REVSTAT – Statistical Journal
Volume 20, Number 5, October 2022, 547–562
https://doi.org/10.57805/revstat.v20i5.390

Plug-in Estimation of Dependence Characteristics of Archimedean Copula via Bézier Curve

Authors: Selim Orhun Susam ✉
– Department of Econometrics, Munzur University, Tunceli, Turkey
orhunsusam@munzur.edu.tr

Mahmut Sami Erdoğan
– Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey
sami.erdogan@medeniyet.edu.tr

Received: June 2019 Revised: January 2021 Accepted: January 2021

Abstract:

- This article introduces measurement of the dependence between variables with a dependence structure defined by Archimedean copulas. The estimation of the dependence measure, such as Kendall's tau as well as the lower and upper tail dependence, is investigated by using estimation of the Kendall distribution function based on the Bézier curve. A Monte Carlo study is performed to measure the performance of the new estimation method. The simulation results showed that the proposed methods has good results in terms of estimation performance. The new estimators are also used to estimate the dependence coefficients for three sets of real data.

Keywords:

- *Archimedean copula; Bernstein polynomials; tail dependence; Kendall's tau.*

AMS Subject Classification:

- 62G32, 62G05.

✉ Corresponding author.

1. INTRODUCTION

In statistical theory and applications, copula models are useful tools for determining the dependence structure between the random variables. For instance, when two random variables of X and Y with joint cumulative distribution function H and marginals of F and G are considered respectively, there exists a copula C such that $H(x,y) = C\big(F(x), G(y)\big)$, for all x, y in $\mathbb{R}$. In the literature, there are many parametric copula families which have different dependence structure. The main focus of this paper was on the Archimedean copula class, which is characterized by generator function φ. Archimedean copula with generator function φ is defined by

$$C(u,v) = \varphi^{[-1]}\left\{\varphi(u) + \varphi(v)\right\}, \qquad u, v \in [0,1], \tag{1.1}$$

where φ is a generator function which is continuous and strictly decreasing convex function defined from $\mathbf{I}$ to $[0,\infty)$ such that $\varphi(1) = 0$.

Genest *et al.* [15] showed that the function φ can be obtained by the univariate distribution function of $K(t) = P\big(C(u,v) \le t\big)$. Remarkably, there is a relationship between the function $\varphi(t)$ and $K(t)$ as

$$K(t) = t - \frac{\varphi(t)}{\varphi'(t)}. \tag{1.2}$$

The Kendall distribution function $K(t)$ has some important properties. These properties are summarized by Nelsen [20] as follows:

1. $K(0) = 0$;
2. $K(1) = 1$;
3. $K(t) > t,\ t \in (0,1)$;
4. $K'(t) > 0,\ t \in (0,1)$.

The dependence structure of the Archimedean copula family is characterized by $K(t)$. Kendall's tau (τ) is designed to describe how large (or small) values of one random variable appear with large (or small) values of the other as defined by Genest *et al.* [13] by

$$\tau = 3 - 4\int_0^1 K(t)\,dt. \tag{1.3}$$

Also, the tail dependence is related to the level of dependence in the upper-right (λ_{U}) or lower-left (λ_{L}) quadrant tail of a bivariate distribution. Michiels *et al.* [19] defined λ_{L} and upper λ_{U} dependence as

$$\lambda_{\mathrm{L}} = 2^{\lim_{t\to 0^+}\left(t - K(t)\right)'}, \tag{1.4}$$

$$\lambda_{\mathrm{U}} = 2 - 2^{\lim_{t\to 1^-}\left(t - K(t)\right)'}. \tag{1.5}$$

Some well-known Archimedean copula functions were proposed by Clayton [4], Frank [11], and Gumbel [17]. The generator functions of $\varphi(t)$ and Kendall distribution functions $K(t)$ of these copulas are summarized in Table 1. And also, Kendall's Tau (τ), Lower λ_{L} and Upper λ_{U} tail dependence coefficients for Gumbel, Clayton, and Frank copula are listed in Table 2.

Table 1: Archimedean Copulas with Generator functions $\varphi(t)$.

Copula	$\varphi(t)$	$K(t)$	Range of θ
Clayton	$\dfrac{t^{-\theta}-1}{\theta}$	$t+\dfrac{t\,(1-t^{\theta})}{\theta}$	$(-1,\infty)-\{0\}$
Frank	$-\log\left(\dfrac{\exp(t\theta)-1}{\exp(\theta)-1}\right)$	$t-\dfrac{(\exp(t\theta)-1)\,\log\left(\frac{\exp(-t\theta)-1}{\exp(-\theta)-1}\right)}{\theta}$	$(-\infty,\infty)-\{0\}$
Gumbel	$(-\log(t))^{\theta}$	$t-\dfrac{t\,\log(t)}{\theta}$	$[1,\infty)$
Independence	$-\log(t)$	$t-t\,\log(t)$	—

Table 2: Kendall's Tau (τ), Lower λ_{L} and Upper λ_{U} tail dependence for some Archimedean copulas.

Copula	$\tau(\theta)$	λ_{L}	λ_{U}
Clayton	$\dfrac{\theta}{\theta+2}$	$2^{-\frac{1}{\theta}}$	0
Frank	$1+4\theta^{-1}\left(D_1^{*}(\theta)-1\right)$	0	0
Gumbel	$\dfrac{\theta-1}{\theta}$	0	$2-2^{\frac{1}{\theta}}$

$^{*}\,D_1(x) = x^{-1}\displaystyle\int_0^x t\,(\exp(t)-1)^{-1}\,dt$

Modern risk management is mainly interested in assessing the amount of Kendall's tau and tail dependence. For this reason, many minimum-variance portfolio models are based on correlation. However, correlation itself is not enough to describe a tail dependence structure and often results in misleading interpretations (Embrechts *et al.* [7]). The importance of this issue has led to some improvements in the estimation of the dependence coefficients. Kollo *et al.* [18] examined tail behavior of skew t-copula considering the bivariate case. They used the method of moments and the maximum likelihood for the estimation of the tail dependence coefficients. Ferreira [10] proposed a nonparametric estimator of the tail dependence coefficient and proved its strong consistency and asymptotic normality in the case of known marginal distribution functions. Schmidt *et al.* [21] proposed a set of nonparametric estimators for the upper and lower tail copula and established results of weak convergence and strong consistency for the tail-copula estimators. Ferreira *et al.* [9] introduced the s,k-extremal coefficients for studying the tail dependence between the s-th lower and k-th upper order statistics of a normalized random vector. Caillault *et al.* [3] introduced nonparametric estimators for upper and lower tail dependence whose confidence intervals are obtained with the bootstrap method as they called these estimators "Naive estimators".

