MATH1318 Semester 1, 2020 - Egg depositions of Lake Huron Bloaters (Coregonus hoyi)

Assignment 2

Millie Woo (s3806940)

Contents

1. Introduction 5

2. Summary Statistics 5

3. Data Analysis on the original time series data 5

4. Data Transformation 8

5. Differencing on the Time series 11

6. Find additional p and q from EACF and BIC table 13

7. Model Estimation 15

  a. Maximum Likelihood 15

  b. Least Square Estimation 17

8. Perform Residual Analysis and find the best model 20

9. Five years predictions 23

10. Conclusion 24

References 25

Appendix 25

  [A1] – Import packages and data preparation 25

  [A2] – Generate descriptive statistics of the time series 26

  [A3] – Produce relevant plots and perform normality tests 26

  [A4] – Box-cox and log transformation 27

  [A5] - Differencing 28

  [A6] – Model Estimation 30

  [A7] – Residual Analysis 31

  [A8] - Forecasting 34

  [A9] – Null Hypothesis of various statistical test 34

    1. Shapiro Normality Test 34

    2. Augmented Dickey-Fuller Test 34

3. Linear Regression Test on Model coefficient 34
4. Ljung-Box Test 35
[A10] – Residual plots for all the possible ARIMA models 35

Table of Figures

Figure 1 - Time series plot of Egg depositions for original data 6
Figure 2 - Scatter plot of Original Egg Depositions with it's first time lag 6
Figure 3- Quantile-Quantile plot of Original Egg Depositions 7
Figure 4 - ACF plot for Original Egg Depositions 7
Figure 5 - PACF plot for Original Egg Depositions 7
Figure 6 - Log Likelihood vs Lambda (Maximum Likelihood Estimation) 8
Figure 7- Log Likelihood vs Lambda (Method of Moments Estimation) 8
Figure 8- Time series plot of Egg depositions after Box Cox Transformation 8
Figure 9- Scatter plot of Egg depositions with its first time lag after Box Cox Transformation 8
Figure 10- Quantile-Quantile plot of Egg Depositions after BoxCox Transformation 9
Figure 11- Time series plot of Egg depositions after Log Transformation 10
Figure 12- Scatter plot of Egg depositions with its first time lag after Log Transformation 10
Figure 13- Quantile-Quantile plot of Egg Depositions after Log Transformation 10
Figure 14 - Time series plot of Egg Depositions after First Differencing 12
Figure 15- Time series plot of Egg Depositions after Second Differencing 12
Figure 16- ACF plot of Egg Deposition after First Differencing 13
Figure 17- ACF plot of Egg Deposition after Second Differencing 13
Figure 18- PACF plot of Egg Deposition after First Differencing 13
Figure 19- PACF plot of Egg Deposition after Second Differencing 13
Figure 20- Time Series Plot of Standard Residuals of ARIMA (0,1,0) 21
Figure 21- Histogram of Standard Residuals of ARIMA (0,1,0) 21
Figure 22- ACF Plot of Standard Residuals of ARIMA (0,1,0) 21
Figure 23- PACF Plot of Standard Residuals of ARIMA (0,1,0) 21
Figure 24 - Q-Q Plot of Standard Residuals of ARIMA (0,1,0) 21
Figure 25 - Ljung Box Plot of Standard Residuals of ARIMA (0,1,0) 21
Figure 26- Forecasts of Egg Depositions using ARIMA(0,1,0) for year 1997 to 2001 24
Figure 27- Time Series Plot of Standard Residuals of ARIMA (0,1,1) 35
Figure 28 - Histogram of Standard Residuals of ARIMA (0,1,1) 36
Figure 29- ACF Plot of Standard Residuals of ARIMA (0,1,1) 36
Figure 30- PACF Plot of Standard Residuals of ARIMA (0,1,1) 37
Figure 31- Q-Q Plot of Standard Residuals of ARIMA (0,1,1) 37
Figure 32- Ljung Box Plot of Standard Residuals of ARIMA (0,1,1) 38
Figure 33- Time Series Plot of Standard Residuals of ARIMA (0,1,2) 39
Figure 34 - Histogram of Standard Residuals of ARIMA (0,1,2) 39
Figure 35- ACF Plot of Standard Residuals of ARIMA (0,1,2) 40

Figure 36 - PACF Plot of Standard Residuals of ARIMA (0,1,2) 40
Figure 37- Q-Q Plot of Standard Residuals of ARIMA (0,1,2) 41
Figure 38- Ljung Box Plot of Standard Residuals of ARIMA (0,1,2) 41
Figure 39 - Time Series Plot of Standard Residuals of ARIMA (1,1,0) 42
Figure 40 - Histogram of Standard Residuals of ARIMA (1,1,0) 43
Figure 41 - ACF Plot of Standard Residuals of ARIMA (1,1,0) 43
Figure 42- PACF Plot of Standard Residuals of ARIMA (1,1,0) 44
Figure 43- Q-Q Plot of Standard Residuals of ARIMA (1,1,0) 44
Figure 44 - Ljung Box Plot of Standard Residuals of ARIMA (1,1,0) 45
Figure 45 - Time Series Plot of Standard Residuals of ARIMA (1,1,1) 46
Figure 46 - Histogram of Standard Residuals of ARIMA (1,1,1) 46
Figure 47- ACF Plot of Standard Residuals of ARIMA (1,1,1) 47
Figure 48 - PACF Plot of Standard Residuals of ARIMA (1,1,1) 47
Figure 49- Q-Q Plot of Standard Residuals of ARIMA (1,1,1) 48
Figure 50 - Ljung Box Plot of Standard Residuals of ARIMA (1,1,1) 48
Figure 51-Time Series Plot of Standard Residuals of ARIMA (1,1,2) 49
Figure 52- Histogram of Standard Residuals of ARIMA (1,1,2) 50
Figure 53- ACF Plot of Standard Residuals of ARIMA (1,1,2) 50
Figure 54- PACF Plot of Standard Residuals of ARIMA (1,1,2) 51
Figure 55 - Q-Q Plot of Standard Residuals of ARIMA (1,1,2) 51
Figure 56 - Ljung Box Plot of Standard Residuals of ARIMA (1,1,2) 52
Figure 57- Time Series Plot of Standard Residuals of ARIMA (2,1,0) 53
Figure 58 - Histogram of Standard Residuals of ARIMA (2,1,0) 53
Figure 59 - ACF Plot of Standard Residuals of ARIMA (2,1,0) 54
Figure 60 - PACF Plot of Standard Residuals of ARIMA (2,1,0) 54
Figure 61- Q-Q Plot of Standard Residuals of ARIMA (2,1,0) 55
Figure 62 - Ljung Box Plot of Standard Residuals of ARIMA (2,1,0) 55
Figure 63- Time Series Plot of Standard Residuals of ARIMA (2,1,1) 56
Figure 64 - Histogram of Standard Residuals of ARIMA (2,1,1) 57
Figure 65 - ACF Plot of Standard Residuals of ARIMA (2,1,1) 57
Figure 66 - PACF Plot of Standard Residuals of ARIMA (2,1,1) 58
Figure 67- Q-Q Plot of Standard Residuals of ARIMA (2,1,1) 58
Figure 68 - Ljung Box Plot of Standard Residuals of ARIMA (2,1,1) 59
Figure 69- Time Series Plot of Standard Residuals of ARIMA (2,1,2) 60
Figure 70 - Histogram of Standard Residuals of ARIMA (2,1,2) 60
Figure 71- ACF Plot of Standard Residuals of ARIMA (2,1,2) 61
Figure 72- PACF Plot of Standard Residuals of ARIMA (2,1,2) 61
Figure 73-Q-Q Plot of Standard Residuals of ARIMA (2,1,2) 62
Figure 74 - Ljung Box Plot of Standard Residuals of ARIMA (2,1,2) 62

Table of Tables

Table 1- Summary statistics of Egg depositions between 1981 and 1996 5
Table 2 - ADF Test result for first differencing 12
Table 3- ADF Test result for second differencing 12

Table 4- ADF test result for third differencing 12
Table 5 - ADF test result for fourth differencing 12
Table 6 - EACF Table 14
Table 7- BIC Table 14
Table 8 - Coefficient Test result on ARIMA(0,1,1) with Maximum Likelihood Estimation 15
Table 9 - Coefficient Test result on ARIMA(0,1,2) with Maximum Likelihood Estimation 15
Table 10 - Coefficient Test result on ARIMA(1,1,0) with Maximum Likelihood Estimation 16
Table 11- Coefficient Test result on ARIMA(1,1,1) with Maximum Likelihood Estimation 16
Table 12- Coefficient Test result on ARIMA(1,1,2) with Maximum Likelihood Estimation 16
Table 13 - Coefficient Test result on ARIMA(2,1,0) with Maximum Likelihood Estimation 16
Table 14 - Coefficient Test result on ARIMA(2,1,1) with Maximum Likelihood Estimation 17
Table 15 - Coefficient Test result on ARIMA(2,1,2) with Maximum Likelihood Estimation 17
Table 16 - Coefficient Test result on ARIMA(0,1,1) with Least Square Estimation 17
Table 17 - Coefficient Test result on ARIMA(0,1,2) with Least Square Estimation 18
Table 18 - Coefficient Test result on ARIMA(1,1,0) with Least Square Estimation 18
Table 19- Coefficient Test result on ARIMA(1,1,1) with Least Square Estimation 18
Table 20- Coefficient Test result on ARIMA(1,1,2) with Least Square Estimation 18
Table 21- Coefficient Test result on ARIMA(2,1,0) with Least Square Estimation 19
Table 22- Coefficient Test result on ARIMA(2,1,1) with Least Square Estimation 19
Table 23- Coefficient Test result on ARIMA(2,1,2) with Least Square Estimation 19
Table 24 - Shapiro Normality and Ljung test for ARIMA(0,1,0) 22
Table 25 - AIC score for all models 22
Table 26- BIC score for all models 22
Table 27- Confidence Intervals for forecasting Egg Depositions (using ARIMA(0,1,0) for year 1997 to 2001 23
Table 28 - Shapiro Normality and Ljung test for ARIMA(0,1,1) 38
Table 29- Shapiro Normality and Ljung test for ARIMA(0,1,2) 42
Table 30- Shapiro Normality and Ljung test for ARIMA(1,1,0) 45
Table 31-Shapiro Normality and Ljung test for ARIMA(1,1,1) 49
Table 32- Shapiro Normality and Ljung test for ARIMA(1,1,2) 52
Table 33 - Shapiro Normality and Ljung test for ARIMA(2,1,0) 56
Table 34- Shapiro Normality and Ljung test for ARIMA(2,1,1) 59
Table 35- Shapiro Normality and Ljung test for ARIMA(2,1,2) 63

1. Introduction

Coregonus hoyi, also known as lake Huron bloater, is a species or form of freshwater whitefish in the family Salmonidae. We have given a data set of its egg depositions between 1981 and 1996. We need to analyze this dataset and choose a best fit model to forecast the egg depositions for the next 5 years.

2. Summary Statistics

Refer to Appendix [A1] and [A2] for coding.

From Table 1, it shows the mean of egg deposition is 0.7869 million, with a standard of deviation of 0.597 million, highest point at 2.0984 and lowest point at 0.0402 million. The median (0.743 million) is close to the mean, with Q1 at 0.3062 and Q3 at 1.0893 million. There are only 16 observations in the dataset.

	Mean	Median	Std.Dev	Q1	Q3	Min	Max	N.Valid
eggs	0.7869	0.743	0.5973049	0.3062	1.08935	0.0402	2.0984	16

Table 1- Summary statistics of Egg depositions between 1981 and 1996

3. Data Analysis on the original time series data

Refer to Appendix [A3] for coding.

The time series plot of Egg Depositions in Figure 1 shows that there is an increasing trend with two major fluctuations (at 1985 and 1990 respectively). There is no repeating pattern in the plot, which means no sign of seasonality. We can also see that there is an obvious jump in 1989 and 1990 (from around 0.5 million to 2.0 million), but drop back to a steady trend at 1993.

From Figure 2, the scatter plot shows that the egg deposition is positively correlated with its previous time lag. The correlation index is found to be 0.7445, which is considerably high. (1 indicates 100% positive correlation, and -1 indicates 100% negative correlation).

From Figure 3, the Quantile and Quantile plot shows that most of the data lie on the straight line which passes thru the first and third quantiles. There are only 3 obvious outliners which are further above the line, but in general most of the data complies with normal distribution assumption. The Shapiro normality test [A9-1] also fail to reject the null hypothesis, which further proves normality.

From Figure 4, the acf plot shows a slow decay in autocorrelation, which indicates non-stationarity in the time series, we need to de-trend it in order to further identify the model.

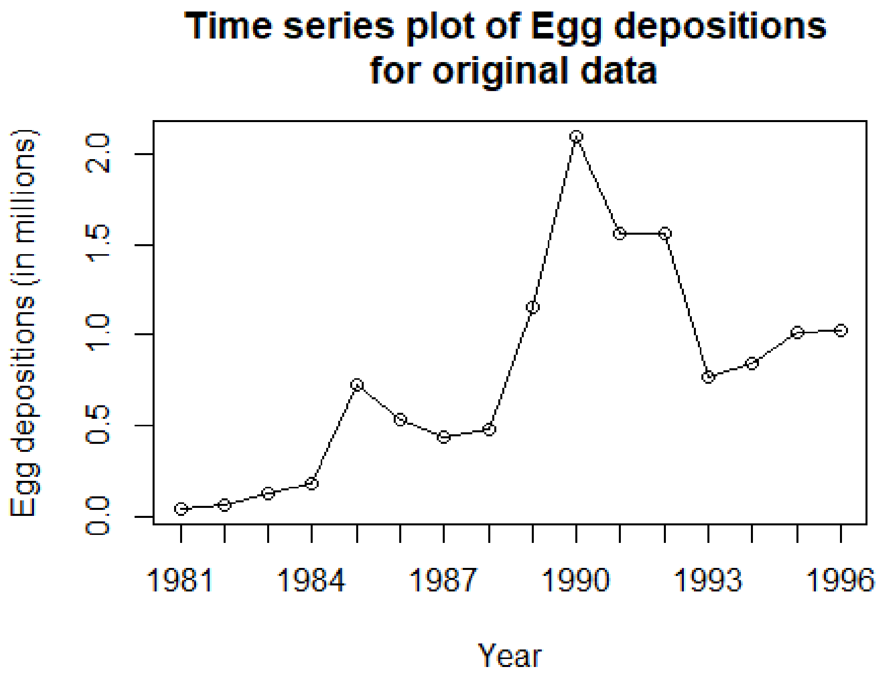


Figure 1 - Time series plot of Egg depositions for original data

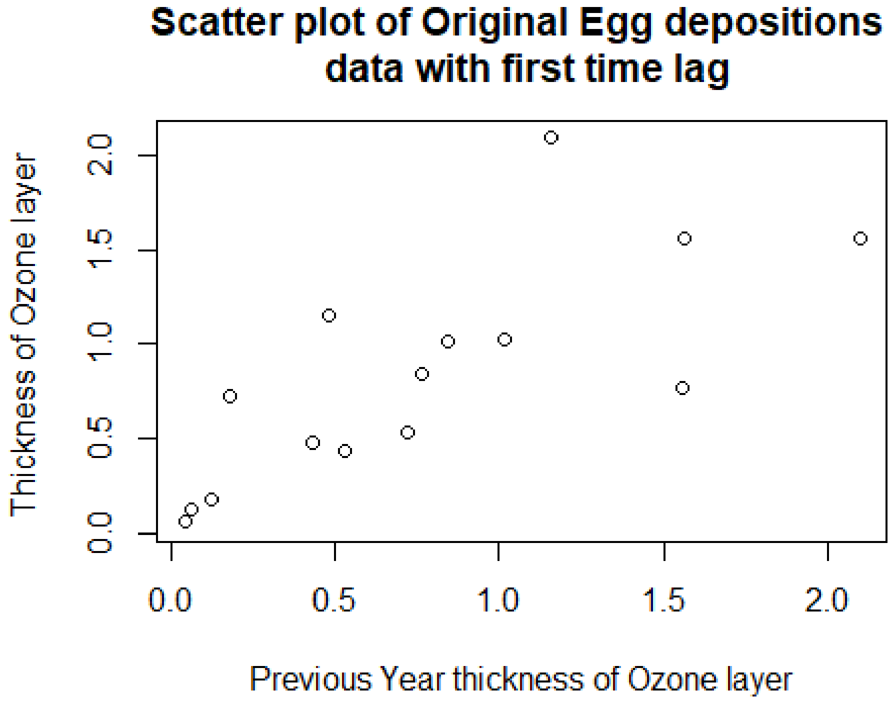


Figure 2 - Scatter plot of Original Egg Depositions with it's first time lag

```
## [1] "Correlation Index:"
## [1] 0.7445657
```

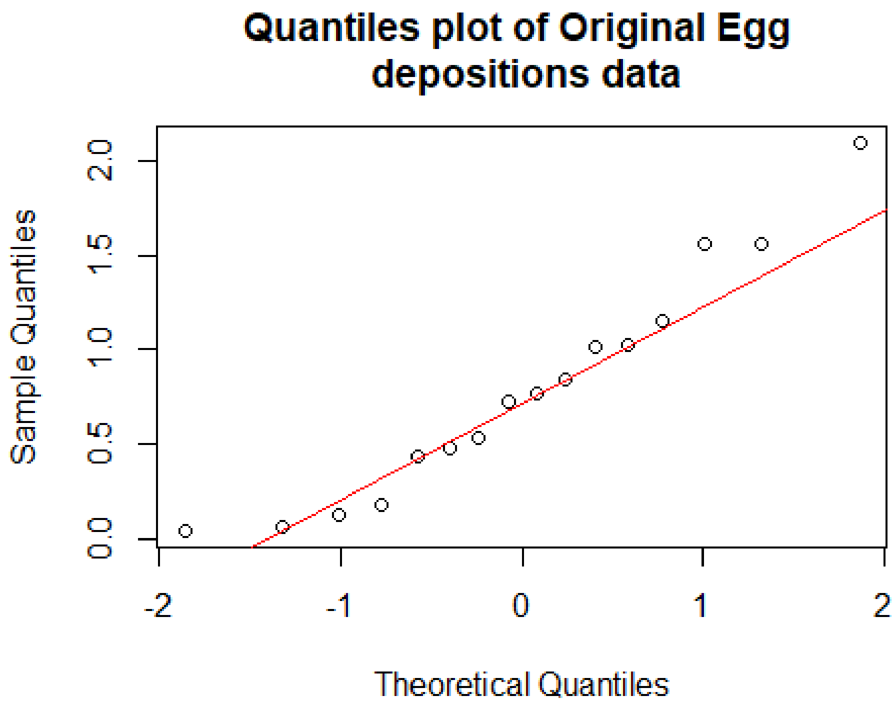


Figure 3- Quantile-Quantile plot of Original Egg Depositions

```
##
##  Shapiro-Wilk normality test
##
## data:  ts_data
## W = 0.94201, p-value = 0.3744
```

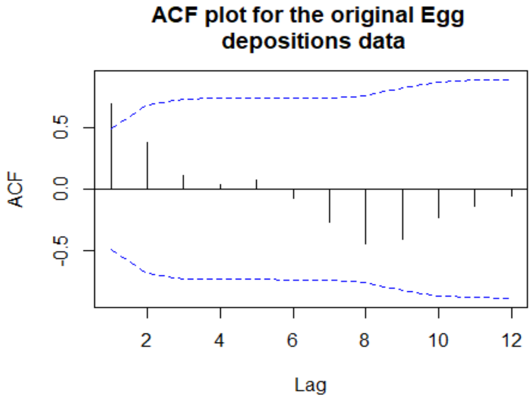


Figure 4 - ACF plot for Original Egg Depositions

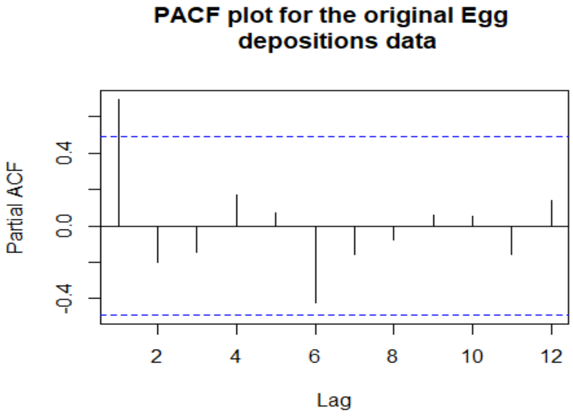


Figure 5 - PACF plot for Original Egg Depositions

4. Data Transformation

Refer to Appendix [A4] for coding.

Let's see whether data transformation could smooth out the fluctuation points, and further improve normality. By plotting the log likelihood with lambda, we could find the lambda for transformation. Confidence Interval is [0.7, 1.0] from maximum likelihood estimation. From method of moment estimation, confidence interval is found to be [0.1, 0.8]. Thus, we chose to use 0.75 as our lambda value for the transformation.

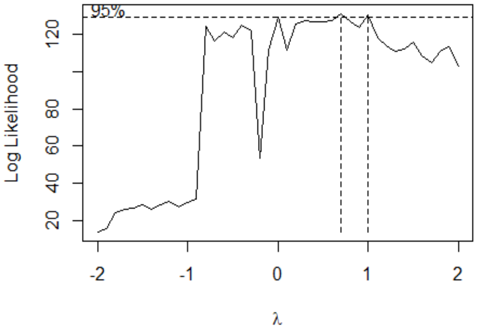


Figure 6 - Log Likelihood vs Lambda (Maximum Likelihood Estimation)

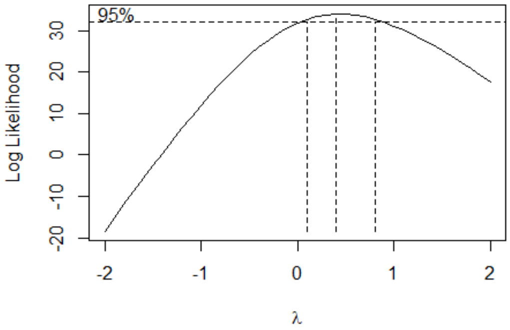


Figure 7- Log Likelihood vs Lambda (Method of Moments Estimation)

By inspecting Figure 8 and Figure 9, we found that after BoxCox transformation, there is no change in the pattern of the time series, as well as the correlation with its first time lag. It shows a bit of improvement in normality, as all the data points are closer to the straight

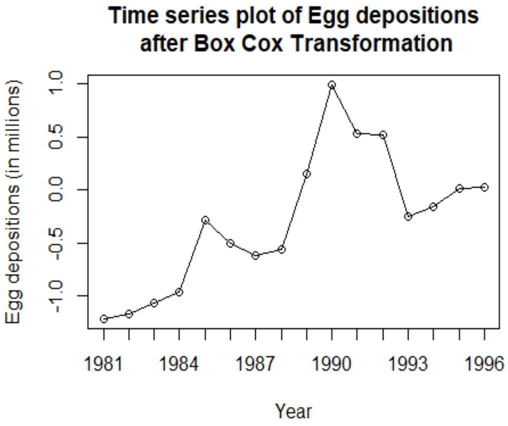


Figure 8- Time series plot of Egg depositions after Box Cox Transformation

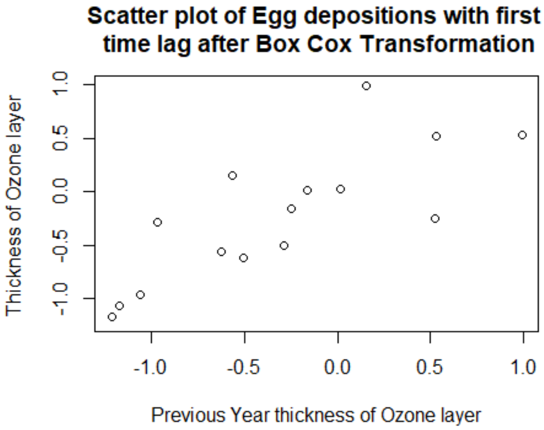


Figure 9- Scatter plot of Egg depositions with its first time lag after Box Cox Transformation

line in *Figure 10*. However, as the price of data transformation is high, and we would need to revert the transformation for forecasting. I would choose not to proceed with Box-Cox Transformation.

```
## [1] "Correlation Index:"
## [1] 0.7862601
```

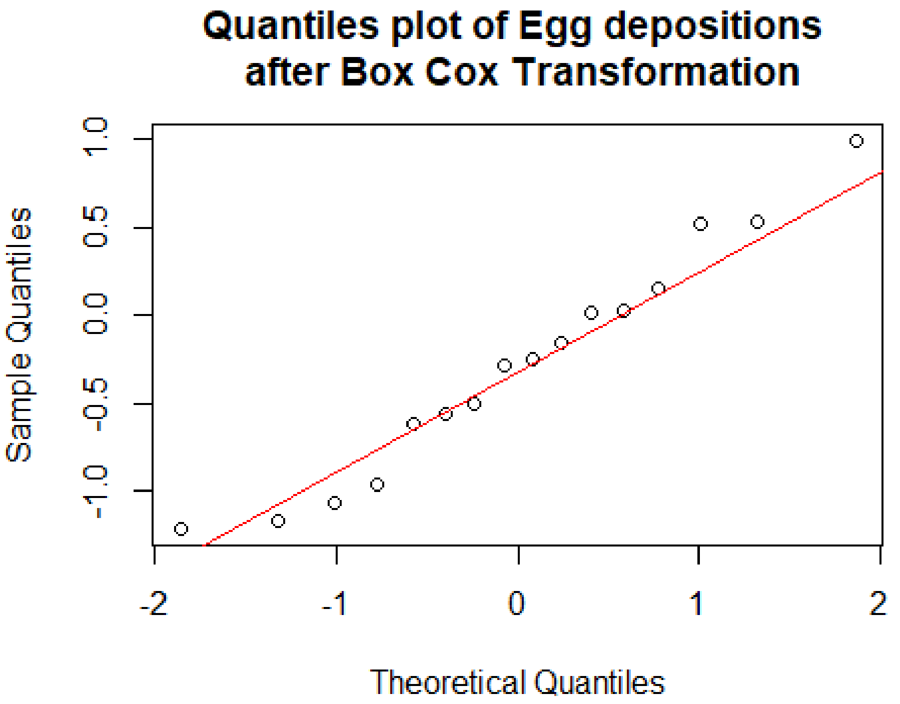


Figure 10- Quantile-Quantile plot of Egg Depositions after BoxCox Transformation

```
## 
##  Shapiro-Wilk normality test
## 
## data:  ts_data
## W = 0.96452, p-value = 0.744
```

I then attempted log transformation. From *Figure 11*, it shows that the fluctuation point in 1990 is not varying as much as in the original data. However, the data becomes more correlated with its first time lag (refer to *Figure 12*) and it is less normally distributed (refer to *Figure 13*) as compare to the original data. Thus I would not proceed with any kind of data transformation and just perform further modelling with the original dataset.

