Forecasting Gold Prices Using Membership Function Model as Probability Density Function of Normal Distribution

S. SAKHA, S. BOONTHIEM, W. KLONGDEE*
Khon Kaen University
THAILAND
*Corresponding author: kwatch@kku.ac.th

Abstract: - In this paper, the new forecasting algorithm for predicting gold prices using concept of fuzzy logic with membership function model as probability density function of normal distribution is proposed and compared to the triangular membership function model and the trapezoidal membership function model. The daily gold prices from January 2015 to May 2018 collected from the London Bullion Market Association are used in this study. The prices are transformed to the rate of return and then the Train : Test percent ratio is used to obtain the training data and the testing data, respectively, and consists of 99:01, 95:05, 90:10, 80:20, 70:30, 60:40, 50:50, 40:50, 30:70, 20:80 and 10:90. The root mean squared error and the mean absolute error are applied to measure the performance of the proposed algorithm and to compare three presented models. The experimental results can conclude that the proposed algorithm can be practical to the gold price forecasting; furthermore, the proposed algorithm with the membership function model as probability density function of normal distribution can better improve itself than other models when it has more training data as well as it can demonstrate better performance than others.

Key-Words: - Fuzzy logic, gold price forecasting, probability density function of normal distribution, trapezoidal membership function, triangular membership function.

1 Introduction

Forecasting is a function in management to help decision making. It is also called as the process of estimation in unknown future situations. Normally, it is known as prediction which refers to the approximation of time series or long-term type data. Price forecasting is an essential part of economic decision making. Forecasts may be used in several ways; particularly, investors may use these forecasts to try to earn income from financial activities, to make business decisions, or to consider and determine appropriate government policies.

Gold is the most popular of the valuable metals and dissimilar from other assets. It plays several characters in the world economy and their relatives with financial and macroeconomic factors are well-accepted [1]. It has a monetary value and it is required by central banks to be part of their international reserves, which fulfill many purposes [2]. It has industrial utilities and it can be transformed into jewelry. In modern finance, investors generally buy gold as a hedge or safe haven harbor during economic, political, or social uncertainty (including investment market declines, burgeoning national debt, currency failure, inflation, war and social unrest). In fact, unlike stocks and bonds, the gold price has been constantly thought as the less risky asset.

It is noticeable that given the significance of gold in the modern world, the ability to provide accurate forecasts into the future price of gold will be of primary importance. Moreover, there are benefits from finding the right model that forecasts the gold price more accurately than others do. Therefore, many researchers have been interested in finding a new method and developing a forecasting model for predicting gold prices.

In this study, the new forecasting algorithm for predicting gold prices using concept of fuzzy logic with membership function (MF) model as probability density function (PDF) of normal distribution (PDFMF) is proposed and compared to the triangular MF model (TRIMF) and the trapezoidal MF model (TRAPMF). In section II, some materials and methodologies used in this study are described and discussed. The proposed algorithm is presented in details in section III. The experimental results are shown in section IV and conclusion is summarized in section V.

2 Materials and Methodologies

In this section, some materials and methodologies used in this study are described and their theoretical backgrounds are discussed.

2.1 Rate of return

Rate of return is a measure of the gain or loss on an investment over a specified period of time. It relates directly to the level of movement rate in investment. In general, the rate of return at time n for $n \in \mathbb{N}$ is given by

$$r_n = \frac{S_{n+1} - S_n}{S_n}, \tag{1}$$

where S_n and S_{n+1} are a gold price at time n and $n+1$ per unit weight, respectively.

