ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 10, No. 6, December 2022, pp.96-13

Estimation of nonparametric regression function using shrinkage wavelet and different mother functions

Saad Kadem Hamza[1] , Shreen Ali [1, 2]

[1] Department of Statistics, College of Administration and Economics, University of Baghdad, Iraq
[2]Department of Quality Assurance and University of Performance, University of Baghdad, Iraq

ABSTRACT

Wavelet reduction is one of the most widely used methods for removing noise from the signal, primarily financial and banking data, and building a non-parametric regression model that enables us to study the phenomenon accurately. The appropriate choice of the wavelet mother, which results in concentrating the bulk of the signal strength on a few wavelet coefficients, is one of the most determining factors in noise removal and obtaining accurate regression function estimates. Given the importance of studying the price index in Iraq, many mother functions within the wavelet transformation have been studied. To determine which of them is more suitable for such a type of data, which gives accurate estimates of the relationship between trading volume as an independent variable and the Iraq market index as a dependent variable, the best or most appropriate functions were determined through the estimates that have less (MSE). It became clear that the best or relevant parts are (Coif1, Coif5, and rbio1.3).
The study was applied to real data represented by the trading volume and price index data for the Iraqi market for the period from (2008) to April (2022). It became clear that the trading volume significantly affects the price index, but other variables must be studied.

Keywords: Contractile wavelet, non-parametric regression, mother function, wavelet transform.

Corresponding Author:

Saad Kadem Hamza
Department of Statistics, College of Administration and Economics, University of Baghdad
Baghdad, Iraq
E-mail: Saad.hamza@coadec.uobaghdad.edu.iq

1. Introduction

The wavelet reduction method is one of the essential methods that can be used in many practical applications, primarily economic, financial, and banking, to address the confusion in those data (the signal) and to obtain the most considerable amount of helpful signal that leads to more accurate results that help in making the necessary decisions and building plans. Many researchers dealt with the factors affecting estimation accuracy using wavelet reduction. The mother wave is to know the wave type or the signal's behavior, such as studying GPS signals, the type of heartbeat signal, or some organ movement signals inside the body. The mother function is an essential part of the wavelet transform. Therefore, the wavelet transform will be passed through, through which the signal can be analyzed into a set of multi-precision levels in both frequency and time domains. The most important part of the wavelet transform is the mother wavelet, which varies according to different time frequencies [1].
The wavelet is defined as a signal of limited time length. It has an average value equal to zero, as the detailed parameters D(s,t) are obtained through the mother wavelet. The essential part contains the information to be known as through the wavelet transformation, the amount of information of the level of importance is reduced—the least amount carried by the signal without prejudice to the highest level of important information. Thus the

signal is filtered and filtered in the event of removing the noise. Practically, this is done by reconstructing the signal after sorting out the complex transactions [2, 3, 4].
Several wavelet types can be used within the wavelet transformation of all kinds (continuous, discontinuous). These types are wavelets (Haar, Daubechies, Coiflets, Meyer) and others. The primary purpose of this research is to choose the optimal wavelet function that achieves the best quality for the selected data and thus get good grades from the phenomenon to be studied.

2. Wavelet transformation

The emergence of the wavelet transform was necessary to address the weakness suffered by the previous transformations, such as the Fourier Transformation, sinusoidal transformations, and the Short Fourier Transformation. The fixed window for all types of data spread [2].
The wavelet transform is a mathematical method in which the signal is analyzed using various extended and transposed versions of an essential function called the mother function used within both the continuous wavelet transformation and the discontinuous wavelet transformation. Discrete wavelet transformation will be adopted because of its advantages over the rest of the other transformations [5].

