Statistics and Applications {**ISSN 2452-7395** (online)}
Volume 17, No. 2, 2019 (New Series), pp 23-36

On A New Bivariate One Parameter Archimedean Copula Function and Its Application to Modeling Dependence in Climate and Life Sciences

Cigdem Topcu Guloksuz and Pranesh Kumar
Department of Mathematics & Statistics, University of Northern British Columbia, V2N 4Z9, Prince George, BC, Canada

Received: January 07, 2019; Revised: November 29, 2019; Accepted: November 29, 2019

Abstract

Copulas provide models to describe the dependence structure between two or more random variables. This study focuses on a special class of copulas namely Archimedean copulas which have some nice mathematical properties. The easiness of generating of Archimedean copula by a generator function and defining a bivariate Archimedean copula by a univariate function are appealing properties which make Archimedean copulas popular to work with them. In this study, a new generator function is proposed to generate a new one parameter bivariate Archimedean copula. The new copula parameter is estimated and the tail dependence properties are presented. In application part of the study, Archimedean copulas are considered to model the dependence structure of the studied data sets. The studied data sets refer to α amylase levels in saliva experiment and the climate change parameters. Simulations to the studies are performed to generate data from the copula-based methodology which is implemented to estimate prediction models. Results are presented.

Key words: Archimedean copulas; dependency; generator function; climate change; radiative forcing; methane; saliva experiment

---

1. Introduction

Copula is a multivariate function of distribution functions which are themselves random variables. Since copulas connect the marginal distributions to their joint distribution function, they can be considered a dependence model for random variables. Abe Sklar first introduced copula as a term in his article Sklar (1959). For a brief introduction to copulas Belgorodski (2010), Frees and Valdez (1998), Genest and Favre (2007), Joe (1997), Matteis De(2001), Nelsen (2006), Sklar (1959), Sklar (1973) can be recommended. Applications of copula in finance and insurance field can be found in Belgorodski (2010), Frees and Valdez (1998), Embrechts et.al. (2002), Cherubini et.al. (2004). Modelling time to event data, competing risks problems and related subjects in survival analysis are discussed in Clayton (1978), Shih and Louis (1995), Wang and Wells (2000). Traditionally, measuring and summarizing dependencies of random variables have centered on correlation measures. However, several shortcomings of the well-known correlation measures such as Pearson correlation, Kendall's tau in modeling dependencies are studied and presented in Embrechts et.al. (2002). In this manner copulas are considered as alternative measures because of the flexibility they possess, Embrechts et.al. (2002). For example, copula functions allow for

Corresponding Author: Cigdem Topcu Guloksuz
E-mail: topcucigdem@gmail.com

describing dependence structure of random variables independently of their marginal, and also allow for asymmetric dependence unlike linear correlation coefficient.

Throughout the study, we focus on modeling dependencies with Archimedean copulas in bivariate context. The dependence structure between two random variables is completely described by known bivariate distributions. Although there are many bivariate distributions in literature, researchers need different models which are able to capture different types of dependence structures. Archimedean copulas can be generated by a generator function which have some particular properties. More detailed information about Archimedean copulas and corresponding generator functions can be found in Frees and Valdez (1998), Genest and McKay (1986), Genest and Rivest (1993), Hennesey and Lapan (2005), Hutchinson and Lai (1990), Nelsen (2006), Smith (2003).

In this study, bivariate Archimedean copula along with their properties and relationship between copula parameter and Kendall's tau are discussed in Section (2). In sub-section (2.1), a new generator function is proposed to generate a new one parameter bivariate Archimedean copula. The properties of the proposed generator function are presented and a new one parameter bivariate Archimedean copula is generated. The method of moments based on Kendall's tau is applied to estimate the parameter of the proposed Archimedean copula. In sub-section (2.2), the algorithm to simulate data from the Archimedean copula is described. The tail behavior of the proposed copula is studied and represented by the scatterplot in sub-section (2.3). The method for fitting copula to the data and comparing copula fits are given in sub-section (2.4). In section (3), copula-based methodology is considered to estimate linear prediction models for two data sets. Three well-known Archimedean copula Clayton, Clayton (1978), Gumbel, Gumbel (1960), Frank, Frank (1979) and the proposed Archimedean copula are employed and the prediction models are estimated by simulating data from the copulas. The minimum distance measure is used to specify an appropriate Archimedean copula which gives best possible fit to the data.

