

RESEARCH ON STOCK PORTFOLIO BASED ON TIME SERIES PREDICTION AND MULTI-OBJECTIVE OPTIMIZATION

Gholamreza Zandi, Rezvan Torabi, Mohammad Amin Mohammad, and Lu Jia

ABSTRACT. Capital markets are characterized by uncertainty. Investors intend to obtain satisfactory investment return; they need to employ a feasible tool-portfolio. Therefore, portfolio theory is an important part of modern finance. However, many traditional models derived from the mean-variance framework focus on the purpose of risk diversification and overlooked predictions of stock prices and market trends. Moreover, the portfolio problem is a non-linear optimization problem. But most traditional analysis methods and models are linear and single objective models. Nonlinear models that can provide external inputs and methods that can quickly search for Pareto optimality can solve the above problems. Fifty stocks from the S&P 500 are randomly selected for this study. The period of the collected samples is from January 1, 2000, to December 31, 2018. Stock prices, S&P 500 indexes, 13 macroeconomic factors and 9 micro factors are employed as inputs to forecasts. First, the influential factors strongly related to the stock price are found by the bivariate correlation analysis. Second, the factors and stock price are inputted to predict the future stock price by NARX model. Stocks with high forecast accuracy are used to form four portfolios. Finally, positive, and negative probabilities and stock returns are employed as objective functions. Pareto optimality of assets allocation is found by GA for multi-objective optimization. The method of combining nonlinear autoregressive exogenous model (NARX) with genetic algorithm (GA) is proved to be effective in making up for the shortcomings of traditional methods. The returns of four portfolios constructed by the methods are higher than the market returns, which is verified by the real data of the quarters of 2018. Moreover, it is worth noting that the univariate model that only enter one macro factor without considering micro factor has less error. In future research, the sample size can be expanded to further improve its effectiveness and reliability.

2020 *Mathematics Subject Classification.* 00A69.

Key words and phrases. Stock Portfolio, Series Prediction, Multi-Objective, Optimization.

Submitted: 26.02.2021; *Accepted:* 09.03.2021; *Published:* 21.03.2021.

1. INTRODUCTION

As is well-known, the success of stocks investment depends on the prediction and judgment of the future trend of stocks prices. Therefore, both individual investors and institutional investors need to find the right forecasting method for securities investment. However, the price of stocks is affected by many factors, the randomness and the volatility are violent, the law of the trend bears the obvious characteristics of non-linear and the uncertainty. Therefore, it is difficult for an investment forecast. For example, nowadays the main methods of investment forecasting contain fundamental analysis, technical analysis, and other forecasting methods. Fundamental analysis method mainly analyses the domestic and international macroeconomic situation, the economic cycle, industry cycle, extent of industry prosperity, market structure, enterprise operating condition, enterprise sales ability, profitability and corporate governance of companies (such as economy, finance, management analysis), thus the prediction of the future business conditions for enterprises. Nevertheless, as previously mentioned, the influencing factors of the price of the stock are varied and complicated, corporate fundamentals are not the sole factor that affects securities market prices. Prediction of price purely relies on analysing fundamental, which is not accurate. Particularly for short-term prediction, that impact of market sentiment and investor psychology would outweigh the fundamentals (Gurav & Sidnal, 2018). Technical analysis method mainly includes k-line, MACD, KDJ, RSI, BOLL, and other indicators, as well as Gann, trend, morphology, and other line methods. Technical analysis prediction method based on actual data, meanwhile, these analysis methods contains the fundamentals and the influence of other factors. However, its disadvantage is that the method itself depends on inventors' experience with stronger subjectivity. Besides, the different markets employ diverse parameters, so that the application of these figure line or indicators depends largely on the investor's subjective judgment [12]. In recent years, a new investment forecasting and trading method, quantitative investment method, has emerged in the securities market of developed countries. Quantitative investment forecasting method is the method based on mathematical models such as mathematics, statistics, and econometrics. Historical data are collected, processed, analysed, modelled, and then subjected to rigorous testing and simulation, finally being used to predict, judge and even trade for the future securities

market [18]. Because the workload of collecting and processing information involved in quantitative investment is too huge to the human brain, therefore, information technology is also one of the foundations of quantitative investment. The computer algorithm technology applied to the conventional data model processing can only deal with the problem of the linear system, and the problem of the nonlinear system such as the securities market is not effective. With the development of artificial intelligence science and technology, people begin to explore the method of artificial intelligence to solve the problem of the complicated nonlinear system. Among them, the neural network is a method of simulating brain neurons through computer software or electronic circuits and obtaining the approximation of complicated nonlinear functions through parallel distributed processing. It has the characteristics of nonlinearity, fuzziness, self-learning, and self-adaptation. These characteristics make it possible to use artificial neural network models to research and explore the prediction of securities market investment [18].

