Exponential Smoothing

Kyla Ayop | Jolia Keziah Balcita

October 14, 2023

Contents

I. Introduction

Crude oil is a naturally occurring liquid petroleum product composed of hydrocarbon deposits and other organic materials formed from the remains of animals and plants that lived millions of years ago. These organisms were covered by layers of sand, silt, and rock, subject to heat and pressure, and eventually turned into a type of fossil fuel that is refined into usable products, including gasoline, diesel, liquefied petroleum gases, and feedstock for the petrochemical industry. Therefore making Crude oil as one of the world's most important commodities, and its price can have ripple effects through the broader economy. It's prices and production are influenced by multitude of factors and may cause differing changes as time passes which makes time series analysis a valuable tool for understanding its behavior.

In this analysis, we will be utilizing the methods of time series exponential smoothing and aim to uncover the patterns in crude oil importation to the United States over the years 2009 to 2021. This involves discussing the cycles, trends, and seasonal patterns of the data. Additionally, we will encompass a variety of exponential smoothing models so that we will be able to process and analyze the behavior of the data. These models are Simple Exponential Smoothing, Holt Smoothing, Damped-Holt Smoothing, Holt-Winter's Smoothing, and Damped Holt-Winter's Smoothing. By the end of this will be our final conclusion for the best fit model to our given data.

II. The Data

The dataset used in this analysis is a historical data of Crude Oil imports from various parts of the world to the United States. It tallied the monthly thousand crude oil barrels imported in 2009 up until the first month of 2021.

III. Type of Decomposition

Historical Plot

```
library(readr)
library(fpp2)

ico <- read_csv("Imports Crude Oil.csv")
icots <- ts(ico$`Imports of all grades of crude oil from World to Total U.S. (US), Monthly (thousand ba

autoplot(icots) + labs(x = "Year", y = "Crude Oil (thousand Barrels)") + theme_bw()
```

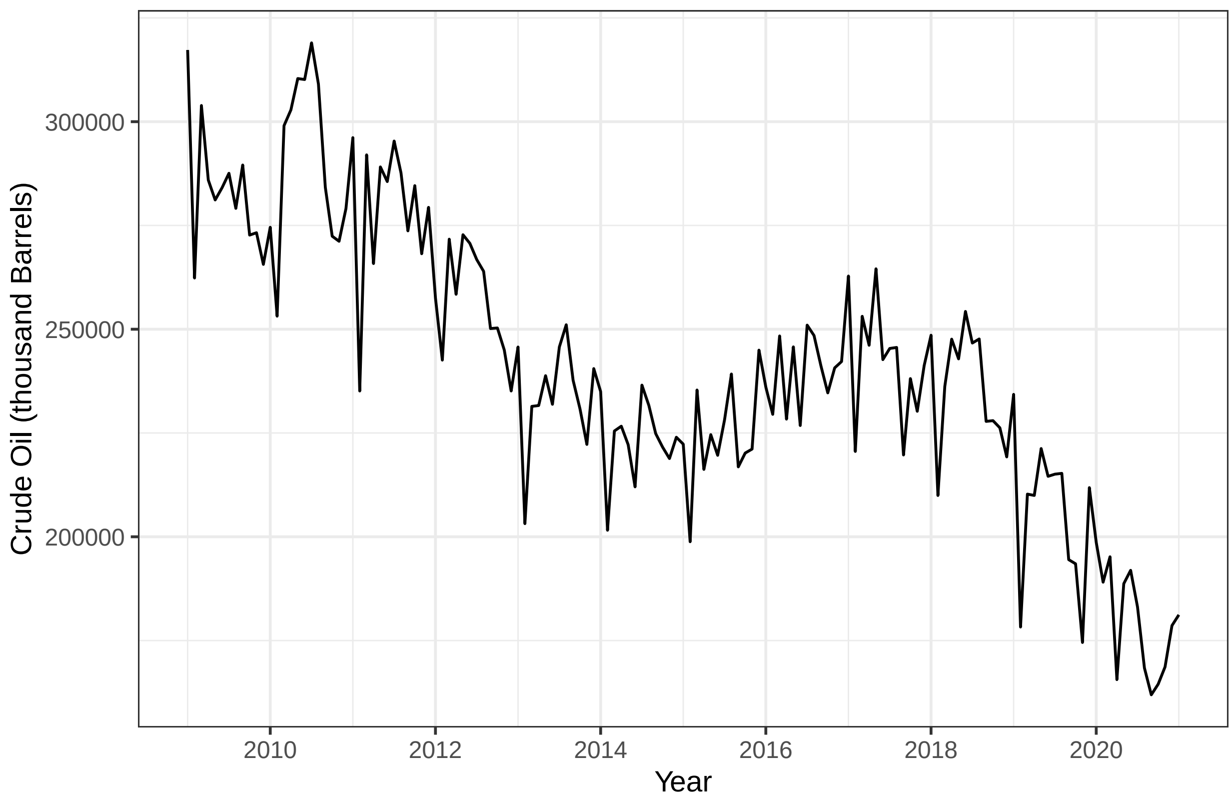


Figure 1: Monthly Totals of all Crude Oil Import from World to US (2009-2021).

We can see from this time series that there is certainly some seasonal variation in the number of crude oil imports in the U.S, per month; at the start of every year shows a visible trough and then peaks throughout the next months. This time series could be described using an additive model as the seasonal fluctuations are roughly constant in size over time and do not seem to depend on the level of the time series, and the random fluctuations seem constant over time.

VI. Components

TREND - CYCLE

```r
data("icots")

decomp_result <- decompose(icots)

autoplot(decomp_result$trend) + xlab("Year") + ylab("Crude Oil (thousand Barrels)") + theme_bw()
```

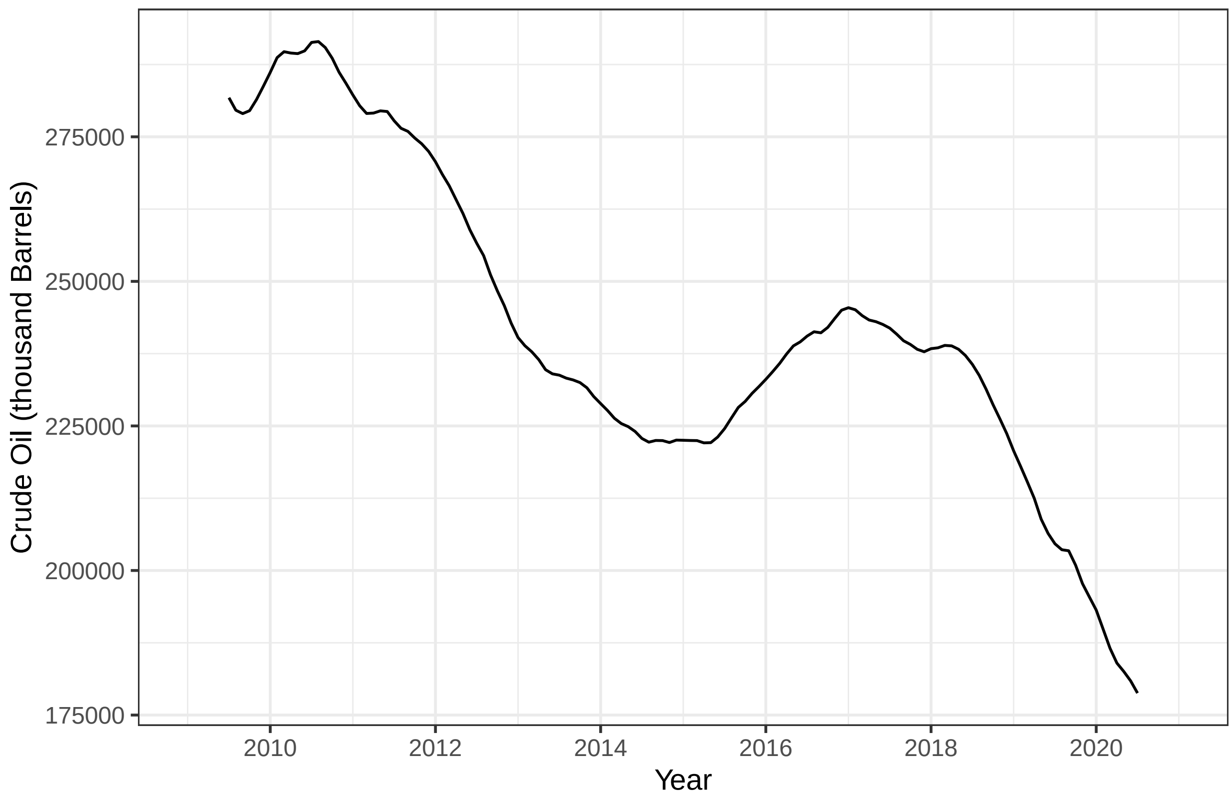


Figure 2: Trend-Cycle of Monthly Totals of all Crude Oil Import from World to US (2009-2021).

It is shown in the plot that the monthly totals of crude oil imports in the U.S. decreased overtime from 2010 to 2014 which then increase in the middle of 2015. However, the imports further decreased again at the start of the year 2017.

DETRENDED SERIES

```
trend_component <- decomp_result$trend

detrended_data <- icots/trend_component

autoplot(detrended_data) + xlab("Year") + ylab("Crude Oil (thousand Barrels)") + theme_bw()
```

