Research on Volatility Prediction of Stock Returns Based on ARIMA Model

Zhuoyan Gong

Department of Statistical Science, University College London, London, United Kingdom

zhuoyan.gong.22@ucl.ac.uk

Abstract. This article illustrates the prediction of stock return volatilities for enhancing confidence and minimizing the risk of investors in making investments. The most suitable ARIMA model for making a forecast is selected with the help of KPSS and portmanteau test. To avoid the unpredictable influences from macroeconomic factors and to prove the validity of the volatility prediction in the real world, a training dataset and a test dataset are selected from a specific time period with minimum variations in these factors. The forecast data tells that in a period of few variations in macroeconomic factors, the selected ARIMA model could make precise predictions for the volatilities. In the period which the volatilities are strongly affected by factors such as inflation and money supply, the model could still be helpful in reducing the risk and enhancing the accuracy of predictions for the investors if they understand the way volatilities change with these factors.

Keywords: ARIMA model, Volatility, Forecasting, Risk management, Portfolio optimization.

1. Introduction

1.1. Research Background and Significance

Volatility is a critical factor that influences investment decisions and portfolio performance. Making an accurate forecast of the volatilities reduces the risk of making investments. Compared with interpreting the future trends of volatilities manually, the use of statistical models and computer software reduces the skill level needed from investors and enhances the accuracy of predictions made by skilled investors. This could encourage citizens to participate in making investments. Positive effects could be generated in the economy, the GDP and living standard could rise by the increasing amount of investment [1].

This article introduces a method of predicting future stock return volatilities with the help of R. Since there are macroeconomic factors influencing the volatilities, they are not able to be predicted based on past data [2,3]. A specific period with the minimum variation in these factors is selected to build our prediction model. The daily CBOE S&P 500 3-Month Volatility Index from August 2018 to August 2019 is taken as the dataset [4]. The forecast of volatilities in this article is more accurate when there are few variations in macroeconomic factors. When economists suggest that there is going to be a significant change in factors such as inflation and money supply, investors are recommended to make their own judgments along with the use of the ARIMA model.

1.2. Research Contents and Framework

In this article, an ARIMA model that aims to describe the autocorrelation in the data is used to model the historical volatility data. In which the autocorrelations are particularly important in making forecast based on past data. In order to prove if the forecast based on the chosen ARIMA model is valid, various tools such as the KPSS test, portmanteau test, AICc values, normal distribution graph, ACF, and PACF graphs are applied to the data. The stationarity of the data is checked before confirming the number of differences. The consistency of white noise of the residuals from the chosen ARIMA model is confirmed before making forecasts. In order to show the accuracy of this method, the forecast data is compared with the real data. The comparison shows that the prediction is accurate enough.

2. Method

2.1. Selection of Data

This article selects the volatility data in a certain period because the stock return volatilities are closely related to several macroeconomic factors. For mature economies such as the US, the volatilities are particularly related to inflation, money supply, and rate of interest [3].

Figure 1 shows that the inflation in the US became relatively stable from 1996 to 2020, and the expectations of inflation became re-anchored since 1996.

