Exploring Stock Market Dynamism in Multi-nations with Genetic Algorithm, Support Vector Regression, and Optimal Technical Analysis

Deng-Yiv Chiu
Chung Hua University
707, Sec.2, WuFu Rd.,
Hsinchu,Taiwan 30012, R.O.C.
Shin-Yi Chian
Chung Hua University
707, Sec.2, WuFu Rd.,
Hsinchu,Taiwan 30012, R.O.C.

Abstract- **in this research, an approach in combination with support vector regression (SVR), genetic algorithm (GA), and optimal technical analysis is proposed to explore stock dynamism of multi-nations under different economical environments. First, we apply full search algorithm to select the optimal number of trading days used to calculate the technical indicator values. Genetic algorithm is then used to search the best combination of parameters for SVR kernel function and technical indicators used as SVR input variables. Finally, support vector regression is then used to classify stock data based on the characteristics of non-linear classification. Also, we apply sliding windows to training data to build a steady stock exploratory approach.**

The data sources are stock data from four countries with different economic development degree. They include United States of America, Singapore, Taiwan, and Indonesia. In empirical results, the input variables of middle-long-term technical indicators can bring stable profits and developed country shows better efficient market.

Keyword: Multi-nation stock dynamism, Genetic Algorithm, Support Vector Regression, sliding window, optimal technical analysis

I. INTRODUCTION

References [1] [2] indicate that there are three characteristics in time series problem. They are as below:

1) Inherently noisy: It means that we can not control the behavior of financial market and the relatedness between the past and the future only according to historical data.

2) Non-stationary: The different periods of data will change according to the moving of time.

3) Deterministically chaotic ： Time series exist specific regularity and this kind of regularity is non-linear.

The three characteristics indicate that the stock market time series problem might exist some kind of memory ability and regularity, as in [3]. Therefore, some researches use related stock market analysis techniques to discover the regularity of non-linear data.

Traditional time series research methods utilize linear regression to construct stock exploration model. The performance of linear regression method has the better result on linear problem, but it is not appropriate for the non-linear time series problems.

Artificial neural network can be used as an approach to resolve non-linear problems. It has great learning ability and flexibility to continually process new data repeatedly. Therefore, many researchers employ ANN to solve financial non-linear problems. But ANN needs huge amount of control parameters. Also, it has problem of over-learning. As a result, the prediction performance of the testing data is not properly proportion to training data.

Support Vector Regression (SVR) is one of the popular learning machines, as in [4]. SVR minimizes structure risk and empirical risk. SVR does not have the local optimal problem.

In this research, we try to explore stock market dynamism in multi-nations of different kind of economic environments with hybrid approach of GA, SVR, sliding window, and different optimal of trading days used to calculate technical indicator values. First, we search the optimal number of trading days used to calculate the technical indicator values. Then we use GA to locate the optimal combination of kernel function parameters for SVR and technical indicators used as the input of SVR. Finally, we use SVR to classify the stock data with moving window training method.

We explore four stock markets, including Taiwan, U.S.A., Singapore, and Indonesia. In empirical results, the input variables of middle-long-term technical indicators can bring stable profits and developed country shows better efficient market.

II. RELATED WORKS

In this section, we introduce some works related to technical analysis theory, genetic algorithms (GA), and support vector regression (SVR).

A. Technical Analysis

Technical analysis theory was proposed by Charles, H. Dow and Willams, P Hamilton. It studies the relatedness between the past and present and reaction of financial market through the records of indicators or graph, in order to explore the trend of future stock index.

Reference [6] applies moving average transaction rules to study the financial market in Bangladesh, India, Pakistan, and Sri Lanka. The research periods are from 1990 to 2000, the experiment result shows that the technical transaction rules have good effect in those countries.

B. Genetic Algorithm

GA is an efficient and good search method. With the simulation of biological evolution phenomenon, the chromosomes with better fitness function value are left. Also, with mechanisms of crossover and mutation, issue of partial minimization during search is avoided and search time is shortened.

Reference [7] employs back propagation neural network, genetic algorithm, and 13 technical indicators to explore the stock index in Korea. The research result shows that the feature discretization (GAFD) with GA method outperforms ANN method. Reference [10] utilizes genetic programming method based on technical indicators to locate the trading rules for stock of TSMC. The experiment results show that the return rate of the proposed approach is better than buy-and-hold strategy.

