Sigma Journal of Engineering and Natural Sciences, Vol. 42, No. 4, pp. XX-XX, August, 2024
Yildiz Technical University Press, Istanbul, Turkey

# Modeling asymmetrically dependent automobile bodily injury claim data using Khoudraji Copulas

Emel KIZILOK KARA[1], Sibel AÇIK KEMALOĞLU[2,*]




## ABSTRACT

Linear-dependent variables are typically modeled through the Spearman correlation, a classical statistical technique. In reality, the dependence between the data cannot always be linear. The copula approach has often been a popular tool for modeling dependent data in these cases. Archimedean copulas, which can model mostly symmetrical data, are also among the copula families used for this purpose. Recently, asymmetric copula models have been developed to model unsymmetrical-dependent variables. The dependency measure is calculated using directional dependency coefficients instead of the Spearman correlation when the data is asymmetrical. Appropriate asymmetric model selection is made with the help of these measurements.

In the study, first, dependency parameters corresponding to different Spearman coefficients were obtained for Archimedean copula families, and asymmetric copulas were derived from them. Then, simulation data were obtained for these parameter values to determine the effect of asymmetry on data modeling, and directional dependency measures were found. In addition, the study methodology was applied to automobile bodily injury claims data, which is a real dataset with an asymmetric structure. Here, we used two different asymmetric models: the Khoudraji copula KC models, which are created by multiplying independent and Archimedean copulas, and the LCC models, which are linear-convex combinations of Archimedean copulas. Finally, the appropriate model was selected according to the directional dependency coefficients, and the results were interpreted.


***Keywords:*** *Asymmetric Copula; Archimedean Copula; Directional Dependence; Automobile Bodily Injury Claims Data*

## INTRODUCTION

Determining the dependency structure of variables is important in many areas of research, including statistics, finance, engineering, and actuaries. In deciding this dependence structure, the Spearman correlation, which is one of the standard statistical methods, is used. This method can model variables with a normal distribution, i.e., linear dependent variables. However, the dependency coefficient may often differ in the real data's lower and upper tail regions. It may also be affected by whether the data are symmetrical or asymmetrical. For situations where the correlation becomes complex, known statistical models are insufficient. Analytical methods have recently been developed for this purpose. In this sense, copulas, first proposed by Sklar [1], are popular methods used to model dependent variables.

Copulas are preferred because they allow for more realistic data modeling and are capable of generating joint distributions for statistical modeling of dependent variables without restrictions on the marginal distributions of each variable. Nelsen [2], Salvadori and De Michele [3], Genest and Favre [4], Joe [5], Durante and Sempi [6], and Hong et al. [7] showed statistical behavior of copulas as dependent variables.

Moreover, there are studies on the use of copulas in fields such as hydrology [8-10], earthquakes [11], finance [12-15], wind [16], ocean [17, 18], climate science [19], and bioinformatics [20, 21].

*This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkilic*
*[1] Department of Actuarial Science, Kırıkkale University, Kırıkkale, Türkiye*
*[2] Department of Statistics, Ankara University, Ankara, Türkiye*
**E-mail address: sibelacik@gmail.com*
*Orcid id: https://orcid.org/0000-0001-7580-5709 Emel Kizilok Kara, 0000-0003-0449-6966 Sibel Acik Kemaloglu*
*Manuscript Received 03 February 2023, Revised 12 March 2023, Accepted 13 May 2023*

On the other hand, most parametric copula models, such as Archimedean copulas, can only be applied to data with symmetric dependence. Some authors have highlighted this shortcoming of existing copula approaches in their studies, such as Genest and Favre [4], Kim et al. [20], and Sungur [22,23]. Indeed, most data have an asymmetric dependence structure. In such data, ignoring asymmetry affects the identification of the dependence structure and subsequent calculations. Some authors such as [24-28] have recently contributed to the development of asymmetric copula construction to eliminate this deficiency. These include various techniques used in multivariate data modeling to capture asymmetric dependence.

Our study focuses on the asymmetric Khoudraji copula (KC) and linear convex combination (LCC) copula families, which can be easily generated using Archimedean copulas. Khoudraji copulas were first developed by Khoudraji [29] and consist of Archimedean copula families and independent (product) copula families. Subsequently, authors such as Nelsen [2], Rodríguez-Lallena and Ubeda-Flores [24], Klement and Mesiar [25], Liebscher [30], Durante [31], Quessy and Kortbi [32], Siburg et al. [33], and Bezak et al. [34] developed asymmetric copulas. In addition, Zhang et al. [35] examined several asymmetric copula functions capable of modeling both linear and nonlinear asymmetric dependence structures using Khoudraji copulas between ocean variables. Moreover, Zhang et al. [36] demonstrated the advantages of asymmetric copulas with Khoudraji copulas and compared them with traditional copula approaches for modeling site soil data. Besides, Lin et al. [37], Bai et al. [38], and Huang and Dong [39] compared the performance of symmetric copula, Khourdaji copula, and traditional conditional modeling methods on bivariate wave data. LCC copulas are also constructed from Archimedean copulas and their linear convex combinations. Authors such as Ma and Zhang [18], Siburg [33], and Wu [40] have investigated the asymmetric properties of these asymmetric families. Recently, some authors ([12], [15], [41]) have also shown that asymmetric copulas provide more realistic and accurate results when modeling asymmetric multivariate data.

While model selection with classical models considers well-known selection criteria such as AIC, KS Cramer-von Mises, and MSE, model selection with asymmetric models uses directional dependence measures calculated based on conditional copula functions. Model selection with directional dependence of copulas is a statistical approach that involves choosing the most appropriate copula function to describe the dependency structure between two or more random variables. The importance of this study is the selection of the most appropriate model with the directional dependency method, which considers the direction of the relationship, among the asymmetric models created with Khoudraji copulas. The advantage of this method is that it allows more accurate and flexible modeling of the dependency structures of asymmetric data.

This study analyzed data on "automobile bodily injury claims" from CASdataset [42]. Frees and Wang [43] modeled these data using the classical copula approach. However, since the data are both dependent and asymmetric, it is necessary to work with asymmetric-dependent models. To this end, unlike the previous study, we apply the asymmetric copula approach to this dataset and determine appropriate models according to directional dependence measures. Thus, this modeling improves the risk assessment and decision-making process by providing valuable information and tools for insurers, researchers, and other stakeholders in the auto insurance industry.

The remainder of this article is organized as follows. The first section introduces copulas and asymmetric copula models. The Archimedean copulas used in this study and the Khoudraji asymmetric copula models derived from them are given. The parameter values corresponding to Spearman correlations were found. The symmetric and asymmetric measurements were analyzed, and the results obtained by simulation for these parameters are presented in tables and graphs. In the next section, one-parameter and two-parameter KCC and LCC models are applied to "Automobile Bodily Injury Claims" data, and model selection is performed according to directional dependence measures. The results are discussed in the last section.

