Some Statistical Pitfalls in Copula Modeling for Financial Applications

Fermanian, Jean-David; Scaillet, Olivier

FACULTE DES SCIENCES ECONOMIQUES ET SOCIALES

HAUTES ETUDES COMMERCIALES

SOME STATISTICAL PITFALLS IN COPULA MODELING FOR FINANCIAL APPLICATIONS

Jean-David FERMANIAN
Olivier SCAILLET

HEC GENÈVE

UNIVERSITÉ DE GENÈVE

SOME STATISTICAL PITFALLS IN COPULA MODELING FOR FINANCIAL APPLICATIONS

Jean-David FERMANIAN[a] and Olivier SCAILLET[b] [1]

[a] CDC Ixis Capital Markets, 47 quai d'Austerlitz, 75648 Paris cedex 13 and CREST, 15 bd Gabriel Péri, 92245 Malakoff cedex, France. jdfermanian@cdcixis-cm.com

[b] HEC Genève and FAME, Université de Genève, 102 Bd Carl Vogt, CH - 1211 Genève 4, Suisse. scaillet@hec.unige.ch

March 2004

Abstract

In this paper we discuss some statistical pitfalls that may occur in modeling cross-dependences with copulas in financial applications. In particular we focus on issues arising in the estimation and the empirical choice of copulas as well as in the design of time-dependent copulas.

Key words: Copulas, Dependence Measures, Risk Management.

[1]The second author receives support by the Swiss National Science Foundation through the National Center of Competence: Financial Valuation and Risk Management. Part of this research was done when he was visiting THEMA and ECARES.

1 Introduction

In modern finance and insurance, the identification and modeling of dependence structures between assets is one of the main challenges we are faced with. Dependence structures must be unveiled for several purposes : control of risk clustering [2], pricing and hedging of credit sensitive instruments, particularly n-th to default credit derivatives and Collateralized Debt Obligations (CDO's) [3], pricing and hedging of basket derivatives and structured products [4], credit portfolio management [5], credit and market risk measurement [6]. In that respect, copulas have been recently recognized as key tools to analyze dependence structures in finance. They are becoming more and more popular among academics and practitioners because it is well known that the returns of financial assets are non-Gaussian and exhibit strong nonlinearities. Clearly, multivariate Gaussian random variables do not provide suitable building blocks from an empirical point of view, and copulas appear as a natural modeling device in a non-Gaussian world [7].

The concept of "copulas" or "copula functions" as named by Sklar [58] originates in the context of probabilistic metric spaces. The idea behind this concept is the following: for multivariate distributions, the univariate margins and the dependence structure can be separated and the latter may be represented by a copula. The word copula, resp. copulare, is a latin noun, resp. verb, that means "bond", resp. "to connect" or "to join". The term copula is used in grammar and logic to describe that part of a proposition which connects the subject and predicate. In statistics, it now describes the function that "joins" one-dimensional distribution functions to form multivariate ones, and may serve to characterize several dependence concepts. The copula of a multivariate distribution can be considered

[2] see Denuit and Scaillet [15] for testing procedures of positive quadrant dependence, and Cebrian et al. [6] for testing procedures of concordance ordering.

[3] see Li [45] for a Gaussian copula-based methodology to price first-to-default credit derivatives as well as Laurent and Gregory [43] and the references therein.

[4] for instance, Rosenberg [51] estimates copulas for the valuation of options on DAX30 and S&P500 indices. Cherubini and Luciano [9] calibrate Frank's copula for the pricing of digital options and options that are based on the minimum on some equity indices. Genest et al. [33] study the relation between multivariate option prices and several parametric copulas, by assuming a GARCH type model for individual returns.

[5] see e.g. Frey and McNeil [29].

[6] see Schönbucher [55], Schönbucher and Schubert [56], among others.

[7] see Embrechts et al. [16], [17] and Bouyé et al. [4] for a survey of financial applications, Frees and Valdez [28] for use in actuarial practice, Patton [46, 47] and Rockinger and Jondeau [50] for introducing copulas in the modeling of conditional dependencies.

as the part describing its dependence structure as a complement to the behavior of each of its margins. Formally Sklar's Theorem states that a d-dimensional cumulative distribution function F evaluated at point $\mathbf{x} = (x_1, ..., x_d)$ can be represented as

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)), \tag{1.1}$$

where C is the copula function and F_i, $i = 1, ..d$, are the margins. In most cases the latter function is uniquely defined by (1.1).

One attractive property of copulas is their invariance under strictly increasing transformations of the margins [8]. Actually, the use of copulas allows to solve a difficult problem, namely to find a whole multivariate distribution, by performing two easier tasks. The first task starts with modeling each univariate marginal distribution either parametrically or non-parametrically. The second task consists of specifying a copula, which summarizes all the dependencies between margins. However this second task is still in its infancy for most of multivariate financial series, partly because of the presence of temporal dependencies (serial autocorrelation, time varying heteroskedasticity,...) in returns of stock indices, credit spreads, or interest rates of various maturities.

Without any doubt, copulas consist of a powerful tool to model dependence structures. They are however subject to some statistical pitfalls if used without appropriate care in financial applications. In this paper our aim is to discuss some of these pitfalls. In particular we analyze issues arising in the estimation (Section 2) and the empirical choice of copulas (Section 3) as well as in the design of time-dependent copulas (Section 4).

