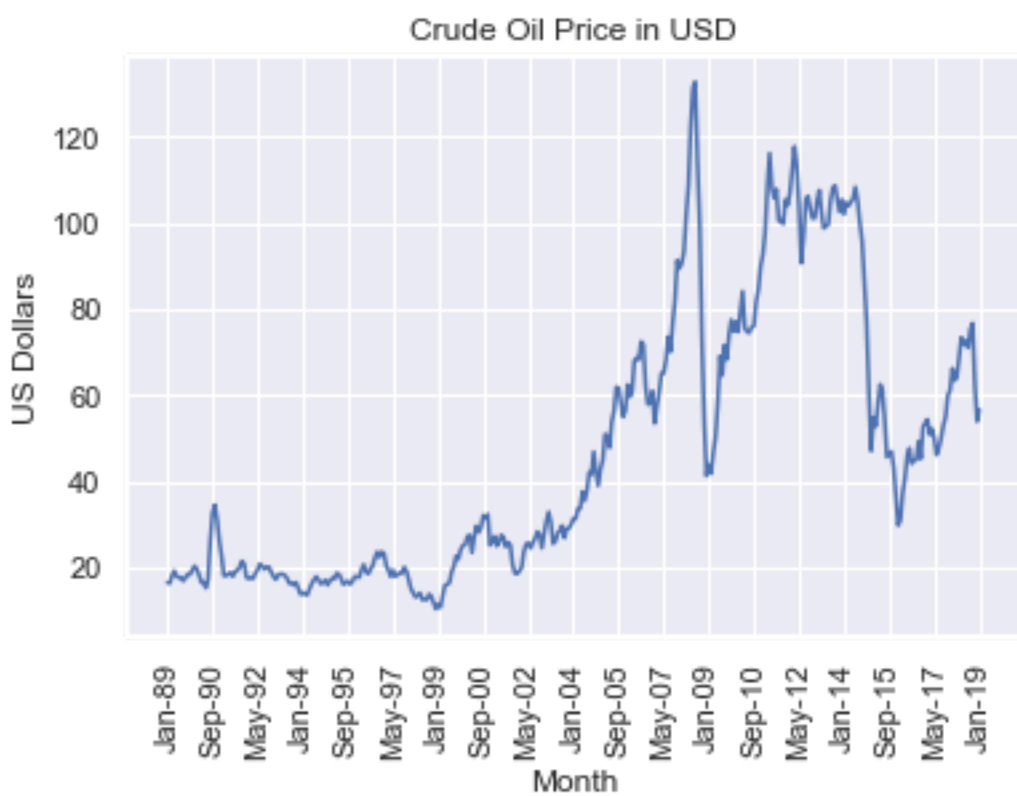
Introduction

In this project, crude oil prices in the United States was forecasted. Data was obtained from the URL https://www.indexmundi.com/commodities/?commodity=crude-oil&months=60. The univariate data are monthly and from January 1989 to January 2019. The AR, MA, and ARIMA models were used for forecasting purposes.

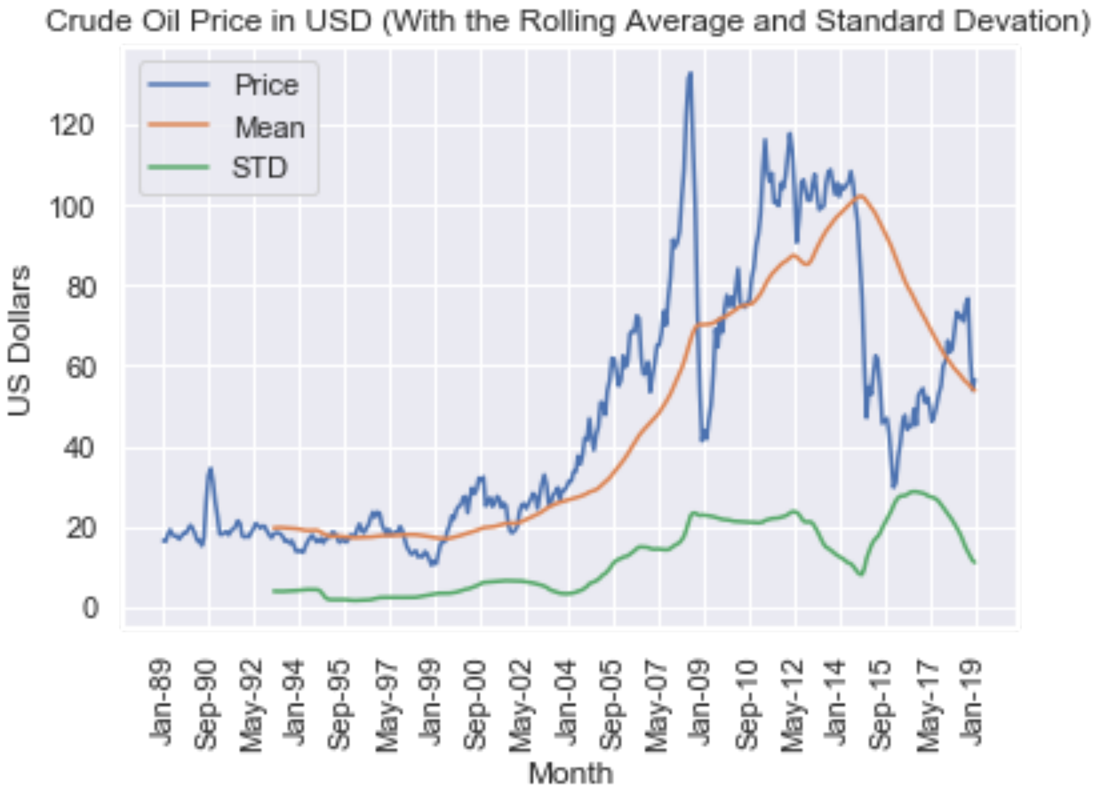
The data were made into a stationary time series set and models were fitted after diagnostics were performed. The models were evaluated using residual sum of squares. The data were then split into training and testing sets and an ARIMA model was built with the optimized parameters. The average error from the test results was used as a metric to evaluate the ARIMA model. The average error was 6.3%.

