Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/25/20230499

LSTM-based Portfolio Optimization Strategy for SP500

Lyucheng Dong[1,a,*]

[1]Shanghai Jiao Tong University, Shanghai 200241, China
a. david_dong@sjtu.edu.cn
**corresponding author*

Abstract: Portfolio optimization is a perennial topic in the field of finance and recent breakthrough in deep learning techniques offers a new perspective to tackle it. This study selects 30 stocks of SP500 in different sectors through various constraints and deploys Long Short-Term Memory and Ledoit-Wolf Shrinkage to estimate returns and covariance respectively. The target portfolio is then obtained by inputting the predicted results into the mean-variance model, which is dynamically updated on a daily basis given evolving market information. The results show that the target model this study proposed surpasses the market benchmark (SP500), 1/N portfolio, and other mean-variance variants in terms of numerous financial metrics. Moreover, the target model exhibits volatility invariance and the capability to mitigate risk while extracting returns. This study showcases the revolutionary and promising applications of deep learning in the financial industry, shedding light on novel portfolio allocation strategies for risk-averse investors seeking stable positive returns in turbulent markets.

Keywords: long short-term memory, portfolio optimization, mean-variance

1. Introduction

Since Markowitz initially put forth the classical mean-variance (MV) theory in 1952, it has served as the cornerstone of contemporary portfolio theory. Markowitz's work provides a rational basis for portfolio management decisions by quantifying the risk-return trade-off. Since then, there has been extensive research into developing model variants that are more adaptable to real-life conditions [1]. However, the limitation of MV model is also clear: since the traditional MV model only exploits past information, it can only present the optimal strategy up to the data input [2]. Moreover, the famous Efficient Market Hypothesis (EMH) stipulates that the stock prices already contain and reflect all available information, and therefore in theory there exists no techniques that can produce excess economic profits in the long run [3]. For a long time, there has been a debate about whether the daily stock price is predictable for its intrinsically chaotic, non-parametric properties [4]. The once-dominated theory of EMH underwent skeptics when more and more economists came to believe that at least some predictable patterns that could lead to excess market profit exist [4]. Furthermore, significant incidents in the financial sector have brought spotlight on the significance of diversity. For instance, the early 2021 GameStop short squeeze illustrated the dangers of holding large amounts of stock in one company. The pandemic-induced market meltdown also underscored how crucial diversification and risk management are when during portfolio construction [5]. Therefore, nowadays, it remains crucial to delve into the subject of portfolio allocation so that investors could better employ it to mitigate risks while boosting profits. And the most pressing issue among all is how to accurately predict future returns so that the mean-variance model can perform perfectly in reality.

Thankfully, the recent advent of advanced machine learning techniques proffers a novel solution to tackle the issue of unpredictability and volatility [4, 6]. Furthermore, studies have shown that a hybrid model of MV with machine learning has a more stellar performance than traditional models [7]. However, with regard to the application of LSTM, few studies aim to integrate LSTM into the portfolio optimization process and the viability of this approach. The majority of the prior studies are put emphasis on the applicability and superiority of LSTM on stock prediction. For instance, Roondiwala et al applied LSTM to predict market indices and discussed accuracy [8]. Fischer & Krauss deployed LSTM networks on portfolios composed of SP500 constituents but primarily emphasized on the contrast with other memoryless machine learning models such as logistic regression and random forest [9]. Some scholars took an innovative departure from classical forecasting and treated it as a classification problem. For example, some studies have focused on using machine learning algorithms to predict the trajectory trend of stock market prices, including neural networks and tree-like classifiers, rather than a regression problem of predicting closing price itself [4, 10].

The purpose of this research is to leverage machine-learning techniques to generate a more informed prediction of future returns and covariance in order to facilitate more effective asset allocation strategies. To achieve this aim, the study first selects 30 distinct sector stocks from the latest SP500 constituents, adhering to certain constraints. The study uses the previous 70 days of stock price data to train a LSTM neural network to project the next day's stock prices. The projection can be easily shifted into percent change, i.e., daily returns, and a shrinkage method is applied to calculate the covariance matrix. The mean-variance optimization method is then utilized to derive the optimal portfolio weights for each day. The study iterates the above step for the next 30 days, updating the portfolio weights each day based on the latest predictions. Ultimately, after the 30-day period has elapsed, the overall portfolio returns are calculated and benchmarked against the returns generated by the 1/n portfolio (equal weights for all stocks) and the SP500 index.

