African Journal of Business Management Vol. 5(15), pp. 6377-6387, 4 August, 2011
Available online at http://www.academicjournals.org/AJBM
DOI: 10.5897/AJBM10.1365
ISSN 1993-8233 ©2011 Academic Journals

Full Length Research Paper

An expanded Adaptive Neuro-Fuzzy Inference System (ANFIS) model based on AR and causality of multi-nation stock market volatility for TAIEX forecasting

Liang-Ying Wei

Department of Information Management, Yuanpei University, 306 Yuanpei Street, Hsin Chu 30015, Taiwan.
E-mail: lywei@mail.ypu.edu.tw.

Accepted 21 January, 2011

For common people, stock investing is one popular way to manage their property. As Information Technology (IT) has risen in recent years, every security company has analyzed computer systems for their customers by developing their own investing. Taiwan is an island nation, and the economy relies on international trade deeply. The fluctuations of international stock markets will impact Taiwan stock market to a certain degree. Therefore, the use of fluctuations of other stock markets as forecasting factors for forecasting the Taiwan stock market is a practical way. In this paper, the proposed model uses the fluctuations of other national stock markets as forecasting factors, employs different discretization methods (Fuzzy C-means clustering, Subtractive Clustering and Cumulative Probability Distribution Approach) to discretize stock data, utilizes a fuzzy inference system to produce understandable rules, and applies an adaptive neural network to optimize model parameters to reach the best forecasting accuracy. To evaluate the forecasting performances, the proposed model is compared with two different models, Chen's model and Yu's model. The experimental results indicate that the proposed model is superior to the listing methods in terms of RMSE (root mean squared error).

Key words: Data mining, Adaptive Neuro-Fuzzy Inference System (ANFIS), stock forecasting, subtractive clustering, cumulative probability distribution approach.

INTRODUCTION

The dependence of a global economic system, such as Taiwan, due to its lack of energy is very high, and further, Taiwan is a viable member of the international economic society. Even though Taiwan had resources before, it is now difficult to find more resources for Taiwan's daily needs and industries. So, Taiwan needs to try another way to get resources, and it is the only way to trade or exchange with other countries. The Taiwanese government has developed international trade in recent years and encourages entrepreneurs to invest abroad to increase their wealth. Therefore, Taiwan will go to find the potential markets and compete with other countries. Taiwan's advantage goes way beyond cheap labor. Fortunately, it is near China, which is the second biggest market in the world and has cheap labor. Since 1993, China has become Taiwan's third largest trading partner, after the United States and Japan. In 2002, Taiwan's trade with the United States, Japan, and China was $44.9 billion, $39.3 billion, and $37.4 billion, respectively. In addition, since 1993, China has also become Taiwan's second largest export market, next to the United States. In 2002, China became Taiwan's largest export market for the first time. In general, the statistics on trade between Taiwan and China should include the transit exports (re-exports) to China via Hong Kong and other places and trans-shipment (goods are consigned directly from Taiwan to a buyer in China, though the goods are transported via Hong Kong). Under such circumstances, the impacts of global economic fluctuations on Taiwan are very huge, especially those of the USA and Hong Kong. Besides, from a study of the literature, Dickinson (2000) proves that the influence of stock markets in different counties on each other is more or less.

The stock market is one of the most exciting and challenging monetary activities, and the accuracy of information for investment planning is crucial. The significance of volatility causality in multi-nation stock markets is an important indicator for forecasting another stock market. There is evidence of the influence of US and Hong Kong stock market volatility. Thus, we employ volatility causality in multi-nation stock markets in this paper.

Conventional time-series models have been applied to forecasting problems for a long time, such as the ARCH (p) (Autoregressive Conditional Heteroscedasticity) model proposed by Engle (1982); Bollerslev (1986) proposed the GARCH (Generalized ARCH) model to refine the ARCH model, and Box and Jenkins (1976) proposed the autoregressive moving average (ARMA) model and the ARIMA model. Autocorrelation (AR) is the correlation (relationship) between members of time series of observations, such as weekly share prices or interest rates and the same values at a fixed time interval. More technically, autocorrelation occurs when residual error terms from observations of the same feature at different time periods are correlated (related). AR, a popular and important method in conventional time-series models, has been applied to time-series forecasting problems. However, the AR technique has limited capabilities for modeling time series data, and more advanced nonlinear methods, including neural networks, have been frequently applied with success (Chatfield, 2003). Further, fuzzy logic-based modeling techniques are appealing because of their interpretability and potential in addressing a broad spectrum of problems.

Furthermore, fuzzy time-series models have been employed for stock price forecasting. Song and Chissom (1993) first proposed the original model of the fuzzy time-series, and the following researchers focused on the two major processes of the fuzzy time-series model: (1) fuzzification and (2) establishment of fuzzy relationships and forecasting. In the fuzzification process, the length of intervals for the universe of discourse could affect forecasting, and Huarng proposed the distribution-based and average-based length to approach this issue (Huarng, 2001). In addition, Chen proposed a new method (Chen and Chung, 2006), in which the length of linguistics intervals is tuned by using genetic algorithms. In the process of establishing fuzzy relationships and forecasting, Yu (2005) argued that recurrent fuzzy relationships should be considered in forecasting and recommended that different weights should be assigned to various fuzzy relationships. Therefore, Yu (2005) proposed a weighted fuzzy time-series method to forecast the TAIEX. Further, Cheng et al. (2006) proposed a methodology that incorporates trend-weighting into the fuzzy time-series model. To take advantage of neural networks (nonlinear capabilities), Huarng and Yu (2006) chose a neural network to establish fuzzy relationships in fuzzy time series, which are also nonlinear, but the process of mining fuzzy logical relationships is not easily understandable, just like a black box (Chen et al., 2008). However, the aforementioned models have been limited to one feature application (Yu and Huarng, 2008). Recently, Chen et al. (2008) proposed a comprehensive fuzzy time-series, in which factors of linear relationships between recent periods of stock prices and fuzzy logical relationships (nonlinear relationships) are mined from time-series into forecasting processes. Jilani and Burney (2008) proposed a simple time-variant fuzzy time-series method to forecast TAIEX and enrollment at the University of Alabama. Cheng et al. (2008) proposed a new fuzzy time-series method, which is based on a weighted-transitional matrix, and also proposed two new forecasting methods: the Expectation Method and the Grade-Selection Method. Yu and Huarng (2008) proposed a bivariate model that applies neural networks to fuzzy time-series forecasting. Further, in the evolution of time series models, many researchers have applied data mining techniques in financial analysis (Takahama and Sakai, 2009; Aihara et al., 2009; Zhang and Chen, 2009; Chang and Chen, 2009).

