IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 6 Issue 9, September 2019
ISSN (Online) 2348 – 7968 | Impact Factor (2019) – 5.248
www.ijiset.com

Longitudinal Data Modeling Using Multiscale Autoregresive (MAR) Wavelet

Suparti[1], Rukun Santoso[2], Alan Prahutama[3], Sudargo[4]

[1,2,3]Departement of Statistics, Faculty Science and Mathematics, Diponegoro University, Semarang, Indonesia
PGRI University Semarang, Indonesia

Corresponding Author: alan.prahutama@gmail.com

Abstract

Longitudinal data is a combination of cross section and time series data, where the subjects are independent and the time of observation of each subject is dependent. The advantages of studies on longitudinal data are knowing individual changes and requiring less subjects because of repeated observations. Moreover, the estimation is more efficient because it is done together for all subjects and observations. Some nonparametric methods for modeling longitudinal data are kernel, spline, local polynomial, fourier and wavelet. The wavelet method was developed to overcome the fourier method which is not good in modeling non-stationary data. In wavelet, Discrete Wavelet Transform (DWT) method is used. But there are limitations in DWT, because it can only be used for data modeling with the number of $N= 2^J$ with J is positive integer numbers. In addition, the number of coefficients in DWT experiencing shrinkage as the scale level increases. To overcome this problem, Maximal Overlap Discrete Wavelet Transform (MODWT) is used, which can be used for any amount of data and the number of constant coefficients is as much as the processed data. By using the MODWT coefficient as an independent variable, longitudinal data modeling will be carried out, which each subject refers to the Autoregresive (AR) model and the model is called the Multiscale Autoregresive (MAR) Model . MAR modeling of longitudinal data is then applied to the simulation data of 3 subjects. Modeling is performed by using Haar (D2) and Daubechies (D4) filters with MAR level and order are 4 and 2 respectively. The modeling generated the best model with the Haar filter because it produces a smaller residual standard error and a greater coefficient of determination (R^2).

Keywords : Longitudinal data, MAR model, residual standard error, determination coefficient.

1. Introduction

In the regression analysis there are two types of data namely time series data and cross section data. Time series data is data from a subject that is observed repeatedly over time. While the cross section data is data from several subjects which are only done one observation on each subject and are mutually independent. According to Weiss (2005), longitudinal data is a special form of data with repeated measurements. The combined time series and cross section data form longitudinal data (Wu and Zhang, 2006). Longitudinal data is data obtained from repeated observations of each subject at different time intervals. This data correlates to the same subject and is independent between different subjects. According to Wu and Zhang (2006), there are several advantages of a study of longitudinal data, including knowing individual changes, and requiring not too many subjects because of repeated observations. In addition, the estimation is more efficient because it can be done together on all subjects and all observations.

Research on longitudinal data include Hu et al (2004) using kernels. Liang and Zeger (1986) with the General Linear Model. Ibrahim and Suliadi (2008) used Smoothing Spline for longitudinal data, and Suparti, et al (2016) used local polynomials with case study modeling in 7 inflation spending groups in Indonesia. The kernel, smoothing spline and local polynomial methods are part of the nonparametric method. Other nonparametric methods that can be used for modeling longitudinal data are the Fourier and wavelet methods.

Fourier transform can detect interference, but fourier transformation has some limitations, which require stationary data in the average so that the trend must be removed first before using the fourier transformation (Popoola, 2007). In addition, the results of the analysis of the data can only provide information about frequencies. This causes fourier transforms cannot be used to analyze non-stationary data.

Wavelet transforms are developed to overcome the weaknesses of Fourier transforms. Wavelet transform is able to represent time and frequency information simultaneously. The representation of time and frequency results in

wavelet transforms that can be used to analyze non-stationary data. Wavelet is a function that mathematically cuts data into different components and studies each component with the appropriate resolution on the scale.
Wavelet transform is divided into two major parts, namely Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). In DWT it is assumed that the sample size N can be expressed as 2^J for a positive integer J. This results in not all data being processed using DWT. A new concept was developed in overcoming the limitations of DWT in the sample size, namely Maximal Overlap Discrete Wavelet Transform (MODWT). MODWT has the advantage that it can be used for any N sample size (Percival and Walden, 2000).

