Proceedings of the 2nd International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/88/20240914

Research on AAPL Stock Price Prediction Using ARIMA Model

Guanglei Zhang[1,a,*]

[1]School of AI and Advance Computing, Xi'an Jiaotong-Liverpool University, Ren'ai Rd, Industrial Park 1111, 215123, China
a. Guanglei.Zhang22@stu.xjtlu.edu.cn
**corresponding author*

Abstract: As the cornerstone of technological development, NYSE share prices have attracted international attention. However, due to the frequency of price fluctuations and the sensitivity to the international economic situation, NYSE share prices are difficult to predict. Due to significant differences in prices before and after the pandemic, this will increase the difficulty of prediction and decrease its accuracy. The ARCH model, as a typical prediction model, is highly favored by scholars from various countries due to volatility of the NYSE returns. This article considers the ARCH model as a best method to forecast stock market volatility compared to the GARCH model. Following the conditional variance predictions, the study proposes Markov-switching model. The conclusion is as follows: in the 2021-2024, the prices will have a slight upward trend, while in the next year, the overall trend will show a relatively increasing trend in NYSE. Therefore, this article believes that it is recommended to invest in stocks in NYSE in the short term. At the same time, this article believes that ensuring the US economic stability can improve the stability in stock prices and thus gain an advantageous position in short-term investments.

Keywords: Prediction, NYSE index, ARCH model, GARCH model, Markov-switching model

1. Introduction

Forecasting volatility holds significant importance within financial market research, playing a pivotal role in both risk management and asset pricing. Thus, either academics or practitioner relevant, tried to find a suitable method to forecasting the volatility in financial market for some certain conditions. Works have been done in many aspects, Solnik et al. used cross-sectional dispersion of stock market returns to estimate the trend of globle correlations [1]. Balaban et al. evaluated eleven models trained by samples in stock market monthly volatility in fifteen stock markets [2]. Martens et al. chose time series model accommodate stylized facts of volatility to forecast the out-of-sample volatility, such as long memory, level shifts [3]. Dar et al. utilized a Markov chain approach to forecast stock trends in the Indian stock market, focusing specifically on Tata Consultancy Services Limited (TCS Ltd.) share prices [4]. Chen and Zhang delved into the repercussions of jumps, co-jumps, and their signed elements on the prediction of oil futures price volatility with heterogeneous autoregressive realized volatility model [5]. Muller et al. proposed a new model class, successfully replicates empirical findings from FX intra-day data [6]. Moreover, the increasing attention from policymakers,

DOI: 10.54254/2754-1169/88/20240914

international investors, and financial researchers underscores the growing concern surrounding shifts in international financial markets, particularly in light of advancing financial globalization.

The fundamental purpose of forecasting the share price volatility using many economic variables is to make the predicted price more accurate and reasonable as the share price volatility is subjected to economic behaviour. In the research process of many scholars, the share price has been subjected to several models. Dar et al. employed Markov processes for share price volatility analysis and predictions revealed that share price fluctuations exhibit substantial swings during periods of market turbulence [4]. Statistical prediction models, such as linear regression, time series models, including Autoregressive Integrated Moving Average Model (ARIMA), Seasonal Autoregressive Integrated Moving Average Model (SARIMA), and Autoregressive Conditional Heteroskedasticity Model (ARCH), are distinct from the Markov chain model [7]. While these models rely on linear time series trends, the Markov chain model can accommodate both linear and nonlinear time series data.

Xi et al. used the principal analysis and MIDAS-GARCH model to calibrate the impact of investor sentiment on the wide range of components of volatility of Shanghai composite stocks [8]. At the same time, more and more scholars have found that the Heterogeneous Market Hypothesis, as proposed by Muller et al. posited that the asymmetric behavior of volatility stems from traders' varying time horizons [6]. Specifically, short-term traders are affected by both short-term and long-term volatility, whereas long-term traders are less influenced by short-term volatility. In assessing predictive accuracy in the forecasting model, Liang et al. adopt the approach which employing the out-of-sample R2 measure [9]. This metric quantifies the percentage decrease in mean squared forecast error (MSFE) achieved by the realized volatility (RV) forecast model compared to a benchmark model.

