STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, August 2026, pp 1110–1124.
Published online in International Academic Press (www.IAPress.org)

Predicting the inflation rate in Iraq using multi-time series models ARIMAX with Daubechies Wavelet

Adhab Nawaf Mahmood Al-Hussein, Najlaa Saad Ibrahim Alsharabi*

Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

Abstract This research utilizes a hybrid model combining the Daubechies Wavelet Transform with the ARIMAX model to eliminate noise and improve the characteristics of the two time series for forecasting the inflation rate in Iraq. Inflation is a major negative economic phenomenon affecting most economies worldwide due to its impact on various economic, social, and other sectors. This study compares the performance of the ARIMAX model before and after applying the wavelet transform to the inflation rate (an internal variable) and oil prices (an external variable) during the period from December 2010 to January 2024. The RMSE and MAE criteria were used to assess the forecast accuracy. The results showed the model's superiority after applying the DaubechiesWavelet, as it exhibited the lowest values for the forecast accuracy criteria. This hybrid approach is preferred for forecasting time series with high volatility.

Keywords Daubechies Wavelet, ARIMAX model, inflation rate, forecasting

DOI: 10.19139/soic-2310-5070-3822

1. Introduction

To study and analyze complex economic phenomena, multivariate time series analysis is used. It is the most suitable analytical tool for such variables because it examines the relationships between them over time [12]. One of the most important models used to analyze these relationships is the ARIMAX model, which incorporates external variables to improve the accuracy of forecasting future values. However, its efficiency remains limited due to irregular fluctuations and noise in economic data. To improve forecasting, modern data processing and enhancement tools have been used, including wavelet transforms, which analyze data in both time and frequency domains [13]. The Daubechies wavelet has demonstrated its exceptional capabilities in analyzing noise-containing time series data into different frequency components. This, in turn, enables researchers to isolate the overall trend from short-term fluctuations [14]. Combining the Wavelet Transform with the ARIMAX model is a modern statistical method aimed at improving estimation efficiency and predictive accuracy. This has been demonstrated by numerous studies that have confirmed that subjecting data to pre-noise processing using wavelets before modeling with a traditional model achieves satisfactory results that are difficult to obtain when using raw data directly [15].
The problem of this study is the existence of complex dynamic relationships between inflation data and oil price fluctuations and the inability of traditional models to absorb and represent these relationships, especially in light of the noise that limits the efficiency of these models. Hence, the aim of the study is to build a multivariable ARIMAX model enhanced by wavelet transformation, specifically using Daubechies wavelets to address the noise, then to compare and evaluate the predictive performance of the model before and after data processing, and to identify the most efficient model in predicting the behavior of the target variables, which in turn provides accurate predictive insights for many decision-makers to develop more effective economic policies

*Correspondence to: Najlaa Saad Ibrahim Alsharabi (Email: najlaa.s.a@uomosul.edu.iq). Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

ISSN 2310-5070 (online) ISSN 2311-004X (print)

2. Literature Review

Many previous studies have focused on time series analysis and examining the relationships between economic variables. Traditional models and wavelet transformations have been used to predict traffic accidents, public revenues and expenditures, oil prices, and inflation. ARIMA and ARIMAX models have played a fundamental role in time series analysis. In 1976, Box and Jenkins presented the time series modeling phases for ARIMA models, which facilitated the use of ARIMAX models to analyze short- and long-term relationships between the study variables.
Several studies have demonstrated that the ARIMAX model possesses a high capacity for explaining the impact of extrinsic variables on intrinsic variables. However, its efficiency remains limited when dealing with data contaminated by noise and outliers, especially in volatile applications such as predicting traffic accidents, oil prices, and inflation [16, 13].
Many researchers have focused on using wavelet transforms, a technique capable of smoothing data by removing noise, for price forecasting, economic series analysis, and financial applications [14, 15]. Despite its high predictive and analytical capabilities, it requires integration with models that can explain the impact of economic variables. Therefore, recent studies have tended to combine traditional models with wavelet transforms to leverage the strengths of both approaches. Several studies have demonstrated the superiority of a model combining wavelet transforms with the ARIMAX model in forecasting, where the ARIMAX model analyzes the relationship between variables and the wavelet transform refines the data.
In the context of forecasting inflation rates in Iraq, scientific bodies concerned with monetary policy are placing increasing importance on econometric models capable of handling high volatility. Despite previous studies demonstrating the effectiveness of wavelet transformation, particularly the Daubechies wavelet, in refining data and improving forecast accuracy when combined with ARIMAX models, the researcher notes a clear scarcity of studies that have applied this hybrid (Wavelet-ARIMAX) model to forecast inflation rates specifically in Iraq. This study represents an attempt to fill this gap.

3. Diagnosing the ARIMAX Model

The Autoregressive Moving Average with External Variables (ARIMAX) model is a multivariate model that can be defined as an extension of the ARIMA model, adding statistically significant external variables to enhance the model's explanatory power and improve prediction accuracy. This model is used to model and predict time series data when the series' behavior is influenced by external factors and is not limited to the effects of its past values. It is widely used in economics, finance, and environmental science [17]. Although the ARIMAX model is very similar to the ARIMA model in terms of its basic structure, the addition of external variables in ARIMAX increases the model's complexity, but it enhances its ability to capture and measure the impact of external factors on the internal variable. These external variables must be independent, but must have a strong and significant correlation with the internal variable [20]. To build an efficient and reliable ARIMAX model, several basic statistical assumptions must be met. Because model building is iterative, these assumptions must be verified and addressed [16]:

1. Both the model residuals (errors) and the time series under study must be stationary. To provide a solid basis for estimation and prediction, the series mean and variance must be constant over time.
2. The residuals of the final model must be "white noise." Accordingly, there should be no obvious patterns or strong autocorrelations over time, and the residuals should be independently and randomly distributed with a mean of zero and a constant variance.
3. The coefficient of the exogenous variable(s) in the final model must be statistically significant.
4. The exogenous variable(s) must show a unidirectional lead-lag relationship with the endogenous variable.
5. When using more than one exogenous variable in the model, it must be ensured that these variables are free from the problem of multicollinearity. That is, there should be no strong correlation between the exogenous variables themselves, as this leads to inaccurate estimates of the coefficients.

The integrated ARIMAX model goes through several sequential methodological stages. These stages can be summarized as follows:

3.1. Stationary

Both the internal and external variables must first undergo time series stationarity tests, namely the extended Dickey-Fuller (ADF) test and the Phillips and Perron (PP) test, since ARIMAX models assume data stationarity. Differential analysis should be used to address any instability [1, 18].

