Time Series Analysis and Prediction of Nike Stock Prices Utilizing SARIMA and GARCH Models

Navin Lo
UC Santa Barbara
navin@ucsb.edu

Abstract

The goal of this project is to forecast and analyze the time series of Nike stock prices using SARIMA (Seasonal Autoregressive Integrated Moving Average) and GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models. In order to accomplish this goal, stock price data from Nike had to be collected and pre-processed, with data going back a year, to June 12, 2023. The closing price of each stock on a given day was used for the time series modeling and predictions. Following this, the SARIMA and GARCH models were then applied to the data. In order to determine which model performed the best, each model underwent a series of parameter estimation and model diagnostic techniques in order to find the best fit. The findings in this project find that the GARCH model was the best-fitting in the prediction of stock prices in the future, with a prediction of a small decrease over the next year. The GARCH model in particular pairs very well with other models such as SARIMA, because it has the ability to forecast the volatility of stock prices, while the SARIMA is better at just predicting short-term future events due to its constant variance assumption which does not account well for later predictions.

The results of this project will be able to give important insight for a real-world context, as it would provide investors with a detailed idea of analyzing the nuances of stock prices and help them find trends that may stem from the data.

1. Introduction

Stock price prediction gives an understanding of the future trends of the stock market and can help investors make informed decisions about buying, holding or selling stocks. This project aims to apply time series analysis to forecast the closing stock prices of Nike, Inc. Nike is a leading multinational corporation specializing in athletic footwear, apparel, and equipment, currently marketing its products under various brands such as Nike, Converse, or Jordan (Carlson 2024). Nike is also a significant player in the stock market, with a ranking of 79th on the S&P 500 Index (Slickcharts, n.d.).

The reason why I chose to analyze and forecast the closing stock prices of Nike is because of the significance the company has had in the sports world over time. Nike is a leading brand in athletic footwear and apparel injury, renowned for its innovative products and strong market presence. I also think its prominence in the market could indicate a higher likelihood for investors to be looking at a stock to invest in, so trying to identify trends and potential growth areas through time series analysis could be valuable. Nike has been a popular sponsor of many athletes and athletic organizations, so I wanted to investigate some of the stock market trends surrounding the company and see its performance in a time series format. This analysis will help me evaluate their investment potential through forecasted models.

In previous studies, researchers have utilized similar models (SARIMA and GARCH) for investigating stock closing prices, and have found that a combination of these models have been a sufficient predictor of stock price over time, and have been able to be applied to investing. In fact, Jie Gao from the Institute of Economics at Shanghai University found that a blend of SARIMA and GARCH, an ARIMA-GARCH model, was the best, and could accurately fit the fluctuation and volatility of stock market trends (Gao, 2022).

This project aims to explore various time series models using Seasonal Autoregressive Integrated Moving Average (SARIMA) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) to identify the best model for forecasting Nike closing stock prices. The models revealed a great insight into understanding the volatility and trends of Nike stock prices, as well as being able to predict future values for investment purposes. Overall, this project paper is a timely and relevant study that contributes to the growing body of research of stock price forecasting and time series analysis, and has practical implications for investors and analysts who are interested in Nike stock prices, or stock prices in general.

2. Data

The dataset used for this time series project is the daily prices of Nike, Inc. (NKE) stock from June 12, 2023 to June 11, 2024. The data comes in a data frame with seven columns: Date, Open, High, Low, Close, Adj. Close, and Volume. This project focuses on the daily closing price of Nike Inc. The frequency of the data is daily, with 252 observations. Though there are 365 days in most years, there are only 252 observations in this dataset, as the stock market is not always open. I chose closing price for this project as it best most closely resembles the performance of the stock on a given day. This dataset is important because of the fact that it allows anyone to clearly visualize the changes and trends in stock prices over time, as there is a closing price value for each day in the dataset.

