

Analysis of Stock Price Forecasts—Taking Midea Group as an Example

Jiashu Zhang

Henan University, Kaifeng, China

Email: Zhang jiashu@qq.com

How to cite this paper: Zhang, J. S. (2022). Analysis of Stock Price Forecasts—Taking Midea Group as an Example. *Open Journal of Social Sciences*, 10, 302-314. <https://doi.org/10.4236/jss.2022.109019>

Received: June 7, 2022

Accepted: August 21, 2022

Published: August 24, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The price of a stock is affected by many factors. On the one hand, it is influenced by various external factors, such as macro factors, microeconomics factors, and on the other hand, it is also affected by internal factors, such as the actual operation of listed companies, at the same time, the impact of domestic and international news will also have an impact on the stock price and stock market, so the prediction of the stock price is very important. Based on the analysis of the technical aspects that affect the stock price, this paper chooses the white horse stock of Midea Group as the research object, from January 1, 2022 to May 27, 2022 on the stock's opening price, the highest price, the lowest price, quantity, turnover as independent variables; The price of the stock at the close of the next day is a dependent variable. The multivariate linear regression model of dependent variable about independent variable is established, and the model is improved by factor analysis, and the final stock price prediction model is obtained.

Keywords

Factor Analysis, Regression Analysis, SAS, Stock Price Forecast

1. Introduction

Since the 1990s, China's stock market has experienced from initial scale to expansion and from immaturity to maturity and perfection through the development of more than 20 years. Due to the large number of investors in the stock market, the stock price affects the quality of people's life, the mood of investors, the good operation of enterprises, the stability of the entire stock market, and to a certain extent, the economic development of the society. At the same time, stock price prediction is a necessary skill for stock investors, not blindly following the majority of investors in the market operation, according to the relevant

factors affecting stock prices, predicting the future trend of stock prices should be the basic skill investors should master; Therefore, it is necessary to study the influencing factors of stock price so as to predict stock price.

This study combining theory with practice, to midea group this white horse, for example, by choosing it in a particular period of time series data, found the stock price itself and stronger correlation between various indicators, through the preliminary analysis of the data, it hopes to build on the dependent variable the day after the stock closing price indicators related to the multiple linear regression model. However, considering that there are serious multicollinearity problems among the original six independent variables, factor analysis is used in this study to solve this problem. Finally, the two common factors extracted through factor analysis are taken as independent variables, and the closing price of the day is the dependent variable to carry out multiple linear regression to obtain the final stock price prediction model. To sum up, the overall research objective of this paper is to effectively improve the accuracy of the stock price prediction model through the analysis and application of a series of empirical research methods, so as to achieve quantitative investment.

At the same time, the model still needs to be improved. This paper selects information related to the stock price of Midea Group. The stock information examined is in the period of relatively stagnant economic operation, so it may also have a certain influence on the stock price change. However, due to the different stock market quotation in different periods, special treatment should be given to the more special stock quotation. In addition, the stock prediction model in this paper is only suitable for the short-term prediction of stock prices, but for the long-term stock price changes it needs to consider the influence of many aspects. In addition, this paper analyzes stock prices from a technical point of view, but under actual circumstances, stock prices are affected by the overall economic environment, macro and micro economic factors, the company's internal operation, investors' irrational deviation, market sentiment. Therefore, the factors taken into account in this model are not particularly sufficient, and it is hoped that various factors can be taken into more in-depth consideration in the future investigation.

2. Research Status

At present, domestic scholars have many analysis methods for stock price prediction, among which many scholars choose to use BP improved algorithm, establish ARMA model, and research stock price prediction based on LSTM and other methods:

Feng Xingjie and Meng Xin (2011) used the BP network optimized by genetic algorithm to predict the maximum stock price by using two kinds of data sets containing the maximum stock price and the average value respectively. According to the characteristics of the data set, the model was established and the two groups of prediction results were fused to improve the prediction accuracy.

Experimental results show that this method can improve the prediction accuracy significantly.

Based on the basic theory of ARMA model, Meng Kun and Li Li (2016) established a reasonable time series model and took Shanghai Composite Index as an example to conduct fitting prediction and empirical analysis on stock prices. Eviews software is used to test the unit root of the original data sequence to judge whether the original data sequence is stationary. If it is not stationary, it is necessary to conduct first-order or even second-order difference processing on the original sequence to test the stationarity of the sequence after difference again. If it is stationary, no difference is needed; secondly, parameters of ARIMA (2, 1, 2) models are estimated by using the model form of autocorrelation graph and partial autocorrelation graph recognition sequence. Finally, the model is used to predict the daily closing price of Shanghai Composite Index.

