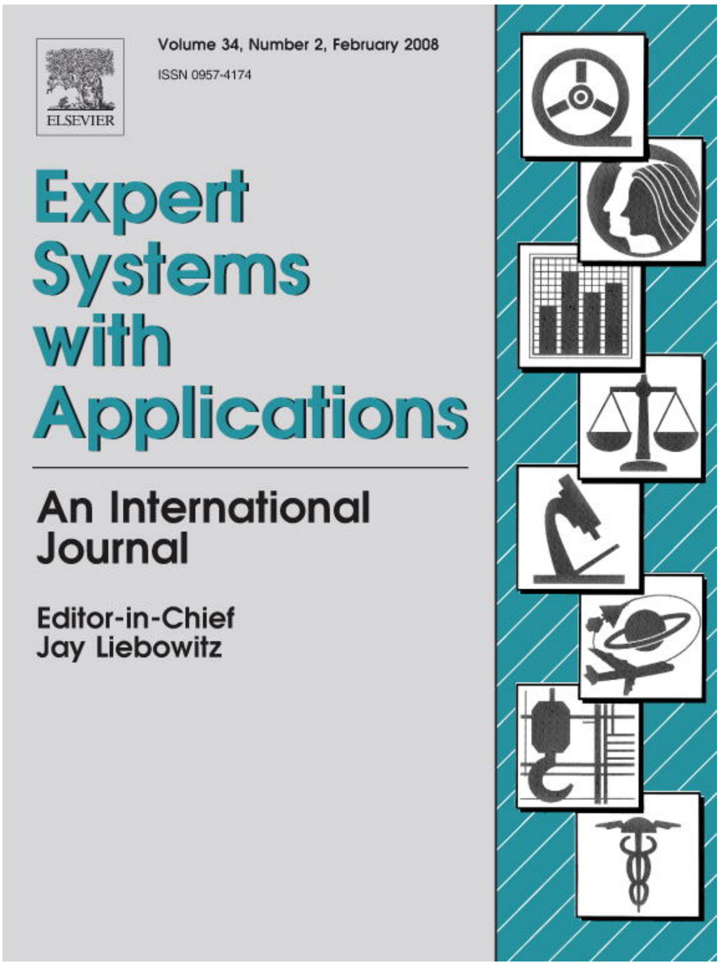
Volume 34, Number 2, February 2008
ISSN 0957-4174
ELSEVIER
Expert Systems with Applications
An International Journal
Editor-in-Chief
Jay Liebowitz

Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 1126–1132
Expert Systems with Applications
www.elsevier.com/locate/eswa

Fuzzy time-series based on adaptive expectation model for TAIEX forecasting

Ching-Hsue Cheng [a], Tai-Liang Chen [a,*], Hia Jong Teoh [a,b], Chen-Han Chiang [a]

[a] Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 64002, Taiwan, ROC

[b] Department of Accounting Information, Ling Tung University 1, Ling Tung Road, Nantun, Taichung 408, Taiwan, ROC

Abstract

Time-series models have been used to make predictions in the areas of stock price forecasting, academic enrollment and weather, etc. However, in stock markets, reasonable investors will modify their forecasts based on recent forecasting errors. Therefore, we propose a new fuzzy time-series model which incorporates the adaptive expectation model into forecasting processes to modify forecasting errors. Using actual trading data from Taiwan Stock Index (TAIEX) and, we evaluate the accuracy of the proposed model by comparing our forecasts with those derived from Chen's [Chen, S. M. (1996). Forecasting enrollments based on fuzzy time-series, *Fuzzy Sets and Systems*, *81*, 311–319] and Yu's [Yu, Hui-Kuang. (2004). Weighted fuzzy time-series models for TAIEX forecasting. *Physica A*, *349*, 609–624] models. The comparison results indicate that our model surpasses in accuracy those suggested by Chen and Yu.

Keywords: Fuzzy time-series model; Adaptive expectation model; TAIEX forecasting; Fuzzy linguistic variable

1. Introduction

Each day individual investors, stock fund managers and financial analysts attempt to predict price activity in stock market on the basis of either their professional knowledge or stock analyzing tools. Higher accuracy is most concerned, because more profit will be made if more accurate predictions are given. So, they have, perennially, strived to discover ways to predict stock price accurately.

