

# Investment Portfolio Construction in the Beijing-Tianjin-Hebei Region of the CSI 300

Zhewei Liu\*

School of Economics, Minzu University of China, Beijing, 100081, China

\*Corresponding author: 19010583@muc.edu.cn

**Abstract.** How to construct an optimal investment portfolio has become the mainstream of research with the increasing demands of investors to preserve values and returns. This paper is based on the background of the construction of the millennium plan of Xiongan New Area and the gradual improvement of the integration of Beijing, Tianjin and Hebei. The author selects 5 stocks related to infrastructure construction in related area from CSI 300 to construct optimal investment stocks. After processing and analyzing the original data, the time series model is used to predict the future stock data, and the mean variance model is used to construct the investment portfolio to obtain the optimal Sharpe ratio. The final simulation result shows the superiority of portfolio construction with a high yield of 0.346 and the Sharpe ratio value is 2.473, which can provide a reasonable reference investment in enterprises in the Beijing-Tianjin-Hebei region. This paper has important reference value in data selection and optimization of time series model parameters.

**Keywords:** ARMA model; Integration of Beijing-Tianjin-Hebei regions; Investment portfolio; Mean-variance model.

## 1. Introduction

"Don't put all your eggs in one basket", building a portfolio in the financial sector can effectively avoid investment risks. There are several ways to construct an optimal portfolio, each with pros and cons.

Time series models have great advantages for short-term stock forecasting. It includes a variety of models, and the parameter settings required by the model need to be determined according to the actual situation. According to the one-year-opening-price of Gujing Gongjiu stock (000596), which estimated and tested the ARMA model based on its time series data. Combined with the results of the error ratio, it was determined that the prediction sequence obtained by using the ARMA model had a higher degree of fitting [1]. There was also a study once used the ARMA model to predict the closing price of Gree Electric's 240th stock. The p,q were determined based on the ACF and PACF. The AIC, SC and other data were used to help prove the accuracy of p, q [2]. A team once used the opening price as the data for time-series stock forecasting, but due to the large volatility and aggregation of the opening price, a more effective stationary sequence could not be obtained, and the operation was difficult [3]. If the data are non-stationary sequence by ADF test, use the first-order difference method to process the non-stationary sequence and test the stationarity again [4]. In addition, random forest, GBDT algorithm, neural network, SVM and other models can predict stock prices. Fits the stock price time series. By analyzing the closing price data of Shanghai A-share Kweichow Moutai (600519), the effect of the final model is obtained and displayed. The experimental results show that the prediction error of the fitting results using the GBDT algorithm is the smallest [5].

As the mean-variance model, the classic Markowitz mean-variance model is generally used to study the most efficient securities and to construct optimal asset allocation. MA has used the data of 10 stocks listed on the Shenzhen Stock Exchange from 2019 to 2020, and used the mean-variance model to implement quadratic programming in EXCEL, and obtained the corresponding weights of each security under the set expected return [6]. By analyzing the impact of historical returns using different time spans on the effective frontier curve of the mean-variance model, ZHANG found that appropriately narrowing the time span can help to obtain a better investment portfolio [7].

Xiongan New Area is an emerging national key construction project in China, and the integration of Beijing-Tianjin-Hebei is a great process that the country has been continuously advancing. The

resources and financial support of the Beijing-Tianjin-Hebei region can be well utilized in the early stage of its construction. A spatial econometric model is established to measure and evaluate the role of listed companies in the Beijing-Tianjin-Hebei region on regional economic growth and industrial structure upgrading [8]. The government needs to actively guide the financial market to ensure that the industrial projects in the Beijing-Tianjin-Hebei coordinated development plan receive financial support and strengthen the role of the capital market [9]. At the same time, the bond and stock market also needs to be paid attention to because it can quickly reflect market changes. The previous research proves that the price of building materials is highly correlated with the volatility of the relevant stock price, and the price change of the material leads the change of the stock price [10]. Due to the construction needs of the Xiongan New Area, the order demand for urban building materials, electricity and communications in the surrounding areas will increase significantly. The stock investment projects in the Beijing-Tianjin-Hebei region will certainly show a growth trend in the future.