Li Guicheng, Xu Li and Zhang Li (2022) studied the prediction results of stock price by referring to various factors that affect the actual stock price. This study proposes a new model M, which combines time series prediction and text sentiment analysis, and adopts the extended MODEL of LSTM, namely DA-RNN and BilSTM-Attention. On the one hand, the effectiveness of DA-RNN in time series prediction experiments is verified by analyzing and comparing the experimental results of XGBoost, LSTM and DA-RNN. On the other hand, bilstm-attention model is adopted as the main method to adjust the stock forecast value, which makes the forecast result more interpretable. In this study, an improved stock price prediction model M is used to predict stock prices.

Therefore, many scholars choose to use BP improved algorithm, establish ARMA model, based on LSTM and other methods for stock price prediction research. In this paper, a new method of predicting stock price model is proposed to improve the model by means of factor analysis. At the same time, considering the influence of factors on stock price, we can also get the short-term prediction model of stock price. That is, on the one hand, it not only considers whether different independent variables have a significant impact on the stock price of the dependent variable, but also considers the prediction model of the future stock price.

In this paper, six independent variables, namely the opening price, closing price, highest price, lowest price, quantity and turnover of Midea Group stock from January 1, 2022 to May 27, 2022, are selected to influence the next day's closing price of the dependent variable. Considering the strong correlation between variables and the possible multicollinearity brought by these six independent variables, factor analysis is undoubtedly a choice for variables with strong correlation. Factor analysis can not only reduce the dimension of the model, but also eliminate the effect of multicollinearity, and simplify the model. Finally, the short-term prediction model of Midea Group stock price is obtained, and the prediction value is compared with the real value. It is found that the establishment of the improved multiple linear regression model based on factor analysis is more accurate, so as to provide feasible suggestions and schemes for short-term inves-

tors.

3. Empirical Analysis

3.1. Data Collation and Variable Selection

In this paper, the daily opening price, highest price, lowest price, closing price, quantity and turnover information of American group stocks in trading days from January 1, 2022 to May 27, 2022 are selected. The closing price of the next day is taken as the dependent variable. The following independent variables are defined: Daily opening price is defined as open, daily highest price is defined as highest, daily lowest price is defined as lowest, daily closing price is defined as pclose, daily quantity id defined as quantity, daily turnover is defined as turnover. All data are from the Wind Database.

3.2. Research Methods and Related Theories

The basic principle of regression analysis: take the linear regression process of one variable as an example, first construct the model: $Y_t = \alpha + \beta X_t + \varepsilon_t$, The least square method is used to estimate the parameters and minimize the “distance” between all observation points and the straight line to determine the values of α and β ; Assuming a regression line: $Y = \alpha^* + \beta^* X$, find the values of α^* and β^* that minimize the function $V = \sum_{i=1}^N (y_i - \alpha^* - \beta^* X_i)^2$, of which $\alpha^* = \bar{y} - \beta^* \bar{x}$; $\beta^* = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}$; The purpose of the regression equation

is to get the statistical rule of the change of Y with X . If $\beta = 0$, So no matter how X changes, $E(Y)$ stays the same, and the regression equation is meaningless; Therefore, the test of regression equation is to test whether β is 0, usually we use F test and T test .At the same time, the regression results should also be diagnosed to observe whether the residual is randomly distributed, whether it is normal, whether there is heteroscedasticity; Whether the highly correlated independent variables cause collinearity, etc.

The basic principle of factor analysis: Factor analysis is a multivariate statistical analysis method which reduces the variables with complex relationship to a few comprehensive factors. We call potential variables that cannot be observed common factors. It is assumed that P indicators are tested for the study sample, and P indicators may be affected by m common factors, plus other factors that have an impact on these indicators, which can be expressed by a linear equation:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + a_{13}F_3 + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + a_{23}F_3 + \cdots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + a_{p3}F_3 + \cdots + a_{pm}F_m + \varepsilon_p \end{cases}$$

