Electricity Demand in Victoria, Australia (2015 – 2022) Analysis

Authors: Kai Ging Yong, Wong Yi Wei (Ethan), Yung Qi Chin

Table of Contents

1.0 Introduction

In this report, we will be analysing the electricity demand data in Victoria, Australia for the years 2015 – 2022. This dataset is obtained from Kaggle and contains 14 variables and 2106 observations (Kozlov, 2020). Additionally, this dataset records the price and electricity demand in megawatt per hour (MWh) for the period 1 January 2015 to 6 October 2020. Although there are 14 variables within the dataset, in this analysis report we will be focusing on the date and demand column, which returns the total daily electricity demand in MWh. The date column returns a daily recording of the observations, which we will convert and summarise the date and electricity demand to show only the monthly data.

The aim of this report is to select the best fitted model for electricity demand data in Victoria for further analysis such as forecasting. To achieve this aim, the objectives set out for the report are to first perform descriptive analysis, propose a set of possible models, fit all models to find the parameter estimation by performing residual analysis, then use goodness-of-fit metrics to select the best model possible and perform forecast with the best model. Furthermore, in this report, we will be using the following hypothesis tests:

Test	Hypothesis
Augmented Dickey-Fuller (ADF) Test	H_0: Distribution is not stationary H_1: Distribution is stationary Alpha level $(\alpha) = 0.05$
Phillips-Perron Unit Root Test	H_0: Distribution is not stationary H_1: Distribution is stationary Alpha level $(\alpha) = 0.05$
Shapiro-Wilk Normality Test	H_0: Distribution is normal H_1: Distribution is not normal Alpha level $(\alpha) = 0.05$

2.0 Data Import and Pre-processing

Before we perform any analysis, we will first import the data and check the structure of the data. This step will allow us to gain a further understanding of the dataset.

```
'data.frame': 2106 obs. of  14 variables:
 $ date         : chr  "2015-01-01" "2015-01-02" "2015-01-03" "2015-01-04" ...
 $ demand       : num  99635 129606 142301 104331 118132 ... $ RRP          : num  25.6 33.1
34.6 25 26.7 ...
 $ demand_pos_RRP : num  97319 121082 142301 104331 118132 ...
 $ RRP_positive   : num  26.4 38.8 34.6 25 26.7 ...
 $ demand_neg_RRP : num  2316 8524 0 0 0 ...
 $ RRP_negative   : num  -7.24 -47.81 0 0 0 ...
 $ frac_at_neg_RRP: num  0.0208 0.0625 0 0 0 ...
 $ min_temperature: num  13.3 15.4 20 16.3 15 17.7 18.9 23.1 16.5 13.6 ...
 $ max_temperature: num  26.9 38.8 38.2 21.4 22 26 37.4 28.2 18 21.7 ...
 $ solar_exposure : num  23.6 26.8 26.5 25.2 30.7 31.6 20.7 13.5 3.1 5.6 ...
 $ rainfall       : num  0 0 0 4.2 0 0 0 19.4 1.2 5.2 ...
 $ school_day     : chr  "N" "N" "N" "N" ...
 $ holiday        : chr  "Y" "N" "N" "N" ...
```

Figure 1: Structure of the Data Set

	date	demand <dbl>	RRP <dbl>	demand_pos_RRP <dbl>	RRP_positive <dbl>
	<chr>				
	2019-08-16	129781.2	100.01159	129781.2	100.01159
	2018-09-26	119238.3	120.86121	119238.3	120.86121
748	2017-01-17	140517.6	129.23766	140517.6	129.23766
	2015-09-26	111064.5	48.78937	111064.5	48.78937
	2018-04-27	119079.5	78.21485	119079.5	78.21485

Figure 2: Random 5 Observations

Above outputs in Figure 1 and Figure 2 shows the structure and random 5 observations within the dataset. From the outputs we can observe that there is no errors or inconsistencies from the data. Now, we will summarise the data using the date and demand column to generate a monthly demand data.

```
year       month       total_demand Min.   :2015   Min.   : 1.000
Min.   :3209217
1st Qu.:2016   1st Qu.: 3.000   1st Qu.:3453621
Median :2017   Median : 6.000   Median :3631206
Mean   :2017   Mean   : 6.304   Mean   :3654578
3rd Qu.:2019   3rd Qu.: 9.000   3rd Qu.:3877235
Max.   :2020   Max.   :12.000   Max.   :4273538
```

Figure 3: Summary for Monthly Data

year <int>	month <int>	total_demand <dbl>
2018	8	3942255
2019	7	3997640
2016	1	3729075
2019	9	3488502
2015	7	4273538

Figure 4: Random 5 Observations for Monthly Data

Above outputs in Figure 3 and 4 shows the summary and random 5 observations for the summarised monthly data. We can observe that the monthly summary is successfully completed and we will proceed with converting the monthly data to a time series object, and perform further analysis.

3.0 Data Visualisation and Descriptive Analysis

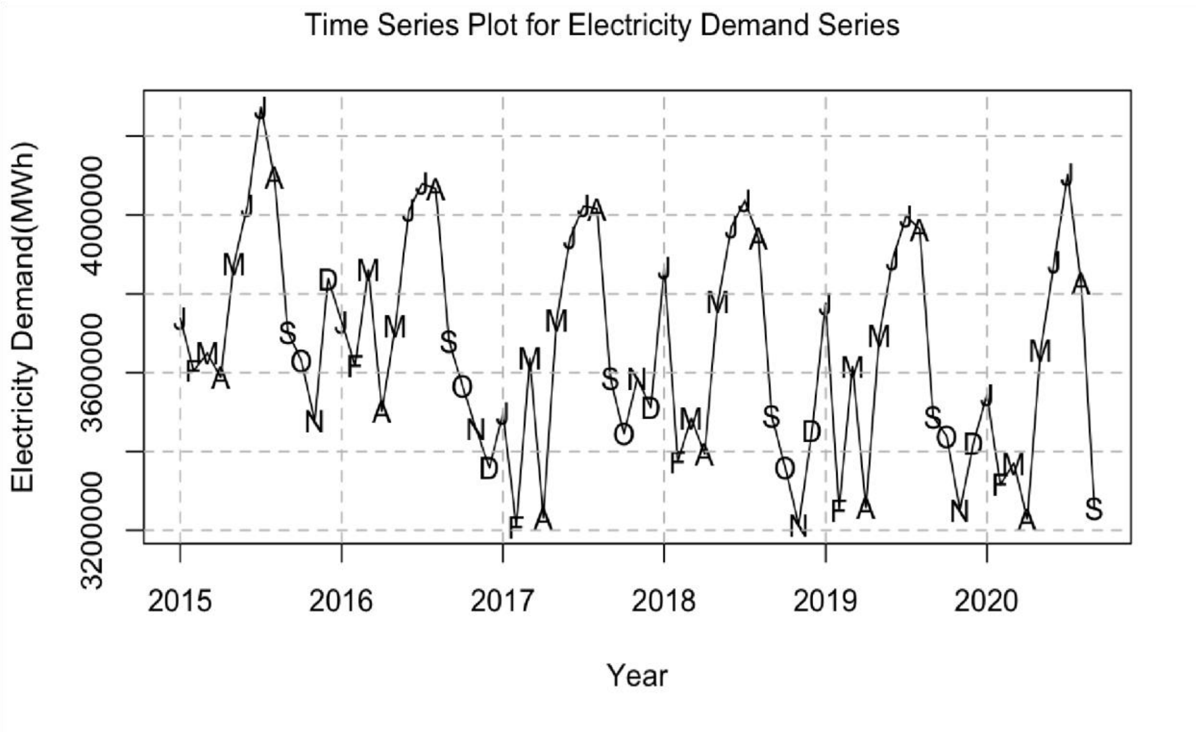

Figure 5: Time Series Plot for Electricity Demand Series

Figure 5 above shows the time series plot for electricity demand series. From the time series plot, we can observe the following characteristics:

Trend: There is a slight downward trend towards 2018 and increases from 2020 but it is not obvious.

Seasonality: There is evidence of seasonality where electricity demand peaks in July every year.

Changing Variance: Since there is seasonality, changing variance is not obvious.

Moving Average Behaviour: There is evidence of autoregression and moving average behaviour within the time series plot.

Intervention/ Change Point: There are no apparent intervention observed in the plot.

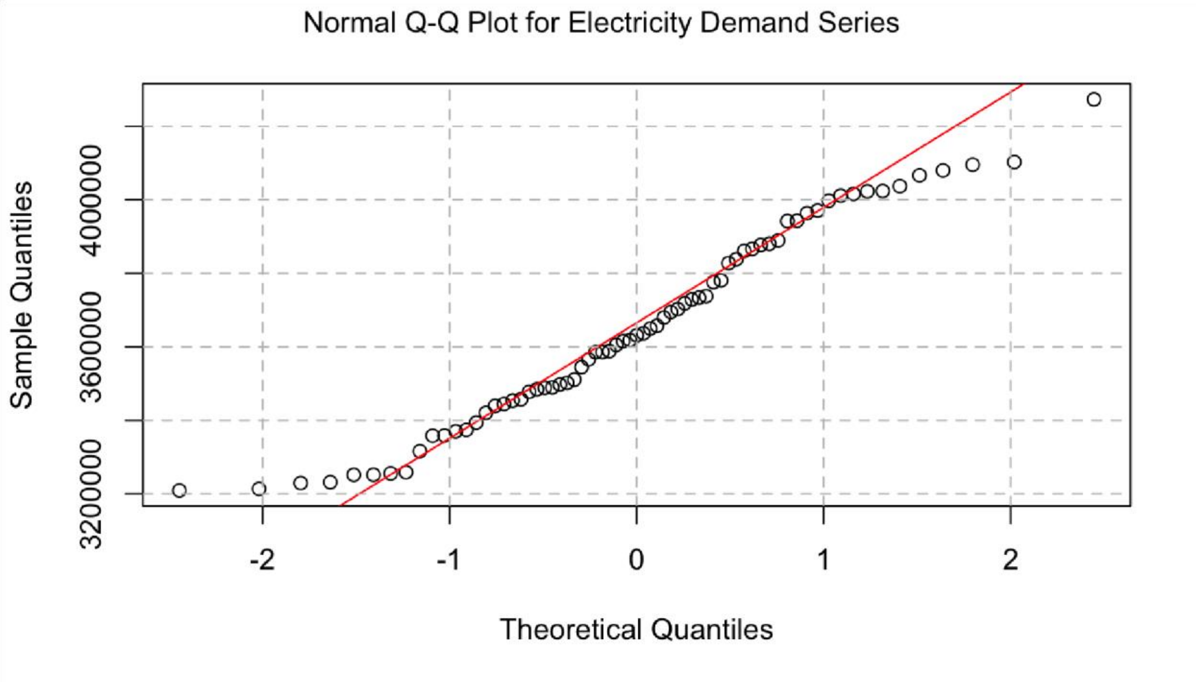

Figure 6: Normal Q-Q Plot for Electricity Demand Series

Figure 6 above shows the normal quantile-quantile (Q-Q) plot for electricity demand series. From the plot, we can observe that the data points follow the reference line closely, which could suggest that the series is not normally distributed.

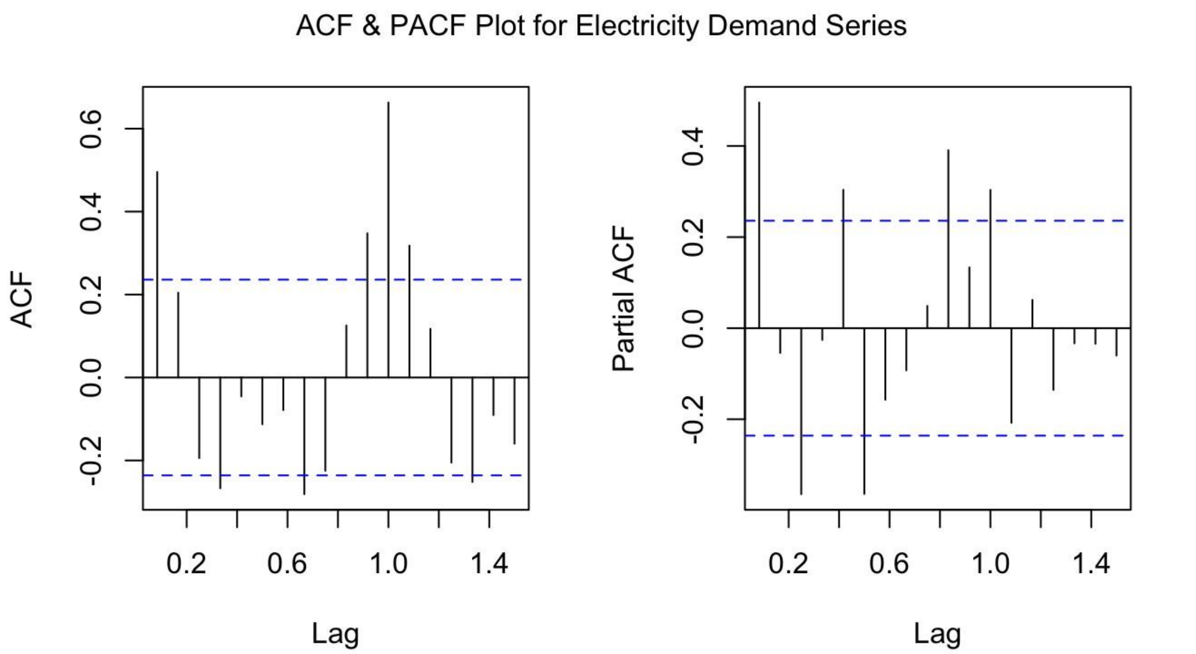

Figure 7: ACF and PACF Plot for Electricity Demand Series

Figure 7 above shows the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots for electricity demand series. From the ACF plot we can observe that there is a decaying patter with intervals between lags. PACF plot shows that there is significant first lag, indicating that the series is nonstationary and not normally distributed.

```
Electricity Demand Series Tests

                    Augmented Dickey-Fuller Test

data:  plotdata
Dickey-Fuller = -3.2417, Lag order = 4, p-value = 0.08841 alternative hypothesis: stationary


                    Phillips-Perron Unit Root Test

data:  plotdata
Dickey-Fuller Z(alpha) = -36.579, Truncation lag parameter = 3, p-value = 0.01 alternative hypothesis: stationary


                    Shapiro-Wilk normality test

data:  plotdata
W = 0.96648, p-value = 0.06078
```

Figure 8: ADF, PP, and Shapiro-Wilk Test for Electricity Demand Series

From Figure 8 above, Augmented Dickey-Fuller (ADF) test returns a p-value of 0.088, where it is greater than the alpha value of 0.05. Hence, we cannot reject the null hypothesis that the electricity demand series is not stationary. Interestingly, Phillips-Perron (PP) Unit Root test returns a p-value of 0.01, which is less than alpha value of 0.05. We can reject the null hypothesis that the electricity demand series is not stationary. Lastly, Shapiro-Wilk normality test returns a p-value of 0.061, which is less than alpha value, 0.05, hence we can reject the null hypothesis that the electricity demand series is normally distributed. Since ADF and PP test returns different results, we will assume that the series is not stationary and perform further analysis.

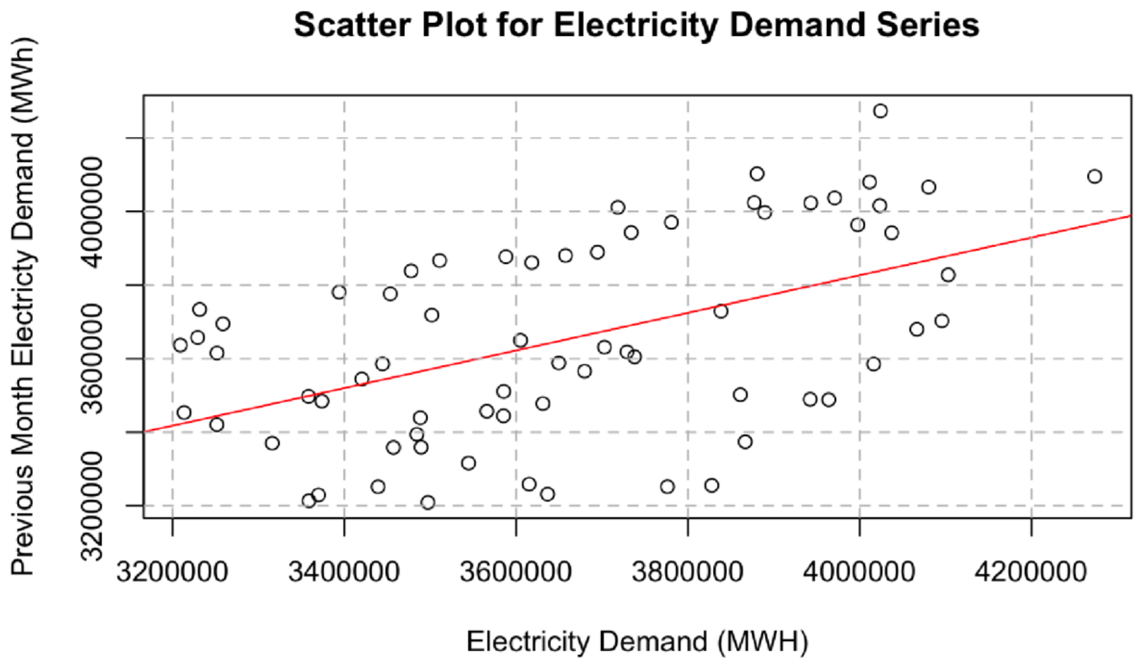

Figure 9: Scatter Plot for Electricity Demand Series

Figure 9 above shows the scatter plot for electricity demand series which plots the electricity demand of the month against the previous month electricity demand. From the plot, we can observe that the points do not follow the regression line closely but is not further away from the line. To gain a further understanding of the correlation, we will calculate the correlation of the electricity demand series.

```
Correlation: 0.5041434
```

Figure 10: Correlation Output for Electricity Demand Series

From Figure 10, we can observe that the electricity demand series has a correlation of 0.504, indicating a moderate correlation between the previous month electricity demand and electricity demand.

4.0 Data Preparation for Modelling

In this section, we will prepare the data for modelling by using transformation methods to try and normalise the data. We will be using Box Cox Transformation and seasonal autoregression integrated moving average (SARIMA) approach.

4.1 Box Cox Transformation

We will now perform the Box Cox transformation to attempt to transform the data, then performing unit root and normality test to ensure that the transformation is suitable for further analysis.

```
Sum of Negative Values: 0

Data Summary:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
3209217 3453621 3631206 3654578 3877235 4273538
```

Figure 11: Sum of Negative Values and Summary Output

For Box Cox transformation, the assumption for the data is to have only positive values, hence in Figure 11 we have calculated the sum of negative values and can confirm that there are no positive values within the data as it would be abnormal if electricity demand contained a negative value.

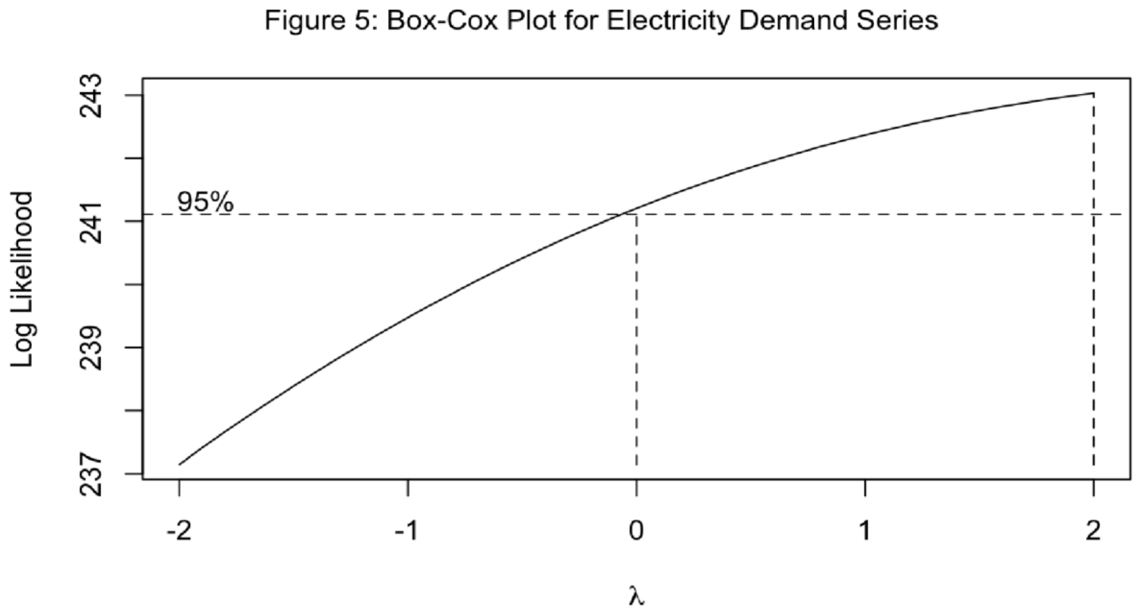

Figure 12: Box-Cox Plot for Electricity Demand Series

```
CI: 0 2
Lambda: 2
```

Figure 13: Confidence Interval and Lambda Value output for Box Cox Transformed Electricity Demand Series

Figure 12 above shows the box-cox plot for electricity demand series. From the plot we can observe that the confidence interval for lambda includes 1, hence no further transformation is necessary. Furthermore, Figure 13 shows that the box cox transformation has a confidence interval between 0 and 2, and a lambda value of 2.

