Development of Accuracy for the Weighted Fuzzy Time Series Forecasting Model Using Lagrange Quadratic Programming

Agus Fachrur Rozy¹, Solimun², Ni Wayan Surya Wardhani³
¹,²,³Departement Statistic, University of Brawijaya, Indonesia
agus.rozy@yahoo.com¹, solimun@ub.ac.id², wswardhani@ub.ac.id³

ABSTRACT

Limitation within the WFTS model, which relies on midpoints within intervals and linguistic variable relationships for assigning weights. This reliance can result in reduced accuracy, especially when dealing with extreme values during trend to seasonality transformations. This study employs the Weighted Fuzzy Time Series (WFTS) method to adjust predictive values based on actual data. Using Lagrange Quadratic Programming (LQP), estimated weights enhance the WFTS model. MAPE assesses accuracy as the model analyzes monthly IHSG closing prices from January 2017 to January 2023. The MAPE value of 0.61% results from optimizing WFTS with LQP. It utilizes a deterministic approach based on set membership counts in class intervals, continuously adjusting weights during fuzzification, minimizing the deviation between forecasted and actual data values. The Weighted Fuzzy Time Series Forecasting Model with Lagrange Quadratic Programming is effective in forecasting, indicated by a low MAPE value. This method evaluates each data point and adjusts weights, offering reliable investment insights for IHSG strategies.

Keywords: Fuzzy Time Series; Weighted Fuzzy Time Series; Lagrange Quadratic Programming; Mean Absolute Percentage Error.

https://doi.org/10.31764/jtam.v7i4.16783

This is an open access article under the CC-BY-SA license

A. INTRODUCTION

Forecasting is a highly relevant technique for projecting data into future periods using historical data as the basis for building mathematical models (Hilhami et al., 2020). The primary goal of forecasting is to provide a solid foundation for planning and addressing potential issues that may arise in the future (Subramanian, 2021). In various sectors, especially in the field of economics, forecasting methods are commonly used as a framework for planning, particularly in the context of trading in the capital markets (Bhowmik & Wang, 2020).

Furthermore, forecasting models also face significant challenges in accommodating and managing highly volatile market conditions (Tschora et al., 2022). Specifically, real-time markets are vulnerable to unexpected price risks, leading to higher price volatility (Assouto et al., 2020). While forecasting errors for forward markets are often reported to be quite accurate with single-digit error rates, comparable studies indicate that real-time markets often report significantly higher error rates (Ferreira et al., 2019).

Forecasting time series data is essential for understanding the uncertainties that may arise in the future. The process of time series data forecasting involves observing data in each
observation period. A commonly utilized conventional method for forecasting is Autoregressive Integrated Moving Average (ARIMA) (Wang et al., 2023). ARIMA is a straightforward statistical method since it only requires the data of the variable to be forecasted. The approach involves iterating on the existing model.

The ARIMA model has a limitation, requiring the assumption of stationarity, which means it cannot be directly applied to time series data with a trend pattern (Moghimi et al., 2023). The presence of a trend pattern causes uncertainty in the average value of the time series data. This issue can be addressed using the Fuzzy Time Series model, which is an extension of fuzzy theory to time series data, representing each data point within an interval (Aliyev et al., 2019). The relationship between fuzzy logic and time series results in accurate forecasting and excellent cluster formation compared to the widely used ARIMA-Garch model, as measured by the Mean Absolute Percentage Error (MAPE) (Haryono et al., 2013).

