Global Journal of Mathematical Analysis, 5 (2) (2017) 29-36

Global Journal of Mathematical Analysis

Website: www.sciencepubco.com/index.php/GJMA
doi: 10.14419/gjma.v5i2.7256
Research paper

The modeling of extreme stochastic dependence using copulas and extreme value theory: case study from energy prices

Ömer Önalan*

Faculty of Business Administration, Marmara University, Istanbul, Turkey
**Corresponding author E-mail: omeronalan@marmara.edu.tr*

Abstract

In this paper, we investigate the properties of tail dependence with an approach which is based on the copula models and extreme value theory to obtain a joint distribution function of extreme events and to quantify the dependence between random variables. To achieve this objective, we quantify the large co-movements between the random variables returns which are based on the data set daily quotes of exceeds the threshold value of random variables. In this study, stochastic dependence was modeled by the copulas which it provides a good approach for constructing multivariate probability distributions with flexible marginal's and different forms of dependence. Choosing the right copula is very important in modeling. The multivariate distributions are easily simulated using the copulas. Finally we can describe the copula family which correctly represents the dependence. To demonstrate the usefulness of the proposed models, we confine our analysis to big price changes of energy commodity spot prices. The empirical findings demonstrated that the copula model which is combined the extreme value theory is a good approach to model the together extreme large changes**.**

Keywords: *Copulas, energy commodity spot prices, extreme value theory, tail dependence.*

1. Introduction

This document can be used as a template for nderstanding the occurrence of extreme events is very important for different scientific fields like energy markets, finance, seismology [1]. A rare event has drastic consequences. The main motivation of this study is research the interaction between complex systems. In this study we review basic ideas on temporal dependence on discrete time series of energy assets. In traditional approach, relationship between the random variables is described by multivariate distribution functions. The main deficit of this approach that is all of the marginal distribution functions must have same distribution family. In the other words each marginal distribution has same type with multivariate distribution function. This approach is not very common in practice. In literature, to model the dependence structure between random variables, as an alternative approach is proposed copula functions approach. Through the use of Sklar theorem, the copula function based on different marginal distribution family provides to create a multivariate distribution [2].

Tail dependence can be used to represent extreme events of between the random variables. Correlation coefficients are attractively metric for the dependence between random variables. The correlation coefficients may only provide qualitative descriptions of the dependence between random variables. They are based on two point expectation with one parameter. Tail dependence has a nonlinear dependence. If we focus on simultaneously large values of random variables, the correlation coefficients may fail. Copula functions can be use to describe the nonlinear, asymmetric and tail dependence structures. Using the copula function we construct the joint distribution of the variables without to the type of their marginal distributions [3].

The word copula is derived by in Latin means link or tie of different things copulare word. Copulas are models for the joint distributions of random variables with arbitrary marginal distributions. In the other words, a copula is n - variety cumulative distribution on $[0,1]^d$ with uniform univariate marginal [2].

Copulas allows us to express dependence on a quantile scale, this led to of risk in terms of quantiles of loss distributions. Copula modeling presents a two-step approach to fitting copulas to observations data. The first the functional form determined of postulated copula models and second Secondly the inverse transform sampling method is applied to obtain the same marginal distributions as in the original data. The tail of a joint distribution is representing the events which are rare occurrence [4]. Understanding the structure of rare events is many important for politics decision makers and risk managers.

In recent years, Oil and its derivatives have huge trade volumes in energy markets. Energy prices are determined on free market with demand and supply. So energy markets have become more volatile. Energy producers and consumers face to face the financial risks. Understanding of the joint dynamics of energy commodity prices have become indispensable part for managing energy risk. When, modeling of energy price process, these financial risks which are originating from different characteristics must be taken into accounts. The fluctuations in prices of the risky assets are led to big lose. The reason of this phenomenon is the extreme events which are in return distribution tails. This kind of risk is called the tail risk. The tail dependence can be use to modeling extreme relationship between random variables.

The goal of this paper is to modeling the stochastic dependence structure with copula functions of between random variables and to characterize and measure extremely dependence which concentrated on the upper and lower quadrant tails of the joint distribu-

Copyright © 2016 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

tion. We achieve this aim by via methods for estimating tail properties from copulas and extreme value theory. These methods applied subsequently to analyze tail behavior of some energy commodity spot price data.

The paper is organized as follows. In section 2, we present a brief review copula theory, in section 3, we explain the dependence models and their relationship copula functions, in section 4, we are modeled the tail dependence with copulas, in section 5, we applied the models to energy data, finally in section 6, we presents our conclusions.

2. Copulas

An A copula is a multivariate distribution that separates the univariate marginal and the multivariate dependence structure. The copula models allow us to combine marginal models with a variety of possible dependence models. To building a multivariate distribution using a copula, the each marginal variable is to transform a marginal variable which has a uniform distribution. The dependence structure can be expressed as a multivariate distribution on the obtained uniforms. For further details, the theory of copulas we refer to [2],[5],[1],[6],[7],[8].

Definition2.1. a $d-$ dimension copula is a distribution function on $[0,1]^d$ with standard uniform marginal distributions. It satisfies the following conditions;

$i)\ C(u_1,u_2,...,u_d)$, it is increasing according to each u_i , $i\in\{1,2,...,d\}$

$ii)\ C(1,...1,u_i,1,...1)=u_i$, $C(u_1,..u_{i-1},0,u_{i+1},..u_d)=0$

$i\in\{1,2,...,d\}$, $u_i\in[0,1]$

$iii)$ For $1\le i\le d$ $a_i\le b_i$ following inequalities is satisfied.

$$\sum_{i_1=1}^{2}...\sum_{i_d}^{2}(-1)^{i_1+...+i_d}C\left(u_{1i_1},...,u_{di_d}\right)\ge 0$$

Definition2.2. If the random variable X has a continuous marginal distribution F then the cumulative distribution $F(X)$ is uniform distributed on interval $(0,1)$ .Let U is standard uniform random variable,

$$F^{-1}(y)=\inf\{x:F(x)\ge y\} \quad (1)$$

$P\left(F^{-1}(u)\le x\right)=F(x)$ .This result is often used in simulation works. If we know inverse of cumulative distribution function, drawing the uniform numbers, we can random sampling from distribution function F .Let $X=(X_1,X_2,...,X_n)$ be a $d-$ dimension random. The vector with joint distribution $F(X)=P(X_1\le x_1,...,X_d\le x_d)$ and continuous marginal distributions $\{F_i(x_i)=P(X_i\le x_i)\}_{i=1}^{d}$ the components of the random vector are uniform distributed [9].