Past forecasting methods based on fuzzy sets rely largely on expert opinions and are difficult to explore directly and correctly utilize valuable information embedded in the collected data. Therefore, instead of utilizing expert opinions to build membership functions, this study uses an objective discritization method to discretize features and construct membership functions. Data discretization has many advantages. For example, the data dimension can be reduced and simplified, and using discrete features is usually more compact and shorter than using continuous ones (Liu et al., 2002). Further, entropy-based discretization (Christensen, 1980) methods, such as the minimal entropy principle approach (MEPA), use class information to discretize features. The entropy-based discretization of quantitative features is a valuable aspect of data mining, particularly in rough set and classification problems.

From the above, there are three major drawbacks in these models: (1) most statistical methods rely upon some assumptions about the features used in the analysis, so it is limited for application to all datasets (Jilani and Burney, 2008). (2) Most conventional time-series models utilize only one nation's stock data for input variables in forecasting. However, financial analysts should consider many market features in forecasting. For this reason above, forecasting models should utilize more features to improve forecasting accuracy (Yu and Huarng, 2008). (3) Rules mined from artificial neural networks (ANNs) are not easily understandable (Chen et al., 2008).

In order to reconcile the drawbacks above, this paper considers that the volatility of American and Hong Kong stock indexes plays an important role in affecting the volatility of TAIEX. Because the forecasting models utilize the relation between the volatility of the American stock index and of TAIEX and the relation between the volatility of the Hong Kong index and of TAIEX, the analytical results approximate the real world. Further, the AR method is applied to the proposed model to enhance it.

Finally, the proposed model, employing fuzzy if-then rules (in the ANFIS method), can model the qualitative aspects of human knowledge and can be applicable for humans to use.

Based on the concepts above, this paper proposes a hybrid model to forecast the Taiwan stock index. First, this paper calculates the volatility of the NASDAQ stock index and Hang Seng stock index by Equations (19)-(20). Second, we test the lag period of AR for the TAIEX to build the AR method. Then, we use the fuzzy inference system to forecast the Taiwan stock index, considering a multi-nation stock index (NASDAQ stock index and Hang Seng stock index) to forecast the TAIEX (t+1). Finally, we optimize the fuzzy inference system parameters by an adaptive network, which can overcome the limitations of statistical methods (the data need obey some mathematical distribution).

LITERATURE REVIEW

This section reviews the related studies of the cumulative probability distribution approach (CPDA), the adaptive network-based fuzzy inference system (ANFIS), fuzzy C-means clustering (FCM), and subtractive clustering (Subclust).

Cumulative probability distribution approach (CDPA)

Probability refers to the study of randomness and uncertainty. In any situation, one out of a number of possible outcomes may occur. The theory of probability provides methods for quantifying the chances, or likelihoods, associated with the various outcomes. Because a probability distribution on the real line is determined by the probability of being in a half-open interval *p(a, b]*, *F(b)–F(a)* if *a<b*. The probability distribution of a real-valued random variable *X* is completely characterized by its cumulative distribution function (CDF) (Acklam, 2004). For every real number *x*, the CDF of *X* is given by

$$x \rightarrow F_X(x) = P(X \le x) \qquad \forall x \in \Re \qquad 1$$

where the right-hand side represents the probability *(p)* that the random variable *X* takes on a value less than or equal to *x*. Capital *F* is used to represent the cumulative distribution function, in contrast to the lower-case *f* , which is used to represent probability density functions and probability mass functions. The CDF of *X* can be defined in terms of the probability density function *f* , as follows:

$$F(x) = P[X \le x] = \int_{\infty}^{x} f(t)dt \qquad 2$$

The inverse of the normal CDF is computed with parameters μ and σ at the corresponding probabilities in *P*, where μ denotes the mean and σ denotes the standard deviation of the data (Acklam, 2004). The normal inverse function in terms of the normal CDF is defined as:

$$x = F^{-1}(p \mid \mu, \sigma) = \{x : F(x \mid \mu, \sigma) = p\} \qquad 3$$

Where,

$$p = F(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt \qquad 4$$

The cumulative probability of normal distribution is used to determine the intervals. The steps of the cumulative probability distribution approach are as follows:

Step 1: Test normal distribution

In this approach, the data must be of normal distribution. This study utilizes the CPDA, because stock market fluctuations and returns tend to be of normal distribution (Anna, 2007) (Tuncay and Stauffer, 2007).

Step 2: Define the universe of discourse U

Let $U = [D_{min} - \sigma, D_{max} + \sigma]$, where D_{min} and D_{max} denote the minimum and maximum values in historical data and σ denotes the standard deviation of the yearly data.

Step 3: Determine the length of intervals and build a membership function

P_{LB}, as the lower boundary of cumulative probability, and P_{UB}, as the upper boundary of cumulative probability of each linguistic value, are computed by:

$$P_{LB} = (2i - 3) / 2n, \ (2 \leqq i \leqq n) \qquad 5$$

$$P_{UB} = i / n, \ (1 \leqq i \leqq n) \qquad 6$$

Where *i* denotes the order of the linguistic values and *n* denotes the number of linguistic values. The lower boundary of the first linguistic value and the upper boundary of the last linguistic value correspond to the lower and upper boundary, respectively. This step computes the inverse of the normal CDF by Equations (3) and (4).

