Journal of Theoretical and Applied Information Technology
15th November 2018. Vol.96. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HIGH-ORDER RTV-FUZZY TIME SERIES FORECASTING MODEL BASED ON TREND VARIATION

[1]NOOR RASIDAH ALI, [2]KU RUHANA KU-MAHAMUD

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Kedah, Malaysia
[2]School of Computing, Universiti Utara Malaysia, Malaysia

E-mail: [1]norrash@kedah.uitm.edu.my, [2]ruhana@uum.edu.my

ABSTRACT

Time series data principally involves four major components which are trend, cyclical, seasonal and irregular, that reflects the characteristics of the data. Ignoring the systematic analysis of patterns from time series components will affect forecasting accuracy. Thus, this paper proposes a high-order ratio trend variation (RTV) fuzzy time series model based on the trend pattern and variations in time series to deal with patterns within the time series data. RTV is used in the fuzzification process to deal with data that contains vagueness, uncertainty and impreciseness. Proper adjustment was also applied to handle the common issues in fuzzy time series model includes determination of length of interval, fuzzy logic relations (FLRs), assigning weight to each FLR and the defuzzification process. Empirical analysis was performed on enrollments data of Alabama University to assess the efficiency of the model. The performance of the proposed model was evaluated by comparing the average forecasting error rate and mean square error values with several fuzzy time series models in the literatures. Experimental results revealed that the proposed model was better than other fuzzy time series models. The use of RTV was able to grip the fuzziness in time series data and reduce the estimation of forecasting errors. In addition, this technique is capable to identify and describe the underlying structure that influence the occurrence of the uncertainty and high fluctuation of the phenomena under investigation.

Keywords: *High-Order Fuzzy Time Series, Ratio Trend Variation, Enrolment, Fuzzy Logic Relation*

1. INTRODUCTION

The time series forecasting approach is a technique used in analyzing the series of observations collected in the past involving single or more variables. The main goal of this system modelling is to predict the future value in complex systems by looking at the existence of patterns in past data and describe its underlying relationship. There are many forecasting models developed, designed and implemented and published in literature for numerous fields such as geology, ecology, hydrology, social sciences, medicine and many more. From the different perspectives, a lot of effort and progressive studies have been carried out over the past decades to develop and improve the time series forecasting models. An adequate and accurate forecasting model is an essential requirement to improve the prediction for better decision-making. Being highly non-stationary, non-linear and uncertainty in nature of the time series are among the main hindrance in achieving efficient forecast. There have been numerous concepts and techniques proposed and discussed in the past few decades to enhance time series forecasting. Generally, time series forecasting techniques are classified into two categories which are conventional and non-conventional. The Regression and Box-Jenkins method are most commonly used in conventional techniques. These statistical approaches required very strict assumption where data must have normal distribution and stationary. In most cases, these requirements may not be fulfilled for all types of data without any adjustment or transformation of the original data. Other aggregation problems such as measurement error, imprecise, vague and uncertain data also may influence the forecasting performance. These techniques would need more historical data that will affect the prediction accuracy, slow convergence and may deviant due to its computational burden. On the contrary, non-conventional techniques have been successfully implemented across extensive applications in various domains with very minimal assumptions. To avoid statistical conflicts, this technique

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attempts to explain the input-output relationship, internal structure and behavior of the time series. This approach is proposed in line with the development of soft computing to forecast uncertain, vague and imprecise data. Nonlinear modelling approaches such as artificial intelligence based has received considerable attention among researchers in the last two decades. For example, [1], [2], [3] and [4] have proposed artificial neural network (ANN) based models (focus time-delay neural network, back propagation neural network, input delay neural network and feed-forward back propagation) for rainfall forcasting. Most of the case studies accommodated either the regression or ANN. ANN is definitely a very powerful tool for the task but greatly depends on the network design. Thus, poor and too complex structures will affect the learning capabilities and lead to overfitting. Difficulties in obtaining stable solution, high tendencies of over fitting and the need for many controlling parameters are several weaknesses of neural network. Alternatively, fuzzy based time series forecasting has achieved a lot of progress in the soft computing approach.

The fuzzy time series (FTS) is capable in dealing deal with historical data in the form of linguistic values. It has received a lot of attention and applied in various domains to describe the non-linear input-output relationship for time series. Initially, the fuzzy set theory was introduced by [5] and [6] which was a pioneer in explaining the basic concept of fuzzy time series. They also proposed the two classes of the FTS model known as time-invariant [7] and time-variant [8]. The first-order fuzzy time series was applied in the enrollment data at the University of Alabama for both classes of the FTS model. Most of the studies developed methods for solving time-invariant fuzzy time series in most literature. For time-invariant, the max-min composition operator was used in the fuzzy relation between the current (t) and previous states (t-1). This technique was very time consuming in the computational process when fuzzy relation is too large. Subsequently, [9] has adopted [7] work by presenting the arithmetic operation which was more simple and more accurate than the max-min composition operator in the first-order fuzzy time series. [10] then promoted the high-order forecasting model using the same method.

There are several different methods in FTS developed by [11], [12], [13], [14], [15] and [16]. Most of the proposed methods proposed were based on fuzzy lagged autoregressive (AR) models. [17] and [18] extended the fuzzy AR model to the fuzzy Autoregressive Moving Average (ARMA) model for IMKB data and gold prices data into first-order and high-order forecasts. For a class of time-variant fuzzy time series, [14] proposed a computational method by introducing the time-variant parameter calculated from the difference between the changes in values of three consecutive years for high order fuzzy time series. This technique was used to deal with the high uncertainty of data, no periodicity and large fluctuation for crop production. Further, [19] extended the fuzzy difference parameter comprehensively by developing a high-order time series model to examine a suitable order. From the drawbacks of previous works of the fuzzy time series model, some improvements have been made in [9]. There were common issues highlighted that might affect the forecasting accuracy such as (1) determination of effective length of interval, (2) managing fuzzy logical relationship (FLRs) into fuzzy logical relationship groups (FLRGs), and (3) Defuzzification process.

