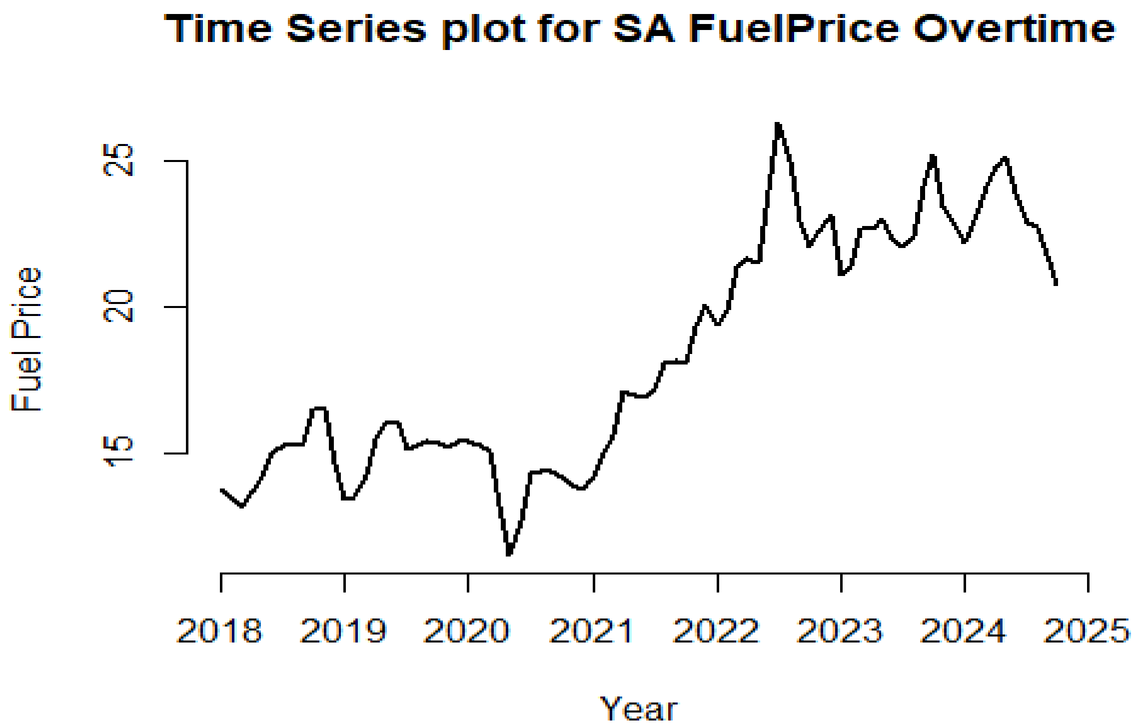
Initials and Surname: MO SEEPAMORE

Student number : 220036766

Data Discussion

Consider the time series plot below.

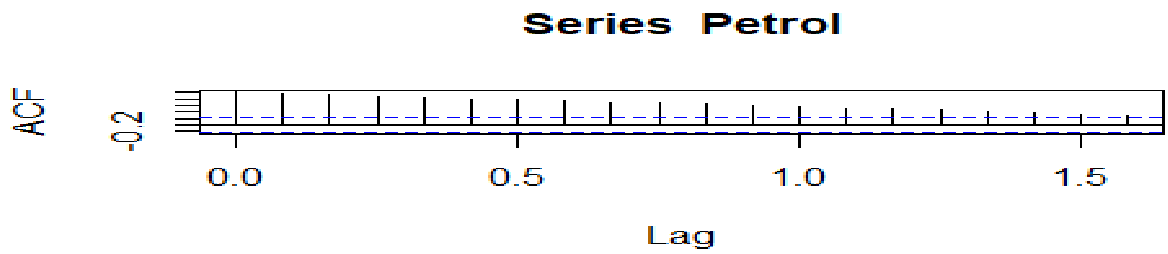
```
plot(Petrol,bty="n",main="Time Series plot for SA FuelPrice Overtime",ylab="Fuel Price",xlab="Year",col="black",lty=1,lwd=2)
```

