

Modelling hierarchical clustered censored data with the hierarchical Kendall copula

Chien-Lin Su^{1,2}, Johanna G. Nešlehová^{1*} and Weijing Wang²

¹*Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan (R.O.C.)*

Email: chien-lin.su@mail.mcgill.ca; wjwang@stat.nctu.edu.tw

²*Dept. of Mathematics and Statistics, McGill University, Montréal (QC) Canada*

Email: johanna.neslehova@mcgill.ca

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 multilevel clustering based on the hierarchical Kendall copula model of Brechmann (2014) with Archimedean clusters. This model easily accommodates clusters of unequal size and multiple clustering levels, without 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. *The Canadian Journal of Statistics* xx: 1–21; xx © xx Statistical Society of Canada

Résumé: Cet article s'inspire des travaux de Brechmann (2014) pour proposer un nouveau modèle de copule de Kendall à grappes archimédiennes adapté aux données de survie dotées d'une structure hiérarchique et sujettes à censure à droite. Ce modèle s'adapte facilement à des grappes de tailles inégales et à des hiérarchies multi-niveaux sans imposer de conditions structurelles sur les différents paramètres et copules qu'il comporte. On propose une procédure d'inférence par étapes qui conduit à des estimations convergentes et asymptotiquement gaussiennes sous des conditions de régularité modérées. L'ajustement du modèle s'appuie sur une procédure d'imputation multiple visant à pallier l'accroissement du taux de censure à mesure qu'on s'élève dans la hiérarchie. On propose aussi une façon de tester si la structure de dépendance est archimédienne. Le comportement à taille finie des estimateurs et du test proposés est évalué par voie de simulation. Le modèle est illustré à l'aide de données de survie à deux niveaux provenant d'une étude clinique portant sur la granulomatose chronique.

La revue canadienne de statistique xx: 1–21; xx © xx Société statistique du Canada

* Author to whom correspondence may be addressed.
E-mail: johanna.neslehova@mcgill.ca

1. INTRODUCTION

Multi-level hierarchical clustered data arise in practice whenever the sampling scheme involves a hierarchical structure. Consider for example 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 and 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 hazards model with two nested independent Gamma frailty variables, while a similar model of Yau (2001) relies on independent normal frailties. Rondeau et al. (2006) also use a model similar to that of Sastry (1997) but handle the baseline estimation nonparametrically through spline-smoothing. Shih and Lu (2009) consider a nested random effect proportional hazards model with an unspecified baseline hazard function, and finally Ha and 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, \dots, X_d) can be expressed in terms of its univariate survival functions $\bar{F}_1, \dots, \bar{F}_d$, and a copula C , i.e., a d -variate distribution function with standard uniform univariate margins. Specifically, for all $x_1, \dots, x_d \in \mathbb{R}$, one has

$$\Pr(X_1 > x_1, \dots, X_d > x_d) = C\{\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)\}. \quad (1)$$

Furthermore, for any copula C and margins $\bar{F}_1, \dots, \bar{F}_d$, the expression on the right-hand side is a valid d -variate survival function with marginal survival functions $\bar{F}_1, \dots, \bar{F}_d$. A copula model thus consists of assuming that in Sklar's representation (1), the copula C and the margins $\bar{F}_1, \dots, \bar{F}_d$ 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 and Sempi, 2016; Mai and 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 and 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 further analyzed using Gaussian copula models by Zhao and Joe (2005) and Othus and Li (2010). Very recently, Prenen et al. (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 et al. (2014), who propose the exchangeable Kendall's tau as a nonparametric intracluster association measure.

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. To model dependence between clusters at subsequent levels of the hierarchy, a different strategy is adopted, however. 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 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 paper 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 through simulations in Section 6. In the same section, the new model is applied to the two-level clustered CGD data. Section 7 contains concluding remarks. The Online Supplement contains details about the marginal estimation, explicit formulas when the Archimedean generators are Clayton, additional simulations, outstanding proofs, R code and implementation instructions.

2. HIERARCHICAL KENDALL COPULAS WITH ARCHIMEDEAN CLUSTERS

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

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 $u_1, \dots, u_d \in [0, 1]$, as $C_{\psi,d}(u_1, \dots, u_d) = \psi\{\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)\}$.

The Archimedean generator $\psi : [0, \infty) \rightarrow [0, 1]$ must be non-decreasing and such that $\psi(0) = 1$ and $\psi(x) \rightarrow 0$ as $x \rightarrow \infty$; $\psi^{-1}(0) = \inf\{x \geq 0 : \psi(x) = 0\}$ by convention. Furthermore $C_{\psi,d}$ is a bona fide copula if and only if ψ is d -monotone (Malov, 2001; McNeil and Nešlehová, 2009). This property means that for all $k \in \{1, \dots, d-2\}$, the k th derivative $\psi^{(k)}$ of ψ exists on $(0, \infty)$ and satisfies $(-1)^k \psi^{(k)} \geq 0$, and further that $(-1)^{(d-2)} \psi^{(d-2)}$ is non-increasing and convex. According to McNeil and 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 $(0, \infty)$. One then has, for all $u_1, \dots, u_d \in (0, 1)$,

$$c_{\psi,d}(u_1, \dots, u_d) = \psi^{(d)}\{\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)\} \prod_{j=1}^d \frac{1}{\psi^{(1)}\{\psi^{-1}(u_j)\}}. \quad (2)$$

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 π of $\{1, \dots, d\}$, the survival copulas of \mathbf{X} and $(X_{\pi(1)}, \dots, X_{\pi(d)})$ are the same. Because $C_{\psi,d}$ is invariant with respect to any permutation of its arguments, it indeed induces ex-

changeable dependence. Another appealing property of Archimedean copulas is that when ψ is a Laplace transform of some positive random variable, $C_{\psi,d}$ is the dependence structure of a multiplicative frailty model (Marshall and Olkin, 1988).

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

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

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

Now consider a partition $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_n\}$ of $\{1, \dots, d\}$ into disjoint subsets $\mathcal{C}_1, \dots, \mathcal{C}_n$, which specify the clusters. For each $i \in \{1, \dots, n\}$, denote the cardinality of cluster \mathcal{C}_i by $|\mathcal{C}_i| = m_i$. 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, \dots, U_d) with the following properties:*

- (i) *For all $i \in \{1, \dots, n\}$, the distribution function of $(U_j, j \in \mathcal{C}_i)$ is an Archimedean copula with generator ψ_i ; let $R_i = \sum_{j \in \mathcal{C}_i} \psi_i^{-1}(U_j)$.*
- (ii) *Given R_1, \dots, R_n , the vectors $(U_j, j \in \mathcal{C}_i)$ defined for all $i \in \{1, \dots, n\}$ are mutually independent, and for each $i \in \{1, \dots, n\}$, the distribution of $(U_j, j \in \mathcal{C}_i)$ given R_1, \dots, R_n depends only on R_i .*
- (iii) *The survival copula of R_1, \dots, R_n is Archimedean with generator ψ_0 .*

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

Remark 1. *For each $i \in \{1, \dots, n\}$, the survival function of the variable R_i defined in (i) in Definition 1 is as in Equation (3) with ψ replaced by ψ_i and d by m_i . Consequently, the distribution of $W_i = \psi_i(R_i) = C_{\psi_i}(U_j, j \in \mathcal{C}_i)$ is the Kendall distribution K_{ψ_i, m_i} . The construction in Definition 1 is thus indeed a hierarchical Kendall copula as defined by Brechmann (2014). In his paper, the variables R_1, \dots, R_n are replaced by W_1, \dots, W_n in (i), and (iii) is replaced by (iii*) stating that the copula of W_1, \dots, W_n is Archimedean with generator ψ_0 . Note also that although Archimedean copulas are considered here, the construction of Brechmann (2014) allows the copulas in (i) and (iii*) to be arbitrary.*

The density of the hierarchical Kendall copula with Archimedean clusters exists provided that the copulas $C_{\psi_0,n}$ and C_{ψ_i, m_i} for $i \in \{1, \dots, n\}$ have densities; these densities are of the form (2). From Theorem 2.8 in Brechmann (2014), the density c of the hierarchical Kendall copula is then as follows. For arbitrary $u_1, \dots, u_d \in (0, 1)$ and each $i \in \{1, \dots, n\}$, let $r_i =$

$\sum_{j \in \mathcal{C}_i} \psi_i^{-1}(u_j)$. Then

$$c(u_1, \dots, u_d) = c_{\psi_0, n} \{ \bar{F}_{\psi_1, m_1}(r_1), \dots, \bar{F}_{\psi_n, m_n}(r_n) \} \prod_{i=1}^n c_{\psi_i, |\mathcal{C}_i|}(u_j, j \in \mathcal{C}_i). \quad (4)$$

Definition 1 can be extended to multiple levels of hierarchy by assuming that the survival copula of R_1, \dots, R_n in (iii) is again a hierarchical Kendall copula with Archimedean clusters; see also Remark 2.7 in Brechmann (2014).

3. 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 that there are n clusters at the upper level (e.g., hospitals), and that for each $i \in \{1, \dots, n\}$, the i th cluster is divided into m_i sub-clusters containing m_{ij} subjects each (e.g., patients' records from the i th hospital). The lengths of the clusters at the upper and lower level respectively need not be the same.

The failure time and the covariate vector of the subject ℓ in the j th sub-cluster of the i th upper-level cluster are denoted by $T_{ij\ell}$ and $\mathbf{Z}_{ij\ell}$, respectively. The covariate vector can depend on time but we write $\mathbf{Z}_{ij\ell}$ instead of $\mathbf{Z}_{ij\ell}(t)$ if no confusion can arise. We assume that for each $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ and $\ell \in \{1, \dots, m_{ij}\}$, the length of the covariate vector $\mathbf{Z}_{ij\ell}$ is the same and equal to p . The vectors $\mathbf{T}_{ij} = (T_{ij1}, \dots, T_{ijm_{ij}})$ defined for all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ form the clusters at the lower level, while the clusters at the upper level are given by $\mathbf{T}_i = (\mathbf{T}_{i1}, \dots, \mathbf{T}_{im_i})$ for all $i \in \{1, \dots, n\}$. Also let $\mathbf{T} = (\mathbf{T}_1, \dots, \mathbf{T}_n)$ 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 $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$, and $\ell \in \{1, \dots, m_{ij}\}$, where $T_{ij\ell}^c$ denotes a censoring variable that is assumed independent of $T_{ij\ell}$ conditional on $\mathbf{Z}_{ij\ell}$. We further observe the censoring indicators $\delta_{ij\ell} = \mathbf{1}(T_{ij\ell} \leq T_{ij\ell}^c)$, so that the entire available data are $\{(X_{ij\ell}, \delta_{ij\ell}, \mathbf{Z}_{ij\ell}), i = 1, \dots, n; j = 1, \dots, m_i; \ell = 1, \dots, m_{ij}\}$.

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

Now for each $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ and $\ell \in \{1, \dots, m_{ij}\}$, let $U_{ij\ell} = \bar{F}_{ij\ell}(T_{ij\ell})$ and set $\mathbf{U}_{ij} = (U_{ij1}, \dots, U_{ijm_{ij}})$, $\mathbf{U}_i = (\mathbf{U}_{i1}, \dots, \mathbf{U}_{im_i})$, and $\mathbf{U} = (\mathbf{U}_1, \dots, \mathbf{U}_n)$. The survival copula of \mathbf{T} given the covariates is then the distribution function of \mathbf{U} . We assume the latter to be as follows.

Assumption 1.

(i) The vectors $\mathbf{U}_1, \dots, \mathbf{U}_n$ are mutually independent.

- (ii) For each $i \in \{1, \dots, n\}$, the distribution function of \mathbf{U}_i is a hierarchical Kendall copula with Archimedean clusters as given in Definition 1; the Archimedean generators at the lower level are denoted by ψ_{ij} , $j \in \{1, \dots, m_i\}$, while the generator at the upper level is ψ_i .
- (iii) For each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, $\psi_{ij} \equiv \psi_1$ and $\psi_i \equiv \psi_2$.

Note that because the cluster lengths need not be the same, the generators ψ_1 and ψ_2 in (iii) of Assumption 1 have to be d -monotone for a sufficiently large d . In particular, for $k \in \{1, 2\}$, ψ_k has to be d_k -monotone, where $d_1 = \max(m_{ij}, i \in \{1, \dots, n\}, j \in \{1, \dots, m_i\})$ and $d_2 = \max(m_1, \dots, m_n)$.

Example 1. In applications to survival data like the one considered in this paper, the generators ψ_1 and ψ_2 in (iii) of Assumption 1 are often taken to be Clayton with parameters θ_1 and θ_2 , respectively. For arbitrary $\theta \in [-1/(d-1), \infty)$, the generator of the d -dimensional Clayton copula is given for all $t > 0$ by $\psi_\theta(t) = \{\max(1 + \theta t, 0)\}^{-1/\theta}$; the case $\theta = 0$ corresponds to the independence copula generator e^{-t} . When $\theta > 0$, ψ_θ is the Laplace transform of a Gamma variable $\mathcal{G}(1/\theta, \theta)$ and hence completely monotone. The k th derivative of ψ_θ with $\theta > 0$ is explicit (Hofert et al., 2012); the formula may be found in Section 2 of the Online Supplement. From Equation (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 θ_1 and θ_2 can typically be interpreted through inter- and intra-cluster associations, respectively. For example, for the bivariate Clayton copula, θ is in one-to-one correspondence with Kendall's τ , viz. $\tau = \theta/(\theta + 2)$, so that $\theta_1 = 2\tau_1/(1 - \tau_1)$, where τ_1 is Kendall's τ between two members of the same cluster at the lower level. Moreover, θ_2 is related to Kendall's τ of two members of different clusters at the lower level that belong to the same cluster at the upper level, but 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 and Lu (2007). First note that the latter model also assumes mutual independence of $\mathbf{U}_1, \dots, \mathbf{U}_n$. Next, for each $i \in \{1, \dots, n\}$, the HAC model takes the distribution function of \mathbf{U}_i to be hierarchical Archimedean. This means that for each $i \in \{1, \dots, n\}$, the distribution function of the variables \mathbf{U}_{ij} with $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, is the m_{ij} -variate Archimedean copula with generator ψ_1 , as in the model proposed here; we denote the latter by C_{ψ_1} for simplicity. However, the distribution function of $\mathbf{U}_{i1}, \dots, \mathbf{U}_{im_i}$ is given, for all $j \in \{1, \dots, m_i\}$ and $\mathbf{u}_{ij} \in [0, 1]^{m_{ij}}$, by $\psi_2[\psi_2^{-1}\{C_{\psi_1}(\mathbf{u}_{i1})\} + \dots + \psi_2^{-1}\{C_{\psi_1}(\mathbf{u}_{im_i})\}]$. The difficulty with the HAC model is that this expression is not necessarily a copula, unless ψ_1 and ψ_2 satisfy further conditions, see, e.g., Joe (1993). When ψ_1 and ψ_2 are Clayton generators with parameters $\theta_1 > 0$ and $\theta_2 > 0$, respectively, as in Shih and Lu (2007), these conditions hold if $\theta_2 < \theta_1$. This additional constraint on the parameters makes estimation of their model more intricate. Another disadvantage is that the HAC model treats the vectors \mathbf{U}_{ij} and \mathbf{U}_{ik} equally for $j \neq k$, even though their dimensions may differ. In contrast, the model advocated here is fully flexible in that ψ_1 and ψ_2 can be any generators of a sufficient degree of monotonicity. Furthermore, the different lengths of \mathbf{U}_{ij} and \mathbf{U}_{ik} are accounted for through the corresponding radial variables, which have different distributions.

