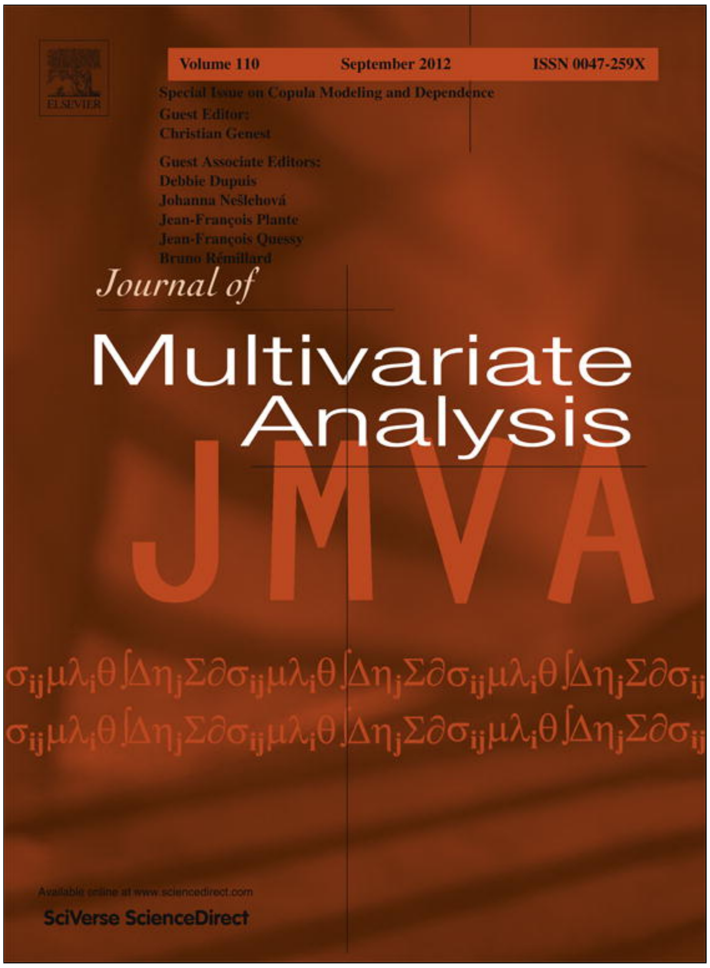
Volume 110
September 2012
ISSN 0047-259X
Special Issue on Copula Modeling and Dependence
Guest Editor:
Christian Genest
Guest Associate Editors:
Debbie Dupuis
Johanna Nešlehová
Jean-François Plante
Jean-François Quessy
Bruno Rémillard
Journal of
Multivariate Analysis
JMVA
SciVerse ScienceDirect

Journal of Multivariate Analysis 110 (2012) 19–29

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Time-dependent copulas

Jean-David Fermanian [a,*], Marten H. Wegkamp [b]

[a] CREST-ENSAE, J120, 3, avenue Pierre-Larousse, 92245 Malakoff cedex, France

[b] Department of Mathematics & Department of Statistical Science, Cornell University, Ithaca, NY, USA

ARTICLE INFO

Article history:
Available online 13 March 2012

AMS subject classifications:
62H
62M

Keywords:
Copulas
Goodness-of-fit tests
Kernel method
Time series

ABSTRACT

For the study of dynamic dependence structures, the authors introduce the concept of a pseudo-copula, which extends Patton's definition of a conditional copula. They state the equivalent of Sklar's theorem for pseudo-copulas. They establish the asymptotic normality of nonparametric estimators of pseudo-copulas under strong mixing assumptions, and discuss applications to specification tests. They complement the theory with a small simulation study on the power of the proposed tests.

1. Introduction

Copulas were introduced in the literature by Sklar [38] in the 1950s. They were rediscovered in the 1980s and have since generated a significant stream of academic literature. They have attracted the attention of practitioners, especially due to their applications in actuarial science, biostatistics, economics, finance, hydrology, etc. By now, they have become a standard tool in finance and insurance; see, e.g., [5,10].

Copulas are useful tools to model multivariate distributions. They are cumulative distribution functions (cdf's) on $[0, 1]^d$ for some integer $d > 1$, with uniform margins. Basically, copulas allow one to replace a rather complex task (the specification of a joint cdf) by two simpler ones: the specification of marginal distributions and of a dependence structure. For a discussion of these ideas and many mathematical properties of copula functions, we refer to the books by Joe [17], Nelsen [29], Embrechts et al. [9], among others. Recent developments in the theory of copulas can be found, e.g., in [6,12,20,22].

One limiting feature of copulas is the difficulty to use them in the presence of multivariate processes, say $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, with $\mathbf{X}_n \in \mathbb{R}^d$. A common practical problem is to specify the law of this process, possibly through copulas. A first idea would be to describe the law of the vectors $(\mathbf{X}_m, \mathbf{X}_{m+1}, \ldots, \mathbf{X}_n)$ for every couple (m, n), $m < n$. This can be done by modeling separately $d(n-m+1)$ unconditional margins plus a $d(n-m+1)$-dimensional copula. This approach seems particularly useful when the underlying process is stationary and Markov; see [3] for the general procedure.

Darsow et al. [7] characterized univariate Markov processes induced by copula families. Their work was extended by Ibragimov [16] in a multivariate framework. Recently, Beare [2] obtained sufficient conditions for geometric rates of mixing. Unfortunately, the description of the law of the process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, or the simulation of its trajectories, is not significantly simplified by this two-step procedure when the size of the relevant copula is "large" (i.e., $d(m-n+1) \geq 3$) and it becomes impractical for dimensions $d > 2$.

An alternative approach would be to use information on the *marginal processes*. This requires us to specify conditional marginal distributions, instead of unconditional margins as above. This approach is very natural because in practice

* Corresponding author.
E-mail addresses: jean-david.fermanian@ensae.fr (J.-D. Fermanian), marten.wegkamp@cornell.edu (M.H. Wegkamp).

doi:10.1016/j.jmva.2012.02.018

univariate processes are often better known and studied than multivariate processes. For instance, financial firms can model and calibrate easily univariate models for individual stock returns, because numerous liquid and reliable quotes exist in the market. These models are implemented in the information systems of these firms. But typically, it is no more the case when dealing with a basket of stocks. Therefore, it is tempting to use "standard" univariate processes as inputs of more complicated multivariate models.

Apart from pricing and hedging of basket derivatives, this problem occurs, e.g., in risk management (joint behavior of several market factors like stock indices, exchange rates, interest rates, etc.) and in credit portfolio modeling (e.g., the joint default of several counter-parties). This means that we need to specify, for every $j = 1, \ldots, d$, the law of $X_{n,j}$ knowing the past values $X_{n-1,j}, X_{n-2,j}, \ldots$ of this univariate process. The main issue is now related to the specification and the estimation of relevant dependence structures, "knowing" these univariate underlying processes, to recover the entire process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$.

Using motivation similar to our second approach above, Patton [32,33] introduced so-called *conditional copulas*, which are associated with conditional laws in a particular way. Specifically, let $\mathbf{X} = (X_1, \ldots, X_d)$ be a d-dimensional random vector from $(\Omega, \mathcal{A}_0, \mathbb{P})$ to $\mathbb{R}^d$. Consider some arbitrary sub-σ-algebra $\mathcal{A} \subset \mathcal{A}_0$. A conditional copula associated to $(\mathbf{X}, \mathcal{A})$ is a $\mathcal{B}([0, 1]^d) \otimes \mathcal{A}$ measurable function C such that, for any $x_1, \ldots, x_d \in \mathbb{R}$,

$$\mathbb{P}(\mathbf{X} \le \mathbf{x}|\mathcal{A}) = C\left\{\mathbb{P}(X_1 \le x_1|\mathcal{A}), \ldots, \mathbb{P}(X_d \le x_d|\mathcal{A})|\mathcal{A}\right\}. \tag{1}$$

The random function $C(\cdot|\mathcal{A})$ is uniquely defined on the product of the values taken by $x_j \mapsto \mathbb{P}(X_j \le x_j \mid \mathcal{A})(\omega)$, $j = 1, \ldots, d$, for every realization $\omega \in \mathcal{A}$. As in the proof of Sklar's theorem (see [29]), $C(\cdot|A)$ can be extended on $[0, 1]^d$ as a copula, for every conditioning subset of events $A \subset \mathcal{A}$.

In the case of multivariate processes, we deal with vector-valued observations $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ and we are often interested in the copula of $\mathbf{X}_n$ given $\mathcal{A}_n = \sigma(\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \ldots)$. Eq. (1) implies that it is necessary to know/model each margin, knowing *all* the past information (vector-valued observations) and not only the past observations of each particular margin.