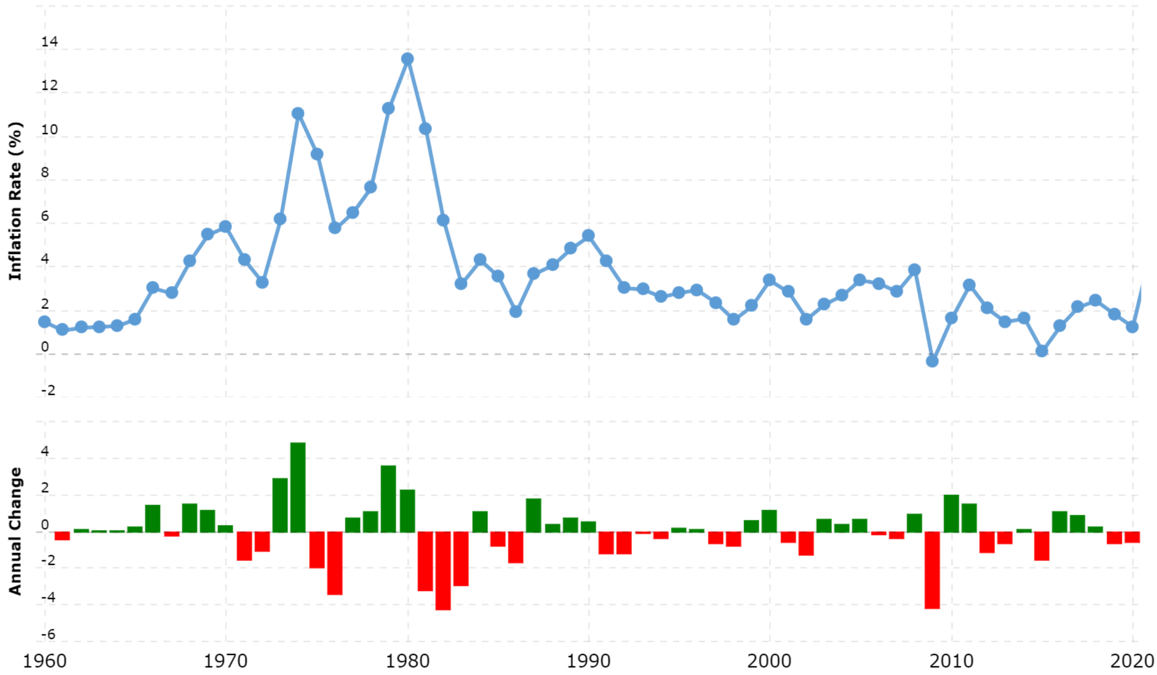

Fig. 1 The US Inflation 1960-2020 [5]

Sources: Minneapolis FED

Fig.2 shows that the money supply in the US has an increasing trend. There are no sudden increases and decreases that cannot be interpreted based on the past data. This means that the US volatility data with stable inflation in the time period 1990 to 2020 can be used to make predictions.

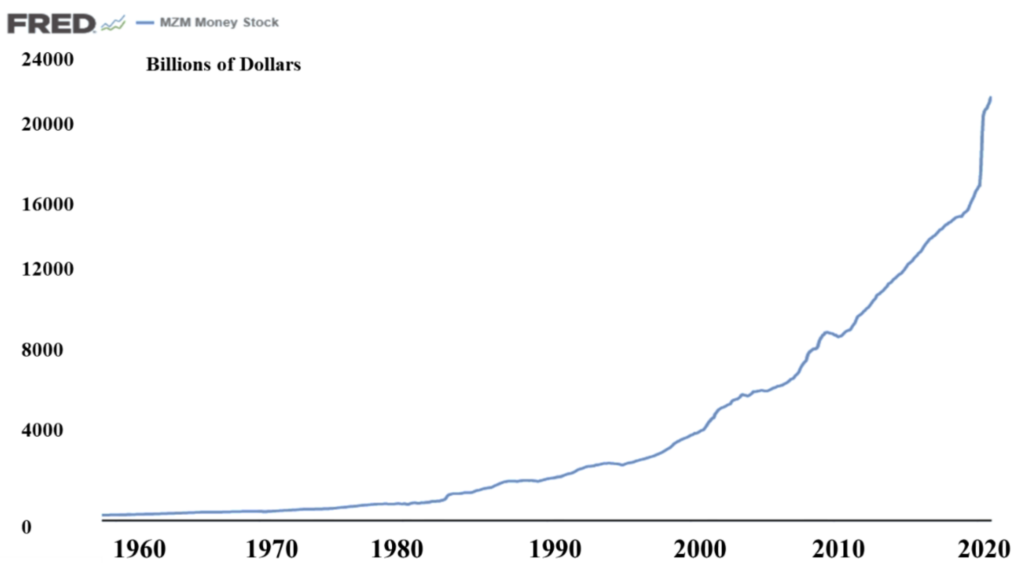

Fig. 2 The US Money Supply [6]

