Fondazione Eni Enrico Mattei Working Papers
Fondazione Eni Enrico Mattei

Year 2009 *Paper* 285

Modelling Asymmetric Dependence Using Copula Functions: An application to Value-at-Risk in the Energy Sector

Andrea Bastianin
Fondazione Eni Enrico Mattei

Modelling Asymmetric Dependence Using Copula Functions: An application to Value-at-Risk in the Energy Sector

Abstract

n this paper I have used copula functions to forecast the Value-at-Risk (VaR) of an equally weighted portfolio comprising a small cap stock index and a large cap stock index for the oil and gas industry. The following empirical questions have been analyzed: (i) are there nonnormalities in the marginals? (ii) are there nonnormalities in the dependence structure? (iii) is it worth modelling these nonnormalities in risk- management applications? (iv) do complicated models perform better than simple models? As for questions (i) and (ii) I have shown that the data do deviate from the null of normality at the univariate, as well as at the multivariate level. When considering the dependence structure of the data I have found that asymmetries show up in their unconditional distribution, as well as in their unconditional copula. The VaR forecasting exercise has shown that models based on Normal marginals and/or with symmetric dependence structure fail to deliver accurate VaR forecasts. These findings confirm the importance of nonnormalities and asymmetries both in-sample and out-of-sample.

Modelling asymmetric dependence using copula functions: an application to Value-at-Risk in the energy sector

Andrea Bastianin*†

February 26, 2009

Abstract

In this paper I have used copula functions to forecast the Value-at-Risk (VaR) of an equally weighted portfolio comprising a small cap stock index and a large cap stock index for the oil and gas industry. The following empirical questions have been analyzed: (i) are there nonnormalities in the marginals? (ii) are there nonnormalities in the dependence structure? (iii) is it worth modelling these nonnormalities in risk-management applications? (iv) do complicated models perform better than simple models? As for questions (i) and (ii) I have shown that the data do deviate from the null of normality at the univariate, as well as at the multivariate level. When considering the dependence structure of the data I have found that asymmetries show up in their unconditional distribution, as well as in their unconditional copula. The VaR forecasting exercise has shown that models based on Normal marginals and/or with symmetric dependence structure fail to deliver accurate VaR forecasts. These findings confirm the importance of nonnormalities and asymmetries both in-sample and out-of-sample.

Keywords: Copula functions, Forecasting, Value-At-Risk

JEL Classification: C32, C52, C53, G17, Q43

1 Introduction

Risk management is used by firms to translate the risks connected to their business activities into competitive advantages. One of the most widely used risk measure is

*Fondazione Eni Enrico Mattei, Corso Magenta, 63, 20123 Milan, Italy. Phone: +39 02 520 36987. E-mail: andrea.bastianin@feem.it

†The author would like to thank Matteo Manera and Eduardo Rossi for insightful discussion.

Value-at-Risk (VaR), defined as the maximum loss of a portfolio within a given time horizon and at a given level of confidence.

VaR can be estimated either parametrically, or non-parametrically. While in the latter case the realizations of past returns are used to estimate their distribution and thus the VaR, parametric techniques rely on distributional assumptions to forecast the mean and the volatility of a portfolio and hence to calculate its VaR [for a survey of the VaR methodology see Jorion (2007)].

The volatility of a portfolio, measured by its variance, is a function of the variance of the individual assets and their correlations. More generally, the distribution of the returns of a portfolio will be function of the marginal distributions of the individual assets in the portfolio and the dependence structure between those assets. It is therefore clear that ill parametric assumptions will lead to poor VaR forecasts. For instance, VaR models based on the Gaussian distribution, such as the J.P. Morgan's RiskMetrics^{TM} approach, could lead to underestimation of risk in the case of returns with excess kurtosis.

More generally, there are at least two kinds of departures from normality that are especially important in the field of risk management: asymmetries and excess kurtosis. A bunch of studies in the empirical finance literature have shown that there is evidence of two types of asymmetries in the joint distribution of stock returns. First, stocks display excess skewness in their marginal distributions [see Harvey and Siddique (1999, 2000)]. Second, also the dependence between stocks seems to be asymmetric: stocks returns are more highly correlated in bear markets that in bull markets [see Hong, Tu and Zhou, (2007), Longin and Solnik (2001)]. As for excess kurtosis, a fat-tailed univariate random variable is more likely to experience extreme events than what we would expect under the assumption of normality [see Hansen, (1994), Hull and White (1998)]. Similarly, when assuming normality for the dependence structure of returns we neglect tail dependence and hence we underestimate the joint likelihood of extreme events [see Jondeau and Rockinger (2003), Patton (2004, 2006c)].

Summing up, it is clear that both in the case of a single asset and in the case of a portfolio, bad parametric assumptions can lead to poor VaR forecasts [see Hull and White (1998)]. The importance of parametric assumptions and the growing body of empirical evidence against the use of the Normal distribution in financial applications motivates my attempt to use copula theory as a tool for improving VaR forecasts. The assumption of joint normality is very often violated and this leads to the problem of finding more appropriate multivariate specifications; copula functions can be a solution to this problem. In fact, the basic idea of the copula approach is that a joint distribution can be factored into the marginals and a dependence function called copula. The dependence relationship is entirely determined by the copula, while the location, scale and shape parameters (i.e. mean, standard deviation, skewness and kurtosis) are completely determined by the marginals [see Sklar (1959)].

Copula functions have been used because they allow us to take simultaneously into account two characteristics of financial data: nonnormalities at the univariate, as well as at the multivariate level. Nonnormalities in the marginals, such as excess skewness and/or excess kurtosis, can be taken into account with a variety of univariate models; however, when considering multivariate modelling, the task of finding an appropriate specification for the data becomes more challenging, either because estimation can suffer from curse of dimensionality, or because models are not flexible enough. On the contrary, the strength of copula functions relies on their flexibility. In fact, these functions can be used to link marginal distributions and to generate a variety of multivariate specifications.

In this paper I have used copula functions to forecast the VaR of an equally weighted portfolio comprising a small cap stock index and a large cap stock index for the oil and gas industry. Such a portfolio represents a very general investment strategy, namely one based on a low-risk/low-return position, the large cap index, and a high-risk/high return position, the small cap index.

It is worth noting that VaR can be a very useful tool for firms in the energy industry (e.g. airlines wishing to hedge the risks due to jet fuel price volatility, or energy traders), and more generally, when dealing with the problem of energy security. Energy security, defined as the availability of a regular supply of energy at an affordable price, is high on the agenda of governments and policy makers around the world. A threat to a country's energy security can originate either from a physical disruption (e.g. when an energy source is exhausted, or its production is stopped), or from an economic disruption. Economic disruptions are due to erratic fluctuations in the price of energy products, which can be caused either by a threat of a physical disruption of supplies, or by speculative activities. In both cases, the result is a sharp price increase, which directly affects business costs and the purchasing power of private consumers. Therefore VaR, measuring the prospect of an extreme price increase, can be used also as an economic measure of energy security.

This paper answers a set of empirical questions: (i) are there nonnormalities in the marginal distributions? (ii) are there nonnormalities in the dependence structure? (iii) is it worth taking these nonnormalities into account for risk-management? (iv) do complicated models perform better than simple models?

As for questions (i) and (ii), I have shown that the data do deviate from the assumption of normality at the univariate, as well as at the multivariate level. The marginal of the small cap index and that of the large cap index display kurtosis and skewness different from what we would expect in the case of normally distributed time series. The most serious problem is represented by excess kurtosis, on the contrary excess skewness does not seem to be relevant, neither in the estimation stage, nor for risk management purposes.

When considering the dependence structure of the data, I have found that they are more correlated in market downturns than in market upturns. Asymmetries show up in their unconditional distribution, as well as in their unconditional copula, that is after having filtered the returns with appropriate specifications.

As for the importance of nonnormalities for risk management purposes, the VaR forecasting exercise has shown that models based on Normal marginals and/or with symmetric dependence structures fail to deliver accurate VaR forecasts. Among the models that properly forecast the VaR, we have very simple models, such as MA models, copula models with Student's T marginals and asymmetric copula functions, as well as a model with T marginals and Normal, symmetric, copula. The analysis of a set of loss functions shows that the T-asymmetric copula models deliver the best VaR forecasts. These findings confirm the importance of nonnormalities and asymmetries both in-sample and out-of-sample. A common finding in the forecasting literature is that complicated models often perform worst than simple, even misspecified, specifications [see González-Rivera, Lee and Mishra (2004), Swanson and White (1995, 1997)]; interestingly, this does not apply to the data I have analyzed.

The rest of the paper is organized as follows: section 2 introduces the theory of copulas; section 3 illustrates how to use copulas to forecast VaR; section 4 is the empirical part of the paper; section 5 concludes.

2 Multivariate models and copulas

A copula function represents a statistical tool that allows to study the dependence between two, or more random variables. The word "copula" comes from the Latin for "link": a collection of marginal distributions can be "linked" together via a copula to form a multivariate distribution. The theory of copulas dates back to Sklar (1959), who showed how to decompose a joint distribution into a set of univariate marginal distributions and a copula which describes the dependence between variables after taking out the effects of the marginals.

Early applications of copulas in statistics focused on random vectors of independently and identically distributed (i.i.d.) data; nowadays, it is common to use them in the context of time series analysis. Following Patton (2006b) we can consider two main areas of applications of copulas to time series modelling. The first is the application to multivariate time series, where the focus is the modelling of the joint distribution of some random vector $\mathbf{X}_t = [X_{1t}, X_{2t}, ..., X_{nt}]'$, conditional on a given information set Ω_{t-1} (i.e. usually it contains past observations on the variates, say $\Omega_{t-1} \equiv \mathbf{X}_{t-j}$, for $j \geq 1$). The second field of application of copulas is the modelling of the joint distribution of a sequence of observations of a univariate time series $\mathbf{X}_i = [X_{it}, X_{it+1}, ..., X_{iT}]'$. In this paper I will focus on the use of copulas for multivariate time series modelling; more details about the application of copulas in time series modelling and in risk management can be found in Dias (2004), Embrechts *et al.* (2001, 2002), and Patton (2006b).

The discussion of the theory of copulas and its application to multivariate time series modelling requires some technical concepts; these technicalities, the main definitions and the properties of copulas will be discussed in the next section. Next, I will go into the details concerning estimation and inference techniques for conditional copulas.

Although copulas are designed to deal with general multivariate distributions, in what follows I restrict my attention to the bivariate case. As for the notation, I will use the following conventions: X and Y denotes two random variables, W is a conditioning variable or vector of variables, F_{XYW} is the joint distribution of (X, Y, W), $F_{XY|W}$ is the conditional distribution of (X, Y) given W and the conditional marginal distributions of $X|W$ and $Y|W$ are denoted $F_{X|W}$ and $F_{Y|W}$, respectively (for unconditional distribution the notation is similar, in this case I simply ignore the conditioning variable). Furthermore, I will adopt the usual convention of denoting cumulative distribution functions (c.d.f.) and random variables using upper case letters, while lower case letters are used for probability density functions (p.d.f.) and realizations of random variables. Through the paper I will assume that F_{XYW} is sufficiently smooth for all required derivatives to exist, and that $F_{XY|W}$, $F_{X|W}$ and $F_{Y|W}$, are continuos.

2.1 Introducing copulas

A copula function can be defined as a multivariate distribution function with uniform $U(0,1)$ univariate marginal distributions. Sklar (1959) showed that copulas are useful not only as a tool for isolating the dependence relationships from the marginal behavior in a multivariate distribution, but also because we can use them to write the mapping from the individual distribution functions to the joint distribution function. This result can be stated as follows:

Theorem 1 (Sklar's theorem) *Given a pair of distribution functions, F_X, F_Y, and a bivariate copula C, the function defined by:*

$$F_{XY}(x,y) = C(F_X(x), F_Y(y)), \qquad \forall (x,y) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}} \tag{1}$$

is a bivariate distribution function with univariate margins F_X and F_Y. $\overline{\mathbb{R}}$ denotes the extended real line, that is $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$.

Equivalently, we can say that given any collection of marginals $(F_1, F_2, \ldots, F_n)$ and any copula C, we can use Sklar's theorem, as stated in Equation (1), to recover the joint distribution from the marginal distributions. This gives a great advantage in terms of flexibility which is very useful in many branches of econometrics. For instance, in portfolio modelling we can use different marginals for each asset and a copula to link them together; given the widespread evidence of nonnormalities in financial data, this flexibility is of great importance also for risk management tools, such as Value-at-Risk [for an application of copulas to VaR see Fantazzini (2004)]. Moreover, what makes copulas really useful in applications involving the joint modelling of two or more variates, is that the linear correlation and the marginal distributions determine a joint distribution only if the variables of interest are elliptically distributed. When this is not the case, the copula will take the place of the correlation.

