Major Players in the West Texas Intermediate Crude Oil Pricing Market

Authors:
Nasim A. Mamaghani [a]
Jongbae Yoon [b]

Supervisor:
Dr. Brian Zaharatos [b]

[a] Department of Civil, Environmental and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309, USA.
[b] Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO 80309, USA.

Abstract

Crude oil is one of the most valuable commodities in the world and its pricing attributes to world economy in different levels. Given the complexities around the oil market, this study aims to understand the major players in the oil market by studying the West Texas Intermediate (WTI) crude oil market data from 2002 to 2015. More than 25 parameters that are commonly known to be affecting oil price are analyzed in this paper using Multiple Linear Regression modeling techniques. Results show that 10 factors spanning across 5 different categories (supply, demand, financial market, commodity market, and geopolitical factors) have significant correlations with the WTI pricing. Among these 10 factors, it is found that Europe and U.S. GDP growth rates have the highest impact on the oil price market. Findings of this study can provide a preliminary insight into the highly intricate and competitive crude oil market and can be used as baseline for future statistical modeling with higher orders of accuracy.

1. Introduction

Oil price fluctuations have an indisputable effect on the world economy at different levels [1]. Khalid and Sarwar [2] studied the impacts of oil price on micro to macro levels in Pakistan and reported significant correlations between oil price and several macro- and micro-economic indexes. Farzanegan and Markwardt [3] found a strong positive relationship between oil price changes and industrial output growth in Iran, a country that heavily relies on its revenue from petroleum exports. According to the 2009 International Energy Outlook report by the U.S. Energy Information Administration (EIA) [4], oil price assumptions attribute to one of the largest uncertainties around long-term energy forecasts. Kang and Ratti [5] reported that oil-market specific demand shocks attribute to 12% of the long-run variability of real stock returns in the U.S. Another study conducted by Tang et al. [6] found that the long-term impacts of oil price shocks overpower the short-term impacts on the Chinese economy.

Many studies have attempted to model oil price fluctuations, given its critical role in shaping the economy. Developing a robust model to predict oil price volatilities is not an easy task due to the intricacies and highly dynamic nature of the oil industry. World supply/demand, geopolitical conflicts, and unpredicted events (e.g., the COVID-19 outbreak) are examples of such intricacies. Mirmirani and Li [7] applied Vector Autoregression (VAR) and Artificial Neural Network (ANN) techniques to forecast the U.S. oil price and found out that the genetic algorithm-based ANN model outperforms the VAR model. In another study, Miao et al. [8] used the Least Absolute Shrinkage and Selection Operator (LASSO) regression method to identify the most influential factors in forecasting oil prices. Safari and Davallou [9] applied a combination of the Exponential Smoothing Model (ESM), Autoregressive Integrated Moving Average Model (ARIMA), and Nonlinear Autoregressive (NAR) Neural Network models in a state-space framework to account for the linear and nonlinear behaviors in crude oil prices. In another study, Abdollahi and Ebrahimi [10] applied a hybrid model, including the Adaptive Neuro Fuzzy Inference System (ANFIS), Autoregressive Fractionally Integrated Moving Average (ARFIMA), and Markov-switching models to forecast Brent crude oil price.

Given the complexities around the oil market, this study aims to understand the major players in the oil market by studying the market data from year 2002 to 2015. 27 parameters that are commonly known to be affecting oil price are considered in this paper. The data was collected from a study conducted by Miao et al. [8]. The data includes weekly records for West Texas Intermediate (WTI) crude oil price along with parameters across six broad dimensions: supply, demand, financial market, commodities market, speculative, and political factors. We applied the Multiple Linear Regression (MLR) model to identify the most influential factors in WTI crude oil price market. The result of this study offers only preliminary insight into the key players in the competitive oil market and can be used as the baseline for statistical modeling with higher accuracy.

2. Dataset

We use West Texas Intermediate (WTI) crude oil spot prices as the response, and 27 potential predictors classified into six broad groups. The description, sample frequency and data source for each predictor are shown in the Table 1 in Appendix. Our data spans the period from January 04, 2002 to September 25, 2015. There are 6 broad groups, and each group has several potential predictors.

Supply factors

Oil price heavily depends on supply and demand. We consider the following supply factors for our potential predictors [8]:

Global crude oil production: This factor includes both OPEC and non-OPEC crude oil production.