The structure of the remainder of this paper is as follows. as follows. Section 2 presents the data utilized in this study and delineates the stock selection process, followed by a descriptive overview of the selected stocks. In Section 3, the methods employed in this study are elaborated in detail, including Long Short-Term Memory, Ledoit-Wolf shrinkage, and other techniques. In Section 4, the article scrutinizes the efficacy of the proposed approach, drawing comparisons with the benchmark asset and other naïve portfolios. Lastly, Section 5 concludes the paper and deliberates on potential avenues for future research.

2. Data Source and Pre-process

The daily stock data used in this article is acquired by the Python package yfinace (https://finance.yahoo.com/) and provided by Yahoo Finance. This study aims to enhance the performance of the SP500 index by constructing an investment portfolio using the latest SP 500 component stocks obtained from Wikipedia. Stocks are selected based on their average daily returns during the 70-day training period, ranked in descending order. The top-performing stocks are chosen accordingly, provided that they also satisfy the following rules and constraints:

(1) To ensure sectoral diversification, a minimum of two and a maximum of five stocks shall be selected from each Global Industry Classification Standard (GICS) sector.

(2) To avoid stocks with high volatility or instability, stocks that experienced a consecutive three-day decline in the past seven days are ruled out. This constraint eventually rules out 39 potential stocks.

A well-diversified portfolio of 30 stocks with the greatest potential to outperform the SP 500 index is constructed. The stocks selected ultimately are listed in the Table 1 below.

Table 1: 30 Stocks Selected for Portfolio Optimization.

Stocks				
NFLX	IPG	HAL	SLB	BKR
RCL	WYNN	CCL	NKE	CZR
LW	KHC	HPE	ORCL	GE
UHS	ISRG	IDXX	GILD	BA
MKTX	IVZ	FCX	STLD	CAT
SPG	EQIX	PCG	AWK	UAL

This study examines a period of 102 market days, from October 1, 2022, to February 28, 2023. The model is trained using the first 70 days of this period, from October 1, 2022, to January 10, 2023. The portfolio's performance is then evaluated during the remaining 32 market days, from January 11, 2023, to February 28, 2023. This article opts for 'Adj Close' as the daily price V_t^i and computes the daily returns of the i-th particular asset R_{it} on day t as follows,

$$R_{it} := \frac{V_{t+1}^i}{V_t^i} - 1 \tag{1}$$

3. Methods

The method used in the study consists of four main steps, which are shown in Fig. 1. Firstly, the study selects 30 stocks from the SP500 index based on certain constraints. Secondly, the study trains a LSTM neural network using the previous 70 days of stock price data to predict the next day's stock prices. Meanwhile, the study utilizes the shrinkage method to estimate the covariance of the assets. Thirdly, the study uses the mean-variance optimization method to calculate the optimal portfolio weights for each day. The above procedures fully describe the target model this study proposes, which is denoted as Model 3. Moreover, this study also entertains the other two MV variants for comparison purposes, namely Model 1 and Model 2. The first one only incorporated the shrinkage method with traditional MV, whereas the latter also deployed exponentially weighted mean (EWM) on the estimation of feature returns to reflect the trend of the assets' price. Lastly, after 32 days of daily portfolio updates, the study compares the performance of the optimized portfolio provided by Model 3 with that of four benchmarks, SP500, a 1/n portfolio (equal weights for all stocks), Model 1, and Model 2 (See Figure 1).

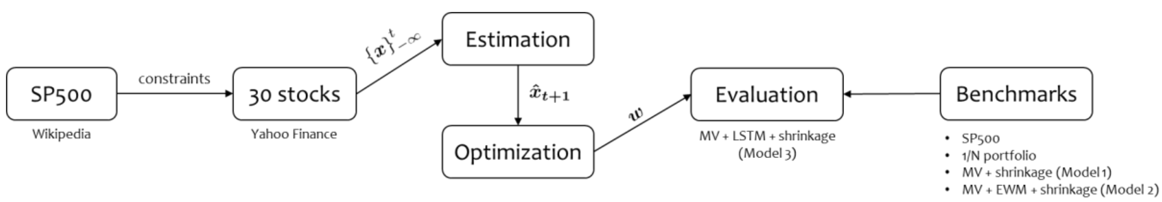

Figure 1: Flowchart of the Study.

