ARTICLE IN PRESS
JID: NEUCOM [m5G;May 7, 2018;23:28]

Neurocomputing 000 (2018) 1–13

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Fuzzy time-series model based on rough set rule induction for forecasting stock price

Ching-Hsue Cheng*, Jun-He Yang

*Department of Information Management, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliu, Yunlin 64002, Taiwan ROC*

A R T I C L E I N F O

*Article history:*
Received 9 November 2016
Revised 15 September 2017
Accepted 3 April 2018
Available online xxx

Communicated by A. Abraham

*Keywords:*
Rough sets
Fuzzy time-series
Financial profits

A B S T R A C T

The stock price prediction is an important issue in stock markets because it will result in significant benefits and impacts for investor. In contrast to traditional time series, fuzzy time series can solve the forecast problem with historical data of linguistic values. In order to improve forecast performance of fuzzy time-series models, this study replaced fuzzy logical relationships with rule-based algorithm to extract forecast rules from time-series observations. Therefore, this paper developed a novel fuzzy time-series model based on rough set rule induction for forecasting stock index, and this study has four contributions to improve forecast accuracy and provide investment point (in right time) to investors:

(1) Proposed a novel fuzzy time-series model to improve forecast accuracy,

(2) rough sets are employed to generate forecasting rules to replace fuzzy logical relationship rules based on the lag period,

(3) utilized adaptive expectation model to strengthen forecasting performance, and based on the meaning of adaptive parameter to observe stock fluctuation and oscillation, and

(4) proposed buy and sell rules to calculate the profit and based on three different scenarios to provide investment suggestion to investor as references.

For evaluating the proposed model, we practically collected TAIEX, Nikkei, and HSI stock price from 1998 to 2012 years as experimental dataset, and compared the listing models under three error indexes and profits criteria. The results show that the proposed method outperforms listing models in error indexes and profits.



## 1. Introduction

Stock investing is an exciting and challenging monetary activity, and forecasting stock trend and price plays an important role in stock market. The stock investors could have a chance to make much money in stock returns with wise decisions; however the most investors keep a pessimistic image with a heavy loss of money in stock investment. Due to stock market behavior is nonlinearity and non-stationary, and the stock price fluctuation is extremely hard to predict correctly if without have experienced or expert knowledge. Up to date, it is difficult to build a general model for forecasting stock price accurately. Nevertheless, many researchers continue to establish feasible model for approximating stock market behaviors.

Forecasting activities play an important role in our daily life; the goal of forecasting activities is to increase accuracy and profit, such as chasing predict [38] and control prediction. In financial engineering, Kuo et al. [17] have demonstrated that the general techniques used for stock market prediction are mathematical and statistical models. In time-series analysis [14], there are many time series models such as ARIMA (Autoregressive Integrated Moving Average model, [3]) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity, [2]), these models have been applied to forecast stock price and trends in the financial market. However, statistical models usually deal with linear forecasting model and variables must obey statistical normal distribution for better forecasting performance. If the research data are represented by linguistic values (also named it, “linguistic intervals” such as linguistic values of age is very young, young, old) or the number of sample data is very little, the traditional forecasting methods maybe generate the bias of forecast or poor results. Therefore, many researchers have proposed different forecasting models based on fuzzy theory [37] to solve time-series problems with linguistic values.

Song and Chissom [27] first proposed a fuzzy time-series model to forecast the enrollments at University of Alabama; the fuzzy time-series model constructed the fuzzy relation $R$ and used a

* Corresponding author.
*E-mail address:* chcheng@yuntech.edu.tw (C.-H. Cheng).

https://doi.org/10.1016/j.neucom.2018.04.014


Please cite this article as: C.-H. Cheng, J.-H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.04.014

Max–Min composition operator to calculate forecasting values. Chen [4] proposed a fuzzy time-series model which used equal interval lengths to partition the universe of discourse and generate forecasting rules with a simplified calculation process. However, in stock forecasting, Huarng [12] extended Chen's model with additional heuristic forecasting rules to produce forecasts. In a subsequent study, Huarng proposed another model to define interval length with distribution based length and average-based length (2001). And Yu [35] proposed a weighted fuzzy time-series model with recurrent fuzzy relationships to produce forecasts. Sun et al. [30] did a prediction of Chinese stock index (CSI) future prices using fuzzy sets and multivariate fuzzy time series method. Aladağ et al. [1] proposed a partial high order Fuzzy lagged variable selection in fuzzy time series with genetic algorithms.

Because of fuzzy time-series are appropriately applied to linguistic values datasets for generating the higher accuracy. Recently, different fuzzy time-series models have been proposed to forecast nonlinear data and various applications, such as enrollment [4,27,28], temperature [6], car road accidents [13], tourism demand [18] and the stock index [5,8], etc. In the same way many researchers presented their fuzzy time-series methods to deal with the stock price forecasting.

To sum up, previous studies have four main limitations as follows: (1) most of previous studies must obey some assumptions about the variables used in the data analysis, so it is limited to be applied to all datasets; (2) most previous time series models used only one variable to forecast stock price; (3) the rules generated from ANN (Artificial neural network) are not easy to understand, and (4) VAR (Vector AutoRegression, [25]) can solve multiple variable time series problem, however there are some disadvantages of VAR: (i) the model selection procedure is complex; (ii) the model assumptions are difficult to validate; (iii) it requires a large amount of data for building model; (iv) it deals with linear model. To improve the deficiencies of previous fuzzy time-series models, and based on the advantages of LEM2 algorithm (learning from examples module, version 2, [11]), this study proposed a fuzzy time-series model based on rough set rule induction for forecasting stock index. This study has some contributions: (1) proposed a novel fuzzy time-series model to improve forecast accuracy, (2) The proposed model utilized rough set LEM2 algorithm to generate forecast rules, it's different from previous fuzzy time series using fuzzy logical relationships rule based on lag period (3) After getting the initial forecast of fuzzy time series, the adaptive expectation model is employed to strengthen forecasting performance, and observes the positive /negative value of adaptive parameter $h_0$ to analyze stock fluctuation and oscillation. (4) The main aim of investment is to obtain profits, this study proposed a profitable unit equation and the rules of selling and buying for investors to determine the trading time of selling and buying as references.

This rest of the paper is organized in the following. Section 2 introduces the related work such as fuzzy time-series model and rough set LEM2 algorithm. Section 3 presents the research concept and the proposed algorithm. The section 4 shows the experimental results and comparison. And the last section is conclusion.

## 2. The related work

In this section, fuzzy time-series model and rough set LEM2 algorithm are introduced briefly in the following.

### 2.1. Fuzzy time-series

In the traditional crisp set, the degree of an element belongs to a set is either one or zero. In order to deal with the uncertain and imprecise data, Zadeh [37] proposed fuzzy set theory. In thirty years ago, fuzzy set theory didn't incorporate into time-series, therefore time-series models had failed to treat the highly nonlinear and linguistic stock market behavior, until Song and Chissom [27] proposed fuzzy time-series model to forecast nonlinear data, that is, fuzzy time series model can treat the nonlinear and nonstationary problem. In recently years, there are many new fuzzy time-series models and applied in stock forecast (Sadaei et al. [31]; Rubio et al. 2017 [24]), these models usually utilized weighted fuzzy-trend or hybrid model.

Based on Song and Chissom's definitions, many fuzzy time-series models have been proposed, and these models have been successfully applied to deal with various applications. The related definitions of fuzzy time series are introduced as follows:

**Definition 1.** Fuzzy set

Let $U$ be the universe of discourse, $U=\{u_1,u_2,...,u_n\}$, A fuzzy set $A$ on $U$ is defined as

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

where $f_A$ is the membership function of fuzzy set A, $f_A$: $U \to [0, 1]$, $u_i$ is an element of fuzzy set A, $f_A(u_i)$ indicates the degree of membership of $u_i$ in A, $f_A(u_i) \in [0, 1]$ and $1 \le i \le n$.

**Definition 2.** Fuzzy time-series

Let $Y(t)$ ($t = \ldots, -2, -1, 0, 1, 2, \ldots$), a subset of a real number, be the universe of discourse in which the fuzzy sets denoted as $f_i(t)$ ($i = 1,2,\ldots$) are defined, and let $F(t)$ be a collection of $f_i(t)$ ($i = 1,2,\ldots$). Then $F(t)$ is a fuzzy time-series of $Y(t)$ ($t = \ldots, -2, -1, 0, 1, 2, \ldots$), where $-2$, $-1$ denotes one-lag period, two-lag period.

**Definition 3.** Fuzzy logical relationship

If there exists a fuzzy logical relationship (FLR), the relationship can be expressed as $F(t) = F(t-1) \times R(t-1, t)$, where $\times$ denotes an operation, then $F(t)$ is said to be caused by $F(t\text{-}1)$. The logical relationship between $F(t)$ and $F(t\text{-}1)$ is denoted as

$$\mathrm{F}(t-1) \to F(t).$$

**Definition 4.** Fuzzy logical relationship for two consecutive fuzzy variables

Let $F(t\text{-}1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive fuzzy variables, $F(t)$ and $F(t\text{-}1)$, referred to as a FLR, can be denoted by $A_i \to A_j$, where $\mathrm{A_i}$ is the left-hand side and $A_j$ is the right-hand side of the FLR.

### 2.2. Rough set theory (RST)

RST is a nonparametric technique that has foundations in mathematical set theory and has been widely applied to decision problems [32]. RST was developed by Pawlak and Skoworn [21] and has been accepted as an effective mathematical tool for modeling vagueness and uncertainty. This approach is particularly importance in artificial intelligence research and cognitive sciences, especially in machine learning, knowledge discovery from databases, data mining, support systems, inductive reasoning and pattern recognition [16,20,22]. Recently, RST has been applied in varied fields such as medical diagnosis, drug research, process control, credit fraud detection, bankruptcy prediction, stock market rule-generation, climate change, and the development of expert systems for The NASA Space Center.

Rough set philosophy is founded on the assumption that, in the universe of discourse associated with every object, some information objects characterized by the same information are indiscernible because of insufficient information. Any set of all indiscernible objects is called an elementary set and forms a basis granule of knowledge about the universe. Any union of elementary sets is referred to as a precise set; otherwise, the set is considered rough. Rough set theory incorporates the use of indiscernibility (equivalence) relations to approximate sets of objects by lower and upper approximations [21]. The lower approximation consists of all objects that surely belong to concept X, and the upper approximation contains all objects that possibly belong to concept X. The difference between the upper and lower approximation constitutes the boundary region of concept X.

The rule induction algorithm LEM2 (Learning from Examples Module version 2, [11]), is most frequently used since in most cases it gives better results. LEM2 explores the search space of attribute-value pairs. In general, LEM2 computes a local covering and then converts it into a rule set. LEM2 learns a discriminant rule set; it learns the smallest set of minimal rules describing a concept. This algorithm can generate both certain and possible rules from a decision table. Therefore, LEM2 algorithm has some advantages as follows:

(1) LEM2 is suitable for rule generation for inconsistent data,
(2) This algorithm can generate both certain and possible rules from a decision table,
(3) LEM2 learns the smallest set of minimal rules for every concept.
(4) LEM2 represent learning from examples, an approach of similarity-based learning.

Lastly, rough set applied in data analysis has many important advantages ([26,39,40], (b). Some of them are listed below:

(1) Rough set discovers important facts hidden in data and expresses them in the natural language of decision rules.
(2) Rough set accepts both quantitative and qualitative attributes and specifies their relevance for approximation of the classification.
(3) Rough set can contribute to the minimization of the time and cost of the decision making process (rough set approach is an information processing system in real time.)
(4) Rough set offers transparency of classification decisions, allowing for their argumentation.
(5) Rough set takes into account background knowledge of the decision-maker.
(6) Rough set can be incorporated into an integrated DSS for the evaluation of corporate performance and viability.

