Research Article GU J Sci 36(2): 933-952 (2023) DOI: 10.35378/gujs.967577

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

Transformed Pair Copula Construction of Pareto Copula and Applications

Friday I. AGU[1*], Salih CELEBIOGLU[2]

[1]*Department of Statistics, University of Calabar, Calabar, Nigeria*
[2]*Department of Statistics, Eskişehir Osmangazi University, Turkey*

Highlights

- The article focuses on the construction of a pair copula.
- The article derived the transformed pair copula model and its submodels.
- The article presents the D-vine structure.
- The numerical application of the transformed pair copula models is explored.

Article Info

Received: 08 July 2021
Accepted: 04 May 2022

Keywords

Copula function
Pair copula
Regular vine
Transformed copula
Kendall's tau

Abstract

The study introduced the transformed copula models and a D-vine structure for three variables. The numerical applications of the models are evaluated using two different sets of real-life data that exhibit nearly no dependence, highly dependent, over, and under dispersed characteristics. We examined only the volatilities of the first data using the exponentiated weighted moving average (EWMA). The parameter estimates of the models were obtained based on the maximum likelihood estimation method for the bivariate copula models and the Dissmann algorithm for sequential top-down estimation for the D-vine structure. The results showed that the introduced copula models outperformed some existing copula models in terms of their fit statistics for both real-life and simulated data sets. In addition, the Gaussian copula model gave a better fit to the D-vine structure than some existing copula models and could be recommended for modeling a D-vine structure comprising of variables that are positively weak correlated and highly correlated.

1. INTRODUCTION

Over the decades, copula models have become one of the popular sophisticated statistical tools in literature known for multivariate modeling in various fields such as finance, insurance, engineering, economics, medicine, and biological sciences among others. Interest in copula models to capture multifarious dependence structures between n-dimensional random variables is increasing exponentially. The variables do not necessarily need to have the same or follow a certain distribution and thus the distinguishing and appealing properties of copula models for exploring good fit to such variables of various characteristics and their dependencies structure cannot be overemphasized. [1] and [2] gave a superb overview of copulas. Thereafter, researchers have achieved some developments in copula, such as [3] exploring the method of selecting an appropriate bivariate copula for statistical data analysis. [4] proposed a bivariate generalized exponential distribution derived from the copula. [5] proposed four types of copula models on the exponentially weighted moving average control chart when observations are from an exponential distribution. [6] explored the theoretical approach of generating copulas via analytical means and gave mathematical treatment to the proposed copula models. [7] discussed the method of constructing bivariate asymmetric copulas. [8] introduced a generalized multiple-step procedure for the full inference of the directional dependence structure. [9] conducted review research on the efficacy of the copula-based methodology and pair copula construction (PCC) and [10] adopted the copula approach to developing collective risk models for flexible dependence structure. [11] introduced a family of the Archimedean Copula via the contagent functions and derived the dependence characteristics. In the study, the Monte Carlo simulation approach is adopted to explore the dependence parameter and the dependence structure of two real-life data sets was examined based on trigonometric copulas and observed that the contagent copula performed well in the real-life data application. [12] used the goodness of fit test as a measurement

*Corresponding author, Friday I. AGU e-mail: agu.friday@unical.edu.ng

parameter to select the copula that gave the best fit to the data set. The study explored this approach on the monthly minimum and maximum barometric data and examined the dependence of the data.

Sometimes, fitting these copula models to higher-dimensional data sets from various fields may be very complicated. Hence, [13,14] introduced a possible decomposition of higher dimensional copulas to bivariate building blocks, the pair copulas, and its margins where a joint distribution can be decomposed into bivariate copulas of the original observations and their unconditional and conditional distribution functions thus making it more convenient to construct and explore multivariate dependence structure as well as exploring goodness of fit. [15] gave an excellent theoretical and applications review of pair copulas decomposition. [16] proposed a nonparametric pair copula construction for modeling complex correlation problems. [17] discussed the pair copula methods to explore complex dependence structure. [18] examined a systematic method to extend the copula to higher dimensions by a vine copula based on the pair copula construction. However, the study of the dependence structure of the vine copula depends solely on the use of an appropriate pair or bivariate copula model that may not have restrictions in their parameter space as well as difficulty in parameter estimation. Therefore, the study of the pair or bivariate copula construction of unrestricted parameter space copula models and vine structure for dependence evaluation and goodness of fit has become an area that demands the attention of researchers. Hence, these become the motivation and interest behind this study. Hence, the study seeks to introduce the pair copula model via the transformation approach using Pareto distribution as the baseline model and the D-vine structure for our variables of interest. The introduced model could be useful in economics, engineering, survival analysis, financial, and actuarial areas to explore dependence structure and goodness of fit among variables of various characteristics.

In most cases, a univariate distribution with two parameters has been proven in the literature to yield a better fit to data sets of various characteristics than the one-parameter models. Pareto distribution is one of the most attractive distributions with scale and shape parameters. It is a right-skewed and heavy-tailed type of distribution with vast applications in incomes and other financial variables. It has been greatly explored in the analysis of extreme observations. Recently, most financial time series-based data are right-skewed with a tail approaching zero so slowly that their bivariate analysis becomes interesting to researchers. Therefore, the Pareto distribution's properties are in many cases a more appropriate choice as a baseline model for the construction of a flexible bivariate copula model that could yield better fits to various time series-based data.

2. COPULA FUNCTION

Copula function C combines multivariate distribution functions with their one-dimensional marginal distribution functions. That is, it is a function with uniformly distributed marginal functions and in the case of $n=2$ links a bivariate distribution function to its univariate marginal distribution functions. It has the following properties.

1. The function is grounded for all $u=u_j\in[0,1]^n$ where $j=1,2,...,n$ such that for at least a coordinate of is 0, then $C(u)=0$.

2. The function C is non-decreasing in the sense that for every $u\in[0,1]^n$ and $v\in[0,1]^n$ such that $u_i\le u_j$ and $v_i\le v_j$ for $j=1,2,...,n$, $i=1,2,...,n$ and $i\le j$. Then, in this case $n=2$, this becomes $C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1)\ge 0$ for $(u_1,u_2),(v_1,v_2)\in[0,1]^2$.

3. For every $u,v\in[0,1]^n$, $C(u,1)=u$ and $C(1,v)=v$, where only one coordinate is u or v and all of the others are 1.

In addition, let $C(u,v)$ be a bivariate copula. For any $u \in [0,1]$ and $v \in [0,1]$, the partial derivative $\frac{\partial C(u,v)}{\partial u}$ exists for almost all $u \in [0,1]$, and for such v and u $0 \le \frac{\partial C(u,v)}{\partial v} \le 1$. Similarly, the partial derivative $\frac{\partial C(u,v)}{\partial v}$ exists for almost all $v \in [0,1]$ and for such u and v, $0 \le \frac{\partial C(u,v)}{\partial v} \le 1$. Thus, the corresponding density function for $C(u,v)$ is the $\frac{\partial C(u,v)}{\partial u \partial v}$.

