

Wavelet Smoothing Spline to Estimate the Nonparametric Regression of Iraqi's Oil Revenue and Inflation

Tahir R. Dikheel Tamheed Ulewi Issa
tahir.dikheel@qu.edu.iq statistics.stp.24.5@qu.edu.iq
University of AL-Qadisiyah

Article history:

Received: 28/1/2025

Accepted: 18/3/2025

Available online: 15 /6 /2025

Corresponding Author : Tamheed Ulewi Issa

Abstract : With the growing energy demand, fluctuations in oil prices have become a key driver of global markets, influenced by economic and political factors. Oil plays a crucial role in the modern economy. Since the 1973 oil crisis, price changes have contributed to rising inflation, prompting central banks to adjust their policies to maintain price stability. This research employs wavelet transforms to enhance oil revenue and inflation data by reducing noise. Additionally, the smoothing splines method is applied to achieve precise regression estimates. The results indicate that the most effective approach is the smoothing spline method with the soft threshold function and the visu threshold value, demonstrating its efficiency in improving forecasts and analysing economic data, ultimately aiding in maximising oil revenues and ensuring financial stability.

Keywords: Wavelet Transformation, Nonparametric Regression, Threshold Functions, Threshold Values, Local Polynomial, Smoothing Splines.

INTRODUCTION: To predict and understand patterns and causal relationships in data, regression analysis models and measures the relationship between variables. However, estimating nonparametric regression models presents several difficulties for researchers because these models need to be able to handle complicated data and nonlinear patterns. The development of precise and adaptable nonparametric estimate methods is required to meet these demands. In the fields of finance, medicine, engineering, etc., this is becoming increasingly significant. where accurate estimation is essential for trustworthy outcomes. Certain techniques, like polynomial regression, are insufficient for handling complicated and time-varying data. Consequently, the smoothing spline method, which is versatile and responsive to changes in contemporary data, is employed. Studies in this field have expanded due to the large increase in interest in non-parametric regression models, particularly with the quick development of applied statistical methods. The research of wavelet shrinkage approaches has demonstrated the efficacy of wavelet transform in model selection and estimation accuracy improvement, as well as in reducing noise in forecasts and enhancing estimate accuracy (Donohoe & Ian ,1994).

In comparison to conventional techniques, wavelet transform has also been employed to eliminate noise from digital photographs, with notable improvements in image quality (Nason, 1995). To deal with missing data and estimate the unknown regression function, wavelet transform techniques enhanced nonparametric estimation techniques. Simulation trials demonstrated how well these techniques worked to increase the estimates' accuracy (Hamza, 2015). Both the wavelet transforms and the AdaBoost model have been used to increase the accuracy of short-range wind forecasts (Shao, 2017). The Wavelet transform increases the precision of medical diagnosis of degenerative changes in the spine, according to a study that used the Wavelet Gabor transform to analyze Magnetic Resonance Images (MRI)(Tao Yang,2020). Selecting the appropriate mother wavelet is crucial as it allows the majority of the signal's energy to be concentrated in a small number of wavelet coefficients. This enhances the ability to remove noise and obtain precise estimates for the regression function. Given the importance of analysing the price index in Iraq, various mother wavelets were evaluated using wavelet transform techniques. The aim was to identify the most suitable wavelet for this data, providing accurate estimates of the relationship between trading volume (independent variable) and the Iraq Stock Market Index (dependent variable). The results indicated that the Coif1, Coif5, and rbio1.3 wavelets gave the best results, with the lowest Mean Squared Error (Hamza & Ali, 2022).

Because they were unable to adjust to changing data and fluctuating noise levels. To improve the estimations, more adaptable techniques like the wavelet transform and smoothing bar were required. As part of a dissertation, this paper, we provide a novel two-step method to enhance nonparametric model estimation. The wavelet transform, which can analyse data at various frequency and temporal levels and enhance signal quality, is employed in the first phase as a fundamental data filtering and noise reduction technique. In order to obtain a more accurate estimation, the filtered data from the inverse of the wavelet transform is then subjected to smoothing using Smoothing Spline regression and local polynomials in the second step. employing both simulated and real data, the results show how effective the suggested approach is when compared to traditional methods employing the mean squared error criterion. This superiority confirms the suitability of the method for applications that require high accuracy in estimating nonparametric models in multiple domains.

