Insurance: Mathematics and Economics 76 (2017) 1–13

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Hierarchical Archimedean copulas through multivariate compound distributions

Hélène Cossette, Simon-Pierre Gadoury, Étienne Marceau *, Itre Mtalai

École d'actuariat, Université Laval, Quebec (Qc), Canada

ARTICLE INFO

Article history:
Received November 2016
Received in revised form June 2017
Accepted 6 June 2017
Available online 19 June 2017

Keywords:
Archimedean copulas
Mixing random variables
Compounding
Marshall–Olkin
Hierarchical structure

ABSTRACT

In this paper, we propose a new hierarchical Archimedean copula construction based on multivariate compound distributions. This new imbrication technique is derived via the construction of a multivariate exponential mixture distribution through compounding. The absence of nesting and marginal conditions, contrarily to the nested Archimedean copulas approach, leads to major advantages, such as a flexible range of possible combinations in the choice of distributions, the existence of explicit formulas for the distribution of the sum, and computational ease in high dimensions. A balance between flexibility and parsimony is targeted. After presenting the construction technique, properties of the proposed copulas are investigated and illustrative examples are given. A detailed comparison with other construction methodologies of hierarchical Archimedean copulas is provided. Risk aggregation under this newly proposed dependence structure is also examined.

1. Introduction

Copulas are now well known tools used for dependence modeling purposes in many research topics. A d-dimensional copula is a d-variate probability distribution function for which the marginals are uniformly distributed on $(0, 1)$, with $d \geq 2$. One important class of copulas is the Archimedean copula family, popular for its simple construction procedure and multivariate generalization. A d-dimensional copula C is said to be an *Archimedean copula* if

$$C(u_1, \ldots, u_d) = \psi\left(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\right), \quad \text{for } (u_1, \ldots, u_d) \in [0, 1]^d. \tag{1}$$

The continuous and strictly decreasing function ψ is called the generator of the copula, where $\psi : [0, \infty) \to [0, 1]$, $\psi(0) = 1$ and $\lim_{t\to\infty}\psi(t) = 0$. In the same manner, $\psi^{-1} : [0, 1] \to [0, \infty)$, for which $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$, where ψ^{-1} is the inverse of the generator ψ. The set of all such functions is denoted by Ψ_∞. In fact, from Kimberling's and Bernstein's theorems, see e.g. Kimberling (1974), Feller (1971), and Hofert (2010), representation (1) leads to a proper copula for all $d \geq 2$ if and only if ψ is the Laplace–Stieltjes Transform (LST) of a strictly positive random variable (rv) Θ with cumulative distribution function (cdf) F_Θ, where the LST of the rv Θ is given by

$$\mathcal{L}_\Theta(t) = \int_0^\infty e^{-tx}\, dF_\Theta(x) = E\left[e^{-t\Theta}\right]. \tag{2}$$

The use of multivariate Archimedean copulas in high dimension can be restrictive due to their exchangeability property. One can resort to another interesting class of copulas, namely vine copulas (see e.g. Bedford and Cooke (2002) and Joe (1997)). They are pair copula constructions allowing a cascade decomposition of a multivariate distribution into the product of bivariate copulas. Vine copulas force the use of $\frac{d(d-1)}{2}$ bivariate copulas which requires a high number of parameters when d, the dimension of the copula, increases.

Hierarchical Archimedean copulas provide an interesting alternative to allow asymmetries. The first approach to construct hierarchical Archimedean copulas was proposed by Joe (1997) who introduced the so-called nested Archimedean copulas in three and four dimensions. They are obtained by nesting into each other Archimedean copulas. They are able to capture different dependence relations between and within different groups of risks with a relatively small number of parameters (see e.g. Górecki et al., 2016a). They were further studied by e.g. McNeil (2008), Hofert (2012), and Hofert (2011), in a general setting. For an Archimedean hierarchical structure to be a proper copula, a given nesting condition must be verified. For example, a 3-dimensional fully nested Archimedean copula can be written as

$$\begin{aligned} C(u_1, u_2, u_3) &= C(u_1, C(u_2, u_3)) \\ &= \psi_0\Big(\psi_0^{-1}(u_1) + \psi_0^{-1} \circ \psi_1\big(\psi_1^{-1}(u_2) + \psi_1^{-1}(u_3)\big)\Big). \end{aligned}$$

For the hierarchical structure to be a proper copula, $\psi_0^{-1} \circ \psi_1$ must have completely monotone derivatives, where ψ_0 and ψ_1 are

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

http://dx.doi.org/10.1016/j.insmatheco.2017.06.001
0167-6687/

generators of the parent and the child copulas respectively. If generators ψ_0 and ψ_1 belong to the same family, the verification of this sufficient nesting condition can be done without much problem, see Hofert (2010) for restrictions on the parameters. However, if generators ψ_0 and ψ_1 of different Archimedean families are combined, the sufficient condition may not hold for any choice of parameters. Also, sampling from these copulas can be difficult.

To circumvent these constraints, other approaches were proposed to build hierarchical Archimedean copulas. Hering et al. (2010) suggest to build hierarchical Archimedean copulas based on Lévy subordinators. Also, Brechmann (2014) used the Kendall distribution function, the multivariate analog to the probability integral transform for univariate random variables, as a tool to construct hierarchical Archimedean copulas.

In this paper, our objective is to propose an alternative approach to build hierarchical Archimedean copulas via multivariate compound distributions. The absence of nesting and marginal conditions leads to major advantages, such as a flexible range of possible combinations in the choice of distributions, the existence of explicit formulas for the distribution of the sum, and computational ease in high dimensions. The approach given here has similarities with the one of Hering et al. (2010). Both approaches derive hierarchical Archimedean copulas from the joint survival function of a multivariate mixed exponential distribution. Also, sampling algorithms are obtained in a similar fashion. The probabilistic arguments at the basis of these two construction strategies however differ. Here, random sums are key elements compared to Lévy processes in Hering et al. (2010). In our opinion, the proposed approach uses simpler mathematical tools to obtain hierarchical Archimedean copulas. We provide in Section 4.2 a detailed comparison of both constructions.

The following sections are organized as follows: in Section 2, we provide the steps for the construction of a multivariate copula with the use of multivariate compound distributions, followed by sampling algorithms, accompanied by relevant notations and properties. Section 3 proposes a representation of the copula by common mixtures. In Section 4, we compare the proposed approach with nested Archimedean copulas and Hering et al. (2010)'s construction method and provide a brief discussion on estimation procedures. Finally, Section 5 is devoted to risk aggregation.

2. Construction of hierarchical Archimedean copulas with multivariate compound distributions

As explained in Marshall and Olkin (1988), multivariate distributions can easily be constructed through the use of exponential mixtures. In this section, we propose an alternative method based on a probabilistic argument framework using multivariate compound distributions to build hierarchical Archimedean copulas. Note that the copula obtained under the proposed approach is represented as a specific function of LSTs of rvs and their inverse. To obtain this representation, we have recourse to rvs with multivariate compound distributions. For this reason and given the context, we will refer either directly to LSTs or to the rvs associated to them. The key element of our construction is that the latter are defined with a vector of mixing rvs, denoted by $\underline{\Theta}$, which follows a multivariate compound distribution. We examine the properties of this new copula and provide simulation algorithms.

2.1. Multivariate compound distributions

Let $\underline{M} = (M_1, \dots, M_d)$ be a vector of discrete strictly positive rvs with joint multivariate probability generating function (pgf) given by

$$\mathcal{P}_{\underline{M}}(t_1, \dots, t_d) = E\left[t_1^{M_1} \dots t_d^{M_d}\right], \tag{3}$$

and marginal pgfs given by

$$\mathcal{P}_{M_i}(t) = E\left[t_i^{M_i}\right], (i = 1, \dots, d), \tag{4}$$

where $|t_1|, \dots, |t_d| \le 1$.

A vector of rvs $\underline{\Theta} = (\Theta_1, \dots, \Theta_d)$ is said to follow a multivariate compound distribution if each component Θ_i can be represented as a random sum i.e. $\Theta_i = \sum_{j=1}^{M_i} B_{i,j}$, where $i = 1, \dots, d$ for $d \ge 2$. For each i, the elements of the sequence $\underline{B_i} = \{B_{i,j}, j = 1, 2, \dots\}$ are assumed independent and identically distributed (iid) strictly positive rvs, while the sequences are independent from each other and from the vector of rvs $\underline{M}$. By convention, for each $i = 1, \dots, d$, $B_{i,j} \sim B_i$ with cdf F_{B_i} and LST $\mathcal{L}_{B_i}$ $(j = 1, 2, \dots)$.

Since the rv Θ_i is defined as a random sum, its LST is given by

$$\begin{aligned}\mathcal{L}_{\Theta_i}(t) &= E\left[e^{-t\Theta_i}\right] = E_{M_i}\left[E\left[e^{-t\Theta_i}|M_i\right]\right] = E_{M_i}\left[E\left[e^{-tB_i}\right]^{M_i}\right] \\ &= \mathcal{P}_{M_i}\left(\mathcal{L}_{B_i}(t)\right),\end{aligned} \tag{5}$$

for $i = 1, \dots, d$.

Similarly, the multivariate LST of the vector of mixing rvs $\underline{\Theta}$ is given by

$$\begin{aligned}\mathcal{L}_{\underline{\Theta}}(t_1, \dots, t_d) &= E\left[e^{-t_1\Theta_1} \dots e^{-t_d\Theta_d}\right] \\ &= E_{\underline{M}}\left[E\left[e^{-t_1\Theta_1} \dots e^{-t_d\Theta_d} \left|\underline{M}\right.\right]\right] \\ &= E_{\underline{M}}[(E[e^{-tB_1}]^{M_1}) \dots (E[e^{-tB_d}]^{M_d})].\end{aligned} \tag{6}$$

With (6) and (3), the expression for the multivariate LST of $\underline{\Theta}$ is given by

$$\mathcal{L}_{\underline{\Theta}}(t_1, \dots, t_d) = \mathcal{P}_{\underline{M}}\left(\mathcal{L}_{B_1}(t_1) \dots \mathcal{L}_{B_d}(t_d)\right). \tag{7}$$

In this paper, we consider the specific case of multivariate compound distribution for $\underline{\Theta}$ assuming that $M_i = M$ $(i = 1, \dots, d)$, where M is a strictly positive discrete rv with

$$\mathcal{P}_M(t) = E\left[t^M\right]. \tag{8}$$

In other words, the components of $\underline{M}$ are comonotonic and identically distributed (with $M_i \sim M, i = 1, \dots, d$). Then, it means that the vector of mixing rvs $\underline{\Theta}$ follows a multivariate compound distribution and its multivariate LST in (7) becomes

$$\begin{aligned}\mathcal{L}_{\underline{\Theta}}(t_1, \dots, t_d) &= E\left[e^{-t_1\Theta_1} \dots e^{-t_d\Theta_d}\right] \\ &= E_M\left[E\left[e^{-t_1\Theta_1} \dots e^{-t_d\Theta_d} \left|M\right.\right]\right] \\ &= E_M\left[E\left[e^{-t_1\Theta_1} \left|M\right.\right] \dots E\left[e^{-t_d\Theta_d} \left|M\right.\right]\right] \\ &= E_M\left[\left(E\left[e^{-t_1B_1}\right] \dots E\left[e^{-t_dB_d}\right]\right)^M\right].\end{aligned} \tag{9}$$

Combining (9) and (8) leads to the following expression for the multivariate LST of $\underline{\Theta}$:

$$\mathcal{L}_{\underline{\Theta}}(t_1, \dots, t_d) = \mathcal{P}_M\left(\mathcal{L}_{B_1}(t_1) \dots \mathcal{L}_{B_d}(t_d)\right). \tag{10}$$

Also, for each $i = 1, \dots, d$, (5) becomes

$$\mathcal{L}_{\Theta_i}(t) = \mathcal{P}_M\left(\mathcal{L}_{B_i}(t)\right). \tag{11}$$

