1. Introduction

The time series I used for this project was Netflix's stock price, using the quantmod package to get the data. Netflix was founded back in 1997 by Reed Hastings and Marc Randolph in California. Netflix originally sold and rented DVD's by mail, but with the advancements in technology updated their business model to a subscription based online streaming service in 2007. Netflix's use of machine learning for personalization of movie and TV shows recommendations along with its ease of use has resulted in the popularity of the streaming services company.

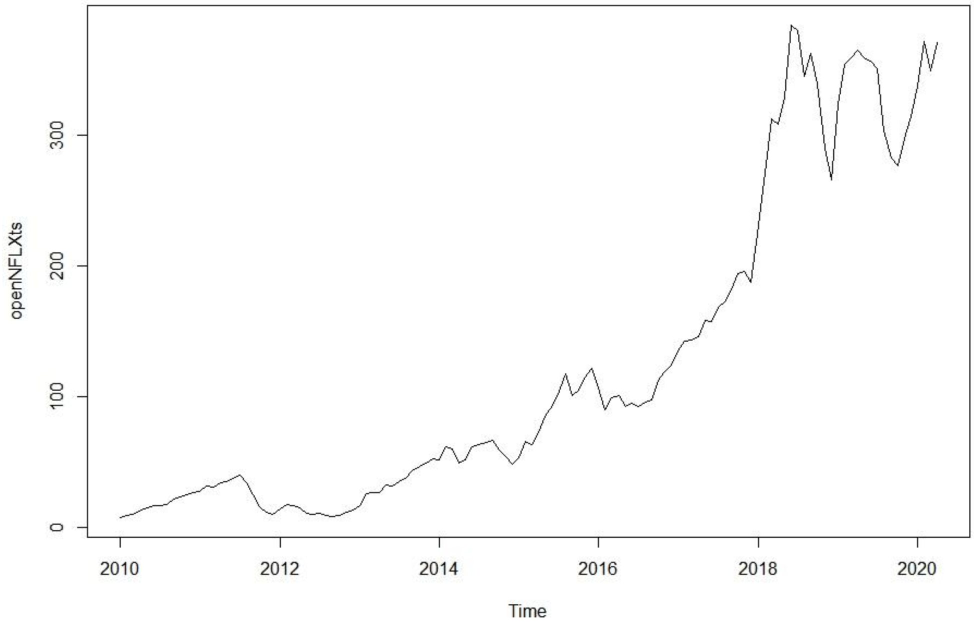
A timeline of important events is listed below, taken directly from Wikipedia:

(https://en.wikipedia.org/wiki/Timeline_of_Netflix)

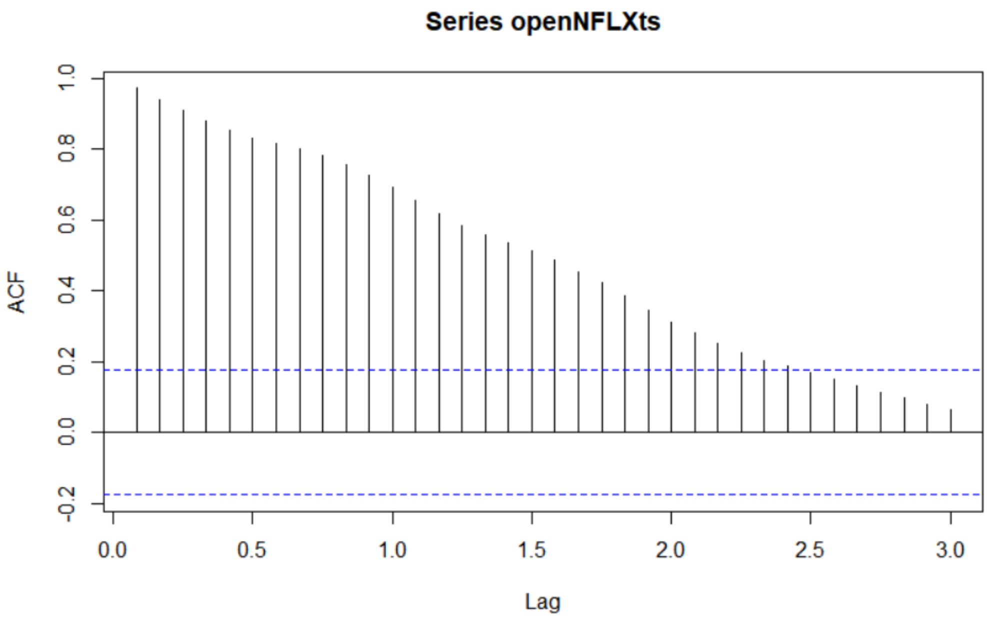
- 1997: Netflix founded in Scotts Valley, California by Marc Randolph and Reed Hastings, who previously had worked together at Pure Software
- 2002: Netflix initiates initial public offering (IPO), selling 5.5 million shares of common stock at the price of US $15.00 per share. It brings in $82.5 million.
- 2006: Netflix offers a $1,000,000 prize to the first developer of a video-recommendation algorithm that could beat its existing algorithm, Cinematch, at predicting customer ratings by more than 10%
- 2011: Reed Hastings says in a Netflix blog post that the DVD section of Netflix would be split off and renamed Qwikster, and the only major change would be separate websites for the services. This change would be retracted a month later. Netflix stock plunges from 42.16/share in July to 9.12/share in November, as 800,000 subscribers quit
- 2012: Netflix starts its expansion in Europe, launching in the United Kingdom and Ireland. By September 18 it has expanded to Denmark, Finland, Norway and Sweden
- 2015: Netflix announces a 7:1 stock split in form of a dividend of six additional shares for each outstanding share, payable on July 14 to stock owners of record at the July 2 close. Trading at the post-split price will start July 15. Netflix announces that its stock has surged to an all-time high (to almost $100/share), a growth of 574% over the past five years
- 2016: Netflix announces a major international expansion into 130 new territories; with this expansion, the company promoted that its service would now be available nearly "worldwide", with the only notable exclusions including China, and regions subject to U.S. sanctions, such as Ukraine, Syria, and North Korea.
- 2017: A study showed that the number of Netflix subscribers now equal that of all the cable subscribers combined; 73% of all US households

1.1 Netflix Data from 2010 to Date

Using the quantmod package, the monthly stock price of Netflix data from January 2010 to April 2020 was downloaded. A plot of the Netflix data is seen below along with the checks to ensure the data is not white noise or a random walk.

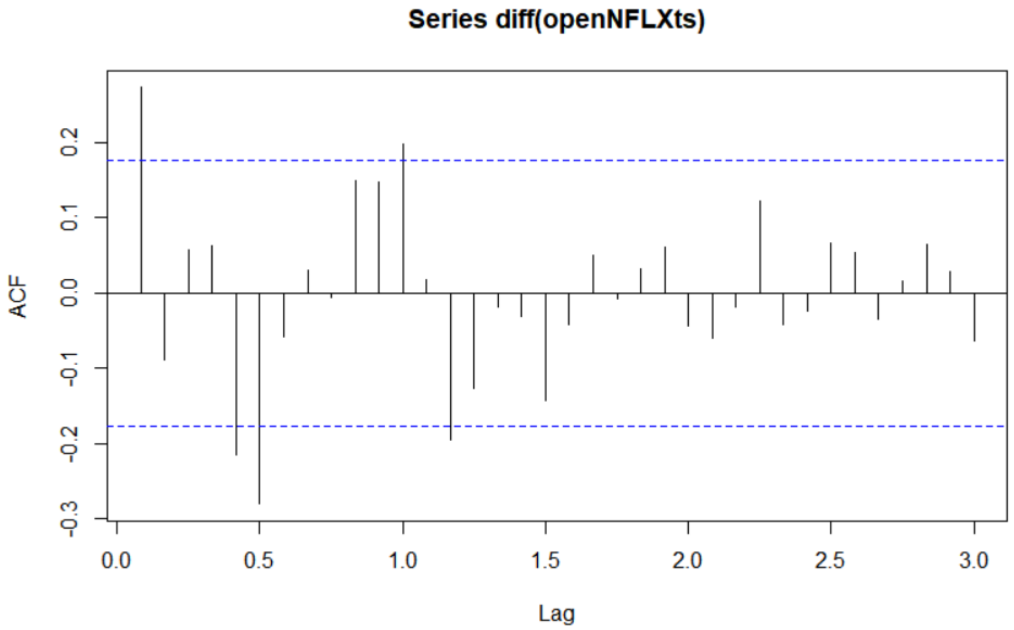


The acf function was applied to the data to investigate if it has any significant autocorrelations. If there are no significant autocorrelations, then the series is white noise.



Since there are significant autocorrelations, the series is not white noise.