The Fuzzy Time Series (FTS) method has a drawback as it defuzzifies prediction data without considering fuzzy logic relationship repetitions, thus disregarding past trends (Yolcu & Lam, 2017). To address this, the weighted Fuzzy Time Series model was developed, repeating fuzzy logic relationships while considering past trends during the defuzzification process (Singh, 2021). A study comparing Weighted Fuzzy Time Series and Fuzzy Time Series by Widiyani et al. (2022) found that the predictions from Weighted Fuzzy Time Series (WFTS) are superior in accuracy based on MAPE values. In contrast to forecasting with fuzzy time series, weighted fuzzy time series forecasting incorporates an additional step post-defuzzification: the introduction of weighting factors. These weights are applied to the fuzzy sets generated during the fuzzification process, influencing the final forecast. This approach enhances the accuracy of predictions by assigning (A’yun et al., 2015) different degrees of importance to various linguistic variables and intervals, reflecting their significance in the forecasting model. It enables a more nuanced and refined forecast, allowing for improved insights into data trends and potential future outcomes. By incorporating weighted elements into the fuzzy time series methodology, this approach advances the precision and reliability of forecasting models.

This study introduces several novel elements compared to previous research. Firstly, it highlights that the Weighted Fuzzy Time Series (WFTS) model outperforms the conventional Fuzzy Time Series (FTS) model in terms of accuracy. This finding is significant as it underscores the superiority of the WFTS approach, providing a fresh perspective in the field. Secondly, the study identifies a limitation within the WFTS model, which relies on midpoints within intervals and linguistic variable relationships for assigning weights. This reliance can lead to reduced accuracy, especially when dealing with extreme values during trend-to-seasonality transformations, as observed by Lucas et al. (2022). To address this limitation and further improve forecasting precision, this study draws inspiration from Rezvani et al. (2021). Rezvani and their team employed the Piecewise Aggregate Approximation (PAA) technique in conjunction with Lagrange Multipliers to detect points in time series data that result from pattern changes. This innovative approach significantly enhanced the model’s accuracy, demonstrating the potential benefits of optimization techniques in time series forecasting.

The main novelty of this study resides in its utilization of the Lagrange Quadratic Programming optimization technique to improve the WFTS model through the estimation of weights as an integral part of the defuzzification process. By mathematically modeling the
estimation of Fuzzy Time Series weighting, the research aims to pinpoint the optimal solution for enhancing forecasting accuracy. This distinctive approach sets the study apart from prior research and presents a promising avenue for enhancing the precision of time series forecasting models.

B. METHODS

In this study, weighting processes are employed in the fuzzy time series method. The purpose of this weighting is to adjust the predictive values generated based on actual data. This research focuses on a case study involving the closing prices of IHSG, where a forecasting model is developed using the Weighted Fuzzy Time Series (WFTS) method. The WFTS model will be enhanced by estimated weights \( w \) calculated using Lagrange Quadratic Programming (LQP). The resulting model will be assessed using MAPE to determine its accuracy. WFTS will be applied to analyze the monthly time series data of IHSG closing prices from January 2017 to January 2023 using a fuzzy logic approach. The IHSG is a stock price index derived from the analysis of trend movements, facilitating the assessment of price fluctuations over various time periods (Daniswara & Daryanto, 2019). The stock market index in Indonesia, referred to as the Indeks Harga Saham Gabungan (IHSG), goes by various names including the Indonesian Composite Index (ICI), the Indonesian Exchange (IDX) Composite, and the Jakarta Stock Exchange (JKSE). LQP will determine optimal weights for each value in the time series, thereby improving the model's predictive accuracy. The analysis steps are illustrated in Figure 1.

![Figure 1](image_url)
Based on Figure 1, it can be observed that the initial process begins with data collection and descriptive analysis. Subsequently, the forecasting process is carried out using the Weighted Fuzzy Time Series algorithm. The weighted fuzzy time series method employs weights as predictors for forecasting computations. These weights are generated from the repetition of Fuzzy Logic Relationships (FLR) to construct a weight matrix (Efendi et al., 2013). In detail, the research steps are shown as:

2. Analyzing descriptive statistic to know characteristic of data.
3. Define $U = \text{Universe of discourse}$:
   \[
   U = [D_{\text{min}} - B_1, D_{\text{max}} + B_2]
   \] (1)
4. Forming partitions into several equal-length intervals using the Sturges formula (Devianto et al., 2022):
   \[
   1 + 3.322 \log n
   \] (2)
5. Defining fuzzy sets $A_1, A_2, \ldots, A_i$ in the universe of discourse $U$ is:
   \[
   A_1 = \frac{a_{11}}{u_1} + \frac{a_{12}}{u_2} + \cdots + \frac{a_{1m}}{u_m} \\
   A_2 = \frac{a_{21}}{u_1} + \frac{a_{22}}{u_2} + \cdots + \frac{a_{2m}}{u_m} \\
   \cdots = \cdots + \cdots + \cdots + \cdots \\
   A_k = \frac{a_{k1}}{u_1} + \frac{a_{k2}}{u_2} + \cdots + \frac{a_{km}}{u_m}
   \] (3)
6. Determining the fuzzy relationship of historical data through Fuzzy Logical Relationship (FLR), where two consecutive fuzzy sets $A_i(t - p)$ and $A_j(t)$ are defined to form the first FLR as $A_i \rightarrow A_j$. $A_i$ can be referred to as Left Hand Sides (LHS), and $A_j$ as Right Hand Side (RHS), with $p = 1, 2, \ldots, p^*$. 
7. Estimating weight based method of Lagrange multipliers which is a widely recognized technique utilized to solve constrained optimization problems (Vadlamani et al., 2020). It involves finding the optimal point (denoted as $x^* \equiv (x, y)$) in multidimensional space that locally optimizes the merit function $f(x)$ while satisfying the constraint $g(x) = 0$:
   \[
   F(y) = f(y) + \sum_{j=1}^{m} \lambda_j g_j(y)
   \] (4)
8. The linguistic variable transformations formed from Step 5 are used as the defuzzification process using the equation:
   \[
   F(t+1) = U(t) \times w(t)^T = \begin{bmatrix} u_{j1}, u_{j2}, \ldots, u_{jk} \end{bmatrix} \times \begin{bmatrix} w'_1, w'_2, \ldots, w'_k \end{bmatrix}^T
   \] (5)
9. Calculating the forecasting value is done by adding a differencing process between the actual data and the midpoint values formed in each interval class (Surono et al., 2022):
\[ \hat{F}(t + 1) = F(t + 1) \pm |diff(Y(t), m_i) | \quad (6) \]

10. Evaluation performance using MAPE: The Mean Absolute Percentage Error (MAPE) is computed by taking the absolute error between the forecasted data and the actual values for each period and then dividing it by the corresponding observed actual value (Kim & Kim, 2016). The calculation of MAPE is shown in the following formula (Khair et al., 2017):

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|\hat{Y}_t - Y_t|}{Y_t} \quad (7)
\]

C. RESULT AND DISCUSSION
1. Descriptive Statistic

Descriptive analysis is conducted to understand the data patterns occurring in the IHSG return rates. Data patterns are examined to understand the movement of closing prices of the IHSG from January 2017 to January 2023. In this research, descriptive statistics are observed from the visualization of actual data as shown in Figure 2.

From Figure 2, it can be observed that there are extreme values. This can be seen in the period of March 2020 when there was a decrease in prices (a downward trend). Meanwhile, the highest IHSG closing price is observed in April 2022, suggesting that the IHSG closing price has experienced an increase in the recent periods, specifically from January 2022 to April 2022.

2. Fuzzyfication Process

The process of forecasting using weighted fuzzy time series is based on the weights generated for each interval class, multiplied by the mid-point value for each interval class. The steps involved in this process are as follows:

a. Forming the universe of discourse using equation (1), which is:

\[
U = [D_{min} - B_1, D_{max} + B_2] = [4538.93 - 0, 7228.91 + 0] = [4538.93, 7228.91]
\]
The calculation process above results in the universe \((U)\) with lower and upper bounds of 4538.93 and 7228.91.

b. Calculate interval class using equation (2), which is:

\[
1 + 3,322 \log 73 = 7.18 \approx 7
\]

The calculation results in 7 interval classes used to group actual data, simplifying the transformation process into linguistic variables in the fuzzification process.

c. Fuzzification of actual data involves categorizing data based on upper and lower bounds, as indicated in the following Table 1.