2.2. Multivariate mixed exponential distributions defined with multivariate compound distributions

Let $\underline{Y} = (Y_{1,1}, \dots, Y_{1,n_1}, \dots, Y_{d,1}, \dots, Y_{d,n_d})$ be a vector of $n_1 + \dots + n_d$ rvs which can be more conveniently represented as $\underline{Y} = \left(\underline{Y_1}, \dots, \underline{Y_d}\right)$, where $\underline{Y_i} = \left(Y_{i,1}, \dots, Y_{1,n_i}\right)$ is the vector of n_i rvs for the subgroup i $(i = 1, 2, \dots, d)$. Given $\underline{\Theta} = \underline{\theta}$, where $\underline{\theta} = (\theta_1, \dots, \theta_d)$, it is assumed that

$$\begin{aligned}&(Y_{1,1} \mid \underline{\Theta} = \underline{\theta}), \dots, (Y_{1,n_1} \mid \underline{\Theta} = \underline{\theta}), \dots, \\ &\quad (Y_{d,1} \mid \underline{\Theta} = \underline{\theta}), \dots, (Y_{d,n_d} \mid \underline{\Theta} = \underline{\theta})\end{aligned}$$

are conditionally independent. The conditional distributions of the components of $\underline{Y_i}$ are only influenced by the component Θ_i of $\underline{\Theta}$, i.e. $(Y_{i,j} \mid \underline{\Theta} = \underline{\theta})$ is identically distributed as $(Y_{i,j} \mid \Theta_i = \theta_i)$, for $i = 1, \ldots, d$ and $j = 1, \ldots, n_i$. We assume that $(Y_{i,1} \mid \Theta_i = \theta_i), \ldots, (Y_{i,n_i} \mid \Theta_i = \theta_i)$ are exponentially distributed with parameter θ_i for $i = 1, \ldots, d$. The univariate distribution of $Y_{i,j}$ is therefore a mixed exponential distribution with survival function given by

$$\overline{F}_{Y_{i,j}}\left(y_{i,j}\right) = \int_0^\infty \overline{F}_{Y_{i,j}|\Theta_i=\theta_i}(y_{i,j})\,\mathrm{d}F_{\Theta_i}(\theta_i) = \int_0^\infty \mathrm{e}^{-y_{i,j}\theta_i}\,\mathrm{d}F_{\Theta_i}(\theta_i) = \mathcal{L}_{\Theta_i}\left(y_{i,j}\right), \tag{12}$$

for $i = 1, \ldots, d$ and $j = 1, \ldots, n_i$. The inverse of the survival function $\overline{F}_{Y_{i,j}}$ in (12) is given by

$$\overline{F}^{-1}_{Y_{i,j}}\left(u_{i,j}\right) = \mathcal{L}^{-1}_{\Theta_i}\left(u_{i,j}\right), \tag{13}$$

where

$$\mathcal{L}^{-1}_{\Theta_i}\left(u_{i,j}\right) = \mathcal{L}^{-1}_{B_i}\left(\mathcal{P}^{-1}_{M}\left(u_{i,j}\right)\right), \tag{14}$$

for $u_{i,j} \in [0, 1]$, $i = 1, \ldots, d$ and $j = 1, \ldots, n_i$. Since

$$\mathcal{P}_M(s) = \mathcal{L}_M(-\ln(s)) \tag{15}$$

and

$$\mathcal{P}^{-1}_M(u) = \exp\left(-\mathcal{L}^{-1}_M(u)\right). \tag{16}$$

We can rewrite $\mathcal{L}^{-1}_{\Theta_i}$ in terms of $\mathcal{L}^{-1}_M$ as follows:

$$\mathcal{L}^{-1}_{\Theta_i}\left(u_{i,j}\right) = \mathcal{L}^{-1}_{B_i}\left(\exp(-\mathcal{L}^{-1}_M\left(u_{i,j}\right))\right), \tag{17}$$

for $u_{i,j} \in [0, 1]$, $i = 1, \ldots, d$ and $j = 1, \ldots, n_i$.

The vector of rvs $\underline{Y_i}$ follows a multivariate mixed exponential distribution, which is defined only in terms of LST of the mixing rv Θ_i given in (11) and for which the multivariate survival function corresponds to

$$\begin{aligned}\overline{F}_{\underline{Y_i}}(\underline{y_i}) &= \int_0^\infty \overline{F}_{\underline{Y_i}|\Theta_i=\theta_i}(\underline{y_i})\,\mathrm{d}F_{\Theta_i}(\theta_i) \\ &= \int_0^\infty \prod_{j=1}^{n_i} \overline{F}_{Y_{i,j}|\Theta_i=\theta_i}(y_{i,j})\,\mathrm{d}F_{\Theta_i}(\theta_i) \\ &= \int_0^\infty \prod_{j=1}^{n_i} \mathrm{e}^{-y_{i,j}\theta_i}\,\mathrm{d}F_{\Theta_i}(\theta_i) \\ &= \mathcal{L}_{\Theta_i}(y_{i,1} + \cdots + y_{i,n_i}),\end{aligned} \tag{18}$$

where $\underline{y_i} = (y_{i,1}, \ldots, y_{i,n_i})$, for $i = 1, \ldots, d$.

Finally, the vector of rvs $\underline{Y}$ follows a multivariate mixed exponential distribution. Most importantly for us, the latter has the interesting feature of being defined by the vector of mixing rvs $\underline{\Theta}$, which follows a multivariate compound distribution. The multivariate survival function of $\underline{Y}$ is represented in terms of the LST (10) of $\underline{\Theta}$ i.e.

$$\overline{F}_{\underline{Y}}\left(\underline{y}\right) = \mathcal{L}_{\underline{\Theta}}\left(y_{1,1} + \cdots + y_{1,n_1}, \ldots, y_{d,1} + \cdots + y_{d,n_d}\right), \tag{19}$$

where $\underline{y} = \left(\underline{y_1}, \ldots, \underline{y_d}\right) = \left(y_{1,1}, \ldots, y_{1,n_1}, \ldots, y_{d,1}, \ldots, y_{d,n_d}\right)$.

2.3. Construction of hierarchical Archimedean copulas defined with multivariate compound distributions

We use the inversion method to identify the proposed hierarchical Archimedean copula C defined with a multivariate compound distribution. Indeed, applying Sklar's Theorem (see e.g. Sklar, 1959 or Nelsen, 2007) and letting $y_{i,j} = \overline{F}^{-1}_{Y_{i,j}}\left(u_{i,j}\right)$ in (13), a hierarchical Archimedean copula C defined with a multivariate compound distribution is obtained from the multivariate survival function given in (19) of the multivariate mixed exponential distribution as follows :

$$\begin{aligned}C\left(\underline{u}\right) &= \overline{F}_{\underline{Y}}\left(\overline{F}^{-1}_{Y_{1,1}}\left(u_{1,1}\right), \ldots, \overline{F}^{-1}_{Y_{1,n_1}}\left(u_{1,n_1}\right), \ldots, \right. \\ &\qquad \left. \overline{F}^{-1}_{Y_{d,1}}\left(u_{d,1}\right), \ldots, \overline{F}^{-1}_{Y_{d,n_d}}\left(u_{d,n_d}\right)\right) \\ &= \mathcal{L}_{\underline{\Theta}}\left(\sum_{j=1}^{n_1} \mathcal{L}^{-1}_{\Theta_1}\left(u_{1,j}\right), \ldots, \sum_{j=1}^{n_d} \mathcal{L}^{-1}_{\Theta_d}\left(u_{d,j}\right)\right),\end{aligned} \tag{20}$$

where $\underline{u} = (\underline{u_1}, \ldots, \underline{u_d})$ with $\underline{u_i} = (u_{i,1}, \ldots, u_{i,n_i})$ for $i = 1, \ldots, d$.

Let us examine more deeply the structure of C. The expression for the multivariate Archimedean copula in (20) is written in terms of the multivariate LST of the multivariate mixing random vector $\underline{\Theta}$ with the sum of the inverse of the corresponding marginal LSTs of Θ_1, …, Θ_d evaluated at each element of a subgroup as components. Hierarchical Archimedean copulas built with our proposed approach can be seen as multivariate extensions to the classical univariate Archimedean copulas which are defined as a univariate LST evaluated at the sum of the inverse of the LST evaluated at each element of a multivariate uniform random vector as follows:

$$C(\underline{u}) = \mathcal{L}_\Theta\left(\sum_{i=1}^{d} \mathcal{L}^{-1}_\Theta(u_i)\right).$$

Given the connection with the mixing random vector $\underline{\Theta}$ and the random variables M, B_i $(i = 1, \ldots, d)$, an alternative representation of the copula C in (20) can be established in terms of $\mathcal{L}_M$ and $\mathcal{L}_{B_i}$ $(i = 1, \ldots, d)$, and $\mathcal{L}_{\Theta_i}$ $(i = 1, \ldots, d)$. Indeed, letting (10) and (15) in (20), the hierarchical Archimedean copula C can also be written as follows:

$$C\left(\underline{u}\right) = \mathcal{L}_M\left(\sum_{i=1}^{d} - \ln\left(\mathcal{L}_{B_i}\left(\sum_{j=1}^{n_i} \mathcal{L}^{-1}_{\Theta_i}\left(u_{i,j}\right)\right)\right)\right). \tag{21}$$

Finally, inserting (17) in (21), the hierarchical Archimedean copula C in (21) can be defined solely in terms of $\mathcal{L}_M$ and $\mathcal{L}_{B_i}$ $(i = 1, \ldots, d)$ as

$$\begin{aligned}&C\left(\underline{u}; \mathcal{L}_M, \mathcal{L}_{B_1}, \ldots, \mathcal{L}_{B_d}\right) \\ &\quad = \mathcal{L}_M\left(\sum_{i=1}^{d} - \ln\left(\mathcal{L}_{B_i}\left(\sum_{j=1}^{n_i} \mathcal{L}^{-1}_{B_i}\left(\exp\left(-\mathcal{L}^{-1}_M\left(u_{i,j}\right)\right)\right)\right)\right)\right).\end{aligned} \tag{22}$$

Note that in (22) the terms $\mathcal{L}_M$, $\mathcal{L}_{B_1}$, …, $\mathcal{L}_{B_d}$ are explicitly inserted as arguments of the function C. We will add these arguments only when we find relevant to insist on their implication in the copula representation.

2.4. Dependence structure of hierarchical Archimedean copulas defined with multivariate compound distributions

Either from (21) or (22), it is clear that the proposed hierarchical Archimedean copulas have explicit forms and offer many possibilities of dependence structures through the choices of distributions for the strictly positive discrete rv M and the strictly positive rvs $B_1, \ldots, B_d$. These copula constructions allow flexibility without requiring a large number of parameters. As stated in Brechmann (2014), "a major issue of any copula model is to find a good balance between parsimony and flexibility".

The dependence structure associated to the hierarchical Archimedean copula defined in either (21) or (22) can be illustrated with a tree representation as shown in Fig. 1 in which $\underline{U} = (\underline{U_1}, \ldots, \underline{U_d})$ with $\underline{U_i} = (U_{i,1}, \ldots, U_{i,n_i})$ for $i = 1, \ldots, d$, is a vector

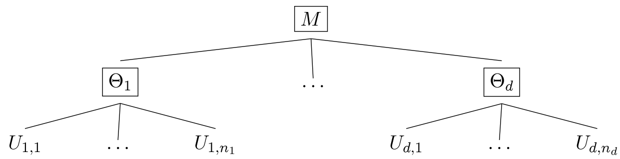

Fig. 1. One level tree structure.

of rvs with multivariate cdf C. This shows a one level tree structure composed of d subgroups joined together by their common link with the rv M. The dependence among the components of $\underline{U_i}$ is captured by Archimedean copulas generated by the mixing rvs Θ_i, for $i = 1, 2, \ldots, d$. Then, the dependence relation between the vectors $\underline{U_1}, \ldots, \underline{U_n}$ is defined through the rv M.

Let us examine the dependence structure for specific subsets of $\underline{U}$. First, the multivariate cdf of $(\underline{U_{i_1}}, \ldots, \underline{U_{i_m}})$ is given by the hierarchical Archimedean copula

$$C(\underline{u_{i_1}}, \ldots, \underline{u_{i_m}}; \mathcal{L}_M, \mathcal{L}_{B_{i_1}}, \ldots, \mathcal{L}_{B_{i_m}}), \tag{23}$$

for $i_1, \ldots, i_m \in \{1, 2, \ldots, d\}$ and for $m = 2, \ldots, d$. Also, if we choose at least two subgroups i_1 and i_2, the multivariate cdf of any vector formed with at least two rvs from each one of them is a hierarchical Archimedean copula of the similar form.

Now we consider the dependence structure within a subgroup i ($i = 1, 2, \ldots, d$). From (20), the multivariate cdf of $(U_{i,j_1}, \ldots, U_{i,j_l})$ (for $j_1, \ldots, j_l \in \{1, 2, \ldots, n_i\}$ and for $l = 2, \ldots, n_i$) is given by

$$C\left(u_{i,j_1}, \ldots, u_{i,j_l}; \mathcal{L}_{\Theta_i}\right) = \mathcal{L}_{\Theta_i}\left(\sum_{k=1}^{l} \mathcal{L}_{\Theta_i}^{-1}\left(u_{i,k}\right)\right) \tag{24}$$

which corresponds to an l-variate Archimedean copula with generator $\mathcal{L}_{\Theta_i}$. As a special case, the multivariate cdf of $\underline{U_i}$ is $C\left(\underline{u_i}; \mathcal{L}_{\Theta_i}\right)$ for $i = 1, \ldots, d$. Due to (11) and (17), copulas of the form as given in (24) are Archimedean copulas based on two generators $\mathcal{L}_M$ and $\mathcal{L}_{B_i}$. Such copulas allow for more flexibility by using at least two parameters (see e.g. Nelsen, 2007, Section 4.5, for other examples of two parameter Archimedean copulas).

