Composite likelihood estimation method for hierarchical Archimedean copulas defined with multivariate compound distributions

Hélène Cossette [a], Simon-Pierre Gadoury [a], Etienne Marceau [a,*], Christian Y. Robert [b]

[a] *École d'actuariat, Université Laval, 2425, rue de l'Agriculture, Québec (Québec), Canada G1V 0A6*
[b] *Univ Lyon, Université Lyon 1, ISFA, LSAF EA2429, 50, avenue Tony-Garnier, F-69007 Lyon, France*

ARTICLE INFO

ABSTRACT

We consider the family of hierarchical Archimedean copulas obtained from multivariate exponential mixture distribution through compounding, as introduced by Cossette et al. (2017). We investigate ways of determining the structure of these copulas and estimating their parameters. An agglomerative clustering technique based on the matrix of Spearman's rhos, combined with a bootstrap procedure, is used to identify the tree structure. Parameters are estimated through a top-down composite likelihood. The validity of the approach is illustrated through two simulation studies in which the procedure is explained step by step. The composite likelihood method is also compared to the full likelihood method in a simple case where the latter is computable.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Copulas are well-known probabilistic tools which are used to model the dependence within a vector $(X_1, \ldots, X_d)$ of random variables with respective marginal distribution functions $F_{X_1}, \ldots, F_{X_d}$. Assuming that the latter are continuous, the unique underlying copula C associated with this random vector is the joint cumulative distribution function (cdf) of the vector $(U_1, \ldots, U_d) = (F_{X_1}(X_1), \ldots, F_{X_d}(X_d))$ of dependent random variables which are uniformly distributed on the interval $(0, 1)$.

Particularly popular is the family of Archimedean copulas [7,23], which have the useful property of being defined solely in terms of a continuous and strictly decreasing Archimedean generator $\psi : [0, \infty] \to [0, 1]$, where $\psi(0) = 1$ and $\psi(t) \to 0$ as $t \to \infty$. Let Ψ_∞ be the class of all such functions ψ which are completely monotone, i.e., such that for all $(k, t) \in \mathbb{N} \times \mathbb{R}_+$, we have $(-1)^k \psi^{(k)}(t) \geq 0$. Then, we can define a d-dimensional copula as Archimedean with completely monotone generator if there exists $\psi \in \Psi_\infty$ such that, for every integer $d \geq 2$ and all $(u_1, \ldots, u_d) \in (0, 1)^d$,

$$C(u_1, \ldots, u_d) = \psi\{\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\}.$$

While Archimedean copulas have an algebraically closed form in every dimension, they are exchangeable, i.e., for any arbitrary permutation $\pi = (\pi_1, \ldots, \pi_d)$ of $\{1, \ldots, d\}$ and $u_1, \ldots, u_d \in [0, 1]$, we have that $C(u_1, \ldots, u_d) = C(u_{\pi_1}, \ldots, u_{\pi_d})$. This is a major drawback of this class of copulas, especially for applications in high dimensions. To

---

* Corresponding author.
E-mail address: etienne.marceau@act.ulaval.ca (E. Marceau).

allow asymmetry, hierarchical Archimedean copulas have been developed, such as nested Archimedean copulas [18], hierarchical Archimedean copulas constructed via Lévy subordinators [15], and hierarchical Archimedean copulas based on the Kendall distribution [1].

Another approach to the construction of such copulas which circumvents the constraints on nesting conditions was proposed by Cossette et al. [2]. It mainly relies on compound distributions. In both their paper and earlier work by Hering et al. [15], a hierarchical Archimedean copula is constructed from the joint survival function of a multivariate mixed exponential distribution. The dependence structure of the multivariate mixed exponential distribution is defined through a vector of dependent mixing random variables which are defined using Lévy subordinators. While a common random time is considered in [15], random sums with a common counting random variable are used in [2]. Section 4.2 in [2] gives a detailed comparison of these two different approaches with a view to proposing new hierarchical Archimedean copulas which circumvent the nesting condition required in the seminal work by Joe [18].

There is little literature on simultaneous estimation of the structure of hierarchical copulas and their parameters. Okhrin et al. [25] perform these tasks using a bottom-up procedure. A multi-stage procedure is proposed to simultaneously determine the structure of the tree and estimate the parameters of the hierarchical Archimedean copulas. The structure determination is done with different approaches (under the assumption that the copula generators are known) and the parameters are mostly estimated by maximum likelihood but the copula densities are not known analytically and hence must be computed numerically. Okhrin et al. [26] study distributional properties of hierarchical Archimedean copulas but do not investigate any estimation procedure for such structures. Segers and Uyttendaele [29] propose an estimation procedure for the target phylogenetic tree of a nested Archimedean copula which relies on the possibility to represent a hierarchical Archimedean copula as a set of trivariate structures which can then be estimated on their own. This procedure, compared to that of Okhrin et al. [25], does not require any assumption in regard to the generators of the nested Archimedean copula.

Górecki et al. [9] and Górecki and Holeňa [12] propose an estimator for both the determination of the structure and the parameters of the hierarchical Archimedean copulas. The algorithm for agglomerative hierarchical clustering relies on Kendall's tau to measure the similarity between pairs of leaves and on complete-, single- and average-linkage clustering to identify the tree structure. In contrast to Segers and Uyttendaele [29], who only focus on the tree structure determination, Górecki et al. [9] and Górecki and Holeňa [12] propose to fully estimate the target nested Archimedean copula. However, their procedure can only output binary tree structures, which makes their estimation procedure unsuitable for tree structures that are not binary. Note that the structure determination algorithm based on a Kendall correlation matrix was originally proposed by Górecki and Holeň [13]. Górecki et al. [10] and Górecki and Holeňa [12] also use Kendall's tau for structure determination (through hierarchical agglomerative methods with different distances, as well as clustering algorithms) and for parameter estimation (through its inversion).

Uyttendaele [30] revisits estimation procedures previously proposed and presents an alternative to the tree structure estimation approach of Segers and Uyttendaele [29] based on supertrees. Zhu et al. [33] extend the work of Hering et al. [15] on Lévy subordinated hierarchical Archimedean copulas. Notably, they propose an estimation procedure to determine the tree structure of LSHACs and estimate the parameters involved. The optimal tree structure is obtained based on a so-called s-Euclidean metric and the parameters are estimated with an augmented inference-for-margin method. Matsypura et al. [22] propose the use of a network approach to the hierarchical Archimedean copulas estimation problem. Most papers discuss the estimation of hierarchical Archimedean copulas constructed by nesting copulas belonging to the same parametric Archimedean family of copulas. In [11], the estimation of hierarchical Archimedean copulas involving different Archimedean families is tackled with goodness-of-fit testing directly in the estimation procedure due to the need to deal with the sufficient nesting condition.

As previously mentioned, Cossette et al. [2] propose a new hierarchical Archimedean copula construction based on multivariate compound distributions. This technique relies on the construction of a multivariate exponential mixture distribution through compounding. The absence of nesting and marginal conditions, contrary to the nested Archimedean copulas approach, leads to major advantages, such as a flexible range of possible combinations in the choice of distributions, or the existence of explicit formulas for the distribution of the sum. Hering et al. [15] also propose hierarchical Archimedean copulas that circumvent a nesting condition but with more complicated mathematical tools than the random sums used in [2].

Assuming that the multivariate compound distributions are characterized by finite-dimensional parameters, we propose an estimation procedure for the class of copulas considered by Cossette et al. [2]. First we use an agglomerative hierarchical clustering method based on the matrix of Spearman's rho values to estimate the tree structure. Combining this method with a bootstrap procedure, we decide whether it is necessary to group the binary structures. Note that the Spearman's rho matrix is used in the clustering method due to the asymptotic statistical test, provided by Gaißer and Schmid [6], for the hypothesis that all pair-wise Spearman's rank correlation coefficients in a multivariate random vector are equal. To our knowledge, this is the only asymptotic statistical test based on a concordance measure that has been derived from its asymptotic behavior. However, we note that Perreault et al. [28] investigate block exchangeability detection in correlation matrices, which is another way to identify the dependence within the tree structure. Unfortunately, their method is designed for partially-exchangeable tree structures (i.e., one-level tree structures) and hence does not immediately apply in the present context.

As our structure is multi-level, it requires a more general technique. We have thus turned our attention to the theory of hypothesis testing, via the Neyman–Pearson paradigm. Second we introduce a top-down composite likelihood approach

using the property that the parameters of the Archimedean copulas are also naturally defined recursively in a new top-down way. Full likelihood estimation is in general not possible for our hierarchical Archimedean copulas, but when it is, it is computationally prohibitive and not necessarily much more efficient.

This paper is structured as follows. In Section 2 we present the family of multi-level hierarchical Archimedean copulas introduced in [2]. We recall the assumptions of the model, introduce notations and definitions, and provide the Archimedean families of copulas of subvectors of leaves within our tree structure.

In Section 3 we present a two-step procedure that provides a tree structure that suits our multi-level hierarchical Archimedean copulas. This procedure only uses the matrix of Spearman's rho values as input. The first step brings into play an agglomerative hierarchical clustering algorithm, while the second step presents a reconstruction algorithm that transforms the tree derived from the first step to a tree compatible with our model. Note that such a reconstruction is not necessary in the case of nested Archimedean copulas. In practice, we can only use empirical Spearman's rho values that differ from the theoretical values and lead to a binary hierarchical tree structure that can be different from the true structure derived from the first step when using the theoretical Spearman's rho.

In Section 4 we propose a bootstrap procedure to decide whether it is necessary to eliminate nodes of the empirical structure or not, and that is based on statistical hypothesis tests of exchangeability developed in [6]. Such a procedure has been implemented in R and is available in the package erhcv; see [5]. Then, we present our innovative top-down composite likelihood approach using the property that the parameters of the Archimedean copulas are defined recursively in a top-down way. We prove the asymptotic normality of the estimators. In Section 5 two numerical examples illustrate the different steps of the procedure, and a comparison with the full maximum likelihood method is made to show the efficiency of our approach in a test case where the full maximum likelihood can be computed.

2. Hierarchical Archimedean copulas defined with multivariate compound distributions

In this section, we first present the family of multi-level hierarchical Archimedean copulas introduced in [2]. The tree structure is derived via the construction of a multivariate exponential mixture distribution through multi-level compounding. We then explain the multilevel construction of the mixing random vector. The estimation procedure proposed for this family of copulas later on requires a modification of the notation used in Section 2.8 of [2]. The tree structure of these hierarchical copulas follows with this new notation. We then provide the families of Archimedean copulas that characterize the distributions of some subvectors of leaves which are used to compute the composite likelihoods. In general, a letter or a number in bold indicates a vector.

2.1. Definition and notation

Let $\mathbf{Y} = (Y_{1,1}, \ldots, Y_{1,n_1}, \ldots, Y_{d,1}, \ldots, Y_{d,n_d})$ be a vector of $n = n_1 + \cdots + n_d$ random variables which can be more conveniently represented as $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_d)$, where $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,n_i})$ is the vector of n_i random variables for subgroup $i \in \{1, \ldots, d\}$. Given $\boldsymbol{\Theta} = \boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, it is assumed that

$$(Y_{1,1} \mid \boldsymbol{\Theta} = \boldsymbol{\theta}), \ldots, (Y_{1,n_1} \mid \boldsymbol{\Theta} = \boldsymbol{\theta}), \ldots, (Y_{d,1} \mid \boldsymbol{\Theta} = \boldsymbol{\theta}), \ldots, (Y_{d,n_d} \mid \boldsymbol{\Theta} = \boldsymbol{\theta})$$

are conditionally independent. The conditional distributions of the components of $\mathbf{Y}_i$ are only influenced by the component Θ_i of $\boldsymbol{\Theta}$, i.e., for all $i \in \{1, \ldots, d\}$ and $j \in \{1, \ldots, n_i\}$, $Y_{i,j} \mid \boldsymbol{\Theta} = \boldsymbol{\theta}$ has the same distribution as $Y_{i,j} \mid \Theta_i = \theta_i$.

Following Marshall and Olkin [21], we assume that for each $i \in \{1, \ldots, d\}$, $(Y_{i,1} \mid \Theta_i = \theta_i), \ldots, (Y_{i,n_i} \mid \Theta_i = \theta_i)$ are exponentially distributed with parameter θ_i, and that $\boldsymbol{\Theta}$ follows a multivariate compound distribution whose construction will be explained later. Let $\mathcal{L}_{\boldsymbol{\Theta}}$ denote the multivariate Laplace–Stieltjes transform of $\boldsymbol{\Theta}$, i.e., for all $t_1 \geq 0, \ldots, t_d \geq 0$,

$$\mathcal{L}_{\boldsymbol{\Theta}}(t_1, \ldots, t_d) = \mathrm{E}[\exp\{-(t_1\Theta_1 + \cdots + t_d\Theta_d)\}].$$

The multivariate survival function of $\mathbf{Y}$ is $\bar{F}_{\mathbf{Y}}$, i.e., $\bar{F}_{\mathbf{Y}}(\mathbf{y}) = \Pr(Y_{1,1} > y_{1,1}, \ldots, Y_{d,n_d} > y_{d,n_d})$. Let $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_d)$ with $\mathbf{U}_i = (U_{i,1}, \ldots, U_{i,n_i})$ for all $i \in \{1, \ldots, d\}$, whose joint cdf $F_{\mathbf{U}}$ corresponds to the hierarchical Archimedean copula $C_{\mathbf{U}} = F_{\mathbf{U}}$ introduced in Section 2.8 of [2] and is obtained from the multivariate survival function of the multivariate mixed exponential distribution of $\mathbf{Y}$ as follows:

$$C_{\mathbf{U}}(\mathbf{u}) = \bar{F}_{\mathbf{Y}}\{\bar{F}_{Y_{1,1}}^{-1}(u_{1,1}), \ldots, \bar{F}_{Y_{1,n_1}}^{-1}(u_{1,n_1}), \ldots, \bar{F}_{Y_{d,1}}^{-1}(u_{d,1}), \ldots, \bar{F}_{Y_{d,n_d}}^{-1}(u_{d,n_d})\}$$

$$= \mathcal{L}_{\boldsymbol{\Theta}}\left\{\sum_{j=1}^{n_1} \mathcal{L}_{\Theta_1}^{-1}(u_{1,j}), \ldots, \sum_{j=1}^{n_d} \mathcal{L}_{\Theta_d}^{-1}(u_{d,j})\right\},$$

where $\bar{F}_{Y_{i,j}}$ is the survival distribution function of $Y_{i,j}$, with its inverse defined as $\bar{F}_{Y_{i,j}}^{-1}$, $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_d)$ with $\mathbf{u}_i = (u_{i,1}, \ldots, u_{i,n_i})$ for all $i \in \{1, \ldots, d\}$,

We now introduce the tree structure describing the joint distribution of $\boldsymbol{\Theta}$ and the new notation. Note that such a change in comparison with [2] is needed in order for all the leaf nodes to have the same depth ℓ (from the tree's root node), implying that all the leaves are now at the same level. The main reason for this change is that, for our top-down
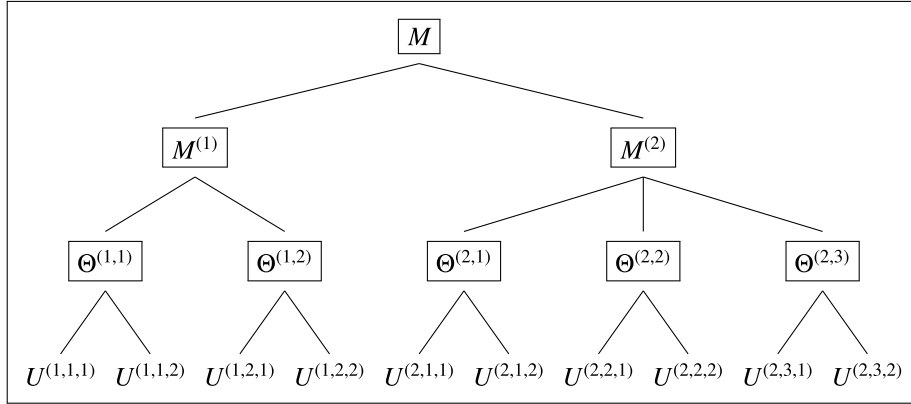
Fig. 1. Illustration of a complex tree structure with $\ell = 3$ levels, discussed in Example 1. Moreover, $a = 2$, $a_{(1)} = 2$, $a_{(2)} = 3$, $b_{\mathbf{i}_2} = 2$, for all $\mathbf{i}_2 \in \mathcal{I}_2$, $d = 5$, and $n = 10$. Note that $r \in \{0, 1, 2\}$.

composite likelihood method, we need to estimate iteratively the parameters of the compound distributions from the root (level 0) to the last node (level $\ell - 1$).

To illustrate the notation, as well as the tree structure, we use the tree depicted in Fig. 1. The root is always associated to a positive and integer-valued random variable M and will be assigned to level $r = 0$. Parent M has children $M^{(1)}$, $M^{(2)}$, ..., i.e., positive random variables associated to the nodes $(1), (2), \ldots$ These children are assigned to level $r = 1$ and are also parents; for example, $M^{(1,1)}, M^{(1,2)}, \ldots$ are the children of $M^{(1)}$. The lineage is then built by iterating until the level $r = \ell - 2$. Finally, the level $r = \ell - 1$ is used for the d children of type Θ born to the parents of type M of level $\ell - 2$. The level $r = \ell$ is for the vector of leaves $\mathbf{U}$. The random variables of type M will only have discrete distributions on $\mathbb{N}_+ = \{1, 2, \ldots\}$, while the random variables of type Θ can have either a discrete distribution on $\mathbb{N}_+$ or a continuous distribution on $\mathbb{R}_+ = (0, \infty)$.

