

# Applying Genetic Algorithm to Support Index Fund Portfolio Strategy

Jui-Fang Chang, Gi-Yi Lai

Department of International Business  
National Kaohsiung University of Applied Sciences  
Kaohsiung, Taiwan

## Abstract

Index funds are popular investment tools currently being used in modern portfolio management; moreover, it has been observed that the performances of index funds are better than those of many other actively managed funds Elton, et al. (1996). The strategy is taken by fund managers when their portfolios will not necessarily outperform the market, thereby allowing fund managers to make necessary adjustments to reach average performance Oh, et al. (2005). In this study, we adopt the model of Oh, et al. (2005), and adjust the stock choosing method. Further, attempting to find the optimal index fund portfolio strategy in the stock market of Taiwan, we also use genetic algorithm to evaluate the performance of the index fund portfolio. Our main purpose is to report that an index fund could improve its performance greatly with the proposed genetic algorithm portfolio strategy, which will be demonstrated for index funds designed to track Taiwan Stock Price Index (TSPI).

**Keywords:** Index Fund Portfolio, Genetic Algorithm

## 1. Introduction

Index funds are designed to mimic the behavior of the given benchmark market indices (e.g. the S&P 500 in New York.). It is a popular investment tool being used in modern portfolios. The index fund is a type of passive investment strategy. Passive management is an investment strategy which does not propose to outperform the market. Hogan (1994) stresses that an index fund is an effective investment tool in modern portfolio theory. Chang (2004) also shows that maximum capital gains and growth funds have done worse than growth and income funds; actively-managed funds underperform a passive investment strategy; low risk funds outperform high risk funds; and no load funds outperform load funds. An index fund portfolio contains a relatively small number of stocks since it is not an effective investment strategy to include all stocks in the index fund portfolio. Thus, this study attempts to use a relatively small number of stocks and mimic the behavior of the benchmark index.

The genetic algorithm (GA) portfolio model was utilized for the index fund optimization. Three fundamental variables—standard error of

portfolio beta given by formula (3), average trading amount, and average market capitalization—were applied to the model. These are the main factors frequently used in analyzing and forecasting the stock market. In general, the purpose of this study is to propose a genetic algorithm (GA) portfolio model. In this research, the model consists of two steps. The first step is to select the stocks for the index fund through the company's variables by utilizing the benchmark index. The second step is that the relative weights of the selected stocks are optimized through the genetic algorithm process.

To demonstrate the usefulness of the proposed GA portfolio scheme, the Taiwan Stock Price Index (a major benchmark index in Taiwan Stock Exchange) from Jan. 2004 to Dec. 2005 is used. For the comparison with Oh et al's model, we used an algorithm which optimizes the weights by minimizing (4) over 5,000 random generations.

## 2. Literature review

### 2.1. Portfolio theory, index funds and tracking error

Markowitz (1952, 1959) proposes a “mean–variance low.” It provides a model for mean-variance within the portfolio and formulates expected returns and risks of a portfolio, respectively. Sharpe (1964) proposes the capital asset pricing model (CAPM), which is the origin of the index fund. In modern index fund portfolio theory, Andrews, et al (1986) propose three index fund models: “Full Replication Model,” “Stratified Model,” “Sampling Model.” Salkin, et al.(1989) provide a well-developed paradigm based on four index fund models “Estimated Coefficients –Nonstratified Model” “Estimated Coefficients–Stratified Model,” “Capitalization Weighted–Nonstratified Model” and “Capitalization Weighted–Stratified” Bogle (1998) examines the relationships among risk, return, and cost—showing that low-cost, passively managed index funds actually deliver the highest risk-adjusted returns in each category of mutual funds.

About index tracking, tracking error (TE) is measured by TE volatility. Fund managers always try to minimize the TE volatility level since it would produce as close as possible returns to the benchmark returns. Markus, et al. (1999) investigate four linear models for

minimizing the tracking error between the returns of a portfolio and a benchmark: (1) The Min-Max Model, (2) The Downside Min-Max Model, (3) The Mean Absolute Deviations (MAD) model, and (4) The Mean Absolute Downside Deviations (MADD) model. The study shows absolute deviations instead of using squared deviations, as is the case in the tracking error volatility model by Roll (1992). Thus, the responsibility of fund managers is to minimize the TE volatility.

### 3. Scheme specification

There are three variables which are frequently used in portfolio management: portfolio beta, trading amount, and market capitalization. In this study, we use these three fundamental variables to construct the function. This function will assist our selection of stocks within the portfolio.