Since $X_i\overset{d}{=}F_{X_i}^{-1}(U_i)$, $i=1,2,...,n$ the marginal distribution family $(X_1,X_2,...,X_n)$ and $\left(F_{X_{1i}}^{-1}(U_1),F_{X_2}^{-1}(U_2),...,F_{X_n}^{-1}(U_n)\right)$ is the same. Every distribution function can be expressed as a copula function. Furthermore a copula unified the marginal distributions is a special multivariate distribution function. Sklar theorem is associated the joint distributions of random vectors to copulas.

Theorem 2.1(Sklar[16]) Let $X=(X_1,X_2,...,X_d)$ be a random vector with marginal distribution $(F_1,...,F_d)$ .Then there is a copula $C:[0,1]^d\to[0,1]$ such that

$$F_{X_1,..,X_d}(x_1,..,x_d)=C(F_1(x_1),...,F_d(x_d)) \quad (2)$$

If each marginal F_i are continuous then copula C is unique. Otherwise C only determine on $F_1\times F_2\times...\times F_d$. We assume that F_i marginal is continuous with $F\circ F^{-1}(x)=x$,

$$C(u_1,u_2,...,u_d)=P\left(X_1\le F^{-1}(u_1),...,X_d\le F^{-1}(u_d)\right) \quad (3)$$

The copulas are invariant under the strictly increasing transforms. The multivariate modeling with copulas consists of two parts:

$i)$ Separate modelling of each marginal distribution

$ii)$ Modelling the multivariate dependence structure

Every copula $C(u_1,u_2,...,u_d)$ has the Frechet-Hoeffding bounds [2].

$$Max\left\{\sum_{i=1}^{d}u_i+1-d,0\right\}\le C(u_1,u_2,...,u_d)\le Min\{u_1,u_2,...,u_d\} \quad (4)$$

Marginal distributions: a marginal distribution is individual probability distribution in multivariate distribution. Marginal distributions are obtained from observed data. Let random variable X has an invertible (CDF) $F_X(x)=P(X\le x)$. Since $F_X(x)$ is uniform distributed on interval $[0,1]$ for any $\alpha\in[0,1]$,

$$P(F(X)\le\alpha)=P\left(X\le F^{-1}(\alpha)\right)=F\left(F^{-1}(\alpha)\right)=\alpha$$

Let random variable U is sampled from a uniform distribution, since $F_X(X)=U$ and $F_X^{-1}(U)\sim X$,

$$\begin{aligned}F\left(F_{X_1}(x_1),...,F_{X_n}(x_n)\right)&=P\left(U_1\le F_{X_1}(x_1),...,U_n\le F_{X_n}(x_n)\right)\\&=P\left(F_{X_1}^{-1}(U_1)\le x_1,...,F_{X_n}^{-1}(U_n)\le x_n\right)\\&=F_{X_1,X_2,...X_n}(x_1,x_2,...,x_n)\end{aligned} \quad (5$$

There is much kind of copula functions.

$i)$ Independent copula; the independent copulas is described by $C_B(u_1,u_2,...,u_d)=\prod_{i=1}^{d}u_i$ .The random variables are independent is if and only if their copula become independent copula.

$ii)$ Copulas is derivate from distribution functions: A copula can be obtain from probability distributions as,

$$C(u_1,...,u_d)=F\left(F^{-1}(u_1),...,F^{-1}(u_d)\right) \quad (6)$$

Gaussian copula:

$$\begin{aligned}C(u_1,u_2,...,u_d)&=\Phi_\tau\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2),...,\Phi^{-1}(u_d)\right)\\&=\Phi_\tau\left(\Phi^{-1}(F_1(x_1)),\Phi^{-1}(F_2(x_2)),...,\Phi^{-1}(F_d(u_d))\right)\end{aligned}$$

For $d=2$, Copula density function of two- variable Gaussian copula is given by

$$c_\tau(u_1,u_2)=\frac{1}{\sqrt{1-\tau^2}}\exp\left(\frac{2\tau\xi(u_1,u_2)-\tau^2\zeta(u_1,u_2)}{2\left(1-\tau^2\right)}\right)$$

$\xi(u_1,u_2)=\Phi^{-1}(u_1)\Phi^{-1}(u_2)$, $\zeta(u_1,u_2)=\Phi^{-1}(u_1)^2\Phi^{-1}(u_2)^2$

τ is Kendall's tau, $\tau=\frac{2}{\pi}\sin^{-1}(\rho)$ and Φ is distribution function of single variety standard normal distribution.

Farlie-Gumbel-Morgenstern (FGM) copula: For $\theta\in[-1,1]$

$$C(u_1,u_2)=u_1u_2+\theta u_1u_2(1-u_1)(1-u_2) \quad (7)$$

$iii)$ *Archimedean copula*: Archimedean copulas are most widely used in applications.