2. Archimedean Copulas

The bivariate cumulative distribution function H of any pair (X,Y) of random variables may be written in the form [22], [23].

$$H(x,y) = C(u,v), \quad u,v \in (0,1) \tag{1}$$

where u and v denote the marginal distributions $F(x)$ and $G(y)$ of X and Y, respectively. Here, C is the copula function with $C:[0,1]^2 \rightarrow [0,1]$ So the equation in (1) can be rewritten

$$H(x,y) = C(F(x),G(y)), \quad F(x),G(y) \in (0,1)\,. \tag{2}$$

It should be noted that, if the marginals are continuous, there is a unique copula representation, Sklar (1973). C copula function has the following properties:

1. C is symmetric, $C(u,v) = C(v,u), \ \forall u, \ v \in (0, 1)$.
2. C is associative $C(C(u,v),w) = C(u,C(v,w)) \ \forall u, \ v, \ w \in (0, 1)$
3. If a is a constant, $a\varphi$ is also a generator of C
4. $C(u,1) = u$ and $C(1,v) = v, \ \forall u, \ v \in (0, 1)$.

Archimedean copulas can be generated by a function that is called generator function. Generator function is defined as follows:

Definition 1: Let Φ be a class of functions $\varphi:[0,1]\to[0,\infty]$ is satisfying that

(i) $\varphi(1)=0$

(ii) $\varphi(0)=\infty$

(iii) $\varphi'(t)<0$, $0<t<1$

(iv) $\varphi''(t)>0$, $0<t<1$.

Thus, φ is a continuous, strictly decreasing and convex function and always has an inverse, φ^{-1}. By using a defined generator function, a bivariate Archimedean copula can be constructed in the way which is given below:

$$C(x,y)=\varphi^{-1}\left\{\varphi(F(x))+\varphi(G(y))\right\}. \tag{3}$$

If $\varphi(0)<\infty$ the generator function is called non-strict and is also capable of generating an Archimedean copula.

The pseudo-inverse of non-strict generator function exists and it is defined by

$$\varphi^{[-1]}(t)=\begin{cases}\varphi^{-1}(t), & 0\le t\le\varphi^{-1}(t)\\ 0, & \varphi^{-1}(0)\le t\le\infty\end{cases} \tag{4}$$

Note that, an Archimedean copula that is generated by a non-strict generator function takes the form

$$C(u,\ v)\ =\ \max(C(u,\ v),\ 0). \tag{5}$$

For Archimedean copulas, Kendal's tau can be written in copula form as follows

$$\tau=\iint_{\mathrm{I}} C(u,\ v)dC(u,\ v)-1=4E\left[C(U,\ V)\right]-1. \tag{6}$$

This relationship is useful to estimate copula parameter. The method of moments based on Kendall's tau can be used to estimate copula parameter. The properties of this method and the estimator are studied and presented in Genest and MacKay (1986), Genest and Rivest (1993), Kojadinovic and Yan (2010).

One of the appealing properties of Archimedean copulas is that a bivariate Archimedean copula can be uniquely determined by a univariate function. This univariate function, $K(t)$, is called Kendall distribution function and defined as:

$$\Pr(C(U,V)\le t)=K(t)=t-\frac{\varphi(t)}{\varphi'(t)},\ \ 0<t<1 \tag{7}$$

Here, $K(t)$ is the distribution function of an Archimedean copula and the expression in (6) can be re-expressed as follows:

$$\tau=4\int_0^1\frac{\varphi(t)}{\varphi'(t)}dt+1 \tag{8}$$

In the following sub-section 2.1, a new generator function is proposed to generate a new bivariate Archimedean copula and the properties of the new Archimedean copula are studied.

2.1. Bivariate Archimedean copula

Considering that different generator functions generate different Archimedean copulas and recalling the Definition 1, new generator function can be defined.

The proposed new generator function, $\varphi:[0,1]\to[0,\infty]$ is defined

$$\varphi(t)=e^{\theta(1-t)}-1,\quad \theta>0,\ \ 0<t<1 \tag{9}$$

The properties in Definition 1 are checked for the proposed function as follows:

(i) $\varphi(1)=0\Rightarrow\varphi(1)=(e^{0}-1)=0$

(ii) $\varphi(0)<\infty\Rightarrow\lim\limits_{t\to0^{+}}(e^{\theta(1-t)}-1)=(e^{\theta}-1)<\infty$

(iii) $\varphi'(t)=-\theta e^{-\theta(t-1)}<0,\quad 0<t<1$

(iv) $\varphi''(t)=\theta^{2}e^{-\theta(t-1)}>0,\quad 0<t<1$

It can be seen that, because of $\varphi(0)<\infty$, the proposed function in (9) is a non-strict generator function. It can be used to generate an Archimedean copula. Its pseudo-inverse, $\varphi^{[-1]}:(0,\infty)\to(0,1)$ is defined as follows

$$\varphi^{[-1]}(t)=\begin{cases}1-\dfrac{1}{\theta}\ln(1+t), & 0\le t\le\varphi(0)\\ 0, & \varphi(0)\le t\le\infty\end{cases} \tag{10}$$

