Frequency density-based partitioning (FDP) for forecasting IHSG

B Irawanto¹,a, R W Ningrum²,b, R Wulandari³, B Surarso¹ and Farikhin¹

¹Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
²Master Program of Mathematics, Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
³Undergraduate Program of Mathematics, Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

*Corresponding author: a) b_irawanto@yahoo.co.id, b) rianiwidyaningrum@student.undip.ac.id

Abstract. Forecasting method based on Fuzzy Time Series (FTS) has been widely developed in recent years. In this paper, we propose a new improvement at determining IHSG data and partitioning stage. At the present study, we define the IHSG data then calculate the basis value to find out how much interval should be used. Secondly, we divide the main interval into several numbers of sub-intervals. The empirical analysis shows that sub-interval caused the fuzzy number getting closer to crisp value. Then, it causes the better forecasting value. We use the data of Jakarta Composite Index (IHSG) for simulation. The Means Squared Error (MSE) value reduces significantly with this modification. The value of Average Forecasting Error (AFER) 0.00386% shows the method outperforms the previous method.

1. Introduction

A method of FTS is growing very rapidly in recent years in order to get better forecasting error rates. The purpose of forecasting activity is to find pattern in a raw historical data and use it for future forecasting. The data should be collected periodically on the order of time in hours, days, weeks, months, quarters and years. There is many forecasting methods-based FTS has been developed to forecast stock market index [1-6]. The IHSG is a reflection of general capital market activities.

In this study, we present FTS method based on metric and linguistic variables. The metric approach based on frequency density-based partitioning [7] with modification at partitioning process-based frequency [2] and calculating length of intervals-based average [8]. We evaluate it by applying the methods on the IHSG data year 2015-2017 and compare with previous metric method FTS without modification. The result of each forecast method will be evaluated by comparing the value of MSE and AFER.

This paper is arranged as follows, Section 1 Introduction, Section 2 Preliminaries, Section 3 Experimental Results, Section 4 Discussion, and Section 5 Conclusion.

2. Fuzzy Time Series (FTS)

The following is some concepts on FTS.
Definition 1 [9]
A fuzzy set $A$ of the IHSG data $P, P = \{p_1, p_2, ..., p_n\}$ defined as $A = \{f_A(p_i)\}, i = 1, 2, ..., n$ where $f_A$ is the member’s func of the fuzzy set $A$, $f_A: P \rightarrow [0,1], f_A(p_i)$ denotes the grade of membership of $p_i$ in the fuzzy set $A$, and $1 \leq i \leq n$.

Definition 2 [10]
Let $P(t)$ be IHSG data and $P(t) \subseteq R$. IHSG data is assumed to be defined by $f_i(t), i = 1, 2, \ldots P(t)$ and $F(t)$ is collection of $f(t_i)$. Then $F(t)$ is called a FTS of $t, i = 1, 2, \ldots$.
We define $F(t) = F(t - 1) \circ R(t, t - 1)$ where $R(t, t - 1)$ is a fuzzy relation and " $\circ$ " is a max-min composition operator, then $F(t)$ is caused by $F(t - 1)$ where $F(t)$ and $F(t - 1)$ are fuzzy set.

Definition 3 [11]
The member’s func. of the triangle is presented by 3 parameters fuzzy $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, with the $\alpha_1, \alpha_2, \alpha_3 \in R$. This curve is essentially a combination of functions as this following equation:

$$\mu_{\alpha}(x) = \begin{cases} \frac{x - \alpha_1}{\alpha_3 - \alpha_1}, & \alpha_1 \leq x \leq \alpha_2 \\ \frac{\alpha_3 - x}{\alpha_3 - \alpha_2}, & \alpha_2 \leq x \leq \alpha_3 \\
0, & \text{else}
\end{cases} \quad (1)$$

3. Results and Discussion
Here are the steps of Metric approaching method. We modify Step 3 Dividing intervals into some sub-intervals using frequency density [7].
Step 1 Determine the universe of discourse $U$
Determine the IHSG data [7]. From the historical monthly data of IHSG we know that $X_{min}, X_{max}$ is minimum and maximum discrete value of the data, $D_1, D_2$ is real number.
$$X_{min} = 4223,908 \quad \text{and} \quad X_{max} = 6355,654$$
Then, $U = [X_{min} - D_1, X_{max} + D_2] = [4000; 7000]$
Step 2 Find the number of intervals
Find the number of intervals with absolute average-based [8] using this following process.
Calculate absolute mean value of the data