The reason I chose to use Nike's stock prices is the same reasoning I mentioned in the introduction. Nike has been one of the biggest players in the sports world for decades, and I wanted to see how this affected the company's market value. Additionally, Nike has been the leader of the pack in terms of the broader sportswear and apparel industry, so studying it can help identify trends in reflecting customer trends and changes in stock price across the entire landscape of the industry.

The data was collected by myself from Yahoo Finance, where I downloaded the stock prices from the "Historical Data" tab on their data for Nike. The data is publicly available, so anyone on the internet could access it. This allows for transparent and repeatable analysis.

Date	Open	High	Low	Close	Adj.Close	Volume
2023-06-12	106.3	106.8	104.6	106.8	105.3	7964600
2023-06-13	107.1	107.5	106.3	106.8	105.3	8460500
2023-06-14	107.5	113.4	107.5	112.9	111.3	17415500
2023-06-15	111.8	112.8	110.2	112.4	110.8	10374200
2023-06-16	114	114.8	113.1	113.6	112	14443700
2023-06-20	111.4	112.8	109	109.5	108	10602400
2023-06-21	109.1	110.4	108.7	110	108.4	8377800
2023-06-22	110.8	111.2	109.8	110.5	108.9	6002900
2023-06-23	107.8	110.3	107.3	109.5	108	14324200
2023-06-26	109.7	112.2	109.7	111.7	110.2	6887100

Figure 1. *Nike, Inc. Stock Price Dataset (See Appendix A.1)*
Link: https://finance.yahoo.com/quote/NKE/history/

3. Methodology

This project uses two time series analysis methods to analyze and forecast Nike stock prices: SARIMA and GARCH models. Each model is often used with financial time series datasets as they are capable enough to explain the intricacies of financial datasets. Furthermore, all time series analysis done with these models were coded in R. The data was downloaded from Yahoo Finance as a .csv file, which was seen above in Figure 1. Before looking into the different methods, the data was manipulated so that the "Closing Price" column was used and any missing values or errors were removed. In addition, the dataset was transformed so that it more closely resembled a stationary dataset. This was done through log transformation and differencing.

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is used to evaluate and predict data that has seasonal patterns, and though this dataset does not have seasonal components, we will utilize its functionalities in R, but with `` `seasonal=False` `` enabled to default to an ARIMA (Autoregressive Integrated Moving Average) model. The model consists of an autoregressive (AR) part, which is based on the idea of a current value depending on its past values, an integrated (I) component, which is the order of difference to make the data stationary, and a moving average (MA) component which represents the dependency between an observation and its past residuals.

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model allows the ability to focus more on the volatility in the dataset. These models are useful for capturing heteroskedasticity that is often present in financial data. So, the GARCH model can help in evaluating and forecasting the level of risk or uncertainty associated with future stock prices. GARCH models also contain the AR and MA from ARIMA or SARIMA models with the degree of volatility.

For each model approach, model estimation and diagnostic techniques are used to analyze the dynamics of Nike Inc.'s stock prices. These techniques include ACF and PACF plots, the Ljung-Box Test, AIC (Akaike information criterion), standardized residual plots, and others. The information collected from these diagnostics are utilized to select the optimal parameters of each model, which are integral in accurate forecasting of future events. Finally, the models that perform the best are put to use to predict Nike Inc.'s stock prices for the next year.

4. Results

4.1 SARIMA Model

After loading in the data and getting an idea of what the data frame looks like, the time series of data for Nike, Inc.'s closing stock prices were plotted to get a visual understanding of the data before applying any time series models. As shown in Figure 2, there seems to be a decent amount of fluctuation of Nike's stock prices over time, with steady climbs and descents throughout the past year.

