Les Cahiers du GERAD ISSN: 0711–2440

Forecasting Time Series with Multivariate Copulas

C. Simard
B. Rémillard

G–2013–85

November 2013

Forecasting Time Series with Multivariate Copulas

Clarence Simard

GERAD & Department of Mathematics and Statistics
Université de Montréal
Montréal (Québec) Canada, H3C 3J7

clarence.simard@hotmail.com

Bruno Rémillard

GERAD & Department of Management Sciences
HEC Montréal
Montréal (Québec) Canada, H3T 2A7

bruno.remillard@hec.ca

November 2013

Les Cahiers du GERAD

G–2013–85

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 and compare it with a univariate version of our forecasting method introduced recently by Sokolinskiy & Van Dijk. Moreover, we also study the influence of the marginal distribution with the help of a new performance measure and lastly we look at the impact of the dependence structure on the predictions performance. We also give an example of practical implementation with financial data.

Résumé : Dans cet article, on présente une méthode de prévision pour des séries chronologiques multidimensionnelles, modélisées à l'aide de copules. On étudie comment la performance des prévisions évolue en fonction de la force de la dépendance et l'on compare aussi nos résultats avec ceux obtenus par Sokolinskiy & Van Dijk dans le cas unidimensionnel. Par ailleurs, nous étudions également l'influence des lois marginales à l'aide d'une nouvelle mesure de la performance, et enfin nous examinons l'impact de la structure de dépendance sur la performance des prévisions. Nous donnons également un exemple de mise en œuvre pratique avec des données financières.

1 Introduction

For many years, copulas have been used for modeling dependence between random variables. See e.g. Genest et al. (2009) 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 their temporal dependence in time series, first in the univariate case, as in Chen and Fan (2006) and Beare (2010), and then in a multivariate setting (Rémillard et al., 2012). 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 Sokolinskiy and Van Dijk (2011), copulas were used to forecast the realized volatility associated with a univariate financial time series, and it was shown there that copula-based forecasts perform better than forecasts based on heterogeneous autoregressive (HAR) model, Corsi (2009). The later method had been proven successful in Andersen et al. (2007), Corsi (2009) and Bush et al. (2011).

In this paper, we extend the methodology of Sokolinskiy and Van Dijk (2011) by proposing a forecasting method using copula-based models for multivariate time series, as in Rémillard et al. (2012). 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 $\{(X_{1,t}, X_{2,t}); t = 0, 1, \ldots\}$ 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 informations than the knowledge of $X_{1,T}$ alone.

As a second matter, 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 Rémillard et al. (2012), which provide the estimation methods, are given for an arbitrary number of time series and most of the theoretical background is going to be presented in the general case. Moreover, the results of our numerical experiments should naturally extend to the multivariate case.

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. Section 3 contains the result of our numerical experiments as well as the analysis of the results. We also give a complete example of practical implementation with financial data 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 about copulas can be found in Nelsen (1999) and Rémillard (2013).

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 $P(U_i \le u) = u_i$ for $i = 1, \ldots, d$ 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 $Range(F_1) \times \cdots \times Range(F_d)$. Hence, if $Range(F_i) = [0, 1]$ for $i = 1, \ldots, d$ the copula is unique.

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

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

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_i of F_i exists for each $i = 1, \ldots, d$, then the density c of C exists if and only if the density h of H also exists. It this case, differentiating equation (1), we get

$$h(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d))\Pi_{i=1}^d f_i(x_i).$$

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

$$c(u_1, \ldots, u_d) = \frac{h(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{\Pi_{i=1}^d f_i(F_i^{-1}(u_i))}. \tag{2}$$

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 $F_1, \ldots, F_{d_1}$ and $\mathbf{Y}$ has marginal distributions $G_1, \ldots, G_{d_2}$.

Setting $\mathbf{F}(X) = (F_1(X_1), \ldots, F_{d_1}(X_{d_1}))$, $\mathbf{G}(Y) = (G_1(Y_1), \ldots, G_{d_2}(Y_{d_2}))$ and defining the random vector $(\mathbf{U}, \mathbf{V}) = (\mathbf{F}(X), \mathbf{G}(Y))$ we can define the copula $C_{\mathbf{UV}}$ of the vector (X, Y) as the joint distribution function of (U, V). Assuming that the density functions exist and applying equation (**??**), one obtains that the conditional copula $C_{\mathbf{U}|\mathbf{V}}$, i.e., the conditional distribution of U given V, is given by

$$C_{\mathbf{U}|\mathbf{V}}(\mathbf{u}; \mathbf{v}) = \frac{\partial_{v_1} \cdots \partial v_{d_2} C_{\mathbf{UV}}(u, v)}{c_{\mathbf{V}}(\mathbf{v})},$$

with density

$$c_{\mathbf{U}|\mathbf{V}}(\mathbf{u}; \mathbf{v}) = \frac{c_{\mathbf{UV}}(u, v)}{c_{\mathbf{V}}(\mathbf{v})},$$

where $c_{\mathbf{V}}$ is the density of the copula $C_{\mathbf{V}}(\mathbf{v}) = C_{\mathbf{UV}}(1, \ldots, 1, v)$ associated to $\mathbf{Y}$ or $\mathbf{V}$.

Having defined conditional copulas, one can now look at how to obtain 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 Soustra (2006) and Rémillard et al. (2012), extending the results of Chen and Fan (2006) to the multivariate case.

Let $\mathbf{X} = \{\mathbf{X}_t; t = 0, 1, \ldots\}$ be a d-dimensional time series and assume that $\mathbf{X}$ is Markovian and stationary. We note F_i the marginal distribution of $X_{i,t}$ for $i = 1, \ldots, d$ 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 $F_1, \ldots, F_d$ 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 and unique copula Q associated to $\mathbf{X}_{t-1}$, viz.

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

where $\mathbf{1}_d$ is the d-dimensional unit vector. Set $\mathbf{U}_t = \mathbf{F}(\mathbf{X}_t)$, for $t \geq 0$.

The next step is to deduce the conditional copula of X_t given X_{t-1}, which is

$$\mathcal{C}(\mathbf{u}; \mathbf{v}) = C_{U_t|U_{t-1}}(\mathbf{u}; \mathbf{v}) = \frac{\partial_{v_1} \cdots \partial v_d C(u, v)}{q(\mathbf{v})},$$

with density

$$c_{\mathbf{U}_t|\mathbf{U}_{t-1}}(\mathbf{u}; \mathbf{v}) = \frac{c(u, v)}{q(\mathbf{v})},$$

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 X_t given 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, for practical implementation, these distributions are unknown and estimations are to be done. This will be treated next.

The use of a copula-based model for time series allows for a more flexible model of the dependence structure.

Let $\mathbf{X} = \{\mathbf{X}_t; t = 0, 1, \ldots, T\}$ be a d-dimensional time series. Our goal is to forecast X_{T+1} based on the information available at time T. Suppose that for all $t \geq 0$, F_i is the marginal distribution of $X_{i,t}$ for $i = 1, \ldots, d$ and the $2d$-dimensional vector $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ as joint distribution H and copula C. Using the preceding 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 observe the value $\mathbf{X}_T = \mathbf{y}$ for the time series at time T. The prediction of $X_{1,T+1}$ goes as follows:

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

$$\hat{\mathbf{X}}_{T+1} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}^{(i)}_{T+1}. \tag{3}$$

   We use $\hat{\mathbf{X}}_{T+1}$ as a predictor for $\mathbf{X}_{T+1}$.

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 $\{X^{(i)}_{1,T+1}; i = 1, \ldots, n\}$. We denote by $\widehat{LB}_{T+1}$ and $\widehat{UB}_{T+1}$ the lower and upper values for the prediction interval.

