การหาจังหวะการลงทุนและการตามดัชนีในประเทศไทยโดยโปรแกรมเชิงพันธุกรรม

นายสมเจตน์ เตชะอินทราวงศ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาการเงิน ภาควิชาการธนาคารและการเงิน

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-17-4462-5

MARKET TIMING AND INDEX TRACKING BY GENETIC PROGRAMMING IN THAILAND

Mr. Somjade Techa-intrawong

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Finance

Department of Banking and Finance

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2005

ISBN 974-17-4462-5

Thesis Title	MARKET TIMING AND INDEX TRACKING BY GENETIC PROGRAMMING IN THAILAND
By	Mr. Somjade Techa-intrawong
Field of study	Finance
Thesis Advisor	Associate Professor Sunti Tirapat, Ph.D.

Accepted by the Faculty of Commerce and Accountancy, Chulalongkorn University in Partial Fulfillment of Requirements for the Master's Degree

Danuja Kunpanitchakit .... Dean of the Faculty of Commerce and Accountancy

(Assistant Professor Danuja Kunpanitchakit, Ph.D.)

THESIS COMMITTEE

.... Chairman

(Assistant Professor Patcharavalai Jayapani, DBA)

Sunti Tirapat .... Thesis Advisor

(Associate Professor Sunti Tirapat, Ph.D.)

Anant Chiarawongse .... Member

(Anant Chiarawongse, Ph.D.)

สมเจตน์ เตชะอินทราวงศ์ : การหาจังหวะการลงทุนและการตามดัชนีในประเทศไทยโดยโปรแกรมเชิงพันธุกรรม (MARKET TIMING AND INDEX TRACKING BY GENETIC PROGRAMMING IN THAILAND) อ. ที่ปรึกษา: รองศาสตราจารย์ ดร.สันติ ถิรพัฒน์, 46 หน้า. ISBN 974-17-4462-5.

วิทยานิพนธ์ฉบับนี้นำเสนอการประยุกต์ใช้โปรแกรมเชิงพันธุกรรม (Genetic Programming) ในการแก้ปัญหาการลงทุนอันได้แก่ การหาจังหวะการลงทุน (Market Timing) และการตามดัชนี (Index Tracking) โปรแกรมเชิงพันธุกรรมเป็นแนวคิดการแก้ปัญหาทางคอมพิวเตอร์เพื่อหาคำตอบที่เหมาะสมที่สุด โดยคำตอบที่ได้จากโปรแกรมเชิงพันธุกรรมจะเป็นคำตอบที่ดีที่สุดเฉพาะช่วงเวลา แต่คำตอบที่ได้จากโปรแกรมก็มีความน่าเชื่อถือว่าเป็นคำตอบที่ใกล้เคียงกับคำตอบที่แท้จริง

ในการศึกษานี้ โปรแกรมเชิงพันธุกรรมจะทำการหากลยุทธ์การซื้อขายที่สามารถสร้างกำไรได้เพื่อแก้ปัญหาการหาจังหวะการลงทุน กลยุทธ์การซื้อขายจะสร้างสัญญาณซื้อและสัญญาณขายจากข้อมูลย้อนหลังของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย ถ้าโปรแกรมเชิงพันธุกรรมสามารถหากลยุทธ์การซื้อขายที่สร้างกำไรส่วนเกินกว่าดัชนีตลาดหลักทรัพย์แห่งประเทศไทยได้ แปลว่านักลงทุนสามารถสร้างกำไรส่วนเกินได้ถ้าลงทุนตามสัญญาณซื้อขายที่ถูกสร้างขึ้น อย่างไรก็ตามในความเป็นจริงแล้วนักลงทุนไม่สามารถลงทุนในดัชนีได้โดยตรง ดังนั้นนักลงทุนจำเป็นต้องสร้างกลุ่มหลักทรัพย์ที่มีผลตอบแทนเสมือนดัชนี ดังนั้นในวิทยานิพนธ์นี้จึงใช้โปรแกรมเชิงพันธุกรรมในการหาหลักทรัพย์และสัดส่วนการลงทุนของกลุ่มหลักทรัพย์ที่มีผลตอบแทนเสมือนดัชนีด้วย โปรแกรมจะหาหลักทรัพย์จากหลักทรัพย์ที่เคยเป็นสมาชิกของดัชนี SET50 ในการศึกษานี้ได้จำกัดจำนวนหลักทรัพย์ในการสร้างกลุ่มหลักทรัพย์ได้แก่ 5, 10 และ 15 บริษัท และคำตอบที่ดีที่สุดของการหาจังหวะการลงทุนและการตามดัชนีจะถูกนำมาทดสอบด้วยกันว่าสามารถนำไปใช้จริงได้หรือไม่ วิทยานิพนธ์นี้ทำการศึกษากับดัชนีตลาดหลักทรัพย์แห่งประเทศไทย ในช่วงปีพ.ศ. 2533 ถึง 2546

ผลการศึกษาพบว่า โปรแกรมเชิงพันธุกรรมสามารถสร้างกลยุทธ์การซื้อขายที่สร้างกำไรส่วนเกินได้ หาหลักทรัพย์และสัดส่วนการลงทุนที่สร้างผลตอบแทนเสมือนดัชนีได้อย่างมีนัยสำคัญในช่วงการเรียนรู้ของโปรแกรม สำหรับในช่วงการทดสอบนอกกลุ่มตัวอย่างเมื่อรวมคำตอบของทั้งการหาจังหวะการลงทุนและการตามดัชนีเข้าด้วยกันแล้ว พบหลักฐานว่าผลตอบแทนส่วนเกินโดยส่วนใหญ่เป็นบวกเมื่อซื้อขายตามกลยุทธ์แต่ไม่มีนัยสำคัญ จึงไม่สามารถสรุปได้ว่าสภาวะตลาดในประเทศไทยมีประสิทธิภาพในระดับต่ำในช่วงเวลาที่ศึกษา

ภาควิชา การธนาคารและการเงิน	ลายมือชื่อนิสิต สมเจตน์ เตชะอินทราวงศ์
สาขาวิชา การเงิน	ลายมือชื่ออาจารย์ที่ปรึกษา สันติ ถิรพัฒน์
ปีการศึกษา 2548	

4682450926: MAJOR FINANCE
KEY WORD: GENETIC PROGRAMMING / MARKET TIMING / INDEX TRACKING

SOMJADE TECHA-INTRAWONG: MARKET TIMING AND INDEX TRACKING BY GENETIC PROGRAMMING IN THAILAND. THESIS ADVISOR: ASSOC. PROF. SUNTI TIRAPAT, Ph.D., 46 pp. ISBN 974-17 -4462-5.

This thesis applies genetic programming to solve the investment problem of market timing and index tracking. Genetic programming is a computer algorithm used to find the optimal solution. The solution found by genetic programming is a local solution but it is recognized that the solution is close to the exact solution.

In this study, genetic programming technique is used to find a profitable trading strategy so as to solve the problem of market timing. The trading strategy will generate both buy and sell signals based on the historical data of Stock Exchange of Thailand (SET) index. If genetic programming is able to find a profitable trading strategy, it means that investors could earn excess return over that of SET index when they apply the strategy with the index. However, in fact, investors cannot invest directly in SET index. So, they must form their index-tracking portfolio. Thus, this thesis will also apply genetic programming to solve for the combinations of index-tracking portfolio that generates daily returns as same as the index. This tracking portfolio is established from the securities having been the members of SET50 index. In this study, the tracking portfolio is consisted of 5, 10 and 15 companies. The combination of solutions for both market timing problem and index tracking problem is applied to the out of sample test in order to verify the possibility of the trading strategy becoming lucrative in the real practice. This thesis applies genetic programming technique to SET index data from year 1995 to 2003.

The final results suggest that genetic programming dramatically help in finding both the profitable trading strategy and the tracking portfolio in training period. However, for out of sample period, the combination of the solutions found by genetic programming generates excess return over SET index with no strong evidence. Almost excess returns are positive but each *t*-statistic is not strongly significant. Hence, the study cannot conclude that the SET is the weak form of the efficient market hypothesis (EMH) in the study period.

Department of Banking and Finance
Field of study: Finance
Academic year 2005

Student's signature Somjade Techa-intrawong
Advisor's signature Sunti Tirapat

Acknowledgements

My appreciation is expresses here to those who have contributed to completion of this thesis.

First of all, I desire to express my gratitude to Assoc. Prof. Sunti Tirapat, my thesis advisor, for his valuable guidance and suggestion throughout the period of research. I am also grateful to Assist Prof Patcharavalai Jayapani and Dr. Anant Chiarawongse my thesis committee for their generous advises and comments. I wish to express my thanks to Mr. Jaturan Raintipayasakun and Miss Chalita Promchan for their help. Also my friends in MSF course for helping anytime upon my request.

Lastly, I would like to express my deep gratitude to my beloved family for their support, understanding and encouragement. Any mistakes in this study are my sole responsibility

Table of Contents

Page

List of Tables

List of Figures

CHAPTER I

Introduction

1.1 Backgrounds and Problem Review

Genetic programming is used to solve an optimal problem. It can solve both linear and nonlinear problems. Besides, genetic programming can find the optimal solution of the problem that is difficult to explain in term of equations. The basic examples of genetic programming are traveling salesman problems, which need the solution of the shortest route to travel all given cities.

Genetic programming is a method developed from adaptive system theory. Genetic programming generates a suitable solution from given component. Genetic programming can find an appropriate for all problems, but its solution is not the global solution. Although a solution found by genetic programming is only local solution, the solution is also acceptable for each problem. In general, the best solution of each genetic programming running is not the same one. However, the solution found by genetic programming has a tendency to usefully answer its problem. Hence, genetic programming users should clearly understand their problems beforehand.

Furthermore, genetic programming has been employed to solve a lot of questions in many areas and absolutely the finance is one of them. There are many foreign studies about the application of using genetic programming to solve the investment timing of both money and capital markets such as Bauer (1994) but there is not the study of this application in Thailand. So, this paper presents the application of genetic programming to solve the main problems of when investors should go into the market, called market timing, and what security and its proportion of investment portfolio.

The major problem of this work is market timing in the Stock Exchange of Thailand. The market timing is an investment strategy that investors can make an abnormal return on buy-and-hold strategy by forecasting the trend of the market. They hope to make a profit with the strategy by sell high and buy low. In this study, SET index is the instrument that the trading strategies apply to but investor cannot directly invest in the index. So, this paper presents other application of genetic programming to find a combination of an investment portfolio that generates daily returns as same as daily index returns, called index tracking. Thus, the trading strategy can exist when both of two solutions are parallel used.

