DEPARTMENT OF ACCOUNTING & FINANCE

A new graphical tool for copula selection

Frederik Michiels & Ann De Schepper

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

EARNED EXCELLENCE
AACSB ACCREDITED
EST. 1916
THE BEST BUSINESS SCHOOLS IN THE WORLD

Stadscampus
Prinsstraat 13, B.213
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
http://www.ua.ac.be/tew

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ACCOUNTING & FINANCE

A new graphical tool for copula selection

Frederik Michiels & Ann De Schepper

RESEARCH PAPER 2010-004
MAY 2010

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.213
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be

The papers can be also found at our website:
www.ua.ac.be/tew (research > working papers) &
www.repec.org/ (Research papers in economics - REPEC)

D/2010/1169/004

A new graphical tool for copula selection

Frederik Michiels*

Faculty of Applied Economics, University of Antwerp, Prinsstraat 13, Antwerp, Belgium

Ann De Schepper

Faculty of Applied Economics & StatUa Statistics Center, University of Antwerp, Prinsstraat 13, Antwerp, Belgium

Abstract

The selection of copulas is an important aspect of dependence modeling. In many practical applications, only a limited number of copulas is tested, and the modeling applications usually are restricted to the bivariate case. One explanation is the fact that no graphical copula tool exist which allows to assess the goodness-of-fit of a large set of (possible higher dimensional) copula functions at once. This paper pursues to overcome this problem by developing a new graphical tool for the copula selection, based on a statistical analysis technique called 'principal coordinate analysis'. The advantage is threefold. In the first place, when projecting the empirical copula of a modeling application on a two-dimensional copula space, it allows us to visualize the fit of a whole collection of multivariate copulas at once. Secondly, the visual tool allows to identify 'search' directions for potential fit improvements (e.g. through the use of copula transforms). Finally, in the bivariate case the tool makes it also possible to give a two-dimensional visual overview of a large number of known copula families, for a common concordance value, leading to a better understanding and a more efficient use of the different copula families. The practical use of the new graphical tool is illustrated for two two-dimensional and two three-dimensional fitting applications.

Keywords: copulas, copula selection, Kendall's tau, graphical tool, multidimensional scaling, dependence models

*frederik.michiels@ua.ac.be.

1 Introduction

In a recent bibliometric study by Genest et al. (2009), one of the conclusions of the authors is that there is a clear need for the further exploration of graphical tools for the selection and validation of copula models. A drawback of the copula visualization tools that are used most commonly, e.g. the level-curve representation or a density-cure plot, is that they become inappropriate when it comes to visualize the goodness-of-fit of a collection of copulas. The alternative of measuring and visualizing the fit of every copula separately is rather tedious and unpractical.

A copula visualization tool which gives a clear overview of the proximity of a (large) collection of copulas in a dependence modeling application is, to the best of our knowledge, non-existent. Nevertheless, from a practical point of view this is an absolute necessity, as it facilitates the consideration of a large number of possible copulas and therefore make the copula selection less arbitrary.

In order to create a two-dimensional representation of a copula space, we use a technique called *principal coordinate analysis*. We start by creating an Euclidean distance matrix, based on the dissimilarities between the values of the cumulative distribution function (cdf) for different copulas, evaluated over a fine grid. Next, we construct a configuration of points in two dimensions, where the Euclidean distances between the different copulas correspond to the original distance matrix. In order to check whether the 2D representation is appropriate, the goodness-of-fit of the two dimensions can be calculated using the stress-function.

The copula selection tool we develop can be used to visualize both bivariate and multivariate goodness-of-fit problems. For the bivariate case we present an overview of 21 well-known one-parameter copula families. The overview is presented as a representation of 'comparable' copulas for a common Kendall's τ value, for 10 % intervals, and can be used as a copula guide for selection or study purposes. The use of the selection tool is shown by means of two examples of well-known data sets. For the multivariate case the selection tool is used to represent the fit of a collection of 18 multivariate copula families in two three-dimensional fitting examples. Finally, the copula selection tool has the extra advantage over existing graphical tools in that it allows to visualize search directions for potential fit improvements.

The paper is organized as follows. In Section 2 an overview of existing (bivariate) graphical tools is presented and the new copula selection tool is developed. In Section 3 the bivariate case is discussed, including the copula guide and two fitting examples. Section 4 discusses the multivariate case and finally, Section 5 contains a conclusion.

2 A new graphical tool for copulas

2.1 Graphical tools: an overview

Recall that a copula C is a multivariate cumulative distribution function (cdf) on the unit n-cube $[0,1]^n$ with uniform margins. By virtue of Sklar (1959), copulas provide a flexible tool for the construction of a n-dimensional cdf H from continuous margins F as $H(\mathbf{x}) = C(F(\mathbf{x}))$. In the continuous case, such a copula is unique. Consequently, multivariate goodness-of-fit problems can be reduced to copula selection problems.

To our knowledge, the availability of graphical tools which can be used for the problem of (bivariate) copula selection is rather limited. We now provide a short overview.

1. Bivariate copulas can well be represented by displaying a 3D image of their cdf or probability density function (pdf), or plotting a 2D level curve representation. Although this technique works fine for the investigation of a single copula, it becomes unpractical for the comparison of the fit of a whole collection of copulas and it clearly is restricted to the bivariate case. In Figure 1 the four representation forms are depicted. Note that for the cdf level curve representation (Figure 1, upper right panel), a possible strategy would be to compare the level curves to those of the co-monotonic, independent and counter-monotonic copulas.

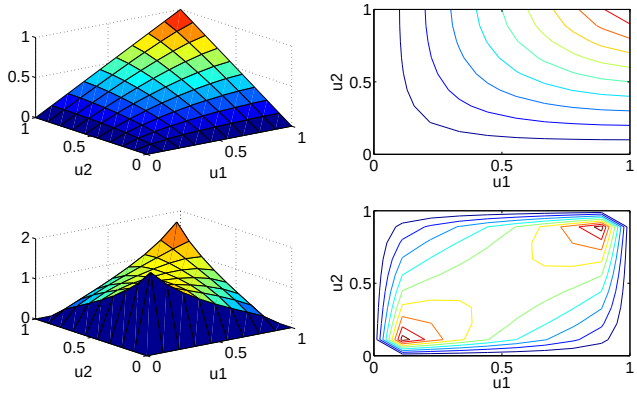


Figure 1: Representation of a Gaussian copula, $\rho = 50\%$, 3D cdf (upper right), 2D cdf level curve (upper left), 3D pdf contours (lower right), 2D pdf level curve (lower left).

2. For bivariate Archimedean copulas, Genest and Rivest (1993) show that the goodness-of-fit can be assessed using quantile-quantile-plots based on the bivariate probability integral transform $K(t) = P(C(U_1, U_2) \leq t)$, for any $t \in [0,1]$, and they develop a procedure to obtain K empirically. In a recent paper we also stress the visual advantage of K in recognizing key dependence character-

istics of an Archimedean copula represented in the lambda function $\lambda(t) = t - K(t)$, see Michiels et al. (2010). In Figure 2 an example is provided; note that $\lambda(t) = t\ln(t)$ corresponds to the independent copula. The downside, however, is that K and λ are only parametrically available for Archimedean copulas, and just as it holds for contour diagrams, the comparison of dissimilarities among several copula families can be unclear and subjective.

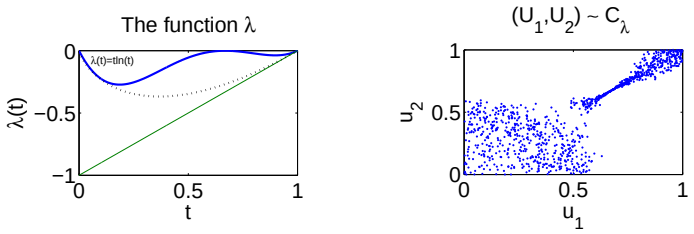


Figure 2: Illustration of the lambda function for a copula with $\lambda(t) = t(t-1)(7.71t^2 - 10.31t + 3.5)$, $\tau = 58.13\%$, $\lambda_L = 8.84\%$, $\lambda_U = 13.39\%$.

3. A final graphical tool for picturing bivariate copulas is based on conditional expectation curves, see e.g. Nikoloulopoulos and Karlis (2008). The visualizations can provide more insight into similarities between copula families. In Figure 3 an example is given. For higher dimensional copulas, the possible regression curves augment quickly (certainly for non-exchangeable copulas), and clearly one graph would not suffice to get a clear image of the copula. Furthermore, displaying a large collection of copulas in this way, again, becomes confusing and unpractical.

