

Indonesian-Malaysian Stock Market Models Using Fuzzy Random Time Series

Riswan Efendi^{1,2}, Nureize Arbaiy², Mustafa Mat Deris²

¹Mathematics Department, UIN Suska Riau, Indonesia

²Faculty of Computer Science and Information Technology, UTHM, Malaysia
riswan.efendi@uin-suska.ac.id; nureize@uthm.edu.my; mmustafa@uthm.edu.my

Abstract

Various fuzzy and non-fuzzy models have been presented to forecast the stock market with multiple inputs data or variables. In other words, some of the researchers have overlooked the key success in financial time series forecasting which is minimizing number of inputs. Moreover, most of the existing time series models have been focused on data consisting of single values, or fuzzy numbers without randomness into consideration. In real situations, there exists a genuine need to cope with data that involves the factors of fuzziness and probability. To address the drawbacks, we propose an enhanced fuzzy random auto-regression model for better stock market forecasting using the low-high procedure. This procedure is able to represent the daily prices variations in stocks. The daily stock markets of Indonesia-Malaysia are used as numerical examples and efficiency of the proposed procedure is compared with baselines models.

Keywords: fuzzy random variable, stock market, low-high data, auto-regression time series, forecasting

1. Introduction

In the present global financial crisis, foreign exchange rates, foreign assets and stock market prices are indispensable indicators in the international financial markets. Some Asian countries' currencies, such as, the Indonesian Rupiah (IDR), the Malaysian Ringgit (RM), and others have sharply decreased in value as compared to the US dollar since 1997. This has led to a serious economic impact resulting in the slowdown of the economic activities in these countries. The continuing depreciation trend of the US dollar has attracted many researchers to explore the currency trading, especially in establishing an efficient method to forecast the volatile exchange rates and stock market prices. Many methods have been introduced ranging from the linear and the non-linear regression to artificial intelligence algorithms [1].

Among these methods, the fuzzy logic methods are found to be suitable due to its ability in pattern recognition and handling of the non-probabilistic uncertainties type in the stock market prices data. Moreover, the fuzzy logic is capable of bridging the gap between the numerical data (quantitative information) and the linguistic statement (qualitative information). For example, the investment using technical analysis and fuzzy logic [2], a type-2 fuzzy rule-based expert system model for stock price analysis [3]. Moreover, various fuzzy techniques have been explored into stock markets predictions, such as, fuzzy time series [4], the hybrid fuzzy with others [5],[6] and fuzzy meta-graph [7].

From previous studies, most of the existing fuzzy and non-fuzzy time series analysis have been focused on data consisting of numerical values (single point), or fuzzy numbers without randomness into consideration. In real situations, there exists a genuine need to cope with data that involves the factors of fuzziness and probability. To resolve the data problems, Watada et al. [8] introduced the randomness of fuzzy data into regression linear model. This model based on fuzzy random variables with confidence intervals. Moreover, Shao et al. [9] implemented fuzzy random regression (FR-R) concept into auto-regression (FR-AR) model and the time series data. Its application can be found in forecasting of weekly Shanghai Composite Index data. However, the procedure of fuzzy inputs and shapes of triangular fuzzy number are not deeply discussed. On the other hands, these shapes are very important to obtain the appropriate variances of fuzzy random variables.

Based on surveying of stock market approaches, the key point to successful of stock market forecasting is achieving best results with minimum required in input data [10]. However, some of the previous studies established the forecasting models with multiple number of input variables. On the other hands, the number of variables may negatively influence the auto correlation and the models' parameters.

Motivated by the existing fuzzy models [1] – [9] and the surveying stock market approaches [10], this paper propose an enhanced of fuzzy random auto-regression for stock market forecasting which include two different inputs, namely; low and high prices data. Both inputs are very important to be investigated in improving the forecasting values and achieving the forecasting accuracy. The main contribution of this paper is to provide a better stock market forecasting model using FR-AR concept.