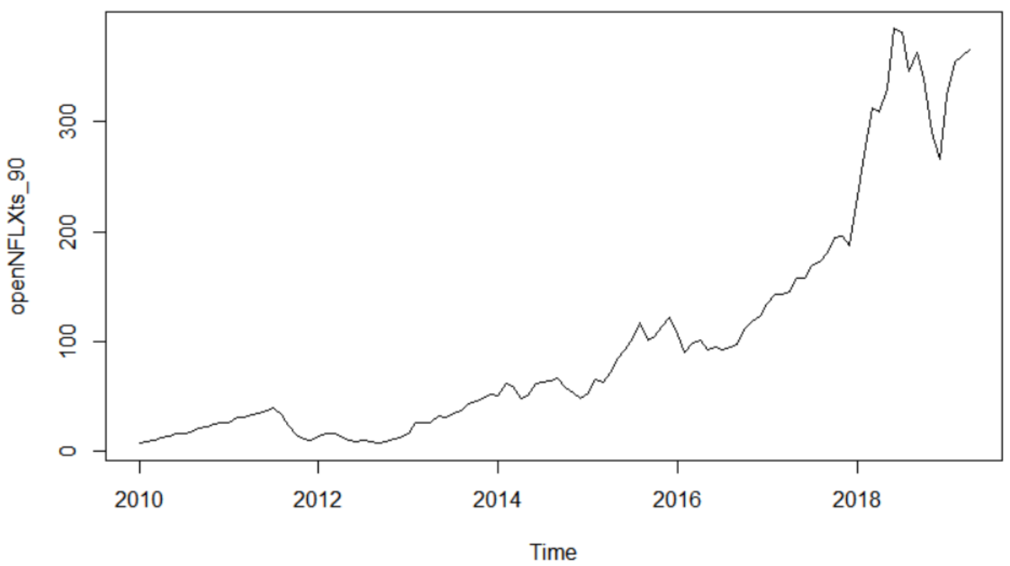
The acf of the differenced series was checked to see if there were any significant autocorrelations, if there were no significant autocorrelations then the series is a random walk.



Since there are significant autocorrelations the time series is not a random walk.

2. Removing 10% of the observations of the data

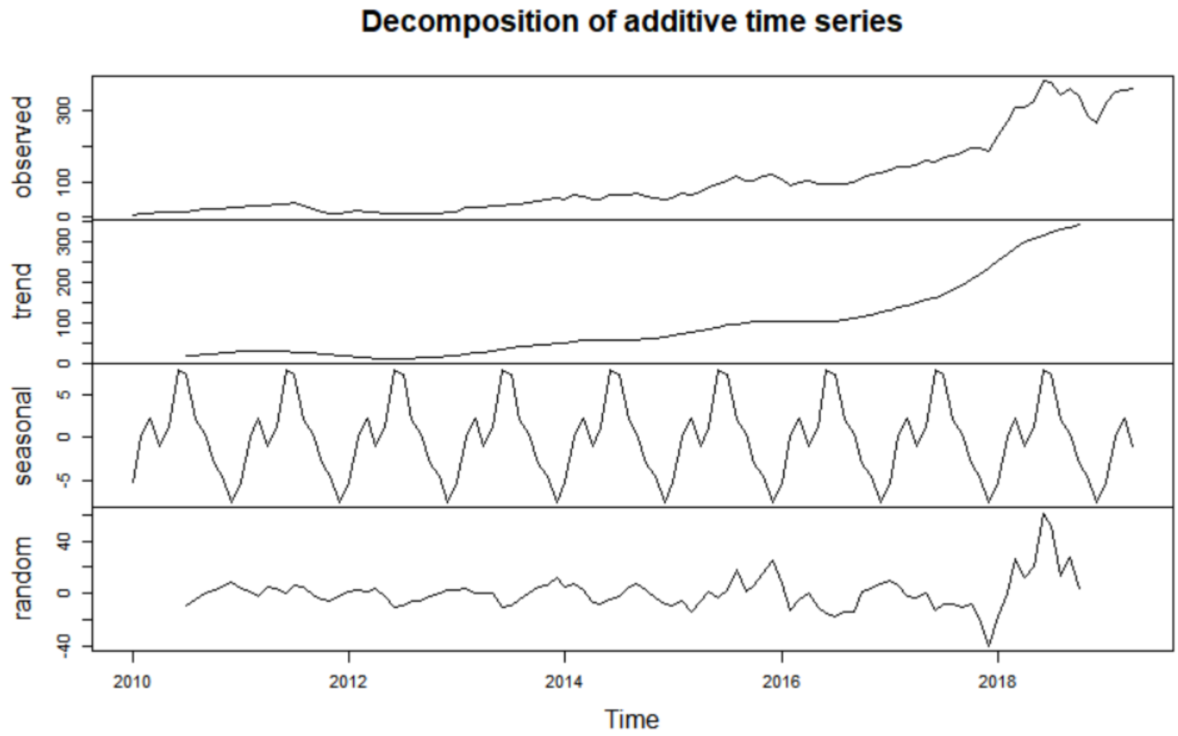
As there is roughly 10 years' worth of data, I removed 1 year from the data, so the time series consists from January 2010 to April 2019, as seen below in the plot.



3. Assessing whether the time series is stationary and applying transformation

Looking at the previous plot we can see there is a strong trend, along with the variance increasing with time, showing us the data is non-stationary. Conducting an Augmented Dickey-Fuller test on the 90% time series data to show this, I got a p-value of 0.8794. As the p-value is non-significant, the data is non-stationary, as expected. I decomposed the 90% time series data to get an idea of what it consisted of.

Decomposed 90% time series

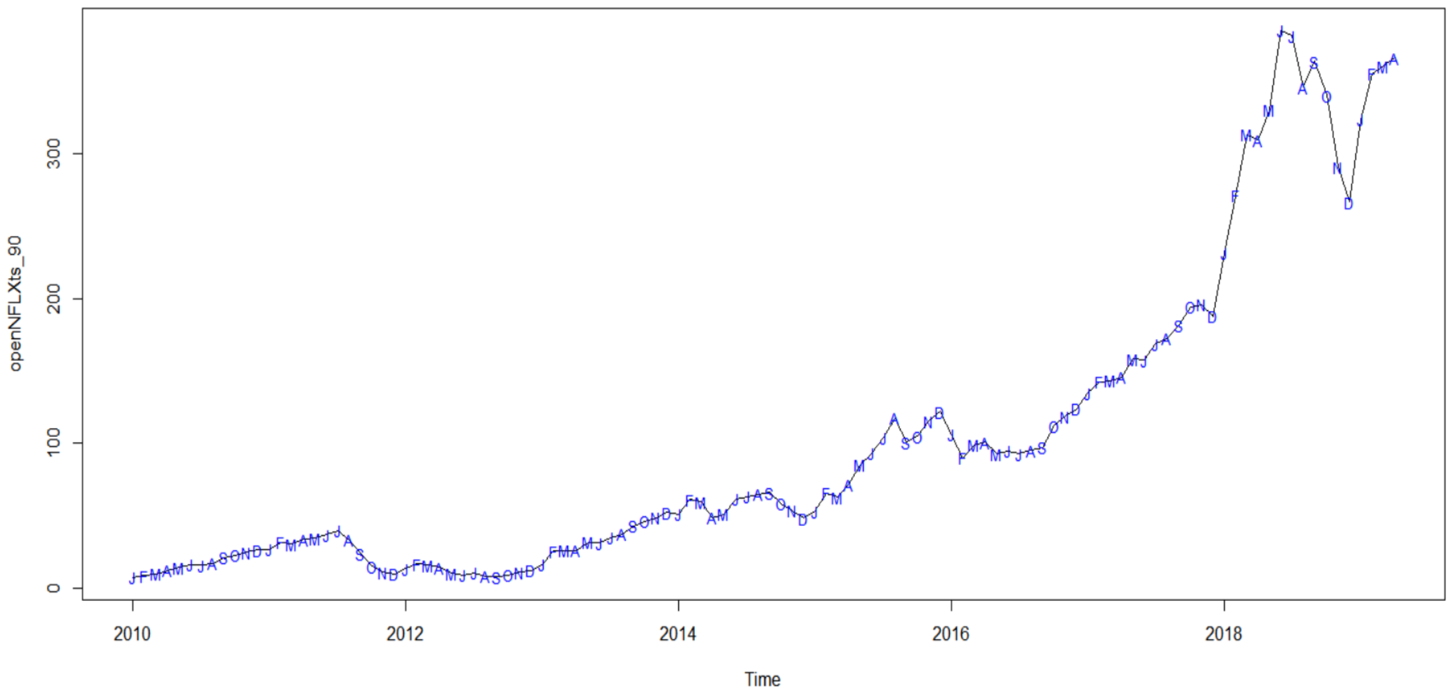


As we can see from the plot of the decomposed data, the data spans about 390 units:

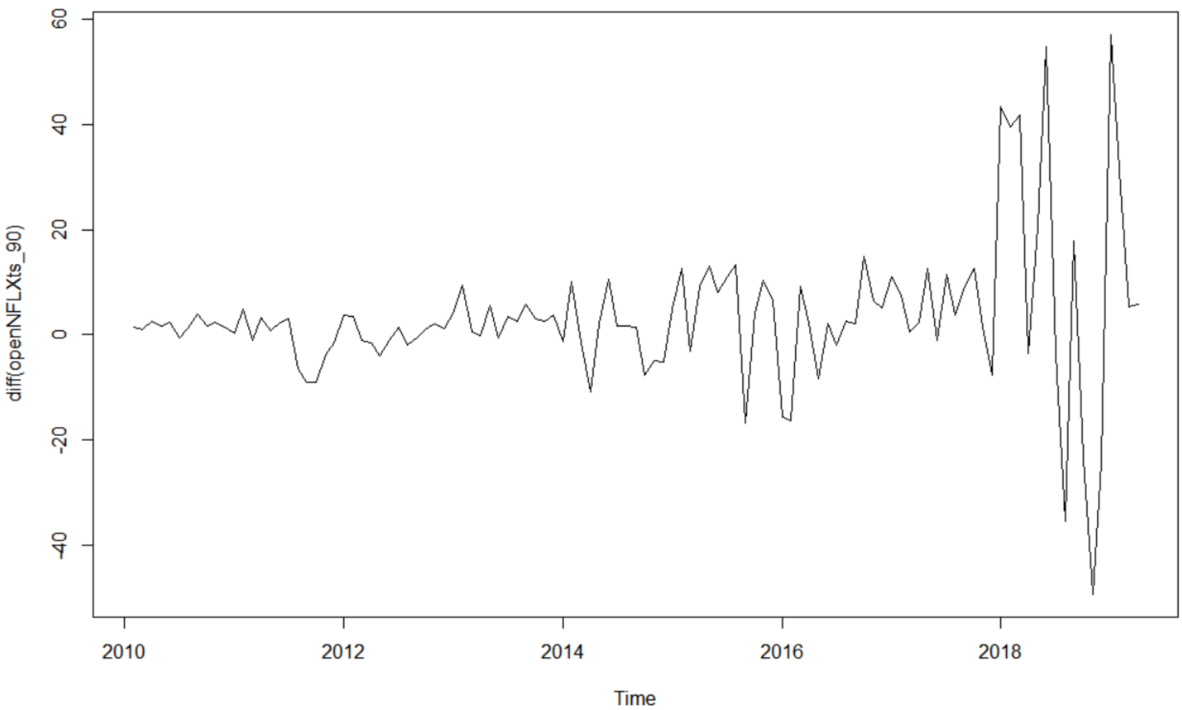
- Trend accounting for approximately 350 units
- Seasonal accounting for approximately 14 units
- Noise accounting for approximately 100 units

