Time Series Forecasting

Managerial Report by Chetan Suvarna

Table of Content	Page No.

2 Project Objectives

The objective of the report is to explore the dataset gas in R using forecast package and generate insights about the data set. This exploration report will consists of the following scenarios:

- Exploratory Data Analysis of the data
- Observations from the dataset
- Inspecting stationarity in the Time Series by performing necessary test
- Build appropriate models using the data
- Check the accuracy of all the models on the test dataset
- Forecasting the data for the next 12 months using the most accurate model

3 Exploratory Data Analysis – Step by step approach

A Typical Data exploration activity consists of the following steps:

1. Environment Set up and Data Import
2. Variable Identification
3. Visualisation Plots

We shall follow these steps in exploring the provided dataset.

3.1 Environment Set up and Data Import

3.1.1 Set up working Directory

Setting a working directory on starting of the R session makes importing and exporting data files and code files easier. Basically, working directory is the location/ folder on the PC where you have the data, codes etc. related to the project.
Please refer Appendix A for Source Code.

3.1.2 Import and Read the Dataset

The dataset 'gas' was imported in r using 'forecast' package. Please refer Appendix A for Source Code.

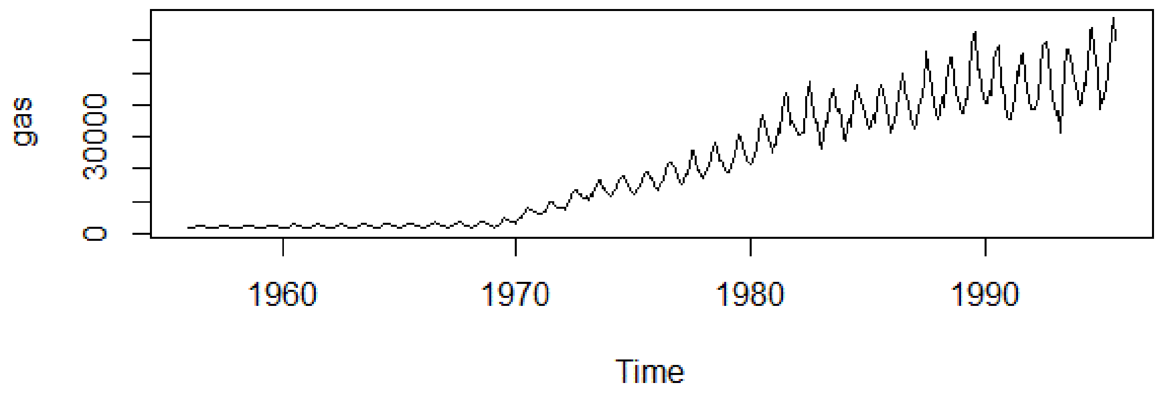
3.2 Variable Identification

The length and breadth of the gas data was examined and the names of the data were pulled from the dataset. The data consisted of 1 time series variable consisting of 476 data points depicting the gas production from 1970 to 1996. The string type of the data was also verified by using the str() function.

3.3 Visualisation Plots

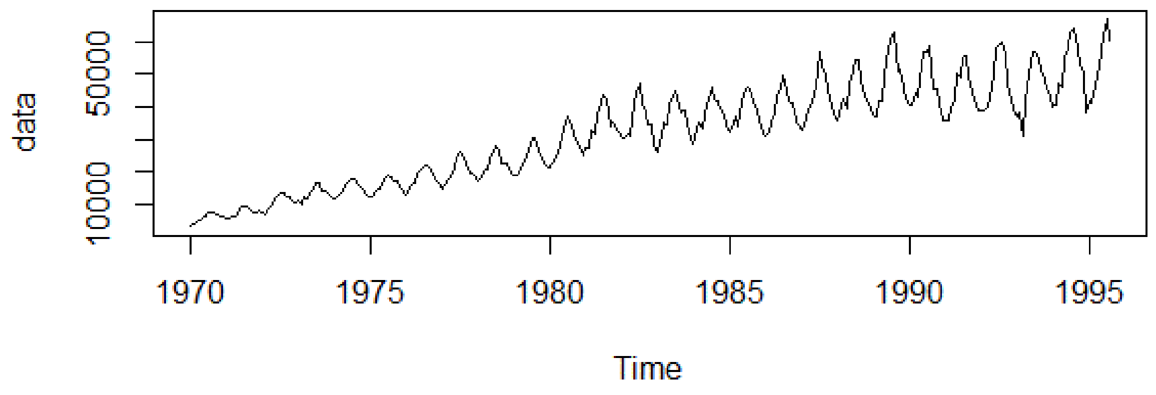
Time Series plot of the gas dataset

Below is the plot for the gas dataset which has been obtained from forecast package.

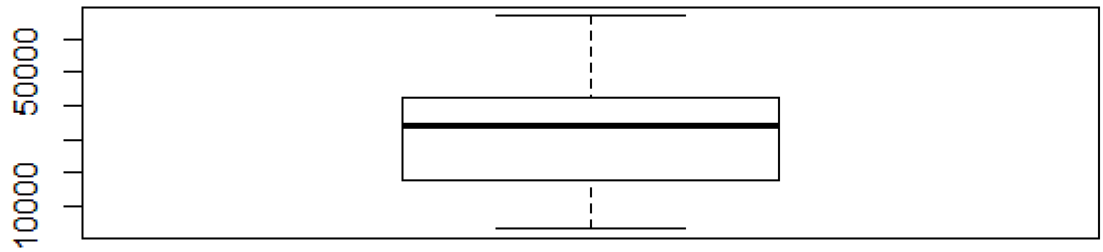


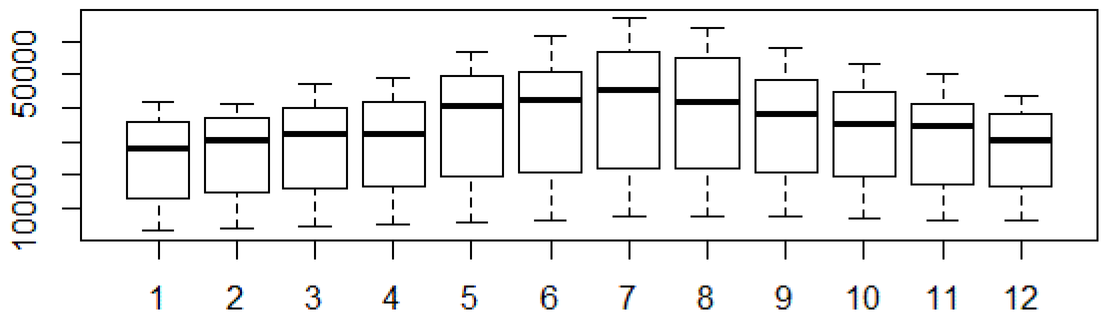
Refining the dataset

The gas dataset has been trimmed until 1970 as there were no changes in the plot of the data. The same has been plotted below.