A method on how data can be partitioned into intervals using an effective length for each fuzzy set is very important. Initially, most of the studies used a same number of intervals to fuzzify the data with no specific basis ([7], [8], [9], [10], [12], [20]). This method is called random partitioning. More progress has been made in determining the effective length of interval using different methods of partitioning the interval later on. Other methods here are distribution based partitioning [20], average based partitioning [20], ratio-based lengths of interval [13] and mean-based discretization [15]. Meanwhile, some other methods are evolutionary algorithm based partitioning, for example the hybrid fuzzy time series as proposed by [21]. Genetic algorithm was used to partition the universe of discourse into unequal interval and then calibrate with the fuzzy time series model. [22] also employed the particle swarm optimization to partition the index 100 data of the Istanbul Stock Exchange. Therefore, the decomposition of universe of discourse into effective lengths of intervals had an impact on the forecasting results [20]. Again, the length of the interval and number of intervals to be fuzzified depended on the nature of data. Some data might have very large fluctuation, non-stationary, additive or multiplicative patterns while some are not. Therefore, the data need to be transformed or modified before the universe of discourse and subintervals are defined. The utilization of differences to define the universe of discourse is very wide in many forecasting methods in the fuzzy

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

time series ([23], [24]). However, the differences were not sufficient to capture the increase and decrease in the time series. Then, [25], [26] and [27] utilized the percentage change and rate of change of time series respectively for the universe of discourse. Eventually, both had the same manner to perform the fuzzy time series.

After all the observations have been fuzzified, the FLRGs were established from the FLRs. Not much difference or improvements were made since [9] at this stage. However, some consideration may influence the forecast values. For example [15] considered the occurrence of FLRs even though they were the same FLRs. How many FLRs appeared will be considered in the FLRG for justification of the model. They introduced the use of weight of which the value was equivalent to the same fuzzy set. In [17] and [18] as well, the frequency of FLRs in FLRG has to be taken into account to improve the fuzzy forecasts. This consideration was relatively important for determining the defuzzification method.

As mentioned above, the application of arithmetic operations by [9] in defuzzification is simple and in fact very efficient. Meanwhile, some studies introduced the application of weight in defuzzification to enhance the forecasting performance. The approach by [28] was more stable when the future trend of time series data is in a irregular manner. To cope with the presence of hidden linguistic values, visible matrix and hidden matrix were considered in the weighted-transitional matrix. An adjustment has been made in the defuzzification forecasts based on this approach. [15] introduced an index-based defuzzification technique. The forecasted value was calculated by multiplying the midpoint and its relative weights and added all possible FLRs in the same FLRG. The midpoint was computed from the average of data values which fall within the sub-intervals of each fuzzy set. Meanwhile, the weight is represented by the 'subscript' of its own fuzzy set. [26] introduced an ordered weighted aggregation (OWA) based on rate of change of outpatient data. The importance of each fuzzy set is represented by its weight designed from the priority matrix. In the final stage, the weighted centroid was used in the defuzzification process for a 3-order fuzzy relation. Primarily, determining the OWA was based on the frequency of data values in each fuzzy set and the OWA was determined in accordance to FLRs. Meanwhile, [17] and [18] implemented the weighted average from middle points of intervals to calculate forecast values in defuzzification. The weight was determined based on the number of the same fuzzy sets with the highest memberships of the defined intervals in the right-hand side (RHS) of FLRs.

The techniques mentioned above have been proposed in developing the FTS model to resolve the drawbacks from other techniques specifically related to many circumstances to enhance forecasting performance. However, there are some of improvements that can be made on the FTS models. The drawback of most forecasting methods is the utilization of actual time series data and differences as universe of discourse. The fluctuation in time series data cannot be captured from the actual data or differences alone. Non of them explore patterns in time series to model the FTS. Therefore, ratio of trend variation (RTV) has been introduced to capture the patterns for FTS model. These patterns appear naturally because of major components involves in time series data which are trend, cyclical, seasonal and irregular. These are the factor that reflects the characteristics of the data itself. The analysis should be treated separately between patterns and true fluctuation because both are confounded in actual time series data. Negligence in analyzing these patterns will affect the forecasting accuracy.

The aim of this study is to develop a high-order forecasting technique that takes into account the characteristics of the time series data. It is suitable for general application related to time series forecasting especially for univariate or multivariate modelling and time-invariant system. This paper presents a forecasting technique based on high-order FTS model which accomodates patterns that exist in time series data. This model is capable to reduce forecasting error by introducing RTV in determining the universe of discourse. The model can be also be applied in a situation of very high fluctuation and uncertainty. The accuracy of the forecasting model is seen to be improved by introducing effective length of interval to define fuzzy linguistic values using RTV. At the beginning stage, determining the universe of discourse is a crucial part in considering the time series data. The utilization of RTV is to overcome the deficiency of differences in capturing long term trend direction in time series data. Rationally, the trend pattern has been considered in the analysis together with the existing problems related to the nature of time series data namely large fluctuation and non-stationary. In this paper, the time-invariant FTS model is proposed by focusing on the data with the presence of trend, seasonal and cyclical

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

patterns. This approach significantly improves the forecasting accuracy in the first-order as well as the high-order forecasting models. To demonstrate the application of the proposed method, the enrollment data from the University of Alabama was used. The developed model has been evaluated by comparing with other existing models to affirm its preeminence.

2. FUZZY TIME SERIES

The basics concepts and theories in the Fuzzy time series (FTS) are almost similar to conventional time series. The difference is in the values represented by the fuzzy set [5], while conventional time series had real values. FTS was proposed by [7] that combined both the fuzzy set and time series. To explain the FTS and its application in the AR model, this section briefly reviews some of the concepts and definitions. Let $U = \{u_1, u_2, \ldots, u_n\}$ denotes the universe of discourse. A fuzzy set A in U is defined as

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \ldots + f_A(u_n)/u \quad (1)$$

where the membership function of the fuzzy set A_i is $f_A : U \to [0,1]$, and $f_A(u_i)$ is the degree of membership u_i that belongs to A_i for $1 \le i \le n$. The definitions related to fuzzy time series in accordance to AR model are given as follows:

Definition 1. Let $Y(t)$ $(t = 1,2,3,\ldots)$ denote a subset of real numbers and the universe of discourse where the fuzzy sets $f_i(t)$ are defined. If $F(t)$ is a collection of $f_i(t)$ $(i = 1,2,3,\ldots)$, then $F(t)$ is called the fuzzy time series on $Y(t)$.

