What Drives WTI Futures Returns?
– An empirical approach

Julian Murillo, Lindsey Benson, Isabella Morrow, Rehjii Martin
Chapman University, ECON452 – Econometrics
Emperical Research Study

Introduction

The West Texas Intermediate Crude Oil Futures serves as a benchmark in oil pricing, which is extremely significant to the United States economy. Oil prices are volatile in nature, so predicting future oil prices may assist in determining the health of the overall economy. Abrupt changes in crude oil prices could negatively alter the growth of the economy, causing harm to investors, consumers, and producers of crude oil. Price sensitivity, possibly caused by geopolitical events, market fluctuations, and policy changes, demonstrates the importance in forecasting an accurate crude oil futures return model. Not only is crude oil one of the most popular traded commodities, it is also very influential to prices of other commodities. Additionally, their volatile nature creates trading opportunities, contributing to its popularity among traders worldwide. This paper will examine economic indicators that may be significant in determining the asset returns of Crude Oil WTI Futures US. Further, this paper analyzes WTI crude oil futures return's predictability from 2000 to 2019. Numerous models will be used to test what variables are truly powerful in determining asset return of Crude Oil WTI Futures US. WTI's market is primarily the United States, as it is extracted from oil fields in Louisiana, Texas, and North Dakota. Moreover, our forecasting model will primarily focus on the predictability of Crude Oil WTI Futures returns in the U.S. market.

Literature Review:

Crude Oil Futures are contracts traded on the exchange in which a buyer agrees to buy a specific quantity of crude oil, typically in barrels, at a predetermined price on a future delivery date. Extensive research has guided us to the following significant variables that may play a key role in predicting West Texas Intermediate crude oil futures.

According to an article written by the European Central Bank, using futures prices to predict oil prices can be a very useful and easily understandable tool; however, they have contributed to large errors in HICP inflation in the past. The harmonized index of consumer prices (HICP), is a consumer price index which is compiled according to a methodology that has been harmonised across EU countries. Oil prices are hard to predict because oil market dynamics are changing substantially all of the time. This specific article gives possible approaches to mitigating risk when forecasting oil prices. Most of the volatility in the Euro area HICP inflation is said to be due mainly to the energy component. A direct correlation can be seen between oil prices and energy; for example, oil prices dropped sharply in 2014, negatively impacting energy exporting nations.[1] There are arguments that suggest current prices of oil may be the best predictors of futures prices because changes in oil prices are inherently unpredictable. During the 2008 financial crisis oil prices dropped by almost 70%, so overall performance of the global economy serves as a large component of the market performance of oil. A change in oil prices inversely affects global GDP; for instance, a 10% increase in price may result in a 0.2% drop in the global GDP.[2] However, the relationship is positively correlated for oil exporting countries. This is most likely because oil profits add to these foreign economies, while taking away from global business. When oil prices are lower, plane tickets are lower and things directly related to cheaper global business. Additionally, there was a sharp decline in oil demand due to the decline in global economic activity during the financial crisis, and this can also be attributed to the drastic fall in oil prices. Further, oil demand; as well as, inventories, and oil supply have been found to largely impact oil prices. Researchers from the European Central Bank argue the fluctuations in oil prices may be caused by alterations in oil demand, production, and oil inventories.

Researchers Miao, Ramchander, Wang, and Yang conducted study utilizing the LASSO regression model and have found a significant relationship between the United States steel production and the price of oil. They used steel production as a demand factor in their oil price forecasting models and found demand factors, such as steel, to be highly influential among most years[3]. As these scholars mentioned, it is important to consider previous oil demand and supply shocks, as these events have attributed to the volatility in oil prices. These disruptions, such as, pipeline issues or plant outages, directly impact the price of oil[4].

---

[1] European Central Bank. (2015)

[2] European Central Bank. (2015)

[3] Miao, Ramchander, Wang, & Yang, (2017)

[4] Miao, Ramchander, Wang, & Yang, (2017)

Economic researchers Lee and Huh suggested factors significantly influencing oil prices have changed over the past few decades. For example, supply-side factors were major determinants until the oil price collapse in the 1980s. Demand factors become more influential in the 1990s, as an oil pricing system directly linked to the oil market was created. As of 2000, financial factors, which includes the financial crisis and the strength of the dollar, have become possible determinants of oil prices, and may be more significant than demand and supply factors[5].

