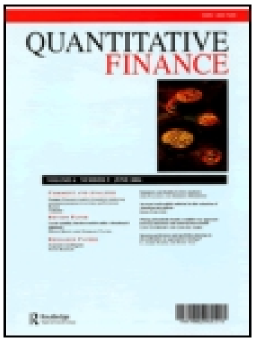

Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Goodness-of-fit tests for parametric families of Archimedean copulas

Cornelia Savu & Mark Trede

To cite this article: Cornelia Savu & Mark Trede (2008) Goodness-of-fit tests for parametric families of Archimedean copulas, Quantitative Finance, 8:2, 109-116, DOI: 10.1080/14697680701207639

To link to this article: https://doi.org/10.1080/14697680701207639

Quantitative Finance, Vol. 8, No. 2, March 2008, 109–116

Routledge
Taylor & Francis Group

Goodness-of-fit tests for parametric families of Archimedean copulas

CORNELIA SAVU and MARK TREDE*

Institute for Econometrics, University of Münster, Am Stadtgraben 9, 48143 Münster, Germany

(*Received 13 July 2005; in final form 9 January 2007*)

1. Introduction

Copulas are a powerful tool in multivariate statistics. They capture the pure dependence structure between random variables irrespective of their marginal distributions. Since modeling the marginal distributions can be severed from modeling dependence, copulas greatly facilitate the development of new multivariate distributions. Practitioners are no longer obliged to use the multivariate Gaussian distribution (and only a few others) for lack of alternatives.

If a copula is used to model the dependence between random variables there is an immediate and obvious need to test whether the model can actually describe the data at hand accurately enough. This paper suggests a goodness-of-fit test for parametric families of Archimedean copulas. The test procedure has three main advantages: (a) it is straightforward, being based on the classical χ^2 statistic, although its asymptotic distribution is not χ^2; (b) one can either fully specify an Archimedean copula or just a parametric family of Archimedean copulas (e.g., the Gumbel–Hougaard family); and, most importantly, (c) it can handle high-dimensional distributions by exploiting the properties of Archimedean copulas. The latter aspect is of particular importance in financial econometrics where high-dimensional multivariate distributions are essential for modeling credit portfolios. In fact, it turns out that the test's performance is better the higher the dimension of the multivariate distribution. Simply testing the goodness-of-fit of the (bivariate) pairs might therefore lead to a false acceptance of misspecified models.

An important question in finance is whether or not there is a family of Archimedean copulas capable of modeling stock return dependence. We address this question in an empirical illustration using monthly returns of six stocks contained in the USA S&P500 Chemicals index.

Although well known in the statistical literature for more than 40 years, applications of the copula theory in statistical modeling are a more recent phenomenon. Today, there is a fast growing literature on copulas. A very good introduction to the copula concept is Nelsen (1999) and, with some extensions, Joe (1997). Examples for studies of copulas in the context of financial problems include Embrechts *et al*. (1999, 2003) who applied copulas to risk management. Another example of an application in finance is Bouyé *et al*. (2001), who review different financial problems and show how these could be solved by copulas. Cherubini *et al*. (2004) present a collection of existing strategies for the use of the copula methodology in empirical finance. Frees and Valdez (1998) and Klugman and Parsa (1999) discuss the relevance of copulas in actuarial applications and perform bivariate goodness-of-fit tests in an insurance context.

In practice, the class of Archimedean copulas is often used, because these copulas are particularly easy to handle and have simple, closed form expressions. Archimedean copulas are discussed in Genest and MacKay (1986a, b), Nelsen (1999) and Joe (1997). Many other authors work with Archimedean copulas, but most of them investigate only the bivariate case, such as, for example, Melchiori (2003).

Another important practical question is the issue of fitting copulas to data. There are a number of articles concerning the statistical inference for copulas. Different approaches have been established for the estimation of copulas: Genest *et al*. (1995) and Joe and Xu (1996) estimate a parametric family by maximum likelihood, whereas Deheuvels (1979) proposed a non-parametric estimator using the empirical copula. For the class of Archimedean copulas, Genest and Rivest (1993)

*Corresponding author. Email: mark.trede@uni-muenster.de

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2008 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/14697680701207639

developed a semiparametric estimator. Empirical copula processes are an important tool for deriving asymptotic distributions of test statistics; they are investigated by Fermanian *et al.* (2004).

An overview of existing methods for calibrating and simulating copula functions with an application to the Italian stock market in the bivariate case is given by Romano (2002). Durrleman *et al.* (2000) asked "Which copula is the right one?" and review this problem in the financial modeling context.