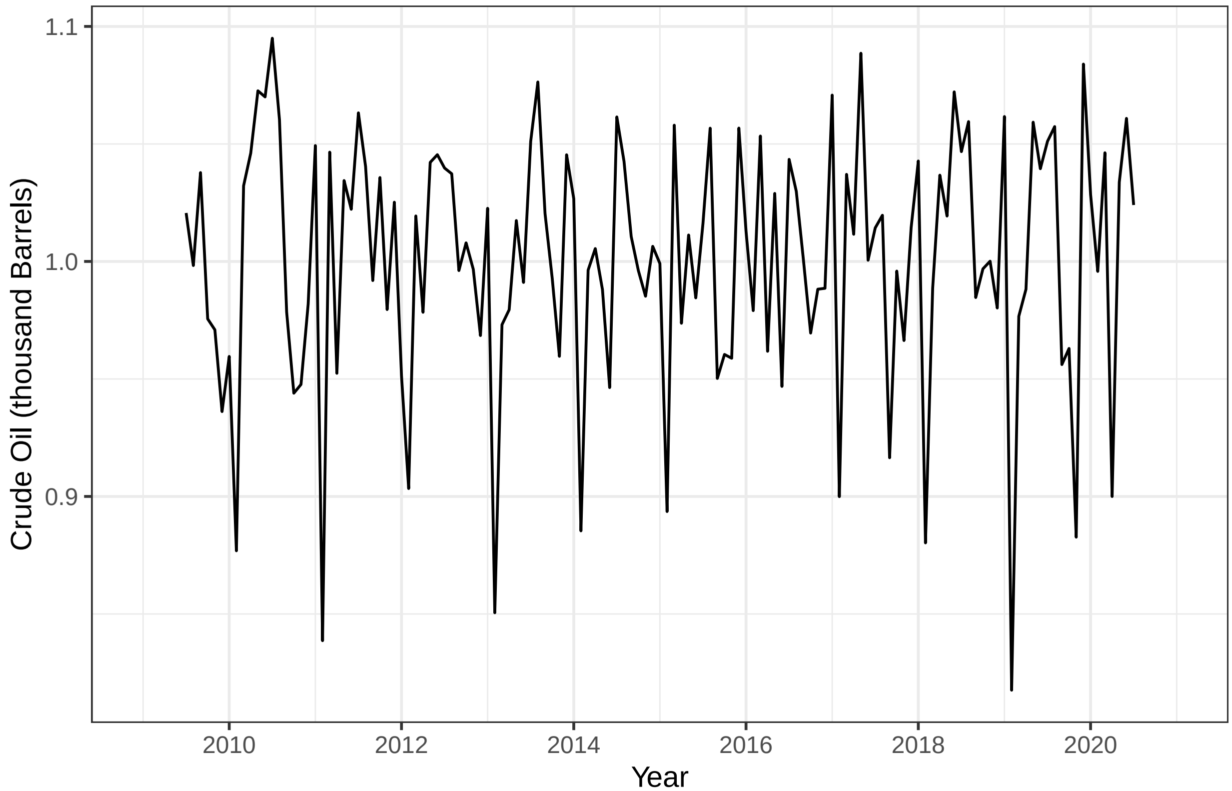


Figure 3: Detrended Monthly Totals of All Crude Oil Import from World to US (2009-2021)

This plot provides an easier visual of subtrends in the data that are seasonal or cyclical. The imports seemed to have a constant fluctuations each year where in the last months of the year until the starting month of the next year have showed decreased in imports and then a sudden increase for the next months after that.

SEASONAL COMPONENT

```r
L <- length(detrended_data)
ff <- 12
periods <- L%/%ff
index <- seq(1, L, by = ff) - 1

sf <- numeric(ff)
for (i in 1:ff) {
  sf[i] <- mean(detrended_data[index + i], na.rm = TRUE)
}

month <- c("January", "February","March","April","May",
          "June","July","August","September","October",
          "November","December")

season_data <- data.frame("Month"= month, "Seasonal.Factor"= sf )

knitr::kable(season_data,
             caption = "Seasonal Factor for each Seasonal Indices (Months)",
             digits = 4)
```

Table 1: Seasonal Factor for each Seasonal Indices (Months)

Month	Seasonal.Factor
January	1.0205
February	0.8928
March	1.0207
April	0.9849
May	1.0360
June	1.0163
July	1.0439
August	1.0435
September	0.9858
October	0.9852
November	0.9669
December	1.0079

```r
plot.ts(sf, ylab = "Seasonal factor", xlab = "Month", cex = 1)
```

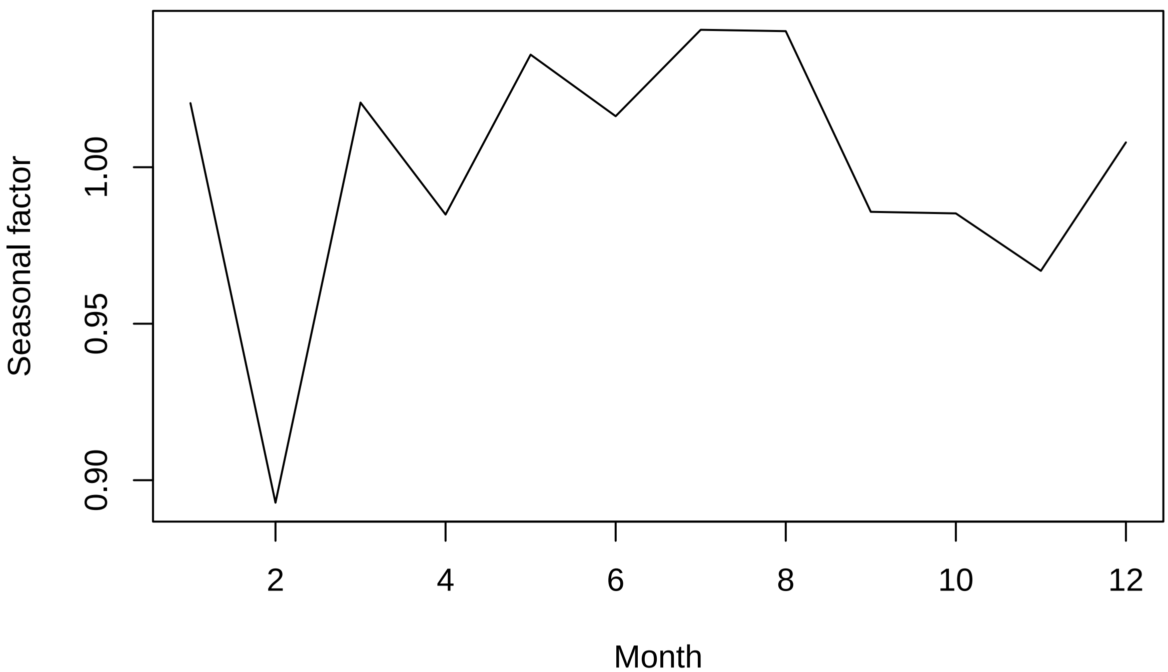


Figure 4: Seasonal Factor of Monthly Totals of All Crude Oil Import from World to US (2009-2021).

```r
L <- length(detrended_data)
ff <- 12
periods <- L%/%ff
index <- seq(1, L, by = ff) - 1


sf <- numeric(ff)
for (i in 1:ff) {
  sf[i] <- mean(detrended_data[index + i], na.rm = TRUE)
}


season_comp <- ts(rep(sf, periods + 1)[seq(L)], start = start(detrended_data),frequency = ff)
autoplot(season_comp) + xlab("Year") + ylab("Crude Oil (thousand Barrels)")
```

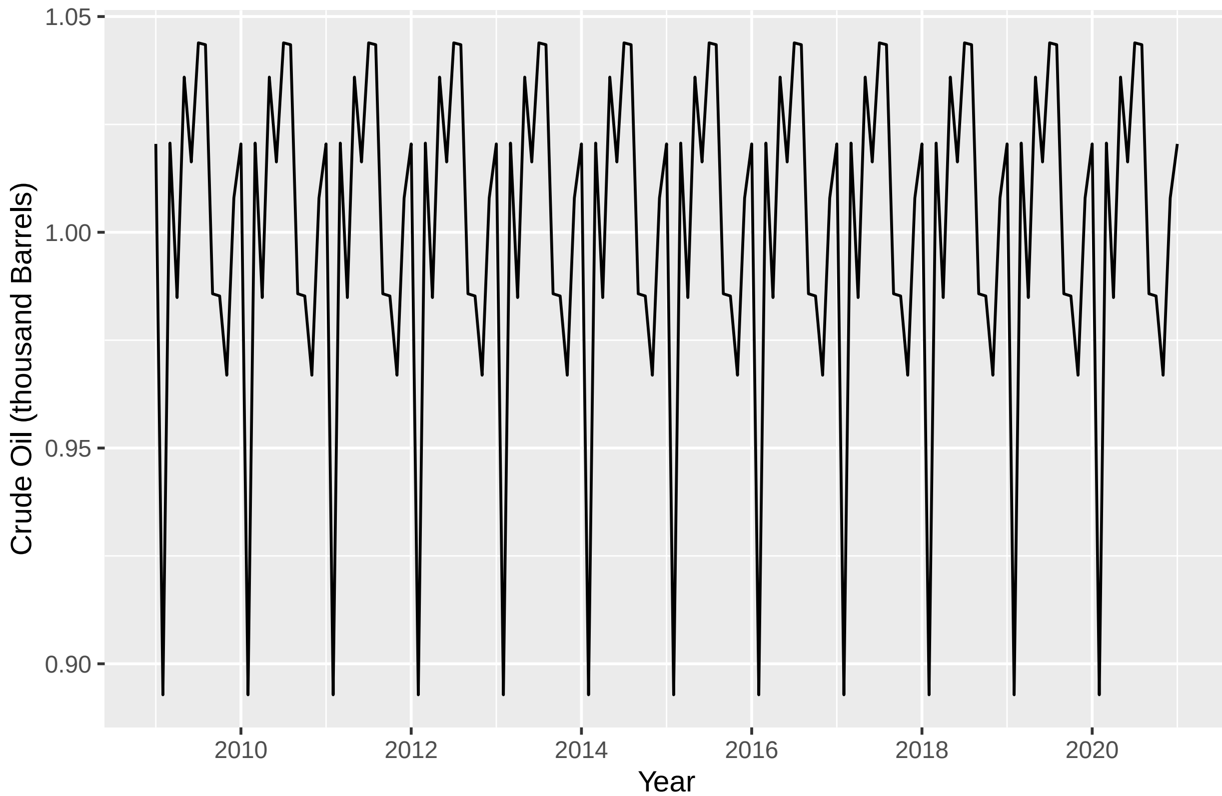


Figure 5: Seasonal Component Estimates of Monthly Totals of All Crude Oil Import from World to US (2009-2021).

The average seasonal pattern is visualized here. It can be identified that the imports are typically lowest in February where it started decreasing during January. Subsequently, the imports increased around February and March which then roughly maintained its peak until August.

