RESEARCH ARTICLE

# A novel stock forecasting model based on High-order-fuzzy-fluctuation Trends and Back Propagation Neural Network

Hongjun Guan[1], Zongli Dai[1], Aiwu Zhao[2]*, Jie He[1]

**1** School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, Shandong, China, **2** School of Management, Jiangsu University, Zhenjiang, Jiangsu, China

* aiwuzh@126.comor, aiwuzh@ujs.edu.cn

## Abstract

In this paper, we propose a hybrid method to forecast the stock prices called High-order-fuzzy-fluctuation-Trends-based Back Propagation(HTBP)Neural Network model. First, we compare each value of the historical training data with the previous day's value to obtain a fluctuation trend time series (FTTS). On this basis, the FTTS blur into fuzzy time series (FFTS) based on the fluctuation of the increasing, equality, decreasing amplitude and direction. Since the relationship between FFTS and future wave trends is nonlinear, the HTBP neural network algorithm is used to find the mapping rules in the form of self-learning. Finally, the results of the algorithm output are used to predict future fluctuations. The proposed model provides some innovative features:(1)It combines fuzzy set theory and neural network algorithm to avoid overfitting problems existed in traditional models. (2)BP neural network algorithm can intelligently explore the internal rules of the actual existence of sequential data, without the need to analyze the influence factors of specific rules and the path of action. (3)The hybrid modal can reasonably remove noises from the internal rules by proper fuzzy treatment. This paper takes the TAIEX data set of Taiwan stock exchange as an example, and compares and analyzes the prediction performance of the model. The experimental results show that this method can predict the stock market in a very simple way. At the same time, we use this method to predict the Shanghai stock exchange composite index, and further verify the effectiveness and universality of the method.

## Introduction

Forecasting is an important means of reducing risk and increase revenue in financial sector. Stock price prediction models can be divided into two categories: statistical model and artificial intelligence model. The former models include ANFIS [1], ARIMA [2], ARCH [3], GARCH [4], and so on. In such models, the variables must strictly obey the restrictive assumptions of linear or normal distribution. However, because of the uncertainty and complexity of the stock market, it is difficult to make out a strict normal assumption for a linear prediction model. Wang [5] studied the relationship between stock price and the changing of investors'

social network. He established a mathematical model based on fuzzy method. However, such models based on external factors are varying from different stock markets. What is more, other external factors, such as economic environment, policy changing and so on also have great relationship with the fluctuation of a stock market. In fact, historical data can somewhat reflect the internal rule for the evolution of a stock market. Artificial intelligence models can reveal the internal rule and therefore achieve the desired results without any strict assumptions. Such models have better nonlinear processing capabilities, so many researchers have applied it to the prediction of various fields [6–8], such as Mishra [9]use it forecast the PM 2.5 during haze episodes. Raza [10]proposed artificial intelligence method to forecast the load demand of smart grid.

Artificial Neural Networks(ANNs) is a machine learning algorithm that simulates human brain learning, which is suitable for the calculation and prediction of complex systems [11]. ANNs can find the mapping relationship between variables in any precision, and it also has good self-adaptability, self-organization, self-learning ability and generalization ability [12]. These characteristics can meet the demand of the general market trend forecast. Dumitru [13] propose that the prediction method based on ANNs is more suitable for multi-variable prediction, especially for wind market forecasting. John [14]analyzed the characteristics of self-adaptive and self-learning to explore the rules of historical rainfall data. As can be seen from these examples, ANNs are especially suitable for general system prediction in various fields. In stock market forecasting, the back propagation Neural Network(BPNN) models, which is a type of the ANNs, is applied to predict the daily Shanghai Stock Exchange Composite Index [15]. However, when the stock market has complicated situation, for example, the fluctuation is more frequent and the fluctuation amplitude is relatively large, the method expose certain limitations. First of all, the stock market trading is not only non-linear, but also chaotic. Therefore, it is difficult to accurately predict the stock market trend by relying on a single neural network. In addition, due to the uncertainty of stock market fluctuation, ANNs is more prone to over-training and over-fitting.

The hybrid models which combined ANNs with other approaches have been applied to stock forecasting area due to superior performance than individual models. At present, the comprehensive model of stock market forecasting based on neural network is divided into two aspects; On the one hand, in view of the original noise problem existing in stock market forecast, some researchers attempt to combine the fuzzy set theory with BPNN to reduce the noise of the stock data. Tu [16] proposed the RSEIT2FNN model based on type-2 fuzzy theory and neural network learning algorithm, which combined more recent research achievements on fuzzy theory with neural network. Other scholars established models with fuzzy classification and prediction methods and studied them in more application areas [17–18]. Given the chaotic state of stock market values, Song and Chissom propose the fuzzy time series (FTS) forecasting model [19–21] based on the fuzzy set theory for the first time.Then, other scholars try to combine neural network and fuzzy time series, such as Aladag et al [22]use BPNN to determine fuzzy relations in their fuzzy time series method. The approach combined FTS with ANNs is effective, and that have been widely applied to stock index forecasting [23–24]. But these methods are prone to over-fuzzy, which leads to the reduction of the regular information contained in the original stock market value. On the other hand, for the overtraining and over-fitting of neural network algorithm in the study of stock data, some scholars have tried to combine the stock volatility with BPNN. Wang [25]propose a new approach to forecasting the stock prices via the Wavelet De-noising-based Back Propagation (WDBP) neural network. The model discusses the accuracy of prediction from the Angle of stock market fluctuation, but it does not solve the problem of noise and chaos. Generally speaking, the stock market is influenced by a variety of factors, dynamic, multivariate complex systems, so it is necessary to

explore the optimal solution through technology integration. So far, in terms of stock market forecasting, the combination of fuzzy, fluctuating and BPNN synthesis is very rare.

The aim of this paper is to propose a new neural model to improve learning efficiency and predictive power. Therefore, we propose a hybrid forecasting method called High-order-fuzzy-fluctuation-Trends-based Back Propagation(HTBP)neural network modal. In such a model, the original data are first decomposed into multiple layers by the High-Order-Fuzzy-Fluctuation series. The algorithm node is consistent with the order of the wave sequence. This paper is the first attempt to utilize the HTBP based algorithm for forecasting the stock prices. The advantages of the model can be summarized as follows: (1)It combines fuzzy set theory and neural network algorithm to avoid overfitting problems existed in traditional models. (2) BP neural network algorithm can intelligently explore the internal rules of the actual existence of sequential data, without the need to analyze the influence factors of specific rules and the path of action. (3)The hybrid modal can reasonably remove noises from the internal rules by proper fuzzy treatment. The HTBP model is used to predict the stock market from 1997 to 2005 using the TAIEX data set and Shanghai Stock Exchange Composite Index (SHSECI) from 2007 to 2015. Furthermore, the superiority of our model is shown by comparing the HTBP with a traditional model based single BP neural network. we also compare the prediction results with several other existing methods, and conclude that the prediction effect of the model is better than the general prediction model.