Goegebeur *et al.* [16] introduced a class of weighted functional estimators for the coefficient of the tail dependence in bivariate extreme value statistics while they also derived the minimum variance asymptotically unbiased estimator.

In this paper, plug-in estimations of Kendall's tau, upper tail dependence and lower tail dependence are introduced. To the author's best knowledge, this is the first study examining the estimation of the dependence coefficients using the plug-in method. The use of Bernstein–Bézier polynomials reduced the complexity of the non-parametric estimation of the tail dependence coefficients. Besides, the proposed estimation method of the dependence coefficient is flexible depending on its polynomial degree while the error of the estimation can be reduced by increasing or decreasing the degree of the polynomial.

The remainder of the study is organized as follows. In Section 2, the estimation of Kendall distribution function based on Bernstein polynomials is discussed. In Section 3, Kendall's tau and tail dependence coefficients are estimated by the plug-in principle. The performance of the new estimation methods for the dependence coefficients is investigated in Section 4. In Section 5, the new estimator of Kendall's tau and tail dependence coefficients are applied to three real data sets. Finally, the conclusion is presented in Section 6.

2. ESTIMATION OF THE KENDALL DISTRIBUTION FUNCTION

Before introducing the estimation of the dependence coefficients for Archimedean copulas, it is important to investigate the estimation of Kendall distribution function since the dependence coefficients of Archimedean copula are closely related to the Kendall distribution function as stated in the last section. First time in the literature, Genest *et al.* [15] investigated the empirical estimate of Kendall distribution function. For the estimation of the random variable of $T = H(x, y)$, univariate distribution function of $K(t) = P\big(H(x, y) \le t\big) = P\big(C(u, v) \le t\big)$ should be estimated within the interval of $[0, 1]$. This estimation process can be accomplished by two steps:

1. Constructing the empirical bivariate distribution function of $H_n(X, Y)$;

2. Obtaining the pseudo observations of $\widehat{T}_i$ by

$$\widehat{T}_i = \sum_{j=1}^{n} \mathrm{I}\big(X_i < X_j\,,\, Y_i < Y_j\big)\big/(n-1)\,, \qquad i = 1, ..., n\,.$$

By using these pseudo observations, $K(t)$ is estimated by the empirical distribution function as

$$K_n(t) = \sum_{i=1}^{n} \mathrm{I}\big(\widehat{T}_i \le t\big)/n\,.$$

Genest *et al.* [15] stated that the empirical estimation of Kendall distribution function is $\sqrt{n}$-consistent estimator while Barbe *et al.* [1] proved consistency of this estimator.

Generally, the classical empirical distribution function has a good performance as an estimator of the distribution function. However, estimating continuous distribution function may not be appropriate (Susam *et al.* [22, 23], Erdogan *et al.* [8]) since it has discontinuities.

Because of this, Susam *et al.* [22] proposed a smooth estimate of Kendall distribution function $K_{n,m}$ given by the following equation:

$$K_{n,m}(t) = \sum_{k=0}^{m} K_n\left(\frac{k}{m}\right) P_{k,m}(t)\,, \qquad t \in [0,1]\,,$$

where $P_{k,m}(t) = \binom{m}{k} t^k (1-t)^{m-k}$ is the Binomial probability. Susam *et al.* [23] proposed the Bézier curve based estimation of Kendall distribution function of $K_{\alpha,m}$ which has lower mean integrated squared error (MISE) scores than $K_{n,m}(t)$. They defined $K_{\alpha,m}$ as it is based on a set of the control points of α_i, $i = 0, ..., m$, as given by the following equation:

$$K_{\alpha,m}(t) = \sum_{k=0}^{m} \alpha_k \, P_{k,m}(t)\,, \qquad t \in [0,1]\,.$$

Also, they state that if the following constraints defined on the control points of α_i ($i = 1, ..., m$) hold, then the Bézier curve based on the estimation of Kendall distribution function of $K_{\alpha,m}$ satisfies all the properties of the Kendall distribution function.

Theorem 2.1 (Susam *et al.* [23])**.** *The estimator $K_{\alpha,m}(t)$ satisfies properties of Kendall distribution function under the following constraints hold:*

1. $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m = 1$;
2. $\alpha_i > \frac{i}{m}$, $i = 1, ..., m-1$.

They used minimum quadratic distance estimator which is based on the empirical Kendall distribution for estimating the control points of α_i ($i = 1, ..., m-1$). Also, Susam *et al.* [24] proposed minimum distance estimator for $K_{\alpha,m}(t)$ based on Bernstein estimate of Kendall distribution function $K_{n,m}(t)$. They stated that the minimum distance method based on Kendall distribution using Bernstein polynomials outperforms the method based on empirical Kendall distribution.

3. ESTIMATION OF DEPENDENCE COEFFICIENTS BASED ON BÉZIER CURVE ESTIMATION OF KENDALL DISTRIBUTION FUNCTION

It is possible to estimate Kendall's tau, lower and upper tail dependence by replacing $K(t)$ with its non-parametric estimation provided in Equations (1.3), (1.4) and (1.5). For a given bivariate random sample of size n, $(X_1, Y_1), ..., (X_n, Y_n)$ from X and Y, plug-in estimation of Kendall's tau, lower and upper tail dependence for Archimedean copula could be derived from the following equations:

$$\widehat{\tau} = 3 - 4\int_0^1 K_{\alpha,m}(t)\,dt\,, \tag{3.1}$$

$$\widehat{\lambda}_{\mathrm{L}} = 2^{\lim\limits_{t\to 0^+}\left(t - K_{\alpha,m}(t)\right)'}\,, \tag{3.2}$$

$$\widehat{\lambda}_{\mathrm{U}} = 2 - 2^{\lim\limits_{t\to 1^-}\left(t - K_{\alpha,m}(t)\right)'}\,, \tag{3.3}$$

where $K_{\alpha,m}(t)$ is the estimation of Kendall distribution function based on the Bézier curve introduced in Section 2. Then, the next lemmas are provided for the estimation of Kendall's tau, lower and upper tail dependence for Archimedean copulas.

Lemma 3.1. *Let $K_{\alpha,m}(\cdot)$ be the estimator of Kendall distribution function based on the Bézier curve while $\widehat{\alpha}_k$ $(k = 1, ..., m-1)$ estimates the control points. The estimator of Kendall's tau for Archimedean copula is obtained by*

$$\widehat{\tau} \;=\; 3 - 4\sum_{k=0}^{m} \widehat{\alpha}_k \binom{m}{k} \beta\big(k+1, m-k+1\big)\,,$$

where $\beta(\cdot,\cdot)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1-1}(1-t)^{v_2-1}\,dt$ for v_1 and v_2 positive integers.