Remark 1 *We would like to mention that for a one period ahead prediction, the simulation of the $X_{i,T+1}$ for $i = 2, \ldots, d$ is useless. We could have built the predictor $\hat{X}_{1,T+1}$ on the conditional distribution of $X_{1,T+1}$ given $(X_{1,T}, \ldots, X_{d,T})$. However, including it allows to extend the methods to predictions of a longer horizon. To get a prediction for the value $X_{1,T+K}$ for $K > 1$, one recursively predicts X_{T+j} for $j = 2, \ldots, K$ by replacing the known value at $T + j - 1$ by $\hat{X}_{T+j-1}$.*

As mentioned previously, we are going to compare our predictions performance with the univariate version presented in Sokolinskiy and Van Dijk (2011). Let D be the copula associated with $(X_{1,t-1}, X_{1,t})$ for $t = 0, 1, \ldots$, i.e., D is the copula of $(U_{1,t-1}, U_{1,t})$.

Suppose we observe the value $X_{1,T+1} = y_1$; The predictor presented in Sokolinskiy and Van Dijk (2011) is defined as

$$\bar{X}_{1,T+1} = n^{-1} \sum_{i=1}^{n} F_1^{-1}(Z^{(i)}) \tag{4}$$

where $Z^{(i)}$ are realizations of $\mathcal{D}(\cdot; v_1)$ where $v_1 = F_1(y_1)$ 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 $\alpha/2$ and $1 - \alpha/2$ quantiles of $\{X_{1,T+1}^{(i)}; i = 1, \ldots, n\}$.

2.3.1 Implementation in practice

For practical implementation, one has to replace the known distributions $\mathbf{F}$ and the copula C by estimated versions. The estimation method for copula-based model for time series is presented in Rémillard et al. (2012) which use non-parametrical estimation for marginal distribution and parametrical estimation for conditional copula where the copula parameters are estimated through pseudo maximum likelihood. Goodness-of-fit tests are also provided to help choose the right copula family, but in the context of forecasting, one can also choose copulas by their prediction power.

2.4 Including more information

The methodology proposed here can also be applied to predict $X_{1,T+1}$ given $\mathbf{X}_t$ and $X_{2,T+1}, \ldots, X_{d,T+1}$, since the joint copula of $(\mathbf{X}_t, \mathbf{X}_{t+1})$ is given.

3 Simulation results

To better understand the performance of our forecasting method we use simulated datas and compare the performance of our predictor with the predictor from Sokolinskiy and Van Dijk (2011). We restrict our simulations to the case of bivariate series but the results can easily be extrapolated to higher dimensions.

The reason why our multivariate method gives better results is obviously because we are using the additional information provided by the second series. Consequently, the gain in performance is affected by the strength of the different dependencies as well as the dependence structure of 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 first simulate datas from the Student copula. This choice is motivated by the fact that we can directly specify 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 $i, j = 1, \ldots, d$ are the elements of the correlation matrix, the Kendall's tau matrix for the Student copula is defined as $\tau_{i,j} = \frac{2}{\pi} arcsin(R_{i,j})$, see Rémillard (2013) proposition 8.7.1. In order to test a different dependence structure we also use datas simulated from the Clayton copula. See Appendix A for details about simulating Student and Clayton copulas.

Another question is the impact of the marginal distributions. To this matter it seems inappropriate to compare directly the predictions when the marginal distributions are different. We can expect that the predictions of a random variable with large variance should be less precise than when the variance is small. To eliminate this effect we propose a new measure of performance.

For pointwise predictions we use two different performance measures. First, the mean absolute error, which is defined as

$$MAE(\hat{X}) = N^{-1} \sum_{t=0}^{N-1} |X_t - \hat{X}_t|$$

for a predictor $\hat{X}$. The second performance measure is designed to compare prediction error through different marginals distributions. We call this measure mean absolute rank error and it is defined as

$$MARE(\hat{X}) = N^{-1} \sum_{i=0}^{N-1} |F(X_t) - F(\hat{X}_t)|$$

where X_t is the observed values, $\hat{X}_t$ are the predictions and F is the marginal distribution of X_t for all $t = 0, 1, \ldots$

In addition to pointwise predictions we also compute confidence intervals at a 95% confidence level for both predictors. To measure the performance of the confidence intervals we take the mean length of the confidence intervals. Let CI_t^u and CI_t^l be the upper and lower value of the confidence interval for $t = 0, \ldots, N-1$, we define the mean length as

$$ML(CI) = N^{-1} \sum_{t=0}^{N-1} \left(CI_t^u - CI_t^d \right).$$

3.1 Impact of the dependencies strength

As we already said, 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 how this impact appears we first simulate $\mathbf{X}_t$ with Student copula and study the impact of each the different possible correlations. We choose a degree of freedom $\nu = 8$ 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 $\nu = 8$ degree of freedom, that is the vector $\mathbf{X}_t$ follows a multivariate distribution. For these experiments we simulate series of length $N = 1000$ and we set $n = 300$ in the definition of the predictors (3) and (4).

The first simulation study the impact of the dependence between both series. We simulate the series using Kendall's tau matrices

$$\tau_\alpha = \begin{bmatrix} 1 & \alpha & 0.1609 & 0.1609 \\ \alpha & 1 & 0.1609 & 0.1609 \\ 0.1609 & 0.1609 & 1 & \alpha \\ 0.1609 & 0.1609 & \alpha & 1 \end{bmatrix}$$

with

$$\begin{aligned} \alpha \in \big\{ & -0.2620, -0.1940, -0.1282, -0.0638, -0.0318, -0.0064, \\ & 0.0064, 0.0318, 0.0638, 0.1282, 0.1940, 0.2620, 0.3333, \\ & 0.4097, 0.4936, 0.5903, 0.7129 \big\}. \end{aligned} \tag{5}$$

As we see in Figure 1 the error of prediction increases when the Kendall's tau ($[\tau]_{1,3}$ and $[\tau]_{3,1}$) increases. The same pattern appears for the mean length of confidence intervals, see Figure 2. To explain this result, suppose the extreme case where the correlation between X_1 and X_2 is one. Then the two series are identical and hence our predictor has no additional information coming from the second serie. This also explain why the difference in the prediction error gets close to zero when the correlation is high, or equivalently the Kendall's tau. Again, as long as the correlation increases, the predictor $\bar{X}_1$ does as good $\hat{X}_1$ since the information from the second serie becomes irrelevant.

