Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

A New Method to Forecast the TAIEX Based on Fuzzy Time Series

Chao-Dian Chen[1] and Shyi-Ming Chen[1, 2]
[1]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R. O. C.
[2]Department of Computer Science and Information Engineering, Jinwen University of Science and Technology, Taipei County, Taiwan, R. O. C.

***Abstract*—In this paper, we present a new method to forecast the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) based on fuzzy time series, where the main factor is the TAIEX and the secondary factors are either the Dow Jones, the NASDAQ, the M_{1b} (Taiwan), or their combinations. First, we fuzzify the historical data of the main factor into fuzzy sets with a fixed length of intervals to form fuzzy logical relationships. Then, we group the fuzzy logical relationships into fuzzy logical relationship groups. Then, we evaluate the leverage of fuzzy variations between the main factor and the secondary factor to forecast the TAIEX. The experimental results show that the proposed method gets a higher average forecasting accuracy rate than Chen's method [1] and Huarng et al.'s method [9] to forecast the TAIEX.**

***Keywords*—fuzzy sets, fuzzy time series, fuzzy logical relationships, fuzzy variation**

I. Introduction

In [11], [12] and [13], Song and Chissom presented the concepts of fuzzy time series based on the fuzzy set theory [21], where the values of a fuzzy time series are represented by fuzzy sets. In recent years, some methods have been presented to handle forecasting problems based on fuzzy time series, such as enrollments forecasting [1], [2], [3], [4], [12], [13], [15], [19], temperature prediction [5], [10], [14], stock index forecasting [6], [7], [8], [9], [14], [16], [17], [18], [20] , …, etc.

In this paper, we present a new method to forecast the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) based on fuzzy time series, where the main factor is the TAIEX and the secondary factors are either the Dow Jones, the NASDAQ, the M_{1b} (Taiwan), or their combinations. First, we fuzzify the historical data of the main factor into fuzzy sets with a fixed length of intervals to form fuzzy logical relationships. Then, we group the fuzzy logical relationships into "fuzzy logical relationships groups". Then, we evaluate the leverage of fuzzy variations between the main factor and the secondary factor to forecast the TAIEX. The experimental results show that the proposed method gets a higher average forecasting accuracy rate than Chen's method [1] and Huarng et al.'s method [9] to forecast the TAIEX.

The rest of this paper is organized as follows. In Section II, we briefly review the definition of fuzzy time series from [11], [12] and [13]. In Section III, we present a new method based on fuzzy time series to forecast the TAIEX. In Section IV, we make a comparison of the experimental results of the proposed method with the existing methods. The conclusions are discussed in Section V.

II. Preliminaries

In [11], [12] and [13], Song and Chissom presented the concepts of fuzzy time series based on the fuzzy set theory [21], where the values of a fuzzy time series are represented by fuzzy sets. Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_n\}$. A fuzzy set A_i in the universe of discourse U is defined as follows:

$$A_i = f_{Ai}(u_1)/u_1 + f_{Ai}(u_2)/u_2 + \ldots + f_{Ai}(u_n)/u_n,$$

where f_{Ai} is the membership function of the fuzzy set A_i, $f_{Ai}(u_j)$ is the degree of membership of u_j belonging to the fuzzy set A_i, $f_{Ai}(u_j) \in [0,1]$ and $1 \leq j \leq n$.

***Definition 2.1* [11]:** Let $Y(t)$ ($t = \ldots, 0, 1, 2, \ldots$) be the universe of discourse and be a subset of R. Assume that $f_i(t)$ ($i = 1, 2, \ldots$) are defined in the universe of discourse $Y(t)$, and assume that $F(t)$ is a collection of $f_i(t)$ ($i = 1, 2, \ldots$), then $F(t)$ is called a fuzzy time series of $Y(t)$ ($t = \ldots, 0, 1, 2, \ldots$).

If a fuzzy relationships $R(t-1,t)$ exists, such that $F(t) = F(t-1) \circ R(t-1,t)$, where the symbol "$\circ$" represents the max−min composition operator, then $F(t)$ is called caused by $F(t-1)$ [11].

***Definition 2.2* [11]:** Let $F(t-1) = A_i$ and let $F(t) = A_j$. The relationship between $F(t-1)$ and $F(t)$ can be denoted by fuzzy logical relationship $A_i \rightarrow A_j$, where A_i is called the left-hand side (LHS) and A_j is called the right-hand side (RHS) of the fuzzy logical relationship.

Fuzzy logical relationships having the same left-hand side can be grouped into a fuzzy logical relationship group (FLRG) [1]. For example, assume that the following fuzzy logical relationships exist:

$$A_i \rightarrow A_{ja},$$
$$A_i \rightarrow A_{jb},$$
$$\vdots$$
$$A_i \rightarrow A_{jm}.$$

then these fuzzy logical relationships can be grouped into a fuzzy logical relationship group, shown as follows:

$$A_i \rightarrow A_{ja}, A_{jb}, \ldots, A_{jm}.$$

III. A New Method for Forecasting the TAIEX Based on Fuzzy Time Series

In this section, we present a new method to forecast the TAIEX from 2000 to 2004 based on fuzzy time series, where

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

the historical data are divided into two parts, i.e., the training data set and the testing data set. The training data set consists of the historical data from January to October for each year, and the testing data set consists of the historical data from November to December for each year. Table I [9] shows the TAIEX, the Dow Jones, the NASDAQ, and the M_{1b} from January 2004 to December 2004. In this paper, the TAIEX is the main factor, where the secondary factors Dow Jones, NASDAQ and M_{1b} are used to forecast the TAIEX. The proposed method is now presented as follows:

Step 1: Define the universe of discourse U, $U = [D_{min}–D_1, D_{max}+D_2]$, where D_{min} and D_{max} are the minimum and the maximum values of the historical data of the main factor, respectively; D_1 and D_2 are two proper positive real values to partition the universe of discourse U into n intervals u_1, u_2, …, and u_n of equal length. For example, from Table I, we can see that the minimum and the maximum values of the training data of the TAIEX of the year 2004 are 5316.87 and 7034.1, respectively. If we let D_1 = 16.87 and D_2 = 65.9, then the universe of discourse U = [5300, 7100]. Let the length of each interval in the universe of discourse U be 100. Then, the universe of discourse U can be divided into 18 intervals, which is defined as follows:

$$u_i = [5300 + (i - 1) \times 100, 5300 + (i) \times 100], \quad (1)$$

where i = 1, 2, …, 18.

