Gazi University Journal of Science
GU J Sci
29(3):593-597 (2016)

The Dependence Structure Between the Monthly Minimum and Maximum Barometric Data in Iran

Vadoud NAJJARI[1,♠]

[1]*Young Researchers and Elite Club, Maragheh branch, Islamic Azad University, Maragheh, Iran*

Received:28/05/2015 Revised:05/11/2015 Accepted: 26/11/2015

ABSTRACT

In this study, the main endeavor is to use a simple and reliable method to select the right copula family that fits best to the data. In this method, the standard goodness of fit test statistic value is posed as a function of copula parameters and then the main problem is to investigate the minimum value of this function. Hereby an estimation of copula parameters is in the minimum point of the mentioned function and also by the minimum value of the mentioned function, we are able to find the right copula that best fits to the data. By modeling dependence structure between the monthly average minimum and maximum barometric data in Tehran/Iran, the mentioned method will be compared with the results of the maximum likelihood estimation and also the nonparametric method proposed by Genest and Rivest (1993).

Keywords:*Data envelopment analysis, stochastic frontier analysis, copulas, efficiency.*

1. INTRODUCTION

Copulas are multivariate distributions modeling the dependence structure between variables, irrespective of their marginal distributions. They allow to choose completely different margins, the dependence structure given by the copula, and merge the margins into a genuine multivariate distribution. The concept of copula has been introduced by Sklar (1959) in the following way,

A copula is a function $C: [0,1]^2 \rightarrow [0,1]$ which satisfies:

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

(b) for every $u_1, \ u_2, v_1, \ v_2$ in $[0, 1]$ such that $u_1 \le u_2$ and $v_1 \le v_2$,

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

The importance of copulas in statistics is described in Sklar's theorem: 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 $\mathbb{R}$. 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. 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, G

♠Corresponding author, e-mail: vnajjari@iau-maragheh.ac.ir

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. For a more formal definition of copulas, the reader is referred to Nelsen (2006).

This paper concentrates on a method which estimates copula parameters and also selects a right copula which fits the best to data. This method is based on standard GOF test. Estimation of copula parameters by this method is explained with an example and its result is compared with the maximum likelihood estimation and also the nonparametric method proposed by Genest and Rivest (1993).

This paper is constructed as follows: in Section 2 we discuss estimation of copula parameters and also copula selection methods. Section 3 explains the discussed method. An application study is given in Section 4 and finally Section 5 summarizes the conclusion of our work.

2. COPULA SELECTION METHODS

Statistical inference on the dependence parameter in copulas is one of the main topics in multivariate statistical analysis. There are several methods for estimation copula parameters, including the methods of concordance (Genest, 1987; Genest and Rivest, 1993), fully maximum likelihood (ML), pseudo maximum likelihood (PML) (Genest et al. 1995), inference function for margins (IFM) (Joe, 1997; Joe, 2005) and minimum distance (MD) (Tsukahara, 2005). There are some discussions about these methods by Kim et al. (2007). Let's just recall that the PML estimator is better than ML and IFM in the most practical situations. However, when the copula dimension increases, in time-consuming point of view the ML, PML, IFM and MD methods require so much computations. Moreover, these methods use the copula density function, that increase complexity of calculations, especially for $d > 2$[1] (Yan, 2007).

Brahimi and Necir (2012) proposed a semi-parametric estimation of copula models based on the method of moments. This method is quick and simple to use with reasonable bias and root mean squared error. Nonetheless, Brahimi and Necir's method has its own complexity.

After estimating copula parameters, the main problem is in selecting the right copula that best fits to data. There are several methods for selecting the best copula some of which are summarized as follows:

Methods of selecting the best copula commonly are based on a likelihood approach, which is used to define indicators of performance, for example, the Akaike Information Criteria (AIC), Pseudo-likelihood ratio test proposed by Chen and Fan (2005) for selecting semi parametric multivariate copula models which the marginal distributions are unspecified. Genest and Rivest (1993) proposed a method in Archimedean copulas as below,

$$K_\theta(t) = P(C(u, v\,|\theta) < t$$

with its non-parametric estimation K_n, given by

$$K_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\,(e_{jn} \le t)$$

where $e_{jn} = (1/\mathrm{n})\sum_{k=1}^{n} \mathbf{1}\,(X_{1k} \le X_{1j}, \dots, X_{pk} \le X_{pj})$. A copula that the function K_θ is closest to K_n, is the best one.

