International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 3 Issue 10, October – 2019, Pages: 9-15

Comparison of Financial Data Analysis with Respect to Copulas and Linear Correlation Coefficient

Ahmed AL-Adilee[1]*, Ola Hasan AL-Lahaibat[2]

[1]Department of Mathematics, Faculty of Education, University of Kufa, Najaf, Iraq
[2]Department of Mathematics, Faculty of CS & mathematics, University of Kufa, Najaf, Iraq

Abstract: *This study aims to present comparison between the classical correlation coefficient and correlation via copula functions. We propose a model of financial data analysis of Iraq index, S&P and Dow jones index in order to demonstrate the weakness points of Iraq index with is comparing to the two worldwide index (S&P or Dow jones). We show several different tables of the results that we have obtained and figures that explain some different relations of the examine data.*

Keywords: Dependencies, Distribution Function, Copulas, Financial Concepts.

1. Introduction

This paper contains two main parts. The first one concerned with the calculations of linear correlation, while the second one deals with the determinations of linear and non-linear correlation via copulas. In general, statistical inference plays an essential role in data analysis, and especially in financial data. Among many statistical tools, correlation coefficient has the major role that is used to identify and calculate the direction and strength between the examined random variable, see [1].
What it is necessary to be mentioned in this paragraph that for elliptical case of data, linear correlation would be valid, elsewhere it would be failed, see [2, 3].
On the other hand, copula function notion is not so old and still fresh idea that has many interesting properties and plays an important role in modern statistics, see [4]. These functions are used to measure the dependence structure between variables with deepest interpretation. Also, it differs from correlation is that it is more general than the concept of standard linear correlation.
Also, it's enable use of calculating nonlinear relationships, see [4].

2. Preliminaries

In this section, we present some basic concepts related to the notion of linear correlation coefficient and nonlinear correlation via copulas.

2.1 Classical Dependencies

Statistical inference is one of the most classic and important approaches used to study these subjects. Whatever, statistical inference is needed for each random experiment, and this can be done by correlation. Indeed, correlation is a measure of dependence between two or more variables. This means that the strength and direction of two random variables can be assessed by correlation. What should be mentioned here is that correlation is only a linear relationship between random variables, see [1].

Definition 2.1 The correlation coefficient is a statistical measure used to assess the strength and direction of a relationship between two or multivariate random variable, see [5].

Indeed, there are three well-known classical dependencies that are widely used in the classical data analysis and they related to each other.

2.1.1 Pearson Correlation Coefficient

It is one of the most common measures of association that are utilized to measure the dependence structure between random variables when they are linearly correlated. The distribution of two random variables is assumed to follow the normal distribution. Its formula defined by the following way, and it is denoted by ν, see [5].

$$\nu = \frac{Cov\ (S,T)}{\sqrt{(Var\ S)\ (Var\ T)}} \qquad (1)$$

Where,

$$\mathrm{Cov(S,T) = E(ST) - E(S)E(T)} \qquad (2)$$

$$Var(S) = E(S^2) - E(S)^2 \qquad (3)$$

$$\mathrm{E(S)} = \int \mathrm{f(s)\ ds}$$

Remark 2.1.1 We note that the strongest correlation coefficient is the closest value to $+1$ or -1, and the weakest one is the one that is closest to zero.

Remark 2.1.2 According to the correlation coefficient formula, ν belongs to the interval $[1, -1]$. Indeed, it is clear that when S and T are independent, so $\mathrm{Cov(S,T)} = 0$, implies that $\nu = 0$. Of course, the converse is not true, which means that we could obtain $\nu = 0$, but this does not mean the random variables are independent, see [1].

2.1.2 Rank Correlations

1- **Spearman Correlation Coefficient**: it is also called rank correlation and it is used to evaluate the correlation value between qualitative data by giving a rank for each value of the random variable. A Spearman correlation is denoted by ρ, see [4].

$$\rho = 3[\Pr\big((S_1 - S_2)(T_1 - T_2) > 0\big) - \Pr\big((S_1 - S_2)(T_1 - T_2) < 0\big)]$$

To calculate the Spearman coefficient of the two random variable, we firstly need to arranges the examined data in ascending or descending order for both random variables. When two or more values of the random variables are equal, we assign each different rank value (as if the values are not equal) and then calculate the average of these ranks, and this is given to each of these equal value, and has the following form, see [6].

