Adv. Math. Fin. App., 2022, **7(**4), P. 849-865

Advances in Mathematical Finance & Applications

www.amfa.iau-arak.ac.ir
Print ISSN: 2538-5569
Online ISSN: 2645-4610
Doi: 10.22034/AMFA.2021.1937022.1622

Research Paper

A New Two-Phase Approach to the Portfolio Optimization Problem Based on the Prediction of Stock Price Trends

Amin Karrabi, Aghileh Heydari. Hamid Reza Yousefzade*

Department of Mathematics, Payame Noor University (PNU), P.O. BOX 19395-4697, Tehran, Iran.

ARTICLE INFO
Article history:
Received 00 December 00
Accepted 00 February 00

Keywords:
Multi-objective optimization
Support vector regression
Multi-objective particle swarm optimization
Efficient Frontier

ABSTRACT

Forming a portfolio of different stocks instead of buying a particular type of stock can reduce the potential loss of investing in the stock market. Although forming a portfolio based solely on past data is the main theme of various researches in this field, considering a portfolio of different stocks regardless of their future return can reduce the profits of investment. The aim of this paper is to introduce a new two-phase approach to forming an optimal portfolio using the predicted stock trend pattern. In the first phase, we use the Hurst exponent as a filter to identify stable stocks and then, we use a meta-heuristic algorithm such as the support vector regression (SVR) algorithm to predict stable stock price trends. In the next phase, according to the predicted price trend of each stock having a positive return, we start arranging the portfolio based on the type of stock and the percentage of allocated capacity of the total portfolio to that stock. To this end, we use the multi-objective particle swarm optimization algorithm to determine the optimal portfolios as well as the optimal weights corresponding to each stock. The sample, which was selected using the systematic removal method, consists of active firms listed on the Tehran Stock Exchange from 2018 to 2020. Experimental results, obtained from a portfolio based on the prediction of stock price trends, indicate that our suggested approach outperforms the retrospective approaches in approximating the actual efficient frontier of the problem, in terms of both diversity and convergence.

1 Introduction

In any trading market, every investment-based decision contains two crucial factors, namely, *risk* and *return*. The risk is defined as a measurable potential loss of an investor. The return is considered as the set of benefits an investor achieves during an investment period. This period can vary from daily to annual, or it can be any deterministic period of time according to the predefined time horizon of an investor. It is clear that in any investment, the investor tries to avoid the risk and tends to make more profit. Uncertainty in the capital market, fluctuations in prices, and different returns of companies' stocks make investors anxious and cautious about

* Corresponding author. Tel.:
E-mail address: *usefzadeh.math@pnu.ac.ir*

the outlook of their investment. One of the most important ways to reduce such concerns is to invest in a set of different stocks, that is, to form a *portfolio*. Choosing a portfolio is one of the most common issues that different investors with different levels of capital always face, from relatively small portfolios with a small number of stocks, properties, etc. to large portfolios that include a variety of assets, and should be managed by professional investors. The main concerns in choosing a portfolio are choosing the best possible combination of assets and determining the appropriate weight for each. Recently, researchers have made extensive efforts to provide methods for determining the optimal portfolio. Harry Markowitz can be considered as a pioneer in the optimization modeling of portfolio selection issues. He was the first researcher to study the concept of portfolio and its varieties, who showed that a diverse portfolio can reduce the investment risk [17]. Markowitz's mathematical modeling was a long way from the real world, but it had a profound effect on improving the portfolio selection procedure, and many researchers thereafter refined his theory. But, so far, no comprehensive model has been proposed to enable investors to choose the optimal portfolio of investments. Markowitz's model has two important drawbacks. First, it is not appropriate for the long-term horizon, and second, it forms a portfolio only based on the price trend of a stock in the past (retrospectively). If a portfolio of different stocks is formed by ignoring the future return trend of those stocks (futurist), the potential loss of the portfolio can be increased. Since investing is a long-term concept, in this article we first use a filter to identify stable stocks.

A stable stock is one whose price trend does not fluctuate sharply, regardless of the general state of the stock market. Using the stability filter in this article, we make the proposed portfolio valid for a longer period of time. After the separation of stable stocks, we use a new two-phase approach to examine the way a portfolio can be formed on the basis of futuristic patterns. To the best of our knowledge, such a study has not been conducted so far. Now, the main questions we aim to answer can be written as follows. Can the proposed approach be used to form a future portfolio with appropriate desirability? To what extent has this approach been able to bring the proposed portfolio closer to the real portfolio of the issue in the medium-term (three-month) period? The results of this research can be used effectively for investment companies, senior managers, financial analysts, investment funds, researchers in the field of finance, and investors. The main part of this paper is organized as follows.

Section 2 describes the research background of the studies conducted on the optimization of portfolio selection. In Section 3, we briefly discuss the structure of the support vector regression (SVR) algorithm and the Hurst exponent as a filter to identify stable stocks. The basics of the portfolio optimization problem and the structure of the multi-objective particle swarm optimization (MOPSO) algorithm are declared in Section 4. In Section 5, the structure of the new two-phase approach is described. In the first step, we use the Hurst exponent as a filter to identify stable stocks, and then use the SVR algorithm to forecast different stock prices in the Tehran Stock Market (TSE) in a period of three months. In the second step, using the forecasted price trends, we calculate the optimal stocks composition and the optimal weight corresponding to each stock in the optimal portfolio. Finally, Section 6 is devoted to the conclusion.

2 Literature Review

In recent years, the trend of economic globalization has resulted in a significant increase in the spread of financial crises from one market to the others, hence the risk of investing in the financial markets such as the securities and the exchange market rises [2]. A strategy to control the investment risk is to form a portfolio of different stocks. However, the main issue in portfolio arrangement is the optimal selection of assets and securities to which a certain amount of total capital can be allocated. Due to the complexities in the context of risk minimization and maximum return on investment, in recent years, extensive efforts have been made to form an optimal stock portfolio; such as Markowitz can be considered as one of the pioneers in this field [17]. Afterwards, numerous researchers developed and modified the Markowitz model. In 2005, Maringer et al., in the Markowitz model, considered an additional constraint on the maximum number of stocks in the portfolio to facilitate portfolio management and to reduce the corresponding costs [16]. In 2007, Fernandez and Gomez by considering the constraints of the minimum and the maximum number of stocks in a portfolio proposed a new model called the cardinality constraints mean-variance model [7]. In 2014, Ghasemi et al. optimized the stock portfolio using a new model where in addition to allowing short selling; they added some constraints on the capital markets such as the maximum amount of short-selling, the maximum number of stocks in the portfolio, and the upper and lower bounds for the total amount of assets to this model [8].

