International Journal of General Systems

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ggen20

Improving the accuracy of predictions in multivariate time series using dynamic vine copulas

Ayyub Sheikhi, Luciana Dalla Valle & Radko Mesiar

INTERNATIONAL JOURNAL OF GENERAL SYSTEMS
https://doi.org/10.1080/03081079.2024.2350542

Improving the accuracy of predictions in multivariate time series using dynamic vine copulas

Ayyub Sheikhi[a], Luciana Dalla Valle[b] and Radko Mesiar[c,d]

[a]Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran; [b]School of Engineering, Computing and Mathematics, Plymouth University, Plymouth, UK; [c]Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, STU Bratislava, Slovakia ; [d]UTIA AV ČR, Prague 8, Czech Republic

ABSTRACT

In this work, we deal with non-stationary multivariate time series, proposing a method which uses copulas to produce more accurate forecasting. The idea is to apply a copula-based approach to identify change points and then split the time series into consecutive segments based on these change points. In each segment, we define the best-fitting copula family and forecast values of the time series of each segment using the corresponding fitting copula. We apply our model to a financial data set to test the predictive power of our approach. A simulation study is also presented for a detailed illustration and assessment of our proposed methodology. Based on the results of numerical analysis, we observed that our proposed approach will help us to improve the accuracy of forecasting in comparison with other existing methods such as traditional time series forecasting as well as neural network forecasting.

ARTICLE HISTORY
Received 19 June 2023
Accepted 27 April 2024

KEYWORDS
Vine copula; dynamic copula; time series; change point

1. Introduction

With the recent developments in information and communication technologies, the currently available data have linked together in a complex manner at various levels, and therefore, treating such dynamic complex systems has gained increased urgency. In this context, multivariate time series analysis has become more prominent due to well-established applications in various domains including environment, health, and economy (Kirchgässner, Wolters, and Hassler 2012; Liu et al. 2015; Zhang, Zhang, and Khelifi 2018). Although many popular approaches such as autoregressive moving average (ARMA), multivariate generalized autoregressive conditional heteroscedasticity (GARCH), and their mixture (ARMA-GARCH) have been applied frequently in time series modelling (Mittnik et al. 2007), the majority of these models are constructed based on the stationarity assumption of the underlying time series, which is very restrictive in real-world data sets (see, e.g. Kleibergen and Dijk 1993; Mikosch and Stărică 2004). Therefore, analysis and forecasting in such non-stationary multivariate time series is a challenging task because of their long/short-term patterns, and hence, adopting time-varying or dynamic approaches has been proposed in

CONTACT Ayyub Sheikhi sheikhy.a@uk.ac.ir

the literature. For more details about using dynamic time warping in multivariate time series classification, we refer to Seto, Zhang, and Zhou (2015) and Li et al. (2019); for exploring flexible copula models with dynamic dependence, see Krupskii and Joe (2020); and finally, for causality structures analysis of time-varying behavior of multivariate time series, see Carlos-Sandberg and Clack (2021).

In order to investigate the dynamic behavior of multivariate time series, determining the relationship structure between variables is an initial problem for both researchers and practitioners. Many of the traditional methods in the literature assume normality and hence linear relationships among variables because of its simplicity in calculation and understanding. However, despite the popularity, it is well known that this assumption is valid only within the Gaussian framework (Dou, Haiyan, and Aivaliotis 2019; Embrechts, McNeil, and Straumann 2002; Patton 2009). For instance, Granger (2003) first reported that the classical linear multivariate modeling (based on the Gaussian distribution assumption) clearly fails to explain the nature of the relationship between time series variables, especially in economic and financial time series (Chen and Fan 2006). Going beyond the linear relationship between time series arises from the fact that the nature of the most multivariate time series is indeed complex, nonlinear, and non-normally distributed, so methodologies modeling nonlinear relationships and focusing on dependence instead of correlation seem to be the most favorable direction, see e.g. (Liu et al. 2022).

In this context, due to the flexible properties of copulas (Section 2 below) in characterizing linear and nonlinear relationships between variables, they have been widely used to modeling multivariate data. Here, the traditional multivariate elliptical copula approach has been widely used in the literature, among others, by Cherubini, Luciano, and Vecchiato (2004), Renard and Lang (2007), Danaher and Smith (2011), Hofert et al. (2018) and Sheikhi, Fereshteh, and Radko (2022), but this type of copula lacks flexibility in modeling complex high-dimensional dependence, especially in multivariate time series analysis. In line with the introduction of vine copulas by Harry (1996) as well as Bedford and Cooke (2001), Aas et al. (2009) have proposed "vine copulas" using pair copulas, namely pair-copula construction (PCC), to achieve more flexibility. See also (Fischer et al. 2009), for more information about the better performance of vine copulas as opposed to the traditional multivariate copulas in modeling the dependence of high-dimensional data. The recent book of Czado (2019) provides a full illustration of vine copulas and their applications.

In this work, we apply vine copulas to time-varying multivariate time series analysis. The use of copula models with time-varying parameters in time series analysis was first introduced by Patton (2006), and then, several attempts have been made in the literature to specify time-varying copulas. Stoeber and Czado (2012) investigated modeling time-varying dependencies with the aid of vine copulas. Acar, Czado, and Lysy (2019) presented flexible dynamic vine copula models for multivariate time series data. Zhou and Ji (2021) adopted regular vine copulas to model mortality dependence. Almeida, Czado, and Manner (2016) used dynamic D-vine copulas to model high-dimensional time-varying dependence. Among many others, see (Candido and Valls 2019; Kreuzer and Czado 2019; Pircalabu and Jung 2017; Smith 2015) for the adoption of D- and C-vines in this field. We refer to Manner and Reznikova (2012) for a survey review of time-varying copulas, Mejdoub and Ghorbel (2018), Oh and Patton (2018), Krupskii and Joe (2020) and Nevrla (2020) for applications of time-varying vine copulas in finance; Feng et al. (2022) for

Li and Li (2021) and Schepsmeier and Czado (2016) for surveys in the context of healthcare time series; Erhardt, Czado, and Schepsmeier (2015) and Ansell and Dalla Valle (2021) for the use of vine copulas with environmental time series data and finally, Ren, Ma, and Han (2022), Feng et al. (2022) and Feng et al. (2020) for using machine learning methods in time series forecasting.

