QBS
Vol. 38, No. 2, pp. 87-92 (2019)
https://doi.org/10.22283/qbs.2019.38.2.87
pISSN 2288-1344 eISSN 2508-7185 **Original Article**

Quantitative Bio-Science
Official Journal of the Institute of Natural Science at Keimyung University
Available Online at https://qbs.kmu.ac.kr:442/

Segmentation Strategy for Asset Groups in the Korean Exchange-Traded Funds Market Using Genetic Algorithm

Han Jun Ko, Kyong Joo Oh*
Department of Industrial Engineering, Yonsei University, Seoul 03722, Korea
(Received May 24, 2019; Revised June 11, 2019; Accepted June 16, 2019)

ABSTRACT

ETFs (Exchange-Traded Funds) have about 400 stocks listed on the financial market in Korea. They are notable for their low transaction costs and high liquidity. However, studies only focus on the arbitrage strategy using the price gap between market price and net asset value of ETF, or the analysis between the ETF market and the underlying asset market. This study proposes a strategy for ETF portfolios based on underlying assets. The weight of each asset group is determined using a genetic algorithm with the Sharpe ratio of the portfolio. The performance of the portfolio is evaluated through the cumulative return and the aforementioned ratio during the test period; the experiment is conducted using the sliding window technique. In conclusion, we show that the cumulative return and Sharpe ratio of a portfolio can be improved through segmentation of the asset groups included therein.

Key words : ETF, Genetic algorithm, Segmentation strategy, Sharpe ratio, Sliding window

1. Introduction

As of 2018, the Korean ETF market totaled KRW 41 trillion, up 15.2 percent from the previous year. The number of newly listed stocks is 96, and the total number of listed stocks is 414. According to Gastineau, ETFs are designed to be linked to price movements or returns of index or asset and are tradable on the stock market [1]. There is lower management fee than general funds and no transaction tax. In addition, it can be said that ETF is an efficient investment product because it has some characteristics of stocks and can enjoy its own diversified investment effects.

Most of studies on ETF are about arbitrage strategy using the price gap between market price and net asset value of ETF, or the analysis between the ETF market and the underlying asset market. Lee and Hong [2] proposed an arbitrage trading strategy using the difference between the market price and the NAV of the ETF. Byun et al. [3] verified the relationship between the KOSPI 200 index and the ETF price and confirmed that there is information factor on the price of future ETFs in the price gap. Jares and Lavin [4] confirmed that there is arbitrage opportunity due to the difference in opening hours using the Japanese and Hong Kong ETFs listed on the US financial markets. Hur et al. [5] examined the price efficiency of the ETF market by analyzing the tracking errors and the gap between the Korean ETFs. Jung [6] analyzed the Korean domestic ETF market to determine the degree of investment efficiency. Seo and Hwang [7] tried to predict the market using a deep learning model.

There are also studies that examine profitability by considering ETF as investment products. Park [8] positively confirmed that ETF can be a competitive investment product compared to other financial products in Korea and examined how ETFs can practically be used by asset managers. Woo and Choe [9] test-

*Correspondence should be addressed to Kyong Joo Oh, Professor, Department of Industrial Engineering, Yonsei University, Seoul 03722, Korea, Tel: +82-2-2123-5720, Fax: +82-2-364-7807, E-mail: johanoh@yonsei.ac.kr
Han Jun Ko is a Master Candidate, Department of Industrial Engineering, Yonsei University, Seoul 03722, Korea.

ed profitability by applying day trading strategy to ETF market. Ryu et al. [10] selected and collected keywords representing trend of the sector through text mining and suggested a configuration strategy. Park et al. [11] presented a strategy to maximize the Sortino ratio, a top-down risk adjusted return, using genetic algorithms.

In this paper, we propose a strategy for constructing an ETF portfolio based on the underlying assets and using the genetic algorithm, which is one of the optimization algorithms, to determine the weight of each asset. The purpose of this study is to examine the test period suitable for the proposed model and to verify that the performance of the portfolio can be improved by segmenting the additional asset group based on the basic model.