3.1. Mean-Variance Model

The MV model provides a mathematical framework for obtaining the optimum weights for each asset for the investor [11]. The key insight of MV is to find the best portfolio that gives the maximum returns for a given rate of risk. Let w_i be the weight of the i-th asset such that $\sum_i w_i = 1$ and μ_i be the expected return of the i-th asset. Then the expected returns of the portfolio read

$$\mu_p = \sum_i w_i \mu_i \tag{2}$$

Denote σ_i as the standard deviation of the i-th asset and ρ_{ij} as the correlation between the returns of the i-th and j-th asset. Then the portfolio return variance is

$$\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij} \tag{3}$$

The most ideal portfolios are those that are situated on the efficient frontier. Also, the above equations can be rewritten in matrix form, which is more convenient to implement in programming and calculate efficient frontiers. The objective function for MV is to

$$min\, w^\top \Sigma w - qR^\top w \tag{4}$$

where w is a vector of asset weights and Σ is the sample covariance matrix for asset returns. The latter is called the sample covariance because it is computed with historical data directly. Some modifications will be introduced in the subsequent part to enhance the performance of the covariance matrix. The parameter q measures investor's risk tolerance.

In the MV model, two portfolios are interesting: the minimum volatility portfolio, and the maximum Sharpe ratio portfolio. The minimum volatility portfolio is where $q = 0$, i.e., the investor is completely averse to risk. The Sharpe ratio is a popular metric to evaluate risk-adjusted return, which is calculated as:

$$Sharpe\ ratio = \frac{R_p - R_f}{\sigma_p} \tag{5}$$

where R_f is the current risk-free rate of the market. This study uses the maximum Sharpe ratio portfolio as target portfolio and compares its performance with other benchmarks, such as SP500, the 1/N portfolio, and other mean-variance variants.

3.2. Exponentially weighted mean

Exponentially weighted mean (EWM) is a statistical method used to calculate a rolling average of time-dependent data. EWM assigns past observations exponentially decreasing weights as opposed to a conventional moving average, which offers equal weights to all data. This means that more recent data points are given greater weight, while older observations have a lesser impact on the average. The formula for calculating EWM is as follows:

$$EWM_t = (1 - \alpha) * EWM_{t-1} + \alpha * X_t \tag{6}$$

where EWM_t is the exponentially weighted moving average at time t, X_t is the value of the time series at time t, and α is a smoothing factor. The smoothing factor α is a value between 0 and 1 that determines the rate at which the weights decay.

3.3. Long Short-Term Memory

The capacity of LSTM to manage long-term dependencies and its improved performance as compared to conventional RNNs have led to its widespread use in a various discipline. The input gate, forget gate, and output gate are three gates that the LSTM introduces to steer the flow of information through and through the memory cell. The model may selectively recall or forget information using the memory cell depending on the input and the state of the cell at the time. The equations for the LSTM gates are as follows:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) && \text{(Input)} \\ f_t &= \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) && \text{(Forget)} \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) && \text{(Output)} \end{aligned} \tag{7}$$

where x_t is the input at time t, h_{t-1} is the hidden state at time $t-1$, $W_i, U_i, b_t, W_f, U_f, b_f, W_o, U_o, b_o$ are the learnable parameters, and σ is the sigmoid activation function. The memory cell is updated based on the input, the previous cell state, and the forget gate and input gate outputs:

$$\begin{aligned} &\text{Candidate cell state} && \widetilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \\ &\text{Cell state} && C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t. \end{aligned} \tag{8}$$

The output of the LSTM is composed of two parts: one is the current hidden state, and the other is from the output gate:

$$\text{Hidden state} \qquad h_t = o_t \odot \tanh(C_t) \tag{9}$$

The above constitutes a single unit of LSTM cell. Each cell transfers its cell state and the hidden state to the next cell. Multiple LSTM modules are connected to form an LSTM layer see Fig. 2. The machine learning model used in the paper is constructed using the TensorFlow Sequential API and consists of two LSTM layers, each with 100 units, and threefold dense layers, each with 100, 50, and 1 connection, respectively. The use of two LSTM layers allows the model to effectively capture long-term relationship that is latent in the data. Meanwhile, the dense layers process non-time-series traits of the data and map the outputs from the LSTM layers to the desired output format (See Figure 2).