Expanding on the Heterogeneous Autoregressive Model of Realized Volatility (HAR-RV) model, Salisu et al. found an appealing framework for understanding and predicting US stock market volatility [10]. Bauer and Vorkink introduced four separate multivariate HAR models denoted as MHAR-RV-BP, MHAR-RV-BPA, MHAR-RV-X, and MHAR-RV-BPAX [8]. These models improved realized volatility by incorporating various elements, including principal components of bi-power co-variation, asymmetry, predictors derived from stock returns, and other models' prediction. Similarly, Cho and Shin introduced the integrated HAR (IHAR) model, which limited the sum of HAR coefficients to be equal to one [11]. Chen et al. suggested a HAR-RV model that considers jumps, co-jumps, and their dissymmetry within the predictive framework [5]. While Yao et al. modified the HAR model by incorporating a hierarchical clustering technique to create a cluster HAR [12]. The Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) model used by María et al. can avoid modal confusion while excluding residual noise in the IMF.

In conclusion, the study of crucial share price data has garnered significant attention from numerous scholars. This article will primarily imply the ARIMA, GAARCH and Markov models to forecast and analyze share prices, providing pertinent recommendations to portfolio risk managers for making economic policies related to asset allocation, asset pricing and risk management investors based on the forecasted results.

2. Methods

2.1. Data Source

The data is taken from Yahoo.finance from 2021 to 2024. This data is the daily average of each New York Stock Exchange (NYSE) Index calculated in US dollars, with a total of 797 observations from January 2021 to February 2024.

DOI: 10.54254/2754-1169/88/20240914

2.2. Variable Selection

The New York Stock Exchange (NYSE) Index reflects fluctuations in share prices, influenced by both the broader trajectory of US economic performance and significant global events. Given the irregular and unpredictable nature of major events both domestically and internationally, price fluctuations can occur frequently and prove challenging to quantify in terms of magnitude, as depicted in Figure 1.

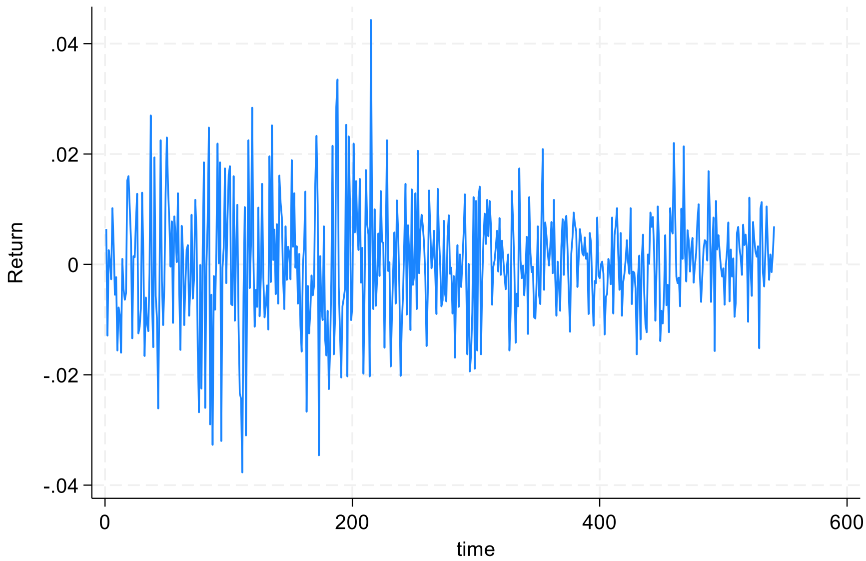

Figure 1: NYSE Index Volatility.

From Figure 1, it can be concluded that before 200th observation, NYSE share prices had entered stable stage without showing a significant increase and decrease. Its highest value is almost 2 times the lowest value, and the period is only 2021-2024 period. After 2022, there was a minimum significant increase or decrease comparative to before 2022, and during these periods, share prices showed a stability.

