ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 4, Issue 1, April 2024, pp. 51-57

Hybrid MODWT-ARMA Model for Indonesia Stock Exchange LQ45 Index Forecasting

Hermansah [a,1,*]

[a] Department of Mathematics Education, Universitas Riau Kepulauan, Batam, Indonesia
[1] hermansah@fkip.unrika.ac.id *
* Corresponding author

ARTICLE INFO

Keywords
time series
LQ45
forecasting
DWT
MODWT-ARMA

ABSTRACT

This research discussed a hybrid Maximal Overlap Discrete Wavelet Transform (MODWT)-Autoregressive Moving Average (ARMA) model by combining the MODWT and the ARMA models to deal with the nonstationary and long-range dependence (LRD) time series. Theoretically, the details series obtained by MODWT are stationary and short-range dependent (SRD). Then, the general form of the MODWT-ARMA model was derived. In the experimental study, the daily Indonesia stock exchange LQ45 index time series was used to assess the performance of the hybrid model. Finally, the Mean Squared Error (MSE) and Mean Absolute Percent Error (MAPE) comparison with DWT-ARMA, ARIMA, and exponential smoothing models indicates that this combined model effectively improves forecasting accuracy. Based on the result of the analysis, the score of MSE of the MODWT-ARMA model was 51.42533, the score of the DWT-ARMA model was 180.1799, the score of the ARIMA model was 168.7863, and the score of the exponential smoothing model was 168.7824. At the same time, the score of MAPE in the MODWT-ARMA model was 0.00580797, the score of the DWT-ARMA model was 0.01106721, the score of the ARIMA model was 0.01070074, and the score of the exponential smoothing model was 0.01069591.

1. Introduction

In statistics and signal processing, a time series is a series of data in the form of observations measured over a period of time. Time series analysis is a method that studies time series, both in terms of theory and forecasting [1], [2]. The method frequently used in time series analysis is the Autoregressive Moving Average (ARMA). This method represents a stationary time series. Stationary means there is no growth or decline in the data. In other words, fluctuations in data are around a constant average value, not dependent on time and variance. However, many time series data are nonstationary, such as daily stock data, inflation, interest rates, rainfall, commodity exchange, and prices. Nonstationary time series must be changed to stationary data through differentiation. Differentiation is calculating changes or differences in observation values. The value

p-ISSN 2798-253X
e-ISSN 2798-3153

of the difference obtained is checked for stationarity. In reality, a lot of data has very high volatility and is difficult to stationer. In this case, the ARMA method is deemed unsuitable [3], [4].

Another method often used in time series analysis is Fourier transformation, a nonparametric method based on the frequency domain. Often, information that cannot be seen in the time domain can be seen in the frequency domain, for example, Electrocardiography (ECG) in the medical field. However, Fourier transforms have drawbacks, namely, being unable to simultaneously represent time and frequency information. It causes Fourier transformation to be unsuitable for analyzing nonstationary data. Another approach was developed to overcome the weaknesses in signal processing, namely by wavelet transformation [5], [6].

Wavelet transforms can represent time and frequency information simultaneously; therefore, they can be used to analyze nonstationary data. Wavelet is a relatively newly developed concept. The word "wavelet," given by Jean Morlet and Alex Grossmann at the beginning of the 1980s, comes from the French "ondelette," meaning small waves. The word "onde" was then translated into English into a "wave," then it was combined with the original word so that a new word, "wavelet," was formed. A wavelet function is defined as a mathematical function with certain properties, including oscillating around zero (such as sine and cosine functions) and localized in the time domain, meaning that when the domain value is relatively large, the wavelet function is zero [7].

Wavelet transform is divided into two parts, namely Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). In DWT, it is assumed that the sample size N can be divided into 2J for a positive integer J. A new concept has been developed to overcome the limitations of DWT in the sample size; this concept is known as the Maximal Overlap Discrete Wavelet Transform (MODWT). MODWT has advantages over DWT. It can be used for each sample size and reduces data to half (downsampling) so that in each decomposition level, there are wavelet and scale coefficients as much as data length [8]–[10].

