Modelling hierarchical clustered censored data with the hierarchical Kendall copula

Chien-Lin SU\textsuperscript{1,2}, Johanna G. NEŠLEHOVÁ\textsuperscript{2*}, and Weijing WANG\textsuperscript{1}

\textsuperscript{1}Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, ROC
\textsuperscript{2}Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada H3A 0G4

Key words and phrases: Archimedean copula; association; censoring; dimension reduction; hierarchical clustered data; Kendall distribution; multiple imputation.

MSC 2010: Primary 62N01; secondary 62N02

Abstract: This article proposes a new model for right-censored survival data with multi-level clustering based on the hierarchical Kendall copula model of Brechmann (2014) with Archimedean clusters. This model accommodates clusters of unequal size and multiple clustering levels, without imposing any structural conditions on the parameters or on the copulas used at various levels of the hierarchy. A step-wise estimation procedure is proposed and shown to yield consistent and asymptotically Gaussian estimates under mild regularity conditions. The model fitting is based on multiple imputation, given that the censoring rate increases with the level of the hierarchy. To check the model assumption of Archimedean dependence, a goodness-of-fit test is developed. The finite-sample performance of the proposed estimators and of the goodness-of-fit test is investigated through simulations. The new model is applied to data from the study of chronic granulomatous disease.

1. INTRODUCTION

Multi-level hierarchical clustered data arise in practice whenever the sampling scheme involves a hierarchical structure. For example, consider the study of chronic granulomatous disease (CGD)
from a multi-center placebo-controlled randomized trial which aimed to investigate the effect of gamma interferon on the disease (Fleming & Harrington, 1991, Section 4.4). There are two levels of clustering; the upper-level and lower-level cluster is formed, respectively, by patients from the same hospital and by the same patient’s infection records.

Hierarchical clustered data are typically analyzed using frailty or copula models. In the frailty approach, the hierarchical dependence structure is accounted for through nested random effect models. For example, Sastry (1997) considers a proportional hazard model with two nested independent Gamma frailty variables, while a similar model of Yau (2001) relies on independent normal frailties. Rondeau, Filleul & Joly (2006) also use a model similar to that of Sastry (1997) but handle the baseline estimation nonparametrically through spline-smoothing. Shih & Lu (2009) consider a nested random effect proportional hazard model with an unspecified baseline hazard function, whereas Ha & Lee (2005) assume an accelerated failure time model with nested and independent frailty variables.

In the copula-based strategy, the dependence structure is modelled explicitly. At the heart of this approach is the decomposition of Sklar (1959), which states that the joint survival function of any random vector \((X_1, \ldots, X_d)\) can be expressed in terms of its univariate survival functions \(\tilde{F}_1, \ldots, \tilde{F}_d\), and a copula \(C\), that is, a \(d\)-variate distribution function with standard uniform univariate margins. Specifically, for all \(x_1, \ldots, x_d \in \mathbb{R}\),

\[
\Pr(X_1 > x_1, \ldots, X_d > x_d) = C\{\tilde{F}_1(x_1), \ldots, \tilde{F}_d(x_d)\}.
\]

(1)

Furthermore, for any copula \(C\) and margins \(\tilde{F}_1, \ldots, \tilde{F}_d\) the expression on the right-hand side of (1) is a valid \(d\)-variate survival function with marginal survival functions \(\tilde{F}_1, \ldots, \tilde{F}_d\). Thus, a copula model consists of assuming that in Sklar’s representation the copula \(C\) and the margins \(\tilde{F}_1, \ldots, \tilde{F}_d\) belong to suitably chosen parametric (or semiparametric) families of copulas and univariate survival functions, respectively. The key advantage of this approach is that the dependence structure can be modelled, fitted, and validated independently of the margins (Nelsen, 2006; Joe, 2015; Durante & Sempi, 2016; Mai & Scherer, 2017).

Although copula models have gained substantial popularity in various fields, their use in the analysis of hierarchical clustered data is less frequent. This is likely because more sophisticated copula models need to be constructed that reflect the hierarchical dependence specific to clustered data. In addition, inference is more complex when the data are incomplete. Shih & Lu (2007) applied hierarchical Archimedean copulas (HAC) of Joe (1993) with Clayton generators to two-level clustered survival data, and proposed a three-stage estimation procedure that can handle right censoring. The HAC model was also used by Andersen (2004) to model familial data, where the siblings and parents are assumed to have an exchangeable dependence structure. Familial data were also analyzed using Gaussian copula models by Zhao & Joe (2005) and Othus & Li (2010). Very recently, Prenen, Braekers & Duchateau (2017) studied a model for one-level clustered data with Archimedean dependence and variable cluster sizes. The copula approach also underlines the work of Romdhani, Lakhal-Chaieb & Rivest (2014), who propose the exchangeable Kendall’s tau as a nonparametric measure of intracluster association.

In this article, we propose a new copula model for hierarchical clustered data using the hierarchical Kendall copula construction of Brechmann (2014) with Archimedean clusters. As in the HAC model, we use exchangeable Archimedean copulas to describe within-cluster dependence at the first level. However, to model dependence between clusters at subsequent levels of the hierarchy, a different strategy is adopted. Particular variables are chosen as cluster representatives and their dependence is captured by an Archimedean copula model. This construction has several advantages over the HAC model. First, clusters of unequal size are easily accommodated. Second, no conditions on the within-cluster dependence are needed, and no restrictions on the parameter space are required. Moreover, the model is easy to simulate.
from and is fitted sequentially, level by level, when the data are complete. Unlike Brechmann (2014), we allow the data to be right-censored, which makes inference more challenging as the amount of censoring of the cluster representatives increases with the level of the hierarchy. To overcome this problem, we obtain complete observations through an imputation algorithm, and base estimation and model validation on the imputed sample.

The article is organized as follows. In Section 2 we recall the hierarchical Kendall copula construction of Brechmann (2014) with Archimedean clusters. Using this approach, we then propose a new model for hierarchical clustered data in Section 3. Parameter estimation for right-censored observations and the accompanying imputation algorithm are discussed in Section 4, where the asymptotic distribution of the proposed estimators is also derived. Section 5 contains a new goodness-of-fit test that can be used to validate whether the proposed hierarchical Kendall dependence structure with Archimedean clusters is indeed well-suited. The performance of the estimators and the goodness-of-fit procedure is investigated in Section 6 via simulations. In the same section, the new model is applied to the two-level clustered CGD data. Section 7 contains concluding remarks. The Supplementary Material contains details about the marginal estimation, explicit formulas when the Archimedean generators are Clayton, additional simulations, outstanding proofs, R source code and implementation instructions.

2. HIERARCHICAL KENDALL COPULAS WITH ARCHIMEDEAN CLUSTERS

Let \(X = (X_1, \ldots, X_d)\) be a random vector with univariate marginal survival functions \(\bar{F}_1, \ldots, \bar{F}_d\). As stated in the Introduction, the joint survival function of \(X\) can be decomposed as indicated in (1); the copula \(C\) appearing therein is unique if and only if \(\bar{F}_1, \ldots, \bar{F}_d\) are continuous. Continuity of the margins is assumed throughout this article; discontinuity of the marginals causes \(C\) to be unidentifiable, thereby invalidating copula inference procedures developed for continuous data; for discussion, see Genest & Nešlehová (2007), for example.

The class of Archimedean copulas plays a special role in the model proposed here. A \(d\)-variate copula is called Archimedean if it can be expressed, for all \(u_1, \ldots, u_d \in [0, 1]\), as \(C_{\psi,d}(u_1, \ldots, u_d) = \psi\{\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\}\).

The Archimedean generator \(\psi : [0, \infty) \to [0, 1]\) must be nondecreasing and such that \(\psi(0) = 1\) and \(\psi(x) \to 0\) as \(x \to \infty\); \(\psi^{-1}(0) = \inf\{x \geq 0 : \psi(x) = 0\}\) by convention. Furthermore, \(C_{\psi,d}\) is a bona fide copula if and only if \(\psi\) is \(d\)-monotone (Malov, 2001; McNeil & Nešlehová, 2009). This property means that for all \(k \in \{1, \ldots, d - 2\}\), the \(k\)th derivative \(\psi^{(k)}\) of \(\psi\) exists on \((0, \infty)\) and satisfies \((-1)^k\psi^{(k)} \geq 0\), and further that \((-1)^d\psi^{(d-2)}\psi^{(d-2)}\) is nonincreasing and convex. According to McNeil & Nešlehová (2009), the density \(c_{\psi,d}\) of \(C_{\psi,d}\) exists if and only if \(\psi^{(d-1)}\) exists and is absolutely continuous on \((0, \infty)\). Then for all \(u_1, \ldots, u_d \in (0, 1)\)

\[
c_{\psi,d}(u_1, \ldots, u_d) = \psi^{(d)}\{\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\} \prod_{j=1}^{d} \frac{1}{\psi^{(1)}(\psi^{-1}(u_j))}.
\]

In the context of clustered data, Archimedean copulas are particularly well-suited for modelling intracluster dependence. The latter is often assumed to be exchangeable, meaning that for any permutation \(\pi\) of \(\{1, \ldots, d\}\), the survival copulas of \(X\) and \((X_{\pi(1)}, \ldots, X_{\pi(d)})\) are the same. Because \(C_{\psi,d}\) is invariant with respect to any permutation of its arguments, it indeed induces exchangeable dependence. Another appealing property of Archimedean copulas is that when \(\psi\) is a Laplace transform of some positive random variable, \(C_{\psi,d}\) is the dependence structure of a multiplicative frailty model (Marshall & Olkin, 1988).

For the hierarchical Kendall copula construction, it is also important to recall that for any \(d\)-monotone Archimedean generator \(\psi\), \(C_{\psi,d}\) is the survival copula of \(R \times (S_1, \ldots, S_d)\), where the
radial random variable $R$ is strictly positive and independent of the random vector $(S_1, \ldots, S_d)$, which is uniformly distributed on the unit simplex $S_d = \{x \in [0, 1]^d : x_1 + \cdots + x_d = 1\}$ (McNeil & Nešlehová, 2009; Genest, Nešlehová & Ziegel, 2011). The survival function $\tilde{F}_{\psi, d}$ of $R$ is uniquely determined by $\psi$, that is, for any $r > 0,$

$$\tilde{F}_{\psi, d}(r) = \Pr(R > r) = \frac{1}{d} \sum_{k=0}^{d-2} \frac{(-1)^k}{k!} r^k \psi^{(k)}(r) + \frac{(-1)^{d-1}}{(d-1)!} r^{d-1} \psi_+^{(d-1)}(r), \quad (3)$$

where $\psi_+^{(d-1)}$ denotes the right-hand derivative of $\psi^{(d-2)}$. If $U = (U_1, \ldots, U_d)$ is distributed as $C_{\psi, d}$, then $R = \psi^{-1}(U_1) + \cdots + \psi^{-1}(U_d)$ where $\overset{\text{d}}{=} \psi$ denotes equality in distribution. The distribution function $K_{\psi, d}$ of $W = C(U) = \psi(R)$ is the so-called Kendall distribution (Genest & Rivest, 1993; Barbe et al., 1996).

