Time Series Analysis - Gas

Muthu Pandian G

January 28, 2020

Time Series Analysis

Project Objective:

This project is to analyze Australian Monthly Gas production dataset “Gas” in package “Forecast”. Monthly gas production of Australia between year 1956-1995 is released by Australian Bureau of Statistics which is in time series format. This dataset is available by default in Forecast package in R.

The package contains methods and tools for displaying and analyzing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling.

1. Explore the gas (Australian monthly gas production) dataset in Forecast package to do the following: Read the data as a time series object in R.

2. Plot the data What do you observe? Which components of the time series are present in this dataset?

3. What is the periodicity of dataset? HINT: Please use the dataset from January 1970 for your analysis. Please partition your dataset in such a way that you have the data for only 1994 in the test data.

4. Is the time series Stationary? Inspect visually as well as conduct an ADF test? Write down the null and alternate hypothesis for the stationarity test? De-seasonalise the series if seasonality is present?

5. Develop an initial forecast for next 20 periods. Check the same using the various metrics, after finalising the model, develop a final forecast for the 12 time periods. Use both manual and auto.arima (Show & explain all the steps) HINT: You can apply auto.arima(Train_data (refers to the train data set), seasonal=TRUE if seasonality is present in the data, FALSE is seasonality is not present.)

6. Report the accuracy of the model

Set the Working Directory

```
setwd("D:/Great Lakes/Projects/Time Series Analysis")
```

Reading Data into R

```
library(forecast)

## Warning: package 'forecast' was built under R version 3.6.2

## Registered S3 method overwritten by 'xts':
## method from
## as.zoo.xts zoo

## Registered S3 method overwritten by 'quantmod':
## method from
## as.zoo.data.frame zoo

## Registered S3 methods overwritten by 'forecast':
## method from
## fitted.fracdiff fracdiff
## residuals.fracdiff fracdiff

gas <- window(gas)
```

Let's check the class,start of the series, end of the series, Frequency of the dataset to be sure that it is a time series data set.

```
class(gas)

## [1] "ts"

start(gas)

## [1] 1956    1

end(gas)

## [1] 1995    8

frequency(gas)

## [1] 12
```

About the Dataset

1. Given timeseries is univariate, has only one variable and 476 observations
2. This series starts from 1956 January and ends in 1995 August.
3. This series is a Monthly series, as the frequency of this dataset is 12.
4. Monthly periodicity from Jan 1956 to Aug 1995

Exploratory Data Analysis (EDA)

Summary

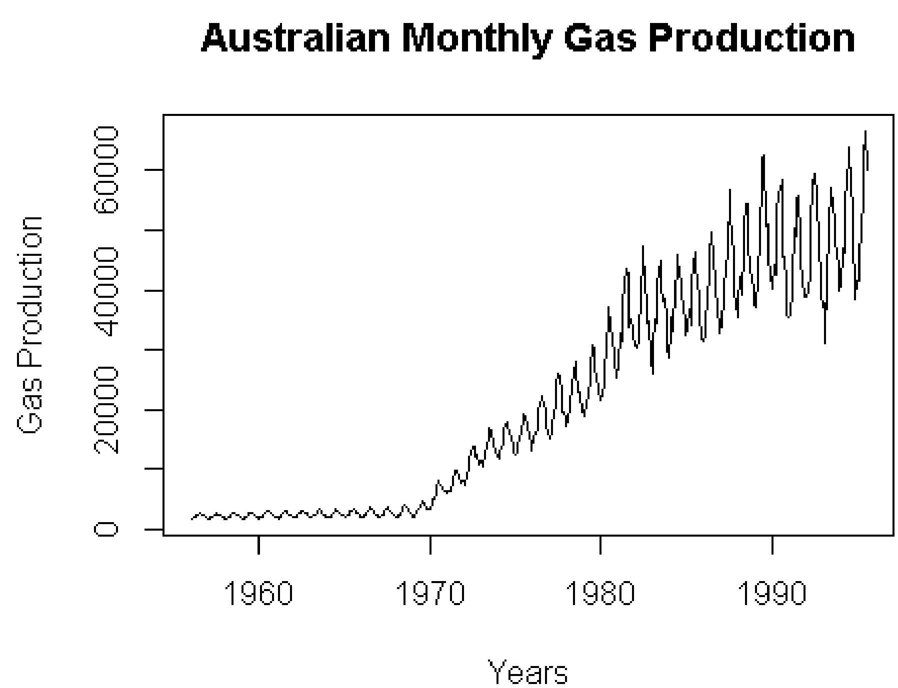
```
summary(gas)

##                Min.    1st    Qu.       Median              Mean    3rd    Qu.               Max.
##    1646    2675   16788   21415   38629   66600
```

Inference

- On an average, 30,394 units of gas is produced during 1956 (Jan) - 1995 (Aug)

Plotting the Time Series

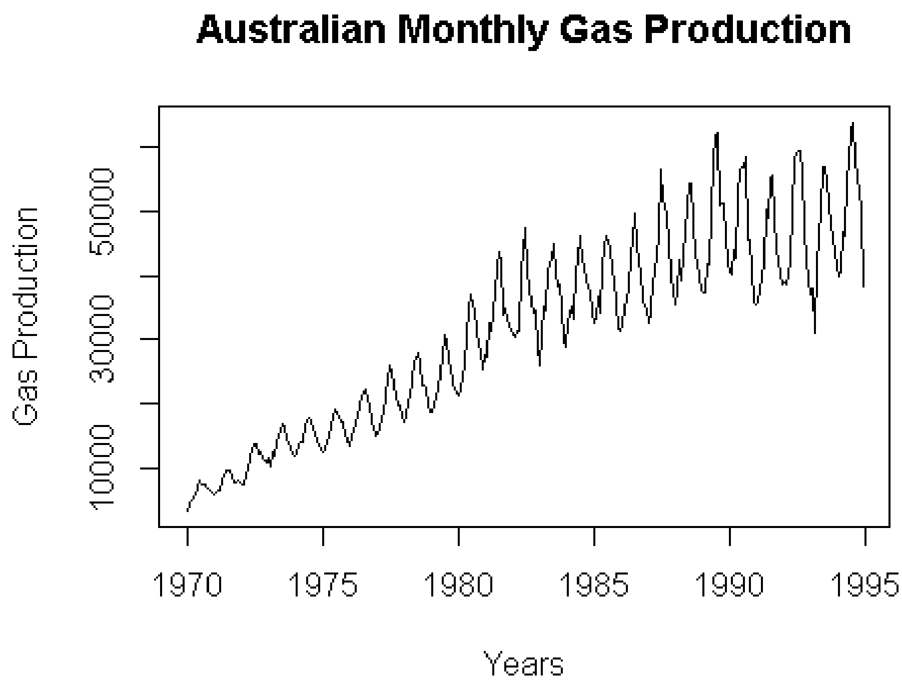
```
plot(gas,main = "Australian Monthly Gas Production",xlab= "Years", ylab = "Gas Production")
```



We can see from the above time plot, that before 1970 there isn't much gas production. This may be due to change in some market conditions, change in company's strategy or due to some other factor.

Hence, it's clear that the data before 1970 won't have much of an impact in forecasting the future. So, we can eliminate the data (1955-1969(December))

```
gas1 <- window(gas,start = c(1970,1), end = c(1994,12), frequency = 12)
plot(gas1,main = "Australian Monthly Gas Production",xlab= "Years", ylab = "Gas Production")
```

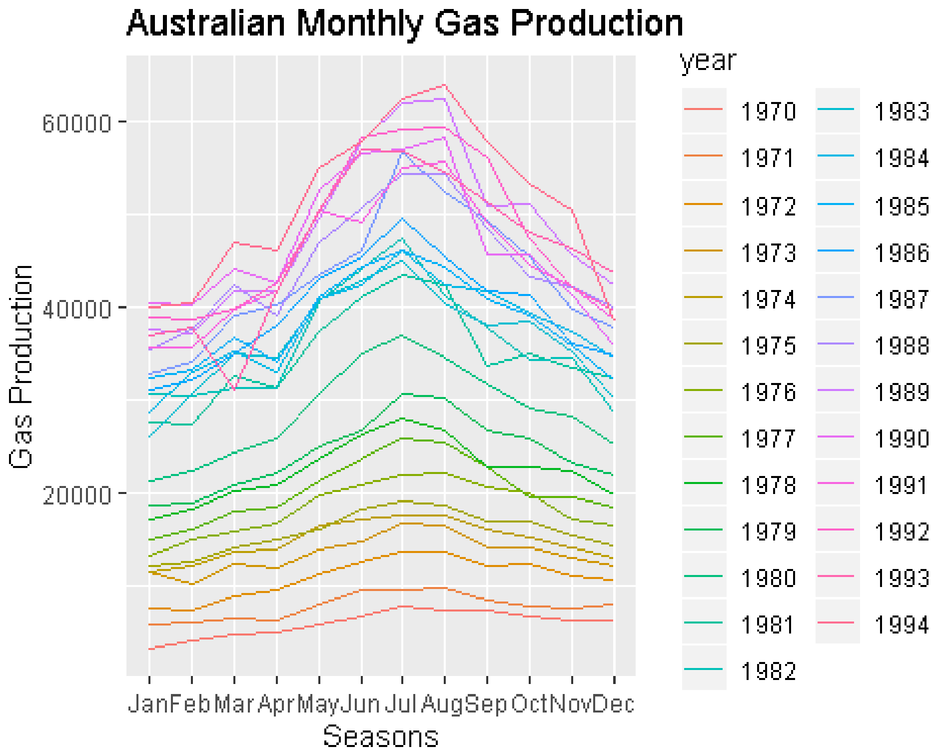


From the above time series plot, it’s evident that there is a increasing linear trend in the series that starts from 1970 - 1994 and this time series could probably be described using an additive model, since the random fluctuations in the data are roughly constant in size over time.