Exploring the Data

As aforementioned, the monthly data are from January 1989 to January 2019. Below is a plot of the time series data.



It appears that the mean and variation of the data changes significantly over time, with both increasing as time moves forward. To get a better look at this, we can visualize the rolling mean and the rolling standard deviation. Below is the moving average and rolling stand deviation plotted along with the price.

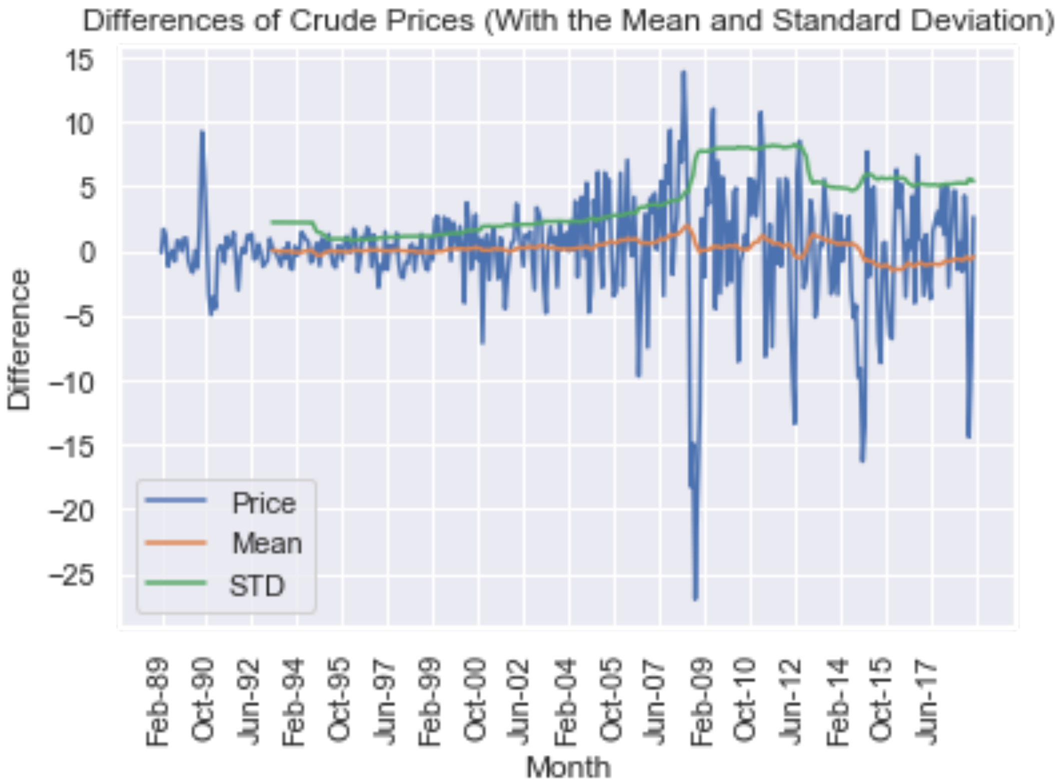


The rolling average and standard deviation do in fact increase as time moves forward. The standard deviation does decrease temporarily in later periods, between May 2014 and late 2014, but it hovers at noticeably higher levels than it does during the early periods. In order to forecast, the data must be stationary, meaning that the mean and variation is not dependent on time. It seems obvious that the data as is is not stationary but to be certain, we can evoke the AD Fuller test. The AD Fuller test gives a test statistic associated with the null hypothesis that the time series data is not stationary and is associated with a model that would have a unit root. A unit root is when present when the price of time period t equals the price of time period t-1 plus an error. Such a model has no predictive ability and its best and only prediction for the price during time period t is the price during time period t-1.

The P value from the AD Fuller test on the base data is 0.23, meaning that the null hypothesis cannot be rejected. Therefore, the data as is cannot be used for forecasting. The data must be put into a stationary state in order to use it for forecasting.