Global crude oil export: This factor measures the potential capacity for production; thus, this can be viewed as an additional measure of the global crude oil supply.

OPEC surplus crude oil production capacity: This factor is an indicator of general market supply conditions. Surplus production capacity can help mitigate the oil price fluctuation, especially due to the shortage of the supply, thus stabilize the global market.

Crude oil inventory: The accumulation of crude oil stock gives the market a great flexibility in responding to short-term supply shortages. we include both global and the U.S. crude oil closing stock.

U.S. refinery utilization rate: Refining rate plays an important role in determining crude oil prices because lower refinery utilization rate will lead to a preference for higher quality crude oil, putting upward pressure on prices.

Baltic exchange dirty and clean tanker index: Baltic exchange dirty tanker index indicates the cost of shipping unrefined petroleum oil, while clean tanker index indicates the cost of shipping refined products without heavy residual components. Lower values generally correspond with lower crude oil prices.

Demand factors

Demand factors have significant and positive influences on crude oil prices just like the supply factors. We consider the below predictors under the demand category [8]:

GDP (Gross Domestic Product): Global economic growth is closely related with the demand of oil. We consider the GDPs of the U.S., China, and Europe, which together account for more than 60% of the world GDP.

Kilian index: This index is an updated and corrected version of the index of global real economic activity in industrial commodity markets. Specifically, the index is based on dry cargo single voyage ocean freight rates.

Steel production: Steel production is a reliable indicator of global economic activity, and we use world steel production, as well as the production in the U.S., China and the Europe.

ISM manufacturing index: ISM manufacturing index is a monthly indicator of U.S. economic activity based on a survey of purchasing managers at more than 300 manufacturing firms. This index is widely viewed as an indicator of the business cycle.

Global crude oil imports: Global crude oil imports reflects the state of the economy. There is a positive relationship between global crude oil imports and the price of the crude oil.

Financial market factors

The relationship between financial market factors and oil prices can be complex. We consider the following three factors for this group [8]:

U.S. interest rates: Crude oil prices have usually a negative relationship with interest rates. We consider the three-month treasury bill rate and the federal fund rate.

Exchange rate: We use the U.S dollar index, computed as the weighted geometric mean of the dollar's value relative to other major currencies.

Stock market: We use S&P 500 index and MSCI world index for the U.S. and world equity markets. Stock market and oil prices are linearly related to each other.

Commodity market factors

Crude oil prices and other industrial commodities prices are correlated with each other. We consider two possible predictors in our forecasting models.

S&P GSCI non-energy index: This is a composite index of commodities that measures the performance of the commodities market. It consists of all commodities included in the S&P GSCI index, but the crude oil and natural gas.

CRB raw industrial materials index: This measures the aggregated price direction of 22 sensitive basic commodities whose markets are believed to be sensitive to changes in economic conditions.

Speculative factors

Speculative factor is the act of conducting a financial transaction that has substantial risk of losing value but also holds the expectation of a significant gain or other major value. Derivatives markets which spawn speculative activities are believed to exert an important influence on oil prices. We use the ratio of trading volume of crude oil futures contracts to the oil production as a potential predictor of the speculative factor.

Geopolitical factor

Crude oil price is sensitive to geopolitical events. We use the total number of terrorist attacks in the middle east Asia and north Africa as an indicator of geopolitical stability. This region is where most of the OPEC countries are located in.

3. Methodology

Multiple Linear Regression (MLR) Modeling

As previously mentioned in the introduction, the goal of this study is to find the predictors which have significant impact on WTI crude oil price. We started with 27 different predictors, and we assumed that these predictors are linearly correlated to the response. Hence, we chose the Multiple Linear Regression (MLR) model to work with. MLR is a statistical technique that uses several explanatory variables to predict the outcome of a response variable.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

Y: Response, dependent variable

X: Predictor, independent variable

β: Coefficient

ε: Error

Model Selection

First, we preprocessed the dataset. The potential predictors with a lot of missing data have been removed from the dataset, and the predictors which turned out to be not necessary for MLR modeling have also been removed from the original dataset. Then, we applied MLR modeling to the remaining potential predictors with the WTI response. We expected some level of collinearity between these potential predictors because most of them, if not all, are the factors or indices of the world economy. Since we chose these predictors because they are related to the global economy as well as the crude oil price, there has to be certain level of correlations with each other. So, first we checked the pairwise collinearity. Due to the background of the predictor selection, we set a higher criterion for collinearity level at 0.6 and removed below 11 predictors with high collinearity:

- Consumer Price Index
- Steel Production in China
- MSCI World Index
- U.S. Crude Oil Closing Stock
- US Dollar Index
- CRB Raw Materials Index
- Crude Oil Exports
- Three-month U.S. Treasury Bill Rate
- Global Crude Oil Production
- Baltic Dirty Index
- Global Crude Oil Imports

After removing the predictors with high collinearity, we applied criterion-based selection to find the best model. We used AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and Adjusted R-squared. AIC and BIC provide measures of model performance that account for model complexity. AIC and BIC combine a term reflecting how well the model fits the data with a term that penalizes the model in proportion to its number of parameters.

$$AIC = 2(p+1) - 2logL(\beta)$$
$$BIC = (p+1)\,log(n) - 2logL(\beta)$$
$$p{:}\,number\ of\ predictors$$
$$n{:}\,number\ of\ data$$

Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model.

$$R_a^2 = 1 - \frac{RSS/(n-(p+1))}{TSS/(n-1)}$$

R-squared result shows that the model with 15 predictors is the best model and AIC tells us that the model with 14 predictors is the best, while BIC result suggests that the model with 10 predictor is the best model. Since both AIC and R squared values converges after the model with 10 predictors, we concluded that the model with 10 predictors is the best model based on the criterion-based selection results.

4. Results and Discussion

The most influential factors

The final model includes 10 predictors which are ordered according to their estimate's absolute value (sign): Europe GDP growth (-), US GDP growth (+), U.S. based federal funds rate (-), OPEC surplus crude oil production capacity (-), capacity utilization rate (-), ISM manufacturing index (+), terror (+), S&P GSCI non-energy index (+), U.S. steel production (+) and global crude oil closing stock (-).

Our analysis showed that the real GDP growth rates of Euro countries has the largest association with the weekly WTI prices among other predictors. However, this association was found to be negative, meaning a one unit increase in real GDP growth rate is associated with a decrease in oil price holding other predictors constant. While this negative correlation seems unexpected at first sight, other studies have found similar adverse relationship between GDP growth rate and oil price. Rodriguez and Sanchez [11] studied the effects of oil price shocks on the real economic activities of the world's main industrialized countries and reported negative impact of oil price increase on economic activity in all cases except Japan.

US GDP growth was found to have the highest positive correlation with WTI weekly prices. This implies a one unit increase in real GDP growth rate is associated with an increase in oil price holding other predictors constant. Similar positive associations were found for ISM manufacturing index and U.S. steel production with WTI price. Together with US GDP growth rate, these predictors are categorized as demand-related factors in this paper. Gadea et al. [12] found significant effects between changes in oil price and US GDP growth rate. However, they reported that this effect has declined over time [12]. Miao et al. [8] conducted a study to identify the influential factors in crude oil price forecasting and reported ISM manufacturing index as one of the most important variables in their analyses. Baffes [13] examined the effect of crude oil prices on the price of internationally traded primary commodities and found that the price of precious metals are strongly correlated with crude oil price.

Amongst the five financial factors included in this study, the U.S. based federal funds rate was the only factor that was found to have the highest explanatory power. A negative correlation between the federal rate and WTI price was observed. On the contrary, in the study conducted by Miao et al. [8] US federal rate was not found to have strong impact on the forecasting power of the oil price model.

As expected, supply-related factors in the final model, naming OPEC surplus crude oil production capacity, capacity utilization rate and stock predictors appeared to have negative correlations with weekly WTI price. Pierru et al. [14] studied the impact of OPEC on oil price and found that OPEC's use of spare capacity has decreased price volatility by nearly half. Dees et al. [15] reported that the refinery capacity utilization rate has a significant explanatory power in their developed oil price prediction model.

