RESEARCH ARTICLE

Forecasting leading industry stock prices based on a hybrid time-series forecast model

Ming-Chi Tsai, Ching-Hsue Cheng, Meei-Ing Tsai, Huei-Yuan Shiu

1 Department of Business Administration, I-Shou University, Dashu District, Kaohsiung City, Taiwan,
2 Department of Information Management, National Yunlin University of Science and Technology, Douliou, Yunlin, Taiwan

☯ These authors contributed equally to this work.
* chcheng@yuntech.edu.tw

Abstract

Many different time-series methods have been widely used in forecast stock prices for earning a profit. However, there are still some problems in the previous time series models. To overcome the problems, this paper proposes a hybrid time-series model based on a feature selection method for forecasting the leading industry stock prices. In the proposed model, stepwise regression is first adopted, and multivariate adaptive regression splines and kernel ridge regression are then used to select the key features. Second, this study constructs the forecasting model by a genetic algorithm to optimize the parameters of support vector regression. To evaluate the forecasting performance of the proposed models, this study collects five leading enterprise datasets in different industries from 2003 to 2012. The collected stock prices are employed to verify the proposed model under accuracy. The results show that proposed model is better accuracy than the other listed models, and provide persuasive investment guidance to investors.

Introduction

The prices forecast of stock is the most key issue for investors in the stock market, because the trends of stock prices are nonlinear and nonstationary time-series data, which makes forecasting stock prices a challenging and difficult task in the financial market. Conventional time series models have been used to forecast stock prices, and many researchers are still devoted to the development and improvement of time-series forecasting models. The most well-known conventional time series forecasting approach is autoregressive integrated moving average (ARIMA) [1], which is employed when the time-series data is linear and there are no missing values [2]. Statistical methods, such as traditional time series models, usually address linear forecasting models and variables must obey statistical normal distribution [3]. Therefore, conventional time series methods are not suitable for forecasting stock prices, because stock price fluctuation is usually nonlinear and nonstationary.

Further, most conventional time-series models utilize one variable (the previous day’s stock price) only [4], when, there are actually many influential factors, such as market indexes, technical indicators, economics, political environments, investor psychology, and the fundamental
financial analysis of companies that can influence forecasting performance [5]. In practice, researchers use many technical indicators as independent variables for forecasting stock prices. How to select the key variables from numerous technical indicators is a critical step in the forecasting process. Investors usually prefer to select technical indicators depending on their experience or feelings for forecasting stock prices despite this behavior be highly risky. However, choosing unrepresentative indicators may result in losing profits for investors. Therefore, selecting the relevant indicators to forecast stock prices is one of the important issues for investors. Financial researchers must identify the key technical indicators that have higher relevance to the stock price by indicator selection. Therefore, proposed models must incorporate indicator selection in the stock forecasting process to enhance forecasting accuracy.

Recently, there have been many new forecasting techniques used to construct efficient and precise machine learning models, but forecasting stock prices is still a hot topic [6, 7, 8]. To overcome the shortcomings of traditional time series models, nonlinear approaches have been proposed, such as fuzzy neural networks [9, 10, 11, 12], and support vector regression (SVR) [13, 14, 15, 16]. SVR utilizes the minimized structural risk principle to evaluate a function by using the minimized the upper bound of the generalized error [17, 18]. The minimized structural risk principle could get better generalization from limited size datasets [19]. Further, SVR has a global optimum and exhibits better prediction accuracy due to its implementation of the structural risk minimization principle, which considers both the training error, and the capacity of the regression model [15, 20]. Although SVR has shown a great number of experimental results in many applications such as economic and financial predictions, the main problem of SVR is the determination of its parameters, which requires practitioner experience [21]. In the literature, genetic algorithms (GA) have been successfully used in a wide range of problems included machine learning, multiobjective optimization problems and multimodal function optimization [22]. GA is a search algorithm inspired by evolution and is usually used to solve optimization problems. Therefore, proposed models utilize GA to optimize the parameters of SVR and obtain better forecasting performance.

From the related work mentioned above, previous studies have shown some drawbacks: (1) Many researches select key technical indicators depending on experiences and ideas [23], (2) Most statistical methods follow some assumptions in different datasets, and obey the statistical distributions [3], (3) Most previous time series models consider only one feature to forecast stock indexes [23], and (4) The parameter of SVR is difficult to determine [24, 25, 26].

This paper proposes a novel GA-SVR time series model based on indicator selection to overcome these problems, and the proposed model contributes the following: (1) In feature selection, this study applies multivariate adaptive regression spline (MARS), stepwise regression (SR), and kernel ridge regression to get the key technical indicators for investors. (2) The proposed model optimizes the parameters of SVR by genetic algorithm (GA) to increase the forecast accuracy. (3) The results could provide persuasive investment guidelines for investors.

The remaining contents of this paper are organized as follows. Section 2 describes the related methodology that incorporate the technical indicator, MARS, genetic algorithm, SVR, and stepwise regression. Section 3 presents the proposed algorithm. Section 4 provides the experimental results and comparisons. Conclusion of this paper are explained in Section 5.

Related work
This section introduces the related work containing the technical indicator, multivariate adaptive regression splines, genetic algorithm, support vector regression, and briefly stepwise regression.
Technical indicator

The technical indicator (TI) is an investment guidance for investors based on evaluating the profits of securities from analyzing trading data of marketing activities, such as past prices and volumes [5]. Stock market data have highly nonlinear, and many researches have focused on the technical indicator to increase the investment return [27, 28]. A technical indicator is a formula, which transfers trading data (open price, the lowest price, the highest price, average price, closing price and volume) into different technical indicators, and try to forecast future prices based on analyzing the past pattern of stock prices [29, 30]. Technical analysis utilizes basic market data, and assumes that the involved factors are included in the stock exchange information [31]. Based on literature review, this paper collected some technical indicators as Table 1. To consider more features, that affect the stock price and volatility, this paper incorporates the microeconomic features that affect the stock price, and these collected factors are listed in Table 2.

