DEPARTMENT OF MATHEMATICS,
STATISTICS AND ACTUARIAL SCIENCES

A Copula Test Space Model: How To Avoid the Wrong Copula Choice

Frederik Michiels & Ann De Schepper

UNIVERSITY OF ANTWERP
Faculty of Applied Economics
Stadscampus
Prinsstraat 13, B.213
BE-2000 Antwerpen
Tel. +32 (0)3 220 40 32
Fax +32 (0)3 220 47 99
http://www.ua.ac.be/tew

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF MATHEMATICS,
STATISTICS AND ACTUARIAL SCIENCES

A Copula Test Space Model:
How To Avoid the Wrong Copula Choice

Frederik Michiels & Ann De Schepper

RESEARCH PAPER 2007-027
JUNE 2008

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.213
phone: (32) 3 220 40 32
fax: (32) 3 220 47 99
e-mail: joeri.nys@ua.ac.be

D/2007/1169/027

Universiteit Antwerpen

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCES

A Copula Test Space Model: How To Avoid the Wrong Copula Choice

Frederik Michiels and Ann De Schepper

June 19, 2008

Research Paper RPS–2007–027

A Copula Test Space Model: How To Avoid The Wrong Copula Choice

Frederik Michiels and Ann De Schepper
University of Antwerp*

We introduce and discuss the test space problem as a part of the whole copula fitting process. In particular, we explain how an efficient copula test space can be constructed by taking into account information about the existing dependence. Although our model is developed in a bivariate environment it can be used for higher dimensional copula fitting applications. This is shown on the 3 dimensional dependence structure of an illustrative porfolio containing the S&P 500 Composite Index, the JP Morgan Government Bond Index and the NAREIT All index.

Keywords: copula, Kendall's τ, goodness-of-fit
AMS Subject Classification: 62H20, 62P05, 62H12

1. INTRODUCTION

In recent years copulas have been hailed in the literature as promising modeling tools able to relax assumptions with regard to the distributional aspects of a multivariate problem. They have been succesfully applied in different fields of research such as biostatistics, e.g. [4], geostatistics, e.g. [6], finance, e.g. [7], decision theory, e.g. [9], insurance, e.g. [16] and hydrology, e.g. [17].

A 'copula' is a dependence function, a mathematical expression which allows modeling the dependence structure between stochastic variables. As such, copulas can be used to construct multivariate distribution functions. The main advantage of the copula approach in this matter is that it is able to split the problem up into a part containing the marginal distribution functions and a part containing the dependence structure. These two parts can be studied and estimated seperately and can then be rejoined to form a multivariate distribution function.

However, the downside of the copula method is that it shifts the problem of identifying the right multivariate distribution function from having a too restrictive solution space to having a vast amount of possible solutions. Indeed, when tackling a dependence modeling problem with copulas, one first has to identify an initial copula space (test space problem) and secondly one has to measure the goodness-of-fit of that space (goodness-of-fit problem). Although it is obvious that these two sub problems are equally important, the first problem has not yet received the attention it deserves in the literature. As a consequence, usually only a small number of recurring copulas is used in a fitting application.

This contribution aims to discuss the test space problem and to tackle a first step in creating an *efficient* copula test space, by making a test space *comparable*, with which we mean that the test space consists of comparable copulas. Although we explain our

*Faculty of Applied Economics, Departement of mathematics, Statistics and Actuarial Sciences

methodolgy for bivariate copulas, it can be easily extented to multivariate fitting applications through the compound method.

The paper is organized as follows. In section 2 we recall basic concepts of copula theory. Section 3 is the main part of this paper. We first encompass some important aspects of the test space problem as well as the pitfalls related to them. Then we elaborate the idea on how to make a test space comparable. In section 4 we illustrate our method by means of an application on financial data, and finally section 5 concludes.

2. BASIC CONCEPTS

We start with the definition and the most important theorem for copulas.

Definition 1 *A bivariate copula is a function* $C : \mathbf{I^2} \to \mathbf{I}$ *with the following properties:*

1. *C is 2-increasing, or for all* $u_1 \leq u_2, v_1 \leq v_2 \in \mathbf{I}$ *it is true that* $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.
2. *C is grounded, or* $C(u, 0) = C(0, v) = 0$ *for all* $(u, v) \in \mathbf{I^2}$.
3. *C has uniform* $[0, 1]$ *margins, or* $C(u, 1) = u$ *and* $C(1, v) = v$ *for all* $(u, v) \in \mathbf{I^2}$.