The author selects 5 stocks related to infrastructure construction in related area from CSI 300 to construct optimal investment stocks, which involved in electric power, gas, electronic communications, new energy technology, etc. The paper will use ARMA model to predict the future stock data, and the mean variance model of the optimal Sharpe ratio is used to construct the investment portfolio. In this paper, various tests will be performed on the original data to ensure the rationality and predictability of the data, including white noise test, correlation test, ADF test, etc. Also the process ensure the accuracy of the prediction results by testing the fitting residuals and calculating the values of MAE and RMSE. The results of ADF test show that the mean and standard deviation of the stocks we select do not change significantly with time and basically tend to be stable. These enterprises can make a good contribution to the construction of Xiongan New Area, while strengthening the integrated construction of Beijing-Tianjin-Hebei. The method of constructing the ARMA model and related tests in this paper can also provide ideas for subsequent research.

This paper will be divided into the following three parts for elaboration: The first part is Data and methodology, including data selection, white noise test and correlation test. The second part will build ARMA model and mean-variance model to predict stocks and construct optimal investment portfolio. After the parameters were calculated by ACF and PACF values, AIC and BIC values were used to verify the goodness of fit of the model. And MAE/RMSE are calculated to judge the rationality of the model. Then calculate the optimal Sharpe ratio portfolio and compare it with equal-weight allocation portfolio. The third part explains the research conclusions and application value, and then analyzes the research problems and proposes solutions.

## 2. Data and methodology

### 2.1 Data selection

Relying on the millennium plan of Xiongan New Area, coupled with the development and rise of Beijing-Tianjin-Hebei integration. The development of Beijing, Tianjin and Hebei should not be underestimated. The urban construction of Xiongan New Area requires a complete industrial chain.

This article selects 5 stocks from the NetEase Finance website. Each stock has 486 closing trading prices from January 2020 to December 2021. Hebei is a strong heavy industry province in China, so we selected a stock that produces and processes basic industrial materials from listed companies in Hebei Province. Tianjin relies on the sea, and its power generation and transmission business is relatively developed, while Beijing is a relatively concentrated high-tech industry concentration area in China, and its technology research and development and the establishment of communication systems are relatively excellent. The selection of these five stocks has integrated various factors for the infrastructure construction such as region, technology, industry and so on. The business scope of these companies includes electricity, gas, communication, new energy technology research and development, etc. of Xiongan New Area. Here are the following five stocks and their lines of business:

300667 Beijing Beotech Inc.(Beijing)—R&D, production and sales of industrial monitoring systems, detection and analysis systems and MEMS sensor chips.

002108 Cangzhou MingZhu Plastic Co.,Ltd.(Hebei)—Production and sales of PE gas and water supply pipes and fittings, as well as production and sales of BOPA films.

000537 Tianjin Guangyu Development Co.,Ltd.(Tianjin)—Power generation, transmission business, supply business; new energy technology research and development.

603803 RAISECOM TECHNOLOGY CO.,Ltd.(Beijing)—Chip concept; Technology development and promotion; sales of communication equipment and industrial automatic control system devices; import and export of goods, agency and technology; computer system integration; manufacturing of communication equipment and industrial automatic control system devices.

000695 Tianjin Binhai Energy & Development Co.,Lt(Tianjin)—Production and sales of heat; electric power, power generation, gas, water system equipment and spare parts; engineering maintenance services and technical consulting services for the above systems.

