DOCUMENTS DE TREBALL
DE LA FACULTAT DE CIÈNCIES
ECONÒMIQUES I EMPRESARIALS

Col·lecció d'Economia

Testing the bivariate distribution of daily equity returns using copulas. An application to the Spanish stock market

Oriol Roch and Antonio Alegre

Adreça correspondència:
Departament de Matemàtica Econòmica, Financera i Actuarial
Facultat de Ciències Econòmiques i Empresarials
Universitat de Barcelona
Avda. Diagonal 690
08034 Barcelona
Tel: +34-93-402-1951; Fax: +34-93-403-4892
Emails: oroch@ub.edu,

Abstract. In this paper we deal with the identification of dependencies between time series of equity returns. Marginal distribution functions are assumed to be known, and a bivariate chi-square test of fit is applied in a fully parametric copula approach. Several families of copulas are fitted and compared with Spanish stock market data. The results show that the t-copula generally outperforms other dependence structures, and highlight the difficulty in adjusting a significant number of bivariate data series.
Keywords: Copulas, Daily equity returns, Bivariate chi-square statistic, Risk Management.

Resum. En aquest article tractem amb la identificació de dependències entre series temporal de rendiments d'accions. Les distribucions marginals se suposen conegudes, i un test ji-quadrat bivariant s'aplica dins d'un enfocament totalment paramètric. Diverses famílies de còpules són ajustades i comparades amb dades de la Bolsa espanyola. Els resultats mostren que la t-còpula generalment supera altres estructures de dependència, i destaca la dificultat d'ajustar un nombre significant de sèries temporals bivariants.

1 Introduction

The distribution of daily equity returns is still a major problem in financial modelling. Kurtosis and asymmetry are usually reported for the individual returns, whereas recent literature has also highlighted the impact of the dependence between financial variables in the total risk of the portfolio, see e.g., Longin & Solnik [21] and Ang & Chen [1].

Embrechts, McNeil & Straumann [8], proposed the application of copula functions in order to model the multivariate distribution of returns. With a copula approach, the modelling of asset returns can be split in two steps:

1. Modelling the marginal distributions.
2. Modelling the dependence structure between marginal distributions.

This approach introduces a wide flexibility in the modelling of equity returns since several dependency structures can be generated via copula functions. Such a parameterization enables to model asymmetric dependence structures with joint fat tails.

For simplicity reasons, in most applications the distribution of equity returns is assumed to be a multivariate Gaussian distribution. This assumption generally leads to an underestimation of the portfolio's risk which can be avoided with alternative hypothesis with respect to the form of the multivariate risk distribution. Models of hedging, asset pricing and portfolio selection could also be improved with a better knowledge of the joint distribution of returns.

There are two important issues when trying to describe empirical data with a copula function. First, copula parameters must be estimated. Secondly, the best fitting copula must be chosen among the copulas available. In this context, some authors have applied statistical tests to copula specifications. In Junker & May [18] a bivariate χ^2 test is applied. However, as pointed out by Klugman & Parsa [19], the critical values of the test are not valid since they consider empirical distributions for the margins. Malavergne & Sornette [22] and Breymann, Dias & Embrechts [2] also deal with the testing of copula specifications. Semiparametric approaches have been proposed in Fermanian [10] and Chen, Fan & Patton [3], where tests based on the kernel smoothing approach are developed. In Genest et al. [12] various goodness of fit tests are proposed, but they are restricted to the arquimedean case. A recent paper similar in spirit to ours is that of Hürlimann [15], who presents some copula fitting results for bivariate daily cumulative returns between a market index and stocks. However, his analysis is restricted to the class of arquimedean copulas and different marginal distributions are considered.

In this paper we present some testing results in a fully parametric copula approach. We apply a bivariate chi-square test against some copula specifications, with marginal distributions parametrically specified. Marginal distributions are allowed to exhibit skewness and kurtosis since this kind of behaviour has been detected in the analyzed data. The chi-square test also enables us to select the best fitting copula for each pair of time series and to compute the number of pairs for which a copula model cannot be rejected.

The remainder of the paper is as follows. Section 2 briefly recalls some concepts about copulas. Section 3 describes the estimation methodology, whereas the chi-square testing procedure is explained in Section 4. Section 5 presents the empirical results obtained on the bivariate testing of different copula functions. Data from the Spanish stock market has been used for the analysis. Section 6 summarizes and concludes the paper.

2 Characterization of copulas

In this section we briefly recall some well known results on copula functions. A full treatment of copulas and their properties can be found in Joe [17] and Nelsen [25].

A copula function is a multivariate distribution function defined on the unit cube $[0,1]^d$, with uniformly distributed margins. Indeed, a function $C : [0,1]^d \to [0,1]$ is a d-dimensional copula if it satisfies the following properties:

1. For all $u_i \in [0,1]\,,\ C(1,...,1,u_i,1,...,1) = u_i$.
2. For all $u \in [0,1]^d\,,\ C(u_1,..,u_d) = 0$ if at least one coordinate u_i equals zero.
3. C is grounded and d-increasing, i.e., the C-measure of every box whose vertices lie in $[0,1]^d$ is non-negative.

The importance of the copula function relies on the fact that it captures the dependence structure of a multivariate distribution as we will see next.