In the real world, we face with the real constraints in modeling the problem, and hence, the exact mathematical methods do not cope with such large-scale problems. In this regard, the meta-heuristic methods are suggested as efficient methods to solve the portfolio optimization problems [9]. To this end, most researchers have called various meta-heuristic methods to solve the problem of determining the optimal stock portfolio. For further studies, we refer the readers to some references such as Unal et al. in 2020 [20] in which they solved the optimization problem of the stock portfolio by the MOPSO algorithm, or Vasiani et al., in [21], solved the problem by the genetic algorithm (GA) and Sahala et al., in 2020 in [19] used the artificial bee colony (ABC) algorithm for this optimization problem. On the other hand, we believe that an investment plan such as stock portfolio selection should consider not only the past performance of the stock but also the future potential of the stock. Each approach considering these perspectives affects the forecast's accuracy of stock price. In recent years, there have been extensive efforts to forecast the stock market and various financial markets.

In 2018, Bernardo et al. combined the SVM and GA with each other and proposed a new algorithm. The results showed the high accuracy of their proposed algorithm in Forex market forecasting [4]. In 2020, Alahmari forecasted the prices of three well-known cryptocurrencies, i.e., Bitcoin, Ripple, and Ethereum, using the SVR algorithm concluding three different kernel methods (linear, polynomial, and radial basis). The numerical results indicated that the prediction accuracy of the SVR algorithm with the kernel of radial basis function was superior to the other kernels [1]. In 2020, Das et al. predicted the price of several stocks on the Bangladesh stock exchange using the SVR algorithm with two linear kernels and a radial base function. They used the sum of squared errors as a measure to determine the prediction accuracy of each algorithm. Their results showed that the prediction accuracy of the SVR algorithm with the linear kernel (the accuracy was about 96.82%) was superior to the kernel of radial base function (the accuracy was about 97.06%) [6]. The study of the literature review regarding the optimal portfolios shows that no approach takes into account the forming of stock portfolio based on forecasting the price trend of stocks. Therefore, we can divide the researches considering the optimal portfolios into two general categories; (i) the studies in which only the price trends forecasting is investigated and (ii) the studies that only concentrated on the retrospective portfolio. The main purpose of this paper is to introduce a two-phase approach to determine the optimal stock portfolio based on predicting different stock price patterns for the medium-term. This approach consists of two main phases. In the first phase, we use the Hurst exponent as a filter to identify stable stocks, and then, we predict the trend of different stock prices in the Tehran stock exchange using the SVR algorithm. In the second phase, by using the obtained results from the first phase and calling the MOPSO as a meta-heuristic algorithm[1], we determine the optimal combination of different stocks as well as the optimal weight corresponding to each stock in the optimal stock portfolio.

3 Prediction and Examination of the Stability Trend of Different Stocks (Phase I)

Stock markets play a key role in the optimal allocation of financial resources, and therefore, in the economic prosperity of countries. Therefore, the stability of these markets is very important for all market factors, especially investors and politicians. In recent years, especially in 2019, various reasons, including the reduction of bank profits, stagnation in other markets, and the government's incentive policies to invest in the stock market, caused a large amount of people's capital to flow to this market. The inflow of this amount of liquidity increased speculative activity and sharp price fluctuations, especially in the shares of small companies. Since in this research we are looking to introduce the optimal stocks in the medium term, according to the specific conditions governing the study period, we study the active and top firms listed on the Tehran Stock Exchange.

Various methods have been proposed to identify stable stocks, one of which is the use of the Hurst exponent. The Hurst exponent is a measure of stability or long-term memory in time series. Various methods have been proposed to calculate the value of the Hurst exponent, one of which is to use the R /S analysis method. The process of calculating the Hurst exponent using the R/S analysis method is described in the following algorithm. We refer the reader to [23] for more details.

Calculation of the Hurst exponent using the R/S analysis method

Input: Insert a set $Points := \{(x_s, y_s) \in R^2 | \ s = 0,1,2,...,N\}$ containing $N+1$ data point due to the stock prices in a predefined time period.

Output: Determine the value of the Hurst exponent (H).

1 **Divide** the set $Points$ into d subsets $Z_{\tau,m}$ $(m = 1,2,...,d)$ of equal length τ with the components $Z_{\tau,m}(i)$ $(i = 1,2,...,\tau)$.

Repeat Steps 1-2 to1-5 for all $m = 1, 2, ..., d$.

2 **Calculate** the mean $E_{\tau,m}$ and the standard deviation $S_{\tau,m}$ for the subset $Z_{\tau,m}$.

3 **Set** $X_{\tau,m}(i) := Z_{\tau,m}(i) - E_{\tau,m}$.

4 **Calculate** the cumulative series $Y_{\tau,m}$ corresponding to $X_{\tau,m}$.

$$Y_{\tau,m}(j) = \sum_{i=1}^{j} X_{\tau.m}(i), \qquad m = 1,2,...,d \quad , \quad j = 1,2,...,\tau$$

5 **Calculate** the deviation of the maximum and minimum values of $Y_{\tau,m}$.

[1] To the best of our knowledge, the performance of MOPSO is better than the other meta-heuristic algorithms in solving multi-objective optimization problems in stock markets. The sensitive analysis about the performance of meta-heuristic algorithms does not studied in this paper.

$$R_{\tau.m} = max_{\,i=1,2,\ldots,\tau}\{Y_{\tau,m}(i)\} - min_{\,i=1,2,\ldots,\tau}\{Y_{\tau,m}(i)\}.$$

6 **Calculate** the ratio $(\frac{R_{\tau,m}}{S_{\tau,m}})$ for each $Z_{\tau,m}$, and **set** for each τ,

$$\left(\frac{R}{S}\right)_{\tau} \coloneqq \frac{\sum_{m=1}^{d}\left(\frac{R_{\tau,m}}{S_{\tau,m}}\right)}{m}.$$

7 **Plot** the values of $(R/S)_{\tau}$ as a function of τ in a logarithmic graph.