Let’s confirm it with the help of seasonal plots

Seasonal Plots

```
ggseasonplot(gas1,main = "Australian Monthly Gas Production",xlab= "Seasons", ylab = "Gas Production")
```



We can see from this time series that there seems to be seasonal variation in the Gas production per month, there is a peak every winter (June - August), and a drop very Summer.

Again, it seems that this time series could probably 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

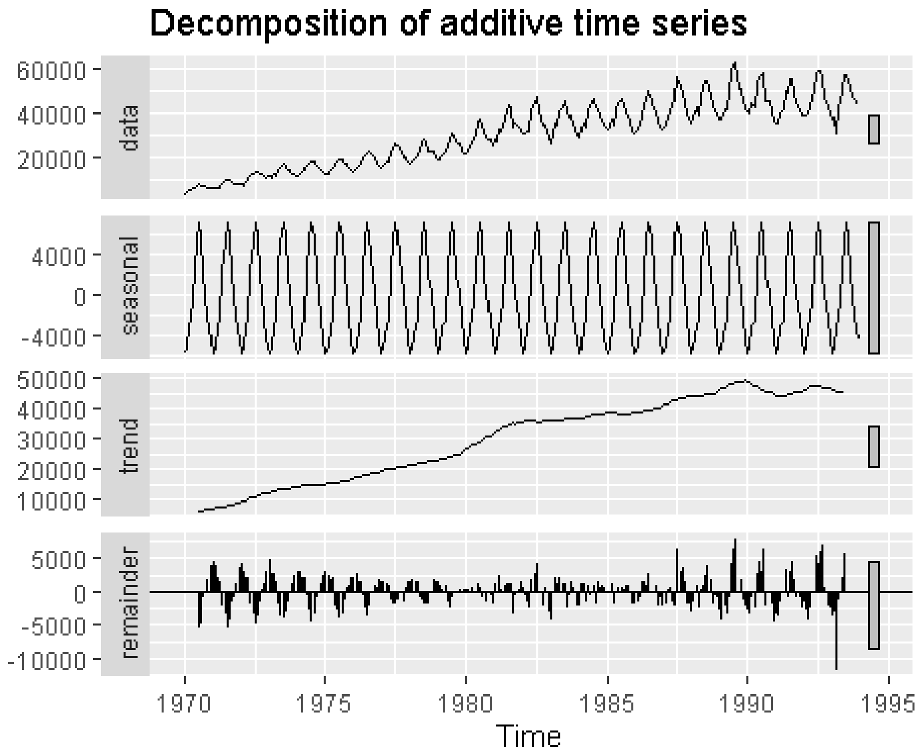
Split data into train and test sets

```
gas1.train <- window(gas1,start = c(1970,1),end = c(1993,12),frequency = 12)
gas1.test<- window(gas1,start = c(1994,1),frequency = 12)
```

Decomposing Time Series

Decomposing Seasonal Data

```
gas_Additivecomponents <- decompose(gas1.train,"additive")

autoplot(gas_Additivecomponents)
```




```
gas_Additivecomponents$seasonal

##              Jan        Feb        Mar        Apr        May        Jun
## 1970  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1971  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1972  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1973  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1974  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1975  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1976  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1977  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1978  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1979  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1980  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1981  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1982  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1983  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1984  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1985  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1986  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1987  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1988  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1989  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1990  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1991  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1992  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
## 1993  -5749.8068 -5132.1419 -3068.5477 -2602.7560  2485.6715  5272.2766
```

```
##            Jul       Aug       Sep       Oct        Nov        Dec
## 1970  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1971  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1972  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1973  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1974  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1975  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1976  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1977  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1978  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1979  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1980  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1981  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1982  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1983  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1984  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1985  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1986  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1987  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1988  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1989  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1990  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1991  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1992  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
## 1993  7284.8889 5919.4233 2010.1045  233.8273 -2141.1999 -4511.7397
```

```
gas_Additivecomponents$trend
```

```
##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 1970        NA        NA        NA        NA        NA        NA  6177.417
## 1971  7201.042  7365.625  7506.875  7596.000  7689.583  7820.000  7975.208
## 1972  9315.042  9646.458  9962.250 10307.458 10646.167 10902.958 11172.375
## 1973 12592.583 12841.250 13044.667 13196.375 13339.292 13472.583 13537.542
## 1974 14431.583 14513.958 14637.875 14769.042 14867.833 14953.458 15016.500
## 1975 15292.833 15398.042 15483.875 15595.250 15718.917 15819.167 15906.750
## 1976 17087.167 17355.667 17657.375 17928.875 18126.792 18296.000 18464.458
## 1977 19491.250 19789.833 20006.333 20077.333 20160.000 20340.167 20511.667
## 1978 21649.625 21787.542 21845.000 21981.500 22230.292 22400.458 22521.417
## 1979 23060.375 23311.417 23616.750 23906.250 24072.750 24200.208 24397.250
## 1980 26813.958 27260.750 27648.458 27999.333 28347.958 28693.583 29094.208
## 1981 32262.917 32862.667 33276.417 33598.125 34050.708 34566.208 34990.458
## 1982 35931.792 36046.000 36170.917 36315.875 36337.542 36229.667 35884.125
## 1983 36118.125 35981.125 35953.792 36142.125 36348.542 36440.458 36614.125
## 1984 36811.583 36941.542 37179.625 37362.708 37477.000 37739.500 38071.708
## 1985 38648.958 38735.625 38781.333 38740.917 38678.875 38520.583 38369.458
## 1986 38819.625 39000.750 39076.333 39196.083 39286.875 39407.875 39586.583
## 1987 40715.708 41301.542 41906.167 42398.292 42731.958 43018.125 43262.083
## 1988 44416.000 44410.875 44462.875 44331.375 44332.333 44512.958 44680.083
## 1989 46013.958 46655.292 47083.167 47507.958 47985.750 48240.292 48467.417
## 1990 49058.083 48680.458 48295.875 47850.583 47437.208 46981.417 46507.292
## 1991 44608.083 44421.250 44452.042 44548.375 44536.792 44683.583 44935.792
```

```
## 1992 46285.375 46606.458 47050.917 47457.083 47579.667 47622.667 47581.167
## 1993 46469.417 46177.667 45786.208 45625.000 45822.250 46160.250        NA
##            Aug       Sep       Oct       Nov       Dec
## 1970  6366.458  6519.917  6650.958  6797.375  7006.958
## 1971  8103.458  8250.458  8482.208  8756.583  9021.500
## 1972 11455.125 11729.500 11971.583 12172.500 12372.750
## 1973 13621.417 13749.750 13888.167 14084.042 14291.333
## 1974 15060.542 15096.083 15152.458 15173.750 15196.292
## 1975 16040.667 16213.250 16363.875 16590.583 16855.875
## 1976 18586.333 18720.917 18881.875 19018.542 19203.958
## 1977 20693.000 20872.792 21061.167 21261.250 21462.875
## 1978 22613.250 22663.458 22742.875 22851.917 22929.792
## 1979 24650.208 24948.417 25255.625 25639.875 26207.750
## 1980 29567.208 30118.708 30687.125 31200.250 31739.000
## 1981 35238.833 35311.458 35257.583 35385.625 35644.208
## 1982 35707.792 35870.792 36154.542 36305.833 36269.333
## 1983 36820.458 36927.333 36881.833 36817.208 36788.167
## 1984 38237.000 38311.375 38417.667 38470.292 38562.708
## 1985 38273.542 38152.667 38237.417 38488.917 38628.292
## 1986 39732.292 39987.375 40250.583 40354.375 40393.583
## 1987 43523.167 43801.542 43888.167 43989.458 44323.458
## 1988 44744.875 44704.875 44787.708 44999.500 45405.208
## 1989 48714.750 48939.208 49080.875 49246.542 49318.375
## 1990 46112.083 45741.208 45523.750 45393.875 44992.875
## 1991 45201.792 45324.333 45354.000 45376.875 45743.542
## 1992 47471.542 47076.708 46677.458 46651.958 46611.708
## 1993        NA        NA        NA        NA        NA
```

```
gas_Additivecomponents$random
```