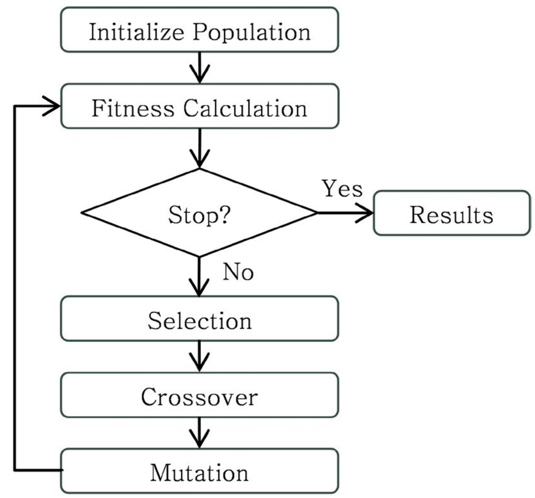

Fig. 1. Genetic algorithm.

2. Methods

2.1 Genetic algorithm

Genetic algorithm was developed by John Holland as an optimization algorithm based on the evolution of biological processes in 1975 [12]. It is based on Darwin's theory of survival of the fittest, and is widely used in natural science, engineering, humanities, and social sciences. In general, if a problem is overly complicated, genetic algorithms can be used to approximate the optimal solution, but not the actual optimal solution.

Genetic algorithm generally uses three kinds of operations: Selection, Crossover, Mutation. Selection is the process of choosing the solution from the previous generation to the next generation, including roulette wheel selection and rank-based selection, but the roulette wheel selection method is widely used [13]. In the roulette wheel selection method, the roulette plate is filled in proportion to the fitness of each chromosome and is then used to select the chromosomes probable next generation. It is a way to maintain the diversity of chromosomes transmitted to the next generation while setting the chromosomes with high fitness to be selected with higher probability.

Crossover is a process of generating new chromosomes from selected chromosomes, which includes simple mating, multipoint mating, and homogenous mating. In simple mating, there is only one crossing between chromosomes as one mating point. Multipoint mating is when there are two or more mating points. Homogenous mating is a method of interchanging genes only when the position is 0 by covering a mask of 0 and 1.

Mutation is a process in which a mutation occurs stochastically on a newly generated chromosome after crossing and is set to occur with a very low probability. Successful variation of a very low probability can rather contribute to the quality improvement of the group.

The process of the genetic algorithm is shown in Fig. 1. First, an initial population is generated, and the fitness of the solution is evaluated through the fitness function. Thereafter, the fitness is evaluated again after selection, crossover and mutation, and the process is repeated until the termination condition is satisfied. Genetic algorithms are solved by randomness, so they can be solved without knowing how to solve the problem, and there is an advantage that the algorithm is relatively simple since differential calculation is unnecessary.

2.2 Sharpe ratio

Sharpe ratio is a measure of the adjusted return on investment risk when assessing investment performance, as suggested by William Sharpe in 1994 [14]. Generally, it is calculated as excess return per unit of standard deviation of returns.

$$S_p = \frac{E[R_p - R_f]}{\sigma_p} \tag{1}$$

In equation (1), S_p means Sharpe ratio of portfolio p, σ_p means standard deviation of returns and R_p means returns. Also, R_f means risk-free rate and $E[R_p - R_f]$ means excess return of portfolio against risk-free rate.

