INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN ACCOUNTING, FINANCE AND MANAGEMENT SCIENCES
www.hrmars.com
ISSN: 2225-8329

A Hybrid Stock Forecasting Model based on EMD -SVR and Technical Indicators Feature Selection

Liang-Ying Wei, Chih-Hung Tsai

To Link this Article: http://dx.doi.org/10.6007/IJARAFMS/v12-i1/12184 DOI:10.6007/IJARAFMS /v12-i1/12184

Received: *11* December 2021*, **Revised:*** *13* January 2022*, **Accepted:*** *23* January 2022

Published Online: *16* February 2022

In-Text Citation: (Wei & Tsai, 2022)

To Cite this Article: Wei, L.-Y., & Tsai, C.-H. (2022). A Hybrid Stock Forecasting Model based on EMD -SVR and Technical Indicators Feature Selection. *International Journal of Academic Research in Accounting Finance and Management Sciences*, *12*(1), 184–200.

Copyright: © 2022 The Author(s)

Published by Human Resource Management Academic Research Society (www.hrmars.com)

This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at: http://creativecommons.org/licences/by/4.0/legalcode

Vol. 12, No. 1, 2022, Pg. 184 - 200

http://hrmars.com/index.php/pages/detail/IJARAFMS | **JOURNAL HOMEPAGE**

Full Terms & Conditions of access and use can be found at

http://hrmars.com/index.php/pages/detail/publication-ethics

INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN ACCOUNTING, FINANCE AND MANAGEMENT SCIENCES
www.hrmars.com
ISSN: 2225-8329

A Hybrid Stock Forecasting Model based on EMD -SVR and Technical Indicators Feature Selection

Liang-Ying Wei, Chih-Hung Tsai
Department of Information Management, Yuanpei University of Medical Technology, Hsin-Chu, Taiwan
Email: lywei@mail.ypu.edu.tw, Corresponding author

Abstract

Stock forecasting is an important and interesting topic in time series forecasting. Accurate stock price is regard as a challenging task of the financial time forecasting process. However, there are three major drawbacks in stock market by traditional time-series model (1) some models can not be applied to the datasets that do not follow the statistical assumptions; and (2) most time-series models which use stock data with many noises involutedly (caused by changes in market conditions and environments) would reduce the forecasting performance; (3) Subjective selected technical indicators as input variable by personal experience would lead to lower forecasting accuracy. For solving above problems, this paper proposes a hybrid time-series support vector regression (SVR) model based on Empirical mode decomposition (EMD) and technical indicators feature selection to forecast stock price for Taiwan stock exchange capitalization weighted stock index (TAIEX). In verification, this paper employs nine year period of TAIEX stock index, from 1997 to 2005, as experimental datasets, and the root mean square error (RMSE) as evaluation criterion. The experimental results indicate that the proposed model is superior to the listing methods in terms of root mean squared error.

Keywords: Support Vector Regression (SVR), Empirical Mode Decomposition (EMD), Feature Selection.

Introduction

If the forecasting method can accurately predict the changes in stock prices, it will bring great profits to investors. Therefore many scholars and stock analysts have devoted a lot of vigor to develop better forecasting models, and technical indicators are also widely used by stock investors in stock forecasting. However, most stock analysts use their own personal experience subjectively to choose technical indicators. Using the inappropriate selection of technical indicators to predict stock prices will reduce the accuracy of stock price forecasts.

In recent years, many time series methods have been proposed and applied to the issue of financial forecasting (Bollerslev, 1986; Box and Jenkins, 1976; Engle, 1982; Huarng, 2001; Huarng and Yu, 2006; Song, and Chissom, 1993). Traditional time series methods needs to meet statistical assumptions, so not all methods can be applied to all data sets, especially some statistical methods that need to meet linear assumptions cannot be used in nonlinear data set forecasting. In addition, most traditional time series methods use a single variable (yesterday's stock price) to predict today's stock price, but a single variable will contain a lot of interference from market environment or conditions, which will reduce the accuracy of the stock price forecasting. Empirical mode decomposition (EMD) (Huang et al., 1998) is a proper method to deal with nonlinear signal analysis (such as stock data) or other related fields (Huang et al., 1999, Vincent et al., 1999, Yu et al., 2005). The complex signal can be decomposed into several intrinsic mode functions (IMFs) by using the empirical mode decomposition method. The correlation between these intrinsic mode functions and the prediction data is higher. Applying IMFs to forecasting models can increase the accuracy of prediction.

In recent years, artificial intelligence techniques have been used in prediction applications related to different fields (Chen et al., 2008, Chen and Chung, 2006; Huarng et al., 2007; Kim, 2003; Kim and Han, 2000; Kimoto et al., 1990; Nikolopoulos, and Fellrath 1994; Roh, 2007; Takagi and Sugeno, 1983; Thawornwong and Enke, 2004; Vapnik, 1995; Vincent et al., 1999; Song and Chissom, 1993). Support vector regression (SVR) is proposed by Huang et al. (2005). The goal of SVR is to find the optimal hyperplane in space and accurately predict the distribution of data. Previous studies have shown that SVR models have been successful in solving related prediction problems in many different research fields (Kim, 2003; Pai & Lin, 2005). In addition, time series forecasting methods have been applied to forecasting business problems with remarkable achievement. Various time series methods have also been proposed, among which the autoregressive (AR) proposed by Box and Jenkins (1976) is one of the most popular and important methods.

Based on the issues mentioned above, this study proposes a model based on the SVR algorithm, combining EMD, AR, and technical index feature selection, and uses EMD to decompose the input variables of the AR. Then proposed model takes the decomposed IMFs and the feature selection technical indicators as input variables and use the SVR forecasting method to forecast the stock trend. The processes of proposed model in this research are as follows:

(1) Establish an AR model: use ordinary least squares estimator (OLS) to establish an AR model.

(2) Use EMD to decompose the input variables of the AR model: Use the empirical modal decomposition method for the input variables of the AR model. For example, if the AR model of Taiwan stock after statistical test is AR(1), that is to select the last period of the stock. The stock price of the stock market can then use the empirical mode decomposition method to decompose the complex signal into several intrinsic mode functions, which are more relevant to the forecast data.