3. Discrete wavelet transformation

The ability of this conversion to achieve accurate positioning in both time and frequency was the main reason that made it superior to the rest of the transformations, as this feature helps in easier processing of the signal by extracting the most significant amount of helpful information without affecting the essential characteristics of the signal [**6**]. Mathematically, the wavelet transform is based on folding the signal to be processed with two functions, the first is called the wavelet function $\psi(t)$ to obtain a set of coefficients, which are called the detailed coefficients D(s,t), and the second is the measurement function (the father function $\phi(t)$ to get the approximate coefficients A(s,t), which is expressed mathematically [7, 8]:

$$D(s,t) = \int_{-\infty}^{+\infty} f(t)\psi_{s,t}(t)d(t) \qquad \text{……….(1)}$$

$$A(s,t) = \int_{-\infty}^{+\infty} f(t)\phi_{s,t}(t)d(t) \qquad \text{……….(2)}$$

Here, s represents the scale, while t represents the transition.
The discontinuous wavelet transformation coefficients denoted by the symbol C(j,k) indicate the correlation between the signal X(n) and the mother wavelet $\psi_j, k(n)$ at the j level and position k. They are given according to the following relationship:

$$C(j,k) = \sum_{n \in z} x(n).\psi_j k(n) \qquad \text{……….(3)}$$

$$\psi_{j.k}(n) = 2^{-j/2}.\psi(2^{-j}n - k) \qquad \text{……….(4)}$$

In practice, the signal with the highest level of important information is obtained through (thresholding) on the detailed coefficients produced by the mother function. Thus, it becomes clear the importance of choosing the wavelet mother function that creates the detailed information through which the most critical data is selected.

4. Mother wavelet

The symbol denotes $\psi(s,t)$ it, and it is the wavelet function that determines the basic wavelet shape, through which the rest of the wavelets are obtained with different accuracy and different frequencies by changing the values of the gradient (s) and the transition (τ), which is expressed mathematically by the following relationship [7]:

$$\psi_{(s,t)}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right), s \in R^{+}, \tau \in R \qquad \text{.........(5)}$$

Where s is the scaling factor and τ time shift or shift factor. Acceptance condition must be based on the following condition must be met in the frequency domain [6, 8].

$$C_{\psi} = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^{2}}{|\omega|} . d\omega < +\infty \qquad \text{.........(6)}$$

Where $\psi(\omega)$ to refer to the transformation of the Fourier function $\psi(t)$.

The number of vanishing moments, which is useful for wavelet pressure, where the wavelet function $\psi(t)$ has p number of vanishing moments if the following is true:

$$\int_{-\infty}^{\infty} t^{k} . \psi(t) dt = 0 \; for 0 \leq k < p \qquad \text{.........(7)}$$

If the wavelet function $\psi(t)$ has more vanishing moments, the wavelet coefficients of the signal x(t) are lower for the largest j scale (more accurate). The total area $|\psi(\omega)|^{2}$ is finite, which indicates that the wavelet energy is limited. The following is achieved:

$$\int_{-\infty}^{\infty} |\psi(t)|^{2} . dt \leq \infty \qquad \text{.........(8)}$$

5. Thresholding rules

Since the noise coefficients are of lower frequency than the frequency of the original signal's coefficients[9], removing noise in the signal is an important step in estimating the regression function using wavelet reduction. The threshold selection also helps preserve the coefficients of the original signal.

There are a few different types of thresholding, the two most prevalent and widely used being soft thresholding and hard thresholding.

5.1. Soft thresholding

It builds upon the prior approach. The difference here is that non-zero items are moved closer to zero after those whose absolute value has been less than a threshold have been zeroed. A mathematical formula for this is [6, 10]:

$$Thr_{\lambda}^{S}(y,\lambda) = \begin{cases} 0 & if \quad |y| \leq \lambda \\ \operatorname{sgn}(y) & if \quad |y-\lambda| > \lambda \end{cases} \qquad \text{.........(9)}$$

The thresholding value is signified by λ.

5.2. Threshold value

Choosing a right threshold value λ is a crucial part of the wavelet reduction technique for lowering signal-to-noise ratio [11] .

5.3. Sure thresholding

Donoho and Johnstone's approach for choosing threshold values (Sure Shrink). (Stein's Unbiased Risk Estimation) stands for an abbreviation for (Stein's Unbiased Risk Estimation) for each wavelet level j [7,12], which is the foundation of this method.