Let $X = (X_1,...,X_d) \in \mathbb{R}^d$ be a random vector with multivariate distribution

$$F(x_1,...,x_d) = P(X_1 \le x_1,...,X_d \le x_d)\,,\ x = (x_1,...,x_d) \in \mathbb{R}^d$$

and continuous margins

$$F_n(x_n) = P(X_n \le x_n)\,,\ x_n \in \mathbb{R}.$$

According to Sklar's theorem [28], there exists a unique copula $C : [0,1]^d \to [0,1]$ of F such that for all x in $\mathbb{R}^d$,

$$\begin{aligned} F(x_1,...,x_d) &= P(X_1 \le x_1,...,X_d \le x_d) \\ &= C(F_1(x_1),...,F_d(x_d))\,. \end{aligned}$$

In general, the copula is affected by some parameters called "copula parameters", which are represented by the vector $\theta = (\theta_1,...,\theta_k)' \in \mathbb{R}^k$. The transformations $X_n \to F_n(x_n)$ used in the above representation are usually referred to as the probability-integral transformations (to uniformity) and form a standard tool in simulation methodology.

A fundamental conclusion of Sklar's theorem is that in multivariate continuous distribution functions the dependence structure and the margins can be separated, and the dependence structure can be represented by a copula. In a financial context, this key property allows us to fit the marginal distribution of each series of equity returns separately in a first step, and to model the dependence structure between different equity returns in a second step.

An important property of copulas is that they are invariant under strictly increasing transformations of the variables. Let $X = (X_1,...,X_d)$ be a vector of continuous random variables with copula C. If $g_1,...,g_d : \mathbb{R} \to \mathbb{R}$ are strictly increasing on the range of $X_1,...,X_d$, then

$(g_1(X_1), ..., g_d(X_d))$ has copula C. This property implies that the dependence structure between the variables is captured by the copula, regardless of the margins.

Three fundamental functions related to copulas are

$$\begin{aligned} M(u_1, ..., u_d) &= \min(u_1, ..., u_d), \\ \Pi(u_1, ..., u_d) &= u_1 \cdots u_d, \\ W(u_1, ..., u_d) &= \max(u_1 + ... + u_d - d + 1, 0). \end{aligned}$$

These functions derive form the Fréchet-Hoeffding bounds, and as a consequence of Sklar's theorem, it can be proved that for any copula $C(u_1, ..., u_d)$,

$$W(u_1, ..., u_d) \leq C(u_1, ..., u_d) \leq M(u_1, ..., u_d),$$

i.e., any copula is constrained between the Fréchet-Hoeffding bounds.

The function $M(u_1, ..., u_d)$ is called the Comonotonic copula and represents perfect positive dependence between the variables, whereas the $\Pi(u_1, ..., u_d)$ function is called the Product copula and represents independence between the variables. Unlike the two previously mentioned functions, $W(u_1, ..., u_d)$ is not a copula for $d \geq 3$, although is the best possible lower bound for d-copulas. For $d = 2$, $W(u_1, ..., u_d)$ is called the Countermonotonic copula, and represents perfect negative dependence between the variables.

Throughout the paper we shall consider several one-parameter and two-parameter copulas to model bivariate daily equity returns. We now introduce the families of copulas that will be employed in our empirical study. We use the notation $C(u, v; \theta)$, $\theta \in \Theta$ to denote the bivariate parametric families of copulas.

Normal copula. The most important copula in financial literature is the Normal copula. Let Φ be the $N(0,1)$ cumulative distribution function with correlation ρ. The bivariate Normal copula is defined by

$$C(u, v; \rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) dsdt,$$

with $-1 \leq \rho \leq 1$.

***T*-copula**. Another important copula in finance is the t-copula. It belongs to the same class of copulas as the Normal copula (elliptical copulas), and add joint fat tails to the Normal copula. It has been recommended by several authors as Mashal & Zeevi [23] and Breymann, Dias & Embrechts [2].

Let t_υ denote the univariate Student's t cumulative distribution function with υ degrees of freedom. The bivariate t-copula with υ degrees of freedom and correlation ρ is given by

$$C(u, v; \rho, \upsilon) = \int_{-\infty}^{t_\upsilon^{-1}(u)} \int_{-\infty}^{t_\upsilon^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{\upsilon(1-\rho^2)}\right)^{-\frac{(\upsilon+2)}{2}} dsdt,$$

with $-1 \leq \rho \leq 1$, $\nu > 2$.

Note that the Normal copula is a limiting case of the t-copula when $\upsilon \to \infty$. For a detailed discussion on the t-copula consult Demarta & McNeil [6].

Another class of copulas, the arquimedean copulas, are also used in this paper. Unlike the copulas presented so far, arquimedean copulas are not derived from multivariate distributions using Sklar's theorem. The bivariate arquimedean copulas are generated by

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

where $\varphi : [0,1] \to [0,\infty]$ is a continuous and strictly decreasing convex function with $\varphi(1) = 0$, and $\varphi^{[-1]}$ is the pseudo-inverse of φ, defined as

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

The function φ is called the *generator* of the copula.

From now onwards we reference the families of arquimedean copulas with numbers for the sake of a subsequent empirical application, giving more references when possible. Fundamental sources are Nelsen [25] for one-parameter families of copulas, and Joe [17] for two-parameter families of copulas. From these sources we have selected the most appropriate copulas for modelling equity returns.

