Modeling and Forecasting Nike Inc Stock Price Through Time Series Analysis

Brock Etzel

December 8, 2023

Abstract

Stock prices are one of the most important and often studied models in the world of finance and business. Stocks allow investors to make consistent gains on investments and be set for the future. Nike is one of the biggest companies in the world and studying its stock can help to further model and predict the stock market as a whole. Through a very in depth analysis of the time series that is Nike's stock price I have found that an ARIMA(2,1,2) model of the form $W_t = -0.8282W_{t-2} + e_t - 0.7042e_{t-2}$ can be used to accurately model the logged data. The original data followed a quadratic model so taking the differences and natural logarithms of the data make it stationary. Furthermore the applied model correctly predicts future values and is clearly a better model than other possible iterations.

Introduction

Stocks

Stocks are the most common form of investment in the United States, over 150 million and 60% of U.S. adults own some sort of stock. Stocks are equity that represent a fractional, often very small, portion of ownership over the company to which the stock corresponds to. There are three main types of shares, each with different functions. The first is an authorized share which is any share that has been authorized to be issued to the public and is the maximum amount of shares a company can issue. Next is an issued share which is any shares owned by the company or the shareholders. Finally an outstanding share is anything held by shareholders. Each classification is a subset of the previous and therefore there are less shares in each sequential type. Shareholders, as owners of the company, have the right to help make decisions for the company, this is often accomplished through voting at shareholders meetings where anyone who is a shareholder has the right to vote on the board of directors who control the company. Shareholders also have the right to receive dividends, which is when a company redistributes its profits back to its shareholders rather than reinvesting in itself. Stocks are usually traded, bought or sold, on a stock exchange which is regulated and overseen by the government. The two most common stock exchanges in the United States are the New York Stock Exchange (NYSE) and the Nasdaq.

Stocks are such a common form of investment because of how easy they are to buy and sell, they can easily be traded from phones nowadays. They are also rather low risk when compared to other forms of investment such as real estate, and they have consistently had higher returns than other methods in the long run. Being able to identify a profitable stock can be extremely beneficial in helping one build wealth. There are a number of factors that determine whether a stock is good or not including specific company activity as well as economic factors of the country as a whole.

Nike

Nike is a shoe and athletic apparel company that was founded in 1964 by Phil Knight in Eugene, Oregon. The company originally started out selling running shoes to track athletes. Today they are the top athletic apparel brand in the world and sell products in nearly every sport imaginable. They also partner with the world's top athletes such as Michael Jordan, LeBron James, Cristiano Ronaldo, Serena Williams, and Tiger Woods in order to sell products and connect to their audience.

When a company first offers stocks to be traded by the public it is known as initial public offering or IPO. Nike first went public in December of 1980 at a price of only 23 cents. Today the stock is worth $114 and has reached prices of $177 prior to the COVID-19 pandemic which dramatically decreased the price of most stocks. The company also currently has a market cap or net worth, which is equal to the price of the stock multiplied by the number of outstanding shares of over $175 billion.

Investing and stocks are extremely important for many people in regards to their future and if one is able to model and predict certain stocks they are able to get a leg up on competition.

