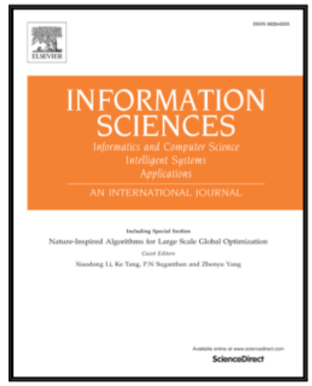
orecasting_based_on_fuzzy_random_auto-regression_time_series.pdf

by

Accepted Manuscript



A New Procedure in Stock Market Forecasting Based On Fuzzy Random Auto-Regression Time Series Model

Riswan Efendia , Nureize Arbaiy , Mustafa Mat Deris

PII: S0020-0255(18)30097-5
DOI: 10.1016/j.ins.2018.02.016
Reference: INS 13421

To appear in: *Information Sciences*

Received date: 14 October 2016
Revised date: 22 December 2017
Accepted date: 8 February 2018

Please cite this article as: Riswan Efendia , Nureize Arbaiy , Mustafa Mat Deris , A New Procedure in Stock Market Forecasting Based On Fuzzy Random Auto-Regression Time Series Model, *Information Sciences* (2018), doi: 10.1016/j.ins.2018.02.016

A New Procedure in Stock Market Forecasting Based On Fuzzy Random Auto-Regression Time Series Model

Riswan Efendi[1,2,*], Nureize Arbaiy[1], Mustafa Mat Deris[1]

[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia

[2]Mathematics Department, Faculty of Science and Technology, State Islamic Univesity of Sultan Syarif kasim Riau, 28293 Panam, Pekanbaru, Indonesia

[1]riswan@uthm.edu.my; [1]nureize@uthm.edu.my; [1]mmustafa@uthm.edu.my;

Abstract. Various models used in stock market forecasting presented have been classified according to the data preparation, forecasting methodology, performance evaluation, and performance measure. However, these models have not sufficiently discussed in data preparation to overcome randomness, as well as uncertainty and volatility of stock prices issues in achieving high forecasting accuracy. Therefore, the focus of this paper is the data preparation procedure of triangular fuzzy number to build an improved fuzzy random auto-regression model using non-stationary stock market data for forecasting purposes. The improved forecasting model considers two types of input, which are data with low-high and single point values of stock market prices. Even though, low-high data present variability and volatility in nature, the single data has to be form in symmetry left-right spread to present variability and standard error. Then, expectations and variances, confidence-intervals of fuzzy random data are constructed for fuzzy input-output data. By using the input-output data and simplex approach, parameters of the model can be estimated. In this study, some real data sets were used to represent both types of inputs, which are the Kuala Lumpur stock exchange and Alabama University enrollment. The study found that variability and spread adjustment are important factors in data preparation to improve accuracy of the fuzzy random auto-regression model.

Keywords: low-high procedure, left-right spread, fuzzy random variable, auto-regression model, stock market

*Correspondent author: riswan@uthm.edu.my;

1 Introduction

With the current world economic crisis, foreign exchange rates, foreign assets and stock market prices are indispensable indicators in the international financial markets. Since 1997, some Asian countries' currencies, such as, the Indonesian Rupiah (IDR), and Malaysian Ringgit (RM), have sharply decreased in value in comparison to the US dollar. This scenario has had a serious economic impact resulting in the slowdown of the economic activities in these countries. The continuing depreciation trend of the

ASEAN currencies against US dollar has attracted many researchers, financial analysts, and academicians to explore the currency trading, especially in the development of an efficient method to forecast volatile exchange rates and stock market prices. Many methods to address these problems have been introduced ranging from linear and the non-linear regression to artificial intelligence algorithms [21].

In the past few decades, most of the stock market analyses were derived using statistical-time series models, such as, regression, exponential smoothing, ARIMA, state space and Kalman filter. However, these models used a single point and they hardly fit non-linear data. Moreover, the stock market is a complex and dynamic system with noisy; non-stationary and chaotic time series [22, 38]. In order to solve both problems, artificial intelligence models, such as, neural network, genetic algorithm, and support vector machine based on the statistics have been suggested. Although these models can produce better forecast values, they cannot be implemented to predict the fuzzy data [35]. Besides that, the fuzzy models are tolerant of imprecision, uncertainty and approximation. As a result, they have become popular among academic community [39].

The conventional fuzzy time series (FTS) models are applied in real data applications, such as, university enrollment [24, 34], stock index price [1, 7, 9, 11, 36, 44, 46], financial-exchange rate [15, 17, 19, 21, 28], and electricity load [20, 23]. Besides that, the recent hybrid fuzzy with particle swarms optimization [13, 36], neural network [49], support vector machine [10], granular computing [7], probabilistic hesitant fuzzy [51], intelligent pattern recognition [14] and data mining [41, 45] models are also explored to improve the accuracy of forecasting stock market. In building the forecasting model, the single point data are still applicable and using as input. However, the single point data are not guaranteed as reliable input. Therefore, validating such single point data to achieve better accuracy in building the forecasting model should be considered. For illustration, if the stock market data are observed more than twice (multiple observations) a day. We will use average data to represent our daily observation generally. There is a possibility when the standard deviation (volatility) of the average data is very huge due to various reasons, such as, human error, machine error, measurement tools are not working well, or other political issues occur during data collecting procedure. If the single point data are to be used in building the forecasting model, eventually, the huge standard deviation problems may contribute to increase the forecasting error indirectly.

Additionally, the existing FTS and non-FTS models have not considered three major issues in data representation, such as, randomness, vagueness, and possibility. For example, different financial analysts may arrive at different conclusions when observing the trend of stock prices within a certain period. There is a possibility that some analysts will evaluate the stock prices performance based their experience and expertise. As a result, there are various possibilities in the evaluation, such as, "low, medium, high, and very high", or other prices. The variations in the performance evaluation are considerably stochastic in nature. In these cases, representing the performance evaluation using a single value is inappropriate. Therefore, the price evaluation would be more reasonable if presented in fuzzy random form [43].

The potential of fuzzy theories in improving forecasting model can be found in various applications due to its well-known capability in bridging the gap between the numerical data (quantitative information) and the linguistic statement (qualitative information). Tanaka *et al*. [37] presented a linear regression analysis with fuzzy model, which treats fuzzy data instead of statistical data whereas Chang [6] discussed a fuzzy least-squares regression by using weighted fuzzy-arithmetic and the least-square fitting criterion. There are other instances when the fuzzy concepts have been applied into the time series data by Watada [42]. From these examples, the researchers and academicians have already applied the fuzzy theories and concepts into the forecasting data, especially, the cross section and time series data.

Watada *et al*. [43] considered problems associated with randomness, fuzziness, and possibility in the presented scenario by introducing a fuzzy random regression model. In another study, Arbaiy and Watada [2] implemented a fuzzy random regression model for oil palm fruit forecasting. These researchers [2] and [43] focused on cross section data whereas, Shao *et al*. and Efendi *et al*. [19, 32] introduced fuzzy random regression (FR-R) to auto-regression (FR-AR) model for time series data applications. FR-AR model was applied the weekly Shanghai Composite Index data. In another application, Efendi *et al*. [18] emphasized the importance of adjusting left-right spread of triangular fuzzy number (TFN) in yearly electricity load data of Taiwan, Wang *et al*. [41] proposed multi-period portfolio selection with dynamic risk/expected return level under fuzzy random uncertainty. The recent studies, the fuzzy and random fuzzy variables theories are applied to solve supply chain master planning and network [40, 48]. The existing literatures presented have illustrated various FR-R and FR-AR models that take the input data directly, without considering any preparation procedure.

The main motivation of this paper is to provide a systematic procedure in handling low-high and single point of stock market prices data using fuzzy random variable. Additionally, the contribution of this paper is to primarily build a FR-AR model that accommodates both, low-high and single point of stock market data. In the case of low-high data set, the importance of variability is emphasized, in the single point data set, the left-right spread procedure is introduced. Besides, influence of the procedure towards forecasting accuracy will be discussed.

The rest of paper is organized in following manner. Section 2 contains the explanation on the fundamentals of fuzzy random variables and fuzzy random auto-regression model. This followed by the construction of proposed low-high procedure is presented in Section 3, the empirical analysis using the daily Kuala Lumpur Stock Exchange (KLSE) and enrollment of Alabama University are presented in Section 4. The final section of the paper is the conclusion.

2 Fundamental Theories of Fuzzy Random Variables

This section provides the basic concepts of fuzzy random variables and fuzzy random auto-regression models. Both concepts are very important in providing information to support the background, methodology, and forecasting procedure discussed in this paper.

2.1 Triangular Fuzzy Number (TFN)

Definition 1: Triangular fuzzy number [31]

A triangular fuzzy number denoted by $M = \langle m, \alpha, \beta \rangle$ has the membership function,

$$\mu_M(x) = \begin{cases} 0, & for\ x < m - \alpha \\ 1 - \frac{m-x}{\alpha}, & for\ m - \alpha < x < m \\ 1, & for\ x = m \\ 1 - \frac{x-m}{\beta}, & for\ m < x < m + \beta \\ 0, & for\ x > m + \beta \end{cases} . \quad (1)$$

The point m, with membership grade of 1, is called the mean value and α, β are the left hand and right hand spreads of M respectively. A TFN is said to be symmetric if both its spreads are equal, i.e., if $\alpha = \beta$ and is sometimes denoted by $M = \langle m, \alpha \rangle$. Based on Definition 1, the illustration is explained as follows:

Sometimes it may happen that some data or numbers cannot be specified precisely or accurately due to the error of a measuring technique or instruments etc. For example, if the height of a person is recorded as 160 cm, it is impossible in practice to measure the height accurately. In reality, the height is actually about 160 cm and it may be a bit more or less than 160 cm. Thus the height of the person can be written more precisely as a triangular fuzzy number $(160 - \alpha, 160, 160 + \alpha)$, where α is the left and right spreads. In general, a symmetry TFN "a" can be written as $(a - \alpha, a, a + \alpha)$, where α is the left and right spreads of a respectively. Alternatively, $(a - \alpha, a, a + \alpha)$ can be represented as $\langle a, \alpha \rangle$ and Figure 1.

