MODELING OF DEPENDENCE STRUCTURES IN METEOROLOGICAL DATA VIA ARCHIMEDEAN COPULAS

Vadoud Najjari[1], Hasan Bal[2], Salih Çelebioğlu [3]

In this study, the main endeavor is to find a copula that fits on meteorological data dependence structure. In order to specify the best copula, we use nonparametric approach and Goodness of fit test. Calculations declare that Archimedean family with cotangent generator (cot-copula) fits to our data truthfully.

Keywords: Archimedean copulas (AC), Generator, Goodness of fit (GOF) test, Kendall's tau, Nonparametric estimation

1. **Introduction**

Actually structure of the dependence between random variables is important. In decision support, properly accounting and modeling these dependencies and correlations are essential in deriving reliable valuations. It has proven, the copulas technique is a superior tool for modeling dependence structures [11, 16].

As Nelsen discussed in his book [11], there are five important international conferences devoted in growing interest on copulas and their applications in statistics and probability which we review them in the following: the Symposium on Distributions with Given Marginals (Fréchet Classes) in Rome in 1990; the conference on Distributions with Fixed Marginals, Doubly Stochastic Measures, and Markov Operators in Seattle in 1993; the conference on Distributions with Given Marginals and Moment Problems in Prague in 1996; the conference on Distributions with Given Marginals and Statistical Modelling in Barcelona in 2000; and the conference on Dependence Modelling: Statistical Theory and Applications in Finance and Insurance in Québec in 2004.

In this study, we take monthly minimum and maximum air temperature records from 1975 to 2010 in Tehran and try on modeling dependence structures of them. Hence we consider three the most widely used Archimedean

[1,2,3] Gazi university, faculty of science, statistics department, Ankara, Turkey.

[1] Department of mathematics, Islamic Azad university, Maragheh branch, Maragheh, Iran.

e-mail: [1] vnajjari@gazi.edu.tr; vnajjari@iau-maragheh.ac.ir
[2] hasanbal@gazi.edu.tr; [3] scelebi@gazi.edu.tr

families of copulas: Clayton, Gumbel and A1 (this family is numbered as 4.2.12 in Nelsen [11]). Also we use an Archimedean family with cotangent generator (*cot*-copula) which recently presented by Pirmoradian and Hamzah [13]. In order to be more familiar with the *cot*-copula, details of this family are demonstrated in Table 1. Calculations declare that this family truthfully fits to our data.

This paper is constructed as follows: in Section 2 we discuss about copulas and AC. Section 3, reviews fitting copulas to bivariate data. And Section 4 shows modeling of dependence structures in the meteorological data via the mentioned AC. Section 5 summarizes the conclusion of our work.

2. Preliminaries

Sklar (1959), for the first time, used the word copula, a function which allows us to combine univariate distributions to obtain a joint distribution with a particular dependence structure [15]. A copula is a function $C : [0,1]^2 \to [0,1]$ which satisfies:

(a) for every u, v in [0,1], $C(u,0) = 0 = C(0,v)$, and $C(u,1) = u$, $C(1,v) = v$;

(b) for every u_1, u_2, v_1, v_2 in [0, 1] such that $u_1 \leq u_2$, and, $v_1 \leq v_2$, $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$.

The importance of copulas in statistics is described by Sklar Theorem [15]: Let X and Y be random variables with joint distribution function H and marginal distribution functions F and G, respectively. Then there exists a copula C such that, $H(x,y) = C(F(x), G(y))$, for all x, y in $\mathbb{R}$. If F and G are continuous, then C is unique. Otherwise, the copula C is uniquely determined on $Ran(F) \times Ran(G)$. Conversely, if C is a copula and F and G are distribution functions then the function H is a joint distribution function with margins F and G. As a result of the Sklar's theorem, copulas link joint distribution functions to their one-dimensional margins.

2.1. Archimedean Copulas

One of the most important classes of copulas is known as AC. This copulas are very easy to construct, many parametric families belong to this class and have a great variety of different dependence structures. In addition, the Archimedean representation allows us to reduce the study of a multivariate copula to a single univariate function. Archimedean copulas originally appeared not in statistics, but rather in the study of probabilistic metric spaces, where they were studied as part of the development of a probabilistic version of the triangle inequality. For details see [11, 14].