To fully understand copulas, we need to introduce the concept of "probability-integral transformation", (PIT). The PIT is a method for generating n values of a non-uniform random variable X which has continuos c.d.f. F_X. The PIT can be introduced as follows[1]:

Definition 1 (Probability integral transformation (a)) *The PIT is the mapping $T : \mathbb{R}^d \to [0,1]^d$, $(x_1, x_2, ..., x_d) \longmapsto (F_1(x_1), F_2(x_1), ..., F_d(x_d))$.*

The PIT exploits the fact that a random variable X with c.d.f. F_X can be transformed into a variable with uniform distribution over the interval $[0,1]$, that is $U = F_X(X)$. Conversely, if U is uniformly distributed over the interval $[0,1]$, then $X = F_X^{-1}(U)$ has c.d.f. F_X. Hence, to generate a value, say x, of the random variable X having continuos c.d.f. F_X, we can generate a value, say u, of the random variable U which is uniformly distributed over $[0,1]$. The value x is then obtained as $x = F_X^{-1}(u)$.

[1] The PIT is due to Rosenblatt (1952). A very intuitive proof is given by Schuster (1976). For its use in goodness-of-fit tests, see for instance Breymann *et al.* (2003), Dias (2004). For the extension of the PIT theory to time series analysis see Diebold *et al.* (1998).

Now that I have introduced the concept of PIT, we are ready to define the density function equivalent of (1). Provided that F_X and F_Y are differentiable and that F_{XY} and C are twice differentiable we have:

$$\begin{aligned} f_{XY}(x,y) &\equiv \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) \\ &= \frac{\partial F_X(x,y)}{\partial x} \frac{\partial F_Y(x,y)}{\partial y} \frac{\partial^2 C(F_X(x), F_Y(y))}{\partial u \partial v} \\ &= f_X(x) f_Y(y) c(F_X(x), F_Y(y)). \end{aligned} \tag{2}$$

where $c(\cdot) \equiv \partial^2 C(F_X(x), F_Y(y)) / \partial u \partial v$ denotes the "copula density", $U \equiv F_X(x)$, $V \equiv F_Y(y)$ are the PIT and $(u,v) \in [0,1]^2$. With this result we can rewrite the Sklar's theorem in terms of density functions:

Theorem 2 (Sklar's theorem (continued)) *Given a pair of density functions, f_X, f_Y, and a bivariate copula density c, the function defined by:*

$$f_{XY}(x,y) = f_X(x) f_Y(y) c(F_X(x), F_Y(y)), \qquad \forall (x,y) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}} \tag{3}$$

is a bivariate density function with univariate margins f_X and f_Y. $\overline{\mathbb{R}}$ denotes the extended real line, that is $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$.

Sklar's theorem written as in (3) is very useful for maximum likelihood estimation, indeed we can state that the joint log-likelihood of (X,Y) can be written as the sum of the univariate marginal likelihoods and the copula likelihoods; additional details will be given below.

Let us now move to the question of conditional copula modelling. Following Patton (2006c), I assume that the dimension of the conditioning variable, W, is one. Hence we can derive the conditional bivariate distribution of $(X,Y)|W$ from the unconditional joint distribution of (X,Y,W) as follows:

$$F_{XY|W}(x,y|w) = f_w(w)^{-1} \frac{\partial F_{XYW}(x,y,w)}{\partial w}, \quad \text{for } w \in \mathcal{W} \tag{4}$$

where f_w is the unconditional density of w and $\mathcal{W}$ is the support of W. However, notice that this type of derivation is not feasible for the conditional copula; in other words, we cannot derive it from the unconditional copula, as we did for the bivariate distribution, because we need the same information set for all the marginal distributions and the copula. For the moment, let us just introduce the notion of conditional copula, without taking the common information problem into consideration. Accordingly, a conditional copula can be defined as follows:

Definition 2 (Conditional copula) *The conditional copula of* $(X,Y)\,|W=w$*, where* $X|\,(W=w)\sim F_{X|W}\,(\cdot|w)$ *and* $Y|\,(W=w)\sim F_{Y|W}\,(\cdot|w)$*, is the conditional distribution function of* $U\equiv F_{X|W}\,(X|w)$ *and* $V\equiv F_{Y|W}\,(Y|w)$ *given* $W=w$*.*

Where U and V are the PIT of X and Y given W; as we have seen, these variates will have Uniform $(0,1)$ distribution, regardless of the original distributions of X and Y. Hence, the conditional copula can be defined as the conditional joint distribution of two conditional Uniform $(0,1)$ variates.

Once again, notice that in the context of conditional copulas the definition of the conditioning set is essential for the validity of the properties listed above. The extension of Sklar's theorem to conditional distributions provided by Patton (2006c) is as follows:

Theorem 3 (Sklar's theorem for conditional copulas) *Let* $F_{X|W}\,(\cdot|w)$ *be the conditional distribution of* $X|\,(W=w)$*,* $F_{Y|W}\,(\cdot|w)$ *be the conditional distribution of* $Y|\,(W=w)$*,* $F_{X,Y|W}\,(\cdot|w)$ *be the joint conditional distribution of* $X,Y|\,(W=w)$*, and* $\mathcal{W}$ *be the support of* W*. Assume that* $F_{X|W}$ *and* $F_{Y|W}$ *are continuous in* x *and* y *for all* $w\in\mathcal{W}$*. Then there exists a unique copula* $C\,(\cdot|w)$ *such that:*

$$\begin{aligned} F_{XY|W}\,(x,y|w) &= C\left(F_{X|W}\,(x|w)\,,F_{Y|W}\,(y|w)\,|w\right), \\ \forall\,(x,y) &\in \overline{\mathbb{R}}\times\overline{\mathbb{R}} \quad \textit{and each} \quad w\in\mathcal{W} \end{aligned} \tag{5}$$

Conversely, if we let $F_{X|W}\,(\cdot|w)$ *be the conditional distribution of* $X|\,(W=w)$*,* $F_{Y|W}\,(\cdot|w)$ *be the conditional distribution of* $Y|\,(W=w)$*, and* $C\,(\cdot|w)$ *be a conditional copula, then the function* $F_{X,Y|W}\,(\cdot|w)$ *is a conditional bivariate distribution with conditional marginal distributions* $F_{X|W}\,(\cdot|w)$ *and* $F_{Y|W}\,(\cdot|w)$*.*

In the context of multivariate time series analysis the converse of Sklar's theorem is very useful, indeed it implies that we can link together any two univariate distributions with any copula and have a valid bivariate distribution. We can think of this flexibility as expanding the set of parametric multivariate distributions we can use in econometric modelling.

As anticipated above, in order to extend Sklar's theorem to conditional copulas the choice of the conditioning set is a delicate matter, indeed this must be the same for both the univariate marginals and the copula. An example of different conditioning sets across variables is represented by situations in which each variable depends on its own first lag, but not on the lags of other variables. Failure to use the same conditioning information set for $F_{X|W}$, $F_{Y|W}$ and C, will in general imply that $F_{X,Y|W}$ is not a proper joint distribution function [see Patton (2006, p. 534)]. The only case in which $F_{X,Y|W}$ is a proper joint distribution function, even thought the conditioning variables are not the same for all marginal distributions, is when some variables affect the conditional distribution of one variable but not the others[2].

To conclude this section, let us see how to use Sklar's theorem, as expressed in Equation (5), and the relation between the distribution and the density function to extract the bivariate conditional copula density $c(\cdot|w)$, associated to the conditional copula function $C(\cdot|w)$:

$$\begin{aligned} f_{XY|W}(x,y|w) &\equiv \frac{\partial^2}{\partial x \partial y} F_{XY|W}(x,y|w), \qquad \forall (x,y,w) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}} \times \mathcal{W} \\ &= \frac{\partial F_{X|W}(x,y|w)}{\partial x} \frac{\partial F_{Y|W}(x,y|w)}{\partial y} \frac{\partial^2 C\left(F_{X|W}(x|w), F_{Y|W}(y|w)|w\right)}{\partial u \partial v} \\ &= f_{X|W}(x|w) f_{Y|W}(y|w) c\left(F_{X|W}(x|w), F_{Y|W}(y|w)|w\right) \end{aligned} \tag{6}$$

where $U \equiv F_{X|W}(x|w)$ and $V \equiv F_{Y|W}(y|w)$.

2.2 Copula modelling

The choice of the copula used to link together the marginals of two variates should be guided by the nature of the data the analyst is going to consider. Indeed, each copula implies a different type of dependence between the variables. Patton (2006c, 541) points out that many of the copulas available in the statistical literature are designed for variables that take on joint extreme values in only one direction. While this kind of functional forms are adequate for some economic variables, for others it is wise to be flexible in the choice of the copula. As for equity returns, we can choose the copula on the basis of the empirical evidence suggesting that "stocks tend to crash together, but not to boom together". In this case we should select a copula that implies greater dependence for joint negative events than for joint positive events. However, for many economic variables it is not easy to select the "right" copula; this is due either to the lack of empirical evidence, or to the fact that we do not have a theoretical model which suggests the sign of the joint dependence for the variable we want to study. In these

[2] For instance, Patton (2006c) reports that, conditional on lags of the DM-USD exchange-rate, lags of the Yen-USD exchange-rate do not impact on the distribution of the DM-USD exchange-rate. Similarly, lags of the DM-USD exchange-rate do not affect the Yen-USD exchange rate, conditional on lags of the Yen-USD exchange rate.

situations the best thing to do is to consider various functional forms for the copula.

The first copula I consider is the Gaussian or Normal one. The Normal copula is the copula function associated to with the bivariate Normal distribution and represents the dependence structure associated to such a distribution. Let us assume that the random vector $(X,Y)|W$ is bivariate Normal, or equivalently that its margins $F_{X|W}$ and $F_{Y|W}$ are Normal and recall that $U \equiv F_{X|W}(x|w)$ and $V \equiv F_{Y|W}(y|w)$. The Gaussian copula can be written as:

$$C_N(u,v|\rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left[\frac{-\left(r^2 - 2\rho rs + s^2\right)}{2\left(1-\rho^2\right)}\right] drds \quad \rho \in (-1,1). \tag{7}$$

where $\Phi^{-1}(\cdot)$ is the inverse c.d.f. of a Normal $(0,1)$ variate. The Gaussian copula depends on a single parameter: the coefficient of linear correlation ρ.

Similarly, the Student's T copula is the dependence structure assumed whenever the bivariate T distribution is used. The T copula depends on the correlation coefficient ρ, and on υ, the shape parameter/degrees of freedom of the distribution. Notice that in analogy to what happens for the c.d.f.s, the Gaussian copula can be thought of as the limiting case of the T copula as $\upsilon \to \infty$. [for more details on the T copula see Demarta and McNeil (2004)].

Both the Gaussian and the T copula depend on the correlation coefficient, but the latter has a different behavior for what concerns tail dependence. In multivariate settings, fat taildness can be referred to both the marginal univariate distributions, or to the joint probability of large market movements. The concept we use to deal with the latter problem is called tail dependence and it can be formally defined as follows:

Definition 3 (Tail dependence) *Let U and V be two random variables uniformly distributed on* $(0,1)$. *If the limit*

$$\begin{aligned}
\tau^L &\equiv \lim_{\varepsilon \to 0^+} \Pr(U \leq \varepsilon | V \leq \varepsilon) \\
&= \lim_{\varepsilon \to 0^+} \frac{\Pr(U \leq \varepsilon, V \leq \varepsilon)}{\Pr(V \leq \varepsilon)} \\
&= \lim_{\varepsilon \to 0^+} \frac{C(\varepsilon,\varepsilon)}{\varepsilon}
\end{aligned}$$

exists, then the copula C exhibits lower tail dependence if $\tau^L \in (0,1]$ and no lower tail *dependence if* $\tau^L = 0$. *Similarly, if*

$$\begin{aligned}\tau^U &\equiv \lim_{\varepsilon \to 1^-} \Pr\left((1-U) > \varepsilon | (1-V) > \varepsilon\right) \\ &= \lim_{\varepsilon \to 1^-} \frac{1 - 2\varepsilon + C(\varepsilon, \varepsilon)}{1 - \varepsilon}\end{aligned}$$

exists, then the copula C *exhibits upper tail dependence if* $\tau^U \in (0,1]$ *and no upper tail dependence if* $\tau^U = 0$.

Notice that τ^U and τ^L are asymptotic measures of dependence focused on bivariate distributions; indeed, we say that two variates are asymptotically dependent in the lower (upper) tail if $\tau^L \in (0,1]$ ($\tau^U \in (0,1]$). Similarly, whenever $\tau^L = 0$ ($\tau^U = 0$) two variables are said to be asymptotically independent in the lower (upper) tail. More informally, we can state that tail dependence captures the behavior of two variates during extreme events, thus it measures the probability that a stock, say ENI, has an extremely low/high return given that another stock, say BP, experiences an extremely low/high return.

It can be shown that the Normal copula has $\tau^L = \tau^U = 0$, meaning that the variables are independent in the tails of the distribution [see Embrechts et al. (2002)]. The tail dependence of two bivariate Student's T variates is determined by the correlation coefficient ρ and the shape parameter, υ. Being a symmetric copula, the dependence between extremely low returns and extremely high returns is the same.