### 2.3. Adaptive expectation model

The Adaptive Expectations model is economic agents who develop forecasts of future inflation based on past actual rates adjusted for their own past expectations. And Adaptive expectations is an economic theory which gives importance to past events in predicting future outcomes. A common example is for predicting inflation. Adaptive expectations states that if inflation increased in the past year, people will expect a higher rate of inflation in the next year. In time series forecasting, the adaptive expectation model [15] has shown to be a reasonable forecast model in stock price, the forecasts are generated by the last one period of stock price and the correction for last one period of forecasting error as Eq. (1).

$$\text{Forecast}\ (t+1) = \text{P}\ (t) + h_0 * \varepsilon\ (t) \quad (1)$$

where forecast $(t+1)$ is the forecasting stock price at time $t+1$, $\text{P}(t)$ is the real stock price at time $t$, $h_0$ is the adaptive parameter for $\varepsilon(t)$, and $\varepsilon(t) = \text{Forecast}(t) - \text{P}(t)$ is forecasting error at time $t$.

## 3. Proposed model

The main objective of forecasting activities is to make correct investment and more profit. In the past, there are many time series models, these models with four main limitations are described in Section 1. To improve the limitations of previous fuzzy time-series models, this study proposed a fuzzy time-series model based on rough set rule induction for forecasting stock index, and the proposed model could improve forecast accuracy and provide investment suggestion to investor as references.

The first Fuzzy time-series [27] could be effectively applied in linguistic values, non-linearity, and non-stationary datasets. And three critical processes will affect forecast accuracy in fuzzy time-series: (1) determine the universe of discourse and the lengths of linguistic intervals, (2) extract fuzzy logical relationships from time-series data, (3) defuzzify linguistic forecasts. In addition, previous studies [4] have ignored the means of recurrent fuzzy relationship hidden in time series, and advantages of LEM2 algorithm have discussed in Section 2.2, hence this study skipped fuzzy logic logical relationships method, direct used rule-based algorithm (rough set LEM2 algorithm) to extract forecast rules from time series observations. Based these reasons, this study proposed a new fuzzy time-series model based on rough set rule induction for forecasting stock index.

For evaluating the forecasting performance of proposed models, three Asian stock markets: TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index), Nikkei (Japanese stock market index), and HIS (Hong Kong - Heng Seng Index) from 1998 to 2012 years are used as experimental dataset. In comparison, five models are compared with proposed model under the root mean square error (RMSE), root relative squared error (RRSE), relative absolute error (RAE) and profit criterion. The five models are: (1) Chen's [4] model, (2) Yu's [36] model, (3) Stepwise regression (SR) combined with adaptive neuro fuzzy inference system [33], named it as SR + ANFIS, (4) SR combined with support vector regression [23], called it as SR + SVR, and (5) Elman recurrent neural network [7,9]. In stock market, sometimes the higher accuracy will not bring more profit, and the key issue for investor considerations is to get more profit. Therefore, this paper proposed a buy-sell rule to calculate the profit of investment as evaluation criteria.

For easily understanding the proposed model, the procedure of proposed model was partitioned into four phases and one block of evaluation/ comparison (as Fig. 1). The phase 1 preprocess contains Step 1 to Step 3, phase 2 is Step 4 LEM2 generating rules, phase 3 is Step 5 initial forecast, phase 4 includes Step 6 adaptive adjustment forecast, and the last phase is evaluation including Step 7 and Step 8. In this proposed model, the Step 4 and Step 8 are new proposed steps, other steps are followed previous works. Moreover, the iterative loops of the proposed procedure in Fig. 1 occurs in single step: (1) Step 4 to get optimal rule-by-rule filter, (2) Step 6 to find the optimal adaptive parameter ($h_0$) in minimal RMSE, and (3) Step 8 to obtain the optimal parameter $\alpha$ for optimal profits.

### 3.1. Proposed algorithm

This study proposed an algorithm for ease of computing, and future results can be followed in the proposed model. The algorithm contained eight steps. The first six step is proposed model; the remained Step 7 and Step 8 are evaluation and comparison. The detailed algorithm is introduced step-by-step in the following.

**Step 1:** *Data collection and transformation.*

This step collected TAIEX, Nikkei, and HSI stock price from 1998 to 2012 years data (each stock market with 15 sub-datasets) to illustrate the proposed model. From the previous studies (Kao, & Chen, 2013), this study selected the importance variable by high occurrence frequency in literature, and the important variables are lag periods, moment, and second moment in statistical method, hence five independent variables are selected and close price as dependent variable as Table 1. After determined the research vari-

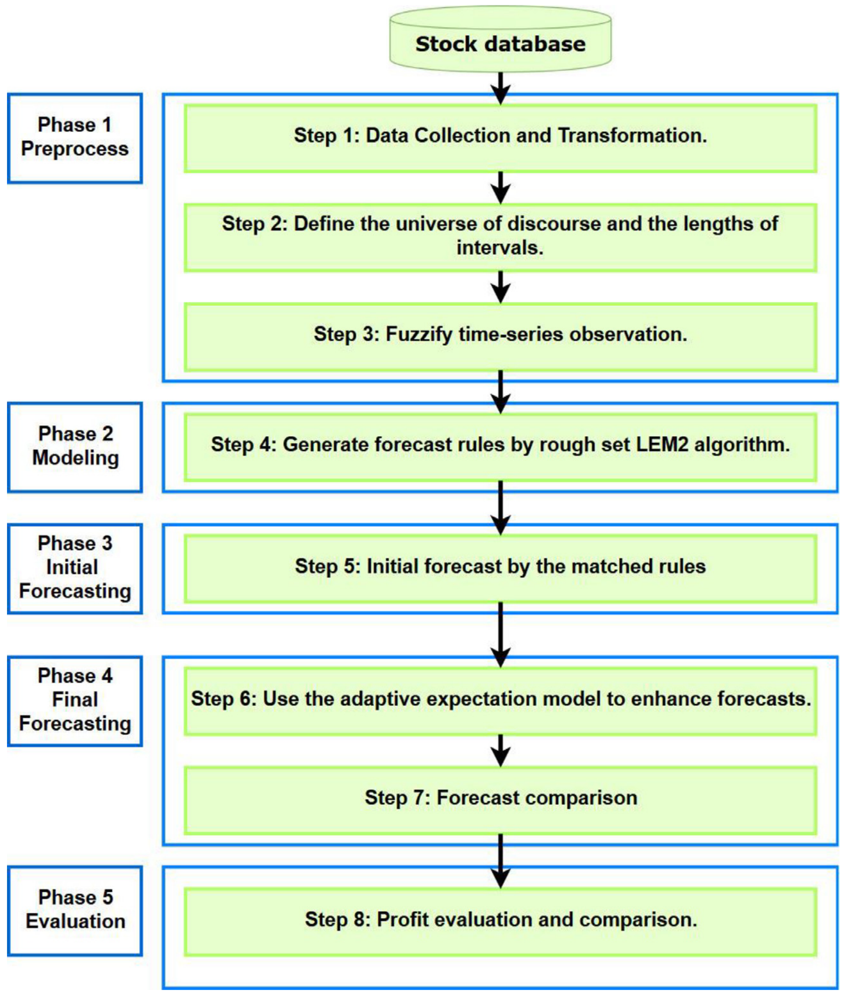


**Fig. 1.** The procedure of proposed model.

**Table 1**
The selected variables.

| Factors | Explanation |
|---|---|
| LAG1 | First-order lagged period. ($\text{Price}_{t-1}$) |
| LAG2 | Second-order lagged period. ($\text{Price}_{t-2}$) |
| LAG3 | Third-order lagged period. ($\text{Price}_{t-3}$) |
| Moment ($V$) | Moment $V_t$ denotes the difference between trading price $\text{day}_t$ and trading price $\text{day}_{t-1}$, it is defined as: $V_t = \text{Price}_t$ - $\text{Price}_{t-1}$ where $\text{Price}_t$ and $\text{Price}_{t-1}$ denote the on trading $\text{day}_t$ close price and trading $\text{day}_{t-1}$ close price, respectively. |
| Moment slope ($S$) | Moment slope $S_t$ is the difference between moment $V_t$ and moment $V_{t-1}$ over the unit time (day), it equation is defined as: $S_t = \frac{(V_t - V_{t-1})}{t-(t-1)}$ |
| Price ($P$) | The close price in trading $\text{day}_t$ |

ables, this step transforms the daily close price into the selected variables of Table 1.

Step 2: Define the universe of discourse and the lengths of intervals.

Define the universe of discourse $U$ and the $U$ is partitioned into n intervals, the detailed sub-steps are described as follows:

Step 2.1: Define the universe of discourse.

Let $D_{\min}$ and $D_{\max}$ be the minimal value and the maximal value of a specific historical dataset, respectively. Then, the universe of discourse $U$ can be defined as $[D_{\min}\text{-}D_1, D_{\max}+D_2]$, where $D_1$ and $D_2$ are two adaptive positive real values (to assure the least and largest interval could cover the $D_{\min}$ and $D_{\max}$) to partition the universe of discourse $U$ into $n$ equal length intervals, $u_1$, $u_2$, ... and $u_n$.

Step 2.2: Define the linguistic intervals.

From the research of Miller [10], the appropriate number of category for human shorten memory function is seven, or seven plus or minus two. Hence, this study employed $7 \pm 2$ as the linguistic values (also named it, linguistic intervals), the universe of discourse U for each variable is partitioned into seven linguistic intervals with equal length, $u_1$, $u_2$, $u_3$, $u_4$, $u_5$, $u_6$, $u_7$ (where use seven as example) and the length $L$ could be defined as follows:

$$L = [D_{\min} - D_1,\ D_{\max} + D_2]/n. \tag{2}$$

**Step 3:** *Fuzzify time-series observation.*

In this step, let fuzzy set **A** denote fuzzy LAG1 variable, the other five fuzzy variables are expressed as **B** (LAG2), **C** (LAG3), **V, S**, and **P**. Fuzzy set **A** can be partitioned into a given linguistic values (the corresponding interval), such as $\mathbf{A_1}$, $\mathbf{A_2}$, $\mathbf{A_3}$, $\mathbf{A_4}$, $\mathbf{A_5}$, $\mathbf{A_6}$ and $\mathbf{A_7}$. Each $\mathbf{A_i}$ is expressed by the intervals $u_i$, which is defined

Please cite this article as: C.-H. Cheng, J.-H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.04.014

as Eq. (3):

$$\begin{aligned}
\mathbf{A_1} &= \{1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7\} \\
\mathbf{A_2} &= \{0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7\} \\
\mathbf{A_3} &= \{0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + 0/u_7\} \\
\mathbf{A_4} &= \{0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0/u_6 + 0/u_7\} \\
\mathbf{A_5} &= \{0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 + 0/u_7\} \\
\mathbf{A_6} &= \{0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6 + 0.5/u_7\} \\
\mathbf{A_7} &= \{0/u_1 + 0/_{u2} + 0/u_3 + 0/u_4 + 0/u_5 + 0.5/u_6 + 1/u_7\} \qquad (3)
\end{aligned}$$

where the symbol "+" denotes union operator.

**Step 4:** *Generate forecast rules by rough set LEM2 algorithm*

This step aims to provide a set of meaningful rules of forecast stock price. Based on decision rules are generated by using rough set LEM2 algorithm, the generated rules are formed "if-then" by composing a several fuzzy conditional value (interval) and fuzzy decision values, and the "supports" denotes the number of records in dataset matching with the generated decision rules. After generating rules of LEM2 algorithm, this study iteratively deletes the rules of less supports to filter the generating rules, and the optimal fuzzy rules is remained by the higher classification accuracy. According to the format of Liu et al. [19], the pseudocode of LEM2 algorithm was listed in Algorithm 1, and the algorithm was described in the followings.

The option LEM2 of LERS (Learning from Examples using Rough Sets), is very frequently employed in many cases due to it gets better results. LEM2 input data is a lower or upper approximation of a concept [11], and its input data is always consistent. The LEM2 algorithm follows a classical greedy scheme, it covers all examples from the given approximation using a minimal set of rules, and furthermore LEM2 calculates a local covering and then converts it into a rule set.

