# The Adequateness of Wavelet Based Model for Time Series

## Rukun S[1], Subanar[2], Dedi Rosadi[2] and Suhartono[3]

[1] Statistics Department of Undip Semarang
[2] Mathematics Department of UGM Yogyakarta
[3] Statistics Department of ITS Surabaya

E-mail: `rukuns@yahoo.com`, `subanar@yahoo.com`, `dedirosadi@ugm.ac.id`,
`gmsuhartono@gmail.com`

**Abstract.**  In general, time series is modeled as summation of known information i.e. historical information components, and unknown information i.e. random component. In wavelet based model, time series is represented as linear model of wavelet coefficients. Wavelet based model captures the time series feature perfectly when the historical information components dominate the process. In other hand, it has low enforcement when the random component dominates the process. This paper proposes an effort to develop the adequateness of wavelet based model, when the random component dominated the process. By weighted summation, the data is carried to the new form which has higher dependencies. Consequently, wavelet based model will work better. Finally, it is hoped that the better prediction of wavelet based model will be carried to the original prediction in reverting process.

## 1. Introduction

George E. P. Box and Gwilym M. Jenkins have become the pioneers of time series analyses with their famous book *Time Series Analysis: Forecasting and Control* (1976). In general the model is called an Autoregressive Integrated Moving Average model (ARIMA). Box-Jenkins method gives a satisfied approximation and forecasting models when the process is in stationary condition. In the opposite, this model can't capture non stationary behaviors well. Engel [6] has proposed a parametric model for non stationary time series which is called Autoregressive Conditional Heteroskedasticity model (ARCH). Bollerslev [2] develops ARCH model become to Generalized Autoregressive Conditional Heteroscedasticity model (GARCH). Almost all economics data as well as stock trading, money exchange and commodity trading are non stationer and non linear. So, it is interesting to explore some solution on non stationer and non linear phenomenon.

Beside of parametric models, there are also available some models in non parametric sense. This method not need special assumption about population distribution, so that easier to be practiced. In general non parametric methods are chosen by the reason of practicability. In 1999, Haykin [8] wrote a useful book about neural network (NN) methods. Some applications of NN had be pointed by Samarasinghe [11]. Wavelet based method becomes the recently non parametric model for time series ( Murtagh et al. [9]). Ciancio [4] pointed the use of discrete wavelet transform for time series modeling. Rukun et al. [12] give a consideration of wavelet method for heteroscedastic time series modeling. In the fact, wavelet model captures time series feature well when the historical information components dominate the process. Pre-processing

and post-processing are needed for adequateness of wavelet based model when the random component dominates the process. This solution will be discussed in this paper. Furthermore, all of these non parametric models must be done along with intensives computation. R system is used as the main computation tool in this paper, and supported with wavelet package by Aldrich [1].

Let $\{X_t\}_{t \in Z}$ is representing a time series data. The value at time $t+1$, that is $X_{t+1}$, depends to historical information component and a random component as described in equation (1).

$$X_t = f(X_{t-i}, \epsilon_{t-j}) + \epsilon_t, \quad i = 1, 2, .., p \quad j = 1, 2, .., q \tag{1}$$

For instance, the model of AR(1) indicate that the value at time $t$ depends to a historical data $X_{t-1}$ and a random component $\epsilon_t$ as mentioned in equation (2).

$$X_t = \phi X_{t-1} + \epsilon_t, \quad -1 < \phi < 1, \quad \epsilon_t \sim N(0, \sigma^2) \tag{2}$$

It has been understood that, the random component will dominate the process when $|\phi|$ decreasing tends to 0 and then the process tends to stationer or white noise process (see figure 1). In other hand, the historical information component will dominate the process when $|\phi|$ tends to 1 , so that the process tends to non stationer (see figure 2). Capability of wavelet based model for capturing time series feature will be discussed in section 3. It will be seen that this feature will be captured by wavelet perfectly when historical information component dominates the process. The capability will decrease when the random component dominates the process. Furthermore, the effort for wavelet based model capability development will be discussed in section 4.
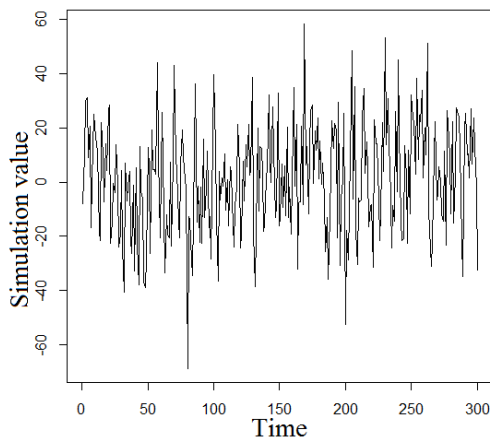