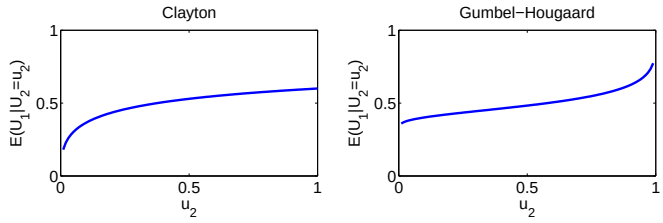


Figure 3: Conditional expectation curves, the Clayton copula vs. the Gumbel-Hougaard copula, $\tau = 20\%$.

In the following section, we show how a visualization based on principal coordinate analysis is much more convenient in order to compare the appropriateness of a large collection of copulas, displayed together, and how it can be used straightforwardly for goodness-of-fit purposes. This visual tool can easily be extended to the multivariate case, see section 4.

2.2 Description of our new graphical tool

The graphical tool developed in this contribution heavily relies on principal coordinate analysis. Principal coordinate analysis or metric multidimensional scaling is a multivariate statistical data analysis technique used to gain insight in the underlying relations which are present in a multivariate dataset. In the context of our investigation, the multivariate dataset corresponds to a collection of copulas with cdf-values. The main objective of the technique is to provide a low-dimensional representation of high-dimensional data, in such a way that the distortion caused by a reduction in dimensionality is minimized.

In this subsection we only briefly recall the main aspects of this technique. We refer to the seminal papers of Kruskal (1964a,b) for a concise description, or to Kruskal and Wish (1978) or Borg and Groenen (1997) for a more elaborate analysis. In the sequel, we adopt the same notations as Johnson and Wichern (1992).

Multidimensional scaling techniques deal with the following problem: For a set of observed similarities between every pair of n copulas, find a representation of the items in few dimensions such that the inter-item proximities nearly match the original similarities. In this analysis we use *classical* multidimensional scaling, which means that we use the Euclidean distance as a proximity measure in the final low-dimensional configuration.

We now come to the basic algorithm. Departing from a set of n copulas $C_1, C_2, \cdots, C_n$ (in cdf form) we compute the (simulated) dissimilarity between the i^{th} and j^{th} copula through $\hat{d}_{(i,j)} = \sqrt{(C_i - C_j)'(C_i - C_j)}$, the Euclidean distance. For the examples, we use 1000 random copula observations, as this yields stable results (i.e. simulated distance corresponds to theoretical distance). The Euclidean distance is chosen in order to give every cdf-observation equal weight.

For the n copulas we obtain $M = n(n-1)/2$ distances between pairs of different copulas. Assuming no ties, these distances can be arranged in a strictly ascending order as

$$\hat{d}_{(i_1,j_1)} < \hat{d}_{(i_2,j_2)} < \cdots < \hat{d}_{(i_M,j_M)} \tag{1}$$

with $1 \leq i_k \neq j_k \leq n$ $(k = 1, \cdots, M)$ where $d_{(i_M,j_M)}$ is the largest of the M dissimilarities and the subscript (i_1, j_1) indicates the pair of copulas which is most similar. The challenge now is to find a q-dimensional configuration of the n items $(q < n)$ such that in terms of the q new dimensions, the distances $d^{(q)}_{(i,j)}$ between pairs of copulas match the ordering in (1). If the new distances are laid out in a manner corresponding to that ordering, a perfect match occurs when

$$d^{(q)}_{(i_1,j_1)} < d^{(q)}_{(i_2,j_2)} < \cdots < d^{(q)}_{(i_M,j_M)}. \tag{2}$$

For our copula applications we strictly want to use 2D representations and consequently we fix $q = 2$. The coordinates (x, y) attached to the new distances $d^{(2)}$ are then found by minimizing the so called 'stress' function over the M coordinates $(x_{i_1}, y_{i_1}), \cdots, (x_{i_M}, y_{i_M})$ which is defined as

$$\text{Stress} = \left\{ \frac{\sum_{k=1}^{M} (d^{(2)}_{(i_k,j_k)} - \hat{d}_{(i_k,j_k)})^2}{\sum_{k=1}^{M} {d^{(2)}_{(i_k,j_k)}}^2} \right\}^{1/2} \tag{3}$$

with distance $d^{(2)}_{(i_k,j_k)} = \sqrt{(x_{i_k} - x_{j_k})^2 + (y_{i_k} - y_{j_k})^2}$, where i_k, j_k $(k = 1, \cdots, M)$ are determined by equation (1).

A stress value of 2.5% is said to provide an excellent fit, see e.g. Kruskal (1964a). Our objective is to find a two-dimensional representation of the copula test spaces, both for creating a copula guide and for goodness-of-fit purposes. For the former (see Section 3.2) it can be seen that a nearly perfect match occurs for all test spaces for 10 % τ intervals and as such, it follows that optimizing equation (3) also leads towards the correct ordering of all $d^{(2)}$'s. However, when comparing the goodness-of-fit of each member of a whole collection of copulas used as a model for an empirical copula (see Sections 3.3 and 4.2), optimizing equation (3) may lead towards violations against equation (2), which becomes problematic. Hence, we propose to add a restriction on the scaling technique for goodness-of-fit applications, with respect to the distances between a copula from the test space and the empirical ocula. Let

$$\hat{d}_{(1,emp)} < \hat{d}_{(2,emp)} < \cdots < \hat{d}_{(n,emp)} \tag{4}$$

be the ordering of the distances between the n parametric copulas from the test space and the empirical copula. Then, in order to assure that equation (4) is always satisfied during the minimization process, add the following restriction to (3):

$$d^{(2)}_{(i,emp)} < d^{(2)}_{(j,emp)} \text{ if } \hat{d}_{(i,emp)} < \hat{d}_{(j,emp)} \tag{5}$$

where $d^{(2)}_{(i,emp)} = \sqrt{(x_i - x_{emp})^2 + (y_i - y_{emp})^2}$.

As mentioned earlier, the representation in this contribution is based on the Euclidean distance. Consequently, the goodness-of-fit metric will be Euclidean too. This assumption, however, can be relaxed, one can take other distance measures like e.g. the Kolmorov-Smirnoff distance or the Cramér-von Mises metric. Without entering into details, the latter will yield more or less the same results as it merely differs in taking the average of all distinct distance pairs. The Kolmorov-Smirnoff will yield different results, but in our opinion it is too sensitive for outliers and therefore not the best choice. Finally, note than when necessary a more robust distance metric than the Euclidean distance can be applied.

2.3 Utilisation

The main idea of our graphical tool consists of making a 2-dimensional representation of the empirical copula together with a test space, which is a set of workable theoretical copula functions. In the bivariate case, the most obvious approach is to choose for this set of copula models the whole comparable test space, corresponding to the estimated τ-value of the empirical copula (see Michiels and De Schepper (2008)). In the multivariate case, we can use a maximum likelihood approach in order to estimate the parameters of some copula models based on the empirical data, and we then define the set of copula functions as the set of these comparable copula models.

In the ideal situation, the stress value is smaller than 2.5% for $q = 2$, allowing to work with a 2D-representation of the copula test space. Then plot the empirical copula in the same plane together with the entire test space; in order to model the dependence in the data, one can choose for the copula with the highest degree of similarity with the empirical copula. This approach will be used for the examples in section 3.3 and 4.2.

Note that the 2D-representation of comparable copula functions also admits mutual comparisons of comparable copula functions. This will be illustrated for the bivariate case in section 3.2. We refer to appendix A for more information on the investigated copulas and on Archimedean copulas in particular.

Finally, the visual tool permits the identification of search directions for potential goodness-of-fit improvements and the inclusion of these improvements on the 2D-representation (see sections 3.3 and 4.2). These improvements can be achieved by making use of transforms preserving or changing certain aspects of the copula model in such a way that a better fit is obtained. We illustrate this possibility with two examples in the bivariate case.

A first strategy for improving the goodness-of-fit could be to create a non-exchangeable alternative of the best fit copula from our set. A straightforward technique can be found in Genest et al. (1998), where the following two-parameter transform is proposed:

$$C_{\kappa,\eta}(u,v) = u^{1-\kappa}v^{1-\eta}C(u^{\kappa},v^{\eta}) \tag{6}$$

where $0 < \kappa, \eta < 1$.

