WEP Warwick Evans Publishing

Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 12
EDI 2024

A Study of Stock Market Volatility Prediction Based on Traditional Regression and Intelligent Algorithms

Ruijie Zong *, Shen Xin

School of Finance, Nankai University, Tian Jin, China

*Corresponding author:

Abstract. China's stock market has been turbulent since the reform and opening up, often with the risk of plummeting, and in recent years it has been even more volatile, while the national level's concern for stock market stability is increasing day by day, and research on the stability of China's stock market is imminent. In this paper, on the basis of previous research and traditional macro fundamental analysis, firstly, we use LASSO regression analysis to initially determine the influence indicators that are causally related to stock market stability, and then test the significance of the model coefficients, and amend the original model by using multivariate linear regression, to further screen out the significant influencing factors; secondly, we use support vector machine regression (SVM) and random forest regression to fit the stock market volatility in the machine learning method. regression to fit the stock market volatility, determine the importance ratio of different characteristic variables in the model, and analyse the factors affecting the prediction of stock market volatility; finally, the LASSO regression is combined with machine learning to establish an improved model, and the screened indicator factors are fitted with machine learning models to further deepen the prediction of stock market volatility. Stock market oscillations and volatility are further deepened to measure the stability of the Chinese stock market.

Keywords: Stock market stability; LASSO regression; Support vector machine regression; Random forest regression.

1. Introduction

Since the opening of the Shanghai market in December 1990, China's A-share market has experienced more than three decades of storms and many rotations, but currently, driven by policy support, the A-share market has a very good momentum for future development [1-3]. This study will further investigate and predict the stability of the Chinese stock market based on the previous experience by using a combination of multiple regression and SVM models. Innovatively combining LASSO regression and multiple linear regression with machine learning, the study explores the complementary combination of traditional regression causality and machine learning intelligent algorithms by using the indicators affecting the volatility of the Chinese stock market such as the weighted average price-earnings ratio, the national urban unemployment rate, and the average daily turnover rate of stocks [4-8].

2. Stock market stability prediction model based on LASSO regression

2.1 Model building and testing

The LASSO regression model is established with the CSI 300 amplitude index as the explanatory variable and the rest of the indicators as the explanatory variables [9]. Firstly, the value of λ is selected through cross-validation to determine that the mean square error is minimised λ=0.02, and the cross-validation diagram is as follow in Fig.2:

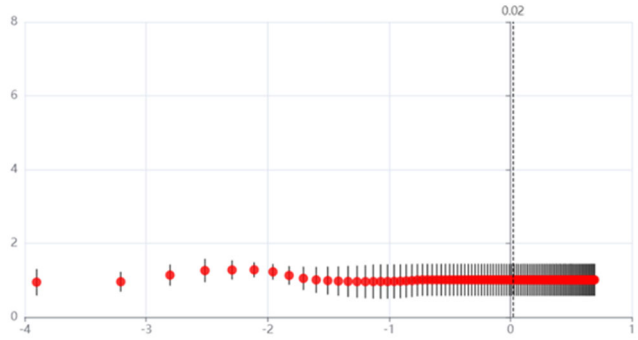

Fig. 1 Cross Validation Chart

On this basis the relationship between λ and the model regression coefficients was determined. As the logarithmic value of λ changes, the model coefficients also change, as shown in Fig. 2:

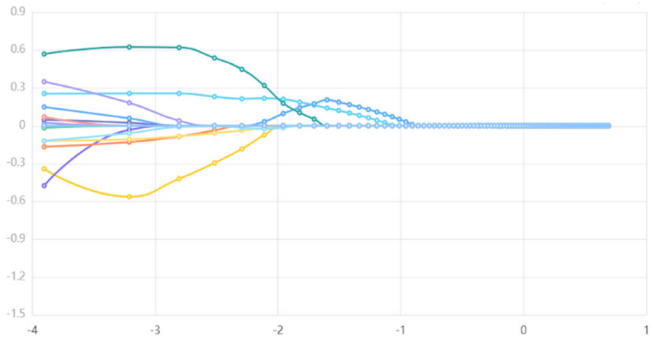

Fig. 2 Plot of λ versus model regression coefficients

The partial results of the model coefficients obtained when the logarithmic value of λ is -3.902 are as follows:

Table 1. Table of model coefficients

Variable	Unstandardized coefficients	R^2
Intercept	0.245	
Domestic and foreign fundraising	0	
Gross Market Value of Equity	0	
Total issued share capital	-0.346	
Market Value of Stocks Outstanding	0	
Equity outstanding share capital	0	
Weighted average price-earnings ratio	0.253	0.468
Average Daily Stock Turnover	0.002	
CSI 300 Closing Index	0	
Average Daily Stock Turnover	0.477	
Year-on-year Growth Rate of Domestic Listed Companies	0	
Year-on-year growth rate of the number of domestic listed stocks	-0.081	
Non-performing Loan Ratio of Commercial Banks (%)	0	
M2 Year-on-Year Growth/M1 Year-on-Year Growth	0	
Investor Sentiment Index (standardised)	0	
Monthlyised risk-free rate (%)	0	
CPI month-on-month	0	
Foreign Exchange Reserves Year-on-Year Growth Rate	-0.091	
RMB Real Effective Exchange Rate Index	0	
National Urban Survey Unemployment Rate (%)	0.057	
Economic Policy Uncertainty Index	-0.019	
Keqiang Index	0	

The unstandardized formula for the LASSO regression model can be obtained from the above table:

$$y = 0.245 - 0.346 * TOS + 0.253 * WAPE + 0.002 * ADV + 0.477 * ADN - 0.081 * SYGR - 0.091 * FXRG + 0.057 * SUR - 0.019 * EPU \quad (1)$$

Fig. 3 Diagram of model results

Observing the results, it can be initially considered that the stability of the stock market is positively correlated with the weighted average price-earnings ratio of the stock market, the average daily stock turnover and volume, and the national urban survey unemployment rate, and negatively correlated with the total issued share capital of the stock, the year-on-year growth rate of domestically-listed stocks, the year-on-year growth rate of the foreign exchange reserves, and the index of economic uncertainty, among which the weighted average price-earnings ratio and the average daily stock turnover have a greater degree of influence on the volatility of the stock market. The model has a goodness of fit of 46.8%, which means that the above variables can be considered to explain 46.8% of the volatility factors in the stock market.

2.2 Lasso regression model significance correction

Through the screening of explanatory variables, the model is corrected by deleting the economic policy uncertainty index and the average daily stock turnover because they do not have a significant effect on stock market volatility [10-11].

Scatter plots between the stock market amplitude index and each explanatory variable were made separately to observe the distribution and determine whether there is heteroskedasticity among the variables.

Fig.4 Scattered distribution of influencing factors

From the scatter distribution of the influencing factors in Figure 4, it can be seen that with the increase in the value of each explanatory variable, the tendency of the scatter distribution of the explanatory variables spreads outwards is more obvious, and there is more obvious heteroskedasticity among the variables. In order to eliminate the influence of heteroskedasticity and autocorrelation on the interpretation of stock market volatility, robust least squares regression is used. The logarithmic and power terms are applied to some of the variables in the regression on the basis of economic significance, and the Eviews regression equations are obtained as follows:

$$\begin{aligned}\log(y) = &-16.45 + 539.97WAPE - 34011.06WAPE^2 \\ &+ 47.70\sqrt{SUR} - 2.66\log(1+FXRG) - 0.94\log(CMV) + 116.26ADV\end{aligned} \tag{2}$$

The equation (2) shows that the amplitude index has a significant positive correlation with the weighted average price-earnings ratio, the national urban unemployment rate, the average daily stock turnover, and a strong negative correlation with the growth rate of foreign exchange reserves and the market value of stocks in circulation, and can be approximated that every 1% increase in the year-on-year growth rate of foreign exchange reserves reduces the amplitude of the volatility of the stock market by about 2.66%; and that every 1% increase in the market value of stocks in circulation reduces the degree of volatility in the stock market by about 0.94 per cent. When observing the stock market, the volatility of the stock market can be predicted by the values of the above three indicators in different periods.

The t-test of the coefficients of each variable passed significantly after the model test, and the standard deviation of the model regression was 0.3896, and the adjusted goodness of fit was 37.32%, which means that the model can explain 37.32% of the volatility sources in the stock market.

