

Modeling Dependence Relationships of Anthropometric Variables Using Copula Approach

Funmilayo Westnand Oshogboye Saporu¹, Isaac Esbond Gongsin^{2, *}

¹National Mathematical Centre, Kwali, Abuja, Nigeria

²Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria

Email address:

saporuf@gmail.com (F. W. O. Saporu), gongsin2013@gmail.com (I. E. Gongsin)

*Corresponding author

To cite this article:

Funmilayo Westnand Oshogboye Saporu, Isaac Esbond Gongsin. Modeling Dependence Relationships of Anthropometric Variables Using Copula Approach. *American Journal of Theoretical and Applied Statistics*. Vol. 9, No. 5, 2020, pp. 245-255.

doi: 10.11648/j.ajtas.20200905.18

Received: September 3, 2020; **Accepted:** September 21, 2020; **Published:** October 22, 2020

Abstract: Copula model is introduced in modeling the co-dependence structures of anthropometric variables-Body mass index (BMI), Abdominal circumference, Adiposity and Percent body fat-because it can capture monotonic dependence. Four copula-based Kumaraswamy-epsilon distributions are derived and used to determine the best fit to the anthropometric data, these are new. These are the Gaussian, Clayton, Frank and Gumbel copulas. Clayton model provided the best fit in four bivariate pairs-BMI and Percent body fat, BMI and Abdominal circumference, Adiposity and Abdominal circumference and Abdominal circumference and Percent body fat-while Gaussian is best for BMI and Adiposity pair and Frank is best for Adiposity and Percent body fat pair. Copula-based Kendall's tau and tail dependence are used as estimates for measuring the strength of the co-dependence. The results strongly recommend the use of BMI as an anthropometric index for estimating human body composition of adiposity. However for individuals with BMI values in the two extreme tails, their adiposity should be measured directly. The results do not find any suitable anthropometric indices for estimating percent body fat and therefore is recommended that for such epidemiological research, percent body fat should be measured directly. The results also clearly show that the Kendall's tau and the corresponding Pearson correlation coefficient estimates are largely at variance whenever the co-dependence structure cannot be described as linear dependence. This can prompt contradictory conclusions. It is therefore suggested that for such research, whenever Pearson correlation coefficient method is in use, a coefficient of determination of a minimum of 75% should be obtained before any anthropometric index can be recommended for body composition substitution.

Keywords: Anthropometric Index, Body Composition, Correlation Matrix, Inference Function for Margin, Kendall's Tau, Kumaraswamy-epsilon Distribution, Monotonic Dependence

1. Introduction

In the study of dependence among variables, Pearson's correlation coefficient is the measure of dependence most widely used. It is actually a measure of linear dependence and not general dependence. The other measures of dependence are the Kendall's tau and Spearman's rho. These are distribution-free methods. When a bivariate distribution can assume an elliptical form (examples multivariate normal or multivariate Student), then the dependence structure among the variables is linear and as such the use of Pearson's

correlation coefficient is appropriate. But when the distribution is non-elliptical, the use of Pearson's correlation coefficient may lead to misleading conclusions [1]. Consequently, there is a need for alternative measures of dependence that is appropriate when the bivariate distribution is non-elliptical. The copula-based ones are in use.

Copula provides a link between a bivariate (multivariate) distribution and its component marginal distributions. It has the advantage that the appropriate marginal distributions can be selected freely and be linked through a suitable copula. Copula-based measures of dependence measure the degree of monotonic dependence between two variables whereas linear

correlation measures the degree of monotonic linear dependence only. The use of copula-based measures for monotonic dependence instead of the linear correlation coefficient is suggested [13]. These authors also opined that a copula is invariant under increasing and continuous transformations of the marginals. Thus, the copula approach is recognized as a powerful tool for modeling dependence between variables. A diagram distinguishing between monotonic dependence and monotonic linear dependence is shown in Figure 1 below.

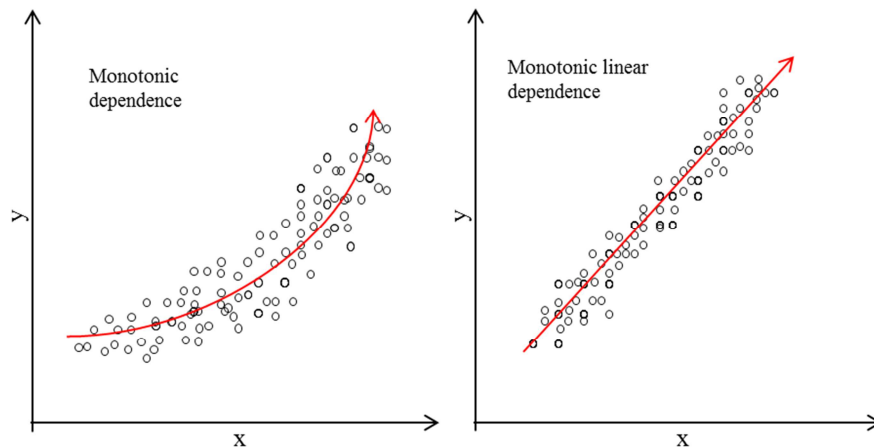


Figure 1. Monotonic and monotonic linear dependence.

2. Review of Copulas

A copula is a multivariate distribution function on a unit cube $[0,1]^k$ in \mathbb{R}^k with uniform marginal distributions. It relates an arbitrary distribution function F on \mathbb{R}^k to a copula C through the marginal distribution functions F_1, \dots, F_k . The name and theory of copula are rooted in Sklar's theorem [36], and the frequency of its appearance in the literature increased as from 1999 [15]. An elaborate article [13] motivated the application of copulas in the financial sector for assessment and management of risk in portfolio investments. Li argued "... why a copula function approach should be used to specify the joint distribution of survival times after marginal distributions of survival times are derived from market information, such as risky bond prices or asset swap spreads" [25]. Today, studies and applications of copulas have become very popular among academicians, engineers, economist, actuarial scientist, dynamic system modelers and more [14].

Many copula families are in use for constructing multivariate distributions; for example, elliptical, Archimedean, Archimax, and order statistics copulas. Comprehensive and elaborate studies on these families and some areas of application can be found in the literature, for examples [31, 18, 30]. Members of the elliptical copula family, for instance, are the normal and Student's t copulas, and they form a class of implicit copulas. These copulas have found application in modeling multivariate relationships in the financial sector [40, 25, 27, 23].

Archimedean copula family has the Clayton, Frank, Gumbel, Ali-Mikhail-Haq copulas as members. These have

In most literature, the study of co-dependence between anthropometric variables is done based on the assumption of a linear correlation. For example, it is found [7] that the linear correlation between body mass index (BMI) and some anthropometric variables is strong and positive. Other examples include [2, 3, 29, 37]. In this study we introduce a copula approach. That is, we construct copula-based bivariate Kumaraswamy-epsilon distributions and apply the same in modeling co-dependence between anthropometric variables. This is new.

enabled the construction of multivariate non-normal distributions that have found applications in finance [17, 38, 32, 21], hydrology [10, 34], wind speed modeling [8, 43, 42, 35], health [28, 22], to mention just a few. The Gumbel copula also belongs in the family of extreme value copulas.