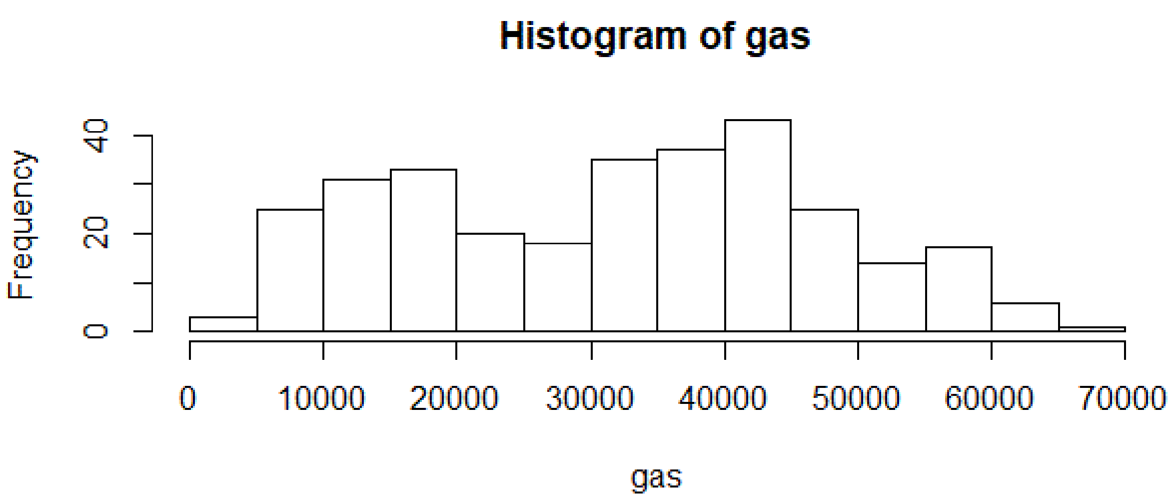
Boxplot





- The above plots signify the overall and monthly distribution of the gas production.
- The maximum production was recorded in the month of July