Choosing the best copula with minimizing the distance (L^2-norm, Kolmogorov, etc) from K_θ to the non-parametric estimation K_n suggested by Durrleman et al. (2000). Genest et al. (1995) proposed a GOF test statistic with a non-truncated version of Kendall's process,

$$\mathbb{K}_n(t) = \sqrt{n}\left\{K_n(t) - K_{\theta_n}(t)\right\}$$

where θ_n denotes a robust estimation of θ. The expression for the statistic is simple and the test has nice properties. An easy way to construct GOF tests for copulas is to consider *p*-dimensional, χ^2 tests. The methodology is presented in Pollard (1979). We recall that the standard χ^2 -test uses the below test-statistic,

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{[o_{ij} - e_{ij}]^2}{e_{ij}}$$

where *n* is the sample size, *I* and *J* are the numbers of classes. o_{ij} is the observed frequency of data in class ij and e_{ij} is the theoretical frequency of data for every class ij. Note also that the power of test increases with growing number of the classes. The main criticism about this approach concerns the arbitrary choice of the subsets that divide the *p*-dimensional space $[0,1]^p$ (Kendall and Stuart, 1983). There are many researchers that have used this method in their activities. As examples, Dobric and Schmidt (2004) used this method in a financial application. Çelebioğlu (2003) in modeling student grades relies on this method. Najjari & Ünsal (2012) used this method in modeling meteorological data, Najjari et al. (2014) applied this method in modeling data of the Danube river, also Şahin Tekin et al. (2014) used this method in their simulation study on copulas.

3. THE SELECTION OF THE RIGHT COPULA

In the Section 2 all of the mentioned tests for selecting the right copula, however, rely on previous estimation of an

[1] Dimensions

optimal parameter set of copulas. The main aim of this section is to concentrate on a method in estimating copula parameters and also selecting the right copula which is independent of the chosen optimal parameter. The method is described for two dimensional data (X, Y) as follows:

After testing time dependency of data, the random samples $(X_1, Y_1), \dots, (X_n, Y_n)$ are converted into normalized ranks in the usual fashion by setting $U_l = rank(X_l)/n$ and $V_l = rank(Y_l)/n$ for each $l \in \{1,2, \dots, n\}$. Then by using r = round($\sqrt[4]{n}$), the data are grouped into $r \times r$ contingency table. Let matrix $O_{I\times J}$, consists of observed frequencies and matrix $E_{I\times J}$ consists of estimation of the expected frequencies $(I = J = r)$ and o_{ij} is an element of i th row and j th column of the matrix O and e_{ij} is expected value of the o_{ij}, calculated by multiplying the number of observations n with appropriate theoretical frequency estimated with copulas, where $i, j \in \{1,2, \dots, n\}$. Let copula parameters θ are in the form $\theta = (\theta_1, \dots, \theta_s)$ where s is the number of copula parameters. Under the assumption that the true underlying

copula is of the form C_θ for some $\theta \in \Theta$ (where Θ is parameters space), expected frequencies $e_{ij}(\theta)$ are computed for the contingency table. So $e_{ij}(\theta)$ is a function of the parameters θ and can be calculated as follows,

$$e_{ij}(\theta) = n \times [C_\theta(u_i, v_j) - C_\theta(u_{i-1}, v_j) - C_\theta(u_i, v_{j-1}) + C_\theta(u_{i-1}, v_{j-1})]$$

where $i, j \in \{1,2, \dots, n\}$ and $e_{11}(\theta) = n \times C_\theta(u_1, v_1)$. So standard GOF statistic value is as follows,

$$h(\theta) = \chi^2_\theta = \sum \frac{[o_{ij} - e_{ij}(\theta)]^2}{e_{ij}(\theta)} . \qquad (3.1)$$

In order to determine appropriate copula, obviously estimation of the expected frequencies calculated by copula must be close to related observed frequencies, and it occurs in the value $\hat{\theta}$ (estimation of θ) that minimizes χ^2_θ in (3.1). Meanwhile, H_0: $C_{\hat{\theta}} \in C_\theta$ is rejected if $\chi^2_{\hat{\theta}}$ in (3.1) yields a low p-value by reference to the chi-squared distribution with $(r-1) \times (r-1)$ degrees of freedom. This means that by calculating minimum point of the function $h(\theta)$ in the range of copula parameters, it is possible to reach to the both aims, estimating the copula parameters and also choosing the right copula that best fits to data. Obviously calculating minimum value of the function $h(\theta)$ is not complicated and also without loss of generality it can be assumed that the minimum value is accrued only on a single point of its parameters range. This method is easily applicable in multiparameter copulas, and also it is able to capture the best fit of copulas to data, so it has advantage to other estimation methods of copula parameters.