Definition 2. Consider that $R(t,t-1)$ as the fuzzy logic relation (FLR) between both fuzzy sets $F(t-1)$ and $F(t)$. If $F(t)$ is only affected by $F(t-1)$ then it can be expressed as $F(t-1) \to F(t)$ and can also be denoted as $F(t) = F(t-1) \circ R(t,t-1)$, where the symbol "$\circ$" is the max-min composition operator. The relationship between $F(t-1)$ and $F(t)$ is called the first-order fuzzy AR(1) time series model which is affected by lag one.

Definition 3. Assume that $F(t-1) = C_i$ and $F(t) = C_j$, the relationship between $F(t-1)$ and $F(t)$ can be represented by $C_i \to C_j$, where C_i (current state) and C_j (next state) are known as the left-hand side (LHS) and the right-hand side (RHD) of FLR respectively.

Definition 4. Assume that the current state, C_i of fuzzified value $F(t-1)$ has FLR with several next states $C_{k1}, C_{k2}, \cdots, C_{kl}$ as shown as follows:

$$\begin{aligned} &C_i \to C_{k1} \\ &C_i \to C_{k2} \\ &\quad \vdots \\ &C_i \to C_{kl} \end{aligned}$$

If the FLRs having same current state of different next states then they are grouped in same fuzzy logical relationship group (FLRG) as

$$C_i \to C_{k1}, C_{k2}, \cdots, C_{kl}$$

Definition 5. $F(t)$ is called *n*-order FTS, if $F(t)$ is influenced by $F(t-1), F(t-2), \cdots, F(t-n)$ which are the lagged FTS. This *n*-order FTS is known as AR(*n*) that can be expressed as,

$$F(t-n), \cdots, F(t-2), F(t-1) \to F(t) \quad (2)$$

Definition 6. Let $R(t,t-1)$ be the fuzzy relation to define the relationship between $F(t-1)$ and $F(t)$. If $R(t,t-1) = R(t-1,t-2)$ for any *t*, then $F(t)$ is called a time-invariant fuzzy time series. Otherwise, it is called a time-variant fuzzy time series.

3. PROPOSED METHOD

In this section, the proposed methodology using the RTV-Fuzzy time series is elucidated in three phases. A new approach known as the RTV is employed to determine the universe of discourse and partitioning it into sub-intervals. The enrollments data of University of Alabama [7] shown in Table 1 was utilized to describe this approach.

Table 1: Enrollments data of University of Alabama

Year	*t*	Enrollment ($Y(t)$)
1971	1	13055
1972	2	13563
1973	3	13867
1974	4	14696
1975	5	15460

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1976	6	15311
1977	7	15603
1978	8	15861
1979	9	16807
1980	10	16919
1981	11	16388
1982	12	15433
1983	13	15497
1984	14	15145
1985	15	15163
1986	16	15984
1987	17	16859
1988	18	18150
1989	19	18970
1990	20	19328
1991	21	19337
1992	22	18876

Phase 1: Trend analysis

Step 1. Estimate the trend equation using regression equation (3) based on the least squares method. The values of *a* and *b* are the parameters to be estimated given by equation (4) and (5). The equations are stated as follows:

$$T = b_0 + b_1 t \quad (3)$$

$$b_1 = \frac{\sum_{i=1}^{n}\left(t_i - \frac{\sum_i^n t_i}{n}\right)\left(y_i - \frac{\sum_i^n y_i}{n}\right)}{\sum_{i=1}^{n}\left(t_i - \frac{\sum_i^n t}{n}\right)^2} \quad (4)$$

$$b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} t_i\right) \quad (5)$$

Where, t is denoted as $1,2,3,\ldots$ of 1971, 1972, 1973, and so on for the corresponding $Y(t)$ $(t = 1,2,3,\ldots)$ as presented in Table 1. The estimated trend equation is $T = 13425.57 + 240.49t$. Based on Figure 1, the trend for the dataset shows an upward trend since the b_1 value is positive.

Step 2: Prepare the trend values (T_t) for each week using (3) as

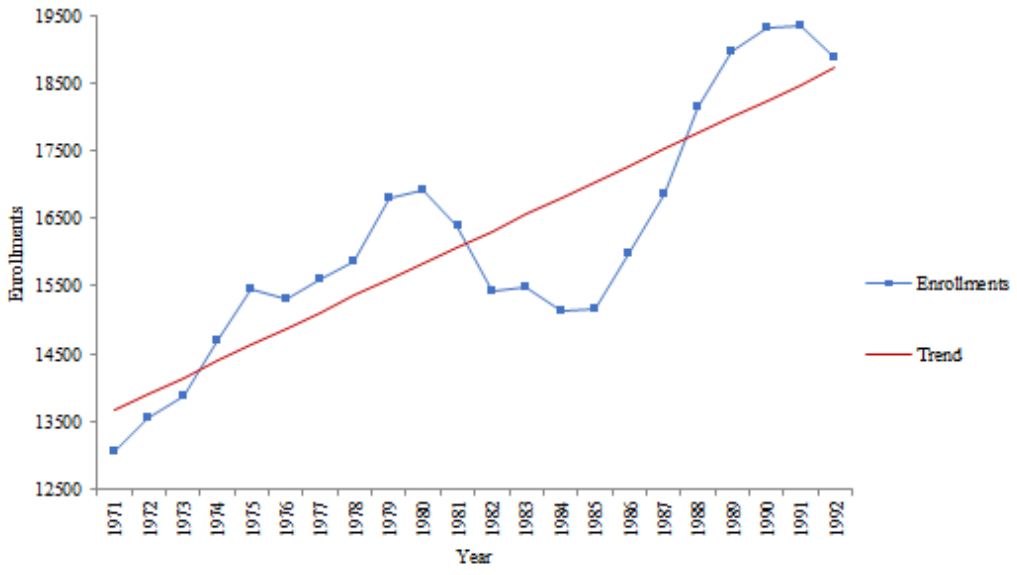

Figure 1: The Enrollments Data And The Trend

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$T_{1(1971)} = 13428.57 + 240.49 \times 1$$
$$= 13669.06$$
$$\vdots$$
$$T_{22(1992)} = 13428.57 + 240.49 \times 22$$
$$= 18719.35$$

Compute the ratio variation (R_t) of original time series value (Y_t) over trend value (T_t) as given by the following equation

$$R_t = \frac{Y_t}{T_t} \times 100\% \quad (6)$$

In this example, ratio variation for the year of 1971 is computed as

$$R_1 = \frac{13055}{13669.06} \times 100\% = 95.51\%$$

The trend values and ratio variations (R_t) are placed in column 4 and 5 of Table 3.