Arq. family 1. This family is known as Frank family, and it is probably the one-parameter arquimedean family of copulas most employed in finance. The Frank family is defined as

$$C(u,v;\alpha) = -\frac{1}{\alpha}\ln\left(1 + \frac{(\exp(-\alpha u) - 1)(\exp(-\alpha v) - 1)}{\exp(-\alpha) - 1}\right), \quad \alpha \in \mathbb{R}\backslash\{0\}.$$

Arq. family 2. This copula can be found in Nelsen [25] as family number 13, and it is given by

$$C(u,v;\alpha) = \exp\left(1 - ((1 - \ln u)^{\alpha} + (1 - \ln v)^{\alpha} - 1)^{\frac{1}{\alpha}}\right), \quad \alpha \in (0,\infty).$$

Arq. family 3. This copula can be found in Nelsen [25] as family number 17, and it is given by

$$C(u,v;\alpha) = \left(1 + \frac{\left((1+u)^{-\alpha} - 1\right)\left((1+v)^{-\alpha} - 1\right)}{2^{-\alpha} - 1}\right)^{-\frac{1}{\alpha}} - 1, \quad \alpha \in \mathbb{R}\backslash\{0\}.$$

Arq. family 4. This two-parameter family of copulas is also known as the Clayton-Gumbel family, and it is referenced in Joe [17] as family BB1.

$$C(u,v;\alpha,\delta) = \left(1 + \left[\left(u^{-\alpha} - 1\right)^{\delta} + \left(v^{-\alpha} - 1\right)^{\delta}\right]^{\frac{1}{\delta}}\right)^{-\frac{1}{\alpha}}, \quad \alpha > 0,\ \delta \geq 1.$$

Arq. family 5. This family can be found in Joe [17] as family BB3, and it is defined as

$$C(u,v;\alpha,\delta) = \exp\left(-\left[\delta^{-1}\ln\left(\exp\left(\delta(-\ln u)^{\alpha}\right) + \exp\left(\delta(-\ln v)^{\alpha}\right) - 1\right)\right]^{\frac{1}{\alpha}}\right),$$

with $\alpha \geq 1$, $\delta > 0$,

Arq. family 6. This family is referenced in Joe [17] as family BB7. It is also used in Patton [26] with the name of Joe-Clayton copula, and it is given by

$$C(u,v;\alpha,\delta) = 1 - \left(1 - \left[(1-(1-u_1)^{\alpha})^{-\delta} + (1-(1-v)^{\alpha})^{-\delta} - 1\right]^{-\frac{1}{\delta}}\right)^{\frac{1}{\alpha}},$$

with $\alpha \geq 1$, $\delta > 0$.

We also introduce a new family of bivariate two-parameter copulas. From $\varphi(t) = ((1-\ln t)^{\alpha} - 1)^{\delta}$, and by applying the definition of bivariate arquimedean copula, we can obtain a new family of copulas that satisfies all conditions required for a function to be a copula. The generator is obtained as a transformation of the generator in arquimedean family 2. The same procedure is applied in Junker & May [18] for the Frank copula generator. This method will be useful to analyze the effect of adding a new parameter into a pre-existing one-parameter copula.
Arq. family 7. This new family of aquimedean copulas is given by

$$C(u,v;\alpha,\delta) = \exp\left(1 - \left[\left(((1-\ln u)^{\alpha}-1)^{\delta} + ((1-\ln v)^{\alpha}-1)^{\delta}\right)^{\frac{1}{\delta}} + 1\right]^{\frac{1}{\alpha}}\right),$$

with $\alpha \geq 1$, $\delta > 0$.

The fitting of the nine families of copulas presented in this section will be compared for the empirical results. In the next two sections, the estimation of the copula parameters and the testing procedure will be detailed.

3 Copula calibration

The first step in order to calibrate the copula is to transform the data using the probability-integral transformation. It can be done in different ways. A nonparametric approach can be followed by computing the empirical cdf for the margins. In that case, the nth marginal cdf is estimated by

$$\hat{F}_n(x) = \frac{1}{T+1}\sum_{t=1}^{T} 1_{\{X_{tn} \leq x\}}, \quad n = 1, ..., d$$

where X_{tn} is the t component of the nth data vector and T is the size of the random sample. Note that the $T+1$ term is needed to avoid points in the boundary of the unit cube, which would cause the impossibility of estimate the copula parameters by maximum likelihood.

Another possibility is to find a parametric distribution which fits the univariate marginal data. This method requires an appropriate distribution for each margin. For the case of equity returns, the distribution of the margins is usually assumed to be Gaussian. However, this generalization often leads to an incorrect specification of the parametric distribution, due to the evidence of heavy tails and skewness showed by the returns. A general approach to estimate the parameters of the margins is the method of maximum likelihood. Given the sample $\{X_t\}_{t=1}^{T}$ we set the log-likelihood function to be

$$L(\gamma; x) = \sum_{t=1}^{T} \ln f(x_t; \gamma).$$

The maximum likelihood estimator is then the value of $\gamma \in \Theta$ maximizing $L(\gamma; x)$, i.e.,

$$\hat{\gamma}_{ML} = \arg\max L(\gamma; x),$$

where $\hat{\gamma}_{ML}$ has the property of asymptotic normality (Davidson & Mackinnon [5]). For large samples, and under certain regularity conditions, it can be shown that

$$\sqrt{T}(\hat{\gamma} - \gamma)' \xrightarrow{d} N\left(0, I^{-1}(\theta)\right),$$

where $I^{-1}(\theta)$ is the inverse of the information matrix whose ij-th element satisfies

$$I_{ij}(\gamma) = E\left[-\frac{1}{n}\frac{\partial}{\partial\gamma_i\partial\gamma_j}L(\gamma; x)\right].$$

It is important that the selected marginal distribution allows for skewness and excess kurtosis, specially when dealing with high-frequency data. A correct selection of the margins is even more important than a correct selection of the copula. Several functional forms of univariate distributions suitable for modelling equity returns can be found in the appendix. Note that in some distributions skewness has been introduced in the way of Fernández & Steel [11].