**Figure 1.** Simulation Data of AR(1) for $\phi = 0.2$ and $\sigma^2 = 400$.
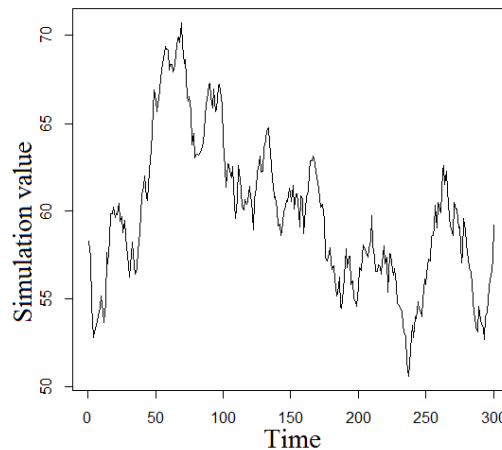
**Figure 2.** Simulation Data of AR(1) for $\phi = 0.9999$ and $\sigma^2 = 1$

There are some measurements which can be used for selecting the adequate model of time series estimation. Wei [14] pointed that model selection can be done based on forecast residuals. For instances, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) as represented in equation (3) and equation (4).

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |\epsilon_t| \tag{3}$$

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{\epsilon_t}{X_t}\right|\right)100\%  \tag{4}$$

It can be seen in section 3 that the domination of random component will increase together with MAPE values. It is hoped that MAPE value can be decreased by a method which is proposed in section 4.

## 2. Wavelet and Filtering

Wavelet is a small wave function which can build a basis for $L_2(R)$, so that every function $f \in L_2(R)$ can be expressed by a linear combination of wavelets as described in equation (5).

$$\begin{aligned} f(t) &= \sum_{k\in Z} c_{J,k}\phi_{J,k}(t) + \sum_{j\leq J}\sum_{k\in Z} w_{j,k}\psi_{j,k}(t) \\ &= S_J + D_J + D_{J-1} + \ldots + D_1 \end{aligned} \tag{5}$$

where $\phi$ and $\psi$ is a father and mother wavelet respectively with dilation and translation indexes

$$\phi_{j,k}(t) = 2^{-\frac{j}{2}}\phi(2^{-j}t - k) \tag{6}$$

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t - k) \tag{7}$$

In the discrete version, wavelet can construct filters so that every discrete realizations of $f \in L_2(R)$ can be transformed into decomposition form of scaling component or smooth component (S) and detail components (Daubechies [5]). This process is called discrete wavelet transform (DWT). There are two kinds of discrete wavelet transforms, i.e. decimated discrete wavelet transforms (DWT) and undecimated wavelet transforms (UDWT). Percival and Walden [10] pointed that UDWT has some superiorities in time series analyses compared to DWT. UDWT has shift invariant properties, so that the choice of starting point in time series is not influencing the result of transformation. In every transformation level, it produces wavelet coefficients length which is equal to the length of time series. Finally, it is able to be applied to arbitrary length of data series.

Let $\mathbf{u} = [u_0, u_1, \ldots, u_{L-1}]$ is a DWT wavelet filter. The DWT scaling filter $\mathbf{v}$ can be derived by formulation appeared in equation (8)

$$v_i = (-1)^{i+1}u_{L-1-i} \tag{8}$$

For example the Haar filter and its scaling filter can be described in equation (9)

$$\mathbf{u} = \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right], \quad \mathbf{v} = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right] \tag{9}$$

Furthermore, wavelet filters for UDWT are symbolized by $\mathbf{h}$ and $\mathbf{g}$. The relationship between DWT filters and UDWT filters can be formulated as $\mathbf{h} = \mathbf{u}/\sqrt{2}$ and $\mathbf{g} = \mathbf{v}/\sqrt{2}$. It means the Haar filters for UDWT are $\mathbf{h} = [-\frac{1}{2}, \frac{1}{2}]$, and $\mathbf{g} = [\frac{1}{2}, \frac{1}{2}]$.

