# Comparative Analysis of ARIMA and Hidden Markov Model Forecasting using Brent Crude Oil Prices

Sean Patrick

4/26/2021

**Introduction**

Brent crude is the most traded of all of the oil benchmarks. It is defined as crude drilled from the North Sea oilfields, the BFOE; Brent, Forties, Oseberg and Ekofisk. Brent crude oil type is widely used because it is both sweet and light, therefore it is easy to refine into gasoline anddiesel fuel. The importance of accurately forecasting oil prices cannot be understated, even though there is starting to be a massive push towards renewables, oil is still a critical part country's energy resources. Due the importance and the necessasity of oil to a country, unexpected, large and constant variations in oil prices can have negative effects on countries if those countries cannot plan properly and minimize risk caused by these variations. It is a helpful forecasting variable in assessing risk on a macroeconomic scale and making macroeconomic projections. If the model is more accurate it makes it easier to have an optimal response to macroeconomic situations and tailoring macroeconomic policies to address these situations. [1]

This paper explores the forecasting capabilities of the ARIMA and Hidden Markov Models using Brent Crude Oil Price data.The goal is to find the model that will have the highest forecasting accuracy. I will make a prediction of the price of Brent Crude Oil using these

[1] https://markets.businessinsider.com/commodities/oil-price

models before the date of the presentation, and then compare the differences on the specific date prices.

## Thesis

Hidden Markov model assumes that future states depend only on the current state, not on the events that occurred before it. ARIMA uses a number of lagged observations of time series to forecast observations. A weight is applied to each of the past term and the weights can vary based on how recent they are. This paper hypothesizes the Hidden Markov model will be able to produce a more accurate forecast based on the current patterns in crude oil prices/demand.

## Literature Review

Forecasting oil prices is a vast field of study, there are constant updates to models and new models being deployed. Pushing the envelope to find and build the most accurate model. There are differing opinions on the approach and results of these models.

Nademi, Arash, and Nademi, Younes conducted a study, using ARIMA, GARCH; Generalized Auto-Regressive Conditional Heteroskedasticity compared to semiparametric Markov switching AR- ARCH models. Using Brent crude, OPEC, and WTI oil prices, they forecasted out 1, 5, 10, and 22 steps ahead. At each step, the forecasts demonstrated that the Markov switching model was more accurate than the ARIMA and GARCH models. [2]

The HMM and ARIMA models I deployed forecasted that the Brent crude oil prices would be down over the 10-month period I tested. US Energy Information Administration

[2] https://www.sciencedirect.com/science/article/pii/S014098831830238X

forecasted that the Brent crude will average $65 a barrel between April 2021 and June 2021. Oil has been on a steady increase since May 2020, March 2021 being the highest since 2019, but there is a slowing in global demand, causing a decline in prices, due to Covid-19 rising again in some parts of the world, Europe specifically.[3]

OPEC wrote in their August 2020 monthly oil report,"[S]tructural changes to the global economy are forecasted to persist,"  and prices “are not expected to achieve pre-COVID-19 levels of activity before the end of 2021."

For the rest of 2020 OPEC predicted total oil demand would fall almost 10% , this a destabilizing number from the industry’s perspective the industry's perspective.  Which were in line with the forecasts that my models produced.[4]

The Hidden Markov and ARIMA models are capable of forecasting oil prices, but with a lower accuracy than some other current models.  Most of the literature on the subject concludes that, the models Hidden Markov and ARIMA are not the strongest models for this work.   The more complicated the model the more accurate the forecasts can be, especially when dealing with oil prices.

**R Packages**

Tidyverse, dplyr, ggplot2, forecast, colortools, Ecdat, fpp2, robustHD, seastests, tseries, x13binary, x12, lmtest, caret, gains, pdfetch, quantmod, depmixS4, XLConnect

---

[3] https://www.eia.gov/outlooks/steo/marketreview/crude.php

[4] https://www.opec.org/opec_web/static_files_project/media/downloads/publications/OPEC_MOMR_August-2020.pdf

**Data and Data processing**

Before modeling of the data can proceed, the data needs to be processed. Tesing for Seasonality, Time Series, Trend and if the data is Stationary. Unless this is done the ARIMA model will not be able to properly process and model the data. And the forecasts they produce will contain these things making the model function inaccuarately. This dataset is the average price of one barrel of Brent Crude oil by month May 1987 to Mar 2021,downloaded from https://www.eia.gov/dnav/pet/hist/rbrtem.htm. [5]

```
oil= read.csv("C:/Users/Sean Patrick/Desktop/Classes/Capstone/Final Proj Info
/RBRTEm.csv")

head(oil)

##       Date Europe.Brent.Spot.Price.FOB..Dollars.per.Barrel.
## 1 May-1987                                            18.58
## 2 Jun-1987                                            18.86
## 3 Jul-1987                                            19.86
## 4 Aug-1987                                            18.98
## 5 Sep-1987                                            18.31
## 6 Oct-1987                                            18.76

tail(oil)

##         Date Europe.Brent.Spot.Price.FOB..Dollars.per.Barrel.
## 403 Nov-2020                                            42.69
## 404 Dec-2020                                            49.99
## 405 Jan-2021                                            54.77
## 406 Feb-2021                                            62.28
## 407 Mar-2021                                            65.41
## 408                                                        NA

colnames(oil)

## [1] "Date"
## [2] "Europe.Brent.Spot.Price.FOB..Dollars.per.Barrel."

str(oil)
```

---

5 https://www.eia.gov/dnav/pet/hist/rbrtem.htm

```
## 'data.frame':    408 obs. of  2 variables:
##  $ Date                                             : Factor w/ 408 levels
"","Apr-1988",..: 273 205 171 35 375 341 307 69 137 103 ...
##  $ Europe.Brent.Spot.Price.FOB..Dollars.per.Barrel.: num  18.6 18.9 19.9 1
9 18.3 ...
```

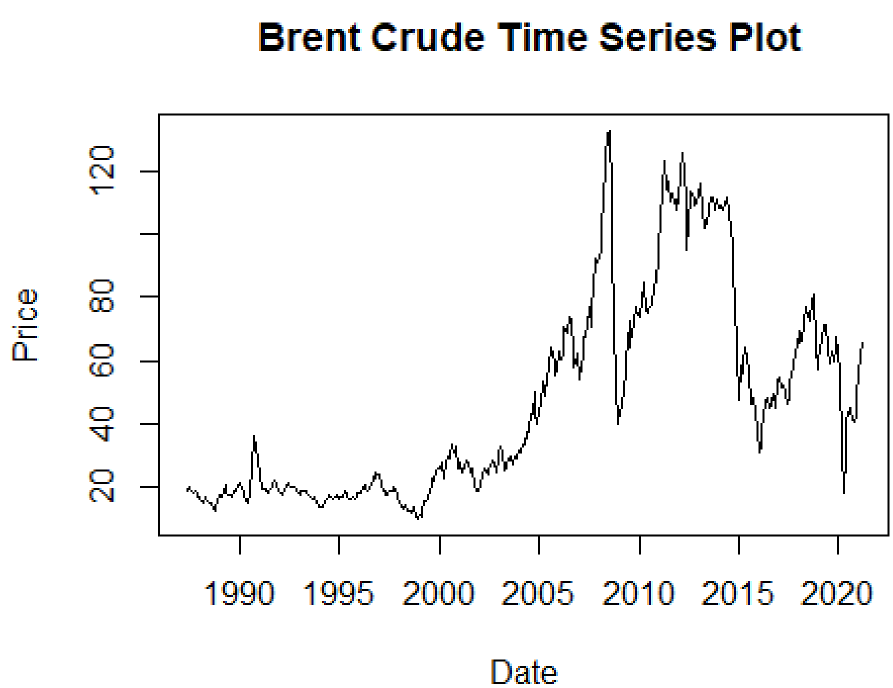
To make this dataset easier to work with, the column

"Europe.Brent.Spot.Price.FOB..Dollars.per.Barrel." was renamed to "Price".

```
oil<-rename(oil,Price=Europe.Brent.Spot.Price.FOB..Dollars.per.Barrel.)
```

Next is testing if the data is a time series and converting to a times series for ARIMA.

```
is.ts(oil)

## [1] FALSE

tsoil = ts(oil$Price, start= c(1987,5), end = c(2021,3), frequency = 12)
plot(tsoil, ylab = "Price", xlab= "Date",main = " Brent Crude Time Series Plo
t")
```