Definition 2: Let φ be a generator function that is defined in (6). Then $C:[0,1]\times[0,1]\to[0,1]$ is a bivariate Archimedean copula that is generated by φ and has the form

$$C(u,v)=\begin{cases}1-\dfrac{1}{\theta}\ln(e^{\theta(1-u)}+e^{\theta(1-v)}-1), & \varphi(u)+\varphi(v)\le\varphi(0)\\ 0, & ow.\end{cases} \tag{11}$$

The copula presented in (11) is indexed by a parameter θ that is called copula parameter. Recalling (7) and (8), the distribution function and the parameter estimation of the new bivariate Archimedean copula are given, respectively.

$$K(t)=t-\frac{e^{\theta(1-t)}-1}{-\theta e^{-\theta(t-1)}} \tag{12}$$

$$\tau=4\int_{0}^{1}\frac{e^{\theta(1-t)}-1}{-\theta e^{-\theta(t-1)}}dt+1 \tag{13}$$

By solving the integral (13),

$$\tau=4\left(\frac{1-e^{-\theta}-\theta}{\theta^{2}}\right)+1 \tag{14}$$

is obtained. Figure 1 illustrate the relationship between the parameter of the proposed copula and Kendall's tau.

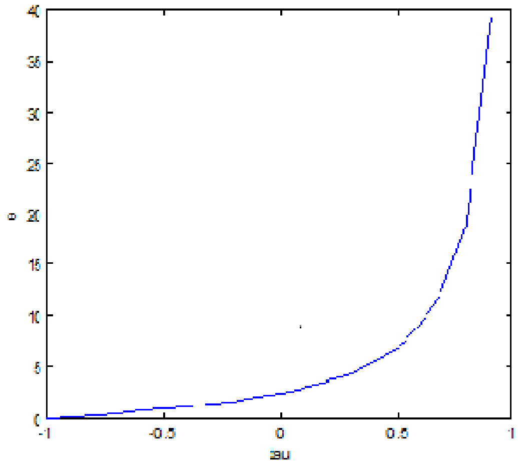

Figure 1: New Copula Parameter and Kendall's τ

The corresponding generator and distribution functions of the studied Archimedean copulas in this study are presented in Table 1. Figure 2 shows the graphs of the distribution functions of the studied Archimedean copulas. The bivariate Archimedean copulas along with the relationship between their parameters and Kendall's tau are presented in Table 2. The plots of the studied copulas are illustrated in Figure 3.

Table 1: Distribution Functions of Archimedean Copulas

Family	$\varphi(t)$	$\varphi(t)$	$K(t)=t-\frac{\varphi(t)}{\varphi'(t)}$
Gumbel	$(-\ln t)^{\theta}$	$-\frac{\theta}{t}(-\ln t)^{\theta-1}$	$t-\frac{(t\ln t)}{\theta}$
Clayton	$\frac{1}{\theta}\left(t^{-\theta}-1\right)$	$-t^{-\theta-1}$	$t-\frac{\left(t^{\theta+1}-t\right)}{\theta}$
Frank	$-\ln(\frac{e^{-\theta t}-1}{e^{-\theta}-1})$	$\frac{\theta}{1-e^{-\theta t}}$	$t-\frac{1}{\theta}\ln(\frac{e^{-\theta t}-1}{e^{-\theta}-1})\left(1-e^{-\theta t}\right)$
NewCop.	$e^{\theta(1-t)}-1$	$-\theta e^{-\theta(t-1)}$	$t-\frac{e^{\theta(1-t)}-1}{-\theta e^{-\theta(t-1)}}$