Stationarity and Differencing the Data

While exploring the data, it was discovered that the data is not stationary. The mean and variation increase as time moves forward making them time dependent. Also the price in time period t is the price of time period t-1 plus an error, as mentioned above. If the coefficient of time period t-1 is not one and in between zero and one, then there would be predictive implications and the data would be stationary. In order to make the data stationary, the data is differenced and a new time series of the differences is created. Below are the differences along with the rolling mean and stand deviation plotted.

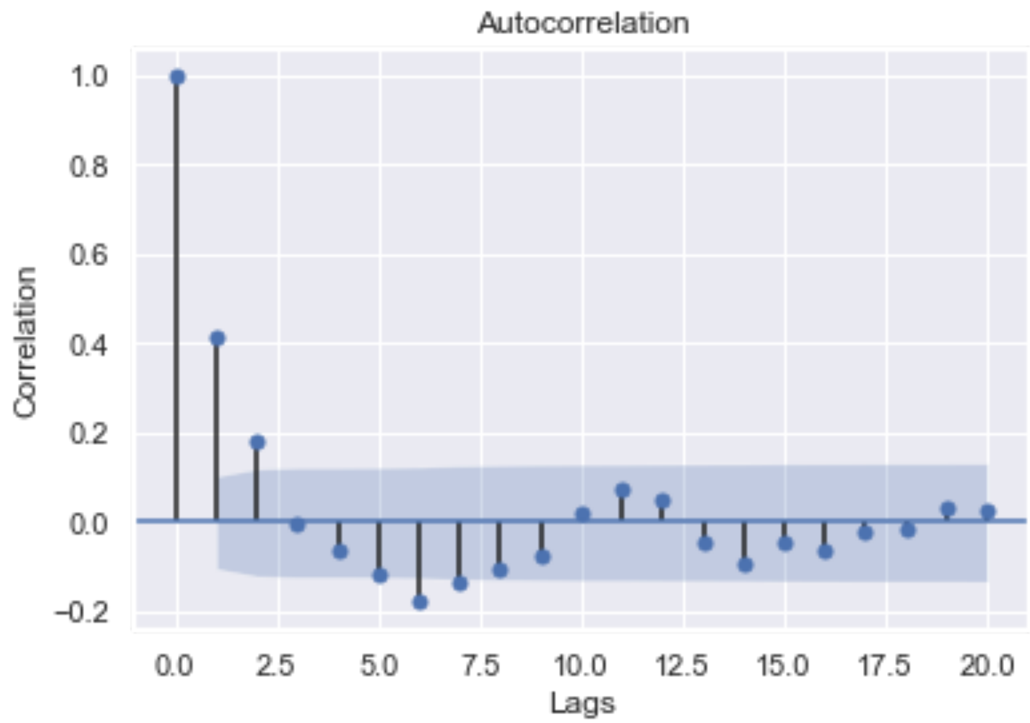


Above it is seen that characteristics of the differenced data differs from the base data, no pun intended. It moves consistently sideways. However, just by eye balling the differences one can notice the increase in variation as time moves forward. The plotted standard deviation illustrates this further with an increase showing between June of 2007 and June of 2012. Though the mean varies where it is obvious (where is the stand deviation increases), it hovers at around the same values throughout. We can again turn to AD Fuller to test the null hypothesis that that data is not stationary and is associated with a unit root to go beyond visual inspection.

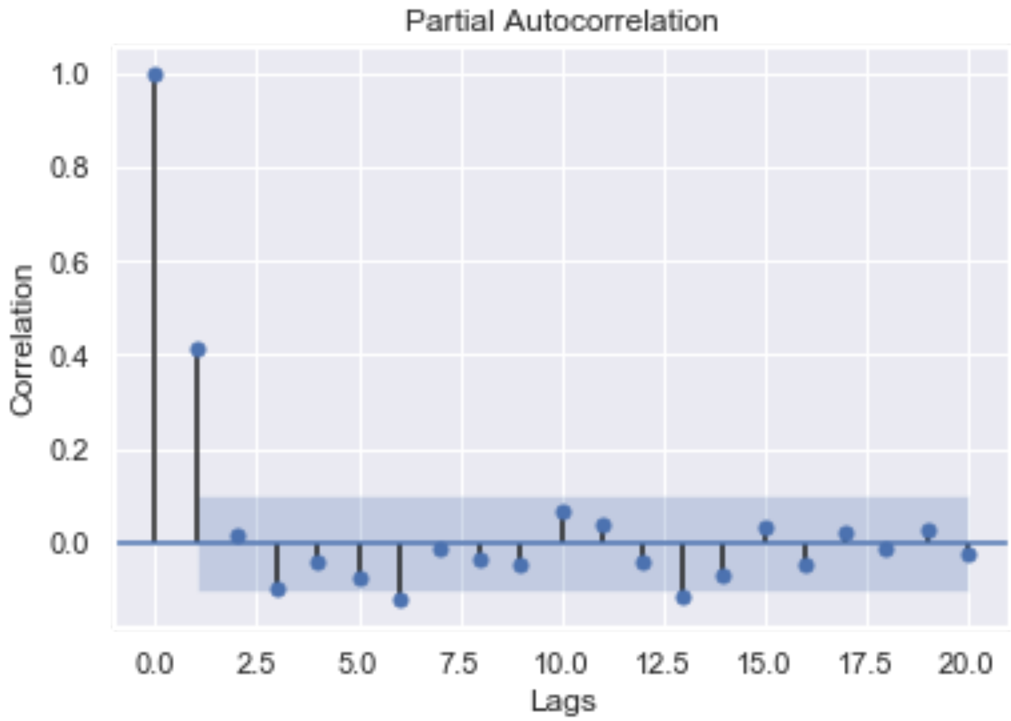
Despite the increase of variation in the differenced data, the P value of the AD Fuller test statistic is 1.4e-14, which makes it highly unlikely that the differenced data is not stationary. Stationarity does not mean that the standard deviation remains constant. It means that the standard deviation is not dependent on time. There is not an apparent trend like association between time and the variation in the differenced data. This confirms that the modified time series data can be utilized for predictive purposes to some extent.