Histogram

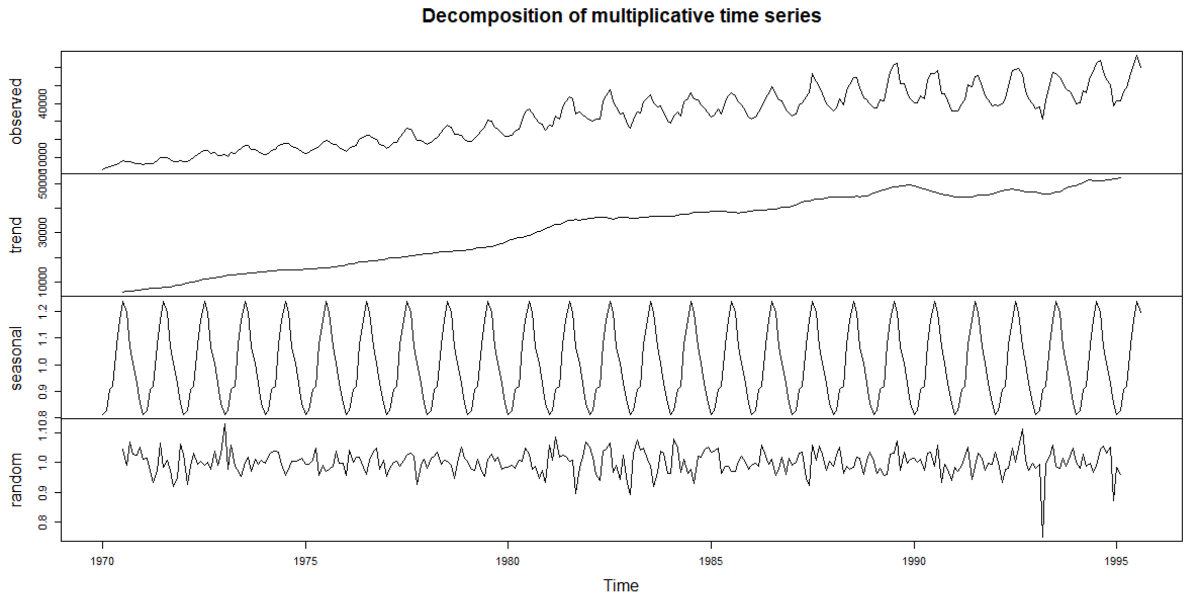


- The above is a histogram plot of the gas production

- The maximum monthly gas production for the data varied between 40000-45000

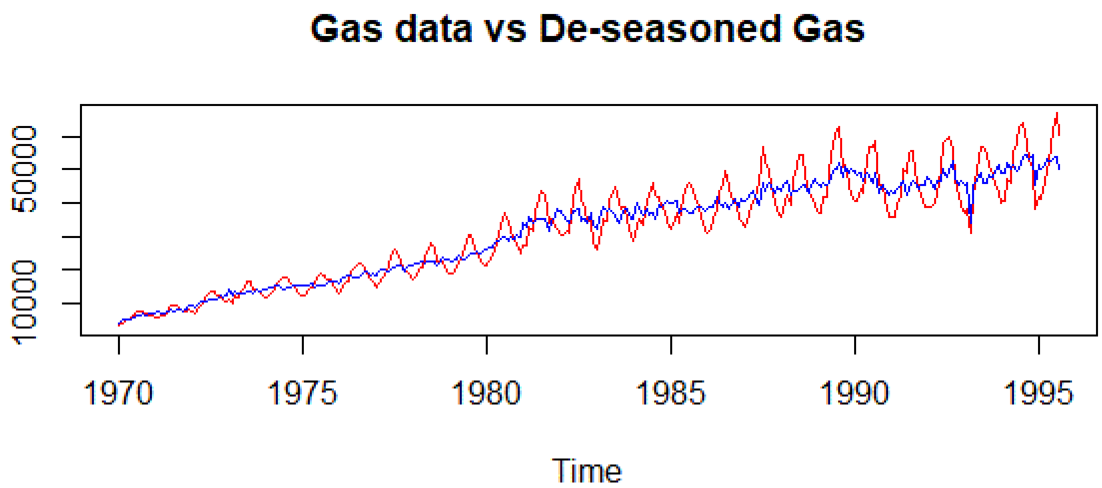
4. Decomposing the dataset

The dataset was decomposed and the graph was plotted as below



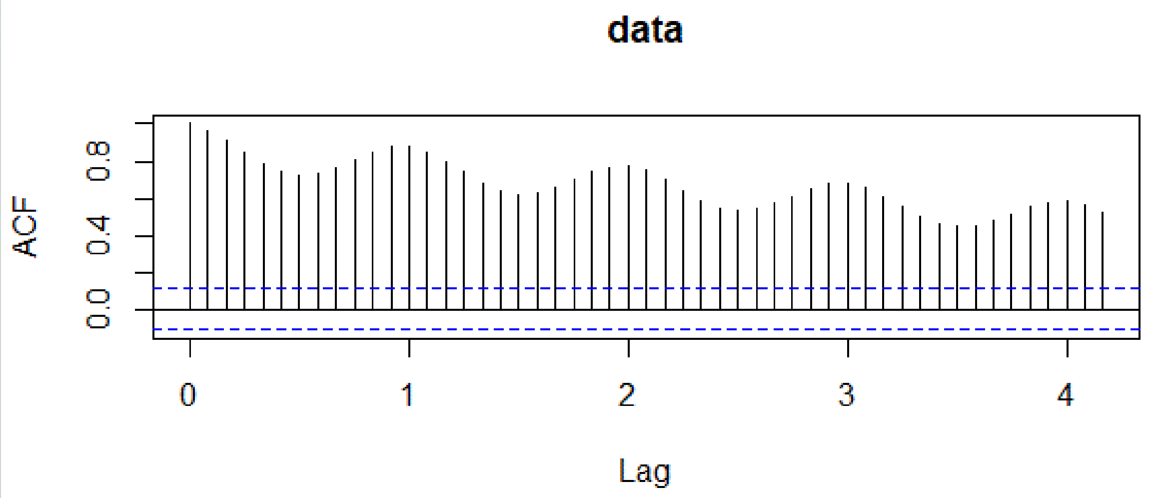
Observations

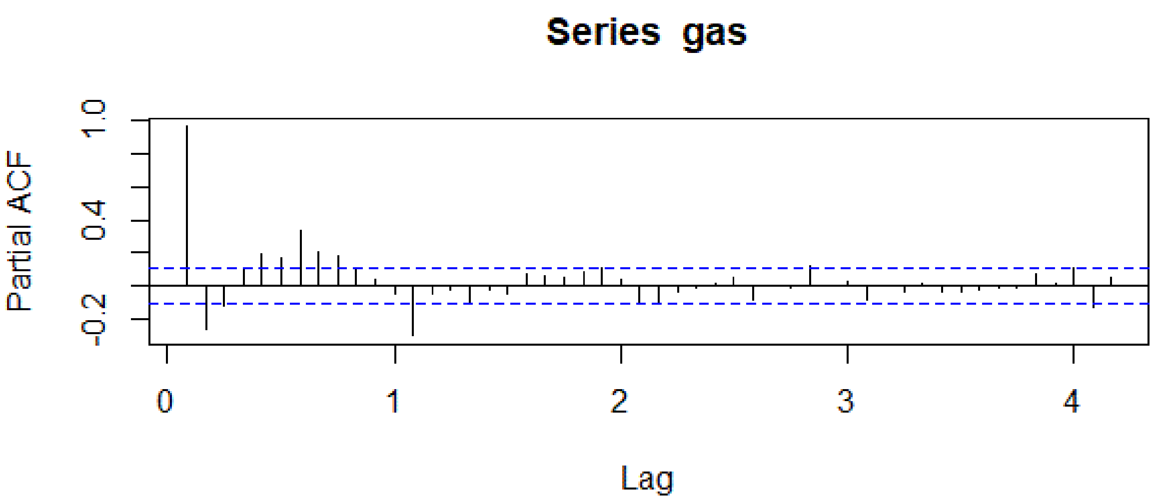
- The decomposition type was selected as multiplicative since the magnitude of seasonality increased with increase in time period.
- Trend and seasonality were observed in the data.
- It is a 12 month periodic dataset with data ranging from 1970-1995.
- Since there is seasonality we will proceed with de-seasonalising the data



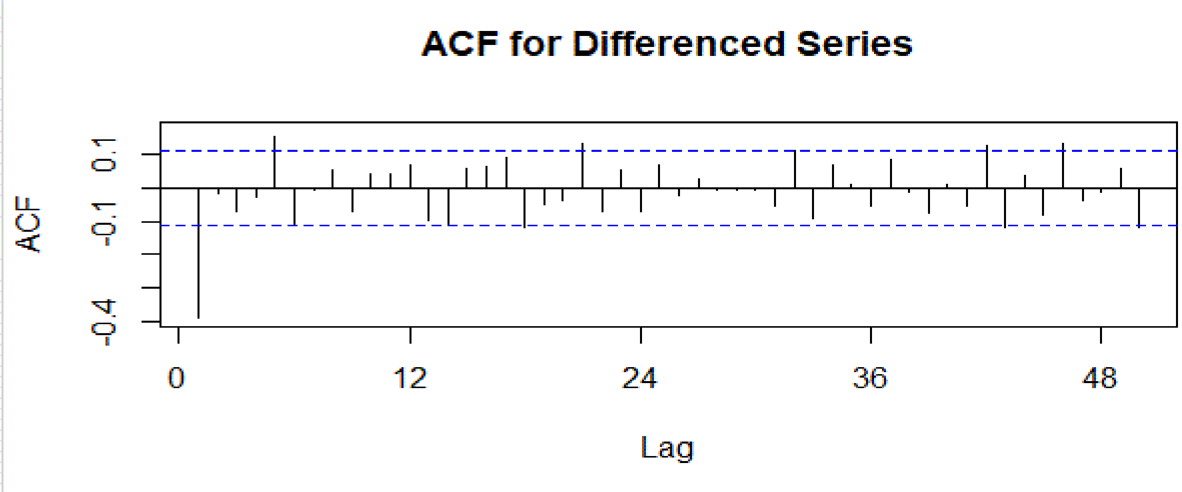
5. Test for stationarity

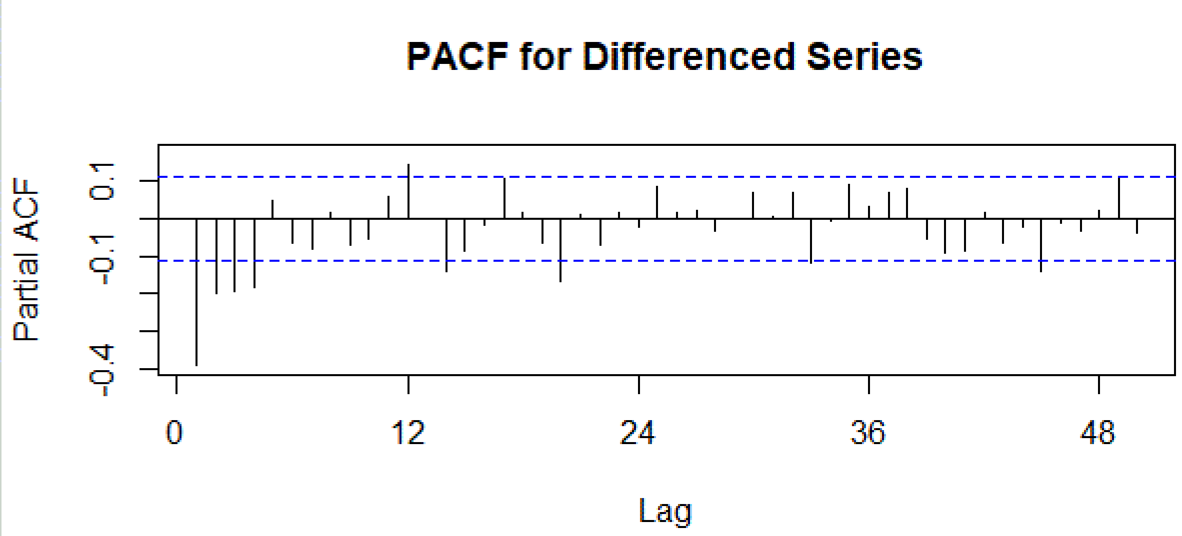
- The Test for stationarity of the data was conducted using Augmented Dickey Fuller Test. Below are its Hypothesis
  H0: Data is not stationary
  H1: Data is stationary
- P value of 0.5826 was obtained from the test which failed to reject the null Hypothesis and concluded that the data is not stationary.
- The below ACF and PACF plots were obtained for the data.





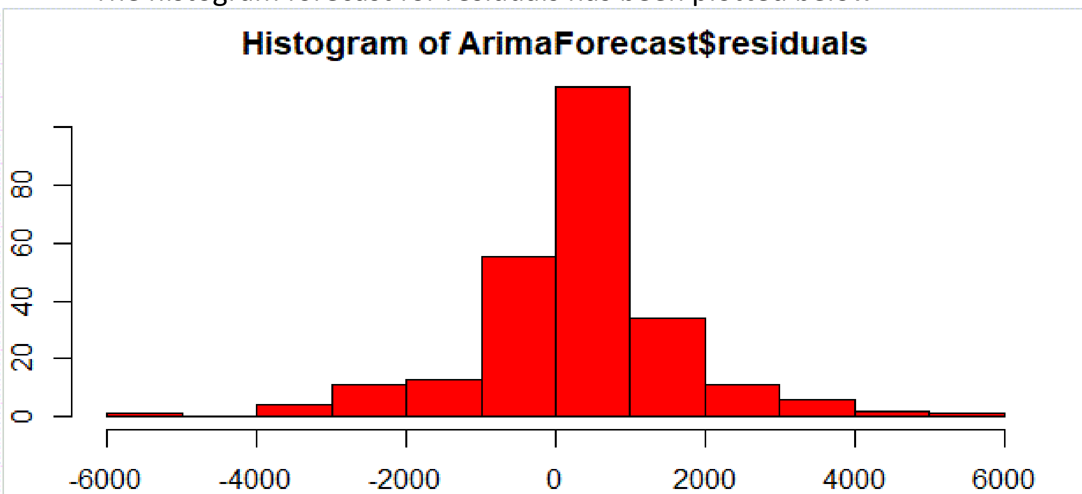
- Differencing was performed for the deseasonalised data and p value of 0.01 was obtained for ADF test.
- ACF and PACF for the differenced data were plotted as below.