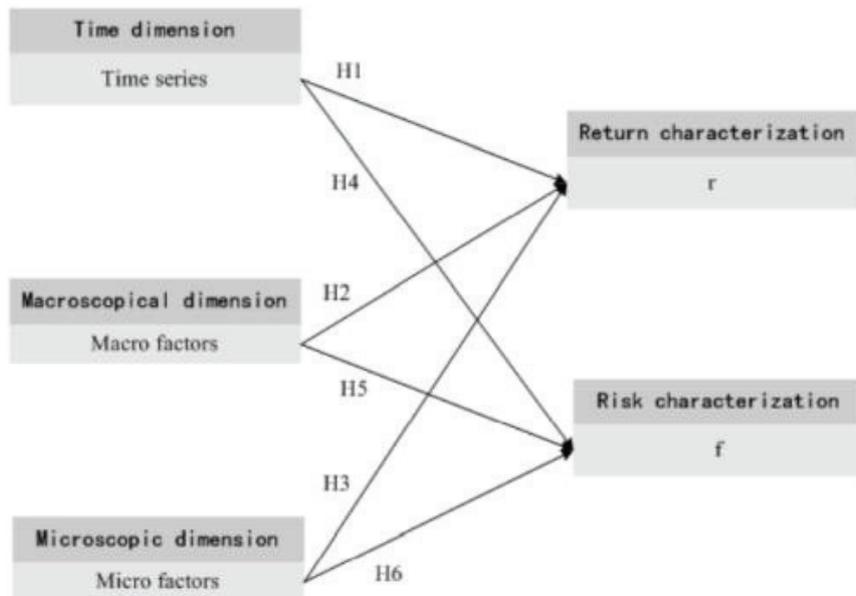
2. LITERATURE REVIEW

Introduction

Efficient markets hypo-study, a major branch of modern portfolio theory, argues that the only way investors can earn higher returns is by taking more risk. According to the hypo-study, in an efficient market, the price of assets reflects all relevant, available, known information, and the market's consistent expectation of unknown information. The Efficient Markets hypo-study (EMH) was advanced and deepened in 1970 by Eugene Fama. The "efficient markets hypo-study" has its roots in the early 20th century, founded by a French mathematician named Louis Bachelor. The hypo-study states that in an efficient capital market, the price of a security changes rapidly in response to new information received. For half a century, economists have been studying whether certain capital markets are efficient. These results of the research are of great practical significance to investors. Up to now, there has been a great deal of theoretical and empirical literature discussing the relationship between risk and expected rate of return based on macroeconomic interpretation. Meanwhile, some characteristics of the security sample can also be used to specify risk indicators in

microeconomic theory. Empirical literature shows that many financial indicators are related to the return rate of common stock, such as enterprise size, earnings/price, cash flow/price, book-to-market ratio, etc. The literature on the determinants of the stock market is divided into two categories: one is about the impact of macroeconomic factors on the stock market; the other is about the impact of microeconomic factors on the stock market. This paper studies both categories. According to the research results, a macro and a micro influencing factor with significant influence on each component stock, as well as the stock closing price, were taken as the input of NARX model, and finally, the predicted value of the model was optimized and combined by genetic algorithm. The dynamic interaction between macro and microeconomic factors and stock prices as well as the research progress of related research methods.

FIGURE 1. Conceptual Framework



H1, H2, H3, H4, H5, and H6 means that r and f have correlations with time, macro, factors, and micro factors, respectively.

Macroeconomic Factors

The stock market allows financial resources to be redistributed among various economic entities. Hence the stock market is considered a common component of the economy. When stocks are employed by governments and companies,

they may acquire the necessary financial approaches and facilities. Other economic entities can invest their savings in these economic sectors because these areas of the economy are reliable and expected to be profitable. Researchers have long been concerned about the correlation between the development of the stock market and changes in the country's economy: when the country's economy is booming, the stock market is active. The performance of the stock market reveals a lot the state of a country's economy: a falling stock price indicates a depression, whereas a rising stock price indicates an improvement. The relationship between macroeconomic indicators and stock prices has been proved in most academic kinds of literature. As the conclusion of King (1966) study, prices of stock suffered the effect of macroeconomic factors by an average of 50% [9]. However, there is a lack of comprehensive evaluation of the causal relationship and dependence between macroeconomic indicators and the stock market in the course of time and changes in the macroeconomic process. This is why the impact of the macroeconomic indicators of NARX model on stock returns makes it a logical extension of academic analysis to reveal the complex causal relationship and to rely on the evaluation of the long-term and = short-term relationship between macroeconomic indicators and stock prices. [9]

Definition of Macroeconomic Factors

Macroeconomic factors are influential fiscal, natural, or geopolitical events that broadly affect the economy of a region or country and these factors are means of reflecting economic conditions. Macroeconomic factors affect a wide range of people, rather than just a few specific individuals. Macroeconomic factors include GDP, inflation and deflation, investment indicators, consumption, economic output, unemployment rate etc. These economic performance indicators are closely monitored by the government, enterprises, and consumers, and play an important role in macroeconomic control analysis and reference.