Now consider a partition $C = \{C_1, \ldots, C_n\}$ of $\{1, \ldots, d\}$ into disjoint subsets $C_1, \ldots, C_n$, which specify the clusters. For each $i \in \{1, \ldots, n\}$, denote the cardinality of cluster $C_i$ by $|C_i| = m_i$. The hierarchical Kendall copula with Archimedean clusters is defined as follows (Brechmann, 2014, Remark 2.5).

**Definition 1.** The hierarchical Kendall copula with Archimedean clusters is the distribution function of $(U_1, \ldots, U_d)$ with the following properties:

(i) For all $i \in \{1, \ldots, n\}$, the distribution function of $(U_j, j \in C_i)$ is an Archimedean copula with generator $\psi_i$; let $R_i = \sum_{j \in C_i} \psi_i^{-1}(U_j)$.

(ii) Given $R_1, \ldots, R_n$, the vectors $(U_j, j \in C_i)$ defined for all $i \in \{1, \ldots, n\}$ are mutually independent, and for each $i \in \{1, \ldots, n\}$, the distribution of $(U_j, j \in C_i)$ given $R_1, \ldots, R_n$ depends only on $R_i$.

(iii) The survival copula of $R_1, \ldots, R_n$ is Archimedean with generator $\psi_0$.

Notice that the key idea behind the construction in Definition 1 is that the intercluster dependence is accounted for through the radial variables $R_1, \ldots, R_n$, which can be thought of as cluster representatives.

**Remark 1.** For each $i \in \{1, \ldots, n\}$, the survival function of the variable $R_i$ defined in part (i) of Definition 1 is as specified in (3) with $\psi$ replaced by $\psi_i$ and $d$ by $m_i$. Consequently, the distribution of $W_i = \psi_i(R_i) = C_{\psi_i}(U_j, j \in C_i)$ is the Kendall distribution $K_{\psi_i, m_i}$. Thus the construction in Definition 1 is indeed a hierarchical Kendall copula as defined by Brechmann (2014). In his article the variables $R_1, \ldots, R_n$ in part (i) are replaced by $W_1, \ldots, W_n$, and part (iii) is replaced by (iii*) which states that the copula of $W_1, \ldots, W_n$ is Archimedean with generator $\psi_0$. Note also that although Archimedean copulas are considered here, the construction of Brechmann (2014) allows the copulas in parts (i) and (iii*) to be arbitrary.

The density of the hierarchical Kendall copula with Archimedean clusters exists provided that the copulas $C_{\psi_i, n}$ and $C_{\psi_i, m_i}$ for $i \in \{1, \ldots, n\}$ have densities; these densities are of the form indicated in (2). From Theorem 2.8 in Brechmann (2014), the density $c$ of the hierarchical Kendall copula is as follows. For arbitrary $u_1, \ldots, u_d \in (0, 1)$ and each $i \in \{1, \ldots, n\}$, let $r_i = \sum_{j \in C_i} \psi_i^{-1}(u_j)$. Then

$$c(u_1, \ldots, u_d) = c_{\psi_0, n} \left\{ \tilde{F}_{\psi_1, m_1}(r_1) \, \cdots \, \tilde{F}_{\psi_n, m_n}(r_n) \right\} \prod_{i=1}^{n} c_{\psi_i, |C_i|}(u_j, j \in C_i). \quad (4)$$

DOI: 10.1002/cjs

The Canadian Journal of Statistics / La revue canadienne de statistique
3. A HIERARCHICAL KENDALL COPULA MODEL FOR CLUSTERED DATA

3.1. Description of the Model

For simplicity, we restrict our attention to two-level clustered data, noting that the model could be extended to the multi-level case. Suppose there are \( n \) clusters at the upper level, for example, hospitals, and that for each \( i \in \{1, \ldots, n\} \), cluster \( i \) is divided into \( m_i \) sub-clusters containing \( m_{ij} \) subjects each, for example, patients’ records from hospital \( i \). The sizes of the clusters at these two levels need not be the same.

The failure time and the covariate vector of subject \( \ell \) in sub-cluster \( j \) of upper-level cluster \( i \) are denoted by \( T_{ij\ell} \) and \( Z_{ij\ell} \), respectively. The covariate vector can depend on time but we will write \( Z_{ij\ell} \) instead of \( Z_{ij\ell}(t) \) if no confusion can arise. We assume that for each \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \), and \( \ell \in \{1, \ldots, m_{ij}\} \), the length of the covariate vector \( Z_{ij\ell} \) is the same and is equal to \( p \). The vectors \( T_{ij \ell} = (T_{ij1}, \ldots, T_{ijm_{ij}}) \) defined for all \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \) form the clusters at the lower level, while the clusters at the upper level are given by \( T_i = (T_{i1}, \ldots, T_{im_i}) \) for all \( i \in \{1, \ldots, n\} \). Also let \( T = (T_1, \ldots, T_n) \) be the entire vector of failure times.

As in the CGD clinical study, clustered survival times are typically right-censored. The observed failure times are thus \( X_{ij\ell} = \min(T_{ij\ell}, T_{ij\ell}^c) \) for \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \), and \( \ell \in \{1, \ldots, m_{ij}\} \), where \( T_{ij\ell}^c \) denotes a censoring variable that is assumed independent of \( T_{ij\ell} \) conditional on \( Z_{ij\ell} \). We also observe the censoring indicators \( \delta_{ij\ell} = 1(T_{ij\ell} \leq T_{ij\ell}^c) \), so that the entire observed data are \( \{X_{ij\ell}, \delta_{ij\ell}, Z_{ij\ell}\}, i = 1, \ldots, n; j = 1, \ldots, m_i; \ell = 1, \ldots, m_{ij}\} \).

In a conditional copula model for \( T \) given the covariates, the univariate survival functions are specified first. For each \( t > 0 \), \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \), and \( \ell \in \{1, \ldots, m_{ij}\} \), let \( \tilde{F}_{ij\ell}(t) = \Pr(T_{ij\ell} > t|Z_{ij\ell}(t)) \) be the conditional survival function of subject \( \ell \) in sub-cluster \( j \) of cluster \( i \) given the covariate vector \( Z_{ij\ell}(t) \). Typically, \( \tilde{F}_{ij\ell} \) is specified through a regression model. For example, Shih & Lu (2007) consider the Cox proportional hazard model with hazard function \( \lambda_{ij\ell}(t|Z_{ij\ell}(t)) = \lambda(t)\exp(Z_{ij\ell}(t)\beta) \), where \( \beta \) is a parameter vector of length \( p \) and \( \lambda(t) \) denotes a baseline hazard function. Even though it is not considered here, the baseline hazard function can be allowed to depend on \( i \) (Spierberman & Lin, 1998). Then, for each \( t > 0 \), the marginal survival function \( \bar{F}_{ij\ell} \) is \( \bar{F}_{ij\ell}(t) = \exp(-\int_0^t \lambda(u)du) \), where \( \lambda(t) = \int_0^t \lambda(u)du \).

Now for each \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \), and \( \ell \in \{1, \ldots, m_{ij}\} \), let \( U_{ij\ell} = \tilde{F}_{ij\ell}(T_{ij\ell}) \) and set \( U_{ij} = (U_{ij1}, \ldots, U_{ijm_{ij}}) \), \( U_i = (U_{i1}, \ldots, U_{im_i}) \), and \( U = (U_1, \ldots, U_n) \). Then the survival copula of \( T \), given the covariates, is the distribution function of \( U \). We assume the latter to be as follows.

**Assumption 1.**

(i) The vectors \( U_1, \ldots, U_n \) are mutually independent.

(ii) For each \( i \in \{1, \ldots, n\} \), the distribution function of \( U_i \) is a hierarchical Kendall copula with Archimedean clusters as described in Definition 1; the Archimedean generators at the lower level are denoted by \( \psi_{ij}, j \in \{1, \ldots, m_i\} \), while the generator at the upper level is \( \psi_i \).

(iii) For each \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m_i\} \), \( \psi_{ij} \equiv \psi_1 \) and \( \psi_i \equiv \psi_2 \).

Note that because the cluster lengths need not be the same, the generators \( \psi_1 \) and \( \psi_2 \) specified in part (iii) of Assumption 1 have to be \( d \)-monotone for a sufficiently large \( d \). In particular, for
$k \in \{1, 2\}$, $\psi_k$ has to be $d_k$-monotone, where $d_1 = \max (m_{ij}, i \in \{1, \ldots, n\}, j \in \{1, \ldots, m_i\})$ and $d_2 = \max (m_1, \ldots, m_n)$.

**Example.** In applications to survival data like the one considered in this article, the generators $\psi_1$ and $\psi_2$ specified in part (iii) of Assumption 1 are often taken to be Clayton with parameters $\theta_1$ and $\theta_2$, respectively. For arbitrary $\theta \in [-1/(d - 1), \infty)$ and all $t > 0$, the generator of the $d$-dimensional Clayton copula is by $\psi_{\theta}(t) = \{\max(1 + \theta t, 0)\}^{-1/\theta}$; the case $\theta = 0$ corresponds to the independence copula generator $e^{-t}$. When $\theta > 0$, $\psi_{\theta}$ is the Laplace transform of a Gamma variable $\mathcal{G}(1/\theta, \theta)$ and hence completely monotone. The $k$th derivative of $\psi_{\theta}$ with $\theta > 0$ is explicit (Hofert, Mächler & McNeil, 2012); the relevant formula may be found in Section 2 of Supplementary Material. From (3), the survival function of the radial variable $R$ of the $d$-variate Clayton copula with $\theta > 0$ is given, for all $r > 0$, by

$$\Pr(R > r) = \sum_{k=0}^{d-1} \frac{(\theta r)^k}{k!} \frac{\Gamma(k + 1/\theta)}{\Gamma(1/\theta)} (1 + \theta r)^{-(k+1/\theta)}.$$

When $\psi_1 = \psi_{\theta_1}$ and $\psi_2 = \psi_{\theta_2}$, the parameters $\theta_1$ and $\theta_2$ can typically be interpreted as intercluster and intracluster associations, respectively. For example, for the bivariate Clayton copula, $\theta$ is in one-to-one correspondence with Kendall’s $\tau$, viz. $\tau = \theta/((\theta + 2)$, so that $\theta_1 = 2\tau_1/(1 - \tau_1)$, where $\tau_1$ is Kendall’s $\tau$ between two members of the same cluster at the lower level. Moreover, $\theta_2$ is related to Kendall’s $\tau$ for two members of different clusters at the lower level that belong to the same cluster at the upper level, but in this case no simple formula is available.

