ASUTOSH COLLEGE

Estd.1916
92, Shyama Prasad Mukherjee Rd, Kolkata - 700026
(Affiliated o **University of Calcutta**)

DEPARTMENT OF STATISTICS

NAME : RAMIT NANDI

STREAM : B.Sc. HONOURS

COLLEGE ROLL NO. : 130

UNIVERSITY ROLL NO. : 193012-21-0381

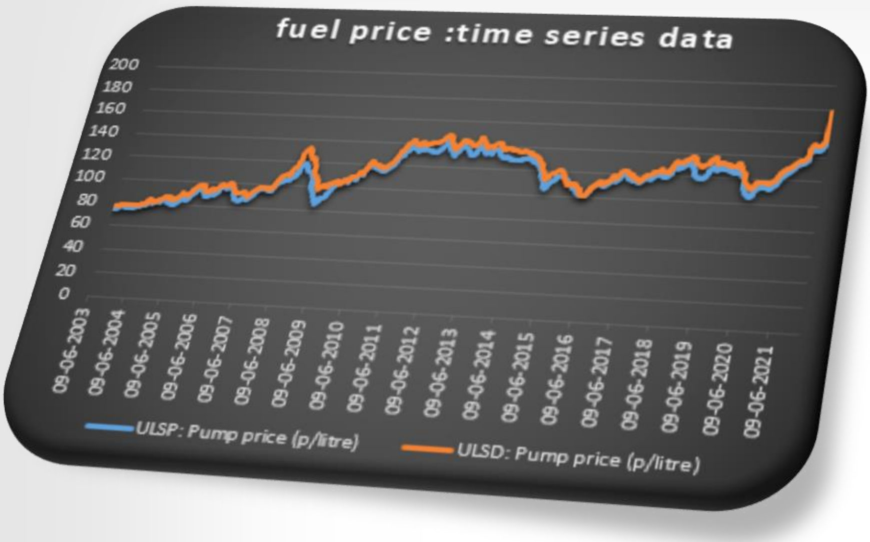
UNIVERSITY REGISTRATION NO. : 012-1111-0558-19

SUBJECT : STATISTICS

PAPER : DSE-B2 (PROJECT WORK)

SEMESTER : VI

YEAR : 2022



PRICE OF PETROL & DIESEL IN UK

- a Time Series analysis

This is a project work on time series ,for UG sem 6 (STS-A-DSE-B-6-2-P),based on BEIS' weekly road fuel price statistics 9/6/2003 – 28/3/2022 where we intend to decompose various components of the data, forecast next few values based on data and compare forecasted values with observed values if available

Date: 14july, 2022

CONTENT

ABSTRACT:

Here in this project, we have collected BEIS' weekly road fuel price statistics providing average UK retail 'pump' prices on a weekly basis. We have this weekly observations for both petrol price and diesel price, for the time period 9june,2003 to 28march,2022. Throughout the project work I intend to ……….

- Obtain the common descriptive measures such as 'mean', 'median', 'quartile', 'range', 'variance' of the data
- Check if any linear or non-linear relationship between petrol and diesel price
- Plot the time series data for easy visualization
- Decompose various components of time series such as—
  - Fitting different 'trend line' to the data by 'Method of Least Square', computing their 'coefficient of determination', constructing corresponding 'ANOVA table'. Comparing such lines and selecting the most appropriate one.
  - Check whether there is 'seasonal' and/or 'cyclical' component. If any then analyse them.
  - Examine the 'irregular component' with the help of 'residual plot'.
- Based on the observed time series data forecasting immediate future values.

INTRODUCTION:

Millions of years ago, algae and plants lived in shallow seas. After dying and sinking to the seafloor, the organic material mixed with other sediments and was buried. Over millions of years under high pressure and high temperature, the remains of these organisms transformed into what we know today as fossil fuels. Coal, natural gas, and petroleum are all fossil fuels that formed under similar conditions. Today, petroleum is found in vast underground reservoirs where ancient seas were located.

Energy is an inseparable component of any economic development. Oil has become the world's most important source of energy since the mid-1950s in such a way that it is named as 'lifeblood of the nations'. In general, low oil prices are considered good for importers of oil because it not only improves consumer spending but also improves the trade balance of a country. Therefore an increase in oil prices has a significant negative impact on the GDP growth in all oil importing countries. On the other side, decrease in Oil price is bad for oil exporters as it could put a depression in revenues of oil exporting countries where oil exports play an enormously important role in supporting economic growth and government finances.

When we think of oil, we tend to think of fuel for our cars, trucks, and planes, and heating oil. However, there are a myriad of uses for crude oil in petro-chemical sector that affects our lives – which means that reducing our dependence on oil may not be as easy as it seems, even after knowing its environmental impacts and non-renewability very well. The European Commission has established a gradual transfer from petroleum-based economy to a more carbohydrate-based economy for 2030, where atleast 20% of transportation fuel should be produced from biomass. The United Kingdom (UK) [the country, not to be confused with Great Britain, its largest island whose name is also loosely applied to the whole country] is a sovereign country in Europe, has the world's sixth-largest economy by nominal gross domestic product (GDP), and the eighth-largest by purchasing power parity (PPP). Having a high-income economy and a very high human development index rating, the United Kingdom ranks 15th in the world for oil consumption, accounting for about 1.6% of the world's total consumption. UK Oil Consumption data is updated yearly, averaging 1,731.920 Barrel/Day from Dec 1965 to 2020. The data reached an all-time high of 2,228.136 Barrel/Day in 1973 and a record low of 1,191.787 Barrel/Day in 2020. The United Kingdom holds 2,754,685,000 barrels of proven oil reserves as of 2016, ranking 30th in the world and accounting for about only 0.2% of the world's total oil reserves of 1,650,585,140,000 barrels. It produces every year an amount equivalent to 14.4% of its total proven reserves (as of 2016), and imports 11% of its oil consumption (177,443 barrels per day in 2016). Through its extensive supply chain, the oil & gas industry employs hundreds of thousands of people and makes a major contribution to the UK economy in terms of tax revenues, technologies and exports. For the oil importer countries, oil price increase and economic growth are negatively correlated while all things being equal.

So, keeping track of oil price in the United Kingdom is essential, what we will try through our project work.

COLLECTED DATA :

BEIS' weekly road fuel price statistics provide average UK retail ('pump') prices on a weekly basis. The data is used to monitor road fuel prices in the UK, and to compare UK road fuel prices with other EU countries. Weekly price data is published on the BEIS website the day after collection.

Here we extract the weekly tabulated data on the time period **9june,2003** to **28march,2022** as follows…

Date	ULSP	ULSD
09-06-2003	74.59	76.77
16-06-2003	74.47	76.69
23-06-2003	74.42	76.62
30-06-2003	74.35	76.51
07-07-2003	74.28	76.46
14-07-2003	74.21	76.41
21-07-2003	75.07	76.9
28-07-2003	75.1	76.86
04-08-2003	75.12	76.81
11-08-2003	75.44	77.08
18-08-2003	75.81	77.44
26-08-2003	76.05	77.68
01-09-2003	76.13	77.59
08-09-2003	76.23	77.67
15-09-2003	76.2	77.63
22-09-2003	76.15	77.55
29-09-2003	76.08	77.53
06-10-2003	76.43	77.66
13-10-2003	75.9	77.43
⋮		
07-02-2022	146.33	150.3
14-02-2022	146.95	151.1
21-02-2022	147.77	151.95
28-02-2022	149.22	153.36
07-03-2022	152.95	158.56
14-03-2022	159.96	169.48
21-03-2022	165.37	177.47
28-03-2022	162.65	176.44

For the data and throughout the project hereafter ……….

- Unit of price = pence/liter
- ULSP =(Ultra low sulphur unleaded) petrol
- ULSD = (Ultra low sulphur unleaded)diesel

We will also denote petrol price and diesel price by 'p' or 'p_t' and 'd' or 'd_t' respectively.

BEIS stands for Department for Business, Energy & Industrial Strategy a department of the government of the United Kingdom .

<remaining entire data set is attached in appendix section ,at the end of this project>

METHODOLOGY:

DESCRIPTIVE MEASURES :

5NUMBER SUMMARY & BOX PLOT |

The five-number summary is a set of descriptive statistics that provides information about a dataset. It consists of the five most important sample percentiles

---- the sample minimum (smallest observation)

---- the lower quartile or Q1 (the value under which 25% of data points are found when arranged in increasing order)

---- the median or Q2 (the middle value)

---- the upper quartile or Q3 (the value under which 75% of data points are found when arranged in increasing order)

---- the sample maximum (largest observation)

The five-number summary gives information about the location [median], spread [midspread or interquartile range IQR : Q3-Q1] ,range [sample maximum – minimum] and skewness [comparing distance between Q1toQ2 & Q2toQ3] of the observations. To readily visualize all such information we draw Boxplot. In addition to the box representing midrange and a midline at median on a box plot, there are lines, called whiskers, extending from the box indicating variability outside the upper and lower quartiles. Conventionally, from above(/below) the upper(/lower) quartile , a distance of 1.5 times the IQR is measured out and a whisker is drawn up(/down) to the largest(/smallest) observed data point from the dataset that falls within this distance. Because the whiskers must end at an observed data point, the whisker lengths can look unequal, even though 1.5 IQR is the same for both sides. All other observed data points outside the boundary of the whiskers are plotted as a dot, a small circle, a star, etc. are called outliers.

It is possible to quickly compare several sets of observations by comparing their five-number summaries and boxplots.

SAMPLE MEAN & VARIANCE |

The sample mean or average is the most commonly used 'measure of location' for a dataset, which is the sum of those values divided by the number of values. Using mathematical notation, if a sample of n observations on variable x is taken, the sample mean is

$$\bar{x} = {\sum_i x_i}/{n}$$

it is easy to interpret but is widely affected in presence of extreme observations.

Sample variance can be defined as the average of the squared difference of data points from the mean of the data set. It is an 'absolute measure of dispersion' and is used to check the deviation of data points with respect to the data's average.The formula is given by

$$s_x^2 = {\sum_i (x_i - \bar{x})^2}/{n-1}$$

. SAMPLE COVARIANCE |

Suppose we have n pairs of observations (x_i , y_i) ;i=1,2,.....n of two variables x , y and we want to know how well or in which direction they vary together. Such a measure is sample covariance , given by

$$cov(x,y) = \sum_i (x_i - \bar{x})(y_i - \bar{y}) \Big/ n-1$$

this measure remains unaffected by origin change, but is not unit free. Note that in particular sample variance of x is $s_x^2 = cov(x,x)$

■ . SAMPLE CORRELATION COEFFICIENT & LINEAR REGRESSION

The degree or extent to which two variables are linearly related or associated is called the correlation between the variables. It is measured by product moment correlation coefficient due to K. Pearson

$$r_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}$$

$$= cov(x,y) \Big/ s_x . s_y \qquad \text{where } |r_{xy}| \le 1$$

Now based on the observations we can fit a straight line such that our model is

$y_i = (a + bx_i) + e_i$;i=1,2,.....n where e_i 's are random errors satisfying assumption $E(e_i) = 0$ & $cov(e_i, e_j) = \begin{cases} \sigma^2, & i = j \\ 0, & i \ne j \end{cases}$ .The unknown parameters a, b are estimated in such a way that error sum of squeares $L = \sum_i {e_i}^2$ is minimized. The normal equations are given by

$$\left.\begin{aligned} \partial L / \partial a &= 0 \\ \partial L / \partial b &= 0 \end{aligned}\right\} \text{ i.e. } \qquad \begin{aligned} \sum_i y_i &= na + b \sum_i x_i \\ \sum_i y_i\, x_i &= a \sum_i x_i + b \sum_i {x_i}^2 \end{aligned}$$

............solving the equations we get

$$\begin{cases} \hat{b} = r_{xy} \frac{s_y}{s_x} \\ \hat{a} = \bar{y} - \hat{b}\bar{x} \end{cases}$$

and MSE = $\sum_i (y_i - \hat{a} - \hat{b}x_i)^2 \Big/ n-2$ gives an estimate of σ^2

Note that $r_{xy} = \begin{cases} +1 & : when\ all\ (x_i , y_i)\ exactly\ lies\ on\ a\ straight\ line\ with + ve\ slope \\ 0 & : when\ x_i , y_i\ 's are\ uncorrelated, i.e\ there\ is\ no\ linear\ relationship \\ -1 & : when\ all\ (x_i , y_i)\ exactly\ lies\ on\ a\ straight\ line\ with - ve\ slope \end{cases}$

in general the quantity $\boldsymbol{r_{xy}}^2$ is termed as 'coefficient of determination' ,which measures how much proportion of the total variability of y is explained by the least square linear regression equation $y_x = (\hat{a} + \hat{b}x)$,means how well the straight line is fitted to the data.

TIME SERIES :

Time series is set of data collected and arranged in accordance of time. A time series is the result of a number of movements which are caused by numerous economic, political, natural and other factors. The analysis of time series means decomposing the past data into components and then projecting them forward. A time series typically has four components, though on occasions only one or two of these may eclipse the others….

- **TREND :-** Over a long period, a time series will have an overall tendency either to move upwards or downwards or sometimes a blend thereof, though the actual movement will not be regular. This general, smooth, long term ,average tendency of the time series data is known as the 'secular trend'. For example, at a glance the sales of a popular soft drink manufacturer is likely to reveal an increasing trend.

It is to be noted that, the term 'long period' has no exact definition but relative to the context of discussion. Trend is the easiest component in a time series data to capture and then the most well discussed one.

- **<u>SEASONAL VARIATION</u> :-** The fluctuations in a time series due to the rhythmic forces which operate in a regular ,periodic manner over a span of less than a year and have the same or almost same pattern year after year are called seasonal variations. For example, for a soft drink manufacturer, while the yearly sales may be on the increase, the sales are likely to be high every summer and low every winter. Clearly, any time series data recorded annually can not have seasonality.

If seasonality is present in a time series data it can be extracted easily as ratio or residual of observed values to trend values.

- **<u>CYCLICAL MOVEMENTS</u> :-** These are oscillatory movements—a series of repeated sequences---superimposed on the original data. These movements are of longer duration than a year. The sales of a company, for example, may be high because the level of economic activity may be high. Similarly, the sales may be low due to overall subdued economic activity.

Although these variations may be recurrent, these are seldom found to be of similar pattern having same period and amplitude of oscillations. Also all cyclic components whose wavelength exceeds the length of the observed data is have to be accumulated in Trend component. A satisfactory method for the direct measurement of cyclical variations is not available. In most of the cases these variations are so intermixed with random variations that it is very difficult, if not impossible, to separate them.