Phase 2: Fuzzification of ratio series

Step 1: From table 3 in column 5, arrange the ratio variation (R_t) in ascending order as

$$R_{(t)} = \{89.01, 90.17, 92.52, \ldots, 107.79\}$$

Where $R_{(t)}$ is the ordered series of R_t .

Step 2: Compute the differences (D) between $r_{(t+1)}$ and $r_{(t)}$ as

$$D_{t+1} = r_{(t+1)} - r_{(t)} \quad (7)$$

Step 3: Find the average of differences values as

$$\overline{D} = \frac{\sum_{i=1}^{n-1} D_i}{n-1} \quad (8)$$

From (7) and (8), average of differences is obtained as

$$\overline{D} = \frac{(90.17 - 89.01) + (92.52 - 90.17) + \ldots + (107.79 - 106.86)}{22 - 1}$$
$$= 0.89$$

Step 4: Define the universe of discourse (U) and sub-interval $\{u_1, u_2, \cdots, u_a\}$ for R_t series $\{r_1, r_2, \cdots, r_n\}$. Let the universe of discourse is $[R_{\min}, R_{\max}]$, where $R_{\min}$ and $R_{\max}$ are minimum and maximum values in R_t . In this example, $R_{\min}$ is 89.01 and $R_{\max}$ is 107.79. Then, U is partitioned into equal sub-interval $u_1, u_2, \cdots$ and u_a by taking $\overline{D}$ as length of interval in step 3 as

$$u_i = \left[R_{\min} + (i-1) \times \overline{D}, R_{\min} + (i) \times \overline{D} \right] \quad (9)$$

From here, 22 sub-intervals and the boundary for each sub-interval are formed to define universe of discourse (U) . The following sub-interval of U is calculated based on (9):

$$u_1 = [89.01, 89.90]$$
$$u_2 = [89.90, 90.79]$$
$$\vdots$$
$$u_{22} = [107.70, 108.59]$$

where $R_{\min}$ and $R_{\max}$ are included in the first interval and last interval respectively.

Step 5: Identify and allocate all observations in R_t series that belongs to each sub-interval determined in step 4. Then, the average of observations in each sub-interval is computed to represent the adjusted midpoint (m_i) of the sub-interval given by

$$m_i = \left(r_{i1} + r_{i2} + \cdots + r_{ik} \right) / k \quad (10)$$

Where, $r_{i1}, r_{i2}, \cdots, r_{ik}$ are observations occurred in sub-interval u_i and k is the number of observations fall within the sub-interval (frequency). Any sub-interval that do not contains any observation will be discarded. After removing empty sub-interval, the number of the remaining sub-intervals was 18 ($\{u_1, u_2, \cdots, u_{18}\}$) which contains at least one observation for each. The number of sub-intervals to be considered, boundaries and observation/s lie in each boundary are arrange in columns 1, 2 and 3 of Table 2. The corresponding midpoints are

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

calculated based on equation (10) with its respective frequency. For example, the midpoint for u_{11} is calculated as

$$m_{11} = \frac{101.95 + 102.12 + 102.21}{3} = 102.09$$

Column 5 of Table 2 shows the corresponding midpoints for all the sub-intervals.

Table 2: The midpoints of sub-intervals

Sub-interval (u_i)	Boundary of u_i	Observation (u_{ik})	Frequency (k)	Midpoint (m_i)
u_1	[89.01, 89.90]	89.01	1	89.01
u_2	[89.90, 90.79]	90.17	1	90.17
u_3	[91.68, 92.57]	92.52	1	92.52
u_4	[93.46, 94.35]	93.61	1	93.61
u_5	[94.35, 95.24]	94.60	1	94.60
u_6	[95.24, 96.13]	95.51	1	95.51
u_7	[96.13, 97.02]	96.24	1	96.24
u_8	[97.02, 97.91]	97.51	1	97.51
u_9	[97.91, 98.80]	98.00	1	98.00
u_{10}	[100.58, 101.47]	100.84	1	100.84
u_{11}	[101.47, 102.36]	101.95, 102.12, 102.21	3	102.09
u_{12}	[102.36, 103.25]	102.96, 103.25	2	103.11
u_{13}	[103.25, 104.14]	103.31	1	103.31
u_{14}	[104.14, 105.03]	104.64	1	104.64
u_{15}	[105.03, 105.92]	105.4, 105.67	2	105.54
u_{16}	[105.92, 106.81]	105.97	1	105.97
u_{17}	[106.81, 107.70]	106.86	1	106.86
u_{18}	[107.70, 108.59]	107.79	1	107.79

Step 6: Define fuzzy sets $C_1, C_2, \cdots, C_a$ accordance to the number of sub-interval for the time series. The fuzzy set is expressed as $C_i = c_{ij}/u_j + c_{ij+1}/u_{j+1} + \cdots + c_{ia}/u_a$ where c_{ij} is defines as

$$c_{ij} = \begin{cases} 1 & , \quad i = j \\ 0.5 & , \quad i = j+1 \text{ or } i = j-1 \\ 0 & , \quad \text{others} \end{cases}$$

c_{ij} indicates the degree of membership for u_j in the fuzzy set C_i for $i, j = 1, 2, 3, \cdots, a$. Since time series ($R_t$) has been divided into 18 sub-intervals ($u_1, u_2, \cdots, u_{18}$), then there are 18 linguistic variables to be defined. In this case, linguistic variables are too many and meaningless to be defined unless they only can be represented by fuzzy sets ($C_1, C_2, \cdots, C_{18}$).

