Portfolio optimization for representative stocks in U.S.

Hanyu Dai[1, a, †], Yitong Wang[2, b, †], Tianle Yang[3, *, †]

[1]Faculty of Science, University of British Columbia, Vancouver, Canada

[2]Guanghua School of Management, Peking University, Beijing, China

[3]School of Artificial Intelligence, Jiangnan University, Jiangsu, China

*Corresponding author:
chris_yangtianle@foxmail.com, [a]dhy666666o@gmail.com, [b]m15611337930@163.com

†These authors contributed equally.

Keywords: Portfolio theory; Monte Carlo simulation; ARIMA; Portfolio Optimization

Abstract: With the rapid development of computer technology, quantitative investment is in the ascendant. To combine the fundamental analysis and quantitative investment together, this paper focuses on 10 representative stocks in the U.S. financial markets. The ARIMA model is adopted to forecast the returns. Based on the Modern Portfolio Theory, the Monte Carlo Simulation and the Convex Optimization are used to analyze the corresponding weights of the assets. The results show that the asset of AAPL accounts for the largest proportion in the Maximum Sharpe Ratio, while when the investor is more risk-averse, the asset of CHTR is the optimized choice. We compare the constructed portfolios with the actual market return and the results show that our portfolios beat the market return with a higher accumulated return and a smaller volatility, thus is beneficial to the investors.

1. Introduction

A portfolio constructing process mainly consists of two steps: selecting the risky assets and determining the allocation. Since Markowitz [1] proposed the modern portfolio theory which firstly used the expectation return and the variance to describe the asset performance, the evolution of portfolio investment became enduring and prosperous. It is now commonly acknowledged that investors are faced with a trade-off between the return and the risk. Besides the principle of allocation, the quadratic utility function which depicts the risk aversion level of an investor is also introduced. In 1992, Fisher Black and Robert Litterman [2] modified the MPT theory by using probabilistic statistical methods and combining investors' views on a broad class of assets with market equilibrium returns to generate new expected returns. After the financial crisis in 2008, institutions put more emphasis on the risk factor allocation and came up with the Risk Parity method. In recent years, in-depth research on portfolio optimization has been massively conducted by scholars around the world. More complex models are used to improve the stability of allocation strategy, for example, Chu and Fang [3] disposed the sample covariance matrix based on random matrix theory and got a better strategy. According to the Black-Litterman theory, scholars have been importing new models to improve the strategy. Wen, Chen and Liang [4] used the macroeconomic variables to predict the returns while catching characters of the variety with GJR-GARCH model. Zhu [5] applied the information entropy compensation investment risk analysis methods and the time sequence AR-TGARCH model to establish the optimized portfolio based on the Black-Litterman model.

Value Investing theory was firstly proposed by Benjamin Graham during 1929-1933, which emphasized stock valuation based on financial analysis. The theory pays close attention to the financial statements of the company and by discounting the predicted future cash flows, it then can discover the stocks that are being undervalued and overvalued. In 1995, Feltham and Ohlson [6] came up with an excess return model which considered the discounting of dividends and residual income from asset value, and took the current accounting fundamentals into consideration of the company's future earnings. Berkshire Hathaway, founded by Warren Buffett in 1956, is by far one of the most successful insurance empires with stocks, bonds, cash and silver, as well as many industrial companies. The founder, Warren Buffet has been known as a famous investor and believer of Value Investing and has made many insightful investment strategies which can be regarded as powerful examples of value investing. However, as the information age arrives, constructed on the fundamental analysis, traditional value investing is considered as lacking agility and ineffective.[7]

Quantitative investment is a trading strategy with the statistical analysis as its tool. Based on the Modern Portfolio Theory and the CAPM model [8], quantitative investment has proved its stability in mature capital markets. What's more, under the background of Artificial Intelligence, machine learning algorithms are also utilized in quantitative investment and portfolio optimization. For instance, Qi, Lin and Wang [9] used genetic algorithm to solve the optimized portfolio and compared with the closed form solution obtained by quadratic programming, finding that within a certain range of stock numbers, genetic algorithm is effective and precise. Zhou and Zhao [10] detailed the design of chromosome, and adopted appropriate selection and hybridization to choose the stocks, promoting the genetic algorithm to a specific empirical level. Li [11] adopted Lasso Regression, Ridge Regression and other 10 machine learning algorithms to construct the price-forecasting model. Chen, Zhang and Liu [12] applied convolutional neural network (CNN) and deep learning technology to quantify financial investment, so as to obtain more robust results.