The remainder of this paper is organized as follows: Section 2 introduces some research on fuzzy time series and the concept and model of BP neural network. Section 3 describes a prediction method based on BP neural network and fuzzy wave trends and logical relationships. In section 4, the model is used to predict the stock market from 1997 to 2005 using different data set. In section 5, summarize the conclusions and potential problems of future research.

## Preliminaries

### Definition of fuzzy-fluctuation time series (FFTS)

Song and Chissom [19–21]combined fuzzy set theory with time series and presented the definitions of fuzzy time series. In this section, we will extend the fuzzy time series to fuzzy-fluctuation time series (FFTS) and propose the related concepts.

**Definition 1.** Let $L = \{l_1, l_2, \ldots, l_g\}$ be a fuzzy set in the universe of discourse U; it can be defined by its membership function, $\mu_L: U \rightarrow [0,1]$, where $\mu_L(u_i)$ denotes the grade of membership of $u_i$, $U = \{u_1, u_2, \ldots u_i, \ldots, u_l\}$.

The fluctuation trends of a stock market can be expressed by a linguistic set $L = \{l_1, l_2, l_3\} =$ {down, equal, up}. The element $l_i$ and its subscript $i$ is strictly monotonically increasing [26], so the function can be defined as follows, $f: l_i = f(i)$. To preserve all of the given information, the discrete $L = \{l_1, l_2, \ldots, l_g\}$ also can be extended to a continuous label $\bar{L} = \{l_a | a \in R\}$, which satisfies the above characteristics. $\bar{L}'$ is defined as forecasting value. M is defined as a constant to scale the range of $\bar{S}(i)$ to facilitate machine learning. $\bar{Q}(i)$ is defined as the s value after scaling.

**Definition 2.** Let $F(t)(t = 1, 2, \ldots, T)$ be a time series of real numbers, where T is the number of the time series $G(t)$ is defined as a fluctuation time series, where $G(t) = F(t) - F(t - 1)$, $(t = 2, 3, \ldots, T)$. Each element of $G(t)$ can be represented by a fuzzy set $S(t)(t = 2, 3, \ldots, T)$ as defined in Definition 1. Then we call time series $G(t)$ to befuzzified into a fuzzy-fluctuation time series (FFTS) $S(t)$.

**Definition 3.** Let $S(t)(t = n + 1, n + 2, \ldots, T, n \geq 1)$ be a FFTS. If $S(t)$ is determined by $S(t - 1), S(t - 2), \ldots, S(t - n)$, then the fuzzy-fluctuation logical relationship is represented by:

$$S(t - 1), S(t - 2), \ldots, S(t - n) \rightarrow S(t) \tag{1}$$

and it is called the nth-order fuzzy-fluctuation logical relationship (FFLR) of the fuzzy-fluctuation time series, where $S(t - n), \ldots, S(t - 2)S(t - 1)$ is called the left-hand side(LHS) and $S(t)$ is called the right-hand side(RHS) of the FFLR, and $S(k)(k = t, t - 1, t - 2, \ldots, t - n) \in L$. The fuzzy-fluctuation logical relationship can also be represented by:

$$S(t - 1), S(t - 2), \ldots, S(t - n) \rightarrow \bar{S}(t) \qquad (2)$$

$\bar{S}(t)$ is introduced to preserve more information, as described in Definition 1.

$$\bar{Q}(i + 1) = \bar{S}(i + 1)/M \qquad (3)$$

$\bar{Q}(i)$ is introduced to help the Machine learning, as described in Definition 1.

$$S(t - 1), S(t - 2), \ldots, S(t - n) \rightarrow \bar{Q}(t) \qquad (4)$$

## Basic concept of BP neural network

BP Neural Network belongs to a hierarchical network with powerful nonlinear processing ability. It doesn't need to know the relationship between the form or the variable of the data distribution. It can spontaneously organize training and learning based on the observed training data. In addition, it establishes a nonlinear mapping between the number of variables and the output. The principle of the network is based on the external feedback of the network, and the weight of the network mapping control variables is realized by adjusting the values of the neural network parameters to minimize errors. Based on BP Neural Network algorithm, we can predict future stock market fluctuations by using algorithms to learn historical fuzzy fluctuations. The model of the activation function is $\tanh(x)$. Compared to the Sigmoid function, the $\tanh(x)$ has been optimized to overcome the shortcomings of Sigmoid's not zero-centered. The value range of $\tanh(x)$ is $[-1, 1]$.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad (5)$$

The number of input layer nodes of BP Neural Network model is 9, which denote the 9th-order historical fuzzy-fluctuation trends(Fig 1). The number of output layer nodes of the model is 1, which denote the RHS. When the number of hidden layer nodes is 5, the learning effect is best.

$x_i$ represents the input value for each node of the input layer, and $i$ represents the corresponding node number of the input layer. $z_j$ represents the hidden layer node, $w_{ij}$ represents the weight between input layer and the hidden layer node, and $y_j$ represents the output layer node.