Nonetheless, as mentioned above, practitioners often specify and estimate marginal models, especially in a dynamic framework. It is thus highly desirable to incorporate these models into a full multivariate model. With the previous notation, users often have good estimates of the conditional distribution of each margin, conditionally given its own past, i.e., $\mathbb{P}(\mathbf{X}_{n,j} \le x_j|\mathcal{A}_{n,j})$, $j = 1, \ldots, d$, by setting $\mathcal{A}_{n,j} = \sigma(X_{n-1,j}, X_{n-2,j}, \ldots)$. And they would like to link these quantities with the (joint) law of $\mathbf{X}_n$ knowing its own past.

It is tempting to write

$$\mathbb{P}(\mathbf{X}_n \le \mathbf{x}|\mathcal{A}_n) = C^*\left\{\mathbb{P}(X_{1,n} \le x_1|\mathcal{A}_{n,1}), \ldots, \mathbb{P}(X_{d,n} \le x_d|\mathcal{A}_{n,d})\right\} \tag{2}$$

for some random function $C^* : [0, 1]^d \longrightarrow [0, 1]$ whose measurability would depend on $\mathcal{A}_n$ and on the $\mathcal{A}_{n,j}$, $j = 1, \ldots, d$. Actually, if the latter function were a copula, then it should satisfy in particular

$$\mathbb{P}\left(X_{1,n} \le x_1|\mathcal{A}_n\right) = C^*\left\{\mathbb{P}(X_{1,n} \le x_1|\mathcal{A}_{n,1}), 1, \ldots, 1\right\} = \mathbb{P}(X_{1,n} \le x_1|\mathcal{A}_{n,1}),$$

for every real number x_1. This means that as far as predicting the current value of the vector-valued process $(\mathbf{X}_n)$ is concerned, its past values do not provide more information than the past values of the univariate process $(X_{1,n})$ only. In other words, the process $(X_{2,n}, \ldots, X_{d,n})_{n\in\mathbb{Z}}$ does not "Granger-cause" the process $(X_{1,n})_{n\in\mathbb{Z}}$. Obviously, a similar reasoning can be made for every margin.

The assumption that each variable depends on its own lags, but not on the lags of any other variable, is clearly strong, even though it can be accepted empirically; see the discussion in [34, pp. 772–773]. In particular, Patton [32] indicates that "in our empirical application we find that, conditional on lags of the DM-USD exchange rate, lags of the Yen-USD exchange rate do not impact the distribution of the DM-USD exchange rate". Therefore, to fulfill (2), the concept of copula itself has to be revisited. This is the main purpose of this article.

The remainder of the paper is organized as follows. In Section 2 we extend Patton's definition to cover a much larger scope of situations, and prove the equivalent of Sklar's theorem. Then we deal with the nonparametric estimation of pseudo-copula models and the important problem of goodness-of-fit tests in Section 3. Finally, in Section 4, we provide a modest simulation study to evaluate the performance of our test statistic.

2. Conditional copulas and pseudo-copulas

A copula is a cdf on $[0, 1]^d$ with uniform margins. We will call a *pseudo-copula* a cdf on $[0, 1]^d$ with arbitrary margins. As we will manipulate these objects in the same way as copulas, and given that they will be compared with the so-called *conditional copulas*, we think it is valuable to introduce the term formally.

Definition 1. A d-dimensional pseudo-copula is a function $C : [0, 1]^d \to [0, 1]$ such that

(a) For every $\mathbf{u} \in [0, 1]^d$, $C(\mathbf{u}) = 0$ when at least one coordinate of $\mathbf{u}$ is zero;
(b) $C(1, \ldots, 1) = 1$;
(c) For every $\mathbf{u}$ and $\mathbf{v}$ in $[0, 1]^d$ such that $\mathbf{u} \le \mathbf{v}$, the C-volume (see [29, Definition 2.10.1]) of $[\mathbf{u}, \mathbf{v}]$ is positive.

Invoking the same type of arguments as in Sklar's theorem (see, e.g., [29]), we get the following result.

Theorem 1. *Let H be a cdf on $\mathbb{R}^d$ and let $F_1, \ldots, F_d$ be d arbitrary univariate cdf's on $\mathbb{R}$. Assume that for every $\mathbf{x} = (x_1, \ldots, x_d), \tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_d) \in \mathbb{R}^d$,*

$$F_j(x_j) = F_j(\tilde{x}_j), \quad 1 \le j \le d \quad \Rightarrow \quad H(\mathbf{x}) = H(\tilde{\mathbf{x}}). \tag{3}$$

Then there exists a pseudo-copula C such that

$$H(\mathbf{x}) = C\left\{F_1(x_1), \ldots, F_d(x_d)\right\}, \tag{4}$$

for every $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. C is uniquely defined on $\mathrm{Ran}\, F_1 \times \cdots \times \mathrm{Ran}\, F_d$, the product of the values taken by the F_j. Conversely, if C is a pseudo-copula and if $F_1, \ldots, F_d$ are some univariate cdf's, then the function H defined by (4) is a d-dimensional cdf.

Therefore, when C is a pseudo-copula, the function H, as defined below by (6), will be a valid cdf on $\mathbb{R}^d$, but its margins can be quite arbitrary (and different from the F_j cdf in general).

Remarks. Obviously, condition (3) is satisfied if the distribution functions F_j are invertible, in particular, if F_j are strictly increasing. If this condition it not satisfied, the knowledge of $(F_1(x_1), \ldots, F_d(x_d))$ is not sufficient to recover $H(\mathbf{x})$, and then the concept of pseudo-copulas fails.

Further note that if the pseudo-copula C in Theorem 1 is a (true) copula, then

$$H(+\infty, \ldots, x_j, \ldots, +\infty) = F_j(x_j), \tag{5}$$

for every $j = 1, \ldots, d$ and $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, i.e., the marginal distributions of the joint cdf H are $F_j, j = 1, \ldots, d$. Conversely, if (5) is satisfied and if the functions $F_j, j = 1, \ldots, d$ are continuous, then C is a true copula.

Finally, observe that pseudo-copulas should not be confused with quasi-copulas dealt with in the literature (see, e.g., Section 6.2 in [29]). Note that the pseudo-copula C above can be extended on $[0, 1]^d$, exactly as in the usual proof of Sklar's theorem; see [29].

Proof of Theorem 1. We prove the result for $d = 2$ only as the extension to higher dimensions is relatively straightforward. Consider the function $\bar{C} : [0, 1]^2 \to [0, 1]$ defined by

$$\bar{C}(u, v) = H\{F_1^{(-1)}(u), F_2^{(-1)}(v)\},$$

where $F_j^{(-1)}(u) = \inf\{t \mid F_j(t) \ge u\}$, $j = 1, 2$ are generalized inverse functions. Mimicking the proof of Sklar's theorem, it is easy to check that $\bar{C}$ is a pseudo-copula and we just need to prove is that $\bar{C}$ satisfies (4). In fact,

$$\bar{C}\{F_1(x_1), F_2(x_2)\} = H\{F_1^{(-1)} \circ F_1(x_1), F_2^{(-1)} \circ F_2(x_2)\}$$

and

$$H\{F_1^{(-1)} \circ F_1(x_1), F_2^{(-1)} \circ F_2(x_2)\} = H(x_1, x_2),$$

because $F_j(x_j) = F_j \circ F_j^{(-1)} \circ F_j(x_j)$ for $j = 1, 2$ by assumption (3). Thus, the existence of a pseudo-copula is obtained. Moreover, when C is unique, $C = \bar{C}$. From Eq. (4), it is obvious that C is uniquely defined on $\mathrm{Ran}\, F_1 \times \cdots \times \mathrm{Ran}\, F_d$ and must be equal to $\bar{C}$. Indeed, if $u = F_1(x_1) = F_1(\tilde{x}_1)$ and $v = F_2(x_2) = F_2(\tilde{x}_2)$, then $H(x_1, x_2) = H(\tilde{x}_1, \tilde{x}_2) = C(u, v)$ for every pseudo-copula C. The converse result is straightforward. $\square$

Now, let us discuss the link between conditional laws and pseudo-copulas. Consider $\mathcal{A}_1, \ldots, \mathcal{A}_d$ and $\mathcal{B}$, some sub-σ-algebras of $\mathcal{A}_0$. For instance, $\mathcal{B}$ could be the σ-algebra induced by $\mathcal{A}_1, \ldots, \mathcal{A}_d$ (and that we denote by $\sigma(\mathcal{A}_1, \ldots, \mathcal{A}_d)$), but this is not required. The only restriction on these sub-σ-algebras is that the "marginal" σ-algebras $\mathcal{A}_j$ do not provide more information than the "global" σ-algebra $\mathcal{B}$. That is, we will assume

Assumption (S). For every $j = 1, \ldots, d$, $\mathcal{A}_j \subseteq \mathcal{B}$.