The application of the wavelet method, specifically the hybrid MODWT-ARMA model, to analyze time series data was discussed in this research. In this case, MODWT was used to decompose time series data into a different scale level at each level. The decomposition produced MODWT coefficients, namely the coefficient of wavelets and scales. The determination of the coefficient of wavelets and scale was calculated using an algorithm called the pyramid algorithm. Prior to this calculation, the wavelet filter and scale filter to be used must be determined first. The wavelet coefficients (detail) and scale coefficients (smooth) cannot be directly utilized to estimate the time series model. The model estimation can be done using the ARMA process. The final results of time series data forecasting were obtained from a combination of detail and smooth forecast values [11].

Furthermore, the hybrid MODWT-ARMA model was applied to this research based on real data, namely the daily Indonesia stock exchange LQ45 index time series. The LQ45 index data is assumed to be suitable due to the assumption that the data is not stationary. Hence, it is in accordance with the research purpose, namely modeling nonstationary time series data. Then, the forecasting accuracy was measured using the Mean Squared Error (MSE) and Mean Absolute Percent Error (MAPE) values. In addition, the performance of the hybrid MODWT-ARMA model was compared to several methods, namely the hybrid DWT-ARMA model described in Paul and Anjoy (2018) [12], the ARIMA model described in Hyndman and Khandakar (2008) [13], and the exponential smoothing model described in Hyndman et al. (2002) [14].

2. Method

2.1. Maximal Overlap Discrete Wavelet Transform (MODWT)

The MODWT is a modified version of the discrete wavelet transform (DWT) [15]. Both DWT and MODWT allow for a multi-resolution analysis, which is a scale-based additive decomposition. The MODWT definition is obtained directly from the DWT: let be, $\{h_{j,l}\}$ is the DWT wavelet filter

p-ISSN 2798-253X
e-ISSN 2798-3153

and $\{g_{j,l}\}$ is the scaling filter, with $l = 1, \dots, L$ is the length of the filter and jth level of decomposition. The MODWT wavelet $\{\tilde{h}_{j,l}\}$ and scaling $\{\tilde{g}_{j,l}\}$ filters are directly defined as follows.

$\tilde{h}_{j,l} = h_{j,l}/2^{j/2}$ and $\tilde{g}_{j,l} = g_{j,l}/2^{j/2}$.

Then, the MODWT wavelet coefficients of level j is defined as the convolution of the time series $X = \{X_t, t = 0, \dots, N-1\}$ and the MODWT filters:

$$\widetilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l}\, X_{t-l \bmod N} \quad (1)$$

$$\tilde{V}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l}\, X_{t-l \bmod N} \quad (2)$$

where $L_j = (2^j - 1)(L-1) + 1$. Note that from the above expressions, the MODWT wavelet coefficients at each scale will have the same length as the original signal X. Now, they can be expressed in matrix notation as follows.

$\widetilde{W}_j = \tilde{\omega}_j\, X$ and $\tilde{V}_j = \tilde{v}_j\, X$

where each row of the $N \times N$ matrix $\tilde{\omega}_j$ has values dictated by $\{\tilde{h}_{j,l}\}$. While $\tilde{v}_j$ has values dictated by $\{\tilde{g}_{j,l}\}$. Then, the original time series X can be recovered from its MODWT via,

$$X = \sum_{j=1}^{J} \tilde{\omega}_j^T\, \widetilde{W}_j + \tilde{v}_J^T\, \tilde{V}_J = \sum_{j=1}^{J} \widetilde{D}_j + \tilde{S}_J. \quad (3)$$

It defines a MODWT-based multiresolution analysis (MRA) of X in terms if jth level MODWT details $\widetilde{D}_j = \tilde{\omega}_j^T\, \widetilde{W}_j$ and the j level MODWT smooth $\tilde{S}_J = \tilde{v}_J^T\, \tilde{V}_J$ [16].