The advantage of the HAC model is the more tractable form of intra-cluster dependence. Let $i \in \{1, \dots, n\}$, $j_1 \neq j_2 \in \{1, \dots, m_i\}$, $\ell_1 \in \{1, \dots, m_{ij_1}\}$, and $\ell_2 \in \{1, \dots, m_{ij_2}\}$. The distribution function of $U_{ij_1\ell_1}$ and $U_{ij_2\ell_2}$ is the bivariate Archimedean copula with generator ψ_2 , and

the parameter of ψ_2 is directly interpretable in terms of intra-cluster 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, Equation (2.13)). However, Brechmann (2014) shows in his Example 2.13 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 $(X_{ij\ell}, \delta_{ij\ell}, \mathbf{Z}_{ij\ell})$ for $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ and $\ell \in \{1, \dots, m_{ij}\}$.

Assume that ψ_1 and ψ_2 in (iii) of Assumption 1 are in 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. To be estimated are thus the marginal parameters β and Λ , and the association parameters θ_1 and θ_2 . As the density of a hierarchical Kendall copula with Archimedean clusters has the product form (4), we propose the following step-wise approach.

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

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

For each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, let $\hat{\mathbf{U}}_{ij} = (\hat{U}_{ij1}, \dots, \hat{U}_{ijm_{ij}})$, $\hat{\mathbf{U}}_i = (\hat{U}_{i1}, \dots, \hat{U}_{im_i})$. Because the marginal survival functions are estimated, $\hat{\mathbf{U}} = (\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_n)$ is a noisy, left-censored observation of the vector \mathbf{U} from Section 3.1. Nonetheless, it can be used to estimate θ_1 and θ_2 . Similarly to Algorithm 3.9 in Brechmann (2014), we proceed sequentially.

4.1. Estimation of θ_1

Under the working assumption that the lower-level clusters are independent, θ_1 can be estimated by maximizing the pseudo-loglikelihood

$$L_1(\theta_1) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} L_{ij}(\hat{\mathbf{U}}_{ij}), \quad (5)$$

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

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

where $d_{ij} = \delta_{ij1} + \dots + \delta_{ijm_{ij}}$ denotes the number of uncensored observations in the sub-cluster. Note that this procedure for estimating θ_1 reduces to the method of Shih and Lu (2007) when the generators are Clayton, and to the one of Preneen et al. (2017) when $m_{ij} = 1$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, which coincides with the approach of Andersen (2005) when $m_i = 2$ for all $i \in \{1, \dots, n\}$. The asymptotic behavior of $\hat{\theta}_1$ that maximizes Equation (5) follows. The proof, based on the ideas of Preneen et al. (2017), is given in the Appendix.

Theorem 1. Suppose that $\bar{F}_{ij\ell}(t) = \exp\{-e^{\mathbf{Z}_{ij\ell}(t)\boldsymbol{\beta}^\top} \Lambda(t)\}$, Assumption 1 holds and the families Ψ_1 and Ψ_2 are correctly specified. Let $(\boldsymbol{\beta}_0, \Lambda_0, \theta_{10})$ denote the true values of $(\boldsymbol{\beta}, \Lambda, \theta_1)$. Under the regularity conditions (C1)–(C5) and conditions (A.2) and (A.4) given in the Appendix, $(\hat{\boldsymbol{\beta}}, \hat{\Lambda}, \hat{\theta}_1)$ is a consistent estimator of $(\boldsymbol{\beta}_0, \Lambda_0, \theta_{10})$ as $n \rightarrow \infty$. Furthermore, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\theta}_1 - \theta_{10})$ converges weakly to a centred Gaussian random variable with variance $\sigma_{\hat{\theta}_1}^2 = \sigma_{\Phi}^2 / \{I_1(\theta_{10})\}^2$; the expressions for $I_1(\theta_{10})$ and σ_{Φ}^2 are given in the Appendix.

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

4.2. Estimation of θ_2

Turning to the estimation of θ_2 , define, for each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, the variables $\hat{R}_{ij} = \psi_{\hat{\theta}_1}^{-1}(\hat{U}_{ij1}) + \dots + \psi_{\hat{\theta}_1}^{-1}(\hat{U}_{ijm_{ij}})$. 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 θ_2 , we first propose to transform the variables \hat{R}_{ij} to the uniform scale, i.e., set, for each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, $\hat{V}_{ij} = \bar{F}_{\psi_{\hat{\theta}_1}, m_{ij}}(\hat{R}_{ij})$, where for any generator ψ and integer $d \geq 1$, $\bar{F}_{\psi, d}$ is as in Equation (3). Note that for each $j \in \{1, \dots, m_i\}$, \hat{V}_{ij} is left-censored with censoring indicator $\delta_{ij}^R = \delta_{ij1} \times \dots \times \delta_{ijm_{ij}}$.

Because the upper-level clusters are taken to be independent and for each $i \in \{1, \dots, n\}$, one can, in principle, maximize a likelihood for censored data based on the pseudo-sample $(\hat{V}_{ij}, \delta_{ij}^R)$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ constructed analogously to the one in Equation (5). According to a preliminary simulation not shown here, the resulting estimator performs poorly, however. This is because for any $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, $\delta_{ij}^R = 0$ whenever at least one of the observations $X_{ij1}, \dots, X_{ijm_{ij}}$ in the j th sub-cluster is censored. The censoring rate of the variables \hat{V}_{ij} is thus much higher than that of the variables $X_{ij\ell}$.

Instead, we propose, for each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, 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 $h_{ij} = m_{ij} - (\delta_{ij1} + \dots + \delta_{ijm_{ij}})$, i.e., $\delta_{ij1} = \dots = \delta_{ijh_{ij}} = 0$ and $\delta_{ij(h_{ij}+1)} = \dots = \delta_{ijm_{ij}} = 1$. Denoting the realizations of $(\hat{U}_{ij1}, \dots, \hat{U}_{ijm_{ij}})$ by $(\hat{u}_{ij1}, \dots, \hat{u}_{ijm_{ij}})$, Algorithm 1 below imputes $\hat{u}_{ij\ell}$ with $\delta_{ij\ell} = 0$ sequentially for $\ell \in \{1, \dots, h_{ij}\}$.

Algorithm 1 Imputation of missing observations

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

$$\Pr(U_{ij1} \leq u | U_{ij1} \leq \hat{u}_{ij1}, \dots, U_{ijh_{ij}} \leq \hat{u}_{ijh_{ij}}, \\ U_{ij(h_{ij}+1)} = \hat{u}_{ij(h_{ij}+1)}, \dots, U_{ijm_{ij}} = \hat{u}_{ijm_{ij}}).$$

- 2) For $k \in \{2, \dots, h_{ij}\}$, draw \tilde{u}_{ijk} from the conditional distribution function

$$\Pr(U_{ijk} \leq u | U_{ij1} = \tilde{u}_{ij1}, \dots, U_{ij(k-1)} = \tilde{u}_{ij(k-1)}, \\ U_{ijk} \leq \hat{u}_{ijk}, \dots, U_{ijh_{ij}} \leq \hat{u}_{ijh_{ij}}, U_{ij(h_{ij}+1)} = \hat{u}_{ij(h_{ij}+1)}, \dots, U_{ijm_{ij}} = \hat{u}_{ijm_{ij}}).$$

The conditional distributions in Algorithm 1 are explicit and given in Section 3.1 of the Online Supplement along with the special case when ψ_{θ_1} is Clayton. Because θ_1 is unknown, it is first replaced by its estimate $\hat{\theta}_1$ from Section 4.1.

Now let $\tilde{U} = (\tilde{U}_1, \dots, \tilde{U}_n)$ denote the imputed complete data set, with $\tilde{U}_i = (\tilde{U}_{i1}, \dots, \tilde{U}_{im_i})$ for $i \in \{1, \dots, n\}$ and $\tilde{U}_{ij} = (\tilde{U}_{ij1}, \dots, \tilde{U}_{ijm_{ij}})$ for $j \in \{1, \dots, m_i\}$. Also, for each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, let $\tilde{R}_{ij} = \psi_{\hat{\theta}_1}^{-1}(\tilde{U}_{ij1}) + \dots + \psi_{\hat{\theta}_1}^{-1}(\tilde{U}_{ijm_{ij}})$ and $\tilde{V}_{ij} = \bar{F}_{\psi_{\hat{\theta}_1}, m_{ij}}(\tilde{R}_{ij})$. To estimate θ_2 , we propose to maximize the pseudo-loglikelihood

$$L_2(\theta_2) = \frac{1}{n} \sum_{i=1}^n \ln\{c_{\psi_{\theta_2}, m_i}(\tilde{V}_{i1}, \dots, \tilde{V}_{im_i})\}, \quad (6)$$

where $c_{\psi_{\theta_2}, m_i}$ denotes the density of the m_i -dimensional Archimedean copula with generator ψ_{θ_2} as given in Equation (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-loglikelihood (Cox and Reid, 2004), viz.

$$L_2^c(\theta_2) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=j+1}^{m_i} \ln\{c_{\psi_{\theta_2}, 2}(\tilde{V}_{ij}, \tilde{V}_{ik})\}, \quad (7)$$

where $c_{\psi_{\theta_2}, 2}$ is the density of the bivariate Archimedean copula with generator ψ_{θ_2} . Maximizing Equation (7) typically leads to a loss in efficiency; this can be partly remedied by adding a weight function such as $1/(m_i - 1)$ (Joe and Lee, 2009).

Asymptotic normality of the estimator $\hat{\theta}_2$ based on the imputed sample is stated next. The proof is similar to that of Theorem 1 and reported in Section 2 of the Online Supplement.

Theorem 2. *Let $\hat{\theta}_2$ be a value that maximizes either L_2 or L_2^c given in Equations (6) and (7), respectively. Also let θ_{20} be the true value of θ_2 . Under the conditions of Theorem 1, condition (C6) in the Appendix, and condition (A.8) from Section 2 of the Online Supplement, $\sqrt{n}(\hat{\theta}_2 - \theta_{20})$ converges weakly to a centred Gaussian random variable whose variance is either $\sigma_{\theta_{20}}^2 = \sigma^2(\theta_{20}, \theta_{10})/\{I_2(\theta_{20})\}^2$ or $\sigma_{\theta_{20}, c}^2 = \sigma_c^2(\theta_{20}, \theta_{10})/\{I_{2,c}(\theta_{20})\}^2$, depending on whether L_2 or L_2^c is used. The expressions for $I_2(\theta_{20})$, $I_{2,c}(\theta_{20})$, $\sigma^2(\theta_{20}, \theta_{10})$, and $\sigma_c^2(\theta_{20}, \theta_{10})$ are given in Section 3.3 of the Online Supplement, with explicit formulas for 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 Equations (6) and (7), respectively. The final estimate of θ_2 is then the average

$$\bar{\theta}_{2,K} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{2,k}. \quad (8)$$

Using Rubin (1987), the variance and asymptotic behavior of $\bar{\theta}_{2,K}$ are as follows.

Theorem 3. *Let $\bar{\theta}_{2,K}$ be as in Equation (8) for some $K \geq 1$ and θ_{20} denote the true value of θ_2 . Under the hypothesis of Theorem 2, $\bar{\theta}_{2,K}$ converges in probability to θ_{20} . Furthermore, $\text{var}(\bar{\theta}_{2,K}) = W_K + (1 + 1/K)B_K$, with $B_K = \sum_{k=1}^K (\hat{\theta}_{2,k} - \bar{\theta}_{2,K})^2/(K - 1)$ and $W_K = \sum_{k=1}^K \text{var}(\hat{\theta}_{2,k})/K$, where $\text{var}(\hat{\theta}_{2,k})$ denotes the estimate of the asymptotic variance of*

$\hat{\theta}_2$ based on the k th imputed sample. Finally, $(\bar{\theta}_{2,K} - \theta_{20})/\sqrt{\text{var}(\bar{\theta}_{2,K})}$ 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 $\bar{\theta}_{2,K}$ given in Theorem 3, W_K and B_K reflect the variation within the estimators $\hat{\theta}_{2,1}, \dots, \hat{\theta}_{2,K}$ and the variation among imputations, respectively, while the factor $(1 + 1/K)$ accounts for the extra variability of $\bar{\theta}_{2,K}$ when the imputation number K is finite. The finite-sample properties of the estimators of the dependence parameters θ_1 and θ_2 proposed in this section are investigated via simulations in Section 6.

5. A GOODESS-OF-FIT TEST

Before proceeding with 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 $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, the distribution of the vector \mathbf{U}_{ij} is an Archimedean copula with generator ψ_{θ_1} , $\theta_1 \in \Theta_1$.
- (b) For each $i \in \{1, \dots, n\}$, the survival copula of the variables $R_{ij} = \sum_{\ell=1}^{m_{ij}} \psi_{\theta_1}^{-1}(U_{ij\ell})$, $j \in \{1, \dots, m_i\}$ is Archimedean with generator ψ_{θ_2} , $\theta_2 \in \Theta_2$.

To construct a suitable test statistic, recall from Section 2 and McNeil and Nešlehová (2009) that if a vector (U_1, \dots, U_d) is distributed as an Archimedean copula $C_{\psi,d}$, $R = \psi^{-1}(U_1) + \dots + \psi^{-1}(U_d)$ and $\mathbf{S} = (\psi^{-1}(U_1), \dots, \psi^{-1}(U_d))/R$ are independent, \mathbf{S} is uniformly distributed on the unit simplex \mathcal{S}_d and R has survival function given in Equation (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 $t > 0$, $\bar{F}_Q(t) = \Pr(Q > t) = (t + 1)^{1-d}$. As this expression is independent of ψ , $Y = (Q + 1)^{1-d}$ is a pivotal statistic. To test that the distribution of (U_1, \dots, U_d) is the Archimedean copula $C_{\psi,d}$, we thus propose to test that Spearman's ρ between R and Q is 0, where $\rho = \text{corr}\{\bar{F}_Q(Q), \bar{F}_{\psi,d}(R)\} = \text{corr}\{Y, \bar{F}_{\psi,d}(R)\}$.