TABLE I. HISTORICAL DATA OF THE TAIEX, THE DOW JONES, THE NASDAQ, AND THE M1B OF 2004 [9]

Date	TAIEX	Dow Jones	NASDAQ	M_{1b}
2004/1/2	6041.56	10409.85	2006.68	6491205
2004/1/5	6125.42	10544.07	2047.36	6487349
2004/1/6	6144.01	10538.66	2057.37	6497906
⋮	⋮	⋮	⋮	⋮
2004/11/1	5656.17	10054.39	1979.87	7040767
2004/11/2	5759.61	10035.73	1984.79	7054955
⋮	⋮	⋮	⋮	⋮
2004/12/31	6139.69	10783.01	2175.44	7370450

Step 2: Define the linguistic terms A_i represented by fuzzy sets, shown as follows:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \ldots + 0/u_{n-2} + 0/u_{n-1} + 0/u_n,$$
$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + \ldots + 0/u_{n-2} + 0/u_{n-1} + 0/u_n,$$
$$\vdots$$
$$A_n = 0/u_1 + 0/u_2 + 0/u_3 + \ldots + 0/u_{n-2} + 0.5/u_{n-1} + 1/u_n.$$

where A_1, A_2, …, and A_n are linguistic terms. For example, based on the obtained 18 intervals, we can define the linguistic terms A_1, A_2, …, and A_{18}, shown as follows:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \ldots + 0/u_{16} + 0/u_{17} + 0/u_{18},$$
$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + \ldots + 0/u_{16} + 0/u_{17} + 0/u_{18},$$
$$\vdots$$
$$A_{18} = 0/u_1 + 0/u_2 + 0/u_3 + \ldots + 0/u_{16} + 0.5/u_{17} + 1/u_{18}.$$

Step 3: Fuzzify each historical datum of the main factor into a fuzzy set defined in **Step 2**. If the historical datum of the main factor belongs to u_i and the maximum membership value of the fuzzy set A_i occurs at u_i, where $1 \le i \le n$, then the historical datum of the main factor is fuzzified into A_i. For example, from Table I, we can see that the TAIEX of 2004/1/2 is 6041.56, which can be fuzzified into A_8. Table II shows the fuzzified TAIEX of the data shown in Table I, respectively.

Step 4: Construct fuzzy logical relationships from the fuzzified historical data of the main factor obtained in **Step 3**. For example, from the fuzzified TAIEX of the training data shown in Table II, we can construct fuzzy logical relationships. For example, because the fuzzified TAIEX of 2004/1/2 is A_8 and because the fuzzified TAIEX of 2004/1/5 is A_9, we can construct the following fuzzy logical relationship:

$$A_8 \rightarrow A_9.$$

In the same way, based on Table II, we can get the fuzzy logical relationships as shown in Table III.

TABLE II. FUZZIFIED TAIEX

Date	Fuzzy Set
2004/1/2	A_8
2004/1/5	A_9
2004/1/6	A_9
⋮	⋮
2004/11/1	A_4
2004/11/2	A_5
⋮	⋮
2004/12/31	A_9

TABLE III. FIRST-ORDER FUZZY LOGICAL RELATIONSHIPS

Date	Fuzzy Logical Relationships
2004/1/2 → 2004/1/5	$A_8 \rightarrow A_9$
2004/1/5 → 2004/1/6	$A_9 \rightarrow A_9$
⋮	⋮
2004/10/28 → 2004/10/29	$A_4 \rightarrow A_5$

Step 5: Fuzzify the variation between the adjacent historical data of the main factor and the secondary factors, respectively, and then group the fuzzy logical relationships of the main factor. The sub-steps are shown as follows:

Step 5.1: Calculate the variation of the close index between the adjacent historical data, where the variation Var_t on day t is calculated as follows:

$$Var_t = \frac{Close_t - Close_{t-1}}{Close_{t-1}} \times 100\% , \quad (2)$$

where the terms $Close_t$ and $Close_{t-1}$ are the close indices on the trading day t and trading day $t-1$, respectively, and the unit of the variation is the percentage. For example, from Table I, we can see that the TAIEX of 2004/1/2 and 2004/1/5 are 6041.56 and 6125.42, respectively. Based on Eq. (2), we can see that the variation of the TAIEX on 2004/1/5 is equal to $\frac{6125.42 - 6041.56}{6041.56} \times 100\% = 1.388052\ \%$. Table IV and Table V show the variation of the TAIEX, the Dow Jones, the NASDAQ and the M_{1b} of the training data and the testing data.