In this paper a study of modeling of longitudinal data using wavelets by utilizing MODWT and applying them to data on 3 inflation sectors in Indonesia, namely (1). Foodstuffs group; (2). Processed foods, beverages, cigarettes and tobacco groups; (3). Clothing group. These three expenditure groups are cases of longitudinal data.

2. Methodology of Research

Compile a longitudinal data layout according to the research case study. Forming MAR modeling using MODWT with J = 4 and Aj = 2 and estimating the parameters. Applying research data modeling to data on sectors / groups of Indonesian inflation expenditure using the Haar (D2) and Daubechies (D4) filter with the initial procedure to calculate the MODWT for data on 3 sectors of inflation spending in Indonesia. Next determine the MODWT coefficient for each inflation spending sector data as input for the independent variable in the multiple regression model. Then estimate the parameters and test the significance of the model formed and calculate the coefficient of determination of the residual error standard. Perform a comparison of MAR models with filters D2 and D4, then choose the best model based on residual error and R2 standards. All calculations use the R software.

3. Results and discussion

3.1. *Model Multiscale Autoregressive (MAR)*

In general, Multiscale Autoregressive wavelet modeling is a modeling method using wavelet transforms, in this case using MODWT. With multiscale decomposition like wavelet, there is a benefit that is obtained automatically separating data components, such as trend components and irregular components in the data. Therefore, this method can be used to make predictions on stationary and non-stationary data (Suhartono, et al, 2010).
Suppose there is a stationary signal $X = (X_1,\ X_2, \dots, X_t)$ and it is assumed to be a predicted valueX_{t+1}. The basic idea used is to use the coefficients obtained from the MODWT decomposition results $w_{j,t-2^j(k-1)}$and $v_{J,t-2^J(k-1)}$, with $k = 1,2, \dots, A_j$ and $j = 1,2, \dots, J$ (Renaud et al, 2003). Predicted model follow the Autoregressive model (AR (p)), is $\hat{X}_{t+1} = \sum_{k=1}^{p} \hat{\phi}_k\, X_{t-(k-1)}$.
Replaced the independent variables X_t, X_{t-1},... $X_{t-(p-1)}$ with coefficient from wavelet decomposition, Renaud *et al.* (2003) given predicted model AR become *Autoregressive* (MAR) model:

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

with $a_{j,k}$ is coefficient MAR (j=1,2,...,J and k=1,2,...,A_j) and A_j is orde of MAR model, $w_{j,t}$ is coefficient of wavelet from the data, and $v_{j,t}$ is scale coefficient from data.
Figure 1 shows that wavelet (2) model formation in level J=4, $A_j = 2$. It is illustrated that to predict the 18-th data with MAR (2) then input variables are coefficient wavelet 1-st level in t=17 and t=15, coefficient wavelet 2-nd level in t=17 and t=13, coefficient wavelet 3-rd level in t=17 and t=9, coefficient wavelet 4-th level in t=17 and t=1, coefficient scale 4-th level in t=17 and t=1.
The MAR model has a shape similar to the multiple regression model, equation (2) can be written as:

$$X_{t+1} = \sum_{j=1}^{J} \sum_{k=1}^{A_j} a_{j,k}\, w_{j,t-2^j(k-1)} + \sum_{k=1}^{A_j} a_{J+1k,}\, v_{J,t-2^J(k-1)} + \varepsilon_{t+1} \quad (3)$$

For example, the number of data N=70, J=4 and A_j=2 ($k = 1,2$), MAR model can be written as:

$$X_{t+1} = \sum_{j=1}^{4} \sum_{k=1}^{2} a_{j,k}\, w_{j,t-2^j(k-1)} + \sum_{k=1}^{2} a_{J+1k,}\, v_{J,t-2^J(k-1)} + \varepsilon_{t+1}$$

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 6 Issue 9, September 2019
ISSN (Online) 2348 – 7968 | Impact Factor (2019) – 5.248
www.ijiset.com

$$X_{t+1} = a_{1,1}w_{1,t} + a_{1,2}w_{1,t-2} + a_{2,1}w_{2,t} + a_{2,2}w_{2,t-4} + a_{3,1}w_{3,t} + a_{3,1}w_{3,t-8} + a_{4,1}w_{4,t} + a_{4,1}w_{4,t-16} + a_{5,1}v_{4,t} + a_{5,2}v_{4,t-16} + \varepsilon_{t+1}$$