Fig.3 shows that from 1990 to 2020, the rate of interest in the US has a decreasing trend. There are no unexpected turning points in the graphs of the three macroeconomic factors in this time period. If unexpected changes appear, manipulations of volatility values based on the factors should be done along with the use of time series models. According to the three figures, the US volatility data in this period of time can be used to make predictions of future values with models alone.

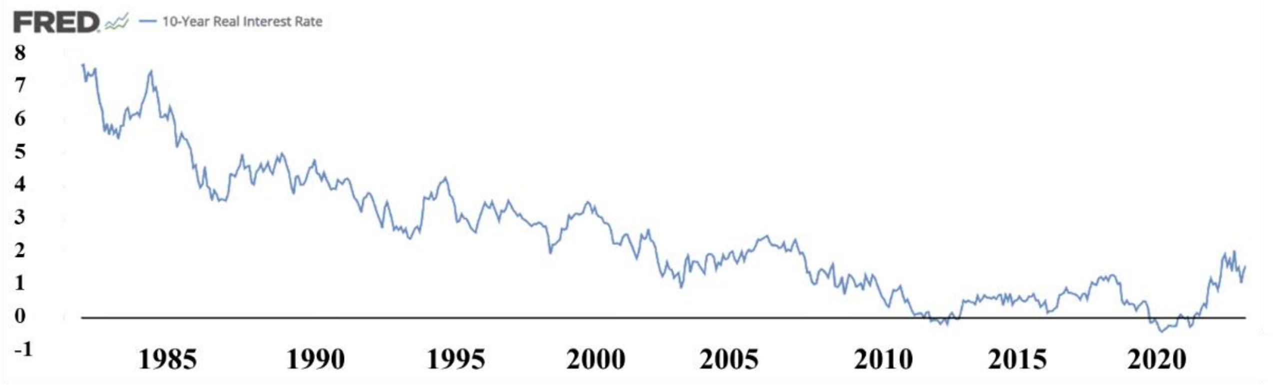

Fig. 3 The US Interest Rates, 1980-2023 [7]

In this paper, the daily dataset CBOE S&P 500 3-Month Volatility Index from August 2018 to August 2019 is used as the training data. The data in the next three months are used as the test data. If the predicted data produced by the time series model which fits the training data well fits the test data well, the model is accurate and can be used to predict future volatility data.

2.2. ARIMA Model

ARIMA model is suitable for the forecast of stock return volatilities. Because it aims to describe the autocorrelations in the data. Autocorrelation refers to the correlation of a stock's price with its past values. It suggests that the stock's previous performance can influence its future performance. It is crucial for technical analysts who use historical price patterns to predict future price movements.

2.3. Selection of ARIMA Model

In this paper, the modelling procedure of ARIMA including the generation of plots and tests is done with the help of R. Before determining the number of differences, the appropriate AR(p) and MA(q) model, the data should be plotted to check if there are any unusual observations or if the variance needs to be stabilized using Box-Cox transformation. When the transformation is done, the plot of transformed data and a KPSS test can be used to check if the data is stationary or not. If the plot is roughly horizontal and has a constant variance, it suggests that the data is stationary and the ACF and PACF plots can be used to select the appropriate ARIMA model. In order to make sure that the observation of the transformed plot leads to the right conclusion, a KPSS test can be done to test the stationarity of the transformed data. If the test statistics from the KPSS test are smaller than the critical values, it doesn't provide significant evidence against the Null hypothesis that the transformed data are stationary and non-seasonal [8,9]. Then a number of first difference of data should be taken until the plot of differenced data and the ACF and PACF plots of the data show that it is stationary. Another KPSS test can be done to support our conclusion from the plots.