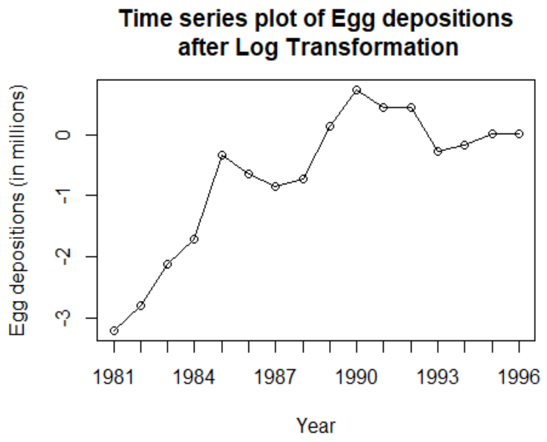


Figure 11- Time series plot of Egg depositions after Log Transformation

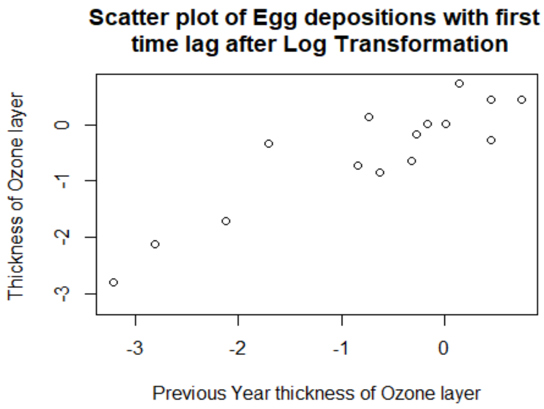


Figure 12- Scatter plot of Egg depositions with its first time lag after Log Transformation

```
## [1] "Correlation Index:"
## [1] 0.9005508
```

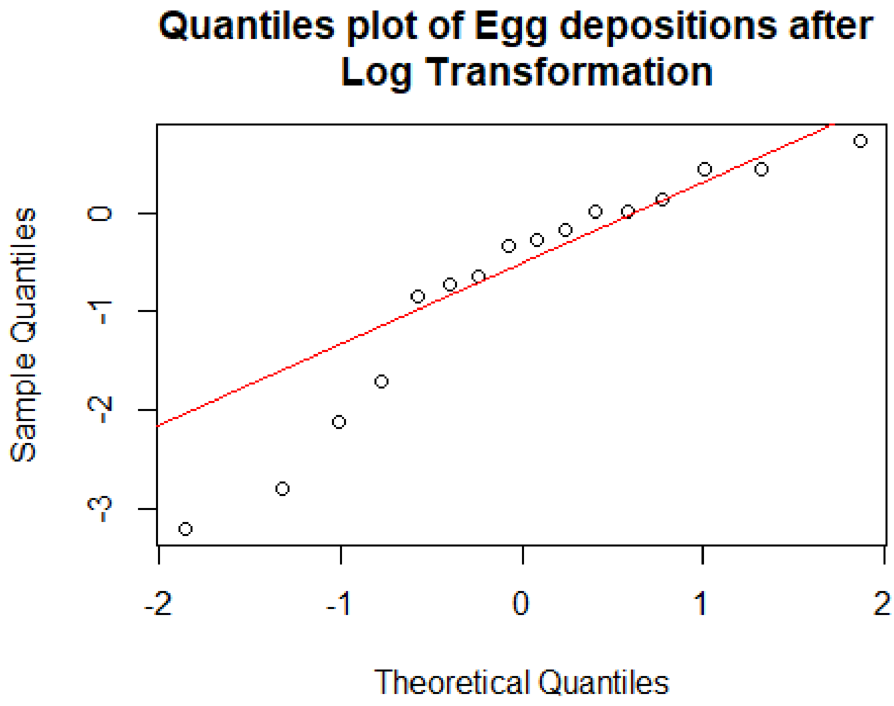


Figure 13- Quantile-Quantile plot of Egg Depositions after Log Transformation

```
##
##  Shapiro-Wilk normality test
##
## data:  ts_data
## W = 0.88762, p-value = 0.05108
```

5. Differencing on the Time series

Refer to Appendix [A5] for coding.

As there is no particular reason to believe the trend in the egg deposition would be exactly the same for another period of time, we would only model this as a stochastic trend.

To confirm there is a trend (or non-stationarity) in the data, we need to perform Augmented Dickey–Fuller (ADF) test [A9-2]. ADF test is sensitive to the parameter "lag". If we set the value of "lag" too high, the power of hypothesis test decreases (i.e. the probability of making the correct decision if the alternative hypothesis <which is the time series is stationary >is true will decrease). If we set it too small, we will be looking at the series at an early lag and the effect on the remaining ones will not be considered for stationarity check.

One of the methods to choose the lag order, is to find out the lag with lowest AIC in the first differenced data. When we apply this method to the original egg dataset, we find lag=0 with the lowest AIC value, we then apply ADF test with lag=0 on the original data set, p-value is 0.452 (refers to Appendix [A5]), which fail to reject null hypothesis at an early lag at 0, the original data is confirmed to be non-stationary.

We then apply first differencing to the data.

To check whether the data is stationary after first differencing, first we need to find the lag with lowest AIC in the second differenced data, lag=4 is returned. Since we only have 16 observations in the dataset, testing lag=4 on the first differenced data would turn out to test only 11 observations(there are only 15 observations left in the data after first differencing), there might not be enough samples, and we might be picking a lag that is too high, and the power of the test will be decreased. To justify this, we have applied the ADF test on the data after first, second, third and fourth differencing. We found out p-value is higher than 0.05 for all the differenced data at lag=4. (refer to Table 2 to Table 5). Also, time series plot on second differencing (refer to Figure 15) is not any better than first differencing (refer to Figure 14) in terms of stationarity (first differencing seems already stabilized the upward trend and fluctuation points). All the lags in ACF plot of first differencing are within the confidence interval (refer to Figure 16), and there is no obvious change compare to the ACF plot of second differencing (refer to Figure 17). Same observation also applies to PACF plots (refer to Figure 18 and Figure 19). Following the principle of parsimony, if applying first differencing is sufficient to remove the non-stationarity, we should avoid over-differencing and keep d = 1.

lag	statistics	p value
1	-2.437	0.018
2	-2.697	0.010
3	-2.161	0.033
4	-0.781	0.360

Table 2 - ADF Test result for first differencing

lag	statistics	p value
1	-2.978	0.010
2	-3.048	0.010
3	-3.921	0.010
4	-1.397	0.164

Table 3- ADF Test result for second differencing

lag	statistics	p value
1	-3.615	0.010
2	-2.529	0.015
3	-4.395	0.010
4	-0.728	0.377

Table 4- ADF test result for third differencing

lag	statistics	p value
1	-4.469	0.010
2	-2.152	0.034
3	-5.214	0.010
4	-0.370	0.490

Table 5 - ADF test result for fourth differencing

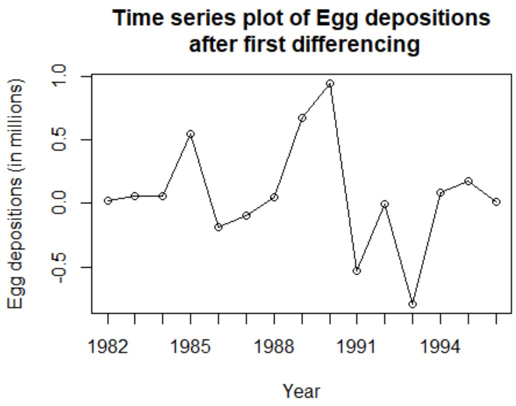


Figure 14 - Time series plot of Egg Depositions after First Differencing

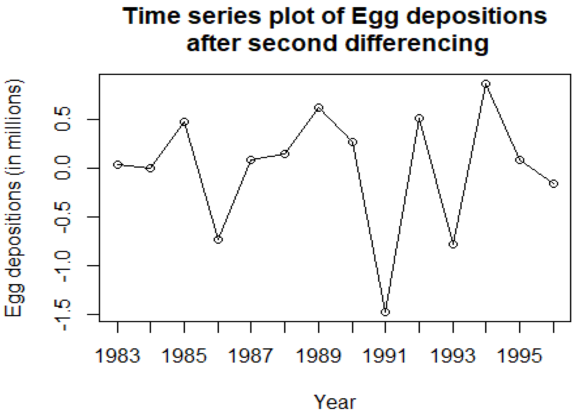


Figure 15- Time series plot of Egg Depositions after Second Differencing

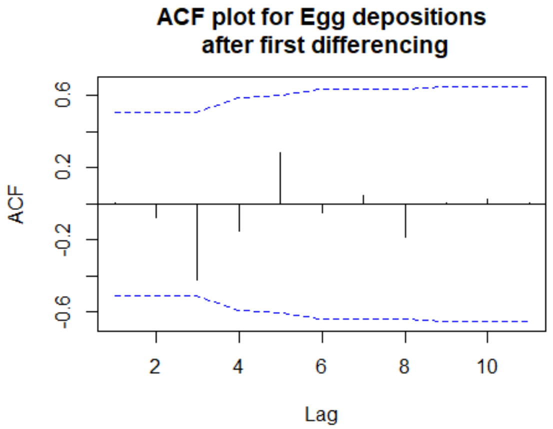


Figure 16- ACF plot of Egg Deposition after First Differencing

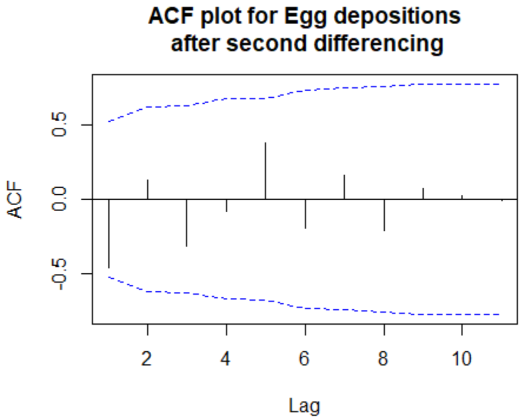


Figure 17- ACF plot of Egg Deposition after Second Differencing

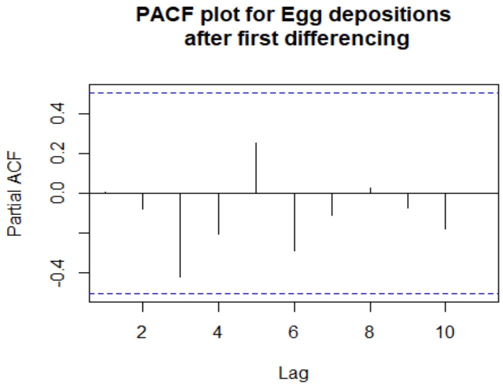


Figure 18- PACF plot of Egg Deposition after First Differencing

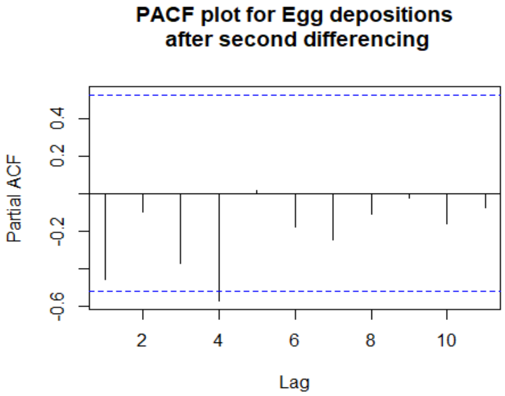
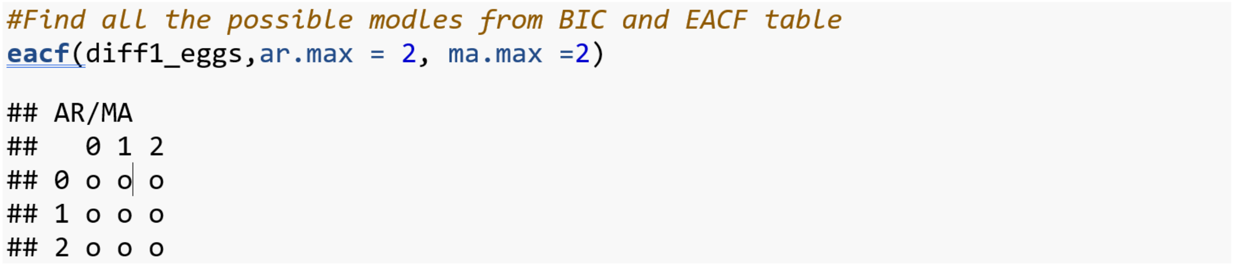


Figure 19- PACF plot of Egg Deposition after Second Differencing

6. Find additional p and q from EACF and BIC table

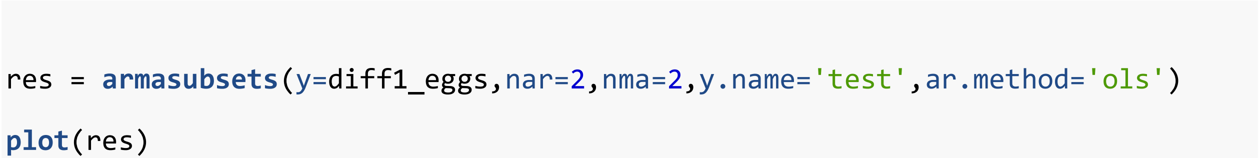
Referring to the EACF table (Table 6), we notice the vertex of the triangle of zero starts at p=0 and q = 0, we can include all the models with p=0, q=0, p=1, q=1, p=2 and q = 2.



```
#Find all the possible modles from BIC and EACF table
eacf(diff1_eggs,ar.max = 2, ma.max =2)

## AR/MA
##   0 1 2
## 0 o o o
## 1 o o o
## 2 o o o
```

Table 6 - EACF Table



```
res = armasubsets(y=diff1_eggs,nar=2,nma=2,y.name='test',ar.method='ols')

plot(res)
```

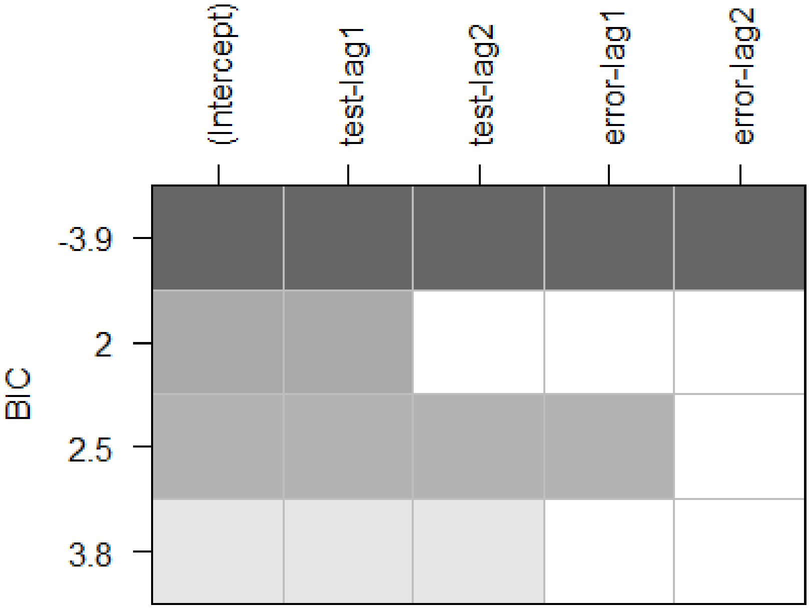


Table 7- BIC Table

From BIC table (refer to Table 7), by limiting p and q to maximum of 2, it tells us that the possible models coincide with the outcomes from EACF. As first differencing (d=1) is already enough to remove the non-stationarity in the time series data, the set of candidate models are {ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(0,1,2), ARIMA(1,1,0), ARIMA(1, 1,1), ARIMA(1,1,2), ARIMA(2,1,0), ARIMA(2,1,1), ARIMA(2,1,2)}

7. Model Estimation

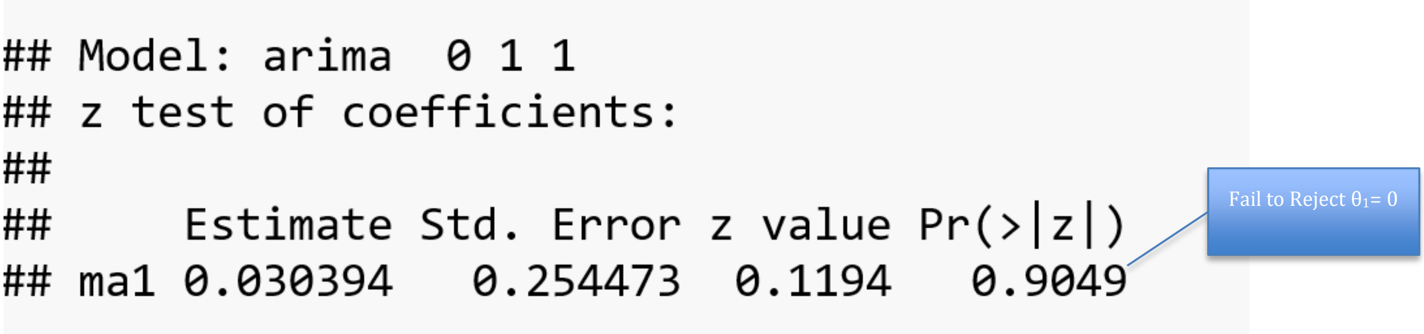
Refer to Appendix [A6] for coding.

I then proceed to model estimation to test whether (φ and θ) are coefficient significant on p and q of each model respectively. As there is no p and q in ARIMA(0,1,0), this is the only candidate model we don't need to perform the coefficient test. [A9-3].

Refer to Table 8 to Table 15, from maximum likelihood estimation, we found that none of the model is **all** coefficients significant, which means all the models fail to reject the hypothesis with **all** coefficient=0.

Least square estimation also turn out with the same result. (refer to Table 16 to Table 23).

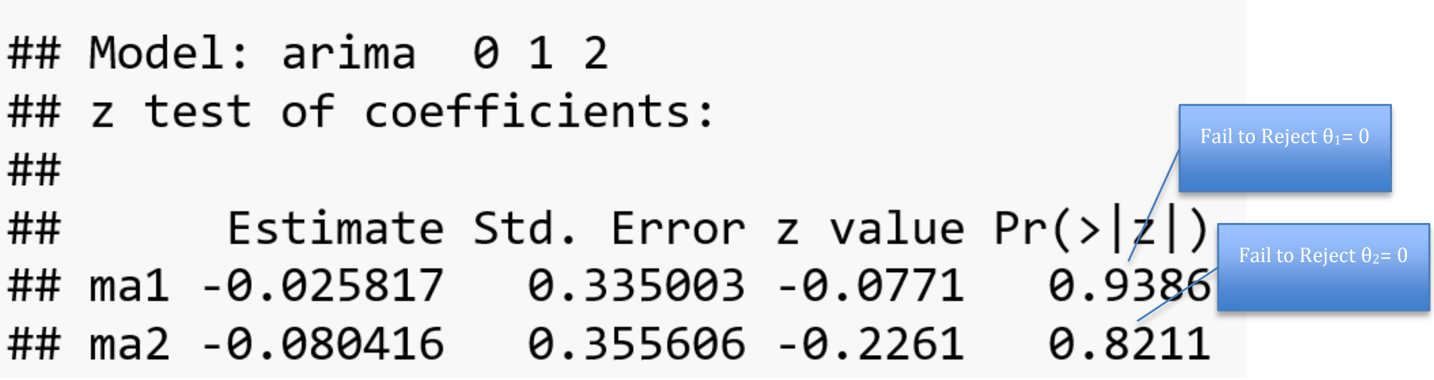
a. Maximum Likelihood



```
## Model: arima  0 1 1
## z test of coefficients:
##
##     Estimate Std. Error z value Pr(>|z|)
## ma1 0.030394   0.254473  0.1194   0.9049
```

Fail to Reject $\theta_1 = 0$

Table 8 - Coefficient Test result on ARIMA(0,1,1) with Maximum Likelihood Estimation

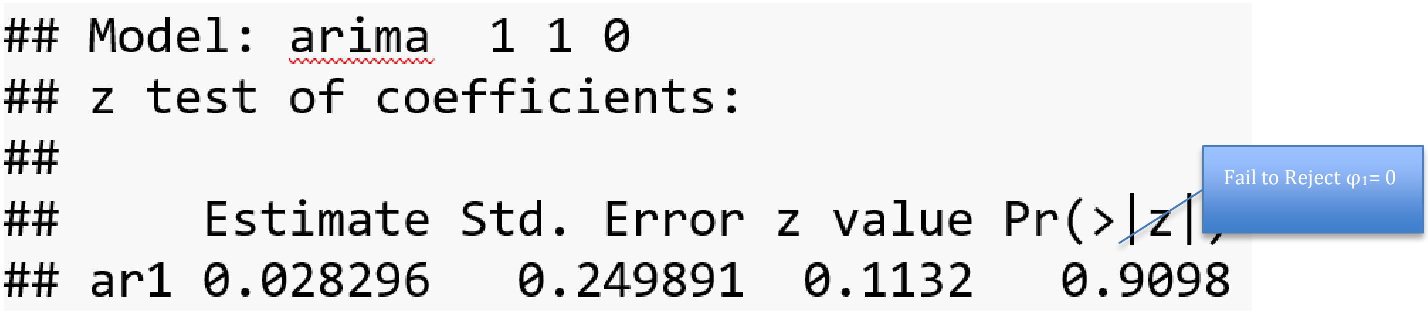


```
## Model: arima  0 1 2
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1 -0.025817   0.335003 -0.0771   0.9386
## ma2 -0.080416   0.355606 -0.2261   0.8211
```

Fail to Reject $\theta_1 = 0$

Fail to Reject $\theta_2 = 0$

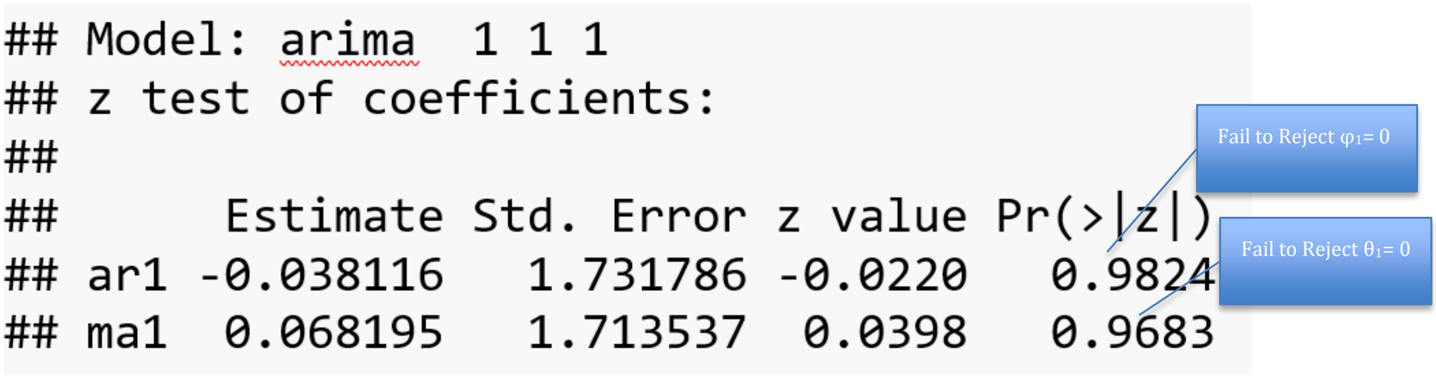
Table 9 - Coefficient Test result on ARIMA(0,1,2) with Maximum Likelihood Estimation



```
## Model: arima  1 1 0
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1 0.028296   0.249891  0.1132   0.9098
```

Fail to Reject $\varphi_1 = 0$

Table 10 - Coefficient Test result on ARIMA(1,1,0) with Maximum Likelihood Estimation

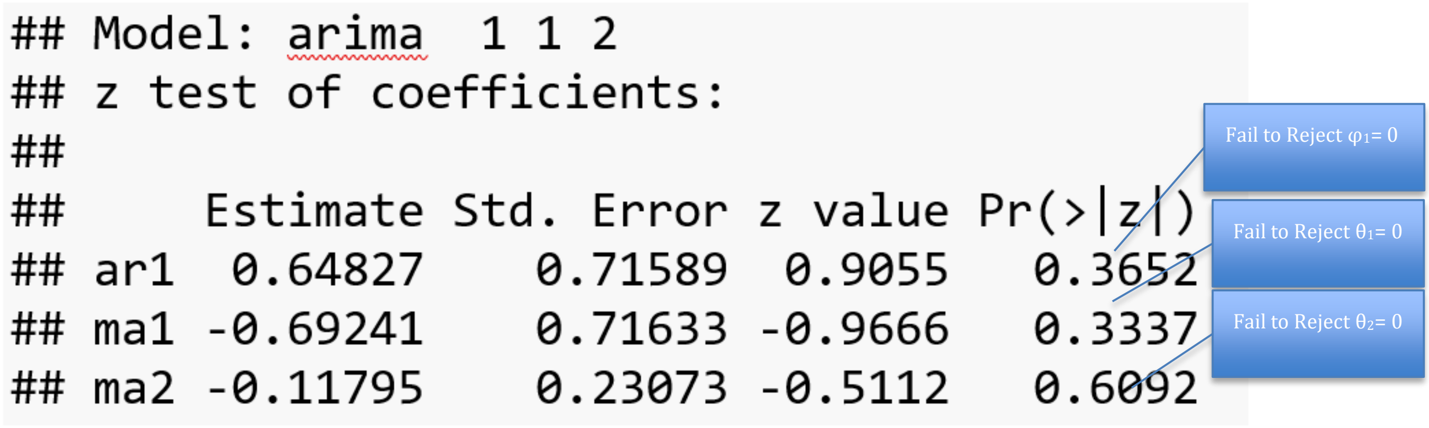


```
## Model: arima  1 1 1
## z test of coefficients:
##
##       Estimate Std. Error z value Pr(>|z|)
## ar1 -0.038116   1.731786 -0.0220   0.9824
## ma1  0.068195   1.713537  0.0398   0.9683
```

Fail to Reject $\varphi_1 = 0$

Fail to Reject $\theta_1 = 0$

Table 11- Coefficient Test result on ARIMA(1,1,1) with Maximum Likelihood Estimation