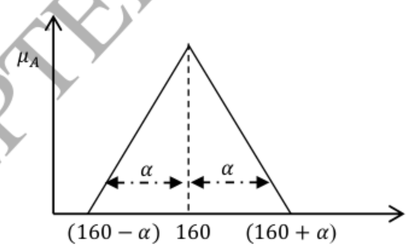


Figure 1. Triangular fuzzy number $\tilde{A} = (160 - \alpha, 160, 160 + \alpha)$

2.2 Fuzzy Random Variables (FRV)

Given some universe, Γ, let Pos be a possibility measure that is defined based on the power set $\mathcal{P}(\Gamma)$ of Γ . Let R be the set of real numbers. A function : $\Gamma \rightarrow R$ is said to be a fuzzy variable defined on Γ [31]. The possibility distribution μ_Y of Y is defined by

$\mu_Y(t) = Pos\{Y = t\}$, $t \in R$, which is the possibility of event $\{Y = t\}$. For fuzzy variable Y, with possibility distribution μ_Y, the possibility, necessity, and credibility of event $\{Y \leq r\}$ are given as follows [32, 43]:

$$Pos\ \{Y \leq r\} = sup\ \mu_Y(t), t \leq r, \tag{2}$$

$$Nec\ \{Y \leq r\} = 1 - sup\ (t), t \geq r, \tag{3}$$

$$Cr\ \{Y \leq r\} = ½\,(1 + sup_{t \leq r}\mu_Y(t) - sup_{t \geq r}\mu_Y(t)). \tag{4}$$

From Eq. (4), 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 (expresses a level of overlap and is highly optimistic in this sense) and necessity (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.

Definition 2. Expected value of fuzzy variable [30]
Let Y be a fuzzy variable and the expected value of Y is defined as:

$$E(Y) = \int Cr\ \{Y \geq r\}\, dr - \int Cr\ \{Y \leq r\}\, dr, \tag{5}$$

under the condition that the two integral are finite. Assume that $Y = [a^l, c, a^r]_T$ is triangular fuzzy variable (TFV = TFN) as explained in Definition 1. Making use of Eq. (5), the expected value of Y to be determined as follows:

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

Definition 3. Fuzzy random variable [29]

Suppose that (Ω, Σ, Pr) is a probability space and F_y is a collection of fuzzy variables defined based on possibility space $(\Gamma, \mathcal{P}(\Gamma), Pos)$, a fuzzy random variable is a mapping $X : \Omega \rightarrow F_y$ 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 of 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, the fuzzy variable $X(\omega)$ is almost surely to be positive. 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 proven to be a measurable function of ω, i.e., it is a random variable. Based on this condition, the expected value of the fuzzy random variable X is defined as a mathematical expectation of the random variable $E(X(\omega))$.

Definition 4. Expected value of fuzzy random variable [29]

If ξ is a FRV defined as the probability space (Ω, A, Pr), then the expected value of ξ is defined by

$$\mathrm{E}[\xi] = \int_0^{\infty} \mathrm{Cr}\{\xi \geq x\}\, dx - \int_{-\infty}^{0} \mathrm{Cr}\{\xi \geq x\}\, dx. \tag{7}$$

Let ξ be a FRV with finite expected value $\mathrm{E}[\xi]$. Then the variance of ξ is

$$\mathrm{Var}\,[\xi] = E[(\xi - E[\xi])^2]. \tag{4}$$

2.3 Fuzzy Random Auto-Regression (FR-AR) Model

In a time series analysis, the stationarity can be recognized from the time plot. If it observed that there are n values $y_1, y_2, \ldots, y_n$ of a time series, then a plot of these values (against time) can be drawn to determine whether the time series is stationary. If n values seem to fluctuate with constant variation around a constant mean μ, then it is reasonable to believe that the time series is stationary. On the other hand, if the n values do not fluctuate around a constant mean or do not fluctuate with constant variation, then it is reasonable to believe that the time series in non-stationary [35]. The stationarity is one of the characteristic features in time series. This concept can be explained by using definitions and properties. The stationarity can be investigated in three linear stationary models such as autoregressive (AR), moving average (MA) and autoregressive-moving average (ARMA). In this paper, the explanation is subject to AR(p) model only.

Let us assume now that Y_t is a stationary series, the autoregressive or AR(p) model can be written as [5]:

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + e_t, \tag{11}$$

where $\emptyset_1, \ldots, \emptyset_p$ are coefficients of $Y_{t-1}, \ldots, Y_{t-p}$, respectively, e_t is an error models at time-t. Based on [32], the input and output data Y_{t-k} for all $k = 0, 1, 2, \ldots, n$ are fuzzy random variables, which are written as:

$$Y_t = \cup_{i=1}^{n} \left[\left(Y_{it}^{l}, Y_{it}^{c}, Y_{it}^{r} \right)_T, P_{it} \right], \tag{12}$$

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]. In this equation, all values are given as fuzzy numbers with probabilities. Thus fuzzy linear regression model is denoted as:

$$\tilde{Y}_t = \varphi_1 Y_{t-1} + \cdots + \varphi_k Y_{t-k} + u, \tag{13}$$

where $\tilde{Y}_t$ denotes an estimate of the output and $\varphi_1, \ldots, \varphi_k$ are coefficients which have real values when triangular fuzzy random data Y_{t-k} is presented in Table 1.

Table 1. Fuzzy random input-output data

Sample	Output	Input		
1	Y_t	Y_{t-1}	Y_{t-2}	$\cdots$ Y_{t-k}
2	Y_{t-1}	Y_{t-2}	Y_{t-3}	$\cdots$ $Y_{t-(k+1)}$
…	…	…	…	…
n	Y_{t-n}	$Y_{t-(n+1)}$	$Y_{t-(n+2)}$	$\cdots$ $Y_{t-(k+n)}$

A simple FR-AR model with coefficients $[\emptyset_1^l, \emptyset_1^r]$ and $[\emptyset_2^l, \emptyset_2^r]$ is written as:

$$(Y_t)_T = [\emptyset_1^l, \emptyset_1^r](Y_{t-1})_T + [\emptyset_2^l, \emptyset_2^r](Y_{t-2})_T + [e_t^l, e_t^r], \qquad (14)$$

Both coefficients in Eq. (14) can be derived by following the steps:

Step 1: Provide a real time series data in the fuzzy data format [min, max] per interval time-*t*, such as, per week, per month, etc. For example, Week-1; [3020, 3050], Week-2; [3000, 3057], etc.

Step 2: Divide the fuzzy data into the fuzzy random data [left, center, right] with probabilities. For example, week-1; FRD1 = [3020, 3030, 3040], $Pr_1 = 0.4$ and FRD2 = [3030, 3040, 3050], $Pr_2 = 0.6$.

Step 3: Calculate the expected value (EV) and standard deviation ($Std.Dev$) of fuzzy random data (FRD) in Step 2 where,

$EV = E(Y) =$ (Center of FRD1$\times Pr_1$) + (Center of FRD2 $\times Pr_2$)

$= (3030 \times 0.4) + (3040 \times 0.6)$

$= 3036$

$Var(Y) = E(Y - e)^2$

Standard deviation $(Std.Dev) = s(Y) = \sqrt{Var(Y)} = 7.4$

Step 4: Determine the confidence-interval (CI) of FRD. For example,

Week -1 : $[(EV - Std.Dev), (EV + Std.Dev)] = [3028.6, 3043.4]$

Step 5: Estimate CI for each coefficient model by using linear programming (LP) approach.

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

Subject to

$$\emptyset_i^r \geq \emptyset_i^l$$

$$a_1\emptyset_{11}^l + \left(a_1 + \frac{1}{3}l\right)\emptyset_{12}^l \leq E_1(Y) - Std.Dev_1(Y)$$
$$a_2\emptyset_{21}^l + \left(a_2 + \frac{1}{3}l\right)\emptyset_{22}^l \leq E_2(Y) - Std.Dev_2(Y)$$
$$\vdots$$
$$a_n\emptyset_{n1}^l + \left(a_n + \frac{1}{3}l\right)\emptyset_{n2}^l \leq E_n(Y) - Std.Dev_n(Y)$$

and

$$\left(a_1 + \frac{2}{3}l\right)\emptyset_{11}^r + (b_1)\emptyset_{12}^r \geq E_1(Y) + Std.Dev_1(Y)$$
$$\left(a_2 + \frac{2}{3}l\right)\emptyset_{21}^r + (b_2)\emptyset_{22}^r \geq E_2(Y) + Std.Dev_2(Y)$$
$$\vdots$$
$$\left(a_n + \frac{2}{3}l\right)\emptyset_{n1}^r + (b_n)\emptyset_{n2}^r \geq E_n(Y) + Std.Dev_n(Y)$$

Step 6: From Step 5, define the confidence interval (CI) estimated for each coefficient model.

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

3 Proposed Procedures in Data Preparation

In real situations, non-stationary time series data frequently occur in our daily life, such as, electricity load consumption, stock market prices, temperature, airline passenger data, and others. It can be measured in many different ways, such as, low, high, average, minimum, maximum, closed and open values. In this section, the stock market data input is introduced. Figure 2 highlights the classification of the input as: double input and single input. It is important to emphasize the classification of the data before deciding appropriate procedures for preparing the data towards building an FR-AR model. A low-high procedure (Type-1) and a left-right procedure (Type-2) are proposed for double input and single input, respectively. Both preparation procedures are presented in Sections 3.1 and 3.2.

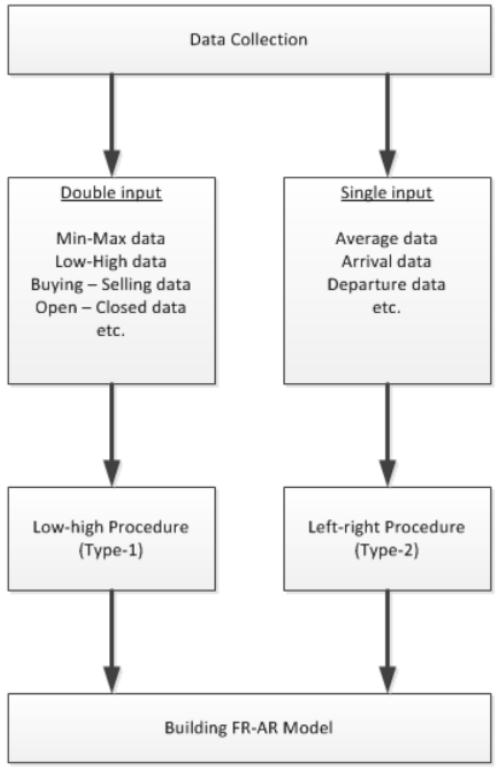


Figure 2. Data classification and data preparation for building FR-AR model

3.1 Low-High Procedure (Type-1)

Generally, the existing models [1, 2, 7, 9, 11, 12, 13, 21, 22, 24, 34, 35, 38, 39] are limited to an application with single point input for building of stock market prediction models such as the ones presented in Figure 3. Additionally, this type of input does not consider variability issue. Apparently, most of the data are obtained from secondary sources. Since the data are obtained from secondary sources, validity, biasness, measurement tools and representation issues or other human errors have to be resolved. Consequently, these limitations contribute to having less accurate forecasting models.

