FUSION ANFIS MODEL BASED ON AR FOR FORECASTING EPS OF LEADING INDUSTRIES

LIANG-YING WEI¹, CHING-HSUE CHENG² AND HSIN-HUNG WU³

¹Department of Information Management
Yuanpei University
No. 306, Yuanpei Street, Hsin Chu 30015, Taiwan
lywei@mail.ypu.edu.tw

²National Yunlin University of Science and Technology
No. 123, Sec. 3, University Rd., Touliu, Yunlin 640, Taiwan
chcheng@yuntech.edu.tw

³Department of Business Administration
National Changhua University of Education
No. 2, Shida Rd., Changhua City 500, Taiwan
hhwu@cc.ncue.edu.tw

Received May 2010; revised October 2010

ABSTRACT. Earnings per share (EPS) represents the profitability of a common stock and the financial performance of a particular company. Therefore, EPS is often regarded as a major indicator for investors to purchase stocks. The traditional approach is to use a conventional linear time series model for EPS prediction. However, the results would be in doubt when the forecasting problems are nonlinear. For this reason, this paper proposes a fusion forecasting model that incorporates an autoregressive model into an adaptive network-based fuzzy inference system (ANFIS) with three facets: (1) test the lag period of EPS; (2) take fuzzy inference systems (FIS) to fuzzify the past periods of EPS based on the AR concept and use adaptive networks to tune optimal parameters; and (3) employ an integrated ANFIS model to predict EPS. To illustrate the proposed model, 15-quarter EPS data are employed. The experimental results indicate that the proposed model outperforms the listing models.

Keywords: Earning per share, Autoregressive model, Adaptive network-based fuzzy inference system

1. Introduction. In a mature industry, the leading company that takes the dominant position in the industry earns greater profits because this kind of companies is able to exploit its economic scale and market power better [1]. Taiwan started its IC design research in 1975 under governmental subsidiaries and holds a key position in the world [2]. The electronic industry is a crucial part of economic development in Taiwan, especially in the semiconductor industry. The share of a leading company, such as High Technology Corporation (HTC), often has a high stock market price and a decisive position in the industry, such that we regard this type of share as an industry-leading share.

In terms of investors, they always focus on earnings per share (EPS). EPS is the earnings return on original investment and represents the profitability of a common stock and the financial performance of this particular company. EPS is generally considered to be the most important variable to determine the market share price. In reality, finding out the profitability of a particular company is essential, such that the accuracy of EPS predictions would be critically important for investors.
Time series methods have been applied to forecast stock markets, and various models, such as ARCH (Autoregressive Conditional Heteroscedasticity), proposed by Engle [3] in 1982, have been used by many financial analysts to forecast the stock market. Later, Bollerslev [4] proposed the GARCH (Generalized ARCH) model to refine the ARCH model. Box and Jenkins [5] proposed the autoregressive moving average (ARMA) model, which combines a moving average process with a linear difference equation to obtain an autoregressive moving average model at a linear stationary condition. The results of this type of linear time series models would be in doubt when the forecasting problems are non-linear.

In the evolution of time series models, many researchers have applied data-mining techniques in financial analysis [6-9]. In 1990, Kinoto et al. [10] developed a prediction system for stock markets by neural networks. In addition, Nikolopoulos and Fellrath [11] combined genetic algorithms (GAs) and neural networks to develop a hybrid expert system for investment advices. However, the rules mined from time series data with the artificial intelligence data-mining techniques, such as GA and artificial neural network (ANN), are difficult to be understood.

Under such circumstances, this study proposes a hybrid adaptive network-based fuzzy inference system (ANFIS) model that can be applied to deal with the nonlinear data when the relationships between EPS and financial indicators are nonlinear. Further, the proposed integrated ANFIS models are rule-based models that can provide the financial-related rules to forecast EPS.

The rest of the paper is organized in the following: Section 2 describes related studies, including EPS, autoregressive models, adaptive network-based fuzzy inference system, fuzzy c-means clustering, and subtractive clustering; the proposed model is presented briefly in Section 3; the verification and comparison of different models are described in Section 4; Section 5 discusses the findings of this work; finally, the conclusions of the study are summarized in Section 6.