F_i is called the common factor, A is called the factor load matrix, ε_i is the special factor of variable X_i , a_{ij} represents the load of the i th variable on the j th common factor. For simplicity, the common factors are assumed to be unrelated

to each other and to have unit variance when the common factors are first extracted. In this case, the covariance matrix of vector X can be expressed as:

$$D(X) = D(AF + \varepsilon) = AA' + D \quad (D = \text{diag}\{d_1^2, d_2^2, \dots, d_p^2\})$$

Generally speaking, the first M common factors extracted make the cumulative contribution rate reach more than 70%; Factor load matrix A needs to be solved. In this paper, we choose SMC method to solve A . In order to make the factors more easily explained, new matrix A^* can be obtained by right-multiplying factor loading matrix A by an orthogonal matrix T through factor rotation, which does not affect the common degree h_i^2 of variable X_i , but will change the variance contribution of the factors. Finally, factor scores need to be calculated. The way to calculate factor scores is to describe factors with original variables, The value of the j th factor on the i th sample can be expressed as:

$$F_{ij} = \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \beta_{jp}x_{ip}, \quad j = 1, 2, \dots, m$$

This formula is also called factor score function, and the factor value coefficient is usually estimated by regression method in the sense of least square.

3.3. Descriptive Statistical Analysis of Data

```
proc means data = meidi mean std uss range;
var tclose open highest lowest pclose quantity turnover;
run;# Call the means process to output the mean, standard deviation, sum of
squares and range of each variable#(All tables and figures below are from the
SAS program output)
```

As can be seen from **Table 1**, the mean values of daily opening price, daily high price, daily low price, daily closing price, daily quantity and daily turnover are respectively about 63.7, 63.8, 64.8, 62.7, 63.7, 31248375.3 and 1979971658. It can be seen that the next day's closing price of the dependent variable is affected by the six factors, among which the first four factors (daily opening price, daily high price, daily low price and daily closing price) have a direct impact on it, while the last two factors (daily quantity and daily turnover) also have an indirect impact on the dependent variable. In addition, as can be seen from the data in the above table, stock prices generally fluctuate constantly within a day, but the fluctuation range is not particularly huge for the stock market with normal

Table 1. Descriptive statistics between variables.

variables	mean	Standard deviation	Sum of squares	range	maximum	minimum	n
tclose	63.45	8.47	381062.04	26.83	78.62	51.79	93
open	63.80	8.47	385161.25	27.47	78.20	50.73	93
highest	64.81	8.72	397699.39	27.53	80.88	53.35	93
lowest	62.69	8.42	372110.43	25.50	77.00	50.50	93
pclose	63.69	8.49	383919.44	26.83	78.62	51.79	93
quantity	31248375.26	11908505.58	1.0385762E17	52495531.00	63204589	10709058	93
turnover	1979971658	7753562455	4.1989596E20	3967600019	4743653586	776053567	93

market conditions. As can be seen from the range of each variable during this period (From January 1 to May 27, 2022), the stock fluctuation range will be relatively large in a certain period of time. Therefore, it is very necessary and meaningful to forecast stock price.

3.4. Multivariate Linear Regression Analysis

The correlation between *tclose* and *open*, *highest*, *lowest*, *pclose*, *quantity* and *turnover* was analyzed. Since these variables were all numerical variables, the paper adopted the method of multiple linear regression analysis to investigate the relationship between them, so as to obtain the regression model. First build the model:

$$tclose_i = \beta_0 + \beta_1 open + \beta_2 highest + \beta_3 lowest + \beta_4 pclose + \beta_5 quantity + \beta_6 turnover + \varepsilon_i$$

```
proc reg data = meidi(drop = date);
  model tclose = open highest lowest pclose quantity turnover;
run;
```

According to the output results in **Table 2** and **Table 3**, the multiple linear regression model of dependent variable and independent variable is:

$$tclose_i = -0.88367 + 0.12898open + 0.07718highest - 0.06118lowest + 0.85836pclose + 4.906933E-8quantity - 6.3784E-10turnover$$

First, look at the economic significance of this model. The stock price closing price of the next day is directly proportional to the opening price of the day, the highest price of the day and the quantity of the day, and inversely proportional to the lowest price of the day. In line with realistic significance; But the next day stock closing price and that day stock turnover is inversely proportional, does not conform to common sense. Secondly, regarding the significance test of the model, the P value of the model as a whole is <0.0001 according to the results of

Table 2. Parameter estimation of model 1.

variables	Parameter estimation	Standard error	Pr > t	vif
intercept	-0.88367	3.44066	0.7979	0
open	0.12898	0.27978	0.6460	225.37774
highest	0.07718	0.31219	0.8053	297.43315
lowest	-0.06118	0.43642	0.8888	542.69790
pclose	0.85836*	0.23497	0.0004	160.00579
quantity	4.906933E-8	1.041713E-7	0.6388	61.75043
turnover	-6.3784E-10	1.6078E-9	0.6926	62.35956

Table 3. Summary of model 1 information.