2. Fundamental Theories of Fuzzy Random

2.1. Fuzzy Random Variable (FRV = FRD)

Given some universe, Γ , let Pos be a possibility measure that is defined on the power set $\mathcal{P}(\Gamma)$ of Γ . Let \mathbb{R} be the set of real numbers. A function: $\Gamma \rightarrow \mathbb{R}$ is said to be a fuzzy variable defined on Γ . The possibility distribution μ_Y of Y is defined by $\mu_Y(t) = Pos\{Y = t\}$, $t \in \mathbb{R}$, which is the possibility of event $\{Y = t\}$. For fuzzy variable Y , with possibility distribution μ_Y , the possibility (Pos), necessity (Nec), and credibility (Cr) of event $\{Y \leq r\}$ are given as follows [8], [9]:

$$Pos\{Y \leq r\} = \sup_{t \leq r} \mu_Y(t), \dots\dots\dots(1)$$

$$Nec\{Y \leq r\} = 1 - \sup_{t \geq r} \mu_Y(t), \dots\dots\dots(2)$$

$$Cr\{Y \leq r\} = \frac{1}{2} (1 + \sup_{t \leq r} \mu_Y(t) - \sup_{t \geq r} \mu_Y(t)). \dots\dots\dots(3)$$

From Eq. (3), we note that the credibility measure is an average of the possibility and the necessity measures, i.e., $Cr\{.\} = \frac{Pos\{.\} + Nec\{.\}}{2}$. The motivation behind the introduction of the credibility measure is to develop a certain measure, which is a sound aggregate of the two extreme cases, such as the possibility (which expresses a level of overlap and highly optimistic in this sense) and necessity (that articulates a degree of inclusion and is pessimistic in its nature). Based on credibility measure, the expected value of fuzzy variable is presented as follows [8].

Definition 1. Expected value of fuzzy variable [9]

Let Y be a fuzzy variable. The expected value of Y is defined as:

$$E(Y) = \int Cr\{Y \geq r\} dr - \int Cr\{Y \leq r\} dr, \dots\dots\dots(4)$$

under the condition that the two integral are finite. Assume that $Y = [a^l, c, a^r]_{\Gamma}$ is triangular fuzzy variable (TFV = TFN) whose possibility distribution is given by

$$\mu_Y(t) = \begin{cases} \frac{x-a^l}{c-a^l}, & \text{if } a^l \leq x \leq c \\ \frac{a^r-x}{a^r-c}, & \text{if } c \leq x \leq a^r \\ 0, & \text{otherwise} \end{cases} \dots\dots\dots(5)$$

Making use of Eq. (4), we determine the expected value of Y to be

$$E(Y) = \frac{(a^l + 2c + a^r)}{4}. \dots\dots\dots(6)$$

Definition 2. Fuzzy random variable [9]

Suppose that (Ω, Σ, Pr) is a probability space and F_v is a collection of fuzzy variables defined on possibility space $(\Gamma, P(\Gamma), Pos)$. a fuzzy random variable is a mapping $X : \Omega \rightarrow F_v$ such that for any Borel subset B of R , $Pos\{X(\omega) \in B\}$ is a measurable function of ω . Let X be a fuzzy random variable on Ω . From the previous definition, we know, for each $\omega \in \Omega$, that $X(\omega)$ is a fuzzy variable. Furthermore, a fuzzy random variable X is said to be positive if, for almost every ω , fuzzy variable $X(\omega)$ is positive almost surely. For any fuzzy random variable X on Ω , for each $\omega \in \Omega$, the expected value of the fuzzy variable $X(\omega)$ is denoted by $E(X(\omega))$, which has been proved to be a measurable function of ω , i.e., it is random variable. Given this, the expected value of the fuzzy random variable X is defined as the mathematical expectation of the random variable $E(X(\omega))$.

Definition 3. Expected value of fuzzy random variable [9]

Let X be a fuzzy random variable defined on probability space (Ω, Σ, Pr) . Then, the expected value of X and variance of X are defined as

$$E(X) = \int \Omega [\int Cr\{\xi(\omega) \geq r\} dr - \int Cr\{\xi(\omega) \leq r\} dr] Pr(\omega), \dots\dots\dots(7)$$

$$Var(X) = E(X - e)^2, \dots\dots\dots(8)$$

where $e = E(X)$ is given by Eq. (7).