Multivariate adaptive regression splines

Friedman [46] proposed multivariate adaptive regression splines (MARS), it is a simple non-parametric regression algorithm. The main advantages of MARS is its capacity to grasp the

Table 1. Technique indicators.

| Indicator | Explanation |
|-----------|-------------|
| MA5       | MA5 (5 days moving average) = \( \frac{1}{5} \sum_{i=1}^{5} p_c \), \( i = 4 \), and \( p_c \) is the closing index of the current day [32] |
| MA10      | MA10 (10 days moving average) = \( \frac{1}{10} \sum_{i=1}^{10} p_c \), \( i = 9 \), and \( p_c \) is the closing price of the trading day [32] |
| 5BIAS     | The difference between the closing price and MA5, which utilizes the stock price nature of returning back to average price for analyzing the stock trends [31] |
| 10BIAS    | The difference between the closing price and MA10, which employs the stock price nature of returning back to average price for analyzing the stock trends [31] |
| RSI       | RSI measures the magnitude of recently gain to recently loss in an trial to determine overbought and oversold conditions of an asset [31] |
| 12PSY     | PSY12 (12 days psychological line) = (\( D_{up12}/12 \) \* 100, \( D_{up12} \) is the number of days when price is going up within 12 days [23]) |
| 10WMS%R   | Williams %R is usually drawn by using negative values. For analysis and discussion, ignore the negative symbols. It is the best to wait the security’s price until change direction before placing your trading [31] |
| MACD      | MACD presents the difference between a fast and slow exponential moving average (EMA) for closing prices. Fast is a short-period average, and slow is a long period one [31] |
| MO1       | MO1(t) = price(t) – price(t – n), \( n = 1 \) [33] |
| MO2       | MO2(t) = price(t) – price(t – n), \( n = 2 \) [33] |
| Transaction volume | Transaction volume presents a basic yet very important element of market timing strategy. Volume gives clues for the intensity of given price moving [34] |
| CDP value | Divide the previous price movement into five values and make the intraday trading decision based on the five value [31] |
| Exponential Moving Average (EMA) | EMA is defined as a linear transformation of time series to a smoother time series by \( \tilde{x}_t = \lambda \sum_{k=0}^{\infty} (1 - \lambda)^k x_{t-k} \). Where 0 < \( \lambda \) < 1 is the timescale. When \( \lambda = 1 \), the EMA is the identity transformation: \( \tilde{x}_t = x_t \); in contrast, many term \( x_{t-k} \) effectively contribute to \( \tilde{x}_t \) when \( \lambda < 1 \) [35]. |
| Company-Daily price change | \( P = \frac{p_c - p_{oT}}{p_{oT}} \times 100 \%), \( P_c \) is the close price of today and \( P_{oT} \) is the open price of today [36]. |
| TAIEX-Daily index change | \( P = \frac{p_{cT} - p_{oT}}{p_{oT}} \times 100 \%), \( P_{cT} \) is the close index of today and \( P_{oT} \) is the open index of today [36]. |

https://doi.org/10.1371/journal.pone.0209922.t001
complicated data mapping and patterns of high-dimensional data, and produce more simple, easy interpretation models, and its can perform analysis on feature relative importance. In concept, MARS integrates the piecewise linear regressions into a flexible model for solving the nonlinear and complex problems. MARS establishes the final model in a two-stage procedure: Firstly, the forward stage, many spline basis functions are built, the feature can be continuous, ordinal, or categorical. Secondly, the backward stage removes the redundant spline basis functions, it uses the generalized cross-validation (GCV) criterion \[46\] to evaluate the performance of model subsets for getting the best subset, the lower \(G_{CV}\) value is better. Moreover, the \(G_{CV}\) is defined as Eq (1).

\[
G_{CV}(M) = \frac{1}{N} \sum_{i=1}^{N} \frac{[y_i - f_M(x_i)]^2}{1 - \frac{C(M)}{N^2}}
\]  

(1)

where \(N\) is the number of data records, \(C(M)\) denotes the penalty cost of a model containing \(M\) basis functions, the numerator is the lack of fit on the \(M\) basis function model \(f_M(x_i)\), the denominator is the penalty for model complexity \(C(M)\) and \(y_i\) denotes the target outputs.

Genetic algorithm

The genetic algorithm (GA) [47] is to search the global optimum by using inspired natural evolve. Moreover, GA has four operators (inheritance, mutation, selection, and crossover) to
evolve repeatedly for obtaining the optimal solution. GA has been applied successfully in economic and financial domain [27, 48]. In specific problem, the GA algorithm encodes a potential solutions into the simple chromosome-like data structure, and applies the re-united operators to preserve critical information [3]. This paper referred the GA steps of Goldberg [49], and reorganized as follows:

Step1: Generate an initial population randomly.
Step2: Evaluate fitness of each chromosome.
Step3: Check the stop criterion.
Step4: Select suitable chromosomes based on the parents’ populations.
Step5: Extend crossover to search a new solution by swapping corresponding to segments of a string representation for the parents.
Step6: Employ mutation randomly to change some of the chosen chromosomes.

Support vector regression

SVR algorithm is a nonlinear kernel-based regression, which tries to find a regression hyperplane with minimized risk in high dimensional space [16]. Compared to the traditional regression model, it estimates the coefficients by minimizing the square loss, SVR uses the \( \varepsilon \)-insensitivity loss function to obtain its parameters. It can be express as:

\[
L_{\varepsilon}(f(x), t) = \begin{cases} 
|f(x) - t| - \varepsilon & \text{if } |f(x) - t| \geq \varepsilon \\
0 & \text{otherwise}
\end{cases}
\]  

where \( t \) is the desired (target) outputs, and \( \varepsilon \) defines the region of \( \varepsilon \)-insensitivity, when the predicted value falls into the band area, the loss is zero. Contrarily, if the predicted value falls outside of the band area, the loss is equal to the difference between the predicted value and the margin.

Considering empirical risk and structural risk, the SVR model can use slack variables to construct a minimal quadratic programming problem.

\[
\text{Min} : \frac{1}{2} \|\mathbf{v}\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)
\]

subject to

\[
\begin{align*}
q_i - (\mathbf{v} \cdot \varphi(x_i)) - b & \leq \varepsilon + \xi_i \\
(\mathbf{v} \cdot \varphi(x_i)) + b - q_i & \leq \varepsilon + \xi_i^* \\
\xi_i, \xi_i^* & \geq 0, \text{ for } i = 1, \cdots, n
\end{align*}
\]  

The symbols \( \xi_i \) and \( \xi_i^* \) are two positive slack variables to calculate the error \((q_i - f(x_i))\) from the boundaries of the \( \varepsilon \)-insensitivity zone. \((\xi_i + \xi_i^*)\) denotes the empirical risk, \( \frac{1}{2} \|\mathbf{v}\|^2 \) is the structural risk to prevent over-learning and the lack of applied universality and \( C \) denotes the regularization constant for specifying the trade-off between the empirical risk and the regularization terms.

Based on the sequentially modifying coefficient \( C \), band area width \( \varepsilon \), and kernel function \( K \), the optimal parameter can be solved by the Lagrange method [19]. This study based on
Vapnik [18] utilized the SVR-based regression function, and it is defined as

$$f(x, z) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)K(x, x_i) + b \quad (4)$$

where $\alpha_i$ and $\alpha_i^*$ are the Lagrangian multipliers that satisfy the equality $\alpha_i \alpha_i^* = 0$, $\alpha_i$ and $\alpha_i^* \geq 0$. $K(x, x_i)$ is the kernel function which represents the inner product $\langle \phi(x_i), \phi(x) \rangle$. The radial basis function (RBF) has been widely used as the kernel function, and this study utilizes RBF because of its capabilities and simple implementation [50].

$$K(x, x_i) = \exp(-\gamma \|x_i - x\|^2) \quad (5)$$

where $\gamma$ is the RBF width.