Model 1	R ²	Adjusted-R ²	value of F	value of P	value of DW
	0.9702	0.9681	466.53	<0.0001	1.860

variance analysis, indicating that the model as a whole is significant. However, for each variable, only the p-value corresponding to the closing price of the variable $pclose$ on the day was less than the significance level, while the p-value corresponding to the other variables were all greater than the significance level, that is, all failed the significance test of the variables. The decision coefficient $R^2 = 0.9702$, indicating that 97% of the change in the next day's closing price of the dependent variable can be caused by the fluctuation of the independent variable. $DW = 1.860$, located near 2, which satisfies the assumption that the residuals are not correlated. Again on the model test in the multicollinearity, each variable variance inflation factor is far greater than 10, often think, variance inflation factor is more than 10, we think that the emergence of multicollinearity, therefore the model 1 very serious collinearity between variables, and many variables are not significant, therefore to improve the model is very necessary.

3.5. Factor Analysis

Factor analysis is a very common dimension reduction method, which can transform multiple variables into a few comprehensive variables, and these few comprehensive variables can reflect the vast majority of the information of the original variables, while these comprehensive variables are not correlated with each other. At the same time, the factor analysis method is to solve the multicollinearity in the model of a kind of feasible method, first by factor analysis of several of the existence of linear variable into a factor, to do with factor analysis of several factors and dependent variable regression analysis, and no significant linear correlation between the factor here, so can well solve the problem of multicollinearity. Therefore, factor analysis can solve the multicollinearity problem between variables on the one hand, and simplify the model by dimensionality reduction on the other hand. Therefore, factor analysis is very feasible in this paper.

Before factor analysis, it is necessary to check whether the requirements of factor analysis method can be met:

```
proc corr data = meidi;
var tclose open highest lowest pclose quantity turnover;
run;
```

Correlation analysis was conducted on the 7 variables in this paper (Pearson correlation coefficient was output by default). As shown in **Table 4**, Pearson correlation coefficient between most variables was greater than 0.3, among which, variable open and tclose; Variables highest and tcloes, open; Var lowest and tclose, open, highest; The correlation coefficients between variable $pclose$ and tclose, open, highest and lowest are all above 0.9. Generally speaking, when most correlation coefficients between variables are greater than 0.3, we believe that factor analysis can be adopted. Because factor analysis uses comprehensive variables to replace original variables, it is meaningless to carry out factor analysis if the correlation between original variables is relatively weak. Thus, this paper is suitable for factor analysis.

Table 4. Correlation coefficients matrix.

	tclose	open	highest	lowest	pclose	quantity	turnover
tclose	1.0000	0.9762	0.9809	0.9803	0.9846	-0.0899	0.3512
open	0.9762	1.0000	0.9946	0.9956	0.9882	-0.1064	0.2375
highest	0.9809	0.9949	1.000	0.9949	0.9938	-0.0698	0.3774
lowest	0.9803	0.9956	0.9949	1.000	0.9948	-0.1471	0.2023
pclose	0.9846	0.9882	0.9938	0.9948	1.000	-0.1096	0.2384
quantity	-0.0899	-0.1064	-0.0698	-0.1471	-0.1096	1.000	0.9297
turnover	0.3512	0.2375	0.3774	0.2023	0.2384	0.9297	1.000

Table 5. Eigenvalue information of correlation matrix (patial data).

	eigenvalue	Cumulative contribution rate
1	6.03807928	0.6709
2	2.90758595	0.9940

```
proc factor data = meidi (drop = date) priors = smc plots = (scree);#consider extract common factors#
```

```
run;
```