3. Univariate Kumaraswamy-epsilon Distribution

The Kumaraswamy-epsilon distribution (henceforth denoted K-epsilon distribution) was introduced [16] as a new probability distribution with shapes similar to most lifetime distributions-for instance, gamma, Weibull and lognormal. It is a continuous probability distribution function of the Kumaraswamy-G family [9] with base epsilon probability distribution [12].

A continuous random variable X is distributed according to the K-epsilon distribution with parameters α, β, δ and λ if its probability density function is given by

$$f_X(x) = \alpha\beta\lambda \left(\frac{\delta^2}{\delta^2 - x^2} \right) (1-z)z^{\alpha-1}(1-z^\alpha)^{\beta-1} \quad (1)$$

where $z = 1 - \left(\frac{x+\delta}{\delta-x} \right)^{-\frac{\lambda}{2}}$, $0 < x < \delta$ and $\alpha, \beta, \delta, \lambda > 0$; α and β control the skewness and tail weight of the distribution, respectively.

Its cumulative distribution and quantile functions are, respectively, given by

$$F_X(x) = 1 - (1 - z^\alpha)^\beta \quad (2)$$

and

$$x_q = \delta \frac{\omega-1}{\omega+1} \quad (3)$$

where $\omega = \left(1 - \left(1 - (1-p)^{\frac{1}{\beta}}\right)^{\frac{1}{\alpha}}\right)^{-\frac{2}{\lambda\delta}}$, $0 < p < 1$.

4. Copula-based Bivariate K-epsilon Distributions

Let $X = (X_1, \dots, X_k)^T$ denote a random vector, whose joint distribution, F , is desired. Let the variables X_1, \dots, X_k be described by the marginal distribution functions F_1, \dots, F_k , respectively. Then, a copula C relates the marginal distributions to F [36] by

$$C(u_1, \dots, u_k) = F\left(F_1^{-1}(u_1), \dots, F_k^{-1}(u_k)\right) \quad (4)$$

where $x_i = F_i^{-1}(u_i)$, $i = 1, \dots, k$, are the quantile functions of the marginals.

When $C(\cdot)$ and $F(\cdot)$ are differentiable, then equation (4) satisfies

$$\frac{f(x_1, \dots, x_k)}{f_1(x_1) \dots f_k(x_k)} = c(u_1, \dots, u_k) \quad (5)$$

where $c(u_1, \dots, u_k) = \frac{\partial^k}{\partial u_1 \dots \partial u_k} C(u_1, \dots, u_k)$, $c(\cdot)$ and $C(\cdot)$ are the copula density and distribution functions,

$$c_\varphi(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\varphi^2}} \exp\left(-\frac{\varphi}{2(1-\varphi^2)}(\varphi\Psi_1^2 + \varphi\Psi_2^2 - 2\Psi_1\Psi_2)\right) \quad (7)$$

where $\Psi_i = \phi^{-1}(u_i)$, $i = 1, 2$, and $\phi^{-1}(\cdot)$ is the inverse standard normal distribution function.

The Gaussian copula-based bivariate K-epsilon probability density function is given by

$$f(x_1, x_2) = f_1 f_2 \frac{1}{2\pi\sqrt{1-\varphi^2}} \exp\left(-\frac{\varphi}{2(1-\varphi^2)}(\varphi\Psi_1^2 + \varphi\Psi_2^2 - 2\Psi_1\Psi_2)\right) \quad (8)$$

4.2. Clayton Copula-based Bivariate K-epsilon Distribution

The Clayton copula distribution and density functions are, respectively, given by

$$C_\varphi(u_1, u_2) = (u_1^{-\varphi} + u_2^{-\varphi} - 1)^{\frac{1}{\varphi}} \quad (9)$$

and

$$c_\varphi(u_1, u_2) = \frac{1+\varphi}{(u_1 u_2)^{1+\varphi}} (u_1^{-\varphi} + u_2^{-\varphi} - 1)^{-2-\frac{1}{\varphi}} \quad (10)$$

where $\varphi > 0$. Thus, the Clayton copula-based bivariate K-epsilon probability density is given by

$$f(x_1, x_2) = f_1 f_2 \frac{1+\varphi}{(u_1 u_2)^{1+\varphi}} (u_1^{-\varphi} + u_2^{-\varphi} - 1)^{-2-\frac{1}{\varphi}} \quad (11)$$

respectively.

The copula-based bivariate K-epsilon distributions that are of interest in this study are based on the normal (Gaussian), Clayton, Frank and Gumbel copulas. The copula functions are derived from equations (4) and (5) where the marginal distribution function $u_i = F_i(x_i)$, $i = 1, 2$, is specified by equation (2). In addition, the univariate K-epsilon density functions, $f_i = f_i(x_i)$, $i = 1, 2$, in each of the copula-based bivariate K-epsilon density function is specified in equation (1) above.

4.1. Gaussian Copula-based Bivariate K-epsilon Distribution

The Gaussian copula is an implicit copula presented in the form of the bivariate Gaussian cumulative distribution function. It is given by

$$C_\varphi(u_1, u_2) = \int_{-\infty}^{\phi^{-1}(u_2)} \int_{-\infty}^{\phi^{-1}(u_1)} \frac{1}{2\pi\sqrt{1-\varphi^2}} \exp\left(-\frac{1}{2(1-\varphi^2)}(t_1^2 + t_2^2 - 2\varphi t_1 t_2)\right) dt_1 dt_2 \quad (6)$$

where $-1 < \varphi < 1$ and φ denotes the copula parameter, which is copula-type specific.

The corresponding Gaussian copula density function is given by

4.3. Frank Copula-based Bivariate K-epsilon Distribution

The Frank copula distribution and density functions are, respectively, given by

$$C_\varphi(u_1, u_2) = -\frac{1}{\varphi} \log\left[1 + \frac{(e^{-\varphi u_1} - 1)(e^{-\varphi u_2} - 1)}{(e^{-\varphi} - 1)}\right] \quad (12)$$

and

$$c_\varphi(u_1, u_2) = \frac{\varphi e^{-\varphi(u_1+u_2)}(1-e^{-\varphi})}{[(1-e^{-\varphi})-(1-e^{-\varphi u_1})(1-e^{-\varphi u_2})]^2} \quad (13)$$

where $\varphi \geq -1 \setminus \{0\}$. Consequently, the Frank copula-based bivariate K-epsilon probability density function is, therefore, given by

$$f(x_1, x_2) = f_1 f_2 \frac{\varphi e^{-\varphi(u_1+u_2)}(1-e^{-\varphi})}{[(1-e^{-\varphi})-(1-e^{-\varphi u_1})(1-e^{-\varphi u_2})]^2} \quad (14)$$