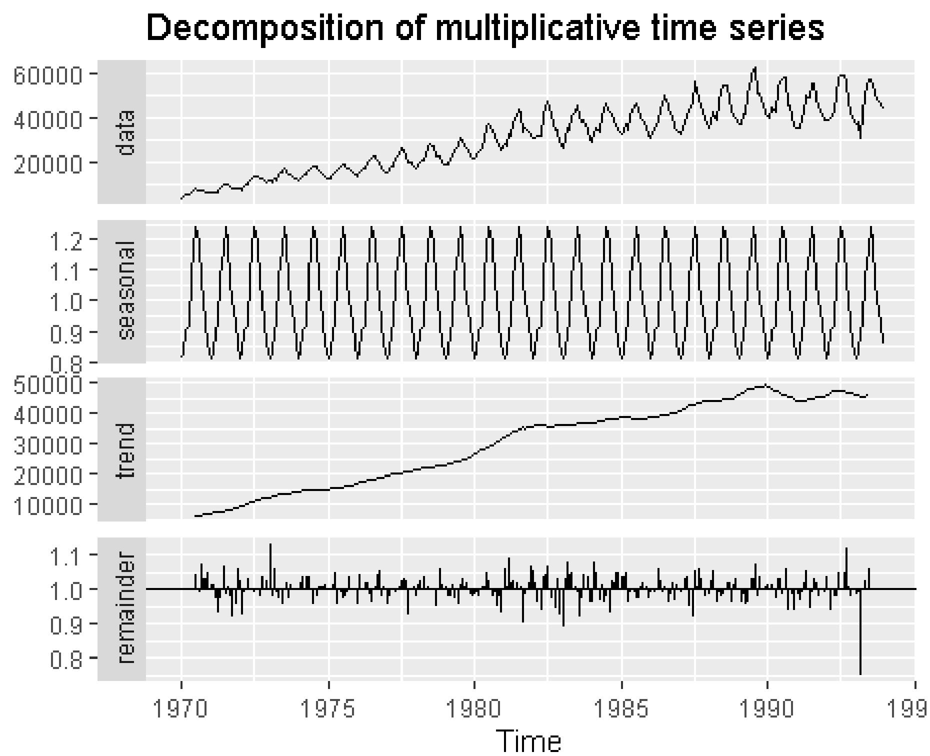
```
##               Jan          Feb          Mar          Apr          May
## 1970           NA           NA           NA           NA           NA
## 1971  4467.765097  3949.516908  2155.672705  1495.756039 -2135.254831
## 1972  4212.765097  2887.683575  2009.297705  2037.297705 -1759.838164
## 1973  4726.223430  2687.891908  2516.881039  1368.381039 -1850.963164
## 1974  3022.223430  2893.183575  2125.672705  1915.714372  -798.504831
## 1975  2810.973430  2416.100242  1725.672705  1996.506039 -2045.588164
## 1976  1922.640097  2766.475242  1386.172705  1443.881039  -793.463164
## 1977  1375.556763  1400.308575  1199.214372   996.422705 -1247.671498
## 1978  1343.181763  1628.600242  1449.547705  1524.256039  -947.963164
## 1979  1528.431763   712.725242   274.797705   908.506039 -1482.421498
## 1980   368.848430   240.391908   -76.910628   508.422705  -228.629831
## 1981  1216.890097  -306.524758  2476.131039   370.631039   922.620169
## 1982   533.015097  -513.858092 -1651.368961 -2407.118961  1768.786836
## 1983 -4230.318237  -103.983092  2132.756039  1009.631039  2145.786836
## 1984 -2260.776570  1224.600242  1182.922705 -1578.952295   834.328502
## 1985  -405.151570  -295.483092  1092.214372 -1917.160628  -144.546498
## 1986 -1830.818237 -1607.608092 -1056.785628  1515.672705  1395.453502
## 1987 -2174.901570 -1963.399758   290.381039   453.464372 -1698.629831
## 1988 -3099.193237 -1582.733092   924.672705 -2591.618961   243.995169
## 1989 -2723.151570 -4246.149758 -2236.618961 -3239.202295  -855.421498
```

```
## 1990   -2849.276570   -3253.316425   -1080.327295   -2550.827295    2638.120169
## 1991   -3266.276570   -3612.108092   -1519.493961    -184.618961    3357.536836
## 1992   -1572.568237   -2784.316425   -4190.368961   -2309.327295      79.661836
## 1993   -3660.609903   -3082.524758  -11674.660628   -1310.243961    2058.078502
##                 Jun            Jul            Aug            Sep            Oct
## 1970             NA   -5465.305556   -4762.881643   -1092.021135      -5.785628
## 1971   -3377.276570   -5546.097222   -4266.881643   -1665.562802    -855.035628
## 1972   -3434.234903   -4724.263889   -3683.548309   -1500.604469     296.589372
## 1973   -3799.859903   -4017.430556   -2953.839976   -1534.854469      35.006039
## 1974   -2886.734903   -4524.388889   -3387.964976    -912.187802     -50.285628
## 1975   -2815.443237   -4034.638889   -3223.089976   -1114.354469     496.297705
## 1976   -2585.276570   -3748.347222   -2168.756643      18.978865     853.297705
## 1977   -1758.443237   -1771.555556   -1133.423309     -78.896135   -1675.993961
## 1978   -1349.734903   -1768.305556   -1756.673309   -1787.562802    -163.702295
## 1979   -2588.484903   -1071.138889    -341.631643    -196.521135     395.547705
## 1980    1018.140097     680.902778    -984.631643    -335.812802   -1645.952295
## 1981    1221.515097    1282.652778    1239.743357   -3494.562802    -529.410628
## 1982    2631.056763    4217.986111    -317.214976      32.103865   -2033.368961
## 1983    1156.265097    1122.986111   -2352.881643    -757.437802    1492.339372
## 1984    -656.776570     741.402778   -1726.423309    1529.520531     679.506039
## 1985     557.140097     518.652778     242.035024     780.228865     797.756039
## 1986     866.848430    2696.527778    -264.714976    -192.479469     796.589372
## 1987   -2153.401570    6162.027778    2863.410024    3585.353865    1378.006039
## 1988     824.765097    2492.027778    3770.701691    1801.020531   -1796.535628
## 1989    4280.431763    6131.694444    7765.826691    -129.312802    1801.297705
## 1990    4318.306763    3065.819444    6331.493357   -2124.312802    -135.577295
## 1991    -826.859903    2845.319444    4549.785024    1723.562198   -1084.827295
## 1992    5269.056763    4168.944444    6017.035024    6901.187198     409.714372
## 1993    5544.473430             NA             NA             NA             NA
##                                        Nov                                 Dec
## 1970                           1832.824879                         3792.781401
## 1971                           1137.616546                         3644.239734
## 1972                           1209.699879                         2967.989734
## 1973                           1073.158213                         2473.406401
## 1974                           1175.449879                         2431.448068
## 1975                            968.616546                         1967.864734
## 1976                            415.658213                         1805.781401
## 1977                            506.949879                         1536.864734
## 1978                           1693.283213                         1376.948068
## 1979                           -170.675121                          233.989734
## 1980                           -754.050121                        -1979.260266
## 1981                            235.574879                         1312.531401
## 1982                            442.366546                        -3028.593599
## 1983                            631.991546                        -2042.426932
## 1984                            998.908213                          463.031401
## 1985                           -446.716787                        -1974.551932
## 1986                          -2145.175121                        -1002.843599
## 1987                          -1991.258454                        -1853.718599
## 1988                           -703.300121                         -898.468599
## 1989                          -1374.341787                        -2278.635266
## 1990                          -1948.675121                        -4465.135266
```

```
## 1991                    -1090.675121                 -2533.801932
## 1992                    -2241.758454                 -2493.968599
## 1993       NA       NA
```