IRREGULAR COMPONENT

```
Irregular = detrended_data - season_comp
plot.ts(Irregular, xlab="Year",ylab="Crude Oil (thousand Barrels)")
```

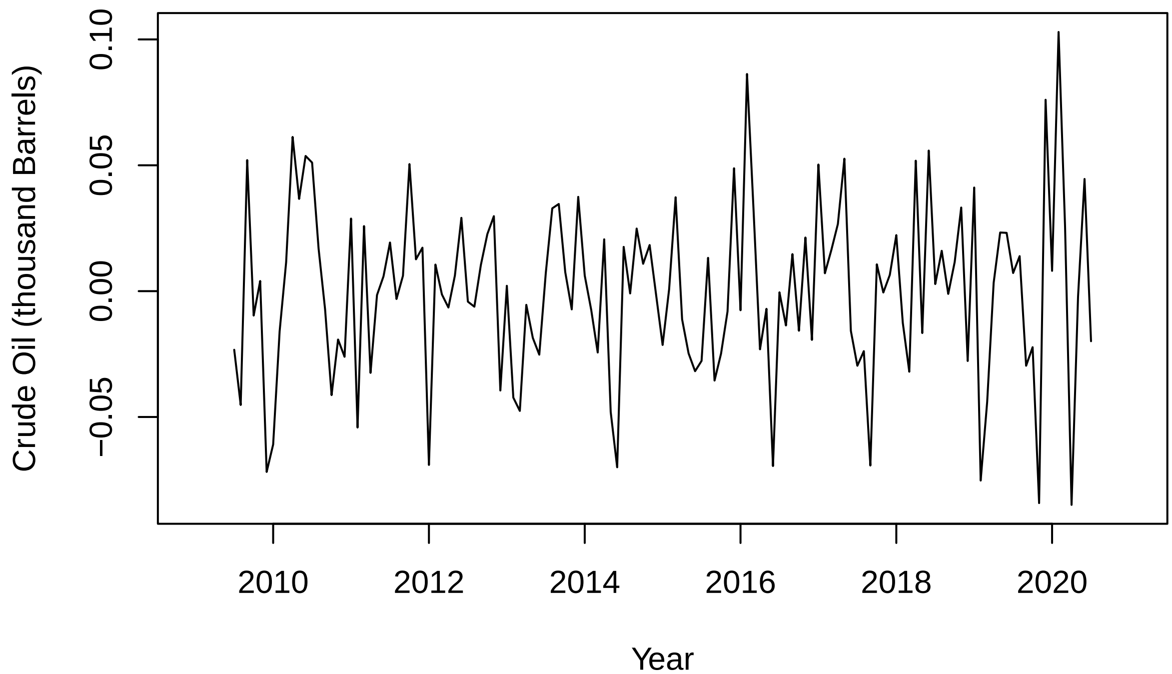


Figure 6: Irregular Component of Monthly Totals of All Crude Oil Import from World to US (2009-2021).

This plot displays the irregular component of the crude oil imports, which comprises extreme values, namely in 2016 and 2020. The outlier evident in 2020 must have been because of the worldwide lockdown in occurence with a pandemic that roughly lasted for about 2 years. This caused the usage of crude oil to decline.

Decomposition Summary

```r
plot(decompose(icots))
```

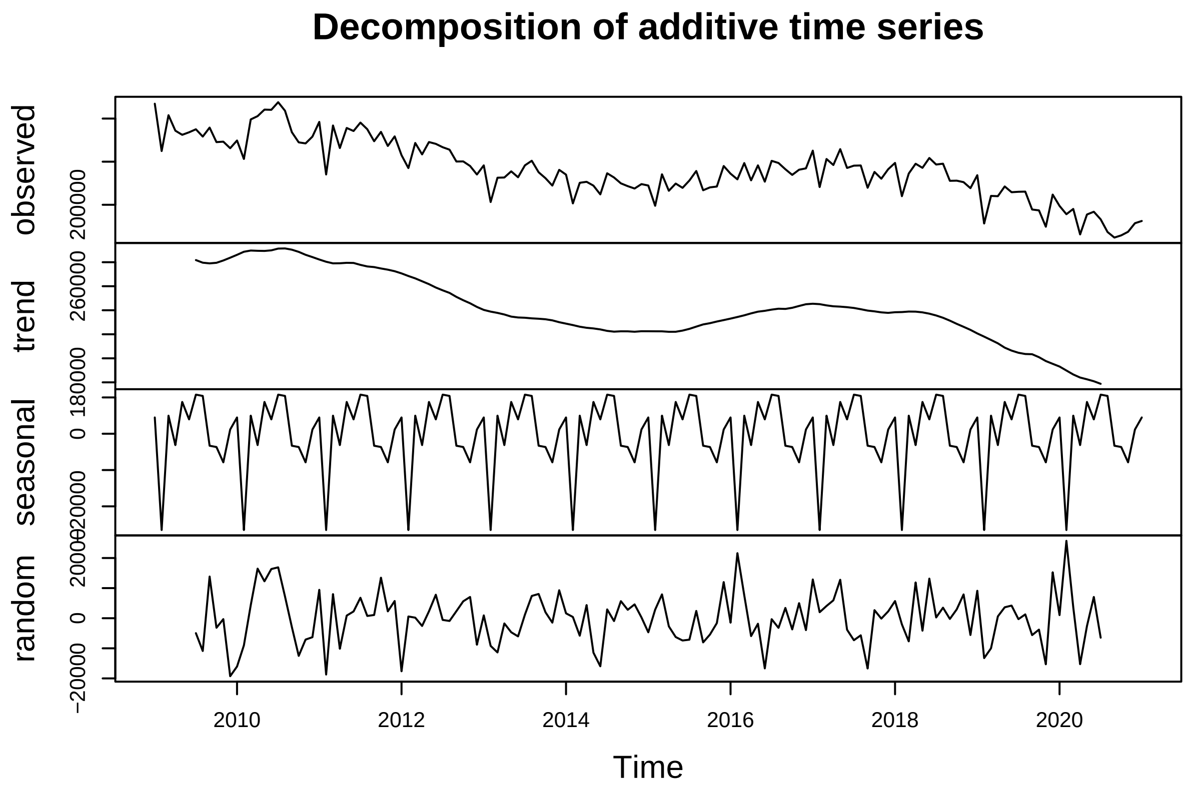


Figure 7: Seasonal Effect of Monthly Totals of All Crude Oil Import from World to US (2009-2021).

IV. Exponential Smoothing and Forecast

```
library(readr)
library(fpp2)
library(forecast)

crude_oil <- read_csv("Imports Crude Oil.csv")

icots <- ts(crude_oil$`Imports of all grades of crude oil from World to Total U.S. (US), Monthly (thousa
plot(icots, main= "Monthly Totals of Crude Oil Import (2009-2021)",
        xlab = "Time", ylab = "Crude Oil (thousand barrels)")
```

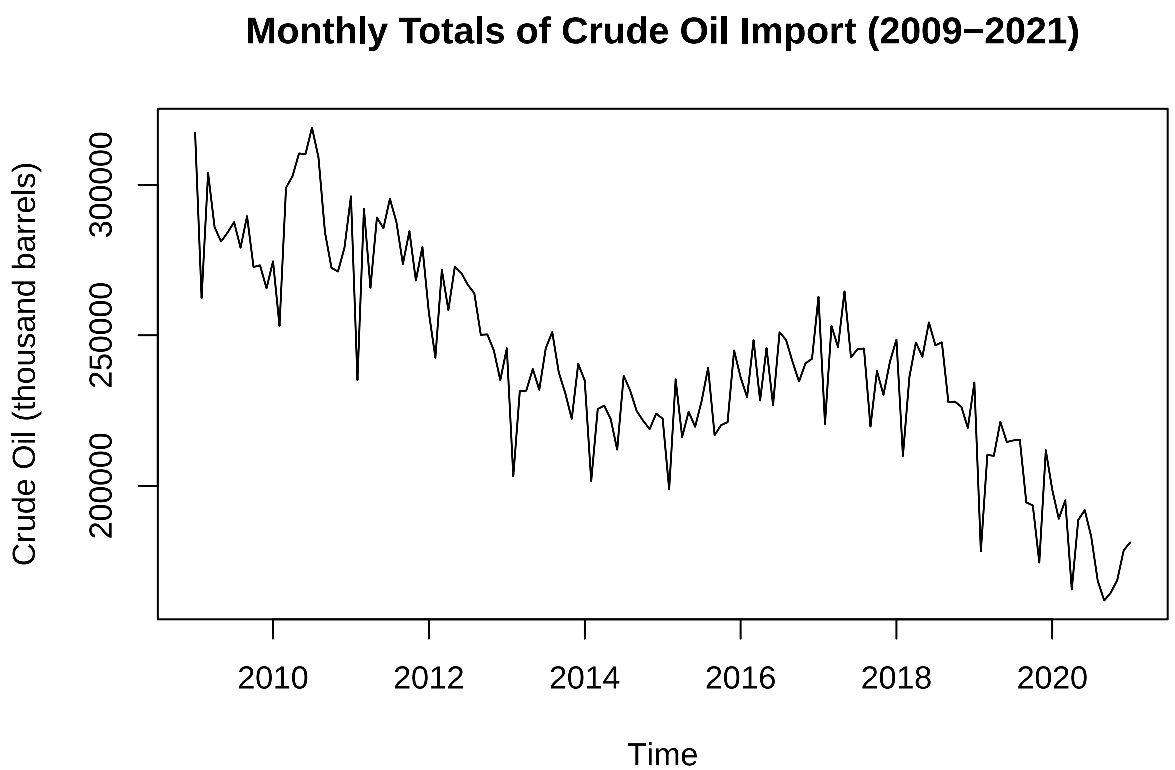
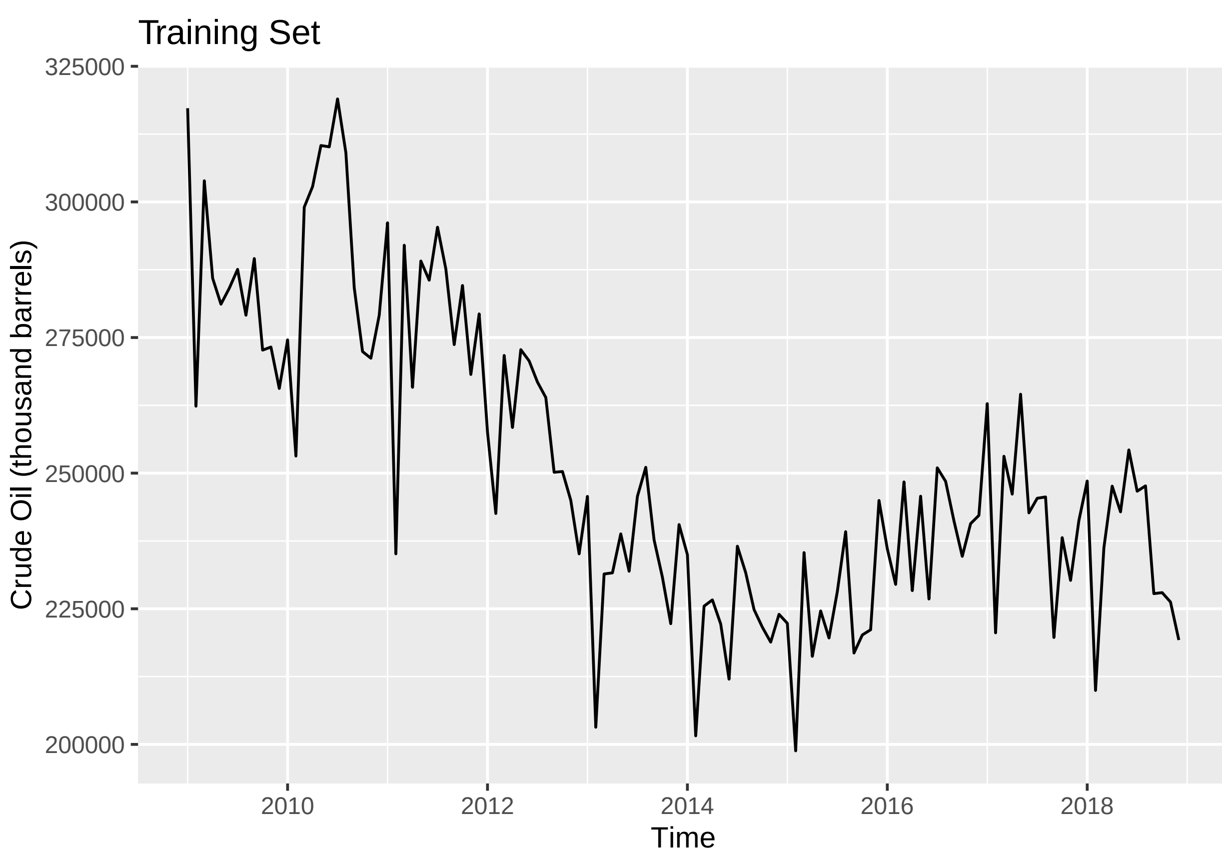


Figure 8: Monthly Totals of All Crude Oil Import from World to US (2009-2021).