As can be seen from **Table 5**, since the eigenvalues of all factors are positive, here is the eigenvalue information of the correlation matrix. Generally, the first m main factors are extracted to make the cumulative contribution rate reach more than 70%. In this paper, the cumulative contribution rate of the first two main factors has reached 0.9940, which can explain most of the information of the original variable. Therefore, further optimization was considered to extract two common factors.

```
proc factor data = meidi priors = smc
```

```
rotate = promax reorder
```

```
plots = (scree initloadings preloadings loadings) nfactors = 2
```

```
score out = datacity; #calculate the factor score coefficient#
```

```
var open highest lowest pclose quantity turnover;
```

```
run;
```

In this paper, the SMC method is used to calculate the principal factor matrix. Let $R = AA' + D$, $R^* = AA'$ is called reduced correlation, matrix. When the eigenvalues and eigenvectors of the reduced-correlation matrix are calculated, A and D are solution of the factor model, which is called the principal factor solution. (Where A is the factor load matrix, $D = \text{diag}\{d_1^2, d_2^2, \dots, d_p^2\}$). Because through orthogonal rotating load values are often not easy to explain the two main factors, so we usually oblique rotation, namely on the basis of maximum variance orthogonal rotation for the optimal rotate oblique factor, and the factor of after rotation in order of from big to small, factor of rotation by changing the coordinate axis, able to redistribute the proportion of each factor to explain the

original variable variance, At this time, variables tend to show the phenomenon of heap aggregation, which makes the leading factor easier to explain.

The analysis of the factor load matrix after oblique rotation in **Table 5** shows that these two main factors can be explained as price factor and quantization factor respectively. Factor1 has a relatively large load value at lowest, open, pclose and highest, which are 1.00118434, 0.99630711, 0.99563705 and 0.9955432, respectively, so Factor1 can be interpreted as a price factor. Factor2 has a relatively large load value in quantity and turnover variables, which are 0.98949435 and 0.9661066 respectively, so it can be interpreted as a quantization factor. Therefore, the higher the lowest price, opening price, closing price and highest price of the stock, the higher the price of the stock will be. The turnover of this stock, clinch a deal the amount is higher, criterion this stock is opposite current quantity is more. See **Table 6**.

The principal factor diagram after the skew matrix transformation in **Figure 1** shows that the six variables can be divided into two groups. One group includes

Table 6. Factor loading array after oblique rotation.

	Factor1	Factor2
lowest	1.00118434	-0.0393359
open	0.99630711	-0.0001631
pclose	0.99563705	-0.0013343
highest	0.9955432	0.03894399
quantity	-0.1691392	0.98949435
turnover	0.17981254	0.9661066

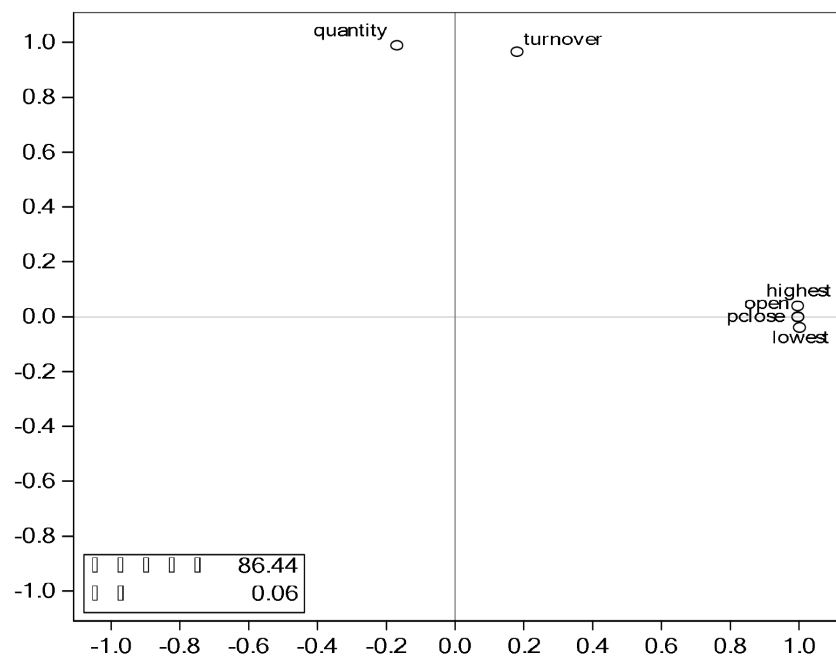


Figure 1. Principal factor diagram of an oblique matrix transformation.