Next, the detailed symbolic system is introduced as follows. For an attribute-value pair $(a;\ v) = t$, a *block* of $t$, denoted by $[t]$, is a set of all instances from $U$ such that for attribute $a$ has value $v$. Let $B$ be a nonempty lower or upper approximation of a concept represented by a decision-value pair $(d;w)$. The set $B$ *depends* on a set $T$ of attribute-value pairs $t = (a;\ v)$ if and only if $\emptyset \neq [T] = \bigcap_{t\in T}[t] \subseteq B$.

Where set $T$ is a minimal complex of $B$ if and only if $B$ depends on $T$ and no proper subset $T'$ of $T$ exists such that $B$ depends on $T'$. Let $\Gamma$ be a nonempty collection of nonempty sets of attribute-value pairs. Then $\Gamma$ is a local covering of B. Furthermore, the detailed tutorial could refer the work of Grzymala-Busse [11,29].

**Algorithm 1**
The pseudocode of LEM2 algorithm [11].

```
input: a set B,
output: a single local covering Γof set B);
begin
      G:=B;
      Γ:=∅;
      while G ≠∅
            begin
            T:=∅;
            T(G):={t|[t]∩G ≠∅};
            while T=∅; or [T] not⊆ B
                 begin
                       select a pair t ∈ T(G) such that |[t]∩G| is
                       maximum; if a tie occurs, select a pair t ∈ T(G)
                       with the smallest cardinality of [t];
                            if another tie occurs, select first pair;
                                 T := T ∪ {t};
                                 G := [t] ∩ G;
                                 T(G):={t|[t]∩G ≠∅};
                                 T(G):=T(G) − T;
            end {while}
                 for each t ∈T do
                   if [T − {t}] ⊆ B then T:=T −{t};
                   Γ := Γ∪{T};
                  G := B − ⋃_{T∈Γ} [T]];
            end {while};
            for each T∈Γ do
                   if ⋃_{S∈Γ−[T]} [S]=B then Γ = Γ − [T];
  end {procedure}.
```

**Step 5:** *Initial forecast by the matched rules*

This step included two sub-steps: forecast and defuzzification, which are introduced as follow.

Step 5.1: Forecast

According to the Step 4, the fuzzy variables of the testing data in trading $\text{Day}_t$ can use the generated rules to find out the decision linguistic values $\boldsymbol{P_i}$ of the matched rule as Table 5. If there are no linguistic values in fuzzy variables of the testing data of trading $\text{Day}_t$, this study will add a new linguistic values (interval) for each fuzzy variables in order to more reasonable forecast consideration.

Step 5.2: Defuzzy

In defuzzification, this step utilizes the middle point of the decision linguistic values $\boldsymbol{P_i}$ as the defuzzified value. That is, if the linguistic values in fuzzy variables of the testing data of trading $\text{Day}_t$ was matched with generated rules, then we can find out the linguistic value of decision value is $\boldsymbol{P_i}$, the linguistic intervals of $\boldsymbol{P_i}$ is [a, b], and the middle point of the $\boldsymbol{P_i}$ is (a + b)/2.

**Step 6**: *Use the adaptive expectation model to enhance forecasts.*

From Eq. (1), this step optimizes the adaptive parameter ($h_0$) under minimal RMSE for the initial forecast of training data. The iterated step is 0.0001 from 0 to 1 to search the optimal adaptive parameter ($h_0$). This step uses minimal RMSE as evaluation criterion to find the optimal $h_0$, and the RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(forecasted\ value_i - actual\ value_i\right)^2}{n}} \qquad (4)$$

where n denotes the number of trading days needed to be forecasted. After the optimal parameter ($h_0$) is obtained, the initial forecast of testing dataset is enhanced by Eq. (1).

From Wei et al. [34], Test the lag period of TAIEX is one-period lag, hence this paper only employs the simple one-period lag adaptive expectation model.

**Step 7:** *Forecast comparison*

In order to compare forecasting performance of the proposed model with those of the listing models, this step uses the daily TAIEX, Nikkei, and HSI closing prices from 1998 to 2012 as the verification dataset. Each year of experimental dataset is splits into two subsets, each year the previous ten-month from January to October is training data, and November and December is testing data. This step also compared the proposed model with the listing models in forecasting performance, and RMSE, RRSE, and RAE are employed as evaluation criterion. The RMSE had been defined in Eq. (4), The RRSE and RAE are defined as Eq. (5), (6).

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}\left(forecasted\ value_i - actual\ value_i\right)^2}{\sum_{i=1}^{n}\left(\ actual\ value_i - average\ of\ actual\ value_i\right)^2}} \qquad (5)$$

$$RAE = \frac{\sum_{i=1}^{n}\left|forecasted\ value_i - actual\ value_i\right|}{\sum_{i=1}^{n}\left|actual\ value_i - average\ of\ actual\ value_i\right|} \qquad (6)$$

Furthermore, Due to the squared value of the correlation coefficient is defined as:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}\left(\ actual\ value_i - average\ of\ actual\ value_i\right)^2}{\sum_{i=1}^{n}\left(forecasted\ value_i - actual\ value_i\right)^2}$$

Based on Eq. (5), we can find $R^2 = 1 - \frac{1}{(RRSE)^2}$, hence this paper only use RMSE, RAE, and RRSE as evaluation criteria.

**Step 8:** *Profit evaluation and comparison.*

This step proposed a profit evaluation method and compared with the listing models. The TAIEX, Nikkei, and HSI from 1998 to 2012 (each market with 15 sub-datasets) were employed as evaluation data, each sub-dataset for the first 10-month data is used for training, and remained November to December as testing data. For evaluating profits, this step set two trade rules to calculate the profit of TAIEX as follows:

*Rule 1: sell rule*

$$\text{IF } \frac{|\text{forecast}(t) - \text{actual}(t)|}{\text{actual}(t)} \leq \alpha \text{ And } \text{ Forecast } (t+1) \\ -\text{actual}(t) > 0 \text{ then sell.}$$

*Rule 2: buy rule*

$$\text{IF } \frac{|\text{forecast}(t) - \text{actual}(t)|}{\text{actual}(t)} \leq \alpha \text{ And } \text{ Forecast } (t+1) \\ -\text{actual}(t) < 0 \text{ then buy.}$$

Where $\alpha$ is daily price fluctuation and threshold parameter, this study set 0.001 as the iterated step, and search the optimal parameter $\alpha$ from 0 to 0.3 ($0 < \alpha \leq 0.3$). Because the daily price fluctuation is smaller than 20% in TAIEX, HIS, and Nikkei stock market, this paper set $\alpha \leq 0.3$ to assure the range of relative absolute error.

And let the profits unit is equal to one, the profit formula is defined as Eq. (7).

$$\text{Profit} = \sum\nolimits_{t\_s=1}^{p} (actual(t+1) - actual(t)) \\ + \sum\nolimits_{t\_b=1}^{q} (actual(t) - actual(t+1)) \quad (7)$$

where $p$ represents the total number of days for selling, $q$ represents the total number of days for buying, $t\_s$ represents the $t$th day for selling and $t\_b$ represents the $t$th day for buying. The optimal threshold parameter $\alpha$ is obtained when maximal profit is reached in training dataset. Furthermore, when the profit is negative in training dataset, this study will not sell and buy in testing data and set the profit is zero.

Furthermore, this paper proposed three different scenario based on buy and sell rule to calculate and compare their profit by proposed model for the best $\alpha$, $\alpha = 0.01$, and $\alpha = 0.1$, respectively, and also provide investment suggestion to investor as references.

## 4. Experiment and comparison

This section verifies the forecast performance of proposed model, and then compares the proposed model with the listing models in three forecast measure indexes and profit. The experimental dataset are practically collected from three Asian stock markets: TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index), HSI (Hong Kong - Heng Seng Index), and Nikkei (Japanese stock market index). Each collected stock dataset from 1998 to 2012 has 15 sub-datasets, each year data is called as sub-dataset. And each sub-dataset is partitioned into ten-month/two-month for training/testing, that is each sub-dataset from January to October is training data, and the remained data from November to December is testing data. The listing comparison model includes Chen's [4] model, Yu's [36] model, Stepwise regression combined with ANFIS (SR + ANFIS), Stepwise regression combined with support vector regression (SR + SVR) and Elman recurrent neural network [7,9], and the comparison criteria are RMSE, RRSE, RAE and profit.

### 4.1. Forecast verification

This section has two parts: (A) demonstrate the proposed model by using one sub-dataset, (B) experiment the forecast performance of 7 and 9 linguistic intervals and whether rules filtered will have the better forecast performance.

**(A) Demonstrating the proposed model**

Based on the proposed algorithm in Section 3.2, the proposed model is Step 1 to Step 6. This part uses 1998-year TAIEX sub-dataset as example to illustrate the proposed model.

(1) The 1998-year sub-dataset, which contains 271 transaction days. Each year the training data is from January to October, and the remaining data (November and December) is used for testing. The research variables as Table 1 in Section 3.2 include five independent variables and close price as dependent.

(2) Define the universe of discourse U and the U is partitioned into 7 intervals, where use the Price as an example, the $D_{\min}$ and $D_{\max}$ in the training dataset (1998/1/4 to 1998/10/31) are 6251.38 and 9277.09, respectively. This study takes percentiles, and then the parameters let $D_1 = 51.38$ and $D_2 = 22.91$, such that the universe of discourse is $U = [6200, 9300]$. Therefore, based on Eq. (2), the seven linguistic intervals were calculated, the results were listed in Table 2.

(3) Fuzzify time-series observation based on Step 3 in section 3.2, and the partial fuzzified variables in sub-dataset are shown in Table 3.

(4) From the selected variable of Table 1 and the linguistic intervals of Table 2, this paper transformed into Table 3. And based on Step 3, the fuzzified observations exclude index variable (second column in Table 3), then Table 3 transform into Table 4 for data format of rough set. Next the decision rules are generated by using rough set LEM2 algorithm, the partial generated rules are shown in Table 5. The generated rules are formed "if-then" by composing a several fuzzy conditional value (interval) and fuzzy decision values, and the "supports" denotes the number of records in dataset matching with the generated decision rules, such as rule #1 in Table 5 is 21. This study deletes the rules of supports $< 2$ to filter the generating rules, *i.e.* the rules of supports $\geq 2$ are the optimal fuzzy rules with the higher classification accuracy. The generalized rules by "if-then" format could be expressed as follows:

$$If\ (condition = \boldsymbol{A_3}\ And\ \boldsymbol{B_3}\ And\ \boldsymbol{C_3}\ And\ \boldsymbol{V_5})\ Then\ (decision = \boldsymbol{P_3}).$$

(5) According to the Step 4, the fuzzy variables of the testing data in trading $\text{Day}_t$ can use the generated rules to find out the decision linguistic values $\boldsymbol{P_i}$ of the matched rule as Table 6. If there are no linguistic values in fuzzy variables of the testing data of trading $\text{Day}_{t,}$ this study will add a new linguistic values (interval) for each fuzzy variables in order to more reasonable forecast consideration. In defuzzification, this step utilizes the middle point of the decision linguistic values $\boldsymbol{P_i}$ as the defuzzified value. For example, if the linguistic values in fuzzy variables of the testing data of trading $\text{Day}_t$ was matched with rule #1 in Table 5, then we can find out the linguistic value of decision value is $\boldsymbol{P_3}$, the linguistic intervals of $\boldsymbol{P_3}$ is 7085 to 7528, and the middle point of the $\boldsymbol{P_3}$ is $(7085 + 7528)/2 = 7307$. Therefore, the defuzzified value of the testing data of trading $\text{Day}_t$ is 7307. The partial forecast results $D(t)$ are shown in the third column of Table 6.