2.2 Fuzzy logic

Fuzzy logic was firstly introduced by [3] as an alternate technique to classify the data that can belong in more than one cluster. Among other methods, the most important advantage of fuzzy logic is the use of linguistic variables represented in 'If-Then' statements that are easily understood by humans. The concept of fuzzy logic can be used in many problems where information is uncertainly, sometimes called fuzziness. This fuzziness can easier consider by its MF. In other words, we can say that MF represents the level of truth in fuzzy logic. In this work, we consider a triangular MF and a trapezoidal MF as models for forecasting gold prices. Both functions are generally given as follows:

2.2.1 Triangular MF

A triangular MF is defined by three parameters $\{a,b,c\}$ such that $a < b < c$ as follows:

$$\text{triangle}(x;a,b,c) = \begin{cases} 0 & ;x \le a, \\ \dfrac{x-a}{b-a} & ;a \le x \le b, \\ \dfrac{c-x}{c-b} & ;b \le x \le c, \\ 0 & ;c \le x. \end{cases} \tag{2}$$

2.2.2 Trapezoidal MF

A trapezoidal MF is defined by four parameters $\{a,b,c,d\}$ such that $a < b < c < d$ as follows:

$$\text{trapezoid}(x;a,b,c,d) = \begin{cases} 0 & ;x \le a, \\ \dfrac{x-a}{b-a} & ;a \le x \le b, \\ 1 & ;b \le x \le c, \\ \dfrac{d-x}{d-c} & ;c \le x \le d, \\ 0 & ;d \le x. \end{cases} \tag{3}$$

2.2.3 PDF of normal distribution

The normal distribution is the most important and most widely used distribution in statistics because it is a distribution that occurs naturally in many situations. The general formula for the PDF of the normal distribution is

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{4}$$

where μ is the location parameter (mean) and σ is the scale parameter (standard deviation).

In this study, the PDF of normal distribution is considered to be MFs of the forecasting model.

3 Proposed Algorithm

The proposed algorithm has steps as follows:

Step 1:

Gold prices are transformed to the rate of return by (1) and are separated into two parts: the training data and the testing data.

Let $\{r_i\}_{i=1,2,3,\ldots,N}$ be a sequence of the rate of return as the training data.

The number of intervals is used in this algorithm derived from Sturge's formula [4]; see [5], and is of the form as follows:

$$k = \lfloor 1 + \log_2 N \rfloor, \tag{5}$$

where N is the number of training data.

For $j = 1,2,3,\ldots,k+1,$ the j^{th} node is given by

$$a_j = L + (j-1)\frac{U-L}{k}, \tag{6}$$

where L and U are minimum and maximum of $\{r_i\}_{i=1,2,3,...,N}$, respectively.

Step 2:

The fuzzification is a fuzziness modeling step constructed by MFs. This study presents three models as follows:

a) TRIMF model

In this TRIMF model, define $[a_l, a_{l+2}]$ the interval of the l^{th} state for $l = 1,2,3,...,k-1$.

The model is created by mixing MF of each state and the MF of l^{th} state based on (2) is of the form as follows:

$$\mathrm{mf}_l\left(x; a_l, a_{l+1}, a_{l+2}\right) = \begin{cases} 0 & ; x \le a_l, \\ \dfrac{x-a_l}{a_{l+1}-a_l} & ; a_l \le x \le a_{l+1}, \\ \dfrac{a_{l+2}-x}{a_{l+2}-a_{l+1}} & ; a_{l+1} \le x \le a_{l+2}, \\ 0 & ; a_{l+2} \le x. \end{cases} \quad (7)$$

The middle point of l^{th} state is a_{l+1}.

The example of triangular MFs is shown in Fig. 1.

b) TRAPMF model

In this TRAPMF model, define $[a_l, a_{l+3}]$ the interval of the l^{th} state for $l = 1,2,3,...,k-2$.

The model is constructed by combining MF of each state and the MF of l^{th} state based on (3) is of the form as follows:

$$\mathrm{mf}_l\left(x; a_l, a_{l+1}, a_{l+2}, a_{l+3}\right) = \begin{cases} 0 & ; x \le a_l, \\ \dfrac{x-a_l}{a_{l+1}-a_l} & ; a_l \le x \le a_{l+1}, \\ 1 & ; a_{l+1} \le x \le a_{l+2}, \\ \dfrac{a_{l+3}-x}{a_{l+3}-a_{l+2}} & ; a_{l+2} \le x \le a_{l+3}, \\ 0 & ; a_{l+3} \le x. \end{cases} \quad (8)$$

The middle point of l^{th} state can compute by $\frac{a_l + a_{l+3}}{2}$.

