Bitlis Eren University Journal of Science and Technology 7(1) (2017) 27–32

Available online at www.dergipark.ulakbim.gov.tr/beuscitech/

Journal of Science and Technology

E-ISSN 2146-7706

Modelling temperature measurement data by using copula functions

Ayşe METİN KARAKAŞ[1*]

[1]Bitlis Eren University, Department of Statistics, TR-13000, Bitlis Turkey

ARTICLE INFO

Article history:
Received 31 August 2016
Received in revised form 11 April 2017
Accepted 17 May 2017

Keywords:
Copula function, Archimedean copula function, Kendall tau, Spearman rho, Temperature, Akaike information criteria

ABSTRACT

In this study, methods of copula estimation are used and the temperature measurement data of the four regions located at the same positions in the range of 01.01.2008 - 30.04.2009 was modeled with copula functions. For dependence structures of the data sets, it is calculated Kendall Tau and Spearman Rho values which are nonparametric. Based on this method, parameters of copula are obtained. A clear advantage of the copula-based model is that it allows for maximum-likelihood estimation using all available data. The main aim of the method is to find the parameters that make the likelihood functions get its maximum value. With the help of the maximum-likelihood estimation method, for copula families, it is obtained likelihood values. These values, Akaike information criteria (AIC) are used to determine which copula supplies the suitability for the data set.

1. Introduction

Copulas are functions that link the marginal distributions to their joint distribution. The notion of copulas is well understand, it is now known that their empirical estimation is stronger. In bivariate status, copulas can be used to description nonparametric measures of dependence for random variables. Asymmetric models of dependence are suited like those that exceed correlation and linear association, and then copulas move a specific role in developing additional status and measures. Copulas are helpful extensions and generalizations of approaches for modeling joint distribution and dependence that have seem in the literature. Succinctly defined copulas are functions that appropriate multivariate distributions to their one-dimensional margins. Copulas are considered in different applications. Especially, these functions are used in the extreme value theory. While theoretical properties of these objects are now fairly understood, inferences for copula models are not extent.

2. Material and Method

2.1. Copula Theory

The copula is defined as a $C:[0,1]^2 \rightarrow [0,1]$ that ensures the limiting conditions

- ✓ $C(u,0)=C(0,u)=0$ and $C(u,1)=C(1,u)=u, \forall u \in [0,1]$.
- ✓ $(u_1,u_2,v_1,v_2) \in [0,1]^4$, such that, $u_1 \le u_2, v_1 \le v_2$

$$C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1) \ge 0$$

Ultimately, for twice differentiable and 2-increasing property can be replaced by the condition

* *Corresponding author.*
Tel.: 05388394839
E-mail address:akarakas@beu.edu.tr

$$c(u,v)=\frac{\partial^2 C(u,v)}{\partial u \partial v}\geq 0 \tag{1}$$

where $c(u,v)$ is the copula density. In the following, for n-uniform random $U_1,U_2,...,U_n$ variables, the joint distribution function C is defined

$$C\left(u_1,u_2,\ldots,u_n,\theta\right)=P\left(U_1\leq u_1,U_2\leq u_2,\ldots U_n\leq u_n\right).$$

Here θ is dependence parameter (Cherubini and Luciano, 2001; Frees and Valdez, 1998; Genest and MacKay, 1986; Genest and Favre, 2007; Genest et. al., 2009; Malevergne and Sornette, 2003; Metin and Çalık, 2012; Sklar, 1959; Schweitzer and Wolff, 1981).

2.1.1. Sklar Theorem

Let X and Y be random variables with continuous distribution functions F_X and F_Y, with $F_X\left(X\right)$ and $F_Y\left(Y\right)$ are uniformly distributed on the interval $[0,1]$. Then, there is a copula such that for all $x,y\in R$,

$$F_{XY}(X,Y)=C(F_X(x),F_Y(y)) \tag{2}$$

The copula C for $\left(X,Y\right)$ is the joint distribution function for the pair $F_X\left(X\right)$, $F_Y\left(Y\right)$ provided F_X and F_Y continuous (Cherubini and Luciano, 2001; Frees and Valdez, 1998; Genest and MacKay, 1986; Genest and Favre, 2007; Genest et. al., 2009; Metin and Çalık, 2012; Naifar, 2010; Nelsen, 1999; Sklar, 1959; Schweitzer and Wolff, 1981; Quesade-Molina, 1992).