Another strategy consists of looking for goodness-of-fit improvements by holding some aspects of the best fit copula constant. In this respect, we first summarize a result, introduced in Michiels and De Schepper (2009), which allows to construct τ-preserving transforms, starting from an arbitrary Archimedean copula function. Let $f_\theta : [0,1] \mapsto [0,1]$ be an increasing bijection with

parameter vector θ. A τ-preserving transform is given by

$$\lambda^t(t) = \frac{df_\theta}{dt}(t) \cdot \lambda(f_\theta(t)) \tag{7}$$

where the range for θ is determined using feasibility conditions on λ, which are $\lambda(0) = 0, \lambda(1) = 0$ and $\lambda(t) < 0, \frac{d\lambda}{dt}(t) < 1$ for all $t \in (0, 1)$.

This means that, for any Archimedean copula and for any increasing bijection on $[0, 1]$, we can create a new copula, with the same value for the concordance parameter, but which differs from the original copula in a certain way. When the transformed copula functions are depicted on the 2D-representation as well, together with the empirical copula and together with the other copula functions of the comparable test space, it allows the choice of an appropriate transform and possibly an appropriate value for the potential parameter of the transforming function f_θ, corresponding to a better overall fit. This technique will also be illustrated in section 3.3.

As an extra benefit of this approach, note that the transform can be chosen such that next to the concordance value, also the tail dependence coefficients of the original copula model are preserved. This will be the case when the function $f_\theta : [0, 1] \mapsto [0, 1]$ satisfies the extra conditions $f'_\theta(0) = 1$, $f'_\theta(1) = 1$, with $f''_\theta(0)$ bounded.

3 The Bivariate Case

As stated in the introduction, the contribution in the bivariate case is threefold. Using our graphical tool we present a copula guide of 21 well-known one-parameter copula families with potential practical use. Secondly, we illustrate the benefits of our tool for copula selection for two fitting applications, and finally we search for fit improvements.

3.1 Selection of copulas

In the remainder of this section the following copula families are used (their functional forms are presented in the first appendix):

- Archimedean class: 14 strict Archimedean families, including the popular Clayton, Frank, Gumbel-Hougaard, Joe and Ali-Mikhail-Haq families $(C_1, \cdots, C_{14})$;
- Extreme value families: Galambos and Hüsler-Reiss family or Gumbel-Hougaard family (C_{15}, C_{16});
- Meta-elliptical class: Normal, Student's t family (C_{17}, C_{18});

- Other families: Farlie-Gumbel-Morgenstern (FGM) family, Plackett family and Raftery family (C_{19}, C_{20}, C_{21}).

For more detailed information about the Archimedean copula class, please consult e.g. Chapter 4 in Nelsen (2006) and McNeil and Nešlehová (2009); for the meta-elliptical class, see Fang and Fang (2002) and for extreme value class, we refer to Joe (1997).

In the bivariate case the copula parameters can well be interpreted in terms of concordance. Indeed, for Kendall's tau we have $\tau = 4\int_0^1 \int_0^1 C(u_1, u_2)dC(u_1, u_2) - 1$ and for the coefficients of tail dependence $\lambda_L = \lim_{u\to 1} \frac{1-2u+C(u,u)}{1-u}$ and $\lambda_U = \lim_{u\to 0} \frac{C(u,u)}{u}$. Note that for the Archimedean class, these quantities can be written straightforwardly by means of the λ-function: $\tau = 1 + 4\int_0^1 \lambda(t)dt$, $\lambda_L = 2^{\lambda'(0^+)}$ and $\lambda_U = 2 - 2^{\lambda'(1^-)}$, see Michiels et al. (2010).

A bivariate copula C is called exchangeable if $C(u_1, u_2) = C(u_2, u_1)$, it is reflection symmetric if $C(u_1, u_2) = C^S = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$, where C^S is called the survival copula, and it exhibits the extreme value property if $C(u_1^k, u_2^k) = \Big(C(u_1, u_2)\Big)^k$ for all $k > 0$.

Remark that all 21 copula families considered here are in fact exchangeable. In order to get an idea of the diversity of this collection of copula families we present an overview of some properties of the 21 copula families in Table 1. Columns two and three contain the coefficients of tail dependence in function of the copula parameter θ. Note that the Student's t copula has two parameters, the correlation coefficient ρ and the degrees of freedom v. The fourth and fifth column display whether or not a copula family possesses the reflection symmetry and/or the extreme value property. Finally, the last column indicates the copula dependence range in terms of Kendall's τ.

3.2 Copula guide

We now visualize the relative differences between one-parameter copula families for a common concordance value (Kendall's tau), using multidimensional scaling. The technique is applied to the 21 copula families introduced earlier, each copula being represented as a vector containing 1000 cdf data points. The goal is to map each of the comparable test spaces (i.e. a set of copulas adequate to model a certain degree of dependence, see Michiels and De Schepper (2008)) into the two-dimensional space, by assigning a point in the (x, y)-plane to each copula of the test space. For each analysis we check the stress function to see whether a 2D representation is feasible. The stress values are summarized in Table 2. It is clear that these stress values indicate an excellent goodness-of-fit for all test spaces.

family	λ_L	λ_U	Reflection symmetry	Extreme value	Kendall's τ range
1	$2^{-1/\theta}$	0			$[0, 1]$
2	0	0			$[\frac{1}{2}(5 - 8\ln 2, \frac{1}{3}]$
3	0	$2 - 2^{1/\theta}$		$\surd$	$[0, 1]$
4	0	0	$\surd$		$[-1, 1]$
5	0	$2 - 2^{1/\theta}$			$[0, 1]$
6	0	0			$[-0.36, 0]$
7	0	0			$[-0.18, 0]$
8	$2^{-1/\theta}$	$2 - 2^{1/\theta}$			$[\frac{1}{3}, 1]$
9	0	0			$[-0.36, 1]$
10	$\frac{1}{2}$	$2 - 2^{1/\theta}$			$[\frac{1}{3}, 1]$
11	$\frac{1}{2}$	0			$[-1, \frac{1}{3}]$
12	0	0			$[-0.61, 1]$
13	1	0			$[\frac{1}{3}, 1]$
14	1	0			$[0, 1]$
15	0	$2^{-1/\theta}$		$\surd$	$[0, 1]$
16	0	$2 - 2\Phi(1/\theta)$		$\surd$	$[0, 1]$
17	0	0	$\surd$		$[-1, 1]$
18	$2t_{v+1}\left(\frac{-\surd(v+1)\surd(1-\rho)}{\surd(1+\rho)}\right)$		$\surd$		$[-1, 1]$
19	0	0	$\surd$		$[\frac{-2}{9}, \frac{2}{9}]$
20	0	0	$\surd$		$[-1, 1]$
21	$\frac{2\theta}{1+\theta}$	0			$[0, 1]$

Table 1: Characteristics for 21 copula families.

τ	stress value
-90%	$< 10^{-3}$
-80%	$< 10^{-3}$
-70%	$< 10^{-3}$
-60%	$< 10^{-3}$
-50%	$< 10^{-3}$
-40%	$< 10^{-3}$
-30%	0.12%
-20%	$< 10^{-3}$
-10%	0.22%

τ	stress value
10%	0.55%
20%	1.01%
30%	0.91%
40%	0.85%
50%	0.97%
60%	0.98%
70%	0.97%
80%	1.59%
90%	1.83%

Table 2: Stress values for principal coordinate analyses for different values of τ.

Negative dependence

In Figure 8, a two-dimensional mapping of the copula families accounting for negative dependence is shown, for nine cases corresponding to $\tau \in \{-90\%, -80\%, -70\%, -60\%, -50\%, -40\%, -30\%, -20\%, -10\%\}$. The following observations can be made:

- For $\tau \in \{-90\%, -80\%, -70\%\}$, the feasible copula families are the Archimedean copulas $\{C_4, C_{11}\}$, the meta-elliptical copulas $\{C_{17}, C_{18}\}$ and copula C_{20}. Note that C_{18} is represented for 3 degrees of freedom. As all bivariate copulas share the same lower bound, it is obvious that the relative differences augment for increasing values of τ, see panel (a)-(b)-(c).

- For $\tau \in \{-60\%, -50\%, -40\%\}$, the set of comparable copulas is extended with the Archimedean copula C_{12}. In general, two main groups remain, but they get closer. See panel (d)-(e)-(f).