**Stepwise regression**

SR is a simple multiple regressions, it establishes a model by adding or removing features based on the statistics of F-test, that is, SR utilized the forward and backward procedures to add or remove features based on F statistics. SR adds the feature to the model if the p-value of variable is less than the given significant level ($p < .05$), and removes the variable from the model if the p-value of variable is greater than the given significant level [51].

**Proposed model**

This study based on previous studies has found some drawbacks in time series forecast: (1) Based on subjective experiences and opinions to select important technical indicators [23]. (2) Previous methods need to follow some assumptions in different datasets, and obey the statistical distributions [3]. (3) Previous time series models consider only one feature to forecast the stock index [23]. (4) The best SVR parameters are difficult to determine [24, 25, 26]. To overcome these problems, this study is based on our conference paper [52] to extend the proposed methods for solving the forecast problems. That is, this paper proposes a GA-SVR time series model based on feature selection to forecast the leading industry stock price. Hence, the proposed model contributes the following. (1) This study utilizes MARS, SR, and KRR to choose the key technical indicators for investors. (2) Use a GA to optimize the SVR parameters for enhancing the forecast accuracy. (3) The results can provide the investment guidance to investors.

The proposed model included three blocks as Fig 1, it can be briefly described as follows:

1. Data preprocessing: this proposed model transforms daily basic stock data (open price, the lowest price, the highest price, average price, closing price and volume) into technical indicators. Then utilizes MARS, SR, and KRR methods to select the key indicators.

2. Modeling: Build a forecast model by using SVR and employs GA to optimize the parameters of SVR.

3. Forecasting: The optimized GA-SVR forecast model is utilized to forecast the stock price, and compare the proposed models with the listing models under the accuracy.

For easy computation, this section proposed an algorithm with six steps, the detailed step is described as follows:

**Step 1:** Transform trading data into technical indicators

This step collected daily stock trading data (open, close, the highest, the lowest price and volume), and transformed these data into technical indicators [31], such as MA, PSY, RSI,
BIAS, and WMS%. In addition, this paper also incorporates other indicators, such as exchange rate, NT dollars to US dollars, and the momentum. These technical indicators used are listed in Tables 1 and 2, respectively.

Step 2: Select key features (MARS, SR, and KRR)

From step 1, the collected data has been transformed into technical indicators; this step utilized MARS, SR, and KRR to choose the key indicators. For removing collinearity, this step
also run SR multi-collinearity to eliminate the high multi-collinearity indicators. For comparing the three feature selection methods, this study selected the number of features as similar as possible.

Step 3: Construct the SVR forecast model

To build the SVR forecast model, this step used the selected features as input features, and the RBF function is used as the kernel function, due to it can handle the nonlinear and high-dimensional data. To build the forecasting model, three parameters that should be set: the loss function $\varepsilon$, the regularization constant $C$, and the RBF width $\sigma$. To obtain a better forecast model, this step utilizes a genetic algorithm to optimize these parameters.

Step 4: Optimize the SVR parameters by GA at minimal RMSE

To get better forecasting accuracy, this step employed genetic algorithm to optimize the SVR parameters $C$ and $\sigma$ under minimal RMSE (Eq 6) for training dataset. The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2}$$  \hspace{1cm} (6)

where $y_t$ is the real stock index, $\hat{y}_t$ is the forecasted stock price, and $n$ is the number of records.

Step 4 has six sub-steps, which is described the operation of the GA processes as follows:

Step 4.1.1: Initialize the parameter for GA.

The initial population was set 80 individual solutions, and randomly generated in this sub-step, the SVR parameters are encoded into a chromosome by a binary string. In addition, the maximal generations, the crossover probability, the mutation probability, and are given as 2000, 0.8, and 0.08 [53] respectively.

Step 4.1.2: Evaluate fitness.

This sub-step uses a pre-defined fitness function (RMSE) to evaluate fitness of each chromosome for determining the goodness of fit for each solution.

Step 4.1.3: Check the stopping criterion

This sub-step sets the stopping rule: If one of the two conditions is got, then the GA process is stopped:

1. The maximal number of generations is reached (2000).
2. The optimal solution is smaller than the given minimal RMSE, the minimal RMSE is set as $10^{-5}$.

If the criterion is not achieved then repeatedly re-run a new iterative process (Step 4.1.2 to 4.1.5).

Step 4.1.4: Select the parents by the fitness function

Selection is to screen out the fit chromosome to be copies for increasing the offspring sharing and eliminating the poorer chromosome for decreasing the offspring sharing. Roulette wheel selection is employed to select the chromosomes for reproduction.

Step 4.1.5: Perform crossover and mutation
Re-unite the parents to produce and mutate offspring, this step one-point crossover is used. Then, it selected randomly a member of the population, and changed one randomly selected bit in its bit string representation [3].

Step5: Forecast the stock price by using the optimized models

From Step 4, the SVR parameters were determined. the testing data is applied to predict the next day’s stock prices by using the optimized forecast model.

Step6: Compare the performance

For evaluating the forecast accuracy, the proposed model will be compared with the listing models under the RMSE criterion. The seven comparison models are as follows: (1) integrated KRR and SR (KRR-SR model), (2) integrated KRR and MARS (KRR-MARS model), (3) integrated KRR and GA-SVR (KRR-GA-SVR model), (4) integrated SR and MARS (SR-MARS model), (5) integrated SR and SR (SR-KRR model), (6) integrated MARS and SR (MARS-SR model), and (7) integrated MARS and KRR (MARS-KRR model). Where KRR-SR model denotes using SR to build the forecast model after selecting the features by KRR, similarly other combined models have the same meanings, and the detailed abbreviations are presented in Table 3.

**Experiment and comparisons**

This study employs Taiwan’s stock as experimental datasets, the selected companies are different leading industries from “business today (www.businesstoday.com.tw)” which published the 1000 largest companies from Mainland China, Taiwan, and Hong Kong. The experimental datasets including Chunghwa Telecom (CHT), China Steel, Hon Hai, Cathay Financial Holdings and Taiwan Semiconductor Manufacturing Company (TSMC), were practically collected from 2003 to 2012. To compare the accuracy of a long test period forecast and short test period forecast, this study implements two experiments for each dataset, and the two experimental designs are listed in Table 4.