After doing the transformation and differencing, an appropriate ARIMA model can be selected by identifying the features shown in the final ACF and PACF figures. In order to select the most accurate model for forecasting, a comparation should be done among the AIC-corrected values of the variations of the selected ARIMA model. The model with the smallest AICc value has the best accuracy in forecasting future volatility.

An appropriate ARIMA model that can be used to make predictions should have residuals consistent with white noise. The residuals should be normally distributed and more than 95% of the autocorrelations should lie within the critical values in the ACF graph. As a complement to the plots, a portmanteau test can be done to check the residuals. If the p-value from the Ljung-Box test is greater than 0.05, it doesn't provide significant evidence against the residuals being white noise [10].

If all procedures above are completed, a forecast of the future volatilities can be done based on the selected model.

3. Result

3.1. Number of Differencing

In the plot of volatilities from August 2018 to August 2019 in Fig.4, there's no significant change in variances over time. No logs need to be taken from the data. The volatilities need no transformation. However, the plot is not rightly horizontal. It has a downward trend. It means that the data is not stationary.

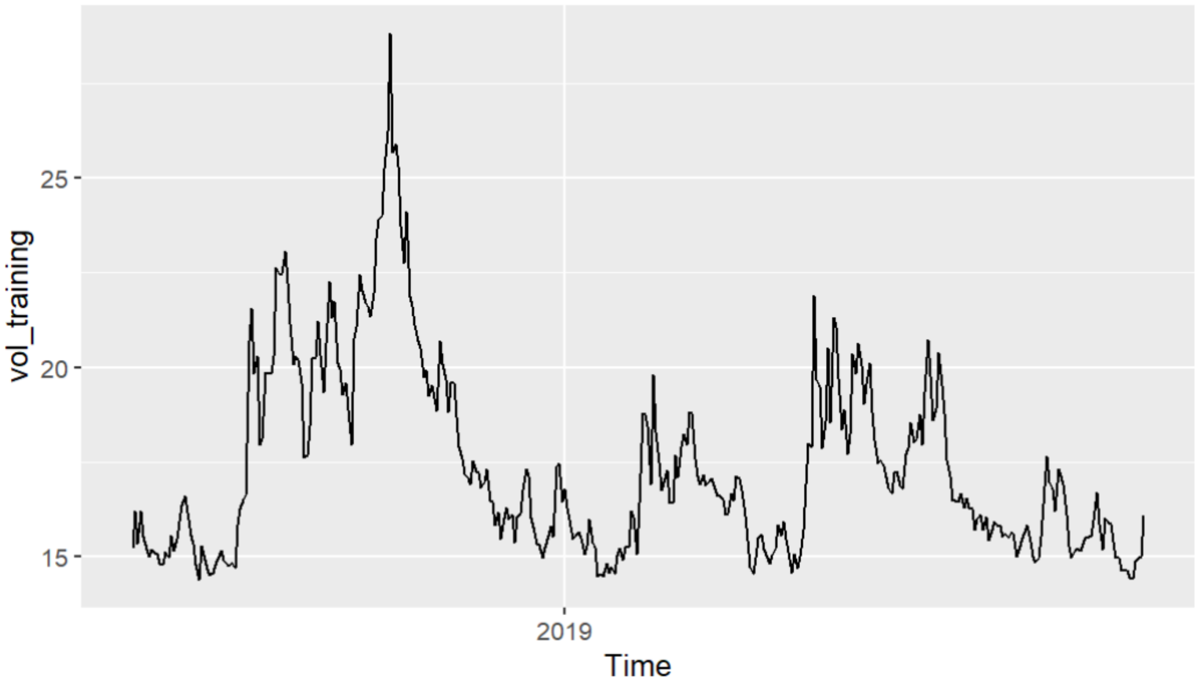

Fig. 4 Volatilities from the August 2018 to the August 2019

In the KPSS test, the p-value 0.8396 is not smaller than the critical values. It provides significant evidence against the Null hypothesis that the volatility data is stationary and non-seasonal (in Table 1).