## A novel forecasting model based on BP Neural Network

In this paper, we propose a novel forecasting model based on High-Order Fuzzy-Fluctuation-Trends and BP Neural NetworkMachine Learning. In order to compare the forecasting results with other researchers' work, the authentic TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index) is employed to illustrate the forecasting process. The data from January 1999 to October 1999 are used as training time series and the data from November 1999 to December 1999 are used as testing dataset. The basic steps of the proposed model are shown (Fig 2).
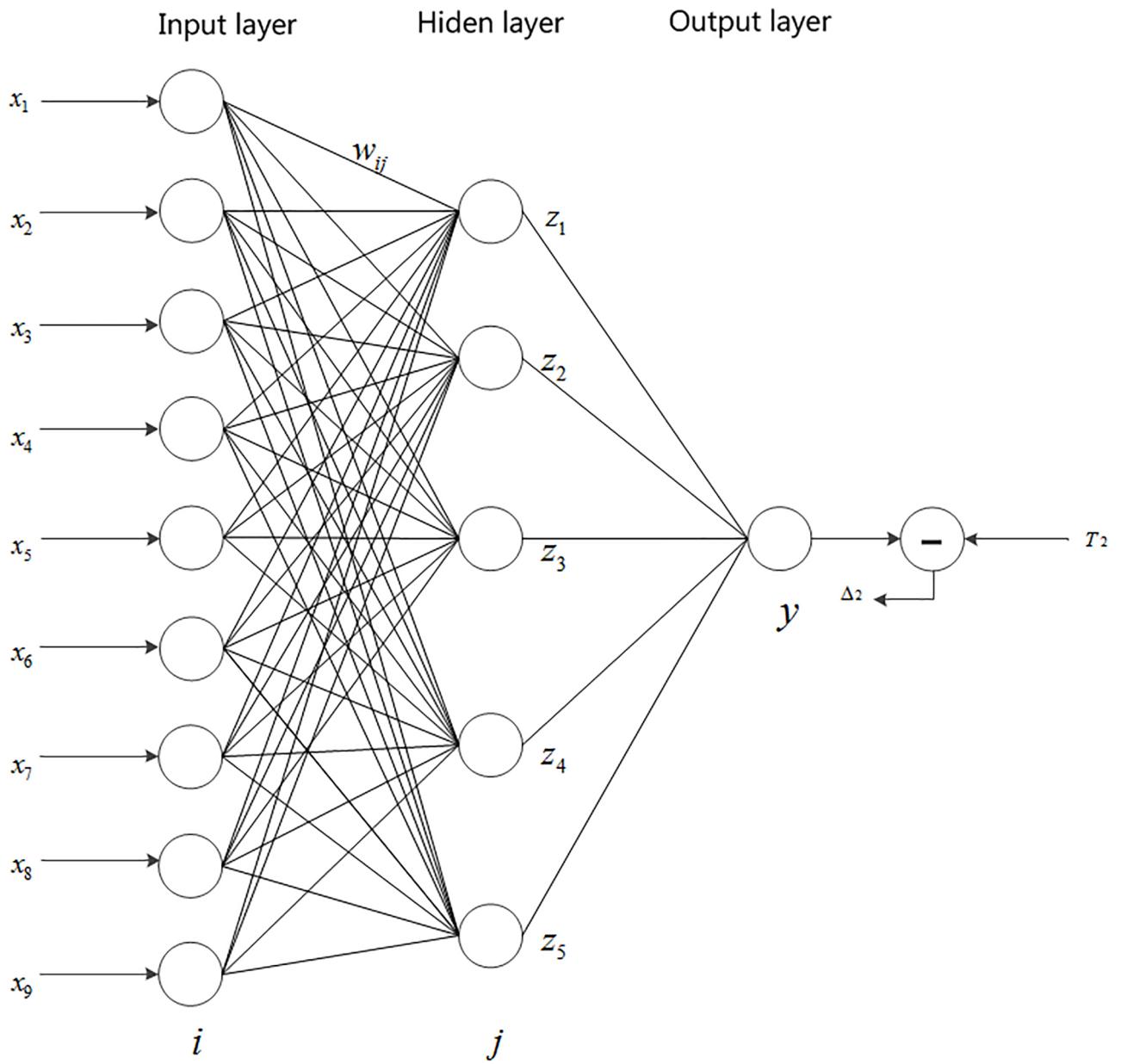
**Fig 1. BP neural network structure.**

**Step 1.** Construct FFTS for historical training data

For each element F(t)(t = 1, 2, . . ., T) in the historical training time series, its fluctuation trend is determined by G(t) = F(t) − F(t − 1), (t = 2, 3, . . ., T). According to the range and orientation of the fluctuations, G(t)(t = 2, 3, . . ., T) can be fuzzified into a linguistic set {down, equal, up}. Let *len* be the whole mean of all elements in the fluctuation time series G(t)(t = 2, 3, . . ., T), define $u_1 = [-\infty, -len/2)$, $u_2 = [-len/2, len/2)$, $u_3 = [len/2, +\infty)]$, then G(t)(t = 2, 3, . . ., T) can be fuzzified into a fuzzy-fluctuation time series S(t)(t = 2, 3, . . ., T).

**Step 2.** Establish nth-order FFLRs for the forecasting model

According to Eq (2), each $S(t)(t \geq n + 2)$ can be represented by its previous n days' fuzzy-fluctuation number. Therefore, the total of FFLRs for historical training data is $p_n = T - n - 1$.
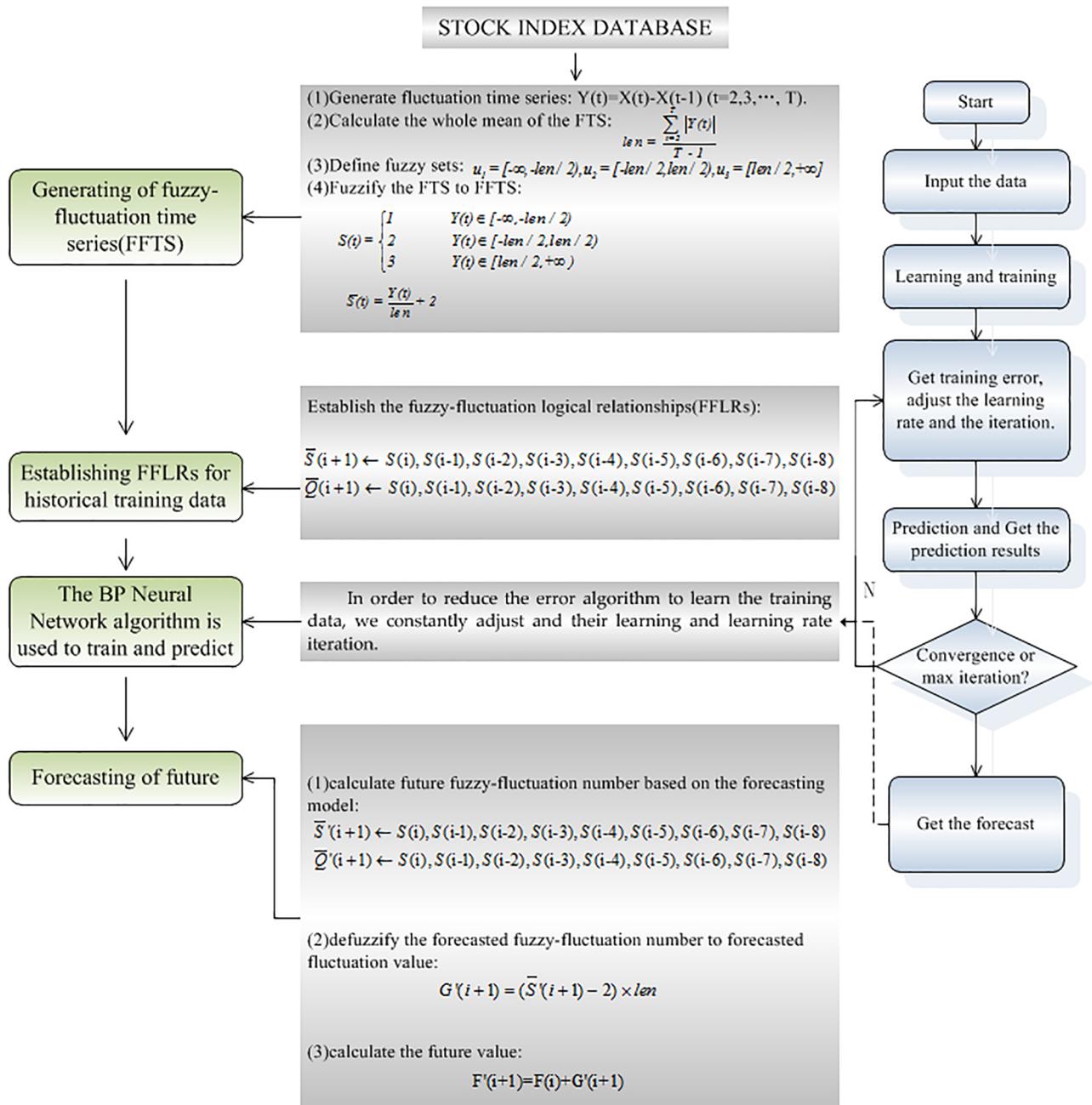
**Fig 2. Flowchart of our proposed forecasting model.**

