"Building conditionally dependent parametric one-factor copulas"

Mazo, Gildas ; Uyttendaele, Nathan

Abstract

So far, one-factor copulas induce conditional independence with respect to a latent factor. In this paper, we extend one-factor copulas to conditionally dependent models. This is achieved through two representations which allow to build new parametric one-factor copulas with a varying conditional dependence structure. Moreover, the latent factor's distribution can be estimated despite it being unobserved. In order to dis- tinguish between conditionally independent and conditionally dependent one-factor copulas, we provide with a novel statistical test which does not assume any parametric form for the conditional dependence structure. Illustrations of the approach are provided through examples, numerical experiments as well as a real data analysis where we capture the intrinsic state of a financial market and the dependence structure of its individual assets.

Document type : *Document de travail (Working Paper)*

Référence bibliographique

Mazo, Gildas ; Uyttendaele, Nathan. *Building conditionally dependent parametric one-factor copulas.* ISBA Discussion Paper ; 2016/04 (2016) 25 pages

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2016/04

Building conditionally dependent parametric one-factor copulas

MAZO, G. and N. UYTTENDAELE

Building conditionally dependent parametric one-factor copulas

Gildas Mazo and Nathan Uyttendaele

Université Catholique de Louvain
Belgium

Abstract

So far, one-factor copulas induce conditional independence with respect to a latent factor. In this paper, we extend one-factor copulas to conditionally dependent models. This is achieved through two representations which allow to build new parametric one-factor copulas with a varying conditional dependence structure. Moreover, the latent factor's distribution can be estimated despite it being unobserved. In order to distinguish between conditionally independent and conditionally dependent one-factor copulas, we provide with a novel statistical test which does not assume any parametric form for the conditional dependence structure. Illustrations of the approach are provided through examples, numerical experiments as well as a real data analysis where we capture the intrinsic state of a financial market and the dependence structure of its individual assets.

Keywords: conditional, factor, copula, latent, independence, test.

1 Introduction

Nowadays, *factor copulas* [9, 11, 12, 15] refer to those copulas which can be expressed by means of unobserved variables, the factors. Often, only one univariate factor, denoted by X_0, is invoked, and thus one talks about one-factor copulas. In the rest of this paper, $(U_1, \ldots, U_d)$ denotes the vector of interest, with uniform margins, whose joint distribution is a copula.

When it comes to build parametric models, the scope of current one-factor copulas is still limited. First, the possibility of considering a factor other than uniformly distributed not allowed. Yet, in applications, the identification of a factor may implicitely assume estimating its distribution, which may be seen as a parameter of interest. Second, studying the factor's impact on the dependence structure is not allowed, too. Indeed, in current one-factor copulas, only conditional independence — that is, the variables $U_1, \ldots, U_d$ are independent conditionally on the factor $X_0 = x_0$ — are permitted. This means that, for all $u_1, \ldots, u_d \in [0, 1]$,

$$P(U_1 \leq u_1, \ldots, U_d \leq u_d | X_0 = x_0) = \prod_{j=1}^{d} P(U_j \leq u_j | X_0 = x_0).$$

As a result, current one-factor copulas write [11]

$$C(u_1, \ldots, u_d) = \int_0^1 \prod_{j=1}^{d} C_{j|0}(u_j | u_0) \, du_0, \tag{1}$$

where the notations are to be understood as $C_{j|0}(u_j|u_0) = \partial C_{0j}(u_0, u_j)/\partial u_0 = P(U_j \leq u_j | U_0 = u_0)$. Therefore, the task of modeling only amounts to choose parametric forms for the $P(U_j \leq u_j | X_0 = x_0)$. What if the practitioner, after the identification of one factor, assumes that the dependence grows with the factor's value? Or, what if the dependence structure remains the same, but is not conditional independence?

This paper is an attempt to overcome these limitations. It introduces two most general representations for one-factor copulas to extend further the parametric models which can be built. These representations being most general, they cover all the models of the literature, as seen in Section 2. Section 3 addresses data generation, estimation, and also proposes a novel test to assess whether conditional independence may hold or not, without assuming any parametric form for the dependence structure. Section 4 presents the numerical experiments used to illustrate our testing procedure as well as a real data analysis.

2 Two useful representations to extend one-factor copulas

This section introduces two representations to build new parametric families of one-factor copulas, which can be grouped into three different categories. It is shown that many standard copulas of the literature can be recovered. Tail dependence questions are also addressed.

2.1 The representations

Consider the *the law of total probability,*

$$\begin{aligned} C(u_1,\ldots,u_d) &= P(U_1 \le u_1,\ldots,U_d \le u_d) \\ &= \int P(U_1 \le u_1,\ldots,U_d \le u_d | X_0 = x_0) f_0(x_0)\, dx_0, \end{aligned} \tag{2}$$

(the integral is taken over the support of X_0 and f_0 denotes its density), from which originated the formula of current one-factor copulas, given by (1). One easily sees that one-factor copulas are a reformulation of the law of total probability in which the factor X_0 is uniformly distributed on $[0,1]$ (hence the change of notation $U_0 = X_0$) and the variables $U_1,\ldots,U_d$ are assumed to be independent conditionally on the factor $U_0 = u_0$.

To extend one-factor copulas, in addition to let the density of X_0, f_0, be unspecified, we propose to reconsider the decomposition of $P(U_1 \le u_1,\ldots,U_d \le u_d | X_0 = x_0)$ in (2). Fix x_0. Given $X_0 = x_0$, certainly the vector $(U_1,\ldots,U_d)$ has a distribution function, but it is not, in general, a copula, because $U_j | X_0 = x_0$ is not, in general, uniformly distributed. By Sklar's theorem [19, 22], $P(U_1 \le u_1,\ldots,U_d \le u_d | X_0 = x_0)$ can be decomposed as a copula and marginal distributions, as

$$\begin{aligned} &P(U_1 \le u_1,\ldots,U_d \le u_d | X_0 = x_0) \\ =&C_{x_0}(P(U_1 \le u_1 | X_0 = x_0),\ldots,P(U_d \le u_d | X_0 = x_0)). \end{aligned} \tag{3}$$

If we let x_0 vary, both the copula C_{x_0} and the margins $P(U_j \le u_j | X_0 = x_0)$, $j = 1,\ldots,d$, will be, in fact, conditional distributions. The following examples illustrate our point.

Example 1. *Consider* (2) *with X_0 following an exponential distribution, as*

$$f_0(x_0) = e^{-x_0}, \qquad x_0 > 0. \tag{4}$$

Moreover, in (3)*, assume that*

$$P(U_j \le u_j | X_0 = x_0) = \int_0^{u_j} \frac{\Gamma(1+x_0)}{\Gamma(x_0)} (1-t)^{x_0 - 1}\, dt,$$

where

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt, \qquad z > 0, \tag{5}$$

is the well known gamma function. Finally, assume that the density of C_{x_0}, c_{x_0}, writes

$$c_{x_0}(u_1,\ldots,u_d) = (det\ R(x_0))^{-1/2} \exp\left[-\frac{1}{2} z^\top ([R(x_0)]^{-1} - I) z \right], \tag{6}$$

where $z = (z_1,\ldots,z_d)$, z_j is the quantile of order u_j of the standard normal distribution, I is the $d \times d$ identity matrix, and

$$R(x_0) = \begin{pmatrix} 1 & & & \\ & \ddots & \beta(x_0) & \\ & \beta(x_0) & \ddots & \\ & & & 1 \end{pmatrix}, \tag{7}$$

where

$$\beta(x_0) = e^{-x_0}.$$

In Example 1, for a fixed x_0, the copula C_{x_0} is a multivariate Gaussian copula with an exchangeable correlation matrix with parameter $\beta(x_0) = e^{-x_0}$. Likewise, the distribution of U_j given $X_0 = x_0$ is a beta distribution with parameters 1 and x_0. By Sklar's theorem, $P(U_j \leq u_j | X_0 = x_0)$ and C_{x_0} can be set independently.

Example 2. *Consider* (2) *with* X_0 *following a Pareto distribution, as*

$$f_0(x_0) = x_0^{-2}, \qquad x_0 > 1. \tag{8}$$

Moreover, in (3)*, assume that*

$$C_{x_0}(u_1, \ldots, u_d) = \exp\left[-\left((-\log u_1)^{x_0} + (-\log u_d)^{x_0}\right)^{1/x_0}\right].$$

In Example 2, for a fixed x_0, the copula C_{x_0} is recognized to be a Gumbel-Hougaard copula with parameter x_0, see e.g. [19] p. 153. The margins $P(U_j \leq u_j | X_0 = x_0)$, $j = 1, \ldots, d$ were not specified.

