2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Estimation of Copula Density Using the Wavelet Transform

Fatimah Hashim Falhi [1,2*], ***Munaf Yousif Hmood*** [2]

[1]Department of Statistics , College of Administration and Economics , University of Basrah , Basrah, Iraq.
[2]Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq.
*Corresponding Author.

Received 25/09/2023, Revised 17/01/2024, Accepted 19/01/2024, Published Online First 20/04/2024, Published 01/11/2024

Abstract

This paper proposes a new method to estimate the copula density function using wavelet decomposition as a nonparametric method, to obtain more accurate results and address the issue of boundary effects that nonparametric estimation methods suffer from. The wavelet method is an automatic method for dealing with boundary effects because it does not take into Consideration whether the time series is stationary or nonstationary. To estimate the copula density function, simulation was used to generate data using five different copula functions, such as Gaussian, Frank, Tawn, Rotation Tawn, and Joe copulas. With five different sample sizes at three positive correlation levels based on multiresolution. The results showed that in estimating the copula density function using the wavelet method when the correlation level $\tau = 0.7$, the Gaussian copula ranked first, followed by the Frank copula, and the Joe copula ranked last. In the case of medium and weak correlation, the Tawn copula was in first place, followed by the Rotation Tawn copula, while Gaussian copula came in last place depending on the measures (Root Mean Square Error, Akiake Information Criteria, and Logarithm likelihood criteria). The real copula functions are shown through drawing (Contour plot) and (3D plot). In addition to the smoothing shapes for each of them using the wavelet method, it is clear from the circular shapes that the distribution of observations of the copula function estimated with the wavelet method was accurate at the edges, while it was less accurate at the center for Gaussian and Tawn functions.

Keywords: Boundary Effects, Copula Function, Dependency, Multiresolution analysis, Wavelets.

Introduction

The concept of dependency is an important tool in modeling joint distributions between variables. Nevertheless, one of the most important problems associated with modeling multivariate functions is the existence of dependencies between observations of the variables of the phenomenon under study. Copula functions are very useful tools for analyzing dependencies between random variables. For the strong boundary effects, the copula density estimator was used according to Sklar's theorem [1]. Copulas are quickly becoming popular as multivariate data modeling tools.

Many studies have been published by researchers to help develop ideas for modeling dependency measures in many fields, especially the challenges encountered during the analysis, such as problems of association between study variables and problems of

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

boundary effects. Therefore, attention was paid to background reviews and literature reviews to help prepare the research and the conclusions reached by the researchers who played a role in enriching the research topic.

There are numerous researchers who use copula in insurance and risk management [2,3]. Copula functions can be a useful tool for analyzing the relationships between random variables [4,5].

Several methods have been proposed to handle copula estimation. Nonparametric tests of independence for many bivariate and multiple variables by highlighting the empirical size and properties of the power in many previous small samples based on copulas, and through the results of the test in small variables, he demonstrated that there are nonparametric dependence structures between the variables of the phenomenon studied [6]. Mohammadi et al. proposed two semiparametric methods to estimate the copula parameter [7]. This method is based on the minimum alpha divergence between the nonparametric estimate of copula density using the local probit transformation method and the true copula density function by relying on simulation experiments to measure the performance of these methods based on the Hellinger distance and Niemann divergence. The results show that the method based on Hellinger distance estimation has good performance in small sample sizes and weak dependency cases. It is then demonstrated by applying parameter estimation methods to a real data set in hydrology. Hmood and Hamza presented four nonparametric methods to estimate the copula density based on the kernel density function after applying simulation experiments on samples of different sizes at two levels of high and low reliability for four types of copula [8]. Comparisons between methods were performed using the integrated mean squared error. Simulation results show that the kernel transformation estimation method is the best among the methods used, and the copula is found to be a very flexible model, especially for high Gaussian dependencies. In addition, many researchers have studied wavelets. Jawad and Abdullah studied the wavelet properties of a series of sunspots [9]. A continuous wave analysis of the series was performed. To increase accuracy, the series was divided into its approximate and detailed coefficients using fixed and non-fixed thresholds. They explained that there is an irregularity in the wavelength and intensity. Genest et al., built a rank-based copula density estimator using the wavelet analysis [10]. This approach can be easily implemented using an off-the-shelf wavelet package that automatically handles boundary effects. They showed that this estimation is optimal for a class of uniform copula densities by applying it to actuarial and financial data.

Ghanbari et al., used wavelet analysis to estimate the copula function for censored data. It has been shown by the correct control model that wavelet-based linear function estimators have accurate convergence rates to the mean integral square error (MISE) [11]. Mohammed used the linear wavelet method to estimate the risk function in a nonparametric method [12]. He adopted the simulation method for two types of bivariate distributions and compared these two types using the mean square error. AlDoori and Mhomod employed variable kernel functions to estimate the risk for censored data[13]. Ahmed et al,proposed a wavelet function by deriving the quotient from two different Fibonacci coefficient polynomials, in addition to comparing ARIMA and wavelet ARIMA. This study uses data that relies on daily wind speed time series data. The obtained results show that the proposed wavelet is the most suitable wavelet for wind speed prediction [14]. Shihab et al, Introduced the new form of polynomials, the orthogonal Boubaker polynomial's useful properties, then defined the Boubaker wavelet depending on the orthogonal Boubaker polynomials. This Boubaker wavelet is utilised along with a collocation method to obtain an approximate numerical solution of the singular linear type of Lane-Emden equations [15].

The purpose of this study is to employ wavelets to estimate copula functions through the use of multiresolution analysis. To remove the boundary effects in nonparametric estimation methods, wavelet analysis, and copula modeling are combined. The process involves dividing the generated data into detailed coefficients and approximate coefficients. Additionally, various correlation levels are considered, along with the utilization of both symmetric and asymmetric copula functions. The subsequent sections of this work are structured as follows: First, a comprehensive definition of copula is presented; second, the concept of wavelet is defined; third, the wavelet-copula estimating technique is introduced; fourth, several performance criteria are discussed; and finally, a simulation study is conducted to highlight the effectiveness of the estimator.

Materials and Methods

Copula Function

Mathematically, the copula is defined as a tool used to represent the relational structure between two or more random variables [3].