Quanta-mental Investing is the combination of quantitative investment and value investing. Recently, this method began to draw more attention from investors. Although there is a large number of papers focusing on analyzing the advantages and drawbacks of quanta-mental investing, as well as developing the measurement of valuation, for example, Novy-Marx [13] found that compared with net profit indicators, gross profit is a better measurement. Previous literature focused mainly on the general traits of Quanta-mental Investing, while limited research focused on elaborating its detailed process and specific result of certain star-stocks.

This paper simulates a common process of Quanta-mental Investing by selecting 10 star-stocks based on Buffet's choice, which shows the essence of value investing, and then calculates the proper weight for each asset based on Quantitative Investment theory.

By studying the Berkshire Hathaway' s positions, we choose the following stocks as our underlying assets: Apple Inc. (AAPL, NASDAQ), Verizon Communications Inc. (VZ, NYSE), Charter Communications Inc. (CHTR, NASDAQ), Bank of America Corp. (BAC, NYSE), American Express Co. (AXP, NYSE), The Coco-Cola Co. (KO, NYSE), Kraft Heinz Co. (KHC, NASDAQ), Moody's Corp. (MCO, NYSE), U.S. Bancorp (USB, NYSE), DaVita, Inc. (DVA, NYSE). In summary, we construct 4 portfolios. The first one is a mimicking portfolio in which the weights of the 10 chosen stocks are proportional to the original ones in Buffet's holdings, but are scaled up so that the sum of them equals to 1. Next, using Monte Carlo Simulation and portfolio theory as theoretical basis, we construct the efficient frontier and find the Maximum Sharpe Ratio Portfolio and the Minimum Volatility Portfolio. Finally, using cxvpy codes in Python and the quadratic utility function, we build a portfolio which corresponds to a certain risk aversion level.

The results in this paper can be generalized as follows. First, the ARIMA model results indicated that the historical return series is a steady time series and can be applied to our stock prediction; Second, the Statsmodel in Python codes helps to forecast the stock returns; Third, through the Maximum Sharpe Ratio Portfolio the authors point out that AAPL has the largest weight, more attention should be paid to AAPL in this portfolio, while through the Minimum Volatility Portfolio, VZ has the largest weight, more attention should be paid to VZ in the portfolio. Fourth, the constructed portfolios based on the weight we obtained above beat the market.

This paper is constructed as follows. Section 2 is the data and methods. Section 3 shows the results and Section 4 concludes the paper.

2. Data and methods

2.1 Data

Our stock data comes from Yahoo Finance (https://www.yahoo.com), and the time duration is from Jun 1, 2020 to Aug 25, 2021. In addition, the specific data we extract is the daily adjusted closing price, and we then calculate the log return of the stock price, in order to obtain more stability and accuracy. We dropped the null value and replaced them with nearest values to maintain the completeness of our data.

$$Y_t = logP_t - logP_{t-1} = log\frac{P_t}{P_{t-1}} \quad (1)$$

The time range is carefully selected. Since our prediction is short-term based and requires promptness, we choose data in 2-year range so as to avoid historic noises from long-term data. However, the Covid-19 has caused great fluctuations in the US stock market in early 2020, thus we choose June as our starting month, before which the stock price has gradually subsided.

The 10 stocks we choose are 10 assets of Berkshire Hathaway's longest position in 2020. In the paper, they refer to the BRKA assets. (AAPL, BAC, AXP, KO, KHC, MCO, VZ, USB, DVA, CHTR) The companies are chosen due to its excellency and its typical significance of representing the preference and investment strategy of Warren Buffet. The descriptive statistics are shown in Table 1.