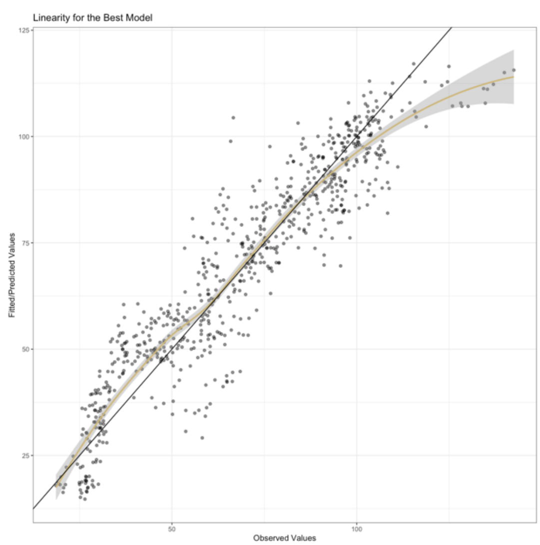
S&P GSCI non-energy index (commodity market factor) and total number of terrorist attacks in the Middle East and North Africa (political factor) were found to have positive association with WTI prices. In a study by Dutta and Noor [16] to examine the volatility effects between oil and three different non-energy commodity markets, they found linkage between oil market volatility to both metal and non-energy aggregate markets. Coleman [17] studied the relationship between crude oil prices and market and geopolitical measures and reported positive links between terrorist activities and oil prices.

Diagnostics

Linear model assumptions

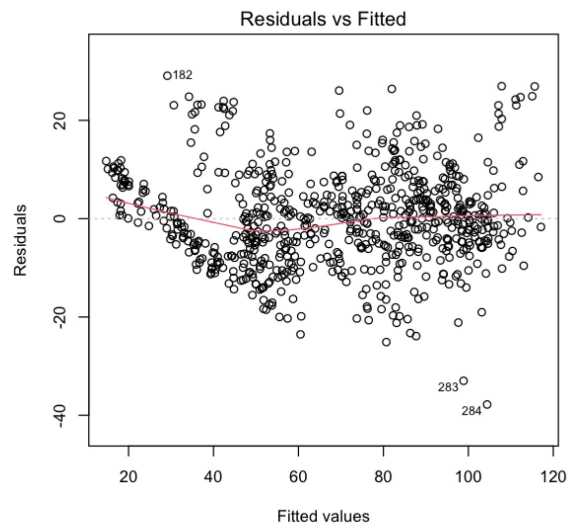
We performed diagnostic tests to detect violations to the linearity, homoscedasticity (constant variance), normality, and independence assumptions on a regression model on real data.

Linearity



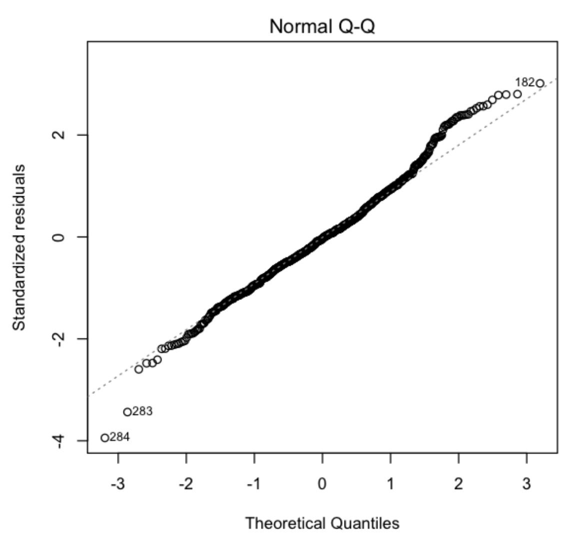
The observed vs predicted value plot should follow the black line, y = x, instead, the plot shows some curvature, as captured by the gold curve. This curvature shows that, for high values of the response, the model is overpredicting, however other than that, there is a linear relationship between predicted and observed values, and this suggests that the model meets linearity assumption.