2. Related Works. In this section, the related works, including EPS, autoregressive model, adaptive network-based fuzzy inference system, fuzzy c-means clustering and subtractive clustering, are briefly reviewed.

2.1. EPS. EPS, the profit performance of a company, can be formulated as below:

\[
\text{Earnings Per Share} = \frac{\text{Profit}}{\text{Weighted Average Common Shares}}
\]

where the profit (loss) after (before) tax of a company in a quarter minus dividends paid on preferred shares is divided by weighted average number of shares (common stock) issued.

Brown [12] used the time series method to predict the quarterly EPS with a sample size of 23 firms. In the long-term investment, EPS reflects the future estimated value of a company [13]. In the portfolio strategy, managers have the responsibility to aim at “profit” in the limited budget, while EPS can be an indicator to pick out potential stock(s) for investors. When two companies generate the same EPS amounts, the company with less investment would be a better company, since the utilization of its capital is more efficient under the same basis. Hence, investors can select the company share with higher earnings for portfolio investment.

Earnings expectations play an important role in the analytical and empirical literature in finance and accounting [14]. For the issue of EPS prediction, we should pick up the suitable variable, such as “return on total assets”, in financial statements to construct the forecasting model, where financial statements are made on behalf of the constitution.
of a company. Besides, the assessment of a company can not reply on a particular financial method. Moreover, statement analysis and other measures should be taken into consideration.

Callen [15] discussed the relationship between EPS and related indicators by pointing out that cash flow-related indicators have some degree of predictability, and the financial statement, such as the balance sheet or income statement, that reveals the financial status of a business in the past period (mainly quarterly or annual) contains the information of a gain or loss. Sloan [16] investigated if stock prices reflected information about future earnings, contained in the accrual and cash flow components of current earnings. Charitou [17] showed that a positive relationship between cash flows and dividend has a strong connection with EPS. Messod [18] took fundamental analysis for the prediction of extreme stock returns. Zhang [19] also used cash flow information to predict EPS based on the research by Jegadeesh [20].

Data-mining techniques have been applied to EPS prediction. Kenneth [21] showed that logit-based financial statement analysis can predict abnormal returns on investments in equity securities. Qi [22] used the neural network to predict the stock return.

2.2. Autoregressive model. In time-series forecast, predictions are practically obtained by forecasting a value at the next time period based on a specific prediction algorithm. Constructing a predictor for forecasting non-periodic short-term time series is much more difficult than for long-term time series. The autoregressive moving-average (ARMA) is a traditional method that is very suitable for forecasting regular periodic data, such as seasonal or cyclical time series [23]. Box and Jenkins (1976) developed a general linear stochastic model by assuming that time series data can be generated by a linear aggregation of random shocks.

An AR model is a model that includes one or more past values of the dependent variable among its explanatory variables. The simplest AR(1) is defined as:

\[ y_t = \phi_1 y_{t-1} \]  

When the random error and constant term are taken into account, the modified AR(1) model becomes

\[ y_t = \mu + \phi_1 y_{t-1} + u_t, \]  

where \( \phi_1 \) is the first-order autoregression coefficient and \( u_t \) is the white noise viewed as a random error. An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. In the AR(1) model, it can be thought that a given value \( y \) in time period \( t \) has a relationship with time period \( t-1 \). If there is an autoregressive model of order \( p \), an AR(p) model can be expressed as:

\[ y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t \]  

2.3. Adaptive network-based fuzzy inference system. An adaptive network-based fuzzy inference system is a fuzzy Sugeno model that adopts the adaptive systems framework to facilitate learning and adaptation [24,25]. To illustrate the system, we assume the fuzzy inference system that consists of five layers of an adaptive network with two inputs, \( x \) and \( y \), and one output, \( z \). The architecture of ANFIS is shown as Figure 1.

