



Forecasts on Best Investment Portfolio for Healthcare Companies Based on ARIMA and GARCH Models

He Gong¹(✉), Rei Hong², and Zhuoran Li³

¹ Faculty of Arts and Science, University of Toronto, Toronto, ON M5S 1A1, Canada
helen.gong@mail.utoronto.ca

² School of Business Administration, Guangdong University of Finance and Economics,
Guangzhou 510320, Guangdong, China

³ School of Physical Sciences and Social Science, University of California, San Diego,
CA 92093, USA

Abstract. Healthcare stocks have increased due to the testing and treatment costs and many other factors caused by the COVID-19 pandemic since 2020. This paper constructs portfolios to minimize the risk and maximize the returns of healthcare stocks. Moreover, to enhance the performance of the initial portfolio, this paper uses time series analysis to forecast the stock price and verify the forecasting outcomes. The paper investigates the stock price of top healthcare companies in the United States using time series analysis to predict their performance and organize an optimal portfolio. Specifically, this research paper first employs the Auto Regressive Integrated Moving Average (ARIMA) and Auto Regressive Conditional Heteroskedasticity (ARCH) models to determine the 30 days stock price forecasting. After that, according to the historical data and forecasts, it evaluates the portfolio's efficient frontier based on Monte Carlo simulations (MCOS), which determines the minimum volatility, maximum Sharpe ratio, and the most suitable portfolio. The results show that the return of the optimal portfolio performs a more significant expected return and has less volatility than the portfolio with equal weights, which proves the validity of our model.

Keywords: ARIMA model · GARCH model · Portfolio optimization · Time series · Healthcare

1 Introduction

Predicting future returns of stocks is of great importance for investors and academics since the equity market's performance is manipulated by macroeconomics, inflation rates, "Black swan" incidents, etc. [1]. The expectation of investors for pharmaceutical companies generally exceeds those for other companies with the development of the epidemic and the emergence of monkeypox. High return is often accompanied by high volatility, and volatility in the macroeconomics and financial sectors often has a cluster

H. Gong, R. Hong and Z. Li—These authors contributed equally.

© The Author(s) 2023

V. Gaikar et al. (Eds.): FMET 2022, AEBMR 227, pp. 399–414, 2023.

https://doi.org/10.2991/978-94-6463-054-1_44

effect [2]. Accurately predicting the price of stocks helps investors make rational decisions and obtain adequate returns, while combining a portfolio helps investors diversify non-systemic risk.

A portfolio combines several securities that can balance return and risk in various uncertain environments [3]. Portfolio optimization is essential since investors hope to achieve portfolio optimization by maximizing portfolio return and minimizing portfolio risk [4]. There are more and more theories and models of the portfolio, such as the Markowitz Mean–Variance Model, Value at Risk, Conditional Value at Risk, Game Theory, etc. In 1952, Harry Markowitz published his portfolio theory about using a mathematical model to evaluate a portfolio by optimizing two conflicting risk and return criteria [5]. Theories of portfolio optimization develop rapidly, M. Ivanova and L. Dospatliev applied Markowitz portfolio optimization on the Bulgarian stock market [4], and Yu et al. used Markowitz's efficient frontier model to find the maximum Sharpe ratio portfolio of 3 stocks [6].

Constructing an investment portfolio is inseparable from predicting asset returns and volatility. Time series is a statistical method of predicting future data based on historical data. In 1976, Box and Jenkins developed the systematic procedures Auto-Regressive Integrated Moving Average model (ARIMA) for identifying and estimating, one of the most widely used models for predicting stock prices. Bollerslev developed the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model in 1986, which is designed to capture volatility clustering behavior. ARIMA models are relatively more robust and efficient than more complex structural models concerning short-run forecasting [7]. Mondall Ayodele et al. A, Adebisi, and Aderemi O. Adewumi use ARIMA models to predict Nokia and Zenith bank stock prices, and 57 companies found that there are some instances of closely related actual and predicted values [8]. Naik et al. N, Mohan B R, and Jha R A illustrated that GARCH helps forecast the stock crisis events by giving stock crisis events as input to the GARCH model [9]. Keying Sun combined the ARIMA and GARCH models, showing that the hybrid ARIMA-GARCH model is appropriate for predicting equity returns [10]. This paper aims to develop a reasonable prediction of stock prices of pharmaceutical companies based on the ARIMA-GARCH model and an optimal portfolio of the security based on the Mean–Variance model. The analysis involved three steps: selecting and processing the stock data and using the ARIMA and GARCH model to forecast and optimizing the portfolio using the Mean–Variance model and Monte Carlo simulation. This paper's innovation lies in predicting pharmaceutical companies' stock prices after the pandemic and constructing optimal portfolios for investors. The forecasts denote the future return will be close to 0, and the optimal portfolio performs well to be compared with the portfolio with equal rights.

