Indonesia’s Inflation Analysis Using Hybrid Fourier - Wavelet Multiscale Autoregressive Method

Suparti¹, R Santoso², A Prahutama³, A R Devi⁴

¹,²,³ Statistics Departement, Diponegoro University, Semarang, Indonesia
⁴ Independent researcher

suparti702@gmail.com

Abstract. Regression analysis is a statistical method to determine the relationship between predictor variables and response variables. Regression approach can be done with parametric and nonparametric approaches. The parametric approach is rigorous with assumptions and must fulfill the assumptions to get a good model. Meanwhile, the nonparametric approach is not rigorous with assumptions because the method is based on an unknown curve shape approach. Nonparametric regression can be done with several approaches including Fourier and Wavelet methods. The Fourier method is a method based on cosine and sine series. It is very suitable for data that has repetitive or periodic patterns. The Fourier series modeling is less efficient because it requires many coefficients to obtain a good model and it is less able to handle data with sharp jumps. Recently, there has been a combination of two methods called hybrid methods that give better results. In this paper, Indonesia’s inflation data modelling is performed by using hybrid-wavelet. First, the data is modeled using Fourier with small K; then the error Fourier model is modeled by using multiscale wavelet autoregressive. In this study, the value of Inflation in Indonesia was modeled from January 2007 to August 2017. The response variable was the inflation value, while the predictor variable was time. The Fourier method with K = 100 generated MSE of 0.846216 and R² of 80.12%. The Fourier-Wavelet hybrid model with K = 1 generated MSE of 0.31 and R² of 95%. So that in inflation modeling in Indonesia, the Fourier-Wavelet hybrid regression model generated a better model than the wavelet model with fewer parameters than the Fourier method.

1. Introduction
The regression model is modeling between predictor and response variable. The parametric model is best to select when the curve shape of the regression forms a specific relationship pattern. Linear regression as one of the parametric regression models is a regression model that the error model must fulfill some model assumptions such as being independent and having a normal distribution with 0 mean and constant variance. On the other hand, when the regression curve does not form a specific pattern, it will be better to use the nonparametric regression model. Some methods related to nonparametric regression models include regression models of the kernel, spline, local polynomial, Fourier and Wavelet.
Suparti [7] has conducted a review on nonparametric regression modeling using Fourier and wavelet model based on IMSE value (Integrated Mean Square Error) [7]. Her study shows that the wavelet regression model has a faster IMSE convergence than the Fourier model. Also, it also reveals that her study used a linear wavelet model using multiresolution analysis of discrete wavelet transformation (DWT). In discrete wavelet distribution, the processed data are limited to the only $2^J$ with $J$ integer. To overcome the limitations of DWT, we can use MODWT (Maximal Overlapping Discrete Wavelet Transform) to apply to the amount of any data without being limited to the only $2^J$.

Inflation data is a time series of fluctuating data that does not take the form of a specific pattern. There has been parametric modeling of Indonesia's inflation data using the annual inflation data from December 2006 to December 2011. However, it appears that there is no suitable Box Jenkins model (both AR, MA or ARIMA) because the assumption of error independence is not fulfilled [6]. It is possible to conduct nonparametric modeling because inflation data does not form a certain pattern. Previously, inflation data modeling using nonparametric models of the local kernel, spline, and polynomial have been discussed in [5], [6] and [10]. In nonparametric modeling, the estimation of the model will adjust itself to its data pattern.

In this paper, the authors discuss the nonparametric model using the Fourier-wavelet hybrid model. The Fourier model uses a combination model of linear additive and cosine function [1], while wavelet model cascading time adopts autoregressive time series (AR (p)) with predictor variables of a lag variable of the MODWT coefficient whose model is known as the Multiscale Autoregressive model [3]. Model performance will be measured using MSE and $R^2$ sizes. Also, it is known that a good model will give a small MSE value and a large $R^2$.