4.4. Gumbel Copula-based Bivariate K-epsilon Distribution

The Gumbel copula is both an Archimedean as well as an extreme-value copula. Its distribution and density functions are given, respectively, by

$$C_\varphi(u_1, u_2) = \exp\left\{-[(-\log u_1)^\varphi + (-\log u_2)^\varphi]^{\frac{1}{\varphi}}\right\} \quad (15)$$

and

$$c_{\varphi}(u_1, u_2) = \frac{C_{\varphi}(u_1, u_2)}{u_1 u_2} \frac{[(-\log u_1)(-\log u_2)]^{\varphi-1}}{[(-\log u_1)^{\varphi} + (-\log u_2)^{\varphi}]^{2-\frac{1}{\varphi}}} \left\{ [(-\log u_1)^{\varphi} + (-\log u_2)^{\varphi}]^{\frac{1}{\varphi}} + \varphi - 1 \right\} \quad (16)$$

where $\varphi \geq 1$. The Gumbel copula-based bivariate K-epsilon probability density function is, therefore, given by

$$f(x_1, x_2) = f_1 f_2 \frac{C_{\varphi}(u_1, u_2)}{u_1 u_2} \frac{[(-\log u_1)(-\log u_2)]^{\varphi-1}}{[(-\log u_1)^{\varphi} + (-\log u_2)^{\varphi}]^{2-\frac{1}{\varphi}}} \left\{ [(-\log u_1)^{\varphi} + (-\log u_2)^{\varphi}]^{\frac{1}{\varphi}} + \varphi - 1 \right\} \quad (17)$$

5. Copula Measures of Dependence

A copula-based measure of dependence measures the degree of monotonic dependence between two variables, whereas Pearson's correlation coefficient measures the degree of monotonic linear dependence only. Copula-based measures of dependence for the Kendall's tau and tail dependence can be derived from relations obtainable from the literature, see for example [1]. The expressions for these derived measures are presented in Table 1 below.

Table 1. Some copula-based dependence measures.

Copula	Copula-based Kendall's τ	Copula-based Tail Dependence
Normal	$\frac{2}{\pi} \sin^{-1}(\varphi)$	$\eta_t = \eta_u = 2 \lim_{x \rightarrow -\infty} \phi\left(x \frac{\sqrt{1-\varphi}}{\sqrt{1+\varphi}}\right) = 0$
Clayton	$\frac{\varphi}{\varphi+2}$	$\eta_t = 2 - \frac{1}{\varphi}$
Frank	$1 - \frac{4}{\varphi} [D_1(\varphi) - 1]$	NA
Gumbel	$1 - \frac{1}{\varphi}$	$\eta_u = 2 - 2^{\frac{1}{\varphi}}$

φ is the copula parameter and $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$, $k = 1, 2$ is the k^{th} order Debye equation

6. Review of the Anthropometric Variables and Impact on Human Health

Anthropometry is the scientific study of the measurements and proportions of the human body. They are applied in the textile industries for design purposes. For instance, there are 41 predefined feature lengths on the body most commonly used by the fashion industry [24] for footwear and clothing design, and they are also used by both working and household environments, to achieve the best match between products and their users [41]. However, the focus of this study is not for such but rather for their use in determining human health indices.

Certain anthropometric measures are used as indicators, or identifiers, of chronic human health risks. For instance, abdominal (waist) circumference is a relative determinant of

adiposity, also known as obesity; and obese persons have high risk of cardiovascular diseases and diabetes mellitus [5]. Abdominal circumference is also used as a complimentary measure to provide information on percent body fat [4]. Percent body fat is also an indicator of human health risks and its assessment is used commonly for categorization in health and sports performance-men and women with more than 25 and 30 percent body fat, respectively are considered obese [39] and stand the risks of hypertension, dyslipidemia and hyperglycemia [44]. Body mass index (BMI) is also a common measure of obesity and studies have shown that it correlates with percent body fat. It is found [29] that the range of correlation between BMI and percent body fat is from 0.61 to 0.85 within location and sex groups in Nigerian, Jamaican and African American populations. Many other references on the interdependence of anthropometric variables and their impact on human health can be found in the literature, see for examples [3, 6, 2, 26, 37].

Body composition (adiposity and percent body fat) can be measured directly. For example, percent body fat can be measured [2] using bioelectrical impedance analysis (BIA), but this is very expensive. Consequently, for health research, it is important to find a reliable easy-to-use method of determining body composition. Research so far has concentrated on exploring the correlation between anthropometric variables and body composition measures by the use of Pearson correlation coefficient. This informed the reasoning behind substituting, for example, BMI for body composition assessment. Here, we introduce copula model alternatives because they have the advantage of capturing general dependence as opposed to the case of Pearson correlation coefficient which assumes only linear dependence.

7. The Model

In most studies on anthropometric variables, Pearson moment correlation coefficient is employed as a measure of dependence. The Pearson's correlation coefficient is a measure of linear dependence and not general dependence as explained earlier. For example, it was found [29] that the functional relationship between percent body fat and BMI were quadratic in all location and sex groups in African American and Jamaican populations. This implies that the use of Pearson moment correlation coefficient to depict dependencies in this, and many other similar scenarios may lead to misleading conclusions. This creates the need for alternative methods for capturing co-dependence such as copula-based measures. A copula-based measure of dependence measures the degree of monotonic dependence between two variables, whereas Pearson's correlation coefficient measures the degree of monotonic linear dependence only.

Consequently, we introduce the copula approach to measure co-dependence among anthropometric variables. What it entails is as follows:

1. We model the co-dependence of variables in anthropometry using copula-based bivariate K-epsilon

distributions derived in equations (8, 11, 14) and (17).

2. We then estimate the copula-based Kendall's tau and tail dependence measures from the estimated copula parameter of the models.

Four copula models are used in this study. This is done in order to provide a greater depth in the search for appropriate models that can best describe the co-dependence structure between pairs of anthropometric variables examined.

8. Application

8.1. The Data

The data used for application are anthropometric measurements on four variables; namely, BMI, percent body fat, adiposity and abdominal circumference. The data were collected for 250 men between the ages 22 years and 81 years (average age is 44.88 years) and obtained from a sample data placed on Dr. John Ralph's statistics website:

[https://www2.statson.edu/jrasp/data.htm/body fat](https://www2.statson.edu/jrasp/data.htm/body%20fat). The data were originally on 252 men but two were removed as a measure of data cleaning. Weight and height were initially in pounds (lb) and inches (inch), respectively, and were converted to kilograms (kg) and metres (m) upon multiplying with respective factors 0.453592 and 0.0254. BMI was computed as weight divided by the square of height.

8.2. Correlation Matrix Plots

Four anthropometric variables were chosen based on their relevance in determining possible human health risks [33, 44, 5]. Preliminary correlation plots to show the trend in the scatter plots of variables in the study are presented in Figure 1 below. This serves as an essential prelude to any statistical analysis.

The scatter plots present varied pictures, suggestive of the possible copula model that can capture their co-dependence.