Financial Ratios

There is a large amount of information related to investment decisions in financial statements obtained through analysis of financial statements. One of the objectives of the analysis is to evaluate the value of the company from the data in the financial statements. To improve the value of financial statement data, many empirical accounting studies have been mining accounting attributes related to corporate value. The study starts with the assumption that market

prices are sufficient to determine the value of a company and assign that as a benchmark for evaluation. Accounting attributes are statistically correlated with stock prices; therefore, it is inferred that they are related to value. As the earlier study has shown, financial ratio analysis could help investors in making an investment decision and predicting the firm's future performance. It can also give early warning about the slowdown of the firm's financial condition [11]. However, the traditional fundamental analysis puts forward a different point of view. The information in the financial statements represents the basic value of the company. Nonetheless, stock prices often deviate from these measures, sometimes showing only a slow trend towards fundamental value. As a result, the scholars argue that the analysis of companies announced financial statements does not reveal the value reflected in stock prices. The "intrinsic value" analysed in the financial statements does not take price as the benchmark for measuring value but takes it as the benchmark for comparing prices. Determine from the comparison whether the stock pricing is too high or too low. Since deviating prices eventually move closer to certain indicators of fundamentals, the results of comparing prices with these fundamental values can be used as a reference for "abnormal return" investment strategies. In this study, some commonly used financial indicators are analysed by bivariate correlation analysis, as well as one macro and one micro factor are selected according to the analysis results. It is adopted as the independent variables to introduce the NARX model. This approach combines the traditional fundamental analysis programme of identifying attributes related to the company's value from the financial statements. The results show that the correlation measure has a great contribution to improving the accuracy of the model for future stock returns. Furthermore, accurate prediction of future returns is the basis of developing a reasonable portfolio.

Rate of Return

The predicted future rate of return of stocks is one of the important factors that determine investors' investment strategies. However, whether the rate of return on the stock can be predicted, the predictability of rate of returns is generally explained by scholars in two areas: 1) some form of general or limited irrationality, such as fads, speculative bubbles or noisy trading or 2) some form of general equilibrium model that provides the real rate of return over time. Although there is considerable literature on "irrational" selection [13] or [15],

they do not pay attention to the work of providing explanations in an efficient and empirically robust data mining model. It must be emphasized that, as Fama and French (1988) and others have pointed out, predictability is not necessarily inconsistent with the concept of market efficiency in the context of an inter-temporal model [5]. Therefore, this study intends to prove that under the efficient market framework, stock prices need not follow a random walk, and the changes in stock returns can predict future returns.

Definition of Rate of Return

In finance, the rate of return is the return on the investment. It includes any change in the value of the investment, and/or the cash flow the investor receives from the investment, such as interest payments or dividends. It can be measured in absolute terms (such as dollars) or as a percentage of the amount invested. The latter is also known as holding period return. Assuming the investment amount is greater than zero, the loss or non-profit is described as a negative return. The rate of return is the profit of an investment over a period, expressed as a percentage of the original investment. In the study, time-frequency is measured in quarters.

Portfolio Selection Theory

In modern financial theory, portfolio selection theory occupies a very important position. It together with capital asset pricing theory, Modigliani-Miller theorem, Black-Scholes option pricing theory, arbitrage pricing theory and efficient market hypothesis constitute the cornerstone of modern financial theory. After half a century of development of the portfolio theory pioneered in the 1950s, theoretical research has yielded fruitful results, which have been widely employed in practice. The establishment and development of the portfolio theory have been through the following stages:

Portfolio Selection Theory of Variance Correction

In 1952, Markowitz's paper Portfolio Selection published in The Journal of Finance, which laid the foundation of portfolio theory and marked the beginning of modern portfolio theory. First, Markowitz's classic portfolio is built on a set of assumptions that fall into four categories: (1) There is no transaction costs and taxes. Asset markets are frictionless. Market liquidity is sufficient. (2) The influence of background risk and investors' liabilities on investors' wealth is not considered. (3) Investors prefer expected utility. (4) Information is free and

can flow freely. Because these assumptions do not accord with the real financial market, it is difficult to apply portfolio theory to practice. Therefore, many scholars make the model more realistic by relaxing the hypothesis. Markowitz's mean-variance model demonstrates both the advantages and disadvantages of diversification. For example, one of the requirements of this model is that the return rate of securities must follow the normal distribution, and then the variance is employed to measure the investment risk. However, in the real security market, this condition is not generally satisfied. Besides, the Markowitz mean-variance model is very computation-intensive for solving large-scale portfolio. Despite its shortcomings, the advent of Markowitz's portfolio theory has led to the rapid development of modern economics. Sharpe presents a single factor model for portfolio selection to reduce the computational burden of model parameter estimation, the model still belongs to mean-variance analysis. This model highlights the importance of return characterization in portfolio selection modelling and is consistent with CAPM and APT in form. Recently, Konno et al have shown that the combined factor model and linear programming model can effectively deal with large-scale portfolio problems.