(3) Feature selection of technical indicators: Use stepwise regression to select important technical indicators.

(4) Establish a prediction model: Proposed model takes the decomposed IMFs and the selected technical indicators as input variables, and uses SVR to build a prediction model, and utilizes the training data to train the parameters of the SVR model to obtain the best prediction model.

Related Works

This section discusses related research, including the introduction of different stock forecasting methods, empirical mode decomposition methods and SVR methods, and stepwise regression methods.

Different Forecast Model in the Stock Market

The stock market is a very exciting and challenging financial activity. The market environment changes dramatically every minute and every second. Gain or lose money can be determined at the click of a finger. In addition, many researchers have proposed many methods to apply to the problem of stock forecasting. Huarng et al (2007) used the stock index of NDSDAQ (the largest electronic stock market in the United States) and the stock index of Dow Jones (Dow Jones Industrial Average) to predict the Taiwan stock market.

Time series models have been used to deal with economic forecasting problems, such as the forecasting of stock indicators. Various forecasting models have been proposed. Engle (1982) proposed the ARCH (Autoregressive Conditional Heteroscedasticity) model, which has been applied in analysis of many financial problems, in addition to methods such as GARCH (Bollerslev, 1986) have also been proposed. Box and Jenkins (1976) proposed the autoregressive moving average (ARMA) model, this method combines the process of moving average and the equation of linear difference to obtain the model of the autoregressive moving average. In the past few years, many researchers have applied data mining technology to financial analysis. Huarng and Yu (2006) applied back propagation neural network to establish fuzzy association rules on time series, and forecasted stock price by proposed model. Kinoto et al. (1990) used neural networks to develop a stock market forecasting system, and Nikolopoulos and Fellrath (1994) combined genetic algorithms and neural networks to develop a hybrid expert system capable of providing investment decision-making suggestions for investors. Kim and Han (2000) applied the concept of genetic algorithm to feature discretization and determine the weight of neural network to predict stocks and Roh (2007) integrate neural network and time series to predict volatility of stock price. Thawornwong and Enke (2004) proposed a redeveloped neural network model to predict stock returns, and Kim (2003) applied SVM (Support vector machine) to the prediction of stock indicators.

Empirical Mode Decomposition

The empirical mode decomposition method is proposed by Huang et al (1998), which is an adaptive time series decomposition technology. It uses the Hilbert-Huang transform (HHT) and can be used for nonlinear and non-stationary time series data. The source data is decomposed into several intrinsic mode functions (IMFs) through empirical mode decomposition (EMD). EMD mainly decomposes the time series into IMFs with inherent adaptability. Different data have different IMFs due to different time series. Decomposition process and the relevant steps of EMD are as follows:

First, EMD finds the local maxima and minima of the original data, and connects the local maxima and minima with a cubic spline to obtain the upper envelope and lower envelope, and then averages the upper and lower envelopes to obtain the mean envelope. Calculation formula is as follows

$$m_1(t) = U(t) + L(t)/2 \qquad (1)$$

EMD subtracts the mean $m_1$from the raw data $X(t)$to get the components$h_1(t)$

$$h_1(t) = X(t) - m_1(t) \qquad (2)$$

If generated $h_1(t)$ is symmetric, and all local maxima are positive and all local minima are negative, the result is IMF, otherwise it is not, and the above steps need to be repeated until the extracted signal is IMF , in this case $h_1(t)$ will be treated as original and repeated.

$$h_{11}(t) = h_1(t) + m_{11}(t) \qquad (3)$$

If the function $h_{11}(t)$still does not meet the requirements of the IMF, EMD repeats the above steps k times until the obtained function satisfies a certain acceptable tolerance. After the above steps, the IMF of the time series is obtained, and its value is $c_1 = h_{1k}(t)$. Then EMD subtracts the first IMF from the original data to get the difference r_1.

$$r_1(t) = X(t) + c_1(t) \tag{4}$$

r_1 is a residual value, EMD takes the residual value $r_1(t)$ as the original data, and then operates the above steps to obtain multiple IMF c_1 . When the final residual value becomes a monotonic function, and the IMF cannot be analyzed any more, the decomposition process of the EMD is finished. The basis of EMD is decomposed from the original signal, so the method is intuitive, direct, a posteriori and adaptable. The above steps can be organized into the formula (5) (6).

$$X(t) = \sum_{i=1}^{n} c_i(t) + r_n \tag{5}$$

$$r_{i-1}(t) - c_1(t) = r_1(t) \tag{6}$$

Stepwise Regression

Stepwise regression is a form of multilinear regression. It can add or remove variables based on the evaluation coefficient of the F test. Stepwise regression combines forward and backward processes. If the F value of the variable is greater than the predicted threshold value, the variable will be added. This variable will be removed if its F value is less than this threshold.

Proposed Model

This research constructs a hybrid SVR forecasting system based on the empirical mode decomposition method and technical index feature selection. The developed system and results will improve the investor's return on investment, and the generated forecasting criteria can provide references for stock market analysts. Research structure of this study is shown in Figure 1.

The proposed model is based on the prediction model of SVR. SVR is a machine learning method and a nonlinear prediction model, it can deal with nonlinear data, which can overcome the limitations of statistical methods (data must obey certain mathematical distribution), which can overcome the limitations of statistical methods (data must obey certain mathematical distributions). Traditional time series methods usually use a single variable (yesterday's stock price) to predict today's stock price, however, yesterday's stock price contains a lot of interference from the market environment or conditions. Therefore, traditional models would get lower prediction accuracy. In order to improve the forecasting ability and overcome the drawback of traditional time series methods, proposed model uses the empirical mode decomposition method to decompose the past stock price into several IMFs, which are more related to the prediction data. Then proposed model combines AR model and technical indicator feature selection to improve the accuracy of prediction.

This research uses stepwise regression to select important technical indicators as input features. The method of regression analysis is simple and fast, and it can be very objective to select important technical indicators, so as to avoid investors from using their own experience and opinions to select technical indicators. Subjective selection of technical indicators may select inappropriate technical indicators and thus affect the accuracy of stock forecasts.