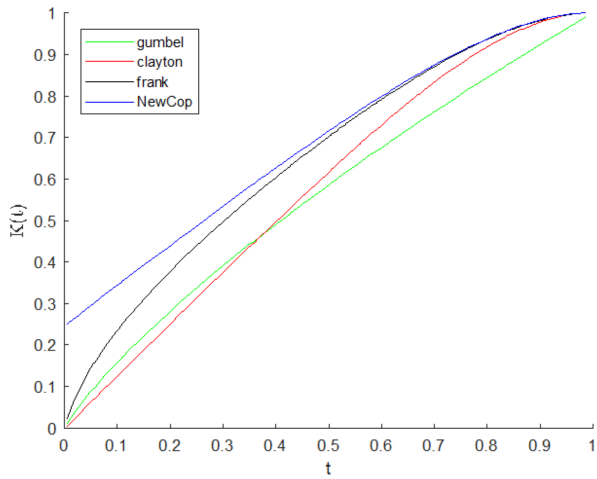

Figure 2: Distribution Functions of Archimedean Copulas

Table 2: Bivariate Archimedean Copulas and the Relationship with Kendall's τ

Family	Bivariate Copula	Dependence Parameter	Kendall's τ
Gumbel	$e^{-((-\ln u)^{\theta}+(-\ln v)^{\theta})^{1/\theta}}$	$\theta \geq 1$	$(\theta-1)/\theta$
Clayton	$((u)^{-\theta}+(v)^{-\theta}-1)^{-1/\theta}$	$\theta > 1$	$\theta/(\theta+2)$
Frank	$\left(-\frac{1}{\theta}\right)\ln\left\{\frac{\left(1-e^{-\theta}\right)-\left(1-e^{-\theta u}\right)\left(1-e^{-\theta v}\right)}{\left(1-e^{-\theta}\right)}\right\}$	$-\infty < \theta < \infty$	$1-\frac{4}{\theta}[D_1(\theta)]$
NewCop	$1-\frac{1}{\theta}\ln(e^{\theta(1-u)}+e^{\theta(1-v)}-1)$	$\theta > 0$	$4\left(\frac{1-e^{-\theta}-\theta}{\theta^2}\right)+1$

Here, $D_n(\theta)=\frac{n}{\theta^n}\int_0^{\theta} t^n/e^t-1dt,\ \ n>0$ is a Debye function.

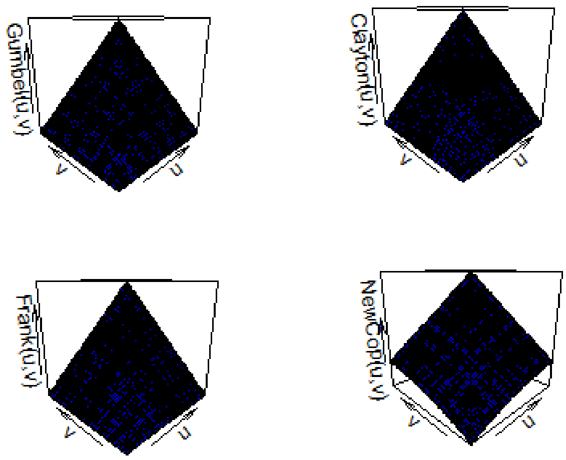

Figure 3: Plots of Archimedean Copulas

The sub-section 2.2 provides a method to simulate data from Archimedean copulas.

2.2. Generating random numbers from bivariate Archimedean copula

Generating random numbers from copulas is important for simulation studies, modelling, selecting random samples, etc. In this study generated random numbers are used to plot a scatter plot, which is a graphical tool to detect the tail dependence. The following steps are listed to generate random numbers from a bivariate copula.

Let (U,V) be a random pair from a bivariate Archimedean copula, $\varphi(t)$ is the generator function and $K(t)$ defined by (8) is the distribution function of copula. A pair of data (x_i, y_i) from a bivariate Archimedean copula can be generated by using the following procedure:

(i) Generate two independent random variables, p and p from Uniform $(0,1)$

(ii) $t = K^{-1}(q)$

(iii) $u=\varphi^{-1}[p\varphi(t)]$ and $v=\varphi^{-1}[(1-p)\varphi(t)]$

(iv) $x = F^{-1}(u)$ and $y = F^{-1}(v)$

Repeating the above steps (i) to (iv), n times n pairs of data (x_i, y_i), $i = 1,2,...,n$ can be generated.

Since the inverse of the distribution function of the proposed Archimedean copula, $K(t)$ which is defined in (12) doesn't have the closed form, Newton-Raphson numerical root finding method is applied to solve the following equation.

$$\left[t-\frac{\varphi(t)}{\varphi'(t)}\right]-q=0 \tag{15}$$

$$\left[t-\frac{e^{\theta(1-t)}-1}{-\theta e^{-\theta(t-1)}}\right]-q=0. \tag{16}$$

To understand the concept of the clustering of extreme event, tail dependence property of copulas is discussed in the following sub-section 2.3.

2.3. Tail dependence properties of bivariate Archimedean copula

The concept of tail dependence refers to clustering of extreme events. Modelling dependence of events, such as economic systems, natural hazards contexts generally exhibit tail dependence. It becomes very important to obtain accurate results especially in tail ends. The definition of tail dependence is the limiting probability that a random variable exceeds a certain threshold, given that another random variable already exceeds that threshold.

More formally definitions of the upper and lower tail dependence of a bivariate copula $C(u,v)$ are given as follows, respectively in (17) and (18).

$$\lambda_U=\Pr\left[F(x)>u\,|\,G(y)>u\right]=\lim_{u\to 1^-}\frac{1-2u+C(u,u)}{1-u} \tag{17}$$

$$\lambda_L=\Pr\left[F(x)<u\,|\,G(y)<u\right]=\lim_{u\to 0^+}\frac{C(u,u)}{u}\ . \tag{18}$$

If the above limits exist, $\lambda_U(\lambda_L)=0$ shows that the copula has no upper (lower) tail dependence. In case of, $\lambda_U(\lambda_L)\neq 0$ there is upper (lower) tail dependence. The upper and lower tails of the new copula in (8) are examined, respectively.

