Scientia Iranica D (2014) 21(3), 815–825

Sharif University of Technology
Scientia Iranica
Transactions D: Computer Science & Engineering and Electrical Engineering
www.scientiairanica.com

A strategy for forecasting option prices using fuzzy time series and least square support vector regression with a bootstrap model

C.-P. Lee[a,*], W.-C. Lin[b] and C.-C. Yang[c]

a. *Department of Information Management, Da-Yeh University, No.168, University Rd., Dacun, Changhua 515, Taiwan.*
b. *Department of Business Administration, National Taipei College of Business, No. 321, Jinan Road, Section 1, Taipei 100, Taiwan.*
c. *Department of Information Management, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei 106, Taiwan.*

Received 27 June 2012; received in revised form 6 November 2012; accepted 23 November 2013

KEYWORDS
Option price;
Fuzzy time series;
Least square support vector regression;
Bootstrap;
Hybrid model.

Abstract. Recently, the strategy for forecasting option price has become a popular financial topic because options are important tools on risk management in financial investments. The well-known Black-Scholes model (B-S model) is widely used for option pricing. In B-S model, the normal distribution assumption is important. However, the financial data in real life may not follow the normal distribution assumption. For this reason, this paper proposes a novel hybrid model, which is a nonlinear prediction model without normal distribution assumptions to forecast the option prices. The proposed model is composed of a Fuzzy Time Series (FTS) model, a Least Square Support Vector Regression (LSSVR), and a bootstrap method. In the proposed model, FTS model is firstly used to fuzzify dataset and to build historical database. Subsequently, the proposed method uses the remainder contractual time to search similar historical trends as training samples. Finally, we use the bootstrap method on LSSVR to enhance the prediction accuracy. The experiment results show that the proposed model outperforms traditional time series models and several hybrid models in terms of the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the correlation coefficient (r) of actual and forecasted option price.

1. Introduction

Many investors enhance their net worth by investing. There are many investment targets in the financial market, such as stocks, bonds, options, futures, funds, and etc., and due to the fact that an option is an important tool in risk management in financial investments [1-3], it has become a popular tool in reducing investment risk [4]. For example, a producer can buy a put option to prevent the loss of profit from a possible decrease in product prices. Similarly, a customer can buy a call option at an expected future price [2,3] of the desired product. Hence, an option is like an insurance policy, where people pay a premium for the option. The premium of option transactions is also known as the price of an option that affects investors regarding whether to purchase the option or not.

As described above, the strategy to forecast option price has become an important financial topic. Traditional strategies in forecasting option price use

*. *Corresponding author. Tel: +886-4-8511888 Ext.: 3132; Fax: +886-4-8511500*
E-mail addresses: chien.pang@gmail.com (C.-P. Lee); wencheng@webmail.ntcb.edu.tw (W.-C. Lin); D10109101@mail.ntust.edu.tw (C.-C. Yang)

statistical or linear models, and the Black-Scholes model (B-S model) [5], first introduced in 1973, is the most well-known model in this regard. In the B-S model, the normal distribution assumption is important. However, the financial data in real life may not follow the normal distribution assumption. Moreover, the price of an option is determined by many factors, such as current stock price, option strike price, expiration time, volatility of stock price and risk-free interest rates [5,6]. Option price forecasting is, therefore, difficult, and in recent years, many researchers have proposed novel prediction models for option price forecasting [7].

The main factors in option pricing are current option price and current stock price. Hence, this paper uses a 2-factor, 3-order Fuzzy Time Series (FTS) model to forecast the option price, as it can search the trends of option by using the two main factors of option pricing. It is, however, still not easy to find a matching trend to forecast option price with the 2-factor 3-order FTS model. Moreover, according to the literature, the expiration time of options would also affect the volition of investors on option price [8-10]. Hence, we use the remaining contractual time to search similar trends, and then, use the Least Square Support Vector Regression (LSSVR) to assist the 2-factor 3-order FTS model on option price forecasting. It is possible that the sample size of the training samples might be small, so we use a bootstrap method on the LSSVR model to enhance prediction accuracy. Accordingly, the proposed model is composed of a 2-factor 3-order FTS model, the LSSVR and a bootstrap method. In this paper, the proposed model is termed; FLSSVR with a bootstrap model.

The remainder of this paper is organized as follows. Section 2 reviews related work, including the definition of time series data and option price forecasting reviews. Section 3 introduces the FTS model, LSSVR, bootstrap method, and the procedure of the FLSSVR with a bootstrap model. Section 4 gives an example of forecasting option price using the proposed model. Section 5 compares the performance of the proposed model with other existing models. Section 6 gives the conclusions of this paper.

2. The related works

2.1. Time series data

A time series data is a sequence of data points such as stock index or stock price. Forecasting time series data is to predict future values based on previously observed time series values. The previous observed time series values are termed a "situation", in this paper. To simplify, this paper defines a situation caused by 3 previously observed time series values, termed a 3-day in this paper. The situation of 3-day time series data can be classified into 9 types of situation, which are shown in Figure 1. Consequently, the 3-day situation is represented as $(\text{day}_1, \text{day}_2, \text{day}_3)$. Furthermore, when forecasting future values on day t, it is represented as $(\text{day}_{t-3}, \text{day}_{t-2}, \text{day}_{t-1}) \rightarrow \text{day}_t$, which is termed a "trend" in this paper. According to the literature [11-14], the main idea of the time series data analysis is to search trends with similar situations from historical data and then to forecast future values.

