ASIAN ACADEMY *of*
MANAGEMENT JOURNAL
of ACCOUNTING
and FINANCE

AAMJAF, Vol. 6, No. 2, 85–114, 2010

FORECASTING THE THAILAND STOCK MARKET USING EVOLUTION STRATEGIES

Phaisarn Sutheebanjard* **and Wichian Premchaiswadi**

Graduate School of Information Technology,
Siam University Bangkok 10163, Thailand

*Corresponding author: mr.phaisarn@gmail.com

ABSTRACT

This paper proposes a new prediction function for the Stock Exchange of Thailand (SET index). Included in the proposed prediction function are the important economic factors: namely, the Dow Jones, Nikkei, and Hang Seng indexes; the minimum loan rate (MLR); and the previous SET index. The tuning coefficients of each factor in this research were calculated using the two-membered evolution strategy (ES) technique. The experiment was conducted by analysing the SET index during three different time periods. The first time period extended from January 2004 to December 2004, and the second time period extended from 9 August 2005 to December 2005. These data were used to evaluate the performance of the proposed prediction function for short-term periods by comparing the results with those achieved using the existing methods. Lastly, the long-term period data extending from January 2005 to March 2009, which covered 1040 days in totals, were used to predict the SET index. The results show that the proposed prediction function not only yields the lowest mean absolute percentage error (MAPE) for short-term periods but also yields a MAPE lower than 1% for long-term periods.

Keywords: Stock market forecasting, stock exchange of Thailand, evolution strategies, prediction function, data analysis

INTRODUCTION

Stock exchange index prediction is an interesting and challenging issue for both investors and academics. The stock market is a highly nonlinear dynamic system. Many factors influence stock market performance, including interest rates, inflation rates, economic environments, political issues, and many others. The Stock Exchange of Thailand (SET) in particular has its own unique characteristics based on to the economic system that it serves, which includes the Dow Jones, the Nikkei, the Hang Seng, gold prices, the minimum loan rate (MLR), the value of the Thai baht and many other factors, as indicated in Chaigusin, Chirathamjaree and Clayden (2008a, 2008b), Rimcharoen and Chongstitvatana (2004), Sutheebanjard and Premchaiswadi (2009), Chaereonkithuttakorn (2005), Chotasiri (2004) and Khumyoo (2000). Therefore, this research has used the historical movement of the SET index itself together with data on the world's major stock market indices [which include the Dow Jones (New York), the Nikkei (Japan), the Hang Seng (Hong Kong)] and the domestic minimum loan rate (MLR) as the factors in the prediction function for the SET index. Of course, the Thai economy is constantly changing. The factors influencing the Thai stock market may be different in different time periods (Khumyoo, 2000). Thus, this research intends to determine the influential factors that have played a role in recent years, specifically from January 2005 to March 2009.

Previous studies have worked to predict the SET index. Towards this end, they used the technique of neural networks (Chaigusin et al., 2008b; Pattamavorakun & Pattamavorakun, 2007; Tunsenee, 2006), ARIMA (Tunsenee, 2006) and evolution strategies (Rimcharoen et al., 2005; Sutheebanjard & Premchaiswadi, 2009). Neural networks are very efficient adaptive forecasting models because of their excellent self-learning capabilities. Unlike other techniques that construct functional forms to represent the relationships between data, neural networks are able to learn patterns or relationships from data themselves (Chaigusin et al., 2008b). However, due to the effects of black-box, slow convergence, local optimal, they are not applicable for some applications (Meesad & Srikhacha, 2008). The autoregressive integrated moving average (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) (Meesad & Srikhacha, 2006). The ARIMA approach is elegant in theory but has been of little practical use in business because of its complexity and the limited increase in accuracy that it provides compared to less sophisticated methods. Evolution strategies were introduced by Rechenberg in 1971. Evolution strategies (ESs) are algorithms that imitate the principles of natural evolution to solve parameter optimisation problems (Back, Hoffmeister & Schwefel 1991; Beyer & Schwefel, 2002; Rimcharoen & Chongstitvatana, 2004). ESs are one of the most popular evolutionary algorithms and are generally used in numerical optimisation for real valued representation. This study investigated impact factors for the SET Index using evolution strategies (Sutheebanjard & Premchaiswadi, 2009). This study also compared the different time periods for the training data to evaluate what time period was suitable to use for the training model.

This paper is organised as follows. The next section describes stock market prediction theories and methods, followed by introduction of the existing work on Stock Exchange of Thailand index prediction. Subsequent section describes the technique used to determine the solution based on evolution strategies, and the experimental results. The last section provides conclusions.

STOCK MARKET PREDICTION

Stock Market Prediction Theories

In predictions of stock market movement, two theories have had a significant impact on market research: the efficient market hypothesis (EMH) and random walk theory.

Efficient market hypothesis

In 1965, Fama developed the EMH. In the EMH, the price of a security reflects complete market information. Whenever a change in financial outlook occurs, the market will instantly adjust the security price to reflect the new information. This theory is also highly controversial and often disputed. The supporters of this model believe that it is pointless to search for undervalued stocks or try to predict trends in the market through fundamental analysis or technical analysis. The EMH involves three different levels of information-sharing: the weak form, the semi-strong form and the strong form. In weak EMH, only historical information is embedded in the current price. The semi-strong form goes a step further by incorporating all historical and currently public information into the price. The strong form includes historical and current public information as well as private information such as insider information in the share price. However, Fama (1991) state that the strong form of the EMH was invalid. Bernstein (1999) suggested that "either the hypothesis has an inherent flaw, or Wall Street and its customer base are in truth totally irrational".

Random walk theory

A random walk is one in which future steps or directions cannot be predicted on the basis of past actions. When this term is applied to the stock market, it indicates that short-run changes in stock prices cannot be predicted. Investment advisory services, earnings predictions, and complicated chart patterns are useless (Malkiel, 1999). The stock's future prices take a random and unpredictable path. Both technical analysis and fundamental analysis are largely considered useless and unproven in outperforming the markets. Random walk theory was originally examined by Kendall and Hill (1953). The theory states that the movement of shares on the stock market is random, i.e., they are as likely to go up on a certain day as down. When first presented, these results were disturbing to some financial economists, and further debate and research then followed. This debate ultimately led to the creation of the random walk hypothesis and the closely related efficient-market hypothesis, which states that random price movements indicate a well-functioning or efficient market. Random walk theory has theoretical underpinnings similar to those of semi-strong EMH, in which all public information is assumed to be available to everyone. However, random walk theory indicates that even with such information, future prediction is ineffective (Schumaker & Chen, 2009).

