PSTAT 174 Final Project

Evan Pei

evanpei@ucsb.edu

University of California, Santa Barbara

Jun 14, 2023

Abstract

This project presents a time series forecasting report using SARIMA and ARIMAX models with the goal of forecasting the future price of Netflix (NFLX), the leading streaming service on the internet, for the next 12 months or 1 year. The dataset contains closing prices from its inception in 2006 to January 2022.

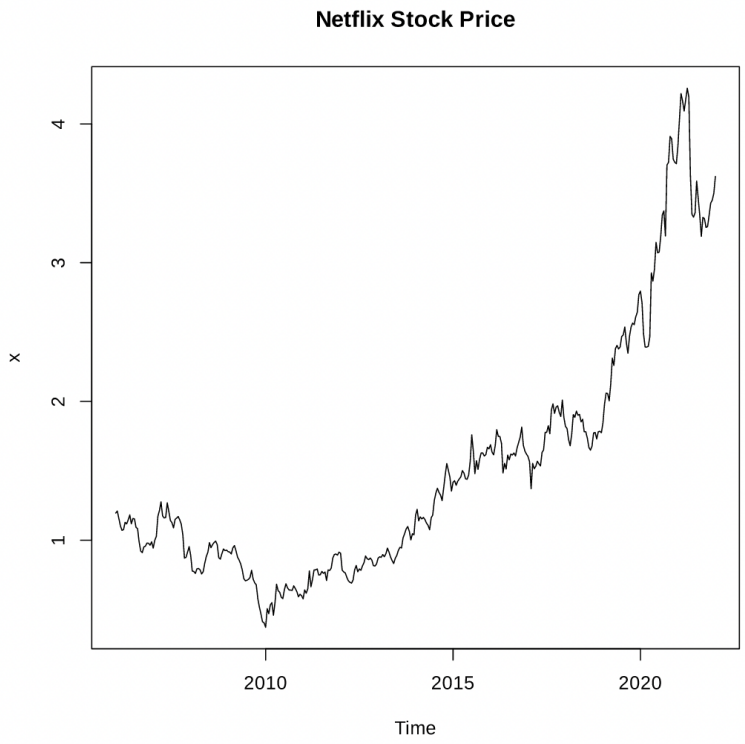
1. Introduction

The stock market is a culmination of buyers and sellers who can buy or sell shares of companies. There are many stocks that are considered risk averse such as blue chip stocks which have existed and thrived over a long period of time and buyers can expect a slow and steady growth. There are also many risk seeking stocks that are highly volatile and buyers can expect high amounts of fluctuations over a short period of time. Stocks in the S&P 500 are considered to be generally safe companies to invest into with little volatility so choosing a stock from that list was a good start. I chose Netflix (NFLX) as it is a large company that has endured the ups and downs of the market and is still very popular to invest into. It has significantly increased in value since its inception and would continue to gain overall value despite recent controversy. Closing price was chosen as the time series since it was the most reliable factor since opening prices can be affected by closed market trading and highs and lows of the days are just fluctuations in the market.

2. Data

The dataset used was pulled from Kaggle.com showing the opening, closing, highest and lowest prices of Netflix Inc (NFLX) from its inception in 2006 to 2022. The data was gathered from Yahoo Finance all by user “Meet Nagadia” and can be found on Kaggle here: https://www.kaggle.com/datasets/meetnagadia/netflix-stock-price-data-set-20022022

This dataset is extremely important as it is the basis of information for the methods we will use to forecast the future stock prices.



3. Methodology

Procedure for SARIMA forecasting

After opening the .csv file of the NFLX stock, we will choose the closing price as our time series data set and then graph to check stationarity, trends, variance, etc and apply the necessary transformations if needed. The graph will very likely not be stationary to start but we can assume stationarity once the graph exhibits an exponential decrease in ACF. Once stationarity is reached we will graph the Autocorrelation function (ACF) and partial autocorrelation function (PACF) and obtain the moving average (q) value by finding the last significant ACF value. The AR (p) value is consequently found by the last significant PACF value.