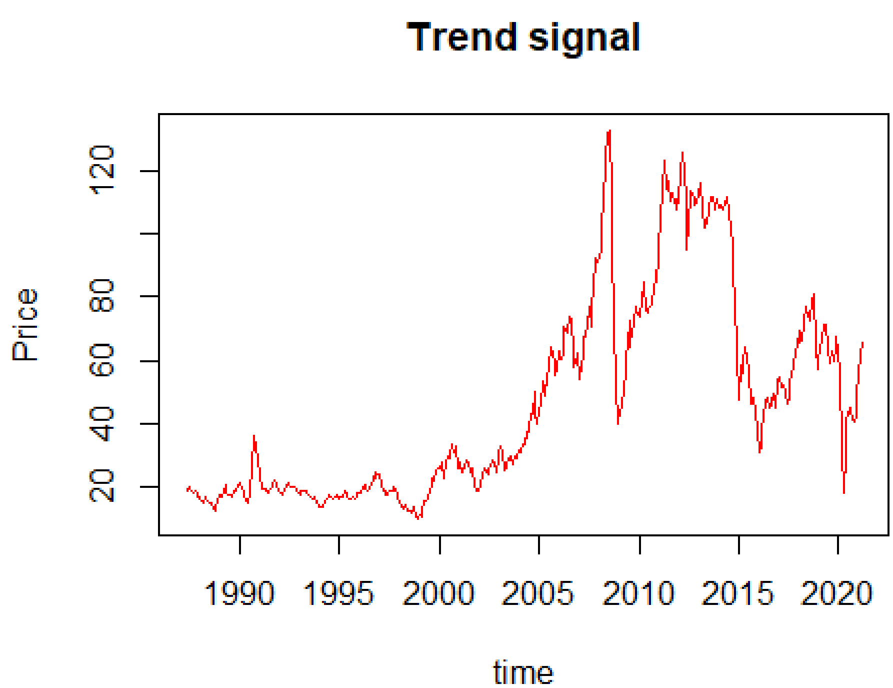



The original data form US Energy Information Administration was not a time series, so it had to be coerced to a time series. Function ts() was used using May 1987 to Mar 2021 with a frequency of 12 because the data is a monthly dataset. The plot illustrates the original dataset converted to a time series.

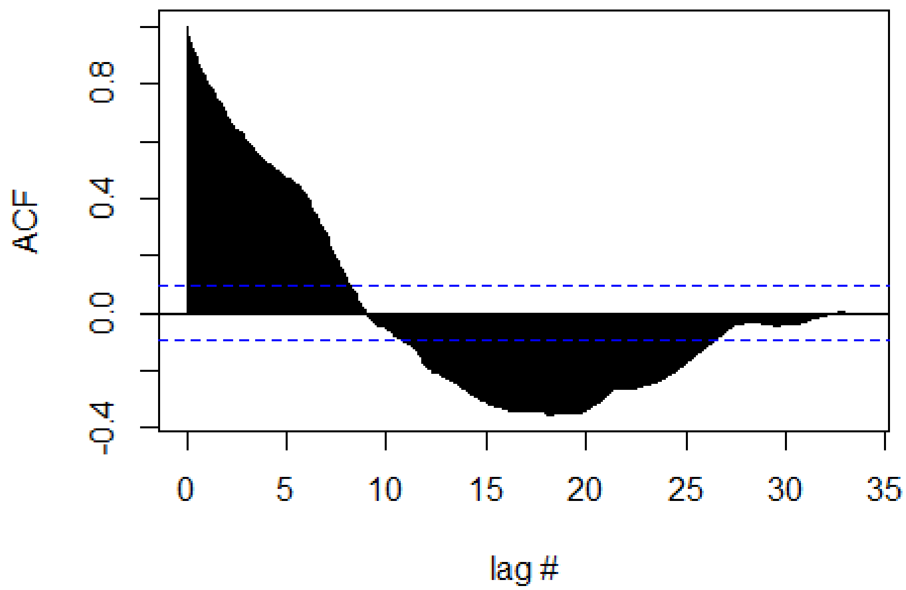
Determining if there is a seasonal or trend in the time series data.

Stationary testing, a stationary time series is one whose properties do not depend on the time at which the series is observed.(3) Autocorrelation Function (ACF)and Partial Autocorrelation Function (PACF) , autocorrelation is eh similarity between a time series and a lagged version of of the given time series over consecutive time intervals. It measures the correlation between the variable's past and current values.(4)

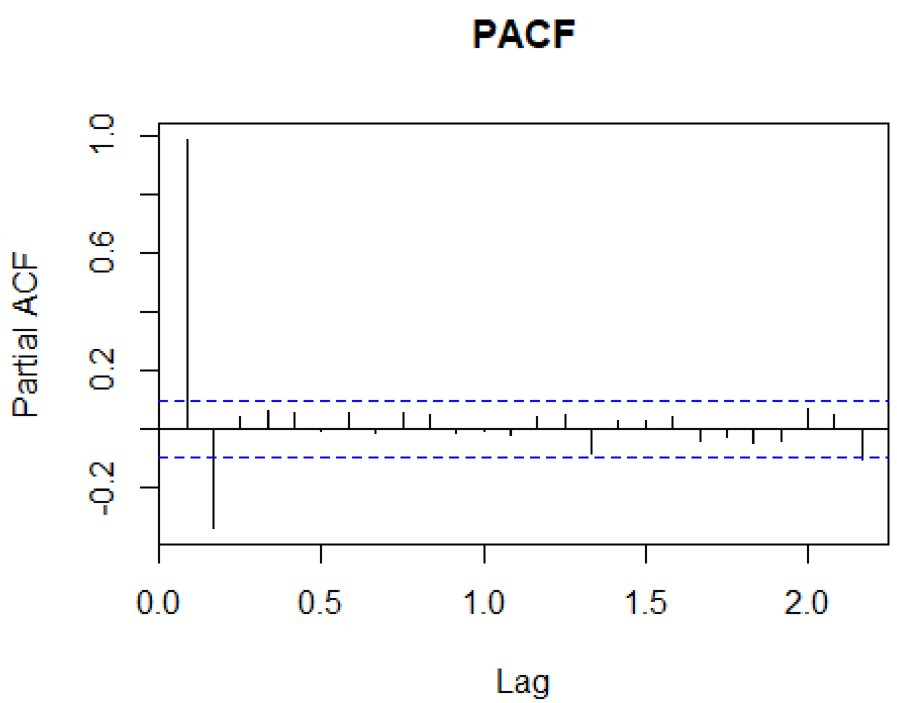
```
plot(tsoil,
     type='l',col='red',
     xlab = "time",
     ylab = "Price",
     main = "Trend signal")
```



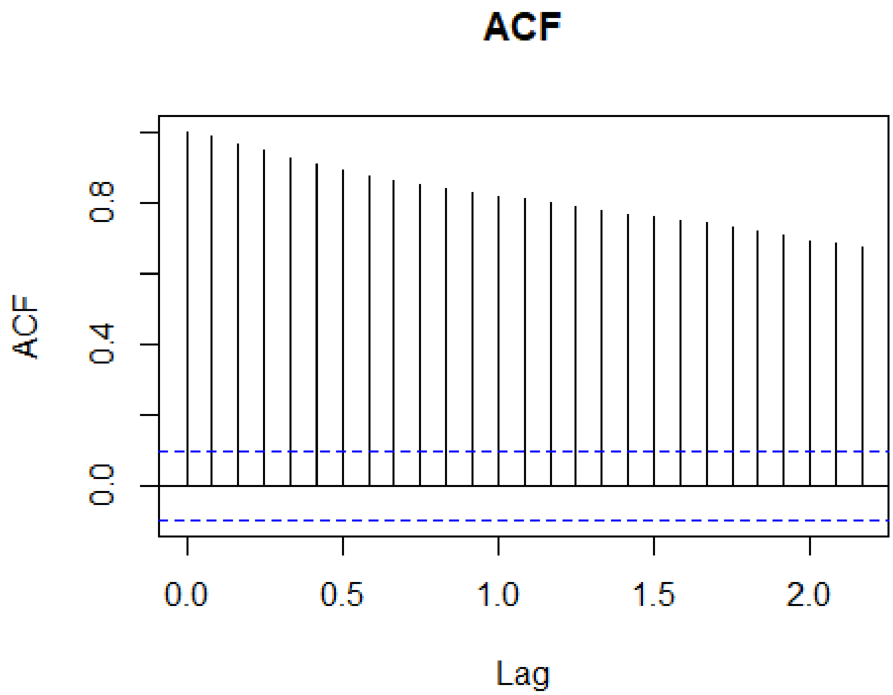

```
acf(tsoil,lag.max = length(tsoil),
    xlab = "lag #", ylab = 'ACF', main=' ')
```




```
pacf(tsoil,  main="PACF")
```




```
acf(tsoil,  main="ACF")
```




Almost all lags exceeding the confidence interval of the ACF, the time series is not stationary. There is a high 1st lag in PACF further evidence of the series not being stationary.