<table>
<thead>
<tr>
<th>Table 1. Fuzzyfication Actual Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

From Table 1, it can be observed that each interval class has a distinct number of set members. The grouping of actual data into interval classes is based on their respective upper and lower bounds.

d. The determination of the fuzzy logic relationship group (FLRG) is based on the number of relationships between linguistic variables across periods in the following Table 2.

<table>
<thead>
<tr>
<th>Table 2. FLRG Actual Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic Variable</td>
</tr>
<tr>
<td>(A_1)</td>
</tr>
<tr>
<td>(A_2)</td>
</tr>
<tr>
<td>(A_3)</td>
</tr>
<tr>
<td>(A_4)</td>
</tr>
<tr>
<td>(A_5)</td>
</tr>
<tr>
<td>(A_6)</td>
</tr>
<tr>
<td>(A_7)</td>
</tr>
</tbody>
</table>

From Table 2, it is evident that the highest total relations exist for variable \(A_4\), which amounts to 24, indicating that the actual data, on average, falls within interval class 4. On the other hand, the lowest number of relations is indicated by variable \(A_2\), which is 4, signifying that there is a limited amount of actual data that falls outside the average and into interval class 2. These relations demonstrate the relationship between linguistic variables across periods as a result of fuzzification.
3. Estimation Weighted using Lagrange Quadratic Programming

Forecasting the IHSG (Indonesia Stock Exchange Composite Index) using WFTS involves weight estimation based on quadratic equation form. The LQP equation is formed based on the objective function and constraints defined in equation (6) as follows:

\[ L(w_{j,i}, \lambda_i) = \sum_{j=1}^{n} w_{j,i}^2 u_{j,i} + 2\lambda_i (\sum_{j=1}^{n} w_{j,i} - 1) \text{ with } i = 1, 2, 3, ..., n \]  

(8)

Where \( w_{j} \) is the weight for the \( j \) member in interval class \( i \), \( u_{j,i} \) is the membership value of the \( j \) member in the universe of discourse for interval class \( i \), and \( \lambda_i \) is the Lagrange multiplier for class \( i \). As for the equations formed from each interval class based on the previously calculated results, there are 7 intervals, and the equations are obtained as follows:

Class interval 1:
\[
L(w_{j,1}, \lambda_1) = \sum_{j=1}^{5} w_{j,1}^2 u_{j,1} + 2\lambda_1 (\sum_{j=1}^{5} w_{j,1} - 1) \tag{9}
\]

Class interval 2:
\[
L(w_{j,2}, \lambda_2) = \sum_{j=1}^{4} w_{j,2}^2 u_{j,2} + 2\lambda_2 (\sum_{j=1}^{4} w_{j,2} - 1) \tag{10}
\]

Class interval 3:
\[
L(w_{j,3}, \lambda_3) = \sum_{j=1}^{5} w_{j,3}^2 u_{j,3} + 2\lambda_3 (\sum_{j=1}^{5} w_{j,3} - 1) \tag{11}
\]

Class interval 4:
\[
L(w_{j,4}, \lambda_4) = \sum_{j=1}^{24} w_{j,4}^2 u_{j,4} + 2\lambda_4 (\sum_{j=1}^{24} w_{j,4} - 1) \tag{12}
\]

Class interval 5:
\[
L(w_{j,5}, \lambda_5) = \sum_{j=1}^{15} w_{j,5}^2 u_{j,5} + 2\lambda_5 (\sum_{j=1}^{15} w_{j,5} - 1) \tag{13}
\]

Class interval 6:
\[
L(w_{j,6}, \lambda_6) = \sum_{j=1}^{9} w_{j,6}^2 u_{j,6} + 2\lambda_6 (\sum_{j=1}^{9} w_{j,6} - 1) \tag{14}
\]

