Frederik Michiels; Ann De Schepper

A copula test space model how to avoid the wrong copula choice

Kybernetika, Vol. 44 (2008), No. 6, 864--878

Persistent URL: `http://dml.cz/dmlcz/135896`

KYBERNETIKA — VOLUME 44 (2008), NUMBER 6, PAGES 864 – 878

A COPULA TEST SPACE MODEL HOW TO AVOID THE WRONG COPULA CHOICE

Frederik Michiels and Ann De Schepper

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, and we present a complete overview of bivariate test spaces for all possible situations. The practical use will be illustrated by means of a numerical application based on an illustrative portfolio containing the S&P 500 Composite Index, the JP Morgan Government Bond Index and the NAREIT All index.

Keywords: copula, Kendall's tau, goodness-of-fit, copula test space, associated copulas

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 successfully applied in different fields of research such as biostatistics, e. g. [5], geostatistics, e. g. [7], finance, e. g. [9], decision theory, e. g. [11], insurance, e. g. [19] and hydrology, e. g. [20].

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 separately 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 compare the goodness-of-fit of the elements 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*. By this we mean that the test space consists of comparable copulas families with respect to their dependence range. When it comes to modeling data in a particular fitting application, the final copula choice has to be based on more information than only the degree of dependence. E. g. any knowledge related to tail dependence, symmetry, extreme value behaviour or to the size of the zero set $Z(C) = \{(u,v) \in \boldsymbol{I} | C_\theta(u,v) = 0\}$ will be helpful to enlarge the possibility of a good fit. In this paper we restrict ourselves to the dependence information as we want to create broad comparable subsets of copula families from which we can choose in all possible situations. In a forthcoming paper we come back to these issues in more detail as we report our investigation of the particular characteristics of the different test spaces.

Although we explain our methodology for bivariate copulas, it can be extended to fitting applications involving multivariate copula extensions like nested Archimedean copula constructions or pair wise copula constructions. The subject of multivariate extension, however, falls beyond the scope of this article. It will be treated in detail in a forthcoming paper.

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. We use the symbol $\boldsymbol{I}$ to denote the unit interval [0, 1].

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

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

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 [35]:

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 and G are continuous, then C is unique. If not, then C is unique on $\mathrm{im}F \times \mathrm{im}G$. Conversely, if C is a copula and F and G are distribution functions, then a joint distribution function H can be 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 joint distribution function, the copula.

An important issue when using copulas in a fitting application, is that a copula family is characterized by usually one or more parameters, allowing it to model a certain dependence range. In this respect, the following definition is noteworthy:

Definition 2. A family of copulas 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 $(u,v) \in \boldsymbol{I}^2$.

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 Archimedean copula, the copula generator φ and its pseudo-inverse are 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.

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

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

A bivariate Archimedean copula with generator φ is the function $C : \boldsymbol{I}^2 \to \boldsymbol{I}$ defined by

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

Archimedean copulas have some other interesting properties, which are useful in the context of the construction of test spaces:

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 $\boldsymbol{I}$.

2. C is associative; i. e. $C(C(u,v),w) = C(u,C(v,w))$ for all u,v in $\boldsymbol{I}$.
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, in the sense that it is easy to create a generator as in Definition 3, resulting in a new Archimedean copula. See e. g. [33] for a list of 22 bivariate one-parameter Archimedean families. However, most of these Archimedean copula families are not capable to describe the whole dependence range. We will come back to this fact in Section 3, when constructing our test space model.

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.

More detailed information about this class can be found e. g. in [17].

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 4. A copula test space $\widetilde{\boldsymbol{C}}$ is a finite subset of the set $\boldsymbol{C}$ consisting of members of all possible copula families of which will be chosen from in a particular fitting application.

Any fitting application typically consists of two fundamentally different inference problems (see [23]). Firstly, for any parametric copula family in the test space, the value of the parameter (vector) needs to be estimated. Secondly, one has to test the goodness-of-fit of the different possible copulas. The fact that we need to tackle both problems may restrict the final test space $\widetilde{\boldsymbol{C}}$ in its composition.