The copulas we have discussed so far belong to the family of elliptical copulas; this definition stems from the fact that they have been derived from elliptical multivariate distributions. A drawback of elliptical copulas is that they cannot account for the fact that in many financial applications it is reasonable to assume that there is a stronger dependence across extremely low returns, than across extremely high returns. For these reasons in the empirical part of the paper I will carry out the analysis by using the Normal copula along with the following copula functions: Clayton copula, symmetrized Joe-Clayton (SJC) copula, Plackett copula and rotated Gumbel copula. Contour plots of some of these copulas, are shown in figure 1. As we can see from figure 1, by linking bivariate Normal (0,1) densities with different copulas, we can generate isoprobability contours of very different shapes. These plots clearly illustrate that different copulas can account for basically any kind of dependence structure. The upper left panel displays the Normal copula with its familiar elliptical contours. In the upper right panel we can see the isoprobability contour of the Student's T copula: we can notice that, although symmetric and elliptically shaped, if compared with the Normal copula, the T copula has a quite different behavior in the first ("positive-positive") and in the third ("negative-

negative") quadrant, where the isoprobability contours are more tightly clustered around the diagonal, suggesting that it allows for (symmetric) non-zero tail dependence. Other copulas that generate symmetric dependence are the SJC and Plackett copulas shown in the lowest panels. Interestingly, the SJC copula, which depends upon two parameters, τ^U and τ^L (that, as we have seen, are measures of tail dependence), is a modification of the Joe-Clayton copula that can generate both symmetric and asymmetric dependence (e.g. it is symmetric for $\tau^U = \tau^L$ and it becomes asymmetric whenever $\tau^U \neq \tau^L$).

The remaining four copulas can generate asymmetric dependence. In particular the rotated Gumbel copula and the Clayton copulas can account for returns more highly correlated in bear markets than in bull markets, which is the case for many financial time-series. This type of behavior has been reported for instance by Carvalho and Amonlirdviman (2008) and Longin and Solnik (2001).

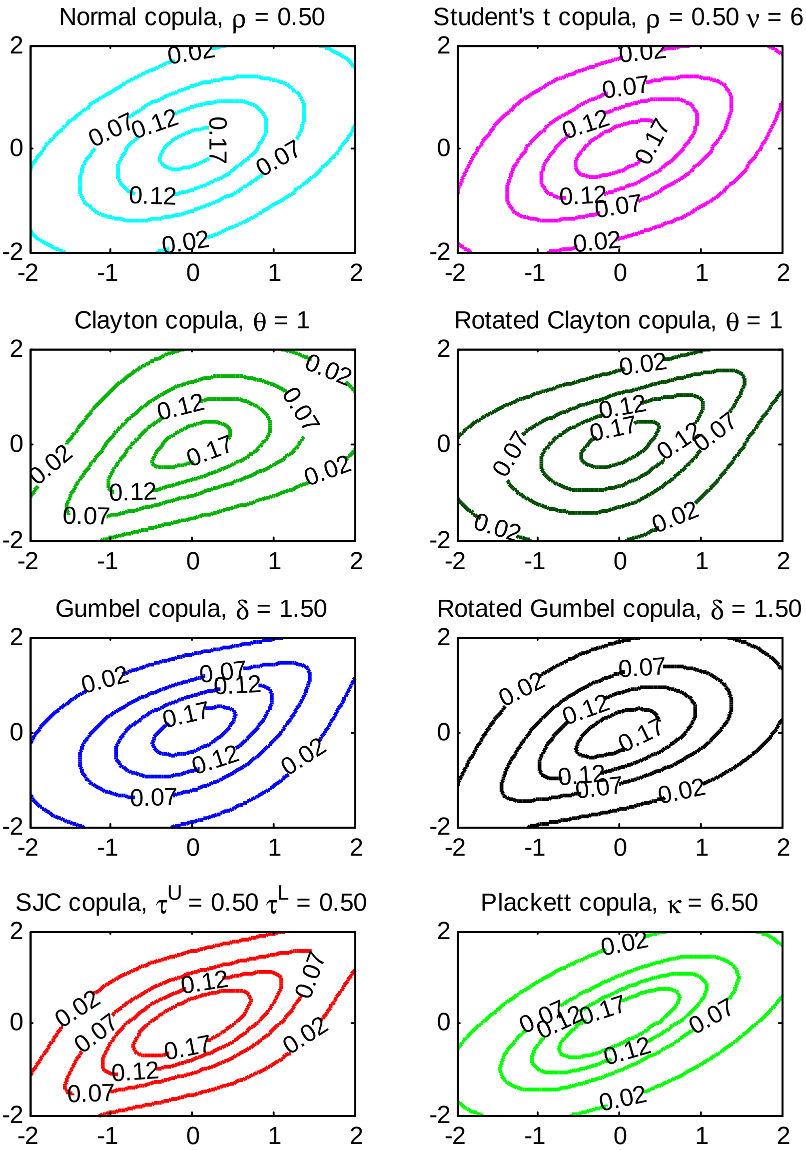
2.3 Multi-stage estimation of copula functions

The methodology to estimate copula functions, known as the Inference Functions for Margins (IMF) method [for details, see Dias (2004)], has been extended to time series analysis by Patton (2006*a*). The author shows that the existing two-stage maximum likelihood estimation framework [see Newey and McFadden (1997) and White (1982)], can be applied to estimate parametric multivariate density models involving variables with histories of different length. Patton (2006*a*) focuses on models with an unknown parameter vector that may be partitioned into elements relating only to the marginals and elements only relating to the copula. This partition is also possible in many common multivariate models, such as vector autoregressions and conditional correlation multivariate GARCH models [see Bollerslev (1990) and Engle *et al.* (2001)].

Let us assume that the conditional distribution of $(X_t, Y_t)|W_{t-1}$ is known and that it is parametrized as $H_t(x, y|w, \theta_0) = C(F_t(x|w, \varphi_0), G_t(y|w, \gamma_0)|w, \kappa_0)$, where $\theta_0 \equiv [\varphi_0', \gamma_0', \kappa_0']'$ must be estimated. In terms of the notation used until now we have that, $H_t \equiv F_{XY|W}$, $F_t \equiv F_{X|W}$, $G_t \equiv F_{Y|W}$ and similarly for the densities. Notice that, when feasible, I suppress the dependence on the conditioning variable W and the subscript denoting time in order to avoid cumbersome notation. From Sklar's theorem we know that the conditional density of $(X_t, Y_t)|W_{t-1}$ can be written as [see Equation (6)]:

$$h_t(\theta_0) = f_t(\varphi_0)\, g_t(\gamma_0)\, c(F_t(\varphi_0), G_t(\gamma_0)|\kappa_0) \tag{8}$$

Figure 1: Contour plots for various copula functions all with normal marginals.



and hence, this implies that the likelihood of $(X_t, Y_t)\,|W_{t-1}$ is given by:

$$\begin{aligned}
\mathcal{L}_{XY}(\theta_0) &\equiv T^{-1}\sum_{t=1}^{T}\log h_t(\theta_0) \\
&= T^{-1}\sum_{t=1}^{T}\log f_t(\varphi_0) + T^{-1}\sum_{t=1}^{T}\log g_t(\gamma_0) + T^{-1}\sum_{t=1}^{T}\log c_t\left(F_t(\varphi_0), G_t(\gamma_0)\,|\kappa_0\right) \\
&\equiv \mathcal{L}_X(\varphi_0) + \mathcal{L}_Y(\gamma_0) + \mathcal{L}_C(\kappa_0) \qquad (9)
\end{aligned}$$

where $\varphi_0 \in int(\Phi) \subseteq \mathbb{R}^p$, $\gamma_0 \in int(\Gamma) \subseteq \mathbb{R}^q$, $\kappa_0 \in int(\mathcal{K}) \subseteq \mathbb{R}^r$ and $\theta_0 \equiv [\varphi_0', \gamma_0', \kappa_0']' \in int(\Theta) \equiv int(\Phi) \times int(\Gamma) \times int(\mathcal{K}) \subseteq \mathbb{R}^{p+q+r} \equiv \mathbb{R}^s$, where $int(\Im)$ is the interior set of $\Im$.

Let the multi-stage maximum likelihood estimator (MSMLE) of θ_0 be denoted as $\hat{\theta}_T$. It is obtained by dividing the estimation process into the following two steps:

1. The parameters φ_0 and γ_0 of the marginal distributions $F_t(x|w, \varphi_0)$ and $G_t(y|w, \gamma_0)$ are estimated as:

$$\hat{\varphi}_T = \arg\max_{\varphi\in\Phi} T^{-1}\sum_{t=1}^{T}\log f_t(x_t|\varphi), \qquad (10)$$

$$\hat{\gamma}_T = \arg\max_{\gamma\in\Gamma} T^{-1}\sum_{t=1}^{T}\log g_t(y_t|\gamma) \qquad (11)$$

2. Given the results in step 1, the copula parameters κ_0 are estimated as:

$$\hat{\kappa}_T = \arg\max_{\kappa\in\mathcal{K}} T^{-1}\sum_{t=1}^{T}\log c\left(F_t(x_t|\hat{\varphi}_T), G_t(y_t|\hat{\gamma}_T)\,|\kappa\right) \qquad (12)$$

Asymptotic results for the MSMLE are obtained as an extension of the two-stage MLE framework discussed for instance in Newey and McFadden (1994) and in White (1982). In particular, it can be shown that under standard regularity assumptions the MSMLE is consistent and that its limiting distribution is given by [Patton (2006a, 166-170)]:

$$B_T^{0-1/2} A_T^0 \sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_s) \qquad (13)$$

where $\mathbf{I}_s$ is an $s \times s$ identity matrix, and:

$$S_T^0 = T^{-1} \begin{bmatrix} \sum_{i=1}^T \nabla_\varphi \log f_t^0 \\ \sum_{i=1}^T \nabla_\gamma \log g_t^0 \\ \sum_{i=1}^T \nabla_\kappa \log c_t^0 \end{bmatrix} \tag{14}$$

$$Hess_T^0 = T^{-1} \begin{bmatrix} \sum_{i=1}^T \nabla_{\varphi\varphi} \log f_t^0 & 0 & 0 \\ 0 & \sum_{i=1}^T \nabla_{\gamma\gamma} \log g_t^0 & 0 \\ \sum_{i=1}^T \nabla_{\varphi\kappa} \log c_t^0 & \sum_{i=1}^T \nabla_{\varphi\kappa} \log c_t^0 & \sum_{i=1}^T \nabla_{\kappa\kappa} \log c_t^0 \end{bmatrix} \tag{15}$$

$$OPG_T^0 = T^{-1} \begin{bmatrix} \sum_{i=1}^T s_{\varphi t}^0 s_{\varphi t}^{0\prime} & \sum_{i=1}^T s_{\varphi t}^0 s_{\gamma t}^{0\prime} & \sum_{i=1}^T s_{\varphi t}^0 s_{\kappa t}^{0\prime} \\ \sum_{i=1}^T s_{\gamma t}^0 s_{\varphi t}^{0\prime} & \sum_{i=1}^T s_{\gamma t}^0 s_{\gamma t}^{0\prime} & \sum_{i=1}^T s_{\gamma t}^0 s_{\kappa t}^{0\prime} \\ \sum_{i=1}^T s_{\kappa t}^0 s_{\varphi t}^{0\prime} & \sum_{i=1}^T s_{\kappa t}^0 s_{\gamma t}^{0\prime} & \sum_{i=1}^T s_{\kappa t}^0 s_{\kappa t}^{0\prime} \end{bmatrix} \tag{16}$$

$$A_T^0 = E\left(Hess_T^0\right), \; B_T^0 = E\left(OPG_T^0\right) \tag{17}$$

where $OPG_T^0 = S_T^0 S_T^{0\prime}$, $s_{\varphi t}^0 \equiv \nabla_\varphi \log f_t^0$, $s_{\gamma t}^0 \equiv \nabla_\gamma \log g_t^0$, $s_{\kappa t}^0 \equiv \nabla_\kappa \log c_t^0$, $f_0^t \equiv f_t\left(x_t|\varphi_0\right)$, $g_0^t \equiv g_t\left(y_t|\varphi_0\right)$, $c_0^t \equiv c_t\left(F_t\left(\varphi_0\right), G_t\left(\gamma_0\right)|\kappa_0\right)$ (when a quantity has a zero in the subscript, or in the superscript it means that this quantity is evaluated at the true vector of parameters θ_0). Equation (14) is the vector of first derivatives, or score vector, Equation (15) is the matrix of second derivatives, or Hessian matrix and Equation (16) is the Outer Product of Gradients (OPG).

Following White (1982), we say that if $\widetilde{V}_T^{-1/2}\sqrt{T}\left(\widetilde{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}\right)$, then the asymptotic covariance matrix of the estimator $\widetilde{\theta}_T$ is $\widetilde{V}_T$, or that $avar\left(\widetilde{\theta}_T\right) = \widetilde{V}_T$. For the MSMLE we have $B_T^{0-1/2} A_T^0 \sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}_s\right)$, thus the asymptotic covariance matrix is $A_T^{0-1} B_T^0 A_T^{0-1\prime}$; equivalently we can write[3]: $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, A_T^{0-1} B_T^0 A_T^{0-1\prime}\right)$. Under standard regularity conditions, the asymptotic covariance matrix can be estimated using the Hessian and the OPG evaluated at the MSMLE, $\hat{\theta}_T$ [see White (1982)].