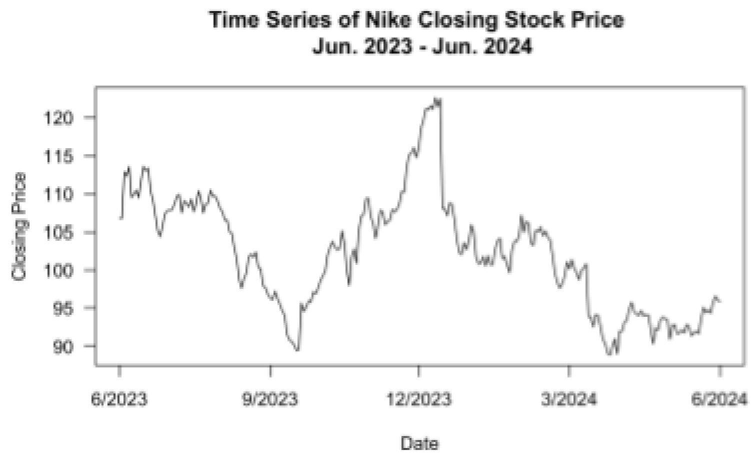

Figure 3. ACF and PACF of Nike Closing Stock Price (See Appendix A.3)

Looking at the ACF (autocorrelation function) and PACF (partial autocorrelation function) plots of this dataset, we can see that there is a significant amount of correlation between current observations and their lagged values (Figure 3). From these plots it is clear that the data needs to be detrended, and this can be done using log transformation and the first difference so that the data can more closely resemble a stationary dataset. This is confirmed by the p-value of the Augmented Dickey-Fuller (ADF) Test, which tests stationarity of a time series. In this case, the p-value was 0.4528, which confirms the non-stationarity of this data.

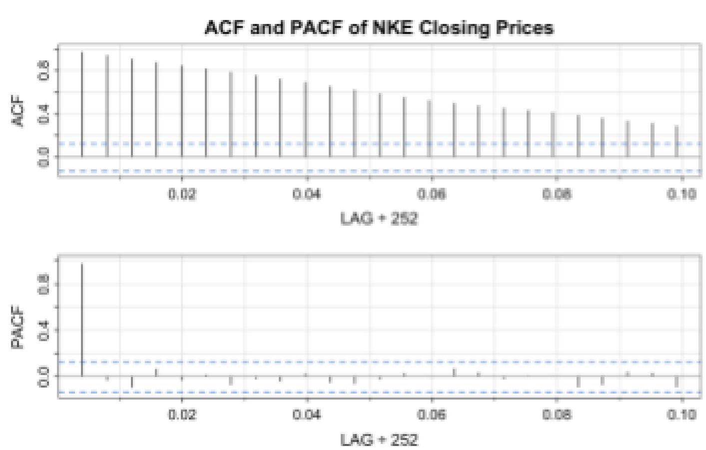

Figure 3. ACF and PACF of Nike Closing Stock Price (See Appendix A.2)

After transforming the data, the stationarity concerns go away for the most part, and there is now no more seasonality or trends in the data. This is confirmed by the time series plot of the transformation (Figure 4), the ACF and PACF plots of the transformed data (Figure 5), as well as the p-value of the ADF test being 0.01, a significant value and determining that the data is stationary.

