DE GRUYTER OPEN Depend. Model. 2015; 3:59–82

Research Article Open Access

Clarence Simard* and Bruno Rémillard

Forecasting time series with multivariate copulas

DOI 10.1515/demo-2015-0005
Received August 17, 2014; accepted May 15, 2015

Abstract: In this paper we present a forecasting method for time series using copula-based models for multivariate time series. We study how the performance of the predictions evolves when changing the strength of the different possible dependencies, as well as the structure of the dependence. We also look at the impact of the marginal distributions. The impact of estimation errors on the performance of the predictions is also considered. In all the experiments, we compare predictions from our multivariate method with predictions from the univariate version which has been introduced in the literature recently. To simplify implementation, a test of independence between univariate Markovian time series is proposed. Finally, we illustrate the methodology by a practical implementation with financial data.

Keywords: Copulas; time series; forecasting; realized volatility

MSC: 62M20

1 Introduction

For many years, copulas have been used for modeling dependence between random variables. See e.g. [16] for a survey on copulas in finance. The possibility to model the dependence structure independently from marginal distributions allows for a better understanding of the dependence structure and a wide range of joint distributions. More recently, copulas have been used to model the temporal dependence in time series, first in the univariate case, as in [9] and [5], and then in a multivariate setting [23]. Once again, the flexibility of copulas allows to model more complex dependence structures and thus to better capture the evolution of the time series. In the recent work of [26], copulas were used to forecast the realized volatility associated with a univariate financial time series and they showed that copula-based forecasts perform better than forecasts based on heterogeneous autoregressive (HAR) model. The later method had been proven successful in [3], [10] and [8].

In this paper, we extend the methodology of [26] by proposing a forecasting method using copula-based models for multivariate time series, as in [23]. As one can guess, we show that forecasting multivariate time series using copula-based models gives better results than forecasting a single time series, since more information means more precision, in general. For example, let $\{(X_{1,t}, X_{2,t}); t \in \mathbb{N}\}$ be two dependent time series with both series showing temporal dependence. Suppose one wants to forecast $X_{1,T+1}$ based on the information available at period T. We show that forecasting the joint values of $(X_{1,T+1}, X_{2,T+1})$ using the observed values $(X_{1,T}, X_{2,T})$ gives significantly better predictions of $X_{1,T+1}$ in general than predictions on $X_{1,T+1}$ based only on the single value of $X_{1,T}$, which of course has to be expected. Since $\{X_{1,t}\}$ and $\{X_{2,t}\}$ are dependent and temporally dependent, the knowledge of $(X_{1,T}, X_{2,T})$ gives more information in general than the knowledge of $X_{1,T}$ alone.

***Corresponding Author: Clarence Simard:** Department of Mathematics and Statistics, Université de Montréal, E-mail: clarence.simard@hotmail.com
Bruno Rémillard: Department of Decision Sciences, HEC Montréal, E-mail: bruno.remillard@hec.ca

We also study the impact of the strength of the different dependencies, the structure of the dependencies as well as the impact of marginal distributions of the vector $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ on the performance of the predictions.

Although our numerical experiments focus on the bivariate case, our presentation can be readily extended to an arbitrary number of dimensions. Actually, the results of [23], which provide basis for the estimation methods, are stated for an arbitrary number of time series and most of the theoretical background is going to be presented here in multivariate case. Moreover, the results of our numerical experiments should naturally extend to the arbitrary dimensions.

The rest of the paper is structured as follows. In Section 2 we give some basic results about copulas and apply the results to model time series. In Section 2.3, we define our forecasting methods and its implementation is discussed in Section 2.4, where a new test of independence between copula-based Markovian time series is given. Section 3 contains the results of the numerical experiments as well as the analysis of the results. Finally, an example of implementation with real data is given in Section 4. The last section contains some concluding remarks.

2 Modeling time series with copulas

2.1 Copulas

We begin by giving some definitions and basic results about copulas. More details can be found in [21] and [22].

Definition 2.1. *(Copula)*
A d-dimensional copula is a distribution function with domain $[0, 1]^d$ and uniform margins.

Equivalently, the function $C : [0, 1]^d \to [0, 1]$ is a d-dimensional copula if and only if there exists random variables $U_1, \ldots, U_d$ such that $P(U_j \le u) = u_j$ for $j \in \{1, \ldots, d\}$ and $C(\mathbf{u}) = P(U_1 \le u_1, \ldots, U_d \le u_d)$ for all $\mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d$. The existence of a copula function for any joint distribution is given by the Sklar's theorem.

Theorem 2.1. *(Sklar's theorem)*
Let $X_1, \ldots, X_d$ be d random variables with joint distribution function H and margins $F_1, \ldots, F_d$. Then, there exists a d-dimensional copula C such that for all $(x_1, \ldots, x_d) \in \bar{\mathbb{R}}^d$,

$$H(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \tag{1}$$

where $\bar{\mathbb{R}} = \mathbb{R} \bigcup \{-\infty, \infty\}$.

We note that the copula function in (1) is uniquely defined on the set $Range(F_1) \times \ldots \times Range(F_d)$. Hence, if $Range(F_j) = [0, 1]$ for $j \in \{1, \ldots, d\}$ the copula is unique.

We define the left continuous inverse of a distribution function F as

$$F^{-1}(u) = \inf\left\{x; F(x) \ge u\right\}, \text{ for all } u \in (0, 1).$$

Using this inverse and Sklar's theorem, we have a way to define the copula function in terms of the quasi-inverses and the joint distribution.

Assuming that the density f_j of F_j exists for each $j \in \{1, \ldots, d\}$, then the density c of C exists if and only if the density h of H also exists. In this case, differentiating one obtains

$$h(x_1, \ldots, x_d) = c\{F_1(x_1), \ldots, F_d(x_d)\} \prod_{j=1}^{d} f_j(x_j).$$

Furthermore, for all $(u_1, \dots, u_d) = (0,1)^d$,

$$c(u_1, \dots, u_d) = \frac{h\left\{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\right\}}{\prod_{j=1}^{d} f_j\left\{F_j^{-1}(u_i)\right\}}. \tag{2}$$

Following the example of conditional distributions it is also possible to define conditional copulas. Let $(\mathbf{X}, \mathbf{Y})$ be a $(d_1 + d_2)$-dimensional random vector with joint distribution H, where $\mathbf{X}$ has marginal distributions $F_1, \dots, F_{d_1}$ and $\mathbf{Y}$ has marginal distributions $G_1, \dots, G_{d_2}$. Setting $\mathbf{F}(\mathbf{X}) = (F_1(X_1), \dots, F_{d_1}(X_{d_1}))$, $\mathbf{G}(\mathbf{Y}) = (G_1(Y_1), \dots, G_{d_2}(Y_{d_2}))$ and defining the random vector $(\mathbf{U}, \mathbf{V}) = (\mathbf{F}(\mathbf{X}), \mathbf{G}(\mathbf{Y}))$, we can define the copula $C_{\mathbf{UV}}$ of the vector $(\mathbf{X}, \mathbf{Y})$ as the joint distribution function of $(\mathbf{U}, \mathbf{V})$. Assuming that the density functions exist and applying (2), one obtains that the conditional copula $C_{\mathbf{U}|\mathbf{V}}$, i.e., the conditional distribution of $\mathbf{U}$ given $\mathbf{V}$, is given by $C_{\mathbf{U}|\mathbf{V}}(\mathbf{u};\mathbf{v}) = \frac{\partial_{v_1}\cdots\partial_{v_{d_2}} C_{\mathbf{UV}}(\mathbf{u},\mathbf{v})}{c_{\mathbf{V}}(\mathbf{v})}$, with density $c_{\mathbf{U}|\mathbf{V}}(\mathbf{u};\mathbf{v}) = \frac{c_{\mathbf{UV}}(\mathbf{u},\mathbf{v})}{c_{\mathbf{V}}(\mathbf{v})}$, where $c_{\mathbf{V}}$ is the density of the copula $C_{\mathbf{V}}(\mathbf{v}) = C_{\mathbf{UV}}(1, \dots, 1, v)$ of $\mathbf{Y}$ or $\mathbf{V}$.

Having defined conditional copulas, one can now look at copula-based models for multivariate time series.

2.2 Modeling time series

In order to get a prediction method, we first need to present how to use copulas for modeling time series. The ideas presented here were developed in [27] and [23], extending the results of [9] to the multivariate case.

Let $\mathbf{X} = \{\mathbf{X}_t; t \in \mathbb{N}\}$ be a d-dimensional time series and assume that $\mathbf{X}$ is Markovian and stationary. We note F_j the marginal distribution of $X_{j,t}$ for $j \in \{1, \dots, d\}$ and H the joint distribution of $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ and assume that all distributions are continuous. From the stationarity assumption, it follows that all distribution functions $F_1, \dots, F_d$ and H are time-independent. Using Sklar's theorem, there is a unique copula C associated to $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ and unique copula Q associated to $\mathbf{X}_{t-1}$, viz.

$$Q(\mathbf{u}) = C(\mathbf{1}_d, \mathbf{u}) = C(\mathbf{u}, \mathbf{1}_d) \text{ for all } \mathbf{u} \in [0,1]^d,$$

where $\mathbf{1}_d$ is the d-dimensional unit vector. Set $\mathbf{U}_t = \mathbf{F}(\mathbf{X}_t)$, for $t \geq 1$.

The next step is to deduce the conditional copula of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$, which is $\mathcal{C}(\mathbf{u};\mathbf{v}) = C_{\mathbf{U}_t|\mathbf{U}_{t-1}}(\mathbf{u};\mathbf{v}) = \frac{\partial_{v_1}\cdots\partial_{v_d} C(\mathbf{u},\mathbf{v})}{q(\mathbf{v})}$, with density $c_{\mathbf{U}_t|\mathbf{U}_{t-1}}(\mathbf{u};\mathbf{v}) = \frac{c(\mathbf{u},\mathbf{v})}{q(\mathbf{v})}$, where q is the density of Q.

Combining the knowledge of the marginal distributions and the conditional copula above, we can get the conditional distribution of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$. This is what we use to define our predictions.

2.3 Forecasting method

To expose our forecasting method, we first make the assumption that the joint distribution as well as the marginal distributions of the time series are known. However, in practice, these distributions are unknown and estimations must be done. This will be covered next. The use of a copula-based model for time series allows for a more flexible model of the dependence structure.

Let $\mathbf{X} = \{\mathbf{X}_t; t = 0, 1, .., T\}$ be a d-dimensional time series. Our aim is to forecast $\mathbf{X}_{T+1}$ based on the information available at time T. Suppose that for all $t \geq 0$, F_j is the marginal distribution of $X_{j,t}$ for $j \in \{1, \dots, d\}$, and the $2d$-dimensional vector $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ as joint distribution H and copula C. Using the results of the previous section, we can define the conditional copula $\mathcal{C}$ of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$, namely $C_{\mathbf{U}_t|\mathbf{U}_{t-1}}(\mathbf{u};\mathbf{v})$.