2.1. Pair Copula Construction

Pair-copula construction pioneered in [13,14] is a more flexible and intuitive approach to expanding the applications of bivariate or pair copula to higher dimensional variables that make the study of dependence structure between variables more convenient by decomposing n-dimensional variables into their conditional and unconditional pair copula. For an n-dimensional variable, the joint density function can be decomposed into the conditional and unconditional components as

$$h(x_1, x_2, ..., x_n) = h(x_n) \times h\left(x_{n-1} \mid x_n\right) \times h\left(x_{n-2} \mid x_{n-1}, x_n\right) \times ... \times h\left(x_1 \mid x_2, ..., x_n\right). \quad (1)$$

2.2. Regular Vınes

For high dimensional variables, there are possible numbers of pair-copula construction. That is, for any $n \ge 2$ we can easily obtain its pair-copula for inference. The graphical representations of the pair-copula construction allow a large number of possible pair-copula constructions called the regular vines which comprise of canonical (C) vine, D-vine, and R-vine. For this study, we presented only the density function of the D-vine as

$$h(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} h(x_i) \prod_{j=1}^{n-1} \prod_{r=1}^{n-j} c_{r,r+j|r+1,...,r+j-1}\left(H(x_r \mid x_{r+1}, ..., x_{r+j-1}), H(x_{r+j} \mid x_{r+1}, ..., x_{r+j-1})\right). \quad (2)$$

Where j represents the trees in the D-vine, r represents the edges in each tree such that there is no node in the tree T_j that connects to more than two edges.

2.3. Transformed Copula

[19] proposed the synchronized transformation approach of copula and its margins. That is, for $u \in [0,1]$ and $v \in [0,1]$ for an initial copula C, the synchronized transformed copula given as

$$C^Q(u,v) = Q\left(C\left(Q^{-1}(u), Q^{-1}(v)\right)\right) = QC(u,v) \quad (3)$$

is a copula where Q is a non-decreasing function and Q^{-1} is its quantile function.

The distribution function of the univariate Pareto distribution is given as

$$F(x) = 1 - \left(\frac{\lambda}{x}\right)^{\beta}. \quad (4)$$

Where $f(x)=\dfrac{dF(x)}{dx}$, $x \ge \lambda$, and $\lambda, \beta > 0$ are scale and shape parameters respectively.

2.4. Measure of Association

For any copula C, Kendall's tau measure of association of a pair (u,v) is defined as

$$\tau = 4\int_0^1\int_0^1 C(u,v)\,dC(u,v) - 1. \tag{5}$$

It measures the strength based on the probability of concordance and discordance of pairs.

3. THE TRANSFORMED COPULA MODEL

This section presents the derived transformed copula model, pair-copula construction of the random variables using regular vines, and parameters estimation.

Let X be a non-negative continuous random variable with distribution function $F(.)$ and density function $f(.)$. Similarly, let another continuous random variable R be defined as $R = e^{-X}$ such that $\ln R = -x$. Then, the distribution function can be expressed as

$Q(r) = P\left[R \le r\right] = 1 - P\left[X < -\ln r\right],$ such that

$$Q(r) = 1 - F\left(-\ln r\right),\ r \in [0,1]. \tag{6}$$

Equation (6) is the survival function of X evaluated related to the expression $\hat{F}(-\ln r)$. Substituting Equation (4) in Equation (6), we have

$$Q(r) = \left(\frac{\lambda}{-\ln r}\right)^{\beta}, \tag{7}$$

such that $\dfrac{dQ(r)}{dr} = \dfrac{\beta\lambda^{\beta}}{r\left(-\ln r\right)^{(\beta+1)}} = q(r)$ and $Q^{-1}(u) = e^{-\left(\lambda u^{-\frac{1}{\beta}}\right)}$ for u is substituted in place of r.

Recall that $\ln r = -x \Rightarrow -\ln r = x$, set $u = r$. By substituting Equation (7) and $Q^{-1}(u)$ into Equation (3), we obtained the proposed transformed pair Pareto (TPP) copula function as

$$C^{Q}(u,v) = \left[\frac{\lambda}{-\ln\left\{C\left(e^{-\left(\lambda u^{-\frac{1}{\beta}}\right)}, e^{-\left(\lambda v^{-\frac{1}{\beta}}\right)}\right)\right\}}\right]^{\beta}, \lambda, \beta > 0. \tag{8}$$

Let us assume that $Q(r)$ is an increasing bijection transformation obtained at $-\ln r = x$. In [18], if $Q \circ \exp : (-\infty, 0) \to [0,1]$ is log-convex, then the transformed copula $C^Q(u,v)$ is a copula.

Similarly, $Q \circ \exp$ is log-convex if $\ln Q \circ \exp$ is convex. To show this, let $Q(u)$ and $Q^{-1}(u)$ be the transformed Pareto and quantile functions respectively. If we set $\ln r = x \Rightarrow r = e^x$, also from Equation (7), we can have that $Q(e^x) = \left(\frac{\lambda}{x}\right)^{\beta}$ such that

$$\eta(x) = \ln\left[Q(e^x)\right] = e^{-\left(\frac{\lambda}{x}\right)^{\beta}}, x \in (-\infty, 0) \tag{9}$$

Since $\frac{d^2\eta(x)}{dx^2} > 0$ and $\lambda, \beta > 0$. Then, the transformed copula function is a copula. Let us consider the initial copula C to be an Archimedean copula with a generator function ϕ. Then, we have that

$$\Phi(u) = \phi\left(Q^{-1}(u)\right) = \phi\left[e^{-\left(\lambda u^{-\frac{1}{\beta}}\right)}\right]. \tag{10}$$

3.1. The TPP-Clayton Copula

This section introduces the TPP-Clayton copula as the sub-model of the TPP model.

Let the Clayton copula function be expressed as

$$C(u,v,\alpha) = \left[u^{-\alpha} + v^{-\alpha} - 1\right]^{-1/\alpha}, \alpha > 0. \tag{11}$$

This copula function has a generator $\Phi(r) = \frac{r^{-\alpha} - 1}{\alpha}$. Then, the proposed TPP-Clayton copula model is given as

$$C^Q(u,v,\alpha,\lambda,\beta) = \left\{\frac{\lambda}{\alpha^{-1}\left\{-\ln\left(e^{-\left\{\lambda\left(e^{-\alpha \ln u}\right)^{-\frac{1}{\beta}}\right\}} + e^{-\left\{\lambda\left(e^{-\alpha \ln v}\right)^{-\frac{1}{\beta}}\right\}} - 1\right)\right\}}\right\}^{\beta}, \alpha, \lambda, \beta > 0 \tag{12}$$

with its generator given as $\Phi(u) = \frac{e^{-\left\{\lambda\left(e^{\ln(u^{-\alpha})}\right)^{-\frac{1}{\beta}}\right\}} - 1}{\alpha}$. The density function corresponding to Equation (12) is given as

$$c^{Q}\left(u,v,\alpha,\lambda,\beta\right)=\frac{\left(\lambda\alpha\right)^{\beta+2}\left(uv\right)^{\frac{\alpha-\beta}{\beta}}e^{-\lambda\left(u^{\frac{\alpha}{\beta}}+v^{\frac{\alpha}{\beta}}\right)}\left[\left(\beta+1\right)+\ln\left(e^{-\lambda u^{\frac{\alpha}{\beta}}}+e^{-\lambda v^{\frac{\alpha}{\beta}}}-1\right)^{\beta+1}\right]}{\beta\left\{\ln\left(e^{-\lambda u^{\frac{\alpha}{\beta}}}+e^{-\lambda v^{\frac{\alpha}{\beta}}}-1\right)^{\beta+1}\left(e^{-\lambda u^{\frac{\alpha}{\beta}}}+e^{-\lambda v^{\frac{\alpha}{\beta}}}-1\right)\right\}^{2}}. \tag{13}$$

Put Equation (12) into Equation (5), Kendall's tau function for the TPP-Clayton copula is

$$\tau=4\int_{0}^{1}\int_{0}^{1}\left\{\frac{\lambda}{\alpha^{-1}\left\{-\ln\left(e^{-\left\{\lambda\left(e^{-\alpha\ln u}\right)^{-\frac{1}{\beta}}\right\}}+e^{-\left\{\lambda\left(e^{-\alpha\ln v}\right)^{-\frac{1}{\beta}}\right\}}-1\right)\right\}}\right\}^{\beta}dudv-1. \tag{14}$$

3.1.1. Parameter estimates of the tpp-Clayton copula

This section presents the parameter estimates of the TPP-Clayton copula model.