Let $\phi=[0,1]\to[0,\infty]$ is continuous and certainly decreasing $\phi(1)=0$ and $\phi(0)=\infty$.

$$C(u_1,u_2)=\phi^{-1}[\phi(u_1)+\phi(u_2)] \tag{8}$$

It is called the Archimedean copula. Where $\phi(.)$ is called the generator function of Archimedean copulas. The generator function solely characterizes the dependence structure of random variables, it described by model parameter θ for details [2].The relationship between $\phi(t)$ and Kendall tau (τ) is given by

$$\tau=1+4\int_0^1\frac{\phi(t)}{\phi'(t)}dt \tag{9}$$

Gumbel copula: Gumbel copula is very sensitive to the change of variables at the upper tail of the distributions. It can be capture upper tail dependence. Its generating function is $\phi(t)=(-\ln t)^{\theta}$ and $\theta\in[1,\infty[$

$$C_{\theta}^{Gum}(u_1,u_2)=\exp\left\{-\left[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}\right]^{\frac{1}{\theta}}\right\} \tag{10}$$

Clayton copula: Clayton copula is very sensitive to the change of variables at the lower tail of the distributions. It can be capture lower tail dependence. Its generating function is

$$\phi(t)=t^{-\theta}-1,\ \theta>0,\ \theta=2\tau/(1-\tau)$$

$$C_{\theta}^{cl}(u_1,u_2)=\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta} \tag{11}$$

If X_1 and X_2 are random variables which are normal distributed, their joint distribution function is given as

$$F_{X_1,X_2}(x_1,x_2)=\left\{\left[\Phi\left(\frac{x_1-\bar{X}_1}{\sigma_{X_1}}\right)\right]^{-\theta}+\left[\Phi\left(\frac{x_2-\bar{X}_2}{\sigma_{X_2}}\right)\right]^{-\theta}-1\right\}^{-1/\theta}$$

Frank copula: $\phi(t)=-\ln\left(e^{-t\theta}-1/e^{-\theta}-1\right)$, $\theta\in(-\infty,\infty)$

$$C_{\theta}^{Frank}(u_1,u_2)=-\frac{1}{\theta}\ln\left[1+\frac{\left(e^{-\theta u_1}-1\right)\left(e^{-\theta u_2}-1\right)}{e^{-\theta}-1}\right] \tag{12}$$

A12copula: $C_{\theta}^{A12}(u_1,u_2)=\left(1+\left[\left(u_1^{-1}-1\right)^{\theta}+\left(u_2^{-1}-1\right)^{\theta}\right]^{1/\theta}\right)^{-\theta}$

$\theta=2/(3-3\tau)$ and $\tau\in[0,3334,1]$

3. Dependence measures

Testing independence between two components of a random vector is an important problem in the empirical analysis. A random sample $(X_{11},X_{12}),\ldots,(X_{n1},X_{n2})$ is collected from the random variable pair (X_1,X_2). Then an estimator is obtained of the dependence parameter. Correlation is the most widely used measure for dependence between random variables. However, the correlation coefficient reflects the dependence structure only in very specific situations. Correlation fails to capture any nonlinear dependencies in a data set.

Linear correlation: The Pearson's linear correlation coefficient which is defined by $r_{X_1,X_2}=Cov(X_1,X_2)/\sigma_{X_1}\sigma_{X_2}$

Rank correlation: Many other popular empirical measures of dependence are based on ranks. The most known rank based measures of dependence are Spearman's rho and Kendall's tau.

Spearman's rho: This correlation measure is used ranks instead of values. $-1\le\rho_n\le1$ and Spearman's rho is not assume about distributions of random variables. Let $x_1,x_2,\ldots,x_n$ and $y_1,y_2,\ldots,y_n$ is observed values from random variables X and Y respectively. R_i Denotes the rank of x_i and T_i denotes the rank of y_i, $\left(\frac{R_i}{n+1},\frac{T_i}{n+1}\right), i=1,2,\ldots,n$, is rank pairs corresponds the observation pairs (x_i,y_i). Spearman's rho is described by

$$\rho_n=\frac{\sum_{i=1}^{n}(R_i-\bar{R})(T_i-\bar{T})}{\sqrt{\sum_{i=1}^{n}(R_i-\bar{R})^2}\sqrt{\sum_{i=1}^{n}(T_i-\bar{T})^2}} \tag{13}$$

$$\bar{R}=\frac{1}{n}\sum_{i=1}^{n}R_i=\frac{n+1}{2}=\bar{T}=\frac{1}{n}\sum_{i=1}^{n}T_i$$

An alternative form

$$\rho_n=\frac{12}{n(n+1)(n-1)}\sum_{i=1}^{n}R_iT_i-3\frac{n+1}{n-1} \tag{14}$$

Empirical copula: Daheuvels was described the tests of interdependence based on empirical copula C_n as follows [15]

$$C_n(u_1,u_2)=\frac{1}{n}\sum_{i=1}^{n}I\left\{\frac{R_i}{n+1}\le u_1,\frac{T_i}{n+1}\le u_2\right\} \tag{15}$$

Where, $u_1,u_2\in[0,1]$ and $C_n\to C$ for $n\to\infty$,

$$\rho=12\int_{[0,1]^2}u_1u_2\,dC_n(u_1,u_2)-3=\frac{n-1}{n+1}\rho_n$$

$\alpha=\%5$ If $\sqrt{n-1}|\rho_n|>Z_{\alpha/2}$, $(Z_{0.05/2}=1.96)$ then random variables X and Y are called independent [10].

Kendall's tau: Kendall's tau τ is a measure of correlation between two ranked variables. It is difference between the probability that the observed data are in the same order versus the probability that the observed data are no in same order.