(6) After the optimal parameter ($h_0$) is obtained, the initial forecast of testing dataset is enhanced by Eq. (1). For example, use the initial forecast of training data (1998/01–1998/10) to

**Table 2**
The seven linguistic intervals of training dataset (1998-year TAIEX).

| Linguistic interval | LAG1 | LAG2 | LAG3 | *V* | *S* | Price |
|---|---|---|---|---|---|---|
| $u_1$ | [6200,6643] | [6200,6643] | [6200,6643] | [−400,−286] | [−400,−257] | [6200,6643] |
| $u_2$ | [6643,7086] | [6643,7086] | [6643,7086] | [−286,−171] | [−257,−114] | [6643,7086] |
| $u_3$ | [7086,7529] | [7086,7529] | [7086,7529] | [−171,−57] | [−114,29] | [7086,7529] |
| $u_4$ | [7529,7971] | [7529,7971] | [7529,7971] | [−57,57] | [29,171] | [7529,7971] |
| $u_5$ | [7971,8414] | [7971,8414] | [7971,8414] | [57,171] | [171,314] | [7971,8414] |
| $u_6$ | [8414,8857] | [8414,8857] | [8414,8857] | [171,286] | [314,457] | [8414,8857] |
| $u_7$ | [8857,9300] | [8857,9300] | [8857,9300] | [286,400] | [457,600] | [8857,9300] |

**Table 3**
The partial fuzzified observations (1998-year TAIEX).

| Date | Index | *Linguistic variables* | | | | | |
|---|---|---|---|---|---|---|---|
| | | LAG1 | LAG2 | LAG3 | *V* | *S* | Price |
| 1998/1/3 | 8028.63 | | | | | | $P_6$ |
| 1998/1/5 | 7966.18 | $A_6$ | | | $V_4$ | | $P_5$ |
| 1998/1/6 | 7835.56 | $A_5$ | $B_6$ | | $V_4$ | $S_4$ | $P_5$ |
| 1998/1/7 | 7778.16 | $A_5$ | $B_5$ | $C_6$ | $V_4$ | $S_5$ | $P_5$ |
| 1998/1/8 | 7737.20 | $A_5$ | $B_5$ | $C_5$ | $V_5$ | $S_4$ | $P_5$ |
| … | … | … | … | … | … | … | … |
| 1998/10/29 | 7101.46 | $A_3$ | $B_3$ | $C_3$ | $V_6$ | $S_5$ | $P_4$ |
| 1998/10/30 | 7165.98 | $A_4$ | $B_3$ | $C_3$ | $V_6$ | $S_4$ | $P_4$ |
| 1998/10/31 | 7218.09 | $A_4$ | $B_4$ | $C_3$ | $V_5$ | $S_4$ | $P_4$ |

**Table 4**
The partial data of rough set format.

| Date | *Linguistic variables* | | | | | |
|---|---|---|---|---|---|---|
| | LAG1 | LAG2 | LAG3 | *V* | *S* | Class |
| 1998/1/7 | $A_5$ | $B_5$ | $C_6$ | $V_4$ | $S_5$ | $P_5$ |
| 1998/1/8 | $A_5$ | $B_5$ | $C_5$ | $V_5$ | $S_4$ | $P_5$ |
| 1998/1/9 | $A_5$ | $B_5$ | $C_5$ | $V_2$ | $S_2$ | $P_4$ |
| 1998/1/12 | $A_4$ | $B_5$ | $C_5$ | $V_6$ | $S_8$ | $P_4$ |
| 1998/1/13 | $A_4$ | $B_4$ | $C_5$ | $V_7$ | $S_5$ | $P_5$ |
| … | … | … | … | … | … | … |

optimize the parameter ($h_0$) and the enhanced forecast results are shown in the last column of Table 6. From adaptive forecast, the best result is RMSE = 111.126 and $h_0 = 0.07$. Within the best adaptive parameter ($h_0 = 0.07$), we forecast TAIEX index for testing data (1998/11–1998/12). The partial forecast results of TAIEX are shown in the last column of Table 6.

(B) *Experiment of linguistic intervals and whether filtered rules*

For verifying the forecast performance of different numbers of linguistic interval, this part employs Miller's [10] magic number 7 (plus or minus two) to determine the lengths of linguistic intervals, and whether rule filtering of the rough set improves the forecasting performance. From proposed algorithm in Section 3.2, the TAIEX and HIS dataset are conducted to compute their forecast results for testing data as Table 7. Each stock market has 9 results of testing sub-datasets in different number of linguistic intervals. From Table 7, we can see that on the whole the nine linguistic intervals with rule filter is better forecast for TAIEX and HIS in 1998–2006 year. From the results of Yu [36], the more linguistic intervals utilize the more get forecast performance. And too many linguistic intervals would disturb human shorten memory function for investors. Therefore, this study suggests that nine linguistic intervals with rule filter is a good approach.

### 4.2. Forecast comparison

To experiment and compare more datasets, the study practically collected TAIEX, Nikkei, and HSI data from 1998 to 2012 (each stock market with 15 sub-datasets). Then this study used five different time series models to compare with the proposed model. The reasons why utilize the five models? Due to Chen's model [4] and Yu's model [36] are typical fuzzy time series model, and many researches would be compared with their models. Furthermore, SR variable selection combined ANFIS [33] and SVR [23] are popular machine learning method, and SVR is the extension of support vector machine for solving nonlinear regression estimation problems. The idea of SVR is based on the computation of a linear regression function in a high dimensional feature space where the input data are mapped via a nonlinear function. Lastly, Elman ANN [9] is a simple example of a recurrent neural network, which is composed of three layers, and its structure contains at least one hidden layer from which the feedback is led. And Elman ANN with layer prediction is possible in sequential order. Therefore, Elman ANN is usually employed to build time series model.

Based on performance comparison, this paper employed RMSE, RRSE, and RAE as evaluation criteria. Based on the proposed algorithm, the three collected stock datasets are trained and tested, the forecast evaluation in RMSE, RRSE, and RAE are shown in Table 8-10. Table 8 indicates that the winning ratio ($\frac{\text{win subdatasets}}{\text{15 subdatasets}}$) is TAIEX = 11/15, HIS = 12/15, and Nikkei = 14/15 in RMSE, respectively. Table 9 shows that the winning ratio is TAIEX = 11/15, HIS = 12/15, and Nikkei = 14/15 in RRSE, respectively. And Table 10 presents that the winning ratio is TAIEX = 13/15, HIS = 14/15, and Nikkei = 11/15 in RAE, respectively. It is clearly, the proposed model outperforms the listing models under the RMSE, RRSE, and RAE. For showing visional view, the forecast trend of actual and predicted values for 1998 testing data of TAIEX is shown in Fig. 2, it shows that the forecast value of proposed model is closer the actual value. Moreover, the forecast results for average RMSE in 1998–2012 year is also illustrated as Fig. 3, clearly, the proposed model is ranked first name, its RMSE is less than the listing models, and the Elman RNN is ranked as second good model.

**Table 5**
The partial generated rules for 1998′ TAIEX dataset.

| No. | Rules | Supports |
|---|---|---|
| 1 | *If (LAG1 = $A_3$) And (LAG2 = $B_3$) And (LAG3 = $C_3$) And (V = $V_5$) Then (Class = $P_3$)* | 21 |
| 2 | *If (LAG1 = $A_5$) And (LAG2 = $B_5$) And (LAG3 = $C_5$) And (V = $V_5$) Then (Class = $P_5$)* | 18 |
| 3 | *If (LAG1 = $A_6$) And (V = $V_5$) Then (Class = $P_6$)* | 13 |
| 4 | *If (LAG1 = $A_8$) And (LAG2 = $B_8$) And (LAG3 = $C_8$) And (V = $V_5$) And (S = $S_4$) Then (Class = $P_8$)* | 10 |
| 5 | *If (LAG1 = $A_6$) And (LAG2 = $B_6$) And (LAG3 = $C_6$) And (V = $V_5$) Then (Class = $P_6$)* | 8 |
| … | … | … |

Please cite this article as: C.-H. Cheng, J.-H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.04.014

**Table 6**
The processes of adaptive forecasts for 1998-year TAIEX testing sub-dataset.

| Date | $P(t)$ | $D(t)$ | $\varepsilon(t)$= $P(t)$-$D(t)$ | $P(t+1)$= $P(t)+h_0{}^*\,\varepsilon(t)$ | Enhanced Forecast |
|---|---|---|---|---|---|
| 1998/11/2 | 7071.44 | 7218.09 | 146.65 | 7071.44 + 0.07*146.65 | 7229.079 |
| 1998/11/3 | 6905.32 | 7071.44 | 166.12 | 6905.32 + 0.07*166.12 | 7061.175 |
| 1998/11/4 | 6957.27 | 7061.11 | 103.84 | 6957.27 + 0.07*103.84 | 6893.692 |
| 1998/11/5 | 6889.65 | 6957.27 | 67.62 | 6889.65 + 0.07*67.62 | 6950.001 |
| 1998/11/6 | 6978.72 | 7061.11 | 82.39 | 6978.72 + 0.07*82.39 | 6884.917 |
| 1998/11/7 | 6957.40 | 6978.72 | 21.32 | 6957.40 + 0.07*21.32 | 6972.953 |
| 1998/11/9 | 6812.30 | 6957.4 | 145.1 | 6812.30 + 0.07*145.1 | 6955.908 |
| 1998/11/10 | 6654.79 | 6812.3 | 157.51 | 6654.79 + 0.07*157.51 | 6802.143 |
| 1998/11/11 | 6829.62 | 6654.79 | −174.83 | 6829.62 + 0.07*−174.83 | 6643.764 |
| … | … | … | … | … | … |

**Table 7**
The results of proposed model for different intervals and whether rules filtered.

| | # interval | Rule filter | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TAIEX | 7 | NO | 141.5 | 164.2 | 175.0 | 154.5 | 168.1 | 95.9 | 129.4 | 65.7 | 84.2 |
| | 7 | YES | 146.8 | 170.7 | 169.8 | 142.4 | 84.5[a] | 95.9 | 68.6 | 56.7[a] | 63.9 |
| | 9 | NO | 137.0 | 180.0 | 373.2 | 136.0 | 103.8 | 60.1 | 104.3 | 60.2 | 101.8 |
| | 9 | YES | **120.8** | **110.7** | **150.6** | **113.2** | **66.0** | **53.1** | **58.6** | **53.5** | **53.1** |
| HIS | 7 | NO | 318.2 | 312.7 | 306.2 | 257.3 | 154.6 | 148.1 | 134.5 | 128.0 | 192.4 |
| | 7 | YES | 224.7 | 252.2 | 261.3 | 231.1 | 129.7 | 144.5[a] | 186.6 | 102.5[a] | 192.4 |
| | 9 | NO | 305.3 | 278.7 | 295.5 | 277.4 | 120.3 | 144.8 | 111.3 | 118.9 | 192.4 |
| | 9 | YES | **202.0** | **231.9** | **251.7** | **156.6** | **106.3** | **118.7** | **105.4** | **104.0** | **189.2** |

*Note:* The bold numeric denotes minimal RMSE in different linguistic values, and # interval represents the number of linguistic interval.