It is known that stationary time series data are easier to forecast and model because it exhibits consistent patterns and behavior over time; on the other hand, non-stationary time series data may exhibit trends, seasonality, non-constant mean, variance, and dependencies over time, which is making it more difficult to forecast and model accurately. One of the most popular approaches to handle time-varying time series data is to compute "change points", see, e.g. (Hofert et al. 2018; Xiong and Cribben 2022). Employing copulas to capture variable dependence, we use the change-point approach to improve the accuracy of predictions in multivariate time series. Our contribution can be summarized in the following points:

- exploiting the dependence structure between variables to split the data into segments based on their relationship;
- improving the accuracy of forecasts compared to other existing methods;
- increase efficiency, especially with big data, by using one part of the time series data, instead of all data, to calculate predictions.

The rest of the article is organized as follows. We review background and preliminaries of the work in the next Section. We illustrate the application of our approach to a real data set as well as simulation studies in Section 3. Finally, Section 4 contains some concluding remarks.

2. Vine copulas

Inspired by Sklar's idea (Sklar 1959), if the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ follows the joint multivariate distribution function $H_{X_1,X_2,\ldots X_d} : \mathbb{R}^d \to [0,1]$ and $F_j : \mathbb{R} \to [0,1]$, $j = 1, 2, \ldots, d$, are the related marginal distribution functions of X_j, $j = 1, 2, \ldots, d$, there exists a grounded, uniformly marginal and increasing function

$$C : [0,1]^d \to [0,1]\,,$$

such that

$$H_{X_1,X_2,\ldots X_d}(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)). \tag{1}$$

If the joint distribution function $H_{X_1,X_2,..,X_d}$ is absolutely continuous, then from (1), one can obtain the joint density f of $(X_1, .., X_d)$ as

$$f\left(x_1, \ldots, x_d\right) = c\left(F_1\left(x_1\right), \ldots, F_d\left(x_d\right)\right) \prod_{j=1}^{d} f_k\left(x_k,\right) \tag{2}$$

where c is the density of the copula C and $f_j, j = 1, .., d$, are the respective densities of the random variables $X_1, \ldots, X_d$.

There are three main types of vine copulas, namely, D-vine, C-vine and R-vine structures. For the D-vine, each tree is a path that goes through all the nodes. For the C-vine, for every layer of the tree there is a central node. The R-vine includes the previous structures, and also all possible additional structures of vines. A general form of these three structures can be sequentially obtained as follows. It is known that the joint density function of the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ can be represented as a product of conditional densities via Bedford and Cooke (2002) as follows:

$$f\left(x_1, \ldots, x_d\right) = f_d\left(x_d\right) \cdot f_{d-1 \mid d}\left(x_{d-1} \mid x_d\right) \cdot \ldots \cdot f_{1 \mid 2 \ldots d}\left(x_1 \mid x_2, \ldots x_d\right). \tag{3}$$

Again, by the Sklar's theorem, the conditional density of $X_{d-1} \mid X_d$ can be easily written as follows:

$$f_{d-1 \mid d}\left(x_{d-1} \mid x_d\right) = c_{d-1,d}\left(F_{d-1}\left(x_{d-1}\right), F_d\left(x_d\right); \boldsymbol{\theta}_{d-1,d}\right) \cdot f_{d-1}\left(x_{d-1}\right), \tag{4}$$

where $c_{d-1,d}$ is a bivariate copula, with parameter vector $\boldsymbol{\theta}_{d-1,d}$. More generally, for a generic element X_j of the vector $\boldsymbol{X}$, a general form of of Equation (4), can be represented as follows:

$$f_{X_j \mid \boldsymbol{V}}\left(x_j \mid \boldsymbol{v}\right) = c_{X_j,V_\ell;\boldsymbol{V}_{-\ell}}\left(F_{X_j \mid \boldsymbol{V}_{-\ell}}\left(x_j \mid \boldsymbol{v}_{-\ell}\right), F_{V_\ell \mid \boldsymbol{V}_{-\ell}}\left(v_\ell \mid \boldsymbol{v}_{-\ell}\right); \boldsymbol{\theta}_{X_j,V_\ell;\boldsymbol{V}_{-\ell}}\right) \cdot f_{X_j \mid \boldsymbol{V}_{-\ell}}\left(x_j \mid \boldsymbol{v}_{-\ell}\right)$$

where $\boldsymbol{V}$ is the conditioning vector, V_ℓ is a generic component of $\boldsymbol{V}$, $\boldsymbol{V}_{-\ell}$ is the vector $\boldsymbol{V}$ without the component V_ℓ. Also, $F_{X_j \mid \boldsymbol{V}_{-\ell}}(\cdot \mid \cdot)$ is the conditional distribution of X_j given $\boldsymbol{v}_{-\ell}$ and $c_{X_j,V_\ell;\boldsymbol{V}_{-\ell}}(\cdot,\cdot)$ is the corresponding conditional bivariate copula density with parameter $\boldsymbol{\theta}_{X_j,V_\ell;\boldsymbol{V}_{-\ell}}$ (Ansell and Dalla Valle 2021). Hence, implementing such conditionals in (3) yields a pair constructed vine copula, which, based on the arrangement of the conditionals (trees), leads to those three R, D, and C-vines.

For instance, in 3-dimension, the joint density can be written as the product of the marginal density functions as well as pair-copula densities as follows:

$$\begin{aligned} f_{X_1,X_2,X_3}(x_1,x_2,x_3) = {} & f_1(x_1)f_2(x_2)f_3(x_3)c_{12}(F_1(x_1),F_2(x_2)) \\ & \times c_{23}(F_2(x_2)F_3(x_3))c_{13;2}\left(F_1(x_1|x_2),F_2(x_3|x_2);x_2\right). \end{aligned}$$

In this work, we consider the 3-dimensional vine copula-desribed above, as it has a unique form regardless of R-vine, D-vine or C-vine structures. Czado (2019) . In this work, we use the vine copula theory to address non-stationary multivariate time series by exploring their change points.