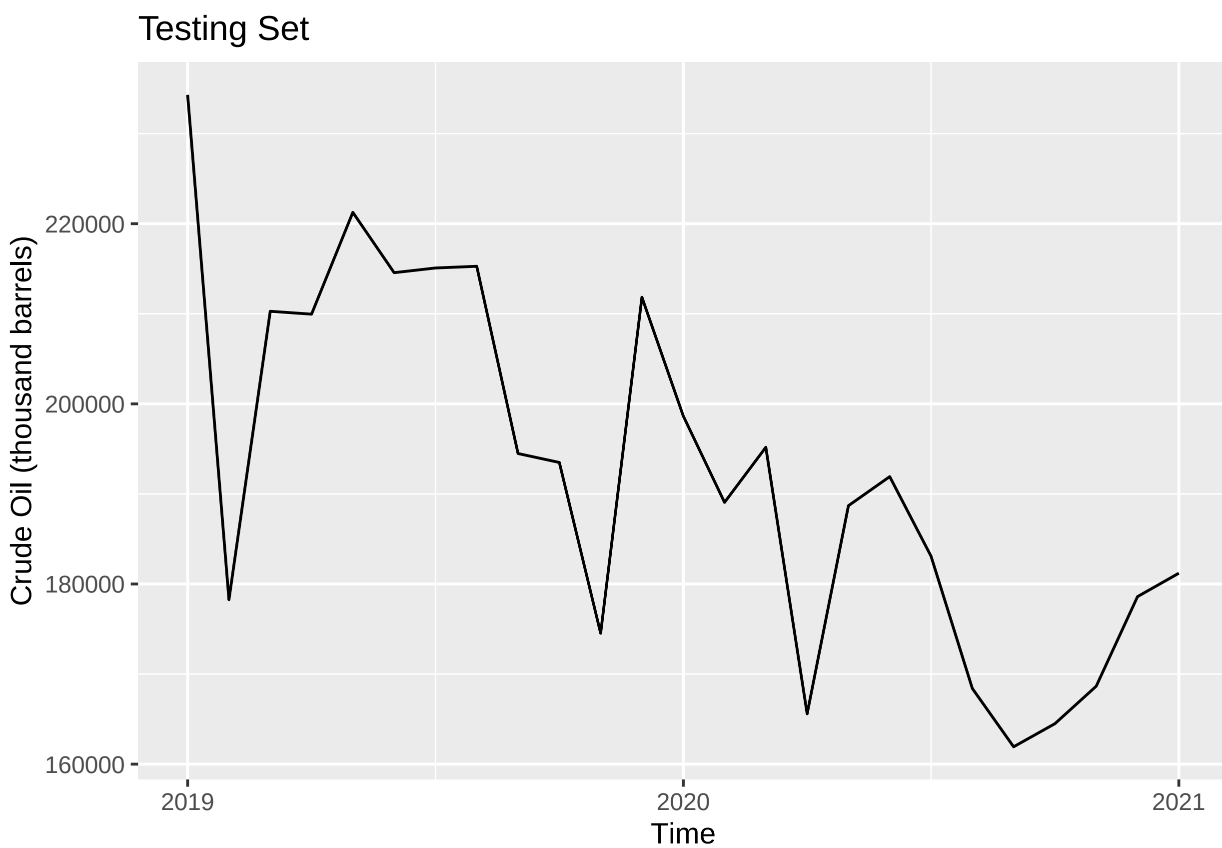
DATA PARTITIONING

The dataset, imports of crude oil from the World to the USA from 2009 to 2021, being prepared for time series analysis contains 145 months of observation, and the data partitioning is done using an 80:20 ratio, which is a common practice for splitting data into training and test sets. The training set, in this context, is the subset of the data that includes the initial 80% of the observations. It corresponds to the “crude_oil_training_set,” which is created using the first 120 out of the 145 total observations. This training set is used to develop and train the time series forecasting model. Allowing the model to learn and make accurate forecasts. The test set, on the other hand, represents the remaining 20% of the data and is defined as the “crude_oil_testing_set” . It consists of the last 25 months of observations. It helps assess how well the model generalizes to unseen data.

```r
#DATA PARTITIONING

crude_oil_training_set <- icots[1:120]
crude_oil_training_set <- ts(crude_oil_training_set,start=2009,frequency=12)
crude_oil_testing_set <- icots[121:145]
crude_oil_testing_set <- ts(crude_oil_testing_set,start=2019, frequency=12)

autoplot(crude_oil_training_set,main="Training Set")+ylab("Crude Oil (thousand barrels)")
```




```r
autoplot(crude_oil_testing_set,main="Testing Set")+ylab("Crude Oil (thousand barrels)")
```


Testing Set
220000
200000
180000
160000
Crude Oil (thousand barrels)
2019
2020
2021
Time

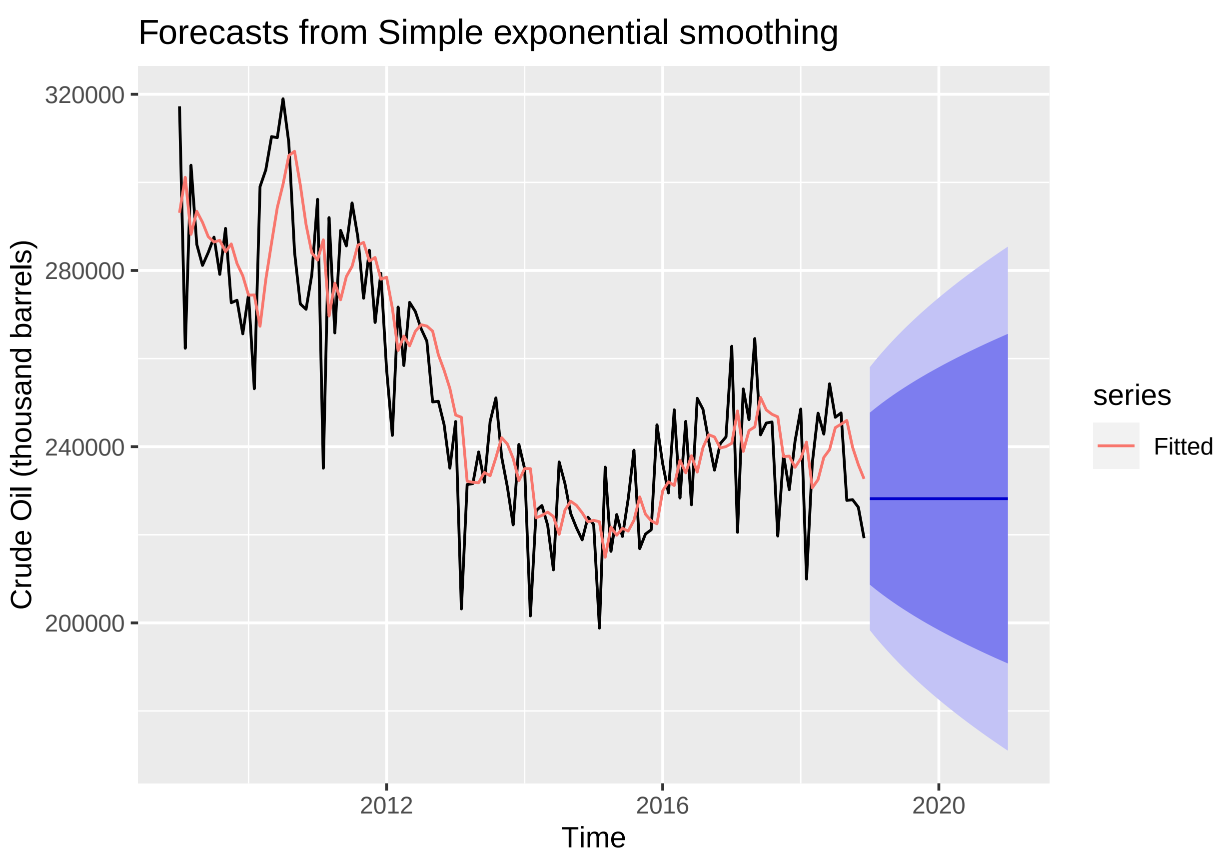
FORECASTING MODELS

In this activity, we are consider five models: Simple Exponential Smoothing, Holt Smoothing, Damped Holt Smoothing, Holt-Winter's Smoothing, and Damped- Holt Winter's Smoothing, in order. All functions used are under the forecast package. The models are then trained using the training set. During this phase, the model learns from past data and internalizes the patterns, trends, and seasonality present in the time series.

Note: In the forecast model plots, the blue line represents the forecasted or predicted values, while the surrounding blue space typically indicates the confidence level or the uncertainty associated with those forecasts. The width of the blue space around the forecasted line represents the confidence interval. A wider blue space generally indicates higher uncertainty, and a narrower blue space suggests higher confidence in the forecasted values.

