HAL
open science

# A parsimonious multivariate copula for tail dependence modeling

Gildas Mazo, Stéphane Girard, Florence Forbes

## To cite this version:

Gildas Mazo, Stéphane Girard, Florence Forbes. A parsimonious multivariate copula for tail dependence modeling. EVT 2013 - Extremes in Vimeiro Today, Sep 2013, Vimeiro, Portugal. hal-00863540

**HAL Id: hal-00863540**

**https://hal.science/hal-00863540v1**

Submitted on 19 Sep 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A parsimonious multivariate copula for tail dependence modeling

Gildas Mazo, Stéphane Girard and Florence Forbes

Inria, Laboratoire Jean Kuntzmann

Grenoble, France

firstname.lastname@inria.fr

**Abstract:** Copulas are increasingly studied both in theory and practice as they are a convenient tool to construct multivariate distribution functions. However the material essentially covers the bi-variate case while in applications the number of variables is much higher. Furthermore, when one wants to take into account tail dependence, a desirable property is to have enough flexibility in the tails while avoiding the exponential growth of the number of parameters. We propose in this communication a one-factor model which exhibits this feature.

**Key words and phrases:** copula, tail dependence, parsimony, flexibility, one-factor model

## 1 Introduction

Copulas have been of increasing interest in the last decade. Nowadays they are widely used in areas like finance and hydrology. They are a convenient tool to construct multivariate distribution functions. Recall that copulas are distribution functions whose marginals are standard uniform. For a detailed account, see, e.g., the book of Nelsen [5]. Most of the material has been developed for the bi-variate case. As soon as the dimension gets higher, constructing multivariate copulas is a tough task. When studying extreme events, this is even more true as one has to take into account tail dependence. A simple way to do this is to consider the so called (upper) *tail dependence coefficient* (tdc) defined as:

$$\lambda_{ij} = \lim_{u \uparrow 1} P(U_i > u | U_j > u) = \lim_{u \uparrow 1} \frac{1 - 2u + C_{ij}(u, u)}{1 - u} \tag{1.1}$$

when the limit exists and where $U_i, U_j$ are two uniform random variables whose joint distribution is the copula $C_{ij}$. A high value of this coefficient is interpreted as the two variables tend to be large together. For the sake of simplicity, we will only consider pairwise tdc's and assume they represent a good picture of the dependence in the tails overall. In this communication, we propose

a multivariate copula which has a closed-form expression and exhibits a good compromise between flexibility in the tails and parsimony. It is generated by pairwise Cuadras-Augé copulas [1] through a one-factor model as described below.

## 2 One-factor copula generated by Cuadras-Augé copulas

Let $U_0$ be a latent standard uniform random variable and $C_{1|0}, \ldots, C_{d|0}$ be conditional copulas given $U_0$. Draw independently $U_i$ from $C_{i|0}$, $i = 1, \ldots, d$, so that the $U_i$ are independent given $U_0$. The distribution function of $U = (U_1, \ldots, U_d)$ at $u = (u_1, \ldots, u_d)$ is then given by

$$C(u) = P\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right) = \int_0^1 C_{1|0}(u_1, x) \ldots C_{d|0}(u_d, x) dx. \tag{2.1}$$

This construction was mentioned by Joe [3] and was called a *one-factor copula*. As noted by the author, it is a particular case of a *vine copula*. It can also be embedded within the framework of Kirshner [4] as a special case of a *latent tree copula*. We consider the case when one plugs Cuadras-Augé copulas into (2.1). Let

$$C_{i|0}(u_i, u_0; \theta_i) = \begin{cases} u_i^{1-\theta_i} \text{ if } u_0 < u_i, \\ (1-\theta_i) u_i u_0^{-\theta_i} \text{ if } u_0 > u_i \end{cases}$$

be Cuadras-Augé conditional copulas, $(u_{(1)}, u_{(2)}, \ldots, u_{(d)})$, $u_{(i)} < u_{(i+1)}$, $i = 1, \ldots, d-1$ be the ordered vector associated with $u$ and $\theta_i \in [0, 1]$, $i = 1, \ldots, d$. Observe that $\lambda_{0i} = \theta_i$ is the tdc of the pair $(U_0, U_i)$. In addition, we can show that

$$C(u) = \sum_{k=1}^{d} \pi_k \tilde{C}_k(u), \tag{2.2}$$

where

$$\tilde{C}_k(u) = \prod_{j=1}^{d} u_{(j)}^{1-\xi_{(j)k}}, \quad \xi_{(j)k} = \begin{cases} 0 \text{ if } j \leq k, \\ -1 + \sum_{i=1}^{k+1} \theta_{(i)} \text{ if } j = k+1, \\ \theta_{(j)} \text{ if } j \geq k+2. \end{cases}$$

and

$$\pi_k = -\theta_{k+1} \frac{\sum_{j=1}^{k} \theta_{(j)} \prod_{j=1}^{k} (1 - \theta_{(j)})}{\left(1 - \sum_{j=1}^{k} \theta_{(j)}\right)\left(1 - \sum_{j=1}^{k+1} \theta_{(j)}\right)}, \tag{2.3}$$

