Determining the Time Period and Amount of Training Data for Stock Exchange of Thailand Index Prediction

Phaisarn Sutheebanjard
Graduate School of Information Technology
Siam University
Bangkok, Thailand
mr.phaisarn@gmail.com

Wichian Premchaiswadi
Graduate School of Information Technology
Siam University
Bangkok, Thailand
wichian@siam.edu

***Abstract*— This paper analyzed the time period and amount of training data that were used in applying the prediction function to forecast the Stock Exchange of Thailand index (SET Index). The training data were grouped into one year and two years over the period of 2003-2008, and the testing data were collected from January 2005 to March 2009. This paper compared the results of using different time periods training data in order to find the best training data set that will most accurately predict the SET index. The results show that there is no significant difference between using one year and two years training data with MAPE lower than 1%.**

Keywords-component; Stock market forecasting, Stock Exchange of Thailand, Evolution Strategies, Prediction function, Data analysis, Financial time series forecasting

I. INTRODUCTION

Recently, the SET Index prediction has been studied by many researchers. They have been using various techniques such as Neural Network [1-3], Autoregressive Integrated Moving Average (ARIMA) [3] and the Evolution Strategies [4-5]. Neural networks are very efficient adaptive forecasting models because of their excellent performance in self learning capability. Unlike other techniques that construct functional forms to represent relationships of data, neural networks are able to learn patterns or relationships from data itself [1]. However, with the effects of black-box, slow convergence, local optimal they are not applicable for some applications [6]. ARIMA was introduced by G. Box and G. Jenkins in the early 1970s. This time series analysis can capture complex arrival patterns, including those that are stationary, non-stationary, and seasonal (periodic) ones [7]. ARIMA approach is elegant in theory but has been of little practical use in business because of its complexity and limited increase in accuracy over less sophisticated methods. Evolution Strategies was introduced by Rechenberg [8]. Evolution Strategies (ES) are algorithm which imitates the principles of natural evolution as a method to solve parameter optimization problems [9-11]. ES is one of the most popular evolutionary algorithms. ES are generally applied to numerical optimization for its real valued representation.

According to the studies, it is assumed that the behaviors of the stock market in the future could be predicted by analyzing the historical stock data. But the question is what time period and amount of the historical data should be used. This research compared the results between using one year and two years training data in order to determine the proper training data set to be used in predicting the SET Index.

II. THE STOCK EXCHANGE OF THAILAND

A. *History of Stock Exchange of Thailand (SET)*

Thai stock market officially started trading on April 30, 1975 and named as “The Securities Exchange of Thailand”. On January 1, 1991 its name was formally changed to “Thai Stock Exchange of Thailand” (SET). The index of Stock Exchange of Thailand is called SET Index. SET Index is a composite market capitalization-weighted price index which compares the current market value (CMV) of all listed common stocks with its market value on the base date of April 30, 1975 (Base Market Value or BMV), which was when the stock market was established. The initial value of SET index on the base date was set to 100 points. The formula of SET index calculation is as follows:

$$\text{SET Index} = \frac{\text{Current Market Value x 100}}{\text{Base Market Value}} \quad (1)$$

B. *Impact Factors to Stock Exchange of Thailand (SET)*

The SET index movement is dependent on both global and domestic economic factors [12]. Since countries are linked together, movement on one stock market may have an impact on other stock markets. Naturally, the Thai stock market has unique characteristics, so the factors influencing the prices of stocks traded in this market are different from the factors influencing other stock markets [12]. An example of factors that influence the Thai stock market are foreign stock indexes, the value of the Thai Baht, the price of oil, the price of gold, the Minimum Loan Rate (MLR) and many others [1, 4, 10, 13-16]. There were some researchers that used these factors to forecast the Stock Exchange of Thailand (SET) index such as Tantinakom [14] who used trading value, trading volume, interbank overnight rate, inflation, net trading value of investment, value of the Thai Baht, price earning ration, the Dow Jones index, the Hang Seng index, the Nikkei index, the Straits Times index and the Kuala Lumpur Stock Exchange Composite index. In 2000, Khumpoo [15] used the Dow Jones index, the price of gold, the Hang Seng index, the exchange rate of the Japanese yen and the Thai baht, the Minimum Loan Rate (MLR), the

