Project – gas (Australian monthly gas production)

Table of Content

1. **Objective**

   The objective of this project work is to build a time series forecasting model to predict the monthly gas production in Australia. ARIMA modeling technique is employed to build the forecasting model.

2. **Reading the Time Series Model and plotting**

   Time series data set **gas** was read and plotted as shown (Fig.1).

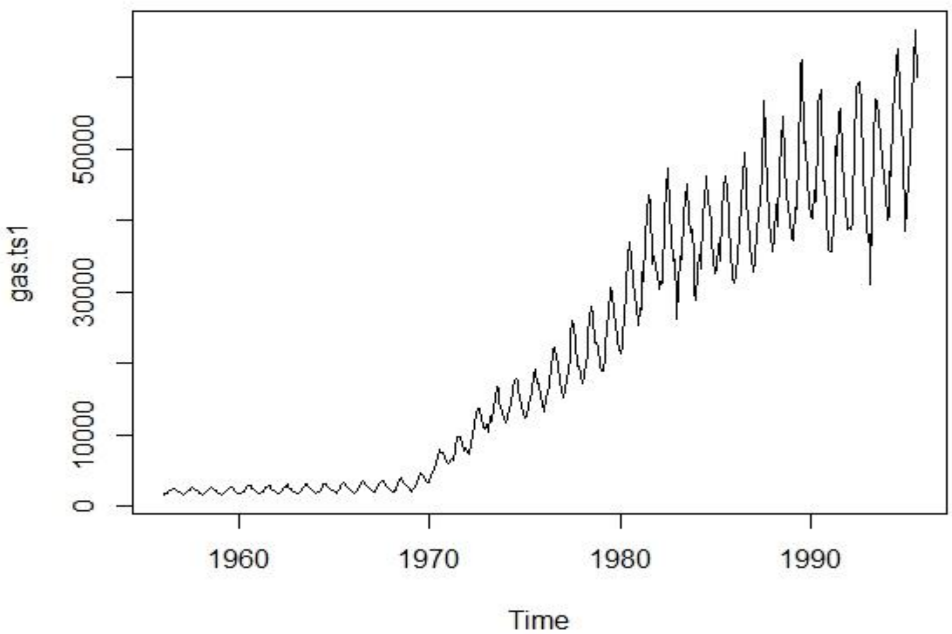

Fig.1: gas data set

Observations

- Times series data given in the **gas** data set contain the trend and seasonality components (Fig.1).
- **Periodicity** of the data set is **yearly** (i.e. data given has twelve observations per year) (Fig.1).

3. **Stationarity**

   From Fig.2 it can be seen that the data set is not stationary as it contains both trend and seasonality. It is also evident from the ACF and PACF plot as it contains some spikes (Fig.3 and Fig.4)

