INSOLOGI: Jurnal Sains dan Teknologi
https://journal.literasisains.id/index.php/INSOLOGI
ISSN 2828-4984 (Media Online) | ISSN 2828-4992 (Media Cetak)
Vol. 2 No. 2 (April 2023) 392-399
DOI: 10.55123/insologi.v2i2.1888
Submitted: 11-04-2023 | Accepted: 21-04-2023 | Published: 29-04-2023

# Comparison of Methods ARIMA and MAR Models with MODWT Decomposition on Non-Stationary Data

**Puce Angreni[1], Melda Juliza[2]**
[1,2]Statistics Study Program, ISTNUBA, Bali, Denpasar, Indonesia
Email: [1]puceangreni@istnuba.ac.id, [2]melda@istnuba.ac.id

***Abstract***

*The forecasting methods used in this study are Autoregressive Integrated Moving Average (ARIMA) and Multiscale Autoregressive (MAR). The ARIMA model does not include predictor variables in the model. The MAR model is a model that performs the transformation process using wavelets. The MAR model adopts an autoregressive time series (AR) model with wavelet coefficients and scale coefficients as predictors. The wavelet coefficient and scale are obtained by decomposition using Maximal Overlap Discrete Wavelet Transformation (MODWT). MODWT functions to describe data based on the level of each wavelet filter. This study aims to determine the best forecasting model using ARIMA and MAR models. The time series data used in this study is data on the rupiah exchange rate against the US dollar. Data on the rupiah exchange rate against the US Dollar for 2019-2020 is non-stationary data, so the ARIMA and MAR models can be used in this study.*

***Keywords:*** *Nonstasioner, ARIMA, MAR, MODWT*

## 1. INTRODUCTION

Each country has its currency used as an exchange tool in trading activities. The exchange rate of rupiah against foreign currencies has a positive or negative effect on the economy Increasing or decreasing the rupiah exchange rate against the US dollar will affect the cost of industrial production be it imports or exports. Prediction or forecasting is an effort to predict what will happen in the future based on past data. Time series data is a set of data obtained from the observation of a phenomenon that occurs based on a time index with a fixed or equal time interval. In the analysis of time series, the assumption that must be fulfilled from the time series data is that the data is stationary.

Research on the analysis of non-stationary time series is the transformation of wavelets by forming a Multiscale Autoregressive (MAR) model. Wavelet transformations are divided into two large sections, continuous wavelet transform (CWT) and Discrete Wavelet Transform (DWT). In DWT it is assumed that the sample size N can be divided into $2^J$ for a positive integer J. A new concept was developed in addressing the limitations of DWT in that sample size, known as Maximal Overlap Discrete Wavelet Transform (MODWT). MODWT has advantages over DWT among others, it can be used for any sample size N.

Based on the description stipulated, the study examined the role of wavelet models in which mar models are contained, to predict un stationary time series data. Therefore, the data will be processed in two ways, namely detrending (trend separation) and differencing which are further decomposed with MODWT method so that mar modeling can be done. The selected MAR model is a model that meets the assumptions of normality and white-noise. By establishing a wavelet forecasting model, forecasting over the next few periods can be applied to rupiah exchange rate data

## 2. METODH

### 2.1. Wavelet Transformation.

Wavelet transformation is one of the methods that can be performed on stationary and non-stationary time series data. This method can automatically separate a trend from a un stationary time series data. Besides, wavelet transformations can model irregular or un linear patterned data. In this method, if the data is a non-stationary time series then it can be decomposed immediately without being stationed first. But based on, the data of the stationed time series provides better results than unstationed time series data. Most non-stationary time series data is marked by a trend. Trends in the time series are two types, namely stochastic trends and deterministic trends. There are two methods for indexing time series data, including differencing and detrending. Stochastic trends are usually addressed by differentiation processes. While deterministic trends are usually overcome by doing trend separation

According to [10], e.g. a non-stationary time series $Z_t = \beta_1 + \beta_2 t + \beta_3 Z_{t-1} + e_t$. If $\beta_1 \neq 0$, $\beta_2 = 0$, $\beta_3 = 1$ then obtained a model $Z_t = \beta_1 + \beta_{t-1} + e_t$ written as

$$Z_t - Z_{t-1} = \beta_1 + e_t$$
$$\Delta Z_t = \beta_1 + e_1 \quad (1)$$

Then $Z_t$ will show a positive trend $(\beta_1 > 0)$ or negative trend $(\beta_1 < 0)$, which is called a stockastic trend.