Homoscedasticity



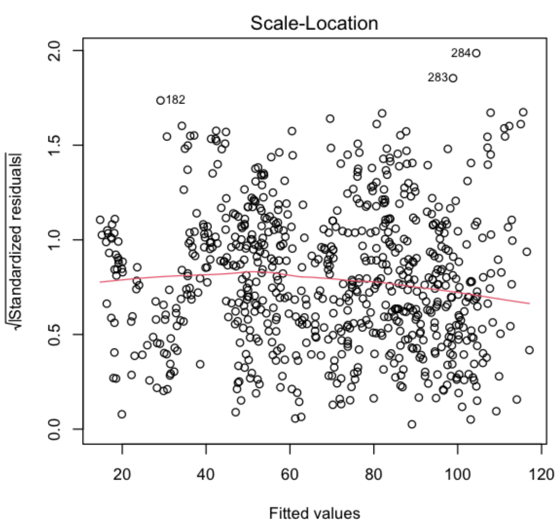
We noted that, other than the outliers (very small residual), we don't seem to see a difference in variability of the residuals across the fitted values. So, we do not have evidence of a violation of homoscedasticity.

Normality



There seems some deviation from normality, especially in the tails of the distribution. This non-normality is likely due to the violation of the linearity assumption in the tails. However, other than the tail part, we do not have evidence of a violation of normality.

Independence



There are no specific patterns in this plot, thus we do not have evidence of a violation of independence.

Multicollinearity

Even though we checked the pairwise collinearity, we tested for multicollinearity to make sure there is no skewness or misleading results. We applied Variance Inflation Factors (VIF) method. Small VIF values indicate low correlation among variables under ideal conditions VIF < 3. However, it is acceptable if it is less than 10. All the VIF of each predictor are well below 3, and this shows that there is no correlation among each predictor and the remaining predictors, and hence the variance of each predictor is not inflated at all.

5. Conclusion

Given the critical importance of crude oil in the world economy and the complexities around its market, this study aimed to understand the major players in the oil market by studying the data from 2002 to 2015. More than 25 parameters that are commonly known to be affecting oil price were considered in this study. Applied the MLR model, we found that 10 predictors have the most impact on the market. These factors are spanning across 5 different categories: supply, demand, financial market, commodity market, and geopolitical factors.

While the results of this study are in good agreement with the available literature, some limitations need to be considered. The data used for this study was collected from a research paper authored by Miao et al. [8]. Although the original data was collected through reliable sources, some errors in the dataset might exist. A portion of the data might be misleading and biased due to several factors such as government corruption and sampling errors. It is also expected that some of the variance in the data might be left unexplained due to excluding factors that were left outside of this study. Example may include other geopolitical factors and unexpected events (e.g., pandemics). Some of the study outputs might be also affected by high sensitivity to sampling data frequency and data interval selection which is a common weakness of crude oil price forecasting models [18]. It is worth noting that these are common limitations to oil and gas price modeling studies and the validity of the results, to the best of authors knowledge, will not be compromised due to these imitations.

Future work may include oil price forecasting analyses with higher orders of accuracy that can better capture the dynamics of the market. Future work might also investigate the time-dependent behaviors in oil price market with applying time series analysis.

References

[1] R. Jammazi and C. Aloui, "Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling," *Energy Econ.*, vol. 34, no. 3, pp. 828–841, 2012, doi: 10.1016/j.eneco.2011.07.018.

[2] M. Khalid, S. Sarwar, R. Waheed, and M. Amir, "Role of Energy on Economy The Case of Micro to Macro Level Analysis," *Econ. Bull.*, vol. 38, no. 4, pp. 1905–1926, 2018.

[3] M. R. Farzanegan and G. Markwardt, "The effects of oil price shocks on the Iranian economy," *Energy Econ.*, vol. 31, no. 1, pp. 134–151, Jan. 2009, doi: 10.1016/J.ENECO.2008.09.003.

[4] U.S. Energy Information Administration, "International energy outlook," 2010.

[5] W. Kang and R. A. Ratti, "Oil shocks, policy uncertainty and stock market return," *J. Int. Financ. Mark. Institutions Money*, vol. 26, no. 1, pp. 305–318, Oct. 2013, doi: 10.1016/J.INTFIN.2013.07.001.

[6] W. Tang, L. Wu, and Z. X. Zhang, "Oil price shocks and their short- and long-term effects on the Chinese economy," *Energy Econ.*, vol. 32, no. SUPPL. 1, pp. S3–S14, Sep. 2010, doi: 10.1016/J.ENECO.2010.01.002.

[7] S. Mirmirani and H. Cheng Li, "A comparison of VAR and neural networks with genetic algorithm in forcasting price of oil," *Adv. Econom.*, vol. 19, pp. 203–223, 2004, doi: 10.1016/S0731-9053(04)19008-7/FULL/XML.