The plot above shows the original time series (top), the estimated trend component (second from top), the estimated seasonal component (third from top), and the estimated irregular component (bottom).

```
###Let's try and make this multiplicative
gas_multiplicativecomponents <- decompose(gas1.train,"multiplicative")

autoplot(gas_multiplicativecomponents)
```




```
gas_multiplicativecomponents$seasonal

##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 1970 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1971 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1972 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1973 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1974 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1975 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1976 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1977 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1978 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1979 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
```

```
## 1980 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1981 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1982 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1983 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1984 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1985 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1986 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1987 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1988 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1989 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1990 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1991 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1992 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
## 1993 0.8132660 0.8304488 0.9045982 0.9183493 1.0757490 1.1666776 1.2415807
##            Aug       Sep       Oct       Nov       Dec
## 1970 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1971 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1972 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1973 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1974 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1975 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1976 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1977 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1978 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1979 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1980 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1981 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1982 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1983 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1984 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1985 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1986 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1987 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1988 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1989 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1990 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1991 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1992 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
## 1993 1.1938198 1.0653672 1.0066845 0.9282892 0.8551699
```

```
gas_multiplicativecomponents$trend
```

```
##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 1970        NA        NA        NA        NA        NA        NA  6177.417
## 1971  7201.042  7365.625  7506.875  7596.000  7689.583  7820.000  7975.208
## 1972  9315.042  9646.458  9962.250 10307.458 10646.167 10902.958 11172.375
## 1973 12592.583 12841.250 13044.667 13196.375 13339.292 13472.583 13537.542
## 1974 14431.583 14513.958 14637.875 14769.042 14867.833 14953.458 15016.500
## 1975 15292.833 15398.042 15483.875 15595.250 15718.917 15819.167 15906.750
## 1976 17087.167 17355.667 17657.375 17928.875 18126.792 18296.000 18464.458
## 1977 19491.250 19789.833 20006.333 20077.333 20160.000 20340.167 20511.667
```

```
## 1978 21649.625 21787.542 21845.000 21981.500 22230.292 22400.458 22521.417
## 1979 23060.375 23311.417 23616.750 23906.250 24072.750 24200.208 24397.250
## 1980 26813.958 27260.750 27648.458 27999.333 28347.958 28693.583 29094.208
## 1981 32262.917 32862.667 33276.417 33598.125 34050.708 34566.208 34990.458
## 1982 35931.792 36046.000 36170.917 36315.875 36337.542 36229.667 35884.125
## 1983 36118.125 35981.125 35953.792 36142.125 36348.542 36440.458 36614.125
## 1984 36811.583 36941.542 37179.625 37362.708 37477.000 37739.500 38071.708
## 1985 38648.958 38735.625 38781.333 38740.917 38678.875 38520.583 38369.458
## 1986 38819.625 39000.750 39076.333 39196.083 39286.875 39407.875 39586.583
## 1987 40715.708 41301.542 41906.167 42398.292 42731.958 43018.125 43262.083
## 1988 44416.000 44410.875 44462.875 44331.375 44332.333 44512.958 44680.083
## 1989 46013.958 46655.292 47083.167 47507.958 47985.750 48240.292 48467.417
## 1990 49058.083 48680.458 48295.875 47850.583 47437.208 46981.417 46507.292
## 1991 44608.083 44421.250 44452.042 44548.375 44536.792 44683.583 44935.792
## 1992 46285.375 46606.458 47050.917 47457.083 47579.667 47622.667 47581.167
## 1993 46469.417 46177.667 45786.208 45625.000 45822.250 46160.250        NA
##            Aug       Sep       Oct       Nov       Dec
## 1970  6366.458  6519.917  6650.958  6797.375  7006.958
## 1971  8103.458  8250.458  8482.208  8756.583  9021.500
## 1972 11455.125 11729.500 11971.583 12172.500 12372.750
## 1973 13621.417 13749.750 13888.167 14084.042 14291.333
## 1974 15060.542 15096.083 15152.458 15173.750 15196.292
## 1975 16040.667 16213.250 16363.875 16590.583 16855.875
## 1976 18586.333 18720.917 18881.875 19018.542 19203.958
## 1977 20693.000 20872.792 21061.167 21261.250 21462.875
## 1978 22613.250 22663.458 22742.875 22851.917 22929.792
## 1979 24650.208 24948.417 25255.625 25639.875 26207.750
## 1980 29567.208 30118.708 30687.125 31200.250 31739.000
## 1981 35238.833 35311.458 35257.583 35385.625 35644.208
## 1982 35707.792 35870.792 36154.542 36305.833 36269.333
## 1983 36820.458 36927.333 36881.833 36817.208 36788.167
## 1984 38237.000 38311.375 38417.667 38470.292 38562.708
## 1985 38273.542 38152.667 38237.417 38488.917 38628.292
## 1986 39732.292 39987.375 40250.583 40354.375 40393.583
## 1987 43523.167 43801.542 43888.167 43989.458 44323.458
## 1988 44744.875 44704.875 44787.708 44999.500 45405.208
## 1989 48714.750 48939.208 49080.875 49246.542 49318.375
## 1990 46112.083 45741.208 45523.750 45393.875 44992.875
## 1991 45201.792 45324.333 45354.000 45376.875 45743.542
## 1992 47471.542 47076.708 46677.458 46651.958 46611.708
## 1993        NA        NA        NA        NA        NA
```

```
gas_multiplicativecomponents$random
```

```
##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 1970        NA        NA        NA        NA        NA        NA 1.0426661
## 1971 1.0106957 1.0108269 0.9710331 0.9302184 0.9719464 1.0648420 0.9810273
## 1972 1.0267165 0.9239923 0.9879233 1.0291736 0.9929620 1.0016322 0.9900223
## 1973 1.1296617 0.9749625 1.0587125 0.9870547 0.9738163 0.9508110 0.9998245
## 1974 0.9972125 1.0184103 1.0342566 1.0382552 1.0350720 0.9938745 0.9534871
## 1975 0.9933151 0.9917665 1.0095894 1.0465799 0.9556105 0.9902542 0.9699986
```

```
## 1976 0.9542032 1.0400339 1.0001359 1.0185260 1.0163654 0.9830159 0.9596898
## 1977 0.9536595 0.9770943 1.0021720 1.0017895 0.9866695 1.0052078 1.0219151
## 1978 0.9793318 1.0105322 1.0235339 1.0354840 0.9938859 1.0072276 1.0027124
## 1979 1.0045207 0.9758800 0.9746921 1.0117385 0.9683260 0.9521907 1.0105590
## 1980 0.9828550 0.9880887 0.9796989 1.0074604 1.0035977 1.0450422 1.0259446
## 1981 1.0568507 1.0048822 1.0857828 1.0165674 1.0226313 1.0181607 1.0026361
## 1982 1.0510879 1.0155555 0.9612121 0.9386921 1.0384222 1.0441148 1.0636088
## 1983 0.8898454 1.0289326 1.0766906 1.0409117 1.0480307 1.0083439 0.9903784
## 1984 0.9620341 1.0767957 1.0493979 0.9670373 1.0119347 0.9619615 0.9752249
## 1985 1.0337911 1.0354405 1.0491278 0.9618667 0.9858501 0.9868472 0.9692314
## 1986 0.9894941 0.9960750 0.9887582 1.0587099 1.0214180 0.9906629 1.0085059
## 1987 0.9902847 0.9972940 1.0321766 1.0337103 0.9467061 0.9192783 1.0557707
## 1988 0.9846348 1.0220993 1.0521609 0.9613210 0.9868221 0.9745385 0.9816683
## 1989 1.0031910 0.9621155 0.9809035 0.9550091 0.9611663 1.0268676 1.0283798
## 1990 1.0140794 0.9967441 1.0104980 0.9716329 1.0299913 1.0321066 0.9846810
## 1991 0.9810841 0.9671297 0.9913647 1.0207776 1.0515460 0.9424082 0.9869978
## 1992 1.0350850 0.9996312 0.9349147 0.9762017 0.9797049 1.0468626 0.9993084
## 1993 0.9806045 0.9899556 0.7495029 0.9955205 1.0217629 1.0579876        NA
##              Aug       Sep       Oct       Nov       Dec
## 1970   0.9898158 1.0708159 1.0274193 1.0283791 1.0493747
## 1971   1.0084692 0.9778416 0.9206096 0.9537879 1.0569138
## 1972   1.0011441 0.9794159 1.0373720 0.9948139 1.0234573
## 1973   1.0200156 0.9710871 1.0125884 0.9955588 1.0025760
## 1974   0.9784437 1.0069097 1.0053925 1.0086877 1.0092794
## 1975   0.9784505 0.9905017 1.0376817 1.0011130 0.9928797
## 1976   1.0066821 1.0403793 1.0505526 0.9795121 1.0045884
## 1977   1.0313835 1.0254894 0.9253394 0.9944474 1.0072787
## 1978   0.9918453 0.9478605 0.9964228 1.0561355 1.0094923
## 1979   1.0271883 1.0068767 1.0181146 0.9801179 0.9784903
## 1980   0.9774514 0.9908225 0.9476486 0.9772862 0.9302107
## 1981   1.0078249 0.8991839 0.9850320 1.0192372 1.0644038
## 1982   0.9690662 0.9920827 0.9439168 1.0268435 0.9262507
## 1983   0.9187845 0.9704846 1.0398517 1.0330919 0.9610259
## 1984   0.9295023 1.0253657 1.0169758 1.0452639 1.0465871
## 1985   0.9724959 1.0072922 1.0201591 1.0048183 0.9730048
## 1986   0.9568615 0.9813095 1.0187899 0.9628267 1.0097160
## 1987   1.0066819 1.0585512 1.0298419 0.9760514 1.0014224
## 1988   1.0190516 1.0186636 0.9587001 1.0091555 1.0300247
## 1989   1.0729645 0.9747167 1.0345493 1.0003493 1.0083558
## 1990   1.0601910 0.9362999 0.9955038 0.9801929 0.9360506
## 1991   1.0316553 1.0159658 0.9747210 1.0005255 0.9892506
## 1992   1.0482692 1.1163222 1.0070553 0.9760426 0.9936046
## 1993 NA NA NA NA NA
```

Still the same pattern, hence series is additive.