By using genetic programming to solve two problems of market timing and index tracking, the investor is going to know both timing and the proportion of an investment portfolio. Moreover, if the combination of solutions of market timing and index tracking can really earn excess return over SET index, SET is concluded that it is under weak form of EMH. In the other word, the Stock Exchange of Thailand is inefficient.

The weak form tests are concerned with the validity of using the past history of prices to predict future prices. Tests of the weak form of efficiency addressed two questions: (a) do prices over time have sufficient serial dependence to allow investors to predict future price movements by studying trend? And (b) can trading strategies based on price movements provide opportunities for abnormal profit? Most of the previous works claim that since the existence of historical price movement, technical analyst can provide insights into future prices. Moreover, the finding suggests the trading strategy make the abnormal return if ignored transaction cost.

Over the years, there are many literatures on the investment strategy with genetic programming. The majority of these literatures have found that the genetic programming as a nonlinear model cannot make excess return on buy-and-hold strategy. Chen and Yeh (1997) and Allen and Karjalainen (1999) find no evidence that the genetic programming can find profitable trading rules on stock market. Neely (2001) also argue that the genetic programming cannot find the excess return on stock market with a risk adjusted return as a fitness measurement.

However, some literatures support that genetic programming can generate a trading rule giving excess return. Neely, Weller and Dittmar (1997) find a strong evidence of profitable trading rules on foreign exchange market. Potvin, Soriano and Vallee (2004) find the condition of profitable and non-profitable trading rules generated by genetic programming on Canadian market. They use not only prices of stocks as other papers but also develop their trading volume in the program. They find that the rule is profitable when the market falls or when it is stable and the rule is non-profitable when market is rising. This result is consistent with Blume, Easley and O'Hara (1994) and Conrad, Hameed and Niden (1994) the results of which indicate investors make better decisions when they base on both prices and volume.

According to genetic programming abilities to solve the optimal problem, it has been applied in other branch of investments. Index tracking, which is famous for the passive portfolio management, is one of them that genetic programming is used to find an optimal tracking portfolio. The program is going to search the combination of portfolios both firm names and weights of each firm. Users can control restricts of programs by given a number of tracking firm, maximum and minimum proportion of each stock. Furthermore, Beasley, Meade and Chang (2003) recommend that investors can apply the costs of portfolio rebalance in the program to find an optimal solution.

Genetic programming is not only a new idea for investments but also a profitable one. It is high probability that Genetic programming has already been applied in Wall Street. Moreover, it is possible that foreign investors and funds have used this idea in Stock Exchange of Thailand (SET) as well. But there are not evidences to confirm that genetic programming is really profitable in Thailand.

In this study, genetic programming finds the trading strategy based on historical data as the technical tool. Although there are many papers reporting that technical analysis cannot earn the abnormal return in an equity market, there are studies indicating that the technical tools can find excess returns in the Asian stock markets such as Bessembinder and Chan (1995) and Gunasekarage and Power (2001).

Hence, genetic programming should be verified whether it has a profitable ability in the SET or not. This paper applies more technical tools as functions of genetic programming in order to generate more profitable and reasonable trading rules. Besides, genetic programming is also used to find a tracking portfolio in order to make the trading strategy coming true in the real world.

1.2 Objectives of this Study

1. Find a technical trading rule having abnormal return on buy-and-hold strategy in the Stock Exchange of Thailand.
2. Study the index tracking in the Stock Exchange of Thailand.
3. Present genetic programming that can be applied with the investment in the Stock Exchange of Thailand.
4. Test the Efficient Market Hypothesis (EMH) in the Stock Exchange of Thailand.

1.3 Research Hypothesis

Main purposes of this paper are using genetic programming to find excess returns from combinations of trading rules and index-tracking portfolios. So, there are three major hypothesis tests. First hypothesis tests whether the trading rule found by genetic programming can earn abnormal returns over buy-and-hold strategy in training period or not. In this study, the excess return is denoted by $\Delta\pi$. Hence, this hypothesis is:

$$H_0 : \Delta\pi = 0$$
$$H_1 : \Delta\pi > 0$$

If the null hypothesis is rejected, it means that genetic programming can establish the trading rule earning abnormal returns in the Stock Exchange of Thailand.

Second hypothesis tests whether the index tracking portfolio found by genetic programming can generate daily returns as same as SET index in training period or not. The error between the tracking portfolio and the index is the parameter indicating how good of the tracking portfolio denoted by ε. Hence, this hypothesis is:

$$H_0 : \varepsilon = 0$$
$$H_1 : \varepsilon \neq 0$$

If null hypothesis is accepted, it means that genetic programming can establish the tracking portfolio that has errors of daily returns between the tracking portfolio and SET index equal to zero. On the other word, the daily return from the tracking portfolio is as same as the return of SET index.

Third Hypothesis tests whether the combination of the trading rule and the tracking portfolio is able to generate excess return over buy-and-hold strategy in the out of sample period or not. The excess return of the combination is denoted by $\Delta\pi$. Hence, this hypothesis is:

$$H_0 : \Delta\pi = 0$$
$$H_1 : \Delta\pi > 0$$

If the null hypothesis is rejected, it means that genetic programming can establish the combination of the trading rule and the tracking portfolio earning excess returns in the Stock Exchange of Thailand. Moreover the last experiment can be interpreted the form of EMH. If the null hypothesis is rejected, it means that the Stock Exchange of Thailand is inefficient market.

1.4 Scope of the Study

This study uses secondary data of the Stock Exchange of Thailand both historical level of SET index and prices of securities from DataStream as inputs of genetic programming in order to generate a profitable trading rule and a tracking portfolio. Besides, this work also requires commercial bank saving rates from Bank of Thailand (BOT) as returns for an out market period. This paper focuses the data between 1995 and 2003. Moreover, this paper has already included the one-way trading commission of 0.25 percent.

1.5 Limitation

This paper bases on assumption “short sell” is not allowed that is consistent with a fact in Thailand. Although some brokers have a short-selling service, this work neglects this transaction.

1.6 Terms and Definition

Genetic programming (**GP**) is an automated methodology inspired by biological evolution to find computer programs that best perform a user-defined task. It is therefore a particular machine learning technique that uses an evolutionary algorithm to optimize a population of computer programs according to a fitness landscape determined by a program's ability to perform a given computational task.

Market timing is the strategy of attempting to predict future price movements through use of various fundamental and technical analysis tools.

Index tracking involves building an investment portfolio designed to track a particular benchmark index. At its simplest, it requires holding all stocks in the index, and weighting each stock holding so each investment is held in proportion to its contribution to the index being tracked. Index tracking is often referred to as 'passive' investment and can be contrasted with 'active' management, where fund managers seek to outperform a market benchmark.

1.7 Contributions

1. Understand genetic programming application with market timing.

This study presents genetic programming to discover the market timing of the SET. In the other word, genetic programming is going to find a trading rule that indicates investors should either invest in or keep out the market. Moreover, investors can determine other constraints or information in order to get an appropriate trading rule for individual.

2. Understand genetic programming application with index tracking.

The contribution of this paper is using genetic programming find a tracking portfolio earning as index returns. Genetic programming is going to select both security names and their weights and has a given number of securities as the constraint. It is the essential part that makes market timing in Stock Exchange of Thailand turn to be possible. Moreover, investors can find their portfolios that are suitable with their conditions

3. Understand Efficient Market Hypothesis (EMH) of the Stock Exchange of Thailand.

The aim of the paper is finding a portfolio trading strategy. If genetic programming can find a market-beating trading rule, it means that the Stock Exchange of Thailand is inefficient. After know the stage of EMH, investors can invest with the right strategy.

1.8 Organization of paper

The structure of the paper is as follow. Chapter II reviews the theoretical of genetic programming, the steps of running and the early empirical work on both applications of market timing and index tracking. Chapter III explains the methodologies of genetic programming applications with both problems. Chapter IV reports results of trading rules and tracking portfolios found by genetic programming. The results present performances of solutions both in sample and out of sample periods. Final chapter of the paper conclude the summaries and suggestions.

CHAPTER II

Literature Review

Chapter II explains the background, concepts of genetic programming and the empirical studies. This chapter goes deep in detail about genetic programming both its algorithm and processes of the solution evolution. Besides, this chapter reviews the previous papers applying genetic programming with the investment and relative studies in Thailand.

2.1 Concept and Theoretical Background

2.1.1 The Concept of Genetic Programming

Genetic programming is a branch of genetic algorithms. Genetic algorithms are ideas of computer processes for an optimal solution based on the principle of natural selection, which originally expounded in Darwin's theory of evolution. These procedures were firstly developed by Holland (1975), who developed genetic algorithms from adaptive system theory. Genetic algorithms are suitable to find the solution with optimal problems. Their solution is neither the exactly answer nor the global solution, however this answer is not only a local solution but it is also an acceptable solution in a wide range of problems.

Genetic programming helpfully uses to solve the optimal problem with unknown length of solutions. Genetic programming is applied in many sciences and also in the finance. Genetic programming was first introduced by John Koza (1992) to solve genetic algorithms' weakness. In Holland's genetic algorithm, which is the original one, genetic structures are represented as fix length character strings. This structure is adequate to solve many problems but it is restrictive when size or forms of solution cannot be assessed beforehand. Koza's extension, solutions of which were the program syntax trees, permits explicitly hierarchical variable length.