Table 1. The KPSS test result for original volatility data.

Value of test statistics	0.8396			
Critical value for a significance level of:	10pct 0.347	5pct 0.463	2.5pct 0.574	1pct 0.739

In order to improve the stationarity of the data, a first difference is taken of the data. The KPSS test is carried out again to check if the differenced data is stationary. The p-value 0.0487 is smaller than the critical values. The time series data after differencing is stationary. No further differencing needs to be taken. The ACF and PACF plots of the differenced data can be used to select the best ARIMA model for forecasting (in Table 2).

Table 2. The KPSS test result for volatility data after taking the first difference.

Value of test statistics	0.0487			
Critical value for a significance level of:	10pct 0.347	5pct 0.463	2.5pct 0.574	1pct 0.739

3.2. Model Selection

In Fig.5, there is no obvious peak in the ACF and PACF figure of the differenced data. This suggests that the ARIMA(0,1,0) model can be an appropriate model.

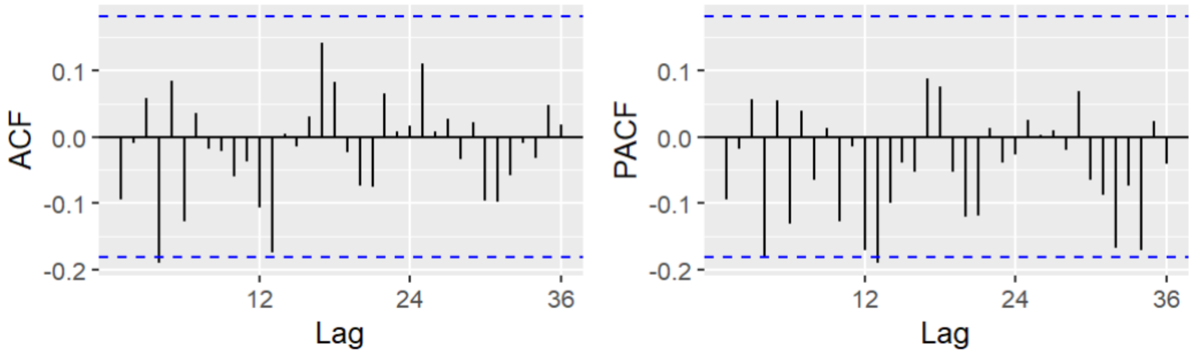

Fig. 5 The ACF and PACF graphs of the volatility data after taking the first difference

In Table 3, the comparison of the AIC corrected values of several variations of ARIMA(0,1,0) shows that the ARIMA(1,1,1) model is the most accurate model. It has the smallest AIC corrected value 904.89.

Table 3. AICc values of various ARIMA models

ARIMA (0,1,0)	ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (1,1,0)	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,0)	ARIMA (2,1,1)	ARIMA (2,1,2)
906.73	905.09	906.74	905.34	904.89	907.85	906.92	907.93	905.41

According to Fig.6, more than 95% of the autocorrelations of the residuals from the selected model lie within the critical values in the ACF graph. The residuals are normally distributed and have a mean zero. The plots above suggest that the residuals are consistent with white noise and can be used to predict future volatility values for the US returns.

Besides, the p-value from the portmanteau test for the residuals is 0.9701. It is greater than 0.5, it doesn't provide us significant evidence against the residuals being white noise.

