MATEMATIKA, 2009, Volume 25, Number 1, 67–78

Modified Weighted for Enrollment Forecasting Based on Fuzzy Time Series

[1]**Muhammad Hisyam Lee,** [2]**Riswan Efendi &** [3]**Zuhaimy Ismail**
[1,2,3]Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia
81310 UTM Skudai, Johor, Malaysia
e-mail: [1]mhl@utm.my, [2]riswan_efendi@yahoo.com, [3]zhi@mel.fs.utm.my

Abstract The enrollment study is main point for the university planning. Many previous studies have been presented for enrollment forecasting. This paper proposed the adoption the weighted and the difference between actual data toward midpoint interval based on fuzzy time series. By using the enrollment of the University of Alabama and Universiti Teknologi Malaysia (UTM), as data sets for training and testing, the performance of the adoption approach has been shown much improvement in terms of MSE (Mean Square Error) and average error of forecasting measurements.

Keywords Recurrence; chronological order; modified weighted; enrollment; fuzzy time series.

1 Introduction

In the last decade, fuzzy time series has been widely used for forecasting data of dynamic and non-linear data in nature. Many previous studies have been discussed for forecasting used fuzzy time series such as enrollment [1-3, 12, 15, 17, 22, 23, 26], the stock index [14, 16, 18, 19, 21, 24, 25, 27], temperature [13] and financial forecasting [20], etc. The accuracy of the proposed forecasting methods is one of the main issues discussed. The enrollment forecasting studies have been presented in many methods and models such as regression models, time series models and artificial intelligences. Most of them were done for real data or numerical data. It is very important to make reasonably accurate estimates of the future enrollment for a university because many decisions can be elaborated from them. However, obtaining accuracy is not easy task, as many variables have impacts on enrollment numbers.

Many different methods and models have been written by researcher used fuzzy time series for enrollments forecasting [1]. Song and Chissom [1] initiated a study on time-invariant and time-variant models for forecasting with fuzzy time series using enrollment data of Alabama University. In this paper, the study is concern to adopt the weighted on fuzzy time series forecasting. Yu [16] proposed the used of weighted fuzzy time series models for Taiwan stock index (TAIEX) forecasting. It is assigned by the recurrent fuzzy logical relationships (FLRs) in fuzzy logical group (FLG). In establishing fuzzy relationship and forecasting are important step to consider the weighted factor. Cheng *et al.* [19] presented the trend-weighted fuzzy time series model for TAIEX forecasting.

In this study was considered between modified weighted factor and difference of the midpoint intervals and the actual data. It can be expected to be a promising way for reducing of MSE (mean square error) and average error of forecasting if it is compared with Yu [16] and Cheng *et al.* [19]. The paper will be detailed as follows. In section 2, the basic theory of fuzzy set and fuzzy time series is described. In section 3, the proposed method is described in detail with few examples. Further, the verification and comparison are done with some current methods available using student enrollment data at Alabama University and Universiti Teknologi Malaysia (UTM) in section 4. In section 5, the conclusions are mentioned.

2 The Basic Theory of Fuzzy Sets and Fuzzy Time Series

2.1 Definition

There are several definitions have been defined for fuzzy set as in [1]:
Definition 1: Let U be the universe of discourse. A fuzzy subset A on the universe of discourse U can be defined as follows:

$$A = \{ (u_i, \mu_A(u_i)) \ \mid u_i \in U \} \tag{1}$$

where μ_A is the membership function of A, $\mu_A : U \rightarrow$[0, 1], and $\mu_A(u_i)$ is the degree of membership of the element u_i in the fuzzy set A.
Definition 2: Let U be the universe of discourse, $U = \{ u_1, u_2, u_3, \ldots , u_n\}$, and U be a finite set. A fuzzy set A can be expressed as follows:

$$A = \sum_{i=1}^{n} \frac{\mu_A(u_i)}{u_i} = \frac{\mu_A(u_1)}{u_1} + \frac{\mu_A(u_2)}{u_2} + ... + \frac{\mu_A(u_n)}{u_n} \tag{2}$$

where the symbol "+" means the operation of union instead of the operation of summation, and the symbol " - " means the separator rather than the commonly used algebraic symbol of division.
Definition 3: Let U be the universe of discourse, where $U$is an infinite set. A fuzzy set A of U can be expressed as follows:

$$A = \int_U \frac{\mu_A(u_i)}{u_i}, \quad \forall u_i \in U \tag{3}$$

There are several definitions have been defined for fuzzy time series as in [1,12-27]:
Definition 4: Let $Y\ (t)$ be the universe of discourse defined by the fuzzy set $\mu_i(t)$. If $F\ (t)$ consists of $\mu_i(t)$ (i = 1, 2, . . .), $F(t)$ is defined as a fuzzy time series on $Y(t)$ (t = . . . , 0, 1, 2, . . .), where $Y(t)$ is a subset of real number.