In other words $V_T\left(\hat{\theta}_T\right)$ is estimated as, $T^{-1} Hess_T\left(\hat{\theta}_T\right)^{-1} OPG_T\left(\hat{\theta}_T\right)^{-1} Hess_T\left(\hat{\theta}_T\right)^{-1}$, which is the so-called "sandwich estimator" of the covariance matrix.

[3]Let $V^{-1/2} = B_T^{0-1/2} A_T^0$, then it follows that:
$C = \left(V^{-1/2}\right)^{-1} = \left(B_T^{0-1/2} A_T^0\right)^{-1} = \left(A_T^0\right)^{-1}\left(\left(B_T^0\right)^{-1/2}\right)^{-1} = A_T^{0-1}\left(B_T^0\right)^{1/2}$,
$C' = \left(A_T^{0-1}\left(B_T^0\right)^{1/2}\right)' = \left(\left(B_T^0\right)^{1/2}\right)'\left(A_T^{0-1}\right)' = B_T^{01/2}\left(A_T^{0-1}\right)'$ (B_T^0 is symmetric)
and
$V = CC' = A_T^{0-1}\left(B_T^0\right)^{1/2} B_T^{01/2}\left(A_T^{0-1}\right)' = A_T^{0-1} B_T^0 A_T^{0-1\prime}$.

2.4 Density functions for the marginals

A probability density function (p.d.f.) is characterized by three parameters: the location, the scale and the shape. The *location* parameter (e.g. mean, median, or mode) specifies the positions on the x-axis of the range of values. For symmetric distributions, the location parameter represents the midpoint and hence, as it shifts, the p.d.f. shifts retaining its shape. The *scale* parameter (e.g. variance) measures the spread of the density and determines the unit of measurement of the values in the range of the p.d.f.. The p.d.f. compresses/expands leaving its shape unchanged, as the scale changes. The *shape* parameter (e.g. skewness and kurtosis) determines how the variation is distributed about the location and the form of the distribution within the general family of distributions to which it belongs. As the shape parameter changes, the properties of the p.d.f. change.

As for time series analysis, we can state that, in general, a p.d.f. should have the following desirable properties:

(i) it must generalize to the standard Normal distribution (e.g. the T distribution converges to the Normal, as its degrees of freedoms tend to infinity);

(ii) it must be sufficiently flexible so as to generate a range of shapes which we think might be relevant in a particular application (e.g. in financial applications, it is desirable that the shape parameter explains the skewness and kurtosis that may be encountered in the data);

(iii) it must be sufficiently parsimonious that the shape parameters can be modeled with time series techniques whenever required;

(iv) it must be available in closed-form in order to facilitate (Quasi-ML) estimation.

The last point is very important, especially in applied work. Indeed in the statistical literature there exist many flexible and parsimonious parametric distribution, but only few of them have closed-form density functions. When the density is unavailable estimation can be carried-out via method of moments, but as Hansen (1994) points out this might involve severe inferential difficulties, especially for IGARCH models. In other words, we want flexible low-dimensional densities with closed-form in order to use QML estimation, which is preferred because of its simplicity and its very well-grounded inference theory.

Let us introduce the notation: the observed sample is $(y_t, w_t : t = 1, ..., T)$ where w_t includes all the past values of y_t. The density of y_t is written as: $f\left(y|\alpha_t\left(w_t, \theta\right)\right)$, where θ is a finite-dimensional vector of parameters and $\alpha_t = \alpha(w_t, \theta)$ is a time-varying parameter. Now assume that is possible to parametrize $f(y|\alpha)$, so that we can partition the time-varying parameter as $\alpha_t = \left(\mu_t, \sigma_t^2, \upsilon_t\right)$, where $\mu_t = \mu(\theta, w_t) = E(y_t|w_t)$ is the conditional mean, $\sigma_t^2 = \sigma^2(\theta, w_t) = E\left[(y_t - \mu_t)^2 |w_t\right]$ is the conditional variance and $\upsilon_t = \upsilon(\theta, w_t)$ is the shape parameter of the distribution. Finally, let us define the normalized variable $z_t \equiv [(y_t - \mu(\theta, w_t)) / \sigma(\theta, w_t)]$ which has density $g(z|\upsilon_t)$. Notice that the densities of y and z are related by $f\left(y_t|\mu_t, \sigma_t^2, \upsilon_t\right) = g(z_t|\upsilon_t) / \sigma_t$.

The first distribution I consider is the standardized Normal p.d.f.. As already highlighted, two of the most common deviations from normality are fat-tails and asymmetry (recall that the implied kurtosis and skewness of the Normal distribution are three and zero, respectively). I use the Student's T density to capture (excess) kurtosis and the skewed Student's T density to capture both skewness and kurtosis. The density of the T-distribution (normalized to have unit variance) depends on the parameter υ which represents the degrees of freedom of the distribution and captures leptokurtosis. It is important to note that the kurtosis is both a measure of the peakedness and the fat taildness of the distribution. The Student's T density allows for variations in the location, scale, and tail thickness. The implied kurtosis of the Student's T distribution is $k = 6/(\upsilon - 4)$ for all $\upsilon > 4$. Notice that the the Student's T is leptokurtic when $4 < \upsilon \leq 25$ and it converges to the Normal as $\upsilon \to \infty$.

A desirable extension with respect to both the Normal and the Student's T density, is to allow for skewness; this can be accomplished by considering the skewed Student's T distribution of Hansen (1994), who underlines the importance of having a density function that can be easily parametrized so that the standardized residuals of a conditional location-scale model have zero mean and unit variance (i.e. otherwise, it might be difficult to separate the fluctuations in the mean and variance from those in the shape of the conditional density). The functional form of the skewed Student's T density is given by:

$$T^{skew}(z|\upsilon, \lambda) = \begin{cases} bc\left[1 + \frac{1}{\upsilon-2}\left(\frac{bz+a}{1-\lambda}\right)\right]^{-(\upsilon+1)/2} & \text{if } z < -a/b \\ bc\left[1 + \frac{1}{\upsilon-2}\left(\frac{bz+a}{1+\lambda}\right)\right]^{-(\upsilon+1)/2} & \text{if } z \geq -a/b \end{cases} \quad (18)$$

where $2 < \upsilon \leq \infty$ and $-1 < \lambda < 1$. The constants a, b and c are defined as:

$$a = 4\lambda c\left(\frac{\upsilon - 2}{\upsilon - 1}\right), \quad (19)$$

$$b = \sqrt{1 + 3\lambda^2 - a^2}, \quad (20)$$

and

$$c = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\pi\left(\upsilon-2\right)}\Gamma\left(\frac{\upsilon}{2}\right)}. \tag{21}$$

Notice that the skewed Student's T distribution encompasses both the (symmetric) Student's T and the Normal distribution; indeed, we get the former when $\lambda = 0$, while the latter is obtained for $\lambda = 0$ and $\upsilon \to \infty$. Like the Student's T distribution, it is well defined only for $\upsilon > 2$, the skewness exists for $\upsilon > 3$ and the kurtosis exists only if $\upsilon > 4$. The parameter λ controls the skewness of the density, which is continuos and has a single mode at $-a/b$. If $\lambda > 0$, the mode of the density is to the left of zero and the variable is skewed to the right, vice-versa for $\lambda < 0$.

3 VaR Forecasting with Copulas

Value-at-Risk measures the worst expected loss under normal market conditions over a specific time interval, at a given confidence level; in other words, VaR estimates market risk, that is the uncertainty of future earnings due to the changes in market conditions. The time period and the confidence level (i.e. the quantile) are two very important parameters that should be chosen in a way appropriate to the overall goal of risk measurement. In this paper I use a 95 percent confidence level and a one day time period.

There are two factors that have contributed to increase the popularity of VaR as a risk management tool. First, its simplicity: VaR reduces the market risk associated with any portfolio to a single number, the loss associated to a given probability. Second, the Basel Capital Accord sets capital requirements of banks as a function of the VaR. In 1988 central bankers from the Group of Ten (G10) countries undertook what is known as the Basel Accord. This agreement, which is now adopted by more than 100 countries, sets the minimum capital requirements that banks must meet to guard against credit and market risks. The market risk capital requirement is a function of the forecasted VaR thresholds. Assuming that returns can be written as $r_t = E\left(r_t|\Omega_{t-1}\right) + \varepsilon_t$ and that ε_t has variance h_t, the VaR threshold is defined as:

$$VaR_t = E\left(r_t|\Omega_{t-1}\right) - q\sqrt{h_t} \tag{22}$$

in which q is the critical value from the distribution of the unpredictable component of returns, ε_t.

For an equally weighted portfolio of two assets, the VaR can be written as:

$$VaR_t = \sqrt{\sum_{i=1}^{2} (VaR_{i,t})^2 + 2\rho_{12,t} VaR_{1,t} VaR_{2,t}} \tag{23}$$

One of the most well known VaR methodologies is J.P. Morgan's RiskMetrics$^{\text{TM}}$. This method assumes that the continuously compounded daily returns of a portfolio follow a conditional Normal distribution, that is $r_t|\Omega_{t-1} \sim N(\mu_t, h_t)$. In addition RiskMetrics$^{\text{TM}}$ assumes that the mean, μ_t, and the variance, σ_t^2, evolve according to:

$$\mu_t = 0, \qquad h_t = \lambda h_{t-1} + (1-\lambda) r_{t-1}^2, \qquad \text{with } \lambda = .94 \tag{24}$$

Therefore, the method assumes that the logarithm of the daily price, $p_t = \ln(P_t)$, of the portfolio satisfies the difference equation $p_t - p_{t-1} = r_t$, where $r_t = \sqrt{h_t}\epsilon_t$, is an IGARCH(1,1) process without drift and λ is a decay factor with a typical value of 0.94 for daily data and of 0.97 for monthly data (these figures are the result of J.P. Morgan's calibration exercises). When using the RiskMetrics$^{\text{TM}}$ methodology on a portfolio of assets, we also need to compute the coefficient of correlation given by $\rho_{tij} = h_{tij}/\sqrt{h_{ti}h_{tj}}$, in which the covariance is estimated using an exponential weighting scheme, that is:

$$h_{ijt} = \lambda h_{ijt-1} + (1-\lambda) r_{it-1} r_{jt-1}, \qquad \text{with } \lambda = .94 \tag{25}$$

Although RiskMetrics$^{\text{TM}}$ permits sizeable computational gains, Zaffaroni (2008) shows that it delivers non-consistent estimates and hence unreliable forecasts of the conditional variances and correlations.

Another simple way to calculate the VaR of a portfolio/asset, is to forecast its volatility as the historical Moving Average of the standard deviations, denoted as MA(m):

$$h(m) = \frac{1}{m} \sum_{t=1}^{m} r_t^2 \tag{26}$$

where m is the length of the estimation window and $r_t \sim N(0, h)$. In the empirical section of the paper, where I deal with daily data, I use two MA models with $m = 20$, and $m = 60$.

Forecasting VaR from copula models is less straightforward. Let us introduce some notation: log-prices are given by $p_t^i = \log\left(P_t^i\right)$ where $i = SC, LC$, log returns are given by $r_t^i = p_t^i - p_{t-1}^i$, standardized residuals after ARMA-GARCH estimation (i.e. ARMA residuals e_t^i divided by the estimated standard deviation $\sqrt{h_t^i}$) are denoted as ε_t^i, the PIT of ε_t^i are given by $u_t = F_t\left(\varepsilon_t^{SC}|\hat{\varphi}\right)$ and $v_t = F_t\left(\varepsilon_t^{LC}|\hat{\gamma}\right)$, where $\hat{\varphi}$ and $\hat{\gamma}$ are the estimated parameters of the marginals and $F_t(.|.)$ denotes the conditional c.d.f. of ε_t^i. Having defined these variables, we can write the value of an equally weighted portfolio containing the small cap index and the large cap index as:

$$V_t = \frac{1}{2}\exp\left(p_t^{SC}\right) + \frac{1}{2}\exp\left(p_t^{LC}\right) \tag{27}$$

The Profit and Loss ($P\&L$) function of this portfolio is given by $L_t = (V_t - V_{t-1})$. Alternatively, the $P\&L$ function can be expressed as:

$$L_t = \frac{1}{2}P_{t-1}^{SC}\left(\exp\left(r_t^{SC}\right) - 1\right) + \frac{1}{2}P_{t-1}^{LC}\left(\exp\left(r_t^{LC}\right) - 1\right) \tag{28}$$

The algorithm I use to obtain the recursive one-step ahead forecasts of the 5 percent VaR implied by copula models is the following:

1. Estimate the marginal distributions of returns using T observations;

2. Forecast returns and variances in $T+1$ and denote these as $\hat{r}_{T+1}^i$ and $\hat{h}_{T+1}^i$, for $i = SC, LC$;

3. Get u_t and v_t and estimate the copula parameters, denoted as $\hat{\kappa}$;

4. Simulate j random variables $\left(u_{T+1}^j, v_{T+1}^j\right)$, where $j = 1, ..., N$, from the copula function[4] estimated in step 3;

5. Get the (simulated) standardized residuals $\varepsilon_{T+1}^{i,j}$ using the fact that $\varepsilon_{T+1}^{SC,j} = F_{T+1}^{-1}\left(u_{T+1}^j|\hat{\varphi}\right)$ and that $\varepsilon_{T+1}^{SC,j} = F_{T+1}^{-1}\left(v_{T+1}^j|\hat{\gamma}\right)$, where $F^{-1}(.)$ is the inverse c.d.f.;

6. Get the simulated (forecasted) returns using the forecasted returns and variances from step 2 (i.e. simulated standardized residuals at time $T+1$ are defined as $\varepsilon_{T+1}^{i,j} = \left(r_{T+1}^{i,j} - \hat{r}_{T+1}^i\right)/\sqrt{h_{T+1}^i}$ therefore $r_{T+1}^{i,j} = \hat{r}_{T+1}^i + \varepsilon_{T+1}^{i,j}\sqrt{h_{T+1}}$);

7. Repeat steps 4-6 N times and use Equation (28) to get a sample of L_{T+1}^j for $j = 1, ..., N$;

8. Sort the j $P\&L$ functions in increasing order;

9. The VaR is the α quantile from the simulated empirical distribution of L_{T+1} (i.e. the $0.05N$-th observation in the sorted sample L_{T+1}).