### 3.2. Comparison with the HAC Model

In this section we compare the model from Section 3.1 to the HAC model of Shih & Lu (2007). First note that the latter model also assumes mutual independence of $U_{1i}, \ldots, U_{ni}$. Next, for each $i \in \{1, \ldots, n\}$, the HAC model takes the distribution function of $U_i$ to be hierarchical Archimedean. This means that for each $i \in \{1, \ldots, n\}$, the distribution function of the variables $U_{ij}$ with $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$ is the $m_{ij}$-variate Archimedean copula with generator $\psi_1$, as in the model proposed here; we denote the latter by $C_{\psi_1}$, for simplicity. However, for all $j \in \{1, \ldots, m_i\}$ and $u_{ij} \in [0, 1]^{|m_j|}$ the distribution function of $U_{ij}, \ldots, U_{im_i}$ is $\psi_2[u_1^{-1}(C_{\psi_1}(u_{1i})) + \cdots + u_p^{-1}(C_{\psi_1}(u_{p,m_i}))].$ The difficulty with the HAC model is that this expression is not necessarily a copula, unless $\psi_1$ and $\psi_2$ satisfy further conditions; for example, see Joe (1993). When $\psi_1$ and $\psi_2$ are Clayton generators with parameters $\theta_1 > 0$ and $\theta_2 > 0$, respectively, as in Shih & Lu (2007), these conditions hold if $\theta_2 < \theta_1$. This additional constraint on the parameters makes estimating their model more complex. Another disadvantage is that the HAC model treats the vectors $U_{ij}$ and $U_{jk}$ equally for $j \neq k$, even though their dimensions may differ. In contrast, the model that we propose is fully flexible in that $\psi_1$ and $\psi_2$ can be any generators that exhibit a sufficient degree of monotonicity. Furthermore, the different lengths of $U_{ij}$ and $U_{ij}$ are accounted for through the corresponding radial variables, which have different distributions.

The advantage of the HAC model is the more tractable form of intracluster dependence. Let $i \in \{1, \ldots, n\}$, $j_1 \neq j_2 \in \{1, \ldots, m_i\}$, $\ell_1 \in \{1, \ldots, m_{ij_1}\}$, and $\ell_2 \in \{1, \ldots, m_{ij_2}\}$. The distribution function of $U_{ij_1\ell_1}$ and $U_{ij_2\ell_2}$ is the bivariate Archimedean copula with generator $\psi_2$, and the parameter of $\psi_2$ is directly interpretable in terms of intracluster correlation. In the model proposed here, the dependence between $U_{ij_1\ell_1}$ and $U_{ij_2\ell_2}$ is more cumbersome (Brechmann, 2014, Eq. (2.13)). However, in his Example 2.13, Brechmann (2014) shows that it is close to the Archimedean copula $C_{\psi_2,2}$ when the generators are Gumbel.
4. ESTIMATION

To fit the hierarchical Kendall copula model with Archimedean clusters to two-level clustered data, first recall that the available data are the triplets \((X_{ij\epsilon}, \delta_{ij\epsilon}, Z_{ij\epsilon})\) for \(i \in \{1, \ldots, n\}, j \in \{1, \ldots, m_i\}, \) and \(\epsilon \in \{1, \ldots, m_{ij}\}\).

Assume that \(\psi_1\) and \(\psi_2\) in part (iii) of Assumption 1 belong to single-parameter classes of generators \(\Psi_1 = \{\psi_{\theta_1}, \theta_1 \in \Theta_1\}\) and \(\Psi_2 = \{\psi_{\theta_2}, \theta_2 \in \Theta_2\}\), respectively; this is the case for most commonly used Archimedean families. Thus, we need to estimate the marginal parameters \(\beta\) and \(\Lambda\), and the association parameters \(\theta_1\) and \(\theta_2\). Since the density of a hierarchical Kendall copula with Archimedean clusters has the product form identified in (4), we propose the following step-wise approach.

The marginal parameters \(\beta\) and \(\Lambda\) are fitted first as in Spiekerman & Lin (1998) under the working assumption that all subjects are independent; the required estimators are given in Section 1 of the Supplementary Material. For formulas for standard errors (SEs) that take clustering into account, see Spiekerman & Lin (1998).

The estimates \(\hat{\beta}\) and \(\hat{\Lambda}\) then serve to obtain the estimates \(\hat{F}_{ij\epsilon}\) of the conditional marginal survival functions for all \(i \in \{1, \ldots, n\}, j \in \{1, \ldots, m_i\}, \) and \(\epsilon \in \{1, \ldots, m_{ij}\}\). Next, set \(\hat{U}_{ij\epsilon} = \hat{F}_{ij\epsilon}(X_{ij\epsilon}) = \max\{\hat{F}_{ij\epsilon}(T_{ij\epsilon}), \hat{F}_{ij\epsilon}(T_{ij\epsilon}^*)\}\). Because \(\hat{F}_{ij\epsilon}\) is decreasing, \(\hat{U}_{ij\epsilon}\) is left-censored whenever \(\delta_{ij\epsilon} = 0\).

For each \(i \in \{1, \ldots, n\}\) and \(j \in \{1, \ldots, m_i\}\), let \(\hat{U}_{ij} = (\hat{U}_{ij1}, \ldots, \hat{U}_{ijmij})\), \(\hat{U}_i = (\hat{U}_{i1}, \ldots, \hat{U}_{im})\). Because the marginal survival functions are estimated, \(\hat{U} = (\hat{U}_1, \ldots, \hat{U}_n)\) is a noisy, left-censored observation on the vector \(U\) identified in Section 3.1. Nonetheless, it can be used to estimate \(\theta_1\) and \(\theta_2\). Analogous to Algorithm 3.9 in Brechmann (2014), we proceed sequentially.

4.1. Estimation of \(\theta_1\)

Under the working assumption that the lower-level clusters are independent, \(\theta_1\) can be estimated by maximizing the pseudo log-likelihood

\[
L_1(\theta_1) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} L_{ij}(\hat{U}_{ij}) ,
\]

where \(L_{ij}(\hat{U}_{ij})\) is the contribution of sub-cluster \(j\) of cluster \(i\), viz.

\[
L_{ij}(\hat{U}_{ij}) = \ln \left[ (-1)^{d_{ij}} \psi_{\theta_1}^{(d_{ij})} \left\{ \sum_{\ell=1}^{m_{ij}} \psi_{\theta_1}^{-1} \left( \hat{U}_{ij\ell} \right) \right\} \right] - \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell} \ln \left[ -\psi_{\theta_1} \left( \psi_{\theta_1}^{-1} \left( \hat{U}_{ij\ell} \right) \right) \right] ,
\]

where \(d_{ij} = \delta_{ij1} + \cdots + \delta_{ijmij}\) denotes the number of complete observations in the sub-cluster. Note that this procedure for estimating \(\theta_1\) reduces to the method of Shih & Lu (2007) when the generators are Clayton, and to the one of Prenen, Braekers & Duchateau (2017) when \(m_{ij} = 1\) for all \(i \in \{1, \ldots, n\}\) and \(j \in \{1, \ldots, m_i\}\), which coincides with the approach of Andersen (2005) when \(m_{ij} = 2\) for all \(i \in \{1, \ldots, n\}\). The asymptotic behaviour of \(\hat{\theta}_1\) that maximizes (5) follows. A proof, based on the ideas of Prenen, Braekers & Duchateau (2017), is given in the Appendix.

**Theorem 1.** Suppose that \(\tilde{F}_{ij\epsilon}(t) = \exp\{ -e^{\psi_{ij\epsilon}(t)} \psi_{\theta_1}(t) \}\), Assumption 1 holds and the families \(\Psi_1\) and \(\Psi_2\) are correctly specified. Let \((\beta_0, \Lambda_0, \theta_{10})\) denote the true values of \((\beta, \Lambda, \theta_1)\). Under the regularity conditions (C1)–(C5) and conditions (A.2) and (A.4) given in the Appendix, \((\hat{\beta}, \hat{\Lambda}, \hat{\theta}_1)\) is a consistent estimator of \((\beta_0, \Lambda_0, \theta_{10})\) as \(n \to \infty\). Furthermore, as \(n \to \infty\) \(\sqrt{n}(\hat{\theta}_1 - \theta_{10})\)
converges weakly to a centred Gaussian random variable with variance \( \sigma^2_{\hat{\theta}_{10}} = \sigma^2_{\theta}/\{I_1(\theta_{10})\}^2 \); the expressions for \( I_1(\theta_{10}) \) and \( \sigma^2_{\theta} \) are given in the Appendix.

The asymptotic variance of \( \hat{\theta}_1 \) specified in Theorem 1 can be estimated consistently by \( \hat{\sigma}^2_{\theta}/\{I_1(\hat{\theta}_1)\}^2 \); an expression for \( \hat{\sigma}^2_{\theta} \) can be found in the Appendix. Explicit formulas when the generator is Clayton are given in Section 3.2 of the Supplementary Material.

4.2. Estimation of \( \theta_2 \)

Turning to the estimation of \( \theta_2 \), for each \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m_i\} \), we define the variables \( \hat{R}_{ij} = \psi_{\hat{\theta}_1}^{-1}\left( \hat{U}_{ij1} \right) + \cdots + \psi_{\hat{\theta}_1}^{-1}\left( \hat{U}_{ijm_i} \right) \). These are noisy, censored pseudo-observations of the radial variables from (i) of Definition 1 whose survival copula is assumed to be \( C_{\psi_{\theta_2}} \). To estimate \( \theta_2 \), we first propose to transform the variables \( \hat{R}_{ij} \) to the uniform scale, that is, for each \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m_i\} \), set \( \hat{V}_{ij} = \hat{F}_{\psi_{\theta_2}^{-1}}\left( \hat{R}_{ij} \right) \), where for any generator \( \psi \) and integer \( d \geq 1 \), \( \hat{F}_{\psi,d} \) is as indicated in (3). Note that for each \( j \in \{1, \ldots, m_i\} \), \( \hat{V}_{ij} \) is left-censored with censoring indicator \( \delta_{ij}^R = \delta_{ij1} \times \cdots \times \delta_{ijm_i} \).

Because the upper-level clusters are assumed to be independent for each \( i \in \{1, \ldots, n\} \), one can, in principle, maximize a likelihood for censored data based on the pseudo-sample \( \left( \hat{V}_{ij}, \delta_{ij}^R \right), \ i \in \{1, \ldots, n\}, \ j \in \{1, \ldots, m_i\} \) that is constructed analogously to the one identified in (5). However, according to a preliminary simulation that we do not report explicitly, the resulting estimator performs poorly. This is because for any \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m_i\} \), \( \delta_{ij}^R = 0 \) whenever at least one of the observations \( X_{ij1}, \ldots, X_{ijm_i} \) in sub-cluster \( j \) is censored. Thus, the censoring rate of the variables \( \hat{V}_{ij} \) is much greater than that of the variables \( X_{ij} \).