For $r \in \{1, \ldots, \ell - 2\}$, the path from the top node (level 0) to a node at level r is denoted $\mathbf{i}_r = (i_1, \ldots, i_r)$ with $\dim(\mathbf{i}_r) = r$, and provides equivalently its node label. The jth component of $\mathbf{i}_r$ is the position of the node in the path at level j with respect to the subgroup of its parents. To simplify the notation, and although $\mathbf{i}_0$ does not have a mathematical meaning or representation, we let $M^{\mathbf{i}_0} = M$.

For $r \in \{0, \ldots, \ell - 2\}$, we define $a_{\mathbf{i}_r}$ to be the number of children of parent $M^{\mathbf{i}_r}$ under a node with path $\mathbf{i}_r$. As an example, $a_{\mathbf{i}_0} = a$ is the number of children of the root $M^{\mathbf{i}_0}$. We define $b_{(\mathbf{i}_{\ell-2},j)}$ to be the number of leaves below $\Theta^{(\mathbf{i}_{\ell-2},j)}$. We assume that $b_{(\mathbf{i}_{\ell-2},j)} \geq 2$ for $\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}$ and $j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}$. This identifiability condition is necessary to estimate the parameters associated to the parent nodes.

Let us denote by $\mathbf{i}_{r|j}$ the vector of the first j components of a specific path $\mathbf{i}_r$ ($j \leq r$). Then,

$$\mathcal{I}_r = \{\mathbf{i}_r = (i_1, \ldots, i_r) : i_j \in \{1, \ldots, a_{\mathbf{i}_{r|j-1}}\}, j = 1, \ldots, r\} \tag{1}$$

is the set of all possible paths, for a given tree, from the root to the hierarchical level $r \in \{1, \ldots, \ell - 2\}$.

Finally, note that the vector of children of type Θ is given by

$$\boldsymbol{\Theta} = (\Theta^{(\mathbf{i}_{\ell-2},j)} : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\})$$

while the vector of the leaves is given by

$$\mathbf{U} = (\mathbf{U}^{(\mathbf{i}_{\ell-2},j)} : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}),$$

where $\mathbf{U}^{(\mathbf{i}_{\ell-2},j)} = (U^{(\mathbf{i}_{\ell-2},j,k)} : k \in \{1, \ldots, b_{(\mathbf{i}_{\ell-2},j)}\})$ for $\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}$ and $j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}$. The number of nodes at level $\ell - 1$ (i.e., the size of $\boldsymbol{\Theta}$) is given by $d = \sum_{\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}} a_{\mathbf{i}_{\ell-2}}$, while the number of leaves of $\mathbf{U}$ is given by

$$n = \sum_{\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}} \sum_{j=1}^{a_{\mathbf{i}_{\ell-2}}} b_{(\mathbf{i}_{\ell-2},j)}.$$

We now proceed by presenting the compound construction derived from the tree structure. For $\mathbf{i}_r \in \mathcal{I}_r$ and $r \in \{0, \ldots, \ell - 2\}$, the stochastic representation of $M^{\mathbf{i}_r}$ is given by the random sum $M^{\mathbf{i}_r} = \sum_{m=1}^{M^{\mathbf{i}_{r|r-1}}} N_m^{\mathbf{i}_r}$, where $\mathbf{N}^{\mathbf{i}_r} = \{N_m^{\mathbf{i}_r} : m \in \mathbb{N}_+\}$ is a sequence of independent and identically distributed (iid), discrete, and strictly positive random variables with $N_m^{\mathbf{i}_r} \sim N^{\mathbf{i}_r}$ for $m \in \mathbb{N}_+$. As a consequence, the Laplace–Stieltjes transform of $M^{\mathbf{i}_r}$ is defined recursively as follows, for $t \geq 0$ with $r \in \{0, \ldots, \ell - 2\}$:

$$\mathcal{L}_{M^{\mathbf{i}_r}}(t) = \mathcal{L}_{M^{\mathbf{i}_{r|r-1}}}\left[-\ln\left\{\mathcal{L}_{N^{\mathbf{i}_r}}(t)\right\}\right].$$

For $r \in \{1, \ldots, \ell - 2\}$, the multivariate Laplace–Stieltjes transform of $\mathbf{M}_r = (M^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r)$ is defined, for $\mathbf{t}_r = (t^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r)$, by

$$\mathcal{L}_{\mathbf{M}_r}(\mathbf{t}_r) = \mathrm{E}\left\{\prod_{\mathbf{i}_r \in \mathcal{I}_r} \exp(-t^{\mathbf{i}_r} M^{\mathbf{i}_r})\right\}$$

and is given recursively, for $r \in \{1, \ldots, \ell - 2\}$, as follows:

$$\mathcal{L}_{\mathbf{M}_r}(\mathbf{t}_r) = \mathcal{L}_{\mathbf{M}_{r-1}}\left[-\sum_{k=1}^{a_{\mathbf{i}_{r|r-1}}} \ln\{\mathcal{L}_{N^{(\mathbf{i}_{r|r-1},k)}}(t^{(\mathbf{i}_{r|r-1},k)})\} : \mathbf{i}_{r|r-1} \in \mathcal{I}_{r-1}\right].$$

For a parent node at level $\ell - 1$, $(\mathbf{i}_{\ell-2}, j)$, with $\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}$ and $j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}$, the stochastic representation of $\Theta^{(\mathbf{i}_{\ell-2}, j)}$ is given by

$$\Theta^{(\mathbf{i}_{\ell-2}, j)} = \sum_{m=1}^{M^{\mathbf{i}_{\ell-2}}} B_m^{(\mathbf{i}_{\ell-2}, j)},$$

where $\mathbf{B}^{(\mathbf{i}_{\ell-2}, j)} = \{B_m^{(\mathbf{i}_{\ell-2}, j)} : m \in \mathbb{N}_+\}$ is a sequence of iid and strictly positive random variables with $B_m^{(\mathbf{i}_{\ell-2}, j)} \sim B^{(\mathbf{i}_{\ell-2}, j)}$ for $m \in \mathbb{N}_+$. It follows that, for all $t \geq 0$,

$$\mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2}, j)}}(t) = \mathcal{L}_{M^{\mathbf{i}_{\ell-2}}}\left[-\ln\left\{\mathcal{L}_{B^{(\mathbf{i}_{\ell-2}, j)}}(t)\right\}\right].$$

The Laplace–Stieltjes transform of Θ is therefore given by

$$\mathcal{L}_\Theta\{(t^{(\mathbf{i}_{\ell-2}, j)} : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j = 1, \ldots, a_{\mathbf{i}_{\ell-2}})\} = \mathcal{L}_{\mathbf{M}_{\ell-2}}\left\{\left(-\sum_{j=1}^{a_{\mathbf{i}_{\ell-2}}} \ln\{\mathcal{L}_{B^{(\mathbf{i}_{\ell-2}, j)}}(t^{(\mathbf{i}_{\ell-2}, j)})\} : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}\right)\right\}.$$

The hierarchical Archimedean copula $C_{\mathbf{U}}$ of $\mathbf{U}$ associated to Θ is finally given by

$$C_{\mathbf{U}}(\mathbf{u}) = \mathcal{L}_\Theta\left\{\left(\sum_{k=1}^{b_{(\mathbf{i}_{\ell-2}, j)}} \mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2}, j)}}^{-1}(u^{(\mathbf{i}_{\ell-2}, j, k)}) : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\right)\right\}.$$

Example 1. In Fig. 1, we have that the number of possible paths at level 1 is 2, with the set of possibilities given by $\mathcal{I}_1 = \{(1), (2)\}$. Then, the first node at level 1 has $a_{(1)} = 2$ children, while the second node has $a_{(2)} = 3$ children. This example refers to the dependence structure associated to the copula

$$C_{\mathbf{U}}(\mathbf{u}) = C_{\mathbf{U}}(u^{(1,1,1)}, u^{(1,1,2)}, u^{(1,2,1)}, u^{(1,2,2)}, u^{(2,1,1)}, u^{(2,1,2)}, u^{(2,2,1)}, u^{(2,2,2)}, u^{(2,3,1)}, u^{(2,3,2)})$$

$$= \mathcal{L}_M\left\{-\sum_{i=1}^{2} \ln\left(\mathcal{L}_{N^{(i)}}\left[-\sum_{j=1}^{a_{(i)}} \mathcal{L}_{B^{(1,j)}}\left\{\sum_{k=1}^{2} \mathcal{L}_{\Theta^{(i,j)}}^{-1}(u^{(i,j,k)})\right\}\right]\right)\right\},$$

where

$$\mathcal{L}_{\Theta^{(i,j)}}^{-1}(u^{(i,j,k)}) = \mathcal{L}_{B^{(i,j)}}^{-1}\{\exp(-\mathcal{L}_{N^{(i,j)}}^{-1}[\exp\{-\mathcal{L}_M^{-1}(u^{(i,j,k)})\}])\},$$

for $u^{(i,j,k)} \in [0, 1]$, $i \in \{1, 2\}$, $j \in \{1, \ldots, a_{(i)}\}$ (with $a_{(1)} = 2$ and $a_{(2)} = 3$), and $k \in \{1, 2\}$.

In our construction, it is possible to consider "degenerate descendants" from a parent $M^{\mathbf{i}_r}$ at level $r \in \{0, \ldots, \ell - 2\}$ to its leaves, such that we are able to "attach" the leaves corresponding to the final descendants of $M^{\mathbf{i}_r}$ only to their parent $M^{\mathbf{i}_r}$. For this, we make a degenerated lineage from $M^{\mathbf{i}_r}$. We proceed in the following way: we impose that the number of children of $M^{\mathbf{i}_r}$ and its descendants is equal to 1, and that the distribution of $N^{\mathbf{i}_r}$ as well as the distribution of the descendants are degenerate, i.e., $a_{\mathbf{i}_r} = 1$, $N^{\mathbf{i}_r} = 1$, $a_{(\mathbf{i}_r, 1)} = 1$, $N^{(\mathbf{i}_r, 1)} = 1$, ..., $a_{(\mathbf{i}_r, 1_{\ell-r-2})} = 1$, $N^{(\mathbf{i}_r, 1, \ldots, 1)} = 1$, $B^{(\mathbf{i}_r, 1_{\ell-r-1})} = 1$, where $\mathbf{1}_m = (1, \ldots, 1) \in \mathbb{R}^d$, such that $\Theta^{(\mathbf{i}_r, 1_{\ell-r-1})} = M^{\mathbf{i}_r}$.

Remark 1. The introduction of degenerated lineage(s) slightly modifies the trees of [2] by restricting the children of type "Θ" of a degenerated lineage to be a discrete random variable.

We denote by $\mathcal{J}_r$ the set of paths $\mathbf{i}_r \in \mathcal{I}_r$ in which the distribution of $N^{\mathbf{i}_r}$ is degenerate at 1, i.e.,

$$\mathcal{J}_r = \{\mathbf{i}_r \in \mathcal{I}_r : N^{\mathbf{i}_r} = 1\}. \tag{2}$$

Note that by convention, $\mathcal{J}_0 = \varnothing$. The set $\mathcal{H}$ is defined as the set of paths $\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}$ in which the distribution of $B^{\mathbf{i}_{\ell-2}}$ is degenerate at 1, i.e.,

$$\mathcal{H} = \{(\mathbf{i}_{\ell-2}, j) : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}, B^{(\mathbf{i}_{\ell-2}, j)} = 1\}. \tag{3}$$
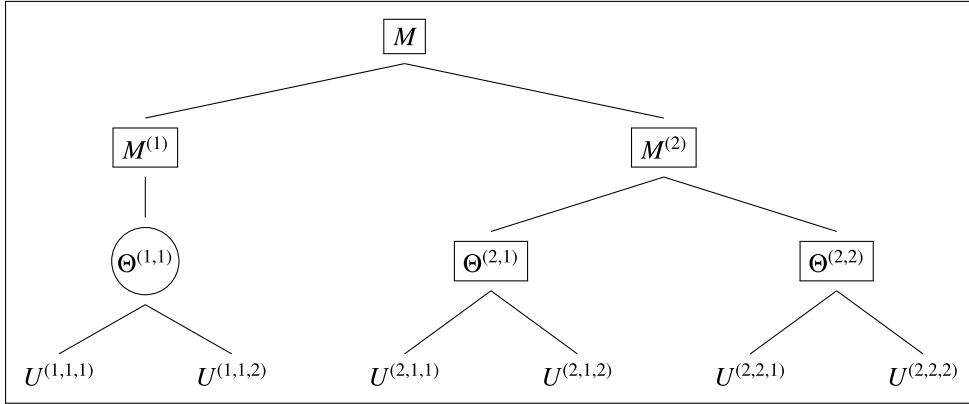
Fig. 2. Tree structure of Example 2, with $\ell = 3$ levels, $a = 2$, $a_{(1)} = 1$, $a_{(2)} = 2$, and $b_{\mathbf{i}_2} = 2$, for all $\mathbf{i}_2 \in \mathcal{I}_2$.
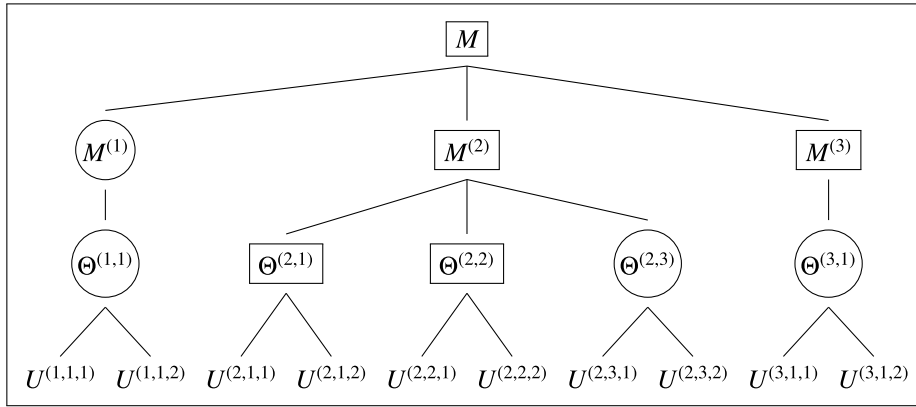

Fig. 3. Tree structure of Example 3, with $\ell = 3$ levels, $a = 3$, $a_{(1)} = 1$, $a_{(2)} = 3$, $a_{(3)} = 1$, and $b_{\mathbf{i}_2} = 2$, for all $\mathbf{i}_2 \in \mathcal{I}_2$.

To illustrate this notation, let us consider the two following additional examples, which will be used throughout the paper. We highlight the degeneracy of a node summands via the use of circles, i.e., circled nodes in all illustrations have degenerate summands.

Example 2. We consider the tree depicted in Fig. 2 in which a degenerated lineage from $M^{(1)}$ to $\Theta^{(1,1)}$ is assumed. This implies that $\Theta^{(1,1)} = M^{(1)}$ by fixing $\Pr(B^{(1,1)} = 1) = 1$. Hence, $\mathcal{H} = \{(1, 1)\}$ and $\mathcal{J}_1 = \emptyset$.

Example 3. We examine the tree shown in Fig. 3 in which degenerated lineages from M to $\Theta^{(1,1)}$, from $M^{(2)}$ to $\Theta^{(2,3)}$, and from $M^{(3)}$ to $\Theta^{(3,1)}$ are assumed. This implies that $\Theta^{(1,1)} = M^{(1)} = M$, $\Theta^{(2,3)} = M^{(2)}$, and $\Theta^{(3,1)} = M^{(3)}$ by fixing $\Pr(B^{(1,1)} = 1) = \Pr(B^{(2,3)} = 1) = \Pr(B^{(3,1)} = 1) = \Pr(N^{(1)} = 1) = 1$. Hence, $\mathcal{H} = \{(1, 1), (2, 3), (3, 1)\}$ and $\mathcal{J}_1 = \{(1)\}$.

2.2. Families of Archimedean copulas for sub-vectors

$C_{\mathbf{U}}$ is not an Archimedean copula, but it is noteworthy that copulas of sub-vectors of $\mathbf{U}$ are Archimedean. Two main classes of sub-vectors have components whose joint distributions are defined by an Archimedean copula. First, within a subgroup with dependence governed by $\Theta^{(\mathbf{i}_{\ell-2} j)}$, the multivariate cdf of any sub-vector of $\mathbf{U}^{(\mathbf{i}_{\ell-2} j)}$ is given by an Archimedean copula. Second, sub-vectors containing only a single component from each $\mathbf{U}^{(\mathbf{i}_{\ell-2} j)}$ under the same parent $M^{\mathbf{i}_r}$ are also linked by an Archimedean copula. The copulas for other sub-vectors are not Archimedean given that they are defined through the Laplace–Stieltjes transform of the random variables $\Theta^{(\mathbf{i}_{\ell-2} j)}$ and $M^{\mathbf{i}_r}$. The dependence structure for specific sub-vectors of $\mathbf{U}$ is discussed in [2]. More precisely, the expressions for the Archimedean copulas which define their multivariate cdf are given in Eqs. (23)–(27) of [2]. We define in more detail below these Archimedean copulas within our tree construction, since they will be used later in the composite likelihood estimation approach.
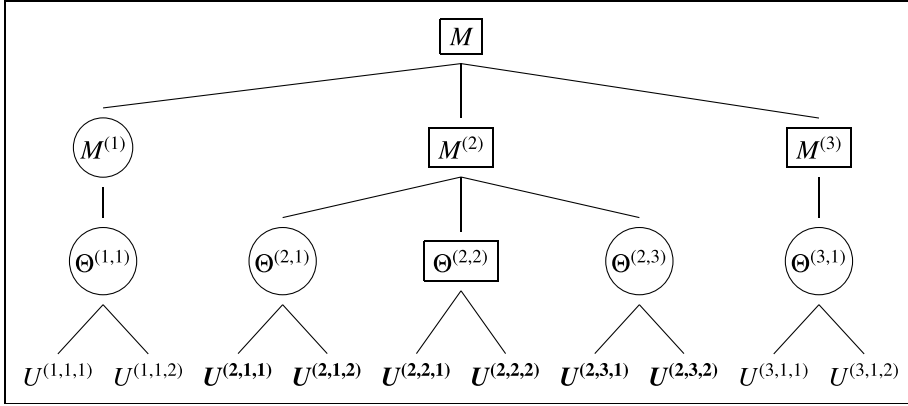
Fig. 4. In Example 3, the leaves associated to $\mathcal{D}_{\mathbf{i}_1}$ for $\mathbf{i}_1 = (2)$ are in bold.