2 How to estimate copulas?

If the true copula is assumed to belong to a parametric family $\mathcal{C} = \{C_\theta, \theta \in \Theta\}$, consistent and asymptotically normally distributed estimates of the parameter θ (assumed to live in some parameter space Θ) can be obtained through maximum likelihood methods. There are mainly two ways to achieve this : a fully parametric method and a semiparametric method. The first method relies on the assumption of parametric univariate marginal distributions. Each parametric margin is then plugged in the full likelihood and this full likelihood is maximized with respect to the parameter θ. Alternatively and without any parametric assumptions for margins, the univariate empirical cumulative distribution functions can be plugged in

[8] Recall that is is not true for the standard Pearson's correlation coefficient.

the likelihood to yield a semiparametric method. These two commonly used methods are detailed in Genest et al. [32] and Shi and Louis [57]. At first glance, the task looks easy. Nonetheless, the result of the first method depends on the right specification of all margins. This may induce too severe constraints, and this aspect lessens the interest of working with copulas. The semiparametric estimation procedure where margins are left unspecified does not suffer from this inconvenient feature, but suffers from a loss of efficiency (see [31]). Note further that there is no guarantee in both cases that the specified copula is indeed the true one. If not the asymptotic variance should be modified adequately (see Cebrian et al. [6] for inference under misspecified copulas). Besides standard inference for independent and identically distributed (i.i.d.) data does not hold with time-dependent data (see Chen and Fan [7] for inference with β-mixing processes). This means that test statistics delivered by standard maximum likelihood routines under the usual assumption of i.i.d. data and/or the assumption of a well specified modeling should be handled with care.

In Table 1 we gather the results of a small Monte Carlo study designed to assess the potential impact of misspecified margins on the estimation of the copula parameter. The assessment relies on measuring estimation performance in terms of Bias and Mean Square Error (MSE). The number of Monte Carlo experiments is set equal to 1000, and the sample size n equal to 200, 500 and 1000, respectively. The true model corresponds to a bivariate Frank copula [9] and two Student margins. The parameter θ of the copula is set equal to one (Kendall's $\rho = 0.1645$, Kendall's $\tau = 0.1100$) and two ($\rho = 0.3168$, $\tau = 0.2139$), while the number ν of degrees of freedom on the Student is set equal to three for both margins. The pseudo model for the margins is chosen to be Gaussian $N(\mu, \sigma)$. This means that we assume the margins to be normal instead of Student to get the misspecified model. We compare the results delivered by a full (one-step) ML procedure and a two-step ML procedure. In the first case the likelihood is simultaneously maximized over the copula parameter and the margin parameters. In the second case the margin parameters are estimated in a first step by optimizing separately the two Gaussian marginal likelihoods. The second step consists then of optimizing the concentrated likelihood over the copula parameter only. We also report the results yielded by the semiparametric method which is by construction well specified. Finally we run the full ML procedure on the true model (likelihood with Student margins) to get the best theoretical benchmark, at least asymptotically.

[9] This copula is often used in actuarial and financial applications and permits quick simulations (Genest [30]).

The results of Table 1 show that the misspecification may lead to a severe bias. The bias is positive and leads to an overestimation of the dependence in the data. This bias seems to be more pronounced in the one step case. MSEs are also much higher when compared to the two well-specified cases. Note further that MSEs of the semiparametric method are close to the ones of the one-step parametric method when the model is true, and the efficiency loss is small in large samples. This suggests that if one has any doubt about the correct modeling of the margins, there is probably little to loose but lots to gain from shifting towards a semiparametric approach.

TABLE 1: Bias and MSE of copula parameter estimators

	Sample size: $n = 200$			
$\theta = 1$	one-step	two-step	semi	true
Bias	0.7012	0.6094	-0.0206	-0.0164
MSE	1.5119	1.0591	0.1754	0.1797
$\theta = 2$	one-step	two-step	semi	true
Bias	1.1144	0.9292	-0.0202	-0.0124
MSE	2.2869	1.4851	0.1913	0.1928
	Sample size: $n = 500$			
$\theta = 1$	one-step	two-step	semi	true
Bias	0.7720	0.6931	-0.0081	-0.0067
MSE	1.1114	0.8184	0.0673	0.0684
$\theta = 2$	one-step	two-step	semi	true
Bias	1.2165	1.0354	-0.0067	-0.0045
MSE	2.0494	1.3977	0.0730	0.0729
	Sample size: $n = 1000$			
$\theta = 1$	one-step	two-step	semi	true
Bias	0.7904	0.7190	-0.0046	-0.0048
MSE	0.9647	0.7397	0.0360	0.0360
$\theta = 2$	one-step	two-step	semi	true
Bias	1.2553	1.0784	-0.0037	-0.0037
MSE	1.9702	1.3853	0.0387	0.0382

3 How to choose copulas empirically?

One of the main issue with copulas is to choose the "best" one, namely the copula that provides the best fit with the data set at hand. The choice among possible copula specifications can be done rigorously via so-called goodness-of-fit (GOF) tests [10]. With the previous notations, usually, a GOF test for multivariate distributions tries to distinguish between two assumptions :

$$\mathcal{H}_0 : F = F_0, \text{ against } \mathcal{H}_a : F \neq F_0, \text{ when the null hypothesis is simple, or}$$

$$\mathcal{H}_0 : F \in \mathcal{F}, \text{ against } \mathcal{H}_a : F \notin \mathcal{F}, \text{ when the null hypothesis is composite.}$$

Here, F_0 denotes some known cumulative distribution function, and $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ is some known parametric family of multivariate cumulative distribution functions.

Let us first work with $d = 1$. The problem is then relatively simple. By considering the transformation of X_1 by F_1, the corresponding empirical process tends weakly to a uniform Brownian bridge under the null hypothesis. Then, a lot of well-known distribution-free GOF statistics are available : Kolmogorov-Smirnov, Anderson-Darling...

