MATH1318 Time Series Analysis / MATH2204 Time Series and Forecasting Final Project Report

“Analysing, Modelling and Evaluating time series Model for Aviation Gasoline CO2 Emissions in the US from 2000-2024"

Declaration of contributions:

No	Name of Team Member	Contribution to the project
1	Akshata Chavan (s3988608)	1/3
2	Kailas Nath Reghunath (s3957282)	1/3
3	Shifali Bhandari Shetty (s3956999)	1/3
	Sum:	1

RMIT University,
School of Science
2024

Index

Introduction

The goal of this assignment is to analyse a public time series dataset and fit a suitable model which can be used to predict future behaviour. As part of modelling the series, we will do a descriptive analysis of the dataset and identify the most defining characteristics of the series. Based on our exploration of the series we will determine a suitable plan of action which involves identifying the type of model we want to apply and find a suitable set of models. We will explore each model's suitability to the series by parameter estimation and residual analysis. Based on each model’s coefficient significance, error metrics and diagnostic checking we will identify the best model out of the bunch and use this for forecasting the next 10 units of time.

In this project we will go through each phase of the time series analysis in detail, provide a justification for each choice made and present our results in the most effective way.

Public Dataset

Global warming is an escalating concern, driven by increasing levels of greenhouse gases in the atmosphere, leading to significant climate changes worldwide. One notable contributor to this environmental issue is aviation. Aviation fuel CO2 emissions significantly contribute to global warming, accounting for approximately 2-3% of total global emissions, and impact climate change through high-altitude emissions. The United States, with its extensive air travel network and high volume of both domestic and international flights, is a major contributor to these emissions. The U.S. aviation sector plays a significant role in the global aviation industry, thus adding substantially to the carbon footprint associated with air travel.

In this project we will study how the CO2 emissions of Aviation Gasoline in the US from 2000-2024 changed over time. Our source is the United States Energy Information Administration. Our time series consists of monthly data of CO2 emissions (in million metric tonnes) from January 2000 to January 2024. Our goal is to determine, **What are the most accurate forecasts for the Aviation fuel CO2 emissions in the United States for the next 10 months?**

Descriptive Analysis

In this section we will read in our data, plot it against time and identify its most defining characteristics. Gathering as much information about our series as possible will help us easily decide on the type of model we need to use. Before we start our analysis, we will first load the libraries and functions we will be using during our project. (See Appendix)

Let us read in our data as shown below.

```
> aviation <- read.csv("aviation.csv", header = TRUE)
```

Missing Values

The first step in our analysis is to check for missing values. This helps us identify whether data imputation measures are required.

```
> sum(is.na(aviation$Jet.Fuel))
[1] 0
```

As seen in the above output, there are no missing values in the series so no further action in that regard is required.

Summary Statistics

Let us also see the summary statistics of the series to get an initial idea of the distribution.

```
> summary(aviation$Jet.Fuel)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.053   0.120   0.149   0.153   0.179   0.311 
```

The spread of the series is between 0.053 to 0.3. The mean is slightly greater than the median indicating a slight skew to the right. The right skew can also be seen in the histogram of the CO2 emissions in fig.1. Most of the values lie between 0.120 (1st quartile) and 0.179 (3rd quartile).

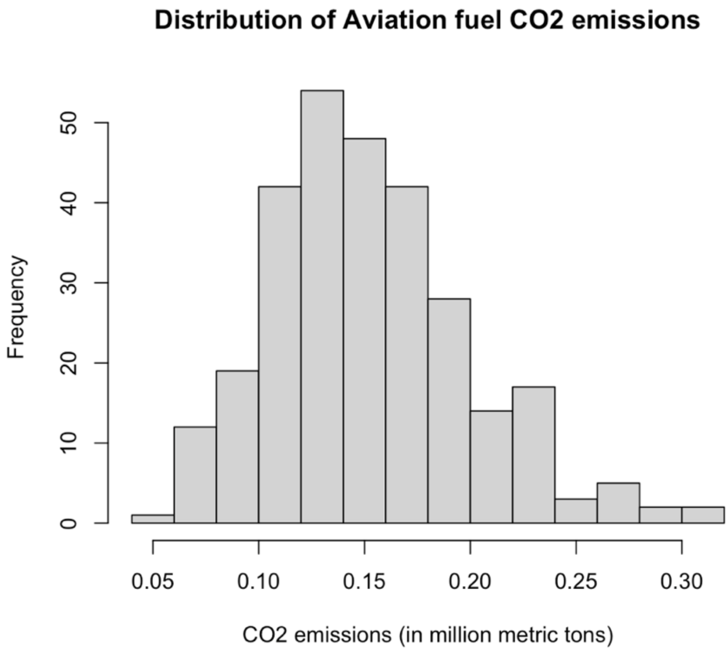


fig1. Histogram of CO2 emissions

We also checked for Skewness and Kurtosis.

```
> skewness <- moments::skewness(aviation$Jet.Fuel)
> print(paste("Skewness: ",skewness))
[1] "Skewness:  0.640987305352397"
> kurtosis <- moments::kurtosis(aviation$Jet.Fuel)
> print(paste("Kurtosis: ",kurtosis))
[1] "Kurtosis:  3.36636715140141"
```

The skewness of the Jet Fuel data is 0.641, indicating a slight right skew, where the tail of the distribution extends more to the right side. While the kurtosis value is 3.366 suggesting that the distribution has higher, and fatter tails compared to a normal distribution

Time Series Plot

To plot the series against time we need to convert `aviation` to a time series object. Here we will set the frequency to 12 as we have monthly data spanning many years, we will set the start and end time as (2000,1) (for January 2000, the first reading) and 2024,1 (for January 2024, the last reading) respectively.

```
> aviationTS <- ts(aviation$Jet.Fuel, start=c(2000, 1), end=c(2024, 1), frequency = 12)
```

Let us now plot the time series of the Aviation gasoline CO2 emissions.

```
> plot(aviationTS, ylab='CO2 emissions (in million metric tons)', xlab='Years', type='o',
+       main = "Time series plot of Aviation fuel CO2 emissions")
```

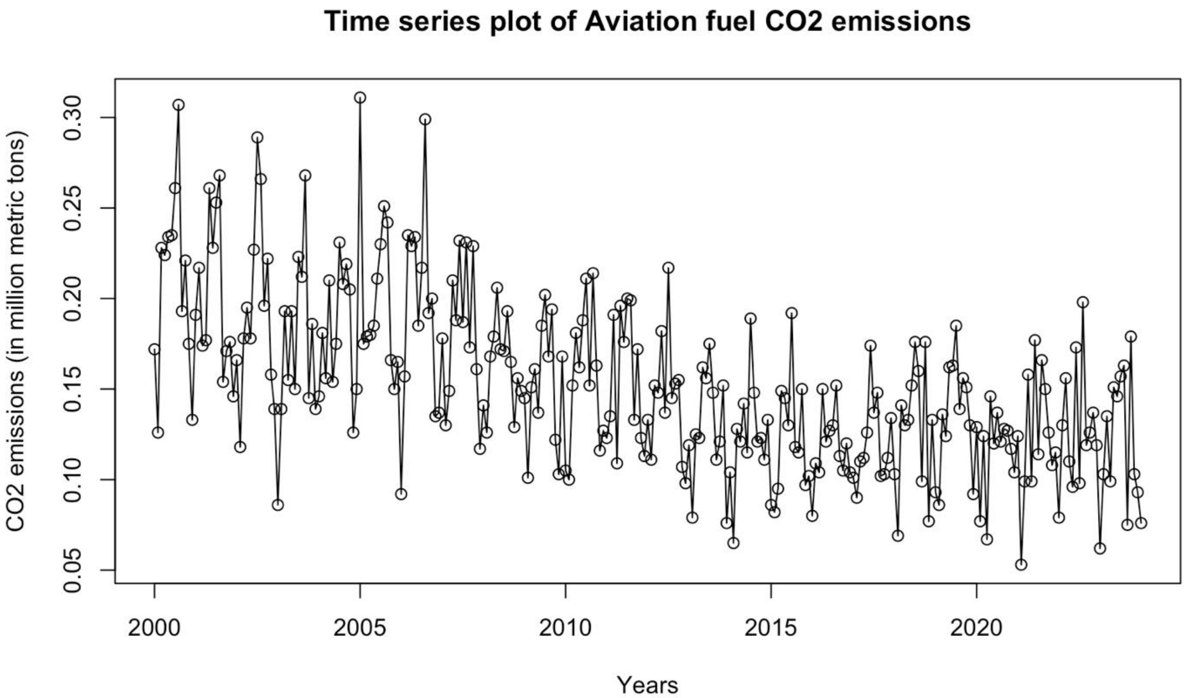


fig2. Time Series of Aviation Fuel emissions

The 5 main characteristics of the series is as follows (fig 2) :

1. **Trend** - There is a gradual downward trend observed in the plot.
2. **Seasonality** - There is a presence of seasonality as we can see repeating up-down patterns after an interval of time. This may be attributed to the seasonal nature in which flight schedules operate.
3. **Changing variance** - There is noticeable changing variance in the plot. The beginning of the plot has a high variance when compared to the end.
4. **Change point** - There is no single significant intervention point in the series. There is however some high fluctuations in value seen in the period around 2005.
5. **Behaviour** - Since there is seasonality in the series we can't comment on its behaviour. However, you can see the presence of both moving average and autoregressive behaviour in the series.

Time Series Vs First Lag Series

We can also compare the series with the first lag of the series and see how an observation is related to the previous observation via a scatter plot.

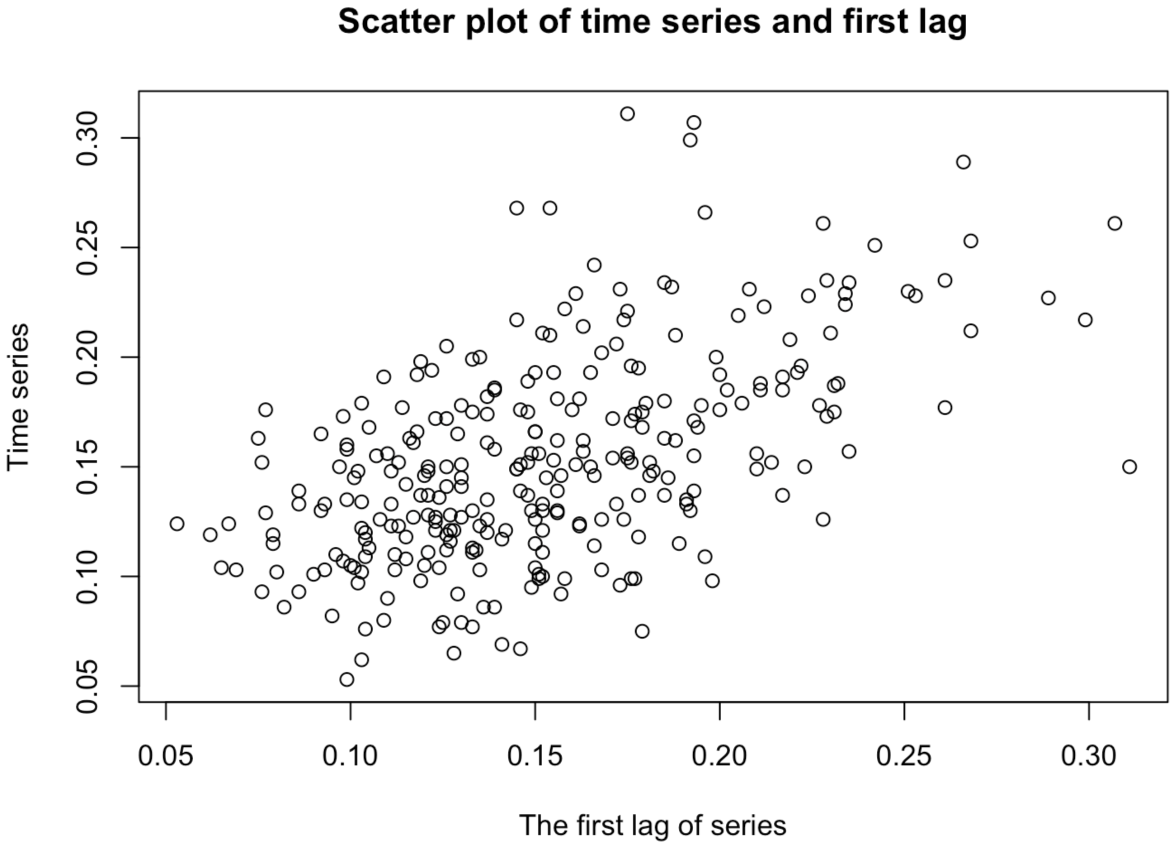


fig3. Scatter Plot of Time series vs First lag

Here, by just glancing at the above plot in Fig 3, we can tell that the time series is not strongly correlated with the lag series as the data points don't follow a straight line. A 0.54 correlation indicates a moderate positive relationship, suggesting that the previous period's values have a noticeable but not strong influence on the current period's values. This can be seen in the Time series plot in Fig 2, as not all data points follow the previous points in value (Autoregressive), some observations are completely opposite to the previous observation (Moving average).

ACF Plot for Frequency

Since the time series plot showed presence of seasonality let us see if we can see the seasonal patterns in the ACF plot.

```
> seasonal_acf(aviationTS,lag.max = 100, main="ACF Plot for Aviation CO2 emissions series")
```

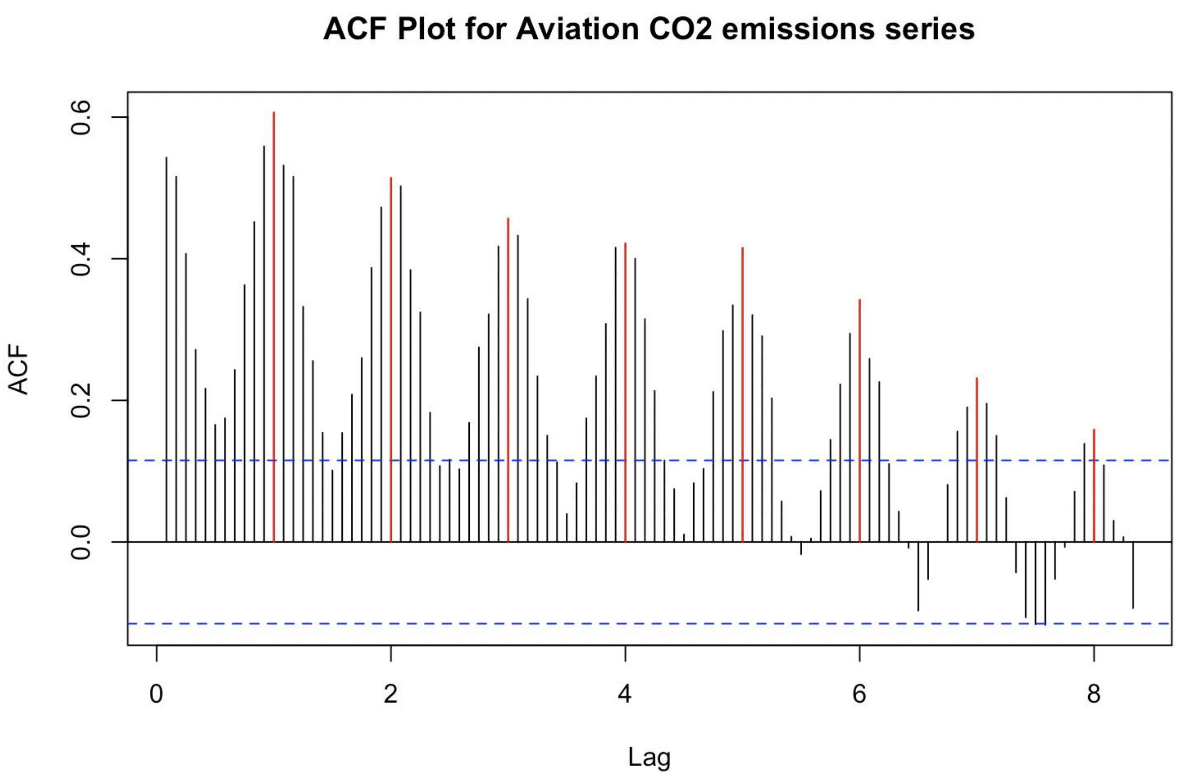


fig4. ACF Plot for aviation Co2 emission series

The seasonality is very clear in the above plot (fig 4). We can see the seasonal wave pattern, with the red line indicating the highest point of the season. If you count between the highest point of one pattern to the next, we will get the frequency i.e. 12 which is in line with what we define in the above sections. Another important thing to note is that there is a downward decline in the seasonal pattern, this clearly shows the presence of **seasonal trend**.

Due to the presence of seasonal trend, we will apply SARIMA model specification to try to find the most suitable model for our series. SARIMA models have two sets of orders to specify (P, D, Q) for seasonal component and (p, d, q) for ordinary component.

