A Hybrid Model Based on ANFIS and Empirical Mode Decomposition for Stock Forecasting

Liang-Ying Wei

Abstract—Time series forecasting is an important and widely interesting topic in the research of system modeling and stock price forecasting is the most important research issues in time series forecasting. Accurate stock price forecasting is regarded as a challenging task of the financial time series forecasting process., This paper proposes a hybrid time-series adaptive network based fuzzy inference system (ANFIS) model based on empirical mode decomposition (EMD) to forecast stock price for Taiwan stock exchange capitalization weighted stock index (TAIEX). In order to evaluate the forecasting performances, the proposed model is compared with autoregressive (AR) model, ANFIS model and support vector regression (SVR) model. The experimental results show that the proposed model is superior to the listing models in terms of root mean squared error (RMSE).

Index Terms—Adaptive network based fuzzy inference system (ANFIS), empirical mode decomposition (EMD), TAIEX forecasting.

I. INTRODUCTION

In recent years, many time series models have been proposed and applied to handle the area of economic forecasting. Engle [1] proposed the ARCH (p) (Autoregressive Conditional Heteroscedasticity) model that has been used by many financial analysts and the GARCH [2] (Generalized ARCH) model is the generalized form of ARCH. Box and Jenkins [3] 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. Models that describe such homogeneous non-stationary behavior can be obtained by supposing some suitable difference of the process to be stationary. Therefore, the autoregressive integrated moving average model (ARIMA) [3] with the assumption of linearity among variables was proposed to handle the non-stationary behavior datasets. From the literatures, the autoregressive (AR) model is the fundamental method in time series models.

However, traditional time series method requires more history data and the data must be normal distribution to get a better forecasting performance. Besides, linguistic expressions are often used to describe daily observations. Hence, Song and Chissom [4] first proposed the original model of the fuzzy time-series. In the fuzzification process, the length of intervals for the universe of discourse could affect forecasting, and Huarng proposed distribution-based and average-based length, to approach this issue [5]. In addition, Chen proposed a new method [6] which the length of linguistics intervals are tuned by genetic algorithms. In the process of establishing fuzzy relationships and forecasting, Yu [7] argued that recurrent fuzzy relationships should be considered in forecasting and recommended that different weights should be assigned to various fuzzy relationships. Therefore, Yu [7] proposed a weighted fuzzy time-series method to forecasting the TAIEX. To take advantage of neural networks (nonlinear capabilities), Huarng and Yu [8] chosen a neural network to establish fuzzy relationships in fuzzy time-series, which are also nonlinear, but the process of mining fuzzy logical relationships is not easily understandable just like a black box [9].

Many researchers have applied the data mining techniques to financial analysis. Kimoto *et al.*, [10] developed a prediction system for stock market by using neural network. Nikolopoulos and Fellrath [11] combined genetic algorithms and neural network to develop a hybrid expert system for investment advising. Kim and Han [12] proposed genetic algorithms approach to feature discretization and the determination of connection weights for artificial neural networks (ANNs) to predict the stock price index. Roh [13] integrated neural network and time-series model for forecasting the volatility of stock price index.

Empirical mode decomposition (EMD), proposed by Huang *et al.*, [14], is suitable for nonlinear signal analysis, which adaptively represents the local characteristic of the given signal [14]. Based on EMD, any complicated signal can be decomposed into a finite and often small number of Intrinsic Mode Functions (IMFs) [14], which have simpler frequency components and stronger correlations, thus are easier and more accurate to forecast. Therefore, the EMD has been widely used in many fields, such as in the analysis of earthquake signal and structure analysis, bridge and constructions state monitoring [15], sea wave data [16], and the fault diagnosis of machines [17].

This paper considers that EMD can decompose the complicated raw data (TAIEX data) into simpler frequency components and highly correlations variables, which is adopted into AR model for building the primary model. Then, the results of primary model will be refined and optimized by an adaptive network based fuzzy inference system (ANFIS), which employs fuzzy if-then rules can model the qualitative aspects of human knowledge and can be applicable for human to use.

II. RELATED WORKS

Stock market is one of the most exciting and challenging monetary activities. The market climates are dramatically

changed in a second and the gain-loss is decided in a twinkling decision. Hence, many researches have presented different methods to deal with forecasting stock price problems. Huarng *et al.*, [18] have used the volatility of NASDAQ (the largest US electronic stock market) stock index and Dow Jones (Dow Jones Industrial Average) stock index to forecast Taiwan stock index.