Therefore, all multivariate CDFs with uniform marginal distributions exhibiting the dependence structure of the random variables X and Y and their marginal CDFs are written as:

$$U = F_X(X) \text{ and } V = F_Y(Y) \qquad 1$$

where U and V are variables with uniform distribution variables; $(U,V) \in [0,1]$. The probability of these variables, $X \le x$ and $Y \le y$, is defined by the joint CDF $F_{XY}(X,Y) = P(X \le x, Y \le y)$.

$$C(u,v) = \Pr(U \le u, V \le v) \qquad 2$$

where $C(u,v)$ is known as a copula and may be uniquely identified when u and v are continuous. It is believed that copulas are of interest to statisticians for two reasons:

- Check if the measure of dependence is scale-free.
- Construct a family of bivariate distributions.

One advantage of this technique is that the copula C, which specifies the dependency between X and Y, can be chosen independently from the marginal models.

Then, the copula density can be expressed as:

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} \qquad u,v \in [0,1] \qquad 3$$

Where $u = F(x)$ and $v = F(y)$

This density exists and is integrable by the unit square.

The following is the formula for a Gaussian copula [4]:

$$C_\theta^{Ga}(u,v) = \frac{1}{2\pi\sqrt{1-\theta^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp[-\frac{u^2 - 2\theta uv + v^2}{2(1-\theta^2)}] dudv \,; where\ \theta\ is\ a\ parameter\ copula$$

Φ Represents the standard normal distribution function while Φ^{-1} represents the inverse of the standard normal distribution function.

A Frank copula is given by formula [3]

$$C(u,v) = \frac{1}{\theta} log\left(1 + \frac{(e^{\theta u} - 1)(e^{\theta v} - 1)}{e^\theta - 1}\right), \theta \in (-\infty, +\infty)$$

Joe copula is provided by [2]

$C_\alpha(u,v) = 1 - [(1-u)^\alpha + (1-v)^\alpha - (1-u)^\alpha(1-v)^\alpha]^{\frac{1}{\alpha}}$ as well as its density

$$c_\alpha(u,v) = [w^\alpha + z^\alpha - wz^\alpha]^{\frac{1}{\alpha}-2} wz^{\alpha-1}[\alpha - 1 + w^\alpha + z^\alpha - wz^\alpha], \alpha \in [1,\infty)$$

Where $w = 1-u\ and\ z = 1-v$. It is distinguished by upper tail dependency. moreover, $\lambda_U = 2 - 2^{\frac{1}{\alpha}}$.

Tawn copula is

$$C = exp\left\{(log(u) + log(v))A\left(\frac{log(v)}{log(uv)}\right)\right\}, where$$

$$A(x) = (1-\alpha_1)x + (1-\alpha_2)(1-x) + \left(\left(\alpha_1(1-x)\right)^\theta + (\alpha_2 x)^\theta\right)^{\frac{1}{\theta}}$$

$and\ (\theta, \alpha_1, \alpha_2) \in (1,\infty) \times [0,1]^2$ $, for\ \alpha_1 = \alpha_2 = 1$, recover the Gumbel copula.

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$At\ any\ time\ \alpha_1 \neq \alpha_2$ it will be asymmetric in its components [4].

Rotation copulas:

Because of their limited parameter space, some of the chosen copula models only allow for positive interdependence, while others only allow for upper or lower tail dependence. To make up for this constraint as previously demonstrated, some copulas do not have complete coverage. Clayton copula, for example, can only capture Kendall's τ between 0 and 1. If Kendall's τ is found to be negative in early analysis, copulas like Clayton will be useless. Copula rotations can correct this. This can be corrected by copula rotation. Displays the copula rotations at 90, 180, and 270 degrees. They are given as

$$C_{90} = v - C(1 - u, v)$$

$$C_{180} = u + v - 1 + C(1 - u, 1 - v)$$

$$C_{270} = u - C(u, 1 - v)$$

where C represents the unrotated copula and u, v represents the margins.

Wavelet

Wavelets are an extension of Fourier analysis in that both seek to express complex functions using the sum of simple ones. Wavelet theory, on the other hand, came considerably later than Fourier analysis [16,17].

Wavelets have accomplished impressive acceptance in earth sciences [18,19]. Wavelets have been used successfully in a variety of applications, including numerical analysis, engineering, signal and image processing, statistics, and geophysics. Using the mathematical construction of a wavelet discrete transform, first provide details of the space $L^2(R)$ in terms of multi-resolution analysis.

Multiresolution is a method for describing the building of spaces and providing an analytical explanation of the components and bases of these spaces. Let us first construct the square-integrable function, often known as the space of Lebesgue measurable functions, which is written as $L^2(R)$ and defined as [20] $L^2(R) = \{f: R \to R; \int_{-\infty}^{\infty} |f(x)|^2 < \infty\}$ [21].

A wavelet is a mathematical function tool used to divide a given function into compounds of different frequencies and explore each configuration using the appropriate solution for each measurement. These tiny waves display information and data in the time and frequency domains. The continuity of their signal is limited by two variables: Unlike the sine function, which extends between $(-\infty, \infty)$, the wavelet function is irregular and asymmetric. A wavelet is defined mathematically as a real value function on the real axis that fluctuates up and down consistently around zero. In other words, it is defined as a signal of limited time length (continuity) with an average value of zero [12, 22]. The wavelet transform is based on the pressure of the wavelet to be processed with two functions: the first is the mother wavelet function $\Psi(x)$ to obtain a set of coefficients characterized by the wavelet coefficients or detailed coefficients D(s,t), and the second is the scaling function $\emptyset(x)$, also called the father's function, to obtain the approximate coefficients A(s,t) [23].

The wavelet is then used to approximate the signal and find a group of wavelet subgroups that are constructed from expanding or compressing and shifting the original wavelet and represent the signal or data that you want to analyze. In other words, the process is the transformation of large-scale measurements into precise measurements by aggregating these data or signals. The main result of the transformation process is the mother wavelet function defined as [24]:

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right) \quad a, b \in R, a \neq 0 \qquad 4$$

Where a and b are dilation and translation parameters,

Ψ refers to the mother wavelet.

$\Psi_{a,b}$ refers to the daughter wavelet.

There are two types of wavelet transforms: continuous wavelet transforms, and discrete wavelet transforms.

Continuous wavelet transform:

The mathematical principle underlying the Continuous Wavelet Transform (CWT) involves partitioning a continuous function over a specified time interval into a collection of wavelets. This transformation has the potential to provide a unified signal representation that encompasses both the time and frequency domains, providing a comprehensive view of the data across these two dimensions. The mathematical expression representing this change may be described as [12,22]:

$$CWT(\tau, s) = \int x(t) \frac{1}{\sqrt{|s|}} \psi\left(\frac{t-\tau}{s}\right) dt$$

Where τ represents the displacement parameter (or withdrawal) for locating the wavelet in the time domain, and s is the scaling factor.