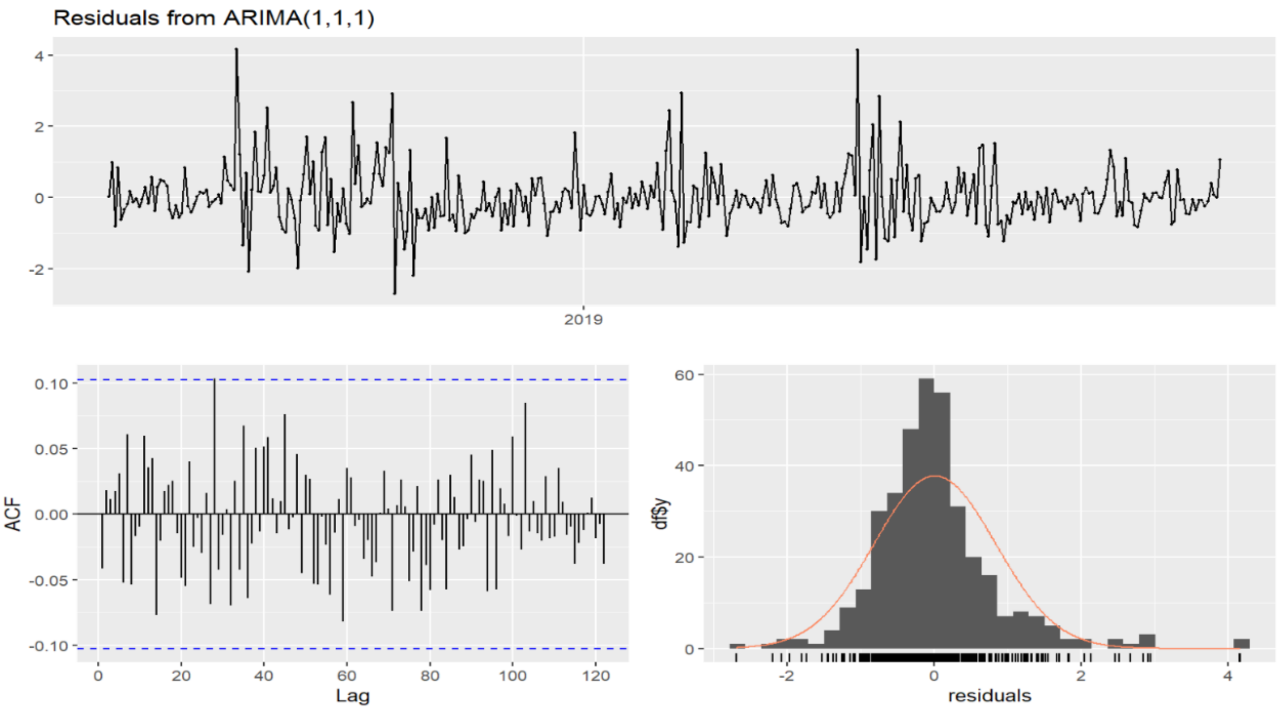

Fig. 6 The time series plot, the ACF plot, and the normal distribution plot of the residuals from ARIMA(1,1,1)

According to Fig.7, the forecast data fits the test data. The test data lie within the prediction interval. It supports that the selected ARIMA model is appropriate for making forecasts.