Malkiel (2003) suggested that the efficient market hypothesis is associated with the idea of a "random walk", which characterises a price series in which all subsequent price changes represent random departures from previous prices. The logic of the random walk idea is that if the flow of information is unimpeded and information is immediately reflected in stock prices, then tomorrow’s price change will reflect only tomorrow’s news and will be independent of the price changes today. However, news is by definition unpredictable, and thus, the resulting price changes must be also unpredictable and random.

Stock Market Prediction Approaches

The efficient market hypothesis and random walk theory have discouraged the prediction of future stock prices. However, there are still many methods intended to make such predictions. These methods can be grouped into two diametrically opposed approaches: fundamental and technical analysis.

Fundamental analysis

Fundamental analysis is a research method that involves studying basic financial information to forecast profits, supply and demand, industry strength, management ability, and other intrinsic factors that affect a stock's market value and growth potential (Thomsett, 1998). Fundamental analysis is concerned with the company that underlies the stock itself. The resulting information can help one to maintain perspective. Certainly some aspects of price movement in the stock market seem illogical. Fundamental analysis also considers both global and domestic economic factors that can influence stock prices (Chaigusin et al., 2008a), including exchange rates, foreign financial markets (global factors) and gross domestic product (GDP), employment, inflation rates, interest rates, public sentiment and the MLR (domestic factors). The goal of analysing a company's fundamentals is to determine a stock's intrinsic value, as opposed to the value at which it is being traded in the marketplace. If the intrinsic value is more than the current share price, then the analysis shows that the stock is worth more than its price indicates and that it makes sense to buy the stock.

Technical analysis

Technical analysis aims to predict financial price movements using information sets limited to a few variables such as past prices (Osler, 2000). Technical analysts use a basic approach to stock investing that involves studying past prices using charts. Technical analysis is not concerned with company fundamentals. Analysts seek to determine the future price of a stock based solely on the (potential) trends reflected in the past price (this is a form of time series analysis). Numerous patterns are employed, including the head and shoulders or the cup and saucer. Together with these patterns, statistical techniques, such as time series analysis, are used. Technical analysts test historical data to establish specific rules for buying and selling securities with the objective of maximising profit and minimising risk of loss. Technical trading analysis is based on two main premises. First, the market's behaviour patterns do not change much over time; this is particularly true of long-term trends. It is assumed that the patterns in market prices will continue to manifest in the future and that these patterns can therefore be used for predictive purposes. From 1884 on, Charles Dow published his ideas about stock market trends in a series of editorials he wrote for the *Wall Street Journal*, which has been known collectively as the Dow Theory. Building on the Dow Theory, the theory includes principles such as that prices follow trends; that prices discount all known information, confirmation and divergence; that volume mirrors changes in price, and that support/resistance is encountered. Of course, the widely followed Dow Jones Industrial Average is a direct offshoot of the Dow Theory.

Both fundamental and technical analysis use trends, but they use them in different ways. The fundamental approach employs historical information (for example, dividend rates, profits, or sales) to forecast financial results. The technical approach uses trends. Technical analysis is highly visual and largely ignores the basic premise of supply and demand, suggesting, instead, that recent price trends (shown in charts) dictate future price movement (Thomsett, 1998).

STOCK EXCHANGE OF THAILAND

History of Stock Exchange of Thailand (SET)

The Thailand stock market officially started trading on 30 April 1975 and was named "The Securities Exchange of Thailand". On 1 January 1991, the exchange's name was formally changed to "Thai Stock Exchange of Thailand". The index of the Stock Exchange of Thailand is called the SET Index. The SET Index is a composite market capitalisation-weighted price index that compares the current market value (CMV) of all listed common stocks with their market value on the base date of 30 April 1975 (base market value or BMV), which was when the stock market was established. The initial value of the SET index on the base date was set to 100 points. The formula for calculating the SET index is as follows:

$$\text{SET Index} = \frac{\text{Current Market Value} \times 100}{\text{Base Market Value}} \tag{1}$$

Impact Factors to Stock Exchange of Thailand (SET)

In economic environments, both global and domestic economic factors can influence stock prices (Chaigusin et al., 2008a). Because countries are linked together, movement on one stock market may have an impact on other stock markets. In developing prediction models for the Thai stock market index, the choice of selection input data is important. 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 (Chaigusin et al., 2008a). Examples of factors that influence the Thai stock market are the foreign stock index, the value of the Thai baht, oil prices, gold prices, the MLR and many others (Rimcharoen et al., 2005; Worasucheep, 2007; Chaigusin et al., 2008b; Chaereonkithuttakorn, 2005; Chotasiri, 2004; Khumyoo, 2000; Tantinakom, 1996). Some researchers have used these factors to forecast the SET index, including Tantinakom (1996), who used trading value, trading volume, interbank overnight rates, inflation, the net trading value of investment, the value of the Thai baht, the price-earnings ratio, the Dow Jones index, the Hang Seng index, the Nikkei index, the Straits Times Industrial index and the Kuala Lumpur Stock Exchange Composite index. Khumpoo (2000) used the Dow Jones index, gold prices, the Hang Seng index, the exchange rate for the Japanese yen and Thai baht, the MLR, the Nikkei index, oil prices, the Straits Times Industrial index and the Taiwan weighted index. Chotasiri (2004) used the interest rates for Thailand and the US; the exchange rates for the USD, JPY, HKD and SKD; the stock exchange indices of the US, Japan, Hong Kong and Singapore; the consumer price index; and oil prices. Chaereonkithuttakorn (2005) used US stock indices, including the Nasdaq index, the Dow Jones index and the S&P 500 index. Rimcharoen et al. (2005) used the Dow Jones index, the Nikkei index, the Hang Seng index, gold prices and the MLR. Worasucheep (2007) used MLR, the exchange rate for Thai baht and the USD, daily effective over-night federal fund rates in the US, the Dow Jones index and oil prices. Chaigusin et al. (2008) used the Dow Jones index, the Nikkei index, the Hang Seng index, gold prices, the MLR and the exchange rate for the Thai baht and the USD. The common factors that researchers used to predict the SET index are summarised in Table 1.

