Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/26/20230575

Application of ARIMA Model in Portfolio Optimization

Yongjie Chen[1,a,*]
[1]*Department of Statistics, University of California Davis, Davis, CA 95616, USA*
a.yjcchen@ucdavis.edu
**corresponding author*

Abstract: This paper investigates the portfolio allocation strategy using the ARIMA time series model with eight companies in four industries: Technology, Healthcare, Financial Services, and New Energy. This research aims to identify the best portfolio allocation approach that maximizes the returns for investors in each industry. This paper collected daily stock prices of Apple and Microsoft in the technology industry, Johnson & Johnson and Pfizer in the healthcare industry, JPMorgan Chase and Goldman Sachs in the financial services industry, and Tesla and NextEra Energy in the new energy industry from January 2016 to December 2019. The logarithmic returns first were calculated for the first three years as the training data. Then, an appropriate ARIMA time series model would be fitted to forecast the stock return for the last year and perform the portfolio optimization based on the predicted return for 2019 following a result of weights for each asset. By comparison, the paper also applied portfolio optimization to the raw logarithmic daily return from 2016 to 2018 without the time series model fitted and got another set of weights for each asset. Finally, both cumulative returns for these two portfolios computed and compared with visualizations. There is a finding that the modified strategy of portfolio with ARIMA model forecasted data could provide a more obvious trade-off relationship between volatility and sharpe-ratio for returns than the traditional one performed with the raw data. The result of this paper provides a potential scope of an application of portfolio optimization worked with time series modeling for investors to make investment decisions that possibly yield a higher return.

Keywords: portfolio optimization, time series analysis, ARIMA model

1. Introduction

It is known that portfolio optimization is a very essential concept in investment and wealth management. This concept was first introduced by Harry Markowitz, which helped many investors recognize and learn the importance of portfolio allocation which can adapt the risk and returns of their investment portfolios to their different goals [1].

There has been so much research on this concept of portfolio management for decades. For example, Michael C. Bartholomew-Biggs introduced nonlinear optimization strategies to solve the financial problems with various in-depth optimization methods such as unconstrained, constrained optimization, and global optimization [2]. Vihayalakshmi pai studied applying metaheuristic algorithms to portfolio optimization, like Genetic Algorithm, Simulated Annealing, and Ant Colony Optimization [3]. Robert Engle and Stephen Figlewski researched on selection of appropriate risk models and covariance forecasting method to improve the performance of the portfolio optimization [4]. Teo Jašić,

DOI: 10.54254/2754-1169/26/20230575

Stoyan Stoyanov, and Dubravko Štimac researched and proposed one new strategy of portfolio optimization named “Factor Constrained Portfolio Optimization” (FCPO) which incorporates factor scores as constraints in the process [5]. Clive N. Enoch focused on the project portfolio management (PPM) and explored PPM maturity models that can help to enhance the portfolio management practices [6]. David Allen, Colin Lizieri, and Stephen Satchell researched on the possible forecasting methods including factor models and machine learning strategies that improve the efficiency and diversification of the portfolio allocation [7]. It can be seen that different algorithms and models were proposed and researched by the past researchers, but there was comparatively little research and few applications with one specific representative time series model of ARIMA, which holds a great capacity of short-term forecasting [8]. For the COVID-19 was a rare and unusual occurrence of global crisis, this paper would not consider this issue. Thus, in the paper, the ARIMA model would be applied to fit the data and the dataset from 2016 to 2019 (before COVID-19 epidemic) would be selected.

The procedure for the whole paper is explained in followings: First, the paper chose eight representative companies in four industries: Technology, Healthcare, Financial Services, and New Energy for adjusted closing prices from January 1, 2016, to December 31, 2019. Second, data cleaning was performed to the raw dataset and data before 2019 was treated as training data while data during 2019 was testing data. Third, eight ARIMA models with appropriate parameters would be fitted to each asset in the training data and the corresponding values of MSE and RMSE would be checked for the model accuracy and indicate a goodness of fit. Fourth, predicted values of returns for 2019 are generated from ARIMA models and used to perform the portfolio optimization. Meanwhile, the true returns from 2016 to 2018 are also used to perform the portfolio optimization as a traditional strategy. Fifth, two typical optimal portfolios are obtained, minimum volatility portfolio and maximum Sharpe-ratio portfolio, with weighted asset and index value of volatility and Sharpe ratio for each portfolio in each strategy. Sixth, the cumulative returns are calculated for both modified and traditional strategies with visualizations. Therefore, it can plainly tell the performance of each portfolio under different strategies with and without application of ARIMA model forecasting.