Discrete Wavelet transform:

The discrete wavelet transform (DWT) is well recognized as a prominent and extensively utilized method for wavelet transformation in many domains such as engineering, mathematics, statistics, and other practical applications. The input and output of this transformation consist of discrete data and imitate the discrete Fourier transformation procedurally. The data undergoes a transformation from the time range, namely the original data field, to the wavelet domain. This transformation yields vector-shaped results that maintain the same size as the original vector. The Discrete Wavelet Transform (DWT) can be mathematically represented using linear equations as well as matrix operations[13]. May be expressing this mathematically by the two equations [23]:

$$D(s,t) = \int_{-\infty}^{\infty} f(X)\, \psi_{s,t}(X) dx$$

$$A(s,t) = \int_{-\infty}^{\infty} f(X) \emptyset_{s,t}(X) dX$$

Where s is the scaling or gradient variable, and t is the transform variable.

To approximate the probability density function, the probability density function is decomposed into a set of infinite functions (daughter wavelets) in the time domain on an orthonormal basis by a scaling function (father wavelet) and a wavelet function (mother wavelet) [10].

The approximation is defined as:

$$\varphi_{jk}(t) = 2^{j/2} \varphi(2^j t - k) \qquad 5$$

and

$$\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k) \qquad 6$$

This study uses mother and father Daubechies wavelets. [25].

Wavelet – Copula

In this section, it will be referred to as "wavelet-copula," and the procedure can be easily performed in two steps:

The first step involves using wavelet analysis to decompose variables. The second step uses the decomposed input variables to estimate the copula density function. Since modeling dependence by copula is sensitive to the marginal model, a major innovation of the procedure is the combination of wavelet analysis with copula models.

Wavelet – Copula Estimation

Copula density estimates are constructed using wavelet analysis. This process is easy to implement using out-of-the-box wavelet tools and is based on algorithms that automatically deal with boundary effects. Pseudo-samples $(R_i/n, S_i/n)$, measured on arbitrary divisions of the unit square, is a more promising approach [10].

Wavelet-based estimation of copula density helps explain the underlying dependence structure. In general, the wavelet analysis of the second-order function is $h(x, y)$, which allows you to analyze this mapping infinitely simultaneously with a number of resolution levels $j = 0, 1, 2, \ldots$

The decomposition at any level $j_0 \in N$ is given by

$$h(x,y) = h_{j0}(x,y) + D_{j0}\, h(x,y), x, y \in R \qquad 7$$

so that

$$h_{j0}(x,y) = \sum_{k \in z^2} \alpha_{j0k} \varphi_{j0k}(x,y) \qquad 8$$

is a trend (or approximation) and

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$$D_{j0}h(x,y) = \sum_{j=j0}^{\infty}\left(\sum_{k\in z^2}\beta_{jk}^{(1)}\psi_{jk}^{(1)}(x,y) + \sum_{k\in z^2}\beta_{jk}^{(2)}\psi_{jk}^{(2)}(x,y) + \sum_{k\in z^2}\beta_{jk}^{(3)}\psi_{jk}^{(3)}(x,y)\right) \quad 9$$

is a collection of three sorts of details: vertical edges, horizontal edges, and oblique (corner of the square). In this form, the coefficients α_{j_0k} and $\beta_{jk}^{(1)}$, $\beta_{jk}^{(2)}$ and $\beta_{jk}^{(3)}$ with $j \geq j_0$ are unique for each choice of $j_0 \in N$. For all $j \in N$ and $= (k_1,k_2) \in z^2$, the functions φ_{j0k} and $\psi_{jk}^{(1)}$, $\psi_{jk}^{(2)}$ and $\psi_{jk}^{(3)}$ are defined as follows:

$$\begin{aligned} \varphi_{jk_1k_2}(x,y) &= \varphi_{jk_1}(x)\varphi_{jk_2}(y) \\ \psi_{jk_1k_2}^{(1)}(x,y) &= \varphi_{jk_1}(x)\psi_{jk_2}(y) \\ \psi_{jk_1k_2}^{(2)}(x,y) &= \psi_{jk_1}(x)\varphi_{jk_2}(y) \\ \psi_{jk_1k_2}^{(3)}(x,y) &= \psi_{jk_1}(x)\psi_{jk_2}(y) \end{aligned} \quad 10$$

in terms of a certain scaling function, a corresponding wavelet, and their location-scale transformations provided by.

$$\varphi_{jk_3}(t) = 2^{j/2}\varphi(2^j t - k_3) \quad 11$$

and

$$\psi_{jk_3}(t) = 2^{j/2}\psi(2^j t - k_3) \quad 12$$

for any $t \in R$, and $k_3 \in Z$. The functions φ and ψ (the father and mother wavelet functions, respectively) are defined by Many technical limitations that have to be achieved. To ensure that the family of position scales they create constitutes an orthonormal system of L^2, the set of square-integrable functions. The selection of each pair (φ, ψ) yields a separate multi-resolution analysis with the required degree of regularity. This study is assumed to have compact support [0, L], as is the case for the widely utilized, and provides an overview of this viewpoint. A wavelet representation is distinguished by the fact that the trend at a level $j_0 + 1$ is Consistent with the trend at level j_0, highlighted by horizontal, vertical, and diagonal features corresponding to a level j_0 in other words, [25, 26].

$$h_{j_0+1} = h_{j_0} + \left(\sum_{k\in z^2}\beta_{j_0k}^{(1)}\psi_{j_0k}^{(1)} + \sum_{k\in z^2}\beta_{j_0k}^{(2)}\psi_{j_0k}^{(2)} + \sum_{k\in z^2}\beta_{j_0k}^{(3)}\psi_{j_0k}^{(3)}\right) \quad 13$$

The actual copula C was detected with h_5 by setting

$$H(x,y) = C(F(x),G(y)), \quad 14$$

Assume that $(X_1,Y_1),\dots,(X_n,Y_n)$ is a random sample from the unknown distribution H. The empirical is represented by F_n and G_n distributions related to F and G

$$\left(\frac{R_i}{n},\frac{S_i}{n}\right) = \left(F_n(X_i),G_n(Y_i)\right), i = 1,\dots,n. \quad 15$$

Where R_i and S_i are the ranks of X_i and Y_i Respectively.