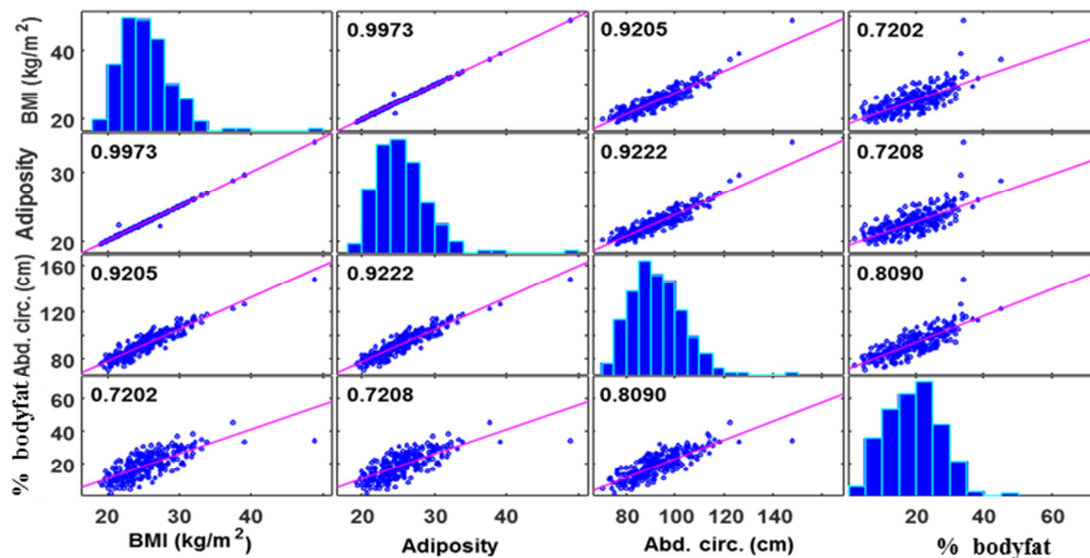


Figure 2. Plots of correlation and linear dependencies among anthropometric variables.

8.3. Fitting the Univariate Distributions

Since interest here is to fit the copula-based bivariate K-epsilon distributions in equations (8, 11, 14) and (17), we need to first find the appropriate marginal distributions. This

is determined by fitting the univariate K-epsilon and normal distributions to the datasets using *fitdistrplus* package in R. The results are presented in Tables 2 and 3, respectively.

Table 2. Estimates of parameters from univariate K-epsilon distribution fit to anthropometric data.

Variable	Parameter estimate (std error)				CvM (p-value)	AIC	Remark
	$\hat{\alpha}$ (se)	$\hat{\beta}$ (se)	$\hat{\lambda}$ (se)	$\hat{\delta}$ (se)			
BMI	48.166 (10.038)	3.815 (0.885)	0.122 (0.011)	57.793 (3.936)	0.3359 (> 0.1)	1349.47	Good fit
Adiposity	43.518 (9.380)	6.183 (1.860)	0.115 (0.011)	82.841 (20.484)	0.2845 (> 0.1)	1343.93	Good fit
Percent body fat	2.801 (0.224)	50.940 (9.609)	0.013 (0.004)	63.604 (17.688)	0.1919 (> 0.2)	1726.25	Good fit
Abd. circum.	154.995 (53.900)	3.933 (1.543)	0.046 (0.005)	234.888 (80.320)	0.0569 (> 0.8)	1887.69	Good fit

CvM denotes Cramer von Mises statistic value, AIC denotes Akaike Information Criterion, Abd. circum. denotes Abdominal circumference

Table 3. Estimates of parameters from univariate normal distribution fit to anthropometric data.

Variable	Parameter estimate (std error)		CvM (p-value)	AIC	Remark
	$\hat{\mu}$ (se)	$\hat{\sigma}$ (se)			
BMI	25.415 (0.229)	3.628 (0.162)	0.3941 (> 0.07)	1357.82	Good fit
Adiposity	25.448 (0.229)	3.615 (0.162)	0.4278 (> 0.06)	1355.99	Good fit

Variable	Parameter estimate (std error)		CvM (p-value)	AIC	Remark
	$\hat{\mu}$ (se)	$\hat{\sigma}$ (se)			
Percent body fat	18.963 (0.483)	7.631 (0.341)	0.0629 (> 0.7)	1729.57	Good fit
Abd. circum.	92.602 (0.675)	10.679 (0.478)	0.1821 (> 0.3)	1897.61	Good fit

CvM denotes Cramer von Mises statistic value, AIC denotes Akaike Information Criterion, Abd. circum. denotes Abdominal circumference

Both distributions fit the data. However, the values of the Akaike information criterion (AIC) suggest that the K-epsilon distribution performed better. Hence the choice of the K-epsilon distribution as marginals for determining the best copula model(s) is appropriate.

8.4. Fitting the Copula-based Bivariate K-epsilon Distributions

As derived above, we are using copula-based bivariate K-

$$L(\theta_1, \theta_2, \phi | x_1, x_2) = \sum_{j=1}^n \log c_{\phi} \left(F_1(x_{1j}, \theta_1), F_2(x_{2j}, \theta_2) \right) + \sum_{j=1}^n \log f_1(x_{1j}, \theta_1) + \sum_{j=1}^n \log f_2(x_{2j}, \theta_2) \quad (18)$$

where $\theta_i = (\alpha_i, \beta_i, \lambda_i, \delta_i)^T$ and $x_i, i = 1, 2$ is a vector of sample values x_{ij} .

Three methods can be applied for estimating the parameters of the log-likelihood function in equation (18). These are exact maximum likelihood (EML), inference function for margins (IFM) and canonical maximum likelihood (CML). The CML is a semi-parametric method that involves the use of the univariate empirical cumulative distribution function for the margins and plug into equation (19) below to estimate the copula parameter. This is not of interest in this study. The EML method involves estimating all the 9 parameters in equation (18) simultaneously. It produces inconsistent parameter estimates when the number of parameters is large and sample size small [20]. The IFM method is a sequential procedure that involves estimating the parameters of the marginal distributions first. That is, we evaluate $\hat{\theta}_1 = \arg \max \sum_{j=1}^n \log f_1(x_{1j}, \theta_1)$ and $\hat{\theta}_2 = \arg \max \sum_{j=1}^n \log f_2(x_{2j}, \theta_2)$, and then estimate the copula parameter using

epsilon distributions as models for studying the co-dependence structure of the anthropometric variables in the study. In order to further understand the nature of these dependences we are fitting four types of copula-based bivariate K-epsilon density functions, namely, Gaussian, Clayton, Frank and Gumbel copula-based density functions. These are given in equations (8, 11, 14) and (17), respectively. The general log-likelihood function of the model is given by

$$\hat{\phi} = \arg \max \sum_{j=1}^n \log c_{\phi} \left(F_1(x_{1j}, \hat{\theta}_1), F_2(x_{2j}, \hat{\theta}_2) \right) \quad (19)$$

The IFM gives parameter estimates that are consistent and asymptotically normal [20]. It produces efficient parameter estimates [19]. Also, in an unpublished paper, Xu suggest “that the IFM method is highly efficient compared with the (exact) MLE method” [18]. Its other advantage is that it turns out to be the best method when the number of parameters in the margins is large. With nine parameters in the bivariate K-epsilon distribution, we considered the IFM method an appropriate choice. Equation (19) was used to estimate the copula parameter by using the estimated parameters values for the margins from Table 2. The estimation was done using *optim* package in R. The results of various copula parameter estimates for dependence among anthropometric variables of BMI, percent body fat, adiposity and abdominal circumference are presented in Table 4 below.

Table 4. Copula parameter estimates for dependence among anthropometric variables.