3. The proposed approach

Time-varying time series data present various challenges, such as seasonality, non-stationarity, time lags, outliers and heteroscedasticity, that can make analysis and prediction particularly complex. Therefore, the main aim of this section is to provide advanced modeling techniques to effectively capture the underlying patterns and relationships in the data (Carlos-Sandberg and Clack 2021). For this purpose, we identify change points and divide the whole time series into segments to have a more accurate analysis and prediction in the segments. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a stretch of a time series. The null hypothesis of stationarity for copulas can be stated as follows:

$$\mathcal{H}_0: \textit{ There exists a copula } C \textit{ such that } \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n \textit{ are associated with } C,$$

where C is the copula associated with joint distribution function H in Equation (1). The hypothesis $\mathcal{H}_0$ is the copula-based test of the so-called tests for change-point detection (Aue and Lajos 2013). Let us assume that pseudo-observations $\boldsymbol{U}_k^{k:l}, \ldots, \boldsymbol{U}_l^{k:l}$ of the random vectors $\boldsymbol{X}_k, \ldots, \boldsymbol{X}_l$, are defined by

$$\boldsymbol{U}_i^{k:l} = \left(F_{k:l,1}(X_{i1}), \ldots, F_{k:l,d}(X_{id})\right) \frac{l-k+1}{l-k+2}, \quad i \in \{k, \ldots, l\},$$

where $F_{k:l,j}$ is the empirical cdf of $X_{kj}, \ldots, X_{lj}$. Then, following (Hofert et al. 2018)), the test statistic for change-point detection for copulas is given by

$$S_n^C = \max_{1 \le k \le n-1} \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{D}_n^C\left(k/n, \boldsymbol{U}_i^{1:n}\right)\right)^2, \tag{5}$$

where

$$\mathbb{D}_n^C(t, \boldsymbol{u}) = \sqrt{n}\lambda_n(0,t)\lambda_n(t,1)\left(C_{1:\lfloor nt \rfloor}(\boldsymbol{u}) - C_{(\lfloor nt \rfloor+1):n}(\boldsymbol{u})\right), \quad (t, \boldsymbol{u}) \in [0,1]^{d+1},$$

with, $\lambda_n(t,t') = (\lfloor nt \rfloor - \lfloor nt' \rfloor)/n$ for $0 \le t \le t' \le 1$ in which $\lfloor . \rfloor$ denotes the floor function, and for any $1 \le k \le l \le n$

$$C_{k:l}(\boldsymbol{u}) = \frac{1}{l-k+1} \sum_{i=k}^{l} 1\left(\boldsymbol{U}_i^{k:l} \le \boldsymbol{u}\right), \quad \boldsymbol{u} \in [0,1]^d$$

is the empirical copula of $\boldsymbol{X}_k, \ldots, \boldsymbol{X}_l$, with the convention that $C_{k:l} = 0$ if $l < k$. In the next section, we use the test statistic (5) to assess the stationarity of a real-world financial multivariate time series.

Our approach follows copula-based time series forecasting and it is illustrated as follows. First, using the test statistic (5), we specify the copula-based change point(s) and divide the data into segments. Then, in each segment, based on the estimated copulas we calculate predictions. More precisely, for each marginal time series variable, we identify the best-fitted ARMA, GARCH or ARMA-GARCH model using the following equations:

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t,$$

$$\varepsilon_t = \sqrt{\sigma_t}\eta_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 \tag{6}$$

where y_t is the marginal time series, a_i, b_i, α_0, α_i and β_i are model parameters, and η_t is a sequence of IID random variables, with mean zero and variance one (Mittnik et al. 2007). We then extract the residuals of the fitted marginal ARMA-GARCH model. The estimated vine copula of transformed residuals will be the model to simulate new realizations from.

Finally, we calculate the predicted values for each simulation, using the inverse cdf and the relevant fitted marginal models (see, e.g. Ansell and Dalla Valle 2021). Algorithm 1 depicts the pseudo code of our approach.

Algorithm 1: Forecasting using change-point detection

Data: $X_1, X_2, ..., X_n$ as a stretch of a time series.
Result: Change point(s) and forecasted values of time series

```
Initialization : Number of simulations = n_s
Finding Change point(s) θ_1 :
    Calculate the test statistic: S_n^C = max_{1≤k≤n−1} (1/n) Σ_{i=1}^n (D_n^C (k/n, U_i^{1:n}))^2
    Determine the change point(s) θ_1
    Split the time series into segments corresponding to the change points θ_1
Forecasting θ_2 :
    Fit ARMA-GARCH models to each marginal:
  y_t = Σ_{i=1}^p a_i y_{t−i} + Σ_{i=1}^q b_i ε_{t−i} + ε_t,   ε_t = sqrt(σ_t) η_t,
  σ_t^2 = α_0 + Σ_{i=1}^r α_i ε_{t−i}^2 + Σ_{i=1}^s β_i σ_{t−i}^2 and extract their residuals η_t
    Find the best-fitted vine copula for the transformed residuals
    For each forecasting point, generate n_s observation form the fitted vine copula
    Calculate forecasts θ_2 using the simulated data and the inverse cdf of the
  marginals
```

Where: $S_n^C = \max_{1\le k\le n-1} \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{D}_n^C\left(k/n, \boldsymbol{U}_i^{1:n}\right)\right)^2$; $y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t, \quad \varepsilon_t = \sqrt{\sigma_t}\eta_t$; $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2$.

4. Numerical results

The following two subsections describe the application of the methodology to a real data and simulation studies, respectively.

4.1. Real data

The time series that we consider in this work are the daily log-return data for the period 1996–2000 of three well known stocks: Intel, Microsoft, and General Electric, see, e.g. (McNeil, Frey, and Embrechts 2015). We adopt the abbreviations INTC, MSFT and GE for Intel, Microsoft, and General Electric, respectively.

For stationarity checking, we follow the approach of Hofert et al. (2018). Applying the test statistic (5), we deduce that the time series is non-stationary and there is a change-point at the date of 1999-05-25. Hence, we split the data into two segments 1 and 2, including the segments before and after the change point, respectively. We make prediction/forecasting of the last 50 points of time series via Algorithm 1 and the difference between the predicted values and the true values will be our criterion to assess the accuracy of the prediction. For this prediction, we set two scenarios. The first scenario makes a prediction in second segment using the best-fitting copula of the segment; while the second scenario carries out these predictions by finding the copula of the entire data. In order to illustrate our approach, we first consider segment 2. We fit ARIMA, GARCH, or ARIMA-GARCH models to the three marginals in this segment. We find that the best models in segment 2 are ARIMA(2,0,2)–GARCH(1,4), ARIMA(1,0,2)–GARCH(1,1) for INTC, MSFT, and GE. On the other hand, the best time series models for these stocks in the whole

Table 1. Estimated vine copulas with relevant copula families and Kendall's τ values using the full series (top panel), segment 1 (central panel) and segment 2 (bottom panel).