Motivated by limitations of previous input data, use of daily low and high stock market data is called low-high (Type-1) procedure has been considered and suggested for this study. The benefit of using Type-1 procedure is useful to handle variability in the data whereby, a more accurate forecasting model can be achieved as illustrated in Figure 4.

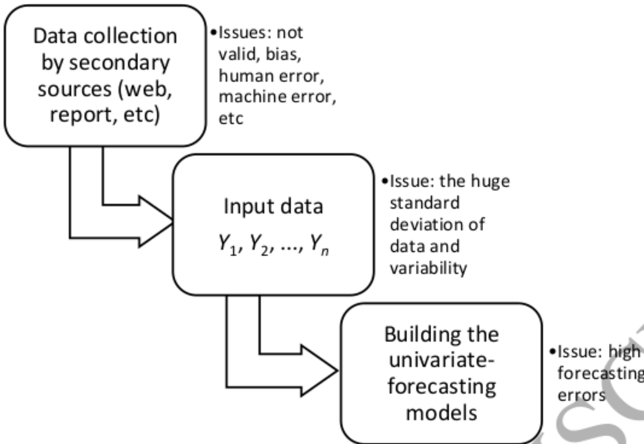


Figure 3. Single input data for building univariate-models

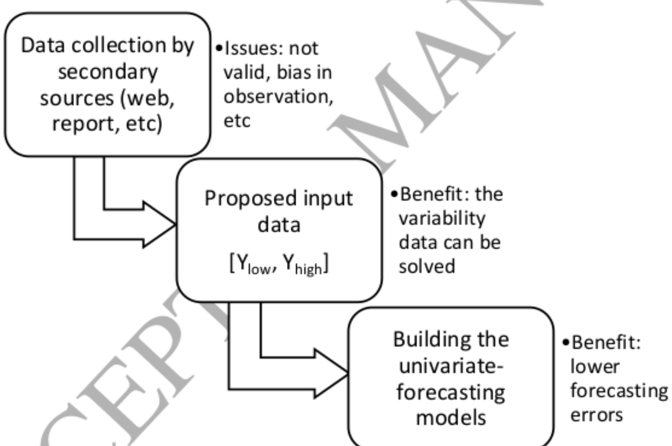


Figure 4. Double input data for building FR-AR model

There are other existing works using the smallest and largest values in the data to build a forecasting model in stock market application [19, 32]. However, the data preparation procedures to obtain fuzzy intervals using the minimum-maximum values from the data have not been extensively. Besides, the ranges of minimum-maximum values obtained are consistent throughout the data. However, the minimum-maximum data presented are lack originality and are not natural. In Type-1 procedure, both values can be obtained using daily low-high data of stock market. Two examples of stock market data are illustrated in Figure 5.

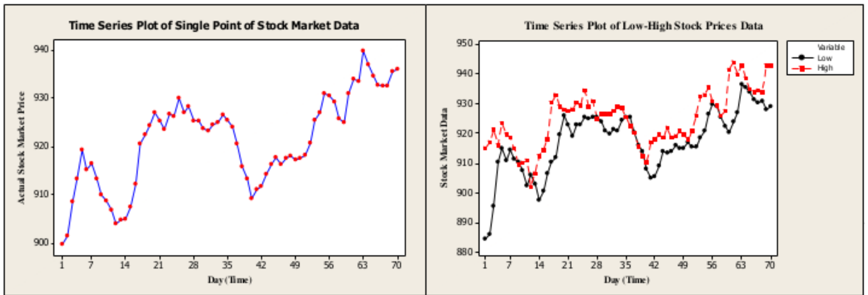


Figure 5. The single point and low-high stock market data

It is observed in Figure 5 that the range of stock market price is not equal for each day (interval time). There are variations in the pattern of stock behavior. The stock behavior pattern is a result of combination between trend and seasonal time series. The fluctuation in the stock market pattern can be observed using low-high procedure. The low-high procedure may allow us to capture the variability in the consumption behavior that leads to better recognition in the pattern. Moreover, the gap between low and high values in the stock market can represent a range of data. By using the values in the range of data, the variability can be determined. Since the low-high values fluctuate every day, thus, result the range values will also fluctuate. Besides the low-high values, fluctuation may be caused by other unpredicted factors. Interestingly, the low-high values in Figure 5 can be transformed into triangular fuzzy number (TFN) directly. The result of the transformation is shown in Figure 6.

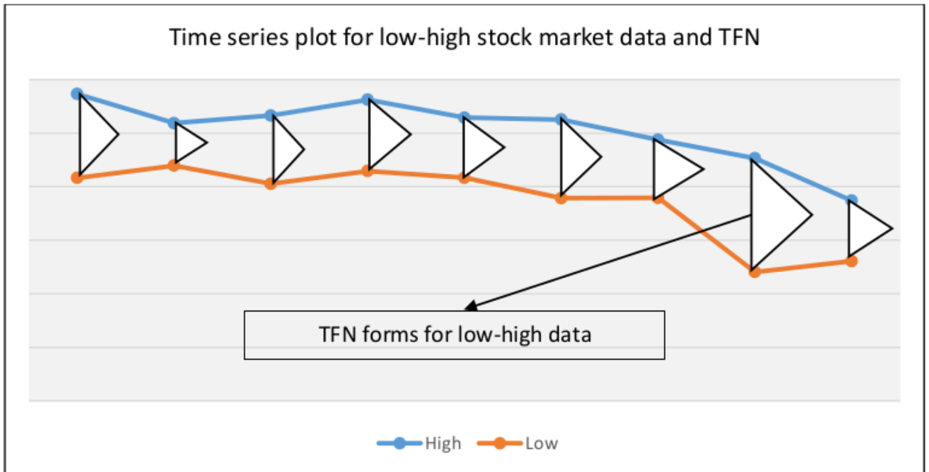


Figure 6. Result of transformation in TFN

Figure 6 shows the min-max data are presented in fuzzy format. Thus, the TFN is obtained directly and naturally without considering left-right spread. The transformation of low-high data to TFN is shown in Figure 7.

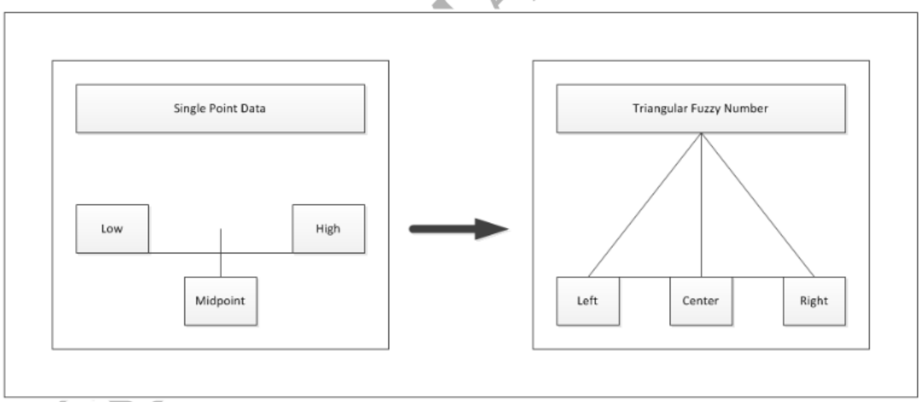


Figure 7. Demonstration of transformation from low-high data into TFN

Based on Figure 7, the equivalence of low-high data with TFN is represented as:

Low data (Y_t^{Low}) ≡ Left spread of TFN (Y_t^l)

Midpoint (Y_t^c) ≡ Center of TFN (Y_t^c)

High data (Y_t^{High}) ≡ Right spread of TFN (Y_t^r)

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

$$[Y_t^{low}, Y_t^{high}] \rightarrow [Y_t^{left}, Y_t^{right}] \rightarrow [Y_t^l, Y_t^c, Y_t^r]_T = \tilde{Y}_t, \quad (15)$$

By using Eq. (15), all the actual low-high load data can be transformed into fuzzy format as presented in Table 2.

Table 2. The low-high prices of stock market and fuzzy data

Day (Time)	Low	High	Fuzzy data (FD)
1	Y_1^l	Y_1^h	$[Y_1^l, Y_1^r]$
…	…	…	…
n	Y_n^l	Y_n^h	$[Y_n^l, Y_n^r]$

The FD data in Table 2 are sufficient to build the desired FR-AR model. In other words, the electricity load consumption data have successfully been prepared using the low-high procedure, referred as Type-1 procedure.

3.2 Symmetry Left-Right Spread Procedure (Type-2)

Motivated by illustration in Section 1, we are really interested to transform the single point data into TFN form using symmetry left-right spread procedure. On the other hands, the concern is to achieve symmetry left-right (LRS) procedure in [2, 18, 19, 32, 43] have not been well discussed. Although, the LRS procedure is commonly used in dealing with single data input, yet the rational of choosing inconsistent left-right spread values.

Let Y_t be time series at $t = 1, 2, \ldots, n$ and let k be possible spread values, $k = 1, 2, \ldots, n$. Y_t can be transformed into TFN format using k values. Note that, although various k values are allowed to adjust the symmetry LRS of TFN, but at any Y_t a consistent k value must be used. As a result, the transformation is written in mathematical form such as in Eq. (16).

$$Y_t \rightarrow [Y_t - k, Y_t, Y_t + k]_T \rightarrow [Y_t^l, Y_t^c, Y_t^r]_T \rightarrow \langle Y_t^c, k \rangle = \tilde{Y}_t, \quad (16)$$

where Y_t^l, Y_t^c, Y_t^r are the results of transformation of Y_t using CLRS procedure. Figure 8 shows an example of Y_t being transformed into TFN at t using symmetry left-right spread $k = 1,2,\dots,n$.