Bayesian Portfolio Theory

The future distribution of variables is uncertain, and the models and parameters used to describe the distribution of variables are unknown due to incomplete information. Therefore, Brennan, Chordia, & Subrahmanyam, (1998) and Xia, & Lemey, (2009) studied the dynamic asset allocation problem with uncertain mean values, starting from the two cases of independent and identical distribution of income and predictable returns [3] [16]. The results show that when there are uncertain parameters, there will be parametric uncertain risk aversion requirements. Yang, Kim, & Ryu, (2018) applied the structural learning method of Bayesian network to find the input variables of target variables based on 488 stocks in S&P500 [17].

Behavioural Portfolio Theory

Kahneman and Tversky founded prospect theory that investors are not completely rational. They sometimes have overconfidence, loss aversion, herd behaviour and so on. Everyone is risk-averse, and everyone is an adventurer. Das, Schneider, Chen, & Smith, (2010) argued that Markowitz's mean-variance theory (MVT) and Shefrin and Statman's behavioural portfolio (BPT) had the same

goal, and they mathematically integrated MVT and BPT into a new mental account model (MA) to prove that the risk management models of MVT, MA and VaR are mathematically equivalent. Nonetheless, in the financial field, there is still a big gap between practice and theoretical research [4]. How to apply theoretical research to practice is an urgent problem to be solved.

Dynamic Neural Network

In addition to being widely used in pattern recognition and intelligent control, the neural network is also used in stock forecasting. P-werbos first applied the artificial neural network model to stock forecasting. Subsequently, the neural network is widely used in the prediction of time series. Neural network models can be divided into a static neural network and dynamic neural network according to different signal transmission processes. Dynamic neural network refers to a neural network with feedback, whether local feedback or global feedback. Through feedback, the neural network can retain the data of the previous moment and add it into the calculation of the data of the next moment, which makes the network not only dynamic but also more complete with the reserved system information. It may have the dynamic characteristics of the network with one or more orders of time-lag delay. The study of dynamic neural network originated from the global feedback neural network proposed by American scientist Hopfield (1982) [6]. On this basis, many scholars put forward their own dynamic neural network models through different signal feedback modes, such as (Bayliss, Jordan, LeMesurier, & Turkel, 1986) compared the corresponding algorithms and theories of these different dynamic network models, as well as their advantages and disadvantages [1]. Wang (2019) studied how neural networks in stock forecasting can reduce the complexity of neural network programs and the computational burden of the computer method. The dynamic neural network allows the output signal to continue to enter the input as feedback to participating in the next iteration of training, and it also can remember the previous output or previous output results, so it has great advantages in processing complex dynamic mapping, especially time-series processing. Roman J. and Jameel A (1996) used dynamic neural networks to model historical trading data for stock exchanges in the United States, United Kingdom, Japan, Canada, and Hong Kong, and to predict the next annual return on investment. Borg, A. (2007) [2] used the Elman neural network to study the S&P500 index

and found that the time series was not random walk and could be predicted by a neural network. Huang, Nakamori, & Wang, (2005) used the Elman artificial neural network to predict the stock market and achieved good results [7]. Jepsen, Solum, McEvelly, Kim, & Rosenfeld, (2007) combined the dynamic neural network and genetic algorithm to predict the stock market, and the results showed that this model had better prediction effect [8]. Ko, P. & Lin, (2008) used a dynamic neural network for portfolio adjustment and selection [10]. Shi, Ouyang, Huang, Yang, & Chen, (2008) used the Elman neural network model to forecast the Chinese stock market, which proved that the network has good approximation ability and prediction ability [14]. Wang, Liang, Shi, Li and Han (2006) introduced the return factor of time-based on Elman neural network to predict China's stock market index and found that applying Elman model to stock market prediction could significantly improve the return rate of investors. Yao, Zhu, & Bai, (2006) modelled two stocks using Elman neural network, and the results show that dynamic artificial neural network is effective in predicting the Chinese securities market [19].

Hypotheses of Bivariate Correlation Analysis

$H_0 : \rho = 0$, there is no linear correlation between the two variables.

$H_1 : \rho \neq 0$, there is a linear correlation between the two variables.