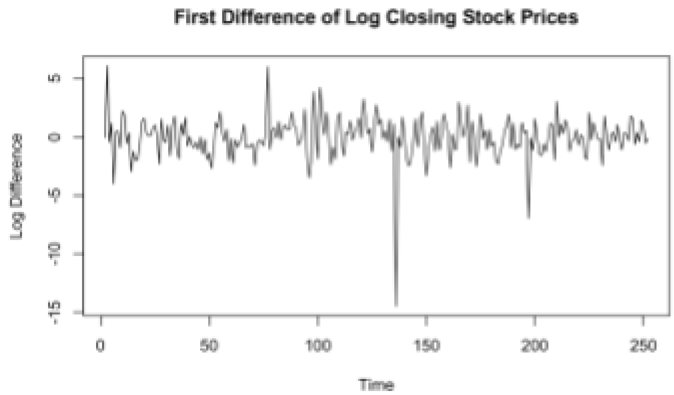

Figure 4. *First Difference of Log Closing Stock Prices Plot (See Appendix A.4)*

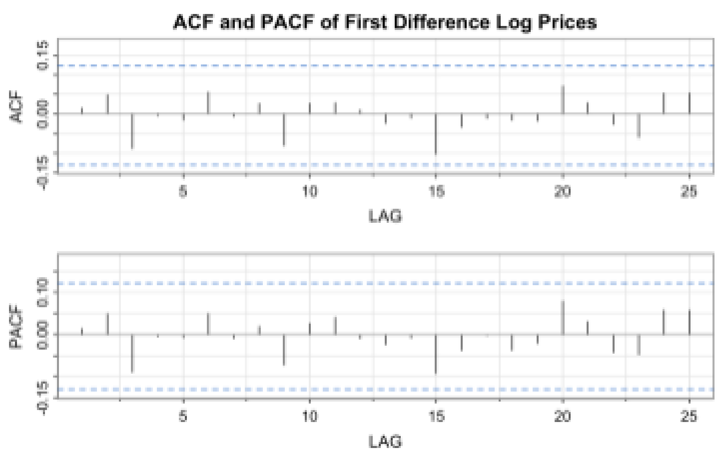

Figure 5. *ACF and PACF of First Difference of Log Closing Stock Prices (See Appendix A.4)*

As mentioned previously, the stock prices do not seem to demonstrate seasonality as the ACF and PACF of the first difference would show some differences in lag in patterns. Therefore instead of a SARIMA model, an ARIMA model was better to use in this situation. From the ACF and PACF plots it was difficult to discern what the AR and MA terms could be, so in experimentation it was found using `auto.arima()` that an ARIMA(2,1,0) gave the lowest AIC value compared to other models, with an AIC value of 4.007792.

Furthermore, the ACF of the residuals and Q-Q plot of the standardized residuals closely follow a normal distribution. And, the p-values of the Ljung-Box statistics show that almost all of the p-values are greater than 0.05 (See Appendix A.4), signifying that there is no autocorrelation present. This shows that the model is sound and can be used in prediction.

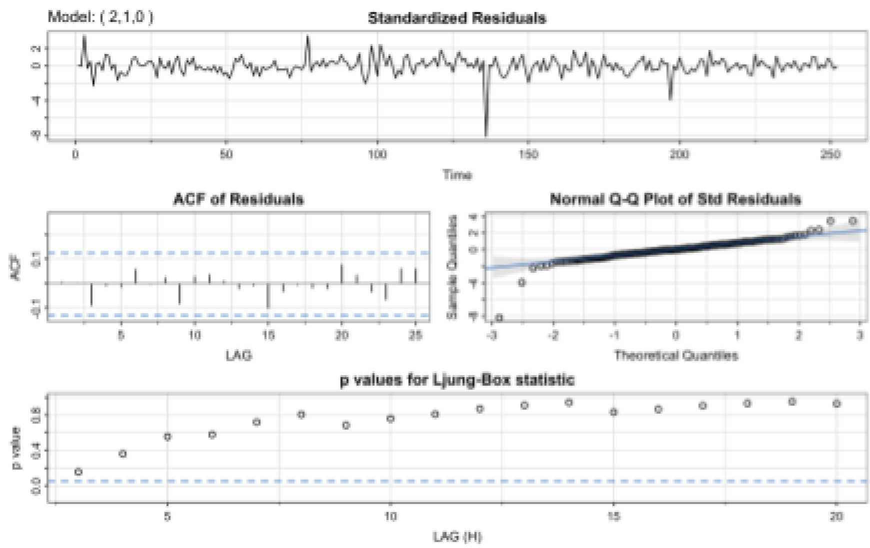

Figure 6. *Standardized Residual Plots (See Appendix A.5)*

Now that the diagnostics of the model have been confirmed to be sound, it can be used to forecast stock prices a year into the future. The forecasted data seems reasonable, and as shown by Figure 7 there seems to be a slight decrease in the Nike stock over time. However ARIMA models tend to be more simple, so it will be necessary to try other models and see if there are better ones that can explain the complexities and nuances of the dataset.