According to JP Morgan's Mid-Year Market Outlook for 2019, "As long as demand does not crater, heightened geopolitical risks and OPEC cuts should still be enough for oil to provide positive returns for the latter part of this year."[6] In 2016 the OPEC (the oil producer group that pumps more than half of the world's crude along with its allies) pledged it would help to balance markets, and has remained consistent with that declaration. Political factors that cannot be quantified will have strong effects on the prediction of oil prices in the future, such as Iranian tensions and conversations between Saudi Arabia and Russia. Since oil production often happens in sanctioned countries or politically sensitive areas, there is a lot of correlation seen between terrorist attacks in these countries and changing political backgrounds. However, global terrorism will not be highly correlated because there are many different variables that contribute to terrorism itself. Terrorism specifically in the Middle East, may give more insight into prices of oil futures. JP Morgan claims a barrel of oil should remain in about the 60 dollar range.[7] US Bank also had some comments about Saudi Arabia's affect on oil prices, as they seemed to allow some flexibility on the OPEC production cut agreement and this caused some worries about oversupply, which is correlated with oil prices being fairly low. There will be another meeting in the near future where the OPEC will decide whether the production cut agreement is extended. If the agreement is not extended, oil prices will likely reset lower. The OPEC seems to have direct effects on oil prices and whether they are to go up or down in the future, which has a large effect on the global economy. "The main reason for the large forecast errors of futures is that the futures curve is usually flat and downward sloping owing to the specific nature of oil as a physical and

[5] Lee, & Huh. (2017)

[6] J.P.Morgan. (2019)

[7] J.P.Morgan. (2019)

storable commodity," according to the ECB.[8] As a result, spot prices are often higher than futures prices, creating a downward slope which is also known as backwardation. J.P. Morgan estimates the S&P GSCI ER index and the BCOM ER index to return 9% and 6% respectively. Bullish second half forecasts for oil, agriculture and precious metals should push broader commodities indices around over the balance of the year.

A study at Columbia University unveiled the influential relationship between oil prices and renewable energy. Researchers Heal and Hallmeyer found low oil prices may impact renewable energy progress by creating competition within the transportation sector. Low oil prices may compete with biofuels as gas replacements, causing a reduction in public appeal of biofuels and electric vehicles (Heal & Hallmeyer, 2015). Renewable energy may influence oil prices further, as demand would shift to the least expensive option for energy consumption. Electric vehicles may be more expensive to purchase, but are overall less expensive to maintain compared to vehicles with internal combustion engines. Researchers Heal and Hallmeyer found electric battery cost to be falling, while gasoline was found to be more expensive per mile driven[9]. Although there are other factors, such as availability of charging stations and charge time, contributing to hesitation on moving away from internal combustion engines, higher oil prices would likely improve competition for vehicles powered by renewable energy[10]. As for the generation of power, the study found that oil only would compete with renewable energy if the price of oil was at the unsustainably low level of $15 per barrel[11]. Oil producers would fail to make a profit, so the price of oil is not significantly related to renewable energy in relation to electric power.

According to CME Group, natural gas and crude oil have continuously demonstrated a direct connection with regards to both the supply and demand side[12]. Natural gas is produced either as a by-product or as the primary energy source, and the quantity of natural gas produced is dependent on the type of well being drilled. Many companies typically produce

[8] European Central Bank. (2015)

[9] Heal, G., & Hallmeyer, K. (2015)

[10] Heal, G., Hallmeyer, K.(2015)

[11] Heal, G., Hallmeyer, K. 2015)

[12] Mchich, A., & CME Group.(2018)

more oil or natural gas if its return on investment is greater and profits can be maximized[13]. For example, as crude oil prices rise and oil drilling increases, natural gas drilling would be expected to decrease, causing natural gas prices to decrease. In relation to the demand side, an increase in the price of one of these energy sources would incentivize individuals to switch to the fuel with the lesser cost[14]. Further, as natural gas prices increase, the demand for crude oil would increase. However, it has been suggested that this relatively stable relationship has changed due to the U.S. Shale Revolution[15]. The Shale Revolution allowed the United States to increase natural gas and oil production from hydraulic fracturing and horizontal drilling[16]. This significant event directly decreased the United States's dependency on oil imports and contributes to delivering a boost to the economy during the 2008 recession[17]. According to CME Group, shale changed crude oil and natural gas's supply structure, causing their price relationship to weaken. Despite a possible weak relationship between natural gas and crude oil, the Shale revolution may influence volatility of crude oil prices due to the growing independence the United States gained.