While examples such as Example 1 and Example 2 could be multiplied endlessly, there is a representation, presented below, which permits to get them all, and build general parametric one-factor copulas quite easily. So, in view of both the law of total probability (2) and the "conditional Sklar's theorem" (3), every one-factor copula writes

$$\begin{aligned} &C(u_1, \ldots, u_d) \\ &= \int C_{x_0}[P(U_1 \leq u_1 | X_0 = x_0), \ldots, P(U_d \leq u_d | X_0 = x_0)] f_0(x_0)\, dx_0, \end{aligned} \tag{9}$$

where, as in Examples 1 and 2, C_{x_0} is to be understood as a collection, running over x_0, of well defined d-variate copulas. The integral is taken over the support of X_0. In representation (9), as well as in Example 1 and Example 2, letting x_0 vary induces a collection of copulas $\{C_{x_0}\}$ which reflects the change in the dependence structure as the factor varies. For instance, in the former example, $C_{x_0} \to \Pi$ (pointwise) as $x_0 \to \infty$, where Π denotes the independence copula, that is, $\Pi(u_1, \ldots, u_d) = u_1 \cdots u_d$ for all $u_1, \ldots, u_d \in [0, 1]$. On the other hand, if $x_0 \to 0$, then $C_{x_0} \to M$, where M denotes the Fréchet-Hoeffding bound for copulas, that is, M represents the complete positive dependence structure, with $M(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d)$ for all $u_1, \ldots, u_d \in [0, 1]$. In sum, as the factor's value varies, the dependence between the variables $X_1, \ldots, X_d$ varies as well, ranging from independence to complete positive dependence. The opposite happens in Example 2. We have that $C_{x_0} \to M$ whenever $x_0 \to 0$ and $C_{x_0} \to \Pi$ whenever $x_0 \to 1$.

Representation (9) can be recast in terms of standard uniform variables only. So, let $Q_0 = F_0^{-1}$ be the inverse of the factor's distribution function F_0. By the change of variables $u_0 = F_0(x_0)$ in (9), we have

$$\begin{aligned} & C(u_1, \ldots, u_d) \\ &= \int C_{x_0}[P(U_1 \le u_1 | U_0 = F_0(x_0)), \ldots, P(U_d \le u_d | U_0 = F_0(x_0))] f_0(x_0)\, dx_0 \\ &= \int_0^1 C_{Q_0(u_0)}[P(U_1 \le u_1 | U_0 = u_0), \ldots, P(U_d \le u_d | U_0 = u_0)]\, du_0 \\ &= \int_0^1 C_{Q_0(u_0)}[C_{1|0}(u_1|u_0), \ldots, C_{d|0}(u_d|u_0)]\, du_0, \end{aligned} \tag{10}$$

where $C_{j|0}(u_j|u_0) = \partial C_{0j}(u_0, u_j)/\partial u_0 = P(U_j \le u_j | U_0 = u_0)$ and $(U_0, U_j) \sim C_{0j}$, $j = 1, \ldots, d$. Examples 1 and 2 can be recast in view of (10).

Example 3 (continuation of Example 1)**.** *From* (4)*, we have* $Q_0(u_0) = -\log(1-u_0)$*, hence, for a fixed* $u_0 \in (0,1)$*,* $C_{Q_0(u_0)}$ *is a multivariate Gaussian copula with correlation matrix given by*

$$R(u_0) = \begin{pmatrix} 1 & & & \\ & \ddots & \beta(u_0) & \\ & \beta(u_0) & \ddots & \\ & & & 1 \end{pmatrix}, \qquad \beta(u_0) = e^{-Q_0(u_0)} = 1 - u_0.$$

Furthermore, since $P(U_j \le u_j | U_0 = u_0) = P(U_j \le u_j | X_0 = Q_0(u_0))$*, we have*

$$C_{j|0}(u_j|u_0) = \int_0^{u_j} \frac{\Gamma(1 + Q_0(u_0))}{\Gamma(Q_0(u_0))} (1-t)^{Q_0(u_0)-1}\, dt,$$

so that the underlying bivariate copula is

$$C_{0j}(u_0, u_j) = \int_0^{u_0} \int_0^{u_j} \frac{\Gamma(1 + Q_0(t))}{\Gamma(Q_0(t))} (1-y)^{Q_0(t)-1}\, dt\, dy,$$

for $j = 1, \ldots, d$*.*

In Example 3, note that $u_0 = 0$ implies $\beta(u_0) = 1$, and thus $C_{Q_0(u_0)}$ is the Fréchet-Hoeffding bound M. Likewise, $u_0 = 1$ implies $\beta(u_0) = 0$ (by continuity), and thus $C_{Q_0(u_0)}$ is the independence copula.

Example 4 (continuation of Example 2)**.** *From* (8)*, we have* $Q_0(u_0) = 1/(1-u_0)$*, hence, for a fixed* $u_0 \in (0,1)$*,*

$$\begin{aligned} & C_{Q_0(u_0)}(u_1, \ldots, u_d) \\ &= \exp\left[-\left((-\log u_1)^{\beta(u_0)} + (-\log u_d)^{\beta(u_0)} \right)^{1/\beta(u_0)} \right], \qquad \beta(u_0) = Q_0(u_0) = \frac{1}{1-u_0}, \end{aligned}$$

that is, $C_{Q_0(u_0)}$ *is a multivariate Gumbel-Hougaard copula with parameter given by* $\beta(u_0) = Q_0(u_0) = 1/(1-u_0)$*.*

In Example 4, $u_0 = 0$ implies $\beta(u_0) = 1$, and thus $C_{Q_0(u_0)}$ is the independence copula. Likewise, $u_0 = 1$ implies $\beta(u_0) = \infty$ (by continuity), and thus $C_{Q_0(u_0)}$ is the Fréchet-Hoeffding bound. In short, we simply replaced the x_0's of Example 1 and Example 2 by $Q_0(u_0)$. Also, note that the vectors $(X_1, \ldots, X_d | X_0 = x_0)$ and $(U_1, \ldots, U_d | X_0 = x_0)$ have the same copula C_{x_0}, while $(U_1, \ldots, U_d | U_0 = u_0)$ has copula $C_{Q_0(u_0)}$.

Mathematically, both representations (9) and (10) are of course equivalent. It is worth stressing that, however, these representations are better not to be taken as plain mathematical results, but rather as a convenient way to generate new parametric one-factor copula models, as was shown in the above examples. The advantage of the representation in (10) is that it involves copulas only and allows an easy comparison with the old versions of the one-factor copulas, given in (1). For example, one sees immediately that they correspond to (10) with $C_{Q_0(u_0)} = \Pi$. But the representation given in (9) is more convenient when one adopts a point of view centered on the factor itself.

Both representations (9) and (10) can be rewritten in terms of densities. Here only the later is given. So, the density of C in (10) is given by

$$c(u_1, \ldots, u_d) = \int_0^1 c_{Q_0(u_0)}\{C_{1|0}(u_1|u_0), \ldots, C_{d|0}(u_d|u_0)\} \prod_{j=1}^{d} c_{0j}(u_0, u_j)\, du_0, \tag{11}$$

where c, c_{0j} and $c_{Q_0(u_0)}$ are the densities corresponding to C, C_{0j} and $C_{Q_0(u_0)}$, respectively.

In the rest of this paper, we will sometimes abuse notation. We shall write C_{u_0} for $C_{Q_0(u_0)}$ and c_{u_0} for $c_{Q_0(u_0)}$, and the notations C_{u_0}, C_{x_0} stand for both the copulas for a fixed x_0 or u_0 *and* for the collection of copulas $\{C_{u_0}\}, \{C_{x_0}\}$, letting x_0 or u_0 run over their respective support. Finally, it is convenient to refer to C_{u_0} or C_{x_0} as the *inner* copula or *conditional* copula, while C or c will be referred to as the *outer* copula.

2.2 Three forms of one-factor copulas

In order to generate new parametric families of one-factor copulas, one can act through 3 ingredients: the bivariate copulas C_{0j}, $j = 1, \ldots, d$, the factor's distribution, represented either by its density f_0 or by its quantile function Q_0, and the set of multivariate copulas $\{C_{x_0}\}$. Depending on the choice for C_{x_0}, three different forms of one-factor copulas can be made. The following example illustrates the method.

Example 5. *Let X_0 follow an exponential distribution with parameter $\lambda > 0$, as*

$$f_0(x_0) = \lambda e^{-\lambda x_0}, \qquad x_0 > 0.$$

For $j = 1, \ldots, d$, let C_{0j} be a Clayton copula so that

$$C_{0j}(u_0, u_j) = [(u_0^{-\alpha_j} + u_j^{-\alpha_j} - 1)]^{-1/\alpha_j} \qquad \alpha_j \geq 0. \tag{12}$$

Finally, let c_{x_0}, the density of C_{x_0}, be as in (6) *where $R(x_0)$ is as in* (7) *and where*

$$\beta(x_0) = e^{-\beta_0 - \beta_1 x_0}, \qquad \beta_0, \beta_1 \geq 0. \tag{13}$$

In Example 5, we have built a parametric model for one-factor copulas which allow for different features. First, the number of parameters, $d+3$, is linear in d, the dimension. While there is no universal rule, this number is seen by many as being about right for moderate to high dimension applications. Second, as will be seen in Section 3, one can estimate λ, the parameter of the factor's distribution, by maximum pseudo-likelihood. But this factor being unobserved, it means that we are able to estimate the distribution of a unobservable variable. Section 4 illustrates this fact. Finally, one can control the growth rate of the dependence structure, relative to the change of the factor's value. Thus, in (13), a decrease in β_1 yields an increase in $\beta(x_0)$, the correlation parameter. In particular, $\beta_1 = 0$ implies that $\beta(x_0) = \exp(-\beta_0)$, and thus the correlation parameter, hence the conditional copula C_{x_0}, does not depend on x_0 anymore: we call this *conditional invariance*, not to be mistaken with conditional independence. This last feature happens when $\beta_0 = \infty$, implying a correlation parameter $\beta(x_0) = 0$.