Let φ and ψ be the corresponding wavelet for a given scaling function. Both functions are considered real-valued and compactly support $[0,L]$ for some $L > 0$. For each $j \in N$, define $\varphi_{jk},\psi_{jk}^{(1)},\psi_{jk}^{(2)}, and\ \psi_{jk}^{(3)}$ as in Eq 13 for each $k = (k_1,k_2) \in z^2$. The set

$$\{\varphi_{j_0k},\psi_{jl}^{(1)},\psi_{jl}^{(2)},\psi_{jl}^{(3)} : j \geq j_0, k \in z^2, l \in Z^2\}$$

is the orthonormal basis of $L(R^2)$ for any arbitrary $j_0 \in N$. Given a copula density c, it may be expanded as Eq 8 with

$$\alpha_{j_0k} = \int_0^1\int_0^1 c(u,v)\,\varphi_{j_0k}(u,v)dvdu, k \in z^2$$
16

According to Eq 15, the change in variables $u = F(x)\ and\ v = G(y)$ yields

$$\alpha_{j_0k} = \int_0^1\int_0^1 \varphi_{j_0k}(F(x),G(y))\,h(x,y)dydx$$

$$= E_h\{\varphi_{j_0k}(F(X),G(Y))\}, \quad 17$$

where E_h is the expectation based on the original observations' common distribution $(X_1,Y_1),\dots,(X_n,Y_n)$.

If $F\ and\ G$ are unknown, a non-parametric is generated by substituting F and G with their empirical distribution function, $F_n\ and\ G_n$.

The estimator is, therefore, rank-based, i.e.

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$$\tilde{\alpha}_{j_0k} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{j_0k}\big(F_n(X_i), G_n(Y_i)\big)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\varphi_{j_0k}(\frac{R_i}{n}, \frac{S_i}{n}) \quad 18$$

A wavelet-based estimate of c is then given by:

$$\tilde{c}_{j_0}(u,v) = \sum_{k\in z^2}\tilde{\alpha}_{j_0k}\ \varphi_{j_0k}(u,v), \qquad u,v \in [0,1] \quad 19$$

where the smoothing index of the technique is denoted by the number j_0. It is worth noting that is not always the copula density, $\tilde{c}_{j_0}$, just as an empirical copula, is not a copula. $\tilde{c}_{j_0}$ In particular, [6] can be negative in the section of the domain so that it cannot be merged into 1. When you want an estimate of the intrinsic copula density, it can be obtained by truncating and normalizing $\tilde{c}_{j_0}$

From a numerical standpoint, it is crucial to notice that the sum over k in (18) is finite since the wavelet is supported by compact support. Consequently, in reality. Only $[L^2(R)]$ c terms must be computed in the special situation when the copula density must be estimated at a single point $(u_0, v_0) \in [0,1]^2$. For these reasons, the procedure's performance is determined by the level of j_0 selected. The latter should be determined using the most efficient method possible [10].

Performance Criteria:

The comparison between the functions is carried out according to the Root Mean Square Error (RMSE) and is done by calculating the mean square error of the copula function estimated for each iteration according to the following formula:

$$MSE(\tilde{c}_{j_0}, c) = \int_0^1\int_0^1\{\tilde{c}_{j_0}(u,v) - c(u,v)\}^2\, dvdu \quad 20$$

$$RMSE(\hat{c}, c) = \sqrt{MSE(\tilde{c}_{j_0}, c)} \quad 21$$

And the Akaike criterion (AIC) is:

$$AIC_n^{(.)} := -2\sum_{i=1}^{n}\ln\left(c_{\theta_n}^{(.)}\left(u_1^{(i)}, \ldots, u_1^{(i)}\right)\right) + 2p, \quad 22$$

Where p is the number of parameters of the family and θ_n is a parameter estimate.The logarithm of maximum likelihood possibility (LOG L).

$$L(\theta;\, u_1, \ldots, u_n) = \prod_{i=1}^{n} c_\theta(u_i) \quad and\ l(\theta; u_1, \ldots, u_n) = \sum_{i=1}^{n} l(\theta; u_i) \quad 23$$

Respectively, where.

$$l(\theta; u_i) = lnc_\theta(u_i)$$

The best method is the one that minimizes root mean square error and minimizes information criterion. Both criteria select the model that gives the highest likelihood.

Discussion and Results

The simulation scheme consists of many steps depending on the R programming package.

It can be described as the estimation algorithm for d=2 simplicity. So, they are given a sequence $\{(X_i, Y_i)\}_{1\le i\le n}$ of n samples, The estimator proposed method in this paper can be summarized in a number of the following steps.

- simulate five different random samples (n= 32,64,128,256, 512) with replication (r=1000).
- Generate X, Y variables from a uniform distribution.
- Rank the X_i, Y_i with

$$R_i = \sum_{i=1}^{n} 1_{X_l < X_i}\ and\ s_i = \sum_{i=1}^{n} 1_{Y_l < Y_i}$$

- Compute the empirical distribution function

$$u = F_n(X_i) = \frac{R_i}{n}\ and\ v = F_n(Y_i) = \frac{S_i}{n}$$

- Compute the empirical scaling factor

$$\hat{\alpha}_{j_0k} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{j_0k}(\frac{R_i}{n}, \frac{S_i}{n})$$

- Compute the empirical wavelet coefficients

$$\hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\psi_{jk}\left(\frac{R_i}{n}, \frac{S_i}{n}\right)$$

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

- Compute the maximum scale index $J_n = \left\lfloor \frac{1}{2}\, log2^{\left(\frac{n}{\log n}\right)} \right\rfloor$, J is integer number represented the finest resolution from a sample of size n.
- Construct the estimated copula density c by the formula

$$\tilde{c}_{j_0}(u,v) = \sum_{k \in z^2} \hat{\alpha}_{j_0 k} \varphi_{j_0 k}(u,v), \quad where\ u,v \in [0,1]$$

At resolution level j=J,J-1,…,0

- Determine the number of vanishing moments at 4 degrees.
- From all steps, it can be estimated the performance depends on the choice of level j_0 , the latter should be determined most optimally.
- The results showed that in estimating the copula density function using the wavelet method when the correlation level τ = 0.7, the Gaussian copula ranked first, followed by the Frank copula, and the Joe copula ranked last. In the case of medium and weak correlation, the Tawn copula was in first place, followed by the Rotation Tawn copula, while Gaussian copula came in last place depending on the measures (Root Mean Square Error, Akiake Information Criteria, and Logarithm likelihood criteria).

Consider five copula functions as dependency structures (Gaussian, Frank, Tawn, Rotation Tawn, and Joe), with Kendall's tau $\tau =$ 0.7,0.5,0.3. as shown in Tables 1-3.