Thus, there is little seasonality in the data, it is mainly made up of trend and some noise.

Looking at the below plot of the 90% time series with monthly labels show no distinctively clear pattern or seasonal trends.

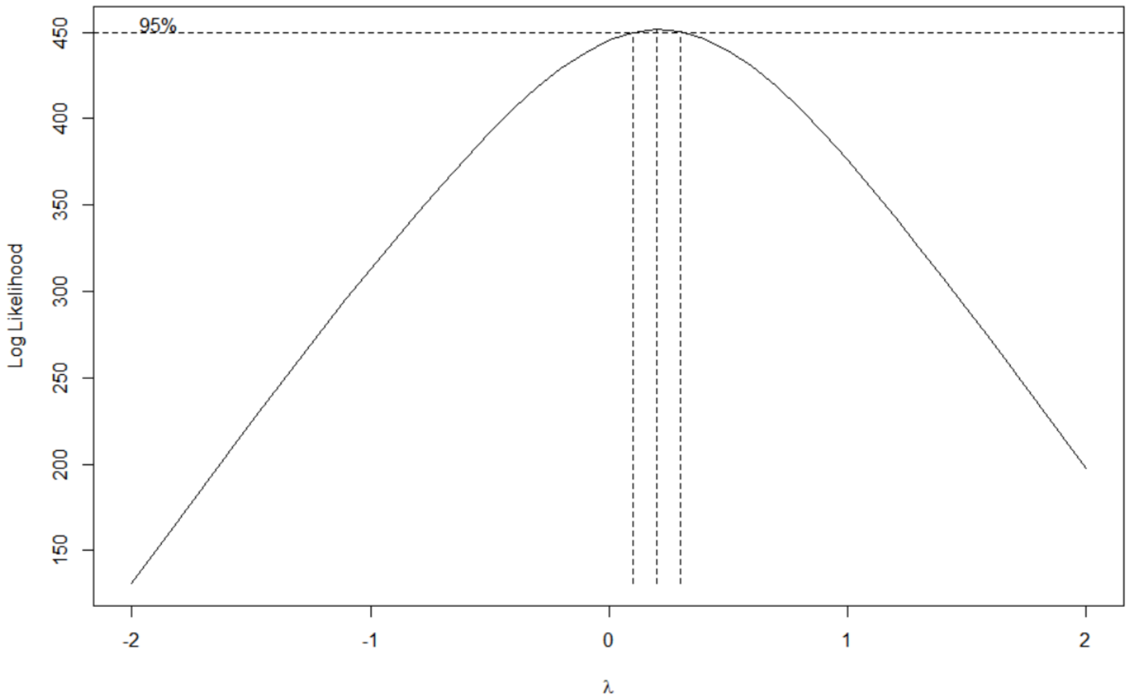


I differenced the 90% time series data to remove the trend in the data and plotted the differenced series, seen below. But as the variance is increasing with time a transformation will be needed.



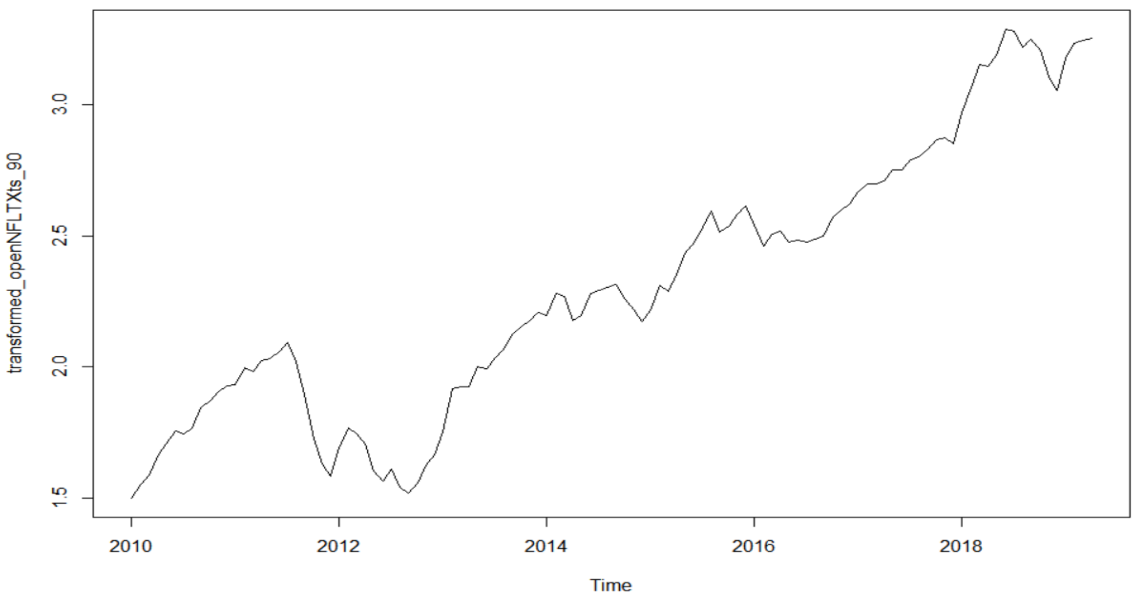
Box-Cox Transformation

I ran a Box-Cox test to decide what transformation, if any, was needed. The result of the Box-Cox test can be seen below, with a score of 1 indicating no transformation is needed. As 1 was not contained in the confidence interval a transformation is need. The 95% confidence interval was [0.1,0.3] and the maximum likelihood estimator (MLE) was 0.2. As there were no common/general transformations in the data, i.e. 0 is a log transform, 0.5 being a square root transform, I used the MLE for the transformation.



Transformed 90% Time Series Data

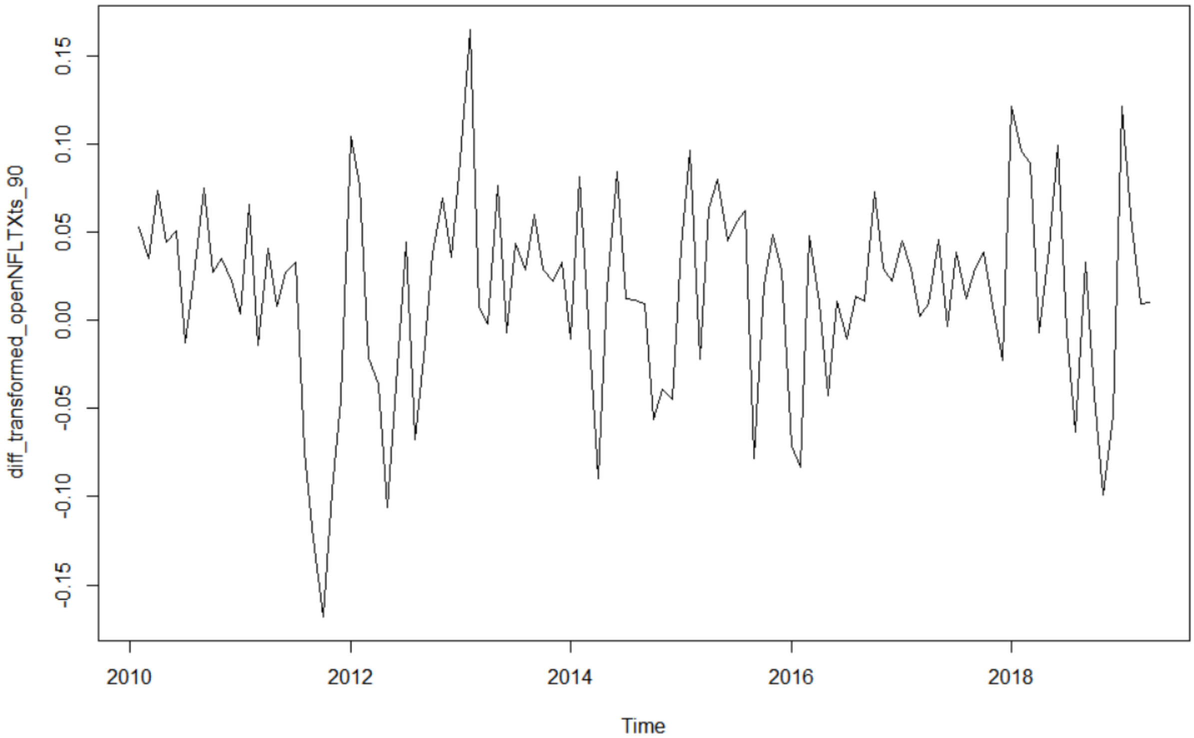
A plot of the transformed data, $x^{0.2}$, is below. There is an increase in the variance at the start of the plot, but the variance is no longer as prevalent at the tail end of the plot.