C. Support Vector Regression

Support vector regression (SVR) is a regression technique extended from support vector machine, as in [5]. It is often applied in the fields of pattern recognition and text classification. Theoretically, it is a learning system using linear-function hypothesis space in a high-dimensional feature space. It is a learning algorithm of optimization theorem and minimized structure risk. In recent years, it has been also used in the research of classification and exploration of finance. Support vector regression consists of linear support vector regression and non-linear support vector regression, as in [9].

Reference [8] applies SVM to study the production time series of mechanical industry in Taiwan. The result shows that the accuracy rate of SVM outperforms seasonal SARIMA and general regression neural networks model.

III. RESEARCH METHOD and STRUCTURE

In this research, an approach in combination with support vector regression (SVR), genetic algorithm (GA), optimal technical analysis, and sliding window is used to explore stock dynamism in different economical environments. The detailed explanation is as follow.

*1)*We collect original data from multi-national stock market, including opening price, highest price, lowest price, transaction amount, and closing price. We use 40 percent of the data for choosing the best number of days used to calculate the values of technical indicators and use 60 percent of data as the training and testing data for GA-SVR approach.

2) We calculate the values of technical indicators.

3) We employ full search method to locate the optimal number of days used to calculate the values of technical indicators of indicators with the top three earning rates.

4) Calculate the values of technical indicators with the top three earning rates.

5) Initialize of chromosomes for GA process. The first generation of GA process is initialized at random. A generation includes 20 chromosomes. In experiment one, each chromosome consists of 27 genes. The former 7 genes represent technical indicators. The following 10 genes represent C value of SVR, and the final 10 genes represent gamma of SVR parameter.

6) Genes are decoded to facilitate the combination of selected technical indicators.

7) Corresponding values of selected technical indicators are extracted as training data to form the input of SVR.

8) The extracted training data is used to train SVR and produce values needed to evaluate fitness function.

9) Evaluate fitness function. To find the optimal combination of technical indicators, training accuracy rate and training error are considered in fitness function. Larger fitness function value means that the chromosome can make better financial earning.

10) The termination criterion is evolution of 100 generations. If the criterion is met, terminate the GA process and then go to step (12).

11) Perform process of genetic algorithm. In genetic algorithm, chromosome evolution includes three processes, selection, crossover, and mutation. In the selection process, we combine roulette wheel selection and tournament selection to complete the evolution process of selection and reproduction. In the crossover process, the crossover rate is 0.8. In mutation, the mutation rate is defined as 0.01.The process is redirected to step (6).

12) Evaluate testing data with trained SVR classifier: The trained SVR classifier is used to classify testing data to determine the proper transaction time point.

13) We compare the performance of three experiments.

A. Search the Optimal Number of Trading Days Used to Calculate Technical Indicator Values

Traditional technical indicators adopt fixed days (such as 6 days RSI, 10 days PSY etc.), but we try to select the optimal number of trading days. We apply full search algorithm to select the optimal number of trading days which is used to calculate the values of technical indicators used in technical rules. The process is as shown in Fig.1. This method examines all possible values to find the optimal solution. In our research, the number of days is from 1 day to 100 days and then we find the buying point and selling point to calculate the earning rate as shown in (1)

$$\text{Earning rate} = \frac{\text{Closing price at sell point} - \text{closing price at buy point}}{\text{Closing price at buy point}} \quad (1)$$

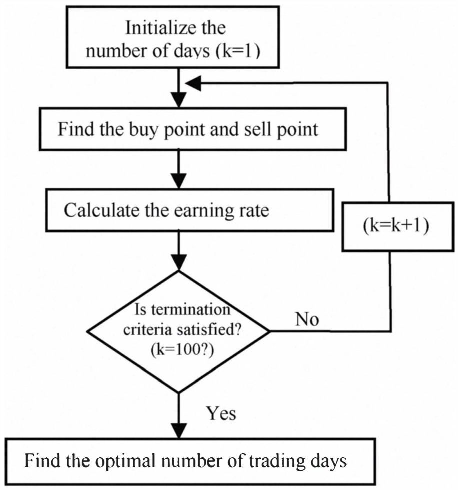

Fig. 1. Search the best number of trading days used to calculate the values of technical indicators

B. *GA-SVR Approach*

This research combines the advantages of genetic algorithms and support vector regression to construct a stock market exploration approach. This approach employs GA to choose the optimal combination of technical indicators and the parameter values of kernel function. We use these combinations of technical indicators to form the input vector of SVR. We calculate the fitness value of GA according to the classification result of SVR. The details are described as below.