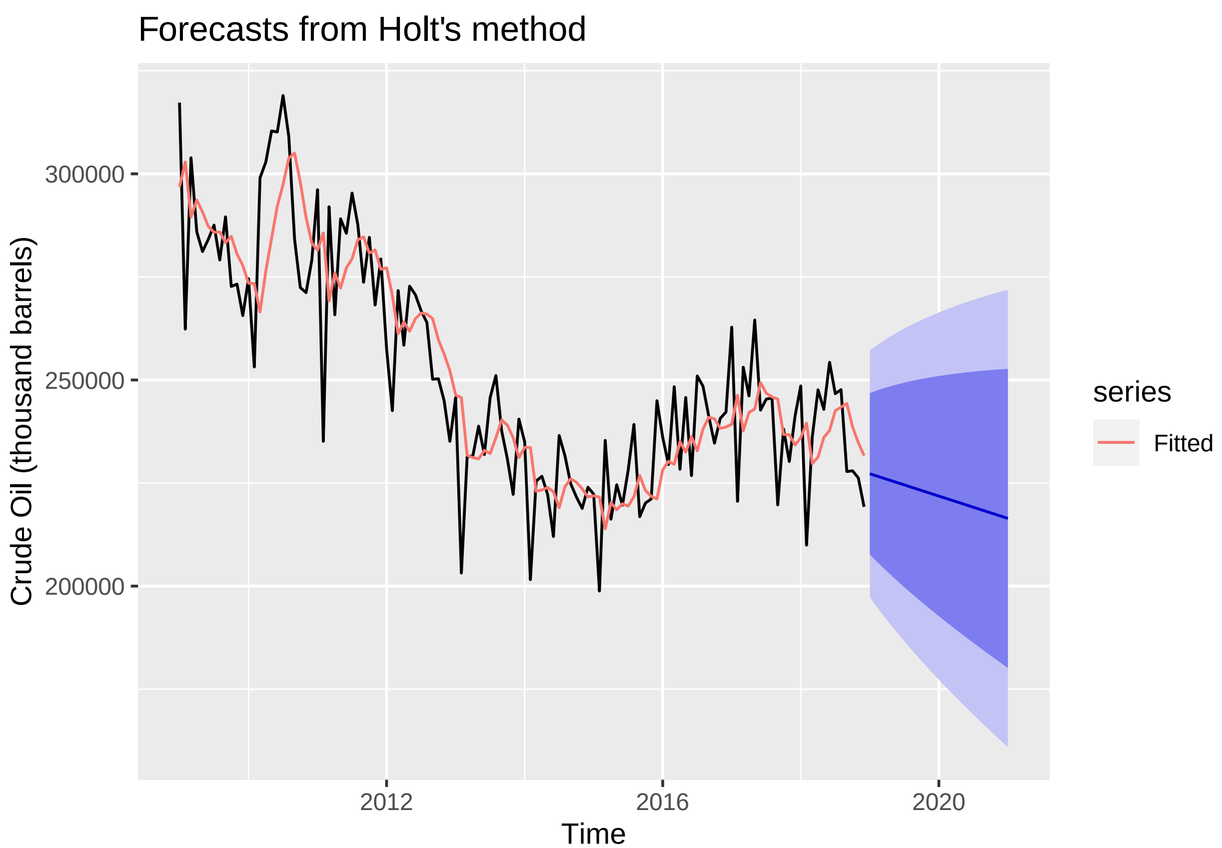
Simple Exponential Smoothing (SES) model has been fitted to the training set. Using the function ses(y,h,...), the model is then used to make forecasts for a period matching the length of the testing set. The SES model is particularly suitable for data with no clear trend or seasonality.

```
#Simple Exponential Smoothing
crude_oil_ses_model <- ses(crude_oil_training_set,h=length(crude_oil_testing_set))
autoplot(crude_oil_ses_model)+ autolayer(fitted(crude_oil_ses_model),series="Fitted")+
ylab("Crude Oil (thousand barrels)")
```

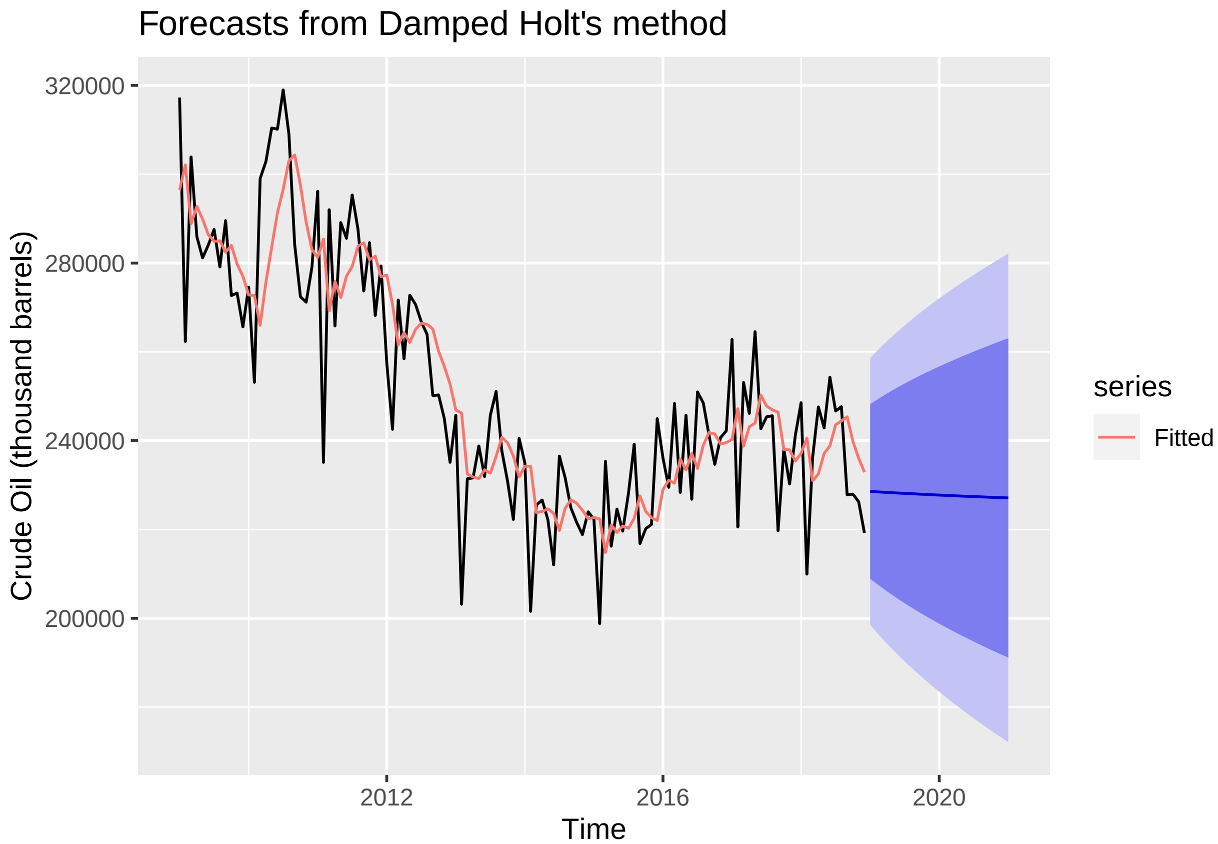


For holt smoothing and damped holt smoothing, holt(y,h,damped) function is use in the training set, where damped holt is utilized when damped = TRUE. It includes a linear trend component in addition to level and smoothing parameters and is designed for time series data with a linear trend but no seasonality.

```
#Holt's Method
crude_oil_holt_model <- holt(crude_oil_training_set,h=length(crude_oil_testing_set))
autoplot(crude_oil_holt_model)+ autolayer(fitted(crude_oil_holt_model),series="Fitted")+ylab("Crude Oil
```

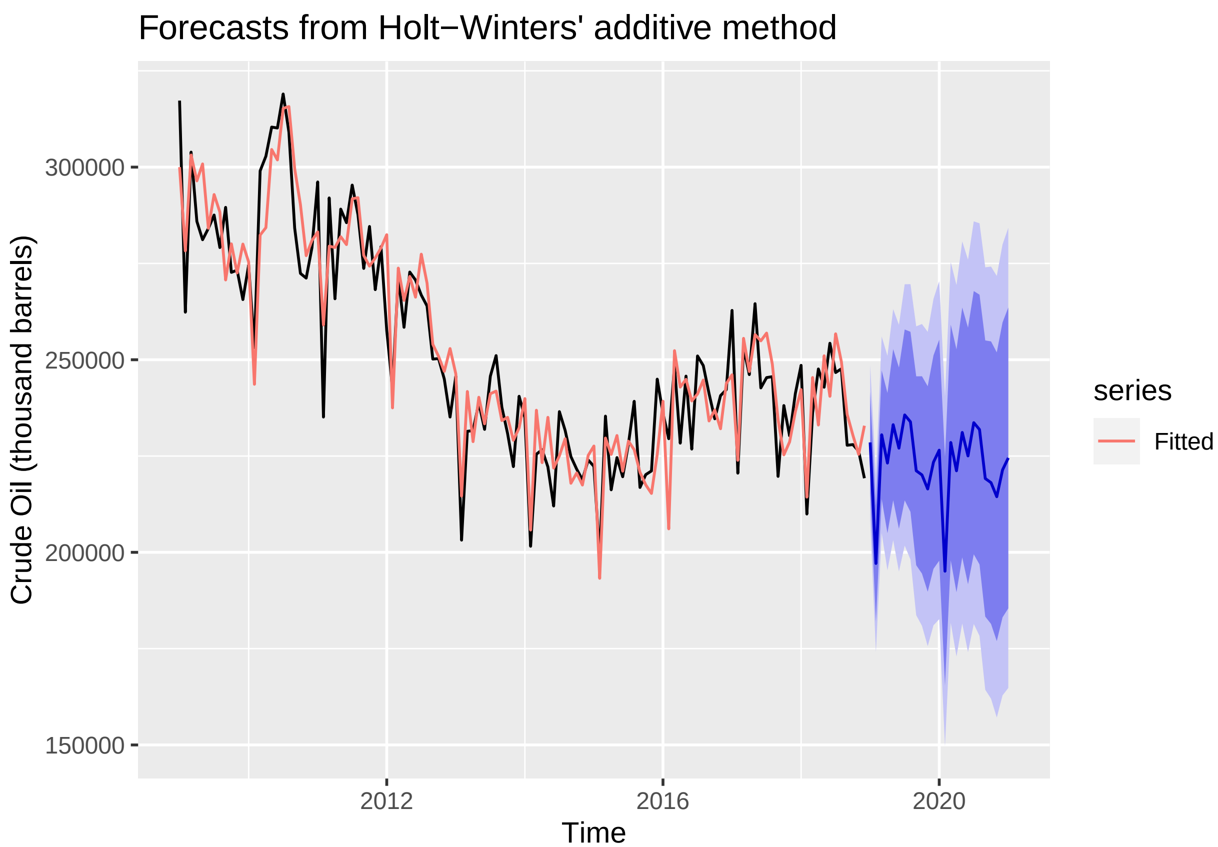



```
#Damped Holt Smoothing (Trend)
crude_oil_holtD_model <- holt(crude_oil_training_set,h=length(crude_oil_testing_set),damped = TRUE)
autoplot(crude_oil_holtD_model)+ autolayer(fitted(crude_oil_holtD_model),series="Fitted")+ylab("Crude Oi
```