We suppose that the system consists of 2 fuzzy if-then rules based on Takagi and Sugeno’s type:

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

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 described below:
Figure 1. The architecture of the ANFIS network

Layer 1: This Layer is the input layer, and every node \( i \) in this layer is a square node with a node function (see Equation (4)). \( O_{1,i} \) is the membership function of \( A_i \) to specify the degree to which the given \( x \) satisfies the quantifier \( A_i \). Typically, we select the bell-shaped membership function as the input membership function (see Equation (5)) with a maximum value of one and a minimum value of zero.

\[
O_{1,i} = \mu A_i(x) \quad \text{for} \quad i = 1, 2 \tag{4}
\]

\[
\mu A_i(x) = \frac{1}{1 + \left(\frac{x - c_i}{a_i}\right)^2 b_i} \tag{5}
\]

where \( a_i, b_i \) and \( c_i \) are the parameters, and \( b_i \) is a positive value, and \( c_i \) denotes the center of the curve.

Layer 2: Every node in this layer is a square node labeled \( \Pi \), which multiplies the incoming signals and sends the product out by Equation (6).

\[
O_{2,i} = w_i = \mu A_i(x) \times \mu B_i(y) \quad \text{for} \quad i = 1, 2 \tag{6}
\]

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 (7). The output of this layer can be called normalized firing strengths.

\[
O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad \text{for} \quad i = 1, 2 \tag{7}
\]

Layer 4: Every node \( i \) in this layer is a square node with a node function (see Equation (8)). Parameters in this layer will be referred to as consequent parameters.

\[
O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i + q_i + r_i) \tag{8}
\]

where \( p_i, q_i \) and \( r_i \) 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 (see Equation (9)).

\[
O_{5,i} = \sum_i \bar{w}_i f_i = \sum_{i=1}^\infty \frac{w_i f_i}{\sum_{i=1}^\infty w_i} = \text{overall output} \tag{9}
\]
2.4. **Fuzzy C-means clustering.** Clustering is one of the efficient tools for data analysis and data structure visualization. The prevalent formulation of this task is to use \( v_j \) \((j = 1, 2, ..., c)\) 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 [26], is the most famous and basic fuzzy clustering algorithm. The FCM depicted in Equation (4) 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 \( J_{m'} \) under the following constraints:

\[
\begin{align*}
(1) & \quad 0 \leq u_{it} \leq 1, \quad \forall i, t, \\
(2) & \quad 0 < \sum_{t=1}^{n} u_{it} \leq n, \quad \forall i, \\
(3) & \quad \sum_{i=1}^{c} u_{it} = 1, \quad \forall t,
\end{align*}
\]

yields 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

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

where \( j \) is a variable on the feature space, i.e., \( j = 1, 2, ..., m \).

2.5. **Subtractive clustering.** Chiu [27] 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 \( W_i \) \((i = 1, 2, ..., N)\) is presented as \( W_i = (x_i, y_i) \), where \( x_i \) denotes the \( p \) input variables and \( y_i \) is the output variable. The potential value \( P_i \) of a data point is calculated by Equation (13)

\[
P_i = \sum_{j=1}^{N} e^{-\alpha ||W_i - W_j||^2}
\]

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 (14):

\[
p_i = p_i - p_1^* \exp \left( -\beta ||W_i - W_1^*||^2 \right),
\]

where \( \beta \) is a positive constant defining the neighborhood that has measurable reductions in potential, \( W_1^* \) is the first cluster center, and \( P_1^* \) is its potential value.
From Equation (14), the method selects the data point with the highest remaining potential as the second cluster center. The generalized equation based on Equation (14) is presented below:

\[ p_i = p_i - p_k^* \exp \left( -\beta \| W_i - W_k^* \|^2 \right) \]  

(15)

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 \( S_i, i = (1, ..., D) \). With the information above, the membership functions can be defined and the spread can be calculated in accordance with \( \beta \).

3. The Proposed Model. In this section, the research framework and algorithms of the proposed model are described.

3.1. Research framework. Because EPS is as an important indicator of share value to assess a company’s performance by investors [28], the accuracy of EPS predictions will bring a huge profit for investors. Many statistical model [12,14] methods have been applied to forecast the EPS of a company. Further, time series methods, such as the popular AR model, have been applied to deal with finance forecasting problems. Callen [15] showed that the relationship between EPS and EPS-related attributes is nonlinear, and conventional linear time series, such as AR, cannot deal with the nonlinear relationships. For this reason, a data-mining method, such as ANN [19,29,30], is more appropriate to deal with nonlinear problems and can overcome the limitations of most conventional time series methods with the linear relationships assumption. However, the rules generated from ANNs are not easily understood.