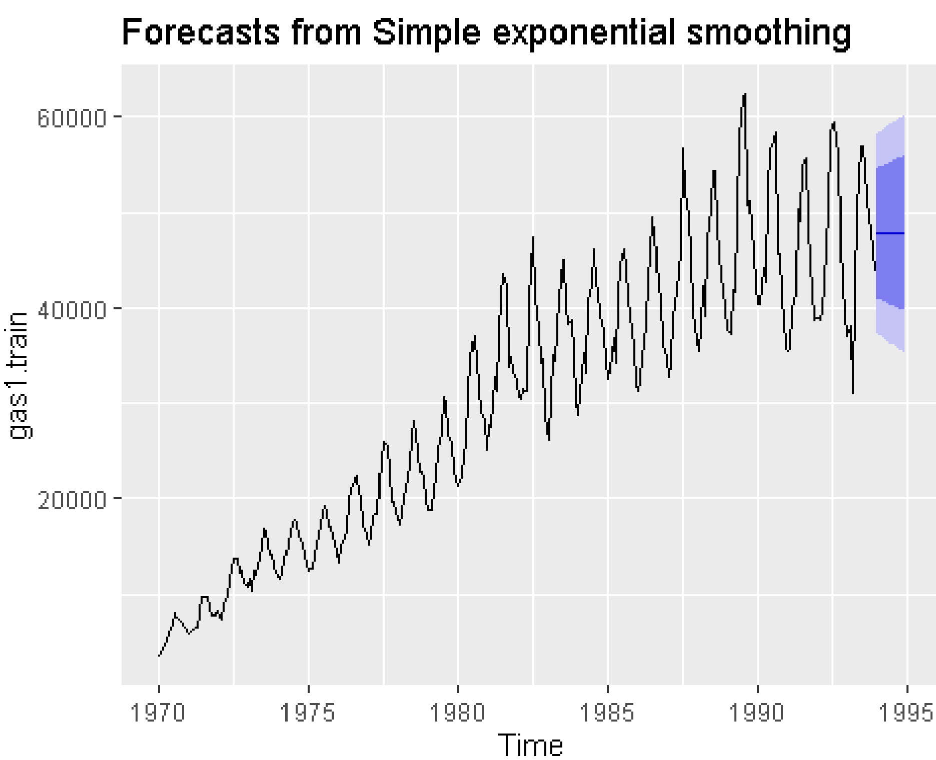
Components Present and Periodicity of the Dataset

1. Trend - Linear Trend as the gas production increases over time
2. Seasonality - Based on seasons (Summer, Winter)
3. Monthly periodicity from Jan 1970 to Aug 1994

Simple Exponential Smoothing on gas production

Exponential smoothing are used to make short-term forecasts for time series data. If we have a time series that can be described using an additive model with constant level and no seasonality, we can use simple exponential smoothing to make short-term forecasts but we have additive model with seasonality, so let's see how the forecasts turns out.

```
ses.gas <- ses(gas1.train, alpha = 0.2, h=12)
autoplot(ses.gas)
```



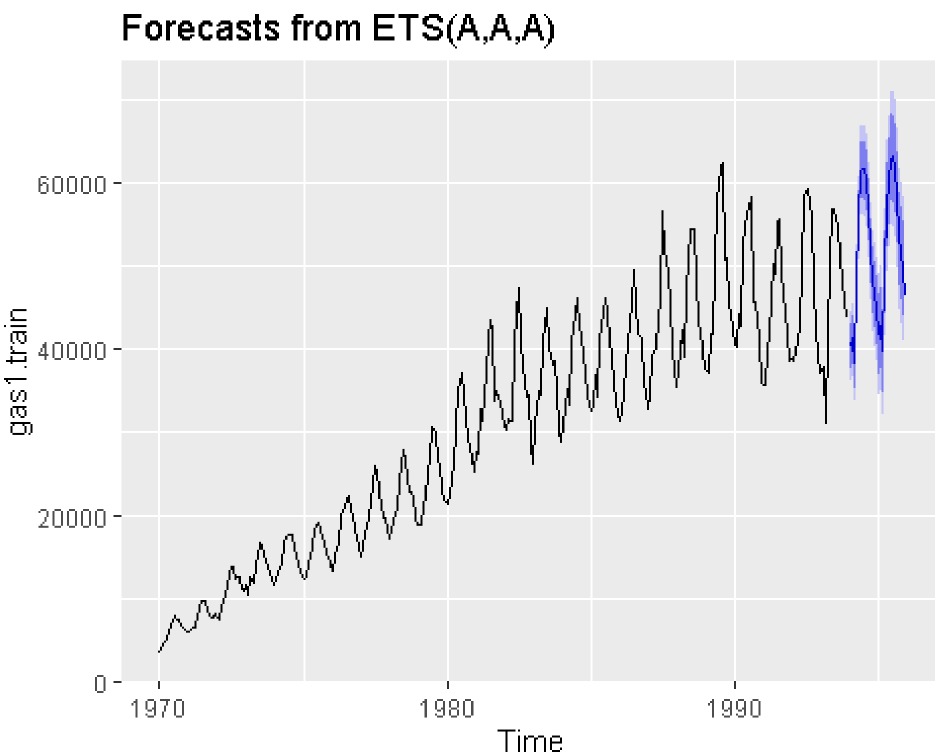
It Gives the forecast for a year, a 80% prediction interval for the forecast, and a 95% prediction interval for the forecast. For example, the forecasted gas production for 1994 is about 47804.79 units, with a 95% prediction interval of (37442.12, 58167.47).

```
accuracy(ses.gas, gas1.test)

##                     ME     RMSE      MAE      MPE     MAPE     MASE
## Training set  737.2333 5268.785 3999.333 1.524847 12.93317 1.517094
## Test set     3231.0385 8979.490 7750.868 3.656030 14.93589 2.940188
```

```
## ACF1 Theil's U
## Training set 0.7686942 NA
## Test set 0.6184060 1.415467
```

Holt winters Model

```
gas.hw <- ets(gas1.train, model = "AAA")
autoplot(forecast(gas.hw))
```




```
summary(gas.hw)

## ETS(A,A,A)
##
## Call:
## ets(y = gas1.train, model = "AAA")
##
## Smoothing parameters:
## alpha = 0.3409
## beta = 1e-04
## gamma = 0.5937
##
## Initial states:
## l = 6253.2029
## b = 119.5505
## s = -4511.743 -2141.073 234.0436 2010.202 5919.329 7284.466
```

```
##                      5272.425   2485.985   -2602.642   -3068.389   -5131.983   -5750.62
##
##                                          sigma:                                2109.357
##
##                            AIC                      AICc                            BIC
##                         6057.255                 6059.521                       6119.525
##
##                 Training               set                error                measures:
##                          ME        RMSE         MAE          MPE        MAPE       MASE
##    Training    set    78.6797    2049.926    1551.876    0.4622781    7.506005   0.5886835
##                                                                                     ACF1
## Training set 0.2737321
```

```r
# forecast the next 12 months
gas.hw.forecast <- forecast(gas.hw, h = 12)
accuracy(gas.hw.forecast, gas1.test)
```

```
##                          ME        RMSE         MAE          MPE        MAPE       MASE
##   Training    set     78.6797    2049.926    1551.876   0.46227811   7.506005   0.5886835
##   Test    set          228.9429   3568.067    2541.069   0.07409708   5.297855   0.9639204
##                                                                      ACF1     Theil's    U
##       Training       set            0.2737321                                          NA
## Test set     -0.1659001 0.7087554
```

Stationarity

Let’s look whether the time series are stationary or not by using ADF Test for Level Stationarity. Null Hypothesis - unit root is present in a time series sample. Alternate Hypothesis - series is stationary

```r
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 3.6.2
```

```r
adf.test(gas1, alternative = "stationary")
```

```
##   Warning   in   adf.test(gas1,   alternative   =   "stationary"):   p-value   smaller   than
## printed p-value

##
##                               Augmented                 Dickey-Fuller                  Test
##
##                           data:                                                        gas1
##      Dickey-Fuller     =     -4.4805,     Lag     order     =     6,     p-value     =     0.01
## alternative hypothesis: stationary
```