In sum, there are 3 types of models, different in nature, that can be built. They are summarized next.

Conditional independence. Conditional independent one-factor copulas are those so that, in (9), $C_{x_0} = \Pi$ for all x_0. They correspond exactly to copulas of the form (1), described in [11], and their interpretation is such that, given the factor's value $X_0 = x_0$, the variables $X_1, \ldots, X_d$ are independent. In Example 5, it corresponds to $\beta_0 = \infty$ and β_1 is finite. Let us note that, even in this simple case, the obtained models are quite reasonable and useful, as was demonstrated not only in [11], but also in view of the vast literature about conditional independent models [23]. In Section 3.3, we provide a novel procedure in order to test the assumption of conditional independence.

Conditional invariance. Conditional invariant one-factor copulas are those so that, in (9), $C_{x_0} = C_{x_0'}$ whatever x_0 and x_0' are. That is, there is a conditional dependence structure, but it remains unchanged whatever the factor's value. Example 5 with $\beta_1 = 0$ enters this setting, and in this case β_0 simply controls the strength of the dependence structure.

Conditional noninvariance. Conditional noninvariant one-factor copulas are those which are not conditionally invariant. Note that, *a fortiori*, they are not conditionally independent either. Here, the conditional dependence structure is allowed to change with the factor's value. For example, in Example 1, $\beta(x_0) \to 0$ as $x_0 \to \infty$ and therefore $C_{x_0} \to \Pi$, the independence copula. On the opposite, $\beta(x_0) \to 1$ as $x_0 \to 0$ and thus $C_{x_0} \to M$, the Fréchet-Hoeffding upper bound, characterizing complete positive dependence. In Example 5, it corresponds to $\beta_1 > 0$.

Natural parametric one-factor copulas can be built with the help of Kendall's tau and Spearman's rho. Recall that, given a bivariate copula C, Kendall's tau is a dependence coefficient in $[-1, 1]$ defined by

$$\tau = 4 \int_{[0,1]^2} C(u, v)\, dC(u, v) - 1. \tag{14}$$

A value of $\tau \approx 0$ hints at independence, and $\tau \approx -1$ (respectively $\tau \approx +1$) indicates complete negative (respectively positive) dependence. Example 6 illustrates the procedure.

Example 6. *Let X_0 follow a standard uniform distribution and let C_{x_0} be as*

$$C_{x_0}(u_1,\ldots,u_d) = (u_1^{-\tau^{-1}(x_0)} + \cdots + u_d^{-\tau^{-1}(x_0)} - d + 1)^{-1/\tau^{-1}(x_0)}$$

where τ^{-1} is the inverse map of

$$\tau(\beta) = \frac{\beta}{\beta+2} \in [0,1], \qquad \beta \geq 0. \tag{15}$$

In Example 6, for a fixed x_0, C_{x_0} is recognized to be a Clayton copula with parameter $\tau^{-1}(x_0) = 2x_0/(1-x_0)$ for $x_0 \in [0,1)$. The procedure works as follows. First, choose a parametric family of copulas, here the family of Clayton copulas

$$C_\beta(u_1,\ldots,u_d) = (u_1^{-\beta} + \cdots + u_d^{-\beta} - d + 1)^{-1/\beta}, \qquad \beta \geq 0. \tag{16}$$

Second, compute Kendall's tau (there is only one, since all pairs have the same distribution), given (15). Third, choose the distribution of X_0 so that its support corresponds to the range of the map induced by (15), here $[0,1]$. Fourth and last, replace β by $\tau^{-1}(x_0)$ in (16).

The conditional dependence structure in Example 6 goes from conditional independence to conditional complete dependence. Indeed, when $x_0 \to 0$, $\beta(x_0) \to 0$ and $C_{\beta(x_0)} \to \Pi$. If $x_0 \to 1$ instead, $\beta(x_0) \to \infty$ and $C_{\beta(x_0)} \to M$, the Fréchet-Hoeffding upper bound for copulas. If one rather defines $\beta(x_0) = -\log(x_0)$, then $\beta(x_0) \to \infty$ when $x_0 \to 0$ and $C_{\beta(x_0)} \to M$. Hence, in one case the dependence increases with respect to the factor, while in the other case it decreases.

2.3 Tail dependence properties

Copulas of the form (10) can successfully address tail dependence questions. Let us remember that the lower tail dependence coefficient, for a bivariate vector $(X_j, X_{j'})$ with marginal distribution functions F_j and $F_{j'}$, denoted by $\lambda^L_{jj'}$, is defined by the limit of $P(F_j(X_j) < u | F_{j'}(X_{j'}) < u)$ as $u \to 0$. Likewise, the upper tail dependence coefficient, denoted by $\lambda^U_{jj'}$, is defined as the limit of $P(F_j(X_j) > u | F_{j'}(X_{j'}) > u)$ as $u \to 1$. It is well known that the Gaussian copula, for instance, is such that $\lambda^L_{jj'} = \lambda^U_{jj'} = 0$, provided the absolute value of its correlation coefficient is not equal to one. For a copula to be able to model a phenomenon where the co-occurrence of extreme values in both dimensions is likely to happen, it is reasonable to demand that $\lambda^L_{jj'}$, $\lambda^U_{jj'}$, or both, be positive. This positiveness property holds for copulas of the form (10), as it is shown now.

Proposition 1. *Suppose that the inner copula C_{u_0} converges to some limit copula C_0 (respectively C_1) as $u_0 \to 0$ (respectively $u_0 \to 1$). Assume also that c_0 (respectively c_1), the density of C_0 (respectively C_1), is such that $c_0(u,v) > 0$ (respectively $c_1(u,v) > 0$), for all u, v. If the lower (respectively upper) tail dependence coefficient of C_{0k} is positive for both $k = j$ and $k = j'$, then $\lambda^L_{jj'} > 0$ (respectively $\lambda^U_{jj'} > 0$).*

The above result is an extension of that in [11] (Proposition 5), see also [9], Chapter 3.

2.4 Links to models in the literature

Many well-known copula models in the literature can be recovered from (10), as shown below.

Archimedean copulas. Let ψ be a completely monotonic function on $[0,\infty]$, that is, $(-1)^k d^k/dt^k \psi(t) \geq 0$ for all integers k and all $t > 0$, and such that $\psi(0) = 1$ and $\psi(\infty) = \lim_{t\to\infty} \psi(t) = 0$. If a copula C can be written as $C(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$, then it is called an Archimedean copula with generator ψ [17]. Let us note that the above-mentioned conditions on ψ are sufficient, but not necessary, in order to make sure that C is a proper copula. For sufficient and necessary conditions, see [18].

Proposition 2. *In* (10)*, let* $C_{x_0} = \Pi$*, assume that the support of* X_0 *is* $[0,\infty]$*, and put*

$$C_{0j}(u_0, u_j) = \int_0^{Q_0(u_0)} e^{-t\psi^{-1}(u_j)} f_0(t)\, dt, \ \text{ where } \psi(x) = \int_0^{\infty} e^{-tx} f_0(t)\, dt,$$

$j = 1, \ldots, d$*, with* f_0 *being the derivative of* F_0*. It can be checked that* ψ *is completely monotonic, see for instance [7]. Then* C*, the left-hand side of equation* (10) *or outer copula, is an Archimedean copula with generator* ψ*.*

Let us note that the above result (as well as its proof in the Appendix) is simply a reformulation of Joe's [7].

Nested Archimedean copulas. Archimedean copulas can be nested in order to get more flexible models. Nested Archimedean copulas were introduced by [7] and have been the main topic of many research papers since, see for instance [17], [5], [21], or [20]. The simplest nested Archimedean copula one can think of is one where a bivariate Archimedean copula $C_{12}(u_1, u_2) = \psi_{12}(\psi_{12}^{-1}(u_1) + \psi_{12}^{-1}(u_2))$ is nested into another bivariate Archimedean copula $C_{123}(\bullet, u_3) = \psi_{123}(\psi_{123}^{-1}(\bullet) + \psi_{123}^{-1}(u_3))$ in order to get a copula of the form

$$\begin{aligned} C(u_1, u_2, u_3) &= C_{123}(C_{12}(u_1, u_2), u_3) \\ &= \psi_{123}\big(\psi_{123}^{-1}(\psi_{12}(\psi_{12}^{-1}(u_1) + \psi_{12}^{-1}(u_2))) + \psi_{123}^{-1}(u_3)\big) \quad (17) \end{aligned}$$

In general, an arbitrary pair of generators (ψ_{123}, ψ_{12}) does not ensure the copula in equation (17) will be a proper copula. In this paper, however, we assume this is always the case. The reader can find more information on this matter in [17].