Let us denote the set of paths to leaves which are the ultimate descendants of $M^{\mathbf{i}_r}$ (which we will call the descendant set of $M^{\mathbf{i}_r}$) by

$$
\mathcal{D}_{\mathbf{i}_r} = \left\{ (\mathbf{i}_r, i_{r+1}, \ldots, i_{\ell-2}, j, k) : \begin{array}{l} i_{r+1} \in \{1, \ldots, a_{\mathbf{i}_r}\}, \ldots, i_{\ell-2} \in \{1, \ldots, a_{(\mathbf{i}_r, i_{r+1}, \ldots, i_{l-3})}\}, \\ j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}, k \in \{1, \ldots, b_{(\mathbf{i}_{\ell-2}, j)}\} \end{array} \right\}.
$$

Let $\mathcal{S}_{\mathbf{i}_r}$ be the set of subsets of paths to leaves of $\mathcal{D}_{\mathbf{i}_r}$ of size $a_{\mathbf{i}_r}$ such that the children of $M^{\mathbf{i}_r}$ are different:

$$
\mathcal{S}_{\mathbf{i}_r} = \left\{ \{(\mathbf{i}_{\ell-2}^{(1)}, j^{(1)}, k^{(1)}), \ldots, (\mathbf{i}_{\ell-2}^{(a_{\mathbf{i}_r})}, j^{(a_{\mathbf{i}_r})}, k^{(a_{\mathbf{i}_r})})\} : \begin{array}{l} (\mathbf{i}_{\ell-2}^{(m)}, j^{(m)}, k^{(m)}) \in \mathcal{D}_{\mathbf{i}_r}, m \in \{1, \ldots, a_{\mathbf{i}_r}\}, \\ \mathbf{i}_{l-2|r+1}^{(1)} < \cdots < \mathbf{i}_{l-2|r+1}^{(a_{\mathbf{i}_r})} \end{array} \right\}.
$$

In the context of Example 3, the set $\mathcal{D}_{\mathbf{i}_1}$ has been illustrated by emphasizing its components (using a bold font) in a tree diagram; see Fig. 4. The same principle has been applied for the set $\mathcal{S}_{\mathbf{i}_1}$, where we used arrows to show admissible paths (i.e., admissible leaves) as an example; see Fig. 5.

For $\mathbf{u}^{(\mathbf{i}_{\ell-2}, j)} \in [0, 1]^{b_{(\mathbf{i}_{\ell-2}, j)}}$, let us define the Archimedean copula $C_{(\mathbf{i}_{\ell-2}, j)}$ of dimension $b_{(\mathbf{i}_{\ell-2}, j)}$ by

$$
C_{(\mathbf{i}_{\ell-2}, j)}(\mathbf{u}^{(\mathbf{i}_{\ell-2}, j)}) = \mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2}, j)}} \{ \mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2}, j)}}^{-1}(u^{(\mathbf{i}_{\ell-2}, j, 1)}) + \cdots + \mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2}, j)}}^{-1}(u^{(\mathbf{i}_{\ell-2}, j, b_{(\mathbf{i}_{\ell-2}, j)})}) \} \tag{4}
$$

which we call the copula associated to $\Theta^{(\mathbf{i}_{\ell-2}, j)}$.

For $\boldsymbol{\sigma} = (\sigma^{(m)} : m \in \{1, \ldots, a_{\mathbf{i}_r}\}) \in \mathcal{S}_{\mathbf{i}_r}$, where $\sigma^{(m)} = (\mathbf{i}_{\ell-2}^{(m)}, j^{(m)}, k^{(m)})$, let $\mathbf{U}^{\{\sigma\}} = (U^{\sigma^{(m)}} : m \in \{1, \ldots, a_{\mathbf{i}_r}\})$. For $\mathbf{u}^{\{\sigma\}} \in [0, 1]^{a_{\mathbf{i}_r}}$, let us define the Archimedean copula $C_{\mathbf{i}_r}$ of dimension $a_{\mathbf{i}_r}$ by

$$
C_{\mathbf{i}_r}(\mathbf{u}^{\{\sigma\}}) = \mathcal{L}_{M^{\mathbf{i}_r}} \{ \mathcal{L}_{M^{\mathbf{i}_r}}^{-1}(u^{\sigma^{(1)}}) + \cdots + \mathcal{L}_{M^{\mathbf{i}_r}}^{-1}(u^{\sigma^{(a_{\mathbf{i}_r})}}) \}, \tag{5}
$$

which we call the copula associated to $M^{\mathbf{i}_r}$.

Proposition 1 (*Families of Archimedean Copulas for Sub-vectors*).

   (i) *For $\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}$ with $j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}$, the joint cdf of $\mathbf{U}^{(\mathbf{i}_{\ell-2}, j)}$ is the Archimedean copula $C_{(\mathbf{i}_{\ell-2}, j)}$ given in* (4).
   (ii) *Furthermore, for $\boldsymbol{\sigma} \in \mathcal{S}_{\mathbf{i}_r}$ and $\mathbf{i}_r \in \mathcal{I}_r$, the joint cdf of $\mathbf{U}^{\{\sigma\}}$ is the Archimedean copula $C_{\mathbf{i}_r}$ given in* (5).

The proof of Proposition 1 can be found in Appendix A.

Example 1 cont'd. We consider the dependence structure associated to the tree depicted in Fig. 1. At level $r = 0$, the Archimedean copula defined with M is given by

$$
\begin{aligned}
&C_{\mathbf{i}_0}(u^{(1,1,k^{(1)})}, u^{(1,2,k^{(2)})}, u^{(2,1,\ell^{(1)})}, u^{(2,2,\ell^{(2)})}, u^{(2,3,\ell^{(3)})}) \\
&= \mathcal{L}_M \{ \mathcal{L}_M^{-1}(u^{(1,1,k^{(1)})}) + \mathcal{L}_M^{-1}(u^{(1,2,k^{(2)})}) + \mathcal{L}_M^{-1}(u^{(2,1,l^{(1)})}) + \mathcal{L}_M^{-1}(u^{(2,2,\ell^{(2)})}) + \mathcal{L}_M^{-1}(u^{(2,3,\ell^{(3)})}) \}
\end{aligned}
$$

for $u^{(1,1,k^{(1)})}, u^{(1,2,k^{(2)})}, u^{(2,1,\ell^{(1)})}, u^{(2,2,\ell^{(2)})}, u^{(2,3,\ell^{(3)})} \in [0, 1]$, $k^{(1)}, k^{(2)}, \ell^{(1)}, \ell^{(2)}, \ell^{(3)} \in \{1, 2\}$.

Then, at level $r = 1$, the two Archimedean copulas defined with $M^{(1)}$ and $M^{(2)}$ are respectively

$$
C_{(1)}(u^{(1,1,k^{(1)})}, u^{(1,2,k^{(2)})}) = \mathcal{L}_{M^{(1)}} \{ \mathcal{L}_{M^{(1)}}^{-1}(u^{(1,2,k^{(1)})}) + \mathcal{L}_{M^{(1)}}^{-1}(u^{(1,2,k^{(2)})}) \}
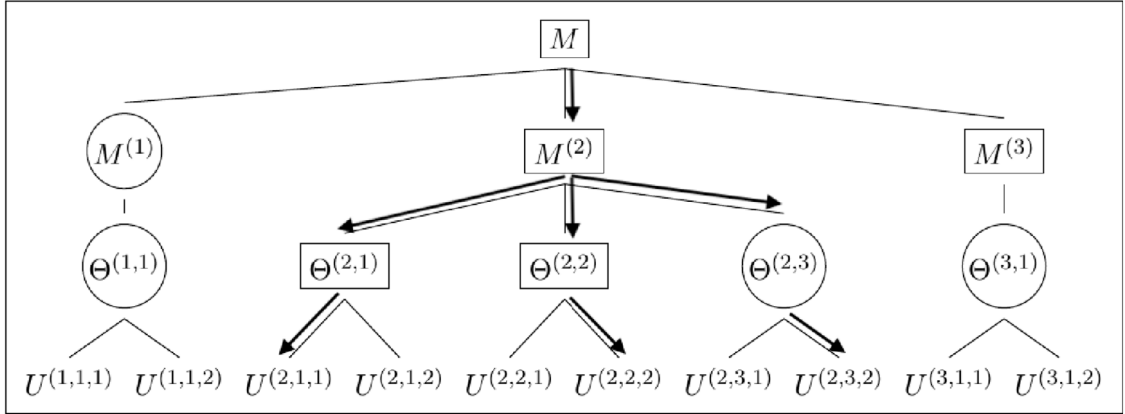$$

Fig. 5. In Example 3, a path associated to $\mathcal{S}_{\mathbf{i}_1}$ for $\mathbf{i}_1 = (2)$ is illustrated with descending arrows.

for $u^{\left(1,2,k^{(1)}\right)}, u^{\left(1,2,k^{(2)}\right)} \in [0, 1]$, $k^{(1)}, k^{(2)} \in \{1, 2\}$, and

$$C_{(2)}(u^{(2,1,\ell^{(1)})}, u^{(2,2,\ell^{(2)})}, u^{(2,3,\ell^{(3)})}) = \mathcal{L}_{M^{(2)}}\{\mathcal{L}_{M^{(2)}}^{-1}(u^{(2,1,\ell^{(1)})}) + \mathcal{L}_{M^{(2)}}^{-1}(u^{(2,2,\ell^{(2)})}) + \mathcal{L}_{M^{(2)}}^{-1}(u^{(2,3,\ell^{(3)})})\}$$

for $u^{(2,1,\ell^{(1)})}, u^{(2,2,\ell^{(2)})}, u^{(2,3,\ell^{(3)})} \in [0, 1]$, $\ell^{(1)}, \ell^{(2)}, \ell^{(3)} \in \{1, 2\}$.

Finally, at level $r = 2$, the five Archimedean copulas defined with $\Theta^{(i,j)}$ are

$$C_{(i,j)}(u^{(i,j,1)}, u^{(i,j,2)}) = \mathcal{L}_{\Theta^{(i,j)}}\left\{\sum_{k=1}^{2}\mathcal{L}_{\Theta^{(i,j)}}^{-1}(u^{(i,j,k)})\right\}$$

for $u^{(i,j,k)} \in [0, 1]$, $i \in \{1, 2\}$, $j \in \{1, \ldots, a_{(i)}\}$, $k \in \{1, 2\}$, $a_{(1)} = 2$, $a_{(2)} = 3$.

3. Determination of the theoretical tree structure through a Spearman's rho matrix

In this section, we explain how we build a tree structure compatible with our model presented in Section 2.1 through a Spearman's rho matrix.

3.1. Spearman's rho matrix and its associated distance matrix

Let (U, U') be a pair of $\mathcal{U}(0, 1)$ random variables with joint distribution C. The value of Spearman's rho for the random pair (U, U') is defined to be

$$\rho\left(U, U'\right) = -3 + 12\,\mathrm{E}\left(UU'\right) = -3 + 12\int_0^1\int_0^1 C(u, v)\,\mathrm{d}u\mathrm{d}v.$$

Let us also denote by $\rho(C_{(\mathbf{i}_{\ell-2},j)})$, resp. $\rho(C_{\mathbf{i}_r})$, Spearman's rho of the Archimedean copula associated to $C_{(\mathbf{i}_{\ell-2},j)}$, resp. $C_{\mathbf{i}_r}$ (i.e., Spearman's rho between any pair of components of $\mathbf{U}^{(\mathbf{i}_{\ell-2},j)}$, resp. of $\mathbf{U}^{\{\sigma\}}$ for $\sigma \in \mathcal{S}_{\mathbf{i}_r}$). The following proposition establishes that Spearman's rho between two leaves of our tree structure is given by the value of Spearman's rho of the Archimedean copula associated to their youngest common ancestor of type M.

Proposition 2. For $\iota, \iota' \in \{(\mathbf{i}_{\ell-2}, j) : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\}$, $k \in \{1, \ldots, b_\iota\}$ and $k' \in \{1, \ldots, b_{\iota'}\}$, the Spearman's rho between two leaves translates into

$$\rho(U^{(\iota,k)}, U^{(\iota',k')}) = \rho(C_{\iota_{|\max\{r:\iota^{(r)}=\iota'^{(r)}\}}}),$$

where $\iota^{(r)}$ is defined as the rth component of ι.

The proof of Proposition 2 is not given since it is a simple consequence of Proposition 1. For nested Archimedean copulas, this result directly follows from Corollary 5.1.4 of [24] or is stated as Lemma 4 in [10].

Example 2 cont'd. Consider the tree shown in Fig. 2. By fixing $\iota = (2, 1, 1)$ and $\iota' = (2, 1, 2)$, we see that $\iota_{|\max\{r:\iota^{(r)}=\iota'^{(r)}\}} = (2, 1)$. It follows that $\rho(U^{(2,1,1)}, U^{(2,1,2)}) = \rho(C_{(2,1)})$.

Example 3 **cont'd.** Consider now the tree shown in Fig. 3. By fixing $\iota = (2, 1, 1)$ and $\iota' = (2, 3, 1)$, we see that $\iota_{|\max\{r : \iota^{(r)} = \iota'^{(r)}\}} = (2)$. It follows that $\rho(U^{(2,1,1)}, U^{(2,3,1)}) = \rho(C_{(2)})$.

For our hierarchical Archimedean copulas model, the Spearman's rho values decrease as we descend the tree structure, and this will be the key to determining the structure.

Proposition 3. *For $\iota \in \{(\mathbf{i}_{\ell-2}, k) : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, k \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\}$, we have that, for all $0 \leq m' < m \leq \ell - 1$, $\rho(C_{\iota_{|m'}}) \leq \rho(C_{\iota_{|m}})$.*

The proof of Proposition 3 is given in Appendix B. Note that such a result is also valid for nested Archimedean copulas. It is discussed notably in Remark 2.3.2 of [16] and more recently generalized in Proposition 2 of [9] and Lemma 1 of [10].

We denote by

$$\mathcal{U} = \{(\mathbf{i}_{\ell-2}, j, k) \, ; \, \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}, k \in \{1, \ldots, b_{(\mathbf{i}_{\ell-2}, j)}\}\} \tag{6}$$

the set of indices of leaves. The Spearman's rho matrix of size $n \times n$ of **U** is defined by

$$K = (\rho(U^{(\mathbf{i}_{\ell-2}, j, k)}, U^{(\mathbf{i}'_{\ell-2}, j', k')}))_{(\mathbf{i}_{\ell-2}, j, k) \in \mathcal{U}, (\mathbf{i}'_{\ell-2}, j', k') \in \mathcal{U}}.$$

We introduce the following distance between leaves $U^{(\mathbf{i}_{\ell-2}, j, k)}$ and $U^{(\mathbf{i}'_{\ell-2}, j', k')}$

$$d\{U^{(\mathbf{i}_{\ell-2}, j, k)}, U^{(\mathbf{i}'_{\ell-2}, j', k')}\} = \max_{(\mathbf{i}''_{\ell-2}, j'', k'') \in \mathcal{U}} \left[|\rho\{U^{(\mathbf{i}_{\ell-2}, j, k)}, U^{(\mathbf{i}''_{\ell-2}, j'', k'')}\} - \rho\{U^{(\mathbf{i}'_{\ell-2}, j', k')}, U^{(\mathbf{i}''_{\ell-2}, j'', k'')}\}| \right].$$

Note that $d(U^{(\mathbf{i}_{\ell-2}, j, k)}, U^{(\mathbf{i}'_{\ell-2}, j', k')})$ is equal to the maximum distance between the columns $(\mathbf{i}_{\ell-2}, j, k)$ and $(\mathbf{i}'_{\ell-2}, j', k')$ of K. The following result is proved in Appendix C.

Proposition 4. *Let $\iota, \iota' \in \{(\mathbf{i}_{\ell-2}, j) : \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\}$. Then, for all $k \in \{1, \ldots, b_\iota\}$ and $k' \in \{1, \ldots, b_{\iota'}\}$, $d\{U^{(\iota, k)}, U^{(\iota', k')}\} = 1 - \rho\{U^{(\iota, k)}, U^{(\iota', k')}\}$.*

A distance matrix D is now introduced as

$$D = [d\{U^{(\mathbf{i}_{\ell-2}, j, k)}, U^{(\mathbf{i}'_{\ell-2}, j', k')}\}]_{(\mathbf{i}_{\ell-2}, j, k) \in \mathcal{U}, (\mathbf{i}'_{\ell-2}, j', k') \in \mathcal{U}}.$$

We deduce from Proposition 4 that $D = 1 - K$.