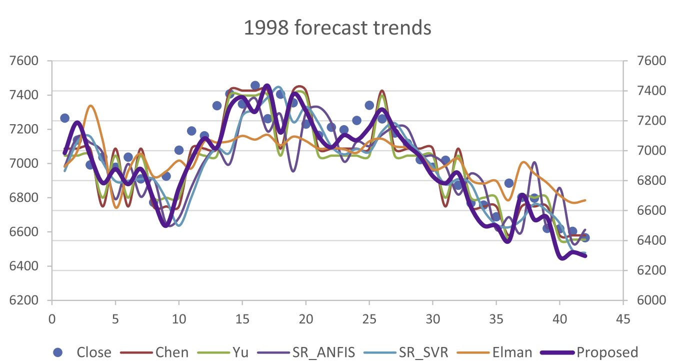


**Fig. 2.** The forecast trend of actual and predicted values for 1998 testing data of TAIEX.

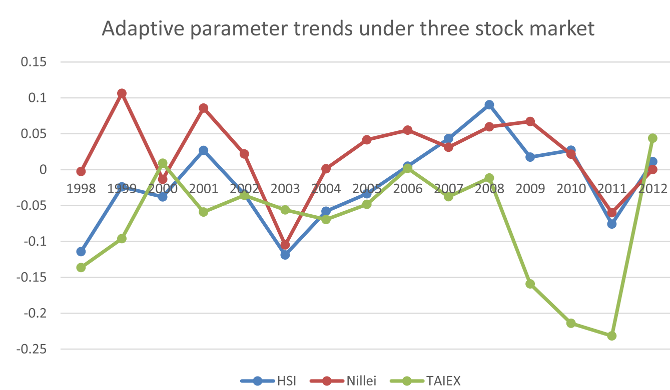


**Fig. 4.** the trends of adaptive parameter $h_0$ for TAIEX, HIS, and Nikkei stock market.

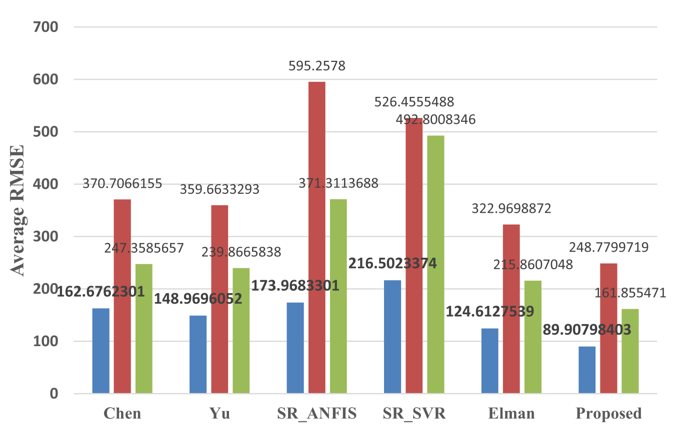


**Fig. 3.** The results for average RMSE of different model for TAIEX, HIS, and Nikkei.

Therefore, the proposed model and feasible forecast model in the collected stock datasets.

In adaptive expectation model [15], this study finds the optimized adaptive parameter $h_0$ and calculates the RMSE of initial and final forecast for TAIEX, HIS, and Nikkei as Table 11. To find which stock market is more fluctuation, this paper also plots three collected stock datasets with the optimized adaptive parameter $h_0$ as Fig. 4. Based on the meaning of adaptive parameter $h_0$ and the results as Table 11 and Fig. 4, we can find that:

(1) adaptive parameter $h_0$ almost is less than zero for TAIEX (only two times $h_0 > 0$), it represents TAIEX almost the downward spiral trend in 1998–2012 periods; and most of the adaptive adjusted forecasts are better than initial forecast in RMSE;
(2) the Nikkei market is frequently upward spiral trend in 1998–2012 year, due to most of its adaptive parameter $h_0$ is greater than zero ($h_0 > 0$ *has* 11 times);
(3) The HIS is fluctuated market because the adaptive parameter $h_0$ oscillates in upward and downward trends ($h_0 > 0$ 7 times and 8 times $h_0 < 0$).
(4) From TAIEX, HIS, and Nikkei stock market, the adaptive expectation model could improve forecast performance, the adjusted forecast is better than initial forecast in RMSE.

**Table 8**
Performance comparison for TAIEX, HSI, and Nikkei under RMSE.

| Year | Chen | Yu | SR + ANFIS | SR + SVR | Elman | Proposed |
|---|---|---|---|---|---|---|
| TAIEX | | | | | | |
| 1998 | 164.74 | 163.95 | 185.29 | 158.08 | 209.38 | **120.79** |
| 1999 | 148.04 | 131.66 | 182.90 | 173.35 | 170.90 | **110.69** |
| 2000 | 413.27 | 419.64 | 454.63 | 255.87 | 154.21 | **150.55** |
| 2001 | 132.74 | 132.40 | 167.55 | 174.74 | **103.14** | 113.17 |
| 2002 | 110.25 | 108.83 | 88.87 | 89.66 | 82.26 | **65.97** |
| 2003 | 81.59 | 60.52 | 76.43 | 86.75 | 117.33 | **53.09** |
| 2004 | 112.29 | 71.63 | 85.30 | 83.39 | 68.82 | **58.60** |
| 2005 | 72.95 | 65.42 | 87.50 | 164.65 | **50.14** | 53.49 |
| 2006 | 112.68 | 95.73 | 187.95 | 743.44 | 94.24 | **53.11** |
| 2007 | 169.31 | 214.86 | 374.11 | 226.11 | 193.97 | **151.94** |
| 2008 | 269.88 | 225.01 | 210.32 | 205.53 | 121.68 | **105.68** |
| 2009 | 295.36 | 177.64 | 102.62 | 109.56 | 131.01 | **78.96** |
| 2010 | 123.58 | 125.36 | 109.59 | 232.45 | **48.31** | 52.72 |
| 2011 | 164.09 | 178.05 | 207.39 | 401.34 | **104.18** | 119.43 |
| 2012 | 69.35 | 63.84 | 89.08 | 142.61 | 219.63 | **60.44** |
| HSI | | | | | | |
| 1998 | 276.76 | 291.40 | 326.62 | 296.67 | 326.24 | **201.99** |
| 1999 | 427.99 | 469.61 | 637.05 | 761.86 | 269.50 | **231.91** |
| 2000 | 280.15 | 297.05 | 356.70 | 356.81 | 302.27 | **251.70** |
| 2001 | 366.90 | 316.85 | 299.43 | 254.07 | 319.89 | **156.58** |
| 2002 | 124.98 | 123.70 | 155.09 | 155.40 | **102.83** | 106.26 |
| 2003 | 248.10 | 186.16 | 226.38 | 199.58 | 128.51 | **118.74** |
| 2004 | 223.46 | 264.34 | 239.63 | 540.19 | 114.65 | **105.38** |
| 2005 | 107.68 | 112.37 | 147.24 | 1127.21 | **103.12** | 103.96 |
| 2006 | 263.14 | 252.44 | 466.24 | 407.89 | **177.32** | 189.20 |
| 2007 | 787.24 | 912.67 | 1847.75 | 1028.66 | 916.46 | **682.08** |
| 2008 | 662.71 | 684.88 | 2178.70 | 593.84 | 613.21 | **460.12** |
| 2009 | 666.98 | 442.64 | 437.24 | 435.18 | 366.60 | **326.65** |
| 2010 | 436.63 | 382.06 | 445.41 | 718.33 | 368.92 | **260.67** |
| 2011 | 453.14 | 419.67 | 688.04 | 578.70 | 455.17 | **346.33** |
| 2012 | 234.74 | 239.11 | 477.34 | 442.44 | 279.86 | **190.13** |
| Nikkei | | | | | | |
| 1998 | 348.64 | 295.54 | 369.30 | 334.02 | 278.40 | **216.26** |
| 1999 | 309.40 | 466.48 | 436.65 | 283.40 | 166.48 | **154.21** |
| 2000 | 407.80 | 265.05 | 404.63 | 355.73 | 263.75 | **231.45** |
| 2001 | 245.44 | 191.74 | 239.94 | 246.27 | 205.87 | **185.58** |
| 2002 | 216.89 | 272.93 | 183.55 | 182.56 | 150.62 | **132.33** |
| 2003 | 196.79 | 216.64 | 261.32 | 275.87 | 222.21 | **176.68** |
| 2004 | 146.02 | 115.94 | 142.65 | 171.31 | 123.00 | **104.84** |
| 2005 | 243.87 | 220.22 | 1511.03 | 3787.93 | 235.94 | **176.71** |
| 2006 | 186.15 | 185.12 | 186.94 | 182.68 | 412.92 | **125.39** |
| 2007 | 401.10 | 383.63 | 573.57 | 536.08 | **196.71** | 216.95 |
| 2008 | 323.20 | 330.52 | 556.13 | 337.12 | 313.38 | **273.73** |
| 2009 | 193.57 | 170.13 | 190.94 | 193.66 | 152.25 | **130.14** |
| 2010 | 124.26 | 131.97 | 154.10 | 131.80 | 246.20 | **99.52** |
| 2011 | 242.89 | 213.37 | 202.26 | 185.77 | 166.68 | **105.30** |
| 2012 | 124.35 | 138.72 | 156.65 | 187.82 | 103.49 | **98.74** |

*Notes:* The bold digital denotes the best performance among 6 models.

**Table 9**
Performance comparison for TAIEX, HSI, and Nikkei under RRSE.

| Year | Chen | Yu | SR_ANFIS | SR_SVR | Elman | Proposed |
|---|---|---|---|---|---|---|
| TAIEX | | | | | | |
| 1998 | 0.5897 | 0.5869 | 0.6633 | 0.5659 | 0.7495 | **0.4324** |
| 1999 | 0.6187 | 0.5502 | 0.7644 | 0.7245 | 0.7142 | **0.4626** |
| 2000 | 1.1083 | 1.1254 | 1.2193 | 0.6862 | 0.4136 | **0.4037** |
| 2001 | 0.2550 | 0.2543 | 0.3218 | 0.3356 | **0.1981** | 0.2174 |
| 2002 | 1.1140 | 1.0997 | 0.8979 | 0.9059 | 0.8312 | **0.6665** |
| 2003 | 0.8177 | 0.6065 | 0.7659 | 0.8693 | 1.1758 | **0.5321** |
| 2004 | 1.1903 | 0.7592 | 0.9042 | 0.8839 | 0.7295 | **0.6212** |
| 2005 | 0.3424 | 0.3070 | 0.4107 | 0.7727 | **0.2353** | 0.2510 |
| 2006 | 0.5164 | 0.4387 | 0.8613 | 3.4071 | 0.4319 | **0.2434** |
| 2007 | 0.4087 | 0.5187 | 0.9031 | 0.5458 | 0.4682 | **0.3668** |
| 2008 | 1.2750 | 1.0630 | 0.9936 | 0.9710 | 0.5749 | **0.4993** |
| 2009 | 1.6022 | 0.9636 | 0.5567 | 0.5943 | 0.7107 | **0.4283** |
| 2010 | 0.5351 | 0.5428 | 0.4745 | 1.0065 | **0.2092** | 0.2283 |
| 2011 | 0.5867 | 0.6366 | 0.7415 | 1.4350 | **0.3725** | 0.4270 |
| 2012 | 0.3202 | 0.2947 | 0.4112 | 0.6583 | 1.0139 | **0.2790** |
| HSI | | | | | | |
| 1998 | 1.0921 | 1.1498 | 1.2888 | 1.1706 | 1.2873 | **0.7970** |
| 1999 | 0.3957 | 0.4342 | 0.5890 | 0.7043 | 0.2492 | **0.2144** |
| 2000 | 0.6543 | 0.6938 | 0.8331 | 0.8334 | 0.7060 | **0.5879** |
| 2001 | 0.7993 | 0.6903 | 0.6523 | 0.5535 | 0.6969 | **0.3411** |
| 2002 | 0.5362 | 0.5307 | 0.6654 | 0.6668 | **0.4412** | 0.4559 |
| 2003 | 1.1468 | 0.8605 | 1.0464 | 0.9226 | 0.5940 | **0.5488** |
| 2004 | 0.7706 | 0.9116 | 0.8263 | 1.8628 | 0.3954 | **0.3634** |
| 2005 | 0.4637 | 0.4839 | 0.6341 | 4.8546 | **0.4441** | 0.4477 |
| 2006 | 0.8327 | 0.7989 | 1.4754 | 1.2908 | **0.5611** | 0.5987 |
| 2007 | 0.7060 | 0.8185 | 1.6571 | 0.9225 | 0.8219 | **0.6117** |
| 2008 | 0.7744 | 0.8003 | 2.5460 | 0.6940 | 0.7166 | **0.5377** |
| 2009 | 1.2234 | 0.8119 | 0.8020 | 0.7982 | 0.6724 | **0.5992** |
| 2010 | 0.6940 | 0.6073 | 0.7080 | 1.1418 | 0.5864 | **0.4143** |
| 2011 | 0.7669 | 0.7103 | 1.1644 | 0.9794 | 0.7703 | **0.5861** |
| 2012 | 0.4888 | 0.4979 | 0.9939 | 0.9212 | 0.5827 | **0.3959** |
| Nikkei | | | | | | |
| 1998 | 0.8255 | 0.6998 | 0.8744 | 0.7909 | 0.6592 | **0.5121** |
| 1999 | 1.1748 | 1.7713 | 1.6580 | 1.0761 | 0.6322 | **0.5855** |
| 2000 | 0.8587 | 0.5581 | 0.8521 | 0.7491 | 0.5554 | **0.4874** |
| 2001 | 1.0696 | 0.8356 | 1.0457 | 1.0732 | 0.8972 | **0.8088** |
| 2002 | 0.8419 | 1.0594 | 0.7125 | 0.7086 | 0.5847 | **0.5137** |
| 2003 | 0.7184 | 0.7909 | 0.9540 | 1.0071 | 0.8112 | **0.6450** |
| 2004 | 0.7781 | 0.6178 | 0.7601 | 0.9128 | 0.6554 | **0.5587** |
| 2005 | 0.3311 | 0.2990 | 2.0517 | 5.1432 | 0.3204 | **0.2399** |
| 2006 | 0.4168 | 0.4145 | 0.4186 | 0.4090 | 0.9246 | **0.2808** |
| 2007 | 0.8964 | 0.8573 | 1.2818 | 1.1980 | **0.4396** | 0.4848 |
| 2008 | 0.8830 | 0.9030 | 1.5194 | 0.9210 | 0.8562 | **0.7478** |
| 2009 | 0.5264 | 0.4627 | 0.5193 | 0.5267 | 0.4141 | **0.3539** |
| 2010 | 0.4032 | 0.4282 | 0.5000 | 0.4276 | 0.7988 | **0.3229** |
| 2011 | 1.5150 | 1.3309 | 1.2615 | 1.1587 | 1.0396 | **0.6568** |
| 2012 | 0.2639 | 0.2944 | 0.3324 | 0.3986 | 0.2196 | **0.2095** |

*Notes:* The bold digital denotes the best performance among 6 models.