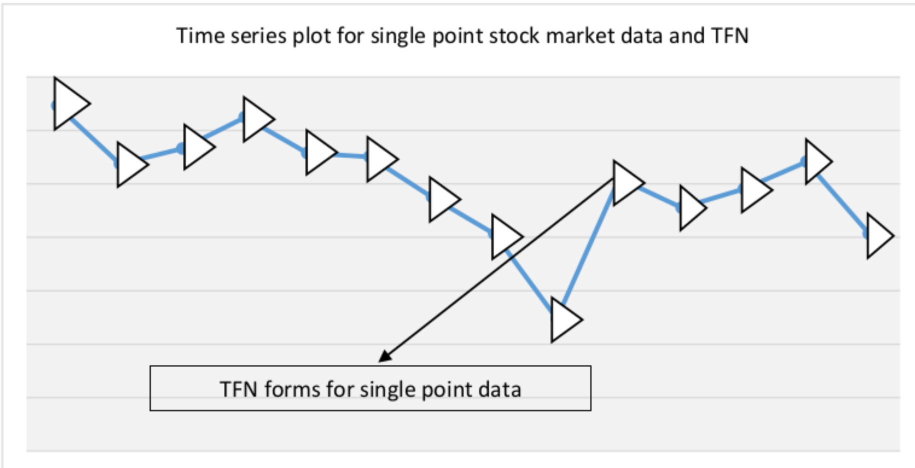


Figure. 8. Transformation of Y_t at t into TFN using LRS procedure

It is important to choose a k value that has the potential to reduce the forecasting error of FR-AR model. Different k values will lead to different TFN shapes. As a result, the TFN shapes will influence the expected value and standard deviation of fuzzy random data (FRD). Essentially, the expected value and standard deviation will influence the parameter estimation FR-AR model. Figure 9 shows the simulation results with four different time series data using $k = 1,2,\dots,10$. Based on data size, there are various initial values of k used. In data set 1, initial value of k is 5, in data set 2, initial value of k is 3, in data set 3, initial value of k is 2 and in data set 4, initial value of k is 1. Note that, for each data set in Figure 9 three k values have been plotted to the highlight the differences in MSE. Apparently, a smaller k is able to produce smaller mean square error (MSE). In other words, there is a significant relationship between spread value of k and MSE.

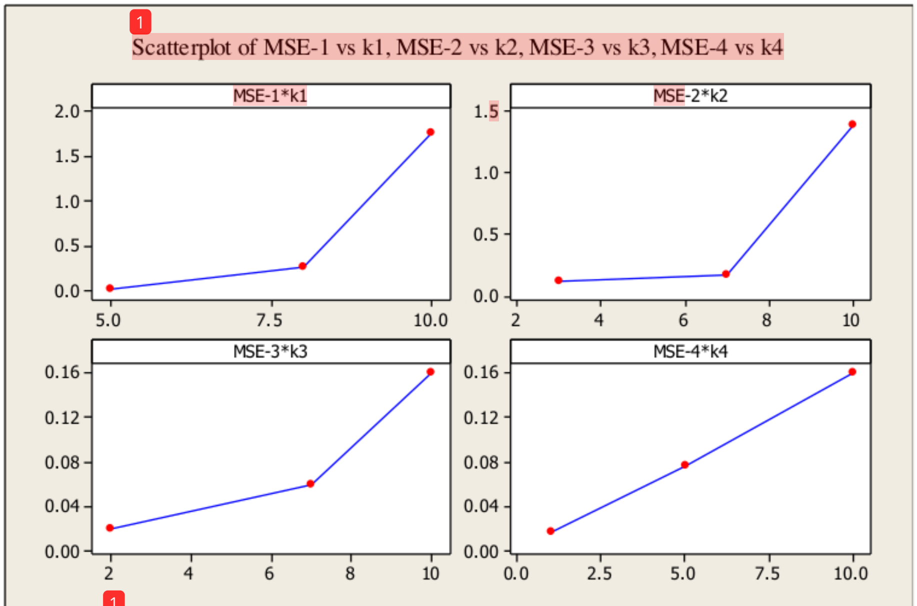


Figure 9. Simulation results to investigate the influence k values on MSE

3.3 Proposed Algorithm for Building FR-AR Model

In this section, the algorithm details in building FR-AR model [21, 32] are presented. The transformation steps for the type of data presented in Sections 3.1 and 3.2 are discussed. Additionally, the parameter estimate using linear programming (LP) approach is provided. The details of the proposed algorithm:

Step 1: Transform the actual time series into TFN based on the procedures in Sections 3.1. and 3.2.

Step 2: Determine fuzzy data (FD) using the TFN results in Step 1 and Table 2 format.

Step 3: Divide FD into two groups of fuzzy random data, FRD_1 and FRD_2. Based on Table 3, the length (l) for each FD can be defined as the difference between upper (b_i) and lower data (a_i), such as $i = 1, 2, 3, \dots, n$. By using value of l, the fuzzy random time series data can be divided into FRD_1 and FRD_2 presented in Table 3.

Table 3. Presentation of TFN results in FD form

Time	Fuzzy data	FRD1, $[Y(t-1)_T]$	FRD2, $[Y(t-2)_T]$
1	$[Y_1^l, Y_1^r]$	$\left[(Y_1^l), \left(Y_1^l + \frac{1}{3}d_1\right), \left(Y_1^l + \frac{2}{3}d_1\right)\right]_T$	$\left[\left(Y_1^l + \frac{1}{3}d_1\right), \left(Y_1^l + \frac{2}{3}d_1\right), (Y_1^r)\right]_T$
2	$[Y_2^l, Y_2^r]$	$\left[(Y_2^l), \left(Y_2^l + \frac{1}{3}d_2\right), \left(Y_2^l + \frac{2}{3}d_2\right)\right]_T$	$\left[\left(Y_2^l + \frac{1}{3}d_2\right), \left(Y_1^l + \frac{2}{3}d_2\right), (Y_2^r)\right]_T$
⋮	⋮	⋮	⋮
n	$[Y_n^l, Y_n^r]$	$\left[(Y_n^l), \left(Y_n^l + \frac{1}{3}d_n\right), \left(Y_1^l + \frac{2}{3}d_n\right)\right]_T$	$\left[\left(Y_n^l + \frac{1}{3}d_n\right), \left(Y_n^l + \frac{2}{3}d_n\right), (Y_n^r)\right]_T$

The TFN of each FRD1 and FRD2 is illustrated in Figure 10.

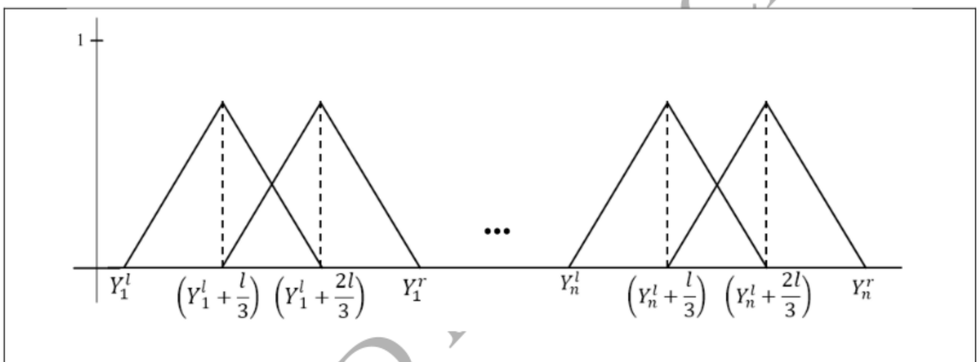


Figure 10. TFN for FRD1 and FRD2

Step 4: Calculate the expected value $E(Y)$ and $Var\ (Y)$ of FRD.

To clarify how $E(Y)$ is obtained, an example from Table 3 at $t = 1$will be used. Let the probabilities, Pr_1 and Pr_2 be equal such as $Pr_1 = Pr_2 = 0.5$ and $Pr_1 + Pr_2 = 1$. Note that, by Definition 3, the values for Pr_1 and Pr_2 may be generated randomly. The formula for $E(Y)$ is written as follows:

$$E(Y) = Pr_1 E(FRD1) + Pr_2 E(FRD2),$$

$$\begin{aligned} E(Y) &= Pr_1\left(Y_1^l + \tfrac{1}{3}d_1\right) + Pr_2\left(Y_1^l + \tfrac{2}{3}d_1\right), \\ &= Pr\left(2Y_1^l + d_1\right), \\ &= \tfrac{1}{2}\left(2Y_1^l + d_1\right), \\ &= Y_1^l + \tfrac{d_1}{2}. \end{aligned} \tag{17}$$

The variance of Y is defined as:

$$Var(Y) = E[(Y-e)^2],$$

where $e = E(Y)$ is given in Eq. (16). To obtain $Var(Y)$, $E[(FRD1-E(Y))^2]$ and $E[(FRD2-E(Y))^2]$ are calculated as follows:

$$(FRD1-E(Y)) = \left(\left(Y_1^l\right), \left(Y_1^l+\frac{1}{3}d_1\right), \left(Y_1^l+\frac{2}{3}d_1\right)\right) - \left(Y_1^l+\frac{d_1}{2}\right),$$

$$= \left(\frac{d_1}{2}, -\frac{d_1}{6}, \frac{d_1}{6}\right)_T,$$

$$(FRD2-E(Y)) = \left(\left(Y_1^l+\frac{1}{3}d_1\right), \left(Y_1^l+\frac{2}{3}d_1\right), \left(Y_1^r\right)\right) - \left(Y_1^l+\frac{d_1}{2}\right),$$

$$= \left(-\frac{d_1}{6}, -\frac{d_1}{6}, \frac{3d_1}{2}\right)_T,$$

Denoting $Y_1 = \left(-\frac{d_1}{2}, -\frac{d_1}{6}, \frac{d_1}{6}\right)_T$ and $Y_2 = \left(-\frac{d_1}{6}, \frac{d_1}{6}, \frac{3d_1}{2}\right)_T$, calculate $\mu_{Y_1^2}$ and $Cr\{Y_1^2 \geq r\}$. Since

$$\mu_{Y_1^2}(t) = Pos\,\{Y_1^2 = t\},$$
$$= max\left\{Pos\{Y_1 = \sqrt{t}\}, Pos\{Y_1 = -\sqrt{t}\}\right\},$$

where $t \geq 0$, we obtain

$$\mu_{Y_1^2}(t) = \begin{cases} \frac{\frac{d_1}{6}+\sqrt{t}}{\frac{l}{3}}, & 0 \leq t \leq \left(\frac{d_1}{6}\right)^2 \\ \frac{\frac{d_1}{2}-\sqrt{t}}{\frac{d_1}{3}}, & \left(\frac{d_1}{6}\right)^2 \leq t \leq \left(\frac{d_1}{2}\right)^2 \\ 0, & otherwise \end{cases}, \qquad (18)$$

Moreover, to compute

$$Cr\{Y_1^2 \geq r\} = \begin{cases} \frac{\left(2-\mu_{Y_1^2}(t)\right)}{2}, & 0 \leq r \leq \left(\frac{d_1}{6}\right)^2 \\ \frac{\left(\mu_{Y_1^2}(t)\right)}{2}, & \left(\frac{d_1}{6}\right)^2 \leq t \leq \left(\frac{d_1}{2}\right)^2 \\ 0, & otherwise \end{cases}. \tag{19}$$