In a multidimensional framework, it is more difficult to build distribution-free GOF tests, particularly because the previous transformation $F(\mathbf{X})$ fails to work. More precisely, the law of the transformed variable is no longer distribution-free. Thus, several authors have proposed some more or less satisfying solutions, see e.g. Justel et al. [39], Saunders and Laud [54], Foutz [27], Polonik [49], and Khmaladze ([40, 41] and especially [42]).

Actually, the simplest way to build GOF composite tests for multivariate random variables is to consider multidimensional chi-square tests, as in D'Agostino and Stephens [10] or Pollard [48]. To do this, it is necessary to choose some disjoint subsets $A_1, \ldots, A_p$ in $\mathbb{R}^d$ and to consider

$$\chi^2 = n \sum_{k=1}^{p} \frac{(P_n(\mathbf{X} \in A_k) - P_0(\mathbf{X} \in A_k))^2}{P_0(\mathbf{X} \in A_k)},$$

computed from a sample of size n. Under the null hypothesis, the test statistic χ^2 tends in law towards a chi-square distribution.

[10] An informal test can be done by examining whether the plot of the estimated parametric copula is not too far from the plot of a copula function estimated nonparametrically via kernel methods, see Fermanian and Scaillet [24] for suggestions. One may also rely on a strong a priori for the type of dependence structure one wishes for the data or the application in mind, and choose the type of copulas accordingly.

To deal with the corresponding composite assumption, it is necessary to consider some estimates of θ in a first step. The simplest solution is to build the above chi-square test statistic based on a cumulative distribution function with a parameter estimated by maximum likelihood. Note that the limiting distribution remains chi-square only if the estimation is performed over grouped data [11]. When it is not the case, the limiting distribution of the test statistic χ^2 is bounded above by a chi-square distribution, which means that the use of the latter distribution leads to reject too often the null hypothesis. Unfortunately this aspect is often overlooked by empirical researchers.

A natural idea would be to handle the GOF problem for copulas in a similar way, say to distinguish between two assumptions :

$$\mathcal{H}_0 : C = C_0, \text{ against } \mathcal{H}_a : C \neq C_0, \text{ when the null hypothesis is simple, or}$$

$$\mathcal{H}_0 : C \in \mathcal{C}, \text{ against } \mathcal{H}_a : C \notin \mathcal{C}, \text{ when the null hypothesis is composite.}$$

Here, C_0 denotes some known copula, and $\mathcal{C} = \{C_\theta, \theta \in \Theta\}$ is some known parametric family of copulas. Since the copula is the cumulative distribution function of $(u_1, ..., u_d) = (F_1(X_1), \ldots, F_d(X_d))$, one may think of designing testing procedures as before with C substituted for F and $\mathbf{u} = (u_1, ..., u_d)$ substituted for $\mathbf{X} = (X_1, ..., X_d)$. The difficulty comes from the univariate cumulative distribution functions F_j being unknown. More specifically, the chi-square testing procedure will not work anymore, after replacing the unknown marginal cumulative distribution functions by their empirical counterparts. The limiting law of the previous test statistic χ^2 will not be a chi-square distribution, and thus any inference made as if will be misleading. Actually, the limiting law is a lot more complex, and depends on C_0 and its derivatives [12]. Note that some authors use crude criteria like

$$\mathcal{S}_p = \int |C_n - C_0|^p,$$

for some $p > 0$, where C_n denotes the so-called empirical copula function (see Deheuvels [11, 12, 13] for a definition). In general these statistics provide testing procedures with poor statistical properties since they suffer from the same drawback as the one of the test statistic χ^2, namely their asymptotic laws are not distribution-free.

[11] which often necessitates some ad-hoc estimation procedures.

[12] The rigorous rationale lies in the difference between the behavior of standard empirical processes and copula empirical processes, see Fermanian et al. [23] for details.

For all these reasons, the general problem of GOF test for copulas has not yet been tackled rigourously from an explicit asymptotic point of view. Some authors however suggest to use the bootstrap procedure to gauge the limiting distribution of the test statistic (Andersen et al. [2], e.g.). Genest and Rivest [31] solve the problem for the case of Archimedian copulas. For this type of copulas the problem can be reduced to a univariate one [13], and some standard methods are available. For instance, Frees and Valdez [28] use Q-Q plots to fit the "best" Archimedian copula. None of these authors have dealt with the case of time-dependent copulas [14]. Recently, some authors have applied Rosenblatt's transformation (cf [52]) to the original multivariate series, like in Justel et al. [39], before testing the copula specification: Breymann et al. [5], Chen et al. [8]. Nonetheless, the use of Rosenblatt's transformation is a tedious preliminary step, especially with high dimension variables, and this step is model specific. Thus the test methodology cannot be really viewed as distribution-free.

Note that we could build testing procedures based on some estimates of the joint cumulative distribution function by modeling the marginal distributions simultaneously. This seems to be a good idea, because some "more or less" standard tests are available to check the GOF of the cumulative distribution function itself. However, this does not directly suit our purpose. Indeed, doing so produces tests for the whole specification - the copula and the margins - but not for the dependence structure itself - the copula only-. A slightly different point of view would be to test each marginal separately in a first step. If each marginal model is accepted, then a test of the whole multidimensional distribution can be implemented (through the previously cited methodologies). Yet, such a procedure is heavy, and it is always necessary to deal with a multidimensional GOF test. Moreover, it is interesting to study dependence in depth first, independently of the modeling of the margins. For instance, imagine the copula links the short term interest rate with some credit spreads. It should be useful to keep the possibility to switch from one model of the short term interest rate to another one without affecting the dependence structure vis-à-vis the credit spreads. Since this research area is very active, new term structure models appear regularly, others are forgotten, and current models are often improved. By choosing the copula independently of the marginal models, such evolutions are clearly easier to incorporate, and modeling maintenance is facilitated.