2.1.2. Gaussian Copula

The normal copula;

$$\begin{aligned}C\ (u_1,u_2;\theta)&=\Phi_G(\Phi^{-1}(u_1),\Phi(u_2);\theta)\\&=\int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)}\frac{1}{2\Pi(1-\theta^2)^{1/2}}\left\{\frac{-(s^2-2\theta st-t^2}{2(1-\theta^2)}\right\}dsdt\end{aligned} \tag{3}$$

where is the cdf of standard distribution, and is the standard bivariate normal distribution with correlation parameter limited to the interval (-1,1) (Cherubini and Luciano, 2001; Schweitzer and Wolff, 1981).

2.1.3. Archimedean Copula

Let φ define a function $\phi:[0,1]\rightarrow[0,\infty]$ which is continuous and provides:

- ✓ $\phi(1)=0, \phi(0)=\infty.$
- ✓ For all $t\in(0,1)$, $\phi^{'}(t)<0$,φ is decreasing, for all $t\in(0,1)$ $\varphi''\left(t\right)\geq 0$,φ is convex.

φ has an inverse $\varphi^{-1}:\left[0,\infty\right]\rightarrow\left[0,1\right]$, which has the same properties out of $\phi^{(-1)}(0)=1$ and $\phi^{(-1)}(\infty)=0$. The Archimedean Copula is defined by

$$C(u,v)=\phi^{(-1)}[\phi(u)+\phi(v)]. \tag{4}$$

2.1.4. Gumbel Copula

This Archimedean copula is defines with the help of generator function $\phi\left(t\right)=\left(-lnt\right)^{\theta}$, $\theta\geq 1$;

$$C_\theta(u,v)=\exp\left(-[(-\ln u)^\theta+(-\ln v)^\theta]^{1/\theta}\right) \tag{5}$$

where θ is the copula parameter restricted to $(1,\infty]$.This copula is asymmetric, with more weight in the right tail. Beside this, it is extreme value copula (Nelsen, 1999).

2.1.5. Clayton Copula

This Archimedean copula is defines with the help of generator function $\phi(t)=\dfrac{t^{-\theta}-1}{\theta}$,$\theta\in\left[-1,\infty\right)/\left\{0\right\}$

$$C_\theta(u,v)=(u^{-\theta}+v^{-\theta}-1). \tag{6}$$

Where θ is the copula parameter restricted to $(0,\infty)$. This copula is also asymmetric, but with more weight in the left tail (Nelsen, 1999).

2.1.6. Frank Copula

This Archimedean copula is defines with the help of generator function; $\phi\left(t\right)=-\ln\dfrac{-e^{-\theta t}-1}{e^{-\theta}-1}$,$\theta\in R/\left\{0\right\}$;

$$C_\theta\left(u,v\right)=-\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{\left(e^{-\theta}-1\right)}\right) \tag{7}$$

where θ is the copula parameter restricted to $\left(0,\infty\right)$ (Nelsen, 1999).

2.1.7. Joe Copula

This Archimedean copula defines with the help of generator function

$$C_\theta(u,v) = 1-\left[\left(1-u\right)^{\theta}+\left(1-v\right)^{\theta}-\left(\left(1-u\right)^{\theta}\left(1-v\right)^{\theta}\right]^{1/\theta} \quad (8)$$

where θ is the copula parameter restricted to $[1,\infty]$. This copula family is similar to the Gumbel. The right tail positive dependence is stronger more than Gumbel (Nelsen, 1999).