Due to the orthogonality of the wavelet transform, the noise transform is likewise orthogonal, that is, the coefficients $d_{j,k}$ are as well orthogonal, and as the noise is distributed (Gaussian), which means d^{*} which Stein used to demonstrate the unbiasedness of this estimate for risk [13] because it follows a (Gaussian) distribution.

$$SURE(\lambda_j, d_{jk}) = N - 2\sum_{k=1}^{N} I\left(\left|d_{jk}\right| \le \lambda_j \right) + \sum_{k=1}^{N} \min\left(\left|d_{jk}\right| \le \lambda \right)^2 \qquad \ldots\ldots\ldots(10)$$

Consequently, following this approach, we can determine the critical value of (SURE).

$$\lambda_{j,SURE} = \arg\min_{0 \le \lambda \le \sqrt{2\log N}} SURE(\lambda_j, d_{jk}) \ldots\ldots\ldots(11)$$

5.4. Visushrink thresholding

It had been presented by (Donoho) and is expressed by the following equation $\sigma\sqrt{2LogN}$, where σ indicates the number of observations and the variance of the noise; this method's efficiency increases with larger samples since it produces a more consistent estimate and a prelude than do global threshold approaches, which operate only when the sample size is small. Some of its major drawbacks include its inefficiency in dealing with signal-to-noise ratios and correlations, both of which grow worse as the sample size increases[14]. This is because more data leads to the loss of wavelet coefficients with noise. The following formula [15] has been used to enhance this method:

$$\lambda = \sigma_n \sqrt{LogN} - \sigma_s^2$$

Since σ_s^2 is the signal's variance, thus if the noise's variance is smaller than the signal's variance, the noise's contribution to the noisy signal's increased average value disappears, while σ_n is the noise level standard deviation, and it may be calculated as follows [16]:

$$\hat{\sigma}_n = MAD(Y)/0.6745$$

Here, MAD is an absolute deviation of the median for the detailed coefficients.

6. Real data

The data of the study represented both the trading volume and the price index in the Iraqi stock market, which are monthly data for the period from 31/1/2009 to 30/4/2022, as they were obtained from the data of the Central Bank of Iraq, and the importance of the study comes from the fact that the stock market considers women Which reflects the state of the national economy, as the state of the financial market generally refers to the development and the state of production in the economy, where the growth of the financial market is linked to economic, industrial and financial development, and given the mentioned importance of the stock market in economic activity through trading in securities listed in the market, this study is to delve into the stock market to get acquainted with the essential and influential elements in it, which is the volume of trading and the trading of the stock market index in Baghdad, that the volume of trading in recent years in various sectors has been on the rise, as it is an essential factor and a real indicator to push the wheel of the economy forward, being a significant factor upon which most stock speculative decisions are taken. Hence, this header illustrates the importance of the trading volume on the Iraq Stock Exchange Index.

By testing the data using the (SPSS) program, version (24), it became clear that the relationship is not linear between the independent and dependent variables and that the moral model is useless in clarifying the relationship between trading volume (the independent variable) and the stock market index (as a dependent variable). This confirms that testing the relationship between the two variables using the parametric model is a non-significant relationship, as the value of (Sig = 0.151) is more significant than (0.05), which indicates the insignificance of the model. The moral model explains the relationship, and accordingly, the non-parametric model represented by the wavelet regression model, referred to in the theoretical aspect, will be used_.Table 1 depicts the MASE criterion comparing the estimates of the confusing ISM index for different mother functions and sample sizes n=128 and n=64.

Table 1. The MASE criterion comparing the estimates of the confusing ISM index for different mother functions and sample sizes n=128 and n=64.

	Dub 2	Dub10	Dub20	Dub30	Symlet2	Symlet4
64	0.231717566	0.147222067	0.186449861	0.212261112	0.229422102	0.23065595
128	0.232037385	0.147606712	0.186994343	0.212861140	0.231245645	0.231654249
	Symlet7	Coif1	Coif3	Coif5	rbio1.3	rbio5.5
64	0.227934535	0.003067601	0.068045951	0.010437419	0.063558631	0.084157162
128	0.227673513	0.003243210	0.068234760	0.008696694	0.024952621	0.074379690