Finally, we turn our attention to the dependence structure for vectors of rvs containing at most one component from each subgroup of $\underline{U}$. For $j_1 \in \{1, 2, \ldots, n_1\}, \ldots, j_d \in \{1, 2, \ldots, n_d\}$, the multivariate cdf of $(U_{1,j_1}, \ldots, U_{d,j_d})$ is obtained from (10) and (20) and is given by

$$\begin{aligned} C\left(u_{1,j_1}, \ldots, u_{d,j_d}\right) &= \mathcal{L}_{\Theta_1,\ldots,\Theta_d}\left(\mathcal{L}_{\Theta_1}^{-1}\left(u_{1,j_1}\right), \ldots, \mathcal{L}_{\Theta_d}^{-1}\left(u_{d,j_d}\right)\right) \\ &= \mathcal{P}_M\left(\mathcal{L}_{B_1}\left(\mathcal{L}_{\Theta_1}^{-1}\left(u_{1,j_1}\right)\right) \ldots \mathcal{L}_{B_d}\left(\mathcal{L}_{\Theta_d}^{-1}\left(u_{d,j_d}\right)\right)\right). \end{aligned} \tag{25}$$

Using (14), (25) becomes

$$\begin{aligned} &C\left(u_{1,j_1}, \ldots, u_{d,j_d}\right) \\ &= \mathcal{P}_M\left(\mathcal{L}_{B_1}\left(\mathcal{L}_{B_1}^{-1}\left(\mathcal{P}_M^{-1}\left(u_{1,j_1}\right)\right)\right) \ldots \mathcal{L}_{B_d}\left(\mathcal{L}_{B_d}^{-1}\left(\mathcal{P}_M^{-1}\left(u_{d,j_d}\right)\right)\right)\right) \\ &= \mathcal{P}_M\left(\mathcal{P}_M^{-1}\left(u_{1,j_1}\right) \ldots \mathcal{P}_M^{-1}\left(u_{d,j_d}\right)\right) \\ &= \mathcal{L}_M\left(-\ln\left(\mathcal{P}_M^{-1}\left(u_{1,j_1}\right)\right) - \cdots - \ln\left(\mathcal{P}_M^{-1}\left(u_{d,j_d}\right)\right)\right). \end{aligned} \tag{26}$$

Then, from (16) and (26), we conclude that the multivariate cdf of $(U_{1,j_1}, \ldots, U_{d,j_d})$ is

$$C\left(u_{1,j_1}, \ldots, u_{d,j_d}; \mathcal{L}_M\right) = \mathcal{L}_M\left(\sum_{i=1}^{d} \mathcal{L}_M^{-1}\left(u_{i,j_i}\right)\right), \tag{27}$$

which corresponds to a single-generator Archimedean copula with $\mathcal{L}_M$. Given (27), it means that the dependence relation between the components of $(U_{1,j_1}, \ldots, U_{d,j_d})$ is only defined through the rv M.

2.5. Dependence properties

In light of the previous results, we aim to measure and compare the strength of the dependence relations between the rvs of a given subgroup and also between rvs from different subgroups. In other words, we want to compare whether the copula generated from $\mathcal{L}_\Theta$ is more dependent than the one with a generator $\mathcal{L}_M$. To do so, we need to introduce the concept of concordance ordering as defined in Joe (1997) page 37.

Definition 1. Let C_1 and C_2 be two d-dimensional copulas. C_2 is more **concordant** that C_1, written $C_1 \prec_c C_2$, if

$$C_1\left(\underline{u}\right) \le C_2\left(\underline{u}\right) \text{ and } \overline{C_1}\left(\underline{u}\right) \le \overline{C_2}\left(\underline{u}\right),$$

for $\underline{u} \in [0, 1]^d$.

A very useful consequence of the concordance ordering is its relation with dependence measures. For instance, if C_1 and C_2 are copulas with respective Kendall taus τ_1, τ_2, Spearman rhos $\rho_S^{(1)}$, $\rho_S^{(2)}$, tail dependence parameters λ_1, λ_2, and $C_1 \prec_c C_2$, then $\tau_1 \le \tau_2$, $\rho_S^{(1)} \le \rho_S^{(2)}$, and $\lambda_1 \le \lambda_2$. (See Joe, 1997.)

Suppose now that C_1 and C_2 are Archimedean copulas with respective generators $\mathcal{L}_1$ and $\mathcal{L}_2$. Theorems 4.1 and 4.7 in Joe (1997) provide the condition that guarantees that $C_1 \prec_c C_2$ is verified. Such a condition states that $\left(\mathcal{L}_1^{-1} \circ \mathcal{L}_2\right)'$ is a completely monotone function.

We recall that our objective is to compare the copula C_Θ, generated from $\mathcal{L}_\Theta$, with the one generated from $\mathcal{L}_M$ (written C_M). Since $\mathcal{L}_\Theta$ is related to $\mathcal{L}_M$, this condition is easily verified as follows:

$$\mathcal{L}_M^{-1} \circ \mathcal{L}_\Theta = \mathcal{L}_M^{-1} \circ \mathcal{L}_M\left(-\ln\left(\mathcal{L}_B\right)\right) = -\ln\left(\mathcal{L}_B\right).$$

Therefore, if $(-\ln(\mathcal{L}_B))'$ is completely monotone, then the dependence within a subgroup is stronger than the outer dependence between subgroups, i.e. $C_M \prec_c C_\Theta$ and hence $\tau_M \le \tau_\Theta$.

It is well known that $(-\ln(\mathcal{L}_{B_i}))'$ is completely monotone if and only if $\mathcal{L}_{B_i}^m$ is the LST of a positive rv i.e. $\mathcal{L}_{B_i}^m \in \Psi_\infty$ for all $m \in \mathbb{N}$ (see Joe, 1997). Table 1 provides a list of distributions, used to generate an Archimedean copula, for which it is easy to directly verify that $\mathcal{L}_{B_i}^m$ is the LST of a positive rv. We denote by $SNB(m; r, q)$ a shifted negative binomial distribution i.e. $B_{i,1} + \cdots + B_{i,m}$ admits a stochastic representation $N + m$, where $N \sim NB(r, q)$, $r \in \mathbb{R}^+$ and $q \in (0, 1)$. For the logarithmic distribution, $\sum_{j=1}^m B_{i,j}$ is an $m-$fold sum of iid logarithmic rvs. If B_i follows a Sibuya distribution, $\sum_{j=1}^m B_{i,j}$ is distributed as an unknown rv N with probability mass function $p_k = \sum_{i=1}^\infty \binom{m}{i}\binom{\alpha i}{k}(-1)^{i+k}$, $k = 1, 2, \ldots$. Finally, for a stable distribution i.e. $B_i \sim S\left(\alpha, 1, \cos(\alpha\pi/2)^{\frac{1}{\alpha}}, \mathbf{1}_{\{\alpha=1\}}; 1\right)$, $\mathcal{L}_{B_i}^m$ is the LST of a positive rv exponentially tilted stable distributed $\tilde{S}\left(\alpha, 1, (\cos(\alpha\pi/2)m)^{\frac{1}{\alpha}}, m\,\mathbf{1}_{\{\alpha=1\}}, h\,\mathbf{1}_{\{\alpha\neq1\}}; 1\right)$ with $h = 0$. For sampling procedures, we refer to Hofert (2011) and Devroye (2009).

2.6. Examples of hierarchical Archimedean copulas

In the following two examples, we give a glimpse of the flexibility of our approach based on multivariate compound distributions in the construction of a large variety of hierarchical Archimedean copulas C. One must first choose the distribution for the strictly positive discrete rv M. Here, we limit ourselves to shifted-geometric and logarithmic distributions for M. Of course, other candidates would have been suitable, such as the Sibuya and shifted-negative binomial distributions (see e.g. Joe, 2014). Next, distributions for the rvs B_i are selected. Either shifted-geometric, logarithmic, or gamma distributions are picked. Among the possible combinations of distributions, we choose to depict 6 hierarchical Archimedean copulas constructed under the proposed

Table 1
Distributions for the sum of common laws.

Distribution of B_i	$\mathcal{L}_{B_i}^{m}(t)$	Distribution of $B_{i,1}+\cdots+B_{i,m}$
ShiftedGeometric (q)	$\left(\frac{qe^{-t}}{1-(1-q)e^{-t}}\right)^{m}$	$SNB(m;m,q)$
ShiftedNegativeBinomial $(r;r,q)$	$\left(\frac{qe^{-t}}{1-(1-q)e^{-t}}\right)^{mr}$	$SNB(rm;rm,q)$
Logarithmic (γ)	$\left(\frac{ln(1-\gamma e^{-t})}{ln(1-\gamma)}\right)^{m}$	$\sum_{i=1}^{m} B_i, B_i \sim Logarithmic(\gamma)$
Sibuya (α)	$\left(1-(1-e^{-t})^{\alpha}\right)^{m}$	$N \sim p_k = \sum_{i=1}^{\infty}\binom{m}{i}\binom{\alpha i}{k}(-1)^{i+k}$
Gamma $(\alpha, 1)$	$\left(\frac{1}{1+t}\right)^{m\alpha}$	Gamma$(m\alpha, 1)$
Stable $\left(\alpha, 1, \cos(\alpha\pi/2)^{\frac{1}{\alpha}}, \mathbf{1}_{\{\alpha=1\}}; 1\right)$	$(\exp(-t^{\alpha}))^{m}$	$\tilde{S}\left(\alpha, 1, (\cos(\alpha\pi/2)m)^{\frac{1}{\alpha}}, m\,\mathbf{1}_{\{\alpha=1\}}, 0; 1\right)$

approach. In the first example, we present 3 copulas with an underlying shifted-geometric distribution for M, while in the second example a logarithmic distribution describes the behavior of the rv M. By convention, when the distributions for the rvs B_i ($i = 1, \ldots, d$) are from the same family of distributions (with different parameters), we refer to the resulting copula as "distribution for M"-"distribution for B_i" hierarchical Archimedean copula.

Example 1. Let $M \sim ShifGeo\,(q)$ with $\mathcal{L}_M(t) = \frac{qe^{-t}}{1-(1-q)e^{-t}}$ and $\mathcal{L}_M^{-1}(t) = -\ln(\frac{1}{\frac{q}{t}+1-q})$. Letting $\alpha = 1-q$, the multivariate cdf of $(U_{1,j_1}, \ldots, U_{d,j_d})$ corresponds to the Ali–Mikhail–Haq (AMH) copula with generator $\mathcal{L}_M$, i.e.,

$$C\left(u_{1,j_1}, \ldots, u_{d,j_d}; \mathcal{L}_M\right) = \frac{\alpha\prod_{i=1}^{n} u_i}{\prod_{i=1}^{n}(\alpha+(1-\alpha)u_i) - (1-\alpha)\prod_{i=1}^{n} u_i}.$$

1. Geometric–Geometric hierarchical Archimedean copula:
   - $B_i \sim ShifGeo\,(q_i)$ with $\mathcal{L}_{B_i}(t) = \frac{q_i e^{-t}}{1-(1-q_i)e^{-t}}$ and $\mathcal{L}_{B_i}^{-1}(t) = -\ln(\frac{1}{\frac{q_i}{t}+1-q_i})$;
   - parameters: $q, q_i, \ldots, q_d \in (0,1)$;
   - copula C (see equation given in Box I).
2. Geometric–Logarithmic hierarchical Archimedean copula:
   - $B_i \sim Logarithmic\,(\gamma_i)$ with $\mathcal{L}_{B_i}(t) = \frac{\ln(1-\gamma_i e^{-t})}{\ln(1-\gamma_i)}$ and $\mathcal{L}_{B_i}^{-1}(t) = -\ln(\frac{(1-(1-\gamma_i)^t)}{\gamma_i})$;
   - parameters: $q, \gamma_1 = (1-e^{-\alpha_1}), \ldots, \gamma_d = (1-e^{-\alpha_d}) \in (0,1)$;
   - copula C:

$$C\left(\underline{u}\right) = \frac{-q\prod_{i=1}^{d}\ln\left(1-(1-e^{-\alpha_i})\prod_{j=1}^{n_i}\left(\frac{1-e^{-\alpha_i\left(\frac{q}{u_{i,j}}+1-q\right)^{-1}}}{(1-e^{-\alpha_i})}\right)\right)}{\prod_{i=1}^{d}\alpha_i+(1-q)\prod_{i=1}^{d}\ln\left(1-(1-e^{-\alpha_i})\prod_{j=1}^{n_i}\left(\frac{1-e^{-\alpha_i\left(\frac{q}{u_{i,j}}+1-q\right)^{-1}}}{(1-e^{-\alpha_i})}\right)\right)}.$$

3. Geometric-Gamma hierarchical Archimedean copula:
   - $B_i \sim Gamma\,(\alpha_i, 1)$ with $\mathcal{L}_{B_i}(t) = \left(\frac{1}{1+t}\right)^{\alpha_i}$ and $\mathcal{L}_{B_i}^{-1}(t) = t^{-\frac{1}{\alpha_i}} - 1$;
   - parameters: $q \in (0,1), \alpha_i, \ldots, \alpha_d \in \mathbb{R}^+$;
   - copula C:

$$C\left(\underline{u}\right) = \frac{q\prod_{i=1}^{d}\left(\sum_{j=1}^{n_i}\left(\frac{q}{u_{i,j}}+1-q\right)^{\frac{1}{\alpha_i}}-(n_i-1)\right)^{-\alpha_i}}{1-(1-q)\prod_{i=1}^{d}\left(\sum_{j=1}^{n_i}\left(\frac{q}{u_{i,j}}+1-q\right)^{\frac{1}{\alpha_i}}-(n_i-1)\right)^{-\alpha_i}}. \quad \square$$

Example 2. We assume that $M \sim Logarithmic\,(\gamma)$ with $\mathcal{L}_M(t) = \frac{\ln(1-\gamma e^{-t})}{\ln(1-\gamma)}$ and $\mathcal{L}_M^{-1}(t) = -\ln(\frac{(1-(1-\gamma)^t)}{\gamma})$. Letting $\alpha = \ln(1-\gamma)$ the multivariate cdf of $(U_{1,j_1}, \ldots, U_{d,j_d})$ corresponds to the Frank copula with generator $\mathcal{L}_M$, i.e.,

$$C\left(u_{1,j_1}, \ldots, u_{d,j_d}; \mathcal{L}_M\right) = \frac{-1}{\alpha}\ln\left(1-\frac{\prod_{i=1}^{n}\left(1-e^{-\alpha u_i}\right)}{\left(1-e^{-\alpha}\right)^{n-1}}\right).$$