Step 7: Fuzzify the ratio series (R_t). Assign each observation (r_t) into fuzzy set C_i with the highest degree of membership of defined sub-intervals. For example, if the ratio value in 1971 (r_1) is 95.51 which falls within the sub-interval u_6 ([95.24, 96.13]), then the fuzzified ratio is C_6. Repeat the same process to fuzzify the R_t series as presented in the last column of Table 3. Finally, all ratio variations (R_t) have been converted into fuzzy sets ($C_1, C_2, \cdots, C_{18}$).

Table 3: Fuzzified Ratio Series

Year	t	Enrollment (Y_t)	Trend (T_t)	Ratio (R_t)	Fuzzified Ratio (C_i)
1971	1	13055	13669.06	95.51	C_6
1972	2	13563	13909.55	97.51	C_8
1973	3	13867	14150.04	98.00	C_9
1974	4	14696	14390.53	102.12	C_{11}

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1975	5	15460	14631.02	105.67	C_{15}
1976	6	15311	14871.51	102.96	C_{12}
1977	7	15603	15112.00	103.25	C_{12}
1978	8	15861	15352.49	103.31	C_{13}
1979	9	16807	15592.98	107.79	C_{18}
1980	10	16919	15833.47	106.86	C_{17}
1981	11	16388	16073.96	101.95	C_{11}
1982	12	15433	16314.45	94.60	C_5
1983	13	15497	16554.94	93.61	C_4
1984	14	15145	16795.43	90.17	C_2
1985	15	15163	17035.92	89.01	C_1
1986	16	15984	17276.41	92.52	C_3
1987	17	16859	17516.90	96.24	C_7
1988	18	18150	17757.39	102.21	C_{11}
1989	19	18970	17997.88	105.40	C_{15}
1990	20	19328	18238.37	105.97	C_{16}
1991	21	19337	18478.86	104.64	C_{14}
1992	22	18876	18719.35	100.84	C_{10}

Step 8: Establish the fuzzy logic relation (FLR) and subsequently form the fuzzy logical relationship groups (FLRGs) for first order autoregressive model (AR(1)) based on definition 3. For example, the relationship between $F(1971)$ and $F(1972)$ can be denoted by $C_6 \rightarrow C_8$ as in column 3 of Table 4. Out of 21 FLRs, 17 FLRGs are formed based on definition 4. Based on FLR in Table 4, there are some LHS having the same current state with different RHS of the next state. For example, $C_{11} \rightarrow C_{15}$, $C_{11} \rightarrow C_5$ and $C_{11} \rightarrow C_{15}$ are group into $C_{11} \rightarrow C_{15}, C_5, C_{15}$ as in column 2 of Table 5.

Table 4: The First Order Fuzzy Logical Relationship (FLRs) Of Fuzzified Ratio

Year	Fuzzified Ratio (C_i)	FLR
1971	C_6	-
1972	C_8	$C_6 \rightarrow C_8$
1973	C_9	$C_8 \rightarrow C_9$
1974	C_{11}	$C_9 \rightarrow C_{11}$
1975	C_{15}	$C_{11} \rightarrow C_{15}$
1976	C_{12}	$C_{15} \rightarrow C_{12}$
1977	C_{12}	$C_{12} \rightarrow C_{12}$
1978	C_{13}	$C_{12} \rightarrow C_{13}$
1979	C_{18}	$C_{13} \rightarrow C_{18}$
1980	C_{17}	$C_{18} \rightarrow C_{17}$
1981	C_{11}	$C_{17} \rightarrow C_{11}$
1982	C_5	$C_{11} \rightarrow C_5$
1983	C_4	$C_5 \rightarrow C_4$
1984	C_2	$C_4 \rightarrow C_2$
1985	C_1	$C_2 \rightarrow C_1$
1986	C_3	$C_1 \rightarrow C_3$
1987	C_7	$C_3 \rightarrow C_7$
1988	C_{11}	$C_7 \rightarrow C_{11}$
1989	C_{15}	$C_{11} \rightarrow C_{15}$
1990	C_{16}	$C_{15} \rightarrow C_{16}$
1991	C_{14}	$C_{16} \rightarrow C_{14}$
1992	C_{10}	$C_{14} \rightarrow C_{10}$

Table 5: The First Order Fuzzy Logical Relationship Groups (FLRGs) Of Fuzzified Ratio

Group	FLRGs
1	$C_1 \rightarrow C_3$
2	$C_2 \rightarrow C_1$
3	$C_3 \rightarrow C_7$
4	$C_4 \rightarrow C_2$
5	$C_5 \rightarrow C_4$
6	$C_6 \rightarrow C_8$
7	$C_7 \rightarrow C_{11}$
8	$C_8 \rightarrow C_9$
9	$C_9 \rightarrow C_{11}$
10	$C_{11} \rightarrow C_5, C_{15}, C_{15}$
11	$C_{12} \rightarrow C_{12}, C_{13}$
12	$C_{13} \rightarrow C_{18}$
13	$C_{14} \rightarrow C_{10}$
14	$C_{15} \rightarrow C_{12}, C_{16}$
15	$C_{16} \rightarrow C_{14}$
16	$C_{17} \rightarrow C_{11}$
17	$C_{18} \rightarrow C_{17}$

Phase 3: Defuzzification of fuzzified ratio variation

Step 1: Obtain the fuzzy time series forecasts after defuzzifying the fuzzy forecasts. The centralization method was used to defuzzify the fuzzy time series forecasts into real values by taking the adjusted midpoints of sub-intervals in Table 2. The calculations attained by the following rules and applicable for the *n*-order FTS model:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Rule 1: If FLRG is $C_p, C_q, \cdots, C_r \rightarrow C_i, C_i, \cdots, C_i$ then the fuzzy forecast is C_i. Where the maximum membership value of C_i occurred in sub-interval u_i with midpoint m_i, then the forecasted value $\hat{r}_{t+1}$ at time t is m_i as given below

$$\hat{r}_{t+1} = m_i \quad (9)$$

For example, if the FLRG is $C_2 \rightarrow C_1$ in 1-order FTS model then the fuzzy forecast is C_1. As for the 2-order FTS model with FLR is C_3, $C_2 \rightarrow C_1$, C_1 the fuzzy forecast is C_1 and applied the same process for the higher order of FTS model.