[4] See Cherubini, Luciano and Vecchiato (2004).

It easy to understand that when using this algorithm a critical variable to be set is N, that is the number of simulations from the copula functions. Obviously, the larger N, the more accurate the VaR; however, copula simulation can be very time-consuming, especially when doing that within a recursive, or rolling forecasting scheme. For this reason, I have carried out a Monte Carlo exercise to determine N on the basis of the trade-off between accuracy of the VaR and CPU time. This exercise demonstrates that setting N=5000 represents a good compromise between accuracy and speed[5].

3.1 Backtesting VaR

I use two tools to evaluate the performance of different VaR models: statistical tests and loss functions. Let us define the following indicator variable as the hit series:

$$I_t = \begin{cases} 1 & \text{if} \quad L_t < VaR_t(q) \\ 0 & \text{if} \quad L_t \geq VaR_t(q) \end{cases} \tag{29}$$

where I_t, which can be written more compactly as $I_t = \mathbf{1}(L_t < VaR_t(q))$, is a dummy variable that takes on value one when the $P\&L$ function exceeds the forecasted VaR threshold.

Recall that the VaR threshold represents the critical value that corresponds to the lower q percent tail of the distribution of returns. Alternatively, q can be defined as the true probability coverage whose sample analogue is given by $\hat{q} = \sum_{t=1}^{T} I_t/T$ in which $\hat{q}$ is called nominal coverage.

With these definitions, I can introduce the trinity of tests due to Christoffersen (1998). These tests are based on the definition of (conditional) efficiency of the sequence of VaR forecasts; more precisely, we say that a series of VaR forecasts is efficient with respect to the information set Ψ_{t-1}, if $E(I_t|\Psi_{t-1}) = q$ for all t. These tests can be done in a likelihood ratio (LR) testing framework. A very convenient feature of Christoffersen's tests is that they can be carried out as a joint test of two properties of the hit series, namely we test separately the correct unconditional coverage and serial independence hypotheses.

The idea behind the unconditional coverage test is straightforward: accurate VaR estimates should exhibit the property that their nominal\unconditional coverage $\hat{q}$ equals the true probability coverage, say $q = 5$ percent. Let $x = \sum_{t=1}^{T} I_t$ be the number of exceptions in a sample of size T, then we can write the probability of x as[6]:

[5] Results are available from the author upon request.

[6] Notice that this corresponds to the probability density function of a Binomial variate. This stems from the fact that x is a sum of T Bernoulli variates I_t.

$$\Pr(x) = \binom{T}{x} q^{x} (1-q)^{T-x}. \qquad (30)$$

From (30) it follows that the maximum likelihood estimate of q can be written as[7] $\hat{q} = x/T$. For a set of 5 percent VaR forecasts, the LR statistic for testing the null hypothesis that $\hat{q} = q = 0.05$ against the alternative $\hat{q} \neq q$ is:

$$LR^{UC} = 2\left\{\log\left[\hat{q}^{x}(1-\hat{q})^{T-x}\right] - \log\left[0.05^{x} 0.95^{T-x}\right]\right\} \qquad (31)$$

As usual, we have $LR^{UC} \overset{asy}{\sim} \chi^2(s-1) = \chi^2(1)$, in which $s = 2$ is the number of possible outcomes of the hit series.

Christoffersen has shown that this test does not have any power against the alternative that the zeros and the ones in the hit series come clustered together in a time-dependent fashion; this explains why we need a test that helps identify the presence of dynamics in higher-order moments. The LR test of independence (LR^{IND}) is used to test the null hypothesis of serial independence against the alternative of first-order Markov dependence. Under null hypothesis LR^{IND} is asymptotically distributed as a $\chi^2(1)$.

Finally, the test of unconditional coverage and test of independence can be combined to form a test of conditional coverage (LR^{CC}). The test of conditional coverage can be written as:

$$LR^{CC} = LR^{UC} + LR^{IND} \overset{asy}{\sim} \chi^2(2) \qquad (32)$$

As we can see the LR^{CC} test is a joint test of unconditional coverage and independence.

The fourth and last statistical test I will use is due to Engle and Manganelli (2002). Let us consider a modified version of the hit series:

$$Hit_t = \mathbf{1}(L_t < VaR_t) - q \qquad (33)$$

and let $X_t = \begin{bmatrix} \iota & Hit_{t-1} & Hit_{t-2} & ... & Hit_{t-p} & VaR_t \end{bmatrix}$, where ι is column of ones. By regressing Hit_t on X_t we get: $\beta = (X_t' X_t)^{-1} X_t' Hit_t$.

The Dynamic Quantile (DQ) test statistic is given by:

[7] Notice that the log-likelihood function is given by: $\log\binom{T}{x} + x\log q + (T-x)\log(1-q)$. Solving the FOC for q yields $\hat{q} = x/T = \sum_{t=1}^{T} I_t/T$.

$$DQ(p) = \frac{\beta' X_t' X_t \beta}{q(1-q)} \sim \chi^2(p). \tag{34}$$

The test exploits the fact that, under the null of a correctly specified VaR model, the Hit series should have mean zero and should be independent of everything in the conditioning information set, including its lagged values and the VaR.

While the previous tests can be used to choose among models on the basis of the number of exceptions, they do not tell us anything about the magnitude of the exceptions. The magnitude of the exceptions can be evaluated using a set of loss functions. The first loss function, proposed by Lopez (1998), can be written as:

$$C_t^L = \begin{cases} 1 + (L_t - VaR_t)^2 & \text{if} \quad L_t < VaR_t \\ 0 & \text{if} \quad L_t \geq VaR_t \end{cases} \tag{35}$$

Notice that the penalty increases with magnitude of the VaR violation. An alternative loss function has been proposed by Blanco and Ihle (1999):

$$C_t^{BI} = \begin{cases} \frac{L_t - VaR_t}{VaR_t} & \text{if} \quad L_t < VaR_t \\ 0 & \text{if} \quad L_t \geq VaR_t \end{cases} \tag{36}$$

This loss function focuses on the average size of the exceptions. The third loss function, used also by González-Rivera et al. (2004), is the loss function used in quantile estimation [see, Kroner and Basset (1978)]:

$$Q = T^{-1} \sum_{i=1}^{T} (q - I_t)(L_t - VaR_t(q)) \tag{37}$$

where $I_t = \mathbf{1}(L_t < VaR_t(q))$, as defined in Equation (29). Q penalizes more heavily, with weight $(1-q)$, the observations for which $L_t - VaR_t(q) < 0$. A smaller Q indicates a better goodness-of-fit.

Lastly I use a loss function based on the Basel Committee's capital charges. The Basel Capital Accord sets the capital charge at the highest of the previous day's VaR, or the average VaR over the last 60 business days times a multiplicative factor k (i.e. this is to be determined by local regulators, but it cannot be less than 3):

$$CC_t = -(3+k) \max\left(\frac{1}{60} \sum_{t=1}^{60} VaR_{t-i}, VaR_{t-1} \right) \tag{38}$$

Notice that CC is the product of the negative of $(3+k)$ and the greatest between the previous day VaR and the mean VaR over the last sixty days. As we can see from table 1, k increases as a function of the number of violations of the VaR threshold over the last 250 trading days.

Table 1: Basel Penalty parameter as a function of the VaR violations in the last 250 trading days

Violations	<5	5	6	7	8	9	≥ 10
Increase in k	0.00	0.40	0.50	0.65	0.75	0.85	1.00

Source: Jorion (2007).

4 Application

In this section of the paper I analyze the forecasting performance of copula models. The next subsection describes the dataset used in the analysis, focusing on the properties of the marginal distributions of the data, as well as on their correlation. Then I analyze the goodness-of-fit of different parametric assumptions. Lastly, I compare the forecasting performance of copula models with a set of statistical tests and loss functions.

4.1 Data description

The theory of copulas is used to forecast the VaR of an equally weighted portfolio composed of two Dow Jones Global Indices (DJGI). These indices, which are available for 47 countries, include stocks that are categorized into 10 industries, 18 super-sectors, 39 sectors and 104 sub-sectors, as defined by the DJ's Industry Classification Benchmark. The empirical application of this paper is based on the DJ's U.S. Oil and Gas Producers Index, which includes companies involved in upstream activities (i.e. exploration and production), as well as integrated oil and gas companies (i.e. companies involved in both upstream and downstream activities, such as refining, marketing, and distribution).

This index is available at the sector level, as well as for three sub-categories of stocks, namely large-cap, mid-cap and small-cap stocks (i.e. indices of stocks sorted by their market capitalization).

The analysis uses the small cap index and the large cap index recorded at daily frequency from January 1992 to May 2008. The sample comprises a total of 4260 observations: 3500 observations are used for estimation purposes, while the remaining 760 observations are used for out-of-sample evaluation. Descriptive statistics for these samples are shown in table 2.

Both the average return and the volatility of small cap stocks are higher than those of large cap stocks over the three periods.

Table 2: Desctriptive statistics

	Full Sample		Estimation Sample		Forecasting Sample	
	Small Caps	Large Caps	Small Caps	Large Caps	Small Caps	Large Caps
Mean	17.5644	11.0880	15.0984	9.1980	28.8288	19.8324
Median	0.0316	0.0162	0.0110	0.0000	0.1821	0.1565
Min	-8.9679	-7.7561	-8.9679	-7.7561	-6.3539	-5.5127
Max	9.5355	9.6772	9.5355	9.6772	5.2323	4.5208
5% VaR	-1.6278	-1.6510	-1.6055	-1.6179	-1.7727	-1.8765
Std. Dev.	23.7975	20.4527	22.83855	19.9336	27.8058	22.6990
Skewness	-0.1866	0.0013	-0.1037	0.1311	-0.4169	-0.4211
Kurtosis	5.2535	5.7232	5.8686	6.4749	3.5452	3.4674
Jarque-Bera	924.1292	1313.6786	1203.5038	1767.2001	31.0172	29.0223
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
T	4260		3500		760	
Correlation	0.6458		0.5793		0.6677	

Notes: The table displays the annualized mean (i.e. the average daily return multiplied by 252) and the annualized volatility (i.e. the daily standard deviation times the squared root of 252). The null hypothesis of the Jarque-Bera statistic is that returns are unconditionally normal.

Notice that these indices have been chosen to represent a portfolio comprising a low-risk/low return position, namely the large cap index, and a high risk/high return position, that is the small-cap index.

Although, for both indices, the empirical 5 percent VaR is quite close to -1.65, that is the 5 percent quantile of a Normal variate, the Jarque-Bera statistic always rejects the null hypothesis of unconditional normality. Furthermore, notice that the small cap index always displays slightly negative excess skewness, while the large cap index displays slightly positive excess skewness. Both indices exhibit excess kurtosis and are positively correlated, with the coefficient of correlation close to 0.6. As I noted above, a common feature of financial data is that they are often more correlated in bear markets than in bull markets. The asymmetric dependence between the two indices can be measured using the framework of Longin and Solnik (2001) and Hong, Tu and Zhou (2007).

Let us define the exceedence correlation between two random variables X and Y, as $\rho^e(q)$:

$$\rho^e(q) = \begin{cases} \rho^- = corr\left(X, Y | X \leq Q_x(q) \cap Y \leq Q_y(q)\right) \\ \rho^+ = corr\left(X, Y | X > Q_x(q) \cap Y > Q_y(q)\right) \end{cases}$$

where $Q_x(q)$ and $Q_y(q)$ are the q-th quantiles of X and Y.