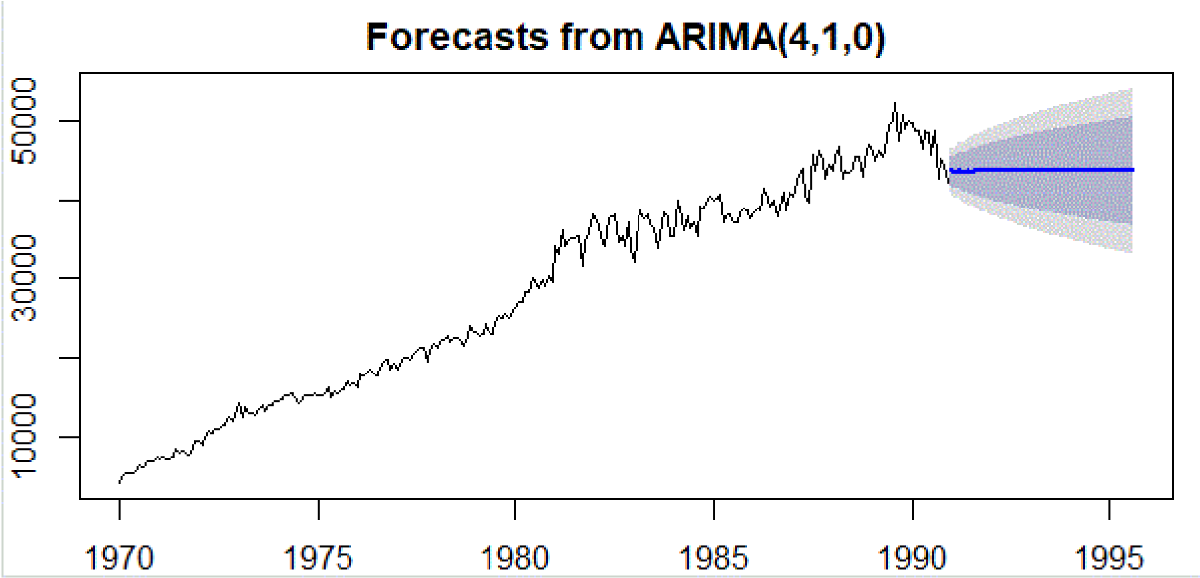


6. Fitting an ARIMA model

- ARIMA model was developed for an order of (4,1,0) for (p,d,q)
- AIC obtained was 4356.41
- The histogram forecast for residuals has been plotted below

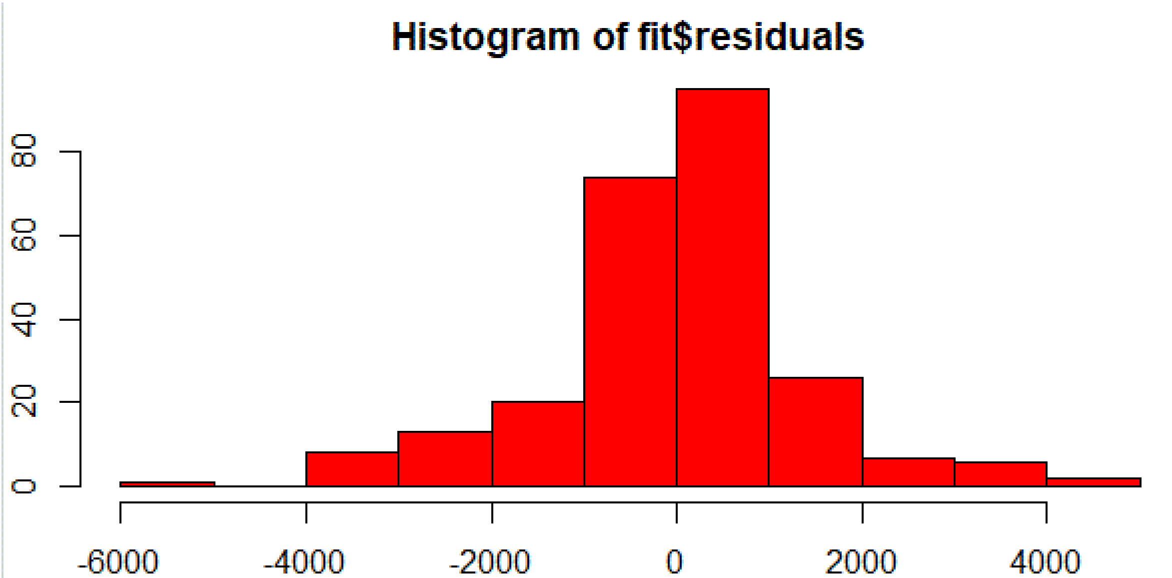


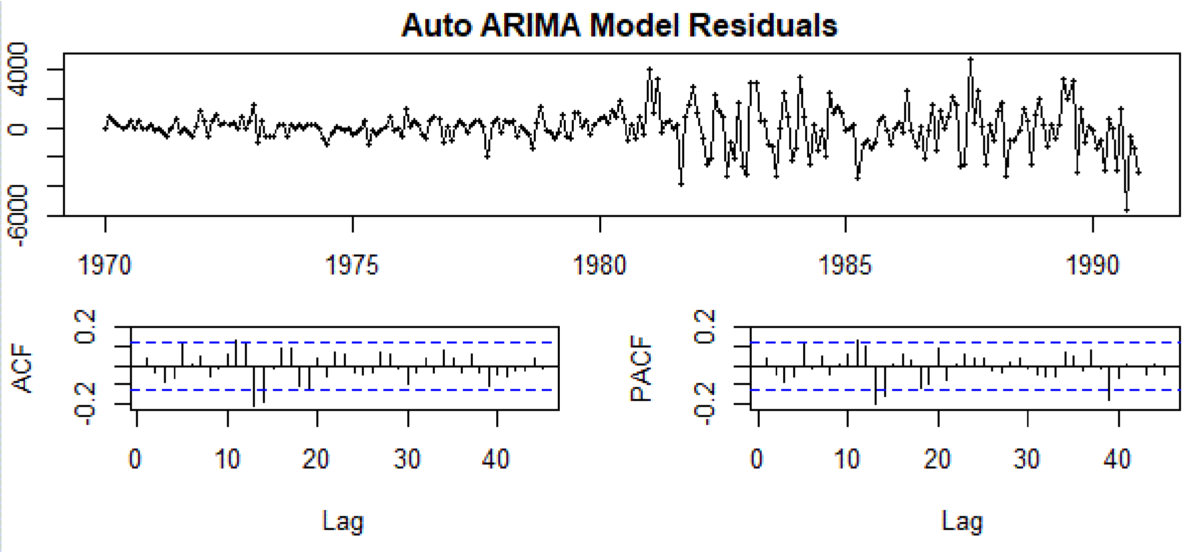
- Box-Pierce test was performed for ArimaForecast$residuals and p value of 0.486 was recorded which concludes that residuals are independent.



