

The Fuzzy Inference System with Least Square Optimization for Time Series Forecasting

Samingun Handoyo¹, Marji²

¹Department of Statistics, Universitas Brawijaya Malang, Indonesia

²Department of Informatics Engineering, Universitas Brawijaya Malang, Indonesia
Jalan Veteran Malang 65145

Article Info

Article history:

Received Mar 29, 2018

Revised May 3, 2018

Accepted Jun 6, 2018

Keywords:

Forecasting

Fuzzy inference system

Rule base optimized

Time series

ABSTRACT

The rule base on the fuzzy inference system (FIS) has a major role since the output generated by the system is highly dependent on it. The rule base is usually obtained from an expert but in this study proposed the rule base generated based on input-output data pairs with generating rule bases using lookup table scheme, then consequent part of each rule optimized with ordinary least square (OLS), so finally formed rule base from model FIS Takagi-Sugeno zero order. The exchange rate dataset of EURO to USD is used for the development and validation of the system. In this study, 12 FISs were developed from a combination of linguistic values of $n = 3, 5, 7, 9$ with the number of lag (k) assumed to have an effect on output for $k = 2, 3, 5$. In training data, the values of R^2 ranged between 0.989 and 0.993, MAPE values ranged between 0.381% and 0.473% where the FIS with $n = 9$ and $k = 5$ has the best performance. In the testing data, the values of R^2 ranged between 0.203 and 0.7858, MAPE values ranged between 0.5136% and 0.9457% where the FIS with $n = 3$ and $k = 2$ perform the best.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Samingun Handoyo,

Department of Statistics, Universitas Brawijaya Malang, Indonesia,

Jalan Veteran Malang 65145, phone +62341571142.

Email: samistat@ub.ac.id

1. INTRODUCTION

Appropriate decision making on an event will have a broad impact on the success of the management. Today there has been a growing number of quantitative methods that can support the decision-making process that has even penetrated or obscured the boundaries of the scientific field. Intelligent computing methods is an option to avoid limitations in conventional quantitative methods laden with assumptions that must be met.

Some of the intelligent computing methods that are currently being applied in many fields are fuzzy logic or fuzzy system or Fuzzy Inference System (FIS). FIS is not only much in demand by some groups of researchers from the fields of engineering, applied sciences, science, and computer science but also has been widely implemented in the field of industry and production. Some applied FIS include power management control to improve the economic and technical performance of a power source by Doaa et al. [1], Arbab et al. [2] identify non-linear functions based on fuzzy logic models with improved gradient search. Expert systems for hypertension use a combination of C5 algorithms, and fuzzy logic [3]. Based on these three studies the fuzzy system has a satisfactory performance. On the other hand, some researchers also adopt hybrid methods by Azriyenni et al. [4] combining fuzzy systems with neural networks for classification of failures in protection systems, and adaptive methods by Abdulrahman and Imad [5], developing a iTraffic simulation based on dynamic fuzzy logic.

Mature planning will greatly affect the process of execution of work or activities to fit the targets that have been set. Knowledge of the circumstances that will occur in the future will support the success of a plan. Time factor affects the events that occur in the future. In the real world, there are many data types collected based on certain time interval known as time series data. Application of fuzzy logic in time series data is often called Fuzzy Time Series (FTS). The main component of FTS is a Fuzzy Logic Relationship (FLR). The output of the FTS model is obtained based on the input data matched with the antecedence part of the FLR, after obtaining an FLR having the antecedent part equal to the inputted data the consequent part of the FLR is automatically generated as the system output [6-12] Usually the FLR is arranged based on one or two lags are considered to affect the output. FTS has been successfully applied to time series data with small amounts of observation, but on data with large numbers of observation, the FTS application is constrained by complexity in the formation of FLR and slow computing process.

Fuzzy systems or FIS which originally developed in the field of control techniques have components for inference, this is what distinguishes it from FTS [13]. The vital component of the FIS is called the fuzzy rule base. Originally this rule base was obtained from an expert in his field [14], but due to many limitations in obtaining the expert fuzzy rule base, the role of the expert is replaced by the generated fuzzy rule base on the input-output pairs data. Several methods have been proposed to generate the fuzzy rule base from the input-output pairs data, namely the establishment of the fuzzy rule base with the fuzzy association rule by Guofang and Yuanchen [15], and the generation of the fuzzy rule base with the mining a priori association rule technique by Muhammad et al. [16], and as well as generating fuzzy rule base with lookup table schemes by Samingun and Ari [14]. Although FIS has been successfully implemented in various ways of generating rule base, but FIS performance for time series forecasting is still very unsatisfactory.

Various attempts were made to improve the performance of the FIS for forecasting among others by forming a neuro-fuzzy hybrid model as performed by Mohammad and Mohammad [17], Mahmut and Mahmud [18] applying a neuro-fuzzy adaptive model to predict hydrological data. Stock market return forecast using ANFIS by Melek and Derya [19]. Angga and Samingun [20] apply FIS for time series forecasting by optimizing membership function using a genetic algorithm. However, these hybrid systems also perform poorly because they only combine two methods directly, and do not to the specific realm to improve the fuzzy rule base of the FIS. In this research, it is proposed that the optimization technique is the most widely used that is Ordinary Least Square (OLS) to produce the better fuzzy rule base, although Samingun et al. [21] have done optimization of parameter model estimation using Particle Swarm Optimization (PSO) with better result than OLS, the fuzzy rule base of the FIS is a linear system, so the OLS is considered appropriate.

Based on the above explanation, this research proposes FIS Takagi-Sugeno zero order with a fuzzy rule base optimized by using OLS. The FIS was built and validated using the exchange rate of EURO to USD dataset. The fuzzy rule base will be generated based on input-output pairs data using lookup table scheme, then consequent part of each rule is constant value resulted using OLS optimization. As many as 12 FIS will be derived from the combination of linguistic values $n = 3, 5, 7, 9$, and the number of lags $k = 2, 3, 5$. The performance of all FIS will be evaluated using MAPE and R^2 value indicators to investigate the influence of the number of linguistic values and the number of lags to the system performance.