3.2. Cross-Correlation Function

After confirming the stationarity of the time series, the cross-correlation function is used to examine the nature of the relationship between the internal variable (Y) and the external variable (X). It can be defined as a function that measures the correlation between two time series at different time intervals. It works by determining the optimal delay of the external variable, which helps to provide the best prediction for the internal variable[21, 22]. The cross-correlation function is mathematically defined as follows: if we have two stationary time series, then the cross-correlation coefficient at time displacement k can be written in the following form [2, 23]:

$$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y} \qquad , for\ k = 0, \pm 1, \pm 2, \ldots \tag{1}$$

Whereas $\rho_{XY}(k)$ represents the correlation coefficient between X at time t and Y at time (t+k), $\sigma_X \sigma_Y$ the standard deviation of the two series Y_t and X_t, respectively and $\gamma_{YX}(k)$ the cross-variance function.

3.3. Box-Jenkins Methodology for Building an Endogenous Variable Model

After confirming a significant relationship between the exogenous variable X and the endogenous variable Y, we move to the next stage: modeling the time series of the endogenous variable using the Box-Jenkins ARIMA methodology. In 1970, Gwilyn Jenkins and George Box presented an integrated statistical methodology that is considered one of the most innovative and straightforward methods for time series analysis. This methodology includes a set of statistical models that have been widely used in analyzing various types of time series (seasonal and non-seasonal) and predicting their future values [3]. Predictions based on this methodology have demonstrated high accuracy in diagnosing and describing the future trajectory of the phenomena under study, thus enhancing its importance in supporting decision-making processes [4]. Reaching the optimal prediction model is often a challenging task, requiring four fundamental and systematic stages [5]:

1. Identification: This stage is fundamental to building Box-Jenkins models. It involves determining the model's rank using the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF), as well as employing certain statistical parameters (AIC, BIC, AICc) to identify the most suitable model that corresponds to the lowest value of these parameters [6].
2. Estimation: After determining the model's rank in this stage, the model's parameters are estimated. Several methods are used to estimate the parameters of the autoregressive (AR) model, but the least squares method is the most commonly used. This method is based on the principle of minimizing the sum of the squared errors of the estimation to its lowest point [7]. However, when estimating the parameters of MA and ARMA models, the estimation process becomes more complex for two reasons: first, the models are non-linear in their parameters, and second, the error variable is not observed in these models. Therefore, the Maximum Likelihood Method is used in the estimation process, and sometimes the Method of Moments is employed [4].
3. Diagnostic Checking: The efficiency and validity of the estimated model in representing time series data are verified. Selecting the ideal ARIMA model requires considerable skill, provided that the estimated residuals of this model appear as white noise [24]. Several tests are used for this purpose, including plotting the autocorrelation function and using the Ljung-Box test.

3.4. Adding an External Variable to the ARIMA Model

After evaluating the ARIMA model and confirming its suitability for the internal variable Y_t, we can now enhance its efficiency by adding the external variable X_t to obtain the ARIMAX model. This model considers the impact of the external variable on the time series and increases the predictive accuracy of the model. The ARIMAX model is expressed mathematically as follows [25]:

$$\emptyset\left(\beta\right)\left(1-\beta\right)^{d}Y_t=\alpha X_t+\theta\left(\beta\right)\varepsilon_t \quad (2)$$

Whereas α is the coefficient of the external variable. $\emptyset\left(\beta\right)$, $\theta\left(\beta\right)$ is a polynomial representing the coefficients of the autoregressive model and the moving averages, respectively.
The ARIMAX model follows the same methodology as the ARIMA model in its formation, as it goes through the three previously mentioned model formation stages (diagnosis, assessment, diagnostic examination) [20].

3.5. Prediction

Reaching this stage signifies the completion of all phases of the Box-Jenkins methodology. When the model successfully passes all previous tests and stages with high efficiency and reaches this point—the primary objective of studying and analyzing any time series–we can then use this model to predict the future values of the phenomenon under study [8, 26].

4. Wavelet Transformation

This transformation is one of the most advanced transformation techniques in signal processing. It is a sophisticated analytical mathematical method used to analyze signals in both the time and frequency domains simultaneously, allowing for the study of local signal characteristics. This transformation is implemented using various wavelet functions [27]. In this context, the Discrete Wavelet Transform is used on the original time series data to obtain more accurate and realistic predictions. The time series data is treated as signals and decomposed into low and high frequencies to obtain approximate and detailed coefficients along the wavelength. This process is called decomposition [28]. The mathematical formula for the Discrete Wavelet Transform (DWT) is [10]:

$$\mathrm{W_Y\left(j,k\right)=\sum_{t} Y\left(t\right)2^{-j/2}.}\psi\left(\mathrm{2^{-j}t-k}\right) \quad (3)$$

Whereas $\mathrm{W_x\left(j,k\right)}$ are the resulting wavelet parameters, j represents the transition, k represents the shift along the time axis.
The general procedures for signal analysis and purification (noise removal) using wavelet transformation include several stages, the most important of which are [9]:

1. **Analysis**: Wavelet selection, selection of the N-plane, and application of the Discrete Wavelet Transform (DWT) to the signal S in the N-plane by:
   (a) Passing the series through a low-frequency filter to obtain the approximation parameters that represent the low-frequency part of the series. The approximation parameters are expressed mathematically [29]:

   $$Y_{scaling}[k]=\sum_{n} g_0\left[n\right].x\left[2k+n\right] \quad (4)$$

   Whereas $Y_{scaling}[k]$ is the approximation coefficient , $\sum_n g_0\left[n\right]$ is the approximation filter (low-frequency pass filter).
   (b) Passing the series through a high-frequency filter yields the detail coefficients that represent the high-frequency portion of the series. The detail coefficients are expressed mathematically [29]:

   $$Y_{wavelet}[k]=\sum_{n} g_1\left[n\right].x\left[2k+n\right] \quad (5)$$

Whereas $Y_{wavelet}[k]$ Detail Parameters and $\sum_n g_1\,[n]$ etail Filter (High-Frequency Pass Filter).

