Received August 24, 2017, accepted September 21, 2017, date of publication September 26, 2017,
date of current version November 7, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2756842

Portfolio Optimization Based on Funds Standardization and Genetic Algorithm

YAO-HSIN CHOU, SHU-YU KUO, AND YI-TZU LO
National Chi Nan University, Puli 54561, Taiwan

Corresponding author: Yao-Hsin Chou (yhchou@ncnu.edu.tw)

ABSTRACT When investing in the stock market, the first problem and one of paramount importance which investors have to face is making the proper stock selection. Selecting the stocks that simultaneously offer high return and low risk is a difficult problem that is worth investigating. However, the traditional risk calculation based on the modern portfolio theory (MPT) of portfolios has some defects. The MPT method requires the calculations of every relationship between each pair of stocks in the portfolio, entailing high computation complexity, which grows exponentially with the increased number of stocks. Besides, the traditional calculation is unable to calculate the coefficient of variation, and merely considers the relationship between each pair of stocks, so it cannot accurately assess portfolio risk. Therefore, this paper proposes a novel method, funds standardization, and utilizes it to represent the portfolio return and calculate portfolio risk. The fluctuation of portfolio funds standardization shows not only the relationships between each pair of stocks, but also the interactions among all stocks. Hence, utilizing funds standardization can accurately assess portfolio risk and completely represent the mood swings of investors. Compared with the traditional method, the proposed method significantly reduces the computation complexity because the complexity does not increase when the portfolios stock number increases. We combine the genetic algorithm, Sharpe ratio and funds standardization to find the optimal portfolio. In addition, we utilize the sliding window to avoid the over-fitting problem, which is common in this field, and test the effect of all kinds of training and testing periods. The experimental results show that the portfolio can spread the risk effectively, and that the portfolio risk can be assessed accurately by utilizing the funds standardization. Comparing with the traditional method, our method can identify the optimal portfolio efficiently and establish a portfolio that has lower risk and stable return.

INDEX TERMS Portfolio, Sharpe ratio, stock selection, funds standardization, low volatility, Genetic algorithm (GA), modern portfolio theory.

I. INTRODUCTION

To make a fortune, most people choose investing. However, how to accurately utilize ones capital to invest in a beneficial target is the first priority for investors. In contrast to other targets, like futures, deposits and real estate, stock investment is more flexible and stable. Unlike real estate and deposits, stocks not only do not require plenty of money, but also permit a free allocation of capital. Moreover, information on the stock market is public. In addition, the fluctuation of stocks is smaller than that of futures; thus, the stock value will not drastically depreciate overnight. Also, some stock markets have a daily price limit, which makes it more stable than other targets. For these reasons, many investors choose to invest in stocks.

Stock selection is the first problem that investors encounter when they invest in the stock market. The Sharpe ratio [1] is an important indicator by which to assess a stock; its concept is to calculate the reward of each stock per unit risk. Investors always expect to choose a target that has less risk for the same reward, or a target that has more reward for the same risk. The higher the Sharpe ratio, the better the target! Therefore, investors like to invest in stocks which have a higher Sharpe ratio. To spread the risk, investors usually invest in more than one stock. As the saying goes, Do not put all your eggs in one basket. The economist Markowitz proposed the Modern Portfolio Theory (MPT) [2] that calculates total portfolio risk by variance and covariance; however, the computation complexity of MPT increases exponentially.

Furthermore, as there are a great number of targets in the market, taking all of the stocks into consideration to create the best portfolio is complicated.

With thousands of stocks in the market, deciding which portfolio should be chosen is quite difficult. The problem of stock selection cannot be solved by an exhaustive method. Computational intelligence (CI) techniques are usually used to solve problems which cannot be exhausted in a limited time. CI techniques have the ability to find a good solution within a large solution space. In recent years, CI has been widely used in the financial field, especially concerning the stock market. Therefore, choosing a portfolio by CI techniques can increase the efficiency of finding a good portfolio.

Most studies on stock selection use CI techniques; these include evolutionary computation, fuzzy theory and artificial neural network. CI techniques choose stocks according to some assessment indicators; the most commonly used indicators are risk, return and the Sharpe ratio. Calculating the Sharpe ratio requires assessing risk. Most of the methods utilize MPT to calculate risk. They all need to calculate the covariance of each pair of variables. As the number of stocks in a portfolio increases, the time for calculating covariance increases exponentially. Some papers propose a simplified way to reduce the risk calculation, but they cannot completely represent the total risk.

Consequently, this paper proposes funds standardization to calculate portfolio risk. In contrast to the traditional calculating method of MPT, ours posits that all the information will reflect the fluctuation of funds [3]. Therefore, it can calculate the risk of the portfolio more simply and also express the mood swings of investors. Then we use the Sharpe ratio where the portfolio assessment method of risk and return is replaced by our method, funds standardization. However, there are thousands of stocks in the market, so we cannot exhaust all the combinations in a limited time. Finding a better portfolio requires powerful computational intelligence. In our method, we use the genetic algorithm (GA) [4] to accurately construct the portfolio which has less risk and more reward. Our method, using GA for a large quantity of stocks, not only reduces the computation complexity to enable investors to easily evaluate the portfolio risk, but also creates a portfolio which has low risk and stable reward.

The rest of this paper is organized as follows. In Section II, other methods of using the Sharpe ratio on choosing a portfolio are discussed. In Section III,we give brief backgrounds of the Sharpe ratio, portfolio and the method Markowitz proposed in 1952 for calculating risk in MPT. Section IV is the basic concept of our method. In Section V, we explain what our method does and how we improved on this issue. The data on the testing effect and the results of comparing other methods with our proposed method are offered in Section VI. Lastly, Section VII presents the conclusion of this paper.

II. RELATED STUDIES

In the investment of stocks, there are three major issues: forecast, market timing and stock selection. Forecasting the stock price or the stock trend, finding the best timing to buy or sell stocks, or selecting good stocks can help investors obtain more profit. Because the feasible solution spaces of these issues are too complex to find the best solution, numerous applications of evolutionary computation are utilized in regard to these issues.

Related studies [5]–[17] on forecasting systems include neural networks [5]–[9], fuzzy time series [10]–[14], dynamic normalization back propagation networks [15], genetic fuzzy systems and artificial neural networks [16], and genetic algorithms and wavelet neural networks [17]. Besides the forecast, many related studies focus on finding the best timing to buy or sell stocks. These studies use different methods, such as fuzzy theory [18], neural network [19], genetic programming [20], ant colony optimization [21], particle swarm optimization [22] and quantum-inspired algorithms [23], [24]. Some of them not only use evolutionary computation, but also apply technical indicators to find the best timing.

In addition to forecasting and market timing, stock selection is a critical issue in investment. Stock selection is the first problem faced before investors can make a sensible investment decision. Selecting the stocks while considering both low risk and high return is difficult; as a result, MPT [2] and the Sharpe ratio [1] were proposed. MPT proposed the method for calculating portfolio risk and portfolio return. The core idea of the Sharpe ratio is maximizing the expected portfolio return per unit of portfolio risk. The Sharpe ratio simultaneously considers the return and the risk. Many studies [25]–[33] use MPT and the Sharpe ratio to assess portfolios.

However, the Sharpe ratio still involves some problems. For instance, sine the calculation of portfolio risk is too complex to comput, some studies [25], [26] focus on simplifying the formula of portfolio risk or reducing the calculation time. Reference [25] exploited particle swarm optimization to find the best portfolio in reasonable computing time. Reference [26] indicated that the calculation time of MPT will increase tremendously with a large number of stocks, so that study proposed a faster heuristic algorithm to reduce the calculation time. Besides, studies [27]–[31] proposed considering not only the Sharpe ratio, but also other factors when assessing an investment portfolio because they think that the Sharpe ratio is not accurate enough. Reference [27] exploited the Sharpe ratio and downside risk to assess a portfolio. Reference [28] utilized a regularized hypervolume selection algorithm and robust statistics to find the best portfolio. Reference [29] incorporated basic, bounding, cardinality and class constraints to assess a portfolio. Reference [30] proposed a fuzzy Sharpe ratio and used the uncertainty of portfolio fuzzy return to calculate the risk. Reference [31] used multi objective particle swarm optimization whose fitness considers the Sharpe ratio and percent return. Some studies [32], [33] do not search for a portfolio, but rather compare different models or algorithms and use the Sharpe ratio as the evaluative criterion. Besides the Sharpe ratio, related studies use methods which include the adjusting of weights, different index or fuzzy theory to improve the accuracy of portfolio assessment.

Besides, many studies [34]–[42] also use the definition of portfolio risk in the MPT. Reference [34] demonstrated the practicality of large-scale numerical portfolio optimization using the covariance matrix estimation process. Reference [35] proved that stocks with low volatility can earn high risk-adjusted return. Reference [37] found that the laws of traditional finance theory are defied, since low-volatility portfolios outperform high-volatility portfolio. Financial theory predicts higher return associated with higher risk. If portfolios have higher return, it is due to higher portfolio risk. However, these studies [34], [35], [37] indicate that actively managed low-volatility portfolios can outperform high-volatility portfolios.

Reference [36] proposed a practical approach to portfolio selection which explicitly considers the conditionally varying volatility and the fat-tailedness of risk factors. Reference [38] discussed the reason why elegant mathematics can lead to disastrous policies and the importance of thinking about the portfolio as a whole. Reference [39] recapitulated the views of Markowitz on the foundations of portfolio theory and hypotheses about actual financial behavior. Reference [40] discussed the different ways of risk measurement. Reference [41] considered the context of rational choice, and applied the mean-deviation analysis in optimal risk sharing. Reference [42] found that rules based on mean-variance have superior performance improvements by using the copula-based model. All of these studies still used the basic concept of MPT: portfolio risk calculated by variance and covariance. Some of these papers claim that the performance of MPT is not very good, so they want to improve it. They use different funds allocation models to enhance portfolio performance. However, these related studies still use variance and covariance to calculate risk, although covariance is unable to consider the interaction of more than two stocks. In other words, these studies want to use complicated calculations or other models to improve portfolio risk assessment, but face risks in still using variance and covariance. Hence, they do not effectively ameliorate the problem. Risk calculation is still complex and does not consider the interactions of more than two stocks.