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \qquad (4)$$

Where,

- d is the difference between the rank.
- n is the sample size

Note that, Spearman rank correlation is also used to measure the degree of association between two variables, and type of relationship. Wherefore it is like Pearson correlation.

2- **Kendall's Tau Coefficient**: it is one of the most powerful tools that is used to determine the correlation between different random variables. It measures the agreement between two separate sets of ranks according to the total changes. The Kendall's correlation coefficient for the rank requires the same conditions to the Spearman correlation coefficient. It is implemented to calculate the correlation coefficient between two random variables expressed in the ranking scale, see [3, 6].

Before we mention the formula of this dependence, we must first mention some properties of data consisting of n pairs of pairs of values (S, T). The first pair $(s_1,\ t_1)$ and the second pair (s_2, t_2), are called concordant $Pr(S \geq s, T \geq t)$, if both s_1 and t_1 greater than s_2 and t_2. While the pairs (s_1, t_1) and (s_2, t_2) are called discordant when $Pr(S \geq s, T \leq t)$,. According to this explanation, a Kendall's tau correlation, denoted by ρ, can be written by the following formal, see [4].

$$\rho = [\Pr\big((S_1 - S_1)(T_1 - T_1) > 0\big) - \Pr\big((S_1 - S_1)(T_1 - T_1) < 0\big)]$$

To calculate Kendall's tau coefficient of two random variables, we use the following formals

$$\rho = \frac{2(N_c - N_d)}{n(n-1)} \qquad (5)$$

Where,

- n is the total number of pairs.
- $N_c$the number of concordant views.
- N_d the number of discordant view.

2.2 Dependencies Via Copulas

The second important part of our measures is associated with the notion of copula. Each copula can be used to describe the dependence structure whether the random variables are linearly or non-linearly correlation, see [2].

It is a well-known concept in nonparametric statistics to concentrate on the ranks of given data rather than on the data itself. This led to important correlation estimators, Spearman's rho and Kendall's tau, which we present directly in a form related to copulas.

The corresponding correlation of Spearman rho with respect to copula can be expressed by the following, see [4].

$$\rho_c = 12\int_0^1\int_0^1 \mathrm{pq}\ \mathrm{dC(p,q)} - 3 \tag{6}$$

where

$$dC(p,q) = \frac{\partial^2 C}{\partial p \partial q} dpdq \tag{7}$$

From equation (6), ρ_c has an equivalent formula that is

$$\rho_c = 12\int_0^1\int_0^1 [\mathrm{C(p,q)} - \mathrm{pq}]\mathrm{dpdq} \tag{8}$$

and

$$\rho_c = \frac{6}{\pi} sin^{-1}(\frac{\nu}{2}) \tag{9}$$

for the Gaussian copula.

These various types of representations in favor of copula concepts as with Spearman's rho. Also, the measure of association that is known as Kendall's tau is based on concordance end through their copulas, see [4].

$$\tau_c = 4\int_0^1\int_0^1 \mathrm{C(p,q)}\ \mathrm{dC(p,q)} - 1 \tag{10}$$

Where $dC(p,q)$ as explained in equation (7), and for the Gaussian copula we have

$$\tau_c = \frac{2}{\pi} sin^{-1}(\nu) \tag{11}$$

Simple calculations over the formulas in (6) and (10) yield new results that depends on the assigned copula.

We apply the formulas above with respect to Farlie - Gumbel- Morgenstern family to calculate the dependencies via copula, see [4].