8 **Fit** a linear regression for the obtained points by using the least squares method, and **define** the slope of the regression line as the Hurst exponent (H).

It is worth mentioning that the Hurst exponent H is usually between 0 and 1; H = 0.5 means that the time series is random or uncorrelated, and there is no significant relationship in the stock price trend; $0 \leq H < 0.5$ indicates that the time series is unstable and has a short-term memory, while $0.5 < H \leq 1$ means that the time series is stable and has a long-term memory. After identifying the stable stocks, we predict their price trends using the past prices. With the rapid development of artificial intelligence and machine learning methods in recent years, some intelligent methods, such as the artificial neural network (ANN) and the SVR, have been widely used to predict financial time series [11,13]. In this section, we detail Phase I of the two-phase approach proposed for determining the optimal portfolio. In this phase, using the SVR algorithm and considering as data points the closing prices per stock in each trading day of a period of time (considered to be a one-year period here), we forecast the medium -term (the three-month) price trend of any stock. In what follows, we briefly describe the SVR algorithm. Suppose that we are given a training data set $S = \{(x_i,y_i): x_i \in R^d, y_i \epsilon R, i = 0,1,2,\ldots,N\}$, where R^d denotes the space of input patterns. The aim of the SVR algorithm is to find a function $f(x)$ whose deviation from the actual y_i is at most ϵ for each of the training data:

$$f(x) = w^T.x + b \quad x,w \in R^d, b \in R. \tag{1}$$

Considering the loss function and slack variables ξ_i^+, ξ_i^- to control the infeasible constraints of the optimization problem [23], it can be written as

$$\text{Minimize } \varphi(w,\xi_i^+,\xi_i^-) = \frac{\|w\|^2}{2} + C(\sum_{i=0}^{N}\xi_i^+ + \sum_{i=0}^{N}\xi_i^-) \tag{2}$$

$$y_i - f(x_i) \leq \varepsilon + \xi_i^+ \qquad i = 0,1,2,\cdots,N$$

$$f(x_i) - y_i \leq \varepsilon + \xi_i^- \qquad i = 0,1,2,\cdots,N$$

$$\xi_i^+,\xi_i^- \geq 0,$$

where the positive constant C has to be selected by the user. It determines the penalty parameter of the error term. The quadratic optimization problem (2) can be easily solved in its dual form. The complete SVR equations are fully expressed in [23], and the summarized equation is given by

$$\text{Maximize } \; -\varepsilon\sum_{i=0}^{N}(\alpha_i^+ + \alpha_i^-) + \sum_{i=0}^{N}y_i(\alpha_i^+ - \alpha_i^-) - \frac{1}{2}\sum_{i=0}^{N}\sum_{j=0}^{N}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)(x_i.x_j) \tag{3}$$

$$\sum_{i=0}^{N} \alpha_i^+ = \sum_{i=0}^{N} \alpha_i^-$$

$$0 \le \alpha_i^+ \le C \qquad i = 0,1,2,\cdots,N$$

$$0 \le \alpha_i^- \le C \qquad i = 0,1,2,\cdots,N,$$

where α_i^+, α_i^- are the Lagrange multipliers. After solving the optimization problem (3) and finding the optimal values $\alpha_i^+$and α_i^-, we calculate the optimal values of the variables w and b. We refer the reader to [23] for further reading.

4 Portfolio Optimization (Phase II)

According to Markowitz's theory, an efficient investment portfolio is one that provides the maximum return for a fixed level of risk and the lowest risk for a fixed level of return [17]. As long as the number of assets to be invested and the number of market constraints are small, Markowitz's model can be solved by quadratic programming methods. But, when the constraints of the real-world problem are taken into account, this problem cannot be solved using classical mathematical methods, and hence the importance of using evolutionary algorithms to determine the optimal portfolio is increased. In portfolio optimization, we usually deal with a two-objective optimization problem whose aims are to maximize the returns and minimize the portfolio risk. In what follows, we explain the way one can build the stock portfolio based on the results obtained in phase II.

Portfolio return. Portfolio return is defined as the weighted average of the expected returns of each stock in a portfolio. It can be calculated by the formula

$$\mu_L = \sum_{i=1}^{N} \mu_i w_i, \tag{4}$$

where μ_L, N and μ_i are the returns of portfolio L, the number of stocks in the portfolio, and the average expected return of stock i, respectively. Also, w_i is the weight of stock i in the optimal portfolio [17].

Portfolio risk. There are some different methods for evaluating portfolio risk, including the *variance* and *semi-variance* methods introduced by Markowitz [17, 3]. Portfolio risk depends not only on the weighted average risk of each stock of the portfolio, but also on the covariance or relationships among the returns of stocks that form the portfolio. According to [17], this can be calculated by the formula

$$\sigma_L^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}, \tag{5}$$

where w_i and w_j are defined as the weights of stocks i and j in portfolio L, σ_{ij} denotes the covariance returns of stocks i and j, and σ_L^2 is the variance (risk) of portfolio L.