#### 3.1. Model Frame

Trying to find the optimal index fund portfolio strategy in Taiwan stock market, the model of Oh, et al. (2005) was used to propose the adjusted portfolio model. The model of Oh, et al. (2005) is shown as follows:

$$P_{i(j)} = v_1 \{B_{i(j)}\}^{-1} + v_2 \bar{A}_{i(j)} + v_3 \bar{M}_{i(j)} \quad (1)$$

where  $v_1$ ,  $v_2$  and  $v_3$  are positively weighted. These are assigned before we process the

experiment.  $B_{i(j)} = \frac{\left\{ \sum_{t \in E} \frac{(r_{i(j)}(t) - \bar{r}_{i(j)})^2}{T-2} \right\}^{\frac{1}{2}}}{\left\{ \sum_{t \in E} \frac{(I_m(t) - \bar{I}_m)^2}{T-2} \right\}^{\frac{1}{2}}}$ , which

is the standard deviation of  $\hat{\beta}_{i(j)}$  (i.e.  $\hat{\beta}_{i(j)}$  is an estimate of  $\beta_{i(j)}$ ). In the experiment,  $\{B_{i(j)}\}^{-1}$ ,  $\bar{A}_{i(j)}$ ,  $\bar{M}_{i(j)}$  are scaled from 0 to 1. Note that  $P_{i(j)}$  eventually indicates the relative importance of individual stock in the portfolio.

Let  $n$  and  $l$  indicate the numbers of stocks for the benchmark index and index fund portfolio, respectively ( $l < n$ ). And we denote that  $C_k$  ( $K=1,2,3,4,5,\dots,l$ ) is the serial code of  $k$ th stock, which is included in the index fund portfolio. In other words, index fund portfolio set is  $\Phi_p = \{c_1, c_2, c_3, \dots, c_l\}$  which is selected from the entire  $n$  stocks. Let  $s$  denote the number of industry sectors comprising the benchmark index and  $d_i$  the number of stocks comprising  $i$ th industry sectors (i.e.  $\sum_{i=1}^s d_i = n$ ). In addition, for each  $j$ th stock of  $i$ th industry sector ( $j=1,2,\dots,d_i$  and  $i=1,2,\dots,s$ ), suppose the portfolio beta is given by  $\beta_{i(j)}$  where the sign  $i(j)$

is used to denote dependence of  $j$  on  $i$ . For that specific stock, allow  $r_{i(j)}(t)$ ,  $A_{i(j)}(t)$  and  $M_{i(j)}(t)$  to denote rate of return, market capitalization and trading amount at time  $t$ , respectively. Further,  $I_m(t)$  denotes the rate of return of the benchmark index  $m$  at  $t$ . Now, we define priority  $Y_{i(j)}$  for each company and  $\bar{X}$  mean  $\sum_{t \in E} X(t)/T$ . Below, unless otherwise stated,  $E$  is the training period of the portfolio,  $a_0$  is the starting point and  $T$  denotes the size. The detailed procedure of the genetic algorithm (GA) portfolio index fund scheme is given as follows:

Step 1. For the selected  $i_k$ , calculate  $Y_{i(j)}$  for  $j=1,2,\dots,d_{i(k)}$  and choose the stock having the highest priority (i.e.  $Y_{i(j)} = \max_{j=1,2,3,4,\dots,d_{i(k)}} Y_{i(k)} = Y_{i(k)}(j)$ ) until the procedure below for  $k=1,2,\dots,l$  selects  $l$  stocks for the portfolio. For  $\Phi_p = \{c_1, c_2, c_3, \dots, c_l\}$  established by Step 1, let  $W_k^m$  ( $k=1,2,3,4,\dots,l$ ) be the entire stock market capitalization. Note that  $\sum_{k=1}^l W_k^m < 1$  if  $l < n$ .

Step 2. Assign the optimal weights  $W_k^p : \sum_{k=1}^l W_k^p = 1$  ( $k=1,2,3,\dots,l$ ) to each stock in the portfolio which will be minimized (2) through the GA process.

$$Q(w_1, w_2, w_3, w_4, \dots, w_l) = \sum_{k=1}^l (w_k^p - w_k^m)^2 \sigma_k^2 \quad (2)$$

where  $\sigma_k^2 = \frac{\left\{ \sum_{t \in E} \frac{(r_{i(j)}(t) - \bar{r}_{i(j)})^2}{T-2} \right\}}{\left\{ \sum_{t \in E} \frac{(I_m(t) - \bar{I}_m)^2}{T-2} \right\}}$ , which is the variance of  $\hat{\beta}_k$  for  $k$ th stock constituting the portfolio set, and  $\beta_p$  is restricted to be approximately 1 (i.e.  $0.995 < \beta_p < 1.005$ ).