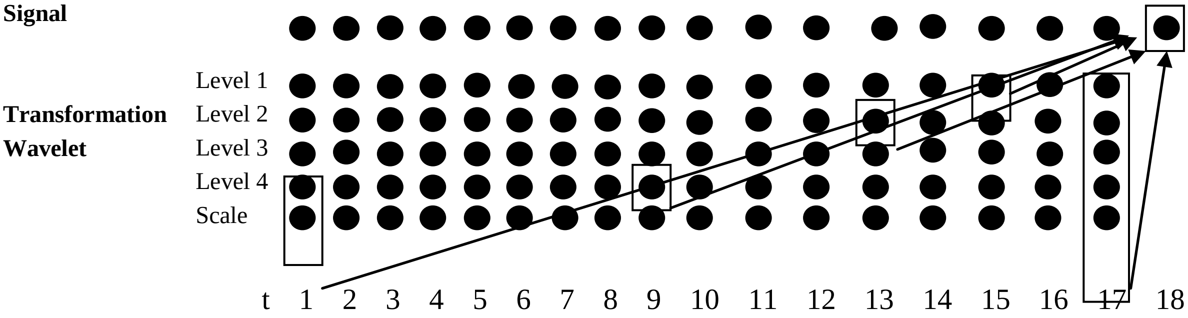

Fig 2. The illustration of Wavelet Model for J=4 and A_j=2

Can be written as matrix:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{16} \\ X_{17} \\ X_{18} \\ X_{19} \\ \vdots \\ X_t \\ \vdots \\ X_{69} \\ X_{70} \end{bmatrix} = \begin{bmatrix} w_{1,0} & w_{1,-2} & w_{2,0} & w_{2,-4} & w_{3,0} & w_{3,-8} & w_{4,0} & w_{4,-16} & v_{4,0} & v_{4,-16} \\ w_{1,1} & w_{1,-1} & w_{2,1} & w_{2,-3} & w_{3,1} & w_{3,-7} & w_{4,1} & w_{4,-15} & v_{4,1} & v_{4,-15} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{1,15} & w_{1,13} & w_{2,15} & w_{2,11} & w_{3,15} & w_{3,7} & w_{4,15} & w_{4,-1} & v_{4,15} & v_{4,-1} \\ w_{1,16} & w_{1,14} & w_{2,16} & w_{2,12} & w_{3,16} & w_{3,8} & w_{4,16} & w_{4,0} & v_{4,16} & v_{4,0} \\ w_{1,17} & w_{1,15} & w_{2,17} & w_{2,13} & w_{3,17} & w_{3,9} & w_{4,17} & w_{4,1} & v_{4,17} & v_{4,1} \\ w_{1,18} & w_{1,16} & w_{2,18} & w_{2,14} & w_{3,18} & w_{3,10} & w_{4,18} & w_{4,2} & v_{4,18} & v_{4,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{1,t-1} & w_{1,t-3} & w_{2,t-1} & w_{2,t-5} & w_{3,t-1} & w_{3,t-9} & w_{4,t-1} & w_{4,t-17} & v_{4,t-1} & v_{4,t-17} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{1,68} & w_{1,66} & w_{2,68} & w_{2,64} & w_{3,68} & w_{3,60} & w_{4,68} & w_{4,52} & v_{4,68} & v_{4,52} \\ w_{1,69} & w_{1,67} & w_{2,69} & w_{2,65} & w_{3,69} & w_{3,61} & w_{4,69} & w_{4,53} & v_{4,69} & v_{4,53} \end{bmatrix} \begin{bmatrix} a_{1,1} \\ a_{1,2} \\ a_{2,1} \\ a_{2,2} \\ a_{3,1} \\ a_{3,2} \\ \vdots \\ a_{j,k} \\ \vdots \\ a_{5,1} \\ a_{5,2} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_{70} \end{bmatrix}$$

$$\boldsymbol{s_1} = A_1\boldsymbol{\alpha} + \boldsymbol{\varepsilon_1} \qquad (4)$$

with:

$\mathbf{s_1}$: vector of time series data with size 70 x 1; A_1 : matrix of coefficient wavelet with size 70 x 10; $\boldsymbol{\alpha}$: vector of estimate parameter with size 10 x 1; $\boldsymbol{\varepsilon_1}$: vector error with size 70 x 1.