$$|X| = \frac{\sum_{i=1}^{n} |\Delta_i|}{n} \quad (2)$$

where,

$|X| :$ Absoluted mean value of data

$|\Delta_i| :$ Absoluted difference value of $i$ with $i = 1, 2, \ldots, n$

Calculate range value:

$$Range(r) = X_{max} - X_{min} = 6355,654 - 4223,908 = 2131,746$$

Calculate the basis with following equation:

$$Basis = \frac{|X|}{2} = 67,53217 \approx 67$$

Calculate the number of intervals:

$$I = \frac{J}{basis} = \frac{2131,746}{67} = 30,4535 \approx 31$$

At single median data, the number of intervals is odd numbers, the value 30.4535 rounded up to the nearest odd number 31 intervals.

Step 3 Divide into sub-interval
Divide into sub-interval with mean-based partitioning [9]. Division returned an average based interval according to their respective frequency, if there is no data that is distributed in the interval (frequency is zero) then the interval will be discarded.
Step 4 define a fuzzy triangle based on table

In this paper, there are 29 linguistic value, as shown in Table 1.

For example, from $u_1$ triangular member’s func can be made from the interval $[4194; 4291]$ as shown Figure 1.

| Sub - Interval | $u_2 = [4194; 4291]$ | $u_3 = [4436.5; 4485]$ | $u_4 = [4485; 4630.5]$ | $u_5 = [4679; 4776]$ | $u_6 = [4776; 4800.25]$ | $u_7 = [4800.25; 4824.5]$ |
|----------------|----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| $u_2 = [4194; 4291]$ | $u_3 = [4436.5; 4485]$ | $u_4 = [4485; 4630.5]$ | $u_5 = [4679; 4776]$ | $u_6 = [4776; 4800.25]$ | $u_7 = [4800.25; 4824.5]$ | $u_1 = [4194; 4291]$ |
| $u_3 = [4436.5; 4485]$ | $u_4 = [4485; 4630.5]$ | $u_5 = [4679; 4776]$ | $u_6 = [4776; 4800.25]$ | $u_7 = [4800.25; 4824.5]$ | $u_2 = [4194; 4291]$ |
| $u_4 = [4485; 4630.5]$ | $u_5 = [4679; 4776]$ | $u_6 = [4776; 4800.25]$ | $u_7 = [4800.25; 4824.5]$ | $u_3 = [4436.5; 4485]$ |
| $u_5 = [4679; 4776]$ | $u_6 = [4776; 4800.25]$ | $u_7 = [4800.25; 4824.5]$ | $u_4 = [4485; 4630.5]$ |
| $u_6 = [4776; 4800.25]$ | $u_7 = [4800.25; 4824.5]$ | $u_5 = [4679; 4776]$ |

Figure 1. Member’s func of triangular fuzzy numbers $A_1$

$$\mu_{A_1}(x) = \begin{cases} 
\frac{x - 4194}{4242.5 - 4194}, & \text{if } 4194 \leq x \leq 4242.5 \\
\frac{4291 - x}{4291 - 4242.5}, & \text{if } 4242.5 \leq x \leq 4291
\end{cases} \quad (3)$$

And further by replacing the $u_i$ as linguistic $A_i$. Based on the fuzzification obtained in step 4, we can get the fuzzy logical relationship group (FLRG).

Step 5 Built FLR and FLRG

FLRG is a relation that will be used to change the value of the fuzzy rule becomes the value of the crisp value. FLR "$A_p, A_q, A_r \rightarrow A_s$" denotes that "if the fuzzified of month $p,q$ and $r$ are $A_p, A_q$ and $A_r$ respectively, then the fuzzified data of month (r) is $A_r$".

Step 6 Defuzzification

For the determination value of forecasting by trend prediction through defuzzification of linguistic forms through the centroid method is represented by the following forms [7]:

$$t_j = \begin{cases} 
\frac{1 + 0.5}{a_1 + 0.5}, & \text{if } j = 1 \\
\frac{0.5 + 1 + 0.5}{a_j + 0.5}, & \text{if } 2 \leq j \leq n - 2 \\
\frac{0.5 + 1}{a_{n-1} + 1}, & \text{if } j = n
\end{cases}$$

(4)
Where \((a_{j-1}, a_j, a_{j+1})\) is the middle value of interval fuzzy \(A_{j-1}, A_j, A_{j+1}\) and respectively.