3. RESEARCH METHODOLOGY

Introduction

Investors must consider which companies to invest in and which stocks have investment potential, to achieve a satisfactory return on investment. Generally, the analysis of the stock market is divided into fundamental analysis and technical analysis. Many investors would ignore the fundamental analysis. In fact, the fundamental is the most useful information that can reflect the potential of stock investment. The fundamental analysis of stock includes macro analysis, medium analysis, and microanalysis. Macro analysis refers to the analysis of national economy, politics and culture, microanalysis refers to the analysis of companies, and medium analysis refers to the analysis of industries and regions. The fundamental analysis of the stock mainly includes four aspects from the micro-level: one is what industry the listed company is in and whether the

industry is encouraged, supported or restricted by policies; Second, the listed company's operating status, operating income, profit level and other financial indicators, year-on-year data is up or down; The third is to analyse whether the listed company has development potential; Fourth, the analysis of stock price fluctuations, whether there are some institutions often pay attention to the main capital and enter it. Therefore, the most important thing is to analyse and evaluate each company and select real stocks with intrinsic value as investment targets. The evaluation of the financial situation of the company is conducted through multi-index analysis, so the first step is to conduct dimensionality reduction processing on the multi-index, and find the decisive and important index as the analysis index, which is crucial to our research. In mathematics, this is a problem of classification and processing of large sample data in high dimensionality, the basic idea of which is to reduce the dimension of the data or find a way to deal with the high-dimensional data. There are also many methods of data dimensionality reduction, such as principal component analysis and the promotion of principal component analysis. However, bivariate correlation analysis was used to keep the data original in this study. Based on the pre-analysis of the above steps, data mining technology and multi-objective optimization method are applied to achieve the research objective.

Research Purpose

Recently, "big data" and "data mining" gradually appear in people's vision. Big data, as its name implies, is data with a large amount of data and high dimension. Big data provides scholars with a lot of research evidence. Scholars no longer have to worry about statistical analysis producing meaningless results, because as we all know, statistical analysis is based on large sample data. Data mining is the process of searching out the information with special relationship hidden in a large amount of data. In general, it can be divided into three steps: data preparation, searching for rules and expressing rules. The tasks of data mining mainly include clustering analysis, evolution analysis, association analysis, etc. With the help of the power of the computer, machine learning, pattern recognition and other methods are used to accurately predict the rate of return and risk of particular stocks. Then, the genetic algorithm is used to realize the multi-objective optimization solution to satisfy the requirements of an investor with minimize risk and maximize return.

Research Approach

The simplest and direct method to analyse stocks is to research the rate of return series. In the study, 50 component stocks of S&P500 are randomly selected as the research samples, covering 11 sectors. The second step is to use a bivariate correlation analysis method to find out the most effective macro and micro factors to predict the return rate of each stock. Third, the NARX model is used to predict the stock's price. Finally, a genetic algorithm is used for multi-objective optimization.

Population and Unit of Analysis

The samples of the study collected from 505 component stocks in the latest list of S&P 500 in 2019. It represents available high-quality large-cap stocks in the United States for investors. At the very least, it means that: (1) the company is the largest in the U.S. by market capitalization. (2) the company must have been profitable in the most recent quarter, and the profitable public company did not enter the index based on its Pre-IPO earnings. (3) the company is very floating and liquid. Companies that are closely held (majority-owned by only a few shareholders) as well as companies that are thinly traded (companies whose shares have very little trading volume), both are excluded from the list; (4) these companies entered the list cautiously. A company must get approval from the index committee to enter the S&P 500, which makes the S&P 500 more "active" than other indices that simply employ mechanical rules to pick stocks. The members of the index committee are among the most influential experts in financial markets, their decisions that could have billions of dollars in consequences for investors. Therefore, they are cautious in their review. Besides, the component stocks of S&P500 were chosen as the research samples, because the financial data of these companies are complete and easy to collect due to the companies must file 10-K annual reports with the SEC. According to the above states, choosing S&P500 as the population and unit of analysis conforms to the relevant assumptions of this study on a portfolio investor, and the sufficiency and integrity of samples are conducive to this study.

Sampling Frame, Technique, Size and Data Collection

The samples selected in this empirical study are S&P500 historical index, 50 component stocks of S&P500, the historical financial statements of a company that issues 50 component stocks and the historical macroeconomic data of the

United States. The period is from January 1st, 2000 to September 30th, 2018. The data downloaded from www.ycharts.com and free.stlouisfed.org. These data range from a closing price of each day, S&P indices, quarterly financial ratios, quarterly macroeconomic indicators. The data of four quarters are employed as empirical data to test the accuracy of the prediction with actual data.