Autocorrelation and Partial Autocorrelation

After making the data stationary, we can focus on working towards making predictions by looking at the correlations between the price in time period t and the prices in time periods prior. This is known as autocorrelation. Below, the autocorrelation of lags one to twenty is plotted.



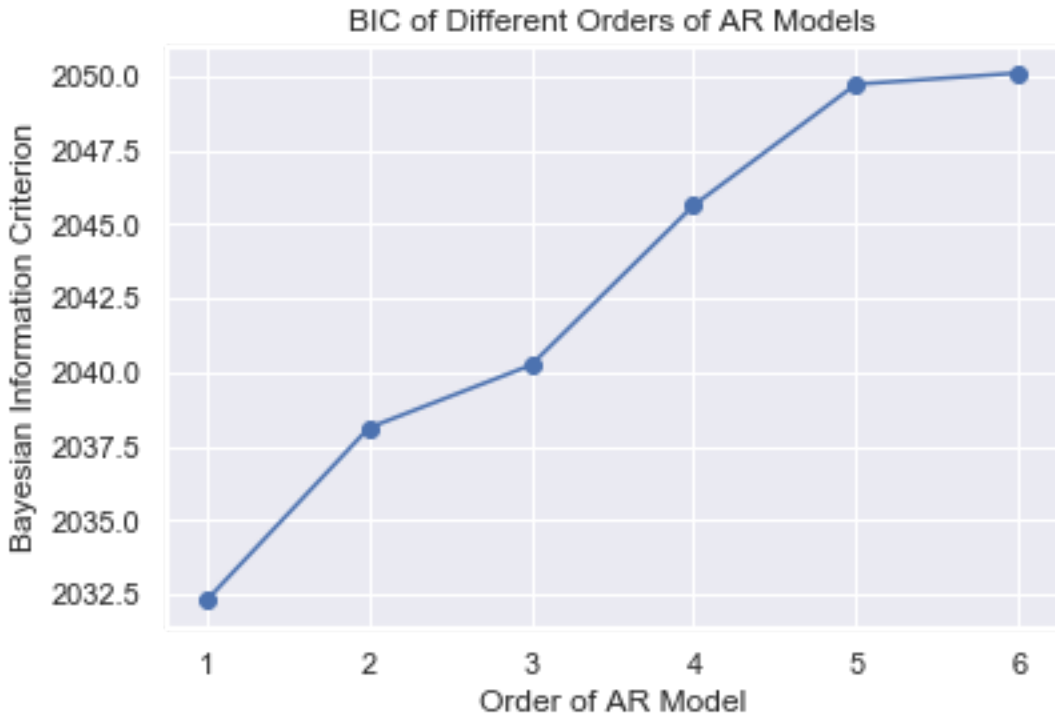
Above we can see that the first and second lags are positively correlated beyond the 95% confidence interval, which is illustrated by the blue shade. Assuming the null hypothesis that the correlation of any given lag is actually zero, the probability of a lag having a correlation outside the 95% confident interval is less than 5%. For this reason, any lag with a correlation outside the 95% confident interval is statistically significant, as is the case for the first and second lags. However, some of the correlation of the second lag is partially due to the correlation of the first lag, due to a sort of spill over effect. For this reason, the partial autocorrelation is used to get a truer and more direct correlation between each lag and the price difference of time period t.