## 2.2 Data processing and testing

### 2.2.1. White noise test

Calculate linear compound returns based on historical closing prices. Its calculation formula is as follows:

$$C_t = \ln\left(\frac{V_{t+1}}{V_t}\right) \tag{1}$$

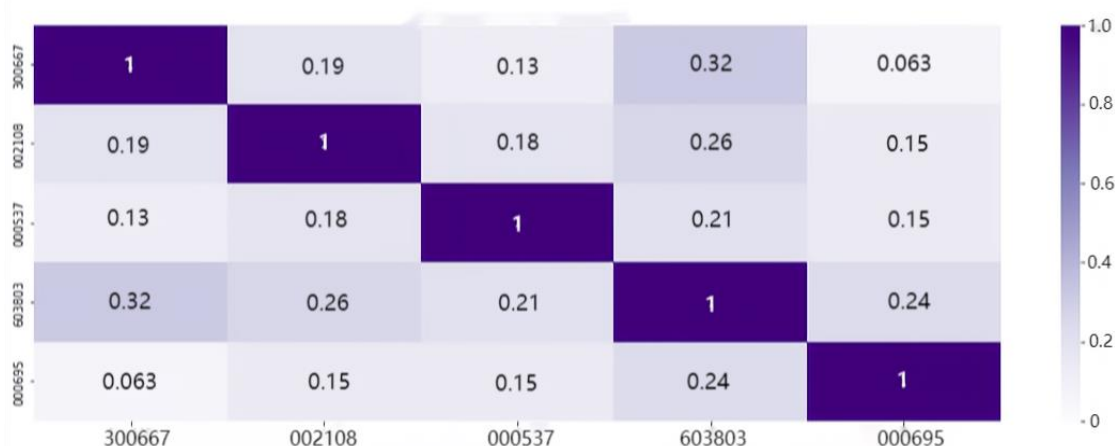
The white noise test is carried out on the compound return, obtained p-values of the five groups of data are all less than 0.05, indicating that the results are all non-white noise, and the data can be predicted. The results are shown in table 1.

**Table1.** White noise test results

	Sz300667	Sz002108	Sz000537	Sh603804	Sz000695
p	0.021	0.006	0.000	0.013	0.001

### 2.2.2. Correlation analysis

Correlation analysis on the returns of different stocks in figure 1 is conducted, which shows the low correlations between most different stocks in the heatmap. However, there is a certain correlation between stocks in the same region, such as 300667 and 603803 in Beijing.



**Fig. 1** Heatmap of returns

### 3. Results

#### 3.1 ARMA model prediction

##### 3.1.1. ADF tests

Before using ARMA model to predict the future data of the stocks, ADF tests on original data are necessary. The result shows that the mean lines are basically horizontal, and the standard deviations don't change much. All sequences is stable.

Taking 300667 as an example for analysis, the red line and black line in the figure 2 indicate that the mean and standard deviation do not change significantly with time and basically tend to be stable. The values have been tested and the results are all less than 0.05, which means that the data passed the stationarity test.

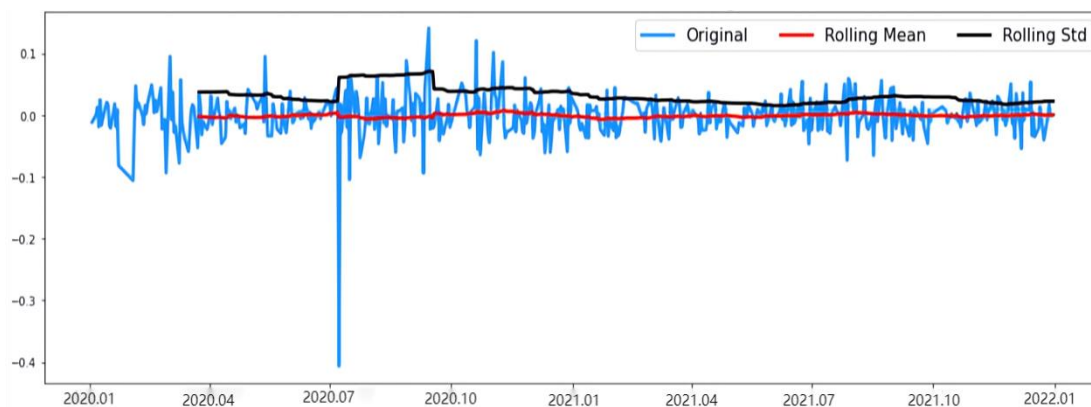


Fig. 2 Mean and standard deviation

The approximate results of the p-value obtained by the calculation are as follows in table 2:

**Table 2.** White noise test results

	Sz300667	Sz002108	Sz000537	Sh603804	Sz000695
<b>P-value</b>	0.000	0.000	0.006	0.000	0.000

##### 3.1.2. p, q selection

The trial-and-error of the model parameters is the highlight of the study. The ACF plot illustrates the degree of correlation between the current series value and the past of the current series. PACF describes the correlation between the residuals after removing the effects already explained by the lags and the next lag value.

From the ACF plot in figure 3, the result shows that the first-order lag is significantly higher than the standard level. Although the second-order lag has the characteristics of censoring, we cannot directly judge whether it is suitable. As for the PACF plot, it is obvious that the partial autocorrelation decreases rapidly after the first-order lag. So as the p and q selection, ARMA (1,1) is more reasonable.

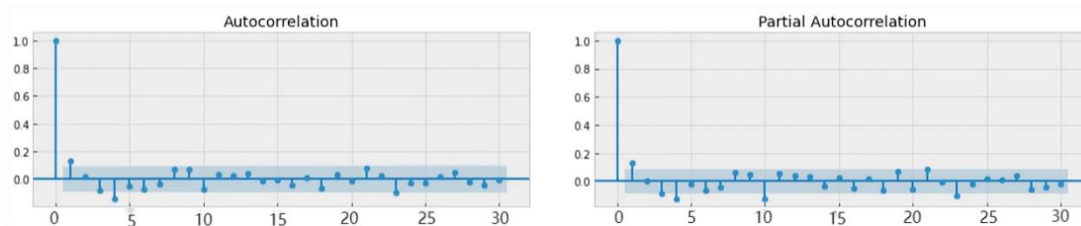


Fig. 3 ACF and PACF plot

AIC is a standard to measure the goodness of fit of statistical models. In general, the smaller the AIC, the better the model. Therefore the model with the smallest AIC is usually chosen. Detecting the BIC value can prevent excessive model complexity caused by excessive model accuracy when the number of samples is too large. In this study, we use AIC as the main criterion and BIC as the secondary criterion.

In the practical application of model selection, it is impossible to verify AIC, BIC and HQ one by one for all models. Therefore, a variety of models with p, q values close to the above models are established through enumeration method as shown in Table 3. ARMA (1,1) is more reasonable comparing these evaluation indexes with several different models.

**Table 3.** Model selection - based on AIC & BIC & HQIC criterion

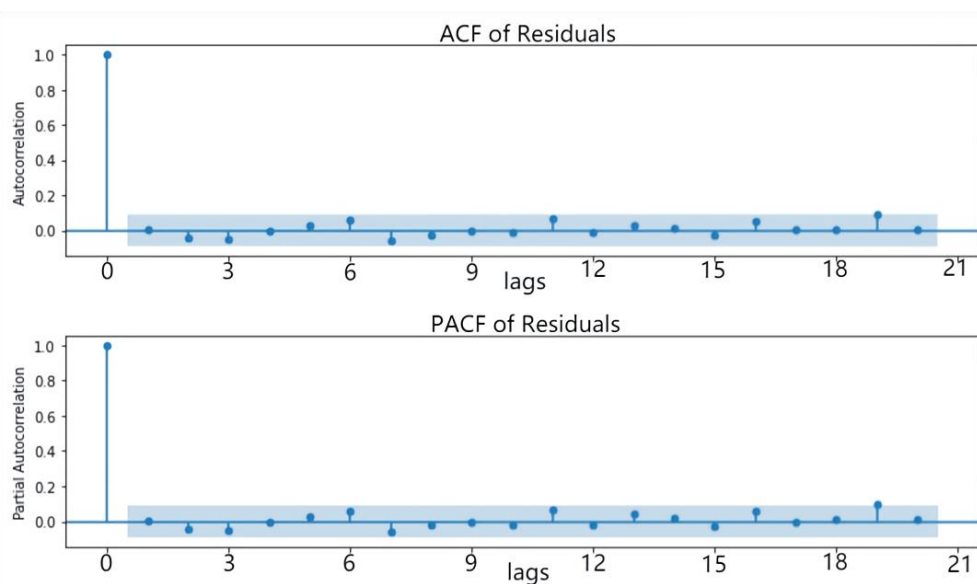
	ARMA(1,1)	ARMA(1,0)	ARMA(0,1)
AIC	-1883.582	-1879.863	-1880.459
BIC	-1866.845	-1867.311	-1867.906
HQIC	-1877.006	-1874.931	-1875.527

