Analysis of Financial Time Series Data Using Adaptive Neuro Fuzzy Inference System (ANFIS)

Tarno¹, Subanar², Dedi Rosadi² and Suhartono³

¹ Statistics Department, Diponegoro University,
Semarang, Indonesia
² Mathematics Department, Gadjah Mada University,
Yogyakarta, Indonesia
³ Statistics Department, Institute of Technology Sepuluh November,
Surabaya, Indonesia

Abstract

The aim of this research is to analyze ANFIS performance for prediction of financial time series data. Financial time series data is usually characterized by volatility clustering, persistence, and leptokurtic data behavior. The financial time series data are usually non-stationary and non-linear. ARIMA has a good performance to predict linear time series data, but its performance is decreasing when applied to predict non-linear times series data. ANFIS as one of hybrid models which composes neural network (NN) and fuzzy system is expected to be able to predict the financial time series data more accurate. Many research conclude that the effectiveness of ANFIS depend on the input selection, the membership function (MF) selection and the rule generation. In this study, the input variables of ANFIS are selected based on preprocessing of original data by using Subset ARIMA model. The rule bases of ANFIS are generated based on a linear Sugeno fuzzy model. The consequent parameters are identified by using least square method and premise parameters are adapted by using gradient descent. For practical assessment of ANFIS performance, ANFIS method is implemented to analyze the Indonesia inflation monthly data from January 1970 up to December 2012.

Keywords: Financial Time Series, Fuzzy Logic, Neural Network, ANFIS.

1. Introduction

Financial time series data is usually characterized by volatility clustering, persistence, and leptokurtic data behavior [2], [13], [14], [20]. The financial time series data is usually non-stationary and non-linear [22]. In financial time series research, ARIMA Box-Jenkins is one of popular mean models [4], [19], [26]. While Autoregressive Conditional Heteroscedasticity (ARCH) proposed by Engle [21] and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) proposed by Bollerslev [25] are the popular variance models. Application of ARIMA-GARCH for modeling and forecasting financial time series data has been done by many researchers [22], [23]. Unfortunately, ARIMA-GARCH still has disadvantages when implemented for modeling and forecasting non-stationary and non-linear time series data.

In recent years, alternative models based on neural network (NN), fuzzy system and its hybrid have been improved to analyze non-stationary and non-linear time series data [7], [10], [24]. ANFIS method is one of the hybrid methods which combines neural network (NN) and fuzzy inference system (FIS) [6], [16]. The many examples of research on financial time series using ANFIS such as financial trading [1]; forecasting automobile sales [3]; prediction of stock market return [5], [11], [17]; prediction government bond yield [9]; forecasting EPS of leading industries [12]; financial volatility [13], [14] and exchange rate forecasting [15], [16], [23], conclude that the performance of ANFIS is better than the other methods.

The remaining parts of the article is organized as follows: Section 2 defines about ANFIS, types of membership function, reviews the architecture of ANFIS and learning algorithm in ANFIS. Section 3 describes the proposed procedure of ANFIS modeling. Section 4 discusses about results of empirical study. Section 5 summarizes the conclusion of the article.

2. Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS is a multilayer feed forward network. This architecture has five layers such as fuzzy layer, product...
layer, normalized layer, de-fuzzy layer and total output layer [6], [7]. The fixed nodes are represented by circle and the nodes represented by square are the adapted nodes. ANFIS gives the advantages of the mixture of neural network and fuzzy logic. The aim of mixing fuzzy logic and neural networks is to design an architecture which uses a fuzzy logic to show knowledge in fantastic way, while the learning nature of neural network to maximize its parameters [18].

2.1 Membership Function in ANFIS

A membership function (MF) is a curve which explains how every point in the input space is mapped to a membership degree between 0 and 1 [16]. Here, four types of membership functions will be used to identify the fuzzy inference system (FIS) parameters, i.e. triangular MF (trimf), trapezoidal MF (trapmf), Generalized Bell MF (gbellmf) and Gaussian MF (gaussmf). The formula of each membership function is given in Table 1 below.