2. RESEARCH METHOD

In this section will be explained theoretically the methods used in this study which include the lookup table scheme, fuzzy system of Takagi-Sugeno zero order optimized with OLS (the proposed method) and indicators of system performance measurement.

2.1. Rule bases generation with lookup table scheme

The fuzzy rule base is ideally sourced from the experts. The basic rule of fuzzy logic in the form of an "IF-THEN" implication consists of two parts, the antecedent, and the consequent. The antecedent part is the proportion or composition of the proportions that follow the "IF" part, whereas the consequent part is the part following "THEN" in the form of a proposition or in the form of a linear equation or a constant. Consider a rule like the following:

$$IF (x_{1,n} \text{ is } A_1^l) \text{ AND } \dots \text{ AND } (x_{p,n} \text{ is } A_p^l) \text{ THEN } y_{1,n} \text{ is } B^l \quad (1)$$

The rule in equation (1) can represent a system consisting of as many p-inputs (X_1, X_2, \dots, X_p) and an output (Y). The system inputs act as proposition in the antecedent part, while the system output acts as a proposition in consequent part. All inputs and output have observations of n, so the input-output pairs can be

expressed in an $n * (p + 1)$ matrix such as in Equation (2). Each input-output variables have a linguistic value of l in the form of a fuzzy set [14].

$$A = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1,k} & y_1 \\ x_{21} & x_{22} & x_{23} & & x_{2,k} & y_2 \\ & \vdots & & \ddots & \vdots & \\ x_{t1} & x_{t2} & x_{t3} & \dots & x_{t,k} & y_t \end{bmatrix} \tag{2}$$

Suppose given a sample of the input-output pairs $x_{11}, x_{21}, \dots, x_{p1}; y_1$ or briefly expressed as (x_0^p, y_0^p) where $x_0^p \in U = [\alpha_1, \beta_1]x[\alpha_2, \beta_2]x \dots x[\alpha_p, \beta_p]$ and $y_0^p \in V = [\alpha_y, \beta_y]$, the values of $[\alpha_1, \beta_1]$ are the range of observed values of the variable X_1 , and $[\alpha_y, \beta_y]$ are the range of values of variable Y . The following will be described how to generate the bases of fuzzy rule based on input-output pairs data [22].

1. Determine the fuzzy set which includes input space (U) and output space (V)

Sets the linguistic value for each input variable as $A^l, l = 1, 2, \dots, n$ in the form of a fuzzy set that satisfies $x_i \in [\alpha_i, \beta_i]$ and for the output variable sets linguistic value as $B^l, l = 1, 2, \dots, n$ satisfying $y \in [\alpha_y, \beta_y]$ related to certain membership function. Input-output transformation into linguistic value (fuzzy set) called fuzzification process. Each observation value in the form of a crisp number is transformed into a fuzzy set of linguistic value and degree of membership. This process can be done if it has been established several things, namely:

- a) The maximum and minimum values of each input-output variable.
- b) The number of fuzzy sets (linguistic values) desired.
- c) Membership function of each fuzzy set.
- d) Parameters of each membership function of each fuzzy set.

Each observation value of the input-output variables is calculated the degree of membership in each fuzzy set, and then search for the highest degree of membership in the linguistic value of each variable. The result of a fuzzification process is a pairs of linguistic value and degree of membership.

2. Create one rule from one input-output pairs known as a rule candidate

The fuzzification results are rearranged in the input-output matrix format so that two input-output matrix pairs are a matrix whose members are linguistic values (matrix S) and matrix whose members are degree of membership (matrix U). Both matrices in equation (3) have the same order as the matrix order of the initial input-output pairs (matrix A). If each row of the matrix S is read using the implication proposition in which columns 1 to k-1 are the antecedent parts and the column of k is the consequent part, then each row in S matrix forms a rule called candidate rule. Since the rows number of matrix S are equal to the number of rows of matrix A then it is obtained number of candidate rules as much as the number of input-output pairs. The U matrix will be used to calculate the degree of each rule candidates.

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} & \dots & s_{1,k} & s_{y1} \\ s_{21} & s_{22} & s_{23} & & s_{2,k} & s_{y2} \\ & \vdots & & \ddots & \vdots & \\ s_{t1} & s_{t2} & s_{t3} & \dots & s_{t,k} & s_{yt} \end{bmatrix} \text{ and } \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \dots & \mu_{1,k} & \mu_{y1} \\ \mu_{21} & \mu_{22} & \mu_{23} & & \mu_{2,k} & \mu_{y2} \\ & \vdots & & \ddots & \vdots & \\ \mu_{t1} & \mu_{t2} & \mu_{t3} & \dots & \mu_{t,k} & \mu_{yt} \end{bmatrix} \tag{3}$$

s_{t1} = the linguistic value (fuzzy set) of the 1st input, and the input-output pair of t
 μ_{t1} = degree of membership of linguistic value s_{t1}

3. Establish the degree of each rule candidates formed

Matrix U is used to calculate the degrees of the rule candidates. The degree is useful to overcome the prospective rule candidates of conflict. Two or more rule candidates are called a set of conflict rule candidates if they have the same antecedent parts but the consequent part is different. A rule candidate with the highest degree is selected as a member of the fuzzy rule bases. Suppose the degree of the first row of matrix U is:

$$\text{Degree}(1) = \mu_{11} * \mu_{12} * \dots * \mu_{1,k-1} * \mu_{y1} \tag{4}$$

The degree of a rule candidate is the multiplication of all membership values per observation on the line as expressed in equation (4). Based on the rule candidate matrix (matrix S) obtained at the previous step, the final fuzzy rule is selected by:

- a) Rule candidates that have both the antecedent and the consequence parts are the same represented by one rule as a member of the fuzzy rule base.
- b) A unique rule candidate automatically becomes a member of the fuzzy rule base.

- c) The set of the conflict rule candidates are selected the rule candidate with the greatest degree as a member of the fuzzy rule base.