- **<u>IRREGULAR COMPONENT</u> :-** Apart from the three systematic components, almost all the time series contain another factor called the 'random' or 'irregular' or 'residual fluctuations' .These fluctuations are purely random ,erratic ,unforeseen, unpredictable and are due to numerous non-recurring and irregular circumstances which are beyond the control of human-hand but obviously a part of our system ,such as earthquakes, wars, floods, strike etc. Irregular component may or may not be significant. As the name suggests, it is not possible to isolate irregular component and study them exclusively, nor one can forecast or estimate them precisely.

- Forecasting time series data implies that predictions about the future values are made only from the past data, and other factors, no matter how important they might potentially be, are not considered.

<u>POLYNOMIAL CURVE FITTING:-</u>

As trend line is a smooth curve , an obvious choice is fitting a polynomial curve(since polynomial functions are continuous and differentiable everywhere).

Here we assume that our suitable trend equation is a polynomial of degree 'p' in time element $'t'$. That is our mathematical model is

$$\boldsymbol{y_t} = (\boldsymbol{a_0} + \boldsymbol{a_1 t} + \boldsymbol{a_2 t^2} + \cdots + \boldsymbol{a_p t^p}) + \boldsymbol{e_t}$$

so here are p covariates t, t^2,...., t^p; response variable y_t ;$(p+1)$ parameters $a_0, a_1, a_2, \dots, a_p$.

e_t 's are random errors satisfying assumption $E(e_t) = 0$ & $cov\left(e_t, e_{t'}\right) = \begin{cases} \sigma^2, & t = t' \\ 0, & t \neq t' \end{cases}$. let there are n observations of the pair $(t\ , y_t)$. The parameters are estimated from the data by 'method of least square' ,which consists of minimizing $L = \sum_t {e_t}^2$

$$= \sum_t (y_t - a_0 + a_1 t + a_2 t^2 + \cdots + a_p t^p)^2$$

the *(p+1)* normal equations are given by $\partial L / \partial a_i = 0 \ \ \forall\, i = 0,1,2, \dots, p$

$$i.e. \begin{cases} \sum_t y_t = n a_0 + a_1 \sum_t t + a_2 \sum_t t^2 + \cdots + a_p \sum_t t^p \\ \sum_t t y_t = a_0 \sum_t t + a_1 \sum_t t^2 + a_2 \sum_t t^3 + \cdots + a_p \sum_t t^{p+1} \\ \vdots \\ \sum_t t^p y_t = a_0 \sum_t t^{p+1} + a_1 \sum_t t^{p+2} + a_2 \sum_t t^{p+3} + \cdots + a_p \sum_t t^{2p} \end{cases}$$

solving these equations we get estimated value of $a_0, a_1, a_2, \dots, a_p$. Then fitted value at time t is

$$\widehat{y_t} = \widehat{a_0} + \widehat{a_1}\, t + \widehat{a_2}\, t^2 + \cdots + \widehat{a_p}\, t^p$$

EXAMPLE : when p=2 , and origin of t is so chosen that $t = 0, \pm 1, \pm 2, \dots, \pm k$; $n = 2k + 1$ we have

$$\widehat{a_2} = {n \sum_t y_t t^2 - \sum_t y_t \sum_t t^2} \Big/ {n \sum_t t^4 - (\sum_t t^2)^2} \ ; \quad \widehat{a_1} = {\sum_t y_t t} \Big/ {\sum_t t^2} \quad ; \quad \widehat{a_0} = {\sum_t y_t - \widehat{a_2} \sum_t t^2} \Big/ n$$

- **MERITS :**
  i. It is possible to compute trend values for all the periods and predict the value for a period lying outside the observed data.
  ii. The functional form is easy to comprehend for layman. Also using suitable transformation, various other kind of equation can be fitted by this method.
  iii. The results of the method of least squares are most satisfactory because the fitted trend satisfies the two most important properties,
  ----- $\sum_t (y_t - \hat{y}_t) = 0$: the position of fitted trend equation is such that the sum of deviations of observations above and below this equal to zero.
  ----- $\sum_t (y_t - \hat{y}_t)^2 = 0$: the sums of squares of deviations of observations, about the trend equations, are minimum among all polynomials of same degree

- **DEMERITS :**
  i. It is not flexible like if some observations are added, then the entire calculations are to be done once again.
  ii. The computation of trend values, on the basis of this method, doesn't take into account the other components of a time series and hence forecasting may not always be reliable except immediate future.
  iii. Since the choice of a particular trend is arbitrary, the method is not, strictly, objective.

■ TESTING OF SIGNIFICANCE:-

Based on our model, mentioned earlier– $y_t = (a_0 + a_1 t + a_2 t^2 + \cdots + a_p t^p) + e_t$

we want to test whether the covariate has any significant effect on response variable or not. So,

our testing problem will be, $\mathbf{H_0}$: $a_i = 0 \ , \forall\, i = 1,2,\dots,p$ vs $\mathbf{H_1}$: $a_i \neq 0 \ , \exists\, atleast\ one\ i\ \ in\ 1,2,\dots,p$
Now, consider the following table

Source of variation	sum of squares Or SS	d.f.	mean squares Or MS	F statistic
regression	$SS_{reg} = \sum_t (\widehat{a_0} + \widehat{a_1}\, t + \widehat{a_2}\, t^2 + \cdots + \widehat{a_p}\, t^p - \bar{y}_t\,)^2$	p	$MS_{reg} = {SS_{reg}}/{p}$	$F = \frac{MS_{reg}}{MSE}$
error	$SSE = \sum_t (\,y_t - \widehat{a_0} - \widehat{a_1}\, t - \widehat{a_2}\, t^2 - \cdots - \widehat{a_p}\, t^p\,)^2$	n-p-1	$MSE = {SSE}/{n-p-1}$	
total	$SS_{total} = \sum_t (\,y_t - \bar{y}_t\,)^2$	n-1	----	

where $\boldsymbol{SS_{total}} = \boldsymbol{SS_{reg}} + \boldsymbol{SSE}$ is an orthogonal splitting ,i.e SS_{reg} and SSE independent.

Under H_0 : $a_i = 0 \ , \forall\, i = 1,2,\dots,p$; the model reduces to $y_t = a_0 + e_t$

$\implies \widehat{a_0} = \bar{y}_t$ and so $SSE_{H_0} = SS_{total}$

Hence a large positive value of $\left(SSE_{H_0} - SSE\right) = SS_{reg}$ indicates deviation from H_0. It can be shown that **F**= ${MS_{reg}}/{MSE}$ $\sim$ F distribution with df (p , n-p-1) . We reject H_0 in favour of H_1 at level α iff $\mathrm{F}_{observed} > F_{\alpha;p,n-p-1}$ [upper α point of F distribution with df (p , n-p-1)]

- **R^2 :-**

The ratio of 'sum square due to regression' to the 'total variation' is an an obvious measure of how well our model is fitted to the data. Then the most general definition of the 'coefficient of determination' is

$$\mathbf{R^2} = {SS_{reg}}/{SS_{total}} \quad = 1 - {SSE}/{SS_{total}}$$

taking value between 0 and 1. More the value of R^2 close to1 indicates a better fit of model.

However each time we add a new covariate to the model the R- squared is guaranteed to increase, with a decrease in error df, even if the predictor variable isn't useful. So, penalizing with error df , we define our modified measure

$$\textbf{adjusted } \mathbf{R^2} = 1 - {MSE}/{MS_{total}} = 1 - (1 - R^2)\,\frac{n-1}{n-p-1}$$

it may be possible that in some case by adding new covariate R^2 is increased but adjusted R^2 is decreased. Same as R^2, the value of adjusted R^2 lies between 0 and 1. More the value of adjusted R^2 close to 1 indicates that all the predictor variables in model have better significant effects.

- *It is to be mentioned that , in time-series analysis when fitting trend line, our goal is not to search for a curve that almost exactly passes through all data points (over-fitting) , but to capture the overall long-term tendency of the time series .*

■ FORECAST ACCURACY :-

It is to be noted that ,size of residuals are not a reliable indication of how large the forecasts error are likely to be, it can only be determined by considering how well a model performs on new data that were not used while fitting. If our training data is given by { $y_1, y_2, \dots, y_T$ } and test data is given by { $y_{T+1}, y_{T+2}, \dots$ } , then 'forecast error' is defined as their difference

$$e_T = y_{T+h} - \hat{y}_{T+h|T}$$

Some useful measures of forecast accuracy are given by

- **Root Mean Squared Error** : $RMSE = \sqrt{mean(e_t^2)}$ ---gives an idea about average distance between observed and forecasted value.
- **Mean absolute percentage error** : $MAPE = mean(\left|\frac{e_t}{y_t} \times 100\right|)$ ---- which measures the average magnitude of proportions of forecast error to observed value.

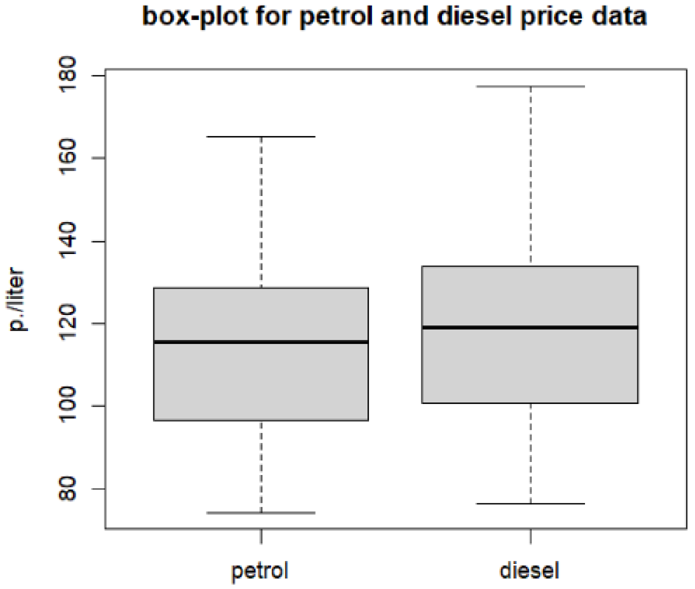
higher these measures ,better the accuracy

ANALYSIS:

DESCRIPTIVE ANALYSIS :

After getting the data , the first thing we intend to study is the ‘5point’ summary of the data,and to visualize easily we add corresponding box-plots.

```
> summary(pricedata[,-1])
 ULSP..Pump.price..p.litre. ULSD..Pump.price..p.litre.
 Min.   : 74.21             Min.   : 76.41
 1st Qu.: 96.57             1st Qu.:100.93
 Median :115.56             Median :119.14
 Mean   :112.71             Mean   :117.22
 3rd Qu.:128.59             3rd Qu.:133.71
 Max.   :165.37             Max.   :177.47
```



the number of observations for both petrol and diesel price is

```
> n
[1] 982
```

.There is no outliar.

for diesel average value of data is slightly larger than petrol ,

in both case mean<median , also 1st quartile is little more apart from median comapared to median & 3rd quartile –-implies the data has slightly -ve skewness for both petrol and diesel.

Using sample variance as measure of dispersion for the data

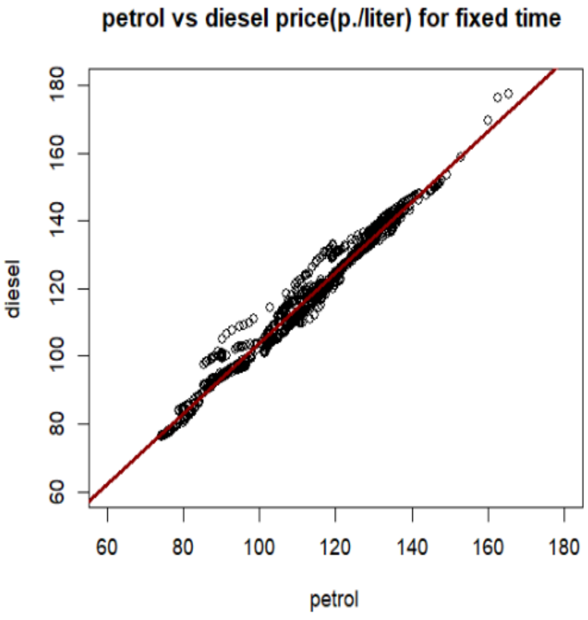
```
> var(p)
[1] 355.0034
```

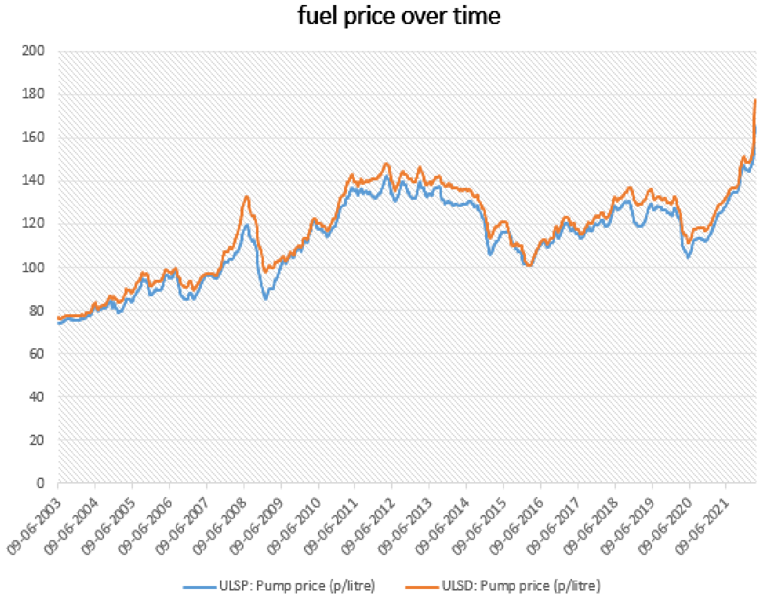
```
> var(d)
[1] 393.7688
```

and

```
> cor(p,d)
[1] 0.9904448
```

see that petrol and diesel price is very highly +ve correlated. If we plot the pair (petrol price, diesel price) for various week , we get an almost straight line. Plotting the prices over time ,we also notice that petrol price and diesel price data looks very similar in nature.