2.3. Model Selection

To begin with, the study aims to prevent potential bias in the findings due to the probable presence of unit roots in the behavior of the NYSE index under examination. Consequently, the study intends to assess the data stationarity using the Augmented Dickey Fuller (ADF) test and the Phillip Perron Test (PP). The ADF methodology adjusts for higher-order correlation by incorporating lagged differences of the dependent variable into the regression model.

Then, this article selects several models for volatility prediction. First, the Autoregressive Conditional Heteroskedasticity (ARCH) model is used to analyze time series data exhibiting volatility clustering, where periods of high volatility tend to cluster together. The basic idea behind the ARCH model is to capture the conditional variance of a return of NYSE index as a function of its past squared observations. In other words, the variance of the NYSE return series at time t depends on the squared errors or residuals from previous time points. This allows the model to capture the volatility clustering phenomenon often observed in return data. The ARCH model is typically expressed as:

$$\sigma^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-1}^2 + \cdots + \alpha_p\varepsilon_{t-p}^2 \tag{1}$$

DOI: 10.54254/2754-1169/88/20240914

In addition, the study used Markov-switching models to predict the volatility of NYSE index return. This model is also recognized as regime-switching models, which is used to analyze time series data that exhibit regime changes or shifts in behaviour over time. These models are particularly useful for capturing non-linearities of NYSE return behaviour and check the changes in underlying dynamics that cannot be adequately captured by ARCH or GARCH models.

3. Results and Discussion

This section addresses three model developments for volatility forecasting. Initially, the study did a test stationary test to mitigate potential bias in the results caused by the probable presence of unit roots in the examined variables, the study assesses the stationarity of the data using the Augmented Dickey Fuller test (ADF) and the Phillip Perron Test (PP) for return of the NYSE index. Thereafter ARCH, GARCH and Markov-Switching Model is fitted for the volatility forecasting.

3.1. Stationary Test

Due to the increasing variability over time and non-stationarity in original series of the NYSE index price, log transformations for NYSE index return were considered. This adjustment aimed to decrease the heteroskedasticity of the sample. The log transformation was conducted for the NYSE index changes, as presented in Table 1. Subsequently, the unit root test was performed on the log-transformed data at the original level, as indicated in Table 1.

Table 1: Results of Unit Root Test for First Difference of Log Return of NYSE Index.

Variable	ADF	P Value	Philiphs Perron	P Value
LNASPI	-13.628	0.0000	-14.06217	0.0000

As shown in Table 1, the null hypothesis can be rejected at a 5% significance level for all variables. Consequently, it can be inferred that the series exhibit stationarity at their original level, as determined by both the ADF and Phillips Perron tests. Given that the returns of the NYSE index are integrated at the same order across all series, an examination of long-run equilibrium relationships between these series is pursued. Hence, the ARCH, GARCH, and Markov-Switching Model analyses are conducted using the original series' logarithmic returns.

3.2. ARCH Model Results

As the daily behaviour in price of NYSE index is volatile the ARCH model is used for forecasting. Findings of Table 2 explains coefficients (α) associated with the lagged squared residuals which the impact of previous volatility shock on current volatility in the ARCH model.

In the ARCH model (Table 2), significant positive coefficients in L1, L2 and L3 suggests persistence in daily volatility in return of NYSE index. This means that past shocks of NYSE return continue to influence volatility in the present.

According to findings in Table 1, larger coefficient in L3 indicate a stronger impact of Lag 3 shocks on current volatility, while smaller coefficient in Lags 1 and 2 suggest a weaker impact, however, statistically significant at 5% level. This implies that volatility shocks are not random but rather tend to occur in clusters over time.

Table 2: ARCH Model Findings.

Return	Coefficient	P-value
Constant	0.0004	0.174
ARCH	-	-
Arch L1	0.167	0.000
Arch L2	0.168	0.009
Arch L3	0.235	0.001
Constant	0.001	0.000
Number of Observation	541	-
Log Likelihood	1724.2	-

3.3. GARCH Model Results

Next, the study uses GARCH model to forecast the volatility of the index return in the USA. Table 3 represents the GARCH model parameters which suggest the viability of the model for prediction of volatility.

Table 3: GARCH Model Findings.