However, since there is still an upward trend differencing is required to remove this trend.

Differenced Transformed 90% of the Time Series Data

After differencing the transformed data, we get the below plot, which looks to have approximately constant mean and variance.



Performing an Augmented Dickey-Fuller test on the differenced transformed 90% time series data, I got a p-value of 0.01. As the p-value is significant, the data is stationary.

4. Deciding on Models

A plot of the acf, pacf, eacf and arma subsets of the differenced transformed 90% time series data is shown below.

From the ACF: as the model cuts out after 1 lag and there are no other significant autocorrelations, this might suggest an MA(1) model.

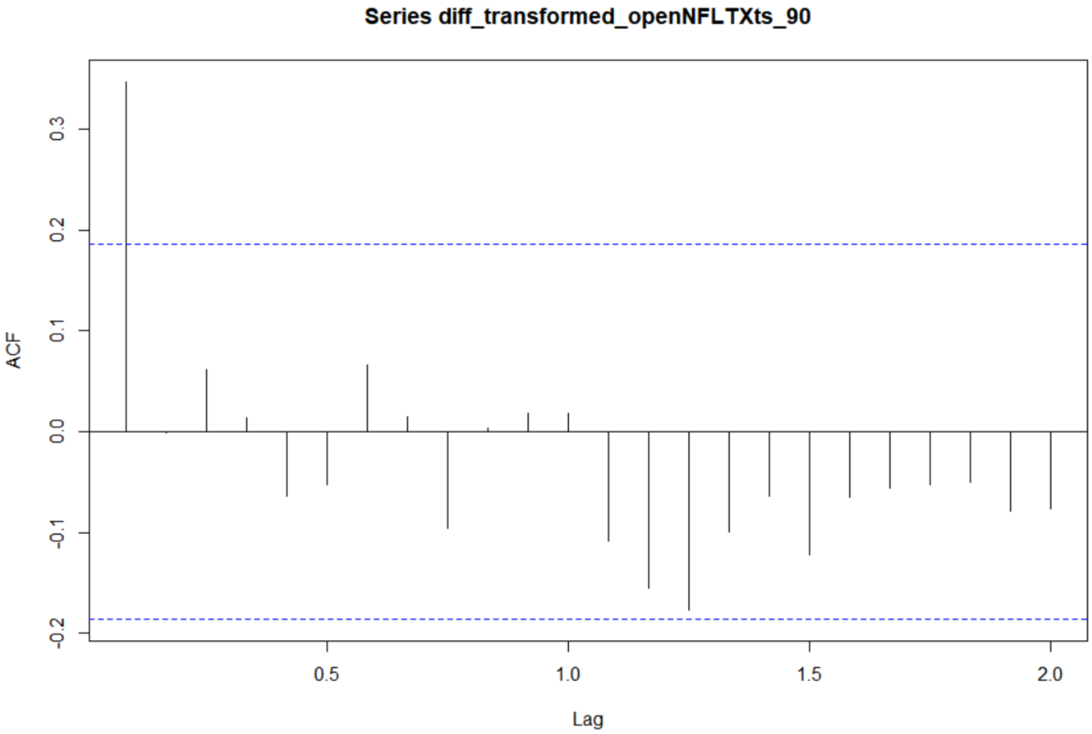
From the PACF: as the model cuts out after 1 lag and there are no other significant autocorrelations, this might suggest an AR(1) model.

From the EACF: the furthers left most upper zero that creates the triangle shape, would suggest an MA(1) model. As there is no "wall" of x's in the 12th column or any adjacent columns this would suggest there is no significant seasonal correlations.

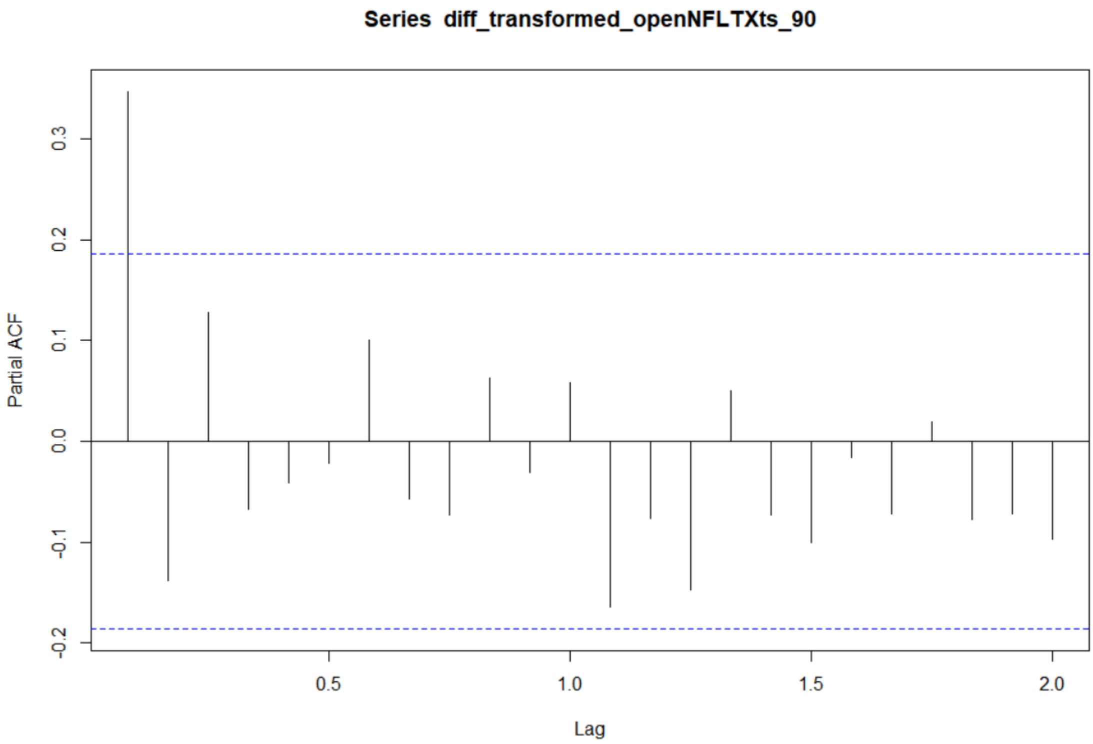
From the arma subsets: the following models are all within ±2 scores of the lowest BIC score, MA(24), ARMA(1,1) and ARMA(15,1).

So potential models are; MA(1), AR(1), MA(24), ARMA(1,1), ARMA(15,1)

ACF of Differenced Transformed Data



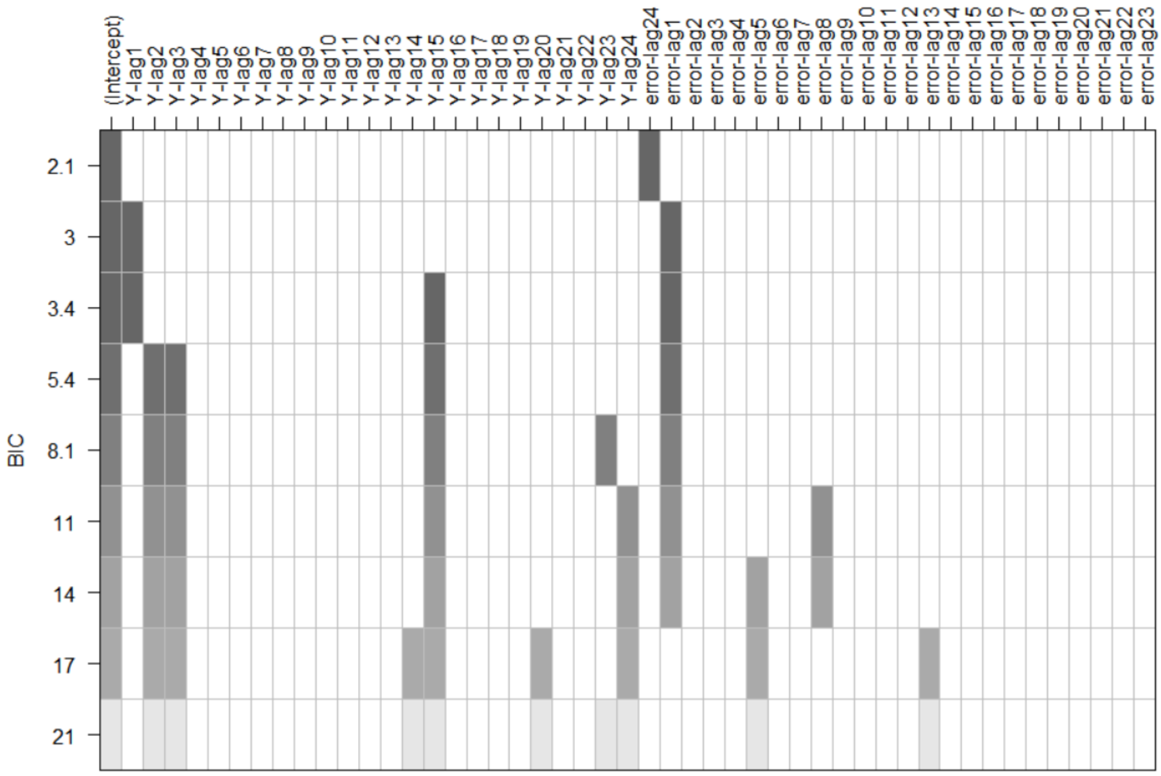
PACF of Differenced Transformed Data



EACF of Differenced Transformed Data

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
2	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
3	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
4	x	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
5	x	o	x	x	o	o	o	o	o	o	o	o	o	o	o	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	o	o	o	o	o	o	o	o	o	o	o
7	x	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
8	x	x	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
9	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
10	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
11	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
12	x	o	x	x	x	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
13	x	x	x	x	o	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
14	x	x	x	x	o	o	o	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
15	x	x	x	x	x	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
16	x	x	x	o	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
17	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
18	o	x	o	o	x	o	o	o	o	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o
19	o	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
20	o	x	o	o	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
21	o	x	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
22	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
23	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
24	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o

