Analysis of Impact of World events on Crude Oil Prices

Syed Farhan Naqvi

Contents

Chapter 1

ABSTRACT

Crude oil is one of the most important fuel sources, it is especially important to businesses that heavily rely on fuel, such as airlines, plastic producers, and agricultural businesses. If we look at the evolution of the oil prices, its price isn’t the same as it was 25 years ago. Crude oil price fluctuation has directly impacted the economic stability. There have been many factors that led to its fluctuation in the past and will so continue in the future too. Application of spline regression has shown great promise in time series analysis and so we will also use it to predict oil prices. This case study shall focus on capturing the fluctuations caused due to financial, commodity factors alongside geopolitics, emergencies and policy changes.

Chapter 2

Introduction

Crude Oil is a very important and critical fuel source to the international community. It has contributed to over one-third of the globe's energy consumption. Oil is critical to businesses that rely heavily on fuel, such as Airlines, Automobile Industry, Farming, etc. Due to its such widespread importance, the evolution of crude oil prices has fluctuated a lot in the past and will further continue in the future too. Its uncertain fluctuation certainly impacts the economic stability of many countries. Crude oil value is affected by financial, commodity factors alongside geopolitics.

Commodity factors

The long-term trend of oil prices was always determined by demand and supply. If demand used to be more and supply used to be less, crude oil prices used to rise. If demand is less and supply is more, crude oil prices used to drop. OPEC – Organisation of Petroleum Exporting Countries – is a cartel made up of 14 countries who export petroleum. It means that the countries within OPEC have come together to regulate the price of oil by controlling supply. It regulates oil production via quotas, which ensures its members get a good price for their oil even if this means producing less in the short term. Outside of OPEC, the world's biggest oil producing countries are the USA, Canada and China. OPEC's decision to cut or increase the supply directly affect the oil prices.

Financial factors

Oil futures contracts are agreements to buy or sell oil at a specific date in the future for an agreed-upon price. Oil derivatives are securities that are based on the underlying price of oil and traded on the exchanges. Commodities traders fall into two categories: hedgers and speculators.

- Hedgers

  Hedgers are representatives of companies who either actually produce oil and those that consume oil for delivery at a future date and fixed price. That way, they know the price of the oil and can plan for it financially—the contracts set the price for the buyer and seller, reducing risk for their companies when prices are rising and falling.

- Speculators

  Traders in the second category are speculators. Their only motive is to make money from changes in the price of oil. Futures speculators are generally the ones that are interested in oil derivatives, trading on small incremental changes in prices.

The price of oil can directly or indirectly increase or decrease costs of goods and services in the economy and can result in inflation. The price of oil can cause a ripple effect in the economy through industries that use crude oil as a raw material. If oil prices rise, it can increase the cost of living for individuals, putting a pressure on wage rates to make up for this. The high labor costs increase the costs of goods and services and increase the rate of inflation in the economy as each dollar can now buy fewer goods and services. When inflation rises the stock prices of the industry decline.

Geopolitical Factors

The geopolitical events have been a major factor in determining its prices since the beginning. Some of the main events are depicted in the following graphs.

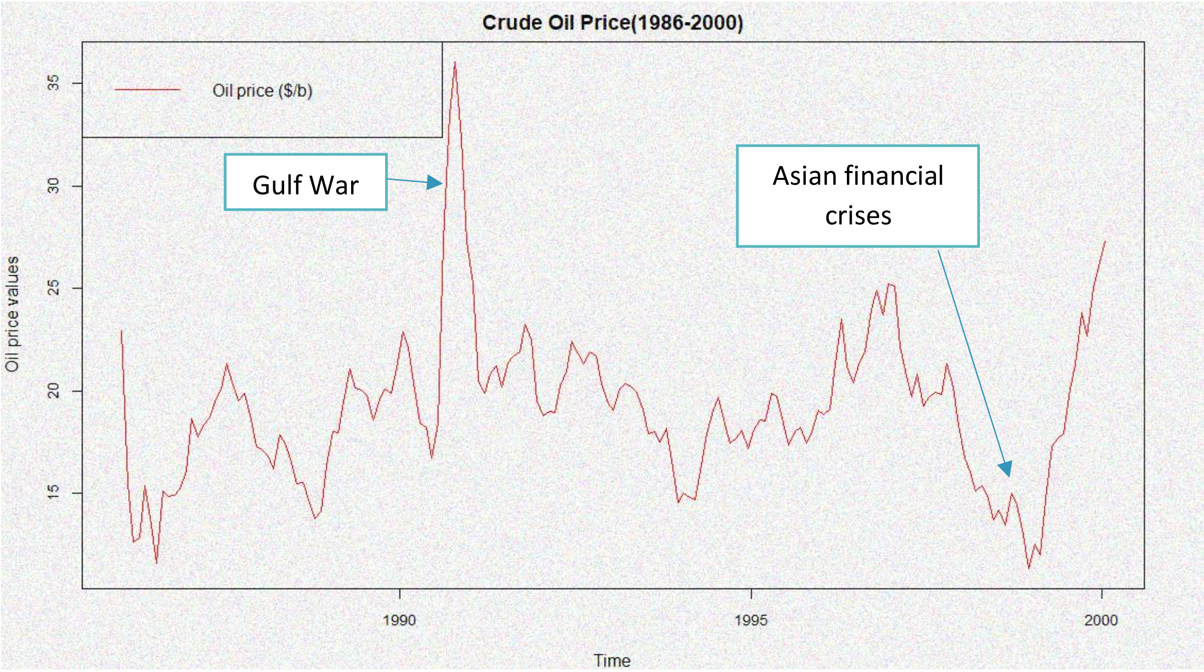


Figure: Events affecting crude oil prices (1986-2000)