The methods which are used to find the lower risk portfolio roughly fall into two categories: portfolios composed of stocks that have the best performance, and portfolios that use variance and covariance to calculate the entire portfolio risk while trying to reduce the calculation complexity. This paper proves that the best portfolio does not include the best stock each time, so the first category has the probability of excluding the best portfolio. In the second category, they consider the whole portfolio risk, but use variance and covariance to calculate the portfolio risk; this not only makes risk calculation more complex, but also ignores the interactions of more than two stocks in a portfolio. However, by using funds standardization to calculate the portfolio risk, the risk calculation becomes simple, and all the interactions in a portfolio are considered. Therefore, this paper does not require deleting any probability to simplify the calculation of portfolio risk. The complexity of stock selection still remains $O(2^n)$ if there are n stocks. Without limiting the selection of stock, this paper uses the evolutionary algorithm to help create the best portfolio in the large solution space in a short time.

In the stock selection problem, a fair comparison of different portfolios is an important criterion. Many studies use the ratio of return to calculate the risk because they think that the calculation method can compare different portfolios with the same criteria. However, the traditional risk calculations entail problems. The traditional risk calculations calculate and average the daily return rate. They use the result to calculate the risk and compare the portfolios with the same criteria. For example, there is a stock in the portfolio, whose price is shown in Table 1. The traditional methods use the stock price to calculate the daily return rate. They then calculate the average return rate which is the average daily return rate, as shown in Table 2. In general, investors believe that the return rate is the stock price of the last day minus the stock price of the first day and divided by the stock price of the first day, also shown in Table 2.

TABLE 1. The example of daily return rate in the traditional method.

	Day 1	Day 2	Day 3
Stock price	100	50	85
Daily return rate	-	-50%	70%

TABLE 2. The calculations of average return rate and return rate.

	Process of calculation
Average return rate	$[(-50) + 70] \div 2 = 10\%$
Return rate	$(85 - 100) \div 100 = -15\%$

We can find that the average return rate is not the same as the return rate. Even when the average return rate is positive, the return rate may be negative. This easily confuses investors and impacts their judgment. In order to solve the problems of calculation and comparison, we propose a novel method, funds standardization, to accurately assess the portfolio and compare the portfolios under the same criteria. Differing from the traditional portfolio risk calculation, which calculates the covariance between pairs of stocks, funds standardization contains all the interactions among the stocks in the portfolio. Hence, funds standardization is able to completely represent the fluctuations of a portfolio and the mood swings of investors. Besides, the risk calculation becomes simple; in contrast with the traditional risk calculation, we only need to calculate the standard deviation of the portfolio funds standardization to derive the portfolio risk. Moreover, this paper uses a simple method, coefficient of variations (CV), to fairly compare different portfolios in regard to risk calculation.

Unlike the traditional methods, funds standardization does not confuse investors. It is able to truly reflect the investment situation, so that investors are able to make the best decision by using our method. Some studies also use CV to compare portfolios regarding the risk calculation, but they still use average return rate. Therefore, their methods also have the problem of average return rate.

In the MPT, the portfolio risk is calculated with the variance and covariance. However, the covariance makes the calculation become complex, and it is not unable to consider the interactions of more than two stocks in a portfolio. Because the complexity in calcuating portfolio risk is high by using the MPT, the related studies have to use some mothods which fixes the best stock in a portfolio or simplified calculations of pairwise covariance to reduce the calculation. However, the problem of ignoring the interactions over for more than two stocks in a portfolio is still not soloved, and this paper proves that simplified ways make cause the best portfolio to be removed in some cases. This paper uses funds standardization to solve the problems of MPT and improves the risk assessment by considering all the relationships in the a portfolio, while reducing the risk calculation computation complexity to O(1) and comparing different portfolios in with the same criteria. Hence, it has the ability to assess portfolio risk with consideration of all the interactions in a portfolio without complicated calculations. After the accurate assessment of portfolio risk, this paper is able to select the proper portfolio which has high Sharpe ratio, via an evolutionary algorithm.

III. BACKGROUND

A. PORTFOLIO

Stock selection is a critically important issue when it comes to investing in the stock market, and the first problem that investors encounter. When investors select stocks, they consider the risk and the return of the stock. However, a single stock with high return usually entails high risk, while a single stock with low risk entails low return. Investors naturally want to invest in stocks with low risk and high return. The fluctuation of stock price can represent the risk of this stock. As the proverb says, ''Do not put all your eggs in one basket.'' In order to stabilize the fluctuation of stocks, we aggregate different stocks in a portfolio because a good combination of stocks is able to reduce the holistic risk and increase the return. Therefore, this paper provides a portfolio which simultaneously considers risk and return to investors.

B. MODERN PORTFOLIO THEORY [2]

Modern Portfolio Theory (MPT), which Markowitz proposed in 1952, assumes that investors are risk averse. It means that given two portfolios with the same return, investors will choose the less risky portfolio. On the other hand, investors will select the higher return portfolio when the portfolios have same risk. Generally, investors want to invest in a portfolio which can balance return and risk. By selecting a proper combination of assets to invest in, investors can acquire higher return with the same amount of the risk. In the mean-variance model, investors need to distribute the funds for every stock when they choose N stocks for investment. The weights of the stocks are $w_1, w_2, \cdots, w_N$, respectively, and the sum of the weights is 1, such as in formula (1). Because the portfolio is the proportion-weighted combination of stocks, the portfolio expected return is the proportion-weighted combination of stock return. Investors can use the weights and the expected return of stocks to calculate the portfolio expected return on investment (ROI) by formula (2). $E(r_p)$ is the expected return of portfolio and r_i is the expected return of stock i, σ_p is the portfolio risk. Besides, the risk of stocks is defined as variance in the MPT, and the covariance signifies the interaction between two stocks. The definition of portfolio risk is given by formula (3), where σ_p is the portfolio risk, and $\sigma_{ij}(i \neq j)$ is the covariance of stocks i and j.

$$\sum_{i=1}^{N} w_i = 1 \textit{ and } 0 < w_i < 1 \tag{1}$$

$$E(r_p) = \sum_{i=1}^{N} w_i r_i \tag{2}$$

$$\sigma_p = \sum_{i=1}^{N} \sum_{i=j}^{N} w_i w_j \sigma_{ij} \tag{3}$$

C. SHARPE RATIO [1]

The Sharpe ratio was proposed by William Sharpe, a winner of the Nobel Memorial Prize in Economic Sciences in 1990. The Sharpe ratio is based on the capital asset pricing model, which is one of the basic and significant investment theories. The Sharpe ratio is utilized to measure the performance of a stock; it is a representative index to assess stocks. Choosing and holding the portfolio which has low risk and high return is the purpose of the Sharpe ratio. The formula of the Sharpe ratio is given by formula (4)where R_f is the risk-free rate of interest. By selecting the higher Sharpe ratio, investors invest in the portfolio which has minimal investment risk with the same amount of return or maximal return with the same investment risk:

$$\textit{Sharpe ratio} = [E(r_p) - R_f]/\sigma_p \tag{4}$$

The Sharpe ratio and MPT are used extensively by investors to select their portfolios, but they possess a lot of defects. For example, the portfolio risk calculation uses covariance to represent the interaction of stocks in a portfolio. When the portfolio includes more than two stocks, there is no well-defined calculation of the portfolio risk. However, covariance only reflects the relationship between two stocks. All the interactions of the stocks in a portfolio need to be considered in regard to portfolio risk, so the traditional calculation cannot assess portfolio risk accurately. Even though the traditional calculation is able to represent portfolio risk accurately, it becomes complicated with the increased number of stocks because the traditional calculation needs to calculate the covariance of each pair of stocks in the portfolio. Besides the accuracy and the complicated calculation, determining how to compare different portfolios with fair criteria is an important problem. Many studies use the rate of return to compare portfolios in regard to risk calculation but the average of daily return rate is inconsistent with the whole return rate in some cases. All of these defects will result in an inaccurate portfolio risk evaluation. Therefore, we propose a new method, funds standardization, to solve the above-mentioned problem. Funds standardization makes a more accurate evaluation, reduces the calculating time no matter how many stocks are chosen, and compares different portfolios fairly.

IV. BASIC CONCEPT

Stock selection is based on the return and the risk of the stock. High return always comes with high risk, and low risk will inevitably have low return. When people invest in a stock, they expect to obtain the return while bearing less risk. As a result, determining how to select stocks with low risk and high return is an important issue. The Sharpe ratio is the common index and has been widely utilized in stock selection. The core concept of the Sharpe ratio is selecting a high return portfolio with the same risk, or selecting a low risk portfolio with the same return; it is able to simultaneously balance risk and return.

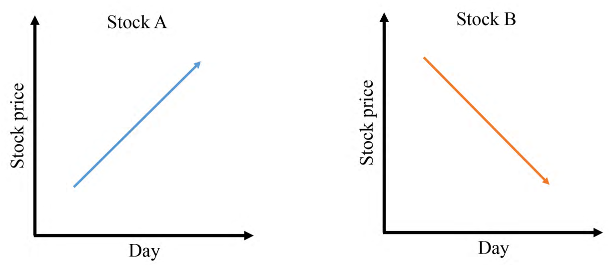

FIGURE 1. Stock A is the upside trend and stock B is the downside trend.

When calculating the portfolio risk, most people think that portfolio risk is the summation of every stock risk; this is wrong. Assume that there are two stocks in the portfolio, and they are following contrary trends, as shown in Fig. 1. Both stocks have high risk. If the portfolio risk is such that the risk of stock A directly adds to the risk of stock B, the portfolio risk will be huge. However, the fact is that two stocks on contrary trends can reduce portfolio risk. Therefore, using the summation of stock risk as the portfolio risk is incorrect. In the MPT, the portfolio risk includes not only the risk of each stock, but also the interaction of each pair of stocks in the portfolio. By adding the covariance, which is the interaction of each pair of stocks in the portfolio, the portfolio risk will be close to the real investment risk, as shown in Fig. 2.