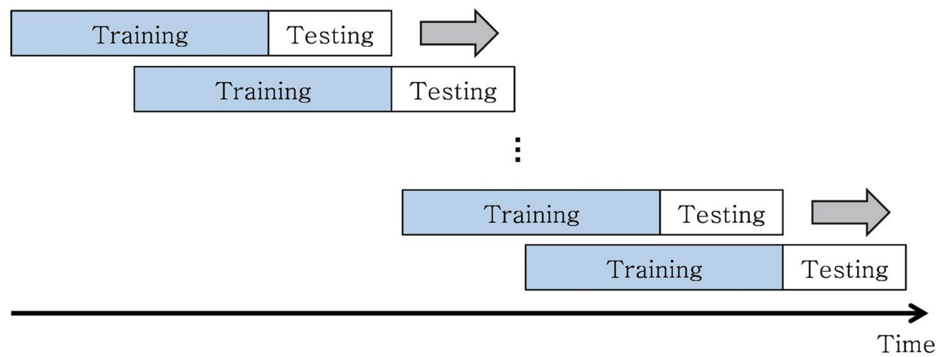

Fig. 2. Sliding window.

Sharpe ratio shows how well the returns are being compensated for the risk of an investor and is usually used to assess the performance of portfolios or funds. The higher Sharpe ratio means less risk for the same returns or higher expected return on the same risk.

2.3 Sliding window

The sliding window technique is one of the methods of dividing the training set and the test set using the time series data. As shown in Fig. 2, the first part of the time series is used as training set and test set, which is considered as a single window, and the back-testing is performed by moving the window for the test period. Many experiments can be conducted on relatively small amounts of data. It is possible to analyze the data of the immediate preceding period and to apply it considering the flow of time which is one of the most important characteristic of the time series data.

3. Results

3.1 Proposed model

In this study, we collect Korean ETF closing price data from January 2014 to September 2018 through xingAPI which is API (Application Programming Interface) provided by eBest Investment Securities. Also, we collect monthly CD rate data from January 2015 to September 2018 through CHECK Expert+provided by KOSCOM. When constructing portfolios on models, we use Sharpe ratio optimization method to determine the weight of each asset groups. The closing price data is converted into return data.

According to Cochran, stratified sampling is a method of dividing a population into multiple layers that are not duplicated and then extracting samples from each layer [15]. More accurate results can be obtained by dividing the population into homogeneous characteristics and heterogeneous characteristics among the layers. In this study, we classify the Korean ETFs into five basic groups and divide the heterogeneous groups into subgroups.

Table 1. Number of asset groups

Model	Number
1	5
2	7
3	8
4	10

There are four models to be covered in this study and the method of determining the weight of asset groups is all same. The method of constructing the asset group in each model is shown in Fig. 3, and the underlined item represents the additional subdivided asset group in the basic model 1.

Model 1 is a model in which the asset groups are divided into stock ETF, bond ETF, currency ETF, real estate ETF and commodities ETF. Model 2 is a subdivision of commodities ETF into oil ETF, metals ETF and agricultural commodities ETF. Model 3 is a model in which stock inverse ETF, bond inverse ETF and currency ETF are added to model 1. Model 4 is a combination of model 2 and model 3, which divides commodities ETF into three asset groups and the inverse asset groups. The number of asset groups constituting each model is shown in Table 1.

Among the ETFs that can be traded from January 2014, the representative ETFs in each asset group are selected as the ETFs with the highest average daily trading volume in the asset group. In other words, the number of items used to construct a portfolio in a model is equal to the number of asset

