

Prediction of the Stock Exchange of Thailand Using Adaptive Evolution Strategies

Sunisa Rimcharoen, Daricha Sutivong and Prabhas Chongstitvatana
*Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University
Pathumwan Road, Bangkok 10330 Thailand
suni16@hotmail.com, daricha.s@chula.ac.th, prabhas@chula.ac.th*

Abstract

In this paper we present a prediction process of the Stock Exchange of Thailand index using adaptive evolution strategies. The prediction process does not require the knowledge of the functional form a priori. In each recursion step, genetic algorithm is used to evolve the structure of the prediction function, whereas the coefficient is evolved by evolution strategies. The proposed method has been shown to successfully predict the Stock Exchange of Thailand and returns an error less than 3%. This methodology is also a tool for knowledge discovery in a specific application. We have found that the SET index can be reasonably forecasted with only two factors: the Hang Seng index and Minimum Loan Rate. The proposed method also achieves a lower prediction error when compared with multiple regression method.

1. Introduction

Prediction of the stock exchange index is an important issue in business and finance. It helps investors to hedge against potential market risks and allows for market speculation. Predicting the stock exchange index is a challenging problem, as it is difficult to determine factors that influence the market and the functional form they take. This paper presents a method to create the prediction function without knowing the functional form. We introduce a method called an adaptive evolution strategy which is a combination of genetic algorithms and evolution strategies.

Genetic Algorithms (GA) [6] and Evolution Strategies (ES) [9] are branches of the Evolutionary Computation (EC) technique. These techniques are based upon the principle of natural evolution and the principle of the survival of the fittest as proposed by Charles Darwin [4]. Evolutionary computation techniques abstract these evolution principles into

algorithms. In an evolutionary algorithm, a representation scheme is chosen by a researcher to define a set of solutions that form the search space for the algorithm. The representation of GA is a fixed-length bit string and that of an ES is a fixed-length real-value vector. A number of candidate solutions are created and evaluated using a fitness function that is specific to the problem being solved. A number of solutions are chosen to be parents for the creation of new individuals or offsprings. The survivors are selected from the original population and the offsprings to form a new population of the next generation using their fitness values. An introduction to GA and ES can be found in [5, 1, 2].

In general, ES requires the knowledge of the functional form a priori. The algorithm then searches for coefficients of the function; for example, the ES can be used to find a set of weights of a linear function that predicts the future inflation rate based on the inflation rate of the previous quarter [7]. On the other hand, if we do not know a functional form, we can use a genetic algorithm to search for a prediction function. For example, Vijayan and Suresh [12] used genetic algorithm to find a forecasting model of linear time series. Chiraphadhanakul *et. al.* [3] used genetic algorithm to search for a model for forecasting commercial bank deposit.

In this work, we propose an adaptive ES with evolving functional forms. In order to search for the best prediction function, GA is used to encode the prediction function and evolves it via the evolution process. In the previous works [10, 11], we successfully use adaptive (1+1)-ES and $(\mu+\lambda)$ -ES to forecast the exchange rate with only time as a parameter. This paper presents an improved $(\mu+\lambda)$ -ES to predict Stock Exchange of Thailand (SET) index that incorporates the knowledge of relevant driving factors, such as other stock indices (Dow Jones, Hang Seng, Nikkei), gold price, Minimum Loan Rate (MLR), while the model still does not require the functional form of these factors. In contrast to a conventional

genetic algorithm and evolution strategy that search only for model or coefficients, the proposed technique can find a functional form and its coefficients concurrently.

The paper is organized as follows: Section 2 defines the stock exchange prediction problem. Section 3 describes the solution technique using an adaptation of evolution strategies. Experiment results and analysis are provided in Section 4. A conclusion is drawn in Section 5.

2. The Stock Exchange Prediction

The stock exchange prediction is complex and difficult because many factors influence the market. In case of the Stock Exchange of Thailand, Khumyoo [8] proposed the study of the factors that affect the SET index during two periods: from January 1994 to December 1995 and from January 1997 to December 1999.

For the period of January 1994 to December 1995, the outcome implied that Dow Jones index, gold price, Hang Seng index, exchange rate of Japanese yen and Thai baht, MLR, Nikkei index, oil price, Singapore SES ST time industrial index and Taiwan weighted index are statistically significant to the Stock Exchange of Thailand.

From the period of January 1996 to December 1999, the outcome implied that Dow Jones index, gold price, Hang Seng index, exchange rate of US dollar and Thai baht, money supply, MLR, interbank overnight rate and Nikkei index are statistically significant to the Stock Exchange of Thailand.