For more than one decade, different fuzzy time-series models have also been applied to solve various domain problems, such as financial forecasting (Faff, Brooks, & Kee, 2002; Huarng & Yu, 2005; Shin & Sohn, 2004; Wang, 2002; Yu, 2004), university enrollment forecasting (Chen, 1996; Song & Chissom, 1993b, 1994), temperature forecasting, etc. As Dourra (2002) notes, it is common practice to "deploy fuzzy logic engineering tools in the finance arena, specifically in the technical analysis field, since technical analysis theory consists of indicators used by experts to evaluate stock price (Dourra & Siy, 2002)".

In this paper, a new fuzzy time-series model based on the adaptive expectation model (Kmenta, 1986) is proposed to improve the forecast accuracy in stock market. In this model, several factors such as a proper weight for the fuzzy logic relationship, a reasonable universe of discourse, a reliable length of intervals and past patterns of stock prices are all considered together for forecasting. Moreover, three refined processes are employed in the forecasting algorithm. Using a nine-year period of Taiwan Stock Index (TAIEX) and the enrollment of the University of Alabama as the data sets for verifications, the results demonstrate that the proposed model outperforms the listing fuzzy time-series models.

The remaining content of this paper is organized as follows: Section 2 introduces the related literature of fuzzy

* Corresponding author. Tel.: +886 920975168.
E-mail addresses: chcheng@yuntech.edu.tw (C.-H. Cheng), g9320817@yuntech.edu.tw (T.-L. Chen), g9320824@yuntech.edu.tw (H.J. Teoh), g9323746@yuntech.edu.tw (C.-H. Chiang).

0957-4174/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2006.12.021

time-series model; Section 3, demonstrates the proposed model and algorithm; Section 4 evaluates the model's performance; and Section 5 concludes the paper.

2. Related works

This section briefly reviews the related literature, including two sections: literature reviews of time-series model and fuzzy time-series definitions and algorithm.

2.1. Literature reviews of fuzzy time-series model

Fuzzy theory was originally developed to deal with the problems involving human linguistic terms (Ross, 1995; Zadeh, 1975a, 1975b, 1976). Time-series methods had failed to consider the application of this theory until fuzzy time-series was defined by Song and Chissom (1993a). In 1993, Song and Chissom proposed the definitions of fuzzy time-series and methods to model fuzzy relationships among observations (Song & Chissom, 1993a). In the following research, they continued to discuss the difference between time-invariant and time-variant models (Song & Chissom, 1994; Tsaur, Yang, & Wang, 2005). Besides the above researchers, Chen (1996) proposed another method to applied simplified arithmetic operations in forecasting algorithm rather than the complicated max–min composition operations presented in Song and Chissom (1993a).

However, in time-series model, when unexpected conditions happen, the historical data can not respond the fluctuations immediately. This would probably results in terrible inaccurate forecast. To deal with the problem, a group decision-making method was employed to integrate the subjective forecast values of all decision makers. Fuzzy weighted method was then combined with subjective forecast values to produce the aggregated forecast value.

Besides, Huarng (2001a, 2001b) pointed out that the length of intervals affects forecast accuracy in fuzzy time-series and proposed a method with distribution-based length and average-based length to reconcile this issue (Huarng, 2001a, 2001b). This method applied two different lengths of intervals to Chen's model and the conclusions showed that distribution-based and average-based lengths could improve the accuracy of forecast. Although the forecasting performance of Huarng's method is excellent, it creates too many linguistic values to be identified by analysts. According to Miller (1956), establishing linguistic values and dividing intervals would be a trade off between human recognition and forecasting accuracy (Miller, 1956).

It becomes apparent that the major drawback of these methods is the lack of consideration in determining a reasonable universe of discourse and the length of intervals. Moreover, the researchers find that the neglected information, which indicates the patterns of price changes in history, should be considered in the processes of forecasting. To reconcile these problems above, a new methodology is hereby proposed.

2.2. Fuzzy time-series definitions and algorithm

Over the past 14 years, many fuzzy time-series models have been proposed by following Song and Chissom's definitions (Song & Chissom, 1993a; Yu, 2005). Among these models, Chen's model is very conventional one because of easy calculations and good forecasting performance (Chen, 1996; Hwang, Chen, & Lee, 1998). Therefore, Song and Chissom's definitions and Chen's algorithm are used for illustrations as follows:

Definition 1. Fuzzy time-series. Let $Y(t)(t = \ldots, 0, 1, 2, \ldots)$, a subset of real numbers, be the universe of discourse by which fuzzy sets $f_j(t)$ are defined. If $F(t)$ is a collection of $f_1(t), f_2(t)\ldots$ then $F(t)$ is called a fuzzy time-series defined on $y(t)$.