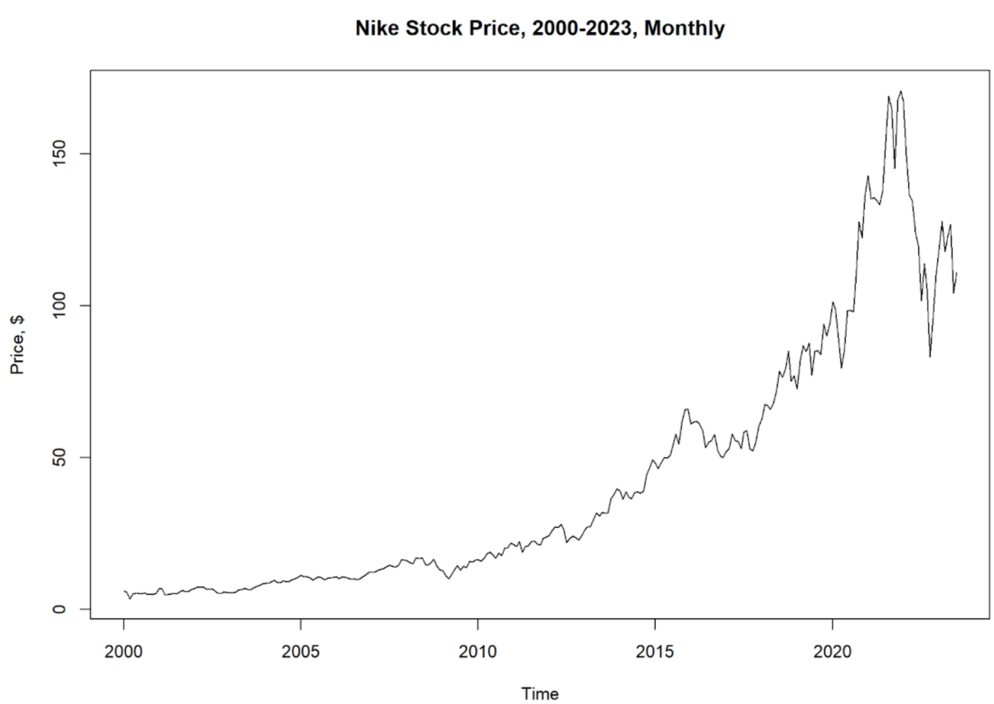
The process of accurately modeling and predicting future values involves much more than just time series but implementing such methods can help predict with some degree of certainty especially when combined with outside knowledge and not simply statistical knowledge. Stocks help many build wealth consistently and save for their future which is very important not just to the individual but for the country as a whole. Having retirees who can afford to live by their own means means the government can focus on other issues. Overall being able to learn as much as possible about the behavior of stocks is both interesting and beneficial to all.

In this paper I will study the changes in Nike's stock price monthly since January 2000. While Nike does not represent all stocks as a whole it is a very good indicator of the market. Nike has a beta value of 1.08 which means that it is 8% more volatile than the market as a whole. The base value for beta is 1 where a stock would move identical with the market. A beta of 1.08 is quite close to that base 1 and thus it is reasonable to assume Nike does represent the market relatively well. Nike is also a part of the Dow Jones Industrial Average, an index which measures 30 large stocks in an attempt to model the market as a whole. Nike's inclusion in the Dow is further proof it most likely follows closely to the true trend of the market.