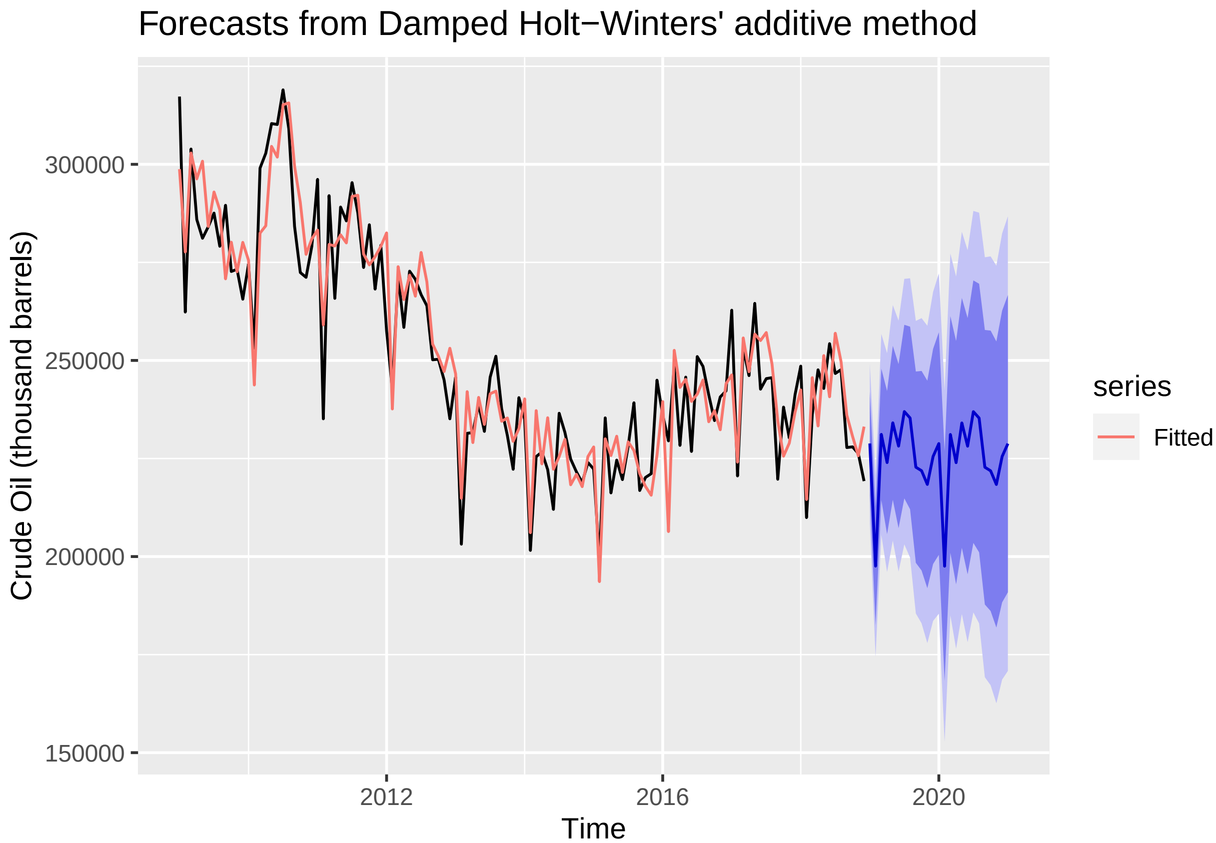

Forecasts from Damped Holt's method
Crude Oil (thousand barrels)
320000
280000
240000
200000
2012
2016
2020
Time
series
Fitted

The Holt-Winter's Smoothing is designed to handle time series data with both a trend and seasonality, making it particularly useful for data that exhibits recurring patterns over time.By using the hw(y,h,damped,seasonal="additive") function to fit a Holt-Winters' model to the training data.

```
#Holt-Winter's Method
crude_oil_hw_model <- hw(crude_oil_training_set,h=length(crude_oil_testing_set))
autoplot(crude_oil_hw_model)+ autolayer(fitted(crude_oil_hw_model),series="Fitted")+
ylab("Crude Oil (thousand barrels)")
```




```
#Damped Holt-Winter Smoothing
crude_oil_hwD_model <- hw(crude_oil_training_set,h=length(crude_oil_testing_set),damped=TRUE)
autoplot(crude_oil_hwD_model)+ autolayer(fitted(crude_oil_hwD_model),series="Fitted")+
ylab("Crude Oil (thousand barrels)")
```



Thus, the forecasts in each model represent the expected values for crude oil imports in the testing set based on the patterns and trends learned from the training set.The fitted values shown represent how well the model fits the training data.

MODEL SELECTION TESTING PERFORMANCE

```r
#MODEL SELECTION TESTING PERFORMANCE

accuracy(crude_oil_ses_model,x=crude_oil_testing_set)
```

```
##                      ME     RMSE      MAE         MPE      MAPE      MASE
## Training set  -1619.926 15103.33 11515.71  -0.9277362  4.629542 0.7503934
## Test set     -35848.081 40898.06 36336.68 -19.8770470 20.085576 2.3677917
##                     ACF1 Theil's U
## Training set -0.08476644        NA
## Test set      0.45313264  2.380562
```

```r
accuracy(crude_oil_holt_model,x=crude_oil_testing_set)
```

```
##                       ME     RMSE      MAE         MPE      MAPE      MASE
## Training set   -401.9192 15019.73 11470.41  -0.4295694  4.591556 0.7474415
## Test set     -29488.9535 34234.10 30053.22 -16.4084304 16.649254 1.9583453
##                    ACF1 Theil's U
## Training set -0.0604972        NA
## Test set      0.3360598  1.980058
```

```r
accuracy(crude_oil_holtD_model,x=crude_oil_testing_set)
```

```
##                      ME     RMSE      MAE         MPE      MAPE      MASE
## Training set   -797.576 15008.60 11444.97  -0.6099592  4.587042 0.7457835
## Test set     -35428.727 40374.36 35888.02 -19.6396161 19.835639 2.3385560
##                     ACF1 Theil's U
## Training set -0.05943503        NA
## Test set      0.43929991  2.348053
```

```r
accuracy(crude_oil_hw_model,x=crude_oil_testing_set)
```

```
##                      ME      RMSE       MAE        MPE      MAPE      MASE
## Training set   -894.617  9569.594  7605.971  -0.442606  3.031433 0.4956247
## Test set     -30880.342 35664.736 31342.046 -17.049820 17.246871 2.0423285
##                     ACF1 Theil's U
## Training set -0.08955484        NA
## Test set      0.60534249  2.091944
```

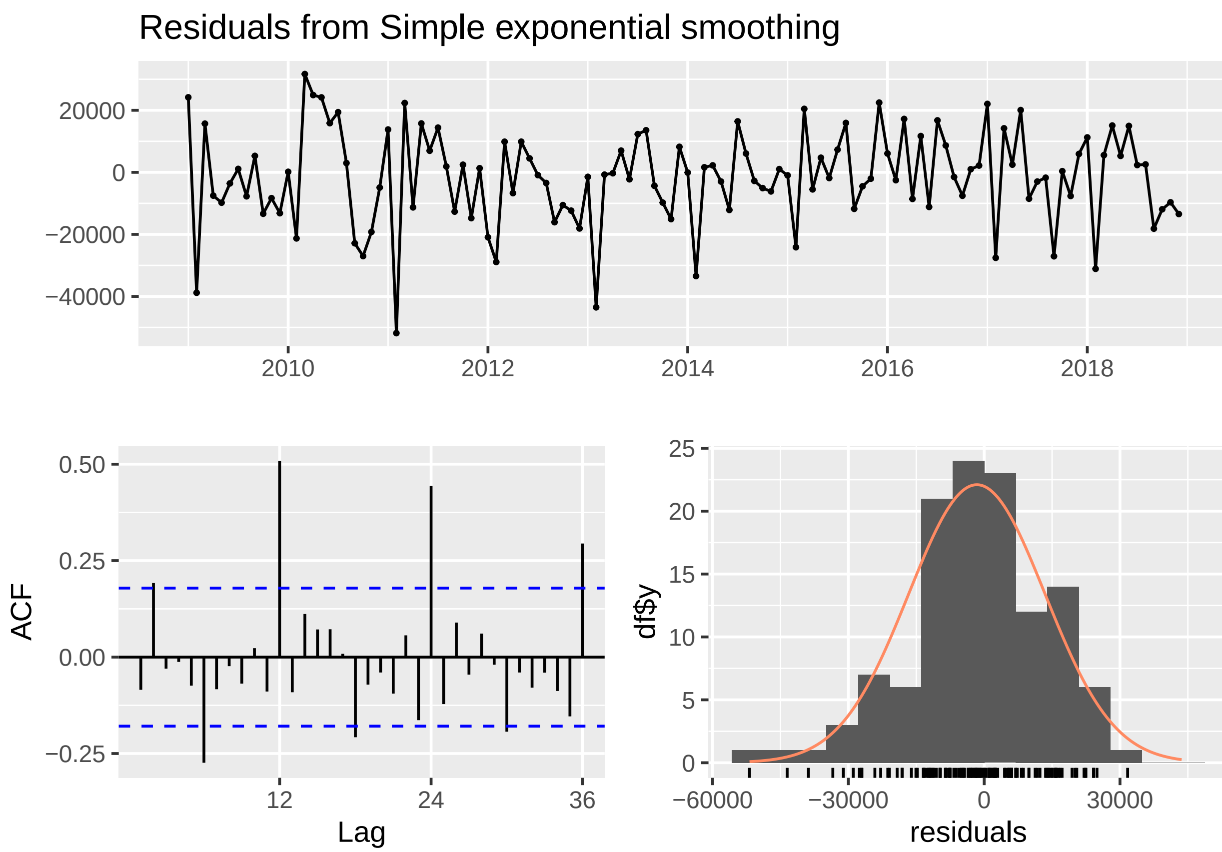
```r
accuracy(crude_oil_hwD_model,x=crude_oil_testing_set)
```

```
##                      ME      RMSE       MAE         MPE      MAPE      MASE
## Training set  -1072.903  9579.497  7626.323  -0.5201111  3.040332 0.4969508
## Test set     -33161.760 38126.362 33599.856 -18.2953792 18.482355 2.1894533
##                     ACF1 Theil's U
## Training set -0.09296205        NA
## Test set      0.63253827  2.238549
```

Noted by our Professor, that a good fit model in the training stage is not necessarily a good fit the the testing stage. Then in comparing the models, we consider its testing performance, hence, lower RMSE and MAE values as better, indicating that the forecasts are closer to the actual values. A lower MPE and MAPE suggest less bias and better accuracy, and a lower MASE indicates that the model's forecasts are more efficient compared to a naïve model. The ACF1 value measures the quality of the model's predictions, and Theil's U is a comprehensive metric that combines forecast accuracy and forecasting efficiency. A general description of common time series forecasting evaluation metrics. Based on these metrics, it appears that the Holt Smoothing model performs better than the others on the testing set. It has the lowest RMSE, MAE, MPE, MAPE, and MASE values among the models, indicating that it provides the most accurate and efficient forecasts. Additionally, it has the highest ACF1 and the lowest Theil's U, further supporting its superior performance. While it's essential to consider other factors, such as the nature of the data and the purpose of the forecast, these results suggest that Holt Smoothing is the best choice among the models considered for this specific forecasting task.