**Table 3. The meaning of model abbreviation.**

| Abbreviation   | Method                                                                 |
|----------------|------------------------------------------------------------------------|
| MARS           | Multivariate adaptive regression splines                               |
| GA             | Genetic algorithm                                                      |
| SR             | Stepwise Regression                                                     |
| KRR            | Kernel Ridge Regression                                                 |
| GA-SVR         | Applying Genetic algorithm to optimize the Support vector regression parameters |
| KRR-SR         | Employing KRR to select the key features and build model by SR         |
| KRR-GA-SVR     | Employing KRR to select the key features and build model by GA-SVR     |
| KRR-MARS       | Employing KRR to select the key features and build model by MARS       |
| SR-KRR         | Employing SR to select the key features and build model by KRR         |
| SR-MARS        | Employing SR to select the key features and build model by MARS        |
| SR-GA-SVR (proposal model B) | Employing SR to select the key features and build model by GA-SVR |
| MARS-KRR       | Employing MARS to choose the key features and build model by KRR       |
| MARS-SR        | Employing MARS to choose the key features and build model by SR        |
| MARS-GA-SVR (proposal model A) | Employing MARS to choose the key features and build model by GA-SVR |

https://doi.org/10.1371/journal.pone.0209922.t003
First, this study conducts an initial experiment to explore the performance of the GA-SVR model. The forecasting performance of the GA-SVR is compared with SR, KRR, and MARS and the results are shown in Table 5. From Table 5, we can see that the GA-SVR model generates the smallest RMSE by the CHT, China Steel and Hon Hai datasets. Therefore, this study combines feature selection with the GA-SVR model as the proposed forecasting model. Second, this study sets the parameters of different forecasting models for the following experiment. In the parameter settings for the MARS model, the training data is utilized to build the MARS model, the maximal number of BFs of the MARS model’s parameter is set as 2000 and the other parameters are set as default [54]. For the KRR model, the parameter lambda for Tikhonov regularization of kernel ridge regression is set as 0.001 to build the forecasting model. In the SR model, the training data is employed to build the forecasting model, the high variance inflation factors (VIF) that are higher than 10 are removed first.

Based on the initial experimental result (GA-SVR model performs better than the other models), this paper combines the GA-SVR model and feature selection method as the proposed model. Then, this study proposes model A and model B based on different feature selection methods. Model A uses MARS to select features, and model B utilizes stepwise regression as the feature selection method.

In comparison, this study selects the same number of features for different feature selection methods. The selected features by MARS, SR and KRR are listed in Table 6. After finding the key features, this study constructs the forecast model by SVR and optimizes the parameters of the MARS-GA-SVR and SR-GA-SVR models by GA, the optimized parameters are listed in Table 7.

### Experimental results

In the section, this study verifies the performance of the proposed model by using five different industry datasets including the Chunghwa Telecom datasets (CHT), China steel datasets, Hon Hai datasets. The experimental design is shown in Table 4. The experiment of the long and short test period.

| Experiment     | Training period      | Testing period      |
|-----------------|----------------------|---------------------|
| Short test period | 2003 to 2011 year | 2012                |
| Long test period | 2003 to 2009 year | 2010 to 2012 year |

Table 5. The initial performance comparisons for three companies in RMSE.

| company     | model | Test period (year) | 1     | 3     |
|-------------|-------|--------------------|-------|-------|
| CHT         | SR    | 0.0662             | 0.7664|       |
|             | KRR   | 15.08679           | 34.3449785 |       |
|             | MARS  | 0.1435             | 2.3455|       |
|             | GA-SVR | 0.001162354       | 0.00329521 |       |
| China Steel | SR    | 0.25348            | 0.34462|       |
|             | KRR   | 3.91307785         | 4.16090037 |       |
|             | MARS  | 0.24370            | 0.32520|       |
|             | GA-SVR | 0.000651508       | 0.000922785 |       |
| Hon Hai     | SR    | 1.89469            | 2.20621|       |
|             | KRR   | 4.27619926         | 12.5219447 |       |
|             | MARS  | 1.9228             | 2.151 |       |
|             | GA-SVR | 0.002353294       | 0.006689822 |       |

Table 4. The experiment of the long and short test period.
Hai datasets, Cathay Financial Holdings datasets, and Taiwan Semiconductor Manufacturing Company datasets. The CHT datasets are employed in the first experiment; the computational process follows the proposed algorithm in Section 3. The predictions of the MARS-GA-SVR and SR-GA-SVR models are demonstrated in Fig 2. Fig 2 shows that the long test period results have more overlap between the forecast line and real closing price line than the results in the short test period in model A (MARS-GA-SVR). For proposed model B (SR-GA-SVR), there is more overlap between the forecast line and real close price line in the short test period results (in Fig 2). The RMSE for the listing models and proposed models, are shown in Table 8, and the results show that the performances of proposed models are better than other models. In the short test period, the proposed model B (SR-GA-SVR) generates the smallest RMSE in listing models as Table 8. Moreover, the proposed model A (MARS-GA-SVR) generates the smallest RMSE in the long test period (in Table 8).

Table 9 and Fig 3 illustrate the numerical results for the China Steel datasets. From Fig 3, the result of model A shows that the real closing price line and the forecast line in the long test period have more overlap than in the short test period. Similarly, compared to the results in the long test period, the results generated by model B have more overlap between closing price line and the forecast line in short test period. From Table 9, the proposed models generate the smallest RMSE in the listing models. In the short test period, the proposed model B generates the smallest RMSE. In addition, the proposed model A generates the smallest RMSE in the long period.

Experiments on the Hon Hai datasets are presented in Table 10 and Fig 4. Fig 4 shows the excellent performance of the proposed model (the forecast line almost completely overlaps the closing price line). Form Table 10, model A and model B generate the smallest RMSE in the long training period and the short training period, respectively.

Table 11 and Fig 5 show the experiments for the Cathay Financial Holdings datasets. In Fig 5, the numerical results clearly show that the forecast line deviates from the closing price line. From Table 11, the KRR-GA-SVR model generates the smallest RMSE in the short test period.

| Company                  | Method | Indicator1       | Indicator2       | Indicator3       | Indicator4       |
|--------------------------|--------|------------------|------------------|------------------|------------------|
| Chunghwa Telecom         | MARS   | CDP value        | MO2              |                  |                  |
|                          | SR     | CDP value        | MO1              |                  |                  |
|                          | KRR    | Sale Month       | P/E              |                  |                  |
| China Steel              | MARS   | Daily price change | TAIEX index | FBAQ             | FBBQ             |
|                          | SR     | CDP value        | MO2              | MO1              | -                |
|                          | KRR    | 10BIAS           | FBBQ             | P/E              | -                |
| Hon Hai                  | MARS   | FBBQ             | In(ROI)          | Transaction volume |
|                          | SR     | EMA              | MACD             | In(ROI)          |
|                          | KRR    | Sale Month       | RS16             | MO1              |
| Cathay Financial Holdings| MARS   | ROI              | FBBQ             | CDP value        |
|                          | SR     | 5BIAS            | MA5              | MO2              |
|                          | KRR    | ROI              | TAIEX -Daily index change | TAIEX index |
| TSMC                     | MARS   | MO2              | FBAQ             | FBBQ             |
|                          | SR     | CDP value        | MO2              | MO1              |
|                          | KRR    | 10W%R            | MA15             | EMA              |

Note: FBAQ denotes The Final Best Ask Quote; FBBQ represents The Final Best Bid Quote

https://doi.org/10.1371/journal.pone.0209922.t006
period, and the proposed model B in the long test period generates a smaller RMSE than other models.

Next, experiments on the TSMC datasets are illustrated in Table 12 and Fig 6. The forecast results in Fig 6 show that the forecast line obviously deviates from the closing price line. From Table 12, the proposed model B generates the smallest RMSE in the short and long test periods.

