Michael Gryncewicz
Nemshan Alharthi

Data 902 - Time Series Project

Data Source:

https://fred.stlouisfed.org/series/MCOILWTICO

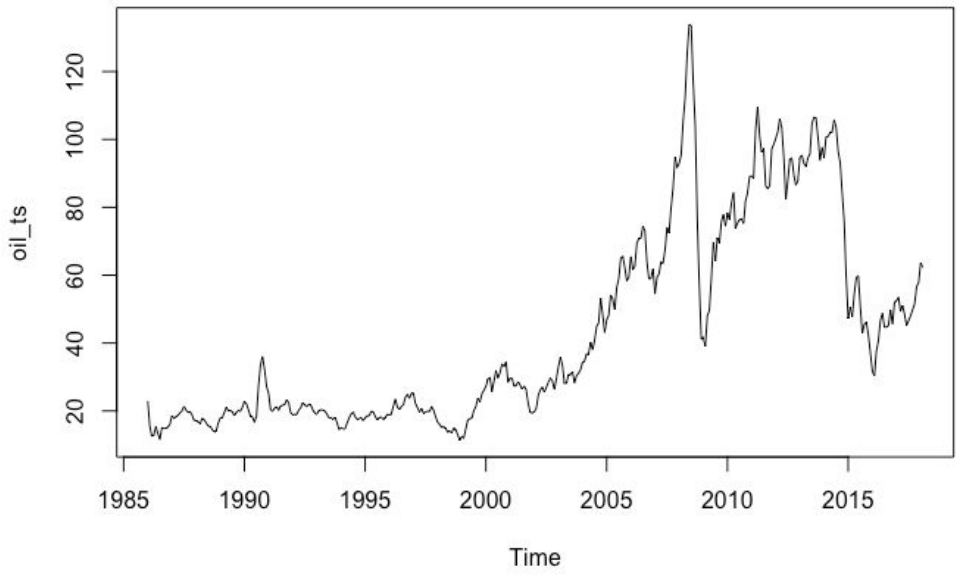
The data used for our time series project came from the link above. This data consists of monthly West Texas Intermediate oil prices from Cushing, Oklahoma. The data extends from January 1986 up to February 2018.

Analysis:

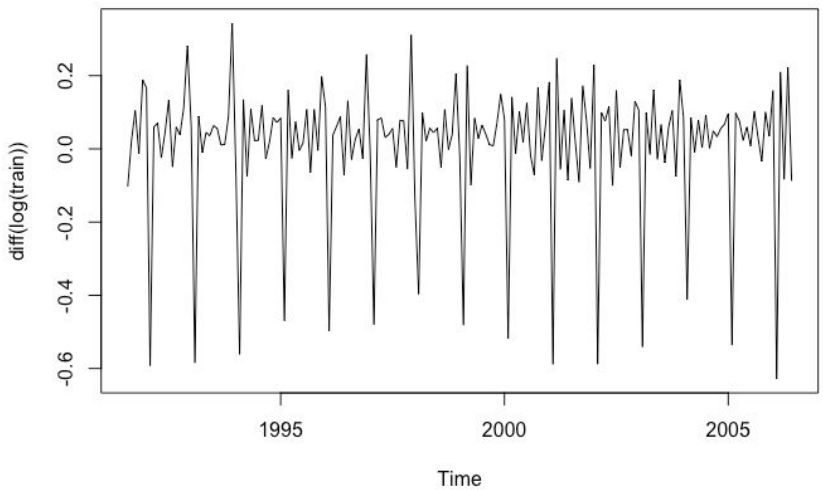
The goal of our analysis was to use different time series methods to predict the oil price for the last 6 months of the data, September 2017 through February 2018, and determine the best prediction model for this data.

ARIMA Model

The first time series model we applied to the data was an autoregressive integrated moving average (ARIMA) model. We began this method by visualizing the time series to look for trends, changes in variance and seasonality.

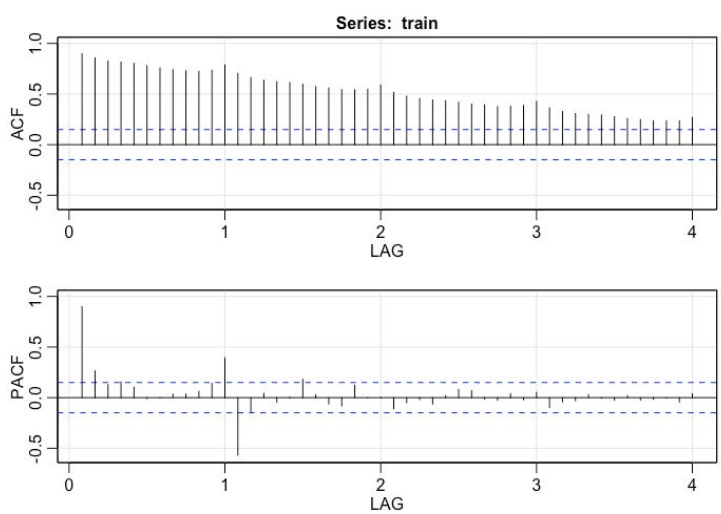


From the visual above, we noticed both a trend and an increase in the variance of the time series. We attempted to stationarize the time series by using a combination of a diff statement and a log statement. This process did introduce a some seasonality to the data which would need to be considered when fitting the model. The stationarized data is shown below.



After the data passed the visual test for stationarity, we used the Augmented Dickey-Fuller to test to help verify that the data was stationary. Running this test gave us a very small p-value which indicated that it passed the test and was ready to continue with the modeling.

The next step we performed was looking at ACF and PACF plots of the data to get an idea of the model parameters.



