Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/25/20230504

Portfolio Optimization Based on LSTM for 5 Stocks

Yize Peng[1,a,*]

[1] *Shanghai University of Finance and Economics, Shanghai 200433, China*
a. pyz2022@163.sufe.edu.cn
**corresponding author*

Abstract: Portfolio is a scientific model that maximizes returns and minimizes risks. It is widely used in production and life, such as stock investment, production optimization, and engineering models. This article selects the stocks of Google, Microsoft, Nvidia, Amazon, and Tesla from September 1, 2022, to March 1, 2023. First, this article uses the LSTM model to learn the top 70% of the data, and predicts the bottom 30% of the data, then use the portfolio strategy for the forecast results. Thus, this paper gets the maximum return model and the minimum risk model on the effective frontier of Monte Carlo. By applying two models of the portfolio and combining two results with the S&P 500 within the same time frame, this article draws a conclusion that models in this paper are better. This article provides a portfolio method using the latest LSTM model as a forecasting model, and the comparison reflects the power of the joint use of the LSTM model and the portfolio model.

Keywords: mean-variance model, portfolio management, LSTM model

1. Introduction

Project portfolio selection is an important issue for many firms, which is also important for investors [1]. A diversified portfolio can provide investors with higher yields at equal risk [2]. Thus, how to obtain higher returns under the lowest risk conditions is undoubtedly investors most important concern.

Inspired by Harry Markowitz's Modern Portfolio Theory [3]. The portfolio has achieved several impressive outcomes. For instance, Bo and Lay researched portfolio optimization under the entire uncertain real financial market [4]. Eyal analyzed portfolio selection in non-stationary markets [5]. Hans, Sahamkhadam and Stephan studied the best portfolio performance of global timber and forestry industry and compared it with the global S&P index [6]. Meanwhile, some researchers have also introduced predictive models into portfolios, which is useful for improving portfolio performance. For example, Ma, Han and Wang introduced a variety of neural network prediction models into portfolio construction in the paper to improve portfolio performance [7]. In recent days, LSTM is a research hotspot, and the related research on introducing LSTM into portfolio construction is still relatively limited.

This paper connects LSTM with portfolio management and makes the following contributions. First, this article selects and processes the stock data of five well-known multinational high-tech companies in the same period. The processed data sets are Google, Microsoft, Nvidia, Amazon, and Tesla from September 1, 2022, to 2023. The closing price of the stock on March 1. Second, this article uses LSTM model to regard top 70% of the data as the set trained, and the bottom 30% of the data as the

DOI: 10.54254/2754-1169/25/20230504

set tested to obtain the set predicted. Third, the predicted closing price of the forecast set is converted into the cumulative rate for returns, the maximum Sharpe ratio pattern and the minimum variance pattern are obtained by using the mean variance model. Fourth, draw cumulative returns of two models using data predicted and cumulative returns of the S&P 500 in the same graph, and get the conclusion that both Sharpe model and Risk model are better than the S&P 500.

The paper is structured in the following patterns, Data section introduces stocks chose and data in the study. Methods section describes means of LSTM and MVA. Results and the conclusions will be presented in Section 4 and Section 5.

2. Data

This paper selects 5 representative stocks based on the aspect of Driven by technology, technological innovation, a global company with high market value and efficient operation that promotes the progress of human society. The ticker of the 5 stocks is GOOGL, MSFT, NVDA, AMZN, TSLA. Closing prices from September 1st, 2022, to March 1st, 2023, are collected from Yahoo finance (https://finance.yahoo.com/) and separated into training set and test set. For further research, we do data cleaning that matches the time. Finally, 615 data are collected (Details are shown in the following Table 1, Table 2).

Table 1: Selected stocks.

Stock Symbol	Company
GOOGL	Alphabet Inc.
MSFT	Microsoft Corporation
NVDA	NVIDIA Corporation
AMZN	Amazon.com, Inc.
TSLA	Tesla, Inc.

We transfer these closing prices to Compounded return (log-return) and calculate some basic information shown in the following table:

Table 2: Descriptive statistics of five stocks.

	'GOOGL'	'MSFT'	'NVDA'	'AMZN'	'TSLA'
Mean	-0.0016	-0.0004	0.0042	-0.0025	-0.0024
Std	0.0257	0.0230	0.0379	0.0290	0.0444
Skew	-0.1404	0.1367	0.3266	0.3525	-0.2175
Kurt	1.950	1.593	1.393	2.087	0.2534

This paper removes the outliers and get cumulative returns of those five stocks in Fig. 1.