Proposition 3. *Define* ψ_{123} *the same way* ψ *was defined in Proposition 2. Also let* C_{0j} *as in Proposition 2. Further define*

$$C_{x_0}(u, v, w) = \exp\Bigg(- x_0 \times \nu\bigg(\nu^{-1}\Big[\frac{1}{x_0}\log\Big(\frac{1}{u}\Big)\Big] + \nu^{-1}\Big[\frac{1}{x_0}\log\Big(\frac{1}{v}\Big)\Big]\bigg)\Bigg) \times w;$$

where $\nu(\bullet) = \psi_{123}^{-1}\big(\psi_{12}(\bullet)\big)$ *and* $\nu(\bullet)^{-1} = \psi_{12}^{-1}\big(\psi_{123}(\bullet)\big)$ *and* $\psi_{12}(\bullet)$ *is equal to the integral between* 0 *and* ∞ *of* $\exp(-t\bullet)dF_{12}(t)$ *with* F_{12} *an arbitrary distribution function. Then* (10) *is the copula given in* (17)*.*

Gaussian copulas. A Gaussian copula is a copula whose density c satisfies

$$\log c(u_1, \dots, u_d) = -\frac{1}{2}\log(\det(R)) - \frac{1}{2} z^\top (R^{-1} - I) z, \tag{18}$$

where R is a $d \times d$ invertible correlation matrix, $z = (z_1, \dots, z_d)^\top$ and z_j is the quantile of order u_j of the standard normal distribution. A Gaussian copula can be represented as in (10), as given below. Let $\beta_0 = (\beta_{01}, \dots, \beta_{0d})^\top$ be a real vector in $[0,1]^d$ and let D be a diagonal matrix with elements given by $1 - \beta_{0j}^2$, $j = 1, \dots, d$. Finally let C_A be a d-variate Gaussian copula with correlation matrix A.

Proposition 4. *Let $C_{u_0} = C_A$ for each u_0 and let C_{0j} be a bivariate Gaussian copula with correlation β_{0j}, $j = 1, \dots, d$. Then the outer copula in* (10) *is a Gaussian copula with correlation matrix given by $R = D^{1/2} A D^{1/2} + \beta_0 \beta_0^T$.*

C-Vine copulas. Let $(U_0, U_1, \dots, U_d)$ be a random vector following a C-Vine copula distribution truncated after the second level. The density of this truncated C-Vine is given by

$$c(u_0, \dots, u_d) = \prod_{j=1}^{d-1} c^*_{1,1+j|0}(C^*_{1|0}(u_1|u_0), C^*_{1+j|0}(u_{1+j}|u_0)|u_0) \prod_{j=1}^{d} c^*_{0j}(u_0, u_j) \tag{19}$$

where $c^*_{0j} = \partial^2 C^*_{0j}(u_0, u_j)/\partial u_0 \partial u_j$, $C^*_{j|0}(u_j|u_0) = \partial C^*_{0j}(u_0, u_j)/\partial u_0$, $\{C^*_{0j}(u_0, u_j)\}$ being a set of arbitrary bivariate copulas and $\{c^*_{1,1+j|0}\}$ is a set of abritrary copula densities for each u_0. Due to their extreme flexibility and ease of use (one only has to specify sets of bivariate copulas), Vine copulas have been used in an increasing number of applications and are still a hot topic of research, see for instance [1], [13] or [2].

Proposition 5. *If, in* (11)*, for each u_0, c_{u_0} is defined as*

$$c_{u_0}(u_1, \dots, u_d) = \prod_{j=1}^{d-1} c^*_{1,1+j|0}(u_1, u_{1+j}|u_0),$$

and $c_{0j}(u_0, u_j) = c^*_{0j}(u_0, u_j)$ for all j, then the outer copula c in (11) *is the d-variate marginal distribution, with respect to u_0, of* (19)*, that is, its density writes*

$$\begin{aligned} c(u_1, \dots, u_d) &= \int_0^1 c_{u_0}(C^*_{1|0}(u_1|u_0), \dots, C^*_{d|0}(u_d|u_0)) \prod_{j=1}^{d} c^*_{0j}(u_0, u_j) du_0 \\ &= \int_0^1 \prod_{j=1}^{d-1} c^*_{1,1+j|0}(C^*_{1|0}(u_1|u_0), C^*_{1+j|0}(u_{1+j}|u_0)|u_0) \prod_{j=1}^{d} c^*_{0j}(u_0, u_j)\, du_0. \end{aligned}$$

If one assumes that, in (19), none of the elements of $\{c^*_{1,1+j|0}\}$ actually depends on u_0, then the inner copula in Proposition 5 becomes

$$c_{u_0}(u_1, \dots, u_d) = \prod_{j=1}^{d-1} c^*_{1,1+j}(u_1, u_{1+j}),$$

which is nothing more than a C-Vine on $(U_1, \dots, U_d)$, truncated at the first level.

p-factor models. Define respectively Π_1-factor and Π_2-factor copulas as copulas of the form

$$C^{(\Pi_1)}(u_1, \ldots, u_d) = \int_0^1 \prod_{j=1}^d C^{(2)}_{j|0}(u_j|v_2)\, dv_2, \text{ and} \tag{20}$$

$$C^{(\Pi_2)}(u_1, \ldots, u_d) = \int_0^1 \int_0^1 \prod_{j=1}^d C^{(2)}_{j|0}(C^{(1)}_{j|0}(u_j|v_1)|v_2)\, dv_2\, dv_1, \tag{21}$$

where $C^{(k)}_{j|0}(u_j|v) = \partial C^{(k)}_{0j}(v, u_j)/\partial v$ for $k = 1, 2$ and $j = 1, \ldots, d$, and where the $C^{(k)}_{0j}$ are (arbitrary) bivariate copulas. Π_1-factor and Π_2-factor copulas have been studied in [11,12] as copula models for conditionally independent variables given respectively one and two latent factors.

The following (trivial) proposition aims at recovering Π_1-factor and Π_2-factor copulas as special cases of the model (10).

Proposition 6. *Consider the copulas given in* (20) *and* (21)*. In* (10)*, put* $C_{0j} = C^{(2)}_{0j}$*. If, moreover,* $C_{u_0} = \Pi$ *for each* u_0*, then the outer copula* C *in* (10) *is the* Π_1*-factor copula given in* (20)*. Likewise, if, in* (10)*,* $C_{0j} = C^{(1)}_{0j}$ *and moreover,*

$$C_{u_0}(u_1, \ldots, u_d) = \int_0^1 \prod_{j=1}^d C^{(2)}_{j|0}(u_j|\tilde{u}_0)\, d\tilde{u}_0,$$

for each u_0*, then the outer copula* C *in* (10) *is the* Π_2*-factor copula given in* (21)*.*

Note that C_{u_0} in the above proposition actually does not depend on u_0 hence the outer copula C is a conditionally invariant model. This restriction can be easily removed as follows. Let, for each u_0, $\widetilde{C}_{0j}(\bullet, \bullet; u_0)$ be bivariate copulas and

$$C_{u_0}(u_1, \ldots, u_d) = \int_0^1 \prod_{j=1}^d \widetilde{C}_{j|0}(u_j|\tilde{u}_0; u_0)\, d\tilde{u}_0,$$

where $\widetilde{C}_{j|0}(u_j|\tilde{u}_0; u_0) = \partial \widetilde{C}_{0j}(\tilde{u}_0, u_j; u_0)/\partial \tilde{u}_0$. The outer copula is then

$$C(u_1, \ldots, u_d) = \int_0^1 \int_0^1 \prod_{j=1}^d \widetilde{C}_{j|0}(C_{j|0}(u_j|u_0)|\tilde{u}_0; u_0)\, d\tilde{u}_0\, du_0. \tag{22}$$

Admittedly, many copulas have a Π_1-factor or Π_2-factor copula representation (Archimedean copulas, structured Gaussian copulas, etc). Our framework however opens the gate to a potentially even larger number of copulas. For instance, to the best of our knowledge, nested Archimedean copulas do not allow for a Π_p-factor copula representation. They can however be recovered in a nontrivial way from (10), as seen in Proposition 3. Moreover, even if, from a mathematical point of view, our framework would turn out to be equivalent to Π_p-factor copula models, it still yields a different perspective. Moreover, we are able to interpret data in a meaningful way, see for instance Section 4.3, and to easily build d-variate models by tapping into the existing pool of both bivariate and multivariate copulas in the literature.

3 Simulation and inference

This section presents a simulation algorithm and procedures to carry out estimation and testing for conditional independence in copula models of the form (10).

3.1 Simulation

To generate one realization $(u_1, \ldots, u_d)$ of the random vector $(U_1, \ldots, U_d)$ with distribution C given by (10), one takes the d-variate margin of $(u_0, u_1, \ldots, u_d)$, a realization of $(U_0, U_1, \ldots, U_d)$, where U_0 is the latent factor. Remembering that, given $U_0 = u_0$, the distribution of $(U_1, \ldots, U_d)$ can be split into the inner copula C_{u_0} and a set of univariate margins $\{C_{j|0}(\bullet|u_0)\}$, with $C_{j|0}^{-1}(\bullet|u_0)$ denoting the inverse function, $j = 1, \ldots, d$, the following algorithm produces the desired output.

Algorithm 1 Generating one observation from (10).