Return	Coefficient	P-value
Constant	0.0004	0.351
ARCH	-	-
Arch L1	0.065	0.001
Garch L1	0.925	0.000
Constant	0.001	0.245
Number of Observation	541	-
Log Likelihood	1735.75	-

According to the findings in Table 3, the constant term is insignificant which represents that the long-term average variance in the time series is insufficient to predict the behaviour of NYSE index return. This suggests lower average volatility levels in 2021-2024 period. ARCH Coefficient at lag 1 (α) in Table 3 capture the impact of past squared errors on current volatility which is a stronger impact of past volatility shocks on current volatility at 5% significance level. A significant ARCH coefficient of 0.065 (P=0.001) suggests the presence of volatility clustering, during 2021-2024 period of high volatility are followed by periods of high volatility in Lag1. GARCH Coefficients at Lag 1 (β) in Table 3 capture the impact of past conditional variances on current volatility. Larger GARCH coefficients indicate a higher persistence of volatility shocks over time. A significant GARCH coefficient at Lag 1 suggests that past volatility continues to influence current volatility at 1% significant level. As all the parameters are not significant in the volatility clustering in GARCH model, that volatility shocks are not random but tend to occur weakly in clusters over time.

Forecasting Volatility Predictions using ARCH and GARCH Model represent in Figure 2 and 3. Volatility prediction is carried out to investigate the difference between actual and predicted behaviour of NYSE return.

DOI: 10.54254/2754-1169/88/20240914

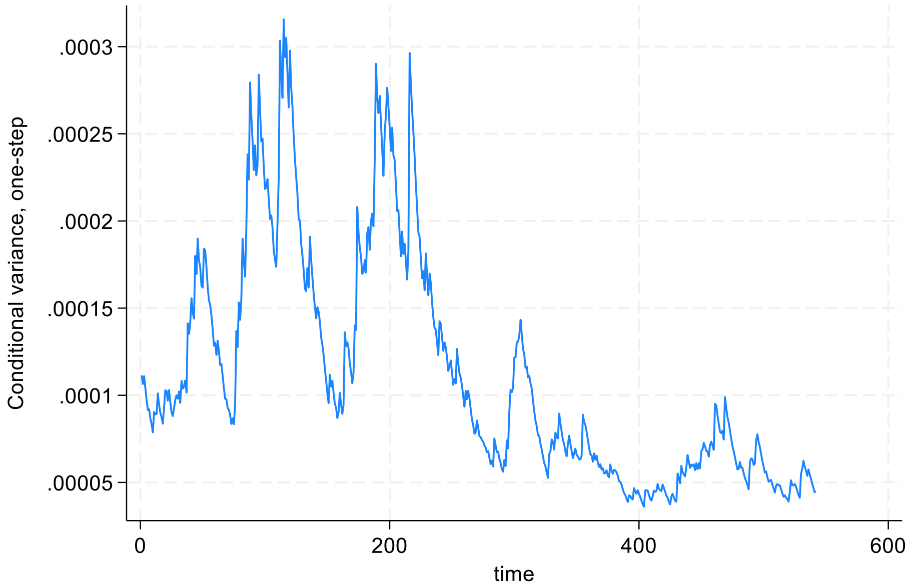

Figure 2: GARCH Variances Forecasting.