**3.1.3. Model fitting**

Take 300667 as an example, the corresponding ARMA (1,1) model based on the coefficients above is established as follows. Predict this stock price for the next 20 days with this formula and others:  
 $Return(t) = -0.0006 + 1,3976 * Return(t - 1) + \varepsilon(t) - 1,2082 * \varepsilon(t - 1)$  (2)

**3.1.4. Model testing — for residual**

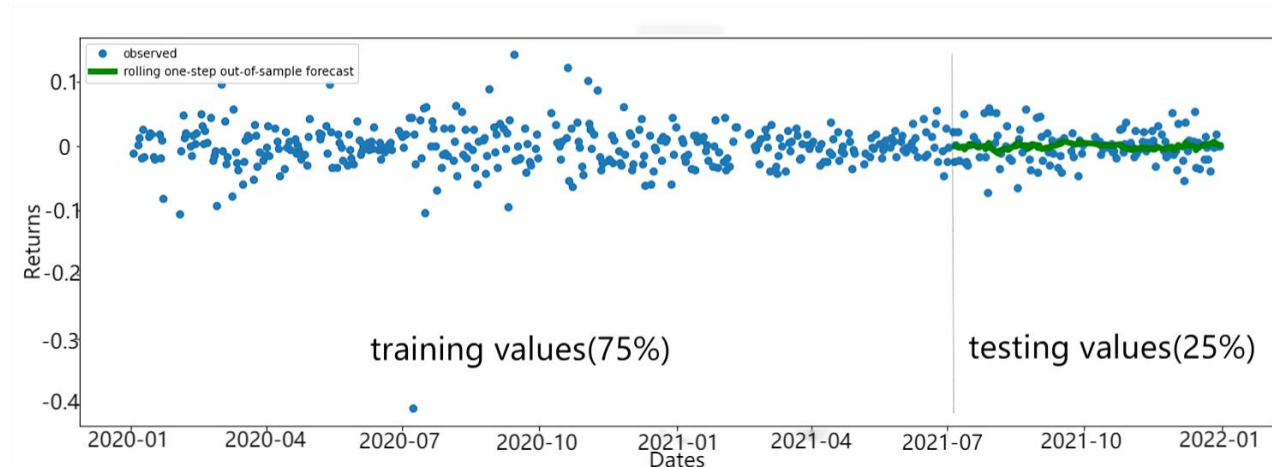
By observing the residual ACF and PACF in Figure 4, it is found that all lag points are in the blue range. The p-value is 0.6946 larger than 0.05, showing that the residuals are basically white noise, which means residual test is passed.



**Fig. 4** Residual ACF and PACF plot

**3.1.5. Model testing and prediction — for RMSE, MAE**

Taking the first 75% of the sequence(365) as training data and the last 25%(121) as testing data in figure 5.



**Fig. 5** Return spot and model test

MAE is the mean of absolute error. It can better reflect the actual situation of the predicted value error. RMSE is a measure of the deviation between the observed value and the true value. These two metrics are the most commonly used metrics to measure the precision of a variable.

Comparing test values and predictions, the evaluation indexes RMSE/MAE index values in table 4 are small, so the model is effective.

