An Empirical Analysis on the Portfolio of Transnational Auto Market Based on ARIMA-LSTM

Yiping Hong(✉)

School of Economics and Management, Hainan University, Haikou 570228, Hainan, China
1114010627@st.usst.edu.cn

Abstract. Due to the rise of raw material prices, the reduction of automobile production capacity caused by COVID-19, the trend of global electrification and other factors, the investment in transnational automobile stocks has gradually become a new target for global investors. Because the general stock price data has both linear and non-linear characteristics, the integrated model is much more favourable than the single models in prediction, resulting in more and more widespread application in financial time series. According to the above, a total of 10 groups of auto stock closing prices listed in the US, Japan, China, Germany and the UK from January 1, 2016 to December 31, 2021 were selected as research samples to establish ARIMA-LSTM model to forecast the closing price in the next 10 working days. Subsequently, the mean variance model is constructed, and the optimal portfolio is obtained in terms of the maximum Sharpe ratio. The results illustrate that: (1) The accuracy of ARIMA-LSTM integrated model is validated to more excellent than that of single ARIMA and LSTM; (2) Based on the stock closing prices in the next 10 working days predicted by the combination model, the performance of the maximum Sharpe ratio portfolio is better than that of the market average.

Keywords: Transnational automobile stock price · ARIMA-LSTM · Mean-variance model · Sharpe ratio

1 Introduction

Due to economic globalization and financial integration, the allocation of financial resources will develop internationally, and the globalization trend of stock market volatility is becoming more and more obvious [1]. At the same time, under the impact of the rising prices of raw materials, production stoppage caused by the Corona Virus Disease 2019 (COVID-19), electrification and other factors, the valuation of the global automotive sector has increased, which has been widely concerned by global investors, so it is of practical significance to build an optimal portfolio by predicting the future transnational auto stock data.

At present, using hybrid models for predicting the future values has become a vitally important research topic, which has been applied to finance, securities, water quality and other fields. Bukhari et al. established autoregressive fractional integrated moving

V. Gaikar et al. (Eds.): FMET 2022, AEBMR 227, pp. 130–142, 2023.
https://doi.org/10.2991/978-94-6463-054-1_16

average and long-short term memory (ARFIMA-LSTM) model to forecast the abrupt stochastic variation of the financial market [2]. An integrated autoregressive integrated moving average (ARIMA) and neural network model is proposed to deal with linear and nonlinear characteristics of water quality time series for prediction by Faruk [3]. Autoregressive integrated moving average and long-short term memory (ARIMA-LSTM) model of housing sales is proved to have the best performance with the lowest error rate compared with single ARIMA and long-short term memory (LSTM) model [4]. Accurate prediction of stock prices can effectively help investors reduce risks and achieve better returns. A combination method bonding exponential smoothing model (ESM), ARIMA, and back propagation neural network (BPNN) is conducted to be the most advantageous for stock index forecasting by Wang et al. [5]. Hybrid Glosten, Jagannathan and Runkle & generalized autoregressive conditional heteroskedasticity (GJR–GARCH) approach provides higher prediction accuracy than other prediction approaches by Yi and Wang [6]. And Yulin & Du predict the future stock price index through autoregressive integrated moving average and back propagation (ARIMA-BP) neural network method to improve accuracy [7]. Besides, more and more literatures show that ARIMA-LSTM has advantages in prediction, for example, Choi found that the ARIMA-LSTM model displayed better prediction sensitivity of the stock price correlation coefficient than the other financial models on a significant scale [8]. Sakshi and Vijayalakshmi revealed that ARIMA-LSTM performed much better than other models [9]. Therefore, ARIMA-LSTM integrated model with greater complexity is increasingly used in the time series prediction.

In addition, Markowitz mean-variance model is widely used by investors to achieve portfolio. Lee et al. use Markowitz model to assess the performance of the Malaysia investment portfolio [10]. Širůček et al. construct the optimal portfolio on the US stock market by applying Markowitz portfolio theory [11]. Based on the calculated risks and returns, draw effective boundaries, then investors determine the appropriate portfolio based on their own risk tolerance.