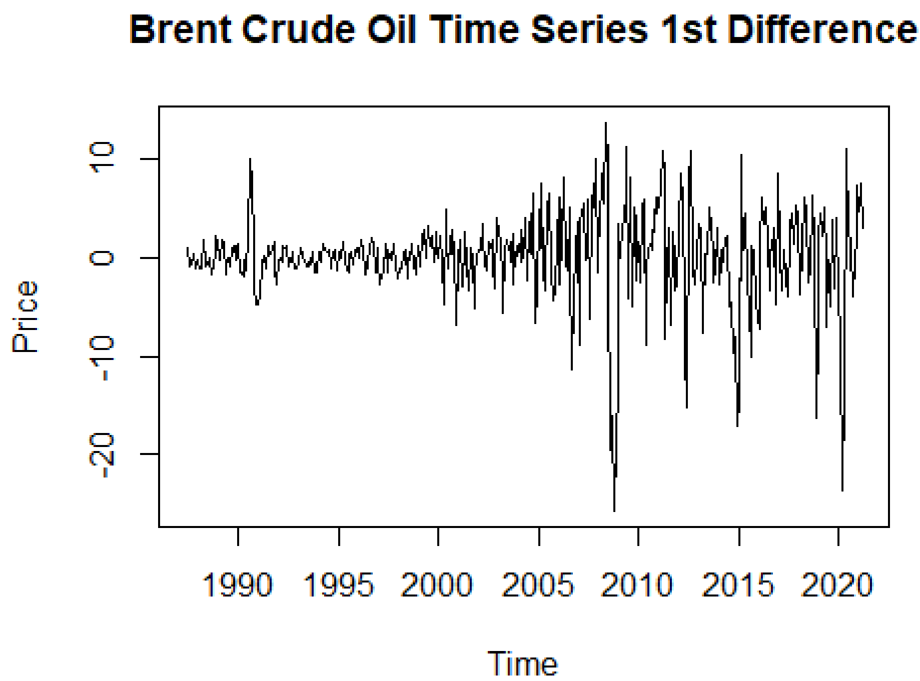
```
adf.test(tsoil)

##
##  Augmented Dickey-Fuller Test
##
## data:  tsoil
## Dickey-Fuller = -2.3908, Lag order = 7, p-value = 0.4123
## alternative hypothesis: stationary
```

Augmented Dickey-Fuller (ADF) t-statistic test for unit root The time series p-value = 0.4123, assuming significance p=0.01, we reject the null hypothesis, and classify this as stationary. The data is confirmed not stationary. Making the Data Stationary, I used the Box-Cox Transformation

```
bcl = BoxCox.lambda(tsoil,method = c("guerrero"), lower = -1, upper = 2)
bcl

## [1] -0.3747872

tsoiltr = (tsoil^bcl)
adf.test(tsoiltr)

##
##  Augmented Dickey-Fuller Test
##
## data:  tsoiltr
## Dickey-Fuller = -2.3018, Lag order = 7, p-value = 0.4499
## alternative hypothesis: stationary
```

Using the Box-Cox transformation failed to make the data stationary,p-value = 0.4499, this method did not reduce p-value So I had to resort to the old fashioned differencing method. It can be made easier by using the function ndiffs(), to determine the number of differences necessary to make the time series stationary.

```
ndiffs(tsoil)

## [1] 1

diff.tsoil<-diff(tsoil)
plot(diff.tsoil, ylab='Price',xlab='Time', main = "Brent Crude Oil Time Serie
s 1st Difference")
```

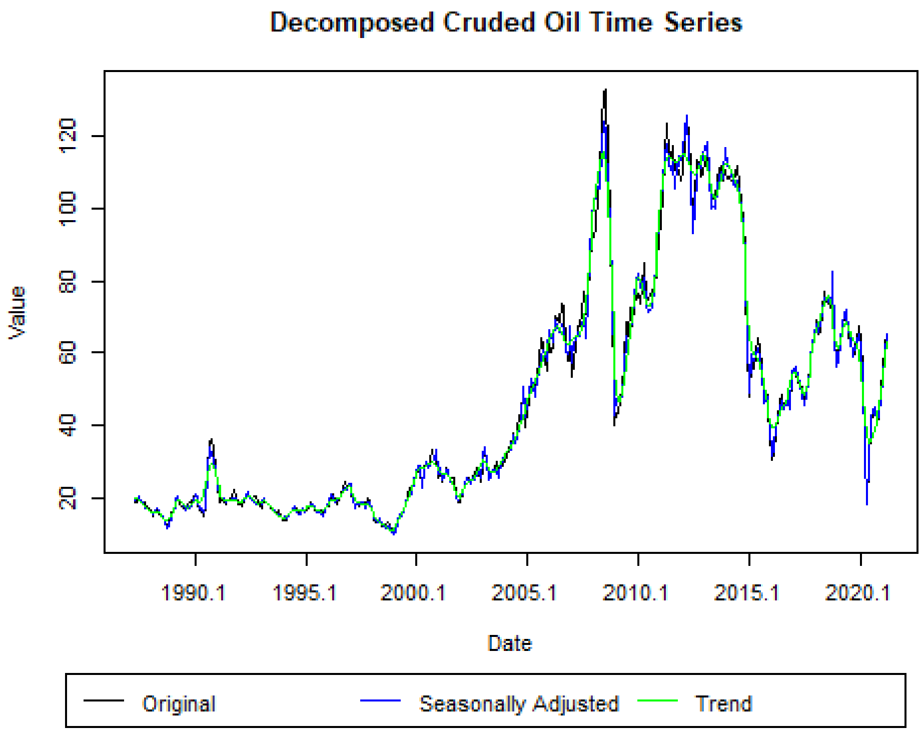



Based off the plot it looks as though the trend has been removed using the 1st difference. It is prudent to test if the 1st difference was effective to minimize inaccuracies in the modeling.

```
adf.test(diff.tsoil)

## Warning in adf.test(diff.tsoil): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  diff.tsoil
## Dickey-Fuller = -7.8344, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

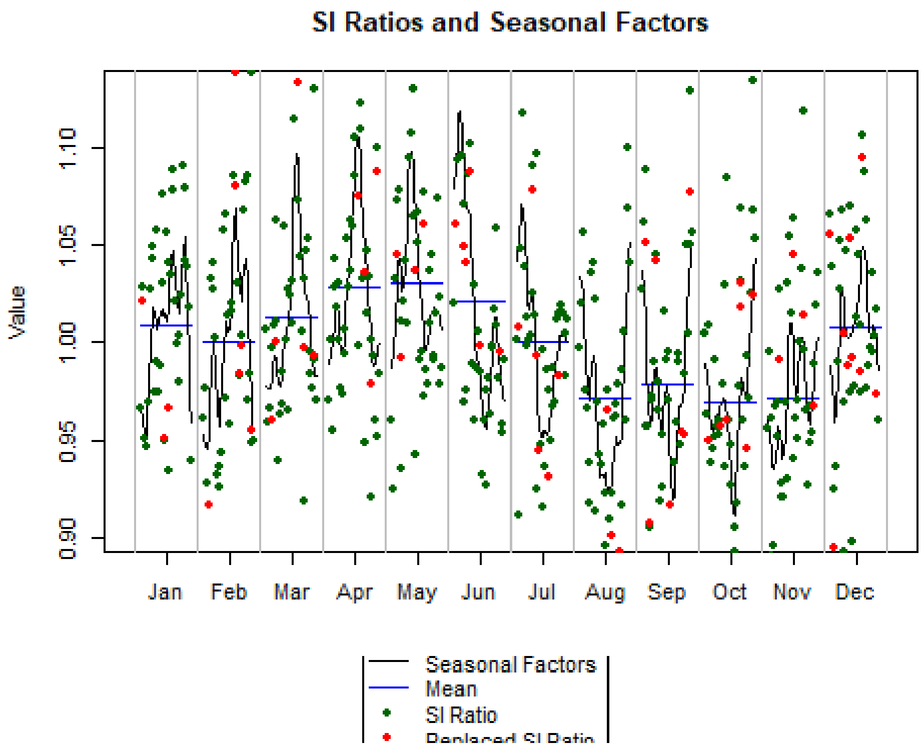
The 1st difference was enough to get the p-value = 0.01 and now the series is stationary and we can move onto testing for seasonality and decomposing the data. The data is monthly so I can use the function x12 to check for seasonality and decompose the data.