Therefore, from Definition (1), the following equation is obtained:

$$E(Y_1^2) = E\left[\left(\left((Y_1^l),\left(Y_1^l+\frac{1}{3}d_1\right),\left(Y_1^l+\frac{2}{3}d_1\right)\right)-E(Y)\right)^2\right]$$

$$E(Y_1^2) = \int_0^\infty Cr\{Y_1^2 \geq r\}\,dr,$$

$$= \int_0^{(d_1/6)^2} \frac{1}{2}\left(2-\frac{\frac{d_1}{6}+\sqrt{t}}{\frac{d_1}{3}}\right)dt + \int_{(d_1/6)^2}^{(d_1/2)^2} \frac{1}{2}\left(\frac{\frac{d_1}{2}-\sqrt{t}}{\frac{d_1}{3}}\right)dt,$$

Similarly, $E(Y_2^2) = \int_0^\infty Cr\{Y_2^2 \geq r\}\,dr,$

$$= \int_0^{(d_1/6)^2} \frac{1}{2}\left(2-\frac{\frac{d_1}{6}+\sqrt{t}}{\frac{d_1}{3}}\right)dt + \int_{(d_1/6)^2}^{(3d_1/2)^2} \frac{1}{2}\left(\frac{\frac{3d_1}{2}-\sqrt{t}}{\frac{d_1}{3}}\right)dt,$$

Thus,

$$Var(Y) = Pr_1 E(Y_1^2) + Pr_2 E(Y_2^2), \tag{20}$$

Before building FR-AR with parameter estimates, a confidence-interval is induced by $E(Y)$ and $Std.Dev(Y)$ of a fuzzy random variable (FRD). Then the one-sigma confidence $(1 \times \sigma)$ interval (CI) of each FRD [43] is considered. The CI of FRD is expressed as follows:

$$CI \cong [E(Y) - Std.Dev\,(Y), E(Y) + Std.Dev\,(Y)]. \tag{21}$$

Note that, the CI adopted in this step is different from the one used to estimate sample mean. The corresponding CI (output) results for each FRD_1 (input) and FRD_2 (input) are presented in Table 4.

Table 4. CI for each FRD

Time	Output, $[Y(t)]_T$ = CI of FRD	Input	
		$\tilde{Y}_{t-1}$ (FRD$_1$)	$\tilde{Y}_{t-2}$ (FRD$_2$)
1	$[E_1(Y) - Std.Dev_1(Y), E_1(Y) + Std.Dev_1(Y)]$	$\left[(a_1), \left(a_1+\frac{1}{3}l\right), \left(a_1+\frac{2}{3}l\right)\right]$	$\left[\left(a_1+\frac{1}{3}l\right), \left(a_1+\frac{2}{3}l\right), (b_1)\right]$
2	$[E_2(Y) - Std.Dev_2(Y), E_2(Y) + Std.Dev_2(Y)]$	$\left[(a_2), \left(a_2+\frac{1}{3}l\right), \left(a_2+\frac{2}{3}l\right)\right]$	$\left[\left(a_2+\frac{1}{3}l\right), \left(a_2+\frac{2}{3}l\right), (b_2)\right]$
⋮	⋮	⋮	⋮
n	$[E_n(Y) - Std.Dev_n(Y), E_n(Y) + Std.Dev_n(Y)]$	$\left[(a_n), \left(a_n+\frac{1}{3}l\right), \left(a_n+\frac{2}{3}l\right)\right]$	$\left[\left(a_n+\frac{1}{3}l\right), \left(a_n+\frac{2}{3}l\right), (b_n)\right]$

Step 5: The general FR-AR model is expressed as:

$$Y_t = [\widehat{\emptyset}_1^l, \widehat{\emptyset}_1^r]\tilde{Y}_{t-1} + [\widehat{\emptyset}_2^l, \widehat{\emptyset}_2^r]\tilde{Y}_{t-2}, \quad (22)$$

Note that, a linear equation system is to be developed using the input (FRD$_1$, FRD$_2$) and output (FRD) in Table 4. There are two linear equation system to be developed for left-right inputs. Then, both linear equation systems will be solved using linear programing (LP), namely, simplex approach to estimate the parameters of FR-AR model. To use this approach, an objective function and it constraints will be defined as follows:

The ultimate goal of the objective function is to minimize the interval length ($J(\emptyset)$) of parameters.

Objective function:

$$\text{Min } J(\emptyset) = \sum_{i=1}^{n}(\emptyset_{i1}^r - \emptyset_{i1}^l) + (\emptyset_{i2}^r - \emptyset_{i2}^l), \quad (23)$$

Subject to

$$\emptyset_{i1}^r \geq \emptyset_{i1}^l, \emptyset_{i2}^r \geq \emptyset_{i2}^l$$

$$a_1\emptyset_{11}^l + \left(a_1 + \frac{1}{3}l\right)\emptyset_{12}^l \leq E_1(Y) - Std.Dev_1(Y)$$

$$\vdots$$

$$a_n\emptyset_{n1}^l + \left(a_n + \frac{1}{3}l\right)\emptyset_{n2}^l \leq E_n(Y) - Std.Dev_n(Y)$$

and

$$\left(a_1 + \frac{2}{3}l\right)\emptyset_{11}^r + (b_1)\emptyset_{12}^r \geq E_1(Y) + Std.Dev_1(Y)$$

$$\vdots$$

$$\left(a_n + \frac{2}{3}l\right)\emptyset^r_{n1} + (b_n)\emptyset^r_{n2} \geq E_n(Y) + Std.Dev_n(Y)$$

Step 6: Determine the predicted FR-AR model from Step 5.

$$\hat{Y}_t = [\hat{\emptyset}^l_{1,}\hat{\emptyset}^r_1]\tilde{Y}_{t-1} + [\hat{\emptyset}^l_{2,}\hat{\emptyset}^r_2]\tilde{Y}_{t-2}, \qquad (24)$$

which $\hat{Y}_t$ is a predicted fuzzy random time series at time-t, $\hat{\emptyset}^l_{1,}\hat{\emptyset}^r_1$ and $\hat{\emptyset}^l_{2,}\hat{\emptyset}^r_2$ are the pair predicted parameters of fuzzy random time series $(\tilde{Y}_{t-1}, \tilde{Y}_{t-2})$, respectively. The uniqueness of the estimated model parameters is that the left-right values are always the same. Therefore, Eq. (24) is capable of forecasting three different values, low, medium and high which are as follows:

$$\hat{Y}_t = \begin{cases} [\hat{\emptyset}^l_1]\tilde{Y}_{t-1} + [\hat{\emptyset}^l_2]\tilde{Y}_{t-2}, & \text{forecast of low data} \\ [\hat{\emptyset}^r_1]\tilde{Y}_{t-1} + [\hat{\emptyset}^r_2]\tilde{Y}_{t-2}, & \text{forecast of high data} \\ \frac{\hat{Y}^l_t + \hat{Y}^r_t}{2}, & \text{forecast of medium data} \end{cases}, \qquad (25)$$

This is unlike existing models which are limited to forecasting single value only.

Step 7: Evaluate and interpret the min-max-average width of the possibility of model based on the following criteria:

$$W_i = \min, \max\left(\hat{Y}^r_i - \hat{Y}^l_i\right), \qquad (26)$$

$$\bar{W} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}^h_i - \hat{Y}^r_i\right). \qquad (27)$$

Eq. (26) can be used to determine the minimum and maximum width of the model's possibility and the average can be determined using Eq. (27). Essentially, if the average is smaller, then the vagueness in the proposed model should also be smaller. The steps for building FR-AR model are summarized in Figure 11.

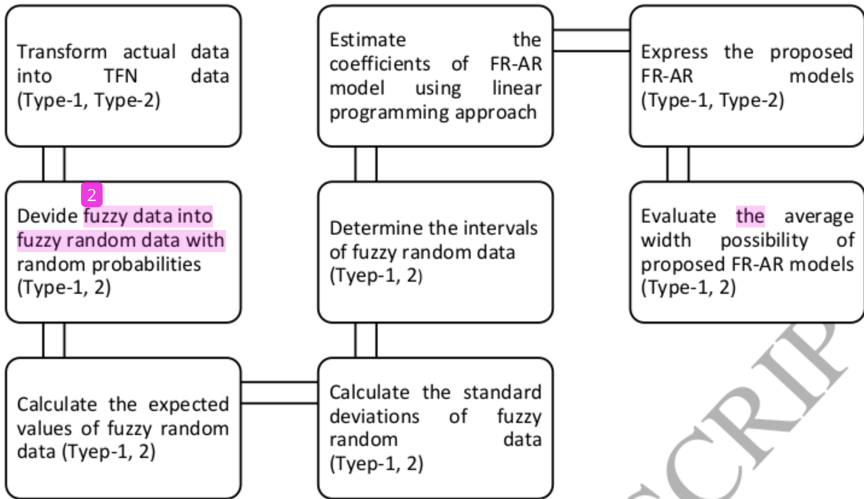


Figure 11. Summary of steps to build the FR-AR model

4. Implementation

The proposed algorithm to build the FR-AR model was applied to real stock market and university enrollment data sets. Note that, the data sets are non-stationary time series and these data may present trend, seasonal, or a combination of both. Such data presentations are found frequently in various problem domains. Thus, a reliable forecasting model is important and needed to improve prediction in the problem domain. Stock prices data of Kuala Lumpur Stock Exchange (KLSE) are used to demonstrate Type-1 procedure and the yearly enrollment of Alabama University as benchmark data are used to demonstrate of Type-2 procedure.

4.1 Implementation of Low-High Procedure (Type-1)

In this section, the daily low-high data of Kuala Lumpur Stock Exchange (KLSE) from 2006-2009 and 2016 were randomly selected for model evaluation. To demonstrate the steps given in Section 3.3, then KLSE 2009 data are used as follows:

Step 1: Transform the daily low-high KLSE data into FD and TFN formats as shown in Table 5.

Table 5. Low-high, FD and TFN data of KLSE 2009

Day	Low-High data	Fuzzy data (FD)	TFN data
1	879.5, 901.0	[879.5, 901.0]	[879.5, 890.3, 901.0]
2	903.5, 928.0	[903.5, 928.0]	[903.5, 915.3, 928.0]
3	916.5, 934.5	[916.5, 934.5]	[916.5, 925.3, 934.5]
…	...		…
145	1165.5, 1177.0	[1165.5, 1177]	[1165.5, 1171.3, 1177]

Step 2: Divide fuzzy data (FD) in Step 1 into FRD1 and FRD2 with random probabilities as shown in Table 6.

Table 6. FRD1 and FRD2 of KLSE 2009

Day	FRD-1	Pr-1	FRD-2	Pr-2
1	[879.5, 886.66, 893.83]	0.4	[886.66, 893.83, 901.0]	0.6
2	[903.5, 911.66, 919.83]	0.2	[911.66, 919.83, 928.0]	0.8
…	…	…	…	…
145	[1165.5, 1169.33, 1173.16]	0.1	[1169.33, 1173.16, 1177.0]	0.9

Step 3: Calculate expected value (EV) and standard deviation ($Std.Dev$) of FRD.