There exists a simple direct way to circumvent the previous difficulties in testing a copula

[13] see the survey of de Matteis [14].

[14] with the exception of Patton [46, 47], but he tests the whole joint specification and not only the copula itself.

specification. This method exploits a smoothing of the empirical copula process, which delivers an estimate of the copula density itself (see Gijbels and Mielniczuc [34]). By definition the kernel estimator of a copula density τ at point $\mathbf{u}$ is

$$\tau_n(\mathbf{u}) = \frac{1}{h^d} \int K\left(\frac{\mathbf{u}-\mathbf{v}}{h}\right) C_n(d\mathbf{v}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{u}-\mathbf{u}_{n,i}}{h}\right), \tag{3.1}$$

where K is a d-dimensional kernel, $h = h(n)$ is a bandwidth sequence and the transformed data $\mathbf{u}_{n,i}$, $i = 1, ..., n$, are obtained from applying the empirical margins to the original data points $\mathbf{X}_i$, $i = 1, ..., n$, component per component [15]. More precisely, we take $\int K = 1$, $h(n) > 0$, and $h(n) \to 0$ when $n \to \infty$. As usual, we denote $K_h(\cdot) = K(\cdot/h)/h^d$. In Fermanian [25], it is proved that:

Theorem 1. *Under some conditions of regularity and $\mathcal{H}_0$, for every m and every vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ in $]0,1[^d$, such that $\tau(\mathbf{u}_k) > 0$ for every k, we have*

$$(nh^d)^{1/2} \left((\tau_n - \tau)(\mathbf{u}_1), \ldots, (\tau_n - \tau)(\mathbf{u}_m)\right) \xrightarrow[T\to\infty]{law} \mathcal{N}(0, \Sigma),$$

where Σ is diagonal, and its k-th diagonal term is $\tau^2(\mathbf{u}_k) \int K^2$.

Now, imagine we want to build a procedure for a GOF test with some composite null hypothesis. Under $\mathcal{H}_0$, the parametric family is $\mathcal{C} = \{C_\theta, \theta \in \Theta\}$. Assume we have estimated θ consistently by $\hat{\theta}$ at the usual parametric rate of convergence, namely

$$\hat{\theta} - \theta_0 = O_P(n^{-1/2}). \tag{3.2}$$

We denote by $\tau(\cdot, \theta_0)$ (or τ in short, when there is no ambiguity) the "true" underlying copula density. Clearly, $\tau(\mathbf{u}, \hat{\theta}) - \tau(\mathbf{u}, \theta_0)$ tends to zero faster than $(\tau_n - \tau)(\mathbf{u})$. Thus, a simple GOF test could be

$$\mathcal{S} = \frac{nh^d}{\int K^2} \sum_{k=1}^{m} \frac{(\tau_n(\mathbf{u}_k) - \tau(\mathbf{u}_k, \hat{\theta}))^2}{\tau(\mathbf{u}_k, \hat{\theta})^2}.$$

Corollary 2. *Under some conditions of regularity, $\mathcal{S}$ tends in law towards a m-dimensional chi-square distribution under the composite null hypothesis $\mathcal{H}_0$.*

The points $(\mathbf{u}_k)_{k=1,\ldots,m}$ are set arbitrarily. They could be chosen in some particular areas of the d-dimensional unit square, where the user seeks a nice fit. For instance, for risk

[15] The vector $\mathbf{u}_{n,i}$ is in fact made of the stack of the ranks componentwise of the i-th observation $\mathbf{X}_i$.

management purposes, it is fruitful to focus on the dependencies in the tails [16]. For the particular copula family $\mathcal{C}$, it is necessary to specify these areas.

It is possible to build another test that does not depend on any particular choice of points. This test is based on the proximity between the smoothed copula density and the estimated parametric density. Under $\mathcal{H}_0$, they will be near each other. To measure such a proximity, we will invoke the L^2-norm. To simplify, denote the estimated parametric $\tau(\cdot, \hat{\theta})$ density by $\hat{\tau}$. Consider the statistic

$$J_n = \int (\tau_n - K_h * \hat{\tau})^2(\mathbf{u})\omega(\mathbf{u})\, d\mathbf{u},$$

where ω is a weight function, viz a measurable function from $[0,1]^d$ towards $\mathbb{R}^+$. Note that we consider the convolution between the kernel K_h and $\hat{\tau}$ instead of $\hat{\tau}$ itself. This trick allows to remove a bias term in the limiting behavior of J_n (see Fan [18]).

The minimization of the criterion J_n is known to produce consistent estimates in numerous situations. These ideas appear first in the seminal paper of Bickel and Rosenblatt [3] for the univariate density with i.i.d. observations. Rosenblatt [53] extends the results in a two-dimensional framework and discusses consistency with respect to several alternatives. Fan [18] deals with various choices of the smoothing parameter. The comparison of some nonparametric estimates-especially estimates of nonparametric regression functions- and their parameter-dependent equivalents has been formalized in a lot of papers in statistics and econometrics : Härdle and Mammen [38], Zheng [59], Fan and Li [19] [17], to name a few.