The coefficients in matrix A, there are negative, zero and positive indexes. Coefficients with zero and negative indices are not found in the results of decomposition with wavelets. The formation of the MAR model is done by not including the coefficients of zero and negative indexes, so that the vectors **s, ε** and matrix A starting from the 18th row can be assumed as

$$\mathbf{s} = \mathrm{A}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \qquad (5)$$

$\mathbf{s}$: vector of time series data with size 53 x 1

A : matrix of coefficient wavelet with size 53 x 10; $\boldsymbol{\alpha}$: vector of estimate parameter with size 10 x 1 $\boldsymbol{\varepsilon}$: vector error with size 53 x 1

As in multiple regression, to estimate the parameters in the MAR model can use the method of least squares (Ordinary least square), namely by minimizing the number of squares error:

$$\sum_{i=1}^{n} \boldsymbol{\varepsilon}_i^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\mathbf{s} - \mathrm{A}\boldsymbol{\alpha})^T(\mathbf{s} - \mathrm{A}\boldsymbol{\alpha}) \qquad (6)$$

By deriving the equation (6) to α then $-2\mathrm{A}^T\mathbf{s} + 2\mathrm{A}^T\mathrm{A}\boldsymbol{\alpha} = 0$,so that can be obtained $\mathrm{A}^T\mathrm{A}\boldsymbol{\alpha} = \mathrm{A}^T\mathbf{s}$ or $\boldsymbol{\alpha} = (\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}^T\mathbf{s}$. So estimate parameter for model (5) is $\hat{\boldsymbol{\alpha}} = (\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}^T\mathbf{s}$ and $\hat{\mathbf{s}} = \mathrm{A}\hat{\boldsymbol{\alpha}}$. By using J = 4 and Aj = 2, there are 17 data that cannot be estimated using model (5), namely data 1 to 17. In its place, the authors estimate the data using the average data to 1-17.

The longitudinal data that will be examined in this study is the longitudinal data with m subjects and observations of each subject are repeated n times. The data structure can be seen in Table 1.

Table 1. Structure of Longitudinal Data

Subject	Observation	Dependent
1st subject	1	$X_{1(1)}$
	2	$X_{2(1)}$
	⋮	⋮

	n	$X_{n(1)}$
2nd subject	1	$X_{1(2)}$
	2	$X_{2(2)}$
	⋮	⋮
	n	$X_{n(2)}$
⋮	⋮	⋮
m-*th* subject	1	$X_{1(m)}$
	2	$X_{2(m)}$
	⋮	⋮
	n	$X_{n(m)}$

3.2. MAR Model for Longitudinal Data

The MAR model of the i-th observation subject to (t + 1) follows the following formula:

$$\hat{X}_{t+1(i)} = \sum_{j=1}^{J} \sum_{k=1}^{A_j} \hat{a}_{j,k(i)}\, w_{j,t-2^j(k-1)(i)} + \sum_{k=1}^{A_j} \hat{a}_{J+1,k(i)}\, v_{J,t-2^J(k-1)(i)} \quad (7)$$

with:

$a_{j,k(i)}$: Coefficient of MAR i-th subject (j=1,2,...,J and k=1,2,...,$A_j$and i =1,2,...,m)
A_j : orde from MAR model
$w_{j,t(i)}$: coefficient of wavelet level j-th from i-th subject in t-th observation
$v_{j,t(i)}$: coefficient of scale level j-th from subject i-th in observation t-th.

The selection of which wavelet coefficients are used to form the MAR Model (7), depends on the values of J and Aj. In this research, to build a prediction model at level J = 4, the order MAR A_j = 2 is used with j = 1,2,3,4. According to equation (7), the MAR model on the i subject becomes:

$$\hat{X}_{t+1(i)} = \hat{a}_{1,1(i)}w_{1,t(i)} + \hat{a}_{1,2(i)}w_{1,t-2(i)} + \hat{a}_{2,1(i)}w_{2,t(i)} + \hat{a}_{2,2(i)}w_{2,t-4(i)} + \hat{a}_{3,1(i)}w_{3,t(i)} + \hat{a}_{3,1(i)}w_{3,t-8(i)} + \hat{a}_{4,1(i)}w_{4,t(i)} + \hat{a}_{4,1(i)}w_{4,t-16(i)} + \hat{a}_{5,1(i)}v_{4,t(i)} + \hat{a}_{5,2(i)}v_{4,t-16(i)} \quad (8)$$