1: Generate one observation u_0 from a standard uniform random variable.
2: Generate one observation $(u_{01}, \ldots, u_{0d})$ from C_{u_0}.
3: Put $u_j = C_{j|0}^{-1}(u_{0j}|u_0)$ for $j = 1, \ldots, d$.

Let us notice that, in the above algorithm and in the presence of conditional invariance, that is if C_{u_0} does not depend on u_0, step 1 is not required for step 2. Needless to say, in the first step, one could have sampled from F_0, the distribution of X_0, and in the second step, one would have sampled from C_{x_0} instead of C_{u_0}.

3.2 Estimation

In this section, we describe likelihood-based methods to perform estimation in one-factor copulas of the form (10). All copulas are assumed to be absolutely continuous with respect to the Lebesgue measure. Moreover, we assume that the built parametric families of one-factor copulas are identifiable. This may not be the case, but this issue is not bounded to representations (9) and (10). Indeed, as we show in the Discussion section, this issue already arised in [11] for conditional independent one-factor copulas.

Thus, for $j = 1, \ldots, d$, we can write $C_{0j}(u_0, u_j) = C_{0j}(u_0, u_j; \alpha_j)$ and $C_{x_0}(u_1, \ldots, u_d) = C(u_1, \ldots, u_d; \beta(x_0))$, where β is a mapping which, to each x_0 in the support of X_0, associates a parameter in the appropriate parameter space. If the mapping β depends on a vector of parameters, as in (13), we denote this vector also by β. Likewise, we denote the parameter vector which contains the parameters of the quantile function Q_0 of X_0 by λ. Accordingly, the notation for the copula of $(U_1, \ldots, U_d | U_0 = u_0)$ becomes $C_{Q_0(u_0)}(u_1, \ldots, u_d) = C(u_1, \ldots, u_d; \beta, \lambda)$. Finally, let us denote by $(x_{i1}, \ldots, x_{id})$, $i = 1, \ldots, n$, the sample of the distribution F with margins $F_1, \ldots, F_d$ and copula C.

The pseudo log likelihood function to maximize is

$$L_n(\theta) = \sum_{i=1}^{n} \log \int_0^1 c\left[C_{1|0}(\widehat{F}_1(x_{i1})|u_0;\alpha_1),\ldots,C_{d|0}(\widehat{F}_d(x_{id})|u_0;\alpha_d);\beta,\lambda\right] \times \prod_{j=1}^{d} c_{0j}(u_0,\widehat{F}_j(x_{ij});\alpha_j)\,du_0, \quad (23)$$

where θ stands for the complete parameter vector, that is, $\theta = (\alpha,\beta,\lambda)$, $\alpha = (\alpha_1,\ldots,\alpha_d)$ and $\widehat{F}_j$ denotes an estimate of F_j, $j = 1,\ldots,d$. There are many ways to estimate F_j. For instance, $\widehat{F}_j$ may be the empirical distribution function, as in [4], or may be a parametric estimate, as in [10].

Regarding the computational aspects, especially in higher dimensions and for datasets of higher sizes, the likelihood 23 may be costly to compute due to the repeated use of integrals (as many as the sample size). A brief discussion on these computational aspects are given in Section 4.1.

3.3 Testing for conditional independence

This section provides procedures to test for conditional independence in models based on the representation (10). Indeed, being able to assess if the variables of interest are dependent or independent conditioned on the latent factor seems a crucial issue. Conditional independence would mean that the factor captures all the dependence in the data whereas no conditional independence would mean that there is a remaining, intrinsic dependence in the variables even though the factor has been accounted for.

Throughout this section, the bivariate copulas C_{0j}, $j = 1,\ldots,d$, are assumed to belong to some parametric families. The inner copula C_{u_0}, however, is left unspecified: it can be parametric or nonparametric. The possibility to carry out a hypothesis test in this setting, is, to the best of our knowledge, new in the literature.

The hypothesis test for conditional independence is of the form

$$H_0 : C_{u_0} = \Pi \text{ for all } u_0$$
$$\text{versus } H_1 : \text{ there exists some } u_0 \text{ such that } C_{u_0} \neq \Pi,$$

(recall that Π stands for the independence copula) where for two functions f and g, $f = g$ means that $f(t) = g(t)$ for all t in their domain. So are to be understood inequalities.

If a certain parametric form is assumed for C_{u_0}, such as in Section 2.2, then most likely the test will reduce to testing for a parameter to equate a certain value, and no conceptual difficulties refrain the task. For instance, in (13), testing for conditional independence amounts to testing for $\beta_0 = \infty$ or $\beta_1 = \infty$ (conceptually). Let us remark that testing for conditional invariance is feasible in this context: in the above example, for instance, it amounts to testing for $\beta_1 = 0$.

If C_{u_0} is left unspecified, the alternative hypothesis needs to be slightly restricted in order for a test to exist. Consider

$$H_0 : C_{u_0} = \Pi \text{ for all } u_0$$
$$\text{versus } H_1 : C_{u_0} > \Pi \text{ for all } u_0 \quad (24)$$

In plain English, the alternative hypothesis is: "conditioned on the factor, the variables of interest are positively dependent".

Now here is our procedure. Let π be the risk of type I error. One rejects H_0 if $T_n \leq c_\pi$, where c_π is chosen so that $P_{H_0}(T_n \leq c_\pi) = \pi$ and where

$$T_n = \sup_{t\in[0,1]^d} M(t) - \widehat{C}(t), \tag{25}$$

where $M(t) = M(t_1, \ldots, t_d) = \min(t_1, \ldots, t_d)$ is the Fréchet-Hoeffding upper bound for copula and

$$\widehat{C}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\widehat{F}_j(X_{ij}) \leq t_j,\, j = 1, \ldots, d) \tag{26}$$

is the empirical estimator of C, $(X_{i1}, \ldots, X_{id})$, $i = 1, \ldots, n$ being the data and $\widehat{F}_j$ being the empirical distribution function of X_{ij}, $j = 1, \ldots, d$.

The heuristic underlying the expression of T_n is as follows. Denote by $C^{(H_0)}$ the copula under H_0, that is, one substitutes Π for the inner copula C_{u_0} in (10) and gets

$$C^{(H_0)}(u_1, \ldots, u_d) = \int_0^1 \prod_{j=1}^{d} C_{j|0}(u_j|u_0)\, du_0. \tag{27}$$

If H_0 is true, $C_{u_0} = \Pi$, trivially implies that the outer copula C in (10) verifies $C = C^{(H_0)}$. But if H_0 is false, $C_{u_0} > \Pi$ implies $C > C^{(H_0)}$ and thus, in view of (25), T_n should take smaller values. The Fréchet-Hoeffding bound M is used in the definition of T_n in order to ensure positiveness.

In order to estimate the distribution of T_n under H_0, bootstrap is required. Note that under H_0, the outer copula C (27) is fully parametric: one can obtain an estimate $\widehat{C}^{(H_0)}$ by maximum pseudo-likelihood [11]. One then can generate bootstrap samples (say N) in order to get N test statistics $T_n^{(1)}, \ldots, T_n^{(N)}$. These can be used, for instance, to compute a p-value as $P_{H_0}(T_n \leq T_n^{(obs)}) \approx N^{-1}\sum_{k=1}^{N} \mathbf{1}(T_n^{(k)} \leq T_n^{(obs)})$.

Comparing the nonparametric estimator $\widehat{C}$ in (26) to a parametric estimator under H_0, say $\widehat{C}^{(H_0)}_{\text{parametric}}$, for instance by considering Kolmogorov-Smirnov or Cramér-von Mises distances, would have been possible but would have required, because of the bootstrap procedure, the computation of $\widehat{C}^{(H_0)}_{\text{parametric}}$ as many times as they are bootstrap samples, which increases the computational needs.

Finally, let us note that the test $H_0 : C_{u_0} = \Pi$ against $H_1 : C_{u_0} < \Pi$ can be carried out by considering (25) again, but this time with a rejection region on the right, that is, we reject H_0 if $T_n \geq c_\pi$, where c_π is chosen so that $P_{H_0}(T_n \geq c_\pi) = \pi$.

4 Illustrations

The purpose of this section is to illustrate how one can take advantage of the framework presented in Section 2 in practice. We first provide a few technical details on how some numerical operations were performed.

4.1 Computational aspects

In this paper, log-likelihoods are maximized using gradient descent algorithms, which can be found in the `optim` function of the statistical software `R`[1]. These algorithms usually require to provide a starting parameter vector. It is advised to try several such points and retain only the one leading to the best result.

In order to numerically evaluate the integral in (23), we relied on our own implementation of numerical integration Newton-Cotes formulas coupled with Romberg's algorithm (see [14], Section 18 and [6]) in `R/C++` using the package `Rcpp` [3]. Alternatively, we also often used Gauss-Legendre quadrature formulas of the `R` package `gaussquad`. In this last case, the number k of function evaluations needed to compute the approximated integral $I(k)$ was chosen upon visual inspection of the graph of $(k, I(k))$. As a rule of thumb, we chose a value $k = k_0$ such that the quantities $I(k)$, $k \geq k_0$ do not vary much.