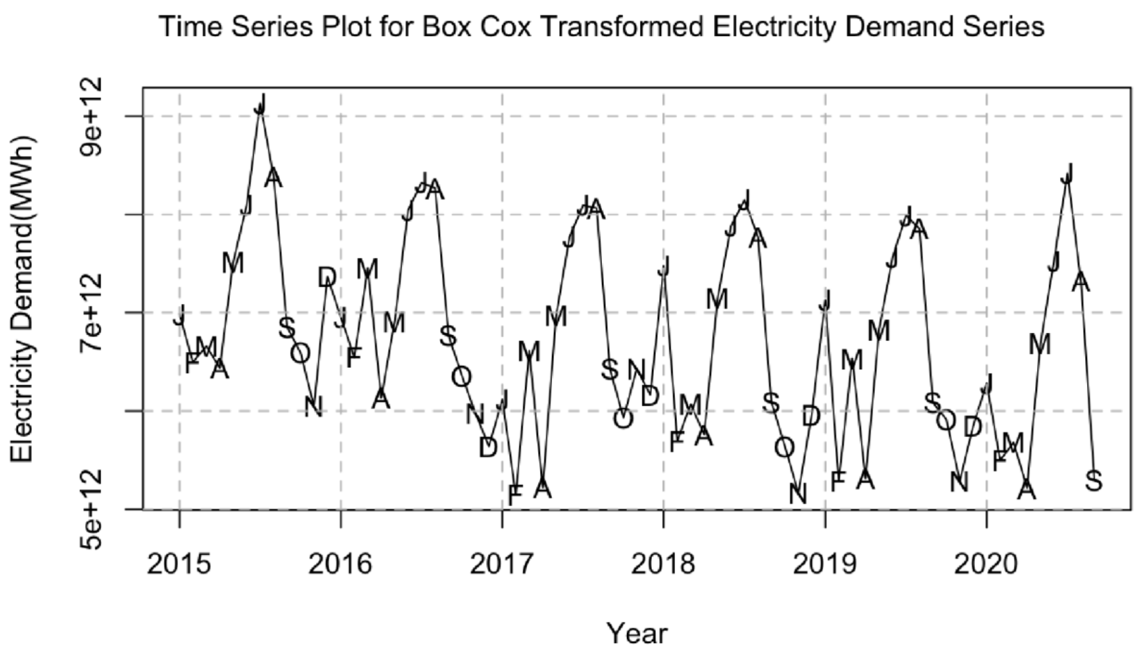

Figure 14: Time Series Plot for Box Cox Transformed Electricity Demand Series

Figure 14 above shows the time series plot for Box Cox transformed electricity demand series. From the plot, the following characteristics can be observed:

Trend: There is a slight downward trend towards 2018 and increases from 2020 but it is not obvious.

Seasonality: There is evidence of seasonality where electricity demand peaks in July every year.

Changing Variance: Since there is seasonality, changing variance is not obvious.

Moving Average Behaviour: There is evidence of autoregression and moving average behaviour within the time series plot.

Intervention/ Change Point: There are no apparent intervention observed in the plot.

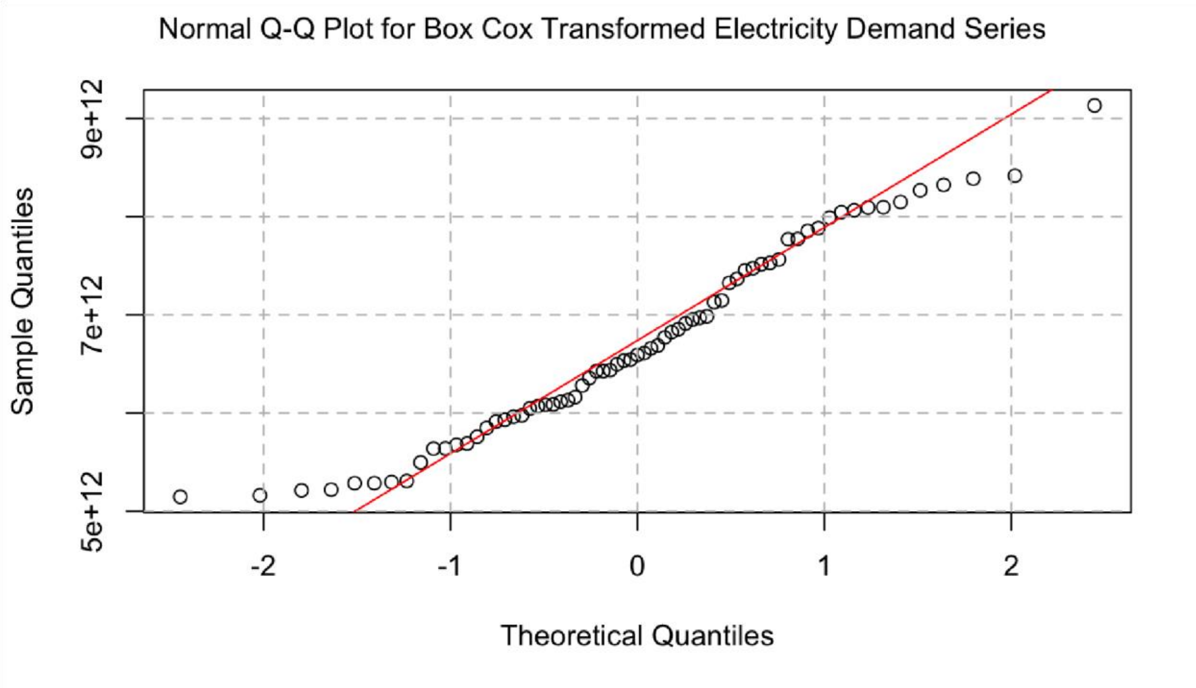

Figure 15: Normal Q-Q Plot for Box Cox Transformed Electricity Demand Series

Figure 15 above shows the normal Q-Q plot for box cox transformed electricity demand series. From the plot, we can observe that similarly to the Q-Q plot in Figure 6, data points follow the reference line closely, which could suggest that the distribution is not normally distributed.

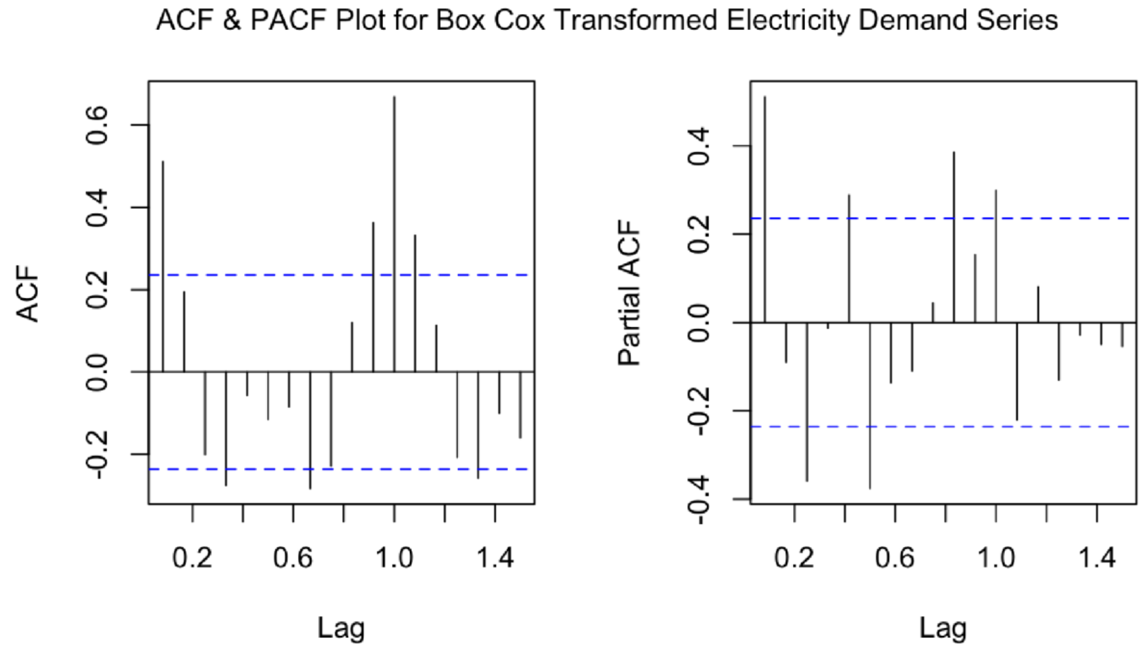

Figure 16: ACF and PACF Plot for Box Cox Transformed Electricity Demand Series

From Figure 16 which shows the ACF and PACF plot for box cox transformed electricity demand series, we can observe that in the ACF plot, there are 5 significant lags with decaying pattern. PACF plot shows that the first lag is significant, indicating that the series is nonstationary and not normally distributed.

Since Box Cox transformation did not address the issue of nonstationary and normality, and there is seasonality observed within the plot, we will proceed with the SARIMA approach.

4.2 SARIMA Approach

Here we will attempt to address the issue of nonstationary and normality using the SARIMA approach. We will first plot the electricity demand series with a seasonal differencing of 1 labelled model 1.

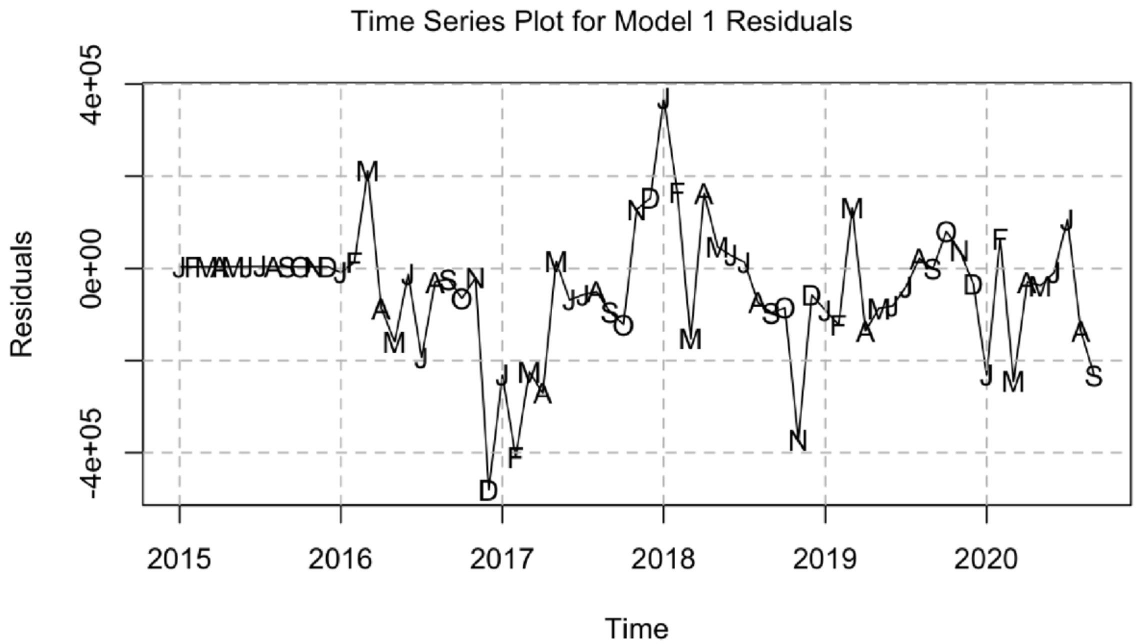

Figure 17: Time Series Plot for Model 1 Residuals

Figure 17 above shows the time series plot for model 1 residuals, with a model fitting of SARIMA(0,0,0)x(0,1,0) and a seasonal differencing component value of 1. From the plot we can see that the residuals do not exhibit any trend, seasonality, changing variance, moving average behaviours or intervention points.

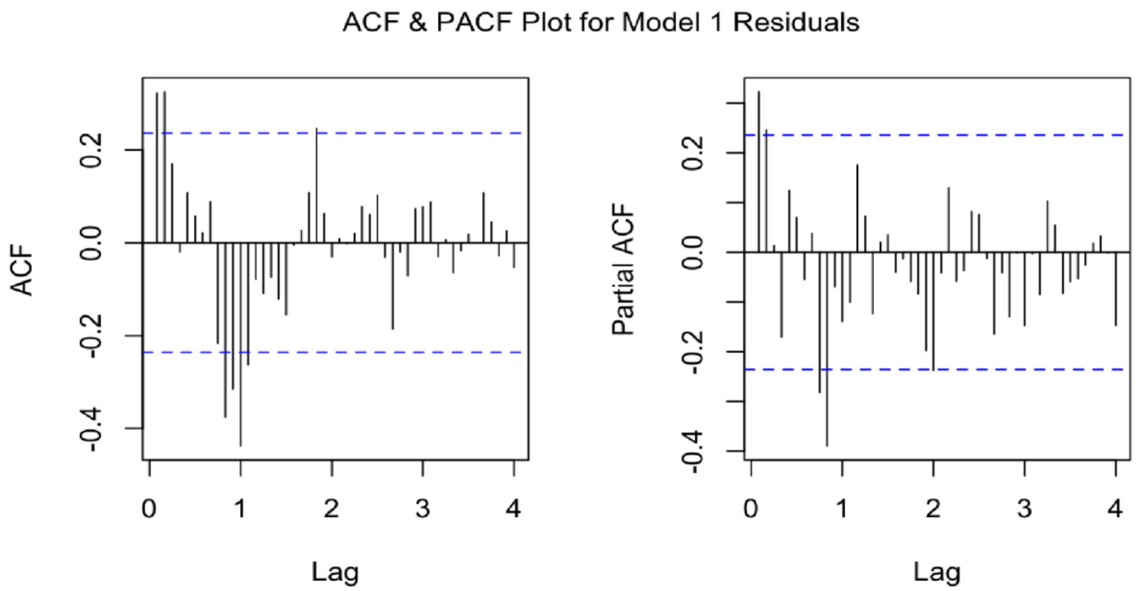

Figure 18 ACF and PACF Plot for Model 1 Residuals

Figure 18 above shows the ACF and PACF plot for the residuals in model 1 with a maximum lag of 48. From the ACF plot, we can observe that there is only 1 seasonal lag is significant, and PACF plot does not have any significant lag, we can conclude that the seasonality is filtered out for now. Therefore, we will try fitting a second model which seasonal P value = 0, and seasonal Q value = 1.

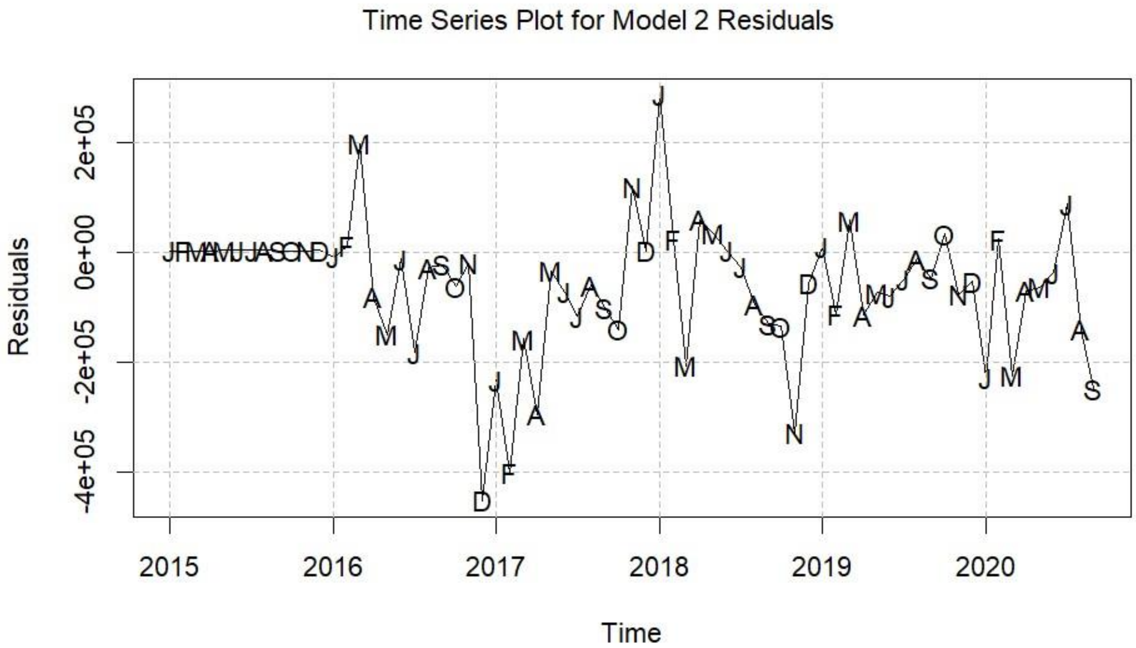

Figure 19: Time Series Plot for Model 2 Residuals

Figure 19 above shows the time series plot for model 2 residuals, with a model fitting of SARIMA(0,0,0)x(0,1,1), a seasonal P value of 0, a seasonal Q value of 1 and a seasonal differencing component value of 1. From the plot we can see that the residuals do not exhibit any trend, seasonality, changing variance, moving average behaviours or intervention points.

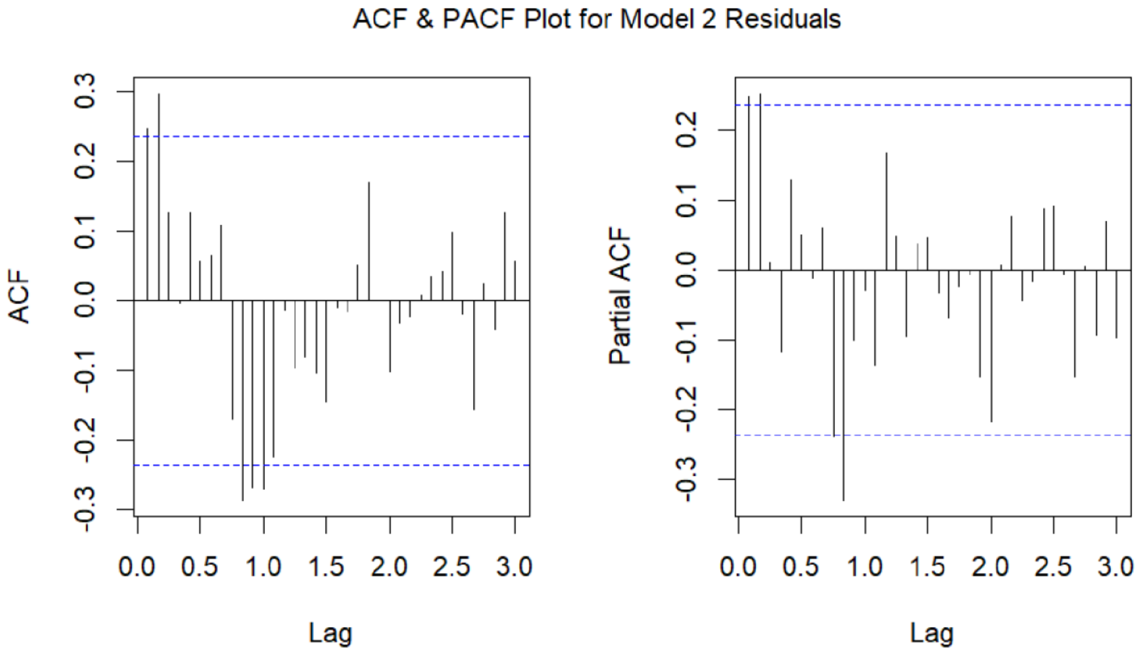

Figure 20: ACF and PACF Plot for Model 2 Residuals

Figure 20 above shows the ACF and PACF plot for the residuals in model 2 with a maximum lag of 48. From the ACF plot, we can observe that there is only 1 seasonal lag is significant, and PACF plot does not have any significant lag, we can conclude that the seasonality is filtered out for now. Next, we will try fitting the Box Cox transformation that we defined in the previous section, to see if we can see the trend more clearly.

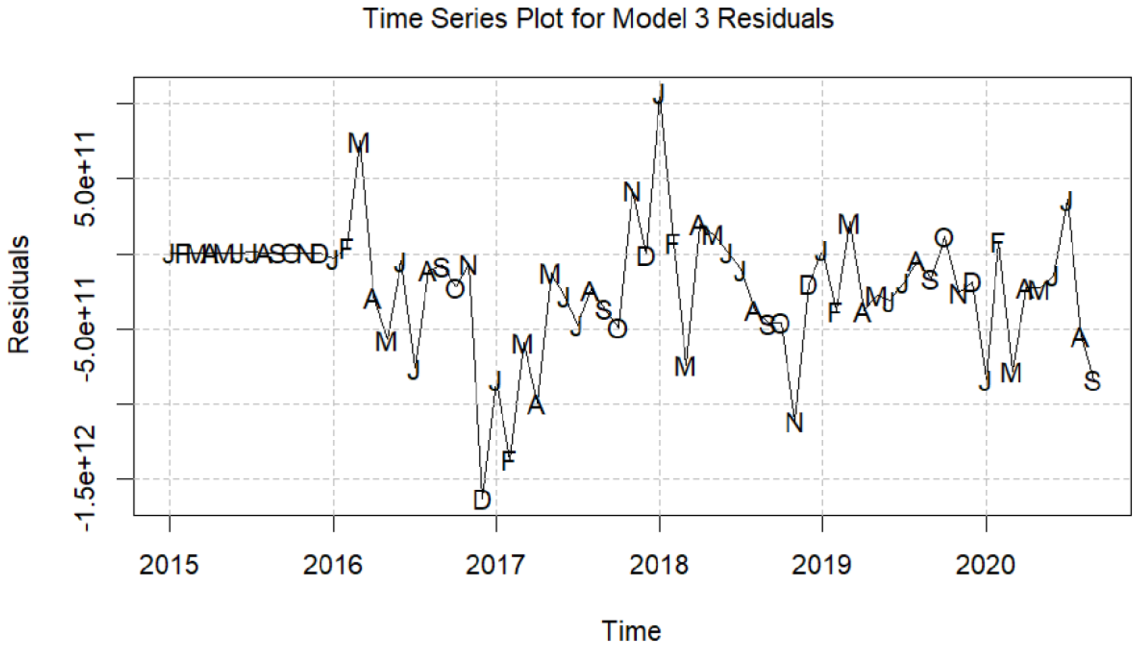

Figure 21: Time Series Plot for Model 3 Residuals

Figure 21 above shows the Box Cox transformed time series plot for model 3 residuals, with a model fitting of SARIMA(0,0,0)x(0,1,1), that is, a seasonal P value of 0, a seasonal Q value of 1 and a seasonal differencing component value of 1. From the plot we can see that the residuals do not exhibit any trend, seasonality, changing variance, moving average behaviours or intervention points.

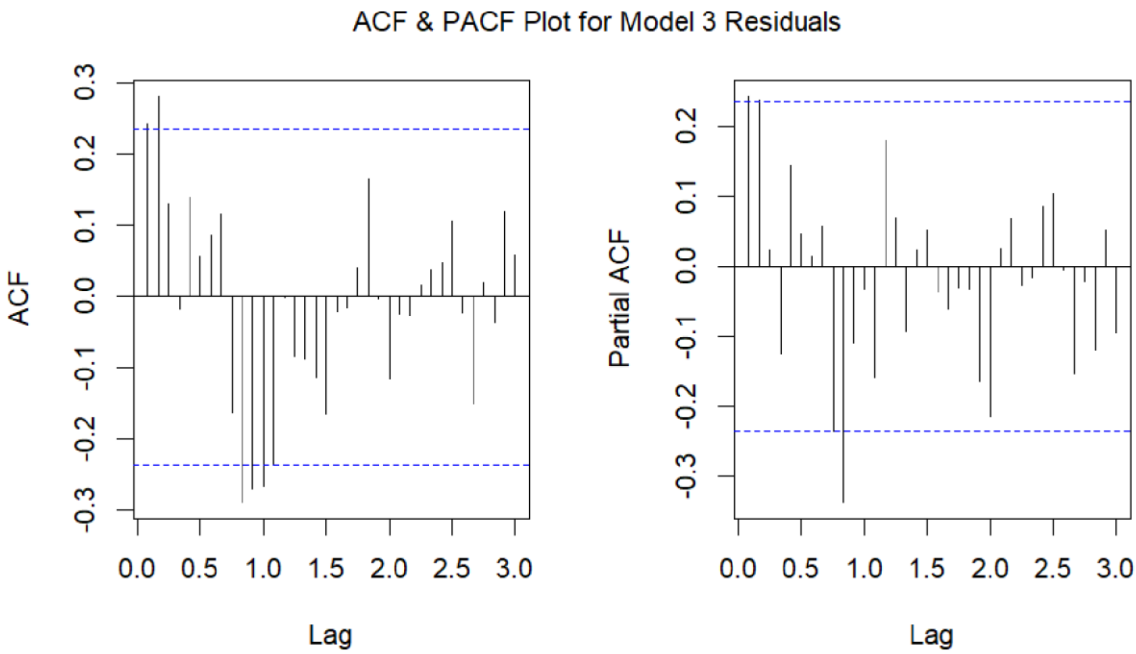

Figure 22: ACF and PACF Plot for Model 3 Residuals

Figure 22 above shows the ACF and PACF plot for the residuals in model 3 with a maximum lag of 48. From the ACF plot, we can observe that there are 5 significant lags between 0 and 1, and PACF plot shows 4 significant lags between 0 and 1. We do not see any patterns in ACF and PACT plots, so we can set ordinal differencing to 0. Next, we will try fitting a fourth model which non-seasonal order p value = 4, and non-seasonal q value = 5.