Example 2.2.1 Let C_θ be a member of the Farlie - Gumbel- Morgenstern family of copulas, where θ is in $[-1,1]$. Then for all $p,q \in [0,1]$,

$$C(p,q) = pq + \theta(p - p^2)(q - q^2) \tag{12}$$

To determine the formula in (6), it is necessary to find $dC(p,q)$. Thus, we obtain that

$$dC(p,q) = \frac{\partial^2 C}{\partial p \partial q} dpdq = 1 + \theta(1-2p)(1-2q)dpdq \qquad (13)$$

Substitute $dC(p,q)$ in the general form for Spearman rho coefficient in equation (6), we obtain that

$$\rho_c = 12\int_0^1\int_0^1 \mathrm{pq}[\,1 + \theta(1-2p)(1-2q)]dpdq - 3$$

$$= 12\int_0^1\int_0^1 [\mathrm{pq}\ + \theta(p-2p^2)(q-2q^2)]dpdq - 3$$

$$= 12\int_0^1 [\frac{pq^2}{2} + \theta(p-2p^2)(\frac{q^2}{2} - \frac{2q^3}{3})\Big|_0^1]dp - 3 = 12\int_0^1 \left[\frac{p}{2} + \theta(p-2p^2)\left(\frac{-1}{6}\right)\right]dp - 3$$

$$= 12[\frac{p^2}{4} + (\frac{-\theta}{6})(\left(\frac{p^2}{2} - \frac{2p^3}{3}\right)\Big|_0^1] - 3 = 12\left[\frac{1}{4} + \frac{\theta}{36}\right] - 3 = \frac{\theta}{3}$$

While the Kendall's tau coefficient with respect to Farlie- Gumbel- Morgenstern family yields the following.

$$\tau_c = 4\int_0^1\int_0^1 (pq + \theta(p-p^2)(q-q^2)\,)(\,1 + \theta(1-2p)(1-2q))]dpdq - 1$$

$$= 4\int_0^1\int_0^1 [pq + \theta(p-2p^2)(q-2q^2) + \theta(p-p^2)(q-q^2) + \theta^2(p-p^2)(q-q^2)(1-2p)(1-2q)\,]\,dpdq - 1$$

$$= 4\int_0^1\int_0^1 [pq + \theta(p-2p^2)(q-2q^2) + \theta(p-p^2)(q-q^2) + \theta^2(p-2p^2-p^2+2p^3)(q-2q^2-q^2+2q^3)]\,dpdq - 1$$

$$= 4\int_0^1\int_0^1 [pq + \theta(p-2p^2)(q-2q^2) + \theta(p-p^2)(q-q^2) + \theta^2(p-3p^2+2p^3)(q-3q^2+2q^3)]\,dpdq - 1$$

$$= 4\int_0^1\int_0^1 \left[\frac{pq^2}{2} + \theta(p-2p^2)\left(\frac{q^2}{2} - \frac{2q^3}{3}\right) + \theta(p-p^2)\left(\frac{q^2}{2} - \frac{q^3}{3}\right) + \theta^2(p-3p^2+2p^3)\left(\frac{q^2}{2} - \frac{3q^3}{3} + \frac{2q^4}{4}\right)\right]\Big|_0^1 dp - 1$$

$$= 4\int_0^1\int_0^1 [\frac{p}{2} + \theta(p-2p^2)\left(\frac{1}{2} - \frac{2}{3}\right) + \theta(p-p^2)\left(\frac{1}{2} - \frac{1}{3}\right) + \theta^2(p-3p^2+2p^3)\left(\frac{1}{2} - \frac{3}{3} + \frac{2}{4}\right)]\Big|_0^1 dp - 1$$

$$= 4\int_0^1\int_0^1 [\frac{p}{2} - \frac{\theta}{6}(p-2p^2) + \frac{\theta}{6}(p-p^2)]dp - 1$$

$$= 4[\frac{p^2}{4} - \frac{\theta}{6}\left(\frac{p^2}{2} - \frac{2p^3}{3}\right) + \frac{\theta}{6}\left(\frac{p^2}{2} - \frac{p^3}{3}\right)\Big|_0^1 - 1 = 4\left[\frac{1}{4} + \frac{\theta}{36} + \frac{\theta}{36}\right] - 1$$

with some arrangement, we obtain that $\tau_C = \frac{2\theta}{9}$

Remake 2.2.1 In fact, one should note that the range of this type of measures still belong to the interval $[-1,1]$.