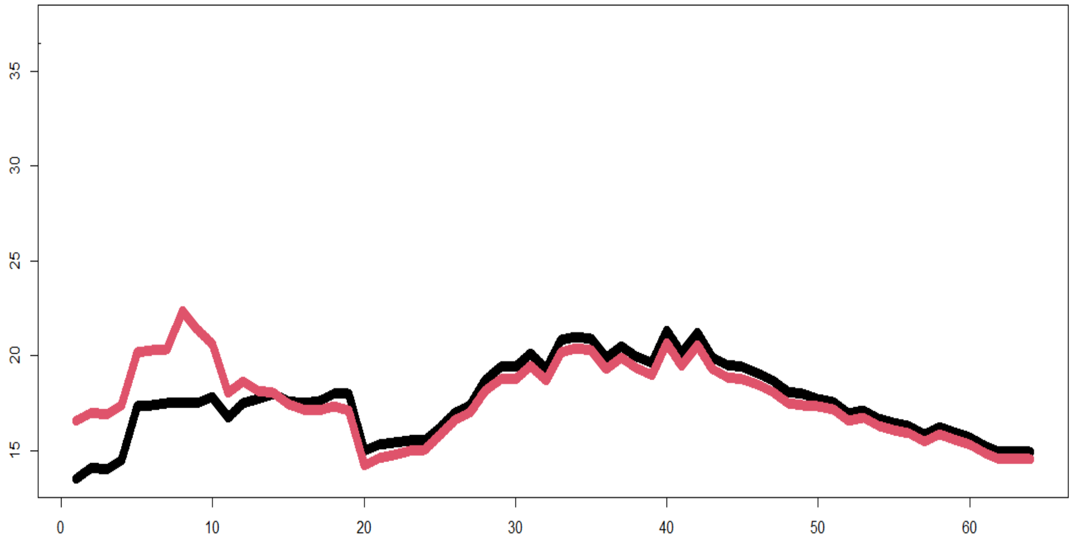

Figure 1. A real and estimated data for a dependent variable y (the Iraq market index) for the estimation methods used using the coif1 function, as the curve in black indicates the real values. In contrast, the red color indicates the estimated values

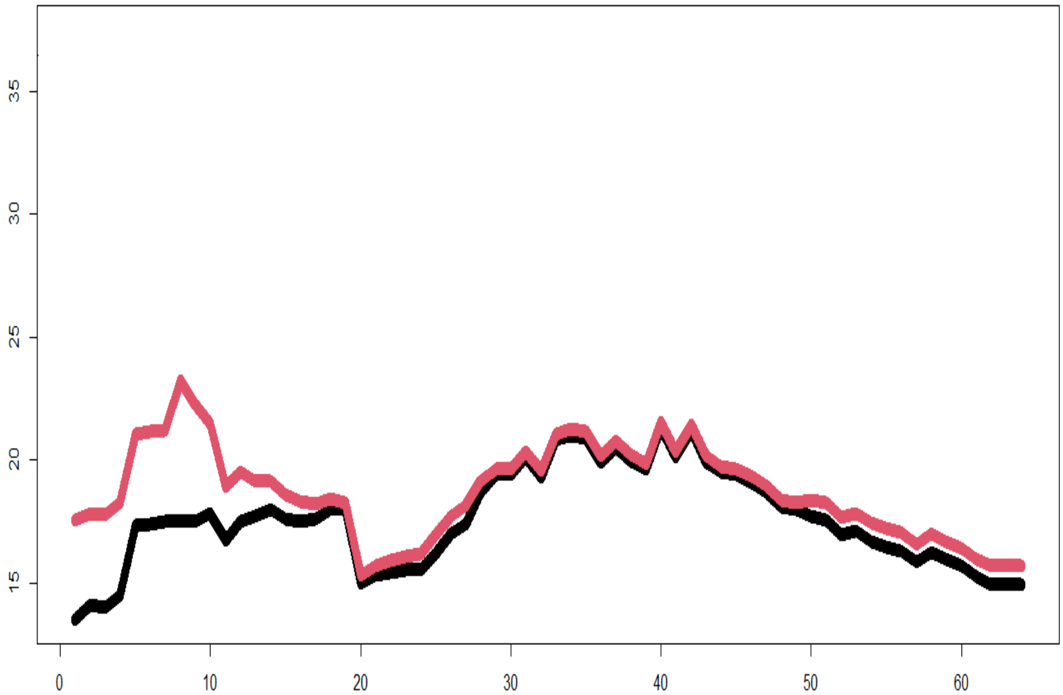

Figure 2. A real and estimated data for a dependent variable y (the Iraq market index) for the estimation methods used using the coif5 function, as the curve in black indicates the real values. In contrast, the red color indicates the estimated values.