2.2. Fuzzy Inference System of Takagi-Sugeno

Consider l as the number of rules on a fuzzy rule base, l ($l = 1, 2, \dots, L$) shows the index of rule and k ($k = 1, 2, \dots, p$) indicates the input index. According to Ajoy and Dobrivoje [23], the inference process on Fuzzy Takagi-Sugeno model is divided into 2 stages. The first stage calculates the degree of activation (fire strength) for each rule on the fuzzy rule base. The second stage is the output calculation of the crisp value obtained by using the average defuzzification method. In general, the rule base of the Fuzzy Takagi-Sugeno model is as follows [16]:

$$\text{Rule}^l : \quad \text{IF } (x_{t1} \text{ is } A_1^l) \text{ AND } (x_{t2} \text{ is } A_2^l) \dots \text{ AND } (x_{t,k-1} \text{ is } A_k^l) \text{ THEN} \quad (5)$$

$$y_t^l = \theta_0^l + \theta_1^l x_{t,1} + \dots + \theta_p^l x_{t,k}$$

for $l = 1, 2, \dots, L$

where $x_{t,k}$ is the value of the k -th input variable, y_t^l is the local output (output generated by the l -th rule), θ_0^l, θ_k^l are the consequent parameters of rule l , and A_k^l is the linguistic value (fuzzy set) of the k -th input variable for the l -th rule represented by a membership function $\mu_{A_k^l}(x_{t,k})$. In this study the representation of the membership function of the fuzzy set is Gaussian (gaussMF) as follows:

$$\mu_{A_k^l}(x_{t,k}) = \exp\left(-\frac{1}{2}\left(\frac{x_{t,k}-c_k^l}{\sigma_k^l}\right)^2\right) \quad (6)$$

where c_k^l and σ_k^l are respectively the parameters of the center (mean) and spread (standard deviation) of gaussMF, corresponding to the antecedent part for the l -th rule. The fire strength of l -th rule for the t -observation is calculated by the multiplication operator as follows:

$$\alpha_t^l = \prod_{k=1}^p \mu_{A_k^l}(x_{t,k}) \quad (7)$$

where p is the number of input variables. To calculate the output as a crisp value, the fire strength is then normalized using the following formula:

$$\bar{\alpha}_t^l = \frac{\alpha_t^l}{\sum_{l=1}^L \alpha_t^l} = \frac{\prod_{k=1}^p \mu_{A_k^l}(x_{t,k})}{\sum_{l=1}^L \prod_{k=1}^p \mu_{A_k^l}(x_{t,k})} \quad (8)$$

The system output is \hat{y}_t (the value of crisp) which is the predicted value on the t -observation calculated by using equation (9) as follows:

$$\hat{y}_t = \sum_{l=1}^L \bar{\alpha}_t^l y_t^l = \sum_{l=1}^L \bar{\alpha}_t^l (\theta_0^l + \theta_1^l x_{t,1} + \dots + \theta_p^l x_{t,p}) \quad (9)$$

The equation (9) is a defuzzification process with weighted average method. The weights used in the Takagi-Sugeno model are normalized fire strength ($\bar{\alpha}_t^l$), consequent part (y_t^l) is a constant number for FIS Takagi-Sugeno of zero order, or in the form of a linear equation for FIS Takagi-Sugeno of one order [22].

2.3. The Proposed Method

In this research, it is proposed a method of FIS Takagi-Sugeno of zero order by establishing constants on the consequent part through Ordinary Least Square (OLS) optimization. The FIS Takagi-Sugeno of zero order based on equation (9) can be written as follows:

$$\hat{y}_t = \bar{\alpha}_t^1(\theta_0^1) + \bar{\alpha}_t^2(\theta_0^2) + \dots + \bar{\alpha}_t^L(\theta_0^L) \quad (10)$$

If there are as many n input-output pairs of $t = 1, 2, \dots, n$, then equation (10) can be expressed in equation (11) as follows:

$$\begin{aligned}
 \hat{y}_1 &= \bar{\alpha}_1^1(\theta_0^1) + \bar{\alpha}_1^2(\theta_0^2) + \dots + \bar{\alpha}_1^L(\theta_0^L) \\
 \hat{y}_2 &= \bar{\alpha}_2^1(\theta_0^1) + \bar{\alpha}_2^2(\theta_0^2) + \dots + \bar{\alpha}_2^L(\theta_0^L) \\
 &\dots \quad \dots \quad \dots \\
 \hat{y}_n &= \bar{\alpha}_n^1(\theta_0^1) + \bar{\alpha}_n^2(\theta_0^2) + \dots + \bar{\alpha}_n^L(\theta_0^L)
 \end{aligned}
 \tag{11}$$

It is obvious that the predicted value of the first input pair is obtained from the linear combination of fire strength in the first row ie $\bar{\alpha}_1^l = (\bar{\alpha}_1^1, \bar{\alpha}_1^2, \dots, \bar{\alpha}_1^L)$ and the vector of parameter $\theta_0^l = (\theta_0^1, \theta_0^2, \dots, \theta_0^L)$, as well as the predicted value of the second input pair is obtained from the linear combination of fire strength in the second row ie $\bar{\alpha}_2^l = (\bar{\alpha}_2^1, \bar{\alpha}_2^2, \dots, \bar{\alpha}_2^L)$ and the parameter vector $\theta_0^l = (\theta_0^1, \theta_0^2, \dots, \theta_0^L)$, and so on. Thus a prediction value can be obtained, provided that the vector of parameter value θ_0^l has been obtained, so in this case, the output of the system depends on the value of the vector parameter θ_0^l .