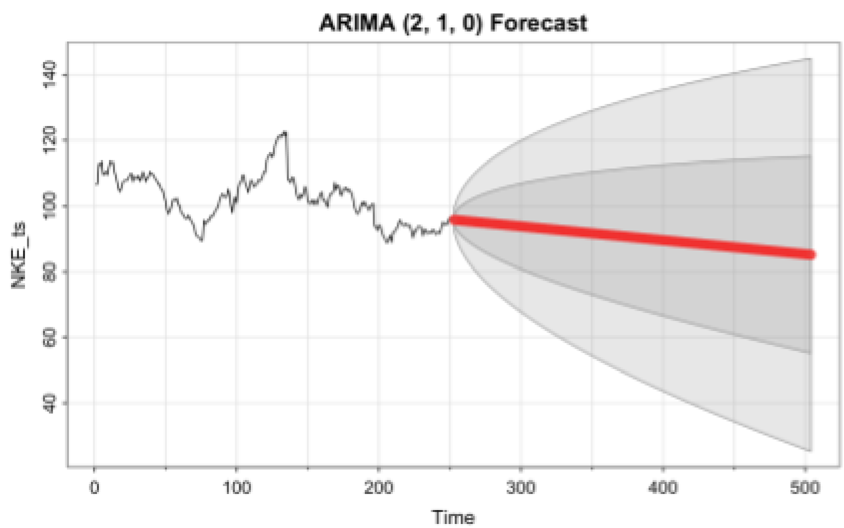

Figure 7. Forecast of ARIMA(2, 1, 0) Model using Nike Inc. Closing Stock Price (See Appendix A.6)

4.2 GARCH Model

In order to fit a GARCH model onto the dataset, we need to first take a look at the ACF and PACF plots of the stock prices as well as the squared residuals of the ARMA model. Figure 5 displays the ACF and PACF of the stock prices with the transformed data. In this case, the ARMA(2,0) model is being used for the GARCH model (See Appendix A.6). For the parameters of the GARCH models, we take a look at Figure 8 to see any significant lags.

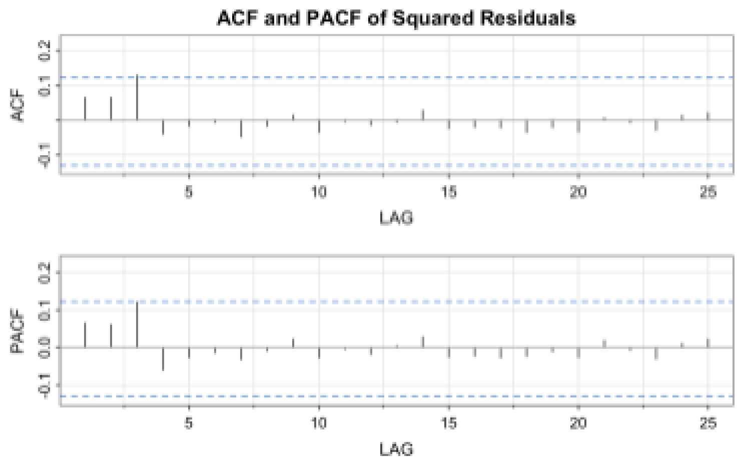

Figure 8. ACF and PACF of Squared Residuals (See Appendix A.7)

There does not seem to be significant lags anywhere for both ACF or PACF, so let's try GARCH models with all combinations of parameters ranging from GARCH(1,0) to GARCH(2,2), and see which one performs best. To see which model is the best fit, taking a look at AIC values and finding the lowest one will be the method of choice in this case. For this data, it is found that ARMA(2,0) + GARCH(1,0) has the lowest AIC value (-5.590), so that will be the model of choice. Here are a few more plots that further show the model is a good fit.