2. **Threshold**: To remove noise, the parameters generated in the first stage are passed through a threshold test. In this case, parameters with values below the specified threshold are removed, while the remaining parameters are used to construct the signal.
3. **Reconstruction**: Using the wavelet parameters after applying the threshold, the signal is reconstructed by applying the Inverse Discrete Wavelet Transform (IDWT) to obtain a purified signal similar to the original signal. In other words, a new, noise-free sequence is obtained while preserving the important data characteristics for use in modeling. This process is called reconstruction [19].

5. Daubechies Wavelet

The Daubechies wavelet family is named after Ingrid Daubechies, who developed the compressed orthogonal wavelet, making discrete-time wavelet analysis possible. The first-order Daubechies wavelet (db1) is known as the Haar wavelet, meaning it is the simplest form of the wavelet. Higher-order Daubechies functions are difficult to describe analytically, as the order of a Daubechies function indicates the number of vanishing moments or zero moments of the wavelet function. This is weakly related to the number of oscillations of the wavelet function. The higher the number of vanishing moments, the better the function can be analyzed,as the correlation between wavelet coefficients at different scales decreases as the wavelet order increases [9, 30]. Daubechies wavelet filters are usually denoted by the symbols DN or dbLi, where d and db are abbreviations for the researcher's name, N is the filter's rank, and Li denotes the number of vanishing moments of the wavelet function. The practical method for using this wavelet in signal analysis and applying it as a filter to remove or reduce noise from the series is done by estimating the measurement function equation ∅(t), which takes the following form [31].

$$\emptyset(x) = \sum_{k=0}^{N-1} C_k\, \emptyset\,(2x-k) \tag{6}$$

Whereas the parameters of the Low-Poss Filter C_k are determined to satisfy the following conditions:

1. Normalization Condition

$$\sum_{k=0}^{N-1} C_k = \sqrt{2} \tag{7}$$

2. Vanishing Moments Condition

$$\sum_{k=0}^{N-1} (-1)^K \;\; K^m\, C_k = 0 \qquad m = 0, 1, \;\ldots\; \frac{N}{2} \; - 1 \tag{8}$$

3. Orthogonality Condition

$$\int \emptyset\,(x)\,\emptyset\,(x-m)\,dx \tag{9}$$

This wavelet has several important properties, the most prominent of which are [11]:

1. The pivot point of the wavelet (dbN) is confined within the interval [0,2N-1].
2. The wavelet (dbN) has a number of vanishing moments equal to N.
3. The shape of the functionals for this wavelet (dbN) is characterized by asymmetry.
4. The regularity or smoothness of the wavelet of type (dbN) increases as the filter length or order increases.

The following steps represent the hybrid methodology proposed in our research:

1. Using the Daubechies wavelet to filter the data because it has a high noise isolation capability.

2. Using discrete wavelet analysis (DWT) because it analyzes the series into approximate and detailed parameters. The optimal level (level 3) was chosen because it achieves a balance between noise removal while preserving the dynamic properties of the series.

3. Using the Rigorous SURE (Stein's Unbiased Risk Estimate) method exact method to determine the threshold because it is an unbiased technique that minimizes the difference MSE between the original and filtered signals.

4. Using a soft thresholding because it removes parameters below the threshold and minimizes parameters above the threshold to maintain signal continuity.

5. Using inverse discrete wavelet transform (IDWT). after applying the soft thresholding to reconstruct the noise-free filtered series.

6. Estimating the ARIMAX model of the filtered series to analyze the relationship between the two variables and predict future values.

6. Practical Aspect

This study examined two economic variables in Iraq using data from January 2010 to December 2024: oil prices (X) and the inflation rate (Y). The Central Statistical Organization of the Ministry of Planning provided the inflation rate data with 2012 as the base year, while OPEC provided the oil price data. The ARIMAX model was used to predict the inflation rate in Iraq. To improve the forecasting process, the Daubechies wavelet model was employed with the ARIMAX model. The training phase was conducted between January 2010 and December 2022. The remaining dataset was used for testing to assess the model's suitability. The analysis was performed using two software programs: MATLAB Version 7 with the Wavelet Toolbox for noise removal using wavelet transform (detailed in Section 6.2), and R Version 4.5.2 with the Forecast, TSeries, and TSA packages for diagnosing the ARIMAX model. The stationarity of the two series—oil prices and the inflation rate—was examined using a graph, as shown in the following figures:

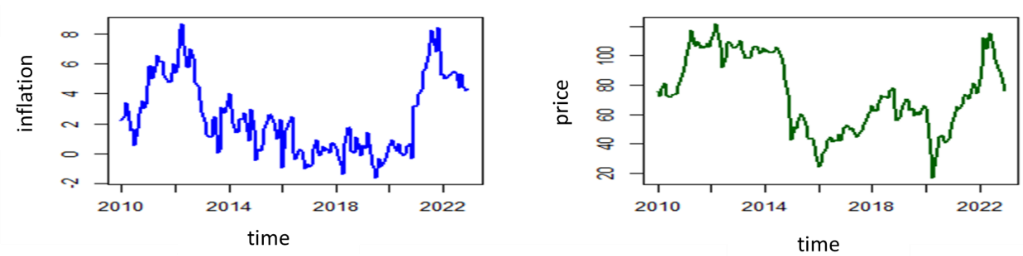

Figure 1. Plotting the original inflation and oil price series.

The aforementioned statistics demonstrate how unstable the two series are, with their graphs showing erratic variations. This suggests that they are unstable. As indicated in Table (1) below, statistical tests were also employed to verify the series' stability.

Table 1 below shows the results of the two tests (ADF and PP) at all levels (constant presence and absence of trend, constant and trend presence, and absence of both) for the inflation and oil series. The P-values shown in parenthesis beneath the computed t-test value are greater than 0.05. This suggests that both series are unstable.

Because both the ADF and PP test values were less than 0.05 after the initial difference, stability was reached. This indicates that no additional differences are needed for the second series since the two series were steady after the initial difference was obtained.