The empirical exceedence correlations of standardized returns[8] along with those implied by a Normal and a Rotated Gumbel copulas[9] are shown in figure 2. Similarly, figure 3 displays the empirical exceedence correlations of transformed standardized residuals[10] (u, v) along with what we would observe if they were linked with the Normal copula, or the Rotated Gumbel copula.

Notice that the information provided by these two plots are substantially different: while the former displays evidence of asymmetries in the unconditional distribution of returns, the latter shows the degree of asymmetry in their unconditional copula, that is after having removed all the asymmetries in the marginal distributions.

Clearly, both figures reveal that the assumption of multivariate normality, that im-

[8] Alternatively, we can define $\rho^e(q)$ as:

$$\rho^e(c) = \begin{cases} \rho^- = corr\left(\bar{X}, \bar{Y} | \bar{X} \leq c \cap \bar{Y} \leq c\right) \\ \rho^+ = corr\left(\bar{X}, \bar{Y} | \bar{X} > c \cap \bar{Y} > c\right) \end{cases}$$

where $\bar{X}$ and $\bar{Y}$ are standardized variables and the correlation at the exceedence level c is defined as the correlation between the two variables when both of them exceed c standard deviations away from their means. Notice that while figure (2) uses $\rho^e(q)$, the J-tests in table (3) are based upon $\rho^e(c)$.

[9] Notice that for these two copulas the parameters have been derived from relationship between Kendall's τ and their parameters. For the Normal copula $\tau = \frac{2}{\pi}\arcsin(\rho)$, while for the rotated Gumbel copula $\tau = 1 - \frac{1}{\delta}$.

[10] These are the Probability Integral Transforms of standardized residuals after normal GARCH(1,1) estimation.

Figure 2: Exceedence correlations between standardized returns.

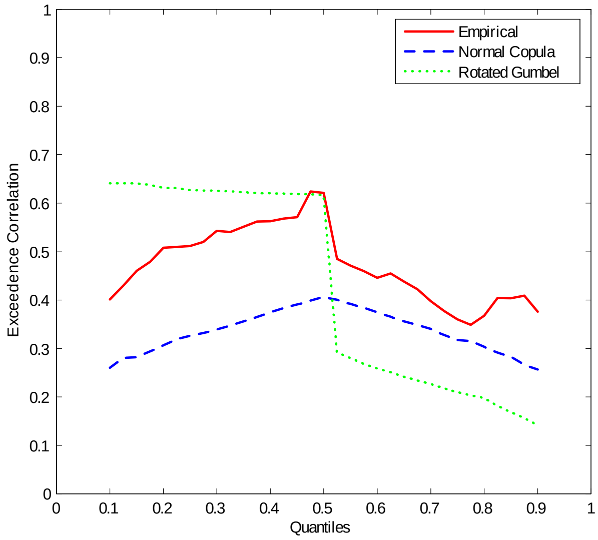
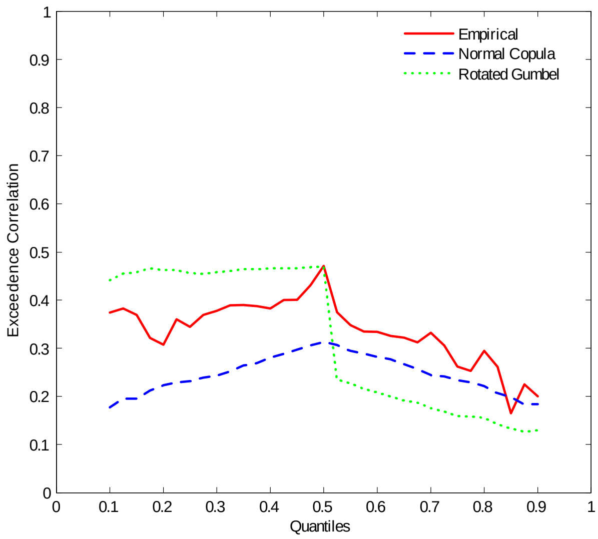


Figure 3: Exceedence correlations between transformed standardized residuals (u, v)



plies symmetric exceedence correlation, is not appropriate for the two indices I am analyzing.

Furthermore, both the F-test and the J-test in table 3 confirm what I have just highlighted: both tests reject the null hypothesis of symmetric correlation (with the exception of the out-of sample J-test). The last four columns of the table, which show the difference between negative and positive exceedence correlations for exceedence levels $c = 0, .5, 1, 1.5$, provide further evidence of the asymmetric dependence between the small cap index and the large cap index.

4.2 Marginal Distribution Models

The evidence presented above coupled with some pre-estimation statistical tests[11] suggest to model the marginal distribution for the two indices as follows:

$$\begin{aligned}
r_t^{SC} &= \beta_1 + \beta_2 r_{t-1}^{SC} + \varepsilon_{1,t} \\
h_t^{SC} &= \beta_3 + \beta_4 \varepsilon_{1,t-1}^2 + \beta_5 h_{t-1}^{SC} \\
r_t^{LC} &= \beta_6 + \beta_7 r_{t-1}^{LC} + \beta_8 \varepsilon_{2,t-1} + \varepsilon_{2,t} \\
h_t^{LC} &= \beta_9 + \beta_{10} \varepsilon_{2,t-1}^2 + \beta_{11} h_{t-1}^{LC} \\
\varepsilon_{j,t} &= \sqrt{h_t^i} \epsilon_{j,t} \text{ for } i = SC,\ LC \text{ and } j = 1,2 \\
& \quad \epsilon_{j,t} \tilde{} T^{skew} \left(\epsilon_{j,t} | \upsilon, \lambda\right)
\end{aligned}$$

The marginal distribution for the small cap index is assumed to be adequately characterized by an $AR(1) - T^{skew} - GARCH(1,1)$, while the marginal distribution for the large cap index is assumed to be characterized by $ARMA(1,1) - T^{skew} - GARCH(1,1)$.

Recall that Hansen's skewed T distribution nests, at least asymptotically, both the Student's T distribution (e.g. when $\lambda = 0$) and the Normal distribution (e.g. when $\lambda = 0$ and $\nu \to \infty$), therefore I can test the adequacy of the parametric assumptions by means of Likelihood Ratio tests (LR). As we can see from table 4, the LR tests reject the normality of the data, but do not reject the null that $\lambda = 0$; in other words, these tests are telling us that the Student's T distribution should be adequate for the data.

Lastly, I evaluate the goodness-of-fit of the three GARCH models with the framework proposed by Diebold, Gunter and Tay (1998). These authors showed that for a times series of PITs to be i.i.d. U(0,1) the sequence of densities must be correct.

[11] Results available from the author upon request.

Table 3: Exceedence correlation.

	F-test$^{(1)}$	P-value	F-test$^{(2)}$	P-value	J-test	P-value	$\hat{\rho}(c)^{-} - \hat{\rho}(c)^{+}$			
							$c_1 = 0.0$	$c_2 = 0.5$	$c_3 = 1.0$	$c_4 = 1.5$
Full-Sample	1528.7851	0.0000	1525.6813	0.0000	11.0247	0.0263	0.1360	0.1289	0.0243	0.0774
In-Sample	885.1013	0.0000	883.8785	0.0000	8.1719	0.0855	0.1102	0.0814	-0.0364	0.0873
Out-of-Sample	1156.4339	0.0000	1159.5031	0.0000	3.5478	0.4706	0.1460	0.2539	0.2303	0.1935

Notes: Columns 1-4 display an F-tests and the associated P-value based on the following regressions: (1) $r_t^{SC} = \alpha_0 + \alpha_1 r_t^{LC} \cdot 1(r_t^{LC} > 0) + \alpha_2 r_t^{LC} \cdot 1(r_t^{LC} \leq 0) + \xi_t$; $H_0 : \hat{\alpha}_1 = \hat{\alpha}_2$; (2) $r_t^{LC} = \beta_0 + \beta_1 r_t^{SC} \cdot 1(r_t^{SC} > 0) + \beta_2 r_t^{SC} \cdot 1(r_t^{SC} \leq 0) + \varepsilon_t$; $H_0 : \hat{\beta}_1 = \hat{\beta}_2$; where $1(\cdot)$ is an indicator function. The remaining portion of the table reports symmetric correlation test between (standardized) returns on the large caps index (r_t^{LC}) and (standardized) returns on the small caps index (r_t^{SC}). The J-test is the statistic for the symmetric hypothesis $H_0 : \hat{\rho}^{+}(c) = \hat{\rho}^{-}(c)$, where $\hat{\rho}^{+}(c) = corr(r_t^{SC}, r_t^{LC} | r_t^{SC} > c, r_t^{LC} > c)$ and $\hat{\rho}^{-}(c) = corr(r_t^{SC}, r_t^{LC} | r_t^{SC} < -c, r_t^{LC} < -c)$ are the exceedence correlations. This statistic has an asymptotic chi-square distribution with 4 degrees of freedom. See Hong et al. (2007).

Therefore, I test the goodness-of-fit of the marginal densities in two separate steps. In the first step I use the Kolmogorov-Smirnov test and an histogram to asses whether the transformed standardized residuals are U(0,1). Then, in the second step, I use the plot of their autocorrelation functions to test the independence hypothesis. For the Normal GARCH, as well as for the T-GARCH model, the Kolmogorov-Smirnov test rejects the null hypothesis that the transformed series are uniformly distributed. Moreover, some of the histograms' bins lie outside their confidence bounds. In the case of the skewed Student's T distribution the Kolmogorov-Smirnov test fails to reject the null and the histogram looks closer to what we would expect if the data were really i.i.d. Uniformly distributed. Lastly, in all cases the correlograms do not reveal any neglected dynamics[12].

All in all, I decided to use all of the three distributions for two reasons: (i) if on the one hand I can quite easily discard the Normal distribution, on the other hand the evidence just presented does not allow to choose between the T and the skewed Student's T distribution; (ii) each distribution allows to keep into consideration different features of the data. The Normal distribution has been used as benchmark model, that I will call the Normal-Normal model, that is a specification with Normal marginals linked with a Normal copula.

4.3 Copula Models

As we have seen above, the estimation of copula models can be broken down into two steps [see Equation (9)]; in the first step, I use Equations (10) and (11) to get QML estimates of the marginal distributions, while in the second step I use Equation (12) to estimate the parameters of the copula and Equation (13) to draw inferences.

For each marginal distribution, I have fitted five copulas: the Normal copula, the Clayton copula, the Plackett copula, the Rotated Gumbel copula and the SJC copula. These copula functions have been chosen because they have functional forms that allow to take into account the characteristics of the data. Recall from figures 2-3 that the indices are more correlated in bear markets than in bull markets. This fact is confirmed by the estimates of the SJC copula that, independently of the marginal distribution, always displays a coefficient of lower tail dependence $\left(\tau^L\right)$ slightly greater than the coefficient of upper tail dependence (τ^U), see table 5. The same table also shows how the models rank according to the value of the maximized log-likelihood function. The first and the third-best copula models are those that link the univariate T distributions with asymmetric copulas, such as the SJC and the Rotated Gumbel.

[12]Results available from the author upon request.

Table 4: Estimation of marginal distribution and goodness-of-fit tests.

	Small Cap	Large Cap
N-Lik	-5734.7868	-5419.5093
T-Lik	-5689.8773	-5366.7107
t^{sk}-Lik	-5689.2082	-5366.6924
AIC(N)	-11469.5718	-10839.0169
AIC(T)	-11379.7524	-10733.4191
AIC(t^{sk})	-11378.4136	-10733.3820
BIC(N)	-11469.5665	-10839.0117
BIC(T)	-11379.7454	-10733.4121
BIC(t^{sk})	-11378.4048	-10733.3732
LR(N vs. T)	89.8188	105.5972
(P-value)	0.0000	0.0000
LR(N vs. t^{sk})	91.1571	105.6338
(P-value)	0.0000	0.0000
LR(T vs. t^{sk})	1.3382	0.0366
(P-value)	0.2473	0.8483

Notes: For each distribution (Normal (N), Student's T (T) and skewed Student's T (t^{sk}) the table shows the value of the log-Likelihood (Lik), the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC) and a Likelihood Ratio test for nested models (e.g. in the case of LR(N vs t) we are testing whether the t-GARCH is better than the N-GARCH model).)

The second best model is characterized by T marginals linked with the (symmetric) Plackett copula. These specifications are followed by the Normal marginals-Plackett copula model and then by the Normal-Normal model. The specifications with T^{sk} marginals are the next in this ranking; however, it should be pointed out that the estimates of the skewness parameter are never statistically different from zero. Interestingly, models based on the Clayton copula are the worst three specifications. This is quite surprising, given that this copula has negative tail dependence as the Rotated Gumbel copula.

4.4 VaR results

The performance of different copula models is now evaluated in terms of their forecasting performances. Table 6 shows the number of VaR violations, the percentage of VaR violations and the results of the Unconditional Coverage $\left(LR^{UC}\right)$, Independence $\left(LR^{IND}\right)$, Conditional Coverage $\left(LR^{CC}\right)$ and Dynamic Quantile (DQ) tests for twelve models. Three of these models, namely the RiskMetrics$^{\text{TM}}$, the MA(20), and the MA(60) model, are single index models and the remaining are portfolio models. In other words, the first three models do not take into account the correlation between the small cap index and the large cap index, while the remaining specifications take diversification into account.