Basic properties of AC are presented below and more information could be found in Nelsen [11]. Let φ be a continuous, strictly decreasing function from I to $[0, \infty]$ such that $\varphi(1) = 0$. The pseudo-inverse of φ is the function

$\varphi^{[-1]}$ given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{(-1)}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases} \tag{1}$$

Copulas of the form $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$, for every u, v in I are called AC and the function φ is called a generator of the copula. If $\varphi(0) = \infty$ we say that φ is a strict generator. In this case, $\varphi^{[-1]} = \varphi^{(-1)}$ and $C(u, v) = \varphi^{(-1)}(\varphi(u) + \varphi(v))$ is said a strict Archimedean copula.

In the AC, one of the main topics is statistical inference on the dependence parameter θ. Several methods of copula parameter estimation have been developed, including the methods of concordance [4, 12], fully maximum likelihood (ML), pseudo maximum likelihood (PML) [5], inference function of margins (IFM) [7, 8] and minimum distance (MD) [16], etc. Nonetheless in this paper, we focus only on nonparametric estimation in the dependence parameter θ which is based on Kendall's tau, and in selecting the right copula we focus on the GOF test and the nonparametric approach using the procedure of Genest and Rivest [6].

3. Fitting copulas to data

In this section for identifying the right copula, we review the nonparametric approach using the procedure of Genest and Rivest [6] and also GOF test.

Table 1. Details of the selected copula families in this study

Family	Generator	Kendall's tau	λ_L	λ_U	θ interval
Clayton	$\frac{1}{\theta}(\frac{1}{t^\theta} - 1)$	$\frac{\theta}{\theta+2}$	$2^{\frac{-1}{\theta}}$	0	$(0, \infty)$
Gumbel	$(-lnt)^\theta$	$\frac{\theta-1}{\theta}$	0	$2 - 2^{\frac{1}{\theta}}$	$[1, \infty)$
A1	$(\frac{1}{t} - 1)^\theta$	$1 - \frac{2}{3\theta}$	$2^{\frac{-1}{\theta}}$	$2 - 2^{\frac{1}{\theta}}$	$[1, \infty)$
cot-copula	$cot^\theta(\frac{\pi t}{2})$	$1 - \frac{8}{\pi^2\theta}$	$2^{\frac{-1}{\theta}}$	$2 - 2^{\frac{1}{\theta}}$	$[1, \infty)$

Note: A1 is a familiy that is numbered 4.2.12 in Table 4.1, Nelsen's book [11].

3.1. Nonparametric approach using the procedure of Genest and Rivest

Assume that we have a random sample of bivariate observations (X_i, Y_i) for $i = 1, ..., n$ available and joint distribution function H has associated Archimedean copula C_φ, we wish to identify the form of φ. First to begin with, define an intermediate (unobserved) random variable $Z_i = H(X_i, Y_i)$ that has distribution function $K(z) = Prob[Z_i \leq z]$ (see [2, 3, 6]). This distribution function is related to the generator of an Archimedean copula through the expression,

$$K(z) = K_\varphi(z) = z - \frac{\varphi(z)}{\varphi'(z)}. \tag{2}$$

In order to identify φ,

(1) find Kendall's tau using the usual (nonparametric or distribution-free) estimate

$$\tau_n = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} Sign[(X_i - X_j)(Y_i - Y_j)]. \quad (3)$$

(2) Construct a nonparametric estimation of K, as follows:
- First, define the pseudo-observations,

$$Z_i = (n-1)^{-1} \sum_{j=1}^{n} if[X_j < X_i \ and \ Y_j < Y_i, 1, 0], \ i = 1, 2, ..., n. \quad (4)$$

- Second, construct the estimate of K,

$$K_n(z) = n^{-1} \sum_{i=1}^{n} if[Z_i \leq z, 1, 0]. \quad (5)$$

(3) Now construct a parametric estimate K_φ by using (2). Illustratively, $\tau_n \to \theta_n \to \varphi_n(t) \to K_{\varphi_n}(z)$, where subscript n denotes estimation.

The step (3) is to be repeated for every copula family that we wish to compare. The best choice of generator then corresponds to the parametric estimate $K_{\varphi_n}(z)$, that most closely resembles the nonparametric estimate $K_n(z)$. Measuring closeness can be done either by a (L_2-norm) distance such as

$$\int_0^1 [K_{\varphi_n}(z) - K_n(z)]^2 dz \quad (6)$$

or graphically by (a) plotting of $z-K(z)$ versus z or (b) corresponding quantile-quantile (Q-Q) plots. Q-Q plots are used to determine whether two data sets come from populations with a common distribution. If the points of the plot, which are formed from the quantiles of the data, are roughly on a line with a slope of 1, then the distributions are the same.