The example of trapezoidal MFs is shown in Fig. 2.

c) PDFMF model

In this PDFMF model, define $[a_l, a_{l+2}]$ the interval of the l^{th} state for $l = 1,2,3,...,k-1$.

Dissimilar to first two models, the mean and the standard deviation of rate of return in each interval of l^{th} state denoted by μ_l and σ_l, respectively, have to be calculated.

The model is produced by blending MF of each state and the MF of l^{th} state based on (4) is of the form as follows:

$$\mathrm{mf}_l(x \mid \mu_l, \sigma_l^2) = \frac{1}{\sigma_l \sqrt{2\pi}} e^{-\frac{(x-\mu_l)^2}{2\sigma_l^2}}. \quad (9)$$

The middle point of l^{th} state is μ_l.

The example of MFs as PDF of normal distribution is shown in Fig. 3.

Step 3:

In this step, the following rule is used to predict the rate of return at time n based on the rate of return at time $n-1$.

Rule: If r_{n-1} is a rate of return at time $n-1$, then a rate of return at time n is

$$\hat{r}_n = \frac{\sum_{l=1}^{K} \left(mf_l\left(r_{n-1}\right) \cdot m_l \right)}{\sum_{i=1}^{k} mf_l\left(r_{n-1}\right)}, \quad (10)$$

where mf_l is the MF of i^{th} state, m_l is the middle point of i^{th} state and K is the number of states.

Step 4:

In the last step, the following formula is to convert the predicted rate of return at time n back to the gold price at time $n+1$.

$$\hat{S}_{n+1} = \left(1 + \hat{r}_n\right) S_n, \quad (11)$$

$\hat{S}_{n+1}$ is called a forecasted gold price.

The testing data are used in step 3 in order to predict the rate of return by (10) and then the predicted rate of return from step 3 are imported into step 4 to approximate the gold prices by (11). Thereafter, the forecasted gold prices are compared with real observations.