Step 4: Fuzzify the historical data

According to the inverse of the normal CDF, the lower boundary, midpoint, and upper boundary as the triangular fuzzy number of each linguistic value can be computed. The triangular fuzzy number is applied to build a membership function. The membership degree of each instance is calculated to determine its linguistic value.

Fuzzy C-means clustering

Clustering has been gaining popularity as an efficient tool of data analysis to understand and visualize data structures. The prevalent formulation of this task is to use *c* feature vectors $v_j (j = 1, 2, ..., c)$ to represent the *c* clusters, such that a sample x_t is classified into the *j-th* cluster according to some measure of similarity and its corresponding objective function. Fuzzy C-means (FCM), proposed by Bezdek (1981), is the most famous and basic fuzzy clustering algorithm. FCM attempts to find a fuzzy partition of the data set by minimizing the following within-group least-squares error objective function with respect to fuzzy memberships u_{it} and center v_i :

$$J_{m'}(X,U,V) = \sum_{i=1}^{c}\sum_{t=1}^{n} u_{it}^{m'} d^2(x_t; v_i)$$

Where $m' > 1$ is the fuzziness index used to tune out the noise in the data, *n* is the number of feature vectors x_t , $c > 2$ is the number of clusters in the set, and $d(x_t; v_i)$ is the similarity measure between a datum and a center. Minimization of $J_{m'}$ occurs under the following constraints:

$$(1) 0 \le u_{it} \le 1, \forall i, t,$$

$$(2) 0 < \sum_{t=1}^{n} u_{it} \le n, \forall i,$$

$$(3) \sum_{i=1}^{c} u_{it} = 1, \forall t,$$

yielding an iterative minimization pseudo-algorithm, well known as the FCM algorithm. The components v_{ij} of each center v_i and the membership degrees u_{it} are updated according to the expressions

$$v_{ij} = \frac{\sum_{t=1}^{n} u_{it}^{m'} x_{kj}}{\sum_{i=1}^{n} u_{it}^{m'}} \text{ and } u_{it} = \frac{1}{\sum_{j=1}^{c} (\frac{d(x_t; v_i)}{d(x_t; v_j)})^{2/m'-1}}$$

where *j* is a variable on the feature space; i.e., *j*= 1,2,...,*m*.

Subtractive clustering

Chiu (1994) developed subtractive clustering, one of the fuzzy clustering methods, to estimate both the number and initial locations of cluster centers. Consider a set *T* of *N* data points in a *D*-dimensional hyper-space, where each data point W_i $(i = 1,2,..., N)$. $W_i = (x_i, y_i)$ where x_i denotes the p input variables and y_i is the output variable. The potential value P_i of data point is calculated by Equation (10)

$$P_i = \sum_{j=1}^{N} e^{-\alpha \| W_i - W_j \|^2} \qquad (10)$$

Where $\alpha = 4/r^2$, r is the radius defining a W_i neighborhood, and ||.|| denotes the Euclidean distance. The data point with many neighboring data points is chosen as the first cluster center. To generate the other cluster centers, the potential P_i is revised for each data point W_i by Equation (11). (7)

$$p_i = p_i - p_1^* \exp(-\beta \| W_i - W_1^* \|^2) \qquad (11)$$

Where, *β* is a positive constant defining the neighborhood that will have measurable reductions in potential. W_1^* is the first cluster center, and P_1^* is its potential value.

From Equation (11), the method selects the data point with the highest remaining potential as the second cluster center. For a general equation, we can rewrite Equation (11) as Equation (12).

$$p_i = p_i - p_k^* \exp(-\beta \| W_i - W_k^* \|^2) \qquad (12)(8)$$

Where $W_k^* = (x_k^*, y_k^*)$ is the location of the k'th cluster center and P_k^* is its potential value.

At the end of the clustering process, the method obtains *q* cluster centers and *D* corresponding spreads S_i , *i = (1, .., D)*. Then, we define their membership functions. The spread is calculated according to *β*.

ANFIS: Adaptive-Network-based Fuzzy Inference System

Jang (1993) proposed ANFIS, which is a fuzzy inference (9) system, implemented in the framework of adaptive networks. For illustrating the system, we assume a fuzzy inference system that consists of five layers of adaptive networks with two inputs *x* and *y* and one output *z*. The architecture of ANFIS is shown in Figure 1.

Then, we suppose that the system consists of 2 fuzzy if-then rules based on Takagi and Sugeno's type (Takagi and Sugeno, 1983):

