

Eviews and SPSS-based Stock Price Analysis and Forecast of Ningde Times

Weifan Kong^{1,a,*}

¹School of Humanities and Social Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China

a. 2462927068@qq.com

**corresponding author*

Abstract: China has emerged as the largest developing country in the world. Many Chinese people use stock speculation to earn income. However, it is widely known that market stock prices fluctuate significantly. If an effective method can be found to forecast stock prices, it will enable investors to reduce unnecessary losses. In this essay, based on Eviews13, a linear regression is conducted using a full year of stock history data of Ningde Times 2022-2023. Its daily opening price, high price, low price, closing price, and total daily after-hours turnover are employed as independent variables, while the stock's closing price on the following day is used as an dependent variable to build a regression model. Subsequently, heteroskedasticity and autocorrelation tests are performed, along with a principal component analysis. The results indicate that historical stock information can be used to predict stock prices in a short period of time in the future.

Keywords: E-views, multiple linear regression, stock price analysis, autocorrelation test, principal component analysis.

1. Introduction

Currently, it is difficult to accurately predict stock prices because China's financial market has not yet reached a semi-strong efficient market, as Chang[1] stated. In such a market efficiency state, investors can only obtain insufficient public and past information, which leads to a compromised level of information disclosure and affects the efficiency of disclosure. Therefore, there is a limitation when predicting stock prices. This essay will use regression analysis research to develop a linear regression model to make short-term forecast of stocks prices using mathematical and logical methods. Additionally, the variables of technical indicators will be added to reduce the impact of the above limitations and analyzed in a factor-principal component regression study. By building effective short-term stock price forecasting models, it improves the science and accuracy of investor decision-making, providing an effective tool to manage risk and optimizing investment strategies. Stock price forecasting models are not only a tool for investors but also provide a rich resource for academic research and financial education. By analyzing model performance and results, researchers can gain a deeper understanding of market dynamics and investment theories.

2. Application of Multiple Linear Regression Models in Stock Price Forecasting

Regression analysis is a method of mathematical and statistical analysis. As Song and Deng [2] stated, it takes the data of a number of observed variables as the starting point and analyzes the structure of the data to find out the dependence among them, thereby studying the correlation between variables.

Regression models can be categorised into linear and non-linear regression models according to whether or not there is linearity between the independent and dependent variables in the model. The function when the regression is linear is:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

beta 0 is the regression constant and beta 1,...,beta n are the regression coefficients.

2.1. Establishment of multiple regression model

This essay will use historical data of the Ningde Times to establish model. The daily opening price, high price, low price, closing price, and total daily after-hours turnover are used independent variables, and the stock's closing price on the following day is used as an dependent variable.

2.2. Prediction research of multiple linear regression model in Ningde Times

In estimating equation, R-squared represents the goodness of fit and is used to measure the degree of fit of the model. The closer its value is to 1, the better the fit. R-squared also represents the coefficient of determination and is used to reflect the percentage of variance in the dependent variable that can be explained by the model. Again, the closer its value is to 1, the better. Adjusted R-squared is the corrected value of R-squared, taking into account the influence of the degree of freedom factor, thus eliminating the influence of the increase in the independent variables, and more accurately assessing the degree of model fit, and its value is closer to 1 the better, as Wang[3] said.

The historical data of the stock is brought to Eviews13 and the data is processed with this application to get the value of the original model adjusted R-squared is 0.985718, Durbin-Watson stat is 1.915915. Besides, the correlation coefficients of daily turnover, daily closing price, daily high, daily low and daily close are 9.652845, -0.000231, -0.025707, 0.109783, -0.035444, 0.928138.

Let the closing price be Y, daily turnover be X1, daily closing price be X2, daily high price be X3, daily low price be X4 and daily close price be X5. The multivariate linear equation is then obtain as:
Y=9.652845-0.000231X1-0.025707X2+0.109783X3-0.035444X4+0.928138X5

squared is 0.986014

The remaining data are shown below(Table 1):

Table1: Results analysed using Eviews.

Dependent Variable: PRECLOSEPRICE				
Method: Least Squares				
Date: 07/14/24 Time: 11:51				
Sample: 1 243				
Included observations: 243				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.652845	3.543363	2.724204	0.0069
AHSHRTRD_D	-0.000231	0.000343	-0.674832	0.5004
CLSPRC	-0.025707	0.084973	-0.302533	0.7625
HIPRC	0.109783	0.098304	1.116777	0.2652
LOPRC	-0.035444	0.091290	-0.388253	0.6982

Table 1: (continued).

OPNPRC	0.928138	0.081464	11.39320	0.0000
R-squared	0.986014	Mean dependent var		441.3595
Adjusted R-squared	0.985718	S.D. dependent var		51.94893
S.E. of regression	6.208184	Akaike info criterion		6.513995
Sum squared resid	9134.348	Schwarz criterion		6.600244
Log likelihood	-785.4504	Hannan-Quinn criter.		6.548735
F-statistic	3341.583	Durbin-Watson stat		1.915915
Prob(F-statistic)	0.000000			

Both R-squared (0.986014) and adjusted R-squared (0.985718) are closed to 1, indicating that the model explains the observed data well and the regression equation is good. Meanwhile, the value of Durbin-Watson test is 1.915915, close to 2, indicating that residual terms are not autocorrelated.