Let u_i and v_i $\left(i=1,2,...,n\right)$ be pair of continuous random variables and $\psi=\left(\alpha,\beta,\lambda\right)$ be the parameter vector. The log-likelihood function for the data $\left\{\left(u_i,v_i\right)\right\}$ using Equation (13) is given as

$$L\left(\alpha,\beta,\lambda;u,v\right)=n\left(\beta+2\right)\ln\left(\alpha\lambda\right)+\frac{\alpha-\beta}{\beta}\sum_{i=1}^{n}\ln\left(u_iv_i\right)-\lambda\sum_{i=1}^{n}\left(u_i^{\alpha/\beta}+v_i^{\alpha/\beta}\right)+\sum_{i=1}^{n}\ln\left[\left(\beta+1\right)+\ln\left(e^{-\lambda u_i^{\alpha/\beta}}+e^{-\lambda v_i^{\alpha/\beta}}-1\right)^{\beta+1}\right]$$
$$-n\ln\beta+2\sum_{i=1}^{n}\ln\left\{\ln\left(e^{-\lambda u_i^{\alpha/\beta}}+e^{-\lambda u_i^{\alpha/\beta}}-1\right)^{\beta+1}\left(e^{-\lambda u_i^{\alpha/\beta}}+e^{-\lambda u_i^{\alpha/\beta}}-1\right)\right\}. \tag{15}$$

However, it is cumbersome to theoretically obtain the maximum likelihood estimates of Equation (15). Thus, the R program is used to obtain the maximum estimates of the parameters $\psi=\left(\alpha,\beta,\lambda\right)$ using the maxLik function in R.

3.2. The TPP-Independent Copula

This section introduced the TPP-Independent copula as the sub-model of the TPP model.

Consider the independence copula to be expressed as $C\left(u,v\right)=uv$. The derived TPP-Independent copula model is given as

$$C_{Ind.}^{Q}\left(u,v,\lambda,\beta\right)=\left\{\lambda\left[e^{-\lambda\left(u^{-\frac{1}{\beta}}+v^{-\frac{1}{\beta}}\right)}\right]^{-1}\right\}^{\beta}. \tag{16}$$

The corresponding density function to Equation (16) is given as

$$c^{Q}_{Ind.}\left(u,v,\lambda,\beta\right)=\lambda^{\beta+2}u^{-\frac{\beta-1}{\beta}}v^{-\frac{\beta-1}{\beta}}e^{\lambda\beta\left(u^{-\frac{1}{\beta}}+v^{-\frac{1}{\beta}}\right)}. \tag{17}$$

The TPP-Independent copula Kendall's tau expression is given as

$$\tau=4\int_{0}^{1}\int_{0}^{1}\left[\frac{\lambda}{e^{-\lambda\left(u^{-\frac{1}{\beta}}+v^{-\frac{1}{\beta}}\right)}}\right]^{\beta}dudv-1. \tag{18}$$

3.2.1. Parameter estimates of the tpp-independent copula

This section presents the parameter estimates of the TPP-Independent copula model.

Let u_i and v_i $\left(i=1,2,...,n\right)$ be pair of continuous random variables and $\psi=\left(\alpha,\beta,\lambda\right)$ be the parameter vector. The log-likelihood function for the data $\left\{\left(u_i,v_i\right)\right\}$ using Equation (17) is given as

$$L\left(\alpha,\beta,\lambda;u,v\right)=n\left(\beta+2\right)\ln\lambda-\left(\frac{\beta-1}{\beta}\right)\left[\sum_{i=1}^{n}\ln(u_i)+\sum_{i=1}^{n}\ln(v_i)\right]-\lambda\beta\sum_{i=1}^{n}\left(u_i^{-1/\beta}+v_i^{-1/\beta}\right). \tag{19}$$

The R program is used to obtain the maximum estimates of Equation (19) parameters by using the maxLik function in R.

3.3. The Pair Copula Construction and The D-vine Structure

This section presents the pair-copula construction for our variables of interest. The variables of interest are the Crude oil, Gold, and Silver data collected from Yahoo Finance between 1st February 2017 to 26th February 2021. We set $x_1=$ Crude oil, $x_2=$ Gold, and $x_3=$ Silver, the 3-dimensional D-vine structure for the pair-copula construction is shown in Figure 1. The vine structure consists of two trees T_j, $j=1,2$. The tree in T_j has $4-j$ nodes and $3-j$ edges, each tree corresponds to a pair-copula density function and the label of the edge corresponds to the subscript of the pair-copula density function. All the pair constructions are defined by $\frac{k\left(k-1\right)}{2}$ edges and the marginal densities of each variable. k are the dimension of variables for pair construction. Note also that if two edges in a tree T_j share a common node, these edges are joined together to form another node in the tree T_{j+1} (proximity condition).

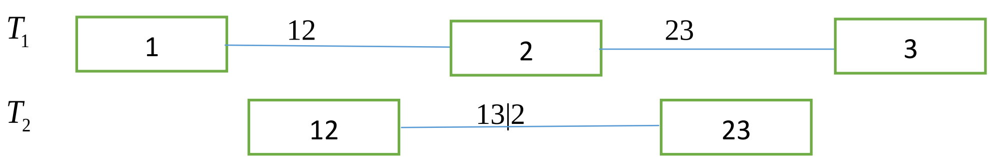

Figure 1. *The D-vine structure of the Crude oil, Gold, and Silver data*

Figure 1 presents the D-vine structure with 3 variables, 3 edges, 5 nodes, and 2 trees. The tree has 3 nodes and T_2 has 2 nodes with a conditional edge. Each edge is associated with a pair-copula. The graph in Figure 1 shows how the variables are linked together. Each linkage (edge) is associated with a pair-copula to explore their dependence. Note that the D-vine structure for the stock data used in section 4 takes the same shape as that of the Crude oil, Gold, and Silver in Figure 1, and hence it is not presented in the study.

Using Equation (2), the expression for the D-vine structure in Figure 1 is given as

$$h(x_1,x_2,x_3)=c_{12}\left(H(x_1),H(x_2)\right)c_{23}\left(H(x_2),H(x_3)\right)c_{13|2}\left(H(x_1\mid x_2),H(x_3\mid x_2)\right)h(x_1)h(x_2)h(x_3)\ , \quad (20)$$

where $h(x_1), h(x_2)$, and $h(x_3)$ are the marginal distributions. The log-likelihood function of Equation (20) is given as

$$L\left(h(x;\Theta,\Omega)\right)=\sum_{j=1}^{n}\ln\left[c_{12}\left(c_{12}\left(H(x_1),H(x_2)\right)c_{23}\left(H(x_2),H(x_3)\right)c_{13|2}\left(H(x_1\mid x_2),H(x_3\mid x_2)\right);\Omega_j\right)\right]+\sum_{j=1}^{n}\ln\left[\left(h(x_1)h(x_2)h(x_3);\Theta_j\right)\right]$$

where Θ_j is the set of parameters in the marginal functions and Ω_j what is the set of parameters in the copula functions?

4. **NUMERICAL APPLICATIONS**

This section presents two numerical applications of the study. Firstly the world's monthly trading volumes of Crude oil, Gold, and Silver were collected from Yahoo Finance between 1st February 2017 to 26th February 2021. The summary statistics for the Crude oil, Gold, and Silver data sets are presented in Table 1a. It is observed that the monthly trading volumes of Crude oil "between" 1st February 2017 to 26th February 2021 are left-skewed and under dispersed. However, the monthly trading volumes of Gold and Silver for the same period are right-skewed and overdispersed. Silver among others has the highest positive skewness value and showed a high level of overdispersion.