More recently similar results have been obtained for dependent processes, see e.g. Fan and Ullah [22], Hjellvik et al. [37], Gouriéroux and Tenreiro [35], and Fan and Li [20]. For instance, Aït-Sahalia [1] applies these techniques to find a convenient specification for the dynamics of the short term interest rate. Besides, Gouriéroux and Gagliardini [36] use such a criterion to estimate possibly infinite dimensional parameters of a copula function [18]. Instead of having inference purposes in mind, we will use J_n as a test statistic, like in Fan [18].

Let us assume that we have found $\hat{\theta}$, a convenient estimator of θ. When τ and its derivatives with respect to θ are uniformly bounded on $[0,1]^d \times \mathcal{V}(\theta_0)$, ω can be chosen arbitrarily. Unfortunately, this is not always the case, a leading example being the bivariate

[16] Generally speaking, the tails are related to large values of each margin, so the $\mathbf{u}_k$ should be chosen near the boundaries. Nevertheless, some particular directions could also be of interest (the main diagonal, for instance).

[17] see numerous references in Fan and Li [21]

[18] for instance, the univariate function defining an Archimedian copula

Gaussian copula density. To avoid technical troubles, we reduce the GOF test to a strict subsample of $[0,1]^d$, say ω's support. In Fermanian [25], it is proved that:

Theorem 3. *Under some technical assumptions and $\mathcal{H}_0$,*

$$nh^{d/2}\left(J_n - \frac{1}{nh^d}\int K^2(\mathbf{t}).(\tau\omega)(\mathbf{u}-h\mathbf{t})\,d\mathbf{t}\,d\mathbf{u} + \frac{1}{nh}\int \tau^2\omega.\sum_{r=1}^{d}\int K_r^2\right) \underset{n\to\infty}{\overset{law}{\longrightarrow}} \mathcal{N}(0,2\sigma^2),$$

$$\sigma^2 = \int \tau^2\omega \cdot \int\left\{\int K(\mathbf{u})K(\mathbf{u}+\mathbf{v})\,d\mathbf{u}\right\}^2 d\mathbf{v}.$$

Thus, a test statistic could be

$$\mathcal{T} = \frac{n^2h^d\left(J_n - (nh^d)^{-1}\int K^2(\mathbf{t}).(\hat{\tau}\omega)(\mathbf{u}-h\mathbf{t})\,d\mathbf{t}\,d\mathbf{u} + (nh)^{-1}\int \hat{\tau}^2\omega.\sum_{r=1}^{d}\int K_r^2\right)^2}{2\int \hat{\tau}^2\omega\cdot\int\{\int K(\mathbf{u})K(\mathbf{u}+\mathbf{v})\,d\mathbf{u}\}^2\,d\mathbf{v}}.$$

Corollary 4. *Under the assumptions of Theorem 3, the test statistic $\mathcal{T}$ tends in law towards a chi-square distribution.*

In the next lines we develop a small Monte Carlo study, where we compare three informal, but simple, methods to choose a parametric family of copulas by exploiting the aforementioned GOF criteria.

Let us consider $\mathcal{C}_j = \{C_{\theta_j}^{(j)}, \theta_j \in \Theta_j \subset \mathbb{R}^{q_j}\}$, $j = 1,\ldots,m$, where each θ_j is estimated consistently by $\hat{\theta}_j$.

The first "crude" rule is based on a choice of the family (and its associated parameter $\hat{\theta}_j$) such that

$$\int |C_{\hat{\theta}_j}^{(j)} - C_n|^p$$

is minimal for some $p > 0$. Here, we take a quadratic distance: $p = 2$.

The second method is based on the "naive" statistics

$$\mathcal{S}_0(\hat{\theta}_j) = n\sum_{k=1}^{p}\frac{(C_{\hat{\theta}_j}^{(j)} - C_n)^2(A_k)}{C_n(A_k)}$$

for some disjoint subsets $A_1,\ldots,A_d \subset [0,1]^d$. Again we use $\mathcal{S}_0$ as a comparative measure, and we choose the family that provides the smallest $\mathcal{S}_0$. As already mentioned such statistics do not tend to a chi-square distribution.

The third method relies on the two proper goodness-of-fit statistics $\mathcal{S}$ and $\mathcal{T}$. As in the two first methods, these statistics are used as comparative measures to discriminate between the parametric families of copulas.

TABLE 2: Percentage of choices for the three different families under the first rule. (The true copula is a Student copula with parameters given in the first column.)

(ρ, ν)	Student (in %)	Gaussian (in %)	Clayton (in %)
(0.2,3)	51	43	6
(0.2,8)	0	96	4
(0.2,13)	0	93	7
(0.5,3)	7	93	0
(0.5,8)	0	100	0
(0.5,13)	0	100	0
(0.8,3)	3	97	0
(0.8,8)	0	100	0
(0.8,13)	0	100	0

For this simulation study, we consider 100 samples of 1000 realizations of some bivariate random variables. The true copula is a Student copula whose parameters are $\nu \in \{3, 8, 13\}$ (degrees of freedom) and $\rho \in \{0.2, 0.5, 0.8\}$ (correlation). The subsets A_k are of the type $[i/10, (i+1)/10[\times[j/10, (j+1)/10[$ for $i, j = 0, \ldots, 9$. For $\mathcal{S}$, we have chosen the points $(i/10, j/10)$, $i, j = 1, \ldots, 9$.