From Table 1, it appears that when the level of correlation is high, Root of the mean square error and for all copula functions decreases. Likewise, the value of the Akaike coefficient is as small as possible, while the value of the logarithm of the maximum likelihood is as large as possible.

Table 1. Refers to root MSE, AIC, and logarithm likelihood criteria for copula density $\tau = 0.7$

Function	ECDWT Sample size	RMSE	AIC	LOGL
	32	0.18843	-57.4329	29.8099
	64	0.16175	-116.785	59.63232
GAU	128	0.13695	-414.451	208.0832
	256	0.0767	-494.185	248.3973
	512	**0.03925**	**-1314.73**	**658.6745**
	32	0.16953	-73.1424	37.44939
	64	0.16593	-122.257	62.16542
FRANK	128	0.22525	-260.514	131.355
	256	0.05901	-458.975	230.9677
	512	**0.05179**	**-1260.37**	**631.4375**
	32	0.19771	-70.7703	40.02418
	64	0.18102	-83.0877	42.97897
TAWN	128	0.17395	-240.125	127.4758
	256	0.17055	-414.966	215.1595
	512	**0.14552**	**-1000.56**	**501.807**
	32	0.23972	-60.1029	31.20294
	64	0.1673	-92.9648	51.16785
RTAWN	128	0.15094	-282.614	142.5703
	256	0.10993	-514.148	258.4948
	512	**0.0898**	**-850.793**	**427.7506**
	32	0.59031	-60.2817	30.79977
	64	0.42713	-63.8449	32.46232
JOE	128	0.38647	-67.469	36.21485
	256	0.37677	-248.321	126.1468
	512	**0.20222**	**-435.536**	**220.0326**

From Table 2 notes that the value of the root of the mean square error and the Akaike coefficient at the medium correlation level is higher than it was at the high correlation level. For all copula functions used in the study, it is clear that the root of the mean square error is inversely proportional to the level of correlation in the copula functions. In contrast, the value of the logarithm of the maximum likelihood is higher than it was at the high correlation level.

Table 2. Refers to root MSE, AIC, and logarithm likelihood criteria for copula density $\tau = 0.5$

Function	ECDWT Sample size	RMSE	AIC	LOGL
	32	0.55886	-18.9267	11.26035
	64	0.46729	-66.2573	34.80296
GAU	128	0.45667	-113.008	58.77383
	256	0.35513	-216.31	111.0381
	512	**0.25761**	**-358.82**	**182.4047**
	32	0.48065	-25.2846	14.28811
	64	0.44904	-81.8056	42.32155
FRANK	128	0.36845	-145.747	74.70922
	256	0.27853	-249.719	127.3506
	512	**0.24437**	**-644.141**	**324.3424**
	32	0.54244	-8.03816	6.34288
	64	0.49021	-45.6272	25.02905
TAWN	128	0.48712	-121.104	62.55991
	256	0.18847	-195.834	100.3944
	512	**0.15348**	**-986.863**	**494.9124**
	32	0.59311	-38.5824	20.66113
	64	0.44009	-60.1973	31.97001
RTAWN	128	0.3904	-209.497	106.0468
	256	0.36051	-289.044	146.3451
	512	**0.19462**	**-788.288**	**395.9537**
	32	0.53243	-19.6327	11.63855
	64	0.48461	-45.4424	25.01778
	32	0.55886	-18.9267	11.26035
JOE	64	0.46729	-66.2573	34.80296
	128	0.45667	-113.008	58.77383
	256	0.37908	-250.979	127.8033
	512	**0.22915**	**-445.204**	**225.2306**

As for Table 3, when the level of correlation is weak, and for all copula functions and at different sample sizes, the value of the root of the mean square error is greater than it was at the high and medium levels, while the value of the Akaike coefficient is the lowest possible, and the value of the logarithm of the maximum likelihood is the highest possible. When comparing the copula functions according to the three comparison criteria, it appears that the Tawn copula function is the most suitable function for the estimation method using wavelets because the amount of change in the root of the mean square error is very slight. Likewise, the Akaike and logarithm criteria are the greatest possibility of the function, especially when the sample size is large, and this indicates that the wavelet transform method has better performance when the sample size is large.

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 3. Refers to root MSE, AIC, and logarithm likelihood criteria for copula density $\tau = 0.3$

Function	ECDWT Sample size	RMSE	AIC	LOGL
GAU	32	0.71163	-8.79708	6.51427
	64	0.55352	-48.6918	26.3745
	128	0.42452	-109.671	56.80845
	256	0.41359	-271.17	137.6521
	512	**0.34671**	**-694.429**	**349.3544**
FRANK	32	0.7231	-10.2541	7.32794
	64	0.66538	-26.0737	15.78511
	128	0.59807	-65.7989	35.52795
	256	0.44442	-253.352	128.7044
	512	**0.3717**	**-415.229**	**209.8027**
TAWN	32	0.66441	-11.8305	8.07624
	64	0.64268	-21.6373	13.47802
	128	0.60952	-63.3696	33.94843
	256	0.46535	-223.006	113.7815
	512	**0.15607**	**-854.443**	**428.785**
RTAWN	32	0.86981	-12.2269	9.2311
	64	0.73668	-19.8288	11.68532
	128	0.56702	-94.2055	49.31526
	256	0.52549	-404.656	203.9782
	512	**0.20583**	**-435.671**	**221.0578**
JOE	32	0.62913	-12.0969	8.09585
	64	0.49163	-42.0206	23.06409
	128	0.41235	-113.394	58.62762
	256	0.39105	-409.191	206.0382
	512	**0.39068**	**-334.387**	**169.9307**

The contour plot and the 3D plot of the real copula functions (Gaussian, Frank, Tawn, Rotation Tawn, and Joe) are illustrated in the Figures below, in addition to the preface shapes for each of them using the estimation copula density wavelet transform method (ECDWT). From Fig. 1, it can be noted, through the three-dimensional figures, that the distribution of the observations of the copula function estimated by the wavelet transform (ECDWT) method was accurate at the edges while it was less accurate at the center for both functions. It is also evident from the three-dimensional figures that the probability density function of the real (Gaussian) copula function is characterized by a similar concentration of observations at the center and at the edges, with the withdrawal of observations towards the tail and relatively little expansion at the center. Through the three-dimensional figure, the (ECDWT) smoothing of the Gaussian function was more flat at the center and more congruent at the tails (extremities) when compared to the real probability density function.