For example,

$EV_1 = E(FRD1) = 886.66,$

$EV_2 = E(FRD2) = 893.83,$

$EV = E(FRD) = Pr_1.E(FRD1) + Pr_2.E(FRD2) = 888.81.$

The other EV and $Std.Dev$ of FRD are shown in Table 7.

Table 7. EV and $Std.Dev$ of FRD of KLSE 2009

Day	EV	$Std.Dev$
1	888.81	4.6
2	915.75	5.7
…	…	…
145	1170.48	2.4

Step 4: Determining CI of FRD is presented in Table 8.

Table 8. CI of FRD

Day	Confidence Intervals
1	[884.2, 893.4]
2	[910.1, 921.4]
…	…
145	[1168.0, 1172.9]

Step 5: Estimate coefficients of FR-AR model using linear programming approach (LP), namely, simplex method. In this case, 80% of actual low-high data have been used to estimate of coefficients FR-AR model. The rest of data were used as model validation (testing data).

Min $=\sum\left(\hat{\emptyset}_{i1}^{r}-\hat{\emptyset}_{i1}^{l}\right)+\left(\hat{\emptyset}_{i2}^{r}-\hat{\emptyset}_{i2}^{l}\right)$, $\hat{\emptyset}_{i1}^{r}\geq\hat{\emptyset}_{i1}^{l}$, $\hat{\emptyset}_{i2}^{r}\geq\hat{\emptyset}_{i2}^{l}$.

Subject to

Inequalities of Left-LP:

$879.5\,\hat{\emptyset}_{11}^{l}+886.66\,\hat{\emptyset}_{12}^{l}\leq 884.2$

$903.5\,\hat{\emptyset}_{21}^{l}+911.66\,\hat{\emptyset}_{22}^{l}\leq 910.1$

… … …

$1052.5\,\hat{\emptyset}_{1161}^{l}+1060\,\hat{\emptyset}_{1162}^{l}\leq 1168.0$

Inequalities of Right-LP:

$893.83\,\hat{\emptyset}_{11}^{r}+901.0\,\hat{\emptyset}_{12}^{r}\leq 893.4$

$919.83\,\hat{\emptyset}_{21}^{r}+928.0\,\hat{\emptyset}_{22}^{r}\leq 921.4$

… … …

$1067.5\,\hat{\emptyset}_{1161}^{r}+1075.0\,\hat{\emptyset}_{1162}^{r}\leq 1172.9$

$\hat{\emptyset}_{i1}^{l}\geq 0$, $\hat{\emptyset}_{i1}^{r}\geq 0$, $\hat{\emptyset}_{i2}^{l}\geq 0$, $\hat{\emptyset}_{i2}^{r}\geq 0$

Step 6: Based on the simplex approach, the estimated parameters and predicted FR-AR model obtained are as follows:

$\widehat{\emptyset}_1 = \widehat{\emptyset}_{i1}^l = \widehat{\emptyset}_{i1}^r = 0.4026755$

$\widehat{\emptyset}_2 = \widehat{\emptyset}_{i2}^l = \widehat{\emptyset}_{i2}^r = 0.5972301$

$$\tilde{Y}_t = 0.4026755\,\tilde{Y}_{t-1} + 0.5972301\,\tilde{Y}_{t-2} \quad . \quad (28)$$

Step 7: Based Equations (26) and (27), evaluate width and ambiguity of proposed FR-AR model for KLSE data as follows:

Width minimum of FR-AR model : 0.666.

Width maximum of FR-AR model : 23.333.

Average width of FR-AR model : 9.907.

Eq. (28) is the proposed FR-AR model for KLSE 2009. Moreover, this is used to estimate the training data from day-3 to day-116. For example, the forecasted values for low-midpoint-high stock market at day-2 is based on day-1 and so on. By using FRD data in Table 6, the predicted $\hat{Y}_2^{low}$, $\hat{Y}_2^{mid}$, and $\hat{Y}_2^{high}$ are follows:

$$\text{FRD1; } (Y_{t-1})_T = [879.50, 886.66, 893.83],$$

$$\text{FRD2; } (Y_{t-2})_T = [886.66, 893.83, 901.00],$$

Based on Eq. (18), the following values are derived:

$$\hat{Y}_2^{low} = 0.4026755(886.66) + 0.5972301(879.50) = 885.15,$$

$$\hat{Y}_2^{mid} = 0.4026755(893.83) + 0.5972301(886.66) = 889.06,$$

$$\hat{Y}_2^{high} = 0.4026755(901.00) + 0.5972301(893.83) = 896.33.$$

The forecasted values of low-mid-high (For-Low-Mid-High) stock are 885.15, 889.06, and 896.33. By following the example at day-2, the results of training (day-3 to day-116) and testing data (day-117 to day-145) are presented in Tables 9-11.

Table 9. Actual low-high of KLSE and training data

Day	Low	High	For-Low	For-High
1	879.50	901.00	*	*
2	903.50	928.00	885.15	896.33
3	916.50	934.50	919.99	930.99
4	920.00	942.50	924.39	939.39
5	910.00	920.50	912.00	917.00
⋮	⋮	⋮	⋮	⋮
115	1053.00	1064.00	1051.09	1060.42
116	1043.50	1064.00	1047.48	1059.15
MSE			109.75	95.06

Table 10. Actual-midpoint, training and testing data of KLSE

Day	Midpoint	M-Training	Day	Midpoint	M-Testing
1	890.25	*	1	1039.25	1036.69
2	915.75	916.46	2	1052.50	1051.98
3	925.50	923.99	3	1067.25	1062.78
4	931.25	929.89	4	1076.75	1074.97
5	915.25	913.50	5	1079.00	1080.16
⋮	⋮	⋮	⋮		⋮
115	1058.50	1054.75	28	1164.25	1167.25
116	1053.75	1050.31	29	1171.25	1168.41
	MSE	45.1875		MSE	47.2054

Table 11. Actual low-high of KLSE and testing data

Day	Low	High	For-Low	For-High
1	1031.00	1047.50	*	*
2	1043.50	1061.50	1041.99	1058.98
3	1057.50	1077.00	1061.28	1072.28
4	1072.00	1081.50	1070.79	1083.12
5	1075.00	1083.00	1072.49	1081.82
⋮	⋮	⋮	⋮	⋮
28	1158.00	1170.50	1160.38	1166.71
29	1165.50	1177.00	1162.68	1175.36
MSE			70.78	75.65

Moreover, the fitting between actual low-mid-high of KLSE data and their forecasted values are illustrated in Figures 12-14.

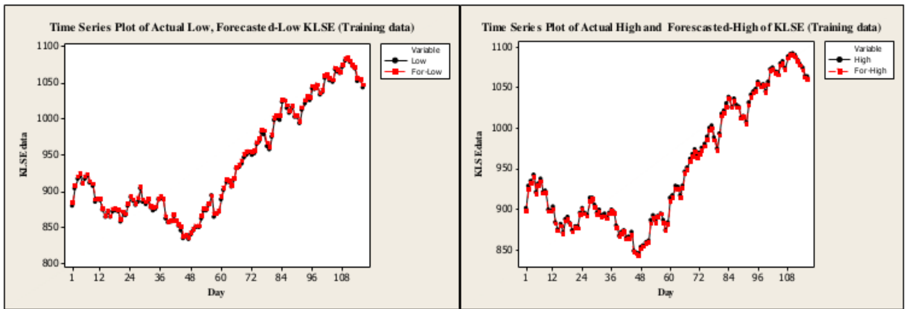


Figure 12. Fitting actual low-high and training of KLSE 2009

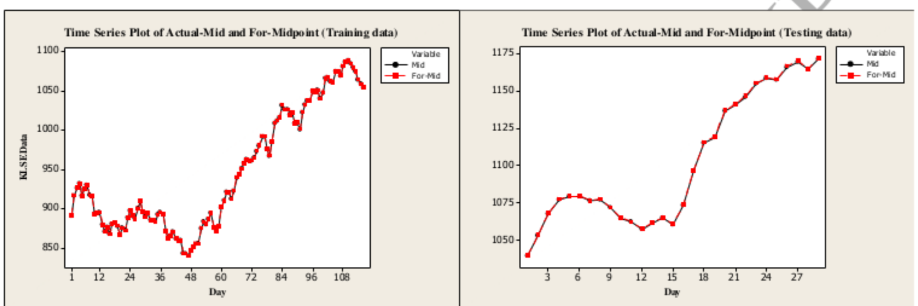


Figure 13. Fitting actual midpoint, training and testing of KLSE 2009

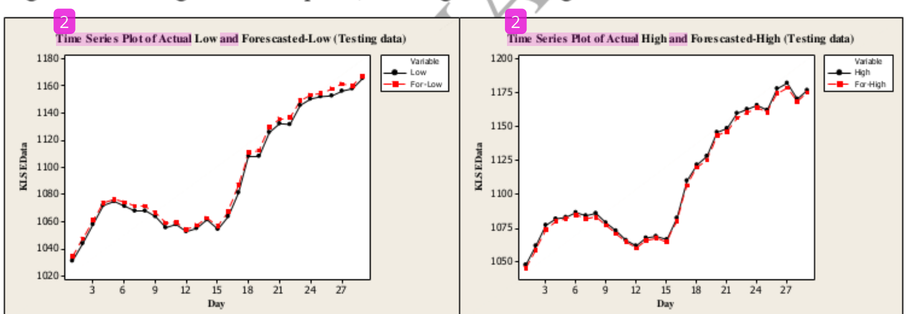


Figure 14. Fitting actual low-high and testing of KLSE 2009

Figures 12-14 show the fitting of low-midpoint-high of actual KLSE 2009 with their forecast values. Apparently, the training data derived from proposed FR-AR model (Eq. 18) have a very close majority of the actual KLSE 2009. As the majority of the data have been captured, therefore, both graphs that are the actual and FR-AR models are very similar. The graph for testing data is the most similar to the graph of actual data. The rest of the models are not capable of producing graphs as good as the FR-AR model. The graph of the training data is close to the actual data as well as, the graph for the testing data. Such finding indicates that proposed FR-AR model using Type-1 procedure outperforms from the other models. Based on due to the performance of the FR-AR model, this model was used to forecast three different values simultaneously that are: low, midpoint and high of KLSE prices.

In a comparison study, the evaluation of mean square error (MSE) was considered using models conventional ones ranging from until soft computing models. Recently, other forecasting algorithms have been employed in some excellent works, such as the multivariate statistical combination forecasting method for product quality evaluation [43]. However, the baseline models are namely the univariate forecasting models, such as, ARIMA, GARCH, single exponential smoothing (SES), double exponential smoothing (DES), fuzzy time series (FTS), exponential fuzzy time series (E-FTS), support vector machine (SVM), fuzzy metagraph (FM), fuzzy auto regression (F-AR) and existing FR-AR models in MSE. Due to the limitation of these univariate models, daily midpoint (average) of KLSE 2009 data were only used for comparison as presented in Table 13.