Class interval 7:
\[
L(w_{j,7}, \lambda_7) = \sum_{j=1}^{11} w_{j,7}^2 u_{j,7} + 2\lambda_7 (\sum_{j=1}^{11} w_{j,7} - 1) \tag{15}
\]

The solution to equations (9) to (15) is carried out through the process of partial derivatives with equation:

Class interval 1
\[
\frac{\partial L(w_{j,1}, \lambda_1)}{\partial w_j} = \sum_{j=1}^{5} 2w_{j,1}u_{j,1} + 2\lambda_1 = 0
\]
\[
\frac{\partial L(w_{j,1}, \lambda_1)}{\partial \lambda_1} = \sum_{j=1}^{5} 2w_{j,1} - 2 = 0 \tag{16}
\]
Class interval 2
\[
\frac{\partial L(w_{j,2}, \lambda_2)}{\partial w_j} = \sum_{j=1}^{4} 2w_{j,2}u_{j,2} + 2\lambda_2 = 0
\]
\[
\frac{\partial L(w_{j,2}, \lambda_2)}{\partial \lambda_2} = \sum_{j=1}^{4} 2w_{j,2} - 2 = 0
\]

Class interval 3:
\[
\frac{\partial L(w_{j,3}, \lambda_3)}{\partial w_j} = \sum_{j=1}^{5} 2w_{j,3}u_{j,3} + 2\lambda_3 = 0
\]
\[
\frac{\partial L(w_{j,3}, \lambda_3)}{\partial \lambda_3} = \sum_{j=1}^{5} 2w_{j,3} - 2 = 0
\]

Class interval 4:
\[
\frac{\partial L(w_{j,4}, \lambda_4)}{\partial w_j} = \sum_{j=1}^{24} 2w_{j,4}u_{j,4} + 2\lambda_4 = 0
\]
\[
\frac{\partial L(w_{j,4}, \lambda_4)}{\partial \lambda_4} = \sum_{j=1}^{24} 2w_{j,4} - 2 = 0
\]

Class interval 5:
\[
\frac{\partial L(w_{j,5}, \lambda_5)}{\partial w_j} = \sum_{j=1}^{15} 2w_{j,5}u_{j,5} + 2\lambda_5 = 0
\]
\[
\frac{\partial L(w_{j,5}, \lambda_5)}{\partial \lambda_5} = \sum_{j=1}^{15} 2w_{j,5} - 2 = 0
\]

Class interval 6:
\[
\frac{\partial L(w_{j,6}, \lambda_6)}{\partial w_j} = \sum_{j=1}^{9} 2w_{j,6}u_{j,6} + 2\lambda_6 = 0
\]
\[
\frac{\partial L(w_{j,6}, \lambda_6)}{\partial \lambda_6} = \sum_{j=1}^{9} 2w_{j,6} - 2 = 0
\]

Class interval 7:
\[
\frac{\partial L(w_{j,7}, \lambda_7)}{\partial w_j} = \sum_{j=1}^{11} 2w_{j,7}u_{j,7} + 2\lambda_7 = 0
\]
\[
\frac{\partial L(w_{j,7}, \lambda_7)}{\partial \lambda_7} = \sum_{j=1}^{11} 2w_{j,7} - 2 = 0
\]

Then, the defuzzification process is performed by multiplying weights and membership values for each interval class using the equation:
\[
F(t) = u_i \times w_i(t)^T
\]
4. Defuzzification

The defuzzification process is carried out by multiplying the LQP weights with each actual data point, which represents membership in each interval class. The results of the defuzzification process are presented in Table 3.