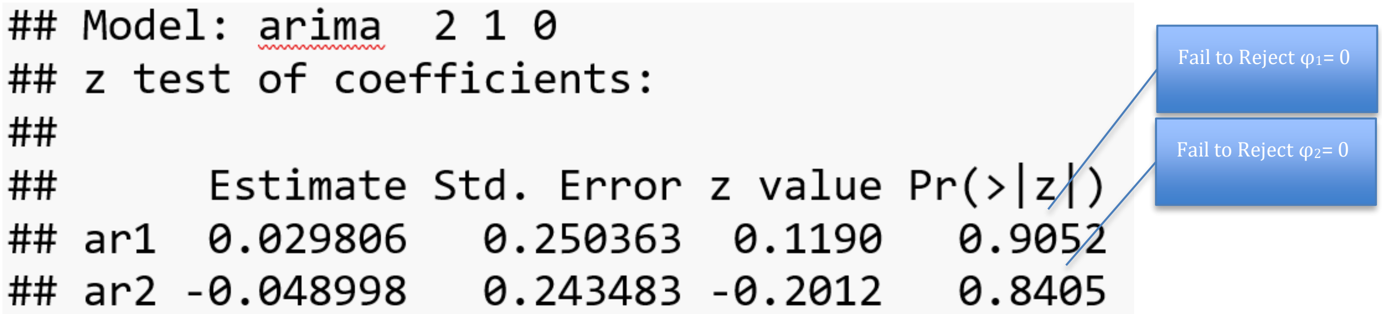
```
## Model: arima  1 1 2
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.64827    0.71589  0.9055   0.3652
## ma1 -0.69241    0.71633 -0.9666   0.3337
## ma2 -0.11795    0.23073 -0.5112   0.6092
```

Fail to Reject $\varphi_1 = 0$

Fail to Reject $\theta_1 = 0$

Fail to Reject $\theta_2 = 0$

Table 12- Coefficient Test result on ARIMA(1,1,2) with Maximum Likelihood Estimation

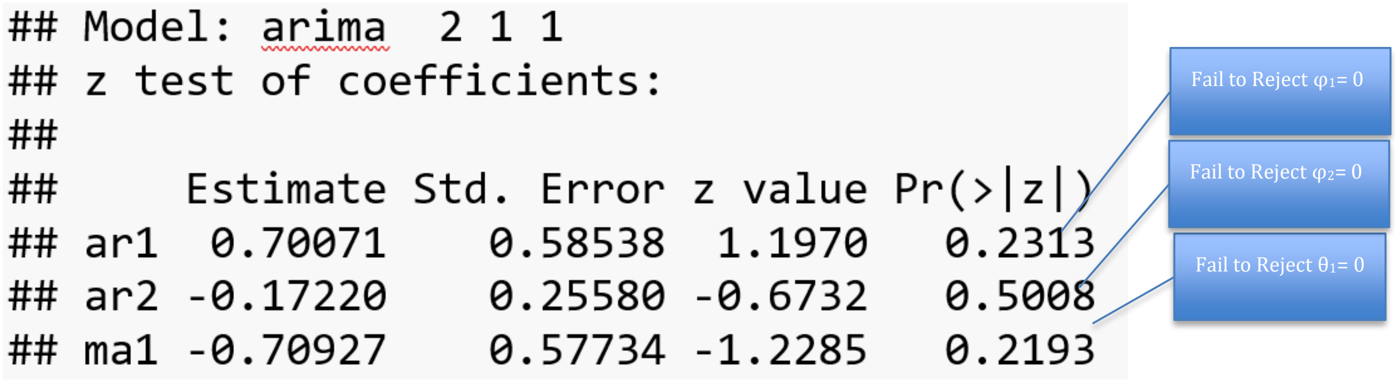


```
## Model: arima  2 1 0
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.029806   0.250363  0.1190   0.9052
## ar2 -0.048998   0.243483 -0.2012   0.8405
```

Fail to Reject $\varphi_1 = 0$

Fail to Reject $\varphi_2 = 0$

Table 13 - Coefficient Test result on ARIMA(2,1,0) with Maximum Likelihood Estimation



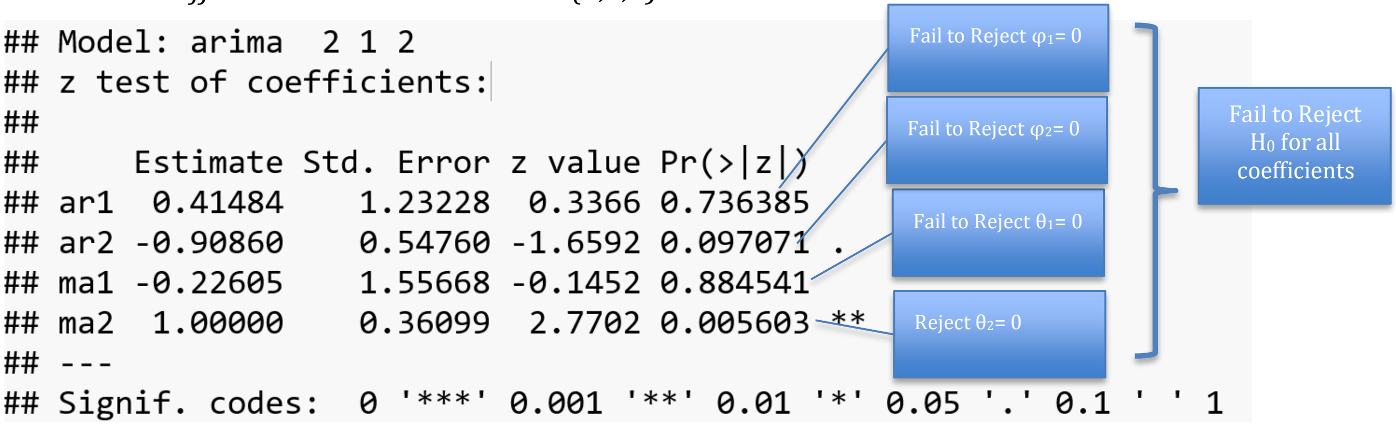
```
## Model: arima  2 1 1
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.70071    0.58538  1.1970   0.2313
## ar2 -0.17220    0.25580 -0.6732   0.5008
## ma1 -0.70927    0.57734 -1.2285   0.2193
```

Fail to Reject $\varphi_1 = 0$

Fail to Reject $\varphi_2 = 0$

Fail to Reject $\theta_1 = 0$

Table 14 - Coefficient Test result on ARIMA(2,1,1) with Maximum Likelihood Estimation



```
## Model: arima  2 1 2
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.41484    1.23228  0.3366 0.736385
## ar2 -0.90860    0.54760 -1.6592 0.097071 .
## ma1 -0.22605    1.55668 -0.1452 0.884541
## ma2  1.00000    0.36099  2.7702 0.005603 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fail to Reject $\varphi_1 = 0$

Fail to Reject $\varphi_2 = 0$

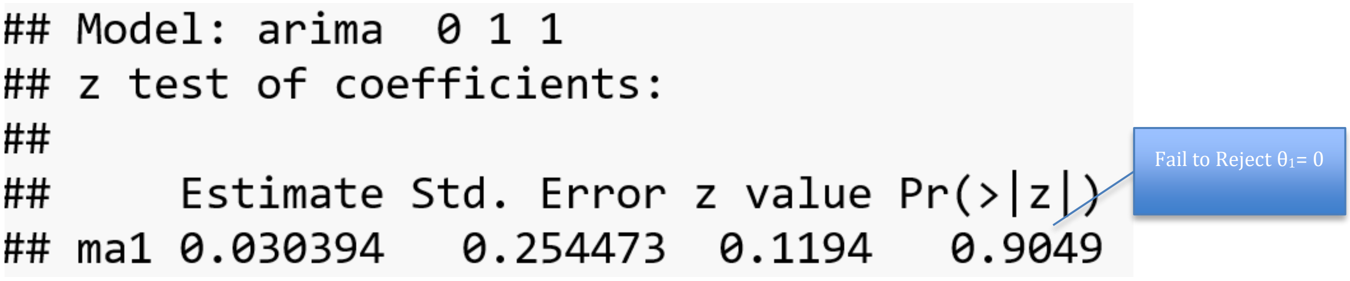
Fail to Reject $\theta_1 = 0$

Reject $\theta_2 = 0$

Fail to Reject H_0 for all coefficients

Table 15 - Coefficient Test result on ARIMA(2,1,2) with Maximum Likelihood Estimation

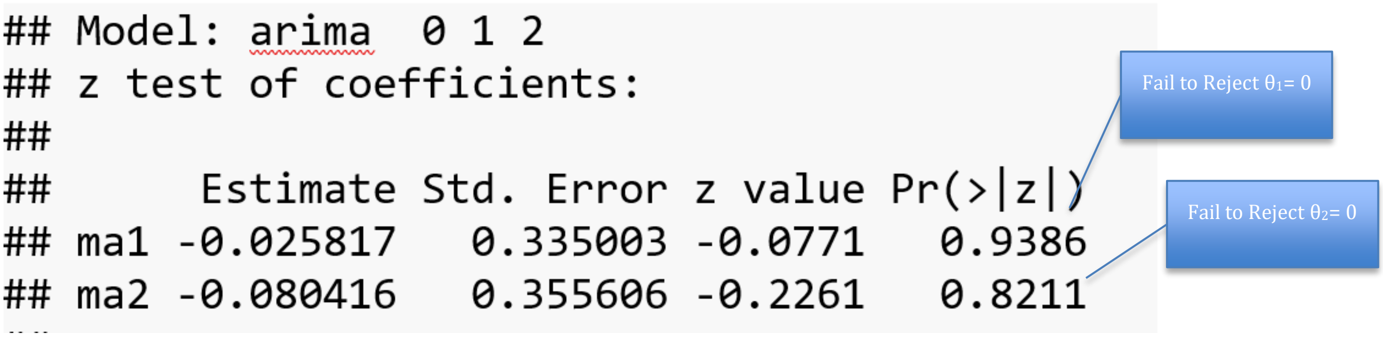
b. Least Square Estimation



```
## Model: arima  0 1 1
## z test of coefficients:
##
##     Estimate Std. Error z value Pr(>|z|)
## ma1 0.030394   0.254473  0.1194   0.9049
```

Fail to Reject $\theta_1 = 0$

Table 16 - Coefficient Test result on ARIMA(0,1,1) with Least Square Estimation

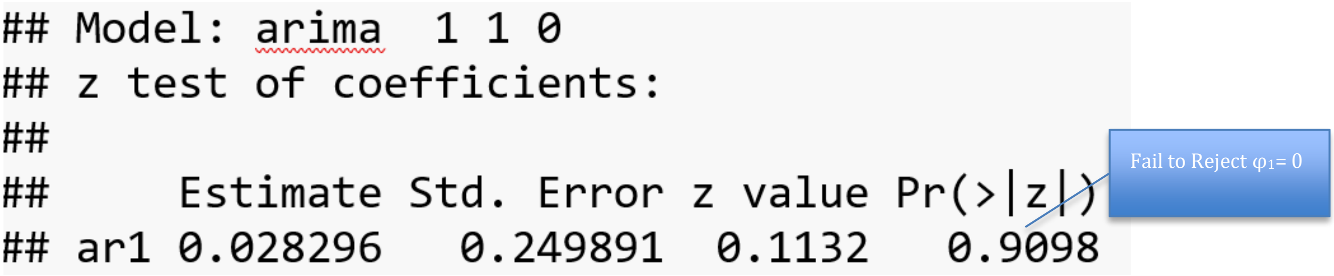


```
## Model: arima  0 1 2
## z test of coefficients:
##
##       Estimate Std. Error z value Pr(>|z|)
## ma1 -0.025817   0.335003 -0.0771   0.9386
## ma2 -0.080416   0.355606 -0.2261   0.8211
```

Fail to Reject $\theta_1 = 0$

Fail to Reject $\theta_2 = 0$

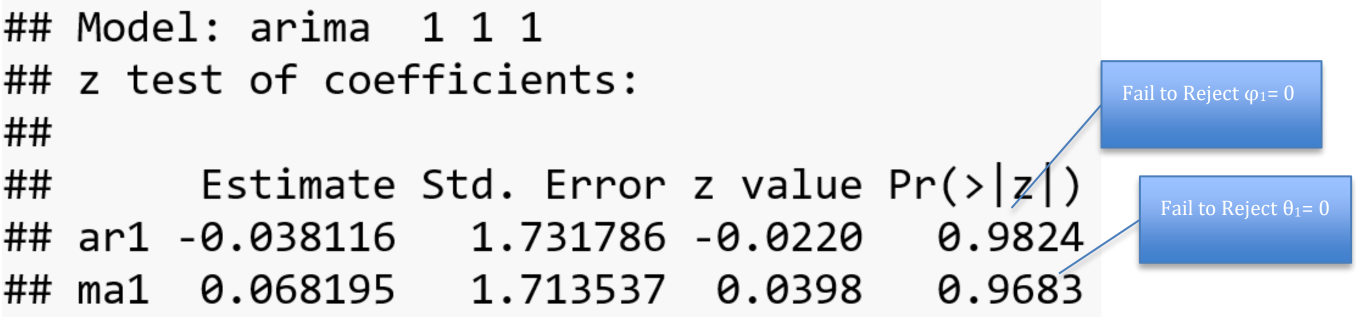
Table 17 - Coefficient Test result on ARIMA(0,1,2) with Least Square Estimation



```
## Model: arima  1 1 0
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1 0.028296   0.249891  0.1132   0.9098
```

Fail to Reject $\varphi_1 = 0$

Table 18 - Coefficient Test result on ARIMA(1,1,0) with Least Square Estimation

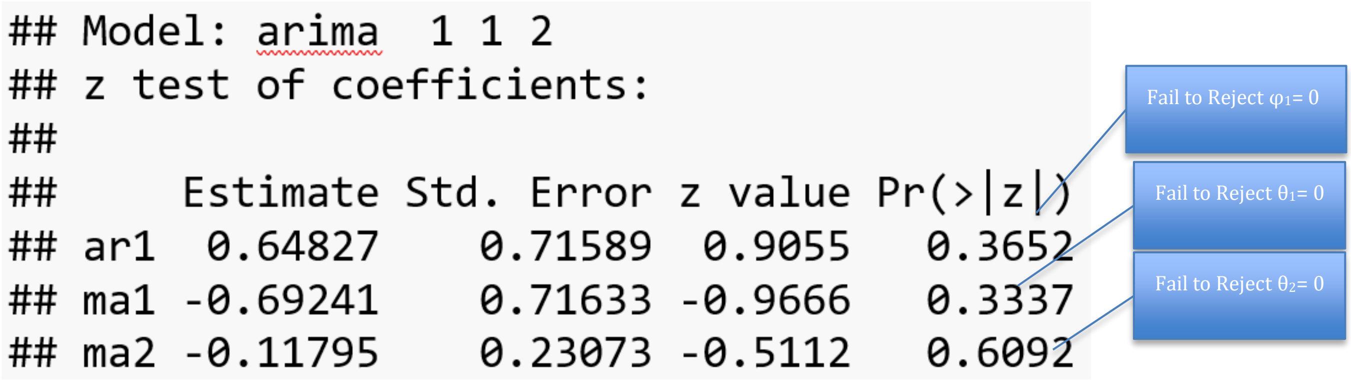


```
## Model: arima  1 1 1
## z test of coefficients:
##
##       Estimate Std. Error z value Pr(>|z|)
## ar1 -0.038116   1.731786 -0.0220   0.9824
## ma1  0.068195   1.713537  0.0398   0.9683
```

Fail to Reject $\varphi_1 = 0$

Fail to Reject $\theta_1 = 0$

Table 19- Coefficient Test result on ARIMA(1,1,1) with Least Square Estimation



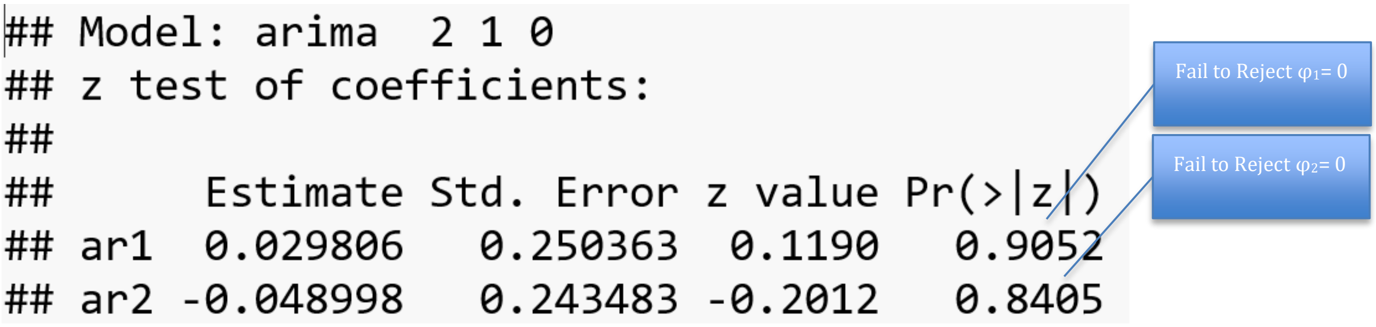
```
## Model: arima  1 1 2
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.64827    0.71589  0.9055   0.3652
## ma1 -0.69241    0.71633 -0.9666   0.3337
## ma2 -0.11795    0.23073 -0.5112   0.6092
```

Fail to Reject $\varphi_1 = 0$

Fail to Reject $\theta_1 = 0$

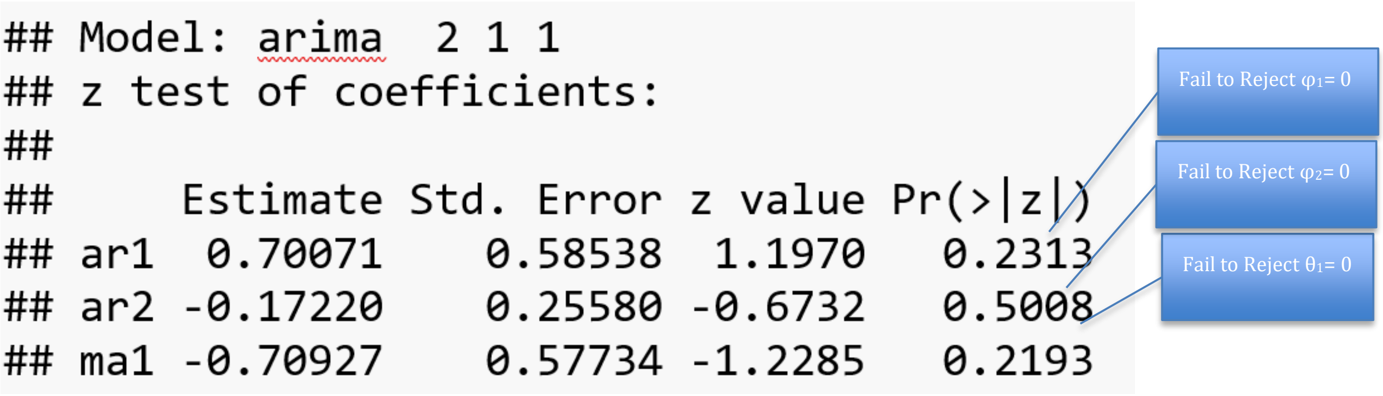
Fail to Reject $\theta_2 = 0$

Table 20- Coefficient Test result on ARIMA(1,1,2) with Least Square Estimation



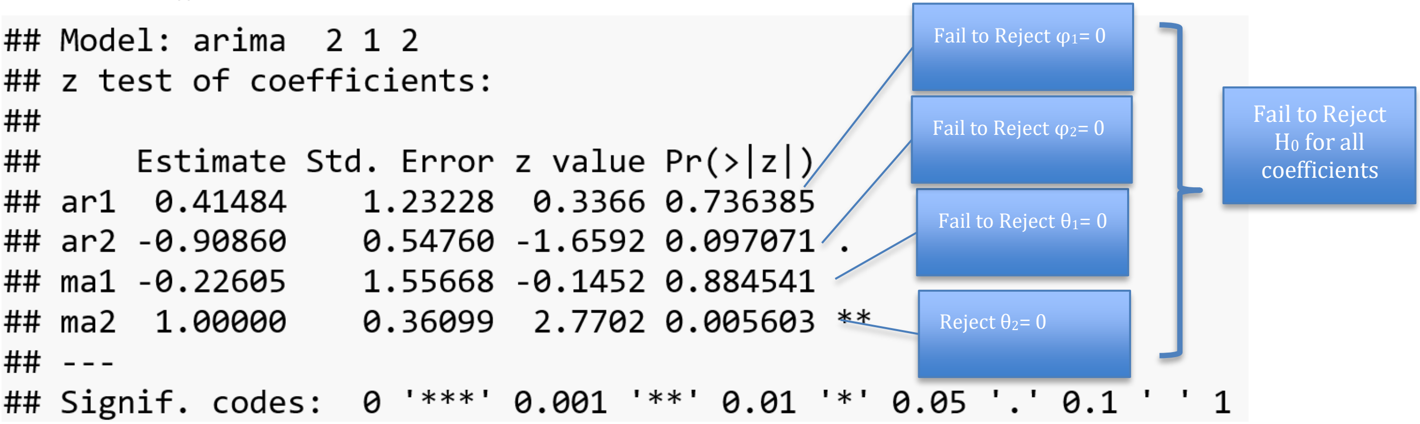
```
## Model: arima  2 1 0
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.029806   0.250363  0.1190   0.9052
## ar2 -0.048998   0.243483 -0.2012   0.8405
```

Table 21- Coefficient Test result on ARIMA(2,1,0) with Least Square Estimation



```
## Model: arima  2 1 1
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.70071    0.58538  1.1970   0.2313
## ar2 -0.17220    0.25580 -0.6732   0.5008
## ma1 -0.70927    0.57734 -1.2285   0.2193
```

Table 22- Coefficient Test result on ARIMA(2,1,1) with Least Square Estimation



```
## Model: arima  2 1 2
## z test of coefficients:
##
##     Estimate Std. Error z value Pr(>|z|)
## ar1  0.41484    1.23228  0.3366 0.736385
## ar2 -0.90860    0.54760 -1.6592 0.097071 .
## ma1 -0.22605    1.55668 -0.1452 0.884541
## ma2  1.00000    0.36099  2.7702 0.005603 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 23- Coefficient Test result on ARIMA(2,1,2) with Least Square Estimation

From all the above results, we notice that:

θ_1 or θ_2 or φ_1 or φ_2 are not significantly different from 0, which suggest ARIMA(0,1,1), ARIMA(0,1,2), ARIMA(1,1,0), ARIMA(1, 1,1), ARIMA(1,1,2), ARIMA(2,1,0), ARIMA(2,1,1), ARIMA(2,1,2) might be overfitting the time series data on egg depositions. ARIMA(0,1,0) might be already sufficient for fitting the data.

8. Perform Residual Analysis and find the best model

Refer to Appendix [A7] for coding.

I then go on to perform residual analysis to confirm if ARIMA(0,1,0) can well capture the time series of egg depositions, so that the residuals are normally distributed and uncorrelated. The histogram (Figure 21) shows that the residuals in this model is normally distributed and the q-q plot also shows most of the datapoints lie on the straight line (Figure 24), shaprio test [A9-2] (Table 24) fails to reject the null hypothesis which further confirm the normality of the residuals. From the ACF plot, we can see that none of the lag has significant correlation, ljung-box test [A9-4] (Table 24)also fails to reject the null hypothesis which further confirm the un-correlated condition.

For our further reference, residual analysis have also been run for all other candidate models and turn out to have similar analytical results as ARIMA(0,1,0) on residual analysis (refers to [A10] for details plots and tests for each model).

By inspecting the AIC and BIC index of all the models (Table 25 and Table 26), ARIMA(0,1,0) has the lowest index which subsequently indicate it is the best model to fit the egg depositions data.