2. Dataset

Our dataset is requested from the website of Yahoo Finance (https://finance.yahoo.com/) with the ticker symbol of eight companies: AAPL, GOOGL, JNJ, PEE, JPM, V, TSLA, and NEE. This paper selected two representative companies for each industry. Four rising and prospering industries were chosen in the 21st century: Technology, Healthcare, Financial Services, and New Energy. The time range for our data is focused on 2016 to the end of 2019 since this paper considered removing the huge negative influence of the COVID-19 epidemic from 2020 to 2021 in our research. The raw data downloaded from Yahoo Finance is the daily price for the eight assets with 1005 observations, and the paper focuses on the value of “Adj Close.” After the data cleaning, the daily returns were obtained from the raw data of stock price for the eight assets. Since it must ensure a more accurate representation of the investment portfolio, the paper uses the logarithmic return in the following methodologies of the time series modeling and portfolio optimization, which is more robust to the high volatility. The descriptive statistics of the daily return for eight assets is shown below in Table 1.

DOI: 10.54254/2754-1169/26/20230575

Table 1: Descriptive statistics of logarithmic return for each asset.

Name	AAPL	GOOGL	JNJ	PFE	JPM	V	TSLA	NÉE
mean	0.0010	0.0006	0.0005	0.0003	0.0009	0.0009	0.0006	0.0010
std	0.0154	0.0141	0.0103	0.0110	0.0130	0.0126	0.0291	0.0097
max	0.0681	0.0919	0.0484	0.0683	0.0800	0.0718	0.1627	0.0324
min	-0.1049	-0.0780	-0.1058	-0.0663	-0.0720	-0.0541	-0.1497	-0.0633

Table 1 clearly shows that GOOGL has the largest mean value while PFE has the lowest mean value. For the standard deviation, TSLA holds the highest value while NEE gets the lowest one, closer to zero value. Overall, all values of the standard deviation for all stock returns are very small, which shows there is a relatively low variability for all daily returns.

3. Methodology

In this paper, the dataset of daily returns from 2016 to 2018 was firstly split as training data and the remaining 2019 data as testing data. Then, an ARIMA model was fitted to the training data and predicted the testing data. The paper measures the model's goodness of fit with mean squared error (MSE). With the predicted data of logarithmic return, the portfolio optimization would be performed with the algorithm of Monte Carlo Simulation and get several different allocations.

3.1. ARIMA Model

"ARIMA" is abbreviated for AutoRegressive Integrated Moving average, which is a combination of AR model, integration, and MA models in time series analysis. ARIMA models are a combination of AR and MA models with differencing, also a subset of linear regression models that attempt to use historical observations to forecast the future values of our response [9]. An autoregressive model of order p, AR(p) utilizes previous observations of the dependent variable y, up to p-th time step, as predictor variables. Here, p signifies the quantity of lagged data points in the model, ϵ is white noise at time t, c represents a constant term and φs denote the model parameters.

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \epsilon_t \tag{1}$$

Where x_t is a series explained as a function of p past values, p is the order of the autoregressive model, $\varphi_1, \dots, \varphi_p$ are the parameters of the model, and ϵ_t is white noise. Integration, also known as differencing, is to take d times to get a stationary series from the original non-stationary one. A time series is considered stationary when its statistical properties remains throughout the course of time. For the first difference$\nabla = (1 - B)$: $\nabla x_t = (1 - B)x_t = x_t - x_{t-1}$is a stationary series after being integrated into the first order. Differences of order d are defined as.

$$\nabla^d x_t = (1 - B)^d x_t \tag{2}$$

Where x_t is a series, d is order of differences, and B is a backshift operator.

A moving average model employs an approach similar to regression analysis which uses linear combination of previous errors to predict future data points, with following formula given below.