Variables	Copula	$\hat{\phi}$ (se)	CvM (p-value)	AIC	Remark
BMI & Adiposity	Clayton	39.470 (3.752)	0.0144 (< 0.0005)	1293.2	Good fit
	Frank	106.000 (NA)	0.0039 (> 0.06)	1328.4	Not fit
	Gaussian	0.996 (0.081)	0.0210 (< 0.0005)	1199.4	Good fit
	Gumbel	52.606 (NA)	0.0027 (> 0.9)	1375.6	Not fit
	Clayton	1.421 (0.097)	0.2526 (< 0.0005)	124.4	Good fit
BMI & Percent body fat	Frank	6.158 (0.628)	0.0310 (< 0.05)	171.0	Good fit
	Gaussian	0.725 (4318)	0.0179 (> 0.3)	180.3	Not fit
	Gumbel	1.854 (22483)	0.0233 (> 0.1)	152.8	Not fit
	Clayton	3.663 (0.277)	0.3356 (< 0.0005)	330.2	Good fit
	Frank	13.090 (1.052)	0.0250 (< 0.04)	400.8	Good fit
BMI & Abd. circum.	Gaussian	0.919 (1282)	0.0156 (> 0.2)	444.0	Not fit
	Gumbel	3.337 (119470)	0.0129 (> 0.4)	395.0	Not fit
	Clayton	1.187 (NA)	0.2573 (< 0.0005)	48.3	Good fit
	Frank	6.130 (0.459)	0.0288 (< 0.05)	47.5	Good fit
	Gaussian	0.742 (143.2)	0.0160 (> 0.5)	51.5	Not fit
Adiposity & Percent body fat	Gumbel	1.250 (708.2)	0.0218 (> 0.1)	35.0	Not fit
	Clayton	3.753 (0.277)	0.3293 (< 0.0005)	333.0	Good fit
	Frank	13.420 (1.113)	0.0239 (< 0.05)	411.8	Good fit
	Gaussian	0.923 (80.36)	0.0151 (> 0.2)	457.4	Not fit
	Gumbel	3.461 (20202)	0.0138 (> 0.3)	415.4	Not fit
Percent body fat & Abd.	Clayton	183.9 (0.142)	0.2503 (< 0.0005)	187.1	Good fit

Variables	Copula	$\hat{\phi}$ (se)	CvM (p-value)	AIC	Remark
circum.	Frank	8.014 (0.770)	0.0303 (< 0.04)	253.0	Good fit
	Gaussian	0.811 (4.096)	0.0180 (> 0.2)	269.6	Not fit
	Gumbel	2.252 (43.75)	0.0344 (< 0.02)	240.6	Good fit

se denotes standard error, CvM denotes Cramer von Mises statistic value, AIC denotes Akaike information Criterion, Abd. circum. denotes Abdominal circumference, $\hat{\phi}$ is the copula parameter estimate

The bolded row in every pair of variables indicates the best fitted copula. The relative measure of Kendall's τ and lower tail dependence based on the best fitted copula in each paired combination are presented in Table 5. The corresponding

estimates of Pearson's correlation coefficient and the assessment of strength of dependence in each case are also indicated in Table 5.

Table 5. Estimates of Copula-based Kendall's τ and Tail Dependence based on copula and Pearson Correlation Coefficient.

Variable	copula	$\hat{\tau}_c$ (strength*)	$\hat{\rho}$ (p-value) (strength*)	$\hat{\eta}_l$
BMI & Adiposity	Gaussian	0.943 (v. strong)	0.997 (< 0.0001) (v. strong)	0.000
BMI & Percent body fat	Clayton	0.415 (moderate)	0.720 (< 0.0001) (strong)	0.614
BMI & Abd. circum.	Clayton	0.649 (strong)	0.921 (< 0.0001) (v. strong)	0.828
Adiposity & Percent body fat	Frank	0.589 (moderate)	0.721 (< 0.0001) (strong)	NA
Adiposity & Abd. circum.	Clayton	0.652 (strong)	0.922 (< 0.0001) (v. strong)	0.831
Percent body fat & Abd. circum.	Clayton	0.479 (moderate)	0.809 (< 0.0001) (v. strong)	0.686

$\hat{\tau}_c$ is the Copula-based Kendall's τ estimate, $\hat{\rho}$ is the Pearson correlation coefficient estimates, $\hat{\eta}_l$ is the estimate of lower tail dependence values, * as classified in www.statisticshowto.datasciencecentral.com.

Scatter plots of the observed and simulated data from the bivariate K-epsilon distribution based on the best copulas, and their respective marginal fit, are presented in Figures 3-5 for the dependence relationship between pairs of anthropometric variables.

9. Discussion of Results

In Table 4, the bolded row in every pair of variables indicates the best fitted copula. The results clearly show that Clayton copula is the best copula model for measuring the co-dependence in four out of six combinations of anthropometric variables examined. The Gaussian and Frank copulas are best in one each. Again from Figure 2, the scatter plots, superimposed

with the simulated data points from the respective best copula-based K-epsilon distribution, depict same.

The Gaussian copula model is compatible with the BMI and adiposity bivariate data. The Gaussian copula has an elliptical distribution as such the co-dependence between BMI and adiposity can be described by a linear dependence. Consequently, the Pearson linear correlation coefficient is an appropriate measure of this dependence. However, it should be noted that for the Gaussian copula there is no tail dependence. Indeed the coefficient of lower and upper tail dependence are zero. This means that irrespective of any high correlation coefficient estimate that may be obtained for BMI and adiposity, extreme events appear to occur independently.

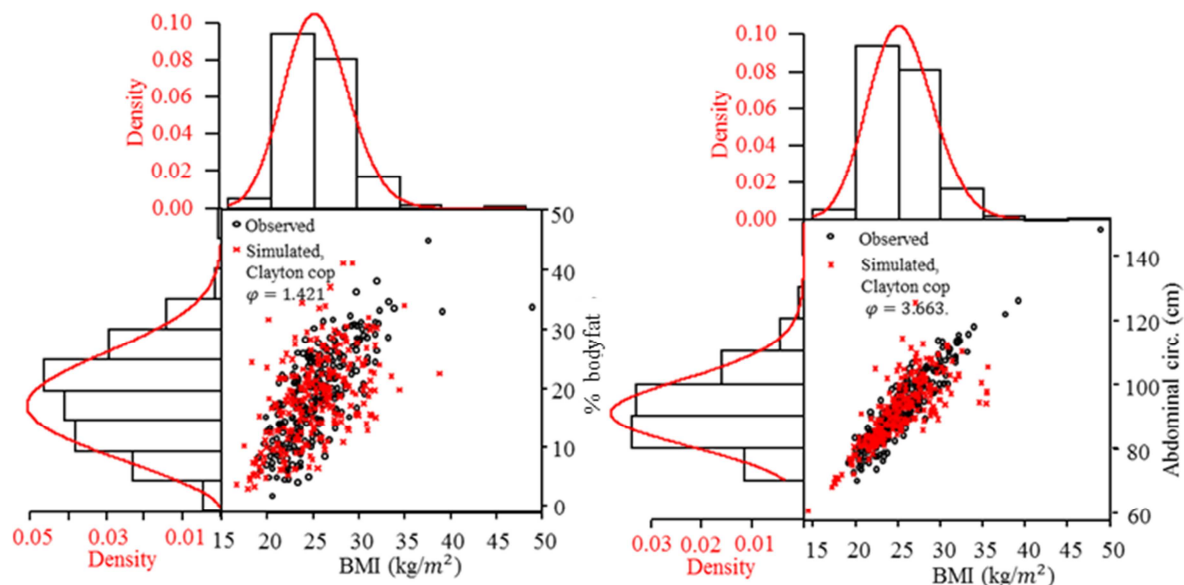


Figure 3. Scatter plots, simulated data points from the fitted bivariate K-epsilon distributions of the pairs BMI and Percent body fat and BMI and Abdominal circumference and their fitted marginal distributions

