**Forecast Time Series Data Project1**

**Alison Yao (yy2564)**

**Data Description**

This project uses the monthly stock price of Netflix from 2002-06-01 to 2022-03-01 (238 observations) for time series forecast. The 2022-03-01 data is the latest entry because I obtained the data in late March from [Yahoo Finance](https://finance.yahoo.com/quote/NFLX/history?period1=1022112000&period2=1647907200&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true). By the time this report is graded, there may or may not be data for 2022-04-01, but I am not able to obtain it at the time being.

The head of the table is shown below:

Table

Description automatically generated

**Time Series Plot**

First, we plot the time series plot of adjusted closing stock price. The Adj Close stock price grows exponentially over time as a general trend. Surprisingly, while Adj Close did not drop during COVID time (probably because streaming was a favorable entertainment during lockdown), it is declining sharply in recent months, starting from 2021-10-01.

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

There is strong evidence of level-dependent volatility, as the stock price is relatively stable in early years but much more volatile in recent years, which is especially easy to see if we take the difference of Adj Close price. Therefore, we take the natural log of adjusted closing stock price and plot another time series plot.

Chart, line chart

Description automatically generated

After taking log, we get rid of level-dependent volatility and linearize the exponential trend at the same time.

**Model Selection**

Now, for the first step of choosing an ARIMA model, we need to determine if we should difference the data and how many times should we difference the data if we need to. We can look at the ACF plot of LogAdjClose.

Chart

Description automatically generated

The hanging behavior is strong evidence that we should difference the data. Differencing once gives us:

Chart

Description automatically generatedChart

Description automatically generated

Differencing twice shows strong evidence of over-differencing as lag-1 is significantly negative and close to -0.5.

Chart

Description automatically generated

Therefore, we conclude that we only need to difference once to make the data stationary, so we choose in our ARIMA model.

As for the values of p and q, we can check the ACF and PACF of DifLogAdjClose, both of which are statistically significant at lag 1. Therefore, we choose p,q = 0,1 and test the following models:

We know N = n – d = 238 – 1 = 237.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Without Constant | | | | | With Constant | | | | |
| p | d | q | SS | AICc | p | d | q | SS | AICc |
| 0 | 1 | 0 | 6.10529 | -865.14335 | 0 | 1 | 0 | 5.95598 | -868.97718 |
| 1 | 1 | 0 | 5.96747 | -868.52041 | 1 | 1 | 0 | 5.85750 | -870.87692 |
| 0 | 1 | 1 | 5.97965 | -868.03717 | 0 | 1 | 1 | 5.86326 | -870.64398 |
| 1 | 1 | 1 | 5.94664 | -867.29740 | 1 | 1 | 1 | 5.85116 | -869.06418 |

The ARIMA(1,1,0) model with constant yields the smallest AICc -870.87692, so we choose ARIMA(1,1,0) with constant for all following procedures.

**Estimation**

The Minitab output for ARIMA(1,1,0) with constant is:

Graphical user interface, text, application, email

Description automatically generated

Both the AR1 parameter and the constant are statistically significant, with p-values less than 0.05.

If we denote as the time series of Adj Close, as LogAdjClose and as DifLogAdjClose. LogAdjClose fits the ARIMA(1,1,0) model with a constant. The best estimate of the AR1 coefficient is 0.1286 and that of the constant is 0.0218. Therefore, the fitted model is where .

**Diagnostic Checking**

The output of the Ljung-Box test is:

Graphical user interface, text, application, email

Description automatically generated

Since all 4 p-values are greater than 0.05, we fail to reject the null hypothesis that the model is inadequate. Therefore, we conclude that the model is adequate.

The plot of the residuals is as follows:

Chart, scatter chart

Description automatically generated

The residuals look quite random, with only a few with absolute values over 0.5.

The ACF and PACF of the residuals are as follows:

Chart

Description automatically generatedChart

Description automatically generated

The residuals look uncorrelated. The plots do not indicate much inadequacy in the model.

**Forecasting**

The forecasts and 95% forecast intervals for 1-150 are as follows:

Table

Description automatically generatedTable

Description automatically generated Table

Description automatically generated Table

Description automatically generated

Chart, line chart

Description automatically generated

The forecast looks reasonable. We can see that the point estimate shows a linear trend because the model is with a constant and . The forecast interval has a tendency to blow up because LogAdjClose is not stationary.

According to the back-testing at t = 150, 100 and 50, the forecast intervals are wide enough to cover all datapoints, but the forecast intervals seem excessively wide.

Chart, line chart

Description automatically generatedChart

Description automatically generated

Chart, line chart

Description automatically generated

**Conclusions**

Since our best model is an ARIMA(1,1,0) with a constant/drift and the constant term is greater than 0, we conclude that Netflix is a good company to buy and hold. We can do an out-of-sample test for the forecast in the future when we have future data.