In fact a bivariate copula represents the link between the one dimensional marginal distribution functions and their bivariate aggregate. This link can be formalized through the theorem of Sklar [25].

Theorem 1 *Let H be a bivariate joint distribution function with margins F and G. Then there exists a copula C in such a way that* $H(x, y) = C(F(x), G(y))$ *for all* $(x, y) \in \bar{\mathbb{R}}$.

If F en G are continuous, then C is unique. If not, then C is unique on imF×imG. Conversely, if C is a copula and F and G are distribution functions, then the joint distribution function H is defined as indicated above.

Sklar's theorem withholds two important facts which are of great value to dependence modeling. The first one includes the observation that copulas facilitate the construction of joint distribution functions, in the sense that any combination of margins can be chosen to build their joint aggregate. The second one entails the observation that any joint distribution function can be split up into a part only containing information related to the respective variables, the margins, and into a part which captures the dependence structure inherent to the distribution function, the copula.

An important issue when using copulas in a fitting application, is that they are characterized by one or more parameters, allowing them to account for a certain dependence range. In this respect, the following definition is noteworthy.

Definition 2 *A copula is called comprehensive if it includes the maximum copula* $M(u, v) = min(u, v)$, *the minimum copula* $W(u, v) = max(u+v-1, 0)$ *and the independence copula* $\Pi(u, v) = uv$ *for all* $(u, v) \in \mathbf{I^2}$

There are only a handful of one parameter copulas who have the property of being comprehensive. All other copulas are only capable of describing a certain part of the dependence range often with M or W as a limiting case. For example in the list of Archimedean copulas in [23] only two have the property of comprehensiveness. We will come back to this fact in section 3, when constructing our test space model.

To end this section we will discuss two important copula classes. The first class, called Archimedean copulas, are characterized by a generator φ. This is a function which facilitates the construction, parameter estimation and simulation of this copula class. Due to this property, Archimedean copulas form a rather popular copula class and find a wide range of applications. The copula generator φ is defined in the following way.

Definition 3 *A generator φ is a continuous, strictly decreasing convex function defined on $\boldsymbol{I}$ and image $[0, +\infty)$. If $\varphi(0) = +\infty$ then the generator is called strict.*

Related to the concept of the generator is its pseudo-inverse, which is defined as follows.

Definition 4 *The pseudo-inverse of φ is the function $\varphi^{[-1]}$ with support $\varphi^{[-1]} = [0, +\infty)$ and image $\varphi^{[-1]} = \boldsymbol{I}$ given by*

$$\varphi^{[-1]} = \begin{cases} \varphi^{[-1]}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t < +\infty \end{cases} .$$

An Archimedean copula is then constructed in the following way.

Definition 5 *A bivariate Archimedean copula with generator φ is the function C : $\mathbf{I^2} \to \mathbf{I}$ defined as:*

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)).$$

The main characteristics the Archimedean class are the following.

Property 1 *Let C be an Archimedean copula with generator φ. Then it has the following properties:*

1. *C is symmetric; i.e. $C(u, v) = C(v, u)$ for all u,v in $\mathbf{I}$.*
2. *C is associative; i.e. $C(C(u, v), w) = C(u, C(v, w))$ for all u,v in $\mathbf{I}$.*
3. *If $c > 0$ is any constant, then $c \cdot \varphi$ is also a generator of C.*

Note that the set of Archimedean copulas is a very large set, see e.g. [23].

Next to the Archimedean copula class, a second important class is the class of meta-elliptical copulas. This class consists of copula families derived from elliptically contoured distributions, such as symmetric Kotz type distributions, symmetric Pearson VII type distributions and symmetric Pearson II type distributions. Well known members are the Normal, Student's t and Cauchy copulas. The class is defined in the following way.

Definition 6 *Let the bivariate vector $Z = (Z_1, Z_2)' \sim EC_2(\mathbf{0}, \mathbf{R}, g)$. Then the copula of Z_1 and Z_2 is given by*

$$C_Z(u, v) = \frac{1}{\sqrt{(1-\rho^2)}} \int_{-\infty}^{Q_g^-1(u)} \int_{-\infty}^{Q_g^-1(v)} g\left(\frac{x^2 + y^2 + 2\rho xy}{\sqrt{1-\rho^2}}\right) dxdy$$

where $\mathbf{R}$ is the bivariate correlation matrix, $\mathbf{R}$= $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ where g is a weighting function and $Q_g(z)$ is the marginal CDF.

More detailed information about this class can be found e.g. in [14].