This paper is organized as follows: Sect. 2 describes the data selected and processes the data. The forecasting procedure is illustrated in Sect. 3. Section 4 and Sect. 5 include portfolio optimization and some concluding remarks, respectively.

2 Initial Data Preprocessing

2.1 Data Collection

2.1.1 Company Selection

This paper strives to find the best possible investment combinations among ten top healthcare companies. Hence, the data is chosen from these companies, attracting growing attention from global investors due to the COVID-19 pandemic. To be specific, the companies are Johnson & Johnson (JNJ), UnitedHealth Group Incorporated (UNH), Eli Lilly and Company (LLY), Pfizer Inc. (PFE), AbbVie Inc. (ABBV), Novo Nordisk A/S (NVO), Merck & Co., Inc. (MRK), Thermo Fisher Scientific Inc. (TMO), AstraZeneca PLC. (AZN), Novartis AG (NVS), and Abbott Laboratories (ABT). The companies are listed in descending order in terms of their market price according to Yahoo Finance on June 19th, 2022, and Yahoo Finance is also the primary data source our research is based upon.

2.1.2 Time Range Set up

Given that the COVID-19 broke out around the start of 2020, our research gets rid of the turbulence time period from 2020/1/1 to 2020/5/31 since that period demonstrates a volatile market price, and most companies were forced to make changes and adaptations. Furthermore, the research is designed to investigate how firms may perform in the near future with the presence of the COVID-19 pandemic, so the reasonable range for data collection is from 2020/6/1–2022/6/1, where COVID-19 existed in both periods, and the market price tends to be stable.

2.2 Data Filtering

2.2.1 Remove Outliers

It can be anticipated that certain outliers could be detrimental to model analysis, thus extracting them from existing model before further investigations is needed. Take the outliers plot from Johnson & Johnson as an example of how outliers for other assets are analyzed. Figure 1 marks two abnormal points around March 2022. Since it's around the time of oil price shock, the market encountered an increase in inflation and led to an overrated increase in return. To make future model more accurate, these two abnormal data points are not used in forecasting for the return of Johnson & Johnson. Total outliers appear to be a small amount, and the other nine assets share similar numbers of outliers.

2.2.2 Incorporate Inflation

During the two years period of market development, the volatile unemployment rate gave rise to an unstable inflation rate. Therefore, data of the Consumer Price Index (CPI) are incorporated into our research, and the inflation rate is calculated by Eq. (1).

$$\text{Inflationrate} = \frac{CPI^{before}}{CPI^{after}} \times 100 \quad (1)$$

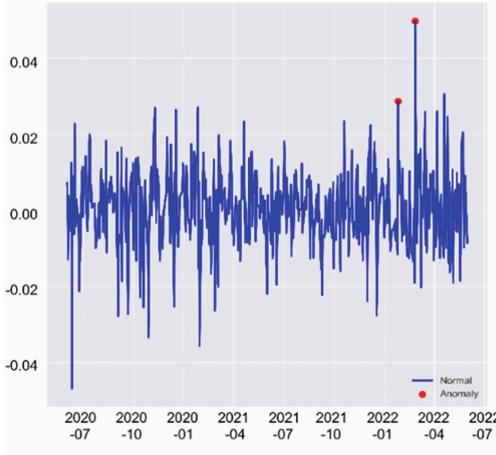


Fig. 1. Outliers plot for Johnson & Johnson

After that, both nominal prices are combined with the inflation rate to find out the real price.

2.3 Return Analysis

2.3.1 Choice of Return

There are two ways to calculate returns: simple return, as in Eq. (2) and log return, as in Eq. (3).

$$R_t = \frac{P_{t+1}}{P_t} - 1 \tag{2}$$

$$R_t = \ln\left(\frac{P_{t+1}}{P_t}\right) \tag{3}$$

where R_t denotes the return, P_t and $P_{t + 1}$ are the current stock price and lag one day stock price respectively.

Our research adopts log return because of the smoothness and symmetrical properties of log functions.

2.3.2 Return Analysis

As shown in the Fig. 2, the autocorrelation of log returns appears to be small, meaning that the future return may be uncorrelated with the return in the past, and that may potentially decrease the accuracy of both ARIMA model and the GARCH model this paper aimed to construct and forecast.