Lemma 3.2. *Let $K_{\alpha,m}(\cdot)$ be the estimator of Kendall distribution function based on the Bézier curve while $\widehat{\alpha}_k$ $(k = 1, ..., m-1)$ estimates the control points. The estimation of the lower tail and the upper tail dependence for the Archimedean copula is obtained by*

$$\widehat{\lambda}_{\mathrm{L}} \;=\; 2^{1-m\widehat{\alpha}_1}\,,$$
$$\widehat{\lambda}_{\mathrm{U}} \;=\; 2 - 2^{1-m(1-\widehat{\alpha}_{m-1})}\,.$$

Proof: First order derivative of Bézier curve is derived by

$$K'_{\alpha,m}(t) \;=\; m\sum_{k=0}^{m-1}(\alpha_{k+1} - \alpha_k)\,P_{k,m-1}\,.$$

From the end-point rule of the Bézier curve, $\lim_{t\to 0^+} K'_{\alpha,m}(t)$ and $\lim_{t\to 1^-} K'_{\alpha,m}(t)$ are equal to $m(\alpha_1 - \alpha_0)$ and $m(\alpha_m - \alpha_{m-1})$ respectively (see Duncan [6]). Because of $\alpha_0 = 0$ and $\alpha_m = 1$, then the desired results are obtained. □

It is observed that $\widehat{\lambda}_{\mathrm{L}}$ and $\widehat{\lambda}_{\mathrm{U}}$ are affected by only the control points of α_1 and α_{m-1}, respectively. The range of the dependence coefficients depending on the polynomial degree m is summarized in Table 3. The results show that the range of dependence coefficients gets wider as the degree of the polynomial increases.

Table 3: Interval of Kendall's Tau (τ), Lower λ_{L} and Upper λ_{U} tail dependence for varying polynomial degrees of m.

Degree (m)	τ	λ_{U}	λ_{L}
5	$[-0.33, 1]$	$[0, 1]$	$[0.0625, 1]$
10	$[-0.64, 1]$	$[0, 1]$	$[0.0019, 1]$
15	$[-0.75, 1]$	$[0, 1]$	$[6.1 \times 10^{-5}, 1]$
20	$[-0.81, 1]$	$[0, 1]$	$[1.9 \times 10^{-6}, 1]$

For estimating the control points of α_i $(i = 0, ..., m-1)$, statistical programming language `R` is used. The package "`nloptr`" is quite handy for optimizing non-linear function. The Augmented Lagrangian algorithm (`auglag`) included in the package "`nloptr`" should be used. Since $K_{\alpha,m}(\cdot)$ has a complex function for higher polynomial degree so that may cause trouble in optimization. In order to overcome such a problem, the number of maximum evaluation number (maxeval) is recommended to be selected as at least 50.000 in the optimization.

4. MONTE CARLO SIMULATION

To determine the performance of the estimation of τ, λ_{U}, and λ_{L}, the Monte Carlo simulation is conducted. 1.000 Monte Carlo samples with $n = 150$ size are generated from each type of Archimedean copulas. For instance, parameters of $\theta = 1.11, 1.25, 1.44$ is used for Gumbel copula while parameters of $\theta = 0.22, 0.50, 0.85$ is used for Clayton copula and $\theta = 0.91, 1.86, 2.92$ is used for Frank copula. Each copula has different shapes and characteristics. Clayton copula exhibits strong left tail dependence. In contrast to Clayton, Gumbel has strong right tail dependence while Frank copula exhibits symmetric and weak tail dependence in both tails. Detailed information about these Archimedean copulas is provided in Nelsen [20]. In all estimation methods, the Bézier curve degrees are selected for $m = 1, ..., 20$. The mean of the estimation of the dependence coefficients for τ, λ_{U}, and λ_{L} Archimedean copulas for the varying degrees of $m = 5, 10, 15$ and 20 are summarized in Tables 4, 5, and 6.

The following results are obtained from Tables 4, 5, and 6:

- For the estimation of Kendall's tau, the mean of the τ estimates is closer to the true value for the polynomial degree of $m = 5$ when the true copula belongs to Gumbel, Clayton, or Frank.
- When the true copula is Gumbel with $\tau = 0.1, 0.2, 0.3$, mean of the λ_{U} estimates is closer to true value for the polynomial degree of $m = 10$ while the mean of the estimation of λ_{L} is closer to true value for the polynomial degree of $m = 20$.
- When the true copula is Clayton with $\tau = 0.1, 0.2, 0.3$, mean of the λ_{U} estimates is closer to true value for the polynomial degree of $m = 20$ while the mean of the estimation of λ_{L} is closer to true value for the polynomial degree of $m = 5$.
- When the true copula is Frank with $\tau = 0.1, 0.2, 0.3$, while the mean of the λ_{U} estimates are closer to true value for the polynomial degree of $m = 20$ while the mean of the λ_{L} estimates is closer to true value for the polynomial degree of $m = 20$.

The results obtained from Figures 1, 2, and 3 are:

- As the dependence level increases for Gumbel, Clayton, and Frank copula, the variance of the estimations of the τ, λ_{U}, and λ_{L} increases as well.
- When the true copula belongs to the Clayton family with $\theta = 0.22, 0.50$ and 0.85, the variance of λ_{U} estimation decreases as the degree of polynomial increases. On the contrary, the variance of λ_{L} increases as the degree of polynomial increases.

- When the true copula is Frank with $\theta = 0.91, 1.86, 2.92$, the variance of λ_U decreases as the degree of polynomial increases. On the other hand, the variance of λ_L does not change as the degree of polynomial increases.
- In all the estimations of dependence coefficients, the estimation of τ, λ_L and λ_U get closure to the real values as the polynomial degree increases.

Table 4: Mean of the estimation of τ of Archimedean copulas.