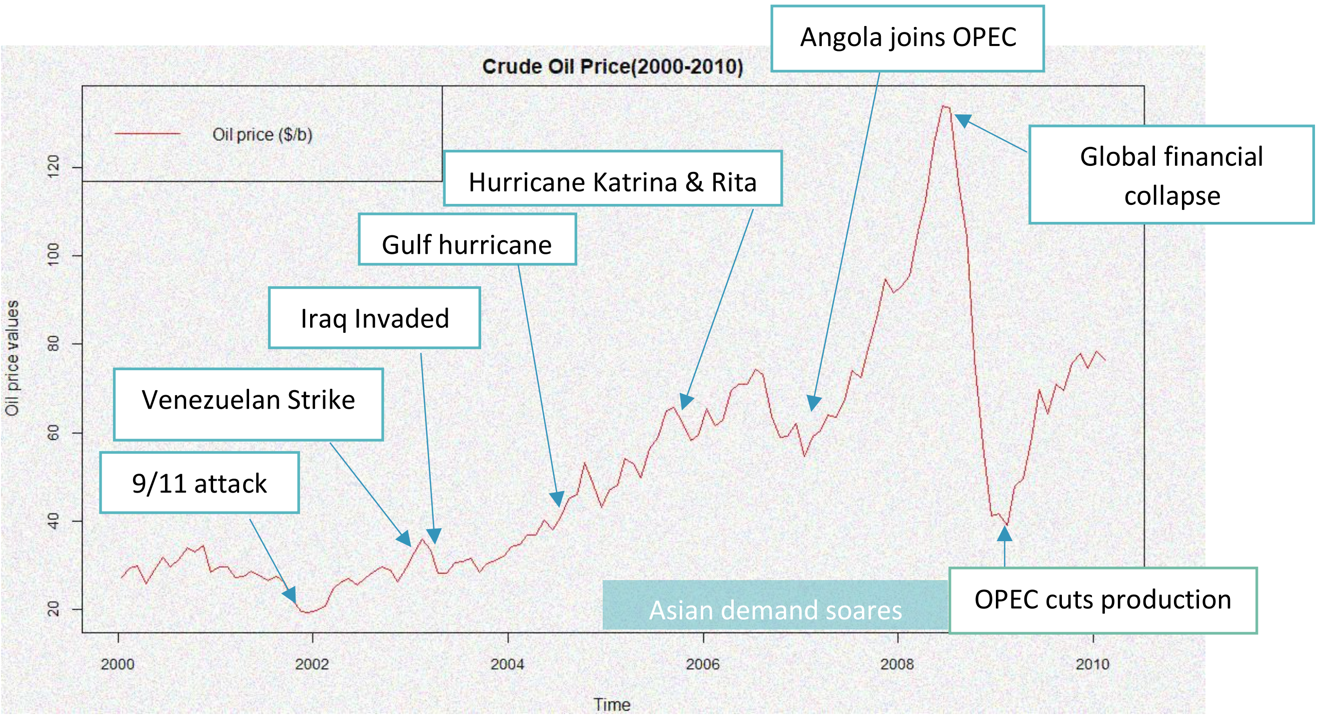


Figure: Events affecting crude oil prices (2000-2010)