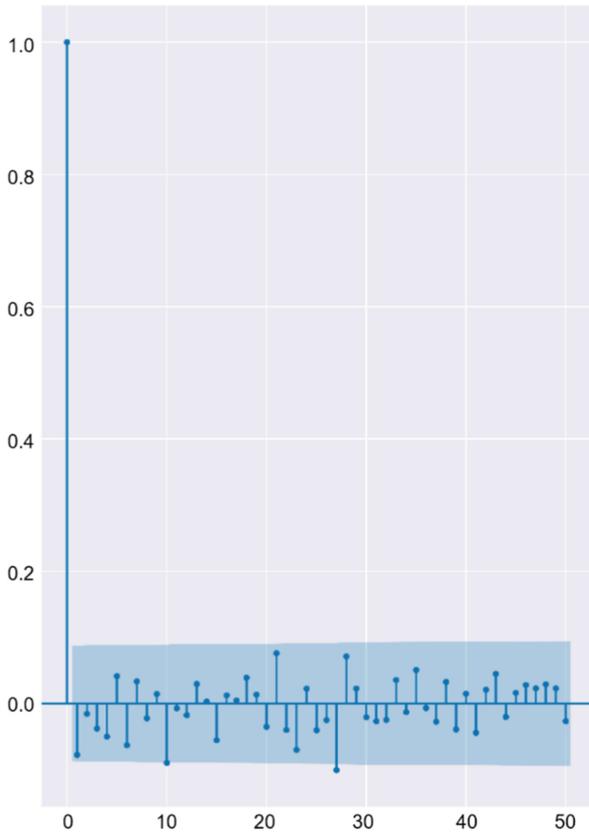


Fig. 2. Autocorrelation plot of log return

3 Time Series Analysis

Our research adopts time series as the primary tool in analyzing data and modeling because time series do well in visualizing data and facilitating the interpretation of trends. Based on time series, two models are used to fit further existing data for future forecasts: the ARIMA and GARCH models.

3.1 ARIMA Model

3.1.1 ARIMA Introduction

There are three parameters required for an ARIMA model: p , d , and q . The parameter p is the number of lags in the Autoregressive model (AR), which is the first half of the ARIMA model. The second part of the ARIMA model is the moving average model (MA), which needs a choice of parameter q . Finally, the parameter d stands for the times we wish to repeat the ARMA process. The reason why the ARIMA process is suitable for forecasting is that it captures effects from both market momentum and white noise,

imitating a natural market as best as it could. The ARMA model for calculation is listed as in Eq. (4):

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + w_t + \beta_1 w_{t-1} \dots + \beta_q w_{t-q} \quad (4)$$

3.1.2 ARIMA Parameters

Under the motive of choosing the best suitable ARIMA parameters, our research adopts the Akaike Information Criterion and selects the parameter combinations with the lowest test error. Akaike Information Criterion aims to predict the difference between the forecasted value and the real value, so a small test error means that the prediction error is small and thus more desirable for the choice of parameters. In our research, we limit the range of parameters p and q to be greater than 0 and smaller than 4, which is a reasonable range for most ARIMA processes to function. Moreover, over choice of parameter d is constrained from 0 to 2, because after d gets greater than 3, the model will be less significant.

3.1.3 Significance Test of ARIMA Parameters

TO determine the statistical significance of ARIMA parameters, we use both the Ljung-Box test and the visualization method to see if the ARIMA parameters calculated in the previous step are valid.

3.1.3.1 The Ljung-Box Test

The Ljung-Box test is helpful in testing whether the parameters in the ARIMA model generate serial correlation. The null hypothesis of the Ljung-Box Test states that the data is independent, and correspondingly the alternative hypothesis states that the data have serial correlation. Here, a good choice of the parameter should produce a p -value greater than 0.05 under the Ljung-Box test, meaning that the residuals of the model should be independently distributed with no interference from further modeling. If it turns out that the p -value is smaller than 0.05, it is possible that residuals give rise to some error in the future forecast.

3.1.3.2 Visualization Method

To further confirm the selection of ARIMA parameters, our research manages to utilize the advantage of time series, that is data visualization. Five plots in total could facilitate our understanding of data: time series plot of log return, autocorrelation plot (ACF), partial autocorrelation plot (PACF), QQ plot and the probability plot. The time series plot of log returns gives an overview of the data trends, providing a basic understanding of data trends. The ACF plot and PACF plot can be used to check the value of parameter p . Notice that p stands for the optimal lag of data, so in either the ACF plot or the PACF plot, there should be a sign of correlation around the lag of p . The QQ plot and the probability plot give the probability distribution of the ARIMA models given chosen parameters. A desirable graph should show signs of normal distribution, meaning that the model is a good fit.

3.1.4 Results and Graphs

For the results, the analysis process is similar across ten assets, so here the interpretation of Johnson & Johnson serves as an example to show how other assets are analyzed. Running the Akaike Information Criterion, the model obtains the parameter value to be ARIMA (2, 0, 0). Then, it follows the step above to check the validity of the parameters. From the Ljung-Box test, the p-value for the parameters is 0.942 which is greater than 0.05. Thus, one can reject the null hypothesis and conclude that the parameters were chosen to form a good fit for the ARIMA model. Then, the visualization process continues to check the parameters by the five plots.