2.2. Fuzzy Random Auto-Regression (FR-AR) Model

Based on [9], the fuzzy random auto-regression (FR-AR) model can be defined as input and output data Y_{t-p} for all $p = 0, 1, 2, \dots, n$ are fuzzy random variables, which are written as:

$$Y_t = \cup_{i=1}^n [(Y_{it}^l, Y_{it}^c, Y_{it}^r)_T, P_{it}], \dots\dots\dots(9)$$

where Y_t is a time series data at time- t and its formatted as a triangular fuzzy number [left, l ; center, c ; ; right, r]. From Eq. (10), all values given as fuzzy numbers with probabilities, P_{it} .

Let a simple FR-AR model with coefficients $[\phi_1^l, \phi_1^r]$ and $[\phi_2^l, \phi_2^r]$ can be written as:

$$(Y_t)_T = [\phi_1^l, \phi_1^r](Y_{t-1})_T + [\phi_2^l, \phi_2^r](Y_{t-2})_T + [e_t^l, e_t^r], \dots\dots\dots(10)$$

To estimate both coefficient and estimated FR-AR model can be derived by following steps:

- Step 1: Transform actual time series data (single point) into fuzzy format.
- Step 2: Divide the fuzzy data (FD) into the fuzzy random data [left, center, right] with their random probabilities (Pr_{i1}, Pr_{i2}) .
- Step 3: Calculate the expected value (EV) and standard deviation ($Std.Dev$) of fuzzy random data (FRD) as determined in Step 2, respectively.
 $EV = E(Y) = (\text{Center of FRD1} \times Pr_1) + (\text{Center of FRD2} \times Pr_2),$
 $Var(Y) = E(Y - e)^2$

Standard deviation (*Std.Dev*) = $s(Y) = \sqrt{\text{Var}(Y)}$.

- Step 4: Determine the confidence-interval (CI) of FRD.
- Step 5: Estimate CI for each coefficient model by using linear programming (LP) approach.
- Step 6: From Step 5, state the confidence interval (CI) estimated for each coefficient model.

$$\hat{Y}_t = [\hat{\theta}_1^l, \hat{\theta}_1^r]Y_{t-1} + [\hat{\theta}_2^l, \hat{\theta}_2^r]Y_{t-2}$$

3. Results and Analysis

3.1. Proposed Data Preparation

In many existing studies, the single point format (left side of Figure 1) is frequently used to forecast the stock market prices data. While, it is insufficient to interpret such the financial or economic systems only based the average values (single points). Based on this format, the uncertainty, the randomness, the possibility or the human behaviors cannot be coped [9]. So it is desirable to propose the low-high procedure (right side of Figure 2) and fuzzy random

variable which can capture the ambiguity of a structure when we analyze and forecast the stock market system which has many vague factors.

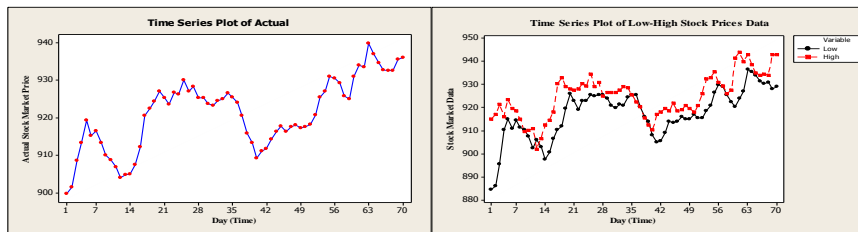


Figure 1. The single and low-high stock markets data

Based on Figure 1 (right side), the actual of low-high data can be transformed into triangular fuzzy number (TFN) as illustrated in Figure 2.