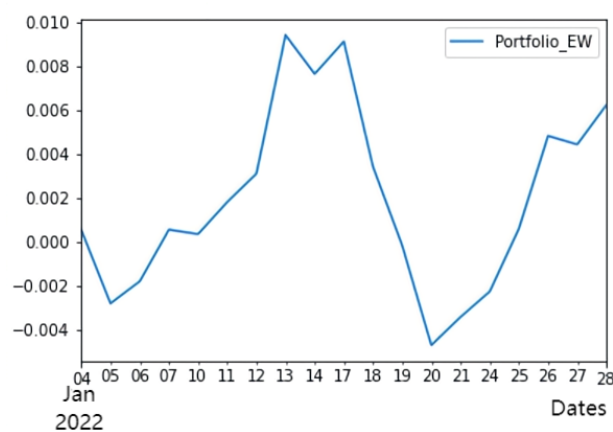
**Table 4.** Evaluation Indexes of 5 Stock sets

	<b>Sz300667</b>	<b>Sz002108</b>	<b>Sz000537</b>	<b>Sh603804</b>	<b>Sz000695</b>
<b>RMSE</b>	0.025088	0.046007	0.049874	0.028771	0.002187
<b>MAE</b>	0.018873	0.036200	0.036012	0.018741	0.036260

### 3.2 Mean-variance model

#### 3.2.1 Return of equal-weight allocation portfolio

Based on the 20-day return preparations by ARMA model, we first set the same weight for different stocks, and the return rate of a single stock is multiplied by the above weight we set to calculate the total return rate of the stock portfolio. And we use it as the benchmark for other portfolios. The figure 6 below shows the 20-day return of this stock portfolio, which is calculated to be 8.37%.



**Fig. 6** The cumulative return of equal weigh portfolio

### 3.2.2. The correlation analysis of the portfolio

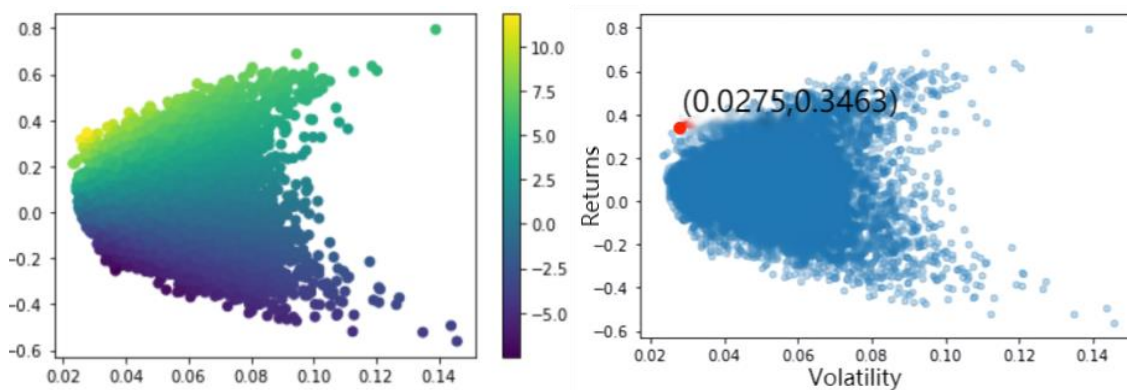
The annualized covariance matrix is in the table 5. It tells us the volatility of the stock and is used for the calculation of volatility.

**Table 5.** Evaluation indexes of 5 stock sets

	Sz300667	Sz002108	Sz000537	Sh603804	Sz000695
Sz300667	0.002165	0.000258	-0.001234	0.001172	-0.000414
Sz002108	0.000258	0.000833	0.000983	0.000912	-0.000210
Sz000537	-0.001234	0.000983	0.028743	-0.006465	0.000532
Sh603804	0.001172	-0.000912	-0.006465	0.034856	-0.000935
Sz000695	-0.000414	-0.000210	0.000532	-0.000935	0.004576