$$\lambda_U=\lim_{u\to 1^-}\frac{2-2u-\frac{1}{\theta}\ln(e^{\theta(1-u)}-e^{\theta(1-v)}-1)}{1-u}\to 2 \tag{19}$$

$$\lambda_L=\lim_{u\to 0^+}\frac{1-\frac{1}{\theta}\ln(e^{\theta(1-u)}-e^{\theta(1-v)}-1)}{u}\to\infty\ . \tag{20}$$

The tail dependence coefficients of studied bivariate Archimedean copulas are listed in Table 3 and Figure 4 demonstrates the tail dependence of the considered copulas. The scatter plots visualize that Gumbel copula has an upper tail, Clayton copula has a lower tail, Frank copula has no tail and the proposed Archimedean copula has both lower and upper tail dependencies.

Table 3: Tail Dependence Coefficients of Considered Archimedean Copulas

Family	λ_L	λ_U
Gumbel	0	$2-2^{1/\theta}$
Clayton	$2^{-1/\theta}$	0
Frank	0	0
New Cop.	∞	2

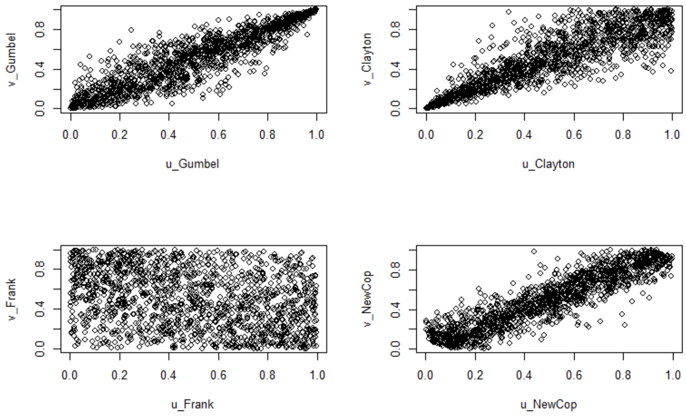

Figure 4: Scatter Plots of Considered Archimedean Copulas

In the following sub-section (2.4), we describe the method of fitting copula.

2.4. Fitting copula to data

Genest and Rivest (1993) suggested a nonparametric approach to select the appropriate bivariate Archimedean copula which gives the best fit to the data. The estimation procedure consists of mainly two steps. First one is to estimate the marginal distributions and the second one is to specify the copula function. Marginal distributions can be estimated by empirical or parametric ways. The procedure which is followed in this study is summarized as follows:

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate population (X, Y) with distribution functions $F(x)$ and $G(y)$, respectively.

(i) Estimate the copula parameter.

(ii) Obtain the empirical estimate of distribution of copula function, say $K_n(t)$. First, define the pseudo-observations,

$$T_i = \sum_{i=1}^{n} \frac{I\left(X_j \le X_i \,\&\, Y_j \le Y_i\right)}{n+1}, \; i = 1, \ldots, n \text{ and then calculate}$$

$$K_n(t) = \frac{\sum_{i=1}^{n} I(T_i \le t)}{n+1}, \; i = 1, \ldots, n$$

(iii) Construct parametric estimate of $K_\varphi(t) = t - \frac{\phi(t)}{\phi'(t)}$

(iv) Compare the distance between $K_n(t)$ and $K_\varphi(t)$

Comparing $K_n(t)$ and $K_\varphi(t)$ can be done in several ways. For instance, by considering information criteria such as Akaike Information Criterion, Bayesian Information Criterion and log likelihood. In this study, we follow Frees and Valdez (1998) and use the following minimum distance measure.

$$MD = \int \left[K_n(t) - K_\varphi(t) \right]^2 dK_n(t). \tag{21}$$

In the following Section 3, we illustrate the applications of copulas with the help of two examples.

3. Application

In this section, we apply the new copula along with other well-known Archimedean copulas namely Clayton, Gumbel and Frank to fit the linear prediction models to the data from two examples described below. We follow the copula-based methodology as described and studied in Kumar and Shoukri (2008, 2007, 2011). Using the copulas, new data sets which have the similar dependence structure and sample sizes as in the actual data sets are simulated choosing the 50, 150, 250 and 350 runs. For each data set, linear prediction models are estimated and the %95 confidence intervals of model parameters are computed. The estimated prediction models from the fitted copulas are compared using the mean absolute prediction error (MAPE) measure.

3.1. Copula applications for α -amylase levels in saliva experiment

This example refers to the study of saliva content which is an enzyme called amylase and which hydrolyses starch into maltose. The experimental data on α -amylase levels in saliva are considered. The data set can be found in Brunner et al. (2004). The α -amylase levels in saliva are measured on different times and in a day. In this study, we have considered the data set which consists of the 14 measurements of α amylase levels in saliva taken on Thursday at 12 a.m. (independent variable, X) and 9 p.m. (dependent variable, Y).