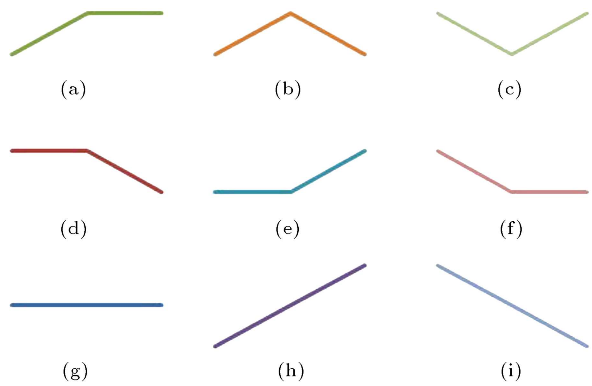

Figure 1. 9 types of the situations of time series data with a 3-day.

2.2. Option prices forecasting reviews

Normally, statistical time series models are widely used in resolving financial problems. However, statistical time series models are often limited by their assumptions, such as the normal distribution assumption. Hence, many researchers proposed novel hybrid models or modify the statistical time series models for option price forecasting. Most existing models are hybrid, based on the Artificial Neural Network (ANN). For example, Tseng et al. [15] proposed a hybrid model to forecast the option price. In Tseng et al.'s model, a grey-exponential generalized autoregressive conditional heteroscedasticity (Grey-EGARCH) is developed to decrease the stochastic and nonlinearity of error term sequence, and then, further, to elevate the predictability of the option-pricing model. Subsequently, the Grey-EGARCH is integrated into ANN to provide functional flexibility in order to capture nonlinearities in financial data. Wang [16] proposed another hybrid model composed of an ANN and a grey Glosten-Jagannathan-Runkle GARCH (Grey-GJR-GARCH). In Wang's model, Grey-GJR-GARCH is developed to reduce the stochastic and nonlinearity of the error term sequence, and then to improve the prediction ability of option-pricing. Subsequently, Grey-GJR-GARCH is integrated into ANN to capture the nonlinearities in financial data. Liang et al. [17] proposed a simple method, which was based on ANN. they first use four different linear models to predict option price. Subsequently, the four predicted option prices are fed into ANN to get the final predicted option price. Compared to traditional time series models, such as

GARCH and GJR-GARCH models, the ANN hybrid models presented a significantly better performance under several performance measures.

3. Methods

In Section 3.1, we briefly introduce the definition of the FTS model. Then, due to the fact that LSSVR and a bootstrap method play important roles in FLSSVR with a bootstrap model, we also briefly review the definition of LSSVR and the bootstrap method in Sections 3.2 and 3.3, respectively. Finally, we introduce the procedure of the FLSSVR with a bootstrap model in Section 3.4.

3.1. Fuzzy time series

The fuzzy time series is based on fuzzy logic. It is a model to forecast problems [18]. The fuzzy time series model was first applied by Song and Chissom [19-21] to forecast enrollment at the University of Alabama. Recently, the fuzzy time series model has been widely used in many financial areas [13]. According to the literature [2,3,12,20-22], the following definitions are given to a fuzzy time series model.

Defenition 1. A fuzzy set, A, defined in the universe of discourse, u, can be represented as follows:

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \cdots + \frac{f_A(u_n)}{u_n}, \tag{1}$$

where f_A is the membership function of fuzzy set A; $f_A(u_i)$ denotes the degree of membership of u_i belonging to the fuzzy set A.

Let $Y(t)$ $(t = \cdots, 0, 1, 2, \cdots)$, a subset of R^1, be the universe of discourse in which fuzzy sets $f_i(t)$ $(i = 1, 2\cdots)$ are defined. If $F(t)$ is a collection of $f_i(t)$ $(i = 1, 2\cdots)$, $F(t)$ is called a fuzzy time series of $Y(t)$. Namely, $f_i(t)$ $(i = 1, 2\cdots)$ denotes the degree of membership of day t belonging to fuzzy set i, and $F(t)$ denotes the fuzzy time series of day t in this paper.

Defenition 2. If, for any $f_j(t) \in F(t)$, there exists an $f_i(t-1) \in F(t-1)$, such that there exists a fuzzy relation $R_{ij}(t, t-1)$ and $f_j(t) = f_i(t-1) \circ R_{ij}(t, t-l)$, where '$\circ$' is the max-min composition, $F(t)$ is said to be caused by $F(t-1)$ only, and it can be represented by $F(t-1) \to F(t)$. Namely, the corresponding fuzzy set of day t is caused by the corresponding fuzzy set of day $t-1$.

Defenition 3. If $F(t)$ is caused by $F(t-1)$, $F(t-2)$, and $F(t-3)$, $F(t)$ is called a 1-factor 3-order fuzzy time series, and it can be represented by $F(t-3)$, $F(t-2)$, $F(t-1) \to F(t)$.

Defenition 4. If $F(t)$ is caused by $(F_1(t-1), F_2(t-1))$, $(F_1(t-2), F_2(t-2))$, $(F_1(t-3), F_2(t-3))$, then $F(t)$ is called a 2-factor 3-order fuzzy time series, where $F_1(t)$ and $F_2(t)$ are the 1st and the 2nd factor fuzzy time series, respectively. $F(t)$ can be represented by $(F_1(t-3), F_2(t-3))$, $(F_1(t-2), F_2(t-2))$, $(F_1(t-1), F_2(t-1)) \to F(t)$. Let $F_1(t) = X_t$ and $F_2(t) = Y_t$, where X_t and Y_t are fuzzy variables whose values are possible fuzzy sets of the first factor and the second factor on day t, respectively. Then, a 2-factor 3-order Fuzzy Logic Relationship (FLR) [12] can be represented by (X_{t-3}, Y_{t-3}), (X_{t-2}, Y_{t-2}), $(X_{t-1}, Y_{t-1}) \to X_t$, where (X_{t-3}, Y_{t-3}), (X_{t-2}, Y_{t-2}) and (X_{t-1}, Y_{t-1}), are referred to as the Left-Hand Side (LHS) of the fuzzy logic relationship, and X_t is referred to as the Right Hand Side (RHS) of the fuzzy logic relationship.