Once the margins have been transformed, they are embedded into the copula in order to estimate the copula parameters. Given a random sample $X = (X_{t1}, X_{t2}, \ldots, X_{td})_{t=1}^{T}$, it is common to calibrate the copula parameters by the method of maximum likelihood. Assuming known margins, this approach implies the maximization of the function

$$L_c(\theta) = \sum_{t=1}^{T} \ln c\left(F_1(x_{t1}; \hat{\gamma}_1), \ldots, F_d(x_{td}; \hat{\gamma}_d); \theta\right),$$

where $c(u_1, ..., u_d)$ is the copula density, which can be obtained from

$$c(u_1, ..., u_d) = \frac{\partial C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}.$$

The two-step procedure explained above is known as *inference functions for the margins* (IFM) and exploits the basic idea of copulas, i.e., the separation between the margins and the dependence structure. More details about this method are studied in McLeish & Small [24] and Xu [30]. Let $\hat{\psi} = \left(\hat{\gamma}_1, \ldots, \hat{\gamma}_d, \hat{\theta}\right)$ be the row vector of parameters estimated by IFM. The IFM method also verifies the property of asymptotic normality, which results (Joe [17])

$$\sqrt{T}\left(\hat{\psi} - \psi\right)' \xrightarrow{d} N(0, V)$$

where $V = D_g^{-1} M_g \left(D_g^{-1}\right)'$, with $g(\psi) = (\partial_{\gamma_1} L_1, \ldots, \partial_{\gamma_d} L_d, \partial_\theta L_c)$, $D_g = E[\partial g'(\psi)/\partial\psi]$ and $M_g = E[g(\psi)'\ g(\psi)]$.

When the empirical distributions of the margins are used, the two-step procedure is called *canonical maximum likelihood method* or CML. This method is similar to IFM, except on the fact that the copula can be calibrated without specifying the margins. Once the data has been transformed into uniform variates $\hat{u}_t = (\hat{u}t_{t1}, ..., \hat{u}_{td})_{t=1}^{T}$ using the empirical distribution, the estimation of the copula parameters implies the maximization of the function

$$L(\theta) = \sum_{i=1}^{T} \ln c\left(\hat{u}_{t1}, ..., \hat{u}_{td}; \theta\right).$$

Alternatively, the parameters of the margins and the parameters of the copula can be jointly estimated. The likelihood would result

$$L(\theta) = \sum_{t=1}^{T} \ln c\left(F_1\left(x_{t1}; \gamma_1\right), \ldots, F_d\left(x_{td}; \gamma_d\right); \theta\right) + \sum_{t=1}^{T} \sum_{n=1}^{d} \ln f_n\left(x_{tn}; \gamma_n\right)$$

However, this method could be high time-consuming in case of high dimensional distributions.

4 Testing procedure

In empirical applications, it is common to calibrate several copulas and to compare the results obtained with one or more selection criteria. Different criteria have been proposed in the literature in order to select the best fitting copula, e.g., the AIC or by computing distances between each considered copula and the empirical copula (Durrleman et al. [7]). However, those criteria are not sufficient to accept a copula as representative enough. Since we are interested in compare both elliptical and arquimeden copulas, and the aim of the paper is to validate different copula specifications, we use a standard bivariate chi-square goodness of fit (GOF) test. The aim is to test the null hypothesis

$$H_0 : C(u, v) = C(u, v; \theta) \quad \text{for a } \theta \in \Theta,$$

i.e., the unknown copula $C(u, v)$ is a member of the parametric family, against the alternative hypothesis

$$H_A : C(u, v) \neq C(u, v; \theta) \quad \text{for a } \theta \in \Theta,$$

i.e., the unknown copula $C(u, v)$ is not a member of the parametric family.

We assume that observations are independent and identically distributed. Otherwise, the chi-square test would not hold. This assumption is also made in Mashal & Zeevi [23], Malavergne & Sornette [22] and implicitly in Hürlimann [15].

As we have seen in previous sections, there are different possibilities for modelling the univariate marginal distributions. When a semiparametric approach is applied, transforming the data with the empirical distribution function, the standard chi-square test won't work . The semiparametric approach has been studied in Fermanian [10] and Chen, Fan & Patton [3]. On the other hand, if known margins are assumed using parametric models for the individual margins, the standard chi-square test will hold. However, each marginal distribution model must be accepted, which requires some previous estimation and validation procedures.