## COPULA THEORY AND MEASURES OF ASYMMETRY

### Definition of the Copula

A copula is a mathematical function that describes the dependence structure between multiple variables, regardless of their marginal distributions. The idea behind copulas is to separate the modeling of the dependencies between variables from the modeling of the marginal distributions of individual variables. This allows for more flexibility in modeling complex dependencies, especially when the underlying data do not follow a normal distribution. Copulas have been widely used in various fields, such as finance, economics, and actuarial science, to model multivariate data and estimate the risk of various financial instruments.

The Sklar theorem, first introduced by economist Sklar [1], plays an important role in copula theory, which is a powerful tool for modeling the dependencies between variables in multivariate data. It is defined as

Sklar's Theorem: Let H be an n-dimensional distribution function with marginal distributions $F_1,\ldots,F_n$. An n-dimensional copula C for all $x \in R_n$ is given by

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{1}$$

If $F_1,\ldots,F_n$ are continuous, then C is unique. Conversely, if $F_1,\ldots,F_n$ are marginal distribution functions and C is a copula, a function $H(x_1, \ldots, x_n)$ with marginal distributions $F_1,\ldots,F_n$ is defined by Eq. (1).

The theorem states that any multivariate distribution function can be represented as the copula of its marginal distribution functions and a uniformly distributed random variable on the unit hypercube. This result allows for flexible modeling of the dependence structure between different variables and has important applications in fields such as finance, insurance, and actuarial science.

A unique bivariate copula $C: [0,1]^2 \to [0,1]$ is defined as $C(u,v) = F_{XY}(F_X^{-1}(u), F_Y^{-1}(v))$, where $F_X^{-1}(u)$ and $F_Y^{-1}(v)$ are the inverse distribution functions of $X$ and $Y$, respectively.

## Asymmetric Copula Models

Different types of copulas have been proposed in the literature, such as Archimedean copulas, elliptical copulas, extreme-value copulas, vine copulas, and empirical copulas. Many commonly used copula families, such as the Gaussian, Clayton, Gumbel, and Frank copulas, have the property of exchangeability, which means that the copula function is symmetric with respect to its arguments. This means that $C(u,v) = C(v,u)$ for all $u, v \in [0,1]$. This property is desirable because it allows a simpler and more intuitive interpretation of the dependence structure. However, this can be a limitation if the data have asymmetric dependence.

To eliminate this deficiency, recently, asymmetric copulas have been constructed in various ways. One simple way to construct an asymmetric copula is by using the rotation method. In this method, using a rotation matrix R, the standard copula is transformed into a new copula with a different dependency structure. The rotation is then defined as $C_R(u,v) = C(R * (u,v))$ where $C_R$ is the rotated copula and $R * (u,v)$ is the rotation of the standard coordinates.

Ma and Zhang [18], and Zhang et al. [35] have described and implemented other methods of creating an asymmetric copula. This study focuses on Khoudraji copulas (KC), developed by multiplying copulas, and LCC copulas constructed using their linear convex combinations. KC and LCC copulas have been widely used in different fields and applications, particularly in finance, in modeling dependency structures in various types of data. They have also been used in bioinformatics, environmental sciences, and engineering to model the dependence between variables. These studies fit the empirical data better than some traditional copulas when the data exhibit asymmetric dependence.

Our study uses these models to analyze automobile bodily injury claim data and determine the best-fit copula models. We choose appropriate models to analyze automobile bodily injury claim data and make inferences.

In the following section, we introduce the mathematical form of the KC and LCC models.

### *Khoudraji copula (KC) model:*

Khoudraji copulas are a class of asymmetrical copulas first introduced by Khoudraji [29]. Later, Liebscher [30] defined its general form. They may model both positive and negative dependency, and a wide variety of dependency constructs can be captured by varying their parameters.

The mathematical model of a Khoudraji copula function $C_{\alpha,\beta}: I^2 \to I$ is defined as

$$C_{\alpha,\beta}(u,v) = C_1\left(u^{\bar{\alpha}}, v^{\bar{\beta}}\right) C_2\left(u^{\alpha}, v^{\beta}\right) \tag{2}$$

where $\phi = (\alpha, \beta, \theta)$, $\alpha, \beta \, \epsilon (0,1)$, $\alpha \neq 1/2$, $\beta \neq 1/2$, $\alpha + \bar{\alpha} = 1$, $\beta + \bar{\beta} = 1$.

In Eq. (1), if $C_1$ is independent copula $C(u,v) = uv$, and $C_2$ is a symmetrical Archimedean copula family with a dependency parameter $\theta$, $C_{\alpha,\beta}$ is called a Khoudraji copula. For $\beta = 1 - \alpha$, one parameter Khoudraji copula families (KC1-model),

$$C_{KC1}(u,v) = u^{1-\alpha} v^{\alpha} C(u^{\alpha}, v^{1-\alpha}) \tag{3}$$

and for $\alpha \neq \beta$, two-parameter Khoudraji copula families (KC2-model)

$$C_{KC2}(u,v) = u^{1-\alpha} v^{1-\beta} C\left(u^{\alpha}, v^{\beta}\right) \tag{4}$$

are expressed as in Eqs. (3) and (4), respectively.

*Linear convex combination (LCC) model:*

It is possible to create asymmetrical copulas using linear and convex combinations of the copulas. However, the resulting pattern remains a symmetrical copula when direct linear-convex combinations are produced with symmetric copula functions. Using the method in Wu [40], basic copulas can be modified to include asymmetrical features. This method involves deriving a new asymmetric copula by modifying the basic copulas using a weighting function that allows for asymmetry. A new asymmetric copula can be derived using this methodology as follows:

$$\breve{C}_k(u_1, \dots, u_n) = C(u_1, \dots, u_{k-1}, 1, u_{k+1}, \dots, u_n) - C(u_1, \dots, u_{k-1}, 1 - u_k, u_{k+1}, \dots, u_n) \tag{5}$$

where $C(\cdot)$ is the n-dimensional base copula.

