MATH1318
TIME SERIES ANALYSIS

Semester 1, 2021

Group Assignment

Harold Li Guo Choi S3866530
Jessica Lin S3852191
Vanessa Jie En Yew S3817207

Table of Content

Introduction

Varicella, also known as Chickenpox, is a highly contagious and common childhood infection among people worldwide. In Hungary, it is reported to have 36,000 to 45,000 cases per year, however, this is likely to be an underestimated amount.

In most cases, the disease is self-limiting, however, it may lead to severe complications such as pneumonia, encephalitis, even in otherwise healthy people (Huber et al., 2020). Infants, elderlies and immunocompromised people are at higher risk of developing serious complications from Chickenpox. These complications may be lethal.

This project aims to comprehensively analyse the Chickenpox dataset by using the analysis methods covered in the Time Series Analysis course and accurately predict the series for the next 10 units of time. This report includes the descriptive analysis, proper visualization, model specification, model fitting and selection, diagnostic checking, and interpretation of the Hungary Chickenpox dataset.

Methods

The Hungary Chickenpox dataset is a .csv extension file that consists of weekly Chickenpox cases in Hungary between 2005 to 2015. It was sourced from the UCI Machine Learning Repository and loaded into R.

A Shapiro Test, Dickey-Fuller Unit-Root Test (ADF) and the Phillips-Perron Unit Root Test (PP) was performed and the chosen level of significance for these tests is 0.05.

The analysis was done with the aid of the TSA library, implementing the ACF, PACF, ARIMA, BoxCox.ar, harmonic, predict, season, zlag and armasubsets function. The analysis also uses the adf.test and the pp.test function from the tseries library.

Functions from the lubridate and dpylr are used for the preprocessing of the data.

The CV function from the fpp library was also used to determine the model's predictive accuracy and base R functions such as qqnorm, qqline, hist, rstudent, shapiro.test, plot, cor, abline, diff and summary were also used for the analysis. The results are then further analyzed to provide insightful information.

Results

The Hungary Chickenpox dataset is aggregated across all cities in Hungary by year. Research has found that chickenpox peaks during the spring season (Rettner, 2019). Therefore, the dataset is further preprocessed and aggregated from weekly data into monthly data. Monthly data is more informative when determining the seasonality of the disease. The following are the results.

Data Exploration

As the processed Hungary Chickenpox dataset only contains one variable, the time series model would include only the number of chickenpox cases in Hungary over time.