2. Wavelets

Wavelet theory becomes an effective tool for analysing complex phenomena in a variety of domains, including engineering, communications, and medicine. Wavelet theory's advancements are essential for more thorough and precise data processing. Two primary filter types are employed while dealing with wavelet transform, high-pass filters are used to extract high-frequency information from the signal, and low-pass filters are used to recover low-frequency information. There are two primary components to wavelets (Daubechies, 1992). In order to determine the form and behavior of the wavelets at various resolution levels, more wavelet functions are constructed using the father function (scaling function). It is obtained using the following formula:

$$f_{(x)} = \sum_{k=0}^N C(K)f(2x - k) \quad (1)$$

where $C(k)$ are scalar coefficients that, when applied to a filter, alter the original signal. Each signal point's impact on its nearby points is determined by these weights. A mathematical function called the Mother function (Wavelet Function) is used to scale and shift a series of wave functions. The signals in the wavelet transform are examined using the functions that are produced.

$$w_{(x)} = \sum_{k=0}^N d(k)f(2x - k) \quad (2)$$

where $d(k)$ is an integer that is used to extract fast changes and precise details from a signal. It can be compared to a "edge finder" that displays abrupt signal shifts. The Mother function's properties (Donald et al., 2004) Dimensions and displacement: All additional wavelet functions are created via transformation and scaling procedures, which allow the signal to be broken down into its constituent parts. It is acquired by applying the subsequent formula:

$$\psi_{a,b}(t) = (1/\sqrt{|a|})\psi(t - b/a) \quad (3)$$

where, a is the scaling factor, b is the displacement coefficient.

Zero periodicity the mother function has the property that its integral over time is zero.

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \quad (4)$$

Local in time and frequency the mother function is local in both time and frequency.

Mother function types. We'll talk about a few popular wavelet functions, such as the Wavelet of Haar. The wavelet transform was invented by mathematician Alfred Haar in 1909. Haar, 1910. One of the most basic waveforms for wavelet processing and signal analysis is the Haar Wavelet. The Daubechies Wavelet, which is sometimes shortened to DbN, is a set of wavelets that were first presented by mathematician Ingrid Daubechies in 1988, The wavelet's order is denoted by N (Cohen et al, 1993). Many mathematicians and researchers found the wavelet transform in the late 20th century, which is a major improvement over earlier transforms. The wavelet transform uses a variable amplitude window, which allows for constant change of the window's size and location across the signal, in contrast to conventional techniques that use a fixed amplitude window. This makes it possible to efficiently extract information about frequency variations, giving signal analysis more flexibility. Therefore, even with noisy signals, wavelet transform may break down signals into components of various frequencies while maintaining the important information. Researchers across various disciplines agree that the wavelet transform is one of the most accurate and efficient methods for signal analysis, the most famous wavelet transforms are Continuous wavelet transform(CWT) (Zaidan, 2022). The continuous wavelet transform divides the signal into an infinite number of wavelets (the mother wavelet) that depend on location and scale. which can be expressed mathematically as follows:

$$\Psi_{a,b(t)=\frac{1}{\sqrt{|a|}}}\psi\left(\frac{t-b}{a}\right) \quad (5)$$

Where a is the Scale factor while b is the time location of the wavelet function, as for the amount $\frac{1}{\sqrt{|a|}}$ Represents the amount of energy of the signal. The continuous wavelet transform is defined as the internal multiplication of the signal $f(t)$ with the basis function of the wavelet (the mother wavelet) (Mallat,1999):

$$w_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi_{a,b} \left(\frac{t-b}{a} \right) f(t) dt \quad (6)$$

The DWT is a discrete set of wavelet scales and translations. It is especially designed for the sampled value (Daubechies, 1992). This transformation decomposes the signal into a mutually orthogonal set of wavelets. This specificity is the main difference between DWT and CWT. In reality, the DWT's dyadic grid scales the mother wavelet by a power of two. ($a = 2^j$) and converted using an integer ($b = k2^j$), where the location index, k , runs from 1 to $2^{-j}N$ (N is the total number of scales, and j ranges from 0 to J (number of observations)). The following formula is used to express the DWT (Rhif et al,2019):

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \quad (7)$$

3. Wavelet threshold functions

The wavelet function estimation process includes removing noise from the signal using a threshold term (Hedao & Godbole, 2011). Using the discrete wavelet transform, a frequency threshold is set to remove noise while keeping the important parts of the signal. Choosing the right threshold is very important. If the threshold is too high, it might remove important details, causing the signal to be distorted. If the threshold is too low, too much unnecessary detail might pass through, making the estimate unstable. So, selecting the right threshold is key to getting good results. The threshold value helps decide whether to keep or remove the wavelet coefficients. There are different types of thresholding methods (Donoho, 1995).