DOI: 10.54254/2754-1169/26/20230575

$$x_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{3}$$

Where μ is the mean of the series, q is order of the model, $\theta_1, .., \theta_q$ denote the model coefficients, and $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-q}$ represent white noise error terms. Therefore, the final ARIMA(p,d,q) model is obtained by merging all three models mentioned above[9].

3.2. Mean Squared Error (MSE) and Root Mean Squared Error (RMSE)

Mean squared error is an estimate which measures the average squares of the errors.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{4}$$

Where n is the number of observations, Y_i is the true value of the response variable, and $\hat{Y}_i$ is the estimate to the true value of the response [10]. Root mean squared error (MSE) is the square root of the value of MSE, with a formula given below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n}} \tag{5}$$

Where n is the number of observations, Y_i is the true value of the response variable, and $\hat{Y}_i$ is the estimate to the true response value.

3.3. Monte Carlo Simulation

Monte Carlo simulation is a statistical method applied to model and analyze one complicated system with its probability distribution. It involves generating a large number of random samples based on assumptions about the underlying data and analyzing the results to estimate the likelihood of different outcomes.

In finance, Monte Carlo simulation is often utilized to model the possible returns and risks associated with different investment strategies or portfolios under various market scenarios. By simulating a large number of possible outcomes, Monte Carlo simulation can provide more accurate estimates of the probabilities and potential outcomes of different investment decisions [11].

3.4. Sharpe Ratio

The Sharpe Ratio serves as a method to assess the performance of an investment while factoring the associated risk. It is derived by subtracting the risk-free rate of return from the average return from one asset or portfolio for a specific time frame and dividing the result by the corresponding standard deviation [12]. The corresponding formula can be found below.

$$SR = \frac{E[R_a - R_b]}{\sigma_a} \tag{6}$$

Where R_a denotes the average return, R_b represents the risk-free rate of return. σ_a is the standard deviation of excess returns.

DOI: 10.54254/2754-1169/26/20230575

4. Results

After appropriate ARIMA models fitted to each asset return, the values of MSE and RMSE are needed to be checked for each model with predicted data and testing data to test the model accuracy. Then there is a finding that both values are extremely small, close to zero, for each model (See Table 2).

Table 2: MSE and RMSE.

	AAPL	GOOGL	JNJ	PFE	JPM	V	TSLA	NEE
MSE	0.0030	0.0029	0.0001	0.0001	0.0001	0.0008	0.0018	0.0002
RMSE	0.0174	0.0171	0.0115	0.0121	0.0137	0.0275	0.0424	0.0123

For example, for the model fitted to "AAPL" returns, the MSE and RMSE have values of 0.0030 and 0.0174 correspondingly. For model fitted to "GOOGL" returns, the values of MSE and RMSE are 0.00295 and 0.0171 respectively. Therefore, it can be concluded that all ARIMA model can well fit the training data and make precise predictions for the latest-year return.

With predicted value of returns for eight assets, the portfolio optimization was performed with results of a portfolio with highest Sharpe-ratio and one with lowest volatility. For both portfolios, they give out the weights of each asset and calculate the cumulative return with given weight under this portfolio allocation for 2019 and the real return in 2019. Details for each portfolio are shown in Table 3 with a graph of efficient frontier in Fig. 1, and portfolio cumulative returns are shown in Fig. 2 and Fig. 3.

Table 3: Information for portfolio allocation using ARIMA forecasted return (2019).

	Minimum Volatility	Maximum Sharpe-ratio
AAPL	3.05%	0.00%
GOOGL	6.09%	0.00%
JNJ	0.00%	0.00%
PFE	47.62%	0.00%
JPM	1.85%	57.15%
V	9.90%	0.00%
TSLA	0.00%	0.00%
NEE	31.48%	42.87%
Volatility	1.79%	7.11%
Sharpe-ratio	756.13%	4288.71%