4. APPLICATION

In this section the discussed method is put into competition with the most used MLE and also nonparametric methods in estimation of copula parameters. Data set comes from $n = 420$ monthly minimum and maximum QFE[2] records between years 1975 and 2010 in Tehran (available online at I.R Of Iran Meteorological Organization). All data set were tested for serial (temporal) independence (not simply by checking significance of autocorrelations but) primarily by a test based on empirical copulas described in Kojadinovic and Yan (2010). To avoid decision about univariate distributions, the observations were transformed to unit interval by their corresponding empirical distribution functions, see Figure 1 for the resulting pairs. Kendall's tau for these data is $\tau = 0.3894$.

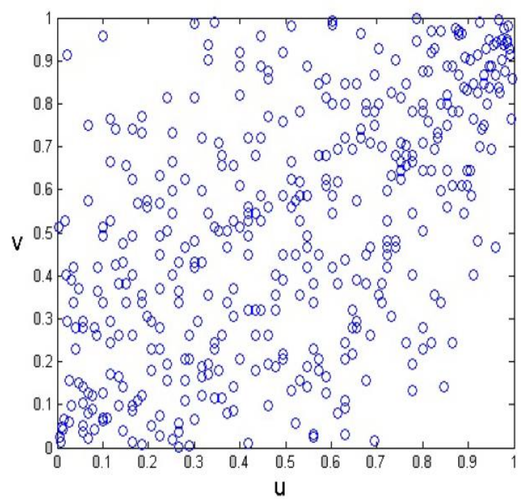

Figure 1. Scatterplots for minimum and maximum QFE data.

In modeling dependence between minimum and maximum QFE, we consider five widely used Archimedean families of copulas: Clayton, Gumbel and A12, A14 (these copula families have numbered 4.2.12 and 4.2.14 respectively in Nelsen, 2006) and also coth copula. coth copula has hyperbolic generator and recently presented to the literature by Najjari et al. (2014). Details of the mentioned copulas are shown in Table 1.

[2] QFE is the barometric altimeter setting that will cause an altimeter to read zero at the reference datum of a particular airfield (in general, a runway threshold). In ISA temperature conditions the altimeter will read height above the datum in vicinity of the airfield.

Table 1. Definition and parameter domain of the copulas used in this paper

Family	$C(u,v) =$	$\tau =$	$\tau \in$	$\theta \in$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\frac{\theta}{\theta+2}$	(0,1)	[-1,∞)-{0}
Gumbel	$exp\left(-\left[(-logu)^{\theta} + ((-logv)^{\theta})\right]^{1/\theta}\right)$	$\frac{\theta-1}{\theta}$	[0,1)	$[1,\infty)$
A12	$(1 + [(u^{-1} - 1)^{\theta} + (v^{-1} - 1)^{\theta}]^{1/\theta})^{-1}$	$1 - \frac{2}{3\theta}$	$[\frac{1}{3},1)$	$[1,\infty)$
A14	$(1 + [(u^{-1/\theta} - 1)^{\theta} + (v^{-1/\theta} - 1)^{\theta}]^{1/\theta})^{-\theta}$	$1 - \frac{2}{1+2\theta}$	$[\frac{1}{3},1)$	$[1,\infty)$
coth	$\frac{1}{\theta}\text{acoth}(coth(\theta u) + coth(\theta v) - coth(\theta))$	$1+\frac{2}{\theta^2} - \frac{2}{\theta}coth(\theta)$	$(\frac{1}{3},1)$	$(0,\infty)$

Note: A12, A14 families have numbered as 4.2.12 and 4.2.14 in Table 4.1 Nelsen (2006).

As it is mentioned in Section 3, to apply GOF test, by using relation $\sqrt[4]{n}$, range of the two variables were divided into 4 intervals for each, so the critical point in GOF test is $\chi^2_{0.05,\ 9}$=16.9190.

Parameters for each copula families (mentioned in Table 1) estimated with the discussed method, maximum likelihood and also nonparametric method by using $\tau = 0.3894$, then GOF statistic value is calculated by the estimated parameters. Meanwhile, results are summarized in Table 2.

In Table 2, it is seen that by the discussed method, MLE and also nonparametric estimation methods, Gumbel family has the best fit to data, i.e., the GOF test statistic values is less than the critical point $\chi^2_{0.05,\ 9} = 16.9190$. The discussed method estimates Gumbel copula's parameter as 1.59 with $\chi^2 = 8.4987$, while by MLE they are $\hat{\theta} = 1.56$ and $\chi^2 = 8.5931$. The corresponding values for the nonparametric estimation method are $\hat{\theta} = 1.64$ and $\chi^2 = 8.8204$. Clearly the lowest GOF test value belongs to the new method and it means that the best fit to data is estimated by the discussed method.