Table 1. Inflation and Oil Series Stability Tests

Series	ADF			PP		
	Presence of a constant and absence of a general trend	Presence of a general and consistent trend	Absence of a general and consistent trend	Presence of a constant and absence of a general trend	Presence of a general and consistent trend	Absence of a general and consistent trend
Original oil data	-2.139962 (0.2295)	-2.405078 (0.3755)	-0.682282 (0.4200)	-1.791292 (0.3838)	-1.931430 (0.6333)	-0.566389 (0.4705)
Oil data after the difference	-8.498339 (0.0000)	-8.465612 (0.0000)	-8.526679 (0.0000)	-8.124041 (0.0000)	-8.090296 (0.0000)	-8.156884 (0.0000)
Original inflation data	-2.478878 (0.1226)	-2.463071 (0.3460)	-1.633570 (0.0965)	-2.391321 (0.1459)	-2.373768 (0.3918)	-1.422488 (0.1439)
Inflation data after the difference	-13.83161 (0.0000)	-13.79257 (0.0000)	-13.87481 (0.0000)	-14.07378 (0.0000)	-14.03511 (0.0000)	-14.12080 (0.0000)

6.1. Diagnosing the ARIMAX model before performing the wavelet transformation

To diagnose the ARIMAX model, there must be a relationship between the study variables. Figure (2) below, plotting the cross-correlation coefficients between the two series after stationarity (taking the first difference of the two series), confirms that there is an effect of the price of oil on inflation. Similarly, Table (2) shows the results of the Granger causality test, indicating a unidirectional causal relationship between oil prices and inflation (at a significance level of 5%). This result indicates the suitability of performing the ARIMAX model.

Table 2. Granger Causality Test.

Null Hypothesis:	F-Statistic	Prob.
DY does not Granger Cause DX	2.59239	0.0782
DX does not Granger Cause DY	3.65716	0.0281

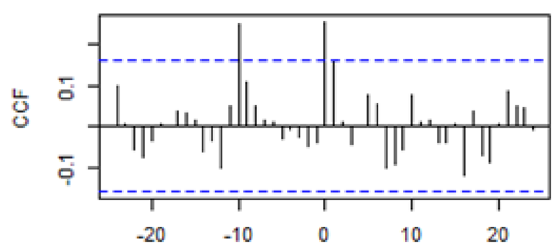

Figure 2. plots the cross-correlation coefficients between the oil and inflation series.

First, the ARIMA model for the inflation series was identified, and to determine the ranks of the ARIMA model, the plotting of the ACF and PACF functions for the stationary time series was used, as shown in the following figure (3) and the statistical parameters (AIC, AICc, BIC) were used to determine the best model. 255 models were tested, and it was concluded that the best model is SARIMA(0,1,1)(1,0,1)12, as it has the lowest values for the statistical parameters, where the value of AIC=439.771, BIC=424.554, AICc=424.821.
The SARIMA $(0,1,1)(1,0,1)_{12}$ seasonal model proposed in the first phase was estimated. Table (3) below shows the results of the model estimation along with its estimated coefficient.

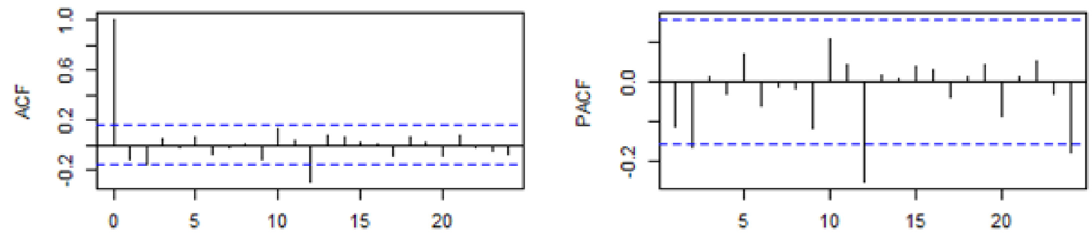

Figure 3. Plotting the ACF and PACF functions for a steady–state time series.

Table 3. Results of estimating the best inflation series model

model	Estimate	Std. Error	T_ statistical	P-Value
SARIMA(0,1,1)(1,0,1)$_{12}$	MA1=0.1682	0.0805	2.09	0.038
	SAR12= 0.3311	0.1505	2.20	0.029
	SMA12= 0.7576	0.1126	6.73	0.000

As shown in the table below, which displays the estimation results for the SARIMA(0,1,1)(1,0,1)$_{12}$ model, all coefficients are statistically significant. The accuracy of the SARIMA(0,1,1)(1,0,1)$_{12}$ model is confirmed and evaluated by examining the series residuals. The following figure (4) illustrates the graphs of both the autocorrelation function and the partial autocorrelation function of the residuals.

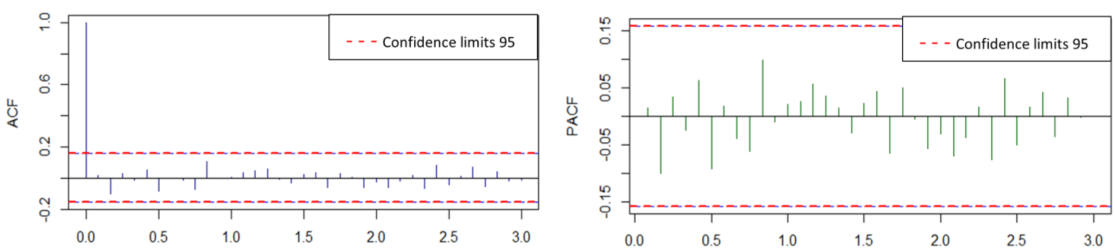

Figure 4. shows the autocorrelation function and the partial autocorrelation function for the residuals of the $\mathrm{SARIMA}(0,1,1)(1,0,1)_{12}$ model.

As shown in Figure (4) above, most of the autocorrelation and partial autocorrelation values for the residuals fall within the confidence intervals or critical lines, indicating randomness of the residuals. Furthermore, the Ljung_Box test was estimated for a number of lagging results, as shown in the following table (4):

Table 4. Results of the Ljung Box test for the model $\mathrm{SARIMA}(0,1,1)(1,0,1)_{12}$ model

Lag	Statistic	P_value
6	3.3779	0.7601
12	6.1593	0.9078
18	7.4845	0.9854
24	9.2642	0.997
30	12.5569	0.9979
36	14.5667	0.9994

The table (4) showing the results of the Ljung_Box test for several lags reveals that the p-value (p-value) is $>$ 0.05 for all lags. This supports the null hypothesis, which states that there is no autocorrelation of the residuals. These results confirm the quality and suitability of the SARIMA(0,1,1)(1,0,1)$_{12}$ model for estimating time series

data. The Breusch-Pagan test was also performed to confirm homoscedasticity, and the results showed that the variance was homoscedastic (p-value = $0.17706 > 0.05$).
ARIMAX models are an extended version of the ARIMA model that incorporates other independent (predictive) extrinsic variables. ARIMAX models are similar to multiple regression models, except that they allow for the exploitation of autocorrelation that may be present in the regression residuals. This process improves the accuracy of the prediction. After obtaining the SARIMA(0,1,1)(1,0,1)$_{12}$ model, we now find a suitable model for predicting the inflation rate with the extrinsic variable of oil prices. Several models were tested and relying on the statistical parameters (AIC, AICc, BIC) to determine the best SARIMAX model, it was concluded that the best model is SARIMAX(2,1,0)(0,0,2)$_{12}$, as it possesses the lowest values for the statistical parameters: AIC = 408.48, BIC = 426.74, and AICc = 409.05. Table (5) below shows the values of the coefficients and their standard errors for the estimated model.