In this work the common factors of the two periods are extracted to form basis elements of the prediction function with an assumption that these factors probably still affect the market nowadays. These factors are:

- Dow Jones index
- Nikkei index
- Hang Seng index
- Gold price
- Minimum Loan Rate (MLR)

Even though many other factors may also influence the market, the proposed method can easily incorporate new factors into the prediction model by extending this set of basis elements.

In the prediction process, the parameters of the previous time period (yesterday) are used to predict today's stock exchange index as follows:

$$SET(t) = \mathcal{F}(t-1) \quad (1)$$

where $SET(t)$ is the stock exchange index at day t , and $\mathcal{F}(t-1)$ is a prediction function of day $t-1$.

The data in the experiment are obtained from the following two sources: The SET index, Dow Jones index, Nikkei index, Hang Seng index and Minimum Loan Rate from Bank of Thailand and gold price data from Gold Trader Association. The series investigated

in the experiment started from January 2003 to December 2004, containing 491 days of data. The data are divided into two groups: 420 days for training and 71 days for testing.

3. An Adaptive ES

This paper proposes using an adaptation of $(\mu+\lambda)$ -ES with an evolution of the functional form in order to predict the stock exchange index. The technique allows forecasting of the stock exchange index without knowing the functional forms of the driving factors mentioned in the previous section. The method involves mutation of the functional form via the evolution process. At the beginning, a function is randomly generated from primitive functions and arithmetic operators. Primitive functions are sine and cosine functions and arithmetic operators are +, -, *, /.

From the previous section, the driving factors of SET index are Dow Jones index, Nikkei index, Hang Seng index, gold price and Minimum Loan Rate. The following primitive elements are used by the adaptive evolution strategies.

- $a \times (\text{Dow Jones index})$
- $a \times (\text{Nikkei index})$
- $a \times (\text{Hang Seng index})$
- $a \times (\text{gold price})$
- $a \times (\text{Minimum Loan Rate})$
- $a \times \sin^b(c \times t)$
- $a \times \cos^b(c \times t)$

where a , b and c denote coefficients.

There are seven primitive elements and four arithmetic operators as basis elements of the adaptive evolution strategies. They are encoded using integer values, as shown in Table 1.

Table 1. Encoding of primitive elements

Code	Operator and Function
0	+
1	-
2	*
3	/
4	$a \times \text{Dow Jones index}$
5	$a \times \text{Nikkei index}$
6	$a \times \text{Hang Seng index}$
7	$a \times \text{gold price}$
8	$a \times \text{Minimum Loan Rate}$
9	$a \times \sin^b(c \times t)$
10	$a \times \cos^b(c \times t)$

To generate a functional form, a uniform random number generator is used to randomly select a function

and an operator from the set. Encoding of a prediction function is shown in Figure. 1. An individual has two chromosomes. The first chromosome is used to encode the coefficients which evolve via evolution strategies. The other chromosome is used to encode the functional form which evolves via genetic algorithm using mutation.

a_0	a_1	b_1	c_1	a_2	b_2	c_2	...	a_n	b_n	c_n	σ
f_1	op_1	f_2	op_2	f_3	...	op_{n-1}	f_n				

Figure 1. Encoding of an individual

In figure 1, a_0 is an inserted constant of the prediction function; n is a number of terms in the function; $a_1...a_n, b_1...b_n, c_1...c_n$ are coefficients of each term; $f_1...f_n$ are functional terms; $op_1...op_{n-1}$ are operators; σ is the variance used in controlling the search strategy.

The following example illustrates how a function is generated. Starting from the first symbol, suppose 4 is selected; the first term of a function is $a \times (\text{Dow Jones index})$. Next, an operator is chosen, suppose it is 0, the operator is '+'. As the operator requires two operands, another function is generated; suppose it is $a \times \sin^b(c \times t)$. Once the function is selected, its coefficients are initialized with random real-values. The number of coefficients depends on the number of terms in the function. In this example, the prediction function is $f(t) = 0.81 + 0.15 \times (\text{Dow Jones index}) + 0.27 \times \sin(0.53 \times t)$. The encoding of this function is shown in Figure 2.

0.81	0.15	1	1	0.27	1	0.53	σ
a_0	a_1	b_1	c_1	a_2	b_2	c_2	
4		0		9			
f_1		op_1		f_2			

Figure 2. Example of a functional form

After a prediction function is initialized, the evolution process begins. The fitness function is used to evaluate an individual. In this work, Mean Absolute Percentage Error (MAPE) is used for evaluation in order to minimize an error of fitting between the prediction function and the training data set (eq.2).