2.2. Dependence Structure of Copulas

In this section, we explore ways in which copulas can be used in the study of dependence or association between random variables. There are varieties of ways to discuss and to measure dependence. Dependence properties and measures of association are interrelated, and so there are many places where we could begin this study.

2.2.1. Spearman's Rho

Similar to approach of Pearson correlation coefficient, to compute the correlation between the pairs (R_i, S_i) of ranks have been used. Thus, Spearman's Rho

$$\rho_n = \frac{\sum_{i=1}^{n}(R_i-\bar{R})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i-\bar{R})^2\sum_{i=1}^{n}(S_i-\bar{S})^2}} \in [-1,1] \quad (9)$$

where

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n}R_i = \frac{n+1}{2} = \frac{1}{n}\sum_{i=1}^{n}S_i \quad (10)$$

write. This coefficient that stated expediently in the form

$$\rho_n = \frac{12}{n(n+1)(n-1)}\sum_{i=1}^{n}R_iS_i - 3\frac{n+1}{n-1} \quad (11)$$

Also, ρ_n is asymptotically unbiased estimator of

$$\rho = 12\int_{[0,1]^2} uv\,dC(u,v) - 3 = 12\int_{[0,1]^2} C(u,v)\,du\,dv - 3 \quad (12)$$

where the second equality is obtained (Genest and Favre, 2007). This statement extended Quesada- Molina (1992)

$$12\int_{[0,1]^2} uv\,dC_n(u,v) - 3 = \frac{12}{n}\sum_{i=1}^{n}\frac{R_i}{n+1}\frac{S_i}{n+1} - 3 = \frac{n-1}{n+1}\rho_n \quad (13)$$

and $C_n \to C$ as $n \to \infty$. Here the null hypothesis $H_0 = C = \Pi$ of independence of X and Y, the distribution of ρ_n is normal with zero mean and variance $1/(n-1)$, thus for H_0 approximate $\alpha = 0.05$, $\sqrt{n-1}\left|\rho_n\right| > z_{\alpha/2} = 1,96$ (Genest and Favre, 2007).

2.2.2. Kendall Tau

Another measure of dependence is Kendall Tau. This measure based on ranks given by

$$\tau_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4}{n(n-1)}P_n - 1 \quad (14)$$

where P_n and Q_n number of concordant and discordant pairs respectively. Here, $(X_i, Y_i), (X_j, Y_j)$ pairs are concordant $(X_i - X_j)(Y_i - Y_j) > 0$ and these are disconcordant $(X_i - X_j)(Y_i - Y_j) < 0$. If $(X_i - X_j)(Y_i - Y_j) > 0$; we can say $(R_i - R_j)(S_i - S_j) > 0$. τ_n is function of copula C_n. As $n \to \infty$, $C_n \to C$

$$W = \frac{1}{n}\sum_{j=1}^{n} I_{ij} = \frac{1}{n}\#\left\{j : X_j \le X_i, Y_j \le Y_i\right\},$$

$$\tau_n = 4\frac{n}{n-1}\bar{W} - \frac{n+3}{n-1} = 4\int_{[0,1]^2} C(u,v)\,dC(u,v) - 1 \quad (15)$$

written. τ_n is asymptotically unbiased estimator of τ and τ_n is normal with zero mean and variance $2(2n+5)/\{9n(n-1)\}$. Here the null hypothesis $H_0 = C = \Pi$ of independence of X and Y, thus for H_0 approximate $\alpha = 0.05$, $\sqrt{9n(n-1)/2(2n+5)}\left|\tau_n\right| > 1.9($ (Genest and Favre, 2007).