3.2. Least square support vector regression

The Support Vector Machine (SVM) was introduced by Vapnik and his coworkers [23]. It is a popular and powerful technique for data classification [24]. SVM was extended to solve a nonlinear regression estimation problem in 1996. The extended SVM, which was proposed by Drucker et al. [25], is called SVR. To reduce the computational complexity of SVR, Least Squares Support Vector Regression (LSSVR) was proposed by Suykens et al. [26]. In LSSVR, function $f(x)$ can be solved by the following equations [27]:

$$\begin{aligned} &\text{Minimize:} \quad \frac{1}{2}\|w\|^2 + \frac{1}{2}\sum_{i=1}^{l} e_i^2. \\ &\text{Subject to:} \quad f(x) = w^T\varphi(x_i) + b + e_i, \\ &\qquad\qquad i = 1, 2, \cdots, l, \end{aligned} \tag{2}$$

where l denotes the number of data.

After resolving the above optimization problem, we can obtain the solution from the following equations [27,28]:

$$f(x) = \sum_{i=1}^{l} \alpha_i K(x, x_i) + b, \tag{3}$$

where $K(x, x_i)$ is the radial basis function shown in the following:

$$\begin{aligned} &K(x, x_i) = \exp\left\{-\frac{(x-x_i)^T(x-x_i)}{2\sigma^2}\right\}, \\ &\quad i = 1, 2, \cdots, l, \end{aligned} \tag{4}$$

where σ^2 denotes the width of the radial basis function.

3.3. Bootstrap method

In 1979, the bootstrap method was first proposed by Efron [29]. Bootstrap is a method to assign the accuracy of measures to estimate samples [30]. Generally, a bootstrap method is classified into the broader class of resampling methods. It can be implemented by generating a large number of resamples of the original dataset, each of which is obtained by random sampling with a replacement from the original dataset. Then, a particular statistic can be calculated from the collected values of the sampling distribution. Through simulations, it is found that the bootstrap method provides less biased statistics [31]. Hence, the bootstrap method can be used to enhance the measures of statistical accuracy.

3.4. FLSSVR with a bootstrap model

FLSSVR with a bootstrap model includes seven steps. The 1st to 5th steps use a 2-factor 3-order FTS model to search similar instances (similar FLRs) and to generate training samples from the historical database. The 6th step uses LSSVR to build a prediction model using the selected training samples. Finally, the bootstrap method is used to enhance the prediction accuracy in Step 7. However, when there are many orders at the LHS of a FLR, it is difficult to find matching FLR for prediction. For this reason, we use two conditions to search similar FLRs in Step 5. The flowchart of FLSSVR with a bootstrap model is shown in Figure 2. The detail procedures of FLSSVR with a bootstrap model are described in the following steps.

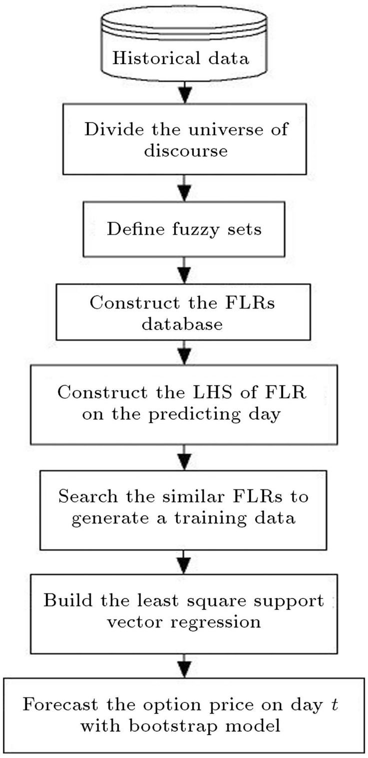

Figure 2. The flowchart of FLSSVR with a bootstrap model.

Step 1: Divide the universe of discourse. The universe of discourse of the first factor is defined as $U = [D_{\min} - D_1, D_{\max} + D_2]$, where $D_{\min}$ and $D_{\max}$ are the minimum and maximum of the first factor, respectively; D_1 and D_2 are two positive real numbers to divide the universe of discourse into n equal length intervals. The universe of discourse of the second factor is defined as $V = [V_{\min} - V_1, V_{\max} + V_2]$, where $V_{\min}$ and $V_{\max}$ are the minimum and maximum of the second factor, respectively. Similarly, V_1 and V_2 are two positive real numbers used to divide the universe of discourse of the second factor into m equal length intervals. Note that the length of the interval of each factor is determined by its largest value of the factor in the historical data.

Step 2: Define fuzzy sets. Linguistic terms, A_i, $1 \le i \le n$, are defined as the fuzzy sets on the intervals of the first factor. They are defined as follows:

$$\begin{aligned}
A_1 =& 1/u_1 + 0.5/u_2 + 0/u_3 + \cdots + 0/u_{n-2} \\
&+ 0/u_{n-1} + 0/u_n, \\
A_2 =& 0.5/u_1 + 1/u_2 + 0.5/u_3 + \cdots + 0/u_{n-2} \\
&+ 0/u_{n-1} + 0/u_n, \\
&\vdots \\
A_{n-1} =& 0/u_1 + 0/u_2 + 0/u_3 + \cdots + 0.5/u_{n-2} \\
&+ 1/u_{n-1} + 0.5/u_n, \\
A_n =& 0/u_1 + 0/u_2 + 0/u_3 + \cdots + 0/u_{n-2} \\
&+ 0.5/u_{n-1} + 1/u_n,
\end{aligned}$$

where u_i denotes the ith interval of the first factor.