Indeed, the estimation of the parameter(s) can be done e. g. by using asymptotic maximum likelihood estimation, where the practitioner can choose among other things between exact maximum likelihood estimation, the IFM method (see [26]) or the canonical likelihood estimation. These three methods assume that the density of the copula exists and as such only absolutely continuous copula families can enter the test space. Another parameter estimation technique is the one based on bivariate sample concordance measures such as Kendall's τ (see [21]) and Spearman's ρ (see [33]). When this parameter estimation technique is used, also copulas with singular components can be taken into account for the creation of the test space. In a similar way, the type of goodness-of-fit statistic that is used to choose between the copulas in the test space may restrict the test space. As an example, we mention the goodness-of-fit statistic based on Rosenblatt's transform (see e. g. [15]); this test again requires the existence of the copula density, and hence, only absolutely continuous copula families can be used. When on the other hand the goodness-of-fit test is based on the cumulative distribution function, like tests based on the empirical copula or Kendall's transform (see e. g. [21]), also copula families with singular components are allowed.

Any attempt to construct a test space can be regarded as the test space problem. Apart from the methodological issues with respect to the inference problems stated above, the following three aspects should be taken into account when creating a test space:

- size (*the number of copula families entering the test space*),
- diversity (*the various copula properties present in the test space*),
- relevance (*only comparable copulas entering 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, and as far as we know, there are only a few examples dealing with copula selection from a large test space. See for example [12] for an application with Archimedean copulas and [20] for an application with various copula classes.

Especially in the financial literature small test spaces are used, usually containing three to five 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[1] and the Gaussian and Student's t family are frequently used in examples or in fitting applications (see e. g. [1, 3, 6, 7, 10, 14] or [28]). The motivation for the particular selection of copula families, however, 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

[1]The names are used as in [33].

will also 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 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. [13], 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 [36], 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.

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 $\widetilde{\boldsymbol{C}}$. Clearly $\boldsymbol{C}$ is unknown and we have to rely on existing classes and families in the literature (especially those described in [33] and [26]). In order to keep our analysis synoptical, at this moment we limit ourselves to one parameter copula families. This is not too restrictive, as there exists a rather wide variety in these 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–Mikhail–Haq families);
- Extreme value families: Galambos and Hüsler and Reiss family (also Gumbel–Hougaard family);
- Meta-elliptical class: Normal, Student's t family (with as special case the Cauchy family if $\mathrm{df} = 1$);
- Other families: Farlie–Gumbel–Morgenstern family, Plackett family and Raftery family.

Based on the 29 families described above we can extend our copula space by adding for each bivariate family the three *associated* copula families, defined as follows (see [26]):

Definition 5. For any bivariate copula C, we can define the following associated copulas:

- $C'(u,v) = u - C(u, 1-v)$,
- $C''(u,v) = v - C(1-u, v)$,
- $\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v)$.

The first two transformations allow to create positive quadrant dependent (PQD) families from negative quadrant dependent (NQD) families and vice versa. The third associated copula is called the 'survival' copula as it is the function which is related to the joint distribution of $(1-U, 1-V)$, where (U,V) is a pair of rv's with copula C. By including these associated copulas into the test space, a more diverse and richer test space is obtained.

Noteworthy is the property of some copula families to stay invariant under the aformentionned transformations. Examples from the 29 families introduced above are the Frank family, the Farlie–Gumbel–Morgenstern family, the Plackett family and the families from the meta-elliptical class. We refer to [31, 33], where the invariance with respect to these specific transformations is discussed in the light of symmetry and to [5, 27], where the invariance is investigated in the light of joint associativity. As a final remark we mention the fact that the first two associated copulas are often asymmetric, i. e. $C(u,v) \neq C(v,u)$. As a consequence, the transformed families based on Archimedean families do not always hold the Archimedean property (except e. g. the Frank family and the independence copula Π).