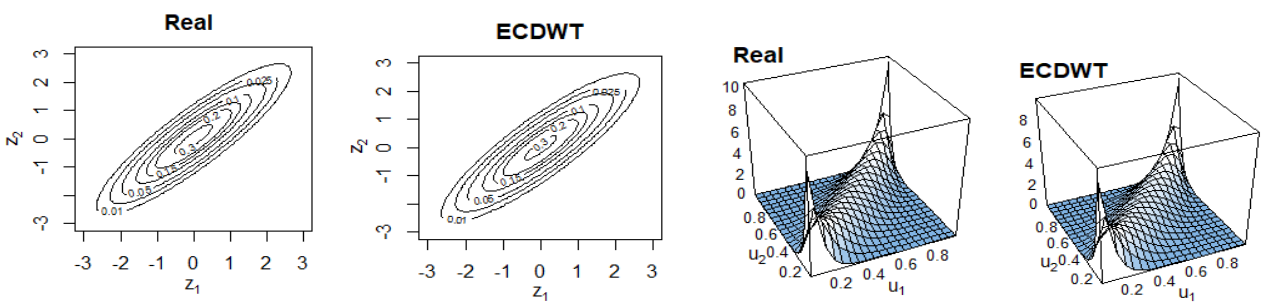

Figure 1. Counter and 3D plot for Gaussian copula density estimation with (ECDWT).

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Besides, Fig. 2 represents the estimates of the Frank function when (n = 128, tau=0.7). It shows that the Frank function is characterized by similar dependency the center and at the edges, noting that the difference in the distribution of observations between the two Gaussian functions and Frank is that the observations at the center in the Frank function are less flat than in the Gaussian function and that the observations with respect to the Frank function are less pulled towards the edges. As for the smoothing of the observations using (ECDWT) method, we noticed fluctuations in the distribution of observations. At the edges, it was better at the center. It can be recognized as the general figure of the densities as well as high- and low-density regions. Quiet, some of the families (e.g., Gaussian and Frank) are very difficult to characterize. Indications of the argued effects can be noted in the simulated sample of the Frank copula see Fig. 2. (ECDWT) tends to overvalue the true density in the corners (0, 0) and (1, 1).

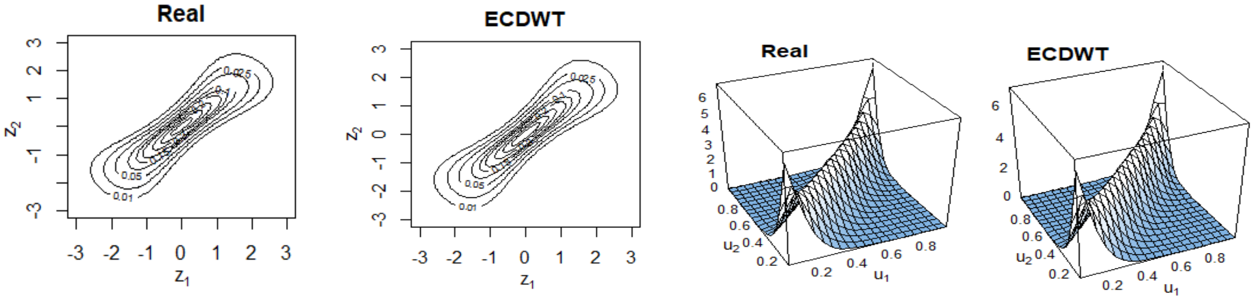

Figure 2. Counter and 3D plot Frank copula density with ECDWT method.

In addition, Fig. 3 represents the assumed and estimated probability density function for the copula (Tawn) at the high level of correlation and the sample size (128), and it is clear that the copula function (Tawn) is characterized by a large concentration of observations at the right side, and the (ECDWT) method remained far from the assumed copula function.

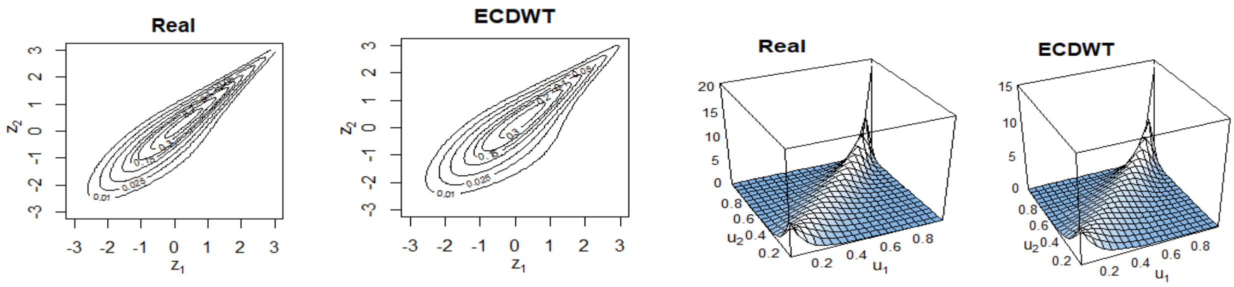

Figure 3. Counter and 3D plot for Tawn copula density with ECDWT method.

Fig. 4 represents the estimates of the copula function (RTawn) at the high level of correlation, and the sample size (128), which was rotated by 90 degrees, assumed the Rtawn copula function. There is a large concentration of observations on the right, but the observations in the middle are not as flat as the Tawn copula). noted that the centering of the observations at the right tail was identical to the real shape, and the centering of the observations at the middle became less fluctuating than it was in the Tawn copula. shown a close estimate of the assumed copula function (Tawn).

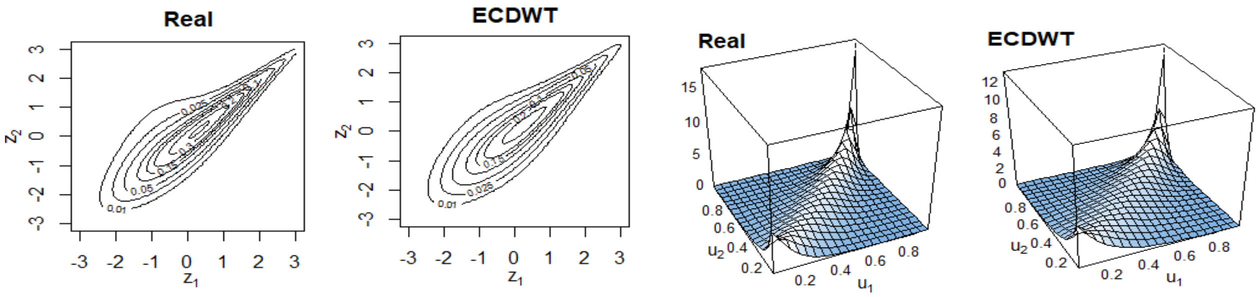

Figure 4. Counter and 3D plot for Rotation Tawn copula density with ECDWT method.