Looking at these plots, we saw that both the ACF and PACF plot seemed to tail off. This indicated to us that the model should have both a AR and MA component.

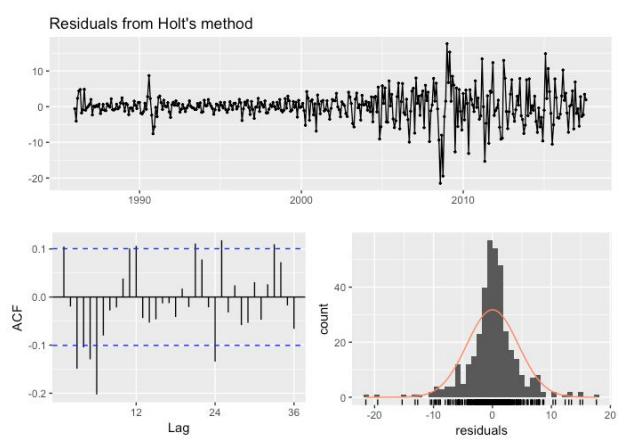
During the modeling process, we tried multiple different modelings. This can be shown in the R code provided. We compared the models on multiple criteria including AIC, BIC, accuracy, and model complexity. The model we chose as our final model consisted of the parameters p=2, d=1, q=1, P=0, D=0, Q=2, S=12. This model did not have the highest accuracy out of all our models, but it was only slightly less accurate while being less complex than the more accurate model. Due to this, we chose this model over the slightly more complex one. The final RSME for our ARIMA model was 3.018395.

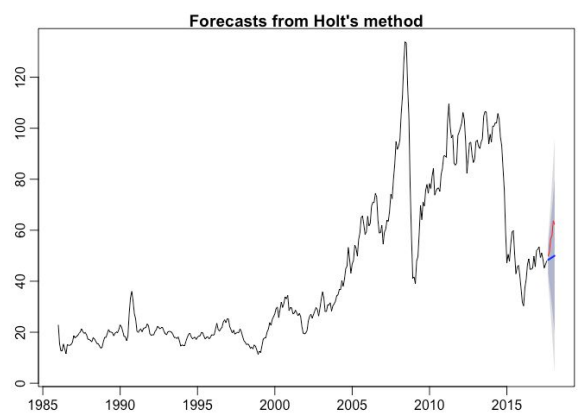
<u>Exponential Smoothing Model</u>

When fitting the exponential smoothing model, we fit three different types of EM models even though we had an idea of which one would predict the best.

We first started with a Simple Exponential Smoothing model. We did not think this model would work well as our data had a clear trend when visualized. After fitting the SES model, the RMSE value for this model was much higher than the ARIMA. For the test dataset it had an RMSE of about 10.27. This was about what was expected from the SES model based on the state of our data.

The second exponential smoothing model we tried was a Holt model. We expected this model to do much better than the SES model because it is much better at handling trend in the data.





The residual plots and the predicted value plots for the Holt model are shown above. The residual plot appears to show some autocorrelation, but this was the best of the three exponential smoothing models.

This Holt model had an RMSE of about 9.03 on the test data. This is worse than our best ARIMA model but it was better than the SES model fit above. In addition to this normal Holt model, we did also try a Holt model with dampening, but this did not improve the models accuracy at all.

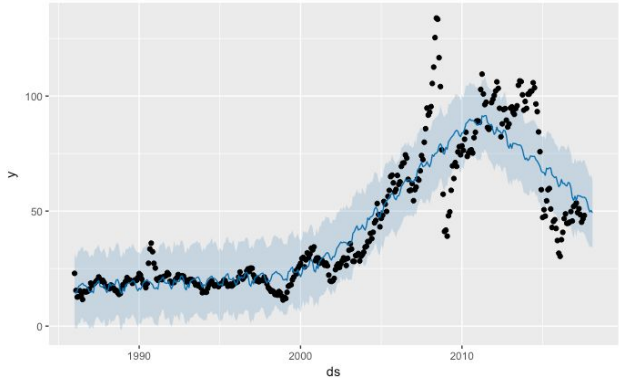
The final exponential smoothing model we fit was a Holt-Winters model. A Holt-Winters model differs from a Holt model as it can handle both trend and seasonality in the data. Before fitting it, we did not expect this model make any improvement over the normal Holt model as our data did not have a clear seasonality. After running the model, we found that we were correct with our assumption as the model predicted worse and had an RMSE of about 13.27 on the test data.

Overall, out of all the exponential smoothing models we tried, the normal Holt model performed the best, although this model did not predict better than our final ARIMA model.

<u>Prophet Model</u>

The final type of model we fit was a Prophet time series model. The prophet model differs from the previous kinds of models we have fit. The Prophet models is an additive regression model but it also fits the data using a piecewise curves. This allows for easy handling of missing data and the data does not need to be imputed. The Prophet model also differs by allowing the analyst to input a custom list of holidays that may help to detect trend and predict the future.

To begin the prophet model we first needed to change our data back to data frames instead of time series objects. Then, the models requires the columns to be named ‘ds’ and ‘y’ for it to work. The final step before fitting the model is to create a future data frame for the period that the model will predict on. Finally, we fit the model and predicted on the test data.



The plot above shows the Prophet model fit. The RMSE for this model was about 8.6 meaning it predicted slightly better than our best exponential smoothing model but not as well as our ARIMA model. The Prophet model may not be the best for our data as it is designed more for weekly and daily time series data.

Conclusion:

After trying three different forms of time series, the best model we found for this data was an ARIMA model with parameters p=2, d=1, q=1, P=0, D=0, Q=2, S=12. This was the best model when comparing AIC, BIC, RSME, and model complexity criterias. This kind of model works better for our data but it may not work the best for a different time series dataset.