Variables	copula	Estimated copula	Estimated τ
		Full series	
INTC and MSFT	c_{12}	*BB*1	0.42
MSFT and GE	c_{23}	*BB*7	0.229
INTC and GE; MSFT	$c_{13;2}$	*t*	0.11
		Segment 1	
INTC and MSFT	c_{12}	*t*	0.43
MSFT and GE	c_{23}	*BB*1	0.31
INTC and GE; MSFT	$c_{13;2}$	*t*	0.11
		Segment 2	
INTC and MSFT	c_{12}	*Gaussian*	0.41
MSFT and GE	c_{23}	*BB*8	0.21
INTC and GE; MSFT	$c_{13;2}$	*Gumbel*	0.08

dataset are estimated as ARIMA(0,0,1)–GARCH(1,1), ARIMA(1,0,0)–GARCH(1,1), and ARIMA(0,0,2)–GARCH(2,1), respectively for INTC, MSFT, and GE. Via Algorithm 1, after fitting the time series models for the marginals in segment 2 and the whole time series, we estimate the vine copula of the transformed marginal residuals via the `VineCopula` package (Schepsmeier et al. 2015), and then we simulate $M = 20$ realizations from the vine copula. Hence, we calculate the predicted values for each simulation, using the inverse cdf and the relevant fitted marginal models (see, e.g. Ansell and Dalla Valle 2021). Denoting the variables INTC, MSFT, and GE as 1, 2, and 3, respectively, the results of vine copulas estimation are reported in Table 1. This Table Shows the best superior copula between variables in each segment as well.

In order to compare the performances of the different scenarios, we compute in-sample predictions for the last 50 observations of the second segment, see Figure 1. As an alternative method, we carry out a neural network forecasting as well. In both forecasting methods, the criteria for assessing our approach are the root-mean-square error (RMSE) as well as the the Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery 2007; Matheson and Winkler 1976).

The results are summarized in Table 2. As this Table reveals, the vine copula and neural network methods yield similar results. However, the vine copula forecasting is superior in terms of both RMSE and CRPS, especially after identifying the change point.

4.2. Simulation study

In order to more deeply investigate the results of our approach, in this subsection, we describe the results of the implementation of a simulation study. We consider the time series of two random variables with 3 000 data points each, that are split, by construction, in three segments. We carry out this simulation study by assuming three types of copula families: Gaussian, Clayton, and Gumbel while the marginals follow a standard normal distribution. We assume a different copula parameter in each of the three segments and the difference between true simulated data and their corresponding forecasted values will be our criterion for the accuracy of the proposed approach. More details about our algorithm are coming in the sequel.

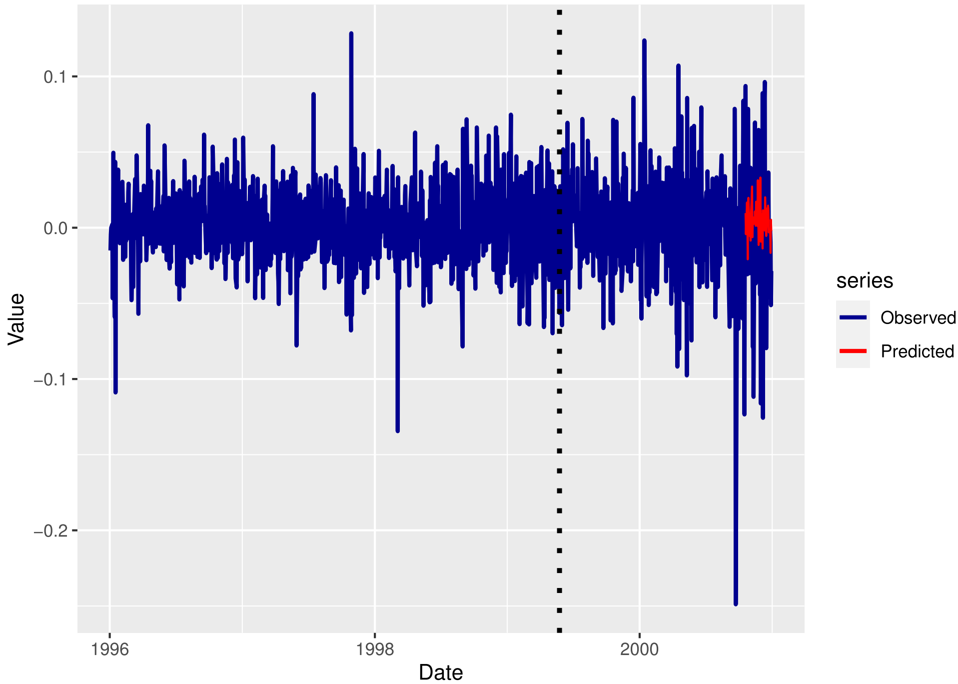

Figure 1. Observed time series (in blue) and the corresponding predicted values (in red) using the fitted copula of the entire series of the INTC stocks. The dotted black vertical line shows the change point.

Table 2. Forecasting the last 50 data points using two methods: vine copulas with relevant copula families and Kendall's τ values and neural networks.

Method	RMSE for INTC, MSFT, GE	CRPS for INTC, MSFT, and GE
	Forecasting using Full series	
Vine copula approach	0.053, 0.049, 0.026	0.011, 0.009, 0.052
Neural network time series forecasting	0.035, 0.048, 0.024	0.013, 0.009, 0.059
	Forecasting using Segment 2	
Vine copula approach	0.051, 0.048, 0.022	0.010, 0.009, 0.052
Neural network time series forecasting	0.052, 0.036, 0.023	0.011, 0.011, 0.065

Note: We used the full series (top panel), and segment 2 (bottom panel).

Case 1: Gaussian copulas (positive correlations)

For the first segment, we generate $n = 1000$ bivariate data (x_1, x_2) from a Gaussian copula where the considered correlation value is equal to $\rho = 0.35$, that we denote by $Gaussian(0.35)$, with standard normally distributed marginals. Subsequently, for the second segment we generate $n = 1000$ bivariate data from a $Gaussian(0.65)$, with standard normal marginals, while for the third segment, we generate $n = 1000$ bivariate data from a $Gaussian(0.95)$, with standard normal marginals. In order to assess our approach, we consider two scenarios for making predictions. The first scenario (scenario A) uses different copulas for each segment separately; while the second scenario (scenario B) uses a single copula of the entire data. For both scenarios, we fit a mean–variance model for each of the marginals x_1 and x_2 and we obtained an ARIMA(0,0,0) for both marginals.

Table 3. RMSE and CRPS criteria for estimated copulas and their estimated parameters using the data of segments 1, 2, and 3 and entire data.