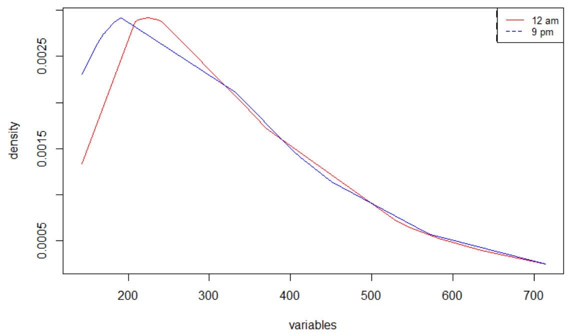

Figure 5: Marginal Fitting to Amylase Levels

To specify the dependence structure between the amylase levels in saliva at two time points, *i.e.*, between Y and X, their marginal distributions are estimated as the Log Normal (mean = 5.6210, sd = 0.6988) for X and the Log Normal (mean = 5.6525, sd = 0.5487) for Y, see Figure 5. Kendall's Tau between two amylase levels is estimated as τ = 0.64835. The copula parameters are estimated based on the Kendall's Tau and the minimum distance measure (MD) are given in Table 4. Fitted copulas are then plotted and compared in Figure 6.

Table 4: Estimated Copula Parameters and Minimum Distance Measure for Amylase Levels

	Clayton	Gumbel	Frank	New Cop
$\hat{\theta}$	3.6875	2.8438	9.3816	10.2671
MD	0.0785	0.0312	0.0281	0.0208*

It may be noted from Table 4 that the minimum distance measure (MD) for the proposed new copula is 0.0208 and thus, new copula is the best fit compared to the Frank, Gumbel and Clayton copulas to represent the dependence structure between two amylase levels. New data sets (X,Y) of size 14 are simulated using the new copula 50, 150, 250 and 350 times. For each data set, linear prediction models are estimated along with the intercept and slope, standard error and the % 95 confidence interval of slope. Results are listed in Table 5.

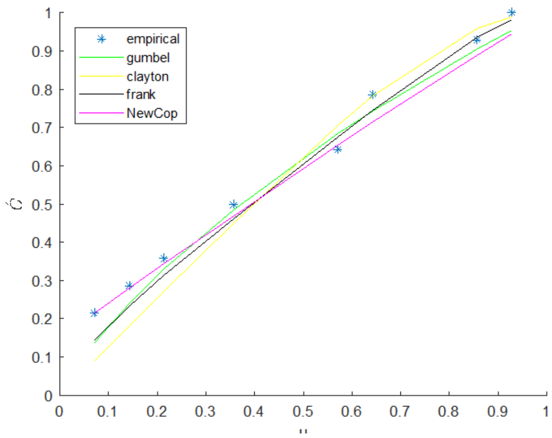

Figure 6: Copula Fitting for Amylase Levels

Table 5: Estimated prediction models and confidence intervals for amylase levels

	Intercept	**Slope (b)**	**Std.Error (b)**	**CI Lower**	**CI Upper**	**CI Width**
Data Model	73.3113	0.7443	0.1011	0.5462	0.9424	0.3962
New Copula Simulations						
50	66.0857	0.7942	0.1460	0.5080	1.0803	0.5723
150	67.3012	0.7741	0.1191	0.5406	1.0076	0.4670
250	68.5123	0.7625	0.0787	0.6082	0.9167	0.3085
350	66.4007	0.7278	0.0701	0.5904	0.8652	0.2748

In Table 5, it is noted that the estimated models from the actual data set and also from the simulated data sets have the intercept and slope estimates in close agreement with each other. For instance, the estimates of intercept, slope and the %95 confidence interval width of the slope in actual data set are 73.3113, 0,7443 and 0.3962, respectively, while these values for the new data set using 350 simulations run are 66.4007, 0.7278 and 0.2748, respectively. With regard to the comparison of the prediction errors, it is noted from Table 6 that the mean absolute prediction errors (MAPE) for actual data set and the new data sets with 50, 150, 250 and 350 simulation runs are 18.2229, 19.8664, 18.2299, 18.1937 and 16.0277, respectively.

However, as expected, when the number of simulation runs increases, the estimated models have smaller standard errors and narrow confidence intervals of slope estimation, and mean absolute prediction errors. Thus, having the smallest confidence interval width of the model parameters and also, the prediction errors, the proposed new copula-based prediction model may be recommended to study the relationship between two amylase levels.