2.3. Copula Estimation Method

2.3.1. Maximum Likelihood Method (MLE)

Maximum likelihood method is the most used for copula. The aim of the method is basic to find the parameters that make the likelihood functions get its maximum value. It is given

$$f(x_1, x_2, ..., x_n) = c(F_1(x_1), F_2(x_2), ..., F_n(x_n)) \prod_{j=1}^{n} f_j(x_j) \quad (16)$$

$$c(F_1(x_1), F_2(x_2), ..., F_n(x_n)) = \frac{\partial^n c(F_1(x_1), F_2(x_2), ..., F_n(x_n))}{\partial F_1(x_1), F_2(x_2), ..., F_n(x_n)}$$

Let $\left\{x_{1t}, x_{2t}, ..., x_{nt}\right\}_{t=1}^{T}$ is the sample data matrix; the likelihood functions can be given

$$l(\theta) = \sum_{t=1}^{T} \ln(c(F_1(x_{1t}), F_2(x_{2t}), ..., F_n(x_{nt})) + \sum_{t=1}^{T} \sum_{j=1}^{n} \ln f_j(x_{jt}) \quad (17)$$

Accordingly, the maximum likelihood estimator is

$$\hat{\theta}_{MLE} = \max_{\theta} l(\theta) \quad (18)$$

2.3.2. Inference for marginal (IFM):

This method is used to overcome the drawbacks of full maximum likelihood function. The aim of copula theory is separate between the univariate margins and the dependence structure. From equation (18)

$$l(\theta) = \sum_{t=1}^{T} \ln(c(F_1(x_{1t}, \theta_1), F_2(x_{2t}, \theta_2), ..., F_n(x_{nt}, \theta_n), \alpha) + \sum_{t=1}^{T} \sum_{j=1}^{n} \ln f_j(x_{jt}, \theta_j) \quad (19)$$

In this equation (19) the vector of the parameters for the univariate marginal $\theta = (\theta_1, \theta_2, ..., \theta_n)$ and α is vector the parameters of copula. Accordingly, the fundamental idea of inference for margins is that it is forecasts the parameters for marginal distributions and copula separately in two stages.

- ✓ Estimate the parameters θ_j from marginal distributions,

  $$\hat{\theta}_j = \arg\max_{\theta_t} \sum_{t=1}^{T} \ln f_j(x_{jt}; \theta_j) \quad (20)$$

- ✓ Estimation of the vector of the copula parameters α, used the $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n)$;

$$\hat{\alpha}_{IFM} = \arg\max_{\alpha} \sum_{t=1}^{T} \ln(c(F_1(x_{1t}, \hat{\theta}_1), F_2(x_{2t}, \hat{\theta}_2), ..., F_n(x_{nt}, \hat{\theta}_n); \alpha) \quad (21)$$

2.3.3. Akaike Information Criteria

For the series, to model dependence structure, other selection criteria are Akaike's information criterion (AIC). This;

$$AIC = -2 \log L + 2k / n \quad (22)$$

Here, k is the number of estimated parameter for each model, n size of sample.

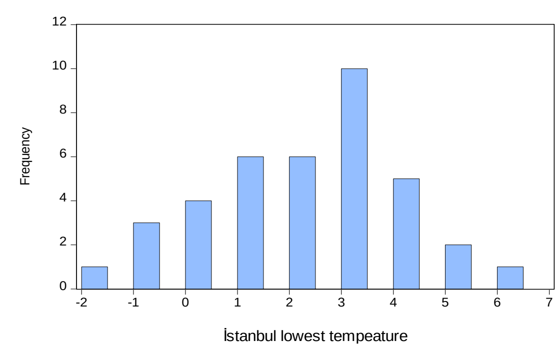
3. Application

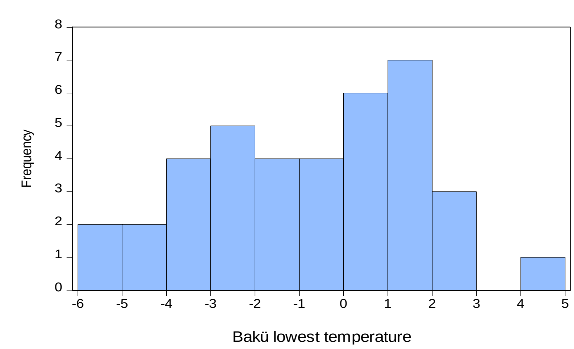
3.1. Data Sets

In this study, dependence structure between at the same time temperature measurements located in similar coordinate İstanbul, Rome, Baku and Tokyo modeled by copula functions. The dependence structure between the selected regions is being done estimated nonparametric method based on using Kendall tau methods. With the help of this method copulas family suitable selected data is determined, accordance with parameters for this family are calculated. In the study, because of pairwise comparisons $\binom{4}{2}$ correlation analysis table were obtained. Kendal Tau values are given in the table.1 below.