3.2. GOF test.

The χ^2-test uses the below test-statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{[f_i - np(x_i)]^2}{np(x_i)} \quad (7)$$

where k is the number of classes, f_i is the absolute frequency of data in the class i and $np(x_i)$ is the theoretical frequency of data for every class i. Note also that power of the test grows with growing number of the classes.

4. Modeling of dependence structures

In this section, we take monthly minimum and maximum air temperature records from 1975 to 2010 in Tehran, then we consider three the most widely used Archimedean families of copulas: Clayton, Gumbel and A1 (this family numbered 4.2.12 in Nelsen [11]) and also *cot*-copula which recently presented by Pirmoradian and Hamzah [13], for details see Table 1. Figure 2 shows scatter plots of the *cot*-copula for several values of its parameter. In

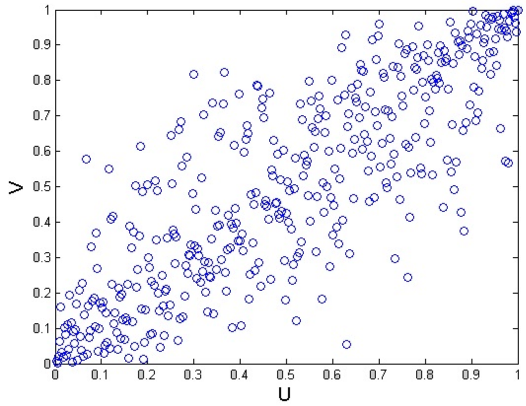

FIGURE 1. Scatter plots of the minimum and maximum air temperature records.

continue, each series of the mentioned data (monthly maximum and minimum air temperature) is arranged and has been divided to total data sample size plus one, as below (see [1], [6])

$$u_i = \frac{R(x_i)}{n+1}\ ,\ v_i = \frac{R(y_i)}{n+1} \tag{8}$$

where $i = 1, 2, 3, ..., n$ and $x_i,\ y_i$ are the minimum and maximum air temperature records respectively and $R(x_i),\ R(y_i)$ are respectively the rank of the related data. It is clear that the final data (u and v) will be on the interval $(0, 1)$. Figure 1 shows scatter plots of the final data. Kendall's tau for this data is $\tau = 0.6373$.

TABLE 2. Results of the nonparametric estimation and the GOF test

Family	Parameter	$d(K_\varphi, K_n)$	GOF
Clayton	3.5135	0.0900	28.2638
Gumbel	2.7568	0.0241	14.7278
A1	1.8378	0.0140	5.3600
cot-copula	2.2345	0.0096	5.1514

Note: A1 is a familiy that is numbered 4.2.12 in Table 4.1, Nelsen's book [11].

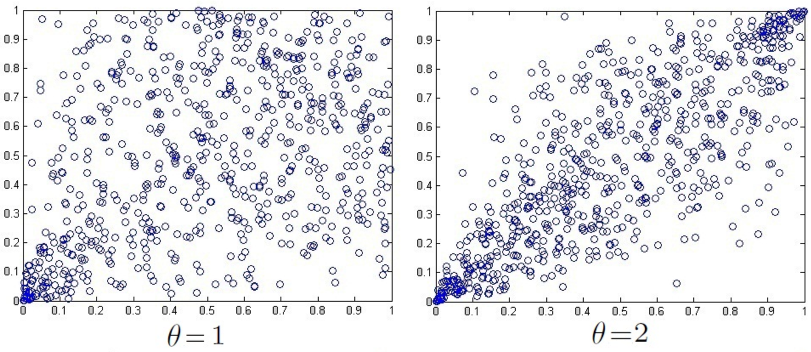

Figure 2. Scatterplots of *cot*-copula for several values of θ, n=700.

With relying on the calculated value of Kendall's tau, for the mentioned any copula family in Table 1, we apply nonparametric estimation of family parameter. Then following the described nonparametric procedure in the Section 3.**1**, we estimate K_n and the parametric estimation for each copula families. Table 2 shows the estimated family parameters and their closeness to K_n numerically. In our process Matlab software has been used. With using relation $\sqrt[4]{n}$, (n number of observations), (see [1, 10]), we divide the range of two variables uniform transformation into 4 intervals each, therefore $df = 9$ and $\chi^2_{0.05,df} = 16.9190$. Thus in any copula family which $\chi^2 < \chi^2_{0.05,df}$, it means that the family is suitable to the mentioned data.

Table 3. Goodness-of-fit test values.