Data	Estimated copula	TS RMSE1	NNs RMSE2	Proposal RMSE3	TS CRPS1	NNs CRPS2	Proposal CRPS3
Case 1: $Gaussian(0.35)$, $Gaussian(0.65)$, $Gaussian(0.95)$							
Segment 3	$Gaussian(0.95)$	.99	.99	.97	.75	.76	.73
Entire data	$t(0.66, 4.46)$	1.01	.99	.98	.71	.76	.70
Case 2: $Gaussian(-0.95)$, $Gaussian(0.05)$, $Gaussian(0.95)$							
Segment 3	$Gaussian(0.95)$	1.07	1.07	1.04	.85	.85	.82
Entire data	$t(0.04, 2)$	1.14	1.13	1.11	.83	.85	.81
Case 3: $Clayton(0.1)$, $Clayton(1)$, $Clayton(5)$							
Segment 3	$bb7(4.84, 0.15)$	.93	.92	.90	.75	.75	.68
Entire data	$bb6(1.2, 1.4)$	.96	.96	.96	.75	.75	.73
Case 4: $Gumbel(1.1)$, $Gumbel(2)$, $Gumbel(5)$							
Segment 3	$bb7(5.5, 1.9)$	.88	.87	.85	.73	.73	.72
Entire data	$t(0.63, 2.18)$	.89	.88	.87	.75	.75	.73

Note: $t(\rho, \nu)$ denotes the t-copula where ρ and ν are, respectively, the correlation coefficient and the degrees of freedom parameters.
$BB1(\theta, \delta)$, $BB6(\theta, \delta)$ and $BB7(\theta, \delta)$ stand for $BB1$, $BB6$ and $BB7$ copulas with parameters θ and δ (Harry 2014).
RMSE1 denotes the RMSE using traditional times series (TS) forecasting
RMSE2 denotes the RMSE using neural network (NN) times series forecasting
RMSE3 denotes the RMSE using the proposed copula approach.
Similar notation holds for CRPS.

Subsequently, using Algorithm 1, we identify the best-fitted copula of the transformed residuals of the two marginal models, and we simulate $M = 20$ realizations from the fitted copula. As the first part of Table 3 reveals, the estimated copulas for these three segments are $Gaussian(0.35)$, $Gaussian(0.65)$, and $Gaussian(0.95)$, respectively. Finally, we obtain the predicted values for each simulation, using the inverse cdfs and the relevant fitted marginal models for the last 100 points of each segment. More precisely, our predicted values will correspond to the points from from 2901 to 3 000 using the fitted copula of the third segment. We consider two other competitor approaches: traditional time series forecasting and neural network forecasting. Considering RMSE and CRPS, we compare predicted values and their original simulated values to check the performance of our algorithm against those two alternatives.

For scenario B, similarly to scenario A, we calculated the predictions for points 2901 to 3 000, but we use the fitted copula for the whole 3 000 observations. As seen in Table 3 the fitted copula for the entire data is a t copula with the estimated Kendall's τ value $\hat{\tau} = 0.45$. In addition, as can be seen from the table, the mean of the RMSEs of the predicted values of x_1 and x_2 are 0.99, 0.99, and 0.97,respectively, for traditional, neural network and our approach forecasting in Segment 3. These values, when we use the entire data, are 1.01, 0.99, and 0.98, which shows that vine copula forecasting yields better results compared to the other two methods. More precisely, the traditional time series and neural network forecasting perform similarly, while our proposed copula approach overcomes the other two approaches. However, results improve when we use the last segment instead of all data, as can be deduced from the CRPS criterion results. See also the first row of Figure 2 for a visualization of the copulas involved. In the first row of the figure, the left graph (a) shows the simulated data (scatter points) as well as the fitted copulas (contour plot) for segment 1. Graphs (b), (c), and (d) are representing the simulated data and their fitted copulas for segments 2, 3 and the whole data, respectively. Comparing the RMSEs and CRPSs from