There exist of course other methods to obtain larger families of copulas starting from the ones described above. Indeed, we can take e. g. a convex linear combination of a positive quadrant dependent family and a negative dependent family. However, the resulting family is no longer one-parametric and thus beyond the scope of this article, and it is therefore left out in our discussions.

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 copulas can not be compared. We suggest to do this by means of a powerfull and well-known measure of concordance, namely Kendall's τ.

In general, the relationship between τ and a copula C can be written as:

$$\tau = 4\int_0^1\int_0^1 C(u,v)\,\mathrm{d}C(u,v) - 1 = 1 - 4\int_0^1\int_0^1 \frac{\partial C}{\partial u}(u,v)\frac{\partial C}{\partial v}(u,v)\,\mathrm{d}u\mathrm{d}v. \qquad (1)$$

For Archimedean copulas, this can be rewritten in terms of the copula generator $\varphi(t)$ and its first derivative $\varphi'(t)$, or (see [22])

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

for Meta-elliptical families, expression (1) can be reduced to (see [17, 29])

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

with ρ the correlation coefficient. Associated copula families have tau-values with the same magnitude as the original family. This is summarized in the following lemma.

Lemma 1. For any bivariate copula (family) C, with the associated families defined as in Definition 5, we have $\tau_{C'} = \tau_{C''} = -\tau_C$ and $\tau_{\hat{C}} = \tau_C$.

Proof. From (1) we know that $\tau_{C'} = 1 - 4\int_0^1\int_0^1 \frac{\partial C'}{\partial u}(u,v)\frac{\partial C'}{\partial v}(u,v)\,\mathrm{d}u\mathrm{d}v$. Together with Definition 5, it easily follows that

$$\begin{aligned}\tau_{C'} &= 1 - 4\int_0^1\int_0^1 \left(1 - \frac{\partial C}{\partial u}(u,1-v)\right)\frac{\partial C}{\partial v}(u,1-v)\,\mathrm{d}u\mathrm{d}v \\ &= 1 - 4\int_0^1\int_0^1 \frac{\partial C}{\partial v}(u,1-v)\,\mathrm{d}u\mathrm{d}v + 4\int_0^1\int_0^1 \frac{\partial C}{\partial u}(u,1-v)\frac{\partial C}{\partial v}(u,1-v)\,\mathrm{d}u\mathrm{d}v.\end{aligned}$$

The first integral is equal to 1/2, and with a substitution $v^* = 1 - v$ in the second integral, it follows that $\tau_{C'} = -\tau_C$. The same applies for C''. The proof for $\hat{C}$ is completely analogous, resulting now in a τ-value with the same size and the same sign. □

As a consequence, the dependence intervals of the first two associated families are the reversed version of the dependence intervals of the original families. For the survival versions of the copula families, the τ ranges remain identical.

Tables 1 to 4 show the results for the range for Kendall's τ that were obtained for the 29 one parameter families mentioned earlier. The parameter values for the 29 families 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.

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 $C_{\#2}$, $C_{\#8}$, $C_{\#15}$, $C_{\#16}$, $C_{\#21}$ and $C_{\#26}$ for df ≤ 2). Secondly, one should take into account the fact that not all copulas are positively ordered. Some of them, like copula $C_{\#9}$, $C_{\#10}$ and $C_{\#11}$, 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 τ 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 τ values. The results are summarized in Table 5. The table shows a restructuring of the copula space into 13 dependence intervals, each entry containing 10 to 25 different copulas, with the feasible copula families marked in grey. If e. g. the data reveal an estimate for Kendall's τ of about −0.5, we have to look at the third test space, which we will denote as $TS_{[-0.5649\,;\,-0.4674)}$.

Table 1. Parameter ranges in terms of Kendall's τ for Archimedean families.