1) Chromosome design: In this research, the chromosome is divided into two parts. A part of chromosome is used to choose technical indicators and the other part of *chromosome* is used to choose the parameters of kernel function of SVR, including C (cost) and γ (gamma).

a. Choose the technical indicators : We use binary code; each gene represents a technical indicator. When the value is 1, it means to choose that indicator as the input of SVR; otherwise, it means the indicator is not used as input.

b. Choose the parameters of kernel function : the kernel function of SVR method used in this study is Radial Basis Function (RBF). The parameters of kernel function are C value and γ value. We refer to some researches and found out that the ideal ranges of the parameters are as follows : $C=[2^{12}\sim2^{-2}]$, $\gamma=[2^{4}\sim 2^{-10}]$. Therefore, we consider the ideal ranges of parameters and the calculation time. We use 10 genes to represent the C value and the other 10 genes to represent the γ value. The value of C is between 1 and 1024, as shown in (2). The value of γ is between 0.02 and 20.48, as shown in (3). For example, suppose the genes for C value are 0001001101, the transformation is as follows. C value= (26*1+23*1+22*1+20*1+1) =78.

$$C = x_9*2^9 + x_8*2^8 + x_7*2^7 + x_6*2^6 + x_5*2^5 + x_4*2^4 + x_3*2^3 + x_2*2^2 + x_1*2^1 + x_0*2^0 + 1 \quad (2)$$

$$\gamma = \begin{pmatrix} y_9*2^9 + y_8*2^8 + y_7*2^7 + y_6*2^6 \\ +y_5*2^5 + y_4*2^4 + y_3*2^3 + y_2*2^2 \\ +y_1*2^1 + y_0*2^0 + 1 \end{pmatrix} / 50 \quad (3)$$

2) Calculation of fitness value: We try to achieve high accuracy rate and low error rate. Therefore, we adjust the weight of training accuracy and training error in fitness function as shown in (4). During the process of experiment, we find that the maximum value of training error is less than 2. Therefore, we use the (2-training error) to adjust the difference between accuracy and error. For example, if the accuracy rate is 0.9 and the error is 0.75, then the fitness value is 2.15 (0.9+ (2-0.75)).

Fitness function= training accuracy rate + (2-training error) (4)

3) Normalization of the input data: In this research, we employ linear transformation method to normalize the input, in order to promote the accuracy rate.

4) Kernel function: The kernel function of SVR used here in this study is RBF since it is a nonlinear function that can convert data from original space to a higher-dimensional space to solve non-linear problems well. When original data and attribute are non-linear, this function has good effect.

5) Application of sliding windows training method: In order to reflect the change of environment, we use different periods of data to simulate the changing of time. We use 300 days of historical data as training data to train SVR and then used the following 30 days of data to test the trained SVR. And then we move the time window forwards along the time-axis for 30 days and perform the training and testing repeatedly.

IV. EXPERIMENT

This research collected the stock market data from four countries, including the United States, Taiwan, Singapore, and Indonesia. The data include open price, close price, the highest price, transaction amount, and the lowest price. This research adopts 7 technical indicators as the input variables, including Relative Strength Index (RSI), K, D, RSV, William Index, Psychological Line, and BIAS. We use 40 percent of historical data to locate the optimal number of trading days used to calculate technical indicator values and 60 percent as the training and testing data of GA-SVR approach.

In the experiment one, we use the indicator with the highest earning rate as the input of GA. In the experiment two, we use the indicators with the top 2 high earning rates as the input of GA. In the experiment three, we use the indicators with the top 3 high earning rates as the input of GA.

A. *Experiment Result-Taiwan*

Here, we describe the experiment results for Taiwan stock market. Table I shows the technical indicators and the optimal number of trading days used to calculate the values of technical indicators for three experiments. Table II shows the parameter values of SVR kernel function and testing earning rate. From the experiments, we conclude the followings:

1) We find out that the numbers of days for indicators RSV, K, and D are suitable for short-term investment; the profit is higher. The numbers of days for other four indicators are suitable for middle-long-term investment.

2) According to the experiment results of Taiwan, the testing earning rate of experiment one is -2% and the testing error is 1.74. Most of the input indicators belong to short-term indicator. The testing earning rates of experiment two and three are 10% or more and the range of testing error is from 0.16 to 0.27. Most of the input indicators belong to middle-term and short-term data periods. Therefore, we conclude that the indicators with short-middle-term data periods have more stable performance.

3) The earning rates of experiment two and experiment three are 10% or more. It shows that the proposed approach is valuable for reference in exploring the stock market in Taiwan.