**Step 3.** Determine the parameters for the forecasting model based on BP Neural Network-Machine Learning algorithm

In this paper, the BP Neural Network method is employed to learnthe fuzzy-fluctuation logical relationship.

$$G'(i + 1) = (\bar{S}'(i + 1) - 2) \times len \qquad (6)$$

$$\mathrm{F}'(i+1) = \mathrm{F}(i) + \mathrm{G}'(i+1) \qquad (7)$$

**Step 4.** Forecast test time series

For each data in the test time series, its future number can be forecasted according to Eq (7), based on the result of the output of the BP Neural NetworkMachine Learning, its n-order fuzzy-fluctuation trends.

## Empirical analysis

### Forecasting TAIEX

Since lotsofstudiesuseTAIEX1999as an example to illustrate their proposed forecasting methods [27–34]. We also use TAIEX1999 to illustrate the proposed method, and then we compared the accuracy with their models.

**Step 1.** Calculate the fluctuation of each element of the history training dataset. Then, the fluctuation trends will be fuzzified into FFTS by the whole mean of the fluctuation numbers of the training dataset. For example, the whole mean of the historical dataset of TAIEX1999 from January to October is 85. That is to say, $len = 85$. For $F(1) = 6152.43$ and $F(2) = 6199.91$, $G(2) = 47.48$, $S(2) = 3$. In this way, the historical training dataset can be represented by a fuzzified fluctuation dataset as shown in S1 Table.

**Step 2.** Based on the FFTS from 5January 1999 to 30October shown in S1 Table, the nth-order FFLRs for the forecasting model are established as shown in S2 Table. The subscript $I$ is used to represent element $l_i$ in the FFLRs for convenience.

For example, suppose n = 6, the 9th-order historical fuzzy-fluctuation trends2,3,1,1,1,2,2,3,3 on 18January 1999 $\bar{S}(11) = 0.7524$, then according to Eq (2), the Mapping relationships can be further expressed as:

$$2, 3, 1, 1, 1, 2, 2, 3, 3 \rightarrow 0.7524$$

Since parameter $\bar{Q}(11) = \bar{S}(11)/15 = 0.05016$, then according to Eqs (3) and (4), the Mapping relationships can also be further expressed as:

$$2, 3, 1, 1, 1, 2, 2, 3, 3 \rightarrow 0.05016$$

**Step 3.** The detailed BP Neural Network Machine Learning processes are shown in Fig 2.

In order to reduce the error algorithm to learn the training data, we constantly adjust and their learning and learning rate iteration, finally determined the iteration times to 8000 times, learning efficiency is set to 0.00008, the momentum factor is set to 0.003.

**Step 4.** Usethe *FFLR* obtained from historical training data to forecast the test dataset from 1 November 1999 to 30 December.

Firstly, the 9th-order historical fuzzy-fluctuation trends 3,2,2,2,2,3,1,2,2 on 1 November 1999 can be forecasted by the result 0.14506. Therefore, the forecasted fuzzy-fluctuation number is:

$$\bar{S}'(i+1) = \bar{Q}'(i+1) \times M = 0.14506 \times 15 = 2.1759$$

The forecasted fluctuation from current value to next value can be obtained by defuzzifying the fluctuation fuzzy number:

$$G'(i+1) = (\bar{S}'(i+1) - 2) \times len = (2.1759 - 2) \times 85 = 14.96$$

**Table 1. Forecasting results from 1 November1999 to 30 December 1999.**

| Date (MM/DD/YYYY) | Actual | Forecast | (Forecast–Actual)$^2$ | Date (MM/DD/YYYY) | Actual | Forecast | (Forecast–Actual)$^2$ |
|---|---|---|---|---|---|---|---|
| 11/1/1999 | 7,814.89 | 7869.81 | 3015.71 | 12/1/1999 | 7,766.20 | 7658.73 | 11550.01 |
| 11/2/1999 | 7,721.59 | 7767.25 | 2085.04 | 12/2/1999 | 7,806.26 | 7797.15 | 83.04 |
| 11/3/1999 | 7,580.09 | 7737.17 | 24674.73 | 12/3/1999 | 7,933.17 | 7926.84 | 40.01 |
| 11/4/1999 | 7,469.23 | 7505.08 | 1285.54 | 12/4/1999 | 7,964.49 | 8041.65 | 5952.94 |
| 11/5/1999 | 7,488.26 | 7405.53 | 6844.71 | 12/6/1999 | 7,894.46 | 8061.73 | 27978.96 |
| 11/6/1999 | 7,376.56 | 7405.23 | 821.78 | 12/7/1999 | 7,827.05 | 7907.67 | 6499.29 |
| 11/8/1999 | 7,401.49 | 7400.98 | 0.26 | 12/8/1999 | 7,811.02 | 7761.22 | 2480.17 |
| 11/9/1999 | 7,362.69 | 7464.39 | 10343.58 | 12/9/1999 | 7,738.84 | 7719.02 | 393.01 |
| 11/10/1999 | 7,401.81 | 7471.79 | 4896.83 | 12/10/1999 | 7,733.77 | 7750.31 | 273.52 |
| 11/11/1999 | 7,532.22 | 7391.31 | 19854.75 | 12/13/1999 | 7,883.61 | 7843.78 | 1586.65 |
| 11/15/1999 | 7,545.03 | 7581.21 | 1309.10 | 12/14/1999 | 7,850.14 | 7919.10 | 4755.54 |
| 11/16/1999 | 7,606.20 | 7535.24 | 5034.73 | 12/15/1999 | 7,859.89 | 7744.21 | 13382.87 |
| 11/17/1999 | 7,645.78 | 7583.48 | 3880.77 | 12/16/1999 | 7,739.76 | 7832.19 | 8542.56 |
| 11/18/1999 | 7,718.06 | 7665.95 | 2715.32 | 12/17/1999 | 7,723.22 | 7698.15 | 628.71 |
| 11/19/1999 | 7,770.81 | 7711.19 | 3554.60 | 12/18/1999 | 7,797.87 | 7639.44 | 25101.33 |
| 11/20/1999 | 7,900.34 | 7833.44 | 4475.04 | 12/20/1999 | 7,782.94 | 7801.32 | 337.75 |
| 11/22/1999 | 8,052.31 | 7924.00 | 16463.38 | 12/21/1999 | 7,934.26 | 7796.21 | 19056.81 |
| 11/23/1999 | 8,046.19 | 8083.08 | 1360.55 | 12/22/1999 | 8,002.76 | 7932.06 | 4999.04 |
| 11/24/1999 | 7,921.85 | 8037.94 | 13476.54 | 12/23/1999 | 8,083.49 | 7998.60 | 7205.82 |
| 11/25/1999 | 7,904.53 | 7935.50 | 992.02 | 12/24/1999 | 8,219.45 | 8099.29 | 14439.44 |
| 11/26/1999 | 7,595.44 | 7833.93 | 56879.02 | 12/27/1999 | 8,415.07 | 8252.86 | 26313.12 |
| 11/29/1999 | 7,823.90 | 7632.06 | 36802.23 | 12/28/1999 | 8,448.84 | 8452.34 | 12.24 |
| 11/30/1999 | 7,720.87 | 7858.64 | 18981.79 | Root Mean Square Error(RMSE) | | | 96.77 |

https://doi.org/10.1371/journal.pone.0192366.t001

Finally, the forecasted value can be obtained by current value and the fluctuation value:

$$\mathrm{F}'(i+1) = \mathrm{F}(i) + \mathrm{G}'(i+1) = 7854.85 + 14.96 = 7869.81$$

The other forecasting results are shown (Table 1 and Fig 3).

This paper compares the difference between the predicted value and the actual value, and the objective is to evaluate the prediction performance. In the comparison of time series model, the broad indexes are the mean squared error (MSE), root of the mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), etc. These indicators are defined by Eqs (8)–(11):

$$MSE = \frac{\sum_{t=1}^{n} (forecast(t) - actual(t))^2}{n} \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (forecast(t) - actual(t))^2}{n}} \tag{9}$$

$$MAE = \frac{\sum_{t=1}^{n} |(forecast(t) - actual(t))|}{n} \tag{10}$$
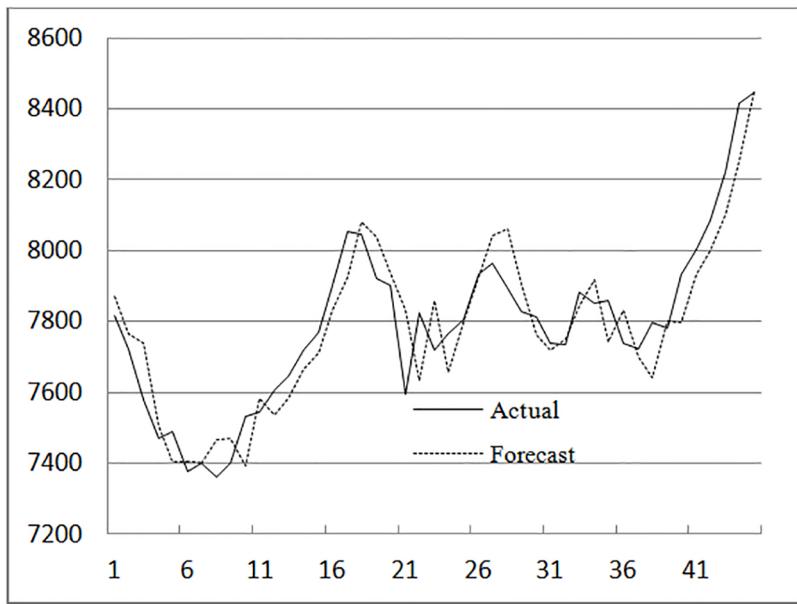
**Fig 3. Forecasting results from 1 November1999 to 30 December 1999.** Based on the method presented in this paper, the data of 1999 is predicted.

$$MPE = \frac{\sum_{t=1}^{n} |(forecast(t) - actual(t))|/actual(t)}{n} \tag{11}$$

where $n$ denotes the number of values forecasted, *forecast(t)* and *actual(t)* denote the predicted value and actual value at time $t$, respectively. With respect to the proposed method for the 9th-order, the MSE, RMSE, MAE, and MPE are 9363.57, 96.76, 79.54, and 0.01, respectively.

Let the order number $n$ vary from 2 to 10, the RMSEs for different nth-order forecasting models are listed in Table 2. The item "Average" refers to the RMSE for the average forecasting results of these different nth-order(n = 2,3,. . .,10) models.

In practical forecasting, the average of results for different nth-order (n = 2,3,. . .,9) forecasting models is adopted to avoid the uncertainty. The proposed method is employed to forecast the TAIEX from 1997 to 2005. The forecasting results and errors are shown (Fig 4 and Table 3).

Table 4 shows the comparison results for RMSEs of different methods for predicting TAIEX1999. As can be seen from this table, the performance of the proposed method is acceptable. The best advantage of this method is that you do not need to determine the target function, nor do you need to determine the mapping rules. Learn from the algorithm and find the rules. Although some other methods of RMSEs are superior to the methods presented in this article, they usually need to determine complex rules to predict the results. In practice, however, it is often difficult to establish proper rules. The method presented in this paper is very simple and easy to implement computer program.