[8] H. Miao, S. Ramchander, T. Wang, and D. Yang, "Influential factors in crude oil price forecasting," *Energy Econ.*, vol. 68, pp. 77–88, 2017, doi: 10.1016/j.eneco.2017.09.010.

[9] A. Safari and M. Davallou, "Oil price forecasting using a hybrid model," *Energy*, vol. 148, pp. 49–58, Apr. 2018, doi: 10.1016/J.ENERGY.2018.01.007.

[10] H. Abdollahi and S. B. Ebrahimi, "A new hybrid model for forecasting Brent crude oil price," *Energy*, vol. 200, p. 117520, Jun. 2020, doi: 10.1016/J.ENERGY.2020.117520.

[11] R. Jiménez-Rodríguez, M. Sánchez, R. Jime´nezjime´nez-Rodrı´guezrodrı´guez, and M. Sa´nchezsa´nchez, "Oil price shocks and real GDP growth: empirical evidence for some OECD countries," *https://doi.org/10.1080/0003684042000281561*, vol. 37, no. 2, pp. 201–228, Feb. 2006, doi:

10.1080/0003684042000281561.

[12] M. D. Gadea, A. Gómez-Loscos, and A. Montañés, "Oil Price and Economic Growth: A Long Story?," *Econom. 2016, Vol. 4, Page 41*, vol. 4, no. 4, p. 41, Oct. 2016, doi: 10.3390/ECONOMETRICS4040041.

[13] J. Baffes, "Oil spills on other commodities," *Resour. Policy*, vol. 32, no. 3, pp. 126–134, Sep. 2007, doi: 10.1016/J.RESOURPOL.2007.08.004.

[14] A. Pierru, J. L. Smith, and T. Zamrik, "OPEC's Impact on Oil Price Volatility: The Role of Spare Capacity," *Energy J.*, vol. 39, no. 2, pp. 103–122, 2018, doi: 10.5547/01956574.39.2.APIE.

[15] S. Dees, A. Gasteuil, R. Kaufmann, and M. Mann, "Assessing the Factors Behind Oil Price Changes," *SSRN Electron. J.*, Jan. 2008, doi: 10.2139/SSRN.1080247.

[16] A. Dutta and M. Hasib Noor, "Oil and non-energy commodity markets: An empirical analysis of volatility spillovers and hedging effectiveness," *http://www.editorialmanager.com/cogentecon*, vol. 5, no. 1, Jan. 2017, doi: 10.1080/23322039.2017.1324555.

[17] L. Coleman, "Explaining crude oil prices using fundamental measures," *Energy Policy*, vol. 40, no. 1, pp. 318–324, Jan. 2012, doi: 10.1016/J.ENPOL.2011.10.012.

[18] Q. Lu, S. Sun, H. Duan, and S. Wang, "Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model," *Energy Informatics*, vol. 4, no. 2, pp. 1–20, Sep. 2021, doi: 10.1186/S42162-021-00166-4/TABLES/4.

Appendix

Table A.1: Time series data for individual crude oil price indicators [8].

Factor Group	Individual indicator	Description	Frequency	Source
Crude oil price (Response)	WTI	Nominal West Texas Intermediate spot price	daily	Energy Information Administration
Supply	Global Production	Global crude oil production	monthly	JODI-Oil Database
	Global Stock	Global crude oil closing stock	monthly	JODI-Oil Database
	Global Export	Global crude oil export	monthly	JODI-Oil Database
	OPEC Surplus	OPEC surplus crude oil production capacity	monthly	Bloomberg database
	US Stock	US crude oil closing stock	monthly	Bloomberg database
	Capacity Utilization Rate	Rate of refining capacity utilization.	weekly	Energy Information Administration
Demand	Baltic Dirty	Baltic exchange dirty tanker index	daily	Bloomberg database
	Kilian Index	Monthly measure of log real global economic activity calculated from monthly change reported by Kilian (2009)	monthly	www.personal.umich.edu/
	GDP Growth China	Growth rate of China real GDP	quarterly	OECD database
	GDP Growth US	Growth rate of US real GDP	quarterly	OECD database
	GDP Growth EURO	Growth rate of EURO real GDP	quarterly	OECD database
	Steel World	World steel production	monthly	Bloomberg database
	Steel China	China steel production	monthly	Bloomberg database
	Steel US	US steel production	monthly	Bloomberg database
	Steel EURO	EURO steel production	monthly	Bloomberg database
	Global Import	Global crude oil import	monthly	JODI-Oil Database
	ISM	ISM manufacturing index	monthly	Bloomberg database
Financial	Federal Funds Rate	Federal Funds Rate, U.S. base rate	daily	Bloomberg database
	T-Bills	Three-month U.S. Treasury bill rate	daily	Bloomberg database
	SP500	S&P 500 index	daily	Bloomberg database
	DXY	US dollar index	daily	Bloomberg database
	MSCI	MSCI world index	daily	Bloomberg database
	GSCI	S&P GSCI non-energy index	daily	Bloomberg database