Nikkei index, the price of oil, the Straits Times Industrial index and the Taiwan weighted index. In 2004, Chotasiri [16] used the interest rate of Thailand and the USA, the exchange rate of USD, JPY, HKD and SKD, the stock exchange indices of USA, Japan, Hong Kong and Singapore, the consumer price index, and the price of oil. In 2005, Chaereonkithuttakorn [17] used United State stock indices including the Nasdaq index, the Dow Jones index and the S&P 500 index. In 2005, Rimcharoen et al. [4] used the Dow Jones index, the Nikkei index, the Hang Seng index, the price of gold and the Minimum Loan Rate (MLR). In 2007, Worasucheep [13] used the Minimum Loan Rate (MLR), the exchange rate of the Thai Baht and the US dollar, daily effective over-night federal fund rates in the USA, the Dow Jones index and the price of oil. In 2008, Chaigusin et al. [1] used the Dow Jones index, the Nikkei index, the Hang Seng index, the price of gold, the Minimum Loan Rate (MLR) and the exchange rate of the Thai Baht and the US dollar.

This research takes into account both internal and external factors to forecast the SET index. The external factors are foreign major stock market indices, while the internal factors are SET index and MLR. From the assumption that both external and internal factors probably have great impact on the SET index, these factors include:

- SET Index (Thailand)
- Dow Jones index (New York)
- Nikkei index (Japan)
- Hang Seng index (Hong Kong)
- Minimum Loan Rate (MLR)

III. EVOLUTION STRATEGIES (ES)

Evolution strategies (ES) is one of the main branches of evolutionary computation. Similar to Genetic Algorithms, Evolution Strategies (ES) are algorithms which imitate the principles of natural Darwinian evolution, generally produce consecutive generations of samples. During each generation, a batch of samples is generated by perturbing the parents' parameters by mutating their genes. A number of samples are selected based on their fitness values, while the less fit individuals are discarded. The winners are then used as parents for the next generation, and so on. This process typically leads to increasing fitness over the generations.

The evolution strategy (ES) was proposed for real value parameters optimization problems developed in 1971 by Rechenberg. In ES, the representation used was one n-dimensional real-valued vector. A vector of real values represented an individual. The standard deviation was used to control search strategy in ES. The main operator in ES was Gaussian mutation, in which a random value from a Gaussian distribution (normal distribution) was added to each element of an individual's vector to create a new offspring. This basic ES framework, though simple and heuristic in nature, has proven to be very powerful and robust, spawning a wide variety of algorithms.

The basic difference between evolution strategy and genetic algorithms lies in their domains (i.e. the representation of individuals). ES represents an individual as float-valued vectors instead of binary representation. This type of representation reduces the burden of converting genotype to phenotype during evolution process.

Evolutionary strategies (ES) introduced by Ingo Rechenberg (1971, 1973) were (1+1)-ES and (μ+1)-ES. And two further version introduced by Schwefel were (μ+λ)-ES and (μ,λ)-ES Schwefel, (1975, 1977). This research uses (1+1)-ES (two-membered ES) for the selection process. The (1+1)-ES consists of one parent individual (a real-valued vector), which produces one offspring by means of adding normal distribution random numbers. The better of both individuals then serves as the ancestor of the following iteration/generation. The (1+1)-ES was used to find the coefficients of function. Firstly, initialize the coefficient of prediction function by mutation operation. Secondly, each child is evaluated to its fitness function for a possible solution in each generation. These evaluations are preserved for creating a new generation.

This research used the prediction function of Sutheebanjard and Premchaiswadi [5] where the important factors namely, the Down Jones index, the Hang Seng index, the Nikkei index and domestic MLR were taken into account as shown in (2).

$$SET_{(t)} = a_0 SET_{(t-1)} + a_1 \left(\frac{a_2 DJ_{(t-1)} + a_3 NK_{(t-1)} + a_4 HS_{(t-1)}}{a_5 MLR_{(t-1)}} \right) \quad (2)$$

where a_0-a_5 denote coefficients.
SET is SET index (Thailand)
DJ is Dow Jones index (New York)
NK is Nikkei index (Japan)
HS is Hang Seng index (Hong Kong)
MLR is Minimum Loan Rate (MLR)

IV. EXPERIMENTAL RESULTS

The experimental data were collected from a reliable source, the Bank of Thailand, which consisted of historical data of the SET index, the Dow Jones index, the Nikkei index, the Hang Seng index, and Minimum Loan Rate. Since the raw data were obtained from different stock markets in different countries, therefore some data were missing because each country has different stock market holidays or non trading days. However, those gaps can be filled by using the data from previous day with no statistically significant difference. So, the assumption underlying this study was that the missing data on non-trading days will be filled with previous day's data.