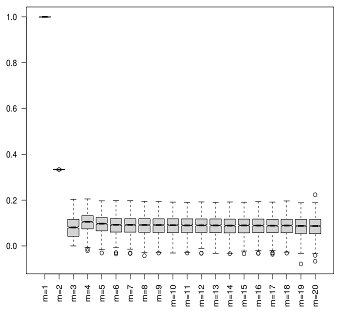
Copula	θ	τ	$\widehat{\tau}^5$	$\widehat{\tau}^{10}$	$\widehat{\tau}^{15}$	$\widehat{\tau}^{20}$
Gumbel	1.11	**0.099**	0.0956	0.0886	0.0876	0.0845
	1.25	**0.200**	0.1923	0.1883	0.1876	0.1858
	1.43	**0.300**	0.2909	0.2885	0.2880	0.2862
Clayton	0.22	**0.099**	0.0937	0.0896	0.0889	0.0871
	0.50	**0.200**	0.1922	0.1903	0.1898	0.1882
	0.85	**0.300**	0.2901	0.2884	0.2879	0.2869
Frank	0.91	**0.099**	0.0992	0.0897	0.0885	0.0861
	1.86	**0.200**	0.1972	0.1894	0.1879	0.1872
	2.92	**0.300**	0.2966	0.2897	0.2886	0.2875

Table 5: Mean of the estimation of λ_U of Archimedean copulas.

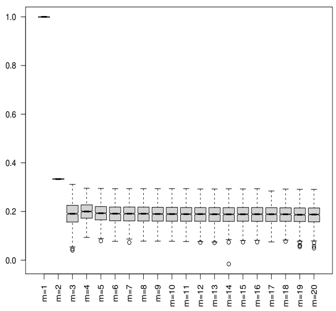
Copula	θ	λ_U	$\widehat{\lambda}_U^5$	$\widehat{\lambda}_U^{10}$	$\widehat{\lambda}_U^{15}$	$\widehat{\lambda}_U^{20}$
Gumbel	1.11	**0.132**	0.0963	0.1384	0.1265	0.1326
	1.25	**0.258**	0.1862	0.2299	0.2080	0.2311
	1.43	**0.376**	0.2912	0.3207	0.2947	0.3418
Clayton	0.22	**0.000**	0.0243	0.0450	0.0346	0.0198
	0.50	**0.000**	0.0448	0.0454	0.0401	0.0251
	0.85	**0.000**	0.0563	0.0518	0.0447	0.0296
Frank	0.91	**0.000**	0.0164	0.0516	0.0395	0.0213
	1.86	**0.000**	0.0338	0.0564	0.0481	0.0344
	2.92	**0.000**	0.0476	0.0721	0.0620	0.0444

Table 6: Mean of the estimation of λ_L of Archimedean copulas.

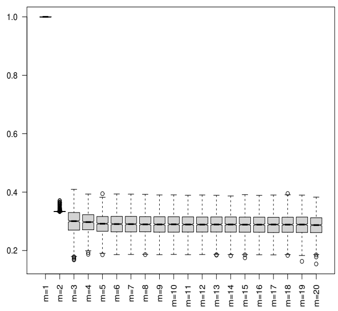
Copula	θ	λ_L	$\widehat{\lambda}_L^5$	$\widehat{\lambda}_L^{10}$	$\widehat{\lambda}_L^{15}$	$\widehat{\lambda}_L^{20}$
Gumbel	1.11	**0**	0.1605	0.0876	0.0637	0.0580
	1.25	**0**	0.1932	0.1132	0.0855	0.0790
	1.43	**0**	0.2330	0.1490	0.1185	0.1082
Clayton	0.22	**0.04**	0.1964	0.1380	0.1161	0.1140
	0.50	**0.25**	0.2985	0.2558	0.2302	0.2405
	0.85	**0.44**	0.4240	0.3943	0.3744	0.4010
Frank	0.91	**0**	0.1694	0.0901	0.0651	0.0578
	1.86	**0**	0.2107	0.1168	0.0852	0.0737
	2.92	**0**	0.2565	0.1484	0.1101	0.0949



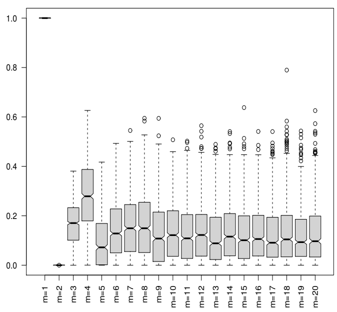
(a) τ estimation for $\theta = 1.11$ ($\tau = 0.1$).



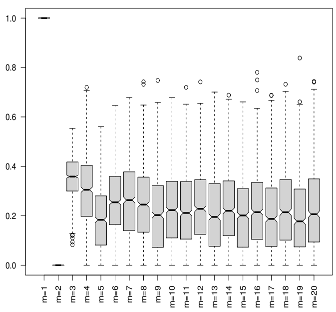
(b) τ estimation for $\theta = 1.25$ ($\tau = 0.2$).



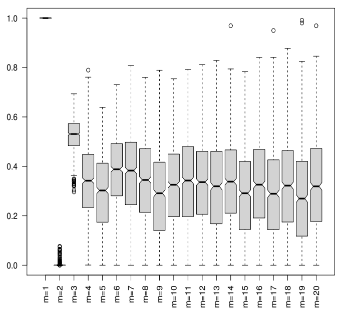
(c) τ estimation for $\theta = 1.43$ ($\tau = 0.3$).



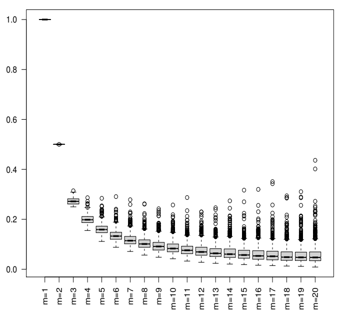
(d) λ_{U} estimation for $\theta = 1.11$ ($\lambda_{\mathrm{U}} = 0.13$).



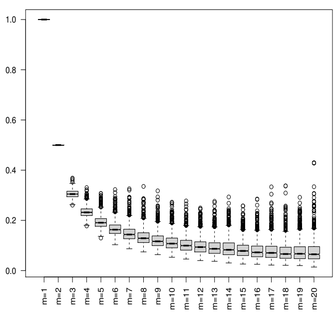
(e) λ_{U} estimation for $\theta = 1.42$ ($\lambda_{\mathrm{U}} = 0.25$).



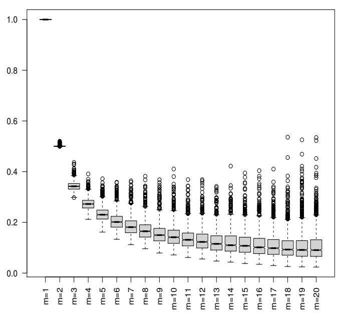
(f) λ_{U} estimation for $\theta = 2$ ($\lambda_{\mathrm{U}} = 0.37$).



(g) λ_{L} estimation for $\theta = 1.11$ ($\lambda_{\mathrm{L}} = 0$).