In the following section, these two large classes of copulas will be used as the major part of our copula test space.

3. TEST SPACE MODEL

In this section, we will first identify (see subsection 3.1) what we believe are the necessary aspects of a good test space. Secondly (subsection 3.2), we will explain how in a specific fitting problem such a test space can be constructed.

Without defining an efficient test space, a fitting application can have an acceptable result, but it is not sure at all that it will result in the best possible estimate for the unknown underlying dependence structure. We are confident that *with* an efficient test space, many modeling problems will end up with a better fit for the applications at hand.

3.1. The test space problem

We first state the concept of a copula test space.

Definition 7 *A copula test space* $\tilde{\mathbf{C}}$ *is a finite subset of the set* $\mathbf{C}$ *of all possible copulas of which will be chosen from to compare the goodness-of-fit in a fitting application.*

Any attempt to construct such a test space can be regarded as the test space problem. The following three aspects should be taken into account when creating a test space:

- size (the number of copulas entering the test space),
- diversity (the various copula properties present in the test space) ,
- relevance (only comparable copulas enter the test space).

The first and second objective, a sufficiently large and diverse test space, are somewhat related. They are both of great importance as they will increase the possibility of a good fit and therefore also improve the understanding of the true underlying dependence structure. Using a large test space is usually not done in the literature. We found only a few examples dealing with copula selection from a large test space. See for example [10] for an application with Archimedean copulas and [17] for an application with various copula classes.

Especially in the financial literature small test spaces are used, usually containing 3 to 5 copula families. Although many papers contain very nice results, we strongly believe that the results would be even better if the fitting would have started from a larger and more diverse test space. In particular, the Gumbel-Hougaard, Frank and Clayton family* and the Gaussian and Student's t family are frequently used in examples or in fitting applications (see e.g. [1], [8], [3], [5], [6], [21] or [12]). However, the motivation for the particular selection of copula families is not always clear. Most of the time, the choice is based on their wide usage in the literature, but it also happens that no motivation is given at all.

Closely related to what precedes, but yet different, the third aspect of the test space problem concerns efficient copula testing, where the goal is to select the relevant copulas from a copula space. With relevant copulas we mean copulas that are effective to describe a given dependence structure, in terms of their parameter range. One option of course is to use only comprehensive copulas, which can describe the whole dependence space, but as this approach limits the modelers' possibilities this will reduce the likelihood of a good fit, as the first and second objective cannot be met. The opposite is to test a large copula sample without taking into account any a priori information. This, however, may lead to inefficiencies, since not all copulas cover the same dependence range. A third

*The names are used as in [23].

possibility, which we prefer, is to include some information concerning the dependence structure into the construction of the copula test space. In this way, only really relevant copulas are selected. Although this approach in the literature is often used as a criterion for choosing *within* a test space, it has never been used as a strategy for *constructing* it.

Indeed, generally the copula test space is constructed *before* the dependence structure is examined. When e.g. the test space includes the Gumbel-Hougaard or Joe family, it is implicitly assumed that the data to be modeled are positive dependent, since both copulas are only able to describe this kind of dependence. A test on positive quadrant dependency, see e.g. [11], or any a priori knowledge of the dependence structure of the data could ascertain this assumption. When the assumption does not hold, the choice of the subset can lead to fitting an inadequate copula. For example in [26], the fit could have been better when another copula than the Joe family was used, as there was negative dependence in the data of the fitting application.

When constructing multivariate families using the compound method, the test space problem becomes even more apparent, as all generators φ must lie in the parameter range. In [21] e.g. the trivariate Gumbel-Hougaard family was selected before the dependence structure was examined; as the data contain negative dependencies-which is not covered by Gumbel-Hougaard- the fit could be improved by starting from an efficient test space.

The above shows that some a priori knowledge concerning the true dependence structure can be a first and important step towards an efficient test space, by eliminating copulas with an unappropriate dependence range. In the next subsection, we will present a possible way of creating universal comparable test spaces which can be used for fitting applications in different fields of research.

3.2. Creating a large comparable test space

We first focus on the size of $\tilde{\mathbf{C}}$. Clearly $\mathbf{C}$ is unknown and we have to rely on existing classes and families in the literature (especially those described in [23] and [20]). In order to keep our analysis synoptical, at this moment we limit ourselves to one parameter copula families. We gather 29 of such families from different classes:

- Archimedean class: 22 Archimedean families (including the popular Clayton, Frank, Gumbel-Hougaard, Joe and Ali-Mikhael-Haq families);
- Extreme value families: Galambos and Husler Reiss family (also Gumbel-Hougaard family);
- Meta-elliptical class: Normal, Student t family (with as special case the Cauchy family if df=1);
- Other families: Farlie-Gumbel-Morgenstern family, Plackett family and Raftery family.

