Stanford | ENGINEERING
Computer Science

Predicting Oil Prices

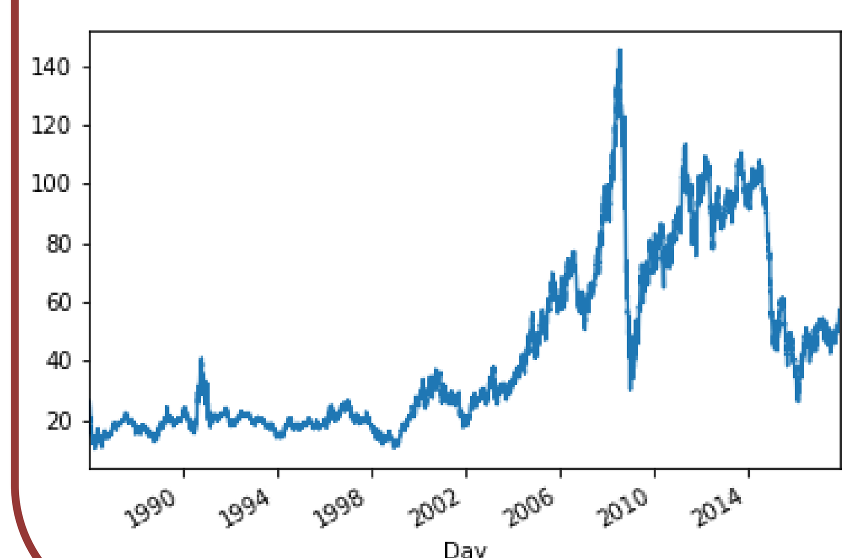
Jeffrey Hale

Summary

The goal of this project is to predict the future pricing of oil based on historical oil prices as well as various factors that affect supply and demand. The end goal is not speculation for financial gain, but rather to predict future pricing for better planning of capital expenditures and estimation of future revenue from the point of view of the supplier. To this end, I built two models to predict oil pricing, one based on an autoregressive integrated moving average (ARIMA) and another based on a neural network (NN). We can see mixed results, where the ARIMA builds a lagging version of the input data and the NN sees overfitting of the training set.