DOI: 10.54254/2754-1169/26/20230575

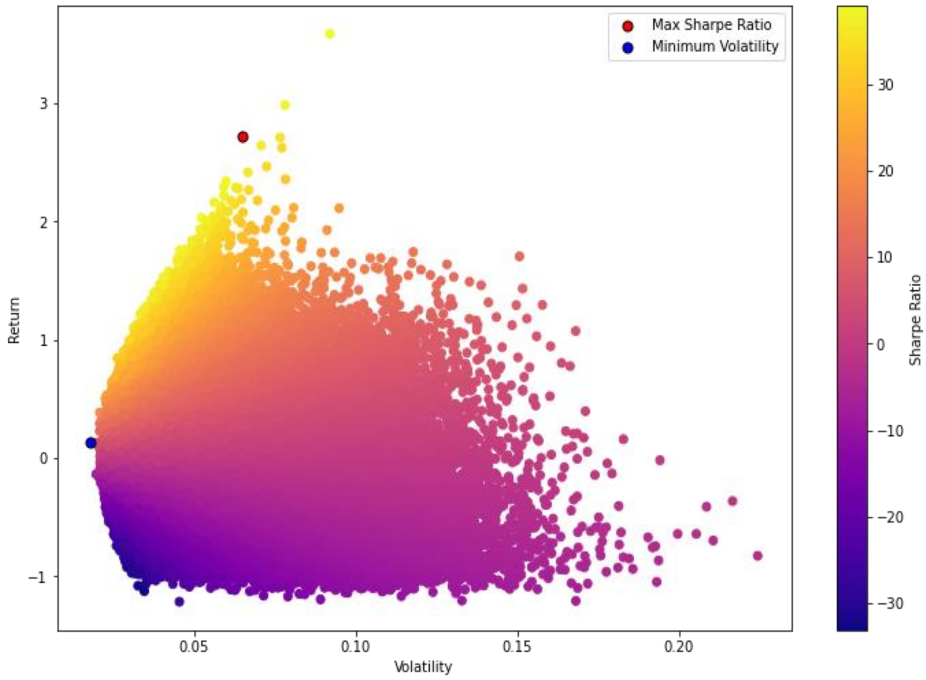

Figure 1: Efficient frontier for portfolio allocation using ARIMA model (2019).

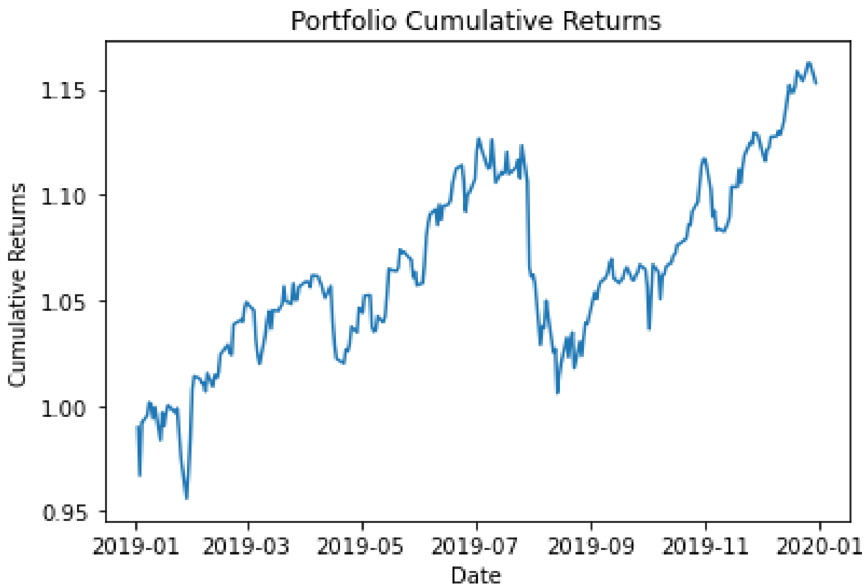

Figure 2: Cumulative returns for minimum volatility portfolio in 2019.