Let us now assume that we do not know the tree structure of the model but only the Spearman's rho matrix. We propose in the following sub-sections a procedure (a combination of two algorithms) that provides a tree which is compatible with the model presented in Section 2.1. Note that such a tree is not unique since it is possible to permute the branches/leaves of the tree. To guarantee uniqueness, it would be necessary to impose an identification rule.

3.2. Agglomerative clustering based on the distance matrix D

As mentioned in, e.g., [17,19], hierarchical clustering schemes refer to techniques aiming to partition objects into optimally homogeneous groups based on a chosen measure of similarity between those objects. We use one of these methods here as a first step in the determination of the tree structure behind our model using Spearman's rho to measure the similarity between the leaves of the tree. More specifically, we perform an average-linkage clustering based on a distance matrix D (but note that the single-linkage or the complete-linkage algorithm would also work).

This agglomerative hierarchical clustering method produces a hierarchy of clusters as follows. Initially, each of the n leaves constitutes a cluster of its own. The aim of the process is to have a single cluster. Thus, as long as there is more than one cluster, the closest pair of clusters, meaning the two clusters separated by the shortest distance, is joined into a single cluster. The average-linkage distance between two clusters of leaves is defined as the distance between two leaves, one from each cluster, which are the farthest apart from each other.

More precisely, the clusterings are assigned sequence numbers $0, \ldots, n-1$ and $L(m)$ is the level of the mth clustering. A cluster, denoted (c), is defined to be a set made of other disjoint clusters and/or leaves, and the proximity between clusters (c) and (s) is defined iteratively to be

$$d\{(c), (s)\} = \frac{1}{\#(c) \, \#(s)} \sum_{(c') \in (c)} \sum_{(s') \in (s)} d\left\{(s'), (c')\right\},$$

until both inputs of d are clusters consisting of a single leaf, in which case the distance function given in Proposition 4 is used. Moreover, the cardinality of a cluster (c), denoted $\#(c)$, represents the number of elements in cluster (c). In Algorithm 1, we provide the average-linkage clustering algorithm; see, e.g., [3,14] for more details.

Algorithm 1: Average-linkage clustering algorithm

> **Result**: A dendrogram
> Begin with the disjoint clustering having level $L(0) = 0$ and sequence number $m = 0$;
> **while** *All objects are not in one cluster* **do**
> > Fix $m = m + 1$;
> > Find the most similar pair of clusters in the current clustering, say pair (c), (s), according to $d\{(c), (s)\}$;
> > Merge clusters (c) and (s) into a single cluster, denoted $\{(c), (s)\}$, to form the next clustering m and set the level of this clustering to $L(m) = d\{(c), (s)\}$;
> > Update the proximity matrix D by deleting the rows and columns corresponding to clusters (c) and (s) and adding a row and column corresponding to the newly formed cluster (the proximity between the new cluster and old cluster (k) is defined as $d[(k), \{(c), (s)\}]$).
> **end**
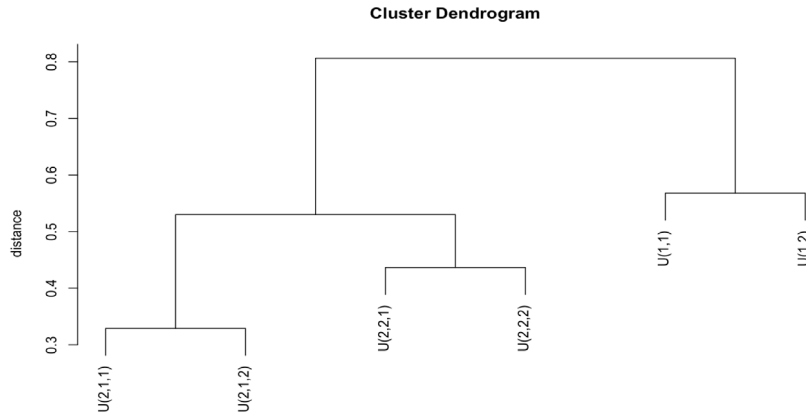
Fig. 6. Average-linkage clustering with the theoretical Spearman's rho matrix of Example 2.

Remark 2. Algorithm 1 does not necessarily provide a binary tree. If, for example, $L(m) = L(m-1)$, the algorithm collapses more than two clusters.

Remark 3. It is noteworthy that Propositions 3 and 4 imply that Algorithm 3 provides a one-to-one correspondence between its output dendrogram and the dependence structure of the model. However, the dendrogram is not compatible with our model specification as defined in Section 2.1; see the next section. Further note that this result is consistent with Theorem 2 in [10].

Let us now have a look at the dendrograms provided by the average-linkage clustering for the trees depicted in Figs. 2 and 3, as well as their corresponding trees.

Example 2 cont'd. We consider the dependence structure defined in Example 2 and illustrated in Fig. 2. Assume that Spearman's rho matrix is given by

$$
K = \begin{bmatrix}
. & 0.407 & 0.192 & 0.192 & 0.192 & 0.192 \\
. & . & 0.192 & 0.192 & 0.192 & 0.192 \\
. & . & . & 0.665 & 0.441 & 0.441 \\
. & . & . & . & 0.441 & 0.441 \\
. & . & . & . & . & 0.545 \\
. & . & . & . & . & .
\end{bmatrix}.
$$

Fig. 6 provides the dendrogram resulting from the average-linkage clustering algorithm. The corresponding tree $\mathcal{T}$ based on the average-linkage clustering of the theoretical matrix of Spearman's rho would be the one given in Fig. 7.

Fig. 7. Tree structure $\mathcal{T}$ derived from the average-linkage clustering for Example 2, where $a = a_{(1)} = a_{(2)} = 2$, and $b_{(2,1)} = b_{(2,2)} = 2$.

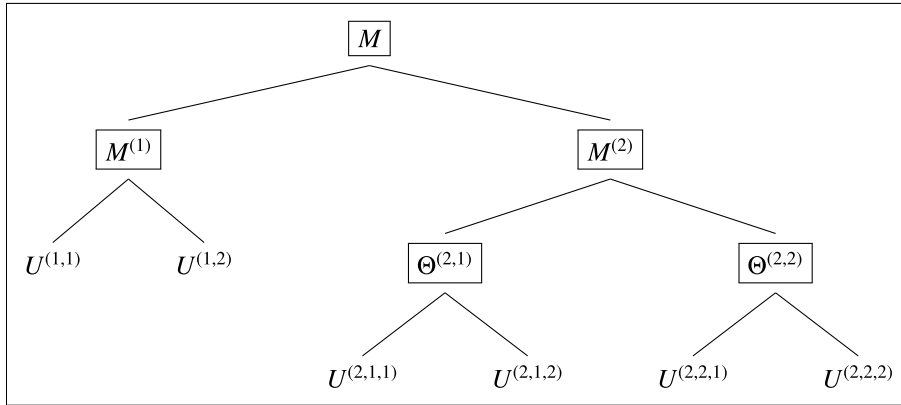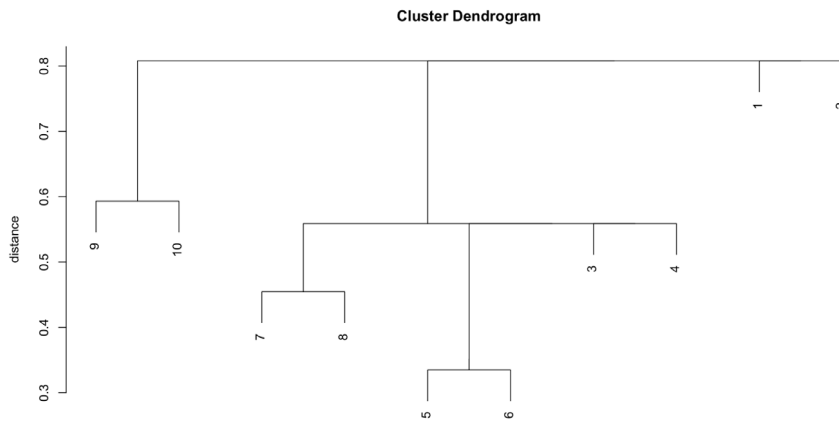**Fig. 8.** Average-linkage clustering with the theoretical Spearman's rho matrix of Example 3.

Example 3 cont'd. We consider the dependence structure defined as in Example 3 and illustrated in Fig. 3. Assume that Spearman's rho matrix is given by

$$
K = \begin{bmatrix}
. & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 \\
. & . & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 & 0.192 \\
. & . & . & 0.441 & 0.441 & 0.441 & 0.441 & 0.441 & 0.192 & 0.192 \\
. & . & . & . & 0.441 & 0.441 & 0.441 & 0.441 & 0.192 & 0.192 \\
. & . & . & . & . & 0.665 & 0.441 & 0.441 & 0.192 & 0.192 \\
. & . & . & . & . & . & 0.441 & 0.441 & 0.192 & 0.192 \\
. & . & . & . & . & . & . & 0.545 & 0.192 & 0.192 \\
. & . & . & . & . & . & . & . & 0.192 & 0.192 \\
. & . & . & . & . & . & . & . & . & 0.407 \\
. & . & . & . & . & . & . & . & . & .
\end{bmatrix}.
$$

Fig. 8 is the dendrogram resulting from the average-linkage clustering procedure. The corresponding tree $\mathcal{T}$ based on the average-linkage clustering of the theoretical matrix of Spearman's rho would be the one given in Fig. 9.

3.3. Reconstruction algorithm

Trees $\mathcal{T}$ derived from the dendrograms are not compatible with our model as defined in Section 2.1 because we need to have leaves at the same (and ultimate) level ℓ. The second step consists in providing trees $\mathcal{T}^*$ as in Fig. 2. For this, we propose a second algorithm. Algorithm 2 works recursively. First, the initial tree is taken as input. The algorithm then finds the elements of the initial cluster that are of dimension 1 (leaves). Those elements are then merged into one large cluster and put to the right of the tree, while creating the appropriate degenerate random variables. Then, the process runs again for all nodes below the main node that are not leaves, i.e., which have more than one element.
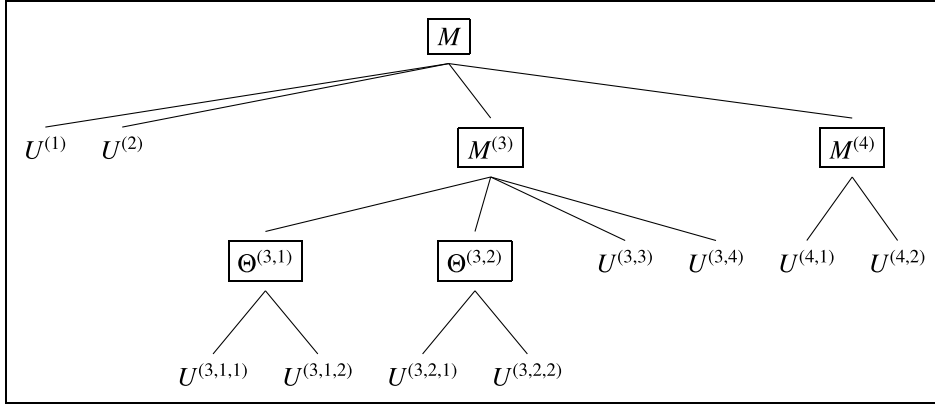
Fig. 9. Tree structure $\mathcal{T}$ derived from the average-linkage clustering for Example 3, where $a = 4$, $a_{(3)} = 4$, $a_{(4)} = 2$, and $b_{\mathbf{i}_2} = 2$, for all $\mathbf{i}_2 \in \{(3, 1), (3, 2)\}$.

Fig. 10. Tree $\mathcal{T}^*$ after reconstruction algorithm for Example 2, where $a = 2$, $a_{(1)} = 2$, $a_{(2)} = 1$, and $b_{\mathbf{i}_2} = 2$, for all $\mathbf{i}_2 \in \mathcal{I}_2$.

Remark 4. Note that an algorithm such as Algorithm 2 is not necessary in general. We need it in the present case to implement our composite likelihood approach for estimating the parameters of the model.

We let $(c^{\mathbf{v}})$ be a cluster with path $\mathbf{v}$. We denote the set of paths to the clusters of a given tree $\mathcal{T}$ by $\mathcal{A}_{\mathcal{T}}$. Then, for $\mathbf{v} \in \mathcal{A}_{\mathcal{T}}$, we can more specifically define $(c^{\mathbf{v}}) = \{(c^{(\mathbf{v},k)}) : (\mathbf{v}, k) \in \mathcal{A}_{\mathcal{T}}\}$, and finally let $\mathcal{T} = (c^{(0)})$. Moreover, we define the hierarchical level of a tree structure $\mathcal{T}$ by

$$\ell = \max\{\dim(\mathbf{v}) : \mathbf{v} \in \mathcal{A}_{\mathcal{T}}\} - 1,$$

and impose that $\ell > 0$; otherwise $C_{\mathbf{U}}$ would be an Archimedean copula.

One should note that the condition $p < \infty$ in the algorithm implies that there is at least one leaf below the node under consideration, since we use the convention that $\inf \emptyset = \infty$. Moreover we suppose that if $x = \emptyset$, then, $(x, y, z) = (y, z)$.

Let us now have a look at the trees provided by the reconstruction algorithm for the trees depicted in Figs. 7 and 9 . Observe that in both Figs. 10 and 11, all leaves are now at the same level with degenerate descendants (identified with circles) on the right side of tree.

Example 2 cont'd. After reconstruction, we obtain the tree structure given in Fig. 10.

Example 3 cont'd. After reconstruction, we obtain the tree structure given in Fig. 11.

4. Statistical procedure

In this section, we assume that we have n_0 observations from $\mathbf{U}$. The empirical Spearman's rho values are then computed for all pairs of $\mathbf{U}$, but they differ from the theoretical values and Algorithm 2 leads to a binary empirical

---

Algorithm 2: Reconstruction algorithm for the tree $\mathcal{T}$

Input:
$\mathcal{T}$: initial tree
Global variable:
$\mathcal{A} \leftarrow \emptyset$

Procedure Reconstruction($\mathcal{T}, \iota = \emptyset$)

  Let $(c_i) \in \mathcal{T}$ for $i \in \{1, \ldots, \#\mathcal{T} = \#(c)\}$, such that $\#(c_{\#\mathcal{T}}) \leq \cdots \leq \#(c_1)$, $p = \min\{j : \#(c_j) = 1\}$;

  if $p < \infty$ **then**

    Let $(c^{(\iota, p, \mathbf{1}_{l-\dim(\iota)-1}, i-p+1)}) = (c_i)$, for $i \in \{p, \ldots, \#\mathcal{T}\}$, and $(c^{(\iota, p, \mathbf{1}_{l-\dim(\iota)-1})}) = \bigcup_{i=p}^{\#\mathcal{T}}(c_i)$;

    Define $(c^{(\iota, p, \mathbf{1}_m)}) = \{(c^{(\iota, p, \mathbf{1}_{m+1})})\}$, for $m \in \{0, \ldots, \ell - \dim(\iota) - 2\}$;

    Define $N^{(\iota, p, \mathbf{1}_m)}$ such that $\Pr(N^{(\iota, p, \mathbf{1}_m)} = 1) = 1$, a.s., for $m \in \{0, \ldots, \ell - \dim(\iota) - 2\}$;

    Define $B^{(\iota, p, \mathbf{1}_{l-\dim(\iota)-1})}$ such that $\Pr(B^{(\iota, p, \mathbf{1}_{l-\dim(\iota)-1})} = 1) = 1$, a.s.;

    Fix

$$\mathcal{A} := \mathcal{A} \bigcup \{(\iota, p, \mathbf{1}_m) : m = 0, \ldots, \ell - \dim(\iota) - 1\} \bigcup$$
$$\{(\iota, p, \mathbf{1}_{l-\dim(\iota)-1}, k) : k = 1, \ldots, \#\mathcal{T} - p + 1\}$$

  else

    Define $p = \#\mathcal{T} + 1$;

  end

  for $k \leftarrow 1$ **to** $p - 1$ **by** 1 **do**

    Define $(c^{(\iota, k)}) = (c_k)$;

    Fix $\mathcal{A} = \mathcal{A} \bigcup \{(\iota, k)\}$;

    if $p > 1$ **then**

      Reconstruction($\mathcal{T} = (c_k), \iota = (\iota, k)$);

    end

  end

  if $\iota = \emptyset$ **then**

    return $\mathcal{A}$.