Consider if equation (11) is expressed in matrix notation as follows:

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_1^1 & \bar{\alpha}_1^2 & \bar{\alpha}_1^3 & \dots & \bar{\alpha}_1^L \\ \bar{\alpha}_2^1 & \bar{\alpha}_2^2 & \bar{\alpha}_2^3 & \dots & \bar{\alpha}_2^L \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{\alpha}_n^1 & \bar{\alpha}_n^2 & \bar{\alpha}_n^3 & \dots & \bar{\alpha}_n^L \end{bmatrix} * \begin{bmatrix} \theta_0^1 \\ \theta_0^2 \\ \vdots \\ \theta_0^L \end{bmatrix}
 \tag{12}$$

If the left-hand side of equation (12) is substituted with the actual output data then the multiple linear regression model will be obtained in the matrix notation as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}
 \tag{13}$$

Both sides of the equation (13) multiplied by \mathbf{X} transpose, the solution will be obtained in equation (14) as follows:

$$\begin{aligned}
 t(\mathbf{X}) * \mathbf{y} &= (t(\mathbf{X}) * \mathbf{X}) * \boldsymbol{\theta} \\
 \boldsymbol{\theta} &= (t(\mathbf{X}) * \mathbf{X})^{-1}(t(\mathbf{X}) * \mathbf{y})
 \end{aligned}
 \tag{14}$$

The parameter vector solution θ in equation (14) above is a solution of the Ordinary Least Square (OLS) optimization method in which:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \bar{\alpha}_1^1 & \bar{\alpha}_1^2 & \bar{\alpha}_1^3 & \dots & \bar{\alpha}_1^L \\ \bar{\alpha}_2^1 & \bar{\alpha}_2^2 & \bar{\alpha}_2^3 & \dots & \bar{\alpha}_2^L \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{\alpha}_n^1 & \bar{\alpha}_n^2 & \bar{\alpha}_n^3 & \dots & \bar{\alpha}_n^L \end{bmatrix}, \text{ dan } \boldsymbol{\theta} = \begin{bmatrix} \theta_0^1 \\ \theta_0^2 \\ \vdots \\ \theta_0^L \end{bmatrix}$$

Parameter θ is a vector consisting of constants that act as a consequent part of the fuzzy rule base of the FIS Takagi Sugeno of zero order.

2.4. Indicators of the System Performance measurement

Assessment of system performance in this study using 2 indicators namely MAPE (Mean Absolute Percentage Error) and R^2 (Coefficient of Determination). These two values are considered very firm and able to provide standard criteria for a good forecasting system or model. The following are given formulas for calculating MAPE and R^2 [14]:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\% \quad \text{and} \quad R^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}$$

The value of MAPE = 0 means that the system output perfectly has the same value as the actual value. The value of MAPE is an interval of [0,100] with percentage unit. The value of $R^2 = 1$ means the system output has accuracy in predicting the actual value is 100%, or all the system output values are equal to the actual values. The value of R^2 has a range of values [0,1]. In this study, the value of R^2 is calculated from the square of the correlation value between the actual and predicted values. Use of this R^2 must be careful because two data are far apart but have a parallel pattern will have the value of R^2 closer to one. Therefore in its use, the R^2 indicator must be accompanied by a plot between the actual and predicted values [14].

3. RESULTS AND ANALYSIS

In this section is discussed the results obtained from this study include descriptions of data with the distribution of training and testing data, parameter settings of the membership function of the fuzzy set used in fuzzification, generating fuzzy rule base with lookup table scheme and consequent part of rules are optimized with OLS, presents the results of FIS implementation on training and testing data, and measures system performance.

3.1. Data Description

The exchange rate of EURO to USD dataset was observed as many as 350 observation points in the period February 2014 until August 2015. The original dataset (all observations) are divided into two parts, ie the first 300 observations are used for modeling which are often referred to as training data and the remaining 50 the last observations are used as cross-validation data or model testing or referred to as testing data. Figure 1 shows the existing pattern on the whole observations (the top of Figure 1). Broadly speaking the data have a negative linear trend with a relatively constant variance. Also visible if the data are plotted in more detail as at the bottom of Figure 1, the actual data pattern tends to be non-linear with a non-constant variance. Training data (the center part of Figure 1) are used to build the sytem or model, while testing data are used for model validation. In this paper, the development of the FIS Takagi-Sugeno of zero order based on fuzzy rule bases established using lookup table scheme. System specifications are based on a combination of many lags that are considered to have an effect on the output and the number of linguistic values (fuzzy sets) used.

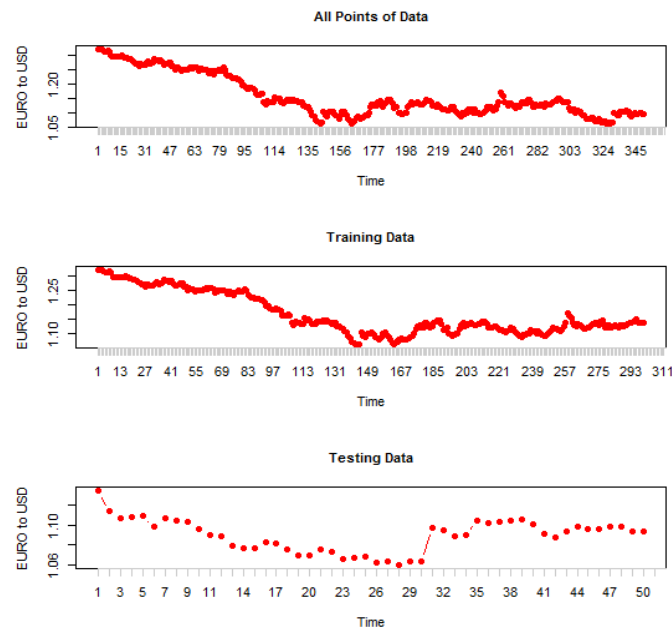


Figure 1. Plot daily exchange rates of EURO to USD for all records, training data, and testing data

3.2. Parameters Setting of Membership Function

In Generating the fuzzy rule base with a lookup table scheme, each observational value must be transformed into a membership degree on a linguistic value (fuzzy set). Selection of membership function and its parameter setting are important steps for fuzzification process. Suppose there are 3 linguistic values that are Small (S), Medium (M), and Big (B). An arrangement of parameters of each membership function is needed in order to the degree of membership of an observational value on each linguistic value can be calculated. The membership of an observation on the fuzzy set is determined by finding the highest degree of membership of the observation in all linguistic values.