Various authors have dealt with the question of goodness-of-fit tests for copulas, but to date there is no unanimously recommended method for checking whether a parametric family of copulas can appropriately describe the dependence structure for the data at hand. Several goodness-of-fit tests have been proposed in the literature. Some use tests based on the probability integral transform introduced by Rosenblatt (1952) such as Chen *et al.* (2004) and Berg and Bakken (2005). Others assume goodness-of-fit tests based on the empirical copula process such as, for example, Fermanian (2005) and Fermanian and Scaillet (2005). Genest *et al.* (2006) and Wang and Wells (2000) propose tests based on Kendall's process of Barbe *et al.* (1996), which implies the use of non-parametric distribution free statistics like those of Kolmogorov–Smirnov, Anderson–Darling or Cramér–von-Mises. Others introduce multivariate tests based on the χ^2 statistic, such as Dobrić and Schmid (2005). Scaillet (2005) suggests a non-parametric goodness-of-fit test based on a kernel smoothed estimator for the copula density. A slightly different approach is proposed by Malevergne and Sornette (2003), who test the Gaussian copula hypothesis for the bivariate case.

The existing empirical research with regard to copulas almost exclusively investigates the bivariate case, but, for financial applications, a multivariate framework is required. Our paper provides an extension to higher dimensions, suggesting a goodness-of-fit test for multivariate Archimedean copulas.

The paper is structured as follows. Section 2 briefly reviews the family of Archimedean copulas and discusses their properties. In section 3 we present a goodness-of-fit test for parametric families of Archimedean copulas. Section 4 is devoted to simulations, both of the test's size and of its power against various alternatives. We illustrate the test procedure with an application to joint return distributions in section 5. Finally, section 6 concludes.

2. Parametric families of Archimedean copulas

This section presents some basic definitions and properties for the class of Archimedean copulas that we use in this paper.

The idea behind the copula concept is very simple and implies separate modeling of the marginal behaviour and dependence structure for the underlying random variates. A d-dimensional copula is a multivariate distribution function on $[0, 1]^d$ with standard uniform marginal distributions. Let $X_1, \ldots, X_d$ be random variables with continuous distribution functions $F_i(x_i) = P(X_i \le x_i)$, $i = 1, \ldots, d$, and joint distribution function

$$F(x_1, \ldots, x_d) = P(X_1 \le x_1, \ldots, X_d \le x_d).$$

Then the joint distribution function of the random vector $(X_1, \ldots, X_d)$ can be represented as

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \tag{1}$$

where the copula C describes the dependence structure between the random variables $X_1, \ldots, X_d$. This result is known in the literature as Sklar's theorem and implies that, for continuous multivariate distributions, the univariate margins can be separated from the dependence structure, represented by the copula. A d-dimensional copula of a d-dimensional random vector $(X_1, \ldots, X_d)$ is the joint distribution function C of the uniform random vector $(F_1(X_1), \ldots, F_d(X_d))$.

One important class of copulas is known as the Archimedean copulas. There are a number of reasons why these find a wide range of application in practice. They are very easy to construct and many parametric families belong to this class. Archimedean copulas allow for a great variety of different dependence structures and all commonly encountered Archimedean copulas have simple closed form expressions. In addition, the Archimedean representation allows us to reduce the study of a multivariate copula to a single univariate function.

2.1. *Definition*

Let $\phi : [0, 1] \to [0, \infty]$ be a continuous, strictly decreasing and convex function such that $\phi(1) = 0$ and $\phi(0) = \infty$. The function ϕ has an inverse $\phi^{-1} : [0, \infty] \to [0, 1]$ with the same properties as ϕ, except that $\phi^{-1}(0) = 1$ and $\phi^{-1}(\infty) = 0$.

The function $C^d : [0, 1]^d \to [0, 1]$, defined by

$$C(u_1, \ldots, u_d) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_d)), \tag{2}$$

is called an Archimedean copula if and only if ϕ^{-1} is completely monotonic on $[0, \infty)$, that is

$$(-1)^k \frac{\partial^k}{\partial u^k} \phi^{-1}(u) \ge 0, \quad \text{for } k = 1, 2, \ldots.$$

The function ϕ is called the generator of the copula. See Schweizer and Sklar (1983, Theorem 6.3.6) and Nelsen (1999, Theorem 4.6.2).

For Archimedean copulas with generator ϕ the density is given by

$$c(u_1, \ldots, u_d) = \phi^{-1(d)}(\phi(u_1) + \cdots + \phi(u_d)) \prod_{i=1}^{d} \phi'(u_i). \tag{3}$$

2.2. *Properties*

Let C be an Archimedean copula with generator ϕ. Then:

1. C is permutation-symmetric in its d arguments, e.g. in the bivariate case $C(u_1, u_2) = C(u_2, u_1)$ for all $u_1, u_2 \in [0, 1]$. This implies that the copula is the distribution function of d exchangeable uniform random variates;
2. C is associative, e.g. in the trivariate case
$$C(C(u_1, u_2), u_3) = C(u_1, C(u_2, u_3)),$$
for all $u_1, u_2, u_3 \in [0, 1]$. The associativity of Archimedean copulas is, in general, not shared by other copulas;
3. if $a > 0$ is any constant then $a\phi$ is also a generator of C. The generator ϕ uniquely determines an Archimedean copula only up to a scalar multiple.