4.2 Testing for conditional independence

In this section, we study the power of the test statistic T_n in (25) by means of a simulation experiment. Recall that the power is the probability of rejecting the null hypothesis H_0 under the alternative hypothesis H_1. We considered the test (24) and set the type I error risk to $\pi = 0.1$. We drew $N = 500$ datasets of size $n = 50, 500$ from the model (10), with $d = 3$ and C_{0j} being Clayton copulas as in (12) with parameters α_j, $j = 1, 2, 3$, such that Kendall's τ coefficients are equal to 0.4 for $j = 1$, 0.5 for $j = 2$ and 0.6 for $j = 3$. The inner copula C_{x_0} was a normal copula as in (18) with correlation matrix

$$R = \begin{pmatrix} 1 & & & \\ & \ddots & \beta & \\ & \beta & \ddots & \\ & & & 1 \end{pmatrix}, \tag{28}$$

for $\beta = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5$. (There are $N = 500$ samples of size n for each β and each n). Note that $\beta = 0$ corresponds to the null hypothesis H_0.

For each k-th sample, $k = 1, \ldots, N$, we calculated a p-value $p^{(k)}$ based on 200 boostrap replications. That is, we calculated the proportion of 200 simulated test statistics that where lower or equal than the observed one. As rejection occurs whenever the p-value is lower or equal to the type I error risk π, we approximated the power by the proportion of the $p^{(k)}$ falling below π. See Section 3.3 for details.

Figure 1 shows the estimated power of T_n in (25). As it was expected, the power of the test increases as n and β grow. Furthermore the power is equal to the type I error risk π under the null, that is when $\beta = 0$.

[1] `https://www.r-project.org/`

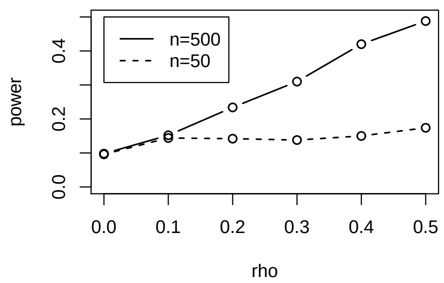

Figure 1: Power of (25) as a function of β.

4.3 Estimating the distribution of a financial market through the dependence of its individual assets

It is commonly assumed that dependence within financial markets is higher in "crisis times" than in "stable times" (see e.g. [24] for a statistical analysis supporting this view). If one wishes to turn this plain English phrase into a statistical model, then certainly the approach developed in this paper would be useful. Indeed, one would let X_0 be the crisis indicator and C_{x_0} account for the dependence in the market as a function of its state — state which would range from "no crisis", represented by the number 0, to "extreme crisis", represented by the number 1.

Once a particular model would have been chosen, many things may be of interest. One may be interested in estimating the latent crisis indicator distribution and study its evolution through time. Or would assess the goodness-of-fit of the model in order to infirm or confirm it, in particular the functional dependence related to the latent factor, or, in other words, how the individual assets respond to the market state.

Data

Our market consists of $d = 10$ arbitrary components[2] of the NASDAQ index. We gathered weekly data from Yahoo! Finance at `http://www.yahoo.com/` between 2005 and 2013. The log-return at the i-th week and k-th year for the j-th component is denoted by $X_{ij}^{(k)} = \log(V_{ij}^{(k)}/V_{i-1,j}^{(k)})$ where the $V_{ij}^{(k)}$ stand for the raw prices. The log-returns are uniformized as $R_{ij}^{(k)}/(n_k + 1)$, where n_k is the number of observations for the k-th year (usually 52) and $R_{ij}^{(k)}$ is the rank of $X_{ij}^{(k)}$ among $X_{1j}^{(k)}, \ldots, X_{nj}^{(k)}$. The latent crisis indicator at the i-th week and k-th year is denoted by $X_{i0}^{(k)}$. For the sake of simplification, we assume that the vectors $(X_{i0}^{(k)}, X_{i1}^{(k)}, \ldots, X_{id}^{(k)})$, $i = 1, \ldots, n$, are independent and identically distributed for each fixed k.

[2] "Alexion", "Apple", "Biogen", "Rob", "Citrix", "Costco", "EA", "Fast", "Garmin" and "Henry"

Models and methods

Our choice for the 3 ingredients required (remember Section 2.2) to build a parametric model are given here. The latent crisis indicator is assumed to be beta distributed, so that it has a flexible distribution over the interval $[0,1]$. So, for each year k,

$$f_0^{(k)}(x_0;\lambda_1^{(k)},\lambda_2^{(k)}) = \frac{\Gamma(\lambda_1^{(k)}+\lambda_2^{(k)})}{\Gamma(\lambda_1^{(k)})\Gamma(\lambda_2^{(k)})} x_0^{\lambda_1^{(k)}-1}(1-x_0)^{\lambda_2^{(k)}-1},$$

where $0 \leq x_0 \leq 1$, $\lambda_1^{(k)},\lambda_2^{(k)} > 0$ and Γ is the gamma function defined in (5). Consequently, the factor's median may also be interpreted as a deterministic crisis indicator: the more the median approaches 1, the more likely we are in a crisis. The copula of $(X_{i0}^{(k)}, X_{ij}^{(k)})$ is assumed to be a Frank copula for all $j = 1,\dots,d$ and all k, that is,

$$C_{0j}^{(k)}(u_0,u_j;\alpha_j^{(k)}) = -\frac{1}{\alpha_j^{(k)}} \log\left(1 + \frac{(e^{-\alpha_j^{(k)}u_0}-1)(e^{-\alpha_j^{(k)}u_j}-1)}{e^{-\alpha_j^{(k)}}-1}\right),$$

where $\alpha_j^{(k)} \neq 0$ and $-\infty < \alpha_j^{(k)} < \infty$ (see e.g. [19] p. 116). The inner copula is a Gaussian copula with an exchangeable correlation matrix, so that c_{x_0} has formula (6) with

$$R(x_0) = \begin{pmatrix} 1 & & & \\ & \ddots & x_0 & \\ & x_0 & \ddots & \\ & & & 1 \end{pmatrix}.$$

In other words, the dependence between the individual assets increases linearly as the crisis becomes more severe.

For each year, there were 12 parameters to estimate: 10 for the 10 bivariate Frank copulas $(\alpha_1^{(k)},\dots,\alpha_d^{(k)})$ and 2 for the latent factor beta distribution $(\lambda_1^{(k)},\lambda_2^{(k)})$. Estimation was performed by pseudo maximum likelihood, as described in Section 3.2.

In order to assess the goodness of fit of our model, we compared the average of the pairwise Kendall's tau coefficient model-based estimates to its empirical counterpart. Denote by $\tau_{jj'}^{(k)}$ (respectively $\hat{\tau}_{jj'}^{(k)}$) the true (respectively empirical) Kendall's tau coefficient between the j-th and j'-th individual assets for the k-th year. Let $\tau^{(k)} = \sum_{j<j'} \tau_{jj'}^{(k)}/(d(d-1)/2)$ and $\overline{\tau}^{(k)} = \sum_{j<j'} \hat{\tau}_{jj'}^{(k)}/(d(d-1)/2)$ be the respective pairwise averages. Recall that the empirical estimate of Kendall's tau coefficient beteween the j-th and j'-th individual assets for the k-th year is given by

$$\hat{\tau}_{jj'}^{(k)} = \binom{n}{2}^{-1} \sum_{i<i'} \text{sign}\left((X_{ij}^{(k)} - X_{i'j}^{(k)})(X_{ij'}^{(k)} - X_{i'j'}^{(k)})\right),$$

where $\text{sign}(x) = 1$ if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$.

Confidence intervals around $\tau^{(k)}$ can be built as follows. Let $\boldsymbol{\tau}^{(k)}$ be the vector whose coordinates are the $\tau_{jj'}^{(k)}$ and let $\hat{\boldsymbol{\tau}}^{(k)}$ be its empirical counterpart. The vector $\sqrt{n}(\hat{\boldsymbol{\tau}}^{(k)} - \boldsymbol{\tau}^{(k)})$ tends to a centered normal distribution of dimension $d(d-1)/2$. The asymptotic variance-covariance matrix can be estimated by bootstrap from formula (12) in [16]. The convergence in distribution of $\sqrt{n}(\overline{\tau}^{(k)} - \tau^{(k)})$ to a centered normal distribution and standard deviation, say $\sigma^{(k)}$, comes after applying the delta method. Again, on can compute an estimate, say $\hat{\sigma}^{(k)}$, of $\sigma^{(k)}$ by bootstrap. As a result, one can easily compute a confidence interval of level 99% as $\overline{\tau}^{(k)} \pm c_{.99}\hat{\sigma}^{(k)}/\sqrt{n}$, where $c_{.99}$ is a number such that the probability of a standard normal variable to lie between $-c_{.99}$ and $+c_{.99}$ is 99%.

Results

Figure 2 pictures the average of the pairwise Kendall's tau coefficient model-based estimates. The shaded area represents the 99% empirical confidence intervals. The results presented in Figure 2 support the plausibility of our model as the curve lies inside the confidence intervals. In particular, these results also demonstrate the model flexibility, as the curve seems to "follow" the empirical confidence intervals.