```r
dec.tsoil = x12(tsoil)
plot(dec.tsoil , sa=TRUE , trend=TRUE, main = "Decomposed Cruded Oil Time Ser
ies ")
```



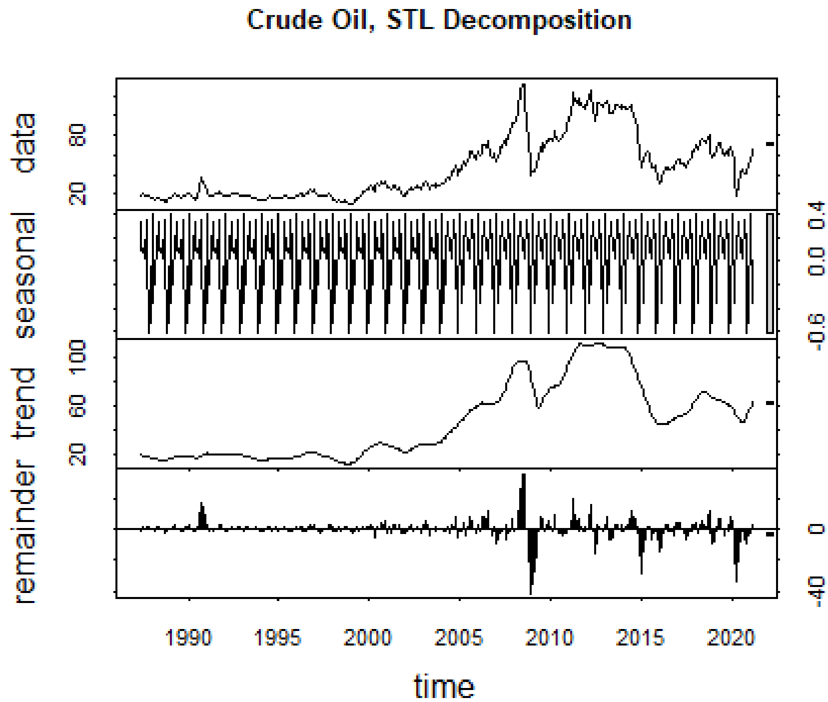

The seasonally adjusted is closely tied to the original series, with no evidence of a significant seasonal effect. But there are slight differences so there is some, because of the deviations with the trend line from the original. I wanted to look at the seasonal factors for the series.

```r
plotSeasFac(dec.tsoil, main = "SI Ratios and Seasonal Factors")
```

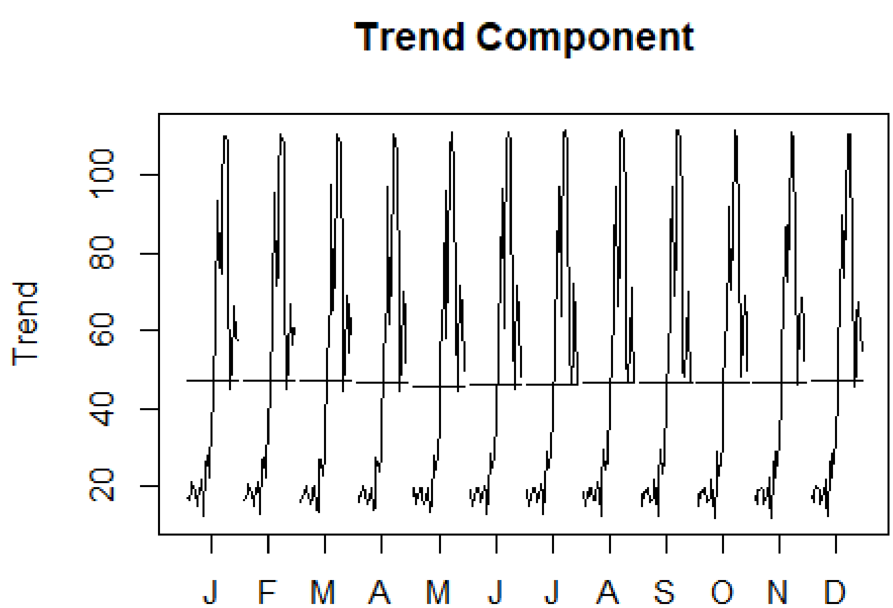



June appears that the expected pattern deviates from the mean. Every month has influential observations, based on the SI ratios. To decompose the data using the function stl(), “Seasonal and Trend decomposition using Loess,” Loess is one way of estimating nonlinear relationships. (3)

```
decomp.oil <- stl(tsoil, t.window=15, s.window="periodic", robust=TRUE)
plot(decomp.oil, main = "Crude Oil, STL Decomposition ")
```