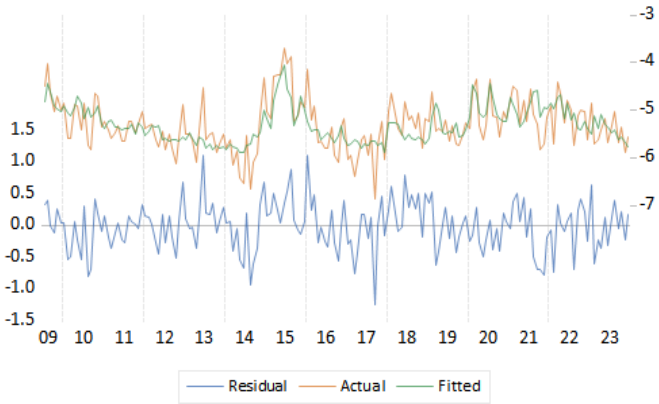

Fig. 5 Fitting effect of the explanatory variables and the distribution of residuals

In Fig. 5, the residuals of the predicted values of the explanatory variables are uniformly distributed on both sides of the 0 point, which is in line with the assumption of the model's random error; the fluctuation trend of the predicted amplitude index is approximately coincident with the original amplitude index, which shows that the model has a good fitting effect and stability.

3. SVM Based Stock Market Stability Prediction Model

3.1 SVM Based Stock Market Stability Prediction Model

After the added independent variables are processed according to the steps of “missing value processing - tailing processing - standardization processin”, each of the above indicators is selected as an independent variable and brought into the model to predict the amplitude index [12]. The particle swarm (PSO) algorithm in the heuristic algorithm is used to build the model, and the model is cross-validated. In the particle swarm algorithm, the initial number of particles is 50, the maximum

number of iterations is 150, the inertia weight is 0.9, and the individual and social learning factors are both 2.

The model kernel function is linear linear kernel function, the kernel function coefficients are scale, the highest number of terms is 3, the penalty coefficient is 1. 10 times cross validation is used, and the final results of the SVR training set and test set are as follows:

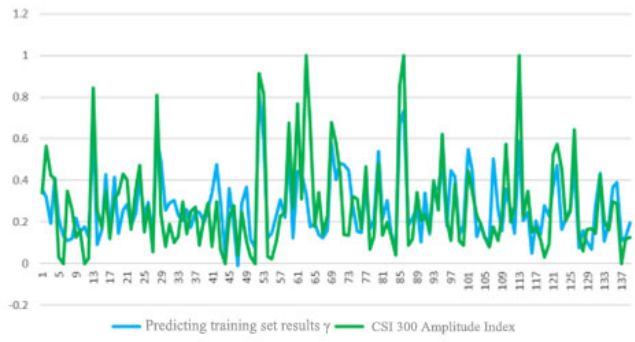

Fig. 6 Training set prediction results

Fig. 7 Test set prediction results display

3.2 Model evaluation and result analysis

In the above table, the prediction evaluation metrics of cross-validation set, training set and test set are shown to measure the prediction effect of support vector regression through quantitative metrics. Among them, the evaluation metrics of the cross-validation set can continuously adjust the hyperparameters to obtain a reliable and stable model. From the evaluation results, it can be seen that the evaluation indexes of the model test set, such as mean square error (MSE) and root mean square error (RMSE), are smaller, indicating that the model is more accurate and the established model is more reliable. The goodness of fit of the model test set is 59.4%, which indicates that the model fits well and the model has good generalisation ability.

Table 2. Assessment of model results

	MSE	RMSE	MAE	MAPE	R^2
Training set	0.024	0.156	0.118	54.766	0.521
Cross Validation Set	0.033	0.177	0.137	102.65	-0.213
Test Set	0.025	0.159	0.111	50.251	0.594

4. A study of stock market stability factors based on random forest

4.1 Model building and solving

On the basis of SVR, the establishment of random from forest regression to determine the weight of different features in machine learning regression [13-14].