Similarly, linguistic term B_j, $1 \le j \le m$, is defined as a fuzzy set on the intervals of the second factor. They are defined as follows:

$$\begin{aligned}
B_1 =& 1/v_1 + 0.5/v_2 + 0/v_3 + \cdots + 0/v_{m-2} \\
&+ 0/v_{m-1} + 0/v_m, \\
B_2 =& 0.5/v_1 + 1/v_2 + 0.5/v_3 + \cdots + 0/v_{m-2} \\
&+ 0/v_{m-1} + 0/v_m, \\
&\vdots \\
B_{m-1} =& 0/v_1 + 0/v_2 + 0/v_3 + \cdots + 0.5/v_{m-2} \\
&+ 1/v_{m-1} + 0.5/v_m,
\end{aligned}$$

$$B_m = 0/v_1 + 0/v_2 + 0/v_3 + \cdots + 0/v_{m-2} + 0.5/v_{m-1} + 1/v_m,$$

where v_i is the ith interval of the second factor.

Step 3: Construct the FLRs database. For historical data on day i, let X_{i-n} and Y_{i-n} denote the fuzzy set of $F_1(i-n)$ and $F_2(i-n)$ of the fuzzy time series, respectively. The FLRs with 2-factor 3-order on day i can be represented by (X_{i-3}, Y_{i-3}), (X_{i-2}, Y_{i-2}), $(X_{i-1}, Y_{i-1}) \rightarrow X_i$. Then, the 2-factor 3-order FLRs database can be constructed as shown in Table 1. In Table 1, the symbols, J and $J-T$, denote expiration times, namely, the remaining J day and $J-T$ day to the expiration date, respectively.

Step 4: Construct the LHS of FLR on the predicting day (assume that day t is the prediction day). If the expiration time is $J-T-1$ day, the LHS of the FLR with 2-factor 3-order on day t can be represented as follows:

$$(X_{t-3}, Y_{t-3}), (X_{t-2}, Y_{t-2}), (X_{t-1}, Y_{t-1}).$$

Note that the LHS of the FLR with 2-factor 3-order on day t is called the LHS of the prediction day below.

Step 5: Search the similar FLRs to generate a training data. Due to the fact that there are 2 factors and 3 orders at the LHS of a FLR, it is difficult to find a matching FLR for prediction. To solve this problem, we use two conditions to search similar FLRs in this paper:

(a) According to the literature, the expiration time of options would affect the volition of investors on option price [8-10]. Accordingly, FLRs are selected from the FLR database when their expiration time and the prediction day's expiration time are the same.

(b) When there are no FLRs selected from the FLRs database, we refer to the literature [13] to search for similar FLRs. Note that in the literature [13], the authors used a Euclidean distance to measure the difference between the LHS of the FLR database and the LHS of the prediction day. Subsequently, they selected k FLRs with the smallest difference from the FLR database. However, the 2 factors of the fuzzy time series model play different roles in predicting option price. To search similar FLRs, we use a Mahalanobis distance to measure the similarities between the LHS of the FLR database and the LHS of the prediction day to balance the weight of the 2 factors in this paper. The Mahalanobis distance between the LHS of the prediction day and the ith LHS of the FLR database can be calculated according to Eq. (5):

$$D_i = \sqrt{(IX_i - TX)^T S^{-1} (IX_i - TX)}, \tag{5}$$

where IX_i denotes the ith LHS of the FLR database; TX denotes the LHS of the prediction day; S is the covariance matrix for IX_i and TX. In this paper, we select 10 FLRs with the smallest Mahalanobis distance as training data for building LSSVR with a bootstrap model.

Step 6: Build the least square support vector regression. With the similar FLR selected, we can train a LSSVR model to forecast option price. The input of LSSVR contains six variables, which are the LHS of the FLRs. The output of LSSVR only contains one variable, which is the RHS of the FLRs. Simply, the 1st to the 3rd input variables are the subscripts of fuzzy sets of LHS's 1st factor of the FLRs, and the 4th to the 6th input variables are the subscripts of fuzzy sets of LHS's 2nd factor of the FLRs.

Step 7: Forecasting option price on day t with bootstrap model. When the LSSVR model is built, we perform the forecasting of option price on day t by feeding the LHS of the predicting day into the built LSSVR model to get the forecasted subscript of the RHS on the prediction day. The input of the built LSSVR model also contains six variables, which are the LHS of the prediction day. Simply, the 1st to the 3rd input variables are the subscripts of fuzzy sets of LHS's 1st factor of the prediction day, and the 4th to the 6th input variables are the subscripts of fuzzy sets of LHS's 2nd factor of the prediction day. However, due to the fact that the size of the training data might be small, this paper uses a bootstrap method to enhance prediction accuracy. Hence, Steps 6 and 7 are repeated

Table 1. An example of the 2-factor 3-order FLRs database.

FLR	LHS		RHS	Expiration time
FLR_1	$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$	$\rightarrow$	X_4	J
FLR_2	$(X_2, Y_2), (X_3, Y_3), (X_4, Y_4)$	$\rightarrow$	X_5	$J-1$
...	...	...	...	
RLR_m	$(X_{m-3}, Y_{m-3}), (X_{m-2}, Y_{m-2}), (X_{m-1}, Y_{m-1})$	$\rightarrow$	X_m	$J-T$

100 times to build 100 different LSSVR models using the bootstrap method, and then 100 subscripts of the RHS are generated on the prediction day. We then calculate the mean of the 100 subscripts of the RHS on the prediction day as the forecasted subscript of the RHS on the prediction day.

Finally, as the forecasted value is a subscript of a fuzzy set, we have to defuzzify it into the option price forecasting value. We use the weighted average method as the defuzzification method, shown by Eq. (6):

$$\text{forecast\_value} = \begin{cases} \frac{M_1+0.5\times M_2}{1+0.5} & k=1, \\ \frac{0.5\times M_{k-1}+M_k+0.5\times M_{k+1}}{0.5+1+0.5} & 1<k\leq n-1, \\ \frac{0.5\times M_{n-1}+M_n}{0.5+1} & k=n, \end{cases} \tag{6}$$

where M_k denotes the midpoint value of the fuzzy set, k. Note that an iteration of the above procedure (Step 1 through Step 7) predicts only one forecasting value.