Rimcharoen et al. (2005) proposed adaptive evolution strategies with an evolving functional form and coefficients for use in predicting the Stock Exchange of Thailand index. The potential parameters that drive the stock exchange index, namely the Dow Jones index, the Nikkei index, the Hang Seng index, gold prices and the MLR, were used in the algorithm based on the adaptive evolution strategies. The adaptive evolution strategies method is a combination of genetic algorithms (GA) and evolution strategies ($\mu+\lambda$)–ES. The genetic algorithm (Holland, 1992) and evolution strategies (Schwefel, 1975) are branches of evolutionary computation (EC). The genetic algorithm was used to randomly select the structure of the prediction function, whereas the coefficient is calculated via evolution strategies. The study was based on daily data from January 2003 to December 2004. It is not necessary to determine the functional form of the prediction function a priori. The experimental data show that their method can be effectively used to forecast the SET index with error less than 3%. The function developed by Rimcharoen et al. is shown in equation (2).

$$\begin{aligned}\mathrm{SET}_{(t)} = {} & 2.3645 + 5.5208\sin^3[0.3138\mathit{SET}_{(t-1)}] - \\ & 1.5430\mathit{HangSeng}_{(t-1)} / -5.2054\mathit{MLR}_{(t-1)} + \\ & 2.8360\cos^2[0.6246\mathit{SET}_{(t-1)}] * 4.6811\sin[0.3651\mathit{SET}_{(t-1)}] - \\ & 1.5380\cos^3[0.7522\mathit{SET}_{(t-1)}] - 1.1618\cos^3[0.7724\mathit{SET}_{(t-1)}] + \\ & 3.3228\sin^3[1.5317\mathit{SET}_{(t-1)}] - 2.4620\cos[0.6676\mathit{SET}_{(t-1)}] * \\ & 2.3144\mathit{MLR}_{(t-1)}\end{aligned} \quad (2)$$

where (t) is today and (t–1) is yesterday.

They eventually found that the SET Index can be adequately explained by only two major factors, the Hang Seng index and MLR, as in equation (3). The graph of equation (3) is shown in Figure 1.

$$SET_{(t)} = \frac{1.5430 HangSeng_{(t-1)}}{5.2054 MLR_{(t-1)}} \quad (3)$$

Table 1
Impact Factors for Stock Exchange of Thailand Index

	Tantinakom (1996)	Khumyoo (2000)	Chotasiri (2004)	Chaereon-kithuttakorn (2005)	Rimcharoen et al. (2005)	Worasucheep (2007)	Chaigusin et al. (2008)
Nasdaq Index				x			
Down Jones Index	x	x	x	x	x	x	x
S&P 500 Index				x			
Nikkei Index	x	x	x		x		x
Hang Seng Index	x	x	x		x		x
Straits Times Industrial Index	x	x	x				
USD		x	x			x	x
JPY		x	x				
HKD			x				
SKD			x				
Gold prices		x			x		x
Oil prices		x	x			x	
MLR		x			x	x	x

Notes: USD indicates the exchanges rate for the Thai baht and the US dollar
JPY is the exchange rate for the Thai baht and Japanese yen
HKD is the exchange rate for the Thai baht and Hong Kong dollar
SKD is the exchange rate for the Thai baht and Singapore dollar

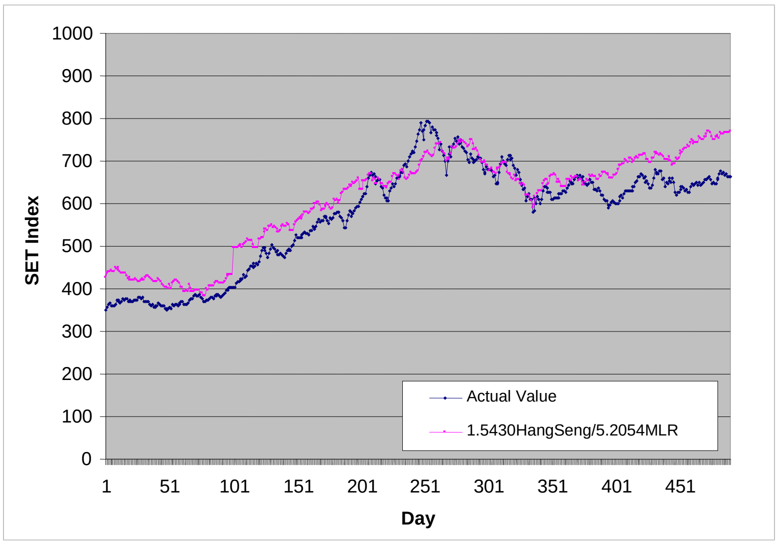

Figure 1. Graph of the SET Index 2003–2004 against (1.5430 HangSeng/5.2054MLR) Rimcharoen et al. (2005).

Chaigusin et al. (2008a) show the SET index plot against the two major terms from function equation (3) using data from January 2005 to December 2006, as shown in Figure 2.

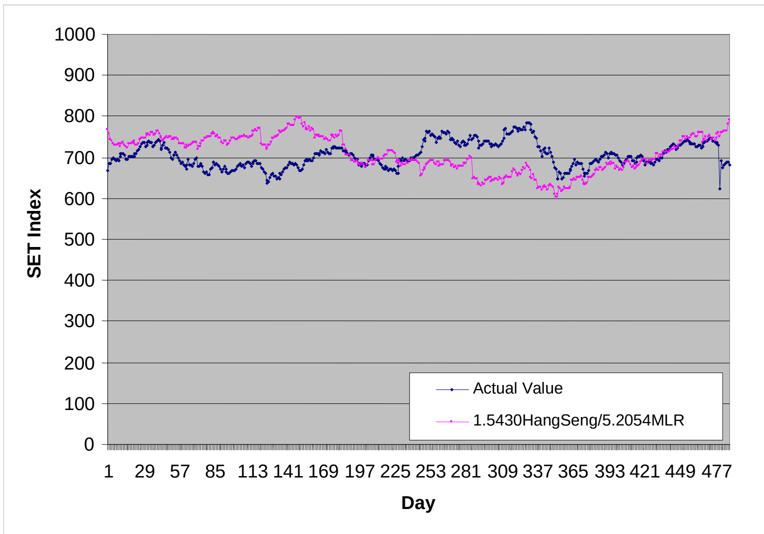

Figure 2. Graph of the SET Index 2005–2006 against (1.5430HangSeng/5.2054MLR) Chaigusin et al. (2008a).