An LCC copula can be constructed to capture the asymmetric characteristics of a multivariate variable based on Eq. (6) as follows:

$$C_{LCC}(u_1, \dots, u_n) = \sum_{k=0}^{n} p_k \breve{C}_k(u_1, \dots, u_n) \tag{6}$$

where $0 \leq p_k \leq 1$ and $\sum_{k=0}^{n} p_k = 1$. Thus, an asymmetric copula can be constructed by linear convex combinations of $\breve{C}_k(\cdot)$. There are many copula families that can be specified for base copula C. For instance, a bivariate copula $C(u,v)$ according to Eq. (5) can be written as follows:

$$C(u,v) = u + v - 1 + C_0(1-u, 1-v)$$
$$\breve{C}_1(u,v) = v - C(1-u, v) = u - C_0(u, 1-v)$$
$$\breve{C}_2(u,v) = u - C(u, 1-v) = v - C_0(1-u, v)$$

Thus, the constructed one-parameter (LCC1-model) and two-parameter (LCC2-model) asymmetric copulas by linear combination can be given in Eqs. (7) and (8), respectively.

$$C_{LCC1}(u,v) = p_0 C(u,v;\theta_1) + p_1 \breve{C}_1(u,v;\theta_2) \tag{7}$$
$$C_{LCC2}(u,v) = p_0 C(u,v) + p_1 \breve{C}_1(u,v) + p_2 \breve{C}_2(u,v) \tag{8}$$

**Directional Dependence Measures**

Directional dependence refers to the ability to measure the degree and direction of dependence between two or more variables. When the coefficient of association between variables is linear, it can be measured by the Spearman correlation, whereas when it is nonlinear, it can be measured by the Spearman coefficient based on the copula, expressed as follows:

$$\rho_C = 12 \iint_0^1 C(u,v)\,du\,dv - 3 \tag{9}$$

Therefore, $\rho_C^2$ can be used to calculate the ratio of variables explaining each other. However, when there is asymmetricity in the data structure, the directional dependence coefficients determined according to regression-based copula functions will not be the same relative to each other. Therefore, this situation will cause the ratio of the explained variance to be different.

Accordingly, when dependence is symmetric, the regression functions for $U$ and $V$ have the same linear form, and the same model can be used to predict both $U$ and $V$. However, when the dependence is asymmetric, the regression functions for $U$ and $V$ will not be the same, and different models will be needed to estimate the regression functions for $U$ and $V$ separately. Detailed information on directional dependence can be found in Sungur [22, 23], Jung et al. [44], and Kim and Kim [45].

The directional dependence coefficients using the copula regression functions (in the directions of $U$ to V ($U \rightarrow V$) and $V$ to U ($V \rightarrow U$)) can be obtained by an approximate calculation method as follows:

$$\tilde{\rho}_{U\rightarrow V}^{(2)} = \frac{12}{S}\sum_{s=1}^{S}\left(\tilde{r}_{V|U}(u_s)\right)^2 - 3 \tag{10}$$

and

$$\tilde{\rho}_{V\rightarrow U}^{(2)} = \frac{12}{S}\sum_{s=1}^{S}\left(\tilde{r}_{U|V}(v_s)\right)^2 - 3 \tag{11}$$

where,

$$\tilde{r}_{V|U}(u) = 1 - \frac{1}{S}\sum_{s=1}^{S} C_u(v_s) \text{ and } C_u(v) \equiv P(V \le v|U = u) = \frac{\partial C(u,v,\Phi)}{\partial u},$$

$$\tilde{r}_{U|V}(v) = 1 - \frac{1}{S}\sum_{s=1}^{S} C_v(u_s) \text{ and } C_v(u) \equiv P(U \le u|V = v) = \frac{\partial C(u,v,\Phi)}{\partial v},$$

$\Phi$ is the parameter set. $\tilde{r}_{V|U}(u)$ and $\tilde{r}_{U|V}(v)$ are approximately calculated copula regression functions over the pseudo-observations, $(u_s, v_s) \in (0,1)^2$ for the size of the pseudo-observation, S.

### Simulation Study of Archimedean and Khoudraji Copula Models

A simulation study was conducted to examine the symmetrical and asymmetrical dependency structures. Archimedean copulas for symmetrical models and Khoudraji copulas for asymmetrical models are considered. First, dependency parameter values corresponding to various correlation values were obtained for the mentioned models (Table 1), and parameter estimates of 1000 data pairs produced with these parameters were made (Table 2). Additionally, the asymmetry test was applied to these simulated data using Cramer-von Misses statistics. As shown in Table 3, simulated data from symmetric models are symmetric (p-values > 0.05), while simulated data from asymmetric models are asymmetric (p-values < 0.05).

**Table 1.** Parameter values of symmetric and asymmetric copulas for various correlations

| $\rho$ | Clayton | KC1-Clayton | | Frank | KC1-Frank | | Gumbel | KC1-Gumbel | |
|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\alpha$ | $\theta$ | $\theta$ | $\alpha$ | $\theta$ | $\theta$ | $\alpha$ | $\theta$ |
| 0.1 | 0.143 | 0.122 | 2.8 | 0.603 | 0.333 | 1.9 | 1.072 | 0.274 | 1.2 |
| 0.2 | 0.311 | 0.287 | 2.7 | 1.224 | 0.459 | 3.7 | 1.156 | 0.160 | 3.0 |
| 0.3 | 0.511 | 0.352 | 5.6 | 1.883 | 0.511 | 6.9 | 1.257 | 0.246 | 6.099 |
| 0.4 | 0.759 | 0.492 | 14.5 | 2.610 | 0.507 | 19.7 | 1.382 | 0.611 | 9.599 |

| $\rho$ | Clayton | KC2-Clayton | | | Frank | KC2-Frank | | | Gumbel | KC2-Gumbel | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\alpha$ | $\beta$ | $\theta$ | $\theta$ | $\alpha$ | $\beta$ | $\theta$ | $\theta$ | $\alpha$ | $\beta$ | $\theta$ |
| 0.2 | 0.311 | 0.4 | 0.6 | 1.9 | 1.224 | 0.4 | 0.7 | 3.4 | 1.156 | 0.2 | 0.4 | 3.8 |
| 0.4 | 0.759 | 0.6 | 0.9 | 2 | 2.610 | 0.9 | 0.9 | 3.1 | 1.382 | 0.6 | 0.9 | 1.7 |
| 0.6 | 1.505 | 0.8 | 0.9 | 2.9 | 4.466 | 0.7 | 0.9 | 8.5 | 1.755 | 0.7 | 0.8 | 3.1 |
| 0.8 | 3.185 | 0.9 | 0.9 | 7.1 | 7.902 | 0.9 | 0.9 | 13.8 | 2.581 | 0.9 | 0.9 | 4.1 |