**Step 7 Calculate Forecasting Error**

In this step, we use Mean Squared Error (MSE) and Average Forecasting Error Rates (AFER) to know the magnitude of the irregularities which occurred in forecasting results. Here is the equation to calculate AFER:

\[
AFER = \frac{|x_i - \hat{r}_i|}{\bar{x}_i} \times 100\% ,
\]  

(5)

The mean Squared Error (MSE) is another method to evaluate forecasting methods

\[
MSE = \frac{\sum_{i=1}^{n} (X_i - F_i)^2}{n}.
\]  

(6)

| AFER Value | Criterion of Forecasting Results |
|------------|----------------------------------|
| <10%       | Very good                        |
| 10%-20%    | Good                             |
| 20%-50%    | Good enough                      |
| >50%       | Poor                             |

**4. Simulation**

In this section, we demonstrated the proposed method to forecast monthly IHSG from year 2015-2017 which can be access in www.idx.com. We compare the result of forecasting between Metric approaching method and Modification of metric approaching method (the modification is using FDP).

| Year | Month | Actual Data | Metric Approaching Method | Propose Modification |
|------|-------|-------------|----------------------------|----------------------|
| 2015 | 1     | 5289.404    | 5295.405965                | 5275.023632          |
|      | 2     | 5450.294    | 5445.374688                | 5454.845495          |
|      | 3     | 5518.675    | 5552.537533                | 5512.639634          |
|      | 4     | 5086.425    | 5059.367487                | 5084.81757           |
|      | 5     | 5216.379    | 5059.367487                | 5222.11098           |
|      | 6     | 4910.658    | 5059.367487                | 4923.691154          |
|      | 7     | 4802.529    | 4816.502654                | 4812.3139            |
|      | 8     | 4509.607    | 4458.726722                | 4532.916208          |
|      | 9     | 4223.908    | 4305.404908                | 4312.83771           |
|      | 10    | 4455.18     | 4458.726722                | 4421.622488          |
|      | 11    | 4446.458    | 4458.726722                | 4421.622488          |
|      | 12    | 4593.008    | 4640.659895                | 4617.332292          |
| 2016 | 1     | 4615.163    | 4640.659895                | 4617.332292          |
At Figure 2 in general, forecasting results from both methods approach the actual value. The metric approached method shown by the green line shows inappropriate results at several points between March-July 2015, July-November 2016 and November 2017. While the proposed modification shows forecasting results that are almost same as the actual data at almost all points.

The proposed method is Metric approach method based on frequency density-based partitioning [7] with modification at partitioning process-based frequency [2] and calculating length of intervals-based average [8]. At modification process, the universe of discourse divided into 29 clusters then we build the fuzzification. At this case, the proposed method gives better forecasting results indicated by the MSE and AFER value is smaller than the classical method. Metric Approached method gives MSE value 4495.346 and AFER value 0.00935%, while proposed modifications gives MSE value 1018.698 and AFER value 0.00386%. Both MSE and AFER value at proposed modification gives smaller value than
the previous method. So, we claimed that the modification has succeeded in producing a better forecasting value.

![Graphic of forecasting results](image)

**Figure 2.** Graphic of forecasting results

5. Conclusion
We have presented a modification of metric approach forecasting method. We modify Step 3 by analogizing the interval as a range of data groups then calculate the frequency. Then, the top 3 frequencies are partitioned again into sub-intervals. The frequency value of new intervals is calculated again. An interval with zero frequency is omitted. From Table 3 and Figure 2, the AFER and MSE values of the modification are smaller than Metric approaching method.

References
[1] Chen M Y and Chen B T 2015 *Inf. Sci.* 294 227
[2] Cheng C H and Yang J H 2018 *Neurocomputing* 302 33
[3] Efendi R, Arbaiy N and Deris M M 2018 *Inf. Sci.* 441 113
[4] Rubio A, Bermudez J D and Vercher E 2017 *Expert Syst. Appl.* 76 12
[5] Su C H and Cheng C H 2016 *Neurocomputing* 205 264
[6] Degiannakis S, Filis G and Hassani H *J. of Empir. Finance* 46 111
[7] Jilani T A, Burney S M A and Ardl C 2010 *Int. J. Comput. Inf. Eng.* 4 1194
[8] Saxena P, Sharma K and Easo S 2012 *Int.J.Comput. Technol.Appl.* 3 957
[9] Cheng S H, Chen S M and Jian W S 2016 *Inf. Sci.* 327 272
[10] Abdullah L 2013 *Int. J. of Eng. and Technol. (IJET)* 5 452
[11] Zeng X Y, Shu L, Huang G and Jiang J 2016 *Appl. Math. Model.* 40 1717