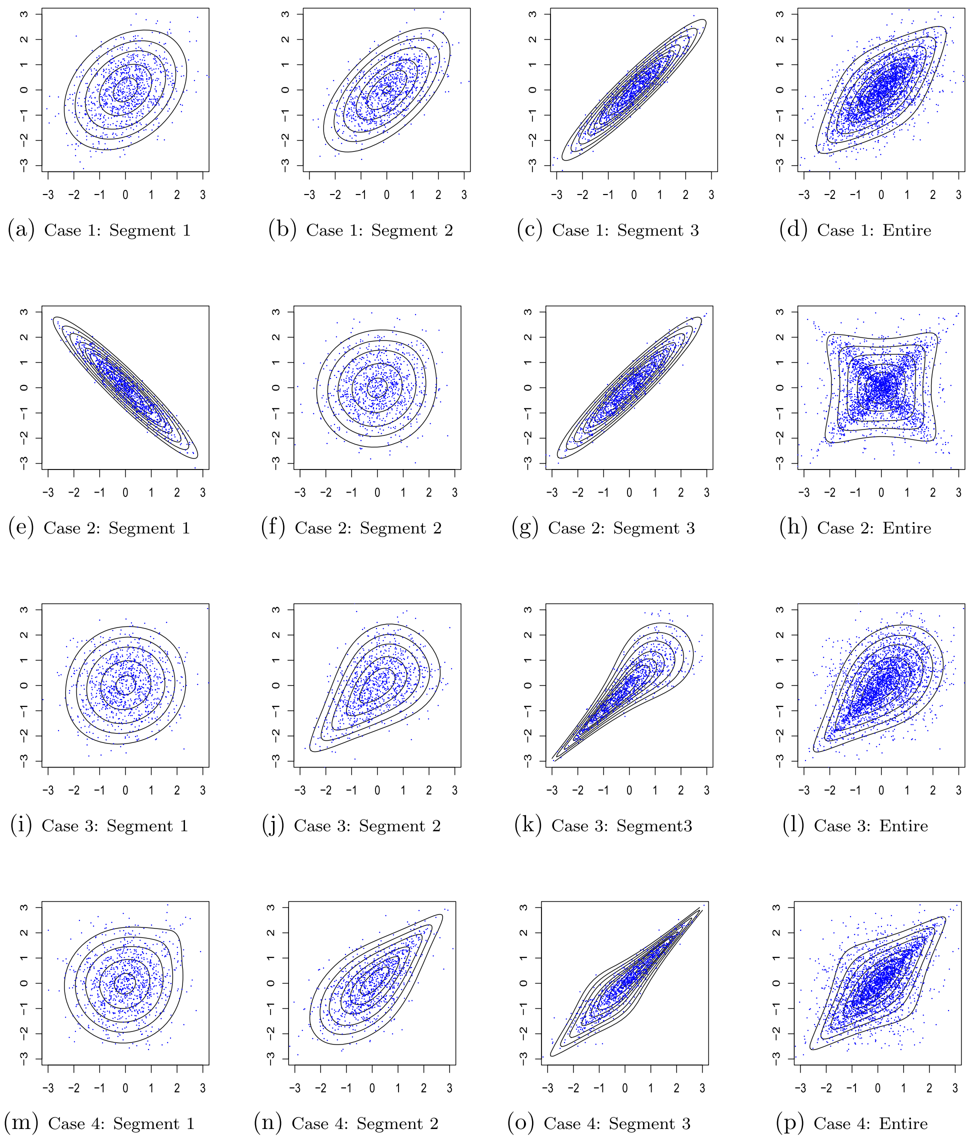

Figure 2. Scatterplot of the simulated data and the fitted copula of residuals. The first row shows case 1, the second row case 2, the third row case 3 and the fourth row case 4. (a) Case 1: Segment 1. (b) Case 1: Segment 2. (c) Case 1: Segment 3. (d) Case 1: Entire. (e) Case 2: Segment 1 (f) Case 2: Segment 2. (g) Case 2: Segment 3. (h) Case 2: Entire. (i) Case 3: Segment 1. (j) Case 3: Segment 2. (k) Case 3: Segment3. (l) Case 3: Entire. (m) Case 4: Segment 1. (n) Case 4: Segment 2. (o) Case 4: Segment 3 and (p) Case 4: Entire.

the first four rows of Table 3, we observe that considering different copulas in the third segment improves the accuracy of prediction.

Case 2: Gaussian copulas (negative and positive correlations)

In case 2, we assume *Gaussian*(-0.95), *Gaussian*(0.05), and *Gaussian*(0.95), respectively, for segments 1, 2, and 3. The results are summarized in the second panel of Table 3 and the second row of Figure 2. In the second row of the figure, the graphs (e), (f), (g), and

(h) depict the simulated data and their fitted copulas for the segments 1, 2, 3 and the whole data, respectively. Again, the traditional and neural network methods provide the similar forecasting performance, while the vine copula proposed approach yields better results, particularly with data split into segments.

Case 3: Clayton copulas

In this case, for the first segment, we generate $n = 1000$ bivariate data (x_1, x_2) from a Clayton copula with parameter equals 0.1, denoted by $Clayton(\theta = 0.1)$, with standard normally distributed marginals. Similarly, we generate the same number of points from a $Clayton(\theta = 1)$ and a $Clayton(\theta = 5)$, respectively, with standard normal marginals. In order to assess our approach, we consider two scenarios for making predictions. The estimated copula of the residuals of the marginals for the entire series is $BB6$; while for segments 1,2, and 3, they are, respectively, $BB1$, *Clayton*, and $BB7$. See the third panel of Table 3 for the estimated parameters as well as RMSEs and CRPSs. The results again indicate the superiority of the change-point approach compared to the traditional approaches. Also, as noticed previously, neural networks and traditional time series forecasts yield analogous results. The copula visualizations are displayed in the third row of Figure 2.

Case 4: Gumbel copulas

In this case, assuming standard normals marginals, we generate $n = 1000$ bivariate data from Gumbel copulas. In this regard, we consider $Gumbel(\theta = 1.1)$, $Gumbel(\theta = 2)$, and $Gumbel(\theta = 5)$, respectively, for segments 1, 2 and 3. The fitted copulas for the residuals of the fitted model for the entire data, segment 1, 2, and 3 are reported in the final row of Figure 2 as well as the fourth panel of Table 3. Comparing the RMSEs and CRPSs it is clear that using copulas for each segment, as well as using copulas as opposed to traditional time series or neural networks, improves the accuracy of the prediction.

In addition to the simulation study illustrated above, we implemented a new simulation to test the flexibility of the vine copula compared to traditional multivariate copulas. Similar to the above study, we considered the time series of three random variables with 2 000 data points each, that are split, by construction, in two segments. The first segment consist of 1000 observations from a three-dimensional Clayton copula with parameter $\theta = 2$ and standard normally distributed marginals. The second segment comprises 1000 observations from a trivariate Clayton copula with parameter $\theta = 5$ as well as standard normal marginals. Here, we only calculate forecasts for the last 100 observations using two types of copulas: a three-dimensional vine copula and a trivariate Clayton copula. We compared the results obtained using only the observations of the second segment with those obtained using the observation of all data. The results are presented in Table 4. As it is clear form the table, both the RMSE and CRPS criteria are larger with the traditional copula, while the vine copula yields better results in all cases.

Table 4. Comparing vine copula and traditional multivariate copula.

Data	Vine Copula $RMSE_v$	Traditional Copula $RMSE_t$	Vine Copula $CRPS_v$	Traditional Copula $CRPS_t$
Segment 2	1.08	2.13	0.38	0.54
Entire data	1.11	2.45	0.39	0.58

$RMSE_v$ denotes the RMSE using vine copula;
$RMSE_t$ denotes the RMSE using traditional multivariate copula.
This notation holds for CRPS as well.

5. Conclusion

In this work, we considered multivariate time series and, in order to produce more accurate predictions, we applied a copula-based technique to identify change points and split the time series into consecutive segments. In each segment, we determined the best-fitting copula and conveyed the predicted values in each segment using the fitted copula. Our results in real data analysis and simulation studies showed that our approach has a better performance than traditional methods which use a single copula family of the whole data.

The main contribution of this work is to support researchers to obtain higher accuracy for prediction/forecasting. Our approach is based on splitting the data of the time series into consecutive segments and using the information of each separate segment to model data dependence. We fitted the best vine copula in each segment and the predictions demonstrated that our algorithm shows a good performance. In summary, the proposed approach helped us to improve the accuracy of forecasting in comparison with other existing methods such as traditional time series forecasting as well as neural network forecasting. In addition, this study showed that when data are copula related, employing a part of the time series to calculate forecasts instead of using all data, especially working with big data, is not only saving time, but improves accuracy.

In addition, this method can be used for nowcasting real-time data streaming and high frequency data. Based on the results of this work, far nowcasting, it is supposed that the copula of the last segment will have a better performance compared to the copula of the entire data. So, when researchers try to carry out a nowcasting, they can use the copula of the last segment. This will have two benefits: first, more accuracy because of using the best-fitted copula, and second, less time-consuming because of using the last segment instead of all data.

To best of our knowledge, this is the first article that implements the approach we propose, and, however, it can be further extended in many fashions. Although we only considered the change-point method for splitting the time series, our approach may be useful in a "sliding window" context (see, for instance, the review work of Miodrag, Milanović, and Stamenković (2014)). In addition, although we used basic univariate marginal distributions as well as basic copula families, with more complex data structures, one may be interested in adopting mixture univariate distributions and mixture copulas.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the EPSRC grant number EP/W021986/1 and the grant VEGA 1/0036/23; Slovak Academic Information Agency [APVV-18-0052].