**Table 2.** Parameter estimation results for data pairs generated by simulation ($N = 1000$)

| $\rho$ | Clayton | KC1-Clayton | | | Frank | KC1-Frank | | | Gumbel | KC1-Gumbel | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\alpha$ | $\theta$ | $\hat{\rho}$ | $\theta$ | $\alpha$ | $\theta$ | $\hat{\rho}$ | $\theta$ | $\alpha$ | $\theta$ | $\hat{\rho}$ |
| 0.1 | 0.1397 | 0.1053 | 4.7777 | 0.0907 | 0.5827 | 0.8922 | 6.9387 | 0.0913 | 0.7191 | 0.1407 | 1.2546 | 0.0935 |
| 0.2 | 0.3356 | 0.2147 | 4.3232 | 0.2007 | 1.2279 | 0.3412 | 4.5981 | 0.2001 | 1.1743 | 0.1728 | 3.0878 | 0.2033 |
| 0.3 | 0.4927 | 0.3915 | 6.9374 | 0.2993 | 1.8870 | 0.5441 | 6.9271 | 0.2970 | 1.2605 | 0.2368 | 5.7554 | 0.3024 |
| 0.4 | 0.7326 | 0.4964 | 13.0285 | 0.4052 | 2.7652 | 0.5165 | 20.6832 | 0.3935 | 1.4015 | 0.6148 | 9.5974 | 0.4033 |

| $\rho$ | Clayton | KC2-Clayton | | | | Frank | KC2-Frank | | | | Gumbel | KC2-Gumbel | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\theta}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\theta}$ | $\hat{\rho}$ | $\hat{\theta}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\theta}$ | $\hat{\rho}$ | $\hat{\theta}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\theta}$ | $\hat{\rho}$ |
| 0.2 | 0.3387 | 0.3354 | 0.9999 | 1.6262 | 0.2044 | 1.2498 | 0.4036 | 0.9965 | 2.7880 | 0.2050 | 1.1670 | 0.2385 | 0.4178 | 3.3837 | 0.2074 |
| 0.4 | 0.7554 | 0.5809 | 0.9924 | 1.8154 | 0.3957 | 2.4613 | 0.7612 | 0.8330 | 3.9251 | 0.4183 | 1.3966 | 0.5867 | 0.8345 | 1.8024 | 0.3970 |
| 0.6 | 1.5020 | 0.7857 | 0.8652 | 2.9955 | 0.5936 | 4.5125 | 0.7183 | 0.9114 | 8.2398 | 0.6039 | 1.7922 | 0.7433 | 0.8660 | 2.6038 | 0.6050 |
| 0.8 | 3.1726 | 0.8895 | 0.8963 | 7.0759 | 0.7897 | 7.9359 | 0.9098 | 0.8739 | 14.5948 | 0.7998 | 2.5988 | 0.9206 | 0.8845 | 4.0058 | 0.7934 |

**Table 3.** Asymmetry test results of simulated data from symmetric and asymmetric models: Cramer-von Misses statistics and p-values

| Symmetric Archimedean models | | | | Asymmetric KC1 models | | | |
|---|---|---|---|---|---|---|---|
| $\rho$ | Clayton | Frank | Gumbel | $\rho$ | KC1-Clayton | KC1-Frank | KC1-Gumbel |
| 0.1 | 0.024344<br>(0.6888) | 0.02432<br>(0.7218) | 0.048299<br>(0.1154) | 0.1 | 0.077899<br>(0.02647) | 0.092821<br>(0.008492) | 0.089669<br>(0.004496) |
| 0.2 | 0.020519<br>(0.7947) | 0.013828<br>(0.9805) | 0.018285<br>(0.8866) | 0.2 | 0.07341<br>(0.02348) | 0.080107<br>(0.01449) | 0.42153<br>(0.0004995) |
| 0.3 | 0.030133<br>(0.3921) | 0.018385<br>(0.8536) | 0.016769<br>(0.9066) | 0.3 | 0.092927<br>(0.005495) | 0.05489<br>(0.04745) | 0.54746<br>(0.0004995) |
| 0.4 | 0.023138<br>(0.6119) | 0.031925<br>(0.3322) | 0.035001<br>(0.2143) | 0.4 | 0.063067<br>(0.02048) | 0.076384<br>(0.00649) | 0.31175<br>(0.0004995) |
| **Symmetric Archimedean models** | | | | **Asymmetric KC2 models** | | | |
| $\rho$ | Clayton | Frank | Gumbel | $\rho$ | KC2-Clayton | KC2-Frank | KC2-Gumbel |
| 0.2 | 0.0237<br>(0.7008) | 0.0230<br>(0.7068) | 0.0462<br>(0.1234) | 0.2 | 0.0866<br>(0.0085) | 0.0739<br>(0.0245) | 0.0787<br>(0.0145) |
| 0.4 | 0.0170<br>(0.8417) | 0.0255<br>(0.504) | 0.0305<br>(0.3072) | 0.4 | 0.1329<br>(0.0015) | 0.0816<br>(0.0025) | 0.0815<br>(0.0035) |
| 0.6 | 0.0174<br>(0.6838) | 0.0234<br>(0.4241) | 0.0101<br>(0.9945) | 0.6 | 0.0648<br>(0.0025) | 0.0805<br>(0.0015) | 0.0663<br>(0.0015) |
| 0.8 | 0.0097<br>(0.9276) | 0.0214<br>(0.1993) | 0.0153<br>(0.515) | 0.8 | 0.0372<br>(0.0195) | 0.0339<br>(0.0245) | 0.0338<br>(0.0335) |

Contour plots showing symmetrical and asymmetrical dependency structures are visually presented in Figure 1 for $\rho = 0.4$. These graphs show that the ones drawn for the Archimedean copulas are symmetrical, and those drawn for the KC1 and KC2 models are irregular and asymmetrical.