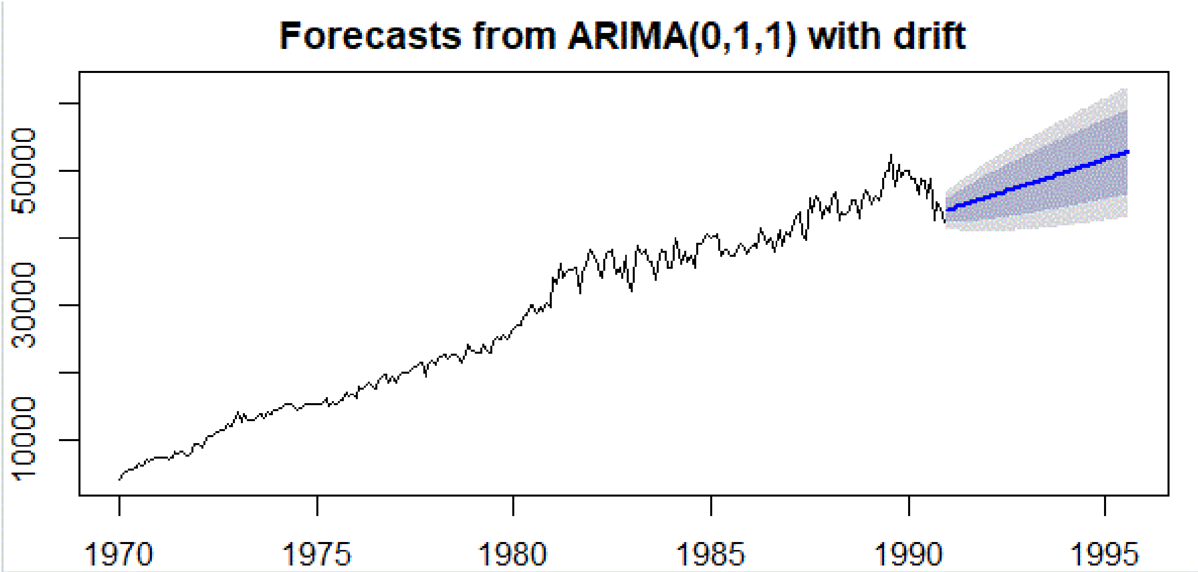
Fitting an Auto ARIMA model

- Auto ARIMA model was developed and AIC value of 4344.64 was obtained.
- The histogram forecast for residuals has been plotted below





- Box-Pierce test was performed for ArimaForecast$residuals and p value of 0.505 was recorded which concludes that residuals are independent.



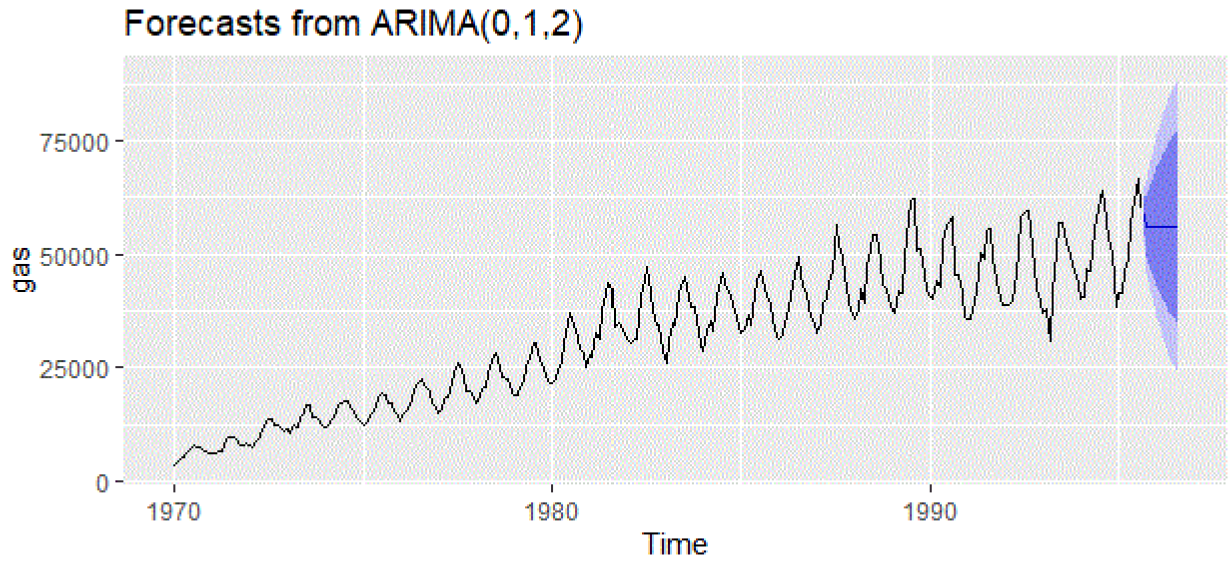
Accuracy

- The accuracy of the models for ARIMA and AUTO ARIMA was checked and found to be below

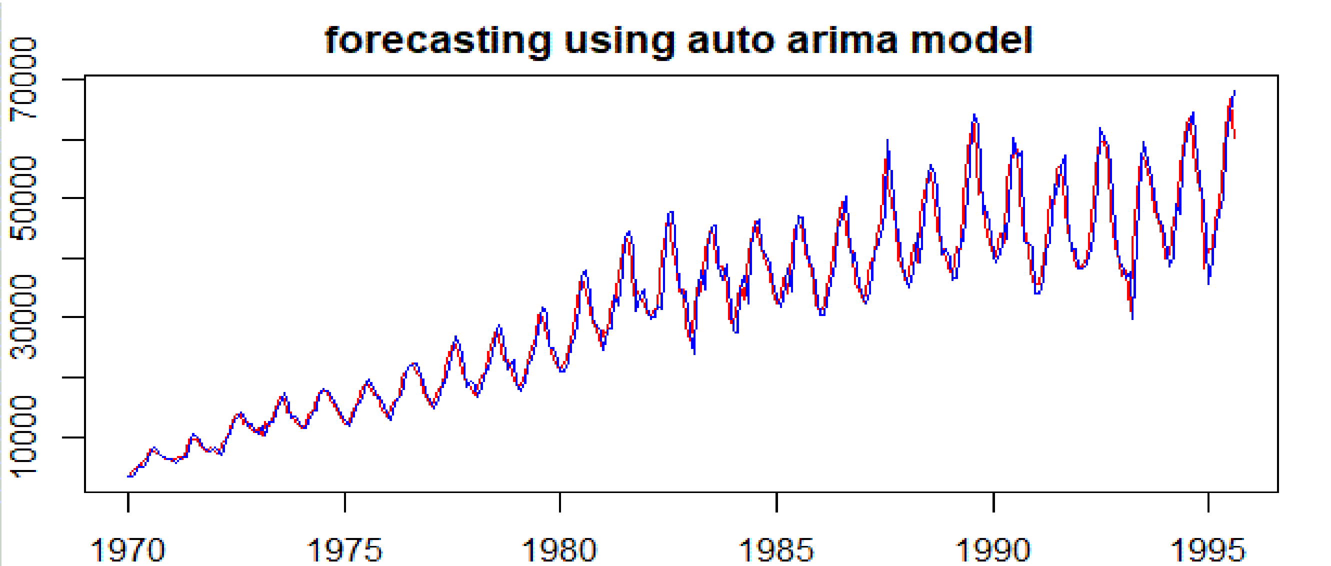
Accuracy	Train	Test	AIC
ARIMA	3.881558	9.97286	4356.41
AUTO ARIMA	3.457422	3.988248	4344.64

- From the above table we can conclude that auto.arima model performs better than arima. Hence we will use the same for forecasting on the future dataset for next 12 months.

7. Forecasting on future dataset



Model Comparison with the original dataset



8. Conclusion

Australia has the 10th largest natural gas exports in the world. It is proud to deliver the economic and environmental benefits of natural gas to homes and businesses throughout Australia and Asia. The core assets provide stable production, long-term revenue streams and significant upside opportunities.