Now, considering the two objective functions of the problem, namely, maximizing the returns and minimizing the risk, the model of the portfolio optimization problem can be formulated as

$$
\begin{aligned}
& min\ \sigma_L^2 = \sum_{i=1}^{N}\sum_{J=1}^{N} w_i w_j \sigma_{ij} \\
& \max\ \mu_L = \sum_{i=1}^{N} \mu_i\, w_i \\
& \max\ \mu_L = \sum_{i=1}^{N} \mu_i\, w_i \\
& \sum_{i=1}^{N} w_i = 1 \\
& w_i \geq 0 \qquad i = 1,2,\cdots,N,
\end{aligned} \tag{6}
$$

where $\sum_{i=1}^{N} \mu_i\, w_i$ is portfolio return, and $\sum_{i=1}^{N}\sum_{J=1}^{N} w_i w_j \sigma_{ij}$ indicates portfolio risk [17]. The constraint $\sum_{i=1}^{N} w_i = 1$ indicates that the amount of capital corresponding to the stocks in a portfolio is equal to the total investment of a stockholder. Finally, the constraint $w_i \geq 0$ means that the investor always has the right to sell the stocks he/she owns. The problem of portfolio optimization is an *NP*-hard problem [15]. This is why it cannot be solved by the classical methods in a reasonable time, and we must use evolutionary algorithms to solve it. Due to the proper performance of the PSO algorithm in solving single-objective optimization problems, many researchers have used this algorithm to solve multi-objective optimization problems (MOOPs). Several versions of the algorithm, improved for solving MOOPs, exist in the literature. (See [17, 18] for example.) Therefore, in this paper we consider this algorithm as a tool for solving the stock portfolio optimization problem. As mentioned before, the aim of this paper is to propose a new approach for determining the optimal stock portfolio. Hence, we do not investigate the way portfolio optimization may be affected by the kind of meta-heuristic algorithm one chooses.

4.1 The Particle Swarm Optimization (PSO) Algorithm

In 1995, the PSO algorithm was first introduced by Kennedy and Eberhart in [16]. This method, which is an iterative process like other meta-heuristic algorithms, is inspired by the social behavior of some animals in a group manner, such as the collective movement of fish school or bird flocks. In this algorithm, several particles are distributed in the search space of functions to be optimized. Each particle calculates the value of the objective function at the position in which it is located, and then, using a combination of the information, namely, its current location, the best position known by itself as well as the best position known by all particles, chooses the direction of movement. Therefore, any particle chooses a direction to move, and hence, one step of the algorithm ends. These steps are repeated until the desired solution is achieved. Each particle in the PSO algorithm consists of three *d*-dimensional vectors, in which *d* is the dimension of the search space. For a particle, say particle i, these vectors are defined as follows.

- The current position of particle *i* in iteration *k* is denoted by x_i^k.
- The speed of particle *i* in iteration *k* is denoted by v_i^k.

- The best position particle i has ever experienced is denoted by $x_i^{i.best}$, and the best position of all particles is $x^{\mathrm{g,\,best}}$.

In any iteration, a solution x_i is computed. If this solution is better than the previous solutions, it is stored as $x_i^{i.best}$. The values of the objective function at x_i and $x_i^{i.best}$ are denoted by f_i and $f_i^{i.best}$, respectively. The positions of particles in the initial iteration are randomly generated, and all initial velocities are set equal to 0. In the next steps, the position and velocity of each particle are updated using

$$v_i^{k+1} = wv_i^k + c_1r_1\left(x_i^{i.best} - x_i^k\right) + c_2r_2\left(x^{\mathrm{g,\,best}} - x_i^k\right) \qquad (7)$$
$$x_i^{k+1} = x_i^k + v_i^{k+1},$$

where w is the inertia coefficient, denoting the tendency of each particle to move along its current velocity vector, r_1 and r_2 are random numbers in [0,1] with a uniform distribution which yields diversity in solutions, and finally, c_1 and c_2 are positive, constant parameters called "acceleration coefficients". The pseudo-code of the PSO algorithm is summarized as follows.

The Pseudo-code of the PSO algorithm

Step 1- **Create** the initial population and **evaluate** its individuals.

Step 2- **Determine** the best position of each particle and the best global position.

Step 3- **Update** the velocity and position of any particle solutions by (7), and **evaluate** the newly generated.

Step 4- **Go** to Step 2 if the stopping criterion is not met.

Step 5- **End**.

4.2 The multi-objective particle swarm optimization (MOPSO) algorithm

The MOPSO algorithm was first introduced by Coello [5] in 2004. This algorithm is a generalization of the PSO algorithm that is used to solve MOOPs. Unlike single-objective optimization problems, in MOOPs, no single solution can simultaneously optimize all objectives. Hence, we can find a set of Pareto solutions (or non-dominated solutions) that form the efficient frontier (or Pareto optimal). Furthermore, in the MOPSO, a new concept called *archive* has been added to the PSO to preserve the non-dominated solutions that are determined during the search process. Initially, the particles are randomly generated in the MOPSO. It is noteworthy that in this paper, each particle in the MOPSO represents a portfolio, and its position represents the corresponding weights for stocks in the portfolio. Then, we compare the individual particles and store the non-dominated individuals of the population in an archive. Since it is not necessarily possible in the MOPSO to select a particle as the best global position, each particle chooses an individual from the archive as the leader. Since the MOPSO always tends to search for more space in the problem, it tries to choose the selected particle from the areas where particles with less dispersion remain in the archive. Therefore, we first prioritize the discovered objective space and then calculate the probability of each selected area; a probability function is considered that satisfies the following conditions.

$$\sum_{i=1}^{n} P_i = 1 \qquad (8)$$
$$0 \le P_i \le 1$$
$$n_i \le n_j \Leftrightarrow P_i \ge P_j.$$

Herein, P_i is the probability that area i be selected, and n_i is the number of archive members in the area. The third condition means that areas with more archive members are less probable to be selected. In this paper, we use the Boltzman method to calculate the probability of each region [9]. The method is based on the formula

$$P_i = \frac{e^{-\beta n_i}}{\sum_j e^{-\beta n_j}}, \tag{9}$$

where P_i is defined as before, and β is the selection pressure parameter, which is determined by the user according to the importance of selecting areas with fewer non-dominated points. After calculating the probability of each area, an area is selected using the Roulette-wheel method, and then, one of the members of the archive which lies in that area is randomly selected as a leader. After selecting a leader by any particle and moving towards it, the best position of any particle must be updated. To compare the best position of each particle with its current position, we consider the following instructions.

a) If the new position dominates the best position of that particle, then substitute the best position with the new one.

b) If the new position is dominated by the best position of that particle, do nothing.

c) If none of the positions dominates the other, randomly choose one of them as the best position.