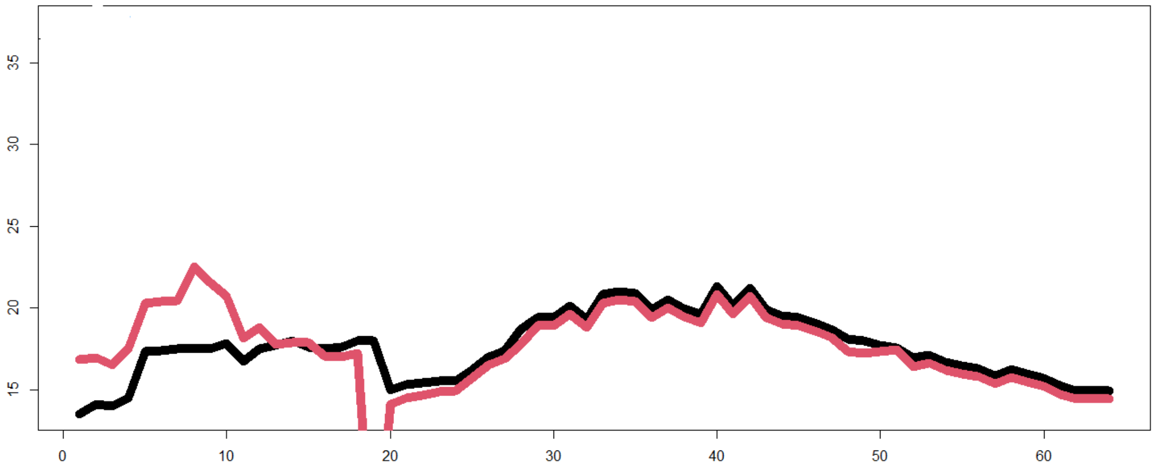

Figure 3. A real and estimated data for a dependent variable y (the Iraq market index) for the estimation methods used using the rbio1.3 function, as the curve in black indicates the real values. In contrast, the red color indicates the estimated values

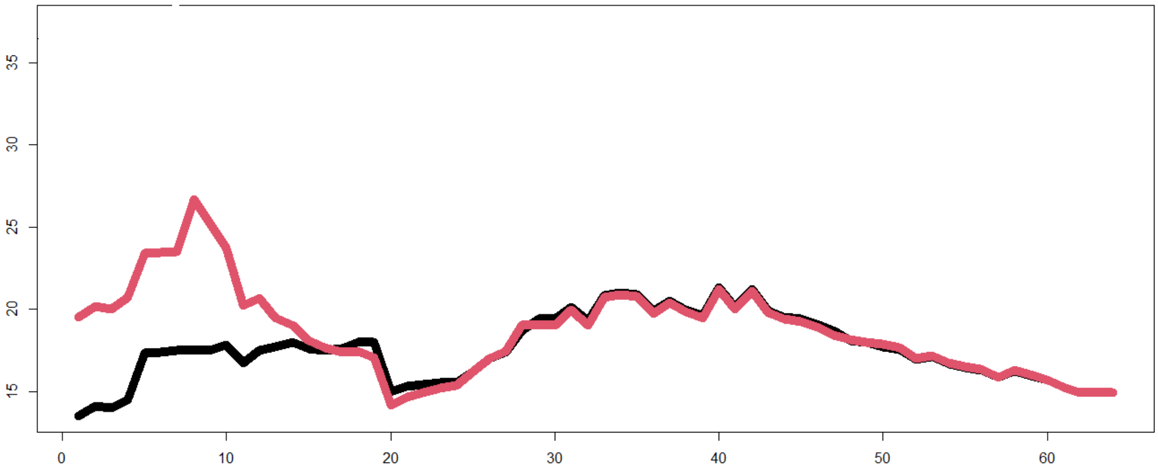

Figure 4. A real and estimated data for a dependent variable y (the Iraq market index) for the estimation methods used using the rbio1.5 function, as the curve in black indicates the real values. In contrast, the red color indicates the estimated values