Table 1. Descriptive Statistics of return series

	AAPL	BAC	AXP	KO	KHC	MCO	VZ	USB	DVA	CHTR
Mean	0.0013	0.0012	0.0011	0.0004	0.0005	0.0007	0.0001	0.0011	0.0011	0.0008
Std	0.0154	0.0160	0.0149	0.0092	0.0118	0.0115	0.0069	0.0166	0.0139	0.0119
50%	0.0016	0.0017	0.0014	0.0003	0.0012	0.0013	0.0000	0.0002	0.0006	0.0006
75%	0.0082	0.0073	0.0084	0.0046	0.0056	0.0050	0.0032	0.0086	0.0064	0.0051
Max	0.0996	0.0606	0.0689	0.0413	0.0511	0.0592	0.0511	0.0629	0.8279	0.0069

Table 1 shows that the returns on mean of the ten assets are positive, which are 0.0013, 0.0012, 0.0011, 0.0004, 0.0005, 0.0007, 0.0001, 0.0011, 0.0011, 0.0008, respectively. Generally, the return series is stable without abnormal value. AAPL has the largest rally among the ten assets, for about 0.0996. And BAC has the largest drop among the ten assets, for about -0.1058.

2.2 ARIMA Model

The ARIMA model was pioneered by Box and Jenkins and is one of the most popular prediction methods in time series prediction [14]. In this study, we selected ARIMA (autorepression integrated moving average model) as the model for time series prediction. The reason why the ARIMA model is uniformly used is that this model is very simple, only endogenous variables are required without the help of other exogenous variables, and our data are stable time series after the visual test.

ARIMA model has three parameters: P, D and Q. P represents the lag of the time series data used in the prediction model, also known as AR/auto regressive term. D is the number of differences which make it a stationary sequence. Q lags represent the prediction error used in the prediction model, also known as MA / moving average term. Suppose p, q, d is known, ARIMA is expressed mathematically as:

$$\widehat{y_t} = \mu + \emptyset_1 * y_{t-1} + \ldots + \emptyset_p * y_{t-p} + \theta_1 * e_{t-1} + \ldots + \theta_q * e_{t-q} \quad (2)$$

Among them, ϕ represents the coefficient of Autoregressive Model and θ represents the coefficient of Moving Average.

2.3 Monte Carlo Simulation Method

In this study, we used the Monte Carlo Simulation to construct optimized stock portfolios.

Monte Carlo simulation is a method done by modelling and simulating. It is an easy-use application in the study of finance and the economy. It is a solid and flexible simulation way when we discuss realistic data like assets allocation in this study [15]. The difference between Monte Carlo simulation with other methods is that Monte Carlo is in the statistical approach. It depends on the statistical method (sampling and inference). Even the previous study, shows that Monte Carlo is mainly focused on the random phenomenon, it does not mean that this method only is applied to random events or processes [16]. Monte Carlo can help when the integral cannot be analytically evaluated, and these methods are used to get the approximately multivariable integrals.

$$\mu = \int_{\chi} f(x)v(dx) \tag{3}$$

Where $f : x \to \mathbb{R}$is a considerable function, x is a measurable set, v is a probability measure. μ is the weighted integrand average. Also, $\mu = E\ [f(X)]$. The X is a random variable with probability measure v. Monte Carlo method take a weighted average of values of f with a finite data site, $x_1, \ldots, x_n$.

$$\hat{\mu} = \sum_{i=1}^{n} f(x_i)w_i = \int_{\chi} f(x)\hat{v}(dx) \tag{4}$$

The sampling measure, $\hat{v}$, assigns a weight w_i to this function value at x_i and lies in the vector space

$$\mu_s := \left\{\sum_{i=1}^{n} w_i\delta_{x_i} :\ w_1, \ldots, w_n, w_n\ \in\ \mathbb{R}, x_1, \ldots, x_n\ \in x, n\ \in\ \mathbb{N}\ \right\}, \tag{5}$$