For the second simulation, Figure 3 and 4, we study the impact of the correlation between $X_{1,t}$ and $X_{1,t-1}$. We simulate the series using Kentall's tau matrices

$$\tau_\alpha = \begin{bmatrix} 1 & 0.1609 & \alpha & 0.1609 \\ 0.1609 & 1 & 0.1609 & 0.1609 \\ \alpha & 0.1609 & 1 & 0.1609 \\ 0.1609 & 0.1609 & 0.1609 & 1 \end{bmatrix}$$

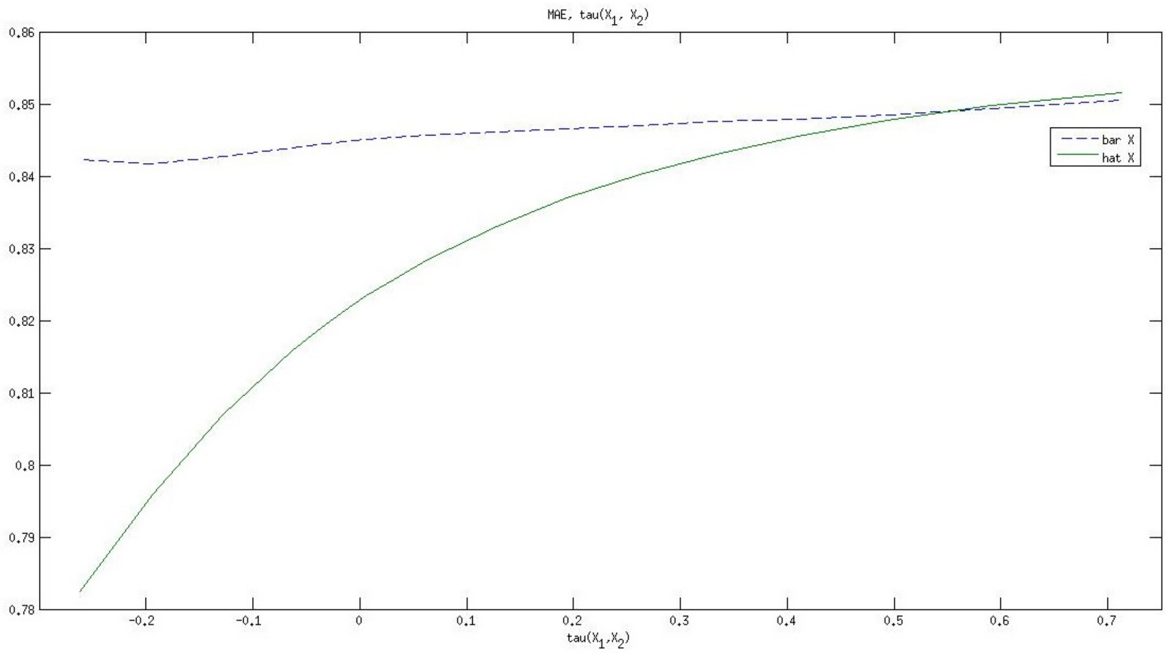

Figure 1: Evolution of MAE in function of the strength of the dependence between X_1 and X_2. The dashed line is the $MAE(\bar{X})$ and the continuous line is $MAE(\hat{X})$.

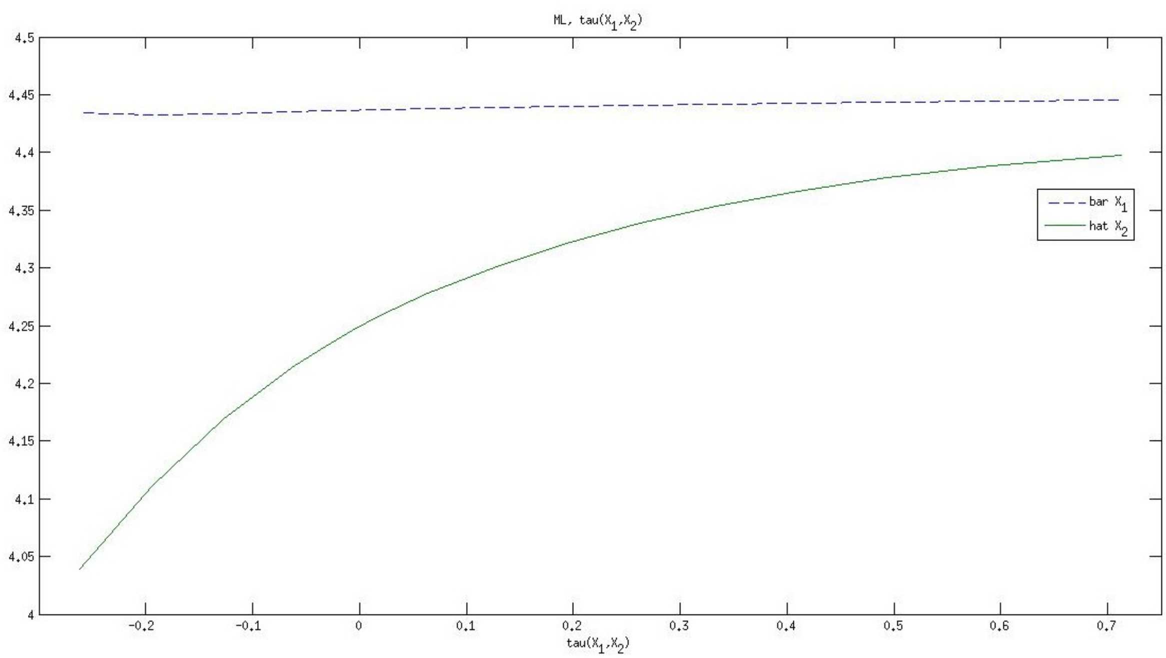

Figure 2: Evolution of $ML(CI)$ in function of the strength of the dependence between X_1 and X_2. The dashed line is the $ML(\bar{C}I)$ and the continuous line is the $ML(\hat{C}I)$.

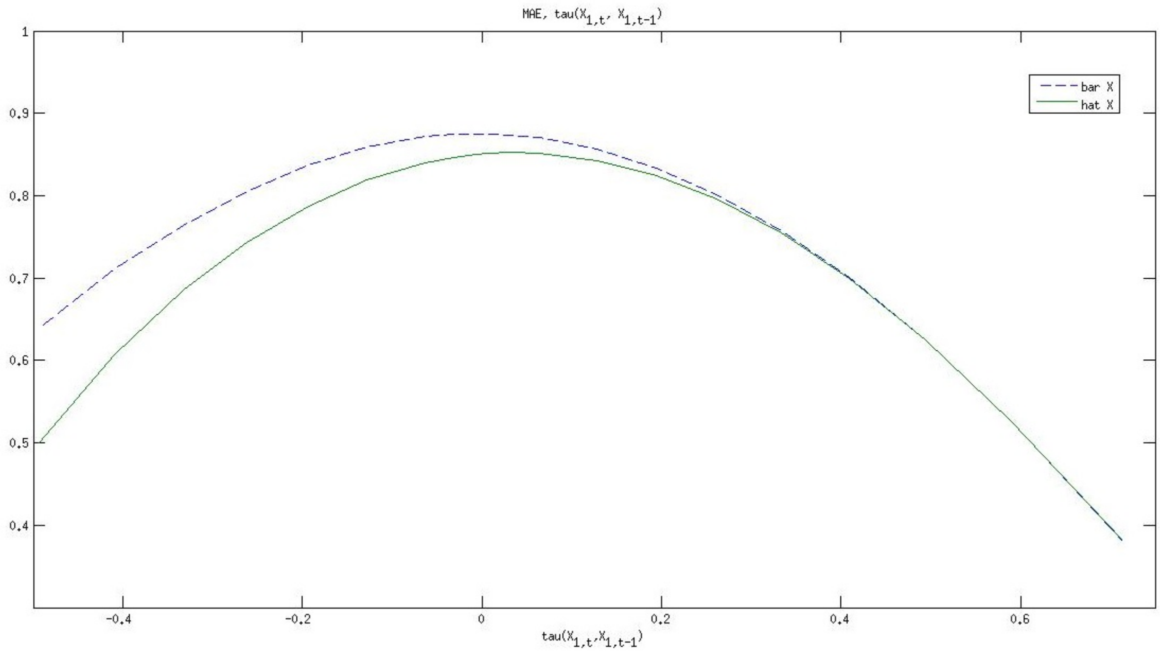

Figure 3: Evolution of MAE in function of the strength of the dependence between $X_{1,t}$ and $X_{1,t-1}$. The dashed line is the $MAE(\bar{X})$ and the continuous line is $MAE(\hat{X})$.