The structure of this research is shown in Figure 1. First, this study collects experiment data, and then the AR model is established by the least squares method, and the input variables of the AR model after EMD decomposition are used to convert it into several IMFs. Then proposed model uses stepwise regression select important technical indicators, and utilizes the decomposed IMFs and the selected technical indicators as input variables of the prediction model. Finally this study uses SVR algorithm to predict stock price. The steps of proposed model are as follows:

Step 1: Data collection

This study collects annual stock market data, and the first 10 months (January to October) of each year's data will be used as training data, and the last two months (November, December) will be used as test data set.

Step 2: Use ordinary least squares estimator (OLS) to build AR model

The AR model is established by using the least squares regression estimation method. The purpose is to detect the correlation between the current stock index and the last period or the last two periods or the last three periods in the past.

Step 3: Use EMD to decompose the input feature of the AR model

Proposed model selects input variables generated by AR model in step 2, and uses the EMD to decompose the complex signal into several IMFs, which are related to the prediction data.

Step 4: Use stepwise regression to select important technical indicators

First, this study collects technical indicators that may be related to stock forecasting through relevant literature, and then utilizes stepwise regression to perform feature selection. The selected technical indicators are more relevant to Taiwan stock price.

Step 5: Train parameters of SVR prediction model

In this step, this study uses SVR and ε -insensitive loss function (ε -SVR) to build a prediction model. The IMF decomposed in step 3 and the technical indicators selected in step 4 will be regarded as SVR's input features. The input features would be used to predict the stock index of *t+1* period. This study will set the radial basis function as the type of the kernel function. This study set ε=0.001 as the tolerance that would be the stopping condition for training.

Step 6: Use the established prediction model to predict test dataset

In step 5, the parameters of the SVR will be determined when the training stop condition is reached. The established training model will be used to predict test data set, and the stock index of period t+1 will be generated.

Step 7: Calculate RMSE and compare with the listing models

In this study, Root Mean Square Error (RMSE) is taken as evaluation criterion to compare the listing models. The process for calculating RMSE can be expressed by equation (7).

$$\text{RMSE}=\sqrt{\sum_{t=1}^{n}|actual(t+1)-forecast(t+1)|^2/_{n}} \qquad (7)$$

Where actual (t+1) denotes the real TAIEX value, forecast (t+1) denotes the predicting TAIEX value and *n* is the number of data.

Vol. 1 2 , No. 1, 2022, E-ISSN: 2225-8329 © 2022 HRMARS

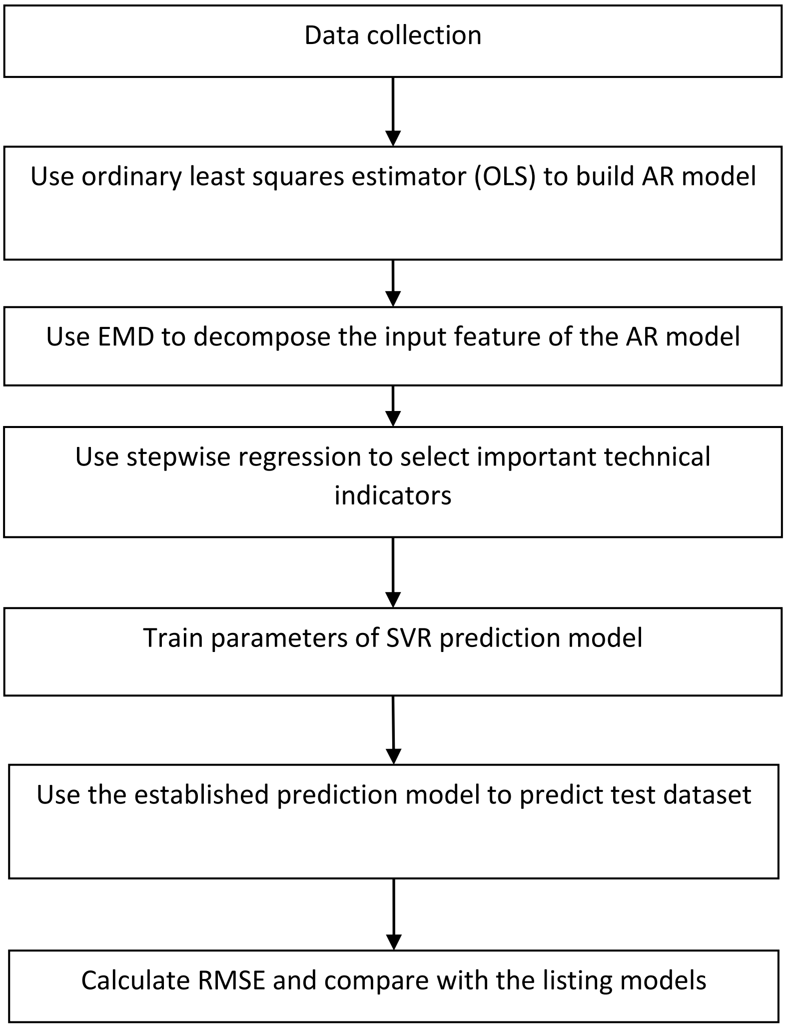

Fig. 1. Flowchart of proposed procedure

Experiments and Comparisons

In this section, the actual data set will be collected, and an actual case will be used to demonstrate our proposed prediction model. The experimental results will be verified with the experimental results of other models to demonstrate the pros and cons of the predictive ability of our proposed prediction model.

Practical Case Study

This section will actually show the various steps of prediction model for predicting the Taiwan stock market index, such as data conversion, AR model establishment, EMD decomposition of AR model variables, use stepwise regression for feature selection, and training of the SVR prediction model. The content of each step is shown as follows:

Step 1: Data collection

This step collected stock data of Taiwan stocks for 9 years from 1997 to 2005, including 4 daily basic variables (daily highest price, daily lowest price, opening price, closing price) to calculate technical indicators. The annual stock data is defined as a subset, the first 10 months of each subset (January to October) are training data, and the remaining two months (November-December) are used as test data.