Table 5. Shows the results of estimating the coefficients of $\mathrm{SARIMAX}(2,1,0)(0,0,2)_{12}$

model	Estimate	Std. Error	T_ statistical	P-Value
SARIMAX(2,1,0)(0,0,2)$_{12}$	AR1= -0.1772	0.0800	-2.21	0.0268
	AR2= -0.1783	0.0799	-2.23	0.0257
	SMA12= -0.4016	0.0810	-4.96	0.0000
	SMA24= -0.1635	0.0826	-1.98	0.0478
	Xreg= 0.0420	0.0101	4.18	0.0000

Table (5) above shows that the parameters are statistically significant, as the P-value < 0.05 for all parameters, which supports the strength of this model. The regression equation can be expressed as follows:

$$y_t = 0.0420x_t + \omega_t$$

$$\omega_t = -0.1772y_{t-1} \; -0.1783y_{t-2} + e_t + 0.4016e_{t-12} + 0.1635e_{t-24} \qquad (10)$$

Figure (5) below shows the autocorrelation and partial autocorrelation functions for the residuals of the SARIMAX (2,1,0)(0,0,2)$_{12}$ model. We observe that most of the autocorrelation values for the residuals fall within the confidence intervals or critical lines. Furthermore, the Ljung_Box test was estimated for a number of lagging values, as shown in Table (6), which indicates that the p-values corresponding to the test values for all lagging values are greater than the significance level of 0.05. This is another indicator of the randomness of the time series and the efficiency of this model. The Breusch-Pagan test was also performed to confirm homoscedasticity, and the results showed that the variance was homoscedastic (p-value = $0.5883 > 0.05$), thus confirming the efficiency and stability of the SARIMAX(2,1,0)(0,0,2)$_{12}$ model.

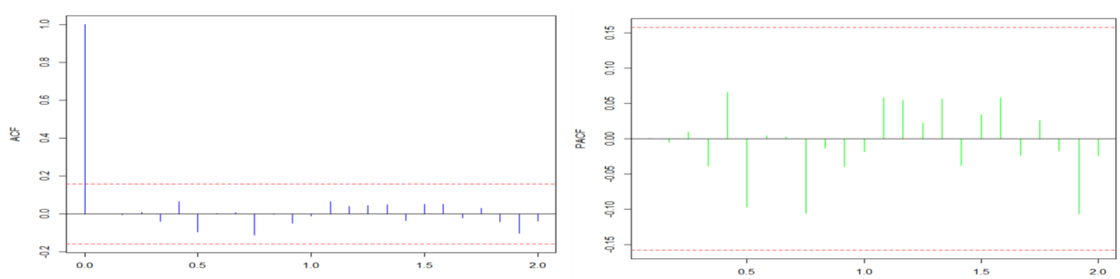

Figure 5. Shows the ACF and PACF for the residuals of the $\mathrm{SARIMAX}(2,1,0)(0,0,2)_{12}$ model.

Table 6. Results of the Ljung_Box test for the model $\mathrm{SARIMAX}(2,1,0)(0,0,2)_{12}$ model

Lag	Statistic	P_value
6	2.4443	0.8747
12	4.9442	0.9598
18	7.3835	0.9865
24	10.5720	0.9918
30	15.7915	0.9844
36	18.6177	0.9926

After passing all the necessary tests, the SARIMAX(2,1,0)(0,0,2)$_{12}$ model is ready to predict future values. Figure (6) below shows the inflation series diagram and expected values according to the SARIMAX(2,1,0)(0,0,2)$_{12}$ model.

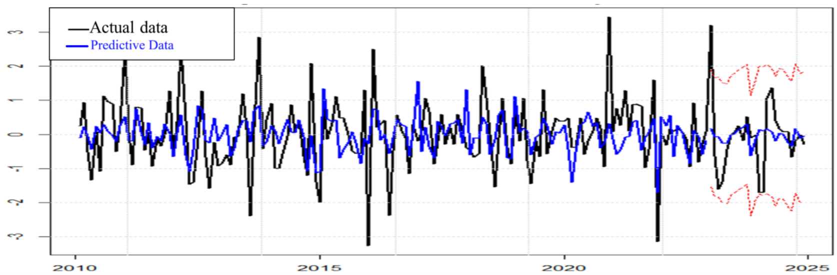

Figure 6. Shows the predicted values according to the $\mathrm{SARIMAX}(2,1,0)(0,0,2)_{12}$ model.

Table 7. The results of the prediction accuracy criteria for the model SARIMAX(2,1,0)(0,0,2)$_{12}$ and SARIMA(0,1,1)(1,0,1)$_{12}$

Data	**SARIMAX(2,1,0)(0,0,2)$_{12}$**		**SARIMA(0,1,1)(1,0,1)$_{12}$**	
	RMSE	MAE	RMSE	MAE
Training phase	0.8586	0.6587	0.91202	0.68192
Testing phase	0.9909	0.6610	0.98656	0.64128

The figure above illustrates the predicted values according to the SARIMAX(2,1,0)(0,0,2)$_{12}$ model. The blue line represents the predicted values, while the black line represents the actual values of the inflation series. The area between the two lines, shown in red, represents the high confidence interval (95%). The figure indicates that the predicted values are close to the actual values, suggesting that the model is well-suited to the data. Table (7) above shows the prediction accuracy parameters (RMSE and MAE) for both the SARIMA and SARIMAX models.