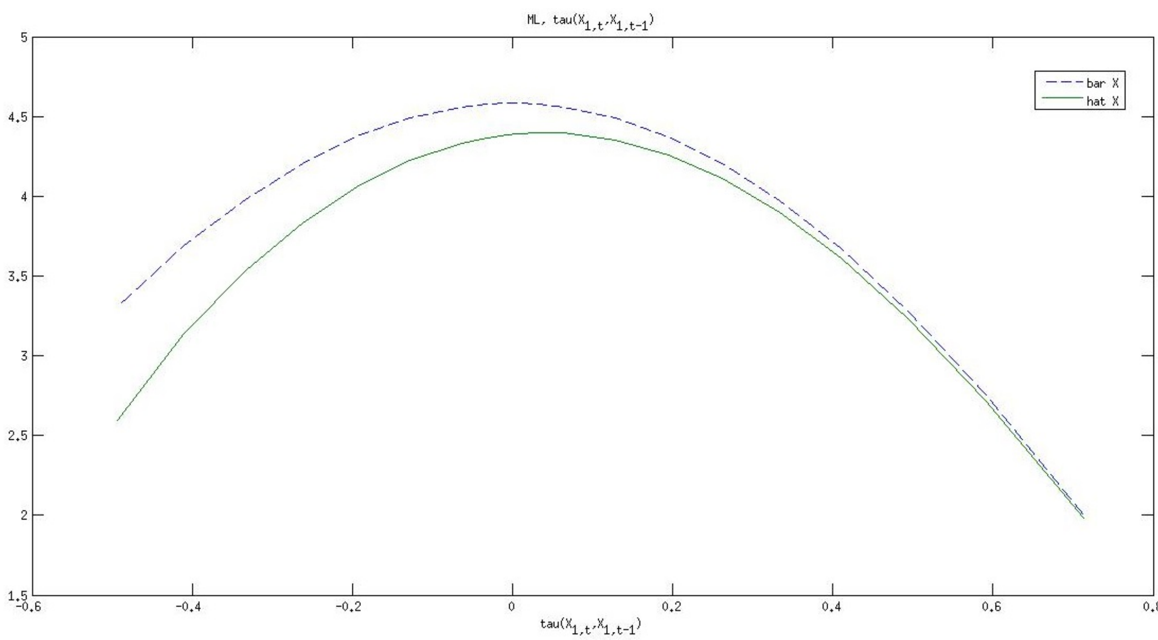

Figure 4: Evolution of $ML(CI)$ in function of the strength of the dependence between $X_{1,t}$ and $X_{1,t-1}$. The dashed line is the $ML(\bar{C}I)$ and the continuous line is the $ML(\hat{C}I)$.

with

$$\alpha \in \big\{ -0.4936, -0.4097, -0.3333, -0.2620, -0.1940, -0.1282, -0.0638, \\ -0.0318, -0.0064, 0.00640.0318, 0.0638, 0.1282, 0.1940, \\ 0.2620, 0.3333, 0.4097, 0.4936, 0.5903, 0.7129\big\}. \quad (6)$$

We remark that range of α is not the same than before. The reason is that we have to keep the correlation matrices positive semi-definite. As we can expect, the predictions are better when the dependence is strong. But it is interesting to notice that $\hat{X}$ benefits more from the negative correlation than $\bar{X}$. 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 serie becomes marginal. That is why the difference in prediction error is close to zero when the correlation is close to one.

The last simulation is about the impact of the strength of the dependence between $X_{1,t}$ and $X_{2,t-1}$, see Figures 5 and 6. As we 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 $\bar{X}$, our predictor gives much better performance when this dependence is strong. In the case where this dependence is strong, the information of $X_{2,t-1}$ almost completely dictate the value of $X_{1,t}$ and that is why we observe a great difference in the performance of $\hat{X}$ and $\bar{X}$.

3.2 Impact of the marginal distributions

Another question we want to tackle is the impact of the marginal distributions. In order to isolate more closely the impact of the correlation in our first simulations we did not really use copulas as it is usually meant to, since a Student copula with ν degree of freedom with Student margins with also ν degree of freedom is simply a multivariate Student distribution. 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 Kendall's tau matrix at $\tau_{i,j} = 0.1609$ for all $i \neq j$.

Table 1: Evolution of MAE and $MARE$ in function of the marginal distributions.

Marginals distributions	$MARE(\bar{X})$	$MARE(\hat{X})$	$MAE(\bar{X})$	$MAE(\hat{X})$
T(·;5)	0.240260	0.207765	0.910358	0.815375
T(·;8)	0.240165	0.207626	0.849395	0.756528
T(·;10)	0.240162	0.207623	0.831341	0.739301
T(·;15)	0.240180	0.207648	0.808801	0.717971
T(·;20)	0.240202	0.207661	0.798150	0.707883
T(·;30)	0.240225	0.207676	0.787869	0.698152
T(·;50)	0.240244	0.207691	0.779895	0.690627
log-normal(·;0,1)	0.271711	0.232255	1.217649	1.095158
Chi-squared(·;8)	0.243123	0.211028	3.015161	2.679748
Exp(·;3)	0.256026	0.221556	2.137274	1.896555
Normal(·;0,1)	0.240274	0.207719	0.768335	0.679747
Normal(·;0,2)	0.240274	0.207719	1.086590	0.961308
Normal(·;0,4)	0.240274	0.207719	1.536670	1.359494
Normal(·;0,8)	0.240274	0.207719	2.173180	1.922615
Normal(·;2,1)	0.240274	0.207719	0.768335	0.679747
Normal(·;4,1)	0.240274	0.207719	0.768335	0.679747
Normal(·;8,1)	0.240274	0.207719	0.768335	0.679747

Our first observation is that for the same marginal distribution the MAE is affected by the change of parameters while the $MARE$ is almost constant. This observation seems to confirm that our performance measure behaves as we intended to. So, looking at the $MARE$ values we conclude that it is the general shape of the distribution that affect the performance of the predictions. We see that for symmetric distributions (normal and Student) the results are very close but they differs for the asymmetric ones (exponential, chi-squares and log-normal). For the same experiment we also computed the mean length of the confidence interval, see Table 2. We see that the behavior is the same as the MAE.