To turn these observations into a test of Hypothesis (a), suppose that $\tilde{\mathbf{U}}$ is an imputed complete data set obtained from Algorithm 1. As in Section 4, define, for each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, $\tilde{R}_{ij} = \psi_{\hat{\theta}_1}^{-1}(\tilde{U}_{ij1}) + \dots + \psi_{\hat{\theta}_1}^{-1}(\tilde{U}_{ijm_{ij}})$ and $\tilde{V}_{ij} = \bar{F}_{\psi_{\hat{\theta}_1}, m_{ij}}(\tilde{R}_{ij})$, set $\tilde{Q}_{ij} = \psi_{\hat{\theta}_1}^{-1}(\tilde{U}_{ij1})/\tilde{R}_{ij}$ and introduce $\tilde{Y}_{ij} = (\tilde{Q}_{ij} + 1)^{1-m_{ij}}$. Let $N = \sum_{i=1}^n m_i$ and set $\tilde{\mathbf{V}}_{\bullet\bullet} = (1/N) \sum_{i=1}^n \sum_{j=1}^{m_i} \tilde{V}_{ij}$ and $\tilde{\mathbf{Y}}_{\bullet\bullet} = (1/N) \sum_{i=1}^n \sum_{j=1}^{m_i} \tilde{Y}_{ij}$. The combined sample version of Spearman's ρ is then

$$\bar{r}_1 = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (\tilde{V}_{ij} - \tilde{\mathbf{V}}_{\bullet\bullet})(\tilde{Y}_{ij} - \tilde{\mathbf{Y}}_{\bullet\bullet})}{\sqrt{\sum_{i=1}^n \sum_{j=1}^{m_i} (\tilde{V}_{ij} - \tilde{\mathbf{V}}_{\bullet\bullet})^2 \sum_{i=1}^n \sum_{j=1}^{m_i} (\tilde{Y}_{ij} - \tilde{\mathbf{Y}}_{\bullet\bullet})^2}}.$$

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

$$\bar{z}_1 = \frac{1}{2} \ln\left(\frac{1 + \bar{r}_1}{1 - \bar{r}_1}\right). \quad (9)$$

To test Hypothesis (b) based on a single imputed sample $\tilde{\mathbf{U}}$, set, for each $i \in \{1, \dots, n\}$, $\tilde{R}_i = \psi_{\hat{\theta}_2}^{-1}(\tilde{V}_{i1}) + \dots + \psi_{\hat{\theta}_2}^{-1}(\tilde{V}_{im_i})$, $\tilde{V}_i = \bar{F}_{\psi_{\hat{\theta}_2}, m_i}(\tilde{R}_i)$, $\tilde{Q}_i = \psi_{\hat{\theta}_2}^{-1}(\tilde{V}_{i1})/\tilde{R}_i$ and $\tilde{Y}_i = (\tilde{Q}_i + 1)^{1-m_i}$. Let also $\tilde{\mathbf{V}}_{\bullet} = \sum_{i=1}^n \tilde{V}_i/n$ and similarly $\tilde{\mathbf{Y}}_{\bullet} = \sum_{i=1}^n \tilde{Y}_i/n$. The sample version of Spear-

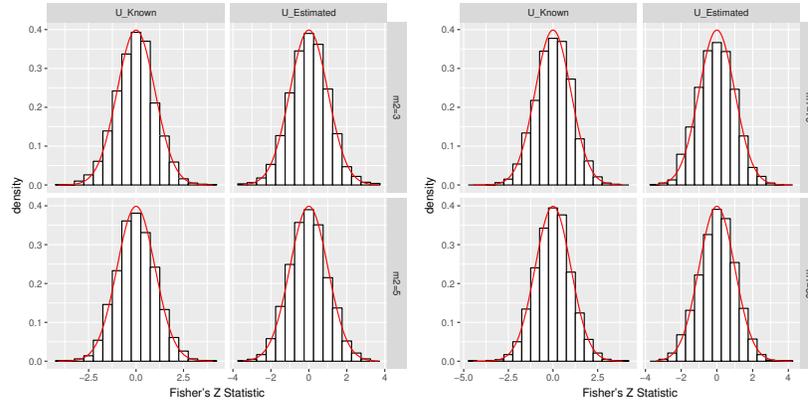


FIGURE 1: Histograms of 2000 values of $\sqrt{n} \bar{z}_1$ (left panel) and $\sqrt{n} \bar{z}_2$ (right panel) when $n = 200$. For each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_{ij}\}$, $m_i = m_1$ and $m_{ij} = m_2$; in the left and right panel these values are set to be $m_1 = 10$, $m_2 \in \{3, 5\}$, and $m_1 \in \{10, 30\}$, $m_2 = 5$, respectively. The generators are Clayton with $\theta_1 = 2$, $\theta_2 = 0.86$, no observations are censored, and the marginal model is as in Section 6.1. In each panel, the left and right columns correspond to the case when the margins are known and estimated.

man's ρ and the resulting Fisher test statistic are then given by

$$\bar{r}_2 = \frac{\sum_{i=1}^n (\tilde{V}_i - \tilde{V}_\bullet)(\tilde{Y}_i - \tilde{Y}_\bullet)}{\sqrt{\sum_{i=1}^n (\tilde{V}_i - \tilde{V}_\bullet)^2 \sum_{i=1}^n (\tilde{Y}_i - \tilde{Y}_\bullet)^2}}, \quad \bar{z}_2 = \frac{1}{2} \ln\left(\frac{1 + \bar{r}_2}{1 - \bar{r}_2}\right). \quad (10)$$

If no censoring is present, $\sqrt{n} \bar{z}_1$ and $\sqrt{n} \bar{z}_2$ are approximately standard Normal under Hypotheses (a) and (b). This is illustrated in Figure 1, which shows histograms of $N = 2000$ values of $\sqrt{n} \bar{z}_1$ and $\sqrt{n} \bar{z}_2$ when $n = 200$, and for each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, $m_i \equiv m_1 \in \{10, 30\}$ and $m_{ij} \equiv m_2 \in \{3, 5\}$. 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 τ , respectively. The marginal model is as in Section 6.1 and the marginal parameters are estimated as in Spiekerman and 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 in Equation (8), the variability introduced through the imputation procedure can be reduced by taking averages over K imputation samples. For each $k \in \{1, \dots, K\}$, let $\bar{z}_1^{(k)}$ and $\bar{z}_2^{(k)}$ denote the statistics in Equations (9) and (10) corresponding to the k th imputed sample, respectively. Let $\bar{z}_{1,K} = \sum_{k=1}^K \bar{z}_1^{(k)} / K$ and $\bar{z}_{2,K} = \sum_{k=1}^K \bar{z}_2^{(k)} / K$. Following Rubin (1987), the variance $D_{1,K}$ of $\bar{z}_{1,K}$ is

$$D_{1,K} = \frac{1}{K} \sum_{k=1}^K \text{var}(\bar{z}_1^{(k)}) + \frac{K+1}{K} B_{1,K}, \quad B_{1,K} = \frac{1}{K-1} \sum_{k=1}^K (\bar{z}_1^{(k)} - \bar{z}_{1,K})^2,$$

while the variance of $\bar{z}_{2,K}$ equals $D_{2,K} = (1/n) + B_{2,K}(K+1)/K$ with $B_{2,K} = \{1/(K-1)\} \sum_{k=1}^K (z_2^{(k)} - \bar{z}_{2,K})^2$. Hypotheses (a) and (b) are then tested using the statistics $T_{1,K} = \bar{z}_{1,K} / \sqrt{\hat{D}_{1,K}}$ and $T_{2,K} = \bar{z}_{2,K} / \sqrt{D_{2,K}}$, respectively. Here, $\hat{D}_{1,K}$ is obtained by estimating $\text{var}(\bar{z}_1^{(k)})$ by the bootstrap procedure of Monaco et al. (2005) in which clusters are sampled with replacement. If Hypotheses (a) and (b) hold and $K \geq 2$, the work of Rubin (1987) implies that

the distribution of $T_{1,K}$ and $T_{2,K}$ is approximately Student t with degrees of freedom ν_1 and ν_2 , respectively, where for $\iota \in \{1, 2\}$, $\nu_\iota = \{[K(K-1)D_{\iota,K}]/\{(K+1)B_{\iota,K}\}]^2$.

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 chronic granulomatous disease (CGD).

6.1. Simulation study of the proposed estimators

Throughout, the parameters of the hierarchical Kendall copula model are fixed as follows. For each $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ and $\ell \in \{1, \dots, m_{ij}\}$, $p = 2$ and $Z_{ij\ell 1}$ is uniform on $(0, 1)$ and independent of $Z_{ij\ell 2}$, which is Bernoulli(0.5). Given $\mathbf{Z}_{ij\ell}$, $T_{ij\ell}$ follows the Cox proportional hazards model with $\beta = (0.5, 1)$ and the Weibull baseline hazard function given, for all $t > 0$, by $\lambda(t) = (at^{a-1})/(b^a)$ with $a = 0.91$ and $b = 52.57$. The latter are the parameter estimates obtained by fitting the Cox proportional hazards model with Weibull baseline hazard function to the CGD data. Finally, the generators are Clayton with $\theta_1 = 2$ and $\theta_2 = 0.857$, corresponding to Kendall's τ of 0.5 and 0.3, respectively.

To generate the vector \mathbf{T} of survival times, first draw the vector \mathbf{U} using Algorithm 3.1 and Algorithm 3.7 of Brechmann (2014). Applying the marginal inverse survival functions then yields $T_{ij\ell} = \{[-\ln(U_{ij\ell})]e^{-\mathbf{Z}_{ij\ell}\beta^\top} b^a\}^{1/a}$. The lifetimes are then censored by independent variables $T_{ij\ell}^c$, uniform on $(0, c)$, where c is chosen to reach the targeted censoring rate 10%, 30%, or 70%.

The censoring rate, the number of upper-level clusters n , and the cluster sizes typically influence the performance of the estimators $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$, $\hat{\theta}_1$, $\hat{\theta}_{2,K}$ and $\hat{\theta}_{2,K}^c$, where the last two are obtained by maximizing Equations (6) and (7), respectively. To investigate this in detail, we consider the following three scenarios:

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

From preliminary simulation results not reported here, we choose $K = 10$, $K = 30$, and $K = 100$ when the censoring rate is 10%, 30%, and 70%, respectively. Throughout, each result is based on 500 simulation runs.

Table 1 shows an excerpt of the results under Scenarios A and B; the complete tables for all scenarios are provided in Tables 3–7 in Section 4 of the Online Supplement. For each estimator, we report 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 of its large-sample 95% confidence interval. Also shown are the average censoring rates δ_{ij}^R of the cluster representatives \hat{R}_{ij} . These are indeed much higher than the censoring rates of the univariate lifetimes, and this becomes worse as the size of the lower-level clusters increases. Overall, the performance of all estimators is reasonable, although the association parameter estimators seem to be systematically negatively biased. It is also reassuring that SE and ESE are close, even for $n = 50$.

As expected, all estimators improve as n increases. The coverage rate is surprisingly good even if $n = 50$, except when the censoring rate is high and $\hat{\theta}_{2,K}$ is used. As the censoring rate increases, so does the standard error of all estimators. The bias of $\hat{\theta}_{2,K}$ and $\hat{\theta}_{2,K}^c$ also increases

TABLE 1: Finite-sample performance of the association parameter estimates under Scenario B.

(n, m_1, m_2)	Est.	Cens. rate	Av. cens. rate of \hat{R}	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%
		70%	96%	-0.0154	0.2579	0.2784	0.2788	94.2%
	$\bar{\theta}_{2,K}$	10%	38%	-0.0694	0.1136	0.1035	0.1246	93.9%
		30%	67%	-0.1246	0.1338	0.1257	0.1770	88.8%
		70%	96%	-0.1674	0.2356	0.2277	0.2826	81.6%
	$\bar{\theta}_{2,K}^c$	10%	38%	-0.0452	0.1636	0.1597	0.1660	93.4%
		30%	67%	-0.0857	0.1732	0.1716	0.1918	93.6%
		70%	96%	-0.1312	0.2864	0.2781	0.3075	85.2%
(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%
		70%	92%	0.0253	0.2354	0.2691	0.2702	94.8%
	$\bar{\theta}_{2,K}$	10%	26%	-0.0657	0.0915	0.0908	0.1121	90.0%
		30%	55%	-0.1133	0.1096	0.1087	0.1570	87.0%
		70%	92%	-0.1571	0.1698	0.1684	0.2303	78.6%
	$\bar{\theta}_{2,K}^c$	10%	26%	-0.0573	0.1141	0.1172	0.1305	92.0%
		30%	55%	-0.0882	0.1227	0.1215	0.1501	89.0%
		70%	92%	-0.1155	0.1938	0.1922	0.2242	84.5%
(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%
		70%	96%	0.0011	0.1961	0.2084	0.2084	93.0%
	$\bar{\theta}_{2,K}$	10%	36%	-0.0614	0.0983	0.0921	0.1107	94.8%
		30%	67%	-0.1066	0.1156	0.1036	0.1486	91.0%
		70%	96%	-0.1411	0.1671	0.1653	0.2173	83.4%
	$\bar{\theta}_{2,K}^c$	10%	36%	-0.0383	0.1269	0.1145	0.1207	94.0%
		30%	67%	-0.0738	0.1407	0.1337	0.1527	93.2%
		70%	96%	-0.1031	0.2031	0.1987	0.2239	88.0%
(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%
		70%	96%	0.0175	0.1785	0.1811	0.1819	93.6%
	$\bar{\theta}_{2,K}$	10%	36%	-0.0701	0.0711	0.0717	0.1003	88.0%
		30%	67%	-0.1369	0.0747	0.0738	0.1555	83.0%
		70%	96%	-0.1701	0.1338	0.1341	0.2166	76.0%
	$\bar{\theta}_{2,K}^c$	10%	36%	-0.0384	0.0839	0.0833	0.0917	94.2%
		30%	67%	-0.0863	0.0876	0.0866	0.1223	93.6%
		70%	96%	-0.1084	0.1457	0.1441	0.1803	93.6%

with the censoring rate. However, the impact of the latter on the bias of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\theta}_1$ is less clear and it may even seem that the bias decreases when the censoring rate is 70%. Although counter-intuitive, the same phenomenon appears in the simulation experiments of Andersen (2005), Shih and Lu (2007), and Prenen et al. (2017). Further, the estimators of β_1 and β_2 perform better than those of θ_1 and θ_2 . Moreover, the estimator of θ_1 performs better than the estimators of θ_2 in terms of bias but not necessarily in terms of the standard error; the RMSE is actually of a similar magnitude.

Scenario B explores the effect of the cluster size when $n = 100$. When the size of the lower-level clusters increases from $m_2 = 3$ to $m_2 = 5$, more information becomes available. This reduces the bias and especially the standard error of the marginal parameter estimates. Also reduced are the bias and standard error of $\hat{\theta}_1$. Because the number of clusters at the upper level is unchanged, the bias and standard error of the estimates of θ_2 remain similar. However, the coverage rate is clearly improved. Similarly, when the size of the upper-level clusters increases from $m_1 = 10$ to $m_1 = 30$, the performance of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\theta}_1$ improves. The RMSE of $\bar{\theta}_{2,K}$ remains about the same while the RMSE of $\bar{\theta}_{2,K}^c$ improves and becomes smaller than the RMSE $\bar{\theta}_{2,K}$ when $m_1 = 30$. This means that when the upper-level clusters are large, the composite likelihood leads to an estimator with a comparable, if not slightly better performance.

When cluster sizes vary under Scenario C, the behaviour of all estimators is similar to the case (100, 10, 5) except for slightly larger standard errors.

6.2. Simulation study of the proposed goodness-of-fit procedures

To examine the goodness-of-fit procedures proposed in Section 5, we consider the same marginal model as in Section 6.1. The hypothesized generators are taken to be Clayton parametrized by Kendall's τ , viz. $\tau_1 = \theta_1/(\theta_1 + 2)$ and $\tau_2 = \theta_2/(\theta_2 + 2)$, respectively. Further, $m_i \equiv m_1$ and $m_{ij} \equiv m_2$ for each $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$. All tests are carried out at the 5% level and the results are based on 500 simulation runs and 500 bootstrap replicates for the computation of $\hat{D}_{1,K}$. The test based on $T_{2,K}$ uses the estimator $\bar{\theta}_{2,K}^c$.

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 investigated in Table 3 in the Online Supplement for various censoring rates, degrees of association, numbers of upper-level clusters, and cluster sizes. The number of imputations is taken to be $K = 0$ if no observation is censored, and $K = 30$ if the censoring rate is 30%. Even though the margins are estimated, the test holds its level fairly well overall. To investigate the power, we take the generators of the lower and upper level copulas to both be either in the Gumbel or the Frank family parametrized by Kendall's τ . The top left panel of Figure 2 shows the power of the tests based on $T_{1,K}$ and $T_{2,K}$ for various values of n , as a function of τ_1 , when $\tau_2 = 0.3$. As expected, power increases with n for each value of τ_1 . The power also increases with τ_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 further see from these plots that the power depends on the alternative, and that it is negatively affected by the censoring rate, although this phenomenon is less pronounced for the test based on $T_{2,K}$. Also apparent is the higher power of the test based on $T_{1,K}$, which is not surprising in view of the fact that this test is based on a higher number of observations. The bottom panel of Figure 2 shows the effect of the cluster sizes m_1 and m_2 on the power of the test based on $T_{1,K}$; as expected, the power increases with the cluster size.