However Fig. 5 represents the probability density function for the (Joe) copula when (tau=0.7, n=128). It is clear from it. The assumed Joe copula function has a right tail, and the concentration of observations was clearly on the left side, while the distribution of observations in the middle appears flat. In the probability density function estimated by the (ECDWT) method, the performance was not good at the center, which was characterized by instability because the observations were too flat or at the right tail, where the concentration of observations was greater, but the concentration of observations at the left end was more similar to the assumed shape of the copula. This is evidence that when estimating the copula function (Frank, Tawn, Rtawn, and Joe), the smoothing of the probability density functions was less flat at the center, but it was more withdrawn towards the tails despite the presence of a great match between the smoothed and the real functions. Additionally, despite having observed that the smoothed and real functions had a significant match, the smoothing of the probability density functions while estimating the copula function (Frank, Tawn, Rtawn, and Joe) was less flat. In general, it can be said that the smoothing when estimating the copula function (Gaussian) is slightly better than the smoothing when estimating the copula functions (Frank, Tawn, Rtawn Joe).

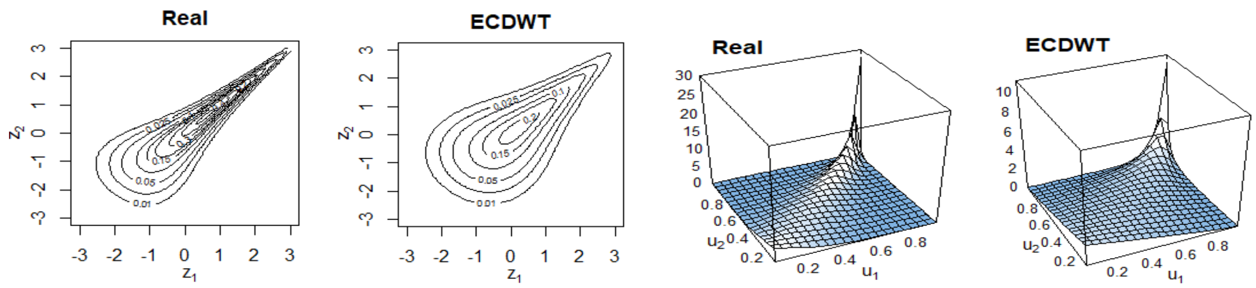

Figure 5. Counter and 3D plot for JOE copula density with ECDWT method.

Conclusion

This paper presents copula estimation based on the wavelet methods, i.e., it is based on Daubechies wavelets from four degrees. The simulation results were reached using five copulas (Gaussian, Frank, Tawn, Rotation Tawn, and Joe) for five sample sizes (n = 32, 64, 128, 256, 512) based on three criteria (RMSE, AIC, and LOGL), represent a statistic for selecting the copula with the best performance when using wavelets to estimate the copula density function when taking high, medium and low correlation levels (τ = 0.7, 0.5, 0.3).

Accordingly, several conclusions can be drawn from the results presented so far.

The present study focused on the estimation and identification of Copula density functions using rank-dependent wavelets. Through simulations, it was shown that the value of the square root of the mean squared error decreases for all relevant functions as the sample size increases. It is also noted that when there is a high level of correlation, the value of the square root of the mean squared error and Akiake coefficient are the smallest possible, and the logarithm of the maximum likelihood is as large as possible.

Wavelet algorithms have efficient computational properties, making them amenable to rapid computation. Additionally, these algorithms possess a straightforward nature that facilitates their ease of updating and adapting to various modeling scenarios. The numerical performance of the recommended linear wavelet density estimator was demonstrated using simulated datasets.

The explanations also included comparisons of complete data and different sample sizes. Nevertheless, wavelet-based copula function estimators fail to satisfy the fundamental requirements of parametric models.

It is clear from the above that when the correlation is medium and low, using wavelet analysis to estimate the copula density function improves the performance of the Tawn copula function. It is clear from the figures that the Tawn copula function is better when using wavelets to estimate the copula density function.

Future studies could include using nonlinear wavelets to estimate copula density and goodness-of-fit testing.

Authors’ Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at University of Basrah.

Authors’ Contribution Statement

F. H. F. contributed to the design, writing the simulation algorithm to generate the data, analysis, the interpretation of the results. M. Y. H. contributed to the conception of the idea of the research. drafting the manuscript, revision, and proofreading.

References

1. Ansari J, Rüschendorf L. Sklar’s theorem, copula products, and ordering results in factor models. Depend Model. 2021 Oct 18; 9(1): 267-306. https://doi.org/10.1515/demo-2021-0113
2. Joe H. Multivariate Models and Multivariate Dependence Concepts. 1st edition. New York: Chapman and Hall/CRC; 1997. 424 p. https://doi.org/10.1201/9780367803896
3. Nelsen RB. An Introduction to Copulas. 2nd edition. New York: Springer; 2006. XIV, 272 p. https://doi.org/10.1007/0-387-28678-0
4. Zhang L, Singh VP. Copulas and their Applications in Water Resources Engineering. Cambridge, UK: Cambridge University Press; 2019. https://doi.org/10.1017/9781108565103
5. Chen L, Guo S. Copulas and Its Application in Hydrology and Water Resources. Singapore: Springer Nature; 2019. X, 290 p. https://doi.org/10.1007/978-981-13-0574-0
6. Herwartz H, Maxand S. Nonparametric tests for independence: a review and comparative simulation study with an application to malnutrition data in India. Stat Pap. 2020 Oct; 61: 2175-201. https://doi.org/10.1007/s00362-018-1026-9
7. Mohammadi M, Amini M, Emadi M. A Simulation Study of Semiparametric Estimation in Copula Models Based on Minimum Alpha-Divergence. Stat Optim Inf Comput. 2020; 8(4): 834-845. https://doi.org/10.48550/arXiv.2009.05247
8. Hmood MY, Hamza ZF. On the Estimation of Nonparametric Copula Density Functions. IJSSST. 2019; 20(2): 1-10. http://doi.org/10.5013/IJSSST.a.20.02.07
9. Jawad LB, Abdullah LT. Wavelet analysis of sunspot series. J Econ Finance Adm Sci. 2007; 13(45): 273-87. https://doi.org/10.33095/jeas.v13i45.1138
10. Genest C, Masiello E, Tribouley K. Estimating Copula Densities Through Wavelets. Insurance: Math Econ.