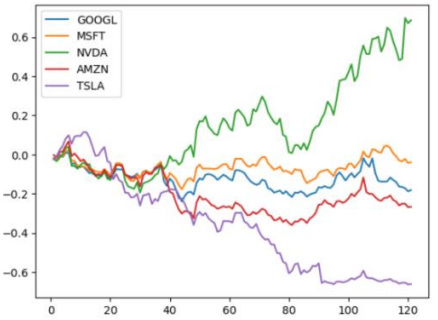

Figure 1: Cumulative returns of selected stocks.

DOI: 10.54254/2754-1169/25/20230504

3. Methods

3.1. Monte Carlo Model

Monte Carlo modelling is a stochastic simulation technique. Based on probability and statistical theory, the problem to be resolved is related to a certain probability pattern, and the computer is used to realize statistics or sampling, to obtain the numerical solution of the problem [8] (See Fig. 2).

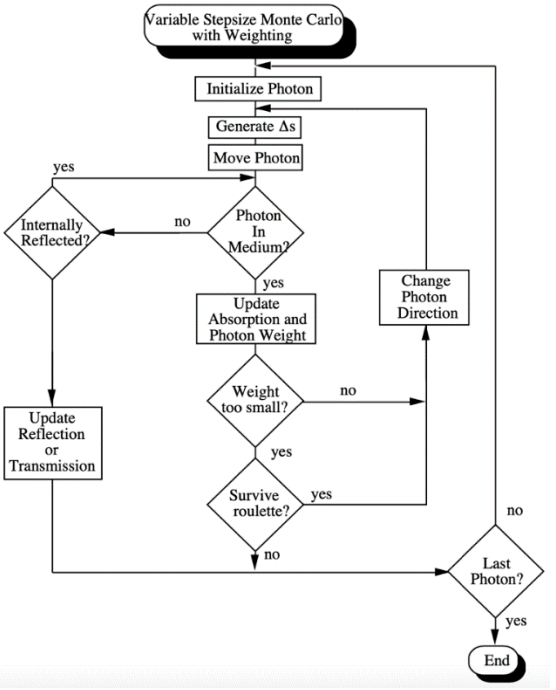

Figure 2: Monte Carlo.

3.2. Efficient Frontier

Efficient frontier is fundamental to contemporary investment theory. The efficient frontier curve plots risks and returns of different portfolios on a graph. The curve is upward sloping, reflecting the positive relationship between risk and return, but it also levels off at a certain point, indicating that additional risk does not lead to higher returns.

3.3. Mean-variance Analysis

Modern Portfolio Theory (MPT) is the portfolio optimization framework developed by Harry Markowitz in 1952. The optimization process used in MPT is often referred as Mean-Variance Optimization (MVO), which aims at helping investors optimize their portfolio by balancing risk and return.

Expected return is an average return that a portfolio of investment is expected by an investor. The weighted average of the expected returns of every asset in the portfolio is its calculation way.

$$Expected\ Return\ of\ Portfolio = \mathrm{E}(R_P) = W^T R \tag{1}$$

In MVO, risk is typically a symbol as the variance of the portfolio returns. The variance of a portfolio is the total of the weighted variances of every asset in the portfolio, plus twice the weighted covariance between each pair of assets.

DOI: 10.54254/2754-1169/25/20230504

$$Variance: \sigma_P^2 = \text{var}\left(\sum w_i r_i\right) = \sum w_i\, w_j \text{cov}(r_i r_j) \tag{2}$$

The Sharpe ratio is a prevalently used tool for measuring and evaluating the risk-adjusted return in investments [8]. When getting positive value, it means that the cumulative rate of return is higher than the risk. As for negative value, it means that the risk is greater than the cumulative rate of return.

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p} \tag{3}$$

The risk-free rate is Rf. The return of portfolio is Rp and $\boldsymbol{\sigma_p}$ is the standard deviation of the portfolio's return [9-10].

3.4. LSTM

LSTM is a new method to predict meaning Long Short-Term Memory, and it is a kind of artificial neural network which is commonly used in deep learning for processing and analyzing sequential data (See Fig. 3, Fig. 4 and Fig. 5).