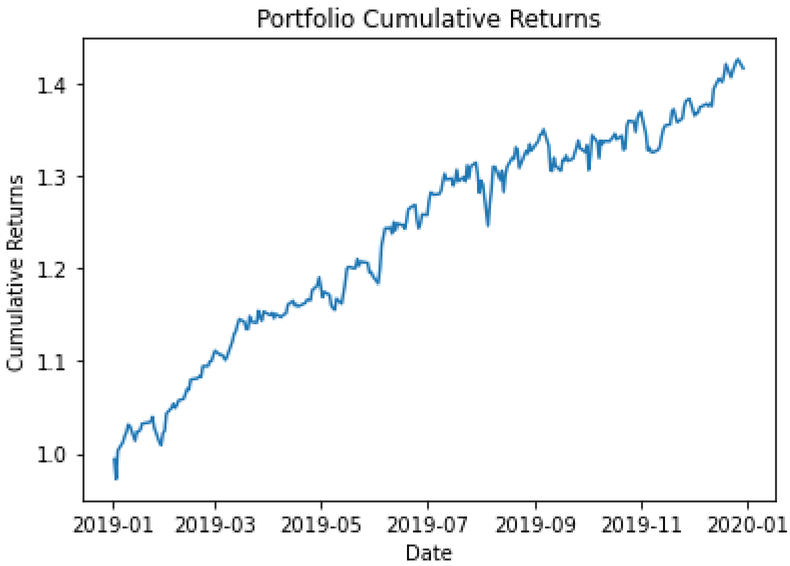

Figure 3: Cumulative returns for maximum Sharpe-Ratio portfolio in 2019.

To measure the performance of our portfolio application with the ARIMA model, this paper also performs the portfolio optimization with the original training data of daily return from 2016 to 2018 and obtain weights for each asset for each portfolio (See Table 4). The cumulative return would be also computed with these new weights for 2019 and compare it with the last one shown above. Details of the new portfolio are shown below with a graph of efficient frontier in Fig. 4. The portfolio cumulative returns are also provided in Fig. 5 and Fig. 6.

Table 4: Information for portfolio allocation using 2016-2018 true returns.

	Minimum Volatility	Maximum Sharpe-ratio
AAPL	4.11%	1.55%
GOOGL	4.35%	0.00%
JNJ	15.24%	0.00%
PFE	16.65%	11.86%
JPM	15.57%	19.42%
V	0.00%	10.66%
TSLA	0.63%	2.16%
NEE	43.44%	54.34%
Volatility	10.94%	11.44%
Sharpe ratio	165.97%	176.43%

DOI: 10.54254/2754-1169/26/20230575

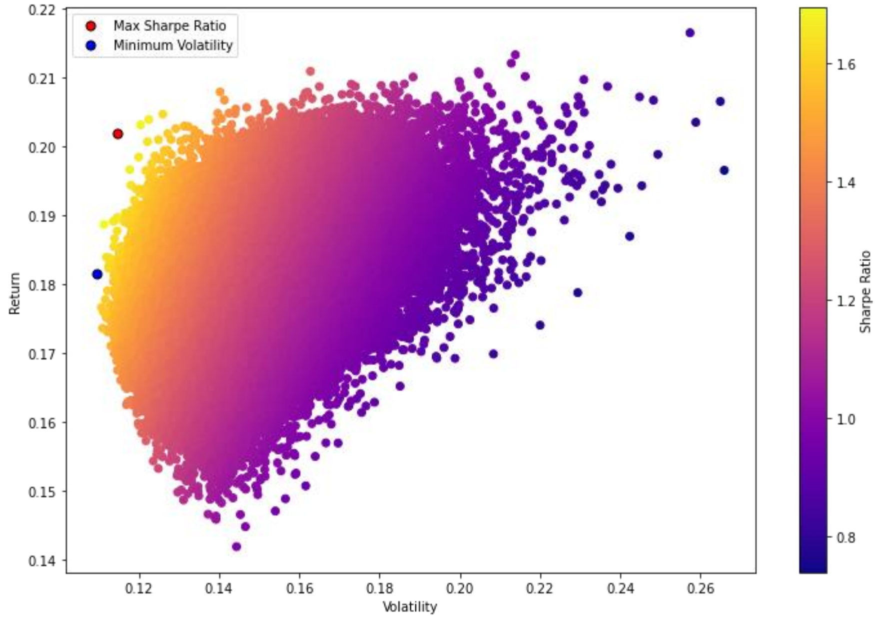

Figure 4: Efficient frontier for portfolio allocation using 2016-2018 true returns.