Model Specification

Seasonal Component (P, D, Q)

We need to filtrate out the effect of seasonality to see the autocorrelation effect more clearly. To do this, we first fit a plain model with only the first seasonal difference with order D = 1.

```
> m1.aviation = Arima(aviationTS,order=c(0,0,0),
+                     seasonal=list(order=c(0,1,0), period=12))
```

We then plot the residuals of this model as shown below.

```
> res.m1 = residuals(m1.aviation)
> plot(res.m1,xlab='Time',ylab='Residuals',main="Time series plot of D=1 residuals")
```

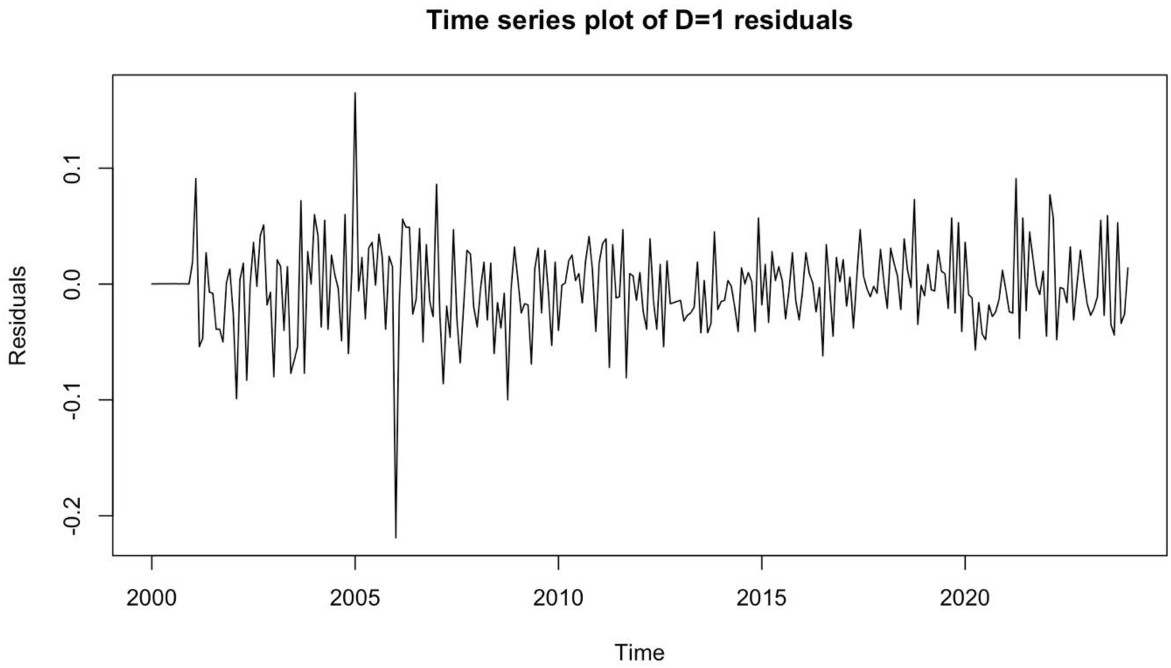


fig5. Time series plot of D=1 residuals

In Fig 5 we no longer see signs of seasonal trend. Also, in the ACF and PACF plot (Fig 6) of the residuals you do not see the declining seasonal pattern as seen in Fig 4 and now the autocorrelations at the seasonal lags are more clearly visible.

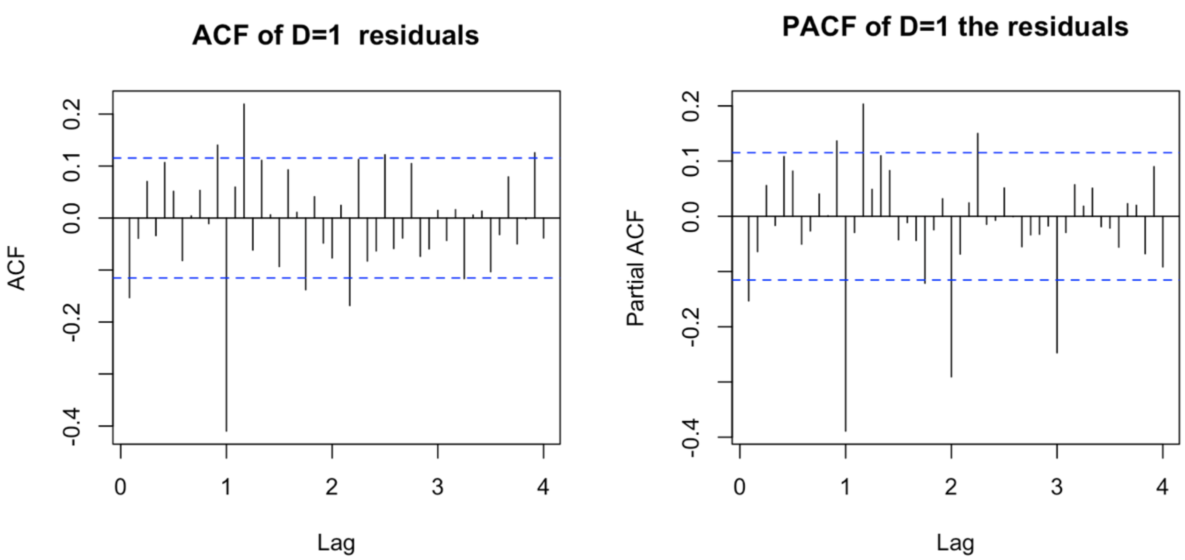


fig6. ACF - PACF Plot for D=1 residuals

We will now determine P and Q from the ACF and PACF plot, for this we need to identify the number of significant bars at seasonal lags. In ACF there is only one significant lag at season 1 so Q=1. In PACF, we see significant lags at seasons 1, 2 and 3. However we noticed a declining pattern here and decided to set P=0.

We will now fit the model P, D, Q = 0,1,1 and plot its residuals.

```
> m2.aviation = Arima(aviationTS,order=c(0,0,0),
+                    seasonal=list(order=c(0,1,1), period=12))
> res.m2 = residuals(m2.aviation)
> plot(res.m2,xlab='Time',ylab='Residuals',main="Time series plot of the PDQ=011 residuals")
```

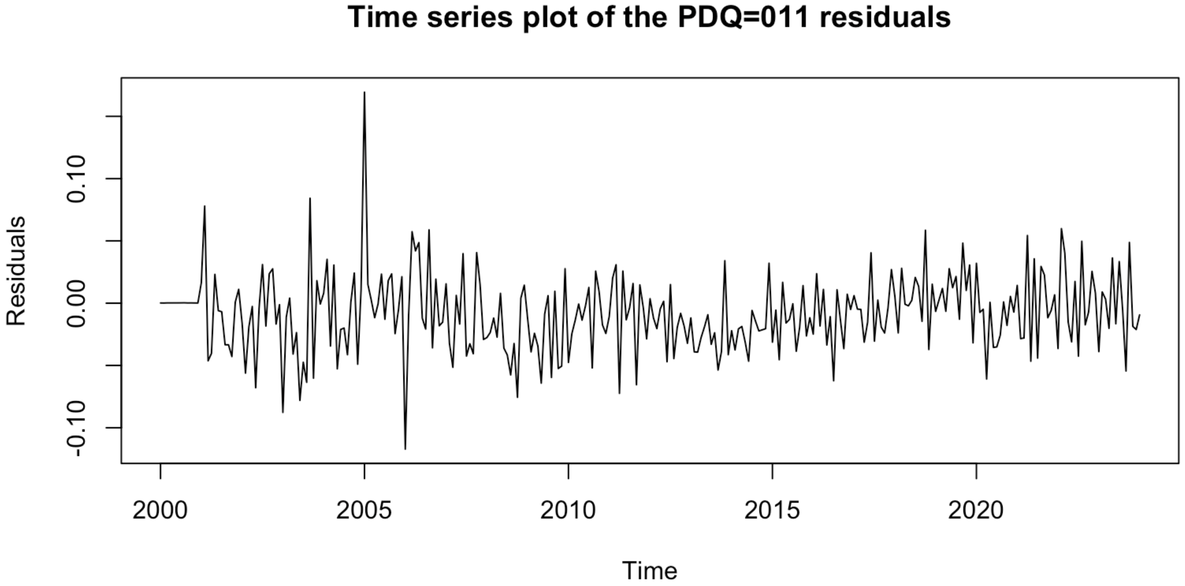
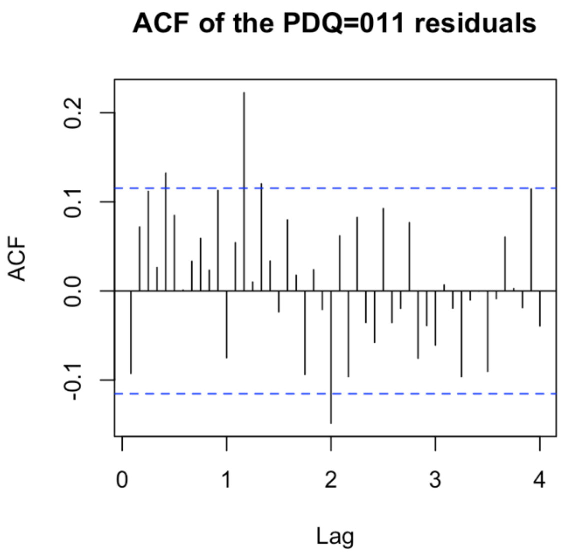


fig7. Time series plot of the PDQ=011 residuals

The Fig 7 plot of the residuals of P, D, Q = 0,1,1 model looks similar to the D=1 model in Fig 5. We don't see trend or seasonality. We do see some high fluctuations around the year 2005. This as we mentioned before was a time when there were some high dips and peaks in values in the original time series.



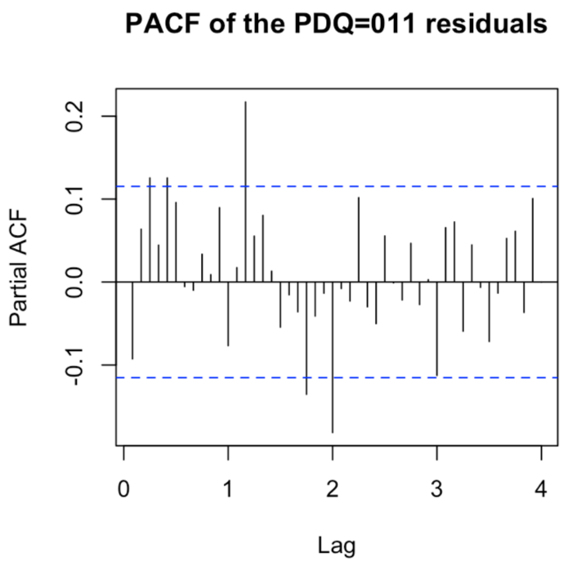


fig8. ACF and PACF plot of the PDQ=011 residuals

If we check the ACF and PACF of the residuals in Fig 8 you will see that although the second seasonal lag is significant, we can still conclude that seasonality has been filtered out since seasonal trend is not evident in these plots. The significant second season may be a result of the changing variance which was seen in the series and the high fluctuations in the period of 2005. This may disappear when we determine the remaining orders of the ARIMA part of the model. Therefore we can conclude that the final orders for Seasonal P,D,Q is 0,1,1.

Ordinary Component (P, D, Q)

In this section of model specification, we will determine the orders of ARIMA p,d,q values. Before we start with various model specification techniques, we need to first determine whether the series is stationary or not.

Check for Stationarity

There are multiple techniques we can apply to check whether the series is stationary or not. If we look at the series in Fig 7 you can tell the series looks like it has a flat mean level with no major trend. Also form the ACF and PACF plot of the residuals in Fig 8 we can see that :

- In ACF plot there are not many significant autocorrelations, no slowly decaying pattern is seen and therefore it looks like there is no trend.
- In PACF there is no significant first lag which is further proof of no trend.

We can also perform statistical tests such as ADF and PP test to confirm stationarity.

```
> adf.test(res.m3)

        Augmented Dickey-Fuller Test

data:  res.m3
Dickey-Fuller = -4.7132, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
> pp.test(res.m3)

        Phillips-Perron Unit Root Test

data:  res.m3
Dickey-Fuller Z(alpha) = -370.06, Truncation lag parameter = 5, p-value = 0.01
alternative hypothesis: stationary
```

Interpretation:

For both tests the null hypothesis and alternate hypothesis are as follows:

- Ho: The series is nonstationary /series has unit root
  Ha: The series is stationary /series has no unit root

From the above ADF test results we observe that the p value is 0.01 which is less than level of significance 0.05 hence we reject the null hypothesis at 5% level of significance and conclude that the series is stationary. This is also confirmed by the pp test.

Based on the evidence collected we can conclude that our residual series is stationary.

Transformation

In the descriptive analysis of the series, we noticed the presence of changing variance. We also noticed that the ACF and PACF plot of residuals in Fig 8 has significant lag at the second season. Hence, we decided that applying a Transformation to our raw series would help stabilise the variations. Let us first determine the normality of the series prior transformation. We will use QQ plot and Shapiro Wilk test to do this.

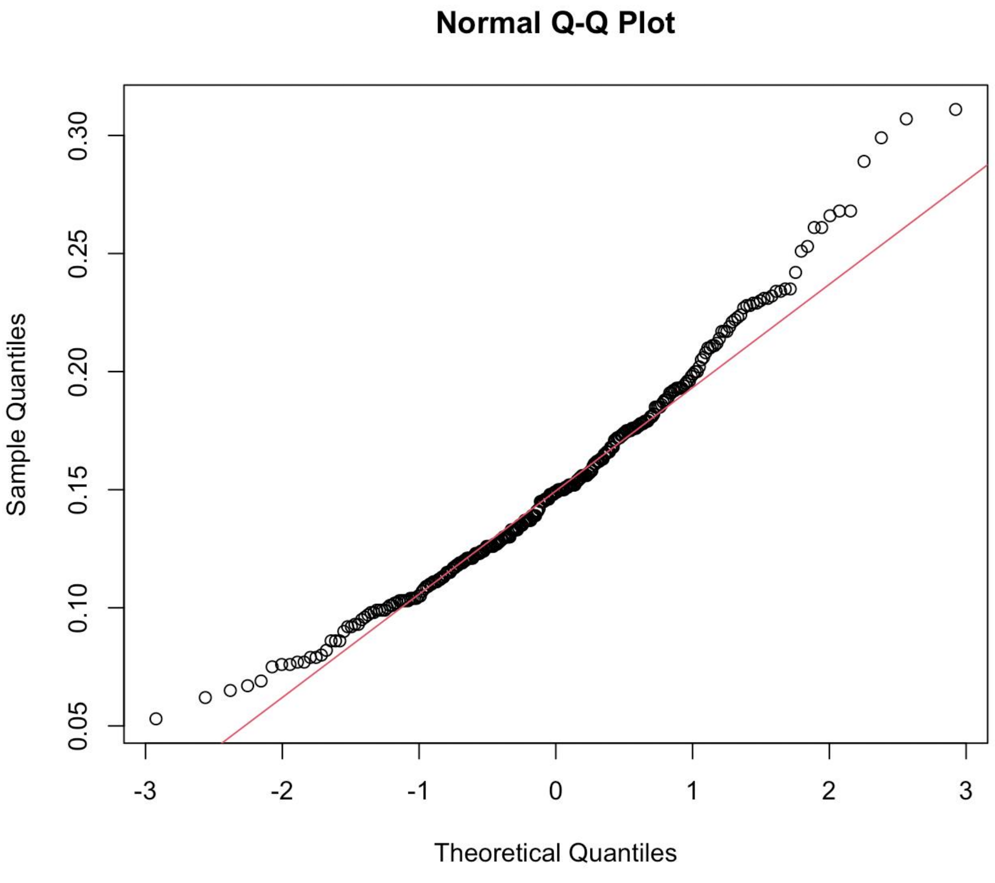


fig 9. QQ Plot of raw series

```
> shapiro.test(aviationTS)

        Shapiro-Wilk normality test

data:  aviationTS
W = 0.97334, p-value = 3.271e-05
```

Interpretation:

- From the qq plot (fig 9) we can observe that the data points don't align with the reference line also from the Shapiro wilk test we get a p-value that is less than 0.05 level of significance hence we reject the null hypothesis that the data is normally distributed. Hence, we can conclude that the dataset is not normally distributed which is also evident by the statistical mean and median value difference.