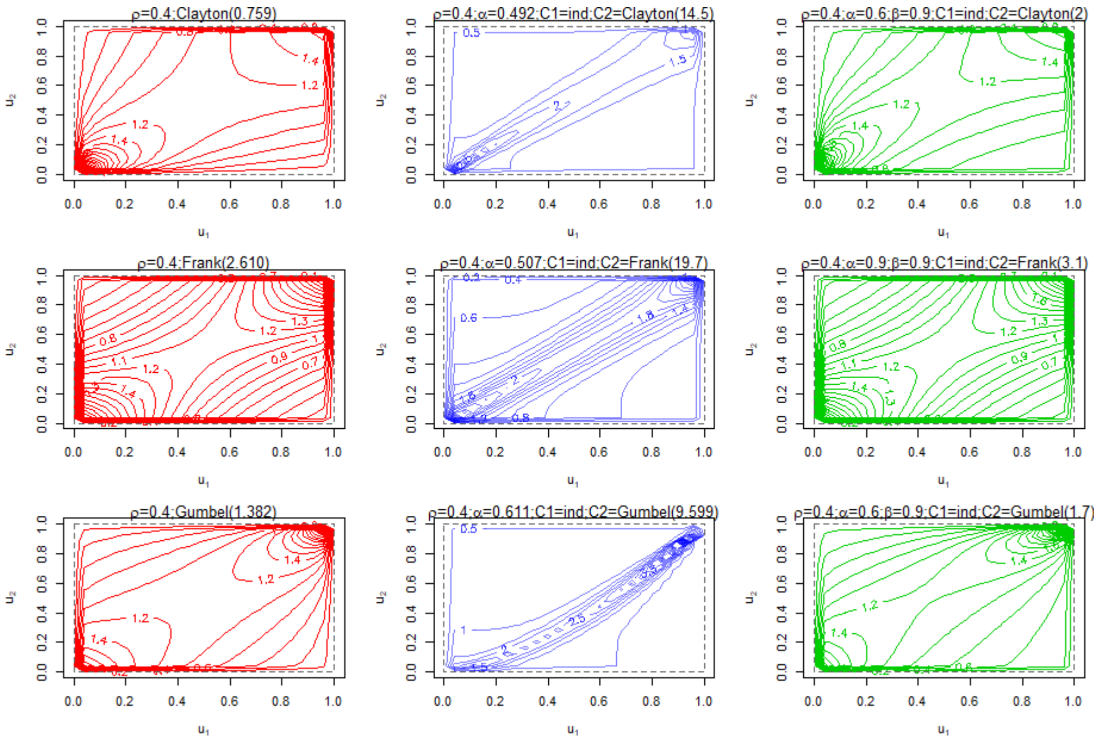


**Figure 1.** Contour plots for the Archimedean copula (red), KC1 (blue), and KC2 (green) models (ρ=0.4 and N=1000)

The $\rho_C^2$, $\rho_{u\to v}^{(2)}$ and $\rho_{v\to u}^{(2)}$ values for the asymmetric KC1 and KC2 models of the data pairs produced in different correlations were calculated using equations (9), (10), and (11) and are presented in Table 4. In this table, $\rho_S$ and $\rho_S^2$ show the Spearman correlation values and coefficients of determination of the produced data, respectively. Here, it can be seen that the directional dependence values from $V$ to $U$ and from $U$ to $V$ are different from the values of $\rho_C^2$. This indicates that $\rho_{u\to v}^{(2)}$ and $\rho_{v\to u}^{(2)}$ values should be used instead of $\rho_C^2$ because the dependency structures of the KC models are asymmetrical. Moreover, Jung et al. [44] showed that the directional dependence coefficients are theoretically different from $\rho_C^2$ asymmetric models for some parameters of the generalized FGM copula family.

**Table 4.** Directional dependence coefficients of the asymmetric KC1 and KC2 models

| | | | **KC1-Clayton** | | | |
|---|---|---|---|---|---|---|
| $\rho$ | $\rho_S$ | $\rho_S^2$ | $\rho_C$ | $\rho_C^2$ | $\rho_{u\to v}^{(2)}$ | $\rho_{v\to u}^{(2)}$ |
| 0.1 | 0.0907 | 0.0082 | 0.1066 | 0.0114 | 0.0357 | 0.0161 |
| 0.2 | 0.2007 | 0.0400 | 0.1951 | 0.0380 | 0.0584 | 0.0241 |
| 0.3 | 0.2993 | 0.0896 | 0.3343 | 0.1117 | 0.1742 | 0.1449 |
| 0.4 | 0.4052 | 0.1642 | 0.3954 | 0.1563 | 0.1430 | 0.1244 |
| | | | **KC1-Frank** | | | |
| $\rho$ | $\rho_S$ | $\rho_S^2$ | $\rho_C$ | $\rho_C^2$ | $\rho_{u\to v}^{(2)}$ | $\rho_{v\to u}^{(2)}$ |
| 0.1 | 0.0913 | 0.0083 | 0.1082 | 0.1170 | 0.0225 | 0.0376 |
| 0.2 | 0.2001 | 0.0400 | 0.2111 | 0.0445 | 0.0634 | 0.0100 |
| 0.3 | 0.2970 | 0.0882 | 0.2981 | 0.0889 | 0.0775 | 0.0867 |
| 0.4 | 0.3935 | 0.1548 | 0.4018 | 0.1614 | 0.0614 | 0.1058 |
| | | | **KC1-Gumbel** | | | |
| $\rho$ | $\rho_S$ | $\rho_S^2$ | $\rho_C$ | $\rho_C^2$ | $\rho_{u\to v}^{(2)}$ | $\rho_{v\to u}^{(2)}$ |
| 0.1 | 0.0935 | 0.0087 | 0.0813 | 0.0066 | 0.0240 | 0.0259 |
| 0.2 | 0.2033 | 0.0413 | 0.2135 | 0.0456 | 0.0211 | 0.1393 |
| 0.3 | 0.3024 | 0.0914 | 0.2909 | 0.0846 | 0.0614 | 0.0895 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.4 | 0.4033 | 0.1626 | 0.3984 | 0.1587 | 0.1845 | 0.1938 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | **KC2-Clayton** | | | |
| $\rho$ | $\rho_S$ | $\rho_S^2$ | $\rho_C$ | $\rho_C^2$ | $\rho_{u\to v}^{(2)}$ | $\rho_{v\to u}^{(2)}$ |
| 0.2 | 0.2044 | 0.0418 | 0.2232 | 0.0498 | 0.0335 | 0.0155 |
| 0.4 | 0.3957 | 0.1566 | 0.3987 | 0.1589 | 0.1964 | 0.0837 |
| 0.6 | 0.5936 | 0.3524 | 0.5783 | 0.3344 | 0.4404 | 0.5771 |
| 0.8 | 0.7897 | 0.6236 | 0.7905 | 0.6249 | 0.7191 | 0.8515 |
| | | | **KC2-Frank** | | | |
| $\rho$ | $\rho_S$ | $\rho_S^2$ | $\rho_C$ | $\rho_C^2$ | $\rho_{u\to v}^{(2)}$ | $\rho_{v\to u}^{(2)}$ |
| 0.2 | 0.2050 | 0.0420 | 0.2120 | 0.0450 | 0.0357 | 0.0708 |
| 0.4 | 0.4183 | 0.1750 | 0.4044 | 0.1636 | 0.1365 | 0.0957 |
| 0.6 | 0.6039 | 0.3647 | 0.6097 | 0.3718 | 0.3584 | 0.4006 |
| 0.8 | 0.7998 | 0.6397 | 0.7965 | 0.6344 | 0.5570 | 0.6732 |
| | | | **KC2-Gumbel** | | | |
| $\rho$ | $\rho_S$ | $\rho_S^2$ | $\rho_C$ | $\rho_C^2$ | $\rho_{u\to v}^{(2)}$ | $\rho_{v\to u}^{(2)}$ |
| 0.2 | 0.2074 | 0.0430 | 0.2250 | 0.0506 | 0.0538 | 0.0874 |
| 0.4 | 0.3970 | 0.1576 | 0.4053 | 0.1643 | 0.1897 | 0.0816 |
| 0.6 | 0.6050 | 0.3660 | 0.6126 | 0.3753 | 0.5773 | 0.4822 |
| 0.8 | 0.7934 | 0.6295 | 0.8007 | 0.6411 | 0.5764 | 0.6408 |

## DATA ANALYSIS

The data contained 174 automobile bodily injury claims collected between 1993 and 1998 in Massachusetts and studied by Frees and Wang [43]. The data can be found in the CASdatasets package [42] of the R program under the name 'usmassBI2'. Descriptive statistics of the data are given in Table 5, and a scatterplot of the data is presented in Figure 2. Here, AC represents the average claims per unit of exposure ($), and PPSM represents the population per square mile of the town.