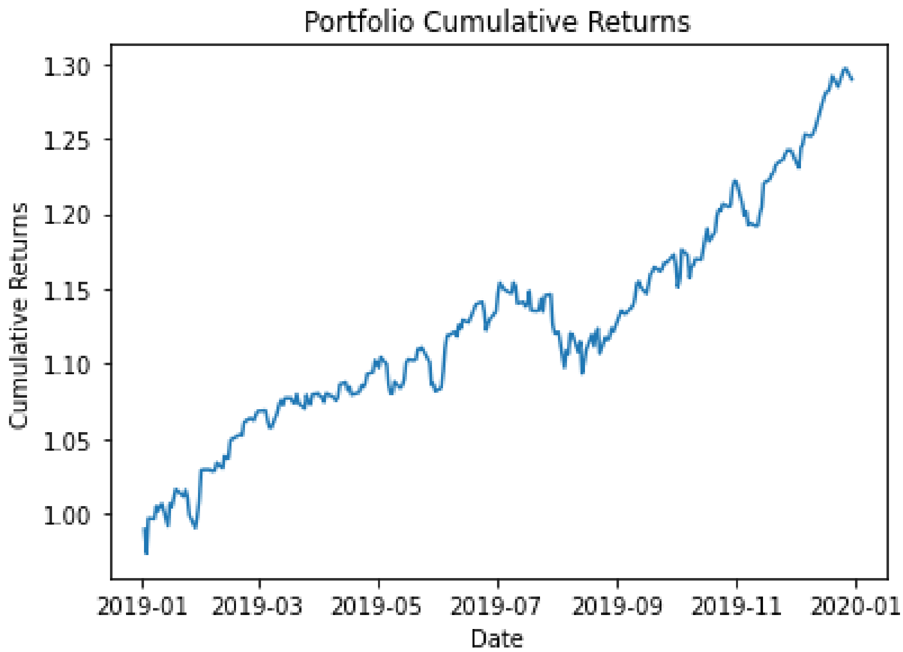

Figure 5: Cumulative returns for minimum volatility portfolio in 2019 with true returns.

DOI: 10.54254/2754-1169/26/20230575

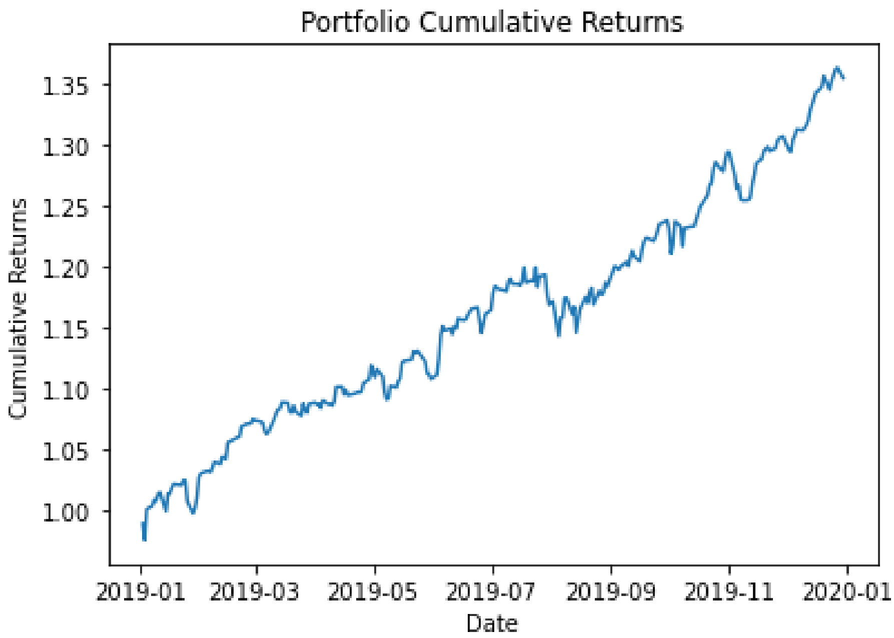

Figure 6: Cumulative returns for maximum Sharpe-Ratio portfolio in 2019 with true returns.