An ANFIS model that integrates the advantages of ANN and fuzzy inference system (FIS) employing fuzzy if-then rules can model the qualitative aspects of human knowledge and can be applicable for humans to use by minimizing the output error with a back propagation algorithm and dealing with uncertainty and nonlinear problems.

To enhance forecasting performance, this paper applies the AR time series method to the proposed model. The AR method can utilize current accuracy values to get the next period’s forecast. Based on the advantage above, this paper proposes a novel forecasting model, which includes three facets: (1) take the least-square method to test the lag periods of EPS, (2) use the past periods of EPS based on the AR concept as the input, and then take discretization methods (FCM and subtractive clustering methods) to fuzzify input values and optimize the FIS parameters by the adaptive network and (3) employ the integrated ANFIS model to predict EPS. In practical operations, the procedure of the proposed model, as shown in Figure 2, can be briefly described in three parts, as follows:

(1) Data preprocessing: The quarterly EPS time series data of industry-leading shares are collected from the TEJ database. Then, we test the lag period of EPS by least-square method.

(2) Rule generation: We partition the input variables and set the type of membership function for output variables. Then we generate fuzzy inference systems and train parameters of FIS from training datasets.

(3) Performance evaluation: We forecast the testing EPS(\( t + 1 \)) by four of forecasting models and use RMSE to evaluate the performance of the proposed model. Then, the best model based on minimal RMSE value is selected and the performances of different models are compared.
3.2. The proposed model. The proposed model consists of eight steps: (1) Collect data, (2) Test the lag period of EPS, (3) Define and partition the universe of discourse for input variables, (4) Set the type of membership functions for output variables, (5) Generate fuzzy inference systems, (6) Train parameters of FIS from training datasets, (7) Forecast the testing EPS($t+1$) by the four types of forecasting models and calculate RMSE values, and (8) Select the best model based on the minimal RMSE value, and compare the performance of different models. The specific descriptions of these eight steps are as follows.

**Step 1: Collect data**
We collected the quarterly EPS time series data of industry-leading shares from the TEJ database.

**Step 2: Test the lag period of EPS**
In this step, one to four orders are evaluated to determine which number of time lags is most fitting for the experiment dataset. To estimate the EPS model, we use the E-Views software package to fit time series models for different orders of EPS in 15 quarters. The least-square method is taken to build models, and then four linear regression variables, namely $(EPS(t-1)$ to $EPS(t-4)$), 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 EPS models can be obtained, and the estimated results would provide the reference resources to determine the lag periods of EPS.

**Step 3: Define and partition the universe of discourse for input variables (see B1 block of Figure 2)**
First, define each universe of discourse for the variable $(EPS(t))$ according to the minimum and maximum values. Second, partition the universe of discourse into three linguistic intervals by both fuzzy c-means clustering and subtractive clustering. Eleven membership functions are applied to input variables, such as the difference of two sigmoid membership functions, two-sided Gaussian curve membership function, Gaussian curve membership function, generalized bell curve membership function, pi-shaped curve membership function, product of two sigmoidal membership function, trapezoidal membership function, triangular membership function, Z-shaped curve membership function, S-shaped curve membership function, and sigmoid curve membership function.

**Step 4: Set the type of membership function for output variables (see B2 block of Figure 2)**
There are two types of membership functions for output variables, as follows.

1. Linear type: a typical rule in a Sugeno fuzzy model has the form as follows: If $x_1(EPS(t)) = A_i$, then $y_j(EPS(t+1)) = c_{j0} + c_{j1}x_1$ where $x_1(EPS(t))$ is a linguistic variable, $A_i$ represents a linguistic value (high or low if the attribute is partitioned into two linguistic values by discretization methods for if-then rule), $y_j(EPS(t+1))$ denotes the $j$-th output value, and $c_{j0}$ and $c_{j1}$ are parameters for $i = 1, 2$ and $j = 1, 2$.