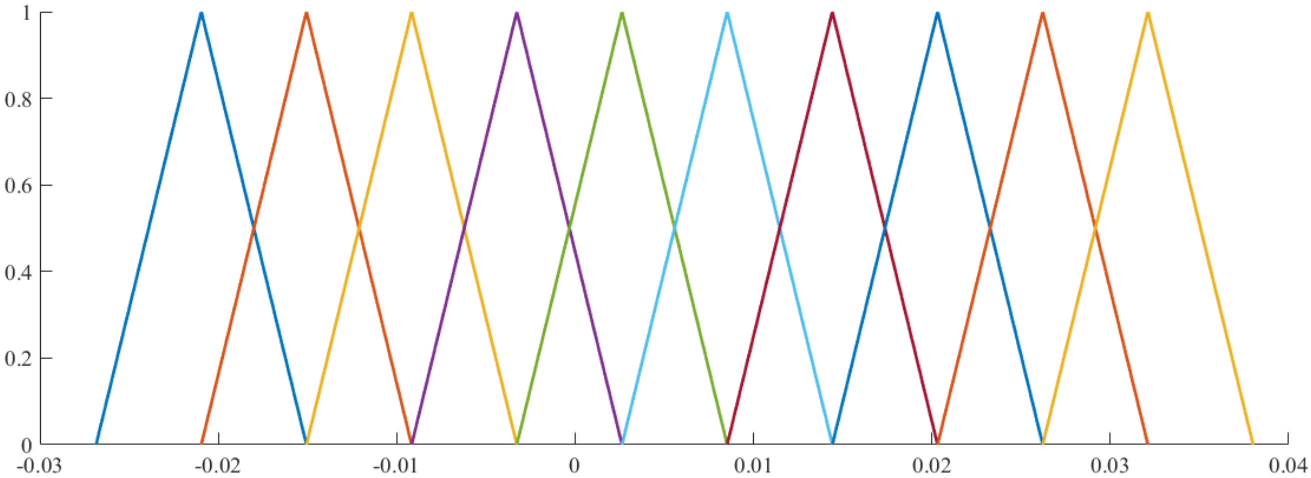

Fig. 1 Example of triangular MFs

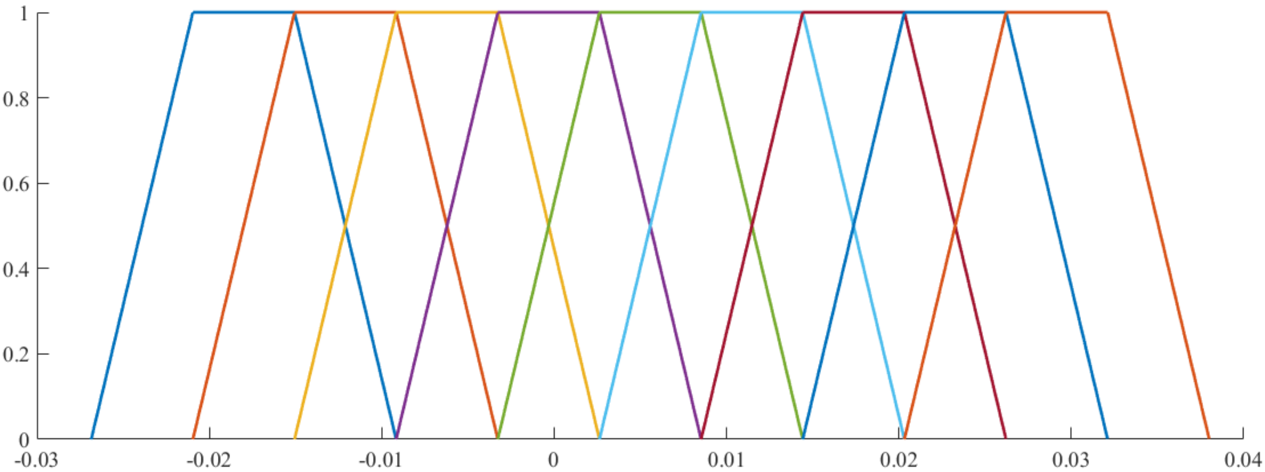

Fig. 2 Example of trapezoidal MFs

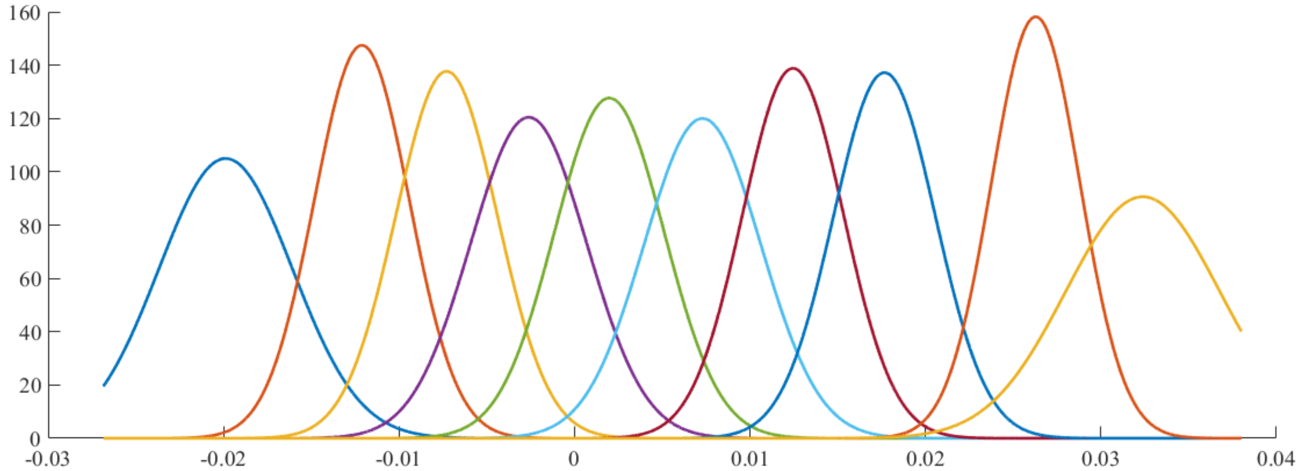

Fig. 3 Example of MFs as PDF of normal distribution

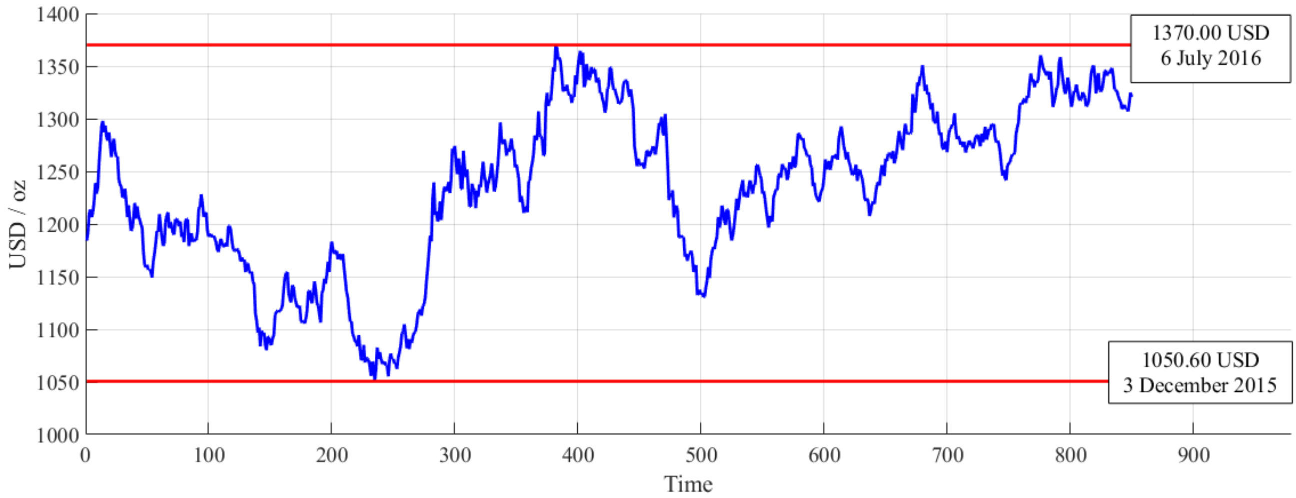

Fig. 4 Daily gold price from January 2015 to May 2018

4 Results

The daily gold prices from January 2015 to May 2018 collected from the London Bullion Market Association (LBMA) are used in this study. The time series of these gold prices are shown in Fig. 4.

This historical period of gold prices is interesting greatly because it has many points of fluctuation and has also signs of two major shocks in the end of 2015 and the middle of 2016, which cause a minimum price and a maximum price in the time series.

In the part of experiment, these gold prices are changed to the rate of return and divided into two parts by using the $\text{Train}:\text{Test}$ percent ratio to obtain the training data and the testing data, respectively. These ratios consist of 99:01, 95:05, 90:10, 80:20, 70:30, 60:40, 50:50, 40:50, 30:70, 20:80 and 10:90.