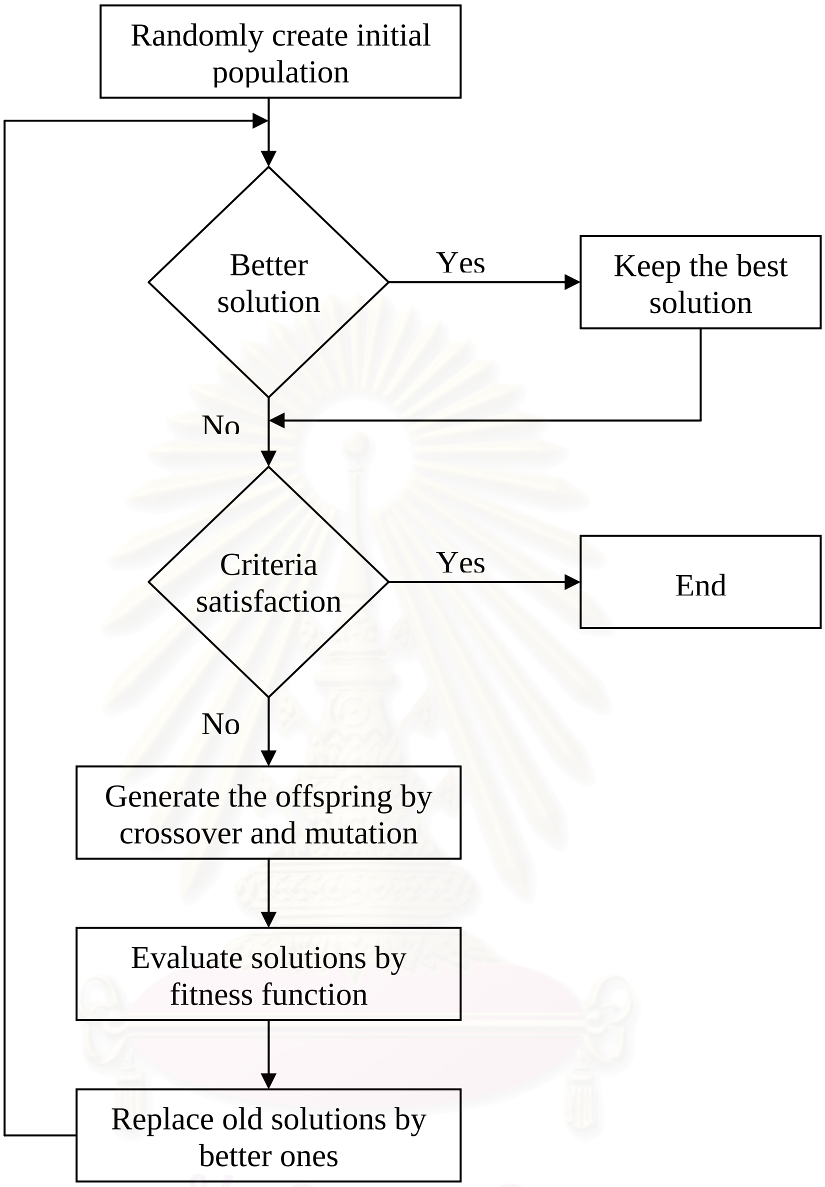


Fig 1 The algorithmic flowchart of genetic programming

The main concepts about evolution of genetic programming are randomly creating and developing solutions to hopefully get a right answer. These algorithms are depicted as a flowchart in figure 1. The first process of genetic programming is randomly creating initial population in the first generation. Besides, genetic programming is about to save the best solution of the generation as the local solution. Next, all solutions are evolved in order to generate their various offspring in new generation. Furthermore, some old worse solutions are replaced by new better offspring. If the new best solution is more appropriate than the old best solution kept in last generation, the new solution is going to replace the old one and becomes new local solution. The processes are going on till the given generation or there is no one better than the local solution. The last best solution is the final solution of each problem.

Genetic programming randomly generates initial population from given function and terminal sets. Function and terminal sets are unique units of solutions that are randomly selected and combined to create initial solutions. After initial population creation, genetic programming transforms solutions into genetic structures that are chromosomes in creatures. Each individual population represents a possible solution and its genetic structures represent the individual characteristic of each solution. Genetic structures are transformed to the tree diagram. For example, a solution of X^2 + 2X + 1 is transformed to a tree diagram as shown in figure 2. This solution has the functions of plus and multiplier and its terminals are variable of x, constants of 1 and 2.

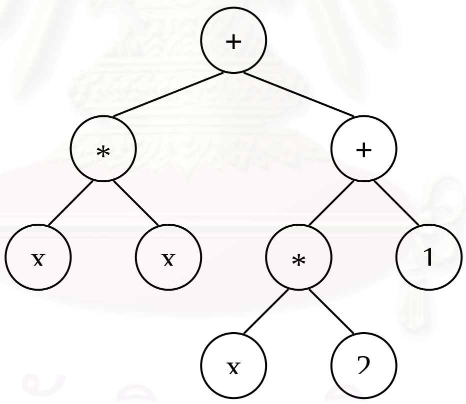


Figure 2   An example of tree diagram of X^2 + 2X + 1

The evolution of genetic programming creates the better offspring in next generation because of crossover, which is a major process making a variety of creatures in the nature. Crossover is the process that recombines genetic structures of parents in order to create various characters of the offspring as shown in figure 3. This method can generate the offspring based on genetic structures of its parents. Crossover is about to randomly select a pair of high fitness parents from population of solutions, randomly separate their genetic structures into two parts—main tree and sub tree—and, finally, cross a sub tree over the other main tree. This may result in offspring that are more fit than the parents.

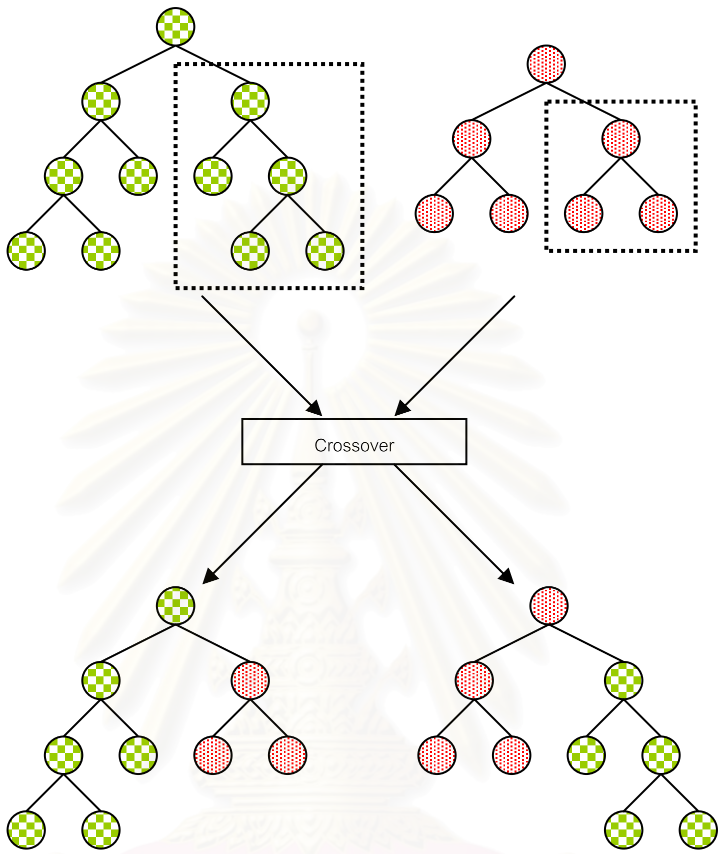


Figure 3 The crossover of genetic programming

After the process of crossover, the offspring of new generations and their parents must be defined how quality of solutions they are. Therefore, we use a fitness measurement in order to indicate the quality of solutions. The several worst solutions are eliminated. Other solutions become new parents for the evolution in the next generation.

Moreover, the natural evolution of creatures is not only use the crossover but also apply the mutation to let the offspring exactly difference from their parents. The mutation is the progress that randomly changes a few genetic structures such as 0 to 1 or vice versa in binary code. Normally, the mutation is very small probability but it is essential for a variety of genetic structures to find the closely optimal solution.

2.1.2 The Steps of Genetic Programming

To clearly understand the operations of genetic programming, this section is about to demonstrate the steps of running in this study. Genetic programming consists of four basic steps as follow:

Step 1: Randomly create the initial solution

Randomly create a solution from given functions and variables.

Run this step 100 times as the initial population.

Calculate the fitness of each rule in the training period.

Step 2: Find the initial best solution

Select the highest fitness solution in the training period

Recalculate fitness of the solution in the selection period.

Save this rule as the initial best rule.

Step 3: Solution evolution

Randomly pick two parent rules, using a probability distribution skewed towards the best rule.

Create a new solution by randomly breaking the parent structures and recombining them.

Compute the fitness of the solution in the training period.

Run this step 100 times as the offspring in each generation.

Replace one of the old rules by the new rules, using a probability distribution skewed towards the worst rule.

Step 4: Find the best solution until matching with the criteria

Select the new highest fitness solution in the training period

Recalculate fitness of the new solution in selection period.

If the new solution improves upon the previous best rule, save as the new best rule.

Stop if there is no improvement for 25 generations or after a total of 50 generations. Otherwise, go back to Step 3.

2.2 Literature Review Empirical Studies

In this section, previous literatures about genetic programming and relative works are grouped into three parts. First group is the study that solves market timing by using genetic programming. Second group is the study that uses computer programs finding a combination of an index-tracking portfolio. The last group is the previous paper that studies about trading strategy and tracking portfolio in Thailand.

2.2.1 Solving Market Timing by Genetic Programming

Firstly, it is foreign literatures applying genetic programming to solve the market timing. They use genetic programming to get a profitable technical strategy that finds the timing of investments and generates excess returns.

F. Allen and R. Karjalainen (1999) used genetic algorithms to find technical trading rules. This paper was the first one that uses genetic programming to identify profitable trading rules in a stock market. They used daily data for the S&P500 index from 1928 to 1995. They divided up the data into successive in-sample periods, consisting of seven years, further sub-divided into five years training and two years for selection for each set of ten runs. Furthermore, they used one-month T-bill as a risk free rate. After transaction costs, the rules did not earn consistent excess return over a buy-and-hold strategy. However, the rules are able to identify periods that traders should be in the index (buy) when daily returns are positive and volatility is low. On the other hand, traders should be out the index (sell) when the reverse is true. Finally, these latter results can largely be explained by low-order serial correlation in stock index return.

C. Neely, P. Weller and R. Dittmar (1997) used the genetic programming technique that Allen and Karjalainen (1995) argued in their working paper. They found strong evidence of economically significant out-of-sample excess return in foreign exchange market. They used daily data of six exchange rate series namely \$/DM, \$/¥, \$/£, \$/SF, DM/¥, and £/SF in the period 1981 to 1995. Furthermore, when the dollar/Deutsche mark rules are allowed to determine trades in other markets, there is significant improvement in performance in all cases, except for Deutsche mark/yen. Moreover, they used the bootstrapping to determine whether observed performance of a trading rule is likely to have been generated under a given model. Bootstrapping results on the dollar/Deutsche mark indicate that the trading rules detect patterns in the data that are not captured by standard statistical models.