TABLE I
THE TECHNICAL INDICATORS AND the OPTIMAL NUMBER of TRADING DAYS USED to CALCULATE the VALUES of TECHNICAL INDICATORS (TAIWAN)

Indicator	Experiment One	Experiment Two	Experiment Three
RSV	RSV(1*)	RSV(1) RSV(2)	RSV(1) RSV(2) RSV(5)
K	K(1)	K(1) K(2)	K(1), K(2) K(5)
D	D(1)	D(1) D(2)	D(1), D(2) D(5)
PSY	PSY(12)	PSY(12) PSY(9)	PSY(12) PSY(9) PSY(11)
BIAS	BIAS(4)	BIAS(4) BIAS(45)	BIAS(4) BIAS(45) BIAS(46)
W%R	W%R(52)	W%R(52) W%R(55)	W%R(52) W%R(55) W%R(53)
RSI	RSI(29)	RSI(29) RSI(30)	RSI(29) RSI(30) RSI(36)
Average days	14.3	17.5	19.3

TABLE II
SVR PARAMETER VALUES and TESTING EARNING RATE (TAIWAN)

Experiment	C	γ	Testing Earning Rate
one	255	5.09	-2%
two	1023	20.46	10%
three	1021	20.42	12%

B. *Experiment Result-U.S.A.*

Here, we describe the experiment results for United States stock market. Table III shows the technical indicators and the optimal number of trading days used to calculate the values of technical indicators for three experiments. Table IV shows the parameter values of SVR kernel function and testing earning rate. The testing earning rate of experiment one is 0.6%. The testing earning rate of experiment two is -2.8%. The testing earning rate of experiment three is 0.6%. According to the data periods of indicators, we find whether the number of days used to calculate the technical indicator values belong to short-term or long-term, their performances are not satisfied. It is obvious that the stock market in U.S.A. is an efficient market. That is, it is difficult to explore the stock market dynamism.

TABLE III
THE TECHNICAL INDICATORS AND the OPTIMAL NUMBER of TRADING DAYS USED to CALCULATE the VALUES of TECHNICAL INDICATORS (U.S.A)

Indicator	Experiment One	Experiment Two	Experiment Three
RSV	RSV(65)	RSV(65) RSV(32)	RSV(65) RSV(32) RSV(30)
K	K(65)	K(65) K(32)	K(65) K(32) K(30)
D	D(65)	D(65) D(32)	D(65) D(32) D(30)
PSY	PSY(3)	PSY(3) PSY(2)	PSY(3) PSY(2) PSY(4)
BIAS	BIAS(45)	BIAS(45) BIAS(46)	BIAS(45) BIAS(46) BIAS(50)
W%R	W%R(45)	W%R(45) W%R(47)	W%R(45) W%R(47) W%R(48)
RSI	RSI(3)	RSI(3) RSI(23)	RSI(3) RSI(23) RSI(2)
Average days	41.6	36.1	33.3

TABLE IV
SVR PARAMETER VALUES and TESTING EARNING RATE (U.S.A)

Experiment	C	γ	Testing Earning Rate
One	255	5.09	0.6%
Two	1023	20.4	-2.8%
Three	1017	20.3	0.6%

C. *Experiment Result-Singapore*

Here we describe the experiment results for Singapore. Table V shows the technical indicators and the optimal number of trading days used to calculate the values of technical indicators for three experiments. Table VI shows the parameter values of SVR kernel function and testing earning rate. We conclude the followings:

1) According to the data periods of indicators (as shown in Table V), we find that most of the data periods of indicators with higher earning rates belong to middle-term and long-term data periods. When we utilize those indicators, the proposed GA-SVR approach can achieve good performance. If we use the technical indicators with longer periods as a transaction strategy, the performance will be more stable.

2) The earning rates of the three experiments achieve 9~12%. It is obvious that the proposed approach can provide stable stock market information in Singapore.

TABLE V
THE TECHNICAL INDICATORS AND the OPTIMAL NUMBER of TRADING DAYS USED to CALCULATE the VALUES of TECHNICAL INDICATORS (SINGAPORE)

Indicator	Experiment One	Experiment Two	Experiment Three
RSV	RSV(74)	RSV(74) RSV(100)	RSV(74) RSV(100) RSV(75)
K	K(74)	K(74) K(100)	K(74) K(100) K(75)
D	D(74)	D(74) D(100)	D(74) D(100) D(75)
PSY	PSY(9)	PSY(9) PSY(13)	PSY(9) PSY(13) PSY(12)
BIAS	BIAS(80)	BIAS(80) BIAS(4)	BIAS(80) BIAS(4) BIAS(78)
W%R	W%R(39)	W%R(39) W%R(37)	W%R(39) W%R(37) W%R(40)
RSI	RSI(57)	RSI(57) RSI(56)	RSI(57) RSI(56) RSI(58)
Average Days	58.1	58.4	58.6