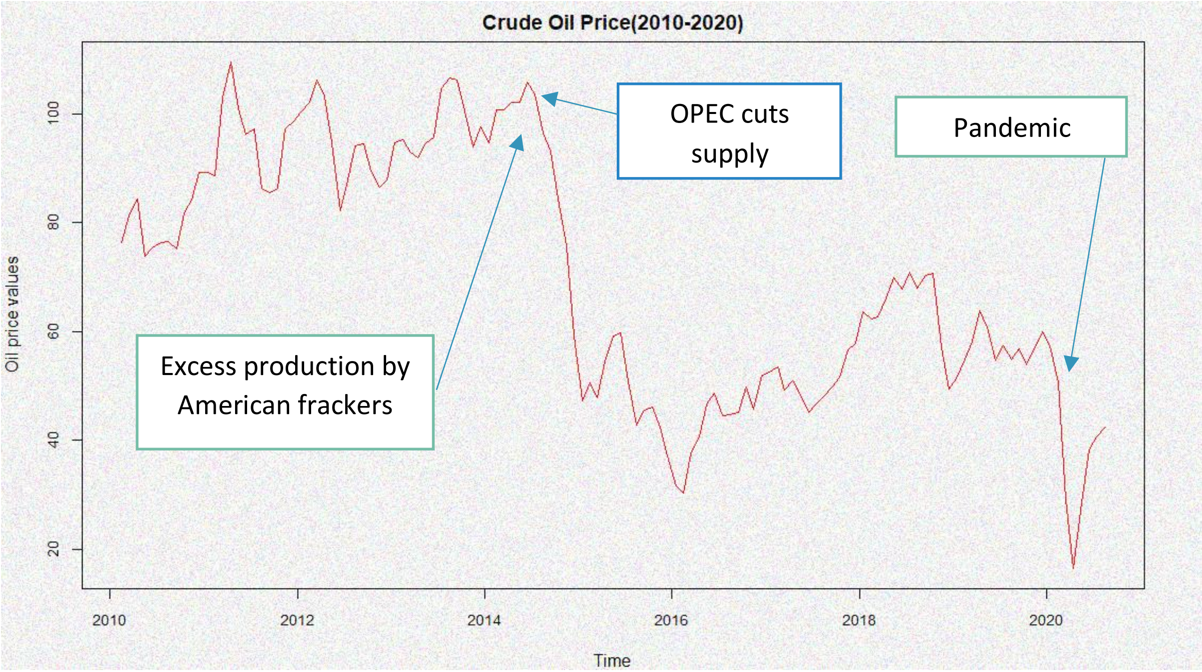


Figure: Events affecting crude oil prices (2010-2020)

In the next section, we shall be using a dataset containing crude oil prices and after exploring it, we will implement spline regression model which will be useful in predicting the oil prices. If predictions are close enough to observed prices, then our model can be useful to determine future oil prices too.

Chapter 3

Methodology

3.1 Data Collection

There is a dataset that consists of West Texas Intermediate (WTI) Crude Oil Prices from 15th January 1986 till 15th August 2020. It includes monthly prices of WTI. It was extracted from DataHub.io - a verified site by SLA. The dataset didn't require any cleaning as it was in a good shape. We will be performing analysis on this data. But before that we need to explore it.

3.2 Data Exploration

The dataset consists of two fields namely- Date and Price. The "Date" field consists of dates recorded on the 15th of every month from 1986-2020. The "Price" field consists of cost of crude oil in dollars per barrel ($/b). R programming language will be for data exploration and applying the spline regression model later.

```
head(wti) #displaying the first 5 values of dataset
tail(wti) # displaying the last 5 rows of the dataset
typeof(wti[,1])
typeof(wti[,2])
```

The first step towards data exploration is getting to know your data. We used 'head(wti)' to display the first 5 rows of our dataset and 'tail(wti)' to display the last 5 rows. We can get to know how many columns and what type of data might be used just by giving a look at it. The typeof() function allows us to understand what type of data is actually stored in each column.

```
> head(wti)
        Date Price
1 1986-01-15 22.93
2 1986-02-15 15.46
3 1986-03-15 12.61
4 1986-04-15 12.84
5 1986-05-15 15.38
6 1986-06-15 13.43
```

head() command displays first 5 rows of the dataset

```
> tail(wti)
          Date Price
411 2020-03-15 29.21
412 2020-04-15 16.55
413 2020-05-15 28.56
414 2020-06-15 38.31
415 2020-07-15 40.71
416 2020-08-15 42.34


> typeof(wti[,1])
[1] "character"

> typeof(wti[,2])

[1] "double
```

tail() command displays last 5 rows of the dataset

Our dataset consists of 416 rows and 2 columns. The Date column consists of data that is of character type and Price column's data is of double type

```
date=as.Date(date) # converting the character data to dates
typeof(date)
```

```
> date=as.Date(date) # converting the character data to dates
> typeof(date)
[1] "double"
```

The data that we have is a time-stamped data. We used 'as.Date()' to character data to dates. The crude oil prices that were recorded over a time interval is used to track its change over the interval. Thus, we can convert the current data into time series.

The "quantmod" package contains the "xts" function that allows you to do it.

xts(x, order.by = index(x))

The function consists of :-

- x = Object containing time series data
- order.by = a corresponding vector of unique times/dates (must be of known time-based class.)

Using the following code, the data was converted to time-series data :-

```
library(quantmod)
timeS<-xts(wti[2],order.by= as.Date(wti$Date))
```

After successfully executing the above code, it was time for visualizing all of the data that we had. We used the plot() function to display the data in the form of a graph.