$k = 1, \ldots, d-1$, $\pi_d = \prod_{j=1}^{d} (1 - \theta_{(j)}) / \left(1 - \sum_{j=1}^{d} \theta_{(j)}\right)$, $\sum_{k=1}^{d} \pi_k = 1$. At a first glance (2.2) looks similar to a standard mixture model, but we can actually show that each $\pi_k \notin (0, 1)$. In addition,

the $\tilde{C}_k$ are not copulas in general. This copula is parsimonious in the sense that it has only $d$ parameters $\theta_1, \ldots, \theta_d$. Another interesting property of this copula is that the tdc $\lambda_{ij}$ between the variable $U_i$ and $U_j$ has a very simple form

$$\lambda_{ij} = \theta_i \theta_j. \tag{2.4}$$

Another copula model derived from bi-variate Cuadras-Augé copulas is the pair-wise multivariate extreme value model in Durante and Salvadori [2] (section 5, p. 150). It has $d(d-1)/2$ parameters $\theta_{ij}$ which must satisfy $d$ constraints (2.5). The tdc's between the variables $U_i$ and $U_j$ are

$$\lambda_{ij} = \theta_{ij}, \quad \text{under the constraints} \quad \forall i = 1, \ldots, d, \sum_{j=1, j \neq i}^{d} \theta_{ij} \leq 1. \tag{2.5}$$

# 3 A study of flexibility

Let $\theta$ be the parameter vector of a copula model $C(\theta)$ of dimension $d$ and let $\mathcal{P} = \big\{\{ij\} : i = 1, \ldots, d-1,\, j = 2, \ldots, d,\, i < j\big\}$ to be the set of all index pairs. The number of pairs is $p = |\mathcal{P}| = d(d-1)/2$. Further let $\lambda = \big(\lambda_{ij},\, \{ij\} \in \mathcal{P}\big)$ be the vector of the true pairwise tdc's and let $\lambda(\theta) = \big(\lambda_{ij}(\theta),\, \{ij\} \in \mathcal{P}\big)$ be the pairwise tdc's under the model of interest $C(\theta)$; for instance, $\lambda_{ij}$ can be one of (2.4) or (2.5). In order to measure the discrepancy between the true $\lambda$ and the model $\lambda(\theta)$, we define the tail dependence (quadratic) loss (tdl) as

$$\ell(\theta, \lambda) = \frac{1}{p} \sum_{\{ij\} \in \mathcal{P}} \big(\lambda_{ij} - \lambda_{ij}(\theta)\big)^2 \in [0, 1]. \tag{3.1}$$

To investigate the tail dependence accuracy, we shall process as follows. For each $k = 1, 2, \ldots$ generate $\lambda^{(k)} \in [0, 1]^p$ from some distribution, compute the minimum loss under the model $\ell_*^{(k)} = \min_\theta \ell(\lambda^{(k)}, \theta)$. Then examine the sample $\ell_*^{(1)}, \ell_*^{(2)}, \ldots$ through a boxplot. Figure 3.1 presents some primary results where each component of $\lambda^{(k)}$ is sampled independently from the uniform distribution on $[0, 1]^p$. We observe that the losses converge to a limit and the structure (2.4) is more accurate than (2.5). Even though our model (2.2) is more parsimonious than the pairwise extreme value model of Durante and Salvadori [6], it offers a largest flexibility for modeling tail dependence.

In future work, we will investigate other sampling schemes for $\lambda^{(k)}$ and compute accuracies for other models. We will study theoretical properties of the tdl (3.1). Note that it can be used as the basis of a fitting strategy by replacing the $\lambda_{ij}$ in (3.1) with empirical estimators $\hat{\lambda}_{ij}$, as in [6].

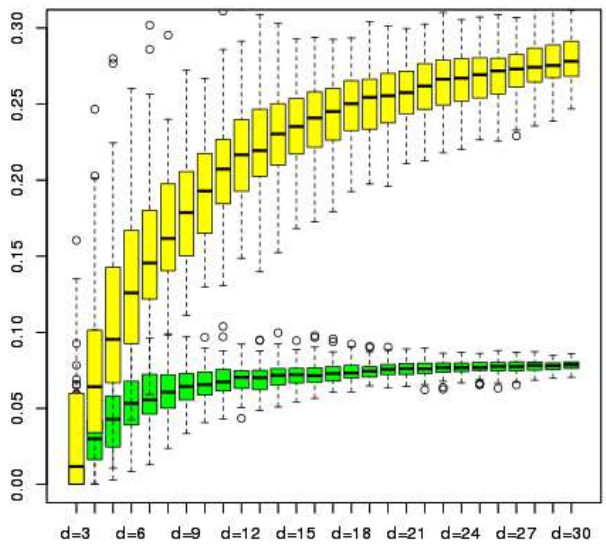


Figure 3.1: Tail dependence loss for two tdc structures. Each box is estimated from a sample of size 100. Green (bottom) is the proposed one-factor model (2.4). Yellow (top) is the Durante and Salvadori model (2.5).

# References


[1] C. M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4):339–353, 1981.

[2] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.

[3] H. Joe. Tail dependence in vine copulae. In *Dependence Modeling Vine Copula Handbook*. World Scientific, D. Kurowicka & H. Joe eds, 2011.

[4] S. Kirshner. Latent tree copulas. *Sixth European Workshop on Probabilistic Graphical Models*, Granada, 2012.

[5] R. B. Nelsen. *An introduction to copulas*. Springer, New York, 2006.

[6] G. Salvadori and C. De Michele. Estimating strategies for multiparameter multivariate extreme value copulas. *Hydrology and Earth System Sciences*, 15(1):141–150, 2011.