Following Definition 3, fuzzy relationships between two consecutive observations can be defined.
Definition 5: Suppose $F\ (t)$ is caused by $F\ (t$-1) denoted by $F\ (t) \quad \rightarrow \quad F\ (t$-1), then this relationship can be represented by

$$F(t) = F(t-1) \circ R(t, t-1) \tag{4}$$

where $R(t, t$-1) is a fuzzy relationship between $F\ (t)$ and $F\ (t$-1) and is called the first-order model of $F\ (t)$.
Definition 6: Let $F\ (t-1) = A_i$ and $F\ (t) = A_j$. The relationship between two consecutive data (called a fuzzy logical relationship, FLR), i.e., $F\ (t)$ and $F\ (t$-1), can be denoted by $A_i \rightarrow \quad A_j$, i, j = 1, 2, . . . , p (where p is interval or subinterval number) is called the left-hand side (LHS), and A_j is the right-hand side (RHS) of the FLR.

The proposed a fuzzy time series model with procedure as follows:

(i) to define the universe of discourse and intervals

(ii) to fuzzify

(iii) to establish fuzzy relationships

(iv) to forecast

Definition 7: Let $A_i \to A_j$, $A_i \to A_k$, ..., $A_i \to A_p$ are FLRs with the same LHS can be grouped into an ordered FLG (called a fuzzy logical group) by putting all their RHS together as on the RHS of the FLG. It can be written as follows:

$$A_i \to A_j, A_k, \ \dots \ , A_p \quad i, j, k, \dots, p = 1, 2, \dots, p \tag{5}$$

3 The Modified Weighted on Fuzzy Time Series

3.1 Fundamental Reason for Weighted

As described in the previous sections, this study proposed the modified weighted for fuzzy time series forecasting. Yu [16] has presented the weighted fuzzy time series models for Taiwan stock index (TAIEX) forecasting where it was considered by using the recurrent of FLR. There are three main reasons why it is necessary to develop these modified weighted for forecasting. Firstly, it is to resolve recurrent fuzzy relationships; secondly, it is to resolve the chronological order on fuzzy relationships, which both of them are used in different interpretation with Yu [16] for weight determining and finally it is to determine a modified weighted based on the chronological number of fuzzy relationships on FLG.

3.2 Recurrence

Some of previous related studies were not clear to describe the recurrent fuzzy relationships. The repeated FLRs were simply ignored when fuzzy relationships were established. The following examples in [16] can be used to explain this. Let there are FLR_s in chronological order as in Table 1.

Table 1. The Recurrent Fuzzy Logical Relationships (FLRs)

FLR_s	Previous studies	Yu's
$A_1 \to A_1$		
$A_1 \to A_2$	$A_1 \to A_1$, A_2	$A_1 \to A_1$, A_2, A_1, A_1
$A_2 \to A_1$	$A_2 \to A_1$	$A_2 \to A_1$
$A_1 \to A_1$		
$A_1 \to A_1$		

Based on Table 1 there are four out of five FLR having the same LHS, A_1. The occurrences of the same FLR in column 2 are regarded as if there were only one occurrence. In other words, the recent identical FLR are simply ignored in [1]. It is questionable if the recurrence is ignored. The occurrence of a particular FLR represents the number of its appearances in the past. For instance, in column 2, $A_1 \to A_1$ appears three times and

$A_2 \rightarrow A_1$ only once. The recurrence can be used to indicate how the FLR may appear in the future. Hence, to cover all of the FLR, an approach to representing the fuzzy relationship is suggested below:

$$A_1 \rightarrow A_1, A_2, A_1, A_1 \tag{6}$$

The various recurrences of FLR are viewpoint to assign of the modified weighted.

3.3 The Chronological Order in Fuzzy Logical Group (FLG)

This is the second type of FLR where there exist a chronological order between A_i and $A_j, A_k, \ldots, A_p$. For example, it will be described as in Table 2.