```
monthplot(decomp.oil, choice = "trend", main="Trend Component", ylab="Trend")
```



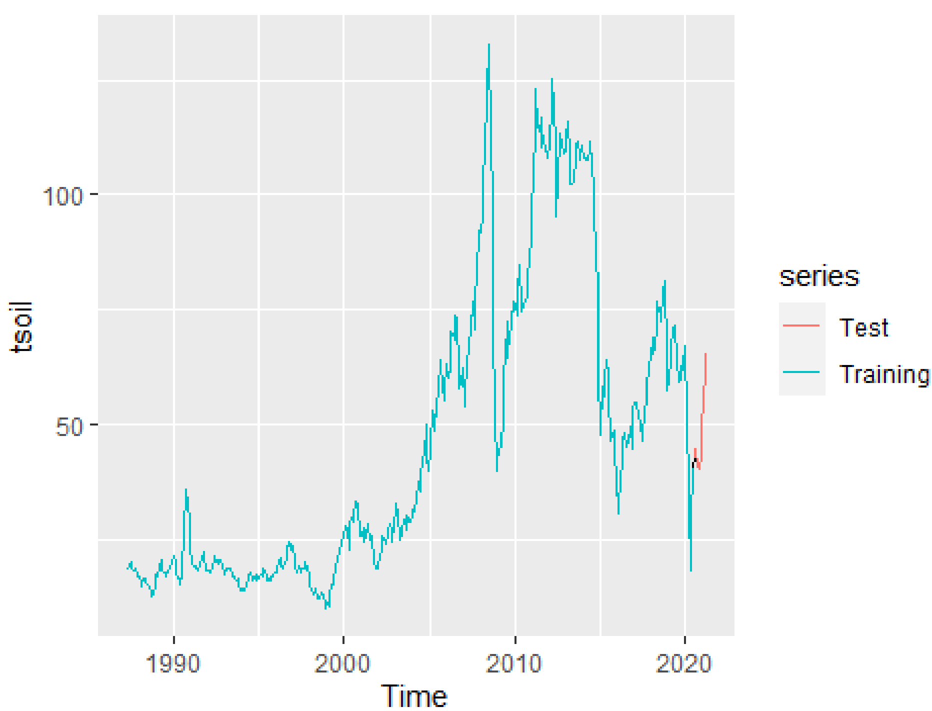
Observing the trend plot and the data plot, that the trend follows closely to the pattern of the original. The remainder shows the spikes and falls, showing the interventions in the series. The mean between the months shows that it stays the same for all 12 months with the variation, between 10 to 120. Demonstrating the trend effect on the original.

**Modelling**

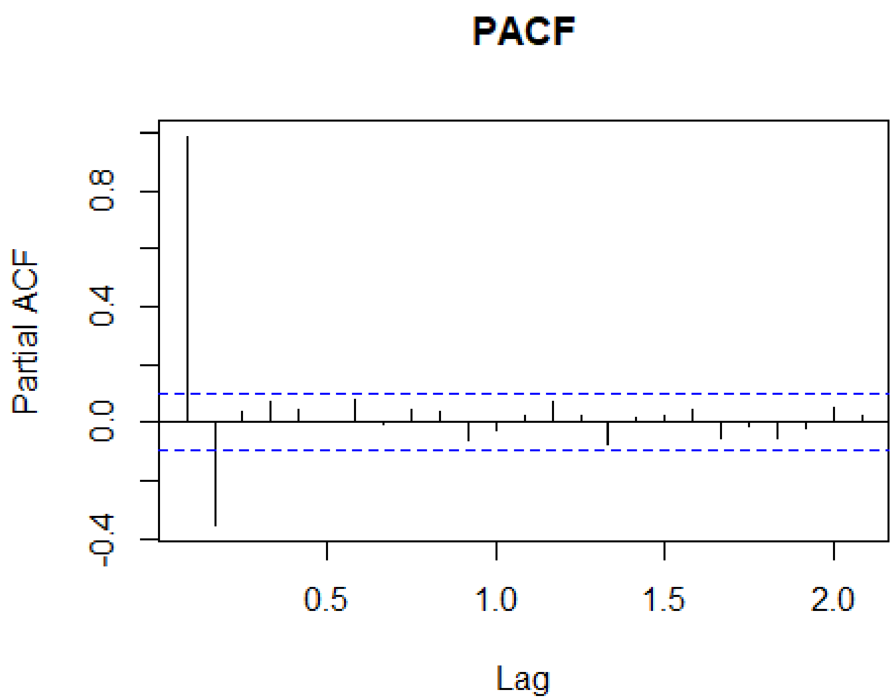
First method is ARIMA;Auto Regressive Integrated Moving Average Arima implementation

ARIMA(p,d,q) mode: p=order of the autoregressive part; d=degree of first differencing involved; q=order of the moving average part.

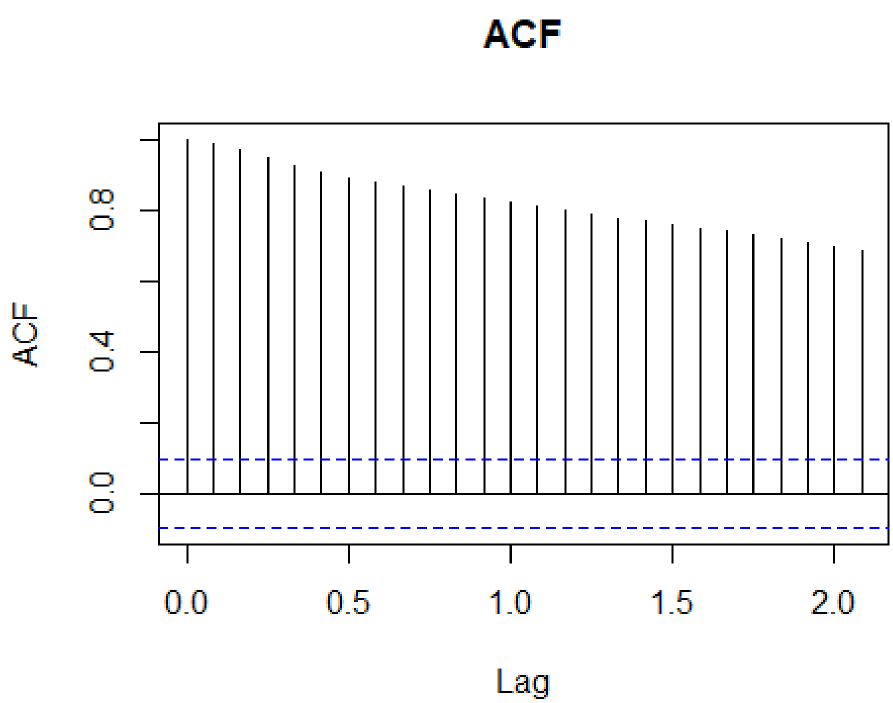
```
oil.train <- window(tsoil, end=c(2020,6))
oil.test <- window(tsoil, start=c(2020,7),end=c(2021,3))

autoplot(tsoil) +
    autolayer(oil.train, series="Training") +
    autolayer(oil.test, series="Test")
```




```
pacf(oil.train,  main="PACF")
```




```
acf(oil.train,  main="ACF")
```



After looking at the ACF and PACF, I determined that my manual model would be Manual Arima, p = 2 auto regressive terms, d = 1 differences, q = 2 is the number of lagged forecast errors in the prediction equation. Further I will test to make sure the models are the best I can make them by using Auto.Arima and Arima(0,2,0).

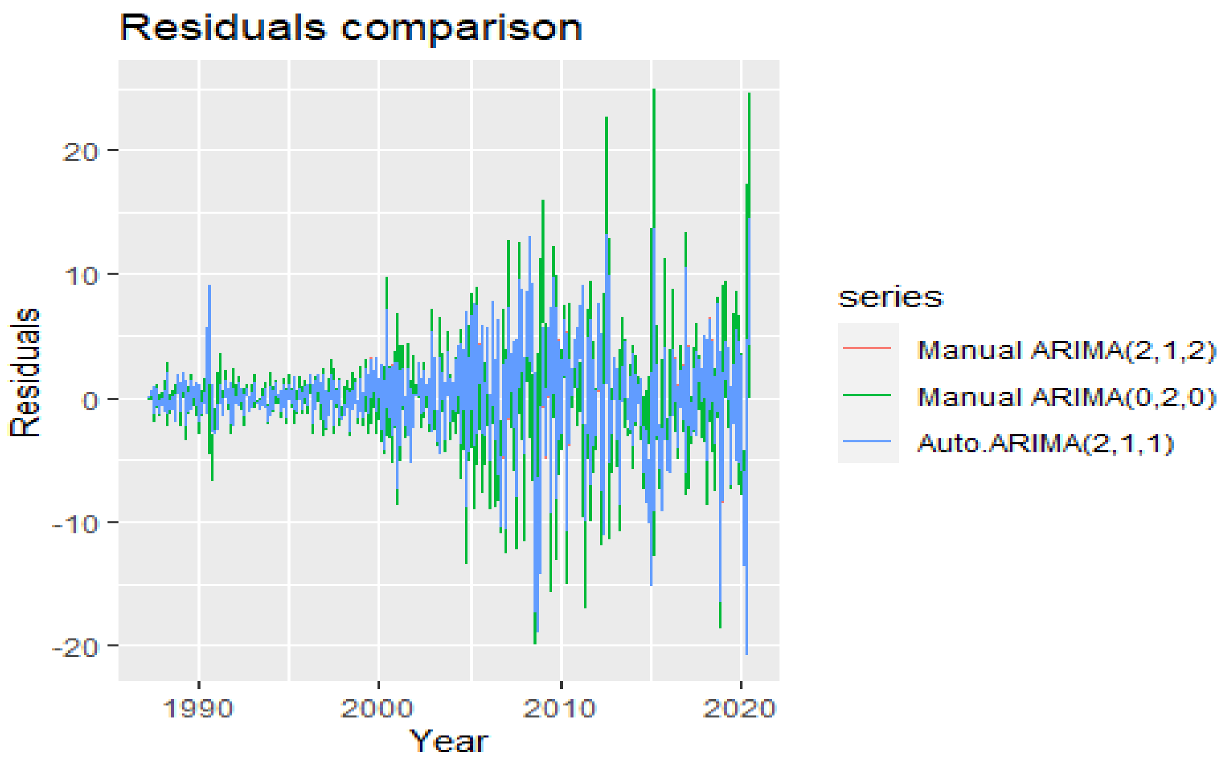
```
fit1 <- Arima(oil.train, order=c(2,1,2))
fit2 <- Arima(oil.train,order = c(0,2,0))
fit3 <- auto.arima(oil.train)
```

Auto.ARIMA(2,1,1)p = 2 auto regressive terms, d = 1 differences, q = 1 is the number of lagged forecast errors in the prediction equation.

Checking Residuals

Check the residuals helps to determine the model that best fits the data.

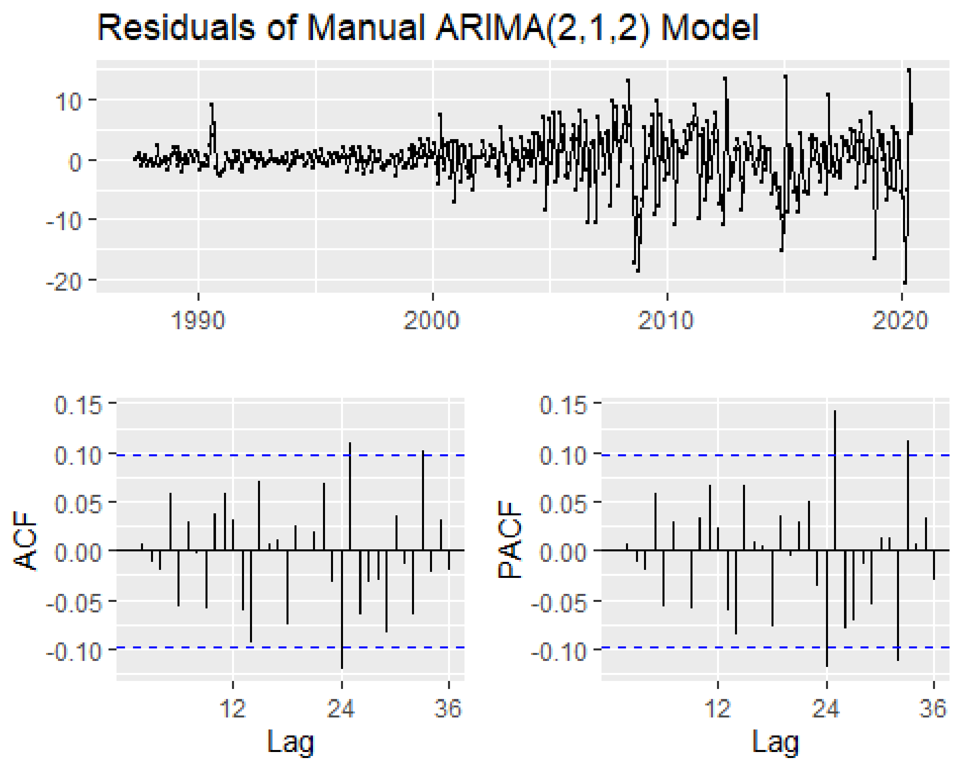
```
f1resid = resid(fit1)
f2resid = resid(fit2)
f3resid = resid(fit3)