Suppose $\mathbf{X} = \{X_t\}$ is a time series which is constructed from discrete time realizations of $f \in L_2(R)$. The scenario of UDWT at $j = 1$ can be shown in Figure 3 as found in Fugal [7]. The wavelet coefficient $\mathbf{w}_{1,k}$ is resulted by convolution of time series $\mathbf{X}$ and $\mathbf{h}$. The first detail component $\mathbf{D_1}$ is resulted by convolution of $\mathbf{w}_{1,k}$ and $\mathbf{h}'$ where $\mathbf{h}'$ is time reverse version of $\mathbf{h}$. The scaling coefficient $\mathbf{c}_{1,k}$ is resulted by convolution of $\mathbf{X}$ and $\mathbf{g}$. The first smooth component $\mathbf{S_1}$ is resulted by convolution of $\mathbf{c}_{1,k}$ and $\mathbf{g}'$ where $\mathbf{g}'$ is time reverse version of $\mathbf{g}$. Furthermore $\hat{\mathbf{X}} = \mathbf{S_1} + \mathbf{D_1}$ will equal to $\mathbf{X}$ regard to wavelet filtering. A higher level of UDWT can be constructed by splitting the scaling coefficient $\mathbf{c}_{j,k}$ into $\mathbf{c}_{j+1,k}$ and $\mathbf{w}_{j+1,k}$. The UDWT algorithm at j=2 can be shown in Figure 4. Furthermore $\hat{\mathbf{X}} = \mathbf{S_2} + \mathbf{D_2} + \mathbf{D_1}$ will equal to $\mathbf{X}$ regard to wavelet filtering.
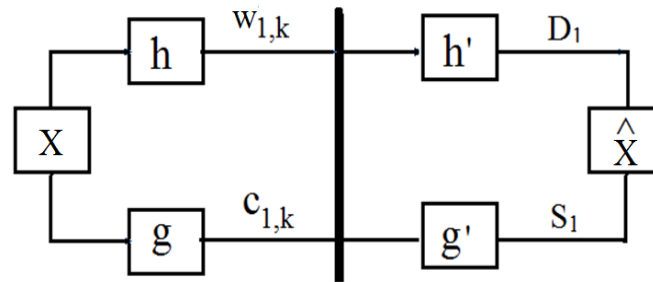
**Figure 3.** UDWT algorithm for producing wavelet coefficients at $j = 1$
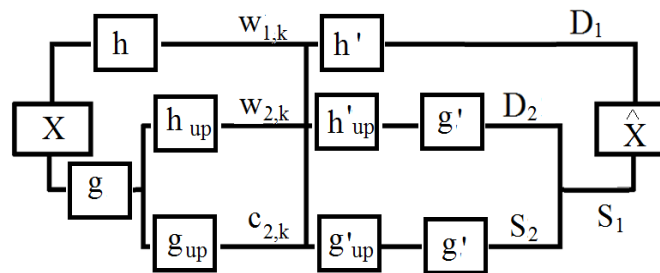
**Figure 4.** UDWT algorithm for producing wavelet coefficients at $j = 2$
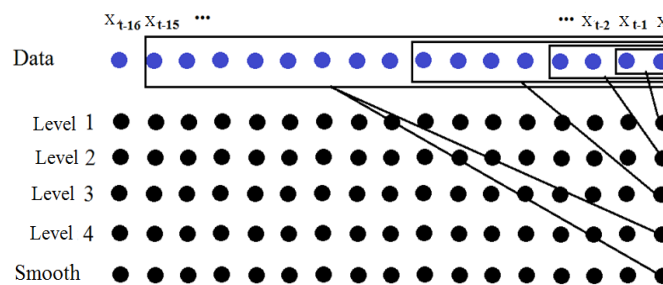
**Figure 5.** Computation of last coefficients at $j = 4$