In this study, the membership function used is Gaussian (gaussMF), so it must be done setting both parameters of the mean value (μ), and standard deviation (std) of each fuzzy set (linguistic value). The number of linguistic values to be used are 3, 5, 7, and 9. Table 1 states the values of both parameters for all linguistic values.

Table 1. Parameters of the Gaussian Membership Function of the Training Data

No	n=3		n=5		n=7		n=9	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	0.0776	1.0594	0.0388	1.0594	0.0259	1.0594	0.0194	1.0594
2	0.0776	1.2147	0.0388	1.137	0.0259	1.1112	0.0194	1.0982
3	0.0776	1.3699	0.0388	1.2147	0.0259	1.1629	0.0194	1.137
4			0.0388	1.2923	0.0259	1.2147	0.0194	1.1758
5			0.0388	1.3699	0.0259	1.2664	0.0194	1.2147
6					0.0259	1.3182	0.0194	1.2535
7					0.0259	1.3699	0.0194	1.2923
8							0.0194	1.3311
9							0.0194	1.3699

Based on Table 1 it can be seen that the minimum value is always the mean of the lowest linguistic value, while the maximum value always being the mean of the highest linguistic value. The standard deviation value is derived from $std = (max-min) / (n-1)$ where n is the number of linguistic values. The mean value of the second linguistic value is $\mu_2 = \min + std$, the mean value of the third linguistic value is $\mu_3 = \min + 2 * std$, and so on. Giving the name of linguistic value using the terms Small (S), Medium (M), and Big (B) for the number of linguistic values of n = 3, while for n = 7 linguistic values are S3, S2, S1, M, B1, B2, and B3.

3.3. Generating of Fuzzy Rule Bases

The fuzzy rule base is generated based on the input-output pairs with the lookup table scheme will initially always form rule candidates as much as the number of input-output pairs data. The training data of 300 observations will be formed the matrices input-output pairs with the dimension of 298x3, 297x4, and 295x6 respectively for the number of lag (k) values that affect the output are k = 2, 3, and 5. In the matrix, the column of k + 1 is output, while the other columns are the input. Using each input-output pair will be generated a rule candidate, so the number of rule candidates will be equal to the number of rows of the input-output pairs matrix with the addition of the column of k + 2 is the degree of each rule. By using the lookup table scheme criteria, the number of rules in each fuzzy rule base are given in Table 2 as follows:

Table 2. Number of Rules on the Fuzzy Rule Bases of 12 FIS

n	k=2		k=3		k=5	
	rows	columns	rows	columns	rows	columns
3	6	4	11	5	27	7
5	11	4	20	5	45	7
7	17	4	31	5	63	7
9	21	4	36	5	77	7

Table 2 is the matrix order of each fuzzy rule base formed, the first column of Table 2 represent the number of linguistic values (n), while the other columns state the number of rules followed by the number of columns of each rule base. Suppose that the linguistic value of n = 5 and lag k = 2 are produced 11 rules with the number of columns is 4 ie columns 1 and 2 are inputs, the third column is the output and the fourth column is the degree of each rule, whereas at n = 5 and k = 3 are generated 20 rules with 5 columns, ie columns 1 to 3 are inputs, column 4 is output, and column 5 is the degree of each rule.

The complete of the fuzzy rule bases from the combination of the number of linguistic values (n) and the number of lag (k) for n = 5, k = 2, and for n = 5, k = 3 are presented in Table 3

The number of rules on the fuzzy rule base with n = 5 and k = 2 are 11 rules, while the number of rules on the fuzzy rule base with n = 5 and k = 3 are as much as 20 rules as exposed in Table 2. Consider of the line 11 in Table 2 The 11th line of the fuzzy rule base for n = 5 and k = 2, can be expressed in terms of the following implication proposition:

“IF input1 is B1 and input2 is B2 THEN Y is B1”,

while the implication proposition of the fuzzy rule base for n = 5 and k = 3 is:

“IF input1 is S1 and input2 is S1 and input3 is M THEN Y is S1”.

Both of the above rules are the form of the rule of the Mamdani FIS model characterized by its consequent part in the form of linguistic value (fuzzy set). The FIS Takagi-Sugeno of zero order model has

the fuzzy rule base which a consequent part is a constant. Therefore a systematic attempt to replace linguistic values on consequent parts of the fuzzy rule base with constants must be made. By transforming each observation of training data into membership degrees by fuzzifying each values of inputs in accordance with the linguistic values that exist in each rule. So the input values are fitted to all rules that eventually generate a vector whose length is equal to the number of rules. These vector elements are normalized by dividing them by the total of all elements. As Illustration on the fuzzy rule base for $n = 5$ and $k = 2$ will form a transformation matrix with dimensions of 298×11 , in which the order denotes the number of input variables as columns and the number of rules as rows. It should be emphasized that the elements of this matrix are the fire strength normalized of each rule.

Reconsider the line number 11 of Table 2, the rule of the Takagi-Sugeno model of the zero order formed is:

“IF input1 is B1 and input2 is B2 THEN Y is 1.3128”,

on the fuzzy rule base for $n = 5$ and $k = 2$, while on the fuzzy rule base for $n = 5$ and $k = 3$, the established rule is:

“IF input1 is S1 and input2 is S1 and input3 is M THEN Y is 1.1454”.

So the fuzzy rule base produced on the system with $n = 5$ and $k = 2$ are as many as 11 rules in which the antecedent parts are located in the columns X1 and X2, while the consequent parts are obtained by replacing column Y with column B. As exposed on the right segment of table 2 that the system with $n = 5$ and $k = 3$ are produced as many as 20 rules in which the antecedent parts are located on columns X1, X2, and X3, whereas the consequent parts of this fuzzy rule base are obtained by replacing column Y with column B. The generation of fuzzy rule bases are performed on all possible combinations of linguistic values of $n = 3, 5, \text{ and } 7$, and the number of lags $k = 2, 3, \text{ and } 5$.