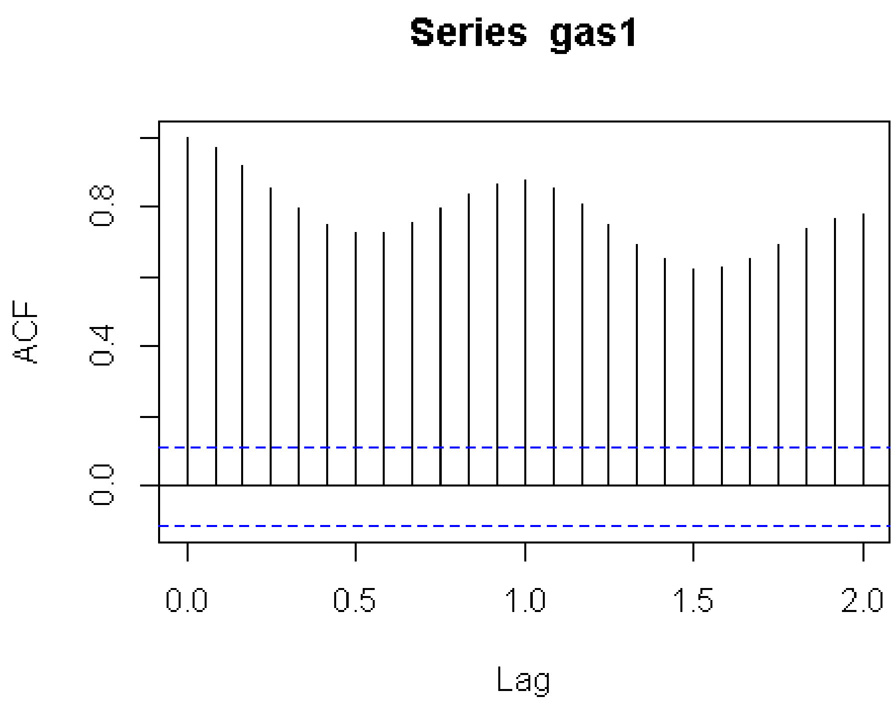
Since the p-value = 0.01, we reject null hypothesis i.e series is non-stationary.which means the series is stationary. In other words, correlation exist in the dataset.

Though exponential smoothing does well in such cases but if we want to make prediction intervals for forecasts made using exponential smoothing methods, the prediction intervals require that the forecast errors are uncorrelated and are normally distributed with mean zero and constant variance.

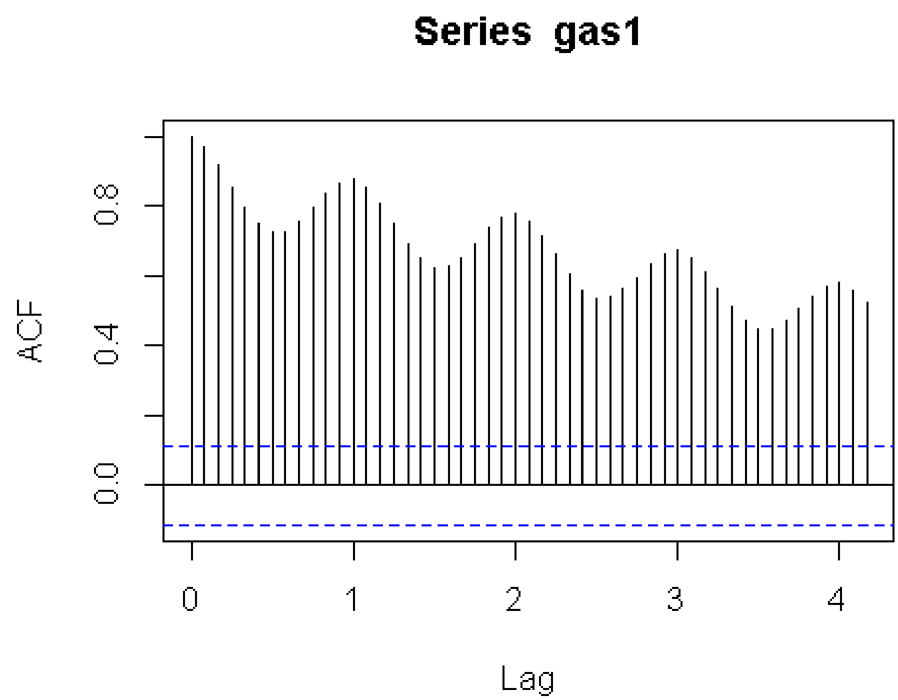
ACF and PACF plots

We use ACF and PACF to determine if corrections(AR and MA terms) are required before building an ARIMA model.

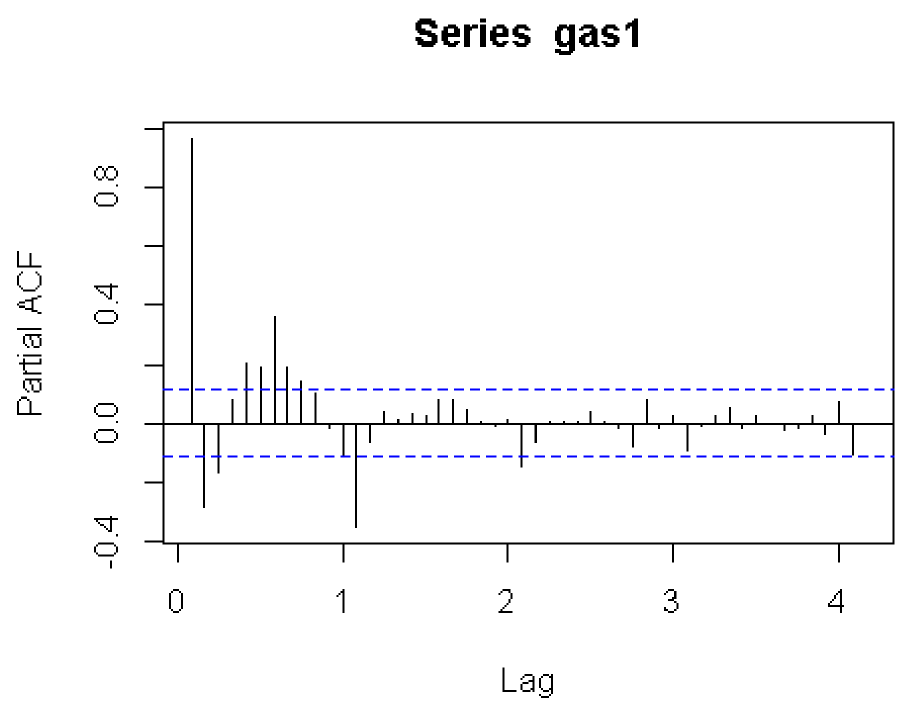
```
acf(gas1)
```




```
acf(gas1, lag.max = 50)
```




```
pacf(gas1, lag.max = 50)
```



There are significant autocorrelations with many lags in our gas time series, as shown by the ACF plot.

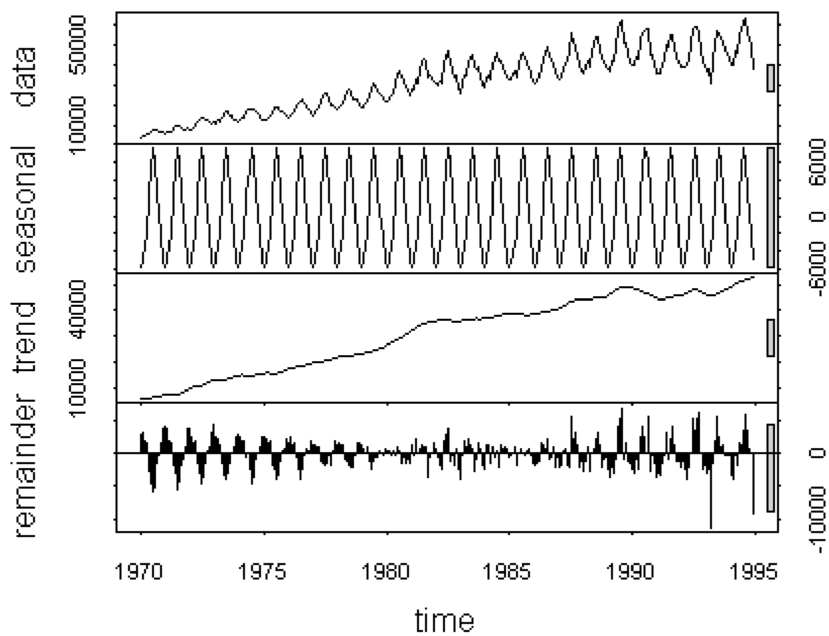
PACF plot shows that there could be monthly seasonality.