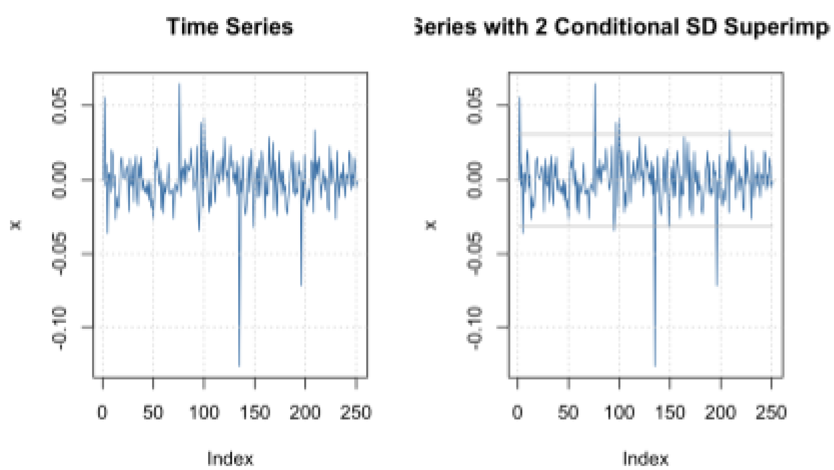

Figure 9. Time Series and 2 Conditional SD Superimposed Plot (See Appendix A.8)

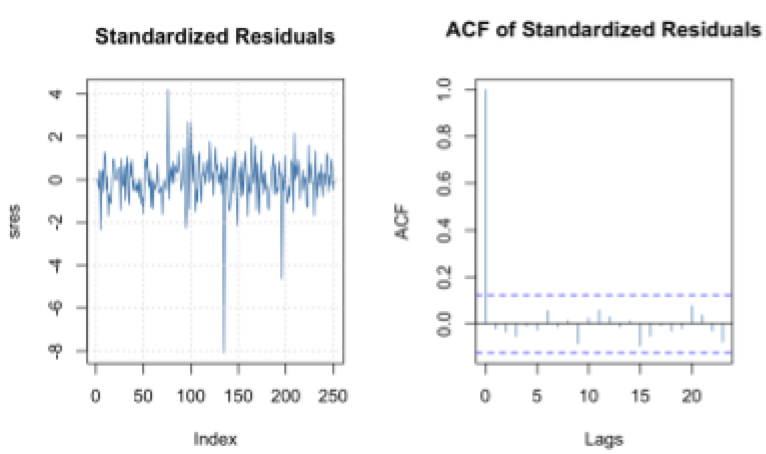

Figure 10. Standardized Residuals and ACF of Standardized Residuals Plot (See Appendix A.8)

Figure 11. *Q-Q Plot of Standardized Residuals (See Appendix A.8)*

Figure 9 displays the time series with two conditional standard deviation estimates, and when compared to the actual time series, the spikes match up. In other words, the estimated conditional standard deviation is able to capture the volatility of the original time series data, which shows the model is a good fit. Furthermore, Figure 10 depicts the standardized residuals and its ACF, which visually indicate that the residuals are somewhat stationary. Lastly, the Q-Q plot of the standardized residuals in Figure 11 displays a normal distribution. All in all, these plots and several model estimation statistics show that the best model is a Garch(1,0) model with an ARMA(2,0) model, performed on the log-transformed and first-differenced data. Now the ensuing step in the process, just like for ARIMA, is to forecast the model for a year.