From Table 3, with application of the ARIMA model fitted, it can see that 'NEE' shares over 30 % of weight in both portfolios. In the Minimum Volatility portfolio, there are two zero-weight assets and six non-zero weight assets, while there are only two non-zero weights for 'JPM' and 'NEE' that both weigh over 40% in Maximum Sharpe-ratio portfolio. Furthermore, the difference of volatility for these portfolios is close to 5% while the difference in Sharpe ratio is over 100%.

From Table 4, without fitting ARIMA model first and using predicted data for 2019, it can tell 'NEE' asset share over 40% of weight for both Minimum Volatility and Maximum Sharpe-ratio portfolio. There is only one zero value for Minimum Volatility portfolio but three zero values for Maximum Sharpe-ratio portfolio. Additionally, the difference of volatility for these portfolios is within 1% while the difference in Sharpe ratio is close to 10%.

By comparison of these two strategies of portfolio optimization, it can conclude from the above Figs that the differences in cumulative returns for 2019 between Minimum Volatility and Maximum Sharpe-ratio portfolios under the first strategy (modified) are greater than those under the second strategy (traditional). More specifically, the cumulative returns at the end of 2019 for Minimum Volatility and Maximum Sharpe-ratio portfolios are close to 1.15 and 1.41 respectively under the modified strategy. However, under the traditional strategy, the cumulative returns at the end of 2019 for Minimum Volatility and Maximum Sharpe-ratio portfolios are close to 1.30 and 1.35 respectively. Also, it can tell under whatever strategies, the weights allocated for asset 'NEE' are always relatively high for all portfolios. For both values for difference in volatility and Sharpe ratio, the difference for these two parameters in the modified strategy is greatly larger than that in the traditional one.

5. Conclusion

This paper discussed the modified strategy of portfolio allocation with one typical time series model – ARIMA. To compare the performance of modified strategy and traditional one, predicted returns in 2019 through ARIMA model and true returns over 2016 to 2018 were used to perform the portfolio optimization. It can be seen from results that the modified strategy yielded portfolios that give better performance with higher returns. Moreover, the result of the modified strategy reveals a more

DOI: 10.54254/2754-1169/26/20230575

significant trade-off relationship between the volatility and Sharpe ratio, which also indicates a trade-off relationship between the returns and risk level.

The goal of this paper is to show the possibility of enhanced portfolio optimization in cooperation with the concept of time series modeling, so this highlight one small scope of this topic. In other words, there are various possible time series models beyond ARIMA that may better fit the data.

References

[1] *Markowitz, H.: Portfolio Selection. The Journal of Finance 7(1), 77–91(1952).*
[2] *Bartholomew-Biggs, M.: Nonlinear optimization with financial applications. Springer Science & Business Media (2006).*
[3] *Pai, G. V.: Metaheuristics for portfolio optimization: an introduction using MATLAB. John Wiley & Sons (2017).*
[4] *Chan, L. K., Karceski, J., Lakonishok, J.: On portfolio optimization: Forecasting covariances and choosing the risk model. The review of financial studies 12(5), 937-974(1999).*
[5] *Jašić, Stoyanov, S., Štimac, D.: Portfolio Optimization Using Factor Scores as Constraints—Factor Constrained Portfolio Optimization Approach. Journal of Portfolio Management 47(2), 145–156 (2021).*
[6] *Enoch, C. N.: Project portfolio management: a model for improved decision making. Business Expert Press (2019).*
[7] *Allen, Lizieri, C., Satchell, S.: In Defense of Portfolio Optimization: What If We Can Forecast? Financial Analysts Journal 75(3), 20–38(2019).*
[8] *Ariyo, A. A., Adewumi, A. O., Ayo, C. K.: Stock price prediction using the Arima model. International Conference on Computer Modelling and Simulation (2014).*
[9] *Newbold, P.: ARIMA model building and the time series analysis approach to forecasting. Journal of Forecasting 2, 23-35 (1983).*
[10] *Allen, D. M.: Mean square error of prediction as a criterion for selecting variables. Technometrics 13(3), 469-475(1971).*
[11] *Harrison, R. L., Granja, C., Leroy, C.: Introduction to monte carlo simulation. AIP Conference Proceedings (2010).*
[12] *Andrew W. Lo.: The Statistics of Sharpe Ratios, Financial Analysts Journal 58(4), 36-52 (2002).*