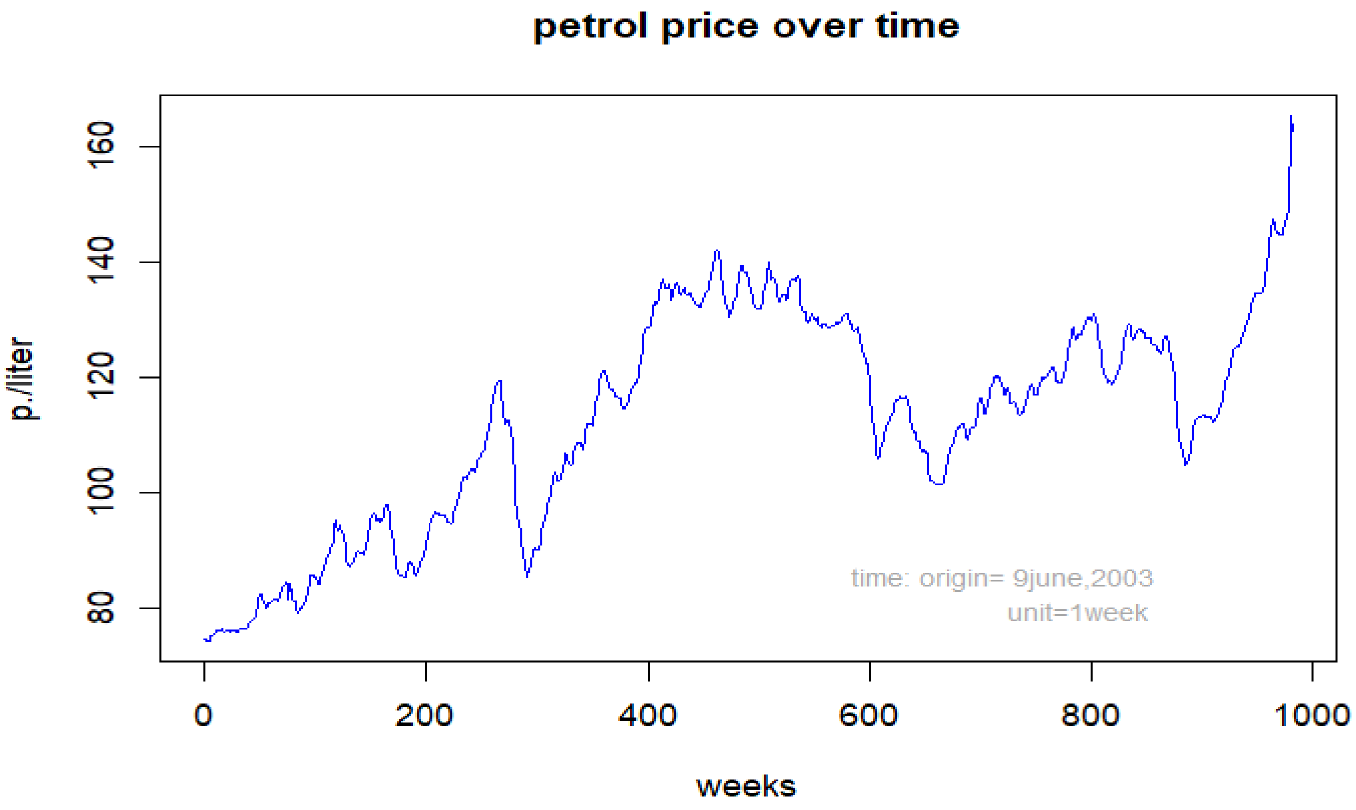




➢ *So from now on, we first obtain time series analysis and forecasting of one data , without loss of generality ,say petrol price data, and later regress the other i.e diesel price based on petrol price*

TIME SERIES ANALYSIS :

As the first step of time series analysis , plot the weekly petrol price data over time, taking 9june 2003 as origin and unit as 1week



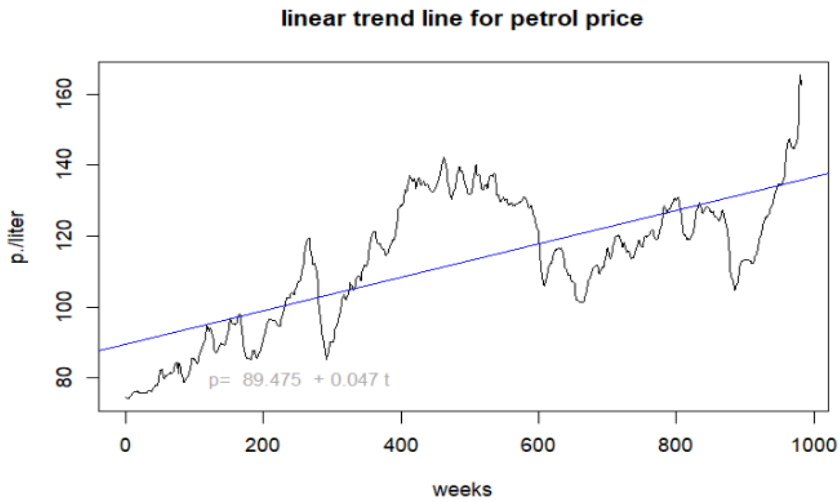
- **TREND :** Lets start with linear trend, the best fitted straight line for this data is given by

```
> lm(p~t)

Call:
lm(formula = p ~ t)

Coefficients:
(Intercept)            t
   89.47510      0.04736
```

```
Residual standard error: 13.22 on 980 degrees of freedom
Multiple R-squared:  0.5083,    Adjusted R-squared:  0.5078
F-statistic:  1013 on 1 and 980 DF,  p-value: < 2.2e-16
```

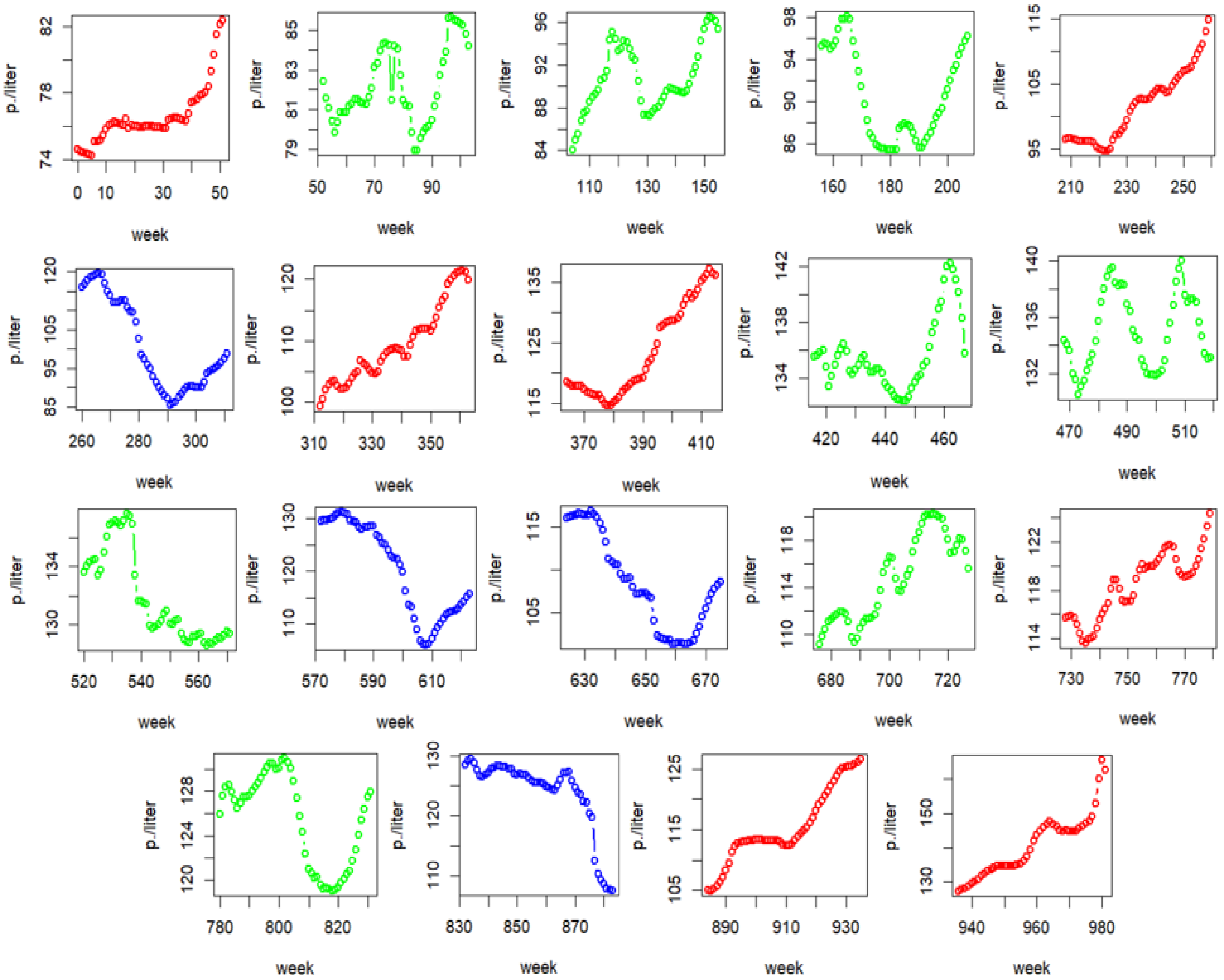


Here ANOVA suggests that corresponding F statistic is significant , which simply means that we have enough evidence to reject the null hypothesis “there is no trend” in favour of alternate hypothesis “there is linear trend” .But from the value of R^2 it is evident that only 50.83% of total variation of the data is explained by linear trend equation ,which is not so satisfying. We can improve our trend equation, but for time being let skip it, only gaining knowledge from the slope of the linear trend that there is an overall increasing trend in the data.

- **SEASONALITY :** If there is seasonality in a time series data, then plot of the data from one year should look alike plot from another year.

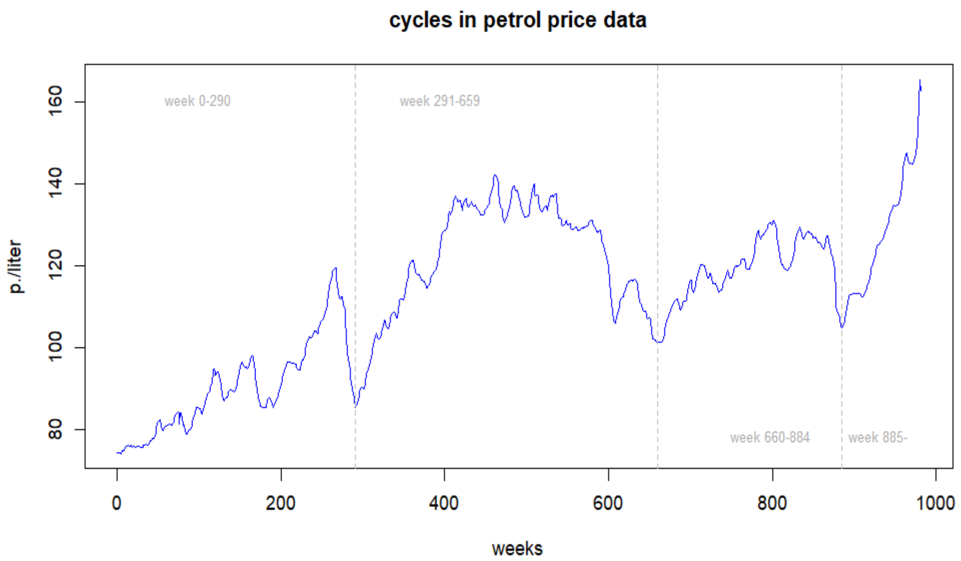
Here after plotting the data for every 52weeks(1year) ,we see increasing pattern in some case as well as decreasing or fluctuating pattern.

seasonal plot



So it is appealing that our data have no seasonality.

- **CYCLICAL COMPONENT :** From the time-plot we notice that the data have three complete cycle of unequal length and one incomplete cycle at the end.



- *we can bypass the hardship of separating 'trend' and 'cyclic' component by capturing so called 'trend-cycle' [The trend-cycle is the component that represents variations of low frequency in a time series, the high frequency fluctuations having been filtered out],and dumping everything others in residual part .*

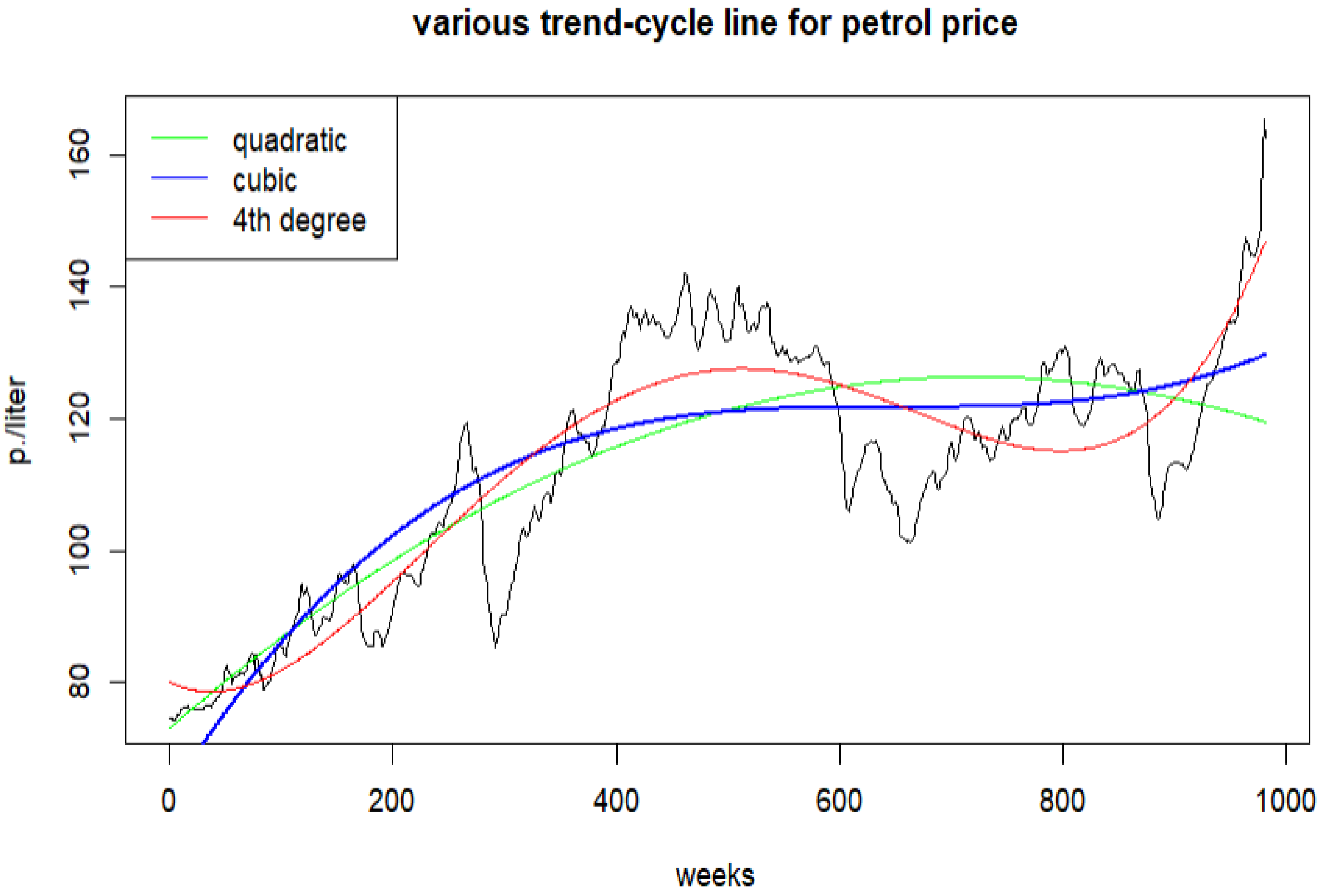
○ TREND-CYCLE DECOMPOSITION:

In the absence of seasonality here, trend-cycle is nothing but the entire systematic part of the time series. The rises and falls in the time-plot suggests that this systematic part can be modelled by either a polynomial curve or by a series of sums of sine and cosine curve (harmonic analysis). For sake of computational simplicity we stick to polynomial model.