Now we will perform Box-Cox Transformation on the raw series.

```
> BC.aviation= (aviationTS^lambda-1)/lambda
> BC = BoxCox.ar(aviationTS)
> lambda <- BC$lambda[which(max(BC$loglike) == BC$loglike)]
> lambda
[1] 0.4
> BC.aviation = (aviationTS^lambda-1)/lambda
```

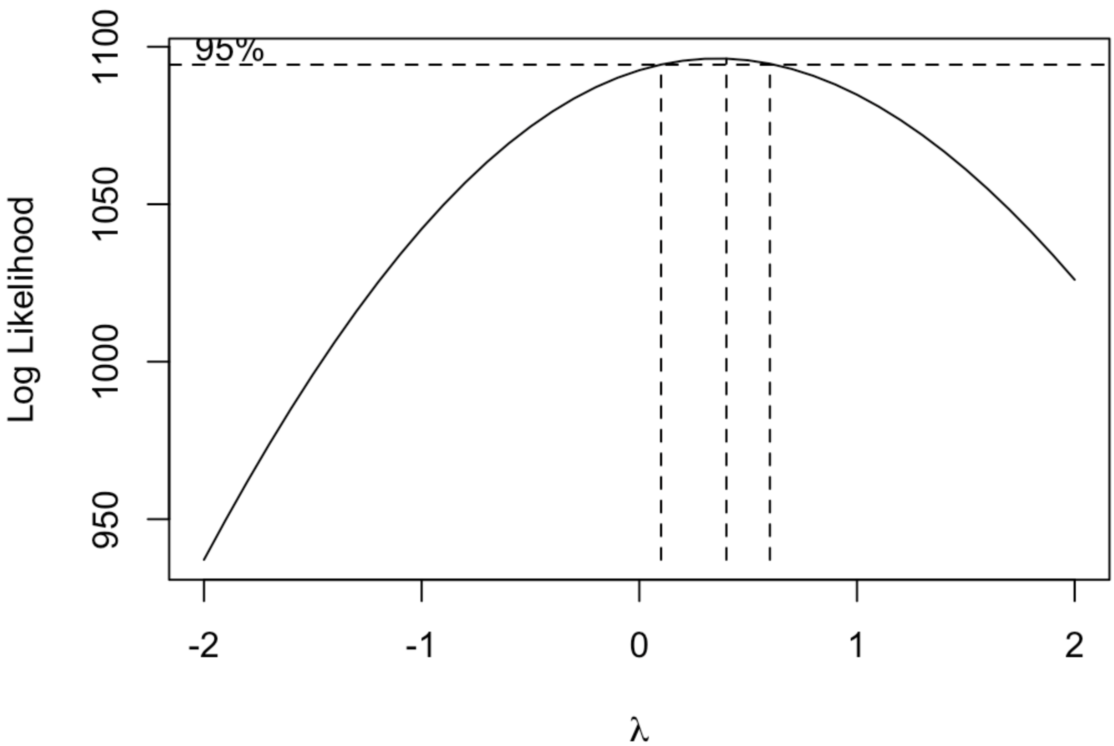


fig10. Lambda value

The lambda value of the transformed series is 0.4. We will now see if transformation had a positive impact on the normality of the series.

```
> shapiro.test(BC.aviation)

	Shapiro-Wilk normality test

data:  BC.aviation
W = 0.99651, p-value = 0.7793
```

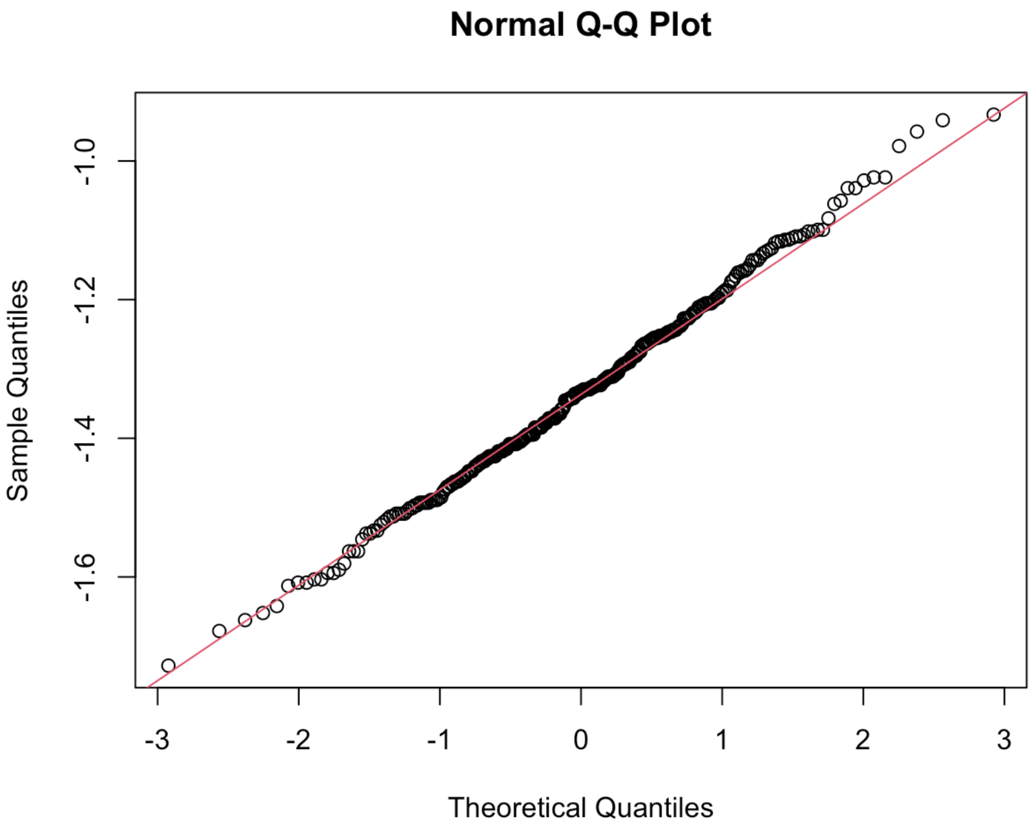


fig11. QQ Plot of transformed series

Interpretation:

- From the qq plot (fig 11) we can see that our original data was right skewed but after transformation we can see that the data values fall in line with the reference line indicating that the transformed data is normally distributed also.
- In Shapiro Wilk test we get a p-value of 0.77 which is greater than 0.05 hence we do not have sufficient evidence to reject the null hypothesis and we can assume normality.

From the above interpretation, we can conclude that Transformation had a positive impact on the series. We will now apply the seasonal PDQ=011 model on the Box-Cox transformed series and use the residuals of this model for p, d, q model specification.

```
> m3.aviation = Arima(BC.aviation,order=c(0,0,0),seasonal=list(order=c(0,1,1), period=12))
> res.m3 = residuals(m3.aviation);
> plot(res.m3,xlab='Time',ylab='Residuals',main="Time series plot PDQ + BC residuals")
```

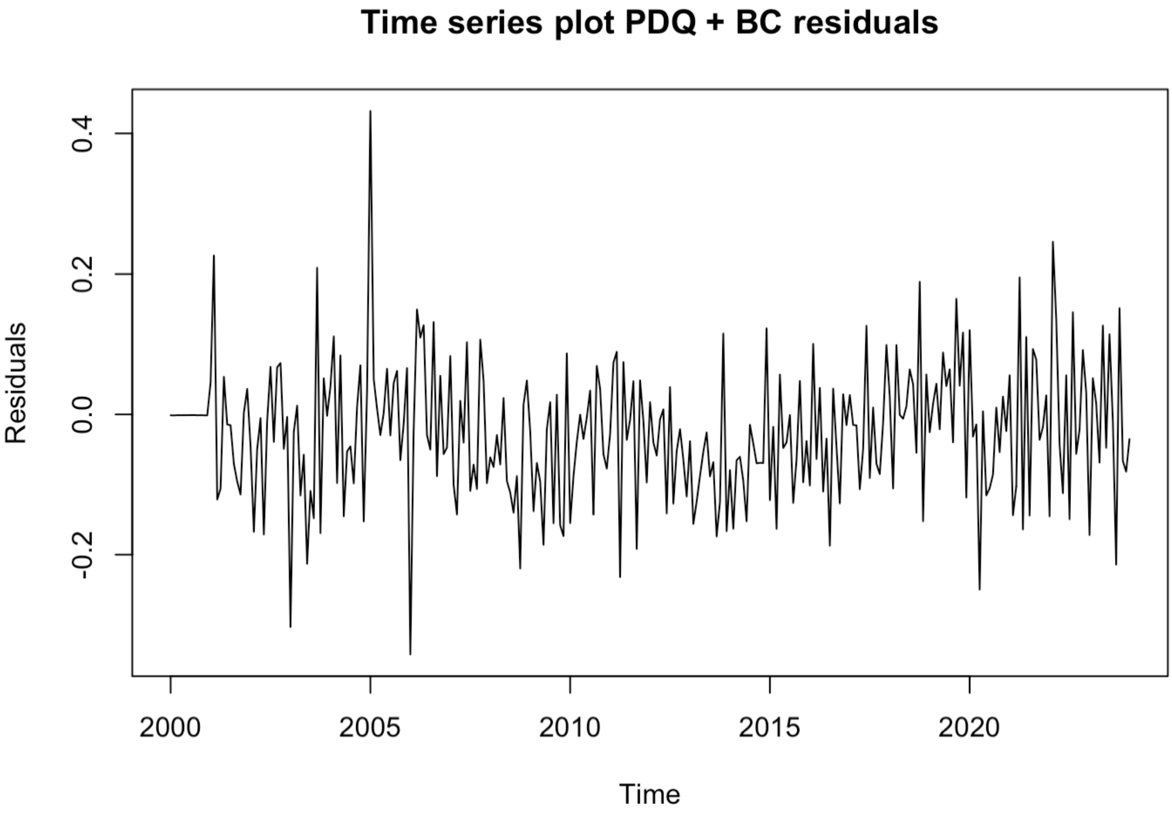


fig12. Time series of PDQ + BC residuals

We will be using the residuals in Fig 12 to determine the orders of p, d and q.

d=0
Since the series was determined to be stationary and therefore has a flat mean level, we do not need to perform differencing and therefore the value of d will be 0.

ACF-PACF Plot

Interpretation:

- From the ACF and APCF plot for residual (fig 13) we observe that there is no seasonal pattern or declining pattern observed in ACF also the first lag in PACF plot is not highly significant hence we can conclude that the series is stationary hence value of d=0.
- From the ACF plot we observe there are two significant lags between first season hence the value of q=2.
- Similarly, from the PACF plot we observe there are two significant lags between first season hence the value of p=2
- Therefore, from ACF and PACF plot the value of ordinary component (p, d, q) is (2, 0, 2)

Set of possible models from ACF and PACF plot are {SARIMA (2,0,2) x (0,1,1)}

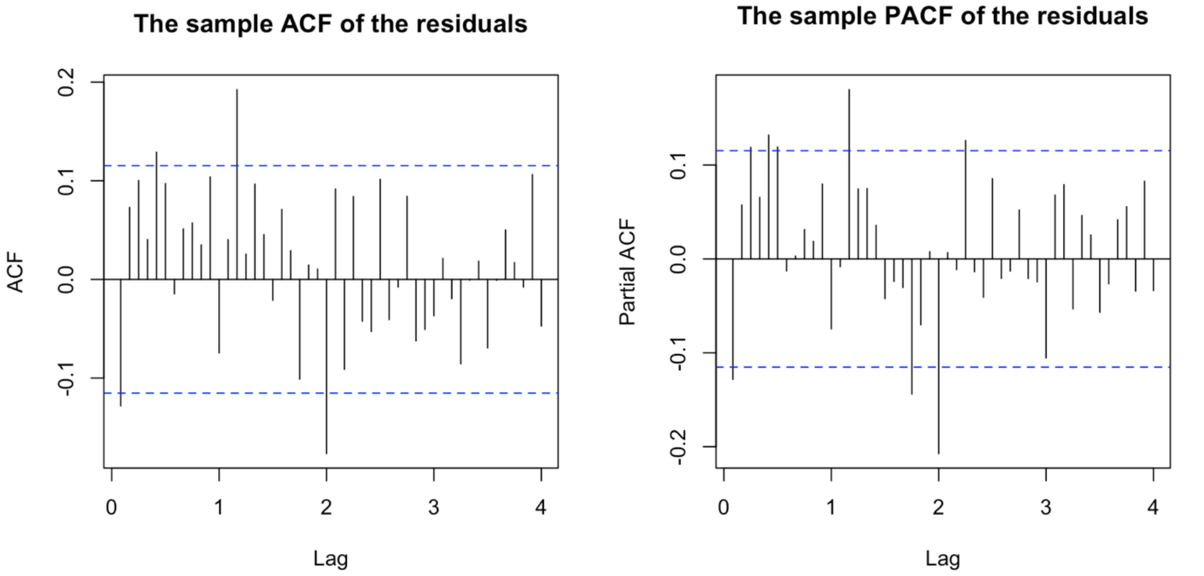


fig13. ACF and PACF of residual (res.m3 SARIMA(0,0,0)x(0,1,1))

EACF Plot

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o x o o o o o o  o  o  x
1 x o o o o x o o o o o  o  o  x
2 x x o o o o o o o o o  o  o  x
3 x x o o o o o o o o o  o  o  x
4 x x o x o o o o o o o  o  o  x
5 x x x x o o o o o o o  o  o  o
6 x o x x x o o o o o o  o  o  x
7 x o x o x o o o o o o  o  o  o
```

fig14. EACF plot

<u>Interpretation:</u>

- From the above EACF plot (fig 14) we see that the top left most 0 not blocked by X falls at (0,1). This vertex will help us get the set of possible models, considering all its nearest neighbours.

Set of possible models from ACF and PACF plot are {SARIMA (1,0,1) x (0,1,1), SARIMA (1,0,2) x (0,1,1), SARIMA (0,0,2) x (0,1,1), SARIMA (0,0,1) x (0,1,1)}

BIC table

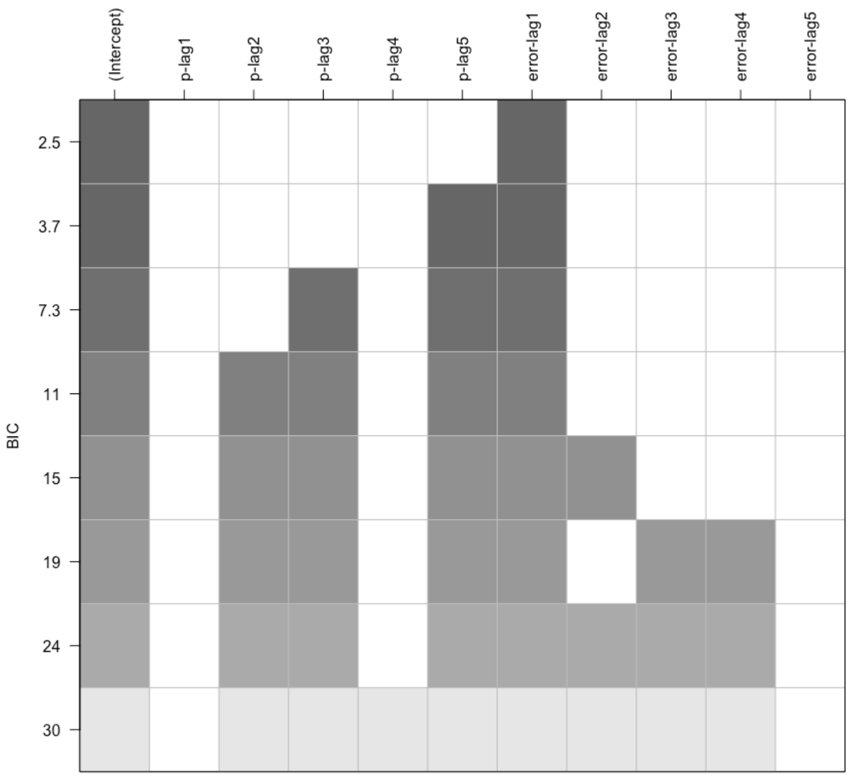


fig15. BIC table

Interpretation:

- From the BIC table (fig 15) we observe that the best model is p=0 and q=1(dark shaded region). We also got this (p, q) model from EACF plot.
- The second-best model is p=5 and q=1. This model is also well supported by the underlying models and the difference in BIC value between this model and the best model is small therefore we ill consider it.
- The third best model is p=3 and q=1. This is also well supported and has a small BIC difference with the 1st and 2nd model which is why we will consider it in our analysis.

Set of possible models from ACF and PACF plot are {SARIMA (5,0,1) x (0,1,1), SARIMA (3,0,1) x (0,1,1),

Thus, the final set of all possible SARIMA models for our time series are:

- SARIMA(2,0,2)x(0,1,1),
- SARIMA(0,0,1)x(0,1,1) ,
- SARIMA(1,0,1)x(0,1,2),
- SARIMA(1,0,2)x(0,1,1),
- SARIMA(0,0,2)x(0,1,1)
- SARIMA(5,0,1)x(0,1,1) ,
- SARIMA(3,0,1)x(0,1,1)

Parameter Estimation

Each model's coefficients were estimated using 3 different methods: Maximum Likelihood (ML), Conditional Sum of Squares (CSS) and a combination of the two (CSS-ML).

SARIMA (0,0,1) x(0,1,1)

Let us first start with the model with 1 MA component and 0 AR component.

```
> m001.aviation = Arima(BC.aviation,order=c(0,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "ML")
> coeftest(m001.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ma1  -0.045753   0.056366  -0.8117    0.417
sma1 -0.594485   0.049055 -12.1188   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m001CSS.aviation = Arima(BC.aviation,order=c(0,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS")
> coeftest(m001CSS.aviation)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1  -0.044771   0.056242  -0.796    0.426
sma1 -0.591118   0.047279 -12.503   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m001CSSML.aviation = Arima(BC.aviation,order=c(0,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS-ML")
> coeftest(m001CSSML.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ma1  -0.045750   0.056366  -0.8117    0.417
sma1 -0.594486   0.049055 -12.1189   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

fig16 SARIMA (0,0,1) x(0,1,1)

Interpretation:

- From the above model SARIMA (0,0,1) x (0,1,1), we see that the non seasonal moving average component does not have a substantial impact on the model which is indicated by the high p-value for ML, CSS and CSS-ML model.