  One clear group can be identified, consisting of the copulas with reflection symmetry. On the other hand, copulas C_{11} and C_{12} are clear outliers. The asymmetric behavior of C_{11} is clear, due to the fact that it has constant lower tail dependence, the asymmetric behavior of C_{12} can be deduced from the plots and is more or less opposite to that of C_{11}, although the former has no upper tail dependence[(1)].

- For $\tau \in \{-30\%, -20\%, -10\%\}$, the comparable test space is enlarged again, now with copulas C_6, C_9, C_{19}, C_7 and C_2 (in that order). See panel (g)-(h)-(i).

  Remark that as τ approaches zero, the relative differences between the copula families entailing independence become smaller and smaller. The clear outliers are C_{11} and C_{18}, and this can be explained by the fact that these copula families do not include the independence copula.

Positive dependence

In Figure 9, a two-dimensional mapping of the copula families accounting for positive dependence is shown, for nine cases corresponding to $\tau \in \{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%\}$. The following observations can be made:

- For $\tau \in \{10\%, 20\%, 30\%\}$, the comparable test space is very large, with the Archimedean copulas $\{C_1, C_2, C_3, C_4, C_5, C_6, C_9, C_{11}, C_{12}, C_{14}\}$, the extreme value copulas $\{C_{15}, C_{16}\}$, the meta-elliptical copulas $\{C_{17}, C_{18}\}$ and the other copula families $\{C_{19}, C_{20}, C_{21}\}$, with copula

[(1)] This fact can also be assured by making λ-plots.

C_{19} disappearing for $\tau > 20\%$, see panel (a)-(b)-(c). We also added the survival versions of C_1 and C_3, for reasons of comparison.

We observe that for $\tau > 20\%$ three regions can be identified: on the left we have copulas with lower tail dependence, in the centre copulas with symmetric tail dependence, on the right copulas with only upper tail dependence. Remark that with increasing τ-values, the group $\{C_1, C_2, C_{11}, C_{14}\}$ can be recognized, all sharing the element $C(u_1, u_2) = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}$ for $\tau = \frac{1}{3}$.

Note also the high similarity between the extreme value copulas $\{C_3, C_{15}, C_{16}\}$, which is apparent throughout the whole analysis: they share the extreme value property, they are also exchangeable and their upper tail dependence parameter can be approximated as $\lambda_U \approx 2^{1-\tau}$.

- For $\tau \in \{40\%, 50\%, 60\%\}$, a new group of Archimedean copulas enters the test space, namely $\{C_8, C_{10}, C_{13}\}$, see panel (d)-(e)-(f).

  Notice the strong similarity between copulas C_{13} and C_{14} (due to their exceptional lower tail dependence property), C_1 and C_{21} (both varying lower tail dependence), C_{10} and C_{18} (both upper and lower tail dependence), and C_1^S and C_5 (relative high upper tail dependence).

- For $\tau \in \{70\%, 80\%, 90\%\}$, relative distances become smaller as all copulas converge to the Fréchet upper bound; therefore, for panel (i), the scale is different.

  As τ approaches 1, the most important and only diversifying factor is λ_L.

How can this copula guide be used in a practical fitting problem?

Suppose that for a particular modeling problem, an estimate for the overall concordance parameter τ has been calculated, and suppose that one wants to choose one or more copulas to model the dependence structure in the data. A good approach would be to start from the pictures in Figures 8-9, in particular the representation corresponding to the situation with the closest τ-value, and to choose some copula families with large dissimilarities. After testing the goodness-of-fit for each of these three or four copula families, one can refine the search procedure, by exploring some more copula families with a higher degree of similarity with the best copula family from the first loop.

3.3 Empirical illustrations

In the previous section we showed that principal coordinate analysis allows to represent comparable test spaces in a two-dimensional space, and their relative differences can be interpreted using known copula properties. We now proceed with a practical application of the visualization tool. The 2D

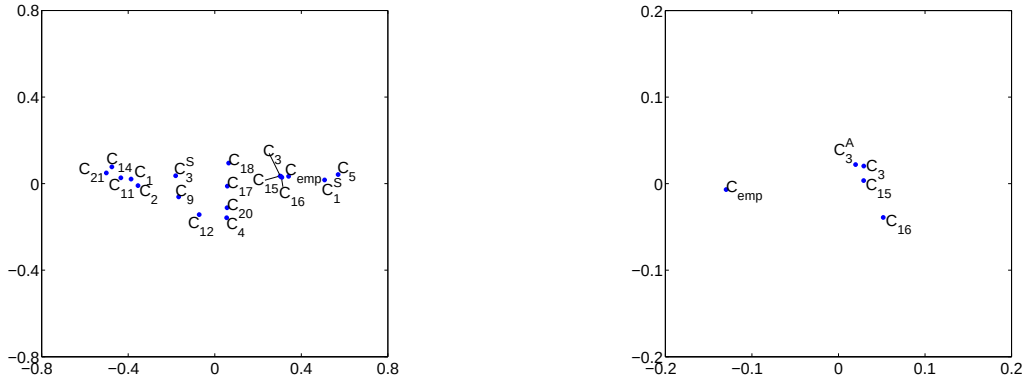


Figure 4: 2D representation goodness-of-fit Loss-ALAE dependence structure (left), 2D representation of goodness-of-fit improvement (right).

mapping allows for the representation of the goodness-of-fit of a complete comparable test space (i.e. a collection of copulas). Indeed, plotting the empirical copula C_{emp}[(2)] into a comparable test space, enables the investigation of the relative distances between the empirical copula and all copulas of the test space at once. Afterwards, we can look for a search strategy to move closer to C_{emp}.

3.3.1 The Loss-ALEA data set of Frees and Valdez (1998)

As a first example we take the well-known Loss-ALEA data set studied by Frees and Valdez (1998). The data comprise 1,500 general liability claims (expressed in USD) randomly chosen from late settlement lags, and were provided by Insurance Services Office, Inc. Each claim consists of an indemnity payment (loss) and an allocated loss adjustment expense (ALAE). Here ALAE are types of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers' fees and claims investigation expenses. In order to price an excess-of-loss reinsurance treaty when the reinsurer shares the claim settlement costs, the dependence between losses and ALAEs has to be taken into account. Departing from the well-known Clayton, Frank and Gumbel-Hougaard test space, Frees and Valdez (1998) conclude that the Gumbel-Hougaard family provides the best fit.

The observed overall dependence is estimated using a rank based estimator of Kendall's τ yielding $\hat{\tau} = 31.54\%$. If we apply our approach to this example, by visualizing the empirical copula in the two-dimensional projection of the comparable test space associated with $\hat{\tau}$, we obtain Figure 4 (left panel). This graph displays the performance of a large set of copulas, and it can be seen

(2) For a definition of the empirical copula, see appendix C.

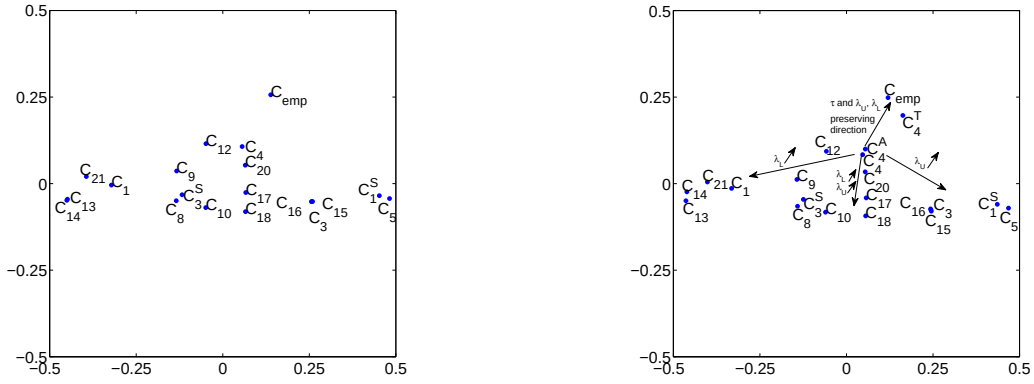


Figure 5: 2D representation goodness-of-fit caesium-scandium dependence structure (left) and improvement strategy for goodness-of-fit caesium-scandium dependence structure (right).

immediately that the distance between C_{emp} and the Gumbel-Hougaard family C_3 is much smaller than for the Clayton family C_1 or the Frank family C_4 – which confirms the choice of Frees and Valdez. However, we see that next to C_3 also the (other) extreme value families C_{15} and C_{16} provide a good fit. When taking a closer look (right panel of Figure 4), and adding the best fit asymmetric version of the Gumbel-Hougaard copula through equation (6) (using a maximum likelihood approach), it becomes clear that the fit can in fact be slightly improved. From the symmetric copulas the Galambos copula (C_{15}) provides the best fit, but the asymmetric copula C_3^A performs better.