Let fit polynomials with degree p= 2,3,…. to the petrol price data by 'method of least square and compare them (remember straight line already have found inefficient)………

Trend-cycle eqn.	Intercept	coeff. Of t	t^2	t^3	t^4	R^2	ANOVA
quadratic	73.0826611	0.1477221	-0.0001023			0.6607	Residual standard error: 10.99 on 979 degrees of freedom Multiple R-squared: 0.6607, Adjusted R-squared: 0.66 F-statistic: 953.3 on 2 and 979 DF, p-value: < 2.2e-16
cubic	6.29E+01	2.73E-01	-4.21E-04	2.16E-07		0.703	Residual standard error: 10.28 on 978 degrees of freedom Multiple R-squared: 0.703, Adjusted R-squared: 0.7021 F-statistic: 771.8 on 3 and 978 DF, p-value: < 2.2e-16
4th degree	8.00E+01	-7.85E-02	1.19E-03	-2.34E-06	1.30E-09	0.797	Residual standard error: 8.506 on 977 degrees of freedom Multiple R-squared: 0.797, Adjusted R-squared: 0.7962 F-statistic: 959 on 4 and 977 DF, p-value: < 2.2e-16

As the degree of the polynomial , the residual variability decreases, R^2 value increases that means the polynomial better fits the data. Quadratic model with 66% explained variability is not a good one. But note that , while the cubic model averages out rises and falls of the cycle , the 4th degree model try to mimic them. So to avoid 'over-fitting' we choose the cubic equation as our suitable trend-cycle equation, though it has smaller R^2 value than the 4th degree one.

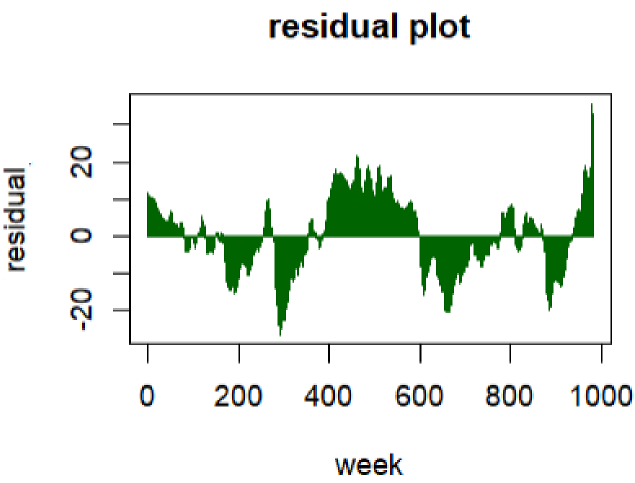


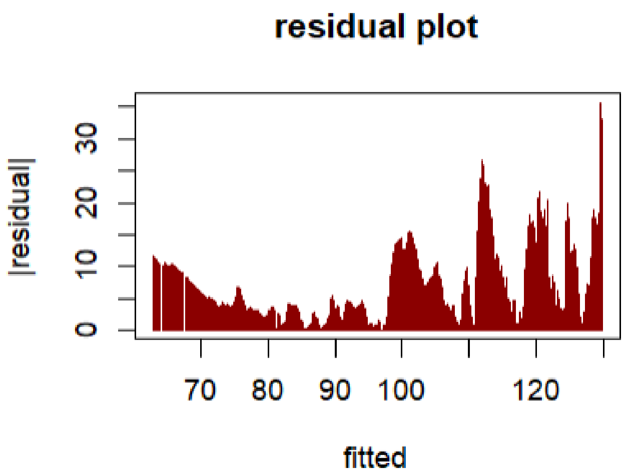
For time t ,the estimated trend-cycle component of petrol price is thus given by

$\widehat{\boldsymbol{\tau}}_{\boldsymbol{t}} = \mathbf{6.29 * 10 + 2.73 * 10^{-1}t - 4.21 * 10^{-4}} \boldsymbol{t}^{\mathbf{2}} + \mathbf{2.16 * 10^{-7}} \boldsymbol{t}^{\mathbf{3}}$,which accounts for explaining 70.3% variation in petrol price data.

- RESIDUALS : Here the magnitude of deviation of observed & fitted values is not proportional to fitted values , so we can use additive residuals ,i.e. for time t observed value of petrol price, $p_t = \widehat{\tau_t} + z_t$; where z_t 's are residuals.

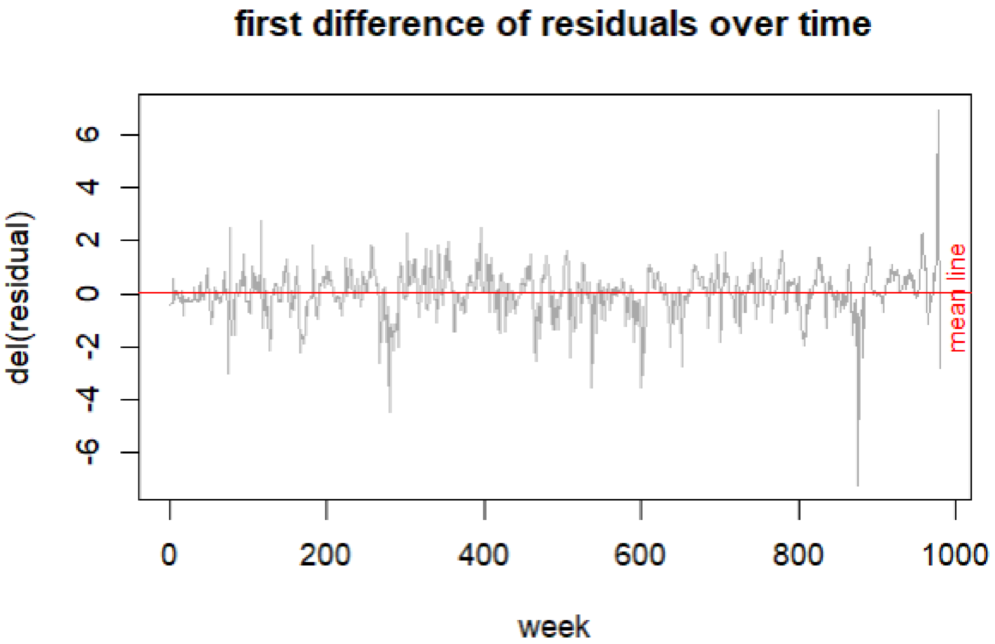
Ploting the residuals over time ,we see that the pattern is not entirely random. Residuals





of same sign have a tendency to occur side by side,i.e residuals are supposed to be highly autocorrelated for initial few lags . The ‘fitted vs |residual| plot’ justifies the additive model.

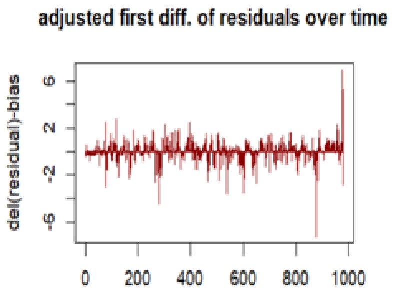
But if we consider the 1st order differences $\Delta z_t = z_{t+1} - z_t$ and plot them over t, we get a more or less random pattern.

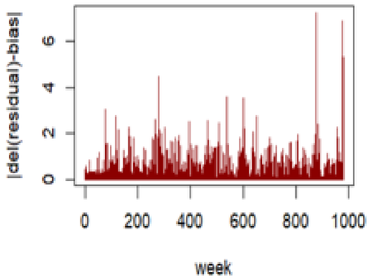


the average value of observed Δz_t's is

```
> mean(diff(z))
[1] 0.02163614
```

, is an estimate of Expectation(Δz_t).





So, let us assume ,

Δz_t = $a\ random\ component\ with\ expectation\ E(\Delta z_t)$

Or,

$(z_{t+1} - z_t)$ = $bias +$
$some\ random\ component\ with\ expectation\ 0$

we can thus predict the non random part of the residual as

$$\widehat{\boldsymbol{z_{t+1}}} = \mathbf{0.0216} + \boldsymbol{z_t},$$

provided $z_t = p_t - \hat{\tau}_t$ is known.

MODELLING PETROL PRICE :

Based on our analysis , we have obtained the following model:

$$\underset{price\ of\ petrol\ at\ time\ t}{\boldsymbol{p_t}} = \underset{trend\ cycle}{\boldsymbol{\tau_t}} + \underset{residual\ :\ a\ non\ random\ part\ +\ random\ error}{\boldsymbol{z_t}}$$

Hence ,

we can forecast the petrol price at time t

as $\widehat{p_t} = \widehat{\tau_t} + \widehat{z_t}$

where the equation is explicitly given by

$\widehat{\tau_t} = 6.29 * 10 + 2.73 * 10^{-1}\mathrm{t} - 4.21 * 10^{-4}t^2 + 2.16 * 10^{-7}t^3$

$\widehat{z_t} = 0.0216 + z_{t-1}$; if z_{t-1} unknown, we can use $\widehat{z_t} = 2 * 0.0216 + z_{t-2}$

Or so on.. with some reduced accuracy.

PETROL vs DIESEL PRICE :

As we have seen earlier, for observations on same time point (petrol price , diesel price) is extremely highly correlated.

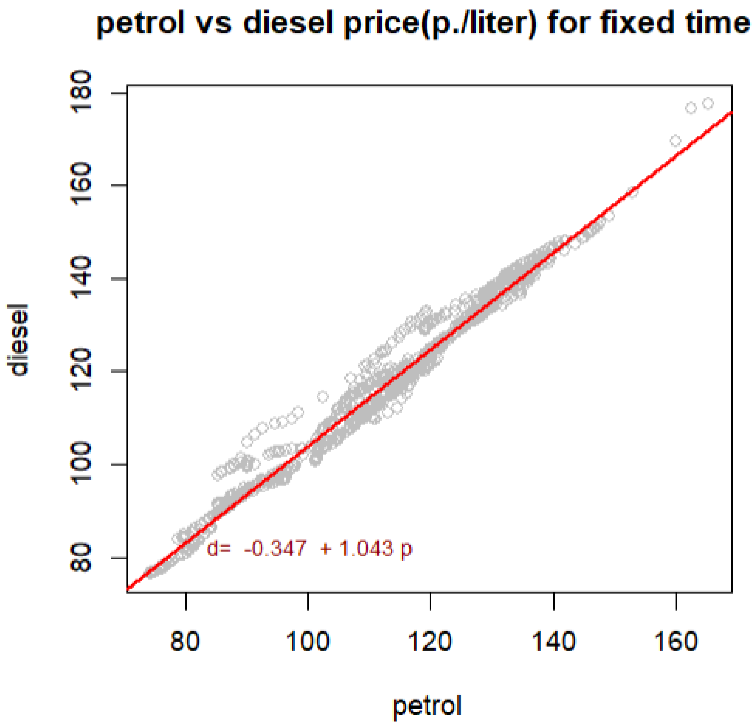
```
> cor(p,d)
[1] 0.9904448
```

So , we can predict diesel price based on petrol price using 'least square linear regression'. The best fitted linear regression equation is given by

```
> lm(d~p)

Call:
lm(formula = d ~ p)

Coefficients:
(Intercept)            p
    -0.3467       1.0431
```