4. Option price forecasting

To forecast the price of the "Taiwan Stock Exchange Stock Price Index Option (TXO)", we select the closing price of TXO as the first factor and the ratio of the spot price divided by the strike price, which is termed S/K (S is the spot price; K is the strike price), as the second factor in FLSSVR with a bootstrap model. Next, we give a simple example to explain the procedures of our proposed model to forecast an option on 4-Mar-2005, with a strike price equal to 6,000 and an expiration date on April, 2005. Part of the historical data is shown in Table 2. In this historical data, U is set at $[0, 1500]$ and is divided into 150 intervals. That is, $u_1 = [0, 10]$, $u_2 = (10, 20], \cdots$, and $u_{150} = (1490, 1500]$. For the second factor, V is set at $[0.6, 1.300]$ and divided into 141 intervals, that is $v_1 = [0.6, 0.605]$, $v_2 = (0.605, 0.610], \cdots$, and $v_{141} = (1.295, 1.300]$. Having defined the intervals, we fuzzify the historical data into fuzzy sets and construct the 2-factor 3-order FLRs database from the fuzzified historical data. A 2-factor 3-order FLRs database for this historical data is shown in Table 3.

Having constructed the FLRs database, the option price can be forecast by our proposed model. For example, if we want to forecast the option price on 4-Mar-2005 for a strike price equal to 6,000 and expiration date in April, 2005, we first construct the LHS of FLR on 4-Mar-2005, as follows:

$$(A_{28}, B_{87}), (A_{28}, B_{88}), (A_{30}, B_{89}).$$

Then, we use the 1st condition to search the similar FLRs from FLRs database. Due to the fact that the expiration time on 4-Mar-2005 is 33, the expiration times of FLRs that equal 33 are selected from the FLR database. Finally, we use these selected FLRs as the training data to build the LSSVR with a bootstrap approach for the prediction, and feed the LHS of the predicting day into the built LSSVR model, as shown in Table 4. After the performing bootstrap method 100 times, the mean of the forecasted option price is 265.1231 and the variance is 43.9782 on 4-Mar-2005. Note that the actual option price on 4-Mar-2005 is 255 in this example.

Table 2. Parts of historical data.

Date	Option prices	S/K	Expiration time
2005/01/03	585	1.0970	12
2005/01/04	481	1.0822	11
2005/01/05	418	1.0694	10
2005/01/06	420	1.0682	9
2005/01/07	368	1.0600	8
...	...	...	...
2005/01/04	630	1.1019	35
2005/01/05	555	1.0888	34
2005/01/06	565	1.0877	33
2005/01/07	510	1.0793	32
2005/01/11	530	1.0865	31
...	...	...	...
2005/01/06	470	1.0682	36
2005/01/07	444	1.0600	35
2005/01/11	437	1.0671	34
2005/01/12	336	1.0498	33
2005/01/13	330	1.0453	32

5. Results and performance

5.1. Dataset

The dataset of this paper are the daily transaction data of TXO and TAIEX from January 3, 2005 to December 29, 2006. This paper investigates a sample of 23,819 call option price data. Call options can be divided into three categories, according to their S/K ratio. The distribution of the dataset, according to categories of moneyness, which is a term describing the relationship between the strike price and the current spot price of an option, is shown in Table 5. We refer to the literature [15,16] for the definition of the categories. In Table 5, in-the-money denotes that the strike price is above the spot price; at-the-money denotes that the current spot price and strike price are the same; and out-of-the-money denotes that the strike price is below the spot price. In this paper, 70% of the dataset are

Table 3. The FLRs database.

LHS		RHS	Expiration time
$(A_{59}, B_{101}), (A_{49}, B_{98}), (A_{42}, B_{95})$	$\rightarrow$	A_{42}	10
$(A_{49}, B_{98}), (A_{42}, B_{95}), (A_{42}, B_{95})$	$\rightarrow$	A_{37}	9
...	...	...	...
$(A_{63}, B_{102}), (A_{56}, B_{99}), (A_{57}, B_{99})$	$\rightarrow$	A_{51}	33
$(A_{56}, B_{99}), (A_{57}, B_{99}), (A_{51}, B_{97})$	$\rightarrow$	A_{53}	32
...	...	...	...
$(A_{47}, B_{95}), (A_{45}, B_{93}), (A_{44}, B_{95})$	$\rightarrow$	A_{34}	34
$(A_{45}, B_{93}), (A_{44}, B_{95}), (A_{34}, B_{91})$	$\rightarrow$	A_{33}	33

Table 4. An example of FLSSVR with a bootstrap model.

	X_1	X_2	X_3	X_4	X_5	X_6		Y	Expiration time
Training samples	57	56	63	98	98	101	$\rightarrow$	51	33
	34	44	45	90	94	92	$\rightarrow$	33	33
	28	26	31	87	87	86	$\rightarrow$	26	33
			...				...	...	...
	28	28	30	87	88	89	$\rightarrow$	26	33
	23	20	14	86	84	81	$\rightarrow$	21	33
Testing sample	28	28	30	87	88	89	$\rightarrow$	?	33

Table 5. Data distribution according to categories of moneyness.

Categories	Moneyness	Number
In-the-money	$S/K > 1.02$	8938
At-the-money	$0.95 < S/K \leq 1.02$	7508
Out-of-the-money	$S/K \leq 0.95$	7373

Notes: S is the spot price; K is the strike price.

selected as the training dataset. The remaining 30% are adopted as the testing dataset.