Table 13. Evaluation of MSE of KLSE 2009

Model	MSE	
	Training data (In-sample forecast)	Testing data (out-sample forecast)
ARIMA [4]	138.10	355.97
GARCH [4]	132.24	287.56
SES [5]	274.77	286.86
DES [5]	287.62	299.45
SVM [1]	124.89	221.78
FTS [20]	74.29	91.90
E-FTS [36]	70.77	145.34
FM [1]	72.67	130.97
F-AR [32]	67.02	88.00
FR-AR [32]	62.26	75.87
FR-AR [19]	50.19	50.21
Proposed FR-AR	45.18**	47.20**

**: smallest MSE

In Table 13, KLSE 2009 is divided into training and testing data. MSEs of the proposed FR-AR model are smaller than the others, training and testing. The MSE values of the proposed model imply that the forecasting error can be reduced significantly. The FR-AR model benefits from the use of low-high data in Type-1 procedure becuase the model parameters obtained are more accurate. Consecutively, the variability and volatility of daily stock market data can be handled using this procedure. The forecasted values tend to be near to the medium data, thus indicating the accuracy has improved. On the other hand, the existing univariate models have bigger MSE values in comparison with the proposed model, because the low-high of stock market data were not implemented in the building of these models. Moreover, only some KLSE data sets using the proposed model from 2006, 2007, 2008 and 2016 were evaluated and presented in Table 14.

Table 14. MSE comparisons of KLSE from 2006 - 2008, and 2016

Model	MSE (KLSE 2006)		MSE (KLSE 2007)		MSE (KLSE 2008)		MSE (KLSE 2016)	
	Training	Testing	Training	Training	Training	Testing	Training	Training
ARIMA [4]	140.932	234.565	137.643	311.899	136.785	139.533	134.634	137.521
GARCH [4]	138.671	210.598	132.765	300.433	131.111	137.231	132.892	135.214
SES [5]	99.190	101.120	274.73	278.43	354.577	360.344	116.324	129.869
DES [5]	98.012	101.101	287.62	290.64	388.866	392.112	115.734	126.553
SVM [1]	129.975	196.658	130.589	287.764	128.451	147.097	140.239	147.198
FTS [20]	118.978	122.987	122.632	198.823	124.776	129.761	135.213	138.835
E-FTS [36]	115.211	132.245	120.543	207.453	134.876	139.675	140.671	147.642
FM [1]	120.764	130.976	129.754	200.634	133.532	137.343	138.832	142.443
F-AR [32]	115.108	119.765	119.452	180.765	125.654	129.107	130.872	139.211
FR-AR [32]	111.734	117.908	115.723	163.865	122.534	125.635	123.223	133.334
FR-AR [19]	105.233	113.551	112.003	143.229	120.111	121.197	120.564	130.113
Proposed FR-AR	95.02**	100.06**	100.49**	120.26**	115.17**	118.14**	114.02**	125.12**

**: smallest MSE

Table 14 shows the MSE comparisons between the proposed model and other existing models. In these comparisons, the daily average of KLSE data from 2006, 2007, 2008 and 2016 were studied. Conventional time series models, ARIMA and GARCH were selected for use in this comparison, because they are frequently implemented in the stock market forecasting models. Moreover, other soft computing models were also considered in error evaluation. From this table, the proposed FR-AR model with Type-1 procedure is able to reduce the forecasting error significantly for the four year data sets if compared with the existing time series and soft computing models. From our perspective, the merit of the proposed procedure can also produce the parameters model in intervals form. These intervals can support the daily range of values that are approximately the expected value of stock market data likely to contain the estimation target. Thus, the forecasting accuracy is achieved significantly if compared with the existing non-fuzzy random models.

4.2 Implementation of Left-Right Procedure (Type-2)

In this section, the implementation of Type-2 procedure to forecast the yearly enrollment of Alabama University from 1972 to 1992 is presented. This data set is the benchmark and frequently used in fuzzy time series forecasting. By following the steps given in Section 3, the proposed FR-AR model of enrolment is written as:

$$\hat{Y}_t = [0.94, 1.076](Y_{t-1})_T. \tag{19}$$

Based on Eq. (14), same parameter estimates for the left and right values ($\hat{\emptyset}_1^l = 0.94$ and $\hat{\emptyset}_1^r = 1.076$) for each model were obtained. Then, the root of mean square error (RMSE) of the models were compared with the existing fuzzy time series (FTS) models in Table 15.

Table 15. RMSE comparison for Alabama enrollment forecasting

Model	RMSE	Rank
FTS-Yolcu *et al*. [3]	805.1	14
FTS-Song &Chissom [34]	650.4	13
FTS-Chen [12]	638.3	12
SES [5]	526.2	11
FTS-Qiu *et al*. [3]	511.3	10
DES [5]	507.8	9
FTS-Lee & Chou [3]	501.2	8
FTS-Huarng [3]	476.9	7
FTS-Cheng *et al*. [3]	478.4	6
FTS-Joshi & Kumar [3]	433.7	5
FTS-Ismail *et al*. [24]	400.2	4
FTS-Kumar & Gangwar [3]	493.5	3
FTS-Bisht & Kumar [3]	428.5	2
Proposed FR-AR	149.2**	1

Due to the limitation of FTS models, only average values for comparison were used. Bearing in mind that the data have been divided into training and testing data. The RMSE for the FR-AR models are smaller than FTS models, the RMSE values of proposed model imply that the forecasting error can be reduced significantly. In other words, the FR-AR model benefits from the use of left-right values in the proposed procedure because the estimated parameters can reduce the randomness, vagueness and possibility of data. Thus, the variation will be minimised too and controlled. As a result, the predicted values tend to be near to the medium data and the accuracy is improved. Moreover, the actual enrollment and the forecasted values derived by the proposed FR-AR and other FTS models are illustrated in Figure 15.

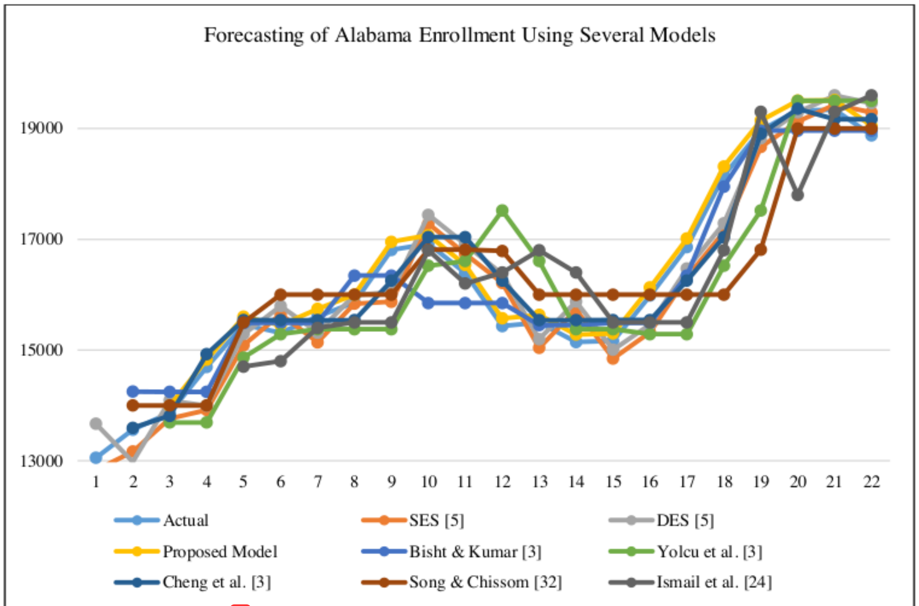


Figure 15. Fitting actual enrollment and forecasted values using some models

Based on Sections 4.1 and 4.2, the advantages of the proposed procedures, Type-1 and Type-2 in improving the performance of FR-AR model on various items are presented in Table 16.

Table 16. The advantages of Type-1 and Type-2 procedures

Item	Type-1	Type-2
Input data	Appropriate for double inputs, such as, min-max values, low-high values, and others.	Appropriate for single input, such as, mean, average, sum, and others.
TFN	Adjustment is unnecessary since data are presented in fuzzy form (low-high interval).	Adjustment is necessary for left-right spread of TFN.
Parameter estimate	Significant parameters can be obtained if the average range between min-max data is small.	Significant parameters can be obtained if smaller spread is considered.
Forecasting model	Appropriate for determining low, medium and high values simultaneously.	Appropriate for determining medium or central value (single value) only.
Vagueness model	Can be reduced if there is a small gap between low-high data.	Can be reduced if the spread is small enough.
Effectiveness	Appropriate for multiple observations with a small gap (range) between minimum and maximum values in a day, a week, etc. For example, stock market, temperature, exchange rate, electricity load consumption data, and others.	Appropriate for limited observations. In this case, the measurement total or average observations are main intention of researchers. For example, enrollment, tourism, arrival-departure airlines passengers data, and others.

5. Conclusion

In this paper, two type procedures in handling the single point data (single record) issues, namely, low-high procedure (Type-1) and left-right spread procedure (Type-2) were discussed. Both procedures are subjected to solve the biasness of the human and machine errors during data collecting, especially, multiple observations data such as, the stock market data. Therefore, the forecasting accuracy of FR-AR model can be improved significantly.

The FR-AR model was implemented to forecast real stock market data sets. In the case of data with a small gap between low and high values, Type-1 procedure is more appropriate. Essentially, the smaller width of the possibility indicates that the model is obtained naturally with Type-1 procedure. Consequently, the vagueness model can be lessened. On the other hand, Type-2 procedure is more appropriate when a large gap exists between the low and high data. In this procedure, a smaller vagueness model can be achieved if a small k value is considered. Mathematically, the better parameters of FR-AR model can be obtained if the gap between minimum and maximum data is not too large.

In the algorithm to build FR-AR model, we used fuzzy random data (input-output data) to develop linear equation systems. The linear equation systems were then solved using linear programming (LP) approach, namely, simplex approach to estimate the parameters of the model. Since the parameter values obtained are the same for left and right, thus, determining three predicted values (low-medium-high) simultaneously is possible.

In the comparison of MSE, non-fuzzy and fuzzy models, namely, SES, DES, ARIMA and FTS have been considered. Based on the applications, the proposed FR-AR model with two different input types (Type-1 and Type-2) outperforms other existing models for both inputs. Based on the finding, this study can be implemented in handling the non-stationary time series data from various domains.

Acknowledgment

This paper work was supported by the Research, Innovation, Commercialization, and Consultancy Management Office (ORICC) under Post-Doctoral Grant Vot. No. D003 at Universiti Tun Hussein Onn Malaysia (UTHM).