Step 2: Use ordinary least squares estimator (OLS) to build AR model

The research results of Chang et al. (2010) show that price patterns in the Taiwan stock exchange are short term. Therefore, the lag period from one to five are evaluated to determine which number of time lag is the most fitting for the experimental dataset. The least square method is utilized to build the model, and then five variables close_price(t-1), close_price(t-2),…,close_price(t-5), are selected to estimate and test their significance. If the p-value is less than the significance level given at 0.05, then reject the null hypothesis. Finally, the estimated TAIEX model is obtained, and the lag periods of TAIEX can be determined.

To demonstrate the proposed model, a 1-year period of TAIEX (Year 1997) is employed as the experimental dataset for the prosed algorithm. This study uses E-Views software package to fit the AR model for different close price orders of TAIEX to be estimated and tested. In the statistical test, the p-values is set to less than the significance level 0.05(95% confidence interval); then the testing variable is significant. Fig. 2 shows that the p-value=0.0000 of close-price (t-1) is less than the significance level of 0.05 among five variables. Therefore, the order of TAIEX is 1, and proposed model apply the AR (1) method to the proposed model for forecasting Taiwan stock.

Dependent Variable: CLOSE_PRICE
Method: Least Squares
Date: 04/21/08 Time: 16:56
Sample (adjusted): 6 238
Included observations: 233 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	118.3171	83.29947	1.420382	0.1569
CLOSE_PRICE(-1)	1.004497	0.066550	15.09386	0.0000
CLOSE_PRICE(-2)	0.038937	0.094724	0.411057	0.6814
CLOSE_PRICE(-3)	-0.064165	0.095438	-0.672327	0.5021
CLOSE_PRICE(-4)	-0.101954	0.096612	-1.055295	0.2924
CLOSE_PRICE(-5)	0.109017	0.068456	1.592510	0.1127

R-squared	0.978643	Mean dependent var	8536.073
Adjusted R-squared	0.978173	S.D. dependent var	817.4751
S.E. of regression	120.7747	Akaike info criterion	12.45114
Sum squared resid	3311143.	Schwarz criterion	12.54001
Log likelihood	-1444.558	F-statistic	2080.363
Durbin-Watson stat	1.984448	Prob(F-statistic)	0.000000

Fig. 2. Testing the lag period of TAIEX in 1997

Step 3: Use EMD to decompose the input feature of the AR

The input variables of the AR (1) model established in step 2 are decomposed into 6 IMFs and 1 residual by EMD. These 7 generated variables (IMFs) will be used as SVR input variables to predict the TAIEX stock index.

Step 4: Use stepwise regression to select important technical indicators

First, this study collects technical indicators that may be related to stock forecasting by referring relevant literature (Cheng et al., (2010)). Then this step chooses moving average (MA), stochastic %K (%K), stochastic %D (%D), relative strength index (RSI), psychology line (PSY), Williams' percent range (%R), and accumulative ratio (AR) as 7 selected technical indicators. Then this step sets the cumulative days as 5, and calculates technical indicators: MA-5, %K-5, %D-5, RSI-5, PSY-5, %R-5, and AR-5. Table 1 shows some basic quantity data of TAIEX, Table 2 shows some data converted from basic quantity data to technical indicators, and then uses stepwise regression to perform feature selection. Table 3 shows the 4 selected features by stepwise regression.

Table 1. The partial four fundamental quantities of TAIEX in 2000 year

Date	Opening price	Highest price	Lowest price	Closing price
2000/01/04	8,644.91	8,803.61	8,642.50	8,756.55
2000/01/05	8,690.60	8,867.68	8,668.02	8,849.87
2000/01/06	8,876.59	9,023.99	8,863.91	8,922.03
⋮	⋮	⋮	⋮	⋮
2000/10/27	5,991.83	6,003.38	5,805.17	5,805.17
2000/10/30	5,644.26	5,666.96	5,615.90	5,659.08
2000/10/31	5,530.80	5,626.03	5,502.67	5,544.18

Table 2. The partial instances of technical indicators data

Date	MA-5(t)	RSI-5(t)	K-5(t)	D-5(t)	R-5(t)	PSY-5(t)	AR-5(t)	TAIEX (t+1)
2000/01/04	8384.68	100	96.46	88.16	0	1	3.72	8,849.87
2000/01/05	8537.95	100	97.64	91.32	0	1	4.81	8,922.03
2000/01/06	8678.47	100	98.42	93.68	0	1	4.86	8,845.47
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2000/10/27	5874.07	65.98	66.88	67.65	63.76	0.6	1.30	5,659.08
2000/10/30	5869.70	48.45	44.58	59.96	100	0.4	1.46	5,544.18
2000/10/31	5794.81	17.98	29.72	49.88	100	0.2	1.13	5,425.02

Table 3. The selected features by stepwise regression

Method	Selected features
Stepwise regression	MA-5, RSI-5, %R-5, AR-5

Step 5: Train parameters of SVR prediction model

The IMFs generated in step 3 and the technical indicators selected in step 4 will be used as the input variables of SVR to predict the stock indicators of the *t+1* period, and the training data will be used to train the SVR model, the kernel function is set as the radial basis function. This step set ε = 0.001 as the tolerance value for the training stop condition.

Step 6: Use the established prediction model to predict test dataset

When the stopping condition of the fifth step is reached, the training model of SVR is established, and then this step uses this trained model to predict the TAIEX (t+1) index of the test set data.

Step 7: Calculate RMSE and compare with the listing models

This step calculates the prediction error of the test data set by using formula (7), and uses RMSE as the evaluation criterion to compare with the listing models.

Model Validation

In order to verify the effectiveness of our proposed forecasting method, from 1997 to 2005, the 9-year TAIEX stock index was selected as the experimental data, each year was a subset, and the first 10 months of each subset were Training data, the last two months are test data. In addition, this study compares the forecasting performance of the proposed method with other time series forecasting methods, Chen's method (Chen, 1996), Yu's method (Yu, 2005), AR(1) (Engle, 1982) and SVR (Vapnik, 1995).