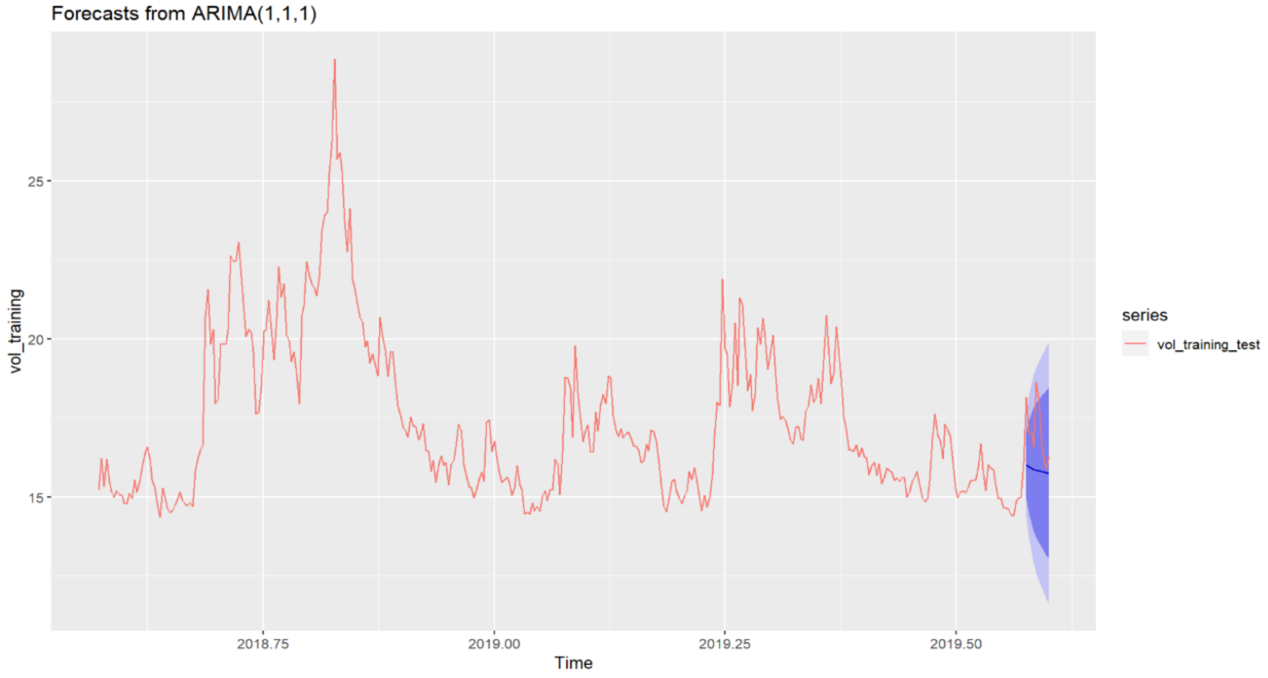

Fig. 7 The combined plot of test data and the predicted data

4. Discussion

Carrying out the procedures above can produce an accurate ARIMA model that fits both the training data and the test data well. This model can make good predictions of future stock return volatilities which can provide valuable insights into risk management and portfolio optimization strategies for investors seeking to navigate the dynamic and uncertain nature of financial markets.

However, the selection of data in this paper avoided the periods with sudden changes in macroeconomic factors such as inflation, rate of interest, and money supply which are relatively closely related to stock return volatilities. Since sudden changes in these factors are not predictable based on past data. For example, the inflation in the 70s rose dramatically in the US because of the Iran-Iraq war, the OPEC Crisis, and the sudden increase in the price of petrol. These factors cannot be noticed by analyzing historical data. ARIMA models can only be used to predict future volatilities when no such events are expected to happen. If economists suggest that any of the three macroeconomic factors are going to change dramatically due to unexpected events, volatility data predicted by the ARIMA model needs to be raised or reduced manually.