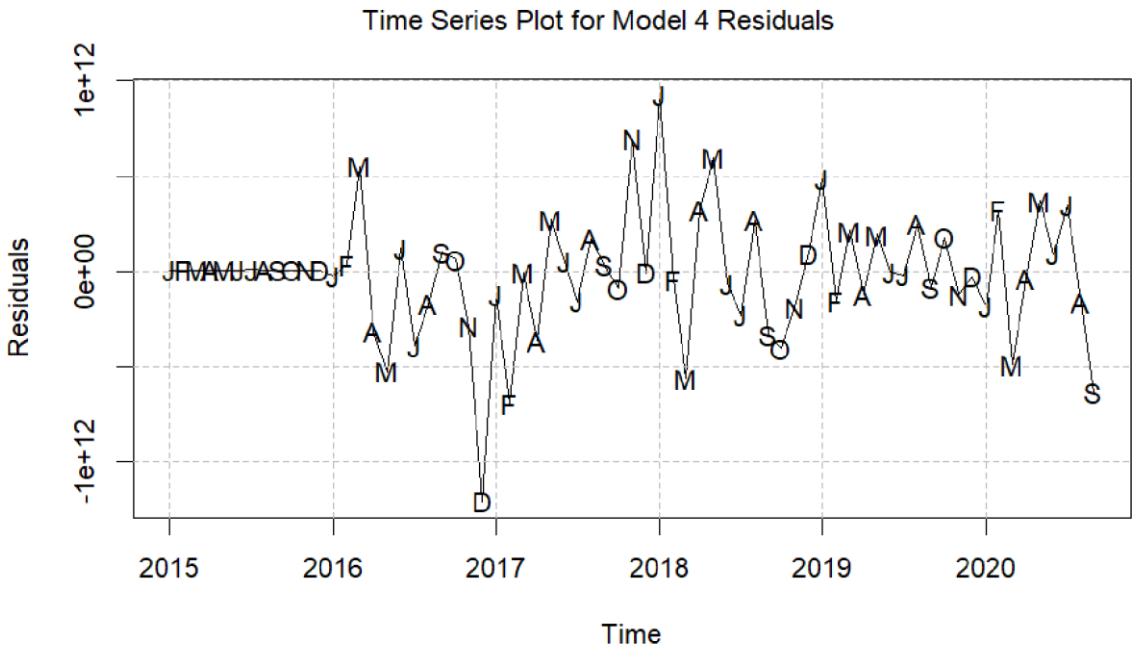

Figure 23: Time Series Plot for Model 4 Residuals

Figure 23 above shows the Box Cox transformed time series plot for model 4 residuals, with a model fitting of SARIMA(4,0,5)x(0,1,1), that is, a non-seasonal p value of 4 , a non-seasonal q value of 5, a seasonal P value of 0, a seasonal Q value of 1 and a seasonal differencing component value of 1. From the plot we can see that the residuals do not exhibit any trend, seasonality, changing variance, moving average behaviours or intervention points.

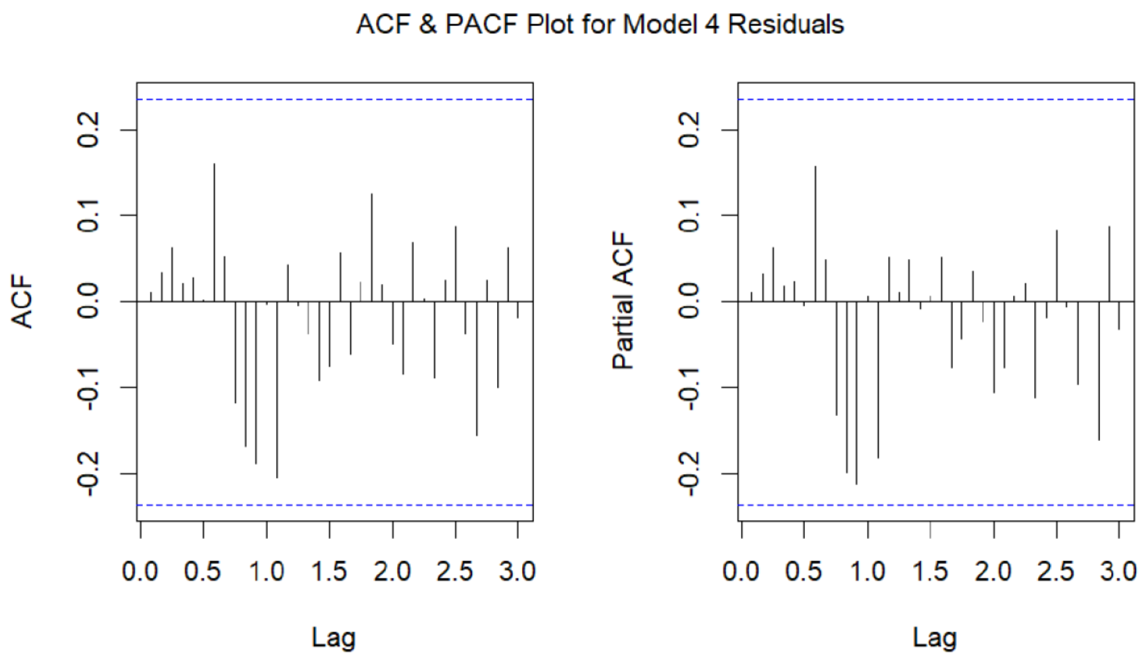

Figure 24: ACF and PACF Plot for Model 4 Residuals

Figure 24 above shows the ACF and PACF plot for the residuals in model 4 with a maximum lag of 48. It is obvious that there is no significant lag between 0 and 1, and there is no significant seasonal lag. Therefore, we can conclude that there is no pattern in the residuals of this seasonal model. We can consider this model as one of the final models for forecasting this time series data.

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

Figure 25: EACF Table for Model 3 Residuals

From the EACF table above, we will select the top left "o" that do not have obstructions for 5 in a row. We will select SARIMA(0,0,2)x(0,1,1)_12 as the origin. The neighbors of the origin are SARIMA(0,0,3)x(0,1,1)_12, SARIMA(1,0,2)x(0,1,1)_12 and SARIMA(1,0,3)x(0,1,1)_12.

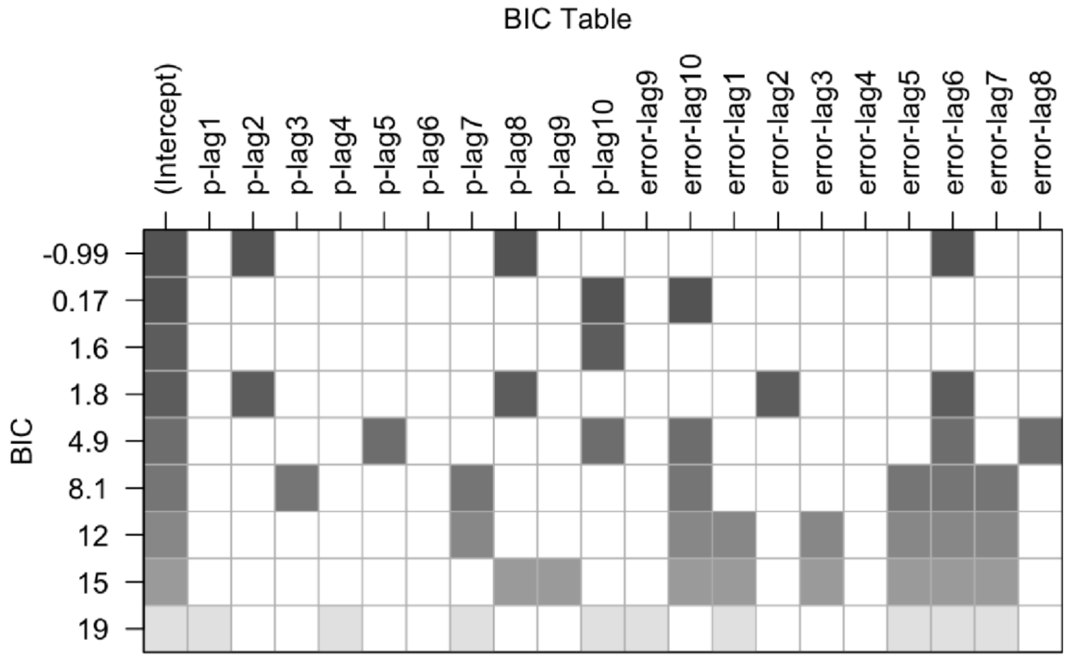

Figure 26: BIC Table for Model 3 Residuals

From the BIC Table of Model 3 residuals, we will pick p = 2 ,8 and q = 6 because their first row in the BIC Table are significant. In addition, we will also pick p = 10 because its second and third rows are significant. In summary, the model that we will pick based on this BIC table is SARIMA(8,0,6)x(0,1,1)_12, SARIMA(2,0,6)x(0,1,1)_12 and SARIMA(10,0,6)x(0,1,1)_12.

5.0 Parameter Estimation

From the ACF, PACF, EACF and BIC tables above we can conclude that the final possible models are:

- SARIMA(4,0,5)x(0,1,1)
- SARIMA(0,0,3)x(0,1,1)_12
- SARIMA(1,0,3)x(0,1,1)_12
- SARIMA(2,0,6)x(0,1,1)_12
- SARIMA(0,0,2)x(0,1,1)_12
- SARIMA(1,0,2)x(0,1,1)_12
- SARIMA(8,0,6)x(0,1,1)_12
- SARIMA(10,0,6)x(0,1,1)_12

We will choose to neglect SARIMA(8,0,6) and SARIMA(10,0,6) as their parameters are much larger than the others, leading to overfitting tendencies.

5.1 SARIMA(4,0,5) x (0,1,1)

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)     ar1  -0.83608    0.23875 -3.5020
0.0004618 *** ar2   0.24009    0.13052  1.8395 0.0658454 .   ar3   0.88726
0.15952  5.5620 2.666e-08 *** ar4   0.70808    0.18705  3.7854
0.0001534 *** ma1   1.04739    0.28264  3.7057 0.0002108 *** ma2
0.30534    0.29426  1.0377 0.2994271     ma3  -0.72453    0.29698 -2.4397
0.0147009 *   ma4  -1.11812    0.32018 -3.4922 0.0004791 *** ma5  -
0.27826    0.22711 -1.2252 0.2205035     sma1 -0.81362    0.32166 -2.5294
0.0114244 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
                         Shapiro-Wilk normality test

data:  res.model
W = 0.94655, p-value = 0.005197
```

Figure 27: Parameter estimation of SARIMA(4,0,5) x (0,1,1) and Shapiro Test using ML method

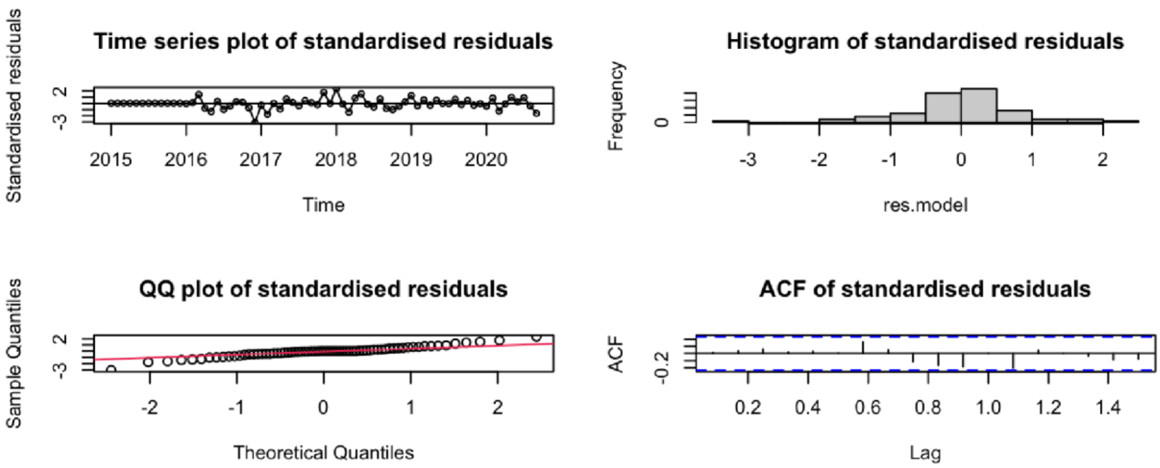

Figure 28: Residuals analysis of SARIMA(4,0,5) x (0,1,1) model using ML method

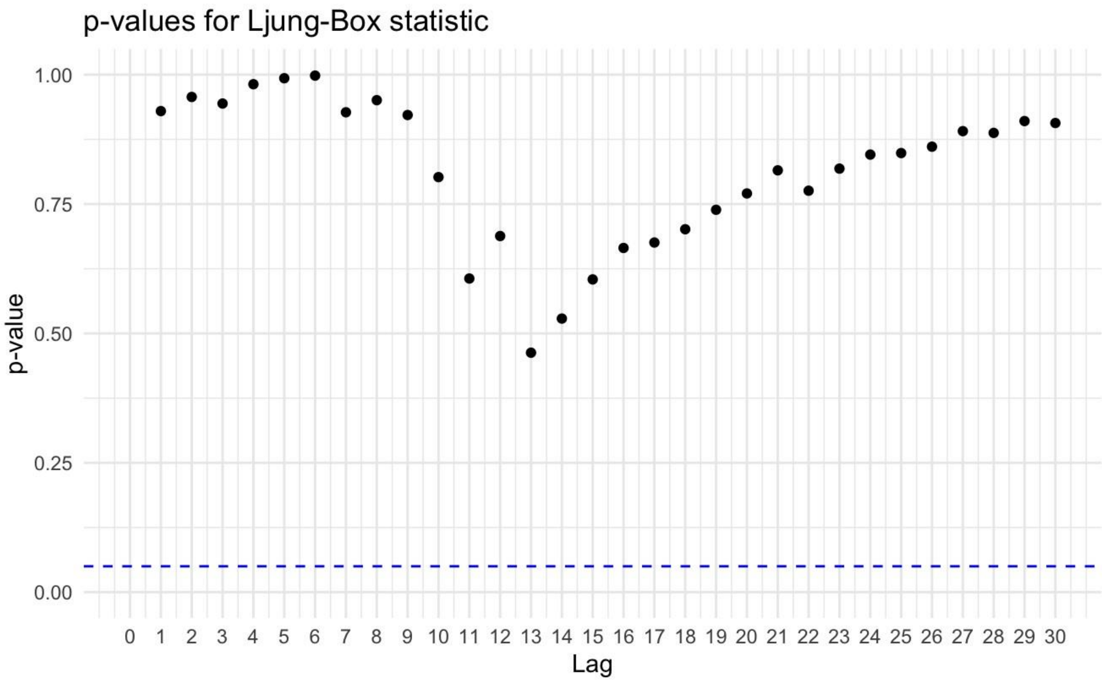

Figure 29: Residuals analysis of SARIMA(4,0,5) x (0,1,1) model using ML method

From Figure 27, 28 and 29 above, we can see that the SARIMA(4,0,5)x(0,1,1) model with the maximum likelihood method (ML) and make the following observations:

- There are 3 insignificant coefficients: ar2, ma2 and ma5.
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are slightly skewed to the left.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)      ar1    0.0386930  0.0134425
2.8784  0.003997 **  ar2   0.0011335  0.0018518   0.6121  0.540472    ar3
0.2460900     0.0015619   157.5551   <   2.2e-16   ***   ar4       0.4960287
0.0036290 136.6834 < 2.2e-16 *** ma1   0.5422815  0.1140982   4.7528
2.007e-06 *** ma2   0.5046185  0.1250595   4.0350 5.460e-05 *** ma3
-0.0646885  0.1331560  -0.4858  0.627102    ma4  -0.7617463  0.1187120
-6.4168 1.392e-10 *** ma5    0.6288237   0.1491431    4.2162 2.484e-05
*** sma1 -0.6434860  0.1222617  -5.2632 1.416e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                         Shapiro-Wilk normality test

data:  res.model
W = 0.95644, p-value = 0.0171
```

Figure 30: Parameter estimation of SARIMA(4,0,5) × (0,1,1) and Shapiro Test using CSS method

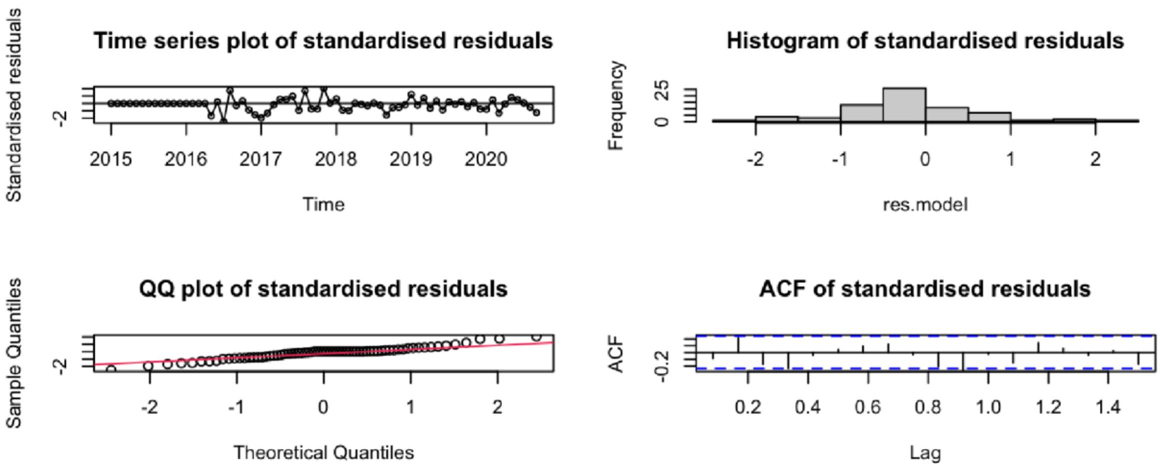

Figure 31: Residuals analysis of SARIMA(4,0,5) × (0,1,1) model using CSS method

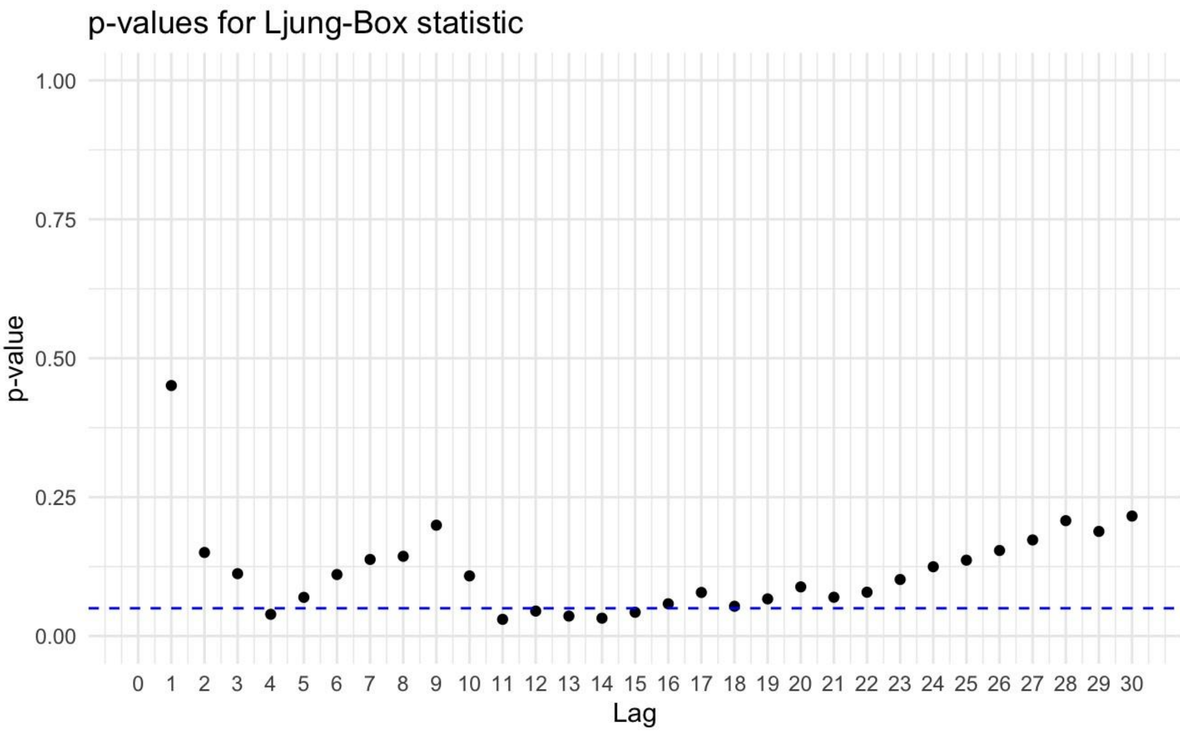

Figure 32: Residuals analysis of SARIMA(4,0,5) x (0,1,1) model using CSS method

From Figure 30, 31, and 32 above, we can see that the SARIMA(4,0,5)x(0,1,1) model with the least square method (CSS):

- There are 2 insignificant coefficients: ar2 and ma3.
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are normally distributed.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that some of the p-values are below the significance level, and there is one significant lag in the ACF plot.
- We can say that there might be some patterns left in the residuals.