Table 1. Formula of fuzzy membership functions [7]

<table>
<thead>
<tr>
<th>Membership function</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triangular MF (trimf)</td>
<td>( \text{trimf}(x;a,b,c) = \max \left( \min \left( \frac{x-1}{b-a}, \frac{c-x}{c-b} \right) 0 \right) )</td>
</tr>
<tr>
<td>Trapezoidal MF (trapmf)</td>
<td>( \text{trapmf}(x;a,b,c,d) = \max \left( \min \left( \frac{x-1}{b-a}, \frac{d-x}{d-c} \right) 0 \right) )</td>
</tr>
<tr>
<td>Generalized Bell MF (gbellmf)</td>
<td>( \text{gbellmf}(x;a,b,c) = \frac{1}{1 + \left( \frac{x-c}{a} \right)^2} )</td>
</tr>
<tr>
<td>Gaussian MF (gaussmf)</td>
<td>( \text{gaussmf}(x;c,\sigma) = \exp \left( -\frac{1}{2} \left( \frac{x-c}{\sigma} \right)^2 \right) )</td>
</tr>
</tbody>
</table>

2.2 ANFIS Architecture

For simplicity, given two inputs of time series data \( Z_{i,1}, Z_{i,2} \) and one output \( \hat{Z}_i \). Assumed that rule-base of Sugeno order-one with two rules as follows.

If \( Z_{i,1} \) is \( A_1 \) and \( Z_{i,2} \) is \( B_1 \), then \( Z_i^{(1)} = p_1 Z_{i,1} + q_1 Z_{i,2} + r_1 \).

If \( Z_{i,1} \) is \( A_2 \) and \( Z_{i,2} \) is \( B_2 \), then \( Z_i^{(2)} = p_2 Z_{i,1} + q_2 Z_{i,2} + r_2 \).

Here, \( Z_{i,1} \) is \( A_1 \) and \( Z_{i,2} \) is \( B_1 \); \( Z_{i,1} \) is \( A_2 \) and \( Z_{i,2} \) is \( B_2 \) are called as the premise section (nonlinear section), while \( Z_i^{(1)} = p_1 Z_{i,1} + q_1 Z_{i,2} + r_1 \) and \( Z_i^{(2)} = p_2 Z_{i,1} + q_2 Z_{i,2} + r_2 \) are called consequent (linear) section. \( p_1, p_2, q_1, q_2, r_1 \) and \( r_2 \) are linear parameters and \( A_1, B_1, A_2 \) and \( B_2 \) are called non-linear parameters. If the firing strength of each rules are \( w_1, w_2 \) for 2 values \( Z_i^{(1)} \) and \( Z_i^{(2)} \), then the output \( \hat{Z}_i \) is computed as weighted mean

\[
\hat{Z}_i = \frac{w_1 Z_i^{(1)} + w_2 Z_i^{(2)}}{w_1 + w_2} = \bar{w} Z_i^{(1)} + \overline{w} Z_i^{(2)}
\]

Architecture of neuro-fuzzy with all fundamental layers is illustrated as Fig.1 [7].

![Fig. 1. Structure of Fundamental Neuro-Fuzzy Network](image-url)

Neuro-fuzzy model as proposed by Jang [6] has 5 layers feed forward, which each layer is described as bellow.

Layer-1: Every node in the first layer is adaptive with one parametric activation function. The output is membership degree of given inputs which satisfy membership function \( \mu_{A_i}(Z_{i,1}) \), \( \mu_{B_i}(Z_{i,2}) \), \( \mu_{A_2}(Z_{i,1}) \) and \( \mu_{B_2}(Z_{i,2}) \). One example of fuzzy membership functions is generalized bell function (gbellmf),

\[
\mu_{A_i}(Z_{i,1}) = \frac{1}{1 + \left( \frac{Z_{i,1} - c}{a_i} \right)^2}, i=1,2;
\]

\[
\mu_{B_i}(Z_{i,2}) = \frac{1}{1 + \left( \frac{Z_{i,2} - c}{a_i} \right)^2}, i=1,2;
\]

where \( c \) is the location parameter and \( a_i \) is the shape parameter. If the value of parameter is change then the shape of bell-function is also change. The parameters in this layer are called premise parameters.