J. Potvin, P. Soriano and M. Vallee (2004) applied the genetic programming as a means to automatically generate such short-term trading rules on the stock markets, because they believed that technical analysis was capable of exploiting short-term fluctuations on the financial markets. In this paper, their trading rules were generated from not only prices and lag of price as Allen and Karjalainen (1999) but also the other technical term as volume and technical functions, namely RSI and ROC, in their genetic programming. Rather than using a composite index for this purpose, the trading rules are adjusted to 14 Canadian companies in on the Toronto stock exchange market. In conclusion, they found that the trading rules generated by GP are generally beneficial when the market falls or when it is stable. On the other hand, these rules do not match the buy-and-hold approach when the market is rising.

S. Chen and C. Yeh (1997) used a biological-based search program, genetic programming, to formalizes the notion of unpredictability in the efficient market hypothesis (EMH). The EMH will be exemplified by the application to the Taiwan and US stock market. A short-term sample of TAIEX and S&P 500 with the highest complexity defined by Rissanen’s minimum description length principle (MDLP) is chosen and tested. They found that a linear model couldn’t predict better than random walk and GP-based search could beat random walk by 50%. Moreover, they concluded that the search costs of discovering the nonlinear regularities might be too high to make the exploitation of these profitable regulations. Hence, the efficient market hypothesis sustained.

C. Neely (2001) used the genetic programming as same as Allen and Karjalainen (1999) in order to test a profitable trading rules on S&P 500 with a risk adjusted return and generate ex ante rules with improved performance. In paper of Allen and Karjalainen, they concluded that genetic programming was not a profitable strategy, however genetic programming can help to reduce volatility. To test this argument, Neely used a risk adjusted return as the fitness measurement of trading rules. He applied Sharpe ratio, Sweeney & Lee's X* statistic, Dacorogna et al's X_{eff} measure, and Jansen's α as a risk adjusted return. Moreover he also changes the day, which is an out of market day earning interest rate, from the business day to the calendar day. He found that the rules were not significantly out performing a buy-and-hold strategy on a risk-adjusted basis. Nevertheless, risk-adjustment techniques should be seriously considered when evaluating trading strategies.

M. Seshadri (2003) presented the genetic programming methodologies to find successful and understandable technical trading rules for the S&P500 index. The thesis used a complexity-penalizing factor to avoid overfitting and improved comprehensibility of the rules produced by genetic programming. This paper also presented different cooperative coevolutionary genetic programming strategies and found that paired collaborator coevolution could give the best result. According to the results, some methods described rules that beat the S&P500 with 99% significance.

2.2.2 Solving Index Tracking by Computer Programs

Second section, it is foreign literatures using computer programs to find an index-tracking portfolio. Computer programs are about to find both what security names and what weights of securities that generate return as close as index return.

J. Shapcott (1992) used genetic algorithms and quadratic programming in order to solve the passive portfolio selection. The genetic algorithm generates the subsets and finds both performance and proportion of available capital that should be invested in each member company. His thesis applied the genetic algorithms to construct 20-company portfolios tracking FTSE-100 share index. The results reported that the genetic algorithms using migration had been found to find a better solution than using isolated subpopulations in most case.

J.E. Beasley, N. Meade and T.J. Chang (2003) presented an evolutionary heuristic for the solution of the index-tracking problem. The index-tracking problem is the problem of reproducing the performance of a stock market index, but without purchasing all of the stocks that make up the index. Their formulation explicitly included transaction costs (associate with buying and selling stock) and a limit on total transaction cost of each portfolio rebalance. They solved the problem of five data sets from major world markets, namely Hang Seng, DAX, FTSE, S&P and Nikkei. They concluded that an evolutionary heuristic is an alternative way to solve the index-tracking problem.

2.2.3 Relative Studies in Thailand

Last section reviews literatures in Thailand that relate with genetic programming, market timing and index tracking. As other countries, Genetic programming is used to find the approximate solution in many ways, but there are a few papers in the financial field, especially the investment.

Leemakdej (2003) presented the portfolio formation on the efficient frontier based on order of expected return of Xia et al (2000). In order to construct the portfolio, he cannot from the portfolio by a normal method as Markowitz (1952) that use expected return from historical data. Hence, he presented genetic algorithm as quadratic programming to solve the portfolio-forming problem. Finally, he suggested how to use GOAL, which is GA freeware, to find the portfolio.

Khanthavit (2000) examined the performance of competing portfolios to track the SET and SET50 indexes. He found that the market capitalization/trading liquidity criterion led to a better tracking performance. These were no clear evidence to suggest the better estimation technique. The tracking error did not grow with high target excess return. Finally, performances could be improved by adding more stocks to portfolio.

CHAPTER III

Methodology

This study outstandingly differs from the previous studies in Thailand. This is the pioneer one that applies genetic programming to solve the investment problems in Thailand. This paper mainly focuses on market timing and finds the profitable technical trading strategy on SET index. Moreover, the index-tracking portfolio is also established by genetic programming in order to make the trading strategy becoming tangible because the investors cannot directly invest in SET index.

This chapter illustrates the algorithms and steps of genetic programming and how it operates to get the best solution for each problem. However, this chapter explains only a general situation of genetic programming. Next chapter is about to explain in detail of its applications to solve the problems of market timing and index tracking.

This chapter explains the applications of using genetic programming to solve the market timing and index tracking. Each application requires the appropriate function and terminal sets to generate the right answer. Furthermore, the right fitness measurement is essential to select the right best solution. So, this chapter is separated into two parts. First part illustrates function and terminal sets of market timing application and its fitness measurement. Second part demonstrates details of market timing application as the first part.

3.1 The Application of Market Timing

This part explains the details of function and terminal sets and the fitness measurement of market timing application. The solution of market timing is the trading strategy beating the market. The trading strategy generates the signals indicating periods of investment as the technical analysis. Normally, the signal is determined from historical prices and/or trading volumes. These variables are terminal sets of genetic programming. The method of technical analysis calculations is not complex. It uses only arithmetic and relational operations that are the function sets. Although the trading strategy uses easy calculations, it can find the excess return over the market when the market is inefficient. So, the performance of trading strategy is how much of its abnormal return earn over the SET index.

3.1.1 Function and Terminal Sets

Function and terminal sets must be suitable with the problem. It is essential that genetic programming is going to find the best solution from the appropriate function and terminal sets. If inputs of genetic programming are not suitable with problem, it is impossible that genetic programming can find the right answer. In market timing application, its terminal sets are a variable of the SET index's level and constants, besides function sets are operators using to compute variables such as plus, minus, function of average, lag, maximum and minimum. Moreover, this study uses the function that is a component in the famous tools of the technical analysis, for example the relative strength function is the unique function of RSI. The function and terminal sets in this study are unique units of almost technical analysis tools. This paper defines function and terminal sets of genetic programming as follow:

Function:

Arithmetic operation: +, -, ×, ÷

Boolean operation: and, or, not

Relational operation: <, >

Real function:

avg(v, n):	average of variable over the past n days
e-avg(v, n):	exponential average of variable over the past n days
lag(v, n):	variable is lagged by n days
max(v, n):	maximum value of variable over the past n days
min(v, n):	minimum value of variable over the past n days
RS(v, n):	relative strength of variable over the past n days
diff(v ,v):	absolute value of the difference between two real numbers

Terminal:

Constant:

Real: constant of interval [1, 250] where 250 is the approximate number of working days in a year

Real Variable:

price: level of the SET index.

3.1.2 Fitness Measurement

This part illustrates the fitness measurement of market timing. In this application, the parameter indicating the performance of genetic programming is the excess return over the SET index. The excess return is difference value between the returns of buy-and-hold strategy and the strategy of solution. Returns of buy-and-hold strategy represent the return of the market. Returns of the solution are calculated from daily returns of the SET index and timing of each solution.

In each trade in the market, returns are generally calculated from buy prices (P_{b_i}), sell prices (P_{s_i}), and one-way commission fee (c) as follow:

$$\pi_i = \frac{P_{s_i}}{P_{b_i}} \times \frac{1-c}{1+c} - 1$$

In this study, a return of each trading is considered in term of a summation of continuous daily returns (r_t) in a holding period. Each trading return is calculated as follow:

$$\begin{aligned} \pi_i &= \exp\left[\sum_{t=b_i+1}^{s_i} r_t\right] \times \frac{1-c}{1+c} - 1 \\ &= \exp\left[\sum_{t=b_i+1}^{s_i} r_t + \ln\frac{1-c}{1+c}\right] - 1 \end{aligned}$$

Returns of a trading rule found by genetic programming come from two sources—the market returns and saving returns. Principally, trading rules is going to generating buy or sell signal. Thus, the daily returns equal the market returns when signal is buy and the daily returns equal the saving rates of commercial bank when signal is sell. This paper employs two dummy variables of $I_b(t)$ and $I_s(t)$. The dummy variable $I_b(t)$ equals 1 when a trading strategy generates buy signal and otherwise is zero. The dummy variable $I_s(t)$ equals 1 when a trading strategy generates sell signal and otherwise is zero. On the other word, two dummy variables have the relationship of $I_b(t) \times I_s(t) = 0 \,\forall\, t$. The total return is subtracted by transaction fee of n times. Hence, the annual return of trading rules found by genetic programming is:

$$\pi = \exp(r) - 1$$

$$r = \sum_{t=1}^{T} r_t \, I_b(t) + \sum_{t=1}^{T} r_f \, I_s(t) + n \;\ln \frac{1-c}{1+c}$$

This paper designs the abnormal returns over buy-and-hold strategy as the fitness of genetic programming in the market timing part. The returns of over buy-and-hold strategy (r_{bh}) are determined by:

$$r_{bh} = \sum_{t=1}^{T} r_t + \ln \frac{1-c}{1+c}$$

So, the fitness of each solution is the excess return ($\Delta\pi$) over buy-and-hold strategy determined by:

$$\Delta\pi = \exp[r - r_{bh}] - 1$$

3.2 The Application of Index Tracking

This section explains the detail of index tracking application. It is not only the function and terminal sets and the fitness measurement must be changed to fit with the characteristic of problem but the processes of genetic programming also might be changed.

In order to use genetic programming solving index-tracking problem, the processes of solution development are difference from using genetic programming to solve market-timing problem. This study defines crossover and mutation process to suitable with the problem. In crossover process, a pair of solutions is selected by their fitness as general. Besides, the same stocks in chromosome of their parents are transferred to their offspring and other stocks are pooled in a bag. Their offspring is full filled by randomly selecting from the bag. The weights of stocks are attached with their names. Although the process of crossover is difference in each problem, the main idea do not still change that crossover is an evolution process by creating offspring from the characteristic of their parents. Moreover, the evolution process that makes offspring changed from their parents as mutation still exists in index tracking as well. The mutation randomly changes names or weights of the offspring.