For d random variables $U_1, \ldots, U_d$ with distribution function C and generator ϕ, we also know the distribution of the random variable $V = C(U_1, \ldots, U_d)$,

$$\begin{aligned} K(t) &= \Pr(C(U_1, \ldots, U_d) \leq t) \\ &= t + \sum_{i=1}^{d-1} (-1)^i \frac{\phi^i(t)}{i!} f_{i-1}(t), \end{aligned} \tag{4}$$

where the auxiliary functions $f_0(t) = 1/\phi'(t)$ and $f_i(t)$ for $i \geq 1$ are defined recursively as $f_i(t) = f'_{i-1}(t)/\phi'(t)$. Appendix A gives these auxiliary functions for a number of Archimedean families.

In the bivariate case, (4) simplifies to $K(t) = t - \phi(t)/\phi'(t)$ (Genest and Rivest 1993). Barbe *et al*. (1996) extended the idea of Genest and Rivest (1993) to higher dimensions and obtained (4) as the distribution function of the copula.

An Archimedean copula is determined by the function $K(t)$ defined on the unit interval [0,1]. This is a very useful result if we want to determine which parametric copula family fits the data best.

2.3. *Examples*

Among the Archimedean copulas, we are going to consider only the one-parameter copulas, which are constructed using a generator $\phi_\theta(t)$, characterized by a scalar parameter θ. Table 1 describes some Archimedean copulas with their generators and the admissible range of parameter θ.

The Cook–Johnson family is the multivariate extension of the well-known Clayton copula family. For a more detailed list of Archimedean copula families and their generators we refer the reader to Nelsen (1999) and Joe (1997).

Table 1. Families of Archimedean copulas.

Family	$\phi_\theta(t)$	Range	$C_\theta(u_1, \ldots, u_d)$
Gumbel–Hougaard	$(-\ln t)^\theta$	$\theta \geq 1$	$\exp\left(-\left[\sum_{i=1}^d (-\ln u_i)^\theta\right]^{1/\theta}\right)$
Cook–Johnson	$\dfrac{t^{-\theta} - 1}{\theta}$	$\theta > 0$	$\left(\sum_{i=1}^d u_i^{-\theta}\right) - n + 1]^{-1/\theta}$
Frank	$-\ln \dfrac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$\mathbb{R}\backslash\{0\}$	$-\dfrac{1}{\theta} \ln\left(1 + \dfrac{\prod_{i=1}^d (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{d-1}}\right)$

3. Goodness-of-fit test

Our goal is to find a goodness-of-fit method in order to test the parametric specification for the copula function only. In doing so we consider composite null hypotheses of the form

$$H_0 : C \in \{C_\theta, \theta \in \Theta\} \quad \text{against} \quad H_0 : C \notin \{C_\theta, \theta \in \Theta\},$$

where $\{C_\theta, \theta \in \Theta\}$ denotes some parametric family of copulas.

Since we deal with composite hypotheses and therefore the right copula is unknown even under the null hypothesis, we need to estimate the exact copula specification. We estimate the parameter θ of the generator function ϕ_θ by maximum likelihood without specifying the marginal distributions (the canonical maximum likelihood method, see Durrleman *et al.* (2000)). The density function belonging to the copula (2) is (3). Let

$$(X_{11}, \ldots, X_{d1}), \ldots, (X_{1n}, \ldots, X_{dn})$$

be a random sample of size n from the random vector $(X_1, \ldots, X_d)$. Define

$$\begin{aligned} U_{ij} &= \frac{1}{n+1} \sum_{k=1}^n 1(X_{ik} \leq X_{ij}) \\ &= \frac{R(X_{ij})}{n+1}, \end{aligned}$$

where $R(X_{ij})$ is the (ascendingly ordered) rank of X_{ij} among $X_{i1}, \ldots, X_{in}$. Hence, the marginal distributions are estimated non-parametrically. The log-likelihood function (also called the pseudo-log-likelihood function) of θ is

$$\ln L(\theta; U_{ij}, i = 1, \ldots, d, j = 1, \ldots, n) = \sum_{j=1}^n \ln c(U_{1j}, \ldots, U_{dj}),$$

and the canonical maximum likelihood estimator of θ is

$$\hat{\theta} = \arg\max \ln L(\theta). \tag{5}$$

Maximizing the likelihood is a standard numerical problem and easily performed. Because of the non-parametric estimation of the margins the asymptotic variance of $\hat{\theta}$ is not the inverse of the information matrix (Genest *et al*. 1995).