Table 3. The Fuzzy Rule Bases of the Linguistic Value of $n = 5$ and the lags $k = 2$ and $k = 3$

No	Combination of $n=5$ and $k=2$				Combination of $n=5$ and $k=3$						
	X1	X2	Y	Degree	B	X1	X2	X3	Y	Degree	B
1	2	2	2	0.9995	1.1504	2	2	2	2	0.9990	1.1486
2	5	5	5	0.9990	1.3739	5	5	5	5	0.9947	1.3796
3	1	1	1	0.9903	1.0637	4	4	4	4	0.9878	1.2953
4	4	4	4	0.9895	1.2989	1	1	1	1	0.9846	1.0523
5	3	3	3	0.9803	1.2181	3	3	3	3	0.9628	1.1999
6	2	1	1	0.4896	1.1517	2	1	1	1	0.4885	1.1393
7	2	3	2	0.4020	1.1283	1	1	2	1	0.3759	1.1159
8	1	2	1	0.3949	1.0221	1	2	1	1	0.3724	1.0400
9	3	4	3	0.3661	1.205	2	2	1	2	0.3697	1.1789
10	4	3	3	0.2921	1.2874	3	3	4	3	0.3382	1.2531
11	4	5	4	0.2530	1.3128	2	2	3	2	0.3301	1.1454
12						1	2	2	1	0.3165	1.0045
13						2	3	3	2	0.2973	1.1616
14						3	4	3	3	0.2542	1.2063
15						3	4	4	4	0.2406	1.1872
16						4	3	4	4	0.2065	1.2666
17						4	3	3	3	0.2001	1.294
18						4	4	5	4	0.1927	1.3589
19						4	4	3	4	0.1750	1.3199
20						4	5	5	4	0.1735	1.2775

3.4. Implementation of FIS Takagi-Sugeno of zero order

In this session will be implemented 12 FIS Takagi-Sugeno of zero order yielded by a combination of the number of linguistic values of $n = 3, 5, 7, \text{ and } 9$ with the number of lags assumed to affect the output of $k = 2, 3, \text{ and } 5$. Furthermore use the abbreviation FIS32 for the combination of $n = 3$ and $k = 2$, FIS33 for the combination of $n = 3$ and $k = 3$, and so on until FIS95 for the combination of $n = 9$ and $k = 5$.

Implementation of 12 FIS Takagi-Sugeno zero order in training data produce output very close to the actual value. An illustration shows the actual values versus prediction results in FIS52 shown in Figure 2. The top of Figure 2 is a plot of 298 actual versus predicted values, while Figure 2 below plots the actual versus the prediction values of the last 50 observed values in training data.

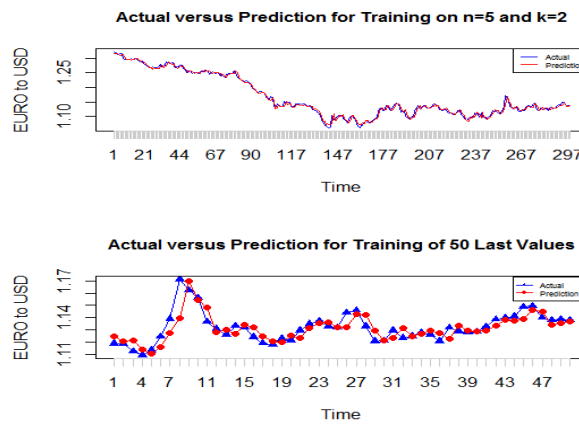


Figure 2. Plot the actual value versus the predicted value of the training data in FIS52

To clarify the system output in the training data, the following 5 recent records of the actual values, predictions, and absolute errors of FIS52, FIS53, FIS93, and FIS95 are shown in Table 3 as follows:

Table 3. Comparison of actual(Yact.), predicted(Ypred.), and absolute error(Abs.err.) values of FIS52, FIS53, FIS93, and FIS95 on 5 Recent Values of Training Data

FIS52			FIS53			FIS93			FIS95		
Yact	Ypred	Abs.err	Yact.	Ypred	Abs.err	Yact	Ypred	Abs.err	Yact	Ypred	Abs.err
1.1494	1.1462	0.0032	1.1494	1.1464	0.003	1.1494	1.148	0.0014	1.1494	1.1475	0.0019
1.1398	1.1447	0.0049	1.1398	1.1457	0.0059	1.1398	1.1452	0.0054	1.1398	1.1458	0.0060
1.1378	1.1344	0.0034	1.1378	1.1367	0.0011	1.1378	1.1327	0.0051	1.1378	1.1312	0.0066
1.1386	1.1357	0.0029	1.1386	1.1373	0.0013	1.1386	1.1362	0.0024	1.1386	1.1307	0.0079
1.1377	1.1369	0.0008	1.1377	1.1373	0.0004	1.1377	1.1373	0.0004	1.1377	1.1384	0.0007

Based on Table 3 shows that the absolute error value of training data for all four FIS is very small ie everything is smaller than 0.01.

The plot of predicted versus actual values of data testing for FIS52, FIS53, FIS93, and FIS95 is shown in Figure 5 as follows:

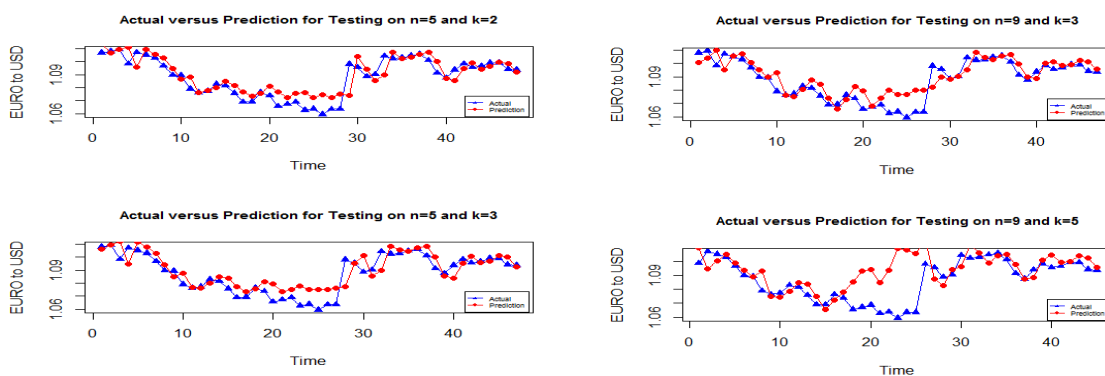


Figure 3. Plots of actual versus predicted values of data testing on FIS52, FIS53, FIS93, and FIS95