**Table 5.** Descriptive statistics of automobile bodily injury claim data

| | Mean | Median | Min | Max | St. D | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| AC | 137.32 | 136.49 | 42.74 | 248.75 | 35.18 | 0.1708 | 0.2445 |
| PPSM | 801.74 | 593.67 | 119.56 | 4636.74 | 815.41 | 3.3309 | 12.4009 |

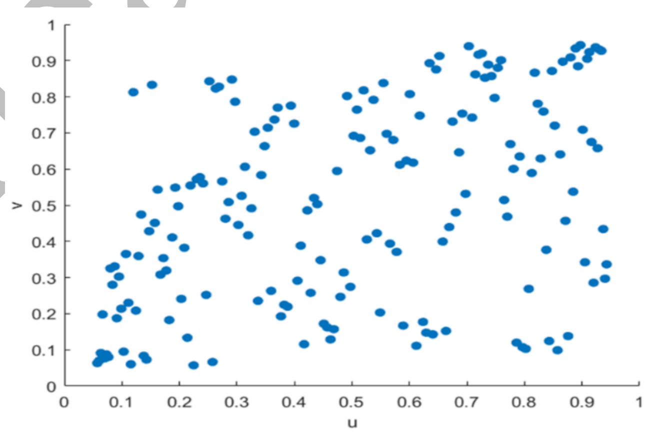


**Figure 2**. Scatterplot of automobile bodily injury claim data

To choose the most appropriate candidate models that can fit the data, an exchangeability test was performed on the data. The results given in Table 6 show that the data did not fit the copula families. The dependency structure of the data is asymmetric because the p-value (=0.01499) is smaller than 0.05. The test of goodness of fit for symmetrical Archimedean copula families also supported this result.

**Table 6.** Parameter estimation summary of symmetric copula families for automobile bodily injury claim data ($S_n$: Cramer-von Mises test statistics)

| Copula Family | Parameter | $LL$ | $S_n$ | $p$-value |
|---|---|---|---|---|
| Clayton | 0.74011 | 20.53453 | 0.080170 | 0.00150 |
| Frank | 2.98366 | 18.34867 | 0.051319 | 0.00649 |
| Gumbel | 1.42846 | 20.28379 | 0.045947 | 0.02747 |

Thus, we can choose candidate copula models from among asymmetric copulas. For this purpose, we use the asymmetric KC and LCC models defined by equations (3-4) and (7-8), respectively. We determine the best-fitting model with the AIC value given by

$$AIC = -2LL + 2p$$

where $p$ is the number of parameters and $LL$ is the maximized log-likelihood. The model selection process according to the directional dependency is as follows: First, the model with the smallest AIC value is considered. The model selection is then continued with the directional dependency measures with the smallest and largest values of $\rho^{(2)}_{u\to v}$ and $\rho^{(2)}_{v\to u}$. Finally, the two selected asymmetric models were tested with GOF (Sn-Cramer-von Mises) and the one with the lowest AIC value was selected as the best-fit model.

The parameter estimations, log-likelihood, and the values of Sn, p, and AIC for the asymmetric KC and LCC models were calculated and are shown in Tables 7 and 8. Here, the maximum pseudo-likelihood (MPL) method is used to obtain parameter estimates.

**Table7.** Estimated parameters, LL, Sn, p, and AIC values of KC models for automobile bodily injury claim data.

| Model | $\hat{\theta}$ | $\hat{\alpha}$ | $\hat{\beta}$ | $LL$ | $S_n$ | p value | AIC |
|---|---|---|---|---|---|---|---|
| $KC_1^C$ | 25.9738 | 0.2853 | 0.7147 | 32.5540 | 0.02829 | 0.2486 | -61.1081 |
| $KC_1^F$ | 33.9303 | 0.2828 | 0.7173 | 32.3130 | 0.03527 | 0.1286 | -60.6260 |
| $KC_1^G$ | 2.98295 | 0.3188 | 0.6812 | 26.1288 | 0.02479 | 0.3114 | -48.2575 |
| $KC_2^C$ | 19.1733 | 0.3341 | 0.8289 | 33.5462 | 0.02829 | 0.2543 | -61.0924 |
| $KC_2^F$ | 30.4350 | 0.3164 | 0.8013 | 32.7218 | 0.03190 | 0.1400 | -59.4436 |
| $KC_2^G$ | 2.2574 | 0.4511 | 0.9987 | 29.3557 | 0.02479 | 0.3171 | -52.7114 |

**Table 8**. Estimated parameters, LL, and AIC values of LCC models for automobile bodily injury claim data

| Model | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{p}_1$ | $\hat{p}_2$ | $\hat{p}_3$ | $LL$ | AIC |
|---|---|---|---|---|---|---|---|---|
| $LCC_1^C$ | 2.9103 | 3.2858 | - | 0.2800 | 0.7200 | - | -165.8824 | 377.6481 |
| $LCC_1^F$ | 4.1575 | 5.1835 | - | 0.1476 | 0.8524 | - | -178.8822 | 363.7643 |
| $LCC_1^G$ | 2.0266 | 2.9032 | - | 0.3216 | 0.6784 | - | -160.7671 | 327.5343 |
| $LCC_2^C$ | 2.3716 | 2.9671 | 3.2964 | 0.2889 | 0.2628 | 0.4483 | -167.8495 | 345.6991 |
| $LCC_2^F$ | 4.6796 | 5.3969 | 5.3969 | 0.1571 | 0.0291 | 0.8138 | -170.3059 | 350.6118 |
| $LCC_2^G$ | 1.7356 | 2.6574 | 3.0947 | 0.2919 | 0.4277 | 0.2804 | -171.6882 | 353.3764 |

KC models are considered first because the AIC values of the KC models are approximately the same among themselves, and they are insignificant compared to the LCC models. The GOF test was performed on selected KC models, and it is seen in Table 7 that all models fit the data (p >0.05).