References

Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. 2009. "Pair-Copula Constructions of Multiple Dependence." *Insurance: Mathematics and Economics* 44 (2): 182–198.

Acar, F. Elif, Claudia Czado, and Martin Lysy. 2019. "Flexible Dynamic Vine Copula Models for Multivariate Time Series Data." *Econometrics and Statistics* 12:181–197. https://doi.org/10.1016/j.ecosta.2019.03.002.

Almeida, Carlos, Claudia Czado, and Hans Manner. 2016. "Modeling High-Dimensional Time-Varying Dependence Using Dynamic D-Vine Models." *Applied Stochastic Models in Business and Industry* 32 (5): 621–638. https://doi.org/10.1002/asmb.v32.5.

Ansell, Lauren, and Luciana Dalla Valle. 2021. "Social Media Integration of Flood Data: A Vine Copula-Based Approach." arXiv:2104.01869.

Aue, Alexander, and Horváth Lajos. 2013. "Structural Breaks in Time Series." *Journal of Time Series Analysis* 34 (1): 1–16. https://doi.org/10.1111/jtsa.2012.34.issue-1.

Bedford, Tim, and Roger Cooke. 2001. "Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines." *Annals of Mathematics and Artificial Intelligence* 32 (1): 245–268. https://doi.org/10.1023/A:1016725902970.

Bedford, Tim, and Roger Cooke. 2002. "Vines–A New Graphical Model for Dependent Random Variables." *The Annals of Statistics* 30 (4): 1031–1068. https://doi.org/10.1214/aos/1031689016.

Candido, Osvaldo, and Pedro Valls. 2019. "Dynamic D-Vine Copula Model with Applications to Value-At-Risk (var)." *Journal of Time Series Econometrics* 11 (2): 1–34.

Carlos-Sandberg, Leo, and Christopher Clack. 2021. "Incorporation of Causality Structures to Complex Network Analysis of Time-Varying Behaviour of Multivariate Time Series." *Scientific Reports* 11 (1): 1–16. https://doi.org/10.1038/s41598-021-97741-2.

Chen, Xiaohong, and Yanqin Fan. 2006. "Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models Under Copula Misspecification." *Journal of Econometrics*135 (1-2): 125–154. https://doi.org/10.1016/j.jeconom.2005.07.027.

Cherubini, Umberto, Elisa Luciano, and Walter Vecchiato. 2004. *Copula Methods in Finance*. Chichester: John Wiley & Sons.

Czado, Claudia 2019. *Analyzing Dependent Data with Vine Copulas*. Lecture Notes in Statistics. Vol. 222. Gewerbestrasse: Springer.

Danaher, J. Peter, and Michael Smith. 2011. "Modeling Multivariate Distributions Using Copulas: Applications in Marketing." *Marketing Science* 30 (1): 4–21. https://doi.org/10.1287/mksc.1090.0491.

Dou, Yali, Liu Haiyan, and Georgios Aivaliotis. 2019. "Dynamic Dependence Modeling in Financial Time Series." arXiv:1908.05130.

Embrechts, Paul, Alexander McNeil, and Daniel Straumann. 2002. "Correlation and Dependence in Risk Management: Properties and Pitfalls." *Risk Management: Value At Risk and Beyond* 1:176–223. https://doi.org/10.1017/CBO9780511615337.

Erhardt, Tobias Michael, Claudia Czado, and Ulf Schepsmeier. 2015. "R-Vine Models for Spatial Time Series with An Application to Daily Mean Temperature." *Biometrics* 71 (2): 323–332. https://doi.org/10.1111/biom.12279.

Feng, S., M. Han, J. Zhang, T. Qiu, and W. Ren. 2020. "Learning Both Dynamic-Shared and Dynamic-Specific Patterns for Chaotic Time-Series Prediction." *IEEE Transactions On Cybernetics* 52 (6): 4115–4125. https://doi.org/10.1109/TCYB.2020.3017736.

Feng, Yi, Yunqiang Yin, Dujuan Wang, and Lalitha Dhamotharan. 2022. "A Dynamic Ensemble Selection Method for Bank Telemarketing Sales Prediction." *Journal of Business Research* 139:368–382. https://doi.org/10.1016/j.jbusres.2021.09.067.

Fischer, Matthias, Köck Christian, Schlüter Stephan, and Weigert Florian. 2009. "An Empirical Analysis of Multivariate Copula Models." *Quantitative Finance* 9 (7): 839–854. https://doi.org/10.1080/14697680802595650.

Gneiting, T., and A. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal Of The American Statistical Association* 102 (477): 359–378. https://doi.org/10.1198/016214506000001437.

Granger, Clive 2003. "Time Series Concepts for Conditional Distributions." *Oxford Bulletin of Economics and Statistics* 65 (s1): 689–701. https://doi.org/10.1046/j.0305-9049.2003.00094.x.

Harry, Joe 1996. "Families of m-Variate Distributions with Given Margins and $m(m-1)/2$ Bivariate Dependence Parameters." *Lecture Notes-Monograph Series*, 120–141.

Harry, Joe 2014. *Dependence Modeling with Copulas*. New York, NY: CRC Press.

Hofert, Marius, Kojadinovic Ivan, Mächler Martin, and Yan Jun. 2018. *Elements of Copula Modeling with R*. Gewerbestrasse: Springer.

Kirchgässner, Gebhard, Jürgen Wolters, and Uwe Hassler. 2012. *Introduction to Modern Time Series Analysis*. Springer Science & Business Media.

Kleibergen, Frank, and Herman K. Van Dijk. 1993. "Non-Stationarity in Garch Models: A Bayesian Analysis." *Journal of Applied Econometrics* 8 (S1): S41–S61. https://doi.org/10.1002/(ISSN)1099-1255.

Kreuzer, Alexander, and Claudia Czado. 2019. "Bayesian Inference for Dynamic Vine Copulas in Higher Dimensions." *arXiv preprint* arXiv:1911.00702.

Krupskii, Pavel, and Harry Joe. 2020. "Flexible Copula Models with Dynamic Dependence and Application to Financial Data." *Econometrics and Statistics* 16:148–167. https://doi.org/10.1016/j.ecosta.2020.01.005.