The gas dataset contains monthly gas production from 1956-1995. Trend and seasonality were observed after decomposing the data. The data was then deseasonalised to test for stationarity which was conducted using Augmented Dickey Fuller (ADF) Test. The Time Series was stationarised by differencing the deseasonalised dataset one time. ACF and PACF were recorded as 0 and 4 respectively and were used to build the manual ARIMA model. Fitting was performed on both manual and auto ARIMA where the latter performed better with an AIC value of 4344.64. The model was then forecasted in the future for the next 12 months and the same has been plotted in the above section.

9. Appendix A – Source Code

Here is the code produced in RStudio for doing an analysis on Product Service Management.

```
> library(forecast)
> library('ggplot2')
> library('forecast')
> library('tseries')

    'tseries' version: 0.10-47

    'tseries' is a package for time series analysis and computational fina
nce.

    See 'library(help="tseries")' for details.

> library(readxl)
> #ts plot of the gas dataset
> plot(gas)
> setwd("C:/Users/Chetan Suvarna/Desktop/project/Time Series-p6")
> gas_total=read_excel("C:/Users/Chetan Suvarna/Desktop/project/Time Serie
s-p6/data.xlsx")
> gas <- ts(gas_total[,3], start=c(1970,1), frequency=12)
> #ts plot of the gas dataset
> #ts plot of the gas dataset
> plot(gas)
> boxplot(gas)
> str(gas)
 Time-Series [1:308, 1] from 1970 to 1996: 3345 4220 4874 5064 5951 ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr "data"
> summary(gas)
      data
```

```
 Min.   : 3345
 1st Qu.:17731
 Median :34017
 Mean   :31752
 3rd Qu.:42455
 Max.   :66600
> seasonplot(gas)
> #gas_data=read_excel("data.xlsx")
> #gas_data=read_excel("data.xlsx")
> monthplot(gas)
> hist(gas)
> #The frequency is given in cycle
> cycle(gas)
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1970   1   2   3   4   5   6   7   8   9  10  11  12
1971   1   2   3   4   5   6   7   8   9  10  11  12
1972   1   2   3   4   5   6   7   8   9  10  11  12
1973   1   2   3   4   5   6   7   8   9  10  11  12
1974   1   2   3   4   5   6   7   8   9  10  11  12
1975   1   2   3   4   5   6   7   8   9  10  11  12
1976   1   2   3   4   5   6   7   8   9  10  11  12
1977   1   2   3   4   5   6   7   8   9  10  11  12
1978   1   2   3   4   5   6   7   8   9  10  11  12
1979   1   2   3   4   5   6   7   8   9  10  11  12
1980   1   2   3   4   5   6   7   8   9  10  11  12
1981   1   2   3   4   5   6   7   8   9  10  11  12
1982   1   2   3   4   5   6   7   8   9  10  11  12
1983   1   2   3   4   5   6   7   8   9  10  11  12
1984   1   2   3   4   5   6   7   8   9  10  11  12
1985   1   2   3   4   5   6   7   8   9  10  11  12
1986   1   2   3   4   5   6   7   8   9  10  11  12
1987   1   2   3   4   5   6   7   8   9  10  11  12
1988   1   2   3   4   5   6   7   8   9  10  11  12
1989   1   2   3   4   5   6   7   8   9  10  11  12
1990   1   2   3   4   5   6   7   8   9  10  11  12
1991   1   2   3   4   5   6   7   8   9  10  11  12
1992   1   2   3   4   5   6   7   8   9  10  11  12
1993   1   2   3   4   5   6   7   8   9  10  11  12
1994   1   2   3   4   5   6   7   8   9  10  11  12
1995   1   2   3   4   5   6   7   8
> boxplot(gas~cycle(gas))
> decomp_gas=decompose(gas,type="multiplicative")
> plot(decomp_gas)
> decomp_gas
$x
       Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   D
ec
1970  3345  4220  4874  5064  5951  6774  7997  7523  7438  6879  6489  62
88
1971  5919  6183  6594  6489  8040  9715  9714  9756  8595  7861  7753  81
54
1972  7778  7402  8903  9742 11372 12741 13733 13691 12239 12502 11241 108
29
1973 11569 10397 12493 11962 13974 14945 16805 16587 14225 14157 13016 122
53
1974 11704 12275 13695 14082 16555 17339 17777 17592 16194 15336 14208 131
16
1975 12354 12682 14141 14989 16159 18276 19157 18737 17109 17094 15418 143
12
1976 13260 14990 15975 16770 19819 20983 22001 22337 20750 19969 17293 164
98
```

```
1977 15117 16058 18137 18471 21398 23854 26025 25479 22804 19619 19627 184
88
1978 17243 18284 20226 20903 23768 26323 28038 26776 22886 22813 22404 197
95
1979 18839 18892 20823 22212 25076 26884 30611 30228 26762 25885 23328 219
30
1980 21433 22369 24503 25905 30605 34984 37060 34502 31793 29275 28305 252
48
1981 27730 27424 32684 31366 37459 41060 43558 42398 33827 34962 33480 324
45
1982 30715 30400 31451 31306 40592 44133 47387 41310 37913 34355 34607 287
29
1983 26138 30745 35018 34549 40980 42869 45022 40387 38180 38608 35308 302
34
1984 28801 33034 35294 33181 40797 42355 46098 42430 41851 39331 37328 345
14
1985 32494 33308 36805 34221 41020 44350 46173 44435 40943 39269 35901 321
42
1986 31239 32261 34951 38109 43168 45547 49568 45387 41805 41281 36068 348
79
1987 32791 34206 39128 40249 43519 46137 56709 52306 49397 45500 39857 379
58
1988 35567 37696 42319 39137 47062 50610 54457 54435 48516 43225 42155 399
95
1989 37541 37277 41778 41666 49616 57793 61884 62400 50820 51116 45731 425
28
1990 40459 40295 44147 42697 52561 56572 56858 58363 45627 45622 41304 360
16
1991 35592 35677 39864 41761 50380 49129 55066 55671 49058 44503 42145 386
98
1992 38963 38690 39792 42545 50145 58164 59035 59408 55988 47321 42269 396
06
1993 37059 37963 31043 41712 50366 56977 56807 54634 51367 48073 46251 437
36
1994 39975 40478 46895 46147 55011 57799 62450 63896 57784 53231 50354 384
10
1995 41600 41471 46287 49013 56624 61739 66600 60054

$seasonal
           Jan       Feb       Mar       Apr       May       Jun       Jul
Aug       Sep
1970 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1971 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1972 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1973 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1974 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1975 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1976 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1977 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1978 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1979 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
```

```