Summarizing, this means that we suggest to construct a test space of up to 29 copula families.

The next step consists of making the test space comparable by choosing the relevant copulas. This is essential in order to know which copulas can be compared and which can not. We suggest to do this by means of a powerfull and well-known measure of concordance, namely Kendall's τ.

The general expression of the relationship between τ and a copula C is:

$$\tau = 4 \int\int_{I^2} C(u,v) dC(u,v) - 1.$$

#	$C_\theta(u,v)$	$\varphi(t)$	$\theta \in$	$\tau \in$
1 (Clayton)	$\max([u^{-\theta}+v^{-\theta}-1],0)^{-1/\theta}$	$\frac{1}{\theta}((t^{-\theta}-1)$	$[-1,\infty)\setminus\{0\}$	[-1,1]
2	$max(1-[(1-u)^\theta+(1-v)^\theta]^{1/\theta},0)$	$(1-t)^\theta$	$[1,\infty)$	$[-1,1]$
3 (Ali-Mikhail-Haq)	$\frac{uv}{1-\theta(1-u)(1-v)}$	$ln\frac{1-\theta(1-t)}{t}$	$[-1,1]$	$[\frac{1}{3}(5-8ln2),\frac{1}{3}]$
4 (Gumbel-Hougaard)	$exp(-[(-lnu)^\theta+(-lnv)^\theta]^{1/\theta})$	$(-lnt)^\theta$	$[1,\infty)$	$[0,1]$
5 (Frank)	$-\frac{1}{\theta}\ln(1+\frac{(e^{-u\theta}-1)(e^{-v\theta}-1)}{e^{-\theta}-1}$	$-ln\frac{e^{-t\theta}-1}{e^{-\theta}-1}$	$(-\infty,\infty)\setminus\{0\}$	$[-1,1]^{\mathrm{b}}$
6 (Joe)	$1-[(1-u)^\theta+(1-v)^\theta-(1-u)^\theta(1-v)^\theta]^{1/\theta}$	$-ln[1-(1-t)^\theta]$	$[1,\infty)$	$[0,1]^*$
7	$max(\theta uv+(1-\theta)(u+v-1),0)$	$-ln[\theta t+(1-\theta)]$	$(0,1]$	$[-1,0]^*$
8	$max[\frac{\theta^2uv-(1-u)(1-v)}{\theta^2-(\theta-1)^2(1-u)(1-v)}\ ,0]$	$\frac{1-t}{1+(\theta-1)t}$	$[1,\infty)$	$[-1,\frac{1}{3}]$
9 (Gumbel-Barnett)	$uvexp(-\theta lnulnv)$	$ln(1-\theta lnt)$	$(0,1]$	$[-0.3613,0]^*$
10	$\frac{uv}{[1+(1-u^\theta)(1-v^\theta)]^{1/\theta}}$	$ln(2t^{-\theta}-1)$	$(0,1]$	$[-0.1817,0]^*$
11	$max([u^\theta v^\theta-2(1-u^\theta)(1-v^\theta)]^{1/\theta},0)$	$ln(2-t^\theta)$	$(0,1/2]$	$[-0.5649,1]^*$
12	$(1+[(u^{-1}-1)^\theta+(v^{-1}-1)^\theta]^{1/\theta})^{-1}$	$(\frac{1}{t}-1)^\theta$	$[1,\infty)$	$[\frac{1}{3},1]$
13	$exp(1-[(1-lnu)^\theta+(1-lnv)^\theta-1]^{1/\theta})$	$(1-lnt)^\theta-1$	$(0,\infty)$	$[-0.3613,1]^*$
14	$(1+[(u^{-1/\theta}-1)^\theta+(v^{-1/\theta}-1)^\theta]^{1/\theta})^{-\theta}$	$(t^{-1/\theta}-1)^\theta$	$[1,\infty)$	$[\frac{1}{3},1]$
15 (Genest-Ghoudi)	$max(1-[(1-u^{1/\theta})^\theta+(1-v^{1/\theta})^\theta]^{1/\theta},0)^\theta$	$(1-t^{1/\theta})^\theta$	$[1,\infty)$	$[-1,1]$
16	$\frac{1}{2}\ (S+\sqrt{S^2+4\theta}),S=u+v-1-\theta(\frac{1}{u}+\frac{1}{v}-1)$	$(\frac{\theta}{t}+1)(1-t)$	$[0,\infty)$	$[-1,0.3333]$
17	$(1+\frac{[(1+u)^{-\theta}-1][(1+v)^{-\theta}-1]}{2^{-\theta}-1})^{-1/\theta}-1$	$-ln\frac{(1+t)^{-\theta}-1}{2^{-\theta}-1}$	$(-\infty,\infty)\setminus\{0\}$	$[-0.6109,1]^*$
18	$max(1+\theta/ln[e^{\theta/(u-1)}+e^{\theta/(v-1)}],0)$	$e^{\theta/(t-1)}$	$[2,\infty)$	$[\frac{1}{3},1]$
19	$\theta/ln(e^{\theta/u}+e^{\theta/v}-e^\theta)$	$e^{\theta/t}-e^\theta$	$(0,\infty)$	$[-0.3333,1]^*$
20	$[ln(exp(u^{-\theta})+exp(v^{-\theta})-e)]^{-1/\theta}$	$exp(t^{-\theta})-e$	$(0,\infty)$	$[0,1]^*$
21	$1-(1-max([1-(1-u)^\theta]^{1/\theta}+[1-(1-v)^\theta]^{1/\theta}-1,0)^\theta)^{1/\theta}$	$1-[1-(1-t)^\theta]^{1/\theta}$	$[1,\infty)$	$[-1,1]^*$
22	$max([1-(1-u^\theta)\sqrt{1-(1-v^\theta)^2}-(1-v^\theta)\sqrt{1-(1-u^\theta)^2}]^{1/\theta},0)$	$arcsin(1-t^\theta)$	$(0,1]$	$[-0.4674,0]^*$