For the vector (X_1,X_2), (R_i,T_i) and (R_j,T_j) being concordant if $(R_i-R_j)(T_i-T_j)>0$ and discordant otherwise.

$$\tau(X_1,X_2)=P\left((X_1-X_1')(X_2-X_2')>0\right)-P\left((X_1-X_1')(X_2-X_2')<0\right)$$

Kendall's tau is defined by

$$\tau_n=\frac{2}{n(n-1)}(C_n-D_n)$$

Where, C_n is the number of is concordant pairs of rank and D_n is the number of discordant pairs. Given a sample $(X_1,X_2,\ldots,X_n)$ where $X_i=(X_{i,1},X_{i,2})$. The sample estimation of Kendall's tau is as,

$$\hat{\tau}=\binom{n}{2}^{-1}\sum_{j<k}sign\left[(X_{j,1}-X_{k,1})(X_{j,2}-X_{k,2})\right] \tag{16}$$

Where, $sign(x)$ is equal to 0 for $x=0$ and $x/|x|$ for $x\neq0$. Kendall's tau is invariant to strictly increasing transformations. Let $u_i=R_i/n+1, v_i=T_i/n+1$ be normalized ranks. To calculate the τ_n drawing a line that is connecting (u_1,v_1) and

(u_2, v_2) points. If this line has the positive slope then it takes the value 1 and has a negative slope it takes value-1. This procedure is applied to all (u_m, v_m) and (u_s, v_s) .The number of such line is $n(n+1)/2$. All of the -1 and 1 is sum and this sum is divided by $n(n+1)/2$ is obtained τ_n .Where n is sample size. To explain the empirical relationship between τ_n and C_n, we define a function as [17], $L_{ij} = I\{x_j \le x_i \, ve \, y_j \le y_i\}$ for $i = 1,2,...,n$

$$L_{11} = 1$$

$$M_n = \frac{1}{n}\sum_{i=1}\sum_{i\neq j}\left(L_{ij} + L_{ji}\right) = -n + \sum_{i=1}^{n}\sum_{j=1}^{n} L_{ij}$$

$$H_i = \frac{1}{n}\sum_{j=1}^{n} L_{ij} = \frac{1}{n} \equiv \{j : x_j \le x_i \, ve \, y_j \le y_i\}$$

$$\overline{H} = \frac{1}{n}(H_1 + ... + H_n), \; M_n = -n + n^2\,\overline{H}$$

$$\tau_n = 4\frac{n}{n-1}\overline{H} - \frac{n+3}{n-1} \tag{17}$$

$$\tau = -1 + 4\int\int C_n(u,v)\,dC_n(u,v)$$

$$\tau = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}\operatorname{sgn}(x_i - x_j)(y_i - y_j)$$

Copula function $C(u_1, u_2)$ is related to the Kendall's tau as

$$\tau = 4\iint_{[0,1]^2} C(u_1,u_2)\,dC(u_1,u_2) - 1$$

Fitting copulas: The building the joint distribution of vectors X are sensitive to $C(u_1, u_2, ..., u_n)$ and the marginal distributions $\{F_i\}_{i=1}^{n}$. In generally two methods used to calibrate the copula to sample data of vector X [9].

i) Parametric method: This method includes two steps, in first step, the parametric families are postulate for copula and marginal distributions, in second step the unknown parameters are estimated. Then copula based joint distribution is obtained by $F(x_1, x_2, ..., x_d) = C_{\hat{\theta}_o}\left(F_1(x_1;\hat{\theta}_1), ..., F_d(x_d;\hat{\theta}_d)\right)$

ii) Semi parametric method: This method includes two steps. First $C_\theta(u_1, u_2, ..., u_d)$ is postulated, second empirical marginal distributions $\hat{F}_i(\xi) = \frac{1}{n+1}\sum_{s=1}^{n} 1\left(x_i^{(s)} \le \xi\right), \; i = 1,2,...,d$ are determined each of observed values rank value $rank_i^{(k)} = \sum_{s=1}^{n} 1\left(x_i^s \le x_i^k\right)$ is calculated. Then this rank is sort in ascending order.

Testing of Copulas: The univariate distribution function $K_c(t)$ is defined as follows [11].

$$K_c(t) = t - \frac{\varphi(t)}{\varphi'(t)} \tag{18}$$

i) Let $(x_i, y_i)_{1\le i\le n}$ is observed values, the distribution functions $F_1(x_i)$ and $F_2(y_i)$ values are obtained.

ii) Let $t_i = C_\theta(F_1(x_i), F_2(y_i))$

iii) t_i is putting in $K_c(t)$ function

iv) Equality $K_c(t) \sim N(0,1)$ is tested

3.1. Tail dependence

The correlation coefficient measures the co movements in central part of two random variables. Whereas we want to know that this two variable how is together behave in tails. Tail dependence can be asymmetric. Tail dependence is probability of large (small) values of one random variable co occurring with large (small) values of other random variable. In this section we interested probability of that given the one of the variables takes an extreme value then other variable taking an extreme value. We will study the lover and upper tail dependences.

Definition. Let $X = (X_1, X_2)$ be a two dimensional random vector with marginal distributions F_1 and F_2 .The coefficient of upper tail dependence of X is defined by

$$\lambda_u = \lim_{q\to 1} P\left(X > F_1^{-1}(q) \middle| Y > F_2^{-1}(q)\right) \tag{19}$$

Definition. The coefficient of lower tail dependence of X is defined as,

$$\lambda_l = \lim_{q\to 0} P\left(X_1 \le F_1^{-1}(q) \middle| X_2 < F_2^{-1}(q)\right) \tag{20}$$

If $\lambda_u, \in [0,1]$ ($\lambda_u, \in [0,1]$), we say that X has upper (lower) tail dependence, if $\lambda_u = 0$ ($\lambda_l = 0$) we say that X is upper (lower) independent. The coefficients of upper and lower tail dependences can be write in copula terms as follows,

$$\lambda_u = \lim_{q\uparrow 1} \frac{1 - 2q + C(q,q)}{1-q}$$

$$\lambda_l = \lim_{q\downarrow 0} \frac{P\left(X_2 \le F_1^{-1}(q), X_1 \le F_1^{-1}(q)\right)}{P\left(X_1 \le F_1^{-1}(q)\right)} = \lim_{q\downarrow 0} \frac{C(q,q)}{q} \tag{21}$$

Table 1: Lower and upper tails

Copula	λ_u	λ_l	Parameter
Clayton	0	$2^{-1/\theta}$	$\theta > 0$
Gumbel	$2 - 2^{1/\theta}$	0	$\theta > 1$
Frank	0	0	
Gaussian	0	0	$-1 \le \rho \le 1$
A12	$2 - 2^{1/\theta}$	$2^{-1/\theta}$	$\theta > 1$

Remark: The Gumbel copula is able to model the upper tail dependence, whereas the Clayton copula can model lower tail dependence so we can use a combination both of them as a mixed copula. Estimating the copula parameter θ : In order to estimate the dependence parameter θ we need the estimated Kendall's tau τ and Pearson correlation coefficient r .