By substituting variables from level 1, $w_{1,t(i)}$ = $Z_{1t(i)}$, $w_{1,t-2(i)}$ = $Z_{2t(i)}$ and coefficient $\hat{a}_{1,1(i)}$ = $\hat{\beta}_{1(i)}$, $\hat{a}_{1,2(i)}$ = $\hat{\beta}_{2(i)}$. From level 2, $w_{2,t(i)}$ = $Z_{3t(i)}$, $w_{2,t-4(i)}$ = $Z_{4t(i)}$ and coefficient $\hat{a}_{2,1(i)}$ = $\hat{\beta}_{3(i)}$, $\hat{a}_{1,2(i)}$= $\hat{\beta}_{4(i)}$. And so on, the last from level 4, $v_{4,t(i)}$ =$Z_{9t(i)}$, $v_{4,t-16(i)}$ =$Z_{10t(i)}$ and coefficient $\hat{a}_{5,1(i)}$ = $\hat{\beta}_{9(i)}$, $\hat{a}_{5,2(i)}$ = $\hat{\beta}_{10(i)}$, then MAR model i-th subject as follow as:

$$\hat{X}_{t+1(i)} = \hat{\beta}_{1(i)} Z_{1t(i)} + \hat{\beta}_{2(i)} Z_{2t(i)} + \hat{\beta}_{3(i)} Z_{3t(i)} + \hat{\beta}_{4(i)} Z_{4t(i)} + \hat{\beta}_{5(i)} Z_{5t(i)} + \hat{\beta}_{6(i)} Z_{6t(i)} + \hat{\beta}_{7(i)} Z_{7t(i)} + \hat{\beta}_{8(i)} Z_{8t(i)} + \hat{\beta}_{9(i)} Z_{9t(i)} + \hat{\beta}_{10(i)} Z_{10t(i)} \quad (9)$$

After removing the zero and negative indexes, there remains a positive index at t = 17.18, ..., N-1.
In matrix, model (8) can be written as $\mathbf{\hat{X}_{t+1(i)}} = \mathbf{Z_{t(i)}\hat{\beta}_{(i)}}$, i = 1, 2,..., m or:

$$\mathbf{\hat{X}_{t+1(1)}} = \mathbf{Z_{t(1)}\hat{\beta}_{(1)}}$$
$$\mathbf{\hat{X}_{t+1(2)}} = \mathbf{Z_{t(2)}\hat{\beta}_{(2)}}$$
$$\vdots$$
$$\mathbf{\hat{X}_{t+1(m)}} = \mathbf{Z_{t(m)}\hat{\beta}_{(m)}}$$

$$\begin{bmatrix} \mathbf{\hat{X}_{t+1(1)}} \\ \mathbf{\hat{X}_{t+1(2)}} \\ \vdots \\ \mathbf{\hat{X}_{t+1(m)}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z_{t(1)}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z_{t(2)}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z_{t(m)}} \end{bmatrix} \begin{bmatrix} \mathbf{\hat{\beta}_{(1)}} \\ \mathbf{\hat{\beta}_{(2)}} \\ \vdots \\ \mathbf{\hat{\beta}_{(m)}} \end{bmatrix}$$

$$\mathbf{\hat{X}_{t+1}} = \mathbf{Z_t\hat{\beta}}$$

$\mathbf{\hat{\beta}}$ can be obtained from OLS with $\mathbf{\hat{\beta}} = (\mathbf{Z_t}^T\mathbf{Z_t})^{-1}\mathbf{Z_t}^T\,\mathbf{X_{t+1}}$ or $\mathbf{\hat{\beta}_i} = (\mathbf{Z_{t(i)}}^T\mathbf{Z_{t(i)}})^{-1}\mathbf{Z_{t(i)}}^T\,\mathbf{X_{t+1(i)}}$

3.3.Case study

The case study in this study is data on 3 sectors / groups of Indonesian yoy inflation spending from January 2007 to August 2017 taken from Bank Indonesia, namely (1). Foodstuffs group; (2). Processed foods, beverages, cigarettes and tobacco groups; (3). Clothing group. Descriptive statistics from data on 3 sectors of inflation expenditure in Indonesia are as follows

Table 2 . Statistic Descriptive statistics from data on 3 sectors of inflation expenditure in Indonesia are as follows