For convenience, denote $\mathcal{A} = \sigma(\mathcal{A}_1, \ldots, \mathcal{A}_d)$. We have the following result.

Theorem 2. *For any sub-algebras $\mathcal{B}, \mathcal{A}_1, \ldots, \mathcal{A}_d$ that satisfy (S), there exists a random function $C : [0, 1]^d \times \Omega \longrightarrow [0, 1]$ such that*

$$\begin{aligned} \mathbb{P}(\mathbf{X} \le \mathbf{x} \mid \mathcal{B})(\omega) &= C\left\{\mathbb{P}(X_1 \le x_1 \mid \mathcal{A}_1)(\omega), \ldots, \mathbb{P}(X_d \le x_d \mid \mathcal{A}_d)(\omega), \omega\right\} \\ &\equiv C\left\{\mathbb{P}(X_1 \le x_1 \mid \mathcal{A}_1), \ldots, \mathbb{P}(X_d \le x_d \mid \mathcal{A}_d)\right\}(\omega), \end{aligned} \tag{6}$$

for every $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and almost every $\omega \in \Omega$. This function C is $\mathcal{B}([0, 1]^d) \otimes \mathcal{B}$ measurable. For almost every $\omega \in \Omega$, $C(\cdot, \omega)$ is a pseudo-copula and is uniquely defined on the product of the values taken by $x_j \mapsto \mathbb{P}(X_j \le x_j \mid \mathcal{A}_j)(\omega)$, $j = 1, \ldots, d$.

Proof. The proof is formally identical to that of Theorem 1, replacing distributions with conditional distributions. We check (3) in this case. For all $\mathbf{x} = (x_1, \dots, x_d)$, $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_d)$ in $\mathbb{R}^d$ and for any arbitrary event $\omega \in \Omega$,

$$\mathbb{P}(X_j \le x_j \mid \mathcal{A}_j)(\omega) = \mathbb{P}(X_j \le \tilde{x}_j \mid \mathcal{A}_j)(\omega), \quad j = 1, \dots, d$$

implies

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

Assume for convenience that $\mathbf{x} \le \tilde{\mathbf{x}}$. If $\mathbb{P}(x_j < X_j \le \tilde{x}_j \mid \mathcal{A}_j) = 0$, then $\mathbb{E}\{\mathbb{P}(x_j < X_j \le \tilde{x}_j \mid \mathcal{B})|\mathcal{A}_j\} = 0$ using Assumption (S). Since the conditional probability is non-negative, we deduce that $\mathbb{P}(x_j < X_j \le \tilde{x}_j|\mathcal{B}) = 0$. Therefore,

$$\left|\mathbb{P}(\mathbf{X} \le \mathbf{x} \mid \mathcal{B}) - \mathbb{P}(\mathbf{X} \le \tilde{\mathbf{x}} \mid \mathcal{B})\right| \le \sum_{j=1}^{d} \mathbb{P}\left(x_j < X_j \le \tilde{x}_j \mid \mathcal{B}\right) = 0,$$

as claimed. □

If C is unique, we will refer to it as the conditional $(\mathcal{A}, \mathcal{B})$-pseudo-copula associated with $\mathbf{X}$ and denote it by $C(\cdot|\mathcal{A}, \mathcal{B})$. In general, $C(\cdot \mid \mathcal{A}, \mathcal{B})$—or more precisely $C(\cdot \mid \mathcal{A}, \mathcal{B})(\omega)$—is *not* a copula. Rewriting condition (5), $C(\cdot|\mathcal{A}, \mathcal{B})$ can be extended (through Sklar's theorem) as a true copula almost everywhere if and only if

$$\mathbb{P}(X_j \le x_j \mid \mathcal{B}) = \mathbb{P}(X_j \le x_j \mid \mathcal{A}_j) \quad \text{a.e.} \tag{7}$$

for all $j = 1, \dots, d$ and $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. This means that $\mathcal{B}$ cannot provide more information about X_j than $\mathcal{A}_j$, for every j. As mentioned above, this is clearly a strong requirement; see the example below. It is closely related to the Granger causality concept in time series, when $\mathcal{B}$ contains $\mathcal{A}_j$ plus some other variables (that may or may not cause X_j). In this case, the constraint (7) can be tested nonparametrically; see, e.g., [14,15,30].

In the sequel, we will use the term *conditional copula* if C is a true copula only. Patton's conditional copula corresponds to the particular case $\mathcal{B} = \mathcal{A}_1 = \cdots = \mathcal{A}_d$, for which (7) is clearly satisfied.

Even when (S) is not satisfied, it is still possible to define a conditional $(\mathcal{A}, \mathcal{B})$-pseudo-copula C by setting, for every $\mathbf{u} \in [0, 1]^d$,

$$C(\mathbf{u}) = \mathbb{P}\{X_1 \le F_1^{(-1)}(u_1|\mathcal{A}_1), \dots, X_d \le F_d^{(-1)}(u_d|\mathcal{A}_d) \mid \mathcal{B}\}, \tag{8}$$

Obviously, we have set $F_j(x_j|\mathcal{A}_j) = \mathbb{P}(X_j \le x_j|\mathcal{A}_j)$ for $j = 1, \dots, d$. In such a case, relation (6) only holds on a subset in $\mathbb{R}^d$ (which may vary with ω).

Typically, when we consider a (stationary or not) d-dimensional process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, the previously considered σ-algebras are indexed by n. In practice, we assume most of the time that they depend on the past values of the process, even if we could consider theoretically future values or both past and future values in the conditioning subsets. For instance, we may set $\mathcal{A}_{n,j} = \sigma(X_{n-1,j}, X_{n-2,j}, \dots)$ and $\mathcal{B}_n = \sigma(\mathbf{X}_{n-1}, \dots)$. Thus, in general, we are dealing with sequences of conditional copulas and pseudo-copulas, that depend on some index n and on the past values $\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \dots$ of the vector-valued process.

When the process $(\mathbf{X}_n)$ is k-order Markov, these conditional pseudo-copulas depend on the last k observed values only. The most common choices for the sub-σ-algebras, especially in finance, are $\mathcal{A}_{n,j} = \sigma(X_{n-1,j})$ for every $j = 1, \dots, d$ and $\mathcal{B}_n = \sigma(\mathbf{X}_{n-1})$. More generally, we could consider $\mathcal{A}_{n,j} = \sigma\{(X_{n-1,j} \in \mathcal{I}_{n,j})\}$ for some or every $j = 1, \dots, d$ and $\mathcal{B}_n = \sigma\{(\mathbf{X}_{n-1} \in \mathcal{I}_n)\}$, where the $\mathcal{I}_{n,j}$ (resp. $\mathcal{I}_n$) denote some measurable subsets in $\mathbb{R}$ ($\mathbb{R}^d$, respectively). For example, the latter quantities may be some intervals or product of intervals as in [8]. The conditioning subsets could be related to several lagged values. All these cases may be mixed yielding a large scope of possibilities concerning the choice of $(\mathcal{A}_{n,1}, \dots, \mathcal{A}_{n,d}, \mathcal{B}_n)$.

The following examples show that even in very natural cases, conditional pseudo-copulas may not be copulas.