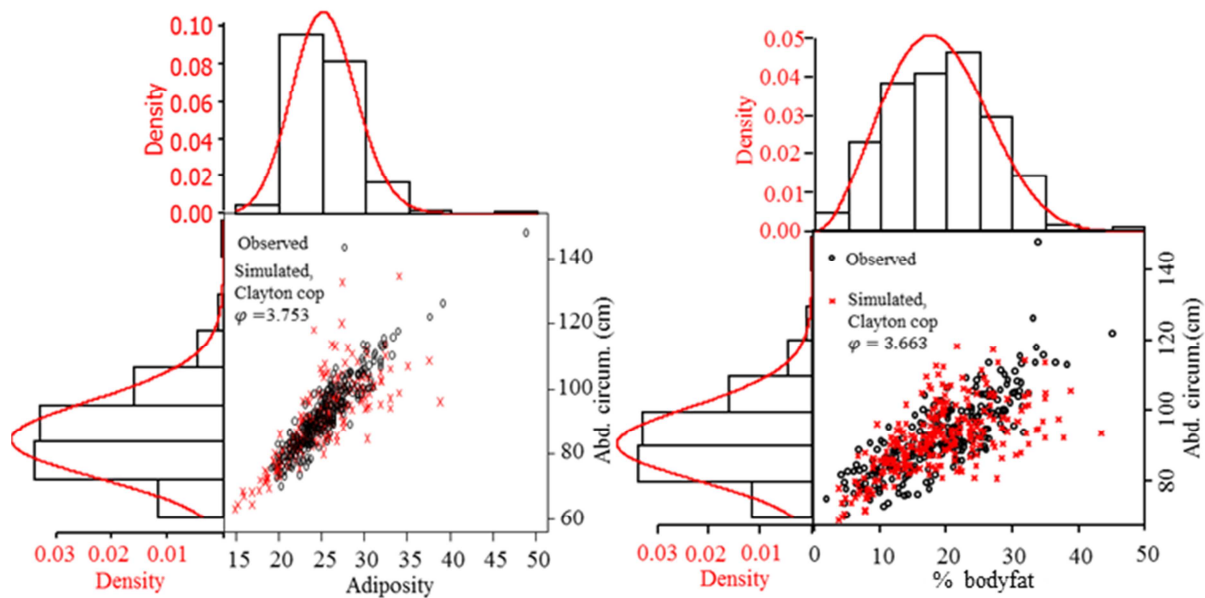


Figure 4. Scatter plots, simulated data points from the fitted bivariate K-epsilon distributions of the pairs Adiposity and abdominal circumference and Percent body fat and abdominal circumference and their fitted marginal distributions.

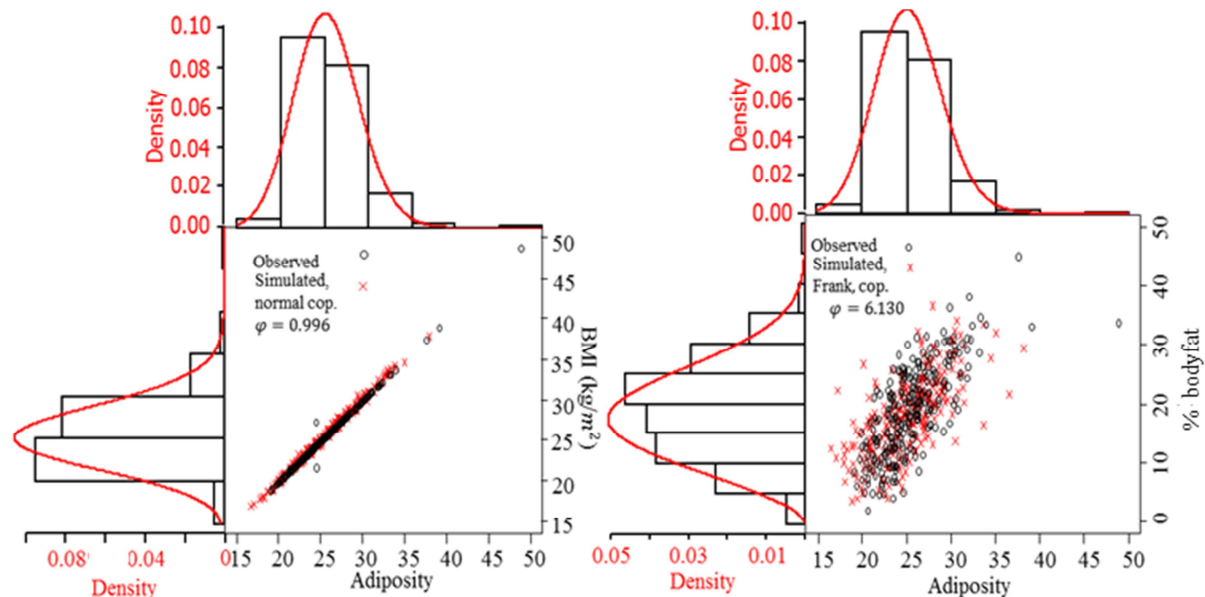


Figure 5. Scatter plots, simulated data points from the fitted bivariate K-epsilon distributions of the pairs Adiposity and Percent body fat and adiposity and BMI and their fitted marginal distributions.

The Clayton copula models are compatible with the bivariate data for the pairs; BMI and percent body fat, BMI and abdominal circumference, adiposity and abdominal circumference, and percent body fat and abdominal circumference. The Frank copula model is compatible with only adiposity and percent body fat bivariate data. The two types of copula models describe a co-dependence that is monotonic but not linear dependence. Copula-based measures of Kendall's tau and tail dependence are better [13] measures for this co-dependence than the linear correlation coefficient. It should be noted that Clayton copula has lower tail dependence. This means that at the lower extremes (not the upper), there is dependence between the two variables in the model in question.

The fitting of the various copula models to the bivariate data has thrown more insight into the nature of their co-dependence structure. But the measure of the co-dependence derivable from the model has far more practical implications. It provides a reliable base for the use of anthropometric measures as surrogates for estimating body composition (adiposity and percent body fat) in human health assessment.

The measures of co-dependence used in this study are the copula-based Kendall's tau and tail dependence, the results of which are tabulated in Table 5. The estimated values for the Kendall's tau range from 0.41 to 0.94; indicating moderate to very strong dependence. The highest value obtained is 0.94 for the co-dependence between BMI and adiposity. This indicates very strong dependence and supports the

substitution of BMI for body composition assessment of adiposity. As explained earlier for the Gaussian copula model there is no tail dependence. Hence, for individuals with values of BMI at the two extreme tails it is better to measure their adiposity directly. Kendall's tau estimates obtained for the measure of co-dependence between adiposity and abdominal circumference is 0.65 indicating strong dependence. Hence abdominal circumference can also be used as a substitute for the assessment of adiposity but BMI is a better substitute. It should be noted that here, the copula model that fits the bivariate data of adiposity and abdominal circumference is Clayton, having a non-elliptical distribution with no upper tail dependence. This suggests that for individuals having abdominal circumference values in the extreme upper tails their adiposity should be measured directly.