TABLE IV. THE VARIATION OF THE TRAINING DATA OF TAIEX, THE DOW JONES, THE NASDAQ, AND THE M1B (UNIT : %)

Date	Variation of the TAIEX	Variation of the Dow Jones	Variation of the NASDAQ	Variation of the M_{1b}
2004/1/5	1.388052 %	1.289356 %	2.027229 %	-0.059403 %
2004/1/6	0.303489 %	-0.051308 %	0.488922 %	0.162732 %
⋮	⋮	⋮	⋮	⋮
2004/10/29	0.182072 %	0.229196 %	-0.037960 %	0.639198 %

TABLE V. THE VARIATION OF THE TESTING DATA OF TAIEX, THE DOW JONES, THE NASDAQ AND THE M1B (UNIT : %)

Date	Variation of the TAIEX	Variation of the Dow Jones	Variation of the NASDAQ	Variation of the M_{1b}
2004/11/1	-0.872075 %	0.268463 %	0.247090 %	-0.348530 %
2004/11/2	1.828799 %	-0.185591 %	0.248501 %	0.201512 %
⋮	⋮	⋮	⋮	⋮
2004/12/30	0.203170 %	-0.266779 %	0.061553 %	0.979241 %

Step 5.2: Define the universe of discourse V, $V = [E_{min}, E_{max}]$, where E_{min} and E_{max} are the minimum and the maximum variation of the main factor and secondary factors, respectively. It should be noted that the minimum and the maximum limited variations of the TAIEX are -7% and 7%, respectively, and the minimum and the maximum limited variations of the Dow Jones, the NASDAQ and the M_{1b} are not limited, respectively. The universe of discourse V is defined as $[-\infty, \infty]$. We let the length of each interval between -6% and 6% be equal to 1. Then, the universe of discourse V can be divided into 14 intervals v_1, v_2, …, and v_{14}, where $v_1 = [-\infty, -6]$, $v_2 = [-6, -5]$, …, and $v_{14} = [6, \infty]$, as shown in Table VI.

TABLE VI. 14 INTERVALS IN THE UNIVERSE OF DISCOURSE V (UNIT : %)

$v_1 = [-\infty, -6]$	$v_4 = [-4, -3]$	$v_7 = [-1, 0]$	$v_{10} = [2, 3]$	$v_{13} = [5, 6]$
$v_2 = [-6, -5]$	$v_5 = [-3, -2]$	$v_8 = [0, 1]$	$v_{11} = [3, 4]$	$v_{14} = [6, \infty]$
$v_3 = [-5, -4]$	$v_6 = [-2, -1]$	$v_9 = [1, 2]$	$v_{12} = [4, 5]$	

Step 5.3: Define the linguistic term B_j represented by fuzzy sets, shown as follows:

$$B_1 = 1/v_1 + 0.5/v_2 + 0/v_3 + \ldots + 0/v_{m-2} + 0/v_{m-1} + 0/v_m,$$
$$B_2 = 0.5/v_1 + 1/v_2 + 0.5/v_3 + \ldots + 0/v_{m-2} + 0/v_{m-1} + 0/v_m,$$
$$\vdots$$
$$B_m = 0/v_1 + 0/v_2 + 0/v_3 + \ldots + 0/v_{m-2} + 0.5/v_{m-1} + 1/v_m.$$

where B_1, B_2, …, and B_m are linguistic terms. For example, based on Table VI, we can define the linguistic terms B_1, B_2, …, and B_{14}, shown as follows:

$$B_1 = 1/v_1 + 0.5/v_2 + 0/v_3 + \ldots + 0/v_{12} + 0/v_{13} + 0/v_{14},$$
$$B_2 = 0.5/v_1 + 1/v_2 + 0.5/v_3 + \ldots + 0/v_{12} + 0/v_{13} + 0/v_{14},$$
$$\vdots$$
$$B_{14} = 0/v_1 + 0/v_2 + 0/v_3 + \ldots + 0/v_{12} + 0.5/v_{13} + 1/v_{14}.$$

Step 5.4: Fuzzify each historical variation of the main factor into a fuzzy set defined in **Step 5.2**. If the historical variation of the main factor on day t belongs to v_j, where $1 \leq j \leq m$, then the historical variation of the main factor on day t is fuzzified into B_j. For example, from Table IV, we can see that the variation of the TAIEX on 2004/1/5 is 1.388052%, which is fuzzified into B_9. Table VII shows the fuzzified variation of the TAIEX of the training data.

TABLE VII. FUZZIFIED VARIATION OF THE TAIEX

Date	Fuzzy Set
2004/1/5	B_9
2004/1/6	B_8
2004/1/7	B_7
⋮	⋮
2004/10/29	B_8

Step 5.5: Based on the linguistic terms of the fuzzified variations, group the fuzzy logical relationships having the same linguistic term of the fuzzified variation into a fuzzy logical relationship group. For example, let us consider the following fuzzy logical relationships: "$A_{a1} \rightarrow A_{aa}$", where the fuzzified variation is B_z; "$A_{a1} \rightarrow A_{ab}$", where the fuzzified variation is B_z; "$A_{a1} \rightarrow A_{ac}$", where the fuzzified variation is B_z; …; "$A_{a1} \rightarrow A_{ak}$", where the fuzzified variation is B_z. Then, these fuzzy logical relationships can be grouped into the same fuzzy logical relationship group in which the fuzzified variation of each fuzzy logical relationship in this group is B_z, shown as follows:

$$A_{a1} \rightarrow A_{aa}, A_{ab}, A_{ac}, \ldots, A_{ak}.$$

For example, from Table III, we can see that the fuzzy logical relationship between the trading days 2004/1/2 and 2004/1/5 is $A_8 \rightarrow A_9$, from Table VII, we can see that the fuzzified variation on 2004/1/5 is B_9; so we can group the fuzzy logical relationship "$A_8 \rightarrow A_9$" into the fuzzified variation B_9 group. Table VIII shows the fuzzy logical relationship groups with respect to different fuzzified variations, respectively.