Arma Subsets of Differenced Transformed Data



5. Fitting the Models to the Data

IMA(1,1)

	AIC	MA(1) Coefficient	Standard Error
arima (0,1,1)	-337.31	0.4672	.0875

ARI(1,1)

	AIC	AR(1) Coefficient	Standard Error
arima (1,1,0)	-334.44	0.3937	.0869

IMA(24,1) Only putting in statistically significant coefficients

	AIC	MA(1) Coefficient	Standard Error MA(1)	MA(12) Coefficient	Standard Error MA(12)
arima (0,1,24)	-314.01	0.3937	.0869	0.3543	0.1631

```
Call:
arima(x = transformed_openNFLTXts_90, order = c(0, 1, 24))

Coefficients:
         ma1     ma2     ma3     ma4      ma5      ma6     ma7     ma8      ma9    ma10    ma11    ma12    ma13    ma14     ma15     ma16
      0.5658  0.0018  0.1875  0.0821  -0.1229  -0.0650  0.2075  0.2631  -0.0846  0.0601  0.1291  0.3543  0.0097  0.1192  -0.0962  -0.0758
s.e.  0.1183  0.1574  0.2082  0.1206   0.1981   0.1263  0.1805  0.1623   0.1530  0.1513  0.1402  0.1634  0.1631  0.1287   0.1348   0.1283
        ma17     ma18     ma19     ma20     ma21    ma22     ma23     ma24
      0.0412  -0.0421  -0.0603  -0.1613  -0.1066  0.0775  -0.1330  -0.0726
s.e.  0.1628   0.1418   0.1284   0.1343   0.1534  0.1398   0.1412   0.1500

sigma^2 estimated as 0.002005:  log likelihood = 181.01,  aic = -314.01
```

ARIMA(1,1,1)

	AIC	AR(1) Coefficient	Standard Error AR(1)	MA(1) Coefficient	Standard Error MA(1)
arima (1,1,0)	-335.31	-0.0181	0.2271	0.4823	0.2055

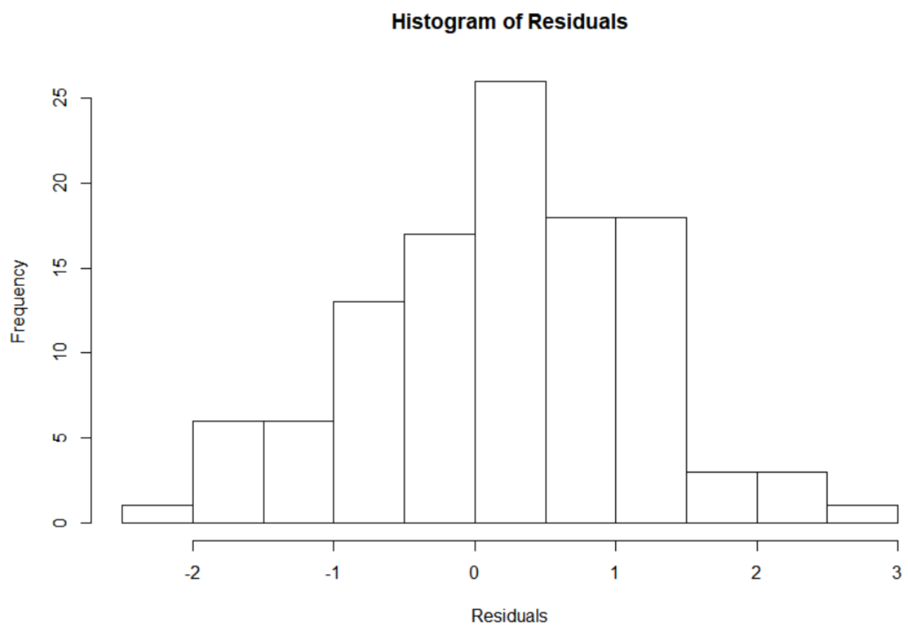
Here the AR(1) coefficient is not statistically significant

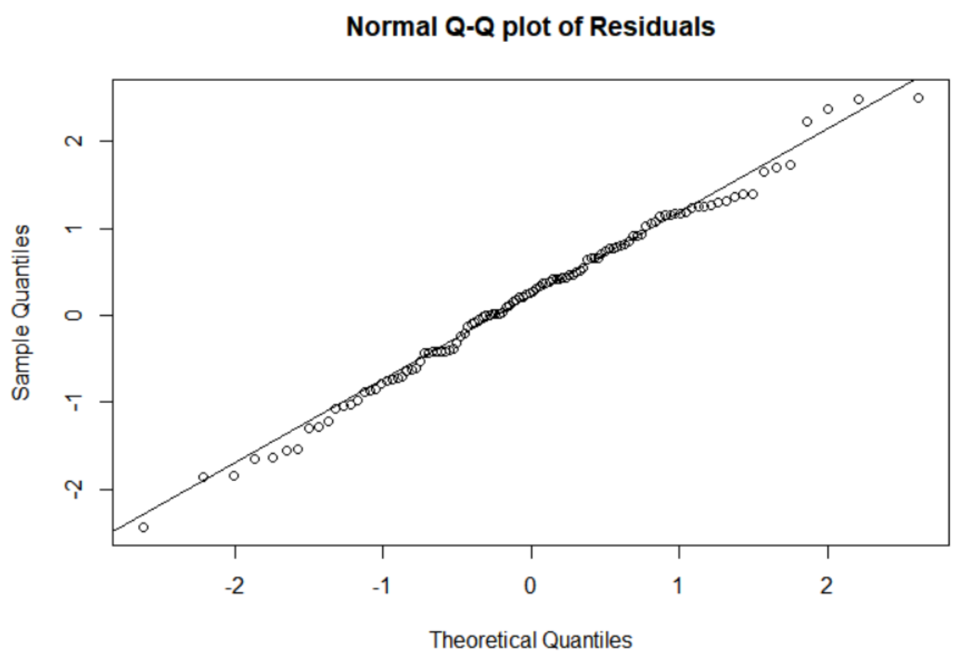
ARIMA(15,1,1)

Yielded an aic of -317.65, but none of the ar or ma coefficients were statistically significant.

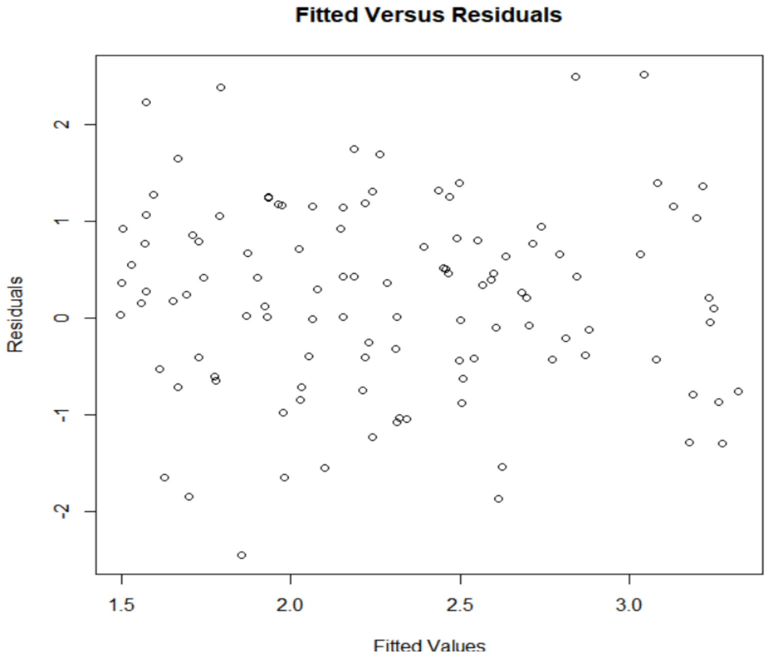
```
Call:
arima(x = transformed_openNFLTXts_90, order = c(15, 1, 1))

Coefficients:
         ar1     ar2     ar3     ar4      ar5      ar6     ar7     ar8      ar9    ar10     ar11    ar12     ar13     ar14     ar15     ma1
      0.0412  0.0066  0.1307  0.0055  -0.0215  -0.0309  0.1171  0.0911  -0.1147  0.0908  -0.0079  0.1443  -0.0921  -0.0363  -0.1449  0.4299
s.e.  0.3901  0.2040  0.1158  0.1210   0.0990   0.0939  0.0957  0.1066   0.0960  0.1071   0.1116  0.0997   0.1138   0.1153   0.0987  0.3878

sigma^2 estimated as 0.00248:  log likelihood = 174.83,  aic = -317.65
```