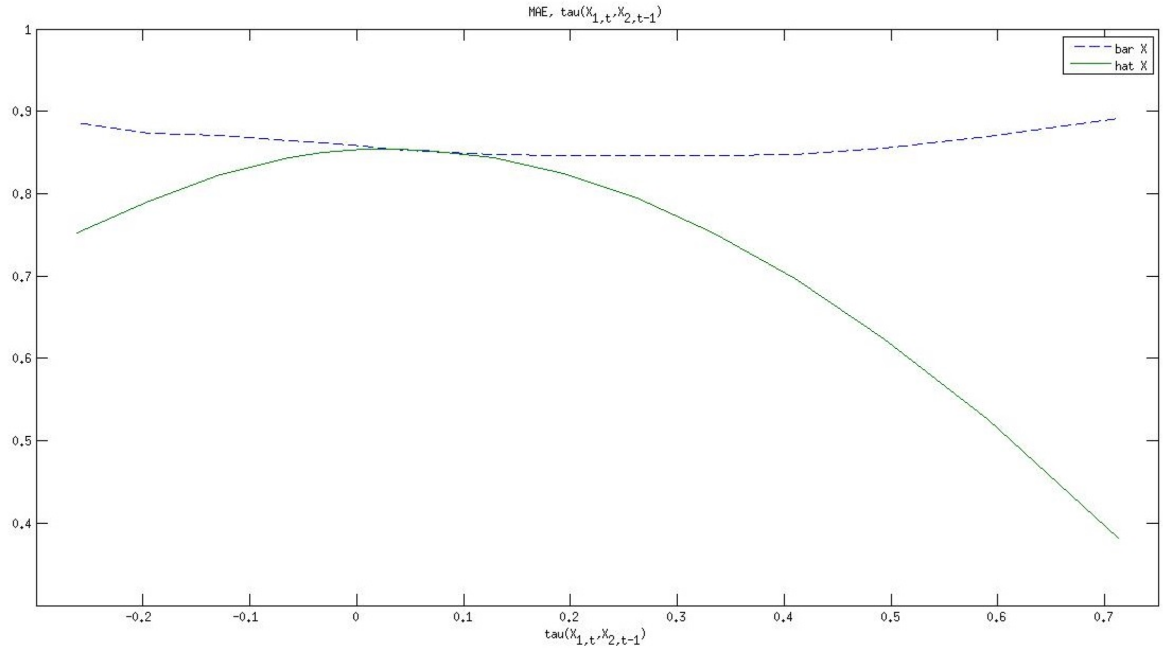

Figure 5: Evolution of MAE in function of the strength of the dependence between $X_{1,t}$ and $X_{2,t-1}$. The dashed line is the $MAE(\bar{X})$ and the continuous line is $MAE(\hat{X})$.

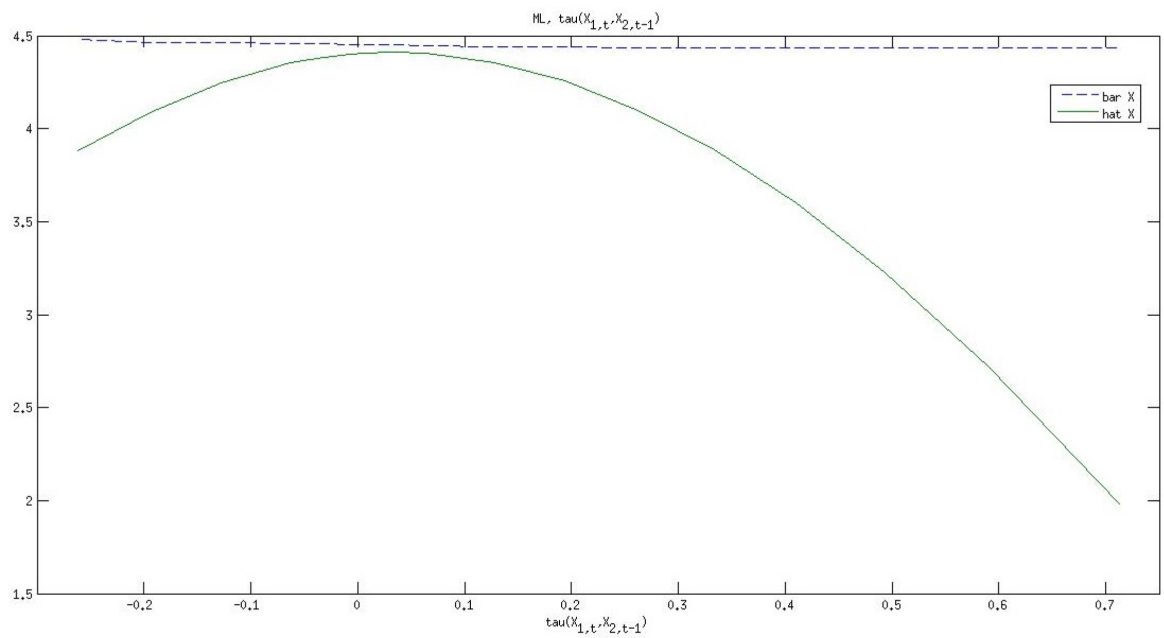

Figure 6: Evolution of $ML(CI)$ in function of the strength of the dependence between $X_{1,t}$ and $X_{2,t-1}$. The dashed line is the $ML(\bar{C}I)$ and the continuous line is the $ML(\hat{C}I)$.

Table 2: Evolution of MAE and $MARE$ in function of the marginal distributions.

Marginals distributions	$ML(\bar{C}I)$	$ML(\hat{C}I)$
T(·;5)	4.917631	4.330145
T(·;8)	4.426599	3.859385
T(·;10)	4.273358	3.755346
T(·;15)	4.088474	3.594310
T(·;20)	4.004253	3.552873
T(·;30)	3.928587	3.458193
T(·;50)	3.859412	3.421150
Log-normal(·;0,1)	6.725636	5.875235
Chi-squared(·;8)	14.721604	13.119912
Exp(·;3)	10.538524	9.452441
Normal(·,0,1)	3.756359	3.337330
Normal(·,0,2)	5.315825	4.732659
Normal(·,0,4)	7.533007	6.676207
Normal(·,0,8)	10.651699	9.433832
Normal(·,2,1)	3.762691	3.338162
Normal(·,4,1)	3.757399	3.342431
Normal(·,8,1)	3.754095	3.346136

3.3 Impact of the dependence structure

Our last numerical experiment shows that the dependence structure has an impact on the gain in performance of our predictor $\hat{X}$ compare to $\bar{X}$. At first sight, we could think that using the information provided by the serie 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 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 increase 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 serie X_2 cannot provides useful informations to our predictor.

The results in Figures 7 and 8 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 that the difference between both prediction errors is slowly decreasing for high values of θ. This is due to the fact that the correlation 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. Obviously, the multivariate version of the predictor will always have an advantage but in the context of the Clayton copula this advantage is minor.

4 Application

In this section we present an application of our method for forecasting 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 the quadratic variation at time T, $[X]_T$. Thus, an estimator of the realized volatility can be defined as the sum of squared returns

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

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 Zhou (1996) but we refer the reader to Andersen et al. (2001) for a detailed justification of the realized volatility estimation. It is well known that each price observation is polluted by some random noise and 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

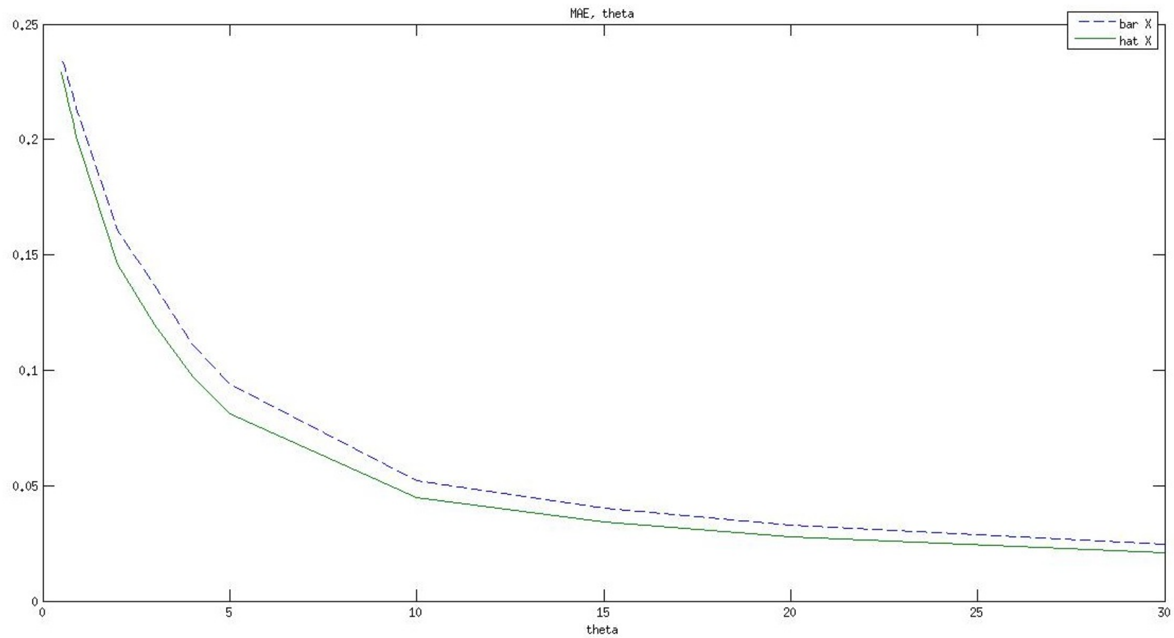

Figure 7: Evolution of MAE for the predictors $\bar{X}$ and $\hat{X}$ in function of the parameter θ in Clayton copula