From above example, note that, when $\theta = 1$ or $\theta = -1$, so we obtain that $\rho_C \le \left|\frac{1}{3}\right|$ and $\tau_C \le \left|\frac{2}{9}\right|$.

3. Dependence structure between financial returns using Classical Correlations and copula functions

This section shows the importance of copula function for the financial analysis where the decisions must be made by the sponsors. Moreover, a model of financial data analysis will be presented for the chosen indices, which are the Iraqi index, S&P, and Dow Jones index, to show the weakness of Iraq index in comparison with the two worldwide indices (S&P or Dow jones).

3.1 Data Initialization and Collection

This part shows the data that has been applied to the ideas, which are financial data from three indices of the daily financial trading and for the period from 2013 to 2018. A first financial index is the Iraqi index for securities and this index was selected for the purpose of examining and improving the financial management system to improve the financial values of the Iraqi index compared with global indices. A second index is the Dow Jones global index which is an industrial index of the top 30 US manufacturing companies in New York stock exchange. While a third index is the global index S&P which is the stock market index of US stock exchanges of more than 500 financial companies from the US banks and financial institutions.

The Dow Jones index and the S&P index are among the most important indices in the world. Moreover, they are important tool for evaluating the performance of economies of different countries and using them in the financial markets to predict the general trend of stock prices and performance.

The collected data has been sorted by using Microsoft Excel for all financial indices. Each index is represented by a random variable, and refer to each chosen sample as follow:

A first sample represents the Iraqi financial index data and is denoted by I, a second sample represents the Dow Jones financial Index is denoted by D. While the last sample represents the S&P financial data and is denoted by S.

3.2 Testing Data for Normality

Essentially, we can start by testing the normality of the examined data, and assign whether the chosen data follows normal distribution or not. To accomplish this testing, we choose two popular methods.

1. Kolmogrove -Semenerov Method
2. The Z-Score Method

Where, the results of both methods showed that the data of Iraq index is nonlinear and this mean that the data does not follow normal distribution. Therefore, when you apply traditional correlation methods, so results will be less accurate than applying correlation within copulas functions, because it has been mentioned that analysis of data within copulas represent a non-parametric approach for testing and analyzing our collected data.

3.3 Comparison Between Classical Correlations and within Copulas

The comparison between the linear correlation and the copula correlation requires calculations of the correlation in two cases: the first case can be done by computing the classic linear correlation among financial indices, mentioned in 3.1. The classical Pearson coefficients can be calculated from the data by calculating the variance and covariance of all bivariate random variable, such as the Iraqi index, S&P and Dow Jones index, using equation (2), (3) and (1) respectively. Subsequently, computing the rank correlation coefficients (Spearman coefficient and Kendall's tau coefficient). This calculation has been done using MATLAB software, and the results were shown in **Table 1**.

On the second case, the correlation coefficients for financial indexes will be calculated using the copula functions. As mentioned earlier, these functions are used to show their importance and impact on the correlation values in case data are not normally distributed. This is followed by comparing the correlation values with those that are calculated by using this function. Where it was adopted the Gaussian copula function, which calculates the Pearson coefficient, using MATLAB software and then followed by computing the rank correlation coefficient using copula function by applying equations (9) and (11), see [7].