  end

$\mathcal{A}_{\mathcal{T}^*} = $ Reconstruction($\mathcal{T}$) ;
return $\mathcal{T}^* = \{(c^\iota) : \iota \in \mathcal{A}_{\mathcal{T}^*} \text{ such that } \dim(\iota) = 1\}$.
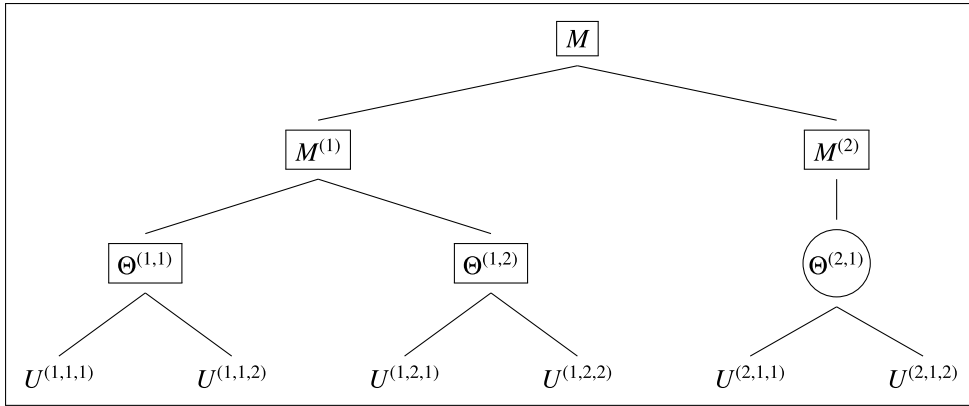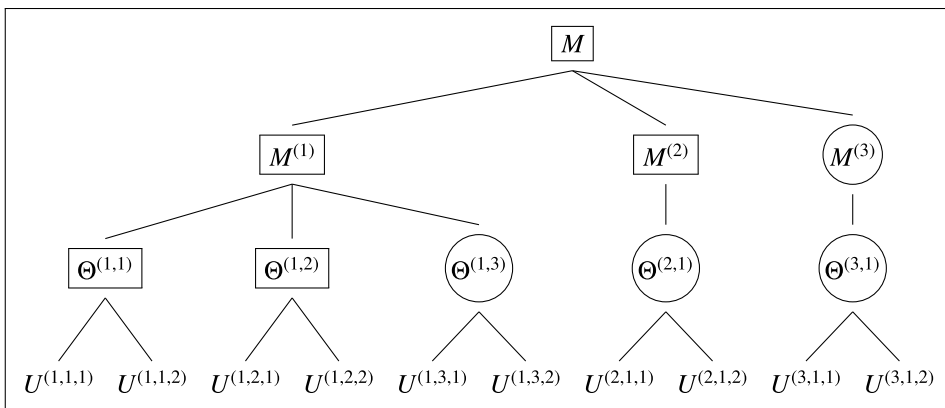

Fig. 11. Tree $\mathcal{T}^*$ after reconstruction algorithm for Example 3, where $a = 3$, $a_{(1)} = 3$, $a_{(2)} = 1$, $a_{(3)} = 1$, and $b_{\mathbf{i}_2} = 2$, for all $\mathbf{i}_2 \in \mathcal{I}_2$.

dendrogram that is different from the true one. The estimation of the parameter is unfeasible as long as the true structure of the dendrogram is not well identified. One can first estimate the structure nonparametrically as in [29] and then estimate the parameters, or do as in [27] and estimate everything together. But note that, in [27], the structure and

the parameters can be estimated all together because the generators of the hierarchical Archimedean copulas belong to the same parametric family.

Here, we propose a bootstrap procedure to decide whether it is necessary to eliminate nodes of the empirical structure or not, based on the statistical test of exchangeability T2 in Eq. (9) of [6]. To our knowledge, it is the first collapsing algorithm for multi-level hierarchical Archimedean copulas based on a statistical test. The only tuning parameter is the test level of confidence which gives an intuition as to the role of the parameter α. Note that this parameter should not be interpreted as the level of confidence for the whole tree. Then we present the hierarchical composite likelihood approach for parameter estimation.

4.1. Equi-rank hierarchical clustering

The average-linkage clustering Algorithm 1 provides a hierarchical binary dendrogram as soon as the empirical values of Spearman's rho are different from each other. The theoretical values of Spearman's rho are equal when leaves belong to the same Archimedean copula. To circumvent this problem, Gaißer and Schmid [6] provide pivotal statistics for testing equality of pair-wise rank correlations in a multivariate random vector. As a consequence, we are able to test the statistical significance of nodes of the dendrogram derived from Algorithm 1.

The goal of this section is to develop a collapsing algorithm (see Algorithm 3) for approximated trees resulting from the average-linkage algorithm that gathers nodes whose leaves can be assumed to have the same Spearman's rho. The main steps of Algorithm 3 are as follows. Algorithm 3 works recursively. Let us consider a cluster $(c^{\mathbf{i}_r})$ with $\mathbf{i}_r \in \mathcal{I}_r$. The goal is to choose whether there is a subcluster of $(c^{\mathbf{i}_r})$ that is not statistically significant. To do so, we consider a subcluster $(c^{(\mathbf{i}_r,j)})$, where $j \in \{1, \ldots, a_{\mathbf{i}_r}\}$. We test the hypothesis for Spearman's rho dependence measure between a leaf under the cluster $(c^{(\mathbf{i}_r,j)})$, i.e., $U^{(\mathbf{i}_r,j)}$ with $\mathbf{i}_r \in \mathcal{D}_{\mathbf{i}_r}$ and $j \in \{1, \ldots, a_{\mathbf{i}_r}\}$, and a leaf under the cluster $(c^{(\mathbf{i}_r,j')})$, i.e., $U^{(\mathbf{i}_r,j')}$ with $\mathbf{i}_r \in \mathcal{D}_{\mathbf{i}_r}$ and $j' \in \{1, \ldots, a_{\mathbf{i}_r}\}$.

The null hypothesis cannot be rejected if the test holds for all such combinations of leaves. If the hypothesis cannot be rejected (at confidence level α), it implies that the node with path $(\mathbf{i}_r, j)$ is not statistically significant, implying that we can eliminate and merge its elements directly into $(c^{\mathbf{i}_r})$. If it is the case, we repeat the process for the new cluster $(c^{(\mathbf{i}_r,j)})$. If the hypothesis is rejected for all subclusters of $(c^{\mathbf{i}_r})$ (i.e., we do the same process for all subclusters, and find that they are all statistically significant), then we go down in the hierarchy and run the algorithm again. The procedure stops when none of the clusters in the tree can be rejected, i.e., all clusters are statistically significant. This procedure has been implemented in the R package erhcv; see [5].

We provide an example of application of the Algorithm 3 for a vector $\mathbf{U}$ of size 10.

Example 4.   Consider the tree

$$\mathcal{B} = \{U_{(1)}, U_{(2)}, \{U_{(3,1)}, U_{(3,2)}, U_{(3,3)}, \{U_{(3,4,1)}, U_{(3,4,2)}\}\}, \{U_{(4,1)}, U_{(4,2)}, U_{(4,3)}\}\},$$

as illustrated in Fig. 12. For the purpose of this example, we let $M \sim \text{ShiftedGeo}(0.5)$, $N^{(1)} \sim \text{ShiftedGeo}(0.3, 1)$, $B^{(1,1)} \sim \mathcal{G}(0.6, 1)$ and $N^{(2)} \sim \text{ShiftedGeo}(0.2)$. These distributions are defined in Appendix E. We sample 1000 vectors of the tree structure $\mathcal{B}$. Fig. 13 shows the tree structure resulting from an average-linkage clustering, but before validation. When we apply the validation algorithm (Algorithm 3) with a 95% confidence level, the resulting tree structure (after validation) is the true one (i.e., the tree structure depicted in Fig. 12). For convenience, we have added a lower index to the leaves in the both figures. The leaf with a lower index i in Fig. 12 is the same as the leaf with a lower index i in Fig. 13.

4.2. Hierarchical composite likelihood approach for parameter estimation

We now develop a parametric estimation approach for our hierarchical Archimedean copula model and assume that the distributions of the summands in the compound sums are characterized by a scalar parameter (without loss of generality). The full likelihood is in general inaccessible for the whole vector $\mathbf{U}$, while it can be computed for some sub-vectors whose distributions are given by an Archimedean copula. Therefore it is natural to turn to the composite likelihood which is an inference function derived by multiplying a collection of component likelihoods. An overview of the composite likelihood estimation method is provided in [31]; see also, e.g., [32] for more details on this estimation method, as well as [4,20] for applications to copula modeling.

Here we propose a top-down composite likelihood approach : the parameters of the Archimedean copulas of the node are functions of the parameter of the node itself and the parameters of its parent. We therefore begin by estimating the parameters of the root, then her children (which may be also parents), etc. In this way, we estimate the parameters for copulas with smaller dimensions than the dimension of the complete copula $C_{\mathbf{U}}$ and we split the estimation problem into simpler sub-problems (based on Archimedean copulas for which likelihoods are known and may be efficiently computed).

Starting from the root of our tree, we denote by $\gamma_{\mathbf{i}_0}$ the parameter of M. As we go down to a level $r \in \{0, \ldots, \ell - 2\}$ of the tree, given a path $\mathbf{i}_r \in \mathcal{I}_r \backslash \mathcal{J}_r$ (see Eqs. (1) and (2)), we assume that $N^{\mathbf{i}_r} \sim F_{N^{\mathbf{i}_r}}$ with parameter $\gamma_{\mathbf{i}_r}$. Let us now define

---

Algorithm 3: Validation algorithm of a tree $\mathcal{T}$ with level ℓ

input

$\mathcal{T}$: tree;

α: confidence level;

Procedure Validation$((c), \nu)$

$\quad$ **if** $\#(c) > 1$ **then**

$\quad\quad$ Let $(s_1), \ldots, (s_{\#(c)}) \in (c)$ be the components of (c);

$\quad\quad$ **for** $k \leftarrow 1$ **to** $\#(c)$ **by** 1 **do**

$\quad\quad\quad$ **if** $\#(s_k) > 1$ **then**

$\quad\quad\quad\quad$ Test the null hypothesis (with confidence α)

$$\mathcal{H}_0 : \rho_1 = \rho_2 = \rho_3,$$

$\quad\quad\quad\quad$ where $\rho_1, \rho_2 \in \hat{\rho}_{(s_k),(s_k)}$, $\rho_3 \in \hat{\rho}_{(s_k),(s_{k'})}$, $k' \in \{1, \ldots, \#(c)\}\setminus\{k\}$, and $\hat{\rho}_{(s_k),(s_k')}$ is the set of all empirical

$\quad\quad\quad\quad$ Spearman's ρ between (s_k) and (s_k'). If $k = k'$, leaves need to be from different children of (s_k);

$\quad\quad\quad\quad$ **if** $\mathcal{H}_0$ *cannot be rejected* **then**

$\quad\quad\quad\quad\quad$ Eliminate the cluster (s_k) of (c), and add all components of (s_k) to (c);

$\quad\quad\quad\quad\quad$ Fix $k = 1$ and return to the beginning of the loop;

$\quad\quad\quad\quad$ **end**

$\quad\quad\quad$ **end**

$\quad\quad$ **end**

$\quad\quad$ **if** $\nu < \ell$ **then**

$\quad\quad\quad$ **for** $w \leftarrow 1$ **to** $\#(c)$ **by** 1 **do**

$\quad\quad\quad\quad$ Validation$((s_w), \nu + 1)$;

$\quad\quad\quad$ **end**

$\quad\quad$ **end**

$\quad$ **end**

Validation$(\mathcal{T}, 0)$;

return $\mathcal{T}$.
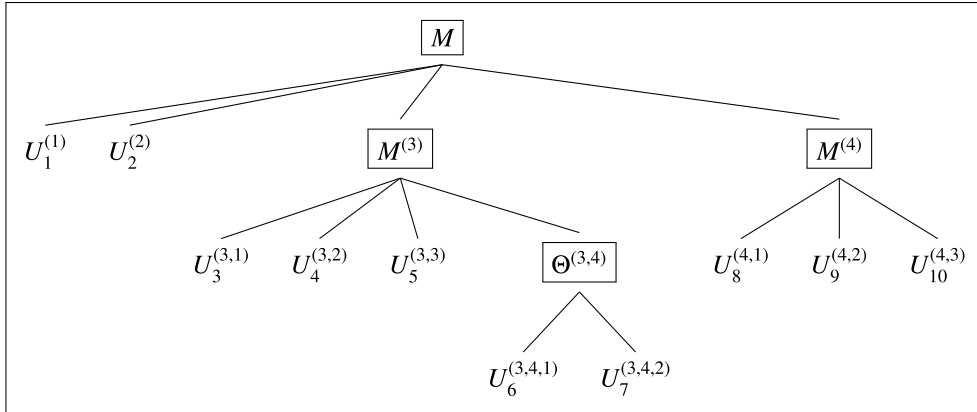
Fig. 12. Tree structure obtained via the clustering of the matrix of theoretical Spearman's rho of Example 4, where $a = 4$, $a_{(3)} = 4$, $a_{(4)} = 3$, and $b_{(3,4)} = 2$.

the vector of parameters $\boldsymbol{\gamma}_{\mathbf{i}_r} = (\gamma_{\mathbf{i}_0}, \gamma_{\mathbf{i}_{r|1}}, \ldots, \gamma_{\mathbf{i}_{r|r}})$ associated to the node $M^{\mathbf{i}_r}$. It is noteworthy that $\mathbf{U}^{\{\sigma\}}$, for $\sigma \in \mathcal{S}_{\mathbf{i}_r}$, has

for multivariate cdf the Archimedean copula

$$C_{\mathbf{i}_r}(\mathbf{u}^{\{\sigma\}}; \boldsymbol{\gamma}_{\mathbf{i}_r}) = \mathcal{L}_{M^{\mathbf{i}_r}} \left\{ \sum_{i=1}^{a_{\mathbf{i}_r}} \mathcal{L}_{M^{\mathbf{i}_r}}^{-1} (u^{\sigma^{(i)}}) \right\}.$$

Fig. 13. In Example 4, an illustration of the estimated tree structure before the validation algorithm: $a = a_{\mathbf{i}_r} = 2$, for all $\mathbf{i}_r \in \mathcal{I}_r$, $r = 1, 2$, and $b_{(1,1,1)} = b_{(2,1,1)} = b_{(2,1,2,1)} = 2$.

It is further assumed that for $(\mathbf{i}_{\ell-2}, j) \in \{\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2} : j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\} \setminus \mathcal{H}$ (see Eq. (3)), $B^{(\mathbf{i}_{\ell-2}, j)} \sim F_{B^{(\mathbf{i}_{\ell-2}, j)}}$ with a scalar parameter $\alpha_{(\mathbf{i}_{\ell-2}, j)}$. It follows that $\mathbf{U}^{(\mathbf{i}_{\ell-2}, j)}$ has for multivariate cdf the Archimedean copula

$$C_{(\mathbf{i}_{\ell-2}, j)}\{\mathbf{u}^{(\mathbf{i}_{\ell-2}, j)}; (\boldsymbol{\gamma}_{\mathbf{i}_{\ell-2}}, \alpha_{(\mathbf{i}_{\ell-2}, j)})\} = \mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2}, j)}} \left[\sum_{k=1}^{b_{(\mathbf{i}_{\ell-2}, j)}} \mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2}, j)}}^{-1}\{u^{(\mathbf{i}_{\ell-2}, j, k)}\} \right].$$

The composite log-likelihood associated to $M^{\mathbf{i}_r}$ to estimate $\boldsymbol{\gamma}_{\mathbf{i}_r}$ is then given by

$$c\ell_{\mathbf{i}_r}(\boldsymbol{\gamma}_{\mathbf{i}_r}; \mathbf{u}) = \sum_{\sigma \in \mathcal{S}_{\mathbf{i}_r}} \ln c_{\mathbf{i}_r}(\mathbf{u}^{\{\sigma\}}; \boldsymbol{\gamma}_{\mathbf{i}_r}),$$

where $c_{\mathbf{i}_r}$ is the density function of copula $C_{\mathbf{i}_r}$, while the log-likelihood associated to $\Theta^{(\mathbf{i}_{\ell-2}, j)}$ to estimate $\alpha_{(\mathbf{i}_{\ell-2}, j)}$ is given by

$$\ell_{(\mathbf{i}_{\ell-2}, j)}(\boldsymbol{\gamma}_{\mathbf{i}_{\ell-2}}, \alpha_{(\mathbf{i}_{\ell-2}, j)}; \mathbf{u}) = \ln c_{(\mathbf{i}_{\ell-2}, j)}(\mathbf{u}^{(\mathbf{i}_{\ell-2}, j)}; \boldsymbol{\gamma}_{\mathbf{i}_{\ell-2}}, \alpha_{(\mathbf{i}_{\ell-2}, j)}),$$

where $c_{(\mathbf{i}_{\ell-2}, j)}$ is the density of the copula $C_{(\mathbf{i}_{\ell-2}, j)}$.