Time-series models have been applied to handle economic forecasting, such as stock index forecasting, and various models have been proposed. Engle [1] proposed the ARCH (Autoregressive Conditional Heteroscedasticity) model that has been used by many financial analysts and the GARCH [2]. Box and Jenkins [3] proposed the autoregressive moving average (ARMA) model that combines a moving average process with a linear difference equation to obtain an autoregressive moving average model. During the past years, many researchers have applied the data mining techniques to financial analysis. Huarng and Yu [8] apply the backpropagation neural network to establish fuzzy relationships in fuzzy time-series for forecasting stock price. Kinoto *et al.*, [10] developed a prediction system for stock market by using neural network. Nikolopoulos and Fellrath [11] combined genetic algorithms and neural network to develop a hybrid expert system for investment advising. Kim and Han [12] proposed genetic algorithms approach to feature discretization and the determination of connection weights for artificial neural networks (ANNs) to predict the stock price index. Roh [13] integrated neural network and time-series model for forecasting the volatility of stock price index. Thawornwong and Enke [19] proposed redeveloped neural network models for predicting the directions of future excess stock return. Kim [20] applied SVM (Support vector machine) to predicting the stock price index.

The empirical mode decomposition (EMD) technique, proposed by Huang *et al.*, [14], is a form of adaptive time series decomposition technique using the Hilbert-Huang transform (HHT) for nonlinear and non-stationary time series data. The basic principle of EMD is to decompose a time series into a sum of oscillatory functions, namely, intrinsic mode functions (IMFs). In the EMD, the IMFs must satisfy two conditions: (1) the number of extrema (sum of maxima and minima) and the number of zero crossing differs only by one, and (2) the local average is zero. The condition that the local average is zero implies that envelope mean of the upper envelope and lower envelope is equal to zero. The first condition is similar to the traditional narrow band requirements for a stationary Gaussian process [14]. The second condition modifies classical global requirement to a local one; it is necessary so that the instantaneous frequency will not have the unwanted fluctuations induced by asymmetric wave forms [14].

III. PROPOSED MODEL

From the literature reviews, there are some major drawbacks in stock forecasting models: (1) for some statistical models, specific assumptions are required for observations, and those models cannot be applied to the datasets that do not follow the statistical assumptions; (2) most time-series models utilize late day stock price as input

variable in forecasting. However, there are many noises involutedly in TAIEX raw data that are caused by changes in market conditions and environments. In order to overcome the drawbacks above, this paper considers that EMD can decompose the noise raw data (TAIEX data) into simpler frequency components and highly correlations variables, which is adopted AR model for building the primary model. Then, the results of primary model will be refined by ANFIS, which can overcome the limitations of statistical methods (data need obey some mathematical distribution) and handle noise data involutedly. The overall flowchart of the proposed model is shown as Fig. 1.
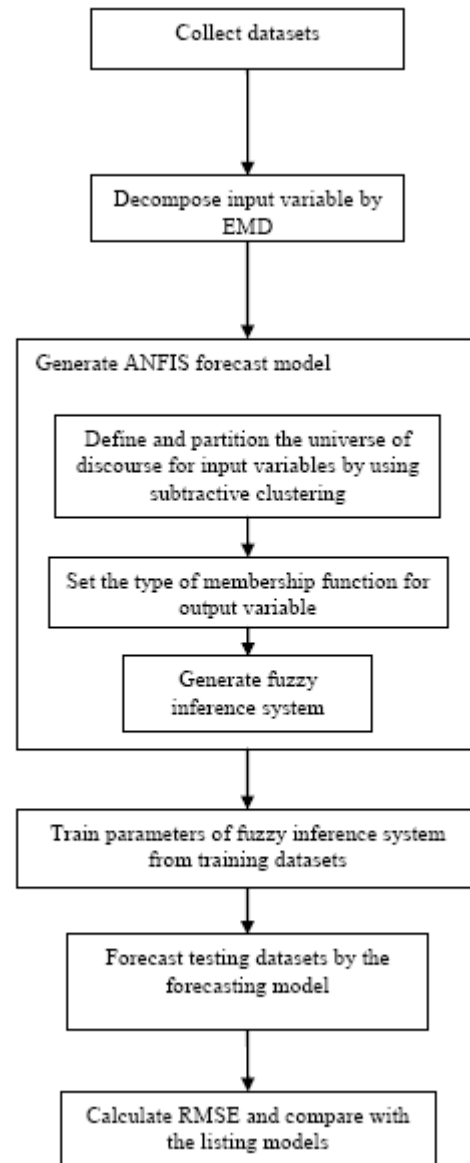

Fig. 1. Flowchart of the proposed procedure.

