2021 International Conference on Networking, Communications and Information Technology (NetCIT) | 978-1-6654-0070-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/NetCIT54147.2021.00073

Research on Stock Price Volatility Prediction Based on Generative Adversarial Network

Lu Wang*
School of Science
Nanjing University of Science and Technology
Nanjing, China
e-mail: 869434454@qq.com

Zhensheng Huang
School of Science
Nanjing University of Science and Technology
Nanjing, China
e-mail: stahzs@126.com

***Abstract*—In order to explore the application effect of the most popular Generative Adversarial Network (GAN) in the field of financial forecasting, this paper proposes to explore the predictive ability of GAN's stock price volatility by taking the daily closing price of the S&P 500 index as the research object. The empirical method takes EGARCH model and Long Short-Term Memory (LSTM) as the benchmark model, MSE and MAE as the prediction error measurement indicators, and empirically compares the prediction results of the three models to analyze the out of sample prediction ability of GAN one day in advance. The empirical results show that GAN has the lowest prediction error and the highest prediction accuracy. LSTM also has a good prediction effect, but it is slightly inferior to GAN. EGARCH model has the largest prediction error. It shows that GAN, as a cutting-edge deep learning technology, has a good application prospect in the field of financial time series prediction.**

Keywords—S&P500 index, generative adversarial network, volatility prediction, deep learning

I. Introduction

Accurately predicting stock price fluctuations is an important but difficult task. The stock market plays an important role in the financial field of all countries. It plays an important role in helping the financing of the real economy, guiding the rational allocation of resources and promoting social and economic development. However, the stock market is complex, and the stock price fluctuation is affected by many factors, including policy changes, news, market sentiment and macroeconomic indicators. It has the characteristics of high complexity, asymmetry and nonlinearity. In addition, stock price fluctuation not only affects the income of stock holders, but also many financial derivatives are derived from basic financial products such as stock index, so their income is closely related to stock price fluctuation. Due to the financial crisis of 2008 and the influence of COVID-19 in 2020, the stock market in various countries fluctuated sharply. Therefore, the tracking and prediction of stock price fluctuation is of great value to investment income, risk management and early warning of major financial crisis[1].

The prediction method of stock price volatility with higher prediction accuracy has always been the research focus of scholars at home and abroad. There have been a lot of research on the prediction method of stock price volatility committed to improving the prediction accuracy. This paper focuses on a more cutting-edge generative countermeasure network model, which was first proposed by Ian J. Goodfellow et al. In 2014[2]. Generative countermeasure network is widely used in computer image synthesis, face recognition and speech generation, and has achieved good results[3]. However, there are only about ten literatures on the application of generation countermeasure network in the financial field. In a few literatures, for example, some scholars extend the application of Gan in image synthesis to missing value interpolation of financial time series, which effectively extends image generation to data generation[4]. Several other scholars have used GAN to predict stock closing prices, predict stock price movements, and capture the properties of financial time series[5]. However, there is little research on the application of Gan to predict the variance series of financial time series, such as stock price volatility.

Based on the above research background and current situation, in order to explore the application effect of the most popular generative confrontation network in the field of financial forecasting, we proposed to take the daily closing price of S &P500 index from January 2010 to April 2021 as the research object to explore the stock price volatility prediction ability of GAN. In addition, EGARCH model and LSTM model are selected as the benchmark models. EGARCH model is the most classic model in the traditional financial time series model, which can capture the leverage effect. Meanwhile, LSTM model is one of the most popular neural networks in recent years, which has been proved by many scholars that the accuracy of predicting the fluctuation of financial time series is very high[6]. Through the empirical comparison of the prediction results of the three models, this paper analyzes Gan's prediction ability of stock price volatility, in order to explore the prediction methods or models with higher prediction accuracy.

II. Data and Model Introduction

A. Data characteristics

Considering that The Chinese stock market is characterized by a policy market and strong speculative nature, the American stock market is more mature and rational in comparison[7]. Therefore, this paper chooses the representative S&P 500 composite stock index in the American stock market as the research object. The daily closing price of S&P 500 index from January 1, 2010 to April 1, 2021 is downloaded from the Choice database, with a total of 2840 daily data. As shown in Figure 1 below, this time series shows an upward trend of volatility, in which there is a very violent fluctuation at the end of 2020, which may be caused by the impact of the COVID-19, which leads to the violent turbulence of the overall economy.