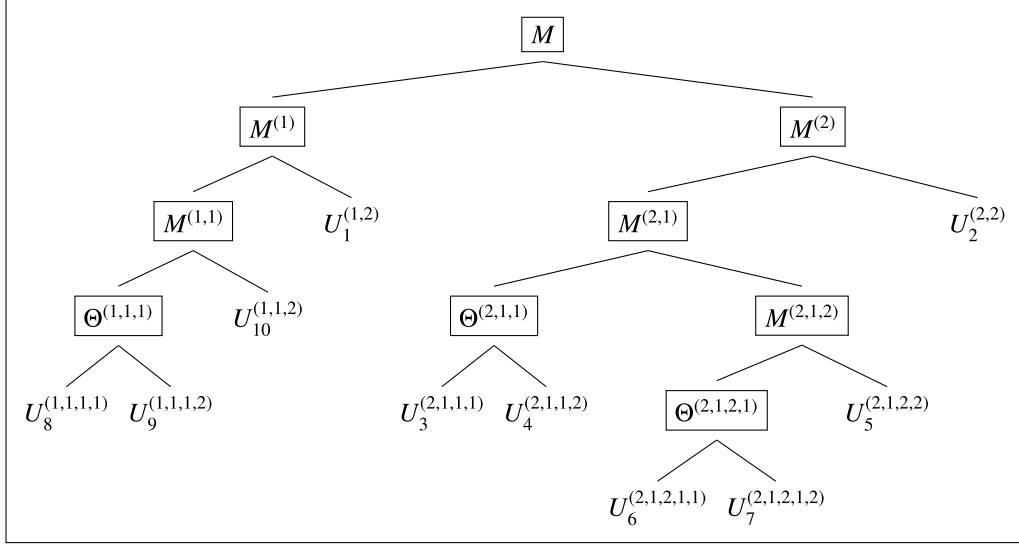For $\mathbf{i}_r \in \mathcal{I}_r \setminus \mathcal{J}_r$ with $r \in \{0, \ldots, \ell-2\}$, we define the score functions

$$g_{\mathbf{i}_r}(\boldsymbol{\gamma}_{\mathbf{i}_r}; \mathbf{u}) = \frac{\partial}{\partial \gamma_{\mathbf{i}_r}} c\ell_{\mathbf{i}_r}(\boldsymbol{\gamma}_{\mathbf{i}_r}; \mathbf{u}),$$

while for $(\mathbf{i}_{\ell-2}, j) \in \{\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\} \setminus \mathcal{H}$, the score functions are given by

$$h_{(\mathbf{i}_{\ell-2}, j)}(\boldsymbol{\gamma}_{\mathbf{i}_{\ell-2}}, \alpha_{(\mathbf{i}_{\ell-2}, j)}; \mathbf{u}) = \frac{\partial}{\partial \alpha_{(\mathbf{i}_{\ell-2}, j)}} \ell_{(\mathbf{i}_{\ell-2}, j)}(\boldsymbol{\gamma}_{\mathbf{i}_{\ell-2}}, \alpha_{(\mathbf{i}_{\ell-2}, j)}; \mathbf{u}).$$

As mentioned earlier, the estimation follows a top-down approach. We begin with $r = 0$ and we estimate the parameter of the Archimedean copula associated to $M = M^{\mathbf{i}_0}$. Then, we carry on the estimation of the other parameters of the Archimedean copulas associated to $M^{\mathbf{i}_r}$ at each level $r \in \{1, \ldots, \ell-2\}$, with the condition that the parameters of the parents of the given node have been estimated.

We assume that we observe n_0 iid sampled vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{n_0}$ from $\mathbf{U}$. The first order condition to determine $\hat{\boldsymbol{\gamma}}_{\mathbf{i}_0}$ is given by $\sum_{m=1}^{n_0} g_{\mathbf{i}_0}(\hat{\boldsymbol{\gamma}}_{\mathbf{i}_0}; \mathbf{u}_m) = 0$. Then, for each $r \in \{1, \ldots, \ell-2\}$ and $\mathbf{i}_r \in \mathcal{I}_r \setminus \mathcal{J}_r$, the first order condition to determine $\hat{\boldsymbol{\gamma}}_{\mathbf{i}_r} = \hat{\boldsymbol{\gamma}}_{\mathbf{i}_r|r}$ is given by

$$\sum_{m=1}^{n_0} g_{\mathbf{i}_r}(\hat{\boldsymbol{\gamma}}_{\mathbf{i}_r}; \mathbf{u}_m) = 0,$$
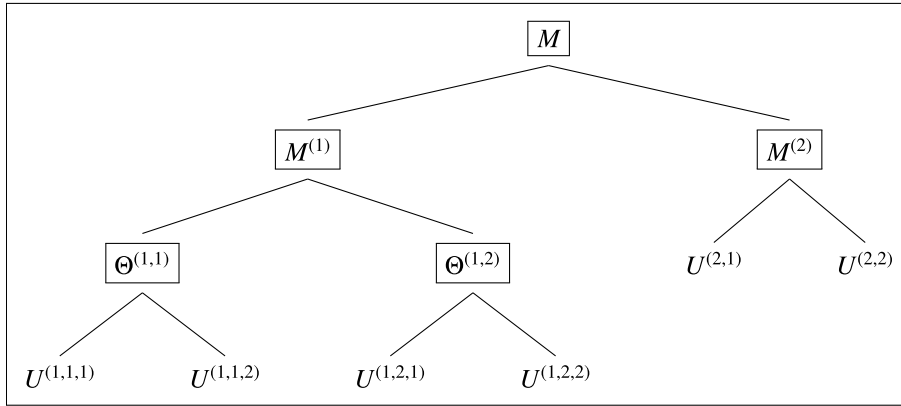
where $\hat{\boldsymbol{\gamma}}_{\mathbf{i}_r} = (\hat{\gamma}_{\mathbf{i}_0}, \gamma_{\mathbf{i}_{r|1}}, \ldots, \hat{\gamma}_{\mathbf{i}_{r|r}})$. Finally, for $(\mathbf{i}_{\ell-2}, j) \in \left\{ \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j = 1, \ldots, a_{\mathbf{i}_{\ell-2}} \right\} \setminus \mathcal{H}$, the first order condition to determine $\hat{\alpha}_{(\mathbf{i}_{\ell-2}, j)}$ is given by

$$\sum_{m=1}^{n_0} h_{(\mathbf{i}_{\ell-2}, j)}(\hat{\boldsymbol{\gamma}}_{\mathbf{i}_{\ell-2}}, \hat{\alpha}_{(\mathbf{i}_{\ell-2}, j)}; \mathbf{u}_m) = 0.$$

Assuming that the tree structure is the true one (which is almost surely the case as $n_0 \to \infty$ given that the test T2 proposed in [6] converges asymptotically), we can prove that the composite likelihood estimators are consistent and asymptotically Gaussian.

Proposition 5. *For each $r \in \{0, \ldots, \ell - 2\}$ and $\mathbf{i}_r \in \mathcal{I}_r \setminus \mathcal{J}_r$, we have, as $n_0 \to \infty$,*

$$\sqrt{n_0} \left(\hat{\gamma}_{\mathbf{i}_r} - \gamma_{\mathbf{i}_r} \right) \rightsquigarrow \mathcal{N}(0, \sigma_{\mathbf{i}_r}^2)$$

and for $(\mathbf{i}_{\ell-2}, j) \in \left\{ \mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}, j \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\} \right\} / \mathcal{H}$, we have as $n_0 \to \infty$,

$$\sqrt{n_0} \left(\hat{\alpha}_{(\mathbf{i}_{\ell-2}, j)} - \alpha_{(\mathbf{i}_{\ell-2}, j)} \right) \rightsquigarrow \mathcal{N}(0, \sigma_{(\mathbf{i}_{\ell-2}, j),}^2),$$

where $\rightsquigarrow$ means convergence in distribution. The expressions for the asymptotic variances $\sigma_{\mathbf{i}_r}^2$ and $\sigma_{(\mathbf{i}_{\ell-2}, j)}^2$ are provided within the proof that is available in the Online Supplement.

5. Simulation study

5.1. Full estimation procedure for Examples 2 and 3

In this sub-section, we proceed to the determinations, validations and rearrangements of the tree structures with two datasets sampled from models of Examples 2 and 3. When the structures have been properly approximated, we estimate the parameters of the chosen distributions for every node of the trees, following the top-down approach described in the previous section.

Example 2 cont'd. The assumptions are $M \sim$ ShiftedGeo($\gamma = 0.5$), $N^{(1)} \sim$ ShiftedGeo($\gamma_{(1)} = 0.1$), $N^{(2)} \sim$ ShiftedGeo($\gamma_{(2)} = 0.2$), $B_{(1,1)} \sim \mathcal{G}(\alpha_{(1,1)} = 0.1, 1)$, $B_{(1,2)} \sim \mathcal{G}(\alpha_{(1,2)} = 0.3, 1)$, $B_{(2,1)} = 1$. The distributions are defined in Appendix E. In total, 1000 samples of size 6 are drawn from the structure. The structure obtained via the average-linkage algorithm is the true one for all samples; see Fig. 14. In addition, the validation algorithm, with a 95% confidence level, does not remove nodes of the tree, implying that the structure stays identical to the true one. We now need to rearrange it for the estimation procedure. Fig. 15 provides the tree structures after rearrangement. We observe that the rearrangement algorithm introduces a degenerate node such that $\mathcal{H} = \{(2, 1)\}$. Once the structure has been properly approximated and rearranged, we estimate the parameters using the top-down composite likelihood method with both observations and pseudo-observations. In keeping with the literature on nonparametric inference for copula models, see, e.g., [8], we refer to observations as realizations of standard uniform distributions and to pseudo-observations as the normalized ranks. The results are given in Table 1, including confidence intervals and standard deviations. Standard deviations of the estimators are rather similar in both cases.
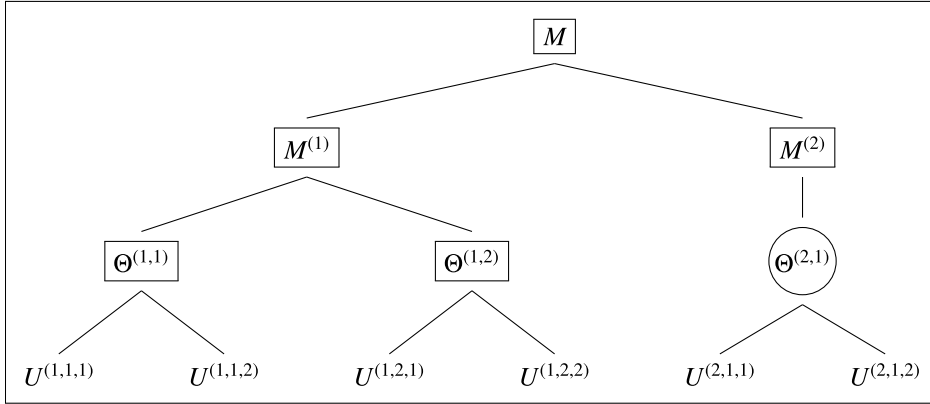
Fig. 15. In Example 2, validated, approximated tree structure after rearrangement: $a = a_{(1)} = 2$, $a_{(2)} = 1$, and $b_{(1,1)} = b_{(1,2)} = b_{(2,1)} = 2$.

Table 1
Estimated parameters of the approximated structure of Example 2 with empirical confidence intervals and standard deviations. Confidence intervals and means have been computed over 1000 simulations.

Parameters	Observations				Pseudo-observations			
	$q_{0.25}$	Mean	$q_{0.75}$	Std	$q_{0.25}$	Mean	$q_{0.75}$	Std
γ	0.421	0.490	0.566	0.045	0.416	0.485	0.562	0.045
$\gamma_{(1)}$	0.076	0.132	0.195	0.036	0.073	0.124	0.182	0.035
$\alpha_{(1,1)}$	0.088	0.121	0.151	0.019	0.081	0.113	0.144	0.019
$\alpha_{(1,2)}$	0.243	0.323	0.395	0.047	0.247	0.322	0.402	0.048
$\gamma_{(2)}$	0.133	0.220	0.322	0.056	0.128	0.216	0.314	0.056

Table 2
Estimated parameters of the approximated structure of Example 3 with the empirical confidence intervals and the standard deviations. Confidence intervals and means have been computed over 1000 simulations.

Parameters	Observations				Pseudo-observations			
	$q_{0.25}$	Mean	$q_{0.75}$	Std	$q_{0.25}$	Mean	$q_{0.75}$	Std
γ	0.475	0.521	0.570	0.029	0.470	0.517	0.566	0.029
$\gamma_{(1)}$	0.066	0.092	0.121	0.017	0.071	0.100	0.129	0.018
$\alpha_{(1,1)}$	0.082	0.103	0.125	0.014	0.080	0.101	0.122	0.013
$\alpha_{(1,2)}$	0.220	0.284	0.355	0.041	0.235	0.299	0.366	0.039
$\gamma_{(2)}$	0.202	0.217	0.234	0.010	0.188	0.202	0.217	0.009

Example 3 cont'd.   The assumptions are that $M \sim$ ShiftedGeo($\gamma = 0.5$), $N^{(1)} \sim$ ShiftedGeo($\gamma_{(1)} = 0.1$), $B_{(1,1)} \sim \mathcal{G}(\alpha_{(1,1)} = 0.1, 1)$, $B_{(1,2)} \sim \mathcal{G}(\alpha_{(1,2)} = 0.3, 1)$, $B_{(1,3)} = 1$, $N^{(2)} \sim$ ShiftedGeo($\gamma_{(2)} = 0.2$), $B^{(2,1)} = 1$, $N^{(3)} = 1$, and $B^{(3,1)} = 1$. The characteristics of these distributions are provided in Appendix E. In all, 1000 samples of size 10 are drawn from the structure. Fig. 16 provides an example of the structure obtained via the average-linkage algorithm. We can see that it is no longer the true one for this sample. However, the validation algorithm, with a 95% confidence level, is able to retrieve it by eliminating all the unnecessary nodes; see Fig. 17. Actually the validation algorithm is able to find the true structure 82 times out of 100; see the discussion in Section 5.3. We now need to rearrange it for the estimation procedure. Fig. 18 provides the tree structure after the validation algorithm and after rearrangement. We observe that the rearrangement algorithm introduces degenerated nodes such that $\mathcal{J} = \{(3)\}$ and $\mathcal{H} = \{(1, 3), (2, 1), (3, 1)\}$. Once the structure has been properly approximated and rearranged, we estimate the parameters and provide additionally confidence intervals and standard deviations. It is important to remark that both Examples 2 and 3 share the same parameters. Actually their tree structures only differ with degenerate nodes. Compared to Example 2, we see, through the confidence intervals and standard deviations, that estimations of the parameters are more precise. This is easily explained by the fact that leaves that have been added to M and $M^{(1)}$ bring additional informations and modify the (composite) likelihoods (see Table 2).

5.2. Composite likelihood method vs. maximum likelihood method

In this sub-section, we briefly compare the results of the top-down composite likelihood method with the results of the maximum likelihood method for a specific example of hierarchical Archimedean copulas.
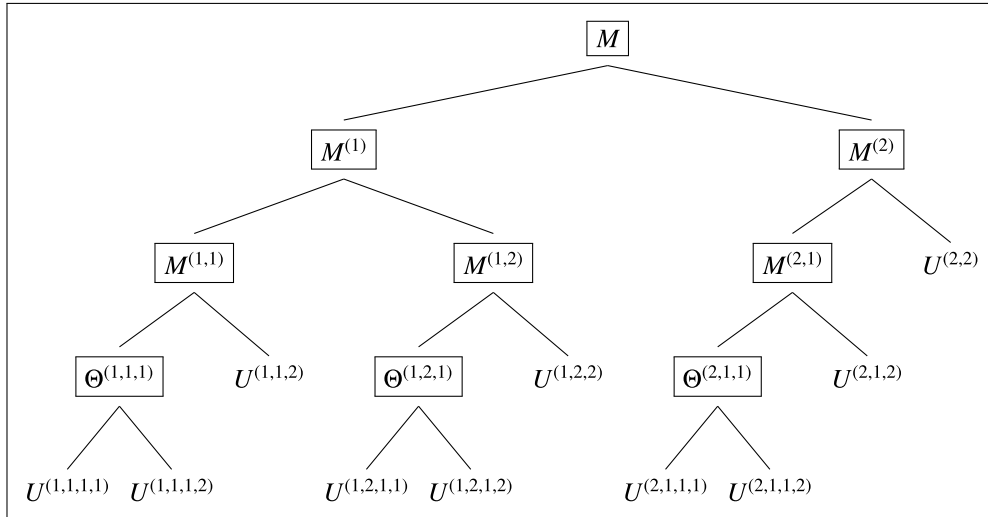
Fig. 16. In Example 3, approximated tree structure before the validation algorithm: $a = a_{\mathbf{i}_r} = 2$, for all $\mathbf{i}_r \in \mathcal{I}_r$, $r = 1, 2$, and $b_{(1,1,1)} = b_{(1,2,1)} = b_{(2,1,1)} = 2$.
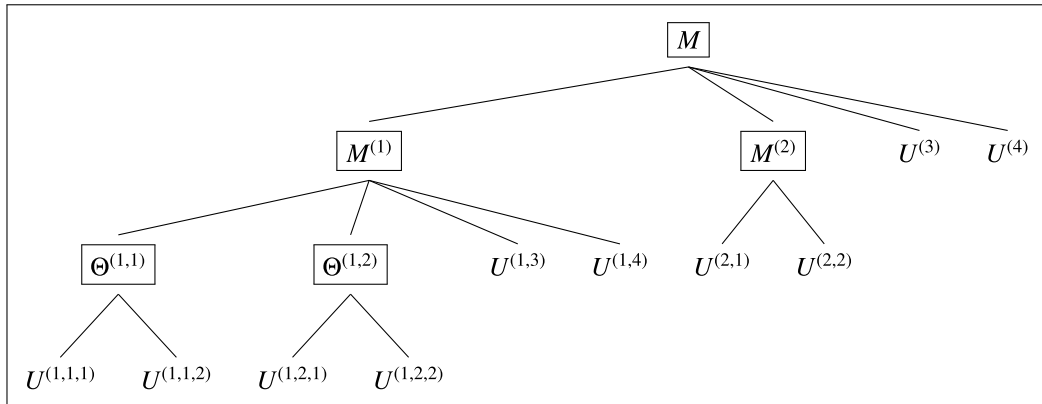
Fig. 17. In Example 3, approximated tree structure after the validation algorithm, but before rearrangement: $a = a_{(1)} = 4$, and $a_{(2)} = b_{(1,1)} = b_{(1,2)} = 2$.
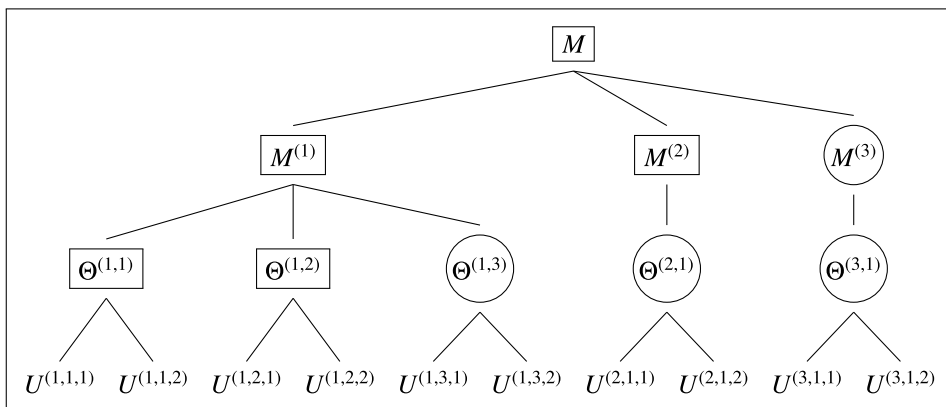
Fig. 18. In Example 3, approximated tree structure after the validation algorithm and after rearrangement: $a = a_{(1)} = 3$, $a_{(2)} = a_{(3)} = 1$ and $b_{\iota} = 2$ for all $\iota \in \{(1, 1), (1, 2), (1, 3), (2, 1), (3, 1)\}$.