## 3. Wavelet Based Prediction

Prediction of $X$ at time $t+1$ will refer to realization of $X$ in the past which represented by wavelet coefficients resulted from decomposition. When the Haar wavelet is used on this necessity, the computational process of last coefficients can be shown in Figure 5. Murtagh et al. [9] and Starck et al. [13] proposed that the wavelet coefficients at each level $j$ which will be used for forecasting at time $t+1$ have the form $w_{j,N-2^j(k-1)}$ and $c_{J,N-2^J(k-1)}$ (see Figure 6). The forecasting model is expressed in equation (10)

$$\hat{X}_{N+1} = \sum_{j=1}^{J}\sum_{k=1}^{A_J}\hat{a}_{j,k}w_{j,N-2^j(k-1)} + \sum_{k=1}^{A_{J+1}}\hat{a}_{J+1,k}c_{J,N-2^J(k-1)} \tag{10}$$

The highest level of decomposition is indicated by $J$, and $A_j$ is indicate the number of coefficients which chosen at level $j$. For example, if $J = 4$ and $A_j = 2$ for $j = 1, 2, 3, 4$ then (10) can be expressed as (11)

$$\hat{X}_{N+1} = \hat{a}_{1,1}w_{1,N} + \hat{a}_{1,2}w_{1,N-2} + \hat{a}_{2,1}w_{2,N} + \hat{a}_{2,2}w_{2,N-4} +$$
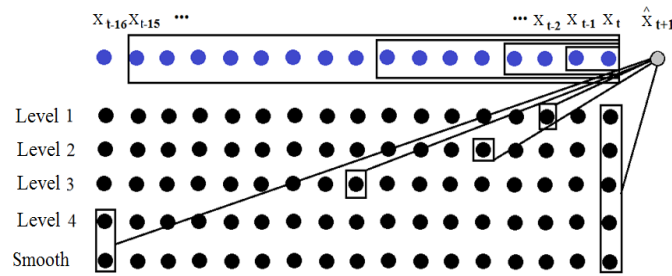
**Figure 6.** Wavelet coefficients which are used on prediction at $j = 4$

$$\hat{a}_{3,1}w_{3,N} + \hat{a}_{3,2}w_{3,N-8} + \hat{a}_{4,1}w_{4,N} + \hat{a}_{4,2}w_{4,N-16} +$$
$$\hat{a}_{5,1}c_{4,N} + \hat{a}_{5,2}c_{4,N-16} \tag{11}$$

The model which is established in equation (11), will be implemented to the AR(1) simulation data sets in Figure 1. Plots of original data and predictions can be seen in Figure 7 and Figure 8. It is seen that the model with low value of $|\phi|$ more difficult to be captured by wavelet based model (Figure 7).
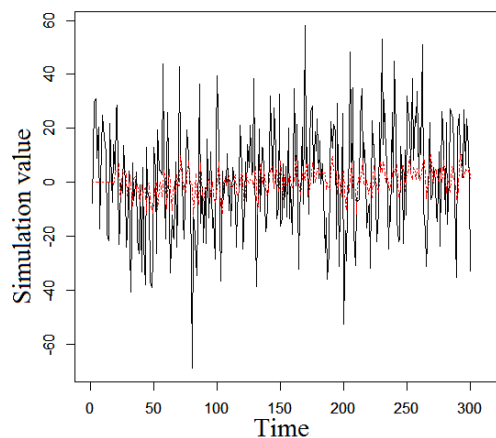


**Figure 7.** AR(1) plot of original data (black line) and its wavelet based predictions (red line) at $j = 4$, $\phi = 0.2$.
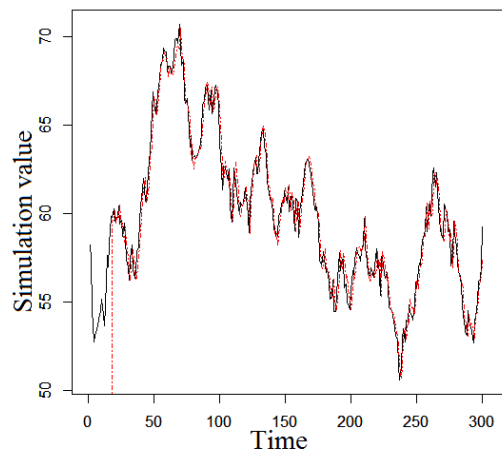


**Figure 8.** AR(1) plot of original data (black line) and its wavelet based predictions (red line) at $j = 4$, $\phi = 0.9999$.