We denote the estimated generator function as $\phi_{\hat{\theta}}$ and the estimated version of (4) as $K_{\hat{\theta}}$. Our goodness-of-fit test procedure is based on the classical χ^2 statistic. Partition [0,1] into B intervals with break points

$$0 = a_0 < a_1 < \cdots < a_{B-1} < a_B = 1.$$

Define

$$\hat{V}_j = \phi_{\hat{\theta}}^{-1}(\phi_{\hat{\theta}}(U_{1j}) + \cdots + \phi_{\hat{\theta}}(U_{dj})); \tag{6}$$

the test statistic is

$$T = \sum_{k=1}^{B} \frac{(|\{j : a_{k-1} < \hat{V}_j \le a_k\}| - n(K_{\hat{\theta}}(a_k) - K_{\hat{\theta}}(a_{k-1})))^2}{n(K_{\hat{\theta}}(a_k) - K_{\hat{\theta}}(a_{k-1}))}.$$

In contrast to the usual χ^2 statistic, T does not follow a χ^2 distribution—unless both the marginal distributions and the parameter θ are fully specified in the null hypothesis. It is well known (Chernoff and Lehmann 1954) that if maximum likelihood estimation of the unknown parameters is based on the entire, original sample (rather than on grouped values), then the test statistic follows a weighted mixture of χ_1^2 distributions, where the weights can be consistently estimated from the data. However, the approach of Chernoff and Lehmann (1954) does not apply in our setting since the non-standard distribution of T is due to the fact that the canonical maximum likelihood estimator $\hat{\theta}$ not only enters the expected number of cell counts, $n(K_{\hat{\theta}}(a_k) - K_{\hat{\theta}}(a_{k-1}))$, but also the observed number $|\{j : a_{k-1} < \hat{V}_j \le a_k\}|$ via $\phi_{\hat{\theta}}^{-1}$ and $\phi_{\hat{\theta}}$.

Note that higher-dimensional distributions are projected into the unit interval via the univariate distribution of $\hat{V}_j$. The projection is random as it depends on the estimator $\hat{\theta}$. Exploiting this property of Archimedean copulas, the test problem is becoming easier in higher dimensions, but we have to pay a price: the test procedure only has trivial power against non-Archimedean copulas with the same $K(t)$. We therefore maintain the assumption that the copulas under consideration are Archimedean without a formal test.

In order to approximate the distribution of T under the null hypothesis we apply a parametric bootstrap procedure. Having estimated θ by its canonical maximum likelihood estimator $\hat{\theta}$ we generate a large number J of artificial samples, each of size n, from the Archimedean copula with generator $\phi_{\hat{\theta}}$ (see section 4 for a description of random number generation). For each sample the parameter is estimated again by canonical maximum likelihood, resulting in $\theta_1^*, \ldots, \theta_J^*$. We then compute the test statistic for each sample T_j^* for $j = 1, \ldots, J$. Let $T_{(1)}^*, \ldots, T_{(J)}^*$ denote their order statistics. The bootstrap critical value for our test is $T_{((1-\alpha)J)}^*$ at significance level α. The null hypothesis is rejected if $T > T_{((1-\alpha)J)}^*$.

The test is consistent as long as plim $\Pr(\hat{V}_j \le a_k) \ne \text{plim} K_{\hat{\theta}}(a_k)$ for at least one $k \in \{1, \ldots, B\}$. Although this condition might not be satisfied for certain combinations of the null model, the alternative model, and break points, it will usually be fulfilled in practice. Trivially, the test cannot distinguish different families of Archimedean copulas when the copula under consideration is the independence copula. In general, the test will be less likely to detect misspecified copulas when the degree of dependence in the data is low.

4. Simulations

In order to assess the performance of our test procedure we conduct a number of simulation studies. The most fundamental question concerns the error probability of the first kind: does the test keep its nominal size? We then proceed to investigate the test's power against various alternatives including alternatives violating the maintained assumption that the copula is Archimedean.

Random number generation from Archimedean copulas is relatively easy because of their simple structure. The simulation method we use draws on Theorem 3 of Barbe *et al*. (1996) which we restate here for convenience. Let $T \in [0, 1]^d$ be a d-dimensional random vector uniformly distributed over the simplex $\{z \in [0, 1]^d : \sum_{i=1}^{d} z_i = 1\}$, and let V be a random variable with distribution function (4). If T and V are independent, then (2) is the distribution function of the vector U with components $U_i = \phi^{-1}(T_i\phi(V))$ for $i = 1, \ldots, d$. Random numbers from V are generated by the conventional inversion method; the inverse of (4) can either be derived analytically or numerically.