Since the MPT was proposed, most studies now use the MPT risk calculation method. However, in the MPT, the portfolio risk calculation ignores some relationships in the portfolio. When there are more than two stocks in a portfolio,

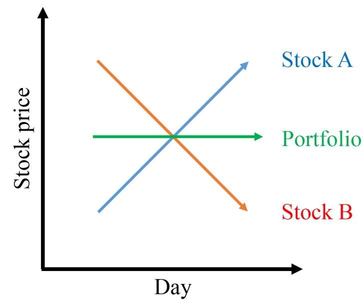

FIGURE 2. An example which only includes two stocks in the portfolio; the portfolio considers the trends of stock A and stock B.

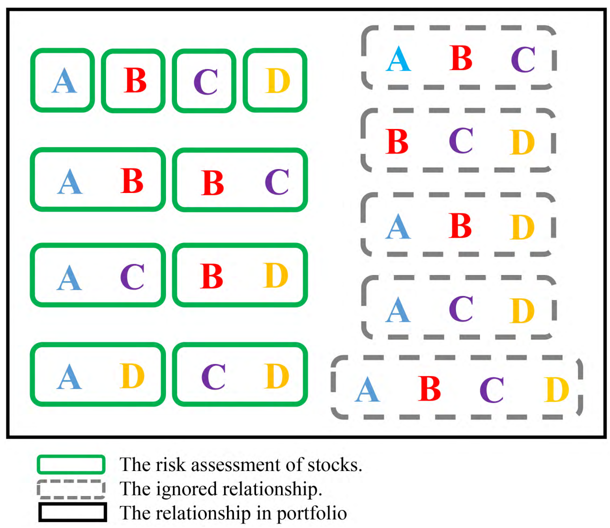

FIGURE 3. The relationships in a portfolio with four stocks. The risk calculation of MPT ignores the relationships in the gray dotted frame, which are the interactions among stocks A, B and C, stocks A, B and D, stocks A, C and D, stocks B, C and D, and stocks A, B, C and D.

the portfolio risk still only considers the variance of each stock and covariance of each pair of stocks. Because the limit of covariance can only represent the relationship of two variables, the calculation is unable to consider the full set of interactions in a portfolio which contains more than two stocks. For example, there are stocks A, B, C and D in a portfolio, but the portfolio risk only considers the interaction between stocks A and B, stocks A and C, stocks A and D, stocks B and C, stocks B and D, and stocks C and D, as shown in Equation 5. The interactions among stocks A, B and C, stocks A, B and D, stocks A, C and D, stocks B, C and D, and stocks A, B, C and D are ignored in MPT, as shown in Fig. 3. Though the computational complexity is diminished from $O(2^n)$ to $O(n^2)$ in this pattern, the risk assessment is inaccurate. Besides, the calculation is still complicated when the portfolio contains more than two stocks. To reduce the calculation complexity, many methods have been proposed. [43] take two different ways for example. One way is selecting a stock whose Sharpe ratio is highest, and then averaging the covariance between this stock and others in the portfolio. The other way is selecting a stock, whose Sharpe ratio is highest, averaging the return rate of other stocks in the portfolio and then calculating the covariance between the highest one and the average result. Although these methods simplify the calculation, they ignore many interactions in the portfolio and even make the risk assessment of portfolio incomplete.

$$\begin{aligned}&\textit{Portfolio risk in MPT}\\&= w^2(\sigma_A^2+\sigma_B^2+\sigma_C^2+\sigma_D^2\\&\quad +2\sigma_{AB}+2\sigma_{AC}+2\sigma_{AD}+2\sigma_{BC}+2\sigma_{BD}+2\sigma_{CD})\end{aligned} \tag{5}$$

Unlike most studies that only focus on reducing the computational complexity, we find that the risk calculation in MPT entails some defects such as high computational complexity and ignoring some relationships among stocks, and therefore adopt a novel method, funds standardization. It not only considers the overall interactions in the portfolio but also reduces the computational complexity of risk from $O(n^2)$ to $O(1)$. It means in our method when calculating the risk of the portfolio, we only need to calculate one standard deviation of one set of data, but traditional studies need to calculate n^2 standard deviations in n sets of data which n indicate the number of stocks in the portfolio.

To solve the problems of risk calculation in MPT, we utilize a novel method, funds standardization, to calculate portfolio risk. Fluctuations of each stock price in the portfolio convert into the fluctuation of funds, as shown in the Fig. 4. It can truly reflect the risk involved in a portfolio. In contrast with the MPT that uses summation of each stock risk and relationship between each pair, this method uses funds standardization, which directly represents the real situation in investing in the portfolio. Because funds standardization entails portfolio fluctuation, we can calculate portfolio risk with portfolio funds standardization. Differing from the risk calculation in MPT, which ignores numerous relationships of each stock in the portfolio when the portfolio contains more than two stocks, funds standardization considers all of the interactions in the portfolio. The relationships of all the stocks in the portfolio are truly shown via the fluctuation of funds standardization, so the standard deviation of funds standardization is able to accurately assess portfolio risk. All of the portfolio interactions are considered in funds standardization, and the fluctuations of portfolio funds standardization reflect the mood swings of investors.

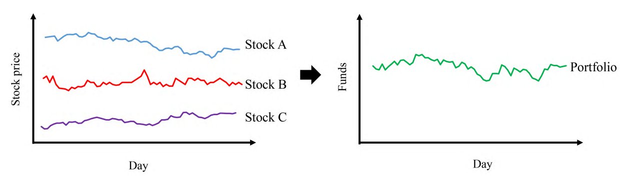

FIGURE 4. **The portfolio includes stocks A, B and C. Each stock price in the portfolio is converted into funds standardization.**

Besides the risk calculation of MPT ignoring the numerous relationships of each stock, the computational complexity of MPT grows exponentially with increased number of stocks. Funds standardization handles the interaction of every stock in the portfolio. Hence, the standard deviation of portfolio funds standardization can accurately represent the portfolio risk. Differing from the risk calculation in MPT, which needs massive calculations, we only need to calculate the standard deviation of portfolio funds standardization, which is far easier. While we only calculate the standard deviation of the portfolio funds standardization, we do not ignore any interactions in the portfolio. Portfolio funds standardization reduces the computational complexity to $O(1)$. In contrast to the traditional risk calculation, our method for calculating the standard deviation of portfolio funds standardization makes it easy to understand the reasons underlying the portfolio risk calculation.

In contrast with other studies which use the average return rate to calculate the risk, this paper uses funds standardization. Funds standardization can avoid the problem of the average return rate and the return rate possibly differing. Differing from traditional studies that individually consider stock risk and interaction between two stocks, portfolio funds standardization can directly reflect portfolio investment risk. Traditional methods use the stock information to calculate every stock risk and the risk of each pair in the portfolio, so they have many sequences to calculate. In general, CV is used in one sequence; therefore, the traditional methods need exceptional calculation to transfer into one sequence if they want to compare portfolio by using CV. Our method only has one sequence (portfolio funds standardization). CV is easy to apply in funds standardization as well as to compare portfolios with the same criterion.

The novel idea of funds standardization offers four important contributions. Firstly, funds standardization properly considers all of the interactions among stocks in a portfolio. Secondly, funds standardization reflects the true investment situation of a portfolio which represents the mood swings of investors. Thirdly, it substantially simplifies the calculation of portfolio risk. Finally, it utilizes the same criterion to compare the portfolios fairly. Funds standardization has the great ability to evaluate a portfolio and can help investors find low-volatility portfolios. In addition, our method does not restrict the number of stocks in a portfolio. If there are n stocks in the stock market, the computational complexity of selecting or not selecting each stock is $O(2^n)$. The problem of stock selection is too complex to perform exhaustive list of all the combinations. So this method uses the evolutionary algorithm to find a proper portfolio from n stocks in a short time. Note that the contribution of this study is proposed a novel method to evaluate a portfolio which significantly reduces the computational complexity and completely represent the interactions of all stocks, so using which optimization algorithms to search a good portfolio will not affect the contribution of fund standardization. In other words, we can use other global optimization techniques to solve stock selection problem, in this method GA is just an example. In future work, we can use faster algorithms such as particle swarm optimization, quantum-inspired tabu search algorithm [44], [45], and jaguar algorithm to search for the best solution.

V. PROPOSED METHODOLOGY

Our method combines the Sharpe ratio and GA to solve the problem of stock selection. We utilize stock price to calculate portfolio funds standardization, further improving the assessment of risk. Using funds standardization can truly reflect portfolio risk as well as consider all of the interactions among stocks in a portfolio. Because this paper does not restrict the stock number of a portfolio, there are massive combinations in the search space. GA is utilized to find the portfolio that has the low risk and high return in the huge search space. Besides, we use the sliding windows to find the best portfolio in the training period and for trading in the testing period, to avoid the over-fitting problem.

A. FUNDS STANDARDIZATION

Funds standardization uses stock price to calculate the funds of each stock in the portfolio. In our method, the initial funds are equally allocated to each stock in the portfolio, as shown in Equation 6, and the remainder of the portfolio, which is given by Equation 7. After the allocation of initial funds, we are able to calculate the affordable shares of each stock and the remainder of the stocks. In general, Taiwans stock market uses shares for calculation; as one share is 1000 lots, we do not consider odd lots. When buying stocks, investors need to pay the handling fee. Hence, the calculation of shares has to consider the handling fee, as shown in Equation 8. The fee rate in Taiwan market is 0.1425%. The remainder of each stock is the allocated funds after deducting the funds for buying the stocks and the handling fee. The calculation of the handling fee and remainder of each stock are shown in Equations 9 and 10, respectively.

$$\begin{aligned} &\mathit{allocated\ funds} \\ &\quad = \lfloor \mathit{initial\ funds} \div N \rfloor \end{aligned} \tag{6}$$

$$\begin{aligned} &\mathit{remainder\ of\ portfolio} \\ &\quad = \mathit{initial\ funds} - \mathit{allocated\ funds} \times N \end{aligned} \tag{7}$$

$$\begin{aligned} \mathit{share} = &\lfloor \mathit{allocated\ funds} \\ &\div (\mathit{stock\ price} \times 1000 + \mathit{stock\ price} \\ &\times \mathit{fee\ rate} \times 1000) \rfloor \end{aligned} \tag{8}$$

$$\begin{aligned} &\mathit{hanlding\ fee} \\ &\quad = \mathit{share} \times \mathit{stock\ price} \times \mathit{fee\ rate} \times 1000 \end{aligned} \tag{9}$$

$$\begin{aligned} &\mathit{remainder\ of\ stock} \\ &\quad = \lfloor \mathit{allocated\ funds} \\ &\qquad - \mathit{funds\ of\ buying\ stocks} - \mathit{handling\ fee} \rfloor \end{aligned} \tag{10}$$

$$\begin{aligned} &\mathit{funds\ standardization}(1) \\ &\quad = \mathit{allocated\ funds} - \mathit{handling\ fee} \end{aligned} \tag{11}$$

$$\mathit{return} = \mathit{share} \times \mathit{stock\ price} \times 1000 \tag{12}$$

$$\begin{aligned} &\mathit{securities\ transaction\ tax} \\ &\quad = \mathit{share} \times \mathit{stock\ price} \times 1000 \times \mathit{rate} \end{aligned} \tag{13}$$

TABLE 3. Stock prices.