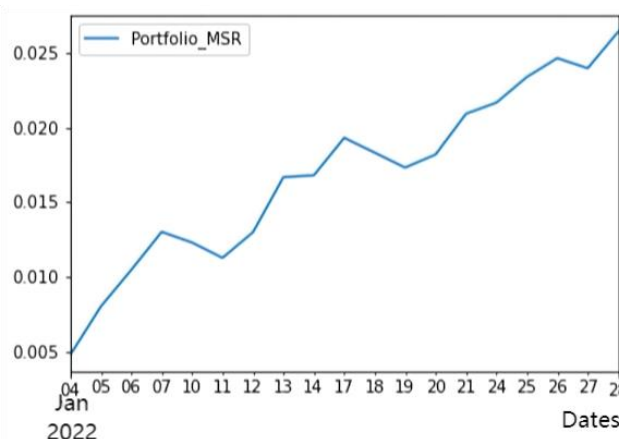
### 3.2.3. Sharp ratio portfolio

Randomly generate a group of weights and repeat the scatter plot for 10000 times to determine the effective boundary portfolio and select the optimal portfolio, which is the portfolio with the maximum Sharpe ratio. We set the risk-free rate of return as the one-year national debt interest rate on the first day of forecast. The red point is the portfolio we want. The figure 7 below shows the random scatter plot.



**Fig. 7** The random scatter plot

The optimal portfolio which is based on the weight composition of the maximum sharp ratio and the return line are shown as follows in figure 8. The 20-day return of the maximum Sharpe ratio portfolio is calculated to be about 34.63%:



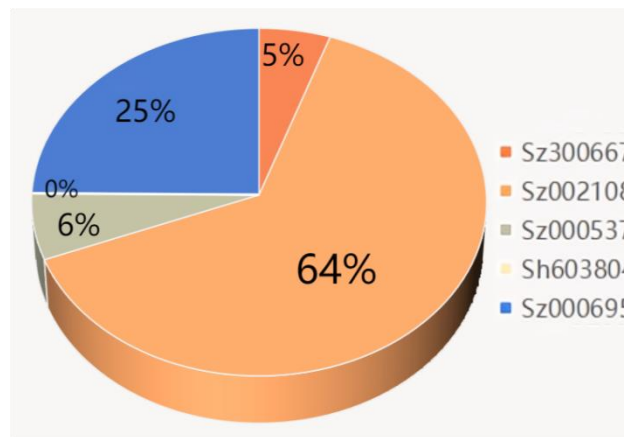
**Fig. 8** The cumulative return of optimal portfolio

The corresponding investment proportions of the five stocks are shown in the following table 6:

**Table 6.** The determined weighted portfolio

	Sz300667	Sz002108	Sz000537	Sh603804	Sz000695
<b>Weight</b>	5.269e-02	6.373e-01	6.143e-02	5.031e-04	2.480e-01

To make the results of the constructed portfolio clearer, a pie chart of the optimal Sharpe ratio portfolio is shown below:



**Fig. 9** Pie chart of the optimal portfolio

### 3.2.3. Comparison of two portfolios

Comparing equal-weight portfolio with sharp ratio portfolio from the perspective of the return rate of portfolio, it is true that equal weight portfolio is inferior to optimal portfolio. Here are the comparison of two portfolios in table 7:

**Table 7.** Comparison of two portfolios

	Equal-weight portfolio	Sharp Ratio Portfolio
Standard deviation	0.35	0.26
Sharpe ratio	-0.054	2.473
Return	0.084	0.346
Annual return	0.181	0.843

## 4. Conclusions

Xiongan New Area and the integration of Beijing-Tianjin-Hebei is a great process that the country has been continuously advancing. The study selects five stocks from industries such as infrastructure constructions and industrial automatic control systems in the Beijing-Tianjin-Hebei region. Use ARMA model to predict the future stock data, and the mean variance model of the optimal Sharpe ratio is used to construct the investment portfolio.