Rule 1: If x is A_1 and y is B_1 , then $f_1 = p_1 x + q_1 y + r_1$

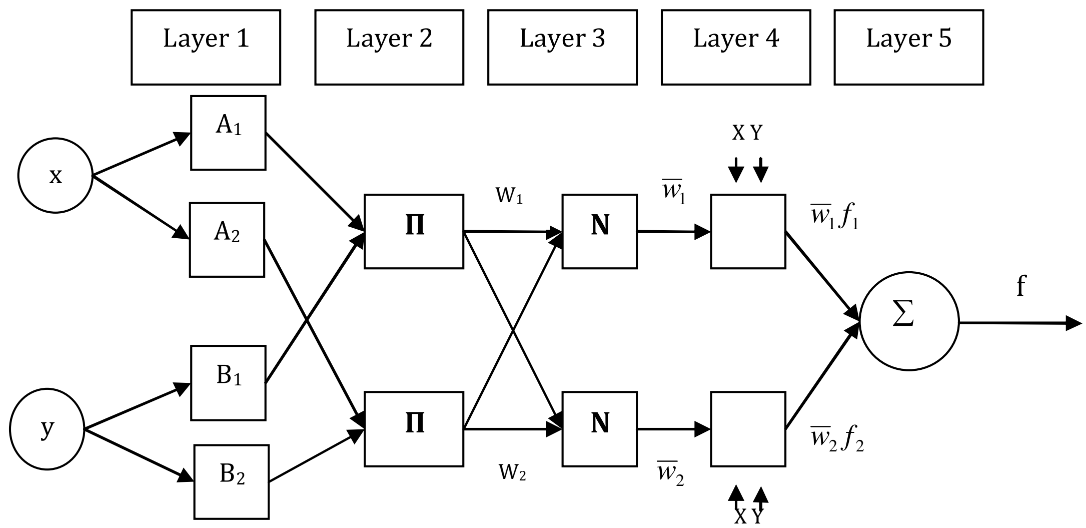

Figure 1. The architecture of ANFIS network.

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2 x + q_2 y + r_2$.

The node in the *i-th* position of the *k-th* layer is denoted as $O_{k,i}$, and the node functions in the same layer are of the same function family as thus described:

Layer 1: This layer is the input layer, and every node *i* in this layer is a square node with a node function (Equation (13)). $O_{1,i}$ is the membership function of A_i, and it specifies the degree to which the given x satisfies the quantifier A_i. Usually, we select the bell-shaped membership function as the input membership function (Equation (14)) with the maximum equal to 1 and the minimum equal to 0.

$$O_{1,i} = \mu A_i(x) \text{ for } i\text{=1, 2} \quad (13)$$

$$\mu A_i(x) = \frac{1}{1+[(\frac{x-c_i}{a_i})^2]^{bi}} \quad (14)$$

Where a_i, b_i, c_i are the parameters, *b* is a positive value, and *c* denotes the center of the curve.

Layer 2: Every node in this layer is a square node labeled Π, which multiplies the incoming signals and sends the product out by Equation (15).

$$O_{2,i} = w_i = \mu A_i(x) \times \mu B_i(y) \text{ for } i\text{=1, 2} \quad (15)$$

Layer 3: Every node in this layer is a square node labeled N. The *i-th* node calculates the ratio of the *i-th* rule's firing strength to the sum of all rules' firing strengths by Equation (16). The output of this layer can be called normalized firing strengths.

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2} \quad \text{for } i\text{=1, 2} \quad (16)$$

Layer 4: Every node i in this layer is a square node with a node function (Equation (17)). Parameters in this layer will be referred to as consequent parameters.

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i + q_i + r_i) \quad (17)$$

Where p_i , q_i , r_i are the parameters.

Layer 5: The single node in this layer is a circle node labeled $\sum$ that computes the overall output as the summation of all incoming signals (Equation (18))

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_{i=1} w_i f}{\sum_{i=1} w_i} = \text{overall output}$$

(18)

METHODOLOGY

From the reviewed literature, there are three major drawbacks mentioned in section 1: (1) time-series models utilize only one feature (Yu and Huarng, 2008); (2) statistical methods rely upon some assumptions (Jilani and, Burney, 2008); and (3) the rules mined from artificial neural networks (ANNs) are not easily understandable (Chen et al., 2008). In order to solve these drawbacks, this paper considers that the volatility of American and Hong Kong stock indexes plays an important role in affecting the volatility of the TAIEX. For this reason, the proposed model utilizes the relation between the volatility of the American stock index and the volatility of the TAIEX and the relation between the volatility of the Hong Kong stock index and the volatility of the TAIEX; the analytical results approximate the real world. Then, we use three different discretization methods for discretizing features. To specify the differences between the three different discretization methods, there are three discussions in this paper as follows: (1) CPDA is a discretization method based on the hypothesized estimations of normal distribution and probability distribution; (2) FCM attempts to find a fuzzy partition of the data set by minimizing the following within-group least-squares error objective function; and (3) the Subclust method is used to estimate both the number and initial locations of cluster centers. The data point with many neighboring data points is chosen as the cluster center. Further, a fuzzy inference system employing fuzzy if-then rules can model the qualitative aspects of human knowledge and can be applicable for humans to use.

Based on the aforementioned concepts, this paper proposes a new hybrid model to forecast the Taiwan stock index. First, this paper calculates the volatility of the NASDAQ stock index and Hang Seng stock index by Equations (19)-(20). Second, we test the lag period of AR for the TAIEX to build the AR model. Because advanced methods, such as CDPA, FCM, and Subclust, are machine learning techniques, the proposed model utilizes these methods for discretizing features to obtain more objective membership functions and enhance forecasting performance. Then, we use the fuzzy inference system to forecast the Taiwan stock index and consider a multi-nation stock index (NASDAQ stock index and Hang Seng stock index) and the AR method to forecast the TAIEX (t+1). Finally, we optimize the fuzzy inference system parameters using an adaptive network, which can overcome the limitations of statistical methods (the data need obey some mathematical distribution).

Based on these advantages of discretization methods and the ANFIS, this study proposes a new stock forecasting model, which incorporates three discretization methods and ANFIS. In the proposed model, three phases of building forecasting processes (9 processes) are provided as follows: (1) preprocess: calculate multivariate volatility and test the lag period of AR, and use three discretization methods to construct more objective and reasonable membership functions of features; (2) generate the ANFIS forecast model: generate a fuzzy inference system and train the fuzzy inference system; and (3) performance evaluation: forecast the testing TAIEX *(t+1)* and compare the results. Then, the overall flowchart of the proposed model is shown in Figure 2.

For easy understanding, this section uses some numerical data, as the step-by-step example shows the core concept in the proposed algorithm.

Step1 (Collect data set): In this section, we choose the TAIEX from 1997 to 2003 (7 sub-datasets) to illustrate the proposed model (such as the year 2000 sub-datasets, which contains 271 transaction days). Each training data is selected from January to October, and the remaining data (from November and December) are used for testing.

