Estimation of the association parameters in hierarchically clustered survival data by nested Archimedean copula functions

Peer-reviewed author version

HASAN, Mirza Nazmul & BRAEKERS, Roel (2021) Estimation of the association parameters in hierarchically clustered survival data by nested Archimedean copula functions. In: COMPUTATIONAL STATISTICS, 36(4), p. 2755-2787.

DOI: 10.1007/s00180-021-01094-3
Handle: http://hdl.handle.net/1942/33926

Computational Statistics manuscript No.
(will be inserted by the editor)

Estimation of the association parameters in hierarchically clustered survival data by nested Archimedean copula functions

Mirza Nazmul Hasan · Roel Braekers

Received: date / Accepted: date

Abstract Statisticians are frequently confronted with highly complex data such as clustered data, missing data or censored data. In this manuscript, we consider hierarchically clustered survival data. This type of data arises when a sample consists of clusters, and each cluster has several, correlated sub-clusters containing various, dependent survival times. Two approaches are commonly used to analysis such data and estimate the association between the survival times within a cluster and/or sub-cluster. The first approach is by using random effects in a frailty model while a second approach is by using copula models. Hereby we assume that the joint survival function is described by a copula function evaluated in the marginal survival functions of the different individuals within a cluster. In this manuscript, we introduce a copula model based on a nested Archimedean copula function for hierarchical survival data, where both the clusters and sub-clusters are allowed to be moderate to large and varying in size. We investigate one-stage, two-stage and three-stage parametric estimation procedures for the association parameters in this model. In a simulation study we check the finite sample properties of these estimators. Furthermore we illustrate the methods on a real life data-set on Chronic Granulomatous Disease.

Keywords Hierarchical survival data · Nested Archimedean copulas · Frailty models · Varying cluster size · One-stage estimation · Two-stage estimation

Mirza Nazmul Hasan
Universiteit Hasselt, Data Science Institute, I-BioStat
Martelarenlaan 42, 3500 Hasselt, Belgium
Tel.: +3211268601
Shahjalal University of Science and Technology, Department of Statistics
Sylhet-3114, Bangladesh
E-mail: mirzanazmul.hasan@uhasselt.be

Roel Braekers
Universiteit Hasselt, Data Science Institute, I-BioStat
Martelarenlaan 42, 3500 Hasselt, Belgium

1 Introduction

Multilevel or hierarchical survival data occur frequently in different research areas. For example, in a mortality study of cancer patients, the patients are clustered in hospitals that, in turn, are clustered within districts/provinces. Hereby, patients within the same hospital/district/province are assumed to be correlated as they share some common practices and care.

In the analysis of clustered multivariate survival times, two approaches are commonly considered when we want to take the association between the survival times within a cluster and/or sub-cluster into account. A first approach is through frailty models (Duchateau and Janssen 2008; Wienke 2011). A frailty model is a conditional model which assumes that different individuals within the same cluster are independent, conditionally on a common frailty term. It is possible to include covariates in this model but the parameters are estimated conditionally on unobserved frailty terms. This frailty term is assumed to be a realization of a random variable with a given frailty distribution. In case of multi-level survival data, nested frailty models can be used to induce multi-level associations (Ma et al. 2003; Rondeau et al. 2006; Sastry 1997). The nested frailty model accounts for the hierarchical clustering of the data by including two nested random effects. Nested frailty models are particularly appropriate when data are clustered at several hierarchical levels naturally or by design (Rondeau et al. 2006).

A second approach to analyze clustered multivariate survival times is by using copula functions. Shih and Louis (1995) first introduce copula models in the analysis of clustered bivariate survival times, in which they specify a (non-)parametric model for the marginal survival function of each lifetime separately and provide a parametric copula function to describe the association between the different lifetimes. Due to the right-censoring of the lifetimes, it is very hard to obtain for a general copula function a closed form expression of the likelihood function to estimate the parameters of the model. Two solutions to this problem can be found in the literature. A first solution is to look only at pairwise associations between the lifetimes within a cluster. Zhao and Joe (2005) establish in this setting a two-stage estimation procedure for both the marginal and association parameters. Li and Lin (2006) consider a second alternative approach by focusing on copula families for which the full likelihood function could be constructed based on the mathematical properties of the family. They consider a Gaussian copula function to model the spatial correlation for spatially correlated survival data. Li et al. (2008) also consider a Gaussian copula function to model the association for bivariate survival data. Othus and Li (2010) extended this model for varying cluster sizes and allow for covariates. Glidden (2000) on the other hand looks within the Archimedean copula family at the Clayton copula function to develop a two-stage estimator for the association parameter for multivariate clustered survival data. Prenen et al. (2017) extended this model for unbalanced clustered data and also consider a general Archimedean copula function with monotonic generators.

Nested Archimedean copula functions were first introduced by Joe (1997) for three and four dimensional variables. Later McNeil (2008) extended these copula functions for general d-dimensional variables. Recently, Hofert and Pham (2013) developed a tractable formula for the density of nested Archimedean copula functions in arbitrary dimensions if the number of nesting levels is not too large. Although they showed a numerically efficient way to evaluate the log-density, their study didn't show any simulation results or real life data application. For failure time data with multiple levels of clustering, Joe (1993) and Bandeen-Roche and Liang (1996) proposed a family of distributional models but these models were not applied in real applications. Shih and Lu (2007) applied this multivariate survival model with a nested Archimedean copula function to model the associations for child mortality data in a vitamin A trial in Nepal. Hereby the data were clustered within households in villages. They introduced and investigated a three-stage semi-parametric estimation method to analyze this hierarchically clustered survival data.

As far as we know, a one-stage or two-stage estimation method such as defined in Prenen et al. (2017) have not been used before for a multilevel multivariate survival copula model. This current study fills up this gap. The three-stage estimation method of Shih and Lu (2007) was furthermore adapted to a parametric setting such that we could compare it with our proposed one-stage and two-stage methods. In this manuscript, we consider hierarchically clustered survival data which are possibly censored and where both the clusters and sub-clusters are allowed to be large and varying in size. A nested Archimedean copula function is used to incorporate the associations in two levels in the model. We also allow covariates in the marginal distributions of this model.

This paper is structured as follows. In Section 2, we describe a nested Archimedean copula model for hierarchical survival data by rewriting the likelihood function in terms of Laplace transformations. We derive in Section 3, the log-likelihood function for the Clayton and the Gumbel copula as members of the tilted outer power family. In Section 4, we present and derive asymptotic properties of the one-stage, two-stage and three-stage parametric estimation procedures while in Sections 5 and 6, we report finite sample simulation results and the results from a real life data-set respectively. We finish this manuscript with a brief discussion in Section 7.

2 Description of the model

In this section we develop a full copula model for hierarchically clustered survival data in which the size of each cluster and/or sub-cluster is unequal, moderate to large. Hereto we suppose that we have L different independent clusters in the data-set ($l = 1, 2, \dots, L$). Within each cluster, there are different sub-clusters ($j = 1, 2, \dots, N_l$), where N_l is the number of sub-clusters in the l^{th} cluster. In each sub-cluster, we denote the lifetime for the different individuals by a positive random variable T_{ijl}, $i = 1, 2, \ldots, n_{jl}$ where n_{jl} is the number of individuals in the j^{th} sub-cluster of the l^{th} cluster.

We will assume that every event time is subject to random right censoring. For each individual, we consider an independent random censoring variable C_{ijl} such that the observed quantities are given by,

$$Y_{ijl} = \min(T_{ijl}, C_{ijl}); \;\; \delta_{ijl} = I(T_{ijl} \leq C_{ijl})$$
$$i = 1, 2, \ldots, n_{jl}; \; j = 1, 2, \ldots, N_l; \; l = 1, 2, \ldots, L$$

The lifetime is allowed to depend on a set of covariates $\mathbf{Z}_{ijl}$. To build up our copula model for hierarchically clustered survival data, we work in different steps. First we assume that the joint survival function for the lifetimes within sub-cluster j of cluster l is given by,

$$\begin{aligned}
&S(t_{1jl}, \ldots, t_{n_{jl},jl} | \mathbf{Z}_{1jl}, \ldots, \mathbf{Z}_{n_{jl},jl}) \\
&\qquad = P(T_{1jl} > t_{1jl}, \ldots, T_{n_{jl},jl} > t_{n_{jl},jl} | \mathbf{Z}_{1jl}, \ldots, \mathbf{Z}_{n_{jl},jl}) \\
&\qquad = \mathcal{C}_1\Big(S(t_{1jl}|\mathbf{Z}_{1jl}), \ldots, S(t_{n_{jl},jl}|\mathbf{Z}_{n_{jl},jl}); \psi_1\Big) \\
&\qquad = \psi_1\Big[\psi_1^{-1}\{S(t_{1jl}|\mathbf{Z}_{1jl})\} + \ldots + \psi_1^{-1}\{S(t_{n_{jl},jl}|\mathbf{Z}_{n_{jl},jl})\}\Big]
\end{aligned}$$

where $S(t_{ijl}|\mathbf{Z}_{ijl}) = P(T_{ijl} > t_{ijl}|\mathbf{Z}_{ijl})$ is the marginal survival model for the lifetime T_{ijl}, given the covariate values $\mathbf{Z}_{ijl}$. $\mathcal{C}_1$ is an Archimedean copula with a generator ψ_1, which describes the association among all individuals within a sub-cluster. Hereby, $\psi_1 \in \boldsymbol{\psi}_\infty : [0, \infty) \to [0, 1]$ is a continuous, strictly decreasing and completely monotonic function with $\psi_1(0) = 1, \psi_1(\infty) = \lim_{t\to\infty} \psi_1(t) = 0$ and $(-1)^m \psi_1^{(m)}(t) \geq 0$ for all $m \in \mathbb{N}_0, t \in (0, \infty)$. The set of all completely monotone Archimedean generators is denoted by $\boldsymbol{\psi}_\infty$.

Next we also assume that the different sub-clusters within a cluster are also correlated. The joint survival function for two event times from two different sub-clusters j and j^* within a cluster l is given by,

$$\begin{aligned}
S(t_{ijl}, t_{i^*j^*l} | \mathbf{Z}_{ijl}, \mathbf{Z}_{i^*j^*l}) &= P(T_{ijl} > t_{ijl}, T_{i^*j^*l} > t_{i^*j^*l} | \mathbf{Z}_{ijl}, \mathbf{Z}_{i^*j^*l}) \\
&= \mathcal{C}_0\Big(S(t_{ijl}|\mathbf{Z}_{ijl}), S(t_{i^*j^*l}|\mathbf{Z}_{i^*j^*l}); \psi_0\Big) \\
&= \psi_0\Big[\psi_0^{-1}\{S(t_{ijl}|\mathbf{Z}_{ijl})\} + \psi_0^{-1}\{S(t_{i^*j^*l}|\mathbf{Z}_{i^*j^*l})\}\Big]
\end{aligned}$$

where $\mathcal{C}_0$ is the Archimedean copula with a generator $\psi_0 \in \boldsymbol{\psi}_\infty$, which describes the association between sub-clusters within a cluster.