Table 2: The relationship the copula and dependence parameter

	Gaussian	Gumbel	Clayton	Frank
τ	$(2/\pi)\arcsin(\theta)$	$\frac{\theta-1}{\theta}$	$\frac{\theta}{\theta+2}$	$1 - \frac{1}{9}[1 - D_1(\theta)]$
ρ	$(2/\pi)\arcsin(\rho/$	No closed form	No closed form	$1 - \frac{12}{\theta}[D_1(\theta) - D_2(\theta)]$
Debye function: $D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^t - 1}dt$				

Copula Matching: The marginal distribution is determining. A proper copula and its fitting method selected. Dependence struc-

ture is representing the copula function. Each pair of observation is transformed into its rank based representation.

$$(rank_1(X_1) < rank_2(X_1) <,...,< rank_n(X_1))$$

$$(rank_1(X_2) < rank_2(X_2) <,...,< rank_n(X_2)), i = 1,2,...,n$$

$$U_{1,i} = \frac{rank_i(X_1)}{n+1}, U_{2,i} = \frac{rank_i(X_2)}{n+1} \quad (22)$$

In order to match the selected copula to experimental data, the joint histogram of U_1 and U_2 variables is comparing to the histogram of selected copula.

4. Modeling of Extreme Dependence

Extreme value is defined as a value that $x-u$ excesses above some suitably high threshold u is observed. Disadvantage of the method is that if the threshold u is small, then the sample includes the observations which in truth are not extreme. If the selected threshold u is large then some real extreme values is not included in sample. The extreme events occur generally in tails of distribution. The behaviour of a variable that takes values grater than a high threshold to model generally is used Peak Over Threshold (POT) method. Let $x_1, x_2, ..., x_n$ are taken from iid random variables a random sample with an unknown distribution function F. The conditional distribution of excess values $x_k - u$ over high threshold u is described by

$$F_u(y) = P(X - u \le y \mid X > u) = \frac{F(u+y) - F(u)}{1 - F(u)} \quad (23)$$

Where $0 \le y \le x_F - u$

Theorem (Pickands (1975), Belkema and de Haan (1974)). For a large class of the underlying distribution F the excess distribution function F_u can be approximated by Generalized Pareto Distribution (GPD) for increasing threshold u [18],

$$F_u(y) \approx G_{\xi,\beta,\mu}(y), u \to \infty$$

Generalized Pareto Distribution (GPD) is given by

$$G_{\xi,\beta,\mu}(x) = \begin{cases} 1 - (1 + \xi((x-\mu)/\beta))^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp(-(x-\mu)/\beta), & \xi = 0 \end{cases} \quad (24)$$

The shape parameter ξ is independent of the threshold u whereas the scale parameter $\beta = \beta(u)$. When the scale parameter, $\xi \ge 0$, $\beta > 0$ and $x > 0$, when parameter $\xi < 0$, $x \le 0$ [1]. In applications, generally the parameter μ is taken as $\mu = 0$.

Parameter estimation: There are many possible approaches for parameter estimation in extreme value analysis. The popular schemes, Maximum likelihood (ML), the method of moments (MOM) and probability weighted moments (PWM)

Moment's method: Let $(x_1, x_2, ..., x_n)$ be a sample of independent observations. The moment estimators of parameters is given

$$\hat{\xi}_{mom} = \frac{1}{2}\left(\frac{(\bar{x})^2}{s^2} - 1\right), \quad \hat{\beta}_{mom} = \frac{1}{2}(\bar{x})\left(\frac{(\bar{x})^2}{s^2} + 1\right)$$

Where $\bar{x}$ and s^2 are sample mean and sample variance.

Probability Weighted Moment Estimators (PWM):

$$\hat{\xi}_{PWM} = \frac{\bar{x}}{(\bar{x} - 2t)} - 2 \quad , \quad \hat{\beta}_{PWM} = \frac{2\bar{x}t}{(\bar{x} - 2t)}$$

Where $t = \frac{1}{n}\sum_{i=1}^{n}(1 - p_{i;n})x_{i,n}$, $p_{i,n} = \frac{1 - 0.35}{n}$ and $x_{i,n}$ is i th order statistics of a sample with size n [12].

Asymptotic Tail Dependence: Let $X = (X_1, X_2)$ be two dimensional random vectors $(x_1^{(k)}, x_2^{(k)})_{k=1}^{n}$. We concerns the tendency that extremely small or extremely large outcomes of X_1 and X_2 occur simultaneously. The widely used and intuitive scalar measures for these tendencies are provided by the coefficients of tail dependence (TDC). To measure between random variables, we will use χ and $\bar{\chi}$.

Now we describe the following conditional distribution,

$$P(U_2 > u_2 | U_1 > u_1) = 2 - \frac{1 - P(U_1 < u_1, U_2 < u_2)}{1 - P(U_1 < u_1)}, u \in [0,1]$$

Quantile-based dependence measure $\chi(u)$ is described as

$$\chi(u) = 2 - \frac{\log C(u,u)}{\log u} \quad (25)$$

When one of the variables is known to be extreme, probability of being extreme in the other variable is described as $\chi = \lim_{u \to 1} \chi(u)$.