In addition, in order to compare the performance generated by stepwise regression feature selection, this study compares the proposed method (4 selected technical indicators) with the model of AR(1)+EMD+7 technical indicators+SVR (without stepwise regression analysis for feature selection) for experimental verification to understand the utility of stepwise regression feature selection. The experimental results are shown in Table 4. From the experimental results, it can be seen that the method proposed in this study is superior to the other prediction methods listed. In addition, the proposed method (The experimental results of using stepwise regression for feature selection) and not using stepwise regression for feature selection, the results show that except for the experimental results in 2000, the experimental results of other years show that the prediction method using stepwise regression for feature selection is better than that without stepwise regression for feature selection.

Table 4. RMSE experimental results by using different prediction methods in the test dataset

Forecasting model	**Year**								
	1997	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Chen'model	154	134	120	176	140	101	74	83	66
Yu'model	165	164	145	191	167	75	66	79	69
AR(1)	141*	114	102	130	115	66	54	55	54
SVR	179	126	113	136	114	66	59	53	55
AR(1) + EMD + 7 technical indicators **+ SVR** (without using stepwise regression for feature selection, 7 technical indicators)	153	84	70	89*	59	49	37	34	33
Proposed model (using stepwise regression for feature selection**,** 4 technical indicators**)**	141*	80*	69*	96	57*	48*	35*	30*	31*

* Indicate the prediction model with the best forecasting performance among the 6 models

In addition, this study uses nonparametric statistical method, Friedman test (Friedman, 1937) to further verify whether the proposed prediction method is significantly better than other prediction methods. This study uses the data set in Table 4 to test the null hypothesis H_0: The prediction performance is the same, and the results show that (p=0.000) The results reject H_0, which means that the prediction ability of different prediction methods is significantly different (as shown in Table 5). Table 6 shows mean rank values generated by Friedman test and the mean rank of the proposed prediction model is 1.17, and the result is superior to other prediction models. Based on the test results obtained in Tables 5 and 6, the result show that proposed model significantly outperform listing model.

Table 5. Result of Friedman test

parameter	
n	9
Chi-Square	39.425
df	5
Asymp. Sig.	.000

Note: n is the number of data points; df represents degrees of freedom

Table 6 . Mean rank generated by Friedman test

Models	Mean Rank
Chen	5.22
Yu	5.56
AR(1)	3.11
SVR	3.94
ANFIS	2.00
Proposed model	1.17

Note: Smaller mean rank indicates higher predictive performance

Discussion

In this section, the fluctuation situation of the experimental data is drawn, and further related discussions are made. The experimental data is collected from the data set of TAIEX from 1997 to 2005. In this study, the data set of these 9 years is drawn in Figure 3-11.

