Stock selection strategy based on double objective programming

Qi Zhang¹, FanYi Yang¹,a and YuBo Wang¹

¹Dujiangyan campus of Sichuan Agricultural University, 288 Jianshe Road, Guankou street, Dujiangyan City, Chengdu City, Sichuan Province, China.

2 First authors: Qi Zhang and Fanyi Yang

Abstract. This paper studies the securities market in which the stock market fluctuates. Because the market fluctuation plays an important role in the analysis of people's risk return, the maximization of shareholders' equity and the effective supervision of the regulators, how to study the law of the securities market fluctuation and what are the causes of the market fluctuation have become the necessary problems in the current study of the securities market. Lagrange interpolation is used to complete the data, and the time series prediction model is established to predict the closing of the next 100 days. The multi-objective programming is carried out by using the prediction data. Under the consideration of multiple constraints, lingo programming is used to solve the equations. The final portfolio is as follows: stock 2 accounts for 0.18, stock 3 accounts for 0.3, stock 5 accounts for 0.22, and stock 10 accounts for 0.29. The model was evaluated and five groups of data were selected. Using MAPE index to judge the error between the predicted result and the real value, the accuracy of the predicted result can be judged if the calculated result is about 0.1. By analyzing the new portfolio and processing the attachments, the investment benefits of the newly selected portfolio are all positive, and the expected income is also positive, which verifies the effectiveness of the model.

1. Introduction
Since the establishment of Shanghai Stock Exchange on December 19, 1990, after 30 years of rapid development, China's stock market has a considerable scale and has made remarkable achievements in many aspects, which plays an increasingly important role in the allocation of resources in the national economy. By the end of 2019, there are about 4000 stocks traded on the Shanghai and Shenzhen stock exchanges. At present, the market trading system, information disclosure system and securities laws and regulations and other supporting systems have been established. Investors are becoming more rational and mature. Institutional investors have developed rapidly and have taken shape. The government's supervision of securities market transactions and the main behavior of listed companies has achieved results.

With the development of China's capital market and the continuous expansion of the scale of securities trading in recent years, more and more funds are invested in the securities market. At the same time, the fluctuation of the market price is also very fierce. As the most essential attribute and feature of the securities market, the fluctuation of the market has a great impact on the analysis of people's risk and return. The maximization of shareholders' equity and rights and interests and the effective supervision of regulators play an important role. Therefore, it is an important content of
theoretical research and empirical analysis of the securities market to study the regularity of securities market fluctuations and analyze the causes of market fluctuations\cite{1}, which can also provide evidence for investors, regulators and listed companies.

2. Problem description
In this paper, according to the highest and lowest, opening and closing, and trading volume of some stocks from January 2019 to March 2020, except weekends and holidays, there are still some missing data. Python is used to obtain the date of the stock market trading day, and the missing date number is determined. Lagrange interpolation is used to complete the data. This paper forecasts the closing of the next ten stocks in 100 days through time series\cite{2}, considers multiple constraints such as maximum return and minimum value at risk\cite{3}\cite{4}, carries out multi-objective programming\cite{5}, and selects the optimal portfolio strategy.

Therefore, author need to evaluate the model in question one, select the evaluation index\cite{6}, consider the psychological needs of consumers, mainly from the aspect of income, and then from the aspect of risk. First of all, the five groups of known data in the selected annex are predicted, and then the average absolute percentage error MAPE between the predicted value and the known real value is calculated to know the accuracy of the prediction. Finally, the combined risk value, expected return, real return and the rationality of the model are analyzed.

Through the comprehensive stock price index based on the proportion of trading volume as the weight of distribution, author can be more accurate to get the market index\cite{7}. Then draw the corresponding graph, and make investment suggestions and strategies according to the fluctuation of the index\cite{8}.

3. Symbol definition and explanation

| Symbol | Definition and explanation |
|--------|---------------------------|
| \(\eta\) | Initial wealth, i.e. total assets |
| \(\alpha_i\) | The difference between the predicted future closing price minus the current known closing price divided by the known closing price of the stock, that is, the stock yield |
| \(x_i\) | Proportion of each stock purchased |
| \(Z_0\) | Predicting the maximum expected value of investment utility under optimal portfolio strategy |
| \(\text{VaR}_p\) | Value at risk of portfolio |
| \(\mu_p\) | Expected return of portfolio |
| \(\sigma_p\) | Standard deviation of portfolio |
| \(Z\) | Sampling quantile of standard normal distribution |
| \(s^2\) | Variance |
| \(b_0\) | Average trading volume of each stock |
| \(b_i\) | Average trading volume of each stock |
| MAPE | Mean absolute percentage error |

To simplify the model, authors define the above parameters.
4. Establishment and solution of model