Example 1. Consider, for instance, the simple stationary bivariate process

$$\begin{cases} X_n = aX_{n-1} + \varepsilon_n \\ Y_n = bX_{n-1} + cY_{n-1} + \nu_n, \end{cases}$$

for every $n \in \mathbb{Z}$, where the sequences of residuals $(\varepsilon_n)_{n\in\mathbb{Z}}$ and $(\nu_n)_{n\in\mathbb{Z}}$ are independent standard Gaussian white noises. Set $\mathcal{A}_{n,1} = \sigma(X_{n-1})$, $\mathcal{A}_{n,2} = \sigma(Y_{n-1})$ and $\mathcal{B}_n = \sigma\{(X, Y)_{n-1}\}$. After some calculations, one gets

$$\mathbb{P}\{Y_n \le y \mid (X, Y)_{n-1} = (x, y)_{n-1}\} = \Phi\,(y - bx_{n-1} - cy_{n-1}) \quad \text{and}$$

$$\mathbb{P}\{Y_n \le y \mid Y_{n-1} = y_{n-1}\} = \Phi\left\{(y - cy_{n-1})\sqrt{\frac{1 - a^2}{1 - a^2 + b^2}}\right\}.$$

It is clearly impossible to make $\mathbb{P}(Y_n \le y \mid \mathcal{B}_n)$ and $C\{1, \mathbb{P}(Y_n \le y \mid \mathcal{A}_{n,2})\}$ equal for all triplets (x_{n-1}, y_{n-1}, y) and any copula C, when it should be true if C were a $\mathcal{B}_n$-conditional copula. In this case, the underlying pseudo-copula is given by

$$C\{u, v|(X, Y)_{n-1} = (x, y)_{n-1}\} = u\Phi\left\{\Phi^{-1}(v)\sqrt{\frac{1-a^2+b^2}{1-a^2}} - bx_{n-1}\right\}.$$

Example 2. Invoking relation (8), it is easy to encounter pseudo-copulas that are not copulas. For instance, consider the Gaussian pseudo-copula family

$$C(\mathbf{u}|\mathbf{y}^*) = \Phi_\Sigma\left\{\Phi^{-1}(u_1) + \eta_1(\mathbf{y}^*), \ldots, \Phi^{-1}(u_d) + \eta_d(\mathbf{y}^*)\right\},$$

where Φ denotes the cdf of a standard Gaussian random variable $\mathcal{N}(0, 1)$, Φ_Σ denotes the cdf of a Gaussian vector with correlation matrix Σ, whose margins are $\mathcal{N}(0, 1)$, and the functions $\eta_1, \ldots, \eta_d$ are arbitrary. The simplest choice of the latter functions could be some "index functions" $\eta_j(\mathbf{y}) = \gamma_j^\top \mathbf{y}, j = 1, \ldots, d$ for vectors γ_j. A similar construction may be used to build the Student pseudo-copula family, by replacing Φ (resp. Φ_Σ) by the cdf of a univariate (resp. multivariate) standard Student distribution.

Example 3. Archimedean pseudo-copulas constitute another example. Let $\psi : [0, +\infty) \to [0, 1]$ be the generator of an Archimedean copula. This function is continuous and strictly decreasing. It satisfies $\psi(0) = 1$, and $\lim_{x\to+\infty} \psi(x) = 0$. Kimberling [19] proved that ψ generates a (true) copula in every dimension if and only if ψ is completely monotone, i.e., iff ψ has derivatives of all orders on $(0, +\infty)$ which alternate successively in sign. For conditions needed in a specific dimension, see [26]. We will impose $\psi > 0$. Then, an Archimedean pseudo-copula with generator ψ is defined by

$$C(\mathbf{u}|\mathbf{y}^*) = \psi\{\psi^{-1}(u_1)\eta_1(\mathbf{y}^*) + \cdots + \psi^{-1}(u_d)\eta_d(\mathbf{y}^*)\},$$

for any positive functions η_j.

Example 4. Tail dependence measures and, more generally, co-movements of extreme fluctuations of random variables and processes are some of the most important concepts from copula theory in financial and economic applications of copulas. They can be extended when models are defined through pseudo-copulas. For instance, consider a bivariate random vector (X_1, X_2), with margins F_1^* and F_2^*. A new concept of lower tail dependence could be defined by

$$\begin{aligned}\lambda_L &= 2\lim_{u\downarrow 0} \frac{\mathbb{P}\{F_1(X_1) \le u, F_2(X_2) \le u\}}{\mathbb{P}\{F_1(X_1) \le u\} + \mathbb{P}\{F_2(X_2) \le u\}} \\ &= 2\lim_{u\downarrow 0} \frac{C(u, u)}{F_1^* \circ F_1^{-1}(u) + F_2^* \circ F_2^{-1}(u)},\end{aligned}$$

if C is the pseudo-copula associated to F_1 and F_2 (no conditioning here). Unfortunately, since λ_L depends on the (true) margins F_j^*, the relevance of such an indicator becomes questionable.

Now, we recall a general recipe to build numerous (discrete time) multivariate processes $(\mathbf{X}_n)_{n\ge 1}$ with time-dependent copulas.

1. Identify a particular parametric *copula* family $\{C_\theta, \ \theta \in \Theta\}$.
2. Knowing the past observations $\underline{\mathbf{X}}_{n-1} = (\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \ldots)$, define a functional relation between θ and $\underline{\mathbf{X}}_{n-1}$.
3. Set the d conditional marginal distributions $x_j \mapsto \mathbb{P}(X_{n,j} \le x_j|\mathcal{A}_{n,j})$, where $\mathcal{A}_{n,j}$ is a subset of $\sigma(\underline{\mathbf{X}}_{n-1})$. For instance, a natural choice could be $\mathcal{A}_{n,j} = \sigma(\underline{X}_{n-1,j}), j = 1, \ldots, d$, but this is not mandatory.
4. The conditional law of $\mathbf{X}_n$ is given by

$$\mathbb{P}\left(\mathbf{X}_n \le \mathbf{x}|\underline{\mathbf{X}}_{n-1}\right) = C_{\theta(\underline{\mathbf{X}}_{n-1})}\left\{\mathbb{P}(X_{1,n} \le x_1|\mathcal{A}_{n,1}), \ldots, \mathbb{P}(X_{d,n} \le x_d|\mathcal{A}_{n,d})\right\}.$$

This strategy allows for the analysis of dynamic dependence structures, but by staying in a given copula family. This natural idea has already been used in the literature; see [32,18,39] among others. In his study of the dependence between Yen-USD and Deutsche mark-USD exchange rates, Patton [32] assumes a bivariate Gaussian conditional copula whose correlation parameter follows a GARCH-type model. Jondeau and Rockinger [18] estimate similar models by specifying a copula parameter that depends linearly on the previously observed joint large deviations. Alternatively, van den Goorbergh et al. [39] postulate Kendall's tau is a function of current conditional univariate variances. We refer to Patton [34] for other references and applications in finance. In risk management particularly, the copula concept is very relevant; see, e.g., [21,23–25,28]. Note that Acar et al. [1] have discussed the nonparametric estimation of the function θ above.

In all these cases, only "true" copula families were used, even after conditioning by the past values of the process. It is then possible to build (often by simulation) multivariate models by stating independently and successively the models of $\mathbf{X}_n$ knowing its past values. This methodology works well because it is assumed (more or less explicitly) that $\mathbb{P}(X_{n,j} \le x_j|\mathcal{A}_{n,j}) = \mathbb{P}(X_{n,j} \le x_j|\mathcal{B}_n)$ for all j and x_j, as discussed above.

We see that, although pseudo-copulas have been used implicitly, they have not been introduced properly. We argue that the previous framework can be revisited: instead of the first two steps, consider conditional $(\mathcal{A}, \mathcal{B})$-pseudo-copulas and follow the same steps. Unfortunately, the statistician no longer has the freedom to choose all these pseudo-copulas arbitrarily.

To illustrate this idea, consider again the previous recipe, but now with a bivariate pseudo-copula family (C_θ). Set $\mathcal{A}_{n,j} = \sigma(\underline{X}_{n-1,j}), j = 1, \ldots, d, \mathcal{B}_n = \sigma(\underline{\mathbf{X}}_{n-1})$. Assume the process is Markov of order 1. Then, we must have

$$\mathbb{P}(\mathbf{X}_n \le \mathbf{y} \mid \mathbf{X}_{n-1} = \mathbf{x}) = C_{\theta(\mathbf{x})}\left\{\mathbb{P}(X_{n,1} \le y_1 \mid X_{n-1,1} = x_1), \mathbb{P}(X_{n,2} \le y_2 \mid X_{n-1,2} = x_2) \mid \mathbf{X}_{n-1} = \mathbf{x}\right\},$$

for all $\mathbf{y}$ and $\mathbf{x}$. This implies, for every $u \in [0, 1]$ and every $\mathbf{x}$,

$$u = \int C_{\theta(\mathbf{x})}\,(u, 1 \mid \mathbf{X}_{n-1} = \mathbf{x})\ \mathrm{d}P_{X_{n-1,2}\mid X_{n-1,1}=x_1}(x_2).$$

Clearly, it is difficult to guarantee that a given parametric family (C_θ) will satisfy such constraints exactly. A solution could be to exhibit highly flexible families that satisfy these constraints approximately. Another solution would be to work in a fully non-parametric way, as detailed in Section 3.