By adding non-dominated members of the current population to the archive, once again, we examine the members of the archive and remove the dominated members. In this algorithm, the cardinality of the archive is restricted by a bound; if the cardinality exceeds the specified value during the processing of the algorithm, the extra members of the archive are omitted. Again, use a mechanism similar to the leader selection mechanism, except that areas with more archive members are more probable to be removed. Therefore, to calculate the probability of removing any particle from the archive due to area i, namely, q_i, we use the equation

$$q_i = \frac{e^{\gamma n_i}}{\sum_j e^{\gamma n_j}}, \tag{10}$$

where γ is the removal pressure parameter whose value is selected by the user according to the priority of areas having greater cardinalities? In what follows, we summarize the main structure of the MOPSO algorithm.

The Pseudo-code of the MOPSO algorithm

Step 1- **Create** an initial population.
Step 2- **Select** the non-dominated solutions and save them in the archive.
Step 3- **Prioritize** the explorative objective space.
Step 4- **Choose** a corresponding leader from the archive for any particle to move.
Step 5- **Update** the best position of each particle.
Step 6- **Add** non-dominated solutions of the current population to the archive.
Step 7- **Remove** dominated solutions from the archive.
Step 8- **Remove** extra archive solutions **if** the cardinality of the archive exceeds the specified bound.
Step 9- **If** the stopping criteria are not met, then **go** to Step 3; **otherwise**, stop.

5 Numerical Results

In this research, we first examined the financial data of active and top companies of the Iran Stock Exchange based on their close prices during the years 2018 to 2020. (These companies have been selected from different industry groups according to their type of activity.) During this period, the market experienced both ascending and descending periods, which made it a very suitable choice for testing the proposed approach. In addition, the market experienced various political and social conditions during this period, including parliamentary elections, sanctions, and the sudden influx of people to the stock market, which had significant psychological effects on the stock prices. The examination of this period can be interesting, because it shows the reaction of the proposed approach to various psychological conditions (both favorable and unfavorable) in the market. We chose the desired stocks from the TSE according to the combination of the following measures.

a) The liquidity of stocks and the market capacity of the stock.

b) Alternation of stock trading in the trading hall (the number of traded days).

c) The effect of a company on the market (the average number of stocks issued and the average current value of the company's stock).

In Phase I of the proposed approach, we first used the Hurst exponent as a filter to identify stable stocks. The results related to the Hurst exponents of the stable stocks in 2018 and 2019 are given in Table 1.

Table 1: The Hurst exponents of active and top companies in the stock market in 2018 and 2019

Company	Hurst exponent		Company	Hurst exponent	
	2018	2019		2018	2019
Iran Tele. Co.	0.5695	0.9146	Atieh Dade Pardaz	0.6558	0.5066
S*Iran Transfo	0.7621	0.7890	Saderat Bank	0.8326	0.5650
Iran Const. Inv.	0.5987	0.8532	Zamyad	0.5880	0.7073
Khorasan Steel Co.	0.8117	0.8147	S*Tejarat Bank	0.6685	0.6925
Khouz. Steel	0.9848	0.7150	S*Mellat Bank	0.6550	0.6285
S*Azarab Ind.	0.8552	0.8551	Inf. Services	0.8996	0.9379
Ind. & M. L.	0.6461	0.6405	Shahrood N.E	0.6559	0.7178
S*IRI Marine Co.	0.7088	0.8891	Pars Int. Mfg.	0.5844	0.6860
Chadormalu	0.7684	0.5664	Shazand Petr.	0.5638	0.8306
Tehran Cement	0.9156	0.7440	Behran Oil	0.7229	0.5923

According to Table 1, since the Hurst exponent of each of the stocks is more than 0.5 in both 1397 and 1398, it is stable and the price trends are predictable. To predict the price trend of each stock, the close prices per stock during the period 2018 to 2019, taken from the website of “Tehran Stock Exchange Technology Management Company (TSETMC)”, were considered as data points for the SVR algorithm to predict the price trend of the stock in the first quarter of 2019 and 2020. In Phase II, using the financial measures of the return and the risk, we determined the optimal stock portfolio and the optimal ratio assigned to each stock of the total portfolio capacity. Then, applying the SVR algorithm to the aforementioned top stocks, we forecasted the price trends of these stocks in 2018 and 2019. We used some measures to evaluate the performance of the SVR algorithm in the accuracy of predicting stock price trends. In light of these measures, we evaluated the performance of the SVR algorithm and the quite popular ANN method.

5.1 Data Normalization

Since non-normalized data reduce the speed and accuracy of prediction, the inputs, and in some cases the outputs, must be normalized. To normalize the input data, we use the equation

$$x_N = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{11}$$

where x_{min} and x_{max} are the minimum and maximum data, respectively, and x_N represents the normalized data corresponding to the data x.

5.2 Predictive Accuracy Measures

In this section, we explain some estimator measures such as the *mean squared error* (MSE), the *relative squared error* (RSE), and the *mean absolute error* (MAE). We used these measures to compare the estimated values (namely, those obtained from the SVR and ANN algorithms) with the actual values. These measures of the quality of an estimator are defined by the formulas

$$MSE = \frac{1}{N+1}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2$$
$$RSE = \frac{\sum_{i=1}^{N}(\hat{x}_i - x_i)^2}{\sum_{i=1}^{N}(\bar{x}_i - x_i)^2} \tag{12}$$
$$MAE = \frac{1}{N+1}\sum_{i=1}^{N}|\hat{x}_i - x_i|\,,$$

where x_i , $\hat{x}_i$ and $\bar{x}_i$ represent the actual values, the predicted values, and the mean of the actual values, respectively. The results obtained by applying the SVR and ANN algorithms to the twenty stocks regarding the aforementioned predictive accuracy measures are presented in Table 2. It is worth mentioning that the neural network used in this article is a three-layer perceptron neural network. There are three input layer neurons, which are the closing price of the stock on the previous trading day, the closing price of the stock on two trading days earlier, and the closing price of the stock on four trading days earlier. Also, there are five hidden layer neurons, and there is only one output layer neuron, which is the closing price of the stock on the next trading day. Moreover, the activation functions used in this perceptron network are the hyperbolic tangent function in the hidden layer and a linear function in the output layer. The neural network structure used in this paper is shown in Figure 1.