Seasonally Adjusting

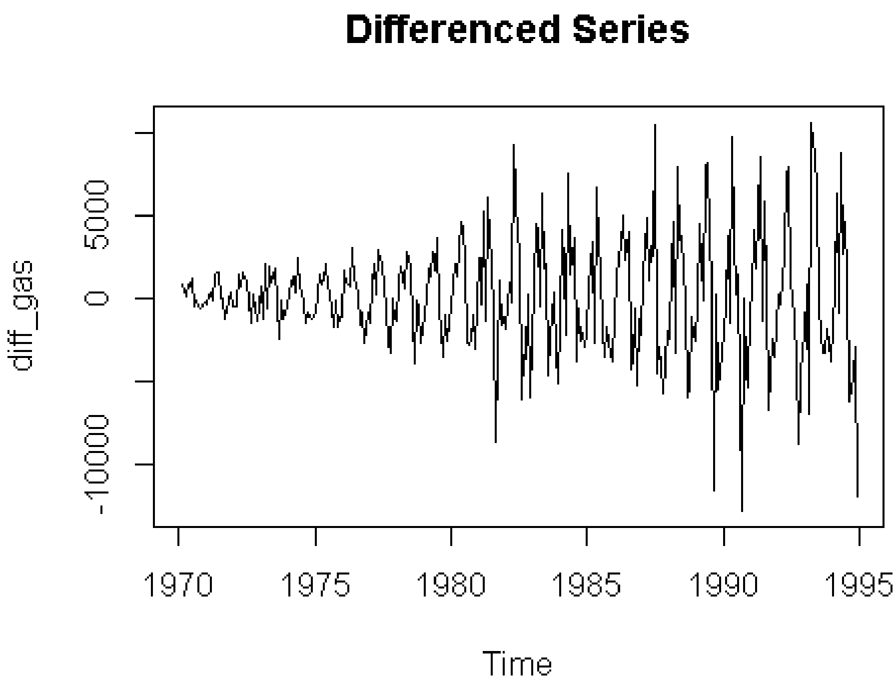
Since we have a seasonal time series that can be described using an additive model, we need to seasonally adjust the time series by estimating the seasonal component, and subtracting the estimated seasonal component from the original time series.

We can do this using the estimate of the seasonal component calculated by the "decompose ()" function.

```
decomp = stl(gas1, s.window = "periodic")
deseasonal_gas=seasadj(decomp)
plot(decomp)
```




```
diff_gas = diff(gas1, differences = 1)
plot(diff_gas,main = "Differenced Series")
```



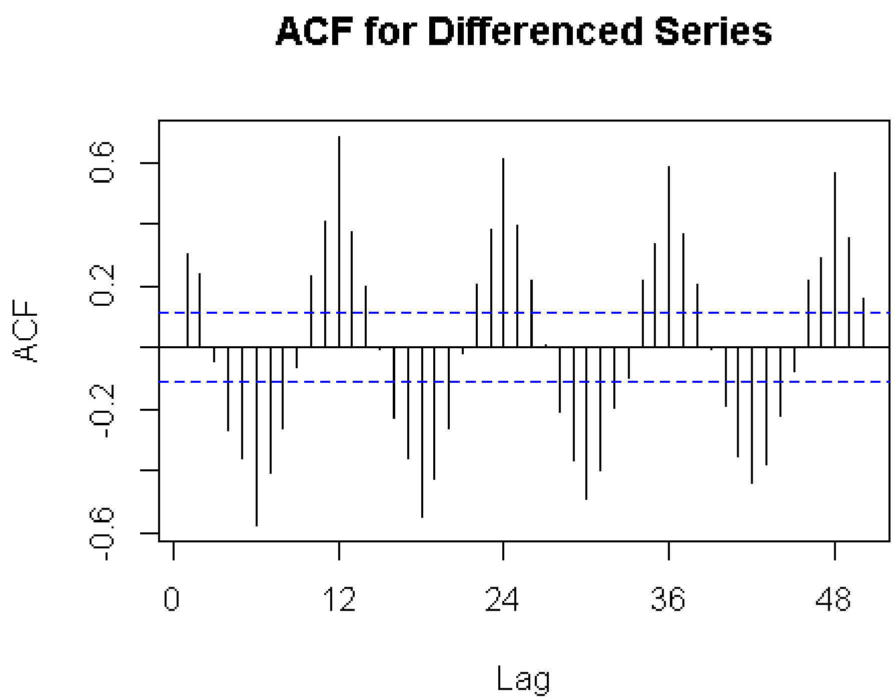

```
adf.test(diff_gas, alternative = "stationary")

## Warning in adf.test(diff_gas, alternative = "stationary"): p-value smaller
## than printed p-value

##
## Augmented Dickey-Fuller Test
##
## data: diff_gas
## Dickey-Fuller = -15.847, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

From the plot, we can see that the time series of first differences appears to be stationary in mean and variance, By, taking the time series of first differences, we have removed the trend component of the time series.

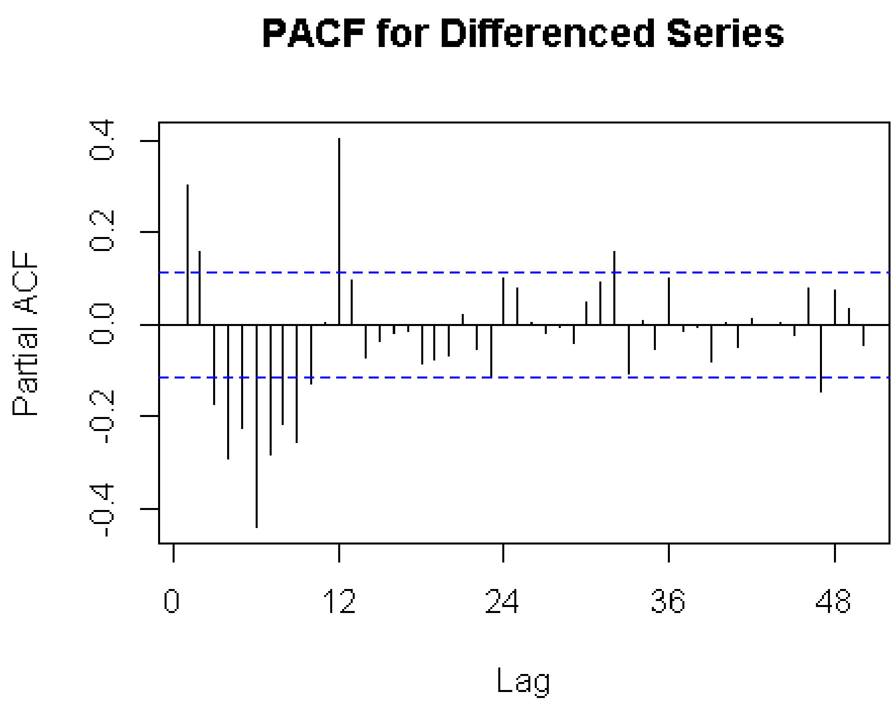
Dickey- Fuller = -15.847 has increased after deseasonality, this means that we will be more accurate than before. Now, let’s check the Acf and Pacf for the Differenced time series. We can now examine whether there are correlations between successive terms of this irregular component.

Acf and Pacf for difference time series

```
Acf(diff_gas,lag.max = 50, main='ACF for Differenced Series')
```




```
Pacf(diff_gas,lag.max = 50, main='PACF for Differenced Series')
```



From the ACF plot, there is a cut off after lag 2. This implies that q=2. PACF cuts off after lag 2. Hence p=2.

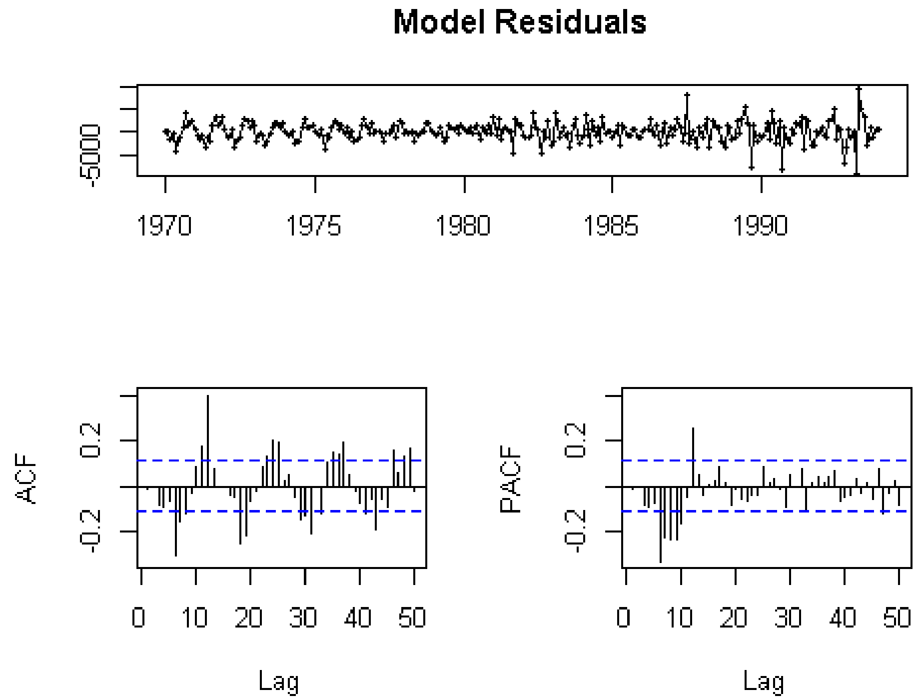
Splitting into training and test sets

Splitting into training and test sets

```
gasTstrain = window(deseasonal_gas, start=c(1970,1), end=c(1993,12),frequency = 12)
gasTStest= window(deseasonal_gas, start=c(1994,1))
```

After multiple Trial and error with p,d,q the following ordered values (p=2,d=1,q=2) are taken as best fit for arima.