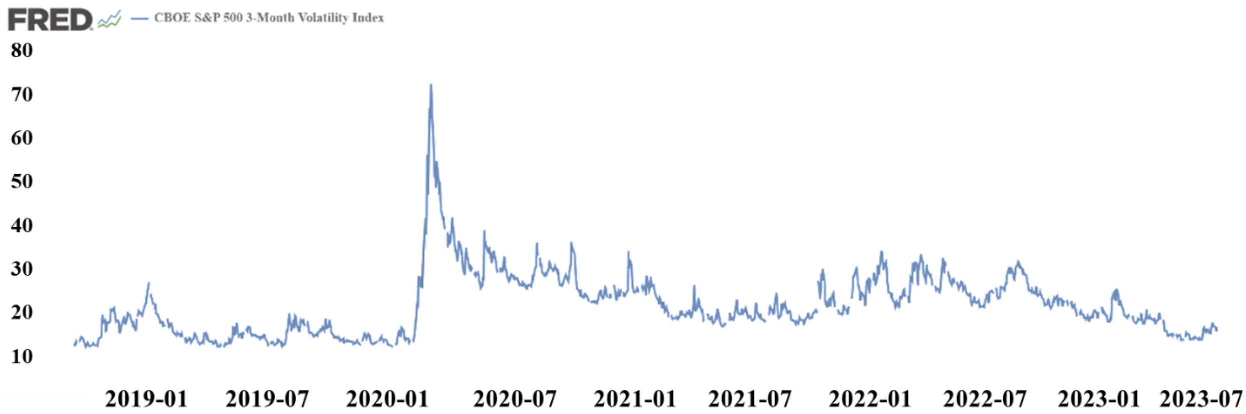

Fig. 8 The daily CBOE S&P 500 3-Month Volatility Index from 2018 to 2023 [4]

According to Fig.1 and Fig.8, the volatility increased dramatically in 2020 due to the increase in inflation. In the predicted data for the year 2020, the dramatic increase was not expected. This supports the idea that the changes in macroeconomic factors need to be noticed and the predicted volatilities by the ARIMA model need to be changed manually according to these factors.

5. Conclusion

In this article, an ARIMA model is selected to model the stock return volatility data and produce an accurate forecast of future volatility. The comparison between test data and forecast data tells that even if the money supply and the interest rate which would influence the volatility value are not stable over time, the ARIMA model could provide an accurate forecast. Confirming the stationarity of the differenced data and the consistency of white noise of residuals from the selected model could ensure that the selected ARIMA model fits the data well and the forecast based on this model is accurate enough. A forecast made by fitting data into a model provides investors insights about how would the volatility change in the future. This could improve confidence in investing for either experienced or inexperienced investors. Due to the increase in investment confidence, further improvement in economic performance might take place in the economy as well.

However, if any of the three macroeconomic factors, inflation, money supply, and interest rate have a sudden change, the forecast data from the model might be quite different from the real data. For example, the increase in inflation caused by war could not be predicted by analyzing the autocorrelations of historical data. Investors are required to make their own judgment about these unpredictable events and alter the predictions from the ARIMA model manually. Recently, there are much more events such as COVID-19 and the discharge of nuclear sewage which could potentially influence macroeconomic factors. It would be crucial for the investors to understand how would these factors affect the volatilities if they want to improve the accuracy of forecast data in these years.

References

[1] Office C B .How Federal Spending for Infrastructure and Other Public Investments Affects the Economy.William S. Hein, 1991.

[2] Guan Wei, He Yali. Empirical Study on the Impact of Macro Variables on Stock Prices and Their Volatility in Various Industries. China Price Journal, 2021 (4): 4

[3] Wang Yange. Research on the Internal Correlation between Macroeconomy and Stock Market Volatility. Wuhan University, 2017

[4] Economic research, URL: https://fred.stlouisfed.org/series/VXVCLS, last assessed 2023/8/25

[5] macrotrends, URL: https://www.macrotrends.net/countries/USA/united-states/inflation-rate-cpi , last accessed 2023/8/22

[6] EconomicGreenfield, URL: https://www.economicgreenfield.com/2021/01/22/money-supply-charts-through-december-2020/#more-26633, last accessed 2023/8/22

[7] Economic research, URL: https://fred.stlouisfed.org/series/REAINTRATREARAT10Y, last accessed 2023/8/22

[8] Shen Hanxi, Lin Jian. Prediction of China's Foreign Trade Import and Export Based on ARIMA Model: 2006-2010. International Trade Issues, 2007

[9] Wang Yurong. Application of ARIMA model in predicting China's export trade. Statistics and Decision, 2004

[10] Forecasting the Increase of Petroleum Consumption in China based on the ARIMA Model. Resources Science, 2007, 29 (005): 69-73