	Stock A	Stock B
Day 1	80	49
Day 2	85	50
Day 3	75	51

$$\begin{aligned} &\mathit{funds\ standardization}(m) \\ &\quad = \mathit{return}_m - \mathit{fee}_m - \mathit{securities\ transaction\ tax}_m \\ &\qquad + \mathit{remainder\ of\ stock}_m \ \mathit{and}\ m > 1 \end{aligned} \tag{14}$$

$$\begin{aligned} &\mathit{portfolio\ funds\ standardization} \\ &\quad = \sum_{i=1}^{N} \mathit{funds\ standardization}_i \\ &\qquad + \mathit{remainder\ of\ portfolio} \end{aligned} \tag{15}$$

After the above calculation, we are able to calculate the daily funds standardization. The funds standardization on the first day is the allocated funds of each stock after deducting the handling fee for buying the stock. The calculation is shown in Equation 11. After the first day, we assume selling each stock in the portfolio to calculate funds standardization. Because selling stocks requires paying a handling fee and securities transaction tax, the funds standardization after selling the stocks subtracts the handling fee as well as the securities transaction tax, and then adds the remainder of stock. The calculation of the funds after selling the stocks, securities transaction tax and funds standardization after the first day, are shown in Equations 12, 13 and 14 respectively, where m is the m_{th} day and the rate of securities transaction tax is 0.3% in the Taiwan market. By using Equations 11-14, we are able to calculate the daily funds standardization of each stock in a portfolio. The portfolio funds standardization is the summation of the funds standardization for each stock in the portfolio with the addition of the remainder of the portfolio. The calculation is shown in Equation 15, where i is the i_{th} stock. In our example, there are two stocks in the portfolio; their stock prices are shown in Table 3. Then our method equally allocates the initial funds 4,000,000 to every stock. By using this information, we can calculate the portfolio funds standardization. The result and the process of calculation are shown in Table 4. The calculation of fund standardization is the basic addition, subtraction, multiplication, and division, and can be presented to investors every day.

B. PORTFOLIO OPTIMIZATION USING GA

GA is a kind of evolutionary algorithm; it is a bio-inspired algorithm. Its core concept is that better parents will breed superior offspring. Imitating the creature evolutionary process, GA generates a high-quality solution through selection, crossover and mutation procedures to solve optimization problems. To evaluate if the solution is good or bad, every solution has a fitness which determines whether the solution

TABLE 4. Portfolio funds standardization.

	Stock A	Stock B	Portfolio
Funds	2000000	2000000	4000000
Shares	24	40	64
Remainders	77264	37207	114471
Day 1	1997264	1997207	3994471
Day 2	2108237	2028357	4136594
Day 3	1869299	2068180	3937479

Shares of stock A : $2000000 \div (80 \times 1000 + 80 \times 1.425) = 24$ shares
Shares of stock B : $2000000 \div (49 \times 1000 + 49 \times 1.425) = 40$ shares
Remainder of stock A : $2000000 - 24 \times 80 \times 1000 - 24 \times 80 \times 1000 \times 0.01425 = 77264$ dollars
Remainder of stock B : $2000000 - 40 \times 49 \times 1000 - 40 \times 49 \times 1000 \times 0.01425 = 37207$ dollars
Funds standardization of stock A on the first day : $2000000 - 24 \times 80 \times 1.425 = 1997264$ dollars
Funds standardization of stock B on the first day : $2000000 - 40 \times 49 \times 1.425 = 1997207$ dollars
Funds standardization of stock A on the second day : $24 \times 85 \times 1000 - 24 \times 85 \times 1.425 - 24 \times 1000 \times 0.003 + 77264 = 2108237$ dollars
Funds standardization of stock B on the second day : $40 \times 50 \times 1000 - 40 \times 50 \times 1.425 - 40 \times 1000 \times 0.003 + 37207 = 2023857$ dollars
Funds standardization of stock A on the third day : $24 \times 75 \times 1000 - 24 \times 75 \times 1.425 - 24 \times 1000 \times 0.003 + 77264 = 1869299$ dollars
Funds standardization of stock B on the third day : $40 \times 51 \times 1000 - 40 \times 51 \times 1.425 - 40 \times 1000 \times 0.003 + 37207 = 2068180$ dollars
Portfolio funds standardization on the first day : $1997264 + 1997207 + 0 = 3994471$ dollars
Portfolio funds standardization on the second day : $24 \times 85 \times 1000 - 24 \times 85 \times 1.425 - 24 \times 1000 \times 0.003 + 77264 + 40 \times 51 \times 1000 - 40 \times 51 \times 1.425 - 40 \times 1000 \times 0.003 + 37207$ $= 2108237 + 2028357 + 0 = 4136594$ dollars
Portfolio funds standardization on the third day : $1869299 + +2068180 + 0 = 3937479$ dollars

fits that optimization problem. At every generation, GA produces a new population and goes through selection, crossover and mutation every time; it then derives a better solution in the end. This paper utilized GA as a tool to demonstrate whether the proposed method can produce a portfolio which has low risk (low-volatility) and stable return simultaneously.

TABLE 5. Chromosome representation.

Stock #	1101	1102	1216	1301	1303
Chromosome	0	0	1	0	1

1) REPRESENTATION

In the encoding with GA in our method, chromosome length represents the number of stocks which are included in Taiwan 50 ETF. Every chromosome in GA symbolizes a portfolio, and a bit represents a stock. We use a binary array (0s and 1s) to show which stock is selected or not in a portfolio. For example, in the chromosome in Table 5, it shows that the third and fifth bit are 1. It represent that this portfolio selects stocks 1216 and 1303, and then invests equal funds in these two stocks. Our method does not limit how many stocks should be chosen; one or zero stocks may be chosen in a

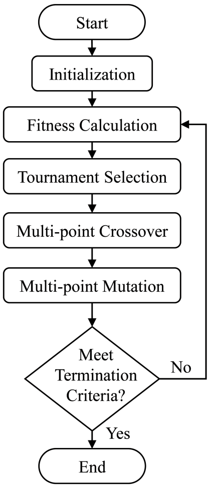

FIGURE 5. Flowchart of GA.

portfolio according to the GA mechanism. The chromosome encode takes all the target stocks into consideration so that the number of stocks in the portfolio is not restricted by the chromosome length. Our flowchart is shown in Fig. 5.

2) INITIALIZATION

We have numbers of chromosomes in one generation. At first, this method initializes the group of chromosomes by randomly giving status to every chromosome. Then a serial number is given to all the chromosomes. For example, we generate ten binary chromosomes whose length is five bits for a portfolio, and randomly choose 0 or 1 for every bit, such as shown in Table 6.

TABLE 6. Gene coding.

Stock	A	B	C	D	E
Bit	1	2	3	4	5
Chromosome I	1	1	0	0	0
Chromosome II	0	1	1	1	0
Chromosome III	1	0	1	0	1
	⋮				
Chromosome IX	0	0	1	1	1
Chromosome X	1	0	0	0	1

3) FITNESS CALCULATION

After initialization, the fitness in GA, the Sharpe ratio is used. The fitness calculation is given in Equation 16. The risk-free rate in Taiwan is the deposit interest rate (0.87%). To fairly compare portfolios, this method utilizes CV to calculate portfolio risk. The definitions of ROI, CV and portfolio risk are given by Equation 17, 18 and 19, respectively. fs is the funds standardization on i_{th} day, $\overline{fs}$ is average of funds standardization, and D is the number of days.

$$fitness = \frac{ROI - risk\text{-}free\ rate}{risk} \tag{16}$$

$$ROI = \frac{final\ funds - initial\ funds}{initial\ funds} \tag{17}$$

$$CV = \frac{standard\ drviation}{average} \tag{18}$$

$$risk = \frac{standard\ deviation\ of\ funds\ standardization}{average\ of\ funds\ standardization}$$
$$= \sqrt{\frac{\sum_{i=1}^{D}(fs_i - \overline{fs})^2}{D}} / \overline{fs} \tag{19}$$

TABLE 7. The stock prices.

Stock	A	B	C	D	E
First Day	80	49	15	35	93
Second Day	85	50	17	38	95

Take the prices of stocks A to E in Table 7. Here, we take ChromosomeI(Chr.I) in Table 6 as an example. If the initial fund is 4000000, the portfolio funds standardization of the first day is 3994471 and the second day, 4136594. In the risk calculation, we calculate the standard deviation and the average by portfolio funds standardization, which are 100496 and 4065532.5, respectively. According to Equation 19, we can derive the portfolio risk by using the standard deviation divided by the average of portfolio funds standardization, which is 1.75%. According to Equation 17, we get the ROI of the portfolio by subtracting the initial funds from final funds and then dividing the initial funds, 3.41%. According to Equation 16, we get the fitness (Sharpe ratio) by subtracting the risk-free rate from the ROI of the portfolio, and then divide the portfolio risk, which is 1.456, such as in Table 8.

TABLE 8. Calculation of evaluation index.

