D214 Executive Summary- Time Series Project

Jacob Polomsky

School of Technology, College of IT, Western Governors University

Data Analytics Graduate Capstone - D214

Dr. Daniel Smith

August 30, 2024

To What extent can the Closing Price of SPY be predicted? We will test the result of our research question, the null hypothesis, to see if we can predict the closing price of SPY with a 90% accuracy. When we make a prediction, we have our predicted values and confidence intervals with our standard error. Our goal for the null hypothesis is to see if 90% of the predicted data can fall into at least two standard deviations from predicted values. Ideally, if 90% of the predicted values fall within one standard deviation away, that would be even better.



A screenshot of a graph

Description automatically generated

This model shows a tailing off ACF, and the PACF cuts off, indicating that it is an AR model.

For finding if the data is stationary for d, with the plot below, we find that the p-value is above 0.05, indicating that this data is non-stationary.



A white background with black text

Description automatically generated

Checking for seasonal components we run a decomposition plot

A close up of a text

Description automatically generated

A graph of different types of waves

Description automatically generated with medium confidence

We see an increasing trend over time, but towards the end of the plot, it starts to trail down; the seasonal component is very repetitive, indicating a seasonal correlation with the plot.

 We use the spectrum plot to see if the data is stationary or not stationary.

A graph of a graph

Description automatically generated

Model 1 is first choosing a 1 0 0 0 1 0 12 model because we have an AR model due to the ACF and PACF plot, and D of 1 because of the seasonality of the decomposition plot.A close up of text

Description automatically generated



A computer screen shot of a number

Description automatically generated



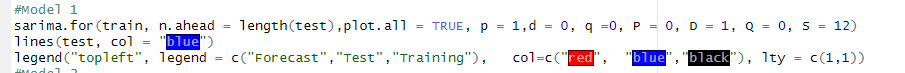
A screenshot of a graph

Description automatically generated



A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated



A graph with a line and a circle

Description automatically generated with medium confidence

Model 2  I did the same thing as model one but noticed that the residual for ACF had a big residual at 12, showing that there might be a missing seasonal component, so I added a Q of 1A white background with black text

Description automatically generated

A screenshot of a computer

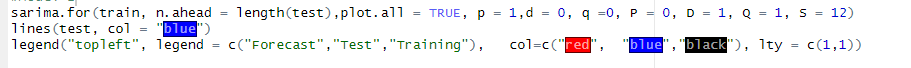
Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated



A graph with a line pointing to the top

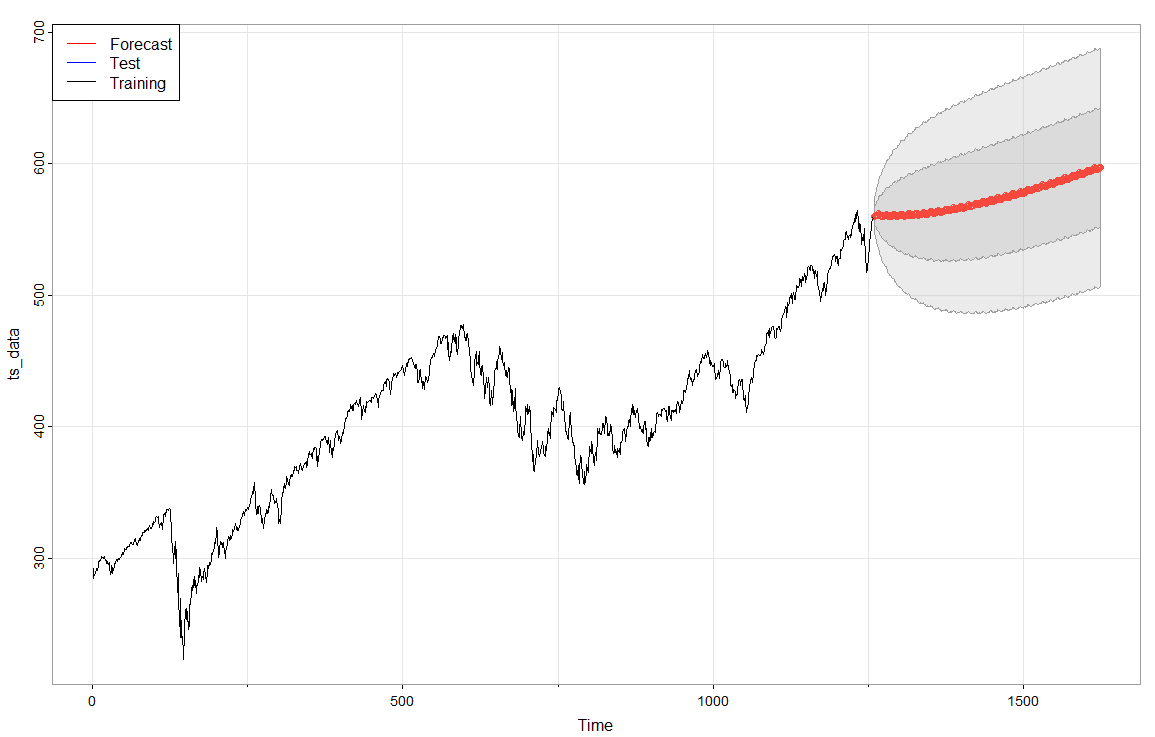
Description automatically generated with medium confidence

For the first model, I went with a 1 0 0 0 1 0 12 model because of the ACF trailing off and PACF cutting off, giving a p of 1 and a q of 0. I went with a seasonal D of 1 instead of d because, in the decomposed plot, the chart showed a seasonal pattern. I plotted it with both D and d, but it threw off the analysis. With the second model, I noticed that Lag 12 had a big stick when ACF residual indicated a possible seasonal component we needed to include. So I did a 1 0 0 0 1 1 12 model, and it looks like the model performed better.

With the findings of model one, we could predict the entire test set within one confidence interval away, but in the second model, the test set fell into the second confidence interval towards the end, with the rest falling into the first confidence interval. By almost every accuracy metric, the second model performed better than the first model. It has a lower AIC, BIC, RMSE, MAE, and MAPE. The confidence interval was narrower.

 The limitation is that the data did not predict where the test set would be for the first and second models. The first model had a very high confidence ratio, showing it could follow the trend. The second model was narrow but had issues following where the test set ran into a higher confidence interval.

 Model 2performed well and would be good enough to aid traders in determining where to expect the price a year from now. The model itself is not good enough to make trades. If you want to do that, the confidence interval needs to be narrowed by doing this model with another stock or using a hybrid model that can factor in other features that allow for better predictions.



The study's benefits are that it will help traders get a general trend of the market and general expectations of where the market will be from now. It lets traders know that the market is increasing, so placing calls for options or buying the stock is a good idea and will most likely lead to profit. It also looks like the market stays stagnant for a bit, then rises pretty fast, and it happens twice: the market grows quickly, then goes back down, then rises quickly, then goes back down. With the prediction, it doesn't go down but stays stagnant, and then the price looks like it goes up quickly. This prediction analysis allows the traders to keep that in mind when they want to place a trade.