This paper attempts to build an optimal cross-national portfolio based on the predicted price of auto stocks obtained from the ARIMA-LSTM hybrid model. First, the ARIMA-LSTM model is established, and the fitting accuracy of the combination model is more advantages than that of the single model according to evaluation indexes. Second, the optimal portfolio with the maximum Sharpe ratio is constructed by using the prediction value obtained from the hybrid model. Third, the performance of the above portfolio is superior by comparing with the average level of the market. The above shows that the portfolio model with the maximum Sharpe ratio based on the predictions of ARIMA-LSTM hybrid model is of positive significance for investors to improve returns and reduce risks.

This paper will elaborate data and methodology, construct the hybrid model, present the interpretation of the results in the next section, and finally draw the conclusion.

2 Data and Methodology

2.1 Data

From the top 20 auto companies in the world in reference to the market value by Wind on December 31, 2021, the closing price data of 10 auto stocks (TSLA.O, 7203.T, 1211.HK, GM.N, VOW.DY, F.N, 0NXX.L, BMW.DF, 2333.HK and 7267.T) from January 1, 2016 to December 31, 2021 (unified in US dollars) are selected for empirical analysis. The sample number of each stock is 1,566, and the data is from the Wind database. Table 1 shows the descriptive statistical results of 10 stocks. Through Augmented Dickey–Fuller (ADF) test, it is not only demonstrably proves the non-stationarity of the original close price sequence, but also certifies the series after 1st-order difference, namely diff (1) sequence, is stable. The p-values of Jarque-Bera (JB) test are significantly < 0.05, which proves that sequence after the first order difference does not meet the normal distribution. Through Ljung-Box (LB) test, the p-values are all < 0.05, which means high autocorrelation of the 1st-order difference sequences.

The Q-Q plot of diff (1) series is shown in Fig. 1. It is obvious that the upper and lower tails are far from the straight line, further proving the series have obvious thick tails and do not meet the normal distribution.

Table 1. Descriptive statistics of stock close prices

		TSLA.O	7203.T	1211.HK	GM.N	VOW.DY	F.N	0NXX.L	BMW.DF	2333.HK	7267.T
Close price	mean	1078.57	74.31	47.87	46.98	191.40	383.87	67.98	87.55	32.92	36.92
	25% quantile	251.24	66.10	24.52	38.61	155.47	338.82	56.90	80.42	16.60	34.18
	median	328.22	71.36	27.43	44.54	169.92	372.62	69.58	87.91	21.28	36.35
	75% quantile	1186.40	78.53	41.19	49.92	199.58	405.43	79.32	97.09	28.38	39.35
p-value of difference tests											
Close price	ADF	1.00	0.93	0.98	0.73	0.71	0.99	0.48	0.30	0.89	0.08
Diff (1) sequence	ADF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	JB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	LB	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.01

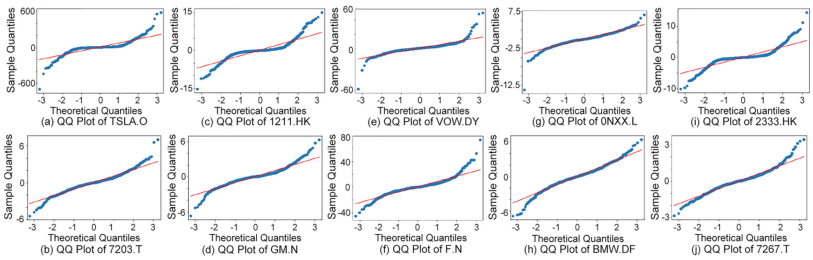

Fig. 1. Q-Q plot of diff (1) sequence

2.2 Methodology

Since the linear and non-linear characteristics of time series, linear model ARIMA and non-linear model LSTM are combined for prediction to better extract data information [12]. Figure 2 shows the research method including ARIMA-LSTM model and mean-variance model. First, the ARIMA models of 10 stock close price are established respectively, and the residual is obtained from the ARIMA results and the real value. Then the LSTM models are established for the residuals. Further, the ARIMA predictions and the LSTM residual predictions are added together, namely the predicted value of the ARIMA-LSTM integrated model.

Finally, based on the predictions obtained from ARIMA-LSTM hybrid model, the optimal portfolio is ultimately achieved by using the mean-variance model and the maximum Sharpe ratio.