These two figures indicate that the SET index can be reasonably described by the Hang Seng index and the MLR for 2003 and 2004. However, from 2005 to 2006, the SET index and 1.5430HangSeng/5.2054MLR moved in different directions. The data for the first 420 days from 2003 to 2004 were used as training data (Rimcharoen et al., 2005), and the SET index in this period is well explained by the term 1.5430HangSeng/5.2054MLR.

Tunsenee (2006) compared the performance of feed-forward neural networks with that of ARIMA with EGARCH-M models in terms of accuracy in predicting the SET index. The neural networks and ARIMA with EGARCH-M models yielded MAPE values of 1.2956 and 0.5972, respectively. He suggests that the neural network model performs more poorly in terms of its predictive capability because of a drawback in model construction: the overuse of data entry and of neurons in the hidden layer.

Chaigusin et al. (2008b) predicted the SET index using multilayer feed-forward back-propagation neural networks. The three suitable neural network models identified in this research were a three-layer (7-3-1), a four-layer (7-7-3-1) and a five-layer (7-13-7-3-1) neural network. Their prediction performance figures, measured using the MAPE, were 1.26594, 1.14719 and 1.14578, respectively.

This research takes into account both internal and external factors in forecasting the SET index. The external factors are foreign major stock market indices, whereas the internal factors are the SET index and domestic MLR. The assumption is that both the external and the internal factors probably have a great impact on the SET index. These factors include the following:

(i) The SET Index (Thailand)
(ii) The Dow Jones index (New York)
(iii) The Nikkei index (Japan)
(iv) The Hang Seng index (Hong Kong)
(v) The Minimum Loan Rate (MLR)

This research proposes a prediction function that does not require the genetic algorithm as in Rimcharoen et al. (2005). The *two-membered* evolution strategy (ES) technique has been used to calculate the tuning coefficients of each factor.

EVOLUTION STRATEGIES

Evolution strategies (ESs) are one of the main branches of evolutionary computation. Like genetic algorithms (Holland, 1992), ES are algorithms that imitate the principles of natural Darwinian evolution, generally producing consecutive generations of samples. In each generation, a batch of samples is generated by perturbing the parents' parameters through mutation of their genes. A number of samples are selected based on their fitness values, and 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 across generations.

Rechenberg (1971) proposed ES as a tool for use in real value parameter optimisation problems. In ES, the representation used was an n-dimensional real-valued vector. A vector of real values represented an individual. The standard deviation was used to control the search strategy in ES. Rechenberg used Gaussian mutation as the main operator in ES, in which a random value from a Gaussian distribution (normal distribution) was added to each element of an individual's vector to create 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 individuals as float-valued vectors instead of using binary representation. This type of representation reduces the burden of converting genotype to phenotype during the evolution process.

The evolutionary strategies (ESs) introduced by Ingo Rechenberg (1971, 1973) were $(1 + 1)$–ES and $(\mu + 1)$–ES. Two further versions introduced by Schwefel were $(\mu + \lambda)$–ES and (μ, λ)–ES Schwefel, (1975, 1977).

(i) $(1 + 1)$–ES or two-membered ES is the simplest form of ES. One parent creates one *n*-dimensional real-valued vector of object variables using mutation with identical standard deviations for each object variable. The resulting individual is evaluated and compared to its parent, and the better of both individuals survives to become the parent of the next generation, whereas the other is discarded.

(ii) $(\mu + 1)$–ES or steady-state ES is the first type of a *multi-membered* ES. There are μ parents at time $(\mu > 1)$ in which one child is created from μ parents. In $(\mu + 1)$–ES, μ parent individuals are recombined to form one offspring, which also underlines mutation. The best of the three is selected as the new current solution, be it the offspring or one of the parents, thus keeping the population size constant.

(iii) $(\mu + \lambda)$–ES, in which $\lambda \geq 1$ descendants are created at a time or in a generation, but to keep the population size constant, the λ worst out of all $\mu+\lambda$ individuals are discarded.

(iv) (μ, λ)–ES, in which the selection process includes the λ offspring only, whereas their parents are "forgotten" no matter how good or bad their fitness is compared to that of the new generation. Obviously, this strategy relies on a birth surplus – i.e., on $\lambda > \mu$ in a strict Darwinian sense.

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) that produces one offspring by adding normal distribution random numbers. The better of the two individuals then serves as the ancestor in the next iteration/generation. The (1 + 1)–ES is used to find the coefficients of the function. First, the coefficient of the prediction function is initialised by mutation operation. Then, each child is evaluated using the fitness function for a possible solution in each generation. These evaluations are used to create a new generation.

In the prediction process, the parameters from the previous time period (yesterday) are used to predict today's SET Index as equation (4).

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

where $SET_{(t)}$ is the SET Index at day t, and $F_{(t-1)}$ is a prediction function of day $t-1$.

This research has proposed a new prediction function for the SET index in which the important economic factors are the Dow Jones, Nikkei, and Hang Seng indexes, the domestic MLR and the previous SET index, as shown in equation (5).

$$SET_{(t)} = a_0 SET_{(t-1)} + a_1 \left(\frac{a_2 DJ_{(t-1)} + a_3 NK_{(t-1)} + a_4 HS_{(t-1)}}{a_5 MLR_{(t-1)}} \right) \tag{5}$$

where $a_0 - a_5$ denote coefficients.

SET is the SET index (Thailand)
DJ is the Dow Jones index (New York)
NK is the Nikkei index (Japan)
HS is the Hang Seng index (Hong Kong)
MLR is the minimum loan rate (MLR)

The mutation operator plays a significant role in global searches and fine-tuning for the ES. It is observed that smaller changes occur more often than larger ones in biological evolution. This type of change in a child can be made easily using a zero-mean Gaussian random number function. The child vector is defined by the mutation operation of a real value coefficient by sampling a real value from the normal distribution and adding it to the coefficient as shown in equation (6)

$$a_c = a_p + N(0,\sigma^2) \tag{6}$$

where a_p is a parent coefficient
a_c is a child coefficient
$N(0,\sigma^2)$ is a Gaussian random (normal distribution) number vector standard deviation σ denotes the standard deviation of the system

In controlling the search strategy for the simple (1 + 1)–ES for two basic objective functions using the convergence rate expressions, Rechenberg derived the optimal value of the single standard deviation (σ') of the mutation operator in a (1 +1)–ES. This rule is called the 1/5-success rule (Rechenberg, 1973), reflecting theoretical results indicating that, on average, one out of five mutations should cause an improvement in objective function values to achieve the best convergence rate as shown in equation (7).

$$\sigma' = \begin{cases} \sigma / 0.817 & \text{if}\,(p > 1/5) \\ \sigma \cdot 0.817 & \text{if}\,(p < 1/5) \\ \sigma & \text{if}\,(p = 1/5) \end{cases} \tag{7}$$

To minimise fit error between the prediction function and the actual value, this research used mean squared errors (MSEs) as the fitness function as shown in equation (8).