1980 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1981 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1982 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1983 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1984 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1985 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1986 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1987 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1988 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1989 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1990 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1991 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1992 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1993 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1994 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
1995 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659
           Oct       Nov       Dec
1970 1.0074817 0.9311643 0.8522505
1971 1.0074817 0.9311643 0.8522505
1972 1.0074817 0.9311643 0.8522505
1973 1.0074817 0.9311643 0.8522505
1974 1.0074817 0.9311643 0.8522505
1975 1.0074817 0.9311643 0.8522505
1976 1.0074817 0.9311643 0.8522505
1977 1.0074817 0.9311643 0.8522505
1978 1.0074817 0.9311643 0.8522505
1979 1.0074817 0.9311643 0.8522505
1980 1.0074817 0.9311643 0.8522505
1981 1.0074817 0.9311643 0.8522505
1982 1.0074817 0.9311643 0.8522505
1983 1.0074817 0.9311643 0.8522505
1984 1.0074817 0.9311643 0.8522505
1985 1.0074817 0.9311643 0.8522505
1986 1.0074817 0.9311643 0.8522505
1987 1.0074817 0.9311643 0.8522505
1988 1.0074817 0.9311643 0.8522505
1989 1.0074817 0.9311643 0.8522505
1990 1.0074817 0.9311643 0.8522505
1991 1.0074817 0.9311643 0.8522505
1992 1.0074817 0.9311643 0.8522505
1993 1.0074817 0.9311643 0.8522505
1994 1.0074817 0.9311643 0.8522505
1995

$trend
```

```
          Jan       Feb       Mar       Apr       May       Jun       Jul
Aug       Sep
1970       NA        NA        NA        NA        NA        NA  6177.417
6366.458  6519.917
1971  7201.042  7365.625  7506.875  7596.000  7689.583  7820.000  7975.208
8103.458  8250.458
1972  9315.042  9646.458  9962.250 10307.458 10646.167 10902.958 11172.375
11455.125 11729.500
1973 12592.583 12841.250 13044.667 13196.375 13339.292 13472.583 13537.542
13621.417 13749.750
1974 14431.583 14513.958 14637.875 14769.042 14867.833 14953.458 15016.500
15060.542 15096.083
1975 15292.833 15398.042 15483.875 15595.250 15718.917 15819.167 15906.750
16040.667 16213.250
1976 17087.167 17355.667 17657.375 17928.875 18126.792 18296.000 18464.458
18586.333 18720.917
1977 19491.250 19789.833 20006.333 20077.333 20160.000 20340.167 20511.667
20693.000 20872.792
1978 21649.625 21787.542 21845.000 21981.500 22230.292 22400.458 22521.417
22613.250 22663.458
1979 23060.375 23311.417 23616.750 23906.250 24072.750 24200.208 24397.250
24650.208 24948.417
1980 26813.958 27260.750 27648.458 27999.333 28347.958 28693.583 29094.208
29567.208 30118.708
1981 32262.917 32862.667 33276.417 33598.125 34050.708 34566.208 34990.458
35238.833 35311.458
1982 35931.792 36046.000 36170.917 36315.875 36337.542 36229.667 35884.125
35707.792 35870.792
1983 36118.125 35981.125 35953.792 36142.125 36348.542 36440.458 36614.125
36820.458 36927.333
1984 36811.583 36941.542 37179.625 37362.708 37477.000 37739.500 38071.708
38237.000 38311.375
1985 38648.958 38735.625 38781.333 38740.917 38678.875 38520.583 38369.458
38273.542 38152.667
1986 38819.625 39000.750 39076.333 39196.083 39286.875 39407.875 39586.583
39732.292 39987.375
1987 40715.708 41301.542 41906.167 42398.292 42731.958 43018.125 43262.083
43523.167 43801.542
1988 44416.000 44410.875 44462.875 44331.375 44332.333 44512.958 44680.083
44744.875 44704.875
1989 46013.958 46655.292 47083.167 47507.958 47985.750 48240.292 48467.417
48714.750 48939.208
1990 49058.083 48680.458 48295.875 47850.583 47437.208 46981.417 46507.292
46112.083 45741.208
1991 44608.083 44421.250 44452.042 44548.375 44536.792 44683.583 44935.792
45201.792 45324.333
1992 46285.375 46606.458 47050.917 47457.083 47579.667 47622.667 47581.167
47471.542 47076.708
1993 46469.417 46177.667 45786.208 45625.000 45822.250 46160.250 46453.833
46680.125 47445.417
1994 49166.208 49787.250 50440.542 50922.833 51308.708 51257.750 51103.542
51212.625 51228.667
1995 52077.833 52090.667        NA        NA        NA        NA        NA
NA
          Oct       Nov       Dec
1970  6650.958  6797.375  7006.958
1971  8482.208  8756.583  9021.500
1972 11971.583 12172.500 12372.750
1973 13888.167 14084.042 14291.333
1974 15152.458 15173.750 15196.292
1975 16363.875 16590.583 16855.875
1976 18881.875 19018.542 19203.958
```

```
1977 21061.167 21261.250 21462.875
1978 22742.875 22851.917 22929.792
1979 25255.625 25639.875 26207.750
1980 30687.125 31200.250 31739.000
1981 35257.583 35385.625 35644.208
1982 36154.542 36305.833 36269.333
1983 36881.833 36817.208 36788.167
1984 38417.667 38470.292 38562.708
1985 38237.417 38488.917 38628.292
1986 40250.583 40354.375 40393.583
1987 43888.167 43989.458 44323.458
1988 44787.708 44999.500 45405.208
1989 49080.875 49246.542 49318.375
1990 45523.750 45393.875 44992.875
1991 45354.000 45376.875 45743.542
1992 46677.458 46651.958 46611.708
1993 48290.708 48669.042 48896.833
1994 51322.750 51509.375 51740.750
1995

$random
           Jan       Feb       Mar       Apr       May       Jun       Jul
Aug       Sep
1970        NA        NA        NA        NA        NA        NA 1.0439267
0.9887837 1.0675877
1971 1.0114000 1.0133269 0.9698804 0.9307009 0.9720504 1.0662952 0.9822134
1.0074176 0.9748937
1972 1.0274320 0.9262775 0.9867506 1.0297074 0.9930683 1.0029992 0.9912192
1.0001001 0.9764632
1973 1.1304489 0.9773738 1.0574557 0.9875667 0.9739205 0.9521085 1.0010332
1.0189519 0.9681595
1974 0.9979075 1.0209291 1.0330288 1.0387937 1.0351828 0.9952308 0.9546398
0.9774235 1.0038742
1975 0.9940073 0.9942193 1.0083909 1.0471228 0.9557128 0.9916056 0.9711713
0.9774303 0.9875157
1976 0.9548682 1.0426061 0.9989487 1.0190543 1.0164742 0.9843575 0.9608501
1.0056324 1.0372428
1977 0.9543240 0.9795109 1.0009823 1.0023091 0.9867752 1.0065796 1.0231506
1.0303080 1.0223979
1978 0.9800143 1.0130314 1.0223189 1.0360211 0.9939923 1.0086021 1.0039246
0.9908111 0.9450029
1979 1.0052207 0.9782936 0.9735351 1.0122633 0.9684297 0.9534901 1.0117808
1.0261172 1.0038412
1980 0.9835399 0.9905325 0.9785359 1.0079830 1.0037051 1.0464683 1.0271850
0.9764322 0.9878355
1981 1.0575872 1.0073675 1.0844938 1.0170947 1.0227408 1.0195502 1.0038483
1.0067740 0.8964731
1982 1.0518204 1.0180672 0.9600711 0.9391790 1.0385334 1.0455397 1.0648946
0.9680557 0.9890919
1983 0.8904655 1.0314774 1.0754125 1.0414516 1.0481429 1.0097199 0.9915758
0.9178264 0.9675589
1984 0.9627045 1.0794589 1.0481521 0.9675389 1.0120430 0.9632742 0.9764039
0.9285330 1.0222746
1985 1.0345115 1.0380014 1.0478824 0.9623656 0.9859556 0.9881939 0.9704032