Layer-2: Every node in the second layer is fixed node, which the output is product of incoming signal. Generally, it uses fuzzy operation AND. Output of each
node represents firing strength $w_i$ of $i$-th rule. 

$$w_i = \mu_a(Z_{i1}) \times \mu_b(Z_{i2}), \quad i = 1, 2.$$  

Layer-3: Every node in the third layer is a fixed node, which compute ratio of firing strength of $i$-th rule relative to sum of firing strengths of rules, 

$$\overline{w}_i = \frac{w_i}{\sum_{j=1}^{2} w_j}, \quad i = 1, 2.$$  

This result is a normalized firing strength.

Layer-4: Every node in the fourth layer is adaptive node, the output of each node is 

$$\overline{w}_i Z_{i}^{(o)} = \overline{w}_i (p_i Z_{i1} + q_i Z_{i2} + r_i).$$  

$\overline{w}_i$ is a normalized firing strength of the third layer and 

$\{p_i, q_i, r_i\}$ is a set of this node parameters. Parameters in this layer are called consequent parameters (forward parameters).

Layer-5: Every node in the fifth layer is a fixed node which adds all of incoming signal. The output of fifth layer is the output of the whole network.

$$\hat{Z}_t = \sum_{i=1}^{2} \overline{w}_i Z_{i}^{(o)}.$$  

$$\hat{Z}_t = \sum_{i=1}^{2} \overline{w}_i (p_i Z_{i1} + q_i Z_{i2} + r_i) + \overline{w}_2 (p_2 Z_{i1} + q_2 Z_{i2} + r_2)$$  

$$\hat{Z}_t = p_1 (\overline{w}_1 Z_{i1}) + q_1 (\overline{w}_1 Z_{i2}) + r_1 \overline{w}_1 + p_2 (\overline{w}_2 Z_{i2}) + q_2 (\overline{w}_2 Z_{i2}) + r_2 \overline{w}_2$$  

(1)

3. The Proposed Procedure of ANFIS Modeling

This proposed procedure is extended based on the procedure which proposed by Wei et al. [12]. There are three basic step blocks of ANFIS modeling, those are preprocessing of data, rule generation and performance evaluation. These steps can be described completely as follows.

1. Collecting the financial time series data

The financial time series data are collected.

2. Preprocessing original data

The original data are preprocessed by using ARIMA method. The ARIMA models are constructed based on significant lags of ACF and PACF. The best model is selected by minimizing the root mean square of error (RMSE), AIC or SBC criterion.

3. Non-linearity test

Non-linearity test is needed to determine the non-linearity properties of data.

4. Determining the input variables

The input variables of ANFIS can be selected based on significant lag periods of ARIMA or Subset ARIMA. The all possible of significant lags yielded from informal preprocessing are selected as input variables of ANFIS.

5. Defining and partitioning the input variables

The selected input variables are classified into some clusters using Fuzzy C-means (FCM).

6. Setting the type of membership functions for input variables

We can set a membership function which can be applied to input variables. In this study, we use four types of membership functions such as triangular function \( (\text{trimf}) \), trapezoidal function \( (\text{trapmf}) \), generalized bell function \( (\text{gbellmf}) \), Gaussian function \( (\text{gaussmf}) \).

7. Generating the fuzzy If-Then rules

The output variables are associated with each input cluster based on the degree of possibility [16]. The fuzzy if-then rules are constructed using a linear Sugeno fuzzy model.

8. Training the parameters of fuzzy inference system (FIS)

FIS parameters are identified from training datasets. Fuzzy reasoning is used to infer new knowledge from identified base [16]. Consequent parameters are estimated using recursive least square method and premise parameters are adapted using back-propagation gradient descent.

9. Forecasting the training data and calculating RMSE value

After the significant models are constructed based on training data, we determine the predicted values and
calculate the RMSE values. The best model is selected by minimizing RMSE.