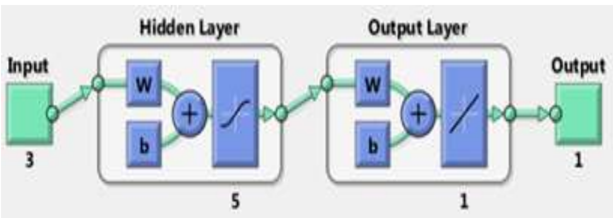

Fig. 1: The structure of the artificial neural network used in this paper

Table 2: The Performance of the SVR and ANN Algorithms On the Twenty Stocks

	2018						2019					
	The SVR algorithm			The ANN algorithm			The SVR algorithm			The ANN algorithm		
	MSE	**RSE**	**MAE**	**MSE**	**RSE**	**MAE**	**MSE**	**RSE**	**MAE**	**MSE**	**RSE**	**MAE**
Iran Tele. Co.	0.0030	0.0947	0.0329	0.0133	0.1435	0.0767	0.0012	0.0152	0.0117	0.0025	0.0217	0.0265
S*Iran Transfo	0.0041	0.0714	0.0466	0.0039	0.0961	0.0472	0.0008	0.0088	0.0201	0.0013	0.0150	0.0345
Iran Const. Inv.	0.0012	0.0203	0.0261	0.0021	0.0221	0.0357	0.0005	0.0083	0.0173	0.0011	0.0095	0.0251
Khorasan Steel Co.	0.0014	0.0282	0.0248	0.0087	0.1042	0.0562	0.0011	0.0165	0.0243	0.0015	0.0169	0.0283
Khouz. Steel	0.0027	0.0420	0.0392	0.0036	0.0548	0.0461	0.0006	0.0163	0.0155	0.0081	0.0879	0.0278
S*Azarab Ind.	0.0041	0.1039	0.0486	0.0037	0.1006	0.0492	0.0007	0.0108	0.0202	0.0008	0.0119	0.0243
Ind. & M. L.	0.0050	0.1164	0.0527	0.0101	0.1676	0.0715	0.0007	0.0122	0.0183	0.0009	0.0217	0.0274
S*IRI Marine Co.	0.0061	0.1060	0.0567	0.0067	0.1198	0.0605	0.0079	0.0597	0.0574	0.0076	0.0580	0.0480
Chadorma-lu	0.0017	0.0208	0.0328	0.0023	0.0327	0.0378	0.0011	0.0149	0.0269	0.0029	0.0340	0.0330
Tehran Cement	0.0008	0.0088	0.0184	0.0024	0.0241	0.0300	0.0011	0.0170	0.0249	0.0022	0.0158	0.0311
Atieh Dade Pardaz	0.0029	0.0574	0.0412	0.0047	0.0896	0.0516	0.0013	0.0231	0.0227	0.0028	0.0306	0.0329
Saderat Bank	0.0076	0.0229	0.0447	0.0015	0.0159	0.0306	0.0004	0.0153	0.0119	0.0014	0.0258	0.0161
Zamyad	0.0033	0.0513	0.0449	0.0046	0.0644	0.0534	0.0001	0.0025	0.0087	0.0010	0.0038	0.0162
S*Tejarat Bank	0.0013	0.0112	0.0291	0.0012	0.0127	0.0262	0.0027	0.0555	0.0441	0.0017	0.0256	0.0232
S*Mellat Bank	0.0010	0.0123	0.0268	0.0032	0.0281	0.0368	0.0019	0.0482	0.0316	0.0021	0.0135	0.0280
Inf. Ser-vices	0.0008	0.0069	0.0210	0.0113	0.0910	0.0431	0.0011	0.0187	0.0233	0.0047	0.0524	0.0303
Shahrood N.E	0.0008	0.0252	0.0231	0.0023	0.0524	0.0408	0.0014	0.0161	0.0292	0.0013	0.0148	0.0277
Pars Int. Mfg.	0.0015	0.0255	0.0297	0.0043	0.0426	0.0430	0.0011	0.0183	0.0229	0.0031	0.0360	0.0336
Shazand Petr.	0.0024	0.0324	0.0381	0.0027	0.0335	0.0418	0.0012	0.0222	0.0202	0.0054	0.0573	0.0304
Behran Oil	0.0028	0.0377	0.0365	0.0026	0.0313	0.0360	0.0009	0.0144	0.0194	0.0015	0.0335	0.0299
Mean devi-ations	0.0027	0.0447	0.0356	0.0047	0.0663	0.0457	0.0013	0.0207	0.0235	0.0026	0.0292	0.0287

The last row of Table 2 shows the average deviations of the forecasted values from the real data for the twenty stocks. Looking closer to the last row of Table 2, we observe that the performance of the SVR algorithm in forecasting the price trend of a stock is better than the ANN algorithm.

5.3 Building the Optimal Portfolio

In this section, we describe Phase II of the two-phase approach proposed for determining the optimal stock portfolio regarding the risk and the return simultaneously. To do so, first, we create an optimal portfolio using the real data of the problem (the close prices of the selected stocks mentioned in Table 1 in the first quarter of 2019) using the mathematical methods of solving the quadratic programming problem. To simplify, considering the new risk aversion parameter $\lambda\epsilon[0,1]$, the model can be described as one objective function:

$$\min \lambda\left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}\right] - (1-\lambda)\left[\sum_{i=1}^{N}\mu_i\, w_i\right]$$
$$s.t \qquad \sum_{i=1}^{N} w_i = 1 \tag{13}$$
$$w_i \geq 0 \qquad i = 1,2,\ldots,N.$$

When λ is 0, the model maximizes the mean return of the portfolio, regardless of the variance (risk). In contrast, when λ equals unity, the model minimizes the risk of the portfolio regardless of the mean return. So, the sensitivity of the investor to the risk increases as λ increases from 0 to unity, while it decreases as λ approaches 0. Each case, with different value of λ, would have a different value for the objective function, which is composed of the mean value and variance (risk). Tracing the mean return and variance intersections with different values of λ, we can draw an efficient frontier. Since each point on an efficient frontier curve indicates an optimum, this indicates that the portfolio optimization problem is a multi-objective optimization problem. The introduction of the parameter λ makes the problem into a single-objective function problem. Since the mathematical method solves the problem in an accurate manner, and in each implementation the same efficient frontier is obtained, the efficient frontier obtained from this method can be used as a basis for comparing the following two approaches.