### 4.3. Profit evaluation and comparison

In order to present that the proposed model not only has good forecast ability, but also has higher profit, this section utilizes the proposed profit Eq. (7) to compare the proposed model with listing models. The profit results are shown in Table 12 and Fig. 5, the results show that the total profit of proposed model outperforms the listing models in 1998–2012, except HIS in Elman RNN model. Therefore, the proposed model is a favorable investment tool in stock market.

Furthermore, this paper sets three scenarios based on the profits of different $\alpha$ (relative absolute error) to provide investment suggestion to investor as references. The three different scenario are buy and sell rule to calculate their profit by proposed model for the best $\alpha$, $\alpha = 0.01$, and $\alpha = 0.1$, respectively. The profit and total profit of different $\alpha$ is showed in Table 13 and Figs. 6–8. From the three scenarios based on the profits of different $\alpha$, we can find that the best total profit comes from the best $\alpha$, and the total profit

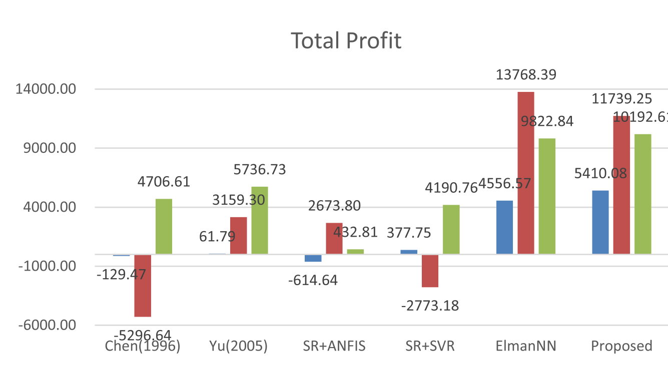


**Fig. 5.** The total profit comparison of different model for TAIEX, HIS, and Nikkei.

**Table 10**
Performance comparison for TAIEX, HSI, and Nikkei under RAE.

| Year | Chen | Yu | SR_ANFIS | SR_SVR | Elman | Proposed |
|---|---|---|---|---|---|---|
| TAIEX (RAE) | | | | | | |
| 1998 | 0.5960 | 0.6046 | 0.6483 | 0.5486 | 0.7180 | **0.4373** |
| 1999 | 0.5967 | 0.5719 | 0.8357 | 0.8141 | 0.7459 | **0.4710** |
| 2000 | 1.1945 | 1.2201 | 1.3524 | 0.7095 | 0.4343 | **0.3798** |
| 2001 | 0.2168 | 0.2079 | 0.2716 | 0.2773 | **0.1800** | 0.1933 |
| 2002 | 1.0603 | 1.0273 | 0.8369 | 0.8741 | 0.7829 | **0.6169** |
| 2003 | 0.8405 | 0.5762 | 0.7535 | 0.8583 | 1.2290 | **0.5029** |
| 2004 | 1.2265 | 0.8274 | 0.9042 | 0.8605 | 0.6546 | **0.5642** |
| 2005 | 0.3246 | 0.2956 | 0.4075 | 0.7856 | **0.2379** | 0.2381 |
| 2006 | 0.4858 | 0.4173 | 0.7936 | 3.1438 | 0.4089 | **0.2258** |
| 2007 | 0.4296 | 0.5370 | 0.8641 | 0.6029 | 0.4617 | **0.3694** |
| 2008 | 1.3212 | 1.0959 | 0.9536 | 0.9451 | 0.5694 | **0.4395** |
| 2009 | 1.9280 | 1.1874 | 0.6432 | 0.6907 | 0.7886 | **0.4329** |
| 2010 | 0.4530 | 0.4644 | 0.3867 | 0.8566 | 0.1759 | **0.2012** |
| 2011 | 0.5427 | 0.6027 | 0.7454 | 1.5260 | 0.3447 | **0.4084** |
| 2012 | 0.2644 | 0.2395 | 0.3304 | 0.6238 | 1.0011 | **0.2209** |
| HIS(RAE) | | | | | | |
| 1998 | 1.1604 | 1.2510 | 1.3791 | 1.1995 | 1.1052 | **0.8443** |
| 1999 | 0.4090 | 0.4322 | 0.6033 | 0.7306 | 0.2367 | **0.2066** |
| 2000 | 0.6562 | 0.7044 | 0.7999 | 0.7798 | 0.6827 | **0.5791** |
| 2001 | 0.8792 | 0.7519 | 0.6900 | 0.5902 | 0.6438 | **0.3347** |
| 2002 | 0.5299 | 0.5101 | 0.6855 | 0.6551 | 0.4304 | **0.4276** |
| 2003 | 1.1694 | 0.8515 | 1.0325 | 0.8543 | 0.5359 | **0.4779** |
| 2004 | 0.8241 | 1.0026 | 0.8922 | 2.0145 | 0.4079 | **0.3572** |
| 2005 | 0.4353 | 0.4535 | 0.6061 | 5.6266 | 0.4259 | **0.4037** |
| 2006 | 0.8885 | 0.8617 | 1.5342 | 1.4582 | **0.5866** | 0.6265 |
| 2007 | 0.6923 | 0.7947 | 1.4967 | 0.9249 | 0.7869 | **0.6052** |
| 2008 | 0.7549 | 0.7740 | 2.6897 | 0.7223 | 0.6871 | **0.5236** |
| 2009 | 1.1993 | 0.7788 | 0.7768 | 0.7227 | 0.5815 | **0.5451** |
| 2010 | 0.7109 | 0.6030 | 0.7165 | 1.0785 | 0.5022 | **0.4175** |
| 2011 | 0.6931 | 0.6422 | 0.8615 | 0.9663 | 0.6362 | **0.5113** |
| 2012 | 0.4784 | 0.4871 | 1.0135 | 0.9520 | 0.4581 | **0.3446** |
| Nikkei (RAE) | | | | | | |
| 1998 | 0.5960 | 0.6046 | 0.6483 | 0.5486 | 0.7180 | **0.4373** |
| 1999 | 0.5967 | 0.5719 | 0.8357 | 0.8141 | 0.7459 | **0.4710** |
| 2000 | 1.1945 | 1.2201 | 1.3524 | 0.7095 | 0.4343 | **0.3798** |
| 2001 | 0.2168 | 0.2079 | 0.2716 | 0.2773 | **0.1800** | 0.1933 |
| 2002 | 1.0603 | 1.0273 | 0.8369 | 0.8741 | 0.7829 | **0.6169** |
| 2003 | 0.8405 | 0.5762 | 0.7535 | 0.8583 | 1.2290 | **0.5029** |
| 2004 | 1.2265 | 0.8274 | 0.9042 | 0.8605 | 0.6546 | **0.5642** |
| 2005 | 0.3246 | 0.2956 | 0.4075 | 0.7856 | **0.2379** | 0.2381 |
| 2006 | 0.4858 | 0.4173 | 0.7936 | 3.1438 | 0.4089 | **0.2258** |
| 2007 | 0.4296 | 0.5370 | 0.8641 | 0.6029 | 0.4617 | **0.3694** |
| 2008 | 1.3212 | 1.0959 | 0.9536 | 0.9451 | 0.5694 | **0.4395** |
| 2009 | 1.9280 | 1.1874 | 0.6432 | 0.6907 | 0.7886 | **0.4329** |
| 2010 | 0.4530 | 0.4644 | 0.3867 | 0.8566 | **0.1759** | 0.2012 |
| 2011 | 0.5427 | 0.6027 | 0.7454 | 1.5260 | **0.3447** | 0.4084 |
| 2012 | 0.2644 | 0.2395 | 0.3304 | 0.6238 | 1.0011 | **0.2209** |

*Notes:* The bold digital denotes the best performance among 6 models.