We introduce two other alternative families of copulas: the Gaussian family and the Clayton family. It has been noted that these three families provide similar prices for credit derivatives (Laurent and Gregory [44]). However, their behavior in the tails are not the same, and a similarity in terms of expectations does not guarantee a suitable identification over the whole space $[0, 1]^2$. Tables 2, 3 and 4 provide the frequency each family has been chosen by the three rules when the true copula is the Student copula.

Clearly, the first "crude" rule is misleading. The Gaussian copula is almost always chosen, even when the tail dependence is strong (small ν). The results under the second rule based on a"naive"chi-square test statistics are less dramatic. In particular, it provides good predictions for the smallest ν and/or strong correlations. However, the lowest the difference between the true underlying Student copula and Gaussian copulas, the less relevant the diagnostics. On the contrary, a classification based on $\mathcal{S}$ or $\mathcal{T}$ looks right for 3 cases over 4 in general. We have checked that these results are robust for a wide range of parameter values, and this strongly supports the use of these two statistics to discriminate between parametric families

TABLE 3: Percentage of choices for the three different families under the second rule. (The true copula is a Student copula with parameters given in the first column.)

(ρ, ν)	Student (in %)	Gaussian (in %)	Clayton (in %)
(0.2,3)	100	0	0
(0.2,8)	12	76	12
(0.2,13)	7	82	11
(0.5,3)	100	0	0
(0.5,8)	35	65	0
(0.5,13)	10	90	0
(0.8,3)	100	0	0
(0.8,8)	68	32	0
(0.8,13)	27	73	0

TABLE 4: Percentage of choices for the three different families under the third rule with $\mathcal{S}$ (resp. $\mathcal{T}$). (The true copula is a Student copula with parameters given in the first column.)

(ρ, ν)	Student (in %)	Gaussian (in %)	Clayton (in %)
(0.2,3)	92 (97)	0 (2)	8 (1)
(0.2,8)	75 (64)	0 (23)	25 (13)
(0.2,13)	73 (50)	0 (29)	27 (21)
(0.5,3)	96 (100)	0 (0)	4 (0)
(0.5,8)	88 (73)	0 (27)	12 (0)
(0.5,13)	78 (63)	0 (35)	22 (2)
(0.8,3)	97 (100)	0 (0)	3 (0)
(0.8,8)	89 (70)	0 (30)	11 (0)
(0.8,13)	89 (63)	0 (37)	11 (0)

of copulas.

4 How to design time-dependent copulas?

To fix the ideas, consider a security whose payoff at some maturity date T depends on the value $S_T = (S_{1,T}, \ldots, S_{d,T})$ of d underlying assets. Then, it is well-known that its price today can be written as

$$P_0 = e^{-rT} E^Q \left[\psi(S_T)|S_0\right], \tag{4.1}$$

where Q denotes the risk-neutral probability, ψ the payoff function and r the (assumed constant) interest rate. To calculate this price, it is necessary to specify the distribution of S_T knowing S_0 under Q. Since S_T is d-dimensional, it seems to be natural to introduce $C_{0,T}(\cdot|S_0)$, the copula associated with the latter multivariate distribution. Note that this copula is time-dependent and is a function of the current value S_0. Moreover, for market and credit risk measures, it is often necessary to simulate the future values of a lot of market factors, say S, at various dates $t > 0$. Knowing the spot value S_0 today, some future realizations of S_t knowing S_0 have to be drawn (sometimes a large number of times). In that respect, the dependence function $C_{0,t}(\cdot|S_0)$ needs to be specified. If we assume a stationary process (S_t), the two previous issues can be subsumed by: how to define $C_{0,t}(\cdot|S_0)$ for every t and S_0?

First, let us fix the horizon t. A first attempt to formalize so-called conditional copulas is due to Patton [46]. His definition is a direct extension of the usual Sklar's Theorem:

Definition 5. *For every sub-algebra $\mathcal{A}$, the conditional copula with respect to $\mathcal{A}$ associated with $\mathbf{X}$ is a random function $C(\cdot|\mathcal{A}) : [0,1]^d \longrightarrow [0,1]$ such that*

$$F(\mathbf{x}|\mathcal{A}) = C(F_1(x_1|\mathcal{A}), \ldots, F_d(x_d|\mathcal{A})|\mathcal{A}), \quad a.e.$$

for every $\mathbf{x} \in \mathbb{R}^d$. Such a function is unique on the product of the values taken by the conditional marginal cumulative distribution functions $F_j(\cdot|\mathcal{A})$.

Even if this definition is useful it should be noticed that other concepts could be more relevant from a practical point of view. Indeed, in practice, the marginal distributions are usually defined with respect to past marginal values, for instance in the case of Markovian processes $(S_{j,t})$ $j = 1, ..., d$. In other words, we work (often easily) with the conditional distributions of $S_{j,t}$ knowing $S_{j,0}$, but not the conditional distributions of $S_{j,t}$ knowing the

full vector S_0. Actually, it is far simpler to model the future returns of the stock index S&P 500 knowing the past values of this index (by some ARCH, GARCH, stochastic volatility, switching regimes models...) than knowing the past values of the S&P 500 and Nikkei indices both. Since users often like to take their marginal models as inputs for multivariate models, it appears to be preferable to define a notion of conditional pseudo-copula (Fermanian and Wegkamp (2004)). To this goal, consider some sub-algebras $\mathcal{A}_1, \ldots, \mathcal{A}_d, \mathcal{B}$ and denote $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_d)$. These sub-algebras cannot be chosen arbitrarily:

Assumption S. Let $\mathbf{x}$ and $\tilde{\mathbf{x}}$ be some d-vectors. For almost every $\omega \in \Omega$,$\mathbb{P}(X_j \leq x_j|\mathcal{A}_j)(\omega) = \mathbb{P}(X_j \leq \tilde{x}_j|\mathcal{A}_j)(\omega)$ for every $j = 1, \ldots, d$ implies

$$\mathbb{P}(\mathbf{X} \leq \mathbf{x}|\mathcal{B})(\omega) = \mathbb{P}(\mathbf{X} \leq \tilde{\mathbf{x}}|\mathcal{B})(\omega).$$

We will assume condition (S) is satisfied. This is the case in particular when all conditional cumulative distribution functions of $X_1, \ldots, X_d$ are strictly increasing. It is also satisfied when $\mathcal{A}_1 = \ldots = \mathcal{A}_d = \mathcal{B}$, as in Patton [46].