2. Constant type: the output level EPS($t+1$) for a zero-order Sugeno model is a constant ($c_{j1} = 0$).

**Step 5: Generate fuzzy inference systems**
From Steps 3 and 4, four types of forecasting models are used to generate fuzzy inference systems; i.e., (1) fuzzy c-means with linear type (FCM.L), (2) fuzzy c-means with constant type (FCM.C), (3) subtractive clustering with linear type (Subclust.L), and (4) subtractive clustering with constant type (Subclust.C). Generate four different fuzzy inference systems according to the four types of forecasting models.

**Step 6: Train parameters of FIS from training datasets**
In the training phase, the least-square method and back-propagation gradient descent method are employed for training the forecasting models. The purpose of the process is to
optimize the parameters of if-then rules. This study sets the error as 0.05 for the training stopping criterion and then obtains the parameters for the if-then rules.

**Step 7:** Forecast the testing EPS(t+1) by the four types of forecasting models and calculate RMSE values
First, the FIS parameters of the four types of forecasting models are determined when the stopping criterion is reached from Step 6. Later, the four training forecasting models are used to forecast EPS\((t+1)\) for the target testing datasets. Second, calculate four RMSE values in testing datasets by Equation (16).

\[ \sqrt{\frac{\sum_{t=1}^{n} (\text{actual}(t) - \text{forecast}(t))^2}{n}}; \quad (16) \]

where \(\text{actual}(t)\) denotes the real EPS value, \(\text{forecast}(t)\) denotes the predicted EPS value, and \(n\) is the number of data.

**Step 8: Select the best model based on the minimal RMSE value and compare the performance of the different models**

Based on the minimal RMSE value for the target testing datasets from Step 7, the best forecasting model among the four models can be obtained. The minimal RMSE value is taken as an evaluation criterion to compare the four different models.

4. Verification and Comparison of Different Models. An empirical case study is provided in this section to illustrate the proposed model, and collected data is used as the experimental dataset to demonstrate the practicability of the proposed model.

4.1. Verification of the proposed model. To verify the proposed model, the experimentation using the EPS-related data from the TEJ database is implemented. The 15-quarter data from 2005 Q1 to 2008 Q3 are used to forecast the eight leading industry shares in seven industries. The company names of these eight shares in the seven industries, including the stock codes are (1) HTC (electronic industry, stock code: 2498); (2) FORMOSA INTERNATIONAL HOTELS CORPORATION, GFRT (tourism industry, stock code: 2707); (3) EVERGREEN MARINE CORPORATION, EMC (shipping industry, stock code: 2603); (4) FENG HSIN and CHIA TA (iron and steel industry, the respective stock codes: 2015 and 2033); (5) HUAKU (construction industry, stock code: 2548); (6) UNI-PRESIDENT (food industry, stock code: 1216); and (7) TAIWAN CEMENT CORPORATION, TCC (cement industry, stock code: 1101).

The procedures of the EPS prediction are summarized below in accordance with the eight steps of the proposed model.

**Step 1: Collect data**

In this step, we choose an EPS dataset for HTC (electronic industry, stock code: 2498). Table 1 shows the 15-quarter EPS data of HTC. When the 15-quarter data are ranked in ascending order, two-thirds of the data, namely, the previous 10 quarters, are used as the training data, whereas the rest of the data are for testing purposes.

**Step 2: Test the lag period of EPS**

To test the lag periods of EPS, take the electronic industry data set HTC as an example. Four linear regression variables, \(\text{EPS}(t-1)\) to \(\text{EPS}(t-4)\), are selected for the statistical test with the \(p\)-value given at 0.05. Table 3 shows that the \(p\)-value of 0.0033 for \(\text{EPS}(t-1)\) is less than 0.05. Therefore, the order of AR is one. When the order of AR for the lag period is set to one, the results generated by the training set and testing test are shown in Table 2 and Table 3, respectively.

**Step 3: Define and partition the universe of discourse for input variables**

We define each universe of discourse for variable (\(\text{EPS}(t)\)) and then partition the universe of discourse into linguistic variables by using the FCM for input variable (\(\text{EPS}(t)\)). In addition, attributes are partitioned into two linguistic values. We also set the triangle-type membership function for input variables.