```
gasarima <- arima(gasTstrain,order = c(0,1,2))
gasarima

##
## Call:
## arima(x = gasTstrain, order = c(0, 1, 2))
##
## Coefficients:
##           ma1     ma2
##       -0.0647  0.0957
## s.e.   0.0588  0.0611
##
## sigma^2 estimated as 4820422:  log likelihood = -2615.48,  aic = 5236.96

tsdisplay(residuals(gasarima), lag.max=50, main='Model Residuals')
```

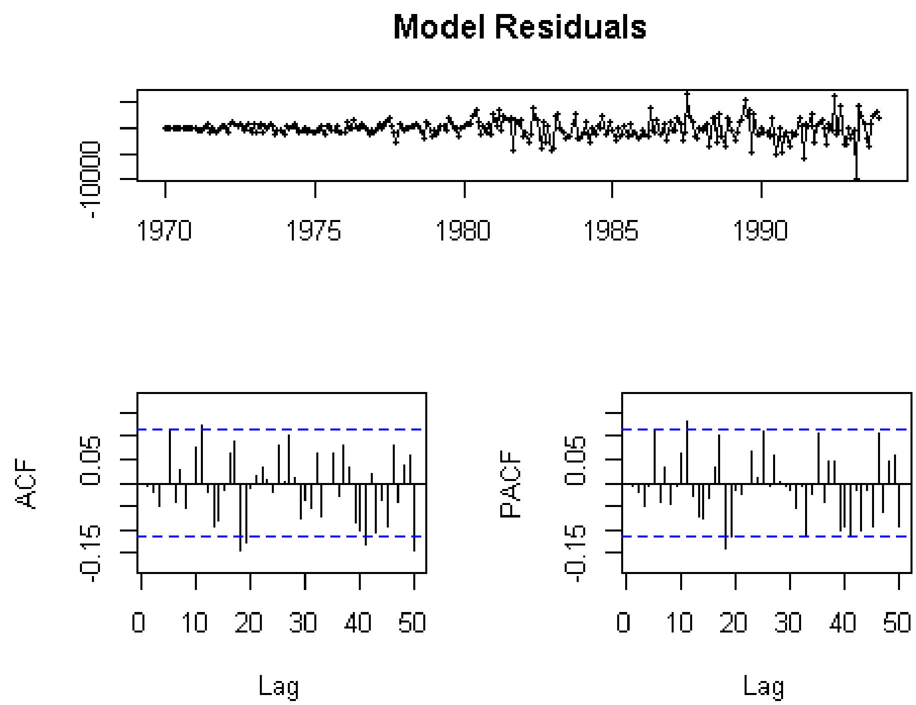


There are no significant autocorrelations present. If, the model is not correctly specified, that will usually be reflected in residuals in the form of trends, skeweness, or any other patterns not captured by the model.

Ideally, residuals should look like white noise, meaning they are normally distributed. Residuals plots show a smaller error range, more or less centered around 0.

Fitting an Auto-ARIMA Model

```
Fit1 <- auto.arima(gasTstrain, seasonal = TRUE)
Fit1

## Series: gasTstrain
## ARIMA(2,1,1)(0,1,2)[12]
##
## Coefficients:
## ar1 ar2 ma1 sma1 sma2
## 0.5017 0.2057 -0.9583 -0.4404 -0.1236
## s.e. 0.0738 0.0722 0.0426 0.0676 0.0639
##
## sigma^2 estimated as 3535011: log likelihood=-2463.67
## AIC=4939.33 AICc=4939.64 BIC=4961.03

tsdisplay(residuals(Fit1), lag.max=50, main='Model Residuals')
```



Auto ARIMA also fits the same p and q parameters for the model, but has a slightly lower AIC.

Ljung box test

H0: Residuals are independent

Ha: Residuals are not independent

```
library(stats)
Box.test(gasarima$residuals)

##
##                          Box-Pierce test
##
## data:  gasarima$residuals
## X-squared = 0.037439, df = 1, p-value = 0.8466
```

Here, p-value = 0.84 which is greater than 0.05. i.e., Residuals are independent.

Forecasting Using an Manual- ARIMA Model

After the best candidate ARIMA(p = 2,d = 1,q = 2) model for our time series data is selected , we can estimate the parameters of that ARIMA model, and use that as a predictive model for making forecasts for future values of our time series.

```
Manual_gasarima_Forecast <- forecast(gasarima, h = 12)
accuracy(Manual_gasarima_Forecast,gasTStest)
```

```
##                       ME     RMSE      MAE        MPE     MAPE      MASE
## Training set   133.2453 2191.731 1611.161 -0.9451071 8.995565 0.6111723
## Test set      2327.1572 4789.675 4204.189  3.8888709 8.100992 1.5948027
##                    ACF1 Theil's U
## Training set -0.01140157        NA
## Test set      0.45842534  1.316523
```

Forecasting Using an Auto-ARIMA Model

```
Auto_gasarima_Forecast <- forecast(Fit1, h = 12)
accuracy(Auto_gasarima_Forecast,gasTStest)
```

```
##                        ME     RMSE      MAE        MPE     MAPE      MASE
## Training set   -79.12886 1820.459 1260.205 -0.3450597 4.192233 0.4780418
## Test set      2370.16919 4226.614 3487.633  4.2382629 6.782591 1.3229869
##                     ACF1 Theil's U
## Training set -0.005840643        NA
## Test set     -0.049336280  1.246271
```

Model Accuracy Report

Manual Arima								
Manual Arima	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**	**Theil's U**
Training set	133.2453	2191.731	1611.161	-0.94511	8.995565	0.611172	-0.0114	NA
Test set	2327.157	4789.675	4204.189	3.888871	8.100992	1.594803	0.458425	1.316523
Auto Arima								
Auto Arima	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**	**Theil's U**
Training set	-79.129	1820.46	1260.21	-0.3451	4.19223	0.47804	-0.0058	NA
Test set	2370.17	4226.61	3487.63	4.23826	6.78259	1.32299	-0.0493	1.24627
Exponential Smoothing								
Exponential Smoothing	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**	**Theil's U**
Training set	737.233	5268.79	3999.33	1.52485	12.9332	1.51709	0.76869	NA
Test set	3231.04	8979.49	7750.87	3.65603	14.9359	2.94019	0.61841	1.41547
Holt's Winter Model								
Holt's Winter Model	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**	**Theil's U**
Training set	78.6797	2049.93	1551.88	0.46228	7.50601	0.58868	0.27373	NA
Test set	228.943	3568.07	2541.07	0.0741	5.29786	0.96392	-0.1659	0.70876

- We have forecasted the Australian gas production by using the following techniques
    - Exponential Smoothing technique
    - Holt's Winter Model
    - Manual Arima
    - Auto Arima
- All these techniques have done well in forecasting the gas production for the year 1994 based on the past data (1970-1993).
- Except, Exponential smoothing all these techniques have accuracy more than 90% in forecasting
- Both Auto arima and Holt's Winter Model done extremely well in both training and test dataset
- But **Auto Arima** is the only model which forecasts with the accuracy of 96% in training dataset and 94% in Test Dataset

+

Conclusion:

In this project we analyzed Australian Monthly Gas production dataset "Gas" which is available in the package "Forecast".

Monthly gas production of Australia between year 1956-1995 is released by Australian Bureau of Statistics which is in time series format. The data is in Univariate time series format

During our initial exploration of the dataset we figured out that there was a sudden rise in production of gas from 1970 Jan and the trend keeps on increase. The 1970 Oil crisis could be one of the major reasons behind this radical change. We figured that from 1970 Jan the oil production has increased by double and this positive trend have been followed till the end (1995).

Therefore, we removed the data before 1970 Jan as it won't help much in forecasting the future. We then identified the key components of the present in the time series and founded out during every year's Australian winter has produced a ton of gas than in the summer. This clearly showed the presence of the seasonal and trend component in the series

Then we decomposed the time series and figured out it follows additive series. After Decomposition we did Simple Exponential Smoothing and Holt's Winter model on both training and test dataset.

We understood that correlation can be one of the reasons behind the lack of model's accuracy. So, we thought of fitting ARIMA model on this time series and check out how it performs.

In order to do ARIMA model we need to stationarize and deseasonlize the Time series so that we will have constant mean, Variance, Covariance which is basic requirement to perform ARIMA model.

After trial and error method we figured out the optimal ARIMA (p,d,q) order and build manual arima model on the same. We then ensured that Auto ARIMA also had the same order.

In the end we figured out that **Auto Arima** model has the **best accuracy** in forecasting the Monthly Production than the other 3 Models.