Computational Statistics and Data Analysis 72 (2014) 190–204

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Nonparametric estimation of the tree structure of a nested Archimedean copula

Johan Segers [1], Nathan Uyttendaele [*,1]

Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

A R T I C L E I N F O

Article history:
Received 29 March 2013
Received in revised form 31 October 2013
Accepted 31 October 2013
Available online 14 November 2013

Keywords:
Archimedean copula
Dependence
Nested Archimedean copula
Hierarchical Archimedean copula
Rooted tree
Subtree
Kendall distribution
Fan
Triple
Nonparametric inference

A B S T R A C T

One of the features inherent in nested Archimedean copulas, also called hierarchical Archimedean copulas, is their rooted tree structure. A nonparametric, rank-based method to estimate this structure is presented. The idea is to represent the target structure as a set of trivariate structures, each of which can be estimated individually with ease. Indeed, for any three variables there are only four possible rooted tree structures and, based on a sample, a choice can be made by performing comparisons between the three bivariate margins of the empirical distribution of the three variables. The set of estimated trivariate structures can then be used to build an estimate of the target structure. The advantage of this estimation method is that it does not require any parametric assumptions concerning the generator functions at the nodes of the tree.

1. Introduction

Archimedean copulas have become a popular tool for modeling or simulating bivariate data. They are however not useful for every application, failing for instance to properly model in high dimensions if the data do not exhibit symmetric dependences. Nested Archimedean copulas (NACs), or hierarchical Archimedean copulas, are an interesting attempt to overcome this problem. They were first introduced by Joe (1997, pp. 87–89) and then have been studied extensively, see for instance McNeil (2008), Hofert (2008, 2010) or Hofert (2011) for sampling algorithms; Hofert and Maechler (2011), who released the first R package devoted to NACs; Hering et al. (2010), who investigated the construction of NACs with Lévy subordinators; Hofert and Pham (2012), who examined their densities; Okhrin et al. (2013a), who were the first to investigate likelihood-based estimation; or Okhrin et al. (2013b), who studied tail dependence properties of NACs.

The hierarchy of variables in a nested Archimedean copula is described through a rooted tree. Most often, the tree is given from the context; see for instance Hofert (2010) or Puzanova (2011). Okhrin et al. (2013a) were the first to address the issue of reconstructing the tree from a sample, offering a parametric answer in which each generator is assumed to be known up to some Euclidean parameter(s). In contrast, the method we propose is completely nonparametric since it does not require the user to make any assumption about the generators of the NAC from which the tree structure must

* Corresponding author. Tel.: +32 497739197.
E-mail addresses: johan.segers@uclouvain.be (J. Segers), nathan.uyttendaele@uclouvain.be (N. Uyttendaele).

1 Order of contributions not necessarily reflected by alphabetical order.

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2013.10.028

be estimated, except for a rather straightforward identifiability condition introduced in Section 4. Although never formally mentioned, this identifiability condition is assumed throughout Okhrin et al. (2013a) as well.

Sections 2 and 3 of this paper review the basics of Archimedean copulas and nested Archimedean copulas. Section 5 shows how the structure of three variables (X_i, X_j, X_k) can be estimated nonparametrically. The idea is to estimate the Kendall distribution associated with each pair of variables within (X_i, X_j, X_k); these estimates allow us then to decide if all pairs of variables have actually the same underlying bivariate distribution or not. If so, then the tree structure of (X_i, X_j, X_k) is the trivial trivariate structure, that is, a structure with one internal vertex and three leaves, also called a 3-fan. If not, determining which pair has a different underlying bivariate distribution allows one to assign the correct tree structure to (X_i, X_j, X_k).

Section 6 introduces a key point, namely that a given tree structure λ can always be represented as a set of trivariate structures. That is, for a random vector of continuous random variables $(X_1, \ldots, X_d)$ with a nested Archimedean copula, it is possible to obtain the tree structure of this nested Archimedean copula provided the tree structure of the nested Archimedean copula associated with any three variables (X_i, X_j, X_k) with distinct $i, j, k \in \{1, \ldots, d\}$ is known. A very similar result was obtained by Ng and Wormald (1996), who showed that a given structure λ can always be represented as a set of *triples* and *fans*, triples and fans being formally defined in Section 6. Another interesting result is offered by Okhrin et al. (2013b) who showed that the structure can be retrieved from the bivariate margins of the nested Archimedean copula.