3. Estimation and goodness-of-fit testing

One key issue is to state if pseudo-copulas depend really on the past values of the underlying process, i.e., to test their constancy. This assumption is often made in practice; see, e.g., [4,36]. Here, we estimate non-parametrically conditional pseudo-copulas, including Patton's conditional copulas as a special case. We test their constancy with respect to their conditioning subsets.

For a stationary process $(\mathbf{X}_n)_{n\in\mathbb{Z}}$, we restrict ourselves to conditional sub-algebras $\mathcal{A}_n$ and $\mathcal{B}_n$ that are defined by a finite number of past values of the process, typically $(\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \ldots, \mathbf{X}_{n-p})$ for some $p \ge 1$. Since we will often identify the random variables and their realizations, the known value of the latter vector will be denoted by $\mathbf{y}$. Obviously, we could set $\mathcal{A}_{n,j} = \sigma(X_{n-1,j})$ and $\mathcal{B}_n = \sigma(\mathbf{X}_{n-1})$, but other specifications are possible, like $\mathcal{A}_{n,j} = \sigma(X_{n-k_j,j})$ for some $k_j \in \{1, \ldots, p\}$ and some $j = 1, \ldots, d$.

Here, the sub-indices n are irrelevant due to the stationarity assumption. The dependence of $\mathcal{A}$ and $\mathcal{B}$ with respect to past values $\mathbf{y}$ will be implicit hereafter. Formally, we would like to test the null hypothesis

$$\mathcal{H}_0^{(1)}: \text{ For every } \mathbf{y}, C(\cdot \mid \mathcal{A}, \mathcal{B}) = C_0(\cdot),$$

against

$$\mathcal{H}_a: \text{ For some } \mathbf{y}, C(\cdot \mid \mathcal{A}, \mathcal{B}) \ne C_0(\cdot),$$

where C_0 denotes a fixed pseudo-copula function. The null hypothesis means that the underlying conditional $(\mathcal{A}, \mathcal{B})$-pseudo-copula is in fact a true copula, independent of the past values of the process.

There are other interesting null hypotheses such as

$$\mathcal{H}_0^{(2)}: \text{ There exists a parameter } \theta_0 \text{ such that } C(\cdot|\mathcal{A}, \mathcal{B}) = C_{\theta_0} \in \mathcal{C}, \text{ for every } \mathbf{y},$$

where $\mathcal{C} = \{C_\theta, \theta \in \Theta\}$ denotes some known parametric family of pseudo-copulas. We may extend this assumption by allowing the parameter θ to depend on past values of the process, to test, say,

$$\mathcal{H}_0^{(3)}: \text{ For some function } \theta(\mathbf{y}) = \theta(\mathcal{A}, \mathcal{B}) \text{ we have } C(\cdot|\mathcal{A}, \mathcal{B}) = C_{\theta(\mathbf{y})} \in \mathcal{C}, \text{ for every } \mathbf{y},$$

where $\mathcal{C} = \{C_\theta,\ \theta \in \Theta\}$ denotes a family of pseudo-copulas.

The latter assumption says that the conditional pseudo-copulas stay inside the same pre-specified parametric family of pseudo-copulas, for different observed values in the past. These three null hypotheses are nested and may be easily rewritten when the conditioning subsets contain more than one past observation.

To our knowledge, only parametric copula families $\mathcal{C}$ whose parameters depend on past values of the underlying process are considered in the literature. We propose a fully nonparametric estimator of the conditional pseudo-copulas, and derive its (Gaussian) limiting distribution. This allows us to build some goodness-of-fit test statistics.

We use the short-hand notation $\mathbf{X}_m^n$ for the vector $(\mathbf{X}_m, \mathbf{X}_{m+1}, \ldots, \mathbf{X}_n)$. Similarly, we write $\mathbf{X}_{m,j}^n = (X_{m,j}, \ldots, X_{n,j})$. Assume that every conditioning set $\mathcal{A}_{n,j}$ (resp. $\mathcal{B}_n$) is related to the vector $\mathbf{X}_{n-p,j}^{n-1}$ (resp. $\mathbf{X}_{n-p}^{n-1}$). Specifically, we consider the events $(\mathbf{X}_{n-p}^{n-1} = \mathbf{y}^*) \in \mathcal{B}_n$, with $\mathbf{y}^* = (\mathbf{y}_1, \ldots, \mathbf{y}_p)$, and $(\mathbf{X}_{n-p,j}^{n-1} = \mathbf{y}_j^*) \in \mathcal{A}_{n,j}$, with $\mathbf{y}_j^* = (y_{1j}, \ldots, y_{pj})$.

With similar techniques, we could extend the framework so that past values of a given margin are integrated into the conditioning subset of other margins. Since this necessitates the introduction of alternative estimates and notations would become tedious, we prefer to concentrate on the most standard situation in practice.

Our nonparametric estimator of the pseudo-copula is based on a standard plug-in technique that requires estimates of the joint conditional distribution

$$m(\mathbf{x} \mid \mathbf{y}^*) = \mathbb{P}\left(\mathbf{X}_p \le \mathbf{x} \mid \mathbf{X}_0^{p-1} = \mathbf{y}^*\right)$$

and of conditional marginal cdf's

$$m_j(x_j \mid \mathbf{y}_j^*) = \mathbb{P}\left(X_{pj} \le x_j \mid \mathbf{X}_{0,j}^{p-1} = \mathbf{y}_j^*\right), \quad j = 1, \dots, d.$$

We will use the following set of assumptions.

Assumption (M). The sequence $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ is stationary and strongly mixing, i.e., there exists a function $\alpha(n)$ defined on $\mathbb{N}$ with $\alpha(n) \downarrow 0$ as $n \to \infty$ and

$$\sup_k |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \le \alpha(n)$$

for all $A \in \sigma(\mathbf{X}_1, \dots, \mathbf{X}_k)$ and $B \in \sigma(\mathbf{X}_{k+n}, \mathbf{X}_{k+n+1}, \dots)$ and n is a positive integer. In addition, we assume that for some $0 < \delta < 1$,

$$\sum_{j=1}^{\infty} j^{2d(p+1)} \alpha^{\delta}(j) < \infty.$$

Assumption (R1). The random vector $(\mathbf{X}_1, \dots, \mathbf{X}_{p+1})$ has a bounded density with respect to the Lebesgue measure. The density of $(\mathbf{X}_1, \dots, \mathbf{X}_p)$ is bounded away from zero in some open neighborhood of a given vector $\mathbf{y}^* \in \mathbb{R}^{pd}$. The vector $(\mathbf{X}_1^p, \mathbf{X}_{1+m}^{p+m})$ has a continuous density in some open neighborhood of $(\mathbf{y}^*, \mathbf{y}^*)$ for every $m \ge 1$. The density of $X_{n,j}$ knowing $\mathcal{A}_{n,j}$ is strictly positive at $x_j = m_j^{-1}(u_j \mid \mathcal{A}_{n,j})$, $j = 1, \dots, d$, where $\mathbf{u} = (u_1, \dots, u_d)$ denotes some given vector.

Assumption (R2). For each $\mathbf{x} \in \mathbb{R}^d$, $m\left\{\mathbf{x} | F_1^{-1}(t_1), \dots, F_d^{-1}(t_{pd})\right\}$ with $0 < t_j < 1$ is twice continuously differentiable in $\mathbf{t} \in V$, where V denotes some open neighborhood of a given pd-dimensional vector

$$(F_1(y_{11}), \dots, F_d(y_{1d}), F_1(y_{21}), \dots, F_d(y_{2d}), \dots, F_1(y_{p1}), \dots, F_d(y_{pd})),$$

and

$$\max_{1\le i,j\le pd} \sup_{\mathbf{t}\in V} \sup_{\mathbf{x}} \left| \frac{\partial^2}{\partial t_i \partial t_j} m\left\{\mathbf{x} | F_1^{-1}(t_1), \dots, F_d^{-1}(t_{pd})\right\} \right| < \infty.$$

Assumption (K). K (resp. $\bar{K}$) is a probability kernel function on $\mathbb{R}^{pd}$ ($\mathbb{R}^p$, respectively), twice continuously differentiable, vanishing outside a compact interval and satisfying $\int v_j K(\mathbf{v})\,d\mathbf{v} = 0$ for $1 \le j \le pd$ ($\int v_j \bar{K}(\mathbf{v})\,d\mathbf{v} = 0$ for $1 \le j \le p$, respectively).