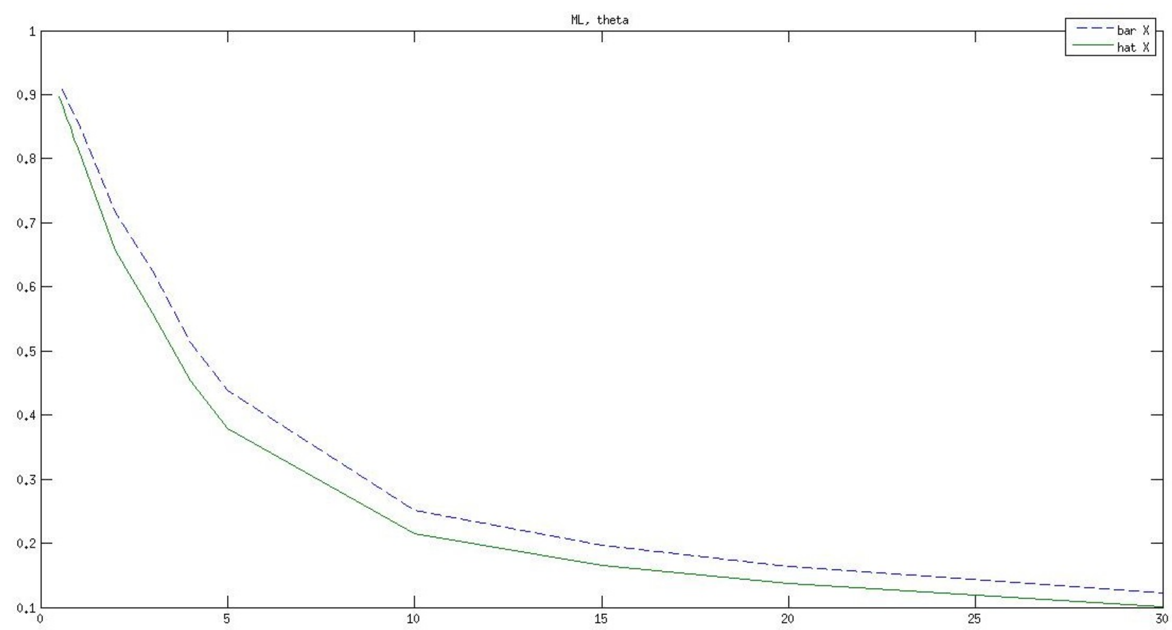

Figure 8: Evolution of $ML(CI)$ for the predictors $\bar{X}$ and $\hat{X}$ in function of the parameter θ in Clayton copula

due to noise is somewhat diminished and the estimation precision becomes acceptable. However, to perform our realized volatility estimation we used the estimator of Zhang et al. (2005) which is an asymptotically unbiased estimator an allows to used high-frequency datas. Another good estimator is given by Martens and van Dijk (2006) which makes use of high and low observed values. The reason we preferred the former estimator is that we used trade prices and it seems that this estimator is less affected by bid/ask spread.

The datas we are using are from de Trade and Quote database. We used Apple.inc (APPL) trade prices from 2006/08/08 to 2008/02/01 which consists of 374 days of trading. In order to avoid periods of lower frequency trading we used datas from 9:00:00 to 15:59:59. In combination with the estimation of realized volatility we also compute the aggregated volume of transactions, see Figure 9. For our time series to satisfy the required hypothesis of stationarity we had to take the first difference of the logarithm of both series. We define $X_{2,t} = Log(\hat{rv}_t) - Log(\hat{rv}_{t-1})$ and $X_{2,t} = Log(\hat{vol}_t) - Log(\hat{vol}_{t-1})$ where $\hat{rv}_t$ is the estimated volatilty, $\hat{vol}_t$ is the aggregate volume of transaction and time is in days. To verify the stationarity assumption of both series we used a non-parametric change point test using the Kolmogorov-Smirnov statistic. The p-values of 0.217 and 0.341 for the series X_1 and X_2 lead us to not reject the null hypothesis of stationarity. We carried parameters estimation and goodness of fit tests for Clayton, Frank, Gaussian and Student copulas. From the p-values given by the goodness of fit test, see Table 3, we selected the Student copula as the best model for $(X_{1,t-1}, X_{2,t-1}, X_{2,t}, X_{2,t})$. The estimated parameters for the Student distribution are the degree of freedom, $\hat{\nu} = 12.6451$, 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 Kendall's tau matrix is then

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

Then we use our algorithm to make one period ahead predictions for out of sample values of the series X_1. The result of the prediction are in Figure 11. It is not clear from the graphics but if we take the mean length of the confidence interval over the 100 forecasts we get 2.1181 for the predictions using bivariate copula and 2.1523 for the univariate, which asserts that the bivariate copula gives a better prediction.

Table 3:

Copula	p-value
Clayton	0
Frank	0
Gaussian	0.037
Student	0.0931

5 Concluding remarks

From the work of Sokolinskiy and Van Dijk (2011) we knew that copulas could be successfully used to forecast time series. In this paper we showed that using copula-based model for multivariate time series it is possible to further improve predictions. To understand the advantages of our multivariate predictor we studied how the performance of predictor evolve in function of the strength of the dependencies as well as the structure of the dependence. Using the $MARE$ performance measure we also showed that the shape of the marginal distribution might affect the performance of the predictions in an absolute manner since the

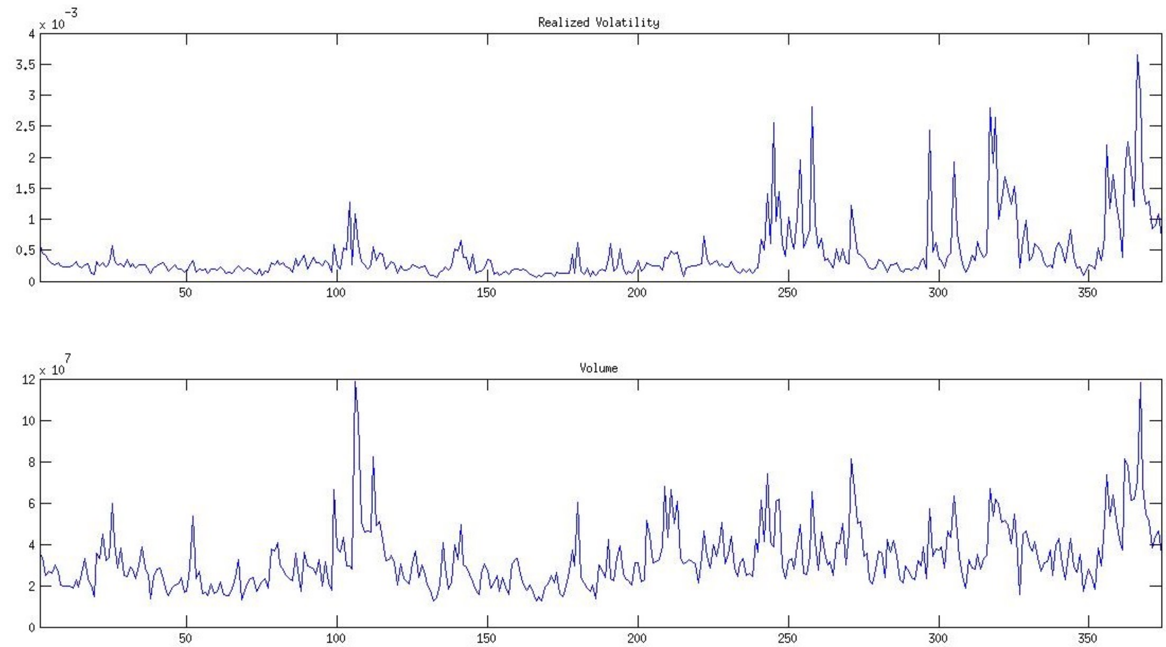

Figure 9: On top figure, the estimated realized and the daily volume of transactions.