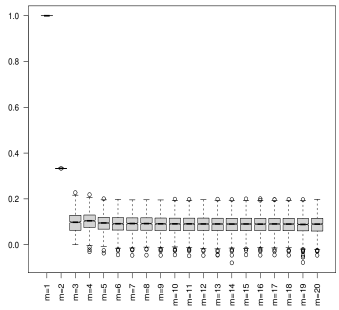


(h) λ_{L} estimation for $\theta = 1.42$ ($\lambda_{\mathrm{L}} = 0$).

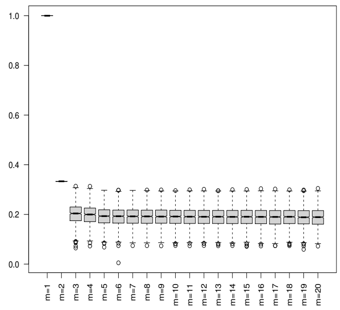


(i) λ_{L} estimation for $\theta = 2$ ($\lambda_{\mathrm{L}} = 0$).

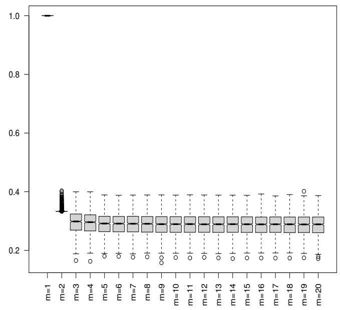
Figure 1: Box-plots of the estimation of the dependence coefficients of Gumbel copula with parameters of $\theta = 1.11, 1.25, 1.43$.



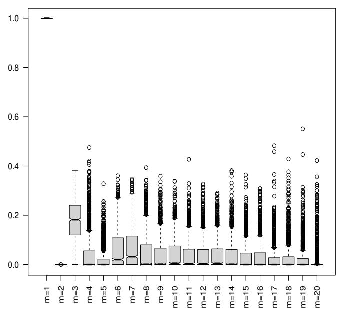
(a) τ estimation for $\theta = 0.22$ ($\tau = 0.1$).



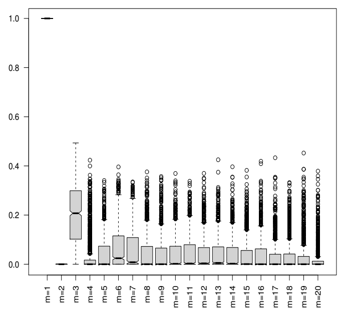
(b) τ estimation for $\theta = 0.50$ ($\tau = 0.2$).



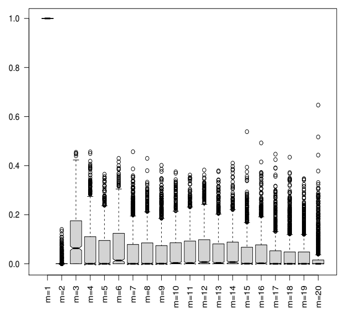
(c) τ estimation for $\theta = 0.85$ ($\tau = 0.3$).



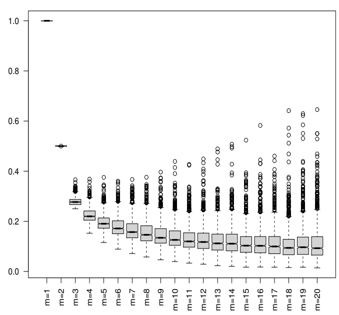
(d) λ_{U} estimation for $\theta = 0.22$ ($\lambda_{\mathrm{U}} = 0$).



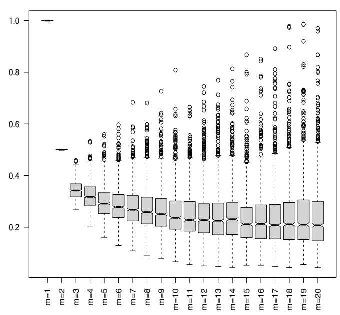
(e) λ_{U} estimation for $\theta = 0.50$ ($\lambda_{\mathrm{U}} = 0$).



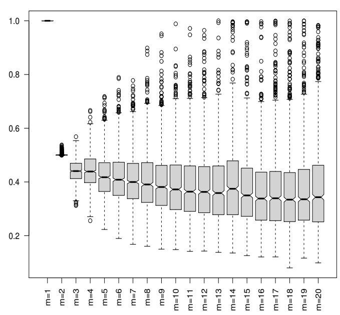
(f) λ_{U} estimation for $\theta = 0.85$ ($\lambda_{\mathrm{U}} = 0$).



(g) λ_{L} estimation for $\theta = 0.22$ ($\lambda_{\mathrm{L}} = 0.04$).

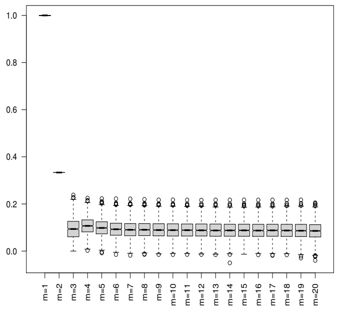


(h) λ_{L} estimation for $\theta = 0.50$ ($\lambda_{\mathrm{L}} = 0.25$).

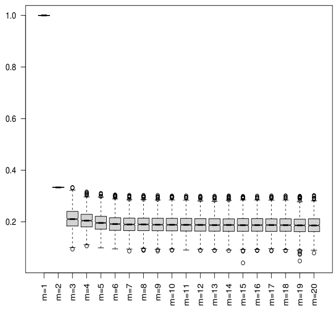


(i) λ_{L} estimation for $\theta = 0.85$ ($\lambda_{\mathrm{L}} = 0.44$).

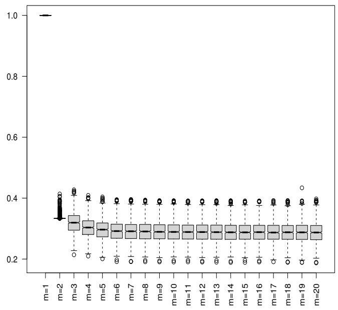
Figure 2: Box-plots of the estimation of the dependence coefficients Clayton copula with parameters of $\theta = 0.22, 0.50, 0.85$.



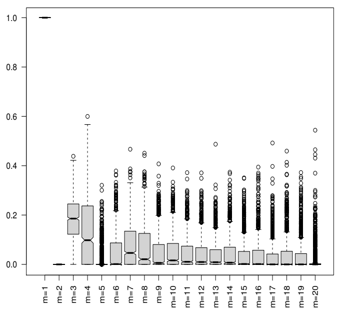
(a) τ estimation for $\theta = 0.91$ ($\tau = 0.1$).