Residual checks of IMA(1,1)

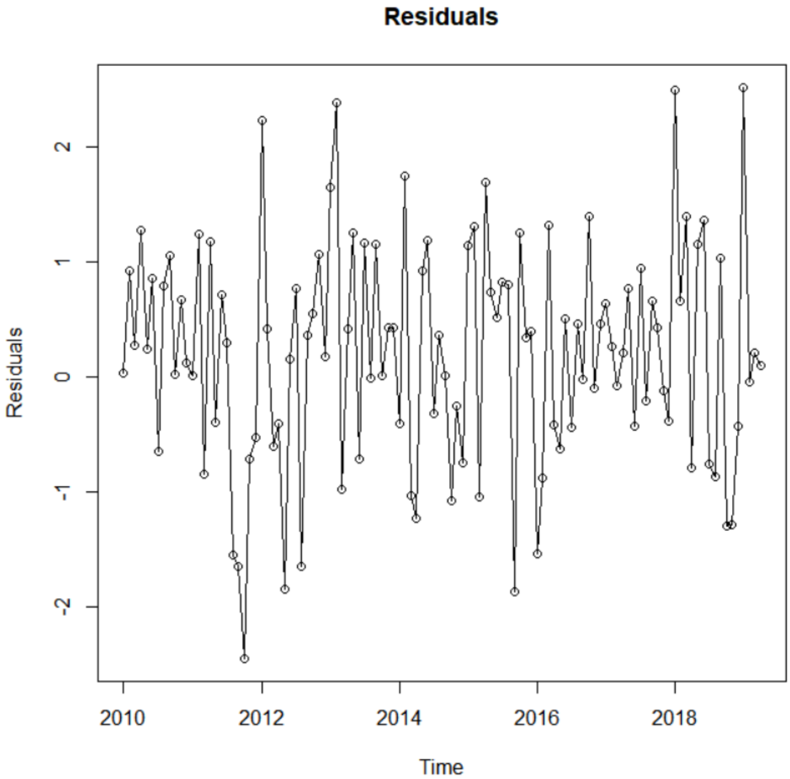




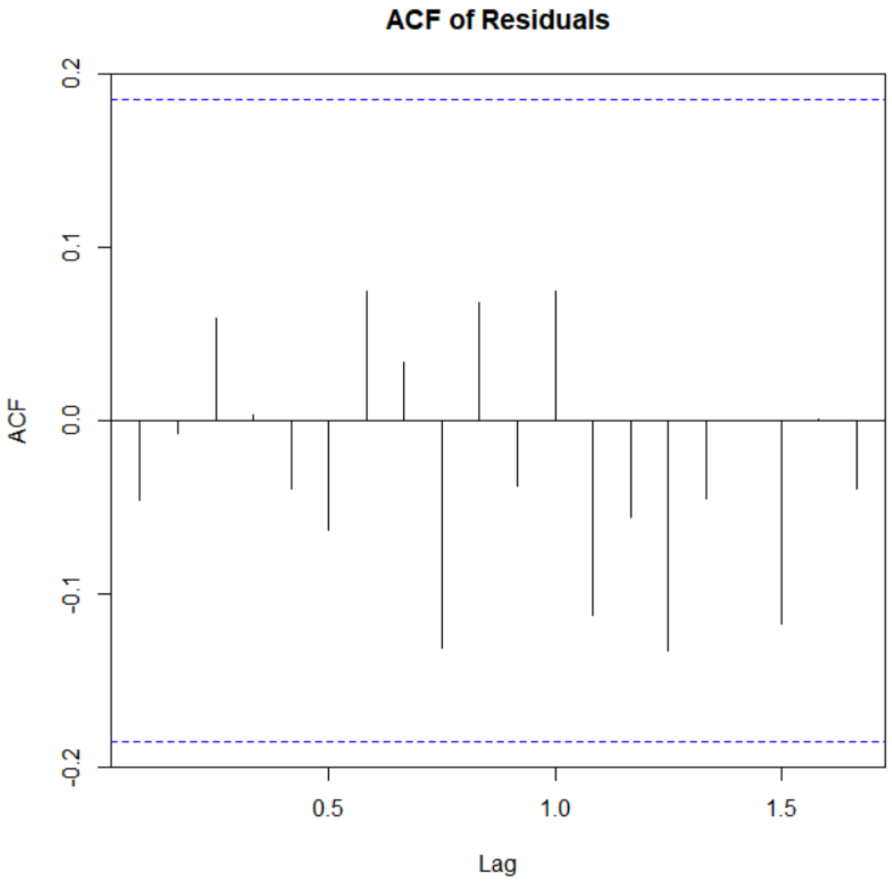
Looking at both the histogram and the q-q plot, the residuals look approximately normally distributed. Running a Shapiro-Wilk normality test yielded a p-value of 0.7336, so this supports that the residuals are approximately normally distributed.

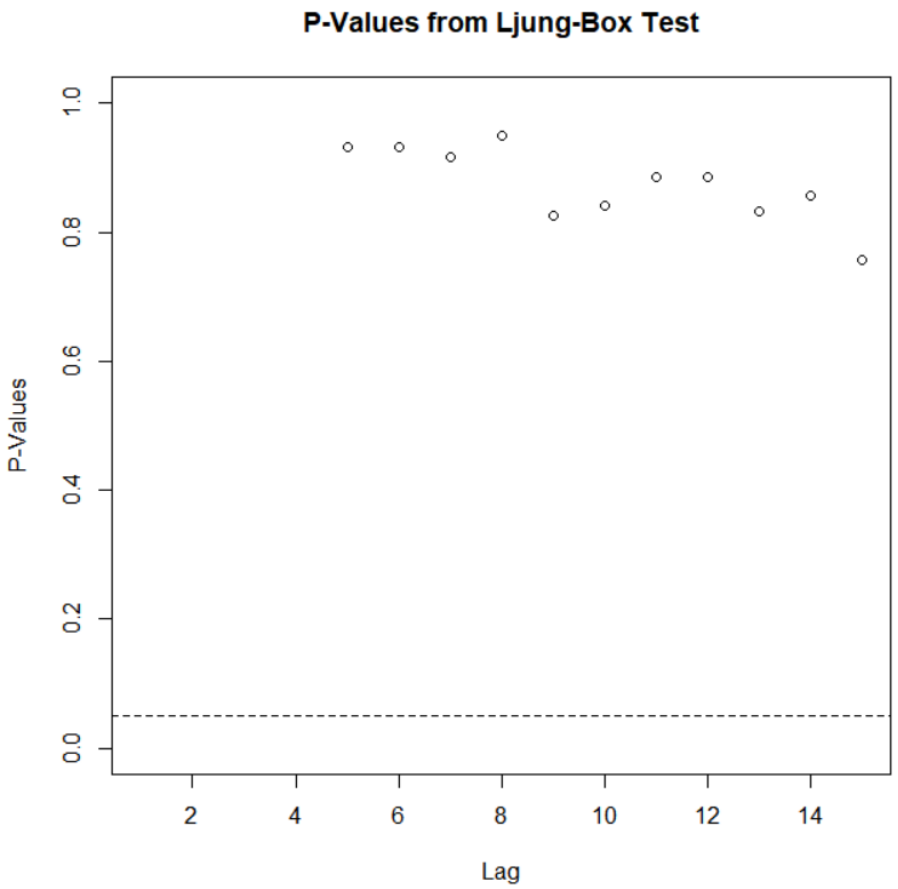


For the fitted values vs the residuals plot, the points look relatively randomly scattered along with no evidence of clear patterns or non-constant variance.



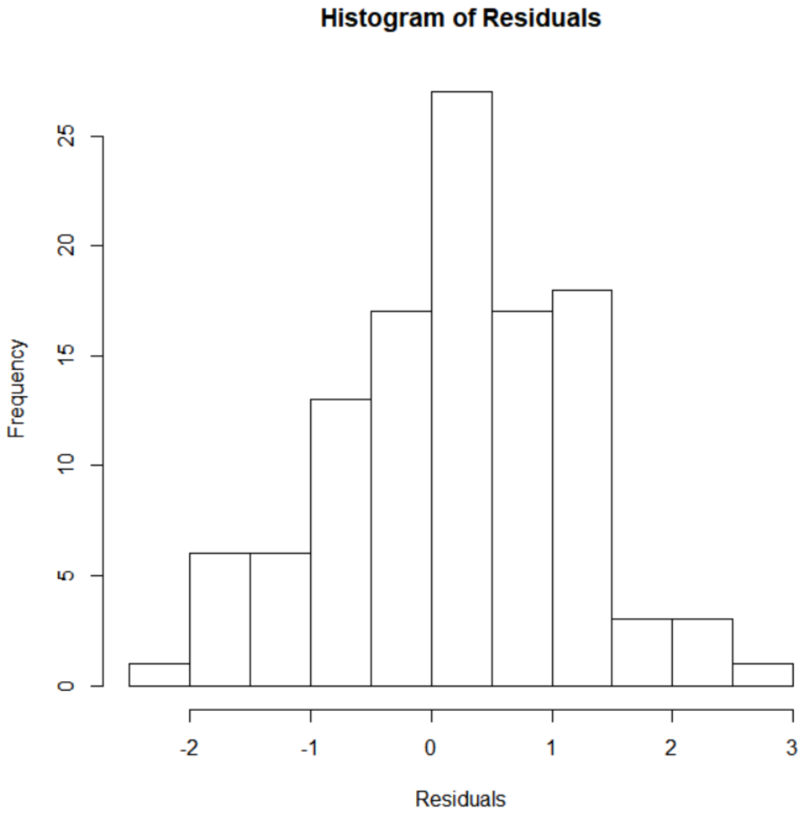
Looking at the plotted residuals, it looks to have approximately constant variance.