Similarly to show other dependence, we will use the notation $\bar{\chi}$

$$\bar{\chi}(u) = 2\frac{\log P(u > U_1)}{\log P(U_1 > u, U_2 > v)} - 1 = 2\frac{\log(1-u)}{\log \bar{C}(u,u)} - 1 \quad , u, v \in [0,1]$$

$$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u) \quad (26)$$

If $\chi \in [-1,1]$ and $\bar{\chi} < 0$ then there is a negative dependence, if $\bar{\chi} = 0$ there are exact independent and if $\bar{\chi} > 0$ there are positive dependence. To calculate the $\bar{\chi}$ empirically, we will use a method that is proposed by [13].Firstly the return pairs (X, Y) are transformed to unit Frechet marginals as,

$$S = -1/\log F_X(X) \quad , \quad T = -1/\log F_Y(Y) \quad (27)$$

Let u be a high threshold value. We define the new variable $Z = \min(S, T)$ and consider the values $Z > u$,

$$\hat{\bar{\chi}} = \frac{2}{n_u}\left(\sum_{j=1}^{n_u} \log\left(\frac{z_j}{u}\right)\right) - 1, \; Var(\hat{\bar{\chi}}) = \frac{1}{n_u}(\hat{\bar{\chi}} + 1)^2 \quad (28)$$

Null hypothesis for this test statistics

H_0 : Extreme values are dependent asymptotically (i.e. $\bar{\chi} = 1$)

If $\hat{\bar{\chi}} + 1.96\sqrt{Var(\hat{\bar{\chi}})} < 1$ then, null hypothesis H_0 is rejected. In this case, it is found that the variables being studied are asymptotically independent. If $\hat{\bar{\chi}} + 1.96\sqrt{Var(\hat{\bar{\chi}})} \ge 1$ then null hypothesis H_0 is not rejected in this case under the assumption that $\bar{\chi} = 1$ the variable χ and its variance are estimated as

$$\hat{\chi} = \frac{u n_u}{n} \quad , \quad Var(\hat{\chi}) = \frac{u^2 n_u (n - n_u)}{n}$$

Table 3: Summarized asymptotic dependence

	Independent	Dependent
χ	0	$]0,1]$
$\bar{\chi}$	$[-1,1]$	1

Threshold selection: The statistical estimation methods of tails and quantiles are very sensitive to substantial changes in the extreme data. A graphical tool for threshold u selection is based on mean excess function as follows $e(u)=E[X-u \mid X>u]$, $u>0$.

In this study, we use an approach that is proposed by [14]. Let $X_1, X_2, \ldots, X_n$ is a random sample that it is taken from a distribution with fat tails. Firstly the observed values are ordered as $X_1 \le X_2 \le \ldots \le X_n$ then

$$u_k^* = \frac{1}{H_{k,n}}\left(\frac{1}{n}\sum_{j=1}^{n} N_{(k-j+1)}.M_j\right) \qquad (30)$$

$$M_j = j.\left(\log X_{(n-j+1)} - \log X_{(n-j)}\right) \quad , j=1,2,\ldots,k$$

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} M_j \quad , \quad N_{(k-j+1)} = 1-\log\left(\frac{j+1}{k+1}\right)$$

We calculate the u_k^* values for $k \in \{3,4,\ldots,n\}$ afterwards we plot of the points $\{(k\ ,\ u_k^*): k=3,4,\ldots,n\} \in R^2$ .The plot should show an approximately linear behavior, when the underlying data follow a GPD. To choose an appropriate threshold for the POT method we search for the smallest k where the plot is approximately linear.

5. Application

In this section, we conduct an empirical study to assertion the performance of our models. We analyse the dependence relationship between oil spot price in Los Angeles and spot price Fuel oil for New York. We use daily data from 02.06.1996 to 29.04.2008. Firstly we research the statistical characteristics of these two energy assets then we put these findings to modelling process.

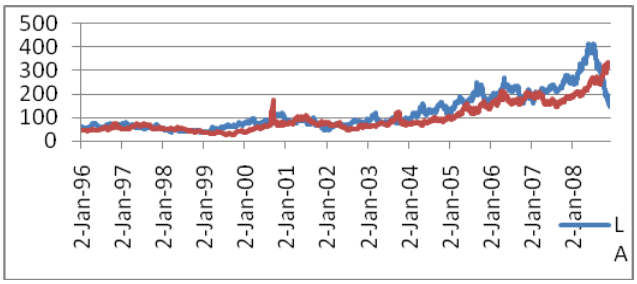


Fig.1: Los Angeles (LA) oil spot price and New York (NY) Fuel oil spot prices

At first sight, the development of the two energy asset spot prices seems to be almost identical, i.e., one could think of an almost perfect correlation. In this study, we research that question whether the joint distribution has a (strong) tendency to generate extreme values simultaneously. Sampling procedure is summarized as (a) standard uniform random variables U_i are simulated (b)inverse cumulative distribution function is applied to each uniform random variable $F_{X_i}^{-1}(U_i)$ .Standardized ranks to eliminate the effects on modelling of marginal distributions we study with the ranks of the observed values instead of original values. Let R_1 denotes the rank of the first series, and R_2 denotes the rank of second series.