Rule 2: If FLRG is
$C_p, C_q, \cdots, C_r \rightarrow (C_i, C_i, \cdots, C_i), (C_j, C_j, \cdots, C_j), \cdots, (C_k, C_k, \cdots, C_k)$
where $C_p, C_q, \cdots, C_r$ influence C_i occurs by α times, C_j by β times, C_k by γ times and so on, then the fuzzy forecast is $(C_i, C_i, \cdots, C_i), (C_j, C_j, \cdots, C_j), \cdots, (C_k, C_k, \cdots, C_k)$. Hence, the maximum membership values of $(C_i, C_i, \cdots, C_i), (C_j, C_j, \cdots, C_j), \cdots, (C_k, C_k, \cdots, C_k)$ occurred in sub-interval $u_i, u_j, \cdots, u_k$ with midpoints $m_i, m_j, \cdots, m_k$. Therefore, the calculation for defuzzification forecasts at time t is using weighted average is as follows:

$$\hat{r}_{t+1} = \frac{\alpha \times m_i + \beta \times m_j + \ldots + \gamma \times m_k}{\alpha + \beta + \ldots + \gamma} \quad (10)$$

In Table 5, one example of FLRG is $C_{11} \rightarrow C_5, C_{15}, C_{15}$ and the fuzzy forecast determined for the FLRG is C_5, C_{15}, C_{15}.

Rule 3. If FLRG is $C_p, C_q, \cdots, C_r \rightarrow$ empty, then the fuzzy forecast is $C_p, C_q, \cdots, C_r$. Where the maximum membership value of $C_p, C_q, \cdots, C_r$ occurred in sub-interval $u_p, u_q, \cdots, u_r$ with midpoints $m_p, m_q, \cdots, m_r$, then the forecasted value $\hat{r}_{t+1}$ at time t is calculated as equation (10). For example, the FLRG is $C_2 \rightarrow$ empty then the fuzzy forecast determined is C_2 and the same process can be applied for the *n*-order FTS model.

Step 2: Forecast the time series data. This stage begins with estimation of the trend using equation (11) and ratio variation $\hat{r}_{t+1}$ in step 1. Meanwhile, the forecasted value for $\hat{y}_{t+1}$ is calculated using equation (12) as given below

$$T_{t+1} = 13425.57 + 240.49 t_{t+1} \quad (11)$$

$$\hat{y}_{t+1} = \frac{T_{t+1}}{100} \times \hat{r}_{t+1} \quad (12)$$

From the application of the proposed method, the results are shown in Table 6. The forecasted ratio variation and students' enrollments are the solution from step 1 and 2 in Phase 3 as in columns 5 and 6. The entire procedures are applicable for the first order FTS model. It may also be extended to high-order FTS models using same dataset as depicted in Figure 2. For meaningfull visualisation, several high-order models are to be compared with first order model and the actual enrollments dataset as well. From the multiple comparison, the forecasting was almost consistent particularly for order 3 or more.

Table 6: Forecasted Ratio Variation And Enrollments

Year	Enrollment (Y_t)	Trend (T_t)	Ratio (R_t)	Forecasted ratio ($\hat{R}_t$)	Forecasted Enrollment ($\hat{Y}_t$)
1971	13055	13669.06	95.51	-	-
1972	13563	13909.55	97.51	97.51	13563.20
1973	13867	14150.04	98.00	98.00	13867.04
1974	14696	14390.53	102.12	102.09	14691.29
1975	15460	14631.02	105.67	102.23	14957.29
1976	15311	14871.51	102.96	104.54	15546.68
1977	15603	15112.00	103.25	103.21	15597.10
1978	15861	15352.49	103.31	103.21	15845.30
1979	16807	15592.98	107.79	105.55	16458.39
1980	16919	15833.47	106.86	106.86	16919.65
1981	16388	16073.96	101.95	102.09	16409.91

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1982	15433	16314.45	94.60	102.23	16678.26
1983	15497	16554.94	93.61	93.61	15497.08
1984	15145	16795.43	90.17	90.17	15144.44
1985	15163	17035.92	89.01	89.01	15163.67
1986	15984	17276.41	92.52	92.52	15984.13
1987	16859	17516.90	96.24	96.24	16858.26
1988	18150	17757.39	102.21	102.09	18128.52
1989	18970	17997.88	105.40	102.23	18399.23
1990	19328	18238.37	105.97	104.54	19066.39
1991	19337	18478.86	104.64	104.64	19336.28
1992	18876	18719.35	100.84	100.84	18876.59

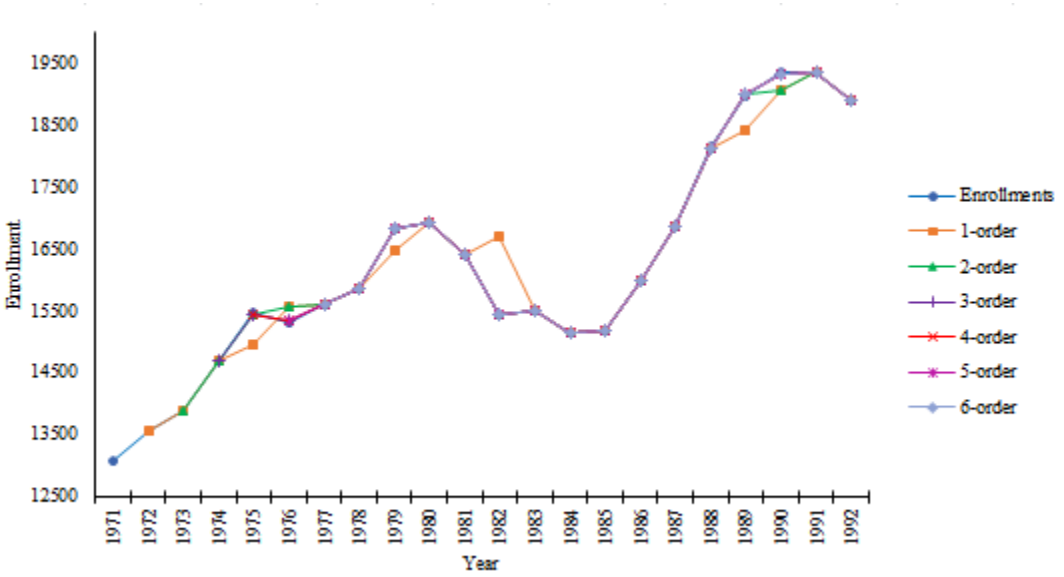

Figure 2: The Comparison Of Actual Enrollments Data And The Forecasting Models For Various Orders

4. PERFORMANCE EVALUATION

In this section, a comparison of accuracy is made to evaluate the performance of the proposed forecasting model with other models. The performance indicators are the average forecasting error rate (*AFER*) and mean square error (*MSE*). It is defined by the following equations:

$$AFER = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t| / y_t}{n} \times 100\% \quad (13)$$

$$MSE = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n} \quad (14)$$

Where, y_t is the actual value of time series data, $\hat{y}_t$ is the forecasted value and n is the total number values in time series data.