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(g_{(t)} - f_{(t)})^2 \tag{8}$$

where $g_{(t)}$ is an actual value
$f_{(t)}$ is a forecasted value
n is the number of data points

The parent vector of the next generation is determined by comparing the objective function W_p for the parent vector and W_c for the child vector (selection) as shown in equation (9) (Horii, Takahashi & Narita, 2000).

$$x'_p = \begin{cases} x_c & (W_c \le W_p) \\ x_p & (W_c > W_p) \end{cases} \tag{9}$$

denotes that if the objective function W_c of the child vector X_c is smaller than W_p for the parent vector X_p, then the child vector X_c is chosen as the parent vector X'_p for the next generation. If W_c for the child vector X_c is larger than W_p for the parent vector X_p, then the child vector X_c is not chosen as the parent vector X'_p for the next generation.

This research used mean squared error (MSE) and mean absolute percentage error (MAPE) to measure the difference between the forecasted value and the actual value. MSE is one of the many ways to quantify the degree to which an estimator differs from the actual value of the quantity being estimated. MSE measures the average of the square of the error. MAPE expresses the error as a percentage. MAPE is commonly used in quantitative forecasting methods because it produces a measure of relative overall fit. The absolute values of all of the percentage errors are summed, and the average is computed. The functions for MSE and MAPE are shown in equations (8) and (10), respectively.

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{g_{(t)} - f_{(t)}}{g_{(t)}} \right|}{n} \times 100 \tag{10}$$

where $g_{(t)}$ is an actual value.
$f_{(t)}$ is a forecasted value
n is the number of data points

This aim of this research is to find the impact of the previous day's SET Index ($SET_{(t-1)}$) on the current SET Index ($SET_{(t)}$). Towards this end, the random coefficient ($a_1 - a_5$) from equation (5) and the fix values of a_0 from 0.01 to 1.06 (34 values in total) were assigned, and the mutation process was performed 34,000 times (1,000 for each a_0 value). The values of a_0 were derived from an empirical analysis done by assigning the weight of the previous day's SET Index from 0.01 to 1.06 in increments of 0.05. The goal of this step was to roughly determine which weight range yields the lowest MAPE value. Based on the empirical results, the value of a_0 from 0.91–1.01 generated the lowest error. Therefore, a detailed analysis was performed using a_0 from 0.90–1.04 in increments of 0.01 to define the value of a_0 that provides the lowest MAPE. The implementation of the pseudo code is shown in Figure 3.

1. Randomly assign standard deviation.
2. Assign coefficient a_0 based on equation (5).
3. Create parent.
   3.1 Randomly assign coefficients ($a_1 - a_5$) based on equation (5) with normal distribution.
   3.2 Evaluate the fitness using MSE.
4. Create new offspring.
   4.1 Mutate and sum up $a_1 - a_5$ equation (6).
   $a_c = a_p + N(0, \sigma^2)$
   where a is 1–5
   4.2 Adjust standard deviation value by applying 1/5 success rule equation (7).
   4.3 Evaluate fitness using MSE equation (8).
5. Select between parent and offspring for the next generation.
6. Repeat step 4 through 5 for 1,000 generations.
7. Evaluate error (MSE and MAPE).

Figure 3. Pseudo Code.

EXPERIMENTAL RESULTS

The experimental data were collected from a reliable source, Bank of Thailand, and consisted of historical data for the SET Index, the Dow Jones Index, the Nikkei Index the Hang Seng Index, and the minimum loan rate. Because the raw data were obtained from different stock markets in different countries, some data are missing because each country has different stock market holidays or non-trading days. However, those gaps can be filled using the data from the previous day with no statistically significant difference. Thus, the assumption underlying this study was that the missing data on non-trading days would be substituted with the previous day's data.

The training data in this experiment were divided into two different time ranges: one-year and two-year periods of training data. There are six different periods of one-year training data: 2003, 2004, 2005, 2006, 2007 and 2008. There are five different periods of two-year training data: 2003–2004, 2004–2005, 2005–2006, 2006–2007 and 2007–2008.

To ensure consistency, the same data periods as in the previous research were used in this experiment. The test data in this experiment were divided into three different time periods: January 2004–December 2004, which includes 244 days in total, was also the period used in Chaigusin et al. (2008b); 9 August 2005–December 2005, which includes 100 days in total, was also used in Tunsenee

(2006); and last, January 2005–March 2009, which includes 1040 days in total, was used for long-term prediction for the proposed function.

The results were compared with those of the existing methods: neural networks (Chaigusin et al., 2008b; Tunsenee, 2006), ARIMA with the EGARCH-M model (Tunsenee, 2006), the simple moving average and random walk using the previous day's SET Index.

The simple moving average is a simple technique in time series forecasting. The weights for the simple moving average used in this research were $0.5_{(t-1)}$, $0.3_{(t-2)}$ and $0.2_{(t-3)}$, as shown in equation (11).

$$SET_{(t)} = 0.5SET_{(t-1)} + 0.3SET_{(t-2)} + 0.2SET_{(t-3)} \tag{11}$$

Random walk was also analysed using the previous day's SET Index ($SET_{(t-1)}$), as shown in equation (12).

$$SET_{(t)} = SET_{(t-1)} \tag{12}$$

Test Data for January 2004–December 2004 (244 Days)

The investigated time series ran from January 2004 to December 2004. It contained 244 days of test data. The data period used in this experiment was the same as in Chaigusin et al. (2008b). The results are shown in Tables 2 and 3 and are used to plot the graph of the function with the lowest MAPE in Figure 4.