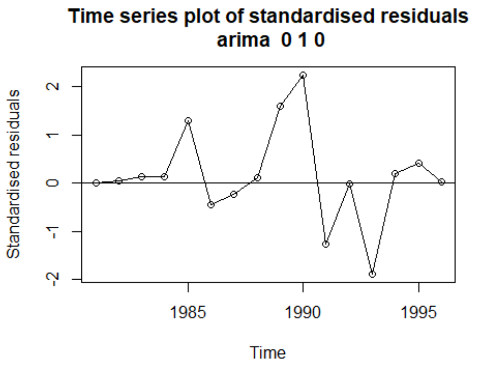


Figure 20- Time Series Plot of Standard Residuals of ARIMA (0,1,0)

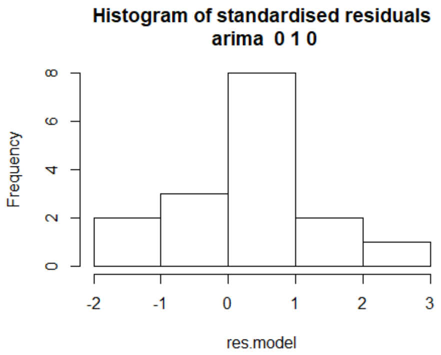


Figure 21- Histogram of Standard Residuals of ARIMA (0,1,0)

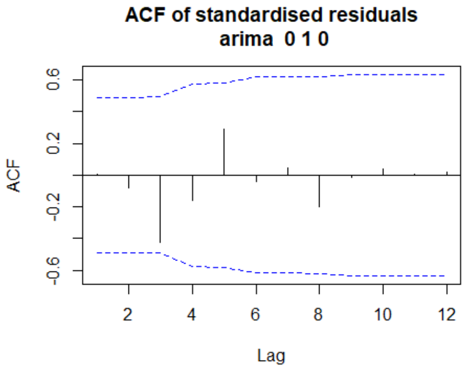


Figure 22- ACF Plot of Standard Residuals of ARIMA (0,1,0)

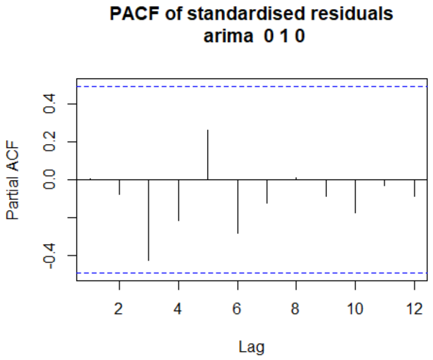


Figure 23- PACF Plot of Standard Residuals of ARIMA (0,1,0)

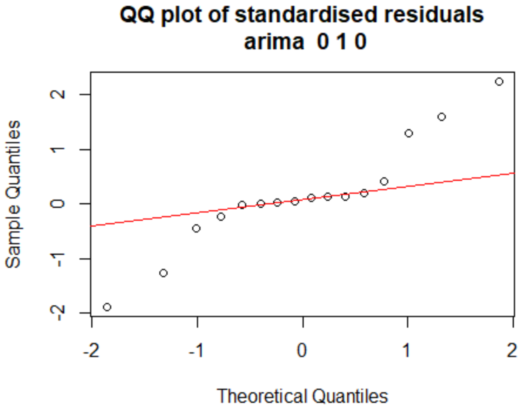


Figure 24 - Q-Q Plot of Standard Residuals of ARIMA (0,1,0)

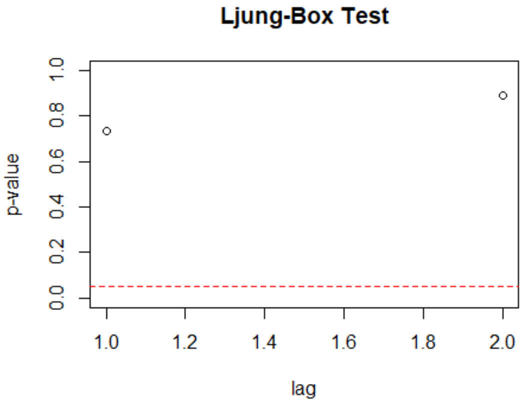
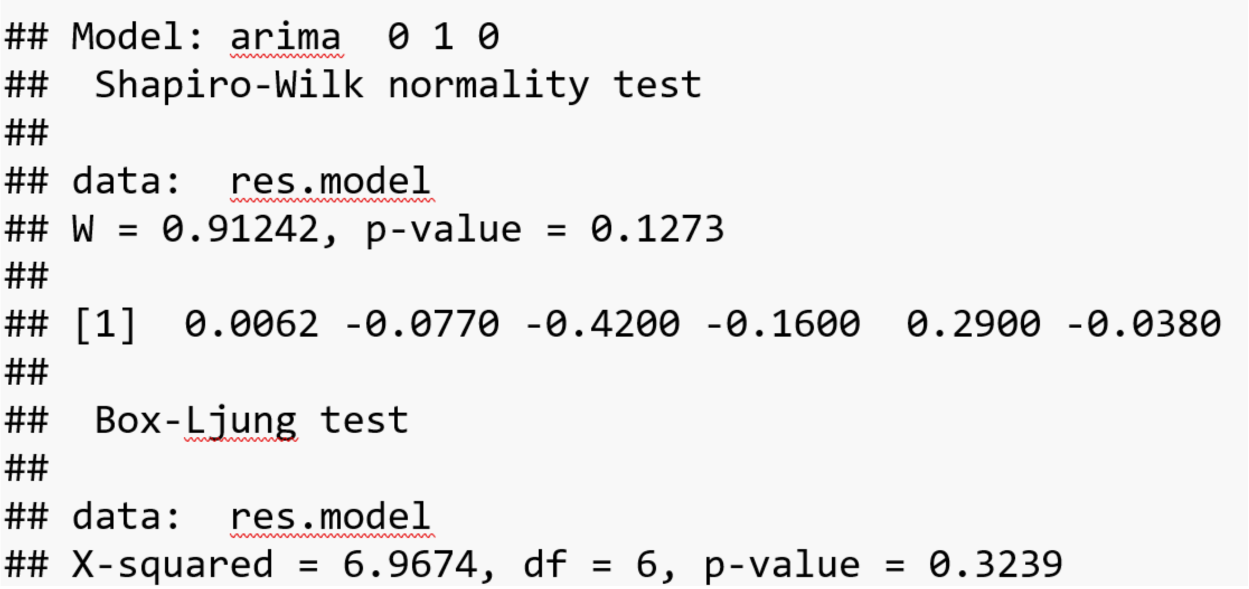
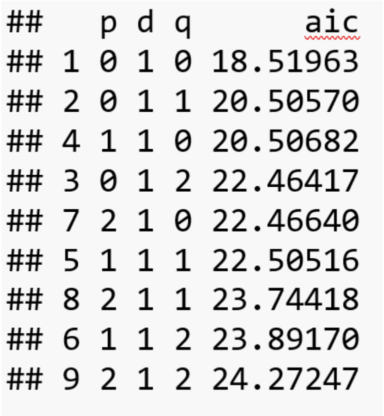


Figure 25 - Ljung Box Plot of Standard Residuals of ARIMA (0,1,0)



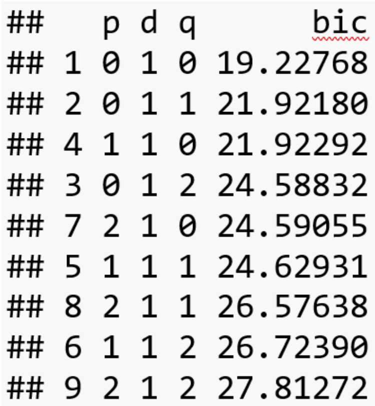
```
## Model: arima  0 1 0
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.91242, p-value = 0.1273
##
## [1]  0.0062 -0.0770 -0.4200 -0.1600  0.2900 -0.0380
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 6.9674, df = 6, p-value = 0.3239
```

Table 24 - Shapiro Normality and Ljung test for ARIMA(0,1,0)



```
##   p d q      aic
## 1 0 1 0 18.51963
## 2 0 1 1 20.50570
## 4 1 1 0 20.50682
## 3 0 1 2 22.46417
## 7 2 1 0 22.46640
## 5 1 1 1 22.50516
## 8 2 1 1 23.74418
## 6 1 1 2 23.89170
## 9 2 1 2 24.27247
```

Table 25 - AIC score for all models



```
##   p d q      bic
## 1 0 1 0 19.22768
## 2 0 1 1 21.92180
## 4 1 1 0 21.92292
## 3 0 1 2 24.58832
## 7 2 1 0 24.59055
## 5 1 1 1 24.62931
## 8 2 1 1 26.57638
## 6 1 1 2 26.72390
## 9 2 1 2 27.81272
```

Table 26- BIC score for all models

9. Five years predictions

Refer to Appendix [A8] for coding.

After reviewing all the candidate models, ARIMA(0,1,0) is most adequate to capture our data without overfitting and all the residuals in the model are within 95% confidence to be normally distributed and uncorrelated.

Using ARIMA(0,1,0), the next five years predictions are shown in Table 27. Lo80 and Hi 80 denotes the prediction boundaries with 80% confidence interval, which is shadowed in dark grey in Figure 26. Lo95 and Hi 95 denotes the prediction boundaries with 95% confidence interval, which is shadowed in light grey in Figure 26.

ARIMA(0,1,0) is also known as random walk series where the mean is constant but its variance is not. Thus, the forecast points for next 5 years are all 1.0241 million, with a widening confidence interval as prediction in subsequent year is highly dependent on current year.

Also, we know that egg deposition will not be negative, we can impute the negative numbers in the following prediction to 0.

```
##      Point Forecast       Lo 80    Hi 80      Lo 95    Hi 95
## 1997         1.0241  0.48626998 1.561930  0.2015600 1.846640
## 1998         1.0241  0.26349349 1.784707 -0.1391473 2.187347
## 1999         1.0241  0.09255108 1.955649 -0.4005811 2.448781
## 2000         1.0241 -0.05156004 2.099760 -0.6209801 2.669180
## 2001         1.0241 -0.17852448 2.226724 -0.8151554 2.863355
```

Table 27- Confidence Intervals for forecasting Egg Depositions (using ARIMA(0,1,0) for year 1997 to 2001

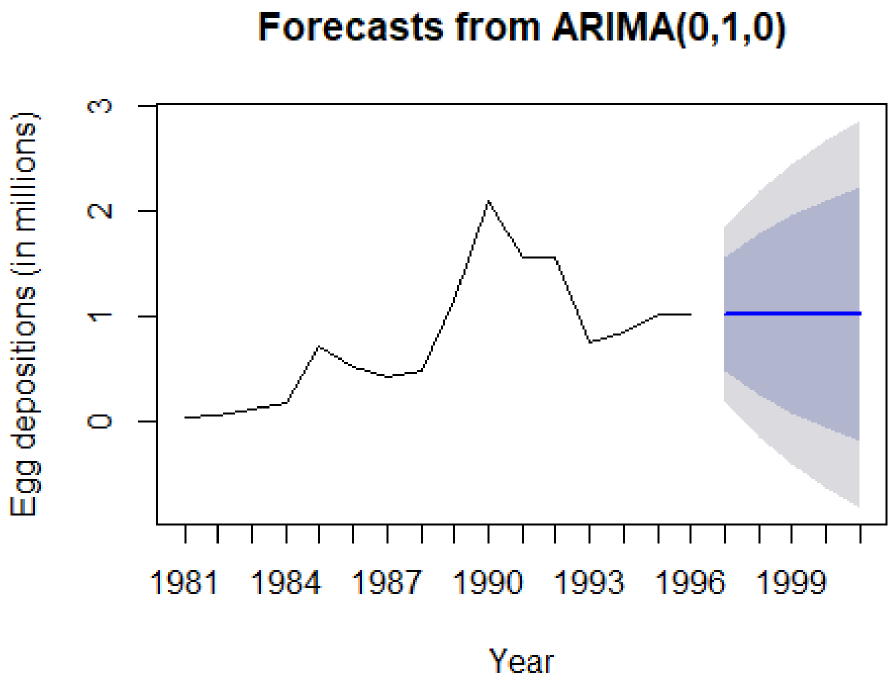


Figure 26- Forecasts of Egg Depositions using ARIMA(0,1,0) for year 1997 to 2001

10. Conclusion

After analyzing the egg deposition dataset, we have the following findings:

* There is an upward trend in this short time series(only 16 observations). It also has a high correlation with its first time lag (r=0.744).

* Data Transformation is not applicable to this time series for modelling as there is no improvement on normality and stationarity.

* As there is no particular reason to believe the trend in the egg deposition would be exactly the same for another period of time, we model it as a stochastic trend.

* The possible ARIMA candidate models are {ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(0,1,2), ARIMA(1,1,0), ARIMA(1, 1,1), ARIMA(1,1,2), ARIMA(2,1,0), ARIMA(2,1,1), ARIMA(2,1,2)}

* None of the model in the set {ARIMA(0,1,1), ARIMA(0,1,2), ARIMA(1,1,0), ARIMA(1, 1,1), ARIMA(1,1,2), ARIMA(2,1,0), ARIMA(2,1,1), ARIMA(2,1,2)} is **all** coefficients significant, they are found to be overfitting the egg deposition dataset.

*Residuals of ARIMA(0,1,0) are within 95% confidence to be normally distributed and uncorrelated. Thus it's the best fit model in our analysis.

* Forecast for the next 5 years will have the same prediction point (1.0241 millions) with a widening confidence interval, as our best fit model, ARIMA(0,1,0) is also known as random walk series where the mean is constant but its variance is not. The prediction in subsequent year is highly dependent on current year in random walk, thus the variance of uncertainty in later years would not be constant and our prediction fall within a much wider interval.

References

#References [1] What is the Augmented Dickey Fuller Test? https://www.statisticshowto.com/adf-augmented-dickey-fuller-test/. Accessed on 10-05-2020.

[2] Comparing ADF test functions in R. https://fabian-kostadinov.github.io/2015/01/27/comparing-adf-test-functions-in-r/. Accessed on 10-05-2020.

Appendix

[A1] – Import packages and data preparation

```r
#The following packages are needed in this assignment:

library(TSA)

library(forecast)

library(tseries)

library(knitr)

library(fUnitRoots)

library(lmtest)

library(FitAR)

library(summarytools)

#Read in the dataset
eggs <- read.csv("D:/RMIT Master of Analytics/semester 2/MATH1318 - Time Seri
es/Assignment 2/eggs.csv", header = TRUE)


#covert to a timeseries object.
eggs <- ts(as.vector(eggs$eggs), start=1981, end=1996)
class(eggs)
```

```
## [1] "ts"
```

[A2] – Generate descriptive statistics of the time series

```
# Put the summary statistics in table format
kable(descr(eggs, stats = c("mean", "med", "sd", "Q1", "Q3", "min", "max", "n
.valid"), transpose = TRUE), caption = "Summary statistics of Egg depositions
between 1981 and 1996")
```

[A3] – Produce relevant plots and perform normality tests

```
#Define a function plot.all, which would plot the following graphs of the pas
s-in time series data:
# 1. Generate the time series Plot
# 2. Scatter plot of the data with its first time lag, also show this correla
tion index
# 2. Normality via QQ-plot and Shapiro test
# 3. Generate the ACF and PACF plot
plot.all <- function(ts_data, ts_plot_title, scatter_plot_title, qq_plot_titl
e, acf_title, pacf_title, isDiff=TRUE){

        #Time series plot

        plot(ts_data, xaxp=c(1981,1996, 15), type='o', xlab = 'Year', ylab='E
gg depositions (in millions)', main = ts_plot_title)

        if(isDiff == FALSE){

                #Scatter Plot and Check correlation of 1st lagging
                plot(y=ts_data,x=zlag(ts_data),ylab='Thickness of Ozone layer
', xlab='Previous Year thickness of Ozone layer' , main = scatter_plot_title)

                y = ts_data
                x = zlag(ts_data)        # Generate first lag of the series
                index = 2:length(x)
                print('Correlation Index:')
                print(cor(y[index],x[index]))

                #QQ Plot and check Normality
                qqnorm(ts_data, main=qq_plot_title)
                qqline(ts_data, col = 2)
                print(shapiro.test(ts_data))
        }
        #ACF and PACF plot
        acf(ts_data, ci.type='ma',xaxp=c(0,20,10), main=acf_title)
        pacf(ts_data, xaxp=c(0,20,10), main=pacf_title)
```

```
}


#Plot all the associated graphs for the Original Time series Data
plot.all(eggs, 'Time series plot of Egg depositions \n for original data', 'S
catter plot of Original Egg depositions \n data with first time lag', 'Quanti
les plot of Original Egg \n depositions data', 'ACF plot for the original Egg
\n depositions data', 'PACF plot for the original Egg \n depositions data', i
sDiff = FALSE )
```

[A4] – Box-cox and log transformation

```
#Apply BoxCox and Log Transformation, check:
# 1. normality has been improved
# 2. correlation with first time lag has been reduced
eggs.transform = BoxCox.ar(eggs) # The default method for fitting is MLE here

eggs.transform$ci

## [1] 0.7 1.0

eggs.transform = BoxCox.ar(eggs, method = "yule-walker") #MoM

eggs.transform$ci

## [1] 0.1 0.8

lambda = 0.75
BC.eggs = (eggs^lambda-1)/lambda

log.eggs = log(eggs)


#Plot all the associated graphs for the BoxCox and Log Transformed Data
plot.all(BC.eggs, 'Time series plot of Egg depositions\n after Box Cox Transf
ormation', 'Scatter plot of Egg depositions with first \n time lag after Box
Cox Transformation', 'Quantiles plot of Egg depositions \n after Box Cox Tran
sformation', 'ACF plot of Egg depositions \n after Box Cox Transformation', '
PACF plot of Egg depositions \n  after Box Cox Transformation', isDiff = FALS
E  )

plot.all(log.eggs, 'Time series plot of Egg depositions\n after Log Transform
ation', 'Scatter plot of Egg depositions with first \n time lag after Log Tra
nsformation', 'Quantiles plot of Egg depositions after \n Log Transformation'
, 'ACF plot of Egg depositions \n after Log Transformation', 'PACF plot of Eg
g depositions \n after Log Transformation', isDiff = FALSE  )
```

[A5] - Differencing

```
#Use ar function to find the Order with lowest AIC from the first differenced
data, use this order to
#apply adfTest on the original data
lagOrder=ar(diff(eggs))$order
cat("lagOrder=",lagOrder)

## lagOrder= 0

adfTest(eggs, lags = lagOrder,  title = NULL,description = NULL)

##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 0
##   STATISTIC:
##     Dickey-Fuller: -0.4911
##   P VALUE:
##     0.452
##
## Description:
##  Sat May 09 13:19:47 2020 by user: millie

#Apply first differecing on the original data, find the order with lowest AIC
from the second differenced
#data

diff1_eggs = diff(eggs, differences = 1)
lagOrder=ar(diff(diff1_eggs))$order




#Define a function adfTest.loop which would perform adfTest for the pass-in m
odel with
#all lag order from 1 to max.lag, print the result in a tabulated dataframe:
adfTest.loop <- function(model, max.lag, table.caption){

        adf.lag <- vector()
        adf.pvalue <- vector()
        adf.stats <- vector()
        for (lag_num in 1:max.lag)
        {
                result=adfTest(model, lags = lag_num,  title = NULL,descripti
```

```
on = NULL)
                adf.lag <- c(adf.lag, lag_num)
                adf.pvalue <- c(adf.pvalue, result@test$p.value)
                adf.stats <- c(adf.stats, result@test$statistic)
                #        adfPstats <- adfTest@test$p.value
                #        cat("pstatistics",adfPstats)
        }

        adf.data <- data.frame(adf.lag, round(adf.stats,3), round(adf.pvalue,
3))
        names(adf.data)[1] <- "lag"
        names(adf.data)[2] <- "statistics"
        names(adf.data)[3] <- "p value"
        kable(adf.data, caption = table.caption)
}



cat("lagOrder=",lagOrder)
```

```
## lagOrder= 4
```

```
#Apply adfTest.loop to first differenced data
adfTest.loop(diff1_eggs, lagOrder, "adfTest for first differencing")

#Apply adfTest.loop to second differenced data
diff2_eggs = diff(eggs, differences = 2)
adfTest.loop(diff2_eggs, lagOrder, "adfTest for second differencing")

#Apply adfTest.loop to third differenced data

diff3_eggs = diff(eggs, differences = 3)
adfTest.loop(diff3_eggs, lagOrder, "adfTest for third differencing")

#Apply adfTest.loop to fourth differenced data

diff4_eggs = diff(eggs, differences = 4)
adfTest.loop(diff4_eggs, lagOrder, "adfTest for fourth differencing")

#Plot all the associated graphs for the first differenced and second differen
ced data

plot.all(diff1_eggs, 'Time series plot of Egg depositions\n after first diffe
rencing', 'Null', 'Null', 'ACF plot for Egg depositions\n after first differe
ncing', 'PACF plot for Egg depositions\n after first differencing', isDiff =
TRUE )

plot.all(diff2_eggs, 'Time series plot of Egg depositions\n after second diff
erencing', 'Null', 'Null', 'ACF plot for Egg depositions\n after second diffe
rencing', 'PACF plot for Egg depositions\n after second differencing', isDiff
= TRUE )
```

[A6] – Model Estimation

```
#Create the 1st arima model for model estimation
model_011_css = arima(eggs,order=c(0,1,1),method='CSS')
model_011_ml = arima(eggs,order=c(0,1,1),method='ML')
coeftest(model_011_css)

##
## z test of coefficients:
##
##     Estimate Std. Error z value Pr(>|z|)
## ma1 0.032491   0.262456  0.1238   0.9015

coeftest(model_011_ml)

##
## z test of coefficients:
##
##     Estimate Std. Error z value Pr(>|z|)
## ma1 0.030394   0.254473  0.1194   0.9049

#The above code chuck needs to run repetitively for all the candidate models
#Streamline this by defining a function model.Estimation which would perform
coeftest on all the possible model which would pass-in as an orderList in c(p
,d,q) format, this function is inspired from
#the myCandidate function developed by Yong Kai, Wong in TSHandy.r

model.Estimation <- function(timeSeries, orderList, methodType = c("CSS-ML",
"ML", "CSS")[1])
{

        model <- list()
        resAnalysis <- list()
        n <- length(orderList)
        for(i in 1:n){
                order <- sapply(orderList,function(x) unlist(x))[,i]
                model[[i]] <- Arima(y = timeSeries, order = order, method = "
ML")
        }
        for(i in 1:n){
                p=0;d=0;q=0
                for(j in 1:3){
                        if(j==1){p <- orderList[[i]][j]}
                        if(j==2){d <- orderList[[i]][j]}
                        if(j==3){q <- orderList[[i]][j]}

                }
```

```
                arimaOrderStr <- paste("arima ", p, d,q)
                print("======================================================
===============")
                cat("Model:",arimaOrderStr)
                print(coeftest(model[[i]]))
        }
}

#No parameter (phi or theta) needs to be estimated for arima(0,1,0)

modelList <- list(c(0,1,1), c(0,1,2), c(1,1,0), c(1,1,1), c(1,1,2), c(2,1,0),
c(2,1,1), c(2,1,2))


#Call the model.Estimation function on all the possible models for the timese
ries data with
# maximum likelihood estimation
model.Estimation(eggs, orderList = modelList, methodType = "ML")

#Call the model.Estimation function on all the possible models for the timese
ries data with
#least square estimation
model.Estimation(eggs, orderList = modelList, methodType = "CSS")
```

[A7] – Residual Analysis

```
#Define a function residual.analysis which would perform the following plots
of the residuals of
# the pass-in arima model
# 1. time series plot of the residuals
# 2, Histogram of the residuals
# 3. ACF plot of the residuals
# 4. PACF plot of the residuals
# 5. Q-Q plot of the residuals
# 6. Ljung-Box plot of the residuals
# 7. Ljung-Box test of the residuals
#
#this function is originated from the residual.analysis function developed by
Yong Kai, Wong
# I just add in the  Ljung-Box test

residual.analysis <- function(model, std = TRUE,start = 2, class = c("ARIMA",
"GARCH","ARMA-GARCH")[1], p, d, q){
        # If you have an output from arima() function use class = "ARIMA"
        # If you have an output from garch() function use class = "GARCH"
        # If you have an output from ugarchfit() function use class = "ARMA-G
ARCH"
        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$residuals
        }else {
                stop("The argument 'class' must be either 'ARIMA' or 'GARCH' ")
        }
        arimaOrderStr <- paste("arima ", p, d,q)
        plot(res.model,type='o',ylab='Standardised residuals', main=paste("Time series plot of standardised residuals", arimaOrderStr))
        abline(h=0)
        hist(res.model,main=paste("Histogram of standardised residuals", arimaOrderStr))
        acf(res.model,ci.type='ma', main=paste("ACF of standardised residuals ",arimaOrderStr))
        pacf(res.model,main=paste("PACF of standardised residuals", arimaOrderStr))
        qqnorm(res.model,main=paste("QQ plot of standardised residuals", arimaOrderStr))
        qqline(res.model, col = 2)
        print("====================================================================")
        cat("Model:",arimaOrderStr)
        print(shapiro.test(res.model))
        print(signif(acf(res.model,plot=F)$acf[1:6],2))
        print(Box.test(res.model, lag = 6, type = "Ljung-Box", fitdf = 0))
        k=0
        LBQPlot(res.model, lag.max = 2, StartLag = k + 1, k = 0, SquaredQ = FALSE)

}

#Define the score.Call.ResAnalysis function which would Call the residual.analysis function
#on all the possible model pass in as an orderList in c(p,d,q) format, at the end it will
#print out the AIC, AICC and BIC score (sorted in ascending order) of all these models
#this function is inspired from the myCandidate function developed by Yong Kai, Wong in TSHandy.r


score.Call.ResAnalysis <- function(timeSeries, orderList)
```

```
{

        model <- list()
        resAnalysis <- list()
        n <- length(orderList)
        for(i in 1:n){
                order <- sapply(orderList,function(x) unlist(x))[,i]
                model[[i]] <- Arima(y = timeSeries, order = order, method = "ML")
        }
        AICTable <- matrix(NA, nrow = n, ncol = 4) # create a matrix to store IC
        AICCTable <- matrix(NA, nrow = n, ncol = 4) # create a matrix to store IC
        BICTable <- matrix(NA, nrow = n, ncol = 4) # create a matrix to store IC

        for(i in 1:n){
                p=0;d=0;q=0
                for(j in 1:3){
                        AICTable[i,j] <- orderList[[i]][j]       # return the ARIMA orders
                        AICCTable[i,j] <- orderList[[i]][j]       # return the ARIMA orders
                        BICTable[i,j] <- orderList[[i]][j]       # return the ARIMA orders

                        if(j==1){p <- orderList[[i]][j]}
                        if(j==2){d <- orderList[[i]][j]}
                        if(j==3){q <- orderList[[i]][j]}

                }
                AICTable[i,4] <- model[[i]]$aic
                AICCTable[i,4] <- model[[i]]$aicc
                BICTable[i,4] <- model[[i]]$bic

                residual.analysis(model[[i]], std = TRUE,start = 1, 'ARIMA', p, d, q)

        }

        AICTable <- data.frame(AICTable)
        AICCTable <- data.frame(AICCTable)
        BICTable <- data.frame(BICTable)
        names(AICTable) <- c('p', 'd', 'q', 'aic')
        AICTable <- AICTable[order(AICTable$aic),]
        names(AICCTable) <- c('p', 'd', 'q', 'aicc')
        AICCTable <- AICCTable[order(AICCTable$aicc),]
        names(BICTable) <- c('p', 'd', 'q', 'bic')
```

```
        BICTable <- BICTable[order(BICTable$bic),]

        print(AICTable)
        print("==============================================================
=======")

        print(AICCTable)
        print("==============================================================
=======")

        print(BICTable)

}

#Include model arima(0,1,0) in the list call score.Call.ResAnalysis
modelList <- list(c(0,1,0), c(0,1,1), c(0,1,2), c(1,1,0), c(1,1,1), c(1,1,2),
c(2,1,0), c(2,1,1), c(2,1,2))
score.Call.ResAnalysis(eggs, orderList = modelList)
```

[A8] - Forecasting

```
fit = Arima(eggs,c(0,1,0))
forecast(fit,h=5)

plot(forecast(fit,h=5), ylab='Egg depositions (in millions)', xlab='Year', xa
xp=c(1981,2001, 20))
```

[A9] – Null Hypothesis of various statistical test

1. Shapiro Normality Test

H_0: Data is normally distributed

H_A: Data is not normally distributed

2. Augmented Dickey-Fuller Test

H_0: Data is non-stationary

H_A: Data is stationary

3. Linear Regression Test on Model coefficient

H_0: Coefficient = 0

H_A: Coefficient ≠ 0

4. Ljung-Box Test

H_0: Data within the specified lags are uncorrelated

H_A: Data within the specified lags are correlated

[A10] – Residual plots for all the possible ARIMA models

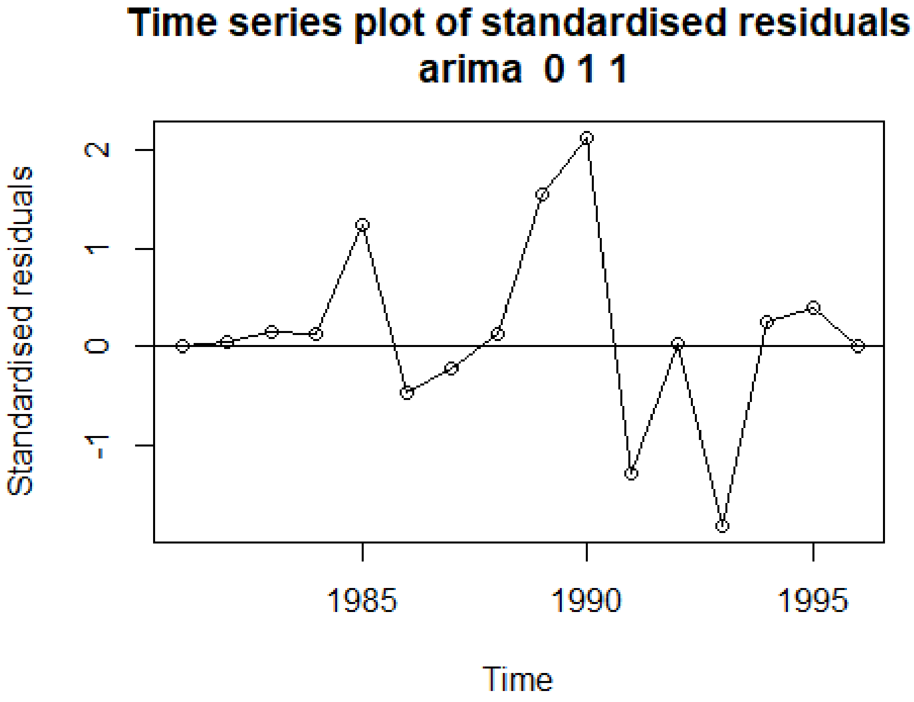


Figure 27- Time Series Plot of Standard Residuals of ARIMA (0,1,1)