Instead, for each \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m_i\} \), we propose to impute the observation \( \hat{V}_{ij} \) whenever \( \delta_{ij}^R = 0 \). To describe the imputation procedure, assume, without loss of generality, that all censored observations are the first \( h_{ij} = m_i - (\delta_{ij1} + \cdots + \delta_{ijm_i}) \), that is, \( \delta_{ij1} = \cdots = \delta_{ijh_{ij}} = 0 \) and \( \delta_{ij(h_{ij}+1)} = \cdots = \delta_{ijm_i} = 1 \). Denoting the realizations of \( \left( \hat{V}_{ij1}, \ldots, \hat{V}_{ijm_i} \right) \) by \( \left( \hat{u}_{ij1}, \ldots, \hat{u}_{ijm_i} \right) \), Algorithm 1, which we outline below, imputes \( \hat{u}_{ij\ell} \) with \( \delta_{ij\ell} = 0 \) sequentially for \( \ell \in \{1, \ldots, h_{ij}\} \).

Algorithm 1

1. Generate \( \tilde{u}_{ij1} \) from the conditional distribution function

\[
\Pr \left( U_{ij1} \leq u \mid U_{ij1} \leq \tilde{u}_{ij1}, \ldots, U_{ijh_{ij}} \leq \tilde{u}_{ijh_{ij}}, U_{ij(h_{ij}+1)} = \tilde{u}_{ij(h_{ij}+1)}, \ldots, U_{ijm_i} = \tilde{u}_{ijm_i} \right).
\]

2. For \( k \in \{2, \ldots, h_{ij}\} \), draw \( \tilde{u}_{ijk} \) from the conditional distribution function

\[
\Pr \left( U_{ijk} \leq u \mid U_{ij1} = \tilde{u}_{ij1}, \ldots, U_{ij(k-1)} = \tilde{u}_{ij(k-1)}, U_{ijk} \leq \tilde{u}_{ijk}, U_{ij(h_{ij}+1)} = \tilde{u}_{ij(h_{ij}+1)}, \ldots, U_{ijm_i} = \tilde{u}_{ijm_i} \right).
\]

The conditional distributions in Algorithm 1 are explicit and given in Section 3.1 of the Supplementary Material, together with the special case when \( \psi_{\theta_1} \) is Clayton. Because \( \theta_1 \) is unknown, it is first replaced by its estimate \( \hat{\theta}_1 \) derived in Section 4.1.

Now let \( \tilde{U} = \left( \tilde{U}_1, \ldots, \tilde{U}_n \right) \) denote the imputed complete dataset, with \( \tilde{U}_i = \left( U_{i1}, \ldots, U_{imi} \right) \) for \( i \in \{1, \ldots, n\} \) and \( \tilde{U}_{ij} = \left( \tilde{U}_{ij1}, \ldots, \tilde{U}_{ijm_i} \right) \) for \( j \in \{1, \ldots, m_i\} \). Also, for each \( i \in \{1, \ldots, n\} \)
and $j \in \{1, \ldots, m_i\}$, let $\tilde{R}_{ij} = \psi_{\theta_j}^{-1}(\tilde{U}_{ij1}) + \cdots + \psi_{\theta_j}^{-1}(\tilde{U}_{ijm_j})$ and $\tilde{V}_{ij} = \tilde{F}_{\psi_{\theta_j}^{-1}}(\tilde{R}_{ij})$. To estimate $\theta_2$, we propose to maximize the pseudo log-likelihood

$$L_2(\theta_2) = \frac{1}{n} \sum_{i=1}^{n} \ln \left\{ c_{\psi_{\theta_2}, m_i}(\tilde{V}_{i1}, \ldots, \tilde{V}_{imi}) \right\},$$

(6)

where $c_{\psi_{\theta_2}, m_i}$ denotes the density of the $m_i$-dimensional Archimedean copula with generator $\psi_{\theta_2}$ specified in (2). Note that $c_{\psi_{\theta_2}, m_i}$ involves the derivative $\psi_{\theta_2}^{(m_i)}$, which may be intractable for some $i$ when $m_i$ is large. To avoid this problem, one can maximize the pair-wise pseudo log-likelihood (Cox & Reid, 2004), viz.

$$L_2^c(\theta_2) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=j+1}^{m_i} \ln \left\{ c_{\psi_{\theta_2}, 2}(\tilde{V}_{ij}, \tilde{V}_{ik}) \right\},$$

(7)

where $c_{\psi_{\theta_2}, 2}$ is the density of the bivariate Archimedean copula with generator $\psi_{\theta_2}$. Maximizing (7) typically leads to a loss in efficiency; this can be partially remedied by adding a weight function such as $1/(m_i - 1)$; see Joe & Lee (2009).

Theorem 2 establishes the asymptotic normality of the estimator $\hat{\theta}_2$ based on the imputed sample. The proof is similar to that of Theorem 1 and is outlined in Section 2 of the Supplementary Material.

**Theorem 2.** Let $\hat{\theta}_2$ be a value that maximizes either $L_2$ or $L_2^c$ given in (6) and (7), respectively. Also let $\theta_{20}$ be the true value of $\theta_2$. Under the conditions stated in Theorem 1, condition (C6) in the Appendix, and condition (A.8) from Section 2 of the Supplementary Material, $\sqrt{n}(\hat{\theta}_2 - \theta_{20})$ converges weakly to a centred Gaussian random variable whose variance is either $\sigma^2_{\theta_{20}} = \sigma^2(\theta_{20}, \theta_{10})/(I_2(\theta_{20}))^2$ or $\sigma^2_{\theta_{20,c}} = \sigma^2_{\theta_{20}}/\sqrt{\var(\hat{\theta}_2, \theta_{10})}$, depending on whether $L_2$ or $L_2^c$ is used. Expressions for $I_2(\theta_{20})$, $I_{2,c}(\theta_{20})$, $\sigma^2(\theta_{20}, \theta_{10})$, and $\sigma^2_{\theta_{20}, c}$ are given in Section 3.3 of the Supplementary Material, together with explicit formulas in the case of Clayton generators.

In order to reduce the effect of randomness in the imputation procedure, the latter can be repeated $K$ times (Rubin, 1987). Let $\hat{\theta}_{2,k}$ denote the estimator based on the $k$th imputed sample and the maximization of either $L_2$ or $L_2^c$ given in (6) and (7), respectively. The final estimate of $\theta_2$ is then the average

$$\bar{\theta}_{2,K} = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}_{2,k}.$$  

(8)

Using the results of Rubin (1987), we obtain the following results concerning the variance and asymptotic behaviour of $\bar{\theta}_{2,K}$.

**Theorem 3.** Let $\bar{\theta}_{2,K}$ be the estimator identified in (8) for some $K \geq 1$ and assume that $\theta_{20}$ denotes the true value of $\theta_2$. Under the conditions specified in Theorem 2, $\bar{\theta}_{2,K}$ converges in probability to $\theta_{20}$. Furthermore, $\var(\bar{\theta}_{2,K}) = W_K + (1 + 1/K)B_K$, where $B_K = \sum_{k=1}^{K} (\hat{\theta}_{2,k} - \bar{\theta}_{2,K})^2/(K - 1)$ and $W_K = \sum_{k=1}^{K} \var(\hat{\theta}_{2,k})/K$; $\var(\bar{\theta}_{2,K})$ denotes the estimate of the asymptotic variance of $\bar{\theta}_{2,K}$ based on the $k$th imputed sample. Finally, $(\bar{\theta}_{2,K} - \theta_{20})/\sqrt{\var(\bar{\theta}_{2,K})}$
converges weakly to the Student t distribution with \(\nu = (K - 1)[1 + {K/(K + 1)}](W_K/B_K)^2\) degrees of freedom.

In the expression for the variance of \(\hat{\theta}_{2,K}\) given in Theorem 3, \(W_K\) and \(B_K\) reflect the variation among the estimators \(\hat{\theta}_{2,1}, \ldots, \hat{\theta}_{2,K}\) and the variation among imputations, respectively, while the factor \((1 + 1/K)\) accounts for the extra variability of \(\hat{\theta}_{2,K}\) when the imputation number \(K\) is finite. The finite-sample properties of the estimators of the dependence parameters \(\theta_1\) and \(\theta_2\) proposed in this section are investigated in Section 6 via simulations.

5. A GOODNESS-OF-FIT TEST

Before proceeding with the aforementioned simulations, we propose a test to validate the assumption that the dependence structure of \(\mathbf{T}\) is indeed a hierarchical Kendall copula with Archimedean clusters as proposed in Section 3.1. Because a fully consistent test would be too cumbersome, we propose to test the following two hypotheses:

(a) For each \(i \in \{1, \ldots, n\}\) and \(j \in \{1, \ldots, m_i\}\), the distribution of the vector \(U_{ij}\) is an Archimedean copula with generator \(\psi_{\theta_1}, \theta_1 \in \Theta_1\).

(b) For each \(i \in \{1, \ldots, n\}\), the survival copula of the variables \(R_{ij} = \sum_{r=1}^{m_i} \psi_{\theta_1}^{-1}(U_{ijr})\), \(j \in \{1, \ldots, m_i\}\) is Archimedean with generator \(\psi_{\theta_2}, \theta_2 \in \Theta_2\).

To construct a suitable test statistic, recall from Section 2 and McNeil & Nešlehová (2009) that if a vector \((U_1, \ldots, U_d)\) is distributed as an Archimedean copula \(C_{\psi,d}\), \(R = \psi^{-1}(U_1) + \cdots + \psi^{-1}(U_d)\) and \(S = \{\psi^{-1}(U_1), \ldots, \psi^{-1}(U_d)\}/R\) are independent, \(S\) is uniformly distributed on the unit simplex \(\mathbf{s}\) and \(R\) has the survival function specified in (3). In particular, \(R\) is independent of \(Q = \psi^{-1}(U_1)/R\) and \(Q\) is distributed as the ratio of two independent Erlang variables with parameters 1 and \(d - 1\), respectively. Hence, for any \(\tau > 0\), \(\tilde{F}_Q(t) = \text{Pr}(Q > t) = (t + 1)^{1-d}\). Since this expression is independent of \(\psi\), \(Y = (Q + 1)^{1-d}\) is a pivotal statistic. Thus, to test that the distribution of \((U_1, \ldots, U_d)\) is the Archimedean copula \(C_{\psi,d}\), we propose to test that Spearman’s \(\rho\) between \(R\) and \(Q\) is 0, where \(\rho = \text{corr}\{\tilde{F}_Q(Q), \tilde{F}_{\psi,d}(R)\} = \text{corr}\{Y, \tilde{F}_{\psi,d}(R)\}\).