Figure 3 pictures three indexes normalized so that their shape through time could be drawn and compared on the same picture. The normalizations are of the form $f_{\text{normalized}}(t) = (f(t) - F_-)/(F_+ - F_-)$ where F_- and F_+ are the minimum and maximum values of f respectively. Thus, up to normalization, the dashed line represents the NASDAQ loss rate, that is, $(X_{ij}^{(k)} - X_{ij}^{(k+1)})/X_{ij}^{(k)}$ for $k \in \{2007, \dots, 2013\}$; the dotted line[3] represents the put-call ratio on the CBOE total exchange volume (see below for an explanation) and the plain line represents the latent factor estimated median.

The put-call ratio[4] on a certain market is, as its name tells, the ratio between the put options and the call options on that market. Put options are contracts on assets which give one the right, but not the obligation, to sell that asset in the future at a price fixed today, so that a profit can be made if the price of the asset goes down. Conversely, a call option is a contract on a asset which allows one to buy in the future at a price fixed today. Thus, arguably, the ratio of the put to the call can be seen as the overall attitude of investors toward a financial market. As such, we might see it as a crisis indicator. Likewise, the index of a market, such as the NASDAQ, is commonly regarded as mirroring the state of some part of the economy, and, therefore, its loss rate can be seen again as a sort of crisis indicator.

Therefore, we have at our disposal, on the one hand, two crisis indicators computed independently from our model, and our latent factor estimated median, which we chose to interpret as a crisis indicator. Arguably, if these three indexes — the NASDAQ loss rate, the put-call ratio and the latent factor estimated median — exhibit a similar behavior, this would support, first, our choice to interpret X_0 as a crisis indicator, second, the functional form of our copula C_{x_0} and third, our model all together.

In Figure 3, all the indexes have a similar shape: that of the letter "M": it goes up, down, up and down again. Moreover, they all pick down at around the same location, corresponding to 2010. Also, they all pick up at around 2008, corresponding to the world financial crisis that took its root into the subprimes

[3]The data were downloaded from `http://www.cboe.com/data/putcallratio.aspx`

[4]see e.g. `http://www.investopedia.com/terms/p/putcallratio.asp`

crisis in the summer of 2007. After 2010, they start to go up again, perhaps corresponding to the European sovereign debt crisis. While we have, admittedly, no expertise to discuss the relevancy or confidence one can have in these indexes, it is still noticeable how they agree with each other, how they tell the same story. In particular, the behavior of our latent factor is consistent with the behaviors of the other crisis indicators. Therefore, we believe that our model passed an important empirical test and we hope that to have convinced the reader of its usefulness in this situation.

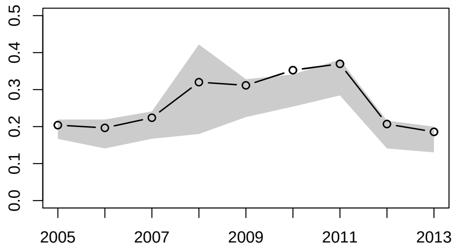

Figure 2: Pairwise Kendall's tau estimated coefficients average under the considered model along with empirical confidence intervals through time.

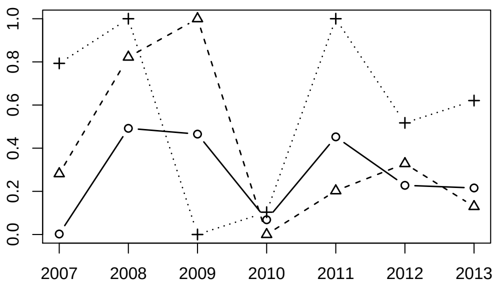

Figure 3: Three normalized financial markets trackers through time: the plain, dashed and dotted curves represent the latent factor estimated median, NASDAQ loss rate and the put-call ratio.

5 Discussion

In this paper, we extended the scope of one-factor copulas by deriving two equivalent representations from which new parametric models can be built. These models can now feature a varying conditional dependence structure and a factor's distribution not restricted to be the standard uniform. This permits to estimate the factor's distribution, despite unobserved. The usefulness of our approach was illustrated by considering the estimation of the behavior of a financial market through the dependence of its individual components. Furthermore, a novel hypothesis test was constructed in order to assess whether conditional independence holds or not.

Nonetheless, open challenges still remain. In our view, one of great importance is the issue of identifiability. Assuming that parametric families have been chosen in (10), different parameter vectors can yield the same distribution. For instance, for $d = 2$, $C_{x_0} = \Pi$ and C_{01}, C_{02} being Farlie-Gumbel-Morgenstern copulas, that is, $C_{0j}(u_0, u_j; \alpha_j) = u_0 u_j + \alpha_j u_j u_0 (1 - u_0)(1 - u_j)$, with $\alpha_j \in [1, -1]$, the copula (10) is easily calculated as $C(u_1, u_2, \alpha_1, \alpha_2) = u_1 u_2 (\alpha_1 \alpha_2 (u_1 - 1)(u_2 - 1) + 3)$. Thus, one can see that $C(u_1, u_2, \alpha_1, \alpha_2) = C(u_1, u_2, \alpha_1', \alpha_2')$ whenever $\alpha_1 \alpha_2 = \alpha_1' \alpha_2'$, and the last equation can be satisfied even if $(\alpha_1, \alpha_2) \neq (\alpha_1', \alpha_2')$. Needless to say, in higher dimensions or for other parametric families, identifiability issues may be tougher to spot.

Acknowledgment. The authors thank Johan Segers for helpful comments which permitted to improve the presentation of this paper. Gildas Mazo thanks Fabrizio Durante for an interesting discussion which triggered the start of the work presented in this paper. The research by G. Mazo was funded by a "Projet de Recherche" of the "Fonds de la Recherche Scientifique – FNRS" (Belgium).

6 Appendix

Proof of Proposition 1.

Assume $d = 2$. Let $\delta_k(v)$ (respectively $\overline{\delta}_k(v)$) be the limit of $C_{k|0}(t|vt)$ (respectively $C_{k|0}(1-t|1-vt)$) as $t \to 0$ for $k = i, j$ and for $0 < v < 1$.

$$\begin{aligned}\lambda^L &= \lim_{u\to 0} \frac{1}{u}\int_0^1 C_{u_0}(C_{1|0}(u|u_0), \dots, C_{d|0}(u|u_0))du_0 \\ &= \lim_{u\to 0}\int_0^{1/u} C_{vu}(C_{1|0}(u|vu), C_{2|0}(u|vu))dv.\end{aligned}$$

Fix $v > 0$. Let $\varepsilon > 0$ and $\eta(v) = \varepsilon e^{-v}$. The triangle inequality yields

$$\begin{aligned}&|C_0(\delta_1(v), \delta_2(v)) - C_{uv}(C_{1|0}(u|uv), C_{2|0}(u|uv))| \\ \le&|C_0(\delta_1(v), \delta_2(v)) - C_0(C_{1|0}(u|uv), C_{2|0}(u|uv))| \\ +&|C_0(C_{1|0}(u|uv), C_{2|0}(u|uv)) - C_{uv}(C_{1|0}(u|uv), C_{2|0}(u|uv))|\end{aligned}$$

for any $0 \le u \le 1$. By definition of δ_1 and δ_2 and continuity of C_0, as $u \to 0$, the first term in the right hand side can be made arbitrarily small. So does the second term, by uniform convergence of $u \mapsto C_{uv}$ to a continuous copula C_0. Therefore, for u small enough, $C_0(\delta_1(v), \delta_2(v)) - \eta(v) < C_{uv}(C_{1|0}(u|vu), C_{2|0}(u|vu))$. Thus,

$$\int_0^{1/u} C_0(\delta_1(v), \delta_2(v))\, dv - \varepsilon(1 - e^{-1/v}) < \int_0^{1/u} C_{uv}(C_{1|0}(u|vu), C_{2|0}(u|vu))\, dv.$$

Passing by the limit $u \to 0$,

$$\int_0^{\infty} C_0(\delta_1(v), \delta_2(v))\, dv - \varepsilon < \lim_{u\to 0}\int_0^{1/u} C_{uv}(C_{1|0}(u|vu), C_{2|0}(u|vu))\, dv = \lambda^L.$$

Note that the integral in the left hand side is finite (otherwise λ^L would not exist). It is also (strictly) positive because of the following arguments. The copula C_0 is (strictly) increasing in each of its arguments as its density is (strictly positive) whenever its arguments are in $(0, 1)$. Moreover, since the lower tail dependence coefficient of C_{0k} is positive, there exists $0 < v < 1$ such that $\delta_k(v) > 0$ for both $k = i$ and $k = j$. See [8] or [11] for a proof.

To conclude that $\lambda^L > 0$, note that ε was arbitrary and therefore could have been taken as small as desired.

The proof for the upper tail dependence coefficient is quite similar to the proof of the first part. Since

$$\int_0^1 C_{i|0}(u_i|u_0)\, du_0 = u_i,$$

$i = 1, \dots, d$, we have

$$\begin{aligned}\lambda^U =& \lim_{u\to 0}\frac{1}{u}\int_0^1 \overline{C}_{u_0}(C_{1|0}(1-u|u_0), C_{2|0}(1-u|u_0))du_0 \\ & \lim_{u\to 0}\int_0^{1/u} \overline{C}_{1-uv}(C_{1|0}(1-u|1-uv), C_{2|0}(1-u|1-uv))dv,\end{aligned}$$

where for any bivariate copula C, $\overline{C}(u,v) = 1-u-v+C(u,v)$. To proceed, one easily adapt the proof for the lower tail dependence coefficient.