**Step 4: Set the type of membership functions for output variables**
Table 1. Fifteen-quarter EPS data for HTC

<table>
<thead>
<tr>
<th>Date</th>
<th>EPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/03</td>
<td>5.55</td>
</tr>
<tr>
<td>2005/06</td>
<td>7.07</td>
</tr>
<tr>
<td>2005/09</td>
<td>7.48</td>
</tr>
<tr>
<td>2005/12</td>
<td>14.35</td>
</tr>
<tr>
<td>2006/03</td>
<td>15.11</td>
</tr>
<tr>
<td>2006/06</td>
<td>13.15</td>
</tr>
<tr>
<td>2006/09</td>
<td>15.37</td>
</tr>
<tr>
<td>2006/12</td>
<td>16.98</td>
</tr>
<tr>
<td>2007/03</td>
<td>12.78</td>
</tr>
<tr>
<td>2007/06</td>
<td>10.42</td>
</tr>
<tr>
<td>2007/09</td>
<td>12.96</td>
</tr>
<tr>
<td>2007/12</td>
<td>17.43</td>
</tr>
<tr>
<td>2008/03</td>
<td>12.12</td>
</tr>
<tr>
<td>2008/06</td>
<td>8.76</td>
</tr>
<tr>
<td>2008/09</td>
<td>9.25</td>
</tr>
</tbody>
</table>

**Figure 3.** Time lag period for HTC

In this step, the constant type of membership functions is set for the output variable (EPS(t + 1)).

**Step 5: Generate fuzzy inference systems**

Attributes are partitioned by FCM into two linguistic values by the following two rules:
If $x_i(EPS(t)) = A_i$, then $y_j(EPS(t + 1)) = c_{j0}$ where $x_i(EPS(t))$ is a linguistic variable, $A_i$ is a linguistic value (high, low), $y_j(EPS(t + 1))$ denotes the $j$-th output value and $c_{j0}$ is a parameter with $i = 1, 2$ and $j = 1, 2$.

**Step 6: Train the parameters of FIS from training datasets**
Table 2. Training set of data set for HTC

<table>
<thead>
<tr>
<th>Date(t)</th>
<th>EPS(t)</th>
<th>EPS(t+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/03</td>
<td>5.55</td>
<td>7.07</td>
</tr>
<tr>
<td>2005/06</td>
<td>7.07</td>
<td>7.48</td>
</tr>
<tr>
<td>2005/09</td>
<td>7.48</td>
<td>14.35</td>
</tr>
<tr>
<td>2005/12</td>
<td>14.35</td>
<td>15.11</td>
</tr>
<tr>
<td>2006/03</td>
<td>15.11</td>
<td>13.15</td>
</tr>
<tr>
<td>2006/06</td>
<td>13.15</td>
<td>15.37</td>
</tr>
<tr>
<td>2006/09</td>
<td>15.37</td>
<td>16.98</td>
</tr>
<tr>
<td>2006/12</td>
<td>16.98</td>
<td>12.78</td>
</tr>
<tr>
<td>2007/03</td>
<td>12.78</td>
<td>10.42</td>
</tr>
</tbody>
</table>

Table 3. Testing set of data set for HTC

<table>
<thead>
<tr>
<th>Date(t)</th>
<th>EPS(t)</th>
<th>EPS(t+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007/06</td>
<td>10.42</td>
<td>12.96</td>
</tr>
<tr>
<td>2007/09</td>
<td>12.96</td>
<td>17.43</td>
</tr>
<tr>
<td>2007/12</td>
<td>17.43</td>
<td>12.12</td>
</tr>
<tr>
<td>2008/03</td>
<td>12.12</td>
<td>8.76</td>
</tr>
<tr>
<td>2008/06</td>
<td>8.76</td>
<td>9.25</td>
</tr>
</tbody>
</table>

In this step, the parameters of the membership functions were optimized on the HTC training dataset through back-propagation, while the consequent parameters were calculated using a linear least-square method. The training error for this optimization was set to 0.05.

Step 7: Forecast the testing EPS(t+1) by the four types of forecasting models and calculate RMSE values

The FIS parameters of the forecasting models are determined when the stopping criterion is reached from Step 6. Later, the four training forecasting models are used to forecast EPS(t + 1) for the target testing data.