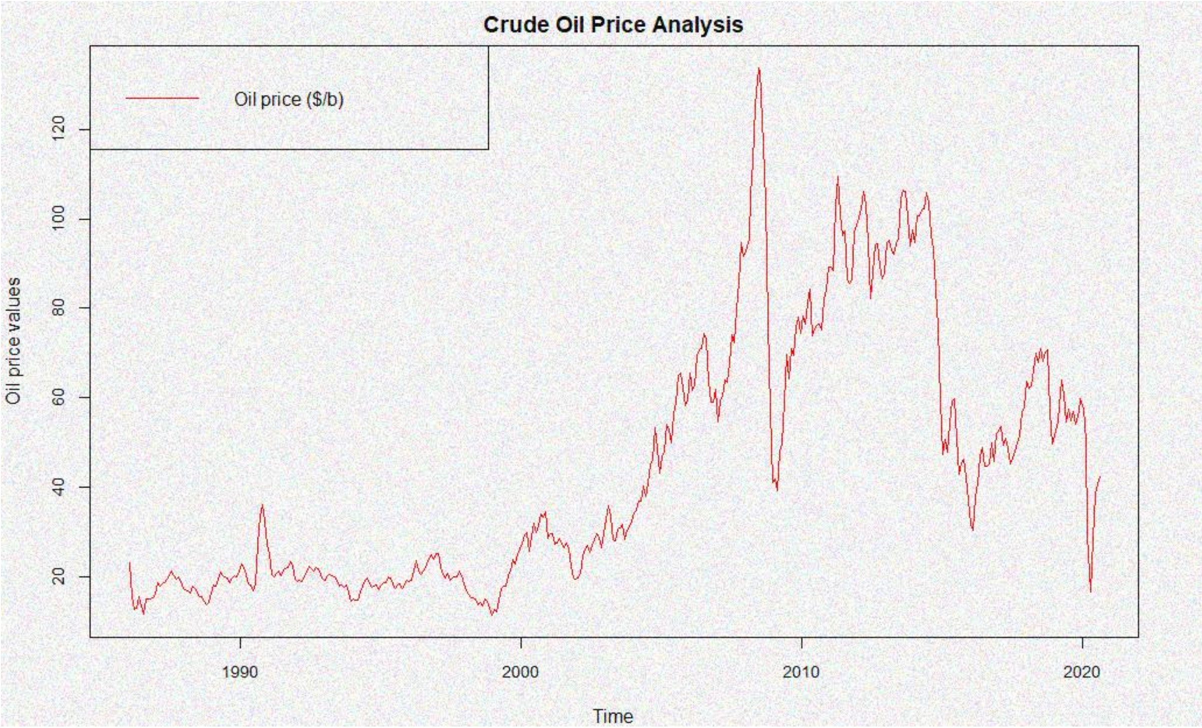


Fig : Crude Oil prices data

The above plot reveals that there were many dips and gains in the price of Crude Oil that led to fluctuation multiple times over a span of 34 years.

The statistical measures for oil prices can be calculated using the "summary" function.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
11.35	19.77	31.41	44.06	62.70	133.88

Table: Summary Of crude oil prices

3.3 Data Analysis

a) Spline regression

Spline Regression is one of the non-parametric regression machine learning models. In this technique the dataset is divided into bins at intervals, called as knots and each bin has its separate fit. Spline is a special function defined piece-wise by polynomials. It refers to a wide class of functions that are used in applications requiring data interpolation and/or smoothing.

A simple linear regression fit would prove to be inefficient as there is no specific linear trend to our data. A spline regression fit would be considered as better option compared to polynomial regression because our time series data has varying oil prices and there are multiple unexpected fluctuations without any specific trend. Polynomial regression only captures a certain amount of curvature in a nonlinear relationship whereas spline is a series of polynomial segments strung together. As we would choose the knots, polynomial regression would be applied between those knots.

The default spline model in R is a third-degree polynomial and that's exactly what we are going to use. This is mainly because it is hard for the eye to detect the discontinuity at the knots. To apply the spline regression fit, we chose the following knot points first.

```
point<-c(1,8,20,35,59,133,157,180,194,248,254,271,279,305,363,375,392,416)
```