This section uses practically collected data as the example step by step shows the core concept of proposed algorithm as follows:

Step 1: collect data set in this step, we collect TAIEX from 2000 to 2003 (seven sub-datasets) to illustrate proposed model. In each year, every training dataset is selected from January to October and the remainder dataset (from November and December) is used for testing.

Step 2: Decompose input variable by EMD

In order to obtain interpretable information of input variable (TAIEX (t)), we use the EMD to decompose the input variable (TAIEX (t)) into a finite set of IMFs (the residual $r_n(t)$ also be considered as an IMF). In this study, the TAIEX (t) is decomposed into seven IMFs and one residue (just seven IMFs in each year dataset), which exhibits a stable and regular variation.

Step 3: Generate ANFIS forecast model in this paper, ANFIS method uses subtractive clustering [21] to partition the universe of discourse for input variables, and then generates the fuzzy inference system.

Step 4: Train fuzzy inference system parameters from training datasets

In this step, we employ the least-squares method and the back-propagation gradient descent method for training the forecasting model. This study sets epoch as 1000 for the training stopping criterion, and then obtains the parameters for the selected output membership function.

Step 5: Forecast testing datasets by the trained model

The fuzzy inference system parameters of the forecasting models are determined when the stopping criterion is reach from step 4, then the training forecasting model are used to forecast TAIEX($t+1$), for the target testing datasets.

Step 6: Calculate RMSE and compare with the listing models

Calculate RMSE values in testing datasets by Equation (16). Then the RMSE is taken as evaluation criterion to compare with the listing models.

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

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

IV. EXPERIMENTS AND COMPARISONS

This section provides accuracy evaluations and comparisons, and the RMSE is taken as the evaluation criterion. To verify the proposed model, the TAIEX datasets from 2000 to 2003 are used as the experiment datasets, each year TAIEX dataset is named as sub-dataset. Each sub-dataset for the previous 10 months is used for training, and those from November to December are selected for testing. Further, this paper compares the performances of the proposed model with the time-series model, AR (1) [1] model, ANFIS [22] and SVR [23] model. The performances of the listing models above used to forecast TAIEX are compared to the proposed model. The forecasting performances of AR (1) model, SVR model, ANFIS model, and the proposed model are listed in the Table I. From the stock performances comparisons shown in Table I, it is clear that the proposed model surpasses the other two models (AR (1), ε -SVR and ANFIS) in each testing period, except 2005. These stock performance evaluations demonstrate the outstanding performance of the proposed model.

TABLE I: THE RESULTS OF DIFFERENT MODELS FOR TAIEX TESTING DATA IN RMSE

Models	Year			
	2000	2001	2002	2003
AR(1) [1]	130	115	66	54
SVR [23]	136	114	66	59
ANFIS[22]	130	120	65	55
Proposed model	129 [a]	110 [a]	52 [a]	49 [a]

[a] The best performance among four models

V. CONCLUSIONS

This paper has presented a TAIEX forecasting model by integrating EMD and ANFIS. The main contribution of the paper is to propose a novel method as well as a simple approach for a stable prediction of fluctuation data. The proposed method preprocesses TAIEX (t) data and decomposes them into more stationary and regular components (IMF or residue) using the EMD technique. Furthermore, the corresponding ANFIS model for each divided component is easier to build. After the IMF components and residue are forecasted in the built SVR model, the forecasting values are as the TAIEX $(t+1)$ forecasting results. This study compared the proposed method with the ANFIS, SVR and AR (1) models, using RMSE as its criteria. Experimental results showed that the proposed EMD-ANFIS model could produce the lowest forecasting error in the datasets. Proposed model outperformed the ANFIS, SVR and traditional AR (1) models. According to the experiments, EMD, which can fully capture the local fluctuations of data, can be used as a preprocessor to decompose the noise raw data into a finite set of IMFs and a residue, which have simpler frequency components and highly correlations. By this preprocessing, we can not only advance the simplification of ANFIS modeling, but also obtain a much precise than AR (1) model based on RMSE. Therefore, the proposed method is very suitable for prediction with nonlinear and strong noise data, and is an efficient method for TAIEX forecasting. Furthermore, by implementing this experiment, two advantages were discovered for the proposed model: (1) the proposed model can produce more reasonable and understandable rules, because the "if-then" rules produced by ANFIS can model the qualitative aspects of human knowledge; and (2) the proposed model can provide stock investors with objective suggestions (forecasts) to make investment decisions in the stock market, because the proposed model produces forecasting rules based on objective stock data rather than subjective human judgments.