Fig. 19. True dependence structure of Example 5 with $a = b_{(1)} = b_{(2)} = 2$.

Table 3
Results of the Example 5 from the composite likelihood method (left) and the maximum likelihood method (right). Means and standard deviations have been computed over 1000 simulations (CT = computational time in seconds).

$E(\widehat{\gamma})$	0.49478	$E(\widehat{\gamma})$	0.50128
$E(\widehat{\alpha}_{(1)})$	0.80339	$E(\widehat{\alpha}_{(1)})$	0.79787
$E(\widehat{\alpha}_{(2)})$	0.19834	$E(\widehat{\alpha}_{(2)})$	0.19976
$\sqrt{\text{var}(\widehat{\gamma})}$	0.05674	$\sqrt{\text{var}(\widehat{\gamma})}$	0.04785
$\sqrt{\text{var}(\widehat{\alpha}_{(1)})}$	0.05899	$\sqrt{\text{var}(\widehat{\alpha}_{(1)})}$	0.05450
$\sqrt{\text{var}(\widehat{\alpha}_{(2)})}$	0.01140	$\sqrt{\text{var}(\widehat{\alpha}_{(2)})}$	0.01000
CT	57.3700	CT	827.2790

Example 5. We consider the 4-dimensional hierarchical Archimedean copula $C_{\mathbf{U}}$ defined by the tree structure depicted in Fig. 19 and fix $M \sim \text{ShiftedGeo}(\gamma = 0.5)$ and $B^{(i)} \sim \mathcal{G}(\alpha_{(i)}, 1)$, where $(\alpha_{(1)}, \alpha_{(2)}) = (0.8, 0.2)$ and $i \in \{1, 2\}$. The properties of those distributions are given in Appendix E.

Let $\mathbf{U} = (U^{(1,1)}, U^{(1,2)}, U^{(2,1)}, U^{(2,2)})$ be the vector of four random variables whose joint cdf is given by the copula $C_{\mathbf{U}}$, where

$$C_{\mathbf{U}}(u^{(1,1)}, u^{(1,2)}, u^{(2,1)}, u^{(2,2)}) = \frac{\gamma \prod_{i=1}^{2} \left\{ \sum_{j=1}^{2} (\gamma/u^{(i,j)} + 1 - \gamma)^{1/\alpha_{(i)}} - 1 \right\}^{-\alpha_{(i)}}}{1 - (1 - \gamma) \prod_{i=1}^{2} \left\{ \sum_{j=1}^{2} (\gamma/u^{(i,j)} + 1 - \gamma)^{1/\alpha_{(i)}} - 1 \right\}^{-\alpha_{(i)}}},$$

for $(u^{(1,1)}, u^{(1,2)}, u^{(2,1)}, u^{(2,2)}) \in [0, 1]^4$. The top-down composite likelihood method and the maximum likelihood method are performed on 1000 samples of size 4 drawn from the random vector $\mathbf{U}$. The results of the top-down composite likelihood method a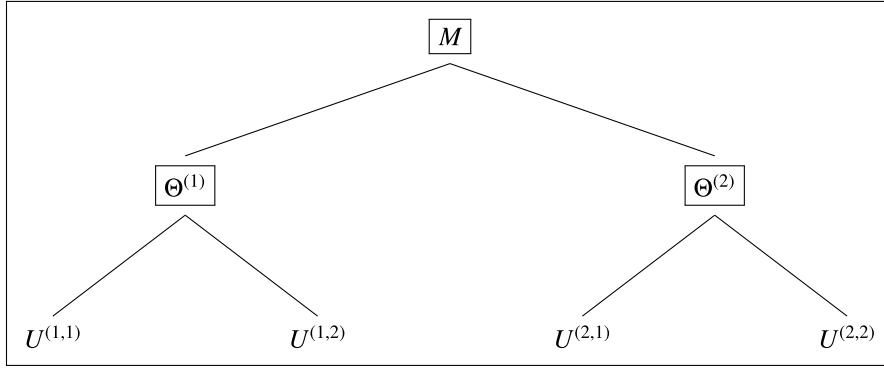nd those of the maximum likelihood method are provided in Table 3. The composite likelihood method offers an attracting and effective alternative to the full likelihood method for the estimation of the parameters of our hierarchical structures. We observe that the computational time (CT), in seconds, is considerably reduced while the variances of the estimators have not so much increased.

5.3. Discussion about the validation algorithm

The tree structure of Example 2 is always well identified by the agglomerative clustering algorithm, and the validation algorithm never proposes to change it. Therefore we decided to study the efficiency of the validation algorithm solely for Example 3.

In Fig. 20, we provide the barplot of the probabilities of validating the true tree structure (these probabilities have been computed over 1000 simulations). We study the influence of the confidence level, the sample size and the choice of the linkage function of the agglomerative clustering algorithm on these probabilities.

We can first observe that there is no significant differences between the average-linkage and the complete-linkage on the probabilities of selecting the true tree structure. Note however that the single, complete and average linkage were also experimentally tested in [9] for some nested hierarchical Archimedean copulas and that the average linkage led to better results in their simulation study. For these reasons, and as in [30], we decided to restrict ourselves to average linkage in this paper. Second, the probabilities of validating the true tree structure naturally increase with sample size. We also observe that these probabilities are very sensitive to the choice of the confidence level. As expected, the optimal choices depend on the sample size. For a size of 1000, the optimal level is 0.95 and leads to a validation algorithm that is able to
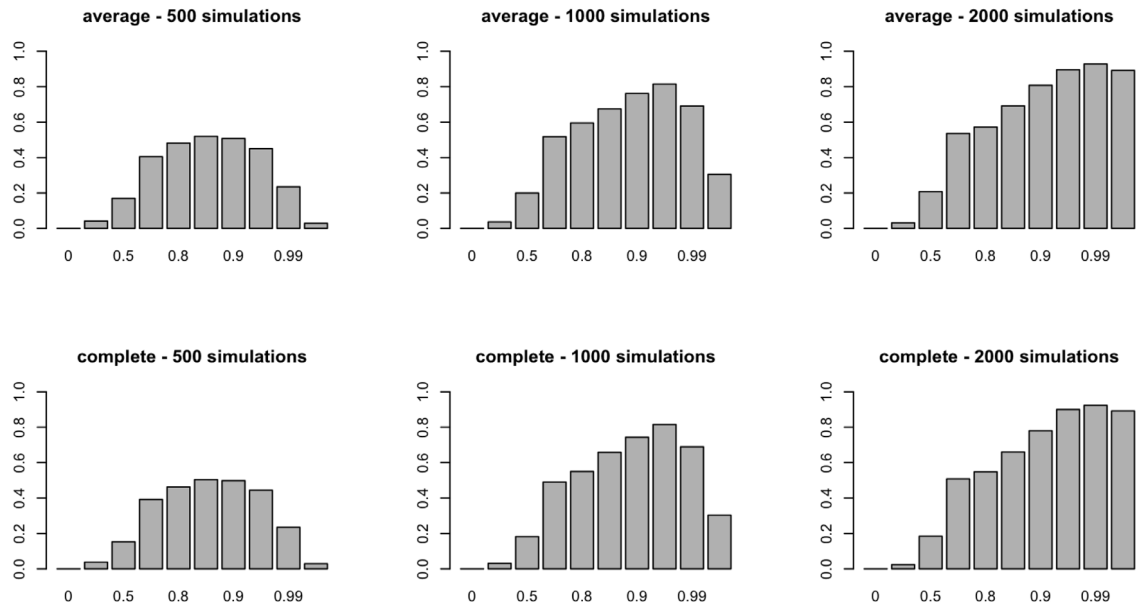
Fig. 20. Probabilities of validating the true tree structure of Example 3 with respect to the confidence level, the sample size and the choice of the linkage function for the agglomerative clustering algorithm.

choose the true structure 82 times out of 100. But, for a sample size of 2000, the optimal level is 0.99 and the validation algorithm is able to choose the true structure 93 times out of 100.

We also compared our procedure described in Algorithm 3 to the alternative procedure proposed by Górecki et al. [11] and implemented in HACopula. Within the context of Example 3, we found that 26.40%, 36.80% and 53.40% of the structures were well estimated, for 500, 1000 and 2000 simulations, respectively.

We observe that the method in HACopula leads to a probability of detecting the true tree than is lower or higher than what is obtained with Algorithm 3 depending on the choice of parameter α. In Example 3, this probability is lower than the results obtained with our procedure for a range of α of [0.75, 0.99]. Note that Górecki et al. [11] use an automated heuristic procedure for choosing an appropriate collapsed hierarchical Archimedean copula meaning that no parameter α is used. However, in our case, no matter how large the tree is, it boils down to the choice of a unique parameter, interpreted as the global severity of the exchangeability hypothesis testing algorithm.

6. Conclusion

In this paper, we have proposed an estimation procedure for the family of hierarchical Archimedean copulas introduced in [2] based on multivariate exponential mixture distributions built through compounding. Our estimation procedure aims to successively identify the tree structure and then estimate parameters of the copula by a top-down composite likelihood method. Every step of the proposed method is detailed and illustrated. Other estimation procedures for families of hierarchical Archimedean copulas have been proposed in the literature. However, the approach proposed in the present paper has the following advantages:

(a) The hierarchical Archimedean copula based on compound distributions leads to a proper copula with heterogeneous copula generators and without nesting conditions. The hierarchical structure allows to implement a top-down composite likelihood method where the parameters of the generators are estimated one after the other. Moreover, we are able to prove the consistency and the asymptotic normality of the estimators.

(b) The concordance measure (Spearman's rho) is used solely to identify the tree structure and not in the parameter estimation as it is done in some previous works. This leads to disconnect the tree structure identification from the estimation of the parameters of the generators. The composite likelihood method does not actually need Spearman's rho but rather appropriate assumptions on the compound distributions. It is noteworthy that different compound distributions can lead to the same Spearman's rho and it is therefore interesting to perform the estimation of the tree structure in the most independent way from the parameters of the compound distributions.

(c) The binary collapsing algorithm for the tree validation is based on a bootstrap procedure and on statistical hypothesis testings of exchangeability. The confidence level used for collapsing branches of the tree is based on asymptotic theoretical results contrarily to other algorithms which use arbitrary critical thresholds.

Acknowledgments

Appendix A. Proof of Proposition 1

The proof is divided in two parts.

Proof of (i). We fix one $\mathbf{i}_{\ell-2} \in \mathcal{I}_{\ell-2}$ and $j \in \left\{1, \ldots, a_{\mathbf{i}_{\ell-2}}\right\}$. We have

$$
F_{\mathbf{U}^{(\mathbf{i}_{\ell-2},j)}}(\mathbf{u}^{(\mathbf{i}_{\ell-2},j)}) = F_{\mathbf{U}}\{(\tilde{\mathbf{u}}^{(\mathbf{i}'_{\ell-2},j')}; \mathbf{i}'_{\ell-2} \in \mathcal{I}_{\ell-2}, j' \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\})\}
$$

$$
= \mathcal{L}_{\Theta}\left[\left\{\sum_{k=1}^{b_{(\mathbf{i}'_{\ell-2},j')}} \mathcal{L}^{-1}_{\Theta^{(\mathbf{i}'_{\ell-2},j')}}(\tilde{u}^{(\mathbf{i}'_{\ell-2},j',k)}); \mathbf{i}'_{\ell-2} \in \mathcal{I}_{\ell-2}, j' \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\right\}\right], \tag{A.1}
$$

where

$$
\tilde{\mathbf{u}}^{(\mathbf{i}'_{\ell-2},j')} = \begin{cases} \mathbf{u}^{(\mathbf{i}_{\ell-2},j)} & \text{if } (\mathbf{i}'_{\ell-2}, j') = (\mathbf{i}_{\ell-2}, j), \\ \mathbf{1}_{b_{(\mathbf{i}'_{\ell-2},j')}} & \text{if } (\mathbf{i}'_{\ell-2}, j') \neq (\mathbf{i}_{\ell-2}, j), \end{cases}
$$

and $\mathbf{1}_m = (1, \ldots, 1) \in \mathbb{R}^d$ with $m = b_{(\mathbf{i}'_{\ell-2},j')}$. Then, (A.1) becomes

$$
F_{\mathbf{U}^{(\mathbf{i}_{\ell-2},j)}}(\mathbf{u}^{(\mathbf{i}_{\ell-2},j)}) = \mathcal{L}_{\Theta}\left[\left\{\sum_{k=1}^{b_{(\mathbf{i}'_{\ell-2},j)}} \mathcal{L}^{-1}_{\Theta^{(\mathbf{i}'_{\ell-2},j')}}(\tilde{u}^{(\mathbf{i}'_{\ell-2},j',k)}) : \mathbf{i}'_{\ell-2} \in \mathcal{I}_{\ell-2}, j' \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}\right\}\right]
$$

$$
= \mathrm{E}\left[\prod_{\mathbf{i}'_{\ell-2} \in \mathcal{I}_{\ell-2}, j' \in \{1, \ldots, a_{\mathbf{i}_{\ell-2}}\}} e^{-\left\{\sum_{k=1}^{b_{(\mathbf{i}'_{\ell-2},j')}} \mathcal{L}^{-1}_{\Theta^{(\mathbf{i}'_{\ell-2},j')}}(\tilde{u}^{(\mathbf{i}'_{\ell-2},j',k)})\right\}\Theta^{(\mathbf{i}'_{\ell-2},j')}}\right]. \tag{A.2}
$$