Definition 2. If there exists a fuzzy logical relationship $R(t-1, t)$, such that $F(t) = F(t-1) \circ R(t-1, t)$, where "$\circ$" represents the max–min composition operator, $F(t-1)$ and $F(t)$ are fuzzy sets, then $F(t)$ is said to be caused by $F(t-1)$. The logical relationship between $F(t)$ and $F(t-1)$ can be represented as: $F(t-1) \to F(t)$.

Definition 3. Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive observations, $F(t)$ and $F(t-1)$, referred to as a fuzzy logical relationship (FLR), can be denoted by $A_i \to A_j$, where A_i is called the left-hand side (LHS) and A_j the right-hand side (RHS) of the FLR.

Definition 4. All fuzzy logical relationships in the training dataset can be further grouped together into different fuzzy logical relationship groups according to the same left-hand sides of the fuzzy logical relationship. For example, there are two fuzzy logical relationships with the same left-hand side (A_i): $A_i \to A_{j1}$ and $A_i \to A_{j2}$. These two fuzzy logical relationships can be grouped into a fuzzy logical relationship group.

Definition 5. Suppose $F(t)$ is caused by $F(t-1)$ only, and $F(t-1) = F(t-1) \times R(t-1, t)$. For any t, if $R(t-1, t)$ is independent of t, then $F(t)$ is named a time-invariant fuzzy time-series, otherwise a time-variant fuzzy time-series.

2.3. The algorithm of Chen's model

Step 1: Define the universe of discourse and intervals for rules abstraction.
Based on the issue domain, the universe of discourse can be defined as: $U = [starting, ending]$. As the length of interval is determined, U can be partitioned into several equal length intervals.

Step 2: Define fuzzy sets based on the universe of discourse and fuzzify the historical data.

Step 3: Fuzzify observed rules.
For example, a datum is fuzzified to A_j if the maximal degree of membership of that datum is in A_j.

Step 4: Establish fuzzy logical relationships and group them based on the current states of the data of the fuzzy logical relationships.
For example, $A_1 \rightarrow A_2$, $A_1 \rightarrow A_1$, $A_1 \rightarrow A_3$, can be grouped as: $A_1 \rightarrow A_1, A_2, A_3$.

Step 5: Forecast.
Let $F(t-1) = A_i$.
Case 1: There is only one fuzzy logical relationship in the fuzzy logical relationship sequence. If $A_i \rightarrow A_j$, then $F(t)$, forecast value, is equal to A_j.
Case 2: If $A_i \rightarrow A_i, A_j, \ldots, A_k$, then $F(t)$, forecast value, is equal to $A_i, A_j, \ldots, A_k$.

Step 6: Defuzzify.
Apply "Centroid" method to get the results. This procedure (also called center of area, center of gravity) is the most often adopted method of defuzzification.

3. Proposed model and algorithm

In this section, we propose a new fuzzy time-series model based on the adaptive expectation model (Kmenta, 1986) (see Fig. 1) and its algorithm. From our review of the literature, there are two major drawbacks: (1) The lack of consideration in determining a reasonable universe of discourse and the length of intervals; (2) Many researchers neglect the information, which indicates patterns of trend changes in the past history. In order to reconcile these problems, three refined processes are factored into the model: (1) To define a reliable length of intervals for linguistic values; (2) To classify recurrent fuzzy relationships into three different types of trends and assign a proper weight to individual fuzzy relationships; (3) To modify the forecasting value with the adaptive expectation model.

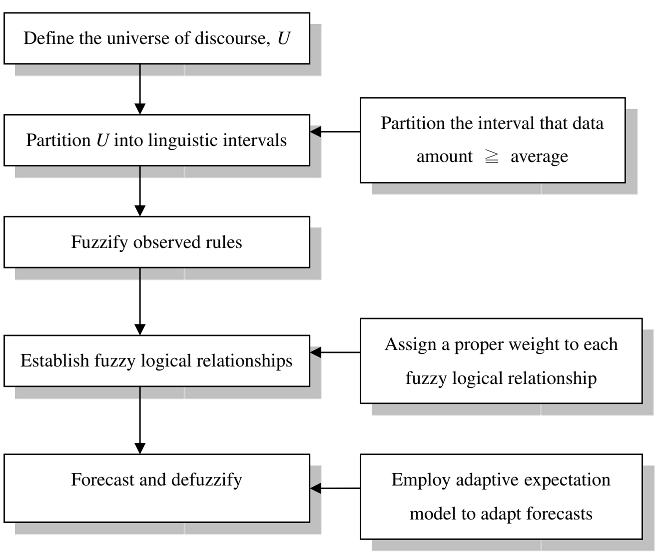

Fig. 1. Research processes of the proposed model.