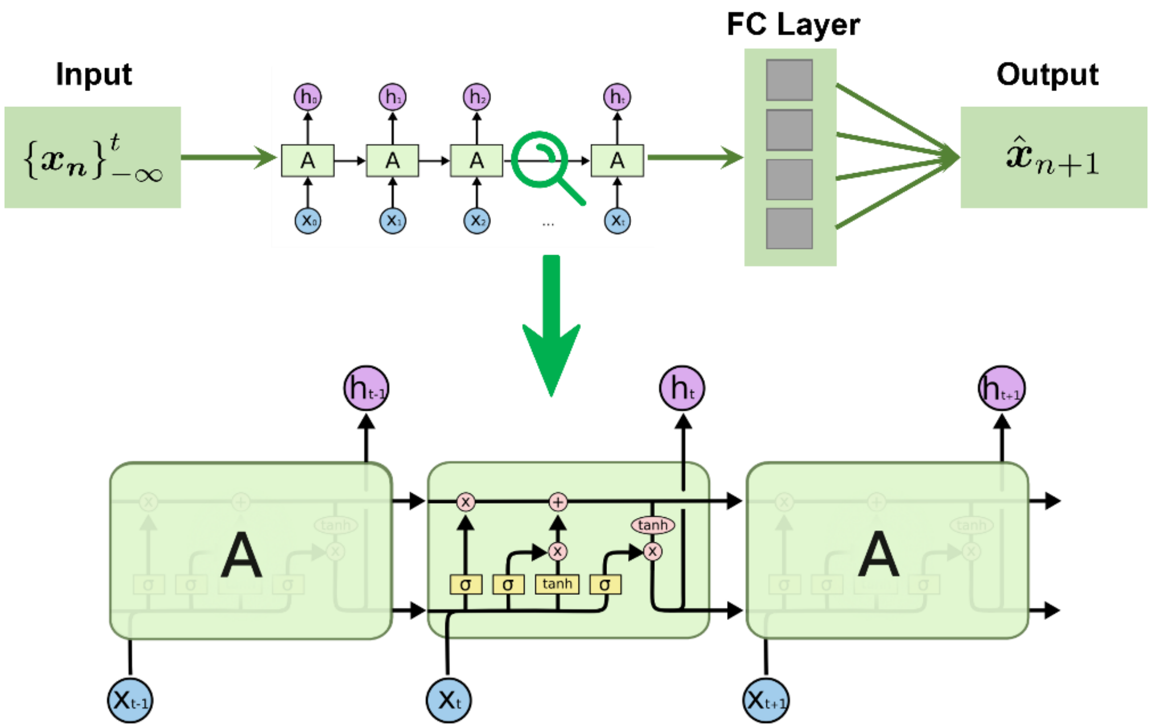

Figure 2: Structure of LSTM.

This study assumes that the most recent data should contain the most relative information to forecast the asset price for the subsequent day. Thus, the model is trained in the following manner:

DOI: 10.54254/2754-1169/25/20230499

Inputs: a vector of 7-dimension, $x_t = (x_{t-6}, \cdots, x_t) \in R^7$, where x_i donotes the adjusted closing price for day t

Outputs: prediction for the next day, $\widehat{x_{t+1}}$

The predicted returns of each asset are updated daily.