Definition 6. *The conditional pseudo-copula with respect to these sub-algebras and associated with* $\mathbf{X}$ *is a random function* $C(\cdot|\mathcal{A}, \mathcal{B}) : [0,1]^d \longrightarrow [0,1]$ *such that*

$$F(\mathbf{x}|\mathcal{B}) = C(F_1(x_1|\mathcal{A}_1), \ldots, F_d(x_d|\mathcal{A}_d)|\mathcal{A}, \mathcal{B}), \quad a.e.$$

for every $\mathbf{x} \in \mathbb{R}^d$. *Such a function is unique on the product of the values taken by the conditional marginal cumulative distribution functions* $F_j(\cdot|\mathcal{A}_j)$.

The (random) function $C(\cdot|\mathcal{A}, \mathcal{B})$ is called a pseudo-copula because it satisfies all the axioms yielding a copula, except for one: its margins are not uniform in general. This means that the theoretical results and the statistical inference developed for standard copulas will not apply *per se* to pseudo copulas in a straightforward manner. Some ad-hoc theories need to be formalized.

To avoid such annoying conditioning procedures, there is another solution that seems to be simpler. It suffices to increase the dimension of the underlying process and to use the stationary marginal distributions $F_1, \ldots, F_d$ instead of $F_j(\cdot|\mathcal{A}_j)$. For instance, in a one-order Markov process (S_t), we could consider the $2d$-dimensional random vector (S_t, S_{t-1}). By denoting its (true) copula by C, we have

$$P(S_t \leq \mathbf{x}, S_{t-1} \leq \mathbf{y}) = C(F_1(x_1), \ldots, F_d(x_d), F_1(y_1), \ldots, F_d(y_d)),$$

for every $\mathbf{x}$ and $\mathbf{y}$. Thus, the knowledge of C and the univariate cumulative distribution functions F_j, $j = 1, .., d$, is sufficient to describe the whole dynamics of the process (S_t). However, such an approach is less rich than the previous one. Since the stationary marginal distributions provide only a small piece of information, most of the specification is induced by the $2d$-dimensional copula. In this sense, this copula is akin to the joint distribution of (S_t, S_{t-1}) itself.

References

[1] Y. Aït-Sahalia, Testing continuous-time models of the spot interest rate, Review Fin. Studies 9 (1996) 385 – 426.

[2] P.K. Andersen, C. Ekstrøm, J.P. Klein, Y. Shu and M.J. Zhang, Testing the Goodness-of-fit of a copula based on right censored data, mimeo (2002) available on the web.

[3] P. Bickel and M. Rosenblatt, On some global measures of the deviation of density function estimates, Ann. Statist. 1 (1973) 1071 – 1095.

[4] E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli, Copulas for finance. A reading guide and some applications, GRO, Crédit Lyonnais, 2000. Available on the web.

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

[6] A. Cebrian, M. Denuit and O. Scaillet, Testing for Concordance Ordering, forthcoming in Astin Bulletin (2003).

[7] X. Chen and Y. Fan, Estimation of Copula-based semiparametric time series models, mimeo (2002).

[8] X. Chen, Y. Fan and A. Patton, Simple tests for models of dependence between financial time series: with applications to US equity returns and exchange rates, mimeo (2003).

[9] U. Cherubini and E. Luciano, Value at Risk trade-off and capital allocation with copulas, Economic notes 30, (2) (2000).

[10] R. B. D'Agostino and M. A. Stephens, Goodness-of-fit techniques, Marcel Dekker, 1986.

[11] P. Deheuvels, La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. Acad. R. Belg., Bull. Cl. Sci., 5 Série 65 (1979) 274-292.

[12] P. Deheuvels, A Kolmogorov-Smirnov type test for independence and multivariate samples, Rev. Roum. Math. Pures et Appl., Tome XXVI, 2 (1981a) 213-226.

[13] P. Deheuvels, A Nonparametric Test of Independence, Publications de l'ISUP 26 (1981b) 29-50.

[14] R. de Matteis, Fitting copulas to data, Diploma thesis, Univ. Zurich (2001).

[15] M. Denuit and O. Scaillet, Nonparametric Tests for Positive Quadrant Dependence, DP FAME (2003).

[16] P. Embrechts, A. McNeil A. and D. Straumann, Correlation and Dependency in Risk Management: Properties and Pitfalls, in Risk Management: Value at Risk and Beyond, eds Dempster M. and Moffatt H., Cambridge University Press, Cambridge (2000).

[17] P. Embrechts, F. Lindskog, and A. McNeil, Modelling dependence with copulas and applications to risk management. Working paper ETHZ (2001).

[18] Y. Fan, Testing the goodness of fit of a parametric density function by kernel method. Econometric Theory 10 (1994) $316 - 356$.

[19] Y. Fan and Q. Li, A general nonparametric model specification test, Manuscript, Departments of Economics, University of Windsor (1992).