Figure 3 describes the overall tendency of the return time series. Note that most return data points lie in the range from -0.025 to 0.025 , and the future forecast should be most likely in this region. According to the Fig. 4 and Fig. 5, the ACF and PACF plots are not significant in confirming the choice for p equals 2, because there is no

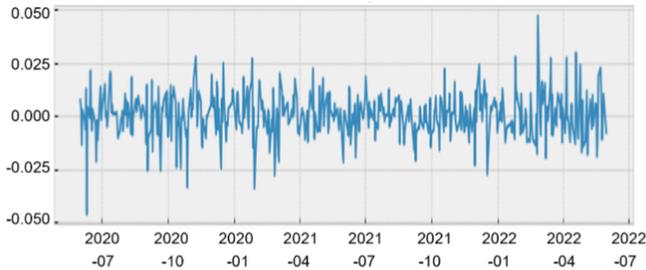


Fig. 3. Time series analysis plot

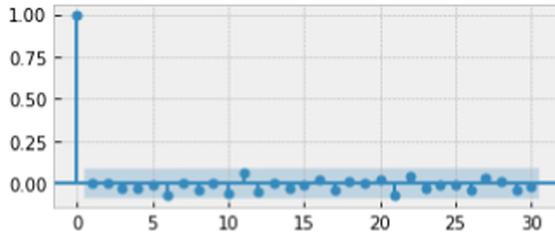


Fig. 4. Autocorrelation plot

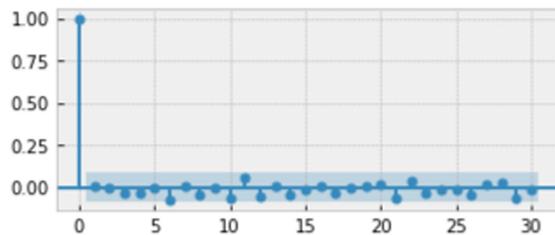


Fig. 5. Partial autocorrelation plot

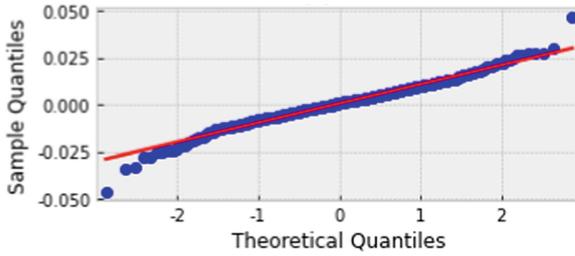


Fig. 6. QQ plot

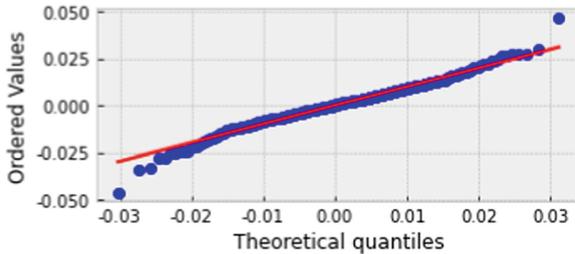


Fig. 7. Probability plot

significant serial correlation in the graph. This conforms with our previous finding in Fig. 2 which shows the autocorrelation is not significant and thus may lead to a less accurate prediction. Figure 6 and Fig. 7 display the QQ plot and the probability plot and show that the sample distribution is close to a normal distribution, which is desired and can enhanced the model accuracy to some extent.

3.2 GARCH Model

3.2.1 GARCH Introduction

According to the residual plot done by the ARIMA model, it is apparent that the ACF and PACF plots do not sufficiently support the parameters chosen. Here, the GARCH model can be used to further analyze the trend of data and make forecasts. The GARCH model is similar to the ARIMA model in terms of its mathematical structure and logic. The GARCH model focuses on the variance of lagged values instead of the original version of lagged values. Equation (5) captures the behaviors of the GARCH model. p determines the lag length, where $p = 0$ indicates white noise.

$$Var(x_t) = \alpha_0 + \alpha_1 Var(x_{t-1}) + \dots + \alpha_p Var(x_{t-p}) + w_t \tag{5}$$

To generalize the GARCH model, our research will use student t-distribution as a primary type of distribution throughout modeling and predictions.

3.2.2 GARCH Parameters

There are two parameters required to fit the GARCH model, p and q . Unlike the ARIMA process, parameter d is left out since no repetition is needed in the GARCH model. However, the function of p and q values here share a similar function with the ARIMA process, in which p gives the time lag, and the q serves for the moving average process. Analogously, Akaike Information Criterion is also valid for computing the best parameter values in the GARCH model.