Sector	min	max	mean	variance
Sector 1	1.452	20.020	8.810	18.63753
Sector 2	4.250	12.930	7.071	3.468268
Sector 3	-0.413	12.270	5.180	8.16835

The first step is to decompose the MODWT of data on 3 groups of inflation spending in Indonesia using the Haar (D2) and Daubechies (D4) filters with a level (J) = 4. The MODWT decomposition process will produce wavelet coefficients (w) and scale (v) that consist of w_1, w_2, w_3, w_4, and v_4. Processing is done using software R. These coefficients will be used to form the MAR model input. The selection of wavelet coefficients used to form the Multiscale Autoregressive Model for each i-th sector follows the formula (8). After getting the coefficients that will be used to form the model, continue modeling with OLS using R software for each subject.

3.4.Model Multiscale Autoregressive (MAR) withFilter Haar (D2)

Sector 1 models are formed by entering ten independent variables $Z_{1(1)}, Z_{2(1)}, \ldots, Z_{10(1)}$ is :

$$\hat{X}_{t+1(1)} = 1.86247\, Z_{1t(1)} + 0.23032\, Z_{2t(1)} + 0.43601\, Z_{3t(1)} + 0.38182\, Z_{4t(1)} + 0.919\,6\, Z_{5t} - 0.42526\, Z_{6t(1)} + 0.95677\, Z_{7t(1)} - 0.16585\, Z_{8t(1)} + 0.95723\, Z_{9t(1)} + 0.02208\, Z_{10t(1)}$$

Together the model is significant with p-value < 2.2 x 10^{-16}, residual error standard of 1.269 and R^2 of 0.9832. The individually significant variable is the variable $Z_{1(1)}$, $Z_{3(1)}$, $Z_{5(1)}$, $Z_{6(1)}$, $Z_{7(1)}$, $Z_{8(1)}$ and $Z_{9(1)}$.

Sector 2 model formed by entering ten independent variables $Z_{1(2)}, Z_{2(2)}, \ldots, Z_{10(2)}$ are :

$$\hat{X}_{t+1(2)} = 1.65308\, Z_{1t(2)} + 0.55415 Z_{2t(2)} + 0.49436\, Z_{3t(2)} - 0.22690 Z_{4t(2)} + 1.17423\, Z_{5t(2)} - 0.36521\, Z_{6t(2)} + 0.99114\, Z_{7t(2)} - 0.18499 Z_{8t(2)} + 1.07749\, Z_{9t(2)} + -0.08326\, Z_{10t(2)}$$

Simultaneous, the model is significant with p-value < 2.2x10^{-16}, residual error standard is 0.3806 and R^2 is 0.9976. The individually significant variable is the variable $Z_{1(2)}$, $Z_{2(2)}$, $Z_{3(2)}$, $Z_{5(2)}$, $Z_{6(2)}$, $Z_{7(2)}$, $Z_{8(2)}$, $Z_{9(2)}$ and $Z_{10(2)}$.

Sector 3 model formed by entering ten independent variables $Z_{1(3)}, Z_{2(3)}, \ldots, Z_{10(3)}$ are :

$$\hat{X}_{t+1(3)} = 1.40250\, Z_{1t(3)} + 0.56226\, Z_{2t(3)} + 0.45962\, Z_{3t(3)} + 0.10331\, Z_{4t(3)} + 0.53969\, Z_{5t(3)} - 0.35859\, Z_{6t(3)} + 1.01910\, Z_{7t(3)} - 0.05276\, Z_{8t(3)} + 0.99058\, Z_{9t(3)} - 0.01733\, Z_{10t(3)}$$

Simultaneous, the model is significant with p-value < 2.2x10^{-16}, residual error standard is 0.9377 and R^2 sebesar 0.9741. The individually significant variable is the variable $Z_{1(3)}$, $Z_{2(3)}$,$Z_{3(4)}$, $Z_{5(3)}$, $Z_{6(3)}$, $Z_{7(3)}$, $Z_{8(3)}$ and $Z_{9(3)}$. Graphically the MAR model with the Haar (D2) filter formed in sectors 1-3 is presented in Figure 2-4.