- The seasonal parameter MA(1) is highly significant across all methods (p-values < 2e-16), highlighting the strong influence of the seasonal component on the aviation time series data.

SARIMA(0,0,2)x(0,1,1)

Let's see if adding another MA component will improve the model coefficients.

```
> # SARIMA(0,0,2)x(0,1,1)
> m002.aviation = Arima(BC.aviation,order=c(0,0,2),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "ML")
> coeftest(m002.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ma1  -0.074013   0.064658  -1.1447  0.25234
ma2   0.137607   0.057950   2.3746  0.01757 *
sma1 -0.625617   0.048579 -12.8785  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m002CSS.aviation = Arima(BC.aviation,order=c(0,0,2),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS")
> coeftest(m002CSS.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ma1  -0.072348   0.064560  -1.1206  0.26244
ma2   0.141178   0.058237   2.4242  0.01534 *
sma1 -0.623233   0.046743 -13.3332  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m002CSSML.aviation = Arima(BC.aviation,order=c(0,0,2),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS-ML")
> coeftest(m002CSSML.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ma1  -0.074091   0.064657  -1.1459  0.25183
ma2   0.137599   0.057948   2.3745  0.01757 *
sma1 -0.625620   0.048575 -12.8794  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

fig17 SARIMA (0,0,2) x(0,1,1)

Interpretation:

- From the above model SARIMA(0,0,2)x(0,1,1) we see that this model captures seasonality very well which is indicated by the highly significant seasonal component SMA(1).

- The first moving average MA(1) parameter is not significant, while the second moving average MA(2) is marginally significant, indicating that first-order moving average component does not contribute significantly to the model.

SARIMA(1,0,2)x(0,1,1)

To the above model we will now add 1 AR component to see if there is any improvement.

```
> m102.aviation = Arima(BC.aviation,order=c(1,0,2),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "ML")
> coeftest(m102.aviation)

z test of coefficients:

       Estimate Std. Error  z value  Pr(>|z|)    
ar1   0.9974194  0.0042668 233.7625 < 2.2e-16 ***
ma1  -1.1347870  0.0601530 -18.8650 < 2.2e-16 ***
ma2   0.2241151  0.0604235   3.7091  0.000208 ***
sma1 -0.9003960  0.0471750 -19.0863 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m102CSS.aviation = Arima(BC.aviation,order=c(1,0,2),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS")
> coeftest(m102CSS.aviation)

z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)    
ar1   0.986184   0.017436  56.5615 < 2.2e-16 ***
ma1  -1.089679   0.060701 -17.9516 < 2.2e-16 ***
ma2   0.228607   0.057056   4.0067 6.158e-05 ***
sma1 -0.790235   0.038652 -20.4450 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m102CSSML.aviation = Arima(BC.aviation,order=c(1,0,2),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS-ML")
> coeftest(m102CSSML.aviation)

z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)    
ar1   0.997380   0.004319 230.9281 < 2.2e-16 ***
ma1  -1.133148   0.060228 -18.8144 < 2.2e-16 ***
ma2   0.222864   0.060465   3.6858  0.000228 ***
sma1 -0.900229   0.047157 -19.0902 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

fig18 SARIMA (1,0,2) x(0,1,1)

Interpretation:

- From the above fitted SARIMA model SARIMA(1,0,2)x(0,1,1) we observe that all the parameters (AR(1), MA(1), MA(2), and seasonal MA(1)) are significant which indicate that autoregressive, moving average and seasonal component is necessary for accurately modelling the aviation time series data which is better captured by this model.

- This looks like a good model, but further residual analysis needs to be performed to determine the best model.

SARIMA(1,0,1) x(0,1,1)

Let us see if decreasing the parameter by 1 i.e. having 1 AR and 1 MA model gives us the same desirable estimates as the previous model.

```
> m101.aviation = Arima(BC.aviation,order=c(1,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "ML")
> coeftest(m101.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1  -0.404777   0.357977  -1.1307   0.2582
ma1   0.332062   0.368186   0.9019   0.3671
sma1 -0.592857   0.047743 -12.4175   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m101CSS.aviation = Arima(BC.aviation,order=c(1,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS")
> coeftest(m101CSS.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1  -0.235283   0.361190  -0.6514   0.5148
ma1   0.160308   0.342022   0.4687   0.6393
sma1 -0.583044   0.046656 -12.4968   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m101CSSML.aviation = Arima(BC.aviation,order=c(1,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS-ML")
> coeftest(m101CSSML.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1  -0.408241   0.358306  -1.1394   0.2546
ma1   0.335626   0.369088   0.9093   0.3632
sma1 -0.592929   0.047742 -12.4194   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

fig19 SARIMA (1,0,1) x(0,1,1)

<u>Interpretation:</u>

- From the above model SARIMA(1,0,1)x(0,1,1), the AR(1) and MA(1) parameters are not significant indicated by the high p-value > 0.05 for all 3 methods. Whereas the seasonal MA(1) parameter is significant which suggest that the seasonal component is crucial, but the non-seasonal components are not well-defined.
- SARIMA(1,0,1)x(0,1,1) is not significant compared to SARIMA(1,0,2)x(0,1,1) which indicates that reducing an MA component may have led to the model not capturing the series well as it did before.

SARIMA (2,0,2) x (0,1,1)

We will now increase the SARIMA(1,0,2)x(0,1,1) by one AR parameter and check model estimates.

```
> m202.aviation = Arima(BC.aviation,order=c(2,0,2),
+                     seasonal=list(order=c(0,1,1),
+                                   period=12),method = "ML")
> coeftest(m202.aviation)

z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)
ar1   1.085268   0.270537   4.0115 6.033e-05 ***
ar2  -0.087414   0.269932  -0.3238    0.7461
ma1  -1.217977   0.258898  -4.7045 2.545e-06 ***
ma2   0.300898   0.241321   1.2469    0.2124
sma1 -0.903630   0.048593 -18.5959 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m202CSS.aviation = Arima(BC.aviation,order=c(2,0,2),
+                        seasonal=list(order=c(0,1,1),
+                                      period=12),method = "CSS")
> coeftest(m202CSS.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1  -0.174301   0.271949  -0.6409   0.5216
ar2  -0.255085   0.220824  -1.1552   0.2480
ma1   0.097906   0.264026   0.3708   0.7108
ma2   0.348712   0.193484   1.8023   0.0715 .
sma1 -0.591347   0.048785 -12.1216   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m202CSSML.aviation = Arima(BC.aviation,order=c(2,0,2),
+                          seasonal=list(order=c(0,1,1),
+                                        period=12),method = "CSS-ML")
> coeftest(m202CSSML.aviation)

z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)
ar1   1.094237   0.267832   4.0855 4.397e-05 ***
ar2  -0.096598   0.267166  -0.3616    0.7177
ma1  -1.226801   0.255168  -4.8078 1.526e-06 ***
ma2   0.309763   0.237373   1.3050    0.1919
sma1 -0.902262   0.048036 -18.7832 < 2.2e-16 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

fig20 SARIMA (2,0 ,2) x(0,1,1)

Interpretation:

- From the above model SARIMA (2,0,2) X (0,1,1) we see that only one pair of AR and MA term is statistically significant at 5% level of significance which suggests that there is significant autocorrelation in the time series.
- The seasonal MA (1) parameter is highly significant (p-value < 2.2e-16) across all methods, highlighting the strong seasonal component in the data.
- The ML and CSS-ML methods yield similar estimates and significance levels for the parameters indicating they produce consistent results. While the CSS approach tends to

give different significant parameters as this approach might not fully capture the underlying process compared to the ML method,

SARIMA(3,0,1)x(0,1,1)
Here we will increase AR to 3 components.

```
> m301.aviation = Arima(BC.aviation,order=c(3,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "ML")
> coeftest(m301.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1   0.752528   0.068772  10.9424  < 2e-16 ***
ar2   0.173016   0.074467   2.3234  0.02016 *
ar3   0.071176   0.066821   1.0652  0.28680
ma1  -0.883991   0.035375 -24.9889  < 2e-16 ***
sma1 -0.899861   0.047285 -19.0308  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m301CSS.aviation = Arima(BC.aviation,order=c(3,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS")
> coeftest(m301CSS.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1   0.688132   0.048830  14.0925  < 2e-16 ***
ar2   0.177399   0.070988   2.4990  0.01245 *
ar3   0.111094   0.059448   1.8687  0.06166 .
ma1  -0.863842   0.034347 -25.1501  < 2e-16 ***
sma1 -0.817250   0.034286 -23.8360  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m301CSSML.aviation = Arima(BC.aviation,order=c(3,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS-ML")
> coeftest(m301CSSML.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1   0.752481   0.068776  10.9411  < 2e-16 ***
ar2   0.173003   0.074465   2.3233  0.02016 *
ar3   0.071223   0.066821   1.0659  0.28648
ma1  -0.883938   0.035391 -24.9766  < 2e-16 ***
sma1 -0.899817   0.047266 -19.0374  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

fig21 SARIMA (3,0,1) x(0,1,1)

Interpretation:

- From the above model SARIMA(3,0,1) x(0,1,1) AR(1), MA(1), and SMA(1) are highly significant with p-values < 2e-16, While AR(2) is marginally significant and AR(3) is not significant. While AR(2) may have some impact, AR(3) does not appear to significantly enhance the model

SARIMA(5,0,1)x(0,1,1)

Here we will increase AR to 5 components.

```
> m501.aviation = Arima(BC.aviation,order=c(5,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "ML")
> coeftest(m501.aviation)

z test of coefficients:

       Estimate Std. Error  z value Pr(>|z|)
ar1   0.7280819  0.0775502   9.3885  < 2e-16 ***
ar2   0.1645226  0.0747293   2.2016  0.02769 *
ar3   0.0565164  0.0754362   0.7492  0.45374
ar4  -0.0088705  0.0755732  -0.1174  0.90656
ar5   0.0559512  0.0697637   0.8020  0.42255
ma1  -0.8598520  0.0514221 -16.7214  < 2e-16 ***
sma1 -0.9054837  0.0492808 -18.3740  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m501CSS.aviation = Arima(BC.aviation,order=c(5,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS")
> coeftest(m501CSS.aviation)

z test of coefficients:

       Estimate Std. Error  z value  Pr(>|z|)
ar1   0.6925196  0.0800807   8.6478 < 2.2e-16 ***
ar2   0.1620373  0.0756490   2.1420  0.032197 *
ar3   0.1750879  0.0664141   2.6363  0.008381 **
ar4  -0.0443698  0.0763075  -0.5815  0.560930
ar5   0.0018758  0.0727067   0.0258  0.979417
ma1  -0.8727006  0.0600507 -14.5327 < 2.2e-16 ***
sma1 -0.8018966  0.0366202 -21.8977 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> m501CSSML.aviation = Arima(BC.aviation,order=c(5,0,1),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS-ML")
> coeftest(m501CSSML.aviation)

z test of coefficients:

       Estimate Std. Error  z value Pr(>|z|)
ar1   0.7281088  0.0775521   9.3886  < 2e-16 ***
ar2   0.1645293  0.0747293   2.2017  0.02769 *
ar3   0.0565146  0.0754361   0.7492  0.45375
ar4  -0.0088783  0.0755733  -0.1175  0.90648
ar5   0.0559277  0.0697649   0.8017  0.42275
ma1  -0.8598658  0.0514226 -16.7216  < 2e-16 ***
sma1 -0.9054741  0.0492816 -18.3735  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

fig22 SARIMA (5,0,1) x(0,1,1)

Interpretation:

- From the above model SARIMA(5,0,1)x(0,1,1) for the ML and CSS-ML method we see that The autoregressive (AR) parameters AR(1) and moving average (MA) parameters MA(1) and seasonal MA (sma1) are highly significant with p-values < 2e-16. The AR(2) parameter is marginally significant. However, AR(3), AR(4), and AR(5) are not significant.
- Similarly, from the CSS method we observe that AR(1), MA(1), and SMA(1) are highly significant AR(2) is marginally significant (p-value = 0.03), and AR(3) becomes significant (p-value = 0.008381), whereas AR(4) and AR(5) remain non-significant.
- The non-significance of the other parameters AR(3), AR(4), and AR(5) across most methods indicates these higher-order autoregressive terms do not contribute meaningfully to the model.

Model Selection

Now that we have fit the different SARIMA models and estimated the coefficients, we can use these models to calculate performance metrics such as AIC & BIC scores and accuracy metrics such as Mean average, Root mean square etc. We can use these metrics to find the best fitting model.

Below code snippet shows the AIC and BIC scores for each model.

```
> sort.score(AIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
+                m202.aviation,m301.aviation,m501.aviation), score = "aic")
              df       AIC
m102.aviation  5 -526.8044
m202.aviation  6 -524.9255
m301.aviation  6 -524.7603
m501.aviation  8 -521.4579
m002.aviation  4 -475.0466
m001.aviation  3 -471.5371
m101.aviation  4 -470.5155
> sort.score(BIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
+                m202.aviation,m301.aviation,m501.aviation), score = "bic")
              df       BIC
m102.aviation  5 -508.6843
m202.aviation  6 -503.1814
m301.aviation  6 -503.0162
m501.aviation  8 -492.4658
m001.aviation  3 -460.6650
m002.aviation  4 -460.5506
m101.aviation  4 -456.0194
```

fig23 AIC BIC score

Interpretation:

- From the above output we observe that SARIMA (1,0,2)x(0,1,1) model has the lowest AIC score of (-526.8044) and has the lowest BIC score of (-508.684) indicating as the best model.

In the below output, we have the performance of the model for each error metric.

```
> df.Smodels
                                   ME       RMSE        MAE          MPE     MAPE      MASE          ACF1
SARIMA(0,0,1)x(0,1,1)_12 -0.026337861 0.09910561 0.07672466  1.609424674 5.749426 0.8405817 -0.0827947424
SARIMA(0,0,2)x(0,1,1)_12 -0.025376129 0.09798915 0.07607199  1.532872279 5.706514 0.8334311 -0.0521043701
SARIMA(1,0,2)x(0,1,1)_12 -0.003971966 0.08690473 0.06703981 -0.012985482 5.078721 0.7344761  0.0047591240
SARIMA(1,0,1)x(0,1,1)_12 -0.026396269 0.09893624 0.07653759  1.615328725 5.738113 0.8385321 -0.0558017688
SARIMA(2,0,2)x(0,1,1)_12 -0.003708607 0.08682947 0.06698300 -0.031767417 5.074996 0.7338537  0.0005117846
SARIMA(3,0,1)x(0,1,1)_12 -0.003863392 0.08691645 0.06696247 -0.021646541 5.072572 0.7336288 -0.0006702371
SARIMA(5,0,1)x(0,1,1)_12 -0.003990772 0.08672372 0.06669304 -0.009542307 5.052560 0.7306770  0.0009187621
```

fig24 Error measure

Interpretation:

- From the above output for error measures, we see that the Mean error that measures the average error forecast value is low for SARIMA(2,0,2)x(0,1,1) followed by SARIMA(3,0,1)x(0,1,1) and SARIMA(1,0,2)x(0,1,1).
- The Root mean squared error which indicates the square root of the average of squared forecast errors has the lowest error value for SARIMA(5,0,1)x(0,1,1) followed by SARIMA(2,0,2)x(0,1,1), SARIMA(3,0,1)x(0,1,1) and SARIMA(1,0,2)x(0,1,1).
- Based on MAE that is mean absolute error which reflects the average absolute error the lowest value is for SARIMA(5,0,1)x(0,1,1) followed by SARIMA(3,0,1)x(0,1,1) and SARIMA(2,0,2)x(0,1,1)
- The error measure Mean percentage error that measures the average percentage deviation of the forecast is lowest for the model SARIMA(5,0,1)x(0,1,1).
- The Mean average percentage error is lowest for the model SARIMA(5,0,1)x(0,1,1) followed by SARIMA(3,0,1)x(0,1,1) and SARIMA(2,0,2)x(0,1,1).
- From the Mean absolute scaled error model SARIMA(5,0,1)x(0,1,1) has the least value lined up by SARIMA(3,0,1)x(0,1,1), SARIMA(1,0,2)x(0,1,1) and SARIMA(2,0,2)x(0,1,1)
- From the ACF value that measures the randomness in the residual we see that SARIMA(2,0,2)x(0,1,1) has a value close to zero indicate no autocorrelation.

Based on the interpretation, SARIMA(5,0,1)x(0,1,1) appears to have the best performance for most error metrics, showing the lowest error. However, this model lacks significant coefficients, is not statistically significant overall, and doesn't achieve the best AIC/BIC scores. Additionally, it has the highest number of parameters among the models we considered, which increases the risk of overfitting. Given our goal to find the best model with the fewest parameters, we concluded that SARIMA(5,0,1)x(0,1,1) is not the most suitable model for our series.