ARresid <- cbind(f1resid,f2resid,f3resid)
colnames(ARresid) <- c("Manual ARIMA(2,1,2)","Manual ARIMA(0,2,0)",
                    "Auto.ARIMA(2,1,1)")
autoplot(ARresid,xlab="Year",ylab="Residuals") +
    ggtitle("Residuals comparison")
```



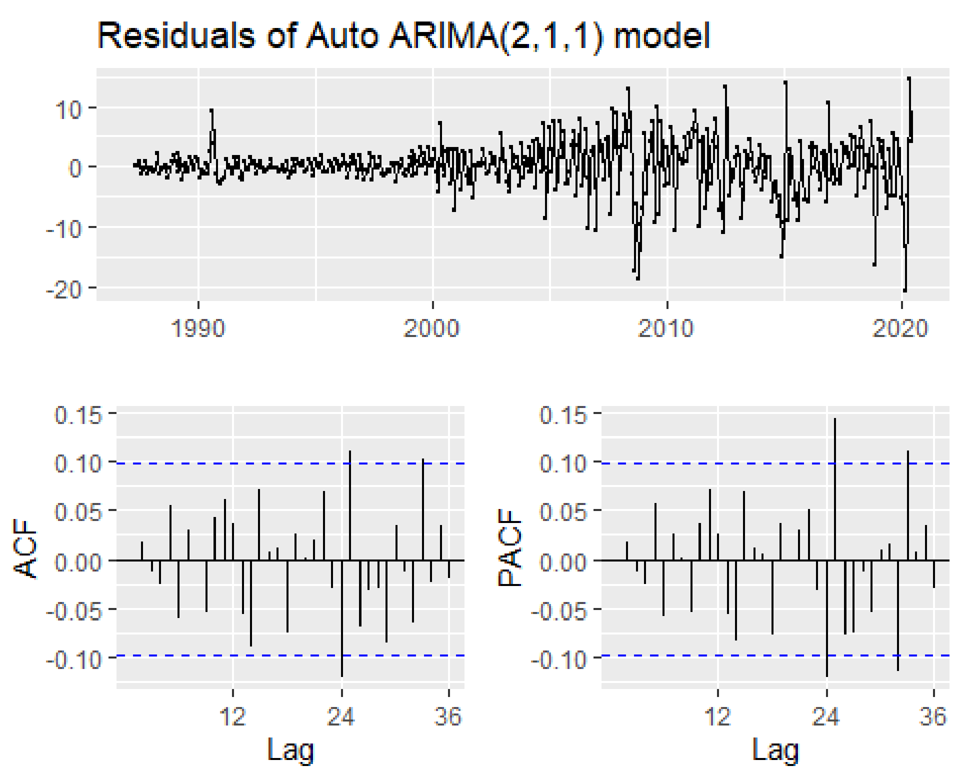

Comparing the residuals plot, ARIMA(0,2,0) was the worst fit in terms of residuals,

ARIMA(2,1,2) and Auto.Arima(2,1,1) perform similarly. Evaluating residuals

```
ggtsdisplay(f1resid,main="Residuals of Manual ARIMA(2,1,2) Model")
```




```
ggtsdisplay(f3resid,main="Residuals of Auto ARIMA(2,1,1) model")
```




More testing is required to determine which model to use. Normality tests

```
shapiro.test(f1resid)

##
##  Shapiro-Wilk normality test
##
## data:  f1resid
## W = 0.93736, p-value = 6.739e-12

shapiro.test(f3resid)

##
##  Shapiro-Wilk normality test
##
## data:  f3resid
## W = 0.93721, p-value = 6.486e-12
```

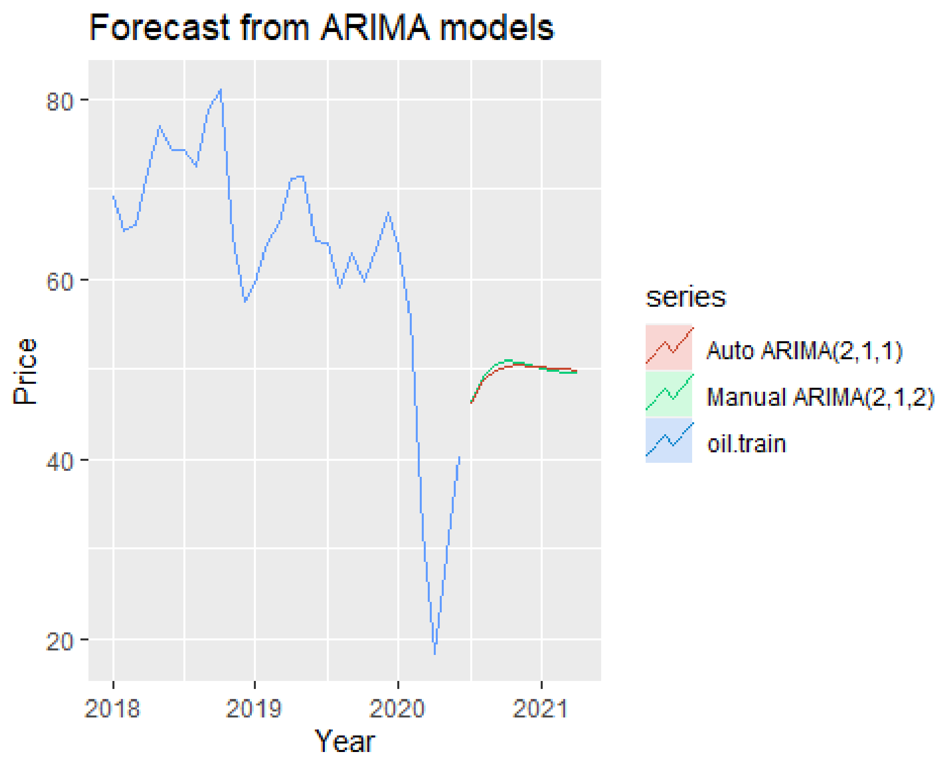
For Manual ARIMA, p-value = 6.739e-12 and Auto.ARIMA,p-value = 6.486e-12. The residuals for both models are significantly heteroscedastic, this means the depedent variable changes significantly from the beginning to the end of the series. The dataset is highly variable between its highest point and lowest. Which has the potential to effect the accuracy of the models overall both for ARIMA and Hidden Markov Model. Due to the models p-values being quite close I decided to move ahead and use both models to forecast. Then determine after which model to move forward with when comparing to the Hidden Markov model.

### Forecasting Manual ARIMA(2,1,2) Auto.ARIMA(2,1,1)