By using partially nested Archimedean copula functions (Joe 1997), we combine both expressions and get that the joint survival functions for all the event times in the l^{th} cluster as:

$$\begin{aligned}
&S(t_{11l}, \ldots, t_{n_{N_l l},N_l,l} | \mathbf{Z}_{11l}, \ldots, \mathbf{Z}_{n_{N_l l},N_l,l}) \\
&\qquad = P(T_{11l} > t_{11l}, \ldots, T_{n_{N_l l},N_l,l} > t_{n_{N_l l},N_l,l} | \mathbf{Z}_{11l}, \ldots, \mathbf{Z}_{n_{N_l l},N_l,l})
\end{aligned}$$

$$
\begin{aligned}
&= \mathcal{C}_0\Big(\Big\{\mathcal{C}_1\big(S(t_{11l}|\mathbf{Z}_{11l}),\ldots,S(t_{n_{1l},1l}|\mathbf{Z}_{n_{1l},1l});\psi_1\big),\\
&\qquad \mathcal{C}_1\big(S(t_{12l}|\mathbf{Z}_{12l}),\ldots,S(t_{n_{2l},2l}|\mathbf{Z}_{n_{2l},2l});\psi_1\big),\ldots,\\
&\qquad \mathcal{C}_1\big(S(t_{1N_l l}|\mathbf{Z}_{1N_l l}),\ldots,S(t_{n_{N_l l},N_l l}|\mathbf{Z}_{n_{N_l l},N_l l});\psi_1\big)\Big\};\psi_0\Big)\\
&= \psi_0\Big[\psi_0^{-1}\circ\psi_1[\psi_1^{-1}\{S(t_{11l}|\mathbf{Z}_{11l})\}+\ldots+\psi_1^{-1}\{S(t_{n_{1l},1l}|\mathbf{Z}_{n_{1l},1l})\}]\\
&\qquad +\psi_0^{-1}\circ\psi_1[\psi_1^{-1}\{S(t_{12l}|\mathbf{Z}_{12l})\}+\ldots+\psi_1^{-1}\{S(t_{n_{2l},2l}|\mathbf{Z}_{n_{2l},2l})\}]+\ldots\\
&\qquad +\psi_0^{-1}\circ\psi_1[\psi_1^{-1}\{S(t_{1N_l l}|\mathbf{Z}_{1N_l l})\}+\ldots+\psi_1^{-1}\{S(t_{n_{N_l l},N_l l}|\mathbf{Z}_{n_{N_l l},N_l l})\}]\Big]\\
&= \psi_0\left[\sum_{j=1}^{N_l}\psi_0^{-1}\circ\psi_1\left\{\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(t_{ijl}|\mathbf{Z}_{ijl})\}\right\}\right]
\end{aligned}
$$

Here, the Archimedean copula functions $\mathcal{C}_0$ and $\mathcal{C}_1$ are called respectively the root copula and the child copula. In order to have a properly defined hierarchical copula function, we have that $\psi_0^{-1}\circ\psi_1 = \mathring{\psi}_{01} \in \boldsymbol{\psi}_\infty^* = \{\omega : [0,\infty)\to[0,\infty)|\omega(0)=0,\omega(\infty)=\infty,(-1)^{j-1}\omega^{(j)}\geq 0; j=1,2,\ldots,\infty\}$ and $\psi_0,\psi_1\in\boldsymbol{\psi}_\infty$ are the Laplace transformations of positive random variables (Joe 1997). Moreover, when the two generators ψ_0 and ψ_1 are of the same Archimedean copula families with corresponding parameters θ_0 and θ_1, they often fulfill the sufficient nesting condition of $\theta_0\leq\theta_1$ (Hofert 2011).

Using that the completely monotonic generator ψ_0 is also the Laplace transformation of a positive distribution function $F_0(x_0)$ with $\bar{F}_0(0)=1$,

$$
\psi_0(t)=\int_0^\infty e^{-tx_0}dF_0(x_0);\quad t\geq 0
$$

we can rewrite the joint survival function for cluster l as,

$$
\begin{aligned}
&S(t_{11l},\ldots,t_{n_{N_l l},N_l,l}|\mathbf{Z}_{11l},\ldots,\mathbf{Z}_{n_{N_l l},N_l,l})\\
&\quad=\psi_0\left[\sum_{j=1}^{N_l}\psi_0^{-1}\circ\psi_1\left\{\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(t_{ijl}|\mathbf{Z}_{ijl})\}\right\}\right]\\
&\quad=\int_0^\infty\exp\left[-x_0\left\{\sum_{j=1}^{N_l}\psi_0^{-1}\circ\psi_1\left(\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(t_{ijl}|\mathbf{Z}_{ijl})\}\right)\right\}\right]dF_0(x_0)\\
&\quad=\int_0^\infty\prod_{j=1}^{N_l}\exp\left[-x_0\mathring{\psi}_{01}\left\{\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(t_{ijl}|\mathbf{Z}_{ijl})\}\right\}\right]dF_0(x_0)\\
&\quad=\int_0^\infty\prod_{j=1}^{N_l}\psi_{01}\left(\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(t_{ijl}|\mathbf{Z}_{ijl})\};x_0\right)dF_0(x_0)
\end{aligned}
\tag{1}
$$

where $\psi_{01}(t;x_0) = \exp\{-x_0\mathring{\psi}_{01}(t)\}$, is called the inner generator. It is a proper generator in t for each $x_0 > 0$ as a composition of the completely monotone function $\exp(-x_0)$ with $\mathring{\psi}_{01}$ which has a completely monotone derivative.

Since each lifetime T_{ijl} is subject to right censoring, we need to take this into account when we construct the likelihood function for this copula model. The contribution of cluster l to the likelihood function corresponds to the derivative of the joint survival function (1) over all uncensored individuals in this cluster. Hence, we get that the contribution of cluster l to the likelihood function is given by,

$$L_l = (-1)^{d_l} \frac{\partial^{d_l}}{\partial\{\delta_{ijl}=1\}} S(y_{11l},\ldots,y_{n_{N_l}l,N_l,l}|\mathbf{Z}_{11l},\ldots,\mathbf{Z}_{n_{N_l}l,N_l,l})$$

where $\partial\{\delta_{ijl}=1\}$ is the set of uncensored individuals in cluster l and $d_l = \sum_{j=1}^{N_l}\sum_{i=1}^{n_{jl}}\delta_{ijl}$ is the total number of uncensored individuals in that cluster.

Using representation (1) of the joint survival function, this derivative is given by

$$\begin{aligned}
L_l &= \int_0^\infty \prod_{j=1}^{N_l} \psi_{01}^{(d_{jl})}\left(\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\};x_0\right) dF_0(x_0) \\
&\qquad \times \prod_{j=1}^{N_l}\prod_{i=1}^{n_{jl}} \left[(\psi_1^{-1})'\{S(y_{ijl}|\mathbf{Z}_{ijl})\} f(y_{ijl}|\mathbf{Z}_{ijl})\right]^{\delta_{ijl}} \\
&= E\left[\prod_{j=1}^{N_l} \psi_{01}^{(d_{jl})}\left(\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\};X_0\right)\right] \\
&\qquad \times \prod_{j=1}^{N_l}\prod_{i=1}^{n_{jl}} \left[(\psi_1^{-1})'\{S(y_{ijl}|\mathbf{Z}_{ijl})\} f(y_{ijl}|\mathbf{Z}_{ijl})\right]^{\delta_{ijl}} \qquad (2)
\end{aligned}$$

where $d_{jl} = \sum_{i=1}^{n_{jl}}\delta_{ijl}$ is the total number of uncensored individuals in the j^{th} sub-cluster under l^{th} cluster. In this expression, we need to find the different derivatives of the inner generators $\psi_{01}(t;x_0)$. Therefore we use the formula of Faà di Bruno's (Craik 2005) which gives the n^{th} derivative of a composition of functions f and g,

$$\begin{aligned}
\frac{d^n}{dx^n} f\{g(x)\} &= (f\circ g)^{(n)}(x) \\
&= \sum_{k=1}^{n} f^{(k)}\{g(x)\} B_{n,k}\{g'(x), g''(x),\ldots,g^{(n-k+1)}(x)\}
\end{aligned}$$

where $B_{n,k}\{g'(x),\ldots,g^{(n-k+1)}(x)\}$ is a Bell polynomial.

In our setting, the inner generator of the nested Archimedean copula in equation (2) given by, $\psi_{01}(t;x_0) = \exp\{-x_0\mathring{\psi}_{01}(t)\}$ with $\mathring{\psi}_{01} = \psi_0^{-1}\circ\psi_1$, we set

$f(y) = \exp(-x_0 y)$ and $g(t) = \mathring{\psi}_{01}(t)$ such that

$$f^{(k)}\{g(t)\} = f^{(k)}\{\mathring{\psi}_{01}(t)\} = (-x_0)^k \exp\{-x_0 \mathring{\psi}_{01}(t)\} = \psi_{01}(t; x_0)(-x_0)^k$$

Hence, we get that for any number of derivatives

$$\begin{aligned}
\psi_{01}^{(n)}(t; x_0) &= \psi_{01}(t; x_0) \sum_{k=1}^{n} B_{n,k}\{\mathring{\psi}'_{01}(t), \dots, \mathring{\psi}_{01}^{(n-k+1)}(t)\}(-x_0)^k \\
&= \psi_{01}(t; x_0) \sum_{k=1}^{n} a_{nk}(t)(-x_0)^k
\end{aligned}$$

where $a_{nk}(t) = B_{n,k}\{\mathring{\psi}'_{01}(t), \dots, \mathring{\psi}_{01}^{(n-k+1)}(t)\}$ are the coefficients of the Bell polynomial. Hereby we note that the sign $\{a_{nk}(t)\} = (-1)^{n-k}$.

By using this general formula for the derivatives of the inner generators $\psi_{01}(t; x_0)$, we can rewrite the product appearing as integrand in equation (2) as follows:

$$\begin{aligned}
&\prod_{j=1}^{N_l} \psi_{01}^{(d_{jl})} \left(\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}; x_0 \right) \\
&= \prod_{j=1}^{N_l} \sum_{k=1}^{d_{jl}} a_{d_{jl},k} \left(\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\} \right) (-x_0)^k \\
&\qquad\qquad \times \prod_{j=1}^{N_l} \psi_{01} \left(\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}; x_0 \right) \\
&= \sum_{k=N_{ul}}^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t_l}(\mathbf{u_l})\}(-x_0)^k \times \prod_{j=1}^{N_l} \exp\left[-x_0 \mathring{\psi}_{01} \left(\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\} \right) \right] \\
&= \sum_{k=N_{ul}}^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t_l}(\mathbf{u_l})\}(-x_0)^k \\
&\qquad\qquad \times \exp\left[-x_0 \left\{ \sum_{j=1}^{N_l} \psi_0^{-1} \circ \psi_1 \left(\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\} \right) \right\} \right] \\
&= \sum_{k=N_{ul}}^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t_l}(\mathbf{u_l})\}(-x_0)^k \times \exp\{-x_0 t_l(\mathbf{u}_l)\}
\end{aligned}$$

In this expression, we denote by $N_{ul} = \sum_{j=1}^{N_l} I(d_{jl} \geq 1)$, the number of subclusters in the l^{th} cluster which includes at least one uncensored observation. Furthermore the coefficients

$$b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t_l}(\mathbf{u_l})\} = \sum_{m \in Q_{\mathbf{d}_l,k}^{N_l}} \prod_{j=1, d_{jl} \geq 1}^{N_l} a_{d_{jl}, m_{jl}} \left[\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\} \right]$$

are the coefficients in the Cauchy product of the polynomials
$\sum_{k=1}^{d_{jl}} a_{d_{jl},k}\Big[\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}\Big](-x_0)^k$.

Hereby, we introduce the following notation to shorten the expressions:

- $\mathbf{d_l} = (d_{1l}, d_{2l}, \ldots, d_{N_l l})^T$
- $\mathbf{t}_l(\mathbf{u}_l) = (t_{1l}(\mathbf{u}_{1l}), \ldots, t_{N_l l}(\mathbf{u}_{N_l l}))^T$ with $t_{jl}(\mathbf{u}_{jl}) = \sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}$
- $t_l(\mathbf{u}_l) = \sum_{j=1}^{N_l}\psi_0^{-1}\circ\psi_1\left(\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}\right)$
- $Q_{\mathbf{d}_l,k}^{N_l} = \Big[m \in \mathbb{N}^{N_l} : \sum_{j=1}^{N_l} m_{jl} = k, m_{jl} \le d_{jl}, j \in \{1, 2, \ldots, N_l\}\Big]$

Taking the expectation of this expression, we get,

$$E\left[\prod_{j=1}^{N_l}\psi_{01}^{(d_{jl})}\left(\sum_{i=1}^{n_{jl}}\psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}; X_0\right)\right]$$
$$= \sum_{k=N_{ul}}^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t_l}(\mathbf{u_l})\}E\Big[(-X_0)^k.\exp\{-x_0 t_l(\mathbf{u}_l)\}\Big]$$
$$= \sum_{k=N_{ul}}^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t}_l(\mathbf{u}_l)\}\psi_0^{(k)}\{t_l(\mathbf{u}_l)\}$$

Hence, we get that the contribution to the likelihood in equation (2) is given by

$$L_l = \left\{\sum_{k=N_{ul}}^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t}_l(\mathbf{u}_l)\}\psi_0^{(k)}\{t_l(\mathbf{u}_l)\}\right\} \times \prod_{j=1}^{N_l}\prod_{i=1}^{n_{jl}}\Big[(\psi_1^{-1})'\{S(y_{ijl}|\mathbf{Z}_{ijl})\}f(y_{ijl}|\mathbf{Z}_{ijl})\Big]^{\delta_{ijl}}$$

Combining the contributions over all clusters, we obtain the following likelihood function

$$L = \prod_{l=1}^{L} L_l = \prod_{l=1}^{L}\left[\left\{\sum_{k=N_{ul}}^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t}_l(\mathbf{u}_l)\}\psi_0^{(k)}\{t_l(\mathbf{u}_l)\}\right\} \times \prod_{j=1}^{N_l}\prod_{i=1}^{n_{jl}}\Big[(\psi_1^{-1})'\{S(y_{ijl}|\mathbf{Z}_{ijl})\}f(y_{ijl}|\mathbf{Z}_{ijl})\Big]^{\delta_{ijl}}\right]$$

and log-likelihood function

$$l = \log L = \sum_{l=1}^{L}\log\left\{\sum_{k=N_{ul}}^{d_l}(-1)^{d_l} b_{\mathbf{d_l},k}^{N_l}\{\mathbf{t}_l(\mathbf{u}_l)\}\psi_0^{(k)}\{t_l(\mathbf{u}_l)\}\right\} + \sum_{l=1}^{L}\sum_{j=1}^{N_l}\sum_{i=1}^{n_{jl}}\delta_{ijl}\Big[\log\Big\{-(\psi_1^{-1})'\{S(y_{ijl}|\mathbf{Z}_{ijl})\}f(y_{ijl}|\mathbf{Z}_{ijl})\Big\}\Big] \quad (3)$$

In the next section, we will show in more details how this log-likelihood function looks like for a specific nested Archimedean copula function.