2.2. MODWT-ARMA Model

By using the MODWT, a discrete time series $\{X_t, t = 1, 2, \dots, N\}$ can be written with the following form:

$$X_t = \tilde{S}_{J_0,t} + \sum_{j=0}^{J_0} \widetilde{D}_{j,t}, t = 1, 2, \dots, N. \quad (4)$$

The first part $\tilde{S}_{J_0} = \{\tilde{S}_{J_0,t}, t = 1, 2, \dots, N\}$ presents the tendency of series and is characterized by slow dynamics. Meanwhile, the second part components $\widetilde{D}_j = \{\widetilde{D}_{j,t}, t = 1, 2, \dots, N\}$, $j = 1, 2, \dots, J_0$, present the local details of the series X_t and is characterized by fast dynamics especially for low levels [17].

To compute the predicted value $\hat{X}_{N+h}$ of X_t, it suffices then to do this for $\tilde{S}_{J_0}$ and the $\widetilde{D}_j$, i.e, to evaluate $\hat{\tilde{S}}_{J_0,N+h}$ and $\hat{\widetilde{D}}_{j,N+h}, j = 1, 2, \dots, J_0$. To do this, let us write:

$$\hat{\tilde{S}}_{J_0,N+h} = f_0(\tilde{S}_{J_0,N}, \tilde{S}_{J_0,N-1}, \dots, \tilde{S}_{J_0,N-p_0})$$

and similarly

$$\hat{\widetilde{D}}_{j,N+h} = f_j\left(\widetilde{D}_{j,N}, \widetilde{D}_{j,N-1}, \dots, \widetilde{D}_{j,N-p_j}\right), j = 1, 2, \dots, J_0$$

where $f_j (j = 0, 1, \dots, J_0)$ is the estimator, each estimator f_j may have its proper order p_j. The choice of $f_0, f_1, \dots, f_{J_0}$ is related to the dynamic behavior of the series to be predicted. Here, we concerned with the linear ARMA model.

The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an ARMA (p, q) process if $\{X_t\}$ is stationary and if for every t,

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \quad (5)$$

where $\{e_t\}$ is a white noise with mean 0 and variance σ^2. We say that $\{X_t\}$ is an ARMA (p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA (p, q) process. Equation (5) can be written symbolically in the more compact form as in (6).

$$\varphi(B)\,X_t = \theta(B)\,e_t, t \in \mathbb{Z} \tag{6}$$

where φ and θ are the pth and qth degree polynomials.

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \tag{7}$$

and

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \tag{8}$$

and B is the backward shift operator defined by:

$$B^j X_t = X_{t-j}, j \in \mathbb{Z}. \tag{9}$$

The polynomials φ and θ will be referred to as the autoregressive and moving average polynomials of the difference (6), respectively.

Based on above theory, the tendency and details can be approximated as following forms.

$$\hat{\tilde{S}}_{J_0,N} = \varphi_1 \tilde{S}_{J_0,N-1} + \cdots + \varphi_{p_0} \tilde{S}_{J_0,N-p_0} + e_N + \theta_1 e_{N-1} + \cdots + \theta_{q_0} e_{N-q_0} \tag{10}$$

and

$$\hat{\widetilde{D}}_{j,N} = \varphi_{j1} \widetilde{D}_{j,N-1} + \cdots + \varphi_{jp_j} \widetilde{D}_{j,N-p_j} + e_N + \theta_{j1} e_{N-1} + \cdots + \theta_{jq_j} e_{N-q_j}. \tag{11}$$

By using the notation above, (8) and (9) can be written as follows.

$$\hat{\tilde{S}}_{J_0,N} = [\varphi(B) - 1]\tilde{S}_{J_0,N} + \theta(B) e_N \tag{12}$$

and

$$\hat{\widetilde{D}}_{j,N} = \left[\varphi_j(B) - 1\right]\widetilde{D}_{j,N} + \theta_j(B) e_N. \tag{13}$$

Hence, the MODWT-ARMA prediction model is as follows.

$$\hat{X}_{N+h} = \hat{\tilde{S}}_{J_0,N+h} + \sum_{j=0}^{J_0} \hat{\widetilde{D}}_{j,N+h}$$

$$\hat{X}_{N+h} = [\varphi(B) - 1]\tilde{S}_{J_0,N+h} + \theta(B) e_{N+h} + \sum_{j=0}^{J_0}\left(\left[\varphi_j(B) - 1\right]\widetilde{D}_{j,N+h} + \theta_j(B) e_{N+h}\right). \tag{14}$$

The flowchart for the hybrid MODWT-ARMA modeling algorithm is presented in Fig. 1.