## 4. Adequateness Development of Wavelet Based Model

Wavelet based model makes time series predictions based on dependencies of series to the historical information. The data pattern will be captured well when the series have high dependencies to the historical information. It is indicated by the high value of $|\phi|$. Wavelet based model will tracks the data pattern inaccurately when the series have high randomness. It means adequateness of wavelet based model will increase along with increasing of time series dependencies.

Suppose $\{X_t\}$, $t = 1, 2, \ldots n$ is a time series. Three transformations form will be proposed so that these results have high enough dependencies to the historical informations. Furthermore,

the transformation results become new inputs for wavelet based modeling. The prediction results of wavelet based model will be reverted to original data predictions. It is hoped that the good prediction of wavelet based model will influence to the original data prediction.

### 4.1. 2-step Weighted Summation

The first, it will be constructed a new series which expresses the weighted summation of series $\{Y_t\} = p_1 X_t + p_2 X_{t-1}$. The dependencies to the historical information will increase in the new series.

Suppose $\{X_t\}$, $t = 1, 2, \ldots n$ is a random process so that $E(X_t) = 0$, $E(X_t X_s) = 0$ for $t \neq s$, and $E(X_t^2) = \sigma^2 < \infty$ . If $Y_t = p_1 X_t + p_2 X_{t-1}$, $p_1, p_2 > 0, p_1 + p_2 = 1$ then $E(Y_t Y_s) > 0$ for $|t - s| = 1$. Suppose $t < s$.

$$
\begin{aligned}
E(Y_t Y_s) &= E(p_1 X_t + p_2 X_{t-1})(p_1 X_{t+1} + p_2 X_t) \\
&\geq p_1 p_2 E(X_t^2) \\
&> 0 \qquad \square
\end{aligned}
$$

### 4.2. k-step Weighted Summation

If the dependencies to the historical informations has not sufficient for wavelet based model yet, then it can be constructed the second opinion form $Y_t = \sum_{i=0}^{k} X_{t-i}$. The dependencies to the historical information will increase in the new series. Suppose $\{X_t\}$, $t = 1, 2, \ldots n$ is a random process so that $E(X_t) = 0$, $E(X_t X_s) = 0$ for $t \neq s$, and $E(X_t^2) = \sigma^2 < \infty$ . If $Y_t = \sum_{i=0}^{k} p_i X_{t-i}$, $p_i > 0, \sum_{i=0}^{k} p_i = 1$ then $E(Y_t Y_s) > 0$ for $|t - s| \leq k$.

Suppose $t < s$ and $p = \min_{0 \leq i \leq k}[p_i]$.

$$
\begin{aligned}
E(Y_t Y_s) &= E(p_1 X_t + \ldots + p_k X_{t-k})(p_1 X_s + \ldots + p_k X_{s-k}) \\
&\geq p(k - |t - s| + 1)E(X_t^2) \\
&> 0 \qquad \square
\end{aligned}
$$

### 4.3. Partial Sum Weighted Summation

The third alternative to increase the dependencies to the historical informations can be done by calculating the weighted partial sum of series $\{X_t\}$ as described in equation (12)

$$
Y_t = \frac{\sum_{i=1}^{t} p_i X_i}{\sum_{i=1}^{t} p_i} \tag{12}
$$

The dependencies to the historical information will increase in the new series.