3 Tilted outer power families

In this section we look at the nested Archimedean copula function based on tilted outer power generators given by

$$\psi_i(t) = \psi\{(c^{\theta_i} + t)^{1/\theta_i} - c\}, i = 0, 1 \tag{4}$$

for a generator $\psi \in \boldsymbol{\psi}_\infty$ and $c \in [0, \infty)$. The range of θ_i depends on specific families of Archimedean copula, for example, $\theta_i \in (0, \infty)$ for a nested Clayton copula and $\theta_i \in [1, \infty)$ for a nested Gumbel copula (see Nelsen 2006 for details). By using Faà di Bruno's formula, we get the k^{th} derivatives of $\psi_0(t)$ (Hofert and Pham 2013),

$$\psi_0^{(k)}(t) = \sum_{j=1}^{k} \psi^{(j)}\{(c^{\theta_0} + t)^{1/\theta_0} - c\}(c^{\theta_0} + t)^{j/\theta_0 - k} s_{kj}(1/\theta_0)$$

where $s_{nk}(x) = \sum_{l=k}^{n} s(n,l)S(l,k)x^l$. Hereby $s(n,l) = (-1)^{n-l}|s(n,l)|$ in which $|s(n,l)|$ are the unsigned Stirling numbers of the 1^{st} kind. These numbers count the number of permutations of n elements with l disjoint cycles. $S(l,k) = \frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^l$ are the Stirling numbers of the 2^{nd} kind, which counts the number of ways of partitioning a set of l elements into k nonempty sets. The different Stirling numbers satisfy following recurrence relations:

$$s(n+1,k) = s(n,k-1) - ns(n,k)$$
$$S(n+1,k) = S(n,k-1) + kS(n,k)$$

for all $k \in \mathbb{N}, n \in \mathbb{N}_0$, with $s(0,0) = S(0,0) = 1$ and $s(n,0) = s(0,n) = S(n,0) = S(0,n) = 0$ for all $n \in \mathbb{N}$.
For the tilted outer power generators of type (4), the nodes are given by,

$$\mathring{\psi}_{01}(t) = \psi_0^{-1} \circ \psi_1 = (c^{\theta_1} + t)^{\alpha_1} - c^{\theta_0}, \quad \text{where} \quad \alpha_1 = \theta_0/\theta_1$$

These generators fulfill the sufficient nesting condition if $\theta_0 \leq \theta_1$ (Hofert and Pham 2013). Moreover we see that the n^{th} derivative is given by

$$\mathring{\psi}_{01}^{(n)}(t) = \alpha_1(\alpha_1 - 1)\dots(\alpha_1 - n + 1)(c^{\theta_1} + t)^{\alpha_1 - n} = (\alpha_1)_n(c^{\theta_1} + t)^{\alpha_1 - n}, \quad n \in \mathbb{N}$$

where $(x)_n = x(x-1)\dots(x-n+1)$ is called a falling factorial. We can write the inner generator as a composition of function,

$$\psi_{01}(t; x_0) = \exp\{-x_0\mathring{\psi}_{01}(t)\} = \exp\Big[-x_0\{(c^{\theta_1} + t)^{\alpha_1} - c^{\theta_0}\}\Big]$$

which has n^{th} derivative,

$$\begin{aligned}\psi_{01}^{(n)}(t;x_0) &= \psi_{01}(t;x_0)\sum_{k=1}^{n} a_{nk}(t)(-x_0)^k \\ &= \psi_{01}(t;x_0)\sum_{k=1}^{n}(c^{\theta_1}+t)^{\alpha_1 k-n}s_{nk}(\alpha_1)(-x_0)^k \end{aligned} \quad (5)$$

where $a_{nk}(t) = B_{n,k}\{\mathring{\psi}'_{01}(t),\ldots,\mathring{\psi}_{01}^{(n-k+1)}(t)\} = (c^{\theta_1}+t)^{\alpha_1 k-n}s_{nk}(\alpha_1)$

Furthermore, we get that $\psi_1^{-1}(t) = \{\psi^{-1}(t)+c\}^{\theta_1} - c^{\theta_1}$ and it's first derivative is $(\psi_1^{-1})'(t) = \theta_1\{\psi^{-1}(t)+c\}^{\theta_1-1}(\psi^{-1})'(t)$.

Using these expressions, we obtain the following log-likelihood function for tilted outer power families:

$$\begin{aligned} l = \log L = &\sum_{l=1}^{L}\log\Bigg\{\sum_{k=N_{ul}}^{d_l}(-1)^{d_l} b_{\mathbf{d}_l,k}^{N_l}\{\mathbf{t}_l(\mathbf{u}_l)\} \\ &\times\left(\sum_{j=1}^{k}\psi^{(j)}\big[\{c^{\theta_0}+t_l(\mathbf{u}_l)\}^{1/\theta_0}-c\big]\{c^{\theta_0}+t_l(\mathbf{u}_l)\}^{j/\theta_0-k}s_{kj}(1/\theta_0)\right)\Bigg\} \\ &+\sum_{l=1}^{L}\sum_{j=1}^{N_l}\sum_{i=1}^{n_{jl}}\delta_{ijl}\Big[\log\Big\{-\theta_1[\psi^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}+c]^{\theta_1-1} \\ &\qquad (\psi^{-1})'\{S(y_{ijl}|\mathbf{Z}_{ijl})\}f(y_{ijl}|\mathbf{Z}_{ijl})\Big\}\Big] \end{aligned}$$

3.1 Clayton copula

In this subsection, we look at the nested Clayton copula. Hereto we take $\psi(t) = 1/(1+t)$ and $c = 1$ in the tilted outer power generator and get

$$\psi_i(t) = \psi\{(c^{\theta_i}+t)^{1/\theta_i} - c\} = \frac{1}{1+(1+t)^{1/\theta_i}-1} = (1+t)^{-1/\theta_i}$$

which is a generator of the Clayton copula with $\theta_i \in (0,\infty)$. Therefore, the k^{th} derivatives of $\psi_0(t)$ is obtained as

$$\begin{aligned}\psi_0^{(k)}(t) &= (-1/\theta_0)(-1/\theta_0-1)\ldots(-1/\theta_0-k+1)(1+t)^{-1/\theta_0-k} \\ &= (-1/\theta_0)_k(1+t)^{-1/\theta_0-k}\end{aligned}$$

We can find the n^{th} derivatives of inner generators, $\psi_{01}(t;x_0) = \exp\{-x_0\mathring{\psi}_{01}(t)\}$ from equation (5) (putting $c = 1$) as

$$\begin{aligned}\psi_{01}^{(n)}(t;x_0) &= \psi_{01}(t;x_0)\sum_{k=1}^{n}a_{nk}(t)(-x_0)^k \\ &= \psi_{01}(t;x_0)\sum_{k=1}^{n}(1+t)^{\alpha_1 k-n}s_{nk}(\alpha_1)(-x_0)^k\end{aligned}$$

where, $a_{nk}(t) = (1+t)^{\alpha_1 k - n} s_{nk}(\alpha_1)$. The inverse of Clayton generator, $\psi_1^{-1}(t) = t^{-\theta_1} - 1$ which has first derivative, $(\psi_1^{-1})'(t) = -\theta_1 t^{-\theta_1 - 1}$. We obtain the following log-likelihood function for the nested Clayton copula:

$$
\begin{aligned}
l = \log L = & \sum_{l=1}^{L} \log \left(\sum_{k=N_{ul}}^{d_l} (-1)^{d_l} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} (-1/\theta_0)_k \{1 + t_l(\mathbf{u}_l)\}^{-1/\theta_0 - k} \right) \\
& + \sum_{l=1}^{L} \sum_{j=1}^{N_l} \sum_{i=1}^{n_{jl}} \log \left[\theta_1 \{S(y_{ijl}|\mathbf{Z}_{ijl})\}^{-(1+\theta_1)} f(y_{ijl}|\mathbf{Z}_{ijl}) \right]^{\delta_{ijl}} \qquad (6) \\
= & \sum_{l=1}^{L} \log \left(\sum_{k=N_{ul}}^{d_l} (-1)^{d_l - k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} (-1)^k (-1/\theta_0)_k \{1 + t_l(\mathbf{u}_l)\}^{-1/\theta_0 - k} \right) \\
& + \sum_{l=1}^{L} \sum_{j=1}^{N_l} \sum_{i=1}^{n_{jl}} \delta_{ijl} \Big[\log(\theta_1) - (1+\theta_1) \log\{S(y_{ijl}|\mathbf{Z}_{ijl})\} + \log\{f(y_{ijl}|\mathbf{Z}_{ijl})\} \Big]
\end{aligned}
$$

3.1.1 Numerical evaluation of the log-likelihood

In the numerical evaluation of the log-likelihood function in equation (6) during for example the optimization process of maximum likelihood, we note that the evaluation of the first term in this log-likelihood expression leads to complicated problems when taking a logarithm of the sum. The second part of the log-likelihood in equation (6) is comparatively trivial to compute. To compute the logarithm of the sum we define,

$$
\begin{aligned}
x_{kl} &= \log \left[(-1)^{d_l - k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} (-1)^k (-1/\theta_0)_k \{1 + t_l(\mathbf{u}_l)\}^{-1/\theta_0 - k} \right] \\
&= \log \left[(-1)^{d_l - k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} \right] + \log \left[(-1)^k (-1/\theta_0)_k \right] \\
&\quad - (1/\theta_0 + k) \log \left[\{1 + t_l(\mathbf{u}_l)\} \right] \qquad (7)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \log \left(\sum_{k=N_{ul}}^{d_l} (-1)^{d_l - k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} (-1)^k (-1/\theta_0)_k \{1 + t_l(\mathbf{u}_l)\}^{-1/\theta_0 - k} \right) \\
&= \log \left(\sum_{k=N_{ul}}^{d_l} \exp(x_{kl}) \right) = (x_l)_{\max} + \log \left(\sum_{k=N_{ul}}^{d_l} \exp\{x_{kl} - (x_l)_{\max}\} \right) \qquad (8)
\end{aligned}
$$

where $(x_l)_{\max} = \max(x_{kl}); N_{ul} \leq k \leq d_l$. The logarithm of the sum in above equation can easily be computed because we take the maximum term out of the sum and the rest of the sum is only a limited contribution where all summands within the sum are in (0,1].