Table 1. For four areas Kendall Tau (τ) rank correlation

	İstanbul	Roma	Bakü	Tokyo
İstanbul	1	0,675	0,741	0,708
Roma	0,675	1	0,655	0,636
Bakü	0,741	0,655	1	0,707
Tokyo	0,708	0,636	0,707	1





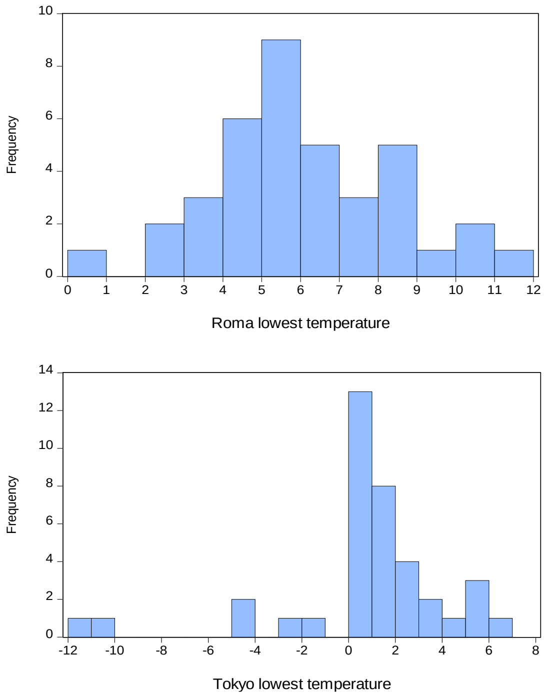


Figure 1. Frequency of Istanbul, Baku, Rome and Tokyo lowest temperature

3.2. Kendal Tau and Gumbel Hougaard Family Estimation

The relationship between Kendall Tau and Gumbel Hougaard family;

$$\tau_K(\theta) = \frac{\theta - 1}{\theta}$$

From this equation given above for the low temperature Gumbel Hougaard family estimations are given in the following table.

Table 2. For Gumbel Hougaard family is the parameter

	θ	Logl	AIC
İstanbul-Roma	3,076923	902,8711	-1805,74
İstanbul-Baku	3,861004	1312,783	-2625,56
İstanbul-Tokyo	3,424658	1074,603	-2149,2
Roma-Baku	2,898551	775,5424	-1551,08
Roma-Tokyo	2,747253	723,4774	-1446,95
Baku-Tokyo	3,412969	1063,465	-2126,93

3.3. Kendal Tau and Clayton Family Estimation

The relationship between Kendall Tau and Clayton family;

$$\tau_K(\theta) = \frac{\theta}{\theta + 2}$$

from this equation given above for the low temperature Clayton family estimations are given in the following table.

Table 3. For Clayton family is the parameter

	θ	Logl	AIC
İstanbul-Roma	4,15384	-88,5239	177,0519
İstanbul-Bakü	5,72008	-50,5097	101,0235
İstanbul-Tokyo	4,84931	-69,3936	138,7913
Roma-Bakü	3,79710	-94,1912	188,3865
Roma-Tokyo	3,49450	-64,4112	128,8265
Bakü-Tokyo	4,825939	-96,792	193,5881

3.4. Kendal Tau and Frank Family Estimation

The relationship between Kendall Tau and Frank family;

$$\tau_K(\theta) = 1 - \left(\frac{4}{\theta} \left[1 - D_1(\theta) \right] \right)$$

Here D is Debye function. From this equation given above for the low temperature Frank family estimations are given in the following table.