To turn these observations into a test of the hypothesis specified in part (a), suppose that \(\tilde{U}\) is an imputed complete dataset obtained from Algorithm 1. As in Section 4, for each \(i \in \{1, \ldots, n\}\) and \(j \in \{1, \ldots, m_i\}\), define \(\tilde{R}_{ij} = \psi^{-1}_{\hat{\theta}_1}(\tilde{U}_{ij1}) + \cdots + \psi^{-1}_{\hat{\theta}_1}(\tilde{U}_{ijn})\) and \(\tilde{V}_{ij} = \tilde{F}_{\psi_{\hat{\theta}_2},m_i}(\tilde{R}_{ij})\), set \(\tilde{Q}_{ij} = \psi^{-1}_{\hat{\theta}_1}(\tilde{U}_{ij1})/\tilde{R}_{ij}\) and introduce \(\tilde{Y}_{ij} = (\tilde{Q}_{ij} + 1)^{1-m_{ij}}\). Let \(N = \sum_{i=1}^{n} m_i\) and set \(\tilde{V}_{**} = (1/N) \sum_{i=1}^{n} \sum_{j=1}^{m_i} \tilde{V}_{ij}\) and \(\tilde{Y}_{**} = (1/N) \sum_{i=1}^{n} \sum_{j=1}^{m_i} \tilde{Y}_{ij}\). The combined sample version of Spearman’s \(\rho\) is then

\[
\tilde{\rho}_1 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (\tilde{V}_{ij} - \tilde{V}_{**})(\tilde{Y}_{ij} - \tilde{Y}_{**})}{\sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (\tilde{V}_{ij} - \tilde{V}_{**})^2 \sum_{i=1}^{n} \sum_{j=1}^{m_i} (\tilde{Y}_{ij} - \tilde{Y}_{**})^2}}.
\]

Following the approach of Fisher (1915), we use the test statistic

\[
\tilde{z}_1 = \frac{1}{2} \ln \left( \frac{1 + \tilde{\rho}_1}{1 - \tilde{\rho}_1} \right) \quad (9)
\]

DOI: 10.1002/cjs
To test the hypothesis specified in part (b) based on a single imputed sample \( \tilde{U} \), for each \( i \in \{1, \ldots, n\} \), set \( \tilde{R}_i = \psi_{\hat{\theta}_2}^{-1}(\tilde{V}_{1i}) + \cdots + \psi_{\hat{\theta}_2}^{-1}(\tilde{V}_{mi}) \), \( \tilde{V}_i = \tilde{F}_{\psi_{\hat{\theta}_2}, m_i}(\tilde{R}_i) \), \( \tilde{Q}_i = \psi_{\hat{\theta}_2}^{-1}(\tilde{V}_{1i})/\tilde{R}_i \), and \( \tilde{Y}_i = (\tilde{Q}_i + 1)^{-m_i} \). Let \( \tilde{V}_* = \sum_{i=1}^n \tilde{V}_i/n \) and \( \tilde{Y}_* = \sum_{i=1}^n \tilde{Y}_i/n \). Then the sample version of Spearman’s \( \rho \) and the resulting Fisher test statistic are

\[
\tilde{r}_2 = \frac{\sum_{i=1}^n \left( \tilde{V}_i - \tilde{V}_* \right) \left( \tilde{Y}_i - \tilde{Y}_* \right)}{\sqrt{\sum_{i=1}^n \left( \tilde{V}_i - \tilde{V}_* \right)^2 \sum_{i=1}^n \left( \tilde{Y}_i - \tilde{Y}_* \right)^2}} \quad \text{and} \quad \tilde{z}_2 = \frac{1}{2} \ln \left( \frac{1 + \tilde{r}_2}{1 - \tilde{r}_2} \right). \tag{10}
\]

If no censoring is present, \( \sqrt{n} \tilde{z}_1 \) and \( \sqrt{n} \tilde{z}_2 \) are approximately standard Normal under the hypotheses specified in parts (a) and (b). This is illustrated in Figure 1, which displays histograms of \( N = 2,000 \) values of \( \sqrt{n} \tilde{z}_1 \) and \( \sqrt{n} \tilde{z}_2 \) when \( n = 200 \); for each \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m_i\} \), \( m_i \equiv m_1 \in \{10, 30\} \) and \( m_i \equiv m_2 \in \{3, 5\} \). The generators are Clayton with parameters \( \theta_1 = 2 \) and \( \theta_2 = 0.86 \), which correspond to values of 0.5 and 0.3 in terms of Kendall’s \( \tau \), respectively. The marginal model is as described in Section 6.1, and the marginal parameters are estimated as in Spiekerman & Lin (1998), given that the data are complete. The overlaid standard normal density curves show that the normality assumption seems reasonable even if the margins are estimated.

As we indicated in connection with (8), the variability introduced through the imputation procedure can be reduced by taking averages over \( K \) imputation samples. For each \( k \in \{1, \ldots, K\} \), let \( \tilde{z}_1^{(k)} \) and \( \tilde{z}_2^{(k)} \) denote the statistics specified in (9) and (10) corresponding to the \( k \)th imputed sample, respectively. Let \( \tilde{z}_{1,K} = \sum_{k=1}^K \tilde{z}_1^{(k)}/K \) and \( \tilde{z}_{2,K} = \sum_{k=1}^K \tilde{z}_2^{(k)}/K \). According to Rubin (1987),
the variance $D_{1,K}$ of $\tilde{z}_{1,K}$ is

$$D_{1,K} = \frac{1}{K} \sum_{k=1}^{K} \text{var} \left( \tilde{z}_{1}^{(k)} \right) + \frac{K + 1}{K} B_{1,K}, \quad B_{1,K} = \frac{1}{K-1} \sum_{k=1}^{K} \left( \tilde{z}_{1}^{(k)} - \tilde{z}_{1,K} \right)^2,$$

while the variance of $\tilde{z}_{2,K}$ equals $D_{2,K} = (1/n) + B_{2,K}(K+1)/K$ with $B_{2,K} = \{1/(K-1)\} \sum_{k=1}^{K} (\tilde{z}_{2}^{(k)} - \tilde{z}_{2,K})^2$. We can test the hypotheses specified in parts (a) and (b) using the statistics $T_{1,K} = \tilde{z}_{1,K} / \sqrt{D_{1,K}}$ and $T_{2,K} = \tilde{z}_{2,K} / \sqrt{D_{2,K}}$, respectively. Here, $\hat{D}_{1,K}$ is obtained by estimating $\text{var}(\tilde{z}_{1}^{(k)})$ by the bootstrap procedure of Monaco, Cai & Grizzle (2005) in which clusters are sampled with replacement. If the hypotheses specified in parts (a) and (b) hold and $K \geq 2$, Rubin (1987) implies that the distributions of $T_{1,K}$ and $T_{2,K}$ are approximately Student $t$ with degrees of freedom $\nu_1$ and $\nu_2$, respectively, where for $i \in \{1, 2\}$, $\nu_i = \{(K(K-1)D_{1,K})/\{(K + 1)B_{1,K}\}\}^2$.

6. SIMULATION STUDY AND DATA ANALYSIS

In this section, we use simulations to evaluate the finite-sample performance of the parameter estimators from Section 4, and of the goodness-of-fit procedure from Section 5. We also use the hierarchical Kendall copula model to analyze data on CGD.

6.1. A Simulation Study of the Proposed Estimators

Throughout this section, the parameters of the hierarchical Kendall copula model were fixed as follows. For each $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m_i\}$, and $\ell \in \{1, \ldots, m_{ij}\}$, $p = 2$ and $Z_{ij\ell1}$ was uniform on $(0, 1)$ and independent of $Z_{ij\ell2}$, which was Bernoulli(0.5). Given $Z_{ij\ell}$, $T_{ij\ell}$ followed the Cox proportional hazard model with $\beta = (0.5, 1)$; for all $r > 0$ the Weibull baseline hazard function was $\lambda(t) = (at^{a-1})/(b^a)$ with $a = 0.91$ and $b = 52.57$. The latter values were the parameter estimates that we obtained by fitting the Cox proportional hazard model with Weibull baseline hazard function to the CGD data. Finally, the generators were Clayton with $\theta_1 = 2$ and $\theta_2 = 0.857$, corresponding to values for Kendall’s $\tau$ of 0.5 and 0.3, respectively.

To generate the vector $T$ of survival times, we first drew the vector $U$ using Algorithms 3.1 and 3.7 of Brechmann (2014). Then, applying the marginal inverse survival functions yielded $T_{ij\ell} = [\{- \ln(U_{ij\ell})\} e^{-Z_{ij\ell}\beta^T\mu}]^{1/\beta}$. These lifetimes were then censored using independent variables $T_{ij\ell}^c$, uniform on $(0, c)$, where $c$ was chosen to reach the targeted censoring rate of 10%, 30%, or 70%.

The censoring rate, the number of upper-level clusters $n$, and the cluster sizes typically influenced the performance of the estimators $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2), \hat{\theta}_1, \hat{\theta}_{2,K}$, and $\hat{\theta}_{2,K}$, where the last two estimates were obtained by maximizing the pseudo log-likelihood functions corresponding to (6) and (7), respectively. To investigate our proposed methods of estimation in detail, we considered the following three scenarios:

A. $n \in \{50, 100, 200\}, \ m_i \equiv m_1 = 10, \text{ and } m_{ij} \equiv m_2 = 5$;
B. $n = 100, \ m_i \equiv m_1 \in \{10, 30\}, \text{ and } m_{ij} \equiv m_2 \in \{3, 5\}$;
C. $n = 100, \text{ and } m_i \text{ and } m_{ij} \text{ drawn once prior to the simulation study from the shifted Poisson distributions with means 10 and 5, respectively.}$

From preliminary simulation results that we have not reported here, we chose $K = 10$, $K = 30$, and $K = 100$ when the censoring rate was 10%, 30%, and 70%, respectively. Each result was based on 500 simulation runs.
| $(n, m_1, m_2)$ | Estimate | Censoring rate (%)$^a$ | Bias   | SE   | ESE  | RMSE  | CR (%) |
|---------------|----------|------------------------|--------|------|------|-------|-------|
| (50, 10, 5)   | $\hat{\theta}_1$ | 10 | 38 | -0.0328 | 0.1722 | 0.1707 | 0.1738 | 94.7 |
|               |          | 30 | 67 | -0.0224 | 0.1957 | 0.1939 | 0.1952 | 93.6 |
|               | $\bar{\theta}_{2,K}$ | 10 | 38 | -0.0154 | 0.2579 | 0.2784 | 0.2788 | 94.2 |
|               |          | 30 | 67 | -0.1246 | 0.1338 | 0.1257 | 0.1770 | 88.8 |
|               | $\bar{\theta}^c_{2,K}$ | 10 | 38 | -0.0452 | 0.1636 | 0.1597 | 0.1660 | 93.4 |
|               |          | 30 | 67 | -0.0857 | 0.1732 | 0.1716 | 0.1918 | 93.6 |
| (100, 10, 3)  | $\hat{\theta}_1$ | 10 | 26 | -0.0250 | 0.1506 | 0.1476 | 0.1497 | 94.0 |
|               |          | 30 | 55 | -0.0326 | 0.1593 | 0.1561 | 0.1595 | 96.0 |
|               | $\bar{\theta}_{2,K}$ | 10 | 26 | -0.1133 | 0.1096 | 0.1087 | 0.1570 | 87.0 |
|               |          | 30 | 55 | -0.1571 | 0.1698 | 0.1684 | 0.2303 | 78.6 |
|               | $\bar{\theta}^c_{2,K}$ | 10 | 26 | -0.0573 | 0.1141 | 0.1172 | 0.1305 | 92.0 |
|               |          | 30 | 55 | -0.0882 | 0.1227 | 0.1215 | 0.1501 | 89.0 |
| (100, 10, 5)  | $\hat{\theta}_1$ | 10 | 36 | -0.0077 | 0.1184 | 0.1080 | 0.1083 | 96.7 |
|               |          | 30 | 67 | -0.0347 | 0.1259 | 0.1244 | 0.1291 | 94.0 |
|               | $\bar{\theta}_{2,K}$ | 10 | 36 | -0.0111 | 0.1961 | 0.2084 | 0.2084 | 93.0 |
|               |          | 30 | 67 | -0.1066 | 0.1156 | 0.1036 | 0.1486 | 91.0 |
|               | $\bar{\theta}^c_{2,K}$ | 10 | 36 | -0.0383 | 0.1269 | 0.1145 | 0.1207 | 94.0 |
|               |          | 30 | 67 | -0.0738 | 0.1407 | 0.1337 | 0.1527 | 93.2 |
| (100, 30, 5)  | $\hat{\theta}_1$ | 10 | 36 | -0.0159 | 0.0989 | 0.0938 | 0.0951 | 95.3 |
|               |          | 30 | 67 | -0.0204 | 0.1161 | 0.1034 | 0.1054 | 96.0 |
|               | $\bar{\theta}_{2,K}$ | 10 | 36 | -0.0175 | 0.1785 | 0.1811 | 0.1819 | 93.6 |
|               |          | 30 | 67 | -0.1369 | 0.0747 | 0.0738 | 0.1555 | 83.0 |
|               | $\bar{\theta}^c_{2,K}$ | 10 | 36 | -0.0384 | 0.0839 | 0.0833 | 0.0917 | 94.2 |
|               |          | 30 | 67 | -0.0863 | 0.0876 | 0.0866 | 0.1223 | 93.6 |