Next we need to calculate x_{kl} in equation (7). To make sure that all terms within logarithms are positive, we note that the signs of the terms $b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\}$

is $(-1)^{d_l-k}$ (Hofert and Pham 2013). Hence we get that $(-1)^{d_l-k}b^{N_l}_{\mathbf{d}_l,k}\{\mathbf{t}_l(\mathbf{u}_l)\} > 0$ for all $k \in \{N_{ul},\ldots,d_l\}$. Also, $(-1)^k\psi_0^{(k)}(t) = (-1)^k(-1/\theta_0)_k\times\ \{1+t_l(\mathbf{u}_l)\}^{-1/\theta_0-k} \geq 0$, as the derivatives, have alternating sign for completely monotonic generators. Since $(-1)^k(-1/\theta_0)_k = (-1)^k(-1/\theta_0)(-1/\theta_0-1)\ldots(-1/\theta_0-k+1) = (-1)^{2k}(1/\theta_0)(1/\theta_0+1)\ldots(1/\theta_0+k-1)$ is positive for all $k \in \{N_{ul},\ldots,d_l\}$, we get that the other part $\{1+t_l(\mathbf{u}_l)\}^{-1/\theta_0-k}$ is also positive. The major difficulty in evaluating x_{kl} lies in the calculation of $\log\left[(-1)^{d_l-k}b^{N_l}_{\mathbf{d}_l,k}\{\mathbf{t}_l(\mathbf{u}_l)\}\right]$. Recall that,

$$
\begin{aligned}
b^{N_l}_{\mathbf{d_l},k}\{\mathbf{t_l}(\mathbf{u_l})\} &= \sum_{m\in Q^{N_l}_{\mathbf{d}_l,k}}\ \prod_{j=1,d_{jl}\geq 1}^{N_l} a_{d_{jl},m_{jl}}(t)\\
\text{where,}\quad a_{d_{jl},m_{jl}}(t) &= (1+t)^{\alpha_1 m_{jl}-d_{jl}} s_{d_{jl}m_{jl}}(\alpha_1);\quad m_{jl}\leq d_{jl}\\
t &= \sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\};\\
\text{and}\quad s_{d_{jl},m_{jl}}(\alpha_1) &= \sum_{l=m_{jl}}^{d_{jl}} s(d_{jl},l)S(l,m_{jl})(\alpha_1)^l
\end{aligned}
$$

In this expression, $s(d_{jl},l)$ and $S(l,m_{jl})$ are the Stirling numbers of first and second kind respectively. The value of both Stirling numbers depend on the total number of events/uncensored individuals in the j^{th} sub-cluster under l^{th} cluster (d_{jl}) and are rapidly increasing with increasing numbers of events in a sub-cluster. As a result, calculating $s_{d_{jl},m_{jl}}(\alpha_1)$ from Stirling numbers is troublesome and is not possible for large number of events in any sub-cluster. Therefore we use an iterative method based on the logarithm of this term and calculate $\log\{(-1)^{d_{jl}-m_{jl}}s_{d_{jl},m_{jl}}(\alpha_1)\}$ since the sign of $s_{d_{jl},m_{jl}}(\alpha_1)$ is $(-1)^{d_{jl}-m_{jl}}$. The details are given in Appendix B. Similar as for the calculation of $s_{d_{jl},m_{jl}}(\alpha_1)$, $a_{d_{jl},m_{jl}}(t)$ and $b^{N_l}_{\mathbf{d_l},k}\{\mathbf{t_l}(\mathbf{u_l})\}$ are also calculated in terms of logarithm because these terms become very close to zero for large d_{jl} and N_l. Hereto we use the following formulas:

$$
\begin{aligned}
&\log\left[(-1)^{d_{jl}-m_{jl}}a_{d_{jl},m_{jl}}(t)\right]\\
&\qquad= (\alpha_1 m_{jl}-d_{jl})\log(1+t) + \log\left[(-1)^{d_{jl}-m_{jl}}s_{d_{jl}m_{jl}}(\alpha_1)\right] \qquad (9)
\end{aligned}
$$

$$
\begin{aligned}
\log\left[(-1)^{d_l-k}b^{N_l}_{\mathbf{d}_l,k}\{\mathbf{t}_l(\mathbf{u}_l)\}\right] &= \log\left[(-1)^{d_l-k}\sum_{m\in Q^{N_l}_{\mathbf{d}_l,k}}\ \prod_{j=1,d_{jl}\geq 1}^{N_l} a_{d_{jl},m_{jl}}(t)\right]\\
&= \log\left[\sum_{m\in Q^{N_l}_{\mathbf{d}_l,k}} \exp\left\{\sum_{j=1,d_{jl}\geq 1}^{N_l}\log\left[(-1)^{d_{jl}-m_{jl}}a_{d_{jl},m_{jl}}(t)\right]\right\}\right]
\end{aligned}
$$

Since $b^{N_l}_{\mathbf{d}_l,k}\{\mathbf{t}_l(\mathbf{u}_l)\}$ is the coefficient in the Cauchy product of the polynomials $\sum_{k=1}^{d_{jl}} a_{d_{jl},k}\Big[\sum_{i=1}^{n_{jl}} \psi_1^{-1}\{S(y_{ijl}|\mathbf{Z}_{ijl})\}\Big](-x_0)^k$, we use an iterative method to calculate the logarithm of the coefficients. For the l^{th} cluster, (ignoring $(-1)^{d_{jl}-m_{jl}}$ for all a's and b's), we have, based on (9), the vectors of the logarithm-coefficients for all sub-clusters

$$
\begin{aligned}
\log\{a_{d_{1l},m_{1l}}(t)\} &= \Big(\log\{a_{d_{1l},0}(t)\}, \log\{a_{d_{1l},1}(t)\}, \ldots, \log\{a_{d_{1l},d_{1l}}(t)\}\Big)\\
\log\{a_{d_{2l},m_{2l}}(t)\} &= \Big(\log\{a_{d_{2l},0}(t)\}, \log\{a_{d_{2l},1}(t)\}, \ldots, \log\{a_{d_{2l},d_{2l}}(t)\}\Big)\\
&\cdots\\
\log\{a_{d_{N_l l},m_{N_l l}}(t)\} &= \Big(\log\{a_{d_{N_l l},0}(t)\}, \log\{a_{d_{N_l l},1}(t)\}, \ldots, \log\{a_{d_{N_l l},d_{N_l l}}(t)\}\Big)
\end{aligned}
$$

First we calculate the logarithm of the coefficients from the first two vectors as follows:

$$
\begin{aligned}
\log\Big[b^2_{\mathbf{d}_1,0}\{\mathbf{t}_l(\mathbf{u}_l)\}\Big] &= \log\Big[\exp\Big\{\log\{a_{d_{1l},0}(t)\} + \log\{a_{d_{2l},0}(t)\}\Big\}\Big]\\
&= \log\{a_{d_{1l},0}(t)\} + \log\{a_{d_{2l},0}(t)\}\\
\log\Big[b^2_{\mathbf{d}_1,1}\{\mathbf{t}_l(\mathbf{u}_l)\}\Big] &= \log\Big[\exp\Big\{\log\{a_{d_{1l},0}(t)\} + \log\{a_{d_{2l},1}(t)\}\Big\}\\
&\quad + \exp\Big\{\log\{a_{d_{1l},1}(t)\} + \log\{a_{d_{2l},0}(t)\}\Big\}\Big]\\
&= \log\Big[\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}\Big]\\
&\quad + \log\left[\sum_{i=0}^{1}\sum_{j=0}^{1}\frac{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}}{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}}\right]\\
&= \Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}\\
&\quad + \log\left[\sum_{\substack{i=0\\ i+j=1}}^{1}\sum_{j=0}^{1}\frac{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}}{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}}\right]\\
\log\Big[b^2_{\mathbf{d}_1,2}\{\mathbf{t}_l(\mathbf{u}_l)\}\Big] &= \Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}\\
&\quad + \log\left[\sum_{\substack{i=0\\ i+j=2}}^{2}\sum_{j=0}^{2}\frac{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}}{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}}\right]\\
&\cdots
\end{aligned}
$$

$$\log\left[b^2_{\mathbf{d}_1,d_l}\{\mathbf{t}_l(\mathbf{u}_l)\}\right] = \Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}$$
$$+ \log\left[\sum_{\substack{i=0\\ i+j=d_l}}^{d_{1l}}\sum_{j=0}^{d_{2l}} \frac{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}}{\exp\Big\{\log\{a_{d_{1l},i}(t)\} + \log\{a_{d_{2l},j}(t)\}\Big\}_{\max}}\right]$$

In this coefficient, we focus on the biggest term of the Cauchy product such that the other terms can be easily computed because all summands within the sum are in (0,1]. The next step is to repeat this procedure between the resulting vector of the previous step with the vector of the next sub-cluster. After we have used the vector of the last sub-cluster, we found the values for the coefficients of the l^{th} cluster.

Finally substituting results from equation (8) in equation (6), we get the log-likelihood function for the nested Clayton copula:

$$\log L = \sum_{l=1}^{L}\left\{(x_l)_{\max} + \log\left(\sum_{k=N_{ul}}^{d_l}\exp\{x_{kl} - (x_l)_{\max}\}\right)\right\} \tag{10}$$
$$+ \sum_{l=1}^{L}\sum_{j=1}^{N_l}\sum_{i=1}^{n_{jl}}\delta_{ijl}\Big[\log(\theta_1) - (1+\theta_1)\log\{S(y_{ijl}|\mathbf{Z}_{ijl})\} + \log\{f(y_{ijl}|\mathbf{Z}_{ijl})\}\Big]$$

3.2 Gumbel copula

In this sub-section, we look at a second nested copula family, the nested Gumbel copula. If we take $\psi(t) = \exp(-t)$ and $c = 0$ in the tilted outer power generator in equation (4), we get,

$$\psi_i(t) = \psi\{(c^{\theta_i} + t)^{1/\theta_i} - c\} = \exp(-t^{1/\theta_i}), \quad i = 0, 1$$

which is a generator of a Gumbel copula with $\theta_i \in [1, \infty)$. To find the k^{th} derivative of $\psi_0(t)$, Faà di Bruno's formula was used which gives,

$$\psi_0^{(k)}(t) = \sum_{j=1}^{k}\psi_0(t)(-1)^j s_{kj}(1/\theta_0)t^{j/\theta_0 - k} = \frac{\psi_0(t)}{t^k}\sum_{j=1}^{k}s_{kj}(1/\theta_0)(-t^{1/\theta_0})^j$$

and we can find the n^{th} derivatives of the inner generators, $\psi_{01}(t; x_0) = \exp\{-x_0\mathring{\psi}_{01}(t)\}$ from equation (5) as

$$\psi_{01}^{(n)}(t; x_0) = \psi_{01}(t; x_0)\sum_{k=1}^{n}a_{nk}(t)(-x_0)^k = \psi_{01}(t; x_0)\sum_{k=1}^{n}t^{\alpha_1 k - n}s_{nk}(\alpha_1)(-x_0)^k$$

with $a_{nk}(t) = t^{\alpha_1 k - n}s_{nk}(\alpha_1)$. The inverse of Gumbel generator, $\psi_1^{-1}(t) = (-\log t)^{\theta_1}$, which has first derivative $(\psi_1^{-1})'(t) = \theta_1(-\log t)^{\theta_1 - 1}(-1/t)$

Hence, from equation (3), after some simplification we obtain the following log-likelihood function for the nested Gumbel copula:

$$\log L = \sum_{l=1}^{L} \log \Bigg(\sum_{k=N_{ul}}^{d_l} (-1)^{d_l-k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} \frac{\psi_0\{t_l(\mathbf{u}_l)\}}{\{t_l(\mathbf{u}_l)\}^k} \sum_{j=1}^{k} (-1)^{j+k} s_{kj}(1/\theta_0)\{t_l(\mathbf{u}_l)\}^{j/\theta_0} \Bigg) \qquad (11)$$

$$+ \sum_{l=1}^{L}\sum_{j=1}^{N_l}\sum_{i=1}^{n_{jl}} \delta_{ijl} \Big[\log \Big(\theta_1 \{-\log S(y_{ijl}|\mathbf{Z}_{ijl})\}^{\theta_1 - 1} \{1/S(y_{ijl}|\mathbf{Z}_{ijl})\} f(y_{ijl}|\mathbf{Z}_{ijl}) \Big) \Big]$$

We compute the logarithm of the sum in the log-likelihood function in equation (11) similar as for the Clayton copula. To do this, let us consider,

$$x_{kl} = \log \left((-1)^{d_l-k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} \frac{\psi_0\{t_l(\mathbf{u}_l)\}}{\{t_l(\mathbf{u}_l)\}^k} \sum_{j=1}^{k} (-1)^{j+k} s_{kj}(1/\theta_0)\{t_l(\mathbf{u}_l)\}^{j/\theta_0} \right)$$

$$= \log \Big[(-1)^{d_l-k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} \Big] + \log \Big[\psi_0\{t_l(\mathbf{u}_l)\} \Big] - k \log \Big\{ t_l(\mathbf{u}_l) \Big\} + \log \left(\sum_{j=1}^{k} (-1)^{j+k} s_{kj}(1/\theta_0)\{t_l(\mathbf{u}_l)\}^{j/\theta_0} \right) \qquad (12)$$