Controlling the magnitude of the interference or noise effect is the main objective of employing a threshold value; selecting the ideal threshold value is a crucial step in achieving the lowest mean square error (MSE). Among the most widely used techniques, which have been applied in this research, is Sure Thresholding

It is a technique for determining a threshold value in wavelet analysis, developed by David Donohue and Ian Johnston (Zaidan, 2022). This technique is based on Stein's Unbiased Risk Estimation (SURE), a principle used to unbiasedly minimize the mean square error (MSE) when estimating noisy signals.

$$SURE(\tau_j, d_{jk}) = N - 2 \sum_{k=1}^N I(|d_{jk}| \leq \tau_j) + \sum_{k=1}^N \min(|d_{jk}| \leq \tau_j)^2 \dots 8$$

where d_{jk} is the wavelet coefficient, τ_j is the threshold value, N is the number of wave coefficients, and I is the indicator function that takes the value 1 if the condition is true and 0 if the condition is false.

Since the wavelet transform is orthogonal, the transformation performed on the noise using the same transform is also orthogonal. This means that the wavelet coefficients d_{jk} The resulting transformation is also orthogonal.

$$\tau_{j,SURE} = \arg \min_{0 \leq \tau \leq \sqrt{2 \log N}} SURE((\tau_j, d_{jk})) \dots 9$$

We determine the threshold value τ that minimizes the Stein Unbiased Risk (SURE). This value is used to differentiate between transactions that represent noise and those that represent real information (Donoho, & Johnstone, 1995), and Visu Thresholding

This method improves the comprehensive threshold method, which is considered the basic building block in calculating the threshold value. This method addresses the shortcomings that the comprehensive threshold suffers from through its good performance even with increased sample size, as it gives a more homogeneous and preliminary estimate (Zaidan, 2022). The reason for this is that the global threshold works under conditions, the most important of which is that it is sensitive to large values of n , which leads to the loss of many wavelet coefficients with noise, and therefore, the threshold does not perform well during interruptions in the signal. This method can be explained according to the following formula:

$$\tau_{visu} = \sigma_n \sqrt{2 \log(n)} \dots 10$$

σ_n It is the standard deviation of the noise level, which can be found through the following relationship: (Donoho & Johnstone, 1995)

$$\sigma_n = \text{MAD}(Y)/0.6745 \quad \dots 11$$

Where MAD represents the absolute median of the wavelet coefficients.

4. Local Polynomial

Techniques for Estimating Nonparametric Regression Models Nonparametric regression models, which rely on an estimated model that offers an approximation of reality and helps with future predictions, are calculated using a variety of techniques. These techniques include the local polynomial method. Whittaker (1923) proposed approximating the regression curve to the true curve, which is where the idea of smoothing originated. The smoothing process, according to Whittaker, is a scale or change of the observations. Recent developments in computing software have led to a major evolution in smoothing approaches. Furthermore, academics have acknowledged that not all data types may benefit from parametric regression curve estimates. (Walker & Wright, 2000). Local Polynomial Regression (LPR) is a nonparametric regression method where a low-order ordinary least squares (OLS) regression is applied at each point of interest, using data from a neighbourhood around that point. The regression model is expressed as:

$$y_i = m(x_i) + e_i \quad (11)$$

Where, $m(x_i)$ is the true regression function.

e_i is a random noise term following a standard normal distribution.

Three crucial parameters determine the quality of the local polynomial fit: bandwidth h regulates the size of the local area surrounding x_0 . While a big h smoothes the fit but may oversimplify the function, a small h produces a more localized fit but may cause overfitting. Polynomial Order p : Establishes the polynomial's degree. Local linear regression ($p=1$) and local quadratic regression ($p=2$) are commonly employed. Kernel Function K : Indicates how observations are given weights according to how far away they are from x_0 . Gaussian, Epanechnikov, and uniform kernels are examples of common kernels. (Stone, C. J. 1977).