Assumption (B). The sequence of bandwidths h_n satisfies $0 < h_n \to 0$, $nh_n^{pd+2+\delta} \to \infty$ and $nh_n^{pd+4} \to 0$ as $n \to \infty$, where δ is as given above. Moreover, the bandwidth sequence $\bar{h}_n$ satisfies $0 < \bar{h}_n \to 0$, $n\bar{h}_n^{p+2+\delta} \to \infty$ and $n\bar{h}_n^{p+4} \to 0$ as $n \to \infty$, where δ is as given above.

Actually, we may assume the sequences mh_n^{pd+4} and $n\bar{h}_n^{p+4}$ tend to some non zero constant as in [27]. Then, the limiting distribution in Theorem 3 would become non centered. Nonetheless, under zero Assumptions $\mathcal{H}_0^{(1)}$ (or $\mathcal{H}_0^{(2)}$), this would change nothing because the latter additional term would be zero; see Theorem 3.1 in [27].

Let F_{nj} be the (marginal) empirical distribution function of X_j, based on the $(X_{1,j}, \dots, X_{n,j})$. Set

$$K_h(\mathbf{x}) = h^{-pd} K\left(\frac{x_1}{h}, \dots, \frac{x_{pd}}{h}\right), \qquad \bar{K}_{\bar{h}}(\mathbf{x}) = \bar{h}^{-p} \bar{K}\left(\frac{x_1}{\bar{h}}, \dots, \frac{x_p}{\bar{h}}\right).$$

For every $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y}^* \in \mathbb{R}^{pd}$, we can estimate the conditional distribution $m(\mathbf{x} \mid \mathbf{y}^*) = \mathbb{P}\left(\mathbf{X}_p \le \mathbf{x} \mid \mathbf{X}_0^{p-1} = \mathbf{y}^*\right)$ by

$$m_n(\mathbf{x} \mid \mathbf{y}^*) = \frac{1}{n-p} \sum_{\ell=0}^{n-p} K_n(\mathbf{X}_\ell^{\ell+p-1}) \mathbf{1}(\mathbf{X}_{\ell+p} \le \mathbf{x}),$$

where

$$K_n(\mathbf{X}_\ell^{\ell+p-1}) = K_h\{F_{n1}(X_{\ell 1}) - F_{n1}(y_{11}), \dots, F_{nd}(X_{\ell d}) - F_{nd}(y_{1d}), \dots, F_{n1}(X_{(\ell+p-1),1}) - F_{n1}(y_{p1}), \dots, F_{nd}(X_{(\ell+p-1),d}) - F_{nd}(y_{pd})\}.$$

Similarly, for all $x_j \in \mathbb{R}$ and $\mathbf{y}_j^* \in \mathbb{R}^p$, the conditional marginal cdf's

$$m_j(x_j \mid \mathbf{y}_j^*) = \mathbb{P}(X_{pj} \le x_j \mid \mathbf{X}_{0,j}^{p-1} = \mathbf{y}_j^*)$$

can be estimated in a nonparametric way by

$$m_{n,j}(x_j \mid \mathbf{y}_j^*) = \frac{1}{n-p}\sum_{\ell=1}^{n-p} \bar{K}_{\bar{h}}\{F_{nj}(X_{\ell,j}) - F_{nj}(y_{1j}), \ldots, F_{nj}(X_{\ell+p-1,j}) - F_{nj}(y_{pj})\}\mathbf{1}(\mathbf{X}_{\ell+p,j} \le x_j),$$

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

We propose to estimate the conditional pseudo-copula by

$$\widehat{C}(\mathbf{u} \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}^*) = m_n\{m_{n,1}^{(-1)}(u_1 \mid \mathbf{y}_1^*), \ldots, m_{n,d}^{(-1)}(u_d \mid \mathbf{y}_d^*) \mid \mathbf{y}^*\}. \tag{9}$$

Obviously, we use pseudo-inverse of conditional empirical functions in the latter equation.

Theorem 3. *Assume that the Assumptions (B), (K), (M), (R1), (R2) and (S) hold. Then, under $\mathcal{H}_0^{(1)}$, for all $\mathbf{u} \in [0,1]^d$ and $\mathbf{y}^* = (\mathbf{y}_1, \ldots, \mathbf{y}_p) \in \mathbb{R}^{dp}$,*

$$\sqrt{nh_n^{pd}}\{\widehat{C}(\mathbf{u} \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}^*) - C_0(\mathbf{u})\} \xrightarrow{\mathrm{d}} \mathcal{N}[0, \sigma(\mathbf{u})]$$

as $n \to \infty$, where $\sigma(\mathbf{u}) = C_0(\mathbf{u})\{1 - C_0(\mathbf{u})\}\int K^2(\mathbf{v})\,\mathrm{d}\mathbf{v}$.

This result can be extended to deal with different vectors $\mathbf{y}^*$ simultaneously. Moreover, we can consider the null hypotheses $\mathcal{H}_0^{(2)}$ and $\mathcal{H}_0^{(3)}$, provided we have an estimator of the true parameter that tends to the true value (in probability) faster than the estimators m_n and $m_{n,j}$.

Corollary 1. *Assume that the above assumptions hold for q different vectors $\mathbf{y}_k^* \in \mathbb{R}^{dp}$, $k = 1, \ldots, q$ and that the vector $(\mathbf{X}_1^p, \mathbf{X}_{1+m}^{p+m})$ has a continuous density in some open neighborhood of $(\mathbf{y}_k^*, \mathbf{y}_\ell^*)$ for every $m \ge 1$, $k, \ell = 1, \ldots, q$. Then, under $\mathcal{H}_0^{(3)}$ and for all $\mathbf{u} \in \mathbb{R}^d$, we have*

$$\sqrt{nh_n^{pd}}\,\{\widehat{C}(\mathbf{u} \mid \mathbf{y}_1^*) - C_{\hat{\theta}_1}(\mathbf{u}), \ldots, \widehat{C}(\mathbf{u} \mid \mathbf{y}_q^*) - C_{\hat{\theta}_q}(\mathbf{u})\} \xrightarrow{\mathrm{d}} \mathcal{N}[0, \Sigma(\mathbf{u}, \mathbf{y}_1^*, \ldots, \mathbf{y}_q^*)],$$

as $n \to \infty$, where

$$\Sigma(\mathbf{u}, \mathbf{y}_1^*, \ldots, \mathbf{y}_q^*) = \operatorname{diag}\left(C_{\theta(\mathbf{y}_k^*)}(\mathbf{u})\{1 - C_{\theta(\mathbf{y}_k^*)}(\mathbf{u})\}\int K^2(\mathbf{v})\,\mathrm{d}\mathbf{v},\ 1 \le k \le q\right),$$

for some consistent estimators $\hat{\theta}_k$ such that

$$\hat{\theta}_k = \theta(\mathbf{y}_k^*) + O_P(n^{-1/2}), \quad k = 1, \ldots, q.$$

Each kth term on the diagonal of Σ can be consistently estimated by

$$\hat{\sigma}_k^2(\mathbf{u}) = C_{\hat{\theta}_k}(\mathbf{u})\{1 - C_{\hat{\theta}_k}(\mathbf{u})\}\int K^2(\mathbf{v})\,\mathrm{d}\mathbf{v}.$$

Note that, in the corollary above, the limiting correlation matrix is diagonal because we are considering different conditioning values $\mathbf{y}_1^*, \ldots, \mathbf{y}_q^*$ but the same argument $\mathbf{u}$. At the opposite, an identical conditioning event but different arguments $\mathbf{u}_1, \mathbf{u}_2, \ldots$ would lead to a complex (non diagonal) correlation matrix, as explained in [11].

Proof. Under the above assumptions, Mehra et al. [27] proved that, for every $\mathbf{x} \in \mathbb{R}^d$ and every real number t, we have

$$\mathbb{P}\left[\sqrt{nh_n^{pd}}\left\{m_n(\mathbf{x} \mid \mathbf{y}^*) - m(\mathbf{x} \mid \mathbf{y}^*)\right\} \le t\tau_{\mathbf{x},\mathbf{y}^*}\right] \to \Phi(t), \tag{10}$$

where $\Phi(t)$ is the cdf of the standard Gaussian distribution, and

$$\tau_{\mathbf{x},\mathbf{y}^*}^2 = m(\mathbf{x} \mid \mathbf{y}^*)\left\{1 - m(\mathbf{x} \mid \mathbf{y}^*)\right\}\int K^2(\mathbf{v})\,\mathrm{d}\mathbf{v}.$$

Similarly, it follows that

$$\sqrt{nh_n^{p}}\left\{m_{nj}(x_j \mid \mathbf{y}_j^*) - F_j(x_j \mid \mathbf{y}_j^*)\right\}$$

converges to a Gaussian limit. Since $K \ge 0$ (K is a probability density), m_n is monotone increasing. Consequently, the convergence in (10) holds uniformly in the argument t.