3.2.3 Significance Test of GARCH Parameters

3.2.3.1 The GARCH Parameters p-value Test

The most straightforward way to check if the obtained GARCH parameters are significant is through a hypothesis t-test. The null hypothesis states that the parameters are equal to 0, or in other words, they are not significant. In the other direction, the alternative hypothesis states that the parameters are different from zero and thus meaningful. Therefore, a p-value smaller than 0.05 will prove the significance of the selected GARCH parameters.

3.2.3.2 Visualization Method

Since the GARCH model should only be applied to a dataset with residuals distributed like discrete white noise, the five-time series plots mentioned in the ARIMA model could also be applied in the GARCH model for data examination. First, the residual time series plot should be clear to be similar to white noise. Next, the value of parameters p and q should be significant in ACF and PACF plots.

3.2.4 Results and Graphs

Now the GARCH equation for Johnson & Johnson is illustrated by Eq. (6). Again, other assets could be interpreted using the similar logic as Johnson & Johnson, here we take JNJ as an example.

$$\text{Var}(x_t) = \alpha_0 + \alpha_1 \text{Var}(x_{t-1}) + w_t \quad (6)$$

As found in the t-test for parameters, α_1 has a p-value smaller than 0.05, but α_2 has a p-value greater than 0.05. The test may not make parameters seem too significant, but that's the combination with the least AIC test error the GARCH model found.

Figure 8 is a visualization of the square of residuals. In this graph, one can notice that the square of residuals is close to zero - what the GARCH model requires. However, in Fig. 9 and Fig. 10, the ACF and PACF plots do not show significant autocorrelation in any lag values. This means that the GARCH model would only explain the data partially for Johnson & Johnson.

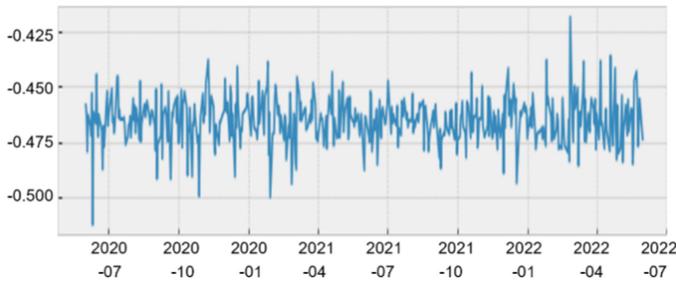


Fig. 8. Square of time series residual plot

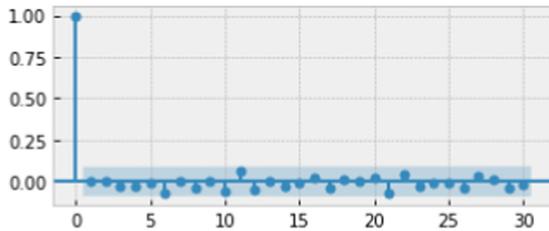


Fig. 9. Autocorrelation plot

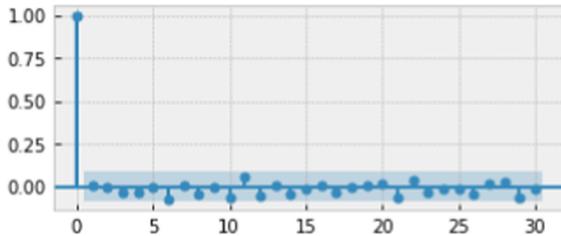


Fig. 10. Partial autocorrelation plot

3.3 Forecast

Given the model obtained from ARIMA and GARCH, one can make predictions about future returns based on existing models and data. The data is presented in two ways: in the table of numbers and plots, as in the Table 1 and Fig. 12. In either case, the model gives 95% and 99% confidence intervals to guarantee room for error. The result of the forecast, once again, will be presented in the same form for all ten assets, and here the forecast of Johnson & Johnson will be used as an example for the other nine assets. Table 1 displays the several top rows of predicted values. Notice that the data is collected up until June 1st, so starting from June 2nd is the forecasted value.