Step 2 (Calculate multivariate volatility (NASDAQ stock index and Hang Seng stock index): In this section, we define two features, namely (1) the NASDAQ (*N)* and (2) the Hang Seng (*H*), and calculate the volatility of the two features by Equations (19)-(20). Table 1 lists the differences in the features the NASDAQ and Hang Seng. From Table 1, some data under the NASDAQ and Hang Seng are empty, because there were no transactions on those days. For this reason, this paper fills in the last volatility as the differences.

$$\text{diff}(N(t)) = N(t) - N(t-1) \tag{19}$$

$$\text{diff}(H(t)) = H(t) - H(t-1) \tag{20}$$

Step 3 (Test the lag period of AR): In this step, orders from 1 to 5 are evaluated to determine which number of time lag is the most fitting for the experimental dataset. The least-square method is taken to build the model, and then four linear regression features (*close_price* (*t* – 1) to *close_price* (*t* – 5)) are selected to be estimated and tested. If the *p*-value is less than the significance level, given at 0.05 here, then reject the null hypothesis. Finally, the estimated TAIEX model is obtained, and the lag periods of the TAIEX can be determined as well.

To demonstrate the proposed model, a one-year period of the TAIEX (Year 1997) is employed as the experimental dataset for the proposed algorithm. We use the E-Views software package to fit the AR model for different lags and orders of the TAIEX, and five linear regression features (that is, from *close_price* (*t* – 1) to *close_price* (*t* – 5)) are selected to be estimated and tested. If the *p*-value is less than the significance level of 0.05, then reject the null hypothesis. Take the TAIEX in 1997 as an example, Figure 3 shows that the *p*-value (0.0000) for *close_price* (*t*–1)) is less than the significance level of 0.05 among the four features, from *close_price* (*t* -1) to *close_price* (*t* –5). Further, the feature *close_price* (*t* -1) is not equal to zero. Therefore, the order of AR is one.

Step 4 [Define and partition the universe of discourse for input features (B1 block of Figure 2)]: Firstly, we define each universe of discourse for three features (*TAIEX (t), diff(N(t)*), *diff(H(t)*)) according to the minimum and maximum value in each feature. Secondly, partition the universe of discourse into three linguistic intervals using fuzzy c-Means clustering (Bezdek, 1981) (triangular membership function), subtractive clustering (Bezdek, 1981) (Gaussian membership function), and the cumulative probability distribution approach (CPDA), respectively.

Step 5 [Set the type of membership function for output features (B2 block of Figure 2)]: There are two types of membership functions (MF) for output features, as follows.

1. Lineal type: a typical rule in a Sugeno fuzzy model has the form, as follows:

If x(*TAIEX (t)*) = A_i , y (*diff(N(t)*)) = B_j and z (*diff(H(t)*)) = C_k , then Output is $f_l = p_l x + q_l y + r_l z + s_l$

Where x (*TAIEX (t)*), y (*diff(N(t)*)), z (*diff(H(t)*)) are linguistic features, A_i, B_j, C_k are the linguistic labels (high, middle, low), f_l denotes the *l-th* output value, and p_l, q_l, r_l, s_l are the parameters (*i=1, 2, 3; j=1, 2, 3; k* =1, 2, 3, and *l*= 1, 2,… , 27).

2. Constant type: a zero-order Sugeno model; the output level f_l is a constant ($p_i = q_i = s_i = 0$).

Step 6 (Generate the fuzzy inference system): From Steps 3, 4 and 5, we can get six types of forecasting models: (1) AR(1) + Fuzzy C-means with linear type (AR(1)_FCM_L), (2) AR(1)+Fuzzy C-means with constant type (AR(1)_FCM_C), AR(1)+Subtractive Clustering with linear type (AR(1)_Subclust_L), (4) AR(1)+Subtractive Clustering with constant type (AR(1)_Subclust _C), (5) AR(1)+CPDA with linear type (AR(1)_CPDA _L), and (6) AR(1)+CPDA with constant type (AR(1)_CPDA _C). Then, generate the six different fuzzy inference systems according to the four types of forecasting models, respectively.

The detailed steps of generating the fuzzy inference system are described as follows: Firstly, from step 4, we can get the linguistic intervals as input membership functions, and the output membership functions are set by step 5. Secondly, generate fuzzy if-then