Now suppose we observed $\mathbf{X}_T = \mathbf{y}$ at time T. The prediction of $\mathbf{X}_{T+1}$ goes as follows:

1. Set $\mathbf{v} = \mathbf{F}(\mathbf{y})$.
2. Simulate n realizations of the conditional copula, $\mathbf{U}^{(i)} \sim \mathcal{C}(\cdot;\mathbf{v})$, $i \in \{1, \dots, n\}$.
3. For each $i \in \{1, \dots, n\}$, set $\mathbf{X}_{T+1}^{(i)} = \mathbf{F}^{-1}\left(\mathbf{U}^{(i)}\right)$.

4. Define
$$\hat{\mathbf{X}}_{T+1} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_{T+1}^{(i)}. \tag{3}$$
We use $\hat{X}_{1,T+1}$ as a predictor for $X_{1,T+1}$.

4′ One can also define a prediction interval of level $1-\alpha \in (0,1)$ for $X_{1,T+1}$ by taking the estimated quantiles of order $\alpha/2$ and $1-\alpha/2$ amongst $\{X_{1,T+1}^{(i)}; i \in \{1,\ldots,n\}\}$. We denote by $\widehat{LB}_{T+1}^{\alpha}$ and $\widehat{UB}_{T+1}^{\alpha}$ the lower and upper values for the prediction interval.

As mentioned previously, we are going to compare our predictions performance with the univariate version presented in [26], when $(X_{1,t})$ is a Markov process. In this case, let D be the copula associated with $(X_{1,t-1}, X_{1,t})$ for $t \in \mathbb{N}$, i.e., D is the copula of $(U_{1,t-1}, U_{1,t})$.

Suppose we observe the value $X_{1,T} = y_1$; The predictor presented in [26] is defined by
$$\bar{X}_{1,T+1} = n^{-1}\sum_{i=1}^{n} F_1^{-1}\left(Z^{(i)}\right), \tag{4}$$
where $Z^{(i)}$ are realizations of $\mathcal{D}(\cdot; v_1)$ with $v_1 = F_1(y_1)$ and $\mathcal{D}$ is the conditional copula associated with $X_{1,t}$ given $X_{1,t-1}$. As before, we can define the prediction interval using the estimated $\alpha/2$ and $1-\alpha/2$ quantiles from the values $F_1^{-1}\left(Z^{(i)}\right)$, $i \in \{1,\ldots,n\}$. We note the upper and lower bound of predictions interval at time $T+1$ by $\overline{UB}_{T+1}^{\alpha}$ and $\overline{LB}_{T+1}^{\alpha}$.

2.4 Implementation in practice

There are several interesting problems for practical implementations. First, there is the choice of the copula and marginal distributions. Then, there is also the problem of choosing the simplest model, i.e. a model with as few explanatory variables as possible. These problems are discussed next, together with a new test of independence between univariate copula-based Markovian time time series.

2.4.1 Choosing the marginal distributions and the copula family

When the marginal distributions $\mathbf{F}$ are not known, they can be estimated by the empirical marginal distributions $F_{T,j}$, $j \in \{1,\ldots,d\}$, where
$$F_{T,j}(x) = \frac{1}{T+1}\sum_{t=1}^{T} \mathbb{I}(X_{j,t} \le x), \quad x \in \mathbb{R}. \tag{5}$$
The validity of the estimation was discussed in [23]; see also [24]. Basically all is needed is alpha-mixing. This hypothesis is met for all models considered here.

Remark 1. *One could also use parametric estimates for the margins. This would possibly lead to narrower prediction intervals.*

Next comes the choice of the copula family. Again, this was discussed in [23], and it was suggested to use non-parametrical estimation for marginal distributions, as given by (5), together with parametric estimation for joint copula models. Goodness-of-fit tests are also provided in [23] to help choose the right copula family. Note that in a forecasting context, one could also choose copulas by their prediction power, using e.g., [11].

Given a copula-based model, one can also ask what is the best set of predictive variables. This is discussed next.

2.4.2 The selection problem

Consider the simple case of a bivariate model with variables X_1 and X_2. How can one choose between the univariate model, i.e., predicting X_1 using only its past values, and the bivariate model, i.e., using the past values of (X_1, X_2) to predict X_1?

In the case of ARMA(p,q) time series, [7] proposed to use the AIC criterion to choose the correct order of p and q and then verify the adequacy of the model using a test of goodness-of-fit. In their setting, the models are imbedded. This is not necessarily the case in our setting. For example, the bivariate series (X_1, X_2) can be a Markov process, without X_1 being Markovian. In this case, to predict X_1, one needs the past values of X_1 and X_2, not only the past values of X_1. Therefore, in order to have to choose between two models, one assumes that X_1 (Model 1) and (X_1, X_2) (Model 2) are both Markov processes.

Remark 2. *If both X_1 and X_2 are Markov processes, one should first test for independence between the two series, as proposed in Section 2.4.3. If the null hypothesis of independence is not rejected, then choose model 1. Otherwise, for any $u_1, v_1 \in (0,1)$, there exists a copula C_{12,u_1,v_1} so that*

$$P(U_2 \le u_2, V_2 \le v_2 | U_1 = u_1, V_1 = v_1) = C_{12,u_1,v_1}\left\{E_1(u_1,u_2), E_2(v_1,v_2)\right\}, \tag{6}$$

where $E_1(u_1,u_2) = P(U_2 \le u_2 | U_1 = u_1)$ and $E_2(v_1,v_2) = P(V_2 \le v_2 | V_1 = v_1)$ are the associated Rosenblatt transforms.

Under the simplifying assumption that the copula C_{12,u_1,v_1} is independent of u_1, v_1, (6) yields

$$P(U_2 \le u_2, V_2 \le v_2 | U_1 = u_1, V_1 = v_1) = C_{12}\left\{E_1(u_1,u_2), E_2(v_1,v_2)\right\}.$$

Therefore, one obtains a model similar to the ones used for vine copulas, under the simplifying assumption. Recall that vine models are graphical structures with building blocks given by bivariate copulas. For example, for a three dimensional model, one could write the joint copula density as a product of the form

$$c(u,v,w) = c_{23|1}\left\{F_{2|1}(v|u), F_{3|1}(w|u)|u\right\} c_{12}(u,v)c_{13}(u,w),$$

where C_{12}, C_{13} and $C_{23|1}(v,w|u)$ are copulas, $F_{2|1}(v|u) = \partial_u C_{12}(u,v)$, $F_{3|1}(w|u) = \partial_u C_{13}(u,w)$, $c_{12}(u,v) = \partial_v F_{2|1}(v|u)$, and $c_{13}(u,w) = \partial_w F_{3|1}(w|u)$. In this context, the simplifying assumption means that the copula $C_{23|1}$ does not depend on u. For more details on the vast topic of vine copula models, see, e.g., [1] and [19].

However, applying our methodology to this model leads to the same predictions as in model 1! For, to generate (U_2, V_2) given $(U_1, V_1) = (u_1, v_1)$, one first generate $(W_1, W_2) \sim C_{12}$, and solve $W_1 = E_1(u_1, U_2)$, $W_2 = E_1(v_1, V_2)$. Hence the prediction of U_2 does not depend on V_1 and it is the same prediction as in Model 1.

As said before, in order to be able to use a selection criteria similar to the AIC, the class of copulas families would have to be more limited, Model 1 being imbedded in Model 2. One such way could be to consider appropriately chosen vine copula models (different from the model presented in Remark 2). From a modeling point of view, there are some preliminary results in this direction; see, e.g., [25], [14], and [6]. This way, the models could be embedded and the AIC criterion could be applied.

If the models are not embedded, then one could base the choice of the prediction power, using a given accuracy measure, as in [11].

2.4.3 A new test of independence between univariate Markovian time series

Suppose that the univariate series $X_1, \ldots, X_d$ are Markovian, with associated bivariate copula families C_{j,α_j}, $j \in \{1, \ldots, d\}$. One way of selecting the simple model based only on X_1, instead of a model based on $(X_1, X_2, \ldots, X_d)$ would be to test independence between the series. Such tests exist for GARCH type models, see e.g., [12]. However these tests cannot be applied in the present context since there are no innovations.

One way to circumvent this problem is to compute the Rosenblatt transforms used for the goodness-of-fit tests. To this end, set

$$e_{T,j,t} = \partial_{u_1} C_{j,\hat{\alpha}_{T,j}}\left(\hat{U}_{T,j,t-1}, \hat{U}_{T,j,t}\right), \quad j \in \{1, \ldots, d\},$$

where $\hat{\alpha}_{T,j}$ is a convergent estimate of α_j, and $\hat{U}_{T,j,t} = F_{T,j}\left(X_{j,t}\right)$, $t \in \{1,\ldots,T\}$, using (5). If the margins and the parameters were known exactly, then $\varepsilon_{j,t} = \partial_{u_1} C_{j,\alpha_j}\left(U_{j,t-1}, U_{j,t-1}\right)$, $U_{j,t} = F_j\left(X_{j,t}\right)$, would be i.i.d. uniform random variables. Since the margins and the parameters are not known, the pseudo-observations $e_{T,j,t}$ will be used instead and the test of independence is based on the empirical processes

$$\mathbb{K}_{T,A}(u_1,\ldots,u_d) = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \prod_{j\in A} \left\{ \mathbb{I}\left(e_{T,j,t} \le u_j\right) - u_j \right\},$$

where A belongs to the set $\mathcal{A}_d$ of all subsets of $\{1,\ldots,d\}$ containing at least two elements. The test of independence is based on the following result whose proof is given in Appendix C. Before stating the result, define

$$\tilde{\mathbb{K}}_T(u_1,\ldots,u_d) = \sqrt{T}\left\{ \frac{1}{T}\sum_{t=2}^{T}\prod_{j=1}^{d} \mathbb{I}\left(\varepsilon_{j,t} \le u_j\right) - \prod_{j=1}^{d} u_j \right\}, \quad u_1,\ldots,u_d \in [0,1].$$

Then it is well-known, e.g., [17], that under the null hypothesis of independence, as $T \to \infty$, $\tilde{\mathbb{K}}_T$ converges in law (in the Skorohod sense) to a centered Gaussian process $\mathbb{E}$, denoted $\tilde{\mathbb{K}}_T \rightsquigarrow \mathbb{E}$.

Theorem 2.2. *Suppose that the copulas $C_{j,\alpha_j}(u_j, v_j)$ are continuously differentiable with respect to v_j, α_j and that they are twice continuously differentiable with respect to u_j, $j \in \{1,\ldots,d\}$.*

Suppose also that the process $\mathbf{X}$ is alpha-mixing, stationary and that the estimators $\sqrt{T}\left(\hat{\alpha}_{T,j} - \alpha_j\right)$, together with $\tilde{\mathbb{K}}_T$, converge jointly in law, as $T \to \infty$, to random vectors $\mathbb{A}_j$, $j \in \{1,\ldots,d\}$, and $\mathbb{E}$. Then under the null hypothesis of independence between the series $X_1,\ldots,X_d$, the sequence of empirical processes $\mathbb{K}_{T,A}$, $A \in \mathcal{A}_d$, converge jointly in law as $T \to \infty$ to independent continuous Gaussian processes $\mathbb{K}_A$, with mean zero and covariance function

$$\Gamma_A(\mathbf{u},\mathbf{v}) = \prod_{j\in A} \left\{ \min(u_j, v_j) - u_j v_j \right\}, \quad \mathbf{u},\mathbf{v} \in [0,1]^d.$$

These processes appear in [17]. The statistics based on the empirical processes $\mathbb{K}_{T,A}$ are basically the same. They also have the same limiting distribution. To perform the test of independence, we propose to use the test based on the Fisher transform, as suggested in [17].