TABLE 1. Financial Ratio

	Micro factors	Number
	Book-to-Market Ratio	1
	Common Shareholder's Equity	2
Profitability	Net Profit Margin	3
	Debt-to-Equity Ratio	4
Liquidity	Current Ratio	5
	Tangible Common Equity Ratio	6
	Receivables Turnover	7
Management Effectiveness Ratio	Return on Equity	8
	Return on Assets	9

One of the objectives of this study is how investors could make use of the huge amount of financial information. Therefore, this study starts from the concern on the intrinsic value of enterprises. The reason why nine financial ratios and one value are used as indicators for quantitative stock selection is that the information they contain can be used to measure the intrinsic value of enterprises and effectively explain the development and change of these values.

Bivariate Correlation Analysis

Bivariate correlation analysis is to study the degree of linear correlation between two or more variables through sample data, reflecting the convergence relationship between two or more variables. Usually follow these steps:

- (1) Making a qualitative judgment on whether there is a correlation between variables.
- (2) Drawing scatter diagram. The degree of scatter density between variables and the future trend can determine and reflect the direction and specific manifestation of the correlation between variables.
- (3) Calculating the correlation coefficient between variables accurately. In practice, the most common is the Pearson correlation coefficient.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

According to the prediction results, the stocks CV < 0.1 and positive predicted return are selected to form four portfolios for four quarters in 2018. The components of each portfolio are indicated in tables 4.10-4.13. Nonetheless, some predicted returns are greater than 10 per cent or 20 per cent. When the stock market is expected to rise (S&P500 index predicates a higher value than the previous period), the predicted returns greater than 20% (if any) will be considered overvalued based on practical experience. Therefore, it will be adjusted to 20%. Conversely, when the stock market is expected to decline (S&P500 index predicates a lower value than the previous period), the predicted returns greater than 10% (if any) will be considered overvalued. Hence, it will be adjusted to 10%. According to the principle of diversification, the manual adjustment process is added to more reasonable allocate weights when multi-objective optimization is performed and to avoid the emergence of super individuals.

TABLE 2. The first portfolio for the first quarter in 2018

No.	Security Symbol	CV	Predicted Return (%)	Adjusted Return (%)
1	AES	0.047897	12.51	10
2	SCHW	0.052256	3.28	3.48
3	BSX	0.055617	14.06	10
4	INTC	0.056981	8.99	8.99
5	CTL	0.064603	31.11	10
6	MRO	0.064782	7.00	7.00
7	AEE	0.067418	4.51	4.51
8	BAX	0.075536	1.97	1.97
9	HON	0.077628	11.42	10
10	MMC	0.088786	0.73	0.73
11	CLX	0.093346	29.03	10
12	DIS	0.097951	17.44	10
Mean Return				7.22

TABLE 3. The second portfolio for the first quarter in 2018

No.	Security Symbol	CV	Predicted Return (%)	Adjusted Return (%)
1	AES	0.042045	13.95	13.95
2	BLL	0.069001	22.86	20
3	BSX	0.089012	9.31	9.31
4	ADP	0.091447	2.08008e-05	2.08008e-05
Mean Return				10.82

Multi-objective Optimization

The solution set of each portfolio is shown in the appendix A.1-A.4. The upper and lower bounds of the domain are set separately, according to the principle of diversification. It can be seen that the CVs are approximately the same size. Furthermore, 60 per cent of the weighting is evenly allocated to each stock in the portfolio. The remaining 40 per cent of the weight are employed for multi-objective optimization. It can be proved that there is no abnormal high weight in each solution set. The positive and negative probabilities are employed as the selection indicators to build the target portfolio. According to the principle of risk minimization, if the predicted market return were negative, the portfolio with the highest positive probability will be selected to protect against losses. If the predicted market rate of return were to rise, the portfolio with the lowest negative probability will be selected to maximize returns.

TABLE 4. Boundary conditions of each portfolio

Quarter	Variables Number	Lower Bound	Upper Bound	Predicted Market Return (\pm)	weight Allocation Method	Remark
1	12	0.05	0.12	-	$[0.6/n, 0.6/n + 0.4/(n/2)]$	
2	4	0.15	0.35	+	$[0.6/n, 0.6/n + 0.4/(n/2)]$	Too few variables
3	11	0.05	0.13	+	$[0.6/n, 0.6/n + 0.4/(n/2)]$	
4	7	0.09	0.20	-	$[0.6/n, 0.6/n + 0.4/(n/2)]$	

Market Trend Prediction

Comparing the results in the two tables, the predicted value of the ARIMA model clearly shows inertia that gradually increases with time and does not respond to the downward trend of the stock market. Therefore, in the case of an incorrect trend forecast, it is meaningless to further compare the accuracy