References

[1] T. Anbalagan and S. U. Maheswari, Classification and prediction of stock market index based on fuzzy metagraph, *Procedia Computer Science*, 47 (2015) 214-221.

[2] N. Arbaiy and J. Watada. Linear fractional programming for fuzzy random based possibilistic programming problem, *Int. J. Simulation Systems Science and Tech*nology, 14 (2) (2014) 24 – 30.

[3] K. Bisht and S. Kumar, Fuzzy time series forecasting method based on hesitant fuzzy sets, *Expert Systems with Applications*, 64 (2016) 577 – 568.

[4] B. L. Bowerman and R. T. O'Connell, *Time series forecasting, unified concepts and computer implementation*. (1987). Boston: Duxbury Press.

[5] P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting*. Second edition. (2002). New York: Springer.

[6] Y-H. O. Chang, Hybrid fuzzy least-squares regression analysis and its reliably measures, *Fuzzy Sets Syst*ems, 119 (2) (2001) 225 – 246.

[7] M. Y. Chen and B. T. Chen. A hybrid fuzzy time series model based on granular computing for stock price forecasting, *Information Sciences*, 294 (2015) 227-241.

[8] M. Y. Chen and B-T. Chen, A hybrid fuzzy time series model based on granular computing for stock price forecasting, *Information Sciences*, 294 (2015) 227 – 241.

[9] S. M. Chen and P.Y. Kao. TAIEX forecasting based on fuzzy time series, particle swarm optimization technique and support vector machines, *Information Sciences*, 247 (2013) 62-71.

[10] Y. Chen and Y. Hao, A feature weighted support vector and K-nearest algorithm for stock market indices prediction, *Expert Systems with Applications*, 80 (2017) 340 – 355.

[11] S. M. Chen, H. P. Chu and T.W. Sheu. TAIEX forecasting using fuzzy time series and automatically generated weight of multiple factors, *IEEE Tran. Systems, Man Cybernetics, Part A Syst. Hum*. 42 (2012) 1485-1495.

[12] S. M. Chen. Forecasting Enrollments based on Fuzzy Time Series. *Fuzzy Sets Systems*, 81 (1996) 311-319.

[13] S. M. Chen and W. S. Jian, Fuzzy forecasting based on two-factor second order fuzzy-trend logical relationship groups, similarity measures and PSO techniques, *Information Sciences*, 391-392 (2017) 65 – 79.

[14] T. L. Chen and F. Y. Chen, An intelligent pattern recognition model for supporting investment decisions in stock market, *Information Sciences*, 346-347 (2016) 261 – 274.

[15] S. H. Cheng, S. H. Chen and W-S. Jian, Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures, *Information Sciences*, 327 (2016) 272-287.

[16] E. Chong, C. Han, and F. C. Park, Deep learning networks for stock market analysis and prediction, Methodology, Data Representations and Case Studies,

Expert Systems with Applications, 83 (2017) 187-205.

[17] R. Efendi, M. M. Deris and Z. Ismail, Implementation of fuzzy time series in forecasting of the non-stationary data, *Int. Journal Computation Intelligence and Applications*, 15 (2016) 50-60.

[18] R. Efendi, N. Arbaiy and M. M. Deris, Estimation of confidence-interval for yearly electricity load consumption based on fuzzy random auto-regression model, S. P. Amnuaisuk *et al.*, (Eds): *Advances Intelligence Systems Computational*, 1 CIIS (2017) 155-164.

[19] R. Efendi, N. Arbaiy and M. M. Deris, Indonesian-Malaysian stock models using fuzzy random auto-regression, *Proceeding Seminar Nasional Teknologi Komunikasi dan Industri*, (2017) 604-611.

[20] R. Efendi, Z. Ismail and M. M. Deris. A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand, *Applied Soft Computing*, 28 (2015) 422-430.

[21] R. Efendi, Z. Ismail and M. M. Deris. Improved weight fuzzy time series used in the exchange rates forecasting of US Dollar to Ringgit Malaysia, *Int. Journal of Computational Intelligence and Applications*, 12 (2013) 71-90.

[22] C. Evans, K. Pappas, P. Xhafa, Utilizing artificial neural networks and genetic algorithms to build an algo-trading model intra-day foreign exchange speculation, *Mathematical and Computer Modelling*, 58 (2013) 1249-1266.

[23] Z. Ismail, R. Efendi and M. M. Deris. Application of fuzzy time series approach in electric load forecasting, *New Mathematics and Natural Computation*, 3 (2015) 229-248.

[24] Z. Ismail, R. Efendi and M.M. Deris. Inter-quartile range approach to length-interval adjustment of enrolment data in fuzzy time series forecasting. *Int. J. Computational Intelligence and Applications*. 3 (2013) 40-50.

[25] J. Jia, A. Zhou, and S. Guan, Forecasting based on high-order fuzzy fluctuation trends and particle swarm optimisation machine learning, *Symetry*, 9 (2017) 1-15.

[26] N. Kashpruk, Comparison idenification of statistical and fuzzy models of time series, *IEEE Int. Conf. Systems, Man and Cybernatics,* (2016) 1935 - 1938 Budapest, Hungary.

[27] H. L. Lee, A. Liu and W. S. Chen. Pattern discovery of fuzzy time series for financial prediction, *IEEE Transaction on Systems, Man, and Cybernetics-Part B: Cybernetics*, 18 (2006) 613-625.

[28] Y. H. Leu, C. P. Lee, Y. Z. Jou. A distance-based fuzzy time series model for exchange rate forecasting, *Expert Systems with Applications*, 36 (2009) 8107-8114.

[29] Y. K. Liu and B. Liu. Fuzzy random variable: A scalar expected value operator, *Fuzzy Optimisation Decision Making*, 2 (2003) 143 – 160.

[30] B. Liu and Y. K. Liu. Expected value of fuzzy variable and fuzzy expected value models, *IEEE Trans. Fuzzy Systems*, 10 (2002) 445 – 450.

[31] S. Nahmias. Fuzzy variables, *Fuzzy Sets Systems*, 1 (1978) 97 – 111.

[32] L. Shao, Y.-H. Tsai, J. Watada and S. Wang. Building fuzzy random auto-regression model and its application, J. Watada *et al.*, (Eds): *Intelligence Decision Technology* 1 SIST, (15) (2012) 155 – 164.

[33] A. K. Shyamal and M . Pal, Triangular fuzzy matrices, *Iranian Journal of Fuzzy Systems*, 4 (2007) 75 – 87.

[34] Q. Song and B. S. Chissom. Forecasting Enrollments with Fuzzy Time Series – Part II. *Fuzzy Sets Systems*, 62 (1994) 1-8.

[35] B. Q. Sun, H. Guo, H. R. Karimi, Y. Ge and S. Xiong, Prediction of stock market futures based on fuzzy sets and multivariate fuzzy time series, *Neurocomputing* 151 (2015) 1528 – 1536.

[36] F. M. Talarposhti, H. J. Sadaei, R. Enayatifar, F. G. Guimares, M. Mahmud, T. Eslami, Stock market forecasting using a hybrid model of exponential fuzzy time series, *International Journal of Approximate Reasoning*, 70 (2016) 79 – 98.

[37] H. Tanaka, S. Uejima and K. Asai, Linear regression analysis with fuzzy model, *IEEE Trans. Systems, Man, Cybernetics*, SMC-12 (2) (2007) 89 – 106.

[38] J. L. Ticknor, A Bayesian regularized artificial neural network for stock market forecasting, *Expert Systems with Applications*, 40 (2013), 5501-5506.

[39] B. J. Vanstone and G. Finnie, Am empirical methodology for developing stock market trading systems using neural networks, *Expert Systems with Applications*, 36 (2005) 6668-6680.

[40] H. Vata Arani and S. A. Torabi, Integrated material-financial supply chain master planning under mixed uncertainty, *Information Sciences*, 423 (2018) 96 – 114.

[41] B. Wang, Y. Li and J. Watada, Multi-period portfolio selection with dynamic risk/expected return level under fuzzy random uncertainty, *Information Sciences*, 385-386 (2017) 1 – 18.

[42] J. Watada, Fuzzy time-series analysis and forecasting of sales volume, *Fuzzy Regression Analysis*, (1992) 211 -227.

[43] J. Watada, S. Wang and W. Pedrycz. Building confidence-interval based fuzzy random regression models, *IEEE Trans. Fuzzy Syst*ems, 17 (6) (2009) 1273 – 1283.

[44] L. Y. Wei, C.H. Cheng and H.H. Wu. A hybrid ANFIS based on-period moving average model to forecast TAIEX stock, *Applied Soft Computing*, 19 (2014) 86-92.

[45] B. Weng, M. A. Ahmed and F. M. Megahed, Stock market one-day ahead movement prediction using disparate data sources, *Expert Systems with Applications*, 79 (2017) 153 – 163.

[46] F. Ye, L. Zhang, D. Zhang, H. Fujita, Z. Gong, A novel forecasting method based on multi-order fuzzy time series and technical analysis, *Information Sciences*, 367-368 (2016) 41 – 57.

[47] S. Yin, L. Liu and J. Hou, A multivariate statistical combination forecasting method for product quality evaluation, *Information Sciences*, 355-256 (2016) 229-236.

[48] B. Zahiri, P. Jula and R. T. Moghaddam, Design of a pharmaceutical supply chain network under uncertainty considering perishability and substitutability, *Information Sciences*, 423 (2018) 257 – 283.

[49] X. Zhong and D. Enke, A comprehensive cluster and classification mining procedure for daily stock market return forecasting, *Neorocomputing*, 267 (2017) 152 – 168.

[50] X. Zhong and D. Enke, Forecasting daily stock market return using dimensionality reduction, *Expert Systems with Applications*, 67 (2017) 126 – 139.

[51] W. Zhou and Z. Xu, Group consistency and group decision making under uncertain probabilistic hesitant fuzzy preference environment, *Information Sciences*, 414 (2017) 276 – 288.

orecasting_based_on_fuzzy_random_auto-regression_time_series.pdf

ORIGINALITY REPORT

20%	%	20%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Riswan Efendi, Noor Azah Samsudin, Nureize Arbaiy, Mustafa Mat Deris. "Fuzzy random auto-regression time series model in enrollment university forecasting", 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), 2017 Publication	14%
2	"Computational Intelligence in Information Systems", Springer Nature, 2017 Publication	4%
3	Barrón Díaz Arturo Joaquín. "Petrogénesis y metamorfismo del magmatismo cámbrico de la formación Cerro Rajón, Sonora, México y su relación con el proceso de rifting en el margen meriodional de Laurentia", TESIUNAM, 2019 Publication	2%

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On