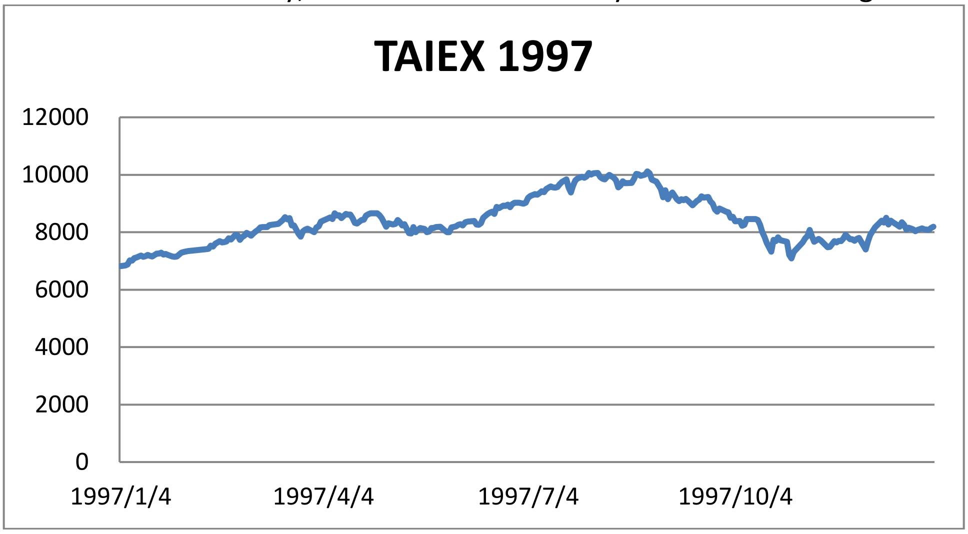

Fig. 3. The actual stock index for TAIEX in 1997

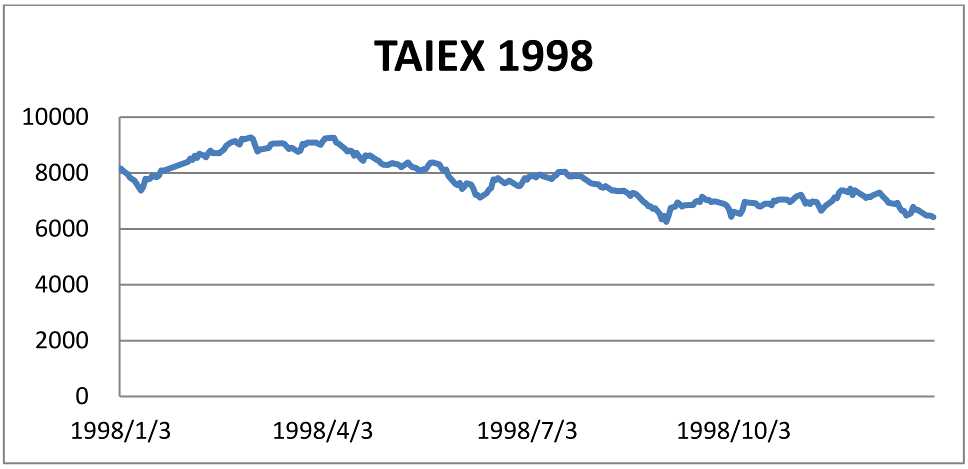

Fig. 4. The actual stock index for TAIEX in 1998

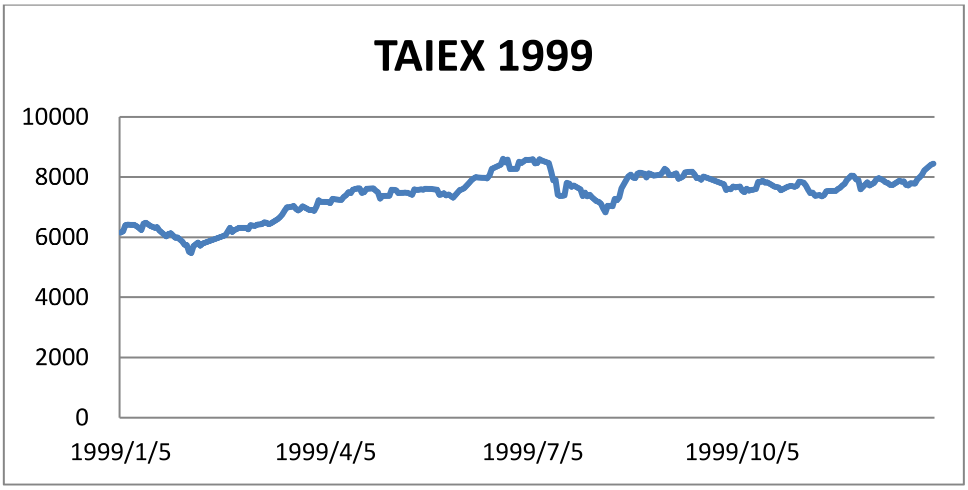

Fig. 5. The actual stock index for TAIEX in 1999

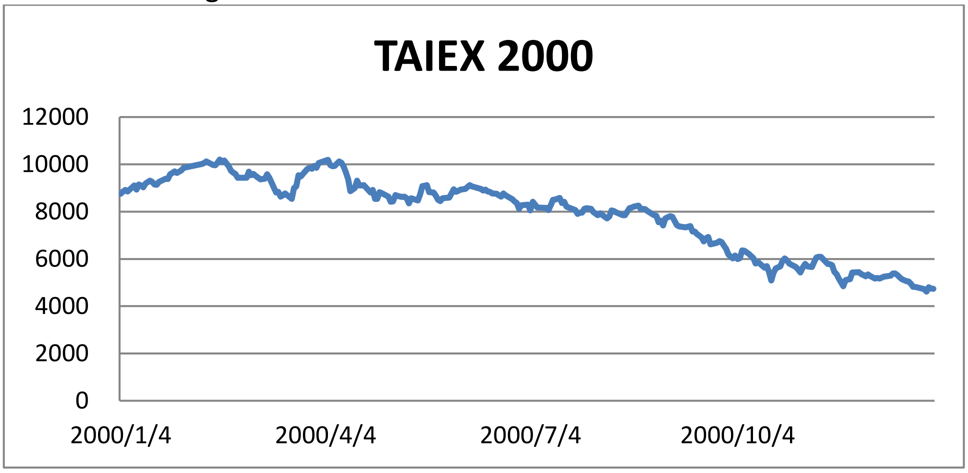

Fig. 6. The actual stock index for TAIEX in 2000

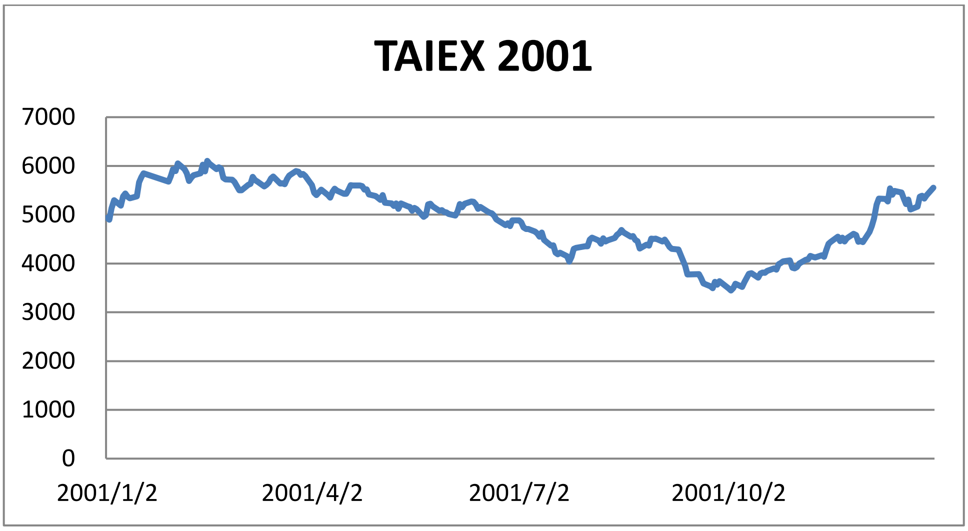

Fig. 7. The actual stock index for TAIEX in 2001

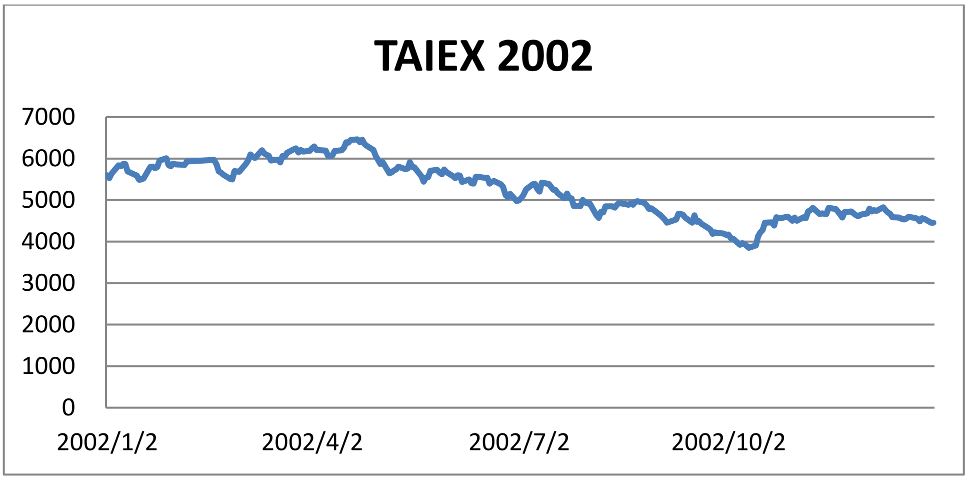

Fig. 8. The actual stock index for TAIEX in 2002

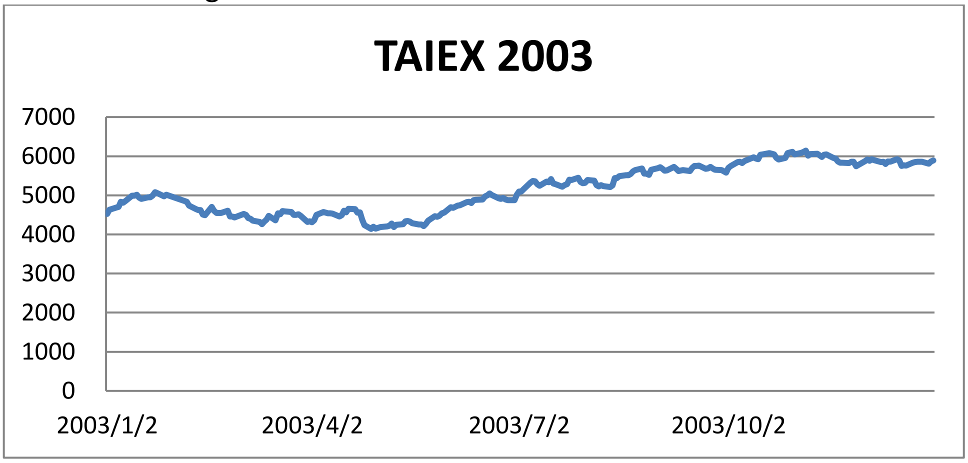

Fig. 9. The actual stock index for TAIEX in 2003

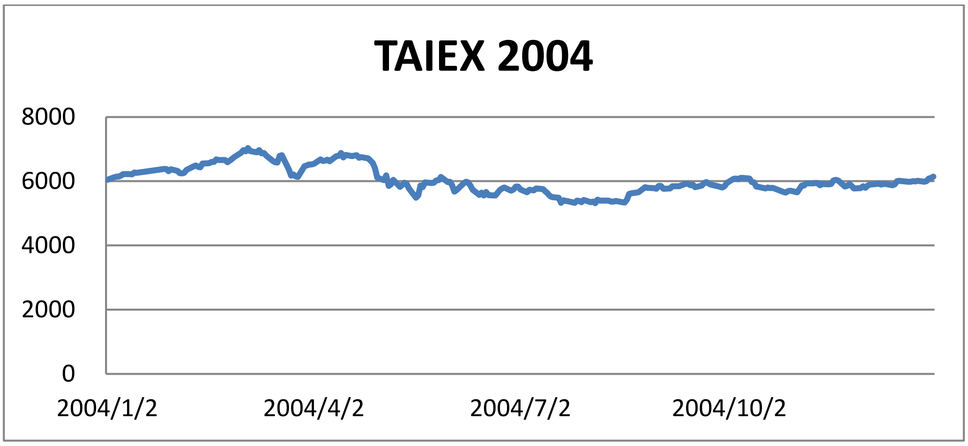

Fig. 10. The actual stock index for TAIEX in 2004

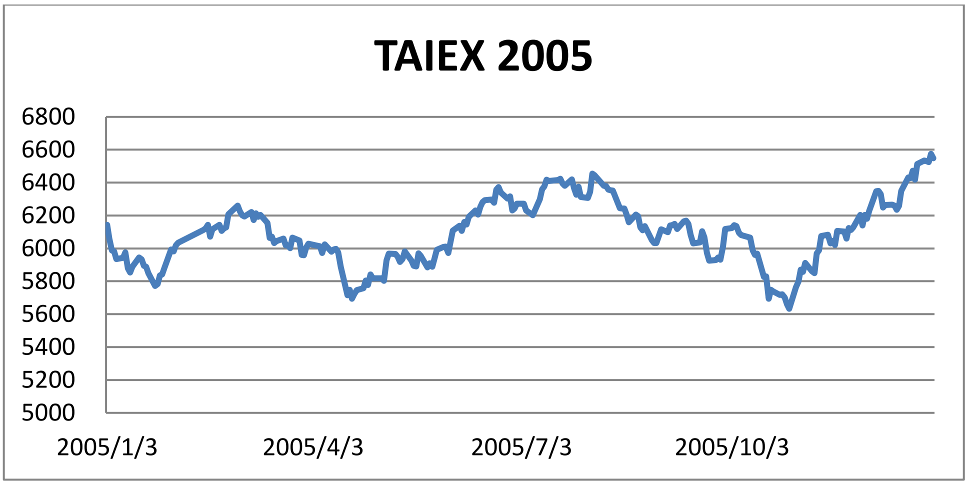

Fig. 11. The actual stock index for TAIEX in 2005

Based on the experimental results and the fluctuation of the TAIEX data set in the past 9 years, we have the following key points of discussion, which are explained as follows:

(1) The method proposed in this study can accurately predict whether the market trend is downward or upward:

The results from Figure 3-11 show that the years for the bear market (November to December of the data set) are 1998, 2000, 2002, and 2003, and the years for the bull market are 2001, 2004, 2005, Through the evaluation of prediction performance (Table 4), it can be shown that the method proposed in this study can predict the volatility pattern of stocks regardless of whether the market trend is bullish or bearish, and obtain a high prediction accuracy.

(2) Proposed model performs well when the stock trend is clearly uptrend:

The results from Figure 7 (TAIEX in 2001) that its fluctuation is an obvious upward trend. From Table 4, it can be seen that the prediction performance (RMSE = 57) proposed in this study is significantly better than other methods (Chen's method (RMSE = 140), Yu's method (RMSE = 167), AR(1) (RMSE = 115), SVR (RMSE = 114)), and from Figure 11 (TAIEX in 2005), it can be seen that the fluctuation is an obvious upward trend. The predictive power (RMSE = 31) is significantly better than other methods (Chen's method (RMSE = 66), Yu's method (RMSE = 69), AR(1) (RMSE = 54), SVR (RMSE = 55)). Therefore, the prediction effect of the forecasting model proposed in this study is particularly significant when the stock index shows an obvious upward trend.

