Time Series Analysis of Netflix Stock

David Tolentino davidtolentino@ucsb.edu

March 25, 2023

Abstract

The goal of this study was to implement and compare the effectiveness of two time series modeling and forecasting techniques: Seasonal autoregressive integrated moving average (SARIMA) and a generalized autoregressive conditional heteroskedasticity in analyzing and predicting adjusted closing price of Netflix Stock from February 2018 to Feburary 2022. Exploratory data analysis was implemented on our data set with various trends and patterns being identified in order to better understand the underlying behavior of the stock price. The SARIMA was used to comprehend the seasonality, while the GARCH model was was used to capture the volatility of our stock. The results of our analysis indicate that Netflix does tend to show some type of seasonality and that the the volatility of the stock was generally pretty stable. Overall, our project was able to demonstrate the proper use of time series techniques for the purpose of creating insights on existing stock data.

Introduction

The purpose of this project is to explore the use of SARIMA and GARCH models on the analysis of Netflix stock data. Specifically, we will be focusing on Netflix's adjusted closing price of stock data, which is commonly used as indicator of a stock's performance. The reason this topic was chose was due to the apparent impact stock prices have in today's economy and financial markets. By being able to understand trends and patterns within a stock's price we will be able to provide general inferences on a stock's performance.

Numerous studies on stock data have been done similarly in the past. For this data set in particular other users from the site of which this data was taken from have also fit and ran the same techniques we will be using in our project.

In this project we will be fitting SARIMA and GARCH models to the adjusted closing price of Netflix stock. The goal will be to identify trends and patterns as well as forecast future prices and analyze levels of volatility through our fitted models.

Overall, the obejctive of this project will be to utilize time series analysis techniques to analyze real-world data and gain deeper insights into their practical applications. By applying advanced analytical methods to relevant data sets, we aim to extract meaningful patterns and trends and ultimately improve our knowledge on practical applications of time series analysis in various fields.

Data

The data being utilized for this project is Netflix stock prices over a 5 year time period from February 2018 to February 2022. The data set was sourced from the website https://www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction, and was scraped by kaggle user Janil Shah. Data was initially obtained off of Yahoo Finance and then uploaded to Kaggle.

The dataset is comprised of irregular frequency values, and our purpose for using it will be to gain insights on the practical applications of time series analysis. Our dataset contains a significant amount of observations, with 1009 total recorded values being accounted for in our data set.

It is important to study financial data like Netflix stock in order to build a better understanding on when investments should be made in order to maximize profit.

Methodology

SARIMA

A SARIMA model will be used to model and forecast the seasonal patterns of our data. The model is derived from the original ARIMA model and involves incorporating seasonal components which capture the regular pattern of variation in data that occurs at fixed intervals.

A SARIMA model is built by specifying (p, d, q) of our non-seasonal component and (P,D,Q, s) for our seasonal component. The parameters represent the order of our autoregessive, integrated, and moving average parameters for both our non-seasonal and seasonal components.

GARCH

A GARCH model will be used to analyze and forecast the volatility of our time series. The model allows for the specification of both autoregressive and moving average terms when calculating conditional variance.

The model is specified using two equations: one that calculates conditional mean and another that represents conditional variance. Conditional variance is calculated based off past variances and passed squared errors. This allows the model to capture the time varying volatility within a time series.

```
library(forecast)
library(astsa)
library(forecast)
library(tseries)
library(xts)
library(fGarch)
library(readr)
```