Commodity market	CRB Rind	CRB raw materials index	daily	Bloomberg database
Speculative	Ratio	The ratio of trading volume of oil futures contracts to global oil production	daily	Quandl database
Political	Terrorist	Total amount of terrorist attacks in the Middle East and North Africa	daily	Global Terrorism Database

R-Code

```
# Import Library and the Data

library(leaps);library(MASS);library(car);library(ggplot2);library(corrplot);library(dplyr)

data = read.csv('/Users/jongbaeyoon/Documents/Documents/CU Boulder/Spring, 2022/STAT 5010_Statistical Mathematics and Application 2/Project/Data/Final_data_set.csv')


# Check pairwise collinearity

df = data.frame(data[,-1]) # Remove "date"

col4 = colorRampPalette(c("black", "darkgrey", "grey","#CFB87C"))

corrplot(cor(df[,2:28]), method = "ellipse", col = col4(100),  addCoef.col = "black", tl.col = "black")


# Remove the predictors with high collinearity (greater than 0.6)

df.col = select(df, -c("CPI", "STEEL_CHI", "MSCI", "CRUDEINV_USA", "DXY", "CRB_RIND", "Exports", "Tbill","Production", "Baltic_d", "Imports"))

corrplot(cor(df.col[,2:17]), method = "ellipse", col = col4(100),  addCoef.col = "black", tl.col = "black")


# Usage of regsubsets for finding out the best combination based on RSS

n = dim(df.col)[1];

reg1 = regsubsets(WTI ~ ., data = df.col, nvmax = 16 )

rs = summary(reg1)

rs$which


# Plotting R-squared vs # of predictors to compare the combinations based on R-squared

plot(1:16, rs$adjr2, xlab = "number of predictors", ylab = "adjusted R-squared")

which(rs$adjr2 == max(rs$adjr2))
```

```
# Plotting AIC vs # of predictors to compare the combinations
AIC = 2*(2:17) + n*log(rs$rss/n)
plot(AIC ~ I(1:16), xlab = "number of predictors", ylab= "AIC") ;AIC[2]
which(AIC == min(AIC))

# Plotting BIC vs # of predictors to compare the combinations
BIC = log(n)*(2:17) + n*log(rs$rss/n)
plot(BIC ~ I(1:16), xlab = "number of predictors", ylab = "BIC");BIC[2]
which(BIC == min(BIC))

# Update the data frame and the linear model in accordance with the latest 10-paramter model.
df.col.BIC = select(df.col, -c("Baltic_c", "STEEL_EURO", "SP500", "Kilian", "china_gdp", "ratio"))
lmod_col_BIC = lm(WTI ~ ., df.col.BIC)
summary(lmod_col_BIC)

# Check multicollinearity
corrplot(cor(df.col.BIC[,2:11]), method = "ellipse", col = col4(100),  addCoef.col = "black", tl.col = "black")
vif(lmod_col_BIC)

# Diagnostic Test
plot(lmod_col_BIC)

# Linearity
ggplot(df.col.BIC, aes(x=data$WTI, y=fitted(lmod_col_BIC))) +
   geom_point(alpha = 0.5) +
   geom_smooth(se = T, col = "#CFB87C")+
   geom_abline(intercept = 0, slope = 1)+
   xlab("Observed Values")+ylab("Fitted/Predicted Values")+
   ggtitle("Linearity for the Best Model")+theme_bw()
```