4.1. Selection of optimal combination strategy

4.1.1 Data analysis and processing
In addition to weekends and holidays, there is also data loss. The data is large, so we can find the clear trading day of the stock market through Python Programming docking with sina finance, find the missing date of each stock by comparison, and complete the data through Lagrange interpolation, making the following K-line chart:

![Figure 1. K-line chart of the first stock](image1)

![Figure 2. local K-line graph of the first stock](image2)

4.1.2 Establish a multi-objective linear programming model
In order to select the optimal stock portfolio, this paper takes the maximization of target benefit and the minimization of risk as the portfolio criteria. In this paper, the multi-objective linear programming is adopted, the constraint conditions are determined, the investment ratio is taken as the solution value of the portfolio strategy, and the relationship between the predicted transaction quantity and the existing transaction quantity is comprehensively considered to establish the multi-objective programming function.

To maximize the target benefit, when the total assets is 1, divide the difference between the predicted future closing price minus the known closing price of the stock by the known closing price...
of the stock, that is, the stock return rate is taken as the coefficient, and the proportion of purchasing
the stock is taken as the unknown quantity. The function is established as follows:

$$\max Z_0 = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_i x_i, i = 1, 2, \ldots, 10$$

(1)

The following constraints are determined to minimize the investment risk.

Avoiding a single stock has a great impact on the economic change of the whole portfolio, this
dpaper restricts the purchase market value of each stock not more than 30% of the initial wealth, and
each stock not less than 15% of the initial wealth.

$$15\% \eta \leq x_i \leq 30\% \eta; i = 1, 2, \ldots, 10$$

(2)

Reducing the similarity of the purchase amount of each stock and ensuring that the purchase
proportion of each stock depends on the corresponding coefficient weight, it is agreed that the variance
of the purchase proportion of all stocks should be less than 0.016.

$$s^2 = \frac{1}{n} [(x_1 - x)^2 + (x_2 - x)^2 + \cdots + (x_n - x)^2] \leq 0.016$$

(3)

Considering that the public stock trading will have a certain impact on the purchasing trend
of the whole stock market, the sum of the proportion of each stock purchased * the average trading
volume of the stock is greater than the average trading volume of 10 stocks $b_0$. The constraint
conditions are as follows:

$$b_i \sum_{i=1}^{10} x_i \geq b_0$$

(4)

For avoiding the instability caused by the economic fluctuation caused by the purchase of only one
or two stocks, authors should restrict the type of stocks purchased to be between 4 and 7, so as to
reduce the investment risk of stocks

$$x_i \neq 0; \quad (4 \leq \text{the number of } i \leq 7)$$

(5)

Considering the wealth value of the fund company, the total amount of the purchase of all the
stocks should be less than or equal to the initial wealth, and the total assets is 1.

$$\sum_{i=1}^{10} x_i \leq 1$$

(6)

All the above constraints are as follows:

$$\begin{align*}
\max Z_0 &= \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_i x_i; i = 1, 2, \ldots, 10 \\
\frac{1}{n} [(x_1 - x)^2 + (x_2 - x)^2 + \cdots + (x_n - x)^2] &\leq 0.016 \\
15\% \eta &\leq x_i \leq 30\% \eta; i = 1, 2, \ldots, 10 \\
b_i \sum_{i=1}^{10} x_i &\geq b_0 \\
x_i \neq 0; \quad (4 \leq \text{the number of } i \leq 7) \\
\sum_{i=1}^{10} x_i &\leq 1
\end{align*}$$

(7)
Value at risk can also be calculated by Var, so in this case, the effect of value at risk and investment utility on independent variables is different. Authors decided to use the linear weighted sum method to transform the investment value at risk into a scalar problem of weighted sum of two objectives. Then, when the final expression of the opposite number of investment utility and investment value at risk is the maximum weighted sum, the author attached the weight of investment utility to 0.6 and the weight of investment value at risk to 0.4

$$\max f = \omega_1 (\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_i x_i) + \omega_2 (VaR_{p_i} = w_0 \cdot (Z \cdot \sigma_{p_i} + \mu_{p_i}))$$  \hspace{1cm} (8)$$

All the above steps are programmed by lingo, and the market value of each stock is obtained as follows: in the case of the highest return and low risk, the optimal purchase strategy is stock 2, stock 3, stock 5 and stock 10, with the proportion of 0.18, 0.3, 0.22 and 0.29.

4.2. Model establishment and solution of problem two

4.2.1 Determine the evaluation index

If investors want to get the highest return when they buy stocks, the author should determine whether the model is reasonable. This paper holds that a stock portfolio should consider the expected income, and the value at risk of the portfolio should be low; the remaining unknown data can be predicted through some known data, and the MAPE value can be calculated to judge the accuracy of the prediction. In fact, it is particularly important how much money the selected stock can actually earn. The rationality of constraint conditions is also the judgment condition of whether the model is reasonable or not. Based on the above considerations, the evaluation indexes in this paper are: the accuracy of prediction, the value at risk and the expected income, the actual income and the expected income, and the rationality of constraints.