The model selection process is then continued according to the directional dependence coefficients. Models with the highest and lowest $\rho^{(2)}_{u\to v}$ and $\rho^{(2)}_{v\to u}$ values are determined, and it is seen in Table 9 that they correspond to the

$KC_1^G$ and $KC_2^G$ models, respectively. Since more than one copula fits the data, $KC_2^G$ the model with the smallest AIC value ($= -52.7114$ from Table 7) is selected as the best-fit model.

**Table 9.** Directional dependence coefficients of asymmetric copula models for automobile bodily injury claim data $(\rho_s = 0.4293, \rho_s^2 = 0.1843)$

| Models | $\rho_C$ | $\rho_C^2$ | $\rho_{u\to v}^{(2)}$ | $\rho_{v\to u}^{(2)}$ |
|---|---|---|---|---|
| $KC_1^C$ | 0.33027 | 0.10908 | 0.11718 | 0.13188 |
| $KC_1^F$ | 0.32924 | 0.10840 | 0.11550 | 0.13131 |
| $KC_1^G$ | 0.32404 | 0.10500 | **0.10619** | **0.12085** |
| $KC_2^C$ | 0.38979 | 0.15194 | 0.16472 | 0.18070 |
| $KC_2^F$ | 0.37266 | 0.13888 | 0.14915 | 0.16970 |
| $KC_2^G$ | 0.43560 | 0.18975 | **0.19208** | **0.21725** |

A comparison scatterplot between the original data and the simulated data from $KC_2^G$ the model is made to further check the suitability. For this purpose, their scatterplots are shown in Figure 3. It can be seen from the scatterplots that the simulated data and the original data fit each other very well.

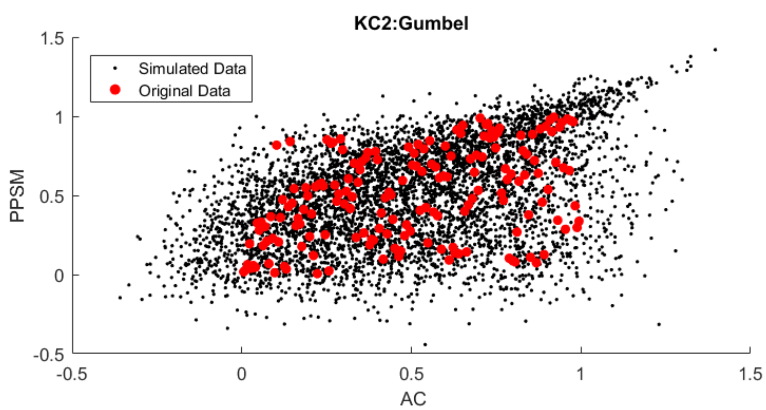


**Figure 3.** Comparison of the scatterplot between the original data and simulated data

Additionally, the values of $\rho_S^2$, $\rho_C^2$, and $\rho_{u\to v}^{(2)}$ the automobile bodily injury claim data in Tables 7 and 9 can be interpreted as follows for the best model chosen ($KC_2^G$). AC was considered the dependent variable in the data, and PPSM was the independent variable. When the relationship between them was accepted as linear, the coefficient of determination was found to be 0.1843 ($\rho_S^2$) based on the known Spearman correlation. In other words, the explanation rate of AC with PPSM was 18.43%. When the dependency structure of these variables is modeled with copulas, this explanation rate is 18.975% ($\rho_C^2$). When modeled with asymmetric copulas, the disclosure rate according to the directional dependency measures was found to be 21.725% ($\rho_{v\to u}^{(2)}$). These results show that the explanation ratio found according to the standard Spearman coefficient without considering dependence and asymmetricity is lower. Thus, the decision-maker will calculate a lower premium than it should be in the case of premium pricing, ignoring the asymmetric model for automobile claim data with an asymmetric dependency.

As a result, the advantage of the model selection method based on directional dependence is that it gives a decision-maker who wants to make actuarial calculations, such as premium pricing in the insurance field, the opportunity to make a more accurate calculation by considering the directional dependence coefficient.

## CONCLUSION

Real data are not always symmetrical. In such cases, it is necessary to use asymmetric models for modeling. Using these asymmetric models, obtaining the desired probabilities and statistical inferences will provide more accurate results. For our data, the first asymmetric tests were performed, and it was found that the dataset was asymmetric. Then, using Clayton Frank and Gumbel copulas, asymmetric models KC1, KC2, LCC1, and LCC2 were created. We used directional dependency coefficients to determine the model that best fits the dataset from among asymmetric models. According to the directional dependency coefficients and AIC, we concluded that the best model that fits our dataset is KC2-Gumbel. This modeling can have important practical implications for the insurance industry, because it can provide more accurate estimates of risk factors and inform risk management and policy decisions. This is particularly important given the increasing importance of data-driven decision-making in the insurance industry and the need to accurately model risk factors to effectively manage risk. In addition, the use of Khoudraji copulas can be applied to other areas of insurance and risk management, where asymmetrically dependent data are common, such as modeling the joint distribution of insurance claims across different lines of business.

## CONFLICT OF INTEREST

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## REFERENCES


[1] Sklar A (1959) Functions de repartition an dimensions at leurs marges. Publ Inst Statist Univ Paris. 8: 229-231

[2] Nelsen RB (2006) An Introduction to Copulas. 2nd edition, Springer, New York

[3] Salvadori G, De Michele C (2007) On the use of copulas in hydrology: theory and practice. J. Hydrol. Eng. 12 (4), 369–380.

[4] Genest C, Favre A-C (2007) Everything you always wanted to know about copula modeling but were afraid to ask. Journal of hydrologic engineering 12(4): 347–368

[5] Joe H (2014) Dependence modeling with copulas. CRC Press.

[6] Durante F, Sempi C (2015) Principles of copula theory. CRC press.

[7] Hong Y, Wang JP, Li DQ, Cao ZJ, Ng CWW, Cui P (2015) Statistical and probabilistic analyses of impact pressure and discharge of debris flow from 139 events during 1961 and 2000 at Jiangjia Ravine, China. Eng. Geol. 187, 122–134.

[8] Salvadori G, De Michele C, Kottegoda NT, Rosso R (2007) Extremes in nature: an approach using copulas. Water Science and Technology Library, Springer Science &Business Media.

[9] Kara EK, Yıldız O (2014). Bivariate analysis of precipitation and using copulas. Istatistik Journal of The Turkish Statistical Association, 7(3), 63-70.