It is observable that in this case, the Kendall's tau estimate of dependence measure and that obtained by Pearson correlation coefficient are approximately the same and suggestive of the same conclusion. It is not surprising because the Gaussian copula model suggests a linear dependence for which the Pearson correlation coefficient is also an appropriate measure. In each of all other pairs, the corresponding values of these estimates are conspicuously at variance-suggestive of contradictory conclusions. For example, consider the pair percent body fat and abdominal circumference, where a Kendall's tau estimate of 0.478 is obtained. This is within the limits of moderate dependence, hence the substitution of abdominal circumference for percent body fat cannot be strongly recommended. On the other hand, the corresponding estimate of Pearson correlation coefficient is 0.809. This is within the limit of very strong dependence and therefore, on the contrary, strongly suggests a substitution. This highlighted contradiction reflects in the results of all the other bivariate pairs. This is not surprising because for these pairs, the copula models have non-elliptical distribution for which Kendall's tau is a more appropriate measure of co-dependence than the Pearson's correlation coefficient. The contradiction highlighted supports the notion expressed in some literatures [1] that the use of Pearson's correlation coefficient in estimating co-dependence for non-elliptical bivariate distributions may lead to misleading conclusions.

Kendall's tau estimates for percent body fat and each of the anthropometric variables (BMI and abdominal circumference) are in the region of 0.4, which indicate only moderate dependence and consequently not suggestive of a substitution. However, it is interesting to look at the result of BMI and percent body fat pair as this provides an opportunity to contribute to the debate [29] of using BMI as a surrogate for percent body fat. Here a Kendall's tau estimate is 0.415, not suggestive of using BMI as a substitute. But on the contrary, Pearson's correlation coefficient estimate, 0.72, strongly supports using BMI as a substitute. It is therefore, tempting to adopt the conclusion from the Pearson correlation result more so when the estimated value is highly significant. However, if we assume a linear regression model

for this dependence, a correlation coefficient estimate result of 0.72 will imply a coefficient of determination of 52%. That is, fifty-two percent of the variation in percent body fat is explained by fitting the linear regression model. The unexplained variation is 48%. This is too high and suggests that there are other important variables in the dependence structure that are not accommodated in the linear model. Hence, this estimated correlation coefficient obtained does not strongly suggest the use of BMI solely as a substitute. We, therefore, go with the suggestion [29] that for epidemiological research, there should be a direct measurement of percent body fat instead of using BMI as a substitute. We go further to suggest that for this study when Pearson correlation coefficient method is used, a coefficient of determination value of 75% should be obtained before any anthropometric variable can be recommended as a substitute for estimating body composition variables.

10. Conclusion

Copula-based bivariate K-epsilon distributions are derived. They are four copula types-Gaussian, Clayton, Frank and Gumbel copulas-fitted as models to anthropometric variables in order to capture the structure of their co-dependence. Appropriate fits were obtained and it was noted that Clayton model was best for four of the six pairs of variables examined. The fitted models suggest that the co-dependence structure can be described as general monotonic dependence in all the pairs considered except for the BMI and adiposity pair, which is a monotonic linear dependence. The results of the copula-based Kendall's tau for estimating the measure of co-dependence very strongly suggest the use of BMI as the best substitute for adiposity in body composition assessment. In this scenario, for individuals having extreme values of BMI, their adiposity should be measured directly. Also, in this scenario, the result gives a strong indication for the use of abdominal circumference as a substitute. For percent body fat, there are no strong indications for the use of either BMI or abdominal circumference as a substitute. Hence, for epidemiological research it is better to measure percent body fat directly instead of using a surrogate estimate as suggested in some research based on Pearson correlation coefficient.

It is noticeable that the Pearson's correlation coefficient estimates of the measure of dependence are conspicuously higher than those of the Kendall's tau, where they are not the appropriate measure to quantify the co-dependence structure. That is, in the scenario where the monotonic dependence is not linear, its use in such circumstance could prompt a misleading conclusion. Consequently, in studies where the Pearson's correlation coefficient is used for determining an appropriate surrogate anthropometric variable, it is suggested that a coefficient of determination derivable from the estimated Pearson's correlation coefficient be at least 75% before any recommendation can be advanced. It must be noted that this is a subjective criterion and as such allows the researcher to adjust the bench mark (75%) as appropriate in order to accommodate the peculiarities inherent in such study.