2. Research Methods

This Research used secondary data taken from the official website of Bank Indonesia with research variables of Indonesia's annual inflation value (year on year/ yoy) from January 2007 to August 2017. In Fourier modeling, the response variable is the value of inflation, and the predictor variable is time, while in MAR wavelet model, the response variable is residual from Fourier model and predictor variable is MODWT coefficient lag from residual Fourier model. The first step of inflation data was modeled using Fourier for a particular K value. The Fourier model was estimated, and the Fourier model residual was calculated. Furthermore, the residuals of Fourier models were modeled using a multiscale autoregressive wavelet with MODWT. The Fourier-wavelet Hybrid model is a sum of the Fourier model and the autoregressive wavelet model of the Fourier model residual. Afterward, the researcher calculated the value of MSE and $R^2$ from the Fourier - wavelet hybrid model. In Fourier modeling, some K values start from 1 to 100. Meanwhile, in the autoregressive multiscale wavelet model, $J = 4$ and $A_j = 2$ with Haar wavelet filter (D2). The researcher made a comparison by seeing the value of MSE and $R^2$. The researcher also selected the parsimony model, that is the model with only a few parameters but having high accuracy.

3. Results and Discussion

Indonesia's annual inflation data (year on year) in January 2007 - August 2017 is presented in figure 1. The data pattern repetitively fluctuates at an interval.
3.1. Nonparametric regression modeling

The nonparametric regression model can be mathematically written in the following equation:

\[ y = f(t) + \varepsilon \]  

with \( \varepsilon \) as residual/error model distributed at approximately 0, \( f(t) \) as a function representing the intrinsic behavior of the data. There are several \( f(t) \) estimation techniques in nonparametric such as kernel, spline, local polynomial, Fourier series and wavelet [10].

3.2. Fourier series modeling

The Fourier series is a series of combined sine and cosine functions [2]. The Fourier series is highly compatible with data patterns that occur in repetitive or periodic events. In Fourier regression, in addition to using a combination of sine and cosine functional additives, it can also use a combination of linear function additives and sinus or cosine functions. Bilodeau (1992) provides nonparametric regression modeling with a combination additive of a linear function and cosine function [1]. This combination is expected to separate data trends and data fluctuations. By taking a combination additive of linear functions, one of the sine or cosine functions will give estimation efficiency. The combination additive of linear and cosine function as in the followings:

\[ f(t) = \frac{1}{2} \alpha_0 + \gamma t + \sum_{k=1}^{K} \alpha_k \cos kt \]  

Having data as much as \( n \), \( (t_i, y_i) \), \( i = 1, 2, ..., n \) with \( f(t) \) as an unknown curve of form and \( f(t) \) as approached by Fourier series (2), thus,

\[ f(t_i) = \frac{1}{2} \alpha_0 + \gamma t_i + (\alpha_{i1} \cos t_i + \alpha_{i2} \cos 2t_i + ... + \alpha_{iK} \cos Kt_i) \]  

The form of the model matrix (3) can be written as follows:

\[ \mathbf{Y} = \mathbf{f}(t) + \varepsilon = \mathbf{A}\theta + \varepsilon \]  

with

\[ \mathbf{Y} = [y_1, y_2, ..., y_n]^T, \theta = [\phi, \gamma, \alpha_{i1}, \alpha_{i2}, ..., \alpha_{iK}]^T \]  

with \( \phi = \frac{n}{2} \alpha_0 \), \( \varepsilon = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]^T \) and

\[ \mathbf{A} = \begin{bmatrix} 1 & t_1 & \cos t_1 & \cos 2t_1 & \cdots & \cos Kt_1 \\ 1 & t_2 & \cos t_2 & \cos 2t_2 & \cdots & \cos Kt_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & \cos t_n & \cos 2t_n & \cdots & \cos Kt_n \end{bmatrix} \]

The estimation of the nonparametric regression model (4) using OLS (Ordinary Least Square) leads to

\[ \hat{\theta} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y} \]  

and

\[ \hat{\mathbf{Y}} = \mathbf{A}\hat{\theta} \]  

\[ (5) \]
In determining the Fourier model, model performance is based only on the value of $K$ as the number of cosines functions formed. The number of estimated parameters in the model (3) is as much as $K + 2$.

3.3. Wavelet Multiscale Autoregressive (MAR) modeling

Generally, wavelet Multiscale Autoregressive (MAR) modeling is a modeling method using wavelet transformation, in this case using MODWT. The benefit of multiscale decomposition such as wavelet is that it can automatically split the data components, such as the components of the trend and the irregular components of the data. Therefore, this method can be used to predict both stationary and non-stationary data [4].