In the sequel, let $A_{ij} = \left[\frac{i-1}{r}, \frac{i}{r}\right] \times \left[\frac{j-1}{s}, \frac{j}{s}\right] \subset [0,1]^2$, $i = 1, ...r$ and $j = 1, ..., s$, be bins of $[0,1]^2$, i.e., the y axis is divided into r equidistant parts and the x axis is divided into s equidistant parts. Let O_{ij} be the observed frequency for bin A_{ij}, and let

$$E_{ij}\left(\hat{\theta}\right) = np_{ij}\left(\hat{\theta}\right)$$

be the expected frequency for A_{ij}, where

$$p_{ij}\left(\hat{\theta}\right) = \int\int_{A_{ij}} dC\left(u, v; \hat{\theta}\right).$$

The chi-square statistic is

$$\chi^2\left(\hat{\theta}\right) = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(O_{ij} - E_{ij}\left(\hat{\theta}\right)\right)^2}{E_{ij}\left(\hat{\theta}\right)},$$

where $\hat{\theta}$ is the vector of parameters estimated.

The test statistic follows, approximately, a chi-square distribution with $(rs - 1 - d)$ degrees of freedom, where rs is the number of non-empty bins and d is the number of estimated parameters. The critical value will be the $(1-\alpha)$ quantile of the χ^2_{rs-1-d} distribution, being the H_0 rejected if $\chi^2\left(\hat{\theta}\right) > \chi^2_{(\alpha, rs-1-d)}$.

Note that for the chi-square approximation to be valid, the expected frequency should be at least 5. Otherwise, the data must be regrouped so that $E_{ij}\left(\hat{\theta}\right) > 5\ \forall i, j$.

5 Numerical Results

In this section we consider daily equity returns for 16 companies of the Spanish stock market from 1/2/1997 to 11/11/2004[1]. The time period practically covers a whole economic cycle. The selected companies are "large caps" within the Spanish stock market, which ensures that the analysis is not affected by liquidity problems. A significant dependence between equity returns is expected by the effect of cross participations, investment strategies as index tracking and the simple fact that some companies belong to the same economic sector.

Our analysis is restricted to the bivariate case. An accurate bivariate distribution of returns is specially important in order to compute prices of exotic options (Cherubini & Luciano [4]) and risk measures (Junker & May [18]). Although we have selected "large caps", similar analysis could be extended, e.g., to indices (Hu [16]), exchange rates (Patton [27]) or "small caps".

Let $S = (S_{t1}, S_{t2}, ..., S_{td})_{t=0}^{T}$ be the time-series of adjusted equity prices. The returns are defined as

$$X_{tn} = \ln\left(\frac{S_{tn}}{S_{t-1n}}\right), \quad n = 1, 2, ..., d, \quad t = 1, ..., T,$$

where d represents the number of equities, T the number of trading days, and being the vector $X_t = (X_{t1}, ..., X_{td})$ the corresponding to the t'th day.

The data yields 1968 observations of daily returns for each equity. Generally, the data exhibits positive skewness and excess kurtosis, which provokes the "fat tails" effect, rejecting the usually assumption of normal distribution of returns. Prior to the calibration and selection of the copula models it seems essential to test the data for independence. A standard procedure, based on Kendall's tau, requires to compute the statistic $\xi_T = 3\sqrt{T}|\tau_T|/2$. The null hypothesis of independence should be rejected if $\xi_T > \Phi^{-1}(0.975) = 1.96$, where τ_T is Kendall's tau statistic

[1]The 16 equities selected are: Abertis, Acciona, Acerinox, ACS, Aguas de Barcelona, Altadis, Banco Popular, Bankinter, BBVA, Corp. Alba, Endesa, FCC, NH Hoteles, Repsol, Santander, and Telefónica.

and T the sample size, which must be large enough. When computing the statistic with our data, ξ_T ranges from 5.2533 to 66.5432, which implies that the null hypothesis of independence is clearly rejected for every data pair. If the null hypothesis of independence was accepted, the product copula could be a good model to represent the dependence structure between the variables. In any case, as some families of copulas include the product copula as special case[2], the acceptance of the null hypothesis would not deeply affect our analysis. More detailed tests of independence related to copulas can be found in Genest & Rémillard [13].

The testing methodology proceeds as follows:

1. Fitting the univariate marginal distribution of each series of equity returns by maximum likelihood estimation. The distributions considered, all of which accept for skewness, have been collected in the appendix. Among the different tested distributions, we select the one that best fits the data. The selection criterion is the higher log-likelihood value. Note that once a distribution has been estimated and selected, it must be accepted in order to employ the standard chi-square test. In our analysis a K-S test was applied, and all the selected marginal distributions were accepted for a significance level $\alpha = 0.05$.

2. Transform each data vector into uniform variates using the probability-integral transformation.

3. For each pair of transformed data vectors, plug the data into a family of copulas and estimate the copula parameters by maximum likelihood as described in section 3. Repeat this procedure with all the families of copulas considered.

4. Apply the bivariate chi-square test of goodness of fit described in section 4 for each family of copulas estimated.

All the results were obtained using Matlab software. The results of the univariate marginal fitting are reported in Table 1. The values of the log-likelihood at the optimum are compared in order to select the best fitting marginal distribution (higher values are highlighted in bold fonts). The SGED distribution is the most selected, followed by the skewed t-location scale distribution. The skewed laplace also performs quite well for high-frequency data, as noted in Linden [20]. The worst performance is obtained by the skewed cauchy, which has only been used for the sake of a comparison.