3.2.1 Function and Terminal Sets

For index tracking application, function and terminal sets are security names and weights of securities. This paper selects stocks that are the member of the Stock Exchange of Thailand. Besides these stocks have ever been member of SET-50 index, which is Thai index consist of 50 stocks selected by market capitalization and trading volume, in 1995 – 2003. This paper focuses of large market capitalization because SET index is weighted-average index. If the stocks have larger market capitalization, it has more effect to index change. The details of securities used in this study are displayed in appendix A.

3.1.2 Fitness Measurement

In index tracking part, this paper identifies that the highest performance of a tracking portfolio is the most similar returns as the market index, because the main purpose of this part is generating a index tracking portfolio with limitation of number of stock holding. This work uses genetic programming to construct a portfolio with n firms where n is 5, 10 and 15 firms.

Daily returns of the tracking portfolio (r_t) are calculated by:

$$r_t = \sum_{i=1}^{N} w_{it-1} r_{it}$$

w_{it-1} is the weight of security i at time t-1

r_{it} is the daily continuous return of security i at time t

Therefore, fitness of genetic programming is defined in term of error of tracking between a tracking portfolio (r_t) and the market index (R_t). This study uses root mean square error (RMSE, ε) as the fitness measurement determined by:

$$\varepsilon = \sqrt{\frac{\sum_{t=1}^{T} \left| r_t - R_t \right|^2}{T}}$$

Lastly, if there is the abnormal return from a trading strategy and a tracking portfolio found by genetic programming, it means EMH of Thai market is inefficient. In other word, investors can invest with the right strategy and possible earn abnormal returns.

CHAPTER IV

The Result

This paper uses genetic programming to solve the investment problem of market timing and index tracking. To find the best solution for each problem, genetic programming is going to randomly create the population of solution from given function and terminate sets. Next step, genetic programming is about to evolve the new solution in next generations. Thank to crossover and mutation processes, genetic programming can crate the varied offspring that probably have higher performance than its parent.

Chapter IV reports the results of market timing and index tracking in terms of returns, standard deviations, and statistic values. The results of this chapter are classified into two parts namely in sample and out sample periods. Both market timing and index tracking are separately tested of in sample period. Besides, the results indicate that genetic programming generates profitable trading rules and tracking portfolios earning as closely as SET index return. The second part is results of out sample period. The results are created from combinations of market timing and index tracking solutions. These results point out whether it is possible or not that the trading rule and the tracking portfolio can come true in the real world. Moreover, this sector can interpret the stage of EMH in the Stock Exchange of Thailand.

4.1 In Sample - Market Timing

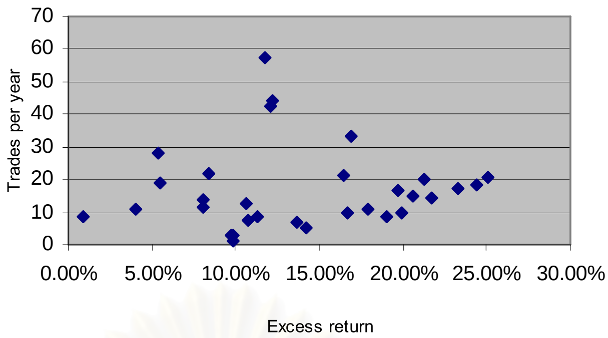
First part reports the results of trading rules found by genetic programming with in sample period. Table 1 depicts the results of trading rules that have 3-year and 5-year training periods in order to compare the effect of training year. In sample period, genetic programming can find strategies generating a positive return in each testing period. The average returns of buy signals are positive and the average returns of sell signals are negative for all period as well. The difference returns between buy and sell signals are significantly positive confirmed by ***t***-statistics both 3-year and 5-year training periods.

In period of 1994-1998, the Stock Exchange of Thailand is the bear market because of the Gulf War. Consequently trading rules found by genetic programming dramatically decline because of inappropriate data in trading periods. The bad rules generating only out of market signals can beat the market because SET index crashes. It is a bad parent in order to create the new offspring with a better performance. Thus, the number of trading rules found by genetic programming is less than other periods. However, genetic programming can also find a few trading rules making abnormal returns in the crisis.

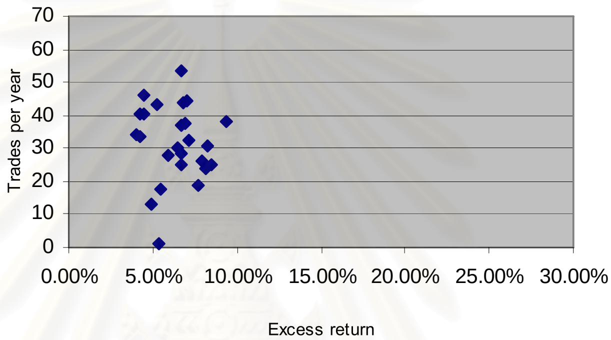
In addition, this paper concerns the numbers of transactions because of commission fee. More trades, more commission investors lose. Figure 4 illustrates relationship between numbers of trades per year and annual excess returns. The results of 3-year training period are obviously more scatter than 5-year training period. It implies that genetic programming can find more stable solution in term of excess returns and number of trading when the training period is longer. Furthermore, all of returns from both training periods are always positive. It means that genetic programming can find the rule generating excess return in the all in sample period.

Table 1
A summary of the results for trading rules found by genetic programming in training period. Panel A shows the results with 3-year training period which start from 1993 to 2000. Panel B shows the results with 5-year training period which start at the same date as panel A. The first column shows the in -sample training and selection period. K is number of trading rules found by the algorithm. Excess denotes the average yearly excess return in the out of sample period above the buy-and-hold strategy after transaction costs. The level of transaction cost is 0.25% which the most reasonable case. K^{+} is the number of rules with a positive excess return. Each trading rule divides into periods 'in' (long in the market) and 'out' of market(earning a risk free rate of return). Average daily returns during 'in' and 'out' periods are denoted by r_b and r_s respectively and standards of daily returns denoted by σ_b and σ_s and number of days by N_b and N_s . T* is the number of t-statistics for r_b-r_s significant at the 5% level.

In-sample	**K**	**Excess**	**K^{+}**	**N_b**	**r_b**	**σ_b**	**N_s**	**r_s**	**σ_s**	**r_b-r_s**	**T***
Panel A : 3-year training period											
1995-1997	3	+ 0.2948	3	65	+ 0.0089	0.0162	654	- 0.0020	0.0173	+ 0.0109	2
1996-1998	9	+ 0.3963	9	142	+ 0.0038	0.0324	578	- 0.0029	0.0208	+ 0.0066	9
1997-1999	10	+ 0.5429	10	205	+ 0.0044	0.0317	514	- 0.0026	0.0227	+ 0.0070	10
1998-2000	10	+ 0.5898	10	331	+ 0.0034	0.0266	388	- 0.0031	0.0221	+ 0.0065	10
Panel B: 5- year training period											
1995-1999	10	+ 0.3760	10	433	+ 0.0025	0.0247	766	- 0.0025	0.0193	+ 0.0050	10
1996-2000	10	+ 0.3427	10	356	+ 0.0027	0.0269	843	- 0.0025	0.0204	+ 0.0052	10



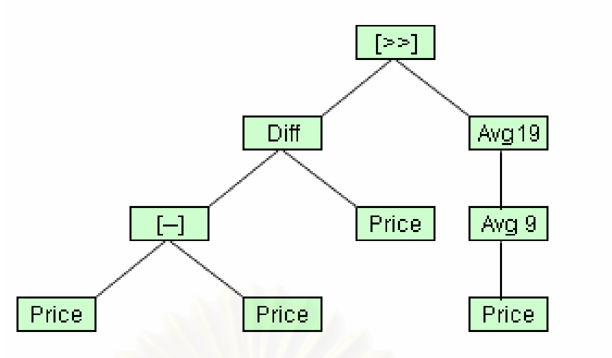
Panel A In-sample results with 3-year training period



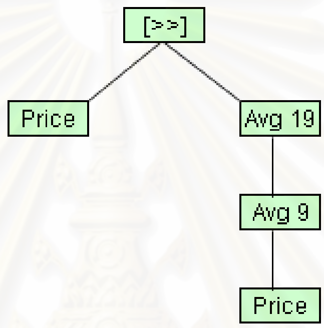
Panel B In-sample results with 5-year training period

Fig 4 The annual excess returns and the number of trades per year of trading rules found by genetic programming.

To clearly understand genetic programming and its applications, this sector is about to explain in more detail with an example of market timing and index tracking. The solution of market timing is a syntax tree. Figure 5 shows the example of the best solution found in 3-year trading period. Each node of the trading rule represents either a function or a variable defined by a user. The example tree diagram in genetic programming is a complex tree. Thank to the complex of solution, the genetic programming can create more varied offspring in order to find a better solution. However, the complex tree can be summarized to the simple one and easy to understand. Figure X shows the example solution that can be reduced to expression "Price > Average 19 day of (Average of 9 day of Price)".



a) The full tree diagram



b) The summary tree diagram

Figure 5 The example tree diagram of the trading rule found by the genetic programming with 3-year training period of 1997-1999.

The rule is surviving with SET index in period of 1997 to 1999 as shown in figure 6. The new better rule comes from more appropriate data as a side way of index. While the index goes side way, genetic programming can more easily find a new solution generating higher excess return than pure bull or bear market. Thank to a side way market, a new solution learns to create a sell position at a peak and buy back at a bottom.

Figure 6 The level of SET index between 1997 and 2000.

The tree is going to generate buy and sell signals from its operation. This tree requires historical price as input and generates two lines of signals. As other technical indicators, the signals of the trading rule compose of two lines—fast and slow lines—displayed in figure 7. It is a buy signal when the fast line crosses the slow one up. On the other hand, it is a sell signal when the fast line crosses the other down. From its signal, this work uses them as criteria of the investment decision in the market or not. Figure 8 shows proportions of a portfolio. The proportion of the portfolio is zero when the trading rule generates a sell signal. In the same way, the proportion is one when the trading rule generates a buy signal. The proportion of zero represents selling the portfolio earning as holding cash and the proportion of one represents holding the portfolio earning as index return.