Based on the plot results of the actual versus the prediction values in Figure 3, it is clearly exposed that in FIS52 and FIS53 the overall gap between the actual and predicted values are relatively small. However, in FIS93 and FIS95 found some actual values that have a very large gap with the prediction values that are on observations of the 23rd to 33rd. The gap are shown in Table 4 as follows:

Table 4. Comparison of actual (Yact.), predicted (Ypred.), and absolute error (Abs.err.) values of FIS52, FIS53, FIS93, and FIS95 on any parts of Testing Data

FIS52			FIS53			FIS93			FIS95		
Yact	Ypred	Abs.err	Yact.	Ypred	Abs.err	Yact	Ypred	Abs.err	Yact	Ypred	Abs.err
1.0656	1.0769	0.0113	1.0656	1.0796	0.014	1.0656	1.0796	0.014	1.0656	1.0854	0.0198
1.0672	1.0723	0.0051	1.0672	1.0736	0.0064	1.0672	1.0678	0.0006	1.0672	1.0932	0.026
1.0688	1.0757	0.0069	1.0688	1.0754	0.0066	1.0688	1.0739	0.0051	1.0688	1.0941	0.0253
1.0627	1.0763	0.0136	1.0627	1.0779	0.0152	1.0627	1.0801	0.0174	1.0627	1.085	0.0223
1.0637	1.0724	0.0087	1.0637	1.0752	0.0115	1.0637	1.0765	0.0128	1.0637	1.0934	0.0297
1.0594	1.0746	0.0152	1.0594	1.0753	0.0159	1.0594	1.0765	0.0171	1.0594	1.1093	0.0499
1.0636	1.0724	0.0088	1.0636	1.0754	0.0118	1.0636	1.0798	0.0162	1.0636	1.108	0.0444
1.0636	1.0752	0.0116	1.0636	1.0764	0.0128	1.0636	1.0803	0.0167	1.0636	1.1058	0.0422
1.098	1.0743	0.0237	1.098	1.0771	0.0209	1.098	1.082	0.016	1.098	1.1147	0.0167
1.0955	1.1041	0.0086	1.0955	1.0951	0.0004	1.0955	1.0899	0.0056	1.0955	1.0872	0.0083

Based on Table 4, the third column of each FIS is the absolute errors. Consider the absolute errors of FIS53, in general, the absolute errors decreases slightly compared to the absolute errors in FIS52. However, very different conditions occur in FIS93 and FIS95, where the absolute errors in both FIS tend to increase compared to the absolute error in FIS52, even the absolute errors on FIS95 is relatively large.

3.5. System Performance Measurement

In this session, it will be shown system performance based on 2 indicators that are MAPE and R^2 both in training and testing data. The MAPE value is expected to be small (close to 0), whereas the R^2 value is expected to be large (close to 1). Both system performance indicators are calculated on the 12 FIS that have been built. MAPE and R^2 values in training data are presented in Table 5 as follows:

Table 5. MAPE and R^2 values of 12 FIS in Training Data

n	Lag k=2		Lag k=3		Lag k=5	
	MAPE(%)	R^2	MAPE(%)	R^2	MAPE(%)	R^2
3	0.4733	0.9894	0.4506	0.9903	0.4199	0.9916
5	0.4528	0.9903	0.4478	0.9904	0.4365	0.9908
7	0.4471	0.9907	0.4446	0.9909	0.3822	0.9931
9	0.4441	0.9909	0.4263	0.9914	0.3814	0.9928

The first column of Table 5 states the number of fuzzy sets and the first row represents the number of lags assumed to affect the output. Table 5 can be read by combining rows and columns. Suppose that in the first column $n = 5$ and in the first row lag $k = 2$, thus the mean is the performance of FIS52 that has the value of MAPE = 0.4528% and the value of $R^2 = 0.9903$. In the first column of $n = 9$ and in the first line of lag $k = 5$ which means that the performance of FIS95 with the value of MAPE = 0.3814% and the value of $R^2 = 0.9928$.

Based on Table 5, it is known that the smallest MAPE indicator occurs in FIS95, while the largest R^2 indicator occurs in FIS75. In this training data MAPE value consistently decreases with increasing linguistic value and number of lags, although MAPE value range of 12 FIS is only 0.09%. The value of R^2 slightly behaves differently in lag $k = 5$ for $n=3$ to $n=7$ increases, but for $n=7$ to $n=9$ decreases. In other hand for lag $k=2$ and $k=5$, the R^2 value continues to increase with increasing linguistic value. The increase and decrease of the R^2 values of 12 FIS only have a range of 0.0037.

The measurement of system performance for testing data is presented in Table 6. This testing data is used to validate the system's ability to predict future observational values. Based on Table 6, it is known that FIS with lag $k = 2$ is generally better performing based on MAPE or R^2 indicator, the smallest MAPE value is 0.5136% of FIS72, while the largest R^2 value is 0.7858 of FIS32.

Table 6. MAPE and R^2 values of 12 FIS in Testing Data

n	Lag k=2		Lag k=3		Lag k=5	
	MAPE(%)	R^2	MAPE(%)	R^2	MAPE(%)	R^2
3	0.5357	0.7858	0.542	0.7674	0.711	0.663
5	0.543	0.7655	0.6014	0.7304	0.8858	0.4077
7	0.5136	0.7849	0.5608	0.7508	0.8933	0.27
9	0.5174	0.7803	0.5694	0.7492	0.9457	0.203

The results in Table 6 also confirm that the largest MAPE value is 0.9457% of FIS95, which is still a very small MAPE value that has the range of 0.4321% in testing data. Very different conditions occur on the value of R^2 that has the range of 0.5828 which is very large value. FIS32 has the accuracy of 0.7858 = 78.58%, whereas FIS95 only has the accuracy of 0.2030 = 20.30%.