On the differencing method, $\nabla Z_t = Z_t - Z_{t-1}$ formed as the first reference. The d-reference for time series stationation when denoted by $W_t$ , i.e. $W_t = \nabla^d Z_t$, with $d$ is an integer $d \geq 1$. A non-stationary time series is considered a deterministic trend when the mean function $(\mu_t)$ can be explained by the polynomial of the $k$-order.

$$Z_t = \sum_{j=0}^{k} \alpha_j t^j + e_t$$
$$= \mu_t + e_t \quad (2$$

The instasioneran achieved by the construction of a new time series in residuals is known as the detrending method. If the data of a stationary time series is detrending symbolized by $Y_t$ then $Y_t = Z_t - \mu_t$ where is the initial data of a time series and μt is a trend model in the form of polynomial. Stationary time series data is decomposed with MODWT method to obtain wavelet coefficients and scale coefficients, so that Multiscale Autoregressive (MAR) modeling can be performed. For forecasting purposes, the model is restored to its original form called the wavelet forecasting model. For example, a time series data isstationed by detrending and trend that has been obtained /separated $(\mu_t)$ in the form of $\alpha_1 + \alpha_2$ then after obtaining the MAR model from $Y_t$, it is then returned to the original form $Z_{t+1} = \alpha_1 + \alpha_2 t + Y_{t+1}$ where $Y_{t+1}$ contains wavelet coefficients or scale coefficients. Multiscale Autoregressive (MAR) model found in wavelet forecasting models, the process follows AR through the convergence of optimal procedures and asymptomatically will be equivalent to the best forecasting.

### *2.2.1 MAR*

MAR model is a model by performing a transformation process using a wavelet, which assumes each scale of a wavelet transformation follows an AR process. Determination of lag-lag that becomes an input variable for the MAR model uses wavelet coefficients and scale coefficients derived from wavelet transformation results. The wavelet coefficient (detail) and scale coefficient of wavelet transformation results through MODWT decomposition are considered to have an influence on predictions at the time $t+1$ will be shaped $w_{j,t-2^J(k-1)}$ and $v_{J,t-2^J(k-1)}$, or can be written as:

$$\hat{X}_{t+1} = \sum_{j=1}^{J}\sum_{k=1}^{A_j} \hat{a}_{j,k} w_{j,t-2^j(k-1)} + \sum_{k=1}^{A_j} \hat{a}_{J+1,k} v_{J,t-2^J(k-1)} + \varepsilon_t \tag{3}$$

with MODWT decomposition level ( $j=1,2,\cdots,J$ ), $A_j$ is the order of the MAR model ($k=1,2,\cdots,A_j$), $a_{j,k}, \hat{a}_{J+1,k}$ is the coefficient value of the MAR model, *t* is the time of the, $w_{j,t-2^j(k-1)}$ is a wavelet coefficient and $v_{J,t-2^J(k-1)}$ is a coefficient of scale.

The determination of mar model input in the forecasting of the $(t+1)$ data is shown in Figure 1, the first input on each scale is the t-data, and the second input on each scale is the $(t-2^j)$ data.

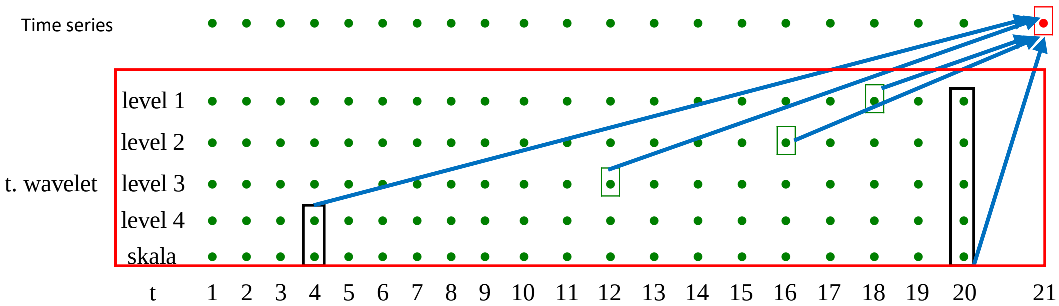

Figure 1. MAR Wavelet Modeling Illustration

Figure 1 is an illustration to predict the 21st data (t+1st data) with the MAR model of order 2 (A_j=2) and level j=4, then the input variables used are the level 1 wavelet coefficient at t=20 and t=18, the level 2 wavelet coefficient at t=20 and t=16, the level 3 wavelet coefficient at t=20 and t=12 , wavelet coefficient level 4 at t=20 and t=4, as well as scale coefficients at t=20 and t=4.