3. Principal component analysis.

In this section, a linear regression and factorial principal component regression analysis of the short-term stock price will be studied using the technical indicators of the stock price, applying SPSS software.

3.1. Basic concepts and data selection.

Principal Component Analysis (PCA) is a “dimensionality reduction” algorithm that transforms multiple indicators into a small number of composite indicators. These composite indicators are linear combinations of the original variables and are independent of each other. PCA effectively retains most of the original data information. Through PCA, “noise” and unimportant indicators can be removed, which effectively reduces the probability of information loss in the transformation and improves the speed of data processing. [3,4,5] There are three important indicators that are used when analysing stock prices using principal components: trend-based indicators, overbought and overbought indicators and popularity-based indicators Zhang and Dong[6].

Based on various literature discussions, selected portions of the three main types of indicators will be analysed. The selected indicators are: moving average (MA), exponentially smoothed moving average of dissimilarity (MACD), relative strength indicator (RSI), divergence indicator (BIAS), Williams indicator (WMS), and energy tide indicator (OBV).

For convenience, in the following essay: X1 is MA X2 is MACD X3 is RSI X4 is OBV X5 is BIAS X6 is WMS.

Since a short-term stock price prediction model is being built, the dependent variable corresponds to obtaining the future stock price, and y is used to denote the opening price on the next trading day. The raw data obtained for the eleven trading days from 29 August 2024 to 27 September 2024 are shown in the table below (Table 2):

Table 2: 29 August 2024 to 27 September 2024 Ningde Times Technical Indicators

Date	Y	X1	X2	X3	X4	X5	X6
2024-08-29	179.1700	182.5680	0.2930	57.1115	6294645.9800	1.6040	21.1712
2024-08-30	180.8000	183.0760	0.6667	69.5113	6548937.6900	3.0082	9.2150
2024-09-02	184.4400	182.5300	0.7006	53.1140	6419085.9300	0.8820	32.5153
2024-09-03	180.3000	182.9600	1.0750	66.5290	6601054.7900	2.7915	6.8138
2024-09-04	185.3200	183.8260	1.4757	69.9939	6766700.0300	3.0727	10.8824

Table 2: (continued).

2024-09-05	187.1500	185.0260	1.7450	68.0223	6643517.8800	2.3197	15.6695
2024-09-06	186.1200	185.3480	1.8682	62.8618	6568263.0200	1.3020	20.5128
2024-09-09	185.1100	186.1400	1.8796	57.9007	6433509.4700	0.5763	25.0712
2024-09-10	185.5000	185.2420	1.5489	39.3256	6279265.3400	-1.5626	55.0164
2024-09-11	182.4000	185.1240	1.7029	60.1092	6491236.1700	1.2505	23.8247
2024-09-12	188.0000	185.6340	2.0267	67.0971	6698336.0300	2.4935	21.6320
2024-09-13	188.8800	185.1440	1.7906	46.5706	6525463.1500	-0.7035	66.6667
2024-09-18	183.6000	185.3940	1.8166	54.7968	6639242.1000	0.6117	44.3418

3.2. Establishment and testing of full variable linear regression model

After modelling using SPSS software and performing multiple linear regression analysis, the regression model coefficients and multiple covariance tests were obtained as follows (Table 3):

Table 3: Regression model coefficients and covariance test results.

model variables	coefficient	tolerance	VIF
constant term	756.938		
x1	-2.852	.031	31.756
x2	9.757	.021	46.798
x3	.529	.012	83.523
x4	-1.489E-5	.040	24.957
x5	-.591	.011	91.809
x6	.224	.040	24.987

This results in a multiple linear regression equation, denoted as model (1):

$$Y = 756.938 - 2.852X_1 + 9.757X_2 + 0.529X_3 - (1.489E - 5)X_4 - 0.591X_5 + 0.224X_6$$

The data in Table 2 was brought into the model (1) for testing and an average relative error of 0.68% was obtained. By observing the multicollinearity test data in Table 3, the tolerance values are less than 0.1 and the VIF values are very large, these two data indicate that there is a serious problem of multicollinearity in the model, in order to improve the prediction accuracy and to build a more effective model, in the following will be used in the factor-principal component analysis

3.3. Principal component regression analysis

Normalisation of the data was first carried out and then Table C was obtained (Table 4)

Table 4: Standardised data

Zx1	Zx2	Zx3	Zx4	Zx5	Zx6
-1.48332	-2.03303	-0.25112	-1.61702	0.17373	-0.51267
-1.08557	-1.3648	1.07632	0.11918	1.1629	1.45865
-1.51308	-1.30419	-0.67906	-0.76739	-0.33488	0.07849
-1.17639	-0.63471	0.75706	0.47502	1.01025	-1.26085
-0.49834	0.08179	1.12798	1.60598	1.20834	-1.04883
0.44124	0.56333	0.91692	0.76494	0.67789	-0.79937
0.69336	0.78362	0.36447	0.25113	-0.03902	-0.54698

Table 4: (continued).