**Table 2. Comparison of forecasting errors for different nth-orders.**

| $n$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|
| RMSE | 99.19 | 98.25 | 95.50 | 98.17 | 94.71 | 98.95 | 99.57 | 96.77 | 96.88 | 97.55 |
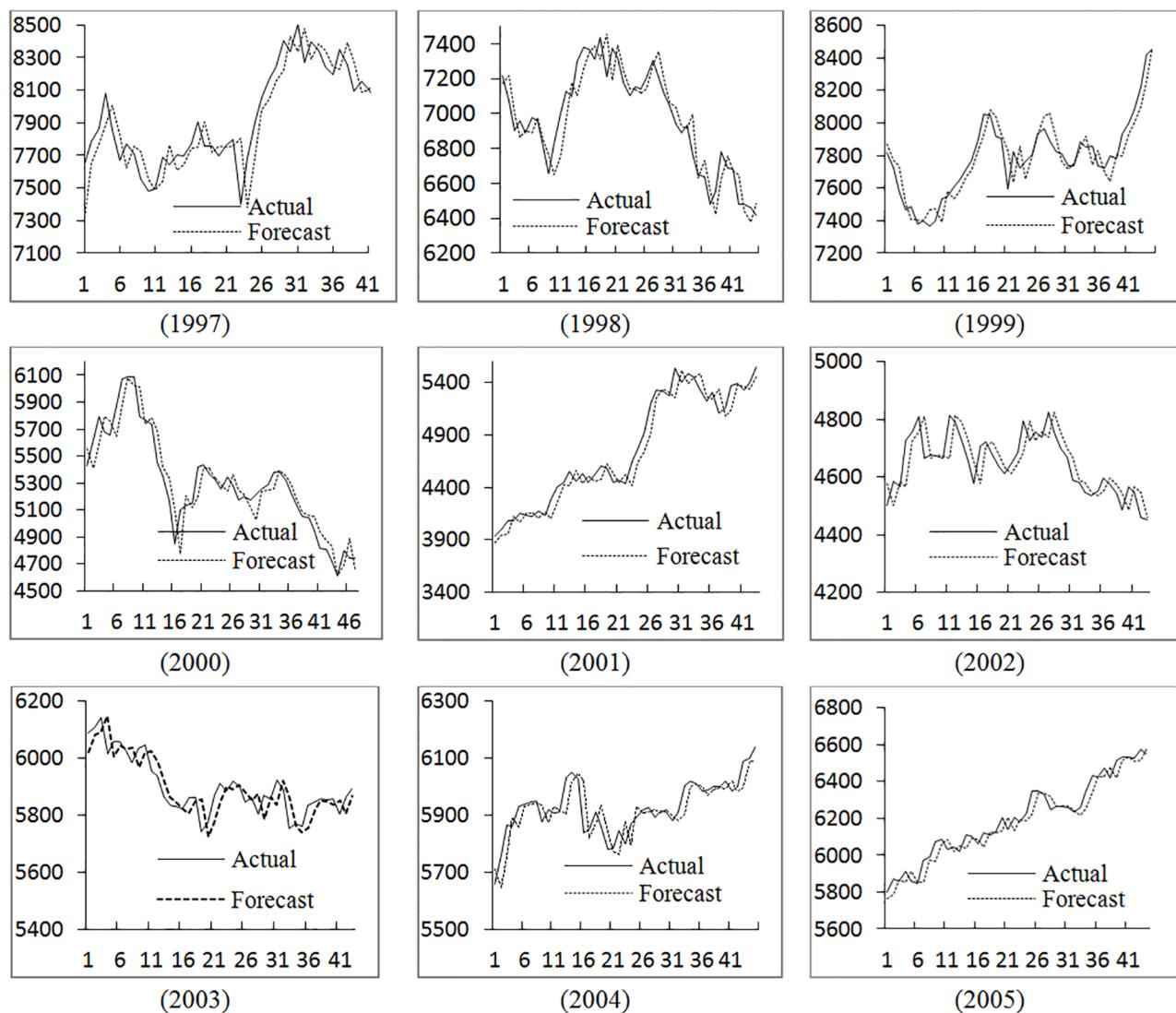
**Fig 4. The stock market fluctuation for TAIEX test dataset (1997–2005).** Based on the method presented in this paper, the results of Taiwan stock market data from 1999 to 2005 are predicted.

## Friedman test

In order to verify the validity of the model proposed in this paper, we applied the Friedman test for the significance test based on JanezDemˇsar's [35] study. The Friedman test was a non-parametric statistical test proposed by Milton Friedman [36–39]. It sequenced the algorithm of each data set, the best algorithm got the rank 1, and the second best was 2. . ., as shown in Table 6. Let $r_i^j$ be the rank of the $j$-th of $k$ algorithms on the $i$-th of N data sets. The Friedman testwill compares the average ranks of algorithms, $R_j = \frac{1}{N}\sum_i r_i^j$. the Friedman statisticis distributed according to $\chi_F^2$ with $k-1$ degrees of freedom, when $N$ and $k$ are big enough.

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{12}$$

**Table 3. RMSEs of forecast errors for TAIEX 1997 to 2005.**

| Year | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|------|------|------|------|------|------|------|------|------|------|
| RMSE | 142.99 | 112.51 | 96.77 | 126.85 | 120.12 | 66.39 | 54.87 | 58.10 | 54.7 |

https://doi.org/10.1371/journal.pone.0192366.t003

Iman and Davenport [40]thought that Friedman's $\chi_F^2$ is undesirably conservative and proposed a better statistic. Which is distributed according to the F-distribution with $k - 1$ and $(k - 1)(N - 1)$ degrees of freedom.

$$F_F = \frac{(N - 1)\chi_F^2}{N(k - 1) - \chi_F^2} \tag{13}$$

Nemenyitest [41] is used when compared between all classifiers. The performance of the two classifiers is very different if the corresponding average level is at least different.

$$CD = q_\alpha \sqrt{\frac{k(k + 1)}{6N}} \tag{14}$$

This article will rank the data sets from 1999 to 2004 and sort the different methods based on the RMSE error, as shown in Table 5.