TABLE VIII. FUZZY LOGICAL RELATIONSHIP GROUPS

B_1 Group	$A_{16} \rightarrow A_{11}$	
B_2 Group	$A_5 \rightarrow A_2$	$A_9 \rightarrow A_6$
B_3 Group	$A_{12} \rightarrow A_9$	
B_4 Group	$A_6 \rightarrow A_4$	$A_8 \rightarrow A_6$
B_5 Group	$A_2 \rightarrow A_1$	$A_9 \rightarrow A_7$
	$A_5 \rightarrow A_3$	$A_{11} \rightarrow A_9$
	$A_6 \rightarrow A_5$	$A_{13} \rightarrow A_{12}$
	$A_7 \rightarrow A_5, A_6$	$A_{16} \rightarrow A_{14}, A_{15}$
B_6 Group	$A_2 \rightarrow A_1$	$A_9 \rightarrow A_8$
	$A_4 \rightarrow A_3, A_3, A_4, A_3$	$A_{11} \rightarrow A_{11}, A_{10}$
	$A_5 \rightarrow A_4, A_4, A_4$	$A_{14} \rightarrow A_{13}, A_{13}$
	$A_6 \rightarrow A_5, A_5, A_5, A_6$	$A_{16} \rightarrow A_{16}, A_{15}$
	$A_7 \rightarrow A_6, A_6$	$A_{17} \rightarrow A_{16}$
	$A_8 \rightarrow A_7$	$A_{18} \rightarrow A_{17}$
B_7 Group	$A_1 \rightarrow A_1, A_1, A_1, A_1, A_1, A_1, A_1, A_1, A_1$	$A_{10} \rightarrow A_{10}, A_{10}, A_{10}, A_{10}, A_9$
	$A_2 \rightarrow A_1, A_1$	$A_{11} \rightarrow A_{11}$
	$A_3 \rightarrow A_3, A_3, A_2$	$A_{12} \rightarrow A_{12}, A_{12}$
	$A_4 \rightarrow A_4$	$A_{13} \rightarrow A_{13}$
	$A_5 \rightarrow A_5, A_5, A_5, A_5, A_5, A_5$	$A_{14} \rightarrow A_{14}, A_{13}, A_{14}, A_{14}$
	$A_6 \rightarrow A_6, A_6, A_5, A_6, A_6, A_6, A_6, A_6, A_6, A_5, A_5$	$A_{15} \rightarrow A_{15}, A_{14}$
	$A_7 \rightarrow A_7, A_7, A_7, A_7, A_6, A_7, A_6, A_7$	$A_{16} \rightarrow A_{15}$
	$A_8 \rightarrow A_8$	$A_{17} \rightarrow A_{17}, A_{17}$
	$A_9 \rightarrow A_9, A_9, A_9, A_8$	
B_8 Group	$A_1 \rightarrow A_1, A_1, A_1, A_1$	$A_{10} \rightarrow A_{10}, A_{10}$
	$A_4 \rightarrow A_5, A_4, A_4, A_4, A_5$	$A_{11} \rightarrow A_{11}, A_{11}$
	$A_5 \rightarrow A_5, A_6, A_5, A_5, A_6, A_5$	$A_{12} \rightarrow A_{12}, A_{12}, A_{13}$
	$A_6 \rightarrow A_7, A_6, A_6, A_6, A_6$	$A_{13} \rightarrow A_{13}, A_{14}, A_{14}, A_{13}, A_{13}$
	$A_7 \rightarrow A_7, A_7, A_7, A_7$	$A_{14} \rightarrow A_{14}, A_{14}, A_{14}, A_{15}, A_{14}, A_{14}$
	$A_8 \rightarrow A_8, A_8, A_9$	$A_{15} \rightarrow A_{16}, A_{15}, A_{15}, A_{16}, A_{15}$
	$A_9 \rightarrow A_9, A_9, A_{10}, A_{10}$	$A_{16} \rightarrow A_{16}$
B_9 Group	$A_1 \rightarrow A_2, A_1, A_2, A_2$	$A_{10} \rightarrow A_{10}, A_{11}, A_{11}$
	$A_2 \rightarrow A_3$	$A_{11} \rightarrow A_{12}$
	$A_3 \rightarrow A_4, A_4$	$A_{12} \rightarrow A_{13}$
	$A_4 \rightarrow A_5$	$A_{14} \rightarrow A_{14}$
	$A_5 \rightarrow A_6, A_5, A_6, A_6$	$A_{15} \rightarrow A_{16}, A_{16}$
	$A_6 \rightarrow A_6, A_7, A_7, A_7, A_7$	$A_{16} \rightarrow A_{17}$
	$A_7 \rightarrow A_8$	$A_{17} \rightarrow A_{18}, A_{17}$
	$A_8 \rightarrow A_9, A_9$	
B_{10} Group	$A_1 \rightarrow A_2$	$A_8 \rightarrow A_9$
	$A_4 \rightarrow A_6$	$A_{13} \rightarrow A_{14}$
	$A_6 \rightarrow A_7$	$A_{14} \rightarrow A_{15}$
	$A_7 \rightarrow A_8, A_8$	$A_{15} \rightarrow A_{16}$
B_{11} Group	$A_2 \rightarrow A_4$	$A_5 \rightarrow A_7$
	$A_3 \rightarrow A_5$	$A_{13} \rightarrow A_{15}$
B_{13} Group	$A_3 \rightarrow A_6$	$A_9 \rightarrow A_{12}$

Step 6: Fuzzify the variation of the secondary factors and evaluate the linguistic terms of the main factor and the secondary factors, respectively, where the sub-steps are shown as follows:

Step 6.1: Select the secondary factors to fuzzify the variation, such as singular kinds "the Dow Jones", "the NASDAQ" and "the M_{1b}"; double kinds like "the Dow Jones and the NASDAQ", "the NASDAQ and the M_{1b}" and "the Dow Jones and M_{1b}"; triple kinds like "the Dow Jones, the NASDAQ and the M_{1b}". We let the variations of the Dow Jones, the NASDAQ and the M_{1b} be denoted by $Var_{Dow\ Jones}$,

Var_{NASDAQ} and Var_{M1b}, respectively. The variation of the secondary factor Var_s is calculated as follows:

Situation 1: If we use one secondary factor for prediction, the variation of the secondary factor Var_s is calculated as follows:

(i) When using "the Dow Jones" for prediction, the variation of the secondary factor $Var_s = Var_{Dow\ Jones}$.