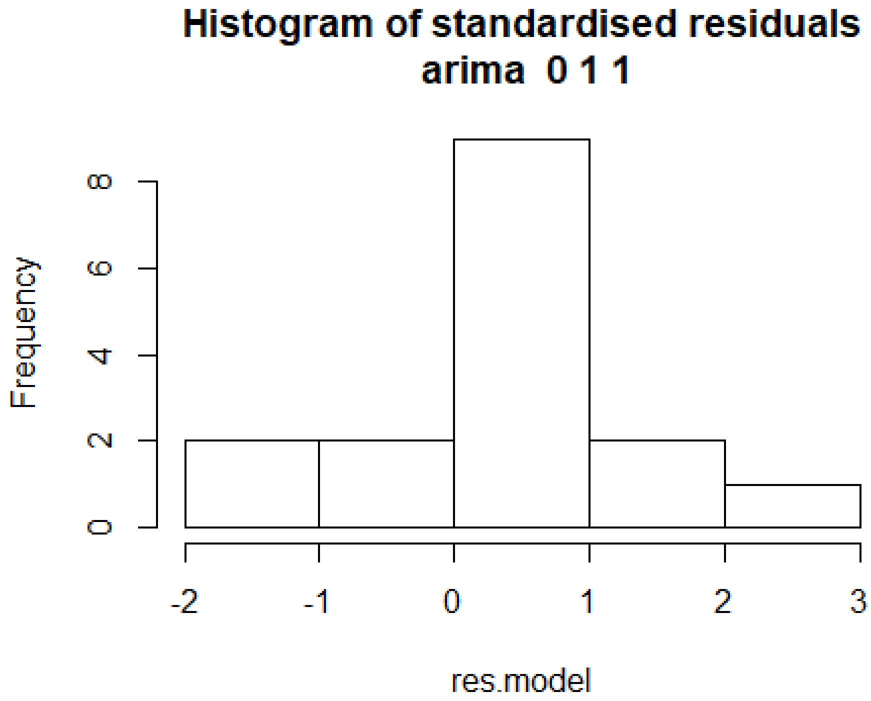


Figure 28 - Histogram of Standard Residuals of ARIMA (0,1,1)

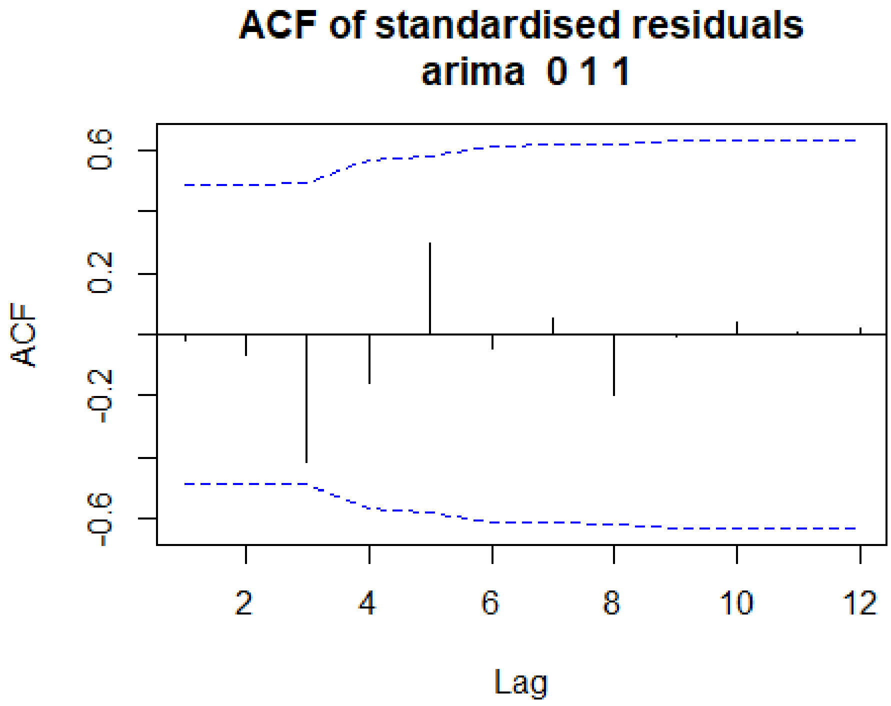


Figure 29- ACF Plot of Standard Residuals of ARIMA (0,1,1)

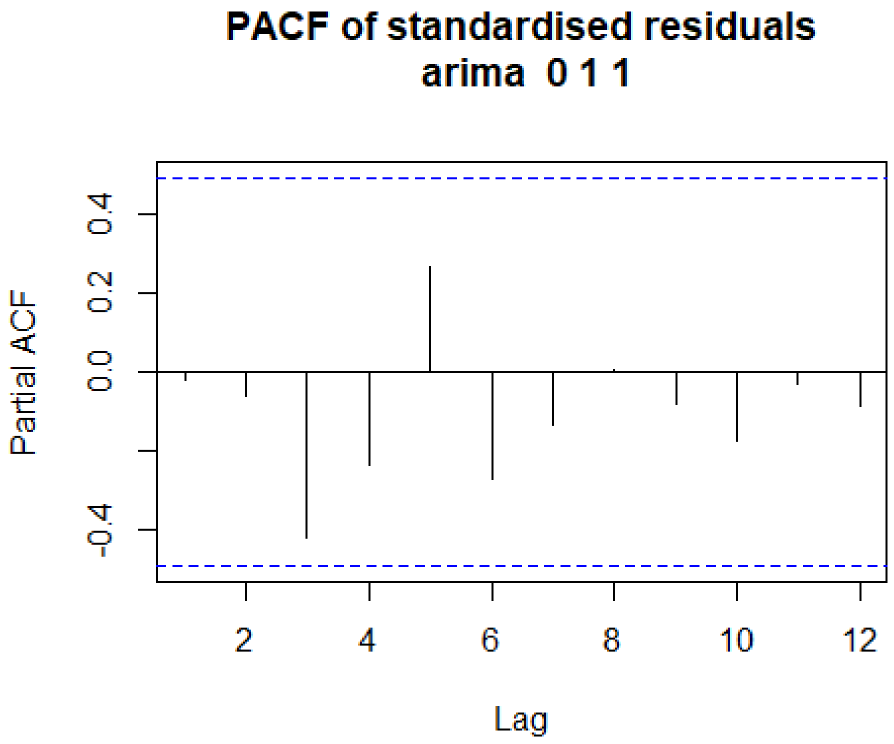


Figure 30- PACF Plot of Standard Residuals of ARIMA (0,1,1)

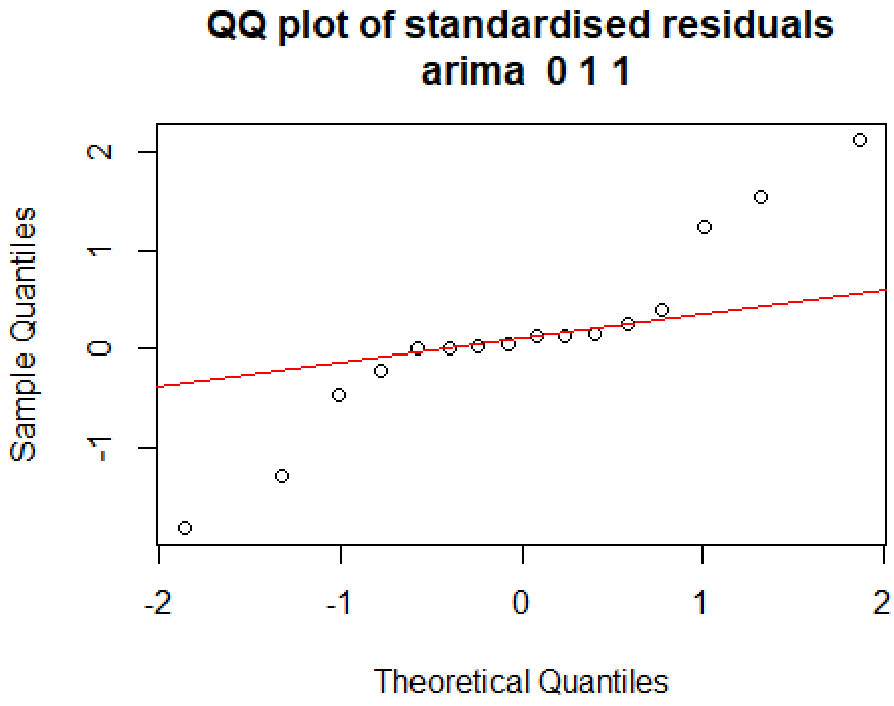


Figure 31- Q-Q Plot of Standard Residuals of ARIMA (0,1,1)

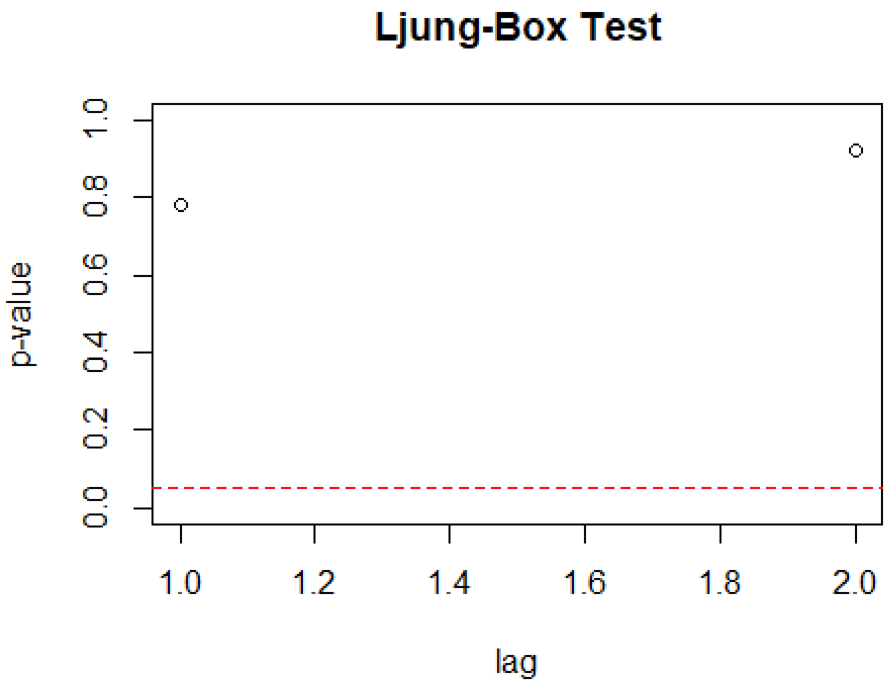


Figure 32- Ljung Box Plot of Standard Residuals of ARIMA (0,1,1)

```
## Model: arima  0 1 1
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.91443, p-value = 0.1372
##
## [1] -0.022 -0.064 -0.410 -0.160  0.300 -0.049
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 6.8838, df = 6, p-value = 0.3317
```

Table 28 - Shapiro Normality and Ljung test for ARIMA(0,1,1)

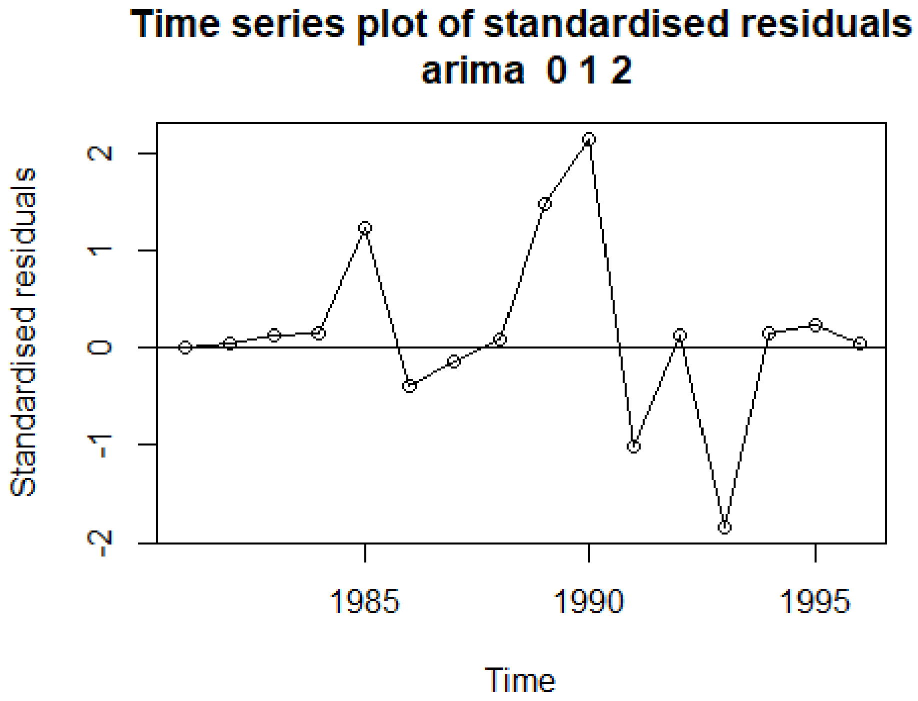


Figure 33- Time Series Plot of Standard Residuals of ARIMA (0,1,2)

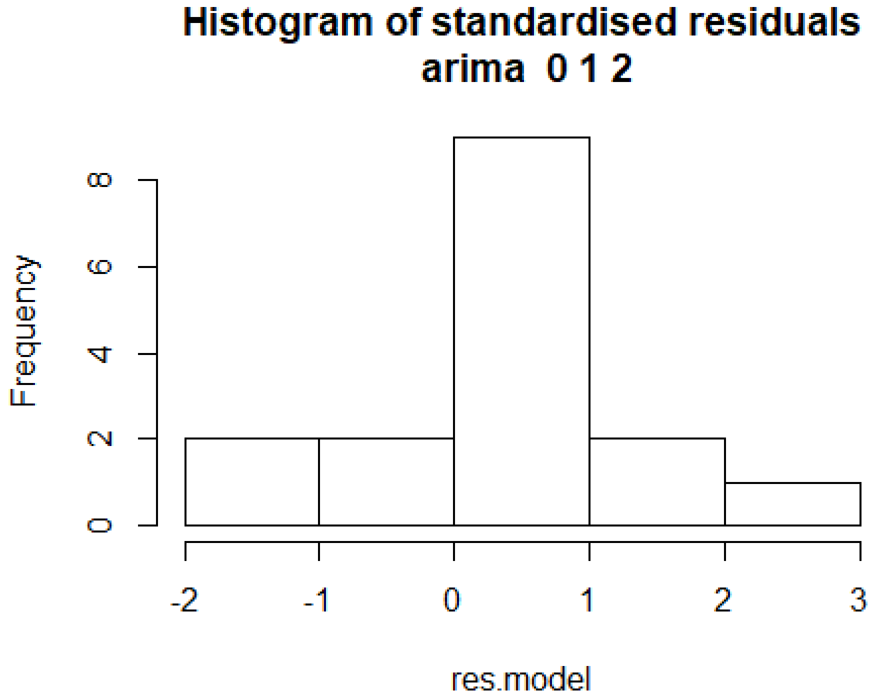


Figure 34 - Histogram of Standard Residuals of ARIMA (0,1,2)

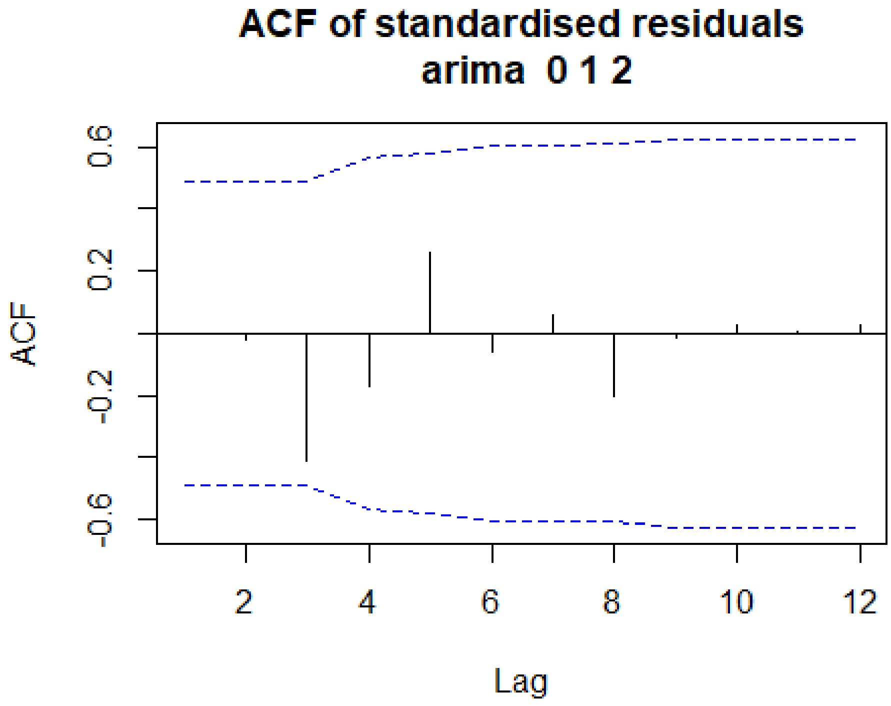


Figure 35- ACF Plot of Standard Residuals of ARIMA (0,1,2)

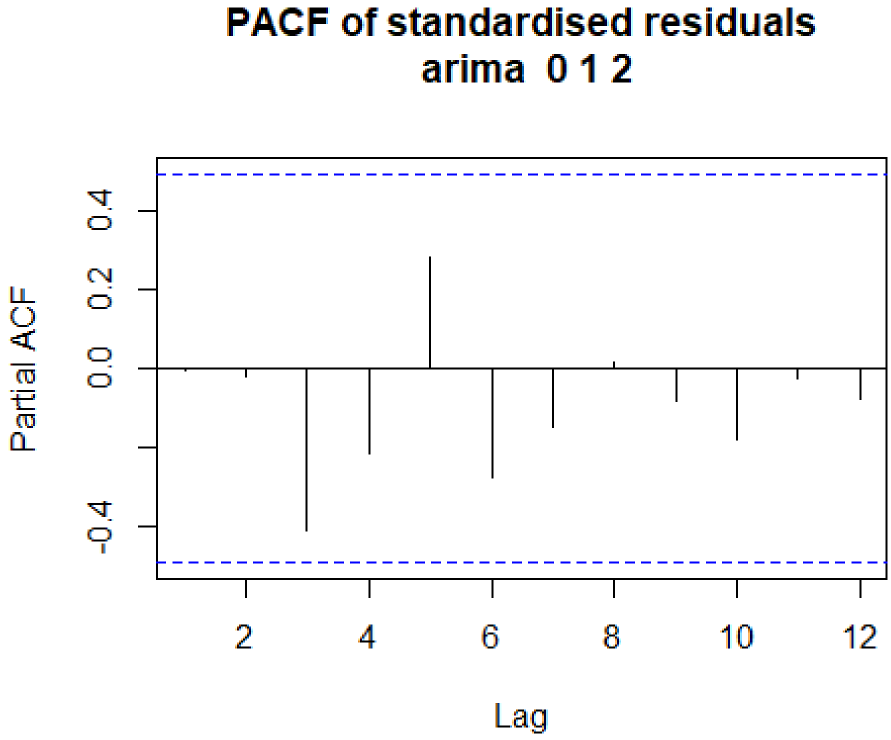


Figure 36 - PACF Plot of Standard Residuals of ARIMA (0,1,2)

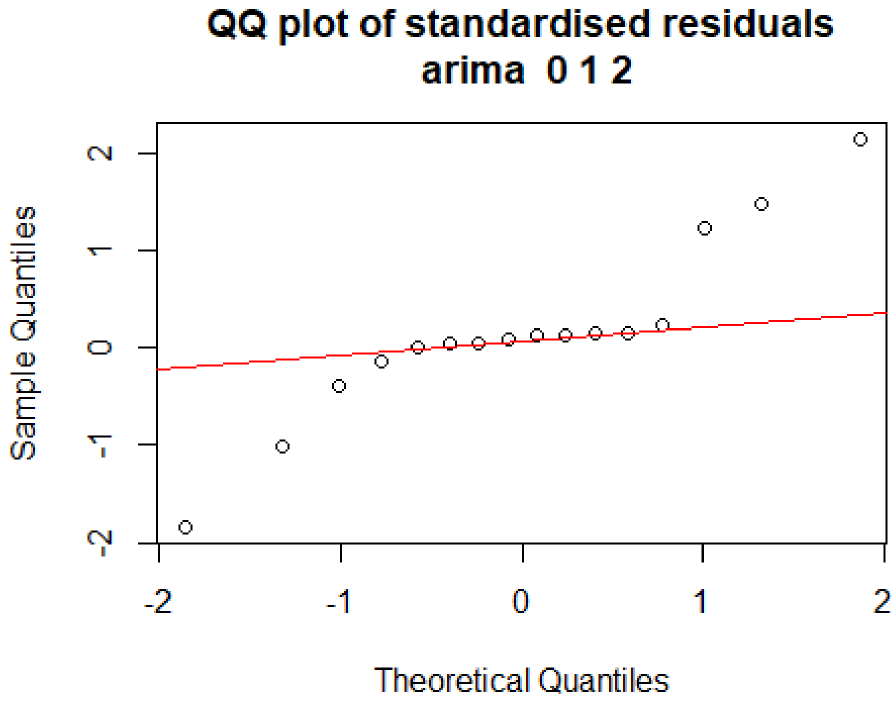


Figure 37- Q-Q Plot of Standard Residuals of ARIMA (0,1,2)

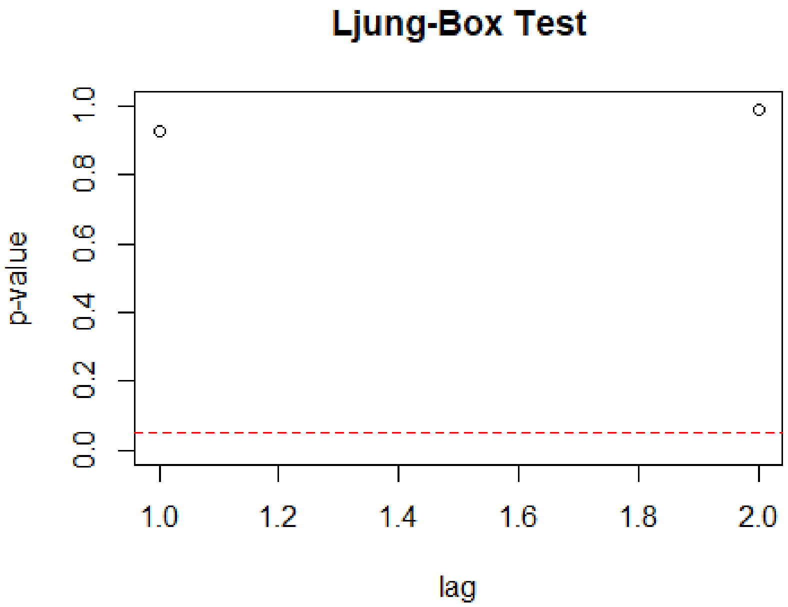


Figure 38- Ljung Box Plot of Standard Residuals of ARIMA (0,1,2)

```
## Model: arima  0 1 2
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.88708, p-value = 0.05009
##
## [1] -0.0031 -0.0210 -0.4100 -0.1700  0.2600 -0.0600
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 6.2703, df = 6, p-value = 0.3936
```

Table 29- Shapiro Normality and Ljung test for ARIMA(0,1,2)

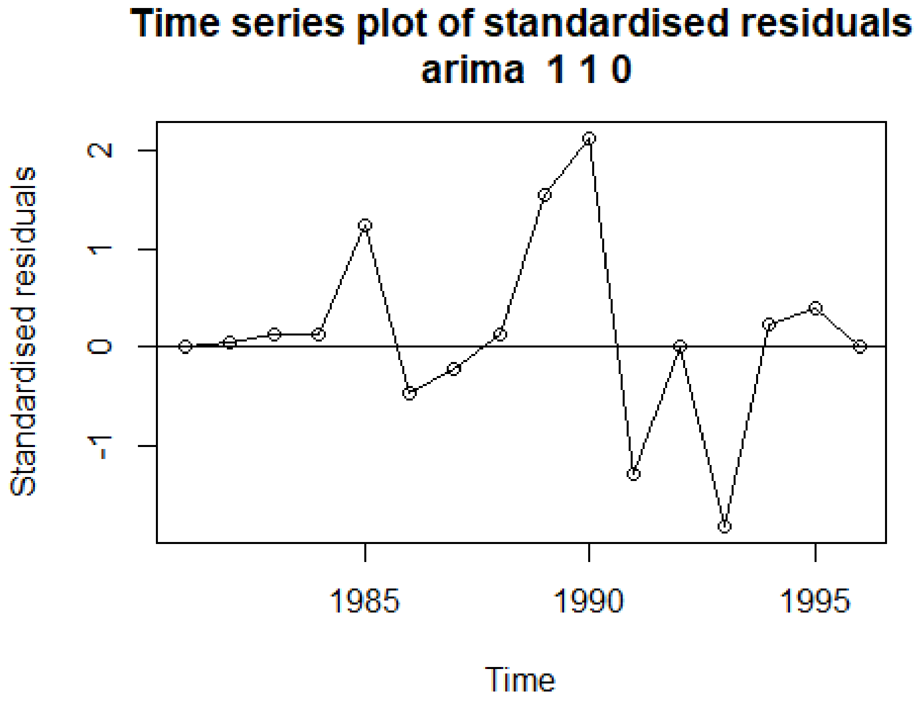


Figure 39 - Time Series Plot of Standard Residuals of ARIMA (1,1,0)

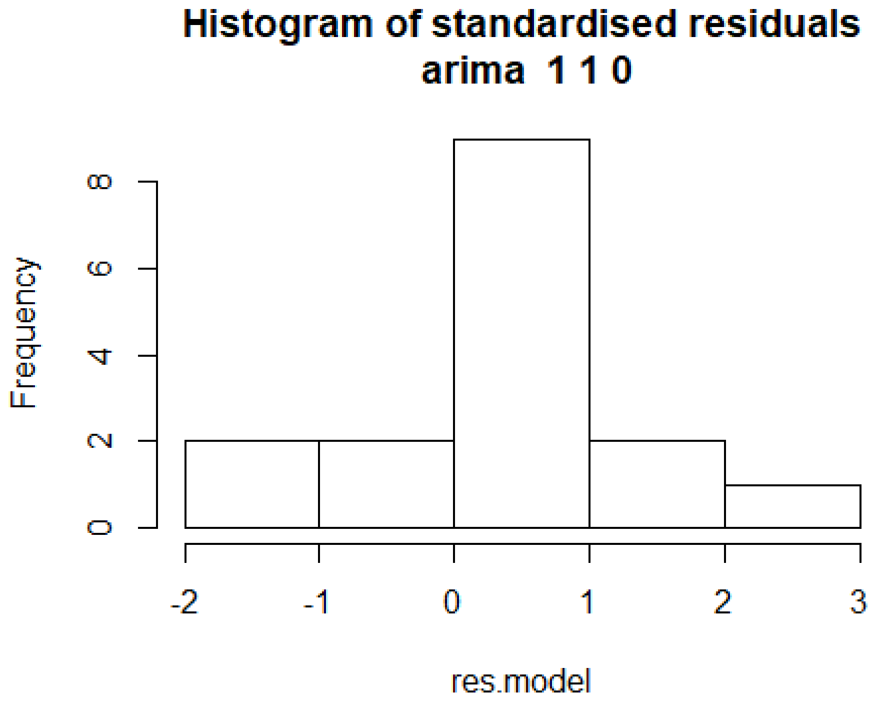


Figure 40 - Histogram of Standard Residuals of ARIMA (1,1,0)