Family	Parameter	GOF
Clayton	2.92	22.3972
Gumbel	2.74	14.7114
A1	1.88	5.2026
cot-copula	2.32	4.7648

Note: A1 is a familiy that is numbered 4.2.12 in Table 4.1, Nelsen's book [11].

To find a suitable family of AC, for all copula families mentioned in the Table 1 we applied the discussed process in previous section and results are shown in Table 2. Moreover we inquired the best copula parameter in the mentioned AC, that minimizes the GOF test statistic value and its results are shown in Table 3.

The second column of the Table 2 consist of the estimated parameter values by the nonparametric method which is based on Kendall's tau, for the mentioned copula families in the Table 1. As an example, parameter of the *cot*-copula is estimated 2.2345, while as it is seen in the Table 3, by minimizing GOF test statistic value, this parameter is estimated 2.32. Moreover, for the mentioned copula families in the Table 1, GOF test statistic values are summarized in the forth column of the Table 2. For example this value for the *cot*-copula is 5.1514, while as it is seen in the Table 3, by minimizing GOF test statistic value, it is estimated 4.7648. The third column of the Table 2 consist of values of the nonparametric approach using the procedure of Genest and Rivest [6]. This value for the *cot*-copula is 0.0096, which is the minimum value between the other mentioned families values.

Obviously with respect to Table 2 and Table 3 the *cot*-copula family is the best one that fits to the mentioned data in the both nonparametric procedure and GOF test. We recall that the *cot*-copula has a trigonometric generator and calculating the dependence measures of this family such as kendall's tau, tail dependences, etc, are not complicated and all of these measures have a closed form, which is an advantage for a copula family to be used in applications.

5. **Conclusion**

For the monthly maximum and minimum air temperature records from 1975 to 2010 in Tehran, we investigate suitable family of AC in modeling of the dependence structures between them. Hence three widely used families of AC and also *cot*-copula family are analyzed. In our process, we focused on the nonparametric approach using the procedure of Genest and Rivest [6] and also the GOF test. Results show that, in the both nonparametric procedure and the GOF test, the *cot*-copula family fits to data the best.

REFERENCES

[1] *S. Çelebioğlu*, Archimedean copulas And An Application, *Selcuk University Journal of Science*, **22** (2003) 43–52.

[2] *V. Durrleman, A. Nikeghbali, T. Roncali*, Which copula is the right one? Groupe de Recherche Operationnelle, *Credit Lyonnais*, 2000.

[3] *E. W. Frees, E. A. Valdez*, Understanding Relationships Using Copulas, *North American Actuarial Journal*, **2** (1998) 1–25.

[4] *C. Genest*, Franks family of bivariate distributions, *Biometrika*, **74** (1987) 549–555.

[5] *C. Genest, K. Ghoudi, L.P. Rivest*, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82** (1995) 543–552.

[6] *C. Genest, L. P. Rivest*, Statistical Inference Procedures for Bivariate Archimedean copulas, *Journal of the American Statistical Association*, **88:423** (1993) 1034–1043.

[7] *H. Joe*, Multivariate Models and Dependence Concepts, *Chapman - Hall*, London, 1997.

[8] *H. Joe*, Asymptotic efficiency of the two-stage estimation method for copula-based models, *J. Multivariate Anal*, **94** (2005) 401–419.

[9] *G. Kim, M.J. Silvapulle, P. Silvapulle*, Comparison of semiparametric and parametric methods for estimating copulas, *Comm. Statist. Simulation Comput*, **51** (2007) 2836–2850.

[10] *V. Najjari, M. G. Unsal*, An Application of Archimedean Copulas for Meteorological Data, *G.U. Journal of Science*, **25(2)** (2012) 301–306.

[11] *R. B. Nelsen*, An Introduction to copulas, *Springer, New York*, second edition, 2006.

[12] *D. Oakes*, A model for association in bivariate survival data, *J. Roy. Statist. Soc. Ser. B*, **44(3)** (1982) 414–422.

[13] *A. Pirmoradian, A. Hamzah*, Simulation of Tail Dependence in Cot-copula, *Proceedings of the 58th WSC of the ISI*, 2011.

[14] *B. Schweizer*, Thirty years of copulas. In: DallAglio G, Kotz S, Salinetti G (eds) Advances in Probability Distributions with Given Marginals. *Kluwer, Dordrecht*, (1991) 13–50.

[15] *A. Sklar*, Fonctions de répartition à n dimensions et leurs marges, *Publ Inst Statist Univ Paris*, **8** (1959) 229–231.

[16] *H. Tsukahara,* Semiparametric estimation in copula models, *Canad. J. Statist*, **33** (2005) 357–375.