- $U_1(t)=\dfrac{(R_1(t)-1/2)}{n}$ and $U_2(t)=\dfrac{(R_2-1/2)}{n}$
- The plot is drawn of pear $(U_1(t), U_2(t))$, $t=1,2,\ldots,n$

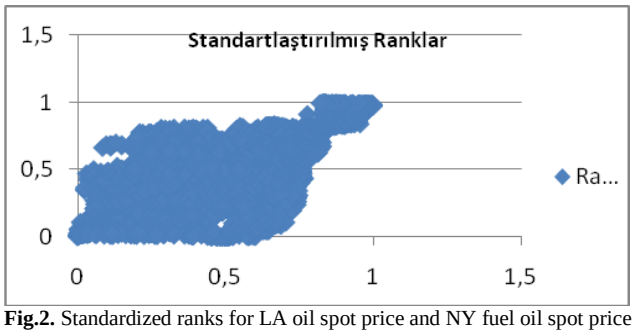


Fig.2. Standardized ranks for LA oil spot price and NY fuel oil spot price

If X_1 and X_2 are independent random variables then the random variable U_1 and U_2 also are independent. In this case, the scattering plot is uniformly distributed on unit square. As can seeing in the above plot there are a strongly correlation between random variables X_1 and X_2 .The power of correlation is measured by closeness to main diagonal.

Table 4: Correlation coefficients that is estimated from observed data

	r	τ	ρ
Corr.	0.864	0.668	0.864

To give a brief insight into the properties of the data, Table 4 summarize the basic descriptive statistics for LA spot oil price and NY fuel oil spot price the sample period.

Table 5: The descriptive statistics for LA spot oil price and NY fuel oil spot price

	n	Range	Mean	Var.	Std.dev	Skew	Kurt
LA	2084	0.335	2.56E-4	6.38E-4	0.025	-0.096	3.69
NY	2084	0.69	7.26E-4	9.18E-4	0.030	-1.993	43.4

Table 6: The estimations of dependence parameter θ for different copulas

Copulas	Kendall's tau and θ relationship	θ
Clayton	$\theta=2\tau/(1-\tau)$	4.0241
Gumbel	$\theta=1/(1-\tau)$	3.0120
Normal	$\theta=\sin(\pi\tau/2)$	0.8671

Table 7: Tail parameter estimations

Copula	λ_u	λ_l
Clayton	0	0.8415
Gumbel	0.7412	0
Frank	0	0
Gaussian	0	0
t-copula	$\lambda_u = \lambda_l = 0.6958, v = 12$	

Let U_1 and U_2 are uniform distributed random variables Clayton Copula,

$$\hat{C}_\theta^{cl}(u_1,u_2) = \left(u_1^{-4.0241} + u_2^{-4.0241} - 1\right)^{-\frac{1}{4.0241}}$$

Gumbel Copula

$$\hat{C}_\theta^{Gum}(u_1,u_2) = \left\{-\left[(-\ln(u_1))^{3.0120} + (-\ln(u_2))^{3.0120}\right]^{\frac{1}{3.0120}}\right\}$$

$$S = (\phi')^{-1}\left(\frac{\phi'(U_1)}{U_2}\right), \qquad T = (\phi')^{-1}[\phi(S) - \phi(U_1)]$$

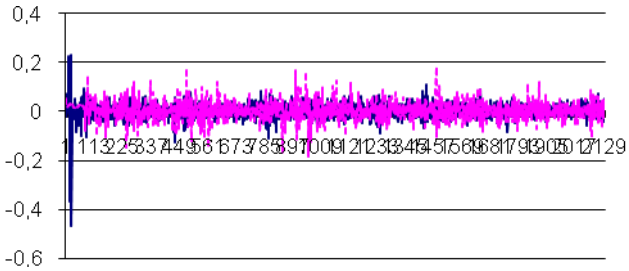
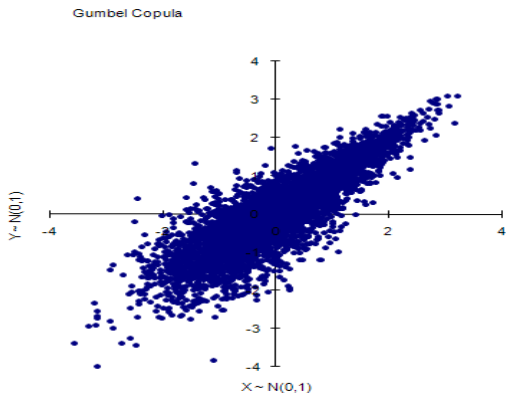
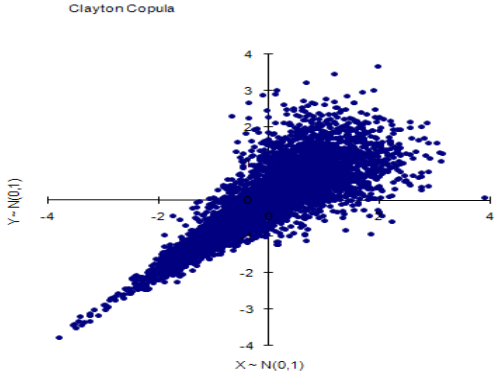
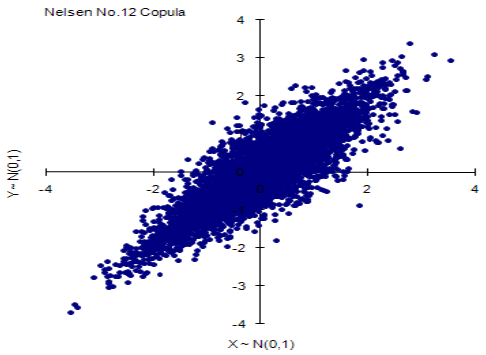