3 Results

3.1 ARIMA-LSTM Model

3.1.1 Find Optimal (p, d, q)

Take TSLA.O as an example, the corresponding autocorrelation function (ACF) and partial autocorrelation function (PACF) of diff (1) sequence are intuitively presented in Fig. 3. Through the image, it can be seen that TSLA.O has obvious autocorrelation and partial autocorrelation, preliminarily judging that the stock price meet the standard of establishing ARIMA model. Similar to TSLA.O, the closing prices of the other 9 stock sets also showed significant autocorrelation and partial autocorrelation.

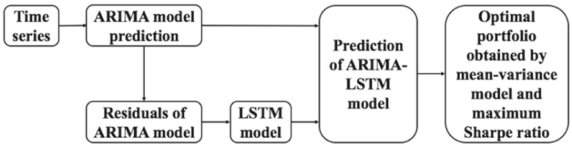

Fig. 2. The flow chart of the research method

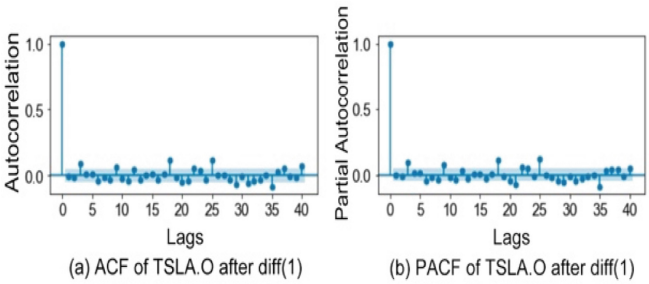

Fig. 3. ACF and PACF plot of TSLA.O

Secondly, the optimal ARIMA model is selected by focusing on the akaike information criterion (AIC), the bayesian information criterion (BIC), and the hannan-quinn information criterion (HQIC), and the optimal parameter (p, d, q) of all stocks are shown in Table 2.

3.1.2 ARIMA Establishment and Residual Test

After determining the optimal (p, d, q), the ARIMA model results of 10 stock sets are shown in Table 3. Then, the residuals are tested by LB test, of which the p-values are all > 0.05, indicating that residuals conform to white noise, which means that the models above are reasonable.

3.1.3 ARIMA-LSTM Model

Firstly, the residuals obtained by ARIMA model are normalized and scaled to $[-1,1]$, which are used as the input to the neural network including input, LSTM, hidden layer, full connection layer and output for training. Owing to the structural complexity of

Table 2. Optimal (p, d, q) of 10 stocks

Stocks	(p, d, q)	Stocks	(p, d, q)
TSLA.O	(4,1,4)	F.N	(2, 1, 2)
7203.T	(3,1,3)	0NXX.L	(4,1,4)
1211.HK	(6,1,0)	BMW.DF	(2,1,2)
GM.N	(4,1,4)	2333.HK	(3,1,4)
VOW.DY	(5,1,5)	7267.T	(0,1,1)

Table 3. ARIMA model results of 10 stocks

Stocks	ARIMA model	Residual p-value
TSLA.O	$Price_t = 3.23-0.64Price_{t-1}+0.44Price_{t-2}-0.51Price_{t-3}-0.89Price_{t-4}+_t-0.65_{t-1}+0.44_{t-2}-0.60_{t-3}-0.97_{t-4}$	0.18
7203.T	$Price_t = 0.02+1.16Price_{t-1}-1.21Price_{t-2}+0.83Price_{t-3}+_t+1.21_{t-1}-1.25_{t-2}+0.90_{t-3}$	0.55
1211.HK	$Price_t = 0.08-0.06Price_{t-1}+0.05Price_{t-2}-0.09Price_{t-3}-0.05Price_{t-4}+0.05Price_{t-5}+0.09Price_{t-6}$	0.62
GM.N	$Price_t = 0.02-0.49Price_{t-1}+0.24Price_{t-2}-0.69Price_{t-3}-0.83Price_{t-4}+_t-0.49_{t-1}+0.24_{t-2}-0.71_{t-3}-0.79_{t-4}$	0.35
VOW.DY	$Price_t = 0.09-1.67Price_{t-1}-1.69Price_{t-2}-1.74Price_{t-3}-1.59Price_{t-4}-0.80Price_{t-5}+_t-1.61_{t-1}-1.56_{t-2}-1.65_{t-3}-1.54_{t-4}-0.72_{t-5}$	0.19
F.N	$Price_t = 0.25-0.38Price_{t-1}-0.88Price_{t-2}+_t-0.35_{t-1}-0.83_{t-2}$	0.06
0NXX.L	$Price_t = -0.005+0.14Price_{t-1}+0.22Price_{t-2}+0.19Price_{t-3}-0.94Price_{t-4}+_t+0.11_{t-1}+0.18_{t-2}+0.21_{t-3}-0.93_{t-4}$	0.47
BMW.DF	$Price_t = -0.004+1.47Price_{t-1}-0.90Price_{t-2}+_t+1.42_{t-1}-0.87_{t-2}$	0.56
2333.HK	$Price_t = 0.04-1.01Price_{t-1}+0.76Price_{t-2}-0.40Price_{t-3}+_t+1.06_{t-1}-0.84_{t-2}+0.59_{t-3}-0.20_{t-4}$	0.38
7267.T	$Price_t = 0.0004+_t+0.09_{t-1}$	0.26