Table 1: parameter ranges in terms of Kendall's τ for Archimedean families

#	$C_\theta(u,v)$	$\theta \in$	$\tau \in$
23 (Galambos)	$uv\exp\left((-\log u)^{-\theta}+(-\log v)^{-\theta}\right)^{-\frac{1}{\theta}}$	$[0,\infty)$	$[0,1]$
24 (Husler and Reiss)	$\exp(\log(u)\Phi(\frac{1}{\theta}+\frac{1}{2}\theta\log\frac{\log u}{\log v})+\log(v)\Phi(\frac{1}{\theta}+\frac{1}{2}\theta\log\frac{\log v}{\log u}))$	$[0,\infty)$	$[0,1]^*$

Table 2: parameter ranges in terms of Kendall's τ for extreme value families

For Archimedean copulas, this relationship can be rewritten in terms of the copula generator $\varphi(t)$, the first derivative $\varphi'(t)$ and τ (see [19]):

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,\mathrm{d}t\,;$$

for Meta-elliptical families, expression (1) can be reduced to (see [14, 22])

$$\tau = \frac{2}{\pi}arcsin(\rho).$$

The relationship between the copula parameter and τ needs analytical or numerical solving, depending on the respective copula. Tables 1 to 4 show the results for the range for Kendall's tau that were obtained for the 29 one parameter families mentioned earlier. The parameter values are displayed with a 4 digit accuracy. The ranges marked with an asterisk are obtained by numerical integration of the ratio $\frac{\varphi(t)}{\varphi'(t)}$, with φ the generator of the corresponding Archimedean copula.

#	$C_\theta(u,v)$	$\rho \in$	$\tau \in$
25 (Normal)	$\Phi_\rho(\Phi(u)^{-1}, \Phi(v)^{-1})$	$[-1,1]$	[-1,1]
26 (Student's t)	$t_{\rho,v}(t_v^{-1}(u), t_v^{-1}(v))$	$[-1,1]$	[-1,1]

Table 3: parameter ranges in terms of Kendall's τ for meta-elliptical families

#	$C_\theta(u,v)$	$\theta \in$	$\tau \in$
27 (Farlie-Gumbel-Morgenstern)	$uv + uv\theta(1-u)(1-v)$	$[-1,1]$	[-2/9,2/9]
28 (Placett)	$\frac{(1+(\theta-1)(u+v))-\sqrt{(1+(\theta-1)(u+v))^2-4uv\theta(\theta-1)}}{2(\theta-1)}$	$(0,\infty)$	$[-1,1]$
29 (Raftery)	$min(u,v) + \frac{1-\theta}{1+\theta}(uv)^{1/(1-\theta)}(1-[max(u,v)]^{-(1+\theta)/(1-\theta)})$	$[0,1]$	$[0,1]^*$

Table 4: parameter ranges in terms of Kendall's τ for other families

Two remarks should be made with respect to these results. First of all, some parameter ranges include $\tau = 0$ but cannot describe the independence case (i.e. for copulas #2, #8, #15, #16, #21 and #26 for df ≤ 2). Secondly, one should take into account that not all copulas are positively ordered. Some of them, like copula #9, #10 and #11, are negatively ordered, which means that a positive increase in their parameter value implies a negative increase in the tau value. In our analysis we start from a commonly (positively) ordered tau range which we then use to calculate the respective copula parameters.