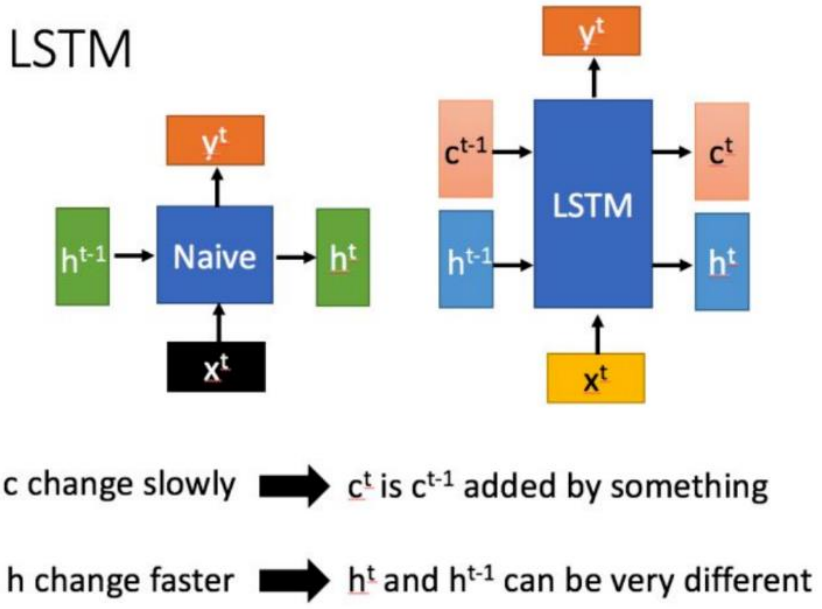

Figure 3: LSTM model.

There are four main points in LSTM which are Time_step, Lstm_units, Epochs and Batch_size. Time_step is steps of time including each input sequence. It is a better example NLP to understand it. Suppose you have a sentence to process, then here sample is one, which means one sentence to read, time_step is the number of words in that sentence. Lstm_units is the number of LSTM units in each layer of the model (default= [50, 50, 50]). For each num_units, LSTM network can regard it as a standard LSTM unit:

DOI: 10.54254/2754-1169/25/20230504

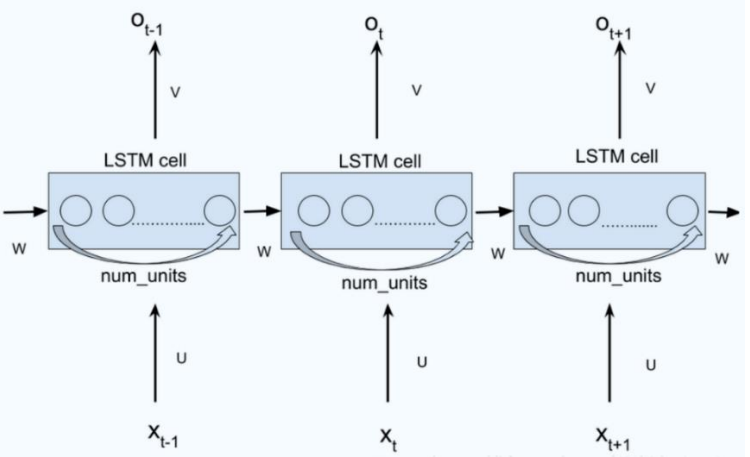

Figure 4: LSTM_units.

Epochs is the number of epochs to train the model (default=50). An epoch indicates that all data is sent to the network, and that a computational process before further retro propagation is completed. The number of epochs is significant in prediction. When epochs are small, the graph is non-fitting status. When epochs lifts, the iterations grow, and the graph becomes optimal fitting state and eventually enters an over-fitting status.

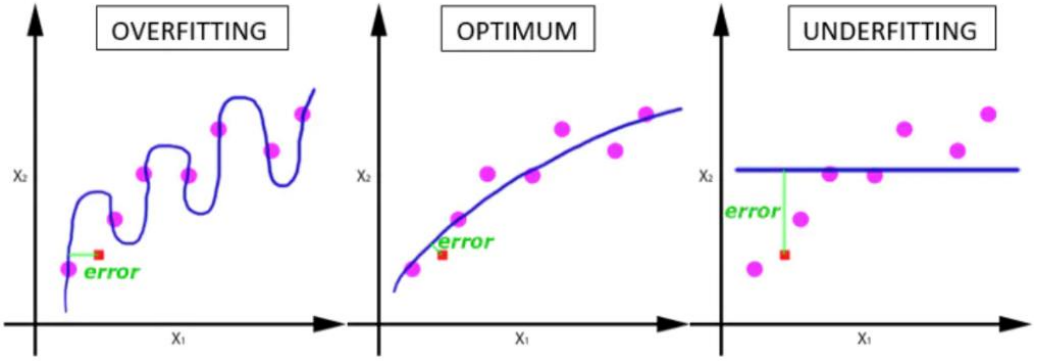

Figure 5: Different State.