**Table 4. A comparison of RMSEs for different methods for forecasting the TAIEX1999.**

| | Methods | RMSE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | I | II | III | IV | V | VI | VII | VIII | IX |
| | | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
| A | Chen and Chang's Method[27] | N | N | **123.64** | 131.1 | 115.08 | 73.06 | 66.36 | 60.48 | N |
| B | Chen and Chen's Method[28] | N | N | **119.32** | 129.87 | 123.12 | 71.01 | 65.14 | 61.94 | N |
| C | Chen et al.'s Method[39] | N | N | **102.34** | 131.25 | 113.62 | 65.77 | 52.23 | 56.16 | N |
| D | Cheng et al.'s method[30] | N | N | **100.74** | 125.62 | 113.04 | 62.94 | 51.46 | 54.24 | N |
| E | Chen and Kao's method[31] | N | N | **87.63** | 125.34 | 114.57 | 76.86 | 54.29 | 58.17 | N |
| F | Guan S's Method[32] | N | N | **101.11** | 127.47 | 114.19 | 61.92 | 53.05 | 53.07 | N |
| G | Jia's method[33] | 143.60 | 115.34 | **99.12** | 125.70 | 115.91 | 70.43 | 54.26 | 57.24 | 54.68 |
| H | Guan H J's method[34] | 141.89 | 119.85 | **99.03** | 128.62 | 125.64 | 66.29 | 53.2 | 56.11 | 55.83 |
| I | The proposed | 142.99 | 112.51 | **96.77** | 126.85 | 120.12 | 66.39 | 54.87 | 58.10 | 54.7 |

https://doi.org/10.1371/journal.pone.0192366.t004

**Table 5. The sorting of different prediction methods based on RMSE for forecasting the TAIEX1999.**

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| III | 123.64(9) | 119.32(8) | 102.34(7) | 100.74(5) | 87.63(1) | 101.11(6) | 99.12(4) | 99.03(2) | 96.77(3) |
| IV | 131.1(8) | 129.87(7) | 131.25(9) | 125.62(2) | 125.34(1) | 127.47(5) | 125.7(3) | 128.62(6) | 126.85(4) |
| V | 115.08(5) | 123.12(8) | 113.62(2) | 113.04(1) | 114.57(4) | 114.19(3) | 115.91(6) | 125.64(9) | 120.12(7) |
| VI | 73.06(8) | 71.01(7) | 65.77(3) | 62.94(2) | 76.86(9) | 61.92(1) | 70.43(6) | 66.29(4) | 66.39(5) |
| VII | 66.36(9) | 65.14(8) | 52.23(2) | 51.46(1) | 54.29(6) | 53.05(3) | 54.26(5) | 53.2(4) | 54.87(7) |
| VIII | 60.48(8) | 61.94(9) | 56.16(4) | 54.24(2) | 58.17(7) | 53.07(1) | 57.24(5) | 56.11(3) | 58.1(6) |
| average rank | 7.83 | 7.83 | 4.5 | 2.17 | 4.67 | 3.17 | 4.83 | 4.67 | 5.33 |

https://doi.org/10.1371/journal.pone.0192366.t005

**Table 6. RMSEs of forecast errors for SHSECI from 2007 to 2015.**

| | Year | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
| RMSE | 123.89 | 57.44 | 48.92 | 47.34 | 28.37 | 25.84 | 21.43 | 50.59 | 59.69 |

Using the data in Table 5, we can calculate:

$$\chi_F^2 = \frac{12 \times 6}{9 \times 10} \left[ 7.83^2 + 7.83^2 + 4.5^2 + 2.17^2 + 4.67^2 + 3.17^2 + 4.83^2 + 4.67^2 + 5.33^2 - \frac{9 \times 10^2}{4} \right]$$
$$= 22.38$$

$$F_F = \frac{5 \times 22.38}{6 \times 8 - 22.38}$$
$$= 4.37$$

With 9 methods and 6 data sets, $F_F$ is distributed according to the F distribution with $9 - 1 = 8$ and $(9 - 1) \times (6 - 1) = 40$ degrees of freedom. The critical value of $F(8,40)$ for $\alpha = 0.05$ is 2.18, so we reject the null-hypothesis. Next, we used the Nemenyi test for pairwise comparisons. The critical value of CD for $\alpha = 0.05$ is 3.102.

$$CD = 3.102 \times \sqrt{\frac{9 \times (9 + 1)}{6 \times 6}}$$
$$= 4.90$$

According to the average order value in the table, the difference between method A and method B exceeds the critical value, and the others are not exceeded. Therefore, there are significant differences between methods A,B and D(7.83–2.17>4.9), and no significant differences among other algorithms. In general, there is no significant difference between the proposed method and the latest methods in predicting the effect of error and predictive value.

### Forecasting Shanghai Stock Exchange Composite Index

TheSHSECI is China's most typical stock market index. In further research, we apply the method to SHSECI's stock market forecast from 2007 to 2015. We use the real data set of SHSECI's closing price from January to October as training data, and data sets from November to December are used as test data. The RMSEs for the prediction error is shown in Table 6.

From Table 6, We can see that this method can successfully predict the SHSECI stock market.

## Conclusions

This paper presents a prediction model based on high order fuzzy fluctuation and BP neural network. This method is based on the high order fuzzy logic relation of time series and then uses the self-learning of BPNN to automatically find the optimal prediction rules to predict the fluctuation trend. The greatest advantage of this approach is that the fuzzy theory, stock market fluctuation model and neural network algorithm are combined to construct a new model, which solves the problem of overfitting and over-fuzzy existing models. Experiments show that the parameters generated from the training data set can also be used for future data sets. To compare the performance of other methods, we take TAIEX1999 as an example. We also predicted the validity and universality of TAIEX 1997-2005and Shanghai Stock Exchange

Composite Index (SHSECI) from 2007 to 2015. The model presented in this paper has a significant advantage in universality, flexibility and comprehensibility. However, because of the influence of changing external factors, the accuracy of the forecasting results is just acceptable comparing with other models. In further research, we will take more consideration of the influence of external factors to improve the accuracy. Moreover, we will consider other factors that may affect the volatility of the stock market, such as trading volume, starting value, final value, etc. We will also consider the impact of other stock markets, such as the Dow Jones, the NASDAQ, and so on.

## Supporting information

**S1 Table. Historical training data and fuzzified fluctuation data of TAIEX 1999.**
(DOCX)

**S2 Table. The FFLRs for historical training data of TAIEX 1999.**
(DOCX)

## Acknowledgments

The authors also would like to express appreciation to the anonymous reviewers and Editors for their very helpful comments that improved the paper.