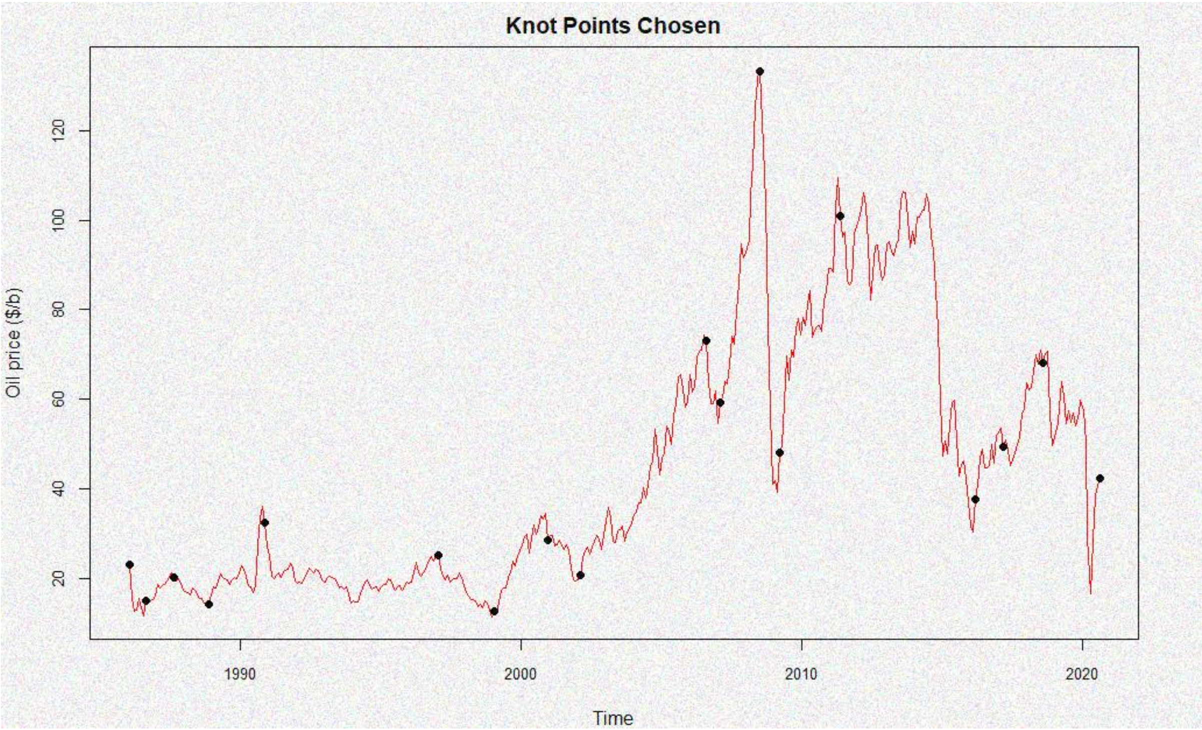


Figure: Key points chosen for spline regression

After choosing the knot points, we used the "lm" function, which is used to create a regression model. Over here we created a cubic regression model with the knot points. The dependent variable is the crude oil value and date is the independent variable.

```
for (i in c(1:(length(point)-1))){

  # we are going to be fitting a model using lm function
  fit<-lm(wti[point[i]:point[i+1],3]~((I(wti[point[i]:point[i+1],1])+I(wti[point[i]:point[i+1],1]^2)+I(wti[point[i]:point[i+1],1]^3))))
  wti[point[i]:point[i+1],4]<-fit$fitted
  wti[point[i]:point[i+1],5]<-fit$residuals
  }

 names(wti)[4]<-"Fit"
 names(wti)[5]<-"Res"
 head(wti)
```

```
  index       Date Price      Fit        Res
1     1 1986-01-15 22.93 22.17864  0.7513636
2     2 1986-02-15 15.46 16.49193 -1.0319264
3     3 1986-03-15 12.61 13.53589 -0.9258874
```

```
4      4 1986-04-15 12.84 12.57082  0.2691775
5      5 1986-05-15 15.38 12.85703  2.5229654
6      6 1986-06-15 13.43 13.65483 -0.2248268
```

We have generated coefficients of 17 equations using the above code. They are as follows:-

Equation	Intercept	Degree1	Degree2	Degree3
1	31.3357	-11.1388	2.1050	-0.1233
2	26.45346	-3.34901	0.29305	-0.00701
3	108.256686	-9.409522	0.330731	-0.003977
4	-7.635e+02	5.232e+01	-1.157e+00	8.467e-03
5	2.481e+01	4.123e-01	-1.022e-02	5.363e-05
6	-4.914e+02	1.185e+01	-8.751e-02	2.070e-04
7	4.188e+03	-8.198e+01	5.273e-01	-1.112e-03
8	-5.889e+04	9.374e+02	-4.966e+00	8.758e-03
9	2.262e+03	-2.905e+01	1.219e-01	-1.625e-04
10	3.359e+06	-3.996e+04	1.585e+02	-2.094e-01
11	-1.278e+05	1.500e+03	-5.874e+00	7.683e-03
12	-8.205e+06	9.010e+04	-3.297e+02	4.021e-01
13	-3.447e+05	3.543e+03	-1.214e+01	1.386e-02
14	2.371e+04	-2.266e+02	7.235e-01	-7.690e-04
15	-6.211e+05	5.027e+03	-1.356e+01	1.219e-02
16	1.269e+06	-9.914e+03	2.581e+01	-2.240e-02
17	1.430e+05	-1.075e+03	2.698e+00	-2.259e-03

From the generated coefficients, equations were created to obtain the best fit. We can see the actual prices trend in black below the predicted values trend in red. This is followed by the graph depicting residuals trend in blue.