More precisely, for any $A \in \mathcal{A}_d$, set

$$T_{A,n} = \frac{1}{n}\sum_{i=1}^{T}\sum_{k=1}^{T}\prod_{j\in A}\left\{ \frac{2T+1}{6T} + \frac{R_{ij}(R_{ij}-1)}{2T(T+1)} + \frac{R_{kj}(R_{kj}-1)}{2T(T+1)} - \frac{\max(R_{ij},R_{kj})}{T+1} \right\},$$

where, for a given $j \in \{1,\ldots,d\}$, R_{ij} is the rank of $e_{T,j,i}$ amongst $e_{T,j,1},\ldots,e_{T,j,T}$. Then, if $F_{|A|,n}$ denote the distribution function of $T_{A,n}$ under the null hypothesis of independence (which only depends on the cardinality $|A|$ of A), the P–values $1 - F_{|A|,n}(T_{A,n})$ are approximately uniform on $[0,1]$. The proposed test statistic is

$$W_n = -2 \sum_{A\subset\mathcal{A}_d,\ |A|>1} \log\left\{1 - F_{|A|,n}(T_{A,n})\right\}.$$

Note that this test is implemented in the R package *copula* under the name *indepTest*.

3 Analysis of the performance

Here we compare the performance of our forecasting method versus the method proposed by [26]. The analysis is restricted to bivariate time series, but the results can easily be extrapolated to higher dimensions. For these experiments we suppose that the copula and the marginal distributions are known so that the predictions are not affected by estimation error. We will see in Section 3.5 the effect of parameters mis-specification.

Theoretically, our proposed methodology should give better results because we are using the additional information provided by a second series, provided the dependence is strong enough. Consequently, the gain in performance should be affected by the strength of the dependencies between and within time series, i.e., the overall dependence associated with the vector $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$.

To understand how these factors come into play, we consider two copulas: The Student copula and the Clayton copula. The choice of the Student copula is motivated by the fact that it seems to fit data well in practice and also that we have a lot of flexibility in specifying the correlation matrix for the Student distribution, which in return defines the strength of the dependencies in the related copula. Actually, there is a bijection between the correlation matrix and the Kendall's tau matrix. If $R = [R_{i,j}]$ for $i, j \in \{1, \ldots, d\}$ are the elements of the correlation matrix, the Kendall's tau matrix for the Student copula is then given by $\tau_{i,j} = \frac{2}{\pi} \arcsin(R_{i,j})$, [15].

In order to test a different dependence structure we also use data simulated from the Clayton copula. See Appendix A for details about simulating multivariate copula-based time series using these two copula families.

We also want to examine the impact of the marginal distributions on the performance. To make things simpler, we chose the same margins for both series. We can expect that the predictions of a random variable with large variance should be less precise than when the variance is small. To try to eliminate the margins effect, we propose a new measure of performance.

3.1 Performance measures

For most of the numerical experiments, we use prediction intervals with $\alpha = 0.05$. To measure the performance of the predictions, we compute the mean length of the prediction intervals, as well as the proportion of observed values outside the prediction intervals. Let UB_t^α and LB_t^α be the upper bound and lower bound of the prediction interval with confidence level $1 - \alpha$, for $t \in \{1, \ldots, N\}$. We define the mean length as

$$ML_N^\alpha = \frac{1}{N} \sum_{t=1}^{N} \left(UB_t^\alpha - LB_t^\alpha \right).$$

We will use $\widehat{ML}_N^\alpha$ for the mean length of prediction intervals based on the bivariate series and $\overline{ML}_N^\alpha$ if predictions are based on the univariate series. Note that the smaller the mean length of prediction intervals, the better the precision. Also, the proportion of observed values outside the prediction intervals should be close to 0.05. For the numerical experiments, we chose $N = 10000$ so that using the normal approximation for the binomial distribution one can find that a 95% confidence interval for the proportion of observed values outside the prediction intervals is approximately 0.05 ± 0.0042.

Remark 3. *Instead of prediction intervals, one can choose to make pointwise predictions using* (3) *and* (4)*. In this case, a natural performance measure for the predictions is the mean squared error. We also used this performance measure for the numerical experiments described in Sections 3.2.1 - 3.2.3 and we found out that the results were the same as for the mean length of prediction intervals. Consequently, we did not include these results in the paper since they did not bring additional information.*

In Section 3.3, we look at the effect of different marginal distributions on the quality of predictions. We expect that the mean length of prediction intervals should be larger for distribution with bigger variance. To eliminate the margins effect, we use pointwise predictions defined at (3) and (4) and we propose the following performance measure called the mean squared rank error (MSRE). Let $\tilde{X}_t$ be pointwise predictions of X_t for $t \in \{1, \ldots, N\}$. Then, MSRE is defined by

$$MSRE = N^{-1} \sum_{t=1}^{N} \left\{ F_1(X_{1,t}) - F_1(\widetilde{X}_{1,t}) \right\}^2,$$

where F_1 is the marginal distribution of $X_{1,t}$ for all $t \in \{1, \ldots, N\}$. As before, we will use $\widehat{MSRE}$ if predictions are based on the bivariate series and $\overline{MSRE}$ for the univariate case.

3.2 Impact of the dependencies strength

As mentioned previously, the structure and the strength of the dependencies of the vector

$$\mathbf{X}_t = (X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$$

should have an impact on the performance of our predictor. In order to understand what is this impact, we first simulate $\mathbf{X}_t$ with a Student copula and study the impact for a set of possible correlations. We choose $\nu = 8$ for the degrees of freedom and fix the initial values $(X_{1,0}, X_{2,0}) = (0, 0)$. In order to isolate the effect of the correlation, we take the margins as Student with $\nu = 8$ degrees of freedom as well, so the distribution of $\mathbf{X}_t$ is a multivariate Student distribution. For these experiments, we take $N = 10,000$ and we generate a sample of $n = 1000$ to compute prediction intervals with $\alpha = 0.05$.

Remark 4. *Recall that our methodology applies to stationary series. In all the following numerical experiments, when we have to generate stationary series, we always discard the first 100 values of the series, so we can consider that the series is stationary. To be precise, for a series of length N, we actually generate $N + 100$ values and discard the first 100 values.*

In all the following numerical experiments, we compare $\widehat{ML}_N^\alpha$ and $\overline{ML}_N^\alpha$ and we see that $\widehat{ML}_N^\alpha$ is always less than or equal to $\overline{ML}_N^\alpha$. This observation shows that predictions based on bivariate series outperform predictions based on univariate series since prediction intervals are narrower. However, this comparison is valid only if the proportion of observed values outside the prediction intervals is close to 0.05. As one will see these proportions are most of the time within the 95% confidence interval $[0.0458, 0.0542]$.

3.2.1 First numerical experiment

The first simulation study the impact of the dependence between both series. We simulate the series using correlation matrices

$$R_\rho = \begin{bmatrix} 1 & \rho & 0.25 & 0.25 \\ \rho & 1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 1 & \rho \\ 0.25 & 0.25 & \rho & 1 \end{bmatrix}$$

with

$$\rho \in \{-0.4, -0.3, -0.2, -0.1, -0.05, -0.01, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}.$$

Here, $\tau = \frac{2}{\pi}\arcsin(\rho)$ is the Kendall's tau between $X_{1,t}$ and $X_{2,t}$.

As seen in Figure 1, the mean length $\widehat{ML}_N^\alpha$ of the prediction intervals increases when the correlation ρ (and Kendall's tau) increases. To explain this result, consider the extreme case where the Kendall's tau between $X_{1,t}$ and $X_{2,t}$ is one. Then, the two series are identical and hence our predictor has no additional information coming from the second series. This also explain why the difference between mean length of prediction intervals gets closer to zero when $\tau = \frac{2}{\pi}\arcsin(\rho)$ is high. Again, as ρ increases, the predictor $\widehat{ML}_N^\alpha$ tends to $\overline{ML}_N^\alpha$, since the information from the second series becomes irrelevant.

3.2.2 Second numerical experiment

For the second simulation, we study the impact of the serial dependence, i.e., the dependence between $X_{1,t}$ and $X_{1,t-1}$ through $\tau = \frac{2}{\pi}\arcsin(\rho)$, Kendall's tau between $X_{1,t}$ and $X_{1,t-1}$. We simulate the series using cor-

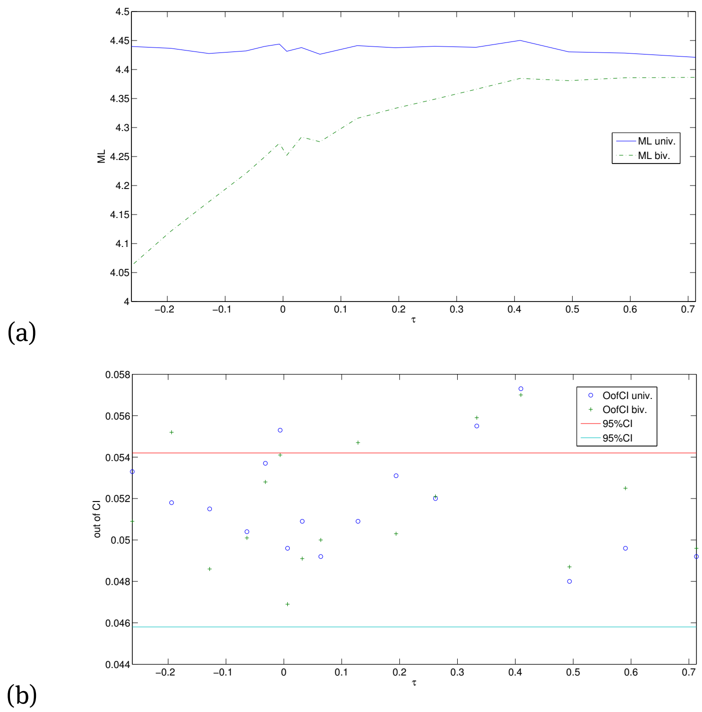


Figure 1: Impact of the dependence between $X_{1,t}$ and $X_{2,t}$, as measured by the average length of the prediction intervals in panel (a). The plain line gives the value of $\overline{ML}_N^\alpha$ and the dashed gives the value of $\widehat{ML}_N^\alpha$. Panel (b) gives the proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series, and the horizontal lines give the 95% confidence interval. For both panels, the x-axis gives the Kendall's tau between $X_{1,t}$ and $X_{2,t}$.