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{g(t_i) - f(t_i)}{g(t_i)} \right|}{n} \times 100 \quad (2)$$

where $g(t)$ is an actual value.
 $f(t)$ is a forecasted value calculated from the prediction function.
 t is time.
 n is the number of data points.

In controlling the search, an adjustment of variance is considered. The mutation of a real value coefficient is accomplished by sampling a value from a normal distribution $Z_i \sim N_i(0, \sigma^2)$ and add that to the coefficient. The mutation of the functional form changes a function or an operator to another function or another operator at a random position. Adaptive ES is implemented in following steps:

1. Randomly select μ parent vectors, each contains 2 chromosomes ($chrom_1, chrom_2$) of coefficients and functions as shown in figure 1.
2. Create λ new offsprings by mutation. Generate the offspring population with following steps:
 - 2.1 Update step-size
 $\sigma' = \sigma \times \text{Exp}(N(0, \Delta\sigma^2))$
where $\Delta\sigma$ denotes learning rate
 - 2.2 Mutate a real-value coefficient
 $chrom_1' = chrom_1 + N(0, \sigma'^2)$
 - 2.3 Mutate the functional form ($chrom_2$)
 $chrom_2' = \text{mutate}(chrom_2)$
Change a function or an operator to another function or operator at a random position.
3. Evaluate the fitness of each individual using MAPE. Select μ most fit solutions for the next generation from all parents and offsprings.
4. Repeat step 2 through 3 until a satisfactory solution is found or the computation time is exhausted.

4. Experiment Results and Analysis

4.1. Prediction by Adaptive ES

The basic elements are from Table 1. The initial functions are randomly generated. Parameters of the algorithm are shown in Table 2.

Table 2. Parameters in the experiment

Parameter	Value
Population size ($\mu + \lambda$)	(80 + 100)
The number of generations	1000
Probability of the functional form mutation	0.3
The number of terms in the prediction function	10

The resulting function from the experiment is shown below, which is the best function of 10 runs:

$$\begin{aligned} SET(t) = & 2.3645 + 5.5208\sin^3[0.3138(t-1)] - \\ & 1.5430\text{hangseng}(t-1) / -5.2054\text{mlr}(t-1) + \\ & 2.8360\cos^5[0.6246(t-1)] * 4.6811\sin[0.3651(t-1)] - \\ & 1.5380\cos^3[0.7522(t-1)] - 1.1618\cos^3[0.7724(t-1)] + \\ & 3.3228\sin^3[1.5317(t-1)] - 2.4620\cos[0.6676(t-1)] * \\ & 2.3144\text{mlr}(t-1) \end{aligned} \quad (3)$$

The prediction function in eq. (3) returns an error of 2.91% on the training data and 2.46% on the testing data. Fig. 3 (a) shows predicted values from the function in eq. (3) compared with the training data. Figure 3 (b) shows predicted values from the function in eq.(3) compared with the testing data.

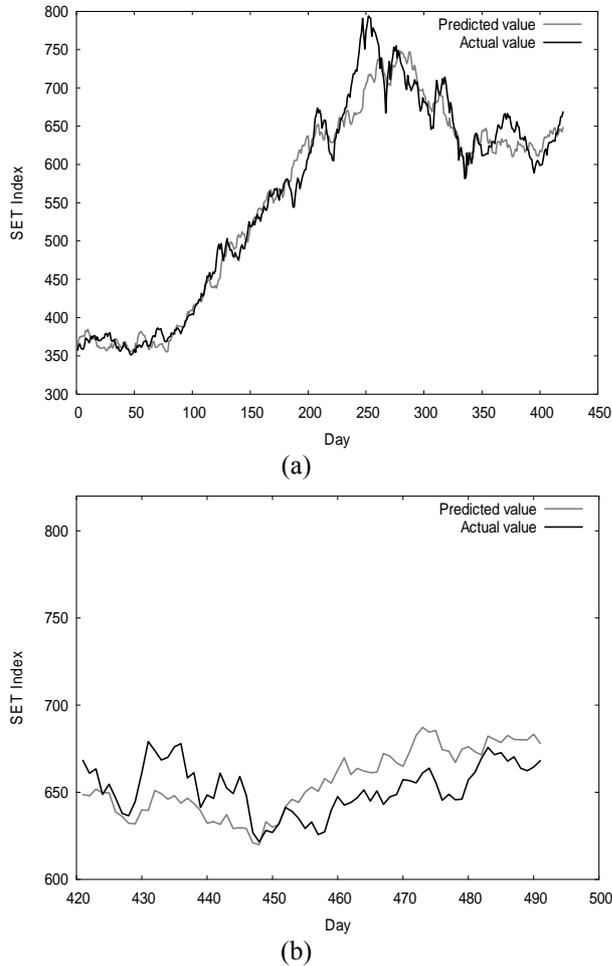


Figure 3. (a) Predicted values and the training data, (b) Predicted values and the testing data