Supposing J=6 and A_j=2 MAR models based on equations (3) can be expressed as:

$$\hat{X}_{t+1} = \sum_{j=1}^{6}\sum_{k=1}^{2} \hat{a}_{j,k} w_{j,t-2^j(k-1)} + \sum_{k=1}^{2} \hat{a}_{J,k} v_{J+1,t-2^J(k-1)}$$

$$\begin{aligned} &= \hat{a}_{1,1}w_{1,t} + \hat{a}_{1,2}w_{1,t-2} + \hat{a}_{2,1}w_{2,t} + \hat{a}_{2,2}w_{2,t-4} + \hat{a}_{3,1}w_{3,t} \\ &\quad + \hat{a}_{3,2}w_{3,t-8} + \hat{a}_{4,1}w_{4,t} + \hat{a}_{4,2}w_{4,t-16} + \hat{a}_{5,1}w_{5,t} + \hat{a}_{5,2}w_{5,t-32} \\ &\quad + \hat{a}_{6,1}w_{6,t} + \hat{a}_{6,2}w_{6,t-64} + \hat{a}_{7,1}v_{6,t} + \hat{a}_{7,2}v_{6,t-64} \end{aligned} \tag{4}$$

or can be written as $\underset{\sim}{s}_1 = \mathbf{A}_1\underset{\sim}{\alpha}$,

The guessing of vector parameters can be solved by the smallest squared method that minimizes squared error, i.e.:

$$\hat{\underset{\sim}{\alpha}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\underset{\sim}{s}$$

Calculation of the alleged value of the MAR model, used α value that has been allegedly using equations. The α value and the wavelet coefficient value and scale obtained through decomposition are incorporated into the MAR model as in the equation.

### *2.1.2 MODWT*

The wavelet transformation seen as more suitable for time series data is MODWT because in each decomposition level there is a wavelet coefficient and a scale of as much data length [3]. The determination of the level (J) for MODWT decomposition depends on the width of the decomposition filter (L) and the amount of data ( n), with the formula:

$$J < \ln\left(\frac{n}{L-1}+1\right)$$

The pyramid algorithm for MODWT is a calculation algorithm to calculate the scale coefficient and wavelet coefficient of MODWT at the j-level. If a data is decompoed with a wavelet filter and a scale filter, it will produce a wavelet coefficient and a scale coefficient. Figure 2 follows a pyramid algorithm for MODWT.

### 2.1.3 The Stepwise Method in Determining the MAR Model Variables

Stepwise is one of the methods that can be used for the process of selecting the best model, so it is useful to obtain significant and appropriate variables to be used in the MAR model. This method is a combination of the forward selection method and the elimination method which are applied alternately. Every time a new variable is entered into the model, all the variables previously entered are re-checked whether they still need to be maintained or excluded, then a selection is made. But if there are no more variables that can be included or removed from the model, the stepwise procedure ends (Sembiring, 1995).

The use of the stepwise MAR model method provides optimal conditions for the lag wavelet coefficients and scaling coefficients which should be used for forecasting a time series data.

### 2.1.4 Examination of the Residual Assumptions of the MAR Model

The MAR model that has been obtained is then examined for its residual assumptions. Mathematically, the residual is defined as:

$$\varepsilon_t = Z_t - \hat{Z}_t \tag{5}$$

where:

$\varepsilon_t$: residual at time t (t = 1, 2, ..., n)
$Z_t$ : observed value t
$\hat{Z}_t$: guessed value t

There are two kinds of residual assumption checks, namely normality testing and white noise checking.

### 2.2. Data analysis method

The time series data used in this study are daily data on the Rupiah exchange rate against the US Dollar from January 2019 to February 2020, totaling 437 observations. it is non-staionarity time series data. The type of wavelet used is the Haar wavelet. The steps taken in this research are:

a) Determine the data to be analyzed, in the form of rupiah exchange rate data
b) Checking data
c) Determine the level according to the filter used, namely Haar with filter 2
d) Performed MODWT decomposition at each level
e) Selected the variables to be input model the MAR based on the significant PACF lag
f) Performed the stepwise method to obtain significant variables
g) Performed MAR modeling on wavelet coefficients and significant scales
h) Formed the MAR model
i) Checking the goodness of fit by selecting the MAR model that has the smallest RMSE.