Table 6: Mean Absolute Prediction Errors (MAPE) of Estimated Models of Amylase Levels

	MAPE
Data Model	18.2299
New Cop Sim.	
50	19.8664
150	18.2299
250	18.1937
350	16.0277

3.2. Copula applications for climate change indicators

The second example is about the climate change indictors [Source: Earth System Research Laboratory, Global Monitoring Division, https://www.esrl.noaa.gov/gmd/aggi]. Radiative Forcing (RF) is one of the climate change indicator which measures heating effect caused by greenhouse gases in the atmosphere. RF is calculated in watts per square meter, which represents the size of the energy imbalance in the atmosphere. Since RF (denoted by Y) is directly associated with the methane (CH4, one the of greenhouse gases), denoted by X, the prediction model of RF and CH4 will be useful to study their cause-and-effect relationship.

We fitted the marginal distributions of RF and CH4 from the given data set as Log Normal (mean = 0.9449, sd = 0.1032) and Log Normal (mean = –0.7279, sd = 0.0261), respectively, as seen in Figure 7. Kendall's tau is estimated as $\tau = 0.9600$ and used to estimate copula parameters. Following the copula fitting procedure in sub-section (2.2), copula parameter estimates and MD measures are given in Table 7 and copulas plotted in Figure 8 for the comparison purpose. From Table 7, the Gumbel copula has mean distance value MD = 0.0110, followed by MD = 0.0128 for the new copula. Thus, for this data set, Gumbel copula gives the best fit followed by the new copula to model the dependence structure between Radiative Forcing and CH4.

Using the Gumbel copula and the new copula, simulation studies are performed to generate data sets by having the number of simulation runs as 50, 150, 250 and 350 and the sample sizes 28. The prediction models are estimated and the results are listed in Table 8.

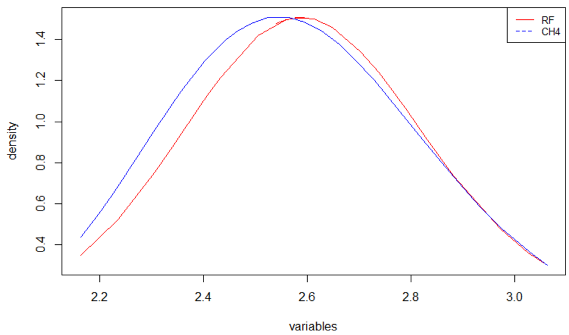

Figure 7: Marginal Distributions for RF and CH4

Table 7: Estimated Copula Parameters and Minimum Distance Measure for RF and CH4 Levels

	Clayton	**Gumbel**	**Frank**	**New Cop**
$\hat{\theta}$	48.0384	25.0198	98.4040	99.0667
MD	0.0138	0.0110*	0.0162	0.0128

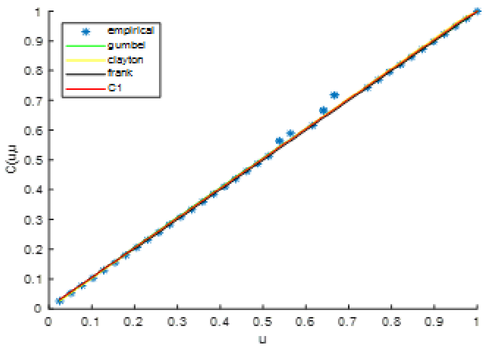

Figure 8: Copula Fitting for RF and CH4

Table 8: Estimated Prediction Models and Confidence Intervals for RF and CH4

	Intercept	**Slope (b)**	**Std.Error (b)**	**CI Lower**	**CI Upper**	**CI Width**
DataModel	–7.3891	20.6498	0.8597	18.8826	22.4170	3.5344
50	–7.4207	20.7121	0.3404	20.0449	21.3793	1.3340
150	–7.4957	20.8635	0.3304	20.2207	21.5160	1.5150
250	–7.5152	20.9089	0.3244	20.2730	21.5447	1.2717
350	–7.5095	20.8964	0.3353	20.2392	21.5536	1.3144
NewCop Simulation						
50	–7.2123	20.2801	0.5471	19.2084	21.3529	2.1444
150	–7.4755	20.8260	0.5189	19.2084	21.8430	2.0340
250	–7.4574	20.7879	0.4946	19.8185	21.7574	1.9389
350	–7.4567	20.7873	0.4929	19.8213	21.7534	1.9322

Table 9: Mean Absolute Prediction Errors (MAPE) of Estimated Models of RF and CH4

	MAPE	
Data Model	0.01924	
	Gumbel Sim.	New Cop Sim.
50	0.01911	0.01941
150	0.01906	0.01908
250	0.01904	0.01909
350	0.01903	0.01924

From Table 8, we note that the estimated model parameters from the actual data set and also from the simulated data sets are in close agreement with each other. For instance, the estimates of intercept, slope and the %95 confidence interval width of the slope in the actual data set are –7.3891, 20.6498 and 3.5344, respectively, while these values using the Gumbel copula and 350 simulations run are –7.5095, 20.8964 and 1.3144, respectively. With regard to the comparison of the prediction errors, it is noted from Table 9 that the mean absolute prediction error (MAPE) for the actual data set is 0.01924, while for the 50, 150, 250 and 350 simulation runs, MAPE values are, respectively, 0.01911, 0.01906, 0.01904 and 0.01903 for the Gumbel copula and 0.01941, 0.01908, 0.01909 and 0.01924 for the new copula. Thus, the Gumbel copula followed by the proposed new copula may be recommended to model the dependence structure between the Radiative Forcing and CH4 and also to make predictions of the Radiative Forcing (heating effect) resulting from the levels of CH4.