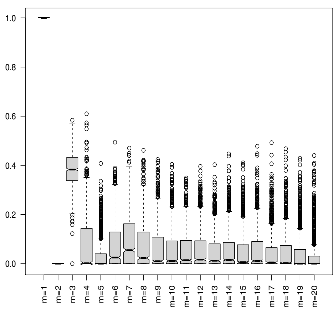
(b) τ estimation for $\theta = 1.86$ ($\tau = 0.2$).



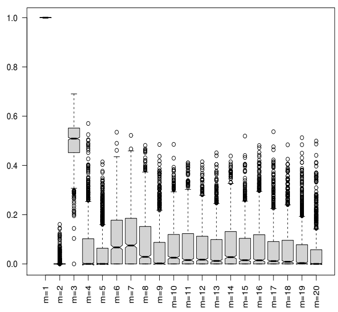
(c) τ estimation for $\theta = 2.92$ ($\tau = 0.3$).



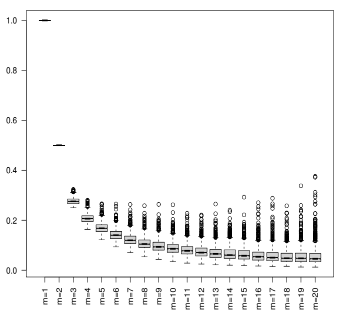
(d) λ_{U} estimation for $\theta = 0.91$ ($\lambda_{\mathrm{U}} = 0$).



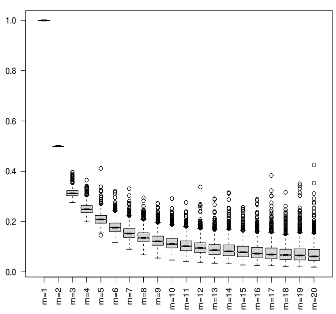
(e) λ_{U} estimation for $\theta = 1.86$ ($\lambda_{\mathrm{U}} = 0$).



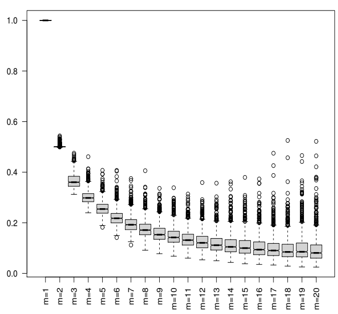
(f) λ_{U} estimation for $\theta = 2.92$ ($\lambda_{\mathrm{U}} = 0$).



(g) λ_{L} estimation for $\theta = 0.91$ ($\lambda_{\mathrm{L}} = 0$).



(h) λ_{L} estimation for $\theta = 1.86$ ($\lambda_{\mathrm{L}} = 0$).



(i) λ_{L} estimation for $\theta = 2.92$ ($\lambda_{\mathrm{L}} = 0$).

Figure 3: Box-plots of the estimation of the dependence coefficients Frank copula with parameters of $\theta = 0.91, 1.86, 2.92$.

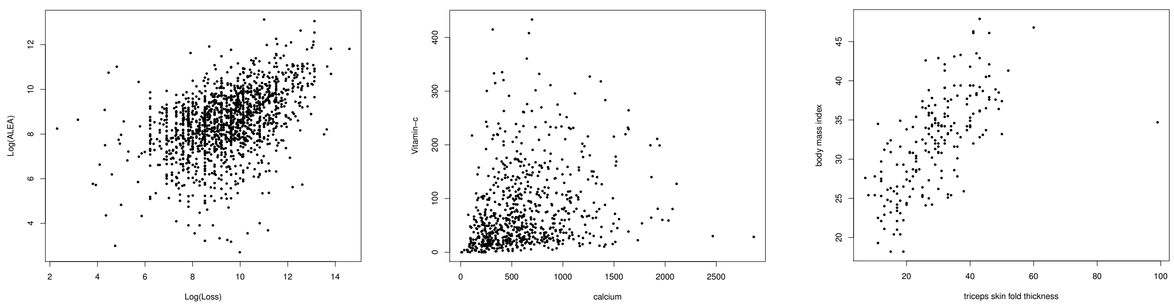
5. APPLICATIONS

To demonstrate the performance of new dependence coefficients estimation in previous sections, the Gumbel, Clayton and Frank copula is fit to the following three real data sets:

- The first data set is comprised of 1500 general liability claims randomly chosen from late settlement lags (Frees *et al.* [12]) and was provided by Insurance Services Office, Inc. Each claim consists of an indemnity payment (the loss) and an allocated loss adjustment expense (ALAE). The data is available in the `R` package “`copula`”. For simplicity, 34 censored data have not been used.

- According to the manual of `R`’s package “`lcopula`”, the nutrient data frame consists of quintuples consisting of four-day measurements for the intake of calcium, iron, protein, vitamin A and C for the women aged from 25 to 50 in the United States as part of the “Continuing Survey of Food Intakes of Individuals” program. The processed data has 737 measurements from a cohort study of the United States Department of Agriculture (USDA) and is available online at the University of Pennsylvania repository. The main concern is to estimate the dependence coefficients of Women’s daily nutrient intake of calcium and vitamin C.

- A population of women who were at least 21 years old, of Pima Indian heritage, and living near Phoenix, Arizona, was tested for diabetes according to World Health Organization criteria by using `R`’s package of “`MASS`”. The data were collected by the US National Institute of Diabetes and Digestive and Kidney Diseases. The training set “`Pima.tr`” contains a randomly selected set of 200 subjects. An application is illustrated for determining dependence coefficients of Triceps skinfold thickness and body mass index in Pima Indian women.



(a) Scatter Plots of Loss-Alea data.

(b) Scatter Plots of Women daily nutrient intake of calcium and vitamin C.

(c) Scatter Plots of triceps skin fold thickness and body mass index in Pima Indian women.

Figure 4: Scatter plots of real data sets.

Figure 4 shows the scatter plots of the three data sets. When Figure 4 is examined, the dependence structure between involved random variables is obvious. In order to assess the goodness-of-fit results, the Cramér von Mises (CvM) statistic is used:

$$CvM = n \int_0^1 \left(\widehat{K}_n(t) - K_{\widehat{\theta}}(t)\right)^2 dK_{\widehat{\theta}}(t)\,, \tag{5.1}$$

where $\widehat{K}_n$ is the empirical Kendall distribution function as a non-parametric estimator of $K(t)$. The dependence parameter θ is estimated by means of the Pseudo-likelihood method. The statistic is evaluated by the relevant p-value obtained by running 10.000 Monte Carlo samples as the method is described in Berg [2] and Genest *et al.* [14]. All goodness-of-fit results and parametric estimation of dependence coefficients are presented in Table 7 while Table 8 provides the estimation results of τ, λ_{L} and λ_{U} based on Bézier curve for three data sets.