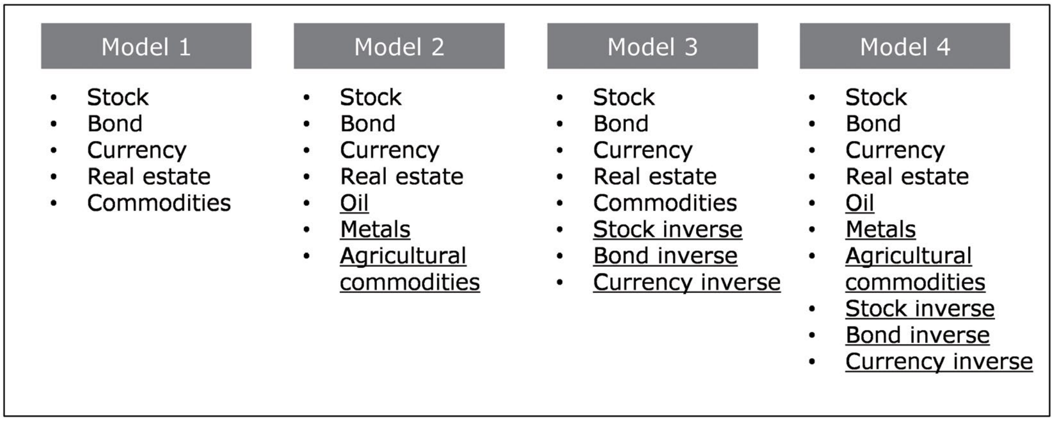

Fig. 3. Asset groups lists.

Table 2. Annual returns of models (test : 1 month)

Year	Model 1	Model 2	Model 3	Model 4
2015	2.21%	1.92%	0.88%	2.87%
2016	0.65%	−3.55%	−0.50%	−0.51%
2017	7.74%	9.59%	6.71%	9.75%
2018	2.44%	1.43%	0.87%	3.19%
Total	13.54%	9.27%	8.05%	15.91%

Table 3. Annual returns of models (test : 2 months)

Year	Model 1	Model 2	Model 3	Model 4
2015	−1.60%	−0.31%	0.70%	0.71%
2016	−0.21%	−2.78%	−0.01%	−1.56%
2017	6.23%	7.56%	5.87%	7.51%
2018	−2.78%	−2.95%	−0.90%	−2.21%
Total	1.41%	1.17%	5.64%	4.23%

Table 4. Annual returns of models (test : 3 months)

Year	Model 1	Model 2	Model 3	Model 4
2015	2.12%	2.84%	−0.55%	1.60%
2016	−0.97%	−4.24%	−4.81%	−0.50%
2017	6.06%	9.73%	7.24%	8.34%
2018	−0.05%	−1.92%	−0.51%	−0.61%
Total	7.20%	5.98%	1.01%	8.86%

Table 5. Annual Sharpe ratios of models (test : 1 month)

Year	Model 1	Model 2	Model 3	Model 4
2015	0.058	0.041	0.019	0.056
2016	0.012	−0.047	−0.005	−0.006
2017	0.090	0.108	0.088	0.143
2018	0.027	0.017	0.014	0.017
Avg.	0.047	0.030	0.029	0.052

classes in the model. For example, model 1 would constitute a portfolio of five ETFs, one for stock ETF, one for bond ETF, one for currency ETF, one for real estate ETF and one for commodity ETF.

The weight of each asset group is determined through genetic algorithm, and the objective function is set as the Sharpe ratio of the portfolio during the training period. In addition, the initial population is set to 500, the crossover rate is set at 0.5 and the mutation rate is set at 0.1. Since the portfolio may be concentrated in one item, the proportion of the asset group is limited to 30% and if it exceeds 30%, the excess is assumed to have cash. When there is a restriction on the weight, the subdivided asset group is considered being included in the existing asset group. For example, in model 2, if the sum of the oil group, the metals group and the agricultural commodities group exceeds 30%, the excess is regarded as cash reserves. Also, in model 3, if the sum of the stock group and the stock inverse group exceeds 30%, the excess is regarded as cash reserves.

3.2 Empirical analysis

Experiments on each model were conducted through a sliding window method. In the training period, the optimum weight of each asset group is determined through the genetic algorithm with the Sharpe ratio of the portfolio according to the weight as an objective function. In the test period, we apply the determined weight to check the return and Sharpe ratio of the portfolio. The training period is fixed at 1 year, and the test period is 1 month, 2 months and 3 months. The return of the representative ETFs is calculated using equation (2) based on the closing price.