To illustrate how well the trading rule beat the index in training period. Figure 8 shows the cumulative daily return of the trading rule. The cumulative return is calculated by continuous summation of daily index returns with one-day lag while the rule generating buy signals. Additional, the cumulative return is added up by the daily saving rate of commercial bank as holding cash while the rule is generating buy signals.

Figure 7 The indicator generated by the trading rule.

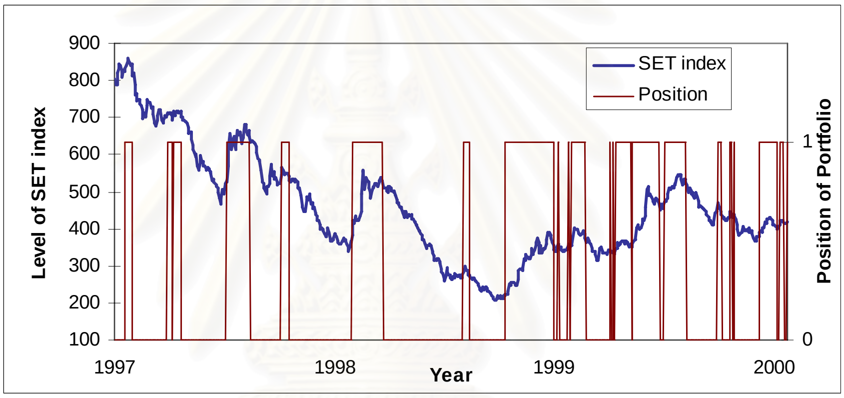


Figure 8 The trading position of the portfolio due to the buy and sell signals.

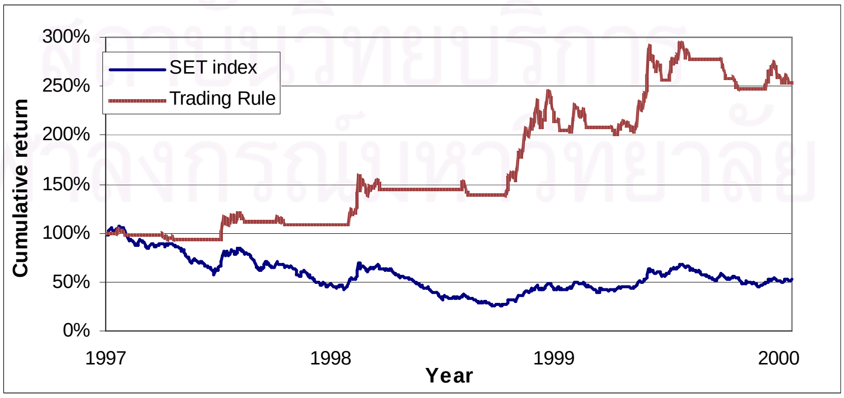
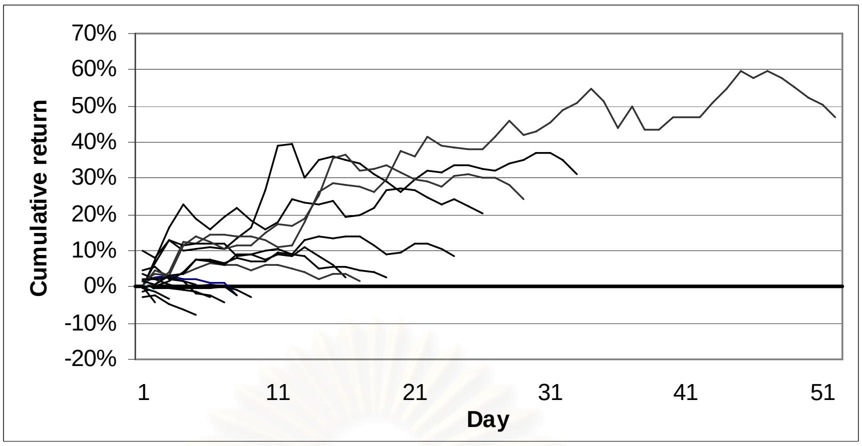
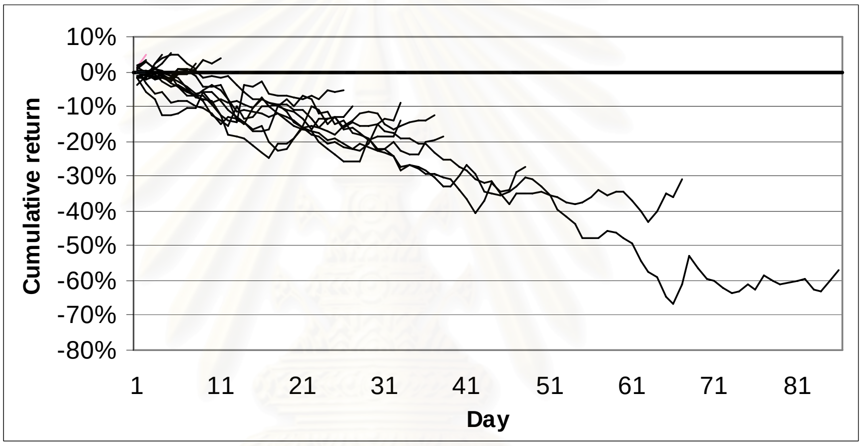


Figure 9 The cumulative returns of trading portfolio compared with SET index between 1997 and 2000.



Panel A Cumulative return of each buy signal



Panel B Cumulative return of each sell signal

Figure 10 The cumulative return of each trading while the trading rule is generating buy or sell signals.

Previous results indicate that the trading rule found by genetic programming generates positive excess returns. It is clearly understand that its signals give investors tendencies of the market. According to figure 10, it shows a cumulative return of each trade. The returns are accumulated until a signal changes from buy to sell or vice versa. Obviously, the cumulative return of a buy signal gives a positive value at the end of holding period. Besides, the cumulative return by a sell signal gives a negative value at the end of the period as well. It can be implied that a trading rule found by genetic programming can classify the bull period out from the bear time.

4.2 In Sample - Index Tracking

Second part presents the results of index tracking by portfolio found by GP. GP is going to form index-tracking portfolios with 5, 10 and 15 companies. Furthermore, GP has 3-year and 5-year training periods as same as the market timing part in order to combine the best solutions of market timing and index tracking for out sample tests. The tracking portfolios are constructed and rolled over every 6 months and used to track the index for next 6 months after 3-year training period and 1-year testing period.

In this part, the results report the error of tracking that is difference daily return between SET index and tracking portfolios. The error represents the performance of a tracking portfolio. Thus, the high performance portfolio interprets the portfolio generating daily returns as same as the index. The error is measured in many ways such as mean of error, a correlation with the index, RMSE and MAE as in table 2.

Table 2 shows the results of error between tracking portfolio found by genetic programming and SET index. T is the number of t-statistics for the error significant at the 5% of confidence levels.

Number of company	Error		ρ	RMSE	MAE	T
	Mean	SD.				
Panel A : 3-year training period						
5	0.0132	0.1007	0.8934	0.1375	0.1094	1
10	-0.0084	0.0831	0.8993	0.1190	0.0956	0
15	0.0173	0.0725	0.9265	0.1079	0.0871	2
Panel B : 5-year training period						
5	-0.0051	0.055	0.9612	0.0572	0.0457	0
10	-0.0001	0.0334	0.9848	0.0345	0.0273	0
15	-0.0016	0.0273	0.9885	0.0281	0.0222	0

All methods consistently indicate that 15-company portfolio has lower tracking error than 10-company and 5-company portfolio respectively. Besides, the tracking portfolio with 5-year training period tends to have higher tracking capability than 3-year training period. For example, the 5-company portfolio with 3-year training period having the correlation with the index of 0.8934, RMSE of 0.0265 and MAE of 0.1094 is dramatically difference from 15-company portfolio with 5-yr training period having the correlation of 0.9885, RMSE of 0.0011 and MAE of 0.0222.

Furthermore, this part has the hypothesis that the portfolio found by genetic programming can track SET index with zero error. Thus, ***t***-statistic is use to verify the hypothesis. According to the statistic values in table 2, there are few solutions having error from zero with 3-year training period, however they are one or two of 56 portfolios that is a small proportion. Moreover, the results illustrate that no error is significantly different from zero with 5-year training period.

In brief, the results of index tracking show that the portfolios found by genetic programming have higher correlations when more companies in portfolio and longer training period. In the same way with correlation, RMSE and MAE have a tendency to decline when numbers of companies and training years increase.

Table 3 shows the member and weight of the example 10-company tracking portfolio in period of July 1997 to June 2000.

No.	Company	% Weight
1	BECL	1.9%
2	BT	5.8%
3	BIGC	17.3%
4	PTTEP	9.6%
5	EGCOMP	15.4%
6	SCC	3.8%
7	NFS	7.7%
8	MAKRO	13.5%
9	TMB	15.4%
10	SSI	9.6%
	Sum	100.0%

In order to understand about the genetic programming application with the index-tracking problem, this part is about to explain the detail of tracking portfolio. The example of tracking portfolio is the best 10-company portfolio with 3-year period of July 1997 to June 2000. The members of the example portfolio and their weight are displayed in table 3. The solution of this problem is not a tree diagram as in market timing part because the order of variables does not affect its performance. On the other word, the members of tracking portfolio can switch their positions with no effect to over all performance.

The tracking portfolio is evaluated by genetic programming to have returns as same as SET index in given period. Figure 11 compares the cumulative daily return between the tracking portfolio and the index. Although, this portfolio is the best solution of the tracking population, there is large error between the tracking portfolio and the index because of low numbers of companies used in the tracking portfolio.