Careful inspection of the proofs of Lemmas 3.1 and 3.2 in [27] reveals that

$$\sup_{x_j \in \mathbb{R}}\left|m_{nj}(x_j \mid \mathbf{y}_j^*) - F_j(x_j \mid \mathbf{y}_j^*)\right| \to 0,$$

in probability, as $n \to \infty$, where we used $F_{nj}(x_j) \to F_j(x_j)$, in probability, as $n \to \infty$, uniformly in x_j, $1 \le j \le d$. The conditions on F_j and a standard technique (see, e.g., [35, Example 1, p. 7]) implies that

$$m_{nj}^{(-1)}(u_j \mid \mathbf{y}_j^*) \to F_j^{(-1)}(u_j \mid \mathbf{y}_j^*), \tag{11}$$

in probability, as $n \to \infty$. The theorem follows immediately from the weak convergence (10), the convergence in probability (11) and the continuity of the normal distribution. □

As in [11], a simple test procedure may be based on

$$\mathcal{T}(\mathbf{u}, \mathbf{y}_1^*, \ldots, \mathbf{y}_q^*) = (nh_n^{pd}) \sum_{k=1}^{q} \frac{\{\widehat{C}(\mathbf{u} \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}_k^*) - C_{\hat{\theta}_k}(\mathbf{u})\}^2}{\hat{\sigma}^2_{\mathbf{y}_k^*}(\mathbf{u})},$$

for different choices of $\mathbf{u}$ and conditioning values $\mathbf{y}_k^*$. Under $\mathcal{H}_0^{(1)}$, the term on the right-hand-side tends to a χ_q^2 distribution under the null hypothesis.

Note that this test is "local" since it depends strongly on the choice of a single $\mathbf{u}$. An interesting extension would be to build a "global" test, based on the behavior of the full process

$$\sqrt{nh_n^{pd}}\,\{\widehat{C}(\cdot \mid \mathbf{X}_{n-1}^{n-p} = \mathbf{y}_k^*) - C_{\hat{\theta}_k}(\cdot)\}.$$

But the task of getting pivotal limiting laws is far from easy, as illustrated in [11]. Most of the alternative GOF test statistics rely on the (non-smoothed) empirical copula process and are bootstrap-based (see [13] for a survey). Whether some of them can be adapted in the pseudo-copula framework remains an open question.

In the case of parametric pseudo-copula models C_θ with $\theta \in \Theta \subseteq \mathbb{R}^q$, a semi-parametric estimation procedure can easily be implemented, as in [3]. Indeed, the log-likelihood can be split still into a part involving a pseudo-copula and a part involving the marginal distributions. For instance, we can estimate the marginal cdf's by the functions $m_{n,j}$ above, and find $\hat{\theta}$ by maximizing (over $\theta \in \Theta$)

$$\sum_{k=p}^{n} \ln \tau_\theta \left\{ m_{n,1}(X_{k,1} \mid \mathbf{X}_{k-p,1}^{k-1}), \ldots, m_{n,d}(X_{k,d} \mid \mathbf{X}_{k-p,d}^{k-1}) | \mathbf{X}_{k-p}^{k-1} \right\},$$

with

$$\tau_\theta(\mathbf{u}|\mathbf{X}_{k-p}^{k-1}) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C_\theta(\mathbf{u}|\mathbf{X}_{k-p}^{k-1}).$$

We leave the study of such a procedure for further research.

4. A small simulation study

We evaluate the power of our test statistics $\mathcal{T}$ in the case of a bivariate stationary process ($d = 2$). We restrict ourselves to $p = 1$, i.e., only the last observed value is assumed to influence the dynamics of the dependence between $X_{n,1}$ and $X_{n,2}$. The null hypothesis states that the bivariate pseudo-copula of $(X_{n,1}, X_{n,2})$ is independent of the past values $(X_{n-1,1}, X_{n-1,2})$ and that it is a Gaussian copula, viz.

$$\mathcal{H}_0 : C(\mathbf{u}|\mathbf{y}) = C_G(\mathbf{u}) \quad \text{for all } \mathbf{u} \in [0, 1]^2 \text{ and } \mathbf{y} \in \mathbb{R}^2 \text{ for some Gaussian copula } C_G.$$

Note that the Gaussian copula above does not depend on $\mathbf{y}$.

The data are generated from the bivariate auto-regressive process

$$\begin{cases} X_{n,1} = aX_{n-1,1} + bX_{n-1,2} + \varepsilon_{n,1}, \\ X_{n,2} = bX_{n-1,1} + aX_{n-1,2} + \varepsilon_{n,2}, \end{cases}$$

where the residuals are standard correlated white noises, with $\mathrm{corr}(\varepsilon_{n,1}, \varepsilon_{n',2}) = \rho\delta_{n,n'}$ for all n and n'. We assume that $a + b < 1$ and that $\rho = 0.5$. The initial vector $(X_{0,1}, X_{0,2})$ is drawn from the stationary law. We choose a set of past observed values to set the conditioning arguments $\mathbf{y}_k^* = (i, j),\ i, j \in \{-1, 0, 1\}$, for $1 \le k \le 9$. This provides a 9-point grid ($q = 9$ in the notation of the previous section).

Moreover, to assess the sensitivity of our results with respect to the choice of $\mathbf{u}$, we choose nine different values for this bivariate vector $\mathbf{u} = (u_1, u_2) : u_1 \in \{0.1, 0.5, 0.9\}, u_2 \in \{0.0, 0.4, 0.8\}$. We consider time series of length N. The bandwidths h and $\bar{h}$ have been chosen by the standard "rule of thumb" [37], as if we were dealing with the densities of the "pseudo-sample" $(F_{N,1}(X_{n,1}), F_{N,2}(X_{n,2}))$, $n = 1, \ldots, N$. All the considered kernels are Gaussian. Since the empirical standard deviation $\hat{\sigma}$ of the two previous marginal distributions are the same, we set

$$h^* = \frac{2\hat{\sigma}}{N^{1/6}}, \qquad \bar{h}^* = \frac{\hat{\sigma}}{N^{1/5}}.$$

Here, with a sample size of $N = 5000$ points, $\hat{\sigma} \simeq 0.288$, $h^* \simeq 0.139$ and $\bar{h}^* \simeq 0.0524$.

Table 1
Percentages of rejection of $\mathcal{H}_0$ at 5% level with $N = 5000$ and 500 replications (test statistic $\mathcal{T}$).

(a, b)	**u**								
	(0.2, 0.1)	(0.2, 0.5)	(0.2, 0.9)	(0.5, 0.1)	(0.5, 0.5)	(0.5, 0.9)	(0.8, 0.1)	(0.8, 0.5)	(0.8, 0.9)
(0, 0)	10.2	11.4	14.2	13.8	12.0	11.8	10.0	9.2	14.4
(0, 0.2)	89.4	97.2	98.4	99.2	98.6	99.4	99.2	98.4	99.2
(0, 0.4)	99.0	99.8	100.0	100.0	99.8	99.8	99.6	99.6	99.6
(0, 0.6)	99.6	100.0	99.8	100.0	100.0	100.0	100.0	100.0	99.4
(0, 0.8)	99.4	100.0	100.0	100.0	100.0	99.8	100.0	99.8	100.0
(0.3, 0.0)	7.0	6.2	9.2	9.0	10.4	6.2	6.8	7.6	5.6
(0.3, 0.2)	97.2	99.8	100.0	99.8	100.0	100.0	100.0	99.6	100.0
(0.3, 0.4)	99.0	100.0	99.8	100.0	100.0	100.0	99.8	100.0	100.0
(0.3, 0.6)	99.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.6
(0.6, 0.0)	6.2	4.8	4.0	6.8	3.0	3.8	4.2	4.2	3.2
(0.6, 0.2)	95.8	100.0	99.8	100.0	99.8	100.0	100.0	100.0	99.4
(0.9, 0.0)	5.0	2.6	6.6	2.8	4.2	3.6	2.0	2.4	4.0

Table 2
Percentages of rejection of $\mathcal{H}_0$ at 5% level with $(a, b) = (0.3, 0.2)$, for several sample sizes N and 500 replications (test statistic $\mathcal{T}$).