5. Spline Regression

Spline methods are among the most widely used techniques for approximating nonlinear functions. They were first employed in the early 20th century as an alternative to local polynomials. Using basis functions and control points, a spline is a continuous local polynomial curve that is used to approximate mathematical and geometric solutions in a smooth manner. Local polynomial functions are applied to each segment of the data once it has been separated into segments. Knots, which indicate the places where changes take place between segments, are used to create these functions. In order to reduce computing complexity and ensure that the model only uses the nodes that were taken from the data, rather than all of the data points, the number and placement of knots are carefully chosen; this speeds up the computations. The continuity of the spline at all locations, including the knots, guarantees an uninterrupted, accurate, and seamless portrayal. In the case of smoothing splines, the model minimizes the discrepancies between the actual and anticipated values by identifying the curve that best fits the data using the theory of nonparametric regression. This increases the estimation's accuracy without running the danger of overfitting or providing too much information. (Wahba, 1990).

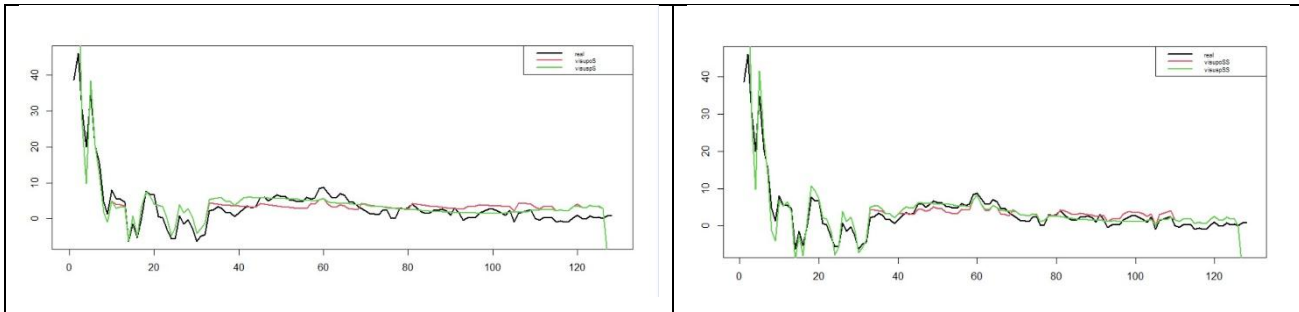
6. Iraqi's Oil Revenue and Inflation Data

Two variables—oil revenues and inflation—are included in the data, which spans the period from May 31, 2007, to December 31, 2017, and has a sample size of 128. The explanatory variable, which represents the entire amount of money the nation makes from the production and sale of oil, is oil revenue. Numerous intricate elements, such as OPEC's production decisions, political upheavals, and shifts in the economy, such as expansion or recession, have an impact on oil profits. Predicting oil prices is very difficult due to international difficulties, market fluctuations, and worldwide supply and demand. Traditional models are not appropriate for predicting oil income because of the unpredictable nature of these components. We chose non-parametric models, particularly the Smoothing Spline method combined with Wavelet Transform, to estimate the relationship. This approach allows for better data cleaning and analysis while adapting to the ongoing changes in the relationship between oil revenues and inflation, improving accuracy and effectively managing data fluctuations. the data is available at the OAPC database link <http://oapc.dbsys.oapec.org:8081/ords/f?p=100:23> and Central Bank database link <https://cbiraq.org/SeriesChart.aspx?TseriesID=423>.

Table 1 MSE values for the inflation rate with the best method.

Estimation Method	Threshold function	Threshold value	MSE
Smoothing Spline	Soft	Visu	12.98345
	Semi Soft		14.67383

Table 1 shows that the smoothing spline with soft threshold function and visu threshold value is the best method due to MSE, followed by the estimation method using smoothing spline with semi-soft threshold function and visu threshold value.

**Figure (1) real and estimated data using local polynomial estimation and Smoothing Spline.**

In the figure (1) the black line represents the original data, while the red line represents the results of estimating the function using the local polynomial method, while the green line indicates the results of estimating the function using the smoothing spline method with the threshold function soft and semi soft the threshold value Visu.

7. Conclusion

In this study, the wavelet transform is combined with two methods (local polynomial and smoothing spline) to evaluate the performance of estimate techniques. Estimation accuracy was the main focus of the comparison, and the findings indicated that performance significantly improved with increasing sample size, leading to more accurate estimations. The findings emphasized how crucial it is to select the right threshold functions because poor selections can result in decreased performance. Furthermore, the outcomes demonstrated the smoothing spline method's superiority over local polynomial regression estimation techniques because of its great capacity to handle complex data and its adaptability to various data properties.