Portfolio I (Chromosome I)	
Initial funds	4,000,000
First day's funds standardization	3,994,471
Second day's funds standardization	4,136,594
Standard deviation	100,496
Average funds standardization	4,065,532.5
Risk	1.75%
Return	3.41%
Sharpe ratio	1.456

4) SELECTION

The concept of evolution involves preserving the better parent generation for breeding superior offspring. Therefore, we select a group composed of the better population with higher fitness for the crossover pool in order to produce better offspring. There are many ways to select a better population, for instance, roulette wheel selection, tournament selection and rank-based wheel selection. In this paper, we use tournament selection to apply our method. Through sorting all the chromosomes by fitness, we preserve some better chromosomes to be a portion of the better population. In this portion, we randomly select a few chromosomes and pick the top two in competition as the parents to produce two offspring. In our example, we order the ten fitness items, and keep 40% of chromosomes: Chr.IX, Chr.II, Chr.III and Chr.VI in Fig. 6. After that, we randomly select three chromosomes from among them: Chr.VI, Chr.IX and Chr.II, and we choose the top two chromosomes to be the parents in competition, namely Chr.IX, and Chr.II.

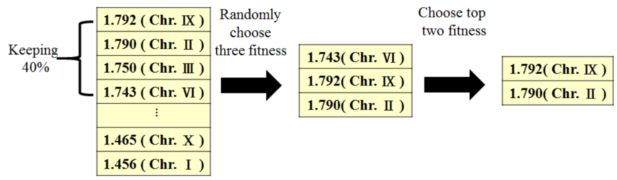

FIGURE 6. Process of selection.

5) CROSSOVER

Likewise, there are many different methods of crossover: single point crossover, multi-point crossover and uniform crossover. In the process of crossover, we use the multi-point crossover technique, randomly deciding how many points and which points should be exchanged between two parent chromosomes. We use the abovementioned parent chromosomes, Chr.IX "00111" and Chr.II "01110", to generate two offspring. Next, we randomly decide how many points will be used for the crossover. If the random number is two, we select two bits of chromosome to be changed. Suppose that we randomly select bits 1 and 5, and then exchange the two parent chromosomes bits with bits 1 and 5 to generate the offspring "00110" and "01111", as shown in Fig. 7.

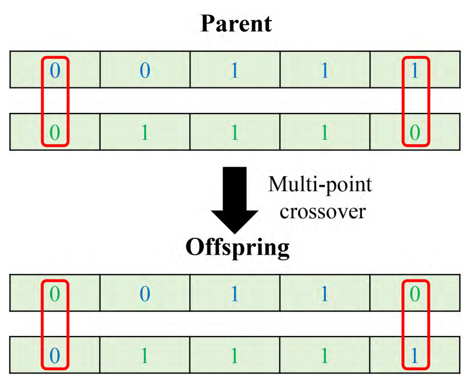

FIGURE 7. Crossover schematic diagram.

6) MUTATION

Mutation is an important part of evolution as it is the source of genetic diversity, enabling the species to become more diversified in order to adapt to different types of environments. Its concept in GA is that chromosomes have some probability of jumping out of the local optimum. Mutation also includes different ways, such as single-point mutation and multi-point mutation. In this paper, we utilize multi-

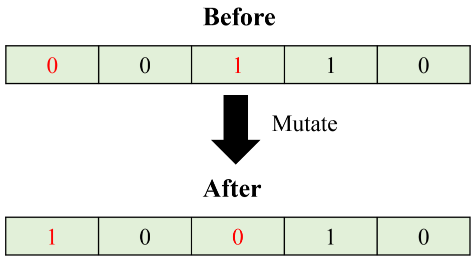

FIGURE 8. Mutating schematic diagram.

point mutation and provide a mutation rate. When every new chromosome is produced, it has some probability to mutate. If the chromosome has to mutate, it will randomly select a few points to be reversed. After the mutation step, our method will repeat the selection, crossover and mutation step, until it fills up the rest of the population. Once we get a new offspring, we generate a number at random to decide whether or not it will mutate. Assume that the mutation rate is 0.5 and we have an offspring chromosome "00110". If the random number is smaller than 0.5, the chromosome does mutate. We generate a number 0.4 randomly for chromosome "00110". Because this number is smaller than the mutation rate, it has to mutate. We randomly select two positions which are 1 and 3 (stock A and stock C), and then reverse the bits. Hence, chromosome 00110 is changed to "10010", as in Fig. 8. After mutation, our method will repeat the steps until it fills up the rest (60%) of the population.

C. SLIDING WINDOWS

There is a great deal of historical information on stock prices in the stock market, and many stocks undergo economic cycles. Analyzing these historical data can help us to find a good investment strategy for investors. However, over-fitting is a common problem in the stock market. We should choose an appropriate length of training period, find the optimal strategy for the training period, and then use the results in the testing period. The length of the training period and testing period are both important factors, and will result in different performance.

To avoid the over-fitting problem, we use the sliding windows in our system. The sliding windows are shown in Fig. 9. Generally, the length of the training period is longer than, or equal to, the testing period. If the length of the training period is too short, we will not have enough information to find a good strategy; if it is too long, it makes us consider historical information that is too dated. Therefore, the goal is to choose periods that are neither too long nor too short. For example, using a one year training period result to test in a one week testing period or a one week training period result to test in a one year testing period is unreasonable. In order to choose the suitable training and testing period, we test for different window sizes in our experiment and analyze the results.

In this paper, we propose 15 types of sliding windows with different training and testing periods. There are three different categories: symmetry, asymmetry and year-on-year, as shown in Table 9. Year-on-year uses the training results in the last year to test in this year. For example, the training result obtained in the first quarter of the 2015 is tested in the first quarter of 2016; it is called Q*. The sliding window of year-on-year is shown in Fig. 10. Appropriate training and testing periods can help investors find a good trading strategy and avoid the over-fitting problem.

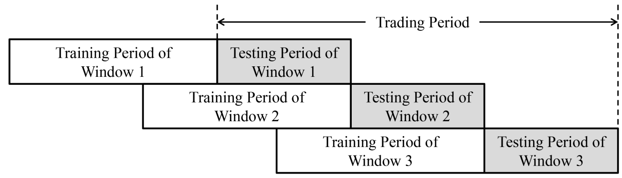

FIGURE 9. Sliding windows.

TABLE 9. The different categories of sliding windows.

Categories	Training and Testing Period
Symmetry	Year-to-Year(Y2Y)
	Half year-to-Half year(H2H)
	Quarter-to-Quarter(Q2Q)
	Month-to-Month(M2M)
	Two weeks-to-Two weeks(B2B)
	Week-to-Week(W2W)
ASymmetry	Year-to-Half year(Y2H)
	Year-to-Quarter(Y2Q)
	Year-to-Month(Y2M)
	Half year-to-Quarter(H2Q)
	Half year-to-Month(H2M)
	Quarter-to-Month(Q2M)
Year-on-year	Half year(H*)
	Quarter(Q*)
	Month(M*)

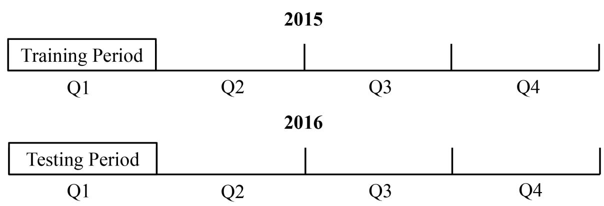

FIGURE 10. Sliding window of Q*.

VI. EXPERIMENT

Our proposed method uses GA combined with the Sharpe ratio to find the best portfolio which has low risk and high return, and we utilize a novel method, funds standardization, to calculate portfolio risk. Moreover, this paper uses sliding windows to avoid the over-fitting problem. In this section, we explain the reason why we chose Taiwan 50 ETF as the investment target and our experimental environment. In addition, we analyze the experimental results for different period tests of sliding windows, and compare the performance with

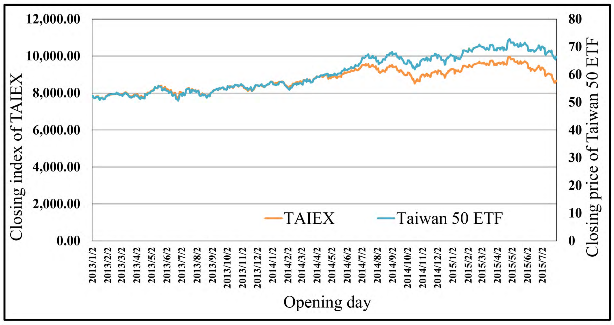

FIGURE 11. The closing prices of Taiwan 50 ETF and the closing index of TAIEX from January, 2013 to July, 2015.

the portfolio which utilized the MPT for calculating risk. It will also be compared with the single stock which has the highest Sharpe ratio, and verify whether the portfolio can spread the risk for investment.

A. INVESTMENT TARGETS

In our experiment, we chose the constituent stocks of Taiwan 50 ETF as our investment target. Taiwan stock market is easily affected by foreign countries, so the stock price of Taiwan often fluctuates. The constituent stocks of Taiwan 50 ETF are the representative stocks in the Taiwan stock market because they are the top fifty stocks in market value. In addition, the total market value of these fifty stocks is up to 70% of the Taiwan stock market, so Taiwan 50 ETF has high relevance with the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), as shown in Fig. 11. Moreover, Taiwan 50 ETF is an exchange-traded fund, and the constituent stocks are verified by the Taiwan Stock Exchange (TWSE) and FTSE. Some constituent stocks of Taiwan 50 ETF will be eliminated because of their lower market value and replaced by other stocks with higher market value. Therefore, the constituent stocks of Taiwan 50 ETF are appropriate investment targets in this paper.

B. EXPERIMENTAL ENVIRONMENT

The source of stock prices in our experiment is the Taiwan Economic Journal (TEJ) during 2010 to June, 2016, and we chose the constituent stocks of Taiwan 50 ETF as our investment target. Some constituent stocks were changed, and in our experiment, we used the constituents stocks of Taiwan 50 ETF which are published by TWSE and FTSE in the first quarter of every year. The constituent stocks that we used are shown in the appendix. Our testing period started from the first trading day in 2010 and ended in June, 2016 (2010/01 to 2016/06). In the following part, we fine-tuned the GA parameter set, and used the appropriate set in our experiment. In our work, we propose 15 different types of sliding windows with different training and testing periods, and describe the 15 different experimental periods, as shown in Table 9 of Section V.

TABLE 10. Testing result of preserved rate at 10 times.

Preserved rate	Number of optimal solutions	Preserved rate	Number of optimal solutions
10%	8	60%	6
20%	6	70%	3
30%	4	80%	1
40%	5	90%	3
50%	5		

TABLE 11. Testing result of preserved rate at 50 times.