Oil Prices

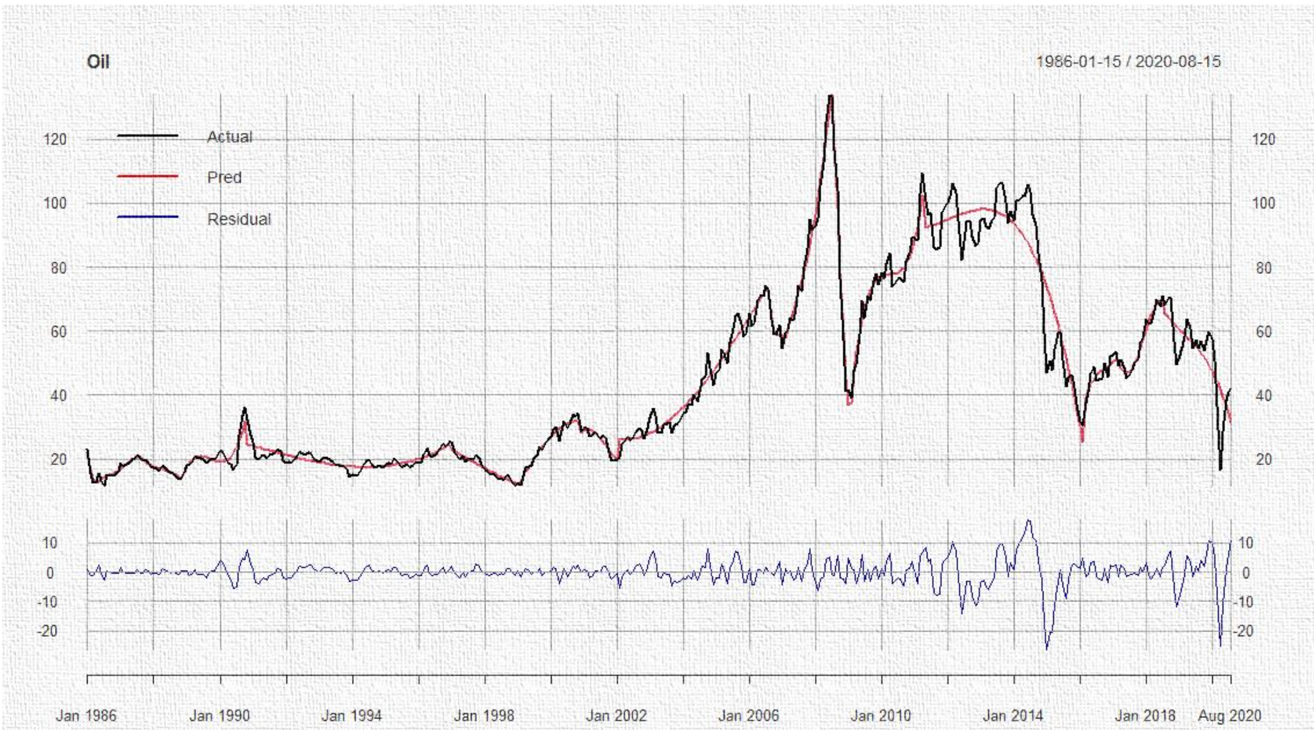


Figure: Crude oil spline regression fit

B. First order derivative

The first order derivative of a function measures its sensitivity with respect to the independent variable. We can find a derivative of a function in R using the "D" function. The following formula was used to obtain derivatives of seventeen cubic expressions(whose coefficients we got in the previous section) between the knot points.

Formula :-

D(function's expression, x)

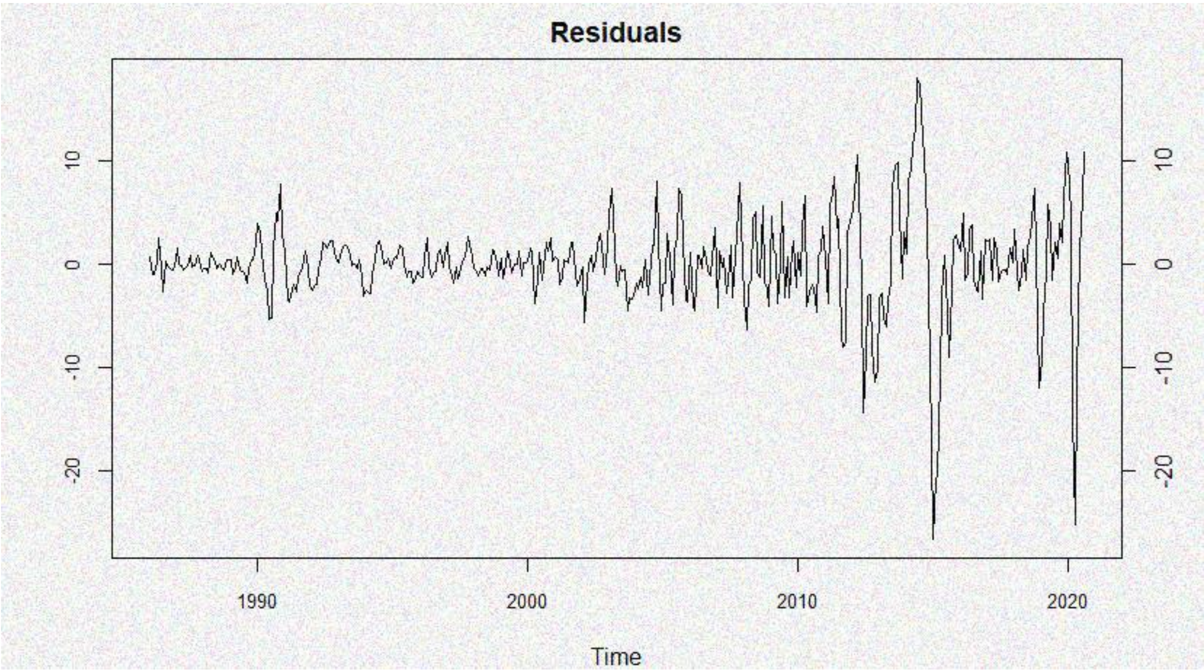
x: variable by which the expression is differentiated

C. Second order derivative

The second order function is used to indicate the rate of change of a function with respect to the independent variable. It will indicate the pace at which oil prices are fluctuating in this stipulated time period. I will again be using the "D()" function on the first order derivatives-that I obtained-to compute the second order derivatives of the functions.