REFERENCES

[1] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, 1982, pp. 987-1008.
[2] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, vol. 31, pp. 307-327, 1986.
[3] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1976.
[4] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time-series Part I," *Fuzzy Sets and Systems*, vol. 54, 1993, pp. 1-10.

[5] K. H. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series," *Fuzzy Sets and Systems*, vol. 123, 2001, pp. 155–162.

[6] S. M. Chen and N. Y. Chung, "Forecasting enrollments using high-order fuzzy time series and genetic algorithms," *International of Intelligent Systems*, vol. 21, 2006, pp. 485-501.

[7] H. K. Yu, "Weighted fuzzy time-series models for TAIEX forecasting," *Physica A.*, vol. 349, 2005, pp. 609-624.

[8] K. H. Huarng and T. H. K. Yu, "The application of neural networks to forecast fuzzy time series," *Physica A.*, vol. 336, 2006, pp. 481-491.

[9] T. L. Chen, C. H. Cheng, and H. J. Teoh, "High-order fuzzy time-series based on multi-period adaptation model for forecasting stock markets," *Physica A.*, vol. 387, 2008, pp. 876-888

[10] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural network," in *Proc. the International Joint Conference on Neural Networks*, San Diego, California, pp. 1-6, 1990.

[11] C. Nikolopoulos and P. Fellrath, "A hybrid expert system for investment advising," *Expert Systems*, vol. 11, no. 4, 1994, pp. 245-250.

[12] K. Kim and I. Han, "Genetic algorithms approach to feature discretization in artificial neural networks for prediction of stock index," *Expert System with Applications*, vol. 19, 2000, pp. 125-132.

[13] T. H. Roh, "Forecasting the volatility of stock price index," *Expert Systems with Applications*, vol. 33, 2007, pp. 916-922.

[14] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, and Q. Zheng, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis,," in *Proc. the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, series A, vol. 454, 1998, pp. 903-995.

[15] H. T. Vincent, S. L. J. Hu, and Z. Hou, "Damage detection using empirical mode decomposition method and a comparison with wavelet analysis," in *Proc. the Second International Workshop on Structural Health Monitoring*, Stanford, 1999, pp. 891-900.

[16] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: the Hilbert spectrum," *Annu. Rev. Fluid Mech.*, vol. 31, 1999, pp. 417 – 457.

[17] D. J. Yu, J. S. Cheng, and Y. Yang, "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings," *Mech. Syst. Signal Process*, vol. 19, no. 2, 2005, pp. 259-270.

[18] K. H. Huarng, H. K. Yu, and Y. W. Hsu, "A Multivariate Heuristic Model for Fuzzy," *IEEE Transactions on Systems*, Man, and Cybernetics-Part B: Cybernetics, vol. 37, no. 4, 2007, pp. 836-846.

[19] S. Thaworwong and D. Enke, "The adaptive selection of financial and economic variables for use with artificial neural networks," *Neurocomputing*, vol. 56, 2004, pp. 205-232

[20] K. J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, 2003, pp. 307-319

[21] S. L. Chiu, "Fuzzy model identification based on cluster estimation," *J. of Intelligent and Fuzzy Systems*, vol. 2, 1994, PP. 267-278.

[22] J. S. Jang, "Adaptive-network-based fuzzy inference systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, 1993, pp. 665-685.

[23] V. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.

Liang-Ying Wei works in the Department of Information Management, Yuanpei University. His research is mainly in the field of fuzzy time series, soft computing, machine learning, data mining and bioinformatics.