relation matrices

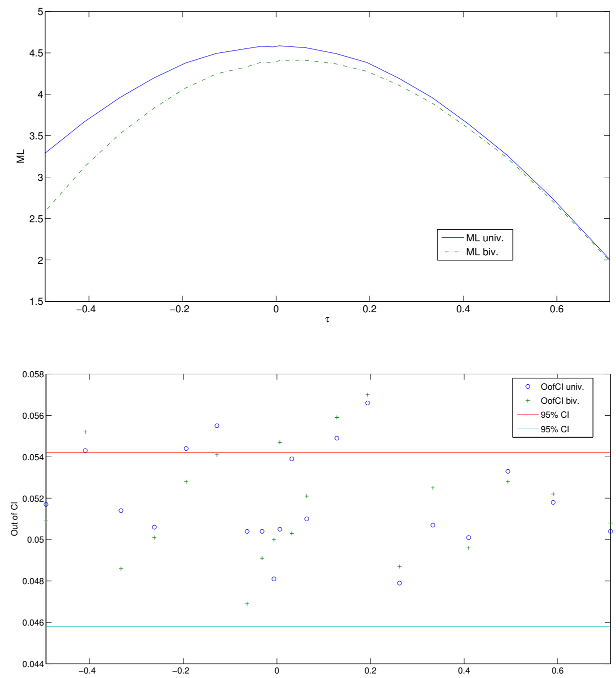
$$R_\rho = \begin{bmatrix} 1 & 0.25 & \rho & 0.25 \\ 0.25 & 1 & 0.25 & 0.25 \\ \rho & 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{bmatrix}$$

with

$$\begin{aligned} \rho \in \; & \{-0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, -0.05, -0.01, \\ & 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}. \end{aligned}$$

Remark 5. *We remark that the range of ρ is not the same than before. The reason is that we have to keep the correlation matrices positive semi-definite.*

As expected theoretically, Figure 2 shows that the predictions are better when the dependence is strong. However, it is interesting to note that, for this dependence structure, $\widehat{ML}_N^\alpha$ benefits more from the negative dependence than $\overline{ML}_N^\alpha$. Finally, when the correlation is close to one, the information given by the first lag dictates almost completely the succeeding value and the information given by the second series then becomes marginal. This is why the difference in prediction error is close to zero when the correlation is close to one.



(a)

(b)

Figure 2: Impact of the dependence between $X_{1,t}$ and $X_{1,t-1}$, as measured by the average length of the prediction intervals in panel (a). The plain line gives the value of $\overline{ML}_N^{\alpha}$ and the dashed gives the value of $\widehat{ML}_N^{\alpha}$. Panel (b) gives the proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series, and the horizontal lines give the 95% confidence interval. For both panels, the x-axis gives the Kendall'tau between $X_{1,t}$ and $X_{1,t-1}$.

3.2.3 Third numerical experiment

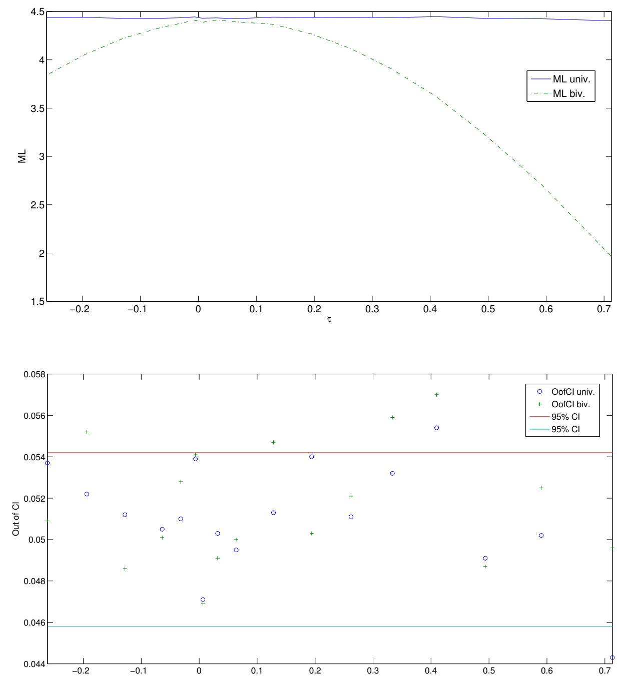
The last simulation is about the impact of the strength of the dependence between $X_{1,t}$ and $X_{2,t-1}$. We simulate the series using correlation matrices

$$R_\rho = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1 & \rho & 0.25 \\ 0.25 & \rho & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{bmatrix}$$

with

$$\begin{aligned} \rho \quad \in \quad & \{-0.4, -0.3, -0.2, -0.1, -0.05, -0.01, 0.01, 0.05, 0.1, 0.2 \\ & 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}. \end{aligned}$$

As one should expect, this is the most important dependence in the comparative performance of our predictor. Since the information given by $X_{2,t-1}$ cannot be used by predictions based only on univariate series, our predictor gives much better performance when this dependence is strong, as seen in Figure 3. In the case where the dependence is strong, the information of $X_{2,t-1}$ almost completely dictates the value of $X_{1,t}$ and this is why we observe a great difference between $\widehat{ML}_N^{\alpha}$ and $\overline{ML}_N^{\alpha}$.



(a)

(b)

Figure 3: Impact of the dependence between $X_{1,t}$ and $X_{2,t-1}$, as measured by the average length of the prediction intervals in panel (a). The plain line gives the value of $\overline{ML}_N^\alpha$ and the dashed gives the value of $\widehat{ML}_N^\alpha$. Panel (b) gives the proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series, and the horizontal lines give the 95% confidence interval. For both panels, the x-axis gives the Kendall'tau between $X_{1,t}$ and $X_{2,t-1}$.

3.3 Impact of the marginal distributions

Another question we want to tackle is the impact of the margins. In order to isolate more closely the impact of the different correlations in the first set of simulations, we used only the Student copula with Student margins. In the next experiment, we still use Student copula, but we are going to use different marginal distributions. The parameters of the Student copula are the same as before and we fix the correlation matrix at $R_{i,j} = 0.25$ for all $i \neq j$. The results are displayed in Table 1.

Remark 6. *As pointed out by a referee, it is expected that the heavier tailed Student distribution as well as normal distributions with higher variance should have an impact on the mean length of prediction intervals. The point of using the MSRE measure here is to have all the values in* $[0, 1]$ *so that we can compare different types of distributions. Looking at the results of Table 1, we see that the* ML_N^α *measure behaves similarly to the MSRE measure, but on a larger scale. Thus, we may conclude that marginal distribution do indeed affect the precision of the predictions.*

3.4 Impact of the dependence structure

Our last numerical experiment shows that the dependence structure has an impact on the gain in performance of predictions based on bivariate series compared to predictions based on univariate series. At first sight, we could think that using the information provided by the series X_2 might always give better predictions but we find that the dependence structure of the Clayton copula almost negate this advantage. From the definition of the Clayton copula we see that the dependence structure is symmetric, that is, all the dependencies of

Table 1: Evolution of $MSRE$ and ML_N^α as a function of the marginal distributions. Numbers in parenthesis are the proportion of observed values out of prediction intervals.

Margins	$\overline{MSRE}$	$\widehat{MSRE}$	$\overline{ML}_N^\alpha$	$\widehat{ML}_N^\alpha$
T_5	0.0819	0.0671	4.9503 (0.0545)	4.3449 (0.0512)
T_8	0.0815	0.0664	4.4407 (0.0526)	3.9101 (0.0513)
T_{10}	0.0812	0.0662	4.2891 (0.0534)	3.7835 (0.0514)
T_{15}	0.0810	0.0658	4.1025 (0.0536)	3.6287 (0.0499)
T_{20}	0.0810	0.0658	4.0171 (0.0545)	3.5547 (0.0521)
T_{30}	0.0808	0.0656	3.9333 (0.0535)	3.4858 (0.0514)
T_{50}	0.0808	0.0655	3.8736 (0.0539)	3.4345 (0.0509)
$LN(0,1)$	0.1114	0.0819	6.6889 (0.0522)	5.7896 (0.0517)
χ_8^2	0.0841	0.0672	14.7845 (0.0522)	13.1147 (0.0512)
Exp_3	0.0956	0.0737	10.5561 (0.0520)	9.2709 (0.0513)
$N(0,1)$	0.0806	0.0654	3.7803 (0.0526)	3.3573 (0.0516)
$N(0.4)$	0.0807	0.0654	7.5583 (0.0520)	6.7141 (0.0520)
$N(0,8)$	0.0807	0.0654	10.6900 (0.0520)	9.4942 (0.0513)
$N(2,1)$	0.0806	0.0654	3.7802 (0.0531)	3.3580 (0.0532)
$N(4,1)$	0.0807	0.0654	3.7796 (0.0536)	3.3566 (0.0509)
$N(8,1)$	0.0806	0.0654	3.7796 (0.0538)	3.3570 (0.0515)

the vector $\mathbf{X}_t = (X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ are the same. Moreover, the strength of the dependencies increases when θ increases. When θ is close to zero, the elements of the vector $\mathbf{X}_t$ are close to be independent and so, there is not much information to use to predict the next value. On the opposite, when θ is high, both series are almost the same and, this time, the series X_2 cannot provide useful information to our predictor.

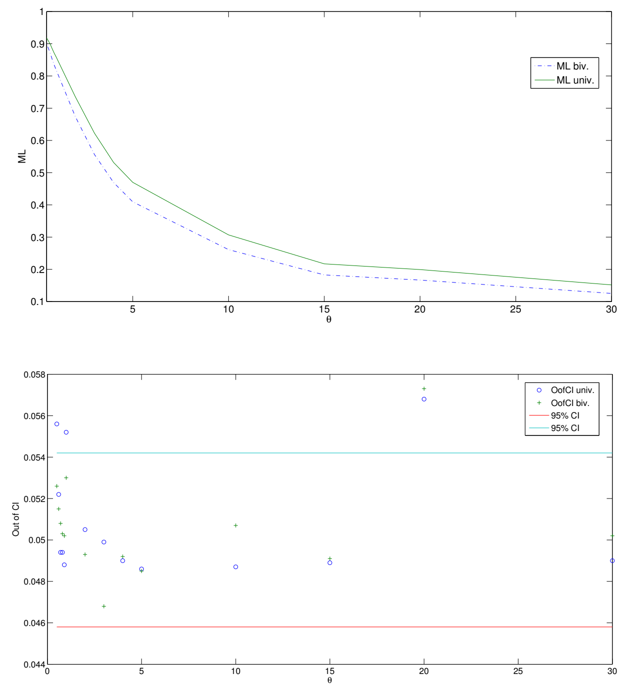
The results illustrated in Figure 4 show the evolution of prediction performances in terms of the parameter θ. We see that both predictors perform badly when θ is small and perform better as long as θ becomes bigger. We also see that the difference between both prediction errors is slowly decreasing for high values of θ. This is due to the fact that the Kendall's tau between $X_{1,t}$ and it's first lag $X_{1,t-1}$ is close to one, and so, the additional information provided by X_2 becomes marginal. As we see, with the dependence structure given by the Clayton copula the advantage of using predictions based on the bivariate series is minor.