(ii) When using "the NASDAQ" for prediction, the variation of the secondary factor $Var_s = Var_{NASDAQ}$.

(iii) When using "the M_{1b}" for prediction, the variation of the secondary factor $Var_s = Var_{M1b}$.

Situation 2: If we use two secondary factors for prediction, the variation of the secondary factor Var_s is calculated as follows:

(i) When using "the Dow Jones and the NASDAQ" for prediction, the variation of the secondary factor $Var_s = \frac{Var_{Dow\ Jones} + Var_{NASDAQ}}{2}$.

(ii) When using "the Dow Jones and the M_{1b}" for prediction, the variation of the secondary factor $Var_s = \frac{Var_{Dow\ Jones} + Var_{M1b}}{2}$.

(iii) When using "the NASDAQ and the M_{1b}" for prediction, the variation of the secondary factor $Var_s = \frac{Var_{NASDAQ} + Var_{M1b}}{2}$.

Situation 3: If we use three secondary factors "the Dow Jones, the NASDAQ and the M_{1b}" for prediction, the variation of the secondary factor $Var_s = \frac{Var_{Dow\ Jones} + Var_{NASDAQ} + Var_{M1b}}{3}$.

In the following, we use "the Dow Jones and the NASDAQ" as the secondary factor for prediction. For example, from Table IV, we can see that the variations of the Dow Jones and the NASDAQ of the trading day 2004/1/5 are 1.289356 % and 2.027229 %, therefore, the variation of the secondary factor Var_s is equal to $\frac{Var_{Dow\ Jones} + Var_{NASDAQ}}{2} = \frac{1.289356\% + 2.027229\%}{2} = \frac{3.316585\%}{2}$ = 1.658292 %. Table IX shows the variation of the secondary factor Var_s "the Dow Jones and the NASDAQ".

TABLE IX. THE VARIATION OF THE SECONDARY FACTOR "THE DOW JONES AND THE NASDAQ"

Date t	Var_t
2004/1/5	1.658292 %
2004/1/6	0.218807 %
2004/1/7	0.447902 %
⋮	⋮
2004/10/29	0.095618 %

Step 6.2: Fuzzify each variation of the secondary factor Var_s into a fuzzy set defined in **Step 5.2**. If the variation of the secondary factor Var_s belongs to v_j, where $1 \le j \le m$, then the variation of the secondary factor Var_s is fuzzified into B_j. For example, from Table IX, we can see that the variation of the secondary factor Var_s of the secondary factor "the Dow Jones and the NASDAQ" on 2004/1/5 is 1.658292 %, which is fuzzified into B_9. Table X shows the fuzzified variation Var_s of the secondary factor "the Dow Jones and the NASDAQ".

Step 6.3: Based on the linguistic terms of the fuzzified variations, group the fuzzy variation of the main factor on trading day t having the same linguistic term of the fuzzy variation of the secondary factor on trading $t - 1$. For example, let us consider the following fuzzy variation between the secondary factor on trading day $t - 1$ and the main factor on trading day t: the fuzzy variation of the main factor on trading day t is "B_{b1}" and the fuzzy variation of the secondary factor on trading day $t - 1$ is B_z; the fuzzy variation of the main factor on trading day $t + m$ is "B_{b2}" and the fuzzy variation of the secondary factor on trading day $t + m - 1$ is B_z; …; the fuzzy variation of the main factor on trading day $t + n$ is "B_{bk}" and the fuzzy variation of the secondary factor on trading day $t + n - 1$ is B_z. Then, these fuzzy variations of the main factor can be grouped into the same fuzzy variation group of the secondary factor in which each fuzzy variation of the secondary factor in this group is B_z, shown as follows:

$$B_{b1}, B_{b2}, \ldots, B_{bk}.$$

For example, from Table X and Table VII, we can see that the fuzzy variation of the secondary factor on 2004/1/5 is B_9 and the fuzzy variation of the main factor on 2004/1/6 is B_8. We can then group the fuzzy variation of the main factor "B_8" into the fuzzified variation of the secondary factor B_9 group. Table XI shows the fuzzy variation groups with respect to different fuzzified variations of the secondary factor.

TABLE X. FUZZIFIED VARIATIONS OF THE SECONDARY FACTOR "THE DOW JONES AND THE NASDAQ"

Date	Fuzzy Set
2004/1/5	B_9
2004/1/6	B_8
2004/1/7	B_8
⋮	⋮
2004/10/29	B_8

TABLE XI. FUZZY VARIATION GROUPS

B_6 Group	B_6, B_8, B_9, B_8, B_6, B_7, B_1, B_5, B_9, B_8, B_5, B_3, B_6, B_4, B_9, B_9, B_8, B_6, B_9, B_6, B_7, B_9, B_7, B_7, B_7, B_7, B_8, B_8, B_7, B_7, B_7, B_7, B_6
B_7 Group	B_7, B_9, B_8, B_8, B_7, B_6, B_8, B_9, B_8, B_9, B_7, B_8, B_6, B_8, B_{10}, B_7, B_9, B_7, B_6, B_8, B_8, B_{13}, B_8, B_8, B_8, B_7, B_5, B_9, B_7, B_{10}, B_7, B_5, B_2, B_7, B_{10}, B_5, B_6, B_{11}, B_8, B_6, B_8, B_6, B_6, B_6, B_6, B_7, B_7, B_5, B_9, B_7, B_{10}, B_7, B_8, B_8, B_{10}, B_8, B_7, B_8, B_9, B_7, B_7, B_7, B_9, B_7, B_7, B_5, B_7, B_7
B_8 Group	B_7, B_8, B_8, B_7, B_7, B_8, B_9, B_7, B_9, B_7, B_7, B_8, B_8, B_8, B_6, B_7, B_7, B_8, B_8, B_7, B_{10}, B_8, B_7, B_8, B_6, B_7, B_6, B_{10}, B_2, B_8, B_{13}, B_7, B_8, B_8, B_9, B_8, B_6, B_4, B_{11}, B_7, B_5, B_5, B_6, B_9, B_7, B_9, B_7, B_9, B_7, B_7, B_7, B_9, B_6, B_8, B_7, B_9, B_8, B_7, B_7, B_9, B_7, B_9, B_7, B_8, B_9, B_7, B_7, B_6, B_8, B_8, B_8, B_8, B_6, B_7, B_8, B_8, B_7, B_8
B_9 Group	B_8, B_9, B_8, B_9, B_5, B_{11}, B_8, B_{10}, B_8, B_9, B_9, B_8, B_8, B_9, B_7, B_7, B_7, B_{11}, B_6, B_8, B_{10}, B_7, B_8
B_{10} Group	B_7