3.3.2 The uranium-caesium data set from Genest and Rivest (1993)

As a second example we take the data set used in Genest and Rivest (1993) where the authors present 655 analyses from uranium and caesium log-concentrations (in parts per billion). In their paper the authors conclude that the Frank copula provides the better fit, compared to the Clayton, Gumbel-Hougaard family and a two-parameter 'log copula' family. We now use our approach to check whether or not a better fit is possible when using a copula test space of 21 families. The estimated τ value equals 47.96%. We proceed as in the previous section by visualizing the empirical copula into the appropriate comparable test space. This is done in Figure 5 (left panel).

The 2D mapping shows that the Frank copula family (C_4) indeed provides the best fit amongst the one-parameter families. However, the best fit distance is quite large (0.2206) and therefore it can be argued that an improvement of this fit should be investigated. As explained in Section 2.3, we can look for an improvement of the fit of the Frank copula by choosing an appropriate τ-*preserving transform*. In this way, we can search in the neighbourhood of C_5 without changing the

overall degree of dependence (and hence, we keep on working in the same comparable test space).

We know from Table 1 that the Frank copula induces no tail dependence. When looking again at Figure 5 (left panel), we see that the fit for copulas which do have tail dependence is worse. As a result, we need a transform which is both concordance and tail preserving.

Following the approach as explained in section 2.3, we look for increasing functions $f : [0,1] \mapsto [0,1]$, with $f(0) = 0$, $f(1) = 1$, $f'(0) = 1$, $f'(1) = 1$ and $f''(0)$ bounded.

In order to create a model with sufficient degrees of freedom, we choose for $f(\cdot)$ a polynomial of order 6. Applying the constraints mentioned above, this leads to the following three-parameter model:

$$f_{\theta_1,\theta_2,\theta_3}(t) = t(\theta_1 t^5 + \theta_2 t^4 + \theta_3 t^3 - (4\theta_1 + 3\theta_2 + 2\theta_3)t^2 + (3\theta_1 + 2\theta_2 + \theta_3)t + 1). \tag{8}$$

For this choice, there are three shape parameters, denoted by θ_1, θ_2 and θ_3. The parameters are estimated using the sum of squares principle on the empirical λ-function. The optimal values are $\hat{\theta}_1 = -6.9149$, $\hat{\theta}_2 = 19.3122$ and $\hat{\theta}_3 = -18.6477$. These values satisfy the feasibility conditions on the function f. As a result of this transform, the minimal Euclidean distance lowers to 0.1436. Figure 5 (right panel) shows the transformed copula as C_4^T. Note also that the best fit asymmetric Frank copula (C_4^A) does not yield a significant improvement in comparison with C_4^T.

4 The Multivariate Case

The 2D representation of a set of copula families can straightforwardly be extended to the multivariate case, given that the stress function results in acceptable values. We show the use of our copula selection in two three-dimensional fitting examples.

4.1 Selection of copulas

Unfortunately, it is not possible to extend all 21 bivariate families from section 3.1 from a bivariate to a multivariate setting. As such, we left out some families because no multivariate extension is known or possible, or no closed form is available. In order to make up for this loss we added other valuable multivariate families.

We use the following 18 multivariate copula families, where we kept the same numbering as in section 3 for the extended copulas:

- Archimedean class: 8 exchangeable Archimedean families, including the popular Clayton, Frank, Gumbel-Hougaard, Joe and Ali-Mikhail-Haq families ($C_1, \cdots, C_5$ and $C_8, \cdots, C_{10}$)

and 4 Archimedean families with partial symmetry (families M3, ..., M6 as introduced in Joe (1997));

- Extreme value families: families MM1, MM2 and MM3 defined in Joe (1997);
- Meta-elliptical class: Normal, Student's t family (C_{17}, C_{18});
- Other families: multivariate Farlie-Gumbel-Morgenstern (MFGM) family as defined in Nelsen (2006) (C_{19}).

For more details about these copula families, we refer to the second appendix.

4.2 Empirical illustrations

4.2.1 Financial data set from Michiels and De Schepper (2008)

We use data from Michiels and De Schepper (2008) consisting of 1499 total returns of the S&P 500 Composite Index, the JP Morgan Government Bond Index and the NAREIT All index, over the period from January 4 2002 until March 13 2008. In Michiels and De Schepper (2008) we only modeled the bivariate margins. We are now interested in the three-dimensional dependence structure. In order to get an idea of the dependence present in the data, we first estimate the concordance matrix, see Table 3.

τ	stocks	bonds	real estate
stocks	1	-0.1721	0.4351
bonds	-0.1721	1	-0.0599
real estate	0.4351	-0.0599	1

Table 3: Estimated values of Kendall's τ for the financial data set.

We show how our graphical tool can be used to get a picture of the goodness-of-fit of the 18 copula families. The parameter estimation is carried out using the maximum likelihood approach. It appears that for the financial data set two out of three bivariate margins actually contain negative dependence. In our copula set, the exchangeable (one parameter) Archimedean copula families, the asymmetric (two parameter) Archimedean families and the (7 parameter) multivariate extreme value copulas can only model positive dependence, whereas the meta-elliptical copulas and the (4 parameter) MFGM family also allow for pairwise negative dependence.

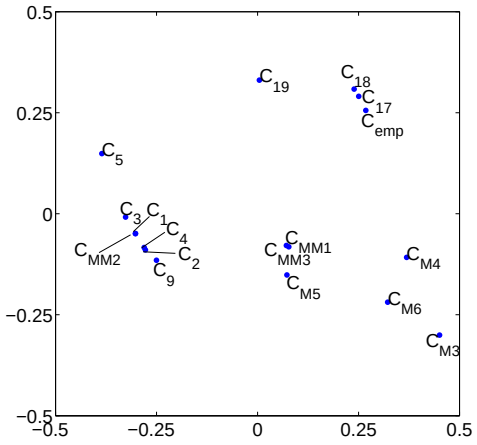


Figure 6: 2D representation of three-variate financial data fitting example.

In figure 6 the result of the fitting example can be found. The stress is calibrated on 5.04% which indicates a good fit. From the graph it follows that the meta-elliptical families C_{17}, C_{18} outperform all other copulas and that they yield a very good fit. The MFGM model C_{19} performs quite good, as indeed in this example the general dependence is rather low and negative pairwise dependence is present. Next, the multivariate extreme value copulas, C_{MM1} and C_{MM3} perform best. Notice that family C_{MM2}, which has as a special case C_1 (see Joe (1997)), actually equals C_1 here. Finally, all symmetric Archimedean copulas ($C_1, \cdots, C_5$ and C_9)[(3)] perform rather bad; this is due to the fact that they only have one parameter and can only be used to model positive dependence, and as a consequence, the estimated parameter values are close to the independence value.

4.2.2 Extension of the data set from Genest and Rivest (1993)

In the previous example, the data set contained bivariate margins with negative dependence, which partially explains why all Archimedean copulas (both symmetric and asymmetric) performed rather poorly. We now visualize a dependence modeling example containing only positive dependence, and we rely once more on the data set from Genest and Rivest (1993). The uranium-caesium pair actually comes from a multivariate data set introduced in Cook and Johnson (1981) containing 6 chemical elements. More specifically, we extend the uranium-caesium pair to a three-dimensional data set including the element potassium. The reason for this choice is the fact

[(3)] Families C_8 and C_{10} are left out because of their very poor performance.