Initially, in the first refined process, the universe of discourse should be partitioned into seven linguistic values (Miller, 1956), and if the data amount of a given linguistic value is larger than the average amount, then the original linguistic value should be further partitioned in half. Because the data occur more frequently in the linguistic value, using once-divided linguistic value to present the data is supposed to be less reliable than twice-divided. However, it would be undesirable to create too many linguistic values and, thereby, ignoring the meaning of fuzzy application.

Secondly, fuzzy relationship weights are determined either based on knowledge which could be elicited from domain experts or their chronological order. Since each fuzzy relationship will reoccur, from the researcher's perspective, classifying them into different trends and converting the counts of trends to incremental weights is reasonable for making more accurate predictions, hence, the frequency-weighted method. The details describing the assignment of weights are listed in Table 1. For example, it is clear that among the Fuzzy Logical Relationships (FLRs) (Tsaur et al., 2005), when $t = 4$ (t denotes time point), then it is assigned the highest value of 3, which means that the probability of its appearance in the near future is 3 times higher than in any of the other cases. The merits of the trend-weighted model are that they can foresee the cycles and events which will eventually occur and relate the fuzzy relationship in a more reasonable manner.

Thirdly, in stock markets, investors usually make their investment decisions according to recent stock information such as late market news, stock technical indicators, or price fluctuations in these two days. Reasonable investors will modify their forecasts with recent forecasting errors. Based on the fact, we argue that the adaptive expectation model (Kmenta, 1986) will be sutitable to simulate this condition. The formula for this model is defined in the following, where $P(t)$ is the price at time t; $P(t-1)$ is the price at time $t-1$; h_0 is a weighted parameter; $\varepsilon(t-1)$ is the forecasting error at time $t-1$.

$$P(t) = P(t-1) + h_0 * \varepsilon(t-1)$$

In the next subsection, we factor these three refined processes into the proposed algorithm to demonstrate the proposed model.

Table 1
Assign weights to different trends

$(t=1)$ $A_1 \rightarrow A_1$	No change	Assign weight 1
$(t=2)$ $A_2 \rightarrow A_1$	Down	Assign weight 1
$(t=3)$ $A_1 \rightarrow A_1$	No change	Assign weight 2
$(t=4)$ $A_1 \rightarrow A_1$	No change	Assign weight 3
$(t=5)$ $A_1 \rightarrow A_1$	No change	Assign weight 4

t denotes time point.

Step 1: Define the universe of discourse and partition it into intervals.
By the problem used in forecasting, the universe of discourse for observations is defined as: $U = [starting, ending]$. Then the average datum that should be in each linguistic value may be calculated. The linguistic value, which the amount of the data falling in is larger than the average amount of all linguistic values, should further be split into smaller linguistic values by dividing them into two.

Step 2: Establish a related fuzzy set (linguistic value) for each observation in the training dataset.
In this step, the fuzzy sets, $A_1, A_2, \ldots, A_k$, for the universe of discourse are defined by Eq. (1), where the value of a_{ij} indicates the grade of membership of u_j in fuzzy set A_i, where $a_{ij} \in [0,1]$, $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant m$. Ascertain the degree of each stock price belonging to each A_i $(i = 1,\ldots,m)$. If the maximum membership of the stock price is under A_k, then the fuzzified stock price is labeled as A_k (Chen, 1996).

$$\begin{aligned} A_1 &= a_{11}/u_1 + a_{12}/u_2 + \cdots + a_{1m}/u_m \\ A_2 &= a_{21}/u_1 + a_{22}/u_2 + \cdots + a_{2m}/u_m \\ &\vdots \\ A_k &= a_{k1}/u_1 + a_{k2}/u_2 + \cdots + a_{km}/u_m \end{aligned} \tag{1}$$

Step 3: Establish fuzzy relationships.
Two consecutive fuzzy sets $A_i(t-1)$ and $A_j(t)$ can be established into a single FLR as $A_i \to A_j$.