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

Table 2. Parameter ranges in terms of Kendall's τ for extreme value 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 (Hüsler and Reiss)	$\exp\left(\log(u)\Phi\left(\frac{1}{\theta}+\frac{1}{2}\theta\log\frac{\log u}{\log v}\right)+\log(v)\Phi\left(\frac{1}{\theta}+\frac{1}{2}\theta\log\frac{\log v}{\log u}\right)\right)$	$[0,\infty)$	$[0,1]^*$

Table 3. Parameter ranges in terms of Kendall's τ for meta-elliptical families.

#	$C_\theta(u,v)$	$\rho \in$	$\tau \in$
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 4. Parameter ranges in terms of Kendall's τ for other 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 (Plackett)	$\frac{(1+(\theta-1)(u+v))-\sqrt{(1+(\theta-1)(u+v))^2-4uv\theta(\theta-1)}}{2(\theta-1)}$	$(0,\infty)$	$[-1,1]$
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 5. Restructuring the copula test space.

τ - range	Archimedean copulas																						extreme value copulas		meta-elliptical copulas		other copulas		
	#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
[-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 , 0.2222]	■	■	■	■	■	■		■					■		■	■	■			■	■		■	■	■	■	■	■	■
(0.2222 , 0.3333)	■	■	■	■	■	■		■					■		■	■	■			■	■		■	■	■	■		■	■
0.3333	■	■	■	■	■	■		■				■	■	■	■	■	■	■	■	■	■		■	■	■	■		■	■
(0.3333 , 1]	■	■		■	■	■						■	■	■	■		■	■	■	■	■		■	■	■	■		■	■

This comparable copula test space consists of Archimedean copulas $C_{\#1}$, $C_{\#2}$, $C_{\#5}$, $C_{\#7}$, $C_{\#8}$, $C_{\#11}$, $C_{\#15}$, $C_{\#16}$, $C_{\#17}$ and $C_{\#21}$, 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 value for τ.

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.

4. EXAMPLE — APPLICATION

In this section, we show how the results of Table 5 can be used in a concrete bivariate fitting application. We have chosen to work with the same data as used in [28], or 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[2]. In particular, our aim is to model all the bivariate margins, i. e. the dependence structures of three pairs of financial data.

As explained earlier, before the actual construction of the test spaces, it is necessary to make decisions with respect to the inference methods for the parameter estimation and for the goodness-of-fit tests. For the present example, we have chosen to use first the sample Kendall's τ for the estimation of the copula parameters, and afterwards the Cramér–von Mises goodness-of-fit statistic based on the empirical copula. For an overview of goodness-of-fit tests, see [18] or [23]. This allows us to use all the copula families of Table 5, without restriction.

We first obtain the empirical concordance matrix which is displayed in Table 6.

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

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

As can be seen in this table, two out of the three τ values indicate a clear negative dependence, and thus 'blindly' creating a test space containing extreme value families like the Gumbel–Hougaard family would be inefficient. When using the empirical τ to construct three different, comparable and efficient test spaces, with the associated copula families this leads to six comparable test spaces, as $\left[C, \hat{C}\right]$ and $\left[C', C''\right]$ respectively have the same τ range. In particular, from Table 5 it follows that for

[2] In the paper of [28], the period covered in the example ended on December 17 2004.

the present application we have to use test spaces $TS_{[-0.1820,0)}$ (for C and $\hat{C}$) and $TS_{[0,0.2222)}$ (for C' and C'') for the stocks-bonds pair, $TS_{[0.3333,1]}$ (for C and $\hat{C}$) and $TS_{(-0.3613,-0.4674]}$ (for C' and C'') for the stocks-real estate pair, and finally $TS_{[-0.1820,0)}$ (for C and $\hat{C}$) and $TS_{[0,0.2222)}$ (for C' and C'') for the bonds-real estate pair. The parameter estimation for the copula members in each test space can be carried out by using their relationship with Kendall's τ.

In order to assess the goodness-of-fit for each copula member we rely on the Cramér–von Mises statistic:

$$S_n = \iint_{\boldsymbol{I}^2} \mathbb{C}_n(\boldsymbol{u})^2 \mathrm{d}C_n(\boldsymbol{u}) \tag{2}$$

where $\mathbb{C}_n = \sqrt{n}(C_n - C_{\theta_n})$. For more details about this technique, we refer to [23].

