)	1.008 (0.116; 93.8%)	1.004 (0.111; 94.7%)	1.002 (0.111; 93.8%)
Gumbel	θ	1.472 (0.071; 90.9%)	1.487 (0.069; 92.5%)	1.485 (0.068; 92.0%)	1.491 (0.066; 93.4%)	1.494 (0.06; 94.3%)	1.495 (0.060; 93.6%)
	τ	0.321 (0.033)	0.328 (0.031)	0.327 (0.031)	0.329 (0.030)	0.331 (0.027)	0.331 (0.027)
	β_1	1.182 (0.136; 71.7%)	1.228 (0.132; 56.5%)	1.227 (0.133; 56.7%)	1.008 (0.102; 94.2%)	1.007 (0.099; 94.6%)	1.009 (0.099; 93.9%)
	β_2	1.173 (0.136; 72.8%)	1.239 (0.132; 53.4%)	1.221 (0.134; 58.0%)	1.006 (0.102; 94.3%)	1.008 (0.099; 94.3%)	1.010 (0.099; 93.1%)

E Example R code

Here we provide an example R code to fit a Frank copula model using a 3 knot spline to the ethanol-induced sleep time data set.

```
library(survival)

markel <- readxl::read_excel("alcgrand.xls")

data <- subset(markel[, c("SUB", "SL1", "SL2", "SEX", "CRO",
    "COA", "WG1", "WG2")], CRO %in% c(44, 45, 53, 54))
dataF2 <- data[complete.cases(data), ]

obs1 <- 1 * (dataF2$SL1 > 0)  # Censoring indicator trial 1
obs2 <- 1 * (dataF2$SL2 > 0)  # Censoring indicator trial 2
sl1 <- ifelse(dataF2$SL1 == 0, 1, dataF2$SL1)  # sleep time 1
sl2 <- ifelse(dataF2$SL2 == 0, 1, dataF2$SL2)  # sleep time 2
alb <- factor(1 * (dataF2$COA == 10))  # Indicator albinism
sex <- factor(dataF2$SEX)  # Indicator sex
wg1 <- dataF2$WG1  # Weight trial 1
wg2 <- dataF2$WG2.  # Weight trial 2

# Royston-Parma spline function with 3 knots
bspline <- function(x, k, gamma) {

    kmin <- k[1]
    k1 <- k[2]
    kmax <- k[3]

    gamma0 <- gamma[1]
    gamma1 <- gamma[2]
    gamma2 <- gamma[3]

    lambda1 <- (kmax - k1)/(kmax - kmin)

    v1 <- pmax(0, (x - k1))^3 - lambda1 * pmax(0, (x - kmin))^3 -
        (1 - lambda1) * pmax(0, (x - kmax))^3
    dv1 <- 3 * pmax(0, (x - k1))^2 - 3 * lambda1 * pmax(0, (x -
        kmin))^2 - 3 * (1 - lambda1) * pmax(0, (x - kmax))^2

    s <- gamma0 + gamma1 * x + gamma2 * v1
    ds <- gamma1 + gamma2 * dv1

    return(list(s = s, ds = ds))
}

# -log likelihood to minimize
llroyston <- function(param, k, cop) {
```

```
theta <- param[1]
gamma1 <- param[2:4]
gamma2 <- param[5:7]
beta1 <- param[8]  # Common intercept
beta2 <- param[9]  # Sex
beta3 <- param[10]  # Albinism
beta4 <- param[11]  # Weight 1
beta5 <- param[12]  # Sex:albinism
beta6 <- param[13]  # Weight 2

k1 <- k[1:3]
k2 <- k[4:6]

b1 <- c(beta1, beta2, beta3, beta4, beta5)
b2 <- c(beta1, beta2, beta3, beta6, beta5)
X1 <- model.matrix(~sex * alb + wg1)
X2 <- model.matrix(~sex * alb + wg2)

status1 <- obs1
status2 <- obs2
resp1 <- sl1
resp2 <- sl2