- z test of coefficients:

```
      Estimate Std. Error z value  Pr(>|z|)
ar1  -0.83608    0.23875 -3.5020 0.0004618 ***
ar2   0.24009    0.13052  1.8395 0.0658454 .
ar3   0.88726    0.15952  5.5620 2.666e-08 ***
ar4   0.70808    0.18705  3.7854 0.0001534 ***
ma1   1.04739    0.28264  3.7057 0.0002108 ***
ma2   0.30534    0.29426  1.0377 0.2994271
ma3  -0.72453    0.29698 -2.4397 0.0147009 *
ma4  -1.11812    0.32018 -3.4922 0.0004791 ***
ma5  -0.27826    0.22711 -1.2252 0.2205035
sma1 -0.81362    0.32166 -2.5294 0.0114244 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


 Shapiro-Wilk normality test

data:  res.model
W = 0.94655, p-value = 0.005197
```

Figure 33: Parameter estimation of SARIMA(4,0,5) × (0,1,1) model using CSS-ML method

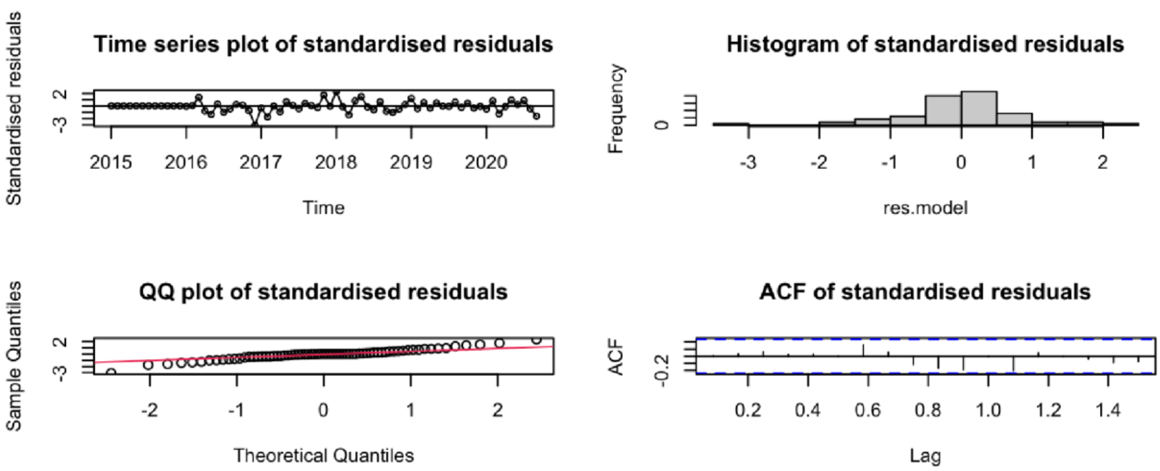

Figure 34: Residuals analysis of SARIMA(4,0,5) × (0,1,1) model using CSS-ML method

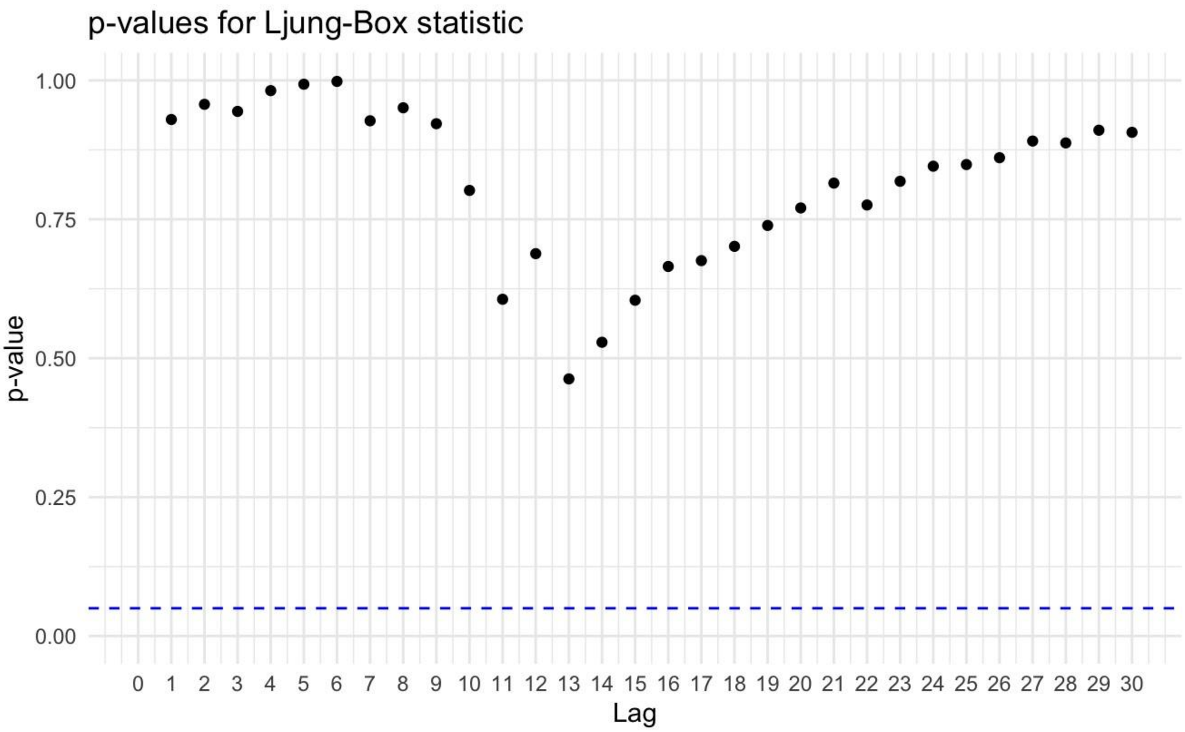

Figure 35: Residuals analysis of SARIMA(4,0,5) x (0,1,1) model using CSS-ML method

The results of the least squared method (CSS) are slightly different from the Maximum likelihood method (ML), so we can try fitting the model using CSS-ML method. From Figure 33, 34, and 35 above, we can see that the SARIMA(4,0,5)x(0,1,1) model with the CSS-ML method:

- There are 3 significant coefficients: ar2, ma2 and ma5.
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are close to normal distribution.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- The histogram of the standardised residuals shows that the residuals are slightly skewed to the left.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

This SARIMA model may not be considered as a good model because they are some insignificant coefficients. The residuals are close to normal distribution.

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

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)      ma1    0.23852     0.13330
1.7894 0.0735519 .   ma2   0.39827    0.11691  3.4067 0.0006576 ***
sma1 -0.44285    0.16413 -2.6982 0.0069705 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                        Shapiro-Wilk normality test

data:  res.model
W = 0.95993, p-value = 0.02643
```

Figure 36: Parameter estimation of SARIMA(0,0,2) x (0,1,1) model using ML method

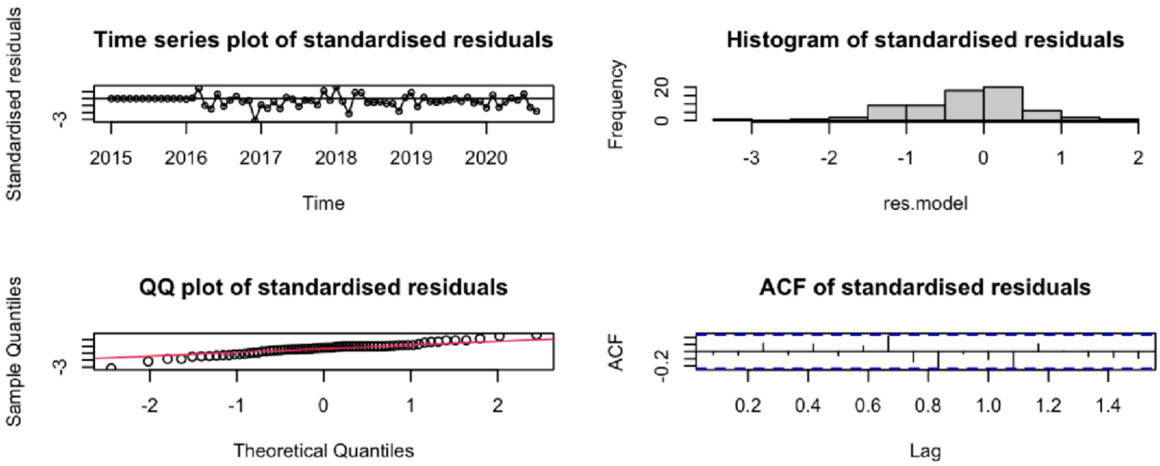

Figure 37: Residuals analysis of SARIMA(0,0,2) x (0,1,1) model using ML method

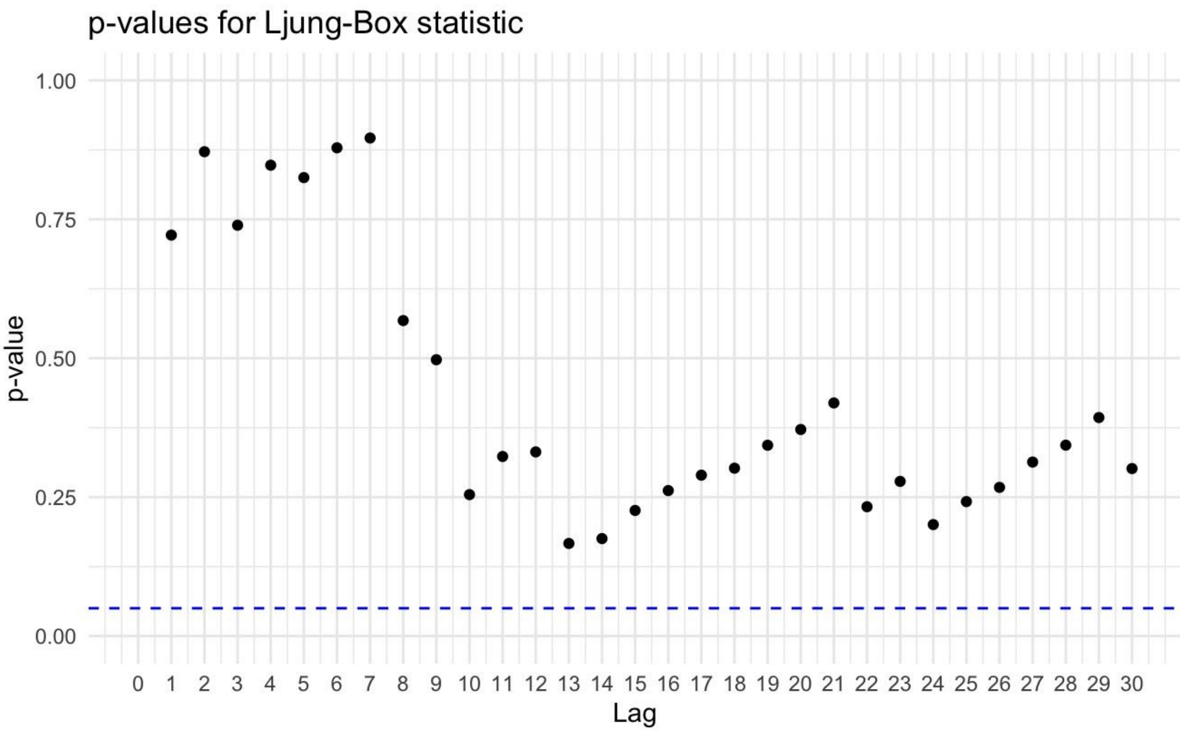

Figure 38: Residuals analysis of SARIMA(0,0,2) × (0,1,1) model using ML method

From Figure 36, 37 and 38 above, we can see that the SARIMA(0,0,2)x(0,1,1) model with the maximum likelihood method (ML):

- There are 1 insignificant coefficient: ma1
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are slightly skewed to the left.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)     ma1    0.24894     0.13302
1.8714 0.0612885 .   ma2   0.40966    0.11499  3.5625 0.0003674 ***
sma1 -0.40053    0.14386 -2.7843 0.0053649 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



                        Shapiro-Wilk normality test

data:  res.model
W = 0.95576, p-value = 0.01572
```

Figure 39: Parameter estimation of SARIMA(0,0,2) × (0,1,1) model using CSS method

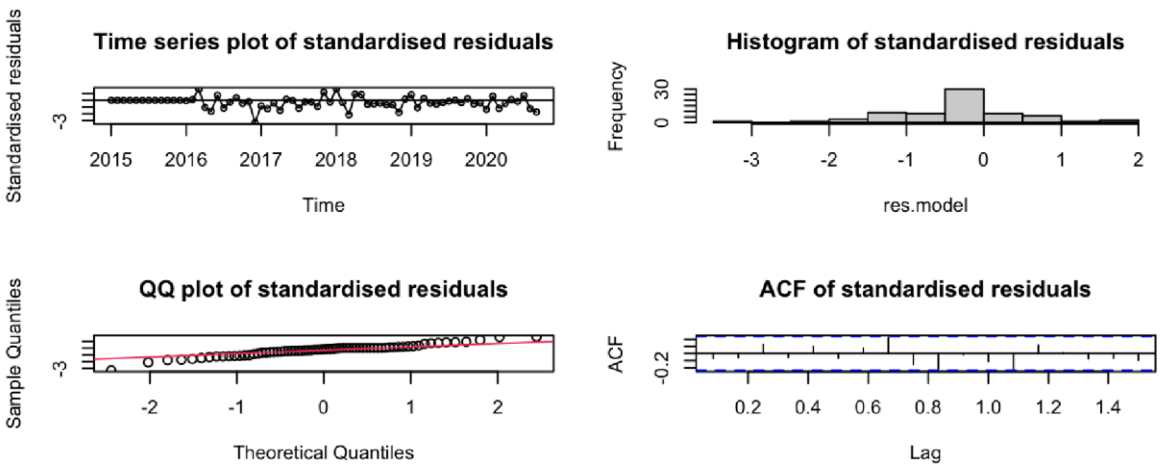

Figure 40: Residual Analysis of SARIMA(0,0,2) × (0,1,1) model using CSS method

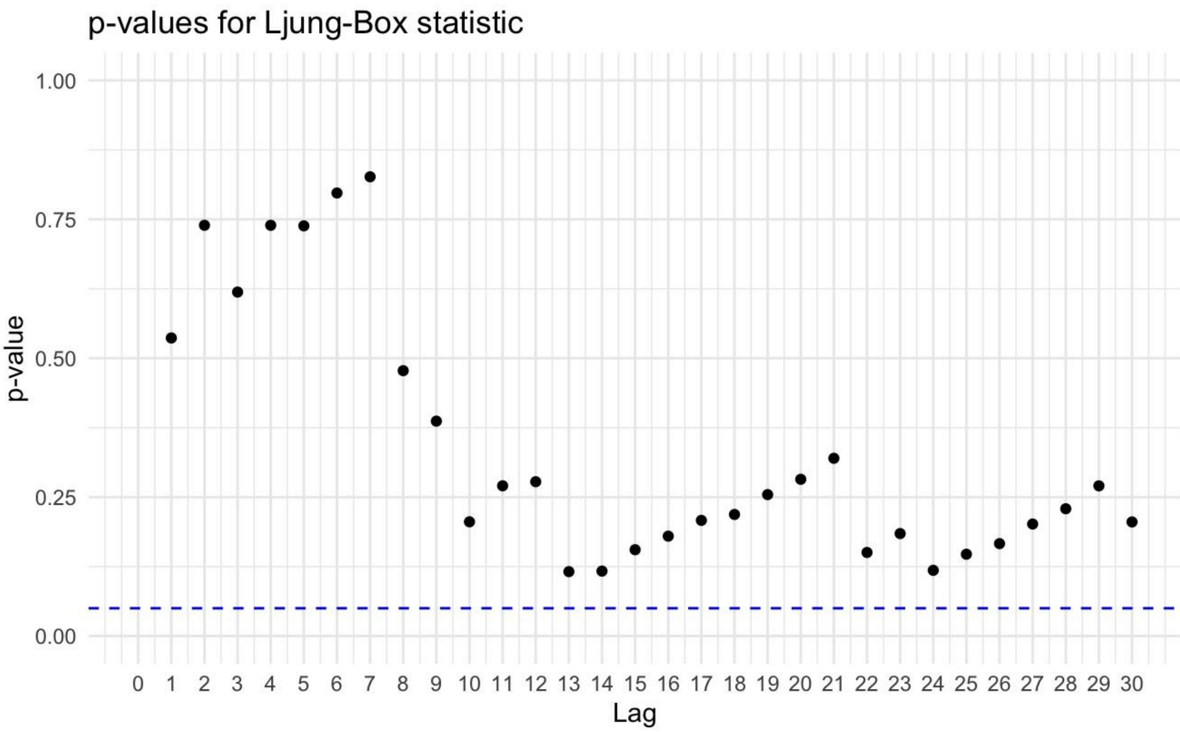

Figure 41: Residual Analysis of SARIMA(0,0,2) x (0,1,1) model using CSS method

From Figure 39, 40, and 41 above, we can see that the SARIMA(0,0,2)x(0,1,1) model with the least square method (CSS):

- There are 1 insignificant coefficient: ma1
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are slightly skewed to the left.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)      ma1    0.23858     0.13329
1.7899 0.0734739 .   ma2   0.39829    0.11691  3.4068 0.0006574 ***
sma1 -0.44291    0.16413 -2.6985 0.0069661 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                        Shapiro-Wilk normality test

data:  res.model
W = 0.95994, p-value = 0.02645
```

Figure 42: Parameter estimation of SARIMA(0,0,2) × (0,1,1) model using CSS-ML method

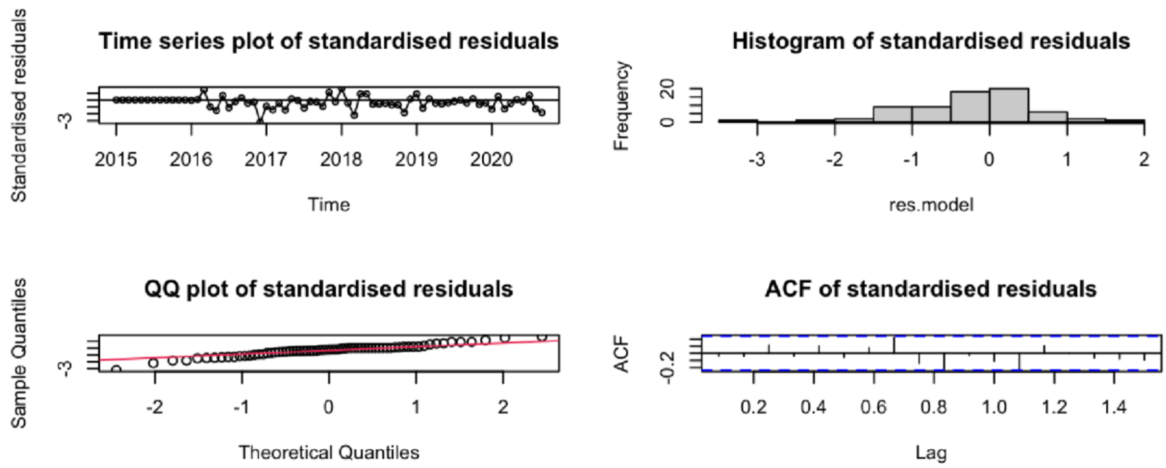

Figure 43: Residuals analysis of SARIMA(0,0,2) × (0,1,1) model using CSS-ML method

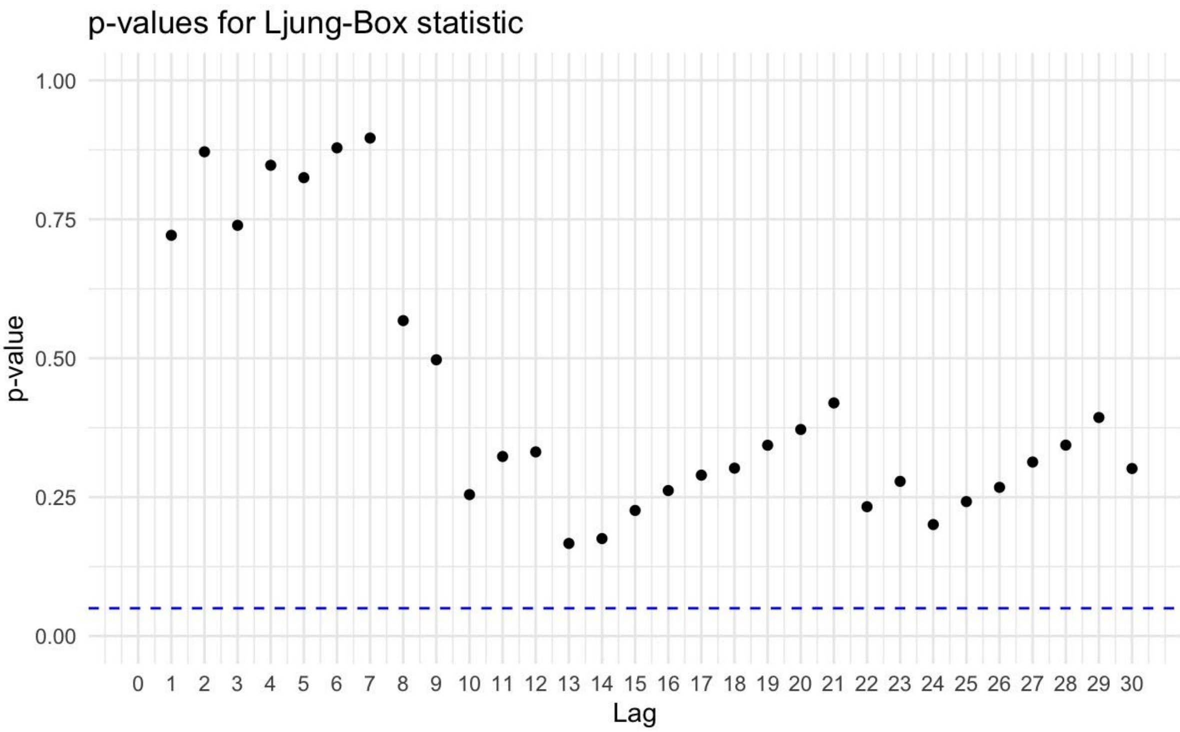

Figure 44: Residuals analysis of SARIMA(0,0,2) x (0,1,1) model using CSS-ML method

From Figure 42, 43, and 44 above, we can see that the SARIMA(0,0,2)x(0,1,1) model with the CSS-ML method:

- There are 1 insignificant coefficient: ma1
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are slightly skewed to the left.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

The results of for all three methods are consistent. In conclusion, this SARIMA model may not be considered as a good model because they are some insignificant coefficients. The residuals are close to normal distribution.

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

```
z test of coefficients:


    Estimate Std. Error z value Pr(>|z|)    ma1   0.27117    0.13383
2.0262 0.042740 *  ma2   0.38020   0.11749  3.2360 0.001212 **
ma3    0.29245     0.14896   1.9632  0.049617  *   sma1  -0.46796
0.16549 -2.8276 0.004689 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                         Shapiro-Wilk normality test


data:  res.model
W = 0.95559, p-value = 0.0154
```

Figure 45: Parameter estimation of SARIMA(0,0,3) x (0,1,1) model using ML method

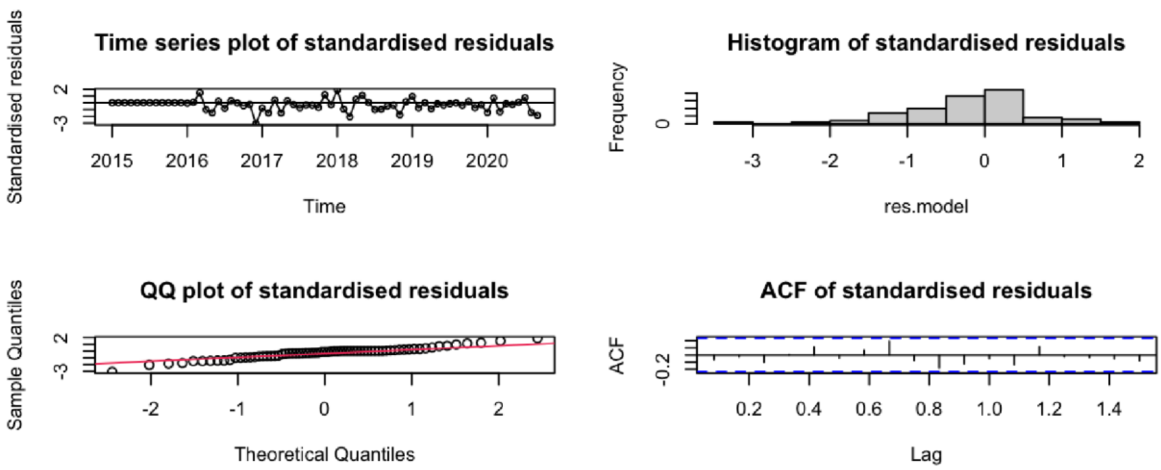

Figure 46: Residual analysis of SARIMA(0,0,3) x (0,1,1) model using ML method

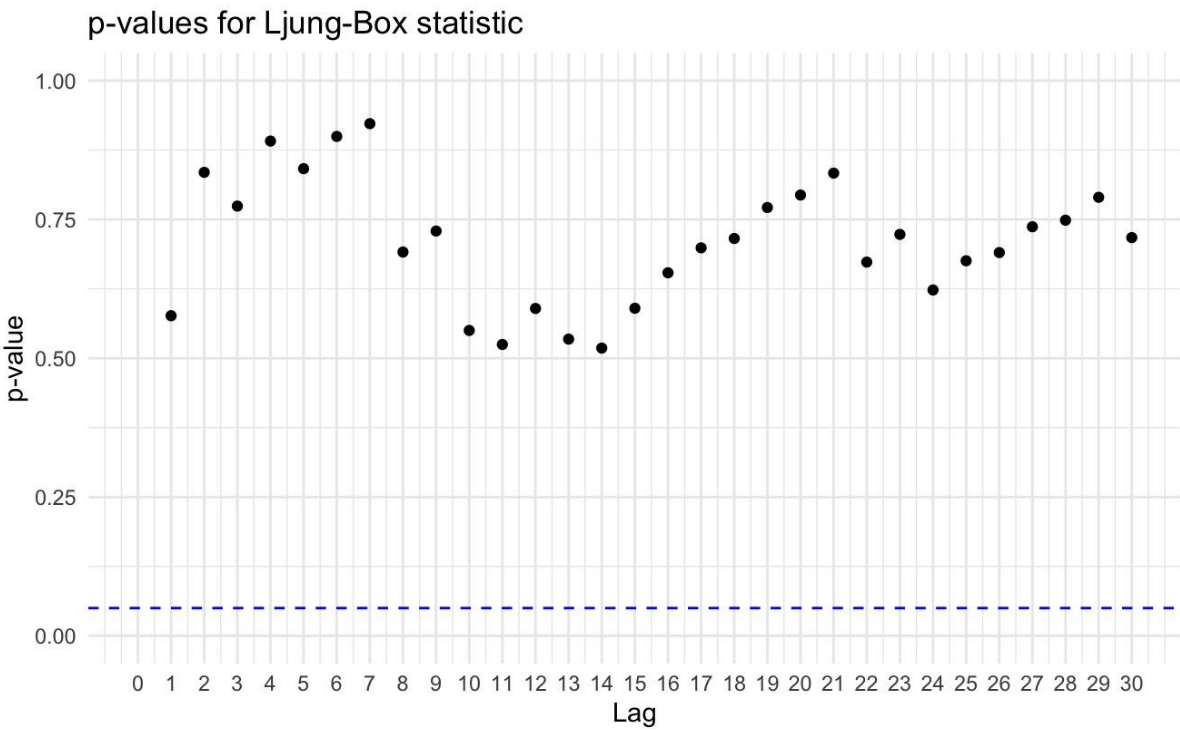

Figure 47: Residual analysis of SARIMA(0,0,3) x (0,1,1) model using ML method

From Figure 45, 46, and 47 above, we can see that the SARIMA(0,0,3)x(0,1,1) model with the maximum likelihood method (ML):

- All the coefficients are significant.
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are slightly skewed to the left.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)     ma1   0.26446    0.13469  1.9635
0.0495927 *   ma2   0.39291    0.11733  3.3488 0.0008116 *** ma3
0.25250    0.16204  1.5583 0.1191733     sma1 -0.39924    0.14800 -2.6975
0.0069870 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


                    Shapiro-Wilk normality test

data:  res.model
W = 0.9415, p-value = 0.002906
```

Figure 48: Parameter estimation of SARIMA(0,0,3) × (0,1,1) model using CSS method

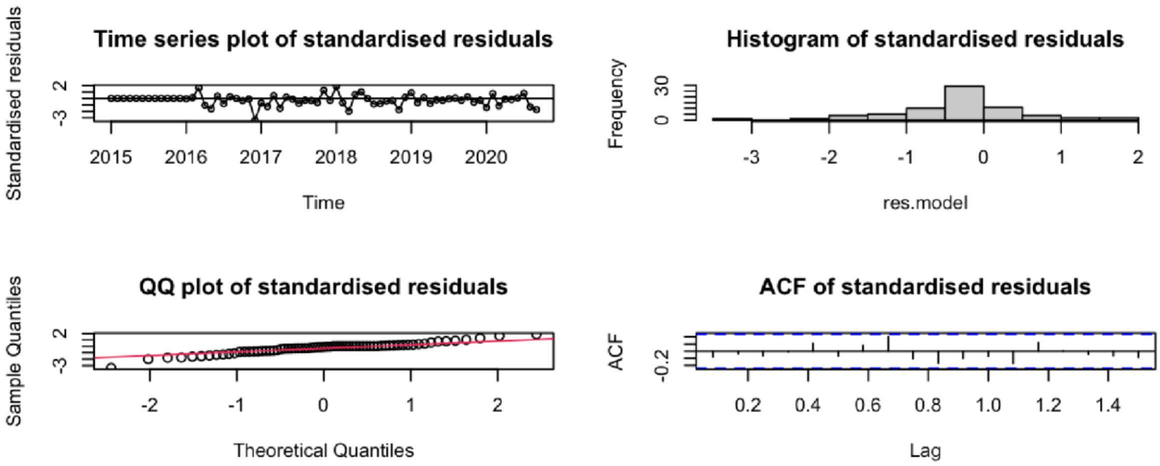

Figure 49: Residual analysis of SARIMA(0,0,3) × (0,1,1) model using CSS method

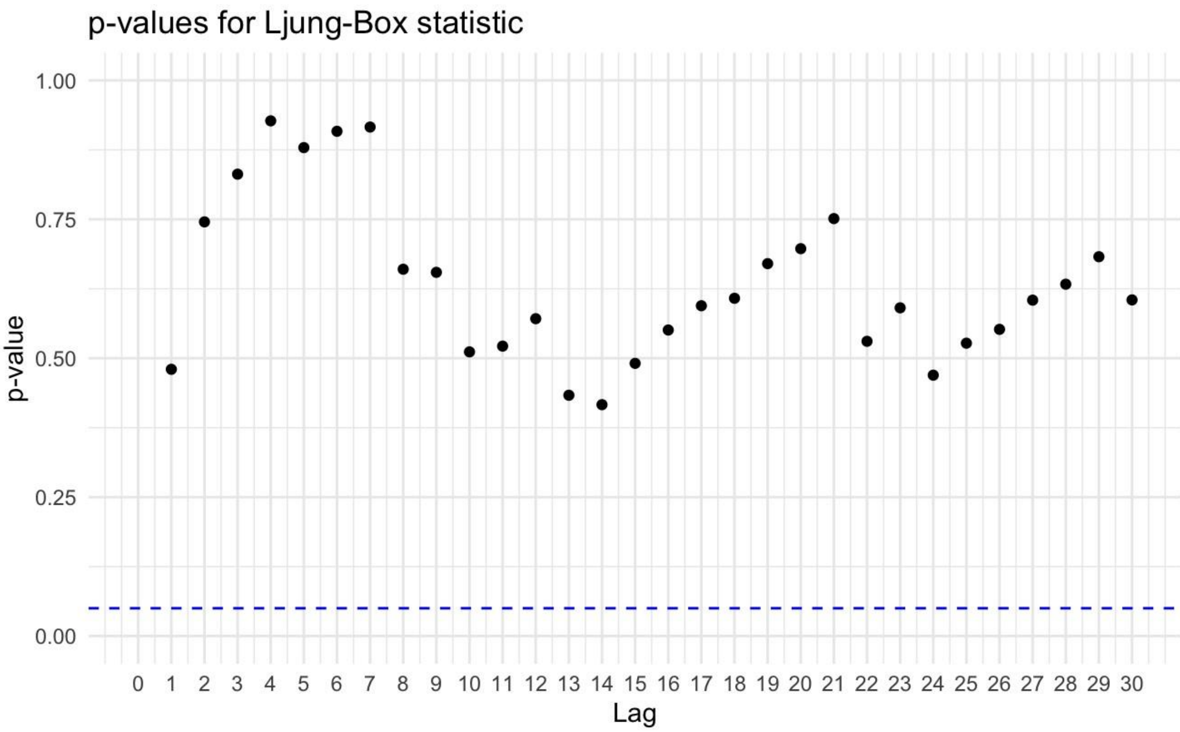

Figure 50: Residual analysis of SARIMA(0,0,3) x (0,1,1) model using CSS method

From Figure 48, 49, and 50 above, we can see that the SARIMA(0,0,3)x(0,1,1) model with the least square method (CSS):

- There is one significant lag: ma3
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are normally distributed.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)    ma1   0.27116    0.13382
2.0262 0.042742 *  ma2   0.38020    0.11748  3.2362 0.001211 **
ma3    0.29252      0.14895   1.9639  0.049546  *    sma1  -0.46791
0.16549 -2.8274 0.004693 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                        Shapiro-Wilk normality test

data:  res.model
W = 0.95559, p-value = 0.0154
```

Figure 51: Parameter estimation of SARIMA(0,0,3) × (0,1,1) model using CSS-ML method

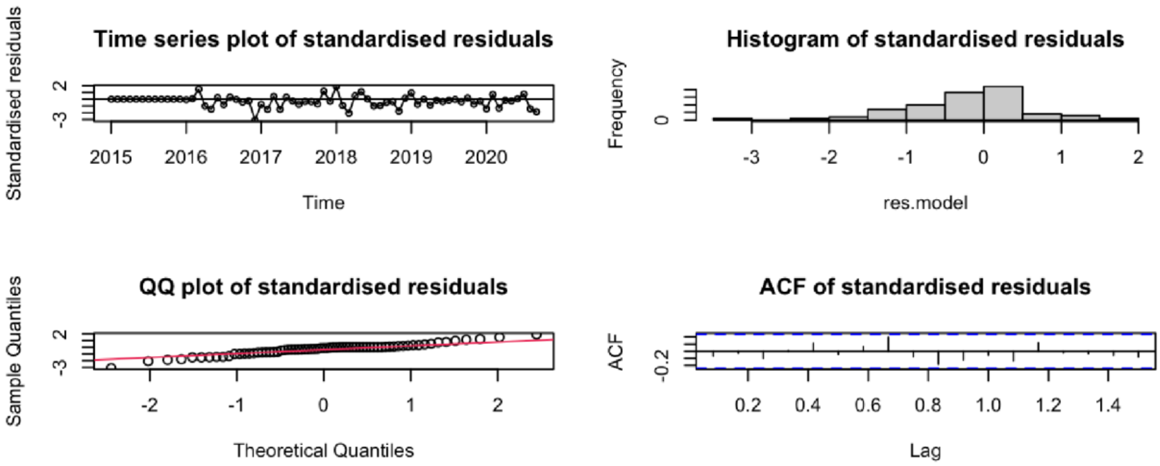

Figure 52: Residual analysis of SARIMA(0,0,3) × (0,1,1) model using CSS-ML method

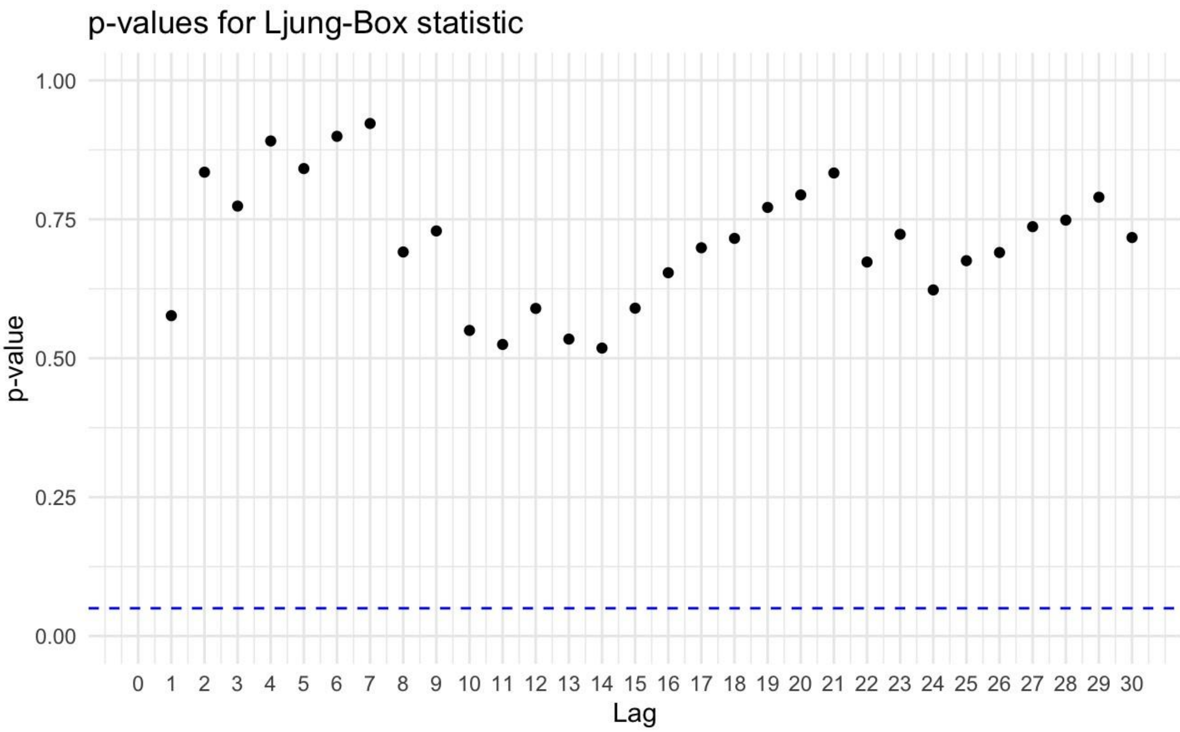

Figure 53: Residual analysis of SARIMA(0,0,3) x (0,1,1) model using CSS-ML method

From Figure 51, 52, and 53 above, we can see that the SARIMA(0,0,3)x(0,1,1) model with the CSS-ML:

- All the coefficients are significant.
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are skewed to the left.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

This SARIMA model may be considered as a good model as all the coefficients are significant and there is no pattern left in the residuals.

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

```
z test of coefficients:


     Estimate Std. Error  z value  Pr(>|z|)      ar1    0.9995874  0.0042171
237.0297 < 2.2e-16 *** ma1  -0.7963119  0.1364214  -5.8371  5.31e-09
*** ma2  -0.0072259  0.1591543  -0.0454 0.9637872     sma1 -0.9405617
0.2786525  -3.3754 0.0007371 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                        Shapiro-Wilk normality test


data:  res.model
W = 0.93997, p-value = 0.002442
```

Figure 54: Parameter estimation of SARIMA(1,0,2) x (0,1,1) model using ML method

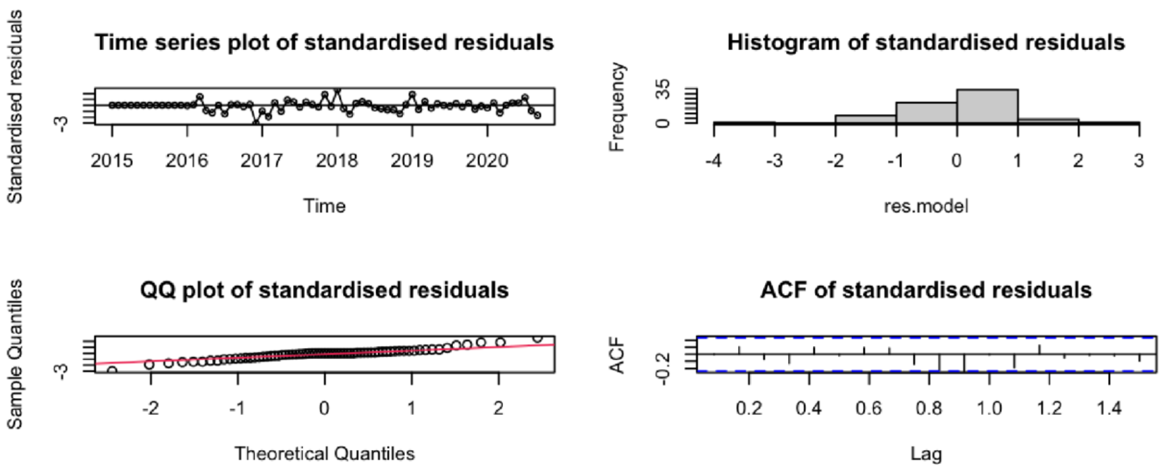

Figure 55: Residual analysis of SARIMA(1,0,2) x (0,1,1) model using ML method

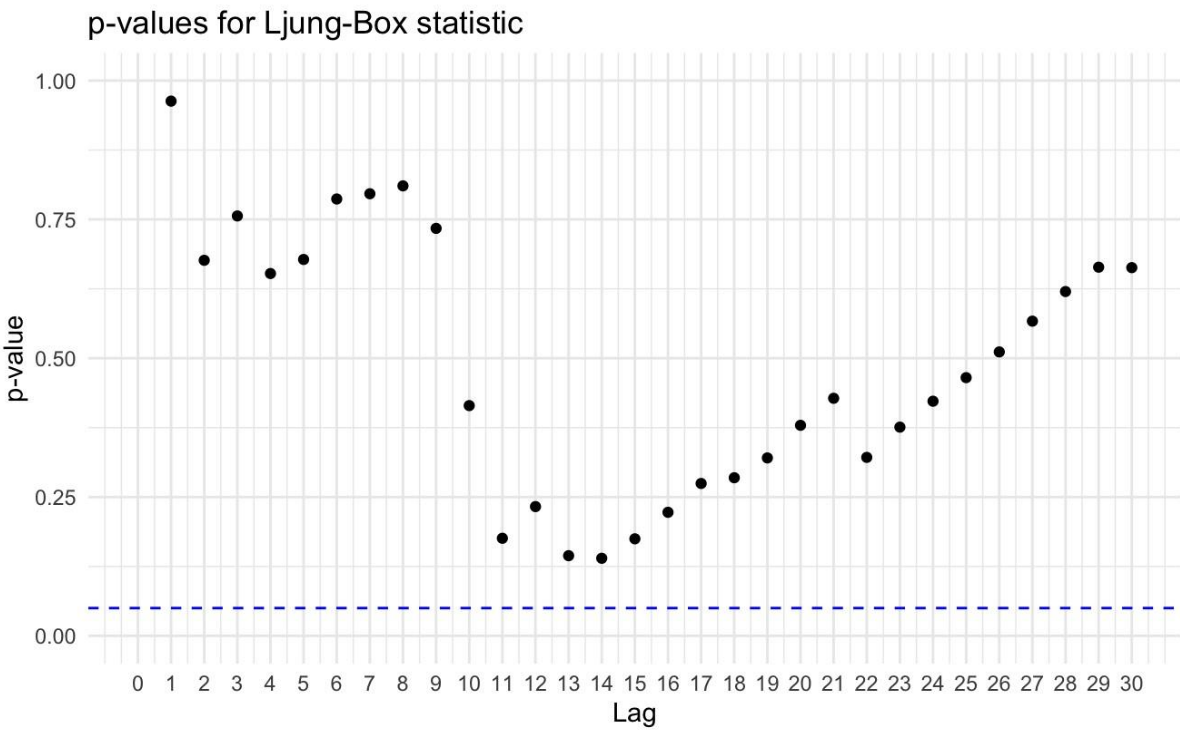

Figure 56: Residual analysis of SARIMA(1,0,2) x (0,1,1) model using ML method

From Figure 54, 55, and 56 above, we can see that the SARIMA(1,0,2)x(0,1,1) model with the Maximum Likelihood (ML):

- There is one insignificant coefficient: ma2
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are skewed to the left.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)     ar1   0.983147   0.047769
20.5811 < 2.2e-16 *** ma1  -0.859151   0.153858 -5.5841 2.350e-08 ***
ma2   0.097856   0.140188  0.6980    0.4852     sma1 -0.703690   0.138640 -
5.0757 3.862e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



                         Shapiro-Wilk normality test


data:  res.model
W = 0.943, p-value = 0.003446
```

Figure 57: Parameter estimation of SARIMA(1,0,2) × (0,1,1) model using CSS method