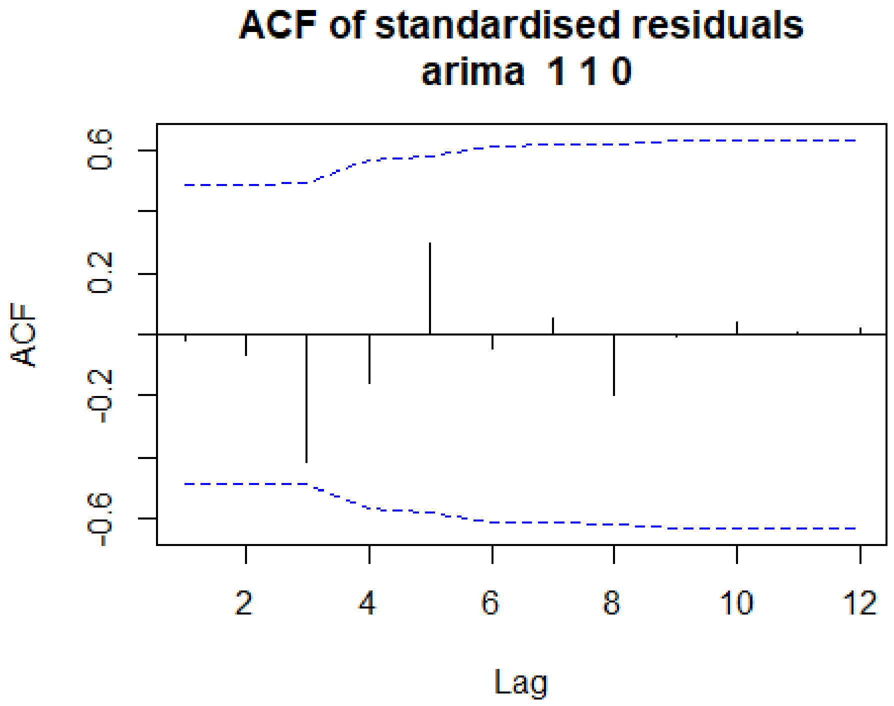


Figure 41 - ACF Plot of Standard Residuals of ARIMA (1,1,0)

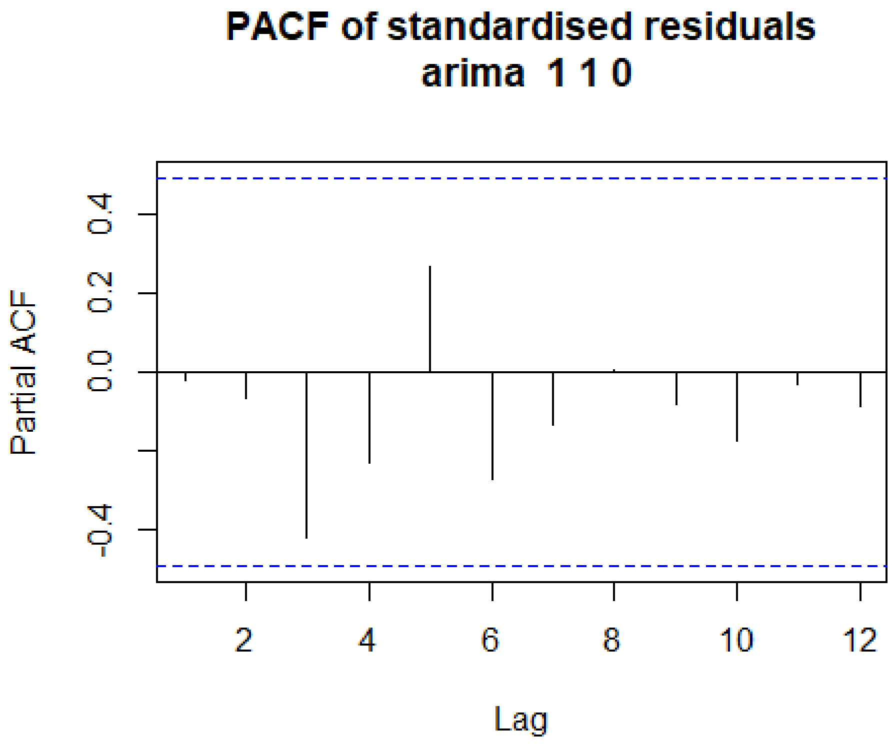


Figure 42- PACF Plot of Standard Residuals of ARIMA (1,1,0)

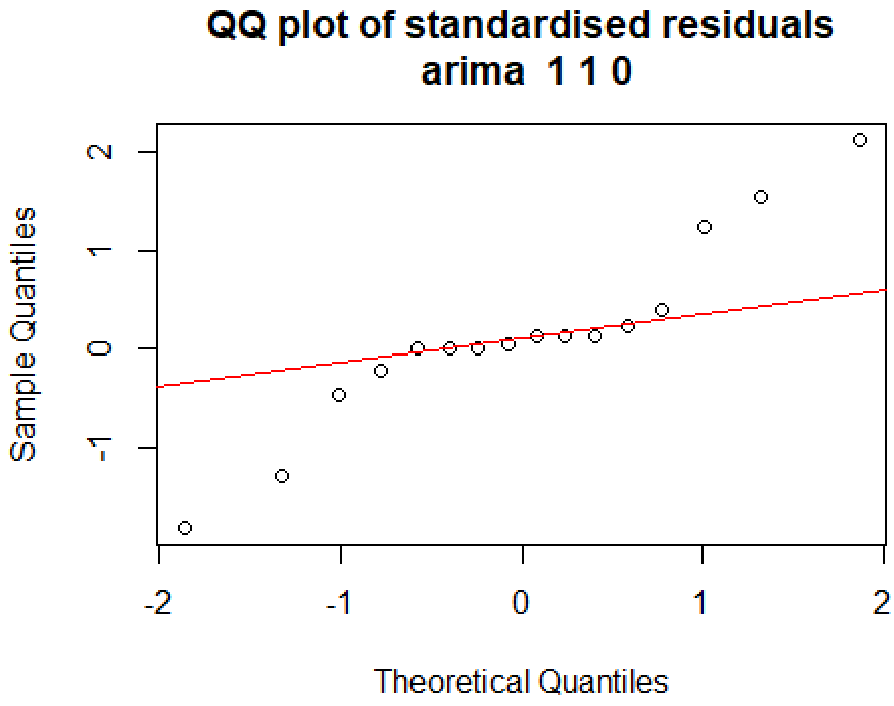


Figure 43- Q-Q Plot of Standard Residuals of ARIMA (1,1,0)

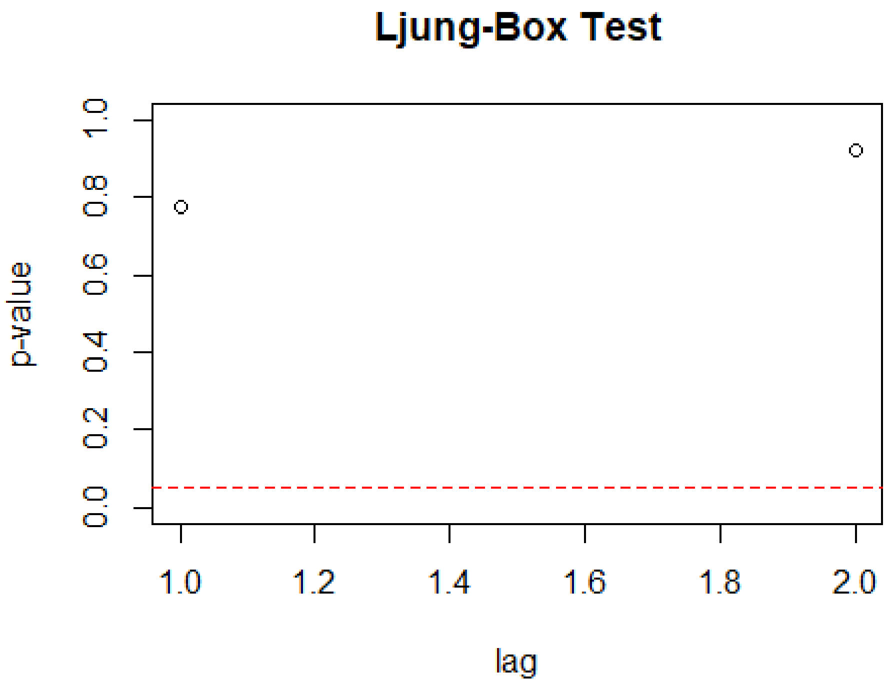
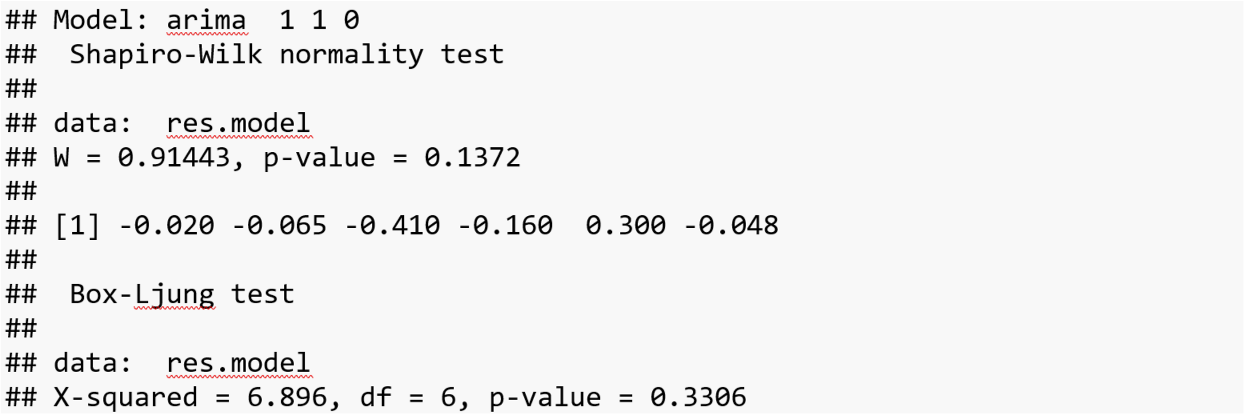


Figure 44 - Ljung Box Plot of Standard Residuals of ARIMA (1,1,0)



```
## Model: arima  1 1 0
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.91443, p-value = 0.1372
##
## [1] -0.020 -0.065 -0.410 -0.160  0.300 -0.048
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 6.896, df = 6, p-value = 0.3306
```

Table 30- Shapiro Normality and Ljung test for ARIMA(1,1,0)

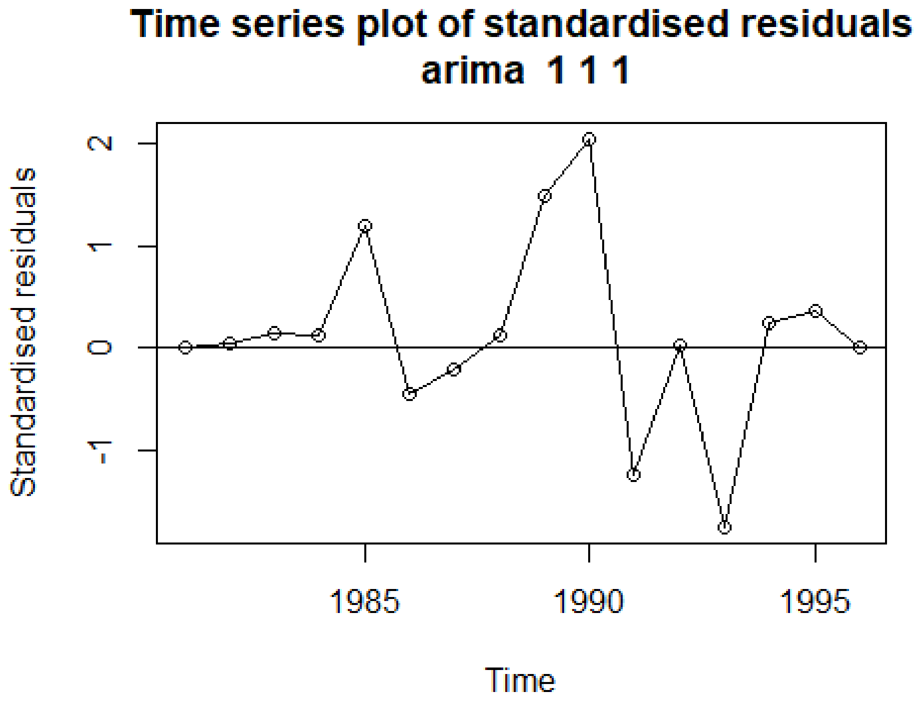


Figure 45 - Time Series Plot of Standard Residuals of ARIMA (1,1,1)

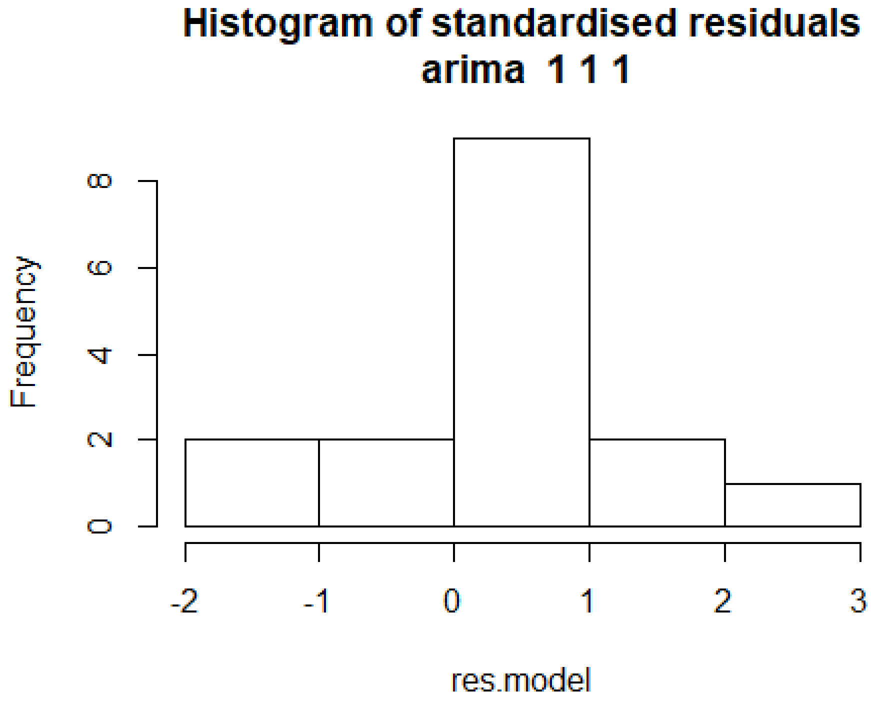


Figure 46 - Histogram of Standard Residuals of ARIMA (1,1,1)

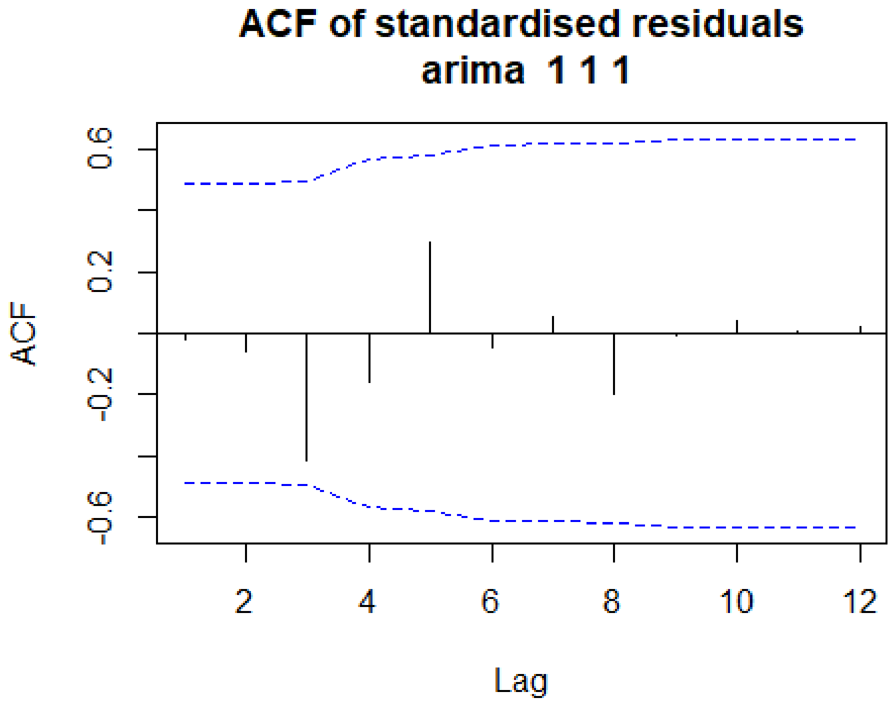


Figure 47- ACF Plot of Standard Residuals of ARIMA (1,1,1)

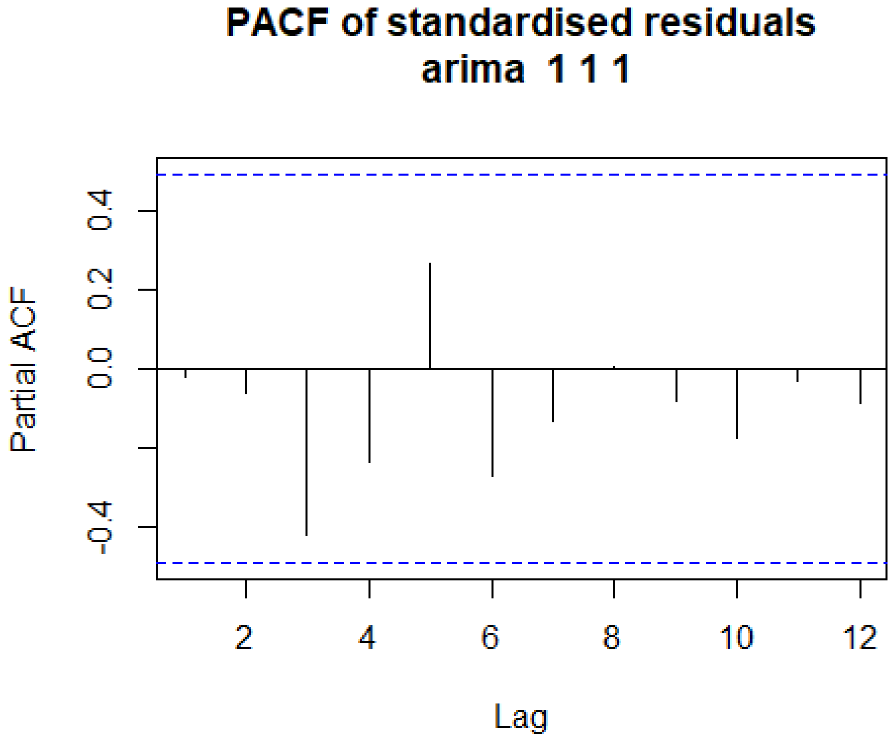


Figure 48 - PACF Plot of Standard Residuals of ARIMA (1,1,1)

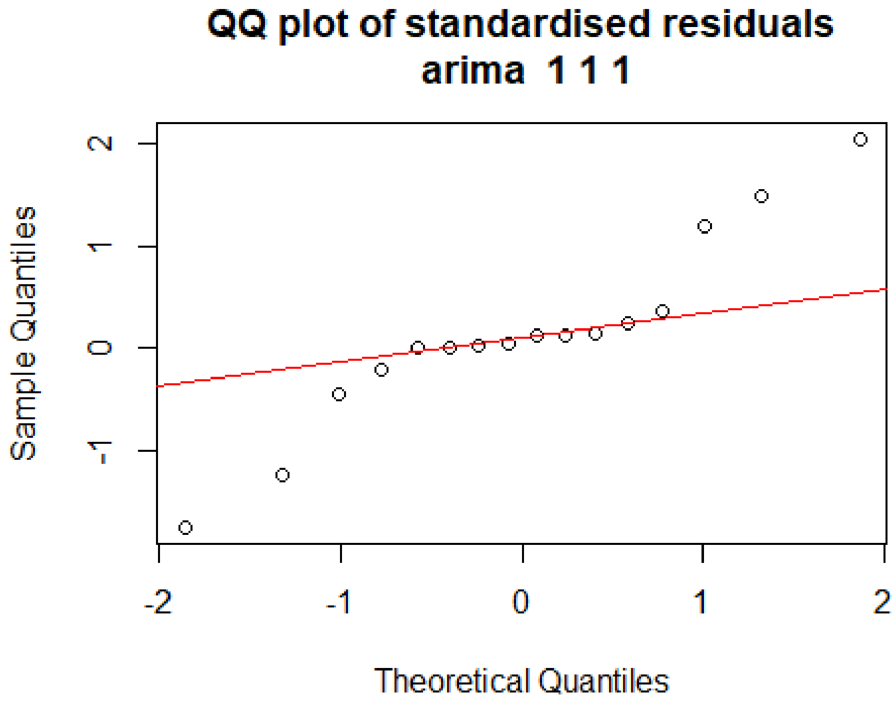


Figure 49- Q-Q Plot of Standard Residuals of ARIMA (1,1,1)

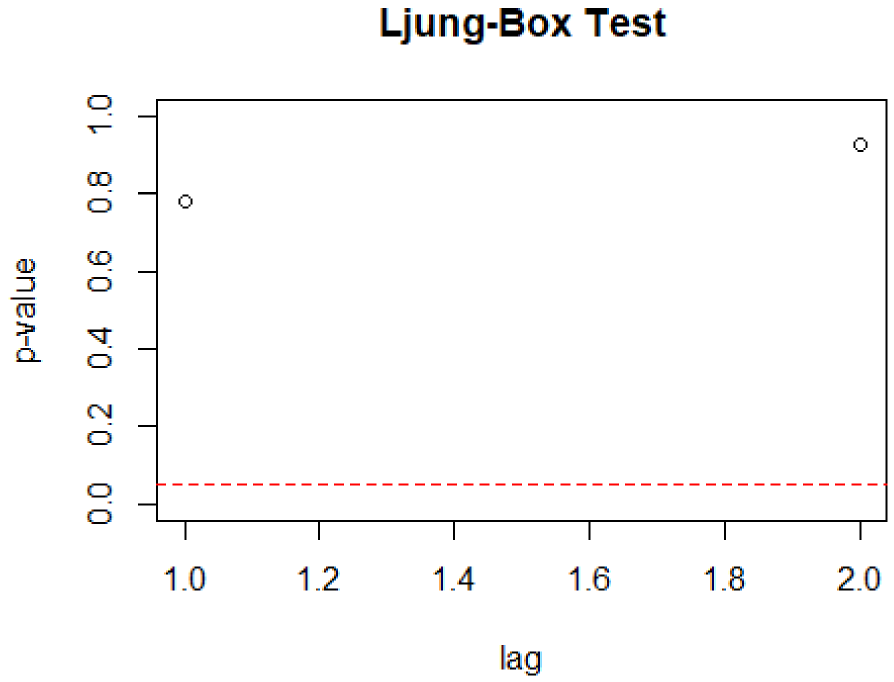


Figure 50 - Ljung Box Plot of Standard Residuals of ARIMA (1,1,1)

```
## Model: arima  1 1 1
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.91417, p-value = 0.1359
##
## [1] -0.022 -0.063 -0.410 -0.160  0.300 -0.049
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 6.8747, df = 6, p-value = 0.3326
```

Table 31-Shapiro Normality and Ljung test for ARIMA(1,1,1)

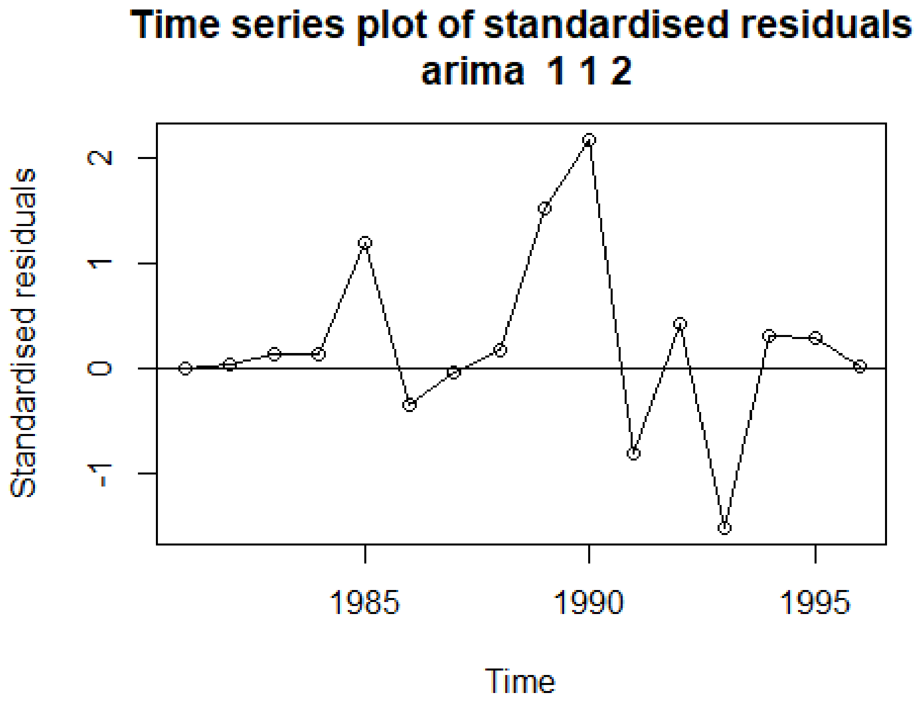


Figure 51-Time Series Plot of Standard Residuals of ARIMA (1,1,2)