Where δ_t is a Dirac measure concentrated at point t. The data sites, the weights and the sample size could be determined or random. [2]]

2.4 Mean-Variance frontier

Portfolio Return

Where n is the number of the assets, w_i is the weight of the i-th asset in the portfolio,R_i is the return of the i-th asset.

$$R_p = \sum_{i=1}^{n} w_i R_i \tag{6}$$

And the variance of the portfolio is,

$$D(R_p) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(R_i, R_j) \tag{7}$$

Where n is the number of the assets, w_i, w_j is the weight of the i-th and j-th asset in the portfolio, R_i, R_j is the return of the i-th and j-th assets.

Sharpe Ratio

The Sharpe ratio (Sharpe 1966) is a measure of return-to-risk that plays an important role in portfolio analysis. It is the most common method that used metrics to analyze the historical performance of financial data, such as funds, stocks, etc.

$$\hat{\zeta} := \frac{\hat{\mu} - r_0}{\hat{\sigma}} \tag{8}$$

That $\hat{\mu}$ is the historical, or sample, mean return of the fund, $\hat{\sigma}$ is the sample standard deviation of returns, and r_0 is some fixed risk-free return [17].

Specifically, a portfolio that maximizes the Sharpe ratio is also the tangency portfolio on the efficient frontier from the mutual fund theorem. The maximum Sharpe ratio portfolio is located on the efficient frontier with the function estimate Max Sharpe Ratio and the dataset object is used to list the assets in this portfolio. In a continuous-time setting, the efficient frontier with the inclusion of a risk-free asset is not tangent to the one with only risky assets. This in turn suggests that a risk-free asset strictly enhances the optimal Sharpe ratio in a continuous-time setting.

The Quadratic Utility Function

The Quadratic Utility Function was firstly proposed by Markowitz in mean-variance analysis. Since every investor has its own preference, that is, investment activities follow a utility function about returns and risks. Among several kinds of utility functions, the quadratic utility function is largely used for its simplicity [18].

$$maxmize: E\big(U(W)\big) = E(W) - \left(\frac{1}{2}\right)\lambda Var(W) \tag{9}$$

Where, W is the return of the underlying asset, thus $E(W)$ is the expected return, representing the returns. $Var(W)$ is the variance of the underlying asset, which represents for risks. λ is the risk aversion coefficient. As λ increases, investors become more and more risk averse.

3. Results

3.1 ARIMA model

We use the statsmodel code in Python, which is based on ARIMA model to forecast the stock returns. We provide 21 steps to predict, and decimal for the alpha argument to specify the confidence intervals. The default setting is 95% confidence. For 99% set alpha equal to 0.01.

Table 2. 5 days forecasting example

Date	Real return	lower_ci_95	lower_ci_99	Upper_ci_95	upper_ci_99
2021-8-20	-0.00165	-0.03141	-0.040777	0.0281211	0.037475
2021-8-21	-0.00087	-0.03066	-0.04002	0.028926	0.038398
2021-8-22	-0.00076	-0.03562	-0.03992	0.029035	0.038397
2021-8-23	-0.00077	-0.03056	-0.03993	0.029030	0.038393
2021-8-24	0.003167	-0.02665	-0.03602	0.032984	0.042353

The ARIMA results shown here are for 5 days ahead forecasting of the asset of AAPL

From Table 2, we can know that the forecasted return is close to the real return. The changes of AAPL of the predicted value in the first three days are consistent with the real return. Also, all forecasted values are within the 95% confidence interval.

3.2 Monte Carlo Simulation

We implement the Monte Carlo simulation, and got the efficient frontier and two certain portfolios, i.e., the Minimum Volatility Portfolio and the Maximize Sharpe Ratio Portfolio, in the following Figure 1.