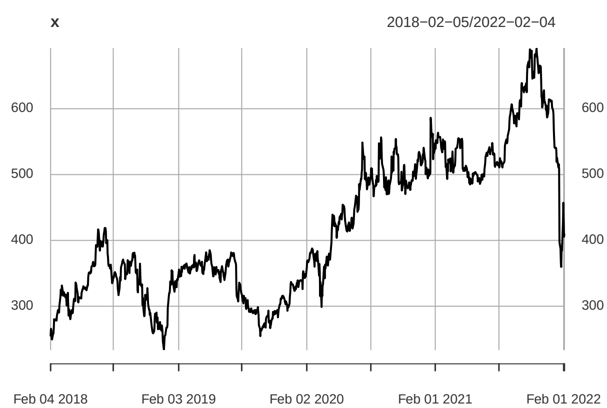
EDA

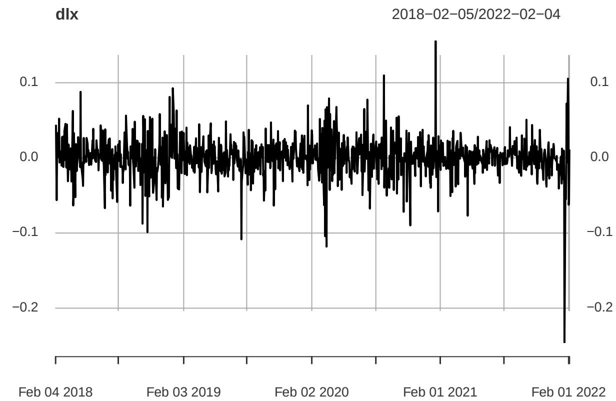
```
NFLX <- read_csv("NFLX.csv")
Netflix = xts(NFLX$`Adj Close`, order.by = NFLX$Date)
```

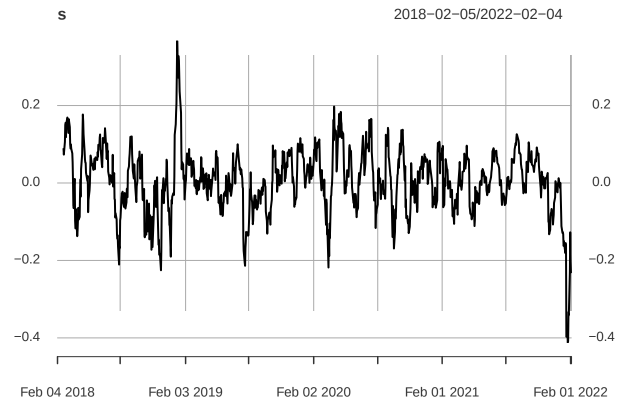
```
x = Netflix #original
lx = log(Netflix) #minimize variance
dlx = diff(log(Netflix)) #detrend
s = diff(log(Netflix),12) #seasonal
```

Transformed Plots

```
plot(x)
plot(lx)
plot(dlx)
plot(s)
```







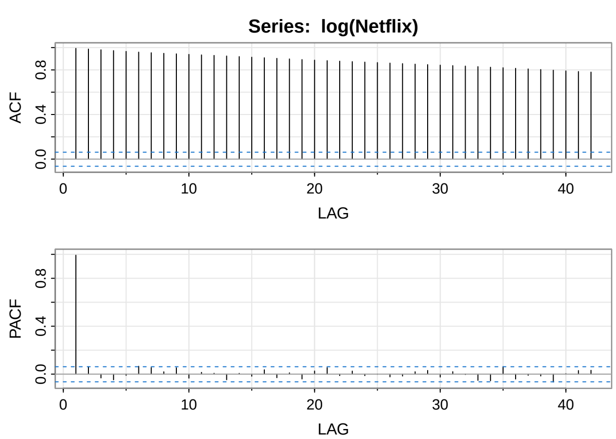
First we must plot our stock data in order to determine if there any transformations that must be done before we carry out our analysis.

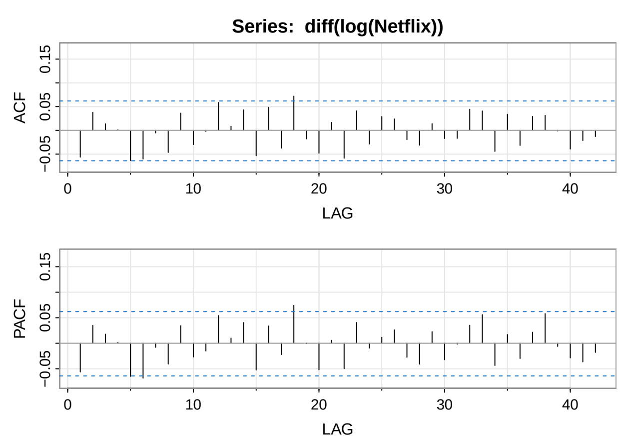
From our initial plot it appears that there is a trend of exponential growth. This is an indication that we should log our data in order to minimize variance. Once we have done that we can take the first difference for the purpose of detrending our data and ensuring we are analyzing a stationary time series. We can then take a look at our data from a seasonal perspective by taking the 12 difference. In doing this we will be observing our data on a monthly cycle.