Finally on Eustockmarket data (1860 observations each) of the Daily Closing Prices of Major European Stock Indices of Germany DAX (Deutscher Aktien Index), Switzerland SMI (Swiss Market Index), and France CAC (Cotation Assistée en Continu) from 1991–1998. These data sets are available on the R data.frame package. The summary statistics for the daily closing prices of the stocks are presented in Table 1b. All the stocks are right-skewed with France CAC having the highest kurtosis value indicating that it is highly leptokurtic among other stocks. The stocks are all overdispersed. Though, France CAC exhibits a high level of overdispersion among others. We only presented the pairwise scatter plot of the daily closing prices of the stocks in this study.

The scatter plots in Figure 2 are constructed to examine the pair dependencies or association of the variables while the pairwise scatter plot of the variables is presented in Figure 3a. The plots showed how close and far the points are from each other. The correlated values of the variables are presented in Table 2a. For the Crude oil, Gold, and Silver data, the result showed that there exists a positive association between the world's monthly trading volumes of Crude oil, Gold, and Silver. Furthermore, it is observed that Gold and Silver seem to have a better association than the association between the monthly trading volumes of Crude oil and Gold, and Crude oil and Silver. Similarly, we only considered the volatility in the world monthly trading volumes of Crude oil, Gold, and Silver. We obtained the log-return rate of the data and used the EWMA model to compute the monthly volatility rate. The volatility rate plots were obtained for the values and are presented in Figure 4 respectively. These plots give clear trends in the monthly trading volumes of the data. Furthermore, the Q-Q plots for the Crude oil, Gold, and Silver data are presented in Figure 7a and it is observed that only Crude oil data points are seen to be close to the straight line indicating the possibility of it being normally distributed.

Similarly, the pairwise scatter plot of the daily closing prices of the stocks is presented in Figure 3b. The pairwise scatter plot showed how close the stock's daily closing prices are to each other. The Q-Q plots of the stocks are presented in Figure 7b and it is observed that the points are far from the straight line indicating the possibility of the data deviating from normal. The correlated values for the daily closing prices of the stocks are presented in Table 2a. It is seen that DAX and SMI are highly correlated followed by DAX and France CAC, and SMI and France CAC stocks seem to have the least correlated value.

In this study, we only fit the Pareto marginal density to the monthly trading volumes of Crude oil, Gold, and Silver using the maximum likelihood estimation method. We obtained the following estimates from the Crude oil marginal, Gold, and Silver marginal $\hat{\Theta}_3 = \left(\hat{\lambda}_{31}, \hat{\beta}_{32}\right) = \left(1127.8524,\ 0.4060\right)$.

Though the use of scatterplot may help us to guess what bivariate copula model might be appropriate for a given data, however, a guess is not the optimal way of selecting the best bivariate copula model that provides a better fit to our data. Thus, we fit the introduced copula models to our data and compared the results with some existing copula models. The parameters estimates, log-likelihood (LL), Akaike Information Criteria (AIC), and the corresponding Kendall's tau values for the fitted bivariate copula models to our data are presented in Tables 3, 4, 5, 10, 11, and 12. Tables 6 and 13 present the parameter estimates of the copula models fitted to the D-vine structure. We use the loglik and maxLik functions in the R package to obtain the parameter estimates of our models. The model with the least AIC value is considered to be the model with a better fit to the data under consideration. In Table 3, the TPP-Clayton model provides a better fit to the data than other copula models. The competing models are the TPP-Independent model and the Clayton model. This could mean that the introduced TPP-Clayton model is quite appropriate for modeling pair variables with right-skewed, one left-skewed, under, and overdispersed, as shown in Tables 1a and 1b respectively. The pair variable under consideration is the Crude oil and Gold data and DAX, SMI, and France CAC stocks. In Table 4, the TPP-Independent model provides a better fit to the data than other copula models. The competing models are the TPP-Clayton and Clayton models. Thus, the TPP-Independent model could be considered appropriate in modeling pair variables with one left-skewed and under dispersed and the other highly right-skewed and overdispersed. Though the TPP-Clayton model has a parameter more than the parameter of the TPP-Independent model it performs poorly to the data under consideration than the TPP-Independent model. This result could also mean that the model with more parameters may not likely provide a good fit for some data on various behavior. The pair variable under consideration is the Crude oil and Silver data. In Table 5, the pair variables are Gold and Silver and the TPP-Clayton model provides a better fit to the data than other copula models. In this case, the competing models are the TPP-Independent and Gaussian models. Hence, the TPP-Clayton model could be considered appropriate in modeling pair variables with both right-skewed and overdispersed. Tables 10, 11, and 12 examined the fitness of the models on the following pairs variables DAX and SMI, DAX and France CAC, and SMI and France CAC respectively. All the results presented in Tables 10 and 11 showed that the proposed TPP-Independent copula model provides the best fit to the pair of the daily closing prices of stock data sets than other models while the TPP-Clayton model seems to be the only competiting model. Furthermore, in Table 12 it is observed that the TPP-Clayton copula provides the best fit for the pair of SMI and France CAC stocks. These stocks are right-skewed and France's CAC is highly overdispersed. This implies that TTP-Clayton copula could be the appropriate option in modeling pairs that are strongly correlated and one is highly skewed and overdispersed. This result is in line with the result obtained in Table 5 in favor of the TPP-Clayton model.

The parameter estimates of the D-vine structure in Figure 1 are based on the [21] algorithm for sequential top-down estimation to maximize the log-likelihood of the tree for a regular vine copula. The steps of the algorithm are presented in Table 7. The results obtained for the fitted copula models to the D-vine structure using Crude oil, Gold, and Silva and DAX, SMI, and France CAC stock data sets are presented in Tables 6 and 13 respectively. The copula models were selected using the RVineStructureSelect() in the R vine package. The choice of the copula models is based on their characteristics (see Aas et al. 2009) for modeling data from various fields. The LL and AIC values for the estimated parameters are reported. The AIC is selected using the select()AIC function in the R vine package. From Tables 6 and 13, the Gaussian copula model provides the least AIC value among other copula models, and hence, it is considered the model that yielded the best goodness of fit to the data sets under consideration. The competiting model to the Gaussian copula is the Student-t copula model for the Crude oil, Gold, and Silver data in Table 6 while the Gumbel copula model performed second best to the stocks data as seen in Table 13. This means that the Gaussian copula model could be considered appropriate for modeling a D-vine structure comprising variables that are right-skewed, left-skewed, over and under dispersed, and positively correlated while the competiting model could be the Gumbel copula for all right-skewed, overdispersed, and highly correlated variables and Student-t copula for right-skewed, left-skewed, nearly no correlated (weak), under and overdispersed variables.

4.1. Simulation Study

This section presents the simulation study to examine the flexibility behavior of the proposed copula models on large sample size. The adopted simulation procedures are presented below.

1. Generate two independent standard-uniform (0,1) variates u and v.
2. Repeat the process for $n = 2000$ times.
3. Fit the bivariate copula models.

The pair and the scatter 3D plots of the generated variates are presented in Figures 5 and 6 respectively. The flexibility behavior of the copula models results obtained based on the generated variates are presented in Table 8.

The LL and AIC values are used as the measure of goodness of fit criteria of the copula model that provides a better fit to the generated data. Based on the generated data of size 2000 from uniformly distributed variates u and v, the introduced TPP-Clayton copula model provides a better fit to the data than other copula models. Furthermore, the TPP-Independent becomes the most competing model in terms of goodness of fit to the generated variates. Ultimately, in both real-life data sets and the generated data, the introduced copula models have proven to yield better fits to the data sets in both cases. Thus could be considered an alternative model to the Gaussian copula, Clayton copula, Frank copula, and Joe copula models respectively.