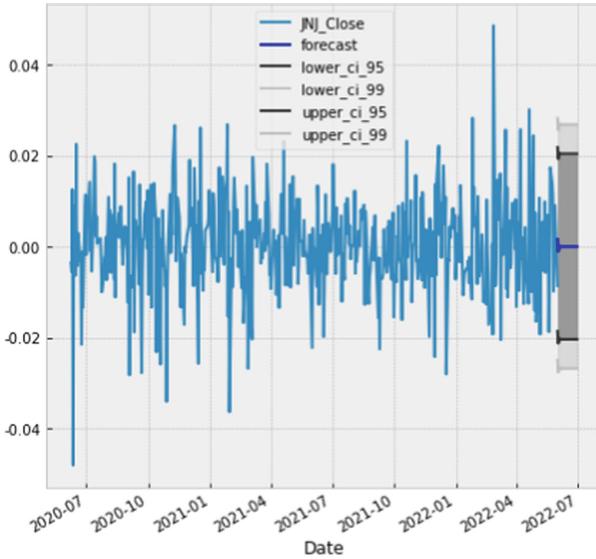


Fig. 11. Return forecast graph

Table 1. Return forecast table

| | Forecast | Lower 95% CI | Upper 95% CI | Lower 99% CI | Upper 99% CI |
|------------|----------|--------------|--------------|--------------|--------------|
| 2022-05-31 | 0.002 | -0.019 | -0.022 | -0.025 | 0.028 |
| 2022-06-01 | -0.001 | -0.021 | 0.019 | -0.028 | 0.026 |
| 2022-06-02 | 0.000 | -0.020 | 0.021 | -0.020 | 0.027 |
| 2022-06-03 | -0.000 | -0.020 | 0.020 | -0.027 | 0.027 |
| 2022-06-04 | 0.000 | -0.020 | 0.020 | -0.020 | 0.027 |

Notes: CI stands for confidence interval

As Fig. 11 says, it is most likely that the future return will be close to 0. Combined with Table 1, one can be 95% confident that the future return is in the range of 2% and 99% confident that the future return is in the range of 2.8%. The asset value could be anywhere in the shaded areas with a high probability.

4 Portfolio Optimization

A modern portfolio theory is one of the essential analytical tools. Based on a security’s expected return and risk and its correlation with other securities in a portfolio, portfolio optimization can determine which stocks should be selected to maximize return for any level of risk [11]. We combine Mean–Variance Models and Monte Carlo simulations to optimize the portfolio.

4.1 Efficient Frontier

4.1.1 Efficient Frontier Introduction

An efficient frontier graph has a Y-axis for returns and an X-axis for volatility. It presents a set of optimal portfolios that provide the highest expected return at a particular risk level or the lowest expected return at a certain risk level [12].

4.1.2 Monte Carlo Simulations

Monte Carlo simulations are served to find out all possible combinations of investment bundle by building models with different parameters. Each time it calculates results, it uses a new set of random values from the probability functions [13].

This paper uses Monte Carlo simulations to simulate the efficient frontier graph and get the maximum Sharpe ratio and minimum volatility.

4.1.3 Covariance Matrix

Covariance matrices explain the relationship between different assets. Covariance between two assets demonstrates the dependence on each other. In this paper, all the assets should show some dependency on each other since they belong to the same healthcare sector. The Table 2 indicates that they have a weak dependency.

4.1.4 Efficient Frontier Simulation

This paper sets 250 days as the yearly trading days, simulates 10^5 times with different weights for each asset, and estimates the return and volatility of that weight combination. Figure 12 is the efficient frontier plot. The black star represents the maximum Sharpe ratio, and the black cross represents the minimum volatility.

Table 2. Covariance matrix

| | ABBV | ABT | AZN | JNJ | LLY | MRK | NVO | PFE | TMO | UNH |
|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| ABBV | 1.89e-4 | 6.8e-5 | 8.3e-5 | 6.3e-5 | 1.21e-4 | 7.7e-5 | 6.2e-5 | 9.3e-5 | 5.2e-5 | 9.1e-5 |
| ABT | 6.8e-5 | 2.35e-4 | 7.7e-5 | 6.9e-5 | 9.6e-5 | 6.2e-5 | 1.02e-4 | 6.1e-5 | 1.64e-4 | 9.6e-5 |
| AZN | 8.3e-5 | 7.7e-5 | 2.74e-4 | 6.2e-5 | 1.13e-4 | 8.3e-5 | 1.30e-4 | 8.6e-5 | 8.2e-5 | 6.8e-5 |
| JNJ | 6.3e-5 | 6.9e-5 | 6.2e-5 | 1.07e-4 | 8.4e-5 | 6.1e-5 | 4.9e-5 | 7.3e-5 | 4.6e-5 | 7.4e-5 |
| LLY | 1.21e-4 | 9.6e-5 | 1.13e-4 | 8.4e-5 | 4.11e-4 | 1.04e-4 | 1.30e-4 | 1.07e-4 | 9.2e-5 | 1.08e-4 |
| MRK | 7.7e-5 | 6.2e-5 | 8.3e-5 | 6.1e-5 | 1.04e-4 | 1.90e-4 | 6.1e-5 | 7.1e-5 | 8.0e-5 | 7.4e-5 |
| NVO | 6.2e-5 | 1.02e-4 | 1.30e-4 | 4.9e-5 | 1.30e-4 | 6.1e-5 | 2.80e-4 | 6.1e-5 | 1.19e-4 | 7.8e-5 |
| PFE | 9.3e-5 | 6.1e-5 | 8.6e-5 | 7.3e-5 | 1.07e-4 | 7.1e-5 | 6.1e-5 | 2.92e-4 | 6.7e-5 | 7.7e-5 |
| TMO | 5.2e-5 | 1.64e-4 | 8.2e-5 | 4.6e-5 | 9.2e-5 | 8.0e-5 | 1.19e-4 | 6.7e-5 | 2.94e-4 | 1.00e-4 |
| UNH | 9.1e-5 | 9.6e-5 | 6.8e-5 | 7.4e-5 | 1.08e-4 | 7.4e-5 | 7.8e-5 | 7.7e-5 | 1.00e-4 | 2.13e-4 |