2009 Apr 1; 44(2): 170-181. https://doi.org/10.1016/j.insmatheco.2008.07.006
11. Ghanbari B, Yarmohammadi M, Hosseinioun N, Shirazi E. Wavelet estimation of copula function based on censored data. J Inequal Appl . 2019; 2019(1): 1-15. https://doi.org/10.1186/s13660-019-2140-5
12. Mohammed AT. Nonparametric estimation of hazard function by using wavelet transformation. phD Thesis. University of Baghdad; 2019.
13. AlDoori EA, Mhomod E. Hazard Rate Estimation Using Varying Kernel Function for Censored Data Type I, Baghdad Sci. J. 2019 Sep. 23; 16(3(Suppl.)): 0793. https://doi.org/10.21123/bsj.2019.16.3(Suppl.).0793
14. Ahmed LA, Mohammed M. A Proposed Wavelet and Forecasting Wind Speed with Application. Ibn al-Haitham J. Pure Appl Sci. 2023; 36(2): 420-9. https://doi.org/10.30526/36.2.3060
15. Ouda EH, Ibraheem SF, Shihab SN. Boubaker Wavelets Functions: Properties and Applications. Baghdad Sci J. 2021; 18(4): 1226-1233. http://dx.doi.org/10.21123/bsj.2021.18.4.1226
16. Abdourahamane ZS, Acar R, Serkan Ş. Wavelet–copula-based mutual information for rainfall forecasting applications. Hydrol Process. 2019; 33(7): 1127-1142. https://doi.org/10.1002/hyp.13391
17. Hmood MY, Hassan YA. Estimate the Partial Linear Model Using Wavelet and Kernel Smoothers. J Econ Finance Adm Sci. 2020; 26(119): 428–443. https://doi.org/10.33095/jeas.v26i119.1892
18. Labat D. Recent advances in wavelet analyses: Part 1. A review of concepts. J Hydrol. 2005 Nov 25; 314(1-4): 275-288. https://doi.org/10.1016/j.jhydrol.2005.04.003
19. Labat D, Ronchail J, Guyot JL. Recent advances in wavelet analyses: Part 2—Amazon, Parana, Orinoco and Congo discharges time scale variability. J Hydrol. 2005 Nov 25; 314(1-4): 289-311. https://doi.org/10.1016/j.jhydrol.2005.04.004
20. Chui CK. An introduction to wavelets. 1st edition. United States: Academic press; 1992 Jan 1. 278 p.
21. Allaoui S, Bouzebda S, Liu J. Multivariate wavelet estimators for weakly dependent processes: strong consistency rate. Commun. Stat. Theory Methods. 2023; 52(23): 8317-8350. https://doi.org/10.1080/03610926.2022.2061715
22. Hmood MY, Hibatallah A. Continuous wavelet estimation for multivariate fractional Brownian motion. Pakistan J Stat Oper Res. 2022 Sep 10; 18(3): 633-41. https://doi.org/10.18187/pjsor.v18i3.3657
23. Rashid DH, Hamza SK. Comparison Some of Methods Wavelet Estimation for Non Parametric Regression Function with Missing Response Variable at Random. J Econ Finance Adm Sci. 2016; 22(90): 382-406. https://doi.org/10.33095/jeas.v22i90.513
24. Kaiser G. A Friendly Guide to Wavelets. 1st edition. USA: Birkhäuser Boston, MA; 2010 Nov 3. XX, 300 p. https://doi.org/10.1007/978-0-8176-8111-1
25. Daubechies I. Ten Lectures on Wavelets. Philadelphia, PA: SIAM; 1992 Jan 1. 369 p. https://doi.org/10.1137/1.9781611970104
26. Meyer Y. Wavelets and Operators: Volume 1. United Kingdom: Cambridge university press; 1993. 239 p. https://doi.org/10.1017/CBO9780511623820

2024, 21(11): 3497-3511
https://doi.org/10.21123/bsj.2024.9673
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

Baghdad Science Journal

تقدير كثافة الرابطة باستعمال التحويل المويجي

فاطمة هاشم فلحي[1]، مناف يوسف حمود[2]

[1]قسم الاحصاء، كلية الإدارة والاقتصاد، جامعة البصرة، بصرة، العراق.
[2]قسم الاحصاء، كلية الإدارة والاقتصاد، جامعة بغداد، بغداد، العراق.

الخلاصة

يقترح هذا البحث طريقة جديدة لتقدير دالة كثافة الرابطة باستخدام تحليل المويجات كطريقة لامعلمية، من أجل الحصول على نتائج أكثر دقة وخالية من مشكلة تاثيرات الحدود التي تعاني منها طرائق التقدير اللامعلمية. اذ تعد طريقة المويجات طريقة اوتماتيكية للتعامل مع تاثيرات الحدود وذلك لانها لا تأخذ بنظر الاعتبار إذا كانت السلسلة الزمنية مستقرة او غير مستقرة. ولتقدير دالة كثافة الرابطة تم استعمال المحاكاة لتوليد البيانات وباستعمال خمسة دوال رابطة مختلفة مثل Gaussian وFrank وTawn وRotation Tawn و Joe وبخمسة أحجام مختلفة للعينات عند ثلاثة مستويات ارتباط موجبة، واعتمادًا على الحلول المتعددة، أظهرت النتائج أن تقدير دالة الكثافة الرابطة بطريقة المويجات عندما يكون مستوى الارتباط $\tau = 0.7$ كانت الرابطة Gaussian في المرتبة الأولى تليها الرابطة Frank واحتلت الرابطة Joe المرتبة الأخيرة. اما في حالة الارتباطات المتوسطة والضعيفة ($\tau = 0.5$,0.3) كانت الرابطة Tawn في المرتبة الأولى تليها الرابطة Rotation Tawn في حين جاءت Gaussian بالمرتبة الأخيرة. بالاعتماد على المعايير (Root Mean Square Error, Akiake Information Criteria, and Logarithm likelihood criteria) ، وتبين من خلال الرسم (Contour plot) والشكل ثلاثي الابعاد (3D plot) لدوال الرابطة الحقيقية. فضلا عن اشكال التمهيد لكل منها باستخدام طريقة (ECDWT)، ويتضح من خلال الاشكال الدائرية ان توزيع مشاهدات الدالة الرابطة المقدرة بطريقة (ECDWT) كان دقيقا عند الاطراف بينما كان اقل دقة عند المركز لكل من الدوال Gaussian وTawn.

الكلمات المفتاحية: تأثيرات الحدود، دالة الرابطة، الاعتمادية، تحليل الحلول المتعددة، المويجات.