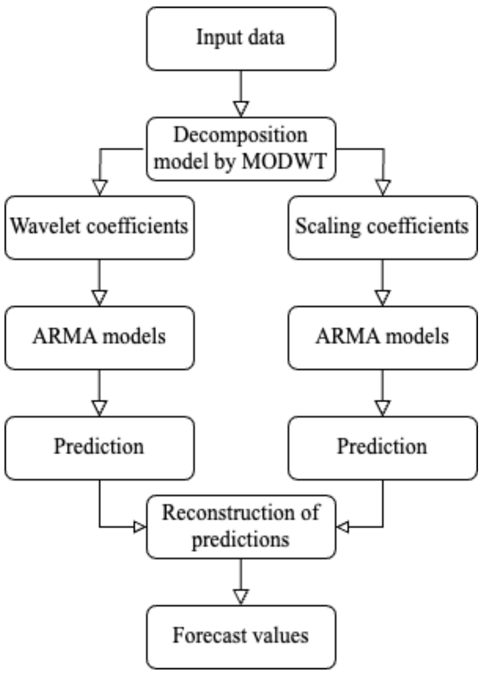

Fig. 1 Flowchart of hybridization of MODWT and ARMA.

p-ISSN 2798-253X
e-ISSN 2798-3153

3. Results and Discussion

The research data used for modeling was the LQ45 index data obtained from finance.yahoo.com. The data was the daily closing stock price collected from January 2, 2015, to January 9, 2019, which amounted to 974 data. The LQ45 index data was divided into training and testing data. The training data, consisting of 779 data, was used to construct a model, while the testing data, consisting of 195 data, was utilized to evaluate the model.

LQ45 is a stock market index for the Indonesia Stock Exchange (IDX) (formerly known as the Jakarta Stock Exchange). It consists of 45 companies. The LQ45 index data was used as research data because the assumption was that the data was not stationary; hence, it aligned with the research's purpose of modeling nonstationary time series data.

The initial modeling procedure used the MODWT-ARMA to check data stationarity. Using the R program, the ADF test statistic obtained was that the p-value = 0.6506, which was greater than the one used, which is 0.05. It can be concluded that the data is not stationary. Decomposition with MODWT was done to station the series details so that the results were stationary. In this case, the wavelet family was Daubechies 6 with six levels. The detail and smooth values obtained were $\widetilde{D}_1$, $\widetilde{D}_2$, $\widetilde{D}_3$, $\widetilde{D}_4$, $\widetilde{D}_5$, $\widetilde{D}_6$ and $\tilde{S}_6$. The original data and the decomposed data using MODWT had the same value, namely:

$$X_t = \widetilde{D}_1 + \widetilde{D}_2 + \widetilde{D}_3 + \widetilde{D}_4 + \widetilde{D}_5 + \widetilde{D}_6 + \tilde{S}_6$$

Subsequently, the stationarity of the data from each of the decomposition results was examined using a hypothesis; namely, there is a root unit (data is not stationary). The ADF test statistic obtained was that all decomposition results have a p-value smaller than α, which was 0.05. Thus, it is concluded H_0 is rejected and the decomposition data is stationary. Furthermore, Detail 1 ($\widetilde{D}_1$) forecasting with ARMA can be done using syntax auto.arima. The best model was then obtained, namely ARMA (0,5). Next step was forecasting with ARMA (0,5). Analog for Detail 2 ($\widetilde{D}_2$) forecasting used the best model ARMA (3,0), Detail 3 ($\widetilde{D}_3$) used the best model ARMA (3,2), Detail 4 ($\widetilde{D}_4$) used the best model ARMA (1,0), Detail 5 ($\widetilde{D}_5$) used the best model ARMA (0,4), Detail 6 ($\widetilde{D}_6$) used the best model ARMA (1,0), and Smooth 6 ($\tilde{S}_6$) used the best ARMA model (2,0).