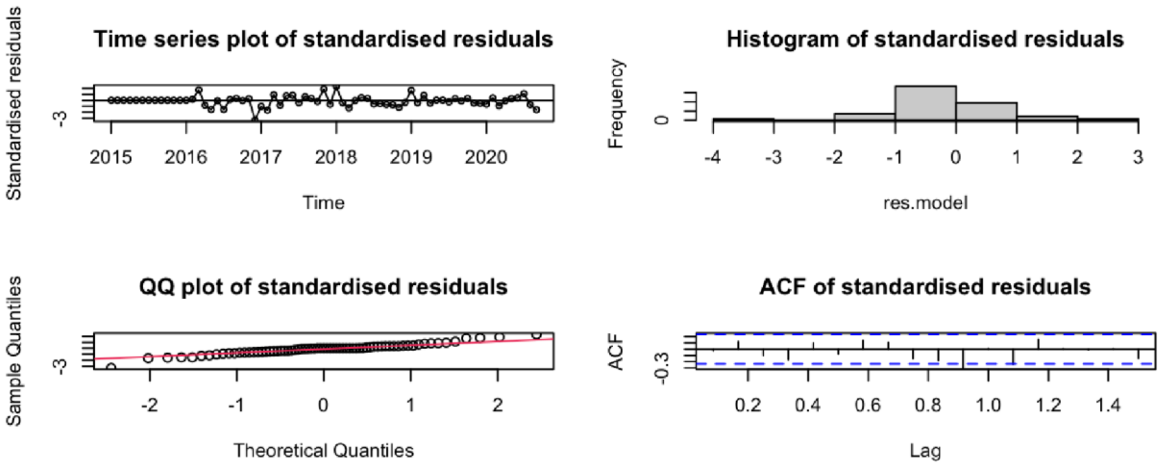

Figure 58: Residual analysis of SARIMA(1,0,2) × (0,1,1) model using CSS method

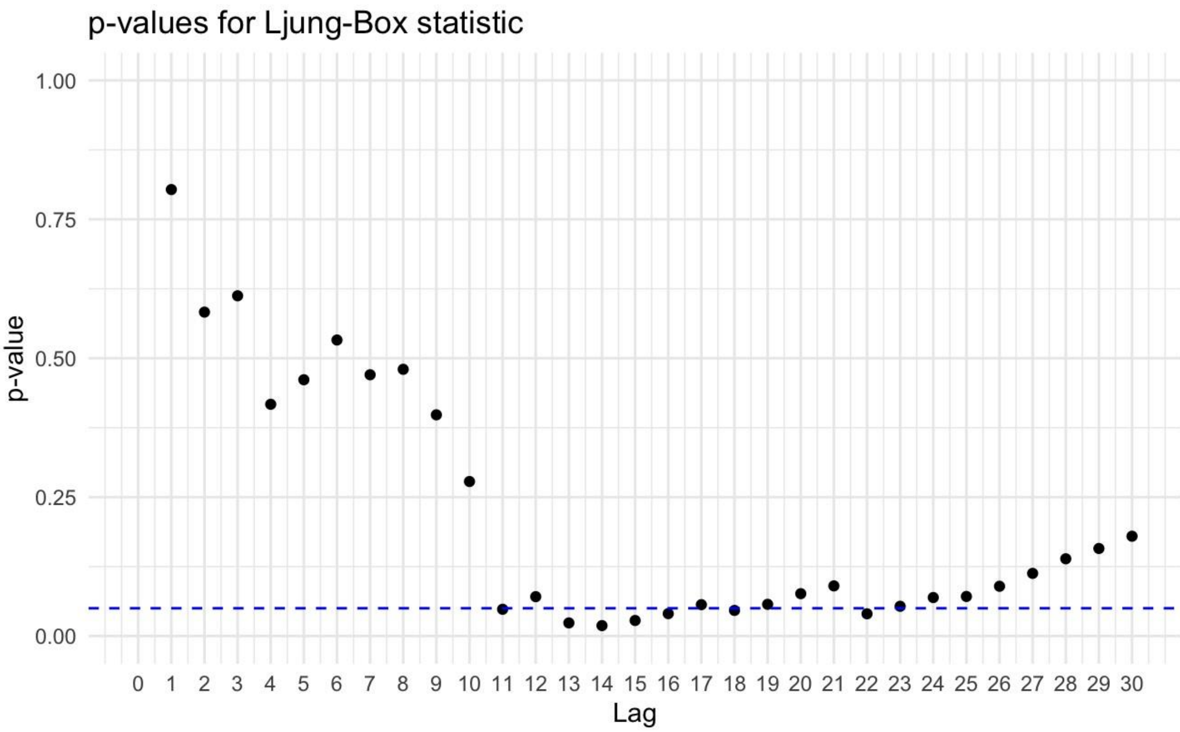

Figure 59: Residual analysis of SARIMA(1,0,2) x (0,1,1) model using CSS method

From Figure 57, 58, and 59 above, we can see that the SARIMA(1,0,2)x(0,1,1) model with the least square method (CSS):

- There is one insignificant coefficient: ma2
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are close to normal distribution.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that some of the p-values are below the significance level, and there is one significant lag in the ACF plot.
- We can say that there might be some patterns left in the residuals.

```
z test of coefficients:

     Estimate Std. Error  z value  Pr(>|z|)     ar1    0.9996989  0.0034235
292.0142 < 2.2e-16 *** ma1  -0.7966874  0.1365757  -5.8333 5.434e-09
*** ma2  -0.0085174  0.1600587  -0.0532 0.9575614     sma1 -0.9483167
0.2735630  -3.4665 0.0005272 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                    Shapiro-Wilk normality test

data:  res.model
W = 0.93975, p-value = 0.002383
```

Figure 60: Parameter estimation of SARIMA(1,0,2) × (0,1,1) model using CSS-ML method

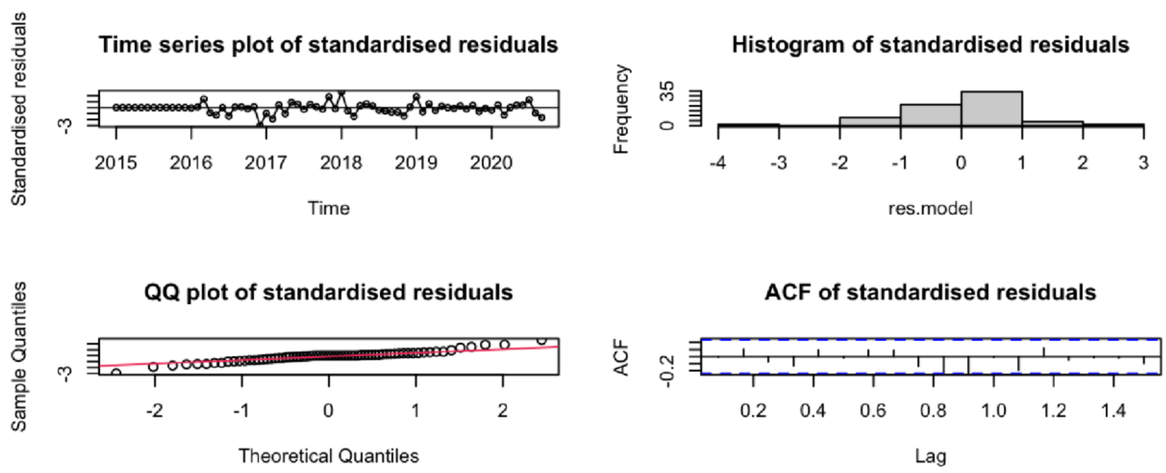

Figure 61: Residual analysis of SARIMA(1,0,2) × (0,1,1) model using CSS method

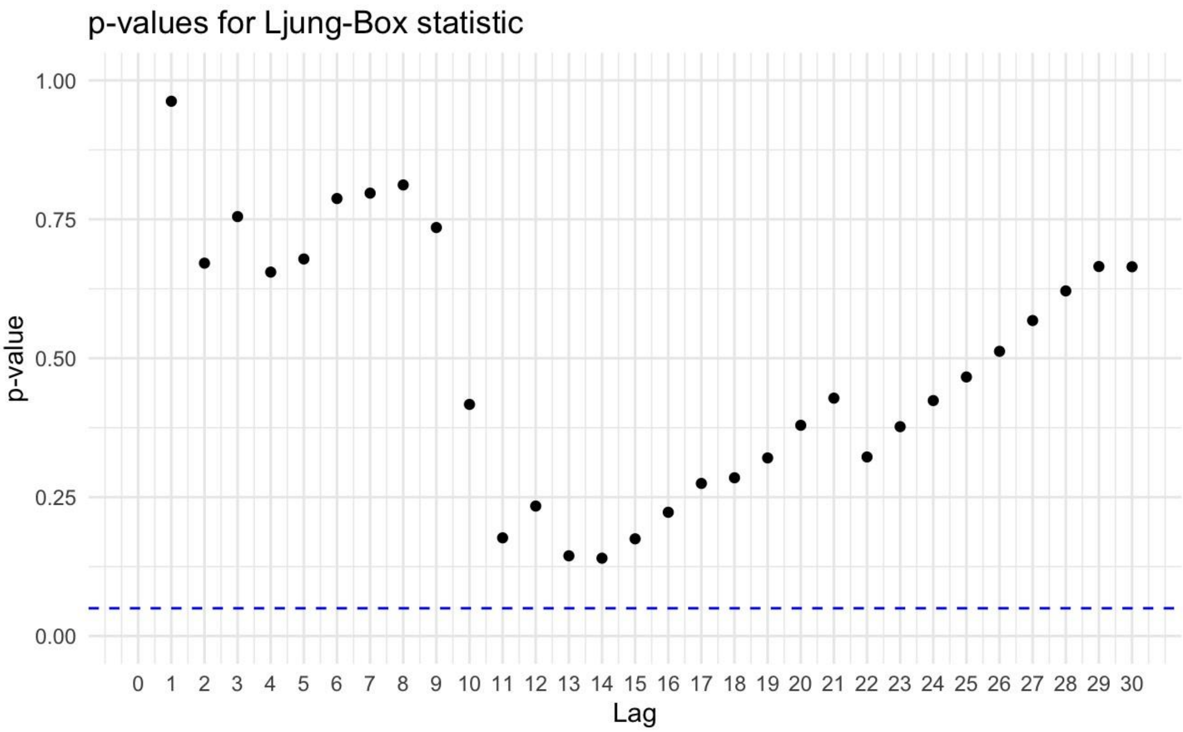

Figure 62: Residual analysis of SARIMA(1,0,2) x (0,1,1) model using CSS method

From Figure 60, 61, and 62 above, we can see that the SARIMA(1,0,2)x(0,1,1) model with the CSS-ML method.

- There is one insignificant coefficient: ma2
- The Shapiro-Walk Normality Test shows a small p-value close to 0, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows that the residuals are close to normal distribution.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that alle of the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there might be some patterns left in the residuals.

The results of all 3 methods are consistent. In conclusion, this SARIMA model may not be considered as a good model because they are some insignificant coefficients. The residuals are close to normal distribution.

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

```
z test of coefficients:


     Estimate  Std. Error   z value  Pr(>|z|)     ar1   0.99998191  0.00019433
5145.8217 < 2.2e-16 *** ma1  -0.86839949  0.14766738   -5.8808 4.083e-09
*** ma2   0.16319159  0.17891851    0.9121   0.36172     ma3  -0.26674875
0.14000390   -1.9053   0.05674 .   sma1 -0.89246603  0.37409069   -2.3857
0.01705 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                        Shapiro-Wilk normality test


data:  res.model
W = 0.93439, p-value = 0.001317
```

Figure 63: Coefficient estimation of SARIMA(1,0,3) x (0,1,1) model using ML method

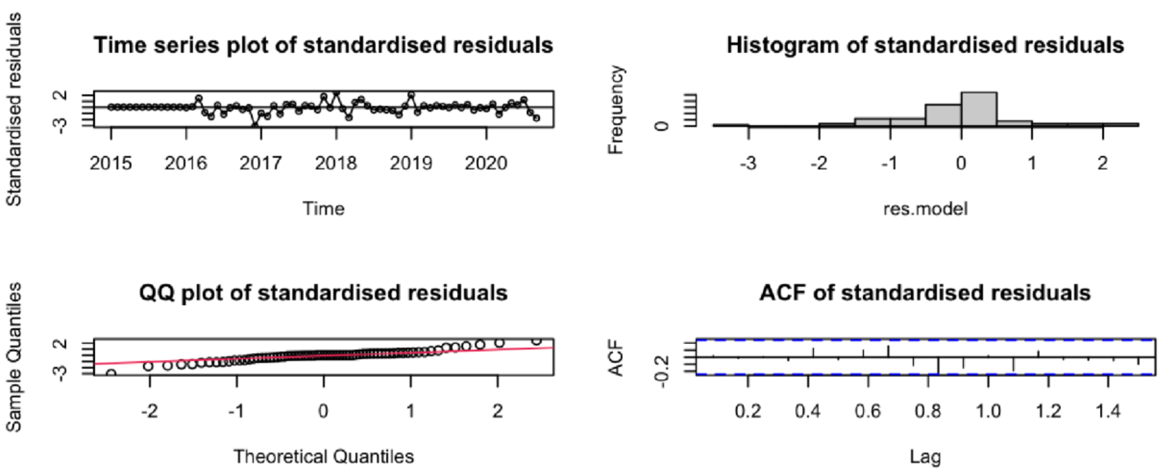

Figure 64: Residual analysis of SARIMA(1,0,3) x (0,1,1) model using ML method

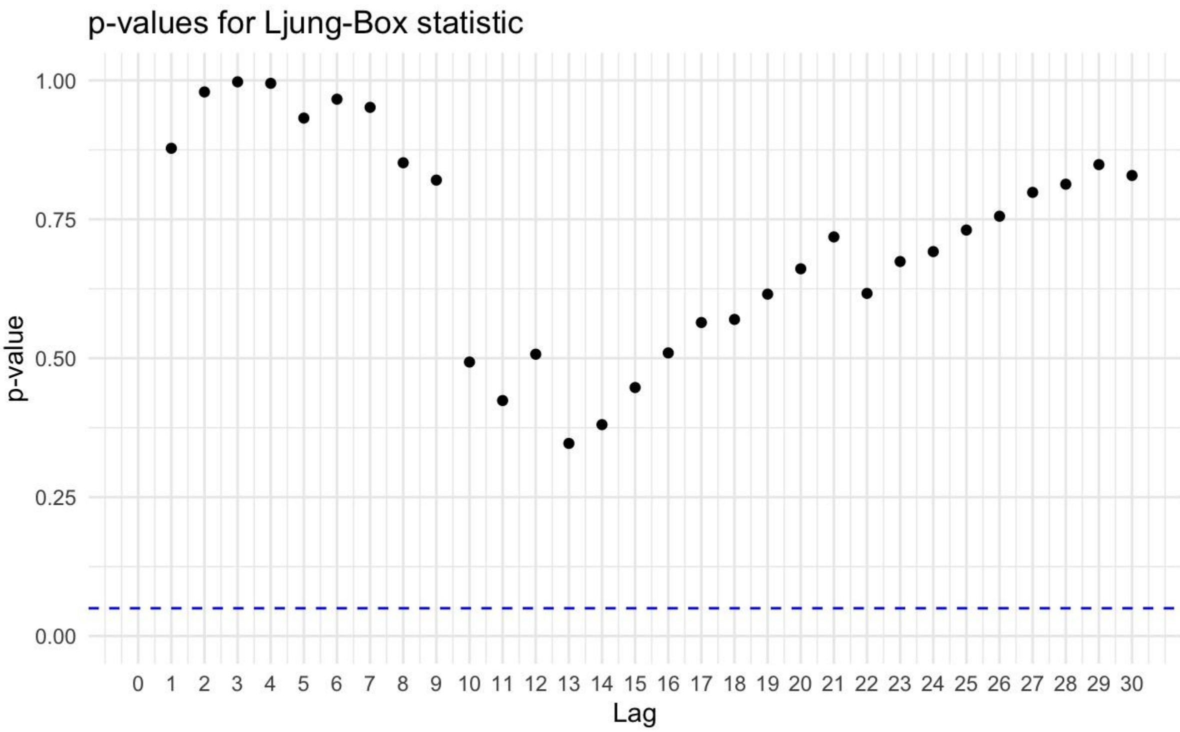

Figure 65: Residual analysis of SARIMA(1,0,3) x (0,1,1) model using ML method

Figure 63, 64 and 65 show SARIMA(1,0,3)x(0,1,1) model with the maximum likelihood method (ML):

- There are 1 slightly insignificant coefficient, ma3, and one insignificant coefficient, ma2.
- The Shapiro-Walk Normality Test shows a p-value < 0.05, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows the model is slightly left skewed.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- There is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)     ar1   1.013841   0.018637
54.4002 < 2.2e-16 *** ma1  -0.897591   0.143401 -6.2593 3.867e-10 ***
ma2   0.260087   0.170444  1.5259     0.127     ma3  -0.304281   0.132852 -
2.2904     0.022 *   sma1 -0.715961   0.135501 -5.2838 1.265e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                        Shapiro-Wilk normality test

data:  res.model
W = 0.92965, p-value = 0.000791
```

Figure 66: Parameter estimation of SARIMA(1,0,3) × (0,1,1) model using CSS method

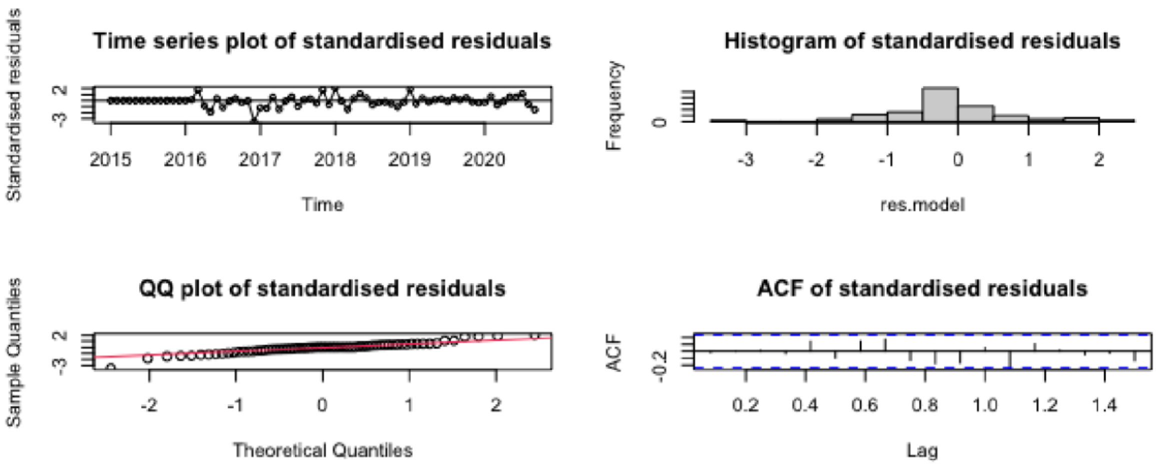

Figure 67: Residual analysis of SARIMA(1,0,3) × (0,1,1) model using CSS method

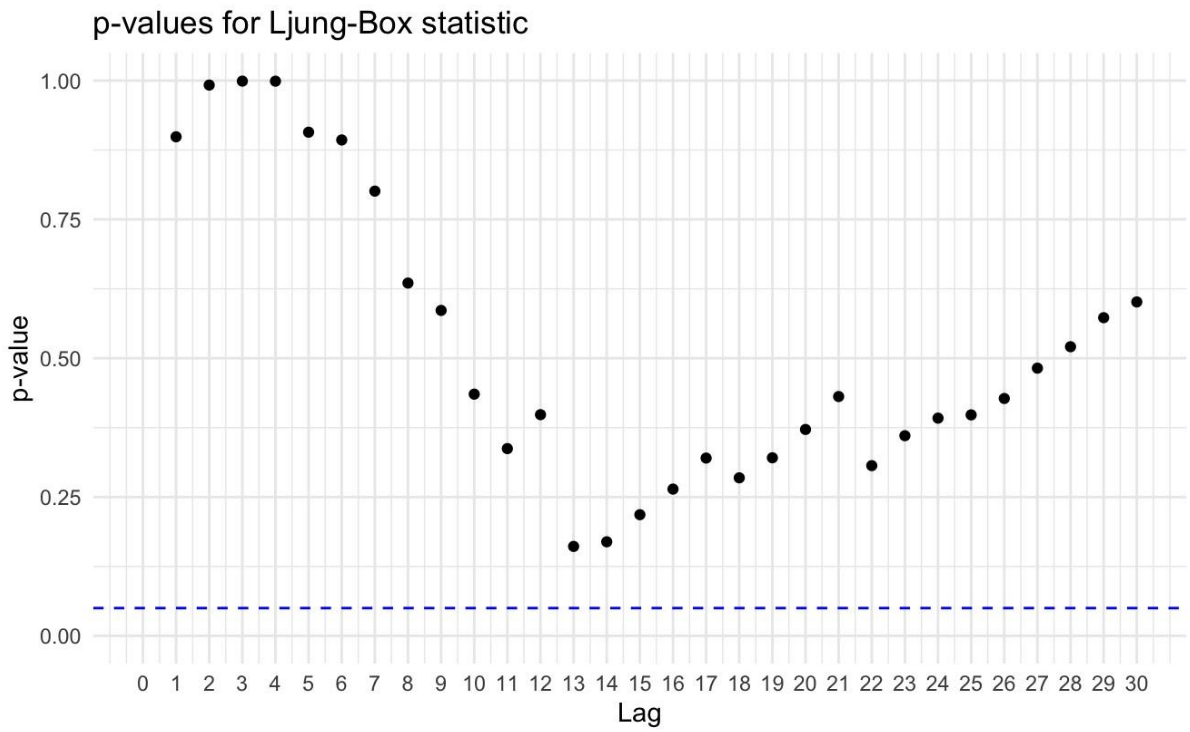

Figure 68: Residual analysis of SARIMA(1,0,3) x (0,1,1) model using CSS method

Figure 66, 67, and 68 show SARIMA(1,0,3)x(0,1,1) model with the least square method (CSS):

- There are 1 insignificant coefficient: ma2
- The Shapiro-Walk Normality Test shows a p-value lower than 0.05, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows the model is not skewed.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

Since CSS and ML results are similar, we will not explore the CSS-ML residual analysis method for SARIMA(1,0,3)x(0,1,1).

5.6 SARIMA(2,0,6) x (0,1,1)

```
z test of coefficients:

     Estimate Std. Error z value Pr(>|z|)   ar1   0.608919   0.570211
1.0679  0.28557   ar2   0.220578   0.414967  0.5316  0.59503   ma1
-0.408607   0.533797 -0.7655  0.44399   ma2  -0.041777   0.304509
-0.1372   0.89088    ma3   -0.200178    0.213015  -0.9397   0.34735
ma4    0.097931    0.212223   0.4615   0.64448    ma5    0.450800
0.268757   1.6773   0.09347 .  ma6   -0.073742    0.298313  -0.2472
0.80476    sma1  -0.677800    0.299243  -2.2650  0.02351 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