### Significance test

To test whether proposed model is superior to the KRR-MARS, KRR-SR, KRR-GA-SVR, MARS-SR, MARS-KRR, SR-MARS and SR-KRR models in the stock price forecasting, this study applies the Wilcoxon signed-rank test. We use RMSE to test the significance between the proposed model and the listed models. Tables 13 and 14 present the Z statistic of the two-tailed Wilcoxon sign test between proposed models and the listed models.

From Table 13, the proposed models have a significant difference (p<0.05) compared to other models in the short test period except for KRR-GA-SVR at the 0.05 significant level.
Therefore, we can conclude that the proposed models are significantly better than KRR-MARS, KRR-SR, MARS-SR, MARS-KRR, SR-MARS, and SR-KRR models. However, we can see that there are no significant differences between proposed model A and B in the short testing period from Table 13. In the long testing period, the proposed models have a higher significance compared with the other models at the 0.05 significant level as shown in Table 14. Therefore, we can conclude that proposed model is significantly better than the listed models.

![Fig 2. Results of forecasting short and long test period for Chunghwa Telecom datasets.](https://doi.org/10.1371/journal.pone.0209922.g002)

Table 8. Performance comparisons for Chunghwa Telecom.

| model            | Test period(year) |   |   |
|------------------|-------------------|---|---|
|                  | 1                 | 2 | 3 |
| KRR-SR           | 25.0596           |   |   |
| KRR-MARS         | 25.4612           |   |   |
| KRR-GA-SVR       | 2.5085            |   |   |
| MARS-SR          | 0.1527            |   |   |
| MARS-KRR         | 8.1484            |   |   |
| SR-MARS          | 0.1434            |   |   |
| SR-KRR           | 7.3393            |   |   |
| Proposed model A | 0.0354            | 0.0586 |   |
| Proposed model B | 0.0342            | 0.0608 |   |

Proposed model A: Integrated MARS and GA-SVR model
Proposed model B: Integrated SR and GA-SVR model

[https://doi.org/10.1371/journal.pone.0209922.t008](https://doi.org/10.1371/journal.pone.0209922.t008)
Findings

Based on the experimental results, this study can summarize the findings as follows.

(1) Datasets quality. From Table 15, we find that of the five datasets with different fluctuations, the highest fluctuation range is the Hon Hai stock price. Despite the Hon Hai datasets

Table 9. Performance comparisons for China Steel.

| model         | Test period(year) |
|---------------|-------------------|
|               | 1                 | 3                 |
| KRR-SR        | 25.0596           | 28.4001           |
| KRR-MARS      | 7.6083            | 0.3205            |
| KRR-GA-SVR    | 1.634339328       | 1.01520238        |
| MARS-SR       | 3.7657            | 3.9520            |
| MARS-KRR      | 0.4634            | 0.4533            |
| SR—MARS       | 0.2468            | 0.3250            |
| SR—KRR        | 0.2452            | 0.3250            |
| Proposed model A | 0.2265      | 0.0381            |
| Proposed model B | 0.0370      | 0.3161            |

Proposed model A: Integrated MARS and GA-SVR model. Proposed model B: Integrated SR and GA-SVR model

https://doi.org/10.1371/journal.pone.0209922.t009

Fig 3. Results of forecasting short and long test period for China Steel datasets.

https://doi.org/10.1371/journal.pone.0209922.g003
having the highest fluctuation range, the proposed models can generate smaller RMSE than
the listing models in the short and long test period as shown in Tables 16 and 17. Further, the
China Steel stock price has the smallest fluctuation in the five datasets, and the proposed mod-
els still achieve better performance in both the short and long test periods, as shown in Tables

Table 10. Performance comparisons for Hon Hai.

| model                  | Test period (year) |
|------------------------|--------------------|
|                        | 1                  | 3                  |
| KRR-SR                 | 22.6070            | 76.1505            |
| KRR-MARS               | 17.8506            | 339.1292           |
| KRR-GA-SVR             | 0.4686             | 0.3136             |
| MARS-SR                | 1.9738             | 2.1778             |
| MARS-KRR               | 1.8837             | 2.2246             |
| SR -MARS               | 1.8768             | 2.1328             |
| SR -KRR                | 1.8528             | 2.1303             |
| Proposed model A       | 0.0040             | 0.0087             |
| Proposed model B       | 0.0030             | 0.0914             |

Proposed model A: Integrated MARS and GA-SVR model. Proposed model B: Integrated SR and GA-SVR model

https://doi.org/10.1371/journal.pone.0209922.t010

Fig 4. Results of forecasting short and long test period for Hon Hai datasets.

https://doi.org/10.1371/journal.pone.0209922.g004
16 and 17. Finally, the results show that the two proposed models fit the forecast stock price for investors.

(2) Short and Long test period. The experimental results of the forecasting models in the short and long test periods are listed in Tables 16 and 17, and we find that the accuracy of

| model                  | Test period (year) |
|------------------------|--------------------|
|                        | 1                  | 3                  |
| KRR-SR                 | 27.9581            | 30.3242            |
| KRR-MARS               | 13.7496            | 24.5962            |
| KRR-GA-SVR             | 0.0010             | 1.0663             |
| MARS-SR                | 2.5218             | 0.8007             |
| MARS-KRR               | 0.4663             | 0.7296             |
| SR-MARS                | 0.4842             | 14.5722            |
| SR-KRR                 | 0.4831             | 0.7367             |
| Proposed model A       | 0.2859             | 0.6552             |
| Proposed model B       | 0.1671             | 0.5280             |

Proposed model A: Integrated MARS and GA-SVR model. Proposed model B: Integrated SR and GA-SVR model

https://doi.org/10.1371/journal.pone.0209922.t011

Fig 5. Results of forecasting short and long test period for Cathay Financial datasets.

https://doi.org/10.1371/journal.pone.0209922.g005
proposed models in the short test period is better than in the long test periods. From Fig 7, the figure shows that the stock indexes change dramatically in the long test periods; and the proposed models has better performance in larger price fluctuation. In the short test period, the results show that the proposed model B generates the smallest RMSE in the TSMC, Hon Hai,

### Table 12. Performance comparisons for TSMC.

| Model       | Test period(year) |
|-------------|-------------------|
|             | 1                 | 3                 |
| KRR-SR      | 2.7440            | 12.3018           |
| KRR-MARS    | 9.8224            | 125.9944          |
| KRR-GA-SVR  | 5.0519            | 6.1786            |
| MARS-SR     | 1.2980            | 0.9255            |
| MARS-KRR    | 32.9380           | 29.0172           |
| SR-MARS     | 1.2494            | 1.8027            |
| SR—KRR      | 30.9380           | 22.4461           |
| Proposed model A | 0.8256 | 0.8395 |
| Proposed model B | 0.7292 | 0.7413 |

Proposed model A: Integrated MARS and GA-SVR model. Proposed model B: Integrated SR and GA-SVR model.

https://doi.org/10.1371/journal.pone.0209922.t012

Fig 6. Results of forecasting short and test long period for TSMC datasets.

https://doi.org/10.1371/journal.pone.0209922.g006
China Steel and CHT datasets, except Cathay dataset as Table 16. Because the fluctuation of Cathay price dataset in short period is smaller than other datasets as Fig 7. Therefore, we conclude that the proposed model B (SR-GA-SVR) has better performance than the listing models, especially in larger price fluctuation.