5.2. Performance measures

There are two different performance measures, the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE), which are used to measure the prediction accuracy of our proposed model and that of existing models. The formulae are shown in the following:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (A_t - P_t)^2}{n}}, \tag{7}$$

$$\text{MAE} = \frac{\sum_{t=1}^{n} |A_t - P_t|}{n}, \tag{8}$$

where A_t and P_t denote the actual option price and the forecasting option price on day t, respectively.

5.3. Performance

This section first compares the performance of FLSSVR with a bootstrap model and that of FLSSVR without a bootstrap model to verify that the bootstrap method is necessary for the proposed model. The performances of these are shown in Table 6. According to Table 6, the bootstrap method is necessary for the proposed model because the performance of FLSSVR with a bootstrap model is better than that of FLSSVR without a bootstrap model in all categories. Subsequently, the FLSSVR with a bootstrap model is compared with support vector Fuzzy Regression Machines (FSVR) [32], as shown in Table 6. According to Table 6, FLSSVR with a bootstrap model outperforms the FSVR in terms of RMSE and MAE in all categories. Finally, the B-S model and the existing models, which were Weighted Fuzzy Time Series Neural Network (WFTSNN), Fuzzy Time Series Neural Network (FTSNN), GJR-GARCH model, Grey-GJR-GARCH model, EGARCH, and Gery-EGARCH, are used to compare the FLSSVR with a bootstrap model. The performances are also shown in Table 6. Table 6 shows that the performances of FLSSVR with a bootstrap model are better than the existing models, except in the in-the-money category. The performance of the B-S model is the best in the in-the-money category, but, that of the B-S model is the worst in the at-the-money category and the out-of-the-money category. Moreover, Table 6 also shows that the

Table 6. The performance in RMSE and MAE.

Methods	RMSE			MAE		
	Category			Category		
	In[a]	At[b]	Out[c]	In	At	Out
FLSSVR1[d]	53.96	20.59*	6.81*	40.12	14.55*	4.06*
FLSSVR2[e]	65.06	21.04	11.31	41.13	15.13	6.48
B-S model	39.52*	55.69	31.61	31.11*	49.52	25.84
WFTSNN [3]	67.94	29.72	8.96	52.13	19.99	6.05
FTSNN [2]	72.79	36.99	16.19	52.11	23.98	9.78
SVFR[f] [32]	89.65	31.76	24.04	48.89	22.70	22.65
GARCH [16]	85.49	44.02	25.73	69.54	34.73	18.78
GJR[g] [16]	76.19	41.06	25.53	59.28	31.67	17.41
Gery-GJR[h] [16]	73.76	40.11	25.89	56.13	30.21	17.26
EGARCH [15]	73.90	41.35	26.34	57.02	32.17	18.30
Gery-EGARCH [15]	72.11	40.26	26.13	57.26	32.51	18.26

'*' denotes the smallest value;

[a] 'In' denotes In-the-money category;

[b] 'At' denotes At-the-money category;

[c] 'Out' denotes Out-of-the-money category.

[d] 'FLSSVR1' denotes FLSSVR with a bootstrap model;

[e] 'FLSSVR2' denotes FLSSVR without a bootstrap model;

[f] 'SVFR' denotes support vector fuzzy regression machines;

[g] 'GJR' denotes GJR-GARCH model;

[h] 'Grey-GJR' denotes Grey-GJR-GARCH model.

Table 7. The test results of RMSE for FLSSVR with a bootstrap model, WFTSNN and B-S model.

Category								
In[a]			At[b]			Out[c]		
	F[d]	W[e]		W	B[f]		W	B
B	$< 0.001^{**}$	$< 0.001^{**}$	F	$< 0.001^{**}$	$< 0.001^{**}$	F	$< 0.001^{**}$	$< 0.001^{**}$
F		$< 0.001^{**}$	W		0.1086	W		$< 0.001^{**}$

Note that "**" denotes significance at $\alpha = 0.01$;

[a] 'In' denotes In-the-money category;

[b] 'At' denotes At-the-money category;

[c] 'Out' denotes Out-of-the-money category;

[d] 'F' denotes FLSSVR with a bootstrap model;

[e] 'W' denotes WFTSNN;

[f] 'B' denotes B-S model.

performance of a traditional model, such as GARCH, is always worse than the other hybrid models. Although the performance of FLSSVR with a bootstrap model is better than that of the B-S model and that of WFTSNN, their performances are close. To compare the performances of FLSSVR with a bootstrap model of the B-S model, and that of the WFTSNN, we use a t-test to evaluate the RMSE and MAE, as shown in Tables 7 and 8, respectively. Table 7 shows that the RMSE of FLSSVR with a bootstrap model is significantly better than that of the other methods in the at-the-money and the out-the-money categories. In addition, the RMSE of the B-S model is significantly better than the other models in the in-the money category. Furthermore, although the RMSE of WFTSNN is better than that of the B-S model in the at-the-money category, they are insignificantly different. Similarly, Table 8 shows that the MAE of FLSSVR with a bootstrap model is also significantly better than that of the other methods in the at-the-money and the out-the-money categories. Furthermore, Figure 3 shows the forecasting results of an option with a strike price equal to 5,600 and an expiration date in Mar, 2005. In Figure 3, the forecasting option price of FLSSVR with a bootstrap model and that of the B-S model are closer to the actual option price, except on the dates where the option prices changed abruptly. Furthermore, Figure 4 shows the scatterplots of actual and forecasted

Table 8. The test results of MAE for FLSSVR with a bootstrap model, WFTSNN and B-S model.

Category								
	In[a]			**At**[b]			**Out**[c]	
	F[d]	**W**[e]		**W**	**B**[f]		**W**	**B**
B	$< 0.001^{**}$	$< 0.001^{**}$	**F**	$< 0.001^{**}$	$< 0.001^{**}$	**F**	$< 0.001^{**}$	$< 0.001^{**}$
F		$< 0.001^{**}$	**W**		$< 0.001^{**}$	**W**		$< 0.001^{**}$

Note that "**" denotes significance at $\alpha = 0.01$;
[a] 'In' denotes In-the-money category;
[b] 'At' denotes At-the-money category;
[c] 'Out' denotes Out-of-the-money category;
[d] 'F' denotes FLSSVR with a bootstrap model;
[e] 'W' denotes WFTSNN;
[f] 'B' denotes B-S model.