3.5 Impact of estimation errors

In all the previous numerical experiments, we supposed that copulas and marginal distributions were known, i.e. there was no need to estimate parameters. This way, we were able to isolate the effect of using the information provided by the second series. We saw that using multivariate predictions outperform univariate predictions, but in some cases the improvement is rather small. In these cases, one can ask if the errors caused by parameters estimation might negate the advantage of the multivariate forecasting method. Mostly when the multivariate method requires more parameters to estimate. In this section, we carry two numerical experiments to test the impact of estimation errors on the performance of the predictions.

For our first experiment we generate a bivariate series $\mathbf{X}_t$ of length $N+1$ from a Student copula with $\nu = 8$ degrees of freedom and correlation matrix

$$R = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1 & 0.7 & 0.25 \\ 0.25 & 0.7 & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{bmatrix}.$$



(a)

(b)

Figure 4: Impact of the parameter θ in the Clayton copula. In panel (a), the plain line represents the value of $\overline{ML}_N^\alpha$ and the dashed gives the value of $\widehat{ML}_N^\alpha$. Panel (b) gives the proportion of observed values outside of prediction intervals. The circles give the results for prediction intervals based on univariate series while plus signs are for prediction intervals based on bivariate series, and the horizontal lines give the 95% confidence interval. The x-axis gives the values of θ.

We estimate the parameters of the copula using the N first values of the series and predict the value $N + 1$. We repeat this experiment 10,000 times and compute the performance of the 10,000 predictions. To generate the series, we also take Student marginal distributions with the same degrees of freedom as the copula.

In order to have different precisions for the estimated parameters we perform this experiment with sample sizes $N \in \{100, 250, 500, 750\}$. The results are displayed in Table 2.

Table 2: Impact of estimation errors for the Student copula. From left to right we have: the length of the series, the mean squared error for the univariate and bivariate methods, and the mean length of prediction intervals for the univariate and bivariate methods. The numbers in parenthesis are the percentage of observed values out of the prediction intervals. They are written in bold when significantly different from 5%, at the 95% level.

	Student copula			
Length	$\overline{MSE}$	$\widehat{MSE}$	$\overline{ML}_N^\alpha$	$\widehat{ML}_N^\alpha$
100	1.2784	0.8240	4.5897 **(5.89)**	3.3043 **(6.43)**
250	1.2426	0.8233	4.4658 (5.32)	3.242 **(5.92)**
500	1.2855	0.8005	4.4505 (5.39)	3.2366 **(5.54)**
750	1.3362	0.7932	4.413 **(5.73)**	3.3241 (5.34)

The first observation from Table 2 is that pointwise predictions are more precise for the bivariate method. Our second observation is that, even though the bivariate method still creates smaller prediction intervals, the proportion of observed values outside the prediction intervals are inside the confidence interval only for the series of length 750. This means that the predictions of the quantiles using the bivariate method are more sensitive to estimation errors than the univariate method. However, increasing the sample size reduces the covering error.

Our second experiment retains the same idea as before, except we generate a bivariate series following a Clayton copula with parameter $\theta = 5$ and uniform marginal distributions. Looking at the results in Table 3, we see that the bivariate method outperforms the univariate method since the mean average error for pointwise predictions as well as the mean length of prediction intervals are smaller and the number of observed values outside prediction intervals are all within the confidence interval.

Table 3: Impact of estimation errors for the Clayton copula. From left to right we have: the length of the series, the mean squared error for the univariate and bivariate methods, and the mean length of prediction intervals for the univariate and bivariate methods. The numbers in parenthesis are the percentage of observed values out of the prediction intervals. These values are not significantly different from 5%, at the 95% level.

Clayton copula				
Length	$\overline{MSE}$	$\widehat{MSE}$	$\overline{ML}_N^{\alpha}$	$\widehat{ML}_N^{\alpha}$
100	0.0174	0.0131	0.4730 (4.90)	0.4129 (4.74)
250	0.0168	0.0129	0.4712 (5.04)	0.411 (4.80)
500	0.0167	0.0129	0.4664 (4.90)	0.4076 (5.00)
750	0.0168	0.0128	0.4683 (5.09)	0.4086 (5.29)

From these results, we conclude that predictions based on the bivariate method still gives better predictions when considering estimation errors. However, in the case of the Student copula, it seems that prediction intervals using the bivariate method are more sensitive to estimation errors, which is probably caused by the fact that it uses more parameters. This explanation is also supported by the results on the Clayton copula. In this case, all the predictions rely on one estimated parameter and the bivariate method always gives the best predictions. It would be interesting to know if this conclusion can be generalized, but it would require an exhaustive study to fully understand the effect of estimation errors.

4 Application

In this section we present an application of our method for forecasting the estimated realized volatility using the bivariate time series of the estimated realized volatility and the volume of transactions. We first compare the performance of the predictions between the univariate and the bivariate version of our method and then use the multivariate GARCH-BEKK model as a benchmark to compare predictions based on a bivariate model. But first, we discuss the estimation of the realized volatility.

Realized volatility might be defined as an empirical measure of returns volatility. In a general setting, if we suppose that the value of an asset is a semimartingale X, then the realized volatility of X over the period $[0, T]$ is its quadratic variation at time T, $[X]_T$. Thus, an estimator of the realized volatility can be defined as the sum of squared returns

$$R\hat{V}(X)_{[0,T]} = \sum_{i=1}^{N} (X_{t_i} - X_{t_{i-1}})^2 , \tag{7}$$

where X_{t_i}, $i = 0, \ldots, n$, are observed values and $0 = t_0 \leq t_1 \leq \cdots \leq t_n = T$. The first mention of realized volatility is probably [29] but we refer the reader to [4] for a detailed justification of the realized volatility estimation.

Since each price observation is noisy (bid-ask spread, etc), a more realistic model for observed price should be $Y_{t_i} = X_{t_i} + \epsilon_{t_i}$, where ϵ_{t_i} is a random variable. In this context, it is easy to show that (7) is an inconsistent estimator. A common practice to estimate realized volatility is to use (7) and to take observations every 5 to 30 minutes. In using less observations the bias due to noise is somewhat diminished and the estimation precision becomes acceptable. However, to perform our realized volatility estimation we used the estimator of [28] which is an asymptotically unbiased estimator that allows using high-frequency data. Another good estimator is given by [20] which makes use of high and low observed values. The reason we prefer the former estimator is that we used trade prices and it seems that this estimator is less affected by bid/ask spread.

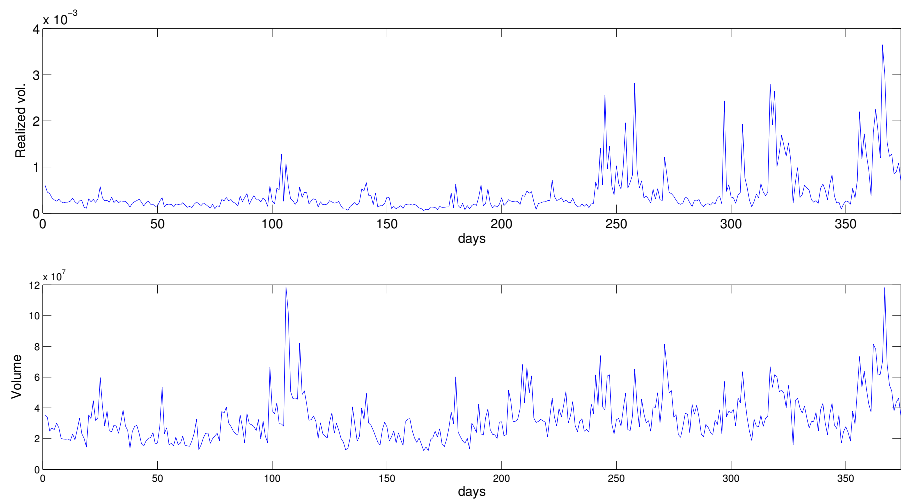


Figure 5: Estimated realized volatility (top panel) and volume of transactions (bottom panel).

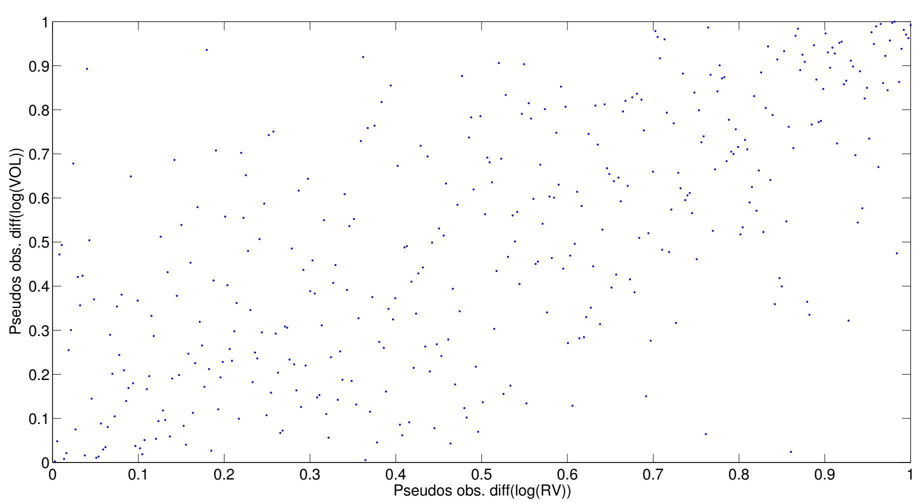


Figure 6: Scatter plot for the normalized rank of the first difference of the log of the realized volatility and the volume of transaction.

The data we are using are from the Trade and Quote database. We used Apple (APPL) trade prices from 2006/08/08 to 2008/02/01, which consists of 374 trading days. In order to avoid periods of lower frequency trading we used data from 9:00:00 to 15:59:59. In combination with the estimation of realized volatility we also computed the aggregated volume of transactions; see Figure 5.

4.1 Predictions based on copula models

For our time series to satisfy the required hypothesis of stationarity, we have to take the first difference of the logarithm of both series. We define $X_{1,t} = \log(\hat{\mathrm{rv}}_t) - \log(\hat{\mathrm{rv}}_{t-1})$ and $X_{2,t} = \log(\hat{\mathrm{vol}}_t) - \log(\hat{\mathrm{vol}}_{t-1})$ where $\hat{\mathrm{rv}}_t$ is the estimated volatility, and $\hat{\mathrm{vol}}_t$ is the aggregated volume of transaction and the time scale is in days.

To verify the stationarity assumption of both series, we used a non-parametric change point test using the Kolmogorov-Smirnov statistic. With P-values of 21.7% and 34.1% for the series X_1 and X_2, one cannot reject the null hypothesis of stationarity. Finally, to try to have a visual interpretation of the dependence between the two series, Figure 6 displays a scatter plot of the normalized ranks. This is a proxy for a sample generated under the underlying copula of $(X_{1,t}, X_{2,t})$.