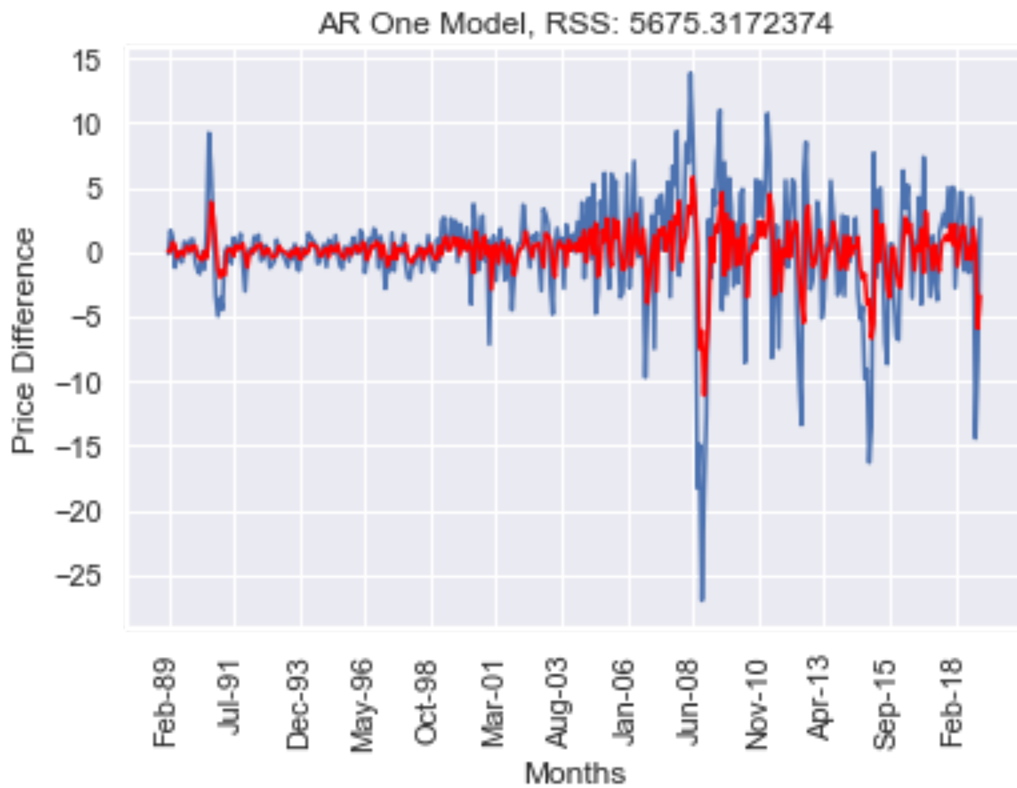
It is seen above that only the first lag period has direct correlation with the price difference of time period t. This indicates that only the first lag period is useful in making an estimation as to what the price in time period t might be. The first model that is built is an autoregression model, also known as AR model, which uses the price differences and coefficients of time periods prior to the one being predicted, whose correlation to the predicted price is statistically significant. The partial autocorrelation plot suggests that only the first lag period has predictive properties. To corroborate, the Bayesian Information Criterion is used and the AR model is further explored.

Autoregression Model

Before building an autoregression model, the partial autocorrelation and the Bayesian Information Criterion are consulted upon. As seen, the partial autocorrelation plot of the differenced data indicates that only the first lag and the price differences in time period t has a direct correlation that is statistically significant. Below is a plot of the BIC of different lag periods.



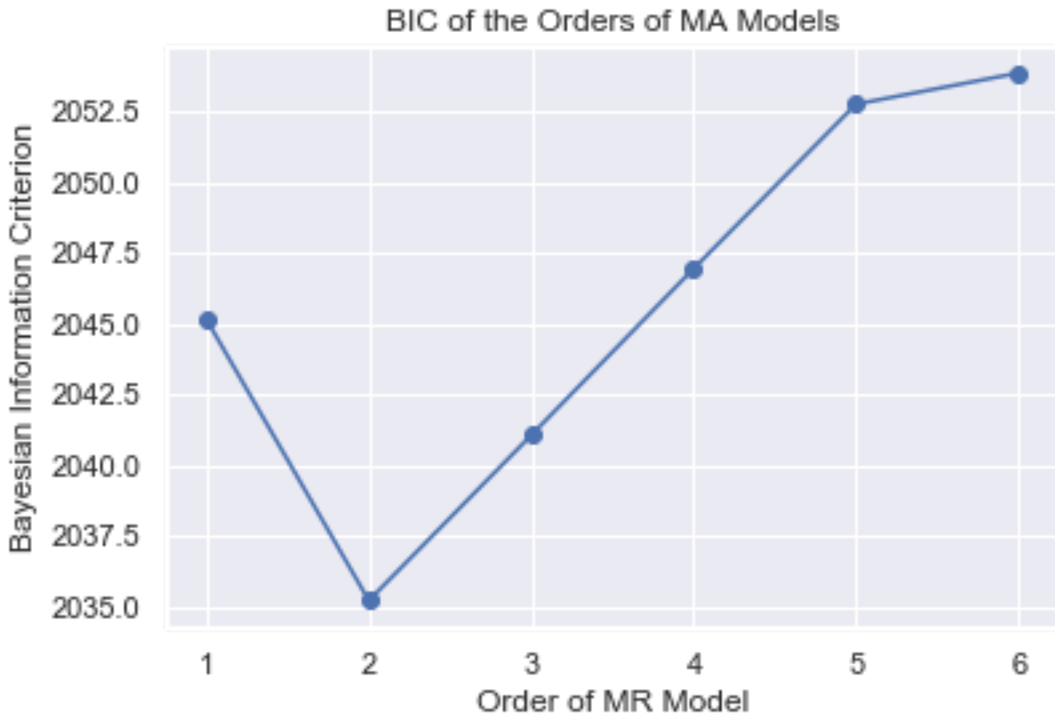
The AR model of order one has the lowest BIC. The lower the BIC, the better. Therefore, according to the BIC, the AR of order one is best, which corroborates the analysis of the partial autocorrelations. This is enough to move forward with the AR model of order one.