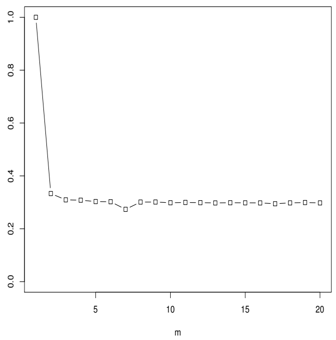
Table 7: Goodness-of-fit results based on $K(t)$ for three reel data sets.

Data	Copula	Parameter	$\widehat{\tau}$	$\widehat{\lambda}_{\mathrm{L}}$	$\widehat{\lambda}_{\mathrm{U}}$	CvM	p-value
Loss-Alea	Gumbel	1.4607	0.3154	0	0.3927	0.0414	0.8291
	Frank	3.0942	0.3154	0	0	0.2293	0.0292
	Clayton	0.9214	0.3154	0.4713	0	1.4181	0.0000
Calcium- Vit. C	Gumbel	1.2665	0.2104	0	0.2714	0.5627	0.0000
	Frank	1.9651	0.2104	0	0	0.3546	0.0011
	Clayton	0.5330	0.2104	0.2724	0	0.0505	0.6073
Thick.-Bmi	Gumbel	2.0933	0.5222	0	0.6074	0.1393	0.0221
	Frank	6.1568	0.5222	0	0	0.0711	0.2252
	Clayton	2.1866	0.5222	0.7283	0	0.2343	0.0014

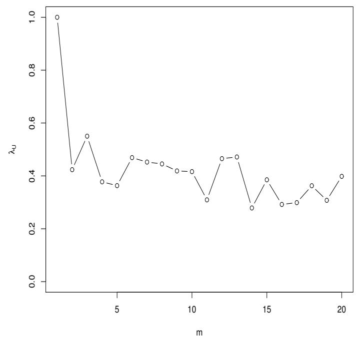
Table 8: The estimation of τ, λ_{U} and λ_{L} for three reel data sets.

Data	Est. Meth.	$m=5$	$m=10$	$m=15$	$m=20$
Loss-Alea	$\widehat{\tau}^m$	0.3030	0.2984	0.2981	0.2979
	$\widehat{\lambda}_{\mathrm{U}}^m$	0.3631	0.4161	0.3852	0.3982
	$\widehat{\lambda}_{\mathrm{L}}^m$	0.2267	0.1248	0.0897	0.0626
Calcium- Vit. C	$\widehat{\tau}^m$	0.2061	0.2051	0.2049	0.2075
	$\widehat{\lambda}_{\mathrm{U}}^m$	0.0523	0.0593	0.0116	0.0865
	$\widehat{\lambda}_{\mathrm{L}}^m$	0.2863	0.2472	0.2568	0.2687
Thick.-Bmi	$\widehat{\tau}^m$	0.4769	0.4694	0.4668	0.4651
	$\widehat{\lambda}_{\mathrm{U}}^m$	0.2517	0.0783	0.0062	0.0001
	$\widehat{\lambda}_{\mathrm{L}}^m$	0.3826	0.2049	0.1143	0.0724

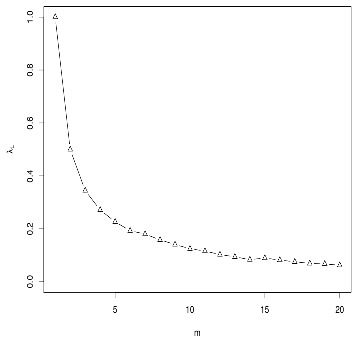
The results in Table 7 represent that Gumbel copula is a good choice for variables Loss-Alea with a p-value of 0.8291. It is concluded from Table 8 that as the degree of polynomial increases, estimation of λ_{U} and λ_{L} approach to the parametric estimate of dependence coefficients of Gumbel copula for Insurance data. For Calcium and Vitamin-C data, Clayton copula fits the data well with p-value of 0.6073. For the estimation of λ_{L}, $\widehat{\lambda}_{\mathrm{L}}^{20}$ is closure to the parametric estimate of $\lambda_{\mathrm{L}} = 0.2714$. Also, it is obtained that the estimation of λ_{U} approaches to the parametric estimate of $\lambda_{\mathrm{U}} = 0$ as polynomial degree increases. For the triceps skinfold thickness and body mass index in Pima Indian women, Frank copula provides the best fit with p-value of 0.2252 from a statistical point of view. Tables 7 and 8 indicate that the estimation of λ_{U} and λ_{L} approaches to the parametric estimate of $\lambda_{\mathrm{U}} = 0$ and $\lambda_{\mathrm{L}} = 0$. In addition, Figure 5 shows the estimations of dependence coefficients for three real data sets depending on the polynomial degree $m = 1, ..., 20$. It can be concluded that, as the polynomial degree increases the estimation of dependence coefficients gets closure to the real values.



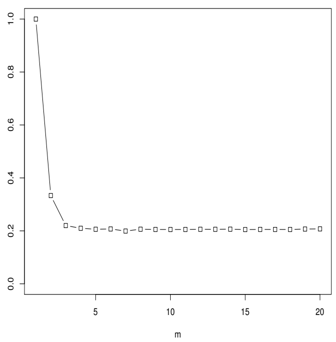
(a) Estimations of τ for Loss-Alea data set.



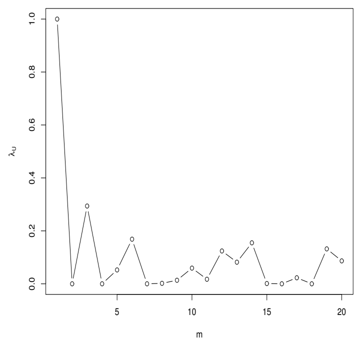
(b) Estimations of λ_{U} for Loss-Alea data set.



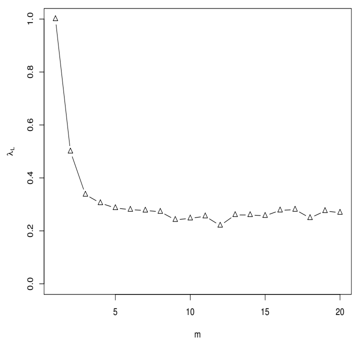
(c) Estimations of λ_{L} for Loss-Alea data set.