A common factor across almost all error metrics and AIC/BIC scores is that SARIMA(3,0,1)x(0,1,1), SARIMA(2,0,2)x(0,1,1), and SARIMA(1,0,2)x(0,1,1) exhibited similar values. These three models rank highest for AIC and BIC scores and demonstrate low error metrics, making them strong candidates for our series. Among them, SARIMA(1,0,2)x(0,1,1) stands out with all significant coefficients, the highest AIC/BIC score, and the fewest parameters. Considering these points, we concluded that SARIMA(1,0,2)x(0,1,1) is the most suitable model for our series based on parameter estimation. In the next section, we will perform residual analysis for each model to verify if these findings are supported and if SARIMA(1,0,2)x(0,1,1) effectively captures the series' properties.

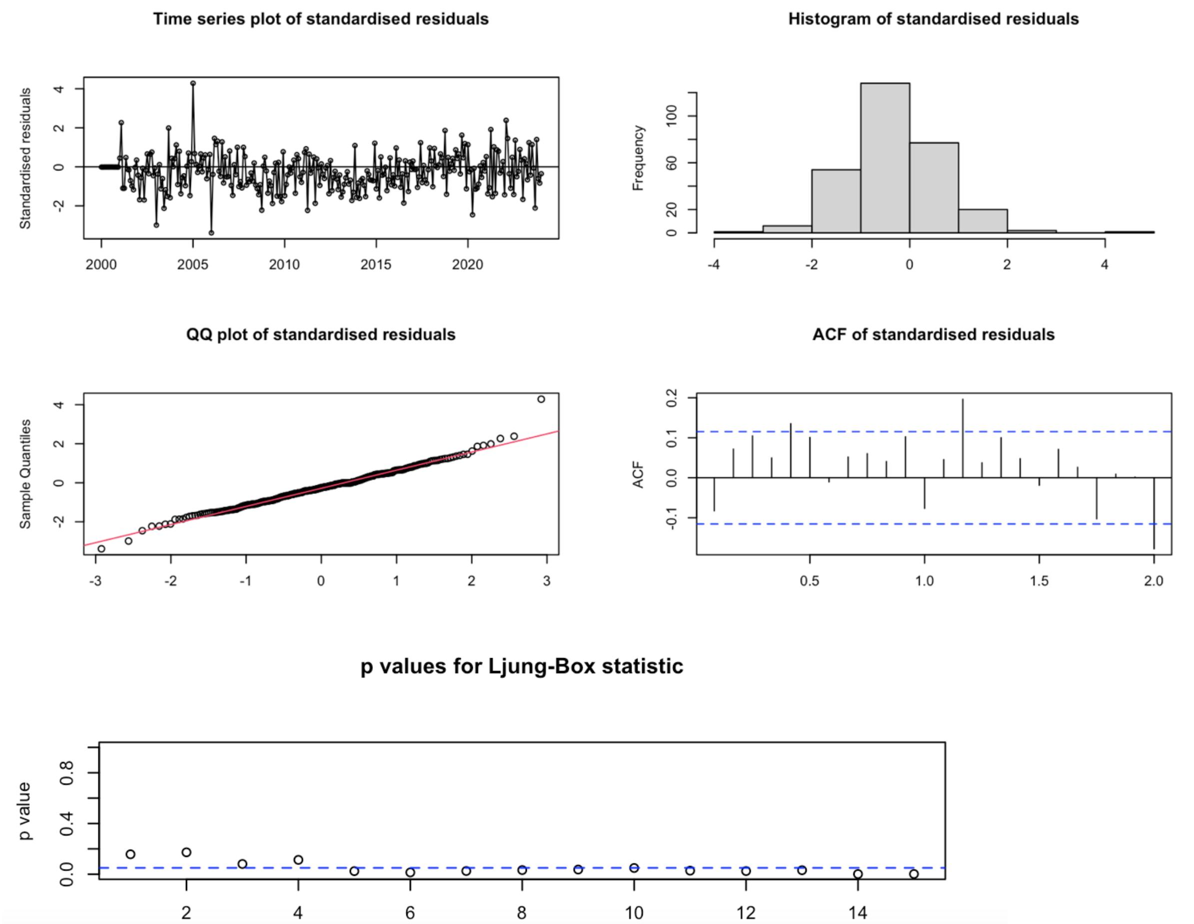
Residual Analysis

A detailed residual analysis was done for each time series model using the following tests :

- **A time series plot of the residuals** which will tell us if any of the 5 main characteristics of the series is left behind in the residuals. An ideal scenario is to see the residuals randomly distributed around 0 with no trend.
- **Histogram of the residuals** which tells us if the residuals a have an asymmetric distribution. A good model will have symmetric distribution between +3 and –3.
- **Normality of the residuals** via QQ plot and Shapiro-Wilk test. A good model will have normally distributed residuals.
- **Ljung-Box test** and **ACF plot** for checking if there is autocorrelation left in residuals.

SARIMA (0,0,1) x (0,1,1)

Below shows the residuals of SARIMA (0,0,1) x (0,1,1).



```
> residual.analysis(model=m001.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.98455, p-value = 0.00332

> Box.test(st.res.m001,type="Ljung-Box")

        Box-Ljung test

data:  st.res.m001
X-squared = 2.0017, df = 1, p-value = 0.1571
```

fig25. Residual Analysis of SARIMA (0,0,1) x (0,1,1)

Interpretation:

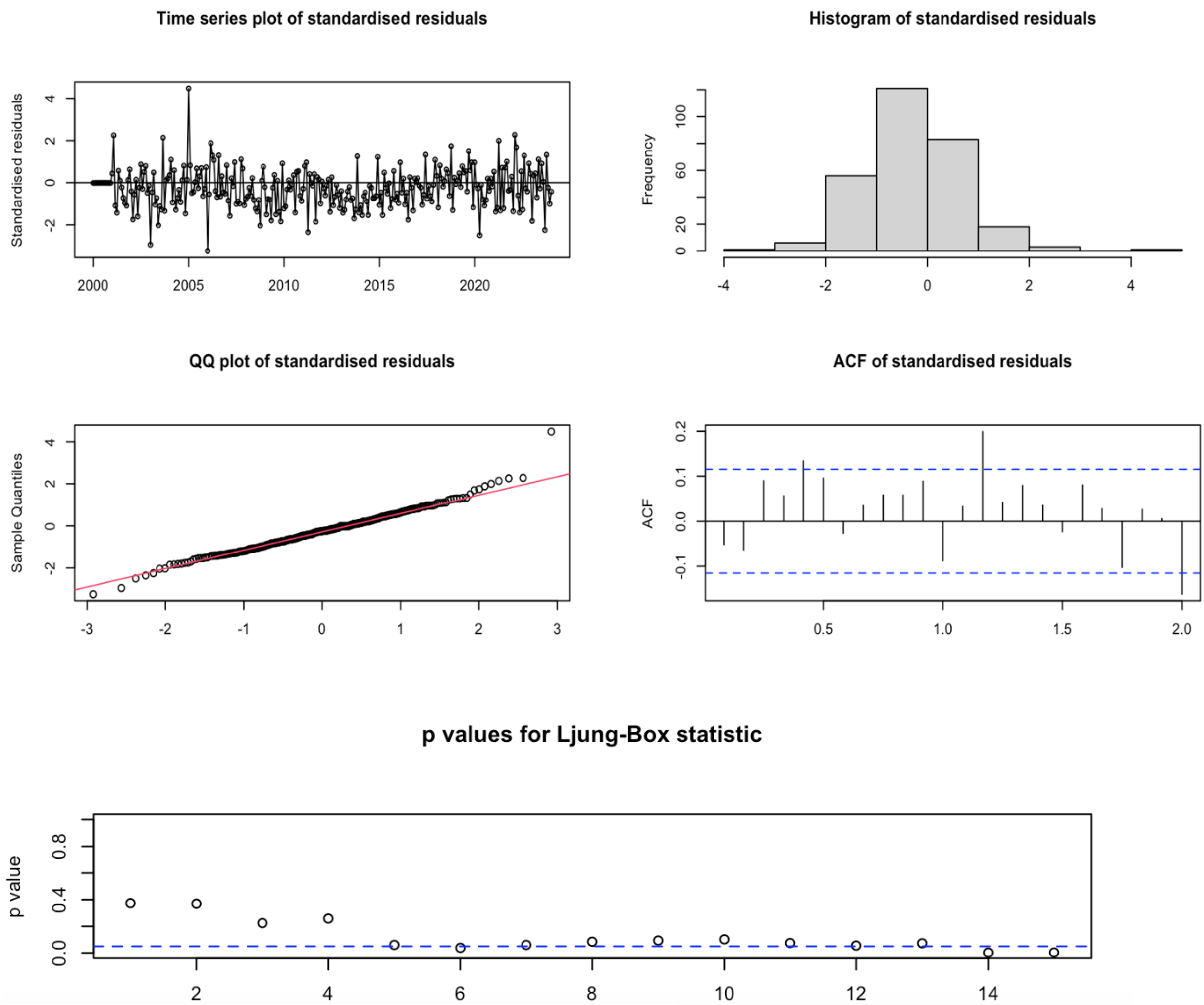
From the above fig 25, we can see that:

- The time series plot is random around 0, no trend, or seasonality is observed
- There are no significant lags in seasons except for season 2, and most are contained within the limits in ACF plot.
- The histogram looks uniformly distributed, however there are outliers seen and it is not within +3 and –3 interval.
- Most of the data points in the QQ-plot follow the line with few deviations in the ends, and the Shapiro-Wilks test p-value is <0.05 so we can reject the null hypothesis and state that the series is not normally distributed.
- Ljung-Box test, which determines where there is autocorrelation left shows most points are above or near the limit. Also p-value is greater than 0.05 which indicates that the model captures autocorrelation.

Residual analysis for SARIMA (0,0,1) x (0,1,1) shows that this model captures autocorrelation, but the residuals aren't normally distributed.

SARIMA (0,0,2) x (0,1,1)

Below image shows the residuals of SARIMA (0,0,2) x (0,1,1).




```
> residual.analysis(model=m002.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.98149, p-value = 0.000853

> Box.test(st.res.m002,type="Ljung-Box")

        Box-Ljung test

data:  st.res.m002
X-squared = 0.79277, df = 1, p-value = 0.3733
```

Fig26. Residuals of SARIMA (0,0,2) x (0,1,1).

Interpretation:

From the above fig 26, we can conclude that:

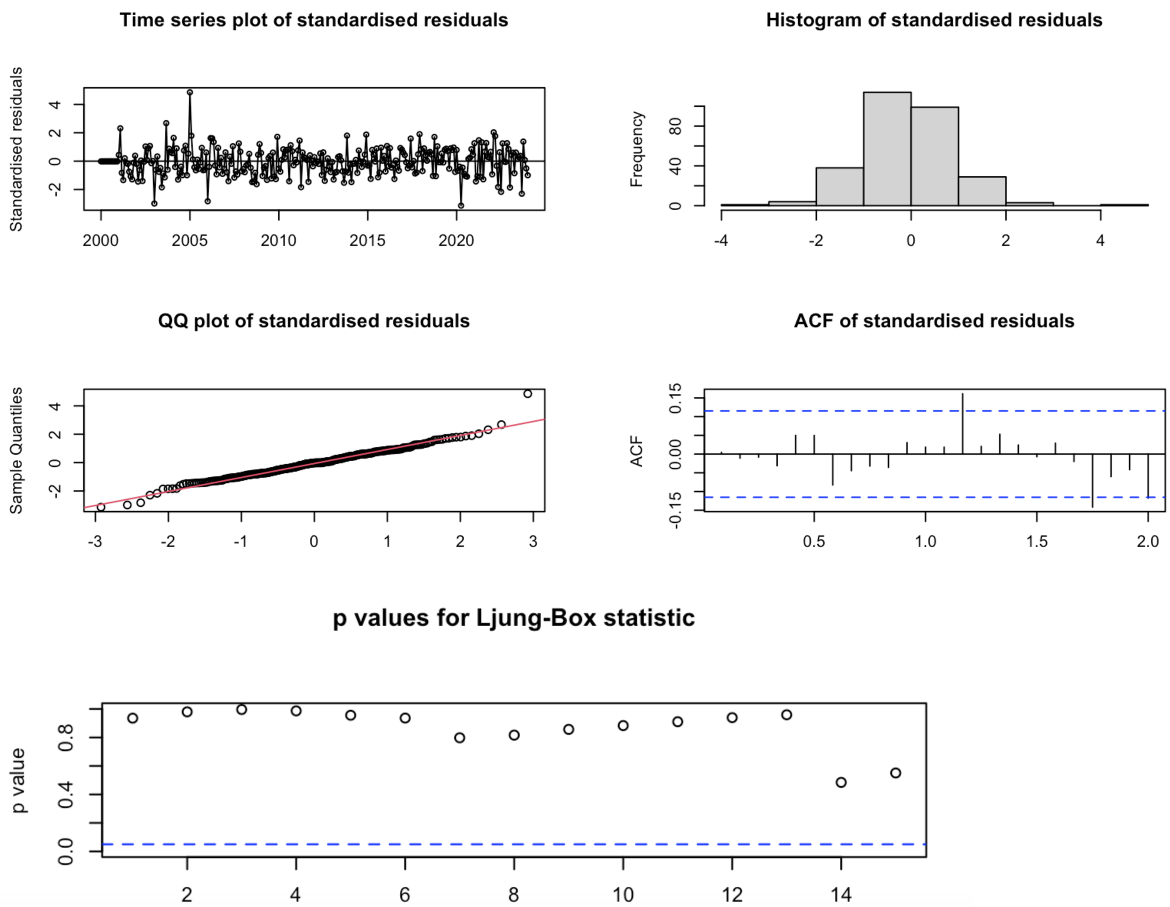
- The time series plot is random around 0 and you can't no trend, or seasonality is observed

- There are not many significant lags in the ACF plot.
- The histogram looks uniformly distributed, however it is not within +3 and –3 interval.
- Most of the data points in the QQ-plot follow the straight line with few deviations in the ends, however the Shapiro-Wilks test p-value is <0.05 so we can state that the series is not normally distributed.
- Ljung-Box test shows majority points above the limit which is a good sign. p-value is greater than 0.05 which indicates that the model captures autocorrelation well.

Residual analysis for SARIMA (0,0,2) x (0,1,1) shows that this model is similar to SARIMA (0,0,1) x (0,1,1) in that it captures autocorrelation but the residuals are not normally distributed.

SARIMA (1,0,2) x (0,1,1)

Below shows the residuals of SARIMA (1,0,2) x (0,1,1).




```
> residual.analysis(model=m102.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.98049, p-value = 0.0005565
```

```
> Box.test(st.res.m102,type="Ljung-Box")

        Box-Ljung test

data:  st.res.m102
X-squared = 0.0066138, df = 1, p-value = 0.9352
```

fig27. Residuals of SARIMA (1,0,2) x (0,1,1).

Interpretation:

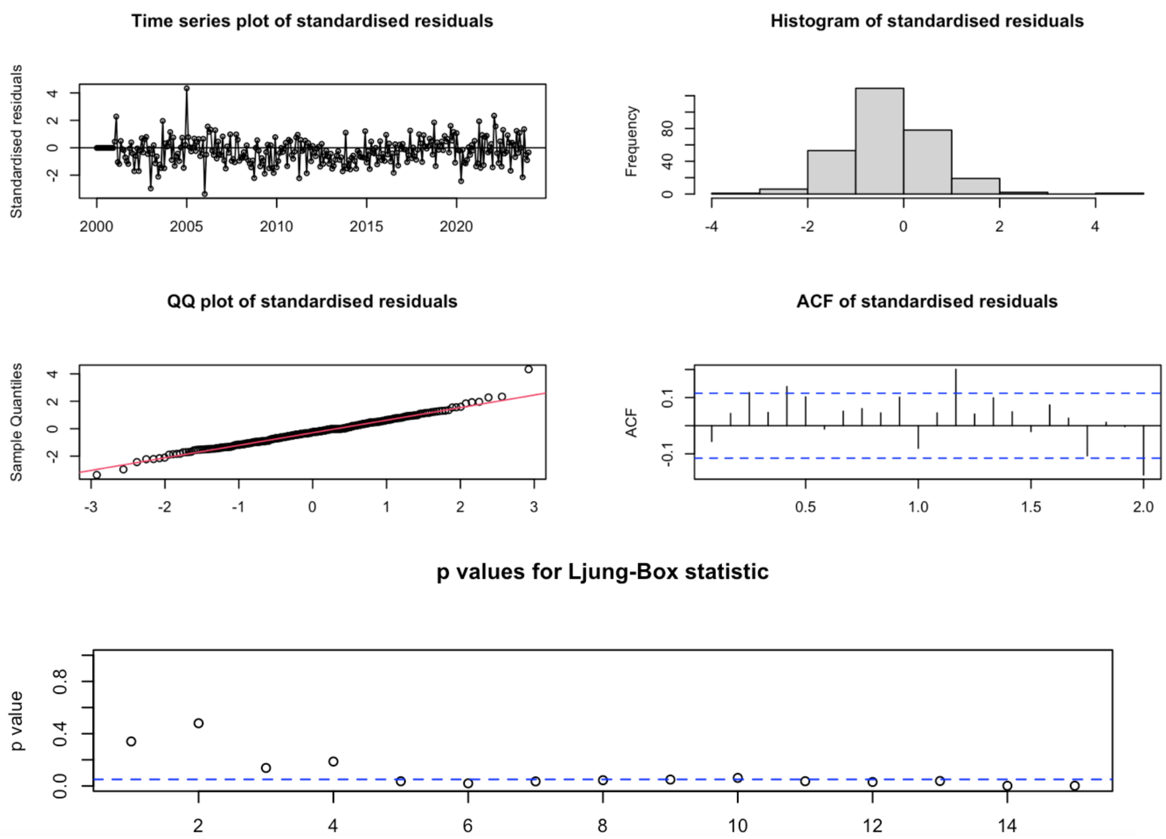
From the above fig 27, we can infer that:

- The time series plot is random around 0, no trend, or seasonality.
- Most of the lags are within the limit. It is also important to note that the lag in the second season is no longer significant and that the model accurately captures it.
- The histogram looks uniformly distributed, but it is not within +3 and –3 interval.
- Most of the data points in the QQ-plot follow the straight line with few deviations in the ends, however the Shapiro-Wilks test p-value is <0.05 so we can conclude that the series is not normally distributed.
- Here Ljung-Box test shows all points above the limit unlike the previous models. p-value is also greater than 0.05 which indicates that the model captures autocorrelation well.