$^a$Left column: study design censoring rate; right column: average observed censoring rate of $\hat{R}$. 

DOI: 10.1002/cjs
Table 1 shows an excerpt of the results observed under Scenario B; complete tables for all scenarios are provided in Tables S.1–S.5 in Section 4 of the Supplementary Material. For each estimator, we have reported the average bias, the empirical standard error (SE), the average estimated standard error (ESE), the root mean squared error (RMSE), and the empirical coverage rate (CR) of its large-sample 95% confidence interval (CI). We also have reported the average censoring rates $\delta_{ij}^R$ of the cluster representatives $\hat{R}_{ij}$. These values were indeed much higher than the censoring rates of the univariate lifetimes, and this phenomenon became worse as the size of the lower-level clusters increased. Overall, the performance of all estimators was reasonable, although the association parameter estimators seemed to be systematically negatively biased. Reassuringly, we observed that the values for SE and ESE were close, even when $n = 50$.

As expected, all estimators improved as $n$, the sample size, increased. The CR was surprisingly good, even for $n = 50$, except when the censoring rate was high and we used $\hat{\theta}_{2,K}$. As the censoring rate increased, so did the SE of all the estimators. The bias of $\hat{\theta}_{2,K}$ and $\hat{\theta}_{2,K}^c$ also increased with the censoring rate. However, the impact of the latter on the bias of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\Theta}_1$ was less clear; it may even seem that the bias decreased when the censoring rate was 70%. Although such an observation appears counterintuitive, the same phenomenon was observed in the simulation experiments of Andersen (2005), Shih & Lu (2007), and Prenen, Braekers & Duchateau (2017). Notice that the estimators of $\beta_1$ and $\beta_2$ performed better than those of $\Theta_1$ and $\Theta_2$. Moreover, the estimator of $\Theta_1$ performed better than the estimators of $\Theta_2$ in terms of bias but not necessarily in terms of the SE; the RMSE was actually of a similar magnitude.

Scenario B explored the effect of the cluster size when $n = 100$. When the size of the lower-level clusters increased from $m_1 = 3$ to $m_1 = 5$, more information became available. This reduced the bias and especially the SE of the marginal parameter estimates. The bias and SE of $\hat{\theta}_1$ were also reduced. Because the number of clusters at the upper level was unchanged, the bias and SE of the estimates of $\theta_2$ remained similar. However, the CR clearly improved. Similarly, when the size of the upper-level clusters increased from $m_1 = 10$ to $m_1 = 30$, the performance of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\Theta}_1$ improved. The RMSE of $\hat{\theta}_{2,K}$ remained about the same while the RMSE of $\hat{\theta}_{2,K}^c$ improved and became smaller than the RMSE of $\hat{\theta}_{2,K}$ when $m_1 = 30$. This means that when the upper-level clusters are large, the composite likelihood should lead to an estimator with a comparable, if not slightly better, performance.

When cluster sizes varied under Scenario C, the behaviour of all estimators was similar to the case (100, 10, 5), apart from slightly larger SEs.

### 6.2. A Simulation Study of the Proposed Goodness-of-Fit Procedures

To investigate the goodness-of-fit procedures that we proposed in Section 5, we used the same marginal model as the one that we specified in Section 6.1. We used Clayton generators parametrized by Kendall’s $\tau$, viz. $\tau_1 = \theta_1/(\theta_1 + 2)$ and $\tau_2 = \theta_2/(\theta_2 + 2)$, respectively. In addition, $m_i \equiv m_1$ and $m_{ij} \equiv m_2$ for each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_1\}$. All tests were carried out at the 5% level and the results that we report were based on 500 simulation runs and 500 bootstrap replicates for the computation of $\hat{D}_{1,K}$. The test based on $T_{2,K}$ used the estimator $\hat{\theta}_{2,K}^c$.

The level of the tests based on $T_{1,K}$ and $T_{2,K}$ pertaining to the association in the lower-level and upper-level clusters, respectively, is summarized in Table S.7 of the Supplementary Material for various censoring rates, degrees of association, numbers of upper-level clusters, and cluster sizes. We assumed that the number of imputations was $K = 0$ if no observation was censored, and $K = 30$ if the censoring rate was 30%. Even though the margins are estimated, our proposed test held its level fairly well overall. To investigate the power, we chose the generators of the lower and upper level copulas to be both either in the Gumbel or the Frank family, parametrized by Kendall’s $\tau$. The upper left panel of Figure 2 displays the estimated power of the tests based on $T_{1,K}$ and $T_{2,K}$ for various values of $n$, as a function of $\tau_1$, when $\tau_2 = 0.3$. As expected, power
increased with $n$ for each value of $\tau_1$. The power also increased with $\tau_1$, which makes sense given that when $\tau_1 = 0$, the Frank and Gumbel copulas become the independence copula, which also belongs to the Clayton family. One can also see from these plots that the power depended on the alternative, and that it was negatively affected by the censoring rate, although this feature was less pronounced for the test based on $T_{2,K}$. We also noted that the test based on $T_{1,K}$ exhibited greater power; this result was not surprising in view of the fact that this test is based on a larger number of observations. The lower panel of Figure 2 indicates the effect of the cluster sizes $m_1$ and $m_2$ on the estimated power of the test based on $T_{1,K}$; as expected, the power increased with the cluster size.

6.3. Data Analysis

In this section, we report on our use of the hierarchical Kendall copula model with Archimedean clusters to analyze data from the study of CGD. The dataset contains 128 patients from $n = 13$
Table 2: Results of analyzing the CGD study data.

| Covariate   | $\hat{\beta}$ | SE  | 95% CI           | $P$-value ($\beta = 0$) | $P$-value (GOF) |
|-------------|----------------|-----|------------------|------------------------|----------------|
| Treatment   | -1.0860        | 0.2680 | (-1.6112, -0.5607) | <0.0001                |                |

Association parameters

| Parameter | Estimate | SE  | 95% CI           | $P$-value ($\theta_i = 0$) | $P$-value (GOF) |
|-----------|----------|-----|------------------|---------------------------|----------------|
| $\theta_1$ | 0.4447   | 0.1916 | (0.0692, 0.8203) | 0.0202                    | 0.1228         |
| $\theta_{2,K}$ | 0.1570   | 0.1510 | (-0.1389, 0.4529) | 0.2984                    | 0.8909         |
| $\theta_{c,2,K}$ | 0.0780   | 0.0719 | (-0.0629, 0.2189) | 0.2779                    | 0.8785         |

CI: confidence interval; GOF: goodness-of-fit test; SE: standard error.

Hospitals. The number $m_i$ of patients from a given hospital ranges from 4 to 26 with mean 9.8. The length $m_{ij}$ of a single patient’s record ranges from 1 to 8 with mean 1.6. The censoring rate for a patient’s record is around 62.5%. Given that $n$ is small, we ran an additional simulation study using the same values of $n$, $m_i$, and $m_{ij}$ as in the dataset, and the same degree of intercluster and intracluster association, as detailed in Section 4 of the Supplementary Material. The results reported in Table 8 of the Supplementary Material show that the performance of $\hat{\theta}_1$ and $\hat{\theta}_2$ is still reasonable, particularly when the pair-wise likelihood is used. This may seem surprising; the tolerable CR can be attributed to the fact that the association at both the lower and upper level is weak. Note also that while the entire dataset is exploited to estimate the marginal parameters, only patients who experienced at least one infection and one censored time (last observation) contributed to the pseudo log-likelihood specified in (5).

We first evaluated the effect of gamma interferon on the disease. From the upper section of Table 2, we observe that the treatment effect of $\gamma$-IFN was significant; a similar result was reported by Ha & Lee (2005) and Rondeau, Filleul & Joly (2006).

The results of the association analysis are summarized in the lower section of Table 2. The numbers of imputations and bootstrap replicates for the calculation of $\hat{D}_{1,K}$ were $K = 1,000$ and $B = 500$, respectively. For both the lower and upper cluster levels, we chose the generators to be Clayton; from the $P$-values of the tests previously described in Section 5 and displayed in the last column of Table 2, this assumption seems reasonable. The penultimate column shows the $P$-values of the test of the hypothesis that the association is not significant, that is, $\theta_1 = \theta_2 = 0$. We can see that the within-patient association is significant albeit weak, with $\hat{\theta}_1 = 0.4447$ corresponding to a value for Kendall’s $\tau_1$ of 0.18. The same conclusion was reached by Ha & Lee (2005) using a random-effects approach. However, $\hat{\theta}_1$ is not significantly different from 0 when the covariates age and gender are also included in the Cox proportional hazard model; this is not surprising given the small sample size (results not reported here). The estimated association between patients treated at the same hospital is $\hat{\theta}_{2,K} = 0.157$ and $\hat{\theta}_{c,2,K} = 0.078$, corresponding to Kendall’s $\tau_2 = 0.07$ and 0.04, respectively. However, none of these values are significantly different from 0. Our results agree with those reported by Ha & Lee (2005) and Rondeau, Filleul & Joly (2006), who concluded from their random effects modelling that there was no hospital effect.
A. APPENDIX

A.1. Regularity Conditions

Let \( T > 0 \) be the end of study time, and \((\bar{\beta}_0, \Lambda_0, \theta_{10}, \theta_{20})\) refer to the true values of the parameters \((\beta, \Lambda, \theta_1, \theta_2)\). Also, for each \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \), \( \ell' \in \{1, \ldots, m_{ij}\} \), and \( t \geq 0 \), define \( Y_{ij\ell} = 1(X_{ij\ell} \geq t) \). Furthermore, following Spiekerman & Lin (1998), introduce the following notation. For an arbitrary vector \( a \), let \( a^{\otimes 0} = 1 \), \( a^{\otimes 1} = a \), and \( a^{\otimes 2} = a^T a \). For \( r \in \{0, 1, 2\} \), write

\[
S^{(r)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{\ell'=1}^{m_{ij}} Y_{ij\ell}(t) \exp \{ Z_{ij\ell}(u) \beta^T \} Z_{ij\ell}^{\otimes r}(t),
\]

and let

\[
E(\beta, t) = S^{(1)}(\beta, t)/S^{(0)}(\beta, t), \quad e(\beta, t) = s^{(1)}(\beta, t)/s^{(0)}(\beta, t),
\]

\[
V(\beta, t) = \{ S^{(2)}(\beta, t)/S^{(0)}(\beta, t) \} - E(\beta, t)^{\otimes 2}, \quad v(\beta, t) = \{ s^{(2)}(\beta, t)/s^{(0)}(\beta, t) \} - e(\beta, t)^{\otimes 2}.
\]

(C1) \( \beta \in B \), where \( B \) is a compact subset of \( \mathbb{R}^p \), \( \Lambda(T) < \infty \), and for \( k \in \{1, 2\} \), \( \theta_k \in O_k \), where \( O_k \) is a compact subset of \( \Theta_k \).