When looking at the number of exceptions, the best model seems to be the one based

Table 5: Copula estimation

		Coefficient	Std. Error	Rank	$-\mathcal{L}_C$	AIC	BIC
N+Normal	ρ	0.5463	0.0099	5	620.41	1240.81	1240.81
N + Clayton	θ	0.7622	0.0314	15	463.96	927.93	927.92
N + Plackett	κ	6.4716	0.2879	4	624.32	1248.63	1248.63
N + Rotated Gumbel	δ	1.5316	0.0209	11	580.48	1160.97	1160.97
N + SJC	τ^U	0.2776	0.0213	12	574.88	1149.77	1149.76
	τ^L	0.3407	0.0178				
T + Normal	ρ	0.5450	0.0129	7	612.64	1225.27	1225.27
T + Clayton	θ	1.0669	0.0374	13	536.71	1073.42	1073.42
T + Plackett	κ	7.2299	0.3346	2	636.99	1273.97	1273.97
T + Rotated Gumbel	δ	1.6894	0.0248	3	631.87	1263.75	1263.74
T + SJC	τ^U	0.4011	0.0208	1	648.65	1297.30	1297.30
	τ^L	0.4529	0.0169				
t^{sk} + Normal	ρ	0.5450	0.0099	8	610.81	1221.61	1221.61
t^{sk} + Clayton	θ	0.8178	0.0309	14	501.62	1003.24	1003.24
t^{sk} + Plackett	κ	5.9125	0.2644	9	606.12	1212.24	1212.24
t^{sk} + Rotated Gumbel	δ	1.5329	0.0204	10	595.29	1190.58	1190.58
t^{sk} + SJC	τ^U	0.3250	0.0206	6	619.50	1238.99	1238.99
	τ^L	0.3602	0.0189				

Notes: For each distribution (Normal (N), Student's T (T) and skewed Student's T (t^{sk}) and for each copula the table shows the estimated copula parameters, its standard error, the ranking of the model (Rank) based on the value of the maximized log-Likelihood function ($\mathcal{L}_C$), the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC).

on the portfolio RiskMetrics^{TM} methodology, which delivers only 19 violations out of 760 VaR forecasts, while the models with the worst performance, at least 55 exceptions, seem to be those using the skewed Student's T distributions for the marginals. These numbers are quite impressive given that the second-best model, namely the T-Rotated Gumbel copula specification, produces 32 VaR exceptions.

A more useful statistics is the percentage of violations; recall that this is obtained as 100 times $\hat{q} = \left(760^{-1} \sum_{t=1}^{760} \mathbf{1}\left(L_t < VaR_t\right)\right)$, where $\hat{q}$ is the empirical, or unconditional coverage. When forecasting the 5 percent VaR, we expect $\hat{q}$ to be close to 5 percent (i.e. the true probability coverage). In this sense, we can see that the portfolio RiskMetrics^{TM} approach corresponds to an unconditional coverage of 2.5 percent, that is a half of what we would expect.

Notice that models using the Student's T distribution for the marginals are those with the best unconditional coverage, ranging from 4.21 percent to 5 percent. Once again, the models using the skewed Student's T distribution are among the worst, implying unconditional coverages slightly higher than 7 percent.

When looking at the LR test for independence, I cannot discriminate among different models, indeed none of them reject the null hypothesis of serial independence of the hit series. However, when moving to the Conditional Coverage LR test, I can confirm what already noticed when I looked at the percentage of violations.

All of the single index models pass the test, while neither the models using the skewed Student's T distribution, nor the portfolio RiskMetrics^{TM} approach pass the test. On the contrary, all of the models based on the T-distribution display correct conditional coverage.

From now on I analyze only those models that have correct conditional coverage, namely the single index RiskMetrics^{TM} model, the two MA models, all the copula models using the Student's T distribution. As usual, I also consider the Normal-Normal model to have a benchmark.

Of this subset of models, those that fail the DQ test are the following: RiskMetrics^{TM}, MA(60), "Normal-Normal" and "T-Plackett".

Summing up, among the models that pass both the LR^{CC} test and the DQ test (at the 5 percent significance level) we find three portfolio VaR models (i.e. the "T-Normal", the "T-Rotated Gumbel" and the "T-SJC" copula models) and one single index model (i.e. the MA(20) model).

The subset of models that, according to the statistical tests, deliver correct conditional coverage are now evaluated in terms of their loss functions. This exercise is carried out in table 7, while they are ranked according to lowest loss function and least number of violations in table 8.

Table 6: Backtesting of 5 percent VaR Forecasts: Statistical Tests

	# Violations	% Violations	LR^{UC}	LR^{IND}	LR^{CC}	DQ
RiskMetrics^{TM} (SI)*	48	6.32	2.5942	1.1463	3.7404	12.9224
p-value			0.1073	0.2843	0.1541	0.0443
MA(20)*,†	51	6.71	4.2843	0.0793	4.3637	11.7127
p-value			0.0385	0.7782	0.1128	0.0687
MA(60)*	51	6.71	4.2843	0.6686	4.9529	17.4732
p-value			0.0385	0.4135	0.0840	0.0077
N + Normal	57	7.50	8.7809	0.5445	9.3254	13.2330
p-value			0.0030	0.4606	0.0094	0.0395
T + Normal*,†	38	5.00	0.0001	0.6166	0.6167	8.8593
p-value			0.9934	0.4323	0.7347	0.1816
T + Rotated Gumbel*,†	32	4.21	1.0348	0.2566	1.2915	8.0954
p-value			0.3090	0.6124	0.5243	0.2312
T + Plackett*	38	5.00	0.0001	0.5419	0.5420	12.9546
p-value			0.9934	0.4617	0.7626	0.0438
T + SJC*,†	35	4.61	0.2476	0.0705	0.3181	6.1299
p-value			0.6188	0.7905	0.8530	0.4088
t^{sk} + Normal†	56	7.37	7.9329	0.0104	7.9433	10.3955
p-value			0.0049	0.9187	0.0188	0.1090
t^{sk} + Rotated Gumbel	55	7.24	7.1234	0.3473	7.4707	12.6529
p-value			0.0076	0.5556	0.0239	0.0489
t^{sk} + Plackett†	56	7.37	7.9329	0.0104	7.9433	11.4632
p-value			0.0049	0.9187	0.0188	0.0751
RiskMetrics^{TM} (P)†	19	2.50	12.1042	1.0278	13.1320	11.1186
p-value			0.0005	0.3107	0.0014	0.0848

Notes: For each model the table shows the number and the percentage of violations of the VaR threshold and three statistical tests: Christoffersen's (1997) tests for Unconditional Coverage (LR^{UC}), Independence (LR^{IND}), Conditional Coverage (LR^{CC}), and the Dynamic Quantile Test of Engle and Manganelli (2002). Models denoted with an "*" pass the Conditional Coverage Test, models denoted with a "†" pass the DQ test.

The joint analysis of these tables reveals that, for three loss functions out of four, two of the T-SJC copula models rank either first, or second. However, when looking at the mean capital charge, asymmetric copula models (i.e. T-SJC and T-RG) rank last and second-last, with the MA models being in the first positions.

Notice however that the MA models are the worst models in the case of three loss functions out of four, as well as for the number of VaR violations.

Moreover, recall that copula models have unconditional coverage closer to 5 percent than MA models. Indeed, none of the MA models pass the unconditional coverage test.

Given that results in tables 7 and 8 are not robust to the choice of the loss function (i.e. MA models outperform copula models when using the mean capital charge), I provide further evidence about the forecasting performance of different models by means

Table 7: Backtesting of 5 percent VaR Forecasts: Loss Functions and Capital Charges

	Lopez	BI	Q	Min CC	Mean CC	Max CC
MA(20)	4701.6366	19.9591	1.5878	40.0664	79.6824	142.1621
MA(60)	5317.1557	20.6035	1.6265	57.7675	81.7763	134.4733
T + Nl	3154.3608	13.2763	1.5316	46.8042	87.1998	156.7110
T + RG	2849.8654	11.6531	1.5338	49.3558	88.9990	160.8449
T + PL	3042.9603	12.5602	1.5296	50.5477	88.0974	158.6559
T + SJC	2809.7689	11.9025	1.5264	47.4040	88.9291	158.6794

Notes: The table uses the following notation to denote models and loss functions: T = Student's T marginal distribution; RG = Rotated Gumbel copula; N = Normal copula; PL = Plackett Copula; BI = Blanco and Ihle; Q = Quantile Loss Function; CC = Capital Charge

Table 8: Ranking of 5 percent VaR Forecasts

Rank	# Violations	Lopez	BI	Q	Mean CC
1	T+RG	T+SJC	T+RG	T+SJC	MA(20)
2	T+SJC	T+RG	T+SJC	T+PL	MA(60)
3	T+N	T+PL	T+PL	T+N	T+N
4	T+PL	T+N	T+N	T+RG	T+PL
5	MA(20)	MA(20)	MA(20)	MA(20)	T+SJC
6	MA(60)	MA(60)	MA(60)	MA(60)	T+RG

Notes: The table uses the following notation to denote models: T = Student's T marginal distribution; RG = Rotated Gumbel copula; N = Normal copula; PL = Plackett Copula.

of the forecast evaluation test developed by Diebold and Mariano (1995). This test, referred to as the DM test, is a test for equal predictive ability of two competiting forecasts. The null of equal predictive accuracy is tested against composite alternatives that suggest which of the two models performs better. The test is based on a sample of loss differentials $d_t = L\left(f_t^A\right) - L\left(f_t^B\right)$, with $L(.)$ being some arbitrary loss function of f_t^i, that is, the time t forecast from model $i = A, B$. Equal predictive accuracy implies $E\left(d_t\right) = 0$ which can be tested using its sample counterpart $\bar{d} = T^{-1}\sum_{t=1}^{T} d_t$. The DM test can be written as:

$$DM = \frac{\bar{d}}{\sqrt{v\hat{a}r\left(\bar{d}\right)}} \tag{39}$$

where $v\hat{a}r\left(\bar{d}\right)$ is the asymptotic variance of $\bar{d}$ computed with the Newey-West (1987) heteroskedasticity and autocorrelation consistent (HAC) estimator. Under the null of equal predictive accuracy, the DM test is normally distributed.

In the paper I use a modified version of the test, denoted as MDM, put forth by Harvey, Leybourne and Newbold (1997), that can improve the power of the test in small samples. This test is given by $MDM = T^{1/2}\left[T + 1 - 2h + T^{-1}h\left(h-1\right)\right]^{1/2} DM$, in which

h is the forecast horizon (i.e. in this case, $h = 1$). Under the null of equal forecast accuracy $MDM \sim T_{(T-1)}$. Given that most of the asymptotics in the forecasting literature requires the loss function to be differentiable, I implement the tests with a differentiable version of the Quantile Loss function, Q [see Equation (37) and González-Rivera et al. (2004)]. From now on, we have $\widetilde{Q} = T^{-1} \sum_{i=1}^{T} (q - m_\delta ((L_t, VaR_t(q)))) (L_t - VaR_t(q))$, where $m_\delta (a, b) = \{1 + \exp [\delta (a - b)]\}^{-1}$. The parameter $\delta > 0$ controls the smoothness[13] of $\widetilde{Q}$.

Table 9 shows that we can reject the null of equal predictive accuracy of the two MA models when these are tested against copula models; on the other hand, the MDM test does not help to choose among copula models. Thus, for this sample, the MDM test has shown once again that copula models outperform simple models, such as the MA specifications.

Lastly, I implement a test introduced by White (2000) and known as "Reality Check" (RC). The null hypothesis of White's RC test is that a given benchmark model performs at least as well as the best competiting alternative model. The test is applied three times on an increasingly smaller set of models and uses loss differentials (i.e. the d_t's of the DM test) calculated with the smoothed quantile loss function, $\widetilde{Q}$.

The first set of models include MA models, RiskMetrics^{TM} models, and the copula specifications excluding those based on skewed Student's T marginals. Results are shown in table 10 where each model in the first column has been used as the benchmark and compared with all the remaining models. As we can see, the table displays two types of p-values labelled as "White" and "Hansen$_L$" [see Hansen, 2001]. The "Hansen$_L$" p-values use the adjustment due to Hansen (2005) that correct for the fact that the "White" p-value tends to become very large whenever a poor model is introduced.

From table 10 we can see that the MA models, as well as the RiskMetrics^{TM} specifications are the least preferred models. The copula models are associated with quite high p-values; notice that the Normal-Normal model represents the worst copula specification.

Given these results I recalculate the RC test on copula models only[14]. In this case, the p-value associated to the Normal-Normal specification is 10 percent, therefore the null hypothesis is rejected and there exist at least one model outperforming that based on the Gaussian distribution.

Lastly, I apply the RC test only to copula models using the Student's T distribution for the marginals. In this case, none of the benchmark models is outperformed by the alternative specifications; however, once again the SJC copula specification appears to

[13] I set $\delta = 25$, however the test does not vary significantly with δ.