Table 2
Experimental Results: January 2004– December 2004 (244 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	47.4644	1.0812	97.3103	1.0964	0.95	0.008659	–1.097766	0.438307	–0.275038	–0.485220
2004	94.8884	1.0981	94.8884	1.0981	0.93	0.161502	0.707495	–0.290776	–0.082473	1.909738
2005	54.1955	0.8449	97.1995	1.1031	0.95	0.005038	1.011462	–0.359281	0.212993	0.256216
2006	118.2925	0.8836	100.8476	1.1130	0.96	0.067321	–0.394087	0.398202	0.664215	4.216546
2007	96.8924	0.9409	98.0089	1.1068	0.93	0.029745	–1.430078	0.343379	–0.896545	–2.640460
2008	144.1077	1.4411	96.9318	1.0971	0.94	–0.049796	–1.986351	1.026574	–0.486922	3.532854
Average	92.6401	1.0483	97.5311	1.1024						
MIN	47.4644	0.8449	94.8884	1.0964						
MAX	144.1077	1.4411	100.8476	1.1130						

Table 3
Experimental Results: January 2004–December 2004 (244 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	78.7869	1.1993	96.4140	1.0916	0.93	–0.112520	–0.623234	0.218552	–0.052978	2.110229
2004–2005	154.0170	1.3657	95.9861	1.1138	0.93	0.197107	0.537165	–0.247805	–0.058315	1.547731
2005–2006	77.9871	0.7942	100.5282	1.1101	0.96	0.033116	–0.089622	0.099849	0.218664	0.702539
2006–2007	129.3140	1.0657	99.0499	1.1056	0.95	–0.001677	1.090188	–0.059179	2.285656	–0.374094
2007–2008	131.5878	1.2892	97.5250	1.1014	0.94	–0.048714	–1.006583	0.530078	–0.406482	2.184182
Average	114.3386	1.1428	97.9006	1.1045						
MIN	77.9871	0.7942	95.9861	1.0916						
MAX	154.0170	1.3657	100.5282	1.1138						

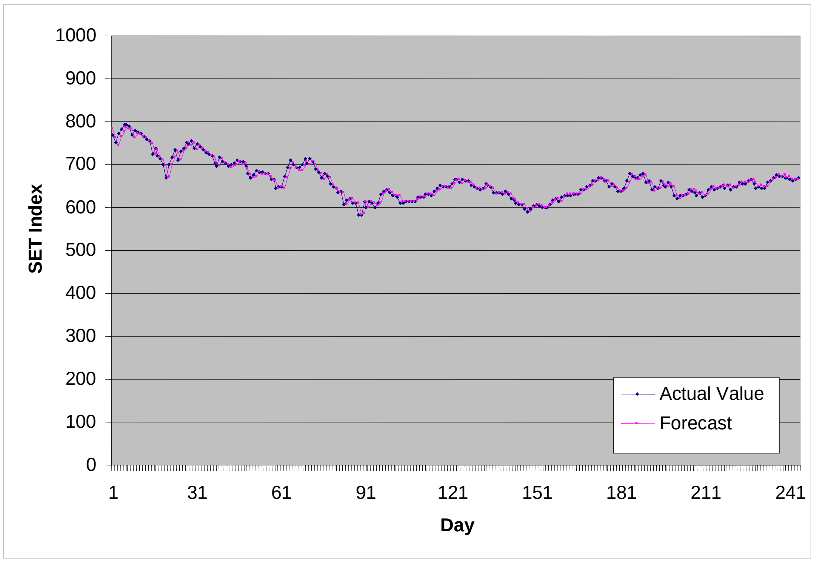

Figure 4. SET Index Comparison Graph for January 2004–December 2004.

Test Data for 9 August 2005–December 2005 (100 Days)

The investigated time series ran from 9 August 2005 to December 2005. It contained 100 days of test data. The data period used in this experiment was the same as in Tunsenee (2006). The results are shown in Tables 4 and 5, and the graph of the function with the lowest MAPE is plotted in Figure 5.

Table 4
Experimental Results: 9 August 2005–December 2005 (100 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	45.3274	1.0342	27.2102	0.5738	0.96	0.011127	0.222573	0.368074	0.202688	0.671492
2004	100.3172	1.1164	27.2325	0.5713	0.96	0.009391	–0.039491	0.862510	0.677917	1.179922
2005	35.9150	0.6760	27.2004	0.5733	0.96	0.006709	0.121044	0.948570	0.694217	0.958766
2006	118.0647	0.8790	27.3677	0.5792	0.97	0.029055	–0.309739	0.271880	0.297817	–1.087940
2007	89.4518	0.8802	27.2286	0.5730	0.96	0.002387	1.168335	1.012049	0.317940	0.429231
2008	144.4205	1.4426	27.2101	0.5726	0.96	0.006128	0.441716	0.970879	0.493224	0.902369
Average	88.9161	1.0047	27.2416	0.5739						
MIN	35.9150	0.6760	27.2004	0.5713						
MAX	144.4205	1.4426	27.3677	0.5792						

Table 5

Experimental Results: 9 August 2005–December 2005 (100 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	73.9844	1.0781	27.2160	0.5750	0.96	0.020723	0.129153	0.691633	0.443156	2.089717
2004–2005	68.5116	0.9008	27.2011	0.5735	0.96	0.009456	–0.039628	0.862833	0.678087	1.180203
2005–2006	77.1363	0.7847	27.2994	0.5776	0.97	–0.083211	–0.412528	–0.398191	–0.063527	6.829231
2006–2007	106.2662	0.9102	27.2050	0.5729	0.96	0.003045	0.384704	0.999451	0.558801	0.461476
2007–2008	117.6047	1.1674	27.2224	0.5745	0.96	0.017930	0.924069	1.298628	0.457218	3.565818
Average	88.7007	0.9682	27.2288	0.5747						
MIN	68.5116	0.7847	27.2011	0.5729						
MAX	117.6047	1.1674	27.2994	0.5776						

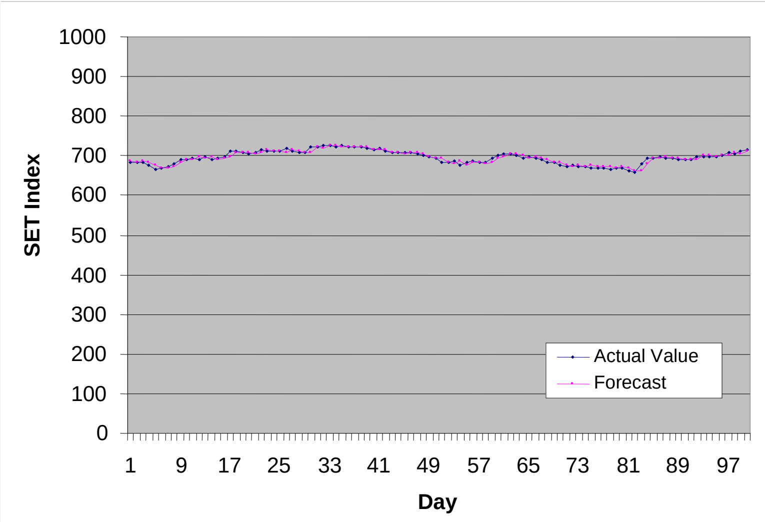

Figure 5. SET Index Comparison Graph for 9 August 2005–December 2005.