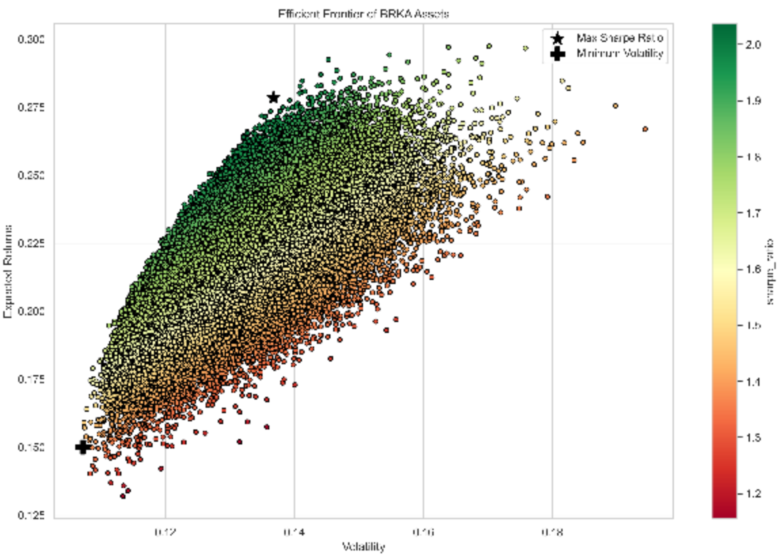

Figure 1. Efficient frontier and the two interested portfolios

By using convex optimization method in Python, we got the other two interested portfolios, i.e., the BRKA portfolio and the Risky portfolio. The BRKA portfolio is a mimicking portfolio in which the weights of the 10 chosen stocks are proportional to the original ones in Buffet's holdings, but are scaled up so that the sum of them equals to 1. The Risky portfolio is an optimized portfolio under the quadratic utility function with λ=1.108, which represents a common and risk aversion level preference.

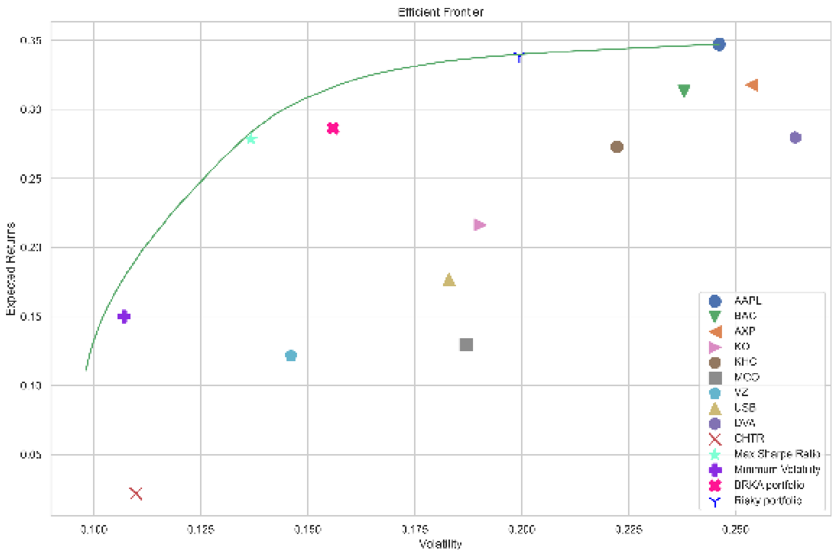

Figure 2. Efficient frontier and the four interested portfolios

Table 3. Portfolio Optimization

	BRKA portfolio	Maximum Sharpe Ratio	Minimum Volatility	Risky portfolio
AAPL	0.4736	0.234	0.0864	0.7470
AXP	0.0977	0.1068	0.0035	0.2530
BAC	0.1624	0.1716	0.0259	0
CHTR	0.0146	0.0273	0.2498	0
DVA	0.0174	0.0123	0.0151	0
KHC	0.0518	0.2254	0.0548	0
KO	0.0843	0.1763	0.1522	0
MCO	0.0348	0.0049	0.0704	0
USB	0.0287	0.0047	0.0449	0
VZ	0.0347	0.0367	0.2969	0