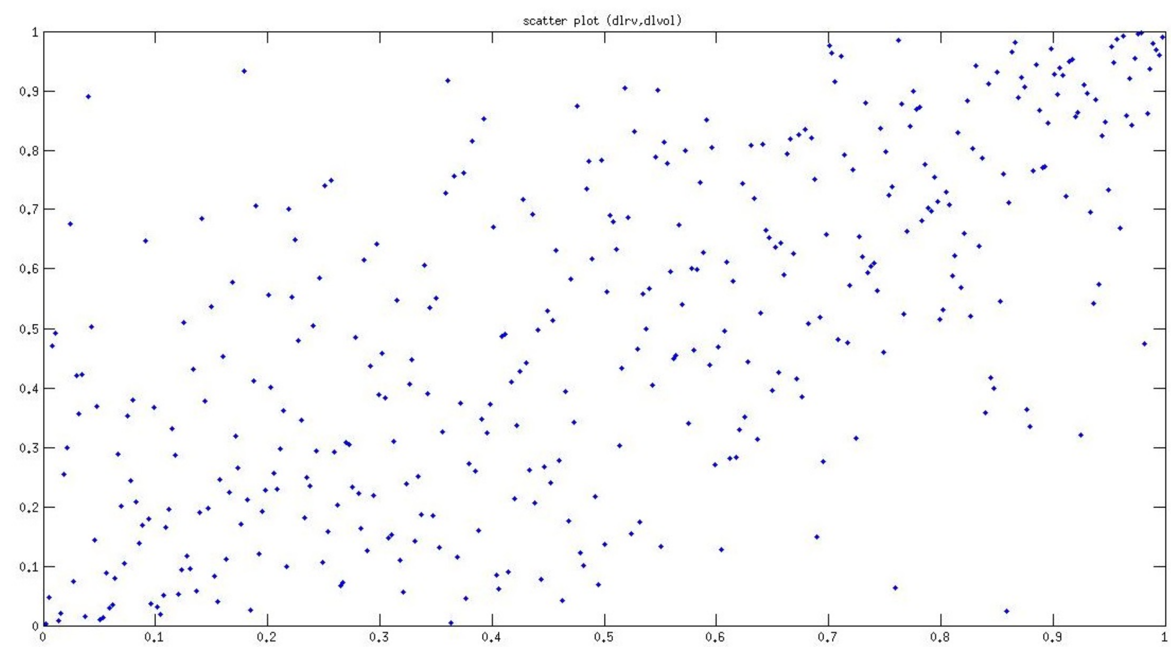

Figure 10: Scatter plot for the first difference of the log for the realized volatility and the volume of transaction.

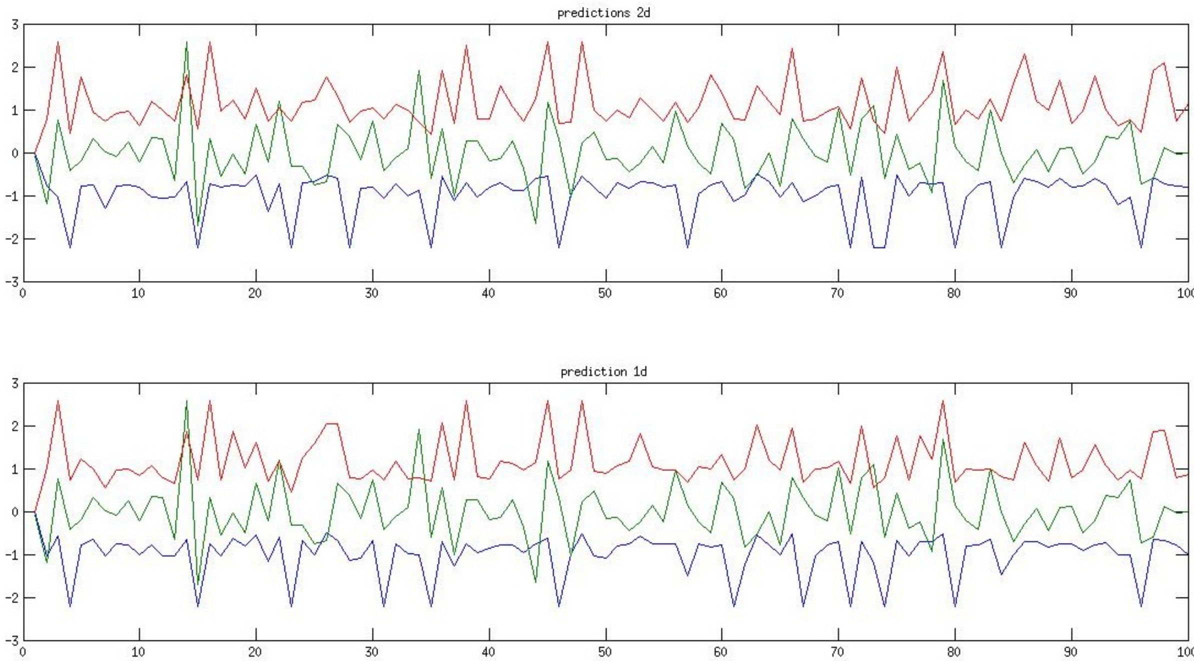

Figure 11: On top figure, predictions of the first difference of the log of the realized volatility using 95% confidence interval with bivariate copula on top panel and univariate copula on the lower panel.

$MARE$ measure remove the scaling effect of the marginal distributions. Finally, based on the estimation methods and goodness-of-fit tests provided by Rémillard et al. (2012) we could present a complete practical implementation of the forecasting method.

A Simulation

A.1 Simulation of time series with Student copula

The Student copula is based on a multivariate Student distribution. Suppose (X, Y) is a $d = (d_1 + d_2)$-dimensional random vector which follows a Student distribution with mean 0, correlation matrix R and ν degree of freedom. We write the matrix R as a block matrix

$$R = \begin{bmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{bmatrix}$$

where R_X, R_Y, R_{YX} and R_{XY} are respectively the correlation matrices of the variables in subscript. It is possible to show that all joint distributions of a multivariate Student vector are also of Student distributions with respective correlation matrix and the same degree of freedom. Let $T_{\nu,R}$ be the distribution function of a multivariate Student vector. The Student copula, noted $C_{\nu,R}$ is defined as

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

Using Schur complement on the correlation matrix it is possible to show that the conditional copula of Y given X is also a Student copula with $\tilde{\nu} = \nu + d_1$ degree of freedom, correlation matrix $\tilde{R} = \frac{\tilde{\nu}}{2}\Omega$ and mean $\mu = XB^T$ where $\Omega = R_Y - R_{XY} R_X^{-1} R_{XY}$ and $B = R_{XY} R_X^{-1}$. The details of the derivations are given in Appendix B.