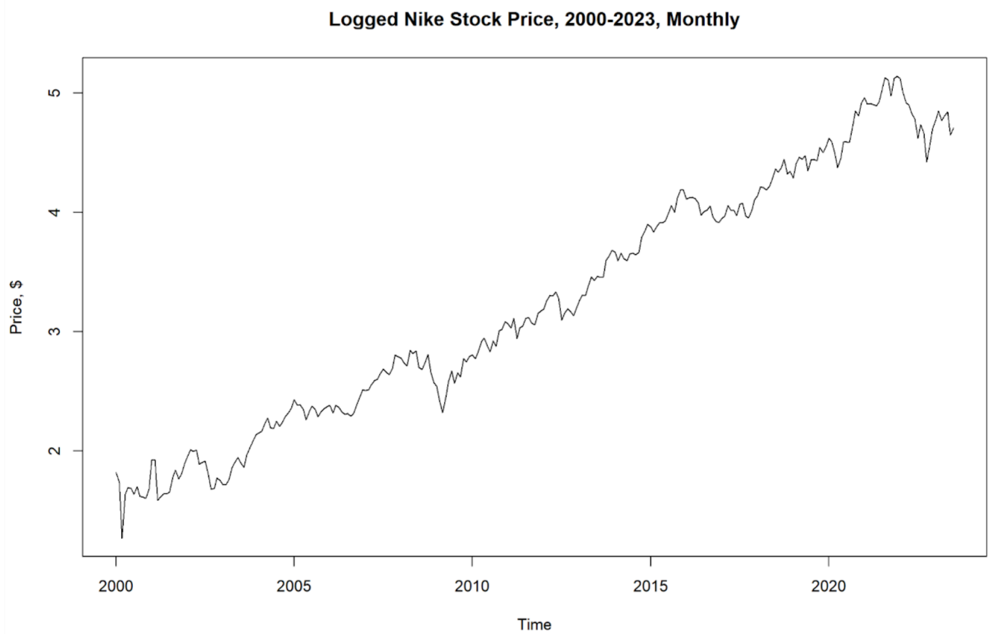
Model Specification

<u>The Data</u>

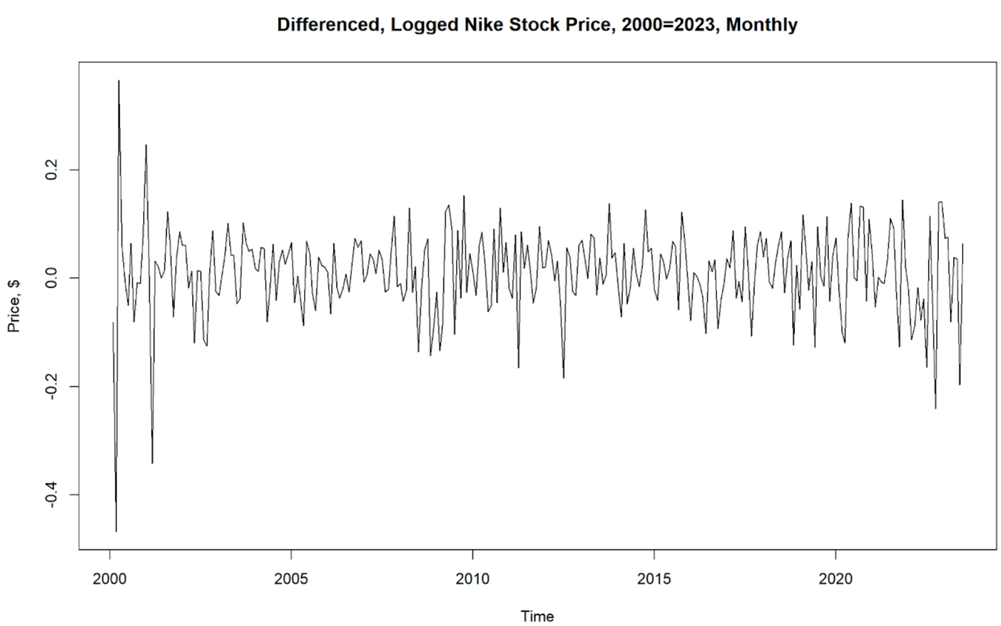
The first goal of working with time series data is always to confirm if the data is stationary or not. If the data is not stationary, transformations of the data should be done in order to make it as such. Stationary data will have both constant variance and mean throughout the entire time series.



The original data set of stock prices very clearly follows a quadratic function. Both the mean and the variance seem to be increasing over time. In order to correct for the issue of the variance the natural logarithm of the data should be taken



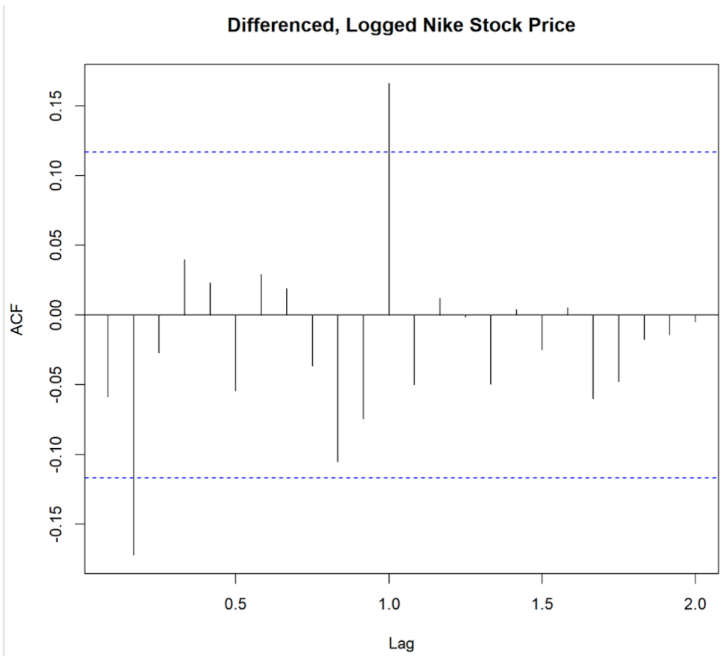
The logged data seems to have constant variance now but clearly still follows a somewhat linear pattern with the mean increasing over time. In order to correct for this the first differences should be taken in order to obtain $\Delta ln(Y_t) = ln(Y_t) - ln(Y_{t-1})$, where Y_t is the stock price at time equal to t.