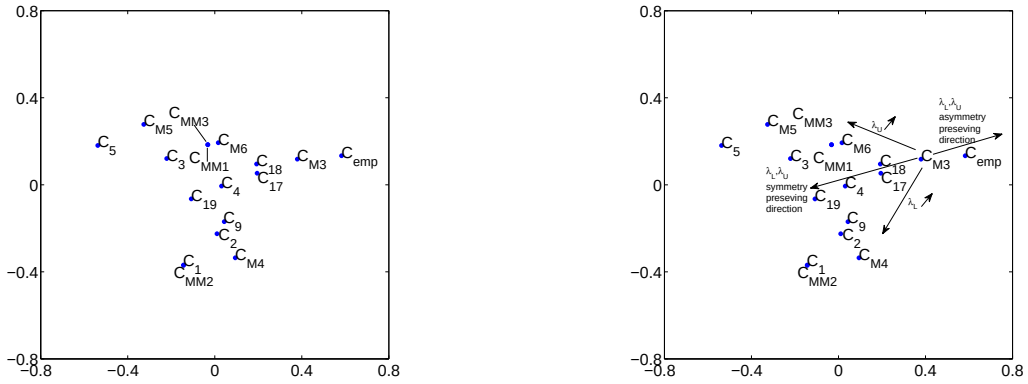


Figure 7: 2D representation of caesium-uranium-potassium fitting example. (left), 2D representation of goodness-of-fit improvement (right).

that potassium is most correlated with (uranium,caesium) among the remaining elements, with $\tau(\text{potassium}, \text{ceasium}) = 30.40\%$ and $\tau(\text{potassium}, \text{uranium}) = 13.45\%$. The fit is visalized in figure 7, left panel. The stress function yields 3.80%, indicating a very good fit.

This example exhibits a totally different goodness-of-fit picture. Note first of all that the difference between fitting results for the meta-elliptical copulas and for the other copula families is not that striking as in the previous example. In fact, an asymmetric version of the Frank copula C_{M3} yields the best fit now, although the meta-elliptical copulas come at the second place. Furthermore, the MFGM copula no longer yields a good fit, as a consequence of its limiting parameter range. Finally, the copula families entailing (asymmetric) tail dependence do not yield a good fit, and an improvement of the fit of C_{M3} should be sought by using tail-preserving and asymmetry-preserving transforms, see Figure 7, right panel.

5 Conclusion

In this contribution we introduced a new graphical copula tool which serves very well in getting an idea about the relative similarities and dissimilarities between different copulas. Therefore, it can also be used as a visual guideline in copula selection problems. The graphical tool is based on the statistical technique of multidimensional scaling, where a low-dimensional mapping of a multivariate dataset is provided. We found that a collection of bivariate copulas can be well represented in a two-dimensional space (a more than excellent fit) and the obtained distances can be explained using well-known copula properties. We also illustrate the possibilities of this graphical tool for two particular empirical examples, where we also visualized the optimization scheme based

on the concordance and tail preserving transforms, introduced in Michiels and De Schepper (2009). Finally, the visual tool can also be used to evaluate the fit of higher dimensional copulas; we illustrated this by two three-dimensional examples.

Appendix

A. Bivariate copula families

A.1 Archimedean copula families

A strict bivariate Archimedean copula is defined as a function $C : [0,1]^2 \to [0,1] : (u_1, u_2) \mapsto C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$, where the generator $\varphi : [0,1] \to [0,\infty)$ is continuous, strictly decreasing and convex with $\varphi(0) = +\infty$ and $\varphi(1) = 0$.
The λ-function for an Archimedean copula is defined by $\lambda(t) = \dfrac{\varphi(t)}{\varphi'(t)}$.

$C_{\#}$	$C_\theta(u,v)$	$\varphi_\theta(t)$	parameter range
1 (Clayton)	$[u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}$	$\varphi_\theta(t)$	$[0,\infty) \setminus \{0\}$
2 (Ali-Mikhail-Haq)	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\ln \frac{1-\theta(1-t)}{t}$	$[-1,1]$
3 (Gumbel-Hougaard)	$\exp(-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta})$	$(-\ln t)^\theta$	$[1,\infty)$
4 (Frank)	$-\frac{1}{\theta}\ln(1 + \frac{(e^{-u\theta}-1)(e^{-v\theta}-1)}{e^{-\theta}-1})$	$-\ln \frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty,\infty) \setminus \{0\}$
5 (Joe)	$1 - [(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta]^{1/\theta}$	$-\ln[1-(1-t)^\theta]$	$[1,\infty)$
6 (Gumbel-Barnett)	$uv\exp(-\theta \ln u \ln v)$	$\ln(1 - \theta \ln t)$	$(0,1]$
7	$\frac{uv}{[1+(1-u^\theta)(1-v^\theta)]^{1/\theta}}$	$\ln(2 - t^{-\theta} - 1)$	$(0,1]$
8	$(1 + [(u^{-1}-1)^\theta + (v^{-1}-1)^\theta]^{1/\theta})^{-1}$	$\left(\frac{1}{t} - 1\right)^\theta$	$[1,\infty)$
9	$\exp(1 - [(1-\ln u)^\theta + (1-\ln v)^\theta - 1]^{1/\theta})$	$(1-\ln t)^\theta - 1$	$(0,\infty)$
10	$(1 + [(u^{-1/\theta}-1)^\theta + (v^{-1/\theta}-1)^\theta]^{1/\theta})^{-\theta}$	$(t^{-1/\theta}-1)^\theta$	$[1,\infty)$
11	$\frac{1}{2}(S + \sqrt{S^2+4\theta}), S = u+v-1-\theta(\frac{1}{u}+\frac{1}{v}-1)$	$\left(\frac{\theta}{t}+1\right)(1-t)$	$[0,\infty)$
12	$(1 + \frac{[(1+u)^{-\theta}-1][(1+v)^{-\theta}-1]}{2^{-\theta}-1})^{-1/\theta} - 1$	$-\ln \frac{(1+t)^{-\theta}-1}{2^{-\theta}-1}$	$(-\infty,\infty) \setminus \{0\}$
13	$\theta/\ln(e^{\theta/u} + e^{\theta/v} - e^\theta)$	$e^{\theta/t} - e^\theta$	$(0,\infty)$
14	$[\ln(\exp(u^{-\theta}) + \exp(v^{-\theta}) - e)]^{-1/\theta}$	$\exp(t^{-\theta}) - e$	$(0,\infty)$

Table 4: Overview of the (strict) Archimedean families (with generator) included in the bivariate test space.

A.2 Other copula families

Next to the Archimedean copulas, we consider some extreme value copulas, some meta-elliptical copulas and a few other copula models.

$C_{\#}$	$C_\theta(u,v)$	parameter range
15 (Galambos)	$uv\exp\left((-\log u)^{-\theta}+(-\log v)^{-\theta}\right)^{-\frac{1}{\theta}}$	$\theta\in[0,\infty)$
16 (Husler and Reiss)	$\exp(\log(u)\Phi(\frac{1}{\theta}+\frac{1}{2}\theta\log\frac{\log u}{\log v})+\log(v)\Phi(\frac{1}{\theta}+\frac{1}{2}\theta\log\frac{\log v}{\log u}))$	$\theta\in[0,\infty)$
17 (Normal)	$\Phi_\rho(\Phi(u)^{-1},\Phi(v)^{-1})$	$\rho\in[-1,1]$
18 (Student's t)	$t_{\rho,\nu}(t_\nu^{-1}(u),t_\nu^{-1}(v))$	$\rho\in[-1,1]$
19 (Farlie-Gumbel -Morgenstern)	$uv+uv\theta(1-u)(1-v)$	$\theta\in[-1,1]$
20 (Plackett)	$\frac{(1+(\theta-1)(u+v))-\sqrt{(1+(\theta-1)(u+v))^2-4uv\theta(\theta-1)}}{2(\theta-1)}$	$\theta\in(0,\infty)$
21 (Raftery)	$\min(u,v)+\frac{1-\theta}{1+\theta}(uv)^{1/(1-\theta)}(1-[\max(u,v)]^{-(1+\theta)/(1-\theta)})$	$\theta\in[0,1]$

Table 5: Overview of the 7 other copula families included in the bivariate test space.

B. Multivariate copula families

B.1 Archimedean copula families

Exchangeable copulas, extensions from bivariate copula families.

Recall that a strict n-dimensional exchangeable Archimedean copula is defined as $C:[0,1]^n\to[0,1]:(u_1,\cdots,u_n)\mapsto C(u_1,\cdots,u_n)=\varphi^{-1}(\sum_{i=1}^n\varphi(u_i))$, where $\varphi:[0,1]\to[0,\infty)$ is continuous, strictly decreasing and convex, with $\varphi(0)=+\infty$ and $\varphi(1)=0$.
The inverse φ^{-1} has to be completely monotone, i.e. $(-1)^k\frac{d^k\varphi^{-1}(t)}{dt^k}\leq 0$ for all $k\in\mathbb{N}_0$. Copulas with this property are $C_1,C_2,C_3,C_4,C_5,C_8,C_9$ and C_{10}.