Table 4. The final parameters of residual from 10 stocks in LSTM model

Stocks	Number of iterations	Stocks	Number of iterations
TSLA.O	450	F.N	50
7203.T	100	0NXX.L	450
1211.HK	100	BMW.DF	50
GM.N	100	2333.HK	250
VOW.DY	150	7267.T	200

LSTM model, the setting of parameters plays a valuable role in improving the accuracy of prediction, so in the 10 LSTM models for training residual, setting the number of iterations to [50,100,150,200,250,300,350,400,450,500], and using the minimum mean square error (MSE) value to confirm best iterations of 10 stocks. In the meanwhile, Adam optimization function is applied, and the initial learning rate is 0.01.

Then, after setting the first 80% of the data as the training set, and the last 20% of the data as the testing set, the final iterations of residuals from 10 stock groups, which are shown in Table 4, are determined through comprehensively considering the root mean square error (RMSE) and mean absolute error (MAE), and the loss decreases significantly after iteration (Fig. 4).

Subsequently, the ARIMA predictions and LSTM residual predictions are added to obtain the predictions of ARIMA-LSTM model. Figure 5 is the comparison graph of true value and corresponding predictions in the test set. It is found that the lines obviously coincide, indicating good fitting effect, intuitively showing that ARIMA-LSTM hybrid model has practical significance in predicting the closing price of stocks.

3.1.4 Comparation

For the purpose of verifying the prediction ability of these three models (ARIMA-LSTM, ARIMA, LSTM), RMSE and MAE, suggested by Dutta et al. [13], are used to measure the prediction effect. The RMSE and MAE values are calculated according to the following formulas:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \tag{2}$$

where y_t and $\hat{y}_t$ are the actual and prediction values, at time t, respectively.

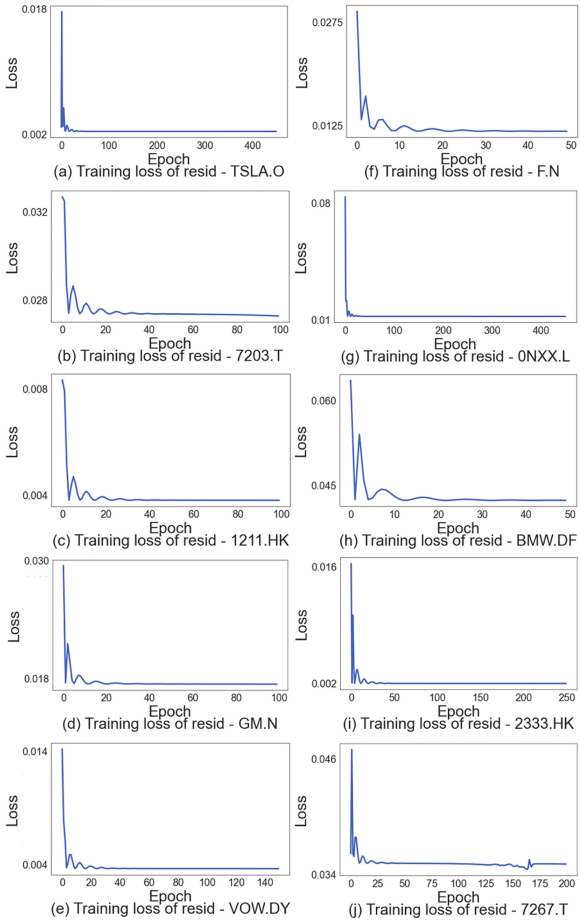

Fig. 4. Loss and epoch of residual from 10 stocks in LSTM model

By comparison, RMSE and MAE results of ARIMA-LSTM are the smallest among the three models, which means the prediction of ARIMA-LSTM hybrid model is much more accurate than that of single ARIMA model and LSTM model, implying that the combination models of 10 stock sets definitely show clear outperformance. Herein, RMSE and MSE values of ARIMA-LSTM, ARIMA and LSTM models are show in Table 5.