nu1 <- bspline(log(sl1), k1, gamma1)$s + X1 %*% b1
nu2 <- bspline(log(sl2), k2, gamma2)$s + X2 %*% b2

F1 <- 1 - exp(-exp(nu1))
F2 <- 1 - exp(-exp(nu2))
f1 <- exp(nu1 - exp(nu1)) * bspline(log(sl1), k1, gamma1)$ds/sl1
f2 <- exp(nu2 - exp(nu2)) * bspline(log(sl2), k2, gamma2)$ds/sl2

theta <- assoc

alpha1 <- exp(-theta * F1) - 1
alpha2 <- exp(-theta * F2) - 1
beta <- exp(-theta) - 1
gamma <- alpha1 * alpha2/beta

L00 <- (-1/theta) * log(1 + gamma)
L10 <- (gamma/(1 + gamma)) * (1 + 1/alpha1) * f1
L01 <- (gamma/(1 + gamma)) * (1 + 1/alpha2) * f2
L11 <- -theta * f1 * f2 * (1 + 1/alpha1) * (1 + 1/alpha2) *
    gamma/(1 + gamma)^2

LL <- log((L00^((1 - status1) * (1 - status2))) * (L10^(status1 *
    (1 - status2))) * (L01^(status2 * (1 - status1))) * (L11^(status1 *
    status2)))
```

```
    return(-sum(LL))
}

# Log transform of uncensored observations
sl1.obs <- sl1[which(obs1 == 1)]
sl2.obs <- sl2[which(obs2 == 1)]
logsl1 <- log(sl1.obs)
logsl2 <- log(sl2.obs)

# Semiparametric Cox model on uncensored observations
H1 <- coxph(Surv(sl1.obs, rep(1, length(sl1.obs))) ~ sex[which(obs1 ==
    1)] * alb[which(obs1 == 1)] + wg1[which(obs1 == 1)])
H2 <- coxph(Surv(sl2.obs, rep(1, length(sl2.obs))) ~ sex[which(obs2 ==
    1)] * alb[which(obs2 == 1)] + wg2[which(obs2 == 1)])

logH1 <- log(predict(H1, type = "expected"))
logH2 <- log(predict(H2, type = "expected"))

# Position of knots for each margin
k1 <- c(min(logsl1), median(logsl1), max(logsl1))
k2 <- c(min(logsl2), median(logsl2), max(logsl2))

# Calculate spline functions
lambda11 <- (k1[3] - k1[2])/(k1[3] - k1[1])
lambda12 <- (k2[3] - k2[2])/(k2[3] - k2[1])
v11 <- pmax(0, (logsl1 - k1[2]))^3 - lambda11 * pmax(0, (logsl1 -
    k1[1]))^3 - (1 - lambda11) * pmax(0, (logsl1 - k1[3]))^3
v12 <- pmax(0, (logsl2 - k2[2]))^3 - lambda12 * pmax(0, (logsl2 -
    k2[1]))^3 - (1 - lambda12) * pmax(0, (logsl2 - k2[3]))^3

# Linear regression: coefficients of intercept, logsl1 and
# v11 are initial values for gamma
fit.lm1 <- lm(logH1 ~ logsl1 + v11 + sex[which(obs1 == 1)] *
    alb[which(obs1 == 1)] + wg1[which(obs1 == 1)])
fit.lm2 <- lm(logH2 ~ logsl2 + v12 + sex[which(obs2 == 1)] *
    alb[which(obs2 == 1)] + wg2[which(obs2 == 1)])

# Starting values for gamma
gamma.init.1 <- fit.lm1$coef[1:3]
gamma.init.2 <- fit.lm2$coef[1:3]

royston <- nlm(llroyston, p = c(4, gamma.init.1, gamma.init.2,
    -0.25, 0.086, -0.05, -0.02, 0.06, -0.005), k = c(k1, k2),
    hessian = TRUE, iterlim = 1000)
se.royston <- sqrt(diag(solve(royston$hessian)))
```