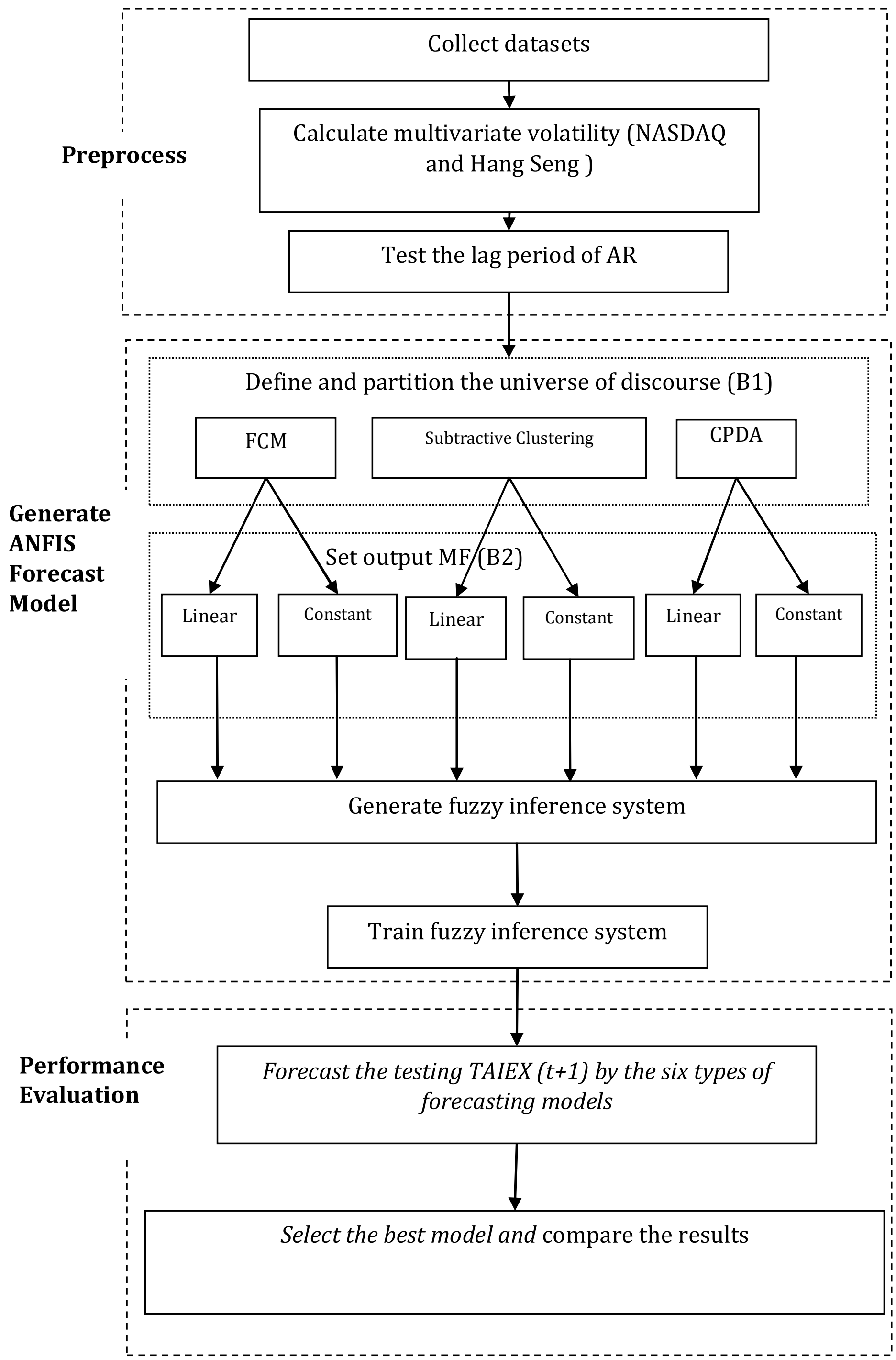

Figure 2. Flowchart of proposed procedure.

rules, where the linguistic values (A_i , B_i C_i) from the input membership functions are used as the 'if-condition' part and the output membership functions (f_i) as the 'then' part.

FCM and CPDA case: The input membership is partitioned by Fuzzy C-Means Clustering or CPDA; we generate 27 rules (3×3×3). The general rule is described as follows:

If x(*TAIEX (t)*) = A_i , y (*diff(N(t))*) = B_j and z (*diff(H(t))*) = C_k , then Output is f_l = p_l x + q_l y + r_l z+ s_l

Where x (*TAIEX (t)*), y (*diff(N(t))*), z (*diff(H(t))*) are linguistic features, A_i , B_j , C_k are the linguistic labels (high, middle, low), f_l denotes the *l-th* output value, and p_l , q_l , r_l , s_l are the

Table 1. Differences in features.

Date	NASDAQ	diff(*N(t)*)	Hang Seng	diff(*H(t)*)
2000/1/3	4131.15		17369.63	
2000/1/4	3901.69	-229.46	17072.82	-296.81
2000/1/5	3877.54	-24.15	15846.72	-1226.1
2000/1/6	3727.13	-150.41	15153.23	-693.49
2000/1/7	3882.62	155.49	15405.63	252.4
2000/1/8		155.49		252.4
2000/1/9		155.49		252.4
2000/1/10	4049.67	167.05	15848.15	442.52
2000/1/11	3921.19	-128.48	15862.1	13.95
2000/1/12	3850.02	-71.17	15714.2	-147.9
2000/1/13	3957.21	107.19	15633.96	-80.24
2000/1/14	4064.27	107.06	15542.23	-91.73
2000/1/15		107.06		-91.73
2000/1/16		107.06		-91.73
2000/1/17		107.06		-91.73
2000/1/18	4130.81	66.54	15789.2	246.97
2000/1/19	4151.29	20.48	15275.34	-513.86
2000/1/20	4189.51	38.22	15215.31	-60.03
2000/1/21	4235.4	45.89	15108.41	-106.9
2000/1/22		45.89		-106.9
2000/1/23		45.89		-106.9
2000/1/24	4096.08	-139.32	15167.55	59.14
2000/1/25	4167.41	71.33	15103.04	-64.51
2000/1/26	4069.91	-97.5	15427.72	324.68
2000/1/27	4039.56	-30.35	15917.81	490.09
2000/1/28	3887.07	-152.49	16185.94	268.13
2000/1/29		-152.49		268.13
2000/1/30		-152.49		268.13
2000/1/31	3940.35	53.28	15532.34	-653.6

parameters (*i=1, 2, 3; j=1, 2, 3; k* =1, 2, 3, and *l*= 1, 2,… , 27). The output membership function is constant when $p_l = q_l = r_l$ =0.

Subclust case: the input membership is partitioned by Subtractive Clustering; we generate three rules. The general rule is described as follows:

If x (*TAIEX (t)*) = A_i , y (*diff(N(t))*) = B_i and z (*diff(H(t))*) = C_i , then f_i = p_i x + q_i y + r_i z+ s_i

Where x (*TAIEX (t)*), y (*diff(N(t))*), z (*diff(H(t))*) are linguistic features, A_i , B_i , C_i are the linguistic labels (high, middle, low), f_i denotes the *i-th* output value, and p_i , q_i , r_i , s_i are the parameters (*i=1, 2, 3*).