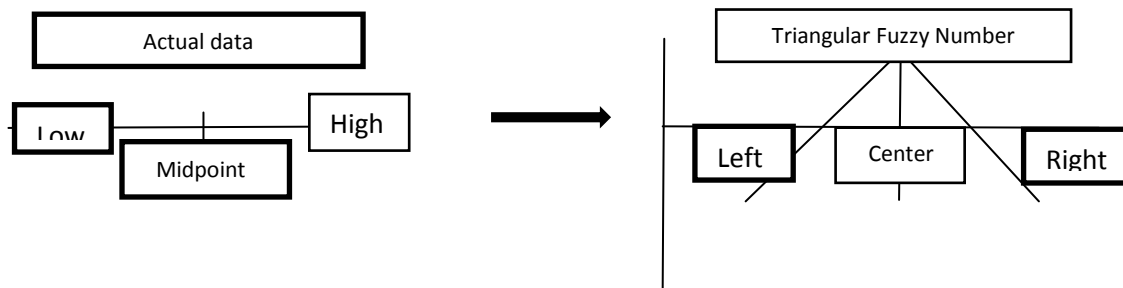


Figure 2. Transformation of actual low-high data into TFN

Based on Figure 2, the transformation between values can be detailed as:

- Low data → Left spread of TFN
- Midpoint → Center of TFN
- High data → Right spread of TFN

Let Y_t^{low} and Y_t^{high} be a lowest and highest stock market data at time- t , then both of them can be written as:

$$[Y_t^{low}, Y_t^{high}] \rightarrow [Y_t^{left}, Y_t^{right}] \rightarrow [Y_t^{left}, Y_t^{center}, Y_t^{right}]_T, \dots\dots\dots(11)$$

By using Eq. (11), all actual of low-high stock market data can be transformed into fuzzy format as presented in Table 1.

Table 1. The low-high prices of stock market and fuzzy data

Day (Time)	Low	High	Fuzzy data (FD)
1	Y_{1l}	Y_{1h}	$[Y_{1l}, Y_{1h}]$
...
n	Y_{nl}	Y_{nh}	$[Y_{nl}, Y_{nh}]$

3.2. Building Stock Market Model Using FR-AR

By considering Sections 2.2 and 3.1, FR-AR of stock market model can be established by following steps:

Step 1. Divide fuzzy data (FD) into two-fuzzy random data (FRD1, FRD2) as input data:

$$[Y_t^l, Y_t^r] \rightarrow \frac{Y_t^r - Y_t^l}{3} = d_t, t = 1, 2, \dots, n. \dots\dots\dots(12)$$

Here, we define FRD 1 and FRD 2 as TFN time series data:

$$\rightarrow [Y(t-1)_T] = \left[(Y_t^l), \left(Y_t^l + \frac{1}{3} d_t \right), \left(Y_t^l + \frac{2}{3} d_t \right) \right]_T, \dots\dots\dots(13) \text{FRD1}$$

$$\text{FRD2} \rightarrow [Y(t-2)_T] = \left[\left(Y_t^l + \frac{1}{3} d_t \right), \left(Y_t^l + \frac{2}{3} d_t \right), (Y_t^r) \right]_T, \dots\dots\dots(14)$$

By using Eq. (12-14), fuzzy random data can be detailed in Table 2.

Table 2. Fuzzy random time series data (FRD1 and FRD2)

Time	Fuzzy data	FRD1, $[Y(t-1)_T]$	FRD2, $[Y(t-2)_T]$
1	$[Y_1^l, Y_1^r]$	$\left[(Y_1^l), \left(Y_1^l + \frac{1}{3} d_1 \right), \left(Y_1^l + \frac{2}{3} d_1 \right) \right]_T$	$\left[\left(Y_1^l + \frac{1}{3} d_1 \right), \left(Y_1^l + \frac{2}{3} d_1 \right), (Y_1^r) \right]_T$
⋮	⋮	⋮	⋮
n	$[Y_n^l, Y_n^r]$	$\left[(Y_n^l), \left(Y_n^l + \frac{1}{3} d_n \right), \left(Y_n^l + \frac{2}{3} d_n \right) \right]_T$	$\left[\left(Y_n^l + \frac{1}{3} d_n \right), \left(Y_n^l + \frac{2}{3} d_n \right), (Y_n^r) \right]_T$