Suppose there is a stationary signal of $y = (y_1, y_2, ..., y_t)$ and it is assumed that the value prediction is $y_{t+1}$, the basic idea used is to use the coefficients obtained from the MODWT decomposition results is $w_{j,t-2^j(k-1)}$ and $v_{j,t-2^j(k-1)}$, with $k = 1, 2, ..., A_j$ and $j = 1, 2, ..., J$ [3]. The predictive model refers to the autoregressive model (AR $(p)$) of $\hat{y}_{t+1} = \sum_{k=1}^{p} \phi_k y_{t-k}$.

By replacing independent variable $y_t, y_{t-1}, ..., y_{t-(p-1)}$ with a coefficient of wavelet decomposition, Renaud et al. (2003) provides an AR prediction model to the Multiscale Autoregressive (MAR) model [3]:

$$\hat{y}_{t+1} = \sum_{j=1}^{J} \sum_{k=1}^{A_j} \hat{\alpha}_{j,k} w_{j,t-2^j(k-1)} + \sum_{k=1}^{J} \hat{\alpha}_{j+1,k} v_{j,t-2^j(k-1)} + \epsilon_{t+1}$$

with:

- $\hat{\alpha}_{j,k}$ : MAR coefficient ($j=1,2,...,J$ dan $k=1,2,...,A_j$)
- $A_j$ : order of MAR model
- $w_{j,t}$ : wavelet coefficient of data
- $v_{j,t}$ : the scale coefficient of data

MAR model has a form similar to a multiple regression model, and thus equation (6) can be written as in the following:

$$y_{t+1} = \sum_{j=1}^{J} \sum_{k=1}^{A_j} a_{j,k} w_{j,t-2^j(k-1)} + \sum_{k=1}^{A_j} a_{j+1,k} v_{j,t-2^j(k-1)} + \epsilon_{t+1}$$

As an example, $J=4$ and $A_j=2$ ($k = 1,2$), MAR is formulated as:

$$y_{t+1} = a_{1,1} w_{1,t} + a_{1,2} w_{1,t-2} + a_{2,1} w_{2,t} + a_{2,2} w_{2,t-4} + a_{3,1} w_{3,t} + a_{3,2} w_{3,t-8} + a_{4,1} w_{4,t} + a_{4,2} w_{4,t-16} + \epsilon_{t+1}$$

Equation (8) components of $w$ and $v$ are worth when the value of $t > 17$. Thus, when $N = 69$ it can be expressed in the following matrix form:

$$\begin{bmatrix}
Y_{18} \\
Y_{19} \\
\vdots \\
Y_{68} \\
Y_{69}
\end{bmatrix} =
\begin{bmatrix}
w_{1,17} & w_{1,15} & w_{1,13} & w_{1,11} & w_{1,9} & w_{1,7} & w_{1,5} & w_{1,3} & w_{1,1} & v_{1,17} & v_{1,15} & v_{1,13} & v_{1,11} & v_{1,9} & v_{1,7} & v_{1,5} & v_{1,3} & v_{1,1} \\
w_{1,18} & w_{1,16} & w_{1,14} & w_{1,12} & w_{1,10} & w_{1,8} & w_{1,6} & w_{1,4} & w_{1,2} & v_{1,18} & v_{1,16} & v_{1,14} & v_{1,12} & v_{1,10} & v_{1,8} & v_{1,6} & v_{1,4} & v_{1,2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_{1,67} & w_{1,65} & w_{1,63} & w_{1,61} & w_{1,59} & w_{1,57} & w_{1,55} & w_{1,53} & w_{1,51} & v_{1,67} & v_{1,65} & v_{1,63} & v_{1,61} & v_{1,59} & v_{1,57} & v_{1,55} & v_{1,53} & v_{1,51} \\
w_{1,68} & w_{1,66} & w_{1,64} & w_{1,62} & w_{1,60} & w_{1,58} & w_{1,56} & w_{1,54} & w_{1,52} & v_{1,68} & v_{1,66} & v_{1,64} & v_{1,62} & v_{1,60} & v_{1,58} & v_{1,56} & v_{1,54} & v_{1,52}
\end{bmatrix} +
\begin{bmatrix}
\hat{a}_{1,1} \\
\hat{a}_{1,2} \\
\vdots \\
\hat{a}_{5,1} \\
\hat{a}_{5,2}
\end{bmatrix} +
\begin{bmatrix}
\epsilon_{18} \\
\epsilon_{19} \\
\vdots \\
\epsilon_{67} \\
\epsilon_{68}
\end{bmatrix}$$