Batch_size is the batch size to use during training (default=64). There are two obviously advantages. The biggest benefit is to make computer run at full capacity, which increases the speed. The second one is to make the direction of gradient descent more accurate.

$$Number\ of\ Batches = \frac{Training\ Set\ Size}{Batch\ Size} \tag{4}$$

4. Results

Firstly, from the size of the training set, the test set has the best effect when Time_Step is 5, Lstm_units is 50, Batch_Size is 64, and Epochs is 50. Also, this paper predicted data of five stocks by LSTM from January 11th, 2023, to March 27th, 2023, shown in Fig. 6.

DOI: 10.54254/2754-1169/25/20230504

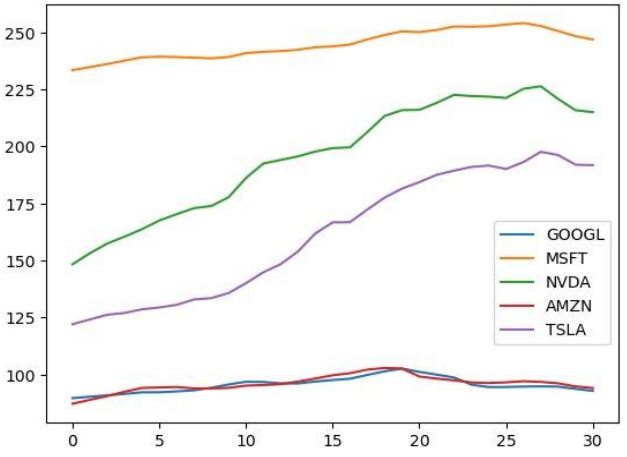

Figure 6: Predicted Data.

Secondly, this article uses the Monte Carlo model to assign 40,000 different portfolio weights to the results and shows the expected returns and volatility of these different portfolios in the same graph shown in the following Fig. 7.

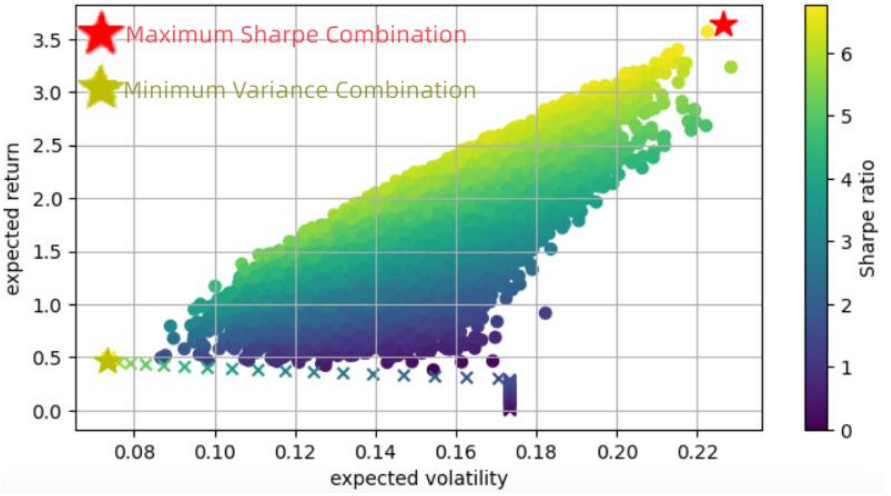

Figure 7: Monte Carlo Graph.