Step 2: Calculate the expected value (EV) and standard deviation ($Std.Dev$) for FRD-output by using FRD1 and FRD2 and their random probabilities (Pr_1, Pr_2). While, (EV) of both FRDs are derived by implementation of integration technique as follows:

$$E(\text{FRD1}) = E(Y_1^2) = \int_0^{\omega} Cr\{Y_1^2 \geq r\} dr, \dots\dots\dots(15)$$

and

$$E(\text{FRD2}) = E(Y_2^2) = \int_0^{\infty} Cr\{Y_2^2 \geq r\} dr \dots\dots\dots(16)$$

Based on Eqs. (15) and (16), the variance of FRD can be written as:

$$\text{Var}(Y) = Pr_1 E(Y_1^2) + Pr_2 E(Y_2^2), \dots\dots\dots(17)$$

Step 3: Define the confidence-interval (CI) of FRD. From Eq. (17), it standard deviation can be written as:

$$\text{Std.Dev}(Y) = \sqrt{Pr_1 E(Y_1^2) + Pr_2 E(Y_2^2)}, \dots\dots\dots(18)$$

By using $E(Y)$ and $\text{Std.Dev}(Y)$ of a fuzzy random variable (FRD), then we consider the one-sigma confidence $(1 \times \sigma)$ interval of each output-fuzzy random variable, they can be expressed (for $t = 1, 2, \dots, n$) in Table 3.

Table 3. Input and output fuzzy random variables (FRD)

t	Output, $[Y(t)]_T$	Input	
		$[Y(t-1)]_T$	$[Y(t-2)]_T$
1	$[E_1(Y) - \text{Std.Dev}_1(Y), E_1(Y) + \text{Std.Dev}_1(Y)]$	$\left[\left(Y_1^l, \left(Y_1^l + \frac{d_1}{3} \right), \left(Y_1^l + \frac{2d_1}{3} \right) \right) \right]$	$\left[\left(Y_1^l + \frac{d_1}{3}, \left(Y_1^l + \frac{2d_1}{3} \right), (Y_1^r) \right) \right]$
\vdots	\vdots	\vdots	\vdots
n	$[E_n(Y) - \text{Std.Dev}_n(Y), E_n(Y) + \text{Std.Dev}_n(Y)]$	$\left[\left(Y_n^l, \left(Y_n^l + \frac{d_n}{3} \right), \left(Y_n^l + \frac{2d_n}{3} \right) \right) \right]$	$\left[\left(Y_n^l + \frac{d_n}{3}, \left(Y_n^l + \frac{2d_n}{3} \right), (Y_n^r) \right) \right]$

Step 4: Estimate the coefficients, $(Y_{t-1})_T$ and $(Y_{t-2})_T$ by using system linear programming (LP) approach as follows:

Objective function: $\min J(\phi) = \sum_{i=1}^n (\phi_i^r - \phi_i^l)$,

Subject to

$$Y_1^l \phi_{11}^l + \left(Y_1^l + \frac{d_1}{3} \right) \phi_{12}^l \leq (E_1(Y) - \text{Std.Dev}_1(Y)) \dots Y_n^l \phi_{n1}^l + \left(Y_n^l + \frac{d_n}{3} \right) \phi_{n2}^l \leq (E_n(Y) - \text{Std.Dev}_n(Y))$$

and

$$\left(Y_1^l + \frac{2d_1}{3} \right) \phi_{11}^r + (Y_1^r) \phi_{12}^r \geq (E_1(Y) + \text{Std.Dev}_1(Y)) \dots \phi_{n1}^r + (Y_n^r) \phi_{n2}^r \geq (E_n(Y) + \text{Std.Dev}_n(Y))$$

Step 5: State the estimated FR-AR model by using result from Step 4.

$$\hat{Y}_t = [\tilde{\phi}_1^l, \tilde{\phi}_1^r](Y_{t-1})_T + [\tilde{\phi}_2^l, \tilde{\phi}_2^r](Y_{t-2})_T \dots\dots\dots(19)$$