Table 2. The Chronological Order in Fuzzy Logical Group (FLG)

FLR_s	FLG
$A_1 \rightarrow A_1$	
$A_1 \rightarrow A_2$	
$A_1 \rightarrow A_3$	$A_1 \rightarrow A_1,\ A_2,\ A_3,\ A_4,\ A_5$
$A_1 \rightarrow A_4$	
$A_1 \rightarrow A_5$	

From column 3, it can be denoted that FLR_s between A_1 with others are chronological order. Each A_i and $A_j, A_k,\ \ldots\ , A_p$ is different linguistics value so that we need to determine the weighted various for fuzzy time series before forecasting. The chronological number in FLG can be used as modified weighted.

3.4 The Modified Weighted on Fuzzy Time Series

In this paper is proposed the weighted fuzzy time series based on the various recurrences of FLR. Its computational is assigned as follows:

Suppose $A_i \rightarrow A_j, A_k,\ \ldots\ , A_p$ is a FLG and the weights are specified as follows: $j = c_1$, $k = c_2, \ldots, p = c_n$. The computational can be determined as below:

$$\begin{aligned}\mathbf{W}(t) = [\, w_1\ w_2\ \ldots\ w_n \,] &= \left[\frac{j}{(j+k+...+p)}\ \frac{k}{(j+k+...+p)}\ \cdots \frac{p}{(j+k+...+p)}\right] \\ &= \left[\frac{c_1}{(c_1+c_2+...+c_n)}\ \frac{c_2}{(c_1+c_2+\cdots+c_n)} \cdots \frac{c_n}{(c_1+c_2+...+c_n}\right] \\ &= \left[\frac{c_1}{\sum_{h=1}^{n} c_h}\ \frac{c_2}{\sum_{h=1}^{n} c_h}\ \cdots\ \frac{c_n}{\sum_{h=1}^{n} c_h}\right]\end{aligned} \tag{7}$$

where n is number of FLR in FLG.

In this section, the modified weighted can be elicited by using recurrences and the chronological number of FLRs in FLG as follows:

Let $A_i \rightarrow A_j, A_k,\ \ldots\ , A_p$is assumed that there are possibility that $i,\ j, \ldots, p$ are equal or can be at least one $i,\ j, \ldots, p$ not equal. Let $A_i \rightarrow A_j,\ A_k,\ A_l,\ A_m$ be a recurrence and can be assumed that $j = k = m \neq l$ where $A_i \rightarrow A_j$ appears two time and $A_i \rightarrow A_l$ only once in FLG. Then, the computational for weights can be done as in Table 3.

Table 3: The Computational of Weighted

FLRs	Chronological number	W_n
$A_i \quad \rightarrow \quad A_j$	$j = c_1$	$w_1 = \dfrac{c_1}{(c_1 + c_2 + c_3 + c_4)}$
$A_i \quad \rightarrow \quad A_k$	$k = c_2$	$w_2 = \dfrac{c_2}{(c_1 + c_2 + c_3 + c_4)}$
$A_i \quad \rightarrow \quad A_l$	$l = c_3$	$w_3 = \dfrac{c_3}{(c_1 + c_2 + c_3 + c_4)}$
$A_i \quad \rightarrow \quad A_m$	$m = c_4$	$w_4 = \dfrac{c_4}{(c_1 + c_2 + c_3 + c_4)}$

Further, from column 3 in Table 3 the weighted factor can be proven that is the necessary condition as below

$$w_1 + w_2 + w_3 + w_4 = \frac{c_1}{\sum_{h=1}^{4} c_h} + \frac{c_2}{\sum_{h=1}^{4} c_h} + \frac{c_3}{\sum_{h=1}^{4} c_h} + \frac{c_4}{\sum_{h=1}^{4} c_h} = \frac{3c + c_3}{3c + c_3} = 1$$

because of c_1, c_2 and c_4 are equal so that they can be denoted as c, or it can be written as below

$$\sum_{h=1}^{4} w_h = 1. \tag{8}$$

In addition, the computational of the weighted based on chronological order can be assigned as follow:

Let $A_i \rightarrow A_j$, A_k, A_l, $A_m$be a chronological order and can be assumed that $j \neq k \neq l \neq m$ where between A_i with A_j, A_k, A_l, and A_m only once in FLG. Then, the computational for weights can be done as like in Table 3. From Table 3 column 3 the weighted can be proven that is the necessary condition as follows

$$w_1 + w_2 + w_3 + w_4 = \frac{c_1}{\sum_{h=1}^{4} c_h} + \frac{c_2}{\sum_{h=1}^{4} c_h} + \frac{c_3}{\sum_{h=1}^{4} c_h} + \frac{c_4}{\sum_{h=1}^{4} c_h} = \frac{\sum_{h=1}^{4} c_h}{\sum_{h=1}^{4} c_h} = 1$$

or it can be written as below

$$\sum_{h=1}^{4} w_h = 1. \tag{9}$$

From the Eq.(8) and (9) it can be denoted that the weighted as below

$$\sum_{h=1}^{n} w_h = 1. \tag{10}$$

In this section, the first modification can be seen from chronological number usage. This is main different from Yu [14] where the determining of weights were assumed that w_1= 1, w_2= 2, ..., $w_m = m$(m is natural number). These weighted values were gradually increased. In addition, if weights were multiplied with the midpoint intervals then the forecast values will increase also. On the other hand, Cheng et al. [18] proposed the trend-weighted fuzzy time series where they considered the recurrent classification of fuzzy relationships into three different types of trends and assign a proper weight to individual fuzzy relationships. The second modification can be described in reversal of transpose matrix elements on forecasting rule.

3.5 Proposed Method on Forecasting

In this proposed method, the forecasting is also used the weighted method and combined with difference actual data and midpoint interval. The weighted are determined according to chronologically number of FLRs in FLG can be seen detail in example. The Forecasting rule on fuzzy time series as follow:

Let $A_i \rightarrow A_j, A_k,\ \dots\ , A_p$ is a FLG and the corresponding weights for $A_j, A_k,\ \dots\ , A_p$are $w_1,\ w_2,\ \dots,\ w_n$. The defuzzied of the midpoints of $A_j, A_k,\ \dots\ , A_p$are $m_j,\ m_k,\ \dots,\ m_p$. It can be denoted in the product of the defuzzified matrix and the transpose of the weight matrix:

$$\begin{aligned} F(t)_{old} = \mathbf{M}(t) \times \mathbf{W}(t)^T &= [m_j\ m_k\ \dots\ m_p]\ \times\ [w_1\ w_2\ \dots\ w_n]^T \\ &= [m_j\ m_k\ \dots\ m_p]\ \times\ \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \end{aligned} \tag{11}$$

where the number elements in matrix $\mathbf{M}(t)$ and $\mathbf{W}(t)$ are equal.

From the Eq.(11) the modification can be done with reversal of weight elements in transpose matrix as follows

$$[m_j\ m_k\ \dots\ m_p]\ \times\ \begin{bmatrix} w_n \\ \vdots \\ w_2 \\ w_1 \end{bmatrix} \tag{12}$$

where $\mathbf{M}(t)$ is a 1 x n matrix and $\mathbf{W}(t)^T$ is a n x 1 matrix, respectively.

Let the actual data is $A(t)$, the midpoint interval is m(t) and the absolute difference between actual and midpoint interval is diff $|A(t),\ m(t)|$ then the forecast can be written as below

$$F(t)_{new} = F(t)_{old} \pm \text{diff}|A(t), m(t)|. \tag{13}$$

3.6 The Forecasting Procedures

In this section, a detailed of procedures can be seen in Figure 1.

4 The Verification and Comparisons

The performance of the proposed method will be compared with Yu's method and Cheng's method by using the enrollment of Alabama University from 1972 to 1992. It can be illustrated in Table 4 as below:

Table 4: The Computational of Weighted

Year	Enrollments	Proposed method	Cheng's method	Yu's method
1971	13055	-	-	-
1972	13563	13863.0	13680.5	14250.0
1973	13867	13933.0	13731.3	14250.0
□	□	□	□	□
1992	18876	18876.0	19033.7	19500.0
MSE		16248.7	192084.3	259357.5
Average error of forecasting		0.00496	0.020870	0.144766

From Table 4, the MSE is 16248.7 and average error of forecasting is 0.05% of the proposed method. Further, the MSE of modified weighted method is 41298.4 and average error of forecasting is 1.08%. On the other hand, the MSE of Cheng's method is 192084.3 and average error of forecasting is 2.08%. In addition, Yu's method given the MSE 259357.5 and average error of forecasting is 14.47%. Thus, it is that obvious that the proposed method has a smaller MSE and less average error than the existing methods. The trends in forecast by above mentioned methods are being illustrated in Figure 2 as given below:

Following the steps then performance of UTM enrollment forecasting can be tested and trained by using the proposed method. The forecasting results can be seen in Table 5 as below:

Table 5: The Enrollment Forecasting of Universiti Teknologi Malaysia (UTM) Based on the Proposed Method

Year	Postgraduate	Proposed Method	Years	Undergraduate	Proposed Method
1990	175	-	1990	4158	-
1991	205	260.0	1991	4615	4615.0
1992	305	310.0	1992	5227	5441.3
1993	439	527.2	1993	5383	5331.3
□	□	□	□	□	□
2007	3958	3958.0	2007	16417	16320.9
2008	4721	4721.0	2008	15010	15010.0
MSE		12546.8	MSE		101.7
Average error forecasting		0.074%	Average error forecasting		0.769%

From Table 5 indicates that MSE is 12546.8 and average error forecasting is 0.074% for enrollment forecasting (posgraduate). In addition, MSE is 101.7 and average error

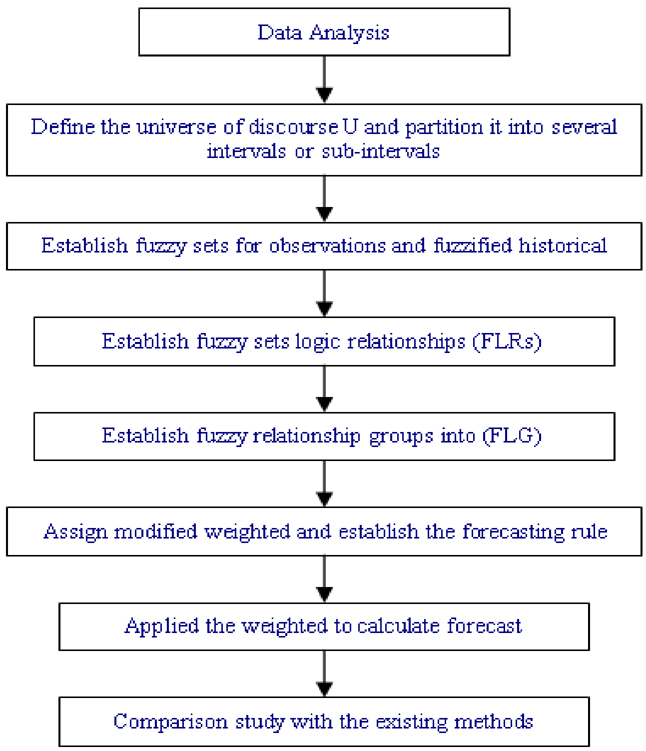


Figure 1: The Forecasting Procedures

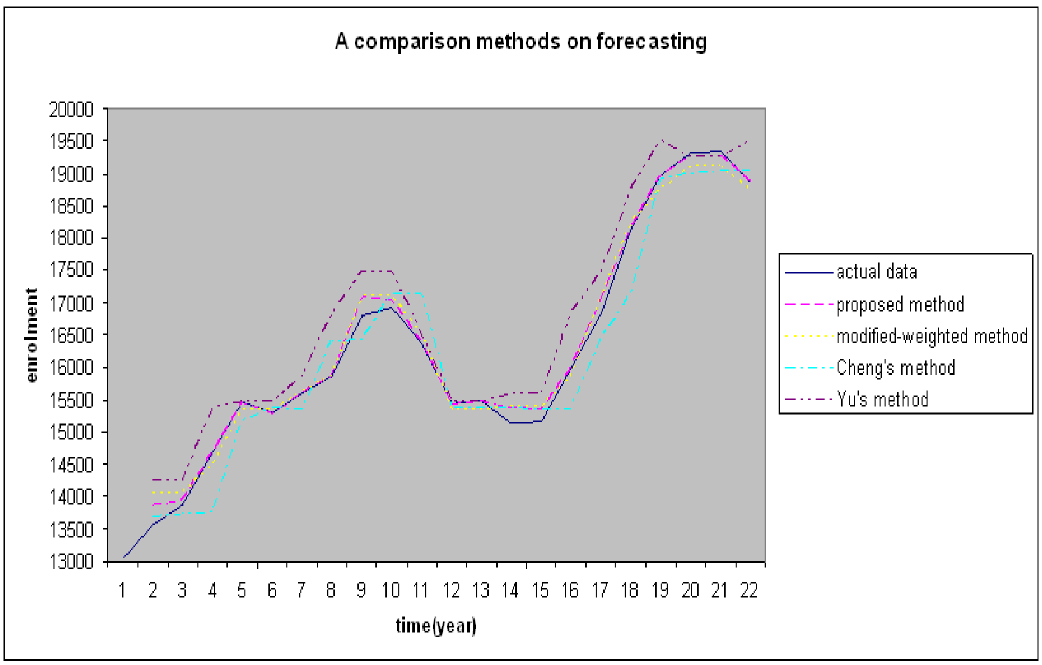
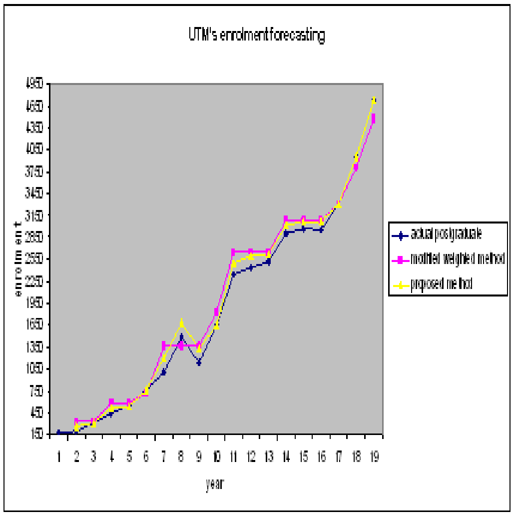


Figure 2: A Comparison Methods on Forecasting

forecasting is 0.769% for enrollment forecasting (undergraduate) by using proposed method. The forecasting results also can be seen in Figure 3 as given below:



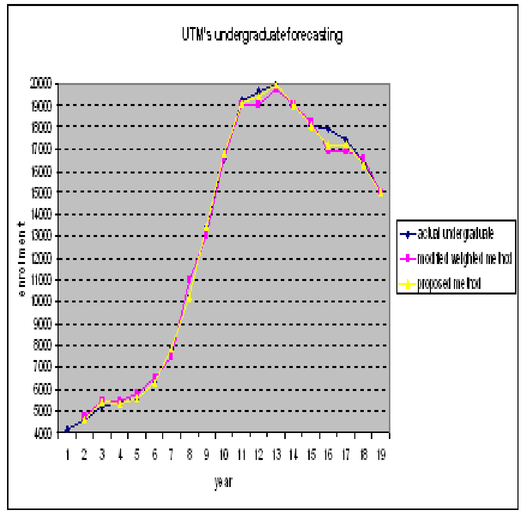


Figure 3: The Enrollment of UTM Forecasting (Undergraduate and Posgraduate)

5 Conclusions