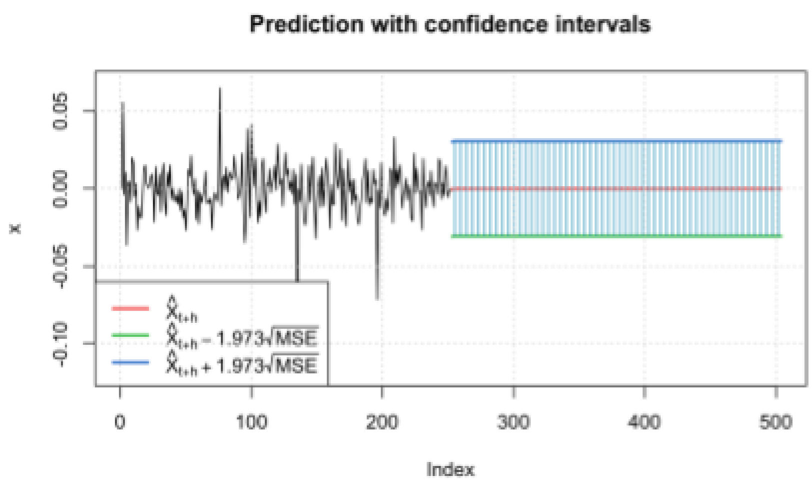

Figure 12. *Forecast of GARCH(1,0) Model (See Appendix A.9)*

This GARCH model is capable of encapsulating the volatility of Nike, Inc. stock prices, which is crucial in forecasting and predicting future values as the previous model showed that the volatility of the dataset was difficult to capture. Based on our findings, it appears that there will be significant volatility in Nike's stock prices in the future, as the potential range of prices is quite large. Therefore, investors might need to exercise caution when examining the company's financial stability in the future. By incorporating this information into ARIMA models, it is possible to make predictions regarding future stock prices as GARCH will take care of the volatility aspect.

5. Conclusion

Given the complexity of financial time series data, it is unrealistic to assume that one model alone can provide the best predictions. Instead, utilizing a combination of different modeling techniques, as demonstrated in this project, can offer valuable insights by exploring various forecasting possibilities. Each modeling approach comes with its own strengths and weaknesses, and the effectiveness of a particular method depends on the specific requirements and preferences of the user.

For example, the SARIMA/ARIMA did a good job at getting a grasp of the autocorrelation in the data, which can help with short-term forecasts. An investor interested in quick profits might consider investing in a stock based on an ARIMA model's predictions, which can provide valuable insights into short-term price movements. Using an ARIMA model, the investor might observe that Nike's stock prices are expected to slowly decrease, suggesting a potentially profitable divestment from Nike's stock leading into the near future.

However, if an analyst wanted to forecast Nike's stock prices over the next few years, they would have some trouble relying only on ARIMA models, as they assume constant variance over time, which would be unrealistic because of how volatile stock prices are, with all of the external factors that can contribute. In this case, using something like a GARCH model in conjunction with an ARIMA model might benefit the analyst because of the ability to analyze volatility. By integrating these models, the analyst could possibly make a well-informed decision to help their firm maximize profits and manage financial risks effectively.

While this project was able to provide reasonable insights on the stock prices of Nike closing stocks, there are many further areas of exploration on this topic. For example, being able to study some of the outside variables that might contribute to a stock's price, like how the rest of the economy is doing, and how other companies in its specific industry are doing, would help forecast the stock prices better.

In general, understanding these external factors will be able to greatly increase the accuracy and preciseness of this time series project, and will be able to provide investors and companies with exact information about how Nike's stock price will change over time, and those investors will be prepared for any increase or decrease in the price.

References

Carlson, Dennie (2022). Nike Inc. Customer Discretionary.Retrieved from https://www.britannica.com/money/Nike-Inc.

Slickcharts (n.d.). S&P 500 - Historical Annual Returns. Retrieved from https://www.slickcharts.com/sp500

Jie Gao *2022). Research on Stock Price Forecast Based on ARIMA-GARCH Model. Retrieved from https://eudl.eu/pdf/10.4108/eai.9-12-2022.2327556

Yahoo Finance (2024). Nike Stock Historical Data. Retrieved from https://finance.yahoo.com/quote/NKE/history/

Appendix

A.1

```{r}
library(astsa)
library(tseries)
library(forecast)
library(dplyr)
library(tidyverse)
library(MASS)
library(fGarch)
library(rugarch)
library(pander)

NKE = read.csv("NKE.csv")
head(NKE, 10)
```