## 3. RESULT AND DISCUSSIONS

In this study, the MAR model was applied to the daily data of rupiah exchange rate against US Dollar which is herely called Exchange Rate from January 01, 2019 – August 1, 2020, consisting of 437 observations.. The description of Exchange Rate data is used to find out an overview of the data, namely how big the average value, data spread, maximum and minimum values, and the amount of Exchange Rate data used in this study.

Table 1. Statistik Deskriptif

| Variabe l | N | Average | st. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Kurs | 437 | 14312 | 518.8394 | 13572 | 16575 |

Based on Table 1. it can be seen that the amount of Exchange Rate data used for modeling is 437 observations. The average exchange rate is 14312. For the spread of exchange rate data is 518.8394. The minimum exchange rate is 13572. While the maximum exchange rate is 16575.

The first step is to test the stationaryity of the data. Stationary testing using dickey-fuller augmented unit root test (ADF) can be seen in Table 2. Following:

Table 2. ADF Exchange Rate Data Test

| Data | Value p before differencing | Value p after differencing | Conclusion |
|---|---|---|---|
| Kurs | 0,1611 | 0,01 | Stationary after differencing to 1 |

Based on Table 2. it is known that after differencing 1 time obtained a value of p = 0.01. This means that the average awareness is achieved after diffrencing 1 time. The plot of exchange rate data after differencing 1 can be seen in Figure 1. Following:

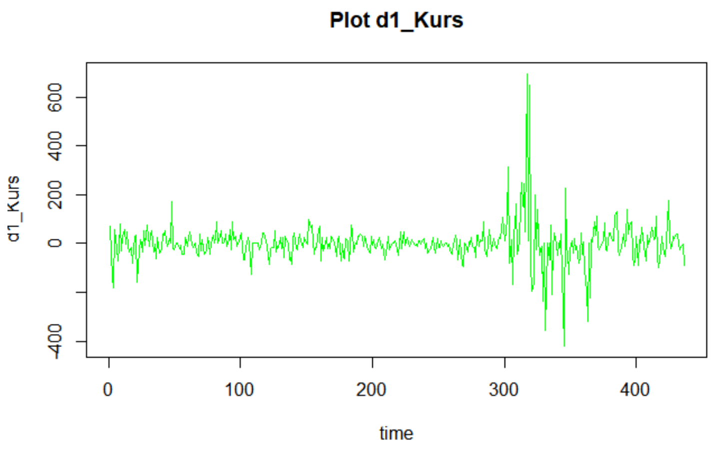


Figure 1. Plot Data Exchange Rate Differencing 1
(left) and Plot Analysis Box-Cox (right)

The Box-Cox transformation shows that the Exchange Rate data returns a value of λ=1. It can be said that the differencing exchange rate data 1 has been stationary against the variety. Thus the results of stationary examination show that the Exchange Rate at differencing 1 has been stationary against average and variety. Average stationary testing is performed by looking at critical values at α=5% compared to the statistical value of t in the Augmented Dickey-Fuller (ADF Test) test. A one-time differentiation that produces a critical value is 5% less than the ADF test statistical value, so it is concluded that the data is stationary to the average.

Table 3. RMSE

| **Data** | **Model** | **RMSE** |
|---|---|---|
| Kurs | ARIMA (0,1,1) | 76.89935 |

76.89935 → RMSE

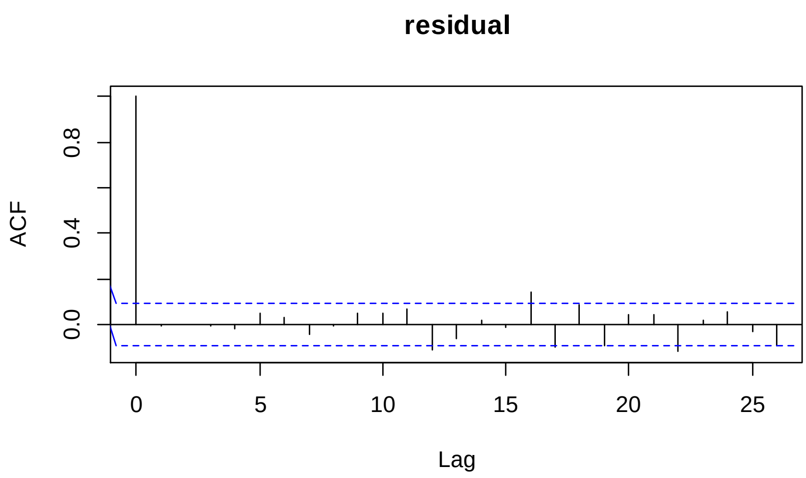


Figure 2. residual

Because there are some lags from the ACF plot of the ARIMA residuals that come out of the blue line, the residuals from ARIMA do not fulfill the nature of white noise.