Table 13. Wilcoxon sign test for different models comparison in short period.

| model           | KRR-MARS | KRR-SR | KRR-GA-SVR | MARS-SR | MARS-KRR | SR-MARS | SR-KRR | Proposed model B | Proposed model A |
|-----------------|----------|--------|------------|---------|----------|---------|--------|-----------------|-----------------|
| KRR-MARS        | -        | -0.135 | -2.023**   | -2.023**| -0.674** | -2.023**| -2.023**| -2.023**        | -2.023**        |
| KRR-SR          | -1.214   | -1.753*| -1.753*    | -0.944  | -0.405   | -0.405  | -0.405  | -0.944          | -2.023**        |
| KRR-GA-SVR      | -        | -0.135 | -0.674    | -0.674  | -0.674   | -0.674  | -0.674  | -0.674          | -2.023**        |
| MARS-SR         | -        | -1.753*| -1.753*    | -0.944  | -0.405   | -0.405  | -0.405  | -0.944          | -2.023**        |
| MARS-KRR        | -        | -0.135 | -0.674    | -0.674  | -0.674   | -0.674  | -0.674  | -0.674          | -2.023**        |
| SR-MARS         | -        | -1.753*| -1.753*    | -0.944  | -0.405   | -0.405  | -0.405  | -0.944          | -2.023**        |
| SR-KRR          | -        | -0.135 | -0.674    | -0.674  | -0.674   | -0.674  | -0.674  | -0.674          | -2.023**        |
| Proposed model B| -        | -1.753*| -1.753*    | -0.944  | -0.405   | -0.405  | -0.405  | -0.944          | -2.023**        |

Note: The digital in parentheses is the corresponding p-value; 
*: p<0.1; 
**: p<0.05

https://doi.org/10.1371/journal.pone.0209922.t013

A time-series model for forecasting leading industry stock prices

Table 14. Wilcoxon sign test for different models comparison in long period.

| model           | KRR-MARS | KRR-SR | KRR-GA-SVR | MARS-SR | MARS-KRR | SR-MARS | SR-KRR | Proposed model B | Proposed model A |
|-----------------|----------|--------|------------|---------|----------|---------|--------|-----------------|-----------------|
| KRR-MARS        | -        | -1.753*| -1.753*    | -1.753* | -0.944   | -2.023**| -2.023**| -2.023**        | -2.023**        |
| KRR-SR          | -        | -1.753*| -1.753*    | -0.944  | -1.214   | -0.405  | -0.405  | -0.405          | -2.023**        |
| KRR-GA-SVR      | -        | -0.135 | -0.674    | -0.674  | -0.674   | -0.674  | -0.674  | -0.674          | -2.023**        |
| MARS-SR         | -        | -1.753*| -1.753*    | -0.944  | -1.214   | -0.405  | -0.405  | -0.405          | -2.023**        |
| MARS-KRR        | -        | -0.135 | -0.674    | -0.674  | -0.674   | -0.674  | -0.674  | -0.674          | -2.023**        |
| SR-MARS         | -        | -1.753*| -1.753*    | -0.944  | -1.214   | -0.405  | -0.405  | -0.405          | -2.023**        |
| SR-KRR          | -        | -0.135 | -0.674    | -0.674  | -0.674   | -0.674  | -0.674  | -0.674          | -2.023**        |
| Proposed model B| -        | -1.753*| -1.753*    | -0.944  | -1.214   | -0.405  | -0.405  | -0.405          | -2.023**        |

Note: The digital in parentheses is the corresponding p-value; 
*: p<0.1; 
**: p<0.05

https://doi.org/10.1371/journal.pone.0209922.t014
Similarly, in the long test period, from Table 17, the proposed model A (MARS-GA-SVR) has the smallest RMSE in the Hon Hai, China Steel and CHT datasets. Therefore, we conclude that MARS can also select better features for the proposed model when the stock price range changes dramatically.

(3) Selected feature. For the MARS selected features shown in Table 18, the feature “The Final Best Bid Quote” was chosen four times in five datasets and the forecasting results of proposed model A are better than the results of proposed model B in the long test period. Based

Table 15. The descriptive statistics for all datasets.

|       | TSMC | Cathay | Hon Hai | China Steel | CHT |
|-------|------|--------|---------|-------------|-----|
| Range | 62.8 | 69.75  | 246     | 34.7        | 64  |
| Minimum | 36.8  | 24.05  | 54      | 19.2        | 46  |
| Maximum | 99.6  | 93.8   | 300     | 53.9        | 110 |
| Mean   | 61.9997 | 53.6333 | 142.352 | 31.6591     | 66.9004 |
| Std. Deviation | 11.1591 | 14.83258 | 50.2475 | 6.48304 | 15.10203 |
| Variance | 142.526 | 220.006 | 2524.807 | 42.03 | 228.071 |

Note: Cathay denotes Cathay Financial Holdings

https://doi.org/10.1371/journal.pone.0209922.t015

Table 16. The RMSE of all experiments for short testing period.

|       | TSMC | Cathay | Hon Hai | China Steel | CHT |
|-------|------|--------|---------|-------------|-----|
| KRR-SR | 2.7440 | 27.9581 | 22.6070 | 25.0596     | 25.0596 |
| KRR-MARS | 9.8224 | 13.7496 | 17.8506 | 7.6083      | 25.4612 |
| KRR-GA-SVR | 5.0519 | 0.0010 | 0.4686  | 1.6343      | 2.5085 |
| MARS-SR | 1.2980 | 2.5218  | 1.9738  | 3.7657      | 0.1527 |
| SR—MARS | 1.2494 | 0.4842  | 1.8768  | 0.2468      | 0.1434 |
| MARS-KRR | 32.938 | 0.4663  | 1.8837  | 0.4634      | 8.1484 |
| SR—KRR | 30.938 | 0.4831  | 1.8528  | 0.2452      | 7.3393 |
| Proposed model A | 0.8256 | 0.2859  | 0.0040  | 0.2265      | 0.0354 |
| Proposed model B | 0.7292 | 0.1671  | 0.0030  | 0.0370      | 0.0342 |

Note: Cathay denotes Cathay Financial Holdings

https://doi.org/10.1371/journal.pone.0209922.t016

Similarly, in the long test period, from Table 17, the proposed model A (MARS-GA-SVR) has the smallest RMSE in the Hon Hai, China Steel and CHT datasets. Therefore, we conclude that MARS can also select better features for the proposed model when the stock price range changes dramatically.