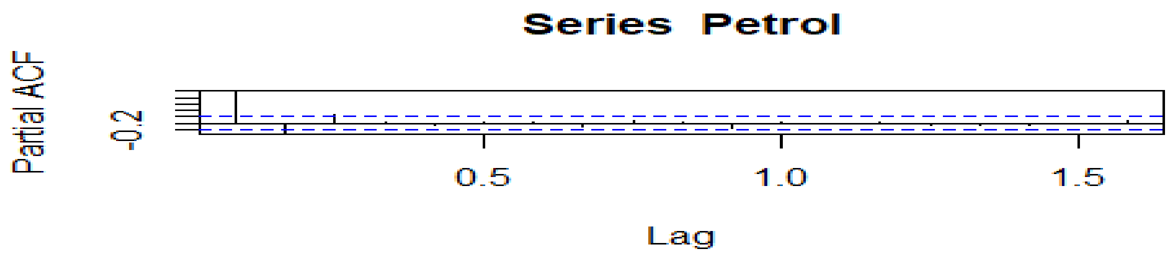


This plot shows the trend of fuel prices from January 2018 until the end of 2024, we see that from 2018 until the middle of 2022 there is an upward trend which indicates that the South African fuel price increases overtime. Then from there until the end of 2024 there is an alternating trend whereby there's a period of decrease in fuel price followed by the period of decrease.

The ACF and PACF Plots

```
par(mfrow=c(2,1))
acf(Petrol)
pacf(Petrol)
```





The ACF plot above shows that as the lag increases there is a decrease in autocorrelation values which indicates that the time series data is non-stationary which might be due to an underlying trend which mean that the mean and variance depends on time. To formalize the stationarity, ADF test will be used. Looking into the ACF and PACF plot they are both tailing off which indicates that ARMA might be appropriate to capture the data. This will be addressed in due course, model identification.

Stationarity test

adf.test(FuelPrice$Price,alternative = "stationary",k=0)

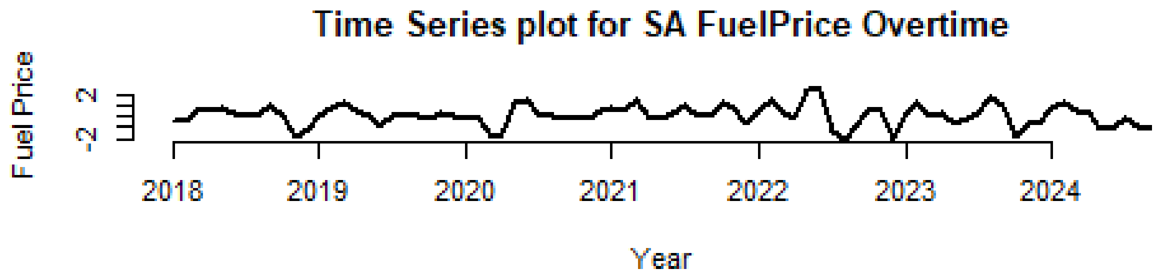
Augmented Dickey-Fuller Test

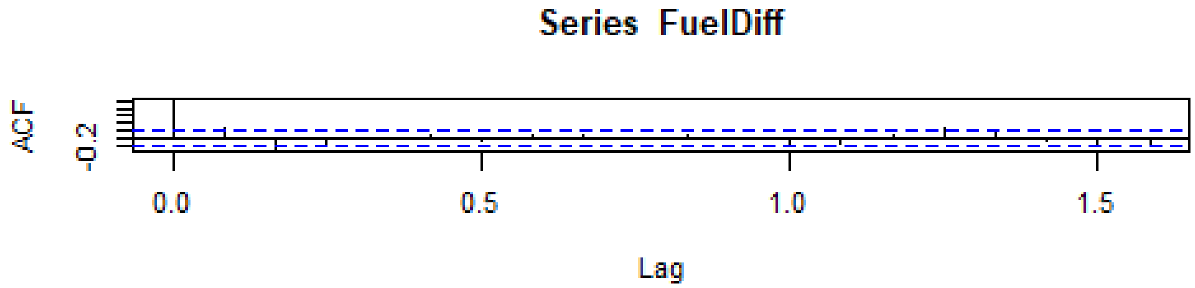
data: FuelPrice$Price
Dickey-Fuller = -1.817, Lag order = 0, p-value = 0.6507
alternative hypothesis: stationary