3.1.5 Prediction

According to ARIMA-LSTM, the closing prices of 10 stocks in the next 10 working days are forecasted. The predictions are visually displayed in Fig. 6.

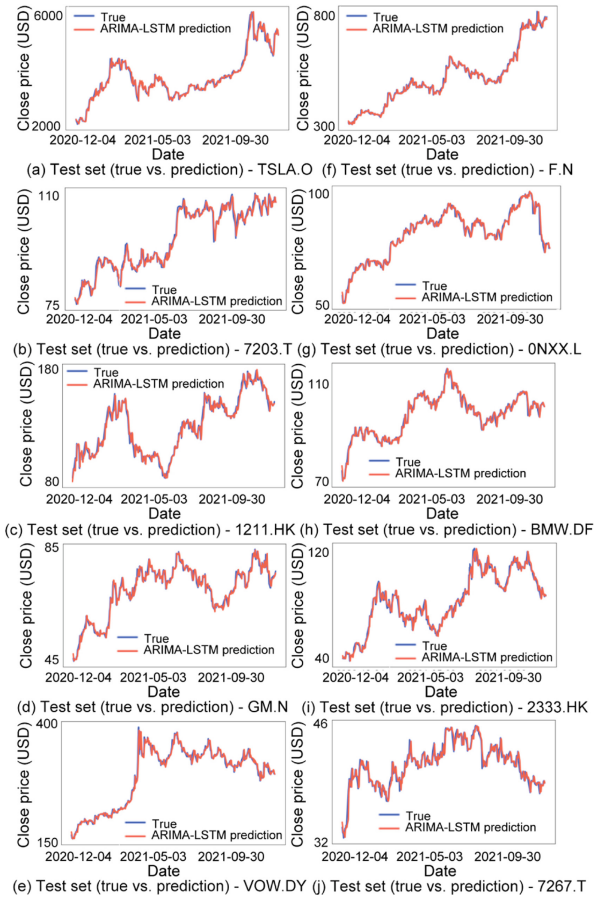

Fig. 5. True value vs. prediction by ARIMA-LSTM in test set

Table 5. Comparation of different models

	Model	TSLA.O	7203.T	1211.HK	GM.N	VOW.DY	F.N	0NXX.L	BMW.DF	2333.HK	7267.T
RMSE	ARIMA-LSTM	130.62	1.35	4.54	1.62	8.55	13.47	1.60	1.61	3.48	0.71
	ARIMA	1543.34	19.64	52.76	26.61	123.35	276.89	32.49	28.09	43.63	8.17
	LSTM	1044.85	9.06	91.96	10.66	44.91	157.70	1.76	2.84	49.50	0.72
MAE	ARIMA-LSTM	91.85	1.01	3.40	1.20	5.43	9.44	1.10	1.20	2.62	0.53
	ARIMA	1327.06	17.52	47.23	25.11	108.34	237.39	29.89	26.01	38.59	7.70
	LSTM	851.25	7.32	89.09	9.52	36.78	124.67	1.19	2.19	45.92	0.54