6.2. Diagnosing the SARIMAX Model for Wavelet-Filtered Data

The Daubechies wavelet (db2) was used to filter the data because of its ability to isolate noise in economic data. Multilevel analysis was used to analyze the two variables (inflation and oil prices). Different levels were used, and we found that level 3 was the best and most suitable for smoothing the two series, as it preserved the relationship between the two variables. When higher levels were used, we found that the series became excessively smoothed, and the smoothing lost short-term fluctuations, leading to a weakening of the relationship between the two variables. The Rigorous SURE (Stein's Unbiased Risk Estimate) method was chosen. This method is considered an unbiased estimation technique, as it selects an appropriate threshold that minimizes the mean squared error between the original and filtered signals. A Soft thresholding was also used, which eliminates wave parameters smaller than the threshold (considered noise) and reduces parameters larger than the threshold to maintain signal continuity.

The stability of the two filtered series (inflation rate and oil prices) was tested using statistical tests (ADF and PP). The p-values were found to be greater than 0.05, indicating that the filtered series were unstable. Stability was achieved after the first difference was taken into account, as the p-values for both ADF and PP became less than 0.05. This indicates that the series are stable after the first difference is taken into account, and therefore, no further differences need to be taken into account. After confirming the stability of the two filtered series, the cross-correlation function (CCF) was estimated as shown in the following figure (7). Significant correlation coefficients were observed, indicating a statistically significant correlation between the two filtered series in the short run. Table (8) shows the results of the Granger causal test, indicating a unidirectional relationship between oil prices and inflation. This demonstrates that the wavelet transformation preserved the relationship between the two variables, and the SARIMAX model can be used to analyze this relationship and predict the inflation rate in Iraq.

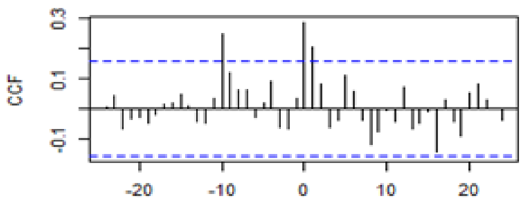

Figure 7. Shows the cross-correlation function between the two filtered series after the first difference.

Table 8. Granger Causality Test after wavelet transformation.

Null Hypothesis:	F-Statistic	Prob.
Wavelet-DX does not Granger Cause **wavelet-**DY	3.72967	0.0263
Wavelet-DY does not Granger Cause **wavelet-**DX	2.21058	0.1132

After applying the Daubechies wavelet transform to filter the data and remove noise, the appropriate model structure was determined using ARIMA models. This was done by observing the graphs of both the autocorrelation and partial correlation functions of the filtered inflation series after stationarity (as shown in Figure 8) and dependence on the statistical parameters (AIC, AICc, BIC). To determine the best ARIMA model, Several models were tested it was found that the best model was Wavelet_SARIMA$(0,1,0)(1,0,2)_{12}$, as it had the lowest values for the information parameters (AIC = 277.4956, AICc = 277.7623, BIC = 289.6693). The following table (9) shows the model estimation results, where we find that all its coefficients are statistically significant, with p-values less than 0.05.

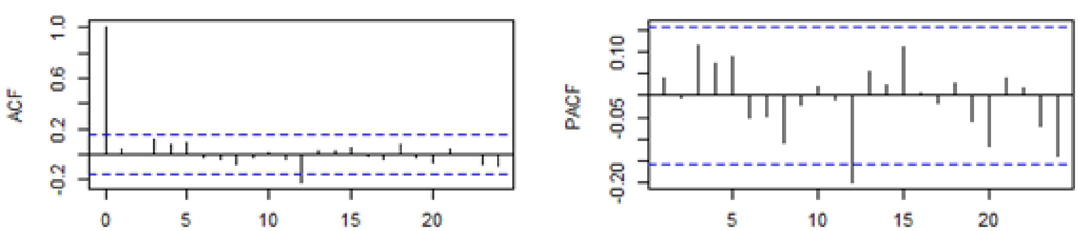

Figure 8. Plotting the ACF and PACF functions for a filtered time series after stabilization.

The suitability of the Wavelet_SARIMA$(0,1,0)(1,0,2)_{12}$ model was confirmed by examining the model residuals. The graphs of both the autocorrelation and partial autocorrelation functions of the residuals showed that most of the autocorrelation and partial autocorrelation values fell within the confidence interval. Furthermore, the following table (10) shows the results of the Ljung_Box test for several lagging data points, with (p-values $>$ 0.05) for

Table 9. Shows the results of estimating the best model for the filtered inflation series

model	Estimate	Std. Error	T_ statistical	P-Value
SARIMA(0,1,0)(1,0,2)$_{12}$	SAR12= -0.7858	0.08364	-9.3939	0.0000
	SMA12= 0.6231	0.13127	4.7463	0.0000
	SMA24= -0.3769	0.09681	-3.8937	0.0000

all lagging data points. This supports the null hypothesis, which states that there is no autocorrelation of the residuals. The Breusch-Pagan test was also performed to confirm homoscedasticity, and the results showed that the variance was homoscedastic (p-value = $0.3958 > 0.05$). These results confirm the quality and suitability of the Wavelet_SARIMA(0,1,0)(1,0,2)$_{12}$ model for estimating the filtered inflation series data.

Table 10. shows the results of the Ljung_Box test for the Wavelet_SARIMA$(0,1,0)(1,0,2)_{12}$ model residuals

Lag	Statistic	P_value
6	5.0695	0.5349
12	7.4181	0.8288
18	9.4754	0.9477
24	13.2028	0.9627
30	16.2371	0.9806
36	23.8815	0.9393

After confirming the suitability of the Wavelet_SARIMA(0,1,0)(1,0,2)$_{12}$ model for the filtered inflation series, the model was improved by adding the external variable representing the filtered oil price series after taking the first equilibrium difference to measure the impact of oil prices on inflation. Using the statistical criteria (AIC, AICc, BIC) to determine the best model, ARIMAX, and testing several models, it was concluded that the best model was Wavelet_SARIMA(0,1,0)(1,0,2)$_{12}$, as it possessed the lowest values for the information criteria (AIC = 264.7977, AICc = 265.2004, BIC = 280.0149). Furthermore, the model estimation results showed that all parameters were statistically significant at the 5% significance level. Table (11) below shows the parameter estimates, their standard errors, and the corresponding p-values.