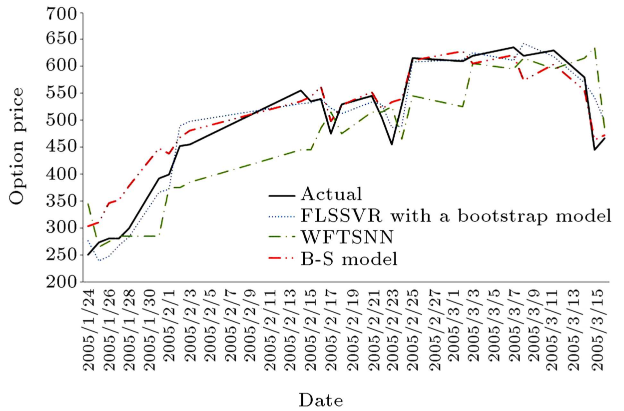

Figure 3. Time series of actual and forecasted price of an option with strike price equal to 5,600 and expiration date of Mar, 2005.

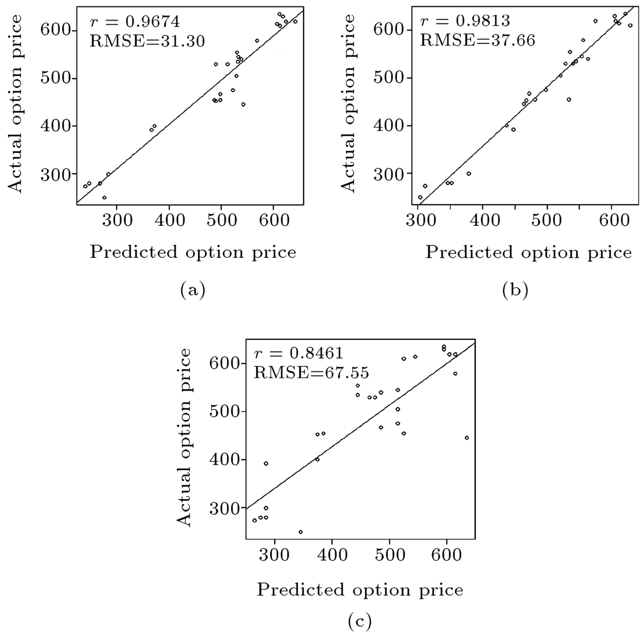

Figure 4. The scatterplot of actual and predicted option price: (a) FLSSVR with a bootstrap model; (b) B-S model; and (c) WFTSNN.

option prices for FLSSVR with a bootstrap model, B-S model, and WFTSNN. Figure 4 also shows RMSE and the correlation coefficient (r) of actual and predicted option prices for FLSSVR with a bootstrap model, B-S model, and WFTSNN. According to Figure 4, FLSSVR with a bootstrap model and the B-S model outperform WFTSNN in the RMSE and the correlation coefficient. Although the correlation coefficient of FLSSVR with a bootstrap model and that of the B-S model are close, they are insignificantly different (p-value=0.96), according to the correlation coefficient test. To summarize, FLSSVR with a bootstrap model outperforms existing methods in terms of RMSE, MAE and the correlation coefficient of actual and forecasted option prices.

6. Conclusion

In this paper, we propose a novel hybrid model to forecast option price. The proposed model is composed of a 2-factor 3-order FTS model, LSSVR, and a bootstrap method. The experiment results showed that FLSSVR with a bootstrap model is more accurate than other existing methods, in terms of RMSE and MAE, for options belonging to the out-the-money and at-the-money categories. The RMSE and MAE of FLSSVR with a bootstrap model belonging to the in-the-money category are worse than those of the B-S model. In addition, the performance of FLSSVR with a bootstrap model is also better than that of WFTSNN, in terms of the correlation coefficient of actual and forecasted option prices. Hence, FLSSVR with a bootstrap model offers a useful alternative for option price forecasting.

References

1. Ko, P.-C. "Option valuation based on the neural regression model", *Expert Systems with Applications*, **36**, pp. 464-471 (2009).
2. Leu, Y., Lee, C.-P. and Hung, C.-C. "A fuzzy time series-based neural network approach to option price forecasting", In *Proceedings of the Second In-*

ternational Conference on Intelligent Information and Database Systems, Vietnam, pp. 360-369 (2010).

3. Leu, Y., Lee, C.-P. and Hung, C.-C. "A weighted fuzzy time series aased neural network approach to option price forecasting", In *Granular Computing and Intelligent Systems*, pp. 237-248 (2011).

4. Huang, S.-C. "Online option price forecasting by using unscented Kalman filters and support vector machines", *Expert Systems with Applications*, **34**, pp. 2819-2825 (2008).

5. Black, F. and Scholes, M. "The pricing of options and corporate liabilities", *Journal of Political Economy*, **81**, pp. 637-654 (1973).

6. Liang, X., Zhang, H., Xiao, J. and Chen, Y. "Improving option price forecasts with neural networks and support vector regressions", *Neurocomputing*, **72**, pp. 3055-3065 (2009).

7. Mercuri, L. "Option pricing in a Garch model with tempered stable innovations", *Finance Research Letters*, **5**, pp. 172-182 (2008).

8. Broughton, J.B., Chance, D.M. and Smith, D.M. "The impact of equity option expirations on the prices of non-expiring options", *Review of Financial Economics*, **4**, pp. 109-123 (1995).