- **a) Markowitz's approach (Approach 1):** This approach only uses the past data related to the stocks to form the optimal portfolio (a retrospective approach).
- **b) The two-phase approach (Approach 2):** This approach uses the predicted data related to the stocks to form the optimal portfolio (a futuristic approach).

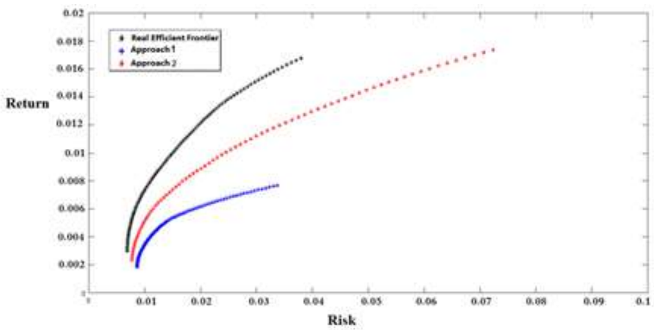

Fig. 2: The efficient frontiers obtained from approaches 1 and 2, and the actual efficient frontier (2019).

Now, to compare the two mentioned approaches, we use the data points (namely, the close prices of the stocks) for a one-year period to form an optimal portfolio for the next year via the MOPSO algorithm. The superiority of an approach is measured by the distance between the approximated points (the points obtained from approaches 1 or 2) and the points on the actual efficient frontier. As shown in Figure 2, the performance of the two-phase approach (depicted in red) is better than Approach 1 (Markowitz's approach, depicted in blue) regarding both convergence (to the real Pareto frontier, depicted in black) and diversity. Now, to confirm and evaluate the results more accurately, in what follows we apply approaches 1 and 2 to the data points in 2019 and form the stock portfolio in 2020.

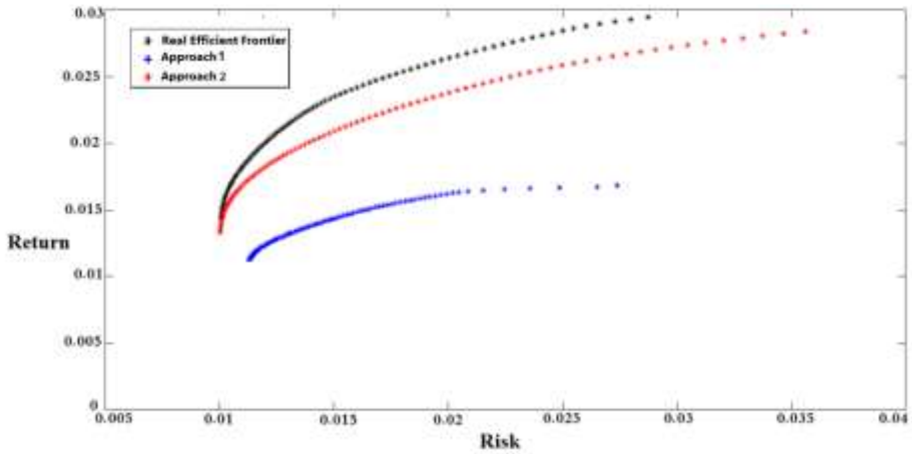

Figure. 3: The efficient frontiers obtained from approaches 1 and 2, and the actual efficient frontier (2020).

As illustrated in Figure 3, like Figure 2, Approach 2 outperforms Approach 1 regarding diversity and convergence points of view. Even more important, at some points, the efficient frontier of Approach 2 coincides with the actual efficient frontier of the problem. Now, we use the *inverted generational distance* (IGD) *indicator* to quantitatively compare two distinct Pareto frontiers which we obtained from the two different approaches. Suppose that M^* is the set of non-dominated points on the actual efficient frontier (the reference data set) of the problem, and that M is the set of non-dominated solutions on the approximate efficient frontier obtained from an approach. The IGD indicator between the sets M^* and M is calculated using the following equation [13].

$$IGD(M,M^*) = \frac{\sum_{\upsilon \epsilon M^*} d(\upsilon,M)}{|M^*|}. \quad (14)$$

Here, $d(\upsilon,M)$ is the Euclidean distance between υ and the points M and $|M^*|$ is the cardinality of M^*. If $|M^*|$ is large enough to provide a good estimate of the Pareto frontier, then the IGD indicator can simultaneously represent both convergence and diversity. Obviously, the smaller the IGD value, the closer the approximate Pareto frontier to the real one.

Table 3: The Values of the IGD Indicator for Approaches 1 and 2

Approach	IGD in 2019	IGD in 2020
Approach 1	0.005	0.125
Approach 2	0.0026	0.0015

It can be seen that the IGD indicator of the two-phase approach (Approach 2) is less than that of

Markowitz's approach (Approach 1) in both 2019 and 2020. Looking closer, we observe that Approach 2 improves the Pareto frontier in comparison with Approach 1 about 48% and 98.8% in 2019 and 2020, respectively, regarding the diversity and the convergence.

Since each non-dominated solution on the efficient frontier represents an optimal portfolio that contains different stocks, it is clear that the investor should choose a solution (or solutions) considering his conditions. In other words, low-risk investors choose portfolios to the left of the efficient frontier (the points closer to the origin). Note that these portfolios have lower risks and of course lower expected returns. In contrast, high-risk investors choose portfolios to the right of the efficient frontier (points farther from the origin), because these portfolios have higher expected returns (and, of course, higher risks).