The combinatory of the data vectors yields 120 pairs. Each pair is fitted with all the nine copulas described in section 2. Then, the bivariate chi-square GOF test is applied to each copula estimated. For every test, the unit square $[0, 1]^2$ is divided into bins of equal area, being $r = s$, and rearranging bins when needed. As the analysis is quite sensitive to the selection of the number of bins, the results are reported for different choices of r and s. In tables 2, 3 and 4 we examine the goodness of fit of the different copula models. Table 2 shows the percentage of rejection for each family of copulas given a significance level $\alpha = 0.1$, table 3 given a significance level of $\alpha = 0.05$, and finally, table 4 given $\alpha = 0.01$.

[2] In our analysis, all the families of copulas except the arquimedean family number 6.

Table 1: Log-likelihood values at the optimum

	Skewed T-Loc	**Kappa**	**Skewed Logistic**	**Skewed Cauchy**	**Skewed Laplace**	**SGED**
Abertis	5432.7	5425.6	5423.6	5242.4	5437.2	**5441.7**
Acciona	5171.9	5117.7	5114.5	5031.5	5171.2	**5177.4**
Acerinox	4819.7	4815.2	4812.2	4610.9	4812.2	**4820.9**
ACS	4970.6	4956.2	4955.5	4779.1	4971.5	**4973.1**
Aguas Bar.	5307.5	5258.8	5296.9	5103.9	5301.9	**5309.1**
Altadis	**5017.9**	5008.8	5009.5	4796.7	4996.7	5010.1
B. Popular	5209.7	5201.2	5200.8	5001.1	5201.6	**5210.3**
Bankinter	4990.3	4961.9	4960	4826.2	4994.2	**4995.1**
BBVA	**4758**	4733.4	4731.6	4570.7	4746.1	4748.4
Corp. Alba	5038.6	5032.7	5031.1	4581.2	5048.3	**5051.7**
Endesa	5177.6	5172.8	5171.1	4973.8	5171.2	**5180.3**
FCC	4954.7	4952.7	4949.3	4763.2	4963.1	**4969.1**
NH Hoteles	**4894.2**	4853.7	4852.3	4729.1	4887.4	4887
Repsol	**5244.4**	5235.1	5231.8	5041	5235.6	5241.9
Santander	**4675.2**	4658.9	4655.9	4487.6	4668.4	4671.6
Telefónica	4652.4	4654.4	4652.4	4420.9	4651.6	**4655.2**

As it was expected, two-parameter copulas fit better than one-parameter copulas. In particular, note the considerable improvement of arquimedean family 2 when it is extended with a new parameter (case of arquimedean familiy 7). The Normal copula, our benchmark due to its relevance in the financial world, is highly rejected. This result contradicts those of Malaverge & Sornette [13] and Chen, Fan & Patton [4] for the bivariate case. If we attend to the best fitting copula, the t-copula is clearly the one that fits better, although it is rejected nearly 30% of the times at $\alpha = 0.05$. Other two-parameter families don't achieve the degree of fitting of the t-copula. However, nearly 50% of the data pairs can be well represented with these families at $\alpha = 0.05$. These results highlight the difficulty in fitting the bivariate distribution of daily equity returns.

The chi-square GOF test can also be used as a criterion for choosing a copula to represent the sample data. Generally, the lower the chi-square statistic, the better the copula represents the data. However, in this case we compare statistics distributed with different degrees of freedom. In order to select the best fitting copula, we compare the p-value of the contrasts. For each pair of data vectors we compare the fitting of the nine copulas, and the selected will be that with higher p-value. Table 5 summarizes the number of times a copula has been selected according to this criterion.

The t-copula is the most selected one. Although other bivariate copulas have been selected, the t-copula clearly outperforms other copula specifications. Items for future research are the extensions of these results to the multivariate case.

The overall results suggest that the t-copula provides good accuracy for the bivariate case, which had been previously reported in the literature. However, an interesting result is obtained when considering the number of pairs that could be accepted given all nine considered copulas.

Table 2: % of rejections for $\alpha = 0.1$

Fam.	r=s=8	r=s=10	r=s=12	r=s=14	r=s=16
Normal	90.00	89.17	84.17	79.17	88.33
T-Copula	45.00	50.83	45.00	44.17	53.33
Arq. 1	91.67	91.67	81.67	87.50	95.83
Arq. 2	80.83	93.33	85.00	89.17	96.67
Arq. 3	90.83	90.00	80.00	80.83	94.17
Arq. 4	67.50	60.00	55.00	53.33	59.17
Arq. 5	67.50	60.83	55.83	55.83	61.67
Arq. 6	70.00	66.67	59.17	56.67	60.00
Arq. 7	72.50	72.50	62.50	60.00	68.33

Table 3: % of rejections for $\alpha = 0.05$

Fam.	r=s=8	r=s=10	r=s=12	r=s=14	r=s=16
Normal	82.50	84.17	75.00	67.50	81.67
T-Copula	28.33	33.33	31.67	31.67	40.83
Arq. 1	86.67	84.17	77.50	77.50	88.33
Arq. 2	73.33	89.17	80.00	79.17	90.00
Arq. 3	84.17	82.50	74.17	75.00	81.67
Arq. 4	49.17	50.83	45.83	46.67	51.67
Arq. 5	53.33	50.83	45.83	48.33	54.17
Arq. 6	55.83	52.50	47.50	48.33	54.17
Arq. 7	66.67	60.38	51.67	52.50	57.50

That result is a crucial result when trying to explore the possibilities of copulas for modelling equity returns. With our data, we found that in the 57.50% of the pairs, the H_0 could not be rejected at a significance level $\alpha = 0.1$, the 76.67% at a significance level of $\alpha = 0.05$ and the 90.83% at a significance level of $\alpha = 0.01$. This means that given a significance level of 5%, in the 23.33% of the pairs none of the nine copula specifications could be accepted. The extreme behaviour of these pairs requires further research in this area.