We first investigate the empirical size of our test procedure by Monte Carlo simulations. For each combination of sample size n and dimension d, 5000 i.i.d. samples from the copula were generated. Table 2 reports the proportion of rejections of the (true) null hypothesis that the data follow the respective

Table 2. Empirical size of the test.

Copula	Dimension d	Sample size $n=100$	$n=200$	$n=500$
Cook–Johnson	$d=2$	0.052	0.046	0.049
	$d=3$	0.048	0.048	0.052
	$d=5$	0.053	0.052	0.052
Frank	$d=2$	0.054	0.047	0.052
	$d=3$	0.052	0.046	0.048
	$d=5$	0.051	0.051	0.048
Gumbel–Hougaard	$d=2$	0.058	0.048	0.056
	$d=3$	0.046	0.049	0.051
	$d=5$	0.048	0.029	0.024

Table 3. Empirical Power of the test.

Data generating copula	Dimension d	Copula family in H_0								
		$n=100$			$n=200$			$n=500$		
		CJ	F	GH	CJ	F	GH	CJ	F	GH
Cook–Johnson	2	–	0.07	0.14	–	0.10	0.25	–	0.23	0.72
Frank		0.14	–	0.10	0.21	–	0.15	0.52	–	0.37
Gumbel–Hougaard		0.46	0.14	–	0.79	0.28	–	1.00	0.72	–
Gaussian		0.18	0.07	0.08	0.27	0.08	0.10	0.67	0.13	0.14
t_2		0.34	0.14	0.06	0.60	0.27	0.07	0.96	0.73	0.09
Cook–Johnson	3	–	0.08	0.27	–	0.16	0.61	–	0.70	1.00
Frank		0.43	–	0.14	0.76	–	0.27	0.99	–	0.79
Gumbel–Hougaard		0.96	0.64	–	1.00	0.87	–	1.00	1.00	–
Gaussian		0.56	0.16	0.10	0.80	0.25	0.14	0.98	0.39	0.36
t_2		0.78	0.50	0.08	0.95	0.75	0.11	1.00	0.98	0.23
Cook–Johnson	5	–	0.09	0.24	–	0.30	0.70	–	0.96	1.00
Frank		0.81	–	0.11	0.99	–	0.24	1.00	–	0.83
Gumbel–Hougaard		1.00	0.87	–	1.00	0.98	–	1.00	1.00	–
Gaussian		0.79	0.23	0.04	0.96	0.34	0.06	1.00	0.59	0.22
t_2		0.87	0.58	0.04	0.97	0.81	0.05	1.00	0.99	0.13

copula family. Critical values were obtained using the parametric bootstrap approach with 5000 replications. In order to make the degree of dependence comparable, the parameter θ was chosen for each distribution such that Kendall's $\tau = 0.3$. The unit interval is divided into $B=20$ intervals $[0, 0.05], \ldots,]0.95, 1]$. The simulation results indicate that the test keeps its size rather accurately, and deviations from the prescribed level, if any, are far more often than not on the conservative side.

Table 3 shows the power of the test. The first entry in each row gives the name of the data generating copula. Again, the parameter θ is chosen such that $\tau = 0.3$. We then tested the null hypothesis that the simulated data follow the copula family given in the column heads at level $\alpha = 0.05$. The sample size is rather small, $n = 100, 200, 500$, the dimensions are $d = 2, 3, 5$, and the number of simulation replications is $R = 1000$. In addition to the Archimedean copulas we also generated data from the equi-correlated multivariate Gaussian copula. The parameter was set such that $\tau = 0.3$ for every pair (U_i, U_j), $i, j = 1, \ldots, d$, $i \neq j$.†

Table 3 reveals that the test's power is, for certain combinations, relatively high even for small samples. The power does not deteriorate with increasing dimension d. On the contrary—and in contrast to other goodness-of-fit test procedures—the higher d the higher, in general, the power. This property of our test is due to the projection step from the d-dimensional unit hypercube into the one-dimensional unit interval. The projections become less and less similar as the dimension increases. Consequently, it becomes easier for the test to detect violations of the null hypothesis. The fact that the power is increasing in the dimension d is a very strong argument in favour of a joint goodness-of-fit test rather than testing each pair of variables separately. The power against the Gaussian distribution (which violates the maintained hypothesis that the copula is Archimedean) does not always follow this pattern; the already low power of the Gumbel–Hougaard copula family decreases when the dimension is increased from $d=3$ to $d=5$.