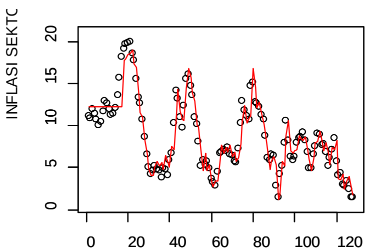

Figure 2. MAR Model with filter D2 in Sector 1

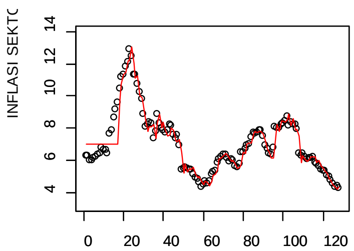

Figure 3. MAR Model with filter D2 in sector 2

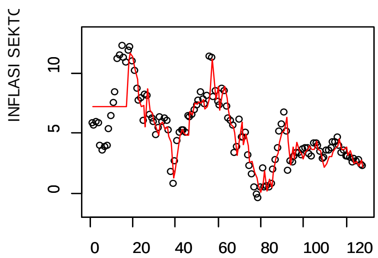

Figure 4 . MAR Model with filter D2 in Sector 3

3.5Model Multiscale Autoregressive (MAR) Filter D4

The MAR sector 1 model is formed by entering ten independent variables $Z_{1(1)}, Z_{2(1)}, \ldots, Z_{10(1)}$ is :

$$\hat{X}_{t+1(1)} = 0.21237\, Z_{1t(1)} + 0.55226 Z_{2t(1)} + 0.08748 Z_{3t(1)} + 0.13201 Z_{4t(1)} - 2.16025 Z_{5t(1)} - 0.31032\, Z_{6t(1)} - 0.84086\, Z_{7t(1)} - 0.02900\, Z_{8t(1)} + 0.57820 Z_{9t(1)} + 0.33102\, Z_{10t(1)}$$

Simultaneously, the model is significant with a p-value < 2.210 x10^{-16}, a residual error standard of 3.4 and R2 of 0.8794. The individually significant variable is the variable $Z_{5(1)}$, $Z_{7(1)}$, $Z_{9(1)}$ and $Z_{10(1)}$.

The MAR sector 2 model is formed by entering ten independent variables $Z_{1(2)}, Z_{2(2)}, \ldots, Z_{10(2)}$ is :

$$\hat{X}_{t+1(2)} = 0.31058\, Z_{1t(2)} + 0.18636\, Z_{2t(2)} - 0.91212\, Z_{3t(2)} + 0.88090 Z_{4t(2)} - 4.15317\, Z_{5t(2)} - 1.38979\, Z_{6t(2)} - 1.33224\, Z_{7t(2)} - 0.34781 Z_{8t(2)} + 1.24311\, Z_{9t(2)} - 0.27673\, Z_{10t(2)}$$

Simultaneously, the model is significant with a p-value < 2.210 x10^{-16}, a residual error standard of 1.151 and R^2 of 0.9776. The individually significant variable is the variable $Z_{5(2)}$, $Z_{6(2)}$, $Z_{7(2)}$, $Z_{9(2)}$ and $Z_{10(2)}$.
The MAR sector 2 model is formed by entering ten independent variables $Z_{1(3)}$, $Z_{2(3)}$,..., $Z_{10(3)}$ is:

$$\hat{X}_{t+1(3)} = 0.020490\ Z_{1t(3)} - 0.002243 Z_{2t(3)} + 0.103098\, Z_{3t(3)} + 0.518868 Z_{4t(3)} - 1.163302 Z_{5t(3)} - 0.147765\, Z_{6t(3)} - 1.125024 Z_{7t(3)} + 0.783924 Z_{8t(3)} + 0.838694 Z_{9t(3)} + 0.059892 Z_{10t(3)}$$

Simultaneously, the model is significant with a p-value < 2.210 x10^{-16}, a residual error standard of 1.693 and R^2 of 0.9156. The individually significant variable is the variable $Z_{5(3)}$, $Z_{7(3)}$, $Z_{8(3)}$ dan $Z_{9(3)}$. Graphically the MAR model with the D4 filter formed in sectors 1-3 is as follows.