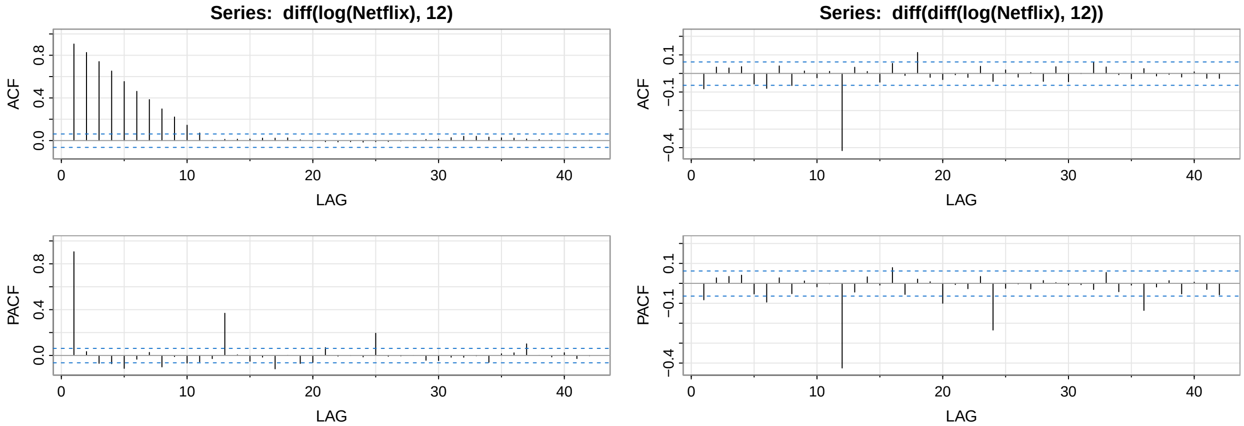
Plot `x` represents our original data, plot `lx` our logged data, plot `dlx` our detrended data, and finally plot `s` our seasonal data.

ACF and PACF Results

```
acf2(log(Netflix))
acf2(diff(log(Netflix)))
acf2(diff(log(Netflix), 12))
acf2(diff(diff(log(Netflix), 12)))
```







Now that we have we properly transformed our data we can analyze their ACF and PACF plots in order to fit a SARIMA model to our data.