Step 4: Establish fuzzy relationship groups for all FLRs.
The FLRs with the same LHSs (Left Hand Sides) can be grouped to form a FLR Group. For example, $A_i \to A_j$, $A_i \to A_k$, $A_i \to A_m$ can be group as $A_i \to A_j, A_k, A_m$. All FLRs will construct a fluctuation-type matrix.

Step 5: Assign weights.
The matrix from Step 4 is further standardized to W_n, and multiplied by the deffuzified matrix, L_{df}, to produce the forecast value. These weights should be standardized to obtain the weight matrix. The weight matrix, $W(t) = \lfloor W'_1, W'_2, \ldots, W'_j \rfloor$, should be normalized by applying the standardized weight matrix equation (shown in Eq. (2))

$$\begin{aligned} W_n(t) &= [W'_1, W'_2, \ldots, W'_i] \\ &= \left[\frac{W_1}{\sum_{k=1}^{i} W_k}, \frac{W_2}{\sum_{k=1}^{i} W_k}, \ldots, \frac{W_k}{\sum_{k=1}^{i} W_k}\right] \end{aligned} \tag{2}$$

Step 6: Calculate forecast value.
From Step 5, we can obtain the standardized weight matrix, to get the forecast value by using Eq. (3) (where $L_{\mathrm{df}}(t-1)$ is the deffuzified matrix, $W_n(t-1)$ is the weight matrix)

$$F(t) = L_{\mathrm{df}}(t-1) \cdot W_n(t-1) \tag{3}$$

Step 7: Employ the adaptive forecasting equation to produce a conclusive forecast. The adaptive forecasting equation is defined in Eq. (4), where $P(t-1)$ is the actual stock index on time $t-1$, $F(t)$ is the initial forecasting value from Eq. (3), and adaptive_forecast(t) is the conclusive forecasting value for the future stock price (t).

$$\text{Adaptive\_forecast}(t) = P(t-1) + h^*(F(t) - P(t-1)) \tag{4}$$

4. Verifications and comparisons

To verify forecasting performance of the proposed model, the TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index) and the enrollment of the University of Alabama, are selected as verification datasets and two conventional fuzzy times-series models, Chen's (1996) and Yu's (2004), are employed as comparison models.

4.1. Forecasting for TAIEX

In first experiment, a nine-year period of TAIEX data, from 1991 to 1999, is selected as the experimental dataset. Previous ten-month of each year, from January to October, is used for training and the rest, from November to December, for testing (Yu, 2004). To inspect forecasting performance for fuzzy times-series models, we employ the RMSE (root of mean squared error) as a performance indicator (defined in Eq. (5), where actual (t) is the actual stock index on time t; forecast (t) is the forecasting value for actual stock index (t), n is the count of forecasts in the testing period).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (\mathrm{actual}(t) - \mathrm{forecast}(t))^2}{n}} \tag{5}$$

To demonstrate the proposed model, the TAIEX of 2000 is used as a numerical example to detail the proposed algorithm as follows.

Step 1: Defining the universe of discourse and partitioning it into intervals.
According to the TAIEX from 2000/1/4 to 2000/10/31, the universe of discourse for observations (U) is defined as [5000, 10,300]. The results of Miller's (1956) research, necessitates the conclusion of defining seven levels of intervals in order to be more humanly recognizable, which lead us to define our intervals for TAIEX data that is showed in Table 2. Then the average datum, which should be in each linguistic value, is calculated. Those data in the linguistic values U_5, U_6 and U_7, which are larger than the average, should further be split

Table 2
Seven linguistic intervals for the TAIEX of 2000

Linguistic interval	Occurrence
$U_1 = [5000, 5757]$	9
$U_2 = [5757, 6514]$	16
$U_3 = [6514, 7271]$	12
$U_4 = [7271, 8029]$	29
$U_5 = [8029, 8786]$	64
$U_6 = [8786, 9543]$	57
$U_7 = [9543, 10300]$	37

Table 3
Linguistic intervals after dividing process

Linguistic interval	Dividing condition	Occurrence
$U_1 = [5000, 5757]$	First	9
$U_2 = [5757, 6514]$	First	16
$U_3 = [6514, 7271]$	First	12
$U_4 = [7271, 8029]$	First	29
$U_5 = [8029, 8408]$	Second	26
$U_6 = [8408, 8786]$	Second	38
$U_7 = [8786, 9165]$	Second	40
$U_8 = [9165, 9543]$	Second	17
$U_9 = [9543, 9921]$	Second	21
$U_{10} = [9921, 10300]$	Second	16