TABLE 5. Target portfolio for 2018

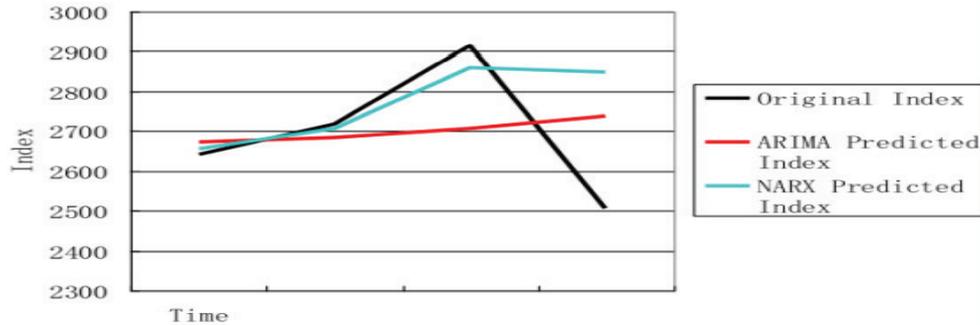
Stock Number	Stock Symbol	Adjusted Predicted Return (%)	Real Return (%)	Weight	Optimized Portfolio Return (%)
1	AES	10	4.99	0.07356	0.74
2	SCHW	3.48	1.65	0.05019	0.17
3	BSX	10	10.21	0.089217	0.89
4	INTC	8.99	12.82	0.077382	0.70
5	CTL	10	-1.5	0.086286	0.86
6	MRO	7.00	-4.73	0.068716	0.48
7	AEE	4.51	-4	0.087542	0.39
8	BAX	1.97	0.62	0.097833	0.19
9	HON	10	-5.77	0.108388	1.08
10	MMC	0.73	1.47	0.050279	0.04
11	CLX	10	-10.51	0.1188	1.19
12	DIS	10	-6.58	0.092643	0.93
Mean/Total		7.22	-0.11	1	7.66

of the prediction. However, the changing trend predicted by the NARX model is the same as the true values, and in the stable state of the stock market, the accuracy of the predicted value is also higher than that of the ARIMA model. In terms of method, compared with the NARX model, ARIMA has the following disadvantages: First, the assumption that the time series data must be a stationary non-white noise sequence does not exist. Second, when using the difference method to smooth the data, subjectively order the parameters and determine the parameters according to the observation charts is rough. Furthermore, each time the difference will lose one original data. Meanwhile, more operating steps have been added. Besides, the variation in the stock market is affected by many external factors, but it is difficult for the ARIMA model to express such a complex relationship.

Discussion and Conclusion

The applicable population of the study findings is identified. Meanwhile, the findings of this study are recommended to investors who are suitable for this method. More importantly, the researcher realized the limitations by analysing

FIGURE 2. Market Trend Prediction Chart



all aspects of this study to lay foundations for further research in this field. Finally, the overall picture of the study is summarized.

Limitation of Research

First, the sample size should be large enough. Additionally, this study selects the monthly frequencies, which is not enough accuracy. However, how much can be considered large enough cannot be clearly defined. Because the sample cannot be infinite, and the stock market does not have the conditions for trial and error. Second, no matter how much historical data is counted, it is impossible to determine which event will occur next time, in the real world. Roughly speaking, it is like the classical principle of inadequate justification in probability theory: all unknown probabilities are equal probabilities.

NARX Model to Predict

First, in the correlation analysis, the selected macro and micro factors are not adequate. This is one of the reasons for the large errors in many prediction results. Second, in the design of NARX model, only several simple combinations of hidden layer and lag order are compared experimentally. That is one reason the number of prediction values is not accurate.

GA to Solve Multi-objective Optimization

First, many parameters need to be set when designing the selection operator, crossover operator and mutation operator. The selected parameters severely affect the quality of the solution. Second, GA possesses a strong dependence on the initial population, which directly affects the convergence of solutions and the quality of optimization results. Third the more iterations a GA has, the better its convergence. However, after increasing the number of genetic iterations,

the computational workload increased. This time-consuming work could cause investors to miss the best time to trade. Fourth, the sample size is too small to build a well-diversified portfolio.

The Findings of the Study

Factors with the strongest correlation to a particular stock price are identified by bivariate correlation analysis. Then, the factors (independent variables) are valid as inputs to the NARX model for predicting stock prices. Finally, Portfolio returns can be increased, and the risks can be reduced by multi-objective optimization.

4. CONCLUSION

The origin of this study is the stock prediction analysis. The function of stock prediction analysis in security investment risk and evaluation of investment value is discussed. Stock predicting analysis is emphasized as the premise and basis for investors to measure their investment risk and evaluate their investment value. Therefore, it is necessary to establish a scientific stock predicting system for effective analysis. The stock market in developed countries has a complex and huge system which is influenced by a variety of uncertainties. This study systematically introduces the technical indicators of stock prediction, the specific design and workflow of NARX model and GA. An additional intention is to explore the strengths and weaknesses of these methods in stock forecasting. This study identified 50 component stocks of S&P500 as the research objects. The correlation between macro, micro factors and stock price is analysed. Scientific stock prediction evaluation indicators system is established. The principles and methods of quantifying and standardizing these indicators are also studied. The NARX model is then employed to predict stock prices. According to the real historical data of sample stocks, the learning samples of NARX-based stock prediction analysis model are constructed, and the analysis and evaluation model is self-learned and trained to make the model suitable for the actual situation. Finally, the effectiveness and reliability of NARX model and GA in stock prediction analysis are proved by case analysis and comparison with the ARIMA model.