6.3. Data Analysis

In this section, we apply the hierarchical Kendall copula model with Archimedean clusters to the data from the study of chronic granulomatous disease (CGD). The dataset contains 128 patients from $n = 13$ hospitals. The number m_i of patients from a given hospital ranges from 4 to 26

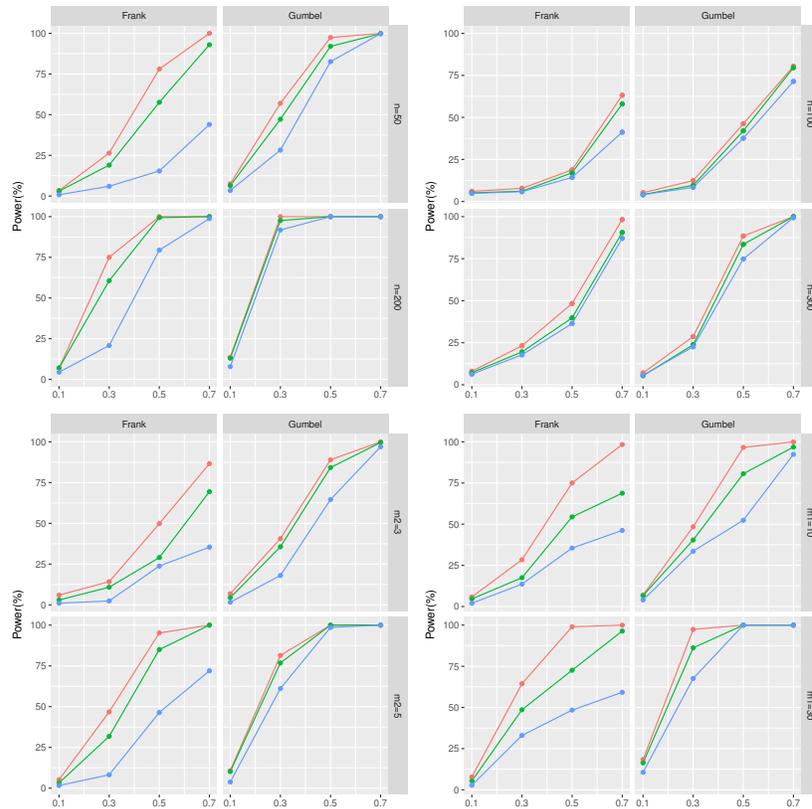


FIGURE 2: Power of the tests based on $T_{1,K}$ (top left, bottom left, and bottom right) and $T_{2,K}$ (top right) for the Frank and Gumbel alternatives as a function of τ_1 when $\tau_2 = 0.3$. In the top panels, $(n, m_1, m_2) = (n, 20, 5)$, while in the bottom left and right panels (n, m_1, m_2) equals $(200, 10, m_2)$ and $(100, m_1, 5)$, respectively. The colours represent the censoring rates 0% (red), 10% (green) and 30% (blue).

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 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 inter- and intra-cluster association, as detailed in Section 4 of the Online Supplement. The results reported in Table 8 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 coverage rate can be attributed to the fact that the association at both the lower and upper level is rather mild. Note also that while the entire data set is exploited to estimate the marginal parameters, only patients who experienced at least one infection and one censored time (last observation) contribute to (5).

We first evaluate the effect of gamma interferon on the disease. From the top part of Table 2, we see that the treatment effect of γ -IFN is significant, as also observed by Ha and Lee (2005) and Rondeau et al. (2006).

The results of the association analysis are summarized in the bottom panel of Table 2. The numbers of imputations and bootstrap replicates for the calculation of $\hat{D}_{1,K}$ are $K = 1000$ and $B = 500$, respectively. For both the lower and upper level, we chose the generators to be Clayton; from the p -values of the tests from Section 5 displayed in the last column, this assumption seems reasonable. The penultimate column shows the p -values of the test of the hypothesis that the

TABLE 2: Analysis of the Chronic Granulomatous Disease (CGD) data.

Covariates	Marginal parameters				
	$\hat{\beta}$	SE	95% CI	p -value ($\beta = 0$)	
Treatment	-1.0860	0.2680	(-1.6112, -0.5607)	< 0.0001	
Estimators	Association parameters				
	Est.	SE	95% CI	p -value (Ind.)	p -value (GOF)
$\hat{\theta}_1$	0.4447	0.1916	(0.0692, 0.8203)	0.0202	0.1228
$\bar{\theta}_{2,K}$	0.1570	0.1510	(-0.1389, 0.4529)	0.2984	0.8909
$\bar{\theta}_{2,K}^c$	0.0780	0.0719	(-0.0629, 0.2189)	0.2779	0.8785

Est.: Estimate; SE: estimated standard error; CI: confidence interval; GOF: Goodness-of-fit test

association is not significant, i.e., that $\theta_1 = \theta_2 = 0$. We can see that the within-patient association is significant albeit mild, with $\hat{\theta}_1 = 0.4447$ corresponding to Kendall's $\tau_1 = 0.18$. The same conclusion was obtained using the random-effect approach by Ha and Lee (2005). 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 $\bar{\theta}_{2,K} = 0.157$ and $\bar{\theta}_{2,K}^c = 0.078$, corresponding to Kendall's $\tau_2 = 0.07$ and 0.04, respectively. None of these values are significantly different from 0, however. This is in agreement with Ha and Lee (2005) and Rondeau et al. (2006), who conclude from their random effects model that there is no hospital effect.

7. CONCLUDING REMARKS

In this paper, we adapted the hierarchical Kendall copula model with Archimedean clusters of Brechmann (2014) to clustered data with multiple levels of clustering. This model aims to reduce dimensionality through clustering. The way it is built reflects the cluster structure, which makes it natural and easy to interpret. Clusters can have unequal size and the model is easily expanded to multiple levels of hierarchy without imposing restrictions on the parameter space. As we discussed, these are clear advantages over the previously proposed hierarchical Archimedean model. Because of the way the model is constructed, association parameters can be estimated sequentially level by level. We extended the work of Brechmann (2014) in that we developed an estimation procedure that can be used for right-censored observations. Because the censoring rate of the cluster representatives increases with the level of the hierarchy, we relied on imputation. The proposed estimators are shown to be asymptotically Gaussian and well-behaved in finite samples. Furthermore, we developed goodness-of-fit tests which can be used to check whether at a given level of the hierarchy, the Archimedean generator indeed belongs to a specific parametric class. As we demonstrated through simulations, the tests hold their level well and have reasonable power in finite samples. We used the proposed model to analyze two-level clustered data and reached similar conclusions as previous analyses based on random effects models.

Acknowledgments

The authors thank the Editor, the Associate Editor and an anonymous referee whose comments led to a substantial improvement of the paper. Chien-Lin Su and Weijing Wang gratefully ac-

knowledge the financial support by the Taiwanese Ministry of Science and Technology, ROC (Su: 104-2811-M-009-067; Wang: 104-2118-M-009-003-MY2). Partial funding in support of Johanna G. Nešlehová's research was provided by the Natural Sciences and Engineering Research Council of Canada (RGPIN/06801–2015), the Canadian Statistical Sciences Institute, and the Fonds de recherche du Québec – Nature et technologies (2015–PR–183236).

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 Anal.*, 11:333–350.
- Barbe, P., Genest, C., Ghoudi, K., and Rémillard, B. (1996). On Kendall's process. *J. Multivariate Anal.*, 58:197–229.
- Brechmann, E. C. (2014). Hierarchical Kendall copulas: properties and inference. *Canad. J. Statist.*, 42:78–108.
- Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.
- Durante, F. and Sempì, 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. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.
- Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bull.*, 37:475–515.
- Genest, C., Nešlehová, J., and Ziegel, J. (2011). Inference in multivariate Archimedean copula models (with discussion). *TEST*, 20:223–256.
- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.*, 88:1034–1043.
- Ha, I. D. and Lee, Y. (2005). Multilevel mixed linear models for survival data. *Lifetime Data Anal.*, 11(1):131–142.
- Hofert, M., Mächler, M., and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *J. Multivariate Anal.*, 110:133–150.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. *J. Multivariate Anal.*, 46:262–282.
- Joe, H. (2015). *Dependence Modeling With Copulas*. CRC Press, Boca Raton, FL.
- Joe, H. and Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *J. Multivariate Anal.*, 100(4):670–685.
- Mai, J.-F. and Scherer, M. (2017). *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*, volume 6 of *Series in Quantitative Finance*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2nd edition. With contributions by Claudia Czado, Elke Korn, Ralf Korn and Jacob Stöber.
- Malov, S. (2001). On finite-dimensional Archimedean copulas. In Balakrishnan, N., Ibragimov, I., and Nevzorov, V., editors, *Asymptotic Methods in Probability and Statistics With Applications*, pages 19–35, Basel. Birkhäuser.
- Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *J. Amer. Statist. Assoc.*, 83:834–841.
- McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d -monotone functions and ℓ_1 -norm symmetric distributions. *Ann. Statist.*, 37:3059–3097.
- Monaco, J., Cai, J., and Grizzle, J. (2005). Bootstrap analysis of multivariate failure time data. *Stat. Med.*, 24:3387–3400.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, 2nd edition.

- Othus, M. and Li, Y. (2010). A Gaussian copula model for multivariate survival data. *Statistics in Biosciences*, 2:154–179.
- Preneen, L., Braekers, R., and Duchateau, L. (2017). Extending the Archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79:483–505.
- Romdhani, H., Lakhel-Chaieb, L., and Rivest, L.-P. (2014). An exchangeable Kendall's tau for clustered data. *Canad. J. Statist.*, 42:384–403.
- Rondeau, V., Filleul, L., and Joly, P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Stat. Med.*, 25:4036–4052.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.
- Sastry, N. (1997). A nested frailty model for survival data, with an application to the study of child survival in Northeast Brazil. *J. Amer. Statist. Assoc.*, 92:426–435.
- Shih, J. H. and 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. and Lu, S.-E. (2009). Semiparametric estimation of a nested random effects model for the analysis of multi-level clustered failure time data. *Comput. Statist. Data Anal.*, 53:3864–3871.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.
- Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.*, 93:1164–1175.
- Yau, K. K. W. (2001). Multilevel models for survival analysis with random effects. *Biometrics*, 57:96–102.
- Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canad. J. Statist.*, 33:335–356.

APPENDIX

Regularity conditions. Let $T > 0$ be the end of study time, and $(\beta_0, \Lambda_0, \theta_{10}, \theta_{20})$ refer to the true values of the parameters $(\beta, \Lambda, \theta_1, \theta_2)$. Also, for each $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$, $\ell \in \{1, \dots, m_{ij}\}$ and $t \geq 0$, $Y_{ij\ell} = \mathbf{1}(X_{ij\ell} \geq t)$. Furthermore, following Spiekerman and Lin (1998), introduce the following notation. For an arbitrary vector \mathbf{a} , let $\mathbf{a}^{\otimes 0} = \mathbf{1}$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}^\top \mathbf{a}$. For $r \in \{0, 1, 2\}$, write

$$S^{(r)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} Y_{ij\ell}(t) \exp\{\mathbf{Z}_{ij\ell}(u) \beta^\top\} \mathbf{Z}_{ij\ell}^{\otimes r}(t), \quad s^{(r)}(\beta, t) = \mathbb{E}\{S^{(r)}(\beta, t)\}.$$

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) $\beta \in \mathcal{B}$, where \mathcal{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 Θ_k .
- (C2) For all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$, $\ell \in \{1, \dots, m_{ij}\}$, $\Pr(\forall t \in [0, T] T_{ij\ell}^c \geq t) > \delta_c > 0$.
- (C3) Writing $\mathbf{Z}_{ij\ell}(t) = \{Z_{ij\ell 1}(t), \dots, Z_{ij\ell p}(t)\}$ one has, for all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$, $\ell \in \{1, \dots, m_{ij}\}$ and $k \in \{1, \dots, p\}$, and some constant B_Z that $|Z_{ij\ell k}(0)| + \int_0^T |dZ_{ij\ell k}(t)| \leq B_Z < \infty$ almost surely.
- (C4) $\mathbf{A} = \int_0^T v(\beta_0, u) s^{(0)}(\beta_0, u) d\Lambda_0(u)$ is positive definite.

- (C5) For all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, $E\{\ln L_{ij}(\theta_1; \beta_0, \Lambda_0)/L_{ij}(\theta_1^*; \beta_0, \Lambda_0)\}$ exists for all $\theta_1, \theta_1^* \in \Theta_1$.
- (C6) For all $i \in \{1, \dots, n\}$, $E\{\ln L_{2i}(\theta_2; \theta_{10}, \beta_0, \Lambda_0)/L_{2i}(\theta_2^*; \theta_{10}, \beta_0, \Lambda_0)\}$ exists for all $\theta_2, \theta_2^* \in \Theta_2$, where L_{2i} denotes the i th summand of the joint likelihood (6) or the composite likelihood (7), as the case may be.

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

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

with

$$Q_{ij\ell} = \exp \left[- \int_0^\tau Y_{ij\ell}(u) \exp \{ \mathbf{Z}_{ij\ell}(u) \beta^\top \} d\Lambda(u) \right]. \quad (\text{A.1})$$

Similarly to Prenen et al. (2017), the Hadamard derivative of L_1 with respect to Λ at $(\Gamma - \Lambda)$ of bounded variation in $[0, T]$ is $\int_0^T \zeta_n(\theta_1; \Lambda)(u) d(\Gamma - \Lambda)(u)$, where for any $u \in [0, T]$,

$$\zeta_n(\theta_1; \Lambda)(u) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} D_{ij\ell} Y_{ij\ell}(u) \exp \{ \mathbf{Z}_{ij\ell}(u) \beta^\top \}.$$

Furthermore, the derivative of $L_1(\theta_1)$ with respect to β is given by

$$\zeta_n(\theta_1; \beta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} D_{ij\ell} \int_0^\tau Y_{ij\ell}(u) \mathbf{Z}_{ij\ell}(u) \exp \{ \mathbf{Z}_{ij\ell}(u) \beta^\top \} d\Lambda(u).$$

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

$$\|D_{ij\ell}\|_\infty < D < \infty, \quad (\text{A.2})$$

which holds, e.g., for the Clayton generator in view of the regularity condition (C2) (Prenen et al., 2017). Consequently, $\|\zeta_n(\theta_1; \Lambda)\|_\infty$ and $\|\zeta_n(\theta_1; \beta)\|$ are bounded. An expansion of $L_1(\theta_1)$ around β_0 and Λ_0 is given by

$$L_1(\theta_1) = L_1^0(\theta_1) + \zeta_n(\theta_1; \beta_0)(\hat{\beta} - \beta_0) + \int_0^T \zeta_n(\theta_1; \Lambda)(t) d(\hat{\Lambda} - \Lambda_0)(t) + R,$$

where L_1^0 denotes the log-likelihood L_1 with β_0 and Λ_0 instead of $\hat{\beta}$ and $\hat{\Lambda}$. As argued in Prenen et al. (2017), $R = o_p(1)$ as $\|\hat{\beta} - \beta_0\|$ and $\|\hat{\Lambda} - \Lambda_0\|_\infty$ are $o_p(1)$ (Spiekerman and Lin, 1998).