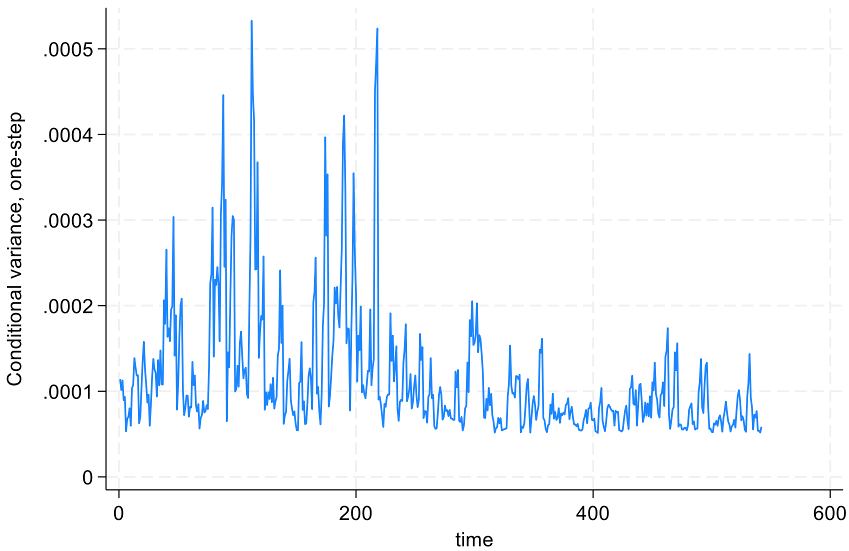

Figure 3: ARCH Variances Forecasting.

Compared to GARCH Variances Forecasting in Figure 2, ARCH variances are minimum after 200th observation in Figure 2. The minimum variance in the ARCH model suggests that the variances between prediction and actual in ARCH model is minimum and recommended for the suitability of volatility prediction in NYSE index return. ARCH model and GARCH model represent two distinct behaviours as seen in Figure 2 and 3. For instance, time after 200th observation, the conditional volatility becomes lower compared to before. As there is a switch in the variance, Markov-switching model is suggested.

DOI: 10.54254/2754-1169/88/20240914

3.4. Markov-Switching Model Results

In a Markov-switching model, different states or regimes are assumed to be switched by the underlying process, each characterized by distinct statistical properties. The transitions between states are governed by a Markov process, meaning that the probability of transitioning to a particular state depends only on the current state and not on the past history of the process.

Table 4: Findings of Markov-switching model.

Return	Coefficient	P-value
State 1	-	-
Constant	15124.12	0.000
State 2	-	-
Constant	16424.11	0.000
Sigma	524.829	0.000
P11	0.988	
P12	0.113	
Number of Observation	541	
AIC	15.459	
HQIC	15.474	
SBIC	15.499	
Log Likelihood	4176.73	
Number of states	2	
Unconditional Probabilities	transition	

Findings in Table 4 represents the Markov-switching model which involves the estimated parameters and their implications for the regime dynamics in state 1 and state 2 in the NYSE index return time series. Regime Transition Probabilities between different states or regimes are significance at 1% level. The findings represent higher transition probabilities which indicate a greater likelihood of transitioning from one regime to another. Understanding these probabilities supports in identifying the frequency and speed of regime switches in NYSE index return. Once the model is estimated in Table 4, the identified regimes can be graphically represented below in Figure 4 with the actual and predicted behaviour.

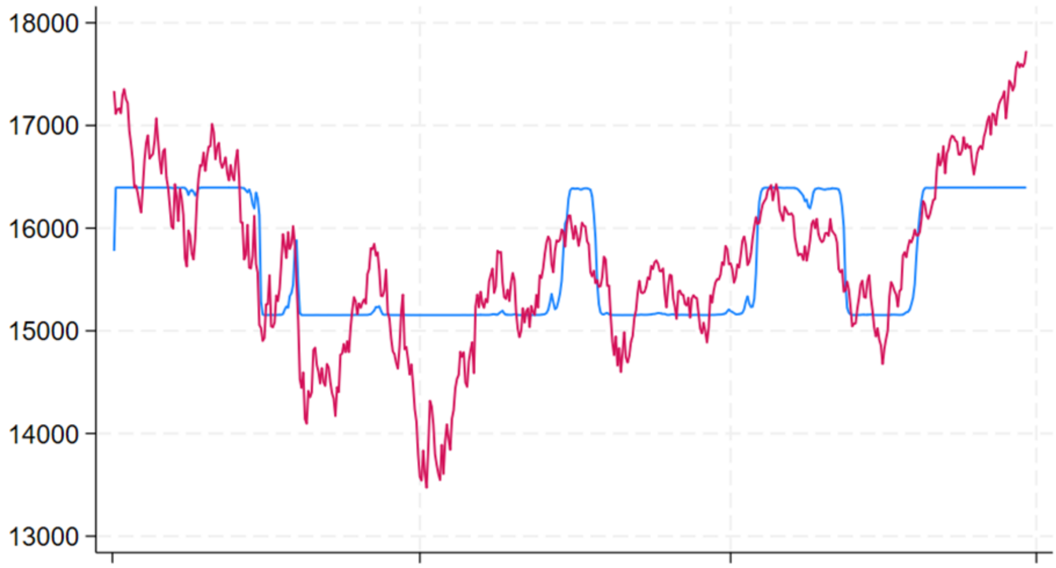

Figure 4: Actual Vs Predicted in Markov Switching Model.