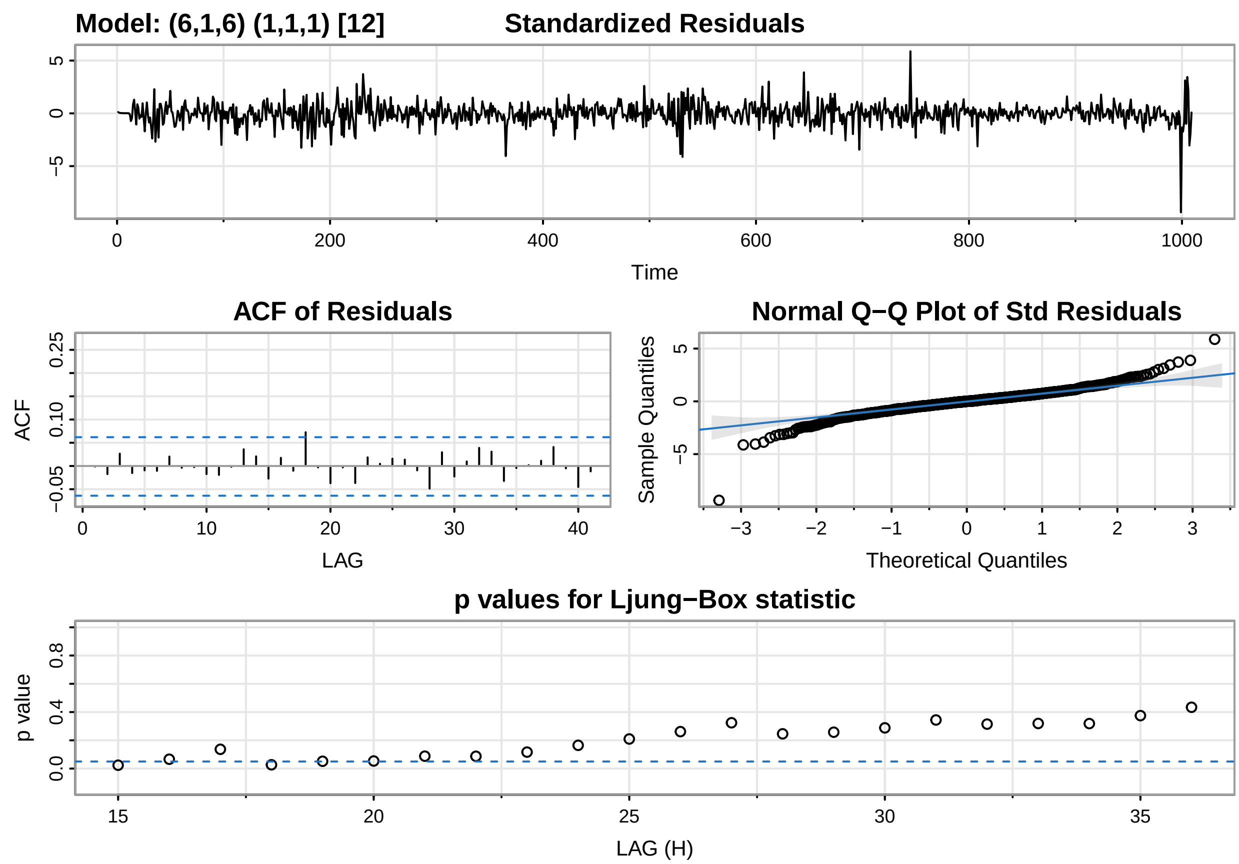
Since the ACF of our logged data presents a decreasing trend between lags it is appropriate to take the first difference of our non seasonal data. The seasonal ACF and PACF plot also exhibits this data so we will take the first difference of our seasonal data.

From our difference plots we can now determine the parameters for both our non seasonal and seasonal parts of our SARIMA model.