Since $\mathcal{L}^{-1}_{\Theta^{(\mathbf{i}'_{\ell-2},j')}}(1) = 0$ for $(\mathbf{i}'_{\ell-2}, j') \neq (\mathbf{i}_{\ell-2}, j)$, Eq. (A.2) becomes

$$
C_{(\mathbf{i}_{\ell-2},j)}(\mathbf{u}^{(\mathbf{i}_{\ell-2},j)}) = \mathrm{E}\left[e^{-\left\{\sum_{k=1}^{b_{(\mathbf{i}_{\ell-2},j)}} \mathcal{L}^{-1}_{\Theta^{(\mathbf{i}_{\ell-2},j)}}(u^{(\mathbf{i}_{\ell-2},j,k)})\right\}\Theta^{(\mathbf{i}_{\ell-2},j)}}\right]
$$

$$
= \mathcal{L}_{\Theta^{(\mathbf{i}_{\ell-2},j)}}\left[\mathcal{L}^{-1}_{\Theta^{(\mathbf{i}_{\ell-2},j)}}\{u^{(\mathbf{i}_{\ell-2},j,1)}\} + \cdots + \mathcal{L}^{-1}_{\Theta^{(\mathbf{i}_{\ell-2},j)}}\{u^{(\mathbf{i}_{\ell-2},j,b_{(\mathbf{i}_{\ell-2},j)})}\}\right] = C_{(\mathbf{i}_{\ell-2},j)}\{\mathbf{u}^{(\mathbf{i}_{\ell-2},j)}\},
$$

which is the desired result. $\quad\square$

Proof of (ii). For a fixed $\sigma \in \mathcal{S}_{\mathbf{i}_r}$, the joint cdf of $\mathbf{U}^{\{\sigma\}}$ is denoted by

$$
F_{\mathbf{U}^{\{\sigma\}}}(\mathbf{u}^{\{\sigma\}}) = \mathrm{Pr}\left[\bigcap_{m=1}^{a_{\mathbf{i}_r}}\{U^{\sigma^{(m)}} \leq u^{\sigma^{(m)}}\}\right],
$$

for $\mathbf{u}^{\{\sigma\}} \in [0, 1]^{a_{\mathbf{i}_r}}$. We need to prove that $F_{\mathbf{U}^{\{\sigma\}}}(\mathbf{u}^{\{\sigma\}}) = C_{\mathbf{i}_r}(\mathbf{u}^{\{\sigma\}})$, for $\mathbf{u}^{\{\sigma\}} \in [0, 1]^{a_{\mathbf{i}_r}}$. For each path $\sigma^{(m)}$ with $m \in \{1, \ldots, a_{\mathbf{i}_r}\}$, given the compounding nature of the dependence structure, and given the definition of $\mathcal{S}_{\mathbf{i}_r}$, the expressions for $\mathcal{L}_{\Theta^{\sigma^{(m)}}}$ and $\mathcal{L}_{\Theta^{\{\sigma\}}}$ are given, for all $m \in \{1, \ldots, a_{\mathbf{i}_r}\}$, by

$$
\mathcal{L}_{\Theta^{\sigma^{(m)}}}(t^{\sigma^{(m)}}) = \mathcal{L}_{M^{\mathbf{i}_r}}[-\ln\{\mathcal{L}_{W^{\sigma^{(m)}}}(t^{\sigma^{(m)}})\}],
$$

and

$$
\mathcal{L}_{\Theta^{\{\sigma\}}}(t^{\{\sigma\}}) = \mathcal{L}_{M^{\mathbf{i}_r}}\left[-\sum_{m=1}^{a_{\mathbf{i}_r}}\ln\{\mathcal{L}_{W^{\sigma^{(m)}}}(t^{(\sigma^{(m)})})\}\right], \tag{A.3}
$$

where the random variable $W^{\sigma^{(m)}}$ is the result of the successive random sums of the random variables N and B along the path $\sigma^{(m)}$, where $m \in \{1, \ldots, a_{\mathbf{i}_r}\}$. We know that

$$F_{\mathbf{U}^{\{\sigma\}}}(\mathbf{u}^{\{\sigma\}}) = \mathcal{L}_{\Theta^{\{\sigma\}}}\left[\mathcal{L}_{\Theta^{\sigma^{(m)}}}^{-1}(u^{\sigma^{(m)}}) : m \in \{1, \ldots, a_{\mathbf{i}_r}\}\right] \tag{A.4}$$

where, from (A.3), we have

$$\mathcal{L}_{\Theta^{\sigma^{(m)}}}^{-1}\{u^{\sigma^{(m)}}\} = \mathcal{L}_{W^{\sigma^{(m)}}}^{-1}\left[\exp\{-\mathcal{L}_{M^{\mathbf{i}_r}}^{-1}(u^{\sigma^{(m)}})\}\right], \tag{A.5}$$

for all $m \in \{1, \ldots, a_{\mathbf{i}_r}\}$. Then, combining (A.1) and (A.5) into (A.4), it follows that, for $\mathbf{u}^{\{\sigma\}} \in [0,1]^{a_{\mathbf{i}_r}}$,

$$
\begin{aligned}
F_{\mathbf{U}^{\{\sigma\}}}(\mathbf{u}^{\{\sigma\}}) &= \mathcal{L}_{\Theta^{\{\sigma\}}}\left[\mathcal{L}_{W^{\sigma^{(m)}}}^{-1}(u^{\sigma^{(m)}}) : m \in \{1, \ldots, a_{\mathbf{i}_r}\}\right] \\
&= \mathcal{L}_{M^{\mathbf{i}_r}}\left[-\sum_{m=1}^{a_{\mathbf{i}_r}} \ln\left[\mathcal{L}_{W^{\sigma^{(m)}}}\left\{\mathcal{L}_{\Theta^{\sigma^{(m)}}}^{-1}(u^{\sigma^{(m)}})\right\}\right]\right] \\
&= \mathcal{L}_{M^{\mathbf{i}_r}}\left(-\sum_{m=1}^{a_{\mathbf{i}_r}} \ln\left[\mathcal{L}_{W^{\sigma^{(m)}}}\left[\mathcal{L}_{W^{\sigma^{(m)}}}^{-1}\left[\exp\{-\mathcal{L}_{M^{\mathbf{i}_r}}^{-1}(u^{\sigma^{(m)}})\}\right]\right]\right]\right) \\
&= \mathcal{L}_{M^{\mathbf{i}_r}}\left\{\sum_{m=1}^{a_{\mathbf{i}_r}} \mathcal{L}_{M^{\mathbf{i}_r}}^{-1}(u^{\sigma^{(m)}})\right\}.
\end{aligned}
$$

This completes the proof of Part (ii). $\quad\square$

Appendix B. Proof of Proposition 3

The concordance ordering between copulas is defined as follows.

Definition A. Let C_1 and C_2 be two d-dimensional copulas. We say that C_2 is more concordant than C_1, denoted $C_1 \prec_c C_2$, if for all $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]$, $C_1(\mathbf{u}) \leq C_2(\mathbf{u})$ and $\bar{C}_1(\mathbf{u}) \leq \bar{C}_2(\mathbf{u})$, where $\bar{C}_1$ and $\bar{C}_2$ are joint survival distribution functions associated to C_1 and C_2, respectively.

First, let $0 \leq m' < m < \ell - 1$. Since $\mathcal{L}_{M^{\iota|m}}(t) = \mathcal{L}_{M^{\iota|m-1}}[-\ln\{\mathcal{L}_{N^{\iota|m}}(t)\}]$, we have

$$\mathcal{L}_{M^{\iota|m-1}}^{-1} \circ \mathcal{L}_{M^{\iota|m}} = \mathcal{L}_{M^{\iota|m-1}}^{-1} \circ \mathcal{L}_{M^{\iota|m-1}}\left\{-\ln\left(\mathcal{L}_{N^{\iota|m}}\right)\right\} = -\ln\left\{\mathcal{L}_{N^{\iota|m}}(t)\right\}.$$

Therefore, if $[-\ln\{\mathcal{L}_{N^{\iota|m}}(t)\}]'$ is completely monotone, then the dependence within a subgroup is stronger than the outer dependence between subgroups, i.e., $C_{\iota_{|m-1}} \prec_c C_{\iota_{|m}}$ and hence $\rho(C_{\iota_{|m-1}}) \leq \rho(C_{\iota_{|m}})$. It is well known that $[-\ln\{\mathcal{L}_{N^{\iota|m}}(t)\}]'$ is completely monotone if and only if $\mathcal{L}_{N^{\iota|m}}^k$ is the Laplace–Stieltjes transform of a positive random variable, i.e., $\mathcal{L}_{N^{\iota|m}}^k \in \Psi_\infty$ for all $k \in \mathbb{N}$. It follows that

$$C_{\iota_{|m'}} \prec_c \cdots \prec_c C_{\iota_{|m}}.$$

Finally, we let $0 < m' < m = \ell - 2$. Since $\mathcal{L}_{\Theta^\iota}(t) = \mathcal{L}_{M^{\iota|m-1}}[-\ln\{\mathcal{L}_{B^\iota}(t)\}]$, it follows, in a similar fashion as above, that

$$\mathcal{L}_{M^{\iota|l-2}}^{-1} \circ \mathcal{L}_{\Theta^\iota} = \mathcal{L}_{M^{\iota|l-2}}^{-1} \circ \mathcal{L}_{M^{\iota|l-2}}\left\{-\ln\left(\mathcal{L}_{B^\iota}\right)\right\} = -\ln\left(\mathcal{L}_{B^\iota}\right).$$

Therefore, if $\{-\ln(\mathcal{L}_{B^\iota})\}'$ is completely monotone, then the dependence within a subgroup is stronger than the outer dependence between subgroups, i.e., $C_{\iota_{|l-2}} \prec_c C_\iota$ and hence $\rho(C_{\iota_{|l-2}}) \leq \rho(C_\iota)$; see [18]. It implies that

$$C_{\iota_{|m'}} \prec_c \cdots \prec_c C_\iota.$$

This completes the Proof of Proposition 3. $\quad\square$

Appendix C. Proof of Proposition 4

Let $\iota, \iota' \in \mathcal{U}$, where $\mathcal{U}$ is the set of indices of leaves defined in (6). Then, we let $\mathcal{R}(\iota, \iota') = \iota_{|\max\{r : \iota^{(r)} = \iota'^{(r)}\}}$ be the deepest path taken by both ι and ι'. Thus,

$$\max_{\iota'' \in \mathcal{U}}\{|\,\rho(U^\iota, U^{\iota''}) - \rho(U^{\iota'}, U^{\iota''})\,|\} = \max_{\iota'' \in \mathcal{U}}\left\{|\,\rho\left(C_{\mathcal{R}(\iota, \iota'')}\right) - \rho\left(C_{\mathcal{R}(\iota', \iota'')}\right)\,|\right\}. \tag{A.6}$$

Then, we define the subset $\mathcal{U}_{\iota'}(\iota) = \{\iota'' \in \mathcal{U} : \rho(C_{\mathcal{R}(\iota, \iota'')}) > \rho(C_{\mathcal{R}(\iota', \iota'')})\}$. The second term in (A.6) can be re-written as the maximum between two scenarios.

The first scenario is when $\rho\left(C_{\mathcal{R}(\iota, \iota'')}\right)$ is maximized and $\rho\left(C_{\mathcal{R}(\iota', \iota'')}\right)$ is minimized, for $\iota'' \in \mathcal{U}$. Thus, it would imply that $\iota'' \in \mathcal{U}_{\iota'}(\iota)$. Also, we note that for all $\iota'' \in \mathcal{U}_{\iota'}(\iota)$, $\rho(C_{\mathcal{R}(\iota', \iota'')})$ is the same. It follows that $\iota'' = \iota$, since $\rho\{C_{\mathcal{R}(\iota, \iota)}\} = 1 \geq \rho\{C_{\mathcal{R}(\iota, \iota'')}\}$, for $\iota'' \in \mathcal{U}_{\iota'}(\iota)$. The resulting equation for the distance is then given by $1 - \rho\left(C_{\mathcal{R}(\iota, \iota')}\right)$.

The second scenario is when $\rho\left(C_{\mathcal{R}(\iota, \iota'')}\right)$ is minimized and $\rho\left(C_{\mathcal{R}(\iota', \iota'')}\right)$ is maximized, for $\iota'' \in \mathcal{U}$. Thus, it would imply that $\iota'' \in \mathcal{U}_t(\iota')$. Also, we note that for all $\iota'' \in \mathcal{U}_t(\iota')$, $\rho(C_{\mathcal{R}(\iota, \iota'')})$ is the same. It follows that $\iota'' = \iota'$, since $\rho\left\{C_{\mathcal{R}(\iota', \iota')}\right\} = 1 \geq \rho\left\{C_{\mathcal{R}(\iota', \iota'')}\right\}$, for $\iota'' \in \mathcal{U}_t(\iota')$. The resulting equation for the distance is then given by $1 - \rho\left\{C_{\mathcal{R}(\iota, \iota')}\right\}$. Thus the proof is complete. $\square$

Appendix D. Proof of Proposition 5

The proof of Proposition 5 is provided in the Online Supplement. $\square$

Appendix E. Probability distributions

Shifted geometric distribution: Let the discrete random variable M follow a shifted geometric distribution with parameter $q \in (0, 1)$ and support $\mathbb{N}_+$, denoted $M \sim \text{ShiftedGeo}(q)$. The probability mass function, the expectation and the variance are $\Pr(M = k) = q(1-q)^{k-1}$ for all $k \in \mathbb{N}_+$, $\mathrm{E}(M) = 1/q$, and $\text{var}(M) = (1-q)/q^2$, respectively. The Laplace–Stieltjes transform and the probability generating function are $\mathcal{L}_M(t) = qe^{-t}/\left\{1 - (1-q)e^{-t}\right\}$ and $\mathcal{P}_M(t) = qt/\{1 - (1-q)t\}$, respectively.

Gamma distribution: Let the continuous random variable X follow a Gamma distribution with parameters $\alpha > 0$, $\beta > 0$, and support $\mathbb{R}_+$, denoted $X \sim \mathcal{G}(\alpha, \beta)$. The probability density function, the expectation and the variance are $h(x; \alpha, \beta) = \beta^\alpha x^{\alpha-1} e^{-\beta x}/\Gamma(\alpha)$ for all $x > 0$, $\mathrm{E}(X) = \alpha/\beta$, and $\text{var}(X) = \alpha/\beta^2$, respectively. The Laplace–Stieltjes transform is $\mathcal{L}_X(t) = \{\beta/(\beta + t)\}^\alpha$.

Appendix F. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2019.03.008.

References

[1] E.C. Brechmann, Hierarchical Kendall copulas: Properties and inference, Canad. J. Statist. 42 (2014) 78–108.
[2] H. Cossette, S.-P. Gadoury, E. Marceau, I. Mtalai, Hierarchical Archimedean copulas through multivariate compound distributions, Insurance Math. Econom. 76 (2017) 1–13.
[3] W.H. Day, H. Edelsbrunner, Efficient algorithms for agglomerative hierarchical clustering methods, J. Classification 1 (1984) 7–24.
[4] T.M. Erhardt, C. Czado, U. Schepsmeier, Spatial composite likelihood inference using local C-vines, J. Multivariate Anal. 138 (2015) 74–88.
[5] S.-P. Gadoury, erhcv: Equi-Rank Hierarchical Clustering Validation, R package version 012, 2018.
[6] S. Gaißer, F. Schmid, On testing equality of pairwise rank correlations in a multivariate random vector, J. Multivariate Anal. 101 (2010) 2598–2615.
[7] C. Genest, R.J. MacKay, Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données, Canad. J. Statist. 14 (1986) 145–159.
[8] C. Genest, J. Nešlehová, Copulas and copula models, in: A.H. El-Shaarawi, W.W. Piegorsch (Eds.), Encyclopedia of Environmetrics, second ed., Wiley, Chichester 2, 2012, pp. 541–553.
[9] J. Górecki, M. Hofert, M. Holeňa, An approach to structure determination and estimation of hierarchical Archimedean copulas and its application to Bayesian classification, J. Intell. Inf. Syst. 46 (2016) 21–59.
[10] J. Górecki, M. Hofert, M. Holeňa, Kendall's tau and agglomerative clustering for structure determination of hierarchical Archimedean copulas, Dep. Mod. 5 (2017) 75–87.
[11] J. Górecki, M. Hofert, M. Holeňa, On structure, family and parameter estimation of hierarchical Archimedean copulas, J. Stat. Comput. Simul. 87 (2017) 3261–3324.
[12] J. Górecki, M. Holeňa, Structure determination and estimation of hierarchical Archimedean copulas based on Kendall correlation matrix, in: International Workshop on New Frontiers in Mining Complex Patterns, Springer, 2014, pp. 132–147.
[13] J. Górecki, M. Holeň, An alternative approach to the structure determination of hierarchical Archimedean copulas, in: 31st International Conference on Mathematical Methods in Economics, pp. 201–206.
[14] T. Hastie, R.J. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed., Springer, New York, 2009.
[15] C. Hering, M. Hofert, J.-F. Mai, M. Scherer, Constructing hierarchical Archimedean copulas with Lévy subordinators, J. Multivariate Anal. 101 (2010) 1428–1433.
[16] M. Hofert, Sampling Archimedean copulas with applications to CDO pricing (Doctoral dissertation), Universität Ulm, Germany, 2010.
[17] G. James, D. Witten, T. Hastie, R.J. Tibshirani, An Introduction to Statistical Learning, With Applications in R, Springer, New York, 2013.
[18] H. Joe, Multivariate Models and Multivariate Dependence Concepts, CRC Press, Boca Raton, FL, 1997.
[19] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley-Interscience, New York, 1990.
[20] D. Lee, H. Joe, Multivariate extreme value copulas with factor and tree dependence structures, Extremes 21 (2018) 147–176.
[21] A.W. Marshall, I. Olkin, Families of multivariate distributions, J. Amer. Statist. Assoc. 83 (1988) 834–841.
[22] D. Matsypura, E. Neo, A. Prokhorov, Estimation of hierarchical Archimedean copulas as a shortest path problem, Econ. Lett. 149 (2016) 131–134.
[23] A.J. McNeil, J. Nešlehová, Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions, Ann. Statist. 37 (2009) 3059–3097.
[24] R.B. Nelsen, An Introduction To Copulas, second ed., Springer, New York, 2006.
[25] O. Okhrin, Y. Okhrin, W. Schmid, On the structure and estimation of hierarchical Archimedean copulas, J. Econometrics 173 (2013) 189–204.
[26] O. Okhrin, Y. Okhrin, W. Schmid, Properties of hierarchical Archimedean copulas, Stat. Risk Model. Appl. Finance Insur. 30 (2013) 21–54.
[27] O. Okhrin, A. Tetereva, The realized hierarchical Archimedean copula in risk modelling, Econometrics 5 (2017) 26.
[28] S. Perreault, T. Duchesne, J.G. Nešlehová, Detection of block-exchangeable structure in large-scale correlation matrices, J. Multivariate Anal. 169 (2019) 400–422.

[29] J. Segers, N. Uyttendaele, Nonparametric estimation of the tree structure of a nested Archimedean copula, Comput. Statist. Data Anal. 72 (2014) 190–204.

[30] N. Uyttendaele, On the estimation of nested Archimedean copulas: A theoretical and an experimental comparison, Comput. Statist. 33 (2018) 1047–1070.

[31] C. Varin, N.M. Reid, D. Firth, An overview of composite likelihood methods, Statist. Sinica 21 (2011) 5–42.

[32] X. Xu, N.M. Reid, On the robustness of maximum composite likelihood estimate, J. Statist. Plann. Inference 141 (2011) 3047–3054.

[33] W. Zhu, C.-W. Wang, K.S. Tan, Structure and estimation of Lévy subordinated hierarchical Archimedean copulas (LSHAC): Theory and empirical tests, J. Bank. Financ. 69 (2016) 20–36.