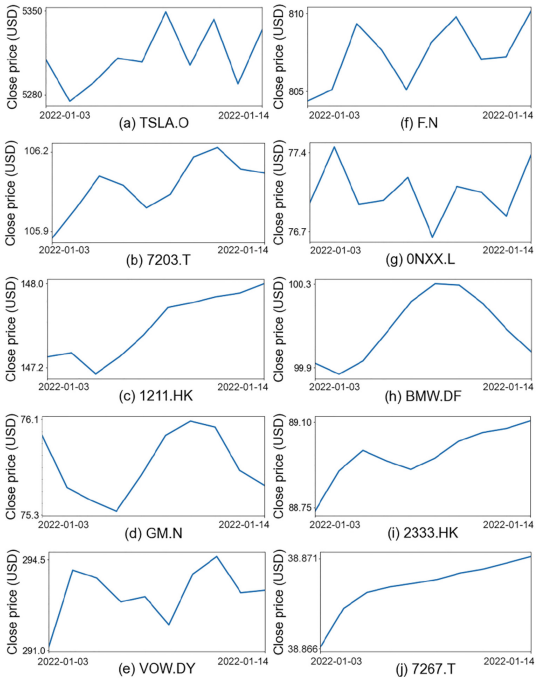

Fig. 6. Predictions of close price in the next 10 working days

3.2 Portfolio Construction

3.2.1 Mean-Variance Model

According to the compound daily returns of 10 stock sets calculated by the above ARIMA-LSTM model, a group of weights is randomly generated, and then the returns and standard deviation are expected to be derived by the mean-variance model. The above process is repeated for 20,000 times. Afterwards, the optimal balance point of risk and return is determined based on the maximum Sharpe ratio. Sharpe ratio is proposed by Sharpe on the basis of modern portfolio theory, which not only pays attention to the return, but also to the risk [14], and it is defied by the following equation, where R_p and $_p$ represent return and risk of portfolio p, R_f represents the risk-free interest rate, herein, R_f is set to 0.

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

Figure 7 shows the Sharpe ratios corresponding to different risks and returns, from which can be preliminarily judged that the point matching the maximum Sharpe ratio is located on the left side of the figure. Meanwhile, the corresponding coordinate (0.014, 0.053) is selected in the scatter plot (Fig. 8), where x = 0.014 represents risk and y = 0.053 represents return. The weights of 10 stocks corresponding with the maximum Sharpe ratio are shown in Table 6.

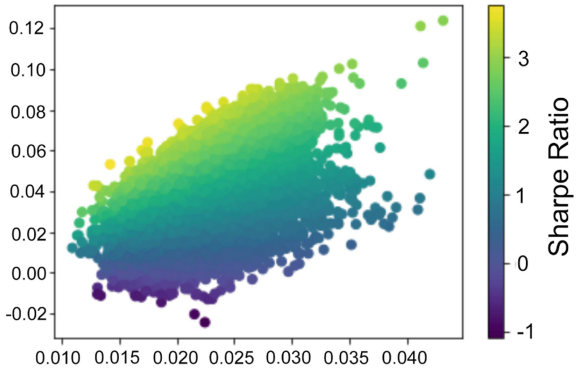

Fig. 7. Scatter plot based on Sharpe ratio

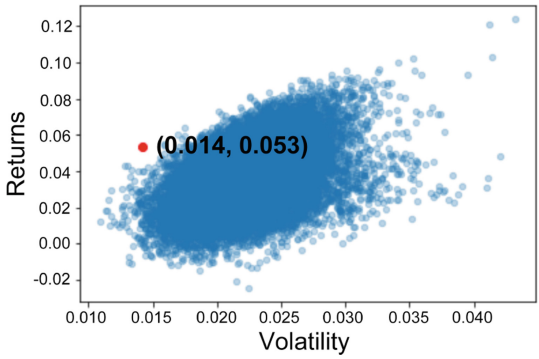

Fig. 8. Coordinates of the maximum Sharpe ratio

Table 6. The weight of optimal portfolio

Stocks	Weight	Stocks	Weight
TSLA.O	0.06	F.N	0.06
7203.T	0.33	0NXX.L	0.02
1211.HK	0.31	BMW.DF	0.06
GM.N	0.03	2333.HK	0.03
VOW.DY	0.06	7267.T	0.05

3.2.2 Performance

By comparing the returns under the same weight of 10 stock sets, it is found that the portfolio based on the maximum Sharpe ratio performs better from the three aspects of variance, Sharpe ratio and return (Table 7), indicating that the portfolio based on

Table 7. Annual return ratio of portfolio

Evaluation index	Annual equal-weight portfolio	Annual maximum Sharpe ratio portfolio
Variance	$3.98e^{-04}$	$2.02e^{-04}$
Sharpe ratio	1.82	3.64
Return	3.62%	5.18%

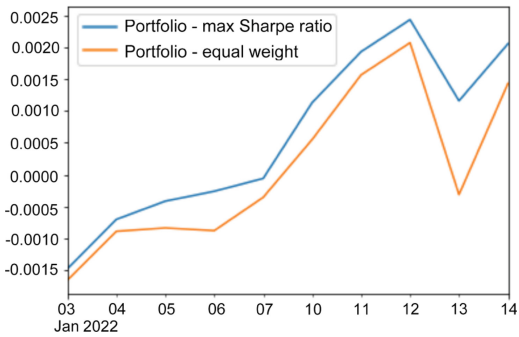

Fig. 9. The cumulative return of optimal portfolio

the maximum Sharpe ratio can better meet the expectations of investors, and the model established is superior.