Data

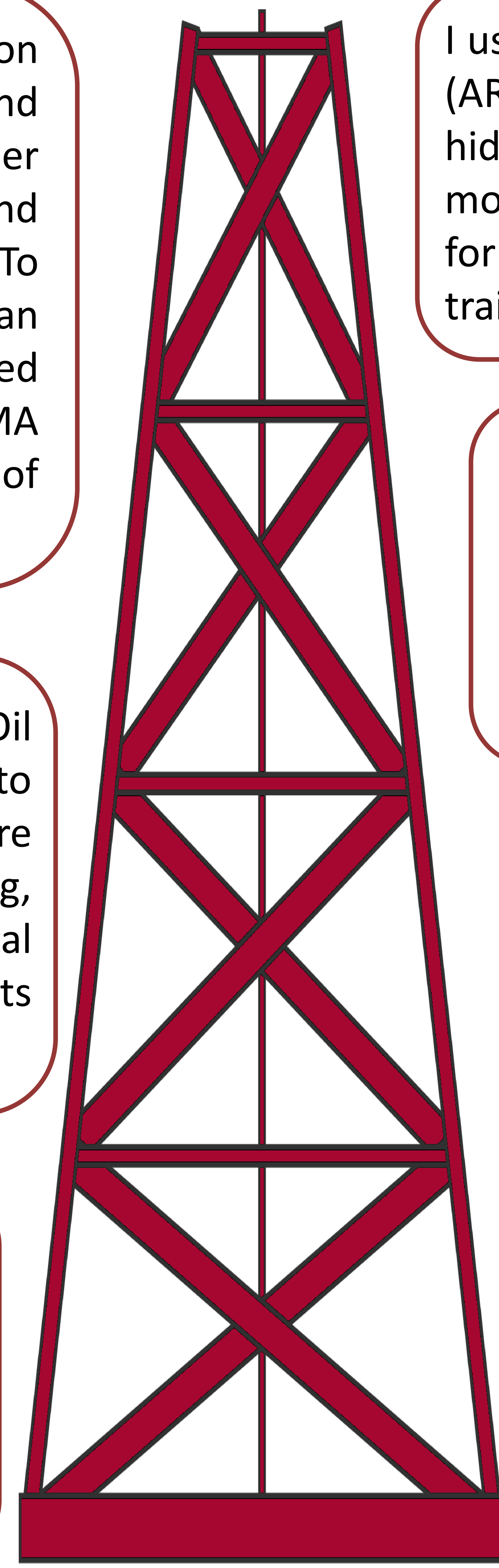
WTI Daily Oil Spot Price



The primary dataset is the WTI Crude Oil daily spot prices from January 1986 to November 2017. Other data considered are historical daily treasury rates, gold pricing, S&P500 pricing, oil company capital expenditures and various industry reports such as the OPEC monthly report.

Features

The features for the ARIMA are the historical oil prices themselves, while the features for the neural network that I tested vary. For the neural network, I added all previous oil price data as features for a given date, as well as trying out various combinations of correlated and uncorrelated historical data such as interest rates, gold prices, and capital expenditures of oil companies.



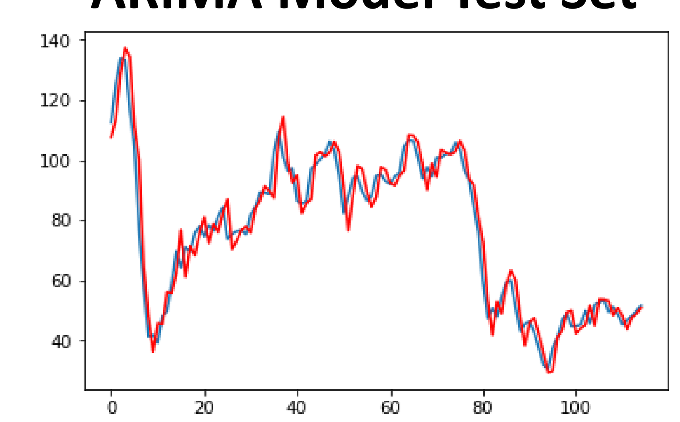
Models

I used two different models, an autoregressive integrated moving average (ARIMA) and a neural network with a learning rate of 0.000001 and one hidden layer with 200 hidden units. Both models used mean squared error (MSE) for the cost function. With both models, 70% of the data was used as the training set and the remaining 30% was used for the test set.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Results

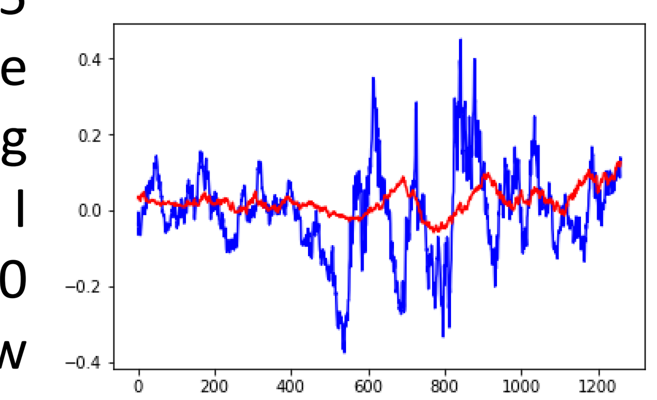
ARIMA Model Test Set



MSE = 39.57

For the ARIMA, the trailing 5 monthly values are used to get the prediction, which leads to a lagging version of the actual. For the NN, I used the percent gain/loss after 30 days to train the model as the raw value lead to some strange behavior.

Neural Network Test Set



MSE = 0.0068

Discussion

It is my belief that commodity cycles like that of the oil market are driven by the mismatch of the capital investment cycle with the demand cycle, leading to periods of oversupply and shortage. In this project, I attempted to model this mismatch by supplying a neural network with some of the indicators of future supply and demand in order to try to predict the effect of this expected mismatch on future oil pricing, but it seems that better indicators are needed.

Future

In the future, I would like to explore using the neural network approach more in depth. I would like to try a deeper network as well as adding more complicated and nuanced features such as the word counts of key words in the monthly OPEC reports.

Jeffrey Hale (jphale@stanford.edu)