the information of the opening price, the highest price, the lowest price and the closing price of the day. The other group includes the daily trading volume, such as the volume of the information. And these two groups are the two groups of information represented by the two common factors extracted by us, showing the phenomenon of aggregation respectively. Therefore, it indicates that the main factor after oblique rotation may be easier to explain. The two principal factor functions obtained after oblique rotation are: (the asterisk represents standardized variables):

$$F1 = 0.70394lowest^* + 0.0171open^* + 0.02138pclose^* + 0.31418highest^* + 0.07126quantity^* - 0.05706turnover^*$$

$$F2 = -0.20930lowest^* + 0.04341open^* + 0.03335pclose^* + 0.12583highest^* + 0.49232quantity^* + 0.50874turnover^*$$

Factor1 (F1) and Factor2 (F2) values of business days since January 1, 2022 to May 27, 2022 can be obtained by calculating the standardized values of the corresponding variables every day and putting them into the two equations respectively. Meanwhile, factor1 (F1) and Factor2 (F2) can be sorted in ascending order respectively. According to the ranking of factor1, the corresponding score on January 10, 2022 is the highest, indicating that the stock price of Midea Group on that day is generally high. According to the ranking of main factor2, the corresponding score is the highest on April 26, 2022, indicating that the trading volume and total number of shares of midea group are the highest on that day. At this time, multiple linear regression analysis was performed on the factor scores corresponding to the extracted two common factors F1 and F2, and the next day's closing price of the dependent variable.

The dependent variable is regressed to the two common factors, and the variance inflation factor value and DW value are output at the same time

```
proc reg data = a(drop = date);
```

```
model tclose = factor1 factor2/vif dw; # The dependent variable is regressed to the two common factors, and vif value and DW value are output at the same time# run;
```

According to **Table 7** and **Table 8**, the multiple linear regression equation of

Table 7. Parameter estimation of model 2.

variables	Parameter estimation	Standard error	Pr > t	vif
Intercept	63.45323*	0.16656	<0.0001	0
Factor1	8.27695*	0.16693	<0.0001	1.00388
Factor2	0.13818*	0.16758	0.04	1.00388

Table 8. Summary of model 2 information

model 2	R ²	Adjusted-R ²	value of F	value of P	value of DW
	0.9649	0.9641	1236.92	<0.0001	1.773

the two common factors F1 and F2 is:

$$tclose = 63.45323 + 8.27695 * factor1 + 0.13818 * factor2 ;$$

According to the regression equation corresponding to Model 2, the two common factor coefficients are both positive, that is, both the price factor and the quantity factor are in direct proportion to the stock closing price of the next day of the dependent variable, which conforms to the practical economic significance. The results of variance analysis of the model showed that the p-values were all lower than the significance level, indicating that the model was significantly established as a whole. The t test of the single variable corresponding to the two variables F1 and F2 passed the significance test, so the two variables were also significant at this time. The decision coefficient R square = 0.9649, the adjustment R square = 0.9641, the model fitting effect is very good. 96.49% of the variation of dependent variables can be caused by the change of common factors F1 and F2. The DW value = 1.773, located near 2, also satisfies the assumption that the residuals are not correlated.

At the same time, it can be seen from **Figure 2** that the QQ map of residual also approximately obeies the assumption of normality, that is, the residual approximately obeies the normal distribution. At this point, the variance inflation factors of the two common factors are both 1.0038 and less than 10. Therefore, extracting common factors through factor analysis can effectively eliminate the influence of multicollinearity between variables on the model. In conclusion, compared with Model 1, model 2 improved by factor analysis is more in line with our expectations. Therefore, the next step is to investigate whether the prediction effect of Model 2 is good. See **Figure 2**.

4. Model Prediction

In order to investigate the effect of model 2, six trading days from May 20 to May 27 were selected as the model test backtest period to observe whether the difference between the predicted value and the real value was significant, so as to judge the effect of model establishment.

```
%let equation = 63.45323 + 8.27695 * factor1 + 0.13818 * factor2;
%put & equation;
data predict;
```

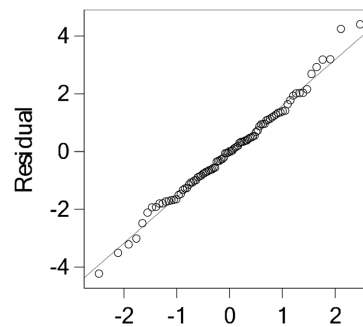


Figure 2. Q-Q chart of residual test corresponding.

```

set meidi;
if factor1^ = . and factor2^ = .;
pred = &equation;
run;

```

Based on the data set predict, the predicted values of the next-day closing prices from May 20 to May 27 were 55.55, 56.03, 55.71, 54.03, 53.28 and 53.11, respectively. The closing prices of the next day from May 20 to May 27 were: 56.9, 54.31, 53.79, 53.4, 53.82, 53.12; The absolute value of error (the true value of the next day's closing price from May 20 to May 27—the predicted value of the next day's closing price from May 20 to May 27) is 1.35, 1.72, 1.92, 0.63, 0.54, 0.011, respectively.