References

- [1] Aas, K.: Modelling the dependence structure of financial assets: A survey of four copulas, *Norwegian Computing Center, Applied Research and Development*, Note no.: SAMBA/22/04 (2004).
- [2] Akindele, M. O., Phillips, J. S. and Igumbor, E. U.: The relationship between percent body fatage and body mass index in overweight and obese individuals in an urban African setting, *Journal of Public Health in Africa* 2016; volume 7: 515, pp 15-19 (2016) <https://doi.org/10.4081/jphia.2016.515>
- [3] Arbel, Y., Birati, E. Y., Shapira, I., Finn, T., Berliner, S. & Rogowski, O.: Comparison of Different Anthropometric Measurements and Inflammatory Biomarkers. *International Journal of Inflammation*, vol 2012, article ID 124639, pp 1-5 (2012) doi: 10.1155/2012/124639.
- [4] Assessing Central Obesity: Waist Circumference (2009) <https://myvmc.com/investigations/assessing-central-obesity-waist-circumference>, accessed January 25, 2020.
- [5] Balkau, B., Deanfield, J. E., Després, J-P., Bassand, J-P., Fox, K. A. A., Smith, S. C., Barter, P., Tan, C-E., Gaal, L. V., Wittchen, H-U., Massien, C., Haffner, S. M.: A Study of Waist Circumference, Cardiovascular Disease, and Diabetes Mellitus in 168 000 Primary Care Patients in 63 Countries, *Circulation*. 2007; 116: 1942-1951. DOI: 10.1161/CIRCULATIONAHA.106.676379, available at <http://circ.ahajournals.org>
- [6] Bergman, R. N., Stefanouski, D. & Buchanan, T. A.: A better index of body adiposity. *Obesity*, vol 19, no. 5, pp 1083-1089 (2011).
- [7] Brannsether, B., Eide, G. E., Roelants, M., Bjerknes, R. & Juliusson, P. B., Interrelationships between anthropometric variables and overweight in childhood and adolescence. *American Journal of Human Biology* 26: 502–510 (2014).
- [8] Carnicero, J. A., Ausin, M. C., & Wiper, M. P.: Non-parametric copulas for circular-linear and circular-circular data: An application to wind directions. *Stochastic Environmental Research and Risk Assessment*, 27, pp. 1991–2002 (2013).
- [9] Cordeiro, G. M. & de Castro, M.: A new family of generalized distributions, *Journal of statistical computation and simulation*. 81, 883-898 (2011) <http://dx.doi.org/10.1080/00949650903530745>
- [10] de Michele, C., Salvadori, G., Vezzoli, R. & Pecora, S.: Multivariate assessment of droughts: Frequency analysis and dynamic return period. *Water Resources Research*, 49, pp. 6985–6994 (2013).
- [11] Dobra, A. & Lenkoski, A.: Copula Gaussian graphical models and their application to modeling functional disability data. *Annals of Applied Statistics*, 5, pp. 969–993 (2011).
- [12] Dombi, J., J'on'as, T. & T'oth, Z., E.: The Epsilon Probability Distribution and its Application in Reliability Theory. *Acta Polytechnica Hungarica*, Vol. 15, No. 1, pp 197-216 (2018).
- [13] Embrechts, P., McNeil A. & Strauman, D.: Correlation and Dependency in Risk Management, Preprint ETH Zurich (1999).
- [14] Fermanian, J-D.: Recent Developments in Copula Models, *Econometrics*, 5, 34 (2017) <https://doi.org/10.3390/econometrics5030034>
- [15] Genest, C., Rémillard, B. & Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44, 199-213 (2009)
- [16] Gongsin, I. E. & Saporu F. W. O.: On the Construction of Kumaraswamy-Epsilon Distribution with Applications, *International Journal of Science and Research*, Volume 8 Issue 11, pp 1199-1204 (2019).
- [17] Hofert, M. & Scherer, M.: Collateralized Debt Obligation pricing with nested Archimedean copulas. *Quantitative Finance* (2011).
- [18] Joe, H.: Multivariate models and dependence concepts, *Monographs on Statistics and Applied Probability*, 73, Chapman & Hall, London (1997).
- [19] Joe, H.: Asymptotic efficiency of the two-stage estimation method for copula-based models, *Journal of Multivariate Analysis*, 94, 401–419 (2005).
- [20] Joe, H. & Xu, J. J.: The estimation method of inference functions for the margins for multivariate models, Department of Statistics, University of British Columbia, *Technical Report*, 166 (1996).
- [21] Jung, Y. S., Kim, J. K. & KIM, J.: New approach of directional dependence in exchange markets using generalized FGM copula function. *Communications in Statistics-Simulation and Computation*, 37, pp. 772–788 (2008).
- [22] Kim, J. M., Jung, Y. S. & Soderberg, T.: Directional dependence of genes using survival truncated FGM type modification copulas. *Communications in Statistics-Simulation and Computation*, 38, pp. 1470–1484 (2009)
- [23] Kojadinovic, I. & Yan, J.: Modeling Multivariate Distributions with Continuous Margins using the copula R Package. *Journal of Statistical Software*, vol. 34, Issue 9, pp 1-20 (2010) <http://www.jstatsoft.org/>
- [24] Leong, I-F, Fang, J-J. & Tsai, M-J.: A feature-based anthropometry for garment industry, *International Journal of Clothing Science and Technology*, Vol. 25 No. 1, pp. 6-23 (2013) DOI: 10.1108/09556221311292183.
- [25] Li, D. X.: On Default Correlation: A Copula Function Approach, *Journal of Fixed Income*, 9: 43-54 (2000).
- [26] Lichtash CT, Cui J, Guo X, Chen Y-DI, Hsueh WA, et al.: Body Adiposity Index versus Body Mass Index and Other Anthropometric Traits as Correlates of Cardio-metabolic Risk Factors. *PLoS ONE* 8 (6): e65954 (2013) <https://doi.org/10.1371/journal.pone.0065954>
- [27] Lojowska, A., Kurowicka, D., Papaefthymiou, G., & Van Der Sluis, L.: Stochastic modeling of power demand due to EVs using copula. *IEEE Transactions on Power Systems*, 27, pp. 1960–1968 (2012).
- [28] Louzada, F., Suzuki, A. K., Cancho, V. G., Prince, F. L. & Pereira, G. A.: The long-term bivariate survival FGM copula model: An application to a Brazilian HIV data. *Journal of Data Science*, 10, pp. 511–535 (2012).

- [29] Luke, A., Durazo-Arvizu, R., Rotimi, C., Prewitt, T. E., Forrester, T., Wilks, R., Ogunbiyi, O. J., Schoeller, D. A., McGee, D. & Cooper, R. S.: Relation between Body Mass Index and Body fat in Black Population Samples from Nigeria, Jamaica, and the United States, *American Journal of Epidemiology*, Vol. 145, No 7, pp 620-628 (1997).
- [30] Nadarajah, S., Afuecheta, E. & Chan, S.: A Compendium of Copulas. *STATISTICA*, anno LXXVII, n. 4 (2017).
- [31] Nelsen, R. B.: *An Introduction to Copulas*. Springer Verlag, New York (2006).
- [32] Pettere, G. & Kollo, T.: Risk modeling for future cash flow using skew t -copula. *Communications in Statistics - Theory and Methods*, 40, pp. 2919–2925 (2011).
- [33] Raimi, T. H. & Oluwayemi, I. O.: Anthropometric Correlates and Prediction of Body fat Measured by Bioelectric Impedance Analysis among Women. *Annals of Medical and Health Sciences Research* (7), pp 32-35 (2017).
- [34] Reddy, M. J. & Ganguli, P.: Risk Assessment of Hydroclimatic Variability on Groundwater Levels in the Manjara Basin Aquifer in India Using Archimedean Copulas. *Journal of Hydrologic Engineering* Vol. 17, 12, 1345-1357 (2012) [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000564](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000564)
- [35] Salvadori, G. & de Michele, C.: Multivariate real-time assessment of droughts via copula-based multi-site Hazard Trajectories and Fans. *Journal of Hydrology*, 526, pp. 101–115 (2015).
- [36] Sklar, A.: Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229-231 (1959).
- [37] Söğüt, M., Altunsoy, K., Varela-Silva, M. I.: Associations between anthropometric indicators of adiposity and percent body fatage in normal weight young adults, *Anthropological Review*, Vol. 81 (2), 174–181 (2018). Available online at: www.degruyter.com/view/j/anre/
- [38] Sun, W., Rachev, S., Stoyanov, S. V. & Fabozzi, F. J.: Multivariate skewed Student's t copula in the analysis of nonlinear and asymmetric dependence in the German equity market. *Studies in Nonlinear Dynamics and Econometrics*, 12 (2008)
- [39] US Department of Health and Human Services, National Institute of Health. WIN Weight Control Network: Understanding Adult Obesity. NIH Publication No. 01-3680. <https://www.win.niddk.nih.gov/publications/PDFs/adultobesbw1201.pdf>
- [40] Valdez, E. & Frees, E.: Understanding Relationships Using Copulas, *Actuarial Research Clearing House*, Proceedings, 32nd Actuarial Research Conference, August 6-8, 1997, 5 (1998).
- [41] Vinué, G.: Anthropometry: An R Package for Analysis of Anthropometric Data, *Journal of Statistical Software*, Volume 77, Issue 6, pp 1-39 (2017) <http://www.jstatsoft.org>
- [42] Xie, K., Li, Y. & Li, W.: Modeling wind speed dependence in system reliability assessment using copulas. *IET Renewable Power Generation*, 6, pp. 392–399 (2012).
- [43] Yu, J., Chen, K., Mori, J. & Rashid, M. M.: A Gaussian mixture copula model based localized Gaussian process regression approach for long-term wind speed prediction. *Energy*, 61, pp. 673–686 (2013).
- [44] Zeng, Q., Dong, S_Y., Sun, X_N, Xie, J. & Cui, Y.: Percent body fat is better predictor of cardiovascular risk factors than body mass index, *Brazilian Journal of Medical and Biological Research*. Vol. 45 (7), pp 591-600 (2012) Doi: 10.1590/S0100-879X2012007500059, www.bjournal.com.br