Conceptual Framework

From the research above, we were able to pull 27 variables that would contribute to creating the best possible model to predict Crude Oil WTI futures prices. Energy prices seemed to be very relevant when predicting prices due to the sustainability aspect of it, if people are using more energy they are most likely to use less gas. We believed this would have an inverse relationship. Another very strong indicator we predicted would be steel prices because pipeline issues or plant outages have very strong influence in the crude oil industry. This would be a strong x-variable in our model. Inflation is predicted to be a key macroeconomic indicator in our model because inflation occurs as the economy experiences expansion. Further, consumer spending is high and people are more inclined to pursue riskier investments. Therefore we can expect a positive correlation between inflation and crude oil

---

[13] Mchich, A., & CME Group. (2018)

[14] Mchich, A., & CME Group. (2018)

[15] Mchich, A., & CME Group. (2018)

[16] The Strauss Center. (n.d.)

[17] The Strauss Center. (n.d.)

WTI futures prices. After compiling the variables, our aim was to be able to predict WTI Futures Returns parsimoniously: relative simplicity and high predictive power. The variables mentioned above were the three most recurring variables in all of the literature we reviewed. We then decided to run various machine learning models such as the LASSO and the Elastic Net over our data set for best subset selection. We wanted to see the strictest method of shrinkage versus the hybrid. From the literature, we discovered that Elastic Net often outperforms LASSO for volatile y-variables such as futures prices. Elastic Net is a hybrid between Ridge Regression and LASSO. You would use LASSO in a scenario where you have many variables and only a few matter whereas you would use Ridge in a model with many variables in which case all matter a little. That's why we elected to use a hybrid Elastic Net model, given that our data was somewhere in-between and given our relative inexperience in dealing with futures returns. After selecting the best variables from Elastic Net we used Variance Inflation Factor (VIF) to rule out any variables that were highly correlated with one another. Using those 11 final variables we then ran Ordinary Least Squares (OLS), Ordinary Least Squares (k-fold CV), Auto Regressive Integrated Moving Average (ARIMA), and Logistic Regression models.

Data Description

Filtering Process Summarized

Initially, we selected monthly variables from the class data bank that contained no missing information. By attempting to replace missing data, it would only create reliability and validity issues within our model. Further, we decided it was best to disregard all variables with incomplete datasets. Our model utilized monthly data, starting from January 1, 2000 until August 31, 2019. We eliminated variables within the data bank that failed to be within our time parameters, and that represented yearly, quarterly or daily time intervals. We immediately identified some variables that are known for being multicollinear; such as the NASDAQ, and flagged them for future reference until we could formally check for collinearity. We then collected additional data from other credible sources. These sources include FRED, IndexMundi, Bloomberg, the Federal Reserve, the United States Energy Information Administration, and the IMF. All data retrieved from these sources was downloaded in monthly format from January 1st 2000 to August 1st 2019. The data was also seasonally and inflation adjusted (real). Further, since price, open, and volume data are not potential target variables, we removed them and ended up with 29 total variables with 27 predictors other than our “date” variable. Following this initial data compilation and cleaning process, we then ran LASSO (strictest) and Elastic Net exclusively for variable selection. The eight coefficients that were selected from LASSO (those that were not reduced to 0) were: inflation rate in U.S.(as a percentage), imports to opec (in thousands of barrels of oil products), primary energy consumption in the U.S. (quadrillion BTU), the Nasdaq Composite Index, Steel Production in the U.S. (million tonnes of crude steel), bloomberg commodity index (b_com), U.S. oil rigs in operation, and the price of natural gas (index). After running Elastic Net, using the optimal alpha of .8 (minimum amount of cross-validated loss) the 11 coefficients that were selected were those listed above as well as 2 more additional variables which were Net Primary Energy Imports (quadrillion BTU) and the 3 month treasury bill yield. We then checked for multicollinearity and collinearity over all 27 predictor variables. We did this by using VIF, or the Variance Inflation Factor Test to generate VIF scores. Once we had VIF scores, we then removed variables with problematic VIF scores from our compiled data set. Finally, we looked to see if any variables selected from our Elastic Net model had high VIF scores. We were extremely fortunate to have all of our Elastic Net variables have relatively low VIF scores. Finally, we were left with the following variables in which to generate models and predictions with: WTI_return, date, inflation_rate, import_opec, primary_energy_cons, primary_energy_import_net, tb_rate, nasdaq_comp, steel_prod, b_com, us_rig, and n_gas.