To measure the performance of the proposed algorithm and to compare three presented models, the root mean squared error (RMSE) and the mean absolute error (MAE) are introduced to use. The RMSE has been used as a standard statistical metric to measure model performance in many research studies and has been approved as a popular measure in a field of forecasting studies in the recent past; see [6-7], while the MAE is another useful measure widely used in model estimations. RMSE and MAE can be calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}\left(\hat{S}_i - S_i\right)^2}{M}}, \tag{12}$$

$$MAE = \frac{\sum_{i=1}^{M}\left|\hat{S}_i - S_i\right|}{M}, \tag{13}$$

where $\hat{S}_i$ is the forecasted gold price, S_i is the actual observation and M is the number of forecasts. A lower RMSE value and a lower MSE value indicate that the forecasting is of higher accuracy. The experimental results are illustrated in Table I.

Table I shows RMSE and MAE values of TRIMF, TRAPMF and PDFMF in all cases of ratio. Obviously, for the same model, RMSE and MAE values of all models decrease when the percentage of the training data increases, especially, the PDFMF. While the RMSE and the MAE values of the TRIMF and the TRAPMF converge to a constant, the RMSE and the MAE values of the PDFMF trend to decline if the training data are increasable; in addition, in case of low percentage of the training data, the PDFMF presents better performance than others in term of RMSE and quite better in term of MAE.

Furthermore, we notice that RMSE and MAE values of the TRIMF and the TRAPMF are rapidly dropped from 50:50 to 60:40 ratio and RMSE values of the PDFMF have two dropped breaks from 30:70 to 40:60 ratio and from 50:50 to 60:40 ratio. These correspond to the gold price data explained at the beginning of this section as well.

Forecasted gold prices predicted by the TRIMF, the TRAPMF and the PDFMF, for instance in case of 10:90 ratio, are demonstrated and compared to actual gold prices by plotting in Fig. 5, Fig. 6 and Fig. 7, respectively.

Table I RMSE and MAE values of three models

Ratio	RMSE			MAE		
	TRIMF	TRAPMF	PDFMF	TRIMF	TRAPMF	PDFMF
99:01	0.14063	0.14063	0.13289	0.00482	0.00482	0.00607
95:05	0.14063	0.14063	0.14438	0.00482	0.00482	0.01559
90:10	0.14063	0.14063	0.16780	0.00482	0.00482	0.03187
80:20	0.14063	0.14063	0.18840	0.00482	0.00482	0.05411
70:30	0.14063	0.15394	0.21136	0.00482	0.00697	0.08064
60:40	0.14063	0.17044	0.28733	0.00482	0.00868	0.14068
50:50	1.68385	1.75205	0.84665	0.10158	0.13519	0.20998
40:60	2.36454	2.42942	0.87291	0.16395	0.20800	0.19697
30:70	3.56348	3.62175	1.63404	0.37226	0.44854	0.38255
20:80	4.22793	4.32538	2.06962	0.59228	0.78969	0.52796
10:90	4.66067	4.79117	2.40234	0.75152	1.04317	0.61107