The phenomenon of MAPE and R^2 values above are very interesting to investigate further. If the two performance measurements are combined with the plot of the actual versus the prediction values of the testing data in Figure 3, or the comparison of the absolute errors of 10 observed values in table 4, it will be obtained that the value of R^2 is very sensitive to the existence of a large gap between the actual and prediction value, but not for MAPE indicator.

4. CONCLUSION

In the exchange rate of EURO to USD dataset used in this study, some important informations are obtained ie, The fuzzy rule bases of 12 FIS Takagi-Sugeno zero order optimized by the least square method have been successfully generated. Implementation of 12 FIS in training data produce MAPE performance measurement indicator between 0.3814% and 0.4733%, and R^2 performance measurement indicator between 0.9894 and 0.9931. Of course, it shows that 12 systems have excellent performance in training data. The number of linguistic values and the number of lags as system input are very influential on system performance in the testing data. The performance measurement indicator of R^2 is very sensitive to the existence of a large gap between the actual and the predicted values, so as to produce a biased performance measurement.

REFERENCES

- [1] Doaa MA, Faten HF, Ninet MA, Hassen TD. Modeling and control PV-wind hybrid system based on fuzzy logic control technique. *Indonesian Journal of Electrical Engineering and Computer Science*. 2012; 10(3): 431-441.
- [2] Arbab NK, Dai Y, Amir MS, Xu XY. Identification of Nonlinear System Based on Fuzzy Model with Enhanced Gradient Search. *Indonesian Journal of Electrical Engineering and Computer Science*. 2014; 12(7): 5261-5267.
- [3] I Gusti Made Ngurah AY, I Ketut Gede Darma P, Ni Md Ika MM. Hypertension Expert System with C5. 0 Algorithm and Fuzzy Logic. *Indonesian Journal of Electrical Engineering and Computer Science*. 2014; 12(7): 5569-5677.
- [4] Azriyenni A, Mohd Wazir M, Naila Z. Fuzzy Neural Network for Classification Fault In Protection System. *Indonesian Journal of Electrical Engineering and Computer Science*. 2014; 12(8): 5969-5975.
- [5] Abdulrahman AA, Imad FA. Implementation of Dynamic Fuzzy Logic Control of Traffic Light with Accident Detection and Action System using iTraffic Simulation. *Indonesian Journal of Electrical Engineering and Computer Science*. 2018; 10(1): 100-109.
- [6] Jeng Ren H, Shyi-Ming C, Chia-Hoang L. Handling forecasting problems using fuzzy time series. *Fuzzy sets and systems*. 1998; 100(1-3): 217-228.
- [7] Shyi Ming C, Jeng Ren H. Temperature prediction using fuzzy time series. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 2000; 30(2): 263-275.
- [8] Kunhuang H. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy sets and systems*. 2001; 123(3): 387-394.
- [9] Shyi Ming C. Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems*. 2002; 33(1): 1-16.
- [10] Shyi Ming C, Nien Yi C. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *International Journal of intelligent systems*. 2006; 21(5): 485-501.
- [11] AB Mutalib, Siti M, Nazirah R, Daud M. Forecasting Unemployment based on Fuzzy Time Series with Different Degree of Confidence. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*. 2017; 9(1-4): 21-24.
- [12] Siti Nor FA, Mahmud O. Forecasting Rainfall Distribution Based on Deseasonalising Fuzzy Time Series Model. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*. 2017; 9(2-4): 89-93.
- [13] Yusri S A, Yasunori M. Seasonal short-term electricity demand forecasting under tropical condition using fuzzy approach model. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*. 2017, 9(1-3): 77-82.
- [14] Samingun H, Ari PSP. Applied Fuzzy System with R Software (*Sistem Fuzzy terapan dengan Software R*). First Edition. Malang. UBPress. 2017: 81-135.
- [15] Guofang K, Yuanchen L. Using fuzzy association rules to design e-commerce personalized recommendation system. *Indonesian Journal of Electrical Engineering and Computer Science*. 2014; 12(2): 1519-1527.
- [16] Muhammad F A, Bima A, Samingun H, Harisaweni N. Algorithm Apriori Association Rule in Determination of Fuzzy Rule Based on Comparison of Fuzzy Inference System (FIS) Mamdani Method and Sugeno Method. *Business Management and Strategy*. 2016; 7(1): 103-124.
- [17] Mohammad Z K, Mohammad T. Using adaptive neuro-fuzzy inference system for hydrological time series prediction. *Applied soft computing*. 2008; 8(2): 928-936.

-
- [18] Mahmut F, Mahmud G. Hydrological time-series modelling using an adaptive neuro-fuzzy inference system. *Hydrological processes*. 2008; 22(13): 2122-2132.
- [19] Melek A B, Derya A. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange. *Expert Systems with Applications*. 2010; 37(12): 7908-7912.
- [20] Angga W P, Samingun H. Optimasi Fungsi Keanggotaan Sistem Fuzzy Time Series Dengan Algoritma Genetika (Studi Kasus Harga Harian Saham Bank CIMB Niaga). *Jurnal Mahasiswa Statistik*. 2015; 3(3): 169-172.
- [21] Samingun H, Achmad E, Ferry J, Agus W. Implementation of particle swarm optimization (PSO) algorithm for estimating parameter of arma model via maximum likelihood method. *Far East Journal of Mathematical Sciences (FJMS)*.2017;102(7):1337-1363.
- [22] LX. Wang, and Jerry MM. Generating fuzzy rules by learning from examples. *IEEE Transactions on systems, man, and cybernetics*. 1992; 22(6): 1414-1427.
- [23] Ajoy KP, and Dobrivoje P. *Computational intelligence in time series forecasting: theory and engineering applications*. Springer Science & Business Media, 2006.