(C2) For all \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \), \( \ell' \in \{1, \ldots, m_{ij}\} \), \( \Pr(\forall \ell' \in [0,T) T_{ij\ell'}^c \geq t) > \delta_c > 0 \).

(C3) Let \( Z_{ij\ell}(t) = (Z_{ij\ell}(t), \ldots, Z_{ij\ell_p}(t)) \). For all \( i \in \{1, \ldots, n\} \), \( j \in \{1, \ldots, m_i\} \), \( \ell' \in \{1, \ldots, m_{ij}\} \) and \( k \in \{1, \ldots, p\} \) there exists a constant \( B_{Z_k} \) such that \( |Z_{ij\ell_k}(0)| + \int_0^T |dZ_{ij\ell_k}(t)| \leq B_{Z_k} < \infty \) almost surely.

(C4) \( A = \int_0^T v(\beta_0, u) s^{(0)}(\beta_0, u) d\Lambda_0(u) \) is positive definite.

(C5) For all \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m_i\} \), \( E[\ln L_{ij}(\theta_1; \beta_0, \Lambda_0)/L_{ij}(\theta_1^*; \beta_0, \Lambda_0)] \) exists for all \( \theta_1, \theta_1^* \in \Theta_1 \).
For all \(i \in \{1, \ldots, n\}\), \(\mathbb{E}[\ln L_{Z_i}(\theta_2; \theta_{10}, \beta_0, \Lambda_0)/L_{2i}(\theta_{20}; \theta_{10}, \beta_0, \Lambda_0)]\) exists for all \(\theta_2, \theta_{20} \in \Theta_2\), where \(L_{Z_i}\) denotes the ith summand of the joint likelihood specified in (6) or the composite likelihood identified in (7), as appropriate.

Proof of Theorem 1. Consistency: For consistency of \(\hat{\beta}\) and \(\hat{\Lambda}\), see Spiekerman & Lin (1998). To establish the consistency of \(\hat{\theta}_1\), we proceed as in the proof of Theorem 2 in Prenen, Braekers & Duchateau (2017), which uses the ideas of Othus & Li (2010). For each \(i \in \{1, \ldots, n\}\), \(j \in \{1, \ldots, m_i\}\), and \(\ell \in \{1, \ldots, m_{ij}\}\), define

\[
D_{ij\ell} = \frac{\psi_{\theta_1}^{(d_{ij\ell}+1)} \left\{ \sum_{\ell' = 1}^{m_{ij}} \psi_{\theta_1}^{-1} (Q_{ij\ell}) \right\}}{\psi_{\theta_1}^{(d_{ij\ell})} \left\{ \sum_{\ell' = 1}^{m_{ij}} \psi_{\theta_1}^{-1} (Q_{ij\ell}) \right\}} + \delta_{ij\ell} \frac{\psi''_{\theta_1} \left\{ \psi_{\theta_1}^{-1} (Q_{ij\ell}) \right\}}{\psi'_{\theta_1} \left\{ \psi_{\theta_1}^{-1} (Q_{ij\ell}) \right\}} - Q_{ij\ell},
\]

with

\[
Q_{ij\ell} = \exp \left[ - \int_0^\tau Y_{ij\ell} (u) \exp \{Z_{ij\ell} (u) \beta^T \} d\Lambda(u) \right].
\]  

(A.1)

Following Prenen, Braekers & Duchateau (2017), the Hadamard derivative of \(L_1\) with respect to \(\Lambda\) at \((\Gamma - \Lambda)\) of bounded variation in \([0, T]\) is \(\int_0^T \xi_n(\theta_1; \Lambda)(u)d(\Gamma - \Lambda)(u)\), where for any \(u \in [0, T]\),

\[
\xi_n(\theta_1; \Lambda)(u) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} D_{ij\ell} Y_{ij\ell} (u) \exp \{Z_{ij\ell} (u) \beta^T \}. 
\]

Furthermore, the derivative of \(L_1(\theta_1)\) with respect to \(\beta\) is

\[
\xi_n(\theta_1; \beta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} D_{ij\ell} \int_0^\tau Y_{ij\ell} (u) Z_{ij\ell} (u) \exp \{Z_{ij\ell} (u) \beta^T \} d\Lambda(u).
\]

From the regularity condition (C5), the variables \(||Y_{ij\ell} \exp \{Z_{ij\ell} (u) \beta^T \}||_\infty\) and \(||\int_0^T Y_{ij\ell} (u) Z_{ij\ell} (u) \exp \{Z_{ij\ell} (u) \beta^T \} d\Lambda(u)||_\infty\) are bounded. Next, assume that for some \(D > 0\),

\[
||D_{ij\ell}||_\infty < D < \infty,
\]

(A.2)

which holds, for example, for the Clayton generator in view of the regularity condition (C2) (Prenen, Braekers & Duchateau, 2017). Consequently \(||\xi_n(\theta_1; \Lambda)||_\infty\) and \(||\xi_n(\theta_1; \beta)||_\infty\) are bounded. An expansion of \(L_1(\theta_1)\) around \(\beta_0\) and \(\Lambda_0\) is given by

\[
L_1(\theta_1) = L_1^0(\theta_1) + \xi_n(\theta_1; \beta_0)(\hat{\beta} - \beta_0) + \int_0^T \xi_n(\theta_1; \Lambda)(t)d \left( \hat{\Lambda} - \Lambda_0 \right)(t) + R,
\]

where \(L_1^0\) denotes the log-likelihood \(L_1\) with \(\hat{\beta}_0\) and \(\Lambda_0\) instead of \(\hat{\beta}\) and \(\hat{\Lambda}\). As argued in Prenen, Braekers, & Duchateau (2017), \(R = o_p(1)\) as \(||\hat{\beta} - \beta_0||\) and \(||\hat{\Lambda} - \Lambda_0||_\infty\) are \(o_p(1)\) (Spiekerman & Lin, 1998).

DOI: 10.1002/cjs  The Canadian Journal of Statistics / La revue canadienne de statistique
Using the same arguments as in the proof of Theorem 2 in Preneken, Braekers, & Duchateau (2017), for all \( \theta_1 \in \Theta_1 \), it follows that

\[
\sup_{\theta_1 \in \Theta_1} |L_1(\theta_1) - \mathbb{E}\{L_1^0(\theta_1)\}| = o_p(1), \quad \mathbb{E}\{L_1^0(\theta_1)\} = \mathbb{E}\{L_1^0(\theta_{10})\} < 0. \tag{A.3}
\]

Because \( \hat{\theta}_1 \) maximizes \( L_1(\theta_1) \), we have \( 0 \leq L_1(\hat{\theta}_1) - L_1(\theta_{10}) = L_1(\hat{\theta}_1) - \mathbb{E}\{L_1^0(\theta_{10})\} + o_p(1) \) so that, in view of (A.3), \( \mathbb{E}\{L_1^0(\theta_{10})\} - \mathbb{E}\{L_1^0(\hat{\theta}_1)\} \leq L_1(\hat{\theta}_1) - \mathbb{E}\{L_1^0(\hat{\theta}_1)\} + o_p(1) = o_p(1) \). Now take \( \theta_1 \) with \( |\theta_1 - \theta_{10}| \geq \varepsilon \) for some fixed \( \varepsilon > 0 \). By (A.3), there must exist some \( \gamma_\varepsilon \) such that \( \mathbb{E}\{L_1^0(\hat{\theta}_1)\} + \gamma_\varepsilon < \mathbb{E}\{L_1^0(\theta_{10})\} \). It follows that \( \lim_{n \to \infty} \mathbb{P}(\hat{\theta}_1 - \theta_{10} \geq \varepsilon) \leq \lim_{n \to \infty} \mathbb{P}[\mathbb{E}\{L_1^0(\hat{\theta}_1)\} + \gamma_\varepsilon < \mathbb{E}\{L_1^0(\theta_{10})\}] = 0 \).

Normality. Let \( S_{1n}(\theta_1; \hat{\beta}, \hat{\Lambda}) \) be the score function, that is, the derivative of \( L_1(\theta_1) \) with respect to \( \theta_1 \). Its Taylor series expansion around \( \theta_1 \) is

\[
S_{1n}(\hat{\theta}_1; \hat{\beta}, \hat{\Lambda}) = S_{1n}(\theta_{10}; \hat{\beta}, \hat{\Lambda}) + \left( \hat{\theta}_1 - \theta_{10} \right) \frac{dS_{1n}(\theta_1; \hat{\beta}, \hat{\Lambda})}{d\theta_1} \bigg|_{\theta_1=\theta_{10}^*},
\]

where \( \theta_{10}^* \) is between \( \hat{\theta}_1 \) and \( \theta_{10} \). Since \( \hat{\theta}_1 \) is a root of \( S_{1n}(\theta_1; \hat{\beta}, \hat{\Lambda}) \), \( \sqrt{n}(\hat{\theta}_1 - \theta_{10}) = (-dS_{1n}(\theta_1; \hat{\beta}, \hat{\Lambda})/\partial \theta_1)^{-1} \sqrt{n}S_{1n}(\theta_{10}; \hat{\beta}, \hat{\Lambda}) \). The Law of Large Numbers and the consistency of \((\hat{\beta}, \hat{\Lambda}, \theta_{10})\) implies that

\[
- \frac{dS_{1n}}{d\theta_1} \left( \theta_{10}^*, \hat{\beta}, \hat{\Lambda} \right) \overset{\text{Pr}}{\to} I_1(\theta_{10}) = - \lim_{n \to \infty} \frac{dS_{1n}}{d\theta_1} (\theta_{10}, \beta_0, \Lambda_0).
\]

Thus it only remains to show the asymptotic normality of \( \sqrt{n}S_{1n}(\theta_{10}; \hat{\beta}, \hat{\Lambda}) \), which can be established as in the proof of Theorem 3 in Preneken, Braekers, & Duchateau (2017).