Then we carried out parameters estimation and goodness-of-fit tests for Clayton, Gaussian and Student copulas, as proposed in [23]. One should note that these goodness-of-fit tests also test the hypothesis that the series are Markovian.

From the P-values listed in Table 4, we selected the Student copula as the best model for the copula of $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$. The estimated parameters for the Student copula distribution are the degree of freedom, $\hat{\nu} = 12.4707$, and the correlation matrix

$$\hat{R} = \begin{bmatrix} 1 & 0.6936 & -0.3628 & -0.1234 \\ 0.6936 & 1 & -0.2960 & -0.3035 \\ -0.3628 & -0.2960 & 1 & 0.6936 \\ -0.1234 & -0.3035 & 0.6936 & 1 \end{bmatrix}.$$

The associated Kendall's tau matrix is then given by

$$\tau = \begin{bmatrix} 1 & 0.4880 & -0.2364 & -0.0788 \\ 0.4880 & 1 & -0.1913 & -0.1963 \\ -0.2364 & -0.1913 & 1 & 0.4880 \\ -0.0788 & -0.1963 & 0.4880 & 1 \end{bmatrix}.$$

It is not true in general that the copula for $(X_{1,t-1}, X_{1,t})$ will also be Student; this will be true if the series $X_{1,t}$ is Markovian. This is indeed the case here and it happens that the best model for $(X_{1,t-1}, X_{1,t})$ is actually the Student copula; see Table 5 for the results of the different goodness-of-fit tests. The estimated degree of freedom is $\hat{\nu} = 4.6751$ and the correlation matrix is defined by taking the corresponding entries from the correlation matrix of the bivariate model. Note also that the test of independence between the two series, as described in Section 2.4.3, has been performed and the null hypothesis of independence is rejected with a P-value equal to zero.

Table 4: P-values of goodness-of-fit tests of $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ computed with 1000 bootstrap samples.

Copula	Clayton	Gaussian	Student
P-value (%)	0	3.2	5.5

Then we used our methodology to make one-period ahead predictions for the next 100 out-of-sample values of the series X_1. The results of the prediction intervals are displayed in Figure 7 while the pointwise

Table 5: P-values of goodness-of-fit tests for $(X_{1,t-1}, X_{1,t})$ and $(X_{2,t-1}, X_{2,t})$, computed with 1000 bootstrap samples.

Model	$(X_{1,t-1}, X_{1,t})$			$(X_{2,t-1}, X_{2,t})$		
Copula	Clayton	Gaussian	Student	Clayton	Gaussian	Student
P-value (%)	1.7	59.9	65.9	58.5	10.9	10.9

predictions are shown in Figure 8. We mention that the pointwise predictions from the univariate and the bivariate model are close and it is difficult to distinguish them. When we compare the performance of the predictions, we find that the bivariate method has a little advantage over the univariate method, see Table 6.

Table 6: Mean squared error and mean length of prediction intervals for predictions based on the univariate and the bivariate method. The means are taken on $N = 100$ predictions and $\alpha = 5\%$. The number in parenthesis is the proportion of observed values outside prediction intervals.

	Univariate	Bivariate
ML_N^α	2.2493 (0.09)	2.2045 (0.09)
MSE	0.4033	0.4023

However, one can asks if this difference is significant. To answer this question, we apply a test for predictive accuracy designed in [11]. Let $X_{1,t}$, $\hat{X}_{1,t}$ and $\bar{X}_{1,t}$ be respectively the series of the observed values, the predictions based on the bivariate model and the predictions based on the univariate model. We define the series of the loss differential of the squared errors, $d_t = (X_{1,t} - \hat{X}_{1,t})^2 - (X_{t,1} - \bar{X}_{1,t})^2$, and test the null hypothesis H_0: $d_t = 0$. Under the hypothesis that the series of the loss differential $\{d_t\}_{t=1}^N$ is covariance stationary and short-memory, [11] state that $\sqrt{N}(\bar{d} - \mu) \to N(0, 2\pi f_d(0))$ where $f_d(0)$ is the spectral density of the loss differential at frequency 0. Defining the statistic

$$S_1 = \frac{\bar{d}}{\sqrt{\frac{1}{N} 2\pi \hat{f}_d(0)}},$$

where $\hat{f}_d(0)$ is a consistent estimator of $f_d(0)$ we find that $S_1 = -0.1966$ so that we don't reject H_0 at level of confidence 95%.

By looking at the estimated correlation matrix $\hat{R}$, we see that the temporal correlation between $X_{1,t}$ and $X_{1,t-1}$ is rather strong. Following the numerical experiments of Section 3, we should only expect a minor advantage of the bivariate method since the correlation between $X_{1,t}$ and the first lag of the second series $X_{2,t-1}$ is not dominant. This observation is a possible explanation why the performance of the bivariate method in this application is better but not significantly better.

On the other hand, in this particular case, the AIC criterion is applicable, since $(X_{1,t-1}, X_{1,t})$ is modeled by a Student copula which is embedded in the Student copula used to model $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$. In the univariate case, from the goodness-of-fit test we get a log-likelihood of 33.34 and there are two parameters to estimate, while for the bivariate case the log-likelihood is 154.84 with 6 parameters to estimate. Since the number of parameters is small with respect to the sample size we use the AIC criterion without the correction term, as defined in [2], that is, $AIC = 2k - 2LL$ where k is the number of parameters and LL is the log-likelihood. Computing the AIC criterion gives -62.69 for the univariate model and -297.68 for the bivariate model which suggests that the bivariate model should be used since its AIC is smaller.

4.2 Comparison with the multivariate GARCH-Bekk model.

After comparing the performance of our method with its univariate version, we would like to use another bivariate model as a benchmark. So, we now model the series $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$ with a multivariate GARCH-BEKK, see [13]. In this model, we suppose that

$$\mathbf{X}_t = \mu + \mathbf{H}_t^{\frac{1}{2}} \epsilon_t$$

where μ is a 2×1 vector of constants and ϵ_t is a 2-dimensional vector of independent standard normals. The covariance matrix $\mathbf{H}_t$ is defined by

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\epsilon_{t-1}\epsilon_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}'$$

where $\mathbf{A}$, $\mathbf{B}$ are 2×2 matrices and $\mathbf{C}$ is a 2×2 lower triangular matrix. To estimate the parameters, we used the USCD_GARCH toolbox for Matlab created by Kevin Sheppard and the result of the esimation is

$$A = \begin{bmatrix} 0.6055 & -0.3053 \\ 0.1126 & 0.1345 \end{bmatrix}, \quad B = \begin{bmatrix} -0.0001 & 0 \\ 0.0001 & 0 \end{bmatrix}$$

and

$$C = \begin{bmatrix} 0.4873 & 0 \\ 0.2277 & 0.2439 \end{bmatrix}.$$

Finally, we have that $\mu = (4.409 \times 10^{-4}, 5.407 \times 10^{-7})$.

For this particular model, pointwise predictions based on the conditional expectation are less relevant since it is a constant. The result of 95% prediction intervals are displayed in Figure 9. To compare the performance with the bivariate copula based model, we look at the mean length of prediction intervals, ML_N^α. For the multivariate GARCH-BEKK model, the mean length of prediction intervals is 3.6489 with 4 observed values outside prediction intervals while for the bivariate copula model we have a mean length of prediction intervals of 2.2045 with 9 observed values outside prediction intervals. In both cases the number of observed values outside prediction intervals is inside the 95% confidence interval. One can see that the predictions based on the bivariate copula model give better predictions than those from the multivariate GARCH-BEKK model.

5 Conclusions

In this paper, we presented a forecasting method for time series based on multivariate copulas and we compared the performance of the predictions with the univariate version. We also discussed the implementation of the proposed methodology, introducing a new test of independence between time series. Using the Student copula, we illustrated the impact of different combinations of dependencies for the vector $(X_{1,t-1}, X_{2,t-1}, X_{1,t}, X_{2,t})$ and we saw that some combinations are more favorable for the multivariate method than others. We also saw that with the symmetrical dependence structure of the Clayton copula, the multivariate forecasting method shows only a minor advantage. In addition, we tested the effect of estimation errors. It was observed that the multivariate method keeps its advantage, but for series following a Student copula, the sample size should be taken sufficiently large in order to provide good estimated parameters if one want to use prediction intervals. These results might be explained by the number of estimated parameters used for the predictions. Finally, to illustrate the methodology, we presented a complete application with parameters estimation and goodness-of-fit tests on the bivariate series of realized volatility and volume of transactions and concluded that predictions based on the bivariate copula model outperform predictions based on the univariate copula model as well as the multivariate GARCH-BEKK model.

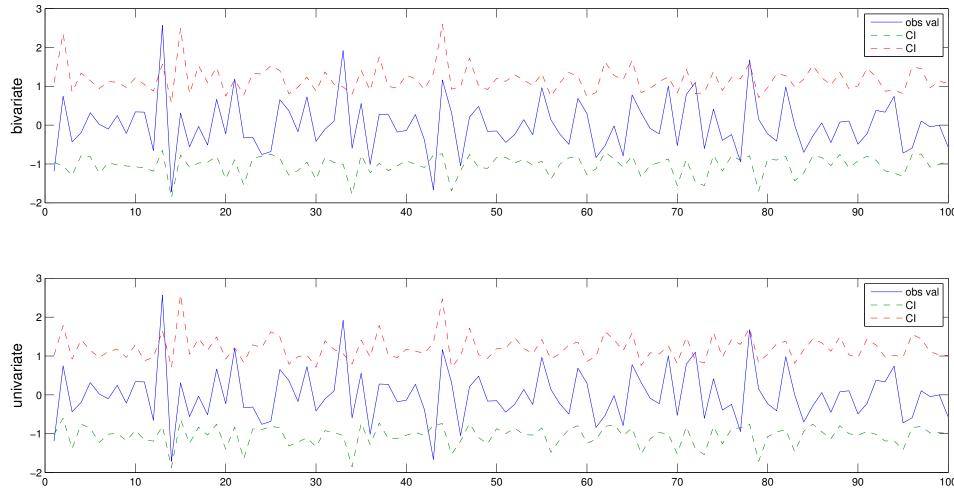


Figure 7: 95% prediction intervals for X_1, using the bivariate model (top panel) and the univariate model (bottom panel).

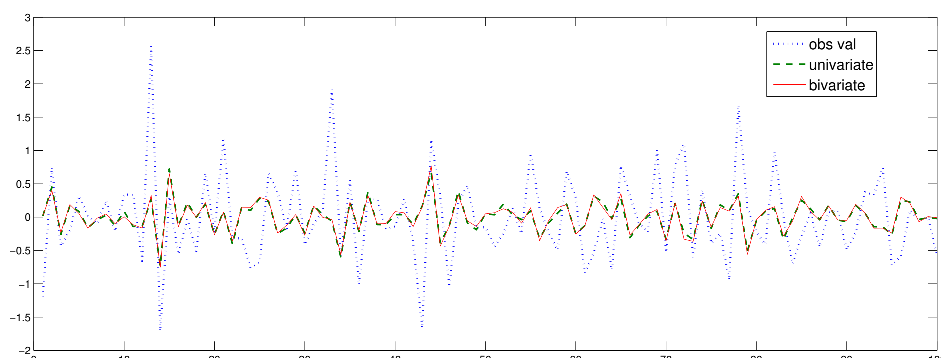


Figure 8: Pointwise predictions using the conditional expectation of the univariate (dashed line) and the bivariate model (plain line). Observed values are shown by the dotted line.