```
fc.fit1  <- forecast(oil.train,model=fit1, h=10)
fc.fit3 <- forecast(oil.train,model=fit3, h=10)
```

Plotting

```
oil.train = window(oil.train, start = c(2018,1), end = c(2020,6))

autoplot(oil.train,xlab ="Year",ylab="Price") +
    autolayer(oil.train) +
    autolayer(fc.fit1,series="Manual ARIMA(2,1,2)",PI =FALSE) +
    autolayer(fc.fit3,series="Auto ARIMA(2,1,1)",PI =FALSE) +
    ggtitle("Forecast from ARIMA models")
```




These two models are very similar in their forecasts.

```
summary(fc.fit1)

##
## Forecast method: ARIMA(2,1,2)
##
## Model Information:
## Series: object
## ARIMA(2,1,2)
##
## Coefficients:
##          ar1      ar2      ma1     ma2
##        1.262  -0.4254  -0.8788  0.0419
## s.e.   0.000   0.0000   0.0000  0.0000
##
## sigma^2 estimated as 19.35:  log likelihood=-1149.5
## AIC=2300.99   AICc=2301   BIC=2304.97
##
## Error measures:
##                      ME     RMSE     MAE        MPE     MAPE      MASE
## Training set 0.07746217 4.370795 2.99556 -0.1351317 6.959521 0.2473435
##                      ACF1
## Training set 0.0002611054
##
## Forecasts:
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
```

```
## Jul 2020       46.19746 40.56054 51.83438 37.57654 54.81838
## Aug 2020       49.22374 39.60270 58.84477 34.50963 63.93784
## Sep 2020       50.52114 37.77497 63.26731 31.02756 70.01472
## Oct 2020       50.87097 35.74012 66.00182 27.73033 74.01161
## Nov 2020       50.76049 33.80481 67.71617 24.82902 76.69196
## Dec 2020       50.47223 32.08722 68.85724 22.35478 78.58968
## Jan 2021       50.15545 30.61041 69.70050 20.26388 80.04702
## Feb 2021       49.87831 29.35318 70.40345 18.48783 81.26880
## Mar 2021       49.66333 28.27754 71.04912 16.95659 82.37008
## Apr 2021       49.50993 27.34303 71.67683 15.60858 83.41128
```

```
summary(fc.fit3)
```

```
##
## Forecast method: ARIMA(2,1,1)
##
## Model Information:
## Series: object
## ARIMA(2,1,1)
##
## Coefficients:
##          ar1      ar2      ma1
##        1.249  -0.3926  -0.8621
## s.e.  0.000   0.0000   0.0000
##
## sigma^2 estimated as 19.3:  log likelihood=-1149.54
## AIC=2301.08   AICc=2301.09   BIC=2305.07
##
## Error measures:
##                      ME    RMSE      MAE        MPE     MAPE      MASE
## Training set 0.08166595 4.37129 2.994678 -0.1345243 6.956477 0.2472706
##                    ACF1
## Training set -0.00399342
##
## Forecasts:
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jul 2020       45.90234 40.27194 51.53274 37.29139 54.51329
## Aug 2020       48.66177 39.03471 58.28883 33.93845 63.38509
## Sep 2020       49.89707 37.17338 62.62076 30.43787 69.35627
## Oct 2020       50.35662 35.27255 65.44068 27.28753 73.42571
## Nov 2020       50.44562 33.54190 67.34933 24.59362 76.29762
## Dec 2020       50.37636 32.03245 68.72026 22.32177 78.43094
## Jan 2021       50.25491 30.73252 69.77729 20.39800 80.11182
## Feb 2021       50.13041 29.60965 70.65117 18.74661 81.51420
## Mar 2021       50.02258 28.62826 71.41691 17.30279 82.74238
## Apr 2021       49.93679 27.75656 72.11702 16.01506 83.85853
```

There is the point forecast, which is what the model predicts the price will be at that time.

Then there are the ranges Lo and High 80 means there is an 80% probability that the price will fall in between those two ranges that were forecasted. Lo and High 95 means there is a 95% probability that price will fall in between those two ranges that were forecasted. Based off the plotting forecast the manual and auto.arima is pretty close. Checking the RMSE Root Mean Squared Error- Manual = 4.370795, Auto =4.37129 Manual ARIMA edges the Manual out marginally.

Manual ARIMA Forecast

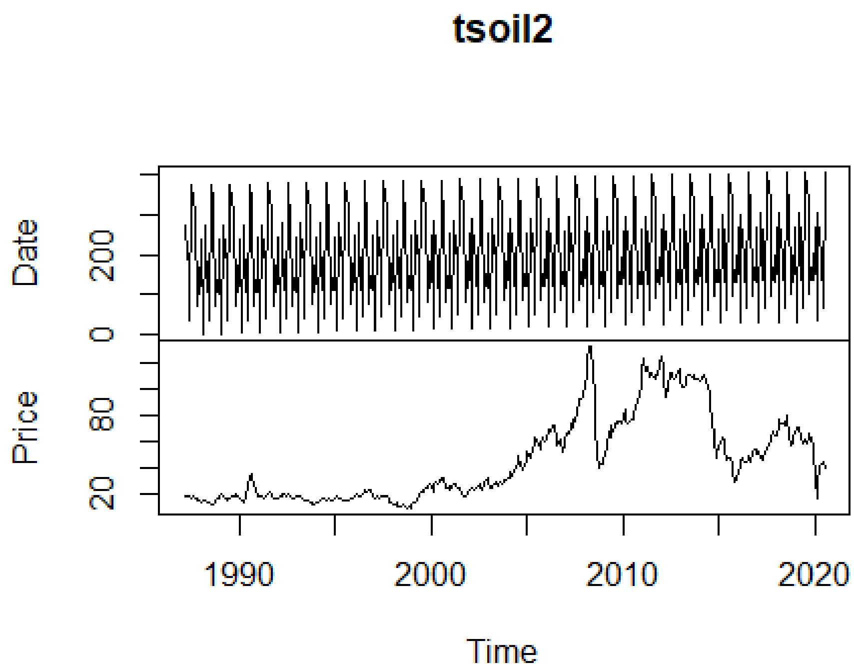
| Date | Price | Actual | Point Forecast | Predicted | Low 80 | High 80 | Low 95 | High 95 | Percentage Difference Actual vs Predicted |
|---|---|---|---|---|---|---|---|---|---|
| 20-Jul | 43.24 | Up | 46.20 | Up | 40.56 | 51.83 | 37.58 | 54.82 | 6.40% |
| 20-Aug | 44.74 | Up | 49.22 | Up | 39.60 | 58.84 | 34.51 | 63.94 | 9.11% |
| 20-Sep | 40.91 | Down | 50.52 | Up | 37.77 | 63.27 | 31.03 | 70.01 | 19.02% |
| 20-Oct | 40.19 | Down | 50.87 | Up | 35.74 | 66.00 | 27.73 | 74.01 | 21.00% |
| 20-Nov | 42.69 | Up | 50.76 | Down | 33.80 | 67.72 | 24.83 | 76.69 | 15.90% |
| 20-Dec | 49.99 | Up | 50.47 | Down | 32.09 | 68.86 | 22.35 | 78.59 | 0.96% |
| 21-Jan | 54.77 | Up | 50.16 | Down | 30.61 | 69.70 | 20.26 | 80.05 | -9.20% |
| 21-Feb | 62.28 | Up | 49.88 | Down | 29.35 | 70.40 | 18.49 | 81.27 | -24.86% |
| 21-Mar | 65.41 | Up | 49.66 | Down | 28.28 | 71.05 | 16.96 | 82.37 | -31.71% |

Using the “Up”, “Down” metric the Manual ARIMA model gave back only two out of nine forecasts correct.  The percentage range between the actual and point forecast -31.71% to 21.00%, is large range to work with when trying to obtain model accuracy.

**Hidden Markov Model**

Reloading the data so it is not processed in the way I did for the ARIMA modeling. HMM requires some steps that are the same such as making it a time series and decomposing the data. I have to create a “bucket” of labels to put the binary series, “Up”,“Down”, to make it the output of the HMM forecast.

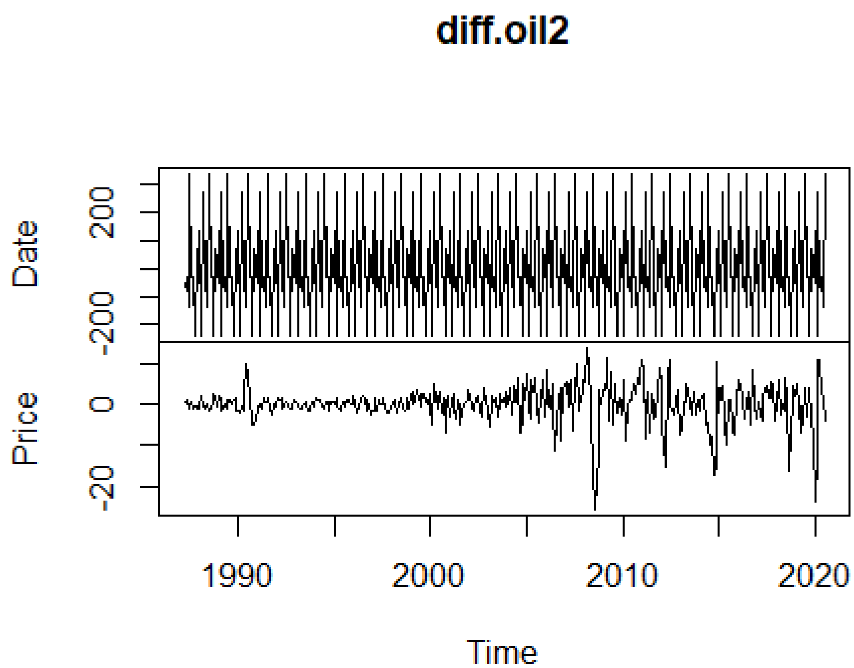
```r
oil2 = read.csv("C:/Users/Sean Patrick/Desktop/Classes/Capstone/Final Proj In
fo/RBRTEm.csv")