To compute x_{kl} in equation (12), we note in the last part of the equation that there is again a logarithm of a sum. We use the same trick as before to calculate the logarithm. Hereto let us consider,

$$y_{kjl} = \log \Big[(-1)^{j+k} s_{kj}(1/\theta_0)\{t_l(\mathbf{u}_l)\}^{j/\theta_0} \Big]$$

such that we get that

$$\log \Big(\sum_{j=1}^{k} (-1)^{j+k} s_{kj}(1/\theta_0)\{t_l(\mathbf{u}_l)\}^{j/\theta_0} \Big) = \log \left(\sum_{j=1}^{k} \exp(y_{kjl}) \right)$$

$$= (y_{kl})_{\max} + \log \left(\sum_{j=1}^{k} \exp\{y_{kjl} - (y_{kl})_{\max}\} \right)$$

where, $(y_{kl})_{\max} = \max(y_{kjl}); \quad 1 \le j \le k, N_{ul} \le k \le d_l$. Therefore, substituting these values in equation (12) we get,

$$x_{kl} = \log \Big[(-1)^{d_l-k} b_{\mathbf{d}_l,k}^{N_l} \{\mathbf{t}_l(\mathbf{u}_l)\} \Big] + \log \Big[\psi_0\{t_l(\mathbf{u}_l)\} \Big] - k \log \Big\{ t_l(\mathbf{u}_l) \Big\} + (y_{kl})_{\max} + \log \left(\sum_{j=1}^{k} \exp\{y_{kjl} - (y_{kl})_{\max}\} \right)$$

Furthermore we get that

$$\log\left(\sum_{k=N_{ul}}^{d_l}(-1)^{d_l-k}b_{\mathbf{d}_l,k}^{N_l}\{\mathbf{t}_l(\mathbf{u}_l)\}\frac{\psi_0\{t_l(\mathbf{u}_l)\}}{\{t_l(\mathbf{u}_l)\}^k}\sum_{j=1}^{k}(-1)^{j+k}s_{kj}(1/\theta_0)\{t_l(\mathbf{u}_l)\}^{j/\theta_0}\right)$$
$$=\log\left(\sum_{k=N_{ul}}^{d_l}\exp(x_{kl})\right)=(x_l)_{\max}+\log\left(\sum_{k=N_{ul}}^{d_l}\exp\{x_{kl}-(x_l)_{\max}\}\right)$$

where, $(x_l)_{\max}=\max(x_{kl}),\quad N_{ul}\le k\le d_l$. Therefore, from equation (11) we obtain the following log-likelihood function for the nested Gumbel copula,

$$\log L=\sum_{l=1}^{L}\left\{(x_l)_{\max}+\log\left(\sum_{k=N_{ul}}^{d_l}\exp\{x_{kl}-(x_l)_{\max}\}\right)\right\}$$
$$+\sum_{l=1}^{L}\sum_{j=1}^{N_l}\sum_{i=1}^{n_{jl}}\delta_{ijl}\Big[\log(\theta_1)+(\theta_1-1)\log\{-\log S(y_{ijl}|\mathbf{Z}_{ijl})\}$$
$$-\log\{S(y_{ijl}|\mathbf{Z}_{ijl})\}+\log\{f(y_{ijl}|\mathbf{Z}_{ijl})\}\Big] \quad (13)$$

4 Parametric estimation

In this section, we investigate one-stage, two-stage and three-stage parametric estimation methods to estimate the parameters of our developed model. Prenen et al. (2017) used one- and two-stage parametric estimation methods to estimate the parameters of the Archimedean copula model for clustered survival data. We extend both of their methods for hierarchically clustered survival data, where both the clusters and subclusters are large and of varying sizes.

4.1 One-stage parametric estimation

Let $\boldsymbol{\beta}$ be the parameter vector for the margins, containing distribution-specific parameters for the baseline survival and covariate effects; and let $\boldsymbol{\theta}$ be the parameter vector for the association parameters based on the nested Archimedean copula. We use the log-likelihood function derived in equation (10) for Clayton copula and in equation (13) for Gumbel copula model. We find the maximum likelihood estimates $(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})$, by solving the following two set of equations:

$$\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\theta})=\frac{\partial\log L(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}=\mathbf{0}$$
$$\boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\beta},\boldsymbol{\theta})=\frac{\partial\log L(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\mathbf{0}$$

simultaneously.

Theorem 1. Let $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ be the solution of $\boldsymbol{U_\beta(\beta, \theta)} = \mathbf{0}$ and $\boldsymbol{U_\theta(\beta, \theta)} = \mathbf{0}$ simultaneously and let $(\boldsymbol{\beta}, \boldsymbol{\theta})$ be the true parameters vector. From maximum likelihood theory (Cox and Hinkley 1974), we know that, under regularity conditions (details in supplementary materials), $\sqrt{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}^{-1}$, where the Fisher information matrix $\mathbf{I}$ is partitioned into blocks:

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{\beta\beta} & \mathbf{I}_{\beta\theta} \\ \mathbf{I}_{\theta\beta} & \mathbf{I}_{\theta\theta} \end{bmatrix}$$

where $L\mathbf{I}_{\beta\beta}$ is the variance-covariance matrix of $\boldsymbol{U_\beta}$, $L\mathbf{I}_{\beta\theta}$ is the covariance matrix between $\boldsymbol{U_\beta}$ and $\boldsymbol{U_\theta}$ and $L\mathbf{I}_{\theta\theta}$ is the variance-covariance matrix of $\boldsymbol{U_\theta}$.

In practical applications, standard errors of parameter estimates can be retrieved from the diagonal elements of the inverse of the Hessian matrix $\mathbf{H}$, where,

$$\mathbf{H}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \left[\frac{\partial^2 \log L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})}{\partial \eta_i \partial \eta_j} \right]_{i,j=1,\ldots,p+2} \quad \text{with } \boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\theta})$$

4.2 Two-stage parametric estimation

In this sub-section, we investigate a two-stage parametric estimation. At the first stage, we estimate the marginal parameters of the parametric model and covariate effects assuming a working assumption of independence. In the second stage, we estimate both the association parameters θ_0 and θ_1 by plugging the estimates for the margins into the log-likelihood functions (10) and (13). Two-stage parametric estimation has been used mainly for multivariate models if the numerical optimization for maximum likelihood estimation is too time consuming or infeasible. Let $\boldsymbol{\beta}$ be estimated by $\bar{\boldsymbol{\beta}}$ at the first stage when assuming that all subjects are independent. That is, $\bar{\boldsymbol{\beta}}$ is the solution of the estimating equations

$$\begin{aligned} \boldsymbol{U^*_\beta(\beta)} &= \sum_{l=1}^{L} \sum_{j=1}^{N_l} \sum_{i=1}^{n_{jl}} \delta_{ijl} \frac{\partial \log\{f(y_{ijl}|\mathbf{Z}_{ijl})\}}{\partial \boldsymbol{\beta}} + (1 - \delta_{ijl}) \frac{\partial \log\{S(y_{ijl}|\mathbf{Z}_{ijl})\}}{\partial \boldsymbol{\beta}} \\ &= \sum_{l=1}^{L} \boldsymbol{U}^*_{l,\boldsymbol{\beta}}(\boldsymbol{\beta}) = \mathbf{0} \end{aligned}$$

Under regularity conditions stated in the supplementary material, $\sqrt{L}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $(\mathbf{I}^*)^{-\mathbf{1}}\mathbf{V}(\mathbf{I}^*)^{-\mathbf{1}}$, where $\mathbf{V}$ is the variance-covariance matrix of the score functions $\boldsymbol{U^*_\beta}$ and $\mathbf{I}^*$ is the Fisher information of $\boldsymbol{U^*_\beta}$. The use of the robust sandwich estimator is required since $(\mathbf{I}^*)^{-\mathbf{1}}$ is not a consistent estimator of the asymptotic variance-covariance matrix because of the correlation between the survival times.

After the margins are estimated at first stage, we estimate the association parameters $\boldsymbol{\theta}$ by solving the estimating equations

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

Theorem 2. Let $\bar{\boldsymbol{\theta}}$ be the solution of $\boldsymbol{U_{\theta}}(\bar{\boldsymbol{\beta}}, \boldsymbol{\theta}) = \mathbf{0}$ and let $\boldsymbol{\theta_0}$ be the true value of the association parameters. Then under regularity conditions (details in supplementary materials), $\sqrt{L}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta_0})$ converges to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix

$$\text{Var-Cov}(\bar{\boldsymbol{\theta}}) = (\mathbf{I}_{\theta\theta})^{-1} + (\mathbf{I}_{\theta\theta})^{-1}\mathbf{I}_{\theta\beta}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\beta\theta}(\mathbf{I}_{\theta\theta})^{-1}$$

The proof of theorem 2 is provided in Appendix A. To estimate this quantity, we make use of $(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}$, which is the robust variance-covariance matrix that is obtained in the first stage, $(\mathbf{I}_{\theta\theta})^{-1}$ is obtained from the hessian matrix at second stage and $\mathbf{I}_{\theta\beta}$ is obtained from the Hessian matrix of the one-stage procedure, which can be estimated numerically by performing one iteration of the one-stage optimization in which we evaluate the Hessian matrix in the two-stage parameter results.

4.3 Three-stage parametric estimation

Based on the estimation method of Shih and Lu (2007), we introduce and investigate in this subsection a three-stage parametric estimation method to analyze the hierarchically clustered survival data.

In the first stage of the estimation procedure, we estimate the marginal parameters under the assumption of independence, similar to the first stage of the two-stage estimation method explained in the previous subsection.

At the second stage, we estimate the association within the sub-clusters at the lowest level of the hierarchy. Hereby we assume that the lifetimes in different sub-clusters are independent of each other such that the clustering reduces to only one level. This allows us to use the model by Prenen et al. 2017. After the margins are estimated at the first stage, we estimate the association parameter θ_1 by solving the following estimating equation

$$U_{\theta_1}(\bar{\boldsymbol{\beta}}, \theta_1) = \frac{\partial \log\{L(\bar{\boldsymbol{\beta}}, \theta_1)\}}{\partial \theta_1} = 0$$

The following result is found in Prenen et al. 2017.

Theorem 3. Let $\bar{\theta}_1$ denote the solution of $U_{\theta_1}(\bar{\boldsymbol{\beta}}, \theta_1) = 0$ and let θ_{10} be the true value of the association parameter. Under regularity conditions, $\sqrt{L}(\bar{\theta}_1 - \theta_{10})$ converges to a normal distribution with mean 0 and variance

$$\sigma_1^2 = \text{Var}(\bar{\theta}_1) = \frac{1}{I_{\theta\theta}} + \frac{\mathbf{I}_{\theta\boldsymbol{\beta}}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\boldsymbol{\beta}\theta}}{I_{\theta\theta}^2}$$

At stage 3, we randomly sample one observation from each subcluster in each cluster, with replacement. The resampled data set of size $\sum_{l=1}^{L} N_l$ contains information about θ_0 but no information about θ_1. The estimate of θ_0 from the resampled data set can be done in a way similar to the second stage estimation. This estimation process is repeated for a large number of times Q, and the within-subcluster resampling estimator, $\bar{\theta}_0$ is the average of the Q resampled-based estimates,

$$\bar{\theta}_0 = Q^{-1} \sum_{q=1}^{Q} \hat{\theta}_0^{(q)}$$

where $\hat{\theta}_0^{(q)}$ denotes the estimate from the qth resampled data set. Originally Hoffman et al. (2001) proposed this clever within cluster resampling method for analyzing clustered data. The asymptotic theory of $\bar{\theta}_0$ is stated below.

Theorem 4. Let $\bar{\theta}_0$ denote the estimate of the true value of the association parameter θ_{00}. Under the regularity conditions, the estimator $\bar{\theta}_0$ is consistent and $\sqrt{L}(\bar{\theta}_0 - \theta_{00})$ converges weakly to a zero-mean normal distribution with variance σ_0^2, which can be consistently estimated by,

$$\hat{\sigma}_0^2 = L\Big\{Q^{-1} \sum_{q=1}^{Q} \hat{\sigma}_0^{2(q)} - (Q-1)Q^{-1} S_{\theta_0}^2\Big\}$$

where $\hat{\sigma}_0^{2(q)}$ is the estimated variance from the qth analysis and $S_{\theta_0}^2 = (Q-1)^{-1} \sum_{q=1}^{Q} \{\hat{\theta}_0^{(q)} - \bar{\theta}_0\}^2$ is the variance among the Q resampled-based estimates $\hat{\theta}_0^{(q)}$'s (Hoffman et al. 2001).

5 Simulation study

In this section, we study the finite sample performance of the different proposed estimation procedures for the nested Archimedean copula model by using simulated data. Hereto we generate 500 data sets with 50, 200 or 500 clusters of size varying between 1 and 10 and sub-clusters of size varying between 1 and 15. The survival times are simulated from a nested Clayton copula with θ_0 (θ_1) = 0.6 (1.5), 1.2 (2.5), 2.0 (6.0) or from a nested Gumbel copula with θ_0 (θ_1) =1.3 (1.8), 1.6 (2.3), 2.0 (4.0) such that $\theta_0 < \theta_1$ for both copula functions. We assume that the marginal survival functions are Weibull distributed $S(t) = \exp\{\lambda t^\rho \exp(\beta' Z)\}$ in which ρ =1.5, λ=0.0045 and Z is a dichotomous covariate with effect $\beta = 0.3$. The values of the association parameters (θ_0 and θ_1) for both copula models are chosen such that the corresponding values of Kendall's τ are comparable. The data are generated by using the sampling algorithm of Hofert and Mächler (2011). We assume that the censoring distribution is also Weibull distributed, with parameters ($\lambda_c = 0.0066$ and $\rho_c = 0.92$) and ($\lambda_c = 0.0120$ and $\rho_c = 0.82$) to get 25% and 50% censoring respectively. For the three-stage estimation method we take $Q = 500$.