The model training set share is 0.8, and the test set data is compared with the actual value of the CSI 300 amplitude index, and the results of the resulting random forest model test set are shown below:

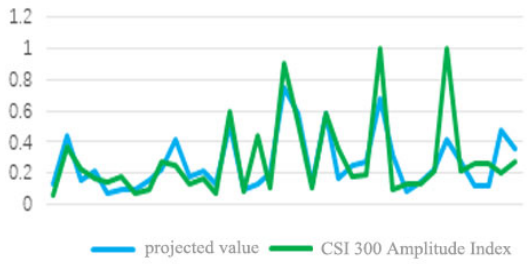

Fig. 8 Random Forest Test Set Results

4.2 Model Evaluation and Analysis of Results

Table 3. Model evaluation (random forest)

	MSE	RMSE	MAE	MAPE	R^2
Training set	0.005	0.068	0.054	24.038	0.912
Cross-validation set	0.031	0.172	0.138	55.084	0.219
Test set	0.026	0.16	0.113	47.713	0.504

By analysing the extent to which each indicator affects the prediction results of stock market amplitude, the effective index of RMB exchange rate, the average daily stock turnover, the weighted average price-earnings ratio and the index of economic policy uncertainty have a greater impact on the prediction results, and the sum of the degree of influence of the four indicators accounts for about 50% of the total degree of influence, and Spearman's correlation analysis is performed on the selected features to check the degree of correlation between the different features as shown in Fig. 9.

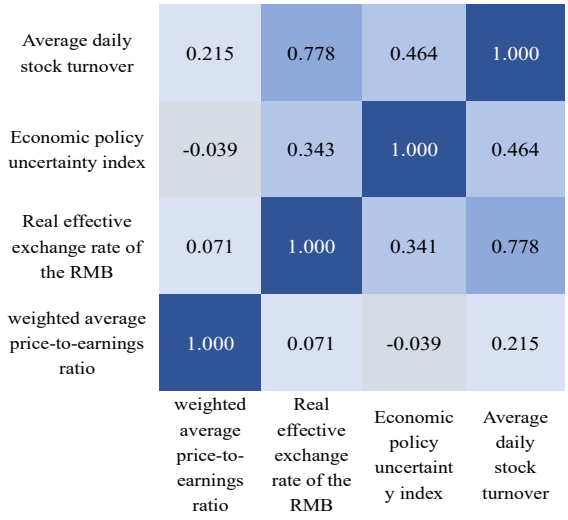

Fig. 9 Correlation coefficient heat map

Fig. 9 shows the correlation coefficients between the different indicators, where there is a more significant correlation between the real effective exchange rate index of the Renminbi (X9) and the average daily stock turnover (X17), and further scatter plots are made to observe the distribution of correlation between the variables.

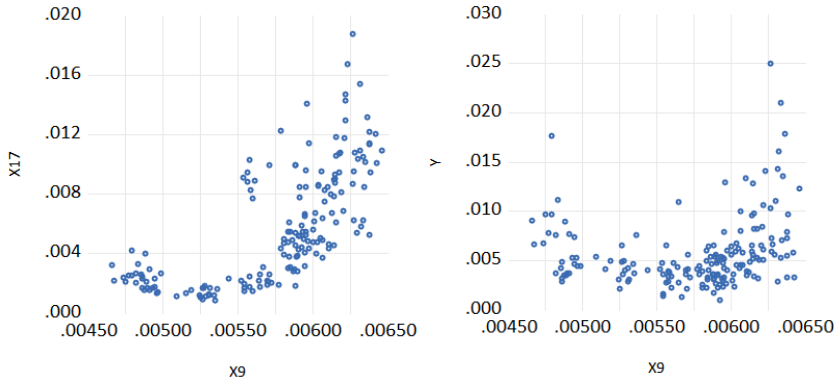

Fig. 10 Variable Correlation Scatterplot

From Fig. 10, it can be seen that the real effective exchange rate index of RMB and the average daily stock turnover are non-linearly correlated, and there is a similar non-correlation with the stock market amplitude index, so the real effective exchange rate index of RMB is excluded, and the remaining three indexes are retained as the analysis of the analytical terms that affect the prediction results. When predicting the degree of stock market volatility, the SVR is used to fit the predictions to the indicators, and the model obtained from the training has a better generalisation ability, and then the importance of the characteristics of the indicators is analysed by the Random Forest regression, and it is found that the weighted average price-earnings ratio, the average daily turnover of the stock, and the index of uncertainty of the economic policy have an impact of nearly 45% on the prediction results. The accuracy of the prediction results can be improved by focusing on the magnitude of changes in the above three indicators when predicting the degree of volatility in the stock market.