Owing to the abnormal distribution of categories within the Taiwan stock market, the method of choosing stocks is based on the pool of 645 shares instead of the whole industry, as in Oh et al's model. We adjust this model and improve its tracking error and volatility of tracking error. The Procedure of Adjusted-GA portfolio model is shown in Figure 1 below:

### 4. Empirical studies

From January 2004 to December 2005, the Taiwan Stock Price Index (a major benchmark index in Taiwan Stock Exchange including 645 shares) was used to display the usefulness of the proposed GA portfolio scheme. We used an algorithm which optimizes the weights by

minimizing (2) over 5,000 random generations. In the process of GA, the crossover and mutation rates are changed to prevent the output from falling into local optima. The crossover rate runs from 0.5 to 0.8 and the mutation rate runs from 0.05 to 0.06, which uses 50 organisms in the population. The GA automatically stops when there is no improvement greater than 1% within the last 5000 trials (See Figure 1). In this section, we will compare the original model of Oh, et al. (2005) referred to as Model 1 and the Adjusted-GA portfolio model which we propose (Model 2).

For both models, we carefully examined the variables commonly involved in two algorithms which may influence their performances. First, the weights of three fundamental variables (standard error of beta, average trading amount, and average market capitalization) ( $v_1, v_2, v_3$ ) as shown in (1), are examined since they are crucial parameters influencing the algorithm performance. Second, the number of stocks in portfolio  $l$  ( $\leq n$ ) is investigated since a large  $l$  value usually implies decreased tracking error (TE). Third, starting points  $a_0$ s are tested to evaluate the performance and stability of both models.

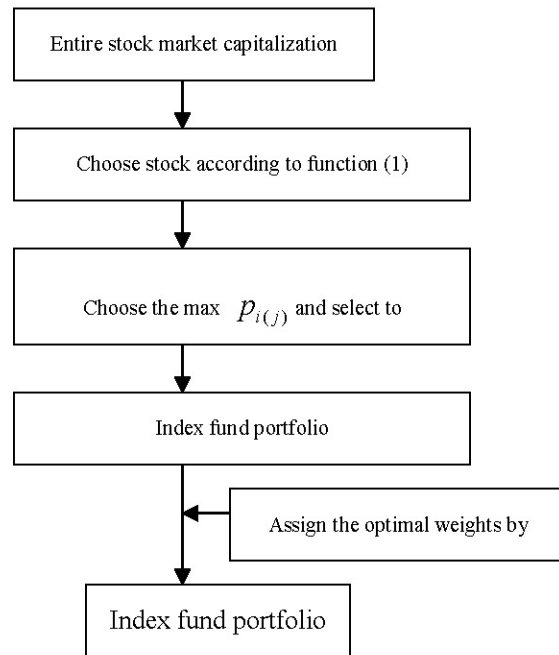


Fig.1: Procedure of Adjusted-GA portfolio model

#### 4.1. Experiments

In order to be compared Model 1,  $l$  is chosen to be not over 60, and  $T$  is set to be 60 days in our adjusted portfolio Model 2. We let  $a_0$  indicate the starting days of the algorithms which were randomly selected among the trading days of October 2004. After the training is done, the trained algorithm is immediately applied to the following trading days of the next

month for its performance evaluation.

For Experiment I, ( $v_1, v_2$  and  $v_3$ ) are weighted as shown in Table 1 and Table 2, with fixed values of  $l=30$ ,  $T=60$ , and  $a_0$  being October 1, 2004. Standard deviation of absolute values of TE and mean absolute deviation (MAD) of TE and movements of TE itself during the test period are provided in Table 1 and Table 2, respectively. It can be noticed that standard errors of Model 2 are uniformly less than Model 1 and mean absolute deviation of Model 2 are uniformly less than Model 1. In particular, Model 2 seems to be less sensitive than Model 1 to change in ( $v_1, v_2, v_3$ ).

Table 1 Standard deviation of absolute values of TEs starting October 1, 2004

$(v_1)$	$(v_2)$	$(v_3)$	Standard deviation of absolute values of TEs	
			Model 1	Model 2
1	1	1	0.004168	0.002997
1	1	2	0.003322	0.002090
1	2	1	0.004242	0.002447
2	1	1	0.006421	0.002997

Table 2 Mean absolute deviation (MAD) of TE

$(v_1)$	$(v_2)$	$(v_3)$	Mean absolute deviation (MAD) of TEs	
			Model 1	Model 2
1	1	1	0.00554	0.00388
1	1	2	0.004537	0.003152
1	2	1	0.004887	0.004009
2	1	1	0.007645	0.00388

(ie.  $v_1, v_2$  and  $v_3$  mean weight of standard deviation of beta, average trading amount and average market capitalization, respectively.)

Experiment II examines  $l$  with fixed  $a_0, T$  and ( $v_1, v_2, v_3$ ). Indeed, when  $a_0$  and ( $v_1, v_2, v_3$ ) are given by Oct 1, 2004,  $T=60$  days and (1,1,2) for various  $l$ , experiments are done which yields Table 3 and Table 4 during the test period. Again TEs of Model 2 are uniformly less than Model 1 and mean absolute deviations of Model 2 are uniformly less than Model 1. Through the experiment, it is easy to find that that Model 2 outperforms the Model 1.