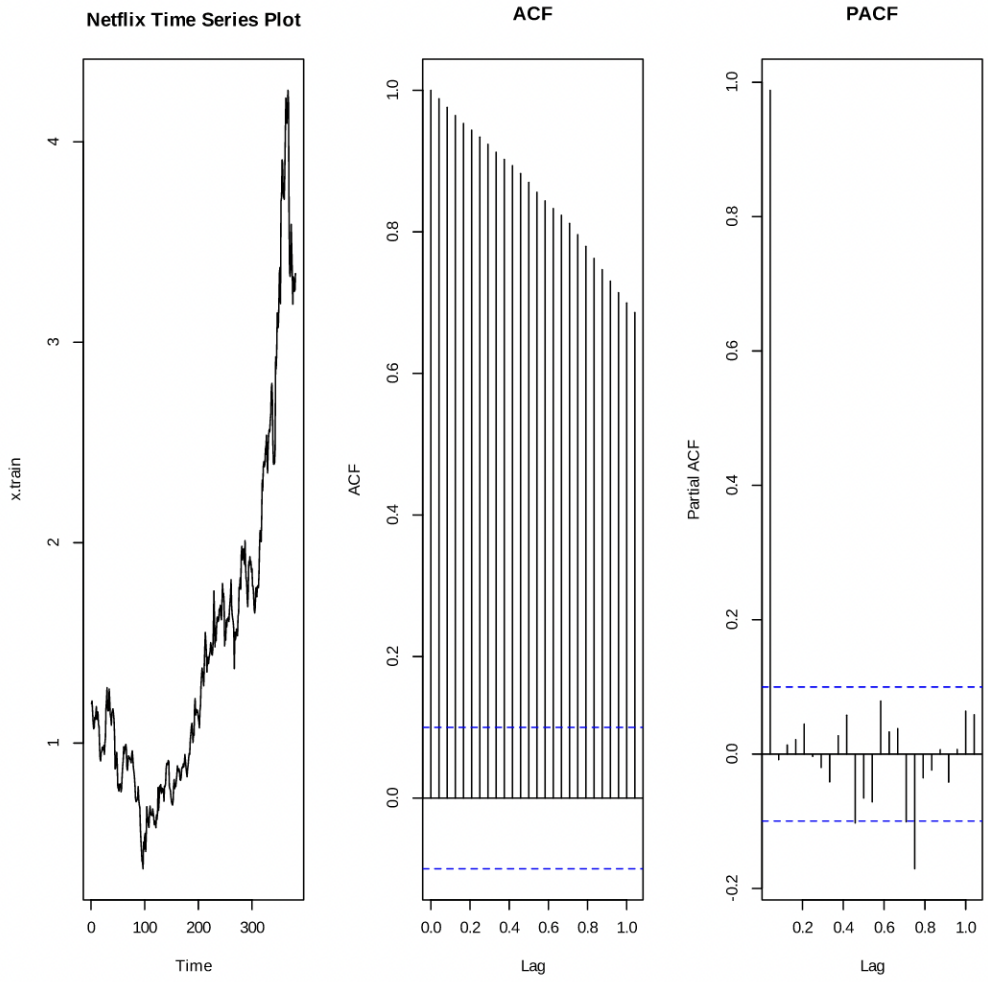
Procedure for ARIMAX forecasting

The purpose of ARIMAX modeling is to see if we can use another stock or factor that can be use to forecast NFLX. Just like ARIMA modeling we load the csv and select the closing price as the time series and set the range dates to that of NFLX setting this to be the exogenous time series. We will use ARIMAX to forecast NFLX's stock using an exogenous time series where we will use Nike (NKE) and select the date range to be the same as NFLX which will be Jan 2020-Jan 2022 with a frequency of 24. We can create the ARIMAX model using Arima() with the parameters of the prior SARIMA model where x is the closing price of NFLX and the xreg is the closing price of NFLX.