The matrix can be written to:

$$s = A\alpha + \epsilon$$

with:

- $s$ : time series data vector of $52 \times 1$
- $A$ : matrix containing $52 \times 10$ wavelet coefficients
- $\alpha$ : vector with estimated parameters of $10 \times 1$
\[ \varepsilon : \text{error vector of 52 x 1} \]

To estimate the parameters in the MAR model, we can use the ordinary least square method resulting in \( \hat{\alpha} \) = \((A^T A)^{-1}A^T s\) and \( \hat{s} = A\hat{\alpha} \). In wavelet MAR model data prediction starts from \( t = 17 \) that is \( \hat{y}_{17}, \hat{y}_{18}, ..., \hat{y}_{70} \) with

\[
\hat{y}_{70} = \hat{a}_{1,1}w_{1,69} + \hat{a}_{1,2}w_{4,67} + \hat{a}_{2,1}w_{2,69} + \hat{a}_{2,2}w_{2,65} + \hat{a}_{3,1}w_{3,69} + \hat{a}_{3,1}w_{3,61} + \hat{a}_{4,1}w_{4,69} + \hat{a}_{4,1}w_{4,53} + \hat{a}_{5,1}v_{4,69} + \hat{a}_{5,2}v_{4,53} \\
\]

(10)

As for \( \hat{y}_{1}, \hat{y}_{2}, ..., \hat{y}_{17} \) the authors used the average of \((y_1, y_2, ..., y_{17})\).

### 3.4 Fourier-wavelet hybrid modeling

A hybrid model is obtained by combining several methods in one model. For example in the nonparametric regression model (1), \( f \) is approached by a local polynomial model and \( \varepsilon \) is approached by spline truncated model or vice versa. The model is a local polynomial mix-spline truncated or local hybrid polynomial spline [8]. In this paper, the researcher selected the hybrid of \( f \) approached with the Fourier model and \( \varepsilon \) approached with the wavelet multiscale autoregressive model called the Fourier-wavelet multiscale autoregressive model.

The Fourier-wavelet multiscale autoregressive hybrid model of the (1) model is:

\[ y = f(t_i) + \varepsilon_i \]

With

\[ f(t_i) = \frac{1}{2} \alpha_0 + \gamma t_i + \sum_{k=1}^{K} \alpha_k \cos \; kt_i \]

\[ \varepsilon_i = \sum_{j=1}^{J} \sum_{k=1}^{K} a_{j,k} w_{j,i-2^j (k-1)-1} + \sum_{k=1}^{K} a_{j+1,k} v_{j,i-2^j (k-1)-1} + u_i \]

(12)

\[ u_i = \text{residual of the error model} \]

For \( J = 4 \) and \( A_j = 2 \), thus,

\[ \varepsilon_i = \sum_{j=1}^{4} \sum_{k=1}^{K} a_{j,k} w_{j,i-2^j (k-1)-1} + \sum_{k=1}^{K} a_{5,k} v_{4,i-2^j (k-1)-1} + u_i \]

(13)

In the Fourier model, a couple of \( K \) values were evaluated. At each intake, the value of \( K \) was then calculated to obtain the residuals which were modeled using wavelet autoregressive. Fourier-Wavelet hybrid model results from Indonesia’s inflation presented in figure 1 was tested to find some \( K \) values in the Fourier model shown in table 1. There was a devaluation of MSE and a significant \( R^2 \) value increase when the Fourier model became the Fourier-wavelet hybrid model. Statistically, modeling results can be seen in table 1 and can be visually seen in figure 2.