Fig. 3: The volatility for LA oil return and NY fuel oil return series







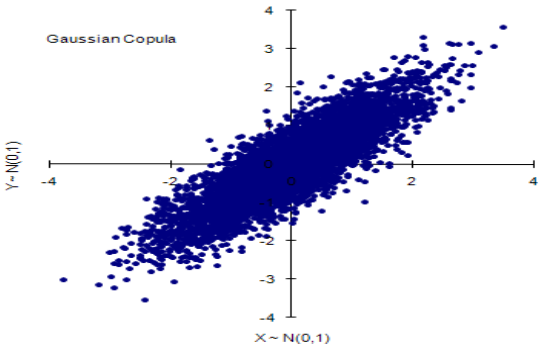


Fig. 4: 5000 simulation for different copulas

Analysis of extreme values: To obtain a cumulative distribution function for LA oil spot price time series, we are drawn the random sample $(x_1, x_2, ..., x_{2100})$ from BURR distribution with $k = 3$, $\alpha = 2$, $\beta = 1$, $\gamma = 0$ parameters. Firstly calculated the cumulative distribution values $(F_1(x_1), F_2(x_2), ..., F_1(x_{2100}))$ and then $S = \left(\frac{-1}{\log(F_1(x_1))}, \frac{-1}{\log(F_1(x_2))}, ..., \frac{-1}{\log(F_1(x_{2100}))}\right)$

Similarly for NY fuel spot price we are drawn a random sample $(y_1, y_2, ..., y_{2100})$ from DAGUM distribution. The calculated cumulative distribution values $(F_2(y_1), F_2(y_2), ..., F_2(y_{2100}))$

$$T = \left(\frac{-1}{\log(F_2(y_1))}, \frac{-1}{\log(F_2(y_2))}, ..., \frac{-1}{\log(F_2(y_{2100}))}\right)$$

A new series is derived from these series as $Z_i = \min(S_i, T_i)$, $i = 1,2,...,2100$ .These series Z is ascending ordered as $Z_{(1)} < Z_{(2)} < ... < Z_{(2100)}$ The term 75%th of the ordered series Z was chosen as the threshold value, i.e. $u = 1.492524$ The number of values exceeding this threshold is $n_u = 531$. Using the values exceeding threshold is calculated following statistics $\hat{\bar{\chi}} = 0.25249$, $Var(\hat{\bar{\chi}}) = 0.0029543$

H_0 : The series are asymptotically dependence $(\bar{\chi} = 1)$

$\hat{\bar{\chi}} + 1.96\sqrt{Var(\hat{\bar{\chi}})} = 1.25249 + 1.96\sqrt{0.0029543} = 0.35904 < 1$

Finally, calculated the test statistics $\hat{\bar{\chi}} = 0.25249$ is a meaningful way from $(\bar{\chi} = 1)$. Thus we can say that the series are asymptotical independent.

6. Conclusion

In this paper, a general dependence model which is based on copulas and extreme value theory is studied to explore the relationship between extreme values of two random variables as a case, LA oil spot price and NY fuel oil spot price. Instead of real data, the standardized ranks are used to remove the effect of marginal distributions. The marginal distributions are fitted to LA and NY time series, as a result, the best fitting distribution for LA oil spot prices time series is Burr distribution and NY fuel oil spot prices time series is Log-logistic distribution. In these series, we found the tail dependence in these series. When, the standardized ranks are analyzed find that both of the series has more values in the lower tail. On the base of these finding, different copulas model are simulated, subsequently the scattering plots of simulated values and observed values are compared. Finally it is found that the Gumbel copula can be capturing the extreme tail dependence between LA oil spot price and NY fuel oil spot price time series. Finally the findings of study show advantages the extreme value copula approach for understanding of extreme tail dependence. The copulas provide a power method of analysing the dependence structure of two or more random variables. The copula function appears in any multivariate distributions as a structure that allows separating the marginal distributions and the dependence model.

References

[1] Embrechts P, Kluppelberg C& Mikosch T(1997),Modeling extremal events: for insurance and finance, *Springer Verlag*, Berlin.

[2] Nelsen RB (2006), an introduction to copulas, 2nd ed., *Springer-Verlag*, New York.

[3] Zhang L& and Singh VP(2007), Gumbel- Hougaard copula for trivariate rainfall frequency analysis, *J. Hydrol. Eng*, 12,pp.409-419.

[4] Trivedi PK& Zimme DM (2005), Copula modeling: an introduction for practitioners, *Foundations and Trends in Econometrics*,Vol.1,no 1, pp.1-111.

[5] Joe J (1997), Multivariate models and dependence concepts, *Chapman& Hall*, London.

[6] Frees EW &Valdez EA (1998), Understanding relationships using copulas, *North American Actuarial Journal*, 2(1), pp.1-26.

[7] Cherubini U,Luciano E & Vecchiato W(2004), Copula methods in finance, *Wiley*, New York.

[8] Mikosch T(2006),Copulas: tales and facts, *Extremes*,9(3),pp.20,.

[9] Grigoriu M (2016),Do seismic intensity measures(IMs) measure up?, *Probability Engineering Mechanics*, 46,pp.80-93.

[10] Sun GH(2015),Tail dependence study of SSE composite index and SZSE component index based on the copula, *Applied Mathematics*, 4, pp.1065-1069.

[11] Halder C &Das K(2016), Understanding extreme stock trading volume by generalized Pareto distribution, *North Caroline Journal of Mathematics and Statistics* ,vol.2,pp.45-60.

[12] Poon S, Rockinger M&Town J (2004), Extreme value dependence in financial markets: diagnostics, models and financial implications, *Review of Financial Studies*,17,pp.581-610.

[13] Beirlant J& Goegebeur Y (2006),A goodness of fit statistics for Pareto-type behavior, *Journal of Computational and Applied Mathematics*,186,pp.99-116.

[14] Daheuvels P(1981), Nonparametric test of independence, *In Analytical methods* 861 of Lecture Notes in Mathematics, Springer,Berlin,pp.42-50.

[15] Sklar A (1981), Functitions de repartition a n dimensions et leurs merges, *Publ. Inst.Statist.Univ.Paris*, 8, pp.229-23.

[16] Genest C& Favre AC (2007), everything you always wanted to know about copula modeling but were afraid to ask, *Journal of Hydrologic Engineering*, 12(4), pp.347-368.

[17] McNeil AJ, Frey R& Embrechts P (2005), Quantitative risk management: concepts, techniques, tools, *Princeton University Press*, Princeton.