                        Shapiro-Wilk normality test


data:  res.model
W = 0.97592, p-value = 0.2033
```

Figure 69: Parameter estimation of SARIMA(2,0,6) x (0,1,1) model using ML method

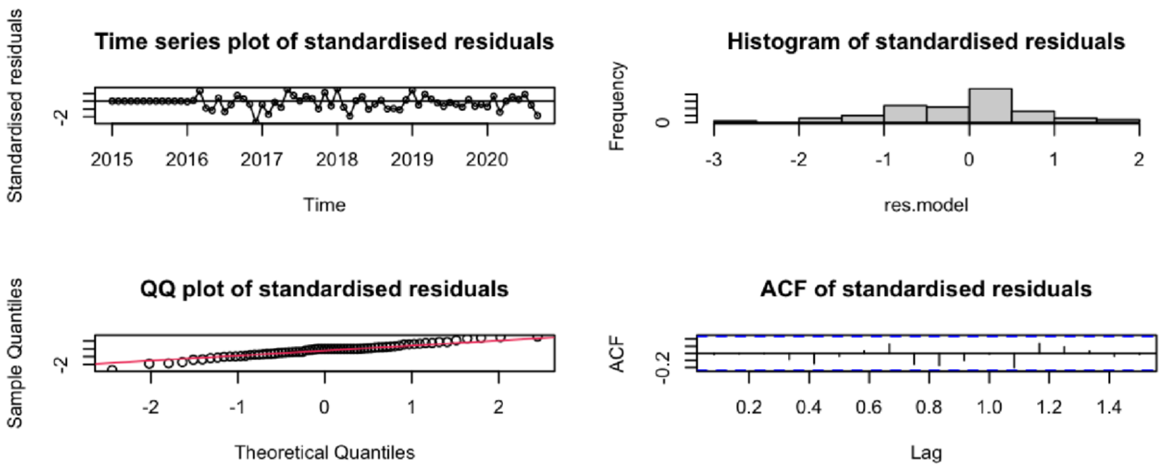

Figure 70: Residual analysis of SARIMA(2,0,6) x (0,1,1) model using ML method

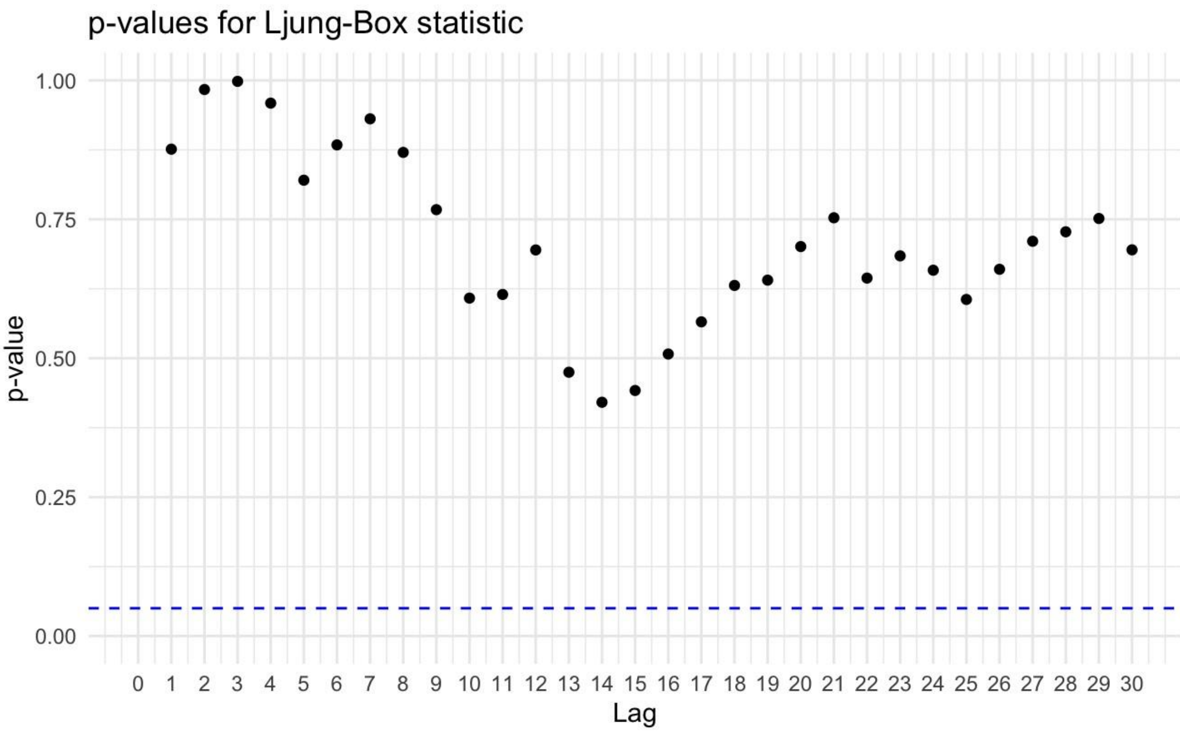

Figure 71: Residual analysis of SARIMA(2,0,6) x (0,1,1) model using ML method

Figure 69,10 and 71 shows SARIMA(2,0,6)x(0,1,1) model with the maximum likelihood method (ML):

- There are 1 slightly insignificant coefficient: ma5
- The Shapiro-Walk Normality Test shows a p-value > 0.05, so we can accept the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows the model is slightly left skewed.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)     ar1  -0.453150   0.368670 -
1.2291 0.2190158      ar2  -0.534124   0.245093 -2.1793 0.0293120 *
ma1    0.780877    0.371605   2.1014 0.0356091 *    ma2    1.263218
0.331810  3.8071 0.0001406 *** ma3    0.657629    0.368354  1.7853
0.0742101 .    ma4    0.502343    0.293493   1.7116 0.0869705 .    ma5
0.303732   0.217477  1.3966 0.1625287     ma6  -0.078214   0.195846
-0.3994 0.6896222      sma1 -0.431274    0.164501 -2.6217 0.0087490
**
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



                        Shapiro-Wilk normality test


data:  res.model
W = 0.95153, p-value = 0.009393
```

Figure 72: Parameter estimation of SARIMA(2,0,6) × (0,1,1) model using CSS method

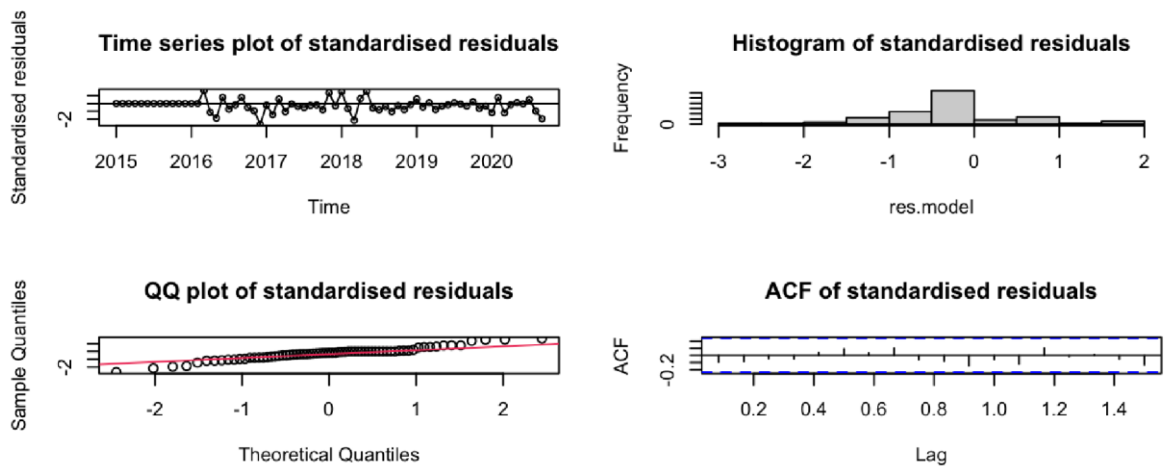

Figure 73: Residual analysis of SARIMA(2,0,6) × (0,1,1) model using CSS method

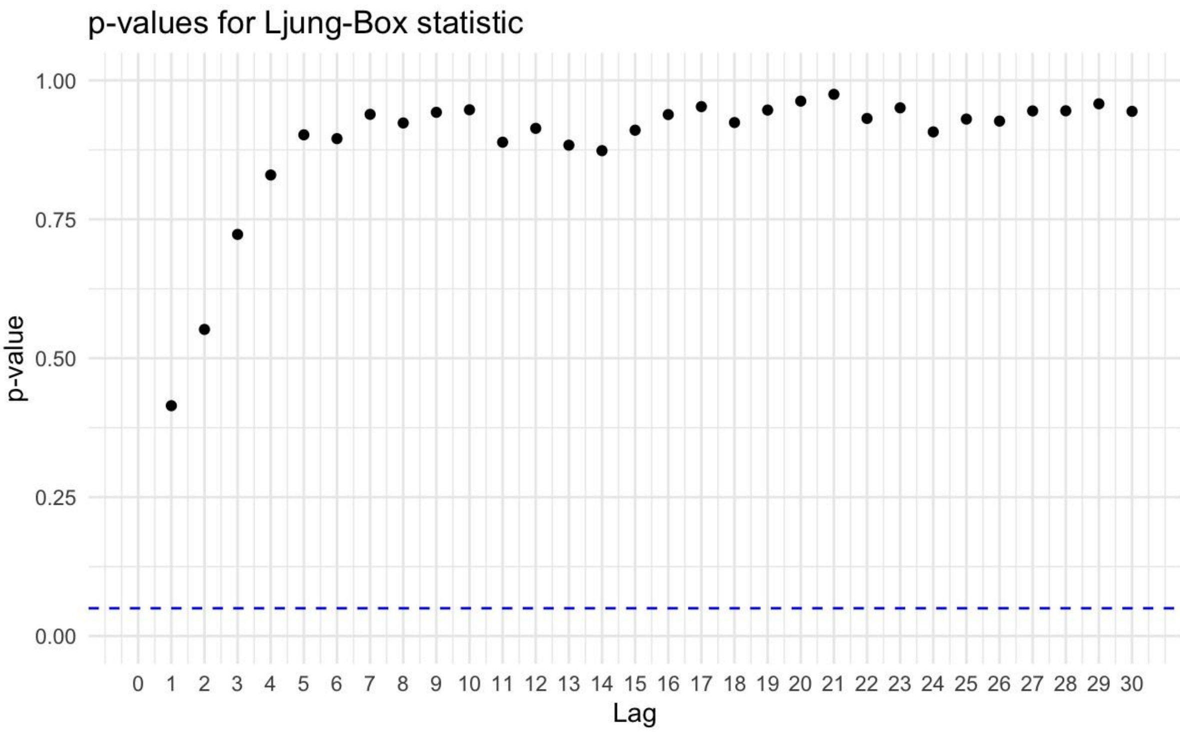

Figure 74: Residual analysis of SARIMA(2,0,6) x (0,1,1) model using CSS method

Figure 72, 73, and 74 show SARIMA(2,0,6)x(0,1,1) model with the least square method (CSS):

- There are 2 insignificant coefficients: ma3 and ma4.
- The Shapiro-Walk Normality Test shows a p-value lower than 0.05, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows the model is slightly left skewed.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- We can say that there is no pattern left in the residuals because all the lags are insignificant.

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)     ar1  -0.535770   0.468171 -1.1444
0.2524629     ar2  -0.529455   0.228046 -2.3217 0.0202489 *    ma1
0.905907   0.487341  1.8589 0.0630448 .   ma2   1.364794   0.390412
3.4958 0.0004727 *** ma3   0.750947   0.495759  1.5147 0.1298373
ma4   0.562650   0.356318  1.5791 0.1143205     ma5   0.391003   0.279579
1.3985 0.1619500     ma6  -0.034411   0.293280 -0.1173 0.9065966     sma1
-0.459155   0.168456 -2.7257 0.0064172 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



                         Shapiro-Wilk normality test


data:  res.model
W = 0.94948, p-value = 0.007348
```

Figure 75: Residual analysis of SARIMA(2,0,6) x (0,1,1) model using CSS-ML method

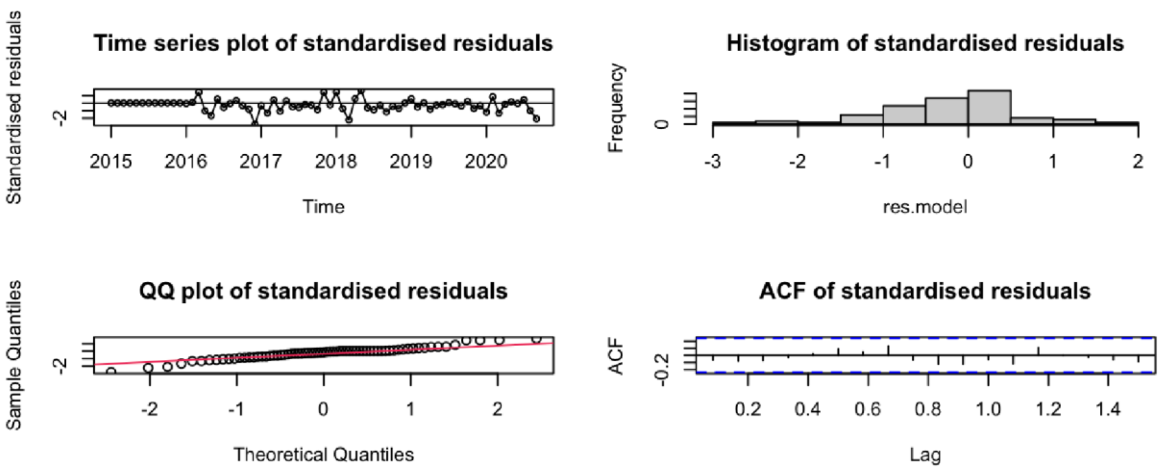

Figure 76: Parameter estimation of SARIMA(2,0,6) x (0,1,1) model using CSS-ML method

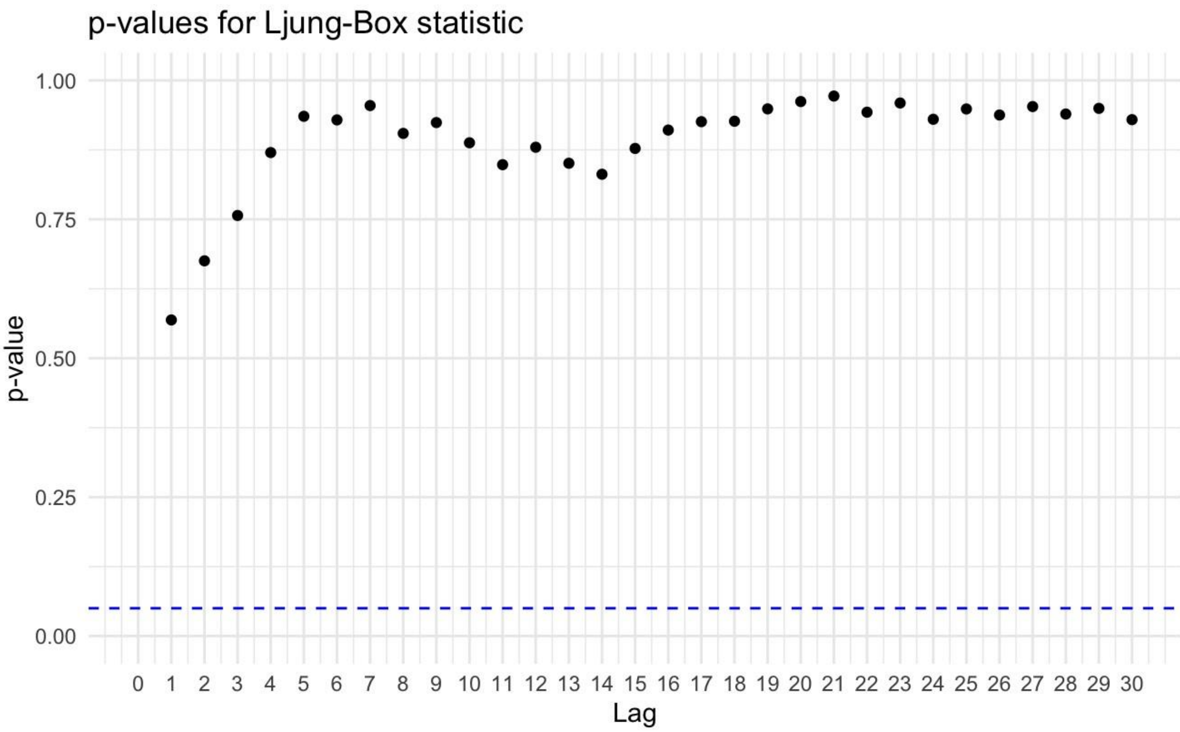

Figure 77: Parameter estimation of SARIMA(2,0,6) x (0,1,1) model using CSS-ML method

Figure 75, 76, and 77 show SARIMA(2,0,6)x(0,1,1) model with the CSS-ML method:

- There are 1 insignificant coefficient: ma1.
- The Shapiro-Walk Normality Test shows a p-value lower than 0.05, so we can reject the null hypothesis that the residuals are normally distributed.
- The histogram of the standardised residuals shows the model is slightly left skewed.
- From the Q-Q plot above, we observe that most data points fall along the straight line.
- From the Ljung-Box test, we can see that all the p-values are above the significance level, and there is no significant lag in the ACF plot.
- There is no pattern left in the residuals because all the lags are insignificant.

6.0 Goodness-of-Fit Metrics

6.1 AIC and BIC Scores

	df <dbl>	AIC <dbl>
model5_103	6	3228.228
model5_102	5	3229.012
model5_405	11	3233.359
model5_003	5	3234.323
model5_002	4	3235.511
model5_206	10	3236.961

Figure 78: Akaike Information Critetion (AIC) Scores Table

	df <dbl>	BIC <dbl>
model5_102	5	3239.228
model5_103	6	3240.486
model5_002	4	3243.683
model5_003	5	3244.538
model5_405	11	3255.832
model5_206	10	3257.391

Figure 79: Bayesian Information Criterion (BIC) Scores Table

From Figures 79 and 79, we conclude that model5_102 and model5_103 are our best models as their AIC and BIC scores are the lowest. Both models have ar1, ma1, and sma1 which are significant coefficients as seen from our previous results. Model5_405 is ranked 3rd in the AIC table but not that great on the BIC table. Other potential models are ruled out as their AIC and BIC rankings vary too much from the 102 and 103 models.

	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>	MASE <dbl>	ACF1 <dbl>
SARIMA(4,0,5)x(0,1,1)_12	-19816987796	325521691926	218220104661	-0.574	3.426	0.517	0.010
SARIMA(1,0,3)x(0,1,1)_12	-24162931458	332104026182	219730779122	-0.658	3.441	0.520	0.018
SARIMA(1,0,2)x(0,1,1)_12	-43344732593	346163560678	231176721197	-0.992	3.645	0.547	-0.005
SARIMA(2,0,6)x(0,1,1)_12	-75719661858	363244467335	265128253043	-1.468	4.176	0.628	-0.018
SARIMA(0,0,3)x(0,1,1)_12	-136175390694	409755889221	282363162299	-2.425	4.520	0.669	-0.066
SARIMA(0,0,2)x(0,1,1)_12	-156446484391	423086910643	298335195607	-2.731	4.778	0.706	-0.042

Figure 80: Accuracy Measures Table

Looking at Figure 80, SARIMA(4,0,5)x(0,1,1) is at the top of the rankings with the best accuracy measures values. Looking back at the model, we see that SARIMA(4,0,5)x(0,1,1) has a lot of significant coefficients. Next, we see that SARIMA(1,0,3)x(0,1,1) is outperforming SARIMA(1,0,2)x(0,1,1), which justifies the reason we will be choosing model5_103 as it has the best performance in AIC, BIC, and accuracy measures table.

Next, we'll be fitting overparametrized models of SARIMA(1,0,3)x(0,1,1) model by adding 1 to p and q respectively, to check if there is a better model.

6.5 Over-parameterized model: SARIMA(2,0,3)x(0,1,1)_12

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)     ar1   0.0019547       NaN     NaN
NaN    ar2   0.9977588       NaN     NaN      NaN     ma1   0.2022079
0.1857427  1.0886 0.2763103     ma2  -0.8074680  0.2115348 -3.8172
0.0001350 *** ma3  -0.0098256  0.1614119 -0.0609 0.9514603     sma1 -
0.9631685  0.2854501 -3.3742 0.0007403 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                        Shapiro-Wilk normality test

data:  res.model
W = 0.93943, p-value = 0.0023
```

Figure 81: Parameter Estimation of SARIMA(2,0,3) x (0,1,1) model using ML method

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)     ar1   0.757770   0.363755  2.0832
0.03723 *   ar2   0.258556   0.367229  0.7041   0.48139     ma1  -0.644104
0.341004 -1.8888   0.05891 .   ma2   0.095299   0.283158  0.3366   0.73645
ma3  -0.297050   0.159857 -1.8582   0.06314 .   sma1 -0.746241   0.142892
-5.2224 1.766e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



                        Shapiro-Wilk normality test

data:  res.model
W = 0.93428, p-value = 0.001301
```

Figure 82: Parameter Estimation of SARIMA(2,0,3) x (0,1,1) model using CSS method

From Figure 81, we can see that there are NaNs produced for ar1 and ar2. This indicates that this model might be overfitting. The added parameter ar2 is insignificant in the CSS method from Figure 82. In addition, the ma1 and ma3 coefficients turned to be insignificant for both the CSS and ML method.

Overall, we can say that this over-parameterized model is overfitting, and it is not better than the SARIMA(1,0,3)x(0,1,1) model.

6.5 Over-parameterized model: SARIMA(1,0,4)x(0,1,1)_12

```
z test of coefficients:

     Estimate Std. Error   z value  Pr(>|z|)    ar1  0.999984  0.000173
5780.3494 < 2.2e-16 *** ma1  -0.849306  0.152628  -5.5646 2.628e-08 ***
ma2  0.135841  0.186790   0.7272  0.46708    ma3  -0.186240  0.229235  -
0.8124  0.41654    ma4  -0.078897  0.185208  -0.4260  0.67011    sma1 -
0.867531  0.421764  -2.0569  0.03969 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


                    Shapiro-Wilk normality test

data:  res.model
W = 0.94002, p-value = 0.002457
```

Figure 83: Parameter Estimation of SARIMA(1,0,4) x (0,1,1) model using ML method

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)    ar1  1.012096  0.020190
50.1291 < 2.2e-16 *** ma1  -0.903756  0.144575 -6.2511 4.076e-10 ***
ma2  0.276244  0.182904 1.5103   0.1310    ma3  -0.356934  0.281184 -
1.2694   0.2043    ma4  0.057312  0.260993  0.2196   0.8262    sma1 -
0.735483  0.158326 -4.6454 3.395e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


                    Shapiro-Wilk normality test

data:  res.model
W = 0.93241, p-value = 0.001062
```

Figure 84: Parameter Estimation of SARIMA(1,0,4) x (0,1,1) model using CSS method

From Figure 84 and Figure 84, the parameter estimation of CSS and ML are consistent, and we can see that added ma4 coefficient is insignificant. Additionally, the ma3 coefficient has turned to be insignificant. In conclusion, we can say that this over-parameterized model is overfitting, and it is not better than the SARIMA(1,0,3)x(0,1,1) model.