The next table displays for each data pair the top three best fit copula families from the different test spaces[3]. The values between brackets are the p-values, based on 10000 simulations.

Table 7. Results goodness-of-fit for 3 pairs of financial data

#	C	(stocks, bonds)	C	(bonds, real estate)	C	(stocks, real estate)
1	$C''_{\#4}$	0.0411 [0.159]	$C_{\#26b}$	0.0127 [0.950]	$C_{\#26b}$	0.0152 [0.734]
2	$C'_{\#1}$	0.0510 [0.078]	$C''_{\#4}$	0.0136 [0.918]	$C_{\#25}$	0.0163 [0.665]
3	$C''_{\#6}$	0.0557 [0.064]	$C'_{\#13}$	0.0156 [0.842]	$C_{\#26a}$	0.0275 [0.281]

From Table 7 the following remarks can be made:

- The method of comparable test spaces results in all three cases in statistically significant best fit copulas.
- The method allows for a wide diversity of copulas; more than 50 percent of the top 3 best fit correspond to copulas that probably would not be used without including a priori dependence information as in our method.
- No Archimedean family is found in the top 3. Conversely, associated copulas from Archimedean families do give a good fit when modeling for negative dependence.
- The pair (stocks, bonds) is best modeled with *asymmetric* copulas, where the associated Gumbel family $C''_{\#4}$ provides the best fit.
- The pair (bonds, real estate) is best modeled with the Student's t family with 10 degrees of freedom, although the asymmetric associated families $C''_{\#4}$ and $C'_{\#13}$ also provide a good fit. This can be due to the low degree of dependence ($\tau = -0.0599$).
- The strong positive dependence structure of the pair (stocks, real estate) is clearly best modeled by means of families from the meta-elliptical class.

[3] The notation $C''_{\#4}$ e.g. means that we take the second associated copula for copula 4 or for the Gumbel–Hougaard copula. The parameter values are estimated by means of the relation between Kendall's τ and the (Archimedean) copula as given in (1). For the Student's t family, 26a corresponds to a situation with 3 degrees of freedom, and 26b to 10 degrees of freedom.

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 how the goodness-of-fit of this technique can be positively influenced by constructing an adequate copula test space before carrying out the fitting question, and it is illustrated through a numerical application. We present an overview of such adequate test spaces in the bivariate case, for all possible dependence ranges in terms of Kendall's tau. The concept of comparable test spaces can be extended to a multivariate situation. This investigation is still ongoing, we hope to present the results in a later paper.

ACKNOWLEDGEMENT

The authors would like to thank two referees of this journal for their appropriate and relevant comments. The paper gained a lot due to their helpful and constructive advice.

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

(Received May 15, 2008.)

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. Insurance: Mathematics and Economics. In press, 2007, doi:10.1016/j.insmatheco.2007.02.001.

[3] F. Abid and N. Naifar: The impact of stock returns volatility on credit default swap rates: A copula study (2005), available at SSRN: http://ssrn.com/abstract=726726.

[4] C. Alsina, M. J. Frank, and B. Schweizer: Associative Functions. World Scientific, Singapore 2006.

[5] 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 J. *47* (2005), 6, 815–826.

[6] T. Ané and C. Kharoubi: Dependence structure and risk measure. J. Business *76* (2003), 3, 441–438.

[7] T. Bacigal and M. Komorníková: Fitting Archimedean copulas to bivariate geodetic data. In: Proc. COMPSTAT, 2006, pp. 649–656.

[8] D. Berg and K. Aas: Models for construction of multivariate dependence: A comparison study (2007).
Unpublished paper, available at http://www.danielberg.no/publications/highdim.pdf.

[9] 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 Crédit Lyonnais 2000.

[10] U. Cherubini, E. Luciano, and W. Vecchiato: Copula Methods in Finance. Wiley, New York 2004.