Step 8: Select the best model based on minimal RMSE value, and compared the performance of the four different models

Based on the minimal RMSE value for the target testing datasets from Step 7, the best forecasting model among the four models can be obtained. A test data set of HTC is provided, and the evaluation metric is the RMSE value by Equation (16). The model performance is examined by RMSE criterion to compare these four models. From Table 4, we can see that the proposed model outperforms the listing models.

Table 4. Results for HTC data set (time lag = 1)

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1) [5]</td>
<td>3.63</td>
</tr>
<tr>
<td>linear-Regression [31]</td>
<td>3.15</td>
</tr>
<tr>
<td>MLP [32]</td>
<td>3.38</td>
</tr>
<tr>
<td>RBF [33]</td>
<td>2.96</td>
</tr>
<tr>
<td>Proposed Model</td>
<td><strong>2.79</strong></td>
</tr>
</tbody>
</table>

4.2. Comparison of different models. The performance of the proposed model is compared with the listing models, including the AR(1) model, linear regression, and neural network (MLP, RBF). The forecasting performance of the four models is listed
in Table 5. From Table 5, the performance of the proposed model (AR(1) + ANFIS) is superior to the other four models in 5 out of 8 testing datasets.

Table 5. Results for the experiment (time lag = 1)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2498</td>
<td>3.63</td>
<td>3.15</td>
<td>3.38</td>
<td>2.96</td>
<td><strong>2.79</strong>*</td>
</tr>
<tr>
<td>2707</td>
<td>1.11</td>
<td>1.09</td>
<td>0.96</td>
<td>1.03</td>
<td><strong>0.85</strong>*</td>
</tr>
<tr>
<td>2603</td>
<td><strong>0.53</strong>*</td>
<td>0.73</td>
<td>0.89</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>2015</td>
<td>1.74</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td><strong>1.24</strong>*</td>
</tr>
<tr>
<td>2033</td>
<td>0.66</td>
<td>0.6</td>
<td>0.64</td>
<td>0.6</td>
<td><strong>0.58</strong>*</td>
</tr>
<tr>
<td>2548</td>
<td><strong>0.86</strong>*</td>
<td>1.43</td>
<td>1.57</td>
<td>1.32</td>
<td>1.24</td>
</tr>
<tr>
<td>1216</td>
<td><strong>0.31</strong>*</td>
<td>0.35</td>
<td>0.32</td>
<td>0.35</td>
<td>1.24</td>
</tr>
<tr>
<td>1101</td>
<td>0.36</td>
<td>0.25</td>
<td>0.27</td>
<td>0.25</td>
<td><strong>0.24</strong>*</td>
</tr>
</tbody>
</table>

*The best performance among the five models

Based on the experimental results, the proposed model, using subtractive clustering for ANFIS, can generate fuzzy rules with a comprehensive way to explain the variation of value [34]. Besides, the linguistic values of the attributes can be used to predict future EPS according to the fuzzy rules. In our experiments, there are three rules generated by subtractive clustering and four rules generated by FCM at most. As a result, the few fuzzy rules can be applied to predict EPS efficiently.

5. Findings. From the experimental results, there are two findings in this paper, as follows:

(1) According to Table 5, it is evident that the proposed model is superior to the listing methods in terms of RMSE. The main reason is that proposed model takes into account the AR model with ANFIS learning for EPS forecasting.

(2) From the empirical results, the best performance of the four types of forecasting models is subtractive clustering with linear type (Subclust_L). In our experiments, at most 4 rules are generated by subtractive clustering and thus a few fuzzy rules can be applied to predict EPS efficiently.

6. Conclusion. In this paper, we propose an integrated ANFIS models, namely AR(1) + ANFIS, by adopting the AR concept (the lag periods of AR is one) into ANFIS in the experiment. In the verification phase, the stock data of Taiwan are used to evaluate the forecasting performance of the proposed model and the other four models. The experimental results reveal that the proposed model outperforms the listing models. Moreover, decision-makers and future research can use the rules generated by the proposed model of this paper to forecast EPS and then to generate huge profits.

In future works, we can apply the fusion models to forecast other financial datasets, such as individual stock price and the futures market. In addition, other defuzzification methods in forecasting processes, such as MEPA (minimal entropy principle approach), can be employed to evaluate performance.

REFERENCES