8. Future Work

From the results of this study, we suggest the following recommendations:

Based on the results of the study confirming the importance of different thresholds in improving the accuracy of the estimates, we recommend further research on the impact of thresholds Visu, Soft and Semi-Soft on non-parametric estimation models. Such studies would contribute to improving the performance of models used in economic applications, especially the estimation of inflation rates and oil revenues. The study also highlights the importance of improving data quality and analysis using advanced mathematical models, such as Smoothing Spline and Wavelet Transform, to enhance the accuracy of estimates and support the effectiveness of fiscal and economic policies. We suggest further research into the impact of different thresholds to improve the accuracy of the model in economic applications. To enhance financial stability, diversification of income sources through agriculture, industry, and tourism is key, while investment in renewable energy can mitigate the effects of oil price fluctuations. In addition, improving data collection and analysis using sophisticated models, supporting central bank independence, and fostering a stable investment environment are vital for long-term financial stability. Stable fiscal strategies should focus on controlling inflation and maximising revenues to address both global and domestic economic challenges.

Table (2) Shows the Actual and Estimated Data of the Dependent Variable Using Estimation Methods

No.	x	y	ypol	yspl
1	16079151	38.6	53.157	53.157
2	19853823	46	69.515	69.515

No.	x	y	ypol	yspl
3	24172796	30.5	27.385	27.385
4	29240702	20	9.770	9.770
5	33375532	34.8	38.471	38.471
6	38439127	20.4	20.475	20.475
7	43948206	15.5	12.385	12.385
8	53162592	4.7	1.743	1.743
9	2503549	1.3	-0.942	-0.942
10	13015534	8.1	4.985	4.985
11	19627396	5.6	3.938	2.849
12	26340364	5.5	3.805	3.415
13	36466983	4.6	3.613	3.165
14	45246926	-6.3	-6.163	-6.163
15	54359419	-1.4	0.826	0.826
16	59746791	-5.2	-4.863	-4.863
17	66451214	0.3	3.114	3.415
18	71902074	7.6	7.486	7.486
19	75535725	6.7	6.370	6.370
20	79131752	6.8	3.985	4.035
21	2468325	0.6	3.715	3.715
22	4861691	0.2	3.315	3.315
23	7982503	-3.1	0.015	0.015
24	11149916	-5.7	-4.690	-4.690
25	14601313	-5.6	-2.590	-2.590
26	19187704	0.7	3.894	3.894
27	24132372	-1.5	1.615	1.615
28	28998911	-0.3	2.815	2.815
29	36250440	-2.7	0.415	0.415
30	41055791	-6.4	-4.133	-4.133
31	46092179	-4.9	-2.922	-2.922
32	51719059	-4.4	-1.285	-1.285
33	4941148	2.2	4.249	5.315
34	9293123	2.5	4.154	5.615
35	16329057	3.4	4.006	5.740
36	22932808	2.9	3.872	5.759
37	27843168	1.6	3.776	4.715
38	33697020	1.7	3.664	4.815
39	38673396	0.6	3.573	3.715
40	43405765	1.7	3.488	4.815
41	49000223	2.7	3.392	5.815
42	54464352	3.6	3.301	5.874
43	59964428	3.1	3.213	5.874
44	66819670	3.3	3.108	5.858