3.4. Ledoit-Wolf Shrinkage

The covariance matrix, which measures the pairwise correlation between asset returns, is a crucial input in the optimization process. Nevertheless, for issues such as small sample size, noisy data, and the non-stationarity of the stochastic process, simply deriving the covariance matrix from historical data might be problematic. The basic notion of the shrinkage method proposed by Ledoit and Wolf is to make a compromise between sample covariance matrix S and structured estimator F. The former is computationally cost-effective and unbiased, while the latter is of less estimation error. By determining the optimal shrinkage constant δ with cross validation, a new covariance matrix is constructed:

$$\Sigma_{LW} = \delta F + (1 - \delta)S \tag{10}$$

4. Results

The study first investigated the fitting performance of LSTM. After tuning the model parameters, the ultimate LSTM model is structured with three input layers, each with 100 dimensions, followed by three dense layers with interface sizes of 100, 50, and 1, respectively, using mean squared error (MSE) as the evaluation metric and trained for 10 epochs. Figure 3 demonstrated partial performance of the LSTM model. Overall, the model is somewhat accurate in predicting the future values of the asset price to some extent, as evidenced by the close alignment between the predictions and the validation values of the test set.

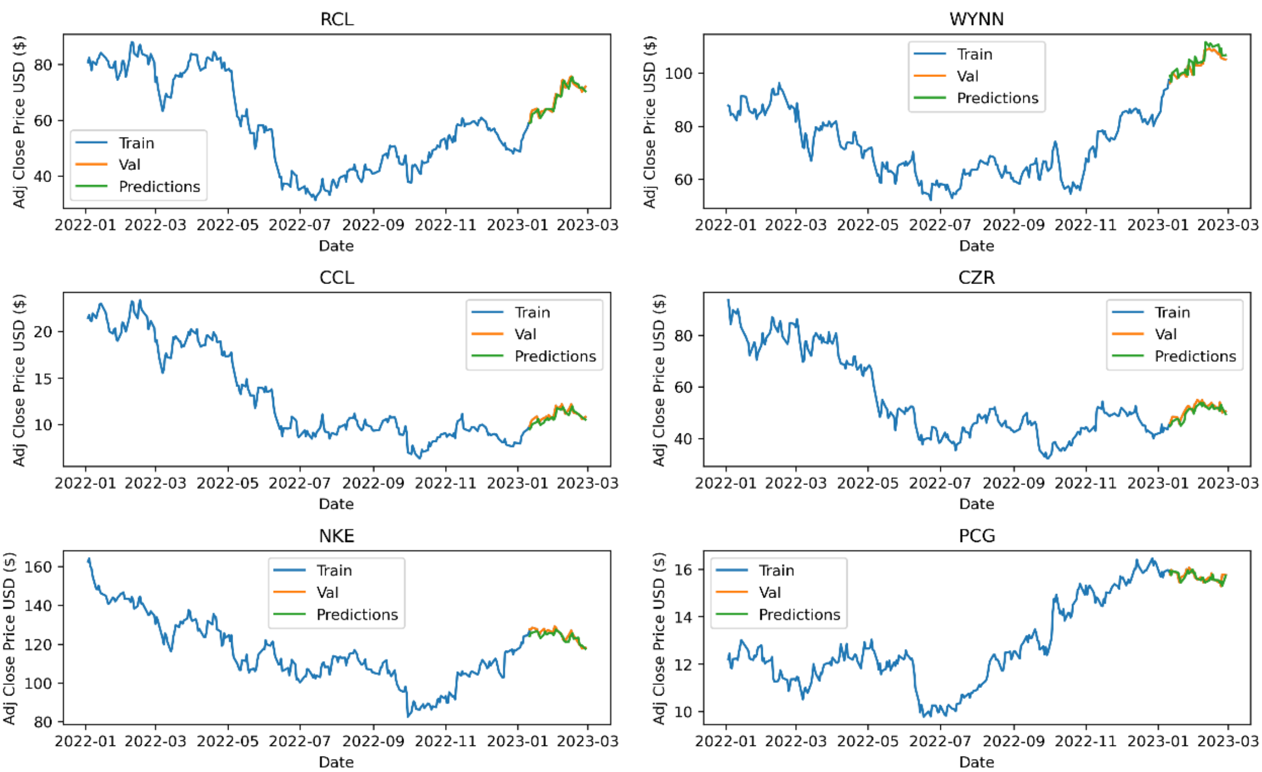

Figure 3: Partial demo of LSTM performance.