Based on the outcomes as presented in Tables 1 to 4, we are now able to reorganize the initial copula space into intervals according to specific dependence ranges, marked out in terms of Kendall's tau values. The result are summarized in Table 5. The table shows a restructuring of the copula space into 13 dependence intervals, where each entry contains 10 to 25 different copulas. If e.g. the data reveal an estimate for Kendall's tau of about -0.5, the comparable copula test space consists of Archimedean copulas #1, #2, #5, #7, #8, #11, #15, #17 and #21, and of the meta-elliptical copulas, but from the extreme value copulas and other copulas, none is qualified for fitting purposes since extreme value copulas always correspond to situations with a non-negative τ value.

Note that there is a significant difference between the negative and the positive dependence space. In general, more copulas can be used to describe positive dependence, whereas a smaller number of copulas can be used to describe strong negative dependence.

The test spaces in Table 5 allow us to solve fitting problems in a more efficient way, since we now have a clear overview of which copula families can be compared for a given degree of dependence. In the bivariate case the practical use of Table 5 is straightforward. After the dependence in the data is examined by means of Kendall's τ, it is possible to choose from the table a comparable test space corresponding to the estimated degree of dependence. Acting in this way, fitting errors will be minimized.

The practical use of Table 5 can also be extended to a multivariate ($d > 2$) framework. This will be illustrated in the example in section 4.

4. EXAMPLE — APPLICATION

In the present section, we will show how the previous results can be used in a multivariate framework. This will be carried out by means of the hierarchical Archimedean copula technique [24]. Note that as a consequence of choosing this technique, the copula space of 29 families is reduced to the first 22 copula families.

We have chosen to work with the same data as used in [21], and to fit a 3-dimensional model to the dependence structure determined by 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‡. Among other results, Kole et al. wanted to test whether the 3-dimensional dependence structure in these data could be described by the survival version of the Gumbel-Hougaard family.

‡In the paper of [21], the period covered in the example ended on December 17 2004.

$\tau \in$	$[-1, -0.6109)$	$[-0.6109, -0.5649)$	$[-0.5649, -0.4674)$	$[-0.4674, -0.3613)$	$[-0.3613, -0.3333)$	$[-0.3333, -0.2222)$	$[-0.2222, -0.1817)$	$[-0.1817, 0)$	0	$[0, .2222)$	$(.2222, 0.3333)$	0.3333	$[0.3333, 1]$
#1	√	√	√	√	√	√	√	√	√	√	√	√	√
#2	√	√	√	√	√	√	√	√	√	√	√	√	√
#3								√	√	√	√	√	
#4									√	√	√	√	√
#5	√	√	√	√	√	√	√	√	√	√	√	√	√
#6									√	√	√	√	√
#7	√	√	√	√	√	√	√	√	√				
#8	√	√	√	√	√	√	√	√	√	√	√	√	
#9						√	√	√	√				
#10								√	√				
#11			√	√	√	√	√	√	√				
#12												√	√
#13					√	√	√	√	√	√	√	√	√
#14												√	√
#15	√	√	√	√	√	√	√	√	√	√	√	√	√
#16	√	√	√	√	√	√	√	√	√	√	√	√	
#17		√	√	√	√	√	√	√	√	√	√	√	√
#18												√	√
#19												√	√
#20									√	√	√	√	√
#21	√	√	√	√	√	√	√	√	√	√	√	√	√
#22				√	√	√	√	√	√				
#23									√	√	√	√	√
#24									√	√	√	√	√
#25	√	√	√	√	√	√	√	√	√	√	√	√	√
#26	√	√	√	√	√	√	√	√	√	√	√	√	√
#27							√	√	√	√			
#28	√	√	√	√	√	√	√	√	√	√	√	√	√
#29									√	√	√	√	√

Table 5: Restructuring the copula space in terms of Kendall's τ. Copulas #1 to #22 correspond to the Archimedean copulas, copulas #23 to #24 to extreme value copulas, copulas #25 to #26 to meta-elliptical copulas, and copulas #27 to #29 to the other copulas.