The above results indicate that the weights of AAPL, AXP, BAC, CHTR, DVA, KHC, KO, MCO, USB, and VZ in the Maximum Sharpe Ratio portfolio are 23.40%, 10.68%, 17.16%, 2.73%, 1.23%, 22.54%, 17.63%, 0.49%, 0.47% and 3.67% respectively. For the Minimum Volatility portfolio, the weights of the ten assets are 8.64%, 0.35%, 2.59%, 24.98%, 1.51%, 5.48%, 15.22%, 7.04%, 4.49%, 29.69% separately. It obvious that AAPL has the largest weight for the Maximum Sharpe Ratio portfolio and VZ accounts for the largest share for the Minimum Volatility portfolio. Therefore, more attention should be paid to AAPL and VZ.

As shown in the Table III, the initial weights of Berkshire Hathaway's positions are 47.36%, 9.77%, 16.24%, 1.46%, 1.74%, 5.18%, 8.43%, 3.48%, 2.87% and 3.47% respectively. The weights of the Risky portfolio are 74.70% for AAPL and 25.3% for AXP. Therefore, it can be concluded that more attention has been paid to AAPL in Berkshire Hathaway and under a risk loving preference, more attention should be paid to AAPL and AXP.

3.3 Future Discussion

The above portfolio optimization results are obtained from certain 10 assets. Thus, it is puzzled whether the constructed portfolios beat the market. In this section, we do further explorations to compare the portfolio. Specifically, we use predicted daily returns for the selected assets and the return of market index (S&P 500). The results for the comparison are shown as follows in Figure 3. The four interested portfolios have smaller volatilities than S&P 500 in our predicting period. Since the period is short- term based, we believe that a smaller volatility shall bring a higher accumulative return and shall reduce the trading costs in reality when conducting high frequency trading. Thus, the constructed portfolio based on the weight we obtained above beat the market.

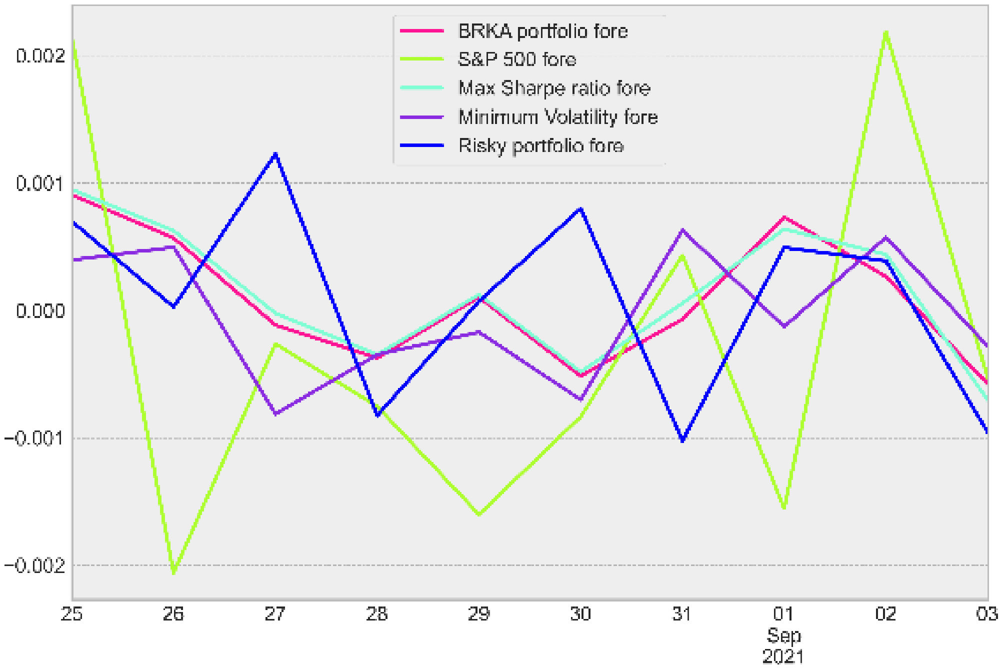

Figure 3. Comparison of the constructed portfolios and the market index return

As shown in Figure 2, we find that though the BRKA portfolio is not exactly on the efficient frontier, it is very close to it. To some extent, we think that it may suggest that what Buffet has chosen is a wise portfolio that combines both the quantitative investment and value investing together.