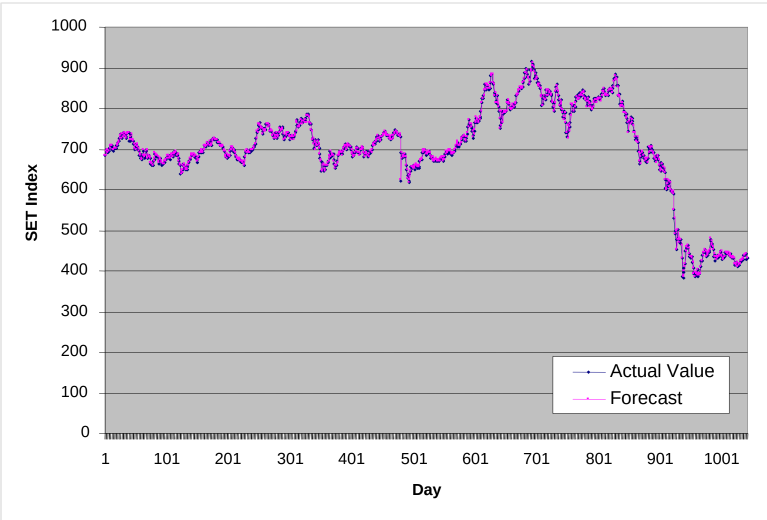

Figure 6. SET index comparison graph for January 2005–March 2009.

Table 6
Experimental Results for January 2005–March 2005 (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 7
Experimental Results for January 2005–March 2005 (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						

Test Data for January 2005–March 2009 (1040 Days)

The investigated time series ran from January 2005 to March 2009. It contained 1040 days of test data. The results are shown in Tables 6 and 7, and the graph of the function with the lowest MAPE is in Figure 6.

Results comparison

As shown in Table 8, the results of the test data for the period of January 2004–December 2004 were obtained by calculating the MAPE of the proposed function (5), the simple moving average (11), and the previous day's SET Index (12), while the results of the three layers, four layers and five layers were taken from Chaigusin et al., (2008b). The results of the comparison show that the proposed function is the best prediction function for the SET Index in this time period with the lowest level of MAPE.

Table 8
MAPE Comparison for the Period January 2004–December 2004

Method	MAPE (%)
Proposed function (1 Year)	1.0964
Proposed function (2 Years)	**1.0916**
Simple Moving Average (11)	1.2795
Previous Day’s SET Index (12)	1.1505
Chaigusin et al. 2008b (Three layers)	1.2659
Chaigusin et al. 2008b (Four layers)	1.1472
Chaigusin et al. 2008b (Five layers)	1.1458

As shown in Table 9, the results of the test data for the period 9 August 2005–December 2005 were measured by calculating the MAPE of function (5), the simple moving average (11), and the previous day's SET Index (12), whereas the results of Tunsenee (2006)'s neural network and ARIMA with EGARCH-M model were taken from his research. The comparison results show that the proposed function is the best prediction function for forecasting the SET Index in this time period because it has the lowest MAPE.

Table 9
MAPE Comparison for the Period of 9 August 2005–December 2005

Method	MAPE (%)
Proposed Function (1Year)	**0.5713**
Proposed Function (2Years)	0.5729
Simple Moving Average (11)	0.7053
Previous Day's SET Index (12)	0.5795
Tunsenee (Neural Network)	1.2956
Tunsenee (ARIMA with EGARCH-M Model)	0.5972

The test data for the period of January 2005–March 2009 were measured by calculating the MAPE of the proposed function [equation (5)](5), the simple moving average (11) and the previous day's SET Index [equation (12)](12) as shown in Table 10. The results show that the proposed function is the best prediction function for the SET index in this time period with the lowest MAPE as well.

Table 10
MAPE Comparison for the Period of January 2005–March 2009

Method	MAPE (%)
Proposed Function (1Year)	**0.9820**
Proposed Function (2Years)	0.9826
Simple Moving Average (11)	1.1752
Previous Day's SET Index (12)	0.9916

Results Analysis

According to Tables 8 and 10, the first period, January 2004–December 2004 (244 days), exhibited a greater degree of error based on all methods than the third period, January 2005–March 2009 (1040 days). The amount of error derived using the previous day's SET Index in the first period (1.1505) is greater than in the third period (0.9916) even though the third period is much longer. This finding indicates that in the year 2004, the SET index relied less on the previous day's SET Index in the first period than in the third period. Therefore, the accuracy of the proposed prediction function depends on the weight of the previous day's SET Index (a_0). If the weight of the previous day's SET Index (a_0) increases, then the proposed method will yield good accuracy. However, regardless, the results achieved using the proposed method are still more accurate than those derived using the previous day's SET Index.

CONCLUSIONS

This research has proposed a prediction function for the SET Index. The proposed prediction function used previous values for important factors include the Dow Jones, Nikkei, and Hang Seng indexes and the MLR to forecast the current SET Index. The proposed prediction function was tested for different three time periods: January 2004–December 2004, 9 August 2005–December 2005 and January 2005–March 2009. The results show that the proposed prediction function yields the lowest errors.

The results achieved using different years of training data from 2003 to 2008 show that there is no significant difference in performance, although many crises took place in Thailand during that time, including the *coup d'état* in 2006, anti-government protests and the sub-prime mortgage crisis that occurred in 2008. This finding suggests that the data used for training can be obtained from any time period between 2003 and 2008 with no significant effect. In addition, the results obtained using training data from time periods of different lengths (one year and two years) are also not significantly different from one another. Therefore, it would seem that the training data time period can be either one year or two years long. However, using a two-year period requires more computation time. Thus, it is best to use data from one-year periods for the prediction function for the SET Index.