Step 6.4: Evaluate the linguistic terms of the fuzzy variation groups between the secondary factor and the main factor. Let the fuzzy variations of the main factor and the secondary factor be B_M and B_S, respectively, where M and S are positive numbers, $1 \le M \le 14$ and $1 \le S \le 14$. Let $B_{S,1}$ be the static counter denoting the number of fuzzy variations in the B_S Group and $M < S$; let $B_{S,2}$ be the static counter denoting the number of fuzzy variations in the B_S Group and $M = S$; let $B_{S,3}$ be the static counter denoting the number of fuzzy variations in the B_S Group and $M > S$, where $B_{S,1}$, $B_{S,2}$ and $B_{S,3}$ are integers whose initial values are 0. Evaluate the fuzzy variation B_S groups in which the effect between the secondary factor and the main factor appear, described as follows:

Situation 1: In the B_S Group, if $M < S$, then we can see that when the secondary factor is B_S, the index M of the fuzzy variation of the main factor is less than S. We add one to the $B_{S,1}$ when the secondary factor is B_S and $M < S$.

Situation 2: In the B_S Group, if $M = S$, then we can see that when the secondary factor is B_S, the index M of the fuzzy variation of the main factor M is equal to S. We add one to the $B_{S,2}$ when the secondary factor is B_S and $M = S$.

Situation 3: In the B_S Group, if $M > S$, then we can see that when the secondary factor is B_S, the index M of the fuzzy variation of the main factor is bigger than S. We add one to the $B_{S,3}$ when the secondary factor is B_S and $M > S$.

For example, we let the TAIEX be the main factor and let "the Dow Jones and the NASDAQ" be the secondary factor. From the B_6 Group shown in Table XI, we can see that the fuzzy variations of the main factor are B_6, B_8, B_9, B_8, B_6, B_7, B_1, B_5, B_9, B_8, B_5, B_3, B_6, B_4, B_9, B_9, B_8, B_6, B_9, B_6, B_7, B_9, B_7, B_7, B_7, B_7, B_8, B_8, B_7, B_7, B_7, B_7 and B_6. Because S is 6 and the situation when $M < S$ are B_1, B_5, B_5, B_3 and B_4, the total number of fuzzy variations is 5, i.e., $B_{6,1} = 5$. In the same way, the situation when $M = S$ are B_6, B_6, B_6, B_6, B_6 and B_6, the total number of fuzzy variations number is 6, i.e., $B_{6,2} = 6$. In the same way, the situation when $M > S$ are B_8, B_9, B_8, B_7, B_9, B_8, B_9, B_9, B_8, B_9, B_7, B_9, B_7, B_7, B_7, B_7, B_8, B_8, B_7, B_7, B_7 and B_7, the total number of fuzzy variations is 22, i.e., $B_{6,3} = 22$.

Table XII shows the statistics of the fuzzy variations of the secondary factor "the Dow Jones and the NASDAQ", where the main factor is the TAIEX.

TABLE XII. THE STATISTICS OF THE FUZZY VARIATIONS OF THE SECONDARY FACTOR "THE DOW JONES AND THE NASDAQ", WHERE THE MAIN FACTOR IS THE TAIEX

S \ Static Counters	$B_{S,1}$	$B_{S,2}$	$B_{S,3}$
$S = 1$	0	0	0
$S = 2$	0	0	0
$S = 3$	0	0	0
$S = 4$	0	0	0
$S = 5$	0	0	0
$S = 6$	5	6	22
$S = 7$	15	21	32
$S = 8$	39	24	15
$S = 9$	14	5	4
$S = 10$	1	0	0
$S = 11$	0	0	0
$S = 12$	0	0	0
$S = 13$	0	0	0
$S = 14$	0	0	0

Step 7: Define the weights of the fuzzy variation B_s of the secondary factor when $S > M$, $S = M$, and $S < M$. According to **Step 6.4** , we let $B_{S,1}$ be the total number of times when the secondary factor is B_S and $S > M$; let $B_{S,2}$ be the total number of times when the secondary factor is B_S and $S = M$; let $B_{S,3}$ be the total number of times when the secondary factor is B_S and $S < M$. Let $W_{Bs,1}$, $W_{Bs,2}$ and $W_{Bs,3}$ be the weights of the fuzzy set B_S at different situations, where $W_{Bs,1}$ denotes the weight of B_S when $S > M$; $W_{Bs,2}$ denotes the weight of B_S when $S = M$; $W_{Bs,3}$ denotes the weight of B_S when $S < M$. The weight of the fuzzy variation $B_{S,k}$ is calculated as follows:

$$W_{Bs,k} = \frac{B_{s,k}}{B_{s,1} + B_{s,2} + B_{s,3}}\ , \tag{3}$$

where k is a positive integer and $k = 1, 2, 3$. For example, from Table XII, we can see that the number of times is 5 when the fuzzy variation is B_6 and $S > M$; the number of times is 6 when the fuzzy variation is B_6 and $S = M$; the number of times is 22 when the fuzzy variation is B_6 and $S < M$. Therefore, the weight $W_{B6,1}$ is equal to $\frac{5}{5+6+22} = 0.151515$, the weight $W_{B6,2}$ is equal to $\frac{6}{5+6+22} = 0.181818$, and the weight $W_{B6,3}$ is equal to $\frac{22}{5+6+22} = 0.666667$. In summary, the weights of the fuzzy variation B_j of the secondary factor are shown in Table XIII, where $1 \le j \le 14$.