[20] Y. Fan and Q. Li, Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing, J. Nonparametric Stat. 10 (1999) $245 - 271$.

[21] Y. Fan and Q. Li, Consistent model specification tests, Econom. Theory 16 (2000) $1016 - 1041$.

[22] Y. Fan and A. Ullah, On goodness-of-fit tests for weakly dependent processes using kernel method, J. Nonparametric Statist. 11 (1999) $337 - 360$.

[23] J-D. Fermanian, D. Radulovic, and M. Wegkamp, Weak convergence of empirical copula process, forthcoming in Bernoulli.

[24] J-D. Fermanian and O. Scaillet, Nonparametric estimation of copulas for time series, Journal of Risk 5, Vol 4 (2002) 25-54.

[25] J.-D. Fermanian, Goodness-of-fit tests for copulas. Working paper Crest (2003).

[26] J.-D. Fermanian, and M. Wegkamp, Time-dependent copulas, Mimeo (2004)..

[27] R. Foutz, A test for goodness-of-fit based on an empirical probability measure, Ann. Statist., 8 (1980) 989 − 1001.

[28] E. Frees and E. Valdez, Understanding Relationships Using Copulas, North American Actuarial J., 2 (1998) 1 − 25.

[29] R. Frey and A. McNeil, Modelling dependent defaults, working paper (2001).

[30] C. Genest, Frank's family of bivariate distributions, Biometrika 74 (1987) 549 − 555.

[31] C. Genest and C. Rivest, Statistical inference procedures for bivariate Archimedian copulas, J. Amer. Statist. Ass. 88 (1993) 1034 − 1043.

[32] C. Genest, K. Ghoudi and C. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika 82 (1995) 543−552.

[33] C. Genest, R.W.J. van den Goorbergh and B. Werker, Multivariate option pricing using dynamic copula models, mimeo (2003).

[34] I. Gijbels and J. Mielniczuc, Estimating the density of a copula function, Comm. Stat. Theory Methods 19 No 2 (1990) 445 − 464.

[35] C. Gouriéroux and C. Tenreiro, Local power properties of kernel based goodness-of-fit tests, J. Multivariate Anal. 78 (2000) 161 − 190.

[36] C. Gouriéroux and P. Gagliardini, Constrained nonparametric copulas, Working paper CREST (2002).

[37] V. Hjellvik, Q. Yao and V. Tjøstheim, Linearity testing using local polynomial approximation. J. Statist. Plann. Inf. 68(1998) 295 − 321.

[38] W. Härdle and E. Mammen, Comparing nonparametric versus parametric regression fits, Ann. Statist., 21 (1993) 1926 − 1947.

[39] A. Justel, D. Pena and R. Zamar, A multivariate Kolmogorov-Smirnov test of goodness of fit, Statist. Prob. Letters 35 (1997) 251 – 259.

[40] E.V. Khmaladze, Martingale approach to the theory of goodness of fit tests, Theory Prob. Appl. 26 (1981) 240 – 257.

[41] E.V. Khmaladze, An innovation approach in goodness-of-fit tests in $\mathbb{R}^m$, Ann. Statist., 16 (1988) 1503 – 1516.

[42] E.V. Khmaladze, Goodness of fit problem and scanning innovation martingales, Ann. Statist. 21 (1993) 798 – 829.

[43] J-P. Laurent and J. Gregory, Basket default swaps, CDO's and factor copulas, mimeo (2002).

[44] J-P. Laurent and J. Gregory, Choice of copula and pricing of credit derivatives, mimeo (2004).

[45] D. Li, On default correlation : A copula function approach. Journal of Fixed Income 9 (2000) 43-54.

[46] A. Patton, Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula, UCSD WP 2001-09 (2001a).

[47] A. Patton, Estimation of Copula Models for Time Series of Possibly Different Lengths, UCSD WP 2001-17 (2001b).

[48] D. Pollard, General Chi-square goodness-of-fit tests with data-dependent cells, Z. Wahrsch. verw. Gebiete, 50 (1979) 317 – 331.

[49] W. Polonik, Concentration and goodness-of-fit in higher dimensions: (Asymptotically) distribution-free methods, Ann. Stat, 27 (1999) 1210-1229.

[50] M. Rockinger and E. Jondeau, Conditional Dependency of Financial Series: An Application of Copulas. HEC Paris DP 723 (2001).

[51] J. Rosenberg, Nonparametric pricing of multivariate contingent claims, NYU, Stern School of Business (2001).

[52] M. Rosenblatt, Remarks on a Multivariate Transformation, Ann. Statist., 23 (1952) 470-472.

[53] M. Rosenblatt, Quadratic measure of deviation of two dimensional density estimates and a test of independence Ann. Statist. 3 (1975) $1-14$.

[54] R. Saunders and P. Laud, The multidimensional Kolmogorov goodness-of-fit test, Biometrika 67 (1980) 237.

[55] P. Schönbucher, Factor models for portfolio credit risk, Univ. Bon, discussion paper 16/2001 (2001).

[56] P. Schönbucher and D. Schubert, Copula dependent default risk in intensity models, Univ. Bonn, mimeo (2001).

[57] J. Shi and T. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics, 51 (1995) $1384-1399$.

[58] A. Sklar, Fonctions de Répartition à n Dimensions et leurs Marges, *Publ. Inst. Stat. Univ. Paris*, **8**, (1959) 229-231.

[59] J. Zheng, A consistent test of functional form via nonparametric estimation technique, J. Econometrics 75 (1996) $263-289$.