The training data in this experiment were divided into two different time ranges: one year and two years training data. The one year training data consisted of six different time periods: 2003, 2004, 2005, 2006, 2007 and 2008, and the two years training data consisted of five different time periods: 2003-2004, 2004-2005, 2005-2006, 2006-2007 and 2007-2008. The test data used in this experiment were collected from January 2005 - March 2009, 1040 days in total.

The investigated time series started from January 2005 to March 2005. It contained 1040 days of test data. The results are shown in table 1 and 2 and the graph of the lowest MAPE is shown in figure 1.

TABLE I. EXPERIMENTAL RESULTS ON JANUARY 2005 – MARCH 2009 (1040 DAYS)

Training Period	Train		Test		a_0 ($SET_{(t-1)}$)	a_1	a_2 ($DJ_{(t-1)}$)	a_3 ($NK_{(t-1)}$)	a_4 ($HS_{(t-1)}$)	a_5 ($MLR_{(t-1)}$)
	MSE	MAPE(%)	MSE	MAPE(%)						
2003	42.2709	1.0079	94.9711	0.9824	0.98	0.006129	0.556757	0.341830	0.031283	0.770467
2004	101.3708	1.1270	95.4627	0.9868	0.98	0.003150	0.041587	1.252696	0.504813	0.940554
2005	36.0749	0.6731	95.0982	0.9832	0.98	-0.007344	-2.109472	-3.448182	-0.757970	6.897148
2006	117.5130	0.8710	95.1273	0.9833	0.98	0.001511	0.381432	0.713091	0.033982	0.238012
2007	90.2289	0.8848	95.0034	0.9830	0.98	0.021602	0.109120	0.116916	0.013107	0.711249
2008	143.5093	1.4312	94.9678	0.9820	0.98	-0.006779	-1.745933	-1.165797	-0.031254	2.679369
Average	**88.4946**	**0.9992**	**95.1051**	**0.9834**						
MIN	**36.0749**	**0.6731**	**94.9678**	**0.9820**						
MAX	**143.5093**	**1.4312**	**95.4627**	**0.9868**						

TABLE II. EXPERIMENTAL RESULTS ON JANUARY 2005 – MARCH 2009 (1040 DAYS)

Training Period	Train		Test		a_0 ($SET_{(t-1)}$)	a_1	a_2 ($DJ_{(t-1)}$)	a_3 ($NK_{(t-1)}$)	a_4 ($HS_{(t-1)}$)	a_5 ($MLR_{(t-1)}$)
	MSE	MAPE(%)	MSE	MAPE(%)						
2003-2004	73.0442	1.0701	95.1106	0.9833	0.98	-0.007340	-2.111214	-3.441897	-0.756813	6.897762
2004-2005	69.1961	0.9052	95.2551	0.9847	0.98	0.002380	0.506854	2.138112	0.656223	1.240233
2005-2006	77.2094	0.7796	94.9636	0.9827	0.98	0.006198	0.556665	0.341234	0.029068	0.770602
2006-2007	104.6126	0.8896	95.0799	0.9873	0.97	-0.011266	-2.083171	-3.497098	-0.815088	6.950055
2007-2008	117.0400	1.1599	94.9791	0.9831	0.98	0.006242	0.557172	0.341211	0.030448	0.772023
Average	**88.2205**	**0.9609**	**95.0776**	**0.9842**						
MIN	**69.1961**	**0.7796**	**94.9636**	**0.9827**						
MAX	**117.0400**	**1.1599**	**95.2551**	**0.9873**						