5. Empirical illustration

We illustrate our test procedure with an application in financial econometrics. Correctly specified multivariate return distributions are of paramount importance for credit portfolio management. However, due to the lack of alternatives, many return models are based on the multivariate Gaussian or the multivariate t-distribution. If the dependence structure of the return vector can sufficiently accurately be described by a parametric Archimedean copula, one could aim at developing credit portfolio models that capture the behaviour of joint returns more realistically. Consequently, estimation of important risk measures (e.g. the value at risk) for credit exposure would improve.

Since random vectors from Archimedean copulas are always equi-correlated, it does not make sense to consider widely diversified portfolios with stocks from many countries and industries. Regarding credit portfolios it is often important to model the joint return distributions of multiple names from the same country and industry.

In our empirical illustration we consider the six stocks comprising the USA S&P500 Chemicals. These are Air

†For the Gaussian and t-copulas the correlation coefficient (in standard notation) is $\rho = \sin(\tau/2) = 0.4540$ (Demarta and McNeil 2005).

Table 4. Descriptive statistics of the returns.

	Stock					
	AirPrds	Dow	DuPont	Eastm	Prax	R&H
Mean	1.030	0.797	0.364	0.189	1.324	1.139
SD	8.304	8.109	7.271	9.155	8.969	9.320
			Kendall's τ			
Air Prds.&Chems	1.000	0.294	0.428	0.336	0.451	0.391
Dow Chemicals	0.294	1.000	0.402	0.427	0.298	0.380
Du Ponte	0.428	0.402	1.000	0.387	0.391	0.431
Eastman Chems	0.336	0.427	0.387	1.000	0.321	0.348
Praxair	0.451	0.298	0.391	0.321	1.000	0.398
Rohm&Hass	0.391	0.380	0.431	0.348	0.398	1.000

Table 5. Goodness-of-fit tests for Archimedean copulas.

Copula family	CML estimate θ	Test statistic	p-Value
Cook–Johnson	0.693	29 580.2	<0.001
Frank	3.534	211.4	<0.001
Gumbel–Hougaard	1.500	14.4	0.710

Prds.&Chems, Dow Chemicals, Du Ponte, Eastman Chemicals, Praxair, and Rohm&Haas. Monthly stock prices from 1 January 1996 to 1 February 2006 were supplied by Datastream. The returns are calculated as the percentage monthly price change. The number of returns is $n = 121$ for each stock.

Table 4 gives some descriptive statistics of the returns. Note that neither the mean nor the standard deviation enter the test statistic, nor do, in fact, any other features of the marginal distributions. The values of Kendall's τ suggest that the implied assumption of equi-correlation is not blatantly violated.

Table 5 reports the test statistics and the p-values for the joint distribution of the six returns. The null hypothesis is that the dependence structure (as described by the copula) of the returns can be modeled by the respective family of Archimedean copulas; p-values were computed by the parametric bootstrap procedure with $R = 5000$ replications.

The test results show that the Gumbel–Hougaard copula fits the data best. The null hypothesis that the data generating process has a Gumbel–Hougaard copula cannot be rejected even at the 10% level. The Frank family and the Cook–Johnson family both fit the data very poorly.

6. Conclusion

The literature on copulas is mainly limited to the bivariate case, applications almost exclusively so. However, many applications in financial econometrics call for multivariate distributions, for example in credit portfolio management. This paper discusses multivariate Archimedean copulas and suggests a goodness-of-fit test. The test projects the (possibly high-dimensional) multivariate data into the unit interval, making use of certain properties of Archimedean copulas. Hence, we successfully avoid the problems in higher dimensions that trouble some goodness-of-fit tests for multivariate copulas. Even though the test statistic is based on the classical χ^2 statistic, its asymptotic distribution is non-standard. The null hypothesis states that the dependence structure of the data at hand can be captured by a particular parametric family of Archimedean copulas. The parameter (vector) need not be specified, but is estimated from the data using the canonical maximum likelihood method.

Simulation studies show that the test keeps its nominal level accurately. Further, the test has good power properties: if the data are generated by an Archimedean copula differing from that specified in the null hypothesis, the test is likely to detect that even for moderate sample sizes. Non-Archimedean data generating processes (e.g. jointly Gaussian or Student's t) are also likely to be detected by our test. A further important advantage of the test is that, for a given sample size, the power increases in the dimension d.

The empirical illustration concerns the joint return distribution of the six stocks contained in the USA S&P500 Chemicals index. We find that the family of Gumbel–Hougaard copulas captures the dependence structure of the return distribution best; both the Frank and Cook–Johnson families are rejected at the usual significance levels.

References

Barbe, P., Genest, C., Ghoudi, K. and Rémillard, B., On Kenndall's process. *J. multivar. Anal.*, 1996, **58**, 197–229.

Berg, D. and Bakken, H., A copula goodness-of-fit test based on the probability integral transform. Working Paper, University of Oslo, 2005.