Table 4
Fuzzy sets for 10 linguistic variables

$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10}$
$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10}$
$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10}$
$A_4 = 0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10}$
$A_5 = 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10}$
$A_6 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6 + 0.5/u_7 + 0/u_8 + 0/u_9 + 0/u_{10}$
$A_7 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0.5/u_6 + 1/u_7 + 0.5/u_8 + 0/u_9 + 0/u_{10}$
$A_8 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0.5/u_7 + 1/u_8 + 0.5/u_9 + 0/u_{10}$
$A_9 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0.5/u_8 + 1/u_9 + 0.5/u_{10}$
$A_{10} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0.5/u_9 + 1/u_{10}$

into smaller linguistic values by dividing them by two. Table 3 shows the results after dividing process.

Step 2: Establish fuzzy sets for observations. Table 4 shows 10 linguistic values A1–A10 by using Eq. (1) and Table 5 shows several example linguistic values for TAIEX.

Step 3: Establishing fuzzy relationships. Take Table 5 as example, the FLRs are demonstrated as follows.

$$A_6 \rightarrow A_7, A_7 \rightarrow A_7, \ldots, A_2 \rightarrow A_1, A_1 \rightarrow A_1$$

Step 4: Establishing fuzzy relationship groups. For example, 10 linguistic values will construct a fluctuation-type matrix (10 * 10) for all FLRs. Each cell in the matrix represents the occurrence of FLR. Table 6 shows an example for 10 linguistic values.

Step 5: Establishing weights. Take Table 6 as example, if $P(t-1) = A_7$, the possible forecasts for $P(t)$ are A_6, A_7 and A_8. After calculating, the weight matrix is [1/5, 3/5, 1/5].

Step 6: Calculating forecast results. The forecasted value can be obtained by matrix multiplication of the defuzzified matrix and weighting matrix as follows:

$$F(t) = L_{\mathrm{df}}(t-1) \circ W_n(t-1)$$
$$= [m_6, m_7, m_8] \circ \left[\frac{1}{5}, \frac{3}{5}, \frac{1}{5}\right]$$

where $L_{\mathrm{df}}(t-1)$ is the deffuzified matrix, $W_n(t-1)$ is the weight matrix.

Step 7: Apply the adaptive forecasting equation, defined in Eq. (4), to adapt the forecast results.

Table 5
Linguistic value time-series for TAIEX

Date	TAIEX	Linguistic value
2000/1/4	8,756.55	A_6
2000/1/5	8,849.87	A_7
2000/1/6	8,922.03	A_7
...	...	...
2000/10/27	5,805.17	A_2
2000/10/30	5,659.08	A_1
2000/10/31	5,544.18	A_1

Table 6
A fluctuation-type matrix for all FLRs

$P(t-1)$	$P(t)$									
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	1	2	0	0	0	0	0	0	0	0
A_2	1	1	1	0	0	0	0	0	0	0
A_3	0	0	1	1	1	0	0	0	0	0
A_4	0	0	0	1	1	1	0	0	0	0
A_5	0	0	0	0	0	1	1	1	0	0
A_6	0	0	0	0	0	0	1	1	0	0
A_7	*0*	*0*	*0*	*0*	*0*	*1*	*3*	*1*	*0*	*0*
A_8	0	0	0	0	0	0	0	1	1	0
A_9	0	0	0	0	0	0	0	1	1	1
A_{10}	0	0	0	0	0	0	0	0	1	2

With seven linguistic values, we generate forecasting values for the TAIEX in nine testing periods. From the literature (Yu, 2004), a performance comparison table (see Table 7) and figure (see Fig. 2) are produced to illustrate

Table 7
Performance comparisons for TAIEX

Models	1991	1992	1993	1994	1995	1996	1997	1998	1999	Average
Chen's model (1996)	80	60	110	112	79	54	148	167	149	107
Yu's model (2004)	61	67	105[a]	135	70	54	133[a]	151	142	102
Proposed model	42[a]	43[a]	105[a]	75[a]	53[a]	51[a]	134	113[a]	109[a]	81[a]

[a] Best performance among 3 approaches.

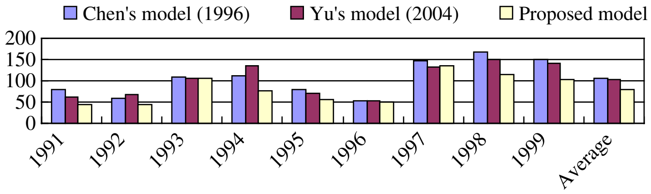

Fig. 2. Performance comparison charts for TAIEX.