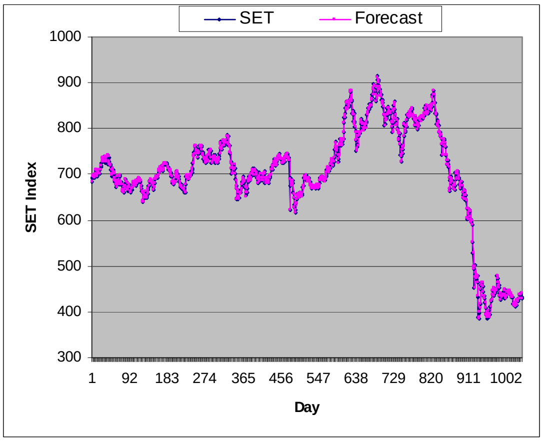


Figure 1. SET index comparison graph on January 2005 - March 2009

A. Methods comparison

The experimental results were compared with the Simple Moving Average and Random Walk by using yesterday's SET index. The simple moving average is a simple technique in time series forecasting. The weights of simple moving average used in this research were 0.5(t-1), 0.3(t-2) and 0.2(t-3) as shown in (3). The Random Walk by using yesterday's SET index (SET(t-1)) was also analyzed as shown in (4). The comparable result was shown in table 3. The result shows that (2) is the best prediction function for predicting the SET index in this time period with the lowest MAPE errors.

$$SET_{(t)} = 0.5SET_{(t-1)} + 0.3SET_{(t-2)} + 0.2SET_{(t-3)} \quad (3)$$

$$SET_{(t)} = SET_{(t-1)} \quad (4)$$

TABLE III. MAPE COMPARISON FOR THE PERIOD OF JANUARY 2005 – MARCH 2009

Method	MAPE (%)
(2) for 1Year training data [5]	**0.9820**
(2) for 2Years training data [5]	0.9826
Simple Moving Average (8)	1.1752
Yesterday SET Index (9)	0.9916

B. Comparison of Different Training Data Periods

The previous section showed that the [5] method yields the lowest MAPE among other methods no matter what training data were used; one year or two years. This section shows the result of using one year and two years training data at different levels weight of SET index (a0) on the test data from January 2005 to March 2009, four years and three months in total. The MAPE on one year training data can be represented as graphs in figure 2. And the MAPE on two years training data can be represented as graphs in figure 3. The graphs of two years training data on figure 3 at different levels of a_0 look smoother with lower error MAPE than the graphs of one year training data on figure 2. This means that the longer period of training data can yield lower error at almost all level of a_0, except for the a_0 level of 98 (the lowest error level). According to table 4, the error of one year training data is a little bit lower than two years training data, but it is not significant.

V. CONCLUSIONS

This paper analyzed the time period and amount of training data that were used in applying the prediction function to forecast the Stock Exchange of Thailand index (SET Index). The experiment was conducted by using one year and two years training data over the period of 2003 to 2008. The results show that there is no significant difference between using different years of training data, although many crises took place in Thailand such as the coup d'état in 2006, the anti-government protest and the subprime mortgage crisis in 2008. In addition, the results also indicate that there was no significant difference between using one year and two years training data. However, using two years training data required more computation time than one year training data, so one year training data is considerably sufficient. In conclusion, the training data used for SET index prediction can be obtained from any time period during 2003-2008 and can be either one year or two years with no significant difference in terms of prediction accuracy.

REFERENCES

[1] S. Chaigusin, C. Chirathamjaree and J. Clayden, "The Use of Neural Networks in the Prediction of the Stock Exchange of Thailand (SET) Index," Computational Intelligence for Modelling Control & Automation. 2008.

[2] S. Pattamavorakun, and S. Pattamavorakun, "Determination the Number of Hidden Nodes of Recurrent Neural Networks for River Flow and Stock Price Forecasting," International Conference on Software Engineering Research, Management and Applications 5th, 2007.

[3] J. Tunsenee, "Accuracy Comparison in Securities Price Forecasting Between Neural Networks Model and EGARCH-M ARIMA Model," Master Thesis in Economics, Chiang Mai University, Chiang Mai, Thailand, 2006.