No.	x	y	ypol	yspl
45	7237156	5.8	4.198	5.857
46	14661596	5.9	4.041	5.743
47	23393615	5	3.863	5.702
48	31851046	5.8	3.699	5.690
49	41695814	6.6	3.519	5.615
50	49765754	6.1	3.379	5.578
51	58071727	6.2	3.243	5.573
52	66520685	5.2	3.113	5.495
53	74723651	5.1	2.994	5.467
54	82894707	4.8	2.884	5.441
55	91123832	4.8	2.780	5.125
56	98090214	6	2.885	5.065
57	8668492	5.4	4.168	5.021
58	14093408	5.7	4.052	4.958
59	22873204	8.3	5.185	5.185
60	37483854	8.7	5.585	5.585
61	50202966	7.2	4.085	4.725
62	59325823	5.8	3.223	4.382
63	66918350	5.7	3.107	4.341
64	72024164	7	3.885	4.253
65	45334668	6.4	3.455	4.177
66	96001586	4.8	2.722	4.148
67	106729085	4.5	2.605	3.997
68	116597076	3.6	2.508	3.964
69	9644874	2.8	4.147	3.918
70	18132528	2.2	3.969	3.816
71	28270812	1.3	3.767	3.455
72	36936615	1.2	3.604	3.445
73	46319198	1.1	3.438	3.423
74	54069509	2.3	3.308	3.332
75	61684583	2.5	3.186	3.014
76	73206555	0.1	3.016	2.917
77	84145401	0.3	2.868	2.860
78	91446283	3.1	2.776	2.692
79	99324469	2.7	2.685	2.677
80	110677542	3.1	2.565	2.665
81	8252744	4	4.176	2.554
82	16798073	3	3.996	2.518
83	23406812	2	3.863	2.295
84	32544939	1.5	3.686	2.222
85	41219920	1.4	3.527	2.134
86	50368089	2.3	3.369	2.039

No.	x	y	ypol	yspl
87	55517837	2.3	3.284	1.988
88	67131218	2.7	3.104	1.969
89	74664750	2.1	2.995	1.866
90	81823595	0.9	2.898	1.826
91	90285779	3	2.790	1.698
92	97072410	1.6	2.710	1.632
93	3684273	-0.4	2.715	1.624
94	6816588	0.3	3.415	1.614
95	11170092	0.2	3.315	1.607
96	17249458	0.5	3.615	1.550
97	21781638	1.7	3.895	1.473
98	28216875	2.2	3.768	1.467
99	32258696	2.6	3.691	1.468
100	37335553	2.6	3.597	1.470
101	41107533	2.1	3.529	1.473
102	45407324	1.6	3.454	1.515
103	48338973	1	3.403	1.700
104	51312621	2.3	3.353	1.806
105	1939372	-0.9	2.215	1.812
106	3857459	1.5	4.272	1.837
107	5971971	1.8	4.226	1.845
108	8585645	2.1	4.169	1.863
109	15186201	2.5	4.030	2.308
110	19350311	0.1	3.215	2.319
111	22537760	-0.4	2.715	2.429
112	26865821	0.2	3.315	2.560
113	29463906	0.3	3.415	2.565
114	34350332	0.2	3.315	2.635
115	38269023	-1	2.115	2.115
116	44267063	-0.7	2.415	2.415
117	4467428	-0.9	2.215	2.215
118	8569894	-0.8	2.315	2.315
119	13390064	0.3	3.415	3.415
120	17723733	1	3.977	3.560
121	22729366	0.1	3.215	3.215
122	27008790	-0.1	3.015	3.015
123	32949964	0.7	3.678	3.766
124	39197493	0.2	3.315	3.315
125	47808419	0.4	3.412	3.515
126	54193699	0.1	2.833	2.833
127	59450680	0.7	-9.895	-9.895
128	65071929	0.8	-10.982	-10.982

References

1. Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3), 613-627.
2. Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3), 425-455.
3. Donoho, D. L., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432), 1200-1224.
4. Hamza, S. K., & Ali, S. (2022). Estimation of nonparametric regression function using shrinkage wavelet and different mother functions. *Periodicals of Engineering and Natural Sciences (PEN)*, 10(6), 96-103.
5. Hamza, S. K. (2024). Comparison of Some Wavelet Transformations to Estimate Nonparametric Regression Function. *Journal of Economics and Administrative Sciences*, 30(142), 532-549.
6. Shao, H., Wei, H., Deng, X., & Xing, S. (2017). Short-term wind speed forecasting using wavelet transformation and AdaBoosting neural networks in Yunnan wind farm. *IET Renewable Power Generation*, 11(4), 374-381.
7. Soomro, T. A., Zheng, L., Afifi, A. J., Ali, A., Soomro, S., Yin, M., & Gao, J. (2022). Image segmentation for MR brain tumor detection using machine learning: a review. *IEEE Reviews in Biomedical Engineering*, 16, 70-90.
8. Stone, C. J. (1977). Consistent nonparametric regression. *The annals of statistics*, 595-620.
9. Zaidan, U. O. (2022). Reduced wavelet estimates using different threshold rules for the nonparametric regression model with application. Master's thesis in statistics, College of Administration and Economics, University of Baghdad.
10. Wahba, G. (1990). Spline models for observational data. Society for industrial and applied mathematics.