Suppose $\{X_t\}$, $t = 1, 2, \ldots n$ is a random process so that $E(X_t) = 0$, $E(X_t X_s) = 0$ for $t \neq s$, and $E(X_t^2) > 0$ . If $\{Y_t\}$ represents the weighted partial sum of $\{X_t\}$ then $E(Y_t Y_s) > 0$. Suppose $t < s$.

$$
\begin{aligned}
E(Y_t Y_s) &= E\left( \frac{\sum_{k=1}^{t} p_k X_k \sum_{k=1}^{s} p_k X_k}{\sum_{k=1}^{t} p_k \sum_{k=1}^{s} p_k} \right) \\
&\geq E\left( \frac{\sum_{k=1}^{t} p_k X_k}{\sum_{k=1}^{s} p_k} \right)^2 \\
&\geq \frac{\sum_{k=1}^{t} p_k^2 E(X_k^2)}{(\sum_{k=1}^{s} p_k)^2} \\
&> 0 \qquad \square
\end{aligned}
$$

**Table 1.** MAPE Simulation Result.

| | Mean Absolute Percentage Error | | |
|---|---|---|---|
| $\phi$ | AR Model | Wavelet Model | Weighted Wavelet Model |
| 0.999 | 0.3002462 | 0.2273194 | 0.2231491 |
| 0.7 | 2.239669 | 2.001092 | 2.121224 |
| 0.6 | 1.911261 | 1.661402 | 1.594556 |
| 0.4 | 1.558648 | 1.403207 | 1.430252 |
| 0.2 | 2.565376 | 2.357589 | 2.196195 |
| 0.1 | 1.027824 | 1.242509 | 1.228172 |
| 0.0001 | 1.078761 | 1.544106 | 1.528038 |

*4.4. Simulation Result*

For example of the proposed methods, it is made a 4-step weighted summation of AR(1) simulation data for various value of $\phi$ and standard deviation equal to 7. The result can be shown in table 1. Actually, various value of $\phi$ and $\sigma$ had been tried. In general, wavelet model and weighted summation model have lower MAPE for $\phi$ tend to 1.

## 5. Concluding Remarks

The appearing of wavelet based model becomes an interesting tool in time series prediction. Beside of easy to use, as a part of non parametric methods, it does not need special assumption about the population distribution. The wavelet based model has superiority to capture non stationer and non linear pattern, thus it becomes a recently alternative solution in time series modeling. In other hand, the Box-Jenkins model has superiority to capture stationer pattern. In general the MAPE of wavelet model is lower than the MAPE of AR model for $\phi$ high enough. The example can be shown in table 1. It is not easy to choose the summation weight which will raise $\phi$ value as well as capable to track the data pattern. The effort aiming to increase the adequateness of wavelet based model has put the little result, but still needs some exploration to gain satisfied achievement.

**References**
[1] Aldrich E 2009 *A package of Functions for Computing Wavelet Filters, Wavelet Transforms and Multiresolution Analyses* http://www.ealdrich.com/wavelets/.
[2] Bollerslev T 1986 generalized autoregressive conditional heteroskedasticity *J. of Econometrics* **31** 307-27
[3] Box G E P and Jenkins G M 1976 *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day)
[4] Ciancio A 2007 analysis of time series with wavelets *Int. J. of Wavelets, Multiresolution and Information Processing* **5**(2) 241-56
[5] Daubhechies I 1992 *Ten Lecture on Wavelets* (Philadelphia: SIAM )
[6] Engel R F 1982 autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *J. Econometrica* **50** 987-1008
[7] Fugal D L 2009 *Conceptual Wavelets in Digital and Signal Processing* (Space and Signals Technologies LLC)
[8] Haykin S 1999 *Neural Networks: A Comprehensive Foundation* Prentice Hall
[9] Murtagh F, Starck J L and Renaud O 2004 on neuro wavelet modeling *Decision Support System* **37** 475-90
[10] Percival D B and Walden A T 2000 *Wavelet methods for time series analysis* (Cambridge: CU Press )
[11] Samarasinghe S 2006 *Neural Network for Applied Science and Engineering* (New York: Auerbach Pub)
[12] Rukun S, Subanar, Rosadi D, and Suhartono 2011 heteroscedastic time series model by wavelet transform *Proc. of "The 6$^{th}$ SEAMS-UGM Conference 2011"*

[13] Starck J L, Jalal F and Murtagh F 2007 the undicimated wavelet decomposition and its reconstruction, *IEEE on Image Processing* **16**(2) 297-309
[14] Wei William W S 1994 *Time Series Analysis: Univariate and Multivariate methods* (Canada: Addison-Wesley)