Table 1a. *The summary statistics for the Crude oil, Gold, and Silver data sets*

Variabes	skewness	kurtosis	mean	median	1^{st}Q	3^{rd}Q	Var.
Crude oil	-0.288	2.117	121	129	993	141	0.000
Gold	0.659	1.790	131	160	920	287	2380
Silver	1.027	3.659	515	647	257	103	3731

Table 1b. *The summary statistics for the daily closing prices of stocks data sets*

Variables	skewness	kurtosis	mean	median	1^{st}Q	3^{rd}Q	Var.
DAX	1.534	4.565	253	214	174	272	1177
SMI	1.309	3.832	338	270	217	381	2766
France CAC	1.946	6.271	223	199	188	227	3368

Table 2a. *The correlated values of Crude oil, Gold, and Silver*

Variables	Crude oil *Gold	Crude oil*Silver	Gold*Silver
	0.0569	0.0755	0.1800

Table 2b. *The correlated values of DAX, SMI, and France CAC stocks*

Variables	DAX *SMI	DAX*France CAC	SMI*France CAC
	0.9912	0.9662	0.9478

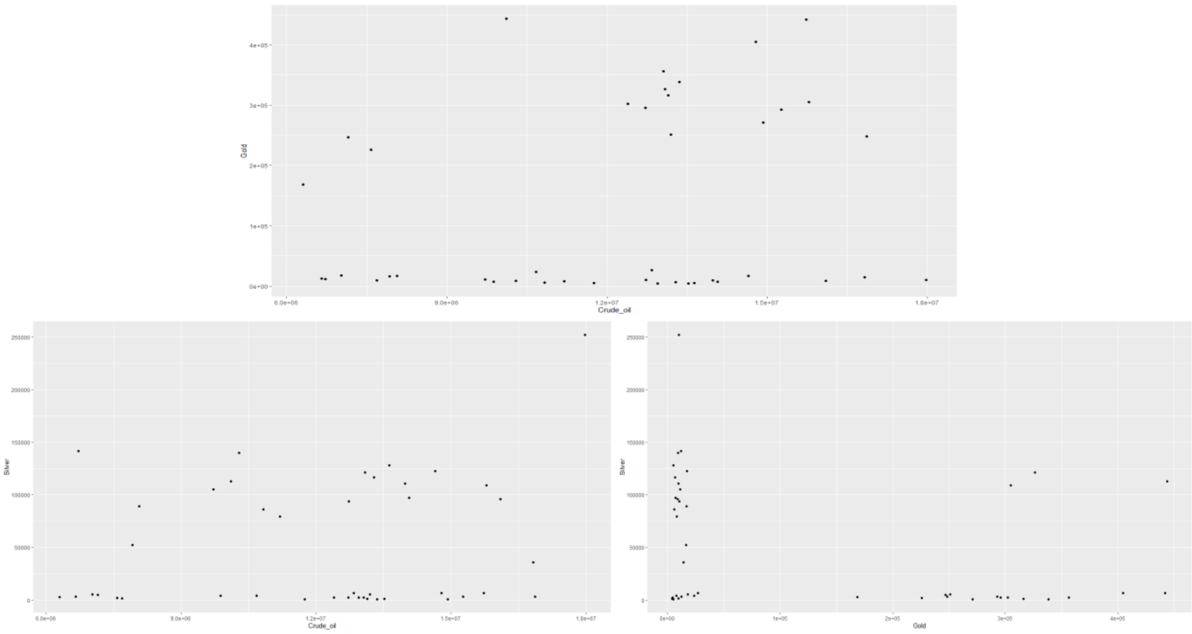

Figure 2. *The scatter plots of the real trading volumes scale of Gold and Crude oil (top), Silver and Crude oil (left), and Silver and Gold (right)*

The real trading volumes scale of Gold, Crude oil, and Silver showed that the points are dispersed or far away from each other. This implies that there exists no such stronger relationship among the real trading volumes scale of Gold, Crude oil, and Silver. Similarly, Figure 3 showed such a trend in the pairwise pattern.

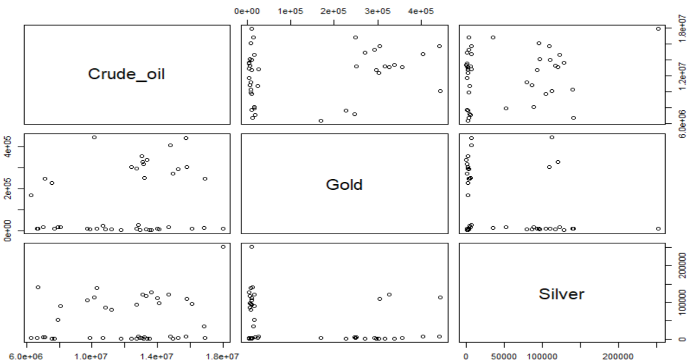

Figure 3a. *The pairwise scatter plot*

The plot presents how each variable in pair tends to form close points. From the plot, the points are nearly dispersed from each other as only a few points seem to be close indicating the level of the association among the variables to be nearly weak. However, in Figure 3b, the points are very close to each other indicating the possibility of a strong association between the pair of stocks.

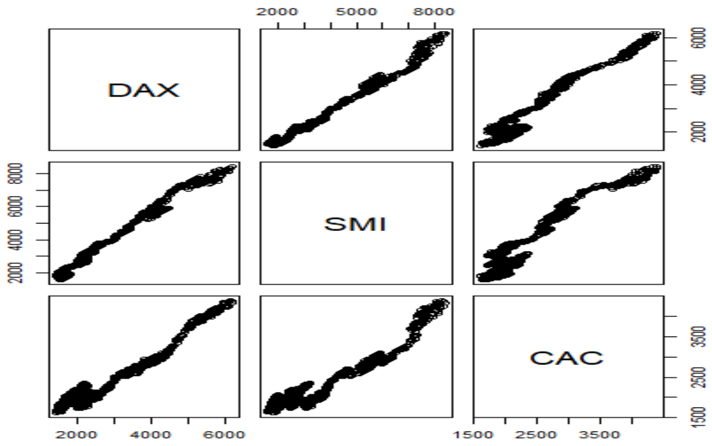

Figure 3b. *The pairwise scatter plot*

4.2. Skewness and Kurtosis Tests for the Data Sets

These are statistical measures that represent the distribution of data. These tests describe the position of many of the data concerning the distribution against the average value of the specific data. The Shapiro-Wilk (W) test is used to test the null hypothesis (H_0) that the data sets are normally distributed against the alternative (H_1) at a 0.05 significance level. The Q-Q plots for the datasets are presented in Figures 7a and 7b respectively. The W test for the variables is presented in Tables 9a and 9b. The W test for Crude oil is in favor of the null hypothesis since the p-value is greater than 0.05. Similarly, the Q-Q plot for Crude oil supports this. However, the W tests for all the daily closing prices of the major stocks, Gold, and Silver are in favor of the alternative hypothesis since their p-values are less than 0.05. Their Q-Q plots showed that the datasets deviate from normal.