Residual analysis for SARIMA (1,0,2) x (0,1,1) shows that this model is a good fit for our series as it captures autocorrelation well.

SARIMA (1,0,1) x (0,1,1)

Below shows the residuals of SARIMA (1,0,1) x (0,1,1).




```
> residual.analysis(model=m101.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.98354, p-value = 0.002102
> Box.test(st.res.m101,type="Ljung-Box")

        Box-Ljung test

data:  st.res.m101
X-squared = 0.90927, df = 1, p-value = 0.3403
```

fig28. Residual of SARIMA (1,0,1) x (0,1,1).

Interpretation:

From the above fig 28, we can infer that:

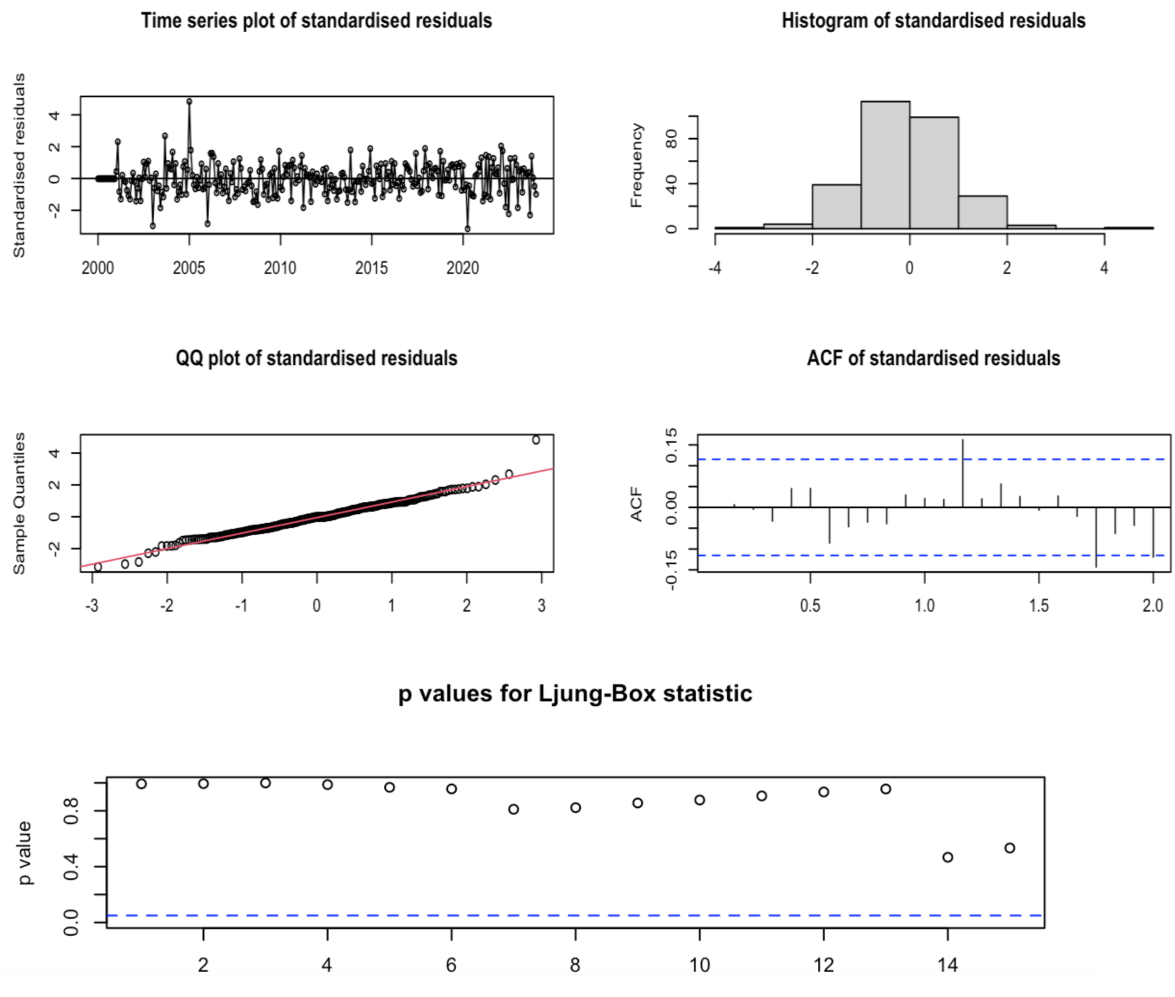
- The time series plot is random around 0 so there is no sign of trend, or seasonality.
- There are not many significant lags in the ACF plot, however this model doesnt capture the 2^{nd} seasonal lag unlike the previous model.
- Histogram appears symmetric; however, it has outliers and is not within the limit.
- QQ plot data points look normally distributed, however Shapiro-Wilks test p-value is <0.05 so the series is not normally distributed.

- Ljung-Box test shows only few points above the limit. The test p-value is greater than 0.05 which indicates that the model captures autocorrelation to a certain extent

Residual analysis for SARIMA (1,0,1)x(0,1,1) re-iterates that this model is a not as good as SARIMA (1,0,2)x(0,1,1) in capturing autocorrelation.

SARIMA (2,0,2) x (0,1,1)

Below image shows the residuals of SARIMA (2,0,2) x (0,1,1).




```
> residual.analysis(model=m202.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.98064, p-value = 0.0005915
```

```
> Box.test(st.res.m202,type="Ljung-Box")

	Box-Ljung test

data:  st.res.m202
X-squared = 7.6484e-05, df = 1, p-value = 0.993
```

fig29. Residuals of SARIMA (2,0,2) x (0,1,1).

Interpretation:

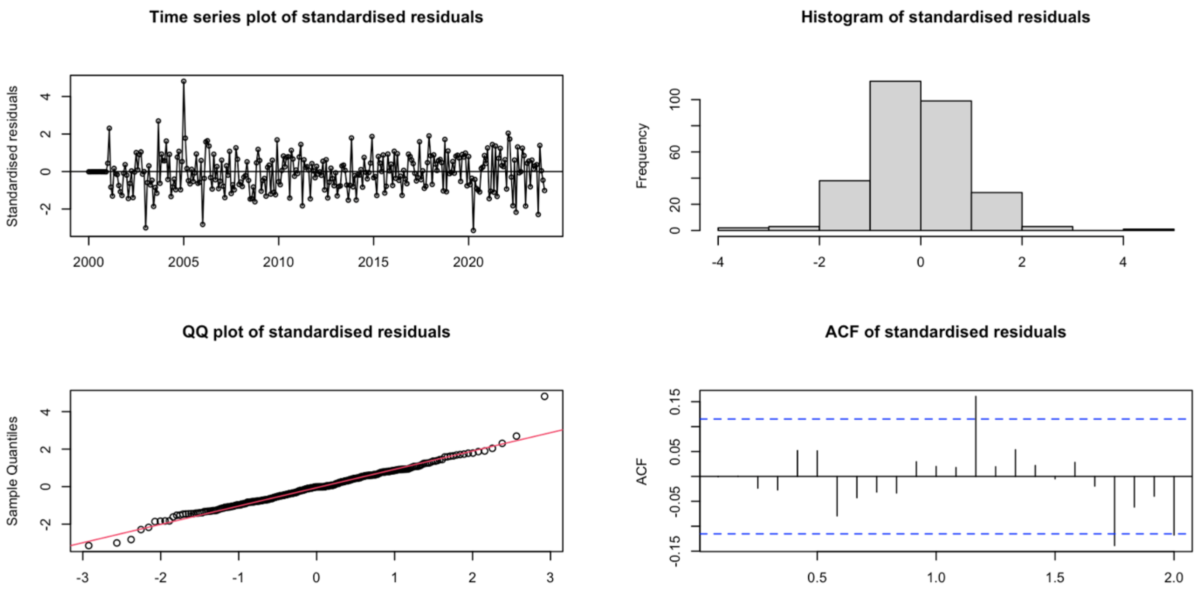
From the above fig 29, we can see that:

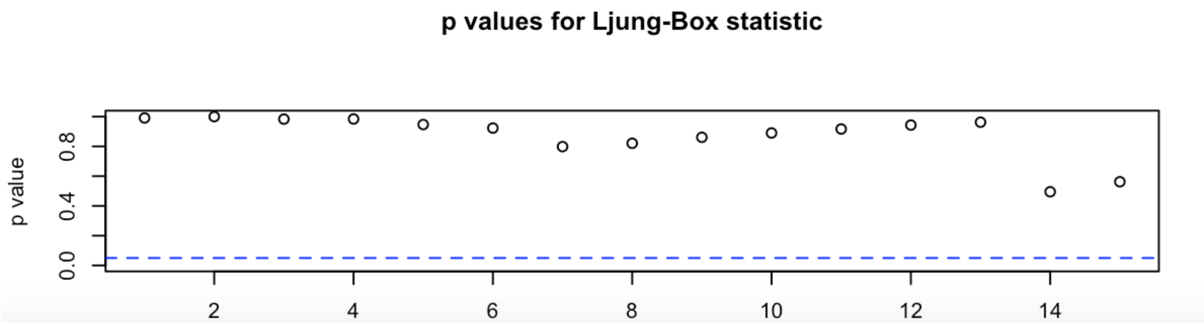
- The time series plot is random around 0, no trend, or seasonality.
- Most of the lags in ACF plot are within the limit and just like SARIMA (1,0,2) x (0,1,1), this model also captures the second seasonal lag.
- The histogram looks uniformly distributed, but it is not within +3 and –3 interval.
- Most of the data points in the QQ-plot follow the straight line with few deviations in the ends, however the Shapiro-Wilks test p-value is <0.05 so we can conclude that the series is not normally distributed.
- Here Ljung-Box test shows all points above the limit. p-value is also much greater than 0.05 which indicates that the model captures autocorrelation well.

Residual analysis for SARIMA (2,0,2) x (0,1,1) shows that this model could be a good fit for our series as is captures autocorrelation well.

SARIMA (3,0,1) x (0,1,1)

Below shows the residuals of SARIMA (3,0,1) x (0,1,1).






```
> residual.analysis(model=m301.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.9811, p-value = 0.0007217

> Box.test(st.res.m301,type="Ljung-Box")

        Box-Ljung test

data:  st.res.m301
X-squared = 0.00013118, df = 1, p-value = 0.9909
```

fig30. Residuals of SARIMA (3,0,1) x (0,1,1).

<u>Interpretation:</u>

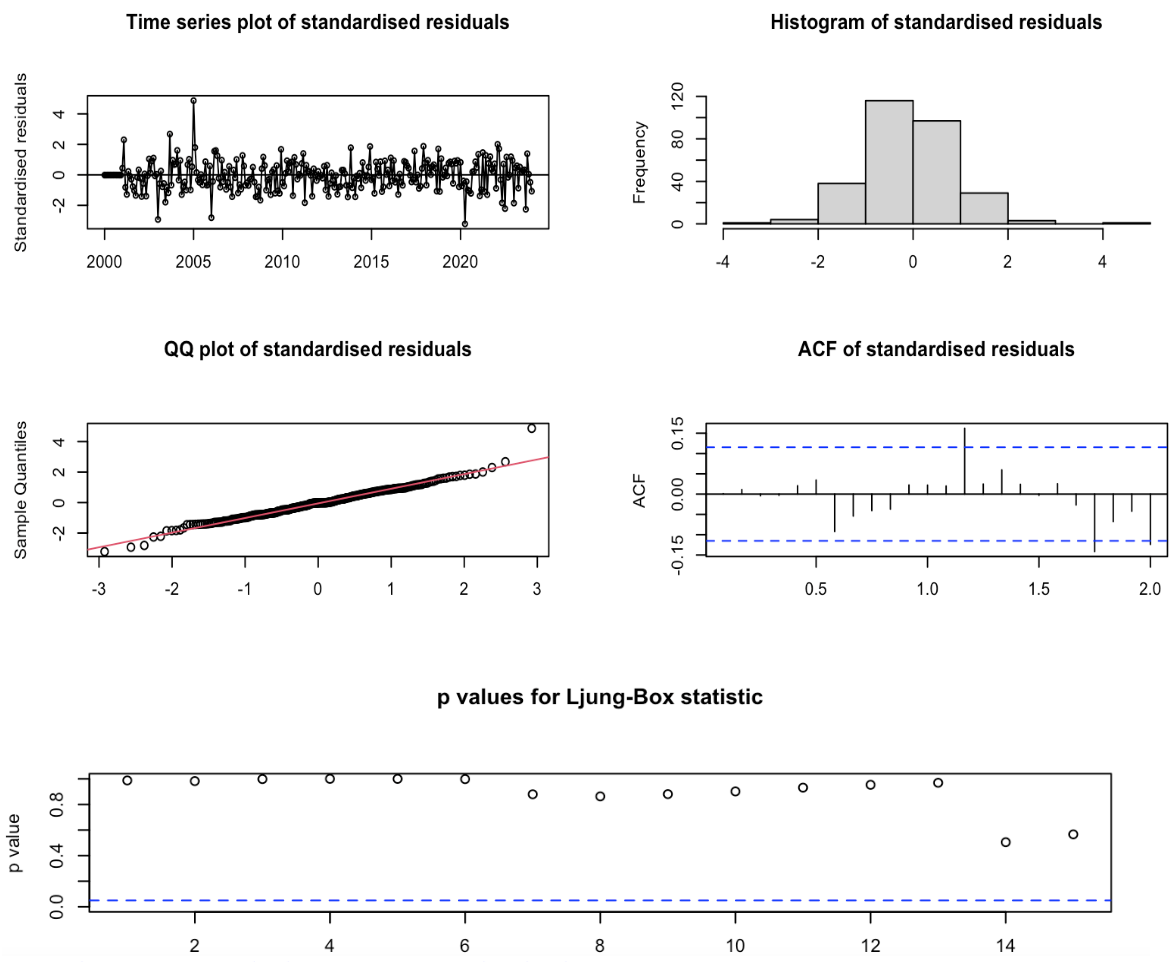
From the above fig 30, we can see that:

- The time series plot is random around 0, no trend, or seasonality.
- Most of the lags in ACF plot are within the limit and just like the previous, this model also captures the second seasonal lag.
- The histogram looks uniformly distributed, but it is not within +3 and –3 interval.
- Most of the data points in the QQ-plot follow the straight line with few deviations in the ends, however the Shapiro-Wilks test p-value is <0.05 so we can conclude that the series is not normally distributed.
- Ljung-Box test shows all points are above the limit and p-value is also much greater than 0.05 which indicates that the model captures autocorrelation well.

Residual analysis for SARIMA (3,0,1) x (0,1,1) shows that this model is a good fit for our series but as we identified in parameter estimation AR (3) is not significant and it does not appear to significantly enhance the model

SARIMA (5,0,1) x (0,1,1)

Below shows the residuals of SARIMA (5,0,1) x (0,1,1).




```
> residual.analysis(model=m501.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.97953, p-value = 0.0003712

> Box.test(st.res.m501,type="Ljung-Box")

        Box-Ljung test

data:  st.res.m501
X-squared = 0.00024649, df = 1, p-value = 0.9875
```

Fig31. Residuals of SARIMA (5,0,1) x (0,1,1).

<u>Interpretation:</u>

From the above fig 31, we can infer that:

- The time series plot is random around 0 and there is no sign of trend or seasonal patterns.
- Most of the lags in ACF plot are within the limit and just like the previous, this model also captures the second seasonal lag.
- The histogram looks uniformly distributed, but it is not within +3 and –3 interval.

- Most of the data points in the QQ-plot follow the straight line with few deviations in the ends, however the Shapiro-Wilks test p-value is <0.05 so we can conclude that the series is not normally distributed.
- Ljung-Box test shows all points are above the limit and p-value is also much greater than 0.05 which indicates that the model captures autocorrelation well.

Residual analysis for SARIMA (5,0,1) x (0,1,1) shows that this model is a good fit for our series as it captures autocorrelation well but as we identified in parameter estimation, AR (4) & AR (5) is not significant, and it does not appear to significantly enhance the model. Furthermore, it has a total of 6 parameters which is unnecessary as model with 4 & 3 parameters performed just as well.