Extension to $d > 2$ simply amounts to look at the bivariate pairs since tail dependence coefficients as understood in this paper are defined for bivariate copulas only.

Proof of Proposition 2

Checking that C_{0i} is a copula is straightforward. Moreover, since

$$\frac{\partial C_{0i}(u_0,u_i)}{\partial u_0} = e^{-Q_0(u_0)\psi^{-1}(u_i)},$$

it holds that

$$C(u_1,\ldots,u_d) = \int_0^\infty e^{-t\sum_{i=1}^d \psi^{-1}(u_i)}\,dt = \psi(\sum_{i=1}^d \psi^{-1}(u_i)).$$

Proof of Proposition 3

First, let us prove that C_{x_0} is a copula. Define for $0 \le u,v,w \le 1$

$$G_{x_0}(u,v,w) = \exp\Bigg(-x_0\times\nu\bigg(\nu^{-1}\Big[\psi_{123}^{-1}\Big(u\Big)\Big]+\nu^{-1}\Big[\psi_{123}^{-1}\Big(v\Big)\Big]\bigg)\Bigg) \times\exp\Bigg(-x_0\psi_{123}^{-1}\Big(w\Big)\Bigg).$$

One can check that C_{x_0} of Proposition 3 is the copula corresponding to G_{x_0}. Now, from [7], page 88, it can be easily deduced that G_{x_0} is a distribution function. Therefore C_{x_0} is a copula.

Let us go on by showing that C in (10) is a nested Archimedean copula. We have

$$C(u_1,u_2,u_3) = \int_0^1 \exp\Bigg(-Q_0(u_0)\times\nu\bigg(\nu^{-1}\Big[\frac{1}{Q_0(u_0)}\log\Big(\frac{1}{C_{1|0}(u_1|u_0)}\Big)\Big] + \nu^{-1}\Big[\frac{1}{Q_0(u_0)}\log\Big(\frac{1}{C_{1|0}(u_2|u_0)}\Big)\Big]\bigg)\Bigg)\times C_{3|0}(u_3|u_0)\,du_0. \quad (29)$$

In the proof of Proposition 2, it was shown that $C_{i|0}(u_i|u_0) = \exp(-Q_0(u_0)\times\psi_{123}^{-1}(u_i))$. Replacing in (29) gives

$$\begin{aligned}&\int_0^1 \exp\Bigg(-Q_0(u_0)\times\nu\bigg(\nu^{-1}\Big[\psi_{123}^{-1}(u_1)\Big]+\nu^{-1}\Big[\psi_{123}^{-1}(u_2)\Big]\bigg)\Bigg)\\ &\qquad\exp(-F_0^{-1}(u_0)\times\psi_{123}^{-1}(u_3))\,du_0\\ =&\int_0^\infty \exp\Bigg(-x_0\bigg[\psi_{123}^{-1}\Big(\psi_{12}\Big(\psi_{12}^{-1}(u_1)+\psi_{12}^{-1}(u_2)\Big)\Big)+\psi_{123}^{-1}(u_3)\bigg]\Bigg)dF_0(x_0)\\ =&\psi_{123}\big(\psi_{123}^{-1}(\psi_{12}(\psi_{12}^{-1}(u_1)+\psi_{12}^{-1}(u_2)))+\psi_{123}^{-1}(u_3)\big).\end{aligned}$$

Proof of Proposition 4.

Let R be a symmetric nonnegative matrix whose diagonal elements are equal to 1 and whose element in the i-th row and j-th column is denoted by β_{ij}. Let $(Z_1, \ldots, Z_d, Z_0)$ be distributed according to a $(d+1)$-variate centered Gaussian distribution with variance-covariance matrix given by

$$\begin{pmatrix} \underset{d\times d}{R} & \underset{d\times 1}{\beta_0} \\ \underset{1\times d}{\beta_0^T} & \underset{1\times 1}{1} \end{pmatrix}$$

so that $(Z_1, \ldots, Z_d | Z_0 = z_0) \sim N(\beta_0 z_0,\, R - \beta_0 \beta_0^\top)$. The partial correlations are given by

$$\mathrm{Cov}(Z_i, Z_j | Z_0 = z_0) = \mathrm{Corr}(Z_i, Z_j | Z_0 = z_0) = \frac{\beta_{ij} - \beta_{0i}\beta_{0j}}{\sqrt{(1-\beta_{0i}^2)(1-\beta_{0j}^2)}},$$

or, in other words, the partial correlation matrix is $A = D^{-1/2}(R - \beta_0\beta_0^\top)D^{-1/2}$. Given $Z_0 = z_0$, the margins $(Z_i|Z_0 = z_0)$ are $N(\beta_{0i}z_0, 1-\beta_{0i}^2)$, hence $P(Z_i \leq z | Z_0 = z_0) = \Phi((z - \beta_{0i}z_0)/\sqrt{1-\beta_{0i}^2})$, and, moreover, the corresponding copula is a Gaussian copula with A as its correlation matrix; let us denote it by C_A.

Let us calculate the copula corresponding to $(Z_1, \ldots, Z_d)$. Let Φ be the cumulative distribution function of the univariate standard Gaussian distribution.

$$\begin{aligned}
&C_{(Z_1,\ldots,Z_d)}(u_1, \ldots, u_d) \\
&= \int P(\Phi(Z_i) \leq u_i,\, i = 1, \ldots, d | Z_0 = z_0)\Phi'(z_0)dz_0 \\
&= \int P(Z_i \leq \Phi^{-1}(u_i),\, i = 1, \ldots, d | Z_0 = z_0)\Phi'(z_0)dz_0 \\
&= \int C_A \left\{ \Phi\left(\frac{\Phi^{-1}(u_i) - \beta_{0i}z_0}{\sqrt{1-\beta_{0i}^2}} \right), i = 1, \ldots, d \right\} \Phi'(z_0)dz_0 \\
&= \int_0^1 C_A \left\{ \Phi\left(\frac{\Phi^{-1}(u_i) - \beta_{0i}\Phi^{-1}(u_0)}{\sqrt{1-\beta_{0i}^2}} \right), i = 1, \ldots, d \right\} du_0
\end{aligned}$$

But this expression corresponds exactly to the copula given in (10).

References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

[2] T. Bedford and R.M. Cooke. Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

[3] D. Eddelbuettel and R. François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

[4] C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.

[5] M. Hofert and D. Pham. Densities of nested Archimedean copulas. *Journal of Multivariate Analysis*, 118:37–52, 2013.

[6] I. Jacques and C. Judd. *Numerical Analysis*. Chapman and Hall, 1987.

[7] H. Joe. *Multivariate models and dependence concepts*. Chapman & Hall/CRC, 2001.

[8] H. Joe. Tail dependence in vine copulae. In *Dependence modeling: Vine Copula Handbook*, chapter 8, pages 165–187. World Scientific, 2011.

[9] H. Joe. *Dependence Modeling with Copulas*. Chapman & Hall, 2014.

[10] H. Joe and J. J. Xu. The estimation method of inference functions for margins for multivariate models. Technical report, Department of Statistics, University of British Columbia, Vancouver, 1996.

[11] P. Krupskii and H. Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101, 2013.

[12] P. Krupskii and H. Joe. Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis*, 138:53–73, 2015.

[13] D. Kurowicka and H. Joe. *Dependence Modeling: Vine Copula Handbook*. World Scientific, 2011.

[14] K. Lange. *Numerical Analysis for Statisticians*. Springer, 2010.

[15] G. Mazo, S. Girard, and F. Forbes. A flexible and tractable class of one-factor copulas. *Statistics and Computing*, pages 1–15, 2015.

[16] G. Mazo, S. Girard, and F. Forbes. Weighted least-squares inference based on dependence coefficients for multivariate copulas. *ESAIM: Probability and Statistics*, 19:746–765, 2015.

[17] A. J. McNeil. Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581, 2008.

[18] A. J. McNeil and J. Nešlehová. Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *The Annals of Statistics*, 37:3059–3097, 2009.

[19] R. B. Nelsen. *An introduction to copulas*. Springer, New York, 2006.

[20] O. Okhrin, Y. Okhrin, and W. Schmid. Properties of hierarchical Archimedean copulas. *Statistics & Risk Modeling*, 30(1):21–54, 2013.

[21] J. Segers and N. Uyttendaele. Nonparametric estimation of the tree structure of a nested Archimedean copula. *Computational Statistics & Data Analysis*, 72:190–204, 2014.

[22] A. Sklar. Fonction de répartition dont les marges sont données. *Inst. Stat. Univ. Paris*, 8:229–231, 1959.

[23] A. Skrondal and S. Rabe-Hesketh. Latent variable modelling: A survey. *Scandinavian Journal of Statistics*, 34(4):712–745, 2007.

[24] N. Tajvidi, S. Kiatsupaibul, S. Tirapat, and C. Panyangam. Behavior of extreme dependence between stock markets when the regime shifts. *Sri Lankan Journal of Applied Statistics*, 16(1):21–40, 2015.