Moreover, Eq. (19) can be used to predict the low (\hat{Y}_{lt}), the medium (\hat{Y}_{mt}) and the high (\hat{Y}_{ht}) stock prices as follows:

$$\hat{Y}_t = \begin{cases} [\tilde{\phi}_1^l](Y_{t-1})_T + [\tilde{\phi}_2^l](Y_{t-2})_T, & \text{for low data} \\ [\tilde{\phi}_1^r](Y_{t-1})_T + [\tilde{\phi}_2^r](Y_{t-2})_T, & \text{for high data} \\ \frac{\hat{Y}_{lt} + \hat{Y}_{ht}}{2}, & \text{for medium data} \end{cases} \dots\dots\dots(20)$$

3.3. Implementation

3.3.1. Indonesian Stock Market Model

In this section, the proposed model is applied for forecasting the IHSG from January to July 2009. By following steps given in Section 3.2, proposed FR-AR model for this stock can be written as:

$$\hat{Y}_t = 0.5547649(Y_{t-1})_T + 0.4452608(Y_{t-2})_T \dots\dots\dots(21)$$

In this case, we obtain the same low-high parameters for $\hat{\theta}_1^l$ and $\hat{\theta}_2^h$. Moreover, the forecasting error of the proposed model is compared with the baseline models as shown in Table 4.

Table 4. MSE comparison between proposed model and baselines models

Model	MSE (Mean Square Error)	
	Training data (80%)	Testing data (20%)
ARIMA (0, 1, 0)	846.04	8807.94
Fuzzy time series (FTS)	993.96	4724.68
Proposed FR-AR	670.53 ^{**}	978.66 ^{**}

Based on Table 4, the error comparison indicates that MSE of proposed FR-AR is smaller than ARIMA and fuzzy time series (FTS) for training and testing data, respectively. By proposed model, the forecasting error can be reduced very significantly if compared with both existing models. In other words, the capability of proposed model is better than others in IHSG forecasting. Moreover, fitting between actual and its training-testing data are also illustrated in Figure 3. While in Figure 4, the illustration between actual-low-medium-high of IHSG and its forecasted values are presented.

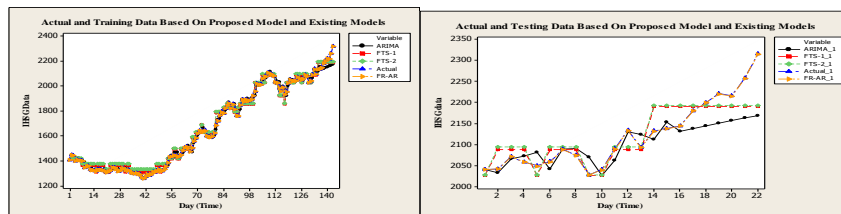


Figure 3. Fitting between actual, training and testing of IHSG data

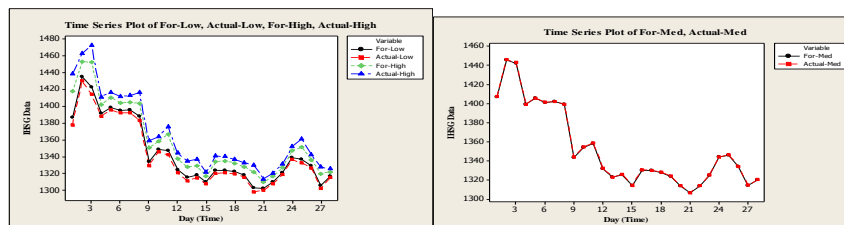


Figure 4. Fitting between actual-low-medium-high and its forecasted values

3.3.2. Malaysian Stock Market Model

In this section, the proposed model is applied for forecasting the KLSE from January to July 2009. By using steps given in Section 3.2, proposed FR-AR for this data can be written as :

$$\hat{Y}_t = 0.4026755(Y_{t-1})_T + 0.5972301(Y_{t-2})_T \dots\dots\dots(22)$$

In this case, we obtain the same low-high parameters for $\hat{\theta}_1^l$ and $\hat{\theta}_2^h$. Moreover, the forecasting error of the proposed model is compared with the baseline models as shown in Table 4.