Using the same arguments as in the proof of Theorem 2 in Prenen et al. (2017), one has, for all $\theta_1 \in \Theta_1$,

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

Because $\hat{\theta}_1$ maximizes $L_1(\theta_1)$, we have $0 \leq L_1(\hat{\theta}_1) - L_1(\theta_{10}) = L_1(\hat{\theta}_1) - E\{L_1^0(\theta_{10})\} + o_p(1)$ so that, in view of Equation (A.3), $E\{L_1^0(\theta_{10})\} - E\{L_1^0(\hat{\theta}_1)\} \leq L_1(\hat{\theta}_1) - E\{L_1^0(\hat{\theta}_1)\} + o_p(1) = o_p(1)$. Now take θ_1 with $|\theta_1 - \theta_{10}| \geq \varepsilon$ for some fixed $\varepsilon > 0$. By Equation (A.3), there must exist some γ_ε such that $E\{L_1^0(\hat{\theta}_1)\} + \gamma_\varepsilon < E\{L_1^0(\theta_{10})\}$. It follows that $\lim_{n \rightarrow \infty} \Pr(|\hat{\theta}_1 - \theta_{10}| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \Pr[E\{L_1^0(\hat{\theta}_1)\} + \gamma_\varepsilon < E\{L_1^0(\theta_{10})\}] = 0$.

Normality. Let $S_{1n}(\theta_1; \hat{\beta}, \hat{\Lambda})$ be the score function, i.e., the derivative of $L_1(\theta_1)$ with respect to θ_1 . Its Taylor series expansion around θ_{10} is

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

where θ_1^* is between $\hat{\theta}_1$ and θ_{10} . Since $\hat{\theta}_1$ is a root of $S_{1n}(\theta_1; \hat{\beta}, \hat{\Lambda})$, $\sqrt{n}(\hat{\theta}_1 - \theta_{10}) = \{-\partial S_{1n}(\theta_1^*; \hat{\beta}, \hat{\Lambda})/\partial \theta_1\}^{-1} \sqrt{n} S_{1n}(\theta_{10}; \hat{\beta}, \hat{\Lambda})$. The Law of Large Numbers and the consistency of $(\hat{\beta}, \hat{\Lambda}, \hat{\theta}_1)$ implies that

$$-\frac{\partial S_{1n}}{\partial \theta_1}(\theta_1^*, \hat{\beta}, \hat{\Lambda}) \xrightarrow{\text{Pr}} I_1(\theta_{10}) = -\lim_{n \rightarrow \infty} \frac{\partial S_{1n}}{\partial \theta_1}(\theta_{10}, \beta_0, \Lambda_0).$$

Thus it only remains to show the asymptotic normality of $\sqrt{n} S_{1n}(\theta_{10}; \hat{\beta}, \hat{\Lambda})$, which can be derived as in the proof of Theorem 3 in Prenen et al. (2017).

Recall the definition of $Q_{ij\ell}$ in Equation (A.1) and introduce $D_{ij\ell}^S = (d/d\theta)D_{ij\ell}$, i.e.,

$$\begin{aligned} D_{ij\ell}^S &= (-Q_{ij\ell}) \left[\delta_{ij\ell} \frac{\psi_{\theta_1}^{(2)}\{\psi_{\theta_1}^{-1}(Q_{ij\ell})\}}{\psi_{\theta_1}^{(1)}\{\psi_{\theta_1}^{-1}(Q_{ij\ell})\}} \frac{\partial}{\partial \theta_1} [\psi_{\theta_1}^{(1)}\{\psi_{\theta_1}^{-1}(Q_{ij\ell})\}]^{-1} \right. \\ &\quad \left. + \delta_{ij\ell} \psi_{\theta_1}^{(1)}\{\psi_{\theta_1}^{-1}(Q_{ij\ell})\} \frac{\partial}{\partial \theta_1} \left[-\frac{\psi_{\theta_1}^{(2)}\{\psi_{\theta_1}^{-1}(Q_{ij\ell})\}}{[\psi_{\theta_1}^{(1)}\{\psi_{\theta_1}^{-1}(Q_{ij\ell})\}]^3} \right] \right. \\ &\quad \left. + \frac{1}{\psi_{\theta_1}^{(d_{ij})} \left\{ \sum_{\ell=1}^{m_{ij}} \psi_{\theta_1}^{-1}(Q_{ij\ell}) \right\}} \frac{\partial}{\partial \theta_1} \left[\frac{\psi_{\theta_1}^{(d_{ij}+1)} \left\{ \sum_{\ell=1}^{m_{ij}} \psi_{\theta_1}^{-1}(Q_{ij\ell}) \right\}}{\psi_{\theta_1}^{(1)} \left\{ \psi_{\theta_1}^{-1}(Q_{ij\ell}) \right\}} \right] \right. \\ &\quad \left. - \frac{\psi_{\theta_1}^{(d_{ij}+1)} \left\{ \sum_{\ell=1}^{m_{ij}} \psi_{\theta_1}^{-1}(Q_{ij\ell}) \right\}}{[\psi_{\theta_1}^{(d_{ij})} \left\{ \sum_{\ell=1}^{m_{ij}} \psi_{\theta_1}^{-1}(Q_{ij\ell}) \right\}]^2} \frac{\partial}{\partial \theta_1} \left[\psi_{\theta_1}^{(d_{ij})} \left\{ \sum_{\ell=1}^{m_{ij}} \psi_{\theta_1}^{-1}(Q_{ij\ell}) \right\} \right] \right]. \end{aligned}$$

The Hadamard derivative of $S_{1n}(\theta_1; \beta, \Lambda)$ with respect to Λ 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

$$\xi_n(\theta_1; \Lambda)(u) = \frac{1}{n} \sum_{i=1}^n \sum_j^{m_i} \sum_{\ell=1}^{m_{ij}} D_{ij\ell}^S Y_{ij\ell}(u) \exp\{\mathbf{Z}_{ij\ell}(u) \beta^\top\},$$

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

$$\xi_n(\theta_1; \beta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} D_{ij\ell}^S \int_0^T Y_{ij\ell}(u) \mathbf{Z}_{ij\ell}(u) \exp\{\mathbf{Z}_{ij\ell}(u) \beta^\top\} d\Lambda(u).$$

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

$$\|D_{ij\ell}^S\|_\infty < D^S < \infty, \quad (\text{A.4})$$

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

$$\begin{aligned} \sqrt{n} S_{1n}(\theta_{10}; \hat{\boldsymbol{\beta}}, \hat{\Lambda}) &= \sqrt{n} \left[S_{1n}(\theta_{10}; \boldsymbol{\beta}_0, \Lambda_0) + \xi(\theta_{10}; \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \right. \\ &\quad \left. + \int_0^T \xi(\theta_{10}; \Lambda_0)(t) d\{\hat{\Lambda}(t) - \Lambda_0(t)\} \right] + o_p(1), \end{aligned} \quad (\text{A.5})$$

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

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{A}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n W_{i++} \right) + o_p(1), \quad (\text{A.6})$$

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

$$W_{i++} = \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} \int_0^T \{\mathbf{Z}_{ij\ell}(u) - e(\boldsymbol{\beta}_0, u)\} dM_{ij\ell}(u)$$

with $M_{ij\ell}(t) = \delta_{ij\ell} \mathbf{1}(X_{ij\ell} \leq t) - \int_0^t Y_{ij\ell}(u) e\{\mathbf{Z}_{ij\ell}(u)\boldsymbol{\beta}_0^\top\} d\Lambda_0(u)$. Now for each $t \in [0, T]$ and $i \in \{1, \dots, n\}$, let $M_{i++} = \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} M_{ij\ell}$ and

$$\mathbf{h}(t) = - \int_0^t e(\boldsymbol{\beta}_0, u) d\Lambda_0(u), \quad \Psi_i(t) = \int_0^t \frac{dM_{i++}(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} + \mathbf{h}^T(t) \mathbf{A}^{-1} W_{i++}.$$

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

$$\sqrt{n} \{\hat{\Lambda}(t, \hat{\boldsymbol{\beta}}) - \Lambda_0(t)\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_i(t) + o_p(t), \quad (\text{A.7})$$

where $o_p(t)$ is uniform in $t \leq T$. Substituting Equations (A.6) and (A.7) into the right-hand side of Equation (A.5) and denoting the i th summand of $S_{1n}(\theta_{10}; \boldsymbol{\beta}_0, \Lambda_0)$ as $\phi_i(\theta_{10}; \boldsymbol{\beta}_0, \Lambda_0)$, i.e. $S_{1n}(\theta_{10}; \boldsymbol{\beta}_0, \Lambda_0) = \sum_{i=1}^n \phi_i(\theta_{10}; \boldsymbol{\beta}_0, \Lambda_0)/n$, we find that $\sqrt{n} S_{1n}(\theta_{10}; \hat{\boldsymbol{\beta}}, \hat{\Lambda}) = \sum_{i=1}^n \Phi_i / \sqrt{n} + o_p(1)$, where for each $i \in \{1, \dots, n\}$,

$$\Phi_i = \phi_i(\theta_{10}; \boldsymbol{\beta}_0, \Lambda_0) + \xi(\theta_{10}; \boldsymbol{\beta}_0) \mathbf{A}^{-1} W_{i++} + \int_0^T \xi(\theta_{10}; \Lambda_0)(t) d\Psi_i(t).$$

By the Central Limit Theorem, $\sqrt{n} S_{1n}(\theta_{10}; \hat{\boldsymbol{\beta}}, \hat{\Lambda}) \rightsquigarrow \mathcal{N}(0, \sigma_{\Phi}^2)$ as $n \rightarrow \infty$, where \rightsquigarrow denotes weak convergence and $\sigma_{\Phi}^2 = \text{var}(\Phi_1)$. By Slutsky's Lemma, $\sqrt{n}(\hat{\theta}_1 - \theta_{10}) \rightsquigarrow \mathcal{N}(0, \sigma_{\hat{\theta}_1}^2)$ as $n \rightarrow \infty$, where $\sigma_{\hat{\theta}_1}^2 = \sigma_{\Phi}^2 / \{I_1(\theta_{10})\}^2$, as claimed. The variance σ_{Φ}^2 can be estimated by $\sum_{i=1}^n \hat{\Phi}_i^2 / n$ where $\hat{\Phi}_i$ is obtained from Φ_i replacing all unknown parameters by their estimators; $I_1(\theta_{10})$ can be estimated by plugging in the estimator $\hat{\theta}_1$. ■

Online Supplement

Modelling hierarchical clustered censored data with the hierarchical Kendall copula

Chien-Lin Su^{1,2}, Johanna G. Nešlehová^{1*} and Weijing Wang²

¹*Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan (R.O.C.)*

Email: chien-lin.su@mail.mcgill.ca; wjwang@stat.nctu.edu.tw

²*Dept. of Mathematics and Statistics, McGill University, Montréal (QC) Canada*

Email: johanna.neslehova@mcgill.ca

1. MARGINAL PARAMETER ESTIMATION

The estimation of β and Λ is done using the step-wise approach of Spiekerman and Lin (1998). In this section, we give the details of the procedure.

Define the counting process $N_{ij\ell}(t) = I(X_{ij\ell} \leq t, \delta_{ij\ell} = 1)$ and $N_{\dots}(t) = \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} N_{ij\ell}(t)$. Let $Y_{ij\ell}(t) = I(X_{ij\ell} \geq t)$ be the at risk process. The maximum follow-up time is denoted by τ .

Under the working independence assumption, the estimate $\hat{\beta}$ of β is obtained by solving the score equation

$$U(\beta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} \int_0^{\tau} \left\{ \mathbf{Z}_{ij\ell}(u) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dN_{ij\ell}(u), \quad (1)$$

where for $r \in \{0, 1, 2\}$, $S^{(r)}(\beta, t)$ is as defined in the regularity conditions section of the Appendix.

Once the estimate $\hat{\beta}$ has been obtained, $\Lambda(t)$ is estimated by the Nelson-type estimator given, for any $t \in (0, \tau)$, by

$$\hat{\Lambda}(t) = \frac{1}{n} \int_0^t \frac{dN_{\dots}(u)}{S^{(0)}(\hat{\beta}, u)}.$$

2. PROOF OF THEOREM 2

In this section, we give the proof of Theorem 2 and the form of the asymptotic variance when the pair-wise pseudo-loglikelihood is used.

Proof of Theorem 2. The consistency and asymptotic normality of $\hat{\theta}_2$ follows similarly as in the proof of Theorem 1. We only sketch the proof of the latter when $L_2(\theta_2)$ is used; the proof for the composite likelihood is analogous.

* Author to whom correspondence may be addressed.

E-mail: Insert your email address here only after your paper has been accepted

Let $S_{2n}(\theta_2, \theta_1)$ be the derivative of $L_2(\theta_2)$. Using Taylor series expansion,

$$\sqrt{n}(\hat{\theta}_2 - \theta_{20}) = \left\{ -\frac{\partial S_{2n}}{\partial \theta_2}(\theta_2^*, \hat{\theta}_1) \right\}^{-1} \sqrt{n} S_{2n}(\theta_{20}, \hat{\theta}_1).$$

where θ_2^* is between $\hat{\theta}_2$ and θ_{20} . By the Law of Large Numbers and the consistency of $\hat{\theta}_1$ and $\hat{\theta}_2$,

$$-\frac{\partial S_{2n}}{\partial \theta_2}(\theta_2^*, \hat{\theta}_1) \xrightarrow{Pr} I_2(\theta_{20}) = -\lim_{n \rightarrow \infty} \frac{\partial S_{2n}}{\partial \theta_2}(\theta_{20}, \theta_{10}).$$