4.2. Prediction by Multiple Regression

In this section the proposed method are compared with multiple regression method. The results from the multiple regression are obtained from SPSS program (eq.4).

$$\begin{aligned}
 SET(t) = & 207.257 + 0.047dowjones(t-1) - \\
 & 0.014nikkei(t-1) + 0.047hangseng(t-1) + \\
 & 0.013gold(t-1) - 90.649mlr(t-1)
 \end{aligned}
 \tag{4}$$

The function in eq.(4) returns the error of 4.51% on the training data and 13.86% on the testing data. Figure 4 (a) shows the prediction function in eq.(4) compared with the training data. Figure 4 (b) shows predicted values from the regression function in eq.(4) compared with the testing data.

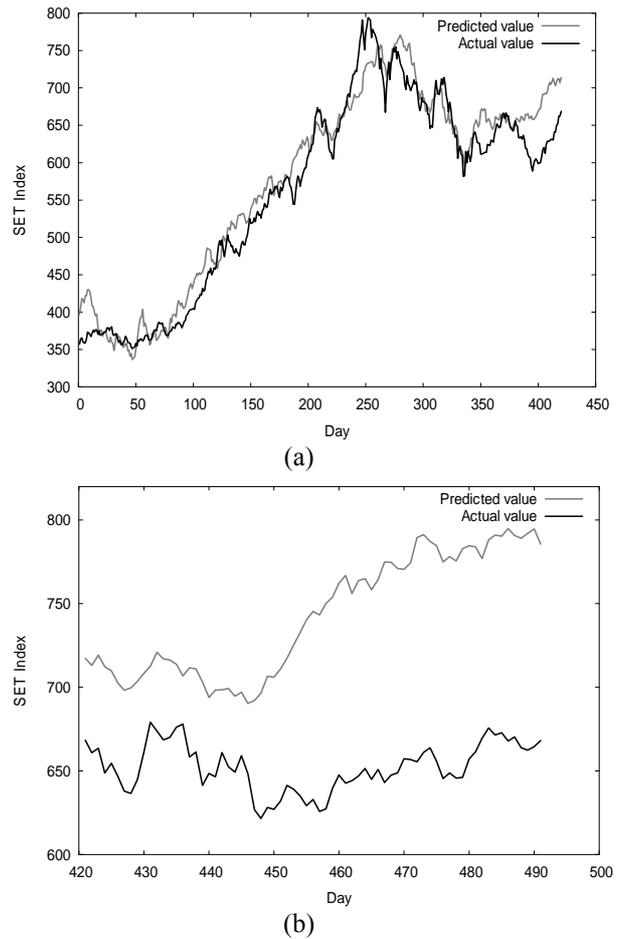


Figure 4. (a) Predicted values from multiple regression and the training data, (b) Predicted values from multiple regression and the testing data

4.3. Analysis

The experiments in section 4.1 and 4.2 show the errors on the training data of 2.91% for the adaptive ES and 4.51% on the multiple regression. The prediction function from both methods are used to forecast the future value of SET index and compare them with the testing data set. The prediction function from the adaptive ES returns an error of 2.46%, while the regression function returns an error of 13.86%. The experimental results show that the prediction function from the adaptive ES returns smaller errors on the training and the testing data than the regression function for the time interval studied.

From Figure 3(a), the proposed algorithm is able to generate a prediction function that captures the movement of the SET index reasonably well and without any knowledge of the functional form a priori. Figure 3(b) also shows that this prediction function can forecast the market movement quite closely, that is within an average error of 2.5%. Even though we wish to achieve a smaller error in financial applications, we note that the prediction function can at least forecast

the index trend (up or down movements) quite accurately, which in itself is sufficient information for day-to-day trading decision to buy or sell.

We also observe from the prediction function in eq. (3) that the significant factors of the Stock Exchange of Thailand index are Hang Seng index and MLR. From five driving factors, which are Dow Jones index, Nikkei index, Hang Seng index, gold price and MLR, the algorithm automatically eliminates Dow Jones index, Nikkei index and gold price during the evolution process. In order to verify the impact of the two factors – Hang Seng index and MLR – on the SET index, we plot the prediction function with only the term $-(1.5430 \cdot \text{hang_seng}) / (-5.2054 \cdot \text{mlr})$ against the actual SET index in Figure 5. Note from the graph that the SET index can be reasonably described with only these two factors.