1. Logarithmic–Geometric hierarchical Archimedean copula:
   - $B_i \sim ShifGeo\,(q_i)$ with $\mathcal{L}_{B_i}(t) = \frac{q_i e^{-t}}{1-(1-q_i)e^{-t}}$ and $\mathcal{L}_{B_i}^{-1}(t) = -\ln(\frac{1}{\frac{q_i}{t}+1-q_i})$;
   - parameters: $\gamma = (1-e^{-\alpha}), q_i, \ldots, q_d \in (0,1)$;
   - copula C:

$$C\left(\underline{u}\right) = -\frac{1}{\alpha}\ln\left(1-\frac{(1-e^{-\alpha})\prod_{i=1}^{d} q_i\prod_{j=1}^{n_j}\left(\frac{q_i(1-e^{-\alpha})}{1-e^{-\alpha u_{i,j}}}+1-q_i\right)^{-1}}{\prod_{i=1}^{d}\left(1-(1-q_i)\prod_{j=1}^{n_i}\left(\frac{q_i(1-e^{-\alpha})}{1-e^{-\alpha u_{i,j}}}+1-q_i\right)^{-1}\right)}\right).$$

2. Logarithmic–Logarithmic hierarchical Archimedean copula:
   - $B_i \sim Logarithmic\,(\gamma_i)$ with $\mathcal{L}_{B_i}(t) = \frac{\ln(1-\gamma_i e^{-t})}{\ln(1-\gamma_i)}$ and $\mathcal{L}_{B_i}^{-1}(t) = -\ln\left(\frac{(1-(1-\gamma_i)^t)}{\gamma_i}\right)$;
   - parameters: $\gamma = (1-e^{-\alpha}), \gamma_1 = (1-e^{-\alpha_1}), \ldots, \gamma_d = (1-e^{-\alpha_d}) \in (0,1)$;
   - copula C:

$$C\left(\underline{u}\right) = \frac{-1}{\alpha}\ln\left(1-(1-e^{-\alpha})\prod_{i=1}^{d}-\frac{1}{\alpha_i}\right.$$
$$\left.\times\ln\left(1-(1-e^{-\alpha_i})\prod_{j=1}^{n_i}\left(\frac{1-e^{-\alpha_i\left(\frac{1-e^{-\alpha\,u_{i,j}}}{1-e^{-\alpha}}\right)}}{1-e^{-\alpha_i}}\right)\right)\right).$$

3. Logarithmic-Gamma hierarchical Archimedean copula:
   - $B_i \sim Gamma\,(\alpha_i, 1)$ with $\mathcal{L}_{B_i}(t) = \left(\frac{1}{1+t}\right)^{\alpha_i}$ and $\mathcal{L}_{B_i}^{-1}(t) = t^{-\frac{1}{\alpha_i}} - 1$;
   - parameters: $\gamma = (1-e^{-\alpha}) \in (0,1), \alpha_1, \ldots, \alpha_d \in \mathbb{R}^+$;
   - copula C:

$$C\left(\underline{u}\right) = \frac{-1}{\alpha}\ln\left(1-(1-e^{-\alpha})\right.$$
$$\left.\times\prod_{i=1}^{d}\left(\sum_{j=1}^{n_i}\left(\frac{1-e^{-\alpha\,u_{i,j}}}{1-e^{-\alpha}}\right)^{\frac{-1}{\alpha_i}}-(n_i-1)\right)^{-\alpha_i}\right). \quad \square$$

Obviously, hierarchical Archimedean copulas can be built assuming that the distributions of the rvs B_i belong to different

$$C\left(\underline{u}\right)=\frac{q\prod_{i=1}^{d}q_i\prod_{j=1}^{n_i}\left(\frac{u_{i,j}}{u_{i,j}-qq_i\left(u_{i,j}-1\right)}\right)}{\prod_{i=1}^{d}\left(1-(1-q_i)\prod_{j=1}^{n_i}\left(\frac{u_{i,j}}{u_{i,j}-qq_i\left(u_{i,j}-1\right)}\right)\right)\left(1-\frac{(1-q)\prod_{i=1}^{d}q_i\prod_{j=1}^{n_i}\left(\frac{u_{i,j}}{u_{i,j}-qq_i\left(u_{i,j}-1\right)}\right)}{\prod_{i=1}^{d}\left(1-(1-q_i)\prod_{j=1}^{n_i}\left(\frac{u_{i,j}}{u_{i,j}-qq_i\left(u_{i,j}-1\right)}\right)\right)}\right)}.$$

Box I.

$$C(\underline{u_1},\underline{u_2})=\frac{q\left(\sum_{j=1}^{n_2}\left(\frac{q}{u_{2,j}}+1-q\right)^{\frac{1}{\alpha}}-1\right)^{-\alpha}\ln\left(1-\gamma\exp\left(\sum_{i=1}^{n_1}\ln\left(\frac{1-\sqrt[q]{\frac{q}{u_{1,i}}+1-q}\,\sqrt[q]{1-\gamma}}{\gamma}\right)\right)\right)}{\ln(1-\gamma)\left(1-\frac{\left(\sum_{j=1}^{n_2}\left(\frac{q}{u_{2,j}}+1-q\right)^{\frac{1}{\alpha}}-1\right)^{-\alpha}\ln\left(1-\gamma\exp\left(\sum_{i=1}^{n_1}\ln\left(\frac{1-\sqrt[q]{\frac{q}{u_{1,i}}+1-q}\,\sqrt[q]{1-\gamma}}{\gamma}\right)\right)\right)}{\ln(1-\gamma)(1-q)^{-1}}\right)}.$$

Box II.

families. Since we cannot list all possible hierarchical Archimedean copulas that can obtained this way, we limit ourselves to providing, in the following example, an illustration of a hierarchical Archimedean copula assuming $d = 2$ subgroups with different families of distributions B_1 and B_2.

Example 3. Let $d = 2$ and $M \sim ShiftedGeo(q)$ with $\mathcal{L}_M(t) = \frac{qe^{-t}}{1-(1-q)e^{-t}}$, $B_1 \sim Log(\gamma)$ with $\mathcal{L}_{B_1}(t) = \frac{\ln(1-\gamma e^{-t})}{\ln(1-\gamma)}$ and finally, $B_2 \sim Gamma(\alpha, 1)$ with $\mathcal{L}_{B_2}(t) = \left(\frac{1}{t+1}\right)^{\alpha}$. Then, the expression of the 3-parameters copula is given by the equation given in Box II. □

2.7. Sampling hierarchical Archimedean copulas

Inspired from Marshall and Olkin (1988), the sampling procedure for a hierarchical Archimedean copula C defined with a multivariate compound distribution follows from its construction. The algorithm given just below aims to simulate samples of a random vector $\underline{U} = (\underline{U_1}, \dots, \underline{U_d})$ (with $\underline{U_i} = (U_{i,1}, \dots, U_{i,n_i})$, for $i = 1, \dots, d$) whose multivariate cdf is the copula C.

Algorithm 1. Let C be a hierarchical Archimedean copula with d subgroups and root M. Define $\underline{U}$ as a vector of standard uniformly distributed rvs with cdf C.

1. Sample M;
2. For each subgroup ($i = 1, \dots, d \geq 2$):
   2.1. Sample $B_{i,k}$ for $k = 1, \dots, M$;
   2.2. Return $\Theta_i = \sum_{k=1}^{M} B_{i,k}$;
   2.3. Sample $R_{i,j} \sim Exp(1)$ for $j = 1, \dots, n_i$;
   2.4. Return $U_{i,j} = \mathcal{L}_{\Theta_i}(R_{i,j}/\Theta_i)$ for $j = 1, \dots, n_i$;
3. Return $\underline{U} = (U_{1,1}, \dots, U_{1,n_1}, \dots, U_{d,1}, \dots, U_{d,n_d})$.

An application of Algorithm 1 is provided in the following example.

Example 4. Let C be the Geometric-Gamma hierarchical Archimedean copula presented in Example 1 with $d = 2$ and $n_1 = n_2 = 2$. To illustrate the pairwise dependence between components of $\underline{U} = (U_{1,1}, U_{1,2}, U_{2,1}, U_{1,2})$, we give, in Fig. 2, a graph of sampled values of $\underline{U}$, where the parameters of the copula C are $q = 0.1$, $\alpha_1 = 0.04$ and $\alpha_2 = 0.2$. Note that τ, in Fig. 2, corresponds to the Kendall's tau.

It is clear that the sampled values of $(U_{1,1}, U_{1,2})$ provide scatter plots that are similar to those of the Clayton copula with a dependence subdued by the shifted geometric distribution. The same phenomenon happens for the sampled values of $(U_{2,1}, U_{2,2})$. Since the dependence structures of the four other pairs $(U_{1,1}, U_{2,1})$, $(U_{1,1}, U_{2,2})$, $(U_{1,2}, U_{2,1})$, $(U_{1,2}, U_{2,2})$ depend on M, their scatter plots are those of an AMH copula. □

2.8. Multi-level hierarchical Archimedean copulas

Until now, we have only considered one-level hierarchical Archimedean copulas, for which the associated typical one-level tree representation is provided in Fig. 1. Naturally, it is possible to design multi-level hierarchical Archimedean copulas by adding ramifications such that their dependence structures can be represented as multi-level trees.

Let $\underline{U} = (\underline{U_1}, \dots, \underline{U_d})$ be a random vector with a cdf defined by a multi-level hierarchical Archimedean copula C, such that each subgroup $\underline{U_i}$, $i = 1, \dots, n_i$, characterized by its genetic code (the path in the hierarchy) e_i and its dimension n_i, can be written as

$$\underline{U_i} = \left(U_{e_i,j},\ j = 1, \dots, n_i\right).$$

This genetic code is represented by a vector in which the first element corresponds to the root of the tree (0 by default) and the kth element represents the position of a node in the $(k-1)$th level, within its mother's direct descendants, assuming that the leaves are always at the far right. A value of 0 at the last position of the vector represents a direct link between a leaf and a mother node.

The design of the hierarchical structure comes from considering the primary distribution of the compound rvs as a random sum. Let $\underline{U_i}$ be a subgroup with genetic code e_i, and let ξ be a right truncated vector of e_i, and k the $(\dim(\xi)+1)th$ element of e_i, where $\dim(.)$ is the dimension of a vector. On one hand, if $\dim(\xi) < \dim(e_i) - 1$, then the rv $M^{(\xi,k)}$ is defined as a random sum, i.e, $M^{(\xi,k)} = \sum_{j=1}^{M^{(\xi)}} N_j^{(\xi,k)}$ where $N_j^{(\xi,k)}$ is distributed as a discrete rv $N^{(\xi,k)}$ which plays the same role as the rv B_i in Section 2.1. On the other hand, if $\dim(\xi) = \dim(e_i) - 1$, then the procedure is similar to the one presented in Section 2.1.

A typical hierarchical structure can be depicted by a tree where the root is always $M^{(0)}$ and all leaves are of type "u". A node of type "M", at a given level of the tree, can have children of all available types, meaning "M","Θ" and "u". Further, Θ nodes can only have u children (leaves), while leaves have no children at all.

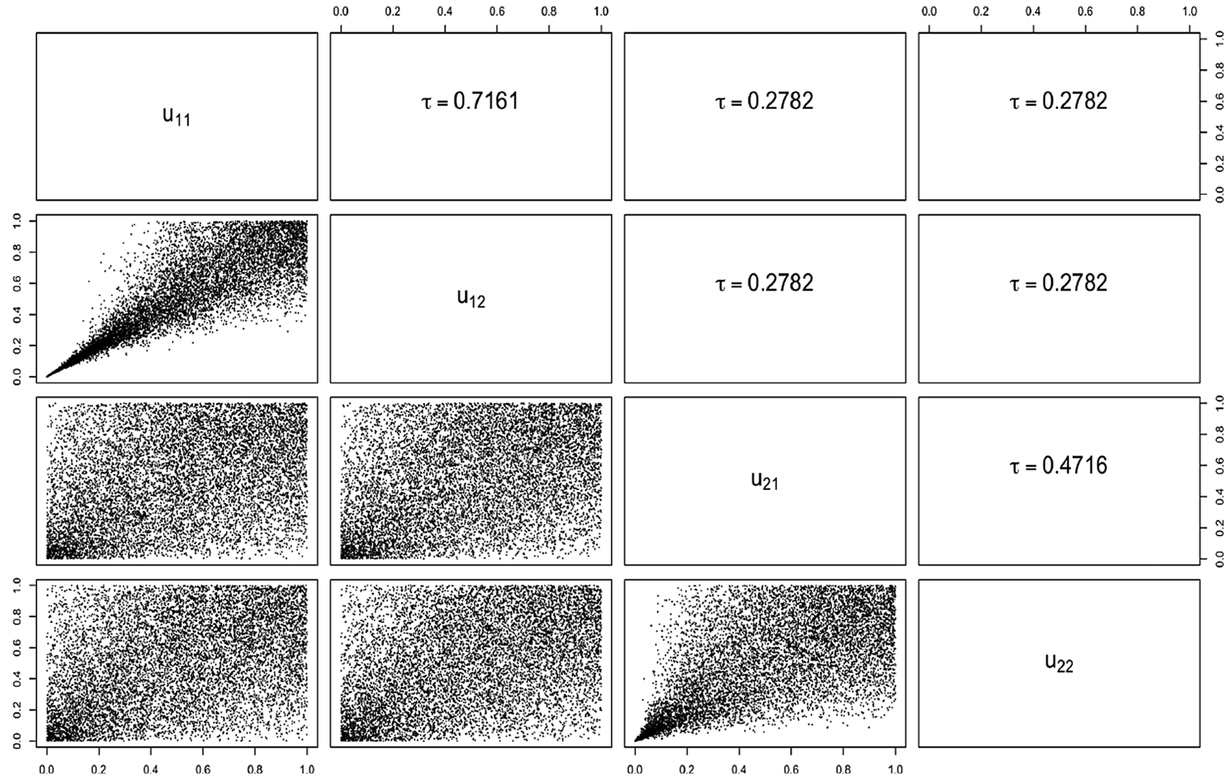

Fig. 2. Pairwise graph of the 10 000 4-dimensional vectors of realizations sampled from the hierarchical Archimedean copula with multivariate compound distributions in Example 4.