The output membership function is constant when $p_i = q_i = r_i$ =0.

Step 7 (Train the fuzzy inference system): In this section, we employ a combination of the least-squares method and the back propagation gradient descent method for training the four types of forecasting models and use FIS membership function parameters to emulate a given training dataset. This paper sets the epoch as 1000 (the process is executed for a predetermined fixed number (1000) of iterations unless it terminates while the training error converges) for the training stopping criterion and then obtains the parameters for the selected output membership function.

Step 8 (Forecast the testing TAIEX (t+1) by the six types of forecasting models): Firstly, the FIS parameters of the four types of forecasting models are determined when the stopping criterion is reached from step 7; then the six training forecasting models are used to forecast *T (t+1)* for the target testing datasets, respectively. Secondly, calculate four RMSE values in the testing datasets by Equation (21).

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n} | actual(t) - forecast(t) |^2 \Big/ n} \qquad (21)$$

here actual (t) denotes the real TAIEX value, forecast (t) denotes the forecasting TAIEX value, and *n* is the number of data.

Step 9 (Select the best model and compare the results): Based on the minimal RMSE for the target testing datasets from Step 8, the best forecasting model among the six models can be obtained

Dependent Variable: CLOSE_PRICE
Method: Least Squares
Date: 04/21/08 Time: 16:56
Sample (adjusted): 6 238
Included observations: 233 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	118.3171	83.29947	1.420382	0.1569
CLOSE_PRICE(-1)	1.004497	0.066550	15.09386	0.0000
CLOSE_PRICE(-2)	0.038937	0.094724	0.411057	0.6814
CLOSE_PRICE(-3)	-0.064165	0.095438	-0.672327	0.5021
CLOSE_PRICE(-4)	-0.101954	0.096612	-1.055295	0.2924
CLOSE_PRICE(-5)	0.109017	0.068456	1.592510	0.1127

R-squared	0.978643	Mean dependent var	8536.073
Adjusted R-squared	0.978173	S.D. dependent var	817.4751
S.E. of regression	120.7747	Akaike info criterion	12.45114
Sum squared resid	3311143.	Schwarz criterion	12.54001
Log likelihood	-1444.558	F-statistic	2080.363
Durbin-Watson stat	1.984448	Prob(F-statistic)	0.000000

Figure 3. Testing the lag period of the TAIEX in 1997.

Table 2. Performance comparison of the different models (TAIEX).

Model	Year						
	1997	1998	1999	2000	2001	2002	2003
Yu's model (2005)	165	164	145	191	167	75	66
Chen's model (1996)	154	134	120	176	148	101	74
Proposed model	129[a]	117[a]	105[a]	148[a]	115[a]	67[a]	49[a]

[a] The best performance among the three models

obtained. Then, the minimal RMSE is taken as an evaluation criterion to compare with different models.

Evaluations and comparisons

To verify the proposed model, a seven-year period of the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), from 1997 to 2003, is selected from the Taiwan Stock Exchange Corporation (TWSE). A one-year period of stock data is defined as one unit of the experimental dataset. A 10-month period of stock data, from January to October, is defined as the training data, and the remaining period, from November and December, is used for testing After each sub-dataset experiment, 7 forecasting performances are generated for the 7 testing sub-datasets. Then, this paper compares the performances of the proposed model with the conventional fuzzy time-series model, Chen's (Takagi and Sugeno, 1983) model. Furthermore, to examine whether the proposed model surpasses the latest fuzzy time-series model, the performance of Yu's (2005) model is compared with the proposed model. The forecasting performances of Chen's model, Yu's model, and the proposed model are listed in Table 2. From Table 2, we can see that the proposed model outperforms the performances of the listing models.

RESULTS AND DISCUSSION

After verification and comparison, the proposed method is superior to the listing methods. However, some opinions can be discussed in this section. The degree of influence of a stock index in different countries, except the USA and Hong Kong, is worth being discussed. Because Taiwan is a typical island country that lacks energy, the degree of dependence on a global economic system is very high. Taiwan is a viable member of the international economic society; therefore, the impacts of world economic fluctuations on Taiwan are very high, too. There are many companies, such as TSMC (Taiwan Semiconductor Manufacturing Company), that publish ADRs (American Depositary Receipts) in American stock markets. Further, China became Taiwan's largest export market in 2002. So, we can see that the relation between the volatility of America and the volatility of the TAIEX and the relation between the volatility of China and the volatility of the TAIEX are very close. In near-neighbor countries, such as Japan and Korea, the degrees of influence on each other are more obvious, particularly. Thus, the proposed method will be verified further for the influence of different stock indexes (like the Nikkei 225, KOSPI) in different countries.

From the experiment results, there are two findings in this paper, as follows:

1. We can see that the best performance of the six types of forecasting models is subtractive clustering with linear type (AR (1)_Subclust_L). For example: one rule of AR(1)_Subclust_L model in the 1997 dataset is shown as follows:

x (*TAIEX (t)*) = A_{low}, y (*diff(N(t))*) = B_{low} and z (*diff(H(t))*) = C_{low}, then f_{low} = 3.306 x + 0.1251 y + 1.022 z -148.5

In the AR(1)_Subclust_L model, the parameters of the model do not equal 0 (output function). The meaning of information proves the relationship in the NASDAQ, Hang Seng, and TAIEX that American and Hong Kong stock indexes play an important role in affecting the volatility of the TAIEX. Further, this result is the same as the study of Dickinson (2000).

2. According to Table 2, the results show that the proposed model outperforms the listing methods in terms of RMSE. The main reasons why the proposed model surpasses Chen's (1996) model and Yu's (2005) model are as follows: (a) it considers that the volatility of American and Hong Kong stock indexes play an important role in affecting the volatility of TAIEX; (b) it utilizes the CDPA, FCM, and Subclust methods for discretizing features to obtain more objective membership functions; and (c) it uses adaptive networks to optimize the fuzzy inference system parameters.