i.e. diesel price at time t,

$$d_t = -0.3467 + 1.0431 p_t$$

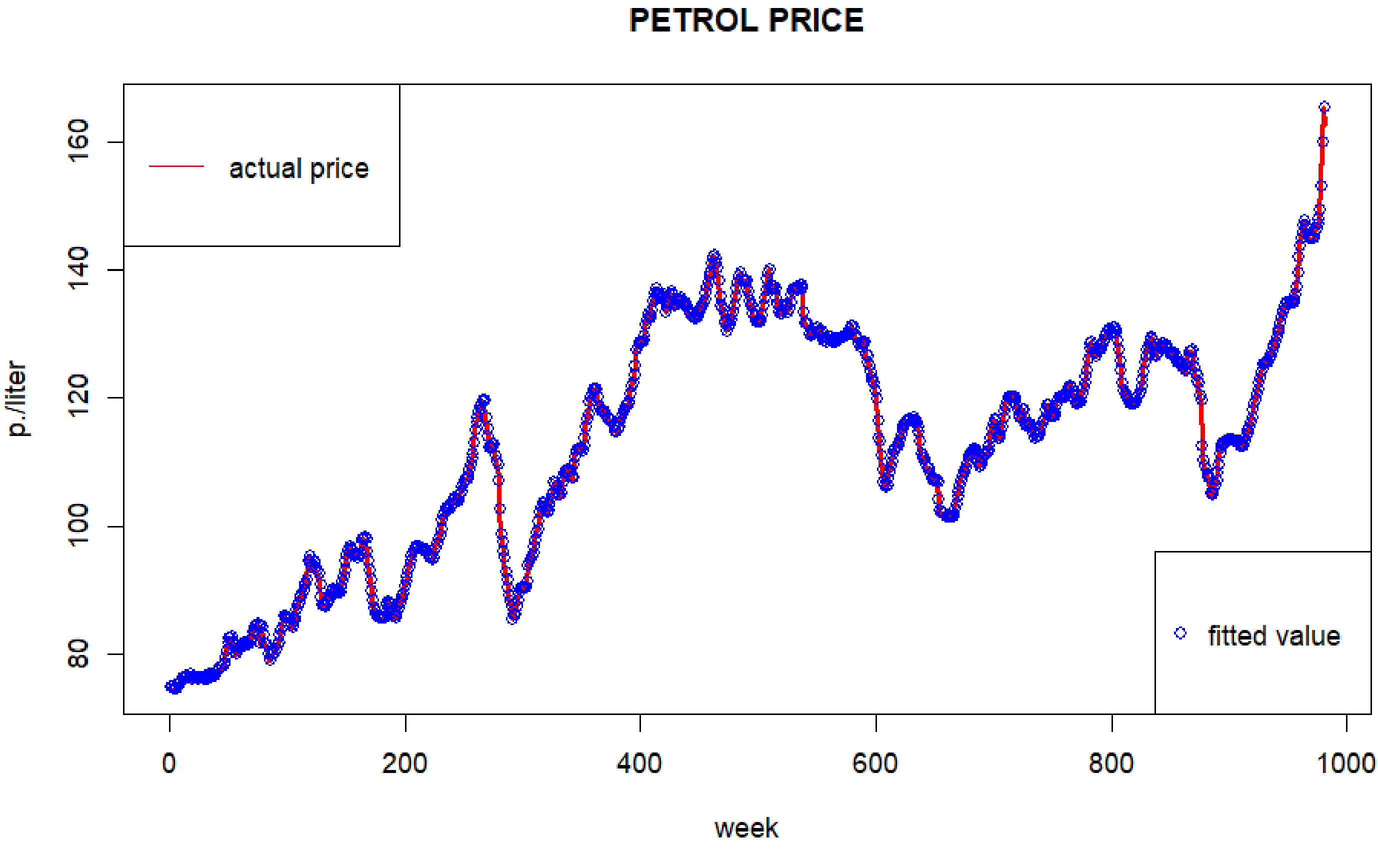
```
Residual standard error: 2.738 on 980 degrees of freedom
Multiple R-squared:  0.981,	Adjusted R-squared:  0.981
F-statistic: 5.055e+04 on 1 and 980 DF,  p-value: < 2.2e-16
```

such regression equation can explain 98.1% variation of the response variable diesel price data.
Thus, we can model the diesel price at time t

as $\widehat{d_t} = -0.3467 + 1.0431\widehat{p_t}$

✓ comparing observed & fitted values:

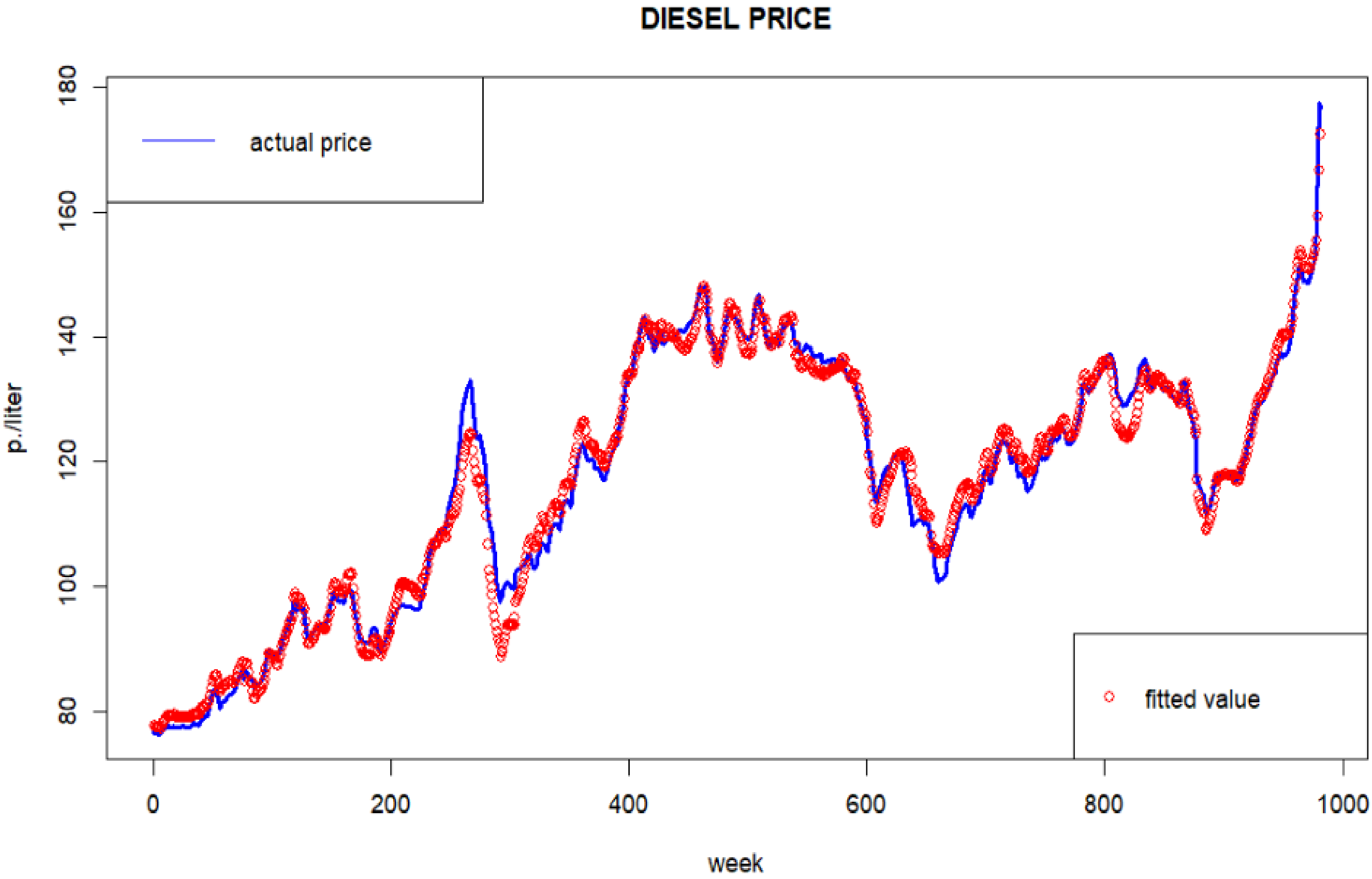
For petrol, we can see that fitted values almost overlap the observed values throughout



the time.The error variance for the model $\widehat{p_t} = \widehat{\tau_t} + \widehat{z_t}$ is 0.799 unit^2 . The magnitude of error is on average 0.54% of the actual petrol price data. So, our derived model fits very well for analysing petrol price data.

```
> summary(p[-1]-estm_p)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
-7.24547 -0.31837  0.02939  0.00000  0.48735  6.91751 
> var(p[-1]-estm_p)
[1] 0.7992405
```

```
> 100*mean(abs(p[-1]-estm_p)/p[-1])
[1] 0.5361433
```



But for diesel price ,we have more error variance as compared to petrol, while petrol and diesel price variation almost same. A model generating error 1.81% of the actual diesel price on average, is not a bad one though.

```
> summary(d[-1]-estm_d)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
-8.146695 -1.762154 -0.494200  0.000703  1.156723 10.862132 
> var(d[-1]-estm_d)
[1] 7.683009
```

```
> 100*mean(abs(d[-1]-estm_d)/d[-1])
[1] 1.81379
```

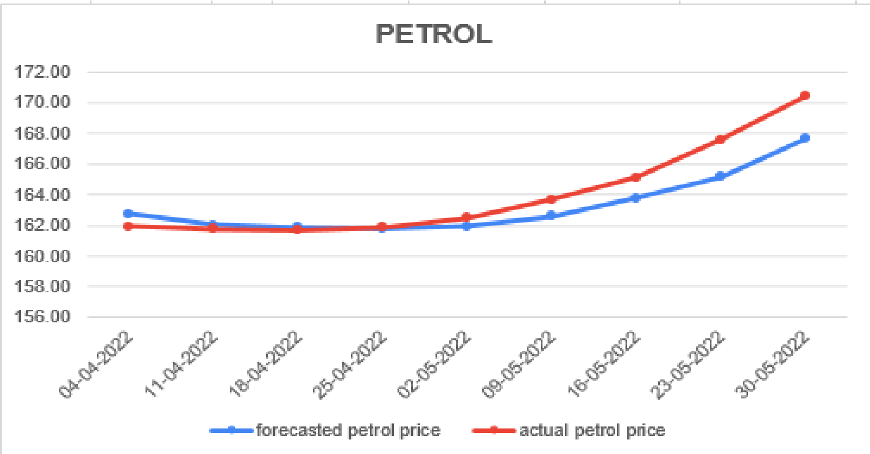
✓cross-checking:

"A model that fits well, not necessarily forecasts well."

We started our project work based on data till 28march,22. But in the meantime of our work being done , few more weekly observations are updated in the original website. So we can now compare these observed values with the predicted values from our model for these weeks……

date	forecasted petrol price	actual petrol price	forecasted diesel price	actual diesel price
04-04-2022	162.74	161.91	169.41	176.00
11-04-2022	162.00	161.78	168.64	176.22
18-04-2022	161.87	161.67	168.51	175.93
25-04-2022	161.77	161.84	168.39	176.33
02-05-2022	161.94	162.48	168.57	177.06
09-05-2022	162.58	163.68	169.24	178.39
16-05-2022	163.78	165.09	170.49	179.67
23-05-2022	165.19	167.59	171.96	181.16
30-05-2022	167.69	170.44	174.57	182.37

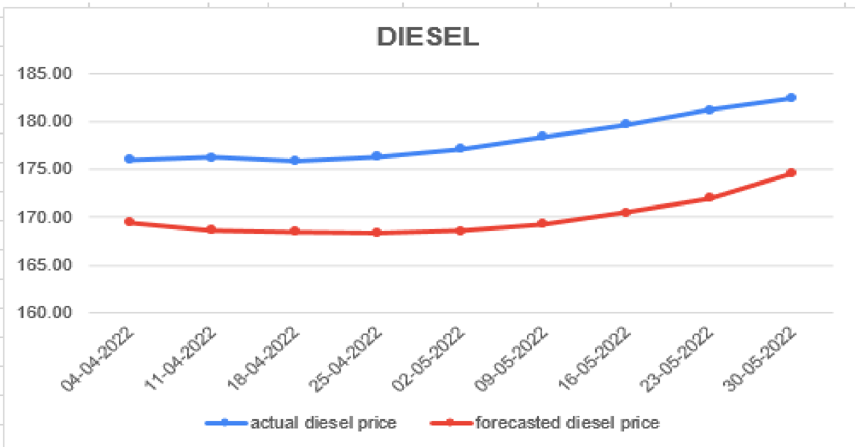
---all prices are in same unit (pence/liter)



$RMSE_{petrol} = 1.390$ $\quad$ $MAPE_{petrol} = 0.631$

As we see, the forecasted prices for petrol match quite well with actual prices. MAPE value tells us that on average the error magnitude in forecasting is only 0.63% of the actual price, and from RMSE we say average distance of forecasted and observed values is 1.39unit.

But as we see, this is not so good for diesel price forecasting. MAPE value tells us that on average the error magnitude in forecasting is 4.57% of the actual diesel price, and from RMSE we say average distance of forecasted and observed values is 8.19unit.



$RMSE_{diesel} = 8.194$ $\quad$ $MAPE_{diesel} = 4.572$

We can also make projection of the average oil price for next few weeks….

date	forecasted petrol price	forecasted diesel price
6june,2022	170.54	177.54
13june,2022	170.64	177.65
20june,2022	170.73	177.75
27june,2022	170.83	177.85
4july,2022	170.93	177.96
11july,2022	171.03	178.06
18july,2022	171.13	178.17
25july,2022	171.23	178.27

---all prices are in same unit (pence/liter)

CONCLUSION:

All we can say ,

- In immediate future there is no tendency to reduce in fuel cost for UK peoples, as the time-plot is increasing in nature.
- In absence of seasonality , we can not say in advance when price will increase or decrease significantly.
- About the correctness of our model, it is obvious that any statistical techniques are based on the assumption that……..

  existing patterns will continue into the future

  ----this assumption is more likely to be correct over the short term than it is over the long term, and for this reason these techniques that provide us with reasonably accurate forecasts for the immediate future, but the same may perform quite poorly further into the future (unless the data patterns are extraordinarily stable).
- And lastly, we can't predict randomness that is why it is named so. We can't model it in advance, if the increasing curve faces a sudden dip with no systematic reason at all.