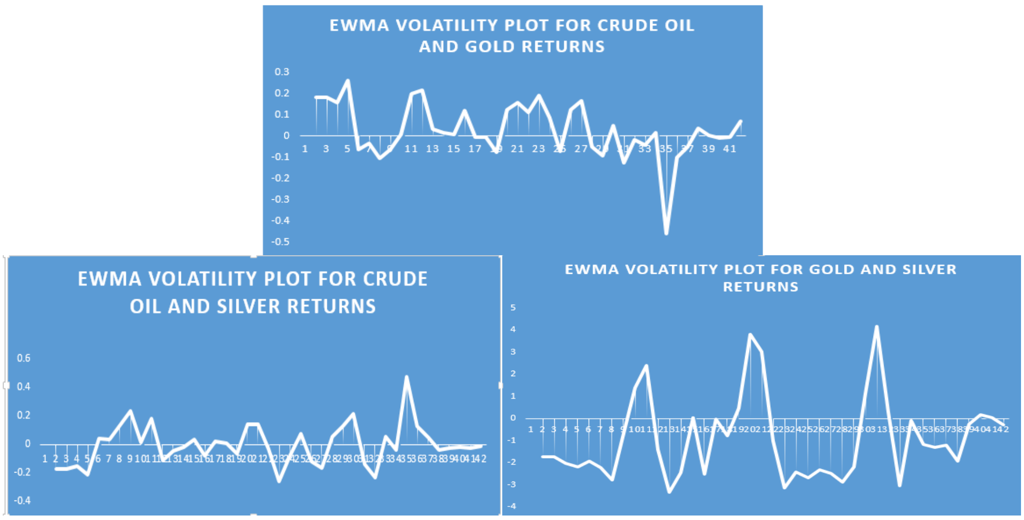

Figure 4. *EWMA plot*

Figure 4 showed that the volatility of the Crude oil, Gold, and Silver returns plots exhibit upward and downward patterns in the trends for the given periods. The first upward trend in Crude oil and Gold returns is seen around the first 5 months where a decline occurred until another upward trend is observed around the 10th month. This process continues as seen in Crude oil and Silver returns and Gold and Silver returns. Though, the upward and downward trends occurred during various periods.

Table 3. *Parameters estimates, LL, AIC, and Kendall's tau values for the copula models fitted to Crude oil and Gold data*

Copula model	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	LL	AIC	tau
TPP-Independent	0.265	-	-1659.389	45142.690	-90281.390	NA
TPP-Clayton	-0.125	-0.047	0.001	161.000	-3224566.000	NA
Gaussian	0.590	-	-	106.620	-211.250	0.400
Clayton	65.530	-	-	134.620	-267.230	0.970
Frank	67.080	-	-	134.320	-266.640	0.940
JOE	50.000	-	-	133.070	-264.130	0.960

[*]Key. NA (not available)

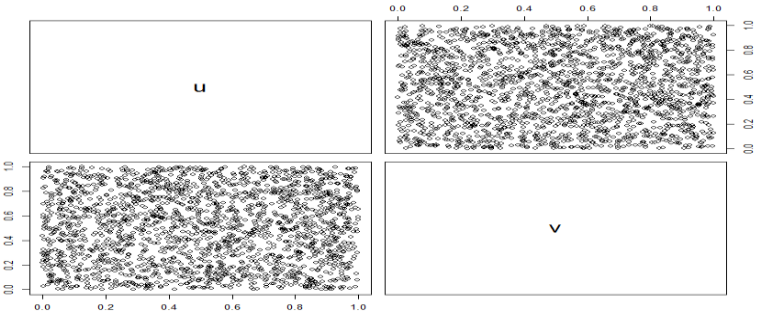

Figure 5. *Pair plot of the independent uniformly generated variates u and v*

Figure 5 presents the pairwise plot of the uniformly simulated variates. It showed that the points are very close to each other and this indicates how strongly the variates could be correlated. Figure 6 showed the visualized 3D plot of the simulated variates.

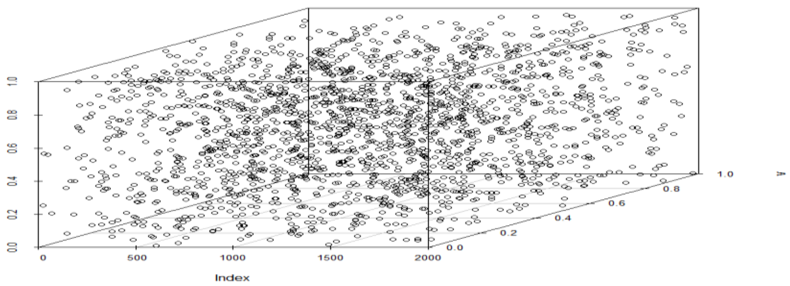

Figure 6. *The 3D scatter plot of the independent uniformly generated variates u and v*

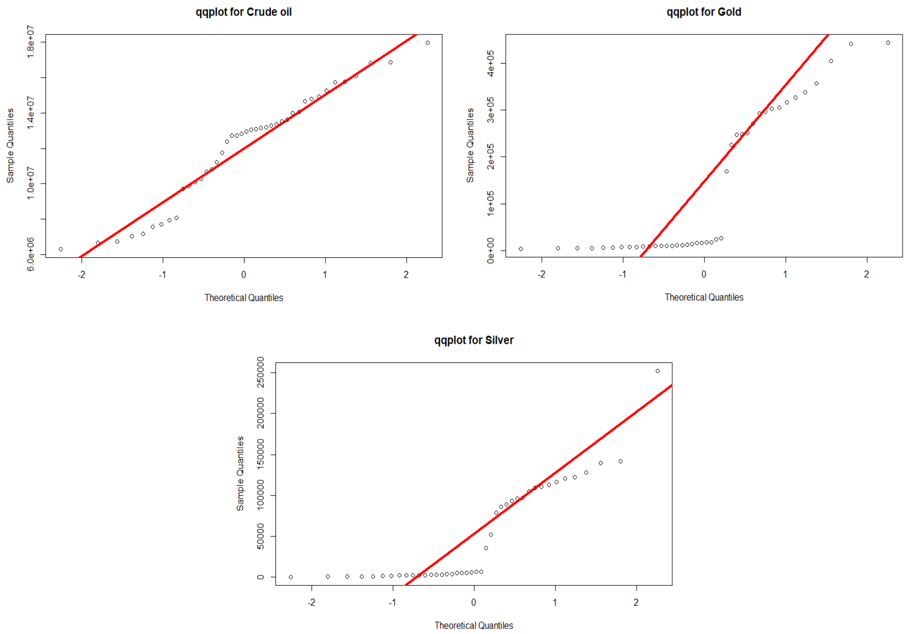

Figure 7a. *Q-Q plot for Crude oil, Gold, and Silver datasets*

The plots show that Crude oil data points are close to the straight line while only a few of the Gold and Silver data points are seen close to the straight. That is, many of the Gold and Silver data points are far away from the straight line. This is an indication that the Gold and Silver datasets are in favor of the alternative hypothesis. Similarly, Figure 7b presents the Q-Q plots for the stocks data and showed how a few data points are on the straight line. This indicates that the data sets are in favor of the alternative hypothesis.