With the "fitted" syntax using the R program, the sum of the simulation values for each decomposition is obtained, visually seen in the plot of the training data with the simulation data in Fig. 2. In Fig. 2, the blue line indicates the in-sample prediction (simulated data), while the black line shows the original data (real data). The graph shows that the in-sample prediction value (simulated data) moves to resemble the pattern of the original data (real data) with an MSE value of 10.62197 and a MAPE of 0.002432078. Meanwhile, the red line in Fig. 2 results from the out-sample prediction (predicted data). The pattern shown from the out-sample prediction data graph (predicted data) almost resembles the original data pattern (real data). The out-sample prediction graph (predicted data) had an MSE value of 51.42533 and a MAPE of 0.00580797.

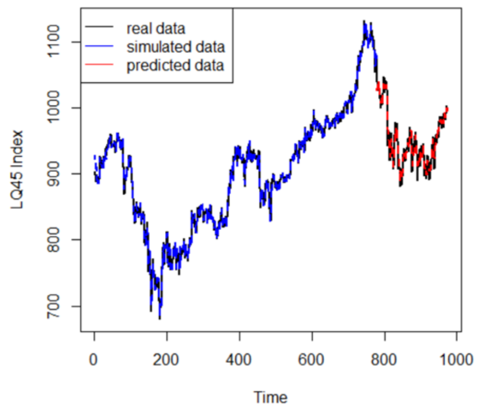

Fig. 2 LQ45 index time series and its simulations and predictions using MODWT-ARMA model.

p-ISSN 2798-253X
e-ISSN 2798-3153

Furthermore, (15) was used to obtain the LQ45 index stock price data forecast.

$$\hat{X}_{N+1} = \hat{\hat{S}}_{6,N+1} + \sum_{j=0}^{6} \widehat{\widehat{D}}_{j,N+1} \tag{15}$$

The results of the visual forecast and the visual testing are shown in Fig. 2. In order to show the proposed method's accuracy, the comparison of MSE and MAPE obtained by MODWT-ARMA model, DWT-ARMA model, ARIMA model and exponential smoothing model are presented in Table 1.

Table 1. Prediction Accuracy Comparison of MODWT-ARMA Model, DWT-ARMA Model, ARIMA Model and Exponential Smoothing Model of the Indonesia Stock Exchange LQ45

Prediction Method	MODWT-ARMA	DWT-ARMA	ARIMA	Exponential Smoothing
MSE simulated	10.62197	117.9819	91.49351	91.49268
MAPE simulated	0.002432078	0.009000032	0.007911403	0.007910197
MSE predicted	51.42533	180.1799	168.7863	168.7824
MAPE predicted	0.00580797	0.01106721	0.01070074	0.01069591

4. Conclusion

MODWT is a wavelet transformation modified from DWT. The sample observed for DWT can only be expressed in the form of 2J with J positive integers, while MODWT can be used for each sample size. Another advantage of MODWT is that it is able to reduce data to half so that for each level of decomposition there are wavelet coefficients (detail) and scale coefficients (smooth) as much as data lengths. Maximal MODWT-ARMA is related to nonstationary time series data. Stationary detail values are obtained by decomposing nonstationary data. It causes the results of decomposition can be predicted directly with ARMA. Based on the case study of the LQ45 index data, the MSE and MAPE were smaller than the DWT-ARMA model, the ARIMA model, and the exponential smoothing model. Based on the result of the analysis, the score of MSE of MODWT-ARMA model was 51.42533, the score of DWT-ARMA model was 180.1799, the score of ARIMA model was 168.7863, and the score of exponential smoothing model was 168.7824. At the same time, the score of MAPE in the MODWT-ARMA model was 0.00580797, the score of DWT-ARMA model was 0.01106721, the score of ARIMA model was 0.01070074, and the score of exponential smoothing model was 0.01069591. The MODWT-ARMA model yielded the lower MSE and MAPE values. It indicates that the hybrid MODWT-ARMA model is effective to increase the accuracy of forecasting.