Checking for Stationary Data

After compiling the data, we needed to start checking for seasonality and stationarity. To check for seasonality, we used the decompose function in R on our time-series data. We discovered our model was not seasonal due to the sheer volatility of the futures market. We also found that due to geopolitical events such as the Gulf War of 1990 and the War on Terror in 2002, that crude oil prices have become much more volatile and that seasonality is harder to detect in a shorter timespan (less than 20 years)[18]. We found there was no trend by analyzing the trend component of the decomposed time-series data visually. To confirm our observations that our data was trend-stationary we utilized the Augmented Dickey-Fuller

[18] Johnson, C. (2008, November 20)

(ADF) Test, in which we used a threshold p-value of .05. The results from the ADF Test yielded a p-value below .01, thus we were able to reject the null hypothesis that a unit-root was present and confirmed that our data was in fact stationary and we could begin constructing our Auto Regressive Integrated Moving Average model.

Econometric Modeling

A.R.I.M.A

The first model we wanted to test our cleaned/filtered data set on was one specifically designed for time-series, which is an Auto Regressive Integrated Moving Average Model (ARIMA). Our optimal ARIMA forecast with parameters (1,0,0)(1,0,0)[12] did not yield the best results, forecasting after 6-12 months essentially flatlined into an average of all WTI_Return. Looking at the Partial AutoCorrelation Function (PACF) which tells us the direct correlation of the WTI_Price N months ago on the WTI_Price today, without considering or removing the effects of the Returns on months between those periods, there was only 1 statistically significant lag period (0). Looking at the AutoCorrelation Function (ACF) which tells you the correlation between the WTI Futures Returns N months ago and the Returns today, which includes both direct and indirect effects of n periods or more potential noise, there was also only 1 statistically significant lag period (0). This model came out very weak and we decided not to continue down this path.

Ordinary Least Squares Regression (non-kfold)

When we ran an OLS model over our compiled data set (prior to applying LASSO, Elastic Net and VIF) we got an adjusted r-squared of .2984. So essentially, with all variables, even the potentially confounding ones, about 30% of the variation in WTI Returns was explained by all 27 variables. After selecting the unproblematic variables (low VIF) with the greatest statistical significance (using LASSO and Elastic Net) and running an OLS model over this data set (WTI_return ~.) we got an adjusted r-squared of .2522. We were pleased with this coefficient of determination because it appeared that our variable selection methods had been successful, and that we had achieved a certain level of parsimony having eliminated 16 variables from our initial data set and only seeing a reduction in adjusted r-squared of only 4.6%.

Ordinary Least Squares Regression (k-fold: nfolds = 10)

If you look at table 1.0, you can see an example of kfold cross validation with nfolds = 5. It is very similar to Leave One Out Cross Validation, except that it partitions the data into "k" distinct groups and excludes k-group from the model. Running an OLS model using k-fold cross validation with 10 folds yielded an adjusted r-squared of .2372 which was even worse than our standard OLS model by 1.5%.

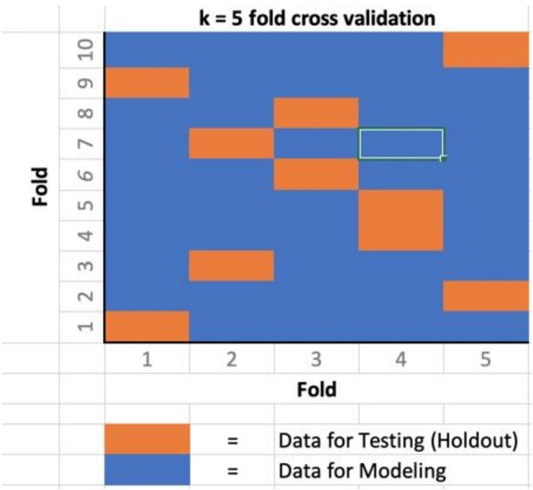


Table 1.0

Logistic Regression (Probability cut-off = .5 or 50%)

Finally, after having attempted to run ARIMA, OLS, and OLS k-fold CV, we decided that it might be of some use to predict the probability of a WTI return being either good or bad as it was proving very difficult to predict an actual return value. We began by creating a binary dummy variable called good_return to split the clean/ filtered data into good and bad returns, with the split being at the median return value of 1.7150. We used this binary variable as our target. Once iterating over our clean data set with 11 variables and exponentiating their coefficients to calculate probabilities we made some interesting observations. Firstly, if you had a return that was produced during a period of high natural gas prices, then you were 18% less likely to have a good return. Further, if you had a return that was produced during a period of high primary energy consumption, then you were 36% more likely to have a good return, if you had a return that was produced during a period of high 3 month Treasury Bill rates then you were 14% more likely to have a good return, and if you had a return that was produced during a high inflationary period you were 23% more likely to have a good return. These observations seem to align with observations made in other academic literature.