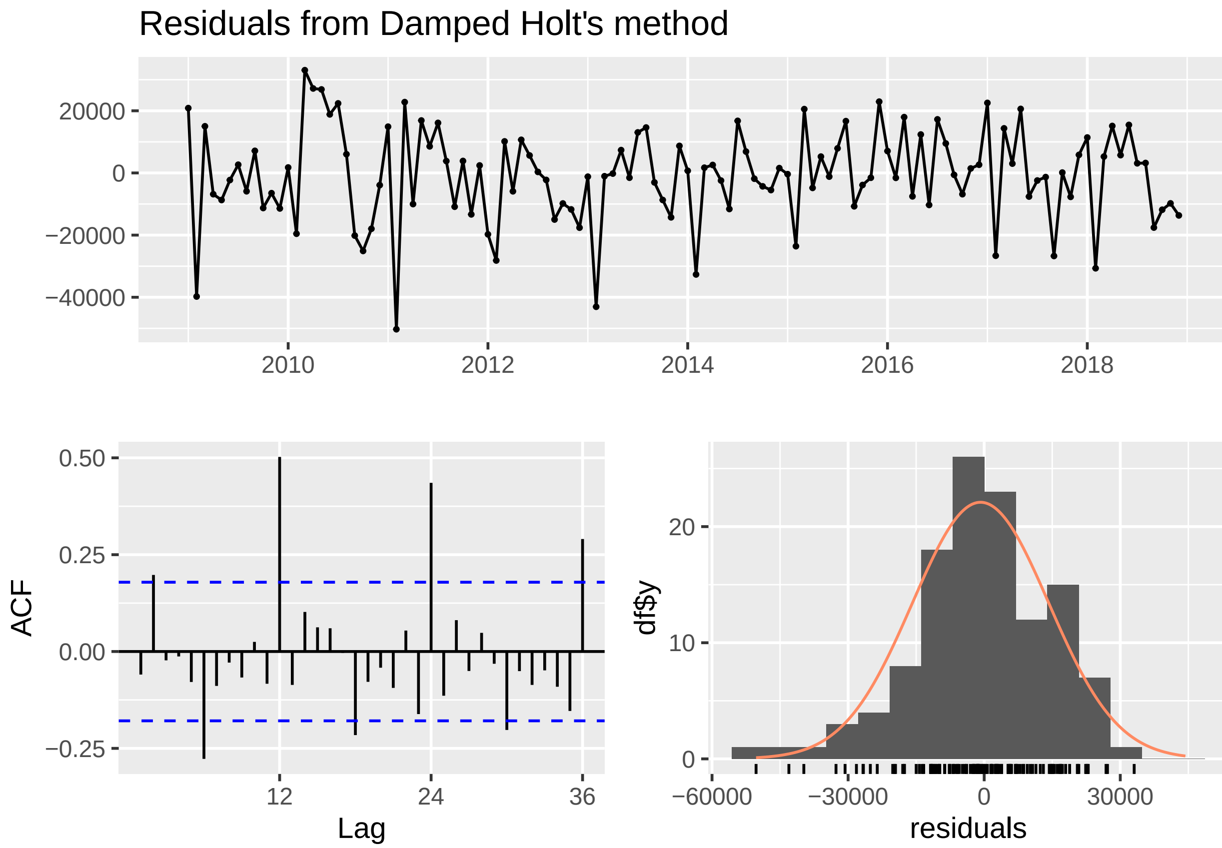
DIAGNOSTIC

```
#DIAGNOSTIC
checkresiduals(crude_oil_ses_model)
```



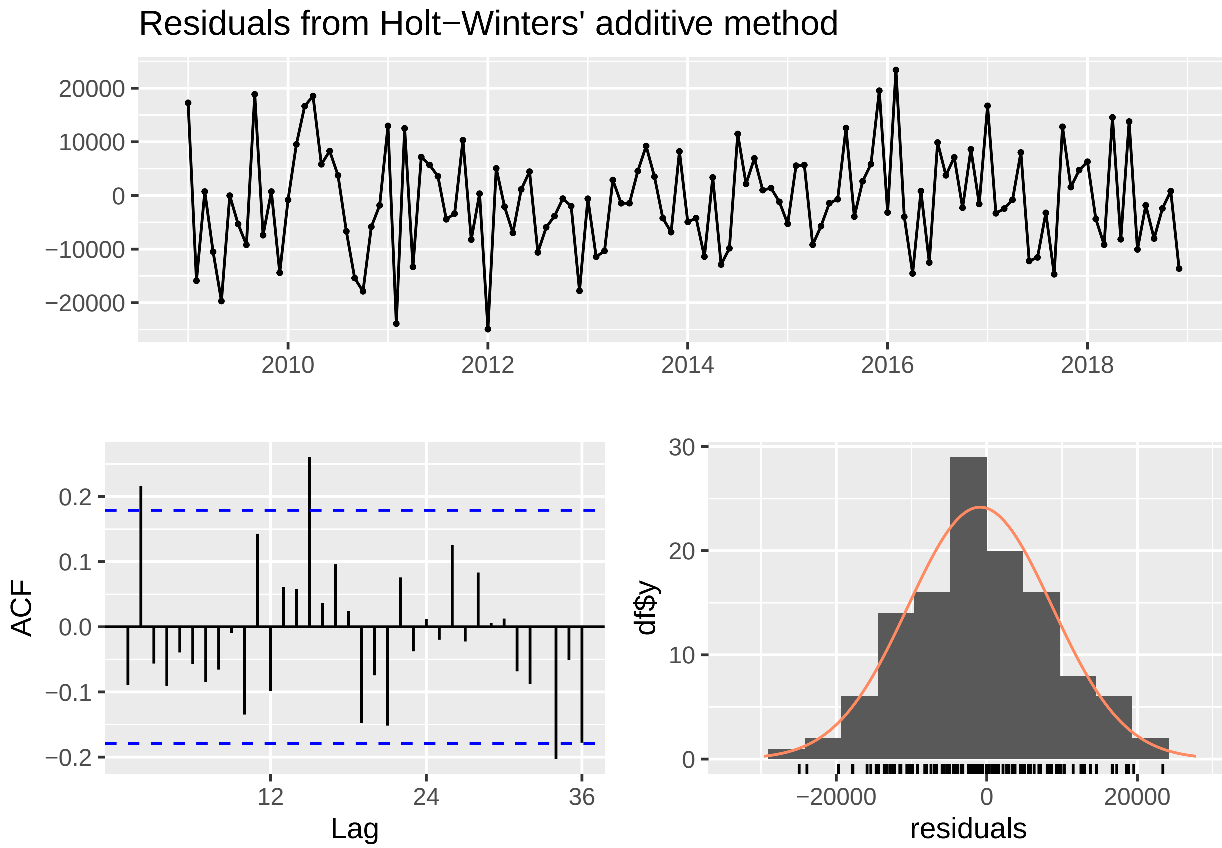

```
##
##  Ljung-Box test
##
## data:  Residuals from Simple exponential smoothing
## Q* = 101.04, df = 24, p-value = 1.999e-11
##
## Model df: 0.   Total lags used: 24
```

```
checkresiduals(crude_oil_holtD_model)
```



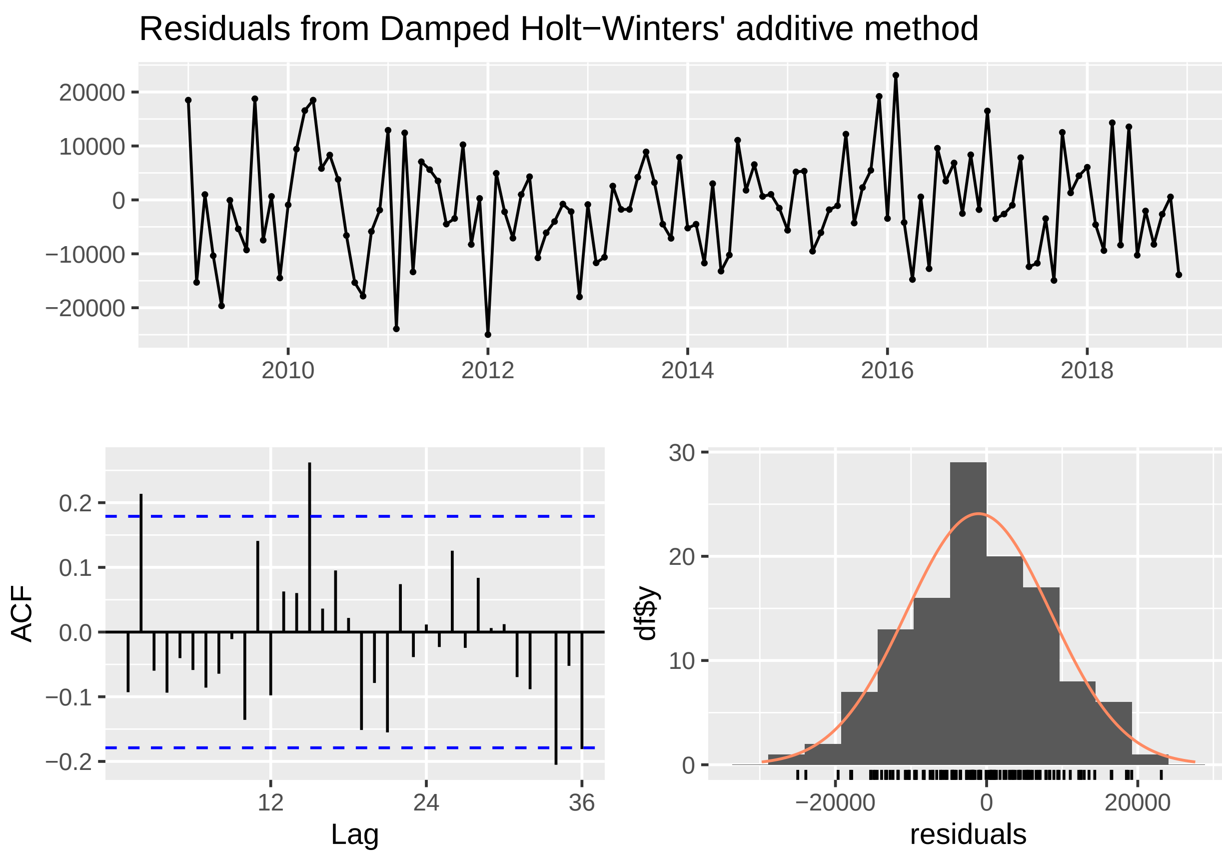

```
##
##  Ljung-Box test
##
## data:  Residuals from Damped Holt's method
## Q* = 98.88, df = 24, p-value = 4.662e-11
##
## Model df: 0.   Total lags used: 24
```

```r
checkresiduals(crude_oil_hw_model)
```