Step 8: Assume that the main factor $F(t-1) = A_i$ and assume that we want to predict the main factor $F(t)$, where A_i is a fuzzy set. Based on the fuzzy variation of the secondary factor $F(t-1) = B_j$, we choose the corresponding fuzzy variation B_j of the weight of the secondary factor of the fuzzy logical relationship groups for prediction. Assume that the fuzzy variation of the secondary factor of the trading day $t-1$ is B_j. We then choose the fuzzy logical relationship: "$A_i \rightarrow A_{i1}, A_{i2}, \ldots, A_{iz}$" in the Group B_j. Let u_{i1}^L, u_{i2}^L, …, and u_{iz}^L be the minimum value of the intervals u_{i1}, u_{i2}, …, and u_{iz}, respectively; let u_{i1}^M, u_{i2}^M, …, and u_{iz}^M be the midpoints of the intervals u_{i1}, u_{i2}, …, and u_{iz}, respectively; let u_{i1}^R, u_{i2}^R, …, and u_{iz}^R be the maximum value of the intervals u_{i1}, u_{i2}, …, and u_{iz}, respectively. The new value u_a* of u_a is calculated as follows:

$$u_a^* = W_{Bj,1} \times u_a^L + W_{Bj,2} \times u_a^M + W_{Bj,3} \times u_a^R, \tag{4}$$

where $a = i1, i2, \ldots, iz$, and the forecasted value FV of day t is calculated as follows:

$$FV = \frac{\sum_{a=i1}^{iz} u_a^*}{z}\ . \tag{5}$$

TABLE XIII. THE WEIGHTS OF THE FUZZY VARIATION $B_{S,k}$ OF THE SECONDARY FACTOR

Weight of Fuzzy Set \ k	$k = 1$	$k = 2$	$k = 3$
$W_{B1,k}$	0	0	0
$W_{B2,k}$	0	0	0
$W_{B3,k}$	0	0	0
$W_{B4,k}$	0	0	0
$W_{B5,k}$	0	0	0
$W_{B6,k}$	0.151515	0.181818	0.666667
$W_{B7,k}$	0.220588	0.308824	0.470588
$W_{B8,k}$	0.5	0.307692	0.192308
$W_{B9,k}$	0.608696	0.217391	0.173913
$W_{B10,k}$	1	0	0
$W_{B11,k}$	0	0	0
$W_{B12,k}$	0	0	0
$W_{B13,k}$	0	0	0
$W_{B14,k}$	0	0	0

For example, assume that we want to forecast the TAIEX of 2004/11/2 by the first order fuzzy logical relationships. From Table I and Table XIII we can see that because the TAIEX on 2004/11/1 is 5656.17 (Note: Its fuzzified TAIEX is A_4) and because from Table V, we can see that the variation of the Dow Jones and the NASDAQ on 2004/11/1 are 0.268463 % and 0.247090 %, respectively, based on **Step 6.1**, the variation of the secondary factor is $\frac{0.268463\% + 0.247090\%}{2} = \frac{0.515553\%}{2} = 0.257777$ % and based on Table VI, we can see that the fuzzy variation of the secondary factor is B_8. We choose the B_8 group shown in Table VIII, where the left-hand side of the fuzzy logical relationship "$A_4 \rightarrow A_5, A_4, A_4, A_4, A_5$" is A_4. The given weights u_5 and u_4 become u_5* and u_4*, respectively, for forecasting. The minimum value of the interval u_5 is 5700, the midpoint of the interval u_5 is 5750 and the maximum value of the interval u_5 is 5800. Based on Eq. (3), we can get the weight of $B_{8,1} = 0.5$, the weight of $B_{8,2} = 0.307692$ and the weight of $B_{8,3} = 0.192308$. Based on Eq. (4), the new value u_5* of u_5 for prediction is calculated as follows: $5700 \times 0.5 + 5750 \times 0.307692 + 5800 \times$

0.192308 = 5734.62. In the same way, we can get u_4^* = 5634.62. Finally, based on Eq. (5), the forecasted TAIEX of 2004/11/2 is calculated as follows:

$$\frac{5734.62+5634.62+5634.62+5634.62+5734.62}{5}=5674.62.$$

IV. Experimental Results

In this section, we apply the proposed method to forecast the TAIEX from 2000 to 2004. We evaluate the performance of the proposed method using the root mean square error (RMSE), which is defined as follows:

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(forecasted\ value_i - actual\ value_i)^2}{n}}, \tag{6}$$

where n denotes the number of dates needed to be forecasted. In Table XIV, we make a comparison of the RMSE and the average RMSE of the proposed method with Chen's method [1] and Huarng et al.'s method [9]. From Table XIV, we can see that the proposed method with "the Dow Jones and the NASDAQ" gets the smallest average RMSE compared to Chen's method [1] and Huarng et al.'s method [9]. It means that the proposed method with the Dow Jones and the NASDAQ gets a higher average forecasting accuracy rate than Chen's method [1] and Huarng et al.'s method [9].