Preserved rate	Number of optimal solutions	Preserved rate	Number of optimal solutions
5%	36	35%	37
10%	37	40%	38
15%	38	45%	27
20%	40	50%	29
25%	38	60%	28
30%	32		

C. PARAMETERS OF GA

We fine-tuned the best parameters of GA that can find the optimal solution in every experiment. We wanted to find the best portfolio steadily during the training period, and test its performance in the testing period; therefore, we trained the parameters for GA, which are the preserved rate of parent chromosomes, mutation rate and the numbers of mutation bits for every chromosome.

To test the preserved rate, we set up the group population at 100, the generation number at 200, and tested it in December 2012. The experiment results after testing 10 times for every preserved rate are shown in Table 10. According to the above experiment, we selected the parameters which are able to find over 5 better solutions to do further tests. In this experiment, we increased the testing number to 50 times, and the result is shown in Table 11.

From the experimental results, we find that the best preserved rate of parent chromosomes is 20%. Therefore, we select 20% as the preserved rate in GA. The mutation rate of every single bit for every chromosome is defined as the mutation rate multiplied by the probability of every bit selected to mutate. For example, it is assumed that the chromosomes mutation rate is 10% and the number of mutation bits for every chromosome is 3. In our experiment, every chromosome has 50 bits. Therefore, the rate of every bit selected to mutate is $\frac{3}{50}$, and the mutation rate of every single bit for every chromosome is $\frac{3}{50} \times 10\%$, which equals 0.6%.

To find the best mutation rate of a single bit and the number of mutation bits, we set the mutation rate at 10% and tested it 10 times. The experimental result are shown in Table 12. We then selected parameters 0.6%, 0.4% and 0.2% as they had steadier performance than the others, and tested them 100 times, as shown in Table 13. The mutation rate of single

TABLE 12. The number of mutation bits tested 10 times.

Mutation rate of single bit	Number of mutation bits	Number of optimal solutions
5%	25	0
2.8%	14	0
1.4%	7	1
1.2%	6	2
1%	5	2
0.8%	4	2
0.6%	3	9
0.4%	2	9
0.2%	1	8

TABLE 13. The number of mutation bits tested 100 times.

Mutation rate	Rate of every bit selected to mutate	Mutation rate of single bit	Number of optimal solutions
10%	$\frac{3}{50}$	0.6%	72
10%	$\frac{2}{50}$	0.4%	89
10%	$\frac{1}{50}$	0.2%	42

TABLE 14. The mutation rate of a single bit is 0.6%.

Mutation rate	Number of mutation bits	Optimal solutions in 10 times tests	Optimal solutions in 100 times tests
1%	30	0	X
2%	15	0	X
3%	10	0	X
5%	6	0	X
6%	5	1	X
10%	3	9	70
15%	2	8	92
30%	1	9	73

TABLE 15. The mutation rate of a single bit is 0.5%.

Mutation rate	Number of mutation bits	Optimal solutions in 10 times tests	Optimal solutions in 100 times tests
1%	25	0	X
5%	5	2	X
25%	1	8	66

bits 0.6% and 0.4% exhibited great performance. Hence, we trained the mutation rate and the number of mutation bits by using the mutation rate of single bit 0.6%, 0.4% and 0.5% to find the best parameter; 0.5% is the average of 0.6% and 0.4%, and we tested it to observe whether or not it had better performance.

Tables 14-16 show the experimental results; the mutation rates of single bits are 0.6%, 0.5% and 0.4%, respectively. In these experiments, we tested it 10 times first, and then selected the best parameters to test it 100 times.

According to Tables Tables 14-16, the mutation rate of single bits 0.6% and 0.4% have better performance. Because this method has different periods in sliding windows, we compared the mutation rate of single bits 0.6% and 0.4% in a year,

TABLE 16. The mutation rate of a single bit is 0.4%.

Mutation rate	Number of mutation bits	Optimal solutions in 10 times tests	Optimal solutions in 100 times tests
1%	20	0	X
2%	10	0	X
4%	5	0	X
5%	4	0	X
10%	2	8	82
20%	1	7	54

TABLE 17. The results of 0.4% and 0.6% in different periods with 100 testing times.

Period	Mutation rate of single bit	Mutation rate	Number of mutation bits	Number of optima
Y	0.6%	15%	2	57
	0.4%	10%	2	33
H	0.6%	15%	2	73
	0.4%	10%	2	65
Q	0.6%	15%	2	63
	0.4%	10%	2	72
M	0.6%	15%	2	92
	0.4%	10%	2	82
B	0.6%	15%	2	35
	0.4%	10%	2	35
W	0.6%	15%	2	86
	0.4%	10%	2	59

TABLE 18. Parameter table of GA.

Parameter	Numerical
Initial funds	10,000,000
Group	100
Generation	200
Preserved rate	20%
Mutation rate	15%
mutation point	2
Crossover	Multiple

half year, quarter year, two weeks, and one week, respectively, when the preserved rate of the parent chromosomes is 20%. Table 17 shows the comparison result of 0.4% and 0.6% in different periods with 100 testing times. When the mutation rate of a single bit is 0.6%, most of the periods have better performance than 0.4%.

We utilized the combination of parameters shown in Table 18 for the following experiment. According to Table 17, the period of two weeks had the worst performance. In only 35 times, could we find the optimal solutions in 100 testing times, so the successful rate of two weeks is 35%. If we do 30 experiments independently, the successful rate at which we can find the best solution is better than 99.9997%. Therefore, we can ensure that we selected the optimal solution by using 30 times of GA in the training period and test the optimal solution in testing period.

TABLE 19. The comparison between MPT and our method.

	ROI		Risk		Sharpe ratio	
	MPT	Ours	MPT	Ours	MPT	Ours
Training period 2009/10-2016/03	19.48	13.90	5.60	3.07	3.25	**4.44**
Testing period 20010/01-2016/06	15.80	96.99	15.30	29.09	0.98	**3.30**

TABLE 20. The execution time comparison between MPT and our method.

	MPT	Ours
Execution time	7281(s)	6768(s)

D. COMPARISON WITH MPT

As mentioned above, our method improves on the ignored relation of MPT. To prove that our method can obtain the results as expected, we did a series of experiments in the actual condition of the stock market by comparing our method with the traditional MPT by risk, ROI and Sharpe ratio. The training-to-testing period in our experiment is Q2Q, and the experimental date is from January, 2010 to June, 2016. The results are shown in Tables 19 and 20.

In this experiment, we considered both the risk and ROI in the training period, and selected the portfolio with the higher ROI per unit risk. The results of the training period demonstrate that the risk of the portfolio our method chose is nearly half that of the MPT, and the performance of ROI is also excellent; with a higher Sharpe ratio than MPT. Although the result of the testing period is unpredictable, our method still shows great performance of the portfolio in the testing period. In particular, the performance of ROI is clearly superior to MPT, over six times greater than MPT, and the risk rising scale is lower than MPT; as a result, the performance of the Sharpe ratio outperforms MPT. The experiment demonstrates that our method is able to effectively find a portfolio which has the higher ROI per unit risk in the stock market; it can not only precisely find the best portfolio, but also works more efficiently. In the next part, we analyze the training and testing periods, respectively.

E. SELF-ANALYSIS

1) TRAINING PERIOD ANALYSIS

In the training period, we tested all the stocks in the Taiwan 50 ETF and found the single stock with the highest Sharpe ratio (SHS). We then compared it with the portfolio which our method chose. The result shows that our method, funds standardization with GA, is truly effective; it can find the Sharpe ratio of a portfolio better than the SHS can. Even if the SHS is the best in some cases, our method will still find the portfolio that only includes that stock. Therefore, the Sharpe ratio of the portfolio that our method chose must be higher than, or equal to, the SHS. In the training-to-testing period experiment from January, 2013 to February, 2014, we found that the performance of B2B and W2W was inferior to other periods, as Fig. 12 shows. The superior training-to-testing periods at least have to train or test for a month; thus, we knew that the system in our method is more appropriate for long-term investment; therefore, we eliminated both periods: B2B and W2W in the following experiment. The experiment results are shown in Fig. 13. The Sharpe ratio of the portfolio is higher than the SHS in every period.

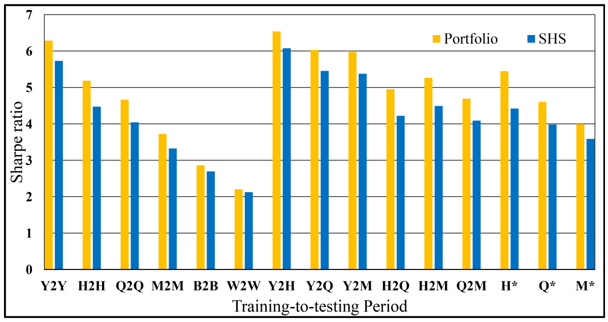

FIGURE 12. Comparison of the Sharpe ratio in every training-to-testing period from January, 2013 to February, 2014.

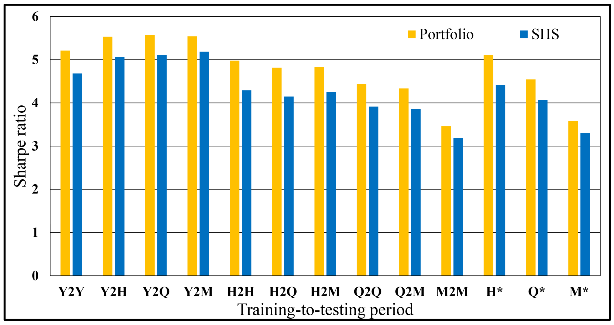

FIGURE 13. Comparison of the Sharpe ratio in every training-to-testing period from January, 2010 to June, 2016.

By observing the fluctuation of risk in every period, the portfolio risks are all lower than with the SHS. The experiment results of risk in every training-to-testing period are shown in Fig. 14. It proves that the portfolio which our method chose can effectively lower the risk, thus avoiding putting all of our eggs in one basket, since the portfolio is able to spread the risk of investment.