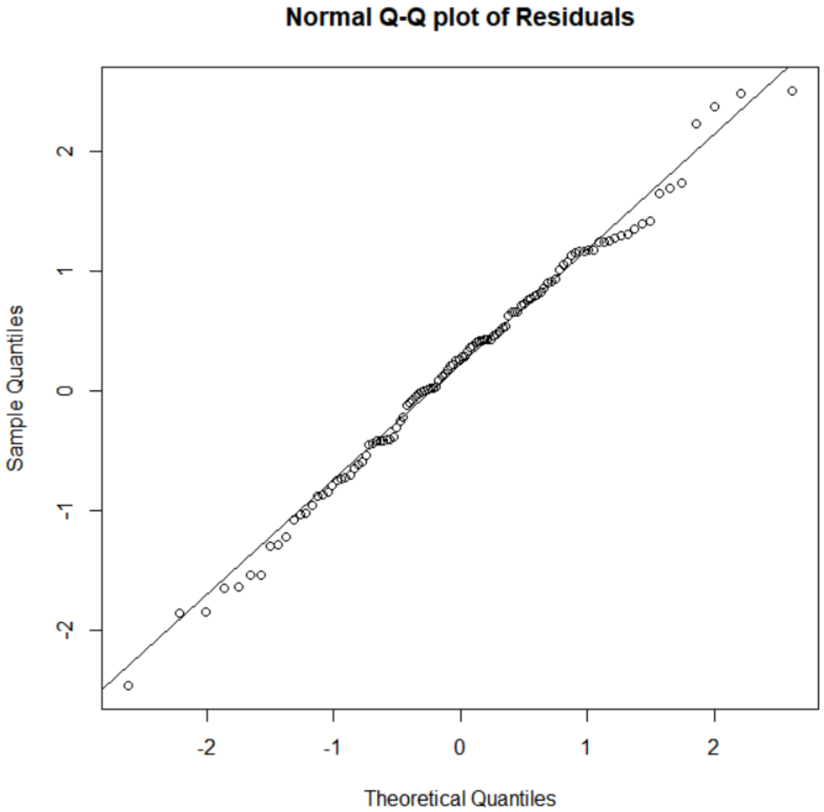

ACF of Residuals
ACF
0.2
0.1
0.0
-0.1
-0.2
0.5
1.0
1.5
Lag


P-Values from Ljung-Box Test
P-Values
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
12
14
Lag

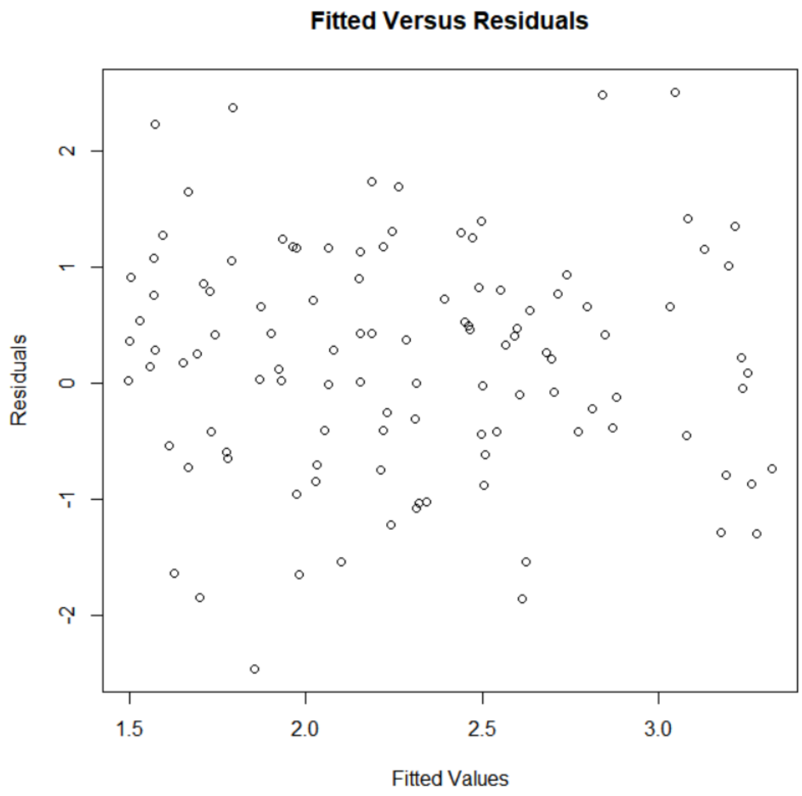
From the above Ljung-Box plot, the p-values for the Ljung-Box test are all above 0.05, this tells us there are no significant autocorrelations, and that the residuals are white noise.

Residual checks of ARIMA(1,1,1)

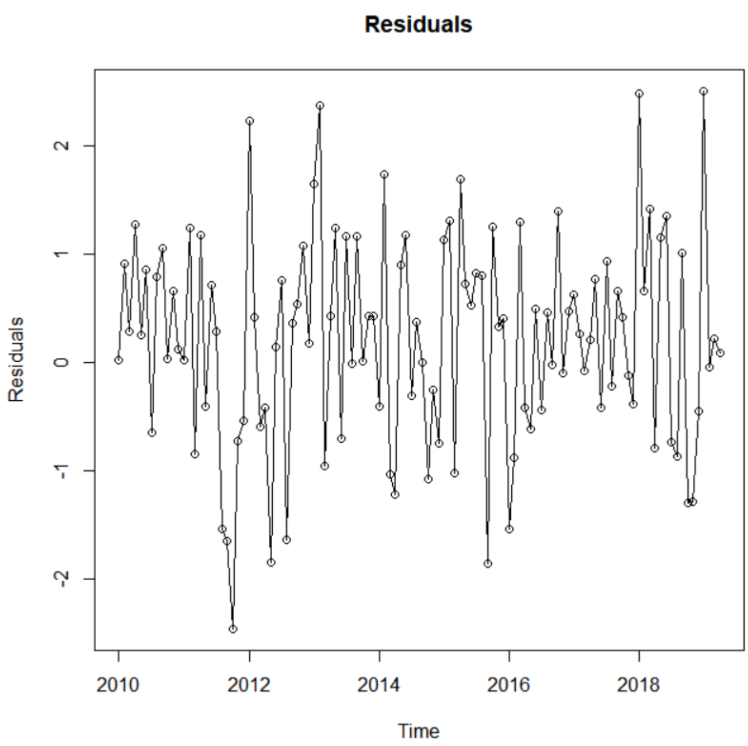




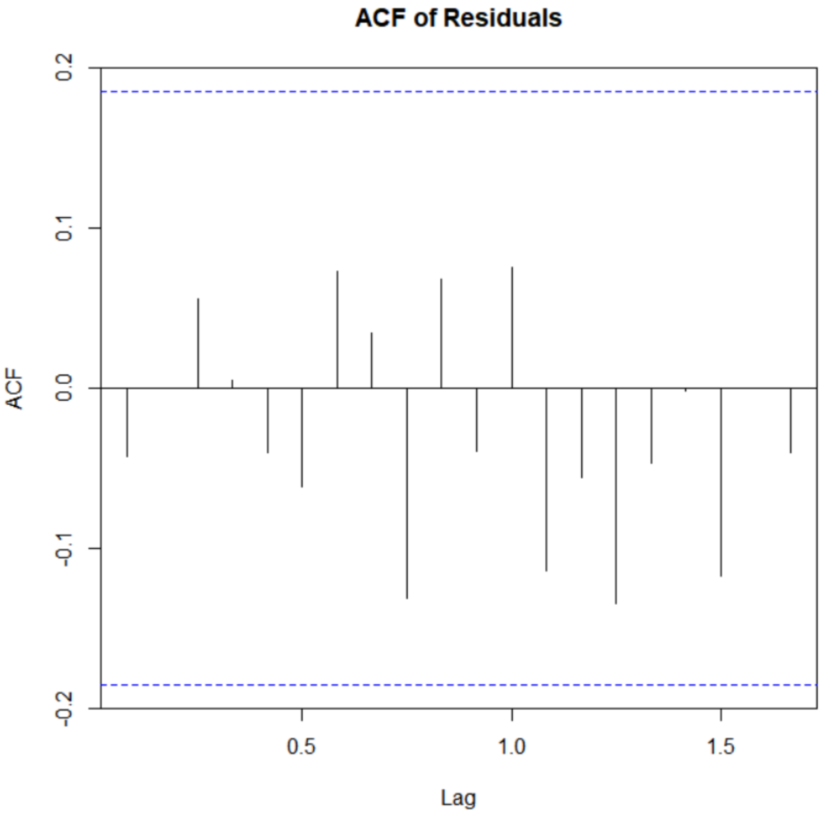
Looking at both the histogram and the q-q plot, the residuals look approximately normally distributed. Running a Shapiro-Wilk normality test yielded a p-value of 0.7357, so this supports that the residuals are approximately normally distributed.