SARIMA

Sarima Model Diagnostics

```
sarima(log(Netflix), 6,1,6, 1,1,1, 12)
```

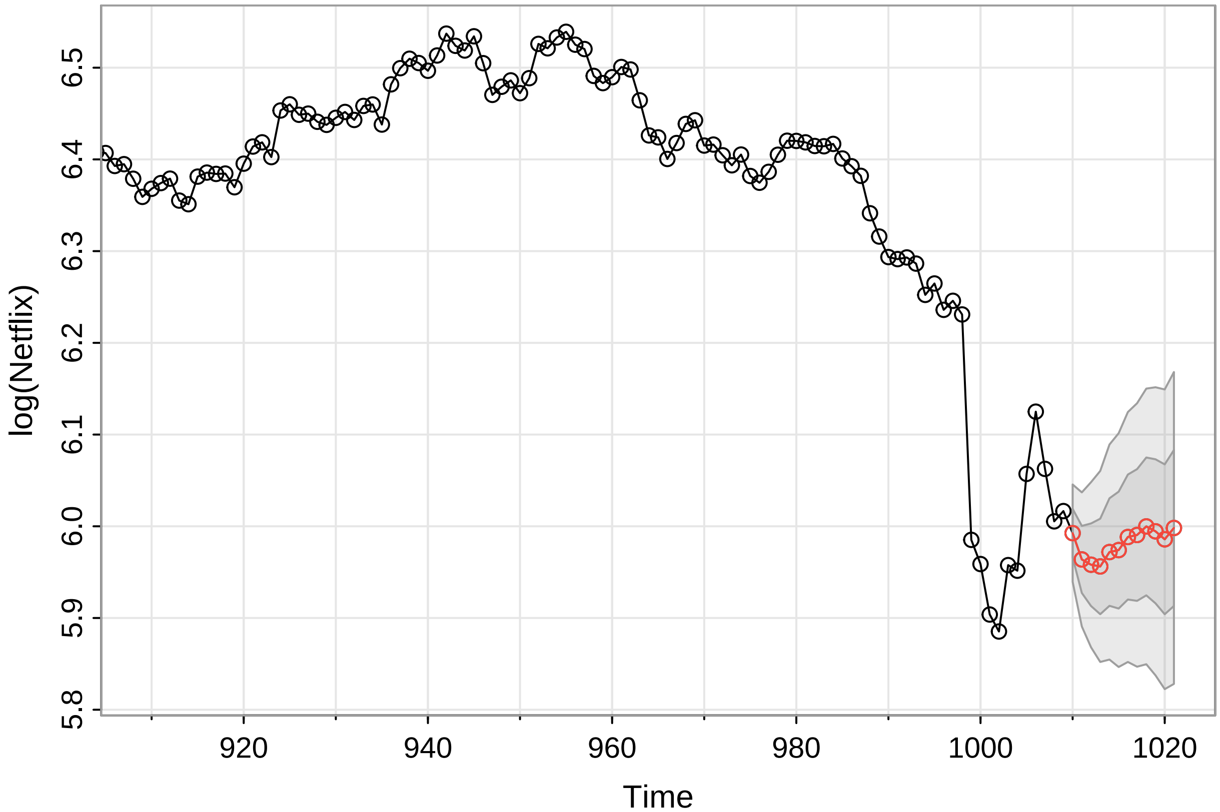


After looking at our ACF and PACF plots, we can test the performance of different models and select a model by evaluating our residuals using the sariama function in R.

I have settled with a model using the parameters (6,1,6)x(1,1,1,12). This model provided the best results from our residual diagnostics.

Forecasted SARIMA model

```
sarima.for(log(Netflix), 12, 6,1,6, 1,1,1, 12)
```



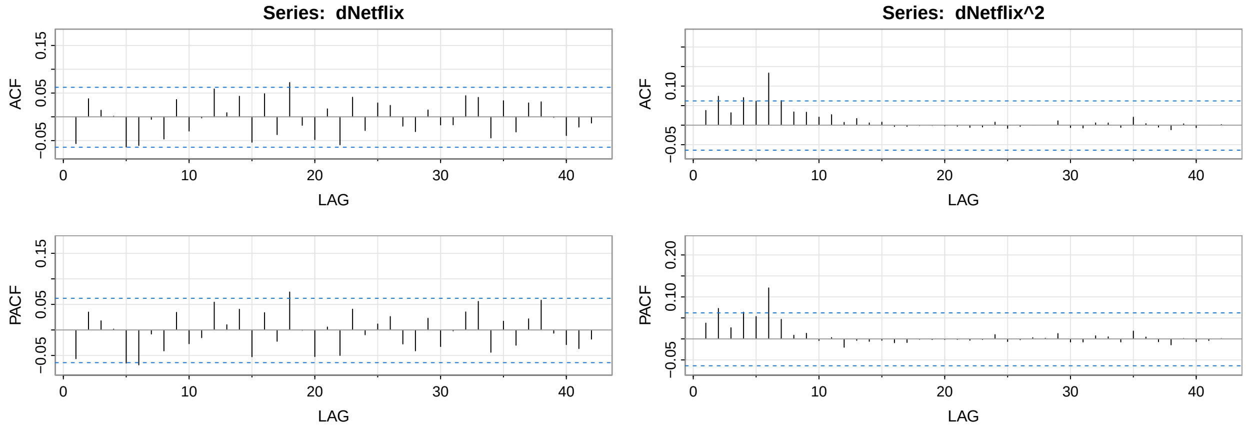
We can then use our model to forecast the next 12 observations of our original data. From the plot we can see that the stock will continue with a downward trend and then slowly to start to rise after.