6 Conclusions

In this paper we have analyzed different models of dependence structures for equity returns within a copula approach. The bivariate distribution of time series has been tested with Spanish stock market data. A total of nine copulas have been fitted and compared. Using a fully parametric chi-square test of fit, we have found statistical evidence against the Normal copula, and a better performance for the t-copula. According to this test procedure, a significant number of data pairs could not be fitted properly, which shows the need of further research in order to generate more flexible distribution functions for the financial modelling.

New solvency regulations and the fast growth of multivariate instruments have turned out this problem into a crucial issue. The knowledge of the multivariate joint distribution function,

Table 4: % of rejections for $\alpha = 0.01$

Fam.	r=s=8	r=s=10	r=s=12	r=s=14	r=s=16
Normal	65.83	64.17	56.67	50.00	50.83
T-Copula	9.17	16.67	18.33	10.00	16.67
Arq. 1	68.33	74.17	61.67	63.33	68.33
Arq. 2	42.50	76.67	64.17	65.00	71.67
Arq. 3	65.00	66.67	53.33	55.83	57.50
Arq. 4	25.83	27.50	25.83	23.33	21.67
Arq. 5	27.50	29.17	26.67	24.17	25.00
Arq. 6	30.00	29.17	28.33	24.17	25.83
Arq. 7	37.5	39.17	36.67	30.38	32.50

Table 5: Number of times a copula has been selected

Fam.	r=s=8	r=s=10	r=s=12	r=s=14	r=s=16
Normal	0	0	0	0	0
T-Copula	93	92	94	86	73
Arq. 1	0	0	0	0	0
Arq. 2	1	0	0	0	0
Arq. 3	0	0	0	0	0
Arq. 4	9	12	10	9	16
Arq. 5	5	6	5	10	11
Arq. 6	6	4	5	4	8
Arq. 7	6	6	6	11	12

and the dynamic dependence structure of equity returns over time keep as an important line of future research.

Acknowledgements

The authors would like to express their thanks to the Spanish *Ministerio de Educación y Cultura* for the project *Becas FPU*, which has partially supported this paper.

Appendix

The appendix collects some functional forms of the distributions used in the univariate marginal fitting. Skewness has been allowed in some distributions introducing inverse scale factors in the positive and the negative orthand. This approach was introduced in Fernández and Steel [11].

The cdf will be denoted by F, and the pdf by f.

a. Skewed logistic:

$$\begin{aligned} f(x) &= \frac{2}{\gamma+\gamma^{-1}}\left(g\left(\frac{x}{\gamma}\right)I_{[0,\infty]}+g\left(x\gamma\right)I_{(-\infty,0)}\right), \text{ where} \\ g\left(x\right) &= \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left(1+\exp\left(\frac{x-\mu}{\sigma}\right)\right)^2}, \quad \sigma>0. \end{aligned}$$

b. Skewed t-location scale

$$\begin{aligned} f(x) &= \frac{2}{\gamma+\gamma^{-1}}\left(g\left(\frac{x}{\gamma}\right)I_{[0,\infty]}+g\left(x\gamma\right)I_{(-\infty,0)}\right), \text{ where} \\ g\left(x\right) &= \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sigma\sqrt{\upsilon\pi}\Gamma\left(\frac{\upsilon}{2}\right)}\left[\frac{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}{\upsilon}\right]^{-\left(\frac{\nu+1}{2}\right)}, \quad \sigma>0,\nu>0. \end{aligned}$$

c. Four-parameter Kappa from Hosking [14]

$$\begin{aligned} F\left(x\right) &= \left(1-h\left[1-k\left(x-\xi\right)/\alpha\right]^{\frac{1}{k}}\right)^{\frac{1}{h}}, \\ f\left(x\right) &= \alpha^{-1}\left(1-k\left(x-\xi\right)/\alpha\right)^{\frac{1}{k}-1}\left[F\left(x\right)\right]^{1-h}, \quad \alpha>0. \end{aligned}$$

d. Skewed Cauchy

$$\begin{aligned} f\left(x\right) &= \frac{2}{\gamma+\gamma^{-1}}\left(g\left(\frac{x}{\gamma}\right)I_{[0,\infty]}+g\left(x\gamma\right)I_{(-\infty,0)}\right), \text{ where} \\ g\left(x\right) &= \frac{1}{\pi}\frac{1}{\lambda^2+\left(x-\alpha\right)^2}, \quad \lambda>0. \end{aligned}$$

e. Skewed Laplace. More details about this distribution can be found in Linden [20].

$$f\left(x\right)=\begin{cases} \frac{b_1}{2}\exp\left(-b_1|x-\mu|\right) & x>0 \\ \frac{b_2}{2}\exp\left(-b_2|x-\mu|\right) & x\leq 0, \end{cases}$$

with b_1, $b_2>0$.

f. Skewed Generalized Error Distribution (SGED). Introduced in Theodossiou [29], it is an asymmetric version of the generalized error distribution.