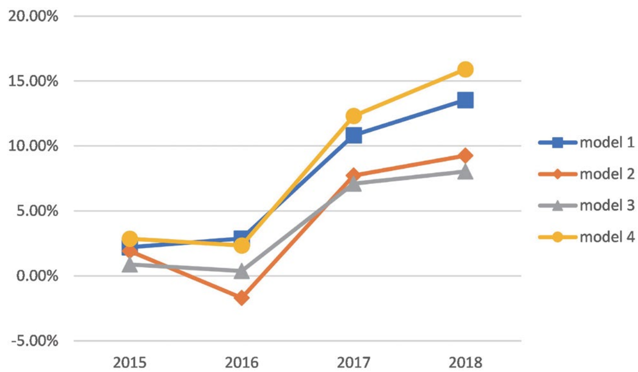

Fig. 4. Cumulative return (test : 1 month).

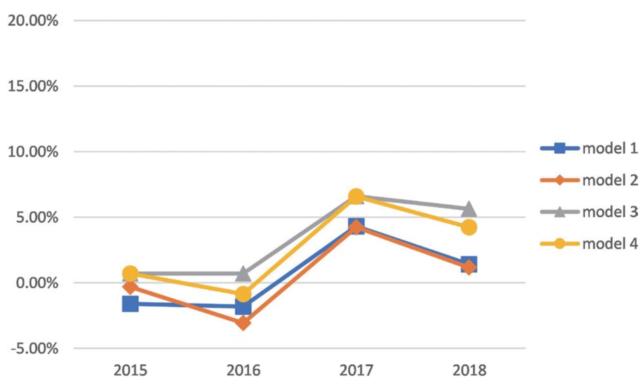

Fig. 5. Cumulative return (test : 2 months).

$$r_n = \frac{(C_n - C_{n-1})}{C_{n-1}} \times 100 \qquad (2)$$

In equation (2), C_n is the closing price of day n, and r_n is the return of day n. The interest rate on cash holdings is assumed to be the average of the monthly CD rate from January 2015 to September 2018, the entire test period. The average CD rate during the test period is 0.13% per month, which is multiplied by 3 for convenience and we use 0.4% for the quarter. The annual and cumulative returns for each model are shown in Table 2 to 4, and the cumulative returns are shown Fig. 4 to 6 and the annual Sharpe ratio are shown in Table 5 to 7.

Table 2 to 4 show the annual returns and the cumulative returns over the entire test period of 1 month, 2 months and 3 months. In the case of different test periods for each model, the cumulative return of the four models is highest when the test period is one month.

If we look at the same test period by model, the cumulative return of model 4 is the highest, except when the test period is two months it is the second highest.

In other words, when the model is 4 and the test period is 1 month, the performance of model 4 is the best in terms of the cumulative return of the portfolio. Fig. 4 to 6 show the cumulative return over time by model and test period and show the same result as the analysis of Table 2 to 4.

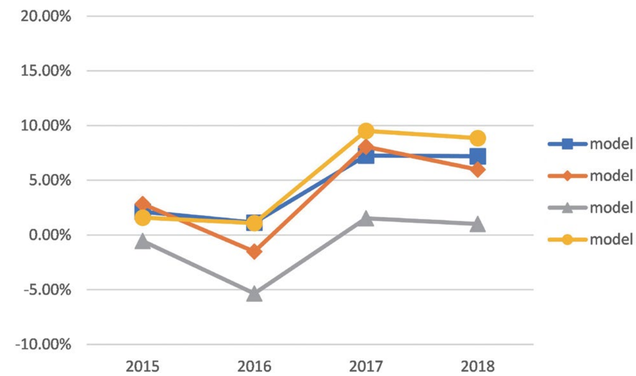

Fig. 6. Cumulative return (test : 3 months).