Recall the definition of \( Q_{ij}^\varepsilon \) in (A.1) and introduce \( D_{ij}^\varepsilon = (d/d\theta)\partial_{ij}^\varepsilon \), that is,

\[
D_{ij}^\varepsilon = (-Q_{ij}^\varepsilon) \left[ \psi^{(2)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \psi^{-1}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \right]^{-1} \left[ \psi^{(1)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) - \psi^{-1}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \right]
\]

\[
+ \frac{1}{\psi^{(d)}_{\theta_1}} \left\{ \sum_{\varepsilon=1}^{m_{ij}} \psi^{(d-1)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \right\} \psi^{(d)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \left[ \psi^{(d)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \right]^{-2}
\]

\[
- \psi^{(d)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \left[ \sum_{\varepsilon=1}^{m_{ij}} \psi^{(d)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \right] \psi^{(d)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \psi^{(1)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \left[ \psi^{(d)}_{\theta_1} \left( Q_{ij}^\varepsilon \right) \right]^{-2} \psi^{(1)}_{\theta_1} \left( Q_{ij}^\varepsilon \right). \]
The Hadamard derivative of $S_{1n}(\theta_1; \beta, \Lambda)$ with respect to $\Lambda$ at $\Gamma - \Lambda$ of bounded variation in $[0, T]$ can be expressed as $\int_0^T \xi_{n}(\theta_1; \Lambda) u d(\Gamma - \Lambda)(u)$, where

$$\xi_{n}(\theta_1; \Lambda)(u) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_j} D_{ij\ell}^s Y_{ij\ell}(u) \exp \{ Z_{ij\ell}(u) \beta^T \},$$

while the derivative of $S_{1n}(\theta_1; \beta, \Lambda)$ with respect to $\beta$ is given by

$$\xi_{n}(\theta_1; \beta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_j} D_{ij\ell}^s \int_0^T Y_{ij\ell}(u) Z_{ij\ell}(u) \exp \{ Z_{ij\ell}(u) \beta^T \} d\Lambda(u).$$

By the regularity condition (C5), the terms $\|Y_{ij\ell}(u)\beta^T\|_\infty$ and $\|\int_0^T Y_{ij\ell}(u) Z_{ij\ell}(u) \exp \{ Z_{ij\ell}(u) \beta^T \} d\Lambda(u)\|$ are bounded. If for some $D^s > 0$

$$\|D_{ij\ell}^s\|_\infty < D^s < \infty,$$

which again holds for the Clayton generator, the terms $\|\xi_{n}(\theta_1; \Lambda)\|_\infty$ and $\|\xi_{n}(\theta_1; \beta)\|$ are bounded. Using a Taylor series expansion and the same arguments as in Prenen, Braekers, & Duchateau (2017),

$$\sqrt{n} S_{1n}(\theta_1, \beta, \hat{\Lambda}) = \sqrt{n} [S_{1n}(\theta_1, \beta_0, \Lambda_0) + \xi(\theta_1, \beta_0)(\hat{\beta} - \beta_0) + \int_0^T \xi(\theta_0, \Lambda_0)(t)d(\hat{\Lambda}(t) - \Lambda_0(t))] + o_P(1),$$

where $\xi(\theta_1, \Lambda)(t)$ is the pointwise limit of $\xi_{n}(\theta_1, \Lambda)(t)$ and $\xi(\theta_1; \beta) = E\{\xi_{n}(\theta_1; \beta)\}$. By the working independence assumption and the results of Spiekerman & Lin (1998), we find

$$\sqrt{n}(\hat{\beta} - \beta_0) = \Lambda^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{i++} \right) + o_P(1),$$

where for $i \in \{1, \ldots, n\}$,

$$W_{i++} = \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_j} \int_0^T \{ Z_{ij\ell}(u) - e(\beta_0, u) \} dM_{ij\ell}(u)$$

with $M_{ij\ell}(t) = \delta_{ij\ell} I(X_{ij\ell} \leq t) - \int_0^t Y_{ij\ell}(u)e \{ Z_{ij\ell}(u) \beta_0^T \} d\Lambda_0(u)$. Now for each $t \in [0, T]$ and $i \in \{1, \ldots, n\}$, let $M_{i++} = \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_j} M_{ij\ell}$ and

$$h(t) = -\int_0^t e(\beta_0, u)d\Lambda_0(u), \quad \Psi_i(t) = \int_0^t \frac{dM_{i++}(u)}{s^{(0)}(\beta_0, u)} + h^T(t)\Lambda^{-1}W_{i++}.$$

From the proof of Theorem 3 in Spiekerman & Lin (1998) it then follows that

$$\sqrt{n} \left\{ \hat{\Lambda}(t, \hat{\beta}) - \Lambda_0(t) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi_i(t) + o_p(t),$$

\[\text{DOI: 10.1002/cjs} \]
where \( o_p(t) \) is uniform in \( t \leq T \). Substitute (A.6) and (A.7) into the right-hand side of (A.5) and denote the \( i \)th summand of \( S_{1n}(\theta_{10}; \beta_0, \Lambda_0) \) by \( \Phi_i(\theta_{10}; \beta_0, \Lambda_0) \), that is, \( S_{1n}(\theta_{10}; \beta_0, \Lambda_0) = \sum_{i=1}^n \Phi_i(\theta_{10}; \beta_0, \Lambda_0) / n \). It follows that \( \sqrt{n}S_{1n}(\theta_{10}; \hat{\beta}, \hat{\Lambda}) = \sum_{i=1}^n \Phi_i / \sqrt{n} + o_p(1) \) where, for each \( i \in \{1, \ldots, n\} \),

\[
\Phi_i = \Phi_i(\theta_{10}; \beta_0, \Lambda_0) + \xi(\theta_{10}; \beta_0)A^{-1}W_i + \int_0^T \xi(\theta_{10}; \Lambda_0)(t)d\Psi_i(t).
\]

By the Central Limit Theorem, \( \sqrt{n}S_{1n}(\theta_{10}; \hat{\beta}, \hat{\Lambda}) \Rightarrow \mathcal{N}(0, \sigma^2_\Phi) \) as \( n \to \infty \), where \( \Rightarrow \) denotes weak convergence and \( \sigma^2_\Phi = \text{var}(\Phi_1) \). By Slutsky’s Lemma, \( \sqrt{n}(\hat{\beta}_1 - \theta_{10}) \Rightarrow \mathcal{N}(0, \sigma^2_\beta) \) as \( n \to \infty \), where \( \sigma^2_\beta = \sigma^2_\Phi / (I_1(\theta_{10}))^2 \), as we previously claimed. The variance \( \sigma^2_\Phi \) can be estimated by \( \sum_{i=1}^n \hat{\Phi}^2_i / n \), where \( \hat{\Phi}_i \) is obtained from \( \Phi_i \) by replacing all unknown parameters by their estimators; \( I_1(\theta_{10}) \) can be estimated by plugging in the estimator \( \hat{\theta}_1 \).

**ACKNOWLEDGEMENTS**

The authors thank the Editor, the Associate Editor and an anonymous referee whose comments led to a substantial improvement in the article. Chien-Lin Su and Weijing Wang gratefully acknowledge the financial support of the Taiwanese Ministry of Science and Technology. Partial funding in support of Johanna G. Nešlehová’s research was provided by the Natural Sciences and Engineering Research Council of Canada, the Canadian Statistical Sciences Institute, and the Fonds de Recherche du Québec – Nature et Technologies.

**BIBLIOGRAPHY**

Andersen, E. W. (2004). Composite likelihood and two-stage estimation in family studies. *Biostatistics*, 5, 15–30.

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

Barbe, P., Genest, C., Ghoudi, K., & Rémillard, B. (1996). On Kendall’s process. *Journal of Multivariate Analysis*, 58, 197–229.

Brechmann, E. C. (2014). Hierarchical Kendall copulas: Properties and inference. *The Canadian Journal of Statistics*, 42, 78–108.

Cox, D. R. & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91, 729–737.

Durante, F. & Sempi, C. (2016). *Principles of Copula Theory*. CRC Press, Boca Raton, FL.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, 10, 507–521.

Fleming, T. R. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons, New York.

Genest, C. & Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, 37, 475–515.

Genest, C. & Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88, 1034–1043.

Genest, C., Nešlehová, J., & Ziegel, J. (2011). Inference in multivariate Archimedean copula models (with discussion). *TEST*, 20, 223–265.

Ha, I. D. & Lee, Y. (2005). Multilevel mixed linear models for survival data. *Lifetime Data Analysis*, 11, 131–142.

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.

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

Joe, H. (2015). *Dependence Modeling with Copulas*. CRC Press, Boca Raton, FL.
Joe, H. & Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, 100, 670–685.

Mai, J.-F. & Scherer, M. (2017). *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*, 2nd ed. World Scientific Publishing, Hackensack, NJ.

Malov, S. (2001). On finite-dimensional Archimedean copulas. In *Asymptotic Methods in Probability and Statistics with Applications*, N. Balakrishnan, I. Ibragimov, & V. Nevzorov, editors. Birkhäuser, Basel, pp. 19–35.

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

McNeil, A. J. & Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and $\ell_1$-norm symmetric distributions. *Annals of Statistics*, 37, 3059–3097.

Monaco, J., Cai, J., & Grizzle, J. (2005). Bootstrap analysis of multivariate failure time data. *Statistics in Medicine*, 24, 3387–3400.

Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed. Springer, New York.

Othus, M. & Li, Y. (2010). A Gaussian copula model for multivariate survival data. *Statistics in Biosciences*, 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. *Journal of the Royal Statistical Society: Series B*, 79, 483–505.

Romdhani, H., Lakhal-Chaieb, L., & Rivest, L.-P. (2014). An exchangeable Kendall’s tau for clustered data. *The Canadian Journal of Statistics*, 42, 384–403.

Rondeau, V., Filleul, L., & Joly, P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine*, 25, 4036–4052.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

Sastry, N. (1997). A nested frailty model for survival data, with an application to the study of child survival in Northeast Brazil. *Journal of the American Statistical Association*, 92, 426–435.

Shih, J. H. & Lu, S.-E. (2007). Analysis of failure time data with multilevel clustering, with application to the child vitamin A intervention trial in Nepal. *Biometrics*, 63, 673–680.

Shih, J. H. & Lu, S.-E. (2009). Semiparametric estimation of a nested random effects model for the analysis of multi-level clustered failure time data. *Computational Statistics & Data Analysis*, 53, 3864–3871.

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

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

Yau, K. K. W. (2001). Multilevel models for survival analysis with random effects. *Biometrics*, 57, 96–102.

Zhao, Y. & Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, 33, 335–356.

Received 14 July 2017
Accepted 20 October 2018

DOI: 10.1002/cjs

*The Canadian Journal of Statistics / La revue canadienne de statistique*