9. Hauser, S., Kamara, A. and Shurki, I. "The effects of randomizing the opening time on the performance of a stock market under stress", *Journal of Financial Markets*, **15**, pp. 392-415 (2012).

10. Xiaoyan Ni, S., Pearson, N.D. and Poteshman, A.M. "Stock price clustering on option expiration dates", *Journal of Financial Economics*, **78**, pp. 49-87 (2005).

11. Chen, S.-M. "Forecasting enrollments based on fuzzy time series", *Fuzzy Sets and Systems*, **81**, pp. 311-319 (1996).

12. Chen, S.-M. "Forecasting enrollments based on high-order fuzzy time series", *Cybernetics and Systems*, **33**, pp. 1-16 (2002).

13. Leu, Y., Lee, C.-P. and Jou, Y.-Z. "A distance-based fuzzy time series model for exchange rates forecasting", *Expert Systems with Applications*, **36**, pp. 8107-8114 (2009).

14. Yu, H.-K. "Weighted fuzzy time series models for TAIEX forecasting", *Physica A: Statistical Mechanics and Its Applications*, **349**, pp. 609-624 (2005).

15. Tseng, C.-H., Cheng, S.-T., Wang, Y.-H. and Peng, J.-T. "Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices", *Physica A: Statistical Mechanics and Its Applications*, **387**, pp. 3192-3200 (2008).

16. Wang, Y.-H. "Nonlinear neural network forecasting model for stock index option price: Hybrid GJR—GARCH approach", *Expert Systems with Applications*, **36**, pp. 564-570 (2009).

17. Liang, X., Zhang, H. and Li, X. "A simple method of forecasting option prices based on neural networks", *in Next-Generation Applied Intelligence*, pp. 586-593 (2009).

18. Wang, C.-C., Tu, Y.-H. and Wong, H.-L. "The comparison between multivariate fuzzy time series and traditional time series modeling to forecasting China exports", In *Proceedings of the 4th International Conference on Innovative Computing, Information and Control, IEEE Computer Society*, pp. 1487-1490 (2009).

19. Song, Q. and Chissom, B.S. "Fuzzy time series and its models", *Fuzzy Sets and Systems*, **54**, pp. 269-277 (1993).

20. Song, Q. and Chissom, B.S. "Forecasting enrollments with fuzzy time series - Part I", *Fuzzy Sets and Systems*, **54**, pp. 1-9 (1993).

21. Song, Q. and Chissom, B.S. "Forecasting enrollments with fuzzy time series - Part II", *Fuzzy Sets and Systems*, **62**, pp. 1-8 (1994).

22. Lee, L.-W., Wang, L.-H., Chen, S.-M. and Leu, Y. "Handling forecasting problems based on two-factors high-order fuzzy time series", *IEEE Transactions on Fuzzy Systems*, **14**, pp. 468-477 (2006).

23. Boser, B.E., Guyon, I.M. and Vapnik, V.N. "A training algorithm for optimal margin classifiers", *Proc. of the 5th Annual Workshop on Computational Learning Theory*, pp. 144-152 (1992).

24. Cervantes, J., Li, X., Yu, W. and Li, K. "Support vector machine classification for large data sets via minimum enclosing ball clustering", *Neurocomputing*, **71**, pp. 611-619 (2008).

25. Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A.J. and Vapnik, V. "Support vector regression machines", In *Advances in Neural Information Processing Systems*, pp. 155-161 (1996).

26. Suykens, J.A.K., De Brabanter, J., Lukas, L. and Vandewalle, J. "Weighted least squares support vector machines: Robustness and sparse approximation", *Neurocomputing*, **48**, pp. 85-105 (2002).

27. Samui, P. and Kothari, D.P. "Utilization of a least square support vector machine (LSSVM) for slope stability analysis", *Scientia Iranica*, **18**, pp. 53-58 (2011).

28. Vapnik, V. "A lerner, pattern recognition using generalized portrait method", *Automation and Remote Control*, **24**, pp. 774-780 (1963).

29. Efron, B. "Bootstrap methods: Another look at the jackknife", *The Annals of Statistics*, **7**, pp. 1-26 (1979).

30. Diaconis, P. and Efron, B. "Computer-intensive methods in statistics", *Scientific American*, pp. 116-130 (1983).

31. Efron, B. and Tibshirani, R. "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy", *Statistical Science*, **1**, pp. 54-75 (1986).

32. Hong, D.H. and Hwang, C. "Support vector fuzzy regression machines", *Fuzzy Sets and Systems*, **138**, pp. 271-281 (2003).

Biographies

Chien-Pang Lee received a BS degree in Applied Statistics from Ming Chuan University, Taiwan, in 2003, an MS degree in Biostatistics from National Chung Hsing University, Taiwan, in 2006, and a PhD degree in Information Management from the National Taiwan University of Science and Technology, Taiwan, in 2010. He is currently Assistant Professor in the Department of Information Management at Da-Yeh University, Taiwan. His research interests include data mining, bioinformatics, statistics, and financial forecasting.

Wen-Cheng Lin obtained a PhD in Management from National Taiwan Ocean University, Taiwan, and is currently Assistant Professor in the Department of Business Administration at National Taipei College of Business, Taiwan. His research interests include performance evaluation, Balanced Scored Card (BSC), intellectual capita and fuzzy set theory.

Chih-Chung Yang obtained an MS degree in Information Management from Chung Yuan Christian University, Chung Li, Tao Tuan, Taiwan, in 2007, and is presently studying for a PhD degree in the same subject at National Taiwan University of Science and Technology, Taipei City, Taiwan. He is also research assistant in the Department of Information Management at the National Taiwan University of Science and Technology and National Taipei University of Nursing and Health Sciences. His current research interests include portfolio optimization models, investment performance, information technology strategy application, investment decision-making support, commonly used methods for genetic algorithms, data envelopment analysis, soft computing, and optimization theory.