Acknowledgement: This research is supported by the Fonds de Recherche du Québéc Nature et Technologies, by the Natural Sciences and Engineering Research Council of Canada, and by the Professorship in Financial Engineering (HEC Montréal). We thank the Editor and the referees for their helpful suggestions that lead to an improvement of the paper.

A Simulation

A.1 Simulation of multivariate time series with Student copula

The Student copula is based on a multivariate Student distribution. Suppose $(\mathbf{X}, \mathbf{Y})$ is a $d = (d_1 + d_2)$-dimensional random vector which follows a Student distribution with mean 0, correlation matrix R and ν

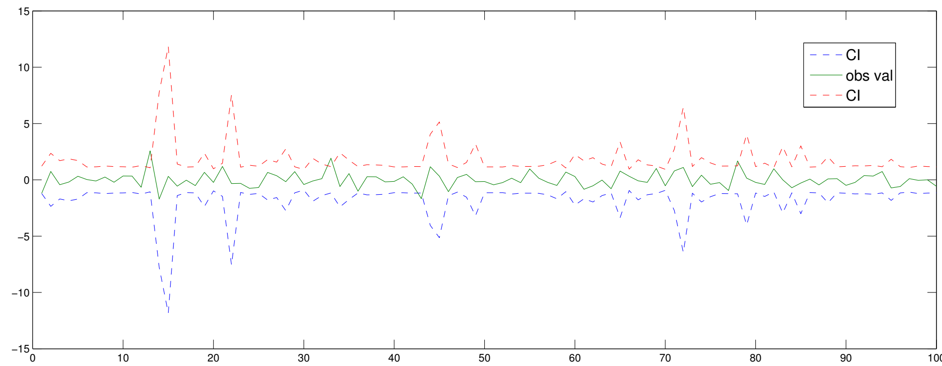


Figure 9: 95% prediction intervals for X_1 using a multivariate Bekk-GARCH.

degrees of freedom. We write the matrix R as a block matrix

$$R = \begin{bmatrix} R_{\mathbf{X}} & R_{\mathbf{XY}} \\ R_{\mathbf{YX}} & R_{\mathbf{Y}} \end{bmatrix}$$

where $R_{\mathbf{X}}$, $R_{\mathbf{Y}}$, $R_{\mathbf{YX}}$ and $R_{\mathbf{XY}}$ are respectively the correlation matrices of the variables in subscript. It is easy to check that all possible joint distributions of a multivariate Student vector are also of Student distributions with respective correlation matrix and the same degrees of freedom. Let $T_{\nu,R}$ be the distribution function of a multivariate Student vector. The Student copula, noted $C_{\nu,R}$ is defined, for all $(\mathbf{u},\mathbf{v}) \in (0,1)^{d_1+d_2}$, by

$$C_{\nu,R}(\mathbf{u},\mathbf{v}) = T_{\nu,R}\left\{T_\nu^{-1}(u_1),\ldots,T_\nu^{-1}(u_{d_1}),T_\nu^{-1}(v_1),\ldots,T_\nu^{-1}(v_{d_2})\right\}.$$

Using Schur's complement on the correlation matrix R, it is possible to show that the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ is also a Student distribution with $\tilde{\nu} = \nu + d_1$ degrees of freedom, mean $\boldsymbol{\mu} = B\mathbf{X}$, and scale matrix $\tilde{R} = \frac{\lambda}{\tilde{\nu}}\Omega$, where $\lambda = \nu + \mathbf{x}^\top \Sigma_{\mathbf{X}}^{-1}\mathbf{x}$, $\Omega = R_{\mathbf{Y}} - R_{\mathbf{YX}}R_{\mathbf{X}}^{-1}R_{\mathbf{XY}}$, and $B = R_{\mathbf{YX}}R_{\mathbf{X}}^{-1}$. The details of the derivations are given in Appendix B.

To generate a d-dimensional time series $\{\mathbf{X}_t\}_{t\in\mathbb{N}}$ such that $(\mathbf{X}_{t-1},\mathbf{X}_t)$ has a Student conditional copula $C_{\nu,R}$ with marginal distributions $F_1,\ldots,F_d$, and

$$R = \begin{bmatrix} R_1 & R_{12} \\ R_{21} & R_1 \end{bmatrix},$$

we use the following algorithm:

1. Generate $\mathbf{Y}_0$ from a d-dimensional Student distribution with ν degrees of freedom and correlation matrix R_1.
2. For all $t \in \mathbb{N}$, generate $\mathbf{Y}_t$ from a d-dimensional Student distribution with $\tilde{\nu}$ degrees of freedom, scale matrix $\tilde{R}$ and mean $B\mathbf{Y}_{t-1}$, where $B = R_{21}R_1^{-1}$.
3. Compute $U_{jt} = T_\nu(Y_{jt})$, for all $j \in \{1,\ldots,d\}$.
4. Set $(X_{1t},\ldots,X_{dt}) = \left(F_1^{-1}(U_{1t}),\ldots,F_d^{-1}(U_{dt})\right)$.

Recall that to generate a d-dimensional random vector Y from the Student distribution $T_{\nu,\mu,R}$, one can generate V from the χ^2_ν distribution and set $Y = Z\sqrt{\nu/V} + \mu$ where Z is a d-dimensional normal vector independent of V with mean 0 and correlation matrix R.

A.2 Simulation of multivariate time series with Clayton copula

The Clayton copula is a member of the Archimedean family. A copula C_ϕ is said to be Archimedean with generator ϕ if

$$C_\phi(u) = \phi^{-1}\left\{\phi(u_1) + \ldots + \phi(u_d)\right\}$$

for any bijection $\phi : [0,1) \to [0,\infty)$ such that $(-1)^i \frac{d^i}{d^i s}\phi^{-1}(s) \geq 0$ for all $s \geq 0$ and all $i = 0,\ldots,d-1$. Archimedean copulas are uniquely defined by the generator, up to a positive scaling factor. The Clayton copula is part of the Archimedean family and is defined by the generator $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$ with $\theta > 0$. Note that more generally it is possible to define a generator for the Clayton copula with parameter $\theta \geq -\frac{1}{d-1}$ but we restrict ourself to the case with positive parameter. Suppose that $(\mathbf{U}, \mathbf{V})$ is a $(d_1 + d_2)$–dimensional random vector which follows a Clayton copula $C_{d,\theta}$, where $d = d_1 + d_2$. Then it is possible to show that the conditional copula of $\mathbf{V}$ given $\mathbf{U}$ is a Clayton copula with parameter $\tilde{\theta} = \frac{\theta}{1+d_1\theta}$.

To generate a $2d$-dimensional time series $\{\mathbf{X}_t\}_{t=0,1,\ldots}$ such that $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ follows a Clayton copula $C_{2d,\theta}$ with marginal distributions $F_1,\ldots,F_d$ we use the following algorithm:

1. Generate $\mathbf{U}_0$ from the distribution $C_{d,\theta}$.
2. For all $t \in \mathbb{N}$ and $j \in \{1,\ldots,d\}$ compute

$$U_{jt} = \left[\left(\sum_{j=1}^{d} U_{j,t-1}^{-\theta} - d + 1\right)\left(V_{j,t}^{-\tilde{\theta}} - 1\right) + 1\right]^{-1/\theta}$$

   where $\mathbf{V}_t \sim C_{d,\tilde{\theta}}$ with $\tilde{\theta} = \frac{\theta}{1+d\theta}$.
3. Set $X_{jt} = F_j^{-1}(U_{jt})$ for all $j \in \{1,\ldots,d\}$ and all $t \in \mathbb{N}$.

Recall that to generate a d-dimensional random vector Y from a Clayton copula $C_{d,\theta}$, we can simulate independently S from a Gamma$(1/\theta, 1)$ and $E_1,\ldots,E_d$ from a Exp(1), and then we set $Y_j = (1 + E_j/S)^{-\theta}$ for $j \in \{1,\ldots,d\}$.

B Conditional Student distribution

Let $\mathbf{Z}^\top = (\mathbf{X}^\top, \mathbf{Y}^\top)$ be a $d = (d_1 + d_2)$-dimensional random vector which follows a multivariate Student distribution $T_d(x; \nu, \mu, \Sigma)$, where ν is the degrees of freedom, $\boldsymbol{\mu}^\top = (\boldsymbol{\mu}_\mathbf{X}^\top, \boldsymbol{\mu}_\mathbf{Y}^\top)$ is a $(d_1 + d_2)$-dimensional real vector which is the location vector and

$$\Sigma = \begin{bmatrix} \Sigma_\mathbf{X} & \Sigma_\mathbf{XY} \\ \Sigma_\mathbf{YX} & \Sigma_\mathbf{Y} \end{bmatrix}$$

is the scale block matrix. The density function of the above multivariate Student distribution is defined as

$$t_d(\mathbf{x}, \mathbf{y}; \nu, \mu, \Sigma) = \frac{\Gamma(\frac{\nu}{2} + \frac{d}{2})}{|\Sigma|^{1/2}\Gamma(\frac{\nu}{2})(\pi\nu)^{-d/2}}\left(1 + \frac{(\mathbf{X} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})}{\nu}\right)^{-(\frac{\nu}{2}+\frac{d}{2})},$$

where $\Gamma(x)$ is the gamma function. Moreover, it is well know that all joint distributions of a multivariate Student distribution are also Student. For our concern, X follows a d_1-dimensional multivariate distribution with parameters ν, μ_1 and Σ_X.