Sample size	**u**								
	(0.2, 0.1)	(0.2, 0.5)	(0.2, 0.9)	(0.5, 0.1)	(0.5, 0.5)	(0.5, 0.9)	(0.8, 0.1)	(0.8, 0.5)	(0.8, 0.9)
200	19.8	41.8	90.2	48.0	23.6	83.6	77.0	63.8	53.2
1000	49.0	92.2	97.2	81.6	80.6	99.0	98.4	99.0	88.4
5000	97.2	99.8	100.0	99.8	100.0	100.0	100.0	99.6	100.0

Table 3
Percentages of rejection of $\mathcal{H}_0$ at 5% level with $(a, b) = (0.3, 0.0)$, for several sample sizes N and 500 replications (test statistic $\mathcal{T}$).

Sample size	**u**								
	(0.2, 0.1)	(0.2, 0.5)	(0.2, 0.9)	(0.5, 0.1)	(0.5, 0.5)	(0.5, 0.9)	(0.8, 0.1)	(0.8, 0.5)	(0.8, 0.9)
200	14.4	14.8	14.0	12.0	9.8	8.4	14.0	16.8	10.2
1000	10.4	10.2	10.0	9.2	8.4	7.2	12.0	6.6	5.8
5000	7.0	6.2	9.2	9.0	10.4	6.2	6.8	7.6	5.6

In each case, we calculate the test statistics $\mathcal{T}(\mathbf{u}, \mathbf{y}^*)$ and compare it with the 0.95 quantile of a chi-square distribution with 9 degrees of freedom. Table 1 provides the proportion of rejection of this test. In average, this proportion is close to 5% of the times when H_0 is true ($b = 0$). More precisely, it is higher for small a, and smaller for high a. Otherwise, the test rejects the null hypothesis most of the time (for different values $b \neq 0$).

We observe that the frequencies of rejection are almost independent of the value of a and that it is slightly increasing with b, what is in line with intuition. The performances depend a bit on the chosen argument $\mathbf{u}$. The best ones are obtained for $\mathbf{u}$ values that lie far from the boundaries of the unit square $[0, 1]^2$. This may be due to the well-known bias of kernel estimates on the boundaries of finite support distributions. Here, the test could surely be improved by modifying our kernels in the lines of Omelka et al. [31]. Overall, the power of our test seems to be reasonable, keeping in mind the complexity of the test procedure.

Clearly, to get reliable results, this test procedure must be led with sample sizes larger than several thousand points, as it is shown in Tables 2 and 3. Of course, as usual in such kernel-based tests, our results depend on the bandwidth choices. We have observed that the performances of our previous test does not seem to be too sensitive to the choices of h^* and $\bar{h}^*$, when they remain "reasonable". This is particularly true concerning the power of this test. When we "over-smooth" the joint underlying distribution and "under-smooth" the margins, we tend to falsely reject more frequently the null hypothesis. In practice, we advise to perform the test $\mathcal{T}(\mathbf{u}, \mathbf{y}^*)$ for several values $\mathbf{u}$ and several bandwidth choices.

Acknowledgments

The authors thank Christian Genest, the Associate Editor, the referees and the participants of the Workshop "Copula Models and Dependence" (Montréal 2011) for their valuable and constructive remarks.

References

[1] E. Acar, R. Craiu, F. Yiao, Dependence calibration in conditional copulas: a nonparametric approach, Biometrics 67 (2010) 445–453.
[2] B. Beare, Copulas and temporal dependence, Econometrica 78 (2010) 395–410.
[3] X. Chen, Y. Fan, Estimation of copula-based semiparametric time series models, Journal of Econometrics 130 (2006) 307–335.
[4] U. Cherubini, E. Luciano, Value at risk trade-off and capital allocation with copulas, Economic Notes 30 (2000) 235–256.
[5] U. Cherubini, E. Luciano, Copula Methods in Finance, Wiley, 2004.
[6] U. Cherubini, S. Mulinacci, F. Gobbi, S. Romagnoli, Dynamic Copula Methods in Finance, Wiley, 2011.

[7] W.F. Darsow, B. Nguyen, E.T. Olsen, Copulas and Markov processes, Illinois Journal of Mathematics 36 (1992) 600–642.
[8] P. Doukhan, J.-D. Fermanian, G. Lang, An empirical central limit theorem with applications to copulas under weak dependence, Statistical Inference for Stochastic Processes 12 (2009) 65–87.
[9] P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, Springer, 1997.
[10] P. Embrechts, A. McNeil, D. Straumann, Correlation and dependence in risk management: properties and pitfalls, in M. Dempster (Ed.), Risk Management: value at Risk and Beyond, Cambridge UP, 2002.
[11] J.-D. Fermanian, Goodness-of-fit tests for copulas, Journal of Multivariate Analysis 95 (2005) 119–152.
[12] C. Genest, A.-C. Favre, Everything you always wanted to know about copula modeling but were afraid to ask, Journal of Hydrologic Engineering 12 (2007) 347–368.
[13] C. Genest, B. Rémillard, D. Beaudoin, Goodness-of-fit tests for copulas: a review and a power study, Insurance: Mathematics and Economics 44 (2009) 199–213.
[14] J. Hidalgo, Nonparametric test for causality with long-range dependence, Econometrica 68 (2000) 1465–1491.
[15] C. Hiemstra, J.D. Jones, Testing for linear and nonlinear Granger causality in the stock price volume relation, Journal of Finance 49 (1994) 1639–1664.
[16] R. Ibragimov, Copula based characterizations for higher-order Markov processes, Econometric Theory 25 (2009) 819–846.
[17] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, 1997.
[18] E. Jondeau, M. Rockinger, The copula-GARCH model of conditional dependencies: an international stock market application, Journal of International Money and Finance 25 (2006) 827–853.
[19] C.H. Kimberling, A probabilistic interpretation of complete monotonicity, Aequationes Mathematicae 10 (1974) 152–164.
[20] N. Kolev, U. dos Anjos, B.V.de M. Mendes, Copulas: a review and recent developments, Stochastic Models 22 (2005) 617–660.
[21] E. Kole, K. Koedijk, M. Verbeek, Selecting copulas for risk management, Journal of Banking and Finance 31 (2007) 2405–2423.
[22] D. Kurowicka, H. Joe, Dependence Modelling, Vine Copula Handbook, World Scientific, 2011.
[23] X. Luo, P. Shevchenko, The t copula with multiple parameters of degrees of freedom: bivariate characteristics and application to risk management, Quantitative Finance 10 (2010) 1039–1054.
[24] T. Markwat, E. Kole, D. van Dijk, Time variation in asset return dependence: strength or structure? Working paper, Erasmus, 2009.
[25] A.J. McNeil, R. Frey, P. Embrechts, Quantitative Risk Management: concepts, Techniques and Tools, Princeton Series in Finance, 2005.
[26] A.J. McNeil, J. Nešlehová, Multivariate Archimedean copulas, d-Monotone functions and ℓ_1-norm symmetric distributions, Annals of Statistics 37 (2009) 3059–3097.
[27] K.L. Mehra, Y.S. Ramakrishnaiah, M. Sudhakara Rao, Asymptotic normality of rank nearest neighbor regression function estimators under strong mixing, Yokohama Mathematical Journal 39 (1991) 49–60.
[28] B.V.de M. Mendes, Computing conditional VaR using time-varying copulas, Brazilian Review of Finance 3 (2005) 251–265.
[29] R.B. Nelsen, An Introduction to Copulas, Springer, 1999.
[30] Y. Nishiyama, K. Hitomi, Y. Kawasaki, K. Jeong, A consistent nonparametric test for nonlinear causality specification in time series, Journal of Econometrics (2011) (in press).
[31] M. Omelka, I. Gijbels, N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing, Annals of Statistics 37 (2009) 3023–3058.
[32] A. Patton, Modelling asymmetric exchange rate dependence, International Economic Review 47 (2006) 527–556.
[33] A. Patton, Estimation of multivariate models for time series of possibly different lengths, Journal of Applied Econometrics 21 (2006) 147–173.
[34] A. Patton, Copula based models for financial time series, in: T.G. Andersen, R.A. Davis, J.-P. Kreiss, T. Mikosch (Eds.), Handbook of Financial Time Series, Springer, 2009, pp. 767–785.
[35] D.B. Pollard, Convergence of Stochastic Processes, Springer, 1984.
[36] J.V. Rosenberg, Nonparametric pricing of multivariate contingent claims, Journal of Derivatives 10 (2003) 9–26.
[37] D. Scott, Multivariate Density Estimation, Wiley, 1992.
[38] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de l'Université de Paris 8 (1959) 229–231.
[39] R.W.J. Van den Goorbergh, C. Genest, B.J.M. Werker, Bivariate option pricing using dynamic copula models, Insurance: Mathematics and Economics 37 (2005) 101–114.