The strategy this study proposes trades on a daily basis. Therefore, during the 32-day testing period, the asset weights for each day are determined based on the price information up to the previous day, using the predictions made by LSTM. When each predicted return is lower than the risk-free rate, all asset weights are automatically set to 0, as any combination of them would be strictly dominated by the risk-free asset. In this way, the portfolio weights on each asset on each day is determined, and ergo the portfolio itself is determined.

To evaluate how each model perform, the historical returns of SP 500 during the test period is acquired to serve as market benchmark. Then, the study conducts an ex-post analysis to ascertain the actual returns of each portfolio. An annualized tear sheet of each portfolio is listed in Table 2 and cumulative returns of each portfolio is plotted in Figure 4.

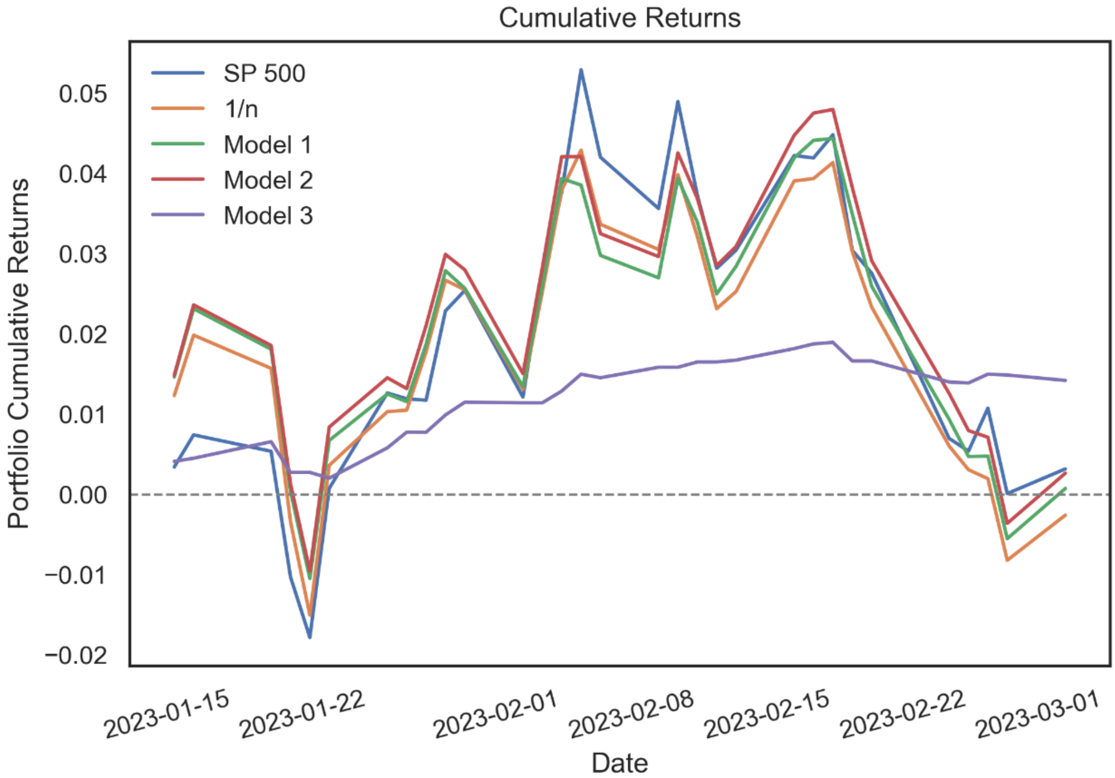

Figure 4: Comparison between SP 500 returns and alternative portfolios.

As shown in Table 2, one could observe such following traits:

1/N portfolio (equally weighted portfolio) is not ideal since it is the only portfolio that gives a negative return of -2.1%. Meanwhile, it lacks the ability to mitigate volatility (15.3%). The same could be said about Model 1, with its barely positive return (0.6%) and bearish risk control (volatility of 15.2%).

The performance of Model 2 is mediocre, roughly on par with the SP 500 with returns of 2.2%, volatility of 15.5% and similar Sharpe Ratio (0.142). While it achieves the basic goal of asset allocation, it does not outperform the market.