[11] R. T. Clemen and T. Reilly: Correlations and copulas for decision and risk analysis. Management Sci. *45* (1999), 208–224.

[12] R. De Matteis: Fitting Copulas to Data. Diploma Thesis, 2001.

[13] M. Denuit and O. Scaillet: Non-parametric tests for positive quadrant dependence. J. Financial Econometrics *2* (2004), 3, 422–450.

[14] J. Dobric and F. Schmid: Testing goodness of fit for parametric families of copulas-application to financial data. Commun. Statist.: Simulation and Computation *34* (2005), 4, 1053–1068.

[15] J. Dobric and F. Schmid: A goodness of fit test for copulas based on Rosenblatt's transform. Comput. Statis. Data Anal. *51* (2007), 4633–4642.

[16] V. Durrleman, A. Nikeghbali, and T. Roncalli: Which copula is the right one? Working paper Groupe de recherche Opérationelle Crédit Lyonnais 2000.

[17] H. Fang, K. Fang, and S. Kotz: The meta-elliptical distributions with given marginals. J. Multivariate Anal. *82* (2002), 1–16.

[18] J. D. Fermanian: Goodness of fit tests for copulas. J. Multivariate Anal. *95* (2005), 1, 119–152.

[19] E. W. Frees and E. A. Valdez: Understanding relationships using copulas. North American Actuarial J. *2* (1998), 1, 1–25.

[20] C. Genest and A. Favre: Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrologic Engrg. *12* (2007), 4, 347–368.

[21] C. Genest and L. Rivest: Statistical inference procedures for bivariate Archimedean copulas. J. Amer. Statist. Assoc. *88* (1993), 1034–1043.

[22] C. Genest and R. J. MacKay: Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. Canad. J. Statist. *14* (1986), 2, 145–159.

[23] C. Genest, B. Rémillard, and D. Beaudoin: Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics 2007, doi:10.1016/j.insmatheco.2007.10.005

[24] M. Hofert: Sampling Archimedean copulas. Comput. Statist. Data Anal. *52* (2008), 12, 5163–5174.

[25] H. Joe: Multivariate concordance. J. Multivariate Anal. *35* (1990), 12–30.

[26] H. Joe: Multivariate Models and Dependence Concepts. Chapman and Hall, London 1997.

[27] E. P. Klement, R. Mesiar, and E. Pap: Invariant copulas. Kybernetika *38* (2002), 275–285.

[28] E. Kole, K. C. G. Koedijk, and M. Verbeek: Selecting copulas for risk management. J. Banking and Finance *31* (2007), 2405–2423.

[29] F. Lindskog, A. J. McNeil, and U. Schmock: Kendall's tau for Elliptical Distributions. Technical Report, Risklab, ETH Zürich 2001.

[30] A. J. McNeil: Sampling Archimedean copulas. J. Statist. Comput. Simul. *78* (2008), 6, 567–581.

[31] R. B. Nelsen: Some concepts of bivariate symmetry. J. Nonparametric Statist. *3* (1993), 95–101.

[32] R. B. Nelsen: Nonparametric measures of multivariate Aassociation. In: Distributions with Fixed Marginals and Related Topics 1996, pp. 223–232.

[33] R. B. Nelsen: An Introduction to Copulas. Second edition. Springer-Verlag, New York 2006.

[34] C. Savu and M. Trede: Goodness-of-fit tests for parametric families of Archimedean copulas. Quantitative Finance *8* (2008), 2, 109–116.

[35] A. Sklar: Fonctions de repartition *à* n dimensions et leur marges. Publ. Inst. Statist. Univ. Paris *8* (1959), 229–231.

[36] M. D. Smith: Modelling sample selection using Archimedean copulas. The Econometrics Journal *6* (2003), 1, 99–123.

Frederik Michiels and Ann De Schepper, University of Antwerp, Faculty of Applied Economics, Department of Mathematics, Statistics and Actuarial Sciences, Prinsstraat 13, Antwerp. Belgium.

e-mail: frederik.michiels@ua.ac.be