Above is the plot of the fitted values of the AR one model along with the differenced data. The residual sum of squares is 5675.32. Which in it of itself is meaningless because the residual sum of squares is used in comparison to other models. A moving average model was then built and compared to the AR model.

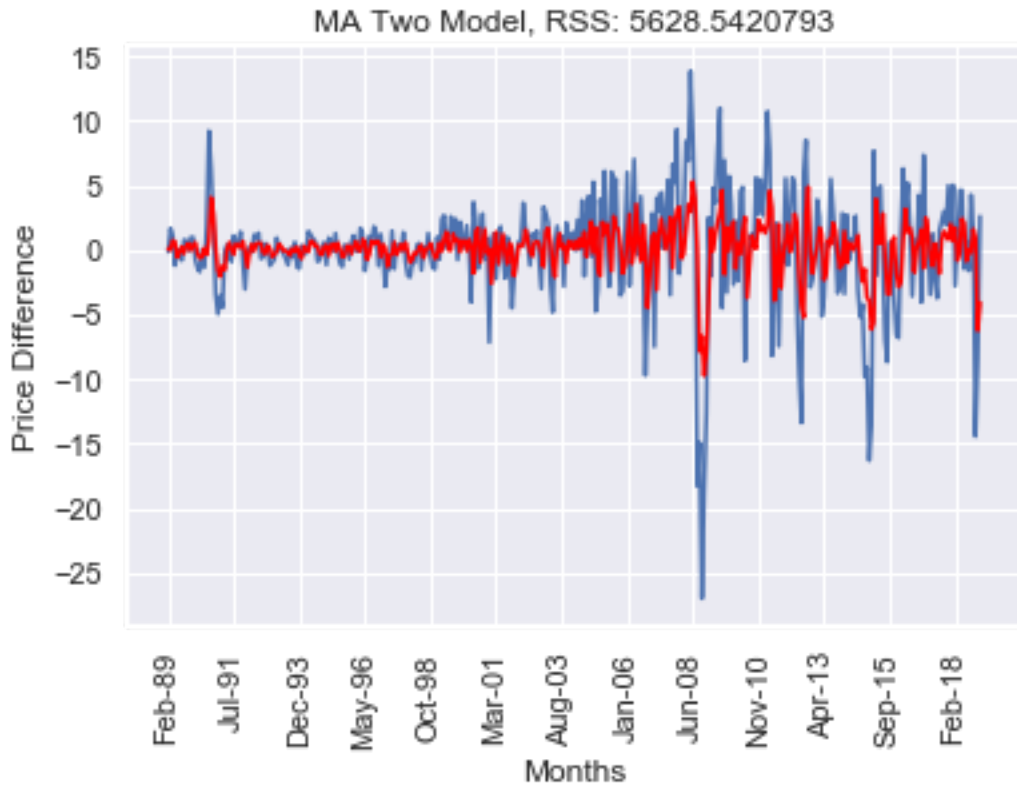
Moving Average Model

The moving average model, also known as the MA model, utilizes the error values of lagged time periods and their correlation to the price of time period t to make estimated predictions. The BIC was again used for diagnostics in regards to the lags of the error terms in relation to making estimations.