[4] S. Rimcharoen, D. Sutivong and P. Chongstitvatana, "Prediction of the Stock Exchange of Thailand Using Adaptive Evolution Strategies," Tools with Artificial Intelligence, 2005. ICTAI 05. 17th 2005.

[5] P. Sutheebanjard and W. Premchaiswadi, "Factors Analysis on Stock Exchange of Thailand (SET) Index Movement," The 7th International Conference on ICT and Knowledge Engineering, ICTKE2009, Bangkok, Thailand, December, 2009.

[6] P. Meesad and T. Srikhacha "Stock Price Time Series Prediction Using Neuro-Fuzzy with Support Vector Gudeline System," ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing 9th. 2008.

[7] P. Meesad and T. Srikhacha, "Universal Data Forecasting with an Adaptive Approach and Seasonal Technique,' Computational Intelligence for Modeling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, 2006.

[8] I. Rechenberg, "Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution," Dr.-Ing. Thesis, Technical University of Berlin, Department of Process Engineering, 1971.

[9] T. Back, F. Hoffmeister and H.P. Schwefel, "A Survey of Evolution Strategies," Proceeding of the Fourth Conference on Genetic Algorithm, 1991.

[10] H.G. Beyer and H.P. Schwefel, "Evolution Strategies-A Comprehensive Introduction," Natural Computing. Vol. 1, Issue 1, 2002.

[11] S. Rimcharoen and P. Chongstitvatana, "An Adaptation of Evolutionary Strategies for Forecasting the Exchange Rate," The 8th Annual National Symposium on Computational Science and Engineering, 2004.

[12] S. Chaigusin, C. Chirathamjaree, and J. Clayden, "Soft computing in the forecasting of the stock exchange of Thailand (SET)," Management of Innovation and Technology, 2008. ICMIT 2008. 4th, 2008.

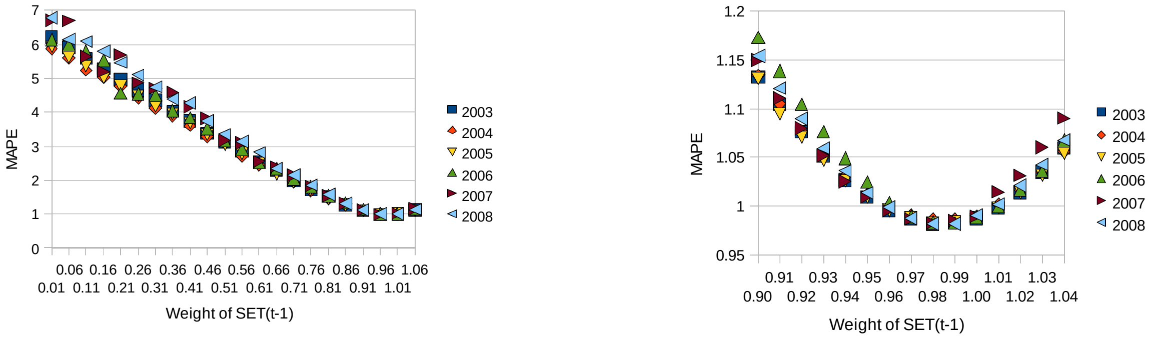
[13] C. Worasucheep, "A New Self Adaptive Differential Evolution: Its Application in Forecasting the Index of Stock Exchange of Thailand," Evolutionary Computation, 2007. CEC 2007, 2007.

[14] T. Tantinakom, "Economic Factors Affecting Stock Exchange of Thailand Index," Master Thesis in Economics, Chiang Mai University, Chiang Mai, Thailand, 1996.

[15] C. Khumyoo, "The Determinants of Securities Price in the Stock Exchange of Thailand," Master Thesis in Economics, Ramkhamhaeng University, Bangkok, Thailand, 2000.