D. Residuals

Residuals in a machine learning model are the differences obtained between the observed and the predicted values. The outcome obtained after applying the Machine learning model on the original data points indicates how well your model's predictions are. After observing the following graph we can conclude that the model came up with predictions that were close to the observed crude oil prices.



Chapter 4
Results

- ## First order derivatives

Function number	First order derivative
1	-11.1388 + 2.1050 * (2 * x) -0.1233 * (3 * x^2)
2	0.29305* (2 * x) -3.34901 - 0.00701 * (3 * x^2)
3	-9.409522 + 0.330731 * (2 * x) -0.003977 * (3 * x^2)
4	-1.157e+00 * (2 * x) + 5.232e+01 + 8.467e-03 * (3 * x^2)
5	4.123e-01 -1.022e-02 * (2 * x) + 5.363e-05 * (3 * x^2)
6	1.185e+01 -8.751e-02 * (2 * x) + 2.070e-04 * (3 * x^2)
7	-8.198e+01 + 5.273e-01 * (2 * x) -1.112e-03 * (3 * x^2)
8	9.374e+02 -4.966e+00 * (2 * x) + 8.758e-03 * (3 * x^2)
9	-2.905e+01 + 1.219e-01 * (2 * x) -1.625e-04 * (3 * x^2)
10	1.585e+02 * (2 * x) -3.996e+04 -2.094e-01 * (3 * x^2)
11	1.500e+03 -5.874e+00 * (2 * x) + 7.683e-03* (3 * x^2)
12	9.010e+04 -3.297e+02* (2 * x) + 4.021e-01 * (3 * x^2)
13	3.543e+03 -1.214e+01* (2 * x) +1.386e-02 * (3 * x^2)
14	-2.266e+02+ 7.235e-01 * (2 * x) -7.690e-04 * (3 * x^2)
15	5.027e+03 -1.356e+01* (2 * x) +1.219e-02 * (3 * x^2)
16	-9.914e+03 + 2.581e+01 * (2 * x) -2.240e-02* (3 * x^2)
17	-1.075e+03 +2.698e+00* (2 * x) -2.259e-03 * (3 * x^2)