It thus remains to show that $\sqrt{n} S_{2n}(\theta_{20}, \hat{\theta}_1)$ converges to a zero-mean normal distribution. The derivative of $S_{2n}(\theta_2, \theta_1)$ with respect to θ_1 can be expressed as

$$\begin{aligned} \xi_n(\theta_2, \theta_1) = & \frac{1}{n} \sum_{i=1}^n \left[\frac{\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \psi_{\theta_2}^{(m_i)} \{ \sum_{j=1}^{m_i} \psi_{\theta_2}^{-1}(V_{ij}) \}}{\psi_{\theta_2}^{(m_i)} \{ \sum_{j=1}^{m_i} \psi_{\theta_2}^{-1}(V_{ij}) \}} \right. \\ & - \frac{\frac{\partial}{\partial \theta_1} \psi_{\theta_2}^{(m_i)} \{ \sum_{j=1}^{m_i} \psi_{\theta_2}^{-1}(V_{ij}) \} \frac{\partial}{\partial \theta_2} \psi_{\theta_2}^{(m_i)} \{ \sum_{j=1}^{m_i} \psi_{\theta_2}^{-1}(V_{ij}) \}}{[\psi_{\theta_2}^{(m_i)} \{ \sum_{j=1}^{m_i} \psi_{\theta_2}^{-1}(V_{ij}) \}]^2} \\ & \left. - \sum_{j=1}^{m_i} \frac{\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \psi_{\theta_2}^{(1)} \{ \psi_{\theta_2}^{-1}(V_{ij}) \}}{\psi_{\theta_2}^{(1)} \{ \psi_{\theta_2}^{-1}(V_{ij}) \}} + \sum_{j=1}^{m_i} \frac{\frac{\partial}{\partial \theta_1} \psi_{\theta_2}^{(1)} \{ \psi_{\theta_2}^{-1}(V_{ij}) \} \frac{\partial}{\partial \theta_2} \psi_{\theta_2}^{(1)} \{ \psi_{\theta_2}^{-1}(V_{ij}) \}}{[\psi_{\theta_2}^{(1)} \{ \psi_{\theta_2}^{-1}(V_{ij}) \}]^2} \right]; \end{aligned}$$

note that θ_1 appears in each of the V_{ij} 's. If the generators are such that

$$\|\xi_n(\theta_2, \theta_1)\| < \infty \tag{A.8}$$

then $\xi(\theta_2, \theta_1) = E\{\xi_n(\theta_2, \theta_1)\}$ is bounded. Thus from the proof of Theorem 1, $\sqrt{n} S_{2n}(\theta_{20}; \hat{\theta}_1) = \sum_{i=1}^n \Pi_i / \sqrt{n}$, where for each $i \in \{1, \dots, n\}$, $\Pi_i = \Xi_i + \xi(\theta_{20}, \theta_{10}) I_1^{-1}(\theta_{10}) \Phi_i$ and Ξ_i denotes the i th summand of $S_{2n}(\theta_{20}, \theta_{10})$, i.e., $S_{2n}(\theta_{20}, \theta_{10}) = \sum_{i=1}^n \Xi_i / n$. The Central Limit Theorem thus implies that as $n \rightarrow \infty$, $\sqrt{n} S_{2n}(\theta_{20}; \hat{\theta}_1) \rightsquigarrow \mathcal{N}[0, \sigma^2(\theta_{20}, \theta_{10})]$, where $\sigma^2(\theta_{20}, \theta_{10}) = \text{var}(\Pi_1)$. By Slutsky's Lemma, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\theta}_2 - \theta_{20}) \rightsquigarrow \mathcal{N}(0, \sigma_{\theta_{20}}^2)$, where $\sigma_{\theta_{20}}^2 = \sigma^2(\theta_{20}, \theta_{10}) / \{I_2(\theta_{20})\}^2$. Note that $\sigma^2(\theta_{20}, \theta_{10})$ can be estimated by $\sum_{i=1}^n \hat{\Pi}_i^2 / n$; $\hat{\Pi}_i$ is obtained from Π_i replacing all unknown parameters by their estimates. ■

Remark 2. When the pairwise likelihood $L_2^c(\theta_2)$ is used, the asymptotic variance $\sigma_{\theta_{20}, c}^2 = \sigma_c^2(\theta_{20}, \theta_{10}) / \{I_{2,c}(\theta_{20})\}^2$ of $\hat{\theta}_2$ can be derived as in the proof of Theorem 2 above. Let $S_{2n}^c(\theta_2, \theta_1)$ be the derivative of $L_2^c(\theta_2)$. Then

$$I_{2,c}(\theta_{20}) = \lim_{n \rightarrow \infty} \left. \frac{-dS_{2n}^c(\theta_2; \theta_{10})}{d\theta_2} \right|_{\theta_2 = \theta_{20}}$$

while $\sigma_c^2(\theta_{20}, \theta_{10})$ is the asymptotic variance of $\sqrt{n} S_{2n}^c(\theta_{20}; \hat{\theta}_1)$. Concretely, $\sigma_c^2(\theta_{20}, \theta_{10}) = \text{var}(\Pi_i^c)$ where for each $i \in \{1, \dots, n\}$, $\Pi_i^c = \Xi_i^c + \xi^c(\theta_{20}, \theta_{10}) I_1^{-1}(\theta_{10}) \Phi_i$, Ξ_i^c denotes the i th summand of $S_{2n}^c(\theta_{20}, \theta_{10}) = \sum_{i=1}^n \Xi_i^c / n$ and $\xi^c(\theta_2, \theta_1)$ is the expected value of the derivative of $S_{2n}^c(\theta_2, \theta_1)$ with respect to θ_1 .

3. EXPLICIT FORMULAS WHEN THE CLAYTON GENERATOR IS USED

3.1. Conditional distribution used in Algorithm 1

For $k \in \{1, \dots, h_{ij}\}$, the conditional distribution function

$$F_{ijk}(u; \theta_1) = \Pr(U_{ijk} \leq u | U_{ij1} = \tilde{u}_{ij1}, \dots, U_{ij(k-1)} = \tilde{u}_{ij(k-1)}, \\ U_{ijk} \leq \hat{u}_{ijk}, \dots, U_{ijh_{ij}} \leq \hat{u}_{ijh_{ij}}, U_{ij(h_{ij}+1)} = \hat{u}_{ij(h_{ij}+1)}, \dots, U_{ijm_{ij}} = \hat{u}_{ijm_{ij}})$$

can be computed to be

$$F_{ijk}(u; \theta_1) = \frac{(-1)^r \psi_{\theta_1}^{(r)} \left\{ \sum_{\ell=1}^{k-1} \psi_{\theta_1}^{-1}(\tilde{u}_{ij\ell}) + \psi_{\theta_1}^{-1}(u) + \sum_{\ell=k+1}^{m_{ij}} \psi_{\theta_1}^{-1}(\hat{u}_{ij\ell}) \right\}}{(-1)^r \psi_{\theta_1}^{(r)} \left\{ \sum_{\ell=1}^{k-1} \psi_{\theta_1}^{-1}(\tilde{u}_{ij\ell}) + \sum_{\ell=k}^{m_{ij}} \psi_{\theta_1}^{-1}(\hat{u}_{ij\ell}) \right\}}, \quad (*)$$

where $r = (k-1) + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}$.

When the generator is Clayton with $\theta > 0$, Hofert et al. (2012) computed that the k th derivative of ψ_{θ} is given, for all $k \geq 1$ and $t > 0$, by

$$(-1)^k \psi_{\theta}^{(k)}(t) = \frac{\Gamma(k+1/\theta)}{\Gamma(1/\theta)} \theta^k (1+\theta t)^{-(k+1/\theta)}.$$

The inverse of $(-1)^d \psi_{\theta}^{(d)}$, say f_d , thus equals

$$f_d(x) = \frac{1}{\theta} \left[\left\{ \frac{x\Gamma(1/\theta)}{\theta^d \Gamma(d+1/\theta)} \right\}^{-\theta/(d\theta+1)} - 1 \right].$$

An observation from \tilde{U}_{ijk} can then be generated by drawing a standard uniform variable V and setting

$$\tilde{U}_{ijk} = \psi_{\theta_1} \left\{ f_r(\kappa V) - \sum_{\ell=1}^{k-1} \psi_{\theta_1}^{-1}(\tilde{u}_{ij\ell}) - \sum_{\ell=k+1}^{m_{ij}} \psi_{\theta_1}^{-1}(\hat{u}_{ij\ell}) \right\},$$

where κ denotes the denominator on the right-hand side of (*).

3.2. Formulas for the asymptotic variance of $\hat{\theta}_1$ when $\theta_1 > 0$

For $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, define

$$L_{ij} = \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) - m_{ij} + 1, \quad A_{1ij} = \frac{\Gamma(1/\theta_1 + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell})}{\Gamma(1/\theta_1)}$$

and $A_{2ij} = (\partial A_{1ij}) / (\partial \theta_1)$. Then

$$I_1(\theta_1) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \left[-\frac{1}{\{\Gamma(1/\theta_1 + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell})\}^2} \left\{ \frac{\partial}{\partial \theta_1} \Gamma\left(\frac{1}{\theta_1} + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}\right) \right\}^2 \right. \\ \left. + \frac{1}{\{\Gamma(1/\theta_1)\}^2} \left\{ \frac{\partial}{\partial \theta_1} \Gamma\left(\frac{1}{\theta_1}\right) \right\}^2 + \frac{1}{\Gamma(1/\theta_1 + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell})} \frac{\partial^2}{\partial \theta_1^2} \Gamma\left(\frac{1}{\theta_1} + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}\right) \right]$$

$$\begin{aligned}
& -\frac{1}{\Gamma(1/\theta_1)} \frac{\partial^2}{\partial \theta_1^2} \Gamma\left(\frac{1}{\theta_1}\right) - \frac{2 \log L_{ij}}{\theta_1^3} + \frac{2}{\theta_1^2 L_{ij}} \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} \\
& + \frac{1/\theta_1 + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}}{L_{ij}^2} \left\{ \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} \right\}^2 \\
& - \frac{1/\theta_1 + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}}{L_{ij}} \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell}^2 - \frac{1}{\theta_1^2} \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell} \Big].
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\phi_i(\theta_1, \beta, \Lambda) &= \sum_{j=1}^{m_i} \left[\frac{A_{2ij}}{A_{1ij}} + \frac{\log L_{ij}}{\theta_1^2} - \frac{1/\theta_1 + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}}{L_{ij}} \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} \right. \\
& \left. + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell} (Q_{ij\ell} + 1/\theta_1) \right].
\end{aligned}$$

Moreover, $\xi(\theta_1; \beta)$ can be estimated by

$$\begin{aligned}
\hat{\xi}(\theta_1; \beta) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \left[\frac{1}{\theta_1 L_{ij}} \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} \mathbf{Z}_{ij\ell} + \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} \right. \\
& + \frac{(1 + \theta_1 \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell})}{L_{ij}^2} \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} \mathbf{Z}_{ij\ell} + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell} Q_{ij\ell} \mathbf{Z}_{ij\ell} \\
& \left. - \frac{(1/\theta_1 + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell})}{L_{ij}} \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} \mathbf{Z}_{ij\ell} (\theta_1 Q_{ij\ell} + 1) \right],
\end{aligned}$$

while $\xi(\theta_1; \Lambda)(t)$ can be estimated by

$$\begin{aligned}
\hat{\xi}(\theta_{10}; \Lambda)(t) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{\ell=1}^{m_{ij}} e^{\mathbf{Z}_{ij\ell}(t)\beta^T} Y_{ij\ell}(t) \left[\theta_1^{-1} L_{ij}^{-1} \exp(\theta_1 Q_{ij\ell}) \right. \\
& + (1 + \theta_1 \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}) \sum_{\ell=1}^{m_{ij}} \exp(\theta_1 Q_{ij\ell}) Q_{ij\ell} L_{ij}^{-2} \exp(\theta_1 Q_{ij\ell}) \\
& \left. - (\theta_1^{-1} + \sum_{\ell=1}^{m_{ij}} \delta_{ij\ell}) L_{ij}^{-1} \exp(\theta_1 Q_{ij\ell}) (1 + \theta_1 Q_{ij\ell}) + \delta_{ij\ell} \right].
\end{aligned}$$

3.3. Formulas for the asymptotic variance of $\hat{\theta}_2$ when $\theta_2 > 0$

Suppose first the joint likelihood (6) is used. For $i \in \{1, \dots, n\}$, let $B_{1i} = \Gamma(1/\theta_2 + \sum_{j=1}^{m_i} \delta_{ij}^R) / \Gamma(1/\theta_2)$ and $B_{2i}(\theta_2) = \{\partial B_{1i}(\theta_2)\} / (\partial \theta_2)$. Then

$$\begin{aligned} \Xi_i &= \frac{B_{2i}(\theta_2)}{B_{1i}(\theta_2)} + \frac{(1/\theta_2 + \sum_{j=1}^{m_i} \delta_{ij}^R)(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} \log \tilde{v}_{ij})}{\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1} \\ &+ \frac{1}{\theta_2} \log\left(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1\right) - \sum_{j=1}^{m_i} \delta_{ij}^v (\log \tilde{v}_{ij} - 1/\theta_2) \end{aligned}$$

and

$$\begin{aligned} I_2(\theta_2) &= -\frac{1}{n} \sum_{i=1}^n \left[-\frac{\left\{ \frac{\partial}{\partial \theta_2} \Gamma(1/\theta_2 + \sum_{j=1}^{m_i} \delta_{ij}^R) \right\}^2}{\Gamma(1/\theta_2 + \sum_{j=1}^{m_i} \delta_{ij}^R)^2} + \frac{\left\{ \frac{\partial}{\partial \theta_2} \Gamma(\theta_2^{-1}) \right\}^2}{\{\Gamma(1/\theta_2)\}^2} \right. \\ &+ \frac{1}{\Gamma(1/\theta_2 + \sum_{j=1}^{m_i} \delta_{ij}^R)} \frac{\partial^2}{\partial \theta_2^2} \Gamma\left(\frac{1}{\theta_2} + \sum_{j=1}^{m_i} \delta_{ij}^R\right) - \frac{1}{\Gamma(1/\theta_2)} \frac{\partial^2}{\partial \theta_2^2} \Gamma\left(\frac{1}{\theta_2}\right) - \sum_{j=1}^{m_i} \frac{\delta_{ij}^R}{\theta_2^2} \\ &- \frac{2 \log(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1)}{\theta_2^3} - \frac{2 \sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} \log(\tilde{v}_{ij})}{\theta_2^2 (\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1)} + \left(\frac{1}{\theta_2} + \sum_{j=1}^{m_i} \delta_{ij}^R\right) \times \\ &\left. \frac{(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} \log \tilde{v}_{ij})^2 - \{\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} (\log \tilde{v}_{ij})^2\} (\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1)}{(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1)^2} \right]. \end{aligned}$$

Finally, $\xi(\theta_2, \theta_1)$ can be estimated by

$$\begin{aligned} \hat{\xi}(\theta_2, \theta_1) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2-1} \left(\frac{\partial}{\partial \theta_1} \tilde{v}_{ij} \right) \left[\frac{(1 + \theta_2 \sum_{j=1}^{m_i} \delta_{ij}^R)}{(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1)^2} \left(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} \log \tilde{v}_{ij} \right) \right. \\ &+ \left. \frac{(1/\theta_2 + \sum_{j=1}^{m_i} \delta_{ij}^R)}{(\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1)} (1 - \theta_2 \log \tilde{v}_{ij}) - \frac{1}{\theta_2 (\sum_{j=1}^{m_i} \tilde{v}_{ij}^{-\theta_2} - m_i + 1)} - \delta_{ij}^R \tilde{v}_{ij}^{\theta_2} \right]. \end{aligned}$$

When the composite likelihood (7) is used, we have

$$\Xi_i^c = \sum_{j=1}^{m_i} \sum_{k=j+1}^{m_i} \frac{\frac{\partial}{\partial \theta_2} B_{i,jk}^c(\theta_2)}{B_{i,jk}^c(\theta_2)} - \frac{\frac{\partial}{\partial \theta_2} D_{ij}(\theta_2)}{D_{ij}(\theta_2)} - \frac{\frac{\partial}{\partial \theta_2} D_{ik}(\theta_2)}{D_{ik}(\theta_2)}$$

where

$$\begin{aligned} B_{i,jk}^c(\theta_2) &= \frac{\Gamma(2 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} (\tilde{v}_{ij}^{-\theta_2} + \tilde{v}_{ik}^{-\theta_2} - 1)^{-(2 + \theta_2^{-1})}, \\ D_{ij}^c(\theta_2) &= \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} (\tilde{v}_{ij}^{1 + \theta_2}) \quad \text{and} \quad D_{ik}^c(\theta_2) = \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} (\tilde{v}_{ik}^{1 + \theta_2}). \end{aligned}$$