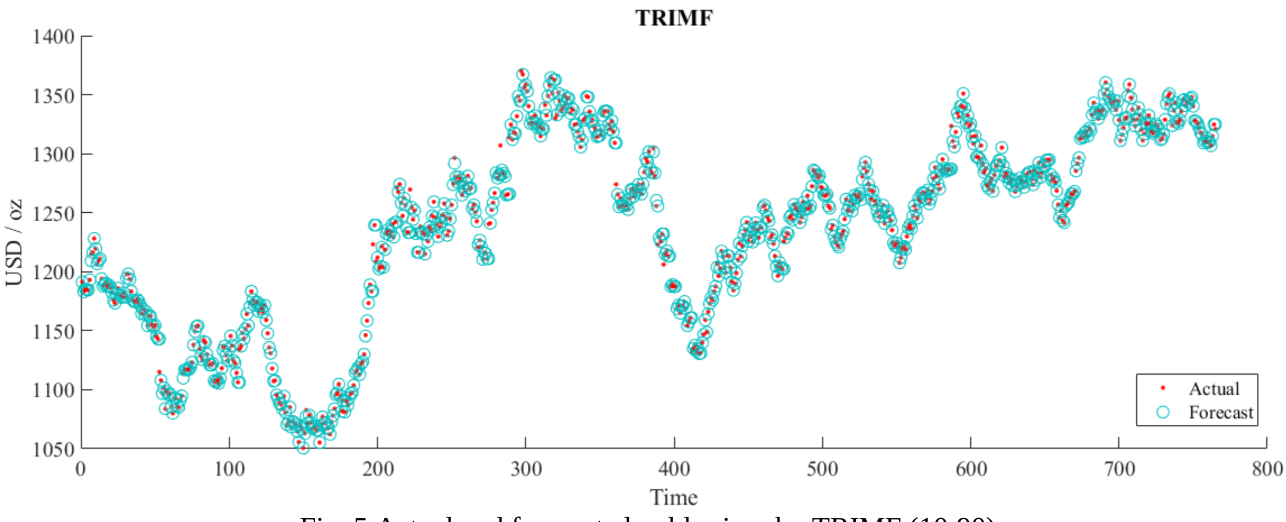

Fig. 5 Actual and forecasted gold prices by TRIMF (10:90)

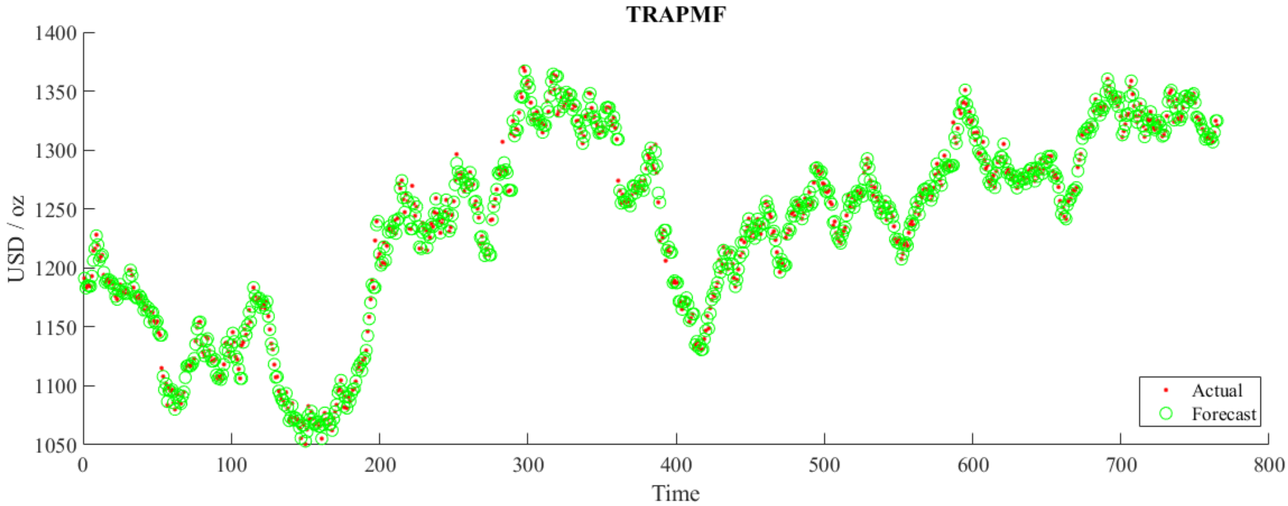

Fig. 6 Actual and forecasted gold prices by TRAPMF (10:90)