Bouyé, E., Durrleman, V., Nikeghbali, A. and Roncalli, T., Copulas: an open field for risk management. Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, 2001.

Chen, X., Fan, Y. and Patton, A., Simple tests for models of dependence between multiple financial time series, with applications to U.S. equity returns and exchange rates. Discussion Paper 483, Financial Markets Group, London School of Economics, 2004.

Chernoff, H. and Lehmann, E.L., The use of maximum likelihood estimates in χ^2-tests for goodness of fit. *Ann. Math. Statist.*, 1954, **25**, 579–586.

Cherubini, U., Luciano, E. and Vecchiato, W., *Copula Methods in Finance*, 2004 (Wiley: Chichester).

Deheuvels, P., La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance. *Académie Royale de Belgique, Bulletin de la Classe des Sciences, 5e série*, 1979, **65**, 274–292.

Demarta, S. and McNeil, A.J., The *t* copula and related copulas. *Int. Statist. Rev.*, 2005, **73**, 111–129.

Dobrić, J. and Schmid, F., Testing goodness of fit for parametric families of copulas—application to financial data. *Commun. Statist. Simul. Comput.*, 2005, **34**, 1053–1068.

Durrleman, V., Nikeghbali, A. and Roncalli, T., Which copula is the right one? Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnaise, 2000.

Embrechts, P., Lindskog, F. and McNeil, A., Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance*, edited by S.T. Rachev, 2003 (Elsevier/North-Holland: Amsterdam).

Embrechts, P., McNeil, A. and Straumann, D., Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, edited by M. Dempster and H.K. Moffatt, 1999 (Cambridge University Press: Cambridge).

Fermanian, J.-D., Goodness of fit tests for copulas. *J. multivar. Anal.*, 2005, **95**, 119–152.

Fermanian, J.-D., Radulovic, D. and Wegkamp, M., Weak convergence of empirical copula processes. *J. Bernoulli Soc. Math. Statist. Probab.*, 2004, **10**, 847–860.

Fermanian, J.-D. and Scaillet, O., Some statistical pitfalls in copula modeling for financial applications. In *Capital Formation, Governance and Banking*, edited by E. Klein, pp. 57–72, 2005 (Nova Science).

Frees, E.W. and Valdez, E.A., Understanding relationships using copulas. *N. Am. Actuar. J.*, 1998, **2**, 1–25.

Genest, C., Ghoudi, K. and Rivest, L.-P., A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 1995, **82**, 543–552.

Genest, C. and MacKay, R.J., Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Can. J. Statist.*, 1986a, **14**, 145–159.

Genest, C. and MacKay, R.J., The joy of copulas: bivariate distributions with uniform marginals. *Am. Statist.*, 1986b, **40**, 280–283.

Genest, C., Quessy, J.-F. and Rémillard, B., Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.*, 2006, **33**, 337–366.

Genest, C. and Rivest, L.-P., Statistical inference procedures for bivariate archimedean copulas. *J. Am. Statist. Assoc.*, 1993, **88**, 1034–1043.

Joe, H., *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability 73, 1997 (Chapman & Hall: London).

Joe, H. and Xu, J.J., The estimation method of inference functions for margins for multivariate models. Technical Report No. 166, Department of Statistics, University of British Columbia, 1996.

Klugman, S.A. and Parsa, R., Fitting bivariate loss distributions with copulas. *Insur. Math. Econ.*, 1999, **24**, 139–148.

Malevergne, Y. and Sornette, D., Testing the Gaussian copula hypothesis for financial asset dependences. *Quant. Finan.*, 2003, **3**, 231–250.

Melchiori, M., Which Archimedean copula is the right one? Working Paper, Universidad Nacional del Litoral, Santa Fe, Argentina, 2003.

Nelsen, R.B., *An Introduction to Copulas*, Lectures Notes in Statistics 139, 1999 (Springer: New York).

Romano, C., Calibrating and simulating copula functions: an application to the Italian stock market. Working Paper No. 12, Centro Interdipartimentale sul Diritto e l'Economia dei Mercati, 2002.

Rosenblatt, M., Remarks on a multivariate transformation. *Ann. Math. Statist.*, 1952, **23**, 470–472.

Scaillet, O., Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. *J. Multivar. Anal.*, 2007, **98**, 533–543.

Schweizer, B. and Sklar, A., *Probabilistic Metric Spaces*, North-Holland Series in Probability and Applied Mathematics, 1983 (Elsevier: New York).

Wang, W. and Wells, M.T., Model selection and semiparametric inference for bivariate failure-time data. *J. Am. Statist. Assoc.*, 2000, **95**, 62–72.