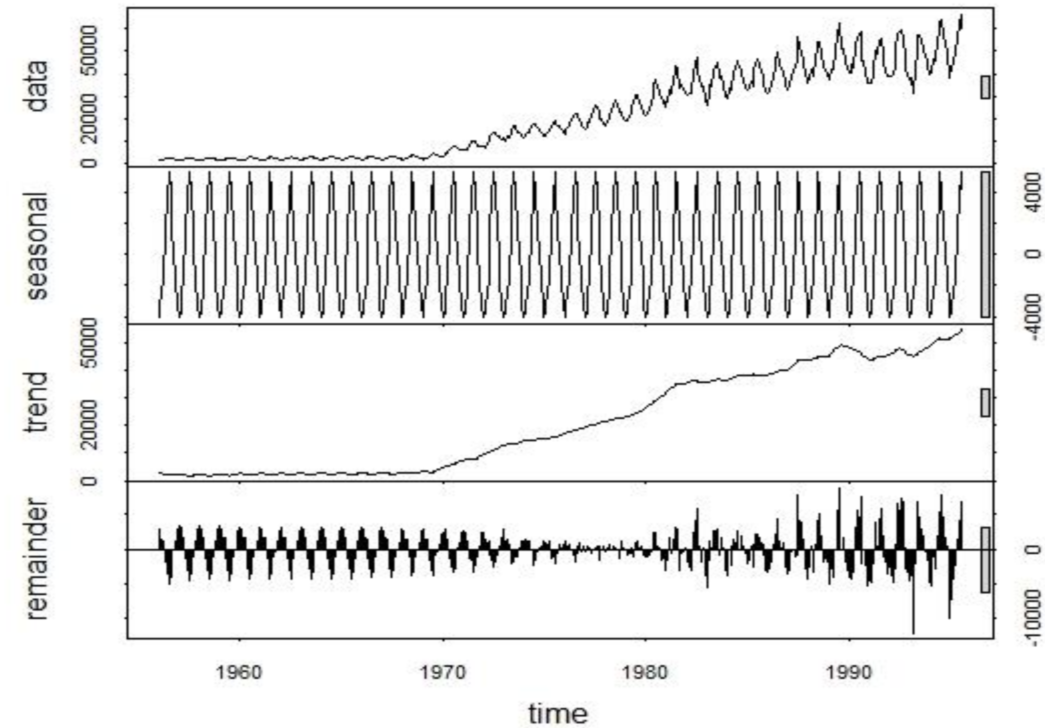

Fig.2: Seasonal, Trend and remainder graph

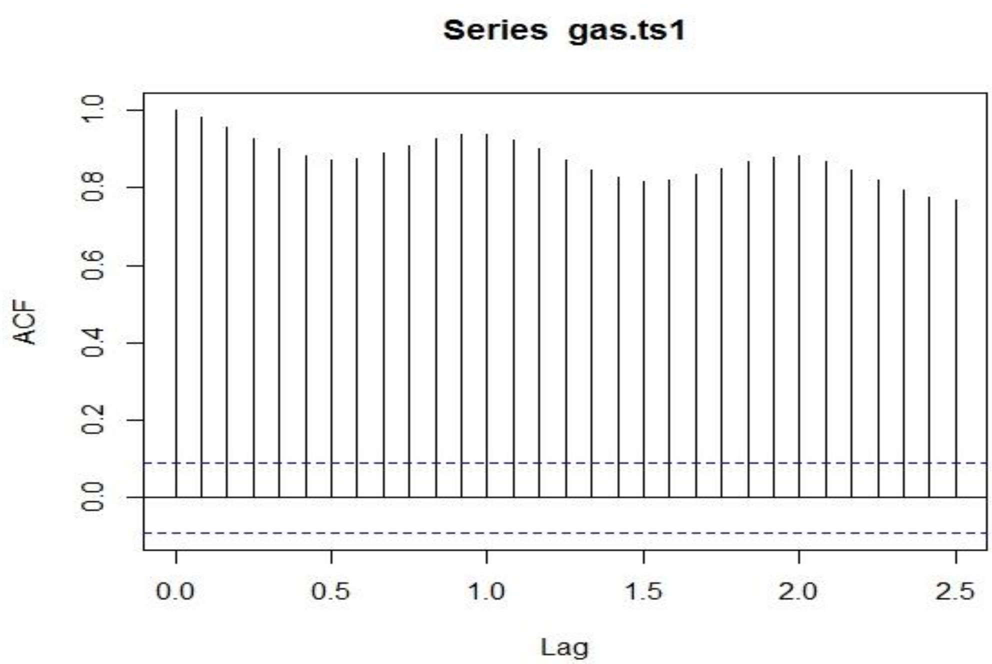

Fig.3: ACF Plot

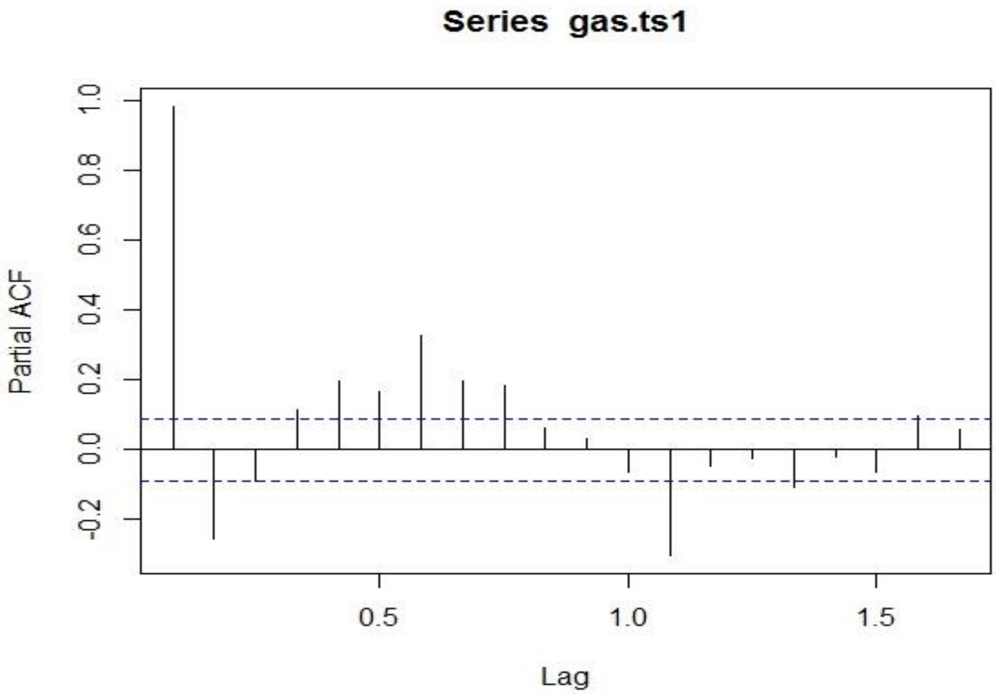

Fig.4: PACF Plot

Statistically, it can also be proved with the help of Augmented Dickey Fuller test (ADF test). To test the stationarity using ADF we can design the null hypothesis (H_0) and alternative hypothesis (H_a)

H_0 = Time series data set is not stationary

H_a = Time series data set is stationary

Results obtained from the ADF test is shown below:

Augmented Dickey-Fuller Test

data: gas.ts1

Dickey-Fuller = -2.7131, Lag order = 7, p-value = 0.2764

alternative hypothesis: stationary

***p is high null will fly, meaning null hypothesis is accepted . Given TS data is not stationary**

4. Removing seasonality and trend

Seasonality and trend is removed by subtracting the seasonality and trend component from the given data set in order to make the data stationary. ADF test proves the stationarity of the data as shown below:

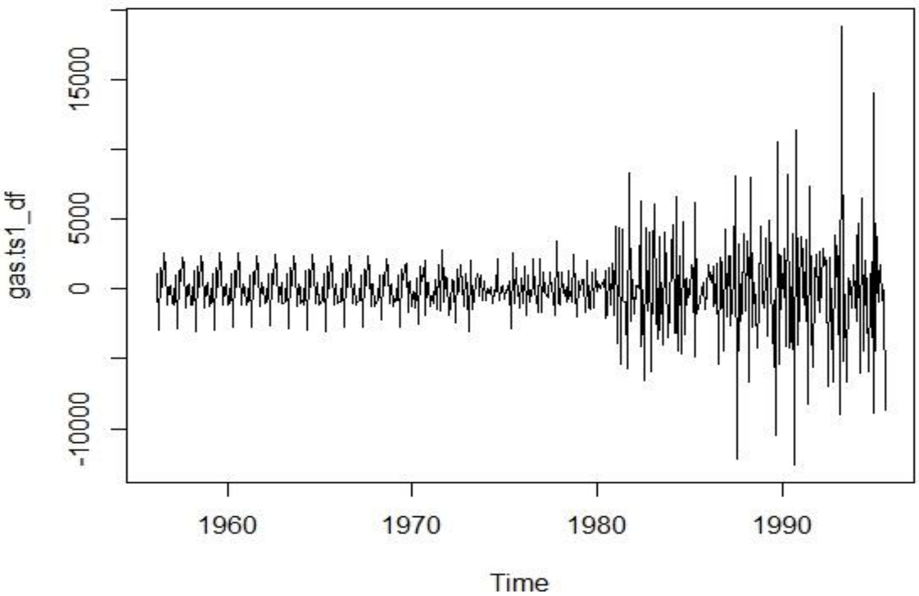

Fig.5: De-seasonal and De-Trend data

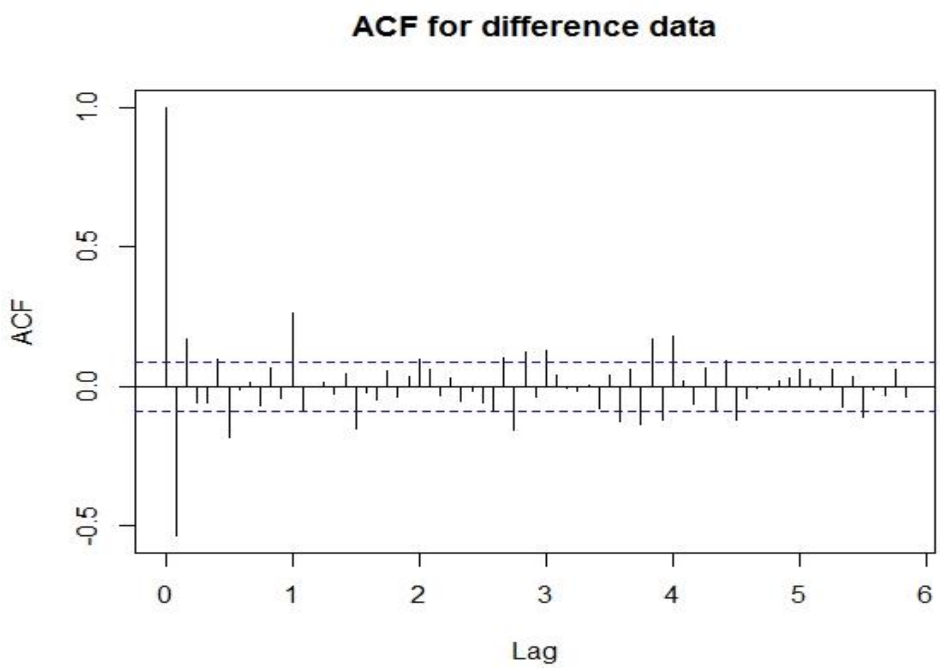

Fig.6: ACF plot on difference data

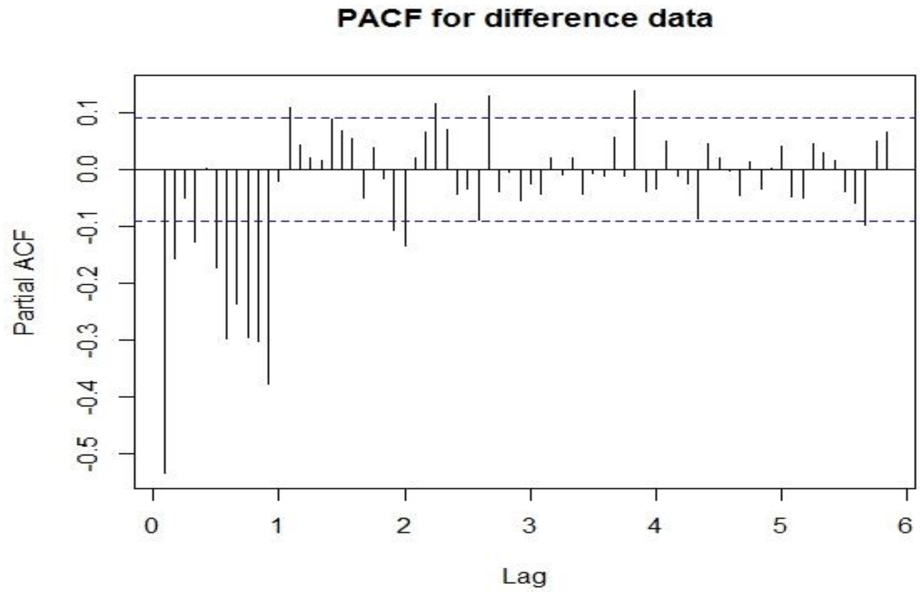

Fig.7: PACF plot on difference data

ADF test on difference gas data

Augmented Dickey-Fuller Test

data: gas.ts1_df
Dickey-Fuller = -15.782, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
***p is low null will go, meaning null hypothesis is rejected i.e. difference gas data is stationary.**

5. ARIMA (Autoregressive Integrated Moving Average)

Stationary data obtained after removing the seasonality and second order differencing as evident from the p value obtained from the ADF test is split into train and test data. An ARIMA model is then built on train data.

Results obtained from the **manual ARIMA** is listed below:

- P = 12
- d = 2
- q = 5
- AIC = 5114.69
- Log likelyhood = -2539.34

Results obtained from auto ARIMA is listed as below:

- P = 0
- d = 1
- q = 0
- AIC = 5263.52
- Log likelyhood = -2630.76

Results obtained from the manual ARIMA and auto ARIMA shows that the AIC value for auto ARIMA is higher as compared manual ARIMA. Hence, choosing auto ARIMA model as a better model for forecasting.