(3) Selected feature. For the MARS selected features shown in Table 18, the feature “The Final Best Bid Quote” was chosen four times in five datasets and the forecasting results of proposed model A are better than the results of proposed model B in the long test period. Based

Table 17. The RMSE of all experiments for long testing period.

|       | TSMC | Cathay | Hon Hai | China Steel | CHT |
|-------|------|--------|---------|-------------|-----|
| KRR-SR | 12.3018 | 30.3242 | 76.1505 | 28.4001     | 28.4000 |
| KRR-MARS | 125.9944 | 24.5962 | 339.1292 | 0.3205 | 37.0234 |
| KRR-GA-SVR | 6.1786 | 1.0663  | 0.3136  | 1.0152      | 1.6955 |
| MARS-SR | 0.9255 | 0.8007  | 2.1778  | 3.9520      | 0.1801 |
| SR—MARS | 1.8027 | 14.5722 | 2.1328  | 0.3250      | 0.6356 |
| MARS-KRR | 29.0172 | 0.7296  | 2.2246  | 0.4533      | 27.6530 |
| SR—KRR | 22.4461 | 0.7367  | 2.1303  | 0.3250      | 75.7406 |
| Proposed model A | 0.8395 | 0.6552  | 0.0087  | 0.0381      | 0.0586 |
| Proposed model B | 0.7413 | 0.5280  | 0.0914  | 0.3161      | 0.0608 |

Note: Cathay denotes Cathay Financial Holdings

https://doi.org/10.1371/journal.pone.0209922.t017
on the reasons above, we can confirm that the feature “The Final Best Bid Quote” influences those stock prices forecasting in long testing period.

For the SR selected features as shown in Table 18, the features CDP, MO1 and MO2 are selected three times in five datasets. In addition, the proposed SR-GA-SVR shows with precise

![Fig 7. The closing prices of five companies from 2003 to 2012.](https://doi.org/10.1371/journal.pone.0209922.g007)

Table 18. The selected features for all datasets.

| Method               | Indicator1          | Indicator2          | Indicator3          | Indicator4          |
|----------------------|---------------------|---------------------|---------------------|---------------------|
| CHT                  | MARS                | CDP value           | MO2                 | -                   | -                   |
|                      | SR                  | CDP value           | MO1                 | -                   | -                   |
| China Steel          | MARS                | Daily price change  | The Final Best Bid Quote | The Final Best Ask Quote | TAIEX index |
|                      | SR                  | CDP value           | MO2                 | MO1                 | -                   |
| Hon Hai              | MARS                | The Final Best Bid Quote | ln(ROI)            | Transaction volume  | -                   |
|                      | SR                  | EMA                 | MACD                | ln(ROI)             | -                   |
| Cathay Financial Holdings | MARS              | ROI                 | The Final Best Bid Quote | CDP value         | -                   |
|                      | SR                  | 5BIAS               | MA5                 | MO2                 | -                   |
| TSMC                 | MARS                | MO2                 | The Final Best Ask Quote | The Final Best Bid Quote | -                   |
|                      | SR                  | CDP value           | MO2                 | MO1                 | -                   |

https://doi.org/10.1371/journal.pone.0209922.t018
accuracy in the short testing period. Therefore, we find that the CDP, MO1 and MO2 have a
great impact on forecasting stock prices for the short test period.

(4) Investor suggestion. After verifying the proposed models, this study can provide
some suggestions to investors as references in the following:

- From Tables 16 and 17, the short test period forecasting is recommended, because it will be
  more accuracy than the long test period forecasting for investment stock.

- In the short period forecasting, we suggest using the proposed model B because it is more
  accuracy than proposed model A (see Table 16). Regarding key features as shown in
  Table 18, we suggest the investors consider the three key features: CDP, MO1 and MO2.

- In the long period forecasting, from Table 17, the proposed model A is recommended
  because it is more accuracy than proposed model B in the long testing period. From
  Table 18, the feature “The Final Best Bid Quote” should be considered as input variables in
  the long period forecasting.

Conclusion

This study has proposed a new time-series model, which considers multifactor and reasonable
selected key features into the GA-SVR model. The results show that proposed models can
improve forecasting accuracy. Furthermore, the proposed models outperform the listed mod-
els in RMSE for Chunghwa Telecom, China Steel, Hon Hai, Cathay Financial Holdings and
Taiwan Semiconductor Manufacturing Company datasets. In addition, from the findings and
discussions, the proposed SR-GA-SVR outperforms the listing models in the short testing
period, except Cathay Financial Holdings. Moreover, in the long testing period, the MARS-
GA-SVR also has better performance. We find that the proposed model B almost has better
performance than the listing models, especially in larger price fluctuation. i.e., the proposed
model is more fit the dataset of larger price fluctuation. Finally, the research results can pro-
vide some suggestion to investors as references.

In future work, several issues from this study can be extended as follows:

1. Consider other features to train the model, such as company news, or government policies.
2. Apply the model to different application fields, such as electric loads and environmental
   pollution forecasting.
3. Employ other methods to improve proposed model, such as feature lags.

Supporting information

S1 Dataset. Five lead-industry-index.
(ZIP)

Author Contributions

Conceptualization: Ching-Hsue Cheng.
Data curation: Ching-Hsue Cheng, Huei-Yuan Shiu.
Formal analysis: Ching-Hsue Cheng.
Methodology: Ming-Chi Tsai, Ching-Hsue Cheng.
Project administration: Ming-Chi Tsai.
Resources: Meei-Ing Tsai.
Software: Huei-Yuan Shiu.
Supervision: Ching-Hsue Cheng.
Validation: Meei-Ing Tsai.
Writing – review & editing: Ming-Chi Tsai.