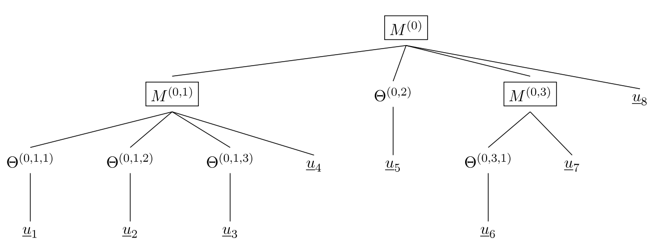

Fig. 3. Example of a multi-level hierarchical structure.

To illustrate the notation, let us consider the tree construction depicted in Fig. 3:

The tree in Fig. 3 corresponds to the dependence structure of a random vector $\underline{U} = (\underline{U}_1, \ldots, \underline{U}_8)$ characterized by the following genetic codes:

$e_1 = (0, 1, 1)$	$e_2 = (0, 1, 2)$	$e_3 = (0, 1, 3)$	$e_4 = (0, 1, 0)$
$e_5 = (0, 2)$	$e_6 = (0, 3, 1)$	$e_7 = (0, 3, 0)$	$e_8 = (0, 0)$

Note that, one can deduce a tree from any given set of genetic codes. This set of codes constitutes a simple tool to summarize all the information in regard to the dependence structure depicted by the tree.

For any given multi-level hierarchical copula, it is possible to derive a tailored sampling procedure by adapting the one described in Algorithm 1. Algorithm 2 provides a sampling procedure for several hierarchical levels, assuming the three different types of children.

Algorithm 2. Let C be a multi-level hierarchical copula with root $M^{(0)}$ and d subgroups. Define $\underline{U}$ as a vector of standard uniformly distributed rvs with cdf C. Let e_i be the genetic code of each subgroup i and let e_i' be the vector e_i without the last element.

1. Fix $j = 2$
2. For each genetic code with $\dim(e_i) = j$ do:
   2.1. Sample $M^{(e_i')}$
   2.2. If the last element of e_i is zero i.e. $e_{i,\dim(e_i)} = 0$:
      i. Sample $R \sim Exp(1)$
      ii. Set the corresponding component of $\underline{U_i}$ as $\mathcal{L}_{M^{(e_i')}}\left(\frac{R}{M^{(e_i')}}\right)$
   2.3. Else:
      i. Sample $\Theta^{(e_i)}$
      ii. Sample $R \sim Exp(1)$
      iii. Set the corresponding component of $\underline{U_i}$ as $\mathcal{L}_{\Theta^{(e_i)}}\left(\frac{R}{\Theta^{(e_i)}}\right)$
3. Set $j = j + 1$
4. Repeat from 2
5. Return $\underline{U}$.

Algorithm 2 proceeds as follows: we start by simulating the root, which is always of type "M". Next, we perform a series of simulations in the following order: simulate all u-children of the root, simulate all Θ-children of the root then simulate all M-children of the root. Finally, we move the root to M and repeat the same logic over again.

In the following illustration, we provide an example of a 2-level hierarchical Archimedean copula.

Example 5 (*Two level tree*)**.** We consider a 2-level hierarchical Archimedean copula C, for which the associated 2-level tree representation is given in Fig. 4. The 6-dimensional copula C is the multivariate cdf of the random vector $\underline{U} = (U_{(0,2),1}, U_{(0,2),2}, U_{(0,1,1),1}, U_{(0,1,1),2}, U_{(0,1,2),1}, U_{(0,1,1),2})$. Let the rvs $\Theta^{(0,2)}$, $\Theta^{(0,1,1)}$, and $\Theta^{(0,1,2)}$

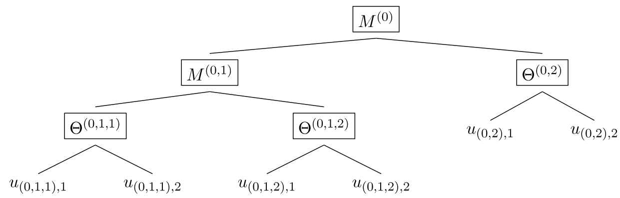

Fig. 4. Two level tree structure.

be associated to the pairs $(U_{(0,2),1}, U_{(0,2),2})$, $(U_{(0,1,1),1}, U_{(0,1,1),2})$ and $(U_{(0,1,2),1}, U_{(0,1,2),2})$.

We define $M^{(0,1)} = \sum_{k=1}^{M^{(0)}} N_k^{(0,1)}$, $\Theta^{(0,2)} = \sum_{k=1}^{M^{(0)}} B_k^{(0,2)}$ and $\Theta^{(0,1,i)} = \sum_{k=1}^{M^{(0,1)}} B_k^{(0,1,i)}$, for $i = 1, 2$.

By first conditioning on $M^{(0,1)}$ and then on $M^{(0)}$, the LST of $\Theta^{(0,1,i)}$ can be written as

$$\mathcal{L}_{\Theta^{(0,1,i)}}(t) = \mathcal{P}_{M^{(0)}}\left(\mathcal{P}_{N^{(0,1)}}\left(\mathcal{L}_{B^{(0,1,i)}}\left(t\right)\right)\right).$$

Similarly, we obtain the trivariate LST of $(\Theta^{(0,2)}, \Theta^{(0,1,1)}, \Theta^{(0,1,2)})$, which is given by

$$\mathcal{L}_{\Theta^{(0,2)},\,\Theta^{(0,1,1)},\,\Theta^{(0,1,2)}}(t_1, t_2, t_3)$$
$$= \mathcal{P}_{M^{(0)}}\left(\mathcal{L}_{B^{(0,2)}}\left(t_1\right)\,\mathcal{P}_{N^{(0,1)}}\left(\mathcal{L}_{B^{(0,1,1)}}\left(t_2\right)\,\mathcal{L}_{B^{(0,1,2)}}\left(t_3\right)\right)\right).$$

Thus, the 6-dimensional copula C can be written as

$$\begin{aligned} &C(u_{(0,2),1}, u_{(0,2),2}, u_{(0,1,1),1}, u_{(0,1,1),2}, u_{(0,1,2),1}, u_{(0,1,2),2}) \\ &= \mathcal{L}_{\Theta^{(0,2)},\,\Theta^{(0,1,1)},\,\Theta^{(0,1,2)}}\Bigg(\sum_{k=1}^{2} \mathcal{L}_{\Theta^{(0,2)}}^{-1}(u_{(0,2),k}), \\ &\quad \sum_{k=1}^{2} \mathcal{L}_{\Theta^{(0,1,1)}}^{-1}(u_{(0,1,1),k}), \sum_{k=1}^{2} \mathcal{L}_{\Theta^{(0,1,2)}}^{-1}(u_{(0,1,2),k})\Bigg). \end{aligned} \tag{28}$$

We assume that $M^{(0)} \sim ShiftedGeo(0.5)$, $N^{(0,1)} \sim ShiftedGeo(0.1)$ and $B^{(0,2)} \sim B^{(0,1,1)} \sim B^{(0,1,2)} \sim Gamma(\alpha = \frac{1}{30}, \beta = 1)$. In Fig. 5, considering that $B^{(0,2)}$, $B^{(0,1,1)}$ and $B^{(0,1,2)}$ are iid, we clearly see the impact of $M^{(0)}$ and $M^{(0,1)}$ on their respective dependence structure. Using the numbered boxes with coordinates (x, y) in Fig. 5, the scatter plot associated to the pair $(U_{(0,2),1}, U_{(0,1),2})$ (see box (5, 1)) has a shape that is the closest to the one of a Clayton copula with a dependence parameter of 30. This phenomenon is explained by the fact that the couple depends only on $M^{(0)}$ with a parameter of 0.5, which in turn has little to no impact on $\Theta^{(0,2)}$. The scatter plot for the pair $(U_{(0,1,1),1}, U_{(0,1,1),2})$ (see box (1, 5)), however, shows a dependence restrained by $M^{(0,1)}$, which has a much bigger impact due to its overall higher values. We can also see that boxes $\{(1, 1), (1, 2), \ldots, (4, 1), (4, 2)\}$ only depend on $M^{(0)}$, which implies a very small dependence. However, boxes $\{(1, 3), (1, 4), (2, 3), (2, 4)\}$ depend on $M^{(0,1)}$, a rv compounded on $M^{(0)}$. As discussed in Section 2.5, the copula generated from $M^{(0,1)}$ is more concordant than the one generated from $M^{(0)}$ and hence $\tau_{M^{(0,1)}} \geq \tau_{M^{(0)}}$.

3. Representation of the copula as a common mixture

The functional symmetry of Archimedean copulas lead to the introduction by Joe (1997) of nested Archimedean copulas. This popular class of copulas results from nesting Archimedean copulas into each other, allowing asymmetries and multiple hierarchy levels. Such constructions are however limited by nesting conditions that must be fulfilled. In contrast, hierarchical Archimedean copulas based on multivariate compound distributions provide very flexible combinations of Archimedean copula families without such restrictions. Given their similar nature, it is of interest to push further their comparison. For that purpose, we use the common mixture representation of a copula to write the proposed hierarchical Archimedean copula. This will also prove useful for further investigation of aggregation methods.

Using (20) and (21), the common mixture representation of a hierarchical Archimedean copula based on a multivariate compound distribution is given by

$$\begin{aligned} C\left(\underline{u}\right) &= \mathcal{L}_{\Theta_1,\ldots,\Theta_d}\left(\sum_{j=1}^{n_1} \mathcal{L}_{\Theta_1}^{-1}\left(u_{1,j}\right), \ldots, \sum_{j=1}^{n_d} \mathcal{L}_{\Theta_d}^{-1}\left(u_{d,j}\right)\right) \\ &= \mathcal{P}_M\left(\mathcal{L}_{B_1}\left(\sum_{j=1}^{n_1} \mathcal{L}_{\Theta_1}^{-1}\left(u_{1,j}\right)\right) \times \cdots \times \mathcal{L}_{B_d}\left(\sum_{j=1}^{n_d} \mathcal{L}_{\Theta_d}^{-1}\left(u_{d,j}\right)\right)\right) \\ &= \sum_{m=1}^{\infty} \prod_{i=1}^{d} \mathcal{L}_{B_i}^{m}\left(\sum_{j=1}^{n_i} \mathcal{L}_{\Theta_i}^{-1}\left(u_{i,j}\right)\right) f_M\left(m\right). \end{aligned} \tag{29}$$

Since $\mathcal{L}_{B_i}^{m}$ is the LST of a positive rv which we denote by V_i, we can write the copula given in (29) as

$$C\left(\underline{u}\right) = \sum_{m=1}^{\infty} \prod_{i=1}^{d} \Big(\int_0^{\infty} e^{-v_i\left(\sum_{j=1}^{n_i} \mathcal{L}_{\Theta_i}^{-1}(u_{i,j})\right)} \mathrm{d}F_{V_i}(v_i)\Big) f_M(m). \tag{30}$$

From the common mixture representation given in (30), we adapt Algorithm 1 and provide below a more efficient one.

Algorithm 3. Let C be a hierarchical Archimedean copula with d subgroups and root M_0. Define $\underline{U}$ as a vector of standard uniformly distributed rvs with cdf C.

1. Sample M_0;
2. For each subgroup ($i = 1, \ldots, d \geq 2$):
   2.1. Sample V_i with $\mathcal{L}_{V_i}(t) = \mathcal{L}_{B_i}^{M_0}(t)$;
   2.2. Sample $R_{i,j} \sim Exp(1)$ for $j = 1, \ldots, n_i$;
   2.3. Return $U_{i,j} = \mathcal{L}_{\Theta_i}(R_{i,j}/V_i)$ for $j = 1, \ldots, n_i$;
3. Return $\underline{U} = (U_{1,1}, \ldots, U_{1,n_1}, \ldots, U_{d,1}, \ldots, U_{d,n_d})$.

In order to confirm the efficiency of Algorithm 3, let us consider a simple one level tree hierarchical Archimedean copula with two groups where $M \sim Geo(q)$, $B_1 \sim Gamma(2.5, 1)$ and $B_2 \sim Gamma(2.5, 1)$. The top graph of Fig. 6 serves as an illustration to compare computation times (in seconds) between Algorithms 1 and 3 for 100 000 realizations with respect to q, while the bottom one compares both algorithms with respect to the sampling size. Evidently, Algorithm 3 is far more efficient. In fact, Algorithm 3 has a very small computation time since we sample directly from the known distribution of the sum in step 2.1 of Algorithm 3.