Robustness Checks

After getting results from our more promising models we decided to check the robustness of the models and compare them to reach a final conclusion on which would be the best to use in a real-world scenario.

Ordinary Least Squares Regression (non-kfold)

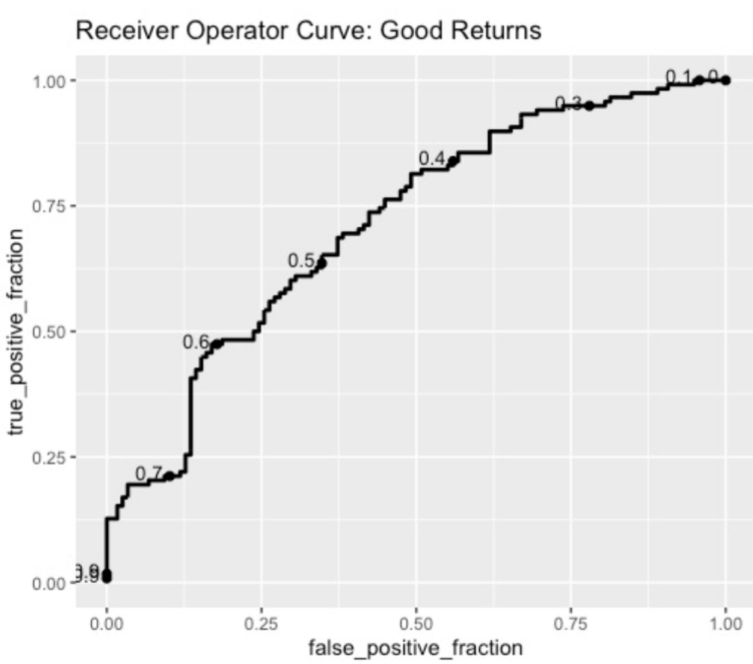
For our standard OLS model, we decided to test the performance of the model by performing a Breusch Pagan Test for Homoscedasticity, in which case the null hypothesis is that the model is homoskedastic and the alternative being that the model is heteroskedastic (non-constant variance in the error term). Our p-value came out to .62 which means that we fail to reject the null and conclude that our model is homoskedastic, which means that our filtering/shrinkage methods appear to have worked. We then calculated the model's Root Mean Square Error (RMSE) which came out to 7.7519 which is relatively high considering the distribution of the WTI Return data column. This RMSE would be placed just above the 3rd quartile (7.07) of the WTI return data column. This model was looking the most promising thus far.

Ordinary Least Squares Regression (k-fold: nfolds = 10)

For our K-fold Cross Validation Model we also tested the performance of the model by performing a Breusch Pagan Test for Homoscedasticity. Our p-value for this model came out to .41 which essentially means that this model is slightly more heteroskedastic than the non-kfold mode, but still homoskedastic. In addition, when we calculated the RMSE the output was 8.161 which was worse than the non-kfold model, this value being placed well above the 3rd quartile of the WTI return data column.

Logistic Regression (Probability cut-off = .5 or 50%)

As for our final model, the way we conducted robustness checks was through calculating sensitivity (True Positive Rate), specificity (True Negative Rate), the False Positive Rate, and the overall accuracy. We felt that a probability cut-off of .5 seemed appropriate due to the fact that the distribution of probability scores was apparently normally distributed and had some observations near the .9 probability and some at the .1 probability as well. A true positive in our case would be a WTI return observation that had a predicted probability over 50% of being a good return and was actually a good return (using the binary good_return variable). A true negative would be a WTI return observation that had a predicted probability over under 50% of being a good return and was actually not a good return. A false positive would be a WTI Return Observation that had a predicted probability over 50% and was actually not a good return. The accuracy is calculated by taking the TP rate and the TN rate and dividing it by all observations (nrow). Using the cut-off probability of .5 our True Positive Rate was .6356, our True Negative Rate was .6525, our false positive rate was .3475, and our accuracy was .6441 (64.41% accurate). If you look at table 1.1 you can see a Receiver Operator Curve which depicts how different cut-off levels affect the TP and FP rates, as well as a confusion matrix.