References
1. Box GEP, Jenkins GM. Time series analysis: forecasting and control. San Francisco: Holden-Day; 1970.
2. Ediger VS, Akar S. ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy. 2007; 35: 1701–1708.
3. Cheng CH, Chen TL, Wei LY. A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. Information Sciences. 2010; 180: 1610–1629.
4. Yu THK, Huarng KH. A bivariate fuzzy time series model to forecast the TAIEX. Expert Systems with Applications. 2008; 34: 2945–2952.
5. Tsai CF, Lin YC, Yen DC, Chen YM. Predicting stock returns by classifier ensembles. Applied Soft Computing. 2011; 11: 2452–2459.
6. Pan Y, Xiao Z, Wang X, Yang D. A multiple support vector machine approach to stock index forecasting with mixed frequency sampling. Knowledge-Based Systems. 2017; 122: 90–102.
7. Su CH, Cheng CH. A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock. Neurocomputing. 2016; 205: 264–273.
8. Wang J, Hou R, Wang C, Shen L. Improved v-support vector regression model based on variable selection and brain storm optimization for stock price forecasting. Applied Soft Computing. 2016; 49: 164–176.
9. Chang PC, Liu CH. A TSK type fuzzy rule based system for stock price prediction. Expert Systems with Applications. 2008; 34: 135–144.
10. Liu CF, Yeh CY, Lee SJ. Application of type-2 neuro-fuzzy modeling in stock price prediction. Applied Soft Computing. 2012; 12: 1348–1358.
11. Oh SK, Pedrycz W, Park HS. Genetically optimized fuzzy polynomial neural networks. IEEE Transactions on Fuzzy Systems. 2006; 14: 125–144.
12. Zarandi MHF, Rezaee B, Turksen IB, Neshat E. A type-2 fuzzy rule-based expert system model for stock price analysis. Expert Systems with Applications. 2009; 36: 139–154.
13. Cao LJ, Tay FEH. Support vector machine with adaptive parameters in financial time series forecasting. IEEE Transactions on Neural Networks. 2003; 14: 1506–1518. https://doi.org/10.1109/TNN.2003.820556 PMID: 18244595
14. Gestel TV, Suykens JAK, Baestaens DE, Lambrechts A, Lanckriet G, Vandaele B, et al. Financial time series prediction using least squares support vector machines within the evidence framework. IEEE Transactions on Neural Networks. 2001; 12: 809–821. https://doi.org/10.1109/72.935093 PMID: 18249915
15. Gavrishchaka VV, Banerjee S. Support vector machine as an efficient framework for stock market volatility forecasting. Computational Management Science. 2006; 3: 147–160.
16. Yeh CY, Huang CW, Lee SJ. A multiple-kernel support vector regression approach for stock market price forecasting. Expert Systems with Applications. 2011; 38: 2177–2186.
17. Vapnik VN. An overview of statistical learning theory. IEEE Transactions on Neural Networks. 1999; 10: 988–999. https://doi.org/10.1109/72.788640 PMID: 18252602
18. Vapnik VN. The nature of statistical learning theory. New York: Springer; 2000.
19. Kao LJ, Chiu CC, Lu CJ, Yang JL. Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. Neurocomputing. 2013; 99: 534–542.
20. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.
21. Kazem A, Sharifi E, Hussain FK, Saberi M, Hussain OK. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Applied Soft Computing. 2013; 13: 947–958.
22. Kabir MMJ, Xu S, Kang BH, Zhao Z. A new multiple seeds based genetic algorithm for discovering a set of interesting Boolean association rules. Expert Systems with Applications. 2017; 74: 55–69.

23. Su CH, Cheng CH, Tsai WL. Fuzzy time series model based on fitting function for forecasting TAIEX index. International Journal of Hybrid Information Technology. 2013; 6: 111–122.

24. Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. Machine Learning. 2002; 46: 131–159.

25. Duan K, Keerthi SS, Poo AN. Evaluation of simple performance measures for tuning SVM hyperparameters. Neurocomputing. 2003; 51: 41–59.

26. Kwok JTY. The evidence framework applied to support vector machines. IEEE Transactions on Neural Networks. 2000; 11: 1162–1173. https://doi.org/10.1109/72.870047 PMID: 18249842

27. Allen F, Karjalainen R. Using genetic algorithms to find technical trading rules. Journal of Financial Economics. 1999; 51: 245–271.

28. Leigh W, Russell R, Ragusa JM. Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support. Decision Support Systems. 2002; 32: 361–377.

29. Gorgulho A, Neves R, Horta N. Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition. Expert Systems with Applications. 2011; 38: 14072–14085.

30. Pring MJ. Technical Analysis Explained. New York: McGraw-Hill; 1991.

31. Chang PC, Liao TW, Lin JJ, Fan CY. A dynamic threshold decision system for stock trading signal detection. Applied Soft Computing. 2011; 11: 3998–4010.

32. Park Ji, Lee DJ, Song CK, Chun MG. TAIEX and KOSPI 200 forecasting based on two-factors high-order fuzzy time series and particle swarm optimization. Expert Systems with Applications. 2010; 37: 959–967.

33. Tanaka-Yamawaki M, Tokuoka, S. Adaptive use of technical indicators for the prediction of intra-day stock prices. Physica A. 2007; 383: 125–133.

34. Lin TN. Using AdaBoost for Taiwan stock index future intra-day trading system. M. Sc. Thesis, National Taiwan University. 2008. https://www.csie.ntu.edu.tw/~lyuu/theses/thesis_R9594016.pdf

35. Hassapis C, Kalyvitis S. Investigating the links between growth and real stock price changes with empirical evidence from the G-7 economies. The Quarterly Review of Economics and Finance. 2002; 42: 543–575.

36. Basher SA, Haug AA, Sadorsky P. Oil prices, exchange rates and emerging stock markets. Energy Economics. 2012; 34: 227–240.

37. WikipediA. Interest rate. 2017. http://en.wikipedia.org/wiki/Interest_rate.

38. Bessembinder H. Quote-based competition and trade execution costs in NYSE-listed stocks. Journal of Financial Economics. 2003; 70: 385–422.

39. Bagella M, Becchetti L, Adriani F. Observed and “fundamental” price-earning ratios: A comparative analysis of high-tech stock evaluation in the US and in Europe. Journal of International Money and Finance. 2005; 24: 549–581.

40. Pontiff J, Schall LD. Book-to-market ratios as predictors of market returns. Journal of Financial Economics. 1998; 49: 141–160.

41. Chen S. The predictability of aggregate Japanese stock returns: Implications of dividend yield. International Review of Economics and Finance. 2012; 22: 284–304.

42. Pan D, Wiersma G, Williams L, Fong YS. More than a number: unexpected benefits of return on investment analysis. The Journal of Academic Librarianship. 2013; 39: 566–572.

43. Politi M, Millot N, Chakraborti A. The near-extreme density of intraday log-returns. Physica A. 2012; 391: 147–155.

44. Business next, not to afraid low interest rates! 8 strategies to preserve capital. 2002. http://www.bnnext. com.tw/article/view/id/7121.

45. Friedman JH. Multivariate adaptive regression splines. The Annals of Statistics. 1991; 19: 1–67.

46. Holland JH. Adaptation in Natural and Artificial Systems. Ann Arbor: University of Michigan Press; 1975.

47. Goldberg DE. Genetic Algorithms in Search Optimization and Machine Learning. Reading: Addison-Wesley; 1989.
50. Huang SC, Chuang PJ, Wu CF, Lai HJ. Chaos-based support vector regressions for exchange rate forecasting. Expert Systems with Applications. 2010; 37: 8590–8598.

51. Chang PC, Liu CH, Fan CY. Data clustering and fuzzy neural network for sales forecasting: A case study in printed circuit board industry. Knowledge-Based Systems. 2009; 22: 334–355.

52. Cheng CH, Shiu HY. A novel GA-SVR time series model based on selected indicators method for forecasting stock price. IEEE Proceedings-ISEEE 2014. Sapporo, Japan. 2014.

53. Asadi S, Hadavandi E, Mehranpazir F, Nakhostin MM. Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. Knowledge-Based Systems. 2012; 35: 245–258.

54. Adoko AC, Jiao YY, Wu L, Wang H, Wang ZH. Predicting tunnel convergence using Multivariate Adaptive Regression Spline and Artificial Neural Network. Tunnelling and Underground Space Technology. 2013; 38: 368–376.