The simulation results for $0\%, 25\%$ and 50% censoring are shown in Table 1, 2 and 3 respectively. For both copulas, simulation results are listed in increasing order of association. For the Clayton copula, higher values of θ correspond to a higher degree of association via $\tau = \theta/(\theta + 2)$ and for Gumbel copula, higher values of θ also correspond to a higher degree of association via $\tau = 1 - 1/\theta$. For each degree of association, we report the mean estimated values of $\hat{\theta}$ in the first row. Mean standard errors together with the coverage of 95% confidence interval are reported in the second row for each combination of association parameters.

As the number of independent clusters increases from $L = 50$ to $L = 200$, standard errors are halved for one-stage estimation procedure since they are proportional to $1/\sqrt{L}$. For the Clayton copula, the biases of the estimates are larger for two- and three-stage estimation methods compare to one-stage parametric estimation. For the nested Gumbel copula, this biases are not noticeably different among the estimation methods. Furthermore we noted that the biases of the estimates are not noticeably affected by an increasing percentage of censoring for both the copulas. The standard errors are smaller for one-stage estimation method compared to both two- and three-stage estimation method for both copula functions. For the one-stage estimation method, the standard errors become a little larger when more censoring is present and they are increasing similarly for both copulas. For both the two- and three-stage estimation methods, the standard errors become more larger with increasing censoring for the nested Gumbel copula compared to the nested Clayton copula. For both copula functions, the coverages are smaller for both the two- and three-stage estimation methods compared to one-stage estimation method, especially for a small number of clusters. The transition from $L = 50$ to $L = 200$ and $L = 500$ leads to a reduction of the biases and we have approximately unbiased estimates of θ_0 and θ_1 for larger cluster size. The coverage of a 95% confidence interval increases with the increasing number of clusters and almost all coverages are greater than 95% for $L = 500$. However, when the number of clusters is small, the two- and three-stage parametric procedures are not recommended. The one-stage parametric procedure yields the best results in every setting.

Table 1 Simulation results for different number of clusters of varying sizes (0% censoring)

Model	# of clusters	$\theta_0(\tau_0)$	$\theta_1(\tau_1)$	Parametric one-stage		Parametric two-stage		Parametric three-stage	
				$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$
Weibull-Clayton	L=50	0.6(0.23)	1.5(0.43)	0.596 (0.093;94.0%)	1.496 (0.170;93.2%)	0.579 (0.109;85.8%)	1.469 (0.185;86.2%)	0.585 (0.105;86.2%)	1.470 (0.187;86.2%)
		1.2(0.38)	2.5(0.56)	1.195 (0.175;94.6%)	2.499 (0.317;94.4%)	1.159 (0.205;88.8%)	2.439 (0.342;88.2%)	1.171 (0.201;89.0%)	2.441 (0.343;88.8%)
		2.0(0.50)	6.0(0.75)	2.000 (0.297;95.2%)	6.004 (0.792;95.4%)	1.930 (0.337;88.4%)	5.808 (0.798;88.8%)	1.942 (0.330;88.8%)	5.809 (0.800;88.8%)
	L=200	0.6(0.23)	1.5(0.43)	0.600 (0.046;94.0%)	1.499 (0.085;94.6%)	0.594 (0.061;91.2%)	1.489 (0.104;91.4%)	0.596 (0.057;91.0%)	1.489 (0.104;91.8%)
		1.2(0.38)	2.5(0.56)	1.200 (0.088;96.0%)	2.502 (0.159;94.6%)	1.187 (0.115;92.8%)	2.481 (0.189;92.0%)	1.190 (0.110;92.8%)	2.481 (0.190;92.2%)
		2.0(0.50)	6.0(0.75)	2.003 (0.149;94.9%)	6.010 (0.396;95.1%)	1.979 (0.185;92.4%)	5.947 (0.435;93.6%)	1.984 (0.179;92.8%)	5.947 (0.435;93.6%)
	L=500	0.6(0.23)	1.5(0.43)	0.600 (0.029;96.4%)	1.499 (0.054;95.6%)	0.596 (0.040;93.0%)	1.493 (0.068;93.0%)	0.596 (0.037;92.6%)	1.493 (0.068;93.6%)
		1.2(0.38)	2.5(0.56)	1.200 (0.056;96.2%)	2.501 (0.101;96.2%)	1.197 (0.075;94.4%)	2.492 (0.123;94.2%)	1.198 (0.072;94.6%)	2.492 (0.123;94.2%)
		2.0(0.50)	6.0(0.75)	2.000 (0.094;96.2%)	5.996 (0.250;96.4%)	1.989 (0.120;94.4%)	5.967 (0.281;95.6%)	1.991 (0.116;94.0%)	5.967 (0.282;95.6%)
Weibull-Gumbel	L=50	1.3(0.23)	1.8(0.44)	1.291 (0.057;91.8%)	1.787 (0.082;92.4%)	1.293 (0.069;89.4%)	1.788 (0.102;88.6%)	1.295 (0.063;89.2%)	1.788 (0.097;87.8%)
		1.6(0.38)	2.3(0.57)	1.585 (0.100;90.8%)	2.279 (0.145;91.0%)	1.585 (0.113;89.8%)	2.276 (0.164;89.4%)	1.588 (0.106;89.2%)	2.276 (0.159;88.8%)
		2.0(0.50)	4.0(0.75)	1.970 (0.159;91.2%)	3.940 (0.313;90.0%)	1.974 (0.168;91.2%)	3.942 (0.331;90.4%)	1.977 (0.165;89.8%)	3.942 (0.327;90.2%)
	L=200	1.3(0.23)	1.8(0.44)	1.298 (0.029;95.2%)	1.797 (0.041;95.4%)	1.300 (0.036;94.0%)	1.799 (0.054;95.0%)	1.300 (0.033;93.2%)	1.799 (0.051;94.4%)
		1.6(0.38)	2.3(0.57)	1.598 (0.050;95.4%)	2.297 (0.073;94.6%)	1.599 (0.059;95.2%)	2.298 (0.086;94.4%)	1.600 (0.055;94.4%)	2.298 (0.083;94.2%)
		2.0(0.50)	4.0(0.75)	1.996 (0.080;95.0%)	3.993 (0.157;94.6%)	1.998 (0.087;95.2%)	3.995 (0.172;95.4%)	1.999 (0.085;95.0%)	3.996 (0.170;95.0%)
	L=500	1.3(0.23)	1.8(0.44)	1.300 (0.018;96.8%)	1.799 (0.026;96.4%)	1.301 (0.023;96.6%)	1.801 (0.034;96.2%)	1.301 (0.021;94.0%)	1.801 (0.033;95.2%)
		1.6(0.38)	2.3(0.57)	1.599 (0.032;96.6%)	2.298 (0.046;96.4%)	1.601 (0.037;95.8%)	2.301 (0.055;95.6%)	1.601 (0.035;94.4%)	2.301 (0.053;94.2%)
		2.0(0.50)	4.0(0.75)	2.000 (0.051;96.8%)	3.998 (0.099;96.2%)	2.001 (0.055;95.8%)	4.001 (0.110;94.6%)	2.001 (0.054;95.0%)	4.001 (0.109;94.6%)

Table 2 Simulation results for different number of clusters of varying sizes (25% censoring)

Model	# of clusters	$\theta_0(\tau_0)$	$\theta_1(\tau_1)$	Parametric one-stage $\hat{\theta}_0$	Parametric one-stage $\hat{\theta}_1$	Parametric two-stage $\hat{\theta}_0$	Parametric two-stage $\hat{\theta}_1$	Parametric three-stage $\hat{\theta}_0$	Parametric three-stage $\hat{\theta}_1$
Weibull-Clayton	L=50	0.6(0.23)	1.5(0.43)	0.596	1.496	0.585	1.479	0.591	1.479
				(0.095;94.2%)	(0.172;94.2%)	(0.108;87.4%)	(0.185;88.6%)	(0.104;88.8%)	(0.187;88.6%)
		1.2(0.38)	2.5(0.56)	1.193	2.497	1.170	2.458	1.182	2.457
				(0.178;94.8%)	(0.320;94.6%)	(0.204;90.4%)	(0.340;89.6%)	(0.200;91.2%)	(0.342;89.6%)
		2.0(0.50)	6.0(0.75)	1.995	6.001	1.948	5.866	1.963	5.866
				(0.300;95.4%)	(0.794;95.8%)	(0.340;90.0%)	(0.805;91.2%)	(0.333;90.0%)	(0.808;91.2%)
	L=200	0.6(0.23)	1.5(0.43)	0.600	1.499	0.595	1.491	0.597	1.491
				(0.047;93.6%)	(0.086;94.2%)	(0.057;93.0%)	(0.099;93.2%)	(0.054;92.4%)	(0.100;94.0%)
		1.2(0.38)	2.5(0.56)	1.197	2.498	1.190	2.488	1.193	2.488
				(0.089;95.4%)	(0.160;93.2%)	(0.109;93.0%)	(0.182;93.0%)	(0.105;92.8%)	(0.182;93.4%)
		2.0(0.50)	6.0(0.75)	2.001	6.007	1.985	5.965	1.992	5.965
				(0.150;94.6%)	(0.397;94.8%)	(0.180;93.0%)	(0.424;93.4%)	(0.175;93.6%)	(0.426;93.8%)
	L=500	0.6(0.23)	1.5(0.43)	0.600	1.500	0.596	1.494	0.597	1.494
				(0.030;96.4%)	(0.054;95.6%)	(0.037;94.0%)	(0.064;94.0%)	(0.035;92.4%)	(0.064;94.2%)
		1.2(0.38)	2.5(0.56)	1.199	2.499	1.199	2.495	1.200	2.495
				(0.056;96.0%)	(0.101;95.2%)	(0.070;94.8%)	(0.117;94.0%)	(0.068;93.4%)	(0.117;94.4%)
		2.0(0.50)	6.0(0.75)	2.001	5.995	1.993	5.974	1.995	5.975
				(0.095;96.0%)	(0.251;95.6%)	(0.115;95.0%)	(0.271;95.4%)	(0.112;94.2%)	(0.272;95.4%)
Weibull-Gumbel	L=50	1.3(0.23)	1.8(0.44)	1.292	1.788	1.293	1.787	1.295	1.787
				(0.059;91.0%)	(0.085;91.6%)	(0.074;87.8%)	(0.110;89.0%)	(0.068;89.2%)	(0.104;87.4%)
		1.6(0.38)	2.3(0.57)	1.584	2.277	1.585	2.276	1.589	2.276
				(0.102;91.6%)	(0.147;91.0%)	(0.121;89.6%)	(0.177;88.8%)	(0.113;88.4%)	(0.171;87.8%)
		2.0(0.50)	4.0(0.75)	1.972	3.942	1.976	3.942	1.980	3.942
				(0.160;91.2%)	(0.316;89.6%)	(0.179;90.6%)	(0.354;89.8%)	(0.174;89.8%)	(0.349;89.8%)
	L=200	1.3(0.23)	1.8(0.44)	1.298	1.797	1.300	1.799	1.300	1.799
				(0.030;95.2%)	(0.043;96.4%)	(0.039;94.0%)	(0.059;94.6%)	(0.035;92.6%)	(0.055;94.2%)
		1.6(0.38)	2.3(0.57)	1.598	2.298	1.599	2.298	1.600	2.298
				(0.052;96.0%)	(0.075;95.4%)	(0.063;93.6%)	(0.093;94.8%)	(0.059;93.6%)	(0.090;94.0%)
		2.0(0.50)	4.0(0.75)	1.996	3.993	1.998	3.995	1.999	3.995
				(0.082;94.8%)	(0.160;94.6%)	(0.093;95.2%)	(0.185;94.4%)	(0.090;95.2%)	(0.183;94.0%)
	L=500	1.3(0.23)	1.8(0.44)	1.300	1.799	1.301	1.801	1.301	1.801
				(0.019;96.6%)	(0.027;96.6%)	(0.025;96.4%)	(0.038;96.0%)	(0.023;94.4%)	(0.035;95.2%)
		1.6(0.38)	2.3(0.57)	1.599	2.299	1.601	2.301	1.601	2.301
				(0.033;96.4%)	(0.047;96.0%)	(0.041;96.6%)	(0.060;96.0%)	(0.038;94.4%)	(0.057;95.2%)
		2.0(0.50)	4.0(0.75)	2.000	3.997	2.001	4.001	2.002	4.001
				(0.052;96.4%)	(0.101;96.0%)	(0.059;96.2%)	(0.118;95.6%)	(0.058;94.8%)	(0.117;95.2%)