Let $\mathbb{I}_d$ and 0_d be respectively the d–dimensional identity matrix and null matrix. Using Schur's method we can write $\Sigma = A \times M \times B$ where

$$A = \begin{bmatrix} \mathbb{I}_{d_1} & 0_{d_1 \times d_2} \\ \Sigma_{YX}\Sigma_X^{-1} & \mathbb{I}_{d_2} \end{bmatrix}$$

$$M = \begin{bmatrix} \Sigma_X & 0_{d_1 \times d_2} \\ 0_{d_2 \times d_1} & \Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY} \end{bmatrix}$$

$$B = \begin{bmatrix} \mathbb{I}_{d_1} & \Sigma_X^{-1}\Sigma_{XY} \\ 0_{d_2 \times d_1} & \mathbb{I}_{d_2} \end{bmatrix}.$$

Then we see that we can write the inverse of Σ the following way,

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_X^{-1} + \tilde{B}\tilde{M}^{-1}\tilde{A} & -\tilde{B}\tilde{M}^{-1} \\ -\tilde{M}^{-1}\tilde{A} & \tilde{M}^{-1} \end{bmatrix} \tag{8}$$

where $\tilde{A} = \Sigma_{\mathbf{YX}}\Sigma_{\mathbf{X}}^{-1}$, $\tilde{M} = \Sigma_{\mathbf{Y}} - \Sigma_{\mathbf{YX}}\Sigma_X^{-1}\Sigma_{\mathbf{XY}}$ and $\tilde{B} = \Sigma_{\mathbf{X}}^{-1}\Sigma_{\mathbf{XY}}$. Using (8), we have the decomposition

$$\begin{aligned}(\mathbf{Z}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{Z}-\boldsymbol{\mu}) &= (\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}}-\tilde{A}\mathbf{X})^\top \tilde{M}^{-1}(Y-\boldsymbol{\mu}_{\mathbf{Y}}-\tilde{A}\mathbf{X}) \\ &\quad +(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})^\top \Sigma_{\mathbf{X}}^{-1}(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}}). \end{aligned} \tag{9}$$

It then follows from (9) and some algebraic manipulation that the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ is a d_2-dimensional Student distribution with degrees of freedom $\tilde{\nu} = \nu + d_1$, location parameter $\tilde{\mu} = \boldsymbol{\mu}_2 + \tilde{A}\mathbf{x}$ and scale matrix $\frac{\lambda}{\tilde{\nu}}\tilde{M}$, where $\lambda = \nu + (\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})^\top \Sigma_{\mathbf{X}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})$.

C Proof of Theorem 2.2

For $j \in \{1,\dots,d\}$ and $x_1, x_2 \in \mathbb{R}$, set

$$H_{T,j}(x_1,x_2) = \partial_{u_1} C_{j,\hat{\alpha}_{T,j}}\left\{F_{T,j}(x_1), F_{T,j}(x_2)\right\},$$

$$H_j(x_1,x_2) = \partial_{u_1} C_{j,\alpha_j}\left\{F_j(x_1), F_j(x_2)\right\}.$$

By assumption, $\sqrt{T}\left(\hat{\alpha}_{T,j}-\alpha_j\right) \rightsquigarrow \mathbb{A}_j$, $\sqrt{T}\left(F_{T,j}-F_j\right) \rightsquigarrow \mathbb{B}_j \circ F_j$, so it is easy to check that as $T \to \infty$,

$$\begin{aligned}\sqrt{t}\left\{H_{T,j}(x_1,x_2) - H_j(x_1,x_2)\right\} &\rightsquigarrow \mathbb{H}_j(x_1,x_2) \\ &= \mathbb{A}_j^\top \partial_{u_1}\dot{C}_j\{F_j(x_1),F_j(x_2)\} \\ &\quad +\mathbb{B}_j \circ F_j(x_1)\partial_{u_1}^2 C_j\{F_j(x_1),F_j(x_2)\} \\ &\quad +\mathbb{B}_j \circ F_j(x_2)\partial_{u_1}\partial_{u_2} C_j\{F_j(x_1),F_j(x_2)\}, \end{aligned}$$

where $\dot{C}_j(u_1,u_2) = \nabla_{\alpha_j} C_{j,\alpha_j}(u_1,u_2)$, and $C_j = C_{j,\alpha_j}$.

Under the null hypothesis of independence, $\varepsilon_{j,t} = H_j(X_{j,t-1}, X_{j,t})$ are i.i.d. and uniformly distributed. Setting

$$D_T(u_1,\dots,u_d) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{d}\mathbb{I}\left(e_{T,j,t} \le u_j\right), \quad u_j \in [0,1], j \in \{1,\dots,d\},$$

and letting Π denotes the d-dimensional independence copula, it then follows from [18] that under the hypothesis of independence, $\sqrt{T}(D_T - \Pi) \rightsquigarrow \mathbb{E} - \sum_{j=1}^d \mu_j(\mathbb{H}_j)$, where

$$\begin{aligned}\mu_j(u_1,\dots,u_d) &= E\left[f\left\{F_j(X_{j,1}),F_j(X_{j,2})\right\}\prod_{k\neq j}\mathbb{I}(\varepsilon_k \le u_k) \,\middle|\, \varepsilon_j = u_j\right] \\ &= E\left[f\left\{F_j(X_{j,1}),F_j(X_{j,2})\right\} \middle| \varepsilon_j = u_j\right]\prod_{k\neq j} u_k, \end{aligned}$$

and $\mathbb{H}_j$ takes values in the set of functions of the form $f\left\{F_1(X_{1,1}),\dots,F_d(X_{d,1})\right\}$, for some functions $f \in M_j$. In our setting, one obtains that

$$\mu_j(\mathbb{H}_j) = \mathcal{B}_j(u_j)\prod_{k\neq j} u_k,$$

where

$$\begin{aligned}\mathcal{B}_j(u_j) &= \mathbb{A}_j^\top E\left[\partial_{u_1}\dot{C}_j\{F_j(\tilde{X}_{j,1}), F_j(\tilde{X}_{j,2})\}\Big|\,\tilde{\varepsilon}_j = u_j\right] \\ &\quad + E\left[\mathbb{B}_j \circ F_j(\tilde{X}_{j,1})\partial^2_{u_1} C_j\{F_j(\tilde{X}_{j,1}), F_j(\tilde{X}_{j,2})\}\Big|\,\tilde{\varepsilon}_j = u_j\right] \\ &\quad + E\left[\mathbb{B}_j \circ F_j(\tilde{X}_{j,2})\partial_{u_1}\partial_{u_2} C_j\{F_j(\tilde{X}_{j,1}), F_j(\tilde{X}_{j,2})\}\big|\,\tilde{\varepsilon}_j = u_j\right],\end{aligned}$$

with $(\tilde{X}_{j,1}, \tilde{X}_{j,2})$ is an independent copy of $(X_{j,1}, X_{j,2})$, independent of $\mathbb{B}_j$, and $\tilde{\varepsilon}_j = H_j\left(\tilde{X}_{j,1}, \tilde{X}_{j,2}\right)$.

Note that for any $j \in \{1, \ldots, d\}$, $\mathcal{B}_j(1) = 0$. As a result, one can apply [17] to obtain that for any $A \in \mathcal{A}_d$, $K_{T,A}$ has the same asymptotic behavior as

$$\tilde{\mathbb{K}}_{T,A}(u_1, \ldots, u_d) = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \prod_{j\in A} \left\{\mathbb{I}\left(\varepsilon_{j,t} \le u_j\right) - u_j\right\}, \quad u_1, \ldots, u_d \in [0, 1],$$

which in turn converge to $\mathbb{K}_A$. This comes from the fact that the so-called Möbius transform of $\mathcal{B}_j(u_j)\prod_{k\neq j} u_k$ is identically 0 [17]. This completes the proof. □

As a Corollary to Theorem 2.2, one obtains the asymptotic behavior of the empirical process

$$\mathbb{D}_{T,j}(u_j) = \sqrt{T}\left\{\frac{1}{T}\sum_{t=1}^{T} \mathbb{I}\left(e_{T,j,t} \le u_j\right) - u_j\right\}, \quad u_j \in [0, 1], j \in \{1, \ldots, d\}.$$

Note that the independence assumption is not needed here.

Corollary 1. *Under the same assumptions as in Theorem 2.2, for any $j \in \{1, \ldots, d\}$, $\mathbb{D}_{T,j} \rightsquigarrow \mathbb{D}_j = \beta_j - \mathcal{B}_j$, where β_j is a Brownian bridge with representation $\beta_j(u_j) = \mathbb{E}(1, \ldots, 1, u_j, 1, \ldots, 1)$, $u_j \in [0, 1]$.*

Remark 7. *In fact, according to [23], the empirical processes $\mathbb{D}_{T,j}$ are the one used to construct tests of goodness-of-fit for the assumption that the univariate series $X_{j,t}$ is a Markov process.*

References

[1] Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.*, 44(2), 182–198.
[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, AC-19(6), 716–723.
[3] Andersen, T., Bollerslev, T., and Diebold, F. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Rev. Econ. Stat.*, 89(4), 701–720.
[4] Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.*, 96(453), 42–55.
[5] Beare, B. (2010). Copulas and temporal dependence. *Econometrica*, 78(1), 395–410.
[6] Beare, B. K. and Seo, J. (2015). Vine copula specifications for stationary multivariate Markov chains. *J. Time. Ser. Anal.*, 36, 228–246.
[7] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York, second edition.
[8] Bush, T., Christensen, B., and M.Ø., N. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *J. Econometrics*, 60(1), 48–57.
[9] Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric model time series models. *J. Econometrics*, 130(2), 307–335.
[10] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *J. Financ. Econ.*, 7(2), 174–196.
[11] Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *J. Bus. Econom. Statist.*, 13(3), 253–263.
[12] Duchesne, P., Ghoudi, K., and Rémillard, B. (2012). On testing for independence between the innovations of several time series. *Canad. J. Statist.*, 40(3), 447–479.
[13] Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Economet. Theor.*, 11(1),122–150.
[14] Erhardt, T. M., Czado, C., and Schepsmeier, U. (2014). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, to appear. DOI:10.1111/biom.12279
[15] Fang, H.-B., Fang, K.-T., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.*, 82(1), 1–16.

[16] Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009). The advent of copula in finance. *Europ. J. Financ.*, 15(7-8), 609–618.
[17] Genest, C. and Rémillard, B. (2004). Tests of independence or randomness based on the empirical copula process. *Test*, 13(2), 335–369.
[18] Ghoudi, K. and Rémillard, B. (2004). Empirical processes based on pseudo-observations. II. The multivariate case. In *Asymptotic Methods in Stochastics*, 381–406. Amer. Math. Soc., Providence, RI.
[19] Kurowicka, D. and Joe, H., editors (2011). *Dependence Modeling. Vine Copula Handbook.* World Scientific, Hackensack, NJ.
[20] Martens, M. and van Dijk, D. (2006). Measuring volatility with the realized range. *J. Econometrics*, 138(1), 181–207.
[21] Nelsen, R. B. (1999). *An introduction to copulas*. Springer-Verlag, New York.
[22] Rémillard, B. (2013). *Statistical Methods For Financial Engineering*. CRC Press, Boca Raton, FL.
[23] Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *J. Multivariate Anal.*, 110, 30–42.
[24] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer-Verlag, Berlin.
[25] Smith, M. (2015). Copula modelling of dependence in multivariate time series. *Int. J. Forecasting*, to appear. DOI:10.1016/j.ijforecast.2014.04.003
[26] Sokolinskiy, O. and Van Dijk, D. (2011). Forecasting volatility with copula-based time series models. Technical report, Tinbergen Institute Discussion Paper.
[27] Soustra, F. (2006). Pricing of synthetic CDO tranches, analysis of base correlations and an introduction to dynamic copulas. Master thesis, HEC Montréal.
[28] Zhang, L., Mykland, P., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472), 1394–1414.
[29] Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *J. Bus. Econom. Statist.*, 14(1), 45–52.