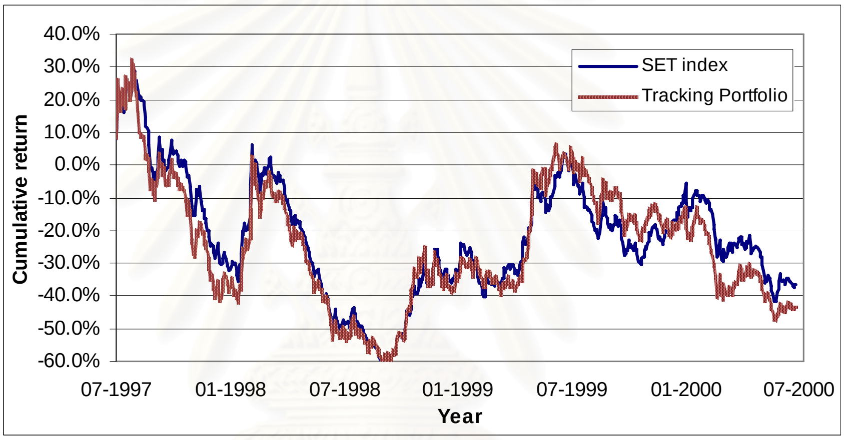


Figure 11 The cumulative daily return of the 10-company tracking portfolio and SET index in period of July 1997 to June 2000

4.3 Out of Sample – The Combination of Market Timing and Index Tracking

This chapter verifies the solutions of market timing and index tracking whether they can really make excess returns or not by out of sample test. The out of sample test is about to combine the best solutions of market timing and index tracking in each period. As in sample part, the solution of market timing generates buy and sell signals in order to beat the market or the index. The return of portfolio in buy signal equals to the return of the tracking portfolio that is formed from the best combination of in sample period. The return in sell signal equals to the saving rate of commercial banks. Furthermore, the overall return includes commission fee of 0.25 percent in each transaction. After the fee, the overall return of portfolio is compared with SET index to find the excess return.

Furthermore, the differences between average daily returns of buy and sell signals are generally positive for all period. This conclusion is confirmed by statistic values. The ***t***-test is applied to verify the capability of trading rules to classify bull and bear periods. Although the results are not significant at confidence level of 95%, the combinations can separate bull and bear market in the out sample period with confidence level of 90%. It is the little evidence that genetic programming can find the combination of a trading rule and a tracking portfolio beating the SET index.

The statistic in the out of sample period is disturbed from many events in the SET. In 1999, the market has just recovered from financial crisis. Investors have the question whether SET index is in the bottom or not. Thus SET index is very volatile, however the solution found by genetic programming can indicate the period of investment with no strongly significant. It is possible that statistic numbers are too low in 1999 because standard deviation of return is high. Furthermore, in 2000, SET index turn to recession again and almost combinations of solutions easily beat the market. Moreover, in 2001, there is an unbelievable event of 9-11, which the world trade centers are sabotaged. The world market panics and over reacts so SET index dramatically drop in next trading day. The tracking portfolio severely decline than the index because it almost invests in big market capitalization that everyone has already had and make selling order.

In point of view of the researcher, the best period of evaluation is 2002. In this year, the index has low volatility and smoothly moves up and down as a natural movement. The statistic number confirms that the combinations of the solution can beat the market in a normal situation.

Moreover, a trading rule having longer training period has a proclivity to generate higher and more stable excess return for all numbers of companies of tracking portfolios.

Table 4 A summary of the results for trading rules found by the genetic programming in out of sample period. Panel A shows the results with 3-year training period which start from 1999 to 2002. Panel B shows the results with 5-year training period which start at the same date as panel A. The first column shows the out sample training and selection period. K is number of trading rules found by the algorithm. Excess denotes the average yearly excess return in the out sample period above the buy-and-hold strategy after transaction costs. The level of transaction cost is 0.25% which the most reasonable case. K+ is the number of rules with a positive excess return. Each trading rule divides into periods 'in' (long in the market) and 'out' of market(earning a risk free rate of return). Average daily returns during 'in' and 'out' periods are denoted by r_b and r_s respectively, with standard of daily return denoted by σ_b and σ_s and number of days by N_b and Ns . T 95% and T 90% are the numbers of t-statistics for r_b- r_s significant at the 5% and 10% of confidence levels.

I. Tracking portfolio consists of 5 companies.

Out of sample	K	Excess	K^+	N_b	r_b	σ_b	N_s	r_s	σ_s	r_b-r_s	T 95%	T 90%
Panel A : 3-year training period												
1999	3	+ 0.0596	3	96	0.0035	0.0268	145	0.0011	0.0226	+ 0.0024	0	0
2000	9	+ 0.4679	9	44	-0.0023	0.0204	197	-0.0026	0.0208	+ 0.0004	0	2
2001	10	+ 0.0131	6	113	0.0014	0.0165	128	-0.0004	0.0180	+ 0.0019	3	4
2002	10	+ 0.0004	4	138	0.0014	0.0159	99	-0.0008	0.0158	+ 0.0022	1	4
Panel B : 5-year training period												
2001	10	+ 0.0829	9	121	0.0021	0.0177	121	-0.0011	0.0213	+ 0.0032	1	4
2002	10	+ 0.0180	9	116	0.0000	0.0097	125	-0.0009	0.0129	+ 0.0009	0	1

II. Tracking portfolio consists of 10 companies.

Out of sample	K	Excess	K^+	N_b	r_b	σ_b	N_s	r_s	σ_s	r_b-r_s	T 95%	T 90%
Panel A : 3-year training period												
1999	3	- 0.0982	0	96	0.0019	0.0228	145	0.0009	0.0221	+ 0.0010	0	0
2000	9	+ 0.4507	9	44	-0.0024	0.0191	197	-0.0027	0.0199	+ 0.0003	0	2
2001	10	- 0.0336	6	113	0.0011	0.0164	128	0.0000	0.0183	+ 0.0011	0	4
2002	10	+ 0.0033	4	138	0.0014	0.0151	99	-0.0005	0.0155	+ 0.0019	1	4
Panel B : 5-year training period												
2001	10	+ 0.0928	9	121	0.0022	0.0163	121	-0.0013	0.0194	+ 0.0034	3	7
2002	10	+ 0.0459	9	116	0.0001	0.0095	125	-0.0018	0.0131	+ 0.0019	3	7

III. Tracking portfolio consists of 15 companies.

Out of sample	K	Excess	K^+	N_b	r_b	σ_b	N_s	r_s	σ_s	r_b-r_s	T 95%	T 90%
Panel A : 3-year training period												
1999	3	- 0.0431	0	96	0.0025	0.0242	145	0.0009	0.0223	+ 0.0016	0	0
2000	9	+ 0.4395	9	44	-0.0028	0.0199	197	-0.0028	0.0205	- 0.0001	0	2
2001	10	- 0.0339	6	113	0.0010	0.0175	128	-0.0004	0.0196	+ 0.0014	1	4
2002	10	+ 0.2186	10	138	0.0030	0.0152	99	0.0001	0.0155	+ 0.0029	4	4
Panel B : 5-year training period												
2001	10	+ 0.0580	9	121	0.0019	0.0153	121	-0.0009	0.0203	+ 0.0028	0	4
2002	10	+ 0.0445	9	116	-0.0001	0.0102	125	-0.0019	0.0125	+ 0.0017	1	6

4.4 Summary

This paper applies genetic programming to solve two problems of the investment. First problem is the market timing. Genetic programming is going to find a trading strategy that can beat the market or classify bull markets from the bear market. Second one is the index-tracking problem. Genetic programming is going to find a combination of companies to form a portfolio generating the daily return as same as the daily index return.

The results indicate that genetic programming is able to find the profitable trading strategy in training periods. The performance of the trading rule depends on training periods. If the index is going side way in training period, the rule tends to generate the accurate signals that divide bull and bear market and also earn higher positive excess return. If the market is either pure bull or bear, the genetic programming will create the poor trading offspring.

Furthermore, genetic programming forms the tracking portfolio earning as same as the index return, although it has a limitation of members such as 5, 10, and 15 companies. The results point out that the errors both RMSE and MAE have a tendency to decline when the members of tracking portfolio increase. Besides, the errors also tend to shrink when genetic programming has longer training periods.

In the out of sample period, the combinations of the market timing and index tracking solutions are generating positive excess returns over SET index in almost periods. Besides, there are little evidences indicating that the trading strategy identifies the bull market from the bear time. Moreover, the combinations also have a proclivity to generate higher and more stable excess return when numbers of training years increase.

In addition, this paper finds the evidence interpreting that genetic programming can find the profitable combination of the trading strategy and the tracking portfolio. The combination earns positive excess returns over SET index in some study periods. Hence, this paper cannot conclude that the Stock Exchange of Thailand (SET) is the weak form of EMH.

CHAPTER V

Conclusion

This paper applies genetic programming to solve the investment problems of market timing and index tracking. Genetic programming is going to create the profitable technical trading strategy and the index-tracking portfolio. The combinations of the trading strategies and the portfolios are tested whether they can generate excess returns or not. If the combinations can beat SET index, this paper concludes that the Stock Exchange of Thailand is under weak form of EMH.

5.1 Conclusion

The major investment problem is the market timing, which is the problem of when the investors should go into or out of the market. Furthermore, the next serious problem is what security should invest. This paper use genetic programming to find the trading strategy on SET index to solve the market timing. Besides, genetic programming also finds the index-tracking portfolio that contents name and weight of securities. Thus, the combination of two solutions can solve the investment problems.

To find the solution for each problem, genetic programming is about to randomly generate the population of answers from given functions and variables. Genetic programming evolves the better solution in next generation by the natural ways such as crossover and mutation processes. New offspring have varieties of characteristics because the processes are recombining the chromosomes of their parent. So, they have both higher and lower performances than their parent. The better solutions are kept and become new parent for the next evolution. The evolutionary cycle is continuously going until users get the desired solution.

Genetic programming has applied to solve the financial problems in many areas. It has also solved the investment problem in both the stock and money markets. The previous papers found that genetic programming can earn the excess return such as Neely, Weller and Dittmar (1997), Seshadri (2003) and Potvin, Soriano and Vallee (2004). On the other hand, there are papers indicated that genetic programming cannot beat the market such as Allen and Karjalainen (1999), Chen and Yeh (1997) and Neely (2001). Moreover, a genetic algorithm and an evolutionary heuristic that are familiar of genetic programming are used to solve the index-tracking portfolio as Shapcott (1992) and Beasley, Meade and Chang (2003). Hence, genetic programming should be applied to solve the market timing and index tracking in Thailand. In addition, if the results indicate that the solutions found by genetic programming are generating excess returns, it means that the Stock Exchange of Thailand is in efficient.

In this paper, genetic programming learns by 3-year and 5-year training periods of SET index to find the profitable technical trading rule. Besides, the tracking portfolio is formed with 5, 10 and 15 companies and learns by 3-year and 5-year training periods as the trading strategy. In next step, the combinations of two solutions are verified that the combinations found by genetic programming can beat SET index or not in out of sample period. If the technical trading rule can generate excess returns and the tracking portfolio can earn as close as the index, it means that investors can beat the Stock Exchange of Thailand in the real world.