4.2.2 Evaluation model based on constraint conditions

(1) The rationality of the model is analyzed by the accuracy of prediction

Select n groups of data in the attachment, the data all contain 49 days, and the date is close to the first half of 2019. The selected five groups of data are brought into the time series model for prediction, and the data values in the next 39 days are obtained. The accuracy of prediction can be known by calculating the average absolute percentage error (MAPE) between the predicted value and the known real value.

The selected data are brought into the time series, and the prediction results are shown in the supporting materials. The MAPE formula is as follows:

$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$  \hspace{1cm} (9)$$

$$APE = \frac{|y_i - \hat{y}_i|}{y_i}$$

The MAPE values of five groups of data were calculated, and the results are as follows:

| Number of groups | 1   | 2   | 3   | 4   | 5   |
|------------------|-----|-----|-----|-----|-----|
| MAPE value       | 0.0869 | 0.0673 | 0.1031 | 0.1002 | 0.16 |

The results show that the MAPE values of the five groups of data are small, indicating that the error between the predicted data and the original data is small and has strong persuasiveness.

(2) The rationality of constraint conditions for multiobjective linear programming is analyzed
All the constraints play an important role in the limitation of this paper, which requires maximizing the utility of investment and minimizing the risk of investment. The maximization of investment utility, that is, the maximization of return, is based on five lower risk conditions, which are coordinated and interrelated. At the same time, considering that each condition is related to reality, it can show that each constraint condition is reasonable.

5. Conclusion
In this paper, the data is supplemented by Lagrange interpolation to make up for the lack of data. The closing price of ten stocks can be predicted through time series. The long-term and short-term memory model (LSTM) with good prediction effect is used to predict the stocks. The prediction results have high reliability and authenticity, which paves the way for the future multi-objective programming. But the time series reflect the average value of the future value, the real value will fluctuate up and down in the forecast value, so it can reflect the sudden situation of the stock in the market. Using multi-objective programming, using LINGO to solve, author can easily and quickly calculate the required answer, make each objective value as close as possible to the optimal solution, and solve the contradiction between the two objectives.

And according to the actual situation, the paper selects n groups of evaluation indexes, selects the data given in the appendix, judges the accuracy of the forecast data by calculating MAPE in the first 49 days and the next 30 days. Author obtains several new groups of investment portfolios by using the multi-objective programming, their income from the appendix data and then calculates the corresponding value at risk and expected income. This paper evaluates the model from multiple perspectives, which makes the logic of this paper close and the evaluation convincing. However, there are many steps in this evaluation method, which can be further simplified or better model can be used for evaluation.

References
[1] Wang Xin, Cheng Jiehai, Si Longwei, Xu Ruirui. Spatial analysis of new housing price based on spatial interpolation: a case study of Jiaozuo City [J]. Surveying and spatial geographic information, 2020,43 (05): 51-55
[2] [J / OL]. Computer engineering and application: 1-9 [2020-06-01]http://kns.cnki.net/kcms/detail/11.2127.TP.20200528.1414.008.html.
[3] Fangfang Yang, Shaohui Zhang, Weihua Li, Qiang Miao. State-of-charge estimation of lithium-ion batteries using LSTM and UKF[J]. Energy,2020, 201.
[4] Vinay Kumar Reddy Chimmula, Lei Zhang. Time series forecasting of COVID-19 transmission in Canada using LSTM networks[J]. Chaos, Solitons and Fractals: the interdisciplinary journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena, 2020, 135.
[5] Yulong Lei, Liguo Hou, Yao Fu et al.. Research on vibration and noise reduction of electric bus gearbox based on multi-objective optimization[J]. Applied Acoustics, 2020, 158.
[6] Fan Chaodong, Ding Changkun, Zheng Jinhua et al.. Empirical Mode Decomposition based Multi-objective Deep Belief Network for short-term power load forecasting[J]. Neurocomputing, 2020 (prepublish).
[7] Zahra Pourreza Movahed, Mohammad Kabiri, Saeed Ranjbar et al.. Multi-objective optimization of life cycle assessment of integrated waste management based on genetic algorithms: A case study of Tehran[J]. Journal of Cleaner Production, 2020, 247.
[8] Li Xianglong, Ma Longfei, Zhao Xiangyang, Zhu Jie, Xu Zhenhua. Time multi scale electric heating load forecasting based on LSTM network [J / OL]. Journal of power system and automation: 1-5 [2020-06-01]https://doi.org/10.19635/j.cnki.csu-epsa.000463.