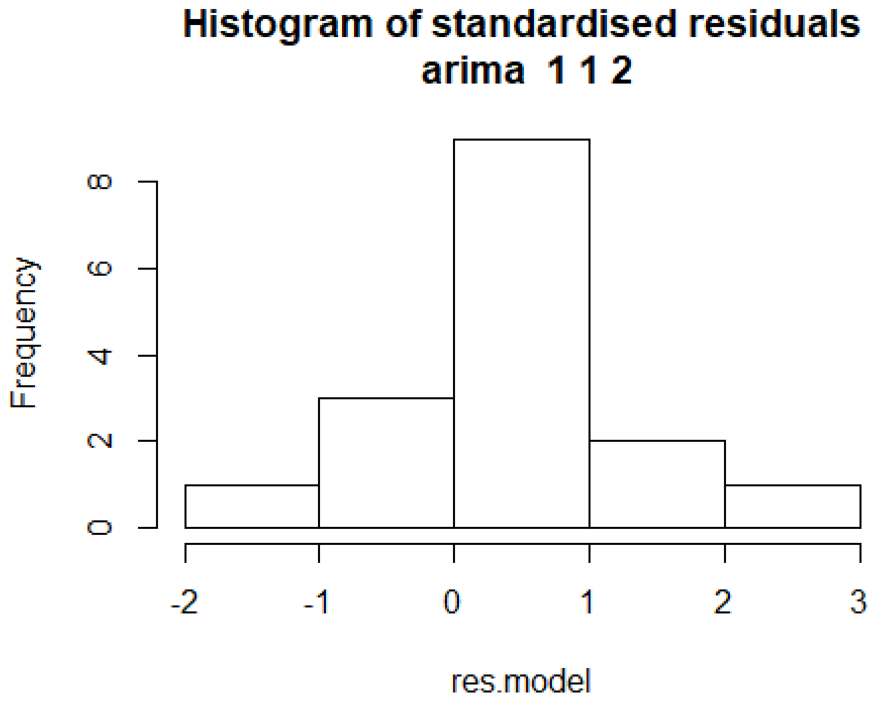


Figure 52- Histogram of Standard Residuals of ARIMA (1,1,2)

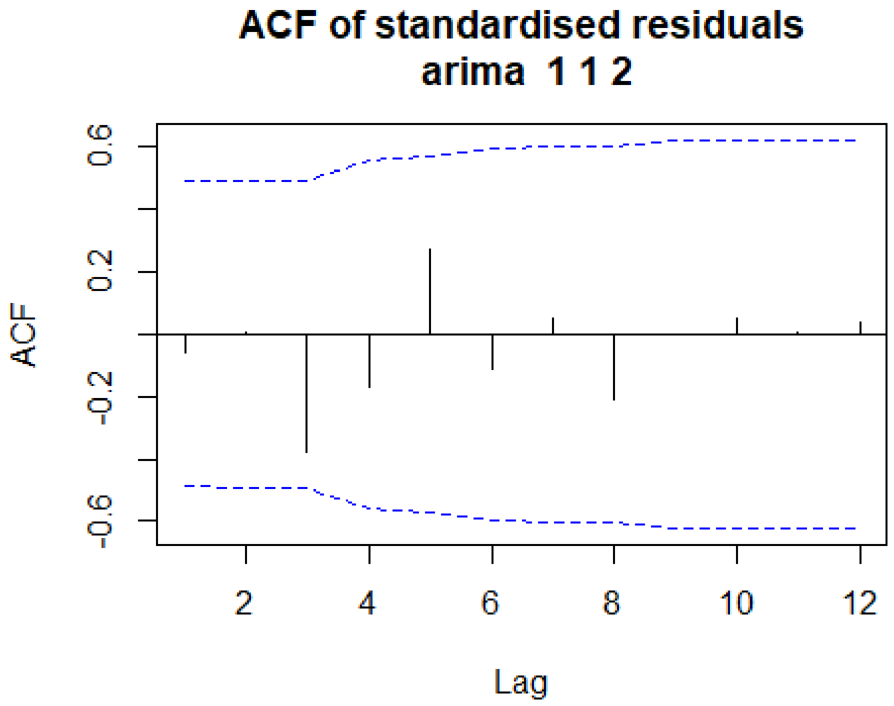


Figure 53- ACF Plot of Standard Residuals of ARIMA (1,1,2)

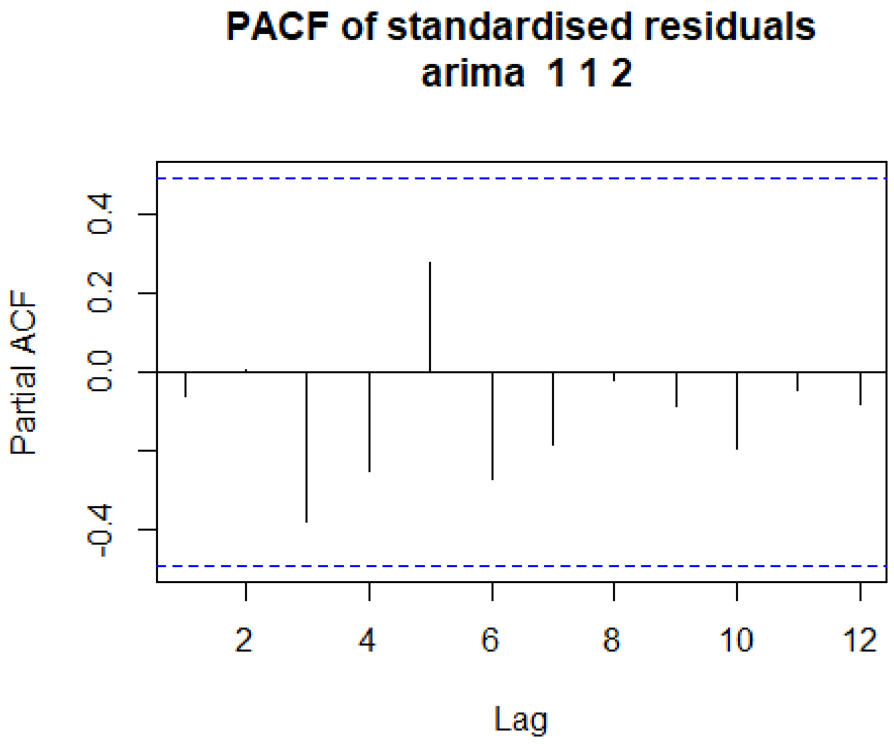


Figure 54- PACF Plot of Standard Residuals of ARIMA (1,1,2)

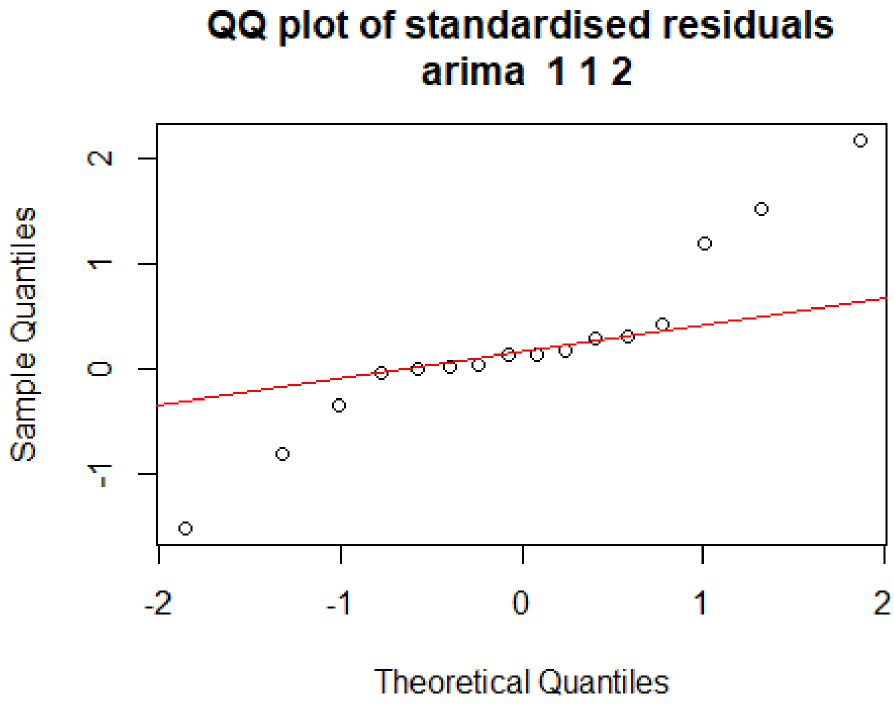


Figure 55 - Q-Q Plot of Standard Residuals of ARIMA (1,1,2)

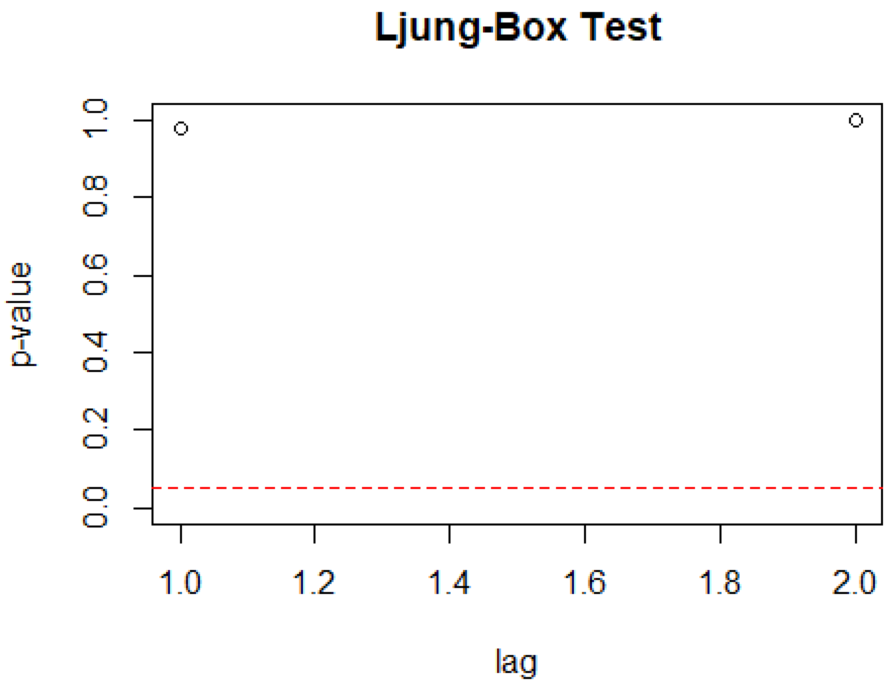
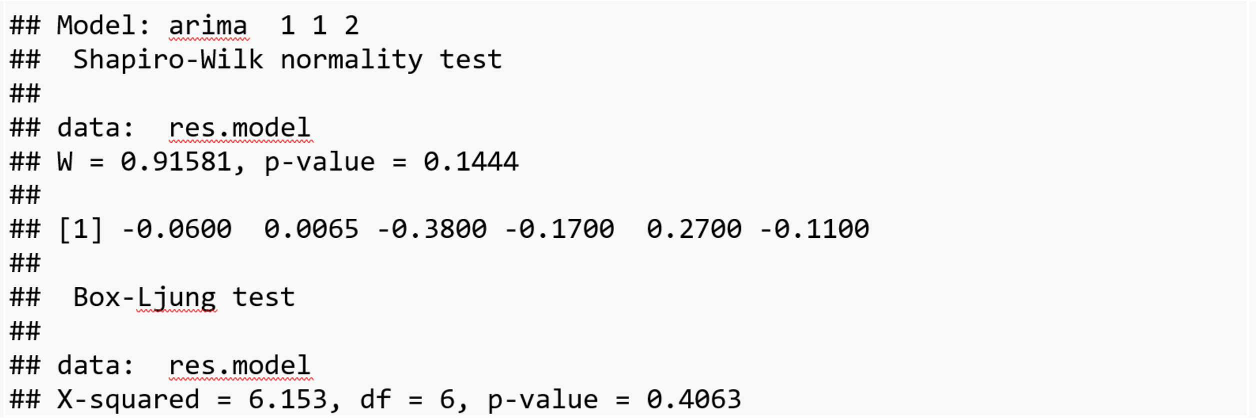


Figure 56 - Ljung Box Plot of Standard Residuals of ARIMA (1,1,2)



```
## Model: arima  1 1 2
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.91581, p-value = 0.1444
##
## [1] -0.0600  0.0065 -0.3800 -0.1700  0.2700 -0.1100
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 6.153, df = 6, p-value = 0.4063
```

Table 32- Shapiro Normality and Ljung test for ARIMA(1,1,2)

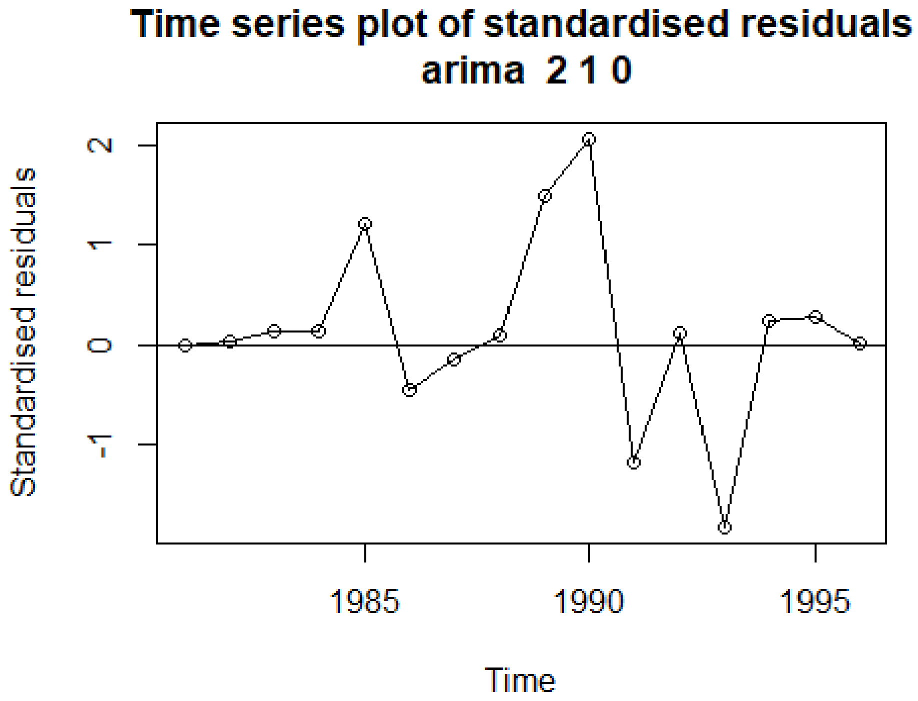


Figure 57- Time Series Plot of Standard Residuals of ARIMA (2,1,0)

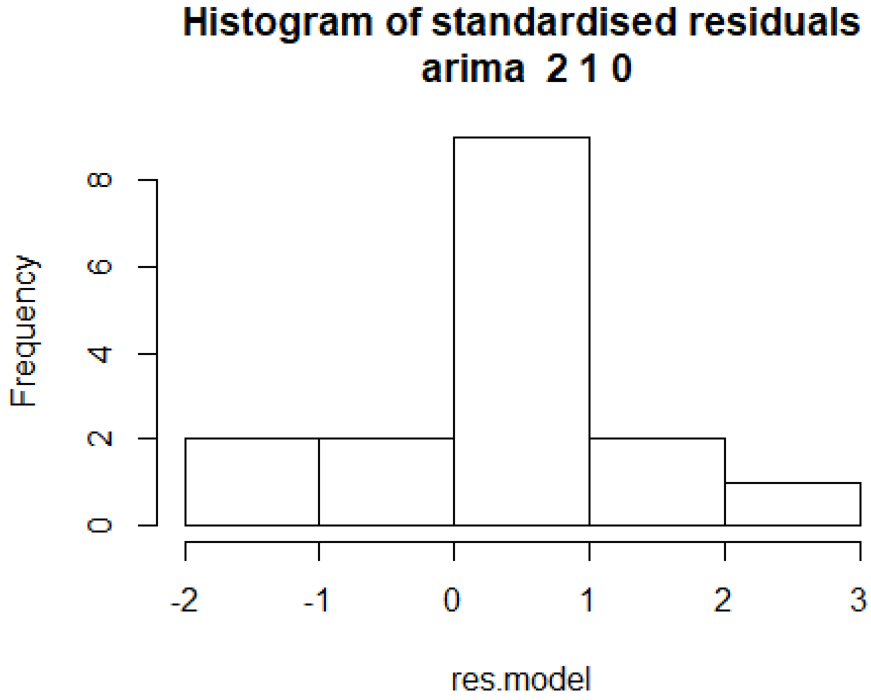


Figure 58 - Histogram of Standard Residuals of ARIMA (2,1,0)

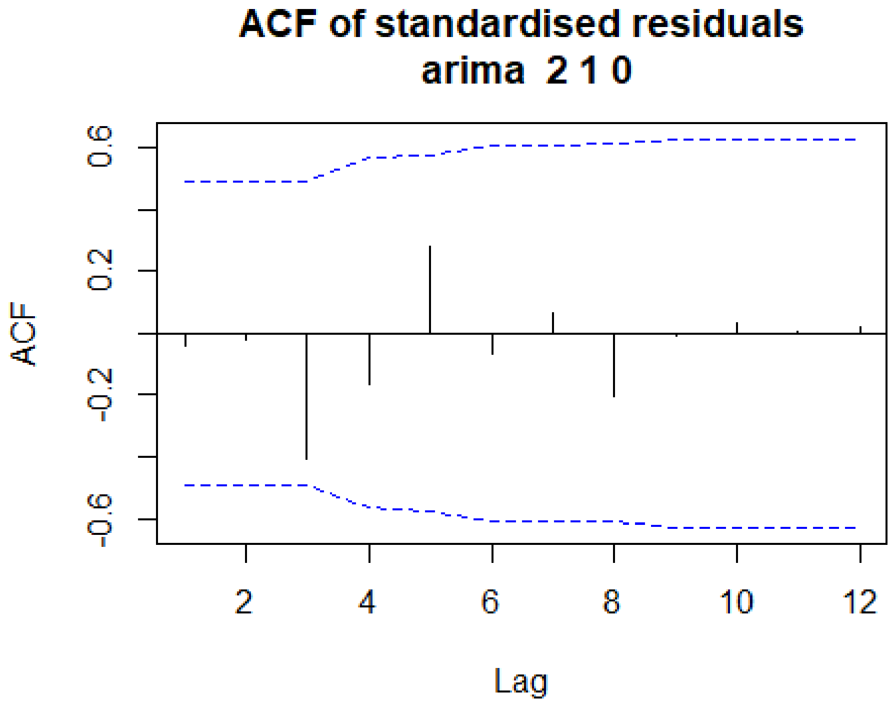


Figure 59 - ACF Plot of Standard Residuals of ARIMA (2,1,0)

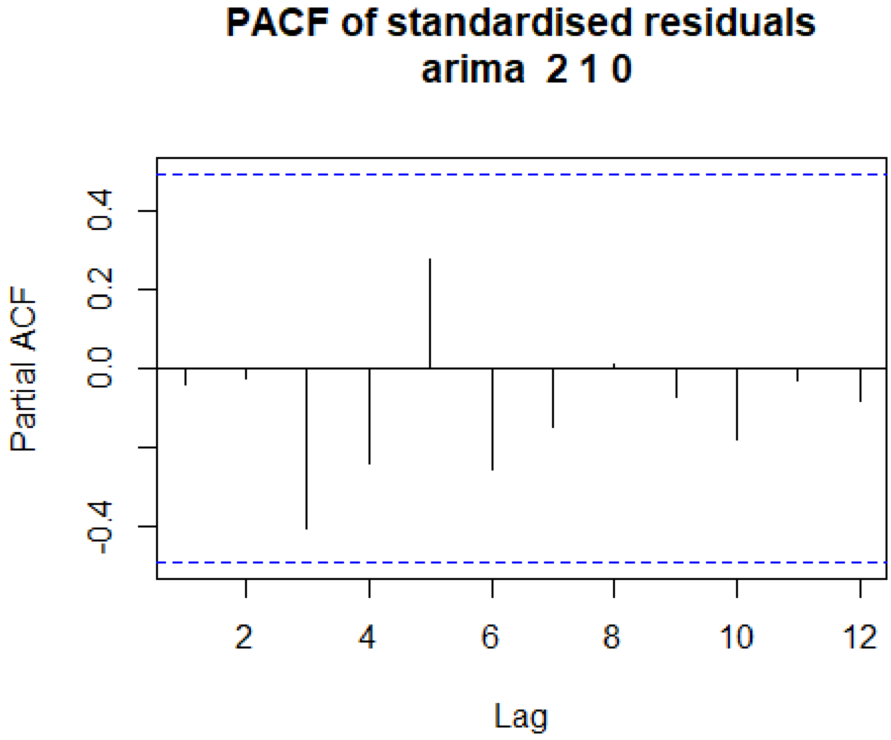


Figure 60 - PACF Plot of Standard Residuals of ARIMA (2,1,0)

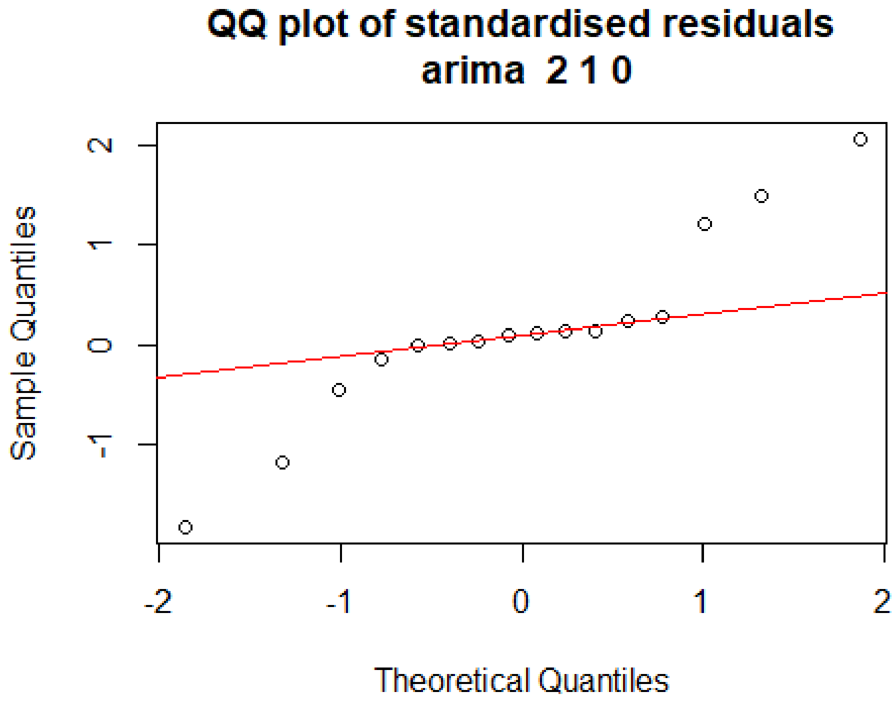


Figure 61- Q-Q Plot of Standard Residuals of ARIMA (2,1,0)

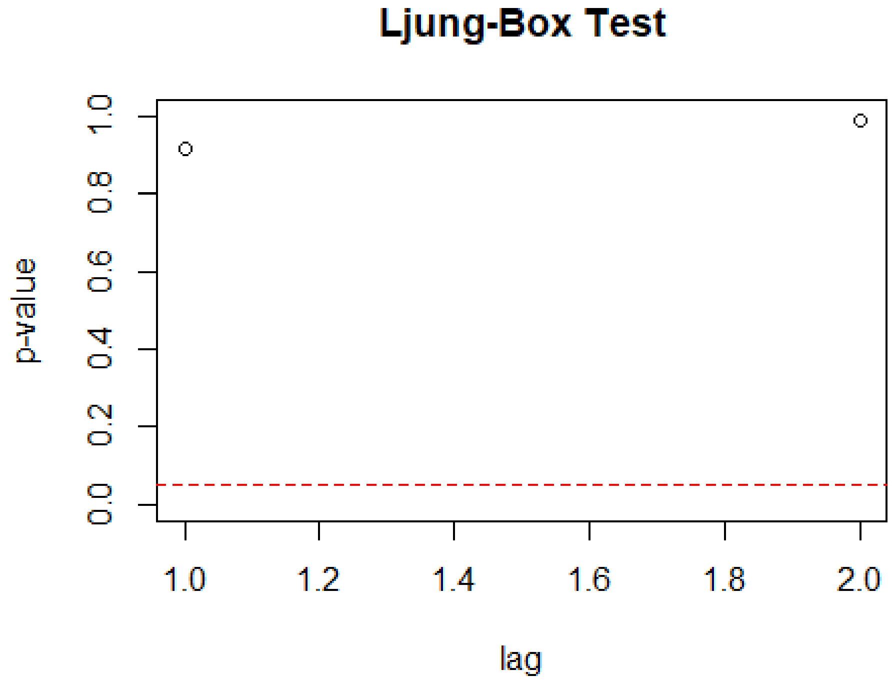
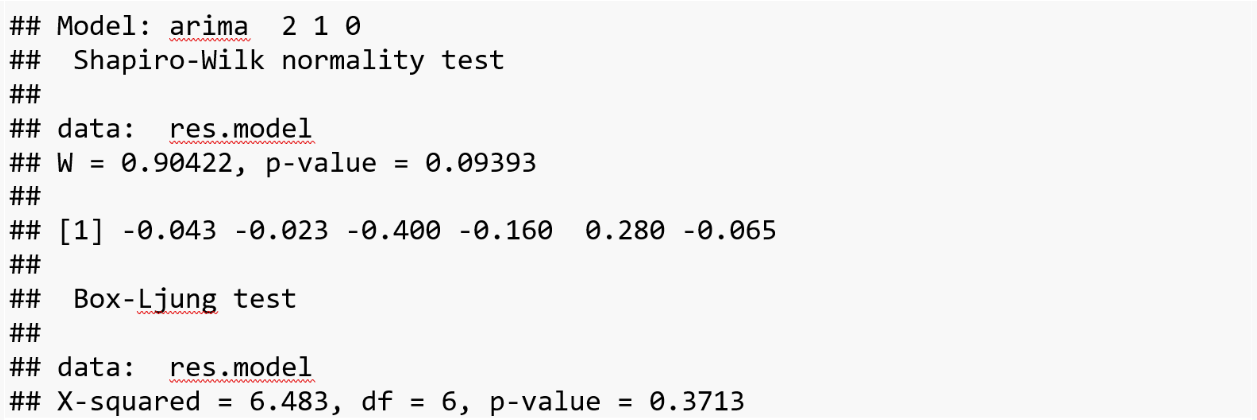


Figure 62 - Ljung Box Plot of Standard Residuals of ARIMA (2,1,0)



```
## Model: arima  2 1 0
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.90422, p-value = 0.09393
##
## [1] -0.043 -0.023 -0.400 -0.160  0.280 -0.065
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 6.483, df = 6, p-value = 0.3713
```

Table 33 - Shapiro Normality and Ljung test for ARIMA(2,1,0)