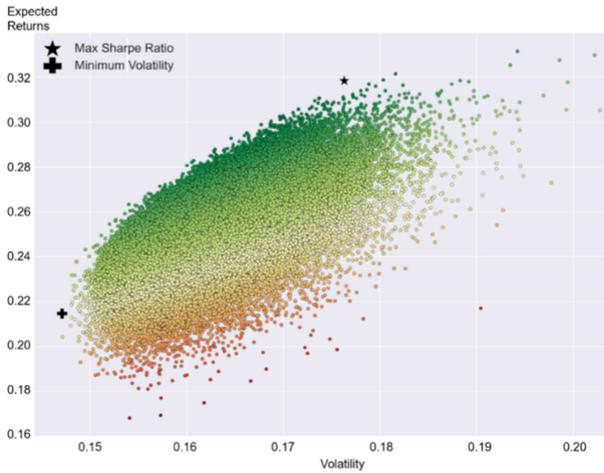


Fig. 12 The efficient frontier plot with Max Sharpe ratio and min volatility

4.1.5 Portfolio Performance

Table 3 displays the portfolio performance and the corresponding weights of each asset. Compared with the minimum volatility portfolio, the maximum Sharpe ratio portfolio is expected to generate a return that is 11% higher. Furthermore, the volatility of the maximum Sharpe ratio is 3% higher than that of minimum volatility. Cautious investors choose the portfolio with less volatility, while risky investors prefer the maximum Sharpe ratio. The calculation of the Sharpe ratio is based on the Eq. (7), where S_a denotes the Sharpe ratio, R_a and R_b denote the asset return and the risk-free return respectively:

$$S_a = \frac{E[R_a - R_b]}{\sigma_a} \tag{7}$$

4.1.6 Comparison of Return Taer Sheet

The paper constructs an equal-weight portfolio to compare to the optimal portfolio. The optimal portfolio’s maximum Sharpe ratio and minimum volatility perform better than the equal-weight portfolio. It denotes that portfolio optimization is indeed valid. Table 4 is the return tear sheet of the portfolio.

Table 3 Performance of maximum sharpe ratio portfolio and minimum volatility portfolio

| | Minimum volatility | Maximum Sharpe ratio |
|----------------|--------------------|----------------------|
| Return | 21% | 32% |
| Volatility | 15% | 18% |
| Sharpe ratio | 146% | 181% |
| Weight of AABV | 15.42% | 25.69% |
| Weight of ABT | 4.96% | 0.24% |
| Weight of AZN | 5.71% | 1.01% |
| Weight of JNJ | 30.75% | 0.35% |
| Weight of LLY | 0.29% | 19.79% |
| Weight of MRK | 11.83% | 1.27% |
| Weight of NVO | 3.85% | 8.62% |
| Weight of PFE | 2.10% | 14.48% |
| Weight of TMO | 8.89% | 17.67% |
| Weight of UNH | 16.21% | 10.88% |

Table 4 Return tear sheet

| | Minimum volatility | Maximum Sharpe ratio | Portfolio of equal rights |
|--------------------|--------------------|----------------------|---------------------------|
| Annual return | 22.80% | 35.70% | 27.10% |
| Cumulative returns | 50.80% | 84.10% | 61.70% |
| Annual volatility | 14.80% | 17.70% | 15.90% |
| Sharpe ratio | 1.47 | 1.81 | 1.59 |
| Calmar ratio | 2.84 | 3.15 | 2.75 |
| Stability | 0.97 | 0.97 | 0.97 |
| Max drawdown | -8.00% | -11.30% | -9.90% |
| Omega ratio | 1.28 | 1.37 | 1.31 |
| Sortino ratio | 2.16 | 2.89 | 2.44 |

5 Conclusion

Due to the rise in healthcare stocks' prices caused by COVID-19, this study intends to predict the returns of healthcare stocks. A portfolio is constructed to minimize risk, and the stock price is forecasted using time series analysis to verify the accuracy of the forecast. Time series analysis is used to determine the stock price of the top ten U.S.

healthcare companies and formulate an optimal portfolio. This study aims to verify the accuracy of the stock forecast by the ARIMA and GARCH models, and this study aims to build a model for portfolio optimization.