TABLE XIV. A Comparison of the RMSEs and the Average RMSE for Different Methods

Methods \ RMSE Year	2000	2001	2002	2003	2004	Average RMSE
Chen's Method [1]	176.32	147.84	101.18	74.46	84.28	116.82
Huarng et al.'s Method (use NASDAQ) [9]	158.7	136.49	95.15	65.51	73.57	105.88
Huarng et al.'s Method (use Dow Jones) [9]	165.8	138.25	93.73	72.95	73.49	108.84
Huarng et al.'s Method (use M_{1b}) [9]	169.19	133.26	97.1	75.23	82.01	111.36
Huarng et al.'s Method (use NASDAQ & Dow Jones) [9]	157.64	131.98	93.48	65.51	73.49	104.42
Huarng et al.'s Method (use NASDAQ & M_{1b}) [9]	155.51	128.44	97.15	70.76	73.48	105.07
Huarng et al.'s Method (use NASDAQ & Dow Jones & M_{1b}) [9]	154.42	124.02	95.73	70.76	72.35	103.46
The Propose Method (use Dow Jones)	127.51	121.98	74.65	66.02	58.89	89.81
The Propose Method (use NASDAQ)	129.87	123.12	71.01	65.14	61.94	90.22
The Propose Method (use M_{1b})	129.87	117.61	85.85	63.1	67.29	92.74
The Propose Method (use Dow Jones & NASDAQ)	124.06	125.12	72.25	57.14	56.95	87.10
The Propose Method (use Dow Jones and M_{1b})	127.75	115.64	79.45	60.41	65.86	89.82
The Propose Method (use NASDAQ and M_{1b})	128.45	126.14	76.03	66.96	65.5	92.62
The Propose Method (use Dow Jones, NASDAQ and M_{1b})	129.57	119.66	73.25	66.8	65.41	90.94

V. Conclusions

In this paper, we have presented a new method to forecast the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) based on fuzzy time series. From Table XIV, we can see that the proposed method with the Dow Jones and the NASDAQ gets a higher average forecasting accuracy rate than Chen's method [1] and Huarng et al.'s method [9] for forecasting the TAIEX. The experimental results show that the proposed method gets a higher average forecasting accuracy rate than the methods, presented in [1] and [9].

Acknowledgment

This work is supported in part by the National Science Council, Republic of China, under Grant NSC 97-2221-E-011-107-MY3.

References

[1] S. M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems*, vol. 81, no. 3, pp. 311-319, 1996.

[2] S. M. Chen, "Forecasting enrollments based on high-order fuzzy time series," *Cybernetics and Systems*, vol. 33, no. 1, pp. 1-16, 2002.

[3] S. M. Chen and N. Y. Chung, "Forecasting enrollments using high-order fuzzy time series and genetic algorithm," *International Journal of Intelligent Systems*, vol. 21, no. 5, pp. 485-501, 2006.

[4] S. M. Chen and C. C. Hsu, "A new method to forecast enrollments using fuzzy time series," *International Journal of Applied Science and Engineering*, vol. 2, no. 3, pp. 234-244, 2004.

[5] S. M. Chen and J. R. Hwang, "Temperature prediction using fuzzy time series," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 30, no. 2, pp. 263-275, 2000.

[6] S. M. Chen and N. Y. Wang, "Handling forecasting problems based on high-order fuzzy time series and fuzzy-trend logical relationships," *Proceedings of the 2008 Workshop on Consumer Electronics*, Taipei County, Taiwan, Republic of China, pp. 759-764, 2008.

[7] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series," *Fuzzy Sets and Systems*, vol. 123, no. 3, pp. 387-394, 2001.

[8] K. Huarng and H. K. Yu, "Ratio-based lengths of intervals to improve fuzzy time series forecasting," *IEEE Transactions on Systems, Man, and, Cybernetics-Part B: Cybernetics*, vol. 36, no. 2, pp. 328-340, 2006.

[9] K. Huarng, H. K. Yu, and Y. W. Hsu "A multivariate heuristic model for fuzzy time-series forecasting," *IEEE Transactions on Systems, Man, and, Cybernetics Part-B: Cybernetics*, vol. 37, no. 4, pp. 836-846, 2007.

[10] L. W. Lee, L. H. Wang, S. M. Chen, and Y. H. Leu, "Handling forecasting problems based on two-factors high-order fuzzy time series," *IEEE Transactions on Fuzzy Systems,* vol. 14 , no. 3, pp. 468-477, 2006.

[11] Q. Song and B. S. Chissom, "Fuzzy time series and its model," *Fuzzy Sets and Systems*, vol. 54, no. 3, pp. 269-277, 1993.

[12] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - Part I," *Fuzzy Sets and Systems*, vol. 54, no. 1, pp. 1-9, 1993.

[13] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - Part II," *Fuzzy Sets and Systems,* vol. 62, no. 1, pp. 1-8, 1994.

[14] N. Y. Wang and S. M. Chen, "Temperature prediction and TAIFEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2143-2154, March 2009.

[15] N. Y. Wang, S. M. Chen, and J. S. Pan, "Forecasting enrollments based on automatic clustering techniques and fuzzy time series," *Proceedings of the 12th Conference on Artificial Intelligence and Applications*, Yunlin, Taiwan, Republic of China.

[16] H. K. Yu, "A refined fuzzy time-series model for forecasting," *Physica A*, vol. 346, no. 3-4, pp. 657–681, 2004.

[17] H. K. Yu, "Weighted fuzzy time-series models for TAIEX forecasting," *Physica A*, vol. 349, no. 3-4, pp. 609–624, 2004.

[18] H. K. Yu and K. H. Huarng, "A bivariate fuzzy time series model to forecast the TAIEX," *Expert Systems with Applications*, vo1. 34, no. 4, pp. 2945-2952, 2008.

[19] I. H. Kuo, S. J. Horng, T. W. Kao, T. L. Lin, C. L. Lee, and Y. Pan, "An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization," *Expert Systems with Applications*, vo1. 36, no. 3, pp. 6108-6117, 2009.

[20] H. J. Teoh, T. L. Chen, C. H. Cheng, and H. H. Chu, "A hybrid multi-order fuzzy time series for forecasting stock markets," *Expert Systems with Applications*, vo1. 36, no. 4, pp. 7888-7897, 2009.

[21] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338-353, 1965.