4. Comparisons with other construction methodologies of hierarchical Archimedean copulas

4.1. Links with nested Archimedean copulas

Now that hierarchical Archimedean copulas with multivariate compound distributions are written as common mixtures, we aim to compare them to nested Archimedean copulas. To do so, we consider a simple example of a one level partially nested Archimedean copula with d children copulas i.e.

$$\begin{aligned} C\left(\underline{u}\right) &= C\left(C\left(\underline{u_1}; \psi_1\right), \ldots, C\left(\underline{u_d}; \psi_d\right); \psi_0\right) \\ &= \psi_0\left(\sum_{i=1}^{d} \psi_0^{-1}\left(\psi_i\left(\sum_{j=1}^{n_i} \psi_i^{-1}\left(u_{i,j}\right)\right)\right)\right), \end{aligned} \tag{31}$$

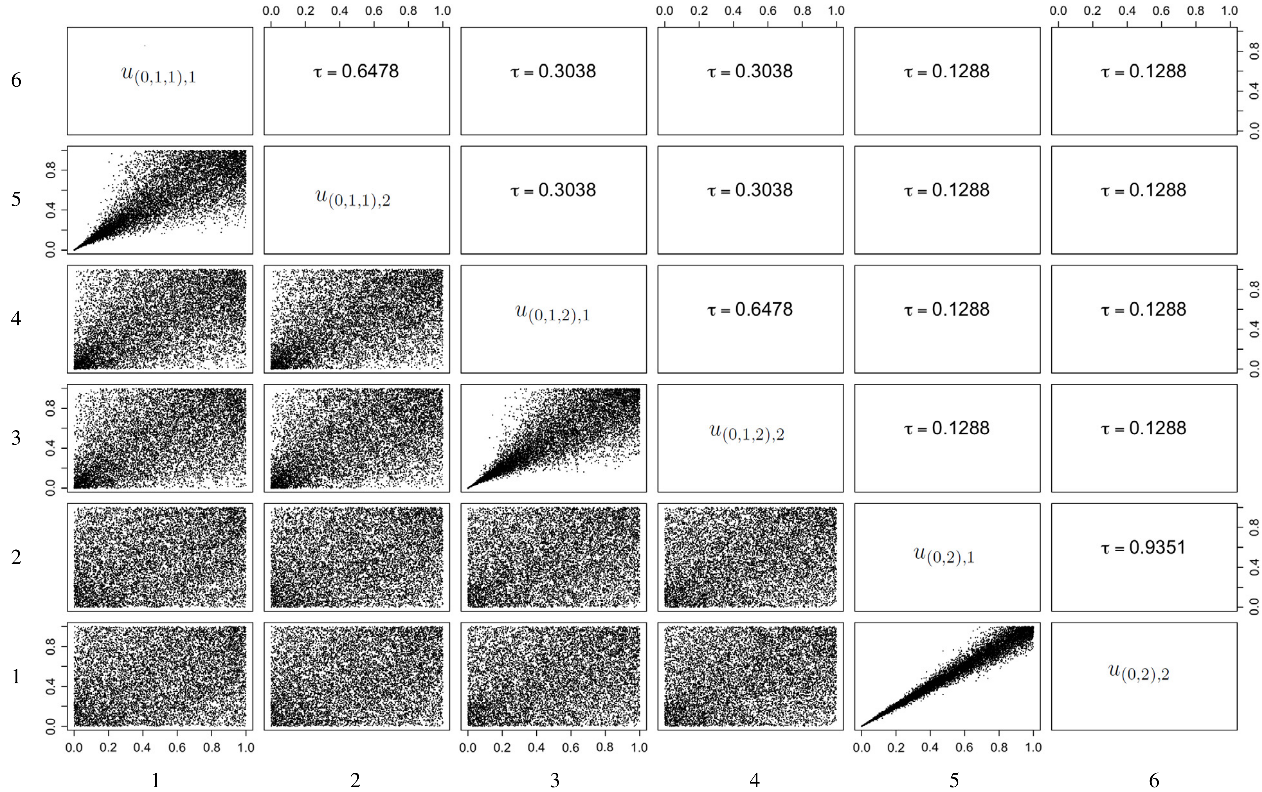

Fig. 5. Pairwise graph of 10 000 6-dimensional vectors of realizations sampled from the hierarchical Archimedean copula with multivariate compound distributions in Example 5.

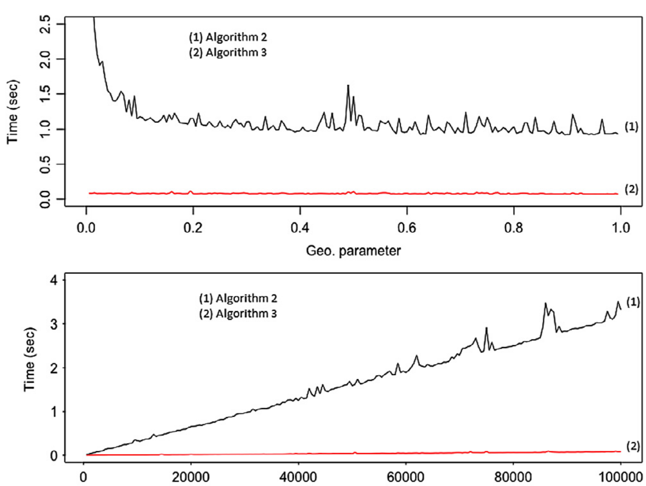

Fig. 6. **Top**: Computation time (sec) for Algorithms 1 and 3 with respect to q. **Bottom**: Computation time (sec) for Algorithms 1 and 3 with respect to the sampling size (n) with $q = 0.01$.

where ψ is the generator of the Archimedean copula C. If we suppose that ψ_0 is the LST of a discrete strictly positive rv R, then, (31) becomes

$$C\left(\underline{u}\right) = \sum_{r=1}^{\infty}\prod_{i=1}^{d} \mathrm{e}^{-r\,\psi_0^{-1}\circ\psi_i\left(\sum_{j=1}^{n_i}\psi_i^{-1}(u_{i,j})\right)} f_R(r)$$
$$= \sum_{r=1}^{\infty}\prod_{i=1}^{d}\psi_{0i}\left(\sum_{j=1}^{n_i}\psi_i^{-1}(u_{i,j});\, r\right) f_R(r), \tag{32}$$

where $\psi_{0i}(t; r) = \exp\left(-r \times \psi_0^{-1} \circ \psi_i(t)\right)$. The representation in (32) is a proper copula if and only if ψ_{0i} is the LST of a positive rv, say W_i, i.e.

$$C\left(\underline{u}\right) = \sum_{r=1}^{\infty}\prod_{i=1}^{d}\left(\int_0^{\infty} e^{-w_i\left(\sum_{j=1}^{n_i}\psi_i^{-1}(u_{i,j})\right)}\mathrm{d}F_{W_i}(w_i)\right) f_R(r). \tag{33}$$

Remark. It is important to highlight that verifying that ψ_{0i} is a LST, and hence that the sufficient nesting condition is verified, is not an easy task. Nesting together different Archimedean families has its limitations and parameter restrictions (see e.g. Hofert, 2010).

We can clearly see that the representation in (30) of the hierarchical Archimedean copula with multivariate compound distributions and the one in (33) of the nested Archimedean copula have the same general form. In order to compare these two representations, we rewrite $\mathcal{L}_{B_i}^m(t)$ as

$$\mathcal{L}_{B_i}^m(t) = e^{-m\mathcal{L}_M^{-1}\circ\mathcal{L}_{\Theta_i}(t)} = \psi_{0i}(t; m). \tag{34}$$

Expressing $\mathcal{L}_{B_i}^m$ as ψ_{0i} allows us to represent the hierarchical Archimedean copula with multivariate compound distributions in (30) as the partially nested Archimedean copula in (32), where the mother copula C_0 is generated from the rv M, and the child copula C_i is generated from Θ_i. Note that C_i is an Archimedean copula derived from $\Theta_i = \sum_{j=1}^{M} B_{i,j}$, which is in general not a well known distribution and expressed in terms of the rv M. Given that $\psi_{0i} = \mathcal{L}_{B_i}^m$, we have that ψ_{0i} solely depends in our structure on the distribution of B_i unlike in the nested Archimedean copulas in which both the mother and the child copulas affect ψ_{0i}, leading to a restrictive nesting condition. Also, since $\mathcal{L}_{B_i}^m$ depends only on the parameters of B_i regardless of the distribution of the rv M, the choice of M's distribution is very flexible in comparison to the nested Archimedean copulas, where nesting different families is a recurrent problem.

The above remarks imply that the hierarchical Archimedean copula C, as defined in (20) can be represented as a nested Archimedean copula as follows:

$$C(\underline{u}) = C(C(\underline{u_1}; \mathcal{L}_{\Theta_1}), \dots, C(\underline{u_d}; \mathcal{L}_{\Theta_d}); \mathcal{L}_M). \tag{35}$$

Since the hierarchical Archimedean copulas proposed in this paper can be written as a nested Archimedean copula, one may wonder if we can produce similar copulas from both approaches. To do so, we have to find the distribution of B_i for which (35) is a known nested Archimedean copula.

Let M be a positive and discrete rv and Θ_i $(i = 1, \dots, d)$ be strictly positive rvs generating known Archimedean copulas. In this case, the distribution of the rvs B_i can be deduced from

$$\mathcal{L}_{B_i}(t) = \mathcal{P}_M^{-1}\left(\mathcal{L}_{\Theta_i}(t)\right) = e^{-\mathcal{L}_M^{-1}\left(\mathcal{L}_{\Theta_i}(t)\right)}. \tag{36}$$

From (34), (36) becomes

$$\mathcal{L}_{B_i}^m(t) = \psi_{01}(t; m), \tag{37}$$

which means that

$$W_i = \sum_{k=1}^{m} B_i,$$

where W_i is the rv generating $\psi_{01}(.; m)$.

To illustrate this idea, let (35) be a nested AMH–AMH copula. To obtain such a copula with our approach, we must have $M \sim Geo(1-\alpha_0)$ and $\Theta_i \sim Geo(1-\alpha_i)$ which implies that $B_i \sim Geo\left(\frac{1-\alpha_i}{1-\alpha_0}\right)$, $\alpha_0 \le \alpha_i$ $(i = 1, \dots, d)$. Similarly, in order for (35) to be a nested Joe–Joe copula, our approach requires that $M \sim Sibuya(\alpha_0)$ and $\Theta_i \sim Sibuya(\alpha_i)$ which implies that $B_i \sim Sibuya\left(\frac{\alpha_i}{\alpha_0}\right)$, $\alpha_i \le \alpha_0$ $(i = 1, \dots, d)$.

The representation in (35) can be easily generalized to multi-level hierarchical structures by adapting the notation presented in Section 2.8. For example, (28) can also be written as

$$\begin{aligned} C(\underline{u_1}, \underline{u_2}, \underline{u_3}) = C\Big(C\left(C\left(\underline{u_1}; \mathcal{L}_{\Theta^{(0,1,1)}}\right), C\left(\underline{u_2}; \mathcal{L}_{\Theta^{(0,1,2)}}\right); \mathcal{L}_{M^{(0,1)}}\right), \\ C\left(\underline{u_3}; \mathcal{L}_{\Theta^{(0,2)}}\right); \mathcal{L}_{M^{(0)}}\Big), \end{aligned} \tag{38}$$

where $\underline{u_1} = (u_{(0,1,1),1}, u_{(0,1,1),2})$, $\underline{u_2} = (u_{(0,1,2),1}, u_{(0,1,2),2})$ and $\underline{u_3} = (u_{(0,2),1}, u_{(0,2),2})$. Notation wise, (38) can be simplified to

$$\begin{aligned} &C(\underline{u_1}, \underline{u_2}, \underline{u_3}) \\ &= C^{(0)}\left(C^{(0,1)}\left(C^{(0,1,1)}\left(\underline{u_1}\right), C^{(0,1,2)}\left(\underline{u_2}\right)\right), C^{(0,2)}\left(\underline{u_3}\right)\right). \end{aligned} \tag{39}$$

4.2. Approach based on Lévy subordinators and random time

In this section, we examine the link between the approach proposed by Hering et al. (2010) to construct hierarchical Archimedean copulas and the one described within this paper. The construction approach by Hering et al. (2010) is based on increasing Lévy processes, also called Lévy subordinators. They are notably important components for building models in financial mathematics (see e.g. Cont and Tankov, 2003).

Let $\underline{\Lambda}^{(i)} = \left\{\Lambda^{(i)}(t), t \ge 0\right\}$ be a Lévy subordinator, for $i = 1, 2, \dots, d$. For all $x, t \ge 0$, the LST of $\Lambda^{(i)}(t)$ is given by

$$\mathcal{L}_{\Lambda^{(i)}(t)}(x) = E\left[e^{-x\Lambda^{(i)}(t)}\right] = e^{-t\zeta_i(x)},$$

where the function ζ_i corresponds to the Laplace exponent of $\underline{\Lambda}^{(i)}$, for $i = 1, 2, \dots, d$.

Let a strictly positive rv V be a random time with LST defined by

$$\mathcal{L}_V(x) = E\left[e^{-Vx}\right] = \psi_0(x),$$

for $x \ge 0$.