| No | Model | K   | MSE  | \( R^2 \) | The number of parameters |
|----|-------|-----|------|----------|------------------------|
| 1  | Fourier | K=1 | 4.237 | 0.143    | 3                      |
|    | Hybrid Fourier-wavelet |      | 0.310 | 0.950    | 13                     |
| 2  | Fourier | K=5 | 4.234 | 0.144    | 7                      |
|    | Hybrid Fourier-wavelet |      | 0.299 | 0.952    | 17                     |
| 3  | Fourier | K=10| 4.067 | 0.177    | 12                     |
|    | Hybrid Fourier-wavelet |      | 0.299 | 0.952    | 19                     |
| 4  | Fourier | K=50| 2.953 | 0.403    | 52                     |
|    | Hybrid Fourier-wavelet |      | 0.208 | 0.976    | 62                     |
| 5  | Fourier | K=100| 0.983 | 0.801    | 102                    |
|    | Hybrid Fourier-wavelet |      | 0.189 | 0.980    | 112                    |
From some selection of K values in the Hybrid Fourier-Wavelet model, i.e., K = 1, 10, 50 and 100, it is revealed that the greater the K value, the smaller MSE value and the greater the $R^2$. However, if the value of K is too large, the model will involve more parameters making it very complex. For Fourier convergence models, it will be achieved if K > 100, but the Fourier Wavelet Hybrid model with K = 5 shows a good model in that even the value of K = 1 was not that different with K = 5. By modeling the parsimony model with K = 1, the variable t and the MODWT coefficient have contributed 95%.

Fourier-Wavelet MAR Model Estimation for K = 1 is:

$$\hat{y} = \hat{f}(t) + \hat{\epsilon}$$

with

$$\hat{f}(t) = 7.31024523 - 0.02276201t - 0.00950163 \cos t$$

$$\hat{\epsilon}_t = 1.71539 w_{1,t-1} + 0.14738 w_{1,t-3} + 0.45581 w_{2,t-1} + 0.06410 w_{2,t-5} + 1.09210 w_{3,t-1} - 0.45823 w_{3,t-9} + 0.96491 w_{4,t-1} - 0.11970 w_{4,t-17} + 0.93886 v_{4,t-1} - 0.08573 v_{4,t-17}$$

(14)
4. Conclusion
Based on the analysis and discussion presented in section 3, it is revealed that the hybrid model of Fourier-wavelet multiscale autoregressive provides a significant decrease in MSE and $R^2$ increment. Fourier-wavelet MAR hybrid model with $K$ value = 1 gives $R^2$ value of 95% meaning that contribution of time variable and MODWT coefficient in the model is 95%. This model is the simplest Fourier-wavelet hybrid model but has contributed to a fairly large model. In the future we would like develop the Fourier Model into Spline truncated and Polynomial local.

Acknowledgments
The author would like to exert gratitude to DRPM (Directorate of Research and Public Infrastructure) of Ministry of Research Technology and Higher Education for funding this research with PTUPT scheme (Higher Education Applied Research) for the year 2018 with contract: 101-124 / UN7.P4.3 / PP / 2018 on February 5, 2018.

References
[1] Bilodeau M 1992 Fourier smoother and additive models The Canadian Journal of Statistics 20 (3) 257-69
[2] Odgen R T 1997 Essential Wavelets For Statistical Applications and Data Analysis (Boston: Birkhauser)
[3] Renaud O, Starck Jand Murtagh F2003 Prediction based on a multiscale decomposition International Journal of Wavelets, Multiresolution and Information Processing 1(2) 217-32.
[4] Suhartono, Ulama B S, and Endharta A 2010 Seasonal time series data forecasting by using neural network multiscale autoregressive model American Journal of Applied Sciences 7(10) 1373-78.
[5] Suparti2013 Analisis data inflasi di Indonesia menggunakan model regresi spline Media Statistika 6(1) 1-9
[6] Suparti2013 Analisis data inflasi di Indonesia pasca kenaikan TDL dan BBM tahun 2013 menggunakan model regresi kernel Media Statistika 6(2) 91-101
[7] Suparti2005 Perbandingan estimator regresi nonparametrik menggunakan metode fourier dan metode wavelet Jurnal Matematika 8(3) 88-94.
[8] Suparti, Santoso R and Prahutama A2018 Mix local polynomial and spline truncated: the development of nonparametric regression model Journal of Physics: Conference Series 1025 1-7.
[9] Suparti, Santoso R., Prahutama A and Devi A R 2018 Regresi nonparametrik (Ponorogo Jawa Timur: Wade Group) p17
[10] Suparti, Warsito B and Mukid M A2014 Analisis data inflasi di indonesiamenggunakan model regresipolinomiallocal IndoMS Journal on Statistics 2(1) 65-78