Table 4. MSE comparison between proposed model and baselines models

Model	MSE	
	Training data (80%)	Testing data (20%)
ARIMA (0, 1, 0)	138.10	355.97
Fuzzy time series (FTS)	78.53	84.56
Proposed FR-AR	60.19**	69.21**

From Table 5, the error comparison indicates that MSE of proposed FR-AR is smaller than ARIMA and fuzzy time series (FTS) for training and testing data, respectively. By proposed model, the forecasting error can be reduced very significantly if compared with both existing models. In other words, the capability of proposed model is better than others in KLSE forecasting. From Sections 3.1, 3.2 and Eq. (20), we can highlight the main advantage of proposed model are: (i). the low-high data are minimum input, (ii). The variation/variability of daily stock prices can be covered, (iii). It can be implemented to forecast three different values; low-medium-high directly. In other words, the existing models have overlooked with these advantages.

4. Conclusion

In this paper, we propose an enhanced FR-AR model for stock market forecasting based on low-high input procedure. In the forecasting results, from numerical examples show the result of the proposed formulation is to obtain FR-AR models easily, giving results that are comparable (or improved) with those obtained with more complex models proposed by other researchers. In the forecasting of KLSE and IHSG, the MSE of proposed model is smaller than ARIMA and FTS models in term of training-testing data. At the same time, proposed FR-AR can directly forecast with three different data, such as, low, midpoint and high values. These values are very important to help the decision makers and financial analyst in the monitoring of the present and future stock market data, respectively.

Acknowledgment

Thanks to Research-Innovation Center (ORICC)-UTHM, Malaysia and UIN Suska Riau, Indonesia.

References

- [1] Efendi, R., Z. Ismail, Z., and Deris, M. M. Improved Weight Fuzzy Time Series Used in the Exchange Rates Forecasting of US Dollar to Ringgit Malaysia. *International Journal of Computational Intelligence and Applications*. 2013; 12: 71-90.
- [2] Yuzie, H. Decision support system for foreign exchange trading. *Proceeding of International Fuzzy Engineering Symposium*. 1991; 971-982.
- [3] Tseng, F. M., Tzeng, G. H., Yu, H. C., and Yuan, B. J. C. Fuzzy ARIMA Model for Forecasting the Foreign Exchange Market. *Fuzzy Sets and Systems*. 2008; 118: 9-19.
- [4] Song, Q., and Chissom, B. S. Forecasting enrollments with fuzzy time series – Part 1. *Fuzzy Sets and Systems*. 1993; 54: 1 – 9.
- [5] Wei, L. Y., Cheng, C. H., and Wu, H. H. A Hybrid ANFIS Based on-Period Moving Average Model to Forecast TAIEX Stock. *Applied Soft Computing*. 2014; 19: 86-92.
- [6] Song, Q. On A New Framework of Autoregressive Fuzzy Time Series Models. *Industrial Engineering Management System*. 2014; 13: 357 – 368.
- [7] Anbalagan, T., and Maheswari, S. U. Classification and Prediction Stock Market Based on Fuzzy Meta-Graph. *Proceeding of Procedia Computer Science*. 2015; 47: 214 – 221.
- [8] Watada, J., Wang, S., and Pedrycz, W. (2009). "Building Confidence-Interval Based Fuzzy Random Regression Models," *IEEE Transaction Fuzzy System*. 2009; 17: 1273 – 1283.
- [9] Shao, L., Tsai, Y-H., Watada, J., and Wang, S. Building Fuzzy Random Auto-Regression Model and Its Application. Watada et al. (Eds): *Intelligence Decision Technology*. 2012; 1: 155 – 164.
- [10] Atsalakis, G. S., and Valavanis, K. V. Surveying Stock Market Techniques-Part II – Soft Computing Methods. *Expert Systems and Applications*. 2009; 36: 5932 – 5941.