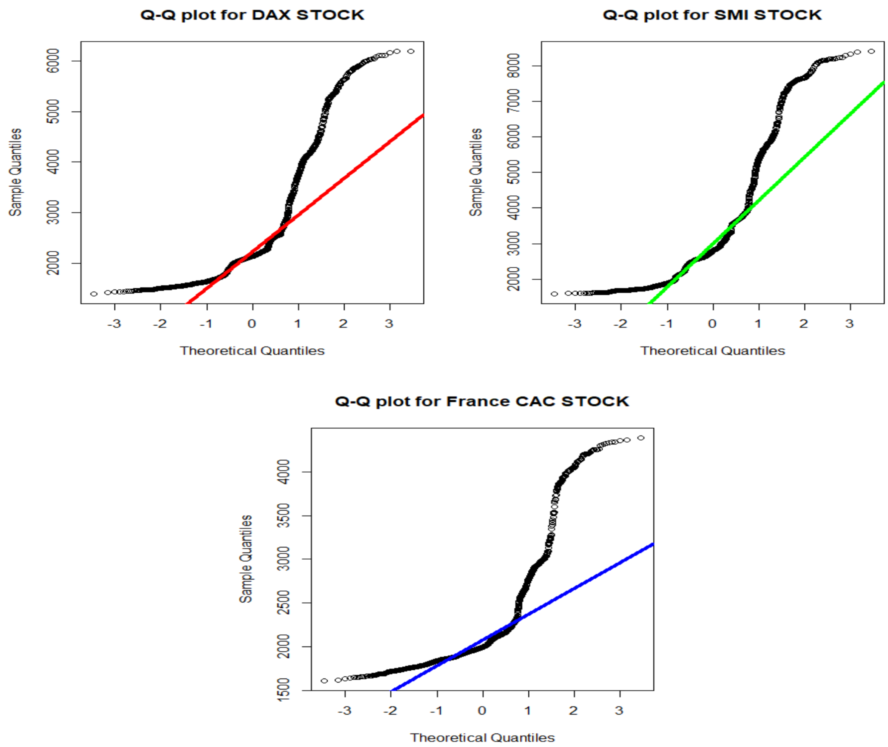

Figure 7b. *Q-Q plot for the DAX, SMI, France CAC, and UK FTSE stocks data*

Table 4. *Parameters estimates, LL, AIC, and Kendall's tau values for the copula models fitted to Crude oil and Silver data*

Copula model	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	LL	AIC	tau
TPP-Independent	0.264	-	-1593.273	43326.230	-86648.460	NA
TPP-Clayton	0.053	0.141	0.058	42014.930	-84023.860	NA
Gaussian	0.550	-	-	105.490	-208.970	0.370
Clayton	65.090	-	-	134.530	-267.070	0.970
Frank	66.630	-	-	134.240	-266.480	0.940
JOE	50.000	-	-	133.050	-264.090	0.960

***Key. NA (not available)**

Table 5. *Parameters estimates, LL, AIC, and Kendall's tau values for the copula models fitted to Gold and Silver data*

Copula model	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	LL	AIC	tau
TPP-Independent	0.265	-	-1241.528	33668.010	-67332.020	NA
TPP-Clayton	0.000	0.195	0.000	631421.600	-1262837.000	NA
Gaussian	0.860	-	-	301.750	-601.490	0.660
Clayton	0.000	-	-	-0.050	-2.110	0.000
Frank	100.000	-	-	192.460	-382.920	0.960
JOE	-1.000	-	-	0.060	-2.130	0.000

***Key. NA (not available)**

Table 6. *Parameter estimates of the copula models fitted to the D-vine structure of the Crude oil, Gold, and Silver data*

Tree	Edge	Copula model	Parameter	tau	LL	AIC
1	2, 3	Gaussian	0.29	0.19		
	1, 2		0.22	0.14		
2	1, 3|2		0.33	0.22	85.50	-165.00
***********						
1	2, 3	Student-t	0.29	0.19		
	1, 2		0.20	0.13		
2	1, 3|2		0.28	0.18	87.60	-163.00
***********						
1	2, 3	Clayton	-	-		
	1, 2		-	-		
2	1, 3|2		0.12	0.06	62.00	-118.00
***********						
1	2, 3	Gumbel	-	-		
	1, 2		-	-		
2	1, 3|2		1.73	0.42	66.80	-128.00
***********						
1	2, 3	Frank	2.37	0.24		
	1, 2		3.34	0.33		
2	1, 3|2		5.61	0.49	54.60	-103.00
***********						
1	2, 3	Joe	-	-		

	1, 2	-	-		
2	1, 3|2	8.41	0.79	52.40	-98.80

Table 7. *The Algorithm for sequential top-down estimation steps*

- Input the data $u_{i,k}$ for all possible variable pairs, $1 \le i < k \le n$
- Select the maximum spanning tree
- **for** each edge say $\left(e \in E\right)$ **do**
- Select a copula say $C_{a,b}$ with estimated parameter(s) $\hat{\Omega}_{a,b}$
- For, generate the pseudo observations
- **end for**
- **for** $j = 2,...,n-1$ **do**
- Calculate the empirical Kendall's tau for all conditional pairs that can be part of the tree T_J. That is, all edges fulfilling the proximity condition
- Among these edges, select the maximum spanning tree.
- **for** each edge say $\left(e \in E\right)$ **do**
- Select a pair copula $C_{a,b}$ with estimated parameter(s) that provides the best fit based on goodness of fit measure criteria.

Table 8. *Parameters estimates, LL, AIC, and Kendall's tau values for the copula models based on generated uniform variates u and v*

Copula model	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	LL	AIC	tau
TPP-Independent	219.000	-	0.686	144.000	-284.000	NA
TPP-Clayton	-0.104	-0.215	0.007	429.000	-852.000	NA
Gaussian	0.050	-	-	2.410	-2.820	0.030
Clayton	0.050	-	-	2.030	-2.050	0.020
Frank	0.270	-	-	1.880	-1.750	0.030
JOE	1.010	-	-	0.060	-1.890	0.000

***Key. NA (not available)**

Table 9a. *Shapiro test of normality for the Crude oil, Gold, and Silver data*

Variables	W	P-value	Decision
Crude oil	0.950	0.067	H_0 (accepted)
Gold	0.753	5.03×10^{-7}	H_0 (not accepted)
Silver	0.779	1.65×10^{-6}	H_0 (not accepted)

Table 9b. *Shapiro test of normality for the DAX, SMI, CAC, and FTSE data*

Variables	W	P-value	Decision
DAX	0.799	2.2×10^{-16}	H_0 (not accepted)
SMI	0.838	2.2×10^{-16}	H_0 (not accepted)
France CAC	0.739	2.2×10^{-16}	H_0 (not accepted)
UK FTSE	0.889	2.2×10^{-16}	H_0 (not accepted)

Table 10. *Parameters estimates, LL, AIC, and Kendall's tau values for the copula models fitted to DAX and SMI data*

Copula model	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	LL	AIC	tau
TPP-Independent	0.207	-	-4658.000	81850.000	- 16370.000	NA
TPP-Clayton	0.059	0.193	0.002	53074.00	-10615.000	NA
Gaussian	0.930	-	-	1899.530	-3797.060	0.770
Clayton	11.570	-	-	29330110	-5864.230	0.850
Frank	25.830	-	-	2519.38	-5036.750	0.860
JOE	12.300	-	-	2938.900	-5875.800	0.850

***Key. NA (not available)**

Table 11. *Parameters estimates, LL, AIC, and Kendall's tau values for the copula models fitted to DAX and France CAC data*

Copula model	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	LL	AIC	tau
TPP-Independent	0.206	-	-4558.450	80169.000	- 16034.000	NA
TPP-Clayton	0.060	0.194	0.003	25152.000	-50305.000	NA
Gaussian	0.870	-	-	1309.560	-2617.120	0.670
Clayton	4.330	-	-	1584.520	-3167.030	0.680
Frank	9.010	-	-	1074.380	-2146.770	0.640
JOE	5.070	-	-	1582.610	-3163.210	0.680

***Key. NA (not available)**

Table 12. *Parameters estimates, LL, AIC, and Kendall's tau values for the copula models fitted to SMI and France CAC data*

Copula model	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	LL	AIC	tau
TPP-Independent	0.032	-	-1242.000	74212.000	- 14842.000	NA
TPP-Clayton	0.002	0.068	0.259	40182.000	-80364.000	NA
Gaussian	0.840	-		1131.120	-2260.240	0.630
Clayton	3.840	-	-	1445.730	-2889.450	0.660
Frank	8.420	-	-	990.240	-1978.480	0.620
JOE	4.570	-	-	1444.190	-2886.380	0.650