GARCH

For fitting a GARCH model we must look at the ACF and PACF our returns and returns squared data for the purpose of determining the parameters of our model.

ACF and PACF of Returns and Returns Squared

```
dNetflix = diff(log(Netflix))[-1]
acf2(dNetflix)
acf2(dNetflix^2)
```

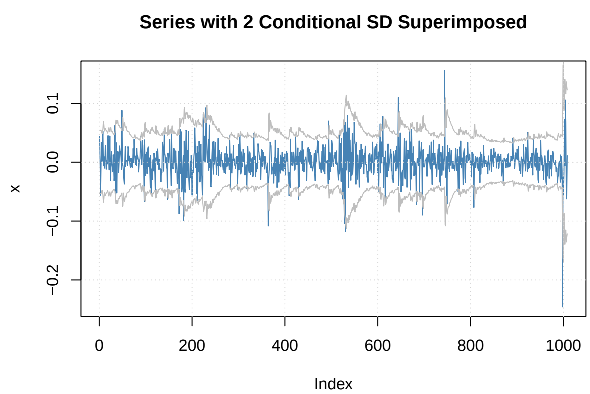


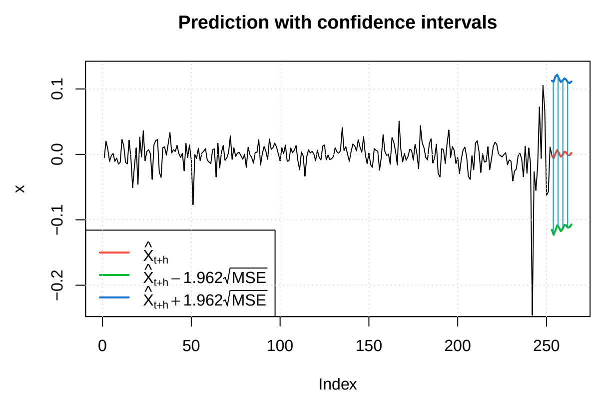
From our plots we can start fitting different models and pick our best GARCH model

Plot of Conditional Variance and Forecasted Conditional Variance

```
dNetflix.g = garchFit(~arma(2,2)+garch(3,4), data = dNetflix, trace = F, cond.dist = "std")
plot(dNetflix.g, which = 3)

predict(dNetflix.g, n.ahead=12, plot = T)
```





From our fitted model we can plot our GARCH model and superimpose over our original returns data. As we can see our data appears to fit within our conditional Standard Deviation predictions. There does exists a few spikes where a data looks much more volatile than what our model predicted.

We can then use our model to forecast the next 12 intervals for the volatility of the Netflix Stock Data.

Conclusion and Future Study

In this project, we applied time series techniques to model and forecast Netflix stock data. We first built a SARIMA model and used it to forecast the next 12 observations of the stock data. Our forecast indicated that the stock price will continue to rise in an upward trend.

Next, we fit a GARCH model to the data to analyze the volatility of the stock. Our GARCH model revealed that Netflix is a relatively volatile stock, as indicated by multiple significant spikes in the model. We then used the GARCH model to forecast the next 12 time intervals of the stock's volatility with a 95% confidence interval. Our analysis showed that the forecasted probability was relatively stable.

Overall, this project provided an opportunity to gain practical experience with time series concepts and their applications. In the future, we plan to explore other types of models, such as state space models and neural networks, to identify more accurate models and improve our forecasting results.

References

Data was retrived from https://www.kaggle.com/code/avantikab/netflix-stock-prediction-with-arima the link includes the data set used for the project. Data was scraped off of Yahoo Finance.

The project used https://www.kaggle.com/code/avantikab/netflix-stock-prediction-with-arima as a reference in order to get an idea of how to perform a time series project.