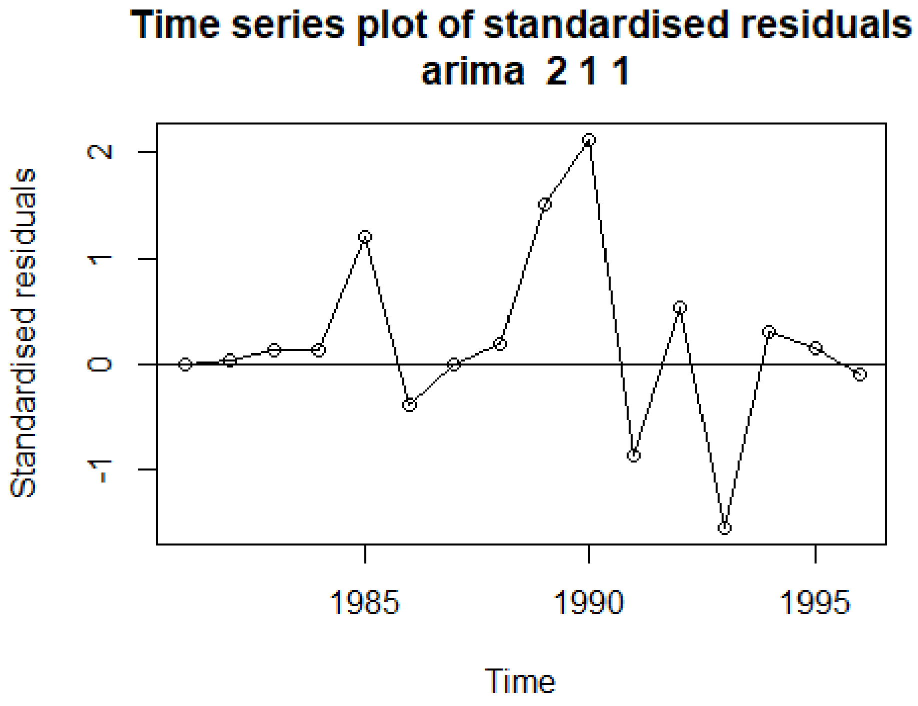


Figure 63- Time Series Plot of Standard Residuals of ARIMA (2,1,1)

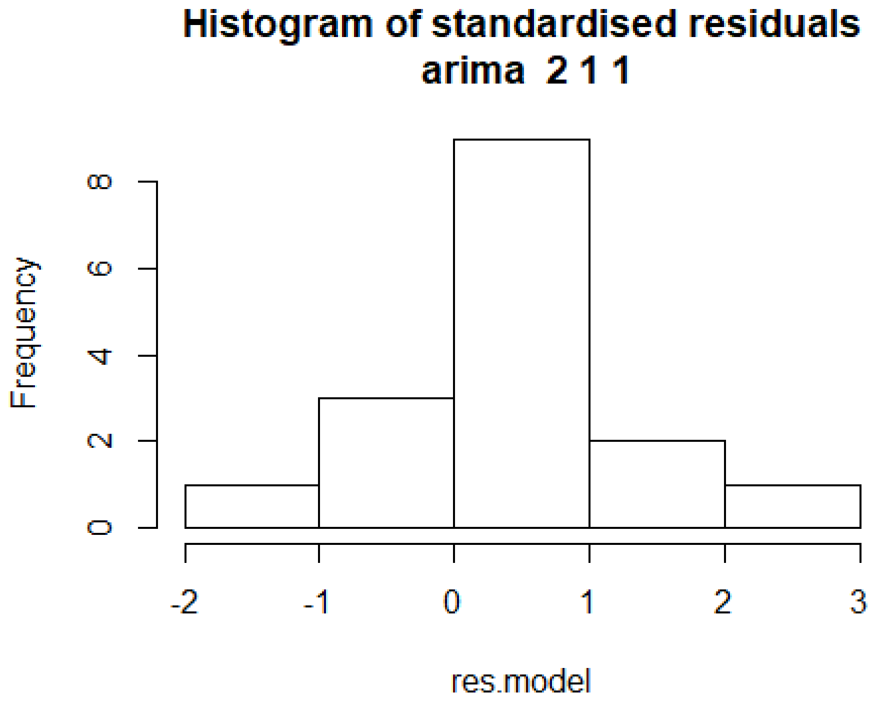


Figure 64 - Histogram of Standard Residuals of ARIMA (2,1,1)

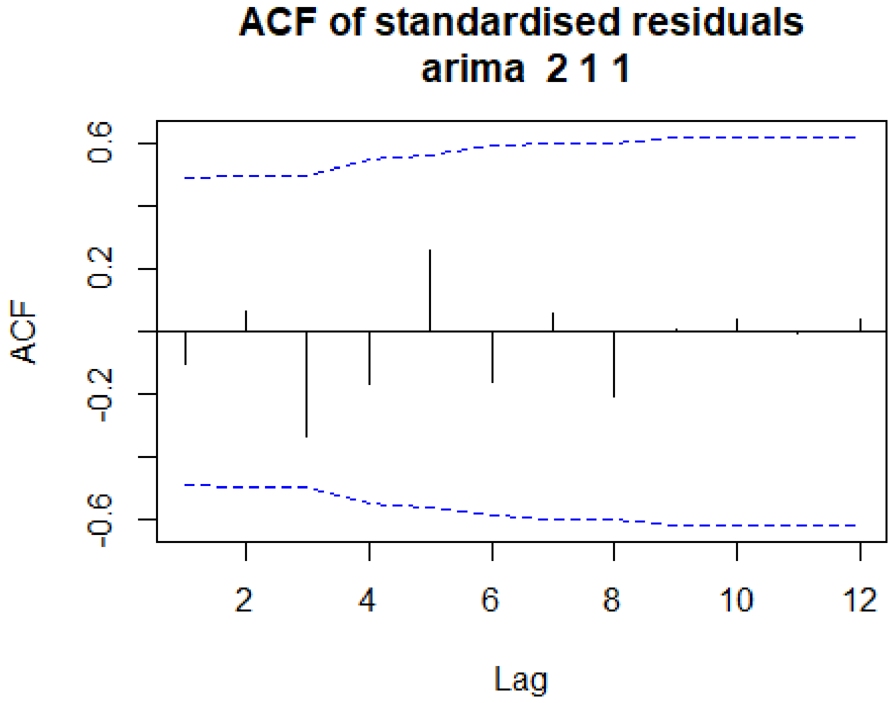


Figure 65 - ACF Plot of Standard Residuals of ARIMA (2,1,1)

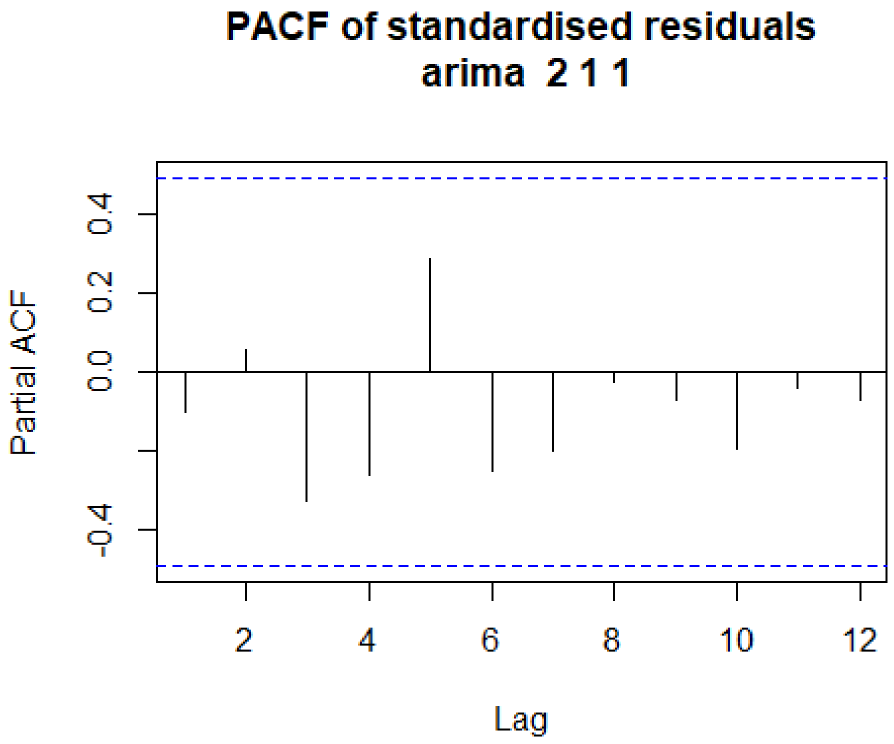


Figure 66 - PACF Plot of Standard Residuals of ARIMA (2,1,1)

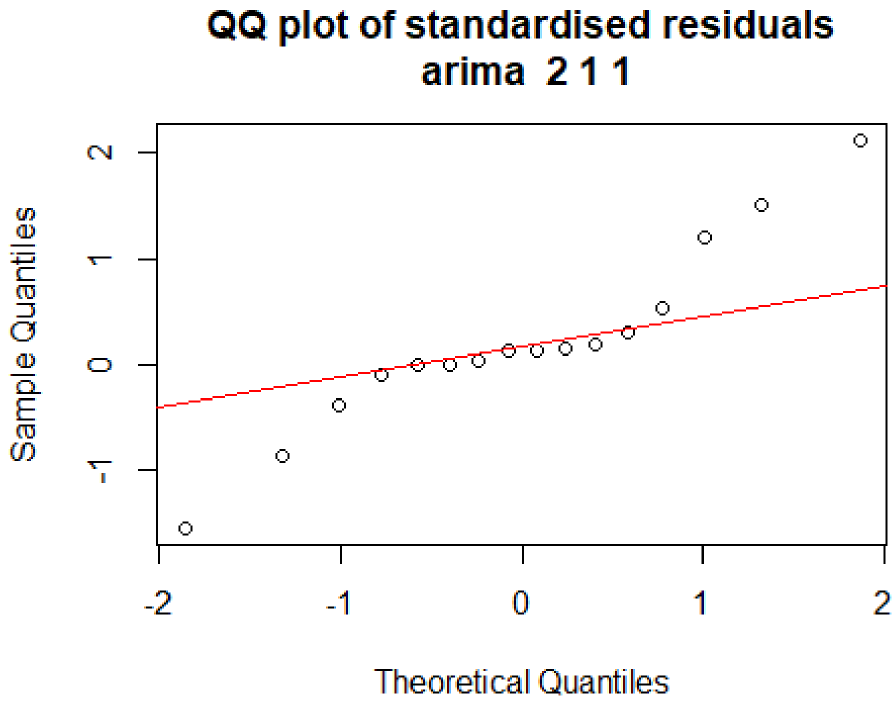


Figure 67- Q-Q Plot of Standard Residuals of ARIMA (2,1,1)

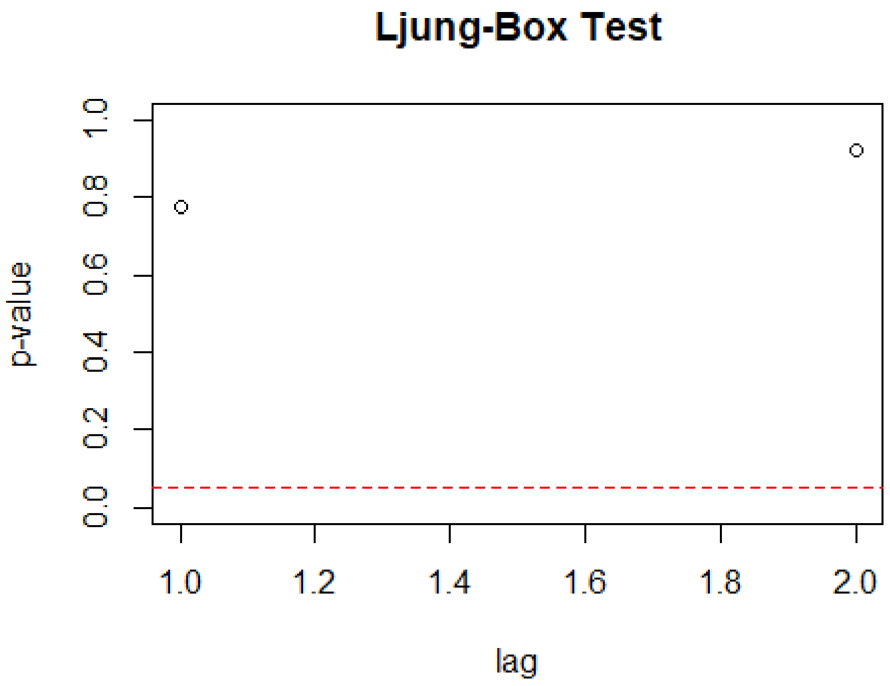
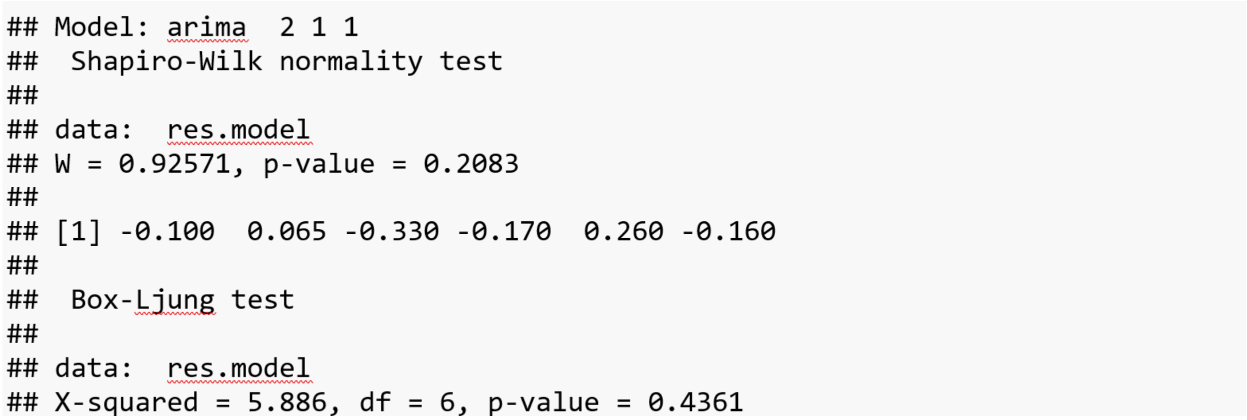


Figure 68 - Ljung Box Plot of Standard Residuals of ARIMA (2,1,1)



```
## Model: arima  2 1 1
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.92571, p-value = 0.2083
##
## [1] -0.100  0.065 -0.330 -0.170  0.260 -0.160
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 5.886, df = 6, p-value = 0.4361
```

Table 34- Shapiro Normality and Ljung test for ARIMA(2,1,1)

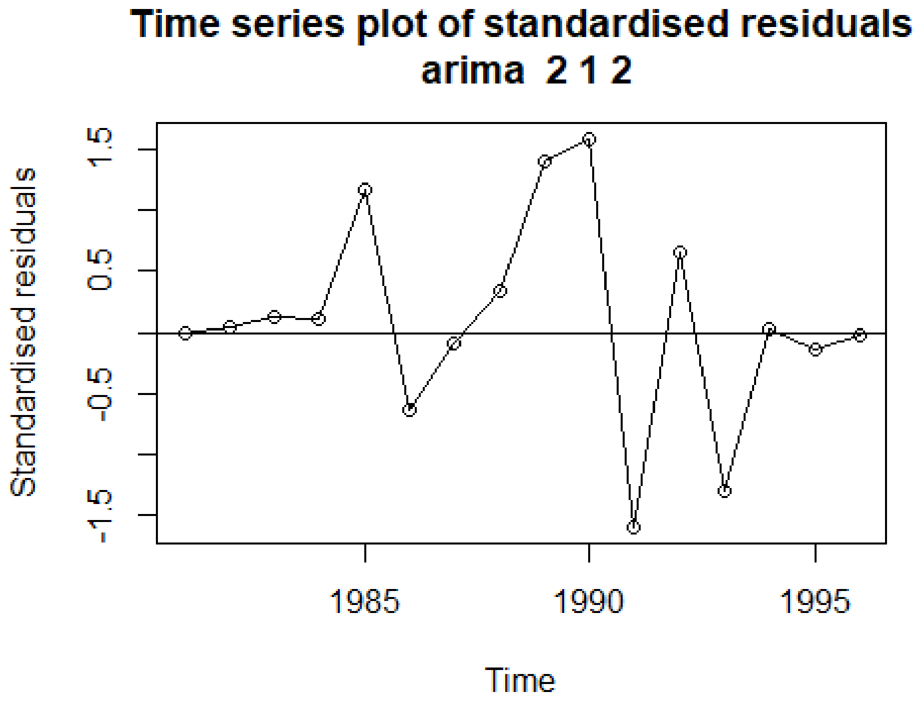


Figure 69- Time Series Plot of Standard Residuals of ARIMA (2,1,2)

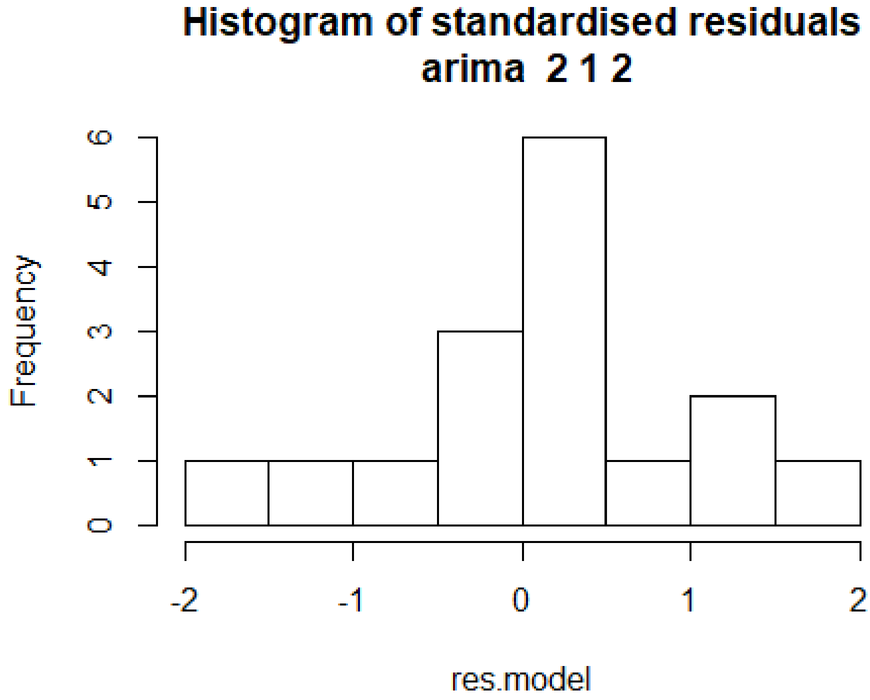


Figure 70 - Histogram of Standard Residuals of ARIMA (2,1,2)

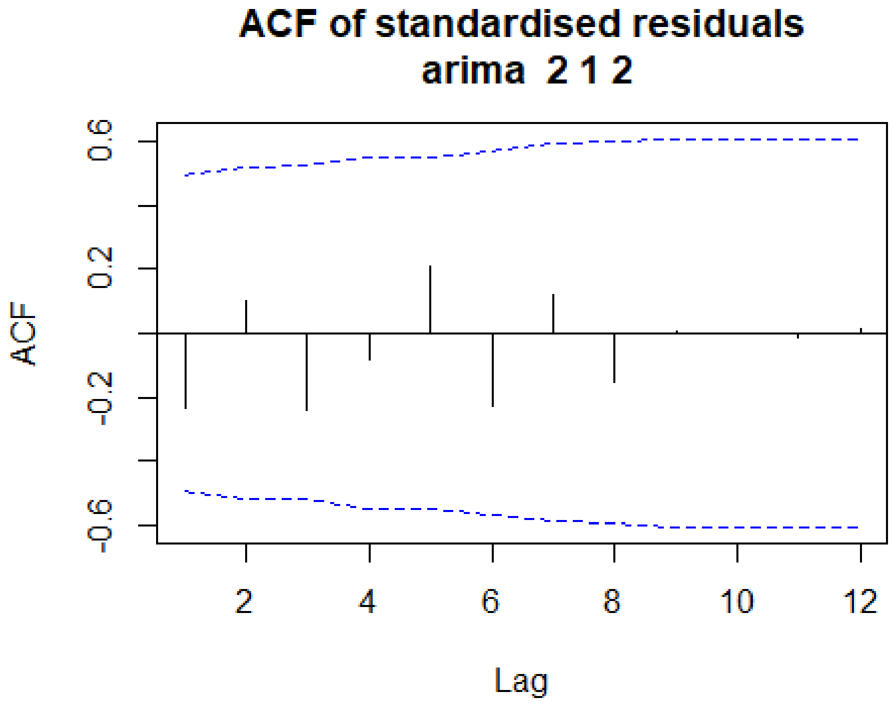


Figure 71- ACF Plot of Standard Residuals of ARIMA (2,1,2)

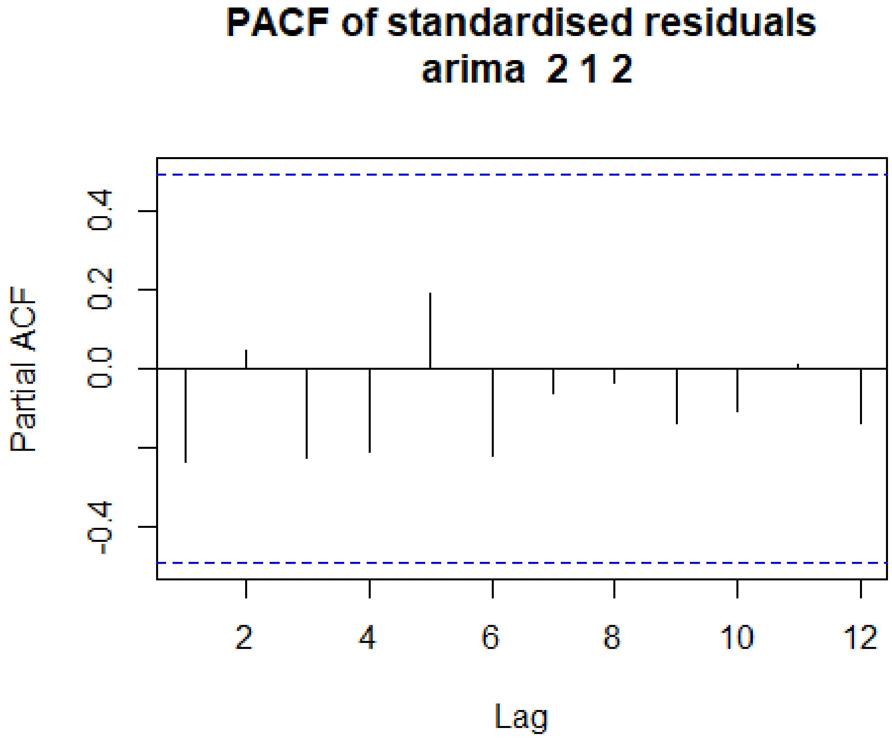


Figure 72- PACF Plot of Standard Residuals of ARIMA (2,1,2)

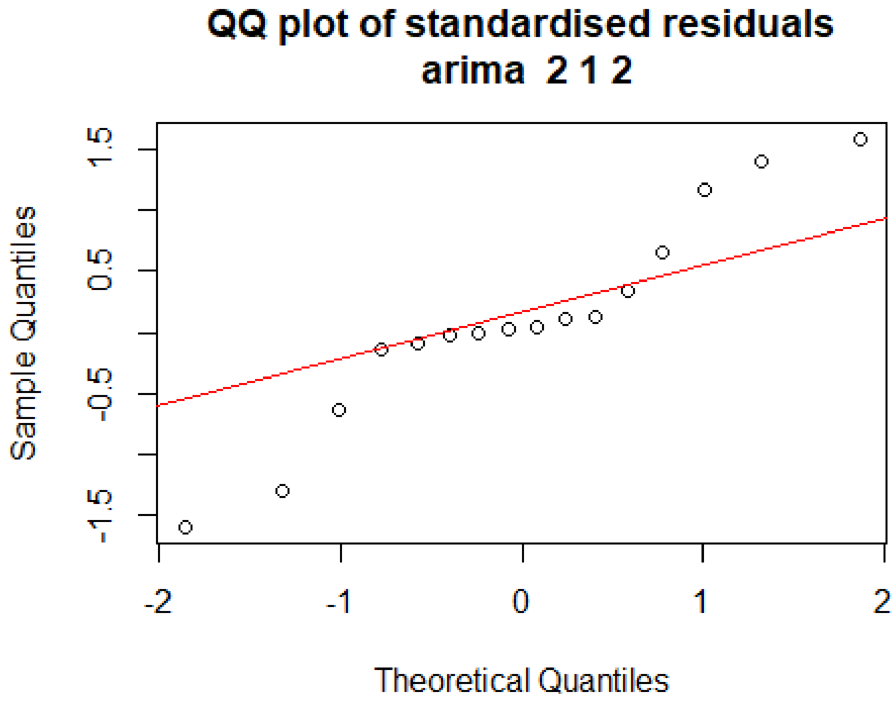


Figure 73-Q-Q Plot of Standard Residuals of ARIMA (2,1,2)

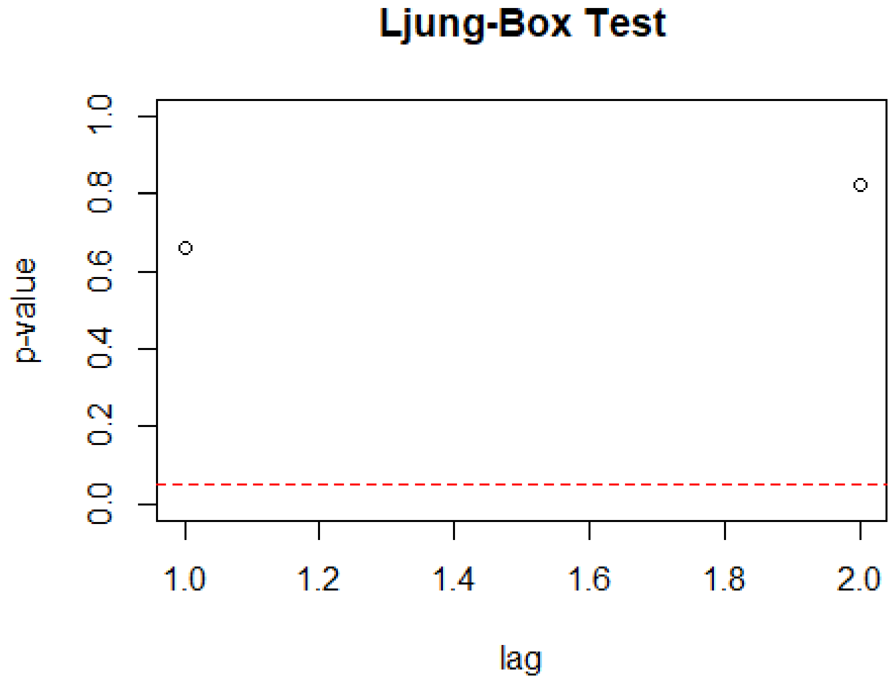


Figure 74 - Ljung Box Plot of Standard Residuals of ARIMA (2,1,2)

```
## Model: arima  2 1 2
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.93167, p-value = 0.2591
##
## [1] -0.230  0.099 -0.240 -0.085  0.210 -0.220
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 5.3376, df = 6, p-value = 0.5013
```

Table 35- Shapiro Normality and Ljung test for ARIMA(2,1,2)