Table 4. For Frank family is the parameter

	θ	Logl	AIC
İstanbul-Roma	10,35253	-2313,32	4626,644
İstanbul-Bakü	13,57225	-3005,15	6010,344
İstanbul-Tokyo	11,78704	-2635,73	5271,464
Roma-Bakü	9,610621	-2165,4	4330,804
Roma-Tokyo	8,976924	-2034,5	4069,004
Bakü-Tokyo	11,73902	-2631,74	5263,484

3.5. Kendal Tau and Joe Family Estimation

The relationship between Kendall Tau and Joe family;

$$\tau_K(\theta) = 1 + \frac{4}{\theta} D_J(\theta)$$

Here D is Debye function. From this equation given above for the low temperature Joe family estimations are given in the following table.

Table 5. For Joe family is the parameter

	θ	Logl	AIC
İstanbul-Roma	4,958317	-83,2145	166,4331
İstanbul-Baku	6,507019	-95,1592	190,3225
İstanbul-Tokyo	5,644045	-105,571	211,461
Roma-Baku	4,607484	-33,1887	60,38152
Roma-Tokyo	4,310508	-26,0271	52,05832
Bakü-Tokyo	5,620964	-58,2653	116,5347

4. Results and Discussion

The studies have focused on copula forecasting methods and using the daily minimum temperatures of four different areas are made in applications. In application it was established to investigate the relationship between the lowest temperatures of the four cities in the world in the time interval 01.01.2008 - 30.04.2009 by using a dataset with 486 units of the temperatures of the four cities on modeling the concept of copulas. In practice, it has been making predictions by choosing different copulas families with non-parametric copulas estimation method. This study has been shown that the Gumbel Hougaard, Clayton, Frank and Joe copulas are statistically better than the other families for our data set. Since the Gumbel Hougaard and the Clayton families are the Archimedean copula classes, they provide easiness in the calculation. Gumbel, Clayton and Gaussian families are more useful in modeling the structure of the dependency for low temperature measurements among the regions. Consequently, the following equalities can be written for each value of the dependency parameter.

References

Cherubini, U., Luciano, E., 2001. Value-at-Risk Trade-off and Capital Allocation with Copulas. Economic Notes 30, 235–256.

Frees, E.W., Valdez, E.A., 1998. Understanding relationships using copulas. North American Actuarial Journal 2, 1-25.

Genest, C., Gendron, M., Boudeau-Brien, M., 2009. The Advent of Copulas in Finance the European Journal of Finance 15, 609-618.

Genest, C., MacKay, V., 1986. The joy of copulas: bivariate distributions with uniform marginal. The American Statistician 40, 280-283.

Genest, C., Rivest, L.P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88(423), 1034-1043.

Genest, C., Favre, A.C., 2007. Everything You Always Wanted to Know About Copula Modelling but Were Afraid to Ask. Journal of Hydrologic Engineering 12, 347-368.

Malevergne, Y., Sornette, D., 2003. Testing the Gaussian Copula Hypothesis for Financial Assets Dependences. Quantitative Finance 3 (4), 231-250.

Metin, A., Çalık, S., 2012. Copula Function and Application with Economic Data. Turkish Journal of Science and Technology 7(2), 199-204.

Naifar, N., 2010. Modeling dependence structure with Archimedean copulas and applications to the iTraxx CDS index. Journal of Computational and Applied Mathematics 235, 2459-2466.

- Nelsen, R.B., 1999. An Introduction to Copulas. First edition. Published by the Springer Verlag, New York, ISBN 978-0-387-28678-5.

Quesade-Molina, J., 1992. A generalization of an identity of Hoeffding and some applications. J. Ital. Stat. Soc. 3, 405-4.

Schweitzer, B., Wolff, E.F., 1981. On nonparametric measures of dependence for random variables. Annals of Statistics 9, 879-885.

Sklar, A., 1959. Fonctions de Repartition an Dimensions et Leurs Marges. Publications de I'lnstitut de Statistique de I'University de Paris 8, 229-231.