Partially exchangeable copulas, introduced by Joe (1997).

A simple multivariate generalization of the Archimedean copula in dimension d results in a dependence structure of partial exchangeability. These copulas are referred to as fully nested, since a higher dimensional copula is obtained by adding one dimension step by step.
For three dimensions, this results in

$$C(u_1,u_2,u_3)=\varphi_2^{-1}\left[\varphi_2\left(\varphi_1^{-1}\Big[\varphi_1(u_1)+\varphi_1(u_2)\Big]\right)+\varphi_2(u_3)\right].$$

In general, this copula can be written as

$$\begin{aligned}&C(u_1,\cdots,u_d)\\&=\varphi_{d-1}^{-1}\left[\varphi_{d-1}\left(\varphi_{d-2}^{-1}\Big[\varphi_{d-2}\Big(\cdots\varphi_2(\varphi_1^{-1}[\varphi_1(u_1)+\varphi_1(u_2)])+\varphi_2(u_3)]\Big)+\right.\right.\\&\left.\left.\quad\cdots+\varphi_{d-2}(u_{d-1})\Big]\right)+\varphi_{d-1}(u_d)\right].\end{aligned}\tag{9}$$

The copula defined by (9) will be a proper d-copula if, in addition to the property of complete monotonicity for the inverse functions of the generators, also the composite generators $\varphi_{i+1}\circ\varphi_i^{-1}$ are completely monotone.

For the three-dimensional case, Joe (1997) section 5.3 proposes to use four models of the form $C_{\theta_1}(u_3, C_{\theta_2}(u_1, u_2))$ with $\theta_2 \geq \theta_1$.

$C_\#$	$C_{\theta_1}(u_3, C_{\theta_2}(u_1, u_2))$	$\theta_2 \geq \theta_1$ range	$\tau_{12}, \tau_{13}, \tau_{23}$ range
M3	$-\theta_1^{-1}\log\{1-(1-e^{-\theta_1})^{-1}(1-[1-(1-e^{-\theta_2})^{-1}(1-e^{-\theta_2 u_1})$ $\cdot(1-e^{-\theta_2 u_2})]^{\theta_1/\theta_2})(1-e^{-\theta_1 u_3})\}$	$[0,\infty)$	$[0,1]$
M4	$[(u_1^{-\theta_2}+u_2^{-\theta_2}-1)^{\theta_1/\theta_2}+u_3^{-\theta_1}-1]^{-1/\theta_1}$	$[0,\infty)$	$[0,1]$
M5	$1-\{[(1-u_1)^{\theta_2}(1-(1-u_2)^{\theta_2})+(1-u_2)^{\theta_2}]^{\theta_1/\theta_2}\cdot$ $\cdot(1-(1-u_3)^{\theta_1})+(1-u_3)^{\theta_1}\}^{1/\theta_1}$	$[1,\infty)$	$[0,1]$
M6	$\exp\{-([(-\log u_1)^{\theta_2}+(-\log u_2)^{\theta_2}]^{\theta_1/\theta_2}+(-\log u_3)^{\theta_1})^{1/\theta_1}\}$	$[1,\infty)$	$[0,1]$

Table 6: Overview of the 4 partially exchangeable copula families included in the trivariate test space.

Note that model C_{M3} is an extension of C_4, C_{M4} is an extension of C_1, C_{M5} is an extension of C_5 and C_{M6} is an extension of C_3.

B.2 Multivariate Extreme value families

We use three multivariate extreme value families $C_{MM1}, \cdots, C_{MM3}$ introduced in Joe (1997), section 5.5. Note that the extreme value copulas are defined for positive dependence, implying $\tau \in [0,1]$.

For the trivariate case we have

- $C_{MM1}(u_1,u_2,u_3) = \exp\Big\{-\Big[((p_1 z_1^\theta)^{\delta_{12}}+(p_2 z_2^\theta)^{\delta_{12}})^{1/\delta_{12}}+((p_1 z_1^\theta)^{\delta_{13}}+(p_3 z_3^\theta)^{\delta_{13}})^{1/\delta_{13}}$
$+((p_2 z_2^\theta)^{\delta_{23}}+(p_3 z_3^\theta)^{\delta_{23}})^{1/\delta_{23}}+\nu_1 p_1 z_1^\theta+\nu_2 p_2 z_2^\theta+\nu_3 p_3 z_3^\theta\Big]^{1/\theta}\Big\}$,
where $\theta \geq 1$, $\delta_{ij} \geq 1$, $\nu_j \geq 0$, $p_j = (\nu_j+2)^{-1}$ and $z_j = -\log u_j$.
- $C_{MM2}(u_1,u_2,u_3) = \Bigg[\Bigg(u_1^{-\theta}+u_2^{-\theta}+u_3^{-\theta}\Bigg)-2-\Bigg((\hat{u}_1^{-\delta_{12}}+\hat{u}_2^{-\delta_{12}})^{-1/\delta_{12}}$
$+(\hat{u}_1^{-\delta_{13}}+\hat{u}_3^{-\delta_{13}})^{-1/\delta_{13}}+(\hat{u}_2^{-\delta_{23}}+\hat{u}_3^{-\delta_{23}})^{-1/\delta_{23}}\Bigg)\Bigg]^{-1/\theta}$,
where $\theta > 0$, $\delta_{ij} > 0$, $\nu_j \geq 0$, $p_j = (\nu_j+2)^{-1}$ and $\hat{u}_j = p_j(u_j^{-\theta}-1)$.

  Note that the special case $[u_1^{-\theta}+u_2^{-\theta}+u_3^{-\theta}-2]^{-1/\theta}$ (C_1) arises for p_1, p_2, p_3 reaching 0.

- $C_{MM3}(u_1,u_2,u_3) = \exp\Bigg\{-\Big[\Big(z_1^\theta + z_2^\theta + z_3^\theta\Big) - \Big((p_1^{-\delta_{12}} z_1^{-\theta\delta_{12}} + p_2^{-\delta_{12}} z_2^{-\theta\delta_{12}})^{-1/\delta_{12}}$

  $+(p_1^{-\delta_{13}} z_1^{-\theta\delta_{13}} + p_3^{-\delta_{13}} z_3^{-\theta\delta_{13}})^{-1/\delta_{13}} + (p_2^{-\delta_{23}} z_2^{-\theta\delta_{23}} + p_3^{-\delta_{23}} z_3^{-\theta\delta_{23}})^{-1/\delta_{23}}\Big)\Big]^{1/\theta}\Bigg\}$,

  where $\theta > 1$, $\delta_{ij} > 0$, $\nu_j \geq 0$, $p_j = (\nu_j + 2)^{-1}$ and $z_j = -\log u_j$.

Note that family C_{MM1} can be seen as an extension of C_3, while C_{MM2} and C_{MM3} are extensions of C_{15}.

B.3 Meta-elliptical families

The meta-elliptical families can straightforwardly be extended to any dimension $n \geq 2$.

B.4 MFGM family

We use the extension described in Nelsen (2006), page 108. In the trivariate case the extended copula can be written as

$$C_{MFGM}(u_1,u_2,u_3) = u_1u_2u_3[1 + \theta_{12}(1-u_1)(1-u_2) + \theta_{13}(1-u_1)(1-u_3) + \theta_{23}(1-u_2)(1-u_3) + \theta_{123}(1-u_1)(1-u_2)(1-u_3)]$$

where $\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123} \in [-1,1]$, which implies τ_{12}, τ_{13} and $\tau_{23} \in [-2/9, 2/9]$.

C. Empirical copula

The empirical copula in d dimensions is a rank-based estimator calculated as

$$C_{emp}(u_1,\ldots,u_d) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_{i1} \leq u_1, \ldots, U_{id} \leq u_d)\ .$$

References

I. Borg and P.J.F. Groenen. *Modern multidimensional scaling.* Springer, Series in Statistics, 1997.

R. Dennis Cook and Mark E. Johnson. A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43:210–218, 1981.

Hong-Bin Fang and Kai-Tai Fang. The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82:1–16, 2002.

E.W. Frees and E.A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2(1), 1998.

C. Genest and L.-P. Rivest. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043, 1993.

C. Genest, K. Ghoudi, and L.-P. Rivest. Discussion of understanding relationships using copulas". *North American Actuarial Journal*, 2:143–149, 1998.

C. Genest, M. Gendron, and M. Bourndeau-Brien. The advent of copulas in finance. *The European Journal of Finance*, 0(00):1–10, 2009.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman and Hall, 1997. ISBN 0-412-07331-5.