Conclusions

A hybrid model, based on the AR method and multi-nation stock volatility causality, joining to fuse with the ANFIS procedure, is proposed to forecast stock price problems in Taiwan; further, the proposed model is compared with Chen's model and Yu's model to evaluate the results. This proposed model uses input features of a stock index (e.g., NASDAQ, Hang Seng and TAIEX) to forecast the TAIEX on the next trading day for investors. To illustrate the proposed model, three practical collected stock index datasets from the US, Hong Kong, and Taiwan stock markets (NASDAQ, Hang Seng, and TAIEX) are employed in this empirical experiment, which all consist of data from 1997 to 2003 (7 years in total). Each dataset includes data for 7 years and is spilt into 7 subdatasets based on year, respectively. The subdataset of each year for the first 10 months, January to October, is used for training, and the last 2 months, November to December, are used for testing. From Table 2, the experimental results indicate that the proposed model outperforms the listing models in terms of RMSE. Moreover, the results of this paper are helpful and feasible for stock investors, decision-makers, and future research. We believe that investors can utilize this forecasting model to discover the superior target of investments with benefits in the stock market.

For subsequent research, we can use other industries of datasets, such as Japan and Korea, to further validate the proposed model. In future work, there are two methods suitable for integration into the proposed model, which will improve the forecasting accuracies: (1) employ the corresponding ordered weighted averaging (OWA) weight to the values of features to enhance the performance of the proposed model (feature orderings are ranked by influence degrees on the TAIEX); and (2) other AI techniques can be joined with the proposed model to evaluate forecasting performance.

REFERENCES

Acklam PJ (2004). An algorithm for computing the inverse normal cumulative distribution function.

Aihara SI, Bagchi A, Saha S (2009). On Parameter Estimation of Stochastic Volatility Models from Stock Data Using Particle Filter-Application to AEX Index. Int. J. Innov. Comput. Inform. Control, 5(1): 17-28.

Anna B (2007). Should normal distribution be normal? The Student's T alternative, Computer Inform. Syst. Ind. Manage. Appl., pp. 3-8.

Bezdek JC (1981), Pattern Recognition with Fuzzy Objective Function Algorithms. NY: Plenum Press.

Bollerslev T (1986). Generalized autoregressive conditional heteroscedasticity. J. Econom., 31: 307-327.

Box G, Jenkins G (1976). Time series analysis: Forecasting and control, San Francisco: Holden-Day.

Chatfield C (2003). The Analysis of Time Series: An Introduction, Sixth Edition. Chapman and Hall/ CRC.

Chen SM (1996). Forecasting enrollments based on fuzzy time-series. Fuzzy Set Syst., 81: 311-319.

Chen SM, Chung NY, (2006). Forecasting Enrollments Using High-Order Fuzzy Time Series and Genetic Algorithms. Int. J. Intel. Syst., 21: 485-501.

Chen TL, Cheng CH, Teoh HJ, (2008). High-order fuzzy time-series based on multi-period adaptation model for forecasting stock markets, Physica A., 387: 876-888.

Chang JF, Chen KL, (2009). Applying New Investment Satisfied Capability Index and Particle Swarm Optimization to Stock Portfolio Selection. ICIC Express Lett., 3(3): 349-354.

Cheng CH, Chen TL, Chiang CH, (2006). Trend-weighted fuzzy time-series model for TAIEX forecasting. Lecture Notes Comput. Sci., 4234: 469-477.

Cheng CH, Wang JW, Li CH, (2008). Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix. Expert Syst Appl., 34: 2568-2575.

Chiu SL (1994). Fuzzy model identification based on cluster estimation. J. Intell. Fuzzy Syst., 2: 267-278.

Christensen R (1980). Entropy minimax sourcebook, general description. Lincoln MA: Entropy Limited.

Dickinson DG (2000). Stock market integration and macroeconomic fundamentals: An empirical analysis 1980-95. Appl. Financ. Econ., 10(3): 261-276.

Engle RF (1982). Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation. Econometrica, 50(4):987-1008.

Huarng KH (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Set Syst., 123: 155-162.

Huarng KH, Yu HK (2006). The application of neural networks to forecast fuzzy time series, Physica A., 336: 481-491.

Jang JS (1993). ANFIS: Adaptive-Network-based Fuzzy Inference Systems, IEEE Transact. Syst. Man Cybern., 23(3): 665-685.

Jilani TA, Burney SMA (2008). A refined fuzzy time series model for stock market forecasting. Physica A., 387: 2857-2862.

Liu HF, Hussain C, Tan MD (2002). Discretization: An enabling technique. Data Min Knowl. Disc, 6(4): 393-423.

Song Q, Chissom BS (1993). Forecasting enrollments with fuzzy time-series Part I. Fuzzy Set Syst., 54: 1-10.

Takagi T, Sugeno M (1983). Derivation of fuzzy control rules from human operator's control actions, in Proc. IFAC Symp. Fuzzy Inform., Knowledge Representation and Decision Analysis, pp. 55-60.

Takahama T, Sakai SA (2009). Hara and Noriyuki Iwane, Predicting Stock Price Using Neural Networks Optimized by Differential Evolution with Degeneration. Int. J. Innov. Comput. I., 5(12): 5021-5032.

Tuncay C, Stauffer D (2007). Power laws and Gaussians for stock market fluctuations. Physica A., 325-330.

Yu HK (2005). Weighted fuzzy time-series models for TAIEX forecasting, Physica A, 349: 609-624.

Yu HK, Huarng KH, (2008). A bivariate fuzzy time series model to forecast the TAIEX. Expert Syst. Appl., 34: 2945-2952.

Zhang Y, Chen Y (2009). Understanding the Price Fluctuations of Stock Markets through Cellular Automata. ICIC Express Lett., 3(3): 307-312.