To generate a $2d$-dimensional time serie $\{X_t\}_{t=0,1,\ldots}$ such that (X_{t-1}, X_t) follows a Student conditional copula $C_{\nu,R}$ with marginal distributions $F_1, \ldots, F_d$ we use the following algorithm:

1. Generate Y_0 from a d-dimensional Student distribution with ν degree of freedom and correlation matrix R_X where R_X is the correlation matrix of X_t.
2. For all $t = 1, 2, \ldots$, generate Y_t from a d-dimensional Student distribution with $\tilde{\nu}$ degree of freedom, correlation matrix $\tilde{R}$ and mean $X_{t-1}B^T$.
3. Set $U_t = T_\nu(Y_t)$.
4. Define $(X_{1t}, \ldots, X_{dt}) = (F_1^{-1}(U_{1t}), \ldots, F_d^{-1}(U_{dt}))$.

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 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

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

for any bijection $\phi : [0, 1) \to [0, \infty)$. Archimedean copulas are uniquely defined by the generator up to a 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 (U, V) is a $(d_1 + d_2)-$dimensional random vector which follows a Clayton copula C_{ϕ_θ}. Then it is possible to show that the conditional copula of V given U is Clayton copula with parameter $\tilde{\theta} = \frac{\theta}{1+d_1\theta}$.

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

1. Generate U_0 from the distribution C_{ϕ_θ}.
2. For all $t = 1, 2, \ldots$ set $A_{U_{t-1}} = C_{\phi_\theta}(U_{t-1}, \mathbf{1}_d)$.
3. Generate V_t from a the distribution $C_{\phi_{\tilde{\theta}}}$ with $\tilde{\theta} = \frac{\theta}{1+d\theta}$.
4. Set $U_{it} = \left[\left(\sum_{i=1}^d U_{it-1}^{-\theta} - d + 1\right)\left(V_{it}^{-\tilde{\theta}} - 1\right) + 1\right]^{-1/\theta}$.
5. Set $X_{it} = F_i^{-1}(U_{it})$ for all $i = 1, \ldots, d$ and all $t = 1, 2, \ldots$.

To generate a d-dimensional random vector Y from a Clayton copula C_{ϕ_θ} we simulate independently S from a $Gamma(1/\theta, 1)$ and $E_1, \ldots, E_d$ from a $Exp(1)$, then we set $Y_i = (1 + E_i/S)^{-\theta}$ for $i = 1, \ldots, d$.

B Conditional Student copula

Let $X^T = (X_1, X_2)$ be a $d = (d_1 + d_2)$-dimensional random vector which follows a multivariate Student distribution $T_d(x; \nu, \mu, R)$, where ν is the degree of freedom, $\mu = (\mu_1, \mu_2)$ is a $(d_1 + d_2)$-dimensional real vector which is the location vector and

$$R = \begin{bmatrix} R_{X_1} & R_{X_1X_2} \\ R_{X_2X_1} & R_{X_2} \end{bmatrix}$$

is the correlation block matrix. Let $\mathbb{I}_d$ and 0_d be respectively the $d-$dimensional identity matrix and null matrix. Using Schur method we can write $R = A \times M \times B$ where

$$A = \begin{bmatrix} \mathbb{I}_{d_1} & 0_{d_1 \times d_2} \\ R_{X_2X_1}R_{X_1}^{-1} & \mathbb{I}_{d_2} \end{bmatrix}$$

$$M = \begin{bmatrix} R_{X_1} & 0_{d_1 \times d_2} \\ 0_{d_2 \times d_1} & R_{X_2} - R_{X_2X_1}R_{X_1}^{-1}R_{X_1X_2} \end{bmatrix}$$

$$B = \begin{bmatrix} \mathbb{I}_{d_1} & R_{X_1}^{-1} R_{X_1 X_2} \\ 0_{d_2 \times d_1} & \mathbb{I}_{d_2} \end{bmatrix}.$$

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

$$R^{-1} = \begin{bmatrix} R_{X_1}^{-1} + \tilde{B}\tilde{M}^{-1}\tilde{A} & -\tilde{B}\tilde{M}^{-1} \\ -\tilde{M}^{-1}\tilde{A} & \tilde{M}^{-1} \end{bmatrix} \tag{8}$$

where $\tilde{A} = R_{X_2 X_1} R_{X_1}^{-1}$, $\tilde{M} = R_{X_2} - R_{X_2 X_1} R_{X_1}^{-1} R_{X_1 X_2}$ and $\tilde{B} = R_{X_1}^{-1} R_{X_1 X_2}$. Using (8) have the decomposition

$$X^T R^{-1} X = (X_2 - \tilde{A}X_1)^T \tilde{M}^{-1} (X_2 - \tilde{A}X_1) + X_1^T R_{X_1}^{-1} X_1. \tag{9}$$

The density function of the above multivariate Student distribution is defined as

$$t_d(x; \nu, \mu, R) = \frac{\Gamma(\frac{\nu}{2} + \frac{d}{2})}{|R_{X_1}|^{1/2} \Gamma(\frac{\nu}{2}) (\pi\nu)^{-d/2}} \left(1 + \frac{(X-\mu)^T R^{-1} (X-\mu)}{\nu}\right)^{-(\frac{\nu}{2} + \frac{d}{2})} \tag{10}$$

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 we have that X_1 follows a d_1-dimensional multivariate distribution with parameters ν, μ_1 and R_{X_1}. Fix $\mu = 0$, then using (9) and some algebraic manipulation it is possible to show that the conditional distribution of X_2 given $X_1 = x_1$ is a d_2-dimensional Student distribution with degree of freedom $\tilde{\nu} = \nu + d_1$, location parameter $\tilde{\mu} = \tilde{A}x_1$ and scale matrix $\frac{\lambda}{\tilde{\nu}}\tilde{M}$, where $\lambda = \nu + x_1^T R_{X_1}^{-1} x_1$ that is

$$t_{d_2}(x_2; \tilde{\nu}, \tilde{\mu}, \frac{\lambda}{\tilde{\nu}}\tilde{M}) = \frac{t_d(x; \nu, 0_d, R)}{t_{d_1}(x_1; \nu, 0_{d_1}, R_{X_1})}.$$

References

Andersen, T., Bollerselv, T., and Diebold, F. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, Vol 89(4):701–720.

Andersen, T., Bollerselv, T., Diebold, F., and Labys, P. (2001). The distribution of realzed exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Beare, B. (2010). Copulas and temporal denpendence. *Econometrica*, 78(1):395–410.

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. *Journal of Econometrics*, 60(1):48–57.

Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric model time series models. *Journal of Econometrics*, 130:307–335.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009). The advent of copula in finance. *The European Journal of Finance*, 15:609–618.

Martens, M. and van Dijk, D. (2006). Measuring volatility with the realized range. *Journal of Econometrics*, 13:181–207.

Nelsen, R. B. (1999). *An introduction to copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis*, 3(1).

Rémillard, B. (2013). *Statistical Methods For Financial Engineering*. Chapman & Hall.

Sokolinskiy, O. and Van Dijk, D. (2011). Forecasting volatility with copula-based time series models. *Tinbergen Institute Discussion Paper*.

Soustra, F. (2006). Pricing of synthetic CDO tranches, analysis of base correlations and an introduction to dynamic copulas. Master's thesis, HEC Montréal.

Zhang, L., Mykland, P., and At-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of The American Statistical Association*, 100(472):1394–1414.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52.