After obtaining the empirical concordance matrix, see Table 6, it can be explained why the fit of the Gumbel-Hougaard family is not as good as possibly expected. Indeed, since Archimedean families are *associative*, it is necessary to have for the respective copulas $C_{\tau_{sr(b)}}(C_{\tau_{sr}}(u_s, u_r), u_b) = C_{\tau_{rb(s)}}(u_s, C_{\tau_{rb}}(u_r, u_b))$, which for the present example is not the case since $\tau_{rb} < 0$, implying the Gumbel-hougaard a parameter $\theta_{rb} < 1$, which is not allowed. It can also be seen when estimating $\tau_{rb(s)}$, resulting in the value -0.1245; from Table 5 it then follows that for the range corresponding to this value, the Gumbel-Hougaard family is not in the list of comparable copulas.

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 6: Estimated values of Kendall's τ for the data of the example

This shows that the a priori creation of a copula test space based on available information about the dependence structure is far more preferable, since then the fitting can be carried out with copulas which are more adequate to describe the specific data.

One possibility to work with more than two dimensions, is to make use of the technique of hierarchical Archimedean copulas. Table 5 then allows to create feasible high dimensional copulas. Hierarchical copulas implicitly assume the existence of partial symmetry, which can be seen as a restriction, since the model then only contains $d-1$ parameters. According to the estimates of Kendall's tau for the data described above, it follows from the previous section that on the first level ($\tau = 0.4351$, the dependence is positive), the efficient comparable test space with strict copulas consists of the copulas #1, #4, #5, #6, #12, #13, #14, #17, #19 and #20. Note that on this level, the Gumbel-Hougaard family (#4) is included. On the second level ($\tau = -0.1245$, the dependence is negative), the efficient comparable test space with strict copulas to choose from consists of the copulas #3, #5, #9, #10, #13, #16 and #17 and here the Gumbel-Hougaard family does no longer appear. Both test spaces can then be combined yielding 70 hierarchical constructions.

When assessing the fit of these constructions, several test statistics can be used. As in [21], we work with the Kolmogorov-Smirnoff distance and the Anderson-Darling distance, defined as follows.

$$D_{KS} = \max_t |C_{emp}(F(u_t)) - C_{hyp}(F(u_t))| \quad (1)$$

$$\overline{D_{KS}} = \int_x |C_{emp}(F(u_t)) - C_{hyp}(F(u_t))| \, dC_{hyp}(u) \quad (2)$$

$$D_{AD} = \max_t \frac{|C_{emp}(F(u_t)) - C_{hyp}(F(u_t))|}{\sqrt{C_{hyp}(u_t)(1 - C_{hyp}(u_t))}} \quad (3)$$

$$\overline{D_{AD}} = \int_x \frac{|C_{emp}(F(u_t)) - C_{hyp}(F(u_t))|}{\sqrt{C_{hyp}(u_t)(1 - C_{hyp}(u_t))}} dC_{hyp}(u) \quad (4)$$

where C_{emp} is the so-called empirical copula and C_{hyp} is the parametric copula family used to model the data.

For each distance metric, the optimal combination of copulas leading to a minimal distance is different, see Table 7. For each of these combinations, the corresponding p-values are mentioned between brackets, based on 10000 simulations. The last line of the table contains the results for the distance metrics in case the data are modeled by means of the survival version of the Gumbel-Hougaard copula.

family	D_{KS}	$\overline{D_{KS}}$	D_{AD}	$\overline{D_{AD}}$
{#4, #5}	0.0239 [0.013]			
{#4, #3}		0.0055[0.008]		
{#13, #3}			0.1061 [0.627]	
{#4, #3}				0.0225 [0.006]
{#4, #4}	0.0467 [$< 5 \cdot 10^{-5}$]	0.0151 [$< 5 \cdot 10^{-5}$]	0.1302 [0.605]	0.0467 [$< 5 \cdot 10^{-5}$]

Table 7: Test results for the distance metrics for the optimal combinations of copulas for the data of the example

The results show that working with a comparable copula test space outperforms the Gumbel-Hougaard model for all distance measures. However, it should be remarked that the obtained p-values are in 3 of the 4 cases below the .05 critical value level. This is due to the partial symmetry restriction which is implicitely imposed by the method of hierarchical Archimedean copulas. There exist other methods, like the technique of pairwise copula construction (see [2]) which allow to model a multivariate dependence structure more freely with respect to the number of parameters. We are exploring the possibilities of this more powerful technique in combination with the results of table 5 at the moment, and we hope to present the results in the near future.