The results show that genetic programming significantly finds the profitable technical trading rules in training period. Genetic programming can more easily find the trading strategy when the index in training period goes sideway. Besides, the returns have tend to more stable when number of training year increases. The tracking portfolios have low error and high correlation with SET index in training period. The error has a tendency to shrink when number of members and number of training years increase. In out of sample, there is a little evidence to confirm the excess return over SET index. Hence, the paper concludes that genetic programming can earn excess return with appropriate condition and cannot summarize that Stock Exchange of Thailand is the weak form of EMH.

5.2 Suggestion

This paper focuses the market-timing problem. Thus, genetic programming is used to solve this problem by generating the trading rule on the index. Now, Thai investors cannot directly invest in SET index, so this works also use genetic programming to find both securities and weights to from the index-tracking portfolio. The result indicates that genetic programming can find the good tracking portfolio with low error and high correlation. Hence, further study should investigate whether the tracking portfolio found by genetic programming is better than the portfolio by other methods or not.

In this paper, the input of genetic programming is broadened in functions of the technical analysis but its variables of information such as close price does not enlarge. Thus, it would apply other information such as fundamental and economic data with genetic programming to solve the same problem. The fundamental trading rule is interesting, however genetic programming requires the desired fundamental function and variable as well.

References

Allen, F., and Karjalainen, R. (1999). Using Genetic Algorithm to Finding Technical Trading Rules. Journal of Financial Economics 51: 245-271.

Banzhaf, W., Nordin, P., Keller, R.E., and Francoen, F.D. (1998). Genetic programming—an introduction: on the automatic evolution of computer programs and its application. San Francisco: Morgan Kaufmann Publisher.

Bauer, R.J. Jr. (1994). Genetic Algorithms and Investment Strategies. New York: John Wiley & Sons.

Beasley, JE., Meade, N., and Chang, TJ. (2003). An evolutionary heuristic for the index tracking problem. European Journal Of Operational Research148: 621-643.

Bessembinder, H., and Chan, K. (1995). The profitability of technical trading rules in the Asian stock markets. Pacific-Basin Finance Journal 3: 257-284.

Blume, L., Easley, D., and O'Hara, M. (1994). Market statistic and technical analysis: The role of volume. Journal of Finance 1: 153-181.

Chen, S., and Yeh C. (1997). Toward a computable approach to the efficient market hypothesis: An application of genetic programming. Journal of Economic Dynamics and Control 21: 1043-1063.

Conrad, J., Hameed, A., and Niden, C. (1994). Volume and Autocovariances in Short-Horizon Individual Security Returns. Journal of Finance 4: 1305-1329.

Gunasekarage, A., and Power, D. (2001). The profitability of moving average trading rules in South Asian stock markets. Emerging Markets Review 2: 17-33.

Holland, J.H. (1975). Adaptation in Natural and Artificial Systems. 2nd ed. Ann Arbor: The University of Michigan Press.

Jacob, C. (2001) Illustrating Evolutionary Computation with Mathematica. San Francisco: Morgan Kaufmann Publisher.

Khanthavit, A. (2000). A shrinkage Weighting Scheme for Index-tracking Portfolio with Application in the Thai Stock Market. Bangkok: Thammasat University.

Koza, J.R. (1992). Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge: MIT Press.

Leemakdej, A. (2003). Stock Portfolio Formulation by Genetic Algorithm. Thammasat University working paper.

Neely, C., Weller, P., and Dittmar, R. (1997). Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach. Journal of Financial and Quantitative Analysis 32: 405-426.

Neely, C., (2001). Risk-Adjusted, Ex Ante, Optimal, Technical Trading Rules in Equity Markets. International Review of Economics and Finance 12: 69–87.

Potvin, J., Soriano, P., and Vallee, M. (2004). Generating trading rules on the stock markets with genetic programming. Computers & Operations Research 31: 1033–1047.

Seshadi, M. (2003). Comprehensibility, Overfitting and Co-Evolution in Genetic Programming for Technical Trading Rules. Master's Thesis. Faculty of Computer Science, Worcester Polytechnic Institute.

Shapcott, J. (1992). Index Tracking: Genetic Algorithm for Investment Portfolio Selection. Report EPCC-SS92-24, Edinburgh Parallel Computing Center, University of Edinburgh.

Appendix

Appendix A

List of companies that have ever been members of SET 50 between 1995-2003

No	Name	Company	Sector
1	ADVANC	Advanced Info Service Public Company Limited	Communication
2	AEONTS	Aeon Thana Sinsap (Thailand) Public Company Limited	Finance & Security
3	AMATA	Amata Corporation Public Company Limited	Property Development
4	AOT	Airports Of Thailand Public Company Limited	Transportation & Logistic
5	AP	Asian Property Development Public Company Limited	Property Development
6	ASL	Adkinson Securities Public Company Limited	Finance & Security
7	ASP	Asia Plus Securities Public Company Limited	Finance & Security
8	ATC	The Aromatics (Thailand) Public Company Limited	Petrochemical & Chemical
9	BANPU	Banpu Public Company Limited	Energy and Utility
10	BAY	Bank Of Ayudhya Public Company Limited	Banking
11	BBL	Bangkok Bank Public Company Limited	Banking
12	BEC	BEC World Public Company Limited	Entertainment & Recreation
13	BECL	Bangkok Expressway Public Company Limited	Transportation & Logistic
14	BIGC	Big C Supercenter Public Company Limited	Commerce
15	BLAND	Bangkok Land Public Company Limited	Property Development
16	BOA	The Bank Of Asia Public Company Limited	Banking
17	BT	Bankthai Public Company Limited	Banking
18	CCET	Cal-Comp Electronics (Thailand) Public Co., Ltd.	Electronic Component
19	CK	Ch. Karnchang Public Company Limited	Property Development
20	CNS	Capital Nomura Securities Public Company Limited	Finance & Security
21	COCO	The Cogeneration Public Company Limited	Energy and Utility
22	CPF	Charoen Pokphand Foods Public Company Limited	Agribusiness
23	DELTA	Delta Electronics (Thailand) Public Company Limited	Electronic Component
24	DTDB	DBS Thai Danu Bank Public Company Limited	Banking
25	NBANK	Thanachart Bank Public Company Limited	Banking
26	EGCOMP	Electricity Generating Public Company Limited	Energy and Utility
27	GOLD	Golden Land Property Development Public Company Limited	Property Development
28	GRAMMY	GMM Grammy Public Company Limited	Entertainment & Recreation
29	HANA	Hana Microelectronics Public Company Limited	Electronic Component
30	IFCT	Industrial Finance Corporation Of Thailand Public Co.,Ltd.	Banking
31	ITD	Italian-Thai Development Public Company Limited	Property Development
32	ITV	ITV Public Company Limited	Entertainment & Recreation
33	JAS	Jasmine International Public Company Limited	Communication
34	KEST	Kim Eng Securities (Thailand) Public Company Limited	Finance & Security
35	KGI	KGI Securities (Thailand) Public Company Limited	Finance & Security
36	KK	Kiatnakin Finance Public Company Limited	Finance & Security
37	KTB	Krung Thai Bank Public Company Limited	Banking
38	LALIN	Lalin Property Public Company Limited	Property Development
39	LH	Land And Houses Public Company Limited	Property Development
40	MAJOR	Major Cineplex Group Public Company Limited	Entertainment & Recreation
41	MAKRO	Siam Makro Public Company Limited	Commerce
42	MS	Millennium Steel Public Company Limited	Construction Material
43	NFS	National Finance Public Company Limited	Finance & Security
44	NPC	National Petrochemical Public Company Limited	Petrochemical & Chemical
45	NSM	Nakornthai Strip Mill Public Company Limited	Construction Material
46	PPPC	Phoenix Pulp And Paper Public Company Limited	Paper
47	PSL	Precious Shipping Public Company Limited	Transportation & Logistic
48	PTT	PTT Public Company Limited	Energy and Utility
49	PTTEP	PTT Exploration And Production Public Company	Energy and Utility
50	QH	Quality Houses Public Company Limited	Property Development

Appendix A (Cont)

List of companies that have ever been members of SET 50 between 1995-2003

No	Name	Company	Sector
51	RATCH	Ratchaburi Electricity Generating Holding Public Co.,Ltd.	Energy and Utility
52	SATTEL	Shin Satellite Public Company Limited	Communication
53	SCB	The Siam Commercial Bank Public Company Limited	Banking
54	SCC	The Siam Cement Public Company Limited	Construction Material
55	SCCC	Siam City Cement Public Company Limited	Construction Material
56	SCIB	Siam City Bank Public Company Limited	Banking
57	SHIN	Shin Corporation Public Company Limited	Communication
58	SIRI	Sansiri Public Company Limited	Property Development
59	SPL	Siam Panich Leasing Public Company Limited	Finance & Security
60	SSI	Sahaviriya Steel Industries Public Company Limited	Construction Material
61	STEC	Sino-Thai Engineering And Construction Public Co.,Ltd.	Property Development
62	SUC	Siam United Services Public Company Limited	Fashion
63	TRUE	True Corporation Public Company Limited	Communication
64	KBANK	Kasikornbank Public Company Limited	Banking
65	THAI	Thai Airways International Public Company Limited	Transportation & Logistic
66	TISCO	Tisco Bank Public Company Limited	Finance & Security
67	TMB	TMB Bank Public Company Limited	Banking
68	TOC	Thai Olefins Public Company Limited	Petrochemical & Chemical
69	TPC	Thai Plastic And Chemicals Public Company Limited	Petrochemical & Chemical
70	TPI	Thai Petrochemical Industry Public Company Limited	Petrochemical & Chemical
71	TPIPL	TPI Polene Public Company Limited	Construction Material
72	TT&T	TT&T Public Company Limited	Communication
73	TTA	Thoresen Thai Agencies Public Company Limited	Transportation & Logistic
74	TUF	Thai Union Frozen Products Public Company Limited	Food & Beverage
75	UBC	United Broadcasting Corporation Public Co., Ltd.	Entertainment & Recreation
76	UCOM	United Communication Industry Public Company Limited	Communication
77	VNG	Vanachai Group Public Company Limited	Construction Material
78	VNT	Vinythai Public Company Limited	Petrochemical & Chemical

BIOGRAPHY

Mr. Somjade Techa-intrawong was born on May 30th, 1981 at Bangkok, Thailand. He graduated a Bachelor of Mechanical Engineering at Mahidol University in 2002. After graduation, he studies in Master of Science in Finance at Chulalongkorn University in 2003.