978-1-6654-0070-1/21/$31.00 ©2021 IEEE
DOI 10.1109/NetCIT54147.2021.00073

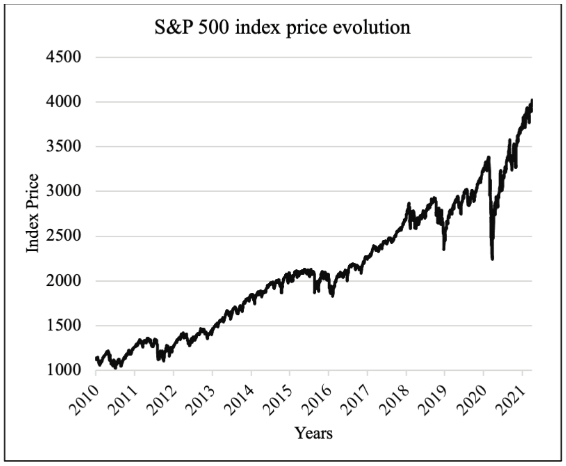

Fig. 1. S&P 500 index price

Then, we test the stationarity of S&P500 index price series. The results show that the ADF test value is 0.437 > - 2.567 (10% critical value), so the original hypothesis is accepted, that is, the S & P500 index series is non-stationary. Then, we calculate the logarithmic return series of S & P 500 index, and get that the ADF test value is -62.125 < -3.433 (1% critical value). Therefore, the original hypothesis is rejected, and it is converted into a stationary series. As shown in Figure 2 below, the logarithmic return series fluctuates around the mean of 0.

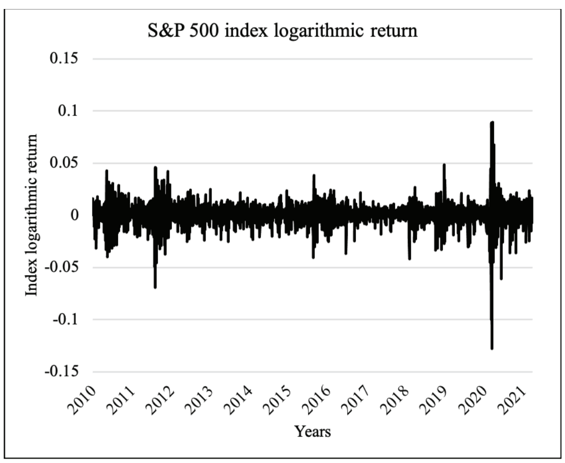

Fig. 2. S&P 500 index logarithmic return

In Figure 3, the histogram of S&P 500 index log return series is shown on the left, and the descriptive statistics of the series are shown on the right. It can be seen that the skewness of the yield series is -0.857, the kurtosis is 19.11, and the value of Jarque-Bera test[8] is 30998.11. In addition, we also do the ARCH-LM test on the residual of the return series, and get the F statistic of the fifth order lag is 336.2650, and the P value is 0. In conclusion, we judge that the log return series has the characteristics of sharp peaks and fat-tailed, non-normal and heteroscedasticity. And the arch effect is significant, so it can be analyzed by financial time series model.

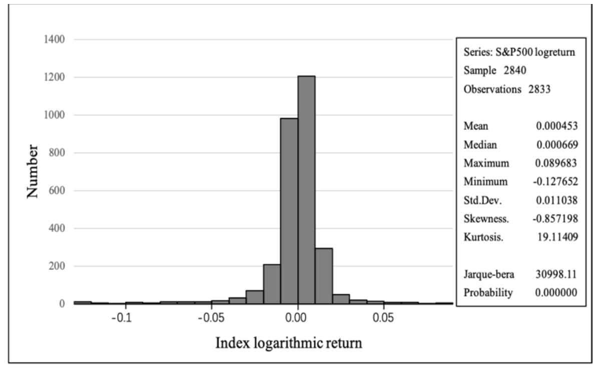

Fig. 3. Histogram of the S&P 500 index logarithmic return

B. Generative adversarial networks (GAN)

Generative adversarial network is an unsupervised non-parametric neural network model composed of two core parts, generator and discriminator, as shown in Figure 4. Compared with the traditional neural network, the generative adversarial network has made a great improvement. It is a new unsupervised antagonism framework, the core of which is to introduce antagonism to realize self-training and reinforcement learning of the model. The generator and discriminator using GAN in this paper are composed of a simple feedforward neural network MLP.