4. Conclusion

In multivariate data sets, studying the dependence or specifying the pattern between random variables is commonly of main interest. Copulas have been used to model different types of dependence patterns between the random variables. Main advantage of working with copulas is that any kind of marginal distributions can be employed in simulating data sets. Therefore, copula-based methodology is an appropriate approach for modelling especially skewed data. Archimedean copulas are preferable in most applications due to their appealing mathematical properties and simple simulation algorithms. In observational studies, researchers may face different kinds of dependence structures and known models may be insufficient to represent the dependency between random pairs. Thus, to generate new and applicable models can be a solution. For this purpose, in this study, we have proposed a new generator function and discussed its properties. Based on this new generator function, a new bivariate Archimedean copula is constructed. Tail dependency of the new copula is examined. Copula based methodology is applied to prediction modeling in two applications, namely, to study α- amylase levels in saliva and to study effect of Methane, one of greenhouse gases on the Radiative Forcing (heating effect) in climate change. The results indicate that the new copula performs well compared to commonly used Archimedean copulas and can be applied in the copula based prediction models.

Acknowledgements

Authors wish to place on record their appreciation to the reviewer for critically examining the paper and providing the beneficial comments. The research was supported by the award of Scientific and Technological Research Council of Turkey (TUBITAK) Grant No. 2219 to the first author. First author also gratefully acknowledges the hospitality extended by the University of Northern British Columbia.

References

Belgorodski N. (2010) *Selecting Pair Copula Families for Regular Vines with Application to the Multivariate Analysis of European Stock Market Indices.* Master Thesis, Technische Universität München.

Brunner E., Domhof S. and Langer F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*. John Wiley & Sons, New York.

Cherubini U., Luciano E.E. and Vecchiato W. (2004). *Copula Methods in Finance*. John Wiley & Sons, New York.

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

Embrechts P., McNeil A. and Straumann D. (2002) Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*. Cambridge University Press, 176-223.

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

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

Genest C. and MacKay J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *Journal of the American Statistical Association - Theory and Methods*, **40**, 280-283.

Genest C. and Rivest L.P. (1993). Statistical inference procedures for bivariate archimedean copulas. *American Statistician*, **88**(**423**), 1034-1043.

Genest C. and Favre A. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, **12**, 347-368.

Gumbel E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**, 698-707.

Henessey D.A. and Lapan H.E. (2005). The use of Archimedean copulas to model portfolio allocations. *Mathematical Finance*, **12**, 143-154.

Hutchinson T.P. and Lai C.D. (1990). *Continuous Bivariate Distributions: Emphasizing Applications*, Rumsby Scientific Publishing, Sydney, Australia,

Joe H. (1997). *Multivariate Models and Dependence Concepts*. (1st Edition), Chapman & Hall Ltd, U.K.

Kumar P. and Shoukri M.M. (2008). Aortic stenosis using the Archimedean copula methodology. *Journal of Data Science*, **6**, 173-187.

Kumar P. and Shoukri M.M. (2007). Copula based prediction models: an application to aortic regurgitation study. *BMC Medical Research Methodology*, **7(21)**, 1-9.

Kojadinovic I. and Yan J. (2010). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance Mathematics and Economics*, **47**, 52-63.

Matteis De R. (2001). *Fitting Copulas to Data*. Diploma Thesis, Institute of Mathematics of University Zurich.

R. Nelsen. (2006). *An Introduction to Copulas*. (2nd Edition), Springer, New York.

Shoukri M.M., Kumar P. and Colak D. (2011). Analyzing dependent proportions in cluster randomized trials: Modeling inter-cluster correlation via copula functions. *Computational Statistics and Data Analysis*, **55**, 11226-1235.

Shih J.H. and Louis T.A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, 1384-1399.

Sklar A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications del'Institut de Statistique de l'Universite de Paris*, **8**, 229-231.

Sklar A. (1973). Random variables, joint distribution functions and copulas. *Kybernetika*, **9**, 449-460.

Smith M.D. (2003). Modelling sample selection using archimedean copulas. *The Econometrics Journal*, **6**, 99-123.

Wang W. and Wells M.T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association*, **95**, 62-76.