Table 11. Shows the results of estimating the Wavelet_SARIMAX $(0,1,0)(1,0,2)_{12}$ coefficients

Model	Estimate	Std. Error	T_ statistical	P-Value
Wavelet_SARIMAX(0,1,0)(1,0,2)$_{12}$	SAR12= -0.8115	0.0826	-9.8286	0.0000
	SMA12= 0.5968	0.1682	3.5476	0.0005
	SMA24= -0.4032	0.1058	-3.8124	0.0002
	Xreg= 0.0280	0.0071	3.9548	0.0001

The regression equation can be expressed as follows:

$$y_t = 0.0280x_t + \omega_t$$

$$\omega_t = -0.8115y_{t-12} + e_t - 0.5968e_{t-12} + 0.4032e_{t-24} \tag{11}$$

To confirm the validity of the Wavelet_SARIMAX(0,1,0)(1,0,2)$_{12}$ model, both the autocorrelation function and the partial autocorrelation function for the residuals were plotted, and it was found that most of the autocorrelation values for the residuals fell within the confidence limits. In addition, the Ljung_Box test estimate for a number of lagging events, as shown in the following table (10), showed that the p-values corresponding to the test values for all lagging events were greater than the significance level of 0.05, which is another indicator of the randomness of the time series and the efficiency of this Wavelet_SARIMAX(0,1,0)(1,0,2)$_{12}$ model. The Breusch-Pagan test was also performed to confirm homoscedasticity, and the results showed that the variance was homoscedastic (p-value = 0.3073 $>$ 0.05), thus confirming the efficiency and stability of the Wavelet_SARIMAX(0,1,0)(1,0,2)$_{12}$ model.

Table 12. Results of the Ljung_Box test for the Wavelet_SARIMAX $(0,1,0)(1,0,2)_{12}$ model

Lag	Statistic	P_value
6	6.4105	0.3788
12	9.5023	0.6595
18	13.3492	0.7704
24	20.0201	0.6956
30	24.7691	0.7362
36	30.5498	0.7252

After the Wavelet_SARIMAX (0,1,0)(1,0,2)$_{12}$ model passed all diagnostic tests, it was used to predict inflation values. Figure (8) below shows the actual (solid blue line) and predicted (dashed red line) values for both the training period (2010-2022) and the test period (2023-2024). The green shaded area represents the 95% confidence intervals for the predictions in the test period, while the vertical gray line indicates the dividing line between the training and test periods. Visual observation of the figure reveals that in the training period (2010-2022), the predicted values show a high degree of convergence with the actual values, indicating the model's ability to simulate the series' behavior during the estimation period. In the test period (2023-2024), the results show that the predicted values follow the general trend of the actual values, with most actual values falling within the 95% confidence intervals, confirming the accuracy of the predictions and the model's goodness in out-of-sample forecasting. For the computational evaluation of the model's performance, predictive accuracy metrics were calculated for both the training and testing periods for both the Wavelet_SARIMA(0,1,0)(1,0,2)$_{12}$ and Wavelet_SARIMAX(0,1,0)(1,0,2)$_{12}$ models., as shown in table (13) below.

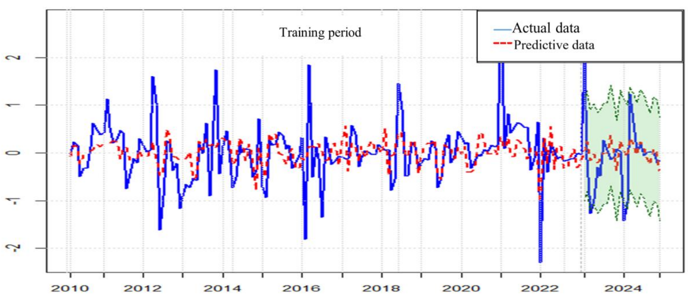

Figure 9. Actual and predicted values according to the Wavelet_SARIMAX (0,1,0)(1,0,2)$_{12}$ model.

Table 13. Results of the prediction accuracy criteria for the Wavelet SARIMAX$(0,1,0)(1,0,2)_{12}$ model and Wavelet SARIMA$(0,1,0)(1,0,2)_{12}$

Data	Wavelet_**SARIMAX(0,1,0)(1,0,2)**$_{12}$		Wavelet_**SARIMA(0,1,0)(1,0,2)**$_{12}$	
	RMSE	MAE	RMSE	MAE
Training phase	0.52731	0.37198	0.55183	0.37209
Testing phase	0.67255	0.47029	0.67641	0.47072

From the table above, we find that the RMSE and MAE values are relatively low compared to the data size. The RMSE was 0.52731 in the training phase and 0.67255 in the testing phase, while the MAE was 0.37198 and 0.47029, respectively. This indicates that the prediction errors are small.

7. Conclusion

1. The results obtained in this research show that the inflation and oil price series in Iraq are characterized by high volatility and clear instability, as demonstrated by stability tests. This reflects the sensitivity of the Iraqi economy to external shocks, especially those related to the oil market. The results proved that applying the wavelet transformation using the Daubechies wavelet was effective in addressing these volatility and removing noise. The prediction accuracy improved significantly after filtering compared to the raw data.
2. The results of the Granger causality test and the cross-correlation functions showed a relationship between the study variables, and the wavelet transformation preserved the relationship.
3. When comparing the models presented in the study, the Wavelet-ARIMAX hybrid model was found to be superior due to having the lowest values for the prediction criteria (RMSE and MAE).
4. The most suitable level for the data filtering was level three, as higher levels were found to weaken the relationship between the two study variables. The strict SURE method with a Soft threshold was also found to be the most effective for noise reduction.

8. Recommendations

1. Given the findings of this research, it is recommended to filter the economic data before conducting the analysis and to use the Wavelet-ARIMAX model to predict future values of the economic data.
2. It was found that using high values in wavelet analysis weakens the relationships between the study variables; therefore, high values should be avoided.
3. To develop more effective monetary and economic policies based on accurate inflation forecasts for Iraq, the Central Bank of Iraq and the Ministry of Planning are recommended to utilize these findings. Furthermore, researchers interested in time series analysis are advised to apply the hybrid methodology (Wavelet-ARIMAX) model to other countries with economic conditions similar to Iraq in order to generalize and compare the results.