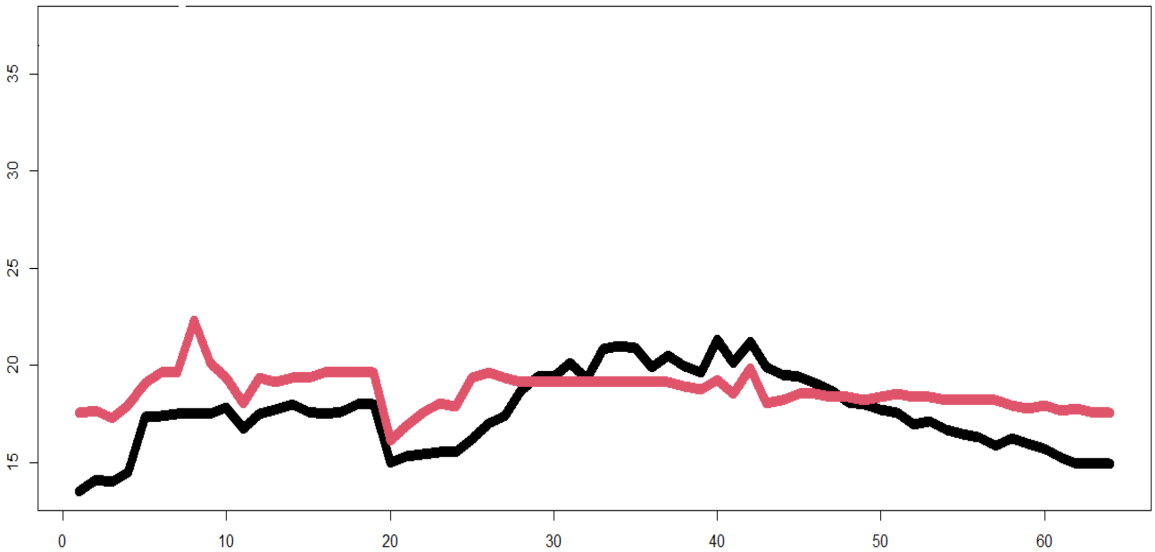

Figure 5. A real and estimated data for a dependent variable y (the Iraq market index) for the estimation methods used using the dob10 function, as the curve in black indicates the real values. In contrast, the red color indicates the estimated values

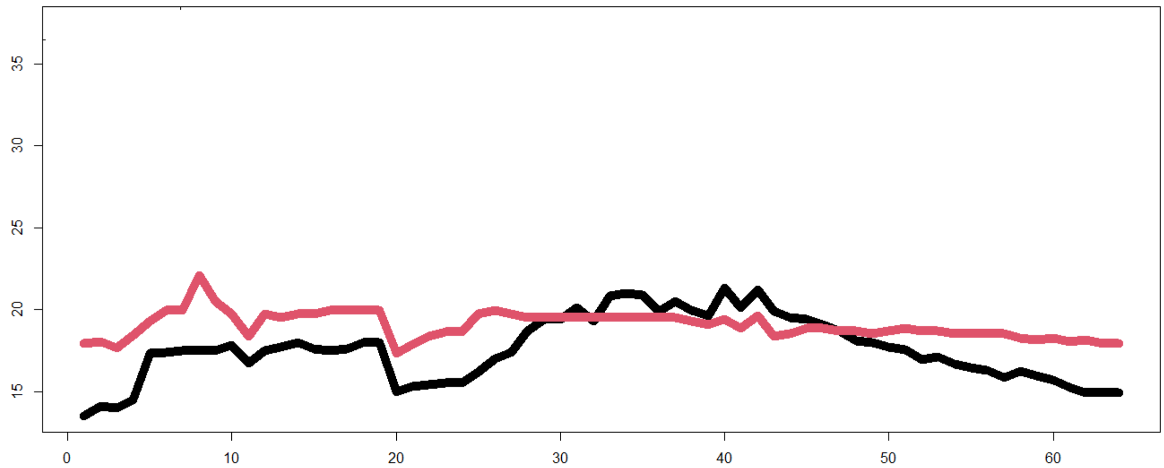

Figure 6. A real and estimated data for a dependent variable y (the Iraq market index) for the estimation methods used using the dob20 function, as the curve in black indicates the real values. In contrast, the red color indicates the estimated values.