(3) Proposed model performs well when the stock trend is clearly down:

From Figure 6 (TAIEX in 2000), it can be seen that its fluctuation is a clear downward trend. From Table 4, it can be seen that the prediction performance (RMSE = 96) proposed in this study is significantly better than other methods (Chen's method (RMSE = 176), Yu's method (RMSE = 191), AR(1) (RMSE = 130), SVR (RMSE = 136)) Therefore, when the stock index shows a significant downward trend, the forecasting model proposed in this study can produce a significant forecasting effect.

Conclusions

The financial tools of the general public in Taiwan include funds, foreign exchange, bonds, stocks, futures, and options. Among them, futures and options are highly leveraged investment methods. The general public invests less, and among other investment tools, stocks are the most. If the stock fluctuations can be accurately predicted, the forecast results will bring great profits to investors. Therefore, stock analysts will look for good forecasting models. However, traditional time series methods in the past were established based on statistical assumptions, so the data set must meet certain statistical assumptions, such as ARMA, which can only be used in linear and steady-state data sets. However, most of the data types of stocks are nonlinear. Non-stationary data sets, so there will be limitations in application, and most stock data sets will contain disturbances caused by market changes or environmental fluctuations. If the stock index is used directly for prediction, the accuracy of the prediction will be reduced. Technical indicators are often used by ordinary investors to invest in stocks, but there are too many types of stock indicators. Investors will subjectively choose technical indicators to make investment decisions according to their own preferences. However, choosing inappropriate stock indicators will lead to wrong investment strategies.

Based on the problems raised above, this study establishes a new hybrid stock indicator forecasting method, which combines the traditional time series model AR and uses the EMD method to decompose stock indicators into several IMFs and residuals related to decision attributes, and The method of attribute screening is used to select technical indicators related to decision attributes. The proposed method is also compared with other methods, Chen's method (Chen, 1996), Yu's method (Yu, 2005), AR(1) (Engle, 1982) and SVR (Vapnik, 1995), as shown by the experimental results (Table 4), the prediction model proposed in this study is superior to the other listed prediction methods, and the results generated by further statistical tests (Table 5-6) show that, The proposed method is significantly better than the other methods listed. The method proposed in this study can effectively select useful technical indicators for prediction of technical indicators objectively. It is of substantial help in stock investment decision-making.

To summarize, the motivations of this study are that traditional time series models have three major drawbacks in stock price forecasting including: (1) some models can not be applied to the datasets that do not follow the statistical assumptions; and (2) most time-series models which use stock data with many noises involutedly would reduce the forecasting performance; (3) subjective selected technical indicators as input variable by personal experience would lead to lower forecasting accuracy. The contributions of this study are that the proposed forecasting method has excellent forecasting performance and can accurately predict the fluctuation of stocks. It is a useful investment decision-making tool for stock analysts.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics. 31, 307-327.

Box, G., Jenkins, G. (1976). Time series analysis: Forecasting and control, San Francisco: Holden-Day.

Chen, T. L., Cheng, C. H., Teoh, H. J. (2008). High-order fuzzy time-series based on multi-period adaptation model for forecasting stock markets, Physica A. 387, 876-888

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

Chen, S. M., Chung, N. Y. (2006). Forecasting Enrollments Using High-Order Fuzzy Time Series and Genetic Algorithms. International of Intelligent Systems. 21, 485-501.

Cheng, C. H., Chen, T. L., & Wei, L.Y. (2010). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting 180, (9), 1610-1629

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation. Econometrica. 50(4), 987-1008.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc.; 675-701.

Huarng, K. H. (2001). Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets and Systems. 123, 155-162.

Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers and Operations Research, 32(10), 2513-2522.

Huang, N. E., Shen, Z., Long, S. R. (1999). A new view of nonlinear water waves: the Hilbert spectrum, Annu. Rev. Fluid Mech. 31, 417-457.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis, in: Proceedings of the royal society of London series a-mathematical physical and engineering sciences, series A, 454, 903-995.

Huarng, K. H., Yu, T. H. K. (2006). The application of neural networks to forecast fuzzy time series, Physica A. 336, 481-491.

Huarng, K. H., Yu, H. K., Hsu, Y. W. (2007). A Multivariate Heuristic Model for Fuzzy, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 37(4), 836-846.

Kim, K. J. (2003) Financial time series forecasting using support vector machines, Neurocomputing. 55, 307-319.

Kim, K., Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for prediction of stock index. Expert System with Applications. 19, 125-132.

Kimoto, T., Asakawa, K., Yoda, M., Takeoka, M. (1990). Stock market prediction system with modular neural network. In Proceedings of the international joint conference on neural networks, San Diego, California, 1-6.

Nikolopoulos, C., Fellrath, P. (1994). A hybrid expert system for investment advising. Expert Systems. 11(4), 245-250.

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497-505.

Roh, T. H. (2007). Forecasting the volatility of stock price index. Expert Systems with Applications. 33, 916-922.

Song, Q., Chissom, B. S. (1993). Forecasting enrollments with fuzzy time-series Part I, Fuzzy Sets and Systems. 54, 1-10.

Takagi, T., and Sugeno, M. (1983). Derivation of fuzzy control rules from human operator's control actions, in Proc. IFAC Symp. Fuzzy Inform., Knowledge Representation and Decision Analysis. 55-60.

Thawornwong, S., Enke, D. (2004). The adaptive selection of financial and economic variables for use with artificial neural networks, Neurocomputing. 56, 205-232

Vapnik, V. (1995). The nature of statistical learning theory. New York: Springer-Verlag.

Vincent, H. T., Hu, S.-L. J., Hou, Z. (1999). Damage detection using empirical mode decomposition method and a comparison with wavelet analysis, in: Proceedings of the second international workshop on structural health monitoring, Stanford 891-900.

Yu, H. K. (2005). Weighted fuzzy time-series models for TAIEX forecasting, Physica A. 349, 609-624.

Yu, D. J., Cheng, J. S., Yang, Y. (2005). Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Signal Process. 19 (2), 259-270.