[10] Callau Poduje AC, Belli A, Haberlandt UA (2014) Dam risk assessment based on univariate versus bivariate statistical approaches: A case study for Argentina. Hydrol. Sci. J., 59, 2216–2232.

[11] Kara EK (2017). The Earthquake Risk Analysis Based on Copula Models for Turkey. Sigma: Journal of Engineering & Natural Sciences, 35(2), 187-200.

[12] Kim JM, Jung YS, Sungur EA (2014) Truncation invariant copulas for modeling directional dependence: Application to foreign currency exchange data. Model Assisted Statistics and Applications. 9(4): 309-324

[13] Kemaloglu SA, Kara EK (2015). Modeling dependent financial assets by dynamic copula and portfolio optimization based on CVaR. Communications Faculty of Science University of Ankara Series A1 Mathematics and Statistics, 64(1), 1-13.

[14] Kara EK, Kemaloglu SA(2016). Portfolio optimization of dynamic copula models for dependent financial data using change point approach. Communications Faculty of Science University of Ankara Series A1 Mathematics and Statistics, 65(2), 175-188.

[15] Kara EK, Kemaloglu SA, Evkaya ÖO (2022). Modeling Currency Exchange Data with Asymmetric Copula Functions. In: Terzioğlu, M.K. (eds) Advances in Econometrics, Operational Research, Data Science and Actuarial Studies. Contributions to Economics. Springer, Cham., 49-62, https://doi.org/10.1007/978-3-030-85254-2_4

[16] Moradian S, Olbert AI, Gharbia S, Iglesias G (2023) Copula-based projections of wind power: Ireland as a case study. Renewable and Sustainable Energy Reviews, 175, 113147

[17] Corbella S, Stretch DD (2013) Simulating a multivariate sea storm using Archimedean copulas. Coastal Engineering 76, 68–78. 10.1016/j.coastaleng.2013.01.011

[18] Ma P, Zhang Y (2022). Modeling asymmetrically dependent multivariate ocean data using truncated copulas. Ocean Engineering, 244, 110226

[19] Li X, Babovic V (2019) Multi-site multivariate downscaling of global climate model outputs: an integrated framework combining quantile mapping, stochastic weather generator and Empirical Copula approaches. Climate Dynamics, 52: 5775-5799, https://doi.org/10.1007/s00382-018-4480-0

[20] Kim JM, Jung YS, Sungur EA, Han KH, Park C, Sohn IA (2008) Copula method for modeling directional dependence of genes. BMC bioinformatics. 9(1): 225

[21] Kim JM, Jung YS, Soderberg T (2009) Directional dependence of genes using survival truncated FGM type modification copulas. Communications in Statistics-Simulation and Computation. 38(7): 1470-1484

[22] Sungur EA (2005a) A note on directional dependence in regression setting. Communications in Statistics-Theory and Methods. 34: 1957–1965

[23] Sungur EA (2005b) Some observations on copula regression functions, Communications in Statistics-Theory and Methods. 34: 1967–1978

[24] Rodríguez-Lallena JA, Úbeda-Flores MA (2004) A new class of bivariate copulas. Statistics & probability letters. 66(3): 315–325

[25] Klement EP, Mesiar R (2006) How non-symmetric can a copula be? Commentationes Mathematicae Universitatis Carolinae. 47(1): 141-148

[26] Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. Adv. Water Resour. 29 (8), 1155–1167.

[27] Mesiar R, Najjari V (2014). New families of symmetric/asymmetric copulas. Fuzzy Sets Syst. 252, 99–110.

[28] Mazo G, Girard S, Forbes F (2015) A class of multivariate copulas based on products of bivariate copulas. Journal of Multivariate Analysis. 140, 363-376.

[29] Khoudraji A (1995) Contributions`al'´etude des copules et`alamod´elisation des valeurs extrˆemes bivari´ees. PhD thesis, Universit´ede Laval, Qu´ebec (Canada)

[30] Liebscher E (2008) Construction of asymmetric multivariate copulas. Journal of Multivariate Analysis 99: 2234–2250

[31] Durante F (2009) Construction of non-exchangeable bivariate distribution functions. Statistical Papers 50(2): 383-391

[32] Quessy JF and Kortbi O (2016) Minimum-distance statistics for the selection of an asymmetric copula in Khoudraji's class of models. In: Statistica Sinica. doi: 10.5705/ss.202014.0082

[33] Siburg KF, Stehling K, Stoimenov PA, Weiß GN (2016) An order of asymmetry in copulas, and implications for risk management. Insurance: Mathematics and Economics. 68: 241-247

[34] Bezak N, Rusjan S, Kramar Fijavž M, Mikoš M, Šraj M (2017) Estimation of suspended sediment loads using copula functions. Water, 9(8), 628.

[35] Zhang Y, Kim CW, Beer M, Dai H, & Soares CG (2018) Modeling multivariate ocean data using asymmetric copulas. Coastal Engineering, 135, 91-111.

[36] Zhang Y, Gomes, AT, Beer M, Neumann I, Nackenhorst U, Kim CW (2019) Modeling asymmetric dependences among multivariate soil data for the geotechnical analysis–The asymmetric copula approach. Soils and Foundations, 59(6), 1960-1979.

[37] Lin Y, Dong S, Tao S (2020). Modelling long-term joint distribution of significant wave height and mean zero-crossing wave period using a copula mixture. Ocean Engineering, 197, 106856.

[38] Bai X, Jiang H, Li C, Huang L (2020) Joint probability distribution of coastal winds and waves using a log-transformed kernel density estimation and mixed copula approach. Ocean Engineering 216, 107937. 10.1016/j.oceaneng.2020.107937.

[39] Huang W, Dong S (2021) Joint distribution of significant wave height and zero-up-crossing wave period using mixture copula method. Ocean Engineering, 219, 108305.

[40] Wu S (2014) Construction of asymmetric copulas and its application in two-dimensional reliability modeling. European Journal of Operational Research. 238: 476–485

[41] Shih JH, Emura T (2021) On the copula correlation ratio and its generalization, Journal of Multivariate Analysis 182, 104708

[42] Dutang C, Charpentier A (2020) "Package 'CASdatasets' ", http://cas.uqam.ca/pub/web/CASdatasets-manual.pdf

[43] Frees EW and Wang P (2005) Credibility using copulas, North American Actuarial Journal, 9(2), 31-48.

[44] Jung YS, Kim JM, Kim J (2008) New approach of directional dependence in exchange markets using generalized FGM copula function. Communications in Statistics—Simulation and Computation 37(4): 772-788.

[45] Kim D, Kim JM (2014) Analysis of directional dependence using asymmetric copula-based regression models. Journal of Statistical Computation and Simulation. 84(9): 1990–2010.