References

[1] G.E.P. Box and G.M. Jenkins, *Time Series Analysis Forecasting and Control*. California, USA: Holden-Day, 1970.

[2] W.W.S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. Boston, USA: Pearson Addison Wesley, 2006.

[3] P.J. Brockwell and A.R. Davis, *Time Series Theory and Methods*. New York, USA: Springer Verlag, 1991.

[4] D. Rosadi, *Analisis Runtun Waktu dan Aplikasinya dengan R*. Yogyakarta, Indonesia: Gadjah Mada University Press, 2014.

[5] G.P. Nason, *Wavelet Methods in Statistics with R*. New York, USA: Springer, 2008.

[6] D.B. Percival and A.T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, United Kingdom: Cambridge University Press, 2000.

[7] J.S. Walker, *A Primer on Wavelets and their Scientific Applications*, 2nd ed. New York, USA: Taylor and Francis Group, 2008.

[8] R.K. Paul, Prajneshu, and H. Ghosh, "Wavelet Frequency Domain Approach for Modelling and Forecasting of Indian Monsoon Rainfall Time Series Data," *Journal of The Indian Society of Agricultural Statistics*, Vol. 67, No. 3, pp. 319–327, 2013.

[9] O. Renaud, J.L. Starck, and F. Murtagh, “Prediction Based on a Multiscale Decomposition,” *International Journal of Wavelet, Multiresolution and Information Processing*, Vol. 1, No. 2, pp. 217–232, 2003.

[10] S.A. Wadi, A. Hamarsheh, and H. Alwadi, “Maximum Overlapping Discrete Wavelet Transform in Forecasting Banking Sector,” *Applied Mathematical Sciences*, Vol. 7, No. 80, pp. 3995–4002, 2013, doi: 10.12988/ams.2013.36305.

[11] L. Zhu, Y. Wang, and Q. Fan, “MODWT-ARMA Model for Time Series Prediction,” *Journal of Applied Mathematical Modelling,* Vol. 38, No. 5–6, pp. 1859–1865, Mar. 2014, doi: 10.1016/j.apm.2013.10.002.

[12] R.K.A. Paul and P. Anjoy, “Modeling Fractionally Integrated Maximum Temperature Series in India in Presence of Structural Break,” *Theory and Applied Climatology*, Vol. 134, pp. 241–249, 2018, doi: 10.1007/s00704-017-2271-x.

[13] R.J. Hyndman and Y. Khandakar, “Automatic Time Series Forecasting: The Forecast Package for R,” *Journal of Statistical Software*, Vol. 27, No. 3, pp. 1–22, 2008, doi: 10.18637/jss.v027.i03.

[14] R.J. Hyndman, A.B. Koehler, R.D. Snyder, and S. Grose, “A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods,” *International Journal of Forecasting*, Vol. 18, No. 3, pp. 439–454, Jul.–Sep. 2002, doi: 10.1016/S0169-2070(01)00110-8.

[15] G.P. Nason and R.V. Sachs, “Wavelets in Time Series Analysis,” Philosophical Transactions: Mathematical, Physical and Engineering Sciences, Vol. 357, No. 760, pp. 2511–2526, Sep. 1999.

[16] M. Aminghafari and J.M. Poggi, “Forecasting Time Series Using Wavelets,” *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 5, No. 5, pp. 709–724, 2007, doi: 0.1142/S0219691307002002.

[17] M. Aminghafari and J.M. Poggi, “Nonstationary Time Series Forecasting Using Wavelets and Kernel Smoothing,” *Communications in Statistics-Theory and Methods*, Vol. 41, No. 3, pp. 485–499, 2012, doi: 10.1080/03610926.2010.529532.