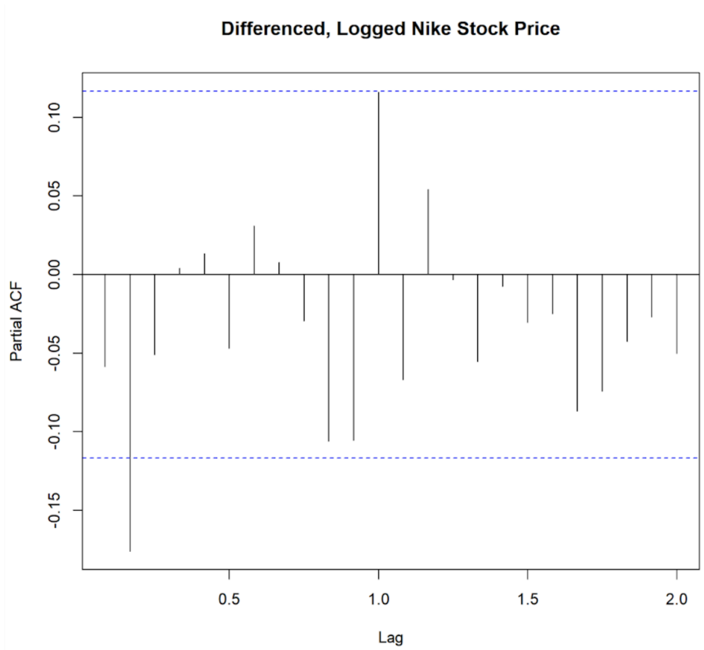
The differenced, logged data seems to be relatively stationary, with possibly a few outliers towards the beginning of the time series. Based on this it should be viable to work with this data when attempting to fit a model to it. Furthermore conducting a Dickey-Fuller Unit Root Test on the data returns a p-value less than 0.01. The Dickey-Fuller Unit Root Test tests the null hypothesis that a time series is nonstationary and the alternate hypothesis that it is stationary. Based on the p-value the null would be rejected and the time series would be concluded to be stationary. Clearly working with this transformed data should provide good results.

<u>Correlation Functions</u>

There are three primary correlation functions that need to be studied when attempting to determine what type of model to fit. Those being the autocorrelation function (ACF), the partial autocorrelation function (PACF), and the extended autocorrelation function (EACF). These three help to identify the correct autoregressive (AR) integrated (I) moving average (MA) or ARIMA model. AR(p) models follow the equation: $Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + e_t$ and MA(q) models follow the equation: $Y_t = e_t - \theta_1 e_{t-1} - ... - \theta_q e_{t-q}$ where e_t is a random white noise process with variance σ_e^2. The ACF is used to determine the order of the MA(q) portion of the model; it should cut off and become zero at any lags beyond q. The PACF is used to determine the order of the AR(p) portion of the model; it should similarly cut off and become zero at any lags beyond p. The I(d) portion of the model has already been determined; it is the number of times the data needs to be differenced in order to be stationary, in this case one.



The ACF shows two significant autocorrelations, at lags of two months and one year. This implies that an MA(2) where the MA(1) term equals zero along with a seasonal MA(1) term would be applicable.