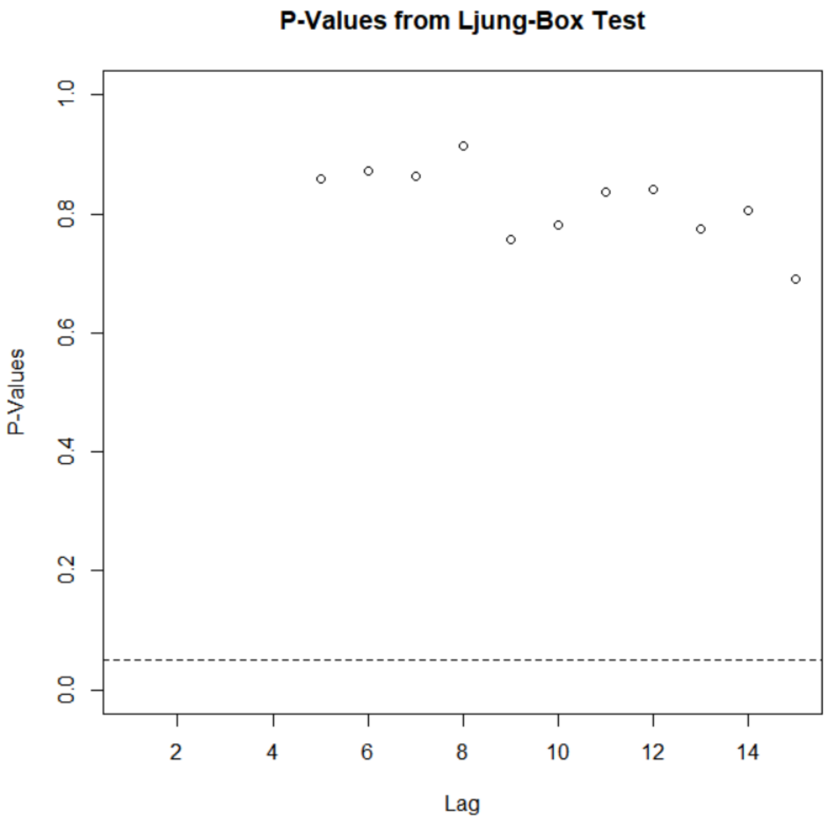


For the fitted values vs the residuals plot, the points look relatively randomly scattered along with no evidence of clear patterns or non-constant variance.



Looking at the plotted residuals, it looks to have approximately constant variance.





From the above Ljung-Box plot, the p-values for the Ljung-Box test are all above 0.05, this tells us there are no significant autocorrelations, and that the residuals are white noise.

Conclusion

So the potential models are; IMA(1,1) and the ARIMA(1,1,1) as they are within ±2 scores of the lowest aic score. They both have approximately normally distributed residuals and constant variance. However, as the ar coefficient is not statistically significant in the arima(1,1,1) and both models are within ±2 scores of the lowest aic score, the IMA(1,1) is chosen as it's the most simplistic model.

Final Model

The final model was the IMA(1,1) which is seen below:

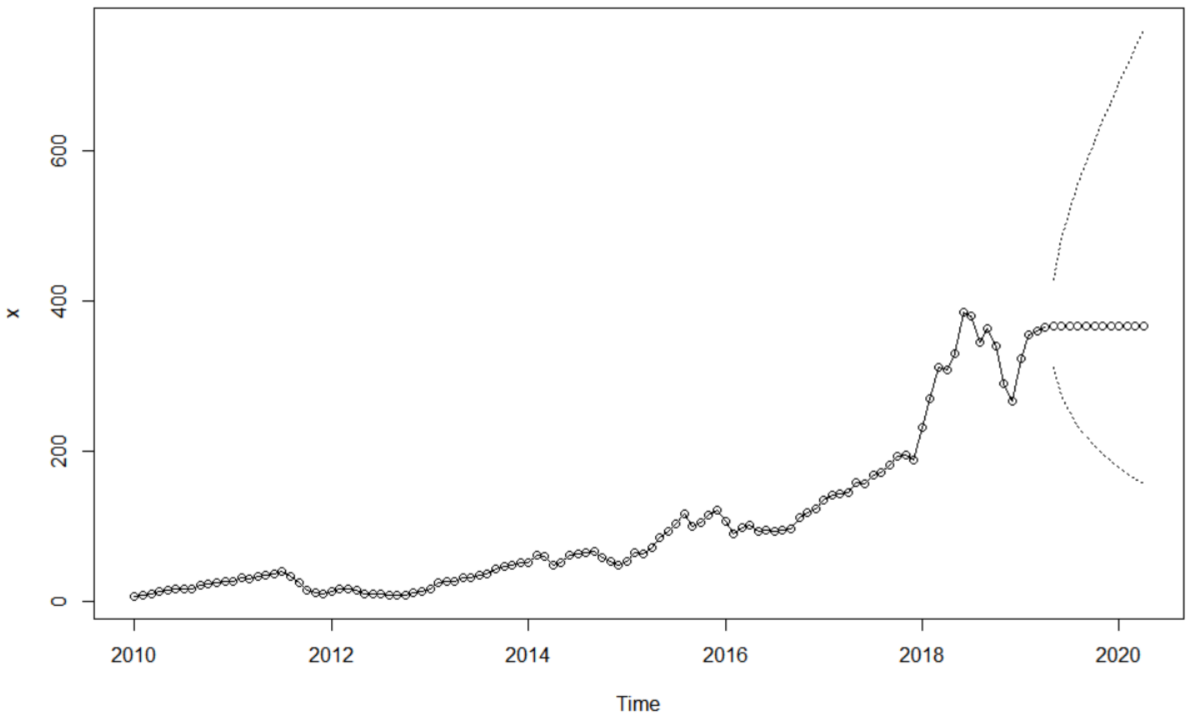
IMA(1,1)

	AIC	MA(1) Coefficient	Standard Error
arima (0,1,1)	-337.31	0.4672	.0875

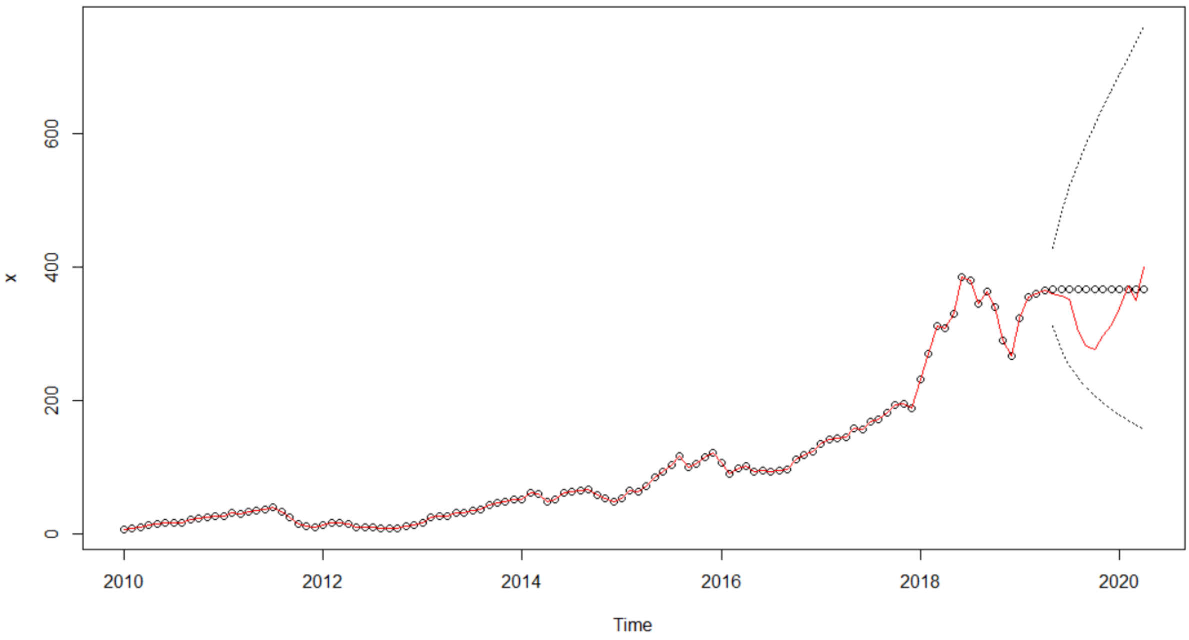
$$Y_t = Y_{t-1} + e_t + 0.4672e_{t-1}$$

6. Prediction

Predicating the next 12 months using the IMA(1,1) model yields the below, with the transform of x^5 to bring the data back to the original scaling. As I used an IMA(1,1) model, the predications are expected to revert back to the process mean after one step, which can also be seen below.



A plot of predicated values with the actual values, in red, is shown below.



There are some obvious differences between the predicated values and the actual values. A table of the predicted values minus the actual values is given below.

May 19'	June 19'	July 19'	Aug 19'	Sept 19'	Oct 19'	Nov 19'	Dec 19'	Jan 20'	Feb 20'	Mar 20'	April 20'
7.36	9.96	15.59	62.88	84.54	90.19	69.36	52.67	28.99	-5.09	17.41	-33.06

Conclusion

As so many factors influence stock prices, the differences between the predicted values and the actual values don't surprise me. More complex models that account for more than just historical stock prices would be needed to predict with more accuracy. An explanation for the dip in the stock price for August 19' and the following months is that Netflix missed its subscriber growth guidance for the previous two quarters, which was more than likely due to a price increase in the spring of 2019. In November 19' Apple released their own premium video platform along with Disney launching its Disney+ platform, both Apple and Disney were at price points well below Netflix's. Looking only at historical stock prices for Netflix means we can not account for the new competition in the market and how it would affect the stock price for Netflix. So, in short, Netflix's stock price is not highly predictable by looking at its historical stock price alone.

Cathaoir Agnew 16171659