(10) **Forecasting the checking data and calculating RMSE value**

For assessment of ANFIS performance, the model result from previous steps is applied to forecast checking datasets and then calculate the RMSE values. The best model can be determined based on both RMSE values of training and checking data.

### 4. Results and Discussion

#### 4.1 Results

For practical implementation of proposed ANFIS model for forecasting financial time series data, Indonesia inflation data are used as case study. The proposed procedure of ANFIS modeling for forecasting the inflation data are used as case study. The proposed procedure of ANFIS modeling for forecasting the Indonesia inflation data are summarized as below.

1. **Collecting financial time series data**
   These data are observed monthly from January 1970 up to February 2012 and obtained from the Indonesia Central Bureau of Statistics (see [www.bps.go.id](http://www.bps.go.id)).

2. **Preprocessing original data**
   Before ANFIS method is employed to forecast the monthly Indonesia inflation data. The original data is preprocessed using ARIMA method. Based on significant lags of autocorrelation function (ACF) and partial autocorrelation function (PACF), we can identify the tentative models such as ARIMA \((1,1,0),(2,2,12,24)\), ARIMA \((1,3,12,0),(2,7,19,24)\) and ARIMA \((1,1,1,0),(2,7,12,24)\). When the residuals of fitted models are verified, all of the models contain ARCH effect and normality assumption of residuals are not satisfied. Based on non-linearity test using RESET test, the inflation data satisfy nonlinear properties. By RMSE, AIC or SBC criterion, ARIMA \((1,3,0),(2,7,12,24)\) is the best model among the tentative models yielded from pre-processing. Furthermore, the significant lags of ARIMA \((1,3,0),(2,7,12,24)\) are considered for determining input variables of ANFIS.

3. **Defining and partitioning of input variables**
   Based on ARIMA \((1,3,0),(2,7,12,24)\) result from preprocessing data, the lag-1 and lag-3 of AR term can be selected as input variables of ANFIS. Besides, by invertibility properties of ARIMA, lag-1, lag-3, lag-2, lag-7, lag-12 or lag-24 or it combinations can also be considered as input variables. In this study, the combination of significant lags, such as lag-1 and lag-3 \((Z_{t-1}, Z_{t-3})\); lag-1, lag-2 and lag-3 \((Z_{t-1}, Z_{t-2}, Z_{t-3})\); lag-1, lag-3 and lag-12 \((Z_{t-1}, Z_{t-3}, Z_{t-12})\) and lag-1, lag-3, lag-7 and lag-12 \((Z_{t-1}, Z_{t-3}, Z_{t-7}, Z_{t-12})\) are selected as inputs variables of ANFIS. The selected input variables are classified into 2, 3 or 4 clusters using FCM. The degree of membership function of input variables are determined by using four types of membership functions, such as triangular function (trimf), trapezoidal function (trapmf), generalized bell function (gbellmf), Gaussian function (gaussmf).

4. **Setting the type of membership function for output variables**
   A linear type of Sugeno fuzzy model is used to set the membership for output variables. For example, if \(Z_{t-1}\) is \(A_1\) and \(Z_{t-2}\) is \(B_2\), then \(Z_{t}^{(i)} = pZ_{t-1} + qZ_{t-2} + r_i\), \(i = 1, 2\);

where \(Z_{t-1}\) and \(Z_{t-2}\) are linguistic variables, while \(A_1\) and \(B_2\) are linguistic values, \(Z_{t}^{(i)}\) denotes the \(i\)-th output value.

5. **Generating the fuzzy If-Then rules**
   Fuzzy inference system is generated by using a linear Sugeno fuzzy model. This step is done for four types of membership functions respectively.

6. **Training parameters of fuzzy inference system (FIS)**
   In this step, the parameters of Fuzzy Inference System for training models are optimized by using recursive least square method and back-propagation gradient descent.

7. **Forecasting the training and the checking data and calculating RMSE values.**
   The predicted values are obtained using 48 types of models and RMSE values are calculated based on training data. These models are verified by using the checking data. The optimal model is selected by minimizing RMSE value of both training and checking data. The results of empirical study of Indonesia inflation data are shown in Table 2.