The PACF shows similar results to the ACF however the one year lag autocorrelation is not quite significant. Based on that an AR(2) with the AR(1) term equal to zero seems applicable for the model. Adding in a seasonal AR(1) term may be reasonable but further tests have to be done to determine the viability of that.

EACF

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	o	o	o	o	o	o	o	o	o	x	o	o
1	x	x	o	o	o	o	o	o	o	o	o	x	o	o
2	x	o	o	o	o	o	o	o	o	o	o	o	o	o
3	x	x	o	o	o	o	o	o	o	o	o	o	o	o
4	x	o	o	o	o	o	o	o	o	o	o	o	o	o
5	x	x	x	x	o	o	o	o	o	o	o	o	o	o
6	x	x	o	x	o	o	o	o	o	o	o	o	o	o
7	o	x	o	x	o	o	x	o	o	o	o	o	o	o

The EACF shows which models would work, denoted by o. The upper leftmost model that is viable is thought to be the best according to the EACF. According to the EACF an ARIMA(0,1,0) model would be the best, this would imply the differenced, logged data would follow a random white noise pattern. However the EACF is not the most accurate or consistent measure so further testing should be done. Based on all of the correlation functions there are a number of possible ARIMA models that should be tested out in order to determine which one is the best. Those models include; ARIMA(0,1,2), ARIMA(1,1,2), ARIMA(2,1,2), and ARIMA(2,1,1) as well as the possibility of including further AR(1) and or MA(1) seasonal terms. All of the estimated models will be based on the logged stock prices moving forward.

An ARIMA(p,1,q) model follows the equation:

$W_t = \phi_1 W_{t-1} + \ldots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$ where $W_t = Y_t - Y_{t-1}$

ARIMA(0,1,0) Model

Using the ARIMA(0,1,0) model as a baseline to compare any further, more complicated models provides an AIC of -615.8. The Akaike Information Criterion is a model selection criteria where lower values correlate to better models. The AIC penalizes models with more parameters so simply adding more parameters does not necessarily make the AIC better.

ARIMA(0,1,2) Model

Estimating the ARIMA(0,1,2) model gives the equation:

$$W_t = e_t + 0.047e_{t-1} + 0.1543e_{t-2}$$

However the θ_1 coefficient has a standard error of 0.0587 meaning it is not significantly different from zero. This determination was expected based on the ACF. The AIC of this model is -619.1. When fixing he θ_1 coefficient at zero the AIC drops further to -620.5 and gives the equation:

$$W_t = e_t + 0.1576e_{t-2}$$

ARIMA (1,1,2) Model

Estimating the ARIMA(1,1,2) model gives the equation:

$$W_t =- 0.1229\ W_{t-1} + e_t - 0.0719e_{t-1} + 0.1657e_{t-2}$$

The ϕ_1 and θ_1 coefficients have standard errors of 0.3454 and 0.3388 respectively meaning neither of them are significantly different from zero, both of which were somewhat expected. The AIC of this model is -617.29. Fixing the ϕ_1 and θ_1 coefficients would simply provide the same model as the fixed ARIMA(0,1,2) model so there is no reason to do so.

ARIMA(2,1,2) Model

Estimating the ARIMA(2,1,2) model gives the equation:

$$W_t = 0.0105\ W_{t-1} - 0.8167W_{t-2} + e_t + 0.0370e_{t-1} - 0.6909e_{t-2}$$

The ϕ_1 and θ_1 coefficients have standard errors of 0.1676 and 0.1992 respectively meaning neither of them are significantly different from zero, both of which were expected. The AIC of the model is -617.4. When fixing the ϕ_1 and θ_1 coefficients at zero the AIC drops to -621.1 and gives the equation:

$$W_t = -0.8282W_{t-2} + e_t - 0.7042e_{t-2}$$

ARIMA(2,1,1) Model

Estimating the ARIMA(2,1,1) model gives the equation:

$$W_t = 0.0118\ W_{t-1} - 0.1743W_{t-2} + e_t + 0.0406e_{t-1}$$

The ϕ_1 and θ_1 coefficients have standard errors of 0.298 and 0.2992 respectively meaning neither of them are significantly different from zero, both of which were expected. The

AIC of this model is -617.9. Fixing the ϕ_1 and θ_1 coefficients at zero drops the AIC to -621.2 and gives the equation which is equal to an ARIMA(2,1,0):

$$W_t = -0.1743W_{t-2} + e_t$$

Model Decision

Overall I have decided to work with the fixed ARIMA(2,1,2) model, while the AIC for the fixed ARIMA(2,1,0) is lower it is only slightly different and I feel more confident with the ARIMA(2,1,2) because it was marked as a viable model in the EACF while the ARIMA(2,1,0) was not.