Table 3 Simulation results for different number of clusters of varying sizes (50% censoring)

Model	# of clusters	$\theta_0(\tau_0)$	$\theta_1(\tau_1)$	Parametric one-stage		Parametric two-stage		Parametric three-stage	
				$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$
Weibull-Clayton	L=50	0.6(0.23)	1.5(0.43)	0.597 (0.099;93.8%)	1.498 (0.177;93.6%)	0.588 (0.107;91.6%)	1.485 (0.185;90.4%)	0.598 (0.105;90.2%)	1.485 (0.185;90.4%)
		1.2(0.38)	2.5(0.56)	1.193 (0.184;94.4%)	2.495 (0.325;94.4%)	1.179 (0.203;92.2%)	2.472 (0.340;91.0%)	1.197 (0.199;91.8%)	2.472 (0.341;91.6%)
		2.0(0.50)	6.0(0.75)	1.994 (0.306;94.8%)	5.989 (0.798;95.0%)	1.958 (0.341;91.6%)	5.901 (0.813;91.2%)	1.981 (0.335;92.2%)	5.901 (0.818;92.0%)
	L=200	0.6(0.23)	1.5(0.43)	0.600 (0.049;94.0%)	1.498 (0.089;94.4%)	0.596 (0.056;94.2%)	1.492 (0.096;95.6%)	0.599 (0.053;92.2%)	1.493 (0.096;95.4%)
		1.2(0.38)	2.5(0.56)	1.199 (0.092;95.2%)	2.503 (0.163;94.8%)	1.193 (0.106;94.4%)	2.494 (0.177;93.4%)	1.197 (0.101;93.6%)	2.494 (0.177;93.6%)
		2.0(0.50)	6.0(0.75)	2.002 (0.154;95.0%)	6.012 (0.401;94.0%)	1.991 (0.177;94.8%)	5.986 (0.421;94.8%)	2.002 (0.172;95.6%)	5.987 (0.423;95.0%)
	L=500	0.6(0.23)	1.5(0.43)	0.600 (0.031;95.6%)	1.499 (0.056;96.6%)	0.597 (0.036;95.0%)	1.497 (0.062;94.8%)	0.598 (0.034;92.2%)	1.497 (0.062;94.6%)
		1.2(0.38)	2.5(0.56)	1.201 (0.058;95.2%)	2.499 (0.103;95.4%)	1.199 (0.067;94.8%)	2.497 (0.113;94.0%)	1.200 (0.065;93.8%)	2.496 (0.113;94.2%)
		2.0(0.50)	6.0(0.75)	1.998 (0.097;95.0%)	5.994 (0.253;96.0%)	1.996 (0.112;94.2%)	5.980 (0.267;95.6%)	2.000 (0.109;94.8%)	5.980 (0.269;95.8%)
Weibull-Gumbel	L=50	1.3(0.23)	1.8(0.44)	1.292 (0.061;92.6%)	1.789 (0.090;92.2%)	1.293 (0.080;86.2%)	1.787 (0.121;85.8%)	1.296 (0.076;88.2%)	1.787 (0.114;85.0%)
		1.6(0.38)	2.3(0.57)	1.585 (0.104;90.8%)	2.279 (0.153;93.4%)	1.584 (0.131;88.0%)	2.273 (0.195;88.0%)	1.590 (0.124;88.6%)	2.273 (0.186;87.6%)
		2.0(0.50)	4.0(0.75)	1.973 (0.163;92.0%)	3.946 (0.323;90.8%)	1.973 (0.193;89.1%)	3.934 (0.387;87.3%)	1.981 (0.189;90.1%)	3.934 (0.380;87.3%)
	L=200	1.3(0.23)	1.8(0.44)	1.299 (0.031;95.2%)	1.799 (0.046;95.6%)	1.299 (0.043;94.2%)	1.799 (0.066;93.6%)	1.300 (0.039;93.0%)	1.799 (0.062;92.4%)
		1.6(0.38)	2.3(0.57)	1.599 (0.053;95.8%)	2.300 (0.078;95.2%)	1.598 (0.070;94.2%)	2.298 (0.104;94.6%)	1.600 (0.065;93.6%)	2.298 (0.100;94.2%)
		2.0(0.50)	4.0(0.75)	1.996 (0.082;95.2%)	3.995 (0.163;94.6%)	1.996 (0.102;94.0%)	3.992 (0.206;94.4%)	1.998 (0.099;94.0%)	3.993 (0.202;93.8%)
	L=500	1.3(0.23)	1.8(0.44)	1.301 (0.020;95.8%)	1.800 (0.029;96.8%)	1.301 (0.027;95.2%)	1.801 (0.042;95.8%)	1.301 (0.025;94.2%)	1.801 (0.040;94.2%)
		1.6(0.38)	2.3(0.57)	1.600 (0.033;96.6%)	2.300 (0.049;96.4%)	1.601 (0.045;95.6%)	2.301 (0.067;96.0%)	1.601 (0.042;94.6%)	2.301 (0.064;95.2%)
		2.0(0.50)	4.0(0.75)	2.000 (0.052;96.8%)	3.999 (0.104;96.0%)	2.001 (0.065;95.6%)	4.000 (0.132;96.0%)	2.002 (0.063;94.6%)	4.000 (0.130;95.6%)

6 Modelling time to Chronic Granulomatous Disease (CGD)

In this section we illustrate our hierarchically clustered Archimedean copula models on a real life data example. Hereto we consider the Chronic Granulomatous Disease data set (CGD) as described by Fleming and Harrington (1991). Chronic Granulomatous Disease (CGD) is a group of inherited rare disorders of the immune function characterized by recurrent pyogenic infections which usually present themselves early in life and may lead to death in childhood. There is evidence that gamma interferon is able to reduce the rate of infections requiring hospitalization. A double-blinded clinical trial was conducted in which patients were randomized to either placebo or gamma interferon. In total, 128 eligible patients with CGD were accrued in 13 hospitals by the International CGD Cooperative Study Group who were followed for about one year between 1988 and 1989. The number of patients per hospital ranged from 4 to 26. Each patient may experience more than one infection. The infection times (times-to-event) are the times between recurrent CGD infections on each patient. There is a minimum of one and a maximum of eight (recurrent) infection times per patient, with a total of 203 records. All recurrent events are clustered within a patient (sub-cluster), which in turn are clustered within a hospital. One covariate, treatment is considered in this study. The objective of this study is to estimate the correlation between time to CGD infection within patient and within hospital. Another objective is to estimate treatment (gamma interferon) effect on CGD infection times. The dataset includes 13 clusters (hospitals) and 128 sub-clusters (patients). cluster size varies between 4 and 26 and sub-cluster size varies between 1 and 8. The censoring percentage is 62.6%.

Table 4 Estimated results for the different parametric estimation method

Copula model	One-stage			Two-stage			Three-stage		
	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\beta}$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\beta}$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\beta}$
Clayton	0.006	1.319	-0.829	0.057	0.771	-1.030	0.540	0.733	-1.030
	(0.107)	(0.597)	(0.285)	(0.141)	(0.410)	(0.140)	(0.317)	(0.440)	(0.140)
Gumbel	1.008	1.142	-0.930	1.025	1.129	-1.030	1.262	1.105	-1.030
	(0.031)	(0.088)	(0.297)	(0.097)	(0.114)	(0.140)	(0.118)	(0.081)	(0.140)

We assume a Weibull distribution for the times to CGD infection,

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

and model the association structure by a nested Clayton copula and a nested Gumbel copula. The estimates for the association parameters and the treatment effect by using all three estimation methods is shown in Table 4.

The treatment effects are similar for all three estimation methods and for both copula models. The estimated association within patients $\hat{\theta}_1$ is significant in all estimation procedures. The estimated value is significantly larger in the one-stage method than in the other methods. However, the association within hospital is significant for three-stage method and insignificant for other two methods. As the cluster size is small we prefer one-stage estimation procedure over other two procedures based on simulation studies. The hazard ratio in the Weibull-Clayton model is 0.44 (95% confidence interval: [0.25,0.76]) and is 0.39 (95% confidence interval: [0.22,0.71]) for the Weibull-Gumbel model by using the one-stage parametric method. The estimated association within patient is much higher compare to between patients for both copula models, which is expected. The estimated Kendall's correlation within patient, $\hat{\tau}_1 = 0.397(0.108)$ is larger for the nested Clayton copula compared to the nested Gumbel copula, $\hat{\tau}_1 = 0.124(0.067)$. The Kendall's correlation within hospital is not significant for both copula models by using one-stage parametric estimation method.

7 Discussion

Most of the existing works for hierarchically clustered survival data use a nested random effects model to induce multi-level association. Although Shih and Lu (2007) considered a nested Archimedean copula model, they used a three-stage estimation methods to estimate the model parameters because finding and evaluating the full likelihood function and it's derivatives is complex and difficult for multilevel survival data. In this article, we solved this problem by using the work of Hofert and Pham (2013). For hierarchical survival data, where the data are clustered in two levels, we developed a general likelihood function to estimate the model parameters for any class of nested Archimedean copulas allowing for any varying (sub-)cluster size and taking censoring into account. Furthermore, we derived this log-likelihood function for two specific families of nested Archimedean copulas, namely, Clayton and Gumbel copulas. We investigated a one-stage, two-stage and three-stage parametric estimation procedure to estimate the model parameters. In the finite sample simulation study we saw that all three estimation methods give approximately unbiased parameter estimates for large numbers of clusters and the coverage of the 95% confidence intervals seems also very good. For small numbers of clusters, the two-stage and three-stage methods are not recommended since they lead to larger bias and less coverage. The one-stage procedure performs better in every settings.

In this article, we used a two-levels of hierarchy but it is possible to extend it for three or more levels of hierarchy. We assumed the same correlation structure for each level but one can assume different correlation structure in future research. The models were developed for (mixtures of) the Clayton copula and (mixtures of) the Gumbel copula, but can be generalized to mixtures of any Archimedean families for which the nesting conditions are met.

References

Bandeen-Roche KJ, Liang KY (1996) Modelling failure-time associations in data with multiple levels of clustering. Biometrika 83:29–39

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

Craik ADD (2005) Prehistory of Faà di Bruno's formula. Am Math Mon 112(2):119–130

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

Fleming TR, Harrington DP (1991) Counting Processes and Survival Analysis. Wiley, New York

Glidden DV (2000) A two-stage estimator of the dependence parameter for the Clayton-Oakes model. Lifetime Data Anal 6:141–156

Hofert M (2011) Efficiently sampling nested Archimedean copulas. Comput Stat Data Anal 55:57–70

Hofert M, Pham D (2013) Densities of nested Archimedean copulas. J Multivar Anal 118:37–52

Hoffman EB, Sen PK, Weinberg CR (2001) Within-cluster resampling. Biometrika 88:1121–1134

Joe H (1993) Parametric families of multivariate distributions with given margins. J Multivar Anal 46:262–282

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

Li Y, Lin X (2006) Semiparametric normal transformation models for spatially correlated survival data. J Am Stat Assoc 101:591–603

Li Y, Prentice R, Lin X (2008) Semiparametric maximum likelihood estimation in normal transformation models for bivariate survival data. Biometrika 95(4):947–960

Ma R, Krewski R, Burnett RT (2003) Random effects Cox models: A Poisson modelling approach. Biometrika 90:157–169

McNeil AJ (2008) Sampling nested Archimedean copulas. J Stat Comput Simul 78(6):567–581

Nelsen RB (2006) An Introduction to Copulas. Springer, New York

Othus M, Li Y (2010) A Gaussian copula model for multivariate survival data. Stat Biosci 2:154–179

Prenen L, Braekers R, Duchateau L (2017) Extending the Archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. J R Stat Soc B 79:483–505

Rondeau V, Filleul L, Joly P (2006) Nested frailty models using maximum penalized likelihood estimation. Stat Med 25:4036–4052

Sastry N (1997) A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. J Am Stat Assoc 92:426–435

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

Shih JH, Lu S (2007) Analysis of failure time data with multilevel clustering, with application to the child vitamin A intervention trial in Nepal. Biometrics 63:673–680