In conclusion, residual analysis showed similar results to parameter estimation where models - SARIMA (2,0,2) x (0,1,1), SARIMA (1,0,2) x (0,1,1), SARIMA (3,0,1) x (0,1,1) - showed good results in terms of standardized residuals capturing autocorrelation. Out of these 3 models, we settle on SARIMA (1,0,2) x (0,1,1) as the best fit for our series due to it having the fewest parameters and based on the results obtained in Parameter estimation.

Over-parameter estimation

Since the SARIMA (1,0,2) x (0,1,1) model has been shown to be a good fit based on parameter estimation & diagnostic checking, we are conducting additional checks to see if over-parameterization (+1 parameters) produces a model with better results. To do this, we are examining neighboring AR and MA values with the SARIMA (2,0,2) x (0,1,1) and SARIMA (1,0,3) x (0,1,1) models. Given that the analysis for SARIMA (2,0,2) x (0,1,1) has already been analyzed, we will now evaluate the SARIMA (1,0,3) x (0,1,1) model and see if it performs better than SARIMA (1,0,2) x (0,1,1).

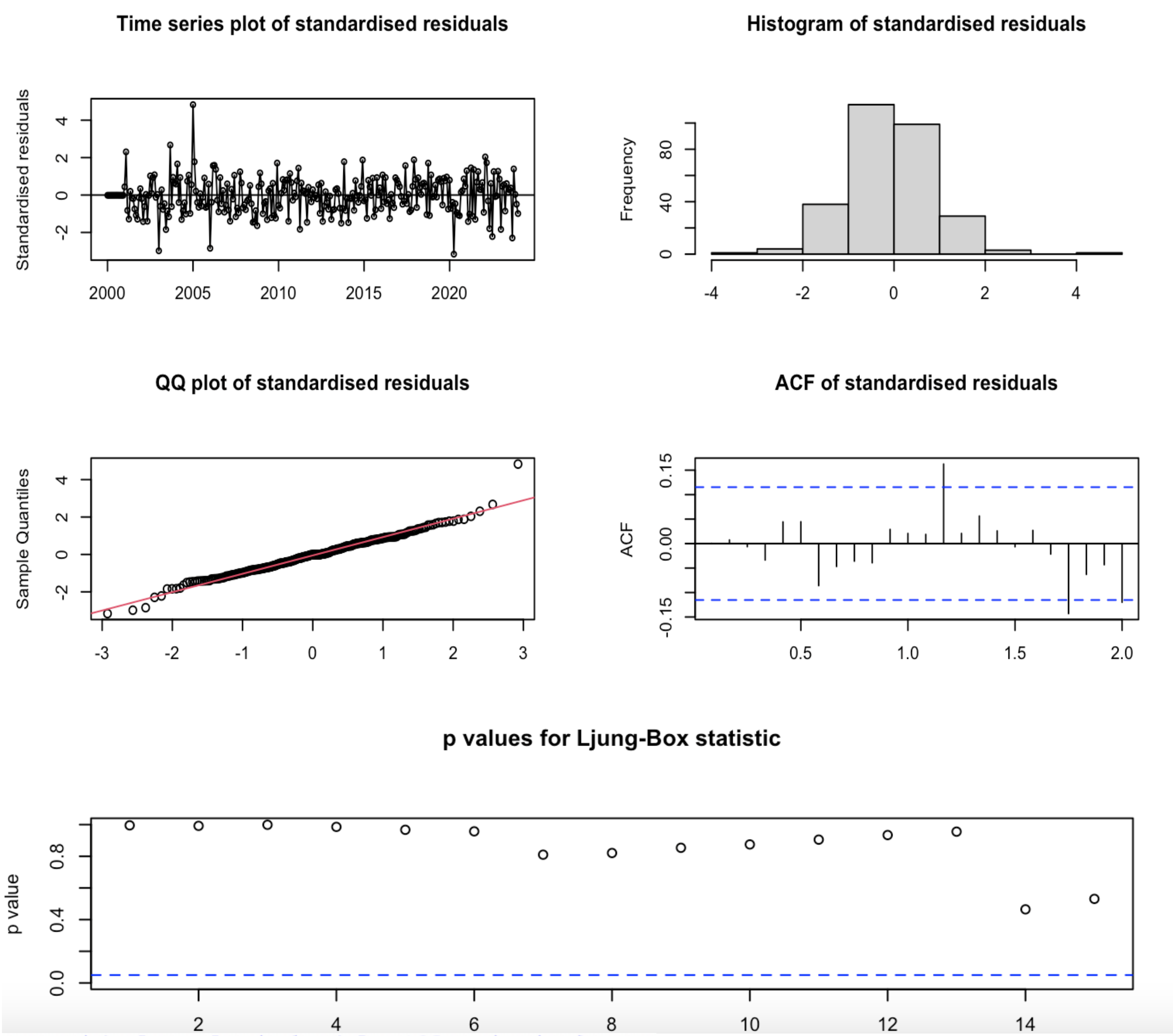
First let us fit the model and see its coefficients.

```
> m103.aviation = Arima(BC.aviation,order=c(1,0,3),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "ML")
> coeftest(m103.aviation)

z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1   0.997456   0.004309 231.4828  < 2e-16 ***
ma1  -1.130251   0.060458 -18.6949  < 2e-16 ***
ma2   0.200584   0.090236   2.2229  0.02622 *
ma3   0.021093   0.060986   0.3459  0.72944
sma1 -0.902435   0.047985 -18.8066  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m103CSS.aviation = Arima(BC.aviation,order=c(1,0,3),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS")
> coeftest(m103CSS.aviation)

z test of coefficients:

       Estimate Std. Error  z value Pr(>|z|)
ar1   0.9860040  0.0177330  55.6027   <2e-16 ***
ma1  -1.0887550  0.0622932 -17.4779   <2e-16 ***
ma2   0.2241536  0.0880508   2.5457   0.0109 *
ma3   0.0039847  0.0593345   0.0672   0.9465
sma1 -0.7899796  0.0388322 -20.3434   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> m103CSSML.aviation = Arima(BC.aviation,order=c(1,0,3),
+                       seasonal=list(order=c(0,1,1),
+                                     period=12),method = "CSS-ML")
> coeftest(m103CSSML.aviation)

z test of coefficients:

       Estimate Std. Error  z value Pr(>|z|)
ar1   0.9974565  0.0043115 231.3476  < 2e-16 ***
ma1  -1.1301192  0.0604611 -18.6917  < 2e-16 ***
ma2   0.2004894  0.0902262   2.2221  0.02628 *
ma3   0.0210875  0.0609859   0.3458  0.72951
sma1 -0.9023985  0.0479921 -18.8031  < 2e-16 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

fig32. SARIMA (1,0,3) x(0,1,1)

According to above results you can see that the newly added parameter MA(3) is not statistically significant. This tells us that the additional MA component when added to SARIMA(1,0,2)x(0,1,1) does not have a positive effect on the model's performance.

Secondly, let us also analyse the model's residuals.




```
> residual.analysis(model=m103.aviation)

        Shapiro-Wilk normality test

data:  res.model
W = 0.98064, p-value = 0.0005922

> Box.test(st.res.m103,type="Ljung-Box")

        Box-Ljung test

data:  st.res.m103
X-squared = 3.4061e-05, df = 1, p-value = 0.9953
```

Fig 33. Residuals of SARIMA (1,0,3) x (0,1,1).

From the above fig 33, you can see that the residuals are very similar to SARIMA(1,0,3) x (0,1,1)

- The time series plot is random around 0, no trend, or seasonality.

- There are no significant lags in seasons, and most are contained within the limits in ACF plot.
- Most of the data points in the QQ-plot follow the line and however the histogram is somewhat not symmetrically distributed, and the Shapiro-Wilks test p-value is <0.05 so the series is not normally distributed.
- Ljung-Box test, which determines where there is autocorrelation left shows all points above the limit and p-value greater than 0.05 which indicates there is no significant autocorrelation left.

As per residual analysis, model SARIMA (1,0,3) x (0,1,1) has similar behavior as SARIMA (1,0,2) x (0,1,1) which shows that this model is a good fit for our series but as we identified MA (3) is not significant and it does not appear to significantly enhance the model.

Lastly, let us also look at the AIC-BIC scores and error metrics for this model.

```
> sort.score(AIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
+                m202.aviation,m301.aviation,m501.aviation,m103.aviation), score = "aic")
              df       AIC
m102.aviation  5 -526.8044
m202.aviation  6 -524.9255
m103.aviation  6 -524.9245
m301.aviation  6 -524.7603
m501.aviation  8 -521.4579
m002.aviation  4 -475.0466
m001.aviation  3 -471.5371
m101.aviation  4 -470.5155
> sort.score(BIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
+                m202.aviation,m301.aviation,m501.aviation,m103.aviation), score = "bic")
              df       BIC
m102.aviation  5 -508.6843
m202.aviation  6 -503.1814
m103.aviation  6 -503.1804
m301.aviation  6 -503.0162
m501.aviation  8 -492.4658
m001.aviation  3 -460.6650
m002.aviation  4 -460.5506
m101.aviation  4 -456.0194
```

fig34. AIC BIC score

The AIC-BIC score of the SARIMA(1,0,3) x (0,1,1) model ranks amongst the top performing models however it isn't better than SARIMA(1,0,3) x (0,1,1).

```
> df.Smodels
                                ME       RMSE        MAE          MPE     MAPE      MASE          ACF1
SARIMA(0,0,1)x(0,1,1)_12 -0.026337861 0.09910561 0.07672466  1.609424674 5.749426 0.8405817 -0.0827947424
SARIMA(0,0,2)x(0,1,1)_12 -0.025376129 0.09798915 0.07607199  1.532872279 5.706514 0.8334311 -0.0521043701
SARIMA(1,0,2)x(0,1,1)_12 -0.003971966 0.08690473 0.06703981 -0.012985482 5.078721 0.7344761  0.0047591240
SARIMA(1,0,1)x(0,1,1)_12 -0.026396269 0.09893624 0.07653759  1.615328725 5.738113 0.8385321 -0.0558017688
SARIMA(2,0,2)x(0,1,1)_12 -0.003708607 0.08682947 0.06698300 -0.031767417 5.074996 0.7338537  0.0005117846
SARIMA(3,0,1)x(0,1,1)_12 -0.003863392 0.08691645 0.06696247 -0.021646541 5.072572 0.7336288 -0.0006702371
SARIMA(5,0,1)x(0,1,1)_12 -0.003990772 0.08672372 0.06669304 -0.009542307 5.052560 0.7306770  0.0009187621
SARIMA(1,0,3)x(0,1,1)_12 -0.003917415 0.08685421 0.06699061 -0.016325267 5.075089 0.7339371  0.0003415287
```

fig 35. Error metrics

Figure 35 displays the error metrics for the SARIMA (1,0,3) x (0,1,1) model alongside those of the other models. While the SARIMA (1,0,3) x (0,1,1) performs well, it does not achieve the lowest error for any metric. Its error metrics are comparable to the best-performing models, including SARIMA (1,0,2) x (0,1,1), SARIMA (2,0,2) x (0,1,1), and SARIMA (3,0,1) x (0,1,1), differing only by a decimal point.

Based on the above analysis we can safely say that while SARIMA(1,0,3) x (0,1,1) performs well, it is not significantly better than SARIMA(1,0,2) x (0,1,1). Therefore, SARIMA(1,0,3) x (0,1,1) will remain the best fit for our data.

Forecasting

The SARIMA (1,0,2) x (0,1,1) model was used to forecast the next 12 observations (i.e 1 year) in the series, as shown below.

```
> preds1
         Point Forecast      Lo 80     Hi 80      Lo 95     Hi 95
Feb 2024     0.08829407 0.06394260 0.1174991 0.05291358 0.1350222
Mar 2024     0.11940478 0.08965274 0.1544078 0.07592930 0.1751580
Apr 2024     0.11488601 0.08575397 0.1492676 0.07235625 0.1696902
May 2024     0.12764482 0.09639606 0.1642947 0.08194026 0.1859784
Jun 2024     0.14028057 0.10699754 0.1791079 0.09152371 0.2020016
Jul 2024     0.14488246 0.11080098 0.1845845 0.09493490 0.2079724
Aug 2024     0.14840939 0.11369284 0.1888138 0.09751727 0.2126014
Sep 2024     0.11612473 0.08630850 0.1513943 0.07262568 0.1723746
Oct 2024     0.12781242 0.09602503 0.1651975 0.08135820 0.1873552
Nov 2024     0.10375517 0.07582632 0.1370897 0.06311844 0.1570306
Dec 2024     0.09958774 0.07229627 0.1322769 0.05992047 0.1518749
Jan 2025     0.08907040 0.06358196 0.1198720 0.05212252 0.1384399
```

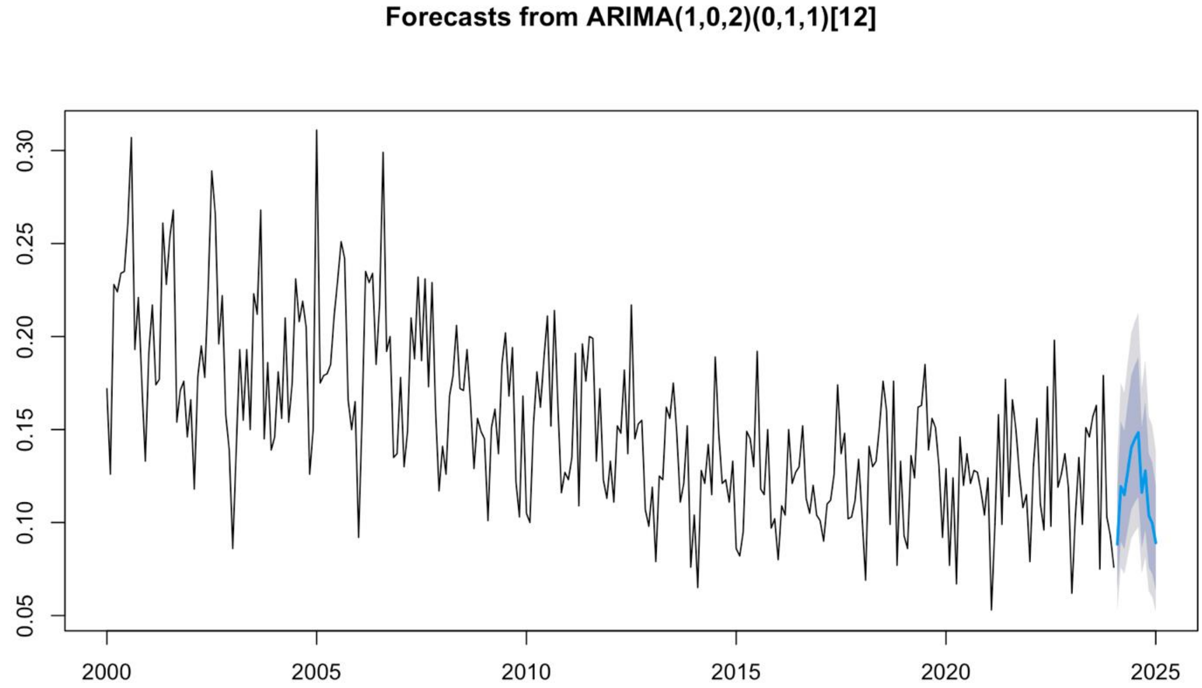


fig36. Forecasts for next 12 months using SARIMA (1,0,2) x (0,1,1) model

If you see the future CO2 emissions in the plot in Fig 36, the predictions follow a cyclic pattern that is similar to the previous seasons. It has the same up-down pattern where from February 2024 to August 2024 we see an incline in the CO2 emissions peaking at 0.148 metric tones and then from September 2024 onwards the emissions decrease ending with a low of 0.089 in January 2025. The 80% and 95% prediction intervals provide a range of expected values, reflecting the uncertainty around these forecasts. This range however is not too wide and seems to correctly encompass the possible forecasts based on past seasonal range.

Conclusion

The objective of this assignment was to analyze a public time series dataset and identify a suitable model for predicting future behavior. We conducted a thorough descriptive analysis to understand the defining characteristics of the series, which informed our modelling strategy. Various models were evaluated based on parameter estimation and residual analysis, considering each model's coefficient significance, error metrics, and diagnostic checks.

Through this comprehensive analysis, the SARIMA (1,0,2) x (0,1,1) model emerged as the best-fitting model. This model was selected due to its significant coefficients, its ability to effectively capture the underlying trends, its favourable accuracy scores and a high adjusted R-squared value that indicates strong explanatory power without overfitting.

Using the SARIMA (1,0,2) x (0,1,1) model, we forecasted the next 12 observations in the series. These forecasts reflect the downward trend observed in the historical data and provide valuable insights into potential future behaviour. However, it is important to note that actual future observations may vary.

Overall, this project successfully demonstrated each phase of time series analysis, from descriptive statistics to model selection and forecasting, justifying each methodological choice and presenting the results effectively.

Reference

Cryer Jonathan D., Chan Kung-Sik (2008) Time Series Analysis: With Application in R,Second Edition, Springer New York, NY, doi:10.1007/978-0-387-75959-3 (doi:10.1007/978-0-387-75959-3)

U.S Energy Information Administration (2024) *Independent Statistics and Analysis,* EIA website, accessed 1st May 2024. https://www.eia.gov/

Dr Heydar Demirhan (2024) 'Module 1: Basic Plots, Examples and Fundamental Concepts' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Dr Heydar Demirhan (2024) 'Module 2: Analysis of Trends' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Dr Heydar Demirhan (2024) 'Module 3: Models for Stationary Time Series' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Dr Heydar Demirhan (2024) 'Module 4: Models for Non Stationary Time Series' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Dr Heydar Demirhan (2024) 'Module 5: Models Specification' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Dr Heydar Demirhan (2024) 'Module 6: Parameter Estimation' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Dr Heydar Demirhan (2024) 'Module 7: Model Diagnostics and Forecasting' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Dr Heydar Demirhan (2024) 'Module 8: Seasonal Models' [PowerPoint slides, MATH1318], RMIT University, Melbourne.

Appendix

Code