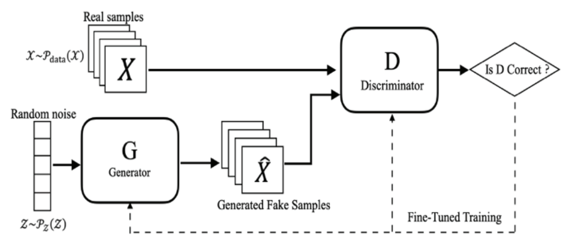

Fig. 4. The architecture of the Generative Adversarial Network

The objective function of GAN model[2] is:

$$\min_{G}\max_{D} V(D,G) = E_{\mathcal{X}\sim P_{data}(\mathcal{X})}\left[log\left(D(\mathcal{X})\right)\right] + E_{\mathcal{Z}\sim P_{\mathcal{Z}}(\mathcal{Z})} (1)$$

The objective function is a maximum-minimum optimization problem, which corresponds to two optimization processes in training. The goal of the generator is to generate "false" data to deceive the discriminator to the greatest extent, and the goal of the discriminator is to minimize the probability of error in distinguishing the two. During training, the discriminator feeds a gradient back to the generator, which updates the parameters accordingly. The optimal state of the model is the game process of continuous confrontation between the two and finally reaching the equilibrium state.

III. Empirical Analyses

A. Empirical Process

We select a total of 2840 S&P 500 closing price series from 2010 to 2021, and first calculate the volatility series of S&P 500 logarithmic return based on a 22-day rolling window. The obtained volatility sequence is divided into two segments, with the first 80% as the training set and the remaining 20% as the test set[9]. Then, EGARCH model, LSTM model and Gan model are used to predict respectively, and the deviation degree between the predicted volatility and the actual volatility is calculated. Among them, EGARCH model and LSTM model are used as the benchmark model, and MSE and MAE, the two loss functions, are used as the evaluation indexes of the prediction effect[10]. The entire empirical process was compiled under Pytorch in Python 3.8.

B. *Empirical Results*

1) *Prediction results based on EGARCH*

TABLE I. PREDICTION ERROR BASED ON EGARCH MODEL

Error	EGARCH Model
MSE	2.41E-01
MAE	8.00E-03

As can be seen from Table I, the mean square error between the true value of S&P 500 volatility and the predicted value obtained by EGARCH model is 2.41E-01, and the mean absolute error is 8.00E-03, indicating a large prediction error.

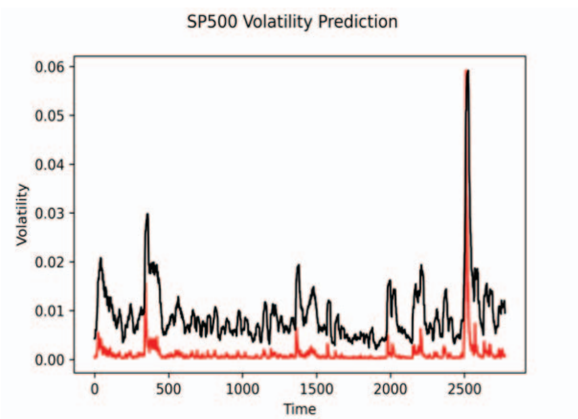

Fig. 5. Prediction effect based on EGARCH model

In Figure 5, we draw a comparison diagram between the volatility predicted by EGARCH model and the actual volatility, where the black line represents the true value of volatility and the red line represents the predicted value of volatility. It can be found that there is a slightly large error between the real value and the predicted value, the prediction effect is not good, and the prediction accuracy is not high enough.

2) *Prediction results based on LSTM*

TABLE II. PREDICTION ERROR BASED ON LSTM MODEL

Error	LSTM Model	
	Train	***Test***
MSE	1.21E-05	6.23E-05
MAE	2.20E-03	4.04E-03

It can be seen from Table II that during the training of LSTM model, the mean square error between the real value of S&P 500 volatility and the predicted value obtained through LSTM is 1.21E-05, and the average absolute error is 2.20E-03. During the testing process, the mean square error is 6.23E-05, the average absolute error is 4.04E-03, and the prediction error is very small.

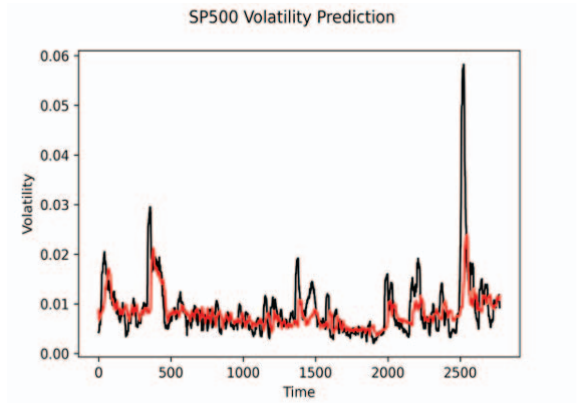

Fig. 6. Prediction effect based on LSTM model

In Figure 6, we draw a comparison diagram between the volatility predicted by LSTM model and the actual volatility, where the black line represents the real value of volatility and the red line represents the predicted value of volatility. It can be found that the error between the real value and the predicted value is not big, the prediction effect is good, and the prediction accuracy is relatively high.