4. Conclusion

This paper constructs portfolios based on the Top 10 longest positions of Berkshire Hathaway, a representative investment company from the United States. We forecast the asset by ARIMA model and analyze the corresponding weights of the assets by the Monte Carlo Simulation and the Convex Optimization. Furthermore, the results show that the asset of AAPL accounts for the largest proportion in the Maximum Sharpe Ratio, while the asset of CHTR is the optimized choice if the investor is more risk-averse. Besides, it is identified that our portfolio with certain asset weight can surely beat the market index return and bring economic benefits for the related investors.

However, deficiencies also exist. For example, the time duration we select is from Jun 1, 2020 to Aug 25, 2021, which is not long, using data from a longer period is worth trying.

References

[1] Markowitz, H. M. (1952). Portfolio selection. The Journal of Finance, 7(1), 77.

[2] Fisher Black & Robert Litterman(1992). Global portfolio optimization. Financal Analysts Journal, 48(5), 28-43.

[3] Chu, C., Fang,Z..(2011),Optimal portfolio project with modified covariance matrix and its stability. Journal of University of Science and Technology of China, 41(12), 1035-1041.

[4] Wen, Q., Chen, M., Liang,B..(2011). The Study of Asset Allocation Strategy Based on Black-Litterman Framework. Journal of Applied Statistics and Management, 30(4), 741-751.

[5] hu, Y. (2016). Asset Allocation Based on Black-Litterman Model with Entropy Compensation. A Dissertation for Master's Degree, University of Science and Technology of China.

[6] Feltham G A, Ohlson J A. (1995). Valuation and clean surplus accounting for operating and financial activities. Contemporary Accounting Research, 61(3),65-79.

[7] Hou, X. Wang, B. (2021). Review and Prospect of Quantitative Investment Based on Analysis of Fundamental. Journal of Northeast Normal University (Philosophy and Social Sciences), (1), 124-131+141.

[8] Sharpe W F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance, 19(3), 425-442

[9] Qi, Y., Lin, L., Wang,Z. (2015). Directly Evaluating Genetic Algorithms for Portfolio Optimization in the Context of Big Data. Chinese Journal of Management Science, 23(S1), 464-469.

[10] Zhou, X. Zhao, M. (2019). Research on Investment Portfolio Optimization in the Background of Artificial Intelligence. Journal of Changzhou Institute of Technology, 32(02), 31-35.

[11] Li, B., Shao, X., Li,X. (2019). Research on Machine Learning Driven Quantamental Investing. China Industrial Economics, (08), 61-79

[12] Chen, C., Zhang, P., Liu, Y., & Liu, J. (2020). Financial quantitative investment using convolutional neural network and deep learning technology. Neurocomputing (Amsterdam), 390, 384-390. https://doi.org/10.1016/j.neucom.2019.09.092

[13] Novy-Marx. (2013). The other side of value: The gross profitability premium. Journal of Financial Economics, 2004(38), 297-331

[14] Pankratz, Alan. (2008). Forecasting with Univariate Box-Jenkins Models: Concepts and Cases.

[15] Pav, S. E. (2021). The sharpe ratio: Statistics and applications. CRC Press LLC.

[16] Brandimarte, P., Wiley Online Library, & O'Reilly for Higher Education. (2014). Handbook in monte carlo simulation: Applications in financial engineering, risk management, and economics. John Wiley & Sons.

[17] Zhan,Y.,TAN,X. (2011). Portfolio Investment Based on Quadratic Utility Function. Institute of Mathematics, Beihua University, Jilin, China,132013.