6. Testing ARIMA model

Manual ARIMA and auto ARIMA

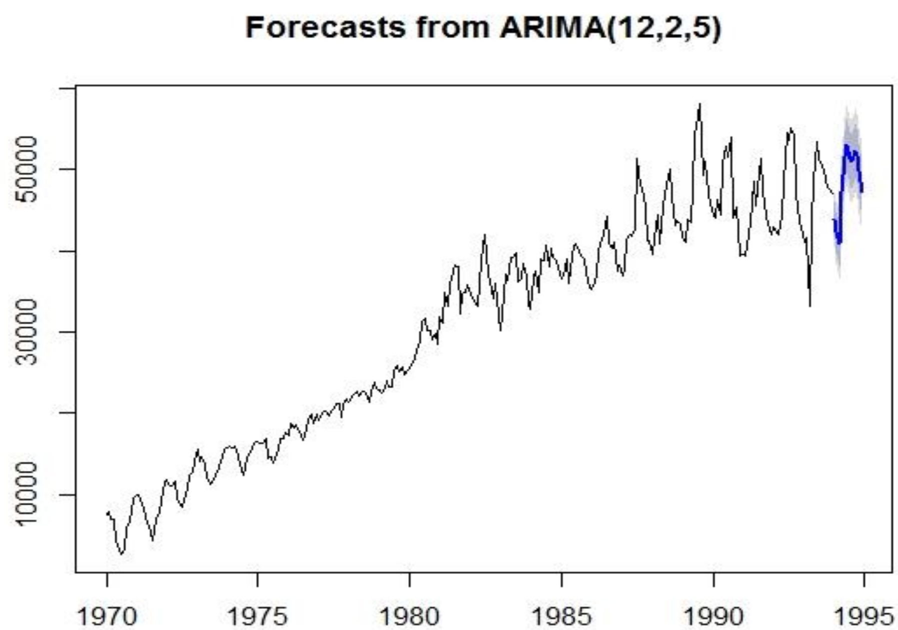

Fig.8: Forecasting manual ARIMA

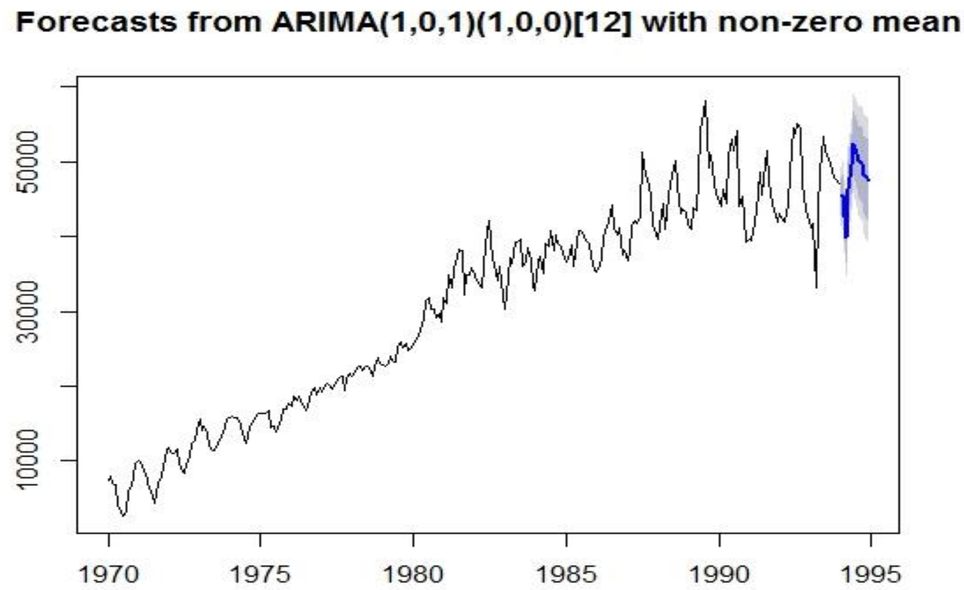

Fig.9: Forecasting auto ARIMA

7. Accuracy

Moving average percentage error (MAPE) is calculated on training and test data sets for the manual ARIMA and auto ARIMA which is tabulated in Table-1 and Table-2 respectively.

Table-1: MAPE for manual ARIMA

	MAPE
Training set	4.230
Test set	6.123

Table-2: MAPE for auto ARIMA

	MAPE
Training set	5.032
Test set	9.476

8. Conclusion

However the AIC value for the auto ARIMA is higher as compared to manual ARIMA model. We will choose the manual ARIMA model as the best model for forecasting since it has less percentage of error (Table-1 and Table-2)