REMARKS :-

- *Note that for petrol price our forecasts are based on time series analysis. But for diesel instead of performing time series analysis once again, we noticed that there is a strong linear relationship between both price and choose the easy wayout of regressing upon the forecasted petrol price as forecasted diesel price of that same time. Though it saved our effort or time , some additional error introduced in this step. So, it is justified that forecast of diesel price would not be as good as that of petrol price.*
- *Here high correlation between petrol & diesel price is because they both are affected by some economic and other factors in a similar way, not due to any cause-effect relationship.*

References:-

https://www.gov.uk/government/statistics/weekly-road-fuel-prices

https://www.ukogplc.com/page.php?pID=72

https://otexts.com/fpp2/

https://medium.com/vitrox-publication/what-is-a-time-series-forecasting-d020d657f11a

https://www.statcan.gc.ca/en/dai/btd/tce-faq

https://medium.com/analytics-vidhya/time-series-decomposition-part-i-trend-cycle-computation-29fac227896a

Books:-

THE ANALYSIS OF TIME SERIES, AN INTRODUCTION –Chris Chatfield

FUNDAMENTALS OF APPLIED STATISTICS – S.C. Gupta & V.K. Kapoor

Softwares used:-

1. RStudio (R version 4.1.3)
2. Microsoft Excel (2013)
3. Microsoft Word (2013)

Useful calculations:-

https://docs.google.com/spreadsheets/d/1gl8HLJNEJOJqcdXzaWP1NFyBFZaJBTcl/edit?usp=sharing&ouid=108980753175037445540&rtpof=true&sd=true

Codes:-

https://drive.google.com/file/d/1gSDbDmIv40ajwGHbfP1PJKMrRSDMEtmg/view?usp=sharing

https://drive.google.com/file/d/1Ahnmzd6Q3nxnwGDjl6_3zR6ZDAmV2YQY/view?usp=sharing

https://drive.google.com/file/d/19EiyuJn9uw0ixkik2FwXJ1haRxYB7t3t/view?usp=sharing

https://drive.google.com/file/d/1_b5M3aCbJsH5_RDr6Y1AIXSiPlQAkHbu/view?usp=sharing

https://drive.google.com/file/d/1vXYsewhtMhvDz34B-ex0VIa-nE84eeiN/view?usp=sharing

-:ACKNWOLEDGEMENT:-

I would like to express my sincere gratitude to several individuals and organizations for supporting me throughout my final semester project. First, I wish to express my sincere gratitude to my supervisor, Professor Oindrila Bose, for her enthusiasm, patience, insightful comments, helpful information, practical advice and unceasing ideas that have helped me tremendously at all times in my research and doing of this project. Her immense knowledge, profound experience and professional expertise in statistics has enabled me to complete this research successfully. Without her support and guidance, this project would not have been possible. I could not have imagined having a better supervisor in my study. I would also wish to thank my Professor Dr. Dhiman Dutta, Head of the Department of Statistics, Asutosh College, and also Dr. Shirsendu Mukherjee, Dr. Sankha Bhattacharya, Dr. Parthasarathi Bera for their guidance through out the project. I also wish to express my sincere thanks to the University of Calcutta for giving us the chance for doing project. Thanks for all your encouragement!

-:DECLARATION:-

I, **Ramit Nandi**, a student of B.Sc. Sem-VI, Statistics Honours, of University of Calcutta, Registration No. **012-1111-0558-19**; Roll No. **193012-21-0381** hereby declare that I have done this piece of project work entitled as "Price of petrol & diesel in UK – a time-series analysis" under the supervision of **Prof. Oindrila Bose** ,Department of Statistics, Asutosh College, as a part of B.Sc. Sem-VI examination according to the syllabus paper **DSE-B2**. I further declare that the piece of project has not been published elsewhere for any degree or diploma or taken from any published project.

APPENDIX

Data-sets:

Date	UL SP	UL SD
09-06-2003	74.59	76.77
16-06-2003	74.47	76.69
23-06-2003	74.42	76.62
30-06-2003	74.35	76.51
07-07-2003	74.28	76.46
14-07-2003	74.21	76.41
21-07-2003	75.07	76.9
28-07-2003	75.1	76.86
04-08-2003	75.12	76.81
11-08-2003	75.44	77.08
18-08-2003	75.81	77.44
26-08-2003	76.05	77.68
01-09-2003	76.13	77.59
08-09-2003	76.23	77.67
15-09-2003	76.2	77.63
22-09-2003	76.15	77.55
29-09-2003	76.08	77.53
06-10-2003	76.43	77.66
13-10-2003	75.9	77.43

1

20-10-2003	76.05	77.54
27-10-2003	76	77.51
03-11-2003	76	77.52
10-11-2003	75.94	77.47
17-11-2003	75.93	77.52
24-11-2003	76.02	77.67
01-12-2003	76	77.66
08-12-2003	75.97	77.63
15-12-2003	75.96	77.62
22-12-2003	75.92	77.61
29-12-2003	75.94	77.6
05-01-2004	75.88	77.54
12-01-2004	75.9	77.54
19-01-2004	76.36	78.06
26-01-2004	76.44	78.08
02-02-2004	76.48	78.13
09-02-2004	76.5	77.98
16-02-2004	76.44	78
23-02-2004	76.38	77.93
01-03-2004	76.32	77.87

2

08-03-2004	76.73	78.25
15-03-2004	77.39	78.8
22-03-2004	77.44	78.92
29-03-2004	77.59	79.01
05-04-2004	77.86	79.23
12-04-2004	77.88	79.27
19-04-2004	78.04	79.43
26-04-2004	78.4	79.75
04-05-2004	79.28	80.64
10-05-2004	80.3	81.68
17-05-2004	81.49	82.83
24-05-2004	82.13	83.48
31-05-2004	82.35	83.61
07-06-2004	82.44	83.65
14-06-2004	81.54	82.6
21-06-2004	81.04	82.09
28-06-2004	80.41	81.04
05-07-2004	79.87	80.51
12-07-2004	80.34	81.05
19-07-2004	80.84	81.68

3

26-07-2004	80.87	81.7
02-08-2004	80.86	81.79
09-08-2004	81.18	82.14
16-08-2004	81.25	82.33
23-08-2004	81.51	82.82
31-08-2004	81.53	82.89
06-09-2004	81.37	82.92
13-09-2004	81.31	82.91
20-09-2004	81.28	83.03
27-09-2004	81.63	83.46
04-10-2004	82.05	83.84
11-10-2004	83.11	85.01
18-10-2004	83.35	85.34
25-10-2004	83.94	86.03
01-11-2004	84.27	86.42
08-11-2004	84.34	86.54
15-11-2004	84.26	86.42
22-11-2004	81.46	85.17
29-11-2004	84.17	86.36
06-12-2004	84.06	86.4

4

13-12-2004	82.73	85.98
20-12-2004	81.46	85.17
27-12-2004	81.2	85.14
03-01-2005	81.17	85.12
10-01-2005	79.84	84.87
17-01-2005	78.93	83.94
24-01-2005	78.93	83.87
31-01-2005	79.54	84.23
07-02-2005	79.84	84.27
14-02-2005	80.03	84.32
21-02-2005	80.15	84.36
28-02-2005	80.45	84.65
07-03-2005	81.16	85.64
14-03-2005	81.64	86.11
21-03-2005	82.7	87.06
28-03-2005	83.4	87.68
04-04-2005	83.91	88.3
11-04-2005	85.61	89.82
18-04-2005	85.63	89.8
25-04-2005	85.51	89.62

5

02-05-2005	85.44	89.3
09-05-2005	85.36	89.46
16-05-2005	85.27	89.4
23-05-2005	84.77	89
30-05-2005	84.18	88.39
06-06-2005	84.01	88.23
13-06-2005	84.99	88.96
20-06-2005	85.52	89.54
27-06-2005	86.73	90.27
04-07-2005	87.49	91
11-07-2005	87.73	91.89
18-07-2005	88.55	92.67
25-07-2005	89	93.03
01-08-2005	89.25	93.26
08-08-2005	89.67	93.63
15-08-2005	90.56	94.5
22-08-2005	90.77	94.68
29-08-2005	91.42	95.33
05-09-2005	94.37	97.19
12-09-2005	95.07	97.89

6

19-09-2005	94.41	97.32
26-09-2005	93.33	96.27
03-10-2005	93.51	96.61
10-10-2005	94.25	97.05
17-10-2005	94.14	97
24-10-2005	93.47	96.8
31-10-2005	92.79	96.57
07-11-2005	92.47	96.59
14-11-2005	90.53	94.92
21-11-2005	88.61	93.22
28-11-2005	87.35	91.8
05-12-2005	87.35	91.8
12-12-2005	87.2	91.48
19-12-2005	87.56	91.83
26-12-2005	87.88	92.12
02-01-2006	88.03	92.27
09-01-2006	88.56	92.94
16-01-2006	89.01	93.19
23-01-2006	89.55	93.62
30-01-2006	89.88	93.93

7

06-02-2006	89.73	93.85
13-02-2006	89.65	93.77
20-02-2006	89.6	93.75
27-02-2006	89.51	93.69
06-03-2006	89.37	93.63
13-03-2006	89.55	93.85
20-03-2006	90.19	94.42
27-03-2006	91.16	94.95
03-04-2006	91.78	95.64
10-04-2006	92.77	96.4
18-04-2006	94.23	97.64
24-04-2006	95.35	98.5
01-05-2006	96.13	98.9
08-05-2006	96.46	98.9
15-05-2006	96.39	98.74
22-05-2006	96.1	98.5
29-05-2006	95.36	97.69
05-06-2006	95.24	97.55
12-06-2006	95.49	97.75
19-06-2006	95.44	97.77

8

26-06-2006	94.97	97.36
03-07-2006	95.27	97.6
10-07-2006	95.69	97.99
17-07-2006	96.88	98.74
24-07-2006	97.78	99.41
31-07-2006	97.83	99.41
07-08-2006	98.05	99.57
14-08-2006	97.82	99.45
21-08-2006	95.74	97.93
28-08-2006	94.4	96.67
04-09-2006	92.88	95.92
11-09-2006	91.48	95.34
18-09-2006	89.78	94.95
25-09-2006	88.2	93.51
02-10-2006	86.95	92.13
09-10-2006	86.6	91.74
16-10-2006	85.91	91.59
23-10-2006	85.74	91.22
30-10-2006	85.6	91.17
06-11-2006	85.54	91.12

9

13-11-2006	85.49	91.15
20-11-2006	85.45	91.17
27-11-2006	85.47	91.17
04-12-2006	85.46	91.5
11-12-2006	87.44	92.96
18-12-2006	87.7	93.26
25-12-2006	87.9	93.41
01-01-2007	87.85	93.4
08-01-2007	87.66	93.11
15-01-2007	87.1	91.55
22-01-2007	86.28	90.45
29-01-2007	85.66	89.7
05-02-2007	85.62	89.63
12-02-2007	86.12	90.11
19-02-2007	86.6	90.55
26-02-2007	87.07	91.02
05-03-2007	87.74	91.59
12-03-2007	88.48	92.17
19-03-2007	88.93	92.61
26-03-2007	89.32	92.91

10

02-04-2007	90.47	93.9
10-04-2007	91.18	94.33
16-04-2007	91.99	94.86
23-04-2007	92.93	95.39
30-04-2007	93.45	95.71
07-05-2007	94.43	96.12
14-05-2007	95.07	96.47
21-05-2007	95.71	96.69
28-05-2007	96.16	96.99
04-06-2007	96.51	97.19
11-06-2007	96.61	97.2
18-06-2007	96.56	97.11
25-06-2007	96.54	97.12
02-07-2007	96.36	96.89
09-07-2007	96.2	96.78
16-07-2007	96.21	96.79
23-07-2007	96.23	96.83
30-07-2007	96.2	96.77
06-08-2007	96.15	96.72
13-08-2007	96	96.7
20-08-2007	95.31	96.59
27-08-2007	95	96.37
03-09-2007	94.8	96.37
10-09-2007	94.63	96.34
17-09-2007	94.65	96.45

11

17-03-2008	106.41	113.08
24-03-2008	106.9	114.4
31-03-2008	107.1	114.79
07-04-2008	107.24	115.47
14-04-2008	107.61	116.51
21-04-2008	108.57	118.06
28-04-2008	109.5	119.1
05-05-2008	110.3	120.37
12-05-2008	111.04	121.49
19-05-2008	112.96	124.5
26-05-2008	114.79	127.21
02-06-2008	116.03	128.82
09-06-2008	116.86	129.68
16-06-2008	117.69	130.66
23-06-2008	118.44	131.45
30-06-2008	118.84	132.03
07-07-2008	119.28	132.55
14-07-2008	119.44	132.87
21-07-2008	119.4	132.89
28-07-2008	116.91	130.71
04-08-2008	115.06	128.19
11-08-2008	113.92	126.2
18-08-2008	112.22	124.1
25-08-2008	112.02	123.74
01-09-2008	112.2	123.89

13

02-03-2009	90.28	100.64
09-03-2009	90.15	100.03
16-03-2009	90.16	99.82
23-03-2009	90.16	99.52
30-03-2009	91.35	99.98
06-04-2009	93.72	102.14
13-04-2009	94.11	102.44
20-04-2009	94.62	102.71
27-04-2009	95.13	102.83
04-05-2009	95.6	102.8
11-05-2009	96.59	103.1
18-05-2009	97.54	103.37
25-05-2009	98.9	103.55
01-06-2009	99.39	103.55
08-06-2009	100.59	103.78
15-06-2009	101.98	104.5
22-06-2009	102.81	104.92
29-06-2009	103.3	105.05
06-07-2009	103.57	105.11
13-07-2009	102.78	104.04
20-07-2009	102.22	102.99
27-07-2009	102.17	102.91
03-08-2009	102.36	103.2
10-08-2009	103.01	103.79
17-08-2009	103.96	104.45

15

15-02-2010	111.89	113.68
22-02-2010	111.56	112.76
01-03-2010	112.42	113.41
08-03-2010	113.71	114.68
15-03-2010	115.46	116.21
22-03-2010	116.59	117.28
29-03-2010	117.28	118.21
05-04-2010	119.29	120.09
12-04-2010	119.94	121.11
19-04-2010	120.55	121.74
26-04-2010	120.99	122.31
03-05-2010	121.18	122.68
10-05-2010	121.31	122.82
17-05-2010	121.3	122.83
24-05-2010	119.93	121.92
31-05-2010	118.54	120.76
07-06-2010	118.2	120.4
14-06-2010	117.87	120.21
21-06-2010	117.8	120.39
28-06-2010	117.88	120.38
05-07-2010	117.91	120.44
12-07-2010	117.39	119.78
19-07-2010	116.85	119.24
26-07-2010	116.61	119.03
02-08-2010	116.51	119.03

17

31-01-2011	128.55	133.25
07-02-2011	128.64	133.47
14-02-2011	128.73	133.71
21-02-2011	128.92	134.04
28-02-2011	129.72	135.06
07-03-2011	131.28	136.9
14-03-2011	132.18	138.25
21-03-2011	133.11	139.53
28-03-2011	132.44	139.04
04-04-2011	132.81	139.46
11-04-2011	133.68	140.33
18-04-2011	135.06	141.34
25-04-2011	135.62	142.11
02-05-2011	136.29	142.53
09-05-2011	137.05	143.06
16-05-2011	136.4	141.5
23-05-2011	136.11	140.77
30-05-2011	135.57	139.81
06-06-2011	135.62	139.58
13-06-2011	135.86	139.77
20-06-2011	136.03	140.05
27-06-2011	134.86	138.77
04-07-2011	133.43	137.7
11-07-2011	134.18	138.4
18-07-2011	134.97	139.36