The blue line in Figure 4 shows the model's anticipated outcomes and the red line shows the behaviour of actual data. From the findings above, it is evident that the predicted share price experienced a slight increase and behave constant in the first quarter of 2024, followed by a downward trend, as suggested by the Markov Switching Model. This suggests that the prediction of NYSE volatility behave relatively same as the actual. According to the model, NYSE return is expected to be constant with improving trend in future. Further studies are required to understand the characteristics of low regimes and high regimes in the above graph which helps in assessing the US economic or market conditions for accurate predictions in future.

4. Conclusion

Through the research in this article, it can be concluded that although the NYSE share price index may slightly increase in the short term, the overall trend of change in the next year will be unpredictable as the characteristics of low and high regimes are undiscovered. However, due to the uncertainty of the US economic situation, predictions made further away will become more inaccurate. Meanwhile, in recent years, the US has been affected by the pandemic, and at the same time technological developments have been impacted and are in a recovery period, which will make NYSE stock index return more stabilize.

Through the predictions and research in this article, some suggestions can be provided. From the behaviour of past performance of NYSE index, it is advisable to check the macroeconomic performance to predict share volatility for share investors to make relevant investments. From a exogenous shock perspective, the model should incorporate the political and global shocks to minimize the variances in the model. This can not only maintain a stability in predicting trend of the NYSE index, protect investors from exogeneous shocks, and maintain investors' confidence. This not only benefits the investing in share market but also puts the investors in an advantageous position in shar market trading.

References

[1] Solnik, B. and Roulet, J. (2000) Dispersion as cross-sectional correlation. Financial Analysts Journal, 56(1), 54-61.

[2] Balaban, E., Bayar, A. and Faff, R. W. (2006) Forecasting stock market volatility: Further international evidence. The European Journal of Finance, 12(2), 171-188.

[3] Martens, M., Van Dijk, D. and De Pooter, M. (2009) Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. International Journal of forecasting, 25(2), 282-303.

[4] Dar, G. F., Padi, T. R. and Rekha, S. (2022) Stock price prediction using a Markov chain Model: a study for TCS Share Values. Advances and Applications in Statistics, 80, 83-101.

[5] Chen, Y., Ma, F. and Zhang, Y. (2019) Good, bad cojumps and volatility forecasting: New evidence from crude oil and the U.S. stock markets. Energy Economics, 81, 52-62.

[6] Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and Von Weizsäcker, J. E. (1997) Volatilities of different time resolutions—Analyzing the dynamics of market components. Journal of Empirical Finance, 4, 213–239.

[7] Kadhem, S. and Thajel, H. (2023) Modelling of crude oil price data using hidden Markov model. Journal of Risk Finance, 24(2): 269-284.

[8] Bauer, G. H. and Vorkink, K. (2011) Forecasting multivariate realized stock market volatility. Journal of Econometrics, 160(1), 93–101.

[9] Salisu, A. A., Gupta, R. and Ogbonna, A. E. (2022) A moving average heterogeneous autoregressive model for forecasting the realized volatility of the US stock market: Evidence from over a century of data. International Journal of Finance & Economics, 27(1), 384-400.

[10] Cho, S. and Shin, D. W. (2016) An integrated heteroscedastic autoregressive model for forecasting realized volatilities. Journal of the Korean Statistical Society, 45(3), 371–380.

DOI: 10.54254/2754-1169/88/20240914

[11] Gong, X. and Lin, B. (2019) Modeling stock market volatility using new HAR-type models. Physica A: Statistical Mechanics and its Applications, 516, 194–211.
[12] Yao, X., Izzeldin, M. and Li, Z. (2019) A novel cluster HAR-type model for forecasting realized volatility. International Journal of Forecasting, 35(4), 1318–1331.