***Key. NA (not available)**

Table 13. *Parameter estimates of the copula models fitted to the D-vine structure of the DAX, SMI, and France CAC data*

Tree	Edge	Copula model	Parameter	tau	LL	AIC
1	2, 3	Gaussian	0.99	0.90		
	1, 2		0.54	0.37		
2	1, 3|2		0.25	0.16	16.50	-270.00
***********						
1	2, 3	Student-t	0.99	0.90		

	1, 2		0.47	0.31		
2	1, 3|2		0.22	0.14	15.10	-240.20
***********						
1	2, 3	Clayton	3.70	0.65		
	1, 2		28.00	0.93		
2	1, 3|2		-	-	14.20	-220.40
***********						
1	2, 3	Gumbel	4.38	0.77		
	1, 2		2.26	0.56		
2	1, 3|2		1.39	0.28	16.00	-260.00
***********						
1	2, 3	Frank	35.00	0.89		
	1, 2		35.00	0.89		
2	1, 3|2		8.67	0.63	14.00	-220.00
***********						
1	2, 3	Joe	4.40	0.64		
	1, 2		30.00	0.94		
2	1, 3|2		-	-	15.00	-240.00

5. SUMMARY AND CONCLUSION

The study introduced the transformed pair Pareto copula model and TPP-Clayton and TPP-Independent copulas as sub-models and a D-vine structure model for our data. The application of the study is illustrated using two sets of real-life data and simulated data of size 2000. The pair plots of the real-life and simulated data sets are presented. The parameter estimates of the models were obtained based on the maximum likelihood estimation method and the [21] algorithm for sequential top-down estimation. The AIC values for the models are used as a goodness of fit criteria to select the appropriate model that provides a better fit to our data. In Tables 3-5 and 10-12, it is observed that the TPP-Independent and TPP-Clayton models performed excellently well to the real-life data than other models in terms of their fit statistics. Similarly, Table 8 presents the parameter estimates of the copula models based on the simulated data. Again, it is observed that the TPP-Independent and TPP-Clayton models performed excellently well to the simulated data than other models in terms of their fit statistics. Furthermore, in Tables 6 and 13, the Gaussian copula model outperformed other models in terms of their fit statistics and could be considered an appropriate model for modeling a D-vine structure. Though we could not obtain results for Kendall's tau values of our introduced models for both the real-life and simulated data sets, however, the future research questions are how to obtain Kendall's tau values for the introduced models and a study on how these models could be used to obtain the estimates of a given vine structure.

Furthermore, in Table 2a, we observed that Gold and Silver have a better correlation value than other pairs. This is an indication of possible dependence or association between Gold and Silver monthly trading volumes. Gold and Silver trading volumes dependence seems to be promising among other pairs' trading volumes. Their pair dependence value indicates that there is a better positive association between their trading volumes. Hence, Gold and Silver trading may likely be promising in pairs compared to other pairs. Similarly, DAX and SMI stocks have a high correlation value than other pairs and it is an indication of possible dependence or strong association between the DAX and SMI stocks. DAX is a stock market index consisting of the largest 40 and most liquid German companies that trade on the Frankfurt Stock Exchange while SMI is a major stock market index in Switzerland and the 20 largest and most liquid stocks based on the Swiss Exchange. Conclusively, the pair trading of these two stocks by investors or stock marketers could be very helpful as it may barricade some undercurrent risks because an unexpected performance of one could be complemented by the other and minimize risk in the market base on their pair dependence level among other pairs.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Sklar, M., “Fonctions de repartition an dimensions et leurs marges”, Public Institute of Statistics, Univiversity of Paris, 8: 229–231, (1959).

[2] Nelsen, R.B., “An Introduction to Copulas Lecture Notes in Statistics 139”, Springer Verlag New York, (1999).

[3] TovarCuevas, J.R., Portilla Y.J., Achcar, J.A., “A method to select bivariate copula Functions”, Revista Colombiana de Estadística, 42(1): 61-80, (2019).

[4] Achcar, J.A., Moala, F.A., Tarumoto, M.H., Coladello, L.F., “A bivariate generalized exponential distribution derived from copula functions in the presence of censored data and covariates”, Pesquisa Operacional, 35(1): 165-186, (2015).

[5] Kuvattana, S., Busababodhin, P., Areepong, Y., Sukparungsee, S., “Bivariate copulas on the exponentially weighted moving average control chart”, Songklanakarin Journal of Science & Technology, 38(5): 569-574, (2016).

[6] Çelebioglu, S., “Generating copulas by analytical means”, Hacettepe Bulletin of Natural Sciences and Engineering, 26: 153-173, (1997).

[7] Mukherjee, S., Lee, Y., Kim, J.M., Jang, J., Park, J.S., “Construction of bivariate asymmetric copulas”, Communications for Statistical Applications and Methods, 25(2): 217-234, (2018).

[8] Kim, D., Kim, J.M., “Analysis of directional dependence using asymmetric copula-based regression models”, Journal of Statistical Computation and Simulation, 84(9): 1990-2010, (2014).

[9] Latif, S., Mustafa, F., “copula-based multivariate flood probability construction: A Review”, Arabian Journal of Geosciences, 13(3): 1-25, (2020).

[10] Oh, R., Ahn, J.Y., Lee, W., “On copula-based collective risk models: from elliptical copulas to vine copulas”, Scandinavian Actuarial Journal, 2021(1): 1-33, (2021).

[11] Susam, S. O., “A new family of archimedean copula via trigonometric generator function”, Gazi University Journal of Science, 33(3): 806-813, (2020).

[12] Najjari, V., “The dependence structure between the monthly minimum and maximum barometric data in Iran”, Gazi University Journal of Science, 29(3): 593-597, (2016).

[13] Bedford, T., Cooke, R.M., “Probability density decomposition for conditionally dependent random variables modeled by vines”, Annals of Mathematics and Artificial Intelligence, 32: 245–268, (2001).

[14] Bedford, T., Cooke, R.M., “Vines - a new graphical model for dependent random variables", Annals of Statistics, 30 (4): 1031–1068, (2002).

[15] Aas, K., Czado, C., Frigessi, A., Bakken, H., “Pair-copula constructions of multiple Dependence", Insurance: Mathematics and Economics, 44(2): 182-198, (2009).

[16] Zhao, Y., Liu, Q., Kuang, J., Xie, K., Du, W., “Modeling multivariate dependence by nonparametric pair-copula construction in composite system reliability evaluation”, International Journal of Electrical Power & Energy Systems, 124: 106-373, (2021).

[17] Deng, Y., Chaganty, N.R., “Pair-copula Models for Analyzing Family Data”, Journal of Statistical Theory and Practice, 15(1): 1-12, (2021).

[18] Takeuchi, T.T., Kono, K.T., “Constructing a multivariate distribution function with a vine copula: towards multivariate luminosity and mass functions”, Monthly Notices of the Royal Astronomical Society, 498(3): 4365-4378, (2020).

[19] Valdez, E.A., Xiao, Y., “On the distortion of a copula and its margins”, Scandinavian Actuarial Journal, 2011(4): 292-317, (2011).

[20] Durante, F., Foschi, R., Sarkoci, P., “Distorted copulas: constructions and tail Dependence”, Communications in Statistics-Theory and Methods, 39(12): 2288-2301, (2010).

[21] Dißmann, J., Brechmann, E., Czado, C., Kurowicka, D., “Selecting and estimating regular vine copulae and application to financial returns”, Computational Statistics and Data Analysis, 59(1): 52-69, (2013).