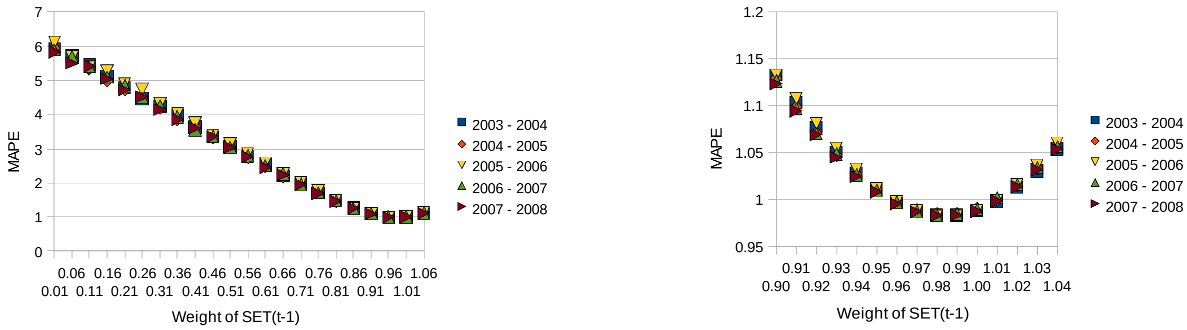
[16] S. Chotasiri, "The Economic Factors Affecting the Fluctuation of The Stock Exchange of Thailand Index," Master Thesis in Economics, Chiang Mai University, Chiang Mai, Thailand, 2004.

[17] K. Chaereonkithuttakorn, "The Relationship between the Stock Exchange of Thailand Index and the Stock Indexes in the United States of America," Master Thesis in Economics, Chiang Mai University, Chiang Mai, Thailand. 2005.



(a) (b)

Figure 2. MAPE of (2) at different a0 from January 2005 to March 2009 with one year training data
(a) scale from 0.01-1.06 (b) scale from 0.90-1.04.



(a) (b)

Figure 3. MAPE of (2) at different a0 from January 2005 to March 2009 with two years training data
(a) scale from 0.01-1.06 (b) scale from 0.90-1.04.

TABLE IV. THE MAPE OF ONE YEAR AND TWO YEARS TRAINING DATA TEST ON JANUARY 2005 – MARCH 2009

SET	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2003 - 2004**	**2004 - 2005**	**2005 - 2006**	**2006 - 2007**	**2007 - 2008**
0.90	1.1328	1.1341	1.1308	1.1732	1.1499	1.1539	1.1330	1.1296	1.1332	1.1265	1.1233
0.91	1.1041	1.1048	1.0946	1.1389	1.1109	1.1209	1.1035	1.0988	1.1078	1.0971	1.0945
0.92	1.0765	1.0767	1.0712	1.1046	1.0797	1.0895	1.0768	1.0715	1.0813	1.0700	1.0690
0.93	1.0519	1.0504	1.0469	1.0768	1.0521	1.0595	1.0508	1.0468	1.0551	1.0489	1.0451
0.94	1.0270	1.0307	1.0262	1.0492	1.0256	1.0360	1.0281	1.0271	1.0321	1.0266	1.0244
0.95	1.0093	1.0124	1.0092	1.0245	1.0097	1.0137	1.0113	1.0111	1.0114	1.0096	1.0083
0.96	0.9961	0.9984	0.9962	1.0040	0.9955	0.9991	0.9969	0.9971	0.9969	0.9972	0.9954
0.97	0.9869	0.9905	0.9881	0.9893	0.9869	0.9878	0.9877	0.9888	0.9878	0.9873	0.9871
0.98	**0.9824**	**0.9868**	**0.9832**	0.9833	**0.9830**	**0.9820**	**0.9833**	**0.9847**	0.9827	**0.9832**	**0.9831**
0.99	0.9826	0.9870	0.9833	**0.9827**	0.9854	0.9822	0.9835	0.9851	**0.9826**	0.9848	0.9839
1.00	0.9872	0.9900	0.9888	0.9884	0.9896	0.9909	0.9881	0.9902	0.9882	0.9896	0.9875
1.01	0.9982	1.0021	0.9986	0.9993	1.0148	1.0023	0.9987	1.0002	0.9993	1.0018	0.9989
1.02	1.0146	1.0172	1.0140	1.0173	1.0315	1.0212	1.0129	1.0151	1.0157	1.0159	1.0138
1.03	1.0354	1.0349	1.0316	1.0361	1.0603	1.0429	1.0306	1.0334	1.0361	1.0341	1.0320
1.04	1.0594	1.0561	1.0538	1.0675	1.0901	1.0681	1.0541	1.0546	1.0606	1.0563	1.0538