Table 1 of Hering et al. (2010) provides a short list of possible distributions for V, which can be either discrete or continuous. The random time rv V is used as a common factor to define the vector of rvs

$$\underline{\Lambda}(V) = \left(\Lambda^{(1)}(V), \dots, \Lambda^{(d)}(V)\right).$$

Given V, the rvs $\Lambda^{(i)}(V)$, for $i = 1, \dots, d$, are assumed to be conditionally independent. This implies that the multivariate LST of $\underline{\Lambda}(V)$ is obtained as follows :

$$\begin{aligned} \mathcal{L}_{\underline{\Lambda}(V)}(x_1, \dots, x_d) &= E\left[e^{-x_1\Lambda^{(1)}(V)} \times \dots \times e^{-x_d\Lambda^{(d)}(V)}\right] \\ &= E_V\left[E\left[e^{-x_1\Lambda^{(1)}(V)} \times \dots \times e^{-x_d\Lambda^{(d)}(V)}|V\right]\right] \\ &= E\left[e^{-V\zeta_1(x_1)} \times \dots \times e^{-V\zeta_d(x_d)}\right] \\ &= E\left[e^{-V(\zeta_1(x_1)+\dots+\zeta_d(x_d))}\right] \\ &= \mathcal{L}_V\left(\zeta_1(x_1) + \dots + \zeta_d(x_d)\right). \end{aligned} \tag{40}$$

Following a similar procedure to the one presented in Section 2.2, Hering et al. (2010) propose to construct multivariate mixed exponential distributions using $\underline{\Lambda}(V)$. Let

$$\underline{Y} = \left(Y_{1,1}, \dots, Y_{1,n_1}, \dots, Y_{d,1}, \dots, Y_{d,n_d}\right)$$

be a vector of $n_1 + \dots + n_d$ rvs. Given $\underline{\Lambda}(V) = \underline{\lambda}$, it is assumed that

$$\begin{aligned} &\left(Y_{1,1}|\underline{\Lambda}(V) = \underline{\lambda}\right), \dots, \left(Y_{1,n_1}|\underline{\Lambda}(V) = \underline{\lambda}\right), \dots, \\ &\left(Y_{d,1}|\underline{\Lambda}(V) = \underline{\lambda}\right), \dots, \left(Y_{d,n_d}|\underline{\Lambda}(V) = \underline{\lambda}\right) \end{aligned}$$

are conditionally independent. Also, given $\underline{\Lambda}(V) = \underline{\lambda} = (\lambda_1, \dots, \lambda_d)$, the rv $\left(Y_{i,j}|\underline{\Lambda}(V) = \underline{\lambda}\right)$ is identically distributed as $\left(Y_{i,j}|\Lambda^{(i)}(V) = \lambda_i\right)$ for $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, d$. We assume that $\left(Y_{i,1}|\Lambda^{(i)}(V) = \lambda_i\right), \dots, \left(Y_{i,n_i}|\Lambda^{(i)}(V) = \lambda_i\right)$ are exponentially distributed with parameter λ_i, for $i = 1, \dots, d$. Then, $\underline{Y}$ follows a multivariate mixed exponential distribution with multivariate survival function given by

$$\begin{aligned} &\overline{F}_{\underline{Y}}\left(y_{1,1}, \dots, y_{1,n_1}, \dots, y_{d,1}, \dots, y_{d,n_d}\right) \\ &= E\Big[e^{-\Lambda^{(1)}(V)y_{1,1}} \times \dots \times e^{-\Lambda^{(1)}(V)y_{1,n_1}} \times \dots \\ &\quad \times e^{-\Lambda^{(d)}(V)y_{d,1}} \times \dots \times e^{-\Lambda^{(d)}(V)y_{d,n_d}}\Big]. \end{aligned} \tag{41}$$

Due to (40), we can write (41) as

$$\begin{aligned} &\overline{F}_{\underline{Y}}\left(y_{1,1}, \dots, y_{1,n_1}, \dots, y_{d,1}, \dots, y_{d,n_d}\right) \\ &= \mathcal{L}_{\underline{\Lambda}(V)}\left(y_{1,1} + \dots + y_{1,n_1}, \dots, y_{d,1} + \dots + y_{d,n_d}\right) \\ &= \psi_0\left(\zeta_1\left(y_{1,1} + \dots + y_{1,n_1}\right) + \dots + \zeta_d\left(y_{d,1} + \dots + y_{d,n_d}\right)\right). \end{aligned} \tag{42}$$

Also, the univariate survival function of $Y_{i,j}$ is

$$\begin{aligned} \overline{F}_{Y_{i,j}}\left(y_{i,j}\right) &= E\left[e^{-\Lambda^{(i)}(V)y_{i,j}}\right] = \mathcal{L}_{\Lambda^{(i)}(V)}\left(y_{i,j}\right) = \mathcal{L}_V\left(\zeta_i\left(y_{i,j}\right)\right) \\ &= \psi_0\left(\zeta_i\left(y_{i,j}\right)\right) = \psi_0 \circ \zeta_i\left(y_{i,j}\right), \end{aligned} \tag{43}$$

for $y_{i,j} \ge 0$, $j = 1, 2, \dots, n_i$, and $i = 1, \dots, d$.

Then, using Sklar's Theorem, the copula associated to $\overline{F}_{\underline{Y}}$ is given by

$$\begin{aligned} C\left(\underline{u}\right) = \overline{F}_{\underline{Y}}\Big(&\overline{F}_{Y_{1,1}}^{-1}\left(u_{1,1}\right), \dots, \overline{F}_{Y_{1,n_1}}^{-1}\left(u_{1,n_1}\right), \dots, \\ &\overline{F}_{Y_{d,1}}^{-1}\left(u_{d,1}\right), \dots, \overline{F}_{Y_{d,n_d}}^{-1}\left(u_{d,n_d}\right)\Big), \end{aligned} \tag{44}$$

where $\overline{F}_{Y_{i,j}}^{-1}$ is obtained from (43) i.e.

$$\overline{F}_{Y_{i,j}}^{-1}\left(u_{i,j}\right) = \mathcal{L}_{\Lambda^{(i)}(V)}^{-1}\left(u_{i,j}\right) = \zeta_i^{-1}\left(\psi_0^{-1}\left(u_{i,j}\right)\right) = \zeta_i^{-1} \circ \psi_0^{-1}\left(u_{i,j}\right), \tag{45}$$

for $u_{i,j} \in [0, 1], j = 1, 2, \ldots, n_i$, and $i = 1, \ldots, d$. Combining (44) and (45) with (42), the expression for $C(\underline{u})$ becomes

$$C(\underline{u}) = \mathcal{L}_{\underline{\Lambda}(V)}\left(\mathcal{L}^{-1}_{\Lambda^{(1)}(V)}(u_{1,1}) + \cdots + \mathcal{L}^{-1}_{\Lambda^{(1)}(V)}(u_{1,n_1}), \ldots, \mathcal{L}^{-1}_{\Lambda^{(d)}(V)}(u_{d,1}) + \cdots + \mathcal{L}^{-1}_{\Lambda^{(d)}(V)}(u_{1,n_d})\right) = \psi_0\left(\sum_{i=1}^{d} \zeta_i\left(\zeta_i^{-1} \circ \psi_0^{-1}(u_{i,1}) + \cdots + \zeta_i^{-1} \circ \psi_0^{-1}(u_{i,n_i})\right)\right).$$

In Hering et al. (2010), conditions for admissible Lévy subordinators are given in Theorem 2.1 and a list of popular Lévy subordinators can be found in Table 2.

We may now compare both approaches and the resulting hierarchical Archimedean copulas. Under both approaches, a hierarchical Archimedean copula is identified from the joint survival function of a multivariate mixed exponential distribution. In both cases, it leads to natural generic sampling algorithms for hierarchical Archimedean copulas. The dependence structure of the multivariate mixed exponential distribution is defined through a vector of dependent mixing rvs,

$$\underline{\Lambda}(V) = \left(\Lambda^{(1)}(V), \ldots, \Lambda^{(d)}(V)\right)$$

in Hering et al. (2010), and through

$$\underline{\Theta} = (\Theta_1, \ldots, \Theta_d)$$

within this paper. Even if these two vectors of rvs are defined through different probabilistic arguments (Lévy subordinators with a common random time and random sums with a common counting rv), both approaches lead to identical generic structures for $\mathcal{L}_{\underline{\Lambda}(V)}$ and $\mathcal{L}_{\underline{\Theta}}$ and for $\mathcal{L}_{\Lambda^{(i)}(V)}$ and $\mathcal{L}_{\Theta_i}$. The expression for $\mathcal{L}_{\Lambda^{(i)}(V)}$ is

$$\mathcal{L}_{\Lambda^{(i)}(V)}(x) = \mathcal{L}_V(\zeta_i(x)),$$

for $i = 1, \ldots, d$. Let us define the function υ_i associated to $\mathcal{L}_{B_i}$ such that

$$\mathcal{L}_{B_i}(x) = \exp(-\upsilon_i(x)),$$

or, equivalently,

$$\upsilon_i(x) = -\ln\left(\mathcal{L}_{B_i}(x)\right),$$

for $i = 1, 2, \ldots, d$. Then, the expression for $\mathcal{L}_{\Theta_i}$ becomes

$$\mathcal{L}_{\Theta_i}(x) = \mathcal{P}_M\left(\mathcal{L}_{B_i}(x)\right) = \mathcal{L}_M\left(-\ln \mathcal{L}_{B_i}(x)\right) = \mathcal{L}_M(\upsilon_i(x)),$$

for $i = 1, \ldots, d$. Note that

$$\mathcal{L}_{\underline{\Lambda}(V)}(x_1, \ldots, x_d) = \mathcal{L}_V(\zeta_1(x_1) + \cdots + \zeta_d(x_d))$$

and

$$\mathcal{L}_{\underline{\Theta}}(x_1, \ldots, x_d) = \mathcal{L}_M(\upsilon_1(x_1) + \cdots + \upsilon_d(x_d)).$$

The rvs V and M play similar roles as a common factor. As mentioned earlier, the rv V can be either discrete or continuous. Under the approach proposed within this paper, M is a strictly positive discrete rv. If V is a strictly positive discrete rv, than $\mathcal{L}_V$ and $\mathcal{L}_M$ are identical. Also, both $\mathcal{L}_V$ and $\mathcal{L}_M$ correspond to the generator ψ_0 in the definition of a nested Archimedean copula.

Finally, both υ_i and ζ_i correspond to the function $\psi_0^{-1} \circ \psi_1$ which can be found in the definition of the nested Archimedean copula and for which it is difficult to verify the conditions of admissibility. In conclusion, the approaches proposed both in Hering et al. (2010) and within the present paper circumvent this difficulty. Note that υ_i plays the same role as ζ_i (the Laplace exponent of the Lévy subordinator $\underline{\Lambda}^{(i)}$).

The construction method provided in this paper and Hering et al. (2010)'s approach have their own advantages even if their strategies appear similar. Both approaches can lead to the same hierarchical Archimedean copula in some cases. However, others can only be obtained with one of the two methods. For all the popular L évy subordinators provided in Table 2 of Hering et al. (2010), it is possible to define the distribution of B_i such that its Laplace exponent υ_i is equal to the Laplace exponent ζ_i of $\underline{\Lambda}^{(i)}$, $i = 1, 2, \ldots, d$. For example, if B_i follows a gamma distribution with parameters β and 1, then $\upsilon_i(x) = \beta \ln(1 + x)$ as in (ii) of Table 2 (η is a scaling factor) then the expressions for $\mathcal{L}_{\Lambda^{(i)}(V)}$ and $\mathcal{L}_{\Theta_i}$ are identical (assuming $\mathcal{L}_V = \mathcal{L}_M$). If B_i follows an inverse Gaussian with parameters $\frac{1}{\eta}$ and 1, $\upsilon_i(x) = \left(\sqrt{\frac{2x}{\eta^2} + 1} - 1\right)$ as in (iii) of Table 2 (β is a scaling factor) then it follows that $\mathcal{L}_{\Lambda^{(i)}(V)}$ and $\mathcal{L}_{\Theta_i}$ have the same expression (assuming $\mathcal{L}_V = \mathcal{L}_M$). Also, under the assumption $\mathcal{L}_V = \mathcal{L}_M$ (i.e. V is a discrete rv) and since the rv B_i can follow many distributions including the ones listed in Table 1 of Hering et al. (2010), the proposed method provides a larger range of possible hierarchical Archimedean copulas in comparison to Hering et al. (2010)'s construction method. If V is a continuous rv, the resulting hierarchical Archimedean copula obtained with Hering et al. (2010)'s approach cannot be obtained with our approach.

Finally, inspired from Hering et al. (2010), the expression of the hierarchical Archimedean copula constructed under our proposed approach given in (20) becomes

$$C(\underline{u}) = \mathcal{L}_{\underline{\Theta}}\left(\mathcal{L}^{-1}_{\Theta_1}(u_{1,1}) + \cdots + \mathcal{L}^{-1}_{\Theta_1}(u_{1,n_1}), \ldots, \mathcal{L}^{-1}_{\Theta_d}(u_{d,1}) + \cdots + \mathcal{L}^{-1}_{\Theta_d}(u_{d,n_d})\right) = \mathcal{L}_M\left(\sum_{i=1}^{d} \upsilon_i\left(\upsilon_i^{-1} \circ \mathcal{L}_M^{-1}(u_{i,1}) + \cdots + \upsilon_i^{-1} \circ \mathcal{L}_M^{-1}(u_{i,n_i})\right)\right).$$

4.3. Estimation procedure and determination of the tree structure

The main objective of the present work is to propose an alternative approach to construct hierarchical Archimedean copulas. The interestingness of these new copulas also relies on the capability to estimate them. Recently, different research works have appeared in the literature on the estimation of hierarchical Archimedean copulas, more specifically on the structure determination and the parameter estimation. Okhrin et al. (2013) appears to be the first paper to address simultaneously both of these tasks through a multi-stage procedure in a bottom-up manner. Maximum-likelihood estimation is used for the parameters and the inversion of Kendall's tau is also suggested. The structure determination is investigated in several ways, with notably approaches based on a goodness-of-fit test or binary trees. In Górecki et al. (2016a) and Górecki and Holeňa (2013), the estimator for both the structure and the parameters of hierarchical Archimedean copulas is based on Kendall's tau, more precisely on the inversion of Kendall's tau estimator. They use agglomerative hierarchical clustering (with three different definitions for the distance between clusters) in which the relationship between two rvs is established through their Kendall's tau. Other research papers, such as Uyttendaele (2016) and Segers and Uyttendaele (2014) estimate the structure differently by considering either trivariate structures or supertrees.