Table 8
Enrollment linguistic values

Lingustic	Intervals
U_1	[13,000, 15,000]
U_2	[13,000, 16,000]
U_3	[14,000, 16,500]
U_4	[14,500, 17,000]
U_5	[15,000, 17,500]
U_6	[15,500, 18,000]
U_7	[16,000, 19,000]
U_8	[17,000, 20,000]
U_9	[18,000, 20,000]

the forecasting performances of the three different models. From Table 7, the proposed model bears the smallest RMSE in nine testing period except 1997. It is obvious that our model surpass Chen's (1996) and Yu's (2004) in forecasting performance. Based on the excellent performance, another experiment using different dataset is warranted.

Table 9
Linguistic values for the enrollments of the University of Alabama

Year	Enrollment	Linguistic value	Year	Enrollment	Linguistic value
1971	13055	A_1	1982	15433	A_3
1972	13563	A_1	1983	15497	A_3
1973	13867	A_1	1984	15145	A_3
1974	14696	A_2	1985	15163	A_3
1975	15460	A_3	1986	15984	A_4
1976	15311	A_3	1987	16859	A_6
1977	15603	A_4	1988	18150	A_8
1978	15861	A_4	1989	18970	A_8
1979	16807	A_6	1990	19328	A_9
1980	16919	A_6	1991	19337	A_9
1981	16388	A_5	1992	18876	A_8

4.2. Forecasting for University enrollment

Besides the TAIEX data set, the enrollments of the University of Alabama are also used to illustrate how the researchers' model outperforms others, since they have been used as the forecasting target in many of the fuzzy time-series forecasting studies. As in this case, the universe of discourse is rendered in Steps 1–6, below, the linguistic intervals for enrollment is showed in Table 8, linguistic values for each year of enrollment is showed in Table 9, and the comparison results with different model, the actual enrollment, the forecasts from Chen's model, Song and Chissom's, Sullivan and Woodall's, Lee's and the forecasts based on the proposed model, are listed in Table 10. Fig. 3 shows the actual and forecasting data for enrollments. The

Table 10
Comparisons of the forecast results for enrollment with different models

Year	Actual enrollment	Song and Chissom (1993a, 1993b)	Sullivan and Woodall (1994)	Chen (1996)	Lee (1996) ($w = 2$)	Proposed model
1971	13,055					
1972	13,563	14,000	13,500	14,000		13680.5
1973	13,867	14,000	14,500	14,000	13,860	13731.3
1974	14,696	14,000	14,500	14,000	13,964	13761.7
1975	15,460	15,500	15231	15,500	14,710	15194.6
1976	15,311	16,000	15563	16,000	15,452	15374.8
1977	15,603	16,000	15,500	16,000	15,311	15359.9
1978	15,861	16,000	15,500	16,000	15,603	16410.3
1979	16,807	16,000	15,500	16,000	15,861	16436.1
1980	16,919	16,813	16684	16,833	16,830	17130.7
1981	16,388	16,813	16684	16,833	16,919	17141.9
1982	15,433	16,789	15,500	16,833	16,388	15363.8
1983	15,497	16,000	15563	16,000	15,417	15372.1
1984	15,145	16,000	15563	16,000	15,497	15378.5
1985	15,163	16,000	15563	16,000	15,145	15343.3
1986	15,984	16,000	15563	16,000	15,163	15345.1
1987	16,859	16,000	15,500	16,000	15,984	16448.4
1988	18,150	16,813	16577	16,833	16,862	17135.9
1989	18,970	19,000	19,500	19,000	18,122	18915.0
1990	19,328	19,000	19,500	19,000	18,970	18997.0
1991	19,337	19,000	19,500	19,000	19,091	19032.8
1992	18,876	*	*	19,000	19,101	19033.7
Average error		3.22%	2.66%	3.11%	2.95%	2.087%
RMSE		650	621	638	615	438

Note: * denotes No answer.

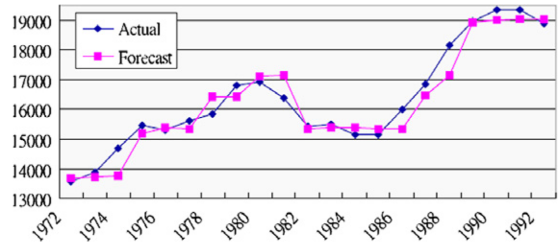

Fig. 3. Actual and forecasting results for the enrollment.

algorithm of this example is the same as that used in forecasting the TAIEX.