Our suggestion to estimate the structure of $(X_1, \ldots, X_d)$ is first to estimate the tree structure of the nested Archimedean copula associated with any three variables (X_i, X_j, X_k) with distinct $i, j, k \in \{1, \ldots, d\}$, and second to use this set of estimated trivariate structures to build an estimate of the structure of $(X_1, \ldots, X_d)$. This suggestion and one important related difficulty make up Section 7.

The performance of our approach is then assessed by means of a simulation study involving target structures in several dimensions (Section 8). As part of this simulation study, the performance of the approach used by Okhrin et al. (2013a) is also investigated.

Finally, Section 9 illustrates how our method could be used to highlight hierarchical interactions in the stock market. Some remaining challenges are outlined in Section 10.

2. Archimedean copulas

Let $(X_1, \ldots, X_d)$ be a vector of continuous random variables. The copula of this vector is defined as

$$C(u_1, \ldots, u_d) = P(U_1 \le u_1, \ldots, U_d \le u_d),$$

where $(U_1, \ldots, U_d) = (F_{X_1}(X_1), \ldots, F_{X_d}(X_d))$, and where $F_{X_1}, \ldots, F_{X_d}$ are the marginal cumulative distribution functions (CDFs) of $X_1, \ldots, X_d$, respectively.

Archimedean copulas are a class of copulas that admit the representation

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

where ψ is called the generator and ψ^{-1} is its generalized inverse, with $\psi : [0, \infty) \to [0, 1]$, a convex, decreasing function such that $\psi(0) = 1$ and $\psi(\infty) = 0$. In order for C to be a d-dimensional copula, the generator is required to be d-monotone on $[0, \infty)$, see McNeil and Nešlehová (2009) for details.

The generators in Table 1 are among the most popular ones. All of them are completely monotone, that is, d-monotone for all integer $d \ge 2$. For the Frank family, $D_1(\theta) = \frac{1}{\theta}\int_0^\theta t/(\exp(t) - 1)\,dt$.

The parameter θ in Table 1 allows one to control the strength of the dependence between any two variables of the related Archimedean copula. This is best understood by expressing Kendall's τ coefficient between any two variables of the related Archimedean copula in terms of θ (Hofert and Maechler, 2011), as done in the last column of Table 1.

All margins of the same dimension of an Archimedean copula are equal, that is, for all $m \in \{2, \ldots, d\}$ and for every subset $\{i_1, \ldots, i_m\}$ of $\{1, \ldots, d\}$ having m elements, the two vectors

$$(U_{i_1}, \ldots, U_{i_m}) \quad \text{and} \quad (U_1, \ldots, U_m)$$

have the same distribution. This result stems from the fact that for Archimedean copulas, $C(u_1, \ldots, u_d)$ is a symmetric function in its arguments and this is why Archimedean copulas are sometimes also called *exchangeable*. For modeling purposes, this exchangeability becomes an increasingly strong assumption as the dimension grows.

Table 1
Some popular generators of Archimedean copulas.

Name	Generator $\psi(x)$	θ	τ
AMH	$(1-\theta)/(e^x-\theta)$	$\theta \in [0, 1)$	$1 - 2\left(\theta + (1-\theta)^2 \log(1-\theta)\right)/(3\theta^2)$
Clayton	$(1+x)^{-1/\theta}$	$\theta \in (0, \infty)$	$\theta/(\theta+2)$
Frank	$-\log(1-(1-e^{-\theta})e^{-x})/\theta$	$\theta \in (0, \infty)$	$1 + 4(D_1(\theta) - 1)/\theta$
Gumbel	$\exp(-x^{1/\theta})$	$\theta \in [1, \infty)$	$(\theta-1)/\theta$
Joe	$1-(1-e^{-x})^{1/\theta}$	$\theta \in [1, \infty)$	$1 - 4\sum_{k=1}^{\infty} 1/(k(\theta k+2)(\theta(k-1)+2))$