It is the first step to data preprocessing by doing various tests on the original data to ensure the rationality and predictability of the data, including white noise test and correlation test. During the prediction process of the ARMA model, the study first conducted an ADF test to ensure the mean and standard deviation do not change significantly with time and basically tend to be stable. Then use ACF and PACF plot in order to determine the value of ARMA(p,q). The AIC\BIC\HQIC also plays the role of auxiliary inspection. Also we ensure the accuracy of the prediction results by testing the fitting residuals and calculating the values of MAE and RMSE. Then the mean-variance model generate a group of weights and repeat the scatter plot for 10000 times to determine the effective boundary portfolio and select the optimal portfolio, which is the portfolio with the maximum Sharpe ratio. Comparing the Equal-weight portfolio with Sharp Ratio portfolio of the standard deviation,

Sharpe ratio and return. We can easily discover that the optimal investment portfolio have higher returns and lower risk. Generally speaking, the Sharpe ratio of stock funds and hybrid funds is greater than 1. The Sharpe ratio of 2.473 indicates that the calculated investment is really worth to take. It can assist investors to make reasonable investments in enterprises in the Beijing-Tianjin-Hebei region.

Combined with historical research and the research of this paper, the premise of successful data prediction is data processing, fitting test, reasonable setting of model parameters, etc. On the premise that these steps meet the established standards, and the research follows the prescribed procedures to ensure that the time series model fits the data well, so the time series model can make short-term more accurate predictions of stock prices. Nowadays, time series models are composed of many models. Models like ARIMA, GRACH and ARCH can be also used for financial forecasting. Take ARIMA as an example, it has a limit parameter more than the ARMA model that can make the model fit and prediction more accurate.

From the point of view of data processing, since the acquisition and recording of data has a certain time lag and some data are missing, there is a certain error between the experimental data and the real data, which will make the subsequent prediction more difficult to be accurate. To solve these problems above, the accuracy of setting the lag order  $p$  and  $q$  of the time series model is crucial. The accuracy of parameter setting can be improved by combining a variety of parameter determination methods, such as ACF and PACF plots, parametric trial and error, bserving the values of AIC, BIC and HQIC,etc. Data processing can compare and select better predictive models. Neural network is a prediction model that can well fit the series data of the fluctuation law of stock prices according to the characteristics of stock time. Among them, the LSTM model is mostly used for stock prediction. The traditional ARIMA model and the LSTM model have their own scope of application. Traditional metrology methods perform better when the data structure is relatively simple, but LSTM neural networks are more suitable for complex nonlinear and unstructured data.

## References

- [1] LI Yi. Short-term prediction of stock price based on ARMA model——Taking Gujing Gongjiu stock as an example. *Journal of Shanxi Finance and Taxation College*, 2022, 24(01): 25-30.
- [2] LIU Jie. Application of ARMA model in stock price forecasting——Taking Gree Electric as an example. *China Management Informatization*, 2021, 24(11): 153-155.
- [3] LIN Lanyu, CHEN Xiufang, ZHANG Defei. Application of ARMA Model in Stocks. *Economic Research Guide*, 2018, (26): 146-148.
- [4] MENG Kun, LI Li. Empirical analysis of stock price prediction based on ARMA model. *Journal of Hebei North University (Natural Science Edition)*, 2016, 32(05): 55-60.
- [5] LI Chao. Application and Comparison of Machine Learning Models in Stock Price Time Series Analysis. *Electronic World*, 2021(09): 66-70.
- [6] MA Tian. Mean-variance model theory and its application in my country's stock market. *Wealth Times*, 2022, (01): 148-150.
- [7] ZHANG Liying. Improvement and application of mean-variance model. Liaoning: Dalian University of Technology, 2021.
- [8] DU Jiahao. Spatial econometric analysis of the impact of Beijing-Tianjin-Hebei listed companies on the regional economy. Hebei: Yanshan University, 2018.
- [9] ZHANG Shuangying, LIU Sheng. Model selection of Beijing-Tianjin-Hebei financial support for industrial structure upgrading. *Finance Teaching and Research*, 2015(06): 10-13+66.
- [10] Xu Jiantao. Research on the price fluctuation of engineering materials based on relevant stock price fluctuations. *Science and Technology Wind*, 2019(17): 211.