Model 3, the target model this study proposes, is has more superior performance. On the one hand, it achieves returns of 12.3%, which is roughly 6 times that of the SP500; on the other hand, it exhibits excellent risk control, resulting in a nearly sevenfold reduction in risk compared to the SP500 and max drawdown of merely -0.5%. Note that in Fig. 4, Model 3 demonstrated a gradual yet steady increase in cumulative returns, whereas other portfolio (asset) experiences extreme fluctuations every

DOI: 10.54254/2754-1169/25/20230499

four or five days. One could say that Model 3 is relatively invariant to volatility. This is largely attributed to the flexibility of daily trading and the accuracy of LSTM predictions.

Table 2: Annualized Tear Sheet for Jan 12 to Feb 27, 2023.

Portfolio	SP 500	1/N	Model 1	Model 2	Model 3
Returns	2.5%	-2.1%	0.6%	2.2%	12.3%
Volatility	16.1%	15.3%	15.2%	15.5%	2.6%
Sharpe Ratio	0.155	-0.137	0.040	0.142	4.731
Max Drawdown	-5.0%	-4.9%	-4.8%	-4.9%	-0.5%

Note:
Model 1: mean-variance + covariance shrinkage
Model 2: mean-variance + exponentially-weighted mean + covariance shrinkage
Model 3: mean-variance + LSTM + covariance shrinkage

5. Conclusion

In summary, this paper offers a novel portfolio optimization strategy that introduces LSTM and shrinkage techniques into the traditional mean-variance model. The study first utilizes a series of rules to select 30 constituent stocks from the SP500 index. The LSTM model is used to predict stock prices, while a shrinkage method is used to estimate covariance, and a mean-variance framework is used to construct a portfolio with daily updated weights. The proposed portfolio is then compared to the market benchmark (SP500), equally weighted portfolio (1/N portfolio) and other mean variance variants (Model 1, Model 2). The flexibility of daily trading, combined with the accuracy of LSTM predictions, has contributed significantly to the success of the model. The cumulative return graph and the tear sheet reveal a notable outperformance against other benchmarks. The findings demonstrate the excellence and the relative volatility invariance of the suggested approach.

Overall, this approach represents a compelling solution for portfolio optimization, as it leverages advanced techniques to effectively manage risk and generate returns. However, this study does not fully consider realistic constraints such as transactional fees. It is also salubrious to investigate whether the strategy remains valid in a wider time horizon and with a higher trading frequency. In this sense, the proposed model can be further refined and expanded in future studies to achieve even better results.

References

[1] *Kalayci, C. B., Ertenlice, O., Akbay, M. A.: A comprehensive review of deterministic models and applications for mean-variance portfolio optimization. Expert Systems with Applications 125, 345–368 (2019).*

[2] *Chaweewanchon, A., Chaysiri, R.: Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning. International Journal of Financial Studies 10(3), 64 (2022).*

[3] *Jensen, M. C.: Some anomalous evidence regarding market efficiency. Journal of Financial Economics 6(2), 95–101 (1978).*

[4] *Basak, S., Kar, S., Saha, S., Khaidem, L., Dey, S. R.: Predicting the direction of stock market prices using tree-based classifiers. The North American Journal of Economics and Finance 47, 552–567 (2019).*

[5] *Zhang, D., Hu, M., Ji, Q.: Financial markets under the global pandemic of COVID-19. Finance Research Letters 36, 101528 (2020).*

[6] *Dixon, M. F., Igor, H., Paul, B.: Machine Learning in Finance. Berlin and Heidelberg: Springer International Publishing (2020).*

[7] *Chen, W., Zhang, H., Mehlawat, M. K., Jia, L.: Mean–variance portfolio optimization using machine learning-based stock price prediction. Applied Soft Computing 100, 106943 (2021).*

[8] *Roondiwala, M., Patel, H., Varma, S.: Predicting stock prices using LSTM. International Journal of Science and Research (IJSR) 6(4), 1754-1756 (2017).*

DOI: 10.54254/2754-1169/25/20230499

[9] *Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 270(2), 654–669 (2018).*

[10] *Khaidem, L., Saha S., Roy, D. S.: Predicting the direction of stock market prices using random forest. arXiv preprint arXiv:1605.00003 (2016).*

[11] *Markowitz, H.: Portfolio Selection. The Journal of Finance 7(1), 77–91 (1952).*