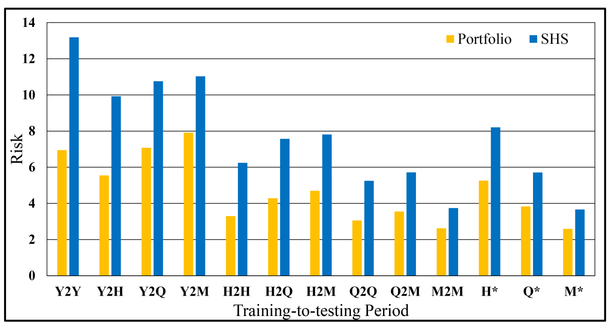

FIGURE 14. Comparison of the risks in every training-to-testing period.

Because of the huge solution space, some methods simplify this problem by taking the SHS with other stocks in the portfolio. The experiment results show that the SHS is unnecessary in the composition of the portfolio. In this period (Q2Q: January, 2016 to March, 2016), the portfolio does not

TABLE 21. Comparison of the best stock, portfolios, and the portfolio which includes the best stock. the rank is sorted by the Sharpe ratio.

	Selected stocks (Rank)	Sharpe ratio
Best stock	2301 (1)	3.21
Portfolio	2912 (3) 2357 (6) 2325 (39)	3.64
Portfolio includes the best stock	2301 (1) 2912 (3) 2357 (6) 2325 (39)	3.39

include the SHS. Even if we pick the SHS and compulsorily add it to the portfolio, it will reduce the Sharpe ratio of the portfolio. Many methods directly select the SHS in a portfolio to simplify the risk calculation and reduce the computing complexity, but our method proves that the SHS is unnecessary in the best portfolio. The simplified methods may not find the best portfolio because of the limit in their stock selection. However, our method not only considers all of the combinations and then finds the best portfolio, but the computational complexity of the portfolio risk calculation is also far lower than the simplified methods. Table 21 shows that the Sharpe ratio will decline if the portfolio includes the SHS in this case.

Traditional studies usually exclude the stocks which have a negative Sharpe ratio to simplify the complexity of this problem. The experimental results show that no stock can be excluded. In Table 22, the portfolio includes the stocks that have lower rank, and their Sharpe ratio is negative; however, if we exclude the negative ones, then the risk will increase substantially and the Sharpe ratio will be reduced. Accordingly, as the portfolio needs to include some declining stocks to balance the whole risk in some cases, our method is unable to exclude any stock when searching for the best portfolio.

TABLE 22. Comparison of the portfolio as to whether or not to exclude the stock with a negative Sharpe ratio.

Stock (Rank)	Sharpe ratio of each stock	Portfolio ROI	Portfolio risk	Portfolio Sharpe ratio
2912 (3) 2357 (6) 2325 (39)	3.06 2.88 -0.30	8.14	2.07	3.64
2912 (2) 2357 (6)	3.06 2.88	12.78	3.50	3.41

The previous simplified methods can reduce a certain amount of calculation, but may exclude the better portfolios. The training period experiment proves that our method, which uses fund standardization, not only fully considers the interactions among the stocks in a portfolio, but also substantially reduces the amount of calculation. Besides, to discover the optimal portfolio, it is unnecessary to include the SHS, and the negative stock should not be excluded arbitrarily. Our method, funds standardization with GA, does not constrain any portfolio combination; therefore, it can find the best portfolio which has the highest Sharpe ratio.

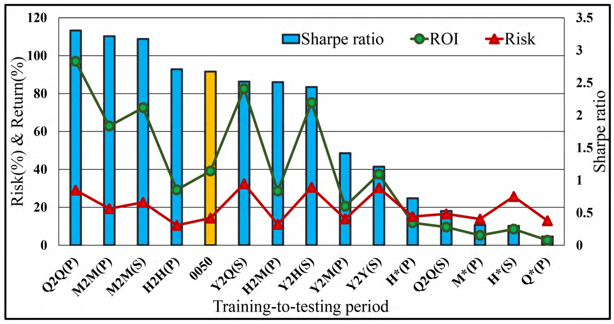

FIGURE 15. Sorted by the Sharpe ratio for every training-to-testing period.

2) TESTING PERIOD ANALYSIS

The testing time is six and a half years, from January, 2010 to June, 2016. The objects of the comparisons are the SHS and Taiwan 50 ETF. We tested them in different periods. Fig. 15 shows the risk, ROI and Sharpe ratio of every training-to-testing period; (P) is the result of the portfolio, which is a combination of stocks (portfolio), and (S) is the result of the SHS.

According to Fig. 15, the best performance of the Sharpe ratio in every training-to-testing period is Q2Q (P). Besides, the results of the portfolio in the majority of the cases show that the Sharpe ratio is higher than in Taiwan 50 ETF. Fig. 16 shows the ranking sorted by ROIs. There are five periods whose ROI are higher than Taiwan 50 ETF. In these five periods, there are three periods whose Sharpe ratio is also higher than Taiwan 50 ETF. The other two periods whose Sharpe ratio are lower than in Taiwan 50 ETF due to their overly-high risks. It proves that the ROI is not the most important assessment in stock selection; risk has to be considered simultaneously.

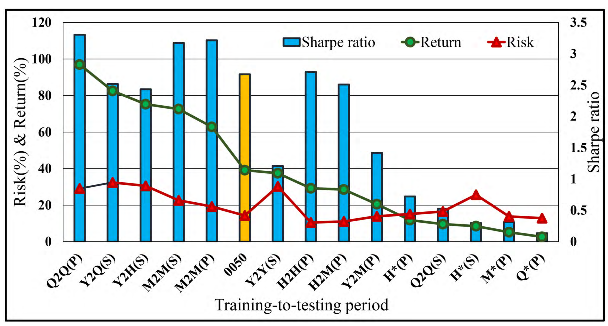

FIGURE 16. Sorted by ROIs in every training-to-testing period.

In the testing periods experiment, the training-to-testing period which has the best performance is Q2Q (P). The performance of the Sharpe ratio is higher than in Taiwan 50 ETF, and in more portfolios than the SHS. Besides, our method can find the portfolio not only with the higher Sharpe ratio but also with the greater return. With our experiments, we know that the system in our method should entail long-term investment, but training-to-testing periods that are too long (e.g. Y2Y) or too short (e.g. B2B and W2W) are inappropriate. Through the experimental results, we find that the better training-to-testing periods are all symmetrical, and the period with the best performance is Q2Q (P).

VII. CONCLUSION

We proposed using funds standardization in order to accurately evaluate the portfolio risk because it is able to effectively overcome the defects of traditional risk calculation. First and foremost, funds standardization properly considers all of the interactions among stocks in a portfolio. Besides, it substantially simplifies the risk calculation. Last but not least, it utilizes the same criteria to compare the portfolios fairly. Combining GA and the Sharpe ratio, which is calculated by funds standardization, can efficiently select low risk and stable return portfolios.

The core idea of the Sharpe ratio is to select the portfolio that has high return at the same risk, or low risk at the same return. However, traditional risk calculation ignores the relationships when there are more than two stocks in a portfolio, so it cannot completely represent portfolio risk. Funds standardization solves this problem because it considers all the interactions among stocks in a portfolio and represents the fluctuation of portfolio investment. Besides the above-mentioned problem, traditional risk calculation includes the problems of complicated calculation and comparisons among the different portfolios. Traditional risk calculation needs to calculate the covariance of each pair of stocks in a portfolio, so the computation complexity increases exponentially. In our method, the portfolio risk barely calculates the standard deviation of portfolio funds standardization because portfolio funds standardization represents the fluctuation of portfolio investment. The computation complexity is reduced to O(1) by the calculation method presented herein. This paper uses CV to avoid the problem of comparing different portfolios. Otherwise, we utilize the sliding window to avoid the over-fitting problem. The experimental results show that our proposed method outperforms the traditional method in all the training and testing periods; it also significantly reduces the complexity of portfolio risk calculation. In addition, we notice that the Sharpe ratio of the optimal portfolio is certainly more than, or equal to, the Sharpe ratio of SHS. The experimental result also proves that a portfolio can effectively spread the risk. In some cases, the optimal portfolio does not necessarily include the single stock with the highest Sharpe ratio, or may include the stock with the worse Sharpe ratio. In other words, with the traditional simplifications of risk calculation, in which a portfolio must include the stocks with the highest Sharpe ratio or exclude the stock with worse Sharpe ratio, defects exist. In the results of the testing periods, Q2Q (P) had the best performance. We find that the training periods and testing periods should not be too long or too short, and all the training-to-testing periods that have great performance are symmetrical. The experimental results show

TABLE 23. The constituent stocks of taiwan 50 ETF in the first quarter of 2009.

The stock symbol of constituent stocks				
1101	1102	1216	1301	1303
1326	1402	1722	2002	2105
2301	2303	2308	2311	2317
2324	2325	2330	2347	2353
2354	2357	2382	2408	2409
2412	2454	2498	2603	2801
2880	2881	2882	2883	2885
2886	2888	2890	2891	2892
2912	3009	3045	3231	3481
4904	5854	6505	8046	9904

that our method can find a portfolio that has lower risk (low-volatility) and higher return in the stock market. As described above, we are able to accurately assess the portfolio, find the portfolio with the optimal Sharpe ratio, and prove that the portfolio can spread risk. As a result, the proposed method is better than the traditional method because funds standardization contains all of the interactions among stocks in a portfolio, reduces the computation complexity, has the ability to compare different portfolios fairly, and represents the mood swings of investors. In future work, we can use other state-of-the-art global optimization algorithms to search for the best portfolio.

APPENDIX
CONSTITUENT STOCKS

Table 23 is the constituent stocks of Taiwan 50 ETF in the first quarter of 2009. 2404, 2603 and 3009 are replaced by 2448, 3474 and 6239 after 2009. 3474, 6239 and 8046 are replaced by 2618, 3673 and 4938 after 2010. 2448, 2618, 2888, 4938, 5854 and 9904 are replaced by 1802, 2201, 2207, 2474, 3008 and 5880 after 2011. 1802 is replaced by 3697 after 2012. 1722, 2347, 2353, 3231, 3673 and 3697 are replaced by 2227, 2884, 2887, 3474, 4938 and 9904 after 2013. 2201 and 2324 are replaced by 2395 and 2408 after 2014. 2498 is replaced by 1476 after 2015.