Li, Jie, and Ping Li. 2021. "Dynamic Copula Analysis of the Effect of Covid-19 Pandemic on Global Banking Systemic Risk." In *Intelligent Computing and Block Chain: First BenchCouncil International Federated Conferences, FICC 2020, Qingdao, China, October 30–November 3, 2020, Revised Selected Papers 1*, 449–460. Springer.

Li, Zhengxin, Fengming Zhang, Feiping Nie, Hailin Li, and Jian Wang. 2019. "Speed Up Dynamic Time Warpingof Multivariate Time Series." *Journal of Intelligent & Fuzzy Systems* 36 (3): 2593–2603. https://doi.org/10.3233/JIFS-181736.

Liu, Bo, Jianqiang Li, Cheng Chen, Wei Tan, Qiang Chen, and MengChu Zhou. 2015. "Efficient Motif Discovery for Large-Scale Time Series in Healthcare." *IEEE Transactions on Industrial Informatics* 11 (3): 583–590. https://doi.org/10.1109/TII.2015.2411226.

Liu, Jianxu, Vicente Ramos, Bing Yang, Mengjiao Wang, and Songsak Sriboonchitta. 2022. "Analysing the Dynamic Co-Movement Between Tourism and Expected Economic Growth Considering Extreme Events." *Tourism Economics* 30 (1): 3–26.

Manner, Hans, and Olga Reznikova. 2012. "A Survey on Time-Varying Copulas: Specification, Simulations, and Application." *Econometric Reviews* 31 (6): 654–687. https://doi.org/10.1080/07474938.2011.608042.

Matheson, J., and R. Winkler. 1976. "Scoring Rules for Continuous Probability Distributions." *Management Science* 22 (10): 1087–1096. https://doi.org/10.1287/mnsc.22.10.1087.

McNeil, Alexander, Rüdiger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition*. Princeton university press.

Mejdoub, Hanène, and Ahmed Ghorbel. 2018. "Conditional Dependence Between Oil Price and Stock Prices of Renewable Energy: A Vine Copula Approach." *Economic and Political Studies* 6 (2): 176–193. https://doi.org/10.1080/20954816.2018.1463600.

Mikosch, Thomas, and Cătălin Stărică. 2004. "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the Igarch Effects." *Review of Economics and Statistics* 86 (1): 378–390. https://doi.org/10.1162/003465304323023886.

Miodrag, Lovrić, Marina Milanović, and Milan Stamenković. 2014. "Algoritmic Methods for Segmentation of Time Series: An Overview." *Journal of Contemporary Economic and Business Issues*1 (1): 31–53.

Mittnik, Stefan, Frank J. Fabozzi, Sergio M. Focardi, Svetlozar T. Rachev, and Teo Jašić. 2007. *Financial Econometrics: From Basics to Advanced Modeling Techniques*. John Wiley & Sons.

Nevrla, Matěj 2020. "Systemic Risk in European Financial and Energy Sectors: Dynamic Factor Copula Approach." *Economic Systems* 44 (4): Article 100820. https://doi.org/10.1016/j.ecosys.2020.100820.

Oh, Dong Hwan, and Andrew J. Patton. 2018. "Time-Varying Systemic Risk: Evidence From a Dynamic Copula Model of Cds Spreads." *Journal of Business & Economic Statistics* 36 (2): 181–195. https://doi.org/10.1080/07350015.2016.1177535.

Patton, J. Andrew 2006. "Modelling Asymmetric Exchange Rate Dependence." *International Economic Review* 47 (2): 527–556. https://doi.org/10.1111/iere.2006.47.issue-2.

Patton, J. Andrew 2009. "Copula–Based Models for Financial Time Series." In *Handbook of Financial Time Series*, 767–785. Springer.

Pircalabu, Anca, and Jesper Jung. 2017. "A Mixed C-Vine Copula Model for Hedging Price and Volumetric Risk in Wind Power Trading." *Quantitative Finance* 17 (10): 1583–1600. https://doi.org/10.1080/14697688.2017.1307511.

Ren, Weijie, Dewei Ma, and Min Han. 2022. "Multivariate Time Series Predictor With Parameter Optimization and Feature Selection Based on Modified Binary Salp Swarm Algorithm." *IEEE Transactions On Industrial Informatics* 19 (4): 6150–6159. https://doi.org/10.1109/TII.2022.3198465.

Renard, Benjamin, and Michel Lang. 2007. "Use of a Gaussian Copula for Multivariate Extreme Value Analysis: Some Case Studies in Hydrology." *Advances in Water Resources* 30 (4): 897–912. https://doi.org/10.1016/j.advwatres.2006.08.001.

Schepsmeier, Ulf, and Claudia Czado. 2016. "Dependence Modelling with Regular Vine Copula Models: A Case-Study for Car Crash Simulation Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 65 (3): 415–429.

Schepsmeier, Ulf, Jakob Stoeber, Eike Christian Brechmann, Benedikt Graeler, Thomas Nagler, Tobias Erhardt, Carlos Almeida, et al. 2015. "Package 'Vinecopula'." *R Package Version* 2 (5): 1–149.

Seto, Skyler, Wenyu Zhang, and Yichen Zhou. 2015. " Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition." In *2015 IEEE Symposium Series on Computational Intelligence*, 1399–1406. IEEE.

Sheikhi, Ayyub, Arad Fereshteh, and Mesiar Radko. 2022. "A Heteroscedasticity Diagnostic of a Regression Analysis with Copula Dependent Random Variables." *Brazilian Journal of Probability and Statistics* 36 (2): 408–419. https://doi.org/10.1214/22-BJPS532.

Sklar, Abe 1959. "Fonctions De Repartition An Dimensions Et Leurs Marges." *Publications de l'Institut de Statistique de l'Universit? de Paris* 8:229–231.

Smith, Michael Stanley 2015. "Copula Modelling of Dependence in Multivariate Time Series." *International Journal of Forecasting* 31 (3): 815–833. https://doi.org/10.1016/j.ijforecast.2014.04.003.

Stoeber, Jakob, and Claudia Czado. 2012. "Detecting Regime Switches in the Dependence Structure of High Dimensional Financial Data." arXiv:1202.2009.

Xiong, Xin, and Ivor Cribben. 2022. "Beyond Linear Dynamic Functional Connectivity: A Vine Copula Change Point Model." *Journal of Computational and Graphical Statistics* 32 (3): 853–872.

Zhang, Zhihua, Khelifi Zhang, and A. Khelifi. 2018. *Multivariate Time Series Analysis in Climate and Environmental Research*. Springer.

Zhou, Rui, and Min Ji. 2021. "Modelling Mortality Dependence: An Application of Dynamic Vine Copula." *Insurance: Mathematics and Economics* 99:241–255.