5. CONCLUSION

The use of copulas as models for dependence in multivariate data is a very popular technique for many researchers nowadays. In the present contribution, it is shown that the goodness-of-fit of this technique can be positively influenced by constructing an adequate copula test space before carrying out the fitting question. An overview of such adequate test spaces is presented for all possible dependence ranges in terms of Kendall's tau.

ACKNOWLEDGEMENT

This work was partially supported by a BOF-NOI grant from the University of Antwerp.

REFERENCES

[1] K. Aas: Modeling the dependence structure of financial assets: A survey of four copulas. Working paper Norwegian computing centre, SAMBA/22/044 (2004).

[2] K. Aas, C. Czado, A. Frigessi and H. Bakken: Pair-copula constructions of multiple dependence. *in print*.

[3] F. Abid and N. Naifar: The impact of stock returns volatility on credit default swap rates: a copula study (2005).

[4] P. K. Andersen, C. T. Ekstrom, J. P. Klein, Y. Shu and M. Zhang: A class of goodness of fit tests for a copula base on bivariate right-censored data. Biometrical Journal 47(6)(2005), 815–826.

[5] T. Ané and C. Kharoubi: Dependence structure and risk measure. Journal of Business 76(3) (2003).
[6] T. Bacigal and M. Komornicova: Fitting Archimedean copulas to bivariate geodetic data (2006).
[7] E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet and T. Roncalli: Copulas for finance: A reading guide and some applications. Working paper Groupe de recherche Opérationelle Crdit Lyonnais (2000).
[8] U. Cherubini, E. Luciano and W. Vecchiato: Copula methods in Finance. John Wiley and Sons Ltd, New York 2004.
[9] R. T. Clemen and T. Reilly: Correlations and copulas for decision and risk analysis. Management Science 45 (1999), 208–224.
[10] R. De Matteis: Fitting copulas to data. diploma thesis (2001).
[11] M. Denuit and O. Scaillet: Non-parametric tests for positive quadrant dependence. Journal of Financial Econometrics 2(3) (2004), 422–450.
[12] J. Dobric and F. Schmid: Testing goodness of fit for parametric families of copulas-application to financial data. Communications in Statistics: Simulation and Computation 34(4) (2005), 1053–1068.
[13] V. Durrleman, A. Nikeghbali and T. Roncalli: Which copula is the right one? Working paper Groupe de recherche Opérationelle Crdit Lyonnais (2000).
[14] H. Fang, K. Fang and S. Kotz: The meta-elliptical distributions with given marginals. Journal of Multivariate Analysis 82 (2002), 1–16.
[15] J.D. Fermanian: Goodness of fit tests for copulas. Journal of Multivariate Analysis 95 (1) (2005).
[16] E. W. Freez and E. A. Valdez: Understanding relationships using copulas. North American Actuarial Journal 2 (1) (1998).
[17] C. Genest and A. Favre: Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering 12(4) (2007), 347–368.
[18] C. Genest and L. Rivest: Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88 (1993), 1034–1043.
[19] C. Genest and R.J. MacKay: Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. Canadian Journal of Statistics 14(2) (1986), 145–159.
[20] H. Joe: Multivariate Models and Dependence concepts. Chapman and Hall, London 1997.
[21] E. Kole, K. C. G. Koedijk and M. Verbeek: Selecting copulas for risk management. Journal of Banking and Finance 31 (2007), 2405 – 2423.
[22] F. Lindskog, A.J. McNeil and U. Schmock: Kendall's tay for elliptical distributions. Technical Report, Risklab, ETH – Zürich (2001).
[23] R. Nelsen: An introduction to Copulas . Second edition. Springer-Verlag, New York 2006.
[24] C. Savu and M. Trede: Hierarchical Archimedean copulas. In: International Conference on High Frequency Finance, Konstanz, Germany (2006).
[25] A. Sklar: Fonctions de repartition *à* n dimensions et leur marges. publications de l'Institut de Statistique de l'Université de Paris 8 (1959), 229–231.
[26] M. D. Smith: Modelling sample selection using Archimedean copulas. The Econometrics Journal 6(1) (2003), 99–123.

Frederik Michiels, University of Antwerp, Department of Mathematics, Statistics and Actuarial Sciences, Prinsstraat 13, Antwerp, Belgium.
e-mail: `frederik.michiels@ua.ac.be`