Table: First order differentiation of the 17 equations

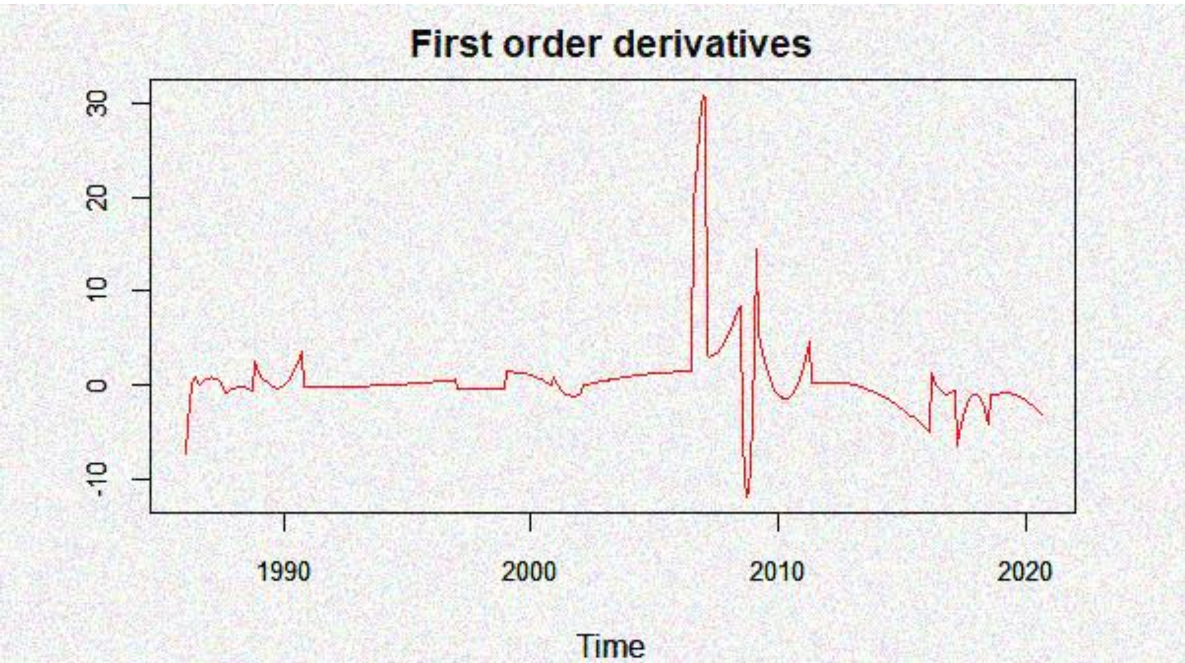


Figure: First order differentiation

- Second order derivatives

Function number	Second order derivative
1	-0.1233 * (3 * (2 * x)) +2.1050 * 2
2	0.29305 * 2 -0.00701 * (3 * (2 * x))
3	-0.003977 * (3 * (2 * x)) + 0.330731 * 2
4	-1.157e+00 * 2 + 8.467e-03 * (3 * (2 * x))
5	-1.022e-02 * 2 + 5.363e-05 * (3 * (2 * x))
6	2.070e-04 * (3 * (2 * x)) -8.751e-02*2
7	-1.112e-03 * (3 * (2 * x)) + 5.273e-01*2
8	8.758e-03 * (3 * (2 * x)) -4.966e+00* 2
9	-1.625e-04 * (3 * (2 * x)) + 1.219e-01 *2
10	1.585e+02 * 2 - 2.094e-01 * (3 * (2 * x))
11	-5.874e+00* 2 + 7.683e-03* (3 * (2 * x))
12	-3.297e+02* 2 + 4.021e-01 * (3 * (2 * x))
13	-1.214e+01* 2 + 1.386e-02* (3 * (2 * x))
14	7.235e-01* 2 + -7.690e-04* (3 * (2 * x))
15	-1.356e+01* 2 + 1.219e-02* (3 * (2 * x))
16	2.581e+01* 2 -2.240e-02 * (3 * (2 * x))
17	2.698e+00* 2 -2.259e-03 * (3 * (2 * x))

Table: Second order differentiation of the 17 equations

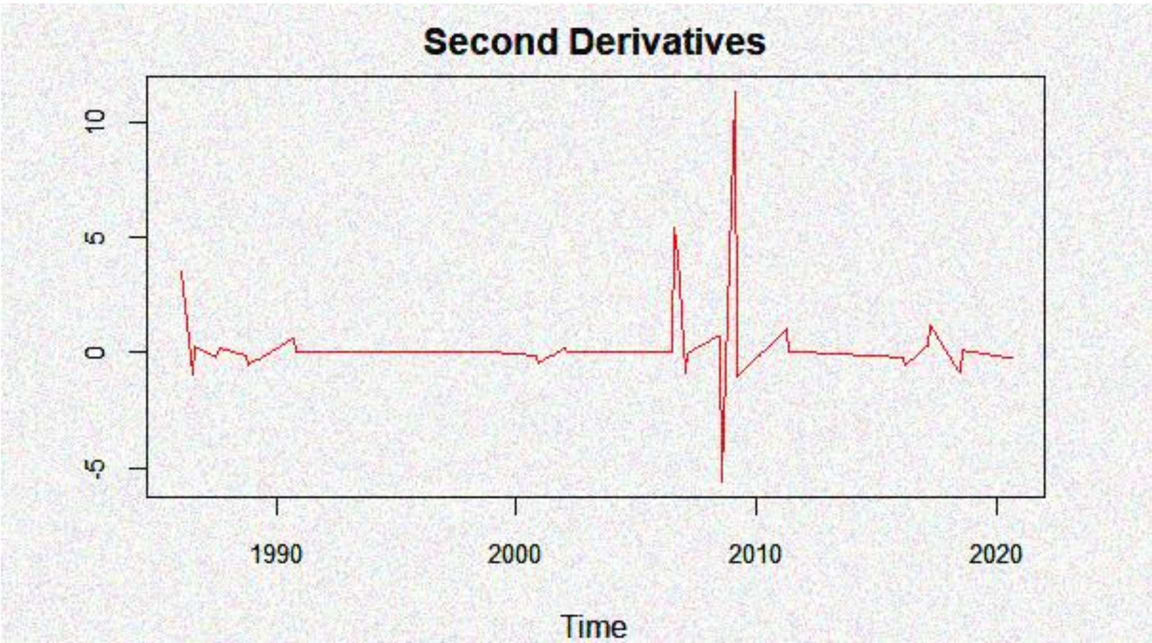


Figure: Second order differentiation

Chapter 5

Conclusion

Residuals of the spline regression model have fitted well and so the future predictions of crude oil price will be easier using this model. Prediction of exact price won't be possible but we will be able to tell how the actual value shall differ from predicted value using the above equations. There have been multiple factors like geopolitics, emergencies, Oil future contracts that are the reason of the fluctuation of the crude oil prices and looking at the graphs, we can conclude that these uncertain events shall continue in the future too.

References

https://synergenog.com/2020/05/03/top-6-events-that-impacted-global-oil-prices/

https://www.investopedia.com/ask/answers/030315/why-did-oil-prices-drop-so-much-2014.asp

https://www.oilandgasiq.com/strategy-management-and-information/articles/history-oil-prices

https://www.investopedia.com/ask/answers/052715/how-did-financial-crisis-affect-oil-and-gas-sector.asp

https://www.investopedia.com/articles/investing/072515/top-factors-reports-affect-price-oil.asp

https://datahub.io/core/oil-prices

https://www.wtrg.com/prices.htm