The higher accuracy of a model will relatively lower the *AFER* and *MSE* values. Using the same dataset, the *AFER* of the proposed partitioning method was compared with other available methods for first order model as presented in Table 7. Results reflected that, most of the random partitioning methods shows larger *AFER* values than other methods which is greater than 2.75. They are using 7 fixed number of intervals to define universe of discourse. However, the increasing in number of intervals does not promises for a better model. Moreover, the proposed model using RTV-based shows the lowest rate as 0.94 and preeminence as compared to other random and non-random partitioning methods. The main key in deciding the length of interval and suitable number of interval must be neither too small nor large. It is important to preserve the fluctuation in time series that make the fuzzy time series meaningful.

Table 7: A Comparison Of First Order Models

Model	Partitioning	#Interval	*AFER*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	method		(%)
[7]	Random	7	3.23
[9]	Random	7	3.11
[12] (MEPA)	Random	7	2.75
[12] (TFA)	Random	7	3.04
[20]	Distribution-based	18	1.51
[20]	Average-based	24	1.31
[13]	Ratio-based	21	1.45
[29]	Frequency density-based	13	1.02
[15]	Mean-based discretization	13	1.19
Proposed model	Ratio trend variation-based	18	0.94

From first order model, it has been extended to high-order models based on definition 5 and step 2 onwards in phase 2. The same results in Figure 3 and Figure 4 also showed that the proposed model using the RTV-based method was more stable and consistent for higher order models based on *AFER* and *MSE* values. There are a significant difference between the order of first, second and third but a slight difference after the third order onward. It is suggested that the third order model is suitable to fit the enrollments data.

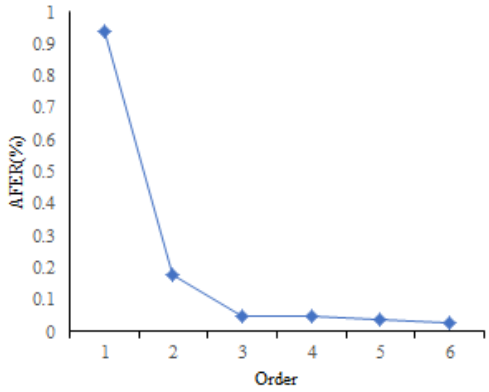

Figure 3: The AFER Of Various Orders Of Proposed Model

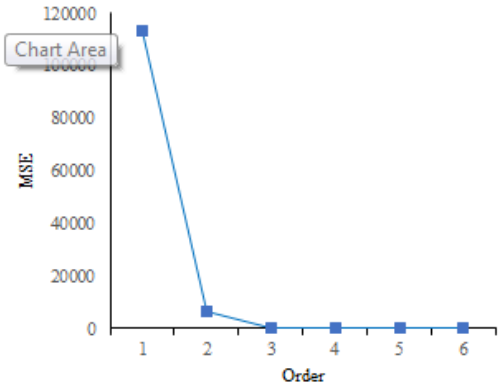

Figure 4: The MSE Of Various Orders Of Proposed Model

Based on Table 8, higher order models do not necessarily give an impact on forecasting errors for certain models. Sometimes, the errors become unstable and inconsistent for higher order models [18]. However, AFER values for the proposed model are small and remain constant at third order onward as compared to others. Since the 3-order model is recommended, a comparison is made with other common models based on the AFER and MSE values. Results in Table 9, show that the proposed model is superior as compared to other models which has the smallest values for both performance indicators.

Table 8: A Comparison Of AFER For High-Order Models

Model	2-order	3-order	4-order	5-order
[10]	2.55	1.93	1.85	1.77
[19]	1.80	1.56	1.74	1.68
[15]	0.66	0.19	0.20	0.20
Proposed model	0.18	0.05	0.05	0.04

Table 9: A Comparison Of AFER And MSE For 3-Order Model

Model	*AFER* (%)	*MSE*
[10]	1.93	185115
[26]	1.65	174391
[19]	1.56	97180
Proposed model	0.05	153

5. DISCUSSION AND CONCLUSION

Evaluation of the proposed model was performed on the students' enrollment data of University of Alabama and results showed that the performance of the proposed model outperformed other models in previous studies. In general, the high-order model is better than the first-order from the aspect of forecasting performance. However, it

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

depends on the nature of the time series data, the factors and how it influences the behavior of historical data. Besides natural variability, specific pattern might be present in the time series data seem like trend, seasonal or cyclical which are very common. In addition, non-stationary, non-linear and high fluctuations are the most impeding factors to achieve better forecasting performance. Those are the factors to be considered in the forecasting process for time series data. Although fuzzy time series is employed in the forecasting model but it is still less efficient if the time series exhibited upward or downward trends as well as seasonality and cyclical patterns. Most literature utilized original historical data and the rate of change in defining the universe of discourse to develop time series forecasting which disregarded the existence of trend and seasonality. Therefore, this paper proposed the utilization of the RTV to define the universe of discourse with an appropriate partitioned number of intervals employed in fuzzy time series model. From there, the trend direction either upwards or downwards can be captured as well as its seasonal fluctuation. Another advantage of using this method is the ability to handle high fluctuations in time series data. In addition, this technique is capable to identify and describe the underlying structure that influence the occurrence of the uncertainty and high fluctuation of the phenomena under investigation. This method proves that the forecasting accuracy depended on how the universe of discourse is defined. To avoid ambiguity, higher order models were derived where the performance of the proposed model has been shown to be more stable and consistent. It was demonstrated that the 3-order RTV-fuzzy time series is appropriate and free from ambiguities. It is also proven that with the *AFER* and *MSE* values, using RTV-fuzzy time series improved the forecasting accuracy.