oil2<-rename(oil2,Price=Europe.Brent.Spot.Price.FOB..Dollars.per.Barrel.)

tsoil2 = ts(oil2,Price, start= c(1987,3), end = c(2020,7), frequency = 12)

plot(tsoil2)
```




Take first differences to try and get time series stationary.

```r
diff.oil2 <- diff(tsoil2)
head(diff.oil2); tail(diff.oil2)

##           Date Price
## Apr 1987   -68  0.28
## May 1987   -34  1.00
## Jun 1987  -136 -0.88
## Jul 1987   340 -0.67
## Aug 1987   -34  0.45
## Sep 1987   -34 -0.98

##           Date  Price
## Feb 2020  -237 -13.63
## Mar 2020   272  11.00
## Apr 2020   -68  10.89
## May 2020   -34   2.97
```

```
## Jun 2020 -136   1.50
## Jul 2020  340  -3.83

diff.oil2 = na.omit(diff.oil2)

plot(diff.oil2)
```




Now the data is stationary again.

Constructing “bucket”

oil.sign <- vector(length=length(diff.oil2)) for (i in 1:length(diff.oil2)) { oil.sign[i] = ifelse(diff.oil2[i] >=0, “Up”, “Down”) } table(oil.sign)

| | |
|---|---|
| oil.sign | |
| Down | Up |
| 453 | 347 |

Transition Matrix

MC.oil <- markovchainFit(data = oil.sign, method=“mle”, name = “nio mle”) MC.oil$estimate@transitionMatrix

| | | |
|---|---|---|
| Transition Matrix | | |
| | Down | Up |
| Down | 0.511062 | 0.488938 |
| Up | 0.636888 | 0.363112 |

A transition matrix has the information about the probability of transitioning between the different states in the model, in this model “Up”, “Down”.

MCoil.pred = predict(object = MC.oil$estimate, newdata = c(“Up”, “Down”), n.ahead=10)
MCoil.pred

| Hidden Markov Model Forecast | | | |
|---|---|---|---|
| Date | Price | Predicted | Actual |
| 20-Jul | 43.24 | Down | Up |
| 20-Aug | 44.74 | Down | Up |
| 20-Sep | 40.91 | Down | Down |
| 20-Oct | 40.19 | Down | Down |
| 20-Nov | 42.69 | Down | Up |
| 20-Dec | 49.99 | Down | Up |
| 21-Jan | 54.77 | Down | Up |
| 21-Feb | 62.28 | Down | Up |
| 21-Mar | 65.41 | Down | Up |

HMM forecast only gave back two out of nine accurate predictions for the average monthly price per barrel of Brent Crude. Yielding

**Conclusion**

Look at the forecasted outcome of the Manual ARIMA model the average price per barrel of the actuals fell within Low and High ranges for the model but the point forecast predictions were somewhat inaccurate. The accuracy range varied from 21% to -31.71% when comparing actual to predicted. Manual ARIMA did manage to get one prediction for December 2020 within 1% of the actual price, giving me my best result. After assessing the predictions of both models, I broke down the Manual ARIMA forecast into “Up”, “Down”, for a more straightforward comparison between the two. The Hidden Markov model forecasted that between July 2020 to March 2021 the average price per barrel would be “Down” for those nine months. HMM successfully forecasted that average Brent oil prices would be “Down” for two months September 2020 and October 2020. Manual ARIMA predicted two successful instances when using my “Up”, “Down” metric, July 2020, and August 2020. Using this metric as a point of comparison Hidden Markov model had a 22.22% forecasting accuracy and Manual ARIMA had 22.22% forecasting accuracy. I would characterize both my models as having a poor forecasting performance.

My interpretation is that the Hidden Markov Model I created was too simple and could not handle the low number of observations accurately. The Covid-19 Recession and the precipitous drop in demand may have had a significant impact on the prediction accuracy of the Hidden Markov model because there was such a large drop in price recently. Another note to add is that the residuals are significantly heteroscedastic, meaning the dependent variable changes significantly from the beginning to the end of the series. This could account for the Hidden Markov Model not being able to handle this dataset as well.

In conclusion, the HMM and Manual ARIMA (2,1,2) were not able to yield a more accurate forecast of the Price than the other, compared to the actuals I tested on. My hypothesis that the Hidden Markov model would outperform the ARIMA modeling is false. And my findings are inconclusive. Further study might involve the use of a different dataset in size and frequency.

## Sources

Nademi, Arash, and Younes Nademi. “Forecasting Crude Oil Prices by a Semiparametric Markov Switching Model: OPEC, WTI, and Brent Cases.” *Energy Economics*, vol. 74, Aug. 2018, pp. 757–766.

“U.S. Energy Information Administration - EIA - Independent Statistics and Analysis.” *Short-Term Energy Outlook - U.S. Energy Information Administration (EIA)*, US Energy Information Administration, 6 Apr. 2021, www.eia.gov/outlooks/steo/marketreview/crude.php.

“Crude and Product Price Movements.” *OPEC Monthly Oil Market Report*, vol. 44, no. 1, 12 Aug. 2020, pp. 1–93., doi:10.1111/opec.12151.

Smith, Tim. “What Is Autocorrelation?” *Investopedia*, Investopedia, 29 Apr. 2021, www.investopedia.com/terms/a/autocorrelation.asp.

Hyndman , Rob J, and George Athanasopoulos. “Forecasting: Principles and Practice (2nd Ed).” *Forecasting: Principles and Practice* , Apr. 2018, otexts.com/fpp2/.

“Crude Oil Price Today | BRENT OIL PRICE CHART | OIL PRICE PER BARREL | Markets Insider.” *Business Insider*, Business Insider, markets.businessinsider.com/commodities/oil-price.

*Europe Brent Spot Price FOB (Dollars per Barrel)*, www.eia.gov/dnav/pet/hist/rbrtem.htm.