5. Summary

Based on the background that the stability of China's stock market needs to be strengthened, this paper, on the basis of past literature and macro-analysis, mainly uses LASSO regression combined with multiple regression model correction to explore the main indicators that have a causal relationship with the stability of the stock market, and then through the SVM model and the Random Forest model to predict the amplitude of the stock market, with the complementary effects of multiple regression and machine learning, we understand that the national Unemployment rate, average daily stock turnover and weighted average price-earnings ratio have a positive impact on the volatility of the stock market, i.e., the larger the value, the more violent the volatility and the worse the stability, of which the unemployment rate has the most significant impact, on the other hand, the growth rate of foreign exchange reserves and the market value of stocks in circulation have a negative impact on the volatility of the stock market and the market value of stocks in circulation has a greater impact. This study innovatively combines multiple regression and machine learning, cleverly integrating the advantages of the former in causal inference with the convenience of the latter in addition to a large amount of data can be further complete optimisation, in order to more accurately predict the volatility of the stock market.

References

[1] Yang Xiaoling, Wang Mengxiao, LI Xiaojuan0, et al. Machine learning-multiple linear regression technique to predict water ecological benchmarks for copper[J/OL]. China Environmental Science:1-11[2024-05-10].https://doi.org/10.19674/j.cnki.issn1000-6923.20240312.002.

[2] Jiang Boat. Mediating and moderating effects in empirical studies of causal inference[J]. China Industrial Economy,2022(05):100-120.DOI:10.19581/j.cnki.ciejournal.2022.05.005.

[3] Zeng Zefan,Chen Siya,Long Zai,et al. A review of causal inference of time series based on observed data[J]. Big Data,2023,9(04):139-158.

[4] Zhou Jiahao, Zhang Mingfu, Len Hongjie. Research on stock selection strategy based on machine learning multi-factor quantitative model[J]. Science and Technology Innovation,2022(05):161-164.

[5] Ouyang Tianhao,Lu Xiaoyong. A new method for measuring the stability of securities market in the context of financial security - A study on market prediction and arbitrage value metrics based on big data support vector machine[J]. Financial Theory and Practice,2019,40(01):77-83.

[6] Liu Haifei, Bai Wei, Li Dongxin, et al. Can the trading system of Shanghai-Hong Kong Stock Connect enhance the stability of Chinese stock market? --A complex network-based perspective[J]. Journal of Management Science,2018,21(01):97-110.

[7] Shao Huaming, Ma Yongtan, Zhu Tao. Stock Market Stability Measurement and Its Role Mechanism - An Analysis Based on the Perspective of Complex Network Model[J]. Financial Science,2017(05):54-66.

[8] Tao Ling, Zhu Ying. Monitoring and Measurement of Systemic Financial Risks--A Study Based on China's Financial System[J]. Financial Research,2016(06):18-36.

[9] Li Zhisheng, Du Shuang, Lin Bingxuan. Short-selling trading and stock price stability-a natural experiment from China's financing and bond market[J]. Financial Research,2015(06):173-188.

[10] Tian Lihui, Wang Guanying, Zhang Wei. Three-factor model pricing:How is China different from the United States? [J]. International Finance Research,2014(07):37-45.

[11] Zhang Xiaoyan,Shen Zhonghua. The impact of stock index futures launch on China's stock market volatility--an empirical analysis based on high-frequency data of CSI 300 stock index futures[J]. Investment Research,2011,30(10):112-122.

[12] Shi Jinfeng, Liu Weiqi, Yang Wei. Financial market stability test based on quantile regression[J]. China Management Science,2011,19(02):24-29.DOI:10.16381/j.cnki.issn1003-207x.2011.02.005.

[13] Malkiel, B.G. and Fama, E.F. (1970), Effi cient Capital Ma rkets: A R eview of Theory and Empir i cal Work. T he Journal of Finance,25: 3 83 41 7.

[14] Zorn T, Dudney D, Jirasakuldech B. P/E Changes: Some New Results[J]. Journal of Forecasting, 2009, 28(4): 358-370.