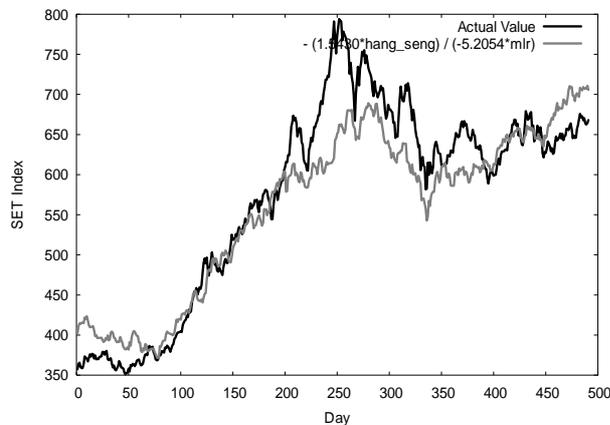


Figure 5. The graph of $-(1.5430 \cdot \text{hang_seng}) / (-5.2054 \cdot \text{mlr})$

The proposed method allows the evolution process to select appropriate functions and driving factors. In our case, we discover that the SET index can be mostly explained by only the Hang Seng index and MLR, while many other relevant factors are not necessary. The practical reasons might arise from the fact that Hong Kong is close to Thailand geographically and economically. Thus, the two markets highly correlate and the movements of Hong Kong market can adequately describe the Thai market. The Minimum Loan Rate (MLR) is obviously an important factor of investment decisions in any market.

5. Conclusions

This paper proposes an adaptation of evolution strategies with evolving functional form and coefficients to predict the Stock Exchange of Thailand index. The potential parameters that drive the stock exchange index, namely Dow Jones index, Nikkei index, Hang Seng index, gold price and MLR are used in the algorithm. By using the proposed adaptive evolution strategies, it is not necessary to determine the functional form of the prediction function a priori. Experimental data have shown that this method can be

effectively used to forecast the Stock Exchange of Thailand index with error less than 3%. We have found that the SET index can be adequately explained with only two major factors, the Hang Seng index and MLR. We test our method and compare the results with multiple regression method. The results have shown that our method can find a prediction function that returns a smaller error than the regression method for the period investigated.

6. References

- [1] Back, T., Hoffmeister, F. and Schwefel, H. P. "A Survey of Evolution Strategies," Proceeding of the Fourth Conference on Genetic Algorithm, 1991.
- [2] Beyer, H.G. and Schwefel, H.P. "Evolution Strategies-A Comprehensive Introduction," Natural Computing, Vol. 1, Issue 1, 2002.
- [3] Chiraphadhanakul, S., Dangprasert, P. and Avatchanakorn, V. "Genetic Algorithm in Forecasting Commercial Banks Deposit," IEEE International Conference on Intelligent Processing Systems, 1997.
- [4] Darwin, C. "On the Origin of Species by means of Natural Selection," John Murray, 1859.
- [5] Goldberg, D. E. "Genetic Algorithms in Search, Optimization," and Machine Learning, Addison-Wesley, 1989.
- [6] Holland, J. H. "Adaptation in Natural and Artificial Systems," University of Michigan Press, 1975.
- [7] Kendall, G., Binner, J. M. and Gazely, A. M. "Evolutionary strategies: A new macroeconomic policy tool?," IFAC Symposium on Modeling and Control of Economic Systems, 2001.
- [8] Khumyoo, C. "The Determinants of Securities Price in the Stock Exchange of Thailand," Master Thesis in Economics, Ramkhamhaeng University, 2000.
- [9] Rechenberg, I. "Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution," Stuttgart: Frommann-Holzboog Verlag, 1973.
- [10] Rimcharoen, S. and Chongstitvatana, P. "An Adaptation of Evolutionary Strategies for Forecasting the Exchange Rate," The 8th Annual National Symposium on Computational Science and Engineering, 2004.
- [11] Rimcharoen, S., Sutivong, D. and Chongstitvatana, P. "Curve Fitting Using Adaptive Evolution Strategies for Forecasting the Exchange Rate," The 2nd Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology Annual Conference, 2005.
- [12] Vijayan, P., Suresh, S. "A Novel Evolutionary Approach to Linear Time-Series Forecasting Model," International Conference on Computational Science, 2003.