At this point, the difference between the predicted value of the model and the real value is very small, so it can be seen that the prediction effect of Model 2 is very good. Therefore, when given the information of the opening price, highest price, lowest price and closing price of the day of the known independent variable, the corresponding factor score function is obtained through the method of factor analysis, and the score of the corresponding extracted common factor is calculated. Linear regression between the dependent variable and the common factor can predict the price of the next day's closing price of the dependent variable. In this paper, the stock price prediction model of Midea Group is finally determined as Model 2: $tclose = 63.45323 + 8.27695 * factor1 + 0.13818 * factor2$.

5. Summary

In this paper, six factors that may affect stock price are considered from the technical aspect, and a multiple linear regression model of dependent variable and six independent variables is established. However, considering that the model at this time has very serious multicollinearity and contains many independent variables, we seek a method that can eliminate multicollinearity and reduce dimension at the same time—factor analysis. Two common factors were extracted through factor analysis, and factor scores corresponding to different dates were calculated respectively. Then multiple linear regression was performed on the dependent variable and the two common factors to obtain model 2. Considering whether the prediction effect of Model 2 is accurate, the next-day stock closing price from May 20 to May 27 is selected as the test backtest period, and the prediction effect of Model 2 is found to be very good by comparing the predicted result with the actual corresponding next-day closing value, so as to obtain the final stock price prediction model. (Yu, Li, & Yin, 2021)

This stock price prediction model successfully predicts the stock price of Midea Group, providing investors with a new method to predict stock price. By analogy, this method can be used to predict related stock prices for different stocks. Meanwhile, feasible suggestions are put forward for investors:

- 1) Price indicators (such as open price, high price, low price and close price) play a crucial role in predicting the future stock price; Relatively speaking, the

quantity index (such as turnover, trading volume, total number of hands) has a relatively small effect on stock price, but still cannot be ignored. Accordingly, what we should consider above all most should be the stock past time period price and whole stock price trend; consider the factor such as rate of total hand of this stock again number, change hands. (Xie & Shang, 2020)

2) This model is suitable for the prediction of short-term stock price of Midea Group, and it is very convenient to use this model in practical operation, which has reference significance for investors engaged in short-term stock. If the stock price rises in the short term or the stock circulation increases, it is highly likely that the stock price will increase in accordance with this trend. Therefore, different investors can make different decisions to sell or buy stocks in different situations. (Xin, 2014)

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Feng, X. J., & Meng, X. (2011). Stock Price Prediction Based on BP Improved Algorithm with Parameter Modification. *Proceedings of 2011 AASRI Conference on Artificial Intelligence and Industry Application (AASRI-AIIA 2011 VI)* (pp. 330-333).
- Li, G. C., Xu, L., & Zhang, L. (2022). Stock Price Prediction Analysis based on LSTM. *Intelligent Computer and Application, 12*, 123-128.
- Meng, K., & Li, L. (2016). Empirical Analysis of Stock price Prediction Based on ARMA Model. *Journal of Hebei North University (Natural Science Edition), 32*, 55-60.
- Xie, H. N., & Shang, Y. (2020). Stock Price Prediction Based on ARMA Model Analysis—Taking China CITIC Bank as an Example. *Northern Economic and Trade, No. 12*, 122-124.
- Xin, C. (2014). Prediction and Analysis of Listed Companies' Stock Prices under the Theory of "Cloaking Effect"—Based on the Performance of Different Industries under Negative Shock. *Finance and Accounting Communications, No. 29*, 4-6.
- Yu, C. L., Li, M. Y., & Yin, W. S. (2021). Stock Price Prediction Analysis Based on PCA-BP Portfolio Model. *Journal of Changchun University of Science and Technology (Natural Science Edition), 44*, 125-130.