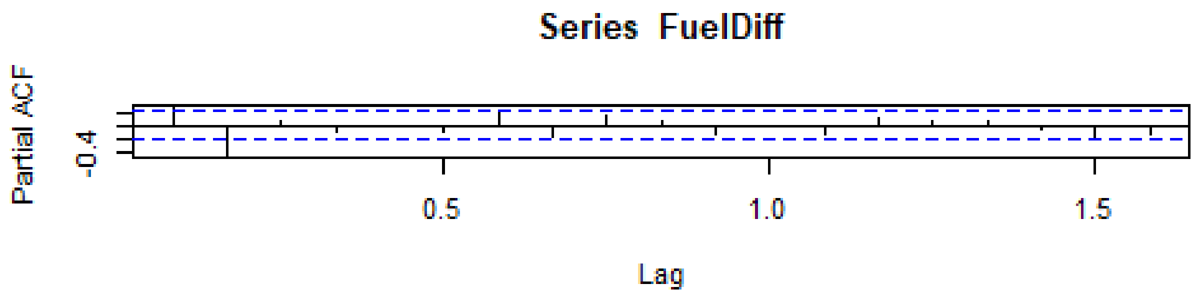
The pvalue suggests that the data series is non-stationary at 1, 5 and 10 percent significance level since well it is sufficiently large, p-value = 0.6507. Since well we have a non-stationarity, ARIMA model cannot be fitted directly. It may be necessary to difference the data to achieve stationarity

Differencing the time series data

```
adf.test(FuelPrice$Price,alternative = "stationary",k=0)
FuelDiff=diff(FuelPrice$Price)
FuelDiff=ts(FuelDiff,start =c(2018,01,03),frequency=12 )

plot(FuelDiff,bty="n",main="Time Series plot for SA FuelPrice Overtime",ylab="Fuel Price",xlab="Year",col="black",lty=1,lwd=2)
```








```
#Checking for Stationarity
> adf.test(FuelDiff,alternative = "stationary",k=0)

        Augmented Dickey-Fuller Test

data:  FuelDiff
Dickey-Fuller = -6.3888, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

INTERPRETATION

From the plots and the ADF test above we can conclude that the differenced series is stationary, a p-value that is less than 0.05 significance level which indicates that the null hypothesis should be rejected that "H0=Series not stationary". Furthermore, looking into ACF and PACF. The ACF is tailing off while PACF plot experiences 0 values at lag greater than 1.4 which indicates that PACF cuts off at lag 2. The proper model to fit is AR (2). To formalize this further, model identification and fitting follows.

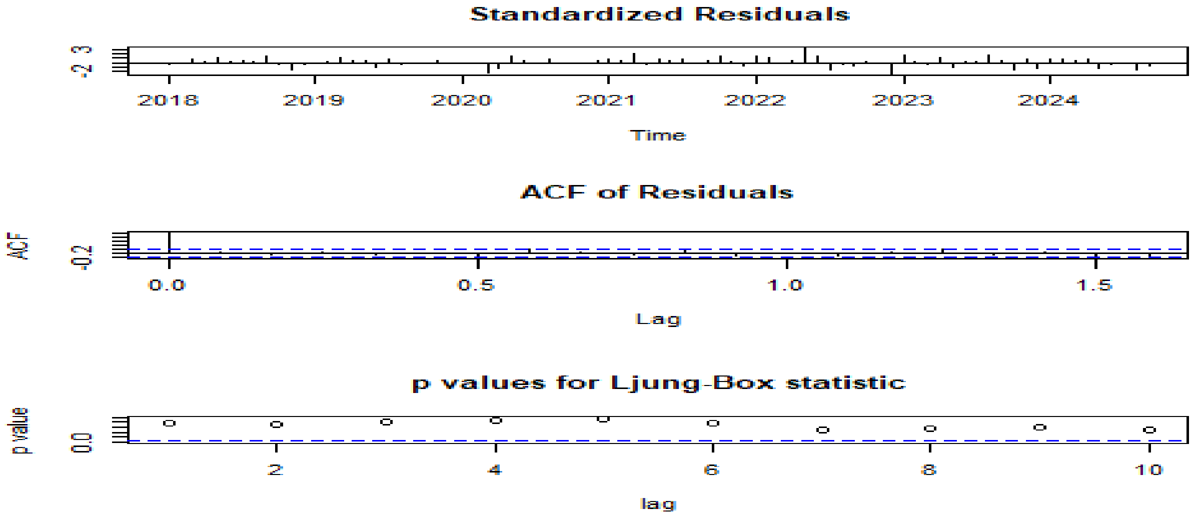
MODEL IDENTIFICATION

```
Fitinitial=auto.arima(FuelPrice$Price,stepwise = FALSE,approximation = FALSE)
Fitinitial