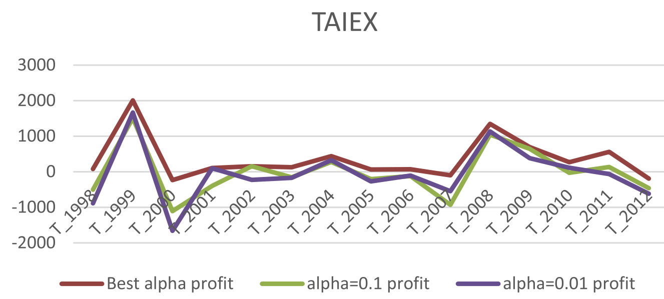


**Fig. 6.** The profit of different scenarios $\alpha$ for TAIEX.

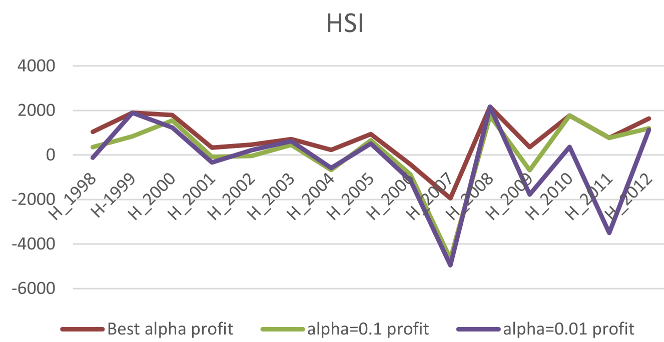


**Fig. 7.** The profit of different scenarios $\alpha$ for HIS.

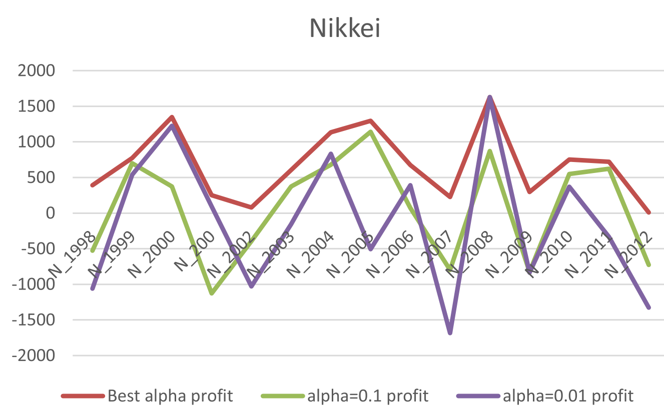


**Fig. 8.** The profit of different scenarios $\alpha$ for Nikkei.

**Table 11**
The results of adaptive expectation model for three Stock market (testing RMSE).

| Year | TAIEX | | | HSI | | | Nikkei | | |
|---|---|---|---|---|---|---|---|---|---|
| | *h* | Initial forecast | Final forecast | *h* | Initial forecast | Final forecast | *h* | Initial forecast | Final forecast |
| 1998 | −0.1363 | 128.80 | **120.79** | −0.1142 | 224.40 | **201.99** | −0.0026 | 173.23 | 216.26 |
| 1999 | −0.0959 | 121.70 | **110.69** | −0.0239 | 242.60 | **231.91** | 0.1063 | 222.42 | **154.21** |
| 2000 | 0.0091 | 135.90 | 150.55 | −0.038 | 252.70 | **251.70** | −0.0134 | 246.52 | **231.45** |
| 2001 | −0.0589 | 122.00 | **113.17** | 0.0268 | 168.60 | **156.58** | 0.0859 | 172.31 | 185.58 |
| 2002 | −0.0358 | 84.50 | **65.97** | −0.0341 | 112.60 | **106.26** | 0.0218 | 131.04 | 132.33 |
| 2003 | −0.056 | 59.10 | **53.09** | −0.1188 | 144.50 | **118.74** | −0.1049 | 151.08 | 176.68 |
| 2004 | −0.0693 | 63.30 | **58.60** | −0.0578 | 103.40 | 105.38 | 0.0012 | 105.60 | **104.84** |
| 2005 | −0.0483 | 56.70 | **53.49** | −0.0336 | 102.50 | 103.96 | 0.0416 | 236.03 | **176.71** |
| 2006 | 0.0019 | 63.20 | **53.11** | 0.005 | 194.20 | **189.20** | 0.0552 | 142.53 | **125.39** |
| 2007 | −0.0376 | 155.28 | **151.94** | 0.0433 | 533.41 | 682.08 | 0.0313 | 146.94 | 216.95 |
| 2008 | −0.0116 | 179.43 | **105.68** | 0.0904 | 566.77 | **460.12** | 0.0596 | 274.06 | **273.73** |
| 2009 | −0.159 | 164.51 | **78.96** | 0.0176 | 468.37 | **326.65** | 0.0672 | 138.00 | **130.14** |
| 2010 | −0.2139 | 72.89 | **52.72** | 0.0272 | 261.00 | **260.67** | 0.0215 | 97.56 | 99.52 |
| 2011 | −0.2315 | 84.49 | 119.43 | −0.0757 | 434.60 | **346.33** | −0.0597 | 107.58 | **105.30** |
| 2012 | 0.0437 | 116.15 | **60.44** | 0.0112 | 182.55 | 190.13 | 0.0002 | 150.09 | **98.74** |

*Notes:* The bold digital denotes final forecast is better than initial forecast.

Please cite this article as: C.-H. Cheng, J.-H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.04.014

**Table 12**
The profits comparisons of different models for TAIEX, HSI, and Nikkei.

| Year | α | Chen [4] | Yu [36] | SR + ANFIS | SR + SVR | Elman | Proposed |
|---|---|---|---|---|---|---|---|
| TAIEX (Profit) | | | | | | | |
| 1998 | 0.001 | 0.00 | 108.83 | −157.51 | 0.00 | −405.25 | 78.38 |
| 1999 | 0.022 | 424.78 | 146.54 | −475.18 | −601.95 | 840.07 | 2007.08 |
| 2000 | 0.001 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | −231.02 |
| 2001 | 0.045 | −2.04 | 294.10 | 865.74 | 850.02 | 1815.53 | 103.11 |
| 2002 | 0.01 | −80.04 | −248.79 | 136.88 | −410.18 | 124.84 | 159.33 |
| 2003 | 0.001 | 0.00 | −69.50 | 51.20 | 142.11 | −20.41 | 131.18 |
| 2004 | 0.02 | −137.96 | 104.15 | 180.01 | −162.82 | 494.99 | 441.42 |
| 2005 | 0.004 | −428.82 | −155.11 | −248.30 | −57.87 | 150.25 | 62.18 |
| 2006 | 0.002 | 123.35 | 119.97 | 112.40 | 0.00 | 76.50 | 74.35 |
| 2007 | 0.021 | 327.52 | −618.40 | 52.68 | 367.61 | 774.95 | −101.96 |
| 2008 | 0.038 | −132.87 | 562.56 | −739.26 | 648.18 | 555.93 | 1351.17 |
| 2009 | 0.009 | −37.91 | 10.94 | −541.00 | −279.48 | −487.50 | 693.25 |
| 2010 | 0.005 | 28.61 | 28.61 | −85.10 | −115.43 | 463.08 | 272.55 |
| 2011 | 0.005 | −198.09 | −197.79 | 187.14 | −2.44 | 173.59 | 559.74 |
| 2012 | 0.001 | −16.00 | −24.33 | 45.66 | 0.00 | 0.00 | −190.68 |
| Total | | −129.47 | 61.79 | −614.64 | 377.75 | 4556.57 | **5410.08** |
| HSI(Profit) | | | | | | | |
| 1998 | 0.008 | −547.28 | 329.00 | −819.69 | −1276.67 | 901.82 | 1043.05 |
| 1999 | 0.032 | −1581.41 | −1436.90 | 400.34 | −1014.37 | 2259.24 | 1894.67 |
| 2000 | 0.012 | −1471.03 | −1368.05 | −602.94 | 190.71 | 2342.38 | 1793.12 |
| 2001 | 0.018 | −317.82 | −278.47 | −414.93 | 269.45 | 1194.91 | 331.26 |
| 2002 | 0.014 | 272.57 | 397.91 | −52.15 | 290.53 | 419.81 | 461.43 |
| 2003 | 0.021 | 328.05 | 320.78 | −85.06 | −1217.18 | 1349.43 | 714.65 |
| 2004 | 0.001 | −156.23 | 0.00 | −155.70 | 0.00 | −351.60 | 228.85 |
| 2005 | 0.005 | 83.27 | −120.19 | 396.15 | 0.00 | 688.33 | 937.34 |
| 2006 | 0.002 | −1593.43 | −1593.43 | −626.20 | −623.25 | −348.72 | −419.44 |
| 2007 | 0.001 | 0.00 | −716.45 | 0.00 | 0.00 | 776.61 | −1948.68 |
| 2008 | 0.082 | −1538.28 | 4072.52 | 416.77 | 359.71 | 4730.24 | 2165.43 |
| 2009 | 0.003 | 442.17 | 4.43 | 206.54 | 664.81 | −198.97 | 352.98 |
| 2010 | 0.01 | 115.89 | 2436.20 | 349.06 | −170.63 | −569.90 | 1771.95 |
| 2011 | 0.01 | 311.13 | 653.12 | 3887.40 | 685.31 | 149.45 | 769.96 |
| 2012 | 0.013 | 355.77 | 458.82 | −225.80 | −931.60 | 425.36 | 1642.67 |
| Total | | −5296.64 | 3159.30 | 2673.80 | −2773.18 | **13,768.39** | 11,739.25 |
| Nikkei (Profit) | | | | | | | |
| 1998 | 0.003 | −294.22 | 296.16 | 322.63 | −2.39 | 340.48 | 392.41 |
| 1999 | 0.009 | 117.37 | −353.06 | −422.31 | −82.04 | 570.70 | 774.84 |
| 2000 | 0.026 | 395.37 | 597.93 | −58.77 | 1590.59 | 1643.55 | 1348.62 |
| 2001 | 0.034 | −60.20 | 151.51 | 1562.57 | 351.76 | 2141.96 | 251.18 |
| 2002 | 0.002 | 162.83 | 0.00 | 51.89 | 284.32 | 300.90 | 78.35 |
| 2003 | 0.025 | 957.80 | 421.88 | −636.03 | 102.52 | 1406.32 | 604.21 |
| 2004 | 0.014 | −57.74 | 850.30 | 8.46 | 422.40 | 755.01 | 1133.37 |
| 2005 | 0.013 | 2002.25 | 2002.25 | 0.00 | 208.10 | 1452.12 | 1295.89 |
| 2006 | 0.004 | −272.64 | 65.73 | 177.60 | 95.06 | −11.96 | 676.32 |
| 2007 | 0.001 | −226.49 | −226.49 | 120.33 | 0.00 | 563.68 | 225.40 |
| 2008 | 0.075 | 2918.96 | 2513.47 | −875.00 | −69.29 | 1125.97 | 1629.46 |
| 2009 | 0.003 | 272.34 | 194.98 | 69.01 | −261.75 | −27.25 | 297.89 |
| 2010 | 0.019 | 178.29 | 35.56 | 381.33 | 1728.53 | 271.87 | 753.32 |
| 2011 | 0.008 | −616.09 | −364.88 | −282.88 | −72.95 | −345.62 | 722.16 |
| 2012 | 0.004 | −771.23 | −448.61 | 13.98 | −104.10 | −364.88 | 9.19 |
| Total | | 4706.61 | 5736.73 | 432.81 | 4190.76 | 9822.84 | **10,192.61** |

*Notes:* The bold digital denotes the best total profits for each stock market under different model.

**Table 13**
The profit of different scenarios α for TAIEX, HIS, and Nikkei.

| Year | TAIEX | | | | HSI | | | | Nikkei | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Best α | Best Profit | α = 0.1 profit | α = 0.01 profit | Best α | Best Profit | α = 0.1 profit | α = 0.01 profit | Best α | Best Profit | α = 0.1 profit | α = 0.01 profit |
| 1998 | 0.001 | 78.38 | −502.06 | −884.74 | 0.008 | 1043.05 | 352.32 | −117.30 | 0.003 | 392.41 | −527.10 | −1061.72 |
| 1999 | 0.022 | 2007.08 | 1507.90 | 1669.47 | 0.032 | 1894.67 | 837.54 | 1894.67 | 0.009 | 774.84 | 698.59 | 539.20 |
| 2000 | 0.001 | −231.02 | −1107.87 | −1657.35 | 0.012 | 1793.12 | 1541.97 | 1231.78 | 0.026 | 1348.62 | 373.95 | 1225.76 |
| 2001 | 0.045 | 103.11 | −399.95 | 96.41 | 0.018 | 331.26 | −87.73 | −332.34 | 0.034 | 251.18 | −1128.01 | 91.98 |
| 2002 | 0.01 | 159.33 | 159.33 | −224.70 | 0.014 | 461.43 | −44.42 | 212.71 | 0.002 | 78.35 | −394.77 | −1030.33 |
| 2003 | 0.001 | 131.18 | −153.76 | −171.14 | 0.021 | 714.65 | 455.22 | 625.53 | 0.025 | 604.21 | 375.11 | −157.07 |
| 2004 | 0.02 | 441.42 | 277.80 | 325.30 | 0.001 | 228.85 | −674.48 | −582.41 | 0.014 | 1133.37 | 677.76 | 834.43 |
| 2005 | 0.004 | 62.18 | −202.24 | −269.49 | 0.005 | 937.34 | 656.75 | 512.97 | 0.013 | 1295.89 | 1143.26 | −506.35 |
| 2006 | 0.002 | 74.35 | −121.15 | −107.13 | 0.002 | −419.44 | −881.29 | −1146.56 | 0.004 | 676.32 | 70.86 | 390.17 |
| 2007 | 0.021 | −101.96 | −930.48 | −549.43 | 0.001 | −1948.68 | −4662.89 | −4954.27 | 0.001 | 225.40 | −796.26 | −1684.50 |
| 2008 | 0.038 | 1351.17 | 1042.35 | 1132.24 | 0.082 | 2165.43 | 1764.83 | 2165.43 | 0.075 | 1629.46 | 871.47 | 1629.46 |

(*continued on next page*)

**Table 13** (*continued*)

| Year | TAIEX | | | | HSI | | | | Nikkei | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Best $\alpha$ | Best Profit | $\alpha=0.1$ profit | $\alpha=0.01$ profit | Best $\alpha$ | Best Profit | $\alpha=0.1$ profit | $\alpha=0.01$ profit | Best $\alpha$ | Best Profit | $\alpha=0.1$ profit | $\alpha=0.01$ profit |
| 2009 | 0.009 | 693.25 | 643.48 | 385.23 | 0.003 | 352.98 | −689.63 | −1781.41 | 0.003 | 297.89 | −823.71 | −837.21 |
| 2010 | 0.005 | 272.55 | −27.59 | 113.47 | 0.01 | 1771.95 | 1771.95 | 365.78 | 0.019 | 753.32 | 547.46 | 370.42 |
| 2011 | 0.005 | 559.74 | 132.18 | −62.03 | 0.01 | 769.96 | 769.96 | −3501.22 | 0.008 | 722.16 | 621.51 | −336.09 |
| 2012 | 0.001 | −190.68 | −457.75 | −615.54 | 0.013 | 1642.67 | 1194.93 | 1127.95 | 0.004 | 9.19 | −728.39 | −1326.93 |
| Total profit | | **5410.08** | −139.81 | −819.43 | | **11,739.25** | 2305.04 | −4278.69 | | 10,192.61 | 981.74 | −1858.78 |

*Notes:* The bold digital denotes the best profits for each stock market under three different $\alpha$.

of $\alpha=0.1$ is better than $\alpha=0.01$, it indicates that forecast accuracy can not represent the higher profit. Since setting the threshold makes trading a proactive investment, the investor must utilize all information to optimize the trading approach.