Index	ν	ρ	τ	ν_C	ρ_C	τ_C
D,I	-0.031	0.112	0.0084	0.0132	-0.0296	-0.0197
D,S	0.968	0.942	0.809	0.9554	0.9649	0.8385
S,I	-0.035	0.023	0.0167	0.0262	-0.0334	-0.0223

Table 1: Different Correlations with Each Pair of Indices

Index	Student's t-copula	Frank copula	Clation copula
D,I	0.0132	0.0756	0.0169
S,I	0.0262	0.1503	0.034

Table 2: Some Families of Copula for Calculate Correlations

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 3 Issue 10, October – 2019, Pages: 9-15

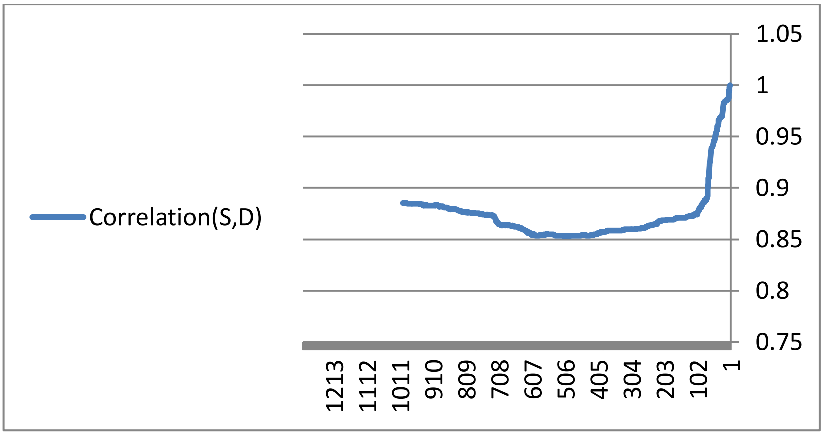

Figure 1: Correlation Coefficient Between S&P index and Dow Jones index

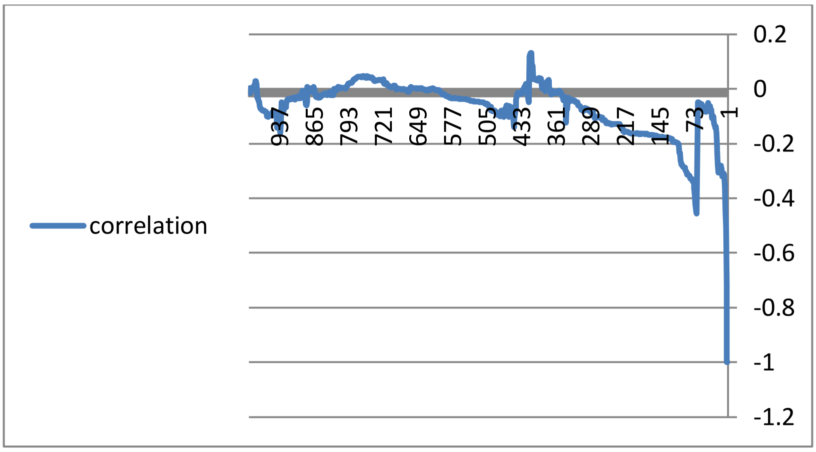

Figure 2: Correlation Coefficient Between S&P index and Iraqi index

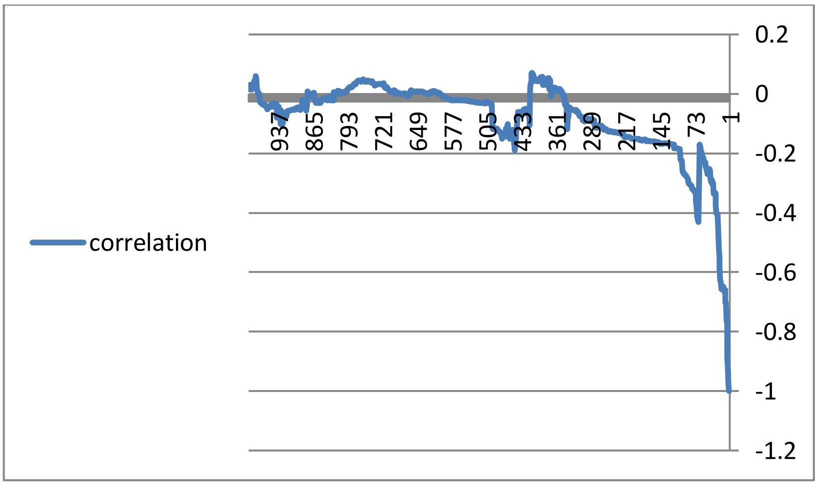

Figure 3: Correlation Coefficient Between Dow Jones index and Iraqi index

Results and Discussion

From **Tables 1** and 2, we can summarize our results as follow:

1. It's clear from that the Pearson coefficient between the Dow and S&P with the Iraqi index is weak. Moreover, they are inversely proportional, since when the global index increases the Iraqi index decreases

due to the negative sign. On the other hand, the value of the correlation coefficient between the global indices is about (0.968), which is an indication of a strong direct correlation.