Diagnostics

The ACF of the residuals from the fitted model show no significant autocorrelations except at a lag of one year. This may imply refitting the model with a seasonal term would make sense

Refitting with Seasonal Term

If a seasonal MA term is added the AIC drops to -631.2 and gives the equation:

$$W_t = -0.4928 + e_t - 0.3218e_{t-2} - 0.2265e_{t-12}$$

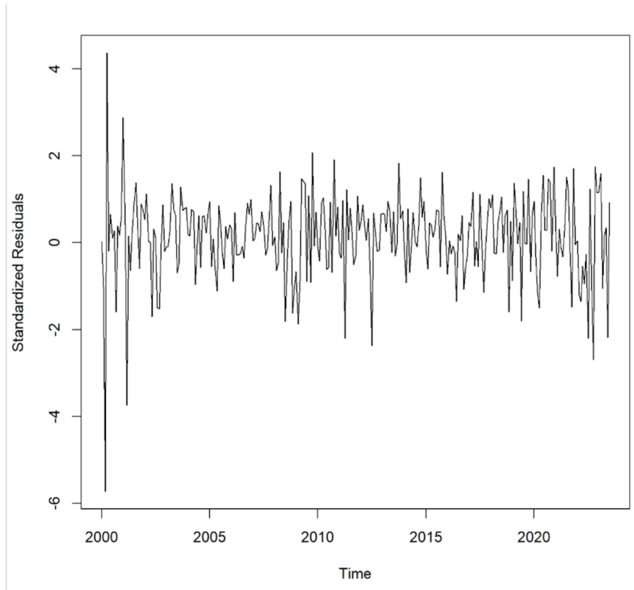
However ϕ_2 and θ_2 have standard errors of 0.3249 and 0.3521 meaning neither of the nonseasonal terms are significant. This data is very clearly not solely seasonal so only including a seasonal term in the model would not make sense, therefore this model will be eliminated.

If a seasonal AR term is added the AIC drops to -632.8 and gives the equation:

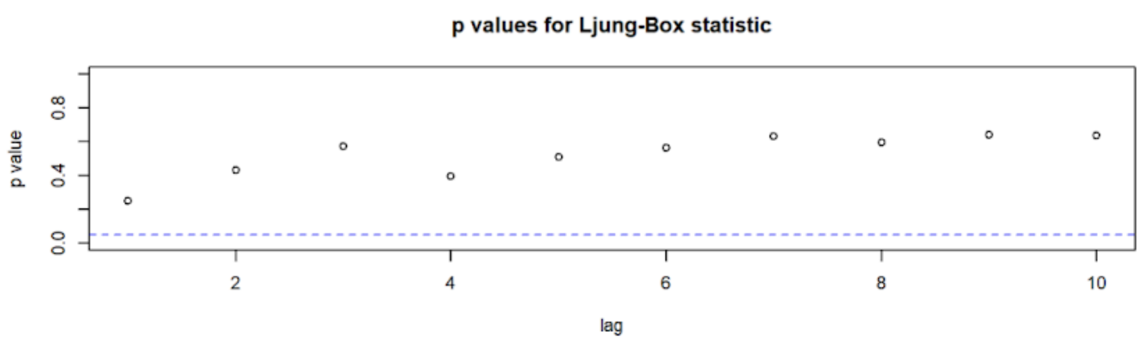
$$W_t = -0.5268 + e_t - 0.3536e_{t-2} - 0.2647Y_{t-12}$$

However once again ϕ_2 and θ_2 have standard errors of 0.2939 and 0.3222 meaning neither of the nonseasonal terms are significant. Based on both of these attempts to add a seasonal component to the data I have concluded that such a term is unnecessary and convolutes the data and I will instead stick with the original ARIMA(2,1,2) model despite the AIC claims that the seasonal models are better.