Table 6. Annual Sharpe ratios of models (test : 2 months)

Year	Model 1	Model 2	Model 3	Model 4
2015	0.037	0.028	0.015	0.048
2016	0.001	−0.051	−0.008	−0.031
2017	0.083	0.103	0.085	0.142
2018	0.020	0.012	0.013	0.016
Avg.	0.035	0.023	0.026	0.044

Table 7. Annual Sharpe ratios of models (test : 3 months)

Year	Model 1	Model 2	Model 3	Model 4
2015	0.050	0.059	−0.013	0.030
2016	−0.018	−0.049	−0.072	−0.005
2017	0.107	0.123	0.097	0.118
2018	0.002	−0.016	−0.008	−0.008
Avg.	0.035	0.029	0.001	0.034

Table 4 to 6 show the average Sharpe ratio for each model over the 1, 2 and 3 months of the test period and the average Sharpe ratio over whole period. In terms of the test period, one month shows the highest average Sharpe ratio for all models, and model 4 shows the highest average Sharpe ratio in comparison to other models. Overall, model 4 is the best performer in terms of both cumulative returns and Sharpe ratio, and the best test period is 1 month.

4. Discussion

This study proposes a strategy for constructing an ETF portfolio constituting an asset group based on underlying assets and tries to improve the investment performance by segmenting additional asset groups. The results show that the test period, that is, 1-month period of the portfolio rebalancing cycle, had the best performance in terms of cumulative return and Sharpe ratio. This result can be interpreted that 1 month is appropriate for each period in which the flow of the underlying asset market changes. And accordingly, a rebalancing cycle of 1 month is appropriate. Also, among the four proposed models, model 4 showed the best performance. This implies that the oil, metals, and agricultural commodities markets are unlikely to be grouped together, and that the addition of inverse asset groups means that even if each asset group falls, the portfolio can still perform well.

In addition, this study constructed portfolios by selecting the ETFs with the highest daily trading volume in each asset group. However, it can be difficult to say that the representative ETFs selected based on the daily average trading volume represent the asset group perfectly. In the future study, it is expected that if more stocks are selected according to the correlation coefficient with the return of the representative ETFs in the asset group, and the number of portfolio constituent ETFs is increased, then the portfolio will perform better.

References

1. Gastineau GL. Exchange-traded funds: An introduction. J Portfolio Manage 2001;27:88-96.
2. Lee JH, Hong JP. Arbitrage in the Korean ETF markets: ETF versus NAV. Kor J Financ St 2004;33:49-93.
3. Byun JC, Jo JI, Lee JW. The information effect of ETF mispricing. Kor Bus Educ Rev 2006;43:119-134.
4. Jares TE, Lavin AM. Japan and Hong Kong exchange-traded funds (ETFs): Discounts, returns, and trading strategies. J Financ Serv Res 2004;25:57-69.
5. Hur CS, Kang HC, Eom KS. Price efficiency of exchange-traded funds in Korea. J Money Fina 2012;26:39-73.
6. Jung HS. A study on the investment efficiency of Korean ETFs. J Digit Conv 2018;16:185-197.
7. Seo YB, Hwang CH. Predicting bitcoin market trend with deep learning models. QBS 2018;37:65-71.
8. Park KS. A study on the utilization of ETFs for financial planners. Financ Plann Rev 2011;4:137-157.
9. Woo MC, Choe H. Analysis of day trading strategy on the ETF market. Kor J Financ St 2012;41:677-704.
10. Ryu JP, Hahn CH, Shin HJ. Sector investment strategies using big data trends. J Inf Tech Archit 2016;13:111-121.
11. Park CH, Lee HJ, Oh KJ. An investment strategy using genetic algorithm to maximize downside risk-adjusted returns for long-term investors. QBS 2017;36:111-117.
12. Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. 1st ed. Cambridge: MIT Press; 1992.
13. Whitley D. A genetic algorithm tutorial. Stat Comput 1994; 4:65-85.
14. Sharpe WF. The Sharpe ratio. J Portfolio Manage 1994;21:49-58.
15. Cochran WG. Sampling techniques. 3rd ed. New York: John Wiley & Sons; 2007.