First, after preprocessing the data, our research embraces the Akaike Information Criterion to determine the most suitable ARIMA parameter combination with the lowest test error and applies the Ljung-Box test and the visualization method to check if the ARIMA parameters are valid. Considering the residual plot created by the ARIMA model, the ACF and PACF plots do not support the parameters selected sufficiently. Then the GARCH model is used for further analysis, and this paper makes predictions about future returns using existing models. Moreover, this paper evaluates the portfolio's efficient frontier based on MCOS.

As a result, even though the square of residuals is close to zero and the ACF and PACF plots do not reveal any significant autocorrelation in lag values, this paper concludes that the GARCH model can only adequately explain the data. Comparing the forecasted data with the actual data, the range of the confidence intervals of the predicted returns includes the actual returns. Moreover, the portfolio optimization shows that the return of the optimal portfolio performs a more significant expected return and has less volatility than the portfolio with equal weights, which suggests that our model is valid.

The research proves the value of time series analysis in stock price forecasting. However, the prediction does not specify the micro fluctuation trend. Involving the LSTM model based on this paper's result may get a more accurate forecast.

References

1. F H Mustapa, M T Ismail, Modelling and forecasting S&P 500 stock prices using hybrid Arima-Garch Model, in: Journal of Physics: Conference Series Journal of Physics, 2019, 1366, pp. 012130. DOI: <https://doi.org/10.1088/1742-6596/1366/1/012130>
2. X He, L Kai, C Wang, Volatility clustering: a nonlinear theoretical approach, in: Journal of Economic Behavior & Organization, 2016, 130, pp. 274-297. DOI: <https://doi.org/10.13140/RG.2.1.1245.8485>
3. X Huang, What is portfolio analysis, in: Portfolio Analysis, Springer, Berlin, Heidelberg, 2010, 180, pp. 1-9. DOI: https://doi.org/10.1007/978-3-642-11214-0_1
4. M Ivanova, L Dospatliev, Application of markowitz portfolio optimization on bulgarian stock market from 2013 to 2016, in: International Journal of Pure and Applied Mathematics, 2017, 117(2), pp. 291-307. DOI: https://doi.org/10.1007/978-3-642-11214-0_1
5. S F Witt, R Dobbins, The Markowitz contribution to portfolio theory, in: Managerial Finance, 1979, 5(1), pp. 3-17. DOI: <https://doi.org/10.1108/eb013433>
6. Linhui Yu, Changjun You, Xin Pan, Tianyuan Zhang, Optimisation of investment portfolios based on Markowitz theory, Proceedings of 4th International Conference on e-Education, in: e-Business and Information Management, 2021, pp. 149-157. DOI: <https://doi.org/10.26914/c.cnkihy.2021.005738>
7. A A Ariyo, A O Adewumi, C K Ayo, Stock price prediction using the ARIMA model, in: 2014 UKSim-AMSS 16th international conference on computer modelling and simulation. IEEE, 2014, pp. 106-112. DOI: <https://doi.org/10.1109/UKSim.2014.67>
8. P Mondal, L Shit, S Goswami, Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices, in: International Journal of Computer Science, Engineering and Applications, 2014, 4(2), pp. 13. DOI: <https://doi.org/10.5121/ijcsea.2014.4202>

9. 10. N Naik, B R Mohan, R A Jha, GARCH model identification for stock crises events, in: *Procedia Computer Science*, 2020, 171, pp.1742-174.DOI: <https://doi.org/10.1016/j.procs.2020.04.187>
10. 11. Kaiying Sun, Equity return modeling and prediction using Hybrid ARIMA-GARCH Model, in: *International Journal of Financial Research*, 2017, 8(3), pp. 154. DOI: <https://doi.org/10.5430/ijfr.v8n3p154>
11. “Portfolio Optimization with Python using Efficient Frontier with Practical Examples,” *Machine Learning Plus*, Oct. 13, 2020. <https://www.machinelearningplus.com/machine-learning/portfolio-optimization-python-example/#:~:text=Efficient%20frontier%20is%20a%20graph> (accessed Jun. 24, 2022).
12. D. Merritt and A. de San Miguel, Portfolio Optimization using Efficient Frontier Theory, *onepetro.org*, Apr. 25, 2000. <https://onepetro.org/SPEAPCIMAM/proceedings-abstract/00APCIMAM/All-00APCIMAM/SPE-59457-MS/133041>
13. Palisade, What is Monte Carlo Simulation?, Monte Carlo Simulation: What Is It and How Does It Work? - Palisade. https://www.palisade.com/risk/monte_carlo_simulation.asp#:~:text=Monte%20Carlo%20simulation%20performs%20risk

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