Table 3 Standard deviation of absolute values of TEs starting Oct 1, 2004

The number of stock ( $l$ )	Standard deviation of absolute values of TEs	
	Model 1	Model 2
30	0.003322427	0.002090436
50	0.003718497	0.002751544
60	0.004871149	0.003412461

In Experiment III, we examine whether or not performance and stability are better than Model 1 in another the period, also using the

starting point  $a_0$  with fixed  $l=30$ ,  $T=60$ . The standard deviation of absolute values of TE and mean absolute deviation (MAD) of TE and movements of TE itself during the test period are provided in Table 5, Table 6 respectively. It can be noticed that standard errors of Model 2 are uniformly less than Model 1 and mean absolute deviation of Model 2 are uniformly less than Model 1.

Table 4 Mean absolute deviation (MAD) of TE

The number of stock ( $l$ )	Mean absolute deviation (MAD) of TEs	
	Model 1	Model 2
30	0.004537512	0.004009879
50	0.004619414	0.00311815
60	0.006264598	0.003869616

#### 4.2. Discussions

The experiment provides useful information for an efficient portfolio scheme. Indeed, the Model 2 scheme shows an improved performance over Model 1. Further, Model 2 is less sensitive to changes in ( $v1$ ,  $v2$ ,  $v3$ ). The results show that Model 2 scheme can improve performance over Model 1.

Table 5 Standard deviation of absolute values of TEs starting Oct 1, 2004

starting point ( $a_0$ )	Standard deviation of absolute values of TEs	
	model_1	model_2
2005/3/16	0.004411875	0.002422528
2005/1/3	0.00458062	0.001316013
2004/10/8	0.004318804	0.00248542
2004/07/06	0.004450691	0.003703511
2004/06/07	0.004346035	0.003260805
2004/02/09	0.004628042	0.003159309

Table 6 Mean absolute deviation (MAD) of TE

starting point ( $a_0$ )	Mean absolute deviation (MAD) of TEs	
	model_1	model_2
2005/3/16	0.00572544	0.003020211
2005/1/3	0.005425585	0.002133218
2004/10/8	0.004887495	0.00315221
2004/07/06	0.003625672	0.005928172
2004/06/07	0.006545482	0.003960065
2004/02/09	0.004930487	0.004580662

#### 5. Concluding remarks

The index fund is a kind type of passive investment strategy. Passive management is an investment strategy which presumes we cannot outperform the market. This research deduces the

model of Oh, et al. (2005) to propose an adjusted-GA portfolio model to support portfolio optimization process. Index TEs of the adjusted-GA portfolio scheme are examined through empirical experiments with Taiwan Stock Price Index under various settings. Our results strongly suggest that adjusted-GA portfolio model has outstanding advantages over the model of Oh, et al. (2005). Indeed, our adjusted-GA portfolio model delivers superior performance, along with other desirable properties.

#### References

- [1] Andrews, C., Ford, D. and Mallison, K., The Design of Index Fund and Alternative Methods of Replication, The Investment Analysis, 82, 123~16, 1986
- [2] Bogle, J. C. The implications of style analysis for mutual fund performance, Journal of Portfolio Management, 24(4), 34~42, 1998
- [3] Chang, K. P. Evaluating mutual fund performance: an application of minimum convex input requirement set approach, Computers and Operations Research, 31, 929-940, 2004
- [4] Elton, E., Gruber, M. and Blake, C, Survivorship bias and mutual fund performance, Review of Financial Studies, 9, 1097~1120, 1996.
- [5] Hogan, P. H. Portfolio theory creates new investment opportunities, Journal of Financial Planning, 7(1), 35~37, 1994
- [6] Markowitz, H., Portfolio selection, Journal of Finance, 7, 77~91, 1952
- [7] Markowitz, H., Portfolio selection: efficient diversification of investments, New York: Wiley, 1959
- [8] Sharpe, W. F, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance, 19, 425-442, 1964
- [9] Meade, N. & Salkin, G. R., Index Fund-Construction and Performance Measurement, Journal of Operational Research Society, Vol. 41, pp.871~879, 1989.
- [10] Oh, J. K., Kim, T. Y., & Min, S, Using genetic algorithm to support portfolio optimization for index fund management. Expert Systems with Applications, 28, 371~379, 2005
- [11] Roll, R, A mean/variance analysis of TE, Journal of Portfolio Management, 18, 13~22, 1992
- [12] Rudolf, M., Wolter, H. J., & Zimmermann, H., A linear model for tracking error minimization, Journal of Banking & Finance, 23, 85-103, 1999