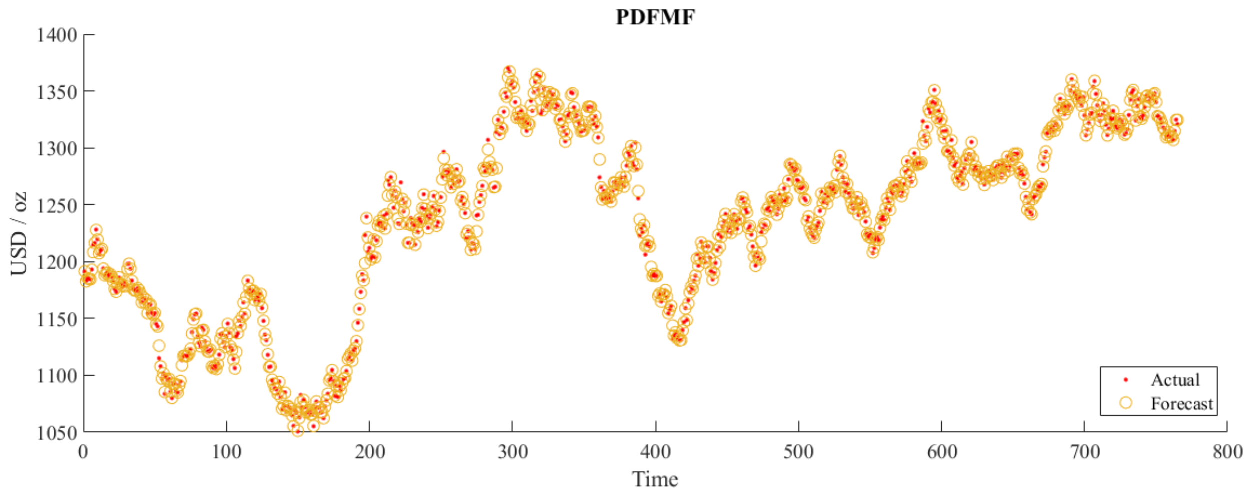

Fig. 7 Actual and forecasted gold prices by PDFMF (10:90)

5 Conclusion

In this paper, the new forecasting algorithm for predicting a gold price using concept of fuzzy logic with the PDFMF is proposed and compared to the TRIMF and TRAPMF. The results can conclude that the proposed algorithm can be applied to the gold price forecasting; besides, the proposed algorithm with the PDFMF can better improve itself than other models when it has more training data and it also show quite better performance than others in term of RMSE and MAE. In the future, the PDFMF will be adopted by updating the data to get a better accuracy in the gold price forecasting.

Acknowledgment I

The authors would like to thank both reviewers for their insightful comments on the paper, as these comments led us to an improvement of the work.

Acknowledgment II

This research is supported by Science Achievement Scholarship of Thailand (SAST).

References:

[1] C. Pierdzioch, M. Risse, and S. Rohloff. "On the efficiency of the gold market: Results of a real-time forecasting approach," *Int. Review of Fin. Analysis*, vol. 32, pp. 95–108. Jan. 2014.

[2] R. Gupta, S. Hammoudeh, W. J. Kim, and B. D. Simo-Kengne. "Forecasting China's foreign exchange reserves using dynamic model averaging: The role of macroeconomic fundamentals, financial stress and economic uncertainty," *N. Am. Journ. of Econ. and Fin.*, vol. 28, pp. 170–189. Feb. 2014.
[3] L. A. Zadeh. "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338 – 353. 1965.
[4] H. A. Sturges. "The Choice of a Class Interval," *Journ. of the Am. Statis.*, vol. 21, No. 153, pp. 65–66. Mar. 1926.
[5] R. J. Hyndman. (1995, January). The problem with Sturges' rule for constructing histograms. Available: https://robjhyndman.com/publications/sturges/
[6] C. Altavilla, and P. de Grauwe. "Non-linearities in the relation between the exchange rate and its fundamentals," *Int.l Journ. of Fin. and Econ.*, vol. 15, No. 1. pp. 1–12. 2010.
[7] C. Beneki, and E. S. Silva. "Analysing and Forecasting European Union Energy Data," *Int. Journ. of Energy and Statis.*, pp. 1–16. Jun. 2013.