<p>| Table 3. Result of Defuzzification using LQP Method |
|-----------------------------|----------------|</p>
<table>
<thead>
<tr>
<th>$F(t)$</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F(1)$</td>
<td>4753.29</td>
</tr>
<tr>
<td>$F(2)$</td>
<td>5201.75</td>
</tr>
<tr>
<td>$F(3)$</td>
<td>5538.93</td>
</tr>
<tr>
<td>$F(4)$</td>
<td>5936.54</td>
</tr>
<tr>
<td>$F(5)$</td>
<td>6285.25</td>
</tr>
<tr>
<td>$F(6)$</td>
<td>6596.58</td>
</tr>
<tr>
<td>$F(7)$</td>
<td>7038.92</td>
</tr>
</tbody>
</table>

From Table 3, it can be seen that $F(t)$ represents the value of defuzzification results for each interval class, and it is known that the higher the criteria in the linguistic variable, namely "Very Low," "Low," "Quite Low," "Medium," "Quite High," "High," and "Very High." The defuzzified values, denoted as $F(t)$, represent the forecasting results for each interval class. The lower defuzzification value is 4753.29 and higher is 7038.92. These defuzzified values have been obtained using the LQP method for forecasting, enabling the transformation of fuzzy outputs into crisp numerical results, making them more interpretable and practical for decision-making processes.

5. The Result of Forecasting

The forecasting values generated are formed through the defuzzification process based on the WFTS LQP model using equation (6). The forecasting results are visualized as:

![Figure 3. Graph of IHSG Closing Prices Actual Data and WFTS LQP Weight Estimation](image)

From Figure 3, it can be observed that there was an increase in prices during the early period, specifically from January 2017 to 2018. The rise in IHSG prices indicates that investor stock purchases were relatively high. During this early period, investors were advised against making investments due to the increasing prices. Investment decisions should ideally be made around March 2020, as indicated by the downward trend in the forecasting graph using the
Lagrange Quadratic Programming weighting. Additionally, towards the end of the period, there is a change in the downward trend, suggesting that investors should consider selling IHSG stocks in April 2022.

6. Model Evaluation using MAPE

The best forecasting model is needed to produce accurate prediction values. This is aligned with the conditions observed in the historical data that has been collected. The best model for forecasting is determined by examining the resulting Mean Absolute Percentage Error (MAPE) to assess the deviations occurring in the forecasting model. In the process, the calculation of the MAPE value using equation (7) is required, and a value of 0.61% is obtained. The resulting MAPE value is relatively small, primarily due to the fact that the weighting of WFTS with the LQP model provides an optimal solution through partial derivative processes via a deterministic approach. The weights generated are adjusted at each actual data point value from the fuzzification process, resulting in a low deviation of forecasting values from actual data values.

D. CONCLUSION AND SUGGESTIONS

Based on the analysis results, Weighted Fuzzy Time Series Forecasting Model using the Lagrange Quadratic Programming weighting method is effective in making forecasts. This can be seen in the level of accuracy based on the MAPE value. Effectiveness is attributed to the evaluation of each actual data point and the adjustment of weights to generate forecast values with relatively low deviations. This method can serve as the basis for decision-making in investment strategies related to the IHSG.

The suggestion provided in this research is that the development of the Weighted Fuzzy Time Series model with weight estimation using the Lagrange Quadratic Programming equation encounters several challenges. The challenges are related to the complexity of the solution process using partial derivatives, especially for long-term data periods, such as daily data. As a result, many weights need to be estimated for each interval class. Additionally, this research focuses on a single variable or univariate, which may limit the flexibility of the model. Therefore, further development is required by modifying the Lagrange equation and adding additional variables. The inclusion of more variables in the model will consider more factors influencing the time data, thus providing more accurate predictions.

ACKNOWLEDGEMENT

The author expresses gratitude to the University of Mataram for providing an opportunity for the publication of scientific writings, and also extends thanks to the lecturers for their valuable advice and constructive feedback during the course at the University of Brawijaya, Malang.

REFERENCES


Hilhami, M. S., Oktavianto, H., & Fitriyah, N. Q. (2020). Forecasting Harga Saham PT. Astra Agro Lestari dengan Metode Simple Moving Average dan Weighted Moving Average. *Jurnal Teknik Informatika Fakultas Teknik UNEJ.*