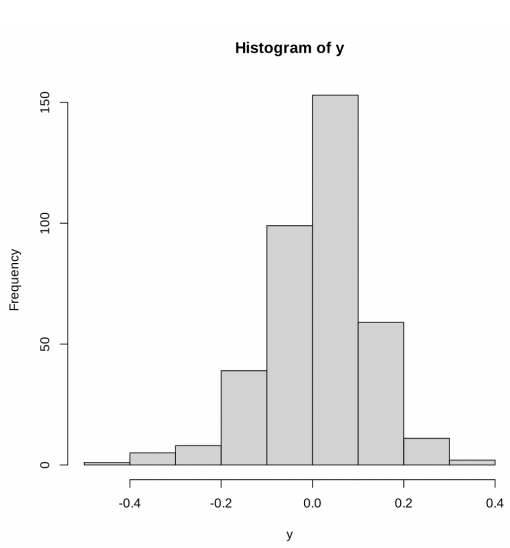
4. Results

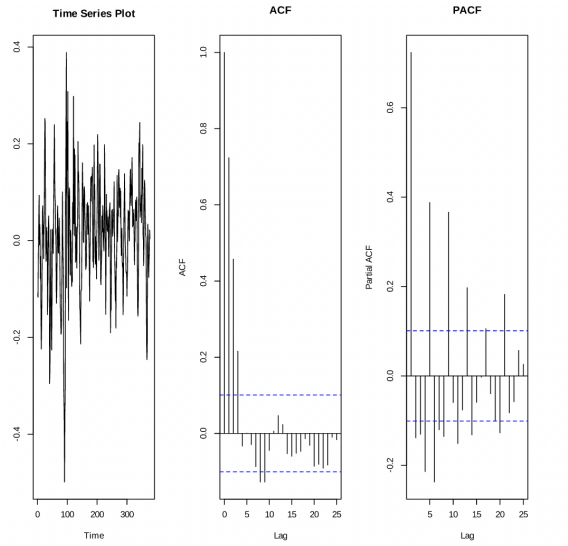
Results for SARIMA forecasting

After loading the .csv we notice that it is not stationary based on the actual plot as well as the slow declining ACF so we must take the difference of the log so that the variance was constant throughout.



We can ensure stationarity has been achieved by noticing that the autocorrelation function (ACF) value of the diff(log()) was rapidly decreasing and checking the plotted histogram and seeing a bell curve indicating normal distribution.





Now that stationarity has been achieved by bell curve and rapidly decreasing ACF, we can now find the best (p,d,q) from the significant lags of the ACF and PACF. I ended up choosing the (p,d,q) to be (2,1,2) based on the plots and ending up with the following ARIMA model with AIC -828.67 AICc -828.51and BIC -808.97:

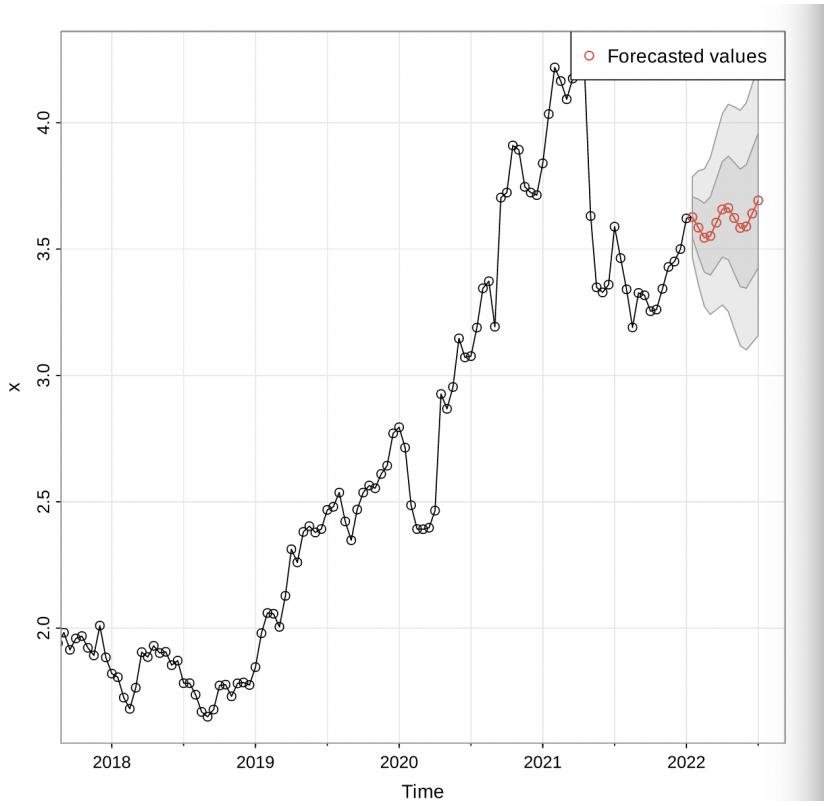
$$X_t - 0.6711X_{t-1} + 0.937X_{t-2} = Z_t + 0.9145Z_{t-1} - 0.19145Z_{t-2}$$

```
Series: x.train
ARIMA(2,1,2)

Coefficients:
         ar1      ar2      ma1     ma2
      0.6711  -0.9371  -0.5865  0.9145
s.e.  0.0439   0.0488   0.0456  0.0611

sigma^2 = 0.006501:  log likelihood = 419.34
AIC=-828.67   AICc=-828.51   BIC=-808.97
```

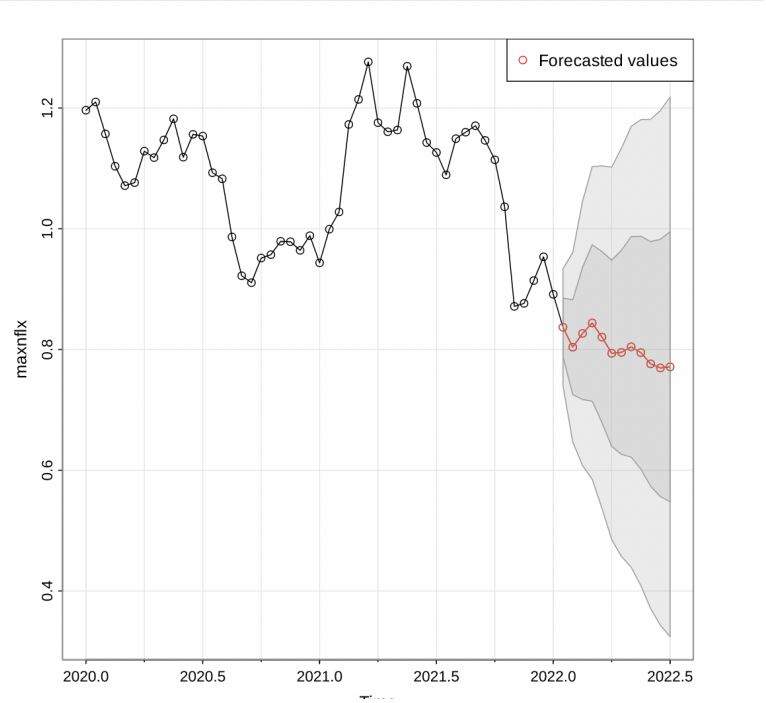
Due to the lack of seasonality of the NFLX stock, P, D, Q, S were not defined when plotting the SARIMA model. The prediction forecast plot for the next 12 months are as follows.



We can conclude that the SARIMA model was correct with the general upward trend from 2022-2023 but it appears to just have followed a sinusoidal wave up and down like previous fluctuations.

Results for ARIMAX Forecasting

After loading the .csv for Nike (NKE) and setting the closing price as the time series, I set the date ranges to be 2 years for both NFLX and NKE. With this exogenous time series we saw that the forecast inaccurately predicted that NFLX would continue to fall. This is most likely because the NKE's value went down at the same time and ARIMAX model assumed that NFLX would go down as well since there are only 2 datasets being taken into account. I believe that ARIMAX would be more accurate if the forecasted stock would have been one that had no controversy so that correlation between the 2 stocks could be made and attributed to the general economical climate. In other words if there were 2 stocks just following normal market volatility without controversy, the monthly trends would be more in line with the market conditions and could more accurately forecast future values.