REFERENCES

1. A. Aboud and A. Al-Mashat Al-Sadiq, *Using ARIMA models to predict economic inflation in the Libyan economy for the period (2016–2025)*, Economic Horizons Journal, vol. 4, no. 8, pp. 1–25, 2018.
2. N.W. Jassam and J.K. Khudair, *Using the DISO two-input transform function model in financial time series*, Journal of the College of Basic Education, vol. 30, no. 126, pp. 192–212, 2024.
3. N. Majid and B.E. Ahmed, *Sales forecasting using the Box–Jenkins methodology in service institutions: a case study of the Algerian insurance company CAAT - Adrar agency*, Journal of Knowledge Groups, vol. 6, no. 1, pp. 156–169, 2020.
4. I.E. Mokrani and A. Sharabi, *Sales forecasting using the Box–Jenkins methodology: a case study of Safely company*, Journal of Human Sciences, vol. 26, no. 2, pp. 241–262, 2015.
5. B.H. Youssef, *A comparative study between the Box–Jenkins methodology and the MLP neural network model in predicting sales volume in economic enterprises*, Master's thesis, University Center Maghnia, Algeria, 2024.
6. W. Bechichi, S. Majlakh, and H. Baali, *Using ARIMA models to predict the exchange rate of the dollar against the Algerian dinar*, Algerian Journal of Economic Development, vol. 5, no. 2, pp. 112–127, 2018.
7. A. Suhaila and J. Khanchour, *Time series modeling of Al Rajhi bank share prices using the Box–Jenkins methodology*, Journal of Human Sciences, no. 48, pp. 59–74, 2017.
8. M.S.A.Q. Agha and R. Zadeh, *Using time series models to predict stock prices in the Saudi stock market*, Journal of Science and Technology, vol. 2, no. 4, pp. 88–107, 2017.
9. Y.F. Al-Rahim and A.M. Sulaiman, *Audio signal purification using wavelet transformation*, Al-Rafidain Journal of Computer Science and Mathematics, vol. 9, no. 2, pp. 143–157, 2012.

10. H.M.M. Nassar, *Time series prediction using the ARIMA-SVR hybrid model based on discrete waveform transform*, Scientific Journal of Trade and Finance, vol. 43, no. 3, pp. 1043–1086, 2023.
11. S.A. Bandar, *Measuring the efficiency of wavelet analysis filters in time series analysis*, PhD thesis, Benha University, Egypt, 2019.
12. M. Coker, *Short-term inflation forecasting in Sierra Leone: a comparison of VAR(p), ARIMAX, and ARIMA models*, International Journal of Scientific Research and Management, vol. 13, no. 5, pp. 9090–9111, 2025.
13. T.H. Ali, A.A. Hamad, S.H. Mahmood, and K.H. Ahmed, *ARIMAX time series analysis for a general budget in the Kurdistan Region of Iraq using wavelet shrinkage*, Communications in Statistics: Case Studies, Data Analysis and Applications, vol. 11, no. 2, pp. 164–188, 2025.
14. T.H. Ali, H.A.A. Hayawi, and H.A. Hamza, *Bayesian time series modelling with wavelet analysis for forecasting monthly inflation*, Iraqi Journal of Statistical Sciences, vol. 22, no. 1, pp. 181–194, 2025.
15. S.K. Othman, B.M. Abdulkareem, and M.A. Badal, *Advanced ARIMAX modeling for OPEC oil forecasting using multivariate wavelet techniques*, International Journal of Environmental Sciences, vol. 11, no. 15s, pp. 2313–2326, 2025.
16. N.M. Ahmed and Q.M. Abdulqader, *Enhancing the ARIMAX model by using bivariate wavelet denoising: application on road traffic accidents data*, Iraqi Journal of Statistical Sciences, vol. 20, no. 2, pp. 15–29, 2023.
17. E.A. Haydier, N.H.S. Albarwari, and T.H. Ali, *The comparison between VAR and ARIMAX time series models in forecasting*, Iraqi Journal of Statistical Sciences, vol. 20, no. 2, pp. 247–260, 2023.
18. H.A.A. Hayawi, N.S. Ibrahim, and L.J. Mohammed, *Using the fuzzy technique to identification stochastic linear dynamic systems*, Journal of Statistics and Management Systems, vol. 24, no. 1, pp. 167–181, 2021.
19. H.A.A. Hayawi and N.S. Ibrahim, *Analysis of multivariate time series for some linear models by using multi-dimensional wavelet shrinkage*, Periodicals of Engineering and Natural Sciences, vol. 10, no. 4, 2022.
20. F. Islam and M.A. Imteaz, *The effectiveness of ARIMAX model for prediction of summer rainfall in northwest Western Australia*, IOP Conference Series: Materials Science and Engineering, vol. 1067, no. 1, pp. 1–11, 2021.
21. A.D. Bolanle and A. Oluwadare, *ARIMA and ARIMX stochastic models for fertility in Nigeria*, International Journal of Mathematics and Computer Applications Research, vol. 7, no. 5, pp. 1–20, 2017.
22. R. Amelia et al., *ARIMAX model for rainfall forecasting in Pangkalpinang, Indonesia*, IOP Conference Series: Earth and Environmental Science, vol. 926, no. 1, pp. 1–6, 2021.
23. W.N. Probst, V. Stelzenmüller, and H. Fock, *Using cross-correlations to assess the relationship between time-lagged pressure and state indicators*, ICES Journal of Marine Science, vol. 69, no. 4, pp. 670–681, 2012.
24. S. Ray and B. Bhattacharyya, *Statistical modeling and forecasting of ARIMA and ARIMAX models for food grains production in India*, Journal of Experimental Biology and Agricultural Sciences, vol. 8, no. 3, pp. 296–309, 2020.
25. C.I. Ugoh, C.A. Uzuke, and D.O. Ugoh, *Application of ARIMAX model on forecasting Nigeria's GDP*, American Journal of Theoretical and Applied Statistics, vol. 10, no. 5, pp. 216–225, 2021.
26. G. Respondek and A. Temming, *Variability of fishing effort for the German brown shrimp fishing fleet*, ICES Journal of Marine Science, vol. 71, no. 7, pp. 1805–1817, 2014.
27. I. Daubechies, *Ten lectures on wavelets*, Springer-Verlag, 1992.
28. R. Aggarwal et al., *Noise reduction of speech signal using wavelet transform with modified universal threshold*, International Journal of Computer Applications, vol. 20, no. 5, pp. 14–19, 2011.
29. W.O. Jarrett, *Congestion detection within multi-service TCP/IP networks using wavelets*, PhD thesis, University of London, 2004.
30. R.J.E. Merry, *Wavelet theory and applications: a literature study*, DCT Report 053, Eindhoven University of Technology, pp. 1–41, 2005.
31. A.C.H. Rowe and P.C. Abbott, *Daubechies wavelets and Mathematica*, American Institute of Physics, vol. 9, no. 6, pp. 635–648, 1995.