A.2

```{r}
NKE$Date = as.Date(NKE$Date, format="%Y-%m-%d")

closing = NKE$Close
NKE_ts = ts(closing)

x_labels = c("6/2023", "9/2023", "12/2023", "3/2024", "6/2024")
label_positions = seq(1, length(NKE_ts), length.out=length(x_labels))
label_positions = round(label_positions)
par(mar = c(5, 5, 5, 2), las = 1)
plot.ts(NKE_ts, main="Time Series of Nike Closing Stock Price \n Jun. 2023 - Jun. 2024",
    ylab="Closing Price", xlab="Date", xaxt="n")
axis(1, at=label_positions, labels=x_labels)
```

A.3

```{r}
adf.test(NKE_ts)
acf2(NKE_ts, main='ACF and PACF of NKE Closing Prices', 25)
```

```{Output}
        Augmented Dickey-Fuller Test

data: NKE_ts
Dickey-Fuller = -2.2922, Lag order = 6, p-value = 0.4528
alternative hypothesis: stationary
```

A.4

```{r}
NKE_log_diff = diff(log(NKE_ts))
adf.test(NKE_log_diff)
plot.ts(NKE_log_diff, main='First Difference of Log Closing Stock Prices',
     ylab="Log Difference")
acf2(NKE_log_diff, main = 'ACF and PACF of First Difference Log Prices', 25)
```

```{Output}
        Augmented Dickey-Fuller Test

data: NKE_ts
Dickey-Fuller = -5.7661, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

A.5

```{r}
auto.arima(NKE_ts)
NKE_sarima = sarima(NKE_ts, 2, 1, 0)
```

```{Output}
Coefficients:
            Estimate      SE    t.value    p.value
ar1           0.0137  0.0629     0.2175     0.8280
ar2           0.0506  0.0643     0.7871     0.4320
ar2          -0.0425  0.1191    -0.3566     0.7217

sigma^2 estimated as 3.120558 on 248 degrees of freedom

AIC  = 4.00772    AICc = 4.008179   BIC = 4.063975
```

A.6

```{r}
fcast_sarima = sarima.for(NKE_ts, n.ahead=252, p=2, d=1, q=0, plot.all=TRUE,
                  main="ARIMA (2, 1, 0) Forecast")
```

A.7

```{r}
arma = arima(NKE_log_diff, order=c(2,1,0))
residuals = residuals(arma)
sqresid = residuals^2
acf2(sqresid, main="ACF and PACF of Squared Residuals", 25)
```

```

A.8
```{r}
par(mfrow=c(1,2))
plot(log_diff_NKE_garch1, which=c(1, 3, 9, 10))
plot(NKE_garch1, which=13)
```

A.9
```{r}
log_diff_fcastGarch = predict(log_diff_NKE_garch1, n.ahead=252, nx=252, plot=TRUE)
summary(log_diff_fcastGarch)
```
```{Output}
meanForecast         meanError       standardDeviation lowerInterval      upperInterval
 Min.   :-5.106e-04   Min.   :0.01552   Min.   :0.01552   Min.   :-0.03113   Min.   :0.03011
 1st Qu.:-6.787e-05   1st Qu.:0.01558   1st Qu.:0.01552   1st Qu.:-0.03080   1st Qu.:0.03066
 Median :-6.787e-05   Median :0.01558   Median :0.01552   Median :-0.03080   Median :0.03066
 Mean   :-7.025e-05   Mean   :0.01558   Mean   :0.01552   Mean   :-0.03080   Mean   :0.03066
 3rd Qu.:-6.787e-05   3rd Qu.:0.01558   3rd Qu.:0.01552   3rd Qu.:-0.03080   3rd Qu.:0.03066
 Max.   :-6.787e-05   Max.   :0.01558   Max.   :0.01552   Max.   :-0.03080   Max.   :0.03066
```
```