5. Conclusion

We can conclude that the SARIMA method did fairly well in predicting the next year and even the next 2 years. While it couldn't predict any FUD and other market influencing factors, it did predict the slight upward trend that did occur in real life. As for the ARIMAX model, it predicted that NFLX would continue to fall like it did from 2021-2022 since NKE also trended downward as it assumed that the general market and economy was falling. So reiterating from the Results section, if there were 2 stocks just following normal market volatility without controversy, the monthly trends would be more in line with the market conditions and could more accurately forecast future values. In terms of future study, using more in depth and complicated parameters and modeling methods could prove successful in terms of predicting stock values. This is a good starting point for one to continue using this method for other stocks tickers to see what models work best. SARIMA/ARIMA may be more beneficial for other time series datasets where forecasting is almost exclusively based on previous history. It is not very accurate for stock forecasting since there are many other factors that influence the value of stock prices like FUD, and political and economical climate.

Project

June 15, 2023

[ ]:
```
library(MASS)
library(astsa)
library(forecast)
require(forecast)
NFLX <- read.csv('NFLX.csv')
x <- ts(NFLX[, c("Close")], start = c(2006,1), end = c(2022,1), frequency = 24)
ts.plot(x,type='l',main= 'Netflix Stock Price')
n <- length(x)
x.train <- x[1:(n-4)]
x.test <- x[(n-3):n]
par(mfrow=c(1,3))
plot.ts(x.train, type='l', main="Netflix Time Series Plot")
acf(x, main="ACF")
pacf(x, main="PACF")
y <- diff(log(x.train), lag=4)
plot.ts(y)
hist(y)
par(mfrow=c(1, 3))
plot.ts(y, type="l", main="Time Series Plot")
acf(y, main="ACF"); pacf(y, main="PACF")
fit1 <- arima(x.train, c(2,1,2), method='ML')
fit1
auto.fit1 <- auto.arima(x.train)
auto.fit1
We have the ARIMA(2,1,2) model:
$$X_t - 0.6711X_{t-1} + 0.937X_{t-2} = Z_t + 0.9145Z_{t-1} - 0.19145Z_{t-2}$$
prediction <- sarima.for(x, n.ahead = 36, plot.all = FALSE,
                         p = 2, d = 1, q = 2)
n <- length(x)
lines(n + 1:(36), prediction$pred, col = "red", main='Netflix Price prediction')
legend("topright", pch = 1, col = c("red"),
       legend = c("Forecasted values"))
We have the ARIMA(2,1,2) model:
$$X_t - 0.6711X_{t-1} + 0.937X_{t-2} = Z_t + 0.9145Z_{t-1} - 0.19145Z_{t-2}$$
NKE.csv <- read.csv('NKE.csv')
NFLX.csv <- read.csv('NFLX.csv')
```

```
maxnflx <- ts(NFLX.csv[, c("Close")], start = c(2006,1), end = c(2022,1),␣
 ↪frequency = 24)
NKE <- ts(NKE.csv[, c("Close")], start = c(2006,1), end = c(2022,1), frequency␣
 ↪= 24)
arimax <- arima(maxnflx, xreg=NKE, order=c(2,1,2))
arimax
prediction <- sarima.for(maxnflx, newxreg=NKE, n.ahead = 12, plot.all = FALSE,
                         p = 2, d = 1, q = 2)
n <- length(x)
lines(n + 1:(12), prediction$pred, col = "red", main='Netflix Price prediction')
legend("topright", pch = 1, col = c("red"),
       legend = c("Forecasted values"))
```

References:

https://www.kaggle.com/datasets/meetnagadia/netflix-stock-price-data-set-20022022