Wienke A (2011) Frailty Models in Survival Analysis. Chapman and Hall, New York

Zhao Y, Joe H (2005) Composite likelihood estimation in multivariate data analysis. Can J Stat 33:335–356

Appendix A: Proof of theorem 2

Let $\boldsymbol{\beta_0}$ denote the true parameter vector for the margins. Expanding the score function $\boldsymbol{U}^*_{\boldsymbol{\beta}}$ in a Taylor series around $\boldsymbol{\beta_0}$ and evaluating it at $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, we get under regularity conditions (stated in the supplementary material) of maximum likelihood theory

$$\boldsymbol{U}^*_{\boldsymbol{\beta}}(\bar{\boldsymbol{\beta}}) = \mathbf{0} = \boldsymbol{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta_0}) + \left.\frac{\partial \boldsymbol{U}^*_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}} (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta_0}) + o_p(\sqrt{L}).$$

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

$$\begin{aligned}\boldsymbol{U_\theta}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\theta}}) = \mathbf{0} = \boldsymbol{U_\theta}(\boldsymbol{\beta_0}, \boldsymbol{\theta_0}) &+ \left.\frac{\partial \boldsymbol{U_\theta}}{\partial \boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},\boldsymbol{\theta})=(\boldsymbol{\beta_0},\boldsymbol{\theta_0})} (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta_0}) \\ &+ \left.\frac{\partial \boldsymbol{U_\theta}}{\partial \boldsymbol{\theta}}\right|_{(\boldsymbol{\beta},\boldsymbol{\theta})=(\boldsymbol{\beta_0},\boldsymbol{\theta_0})} (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) + o_p(\sqrt{L}).\end{aligned}$$

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

$$\begin{aligned}-\frac{1}{L}\left.\frac{\partial \boldsymbol{U}^*_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}} &= \frac{1}{L}\sum_{l=1}^{L}\Big[-\frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{U}^*_{l,\boldsymbol{\beta}}(\boldsymbol{\beta_0})\Big] \to \mathbf{I}^* \\ -\frac{1}{L}\left.\frac{\partial \boldsymbol{U_\theta}}{\partial \boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},\boldsymbol{\theta})=(\boldsymbol{\beta_0},\boldsymbol{\theta_0})} &= \frac{1}{L}\sum_{l=1}^{L}\Big[-\frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{U}_{l,\boldsymbol{\theta}}(\boldsymbol{\beta_0}, \boldsymbol{\theta_0})\Big] \to \mathbf{I}_{\theta\beta} \\ -\frac{1}{L}\left.\frac{\partial \boldsymbol{U_\theta}}{\partial \boldsymbol{\theta}}\right|_{(\boldsymbol{\beta},\boldsymbol{\theta})=(\boldsymbol{\beta_0},\boldsymbol{\theta_0})} &= \frac{1}{L}\sum_{l=1}^{L}\Big[-\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{U}_{l,\boldsymbol{\theta}}(\boldsymbol{\beta_0}, \boldsymbol{\theta_0})\Big] \to \mathbf{I}_{\theta\theta}\end{aligned}$$

Hence

$$\frac{1}{\sqrt{L}}\begin{pmatrix}\boldsymbol{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta_0}) \\ \boldsymbol{U_\theta}(\boldsymbol{\beta_0}, \boldsymbol{\theta_0})\end{pmatrix} \to \sqrt{L}\begin{pmatrix}\mathbf{I}^* & \mathbf{0} \\ \mathbf{I}_{\theta\beta} & \mathbf{I}_{\theta\theta}\end{pmatrix}\begin{pmatrix}\bar{\boldsymbol{\beta}} - \boldsymbol{\beta_0} \\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta_0}\end{pmatrix}$$

By the central limit theorem, $\frac{1}{\sqrt{L}}\begin{pmatrix}\boldsymbol{U}^*_{\boldsymbol{\beta}}(\boldsymbol{\beta_0}) \\ \boldsymbol{U_\theta}(\boldsymbol{\beta_0}, \boldsymbol{\theta_0})\end{pmatrix}$ converges to multivariate normal with mean $\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}$ and variance-covariance matrix $\begin{pmatrix}\mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\theta\theta}\end{pmatrix}$ with $\mathbf{V} = \mathrm{Var}\Big(\boldsymbol{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta_0})\Big) = E\Big[\boldsymbol{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta_0})(\boldsymbol{U}^*_{1,\boldsymbol{\beta}}(\boldsymbol{\beta_0}))'\Big]$.

Thus, $\sqrt{L}\begin{pmatrix}\bar{\boldsymbol{\beta}}-\boldsymbol{\beta_0}\\ \bar{\boldsymbol{\theta}}-\boldsymbol{\theta_0}\end{pmatrix}$ converges to multivariate normal with mean vector zero and variance-covariance matrix

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

$$=\begin{pmatrix}(\mathbf{I}^*)^{-1} & \mathbf{0}\\ -\mathbf{I}_{\theta\beta}(\mathbf{I}^*)^{-1}(\mathbf{I}_{\theta\theta})^{-1} & (\mathbf{I}_{\theta\theta})^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{V} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{\theta\theta}\end{pmatrix}\begin{pmatrix}(\mathbf{I}^*)^{-1^T} & -(\mathbf{I}_{\theta\theta})^{-1^T}(\mathbf{I}^*)^{-1^T}\mathbf{I}_{\theta\beta}^T\\ \mathbf{0} & (\mathbf{I}_{\theta\theta})^{-1^T}\end{pmatrix}$$

$$=\begin{pmatrix}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T} & -(\mathbf{I}_{\theta\theta})^{-1^T}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T}\mathbf{I}_{\beta\theta}\\ -\mathbf{I}_{\theta\beta}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T}(\mathbf{I}_{\theta\theta})^{-1} & V^*\end{pmatrix}$$

$$\text{with } V^*=(\mathbf{I}_{\theta\theta})^{-1}+(\mathbf{I}_{\theta\theta})^{-1^T}\mathbf{I}_{\theta\beta}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1^T}\mathbf{I}_{\beta\theta}(\mathbf{I}_{\theta\theta})^{-1}$$

The lower right element of this matrix is the asymptotic variance of $\sqrt{L}(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta_0})$,which is given by

$$\text{Var-Cov}(\bar{\boldsymbol{\theta}})=V^*=(\mathbf{I}_{\theta\theta})^{-1}+(\mathbf{I}_{\theta\theta})^{-1}\mathbf{I}_{\theta\beta}(\mathbf{I}^*)^{-1}\mathbf{V}(\mathbf{I}^*)^{-1}\mathbf{I}_{\beta\theta}(\mathbf{I}_{\theta\theta})^{-1}$$

[Since $(\mathbf{I}_{\theta\theta})^{-1^T}=(\mathbf{I}_{\theta\theta})^{-1}$ and $(\mathbf{I}^*)^{-1^T}=(\mathbf{I}^*)^{-1}$]

Appendix B: Compute $s_{nk}(\alpha_1)$ for Clayton copula

The n^{th} and $(n+1)^{th}$ derivatives of the inner generator for the nested Clayton copula function are as follows,

$$\psi_{01}^{(n)}(t;x_0)=\psi_{01}(t;x_0)\sum_{k=1}^{n}(1+t)^{\alpha_1 k-n}s_{nk}(\alpha_1)(-x_0)^k \tag{14}$$

$$\psi_{01}^{(n+1)}(t;x_0)=\psi_{01}(t;x_0)\sum_{k=1}^{n+1}(1+t)^{\alpha_1 k-(n+1)}s_{n+1,k}(\alpha_1)(-x_0)^k \tag{15}$$

We also get the $(n+1)^{th}$ derivative of the inner generator by differentiating equation (14), which gives,

$$\begin{aligned}\psi_{01}^{(n+1)}(t;x_0)=&\psi_{01}^{(1)}(t;x_0)\sum_{k=1}^{n}(1+t)^{\alpha_1 k-n}s_{nk}(\alpha_1)(-x_0)^k\\ &+\psi_{01}(t;x_0)\sum_{k=1}^{n}(\alpha_1 k-n)(1+t)^{\alpha_1 k-n-1}s_{nk}(\alpha_1)(-x_0)^k\\ &=\psi_{01}(t;x_0)\alpha_1(1+t)^{\alpha_1-1}(-x_0)\sum_{k=1}^{n}(1+t)^{\alpha_1 k-n}s_{nk}(\alpha_1)(-x_0)^k\\ &+\psi_{01}(t;x_0)\sum_{k=1}^{n}(\alpha_1 k-n)(1+t)^{\alpha_1 k-(n+1)}s_{nk}(\alpha_1)(-x_0)^k\end{aligned}$$

$$
\begin{aligned}
&= \psi_{01}(t;x_0)\sum_{k=1}^{n}(1+t)^{\alpha_1(k+1)-(n+1)}\alpha_1 s_{nk}(\alpha_1)(-x_0)^{(k+1)} \\
&+ \psi_{01}(t;x_0)\sum_{k=1}^{n}(\alpha_1 k-n)(1+t)^{\alpha_1 k-(n+1)}s_{nk}(\alpha_1)(-x_0)^{k} \\
&= \psi_{01}(t;x_0)\sum_{k^*=2}^{n+1}(1+t)^{\alpha_1 k^*-(n+1)}\alpha_1 s_{n,k^*-1}(\alpha_1)(-x_0)^{k^*} \\
&+ \psi_{01}(t;x_0)\sum_{k=1}^{n}(\alpha_1 k-n)(1+t)^{\alpha_1 k-(n+1)}s_{nk}(\alpha_1)(-x_0)^{k}
\end{aligned} \tag{16}
$$

Now, by comparing equations (15) and (16), we get,

$$
\begin{aligned}
&s_{n+1,1}(\alpha_1) = (\alpha_1 - n)s_{n,1}(\alpha_1) \quad \text{for} \quad k=1 \\
&s_{n+1,n+1}(\alpha_1) = \alpha_1 s_{n,n}(\alpha_1) \quad \text{for} \quad k=n+1 \\
&s_{n+1,k}(\alpha_1) = (\alpha_1 k - n)s_{n,k}(\alpha_1) + \alpha_1 s_{n,k-1}(\alpha_1) \quad \text{for} \quad k=2,3,\dots,n
\end{aligned}
$$

Using that for $n=0$, $s_{0,0}(\alpha_1)=1$ we get that

$$
\begin{aligned}
&s_{1,1}(\alpha_1) = \alpha_1 s_{0,0}(\alpha_1) = \alpha_1 \\
&\Rightarrow s_{2,2}(\alpha_1) = \alpha_1 s_{1,1}(\alpha_1) = \alpha_1^2 \\
&\Rightarrow \dots \Rightarrow s_{n,n}(\alpha_1) = \alpha_1 s_{n-1,n-1}(\alpha_1) = \alpha_1^n
\end{aligned}
$$

Therefore, $\log\left\{(-1)^{n-n}s_{n,n}(\alpha_1)\right\} = (-1)^{n-n}n\log(\alpha_1) = n\log(\alpha_1)$ for all $n\in\mathbb{N}$. Furthermore we get that

$$
\begin{aligned}
&s_{2,1}(\alpha_1) = (\alpha_1-1)s_{1,1}(\alpha_1) = \alpha_1(\alpha_1-1) = (\alpha_1)_2 \\
\Rightarrow\ &s_{3,1}(\alpha_1) = (\alpha_1-2)s_{2,1}(\alpha_1) = (\alpha_1)_3 \\
&\dots \\
\Rightarrow\ &s_{n,1}(\alpha_1) = (\alpha_1)_n = \alpha_1(\alpha_1-1)(\alpha_1-2)\dots(\alpha_1-n+1)
\end{aligned}
$$

Hence, we get that $\log\left\{(-1)^{n-1}s_{n,1}(\alpha_1)\right\} = \log\left\{(-1)^{n-1}(\alpha_1)_n\right\}$ in which $(-1)^{n-1}(\alpha_1)_n > 0$ and $\log\left\{(-1)^{n+1-k}s_{n+1,k}(\alpha_1)\right\} = \log\left\{(-1)^{n+1-k}(\alpha_1 k - n)s_{n,k}(\alpha_1) + (-1)^{n+1-k}\alpha_1 s_{n,k-1}(\alpha_1)\right\}$ for all $n\in\mathbb{N}$

Acknowledgements We would like to thank the editors and anonymous referees for valuable comments and insightful suggestions, which helped us to improve the manuscript. For the simulations we used the infrastructure of the Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government-department Economics, Science and Innovation.

Declarations

Funding

Not applicable.

Conflicts of interest

The authors declare that they have no potential conflicts of interest.

Availability of data and material

Not applicable.

Code availability

The computational code added as a supplementary text.