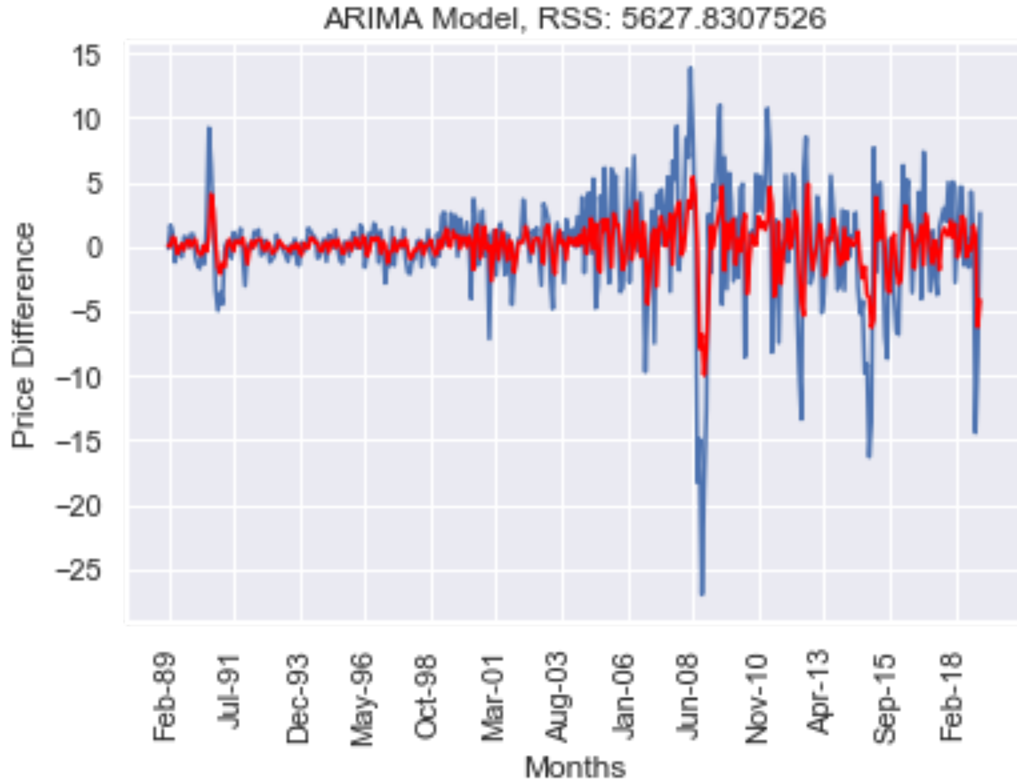
As shown above, the MA two model has the lowest BIC. Which implies that the second order of the MA model is best. To confirm this, the residual sum of squares of both MA one and MA two models are compared.

The residual sum of squares of the MA models of order one and order two are 5882.77 and 5628.54, respectively. According to this metric, the MA two model fits better. We could check for overfitting by testing a model with the order of the MA as one and another with the order of the MA as two and compare their average errors on the same test set, but that is not done here. Below, the MA models of order two is plotted. It is interesting to note the lower residuals sum of squares of the MA two model in comparison to the AR one model. Next, the AR and MA models are combined and it is seen if the residual sum of squares lowers even more.



AR(I)MA Model

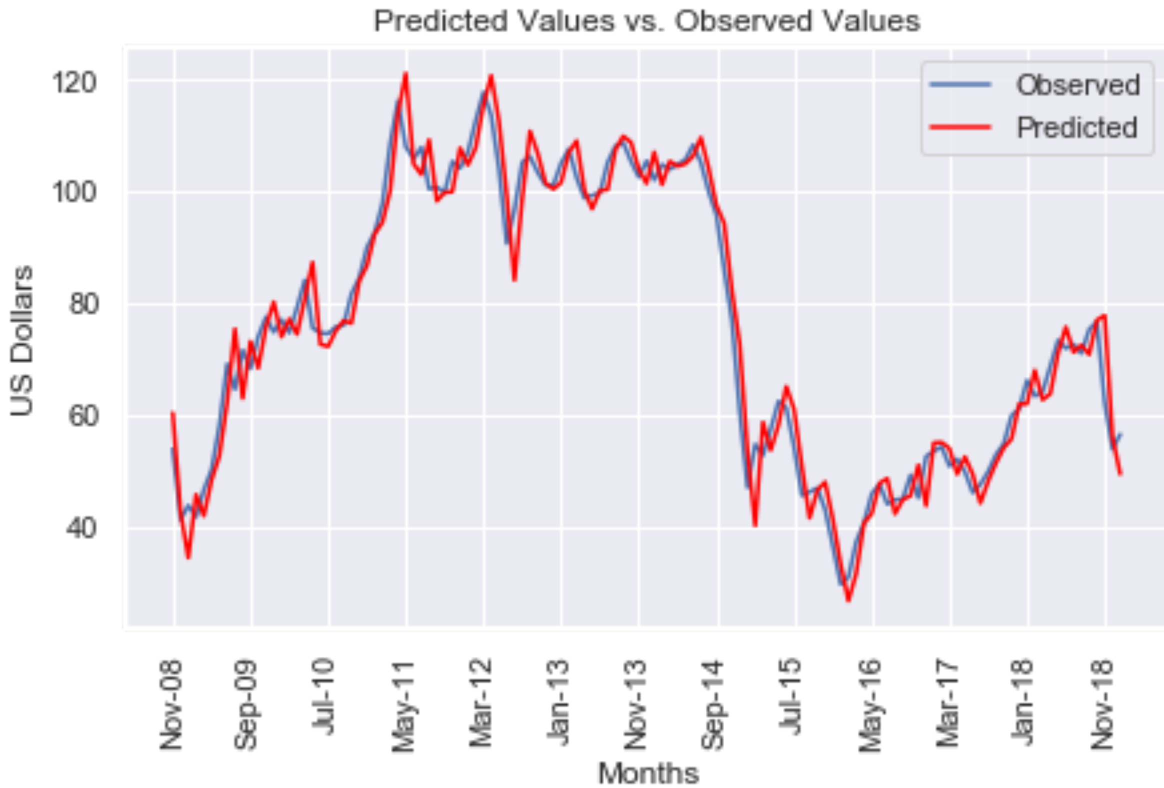
After diagnostics and fitting the AR and MA models using the parameters of best choice, the AR and MA models are combined. Below, the fitted values of the combined models are plotted along with the actual differences. The AR order is one and the MA order is two.