	0	1
Predicted No	77	43
Predicted Yes	41	75

Table 1.1

The zeros and the ones indicate which observations were actually good or bad returns, and the "Predicted No / Predicted Yes" indicate the probability scores as a percentage, and depending on whether or not the predictions were over or under 50% likelihood, they would fall into the "Predicted Yes" or "Predicted No" category. Once we generated the R.O.C plot, we could conduct one final test to evaluate the precision of our model, the Area Under the Curve (AUC). An AUC will tell us how much better our model is than chance. As a rule of thumb that most researchers use, an AUC of .7-.8 is acceptable, an AUC of .85-.9 is good, and an AUC of .9-1 is excellent. Our AUC came out to .7124 which is not the strongest, but acceptable. In a real life economic scenario, it would be more costly to make a decision to invest during a period that predicted a bad return but was actually a good return if you were to do something like short-sell, because then your losses on that short-sell could be infinite.

Conclusion and Recommendations

In conclusion, the models that performed the best were (in order from worst to best): OLS using K-Fold Cross Validation (Adj. r-squared: .2372), Standard OLS (Adj. r-squared: .2522) , and Logistic Regression (AUC: .7124). Once again, all models were generated using the filtered data set containing 11 predictor variables, 1 target (WTI_return), and monthly time (date). In brief summary, after reviewing academic literature, performing best subset selection, developing 4 different models, generating predictions, and evaluating their performance against one another we discovered first-hand how difficult it is to predict West Texas Intermediate Crude Oil Futures returns. We noticed that there were some variables that could have been of some importance to us when predicting WTI returns such as fatalities from terrorist attacks in OPEC, oil spills, fatalities from oil spills, etc. But a dilemma we ran into many times was that to obtain this data at all we would have had to pay for the data or pay for an API. In the future, if we choose to do further analysis, we could consider using other methods such as Leave One Out Cross Validation, Random Forest, or Boot-Strap Aggregation to iterate generate predictions for our Logistic Model or our OLS model. If we were to give advice to an investor or to a firm looking to invest in crude oil contracts we would say the following: We recommend purchasing futures during a high inflationary period (when the economy is experiencing an expansionary period), when the 3 month T-bill yield is high (signaling economic expansion as well), when the price of natural gas is low (greater volume trades in Oil/other sources of energy), when energy consumption is high (both primary and distillate), and when the Bloomberg commodity index is high as well. But overall, we advise again that investing in oil futures contracts is very risky and should not be the first option for a novice investor.

References:

Heal, G., & Hallmeyer, K. (2015, October 15). SIPA Center on Global Energy Policy. Retrieved from https://energypolicy.columbia.edu/research/report/how-lower-oil-prices-impact-competitiveness-oil-renewable-fuels.

Johnson, C. (2008, November 20). TIMELINE: Half a century of oil price volatility. (2008, November 20). Retrieved from https://www.reuters.com/article/us-oil-prices/timeline-half-a-century-of-oil-price-volatility-idUKTRE4AJ3ZR20081120.

J.P.Morgan. (2019). Mid-year Market Outlook 2019: Central Banks Stay Easy and in Sync. Retrieved from https://www.jpmorgan.com/global/research/2019-mid-year-outlook.

Lee, C.-Y., & Huh, S.-Y. (2017). Forecasting Long-Term Crude Oil Prices Using a Bayesian Model with Informative Priors. *Sustainability*, *9*(2), 190.

Mchich, A., & CME Group. (2018). Are Crude Oil & Natural Gas Prices Linked? Retrieved from https://www.cmegroup.com/education/articles-and-reports/are-crude-oil-natural-gas-prices-linked.html.

Miao, H., Ramchander, S., Wang, T., & Yang, D. (2017). Influential factors in crude oil price forecasting. *Energy Economics*, *68*, 77–88.

The Strauss Center. (n.d.). The U.S. Shale Revolution: Energy and Security. Retrieved from https://www.strausscenter.org/energy-and-security/the-u-s-shale-revolution.html

U.S. Bank. (2019). Weekly Market Update: 12.9.19: U.S. Bank. Retrieved from https://www.usbank.com/investing-and-retirement-planning/financial-perspectives/weekly-market-update.html.