In the proposed method, it has been presented the weighted factors for analyzing fuzzy time series forecasting. The modified weighted are assigned by using the chronological number of FLRs in FLG and a modification is also done in reversal of weight elements on transpose matrix for forecasting rule. In addition, the rules of forecasting have been adopted using the difference of actual data and midpoint interval. The result is better than the existing methods. It can be proven with MSE and average error of forecasting values.

However, the error of forecasting result is also influenced by the length of intervals because of the proposed method is established based on fuzzy logical relationships and the midpoint intervals. Each method has been designed in different form so that no standard rule to be followed.

In the future, the weighted factors will be extended to reach a higher forecasting accuracy rate and investigate a new method which is adapted with multivariate fuzzy time series and out-sample forecasts.

References

[1] Q. Song & B. S. Chissom, *Forecasting Enrollments with Fuzzy Time Series – Part I,* Fuzzy Sets and Systems 54(1993), 1-9.

[2] Q. Song & B. S. Chissom, *Forecasting enrollments with fuzzy time series – Part II,* Fuzzy Sets and Systems 64(1994), 1-8.

[3] Q. Song & B. S. Chissom, *Fuzzy time series and its models,* Fuzzy Sets and Systems 54(1993), 269-277.

[4] J. Sullivan & W.H. Woodall, *A comparison of fuzzy forecasting and Markov modelling,* Fuzzy Sets and Systems 64(1994), 279-293.

[5] R.C. Shaw, *Enrollment forecasting: what methods work the best?* NASSP Bull, 68(1984), 52-58.