19

24-09-2007	94.91	96.9
01-10-2007	96.39	98.51
08-10-2007	97.14	99.1
15-10-2007	97.19	99.18
22-10-2007	97.81	99.89
29-10-2007	98.26	100.69
05-11-2007	99.36	102.16
12-11-2007	100.74	104.17
19-11-2007	101.47	105.28
26-11-2007	102.13	106.25
03-12-2007	102.58	107.19
10-12-2007	102.67	107.47
17-12-2007	102.6	107.56
24-12-2007	102.53	107.5
31-12-2007	102.66	107.61
07-01-2008	103.37	108.29
14-01-2008	103.82	108.77
21-01-2008	104.22	109.09
28-01-2008	104.22	109.11
04-02-2008	104.03	108.99
11-02-2008	103.67	108.86
18-02-2008	103.7	108.99
25-02-2008	104.65	110.21
03-03-2008	105.24	110.9
10-03-2008	105.96	112.23

12

08-09-2008	112.55	124.2
15-09-2008	112.51	124.14
22-09-2008	110.81	122.61
29-09-2008	109.85	121.65
06-10-2008	109.45	121.14
13-10-2008	107.04	118.42
20-10-2008	102.66	114.31
27-10-2008	98.62	111.01
03-11-2008	97.4	109.87
10-11-2008	95.89	109.13
17-11-2008	94.83	108.73
24-11-2008	92.79	107.78
01-12-2008	91.47	106.56
08-12-2008	90.16	104.87
15-12-2008	89.07	101.24
22-12-2008	88.01	99.94
29-12-2008	87.34	99.17
05-01-2009	85.42	97.57
12-01-2009	86.02	98.32
19-01-2009	86.19	98.45
26-01-2009	87.42	99.39
02-02-2009	88.16	99.82
09-02-2009	89.28	100.31
16-02-2009	90.12	100.86
23-02-2009	90.33	100.81

14

24-08-2009	104.7	105.11
31-08-2009	105.11	105.5
07-09-2009	106.85	107.3
14-09-2009	106.43	107.09
21-09-2009	105.96	106.79
28-09-2009	105.44	106.39
05-10-2009	104.87	105.85
12-10-2009	104.71	105.69
19-10-2009	105.08	106.28
26-10-2009	106.76	107.78
02-11-2009	107.43	108.51
09-11-2009	108.1	109.35
16-11-2009	108.48	109.65
23-11-2009	108.73	109.87
30-11-2009	108.79	109.99
07-12-2009	108.64	109.87
14-12-2009	108.5	109.57
21-12-2009	107.47	109.13
28-12-2009	107.44	109.17
04-01-2010	109.34	111.08
11-01-2010	110.72	112.59
18-01-2010	111.6	113.39
25-01-2010	111.87	113.71
01-02-2010	111.91	113.75
08-02-2010	111.95	113.79

16

09-08-2010	116.35	118.94
16-08-2010	116.34	118.79
23-08-2010	115.64	117.9
30-08-2010	114.98	117.37
06-09-2010	114.59	117.09
13-09-2010	114.66	117.17
20-09-2010	115.06	117.72
27-09-2010	115.39	118.11
04-10-2010	116.01	118.97
11-10-2010	116.78	119.77
18-10-2010	117.28	120.6
25-10-2010	117.74	121.17
01-11-2010	118.18	121.66
08-11-2010	118.67	122.19
15-11-2010	118.8	122.51
22-11-2010	118.99	122.74
29-11-2010	119.21	123.2
06-12-2010	120.54	124.62
13-12-2010	121.68	125.77
20-12-2010	122.31	126.52
27-12-2010	123.44	127.68
03-01-2011	124.85	129.11
10-01-2011	127.4	131.6
17-01-2011	127.87	132.33
24-01-2011	128.35	133

18

25-07-2011	135.66	139.95
01-08-2011	136.07	140.55
08-08-2011	136.46	140.84
15-08-2011	135.95	140.36
22-08-2011	134.56	139.04
29-08-2011	134.32	138.72
05-09-2011	134.62	138.96
12-09-2011	135	139.4
19-09-2011	135.4	139.86
26-09-2011	135.65	140.14
03-10-2011	135.02	139.77
10-10-2011	134.47	139.52
17-10-2011	134.49	139.73
24-10-2011	134.66	140.19
31-10-2011	134.73	140.27
07-11-2011	134.41	140.42
14-11-2011	133.68	140.53
21-11-2011	133.38	141.06
28-11-2011	133.17	141.1
05-12-2011	132.78	140.99
12-12-2011	132.53	140.95
19-12-2011	132.49	140.93
26-12-2011	132.41	140.77
02-01-2012	132.4	140.84
09-01-2012	132.68	141

20

16-01-2012	133.29	141.69
23-01-2012	133.72	142.06
30-01-2012	134.1	142.45
06-02-2012	134.25	142.24
13-02-2012	134.9	142.83
20-02-2012	135.21	143.15
27-02-2012	136.21	143.92
05-03-2012	137.3	144.71
12-03-2012	137.9	145.15
19-03-2012	138.96	146.16
26-03-2012	139.48	146.56
02-04-2012	140.99	147.68
09-04-2012	141.97	148
16-04-2012	142.17	148.04
23-04-2012	141.76	147.96
30-04-2012	141	147.05
07-05-2012	140.14	146.4
14-05-2012	138.29	144.52
21-05-2012	135.77	141.75
28-05-2012	134.37	140.49

21

04-06-2012	134.06	139.96
11-06-2012	133.67	139.58
18-06-2012	131.99	137.71
25-06-2012	131.55	137.2
02-07-2012	130.5	135.91
09-07-2012	131.06	136.44
16-07-2012	131.49	136.85
23-07-2012	132.21	137.72
30-07-2012	132.78	138.25
06-08-2012	133.39	138.79
13-08-2012	134.27	139.46
20-08-2012	135.72	140.79
27-08-2012	137.1	142.17
03-09-2012	138.03	142.87
10-09-2012	138.91	143.66
17-09-2012	139.42	144.18
24-09-2012	139.54	144.36
01-10-2012	138.43	143.28
08-10-2012	138.24	143.05
15-10-2012	138.35	143.23

22

22-10-2012	138.29	143.23
29-10-2012	136.98	143
05-11-2012	136.43	142.75
12-11-2012	135.09	141.74
19-11-2012	134.56	141.33
26-11-2012	134.38	141.13
03-12-2012	133.05	140.91
10-12-2012	132.55	140.76
17-12-2012	131.98	140.08
24-12-2012	131.95	139.84
31-12-2012	131.92	139.81
07-01-2013	131.86	139.75
14-01-2013	132	139.68
21-01-2013	132.21	139.95
28-01-2013	132.95	140.77
04-02-2013	134.38	141.91
11-02-2013	135.56	143.03
18-02-2013	136.86	144.17
25-02-2013	138.5	145.55
04-03-2013	139.47	146.34

23

11-03-2013	140	146.71
18-03-2013	137.58	144.84
25-03-2013	137.09	144.33
01-04-2013	137.3	143.51
08-04-2013	137.31	143.13
15-04-2013	137.11	141.67
22-04-2013	135.68	140.64
29-04-2013	134.66	139.59
06-05-2013	133.42	138.54
13-05-2013	133.12	138.27
20-05-2013	133.17	138.42
27-05-2013	133.64	138.82
03-06-2013	134.04	139.17
10-06-2013	134.3	139.43
17-06-2013	134.46	139.54
24-06-2013	134.52	139.64
01-07-2013	133.44	138.51
08-07-2013	133.77	138.73
15-07-2013	134.95	139.84
22-07-2013	136.04	140.72

24

29-07-2013	136.87	141.44
05-08-2013	137.02	141.67
12-08-2013	137.13	141.82
19-08-2013	137	141.73
26-08-2013	136.83	141.78
02-09-2013	137.14	142.09
09-09-2013	137.6	142.67
16-09-2013	137.47	142.52
23-09-2013	136.95	142.09
30-09-2013	133.4	140.19
07-10-2013	131.62	138.83
14-10-2013	131.67	139.15
21-10-2013	131.53	138.99
28-10-2013	131.42	138.88
04-11-2013	129.92	137.51
11-11-2013	129.71	137.32
18-11-2013	129.89	137.47
25-11-2013	130.03	137.72
02-12-2013	130.25	137.9
09-12-2013	130.76	138.55

25

16-12-2013	130.96	138.84
23-12-2013	130.05	138.17
30-12-2013	130.03	137.98
06-01-2014	130.26	138.11
13-01-2014	130.36	138.34
20-01-2014	129.37	137.23
27-01-2014	129.02	136.94
03-02-2014	128.85	136.84
10-02-2014	128.81	136.84
17-02-2014	129.16	136.81
24-02-2014	129.2	136.95
03-03-2014	129.36	137.15
10-03-2014	129.39	137.08
17-03-2014	128.76	136.27
24-03-2014	128.57	136
31-03-2014	128.76	135.89
07-04-2014	128.71	136.01
14-04-2014	128.89	136.05
21-04-2014	129.11	136.12
28-04-2014	129.09	135.93
05-05-2014	129.27	136.16
12-05-2014	129.5	136.37
19-05-2014	129.42	136.17
26-05-2014	129.51	136.36
02-06-2014	129.63	136.38

26

01-12-2014	121.18	126.11
08-12-2014	119.83	124.79
15-12-2014	116.3	122.43
22-12-2014	113.66	120.81
29-12-2014	113.16	120.36
05-01-2015	111.06	117.93
12-01-2015	108.87	116.22
19-01-2015	106.83	114.33
26-01-2015	106.32	113.76
02-02-2015	106.04	113.61
09-02-2015	106.35	113.91
16-02-2015	107.26	114.73
23-02-2015	108.34	115.74
02-03-2015	109.2	116.64
09-03-2015	110.07	117.51
16-03-2015	111.04	118.26
23-03-2015	111.65	118.71
30-03-2015	112.06	118.97
06-04-2015	112.37	119.09
13-04-2015	112.52	119.15
20-04-2015	112.87	119.14
27-04-2015	113.55	119.58
04-05-2015	114.27	120
11-05-2015	115.1	120.59
18-05-2015	115.74	121.04

28

16-11-2015	107.36	110.32
23-11-2015	107.29	110.27
30-11-2015	106.97	110.01
07-12-2015	106.75	109.92
14-12-2015	104	108.07
21-12-2015	102.33	107.04
28-12-2015	102.11	106.43
04-01-2016	102	106.32
11-01-2016	101.89	103.41
18-01-2016	101.8	102.81
25-01-2016	101.42	101.54
01-02-2016	101.36	100.84
08-02-2016	101.51	101.14
15-02-2016	101.45	101.21
22-02-2016	101.39	101.13
29-02-2016	101.42	101.38
07-03-2016	101.46	101.67
14-03-2016	101.71	102.52
21-03-2016	102.58	103.64
28-03-2016	103.4	104.4
04-04-2016	104.51	105.32
11-04-2016	105.39	105.98
18-04-2016	106.41	106.97
25-04-2016	107.14	107.67
02-05-2016	107.82	108.58

30

31-10-2016	116.02	118.42
07-11-2016	116.56	119.03
14-11-2016	116.5	118.92
21-11-2016	114.71	117.42
28-11-2016	113.72	116.57
05-12-2016	113.66	116.75
12-12-2016	114.23	117.51
19-12-2016	115	118.61
26-12-2016	115.45	119.03
02-01-2017	117	120.03
09-01-2017	117.98	121.3
16-01-2017	118.63	122.1
23-01-2017	119.43	122.75
30-01-2017	120.01	123.19
06-02-2017	120.18	123.25
13-02-2017	120.15	123.17
20-02-2017	120.2	123.15
27-02-2017	120.12	123.06
06-03-2017	120.02	122.9
13-03-2017	119.8	122.77
20-03-2017	118.95	121.84
27-03-2017	118.03	121.08
03-04-2017	116.93	119.92
10-04-2017	117.02	119.98
17-04-2017	117.54	120.36

32

16-10-2017	117.01	120.42
23-10-2017	117.04	120.65
30-10-2017	117.06	120.63
06-11-2017	117.57	121.27
13-11-2017	118.91	122.69
20-11-2017	119.67	123.32
27-11-2017	120.13	123.79
04-12-2017	119.75	123.21
11-12-2017	119.91	123.69
18-12-2017	119.98	123.81
25-12-2017	120	123.43
01-01-2018	120.19	123.51
08-01-2018	120.52	123.97
15-01-2018	120.96	124.58
22-01-2018	121.5	125.07
29-01-2018	121.69	125.32
05-02-2018	121.73	125.36
12-02-2018	121.58	125.02
19-02-2018	120.52	124.18
26-02-2018	119.55	123.09
05-03-2018	119.29	122.92
12-03-2018	119.11	122.96
19-03-2018	119.17	122.84
26-03-2018	119.23	122.9
02-04-2018	119.43	123.05

34

09-06-2014	129.69	136.04
16-06-2014	129.88	135.54
23-06-2014	130.14	135.5
30-06-2014	130.69	136.04
07-07-2014	131.1	136.22
14-07-2014	131.11	136.17
21-07-2014	131.04	135.69
28-07-2014	130.83	135.5
04-08-2014	129.63	134.31
11-08-2014	129.41	133.85
18-08-2014	129.29	133.74
25-08-2014	128.34	133.16
01-09-2014	128.01	132.77
08-09-2014	128.34	133.09
15-09-2014	128.43	133.13
22-09-2014	128.62	133.32
29-09-2014	128.59	133.23
06-10-2014	126.78	131.43
13-10-2014	126.5	131.08
20-10-2014	125.4	130
27-10-2014	125.11	129.72
03-11-2014	123.94	128.56
10-11-2014	122.94	127.59
17-11-2014	122.5	127.31
24-11-2014	122.29	127.17

27

25-05-2015	116.02	121.32
01-06-2015	116.13	121.3
08-06-2015	116.26	121.34
15-06-2015	116.39	121.3
22-06-2015	116.6	121.21
29-06-2015	116.55	120.99
06-07-2015	116.34	120.76
13-07-2015	116.41	118.83
20-07-2015	116.79	118.43
27-07-2015	116.44	116.95
03-08-2015	116.12	115.17
10-08-2015	115.48	114.2
17-08-2015	114.58	112.05
24-08-2015	113.23	111.05
31-08-2015	111.21	109.92
07-09-2015	110.9	109.78
14-09-2015	110.6	109.95
21-09-2015	110.61	110.52
28-09-2015	109.47	110.45
05-10-2015	108.94	110.57
12-10-2015	108.97	110.78
19-10-2015	109.01	111.17
26-10-2015	108.04	110.61
02-11-2015	107.21	110.14
09-11-2015	107.14	110.08