[14] Results available from the author upon request.

be the most preferred.

Table 9: Modified Diebold and Mariano Test

	T+N	T+PL	T+RG	T+SJC
MA(20)	-1.9121	-1.9048	-1.6956	-1.8440
p-value	0.0562	0.0572	0.0904	0.0656
MA(60)	-2.0547	-2.0591	-1.8819	-1.9819
p-value	0.0403	0.0398	0.0602	0.0479
T+N		-0.3683	0.2820	-0.6734
p-value		0.7127	0.7780	0.5009
T+PL			0.4950	-0.3624
p-value			0.6207	0.7172
T+RG				-1.1529
p-value				0.2493

Notes: The table uses the following notation to denote models: T = Student's T marginal distribution; RG = Rotated Gumbel copula; N = Normal copula; PL = Plackett Copula.

Table 10: Reality Check: Step 1

	White	Hansen_L
MA(20)	0.1440	0.0390
MA(60)	0.0720	0.0460
RiskMetrics^{TM} (SI)	0.2090	0.0690
N+N	0.3050	0.1030
T+N	0.9260	0.4430
T+PL	0.9270	0.4610
T+RG	0.8880	0.3430
T+SJC	0.9820	0.6480
RiskMetrics^{TM} (P)	0.0230	0.0230

Notes: Each model in the first column has been used as the benchmark and compared with all the remaining models. The Null of the Reality Check (RC) is that the benchmark is at least as good as the other models. "White" and "Hansen" denote the p-values of the RC. The bootstrap RC p-values are computed with 1000 bootstrap resamples using the block bootstrap of Politis and Romano (1994). The block length has been determined with automatic-block length selection method of Politis, White and Patton (2007) and is equal to 12. The table uses the following notation to denote models: T = Student's T marginal distribution; RG = Rotated Gumbel copula; N = Normal marginal/copula; PL = Plackett Copula.

5 Conclusions

In this paper I have used copula functions to forecast the VaR of an equally weighted portfolio comprising a small cap stock index and a large cap stock index for the oil and gas industry. Such a portfolio represents a very general investment strategy, namely one based on a low-risk/low-return position, the large cap index, and a high-risk/high return position, the small cap index.

Copula functions have been used because they allow us to take simultaneously into account two characteristics of financial data: nonnormalities at the univariate, as well at the multivariate level. Nonnormalities in the marginals, such as excess skewness and/or excess kurtosis, can be taken into account with a variety of univariate models. However, when considering multivariate modelling, curse of dimensionality comes into play. On the contrary, the strength of copula functions relies on their flexibility. In fact, these functions can be used to link marginal distributions and to generate flexible multivariate specifications.

With this paper I have answered a set of empirical questions: (i) are there nonnormalities in the marginals? (ii) are there nonnormalities in the dependence structure? (iii) is it worth taking these nonnormalities into account for risk-management applications? (iv) do complicated models perform better than simple models?

As for questions (i) and (ii), I have shown that the data do deviate from the null of normality at the univariate, as well as at the multivariate level. The marginal of the small cap index, as well as that of the large cap index display kurtosis and skewness different from what we would expect in the case of normally distributed time series. The most serious problem is represented by excess kurtosis; on the contrary, excess skewness does not seem to be important neither in the estimation stage, nor for risk management purposes.

When considering the dependence structure of the data, I have found that they are more correlated in market downturns than in market upturns. Asymmetries show up in their unconditional distribution, as well as in their unconditional copula, that is after having removed the nonnormalities from their marginal distributions.

As for the importance of nonnormalities for risk management purposes, the VaR forecasting exercise has shown that models based on Normal marginals and/or with symmetric dependence structures fail to deliver accurate VaR forecasts. Among the models that deliver correct VaR forecasts, we have both very simple models, such as MA models, and copula models with Student's T marginals and asymmetric copula functions, as well as a model with T marginal and Normal, symmetric, copula. By means of a set of loss functions, I conclude that the T-asymmetric copula models deliver the best VaR forecasts. This fact is illustrated also with the Diebold and Mariano test and with White's Reality Check test. These findings confirm the importance of nonnormalities and asymmetries both in-sample and out-of-sample. This last observation is not a trivial one, indeed a common finding in the forecasting literature is that complicated models often provide poorer performance than simple, even misspecified, models [Swanson and

White (1995, 1997)].

The paper leaves unanswered at least two questions. First, the importance of time varying correlations which can be addressed with time-varying copulas [see Fantazzini (2004)]. Second, the forecasting performance of different models could be evaluated by means of different loss functions and with a wider set of statistical tests, such as the Conditional Quantile Forecast Encompassing test of Giacomini and Komunjer (2005). Alternative, modelling techniques could also be used in place of GARCH models. Interesting alternatives are represented by the use of copula functions in conjunction with Extreme Value Theory [see McNeil and Frey (2000)], or with quantile regression methods [see Chen, Koenker and Xiao (2008) and Gourieroux and Jasiak (2008)].

References

[1] Artzner P., F. Delbaen and J.M. Eber, 1999, "Coherent measures of risk", *Mathematical Finance*, 9, 203-228.

[2] Bollerslev T., 1990, "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model", *Review of Economics and Statistics*, 72, 498-505.

[3] Blanco C. and G. Ihle, 1999, "How good is your VaR? Using backtesting to assess system performance", *Financial Engineering News*, 1-2.

[4] Breymann W., A. Dias, P. Embrechts, 2003, "Dependence structures for multivariate high–frequency data in finance", *Quantitative Finance*, 3 ,1-16.

[5] Carvalho C. and K. Amonlirdviman, 2008, "Myopic loss aversion, asymmetric market comovements and the home bias", working paper, Princeton University.

[6] Chen X., R.W. Koenker and Xiao Z., 2008, "Copula-Based Nonlinear Quantile Autoregression", *Cowles Foundation Discussion Paper*, No. 1679.

[7] Cherubini U., E. Luciano and W. Vecchiato, 2004, *Copula models in finance*, John Wiley & Sons, Chichester, UK.

[8] Christoffersen P.F., 1998, "Evaluating Interval Forecasts", *International Economic Review*, 39, 841-862.

[9] Demarta S. and A.J. McNeil, 2004, "The t and related copulas", available at: http://www.math.ethz.ch/~mcneil.

[10] Dias A., 2004, "Copula inference for finance and insurance", Doctoral Thesis ETH No. 15283, Swiss Federeal Institute of Technology, Zurich.

[11] Diebold F.X., T. Gunther, and A. Tay, 1998, "Evaluating density forecasts with applications to financial risk management", *International Economic Review*, 39, 863-883.

[12] Diebold F.X. and R. Mariano, 1995, "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.

[13] Embrechts P., F. Lindskog and A.J. McNeil, 2001, "Modelling dependence with copulas and applications to risk management", available at: http://www.math.ethz.ch/~mcneil.

[14] Embrechts P., A. McNeil and D. Straumann, 2002, “Correlation and Dependence in Risk Management: Properties and Pitfalls” in Risk Management: Value at Risk and Beyond, ed. M.A.H. Dempster, Cambridge University Press, Cambridge, 176-223.

[15] Embrechts P., A.J. McNeil and D. Straumann, 1998, "Correlation and dependency in risk management: properties and pitfalls", available at: http://www.math.ethz.ch/~mcneil.

[16] Engle R.F. and K. Sheppard, 2001, "Theoretical and empirical properties of Dynamic Conditional Correlation multivariate GARCH", *University of California Discussion Paper*, 2001-15.

[17] Fantazzini D., 2004, "Dynamic copula modelling for Value at Risk", *Frontiers in Finance and Economics* (forthcoming), available at: http://ssrn.com/abstract=944172.

[18] Giacomini R. and I. Komunjer, 2005, "Evaluation and combination of conditional quantile forecasts", *Journal of Business and Economic Statistics*, 23, 416-431.

[19] González-Rivera G., T.H. Lee and S. Mishra, 2004, "Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood", *International Journal of Forecasting*, 20, 629-645.

[20] Gourieroux C. and J. Jasiak, 2008, "Dynamic quantile models", *Journal of Econometrics*, 147, 198-205.

[21] Granger C.W.J., 2003, "Time series concepts for conditional distributions", *Oxford Bulletin of Economics and Statistics*, 65, 689-701.

[22] Granger C.W.J., T. Teräsvirta and A.J. Patton, 2006, "Common factors in conditional distributions for bivariate time series", *Journal of Econometrics*, 132, 43-57.

[23] Hansen B.E., 1994, "Lagrange Multiplier test for parameter instability in non-linear model", working paper, available at: http://www.ssc.wisc.edu/~bhansen/papers/LMTests.pdf.

[24] Hansen B.E., 1994, "Autoregressive conditional density estimation", *International Economic Review*, 35,705-730.

[25] Hansen P.R., 2005, "A test for superior predictive ability", Woking Paper, 01/06, Brown University Dept. of Economics.

[26] Harvey C.R. and A. Siddique, 1999, "Autoregressive conditional skewness", *Journal of Financial and Quantitative Analysis*, 34, 465-478.

[27] Harvey C.R. and A. Siddique, 2000, "Conditional skewness in asset pricing tests", *Journal of Finance*, 55, 1263-1295.

[28] Harvey D.I., S.J. Leybourne and P. Newbold, 1997, "Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, 13, 281-291.

[29] Hong Y., J. Tu and G. Zhou, 2007, "Asymmetries in stock returns: statistical tests and economic evaluation", *Review of Financial Studies*, 20, 1547-81.

[30] Hull J. and A. White, 1998, "Value-at-Risk when daily changes in market variables are not normally distributed", *Journal of Derivatives*, 5, 9-19.

[31] Jondeau E. and M. Rockinger, 2003, "Conditional volatility, skewness, and kurtosis: existence, persistence and comovements", *Journal of Economic Dynamics and Control*, 27, 1699-1737.

[32] Jorion P., 2007, *Value at Risk: the new benchmark for managing financial risk*, 3rd ed., McGraw-Hill, New York.

[33] Kroener R. and G. Basset, 1978, "Regression quantiles", *Econometrica*, 46, 33-50.

[34] Kupiec P., 1995, "Techniques for verifying the accuracy of risk measurement models", *Journal of Derivatives*, 3, 73-84.

[35] Longin F. and B. Solnik, 2001, "Extreme correlation of international equity markets", *Journal of Finance*, 56, 649-676.

[36] Lopez J., 1998, "Methods for evaluating Value-at-Risk estimates", Federal Reserve Bank of New York, research paper no. 9802.

[37] McAleer M. and B. da Veiga, 2005, "Forecasting Value-at-Risk with a parsimonious portfolio spillover GARCH model", working paper.

[38] McNeil A.J. and R. Frey, 2000, "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach", *Journal of Empirical Finance*, 7, 271-300.

[39] Newey W.K. and D. McFadden, 1994, "Large sample estimation and hypothesis testing", in *Handbook of Econometrics*, chapter 36, vol. 4, Elsevier North Holland.

[40] Newey W.K. and K.D. West, 1987, "A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703-708.

[41] Patton A.J., 2004, "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation", *Journal of Financial Econometrics*, 2, 130-168.

[42] Patton A.J., 2006*a*, "Estimation of multivariate models of possibly different lengths", *Journal of Applied Econometrics*, 21, 147-173.

[43] Patton A.J., 2006*b*, "Copula-Based Models for Financial Time Series", working paper, available at: http://fmg.lse.ac.uk/~patton.

[44] Patton A.J., 2006*c*, "Modelling asymmetric exchange rate dependence", *International Economic Review*, 47, 527-556.

[45] Politis D. and J. Romano, 1994, "The stationary bootstrap", *Journal of the American Statistical Society*, 89, 1303-1313.

[46] Politis D. and H. White, 2004, "Automatic block-length selection for the dependent bootstrap", *Econometric Review*, 23, 53-70.

[47] Politis D., H. White and Patton A.J., 2007, "Correction: Automatic block-length selection for the dependent bootstrap", working paper, available on-line at: http://www.economics.ox.ac.uk/members/andrew.patton.

[48] Rosenblatt M., 1952, "Remarks on a multivariate transformation", *The Annals of Mathematical Statistics*, 23, 470-472.

[49] Schuster E.F., 1976, "The probability integral transformation: a simple proof", *Mathematics Magazine*, 49, 242-243.

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

[51] Swanson N.R. and H. White, 1995, "A model selection approach to assesing the information in the term structure using linear models and artificail neural networks", *Journal of Business and Economic Statistics*, 13, 265-275.

[52] Swanson N.R. and H. White, 1997, "A model selection approach to real-time macroeconomic forecasting using linear models and artificail neural networks", *Review of Economics and Statistics*, 79, 540-550.

[53] White H., 1982, "Maximum likelihood estimation of misspecified models", *Econometrica*, 50, 1-25.

[54] White H., 2000, "A reality check for data snooping", *Econometrica*, 68, 1097-1126.

[55] Zaffaroni P., 2008, "Large-scale volatility models: theoretical properties of professionals' practice", *Journal of Time Series Analysis*, 29, 581-599.