2. **Table 1** shows that results of the Spearman coefficient are different from those computed by the Pearson coefficient, which is clear indication of the inaccuracy of the Pearson index for measuring the correlation between the Iraqi index and the global indices. The results of the Spearman index were better than the Pearson index, wherein the results of the correlation between the global index (Dow and S&P) with the Iraqi index (0.112), (0.023) respectively have shown that the value of the correlation is directly proportional even its weak, while the value of the correlation between the two global indices was (0.942), which is a strong direct proportional.
3. The obtained results in **Table 1** by using Kendal tau coefficient were found better than those obtained by using Pearson correlation coefficients.
   As by in Pearson and Spearman, the value of the correlation between the two global indices is better and same direction. This is due to the strength of these indicators and their linear relationship (i.e. data are normally distributed).
4. **Table 1** shows the correlation that is resulted by using Copulas function is better than others, wherein obtained values are between the Dow Jones index and S&P index with the Iraqi index (0.0132), (0.0262) respectively. This is due to the functions can which uses normal distribution or not. While the results of the correlation between the global index (Dow Jones and S&P) are (0.9554) this value close to the values resulting from the classical correlation because the data two indices are followed normal distribution.
5. **Table 2** shows the correlation by using some functions: Student's t-copula, Frank copula and Clation copula.
6. **Table 1** and **2** illustrate the improvement within the values of correlation coefficients resulting from copula functions. It also indicates that copula function is much better than traditional correlation functions. Moreover, such function has no restrictions, which allow for better calculations. Consequently, copula functions are more general than the traditional correlation function.
7. **Figure 1, 2** and **3** displays the correlation coefficients between the three indices.

4. Conclusion

From this study and according to the statistical inference of the collected data, we can set the following conclusions:

1- Copulas are much better structures and smother than the ordinary correlation coefficient. In other words, copulas are more general structures than classical measures of association.
2- Linear correlation coefficient is not sufficient to analyze data when they do not follow normal distribution. While copulas are very efficient and modern phenomena in the analysis of various aspects of science. They do not describe the correlation only, but they are very good in describing the whole dependence structure.
3- The concept of dependency embedded in copula functions is much more general than the standard linear correlation concept, and it is a much powerful tool to counting the non-linear relationships.

There are several future works that we can propose within this study. Apply the methods, techniques, and concepts on different type of data like insurance data, bi medical data. And investigate the results of the calculated measures of association through copulas with respect to Reni axioms, see [8].

References

1- W. H. Kruskal, (1958): Ordinal measures of association, Journal of the American Statistical Association 53, 814–861.
2- Cherubini U, Luciano E, Vecchiato W (2004): Copula methods in finance. John Wiley, Sons Ltd.
3- H. E. Daniels,(1950): Rank correlation and population models, Journal of the Royal Statistical Society Series B 12, 171–181.
4- Nelsen R. B. (2006): An Introduction to Copulas, Springer Series in Statistics, 139, 270.
5- Al-Adilee A, Mohammed K. (2014): Copula Functions and correlations Personal Contribution, Journal of Kerbala University , Vol. 12 No.4 Scientific.
6- Blomqvist N. (1950): On a measure of dependence between two random variables. Ann. Math. Stat. 21(4). Pp. 593-600.
7- Habiboellah, F. (2007): Copulas, modeling dependencies in Financial Risk Management, BMI, university of Amesterdam.
8- Renyi, A. (1959): On measures of dependence. Acta. Math. Acad. Sci. Hungar. 10 441-451.