REFERENCES

[1] W. F. Sharpe, *Investors Markets: Portfolio Choices, Asset Prices, Investment Advice*. Princeton, NJ, USA: Princeton Univ., 2011.

[2] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[3] R. D. Edwards, J. Magee, and W. C. Bassetti, *Technical Analysis of Stock Trends*. Boca Raton, FL, USA: CRC, 2007.

[4] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA; MIT, 1998.

[5] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 1990, pp. 1–6.

[6] J. C. Patra, N. C. Thanh, and P. K. Meher, "Computationally efficient flann-based intelligent stock price prediction system," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2009, pp. 2431–2438.

[7] C.-M. Hsu, "Forecasting stock/futures prices by using neural networks with feature selection," in *Proc. Int. Joint Inf. Technol. Artif. Intell. Conf. (ITAIC)*, vol. 1. Aug. 2011, pp. 1–7.

[8] F. A. de Oliveira, L. E. Zárate, M. de Azevedo Reis, and C. N. Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices," in *Proc. Int. Conf. Syst., Man (SMC)*, Oct. 2011, pp. 2151–2155.

[9] D. Reid, A. J. Hussain, and H. Tawfik, "Spiking neural networks for financial data prediction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 1–10.
[10] A. Stetco, X.-J. Zeng, and J. Keane, "Fuzzy cluster analysis of financial time series and their volatility assessment," in *Proc. Int. Conf. Syst., Man (SMC)*, Oct. 2013, pp. 91–96.
[11] C.-D. Chen and S.-M. Chen, "A new method to forecast the TAIEX based on fuzzy time series," in *Proc. Int. Conf. Syst., Man (SMC)*, Oct. 2009, pp. 3450–3455.
[12] W.-K. Wong, E. Bai, and A. W.-C. Chu, "Adaptive time-variant models for fuzzy-time-series forecasting," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 6, pp. 1531–1542, Dec. 2010.
[13] S.-M. Chen, G. M. T. Manalu, S.-C. Shih, T.-W. Sheu, and H.-C. Liu, "A new method for fuzzy forecasting based on two-factors high-order fuzzy-trend logical relationship groups and particle swarm optimization techniques," in *Proc. Int. Conf. Syst., Man (SMC)*, Oct. 2011, pp. 2301–2306.
[14] S.-M. Chen and S.-W. Chen, "Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups and the probabilities of trends of fuzzy logical relationships," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 391–403, Mar. 2015.
[15] C.-C. Chen, C. Kuo, S.-Y. Kuo, and Y.-H. Chou, "Dynamic normalization BPN for stock price forecasting," in *Proc. Int. Conf. Syst., Man (SMC)*, Oct. 2015, pp. 2855–2860.
[16] E. Hadavandi, H. Shavandi, and A. Ghanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting," *Knowl.-Based Syst.*, vol. 23, no. 8, pp. 800–808, 2010.
[17] Y. Fang, K. Fataliyev, L. Wang, X. Fu, and Y. Wang, "Improving the genetic-algorithm-optimized wavelet neural network for stock market prediction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 3038–3042.
[18] I.-C. Yeh and C.-H. Lien, "Fuzzy rule-based stock trading system," in *Proc. Int. Conf. Fuzzy Syst. (FUZZ)*, Jun. 2011, pp. 2066–2072.
[19] J.-H. Wang and J.-Y. Leu, "Dynamic trading decision support system using rule selector based on genetic algorithms," in *Proc. IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process. (VI)*, Sep. 1996, pp. 119–128.
[20] J.-Y. Potvin, P. Soriano, and M. Vallée, "Generating trading rules on the stock markets with genetic programming," *Comput. Oper. Res.*, vol. 31, no. 7, pp. 1033–1047, 2004.
[21] M.-C. Chen, E. Y.-H. Lin, and R.-J. Kuo, "The application of ant colony optimization system on the investment strategies at Taiwan stock," *J. Inf. Manage.*, vol. 15, no. 1, pp. 153–177, 2008.
[22] F. Wang, P. L. H. Yu, and D. W. Cheung, "Combining technical trading rules using parallel particle swarm optimization based on hadoop," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 3987–3994.
[23] A. Narayanan and M. Moore, "Quantum-inspired genetic algorithms," in *Proc. Int. Conf. Evol. Comput.*, May 1996, pp. 61–66.
[24] Y.-H. Chou, S.-Y. Kuo, C.-Y. Chen, and H.-C. Chao, "A rule-based dynamic decision-making stock trading system based on quantum-inspired tabu search algorithm," *IEEE Access*, vol. 2, pp. 883–896, 2014.
[25] M. A. Dashti, Y. Farjami, A. Vedadi, and M. Anisseh, "Implementation of particle swarm optimization in construction of optimal risky portfolios," in *Proc. Int. Conf. Ind. Eng., Eng. Manage.*, Dec. 2007, pp. 812–816.
[26] A. Raudys and Z. Pabarskaite, "Discrete portfolio optimisation for large scale systematic trading applications," in *Proc. Int. Conf. Biomed. Eng. Inform. (BMEI)*, Oct. 2012, pp. 1566–1570.
[27] K. K. Hung, C. C. Cheung, and L. Xu, "New sharpe-ratio-related methods for portfolio selection," in *Proc. IEEE/IAFE/INFORMS Conf. Comput. Intell. Financial Eng. (CIFEr)*, Mar. 2000, pp. 34–37.
[28] C. R. B. Azevedo and F. J. Von Zuben, "Regularized hypervolume selection for robust portfolio optimization in dynamic environments," in *Proc. Congr. Evol. Comput. (CEC)*, Jun. 2013, pp. 2146–2153.
[29] G. A. V. Pai and T. Michel, "Fuzzy decision theory based optimization of constrained portfolios using metaheuristics," in *Proc. Int. Conf. Fuzzy Syst. (FUZZ)*, Jul. 2013, pp. 1–8.
[30] T. T. Nguyen, L. Gordon-Brown, A. Khosravi, D. Creighton, and S. Nahavandi, "Fuzzy portfolio allocation models through a new risk measure and fuzzy sharpe ratio," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 3, pp. 656–676, Jun. 2015.
[31] A. C. Briza and P. C. Naval, Jr., "Design of stock trading system for historical market data using multiobjective particle swarm optimization of technical indicators," in *Proc. Annu. Conf. Companion Genetic Evol. Comput.*, 2008, pp. 1871–1878.
[32] L. Jia-Long, L. Bo-Wei, and L. Min, "Model contest and portfolio performance: Black–Litterman versus factor models," in *Proc. Int. Conf. Manage. Sci. Eng. (ICMSE)*, Jul. 2013, pp. 507–516.
[33] A. Agarwal, E. Hazan, S. Kale, and R. E. Schapire, "Algorithms for portfolio management based on the Newton method," in *Proc. Int. Conf. Mach. Learn.*, 2006, pp. 9–16.
[34] R. G. Clarke, H. De Silva, and S. Thorley, "Minimum-variance portfolios in the us equity market," *J. Portfolio Manage.*, vol. 33, no. 1, pp. 10–24, 2006.
[35] D. C. Blitz and P. Van Vliet, "The volatility effect," *J. Portfolio Manage.*, vol. 34, no. 1, pp. 102–113, 2007.
[36] T. Doganoglu, C. Hartz, and S. Mittnik, "Portfolio optimization when risk factors are conditionally varying and heavy tailed," *Comput. Econ.*, vol. 29, nos. 3–4, pp. 333–354, 2007.
[37] A. Banner, V. Papathanakos, and P. Whitman, *Enjoying the Benefits of the Low-Volatility Anomaly*. New York, NY, USA: Institutional Investor, 2012.
[38] H. Markowitz, "Modern portfolio theory, financial engineering, and their roles in financial crises," in *Proc. CFA Inst. Conf. Quart.*, 2009, vol. 26. no. 4, pp. 1–6.
[39] H. M. Markowitz, "Portfolio theory: As I still see it," *Annu. Rev. Financial Econ.*, vol. 2, no. 1, pp. 1–23, 2010.
[40] H. Markowitz, "Risk–return analysis," in *Encyclopedia of Quantitative Finance*. Hoboken, NJ, USA: Wiley, 2006.
[41] B. Grechuk, A. Molyboha, and M. Zabarankin, "Mean-deviation analysis in the theory of choice," *Risk Anal.*, vol. 32, no. 8, pp. 1277–1292, 2012.
[42] R. K. Y. Low, R. Faff, and K. Aas, "Enhancing mean–variance portfolio selection by modeling distributional asymmetries," *J. Econ. Bus.*, vol. 85, pp. 49–72, May/Jun. 2016.
[43] S.-Y. Chen and P.-E. Cheng, *Statistics*, 4th ed. Taipei, Taiwan: Hwa Tai Publishing, 2004.
[44] H.-P. Chiang, Y.-H. Chou, C.-H. Chiu, S.-Y. Kuo, and Y.-M. Huang, "A quantum-inspired tabu search algorithm for solving combinatorial optimization problems," *Soft Comput.*, vol. 18, no. 9, pp. 1771–1781, 2014.
[45] S.-Y. Kuo and Y.-H. Chou, "Entanglement-enhanced quantum-inspired tabu search algorithm for function optimization," *IEEE Access*, vol. 5, pp. 13236–13252, 2017.

YAO-HSIN CHOU received the Ph.D. degree from the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, in 2009. He is currently an Associate Professor with the Department of Computer Science and Information Engineering, National Chi Nan University, Nantou, Taiwan. He has authored over 60 papers in journals and conference proceedings. His current research interests include artificial intelligence, evolutionary computation, combinatorial optimization, financial technology, and quantum information science.

SHU-YU KUO received the B.S. and M.S. degrees in computer science and information engineering from National Chi Nan University, Nantou, Taiwan, in 2012 and 2013, respectively, where she is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Engineering. Her current research interests are automated trading systems, evolutionary computation, portfolio optimization, and function optimization.

YI-TZU LO received the B.S. degree in computer science and information engineering from National Chi Nan University, Nantou, Taiwan, in 2015, where she is currently pursuing the master's degree with the Department of Computer Science and Information Engineering. Her current research interests are evolutionary computation, portfolio optimization, and reversible circuits.