#### Table 2. Empirical Results

<table>
<thead>
<tr>
<th>Input</th>
<th>MFs</th>
<th>Cluster</th>
<th>RMSE Training</th>
<th>RMSE Checking</th>
</tr>
</thead>
<tbody>
<tr>
<td>((Z_{t-1}, Z_{t-3}))</td>
<td>trimf</td>
<td>[2, 2]</td>
<td>1.3425</td>
<td>35.3753</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3, 3]</td>
<td>1.2772</td>
<td>31.7801</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4, 4]</td>
<td>1.2484</td>
<td>97.9147</td>
</tr>
<tr>
<td></td>
<td>trapmf</td>
<td>[2, 2]</td>
<td>1.3107</td>
<td>2.3659</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3, 3]</td>
<td>1.3029</td>
<td>2.8290</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4, 4]</td>
<td>1.2676</td>
<td>9.4467</td>
</tr>
<tr>
<td></td>
<td>gbellmf</td>
<td>[2, 2]</td>
<td>1.3233</td>
<td>3.8086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3, 3]</td>
<td>1.2688</td>
<td>38.6644</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4, 4]</td>
<td>1.2589</td>
<td>55.1295</td>
</tr>
<tr>
<td></td>
<td>gaussmf</td>
<td>[2, 2]</td>
<td>1.3259</td>
<td>2.8368</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3, 3]</td>
<td>1.2759</td>
<td>15.3602</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4, 4]</td>
<td>1.2574</td>
<td>95.2298</td>
</tr>
</tbody>
</table>

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.
In this study, we set four combinations of input variables \((Z_{t-1}, Z_{t-3}, Z_{t-12})\) and lag-1, lag-3, lag-7 and lag-12 \((Z_{t-1}, Z_{t-3}, Z_{t-7}, Z_{t-12})\). Every input variable is classified into 2, 3 or 4 clusters. Four types of membership functions are applied to each combination of input. So, we have 48 types of ANFIS models which should be evaluated. The evaluation to these models are classified into four groups based on the combination of inputs. It can be summarized as follows.

- Corresponding to input variables \((Z_{t-1}, Z_{t-3})\), optimal RMSE are 1.3020 for training data and 2.8290 for checking data. These values related to the number of clusters [3,3] and trapezoidal membership function.
- Corresponding to input variables \((Z_{t-1}, Z_{t-3}, Z_{t-12})\), optimal RMSE are 1.1454 for training data and 1.4679 for checking data. These values related to the number of clusters [3,3,3] and trapezoidal membership function.
- Corresponding to input variables \((Z_{t-1}, Z_{t-3}, Z_{t-7}, Z_{t-12})\), optimal RMSE are 0.99922 for training data and 2.4467 for checking data. These values related to the number of clusters [3,3,3] and trapezoidal membership function.
- Corresponding to input variables \((Z_{t-1}, Z_{t-3}, Z_{t-7}, Z_{t-12})\), optimal RMSE are 0.72396 for training data and 4.9961 for checking data. These values related to the number of clusters [3,3,3] and trapezoidal membership function.

If we compare among minimal RMSE values of 4 groups, we obtain the optimal RMSE are 1.1454 for training data and 1.4679 for checking data related to input variables \((Z_{t-1}, Z_{t-3}, Z_{t-7}, Z_{t-12})\), the number of clusters [3,3,3] and the type of membership function is trapezoidal MF.

### 5. Conclusion

The accuracy of ANFIS model depends on many factors. There are three important roles which can determine the accurate model, i.e. selection of input variables, determining the number of clusters and selection of membership function. Based on the results of empirical study of Indonesia inflation data, we can obtain the optimal ANFIS model with three input variables \((Z_{t-1}, Z_{t-3}, Z_{t-7})\), the number of clusters [3,3,3] and trapezoidal membership function. Using this model, Indonesia inflation data can be predicted accurately.

### References

Model, World Academy of Science, Engineering and Technology 58, 2009.