The circles Fig. 7 represent the combination distribution randomly generated by Monte Carlo, the cross marks represent the efficient frontier, the red star marks the highest Sharpe combination, and the yellow star marks the minimum variance combination. According to the portfolio investment model, we have two allocation options, the first is the maximum Sharpe ratio model, and the second is the variance minimum model. For the first case, choose 22% of Nvidia, 78% of Tesla, the expected return rate is 3.645, the expected volatility is 0.226, and the optimal Sharpe index is 16.111. In the model with the smallest variance, all choose Microsoft, and the expected return rate 0.47, volatility 0.073, Sharpe index 6.41. In the real situation, all the models with the largest Sharpe rate choose Nvidia, the expected return rate is 2.965, the expected volatility rate is 0.666, the optimal Sharpe index is 4.455, and the combination with the smallest variance is Google at 0.005, Microsoft at 0.994, and the expected return rate at 0.398. Rate 0.322, Sharpe index 1. 235.Based on these asset weights, this paper evaluates the portfolio performance and shows the results as follow in Fig .8.

DOI: 10.54254/2754-1169/25/20230504

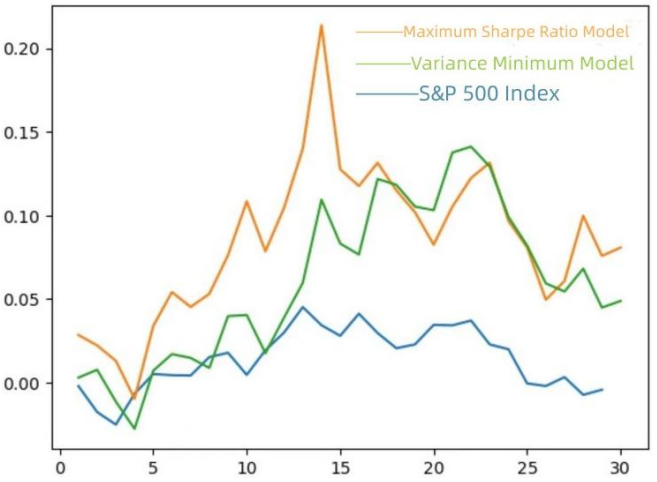

Figure 8: Cumulative Returns of two models and S&P 500 index.

The blue line represents the S&P 500. Clearly, both of our forecasts beat the S&P 500. So, Maximum Sharpe Ratio Model is better.

5. Conclusion

Portfolio is a scientific and popular model which maximizes returns and minimizes risks. This article selects five famous high tech companies' stocks: Google, Microsoft, Nvidia, Amazon, and Tesla from September 1, 2022, to March 1, 2023. First, the article uses LSTM model to learn the first 70% of the data and predicts the bottom 30% of the data. Second using the portfolio pattern for the forecast results, the paper gets the maximum return model and the minimum risk model on the effective frontier of Monte Carlo. Finally, by applying two models to the portfolio and combining two results comparing the yield with the S&P 500 within the same time frame, the paper draws the conclusion. This article provides a portfolio method using the latest LSTM model as a forecasting model, and the comparison reflects the power of the joint use of the LSTM model and the portfolio model.

References

[1] *Archer, N. P., Ghasemzadeh, F.: Project portfolio selection techniques: a review and a suggested integrated approach (1996).*
[2] *ASLAN, Y., ÖZKAN, Ö.: Innovative Approaches to Accounting, Finance and Auditing-4 (2022).*
[3] *Markowitz, H.: Foundations of portfolio theory. The journal of finance 46(2), 469-477 (1991).*
[4] *Li, B., Shu, Y., Sun, Y., Teo, K. L.: An optimistic value–variance–entropy model of uncertain portfolio optimization problem under different risk preferences. Soft Computing 25, 3993-4001 (2021).*
[5] *Kenig, E.: Portfolio selection in non-stationary markets. Algorithmic Finance 9(1-2), 35-47 (2021).*
[6] *Lööf, H., Sahamkhadam, M., Stephan, A.: Incorporating ESG into optimal stock portfolios for the global timber & forestry industry. Royal Institute of Technology (490), (2022).*
[7] *Ma, Y., Han, R., Wang, W.: Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications 165, 113973 (2021).*
[8] *Prahl, S. A.: A Monte Carlo model of light propagation in tissue. Dosimetry of laser radiation in medicine and biology 10305, 105-114) (1989).*
[9] *Amenc, N., Martellini, L., Vaissié, M.: Benefits and risks of alternative investment strategies. Journal of Asset Management 4, 96-118 (2003).*
[10] *Markowitz, H., Todd, G. P.: Mean-variance analysis in portfolio choice and capital markets. John Wiley & Sons (66), (2000).*