***MAR Model Determination***

Significant lag-lag retrieval is significantly limited to 0.2

Table 4. MAR model inputs based on significant PACF lag-lag in wavelet coefficients and scaling coefficients.

| **Coefficient** | **Significant lag** | **n Input** |
|---|---|---|
| W1 | Lag 1, 3 | 2 |
| W2 | Lag 1, 2, 3, 4, 6 | 5 |
| W3 | Lag 1, 2, 3, 4, 5, 7, 9, 10 | 8 |
| W4 | Lag 1, 3, 5, 6, 9, 10 | 6 |
| W5 | Lag 1, 2, 3, 5, 10 | 5 |
| W6 | Lag 1, 3, 5 | 3 |
| V1 | Lag 1, 2, 3 | 3 |
| V2 | Lag 1, 3, 5, 10 | 4 |
| V3 | Lag 1, 3, 9, 10 | 4 |
| V4 | Lag 1, 2, 3, 5 | 4 |
| V5 | Lag 1, 3, 5 | 3 |
| V6 | Lag 1, 3 | 2 |

Form mar models on significant wavelet and scaling coefficients. Next, choose the forecasting model with the smallest RMSE, which is presented in Table 4. RMSE Model ARIMA and MAR Based on SIGNIFICANT PACF and Renaund Et al. Proposal. Check the accuracy of the model by selecting the MAR model that has the smallest RMSE.

Table 5. RMSE values at Levels 1-6

| **Level** | **AIC** | **RMSE** |
|---|---|---|
| 1 | 4996.024 | 76.77699* |
| 2 | 4918.201 | 76.31107 |
| 3 | 4909.316 | 74.63827 |
| 4 | 4909.522 | 74.30664* |
| 5 | 4907.36 | 73.94456 |
| 6 | 4898.652 | 73.53704 |

*) based on stepwise level 1 and level 4 results all have significant variables at a rate of significance of 5%.

Thus, the RMSE value of the MAR model level 1 is 76.77699. while the RMSE value of the MAR model level 4 is 74.30664. So, the better MAR model used is the MAR model at level 4 because it has a smaller RMSE value.

## 4. CONCLUSION

Based on the results of the analysis can be concluded that for the problem of data Exchange rate the best model used is the mar model which is best is the MAR model in accordance with renaund et al proposal with RMSE value. The RMSE MAR level 4 value is smaller than the RMSE ARIMA value, which is 74.30664 < 76.89935. And also the residual ARIMA is not white noise. So the level 4 MAR model is better to use than the ARIMA model

## 5. REFERENCES

Graps, A. 1995. An Introduction to Wavelets. *IEEE Computational Science and Engineering*. **2** (2).

Renaud, O., Starck, J. L dan Murtagh, F. 2003. Prediction Based on a Multiscale Decomposition. *International Journal of Wavelets, Multiresolution and Information Processing.* **1**(2), 217-232.

Percival, D. B. dan Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge.

Montgomery, D. C., Jennings, C. L. dan Kulahci, M. 2008. *Introduction Time Series Analysis and Forecasting*. John Wiley & Sons, Inc., New Jersey.

Cryer, J. D., dan Chan, K. S. 2008. *Time Series Analysis with Application in R.* Second Edition. Spinger Science dan Business Media., USA.

Box, G. E. P. dan Jenkins, G. M. 1994. *Time Series Analysis: Forecasting and Control.* Third Edition. Prentice Hall, New Jersey.

Makridakis, S., Wheelwright, S. C. dan McGee, V. 1999. *Metode dan Aplikasi Peramalan*. Terjemahan Hari Suminto. Edisi Revisi. Binarupa Aksara, Jakarta.

Popoola, A. O. 2007. *Fuzzy-Wavelet Method for Time Series Analysis*. Submitted for the degree of doctor of philosophy in Surey University, England.