1.31348	0.80401	-0.16663	-0.66891	-0.55023	-0.30943
0.61036	0.21268	-2.15515	-1.72203	-2.05696	1.25106
0.51797	0.48805	0.0698	-0.27478	-0.07529	-0.37439
0.91729	1.06704	0.81787	1.13922	0.80033	-0.48865
0.53363	0.64487	-1.37955	-0.04109	-1.45177	1.85817
0.72937	0.69136	-0.49891	0.73575	-0.52529	0.69479

The standardised data were subjected to SPSS factor analysis to obtain Table 5 for the factors explaining the total variance of the original variables, and also the factor component matrix Table 6.

Table 5: Cumulative rate of variance of components.

Number of components	Variance contribution accumulation %
1	50.572
2	86.094
3	96.449
4	99.747
5	99.913
6	100.000

As can be seen from Table 5, the variance contribution ratio of the first three factors reached 96.449%, which is greater than 80%, and the result is ideal, which can be seen that the information of the original variables is less lost. Meanwhile, the three principal components obtained through principal component analysis were set as z_1 , z_2 , z_3 .

Table 6: Component matrix.

Component matrix	components		
	z_1	z_2	z_3
x_1	-.351	.905	-.101
x_2	-.151	.984	-.052
x_3	.963	.112	.089
x_4	.694	.561	.364
x_5	.979	-.097	.108
x_6	-.721	-.090	.676

From Table E, the relationship between the principal component factors and the original variables is obtained, denoted as system of equations I:

$$z_1 = -0.351X_1 - 0.151X_2 + 0.963X_3 + 0.694X_4 + 0.979X_5 - 0.721X_6$$

$$z_2 = 0.905X_1 + 0.984X_2 + 0.112X_3 + 0.561X_4 - 0.097X_5 - 0.090X_6$$

$$z_3 = -0.101X_1 - 0.052X_2 + 0.089X_3 + 0.364X_4 + 0.108X_5 + 0.676X_6$$

X_1, X_2, X_3, X_4, X_5 and X_6 correspond to the variables in Table 2.

The three factors obtained: z_1 , z_2 , and z_3 , are used as independent variables in the regression model to establish a multiple linear regression equation, which yields Table 7:

Table 7: The coefficients of z1, z2 and z3.

Model variables	Variation coefficients
Constant term	184.368
z1	-0.077
z2	2.207
z3	0.643

Thus model (2):

$$y = 184.368 - 0.077z_1 + 2.207z_2 + 0.643z_3$$

The final test using model (2) yielded an average relative error of 0.146% which is much lower than the previous margin of error.

4. Conclusion

In this essay, historical information of Ningde Times stock was selected to establish a linear regression model on Eviews software. Then, technical indicators were selected to establish two short-term stock price change models. Results were tested and compared by means of all-variable linear regression analysis and factor-principal component analysis. The results show that linear regression has a good fitting effect for the model where independent variables are independent of each other. If there is multicollinearity between variables, it will lead to a large error. Multicollinearity between variables can be eliminated by principal component analysis. In addition, stock price is influenced by many factors. As Jiang and Wan[7,8] wrote in the journal, investor bullish sentiment index and attention index have significant predictive power for short-term stock price changes. In the future, it is possible to develop models that incorporate more influencing factors.

References

- [1] Chang, J. X.(2022) 'A study on the semi-strong efficiency of the Shanghai securities market in China: An empirical analysis based on listed companies' dividends', *China Market*, 1, pp. 42-43. doi:10.13939/j.cnki.zgsc.2022.01.042.
- [2] Song, Y. and Deng, S. (2008) 'Application of multiple linear regression analysis in fixed asset investment', *Journal of Shenyang Normal University (Natural Science Edition)*, 2, pp. 160-162.
- [3] Wang, P. D. (2020) 'Analysis and prediction of stock prices based on multiple linear regression', *Science and Technology Economy Market*, 1, pp. 84-85.
- [4] Liu, L. M. and Zhang, Y. N.(2005) 'A simple application of factor analysis and cluster analysis', *Journal of Shenyang Normal University (Natural Science Edition)*, 2, pp. 130-132.
- [5] Liu, L. M.(2008) 'Using principal component regression analysis to solve multicollinearity problems in regression models', *Journal of Shenyang Normal University (Natural Science Edition)*, 1, pp. 42-44.
- [6] Zhang, P. and Dong, L. (2010) 'Technical analysis of stock prices based on factor and principal component regression analysis', *Journal of Shenyang Normal University (Natural Science Edition)*, 28(2), pp. 169-172.
- [7] Jiang, Y. and Qiao, Y. Y. (2024) 'Evaluation of operational performance of logistics enterprises based on PCA-DEA model', *Logistics Technology*, 5, pp. 44-47, 54. doi:10.13714/j.cnki.1002-3100.2024.05.011.
- [8] Wan, Y. P., Zhao, Y. and Zhao, L. (2024) 'Investor sentiment and stock investment returns: Text analysis of social media using machine learning', *Financial Market Research*, 5, pp. 65-77. doi:10.20134/j.cnki.fmr.2024.05.012.