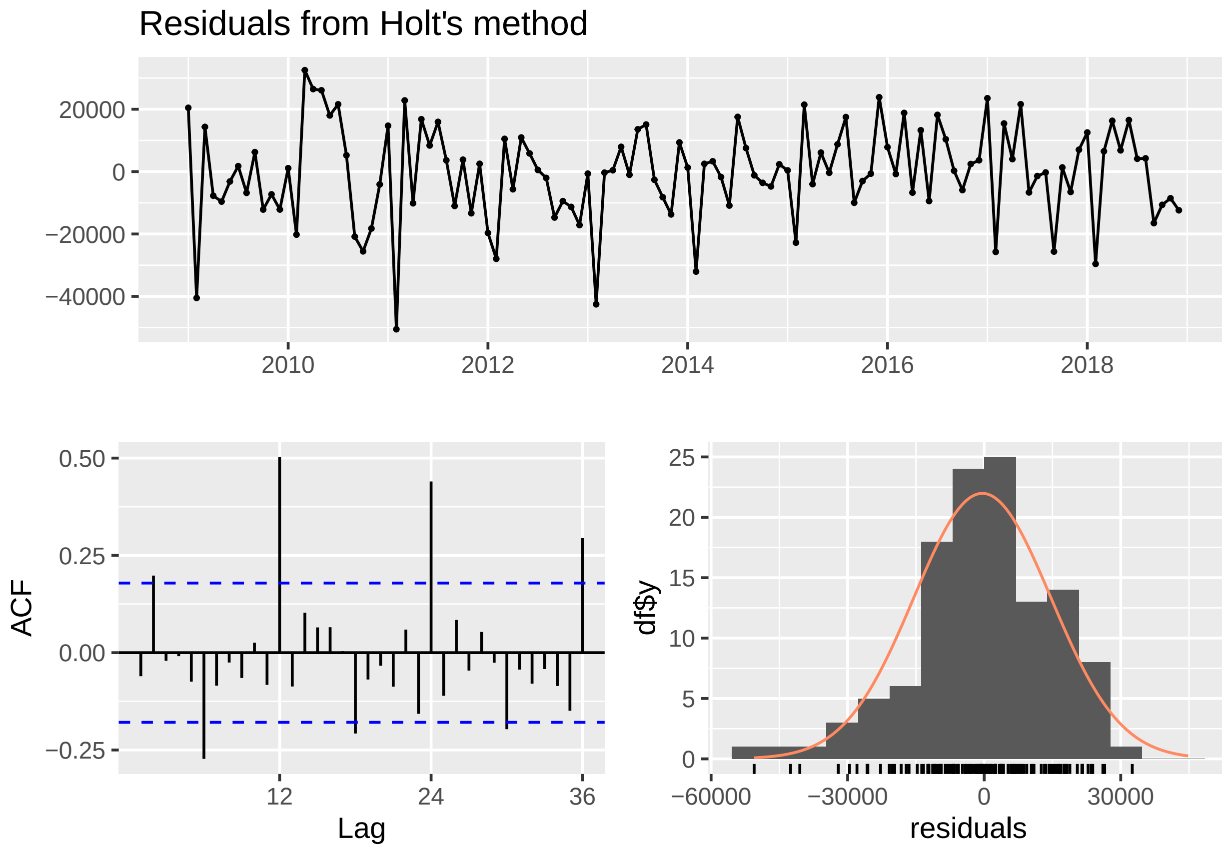

```
##
##  Ljung-Box test
##
## data:  Residuals from Holt-Winters' additive method
## Q* = 37.312, df = 24, p-value = 0.04072
##
## Model df: 0.   Total lags used: 24
```

```r
checkresiduals(crude_oil_hwD_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from Damped Holt-Winters' additive method
## Q* = 37.879, df = 24, p-value = 0.03568
##
## Model df: 0.   Total lags used: 24
```

```
checkresiduals(crude_oil_holt_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from Holt's method
## Q* = 98.115, df = 24, p-value = 6.285e-11
##
## Model df: 0.   Total lags used: 24
```

To show how similar all models are with respect on checking its residuals. It has been observed that all models depict the following but we will specify on Holt's Smoothing. • Residual plot: In this dataset, we can observed that the error fluctuates around zero but between the values of -40000 and 20000. Ten thousand more in using the histogram of the residuals. Large spikes where observed in the first month of 2010 and more likely the end or the start of 2011. • Autocorrelation Function (ACF) plot: For the Auto Correlation Plot there is an indication of autocorrelation in the residuals. In fact all models exhibit that the vertical line exceed at the blue line, which the assumption of the model is violated. • Ljung-Box test: With the null hypothesis that states that there is no serial correlation. With an alpha of 0.05, the Ljung-Box test results provide strong evidence against the null hypothesis that the residuals from Holt's method are not autocorrelated. The small p-value of approximately 6.285e-11 (or 0.00000000006285) (very close to zero) indicates that there is a significant level of autocorrelation in the residuals.

Conclusion

FINAL MODEL FOR IMPORTS OF CRUDE OIL FROM WORLD TO U.S. THOUSAND BARRELS

The Holt Smoothing model appears to be performing better than the other forecasting models, according to the evaluation of the forecasting models, giving it a strong candidate to be the final model for predicting the monthly import of crude oil to the US. Holt Smoothing is found to be the best fit because there is an obvious trend and no seasonal pattern. Additionally, we will anticipate the next five months of import (February 2021 to June 2021) using the Holt Smoothing Method, as displayed below.

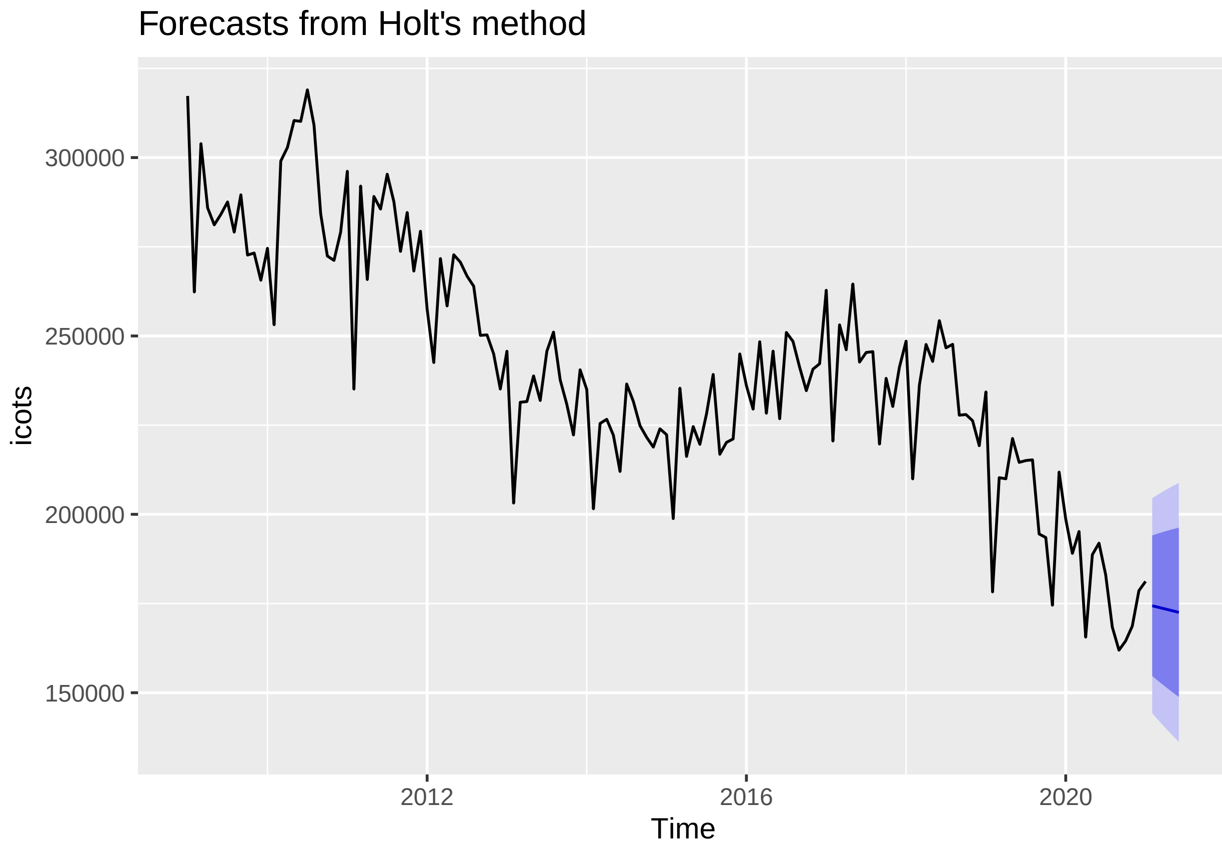
```r
#FINAL MODEL FOR IMPORTS OF CRUDE OIL FROM WORLD TO US. THOUSAND BARRELS

crude_oil_holt_model_final <- holt(icots,h=5,damped=FALSE)
crude_oil_holt_model_final$model
```

```
## Holt's method
##
## Call:
##  holt(y = icots, h = 5, damped = FALSE)
##
##   Smoothing parameters:
##     alpha = 0.3352
##     beta  = 1e-04
##
##   Initial states:
##     l = 297256.902
##     b = -446.2052
##
##   sigma:  15375.29
##
##      AIC     AICc      BIC
## 3523.320 3523.752 3538.204
```

```r
crude_oil_holt_model_final
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Feb 2021       174381.2 154677.0 194085.4 144246.2 204516.2
## Mar 2021       173918.0 153135.5 194700.6 142133.9 205702.2
## Apr 2021       173454.9 151646.7 195263.1 140102.2 206807.6
## May 2021       172991.8 150203.5 195780.1 138140.1 207843.5
## Jun 2021       172528.6 148800.1 196257.1 136239.0 208818.2
```

```r
autoplot(crude_oil_holt_model_final)
```


Forecasts from Holt's method
300000
250000
200000
150000
icots
2012
2016
2020
Time

Reference

Babiera, J. (2023). *Exponential Smoothing.*

Sharma, R. (2021). *10 Time Series Datasets for Practice.* https://medium.com/analytics-vidhya/10-time-series-datasets-for-practice-d14fec9f21bc?fbclid=IwAR1ItNNsvVfbqsXAhZLbdXiasgUlFYomUeDzjnUjoFQxNy8XdbCB4FD

U.S. Energy Information Administration. (2023, October 2). *Oil and Petrolium products explained.* https://www.eia.gov/energyexplained/oil-and-petroleum-products/imports-and-exports.php?fbclid=IwAR1Y2vPWIYFa6ZNmUfEbOiCEkX9TstkR99aKO3Gsrjf7LpCEGsn9SCUPaZw