```
# All packages and function

library(TSA)
library(fUnitRoots)
library(forecast)
library(lmtest)
library(tseries)

sort.score <- function(x, score = c("bic", "aic")){
  if (score == "aic"){
    x[with(x, order(AIC)),]
  } else if (score == "bic") {
    x[with(x, order(BIC)),]
  } else {
    warning('score = "x" only accepts valid arguments ("aic","bic")')
  }
}

residual.analysis <- function(model, std = TRUE,start = 2, class =
c("ARIMA","GARCH","ARMA-GARCH", "fGARCH")[1]){
  library(TSA)
  #library(FitAR)
  if (class == "ARIMA"){
    if (std == TRUE){
      res.model = rstandard(model)
    }else{
      res.model = residuals(model)
    }
  }else if (class == "GARCH"){
    res.model = model$residuals[start:model$n.used]
  }else if (class == "ARMA-GARCH"){
    res.model = model@fit$residuals
  }else if (class == "fGARCH"){
    res.model = model@residuals
  }else {
    stop("The argument 'class' must be either 'ARIMA' or 'GARCH' ")
  }
  par(mfrow=c(3,2))
  plot(res.model,type='o',ylab='Standardised residuals', main="Time series plot of
standardised residuals")
  abline(h=0)
  hist(res.model,main="Histogram of standardised residuals")
  qqnorm(res.model,main="QQ plot of standardised residuals")
  qqline(res.model, col = 2)
  acf(res.model,main="ACF of standardised residuals")
```

```
  print(shapiro.test(res.model))
  k=0
  par(mfrow=c(1,1))
}

 helper <- function(class = c("acf", "pacf"), ...) {

  # Capture additional arguments
  params <- match.call(expand.dots = TRUE)
  params <- as.list(params)[-1]

  # Calculate ACF/PACF values
  if (class == "acf") {
    acf_values <- do.call(acf, c(params, list(plot = FALSE)))
  } else if (class == "pacf") {
    acf_values <- do.call(pacf, c(params, list(plot = FALSE)))
  }

  # Extract values and lags
  acf_data <- data.frame(
    Lag = as.numeric(acf_values$lag),
    ACF = as.numeric(acf_values$acf)
  )

  # Identify seasonal lags to be highlighted
  seasonal_lags <- acf_data$Lag %% 1 == 0

  # Plot ACF/PACF values
  if (class == "acf") {
    do.call(acf, c(params, list(plot = TRUE)))
  } else if (class == "pacf") {
    do.call(pacf, c(params, list(plot = TRUE)))
  }

  # Add colored segments for seasonal lags
  for (i in which(seasonal_lags)) {
    segments(x0 = acf_data$Lag[i], y0 = 0, x1 = acf_data$Lag[i], y1 = acf_data$ACF[i], col =
"red")
  }
}

seasonal_acf <- function(...) {
  helper(class = "acf", ...)
}

#---------

aviation <- read.csv("aviation.csv", header = TRUE)

# Descriptive Analysis
```

```
sum(is.na(aviation$Jet.Fuel))
summary(aviation$Jet.Fuel)

hist(aviation$Jet.Fuel,main="Distribution of Aviation fuel CO2 emissions",xlab ="CO2
emissions (in million metric tons)")

skewness <- moments::skewness(aviation$Jet.Fuel)
print(paste("Skewness: ",skewness))
kurtosis <- moments::kurtosis(aviation$Jet.Fuel)
print(paste("Kurtosis: ",kurtosis))

# Create Time series object
aviationTS <- ts(aviation$Jet.Fuel, start=c(2000, 1), end=c(2024, 1), frequency = 12)
aviationTS

# Plot Time series against time
plot(aviationTS, ylab='Aviation fuel CO2 emissions', xlab='Months', type='o',
    main = "Time series plot of Aviation fuel CO2 emissions")

# Time series vs First lag
y = aviationTS
x = zlag(aviationTS)
index = 2:length(x)
cor(y[index],x[index])

plot(y[index],x[index],ylab='Time series', xlab='The first lag of series', main = "Scatter plot
of time series and first lag")


# Seasonal ACF
seasonal_acf(aviationTS,lag.max = 100, main="ACF Plot for Aviation CO2 emissions
series")


# Model Specification

# Seasonal Component

#D=1, because there is seasonal trend
m1.aviation = Arima(aviationTS,order=c(0,0,0),
              seasonal=list(order=c(0,1,0), period=12))

res.m1 = residuals(m1.aviation)
plot(res.m1,xlab='Time',ylab='Residuals',main="Time series plot of D=1 residuals")

# To see if there is more seasonal trend.
par(mfrow=c(1,2))
acf(res.m1, lag.max = 48, main = "ACF of the residuals")
pacf(res.m1, lag.max = 48, main = "PACF of the residuals")
```

```
# D=1, P=0, Q=1
m2.aviation = Arima(aviationTS,order=c(0,0,0),
                seasonal=list(order=c(0,1,1), period=12))
res.m2 = residuals(m2.aviation);
plot(res.m2,xlab='Time',ylab='Residuals',main="Time series plot of the residuals")

acf(res.m2, lag.max = 48, main = "ACF of the residuals")
pacf(res.m2, lag.max = 48, main = "PACF of the residuals")

# Ordinary ARIMA part

# Stationarity check
adf.test(res.m3)
pp.test(res.m3)

# BC transformation
# Pre-Transformation normality check
qqnorm(aviationTS)
qqline(aviationTS, col = 2)

BC = BoxCox.ar(aviationTS)

lambda <- BC$lambda[which(max(BC$loglike) == BC$loglike)]
lambda
BC.aviation = (aviationTS^lambda-1)/lambda

# Post Transformation normality check
qqnorm(BC.aviation)
qqline(BC.aviation, col = 2)


m3.aviation = Arima(BC.aviation,order=c(0,0,0),seasonal=list(order=c(0,1,1), period=12))
res.m3 = residuals(m3.aviation);
plot(res.m3,xlab='Time',ylab='Residuals',main="Time series plot of the residuals")

# Model Specification
# ACF-PACF
acf(res.m3, lag.max = 48, main = "The sample ACF of the residuals")
pacf(res.m3, lag.max = 48, main = "The sample PACF of the residuals")

# EACF
eacf(res.m3)

 # BIC table
plot(armasubsets(y=res.m3 , nar=5 , nma=5, y.name='p', ar.method='ols'))


# Model Fitting
```

```
# SARIMA(2,0,2)x(0,1,1)
m202.aviation = Arima(BC.aviation,order=c(2,0,2),
                seasonal=list(order=c(0,1,1),
                           period=12),method = "ML")
coeftest(m202.aviation)

m202CSS.aviation = Arima(BC.aviation,order=c(2,0,2),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS")
coeftest(m202CSS.aviation)

m202CSSML.aviation = Arima(BC.aviation,order=c(2,0,2),
                    seasonal=list(order=c(0,1,1),
                               period=12),method = "CSS-ML")
coeftest(m202CSSML.aviation)

residual.analysis(model=m202.aviation)

st.res.m202 = rstandard(m202.aviation)
Box.test(st.res.m202,type="Ljung-Box")
tsdiag(m202.aviation,gof=15,omit.inital=F)

# SARIMA(0,0,1)x(0,1,1)
m001.aviation = Arima(BC.aviation,order=c(0,0,1),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "ML")
coeftest(m001.aviation)

m001CSS.aviation = Arima(BC.aviation,order=c(0,0,1),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS")
coeftest(m001CSS.aviation)

m001CSSML.aviation = Arima(BC.aviation,order=c(0,0,1),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS-ML")
coeftest(m001CSSML.aviation)

residual.analysis(model=m001.aviation)
st.res.m001 = rstandard(m001.aviation)
Box.test(st.res.m001,type="Ljung-Box")
tsdiag(m001.aviation,gof=15,omit.inital=F)


# SARIMA(1,0,1)x(0,1,1)
m101.aviation = Arima(BC.aviation,order=c(1,0,1),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "ML")
coeftest(m101.aviation)
```

```
m101CSS.aviation = Arima(BC.aviation,order=c(1,0,1),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS")
coeftest(m101CSS.aviation)

m101CSSML.aviation = Arima(BC.aviation,order=c(1,0,1),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS-ML")
coeftest(m101CSSML.aviation)

residual.analysis(model=m101.aviation)
st.res.m101 = rstandard(m101.aviation)
Box.test(st.res.m101,type="Ljung-Box")
tsdiag(m101.aviation,gof=15,omit.inital=F)

# SARIMA(1,0,2)x(0,1,1)
m102.aviation = Arima(BC.aviation,order=c(1,0,2),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "ML")
coeftest(m102.aviation)

m102CSS.aviation = Arima(BC.aviation,order=c(1,0,2),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS")
coeftest(m102CSS.aviation)

m102CSSML.aviation = Arima(BC.aviation,order=c(1,0,2),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS-ML")
coeftest(m102CSSML.aviation)

residual.analysis(model=m102.aviation)
st.res.m102 = rstandard(m102.aviation)
Box.test(st.res.m102,type="Ljung-Box")
tsdiag(m102.aviation,gof=15,omit.inital=F)


# SARIMA(0,0,2)x(0,1,1)
m002.aviation = Arima(BC.aviation,order=c(0,0,2),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "ML")
coeftest(m002.aviation)

m002CSS.aviation = Arima(BC.aviation,order=c(0,0,2),
                  seasonal=list(order=c(0,1,1),
                             period=12),method = "CSS")
coeftest(m002CSS.aviation)

m002CSSML.aviation = Arima(BC.aviation,order=c(0,0,2),
                  seasonal=list(order=c(0,1,1),
```

```
                              period=12),method = "CSS-ML")
coeftest(m002CSSML.aviation)

residual.analysis(model=m002.aviation)
st.res.m002 = rstandard(m002.aviation)
Box.test(st.res.m002,type="Ljung-Box")
tsdiag(m002.aviation,gof=15,omit.inital=F)


#SARIMA(5,0,1)x(0,1,1)
m501.aviation = Arima(BC.aviation,order=c(5,0,1),
                  seasonal=list(order=c(0,1,1),
                              period=12),method = "ML")
coeftest(m501.aviation)

m501CSS.aviation = Arima(BC.aviation,order=c(5,0,1),
                  seasonal=list(order=c(0,1,1),
                              period=12),method = "CSS")
coeftest(m501CSS.aviation)

m501CSSML.aviation = Arima(BC.aviation,order=c(5,0,1),
                  seasonal=list(order=c(0,1,1),
                              period=12),method = "CSS-ML")
coeftest(m501CSSML.aviation)


residual.analysis(model=m501.aviation)
st.res.m501 = rstandard(m501.aviation)
Box.test(st.res.m501,type="Ljung-Box")
tsdiag(m501.aviation,gof=15,omit.inital=F)


#SARIMA(3,0,1)x(0,1,1)
m301.aviation = Arima(BC.aviation,order=c(3,0,1),
                  seasonal=list(order=c(0,1,1),
                              period=12),method = "ML")
coeftest(m301.aviation)

m301CSS.aviation = Arima(BC.aviation,order=c(3,0,1),
                  seasonal=list(order=c(0,1,1),
                              period=12),method = "CSS")
coeftest(m301CSS.aviation)

m301CSSML.aviation = Arima(BC.aviation,order=c(3,0,1),
                  seasonal=list(order=c(0,1,1),
                              period=12),method = "CSS-ML")
coeftest(m301CSSML.aviation)

residual.analysis(model=m301.aviation)
st.res.m301 = rstandard(m301.aviation)
```

```
Box.test(st.res.m301,type="Ljung-Box")
tsdiag(m301.aviation,gof=15,omit.inital=F)

# AIC-BIC Score
sort.score(AIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
          m202.aviation,m301.aviation,m501.aviation), score = "aic")

sort.score(BIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
          m202.aviation,m301.aviation,m501.aviation), score = "bic")

# Error Metrics

Sm001.aviation <- accuracy(m001.aviation)[1:7]
Sm002.aviation <- accuracy(m002.aviation)[1:7]
Sm102.aviation <- accuracy(m102.aviation)[1:7]
Sm101.aviation <- accuracy(m101.aviation)[1:7]
Sm202.aviation <- accuracy(m202.aviation)[1:7]
Sm301.aviation <- accuracy(m301.aviation)[1:7]
Sm501.aviation <- accuracy(m501.aviation)[1:7]
df.Smodels <- data.frame(
  rbind(Sm001.aviation,Sm002.aviation,Sm102.aviation,
      Sm101.aviation,Sm202.aviation,Sm301.aviation,Sm501.aviation)
)
colnames(df.Smodels) <- c("ME", "RMSE", "MAE", "MPE", "MAPE",
                  "MASE", "ACF1")
rownames(df.Smodels) <- c("SARIMA(0,0,1)x(0,1,1)_12",
                  "SARIMA(0,0,2)x(0,1,1)_12",
                  "SARIMA(1,0,2)x(0,1,1)_12",
                  "SARIMA(1,0,1)x(0,1,1)_12",
                  "SARIMA(2,0,2)x(0,1,1)_12",
                  "SARIMA(3,0,1)x(0,1,1)_12",
                  "SARIMA(5,0,1)x(0,1,1)_12")

df.Smodels

# Over-parameter estimation

#SARIMA(1,0,3)x(0,1,1)
m103.aviation = Arima(BC.aviation,order=c(1,0,3),
               seasonal=list(order=c(0,1,1),
                       period=12),method = "ML")
m103.aviation = Arima(BC.aviation,order=c(1,0,3),
               seasonal=list(order=c(0,1,1),
                       period=12),method = "ML")
coeftest(m103.aviation)

m103CSS.aviation = Arima(BC.aviation,order=c(1,0,3),
                seasonal=list(order=c(0,1,1),
                        period=12),method = "CSS")
coeftest(m103CSS.aviation)
```

```
residual.analysis(model=m103.aviation)
st.res.m103 = rstandard(m103.aviation)
Box.test(st.res.m103,type="Ljung-Box")
tsdiag(m103.aviation,gof=15,omit.inital=F)


# Score
sort.score(AIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
           m202.aviation,m301.aviation,m501.aviation,m103.aviation), score = "aic")

sort.score(BIC(m001.aviation,m002.aviation,m102.aviation,m101.aviation,
           m202.aviation,m301.aviation,m501.aviation,m103.aviation), score = "bic")

# Error metrics

Sm001.aviation <- accuracy(m001.aviation)[1:7]
Sm002.aviation <- accuracy(m002.aviation)[1:7]
Sm102.aviation <- accuracy(m102.aviation)[1:7]
Sm101.aviation <- accuracy(m101.aviation)[1:7]
Sm202.aviation <- accuracy(m202.aviation)[1:7]
Sm301.aviation <- accuracy(m301.aviation)[1:7]
Sm501.aviation <- accuracy(m501.aviation)[1:7]
Sm103.aviation <- accuracy(m103.aviation)[1:7]
df.Smodels <- data.frame(
  rbind(Sm001.aviation,Sm002.aviation,Sm102.aviation,
      Sm101.aviation,Sm202.aviation,Sm301.aviation,Sm501.aviation,Sm103.aviation)
)
colnames(df.Smodels) <- c("ME", "RMSE", "MAE", "MPE", "MAPE",
                  "MASE", "ACF1")
rownames(df.Smodels) <- c("SARIMA(0,0,1)x(0,1,1)_12",
                  "SARIMA(0,0,2)x(0,1,1)_12",
                  "SARIMA(1,0,2)x(0,1,1)_12",
                  "SARIMA(1,0,1)x(0,1,1)_12",
                  "SARIMA(2,0,2)x(0,1,1)_12",
                  "SARIMA(3,0,1)x(0,1,1)_12",
                  "SARIMA(5,0,1)x(0,1,1)_12",
                  "SARIMA(1,0,3)x(0,1,1)_12")

df.Smodels


# Forecasting
m4_102.aviation = Arima(aviationTS,order=c(1,0,2),seasonal=list(order=c(0,1,1),
period=12),
                 lambda = 0.4)
preds1 = forecast(m4_102.aviation, lambda = 0.4, h = 12)
preds1
par(mfrow=c(1,1))
```

```
plot(preds1)
```