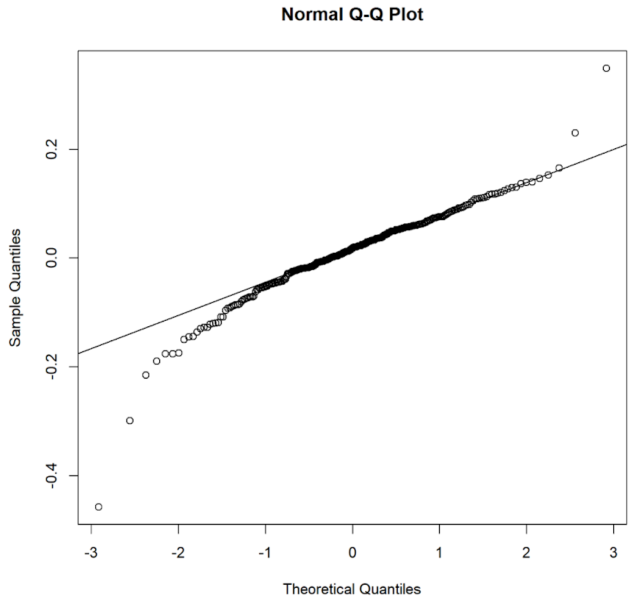
Continuing with the diagnostic techniques looking at the standardized residuals should be done next.



The standardized residuals plot shows two to four significant outliers very early in the plot but otherwise the residuals seem to be good. Furthermore it is good to see that there are no significant outliers in 2008 or 2020 during the Great Recession and COVID-19 pandemic which greatly reduced most stock prices. The next plot that should be looked at is the Ljung-Box Test which tests the null hypothesis that there is no residual autocorrelation.



The Ljung-Box Test shows no significant autocorrelation between residuals at any lag. Similar to the standardized residuals the QQ-plot can be used to check normality of the residuals.



The QQ-plot once again shows the significant outliers as well as some slight to moderate left skewing. Finally, based on a runs test, which tests the null hypothesis that the errors are independent, a large p-value of 0.515 is obtained meaning the null is not rejected.

Based on all of this information it seems very reasonable that the fixed ARIMA(2,1,2) model is a good model and the residuals are not correlated. One final overfitting technique can be done by adding a third AR or MA term to the model. When an extra AR term is added $\phi_3 = -0.002$ with a standard error of 0.042 meaning it is not even close to significantly different from zero. When an extra MA term is added $\theta_3 = -0.0097$ with a standard error of 0.0529 meaning that, once again the addition term is not significantly different from zero whatsoever. Based on overfitting there is no reason to believe that a model of any higher orders is necessary.

Adjusting for Outliers

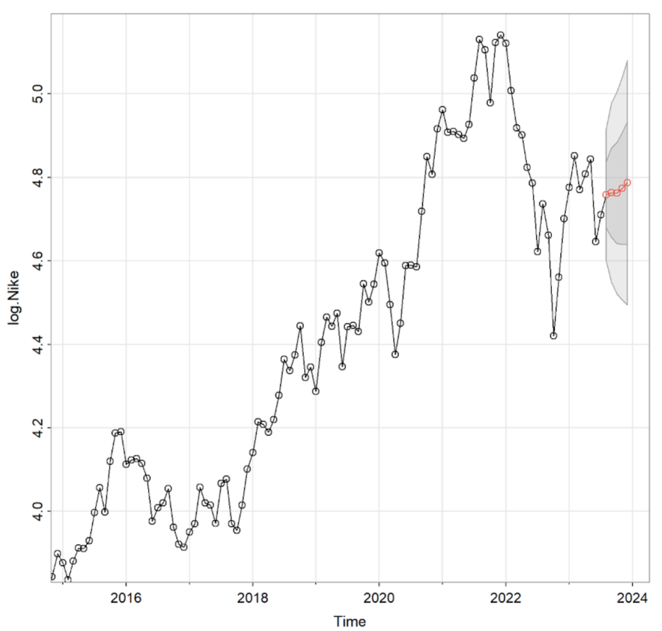
The first thing to do when adjusting the model for outliers is to determine exactly which points are outliers and what type of outliers they are. An outlier can either be innovative (IO) or additive (AO). An IO affects the series both at and after the time which it occurs while an AO only affects the series at the time it occurs. When testing the model for outliers there are significant AOs and IOs at March 2000, April 2000, and March 2001. When adding the three IO terms to the model the AIC drops to -693.2 and has the equation:

$W_t = -0.7443W_{t-2} + e_t - 0.667e_{t-2}$ with IO coefficients

$\psi_3 = -0.4683, \psi_4 = 0.3587, \psi_{15} = -0.3185$ where ψ_t turn on at time t and decay towards zero following t. It is very difficult to forecast with this model however these three outliers all occur at the beginning of the series so it a valid assumption to believe that all three ψ_t terms are equal to zero by the end of the series. Thus using the previously determined model seems viable for predicting future values.

Forecasting

Using the SARIMA function to plot the next five predicted values of the time series as well as their confidence intervals gives the following plot:



Predicting the specific future values and their 95% confidence intervals gives the following values:

Date	Point Estimate	Standard Error	Lower Bound	Upper Bound
August 2023	4.751075	0.07984339	4.594582	4.907568
September 2023	4.75006	0.11291561	4.528745	4.971374
October 2023	4.716101	0.13282222	4.455769	4.976432
November 2023	4.716942	0.15011178	4.422723	5.011161
December 2023	4.745066	0.16923272	4.413369	5.076762

These are the logged prices so when raising each estimated value to e will give the actual predicted values of Nike's stock price, those values, with their 95% confidence intervals, and the true value of Nike's stock at that time are as follows (the final five values were withheld in order to confirm or deny the plausibility of the forecasts):

Date	Point Estimate	Lower Bound	Upper Bound	Actual Price
August 2023	115.7086	98.94678	135.32	110
September 2023	115.5912	92.64223	144.225	101.97
October 2023	111.7317	86.12238	144.9563	96.2
November 2023	111.8258	83.32285	150.0789	102.55
December 2023	115.0153	82.54713	160.2543	110.33

While all of the actual prices are below the predicted price they all fall well within the 95% confidence interval meaning this is probably a relatively good model overall and for forecasting future values.

Discussion

Overall fitting an ARIMA(2,1,2) while keeping the ϕ_1 and θ_1 terms constant makes for a good model of the logged price of Nike's stock. While the model is not perfect, primarily it has a few outliers that may alter the model slightly, it is definitely the best model and provides good insight to the data. The model is able to accurately predict the future values of the time series within an acceptable interval and the residuals of the model seem uncorrelated. Based on that confirming this model makes sense. While the residuals may not be perfectly normally distributed that should not be too much of an issue as virtually every other test points towards there being no need to alter the model. Finally when looking at outliers it is reasonable to believe they will have a negligible effect on the data despite their large residuals. This is because the outliers occur early in the series and should not have too much of an affect on future values when predicting prices.

The model $W_t = -0.8282W_{t-2} + e_t - 0.7042e_{t-2}$ should be used to model the logged stock price for Nike.

Works Cited

Hayes, Adam. “Stocks: What They Are, Main Types, How They Differ from Bonds.” *Investopedia*, Investopedia, 21 Apr. 2023, www.investopedia.com/terms/s/stock.asp.

“Nike, Inc. (NKE) Stock Price, News, Quote & History.” *Yahoo! Finance*, Yahoo!, 8 Dec. 2023, finance.yahoo.com/quote/NKE?p=NKE.

“Nike, Inc..” *Wikipedia*, Wikimedia Foundation, 28 Nov. 2023, en.wikipedia.org/wiki/Nike,_Inc.