From Table 10, the average forecasting error in Song and Chissom's model is 3.22%, with the RMSE of 650; in the Sullivan and Woodall model, the average error is 2.66% with the RMSE of 621; in Chen's model, the average error is 3.11% with the RMSE of 638; and in Lee's model, the average error is 2.95% with the RMSE of 615. Based on the results, the proposed model has a smaller RMSE and less average error than the other models.

5. Conclusions and future research

In stock markets, investors make their decisions not only historical price patterns but also recent price fluctuations. Conventional fuzzy time-series models focus on mining fuzzy logical relationships from time-series which ignore the correlation between recent prices. Therefore, we propose a new fuzzy time-series model, which can modify the forecasts based on recent price changes, to puzzle out this problem. Based on the comparison results in the case of TAIEX, we conclude that, by incorporating the adaptive expectation model into forecasting process, the proposed model can improve the performance of time-series model for stock price forecasting. Besides, the refined process employed in our model, which assingns a proper weight based on occurrence frequency to each fuzzy logical relationship, is a more reasonable approach than Chen's (1996) model which gives an equal weight to all fuzzy logical relationships. Although the proposed model outperforms the listing models, there is room for testing and improving the hypothesis of this model as follows:

1. Using other individual stocks and financial materials as data sets to evaluate the performance of accuracy.
2. Simulating the trading by the proposed method in actual stock market, and sum up the profits of these tradings to evaluate the performance of profit making.
3. Reconsidering the factors affecting the behavior of the stock markets, such as trading volume, news and finical reports which might impact it in the future.

References

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

Dourra, Hussein, & Siy, Pepe (2002). Investment using technical analysis and fuzzy logic. *Fuzzy Sets and Systems, 127*, 221–240.

Faff, R. W., Brooks, R. D., & Kee, Ho Yew (2002). New evidence on the impact of financial leverage on beta risk: a time-series approach. *North American Journal of Economics and Finance, 13*, 1–20.

Huarng, Kunhuang (2001a). Effective lengths of intervals to improve forecasting in fuzzy time-series. *Fuzzy Sets and Systems, 123*, 387–394.

Huarng, K. (2001b). Heuristic models of fuzzy time-series for forecasting. *Fuzzy Sets and Systems, 123*, 137–154.

Huarng, Kunhuang, & Yu, Hui-Kuang (2005). A Type 2 fuzzy time-series model for stock index forecasting. *Physica A, 353*, 445–462.

Hwang, J. R., Chen, S. M., & Lee, C. H. (1998). Handling forecasting problems using fuzzy time-series. *Fuzzy Sets Systems, 100*, 217–228.

Kmenta, J. (1986). *Elements of econometrics*. MacMillan.

Miller, George A. (1956). The magical number seven, plus or minus two: some limits on our capacity of processing information. *The Psychological Review, 63*, 81–97.

Ross, Timothy J. (1995). *Fuzzy logic with engineering applications*. New York: McGraw-Hill.

Shin, Hyungwon, & Sohn, So Young (2004). Segmentation of stock trading customers accord ing to potential value. *Expert Systems with Applications, 27*, 27–33.

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

Song, Q., & Chissom, B. S. (1993b). Forecasting enrollments with fuzzy time-series – Part I. *Fuzzy Sets and Systems, 54*, 1–10.

Song, Q., & Chissom, B. S. (1994). Forecasting enrollments with fuzzy time-series – Part II. *Fuzzy Sets and Systems, 62*, 1–8.

Tsaur, Ruey-Chyn, Yang, Jia-Chi O., & Wang, Hsiao-Fan (2005). Fuzzy relation analysis in fuzzy time-series model. *Computers and Mathematics with Applications, 49*, 539–548.

Wang, Y. F. (2002). Predicting stock price using fuzzy grey prediction system. *Experts Systems with Applications, 22*, 33–39.

Yu, Hui-Kuang (2004). Weighted fuzzy time-series models for TAIEX forecasting. *Physica A, 349*, 609–624.

Yu, Hui-Kuang (2005). A refined fuzzy time-series model for forecasting. *Physica A, 346*, 657–681.

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

Zadeh, L. A. (1975b). The concept of a linguistic variable and its application to approximate reasoning II. *Information Science, 8*, 301–357.

Zadeh, L. A. (1976). The concept of a linguistic variable and its application to approximate reasoning III. *Information Science, 9*, 43–80.