Then the cumulative income of the equal weight portfolio and the maximum Sharp ratio portfolio is calculated respectively, and the line chart is represented in Fig. 9. It is apparent that the cumulative return of the portfolio under the maximum Sharp ratio is greater than that under the equal weight, which visually demonstrates that the maximum Sharpe ratio portfolio performs better.

4 Conclusion

This paper first presents a novel ARIMA-LSTM model and then designs mean variance model to build the next 10 days optimal portfolio based on 10 automobile stocks listed in the US, Japan, China, Germany and the UK from January 1, 2016 to December 31, 2021. One of the purposes is to illustrate that ARIMA-LSTM model is more superior and reliable than single ARIMA and LSTM in prediction accuracy based on RMSE and MSE, which is noteworthy that the hybrid forecasting model is definitely powerful and meaningful. The other purpose is to build an optimal cross-national auto stock portfolio based on the predicted data through the mean variance model, which plays a positive role in reducing risks and enhancing returns for global investors. Ultimately, 5.18% average annualized return is obtained, which means the performance of the portfolio based on

the maximum Sharpe ratio is indeed better than the market average level. To sum up, the model established in this paper is of great significance for investors who intend to invest in trans-national automotive stocks.

However, the model constructed in this paper indeed has some limitations. One the one hand, this study did not consider the seasonality of financial data, on the other hand, the influence of subjective factors such as the emotional tendency of real news on the stock price has not been taken into account in the construction of the model. Therefore, a possible research topic in the future is to eliminate the impact of seasonal factors on financial data. Another research direction is to consider some subjective factors, such as news and events, and then thinking about establishing a relevant news emotion dictionary to form a model integrating emotion analysis, which will have a better effect on improving the prediction accuracy of future stock prices.

References

1. Y. Jiang, J. Zhang, Research on the volatility spillover effect between China and major world stock markets-An empirical test based on samples from 2002 to 2017. China Economic Issues, 2019 (06): 28–43. https://doi.org/10.19365/j.issn1000-4181.2019.06.03.
2. A H. Bukhari, M. Raja, M. Sulaiman, et al., Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. IEEE Access, 2020, PP(99):1–1.
3. D O. Faruk, A hybrid neural network and ARIMA model for water quality time series prediction. Engineering Applications of Artificial Intelligence, 2010, 23(4):586–594.
4. A S. Temür, M. Akgün, G. Temür, Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models. Journal of Business Economics and Management, 2019, 20(5):920–938.
5. J. Wang, J. Wang, Z. Zhang, et al., Stock index forecasting based on a hybrid model. Omega-international Journal of Management Science, 2012, 40(6):758–766.
6. Yi, H. Wang, Nonlinear neural network forecasting model for stock index option price: hybrid GJR–GARCH approach. Expert Systems with Applications, 2009, 36(1):564–570.
7. Yulin, Du, Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network. 2018 Chinese Control and Decision Conference (CCDC). IEEE, 2018: 2854–2857.
8. H K. Choi, Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. Papers, 2018.
9. K. Sakshi, A. Vijayalakshmi, An ARIMA-LSTM hybrid model for stock market prediction using live data. Journal of Engineering Science and Technology Review, 2020, 13(4):117–123.
10. H. Lee, F. Cheng, S. Chong, Markowitz portfolio theory and capital asset pricing model for Kuala Lumpur stock exchange: A case revisited. International Journal of Economics and Financial Issues, 2016, 6(3): 59–65.
11. M. Širůček, L. Křen, Application of Markowitz portfolio theory by building optimal portfolio on the US stock market. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2015, 63.
12. D. Fan, H. Sun, J. Yao, et al., Well production forecasting based on ARIMA-LSTM model considering manual operations. Energy, 2021, 220: 119708.
13. G. Dutta, P. Jha, A K. Laha, et al., Artificial neural network models for forecasting stock price index in the Bombay Stock Exchange. Journal of Emerging Market Finance, 2006, 5.
14. W F. Sharpe, Mutual fund performance. Journal of Business, 1966, 39(1): 119–138.