The results for the January 2005–March 2009 period indicate that the previous day's SET Index$_{(t-1)}$ has greatest effect on the current SET Index$_{(t)}$; the other factors are the Dow Jones Index$_{(t-1)}$, Nikkei Index$_{(t-1)}$, Hang Seng Index$_{(t-1)}$ and the domestic MLR$_{(t-1)}$. Additionally, the Dow Jones Index$_{(t-1)}$, Nikkei Index$_{(t-1)}$, and Hang Seng Index$_{(t-1)}$ has a direct variation relationship to the current SET Index$_{(t)}$, whereas the domestic MLR$_{(t-1)}$ has an inverse variation relationship to the current SET Index$_{(t)}$. Therefore, this study supports the assumption that the previous movements of the SET Index were sensitive to the Dow Jones Index, Nikkei Index, Hang Seng Index and domestic MLR.

This paper predicts the closing SET Index by incorporating historical and public information. The performance of the proposed prediction function is slightly better than that of the method using the previous day's SET Index (based on random walk theory). Intraday individual stock price prediction will be the challenge for future work to undertake because it will tell us whether or not the proposed prediction function can be used to identify undervalued stocks. If it can, then this will not support the semi-strong EMH.

REFERENCES

Back, T., Hoffmeister, F., & Schwefel, H. P. (1991). A survey of evolution strategies. *Proceeding of the Fourth Conference on Genetic Algorithm*, CA, 2–9.

Bank of Thailand. (2009). Retrieved October 1, 2009, from http://www.bot.or.th.

Bernstein, P. L. (1999). A new look at the efficient market hypothesis. *The Journal of Portfolio Management*, *25*(2), 1–2.

Beyer, H. G., & Schwefel, H. P. (2002) Evolution strategies-a comprehensive introduction. *Natural Computing*, *1*(1), 3–52.

Box, G., & Jenkins, G. (1976). *Time series analysis: Forecasting and control*. CA: Holden Day.

Chaereonkithuttakorn, K. (2005). The relationship between the Stock Exchange of Thailand Index and the Stock Indexes in the United States of America. Unpublished master's thesis, Chiang Mai University, Thailand.

Chaigusin, S., Chirathamjaree, C., & Clayden, J. (2008a). Soft computing in the forecasting of the stock exchange of Thailand (SET). *Management of Innovation and Technology, 2008. ICMIT 2008*, September 21–24, Bangkok Thailand.

———. (2008b). The use of neural networks in the prediction of the stock exchange of Thailand (SET) Index. *Computational Intelligence for Modelling Control & Automation*, December 10–12, Vienna, 670–673.

Chotasiri, S. (2004). The economic factors affecting the fluctuation of the stock exchange of Thailand index. Unpublished master's thesis, Chiang Mai University, Thailand.

———. (1965) The behavior of stock-market prices. *Journal of Business*, *38*(1).

Fama, E. F. (1991). Efficient capital markets. *Journal of Finance*, *XLVI*(5), 1575–1617.

Holland, J. H. (1992). *Adaptation in natural and artificial systems*. MA: MIT Press Cambridge.

Horii, M., Takahashi, N., & Narita, T. (2000). Investigation of evolution strategy and optimization on induction heating model. *IEEE Transactions on Magnetics*, *36*(4), 1085–1088.

Kendall, M. G., & Hill, A. B. (1953). The analysis of economic time-series-part I: prices. *Journal of the Royal Statistical Society*, *116*(1), 11–34.

Khumyoo, C. (2000). The determinants of securities price in the stock exchange of Thailand. Unpublished master's thesis, Ramkhamhaeng University, Thailand.

Malkiel, B. G. (1999). A random walk down wall street: Including a life-cycle guide to personal investing. New York: W.W.Norton & Company.

———. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives 17*(1), *59–82.*

Meesad, P., & Srikhacha, T. (2008). 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* (SNPD2008), August 6–8, Phuket Thailand, 422–427.

———. (2006). 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*, November 28–December 1, NSW, Australia.

Osler, C. (2000). Support for resistance: Technical analysis and intraday exchange rates. *Economics Policy Review, July*, 53–68.

Pattamavorakun, S., & Pattamavorakun, S. (2007). 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,* August 20–22, Busan South Korea, 184–191.

Rimcharoen, S., Sutivong, D. & Chongstitvatana, P. (2005). Prediction of the stock exchange of Thailand using adaptive evolution strategies. *Tools with Artificial Intelligence, 2005. ICTAI 05, 17th*. November 16, Hong Kong.

Rimcharoen, S., & Chongstitvatana, P. (2004). An adaptation of evolutionary strategies for forecasting the exchange rate. *The 8th Annual National Symposium on Computational Science and Engineering* (ANSCSE8), July 21–23, Nakhon Ratchasima, Thailand.

Rechenberg, I. (1971). Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Dr.-Ing. Thesis, Technical University of Berlin, Germany.

———. (1973) *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Stuttgart: Frommann-Holzboog Verlag.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFinText system. *Transactions on Information System*, *27*(2), 1–19.

Schwefel, H. P. (1975) Evolutionsstrategie und numerische Optimierung. Dr.-Ing. Thesis, Technical University of Berlin, Germany.

———. (1977) Numerische optimierung von computer-modellen mittels der Evolutionsstrategie. *Interdisciplinary Systems Research, 26.* Birkhäuser, Basel, Switzerland.

Sutheebanjard, P., & Premchaiswadi W. (2009). Factors analysis on stock exchange of Thailand (SET) index movement. *The 7th International Conference on ICT and Knowledge Engineering, ICTKE2009*, Bangkok, Thailand, December 1–2, 2009.

Tantinakom, T. (1996). Economic factors affecting stock exchange of Thailand Index. Unpublished doctoral dissertation, Chiang Mai University, Thailand.

The Stock Exchange of Thailand. Retrieved from http://www.set.or.th. October 1, 2009.

Thomsett, M. C. (1998). *Mastering fundamental analysis: Kaplan AEC Education*. Chicago: Dearborn Financial Publishing, Inc.

Tunsenee, J. (2006) Accuracy comparison in securities price forecasting between neural networks model and EGARCH-M ARIMA model. Unpublished doctoral dissertation, Chiang Mai University, Thailand.

Worasucheep, C. (2007). A New self adaptive differential evolution: Its application in forecasting the Index of stock exchange of Thailand. *Evolutionary Computation, 2007. CEC 2007.*