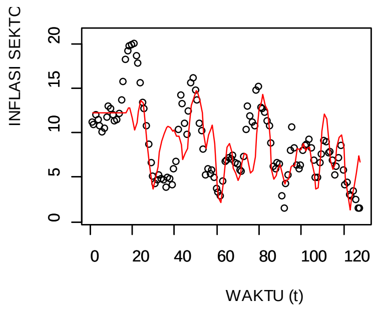

Figure 5. MAR Model with filter D4 in Sector 1

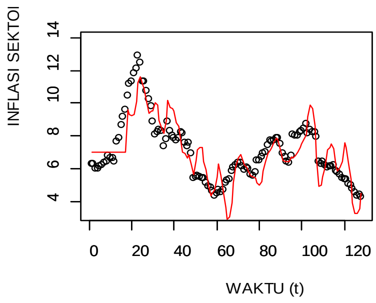

Figure 6. MAR Model with filter D4 in sector 2

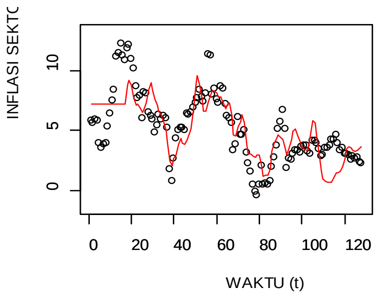

Figure 7. MAR Model with filter D4 in sector 3

Table 3. Comparison of MAR model results with Haar (D2) and Daubechies (D4) filters

Sector	Standar Error Residual		The number of Significant Coefficient		Coefficient of Determination (R^2)	
	MAR D2	MAR D4	MAR D2	MAR D4	MAR D2	MAR D4
1	1.269	3.4	7	4	0.9832	0.8794
2	0.3806	1.151	9	5	0.9976	0.9776
3	0.9377	1.693	8	5	0.9741	0.9156

From the results of MAR modeling with Haar (D2) and Daubechies (D4) filters, it can be seen that the residual error standard for MAR models with filter D2 is always smaller than D4, the number of model coefficients is significant, MAR models with D2 tend to be more numerous than D4, and the magnitude of the coefficient of determination (R2) of the MAR model with filter D2 is greater than the model of MAR with filter D4. From Figure 2-4 and Figure 5-7, visually it can also be seen that the MAR model with the D2 filter is closer to the actual data. So the results of data modeling of 3 sectors / inflation expenditure in Indonesia using the Haar filter (D2) are better than the MAR model with filter D4. From the whole model formed using either the D2 or D4 filters the coefficients derived from the scale coefficient v4 are always significant. Because the scale coefficient provides the largest contribution to modeling.

4. Conclusion

MAR modeling in data on 3 sectors of inflation expenditure in Indonesia using Haar filter (D2) both statistically and visually shows better results than MAR models with Daubechies filter (D4)

Acknowledgment

Thank you to Ministry of Research and Higher Education for funding this research with the PTUPT (Higher Education Applied Research) 2019.

Daftar Pustaka

[1] Hu Z, Wang N, dan Carroll RJ 2004 Profile Kernel Versus Backfitting In The Partially Linier Models For Longitudinal Or Clustered Data *Biometrika* Vol 91 No 2 p 251-262

[2] Ibrahim NA dan Suliadi 2008 Analyzing Longitudinal Data Using Gee –Smoothing Spline *Proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational science* p 26-33

[3] Liang KY dan Zeger SL 1986 Longitudinal Data Using Generalized Linier Models. *Biometrika* Vol 78 No I p 13-22

[4] Percival DB danWalden AT 2000 *Wavelets Methods for Time Series Analysis 1st Edn* Cambridge University Press Cambridge p 620

[5] Popoola AO 2007 *Fuzzy-Wavelet Method for Time Series Analysis* (UK: University of Surrey)

[6] Renaud O, Starck J & Murtagh F 2003 Prediction based on a multiscale Decomposition. *International Journal of Wavelets (Multiresolution and Information Processing 1(2)*) p 217-232

[7] Suhartono, Ulama BSS dan Endharta AJ 2010 *Seasonal Time Series Data Forecasting by Using Neural Network Multiscale Autoregresive Model (*American Journal of Applied Sciences 7 (10)) p 1373-1378

[8] Suparti, Prahutama A, Rahmawati R, dan Utami TW 2016 Modelling Local Polynomial for Longitudinal Data A Case Study; Inflation sectors in Indonesia (ARPN Journal of Engineering and Aplied Science (JEAS) Vol 11 No 23)

[9] Weiss R E 2005 *Modeling Longitudinal Data* (New York: Springer)

[10] Wu H dan Zhang JT 2006 *Nonparametric Regression Methods for Longitudinal Data Analysis* (New Jersey : John-Wiley and Sons Inc.)