0.9714818 1.0042555
1986 0.9901836 0.9985385 0.9875845 1.0592591 1.0215273 0.9920148 1.0097251
0.9558638 0.9783512
1987 0.9909748 0.9997605 1.0309513 1.0342465 0.9468074 0.9205328 1.0570471
1.0056322 1.0553600
1988 0.9853210 1.0246272 1.0509119 0.9618196 0.9869277 0.9758685 0.9828552
1.0179890 1.0155927
```

```
1989 1.0038901 0.9644950 0.9797391 0.9555045 0.9612692 1.0282689 1.0296231
1.0718456 0.9717782
1990 1.0147861 0.9992093 1.0092984 0.9721369 1.0301016 1.0335151 0.9858715
1.0590854 0.9334772
1991 0.9817678 0.9695216 0.9901879 1.0213070 1.0516586 0.9436943 0.9881911
1.0305795 1.0129030
1992 1.0358063 1.0021035 0.9338048 0.9767081 0.9798098 1.0482913 1.0005165
1.0471761 1.1129568
1993 0.9812878 0.9924040 0.7486131 0.9960369 1.0218722 1.0594314 0.9861207
0.9793527 1.0131632
1994 1.0004415 0.9814334 1.0265385 0.9872979 0.9967678 0.9678369 0.9854424
1.0440103 1.0555626
1995 0.9829023 0.9610468        NA        NA        NA        NA        NA
NA
           Oct       Nov       Dec
1970 1.0266063 1.0252038 1.0529694
1971 0.9198811 0.9508429 1.0605343
1972 1.0365511 0.9917423 1.0269632
1973 1.0117871 0.9924849 1.0060103
1974 1.0045969 1.0055732 1.0127367
1975 1.0368606 0.9980219 0.9962808
1976 1.0497213 0.9764878 1.0080296
1977 0.9246072 0.9913769 1.0107291
1978 0.9956343 1.0528745 1.0129503
1979 1.0173090 0.9770916 0.9818421
1980 0.9468987 0.9742687 0.9333972
1981 0.9842526 1.0160902 1.0680499
1982 0.9431699 1.0236729 0.9294236
1983 1.0390289 1.0299021 0.9643179
1984 1.0161711 1.0420365 1.0501722
1985 1.0193519 1.0017158 0.9763379
1986 1.0179838 0.9598539 1.0131748
1987 1.0290270 0.9730377 1.0048528
1988 0.9579415 1.0060396 1.0335531
1989 1.0337307 0.9972605 1.0118100
1990 0.9947160 0.9771664 0.9392571
1991 0.9739497 0.9974363 0.9926394
1992 1.0062585 0.9730290 0.9970082
1993 0.9880990 1.0205681 1.0495209
1994 1.0294791 1.0498358 0.8710525
1995

$figure
 [1] 0.8126996 0.8283999 0.9056733 0.9178732 1.0756338 1.1650876 1.2400814
1.1950659 1.0685886
[10] 1.0074817 0.9311643 0.8522505

$type
[1] "multiplicative"

attr(,"class")
[1] "decomposed.ts"
> deseasonal_gas=seasadj(decomp_gas)
> ts.plot(gas, deseasonal_gas, col=c("red", "blue"),
+         main="Gas data vs De-seasoned Gas")
> #Check for stationarity using the Augmented Dickey-Fuller test
>
> adf.test(deseasonal_gas, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  deseasonal_gas
```

```
Dickey-Fuller = -1.9851, Lag order = 6, p-value = 0.5826
alternative hypothesis: stationary

> #0.5826 for 95% confidence level. < 0.05
>
> #Series is not stationary
>
> #Check autocorrelation plots
> #ACF and PACF plots
> acf(gas)
> pacf(gas)
> acf(gas, lag.max = 50)
> pacf(gas, lag.max = 50)
> #There are significant autocorrelations with many lags in our gas series, as shown
by
> #the ACF plot.
>
> #PACF plot shows that there could be monthly seasonality since the plot peaks at
> #at various intervals
>
> #Differencing the time series data
> #deseasonal_gas=seasadj(gas)
> count_d1 = diff(deseasonal_gas, differences = 1)
> plot(count_d1)
> adf.test(count_d1, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  count_d1
Dickey-Fuller = -8.8861, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(count_d1, alternative = "stationary") :
  p-value smaller than printed p-value
> #acf and pacf for dif time series. ARIMA(p,d,q)
> Acf(count_d1, main='ACF for Differenced Series',lag=50) #--- q
> Pacf(count_d1, main='PACF for Differenced Series',lag.max = 50) #--- p
> #From the ACF plot, there is a cut off after lag 0. This implies that q=0.
> #PACF cuts off after lag 10. Hence p=4.
>
> #p=4,d=1,q=0
> #Splitting into training and test sets
> gas_train <- window(deseasonal_gas, start=c(1970,1), end=c(1990,12), frequency=12)
> gas_test <- window(deseasonal_gas, start=c(1991,1), frequency=12)
> #Fitting an ARIMA model
>
> gas_ARIMA = arima(gas_train, order=c(4,1,0))
> gas_ARIMA

Call:
arima(x = gas_train, order = c(4, 1, 0))

Coefficients:
          ar1      ar2      ar3      ar4
      -0.4483  -0.2331  -0.1734  -0.1442
s.e.   0.0630   0.0681   0.0678   0.0647

sigma^2 estimated as 1938176:  log likelihood = -2173.2,  aic = 4356.41
> ArimaForecast <- forecast(gas_ARIMA, h=56)
> plot(ArimaForecast)
> VecA01 <- cbind(gas_test,ArimaForecast)
> par(mfrow=c(1,1), mar=c(2, 2, 2, 2), mgp=c(3, 1, 0), las=0)
> ts.plot(VecA01[,1],VecA01[,2], col=c("blue","red"),xlab="year", ylab="gas", main="g
as: Actual vs Forecast")
> VecA01 <- cbind(gas_test,ArimaForecast)
> par(mfrow=c(1,1), mar=c(2, 2, 2, 2), mgp=c(3, 1, 0), las=0)
> ts.plot(VecA01[,1],VecA01[,2], col=c("blue","red"),xlab="year", ylab="gas", main="g
```

```
as: Actual vs Forecast")
> Box.test(ArimaForecast$residuals)

        Box-Pierce test

data:  ArimaForecast$residuals
X-squared = 0.48533, df = 1, p-value = 0.486

> #p 0.486
> hist(ArimaForecast$residuals,col="red")

> #Fitting with Auto ARIMA
> fit<-auto.arima(gas_train, seasonal=FALSE)
> fit
Series: gas_train
ARIMA(0,1,1) with drift

Coefficients:
          ma1     drift
      -0.5369  156.4853
s.e.   0.0591   40.2627

sigma^2 estimated as 1894002:  log likelihood=-2169.32
AIC=4344.64   AICc=4344.74   BIC=4355.22
> tsdisplay(residuals(fit), lag.max=45, main='Auto ARIMA Model Residuals')
> Box.test(fit$residuals)

        Box-Pierce test

data:  fit$residuals
X-squared = 0.44444, df = 1, p-value = 0.505

> #p 0.505
> hist(fit$residuals,col="red")
> #Residual analysis using Ljung-Box test
> #H0: Residuals are independent
> #Ha: Residuals are not independent
> library(stats)
> #Validate both the manual and automatically fitted ARIMA models
> fcast <- forecast(gas_ARIMA, h=56)
> fcast1 <- forecast(fit, h=56)
> plot(fcast)
> plot(fcast1)
> autoplot(fcast)
> autoplot(fcast1)
> accuracy(fcast,gas_test)
                   ME     RMSE       MAE      MPE     MAPE      MASE        ACF1 Thei
l's U
Training set  311.748 1389.419  989.4459 1.654326 3.749744 0.3805246 -0.04388505
NA
Test set     4407.324 5733.296 4806.1859 8.606601 9.733419 1.8483801  0.59603116  1.5
56219
> accuracy(fcast1,gas_test)
                      ME     RMSE      MAE        MPE     MAPE      MASE       ACF1 T
heil's U
Training set    2.672296 1368.011  943.263  0.1356263 3.347516 0.3627634 0.04199564
NA
Test set     -325.586994 2684.184 1735.379 -1.0403617 3.809315 0.6673983 0.26269287 0
.7724838
> #Forecast into the future
> fit1<-auto.arima(gas, seasonal=FALSE)
> fcast2=forecast(fit1, h=12)
> plot(fcast2)
> autoplot(fcast2)
> ts.plot(gas,fcast2$fitted,col=c("red","blue"))


>
```