6 Conclusion

In stock investing, choosing the optimal portfolio is one of the most important issues. In this regard, the creation of models which lead to the selection of the best portfolio is of great importance. The studies conducted so far with the aim of determining the optimal portfolio are based solely on the price trends of stocks in the past (retrospective approaches). In this paper, a new two-phase approach was designed to determine the optimal portfolio by predicting the price trends of different stocks (a futuristic approach). Also, in order to obtain better results, the Hurst exponent filter was used for the appropriate initial selection of stable stocks, as part of the research innovations compared to the previously performed studies. In the proposed approach, we first used the Hurst exponent as a filter to identify stable stocks, and then used the SVR algorithm to predict the price trends of the stable stocks. In the second phase, according to the predicted price trend of each stable stock and using the return and risk measures as well as implementation of the multi-objective particle swarm optimization (MOPSO) as a meta-heuristic algorithm, the optimal Pareto portfolios and the optimal weight corresponding to each stock in that portfolio were determined. The numerical results obtained from portfolio optimization using the proposed two-phase approach indicate that this new approach outperforms Markowitz's approach (as a retrospective approach) in the approximation of the efficient frontier and in terms of diversity and convergence. One of the most important and complex issues in the field of investment is the duration of investment on each stock. The present study enabled the investors to create a future portfolio according to their desired risk by changing the minimum acceptable value of the Hurst exponent. Currently, the authors are working on the relationship between the position of each portfolio on the efficient frontier and the stock information in that portfolio.

References

[1] Alahmari, S., *Predicting the Price of Cryptocurrency using Support Vector Regression Methods*, Journal of Mechanics of Continua and Mathematical Sciences, 2020, **15**(4), P. 313-322.

[2] Asgharpur, H., Rezazadeh, A., *Determining the Stock Optimal Portfolio using Value at Risk*, Journal of Applied Theories of Economics, 2016, **2**(4), P.93-118, (in Persian).

[3] Ballestero, E., *Mean-Semi-Variance Efficient Frontier: A Downside Risk Model for Portfolio Selection*, Applied Mathematical Finance, 2005, **12**(1), P. 1-15.

[4] Bernardo, J. A., Rui Ferreira, N., Nuno, H., *Combining Support Vector Machine with Genetic Algorithms to optimize investments in Forex markets with high leverage*, Applied Soft Computing, 2018, **64**(2), P. 596–613.

[5] Coello, A. C., Pulido, G. T., Lechuga, M. S., *Handling multiple objectives with particle swarm optimization*, In IEEE transactions on evolutionary computation, 2004, **8**(3), P. 256-279.

[6] Das, S., Arman, M. S., Hossain, S.S., Islam, S., *Bangladeshi Stock Price Prediction and Analysis with Potent Machine Learning Approaches*, Cyber Security and Computer Science, 2020, **325**(3), P. 230-240.

[7] Fernandez, A., Gomez, S., *Portfolio selection using neural networks*, Computers operation Research, 2007, **34**(4), P. 1177-1191.

[8] Ghasemi, H. R., Najafi, A. A., *Portfolio Optimization in terms of Justifiability Short Selling and Some Market Practical Constraints*, Financial Research Journal**,** 2014, **14**(2), P. 117-132, (in Persian).

[9] Gilli, M., Kellezi, E., *The Threshold Accepting Heuristic for Index Tracking*, Financial Engineering, E-commerce and Supply Chain, 2001, **70**(2), P. 1-18.

[10] Guang-Feng, D., Woo-Tsong, L., *Ant Colony Optimization for Markowitz Mean-Variance Portfolio Mode*, Swarm, Evolutionary and Memetic Computing Lecture Notes in Computer Science, 2010, **6466**(2), P. 238-245.

[11] Kara, Y., Boyacioglu, M. A., Baykan, O. K., *Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul Stock Exchange*, Expert Systems with Applications, 2011, **38**(5), P. 5311-5319.

[12] Kennedy, J., Eberhart, R., *A New Optimizer Using Particle Swarm Theory*, In Sixth international symposium on micro machine and human scienc, 1995, P. 39-43.

[13] Kim, K., *Financial time series forecasting using support vector machines*, Neuro-computing, 2003, **55**(1), P. 307-319.

[14] Li G.Z., Huang J.B. & Huang H.Y. *Calculating method of contraction operators in fractal interpolation based on the B-spline*, Journal of Ordnance Engineering College, 2006, **18**(2), P.76-78.

[15] Mansini, R., Speranza, M. G., *Heuristic Algorithms for the Portfolio Selection Problem with Minimum Transaction Lots,* European Journal of Operational Research, 1999, **114**(2), P. 219–233.

[16] Maringer, D., *Portfolio Management with Heuristic Optimization*, Advances in Computational Management Science, Published by Springer-Verlag, Berlin, Heidelberg, 2006, P. 1–237.

[17] Markowitz, H. M., *Portfolio Selection, The Journal of Finance*, 1952, **7**(1), P. 77-91.

[18] Reyes Sierra M., Coello, C., *Multi Objective Particle Swarm Optimizers: A Survey of the State of the Art*, International Journal of Computational Intelligence Research, 2006, **2**(3), P.287-308.

[19] Sahala, A. P., Hertono, G. F., Handari, B. D., *Implementation of improved quick artificial Bee Colony Algorithm on portfolio optimization problems with constraints*, Proceedings of the 5th International Symposium on Current Progress in Mathematics and Sciences, 2020, **2242**(1), P.1–8.

[20] Ünal, A., Kayakutlu, G., *Multi-objective particle swarm optimization with random immigrants*, Complex & Intelligent Systems, 2020, **6**(2), P. 635–650.

[21] Vasiani, V., Handari, B.D., Hertono, G.F., *Stock portfolio optimization using priority index and genetic algorithm*, Journal of Physics: Conference Series, 2020, **1442**, P. 1-5.

[22] Xu Jun., Effectiveness of the Securities Market Analysis Efficient Market Hypothesis Testing of Shanghai Stock Exchange by the Method of Rescale Range. Wuhan University, 2004.

[23] Yoosefzade, H.R., Karrabi, A., Heydari, A., *Fracsion: New Hybrid Algorithm Predicting the Trend of Tehran Stock Exchange Industries Index*, Journal of Mathematical Researches, 2021, Accepted paper (in Persian).