## 5. Conclusion

This paper has proposed a novel fuzzy time-series model based on rough set rule induction for forecasting stock index, the proposed model utilized rough set LEM2 algorithm to generate forecast rules, it's different from previous fuzzy time series using FLRs and employed adaptive expectation model to strengthen forecasting performance. The results have shown the proposed model with better forecast performance in accuracy and profit. From three stock markets data with 45 sub-datasets, we can conclude that there are four findings as follows.

(1) The number of linguistic interval
Table 7 show that the nine linguistic intervals will result in the higher forecasting performance. From Yu [36], the more linguistic intervals utilize the more get forecast performance. However, based on Miller [10], the appropriate number of category for human shorten memory function is seven, or seven plus or minus two. That is, many linguistic intervals would disturb human shorten memory function for investors. Therefore, this study suggests that nine linguistic intervals with rule filter is a good approach.

(2) Whether rule filtered
After filtering the generated rules (remove the "support" less than 2), the higher support rules can get the better forecasting performance as Table 7. i.e., the higher support rules in the training data have higher data matching, and then the forecast performance of proposed model will be better for testing data.

(3) Adaptive expectation model
This study finds the optimal parameter ($h_0$) of adaptive expectation model in each training dataset, then the proposed model can effectively reduce the RMSE in all testing datasets. And in different model comparisons, Tables 8–10 show that the proposed model is better than the listing models in three evaluation indexes. From Fig. 4, we can find that: (1) adaptive parameter $h_0$ usually is less than zero for TAIEX, it represents TAIEX almost the downward spiral trend in 1998–2012 periods; (2) the Nikkei market is frequently upward spiral trend in 1998–2012 year, due to its adaptive parameter $h_0$ is most of greater than zero; and (3) The HIS is fluctuated market because the adaptive parameter $h_0$ oscillates in upward and downward trends.

(4) In profit
The profit comparison shows that sometimes the better forecast accuracy could not get the better profit. Because the timing of sell/buy depends on the optimal threshold parameter $\alpha$ in training dataset, when the profit is negative in training dataset, this study will not sell and buy transaction in testing data, and the profit is zero. From the results, it presents that the higher forecast accuracy can't get the higher profit, because the best threshold to obtain the best profits is proactive investment approach, this study suggest that the investors must utilize all available information to optimize their benefits.

(5) In forecast
This paper employed RMSE, RRSE, and RAE to evaluate and compare the proposed model with the listing model, the results indicates that the proposed model outperforms the listing models under the RMSE, RRSE, and RAE. And the proposed model is ranked first name in three error indexes, however the Elman RNN has competitive in profit of HSI. Therefore we can find that the fluctuated HIS market is more fit in Elman RNN model.

From above these findings, two conclusions are listed as follows: (1) The proposed model can effectively improve accuracy of forecasting, and the proposed model outperforms the listing models for TAIEX, Nikkei and HSI dataset. (2) The proposed model not only can get better accuracy but also has higher profit, and the results and findings could provide the investors to utilize all available information for optimizing their benefits.

In future, there are still several issues can be extended this study as follows: (1) In more experiments, other stock databases such as NASDAQ or Dow Jones would be implemented as experiment data. (2) In accuracy, we can utilize other factors; let the universe of discourse be partitioned into more linguistic values; and consider other AI techniques to optimize the proposed model such as Fourier transform, particle swarm optimization, genetic algorithm, and so on. (3) In applications, Apply to different application field such as electric loads forecasting, oil price forecasting, tourism demand forecasting, and so on.

## References

[1] Ç.H. Aladağ, U. Yolcu, E. Egrioglu, E. Bas, Fuzzy Lagged Variable Selection In Fuzzy Time Series With Genetic Algorithms, Appl. Soft Comput. 22 (2014) 465–473.
[2] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity, J. Econom. 31 (3) (1986) 307–327.
[3] G.E.P. Box, G.M. Jenkins, Time Series analysis: Forecasting and Control, Holden-Day, San Francisco, CA, USA, 1976.
[4] S.M. Chen, Forecasting enrollments based on fuzzy time-series, Fuzzy Sets Syst. 81 (1996) 311–319.
[5] S.M. Chen, P.Y. Kao, TAIEX forecasting based on fuzzy time-series, particle swarm optimization techniques and support vector machines, Inf. Sci. 247 (2013) 62–71.
[6] S.M. Chen, J.R. Hwang, Temperature prediction using fuzzy time series, IEEE Trans. Syst. Man. Cybern. Part B 30 (2) (2000) 263–275.

Please cite this article as: C.-H. Cheng, J.-H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.04.014

[7] X. Chen, J. She, M. Wu, A hybrid time series prediction model based on recurrent neural network and double joint linear–nonlinear extreme learning network for prediction of carbon efficiency in iron ore sintering process, Neurocomputing 249 (2017) 128–139.
[8] C.H. Cheng, T.L. Chen, C.H. Chiang, Fuzzy time-series based on adaptive expectation model for TAIEX forecasting, Expert Syst. Appl. 34 (2008) 1126–1134.
[9] J.L. Elman, Finding structure in time, Cogn. Sci. 14 (1990) 179–211.
[10] G.A. Miller, The magical number seven, plus or minus two: some limits on our capacity for processing information, Psychol. Rev. 63 (1956) 81–97.
[11] J.W. Grzymala-Busse, A new version of the rule induction system LERS, Fundam. Inform. 31 (1997) 27–39.
[12] K. Huarng, Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets Syst. 123 (2001) 387–394.
[13] T.A. Jilani, S.M.A. Burney, C. Ardil, Multivariate high order fuzzy time series forecasting for car road accidents, Int. J. Comput. Intell. 4 (1) (2007) 15–20.
[14] S.M. Kendall, K. Ord, An intelligence stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network, Fuzzy Sets Syst. 118 (1990) 21–45.
[15] J. Kmenta, Elements of Econometrics, Second ed., MacMillan, 1986.
[16] P. Kumar, P.R. Krishna, R.S. Bapi, S.K. De, Rough clustering of sequential data, Data Knowl. Eng. 63 (2007) 183–199.
[17] J.R. Kuo, C.H. Cheng, Y.C. Hwang, An intelligence stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network, Fuzzy Sets Syst. 118 (2001) 21–45.
[18] M.H. Lee, M.E. Nor, H.J. Sadaei Suhartono, N.H.A Rahman, et al., Fuzzy time series: An application to tourism demand forecasting, Am. J. Appl. Sci. 9 (2012) 132–140.
[19] L. Liu, A. Wiliem, S. Chen, B.C. Lovell, Automatic Image Attribute Selection for Zero-Shot Learning of Object Categories, in: Proceedings of the Twenty Second International Conference on Pattern Recognition, Stockholm, 2014, pp. 2619–2624.
[20] D. Parmar, T. Wu, J. Blackhurst, MMR: an algorithm for clustering categorical data using rough set theory, Data Knowl. Eng. 63 (2007) 879–893.
[21] Z. Pawlak, A. Skoworn, Rudiments of rough sets, Inf. Sci. 177 (2007) 3–27.
[22] Z. Pawlak, Rough set approach to knowledge-based decision support, Eur. J. Oper. Res. 99 (1997) 48–57.
[23] Y. Ren, P.N. Suganthan, N. Srikanth, A novel empirical mode decomposition with support vector regression for wind speed forecasting, IEEE Trans. Neural Netw. Learn. Syst. 27 (8) (2016) 1–6.
[24] A Rubio, JD Bermúdez, E Vercher, Improving stock index forecasts by using a new weighted fuzzy-trend time series method, Expert Syst. Appl. 76 (2017) 12–20.
[25] C. Sims, Macroeconomics and Reality, Econometrica 48 (1) (1980) 1–48.
[26] Y. Siskos, C. Zopounidis, A. Pouliezos, An integrated DSS for financing firms by an industrial development bank in Greece, Decis. Support Syst. 12 (1994) 151–168.
[27] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time-series–Part I, Fuzzy Sets Syst. 54 (1993) 1–10.
[28] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time-series–Part II, Fuzzy Sets Syst. 62 (1994) 1–8.
[29] J. Stefanowski, On rough set based approaches to induction of decision rules, Rough Sets Knowl. Discov. 1 (1998) 500–529.
[30] B. Sun, H. Guoa, H.R. Karimib, Y. Gea, S. Xiong, Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series, Neurocomputing 151 (2015) 1528–1536.
[31] FM Talarposhti, HJ Sadaei, R Enayatifar, FG Guimarães, M Mahmud, T Eslami, Stock market forecasting by using a hybrid model of exponential fuzzy time series, Int. J. Approx. Reas. 70 (2016) 79–98.
[32] A. Wakulicz-Deja, P. Paszek, International diagnose progressive encephalopathy applying the rough set approach, J.f Med. Inform. 46 (1997) 119–127.
[33] Y. Wan, Y.-W. Si, Adaptive neuro fuzzy inference system for chart pattern matching in financial time series, Appl. Soft Comput. 57 (2017) 1–18.
[34] L-Y Wei, C-H Cheng, S-H Wu, A hybrid ANFIS based on n-period moving average model to forecast TAIEX stock, Applied Soft Computing 19 (2014) 86–92.
[35] H.K. Yu, Weighted fuzzy time-series models for TAIEX forecasting, Physica A 349 (2001) 609–624.
[36] H.-K. Yu, A refined fuzzy time-series model for forecasting, Physica A 346 (3/4) (2005) 657–681.
[37] L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, IEEE Trans. Syst. Cybern. 3 (1973) 22–48.
[38] T. Zhang, L. Liu, A. Wiliem, B. Lovell, Is alice chasing or being chased? Determining subject and object of activities in videos, in: Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016, pp. 1–7.
[39] C. Zopounidis, M. Doumpos, N.F. Matsatsinis, Developing a multicriteria knowledge-based decision support system for the assessment of corporate performance and viability: The FINEVA system, Fuzzy Econ. Rev. 1 (2) (1996a) 35–53.
[40] C. Zopounidis, A. Pouliezos, D. Yannacopoulos, Designing a DSS for the assessment of company performance and viability, Comput. Sci. Econ. Manag. 5 (1992) 41–56.

**Ching-Hsue Cheng** Who received the Bachelor's degree in Mathematics from Chinese Military Academy in 1982, the Master's degree in Applied Mathematics from Chung-Yuan Christian University in 1988, and the Ph.D. degree in System Engineering and management from National Defense University in 1994. He is now professor of information management department in National Yunlin University of Science and Technology. His research is mainly in the field of fuzzy logic, fuzzy time series, soft computing, reliability, medical information management, and data mining. He has published more than 200 papers (include 195 significant journal papers).

**Yang Jun-He** Who was born in Kaohsiung, Taiwan. He got his master degree in Management Information System from National Pingtung University of Science and Technology, Taiwan. Currently he is a PhD student at the National Yunlin University of Science and Technology, Taiwan, and now his is working in College of Life and Creativity, Cheng Shiu University.