Moreover,

$$I_{2,c}(\theta_2) = \frac{-1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=j+1}^{m_i} \left\{ \frac{\left(\frac{\partial^2}{\partial \theta_2^2} B_{i,jk}^c(\theta_2) \right) B_{i,jk}^c(\theta_2) - \left[\frac{\partial}{\partial \theta_2} B_{i,jk}^c(\theta_2) \right]^2}{\left(B_{i,jk}^c(\theta_2) \right)^2} \right. \\ \left. - \frac{\left(\frac{\partial^2}{\partial \theta_2^2} D_{ij}^c(\theta_2) \right) D_{ij}^c(\theta_2) - \left[\frac{\partial}{\partial \theta_2} D_{ij}^c(\theta_2) \right]^2}{\left(D_{ij}^c(\theta_2) \right)^2} \right. \\ \left. - \frac{\left(\frac{\partial^2}{\partial \theta_2^2} D_{ik}^c(\theta_2) \right) D_{ik}^c(\theta_2) - \left[\frac{\partial}{\partial \theta_2} D_{ik}^c(\theta_2) \right]^2}{\left(D_{ik}^c(\theta_2) \right)^2} \right\}$$

where

$$\frac{\partial}{\partial \theta_2} B_{i,jk}^c(\theta_2) = \left(\frac{\partial}{\partial \theta_2} \frac{\Gamma(2 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \right) (\tilde{v}_{ij}^{-\theta_2} + \tilde{v}_{ik}^{-\theta_2} - 1)^{-(2+\theta_2^{-1})} \\ + \frac{\Gamma(2 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \frac{\partial}{\partial \theta_2} (\tilde{v}_{ij}^{-\theta_2} + \tilde{v}_{ik}^{-\theta_2} - 1)^{-(2+\theta_2^{-1})},$$

$$\frac{\partial^2}{\partial \theta_2^2} B_{i,jk}^c(\theta_2) = \left(\frac{\partial^2}{\partial \theta_2^2} \frac{\Gamma(2 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \right) (\tilde{v}_{ij}^{-\theta_2} + \tilde{v}_{ik}^{-\theta_2} - 1)^{-(2+\theta_2^{-1})}, \\ + 2 \left(\frac{\partial}{\partial \theta_2} \frac{\Gamma(2 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \right) \left(\frac{\partial}{\partial \theta_2} (\tilde{v}_{ij}^{-\theta_2} + \tilde{v}_{ik}^{-\theta_2} - 1)^{-(2+\theta_2^{-1})} \right) \\ + \left(\frac{\Gamma(2 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \right) \left(\frac{\partial^2}{\partial \theta_2^2} (\tilde{v}_{ij}^{-\theta_2} + \tilde{v}_{ik}^{-\theta_2} - 1)^{-(2+\theta_2^{-1})} \right),$$

$$\frac{\partial}{\partial \theta_2} D_{ij}^c(\theta_2) = \tilde{v}_{ij}^{1+\theta_2} \left[\frac{\partial}{\partial \theta_2} \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} + \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \log \tilde{v}_{ij} \right],$$

$$\frac{\partial^2}{\partial \theta_2^2} D_{ij}^c(\theta_2) = \tilde{v}_{ij}^{1+\theta_2} \left[\frac{\partial}{\partial \theta_2^2} \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} + 2 \log(\tilde{v}_{ij}) \frac{\partial}{\partial \theta_2} \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \right. \\ \left. + (\log \tilde{v}_{ij})^2 \frac{\Gamma(1 + \theta_2^{-1})}{\theta_2^{-1}} \right]$$

$$\frac{\partial}{\partial \theta_2} D_{ik}^c(\theta_2) = \tilde{v}_{ik}^{1+\theta_2} \left[\frac{\partial}{\partial \theta_2} \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} + \frac{\Gamma(1 + \theta_2^{-1})}{\Gamma(\theta_2^{-1})} \log \tilde{v}_{ik} \right],$$

$$\begin{aligned} \frac{\partial^2}{\partial \theta_2^2} D_{ik}^c(\theta_2) &= \tilde{v}_{ik}^{1+\theta_2} \left[\frac{\partial}{\partial \theta_2^2} \frac{\Gamma(1+\theta_2^{-1})}{\Gamma(\theta_2^{-1})} + 2 \log(\tilde{v}_{ik}) \frac{\partial}{\partial \theta_2} \frac{\Gamma(1+\theta_2^{-1})}{\Gamma(\theta_2^{-1})} \right. \\ &\quad \left. + (\log \tilde{v}_{ik})^2 \frac{\Gamma(1+\theta_2^{-1})}{\theta_2^{-1}} \right]. \end{aligned}$$

4. ADDITIONAL SIMULATION RESULTS

This section gathers results of additional simulation studies. Table 3 reports the finite sample performance of the marginal and association parameter estimators under Scenario B. Tables 4 and 5 show the finite sample performance of the marginal and association parameter estimators under Scenario B. Finally, Tables 6 and 7 display results under Scenario C. The design and both scenarios are described in the main paper.

Table 8 displays results of a simulation study that was run in order to assess the performance of the estimators of θ_1 and θ_2 , and the coverage rate of the confidence intervals for these parameters for sample size n and cluster sizes that match those observed in the CGD dataset. Specifically, we set $n = 13$ and $(m_1, \dots, m_{13}) = (4, 16, 4, 26, 8, 9, 4, 4, 6, 16, 8, 19, 4)$; the average of the cluster sizes m_i is 9.8. The sizes of the lower-level clusters m_{ij} were also chosen exactly as in the data; their values range from 1 to 8 with an average of 1.6. For each $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m_i\}$ and $\ell \in \{1, \dots, m_{ij}\}$, we generated a Bernoulli(0.5) variable $Z_{ij\ell}$ and drew $T_{ij\ell}$ conditionally on $Z_{ij\ell}$ from the Cox proportional hazard model with $\beta = -1.0860$ and a Weibull baseline hazard function with $a = 0.91$ and $b = 52.57$. The survival times $T_{ij\ell}$ were subsequently censored by an independent censoring variable, uniform on $(0, c)$; c was chosen as to achieve the censoring rate of 62.5% observed in the CGD data. Furthermore, the generators are Clayton with parameters $\theta_1 = 0.4447$ and $\theta_2 = 0.0780$ which correspond to Kendall's τ of 0.18 and 0.04, respectively. These parameter values are equal to the estimates $\hat{\theta}_1$ and $\hat{\theta}_{2,K}^c$ from the CGD data analysis.

Finally, Table 9 shows results of a simulation study assessing the level of the tests based on $T_{1,K}$ and $T_{2,K}$.

TABLE 3: Finite-sample performance of the association parameter estimates under Scenario A.

(n, m_1, m_2)	Est.	Cens. rate	Av. cens. rate of \hat{R}	Bias	SE	ESE	RMSE	CR	
(50, 10, 5)	$\hat{\beta}_1$	10%	38%	0.0109	0.0526	0.0515	0.0526	93.3%	
		30%	67%	0.0025	0.0574	0.0534	0.0535	92.4%	
		70%	96%	0.0035	0.0781	0.0746	0.0747	93.4%	
	$\hat{\beta}_2$	10%	38%	0.0106	0.0953	0.0941	0.0947	92.4%	
		30%	67%	0.0126	0.1008	0.0952	0.0960	93.4%	
		70%	96%	0.0060	0.1233	0.1306	0.1307	95.8%	
	$\hat{\theta}_1$	10%	38%	-0.0328	0.1722	0.1707	0.1738	94.7%	
		30%	67%	-0.0224	0.1957	0.1939	0.1952	93.6%	
		70%	96%	-0.0154	0.2579	0.2784	0.2788	94.2%	
	$\bar{\theta}_{2,K}$	10%	38%	-0.0694	0.1136	0.1035	0.1246	93.9%	
		30%	67%	-0.1246	0.1338	0.1257	0.1770	88.8%	
		70%	96%	-0.1674	0.2356	0.2277	0.2826	81.6%	
	$\bar{\theta}_{2,K}^c$	10%	38%	-0.0452	0.1636	0.1597	0.1660	93.4%	
		30%	67%	-0.0857	0.1732	0.1716	0.1918	93.6%	
		70%	96%	-0.1312	0.2864	0.2781	0.3075	85.2%	
	(200, 10, 5)	$\hat{\beta}_1$	10%	38%	0.0010	0.0301	0.0294	0.0294	94.6%
			30%	67%	0.0029	0.0322	0.0316	0.0317	94.0%
			70%	96%	-0.0018	0.0379	0.0377	0.0377	93.6%
$\hat{\beta}_2$		10%	38%	0.0097	0.0605	0.0622	0.0630	94.4%	
		30%	67%	0.0080	0.0636	0.0631	0.0636	95.0%	
		70%	96%	0.0029	0.0664	0.0656	0.0656	94.0%	
$\hat{\theta}_1$		10%	38%	-0.0116	0.1073	0.1038	0.1044	96.4%	
		30%	67%	-0.0135	0.1147	0.1128	0.1136	95.2%	
		70%	96%	-0.0015	0.1257	0.1236	0.1236	94.0%	
$\bar{\theta}_{2,K}$		10%	38%	-0.0523	0.0813	0.0817	0.0970	94.9%	
		30%	67%	-0.0978	0.0854	0.0844	0.1292	92.8%	
		70%	96%	-0.1362	0.1348	0.1337	0.1908	88.2%	
$\bar{\theta}_{2,K}^c$		10%	38%	-0.0415	0.0855	0.0867	0.0961	94.2%	
		30%	67%	-0.0657	0.0938	0.0929	0.1138	94.6%	
		70%	96%	-0.1012	0.1783	0.1766	0.2035	92.8%	

TABLE 4: Finite-sample performance of the association parameter estimates under Scenario B.

(n, m_1, m_2)	Est.	Cens. rate	Av. cens. rate of \hat{R}	Bias	SE	ESE	RMSE	CR	
(100, 10, 3)	$\hat{\beta}_1$	10%	26%	0.0082	0.0517	0.0494	0.0501	93.6%	
		30%	55%	0.0109	0.0525	0.0514	0.0525	94.0%	
		70%	92%	0.0016	0.0674	0.0689	0.0689	94.0%	
	$\hat{\beta}_2$	10%	26%	0.0052	0.0904	0.0891	0.0893	94.0%	
		30%	55%	0.0128	0.0915	0.0907	0.0916	95.0%	
		70%	92%	-0.0096	0.1219	0.1204	0.1208	94.3%	
	$\hat{\theta}_1$	10%	26%	-0.0250	0.1506	0.1476	0.1497	94.0%	
		30%	55%	-0.0326	0.1593	0.1561	0.1595	96.0%	
		70%	92%	0.0253	0.2354	0.2691	0.2702	94.8%	
	$\bar{\theta}_{2,K}$	10%	26%	-0.0657	0.0915	0.0908	0.1121	90.0%	
		30%	55%	-0.1133	0.1096	0.1087	0.1570	87.0%	
		70%	92%	-0.1571	0.1698	0.1684	0.2303	78.6%	
	$\bar{\theta}_{2,K}^c$	10%	26%	-0.0573	0.1141	0.1172	0.1305	92.0%	
		30%	55%	-0.0882	0.1227	0.1215	0.1501	89.0%	
		70%	92%	-0.1155	0.1938	0.1922	0.2242	84.5%	
	(100, 10, 5)	$\hat{\beta}_1$	10%	36%	0.0050	0.0377	0.0371	0.0374	91.3%
			30%	67%	0.0040	0.0392	0.0388	0.0390	94.1%
			70%	96%	0.0023	0.0554	0.0535	0.0535	94.0%
		$\hat{\beta}_2$	10%	36%	-0.0007	0.0680	0.0669	0.0669	91.7%
			30%	67%	0.0161	0.0721	0.0696	0.0714	93.3%
			70%	96%	0.0017	0.0997	0.0945	0.0945	93.5%
$\hat{\theta}_1$		10%	36%	-0.0077	0.1184	0.1080	0.1083	96.7%	
		30%	67%	-0.0347	0.1259	0.1244	0.1291	94.0%	
		70%	96%	0.0011	0.1961	0.2084	0.2084	93.0%	
$\bar{\theta}_{2,K}$		10%	36%	-0.0614	0.0983	0.0921	0.1107	94.8%	
		30%	67%	-0.1066	0.1156	0.1036	0.1486	91.0%	
		70%	96%	-0.1411	0.1671	0.1653	0.2173	83.4%	
$\bar{\theta}_{2,K}^c$		10%	36%	-0.0383	0.1269	0.1145	0.1207	94.0%	
		30%	67%	-0.0738	0.1407	0.1337	0.1527	93.2%	
		70%	96%	-0.1031	0.2031	0.1987	0.2239	88.0%	

TABLE 5: Finite-sample performance of the association parameter estimates under Scenario B continued.

(n, m_1, m_2)	Est.	Cens. rate	Av. cens. rate of \hat{R}	Bias	SE	ESE	RMSE	CR
(100, 30, 5)	$\hat{\beta}_1$	10%	36%	0.0054	0.0285	0.0277	0.0282	94.0%
		30%	67%	0.0096	0.0319	0.0295	0.0310	94.6%
		70%	96%	0.0011	0.0413	0.0405	0.0405	94.0%
	$\hat{\beta}_2$	10%	36%	0.0045	0.0484	0.0472	0.0474	95.2%
		30%	67%	0.0089	0.0578	0.0577	0.0584	94.2%
		70%	96%	0.0024	0.0779	0.0738	0.0738	93.3%
	$\hat{\theta}_1$	10%	36%	-0.0159	0.0989	0.0938	0.0951	95.3%
		30%	67%	-0.0204	0.1161	0.1034	0.1054	96.0%
		70%	96%	0.0175	0.1785	0.1811	0.1819	93.6%
	$\bar{\theta}_{2,K}$	10%	36%	-0.0701	0.0711	0.0717	0.1003	88.0%
		30%	67%	-0.1369	0.0747	0.0738	0.1555	83.0%
		70%	96%	-0.1701	0.1338	0.1341	0.2166	76.0%
	$\bar{\theta}_{2,K}^c$	10%	36%	-0.0384	0.0839	0.0833	0.0917	94.2%
		30%	67%	-0.0863	0.0876	0.0866	0.1223	93.6%
		70%	96%	-0.1084	0.1457	0.1441	0.1803	93.6%

TABLE 6: Finite sample performance of the association parameter estimates under Scenario C.

(n, m_1, m_2)	Est.	Cens. rate	Av. cens. rate of \hat{R}	Bias	SE	ESE	RMSE	CR
(100,10,Poi(5))	$\hat{\beta}_1$	10%	42%	0.0021	0.0364	0.0347	0.0348	90.0%
		30%	70%	0.0033	0.0370	0.0357	0.0358	92.3%
		70%	97%	-0.0022	0.0488	0.0492	0.0492	93.6%
	$\hat{\beta}_2$	10%	42%	0.0058	0.0669	0.0647	0.0650	92.8%
		30%	70%	0.0143	0.0706	0.0656	0.0671	93.6%
		70%	97%	0.0012	0.0864	0.0856	0.0856	94.6%
	$\hat{\theta}_1$	10%	42%	-0.0044	0.1209	0.1302	0.1303	93.2%
		30%	70%	-0.0146	0.1302	0.1321	0.1330	94.7%
		70%	97%	0.0083	0.1668	0.1742	0.1744	96.0%
	$\bar{\theta}_{2,K}$	10%	42%	-0.0638	0.0972	0.0966	0.1183	94.0%
		30%	70%	-0.1166	0.1054	0.1032	0.1557	91.6%
		70%	97%	-0.1476	0.1625	0.1618	0.2190	84.2%
	$\bar{\theta}_{2,K}^s$	10%	42%	-0.0434	0.1096	0.1078	0.1162	93.8%
		30%	70%	-0.0754	0.1436	0.1411	0.1600	93.6%
		70%	97%	-0.1081	0.2066	0.2054	0.2321	86.6%
(100,Poi(10),5)	$\hat{\beta}_1$	10%	36%	0.0032	0.0371	0.0355	0.0356	93.0%
		30%	67%	0.0068	0.0381	0.0363	0.0369	94.0%
		70%	96%	-0.0010	0.0521	0.0516	0.0516	93.6%
	$\hat{\beta}_2$	10%	36%	0.0115	0.0647	0.0654	0.0664	92.6%
		30%	67%	0.0147	0.0709	0.0687	0.0702	92.6%
		70%	96%	0.0028	0.0927	0.0884	0.0884	92.8%
	$\hat{\theta}_1$	10%	36%	-0.0069	0.1148	0.1068	0.1070	94.5%
		30%	67%	-0.0242	0.1229	0.1167	0.1192	93.6%
		70%	96%	0.0026	0.1813	0.2000	0.2000	97.1%
	$\bar{\theta}_{2,K}$	10%	36%	-0.0653	0.1030	0.1132	0.1306	93.6%
		30%	67%	-0.1168	0.1192	0.1200	0.1674	90.2%
		70%	96%	-0.1488	0.1682	0.1697	0.2257	83.6%
	$\bar{\theta}_{2,K}^s$	10%	36%	-0.0401	0.1395	0.1266	0.1327	94.5%
		30%	67%	-0.0787	0.1607	0.1533	0.1723	93.6%
		70%	96%	-0.1164	0.2136	0.2077	0.2381	88.2%

TABLE 7: Finite sample performance of the association parameter estimates under Scenario C continued.