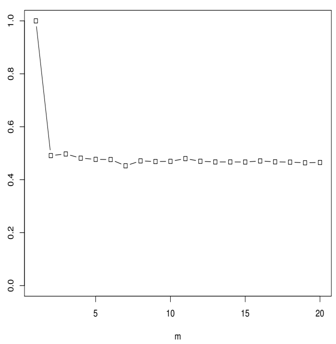
(d) Estimations of τ for Calcium- Vit. C data set.



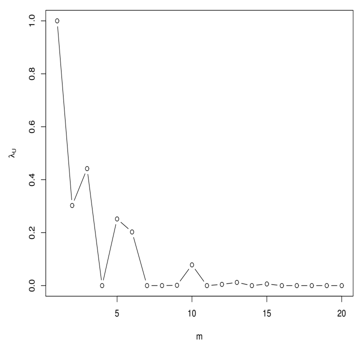
(e) Estimations of λ_{U} for Calcium- Vit. C data set.



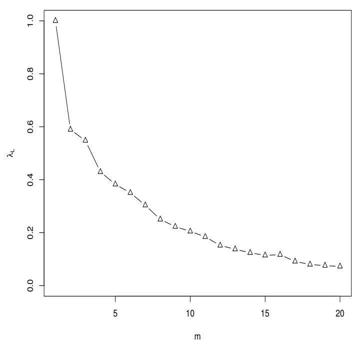
(f) Estimations of λ_{L} for Calcium- Vit. C data set.



(g) Estimations of τ for Thick.-Bmi data set.



(h) Estimations of λ_{U} for Thick.-Bmi data set.



(i) Estimations of λ_{L} for Thick.-Bmi data set.

Figure 5: Estimations of dependence coefficients of data sets for degree $m = 1, 2, ..., 20$.

6. CONCLUSION

In this study, a method of estimating the dependence coefficients of bivariate Archimedean family of copula is proposed. The Kendall's tau, lower tail dependence and upper tail dependence are estimated by using the Bézier curve. The new estimator of the dependence coefficients are flexible by the polynomial degree of m. A Monte Carlo simulation study is performed to measure the performance of the proposed estimation method for τ, λ_{U}, and λ_{L}. The performance according to the different levels of dependence size is investigated as well. The simulation results show that the new estimator of τ, λ_{U}, and λ_{L} presented a good performance. Besides, the new estimators of τ, λ_{U}, and λ_{L} indicated a satisfactory performance for the three data sets examined.

REFERENCES

[1] Barbe, P.; Genest, C.; Ghoudi, K. and Rémillard, B. (1996). On Kendall's process, *Journal of Multivariate Analysis*, **58**, 197–229.

[2] Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison, *The European Journal of Finance*, **15**, 675–701.

[3] Caillaulty, C. and Guégan, D. (2007). Empirical Estimation of Tail Dependence Using Copulas. Application to Asian Markets, *Quantitative Finance*, **5**, 489–501.

[4] Clayton, D.G. (1978). Model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, **65**, 141–151.

[5] Dimitrova, D.; Kaishev, K. and Penev, I. (2008). GeD spline estimation of multivariate Archimedean copulas, *Computational Statistics and Data Analysis*, **52**, 3570–3582.

[6] Duncan, M. (2005). *Applied Geometry for Computer Graphics and CAD*, Springer-Verlag, London.

[7] Embrechts, P.; McNeil, A. and Straumann, D. (2002). *Correlation and dependence in risk management: properties and pitfalls*. In "Risk Management: Value at Risk and Beyond?" (M.A.H. Dempster, Ed.), Cambridge, 176–223.

[8] Erdoğan, M.S.; Dişibüyük, Ç. and Oruç, Ö.E. (2019). An alternative distribution function estimation method using rational Bernstein polynomials, *Journal of Computational and Applied Mathematics*, **353**, 232–242.

[9] Ferreira, H. and Ferreira, M. (2012). Tail dependence between order statistics, *Journal of Multivariate Analysis*, **105**(1), 176–192.

[10] Ferreira, M. (2013). Nonparametric estimation of the tail-dependence coefficient, *REVSTAT – Statistical Journal*, **11**(1), 1–16.

[11] Frank, M.J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$, *Aequationes Mathematicae*, **19**, 194–226.

[12] Frees, E. and Valdez, E. (1998). Understanding relationships using copulas, *North American Actuarial Journal*, **2**, 1–25.

[13] Genest, C. and Mackay, R.J. (1986). Copules Archimédiennes et familles de lois bidimensionnelles dont les marges sont données, *The Canadian Journal of Statistics*, **14**, 145–149.

[14] Genest, C.; Quessy, J.F. and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transform, *Scandinavian Journal of Statistics*, **33**, 337–366.

[15] Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, **88**(423), 1034–1043.

[16] Goegebeur, Y. and Guillou, A. (2012). Asymptotically unbiased estimation of the coefficient of tail dependence, *Scandinavian Journal of Statistics*, **40**(1), 174–189.

[17] Gumbel, E.J. (1960). Distributions des valeurs extrémes en plusiers dimensions, *Publications de l'Institut de Statistique de l'Université de Paris*, **19**(9), 171–173.

[18] Kollo, T.; Pettere, G. and Valge, M. (2017). Tail dependence of skew t-copulas, *Communications in Statistics – Simulation and Computation*, **46**(2), 1024–1034.

[19] Michiels, F.; Koch, I. and De Schepper, A. (2011). A new method for the construction of bivariate Archimedean copulas based on the λ function, *Communications in Statistics – Theory and Methods*, **40**(15), 2670–2679.

[20] Nelsen, R.B. (2006). *An Introduction to Copulas*, Springer-Verlag, New York (NY).

[21] Schmidt, R. and Stadtmüller, U. (2006). Nonparametric estimation of tail dependence, *Scandinavian Journal of Statistics*, **33**, 307–335.

[22] Susam, S.O. and Ucer Hudaverdi, B. (2018). Testing independence for Archimedean copula based on Bernstein estimate of Kendall distribution function, *Journal of Statistical Computation and Simulation*, **88**(13), 2589–2599.

[23] Susam, S.O. and Ucer Hudaverdi, B. (2020). A goodness-of-fit test based on Bézier curve estimation of Kendall distribution, *Journal of Statistical Computation and Simulation*, **90**(13), 1194–1215.

[24] Susam, S.O. and Ucer Hudaverdi, B. (2022). On construction of Bernstein–Bézier type bivariate Archimedean copula, *REVSTAT – Statistical Journal*, **20**(3), 337–351.