6.6 AIC and BIC Scores (Overparameterized model added)

	df <dbl>	AIC <dbl>
model5_103	6	3228.228
model5_102	5	3229.012
model5_104	7	3230.059
model5_203	7	3232.997
model5_405	11	3233.359
model5_003	5	3234.323
model5_002	4	3235.511
model5_206	10	3236.961

Figure 85: Akaike Information Criterion (AIC) Scores Table

	df <dbl>	BIC <dbl>
model5_102	5	3239.228
model5_103	6	3240.486
model5_002	4	3243.683
model5_104	7	3244.360
model5_003	5	3244.538
model5_203	7	3247.299
model5_405	11	3255.832
model5_206	10	3257.391

Figure 86: Bayesian Information Criterion (BIC) Scores Table

From Figure 85 and 86, we can see that the AIC and BIC scores our best SARIMA(1,0,3)x(0,1,1) model are still higher than the AIC and BIC scores of the overparameterized model. We will choose SARIMA(1,0,3)x(0,1,1) model by AIC and BIC.

In conclusion, we will still select SARIMA(1,0,3)x(0,1,1) as the best model because we do not find any overparameterized model with better performance.

7 Forecasting

In this section we will use the best model SARIMA(1,0,3)x(0,1,1) to forecast the electricity time series for the next 10 months.

	Point Forec... <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2020	3278151	3114198	3434286	3023810	3514132
Nov 2020	3168759	2996857	3331804	2901737	3414965
Dec 2020	3360444	3191012	3521733	3097572	3604194
Jan 2021	3516889	3355895	3670830	3267460	3749764
Feb 2021	3203131	3025456	3371455	2927038	3457245
Mar 2021	3420554	3254741	3578692	3163449	3659641
Apr 2021	3171860	2992268	3341815	2892687	3428376
May 2021	3568653	3409995	3720551	3322942	3798503
Jun 2021	3787434	3638304	3930911	3556829	4004783
Jul 2021	3926826	3783170	4065409	3704869	4136891

Figure 87: Forecast table of electricity time series using SARIMA(1,0,3)x(0,1,1) model (10 months)

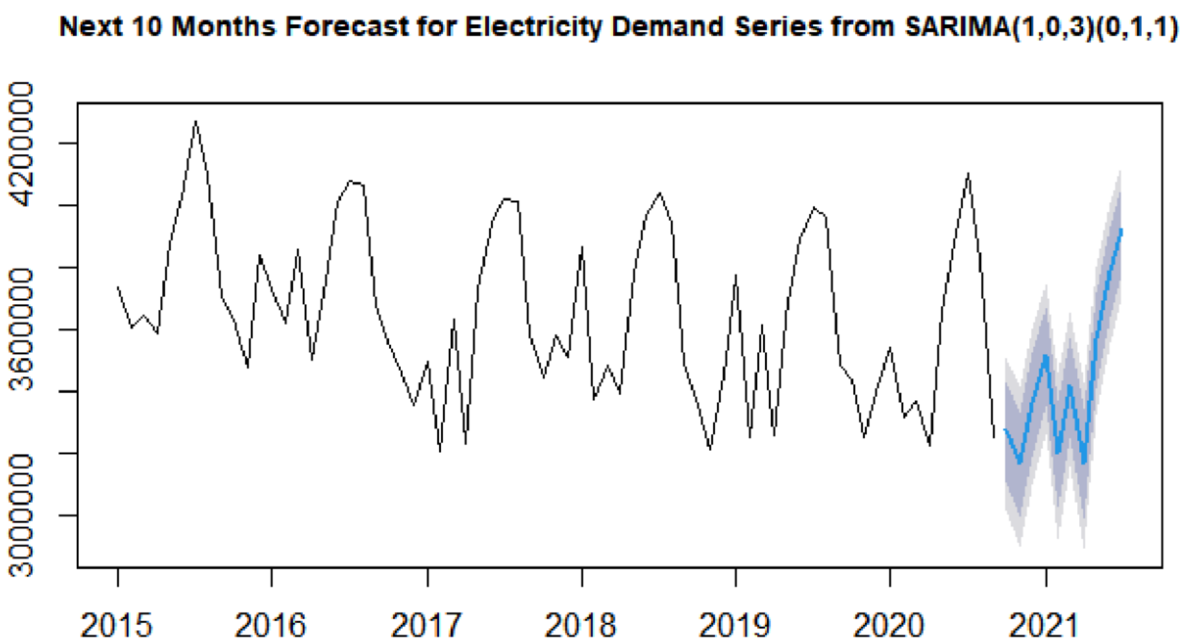

Figure 88: Forecast Plot of electricity time series using SARIMA(1,0,3)x(0,1,1) model (10 months)

Figure 87 above shows the next 10 months point forecast values, with the low and high 80% and 95% confidence intervals. Figure 88 above shows the forecast plot based on the forecasted values. From Figure 88 the trend for next 10 months seems to be increasing and confirmation with the actual results may be needed to confirm that this model is accurate.

8 Conclusion

In summary, our chosen model is SARIMA(1,0,3)x(0,1,1), as it has 3 significant coefficients. Even though SARIMA(4,0,5)x(0,1,1) has more significant coefficients, the performance as seen from AIC and BIC tables are not enough to justify picking it, and it might be prone to overfitting as well. Furthermore, in this report we have achieved our goal set out in the introduction to find the best model for electricity demand series. We have first visualized the data and found that the series is nonstationary and not normally distributed. Then, we performed Box Cox transformation, by using the SARIMA approach we have found that a seasonal differencing of 1 has addressed the issue with the nonstationary and normality. By using the EACF and BIC table, we then selected the parameters of the models and performed residual analysis with the model. However, limitations of this report were that if there were more data from previous years it would be beneficial in aiding to find the best models for analysis.

9 References

Kozlov A (2020) Daily Electricity Price and Demand Data [data set], Kaggle website, accessed 23 May 2024. https://www.kaggle.com/datasets/aramacus/electricity-demand-in-victoriaaustralia?resource=download

10.0 Appendix: R Code

````
---
author: "Kai Ging Yong, Wong Yi Wei (Ethan), Yung Qi Chin"
date: "`r Sys.Date()`"
output: html_document
---

# **Daily Electricity Demand in Victoria (2015 – 2022) Analysis**

## Setup

```{r} #Libraries
knitr::opts_chunk$set(echo = TRUE)
library(dplyr) library(TSA)
library(fUnitRoots)
library(lmtest)
library(tseries)
library(forecast)
library(LSTS)
```

```{r}
#Code Setup
#Function 1: Multiple plots with test function

plots<-function(plotdata, plotarg){

  #Time Series Plot   plot(plotdata,
       ylab="Electricity Demand(MWh)",
xlab="Year",        type='l',
main="",        cex.main=1)
  points(plotdata,x=time(plotdata), pch=as.vector(season(plotdata)))
grid(nx=NULL, ny=NULL, lty=2, col="gray", lwd=1)   mtext(paste("Time
Series Plot for",plotarg),
        side=3,line=-2.8,outer=TRUE,cex=1)

  #Q-Q Plot
qqnorm(plotdata,
main="",
cex.main=1)
qqline(plotdata,
col="red",        lwd=1)
grid(nx=NULL,
ny=NULL, lty=2,
col="gray", lwd=1)
mtext(paste("Normal
Q-Q Plot
for",plotarg),
````
````

```
        side=3,line=-2.8,outer=TRUE,cex=1)

  #ACF & PACF Plot
par(mfrow=c(1,2))
acf(plotdata,main="")
pacf(plotdata,main="")
  mtext(paste("ACF & PACF Plot for",plotarg),                side=3,line=-
2.8,outer=TRUE,cex=1)

  #Stationary & Normality Test
cat(paste(plotarg),"Tests","\n")
print(adf.test(plotdata))
print(pp.test(plotdata))
  print(shapiro.test(plotdata))

}

#Function 2: Residual Plot
res.plots<-function(plotdata, plotarg, lagmax){

  #Time Series Plot
  plot(plotdata,xlab='Time',ylab='Residuals',main="",type="l")
grid(nx=NULL, ny=NULL, lty=2, col="gray", lwd=1)
points(plotdata,x=time(plotdata), pch=as.vector(season(plotdata)))
mtext(paste("Time Series Plot for",plotarg),
        side=3,line=-2.8,outer=TRUE,cex=1)

  #ACF & PACF Plot   par(mfrow=c(1,2))
acf(plotdata,main="",lag.max=lagmax)
pacf(plotdata,main="",lag.max=lagmax)
mtext(paste("ACF & PACF Plot for",plotarg),
        side=3,line=-2.8,outer=TRUE,cex=1)

}

#Function 3: Residual Analysis
residual.analysis <- function(model, std = TRUE,start = 2, class =
c("ARIMA","GARCH","ARMA-GARCH", "fGARCH")[1]){

  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))
  Box.Ljung.Test(res.model, lag = 30)
}

#Function 4: AIC & BIC Score sort function
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")')   } }
```

---

## 1.0 Import Data and Preprocessing

```{r}
#Import data
electricity<-read.csv("complete_dataset.csv")

#Check data class class(electricity)

#Check data structure str(electricity)

#Random 5 observations
electricity[sample(nrow(electricity),5), ]
```

```{r}
#Create year & month columns

````
electricity$date <- as.character(electricity$date) electricity$year
<- as.integer(substr(electricity$date, 1, 4))
electricity$month <- as.integer(substr(electricity$date, 6, 7))

#Generate new dataframe electricity <-
electricity %>%   group_by(year,
month) %>%   summarize(total_demand =
sum(demand))
electricity <- electricity[-nrow(electricity),]

#Random 5 observations
electricity[sample(nrow(electricity),5),]

#Summary of dataframe
summary(electricity)
```

```{r}
#Convert to time series object
electricity.ts = ts(electricity$total_demand,start=c(2015,1),frequency=12)
```

2.0 Data Visualization and Descriptive Analysis

```{r}
plots(electricity.ts,"Electricity Demand Series")
```

```{r} #Scatter plot
plot(y=electricity.ts,
    x=zlag(electricity.ts),      xlab="Electricity Demand
(MWH)",      ylab="Previous Month Electricty
Demand (MWh)",      main="Scatter Plot for
Electricity Demand Series")

abline(lm(electricity.ts~zlag(electricity.ts), data=electricity.ts),col="red")

grid(nx=NULL, ny=NULL, lty=2, col="gray", lwd=1)
```

```{r}
#Correlation y=electricity.ts
x=zlag(electricity.ts)
index=2:length(x)
````

cat("Correlation:",cor(y[index],x[index])) ```

3.0 Data Preparation & Modelling

3.1 Box Cox Transformation

```{r}
#Check for negative values
cat("Sum of Negative Values:", sum(electricity.ts<0),"\n\n")
cat("Data Summary:","\n") summary(electricity.ts)
```

```{r}
#Box Cox transformation
electricity.bc <- BoxCox.ar(electricity.ts)

mtext(paste("Figure 5: Box-Cox Plot for Electricity Demand Series"),
side=3,
     line=-2.8,
outer=TRUE,       cex=1)
```

```{r}
#Lambda & CI values
lambda <- electricity.bc$lambda[which(max(electricity.bc$loglike) == electricity.bc$loglike)]
cat("CI:",electricity.bc$ci,"\n")
cat("Lambda:",lambda)

BC.electricity = ((electricity.ts^lambda)-1)/lambda
```

```{r}
#BC Time Series plot
plots(BC.electricity, "Box Cox Transformed Electricity Demand Series")
```

3.2 SARIMA Approach

```{r} #(0,
0, 0)
model1     =     Arima(electricity.ts,order=c(0,0,0),seasonal=list(order=c(0,1,0),     period=12))
res.model1 = residuals(model1)
res.plots(res.model1,"Model 1 Residuals",48)
```

```{r} #(0, 1,
1) BC
model2 = Arima(electricity.ts,order=c(0,0,0),seasonal=list(order=c(0,1,1), period=12))
res.model2 = residuals(model2)
res.plots(res.model2, "Model 2 Residuals", 36)
```

```{r} #(0, 1,
1) TS
model3 = Arima(BC.electricity,order=c(0,0,0),seasonal=list(order=c(0,1,1), period=12))
res.model3 = residuals(model3)
res.plots(res.model3, "Model 3 Residuals", 36)
```

```{r} #(5,0,4)
BC
model4 = Arima(BC.electricity,order=c(5,0,4),seasonal=list(order=c(0,1,1), period=12))
res.model4 = residuals(model4)
res.plots(res.model4, "Model 4 Residuals" , 36)
```

4.0 Model Specification

4.1 Extended Correlation Function (EACF)

```{r}
#EACF
cat("EACF Table","\n")
eacf(res.model3)
```

4.2 Subset ARMA Model

```{r}
#BIC
par(mfrow=c(1,1))
bic_table = armasubsets(y=res.model3,nar=10,nma=10,y.name='p',ar.method='ols')
plot(bic_table)
mtext("BIC Table",side=3,line=-1.5,outer=TRUE,cex=1)
```

5.0 Parameter Estimation

5.1 SARIMA (4,0,5)

```{r}
#ML
model5_405 = Arima(BC.electricity,order=c(4,0,5),seasonal=list(order=c(0,1,1),
period=12),method = "ML")

coeftest(model5_405)
residual.analysis(model = model5_405)
```

```{r}
#CSS
model5_405_CSS = Arima(BC.electricity,order=c(4,0,5),seasonal=list(order=c(0,1,1),
period=12),method = "CSS")

coeftest(model5_405_CSS)

residual.analysis(model = model5_405_CSS)
```

```{r}
#CSS-ML
model5_405_CSSML = Arima(BC.electricity,order=c(4,0,5),seasonal=list(order=c(0,1,1),
period=12),method = "CSS-ML")

coeftest(model5_405_CSSML)

residual.analysis(model = model5_405_CSSML)
```

5.2 SARIMA (0,0,2)

```{r}
#ML
model5_002 = Arima(BC.electricity,order=c(0,0,2),seasonal=list(order=c(0,1,1),
period=12),method = "ML")

coeftest(model5_002)

residual.analysis(model = model5_002)
```

```{r}
#CSS
model5_002_CSS = Arima(BC.electricity,order=c(0,0,2),seasonal=list(order=c(0,1,1),
period=12),method = "CSS")

coeftest(model5_002_CSS)
```

residual.analysis(model = model5_002_CSS)
```

```{r}
#CSS-ML
model5_002_CSSML = Arima(BC.electricity,order=c(0,0,2),seasonal=list(order=c(0,1,1),
period=12),method = "CSS-ML")

coeftest(model5_002_CSSML)
residual.analysis(model = model5_002_CSSML)
```

5.3 SARIMA (0,0,3)

```{r}
#ML
model5_003 = Arima(BC.electricity,order=c(0,0,3),seasonal=list(order=c(0,1,1),
period=12),method = "ML")

coeftest(model5_003)

residual.analysis(model = model5_003)
```

```{r}
#CSS
model5_003_CSS = Arima(BC.electricity,order=c(0,0,3),seasonal=list(order=c(0,1,1),
period=12),method = "CSS")

coeftest(model5_003_CSS)

residual.analysis(model = model5_003_CSS)
```

```{r}
#CSS-ML
model5_003_CSSML = Arima(BC.electricity,order=c(0,0,3),seasonal=list(order=c(0,1,1),
period=12),method = "CSS-ML")

coeftest(model5_003_CSSML)

residual.analysis(model = model5_003_CSSML)
```

5.4 SARIMA (1,0,2)

```{r}

#ML
model5_102 = Arima(BC.electricity,order=c(1,0,2),seasonal=list(order=c(0,1,1), period=12),method = "ML")

coeftest(model5_102)

residual.analysis(model = model5_102)
```

```{r}
#CSS
model5_102_CSS = Arima(BC.electricity,order=c(1,0,2),seasonal=list(order=c(0,1,1), period=12),method = "CSS")
coeftest(model5_102_CSS)

residual.analysis(model = model5_102_CSS)
```

```{r}
#CSS-ML
model5_102_CSSML = Arima(BC.electricity,order=c(1,0,2),seasonal=list(order=c(0,1,1), period=12),method = "CSS-ML")

coeftest(model5_102_CSSML)

residual.analysis(model = model5_102_CSSML)
```

5.5 SARIMA (1,0,3)

```{r}
#ML
model5_103 = Arima(BC.electricity,order=c(1,0,3),seasonal=list(order=c(0,1,1), period=12),method = "ML")

coeftest(model5_103)

residual.analysis(model = model5_103)
```

```{r}
#CSS
model5_103_CSS = Arima(BC.electricity,order=c(1,0,3),seasonal=list(order=c(0,1,1), period=12),method = "CSS")

coeftest(model5_103_CSS)

residual.analysis(model = model5_103_CSS)

```

5.6 SARIMA (8,0,6)

```{r}
#ML
model5_806 = Arima(BC.electricity,order=c(8,0,6),seasonal=list(order=c(0,1,1),
period=12),method = "ML")

coeftest(model5_806)

residual.analysis(model = model5_806)
```

```{r}
#CSS
model5_806_CSS = Arima(BC.electricity,order=c(8,0,6),seasonal=list(order=c(0,1,1),
period=12),method = "CSS")

coeftest(model5_806_CSS)

residual.analysis(model = model5_806_CSS)
```

```{r}
#CSS-ML
model5_806_CSSML = Arima(BC.electricity,order=c(8,0,6),seasonal=list(order=c(0,1,1),
period=12),method = "CSS-ML")

coeftest(model5_806_CSSML)

residual.analysis(model = model5_806_CSSML)
```

5.7 SARIMA (2,0,6)

```{r}
#ML
model5_206 = Arima(BC.electricity,order=c(2,0,6),seasonal=list(order=c(0,1,1),
period=12),method = "ML")

coeftest(model5_206)

residual.analysis(model = model5_206)
```

```{r}
#CSS

model5_206_CSS = Arima(BC.electricity,order=c(2,0,6),seasonal=list(order=c(0,1,1), period=12),method = "CSS")

coeftest(model5_206_CSS)

residual.analysis(model = model5_206_CSS)
```

```{r}
#CSS-ML
model5_206_CSSML = Arima(BC.electricity,order=c(2,0,6),seasonal=list(order=c(0,1,1), period=12),method = "CSS-ML")

coeftest(model5_206_CSSML)

residual.analysis(model = model5_206_CSSML)
```

5.8 SARIMA (10,0,6)
```{r}
#ML
model5_1006 = Arima(BC.electricity,order=c(10,0,6),seasonal=list(order=c(0,1,1), period=12),method = "ML")

coeftest(model5_1006)

residual.analysis(model = model5_1006)
```

```{r}
#CSS
model5_1006_CSS = Arima(BC.electricity,order=c(10,0,6),seasonal=list(order=c(0,1,1), period=12),method = "CSS")

coeftest(model5_1006_CSS)

residual.analysis(model = model5_1006_CSS)
```

```{r}
#CSS-ML
model5_1006_CSSML = Arima(BC.electricity,order=c(10,0,6),seasonal=list(order=c(0,1,1), period=12),method = "CSS-ML")

coeftest(model5_1006_CSSML)

residual.analysis(model = model5_1006_CSSML)
```
```

6.0 Goodness-of-Fit Metrics

6.1 Akaike Information Criterion (AIC) Scores

```{r}
sort.score(AIC(model5_405,
model5_002,                model5_003,
model5_102,                model5_103,
           model5_206), score="aic")
```

6.2 Bayesian Information Criterion (BIC) Scores

```{r}
sort.score(BIC(model5_405,
model5_002,                model5_003,
model5_102,                model5_103,
           model5_206),score="bic")
```

6.4 Accuracy Measures

```{r}
model5_504_a<-accuracy(model5_405)[1:7] model5_002_a<-accuracy(model5_002)[1:7]
model5_003_a<-accuracy(model5_003)[1:7] model5_102_a<-accuracy(model5_102)[1:7]
model5_103_a<-accuracy(model5_103)[1:7] model5_206_a<-accuracy(model5_206)[1:7]

models_a<-data.frame(rbind(model5_504_a,
                   model5_002_a,
model5_003_a,
model5_102_a,
model5_103_a,
                   model5_206_a))

colnames(models_a) <- c("ME", "RMSE", "MAE", "MPE", "MAPE", "MASE", "ACF1")

rownames(models_a) <- c("SARIMA(4,0,5)x(0,1,1)_12",
                 "SARIMA(0,0,2)x(0,1,1)_12",
                 "SARIMA(0,0,3)x(0,1,1)_12",
                 "SARIMA(1,0,2)x(0,1,1)_12",
                 "SARIMA(1,0,3)x(0,1,1)_12",
                 "SARIMA(2,0,6)x(0,1,1)_12")

models_a<-round(models_a, digits = 3)
```

```
models_a %>% arrange(MASE)
```

6.5 Over-parameterized model: SARIMA(2,0,3)x(0,1,1)_12

```{r}
#ML
model5_203 = Arima(BC.electricity,order=c(2,0,3),seasonal=list(order=c(0,1,1), period=12),method = "ML")

coeftest(model5_203)

residual.analysis(model = model5_203)
```

```{r}
#CSS
model5_203_CSS = Arima(BC.electricity,order=c(2,0,3),seasonal=list(order=c(0,1,1), period=12),method = "CSS")

coeftest(model5_203_CSS)

residual.analysis(model = model5_203_CSS)
```

6.5 Over-parameterized model: SARIMA(1,0,4)x(0,1,1)_12

```{r}
#ML
model5_104 = Arima(BC.electricity,order=c(1,0,4),seasonal=list(order=c(0,1,1), period=12),method = "ML")

coeftest(model5_104)

residual.analysis(model = model5_104)
```

```{r}
#CSS
model5_104_CSS = Arima(BC.electricity,order=c(1,0,4),seasonal=list(order=c(0,1,1), period=12),method = "CSS")

coeftest(model5_104_CSS)

residual.analysis(model = model5_104_CSS)
```

6.6 AIC and BIC Scores (With Overparameterized Models)

```{r}
sort.score(AIC(model5_405,
model5_002,              model5_003,
model5_102,              model5_103,
model5_206,              model5_203,
             model5_104), score="aic")
```

```{r}
sort.score(BIC(model5_405,
model5_002,              model5_003,
model5_102,              model5_103,
model5_206,              model5_203,
model5_104), score="bic")
```

7.0 Forecasting

```{r}
model5_103.electricity     =     Arima(electricity.ts,order=c(1,0,3),seasonal=list(order=c(0,1,1),
period=12), lambda = 2, method = "ML")

model5_103_frc = forecast(model5_103.electricity, lambda = 2, h = 10)
model5_103_frc plot(model5_103_frc,
    main="Next 10 Month’s Forecast for Electricity Demand Series from
SARIMA(1,0,3)(0,1,1)", cex.main=0.9)
```