## Author Contributions

**Data curation:** Jie He.

**Writing – original draft:** Zongli Dai.

**Writing – review & editing:** Hongjun Guan, Aiwu Zhao.

## References

1. Chen YS, Cheng CH, Chiu CL, Huang ST. A study of ANFIS-based multi-factor time series models for forecasting stock index. Applied Intelligence, 2016; 45(2):1–16.

2. Box GE, Jenkins GM. Time series analysis: forecasting and control rev. Oakland, California, Holden-Day.1976; 303–303.

3. Engle RF. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica.1982; 50(4):987–1007.

4. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics.1986; 31(3):307–327.

5. Wang LX. Modeling Stock Price Dynamics With Fuzzy Opinion Networks. IEEE Transactions on Fuzzy Systems.2017; 25(2):277–301.

6. Kisi O, Shiri J, Nikoofar B. Forecasting daily lake levels using artificial intelligence approaches. Computers & Geosciences. 2012; 41(2):169–180.

7. Badrzadeh H, Sarukkalige R, Jayawardena AW. Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. Journal of Hydrology. 2013; 507 (507):75–85.

8. Daut MAM, Hassan MY, Abdullah H, Rahman AH, Abdullah MP, Hussin F. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renewable & Sustainable Energy Reviews. 2017; 70:1108–1118.

9. Mishra D, Goyal P, Upadhyay A. Artificial intelligence based approach to forecast PM 2.5, during haze episodes: A case study of Delhi, India. Atmospheric Environment. 2015; 102:239–248.

10. Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable & Sustainable Energy Reviews. 2015; 50:1352–1372.

11. Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks:: The state of the art. International Journal of Forecasting.1998; 14(1):35–62.

12. Rathnayaka RMKT, Seneviratna DMKN, Wei J, Arumawadu H. A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models[C]// International Conference on Behavioral, Economic and Socio-Cultural Computing. IEEE, 2015:54–60.

13. Dumitru CD,Gligor A. Daily Average Wind Energy Forecasting Using Artificial Neural Networks. Procedia Engineering. 2017; 181: 829–836.

14. Abbot J, Marohasy J. Skilful rainfall forecasts from artificial neural networks with long duration series and single-month optimization. Atmospheric Research. 2017; 197:289–299.

15. Bing Y, Hao JK, Zhang SC. Stock Market Prediction Using Artificial Neural Networks. Information Technology for Manufacturing Systems III. 2012; 6–7:1055–1060.

16. Tu CC, Juang F. Recurrent type-2 fuzzy neural network using Haar wavelet energy and entropy features for speech detection in noisy environments. Expert Systems with Applications. 2012; 39(3):2479–2488.

17. Jiang Y, Deng Z, Chung FL, Wang S. Realizing Two-View TSK Fuzzy Classification System by Using Collaborative Learning. IEEE Transactions on Systems Man & Cybernetics Systems.2017; 47(1):145–160.

18. Deng Z, Cao L, Jiang Y, Wang S. Minimax Probability TSK Fuzzy System Classifier: A More Transparent and Highly Interpretable Classification Model. IEEE Transactions on Fuzzy Systems. 2015; 23 (4):813–826.

19. Song Q, Chissom BS. Forecasting enrollments with fuzzy time series—part I. Elsevier North-Holland; 1993.

20. Song Q, Chissom BS. Fuzzy time series and its models. Fuzzy Sets Syst; 1993.

21. Song Q,Chissom BS. Forecasting enrollments with fuzzy time series—Part II. Fuzzy Sets Syst;1994.

22. Aladag CH, Basaran MA, Egrioglu E, Yolcu U, Uslu VR. Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations. Expert Systems with Applications. 2009; 36(3):4228–4231.

23. Chen SM, Jian WS. Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups, similarity measures and PSO techniques. Elsevier Science Inc; 2017.

24. Sadaei HJ, Guimares FG, Silva CJD,Lee MH, Eslami T. Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. International Journal of Approximate Reasoning. 2017; 83:196–217.

25. Wang JZ, Wang JJ, Zhang ZG, Guo SP. Forecasting stock indices with back propagation neural network. Expert Systems with Applications. 2011; 38(11):14346–14355.

26. Herrera F, Herrera-Viedma E, Verdegay JL. A model of consensus in group decision making under linguistic assessment. Fuzzy Sets & Systems. 1996; 78(1):73–87.

27. Chen SM, Chang YC. Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. Information Sciences. 2010; 180(24):4772–4783.

28. Chen SM, Chen CD. TAIEX Forecasting Based on Fuzzy Time Series and Fuzzy Variation Groups. IEEE Transactions on Fuzzy Systems. 2011; 19(1):1–12.

29. Chen SM, Manalu GM, Pan JS, Liu HC. Fuzzy Forecasting Based on Two-Factors Second-Order Fuzzy-Trend Logical Relationship Groups and Particle Swarm Optimization Techniques. IEEE Transactions on Cybernetics. 2013; 43(3):1102–1117. https://doi.org/10.1109/TSMCB.2012.2223815 PMID: 23193240

30. Cheng SH, Chen SM, Jian WS. Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures. Information Sciences. 2016; 327:272–287.

31. Chen SM, Kao PY. TAIEX forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines. Elsevier Science Inc; 2013.

32. Guan S, Zhao AW. A Two-Factor Autoregressive Moving Average Model Based on Fuzzy Fluctuation Logical Relationships. Symmetry. 2017; 9(10):207.

33. Jia J, Zhao AW, Guan S. Forecasting Based on High-Order Fuzzy-Fluctuation Trends and Particle Swarm Optimization Machine Learning. Symmetry. 2017; 9(7):124.

34. Guan H, Guan S, Zhao AW. Forecasting Model Based on Neutrosophic Logical Relationship and Jaccard Similarity. Symmetry. 2017; 9(9):191.

35. Janez Demšar. Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research. 2006; 7(1):1–30.

36. Ar J. Statistical Comparisons of Classifiers over Multiple Data Sets[J]. Journal of Machine Learning Research, 2006, 7(1):1–30.

**37.** Friedman Milton. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association. American Statistical Association. 1937; 32 (200): 675–701

**38.** Friedman Milton. A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association. American Statistical Association. 1939; 34 (205):109.

**39.** Friedman Milton. A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics. 1940; 11 (1): 86–92.

**40.** Iman RL, Davenport JM. Approximations of the critical region of the fbietkan statistic. Communications in Statistics. 1979; 9(6):571–595.

**41.** Nemenyi P. Distribution-free multiple comparisons. PhD thesis, Princeton University, 1963.