29

09-05-2016	108.19	109
16-05-2016	108.53	109.3
23-05-2016	109.12	109.9
30-05-2016	109.79	110.7
06-06-2016	110.4	111.43
13-06-2016	111.03	111.99
20-06-2016	111.22	112.31
27-06-2016	111.44	112.56
04-07-2016	111.65	112.86
11-07-2016	111.89	113.09
18-07-2016	111.92	113
25-07-2016	111.69	112.89
01-08-2016	111.05	112.43
08-08-2016	109.75	111.4
15-08-2016	109.28	111.02
22-08-2016	109.63	111.71
29-08-2016	110.43	112.55
05-09-2016	110.97	113.1
12-09-2016	111.31	113.42
19-09-2016	111.29	113.23
26-09-2016	111.36	113.38
03-10-2016	111.65	113.71
10-10-2016	112.35	114.4
17-10-2016	113.72	116
24-10-2016	115.2	117.7

31

24-04-2017	118.15	120.77
01-05-2017	118.08	120.78
08-05-2017	117.06	119.34
15-05-2017	115.6	117.65
22-05-2017	115.67	117.76
29-05-2017	115.79	117.88
05-06-2017	115.82	117.95
12-06-2017	115.69	117.76
19-06-2017	115.12	117.18
26-06-2017	114.43	116.26
03-07-2017	113.74	115.63
10-07-2017	113.58	115.29
17-07-2017	113.89	115.58
24-07-2017	113.99	115.73
31-07-2017	114.16	115.99
07-08-2017	114.78	116.64
14-08-2017	115.53	117.43
21-08-2017	116.04	118.05
28-08-2017	116.4	118.24
04-09-2017	116.95	118.77
11-09-2017	118.14	119.94
18-09-2017	118.85	120.59
25-09-2017	118.87	120.92
02-10-2017	118.13	120.37
09-10-2017	117.16	120.51

33

09-04-2018	119.96	123.58
16-04-2018	120.56	124.29
23-04-2018	121.44	125.39
30-04-2018	122.22	126.01
07-05-2018	123.27	126.99
14-05-2018	124.32	128.28
21-05-2018	125.93	130.04
28-05-2018	127.58	131.66
04-06-2018	128.43	132.34
11-06-2018	128.61	132.62
18-06-2018	127.92	131.98
25-06-2018	127.18	131.27
02-07-2018	126.54	130.83
09-07-2018	126.97	131.53
16-07-2018	127.47	131.85
23-07-2018	127.5	131.9
30-07-2018	127.54	131.94
06-08-2018	128.02	132.39
13-08-2018	128.39	132.56
20-08-2018	128.79	132.97
27-08-2018	129.18	133.2
03-09-2018	129.76	133.67
10-09-2018	130.24	134.31
17-09-2018	130.59	134.57
24-09-2018	130.59	134.77

35

01-10-2018	129.98	134.86
08-10-2018	130.15	135.53
15-10-2018	130.81	136.63
22-10-2018	130.98	136.93
29-10-2018	130.64	136.99
05-11-2018	130.11	137.03
12-11-2018	128.94	137.08
19-11-2018	127.4	136.38
26-11-2018	125.77	135.43
03-12-2018	124.3	134.38
10-12-2018	122.37	132.5
17-12-2018	121.02	131.26
24-12-2018	120.61	130.91
31-12-2018	120.18	130.26
07-01-2019	120.27	130.33
14-01-2019	119.53	129.47
21-01-2019	119.12	128.92
28-01-2019	119.29	129.1
04-02-2019	119.13	129.13
11-02-2019	118.97	129.17

36

18-02-2019	119.05	129.23
25-02-2019	119.22	129.66
04-03-2019	119.72	130.25
11-03-2019	120.1	130.59
18-03-2019	120.48	130.85
25-03-2019	120.83	131.15
01-04-2019	121.7	131.48
08-04-2019	122.66	132.08
15-04-2019	124.06	132.96
22-04-2019	125.43	133.99
29-04-2019	126.36	134.6
06-05-2019	127.5	135.41
13-05-2019	127.97	135.36
20-05-2019	128.51	135.82
27-05-2019	129.14	136.45
03-06-2019	129.41	136.39
10-06-2019	128.89	135.4
17-06-2019	127.66	133.76
24-06-2019	126.66	131.81
01-07-2019	126.49	131.55

37

08-07-2019	126.86	131.68
15-07-2019	127.13	131.86
22-07-2019	127.81	132.21
29-07-2019	128.03	132.6
05-08-2019	128.37	132.61
12-08-2019	128.36	132.59
19-08-2019	128.17	132.6
26-08-2019	128.22	132.51
02-09-2019	127.86	132.29
09-09-2019	127.79	131.89
16-09-2019	126.92	131.35
23-09-2019	126.78	131.52
30-09-2019	126.92	131.83
07-10-2019	126.87	131.82
14-10-2019	126.91	131.89
21-10-2019	126.4	131.28
28-10-2019	125.77	130.6
04-11-2019	125.56	130.38
11-11-2019	125.59	130.42
18-11-2019	125.58	130.35

38

25-11-2019	125.32	130.08
02-12-2019	124.81	129.79
09-12-2019	124.75	129.79
16-12-2019	124.33	129.56
23-12-2019	124.16	129.81
30-12-2019	124.96	130.54
06-01-2020	126.09	131.56
13-01-2020	127.17	132.69
20-01-2020	127.18	132.74
27-01-2020	127.33	132.88
03-02-2020	125.85	131.48
10-02-2020	124.73	129.7
17-02-2020	123.8	128.2
24-02-2020	123.45	127.67
02-03-2020	122.43	126.62
09-03-2020	122.24	126.25
16-03-2020	120.33	124.36
23-03-2020	119.64	123.41
30-03-2020	112.45	118.6
06-04-2020	110.23	116.83

39

13-04-2020	109.27	116.15
20-04-2020	108.63	115.73
27-04-2020	107.88	115.22
04-05-2020	107.56	114.94
11-05-2020	107.45	114.83
18-05-2020	105.09	112.22
25-05-2020	104.87	111.7
01-06-2020	105.17	111.76
08-06-2020	105.74	111.98
15-06-2020	106.37	112.53
22-06-2020	107.13	112.98
29-06-2020	108.31	114.1
06-07-2020	109.43	115.05
13-07-2020	111.24	116.75
20-07-2020	112.23	117.35
27-07-2020	112.66	117.76
03-08-2020	112.91	117.97
10-08-2020	113.02	118.01
17-08-2020	113.04	118
24-08-2020	113.22	118.14

40

31-08-2020	113.29	118.18
07-09-2020	113.37	118.22
14-09-2020	113.32	118.18
21-09-2020	113.31	118.16
28-09-2020	113.3	118.14
05-10-2020	113.26	118.11
12-10-2020	113.19	118.05
19-10-2020	113.18	118.08
26-10-2020	113.14	118.08
02-11-2020	113.11	118.06
09-11-2020	112.5	117.28
16-11-2020	112.35	117.07
23-11-2020	112.42	117.13
30-11-2020	112.61	117.41
07-12-2020	113.17	117.99

41

14-12-2020	113.82	118.57
21-12-2020	114.43	119.17
28-12-2020	114.91	119.65
04-01-2021	115.39	119.97
11-01-2021	116.14	120.61
18-01-2021	116.93	121.52
25-01-2021	118.1	122.7
01-02-2021	119.14	123.7
08-02-2021	119.67	124.09
15-02-2021	120.53	125.01
22-02-2021	121.27	125.71
01-03-2021	122.17	126.62
08-03-2021	122.94	127.4
15-03-2021	123.9	128.24
22-03-2021	124.57	129.02

42

29-03-2021	125.13	129.32
05-04-2021	125.24	129.35
12-04-2021	125.4	129.39
19-04-2021	125.48	129.49
26-04-2021	125.8	129.77
03-05-2021	126.09	130
10-05-2021	126.53	130.43
17-05-2021	127.19	131.03
24-05-2021	127.89	131.52
31-05-2021	128.15	131.82
07-06-2021	128.69	132.38
14-06-2021	129.47	133.17
21-06-2021	130.1	133.69
28-06-2021	130.73	134.27
05-07-2021	131.7	134.13

43

12-07-2021	132.47	135.29
19-07-2021	133.2	135.95
26-07-2021	133.48	136.2
02-08-2021	134.21	136.89
09-08-2021	134.7	137.31
16-08-2021	134.77	136.77
23-08-2021	134.68	137.15
30-08-2021	134.66	137.11
06-09-2021	134.76	137.19
13-09-2021	134.75	137.19
20-09-2021	134.86	137.35
27-09-2021	135.19	137.95
04-10-2021	136.1	139.2
11-10-2021	137.17	140.66
18-10-2021	139.46	143.19

44

25-10-2021	141.81	145.9
01-11-2021	143.7	147.48
08-11-2021	144.9	148.84
15-11-2021	145.87	149.76
22-11-2021	146.89	150.73
29-11-2021	147.53	151.31
06-12-2021	146.89	150.61
13-12-2021	146.22	149.88
20-12-2021	145.16	148.89
27-12-2021	144.92	148.82
03-01-2022	145.04	148.85

45

10-01-2022	144.82	148.65
17-01-2022	144.8	148.7
24-01-2022	144.87	148.81
31-01-2022	145.74	149.68
07-02-2022	146.33	150.3
14-02-2022	146.95	151.1
21-02-2022	147.77	151.95
28-02-2022	149.22	153.36
07-03-2022	152.95	158.56
14-03-2022	159.96	169.48
21-03-2022	165.37	177.47
28-03-2022	162.65	176.44

46

descriptive.R

```
library("readxl")
setwd("F:\\PROJECT WORK _OB") #change directory as per convenience
pricedata=data.frame(read_excel("fuel prices time series.xlsx"))

summary(pricedata[,-1])
boxplot(pricedata[,-
1],names=c("petrol","diesel"),ylab="p./liter",main="box-plot for petrol and 
diesel price data")
n=length(pricedata[[1]]) #no. of obs
t=0:(n-1)
p=pricedata[[2]] #petrol price
d=pricedata[[3]] #diesel price
var(p)
var(d)
cor(p,d)


plot(t,p,type="l",col="blue",xlab="weeks",ylab="p./liter",main="petrol 
price over time")
lines(t,d,col="red")
mtext("time: origin= 9june,2003 ",side=1,line=-
3,adj=.8,col="darkgrey",cex=.8)
```

```
mtext("                unit=1week ",side=1,line=-
2,adj=.8,col="darkgrey",cex=.8)
legend("topleft",c("petrol","diesel"),col=c("blue","red"),lwd=1)
```

trend.R

```
modelp1=lm(p~t)#linear trend
summary(modelp1)
modelp2=lm(p~poly(t,2,raw=T))#quadratic trend
summary(modelp2)
modelp3=lm(p~poly(t,3,raw=T))#cubic trend
summary(modelp3)
modelp4=lm(p~poly(t,4,raw=T))#4th degree polynomial trend
summary(modelp4)


plot(t,p,type="l",xlab="weeks",ylab="p./liter",main="linear trend line for
petrol price")
abline(modelp1,col="blue")
mtext(paste("p= ",round(coef(modelp1)[1],3),"
+",round(coef(modelp1)[2],3),"t"),side=1,line=-
2,adj=.2,col="darkgrey",cex=.95)

plot(t,p,type="l",xlab="weeks",ylab="p./liter",main="various trend-cycle
line for petrol price")
lines(t,fitted(modelp2),col="green")
lines(t,fitted(modelp3),col="blue",lwd=2)
lines(t,fitted(modelp4),col="red")
legend("topleft",c("quadratic","cubic","4th
degree"),col=c("green","blue","red"),lwd=1)

plot(t,residuals(modelp3),type="h",xlab="week",ylab="residuals",
     main="residual plot of cubic trend",col="darkgreen")
plot(fitted(modelp3),abs(residuals(modelp3)),type="h",xlab="fitted",ylab="|
residuals |",
     main="fitted vs residual",col="darkgreen")
```

residual.R

```
z=residuals(modelp3)
plot(t,z,type='h',col="darkgreen",xlab="week",ylab="|residual|",main="resid
ual plot")
plot(fitted(modelp3),abs(z),type='h',col="darkred",xlab="fitted",ylab="|res
idual|",main="residual plot")

plot(t[-n],diff(z)-
>dz,type='l',lwd=.8,col="darkgrey",,xlab="week",ylab="del(residual)",main="
first difference of residuals over time")
abline(h=mean(dz),col="red")
mtext("mean line",side=4,line=-1,cex=.85,col="red")

plot(t[-n],((dz-
mean(dz))),type='h',col="darkred",,xlab="week",ylab="del(residual)-
bias",main="adjusted first diff. of residuals over time")
```

```
plot(t[-n],abs((dz-
mean(dz))),type='h',col="darkred",,xlab="week",ylab="|del(residual)-
bias|",main="adjusted first diff. of residuals over time")

estm_z= (z+mean(dz))[-n] # estimated z1 is based on z0 ...,zn is based on
zn-1
```

PvsD.R

```
plot(p,d,col="grey",xlab="petrol",ylab="diesel",main="petrol vs diesel
price(p./liter) for fixed time ")
abline(lm(d~p)->donp,col="red",lwd=2)
mtext(paste("d= ",round(coef(lm(d~p))[1],3),"
+",round(coef(lm(d~p))[2],3),"p"),side=1,line=-
2,adj=.2,col="darkred",cex=.8)
```

comparison.R

```
estm_p= (fitted(modelp3))[-1] +estm_z
estm_d= coefficients(donp)[1] + estm_p*coefficients(donp)[2]

plot(t,p,main="PETROL PRICE",xlab="week",ylab="p./liter",
     col="red",type="l",lwd=3)
points(t[-1],estm_p,col="blue",lwd=.05)
legend("topleft","actual price",col="red",lwd=1.5)
legend("bottomright","fitted value",col="blue",pch=1)
summary(p[-1]-estm_p)
var(p[-1]-estm_p)
100*mean(abs(p[-1]-estm_p)/p[-1])

summary(d[-1]-estm_d)
var(d[-1]-estm_d)
100*mean(abs(d[-1]-estm_d)/d[-1])
plot(t,d,main="DIESEL PRICE",xlab="week",ylab="p./liter",
     col="blue",type="l",lwd=3)
points(t[-1],estm_d,col="red",lwd=.05)
legend("topleft","actual price",col="blue",lwd=1.5)
legend("bottomright","fitted value",col="red",pch=1)
```