TABLE VI
SVR PARAMETER VALUES and TESTING EARNING RATE (SINGAPORE)

Experiment	C	γ	Testing Earning Rate
One	255	5.09	9.2%
Two	1023	20.4	11.2%
Three	1023	20.4	12.6%

D. Experiment Result- Indonesia

Here, we describe the experiment results for Indonesia stock market. Table VII shows the technical indicators and the optimal number of trading days used to calculate the values of technical indicators for three experiments. Table VIII shows the parameter values of SVR kernel function and testing earning rate. The earning rate of the first experiment is 2%, the testing error is 1.83. The earning rate of the second experiment is 6%, the testing error is 0.42. The earning rate of the third experiment is 8%, the testing error is 0.30. According to the data periods and the number of indicators, the experiment result shows that if GA chooses more indicators with longer periods, the performance will be better. Furthermore, the more input variables are chosen, the less testing error will be.

TABLE VII
THE TECHNICAL INDICATORS AND the OPTIMAL NUMBER of TRADING DAYS USED to CALCULATE the VALUES of TECHNICAL INDICATORS (INDONESIA)

Indicator	Experiment One	Experiment Two	Experiment Three
RSV	RSV(4)	RSV(4) RSV(65)	RSV(4) RSV(65) RSV(59)
K	K(4)	K(4) K(65)	K(4) K(65) K(59)
D	D(4)	D(4) D(65)	D(4) D(65) D(59)
PSY	PSY(13)	PSY(13) PSY(7)	PSY(13) PSY(7) PSY(12)
BIAS	BIAS(100)	BIAS(100) BIAS(94)	BIAS(100) BIAS(94) BIAS(99)
W%R	W%R(21)	W%R(21) W%R(20)	W%R(21) W%R(20) W%R(19)
RSI	RSI(49)	RSI(49) RSI(50)	RSI(49) RSI(50) RSI(48)
Average Days	27.9	40.1	43.6

TABLE VIII
SVR PARAMETER VALUES and TESTING EARNING RATE (Indonesia)

Experiment	C	γ	Testing Earning Rate
One	255	5.09	2%
Two	1023	20.4	6%
Three	1023	20.4	8%

V. CONCLUSION

We utilize genetic algorithm to locate the optimal combination of technical indicators and the parameters of SVR kernel function. And we perform the proposed GA-SVR approach with the optimal trading days used to calculate the values of technical indicators and sliding window technique. We examine the stock markets of the United States, Taiwan, Singapore and Indonesia.

The experiment results indicate that the countries with immature economic development environment have better performance. In other hand, the economic development in the U.S.A is more mature, it is more difficult to explore the stock market dynamism. Therefore, the performance of the approach is not as good as the other three countries.

Also, we find that the performance of middle-long-term technical indicators is better than short term indicators. It is possible that the value of short term technical indicators can not reflect the real trend of stock market.

REFERENCES

[1] Hall, J. W., "Adaptive selection of U.S. stock with neural nets," In: GJ Deboeck (Eds.), Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets, 1994

[2] Yaser, S. and Atiya, A. F., "Introduction to financial forecasting," Applied Intelligence, vol. 6, 1996, pp. 205-13.

[3] Tay, F. E. H. and Cao, L., "Application of support vector machines in financial time series forecasting," Omega, vol. 29, no. 4, 2001, pp.309-317.

[4] Vapnik, V. N., *The Nature of Statistical Learning Theory*, Springer, 1995

[5] Vapnik, V. N., Golowich, S. and Smola, A., "Support vector method for function approximation, regression estimation and signal processing,"Advantage Neural Information Proceedings System Cambridge, 1997, pp. 281-287.

[6] Gunasekarage, A. and Power, D. M., "The profitability of moving average trading rules in South Asian stock markets," Emerging Markets Review, vol. 2, 2001, pp.17-33.

[7] Kim, K. J. and Han, I., "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index," Expert Systems with Application, vol. 19, 2000, pp. 125-132.

[8] Pai, P. F. and Lin, C. S., "Using support vector machines in forecasting production values of machinery industry in Taiwan," International Journal of Advanced Manufacturing Technology, vol. 27, 2006, pp.205-210.

[9] Huang, C. L. and Tsai, C. Y. "A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting," Expert Systems with Applications," vol.36 no.2 part 1, 2009, pp.1529-1539.

[10] Yao-Tang Lin, "An Application of Genetic Programming on Trading Rules," National Central University Master Thesis, 2001