7. Results and discussion

The best estimate of the price index function in the Iraqi market by the effect of trading volume was when using the mother function of type (coif1), followed by (coif5), followed by a mother function of type (rbio1.3) and (coif3) respectively.

There is an apparent effect of the trading volume on the price index in the Iraqi Stock Exchange, as the increase in trading volume leads to market activity, but it is insufficient. The change in the price index is explained by other variables that were not included in the model.

Both the mother function (Dub) and (Symlet) of all kinds regressed in such type of evidence in the efficiency of estimating the function.

The parametric models are more efficient in such studies, primarily financial and banking studies because the nature of these data is not controlled and contains many problems, such as heterogeneity and correlation, which the parametric models fail to estimate. In contrast, the non-parametric models are more adaptive due to their adaptive properties, In particular, the methods of wavelet reduction using different mother functions.

8. Conclusions

The best suitable waveform mother functions for the financial data are the mother functions of the type (coif 5), which were more accurate in choosing the detailed information, which is the decisive part in the accuracy of the estimate, in addition to the fact that the model needs to enter other variables, as the variable of the trading volume that was entered into the model Explain 40% of the change in the Iraqi stock market index, as we show the efficiency of the parameter model in the estimation process, being more efficient in the estimation process, being more adaptive, and specifically the methods of reducing wavelengths.

Declaration of competing interest

The authors declare that they have no any known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

[1] F. Abramovich, T. C. Bailey, and T. Sapatinas, “Wavelet Analysis and its Statistical Applications,” *J R Statist Soc D*, vol. 49, no. 1, pp. 1–29, 2000.

[2] U. Amato, A. Antoniadis, and I. De Feis, “Dimension reduction in functional regression with applications,” *Comput. Stat. Data Anal.*, vol. 50, no. 9, pp. 2422–2446, 2006.
[3] A. Antoniadis, “Wavelet methods in statistics: some recent developments and their applications,” *Stat. Surv.*, vol. 1, no. none, pp. 16–55, 2007.
[4] C. Burrus, R. Gopinath, and H. Guo, *Introduction to wavelets and wavelet transforms a primer-Prentice hall Upper Saddle River*. New Jersey, 1998.
[5] D. 5- Donoho and I. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American statistical association*, pp. 1200–1224, 1995.
[6] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
[7] -. J. Rafiee, P. W. Tse, A. Harifi, and M. H. Sadeshi, “A novel Technique for Selecting Mother Wavelet Function Using an Intelligent Fault Diagnosis System,” *Export Systems with Applications*, pp. 4862–4875, 2009.
[8] I. Johnstone, “Wavelet Shrinkage for Correlated Data and Inverse Problems: A Captivity Results,” *Statistica Sinica*, vol. 9, pp. 51–83, 1999.
[9] -. N. Waikeng and M. Salmenleong, “Wavelet Analysis: Mother Wavelet Selection Methods" 2013, Applied Mechanics and Materials,” pp. 953–958.
[10] M. E. Souza and G. Monico, *Wavelet Shrinkage: High-Frequency Multipath Reduction from GPS Relative Positioning” GPS Solutions*, vol. 8. Springer - Verlag, 2004.
[11] S. K. Hamza, “Using the wavelet analysis to estimate the non-parametric regression model in the presence of associated errors,” *International journal of nonlinear analysis and applications*, vol. 13, no. 1, pp. 1855–1862, 2022.
[12] M. Sharie, M. R. Mosavi, and N. Rahemi, “Determination of an Appropriate Mother Wavelet for De-noising of Weak GPS Correlation Signals Based on Similarity Measurements,” *2020, Engineering Science and Technology and International Journal*, pp. 281–288.
[13] G. P. Nason and B. W. Silverman, *Wavelet for regression and other statistical problems” school of mathematics*. 1997.
[14] S. C. Mallat, “A Theory for multiresolution signal De composition: The wavelet Representation,” *IEEE transactions on pattern and Machine intelligence*, vol. 11, 1989.
[15] N. A. Z. Abdullah, M. Abduljaleel , “Adaptive medical image watermarking technique based on wavelet transform,” *Iraqi journal of science*, vol. 22, no. 2, pp. 548–555, 2014.
[16] “Numerical solution for linear state space systems using Haar wavelets method,” *Baghdad Sci. J.*, vol. 19, no. 1, 2022.