(n, m_1, m_2)	Est.	Cens. rate	Av. cens. rate of \hat{R}	Bias	SE	ESE	RMSE	CR
(100, Poi(10), Poi(5))	$\hat{\beta}_1$	10%	42%	0.0038	0.0329	0.0342	0.0344	95.4%
		30%	70%	0.0055	0.0378	0.0362	0.0366	93.6%
		70%	97%	0.0016	0.0482	0.0465	0.0465	92.6%
	$\hat{\beta}_2$	10%	42%	0.0060	0.0674	0.0637	0.0640	92.8%
		30%	70%	0.0088	0.0705	0.0687	0.0691	92.6%
		70%	97%	-0.0014	0.0856	0.0812	0.0812	93.0%
	$\hat{\theta}_1$	10%	42%	-0.0316	0.1221	0.1278	0.1316	91.4%
		30%	70%	-0.0182	0.1278	0.1388	0.1400	95.0%
		70%	97%	0.0078	0.1614	0.1683	0.1684	94.8%
	$\bar{\theta}_{2,K}$	10%	42%	-0.0658	0.1069	0.1033	0.1224	94.5%
		30%	70%	-0.1192	0.1244	0.1201	0.1692	90.0%
		70%	97%	-0.1557	0.1793	0.1755	0.2346	83.8%
$\bar{\theta}_{2,K}^c$	10%	42%	-0.0399	0.1327	0.1298	0.1357	95.0%	
	30%	70%	-0.0883	0.1631	0.1611	0.1837	93.6%	
	70%	97%	-0.1159	0.2284	0.2156	0.2448	88.4%	

TABLE 8: Small sample performance of the association parameter estimates with sample and cluster sizes as in the CGD data.

(n, m_1, m_2)	Est.	Cens. rate	Av. cens. rate of \hat{R}	Bias	SE	ESE	RMSE	CR
	$\hat{\beta}_1$	62.5%	90%	-0.0252	0.2496	0.2295	0.2308	95.8%
	$\hat{\theta}_1$	62.5%	90%	-0.0156	0.3225	0.3237	0.3241	90.0%
	$\bar{\theta}_{2,K}$	62.5%	90%	-0.0367	0.0199	0.0211	0.0423	84.5%
	$\bar{\theta}_{2,K}^c$	62.5%	90%	-0.0322	0.0288	0.0233	0.0397	88.6%

TABLE 9: Level of the tests based on $T_{1,K}$ and $T_{2,K}$.

Cens. rate	(τ_1, τ_2)	$(200, 10, m_2)$			$(100, m_1, 5)$			$(n, 20, 5)$		
		2	3	5	10	30	50	100	200	300
Test based on $T_{1,K}$										
0%	(0.1,0.3)	5.0%	6.6%	4.8%	6.0%	4.6%	5.8%	5.4%	5.0%	6.0%
	(0.3,0.3)	6.4%	5.0%	6.6%	3.6%	4.5%	6.4%	6.4%	6.4%	4.8%
	(0.5,0.3)	6.4%	6.0%	6.4%	5.6%	5.2%	6.4%	6.4%	5.8%	5.2%
	(0.7,0.3)	5.6%	6.0%	4.4%	6.4%	6.2%	5.8%	6.0%	6.0%	4.2%
30%	(0.1,0.3)	6.4%	4.8%	3.8%	3.4%	4.0%	4.0%	4.6%	4.8%	4.5%
	(0.3,0.3)	4.2%	5.4%	4.0%	3.8%	3.8%	4.4%	3.7%	3.2%	3.3%
	(0.5,0.3)	6.0%	3.8%	3.6%	3.2%	4.0%	3.6%	4.2%	4.2%	4.4%
	(0.7,0.3)	3.6%	4.0%	4.2%	3.6%	3.4%	3.2%	3.6%	4.8%	4.6%
Test based on $T_{2,K}$										
0%	(0.5,0.1)	5.2%	5.8%	4.4%	5.2%	3.6%	5.2%	6.8%	4.4%	4.8%
	(0.5,0.3)	7.2%	4.0%	5.4%	5.0%	6.8%	5.0%	5.8%	5.2%	6.0%
	(0.5,0.5)	6.0%	5.4%	4.0%	3.6%	6.0%	5.0%	6.2%	5.8%	5.0%
	(0.5,0.7)	4.2%	4.6%	5.8%	5.2%	4.6%	6%	6.2%	5.4%	4.2%
30%	(0.5,0.1)	3.4%	3.8%	3.6%	3.2%	4.6%	5.4%	4.0%	5.4%	4.0%
	(0.5,0.3)	6.3%	5.2%	4.6%	6.4%	7.0%	5.6%	6.4%	4.2%	4.7%
	(0.5,0.5)	5.7%	4.4%	4.0%	3.6%	4.2%	4.3%	3.8%	3.6%	4.0%
	(0.5,0.7)	3.8%	3.6%	5.6%	4.3%	5.8%	5.4%	4.2%	3.8%	4.2%

5. IMPLEMENTATION

To run the procedures proposed in this paper and to analyze the data from the study of chronic granulomatous disease (CGD), load the following packages and source functions; the code below is available in `DemoSNW.R`.

```
library(survival)
library(copula)
library(numDeriv)
source('CodeSNW.R')
```

Data from the hierarchical Kendall copula with Clayton generators can be generated using the function `rCopulaHKCC`. First, determine the cluster sizes.

```
culevel <- c(10, 4, 4, 10, 4, 6, 6)
cllevel <- c(rep(5, 10), 1, 4, 5, 6, rep(2, 4),
             rep(5, 10), 2, 3, 4, 2, rep(3, 6), rep(2, 6))
```

In this example, there are 7 clusters at the upper level with sizes given in `culevel`; the sizes of the lower level clusters are stored in `cllevel`. A sample from the hierarchical Kendall copula with Clayton clusters is then

```
U <- rCopulaHKCC(ulevel=culevel, llevel=cllevel, theta1=2, theta2=1,
                 method="simple")
```

The parameter `method` has one of the values "default", "simple", "sort", "discrete", "monoH.FC" and is passed on to the function `qK` from the package `copula` for the inversion of the Kendall function. The output `U` is in the form of a list, the i -th element is again a list such that U_{ij} can be recovered as `U[[i]][[j]]`.

Next, `U` can be transformed into `T`, which corresponds to the vector T of failure times using the function `rFailure`. Here, only the Weibull hazard rate is implemented and two covariates, one uniformly distributed and the other Bernoulli with parameter `prob`.

```
T <- rFailure(U, a=0.91, b=52.5779, prob=0.5, beta=c(0.5, 1))
```

The function `rFailure` can be easily modified to accommodate a different number of covariates and/or their distributions. The output `T` is in the same format as `U`, except that the values of the covariates are also stored. The vector T_{ij} of failure times can be retrieved as `T[[i]][[j]]$failure` and the corresponding covariates are stored as a $m_{ij} \times 2$ matrix in `T[[i]][[j]]$covariates`.

The failure times can then be censored by independent censoring variables that are assumed to be uniform on $(0, c)$ using the function `rCensor`.

```
X <- rCensor(T, c=300)
```

The format of `X` is again a list of lists as `T` and `U`. The observed failure times X_{ij} , censoring indicators δ_{ij} and the corresponding covariates can be retrieved as `X[[i]][[j]]$failure`, `X[[i]][[j]]$delta`, and `X[[i]][[j]]$covariates`, respectively. The observed censoring rate can be calculated as follows.

```
cens <- unlist(lapply(X, FUN=function(l) {
  lapply(l, FUN=function(ml) {
    length(ml$delta) - sum(ml$delta)
  })
}))
sum(cens) / sum(clevel)
```

Finally, X can be turned into a data frame using the function `getMyDF`.

```
data <- getMyDF(X)
# head(data)
#   ulevel llevel   failure delta      cov1 cov2
#1      1      1  4.5385334      1 0.20103526  1
#2      1      1  6.2049466      1 0.01022154  0
#3      1      1  1.1793810      1 0.71061659  1
#4      1      1 12.9517248      1 0.42965971  1
#5      1      1 16.7325076      1 0.11957299  0
#6      1      2  0.5041147      1 0.28785984  0
```

To fit the hierarchical Kendall copula model with Clayton clusters, first fit the Cox proportional hazard model. The example below assumes that the data frame contains the columns `ulevel` and `llevel` (indicating the cluster structure), `failure` (failure times), and `delta` (censoring indicator), and that the covariates are stored in columns 5 and above in the data frame.

```
xnam <- c(names(data)[5:ncol(data)], "cluster(ulevel)")
fmla <- as.formula(paste("Surv(failure,delta) ~ ",
  paste(xnam, collapse=" +")))
results <- coxph(fmla, robust=T, data=data, x=TRUE)
est_beta <- as.numeric(summary(results)$coefficients[,1])
# est_beta
# [1] 0.7435947 1.4327068
```

The pseudo observations \hat{U} can then be computed as follows; the format of `Uhat` is again a list of lists as X above.

```
covariates <- as.matrix(results$x)
pseudos <- getPseudos(failure=data$failure, delta=data$delta,
  covariates=covariates, beta=est_beta)
Uhat <- getMylist(failure=pseudos, delta=data$delta,
  covariates=covariates, ulevel=data$ulevel, llevel=data$llevel)
# Uhat[[1]][[2]]
# $failure
# [1] 0.4995937 0.5072516 0.7129800 0.4911173 0.2094866
#
# $delta
# [1] 1 1 0 1 1
#
# $covariates
#           [,1] [,2]
# [1,] 0.9619747  0
# [2,] 0.6622585  1
# [3,] 0.8820604  1
# [4,] 0.3467289  0
# [5,] 0.7248627  1
```

The estimate of θ_1 can be calculated using the function `fitTheta1`. The input are the pseudo-observations `Uhat` and the parameter method is passed on to the R optimization routine `optim`.

```
theta1 <- fitTheta1(Uhat, start=c(1), method="BFGS")
# theta1
# [1] 1.626126
```

The estimate of θ_2 can be calculated using the function `fitTheta2`. The input is again `Uhat`, as well as the estimate `theta1` of θ_1 , a boolean variable `pairwise` indicating whether the pair-wise likelihood is used, the number of imputations `nimp`, the starting value `start` and the optimization method that is passed on to `optim`.

```
theta2 <- fitTheta2(Uhat, theta1=theta1, pairwise=FALSE, nimp=100,
  start=1, method="BFGS")
theta2p <- fitTheta2(Uhat, theta1=theta1, pairwise=TRUE, nimp=100,
  start=1, method="BFGS")
# theta2
# [1] 0.8782442
# theta2p
# [1] 0.5830707
```

The test of the hypothesis that the lower level cluster copula is Clayton is implemented through the function `gofLlevel`. Its input is `Uhat`, the estimate `theta1` of θ_1 , the number `nimp` of imputations and the number `nboot` of bootstrap samples.

```
gofLlevel(Uhat, theta1=theta1, nimp=100, nboot=100)
# $statistic
# [1] -0.8574091
#
# $p.value
# [1] 0.3912202
```

The test of the hypothesis that the upper level cluster copula is Clayton is implemented in the function `gofUlevel`. Its input is `Uhat`, the estimate `theta1` of θ_1 , the estimate `theta2` of θ_2 , and the number `nimp` of imputations.

```
gofUlevel(Uhat, theta1=theta1, theta2=theta2, nimp=100)
# $statistic
# [1] -1.669194
#
# $p.value
# [1] 0.09508078
```

Finally, the standard errors of $\hat{\theta}_1$ and $\hat{\theta}_2$ may be computed as follows.

```
# Standard error of theta1
output_se_theta1=se_Theta1(data, method="BFGS", start=1)
# as.vector(output_se_theta1$se_theta1)
# [1] 0.5968761
# Standard error of theta2 using the full likelihood
output_se_theta2=se_Theta2(data, pairwise=FALSE, nimp=100,
  es_theta1=theta1, Ky=output_se_theta1$Ky, output_se_theta1$Info_theta1)
# output_se_theta2$se_theta2
```

```
# [1] 0.2520027
# Standard error of theta2 using the pair-wise likelihood
output_se_theta2=se_Theta2(data, pairwise=TRUE, nimp=100,
  es_theta1=theta1, Ky=output_se_theta1$Ky, output_se_theta1$Info_theta1)
# output_se_theta2$se_theta2
# [1] 0.2986248
```

The data from the study of chronic granulomatous disease (CGD) is freely available in the R package survival.

```
library(survival)
data(cgd)
```

It can be put in the format that is required to run the above code as follows.

```
len=length(cgd[,1])
sex=rep(0, len)
treatment=rep(0, len)
Hospital=rep(0, len)
Hospital[which(levels(cgd$center)[1]==as.character(cgd$center))]=1
Hospital[which(levels(cgd$center)[2]==as.character(cgd$center))]=2
Hospital[which(levels(cgd$center)[3]==as.character(cgd$center))]=3
Hospital[which(levels(cgd$center)[4]==as.character(cgd$center))]=4
Hospital[which(levels(cgd$center)[5]==as.character(cgd$center))]=5
Hospital[which(levels(cgd$center)[6]==as.character(cgd$center))]=6
Hospital[which(levels(cgd$center)[7]==as.character(cgd$center))]=7
Hospital[which(levels(cgd$center)[8]==as.character(cgd$center))]=8
Hospital[which(levels(cgd$center)[9]==as.character(cgd$center))]=9
Hospital[which(levels(cgd$center)[10]==as.character(cgd$center))]=10
Hospital[which(levels(cgd$center)[11]==as.character(cgd$center))]=11
Hospital[which(levels(cgd$center)[12]==as.character(cgd$center))]=12
Hospital[which(levels(cgd$center)[13]==as.character(cgd$center))]=13
Patient_id=cgd[,1]
sex[which("male"==as.character(cgd[,5]))]=1
treatment[which("rIFN-g"==as.character(cgd[,4]))]=1
age=cgd[,6]
failure=(cgd[,15]-cgd[,13])
delta=cgd[,16]
data=data.frame(ulevel=Hospital, llevel=Patient_id, failure=failure,
  delta=delta, cov1=treatment)
# using more covariates
# data=data.frame(ulevel=Hospital, llevel=Patient_id, failure=failure,
# delta=delta, cov1=sex, cov2=age, cov3=treatment)
```

BIBLIOGRAPHY

- Hofert, M., Mächler, M., and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *J. Multivariate Anal.*, 110:133–150.
- Speikerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.*, 93:1164–1175.