These works mainly focus on hierarchical Archimedean copulas in which all copulas in the tree are from the same Archimedean family. An excellent paper, Górecki et al. (2016b), paves the way to the estimation of the structure and the parameters of hierarchical Archimedean copulas involving different Archimedean families. This new estimation procedure, which is in part based on goodness-of-fit tests, adopts estimation algorithms from previous papers in a way that guarantees the verification of the nesting condition. In the same context of different family generators, Zhu et al. (2016) recently proposed an approach using a three stage estimation approach based on a clustering procedure to choose the optimal hierarchical structure as an estimation of a Lévy subordinated hierarchical Archimedean copula.

Given the similarities of our construction with the one of Hering et al. (2010) discussed in Section 4.2, the approach suggested in Zhu et al. (2016) is evidently a good starting point for the estimation of our structure based on maximum likelihood. This should be compared with the method proposed in Górecki et al. (2016b), based on Kendall's tau estimator, which allows the use of Archimedean copulas from different copula families as the proposed construction here. Contrarily to the nested Archimedean copulas approach, we will not be confronted with the difficulties surrounding the nesting condition. However, the flexibility of our new structure leads to new challenges in regard to finding the optimal solution among all possible structures. This mainly concerns for us, firstly, the choice of the appropriate LST of M and then the choice of the LSTs of the rvs B_i ($i = 1, \ldots, d$). This will be followed by the estimation of the parameters of the hierarchical Archimedean copulas. These questions are at the core of another ongoing research project.

5. Aggregation method

Risk aggregation comes into play for tasks such as the analysis of a risk portfolio and regularly capital calculations within and between risk categories. In this section, we examine the behavior of aggregated dependent risks with multivariate cdf defined through a hierarchical Archimedean copula based on multivariate compound distributions.

Let $\underline{X} = (X_{1,1}, \ldots, X_{1,n_1}, \ldots, X_{d,1}, \ldots, X_{d,n_d})$ be a random vector with the following multivariate cdf defined with an underlying hierarchical Archimedean copula as given in (21) and marginals $F_{X_{i,j}}$ (for $i = 1, \ldots, d$ and $j = 1, \ldots, n_i$)

$$\begin{aligned} &F_{\underline{X}}\left(x_{1,1}, \ldots, x_{1,n_1}, \ldots, x_{d,1}, \ldots, x_{d,n_d}\right) \\ &= C\left(F_{X_{1,1}}\left(x_{1,1}\right), \ldots, F_{X_{1,n_1}}\left(x_{1,n_1}\right), \ldots,\right. \\ &\quad \left. F_{X_{d,1}}\left(x_{d,1}\right), \ldots, F_{X_{d,n_d}}\left(x_{d,n_d}\right)\right) \\ &= \mathcal{L}_{\underline{\Theta}}\left(\sum_{i=1}^{n_1} \mathcal{L}_{\Theta_1}^{-1}\left(F_{X_{1,i}}\left(x_{1,i}\right)\right), \ldots, \sum_{j=1}^{n_d} \mathcal{L}_{\Theta_d}^{-1}\left(F_{X_{d,j}}\left(x_{d,j}\right)\right)\right). \end{aligned} \tag{46}$$

Since the vectors of rvs $(X_{i,1} \mid \Theta_i = \theta_i), \ldots, (X_{i,n_i} \mid \Theta_i = \theta_i)$, for $i \in 1, \ldots, d$, are conditionally independent given $\Theta_i = \theta_i$ and $(X_{1,j_1} \mid M = m), \ldots, (X_{d,j_d} \mid M = m)$, for $j_i \in 1, \ldots, n_i$, are also conditionally independent given $M = m$, $F_{\underline{X}}$ can be represented as a common mixture of conditional cdfs as

$$\begin{aligned} &F_{\underline{X}}\left(x_{1,1}, \ldots, x_{1,n_1}, \ldots, x_{d,1}, \ldots, x_{d,n_d}\right) \\ &= \sum_{m=1}^{\infty} F_{\underline{X}|M=m}\left(x_{1,1}, \ldots, x_{1,n_1}, \ldots, x_{d,1}, \ldots, x_{d,n_d}\right) f_M(m) \\ &= \sum_{m=1}^{\infty} \prod_{i=1}^{d}\left(\int_0^{\infty} \prod_{j=1}^{n_i} F_{X_{i,j}|M=m,\Theta_i=\theta_i}\left(x_{i,j}\right) \mathrm{d}F\Theta_i(\theta_i)\right) f_M(m). \end{aligned} \tag{47}$$

To find the expression of the conditional cdfs $F_{X_{i,j}|M=m,\Theta_i=\theta_i}$, we use the common mixture representation given in (30). We obtain

$$\begin{aligned} &F_{\underline{X}}\left(x_{1,1}, \ldots, x_{d,n_d}\right) \\ &= \sum_{m=1}^{\infty} \prod_{i=1}^{d}\Big(\int_0^{\infty} e^{-v_i\left(\sum_{j=1}^{n_i} \mathcal{L}_{\Theta_i}^{-1}\left(F_{X_{i,j}}\left(x_{i,j}\right)\right)\right)} dF_{V_i}(v_i)\Big) f_M(m). \end{aligned} \tag{48}$$

Given (47) and (48), we have $F_{X_{i,j}|M=m,V_i=v_i}\left(x_{i,j}\right) = e^{-v_i\left(\mathcal{L}_{\Theta_i}^{-1}\left(F_{X_{i,j}}\left(x_{i,j}\right)\right)\right)}$, for $i = 1, \ldots, d \geq 2$ and $j = 1, \ldots, n_i$.

Table 2
Values of the variance, VaR and TVaR of $S = X_{1,1} + \cdots + X_{1,40} + X_{2,1} + \cdots + X_{2,40}$ where the joint cdf $F_{X_{1,1},\ldots,X_{1,40},X_{2,1},\ldots,X_{2,40}}$ is as defined in Example 6.

Risk measures	Cossette et al. (2017)	Sampling (1M)
$Var(S)$	1157.4461	1156.8011
$VaR_{0.9}(S)$	193.0000	193.0000
$TVaR_{0.9}(S)$	214.4829	214.4334
$VaR_{0.99}(S)$	240.0000	240.0000
$TVaR_{0.99}(S)$	252.1244	252.2950
$VaR_{0.999}(S)$	267.0000	268.0000
$TVaR_{0.999}(S)$	276.1494	276.4880
$VaR_{0.9999}(S)$	287.0000	288.0000
$TVaR_{0.9999}(S)$	293.5822	294.8900

Let $S = \sum_{i=1}^{d} S_i$, where $S_i = X_{i,1} + \cdots + X_{i,n_i}$, for $i = 1, 2, \ldots, d$. To find the desired cdf of S, one may apply the approach proposed in Cossette et al. (2017). To do so however, we must assume V_i to be a strictly positive discrete rv defined on $\{1, 2, \ldots\}$, with probability mass function $f_{V_i}(v_i) = \Pr(V_i = v_i)$, $V_i = 1, 2, \ldots$ In such a case, (47) becomes

$$\begin{aligned} &F_{\underline{X}}\left(x_{1,1}, \ldots, x_{d,n_d}\right) \\ &= \sum_{m=1}^{\infty} \prod_{i=1}^{d}\left(\sum_{v_i=1}^{\infty}\Big(\prod_{j=1}^{n_i} F_{X_{i,j}|M=m,V_i=v_i}\left(x_{i,j}\right)\Big) f_{V_i}(v_i)\right) f_M(m). \end{aligned}$$

We provide next an example to illustrate the applicability of the aggregation procedure proposed in Cossette et al. (submitted for publication) in the context of the new construction given in this paper. Since the method of Cossette et al. (submitted for publication) leads to exact results, the following example also shows the precision of the values obtained with Algorithm 3.

Example 6. Consider a portfolio of 80 risks $\underline{X} = \big(X_{1,1}, \ldots, X_{1,40}, X_{2,1}, \ldots, X_{2,40}\big)$ with a multivariate cdf defined as in (46) with $d = 2$ and $n_1 = n_2 = 40$. We assume $M \sim Logarithmic(0.5)$, $B_1 \sim Geo(0.8)$ and $B_2 \sim Geo(0.9)$. Let $X_{s,i} \sim Binom\left(10, q_{s,i}\right)$ where $q_{s,i} = 0.05 \times s + 0.005i$, $s = 1, 2$ and $i = 1, 2, \ldots, 40$. It implies that $E[S] = 142$. Relevant measures of $S = \sum_{s=1}^{2}\sum_{i=1}^{40} X_{s,i}$ can be obtained with the approach of Cossette et al. (submitted for publication) or with MC simulations using Algorithm 3. Note that the distributions of M and B_i are not from the same families. Both risk measures, VaR and $TVaR$, are given in Table 2 to illustrate relevant results for both methods as a mean of comparison. Note that the approach of Cossette et al. (submitted for publication) always provides exact results for discrete marginals $F_{X_{i,j}}$. Moreover, the simulation results are very close to the actual results, considering 1 million simulations were done.

6. Conclusion

A new hierarchical Archimedean copulas construction method involving multivariate compound distributions was presented. The absence of nesting and marginal conditions enlarges the possibilities of nested copulas by improving the flexibility in both the choice of families and parameters. Moreover, new copulas with multiple parameters were derived allowing for a large variety of dependence structures. In addition to the new parametric copulas, well-known Archimedean copulas can be obtained as special cases. Closed form expressions have been derived for the copulas. To further complement our theoretical results, efficient sampling algorithms have been developed for computational applications.

Acknowledgments

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (Cossette: 054993;

Marceau: 053934) and by the Chaire en actuariat de l'Université Laval (Cossette, Gadoury, Marceau, and Mtalai: FO502323). Helpful comments by two anonymous referees on earlier versions of the manuscript have been highly appreciated.

References

Bedford, T., Cooke, R.M., 2002. Vines: A new graphical model for dependent random variables. Ann. Statist. 1031–1068.

Brechmann, E.C., 2014. Hierarchical kendall copulas: Properties and inference. Canad. J. Statist. 42 (1), 78–108.

Cont, R., Tankov, P., 2003. Financial Modelling with Jump Processes, vol. 2, 2003. Chapman & Hall/CRC, London.

Cossette, H., Marceau, E., Mtalai, I., Veilleux, D., 2017. Archimedean copulas: aggregation, capital allocation and other applications. Available At http://Ssrn.Com/Abstract=2957181.

Devroye, L., 2009. Random variate generation for exponentially and polynomially tilted stable distributions. ACM Trans Modeling Comput. Simul. 19 (4), 18.

Feller, W., 1971. An Introduction to Probability and Its Applications, Vol. II. Wiley, New York.

Górecki, J., Hofert, M., Holeňa, M., 2016a. An approach to structure determination and estimation of hierarchical archimedean copulas and its application to bayesian classification. J. Intell. Inf. Syst. 46 (1), 21–59.

Górecki, J., Hofert, M., Holeňa, M., 2016b. On structure, family and parameter estimation of hierarchical archimedean copulas. ArXiv Preprint arXiv:1611.09225.

Górecki, J., Holeňa, M., 2013. Structure determination and estimation of hierarchical archimedean copulas based on kendall correlation matrix. In: International Workshop on New Frontiers in Mining Complex Patterns. Springer, pp. 132–147.

Hering, C., Hofert, M., Mai, J.-F., Scherer, M., 2010. Constructing hierarchical archimedean copulas with lévy subordinators. J. Multivariate Anal. 101 (6), 1428–1433.

Hofert, M., 2010. Sampling Nested Archimedean Copulas with Applications to cdo Pricing. (Ph.D. Thesis), Universität Ulm.

Hofert, M., 2011. Efficiently sampling nested Archimedean copulas. Comput. Statist. Data Anal. 55 (1), 57–70.

Hofert, M., 2012. A stochastic representation and sampling algorithm for nested archimedean copulas. J. Stat. Comput. Simul. 82 (9), 1239–1255.

Joe, H., 1997. Multivariate Models and Multivariate Dependence Concepts. CRC Press.

Joe, H., 2014. Dependence Modeling with Copulas. CRC Press.

Kimberling, C.H., 1974. A probabilistic interpretation of complete monotonicity. Aequationes Math. 10 (2), 152–164.

Marshall, A.W., Olkin, I., 1988. Families of multivariate distributions. J. Amer. Statist. Assoc. 83 (403), 834–841.

McNeil, A.J., 2008. Sampling nested archimedean copulas. J. Stat. Comput. Simul. 78 (6), 567–581.

Nelsen, R.B., 2007. An Introduction to Copulas. Springer Science & Business Media.

Okhrin, O., Okhrin, Y., Schmid, W., 2013. On the structure and estimation of hierarchical archimedean copulas. J. Econometrics 173 (2), 189–204.

Segers, J., Uyttendaele, N., 2014. Nonparametric estimation of the tree structure of a nested archimedean copula. Comput. Statist. Data Anal. 72, 190–204.

Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Université Paris 8.

Uyttendaele, N., 2016. On the estimation of nested archimedean copulas: A theoretical and an experimental comparison.

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