3) *Prediction results based on GAN*

TABLE III. PREDICTION ERROR BASED ON GAN MODEL

Error	GAN Model	
	Train	***Test***
MSE	9.43E-07	3.72E-06
MAE	8.10E-04	1.10E-03

As can be seen from Table III, in the training process of GAN model, the mean square error between the real value of S&P 500 volatility and the predicted value obtained through GAN is 9.43E-07, and the mean absolute error is 8.10E-04. In the testing process, the mean square error is 3.72E-06, the mean absolute error is 1.10E-03, and the prediction error is very small.

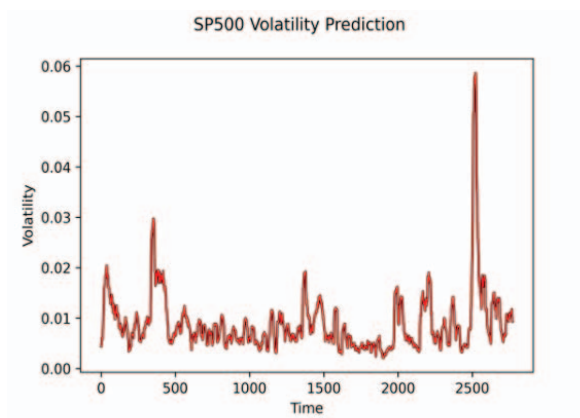

Fig. 7. Prediction effect based on GAN model

In Figure 7, we draw a comparison diagram between the volatility predicted by GAN model and the actual volatility, where the black line represents the true value of volatility and the red line represents the predicted value of volatility. It can be found that the red line almost completely overlaps the black line, so that the error between the two is very small, and the prediction effect is very good, and the prediction accuracy is very high.

IV. Conclusion

In this paper, the 10-year time series data of the daily closing price of the S&P 500 are taken as samples, and the prediction accuracy of stock price volatility of GAN, EGARCH and LSTM models is compared to analyze the prediction effect of GAN model. The empirical results show that the order of MSE index from small to large is GAN<LSTM<EGARCH. Specifically, the prediction accuracy of GAN (MSE=3.72E-06) is nearly ten times higher than that of LSTM (MSE=6.23E-05), and, 10^5 times higher than that of EGARCH model (MSE=2.41E-01). Combined with the effect chart of volatility prediction, we conclude that GAN has the lowest prediction error and the highest prediction accuracy among the three models. LSTM also has a good prediction effect, but it is slightly inferior to GAN. EGARCH model has the largest prediction error. It can be said that GAN, as a cutting-edge deep learning technology, is a very attractive volatility prediction method and is expected to be applied more widely in the field of financial time series prediction. In addition, this paper selects daily data. Because the stock price changes continuously, daily data may lose useful volatility related information. Therefore, the application of GAN in predicting more high-frequency stock price data can be further explored in the future.

References

[1] Markowitz H M. Portfolio selection. The Journal of Finance, 1952, 7(1):77.

[2] Goodfellow I J , Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.

[3] Karamchandani S, Shukla G. Face Sketch-Image Recognition for Criminal Detection Using a GAN Architecture//International Conference on Information and Communication Technology for Intelligent Systems. Springer, Singapore, 2020.

[4] Luo Y, Cai X, Zhang Y , et al. Multivariate time series imputation with generative adversarial networks. 2018.

[5] Zhang, K., Zhong, G., Dong, J., Wang, S., & Wang, Y. (2019). Stock Market Prediction Based on Generative Adversarial Network. Procedia Computer Science, 147, 400–406.

[6] Chen K, Zhou Y, Dai F. A LSTM-based method for stock returns prediction: A case study of China stock market. 2015.

[7] Chang R. Financial Technology: China's Stock Markets vs U.S. Stock Markets. E3S Web of Conferences, 2021, 275(1):01006.

[8] Jarque C M. Jarque-Bera Test. Springer Berlin Heidelberg, 2011.

[9] J G I, J P-A, MIRZA M et al. Generative adversarial networks. Advances in Neural Information Processing Systems, 2014, 3: 2672–2680.

[10] Wen J, University W B. Comparison of the Effects of MSE and MAE on Machine Learning Performance Optimization. China Computer & Communication, 2018.