REFERENCES

- [1] BAYLISS, A., JORDAN, K. E., LEMESURIER, B. J., & TURKEL, E. (1986: *A fourth-order accurate finite-difference scheme for the computation of elastic waves*, Bulletin of the Seismological Society of America, 76(4), 1115-1132.
- [2] BORG, A. (2007: *Towards a history and typology of color categorization in colloquial Arabic*, EDITED BY ROBERT E. MACLAURY GALINA V. PARAMEL.
- [3] BRENNAN, M. J., CHORDIA, T., & SUBRAHMANYAM, A. (1998: *Alternative factor specifications, security characteristics, and the cross-section of expected stock returns*, Journal of financial Economics, 49(3), 345-373.
- [4] DAS, D., SCHNEIDER, N., CHEN, D., & SMITH, N. A. (2010, JUNE: *Probabilistic frame-semantic parsing. In Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics* (pp. 948-956).
- [5] FAMA, E. F., & FRENCH, K. R. (1988: *Permanent and temporary components of stock prices*, Journal of political Economy, 96(2), 246-273.
- [6] HOPFIELD, J. J. (1982: *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the national academy of sciences, 79(8), 2554-2558.
- [7] HUANG, W., NAKAMORI, Y., & WANG, S. Y. (2005: *Forecasting stock market movement direction with support vector machine*, Computers & operations research, 32(10), 2513-2522.
- [8] JEPSEN, K., SOLUM, D., ZHOU, T., MCEVILLY, R. J., KIM, H. J., GLASS, C. K., ... & ROSENFELD, M. G. (2007: *SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron*, Nature, 450(7168), 415-419.
- [9] KING, B. F. (1966: *Market and industry factors in stock price behavior*, the Journal of Business, 39(1), 139-190.
- [10] KO, P. C., & LIN, P. C. (2008: *Resource allocation neural network in portfolio selection*, Expert Systems with Applications, 35(1-2), 330-337.
- [11] OHLSON, J. A. (1980: *Financial ratios and the probabilistic prediction of bankruptcy*, Journal of accounting research, 109-131.
- [12] PETRUSHEVA, N., & JORDANOSKI, I. (2016: *Comparative analysis between the fundamental and technical analysis of stocks. Journal of Process Management, New Technologies*, 4(2), 26-31.
- [13] POTERBA, J. M., & SUMMERS, L. H. (1988: *Mean reversion in stock prices: Evidence and implications*, Journal of financial economics, 22(1), 27-59.
- [14] SHI, L. H., MIAO, Y. L., OUYANG, Y. C., HUANG, J. C., LEI, Z. L., YANG, J. W., ... & CHEN, D. Y. (2008: *Trichostatin A (TSA) improves the development of rabbit-rabbit intraspecies cloned embryos, but not rabbit-human interspecies cloned embryos*, Developmental dynamics an official publication of the American Association of Anatomists, 237(3), 640-648.

- [15] WEST, K. D. (1988: *Dividend innovations and stock price volatility*. *Econometrica*, Journal of the Econometric Society, 37-61.
- [16] XIA, X., & LEMEY, P. (2009: *Assessing substitution saturation with DAMBE*, The phylogenetic handbook: a practical approach to DNA and protein phylogeny, 2, 615-630.
- [17] YANG, E., KIM, S. H., KIM, M. H., & RYU, D. (2018: *Macroeconomic shocks and stock market returns*, The case of Korea. *Applied Economics*, 50(7), 757-773.
- [18] YANG, X. L., ZHANG, Y. Z., JIANG, R. D., GUO, H., ZHANG, W., LI, B., ... & LI, S. Y. (2017: *Genetically diverse filoviruses in Rousettus and Eonycteris spp. bats, China, 2009 and 2015*, *Emerging infectious diseases*, 23(3), 482.
- [19] YAO, Y., ZHU, J. F., & BAI, M. G. (2006: *Extension Early Warning Model for Irregular Flight Management [J]*, *Operations Research and Management Science*, 1.

UNIVERSITI KUALA LUMPUR (UNIKL) BUSINESS SCHOOL, MALAYSIA.

Email address: zandi@unikl.edu.my

SEGI UNIVERSITY

Email address: torabi.economic@gmail.com

UNIVERSITI KUALA LUMPUR (UNIKL) BUSINESS SCHOOL, MALAYSIA.

Email address: aminmohd8844@gmail.com

SEGI UNIVERSITY