Series: FuelPrice$Price
ARIMA(2,1,0)

Coefficients:
         ar1      ar2
      0.4350  -0.4365
s.e.  0.0998   0.0991

sigma^2 = 0.6645:  log likelihood = -97.63
AIC=201.26   AICc=201.57   BIC=208.44

FitDiff=auto.arima(FuelDiff,stepwise = FALSE,approximation = FALSE)
FitDiff

Series: FuelDiff
ARIMA(2,0,0) with zero mean

Coefficients:
         ar1      ar2
      0.4350  -0.4365
s.e.  0.0998   0.0991

sigma^2 = 0.6645:  log likelihood = -97.63
AIC=201.26   AICc=201.57   BIC=208.44
```

The two models look to be statistically equivalent; they have the same BIC and AIC values. Both models might fit the data well, for simplicity ARIMA (2,0,0) on differenced data would be a valid choice. Thus the fitted model is $x_t = 0.4350\ x_{t-1} - 0.4365\ x_{t-2} + w_t$.

<u>Model Diagnostics</u>

```
tsdiag(FitDiff)
```



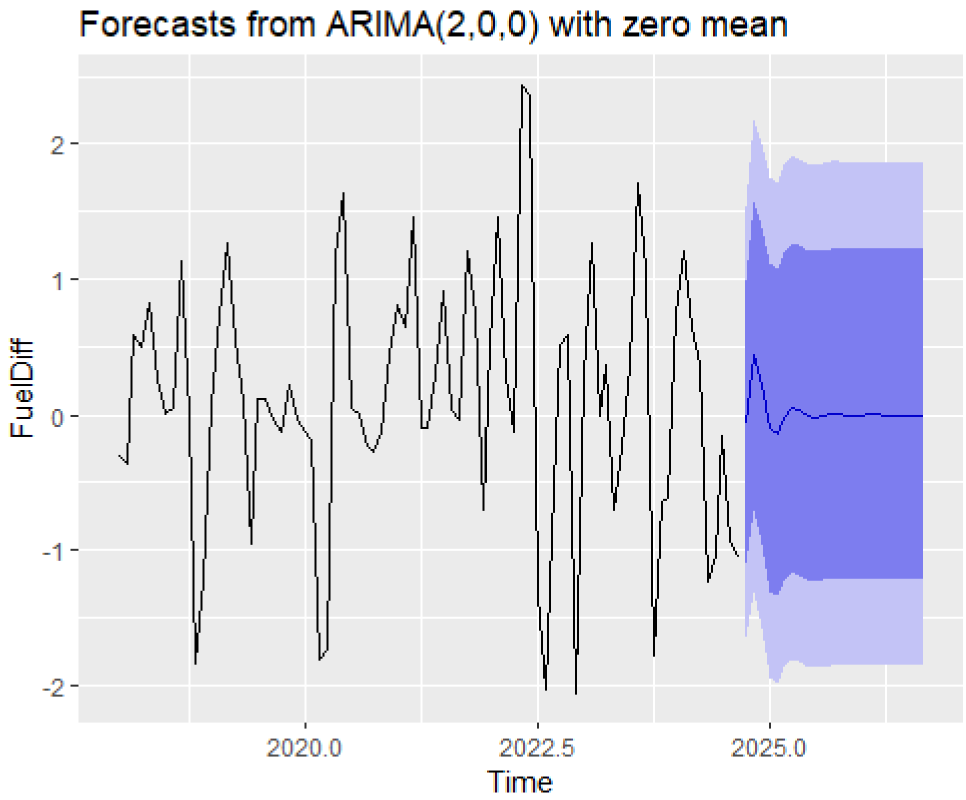
Interpretation

- Standardized residuals look randomly scattered around the horizontal line but around May 2022 there is a possible outlier.
- The ACF of Residuals indicates that there is only one significance at lag 0 and at non-zero lags the autocorrelation is zero which shows that the residuals behaves like white noise.
- Ljung-Box, the p values are mostly above 0.05 suggesting that there is no significance correlation at those lags. Furthermore, the residuals are independently distributed and thus this model fits data well.

Since the first model ARMA needed differencing to reach stationarity assumption, there is a potential flaw in the ARIMA (2,1,0) model

FORECASTING

autoplot(forecast(FitDiff))



The forecasted values suggests that for the next 12 months the fuel prices might capture the same trend as previous years.