Appendix A: Generator functions, inverses and derivatives

This appendix lists four important generator functions ϕ, their ith order derivatives $\phi^{(i)}$, their inverse function ϕ^{-1}, and the ith order derivatives of the inverse function $\phi^{-1(i)}$. Further, it gives the auxiliary function $f_i(t)$ necessary to compute the cdf $K(t)$ in (4).

Cook–Johnson family with parameter $\theta \geq 0$

- Generator

$$\phi(t) = \frac{1}{\theta}(t^{-\theta} - 1).$$

- Derivative of order i

$$\phi^{(i)}(t) = (-1)^i t^{-\theta-i} \prod_{j=1}^{i-1} (\theta + j).$$

- Inverse generator

$$\phi^{-1}(x) = (1 + x\theta)^{-1/\theta}.$$

- Derivative of inverse generator of order i

$$\phi^{-1(i)}(x) = (-1)^i (1 + \theta x)^{-(1+i\theta)/\theta} \prod_{j=0}^{i-1} (1 + j\theta).$$

- Auxiliary function

$$f_i(t) = (-1)^{i+1} t^{1+(i+1)\theta} \prod_{j=0}^{i} (1 + j\theta).$$

Frank family with parameter $\theta \in \mathbb{R}\backslash\{0\}$

- Generator

$$\phi(t) = -\ln\left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}\right).$$

- Derivative of order i

$$\frac{\mathrm{d}^i\phi(t)}{\mathrm{d}t^i}=\frac{\theta^i e^{-\theta t}}{(e^{-\theta t}-1)^i}\sum_{j=1}^{i-1}B_{i-1,j}\exp(-(j-1)\theta t),$$

where $B_{ij}=(i-j+1)B_{i-1,j-1}+jB_{i-1,j}$ and $B_{i1}=B_{ii}=1$.

- Inverse generator

$$\phi^{-1}(x)=1+\frac{x-\ln(1-\exp(\theta)+\exp(\theta+x))}{\theta}.$$

- Derivative of inverse generator of order i

$$\frac{\mathrm{d}^i\phi^{-1}(x)}{\mathrm{d}x^i}=\frac{(\exp(\theta)-1)\exp(\theta+x)}{(\exp(\theta)-\exp(\theta+x)-1)^i\theta}\times\sum_{j=1}^{i-1}B_{i-1,j}\exp((i-1-j)(x+\theta))(\exp(\theta)-1)^{j-1},$$

where $B_{ij}=(i-j+1)B_{i-1,j-1}+jB_{i-1,j}$ and $B_{i1}=B_{ii}=1$.

- Auxiliary function

$$f_i(t)=-\frac{(e^{-\theta t}-1)}{\theta e^{-\theta t}}e^{i\theta t}\sum_{j=1}^{i}(-1)^{j+1}B_{ij}\exp(-(i-j)\theta t),$$

where $B_{ij}=j(B_{i-1,j-1}+B_{i-1,j})$ and $B_{i1}=1$ and $B_{ii}=i!$

Gumbel–Hougaard family with parameter $\theta\geq 1$

- Generator

$$\phi(t)=(-\ln t)^{\theta}.$$

Derivative of order i

$$\frac{\mathrm{d}^i\phi(t)}{\mathrm{d}t^i}=\frac{(-\ln t)^{\theta}\theta}{(t\ln t)^i}\sum_{j=1}^{i}(-1)^{i+j}B_{ij}\cdot(\ln(t))^{i-j}\cdot\prod_{k=1}^{j-1}(\theta-k),$$

where $B_{ij}=(i-1)B_{i-1,j}+B_{i-1,j-1}$ and $B_{i1}=(i-1)!$ and $B_{ii}=1$.

- Inverse generator

$$\phi^{-1}(x)=\exp(-x^{1/\theta}).$$

- Derivative of inverse generator of order i

$$\frac{\mathrm{d}^i\phi^{-1}(x)}{\mathrm{d}x^i}=(-1)^i\frac{\exp(-x^{1/\theta})}{(\theta x)^i}x^{1/\theta}H_i(x^{1/\theta},\theta),$$

where
$H_i(s,t)=s(H_{i-1}(s,t)-(\mathrm{d}/\mathrm{d}s)H_{i-1}(s,t))+((i-1)t-1)\times H_{i-1}(s,t)$ and $H_1(s,t)=1$.

- Auxiliary function

$$f_i(t)=\frac{(-\ln t)^{-(i+1)\theta}t\ln t}{\theta^{i+1}}H_{i+1}(\ln(1/t),\theta),$$

where
$H_i(s,t)=s(H_{i-1}(s,t)-(\mathrm{d}/\mathrm{d}s)H_{i-1}(s,t))+((i-1)t-1)\times H_{i-1}(s,t)$ and $H_1(s,t)=1$.