Although the discussed method tries to minimize the value of GOF test, unfortunately the other four copula families have not compatibility with the mentioned data. It is notable that also for the GOF test in all the other four families, the minimum values of GOF test belong to the discussed method. For example in the Clayton family, GOF test value by the discussed method is 46.287, while these values are 48.2855 and 53.4373 by the MLE and nonparametric estimation methods, respectively.

Table 2. Copula parameters estimations and GOF test values

Family	Discussed Method		MLE		Nonparametric	
	$\hat{\theta}$	χ^2	$\hat{\theta}$	χ^2	$\hat{\theta}$	χ^2
Clayton	0.91	46.287	0.74	48.2855	1.28	53.4373
Gumbel	1.59	8.4987	1.57	8.5931	1.64	8.8204
A12	1.11	40.9745	1.08	41.2886	1.09	41.0929
A14	1.23	34.1538	1.07	39.7557	1.14	35.8295
coth	1.38	37.9248	1.57	38.5943	1.19	38.3566

Obviously, in the discussed method, $\hat{\theta}$ (estimation of θ) minimizes χ^2_{θ} in (3.1) and in Table 2 this fact clearly is seen. With the discussed method, any five copula families given in Table 1 have minimum GOF test values and this means that there are more compatibility between the observations and expected theoretical frequencies estimated with copulas.

5. CONCLUSION

This study discusses GOF test to be as a function of copula parameters. Then by calculating minimum point of this function in the range of copula parameter, it is possible to have an estimation for the copula parameters and also choosing the right copula that fits the best to data. In the application section, data set comes from n = 420 monthly minimum and maximum barometric data, and by using the most used five copula families (mentioned in Table 1) it is tried to model the dependency of the data. Results show that with the discussed method, there are more compatibility between the observations and expected theoretical frequencies estimated with any five copula families given in Table 1. There are two advantages of the discussed method, it is easily applicable in multiparameter copulas, and also it captures the best fit of copulas to data.

ACKNOWLEDGEMENTS

The work on this paper was supported by grants of the Islamic Azad University-Maragheh Branch. The author thanks anonymous referees for insightful comments on the previous versions of the manuscript.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] B. Brahimi., A. Necir., A semiparametric estimation of copula models based on the method of moments, Statistical Methodology .9 (2012) 467-477.

[2] X. Chen., Y. Fan., Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection, La Revue Canadienne de Statistique. 33 (3) (2005) 389-414.

[3] S. Çelebioğlu., Archimedean copulas and an application, Selcuk University Journal of Science, 22 (2003) 43-52.

[4] J. Dobric., F. Schmidt., Testing goodness of fıt for parametric families of copulas application to fınancial data. Seminar of economic and social statistics, University of Cologne. 2004.

[5] V. Durrleman., A. Nikeghbali., T. Roncalli., Which copula is the right one? Working document, Groupe de Recherche Operationnelle, Credit Lyonnais, 2000.

[6] C. Genest., Frank's family of bivariate distributions, Biometrika, 74 (1987) 549-555.

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

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

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

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

[11] M. Kendall., A. Stuart., The Advanced Theory of Statistics, fourth ed. Oxford University Press, New York, vol 2, 1983.

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

[13] Kojadinovic., J. Yan., Tests of serial independence for continuous multivariate time series based on a Mobius decomposition of the independence empirical copula process, Ann. Inst. Stat. Math, 2010.

[14] V. Najjari., T. Bacigal., H. Bal., An Archimedean Copula Family with Hyperbolic Cotangent Generator, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 22(5):(2014) 761-768.

[15] V. Najjari., M. G. Ünsal., An Application of Archimedean Copulas for Meteorological Data, G. U. Journal of Science, 25:2 (2012) 301-306.

[16] R. B. Nelsen., An introduction to copulas, Springer, New York, Second edition, 2006.

[17] D. Pollard., General chi-square goodness-of-fıt tests with data-dependent cells. Z. Wahrscheinlich keitstheorie und verwandte Gebiete, 50 (1979) 317-331.

[18] A.Sklar., Fonctions de rèpartition á n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris, 8 (1959) 229-231.

[19] S. T. Şahin Tekin., V. Najjari., H. H. Örkcü., Simulation study on copulas, Sahand Communications in Mathematical Analysis, 1(2) (2014) 55-63.

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

[21] J. Yan., Enjoy the joy of copulas: with a package copula, J. Stat. Software, 21(4) (2007) 1-21.