[6] D.E. Gardner, *Weight factor selection in double exponential smoothing enrollment forecasts,* Res. Higher Educ., 14(1981), 49-56.

[7] W.C. Weiler, *A model for short-term institutional enrollment forecasting,* J. Higher Education, 51(1980), 314-327.

[8] S.P. Chatman, *Short-term forecasts of the number and scholastic ability of enrolling freshmen by academic divisions,* Res. Higher Educ., 25(1986), 68-81.

[9] J. A. Pope & J.P. Evans, *A forecasting system for college admissions,* College and University, 60 (1985 113-131.

[10] B.J. Warrack & C.N. Russel, *Forecasting demand for postsecondary education in Manitoba: the motivational index and the demand index as an enrollment forecasting tool,* Res. Higher Educ., 19(1983), 335-349.

[11] S.A. Hoenack & W.C. Weiler, *The demand for higher education and institutional enrollment forecasting,* Economic Inquiry, 17(1979), 89-113.

[12] S.M. Chen, *Forecasting enrollments based on fuzzy time series,* Fuzzy Sets and Systems, 81(1996), 311-319.

[13] S.M. Chen, *Temperature prediction using fuzzy time series,* IEEE Transactions on Systems, Man, and Cybernetics, 30(2000), 263-275.

[14] K. Huarng, *Heuristic models of fuzzy time series for forecasting,* Fuzzy Sets and Systems, 123(2001 369-386.

[15] S.M. Chen & C.C. Hsu, *A new method to forecast enrollments using fuzzy time series,* International Journal of Applied Science and Engineering, 3(2004), 234-244.

[16] H-K. Yu, *Weighted fuzzy time series models for Taiex forecasting,* Physica A, 349(2005), 609-624.

[17] M. Sah & Y.D. Konstantin, *Forecasting enrollment model based on first-order fuzzy time series,* Proceeding of World Academy of Science, Eng. and Tech., 1(2005), 375-378.

[18] T.A Jilani & S.M.A. Burney, *A refined fuzzy time series model for stock market forecasting,* Physica A, ScienceDirect, 387(2008) 2857-2862.

[19] C.H. Cheng, T.L. Chen & C.H. Chiang, *Trend-weighted fuzzy time series model for TAIEX forecasting,* ICONIP, 2006, Part III, LNNC 4234, 469-477.

[20] C.H.L.Lee, A.Liu & W.S.Chen, *Pattern Discovery of Fuzzy time series for financial prediction,* IEEE Transactions on Knowledge and data Engineering, 18(2006), 613-625.

[21] K. Huarng, Tiffany H-K Yu & Yu W-S, *A multivariate heuristic model for fuzzy time series forecasting,* IEEE Transactions on Systems, Man, and Cybernetics, 37(2007), 263-275.

[22] S.R. Singh, *A simple method of forecasting on fuzzy time series,* Applied Mathematics and Computation Science Direct, 186(2007), 330-339.

[23] S.R. Singh, *A robust method of forecasting based on fuzzy time series,* Applied Mathematics and Computation ScienceDirect, 188(2007), 472-484.

[24] T.A Jilani & S.M.A. Burney, *A refined fuzzy time series model for stock market forecasting,* Physica A, Science Direct, 387(2008), 2857-2862.

[25] T.H.K. Yu & K.H. Huarng, *A bivariate fuzzy time series model to forecast the TAIEX,* Expert Systems with Application 34(2008), 2945-2952.

[26] I.H. Kuo, S.J. Horng, T.W. Kao, T.L. Lin C.L. Lee & Y. Pan, *An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization,* Expert Systems with Applications, 36(2009), 311-319.

[27] H.H Chu, T.L. Chen, C.H. Cheng & C.C. Huang, *Fuzzy dual-factor time series for stock index forecasting,* Expert Systems with Applications, 36(2009), 165-171.

[28] Zadeh, L.A., Fuzzy Sets, *Information and Control*, 8(1965), 338-353.

[29] Zadeh, L.A., *The concept of a linguistic variable and its application to approximate reasoning - Part I,* Information Science, 8(1975), 199-249.

[30] Zadeh, L.A., *The concept of a linguistic variable and its application to approximate reasoning - Part II,* Information Science, 8(1975), 301-357.

[31] Zadeh, L.A., *The concept of a linguistic variable and its application to approximate reasoning – Part III,* Information Science, 9(1975), 43-80.