R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall Inc., 1992.

J.B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27, 1964a.

J.B. Kruskal. Nonmetric multidimensional scaling : a numerical method. *Psychometrika*, 29: 115–129, 1964b.

J.B. Kruskal and M. Wish. *Multidimensional scaling*. Sage Publications, Series in Quantitative Applications in the Social Sciences, 1978.

A.J. McNeil and J. Nešlehová. Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *Annals of Statistics*, 37:3059–3097, 2009.

F. Michiels and A. De Schepper. Understanding copula transforms: A review of dependence properties. *Research paper University of Antwerp, Faculty of Applied Economics*, 2009:012, 2009.

F. Michiels and A. De Schepper. A copula test space model: How to avoid the wrong copula choice. *Kybernetika*, 44(6):864–878, 2008.

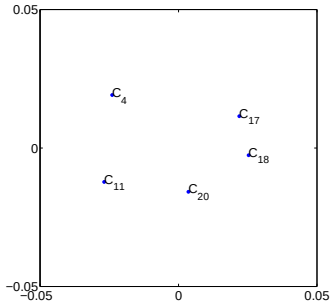
F. Michiels, I. Koch, and A. De Schepper. A new method for the construction of bivariate Archimedean copulas based on the λ-function. *Communications in Statistics: Theory and Applications*, accepted for publication, 2010.

R.B. Nelsen. *An Introduction to Copulas*. Springer-Verlag, 2006.

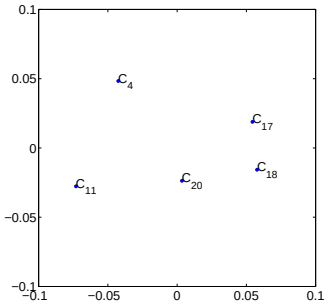
A. K. Nikoloulopoulos and D. Karlis. Copula model evaluation based on parametric bootstrap. *Computational Statistics and Data Analysis*, 52:3342–3353, 2008.

A. Sklar. Fonctions de repartition n dimensions et leurs marges. *publications de l'Institut de Statistique de l'Universit de Paris*, 8:229–231, 1959.

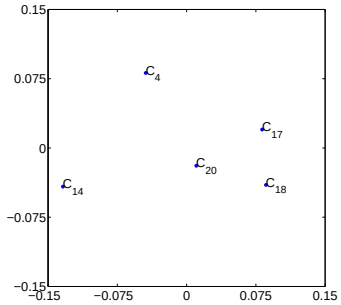
Figure 8: 2D mapping of bivariate comparable test spaces for $\tau < 0$.



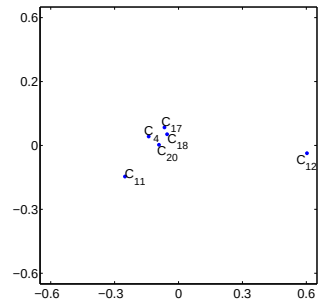
(a) $\tau = -90\%$



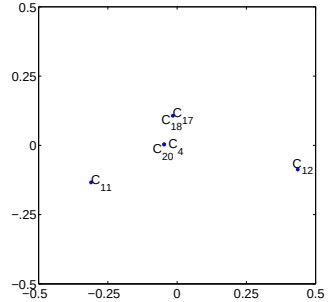
(b) $\tau = -80\%$



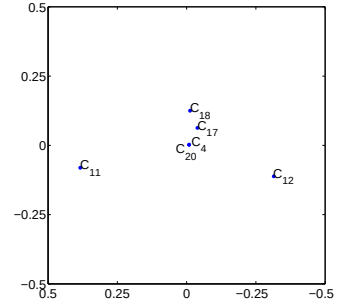
(c) $\tau = -70\%$



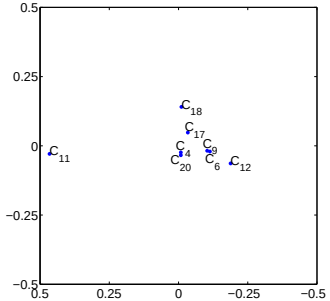
(d) $\tau = -60\%$



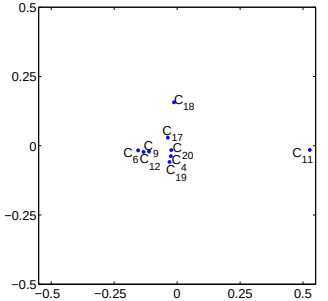
(e) $\tau = -50\%$



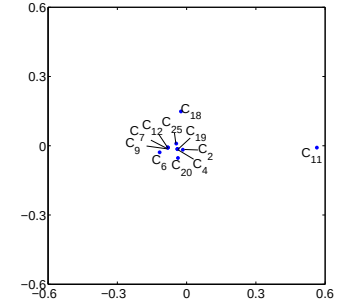
(f) $\tau = -40\%$



(g) $\tau = -30\%$

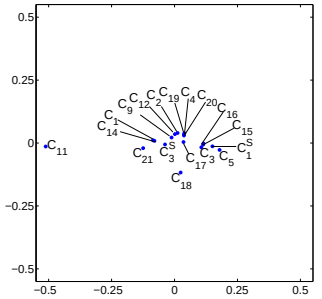


(h) $\tau = -20\%$

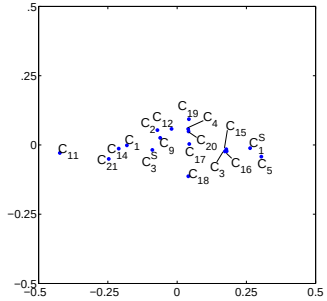


(i) $\tau = -10\%$

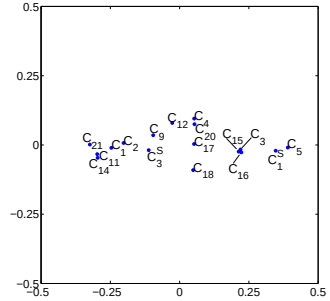
Figure 9: 2D mapping of bivariate comparable test spaces for $\tau > 0$.



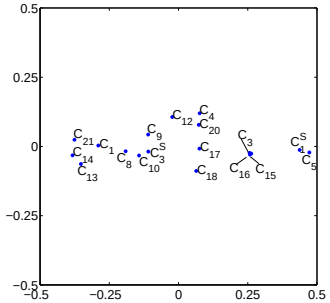
(a) $\tau = 10\%$



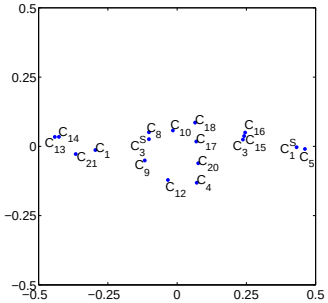
(b) $\tau = 20\%$



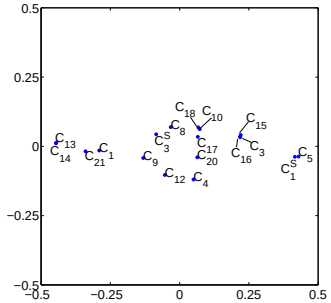
(c) $\tau = 30\%$



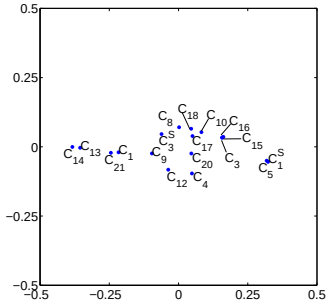
(d) $\tau = 40\%$



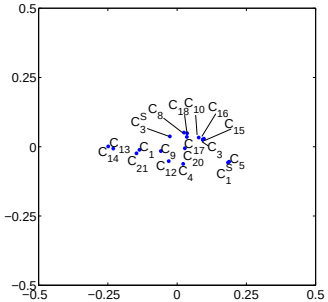
(e) $\tau = 50\%$



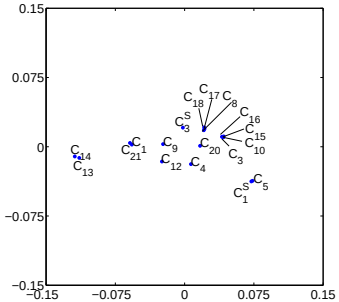
(f) $\tau = 60\%$



(g) $\tau = 70\%$



(h) $\tau = 80\%$



(i) $\tau = 90\%$