ACKNOWLEDGEMENTS

The authors thank the Ministry of Higher Education Malaysia for funding this study under the Long Term Research Grant Scheme (LRGS/b-u/2012UUM/Teknologi Komunikasi dan Informasi) and the Department of Irrigation and Drainage Malaysia for supplying hydrology and reservoir operation data.

REFERENCES

[1] K. K. Htike and O. O. Khalifa, "Rainfall forecasting models using focused time-delay neural networks", in *International Conference on Computer and Communication Engineering*, 2010, pp. 11–13.

[2] E. Vamsidhar, K. V. S. R. P. Varma, P. S. Rao, and R. Satapati, "Prediction of rainfall using backpropagation neural network model", *Int. J. Comput. Sci. Eng.*, Vol. 02, No. 04, 2010, pp. 1119–1121.

[3] A. El-Shafie, A. Noureldin, M. Taha, A. Hussain, and M. Mukhlisin, "Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia", *Hydrol. Earth Syst. Sci.*, Vol. 16, No. 4, 2012, pp. 1151–1169.

[4] I. 'Izzati Abdul Rahman and N. M. A. Alias, "Rainfall forecasting using an artificial neural network model to prevent flash floods", *International Conference on High-capacity Optical Networks and Emerging Technologies*, 2011, pp. 323–328.

[5] L. A. Zadeh, "Fuzzy sets", *Inf. Control*, Vol. 8, No. 3, 1965, pp. 338–353.

[6] Q. Song and B. S. Chissom, "Fuzzy time series and its models", *Fuzzy Sets Syst.*, Vol. 54, No. 3, 1993, pp. 269–277.

[7] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - part I", *Fuzzy Sets Syst.*, Vol. 54, No. 1, 1993, pp. 1–10.

[8] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - Part II", *Fuzzy Sets Syst.*, Vol. 62, 1994, pp. 1–8.

[9] S. M. Chen, "Forecasting enrollments based on fuzzy time series", *Fuzzy Sets Syst.*, Vol. 81, No. 3, 1996, pp. 311–319.

[10] S. M. Chen, "Forecasting enrollments based on high-order fuzzy time series", *Cybern. Syst.*, Vol. 33, No. 1, 2002, pp. 1–16.

[11] M. Sah and Y. D. Konstantin, "Forecasting enrollment model based on first-order fuzzy time series", *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 1, No. 1, 2005, pp. 375–378.

[12] C. H. Cheng, J. R. Chang, and C. A. Yeh, "Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost", *Technol. Forecast. Soc. Change*, vol. 73, no. 5, pp. 524–542, 2006.

[13] K. Huarng and T. H. K. Yu, "Ratio-based lengths of intervals to improve fuzzy time series forecasting," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, Vol. 36, No. 2,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2006, pp. 328–340.

[14] S. R. Singh, "A computational method of forecasting based on fuzzy time series", *Math. Comput. Simul.*, Vol. 79, No. 3, 2008, pp. 539–554.

[15] P. Singh and B. Borah, "An efficient time series forecasting model based on fuzzy time series", *Eng. Appl. Artif. Intell.*, Vol. 26, No. 10, 2013, pp. 2443–2457.

[16] M. Othman and S. N. F. Azahari, "Deseasonalised forecasting model of rainfall distribution using fuzzy time series", *J. Inf. Commun. Technol.*, Vol. 2, No. 2, 2016, pp. 153–169.

[17] C. Kocak, "First-order ARMA type fuzzy time series method based on fuzzy logic relation tables", *Math. Probl. Eng.*, Vol. 2013, 2013, pp. 1–12.

[18] C. Kocak, "ARMA(p,q) type high order fuzzy time series forecast method based on fuzzy logic relations", *Appl. Soft Comput.*, Vol. 58, 2017, pp. 92–103.

[19] S. R. Singh, "A computational method of forecasting based on high-order fuzzy time series", *Expert Syst. Appl.*, Vol. 36, No. 7, 2009, pp. 10551–10559.

[20] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series", *Fuzzy Sets Syst.*, Vol. 123, 2001, pp. 387–394.

[21] S. Sakhuja, V. Jain, S. Kumar, C. Chandra, and S. K. Ghildayal, "Genetic algorithm based fuzzy time series tourism demand forecast model", *Ind. Manag. Data Syst.*, Vol. 116, No. 3, 2016, pp. 483–507.

[22] E. Eğrioglu, C. H. Aladag, U. Yolcu, and A. Zafer Dalar, "A hybrid high order fuzzy time series forecasting approach based on PSO and ANNs methods", *Am. J. Intell. Syst.*, Vol. 6, No. 1, 2016, pp. 22–29.

[23] J. Hwang, S. Chen, and C.-H. Lee, "Handling forecasting problems using fuzzy time series," *Fuzzy Sets Syst.*, vol. 100, no. 1–3, pp. 217–228, 1998.

[24] M. Bose and K. Mali, "A novel data partitioning and rule selection technique for modeling high-order fuzzy time series", *Appl. Soft Comput. J.*, Vol. 63, 2018, pp. 87–96.

[25] B. Garg, M. M. S. Beg, and A. Q. Ansari, "Fuzzy time series model to forecast rice production", *EEE International Conference on Fuzzy Systems*, 2013, pp. 1–8.

[26] B. Garg and R. Garg, "Enhanced accuracy of fuzzy time series model using ordered weighted aggregation", *Appl. Soft Comput.*, Vol. 48, 2016, pp. 265–280.

[27] P. Jiang, Q. Dong, P. Li, and L. Lian, "A novel high-order weighted fuzzy time series model and its application in nonlinear time series prediction", *Appl. Soft Comput.*, Vol. 55, 2017, pp. 44–62.

[28] C. H. Cheng, J. W. Wang, and C. H. Li, "Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix", *Expert Syst. Appl.*, Vol. 34, No. 4, 2008, pp. 2568–2575.

[29] T. A. Jilani, syed M. A. Burney, and C. Ardil, "Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning", *Int. J. Comput. Inf. Eng.*, Vol. 4, No. 7, 2010, pp. 1194–1199.