$$
\begin{aligned}
f(x) &= \frac{C}{\sigma}\exp\left(-\frac{1}{\left[1-sign\left(x-\mu+\delta\sigma\right)\lambda\right]^{k}\theta^{k}\sigma^{k}}\left|x-\mu+\delta\sigma\right|^{k}\right), \\
\sigma &> 0,\ k>0,\ -1<\lambda<1, \quad \text{where} \\
C &= \frac{k}{2\theta}\Gamma\left(\frac{1}{k}\right)^{-1}, \\
\theta &= \Gamma\left(\frac{1}{k}\right)^{\frac{1}{2}}\Gamma\left(\frac{3}{k}\right)^{-\frac{1}{2}}S\left(\lambda\right)^{-1}, \\
\delta &= 2\lambda AS\left(\lambda\right)^{-1},\ \text{with} \\
S\left(\lambda\right) &= \sqrt{1+3\lambda^{2}-4A^{2}\lambda^{2}}, \\
A &= \Gamma\left(\frac{2}{k}\right)\Gamma\left(\frac{1}{k}\right)^{-\frac{1}{2}}\Gamma\left(\frac{3}{k}\right)^{-\frac{1}{2}}.
\end{aligned}
$$

References

[1] Ang, A. and Chen, J., 2002. Correlations in equity portfolios. Journal of Financial Economics, 63(3) 443-494.

[2] Breymann, W., Dias, A. and Embrechts, P., 2003. Dependence structures for multivariate high-frequency data in finance. Quantitative Finance, 3(1) 1-16.

[3] 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. Mimeo. (Financial Markets Group, London School of Economics, Discussion paper 483, 2004).

[4] Cherubini, U. and Luciano, E., 2002. Bivariate option pricing with copulas. Applied Mathematical Finance, 8 69-85.

[5] Davidson, R. and MacKinnon, J., 1993. Estimation and Inference in Econometrics. Oxford University Press, Oxford.

[6] Demarta, S., McNeil, A.J., 2005. The t copula and related copulas. International Statistical Review (to appear).

[7] Durrleman, V., Nikeghbali, A. and Roncalli, T., Which copula is the right one?. Working Paper (Groupe de Recherche Opérationnelle Crédit Lyonnais, 2000).

[8] Embrechts, P., McNeil, A.J. and Straumann, D., 2001. Correlation and dependency in risk management: properties and pitfalls. In: M. Dempster, H. Moffatt (Ed.), Risk Management: Value at Risk and Beyond, Cambridge University Press, 176-223.

[9] Fees, E.W. and Valdez, E.A., 1998. Understanding relationships using copulas. North American Actuarial Journal, 2 1-25.

[10] Fermanian, J-D., Goodness of Fit Tests for Copulas. Working paper (CREST, 2003). To appear in Journal of Multivariate Analysis.

[11] Férnandez, C. and Steel, M.F.J., 1998. On Bayesian modelling of fat tails and skewness. Journal of the American Statistical Association, 93 359-371.

[12] Genest, C., Quessy, J.F. and Rémillard, B., 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. Scandinavian Journal of Statistics (to appear).

[13] Genest, C. and Rémillard, B., 2004. Tests of independence and randomness based on the empirical copula process. TEST, 13 335-370.

[14] Hosking, J.R.M., 1994. The Four-Parameter Kappa Distribution. IBM Jounal of Research and Development, 38(3) 251-258.

[15] Hürlimann, W., 2004. Fitting bivariate cumulative returns with copulas. Computational Statistics & Data Analysis, 45 355-372.

[16] Hu, L., 2005. Dependence patterns across financial markets: a mixed copula approach. Applied Financial Economics (to appear).

[17] Joe, H., 1997. Multivariate models and dependence concepts. Chapman and Hall, London.

[18] Junker, M. and May, A., Measurement of aggregate risks with copulas. Preprint (Caesar, 2002).

[19] Klugman, S.A. and Parsa, R., 1999. Fitting bivariate loss distributions with copulas. Insurance: Mathematics & Economics, 24 139-148.

[20] Linden, M., 2001. A model for stock return distribution. International Journal of Finance and Economics, 6 159-169.

[21] Longin, F. and Solnik, B., 2001. Extreme correlation of international equity markets. Journal of Finance, LVI(2) 649-676.

[22] Malavergne, Y. and Sornette, D., 2003. Testing the Gaussian copula hypothesis for financial asset dependences. Quantitative Finance, 3 231-250.

[23] Mashal, R. and Zeevi, A., Beyond correlation: extreme co-movements between financial assets. Mimeo. (Columbia Graduate School of Business, 2002).

[24] McLeish, D.L. and Small, C.G., 1988. The theory and applications of statistical inference functions. In: Lecture Notes in Statistics, Vol. 44. Springer, New York.

[25] Nelsen, R., 1999. An introduction to copulas. Springer, New York.

[26] Patton, A.J., 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. Journal of Financial Econometrics, 2(1) 130-168.

[27] Patton, A.J., 2005. Modelling asymmetric exchange rate dependence. International Economic review (to appear).

[28] Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris, 8 229-231.

[29] Theodossiou, P. Skewness and kurtosis in financial data and the pricing of options. Working paper (Rutgers University, 2002).

[30] Xu, J.J., 1996. Statistical modelling and inference for multivariate and longitudinal discrete response data. Ph.D. Thesis, University of British Columbia.