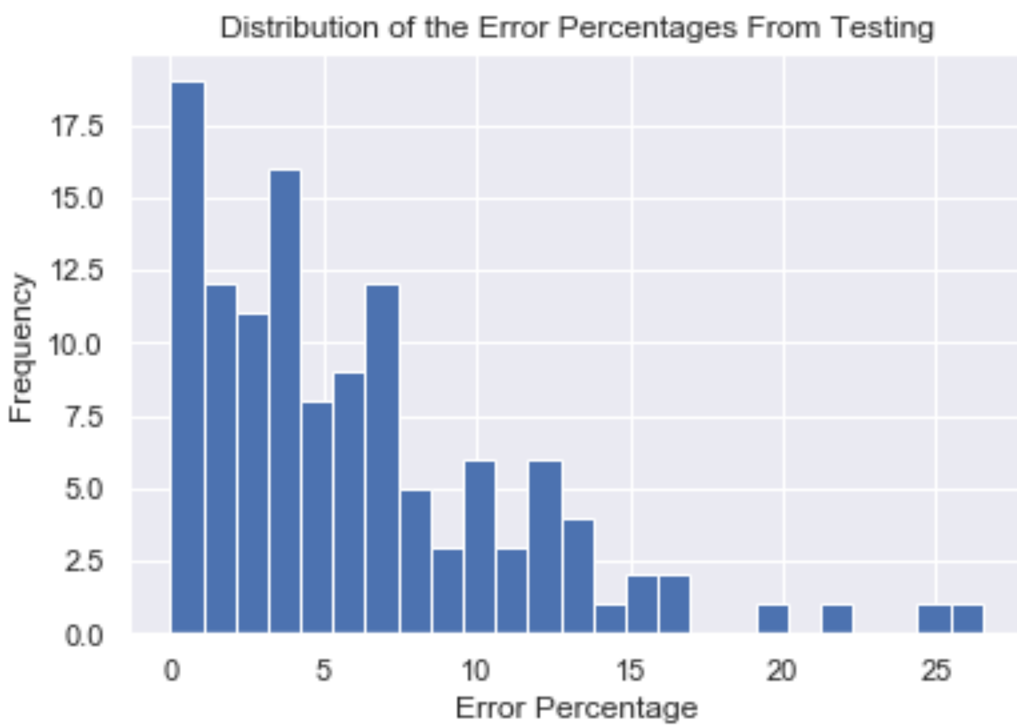
Interestingly, the residuals sum of squares for the AR(I)MA model is lower than that of the MA two model by less than 1. It is 5628.54 for the MA two model and 5627.83 for the AR(I)MA model. Given the residuals sum of squares of the models, the AR(I)MA model is used to forecast and tested.

Evaluation

The next step is to test the model using the parameters determined as most optimal. This is done by first splitting the data into a training set and a test set. The last 33.3% of the data is allocated to the test set. Then the ARIMA model is used to make predictions of differences, which are then undifferenced and put into a list of predicted prices. These predicted prices are plotted along with the observed prices. Finally, the percentage differences between the predicted values and the observed values, the errors, are calculated for each pair and averaged to obtain the final metric used for evaluation, the mean error. Below is the plot of the predicted and observed values.



The average percentage error of the test set, or error mean, is 6.09%, with the max of 28.52%. Below is the distribution of the error percentages, which provides some insight in regards to model performance.



As seen, the model error percentages are heavily distributed towards the lower end. Overall, the model performed well with 6.09% average error on the test set.

Conclusion

A model to forecast crude oil prices in the United States was developed and evaluated. In preparation, the monthly time series data of crude oil prices in the United States was made stationary by differencing. This stationarity was confirmed using the AD Fuller test. Autocorrelation was then analyzed as part of diagnostics.

When examining the autocorrelation it was found that the first two lag periods were correlated with the price of lag zero. Upon further investigation, it was found that only the first lag had direct correlation. The second lag was only correlated due to the spill over effect of the first lag. An AR model of order one was then fitted after corroboration from the Bayesian Information Criteria.

The BIC was again used in regards diagnostics concerning the moving average model. It was determined by the BIC and the residuals sum of squares that the MA two model was optimal. The AR and MA models were then combined, and the combined model was compared to the AR and MA models to see which one should be used for testing and forecasting.

The AR(I)MA model had the lowest residual sum of squares and was tested. The data was split into training and testing sets. From the test, the model was calculated to have a mean error of 6.09%. The distribution of the error percentages was plotted to get a better understanding of how the model performed. Overall, the model performed well and could be used to forecast the price of crude oil in the United States one time period ahead with 6.09% average error.