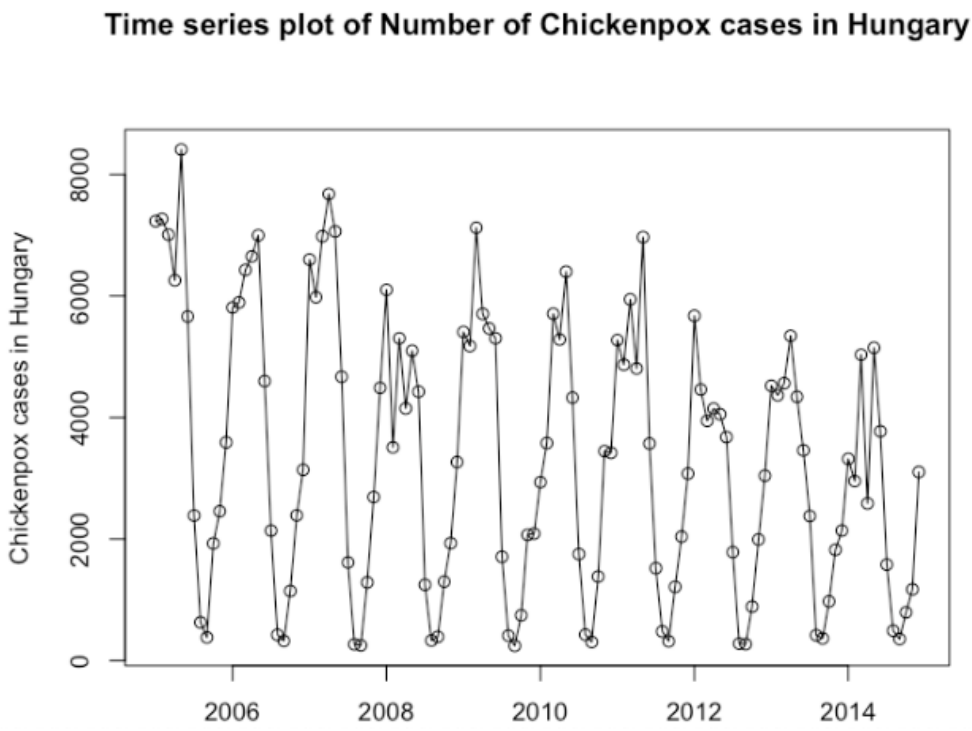

Figure 1 – Time series plot of the monthly number of chickenpox cases in Hungry

R code:

```
cpts <- ts(cps[,3],start=c(2005,1), end=c(2014,12), frequency=12)
plot(cpts ,ylab='Chickenpox cases in Hungary',xlab='Year',type='o',
     main = "Time series plot of Number of Chickenpox cases in Hungary")
```

Figure 1 displays the plot of the total number of monthly cases of chickenpox in Hungary from 2005 to 2015. The following properties are deduced from the chickenpox series:

1. Trend: No trend or a slight downward trend is observed in the series.
2. Changing Variance: There are hints of decreasing variance as the size of fluctuations appear to decrease with time.
3. Behaviour: The plot suggests both Autoregressive (AR) and Moving Average (MA) behaviours as succeeding time points are similar in value. Large changes in the number of chickenpox cases can also be seen occurring from one month to the next in certain months.
4. Change Point: There is no existence in a change point
5. Seasonality: Seasonality seemed to be apparent in the time series plot.

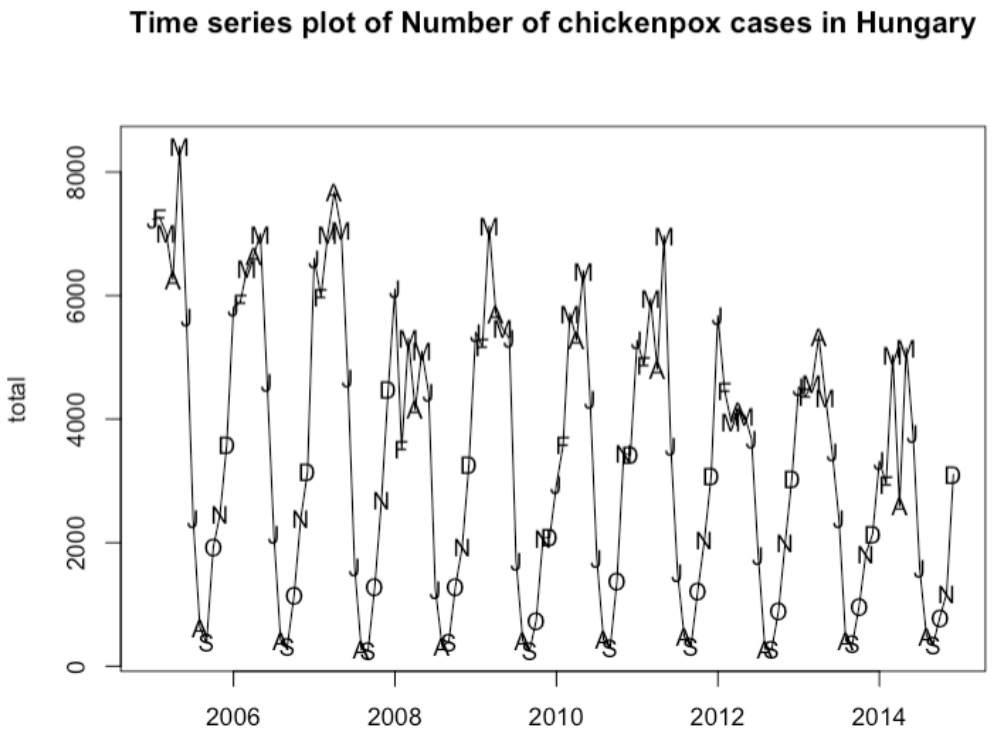

Figure 2 – Seasonal Time series plot of the monthly number of chickenpox cases in Hungry

R code:

```
#look out for seasonlity
plot(cpts ,type='l', main = "Time series plot of Number of chickenpox cases in Hungary")
points(y=cpts ,x=time(cpts), pch=as.vector(season(cpts)))
```

The labelled data points in Figure 2 suggests there is seasonality in the series as corresponding months have similar values. The number of chickenpox cases exponentially increases up from December to February and eventually peaks during May and troughs during August and September. Given seasonality can mask the true nature of the autoregressive or moving average behaviour of a series, it can not be definitively determined from this plot alone and therefore further exploration is required.

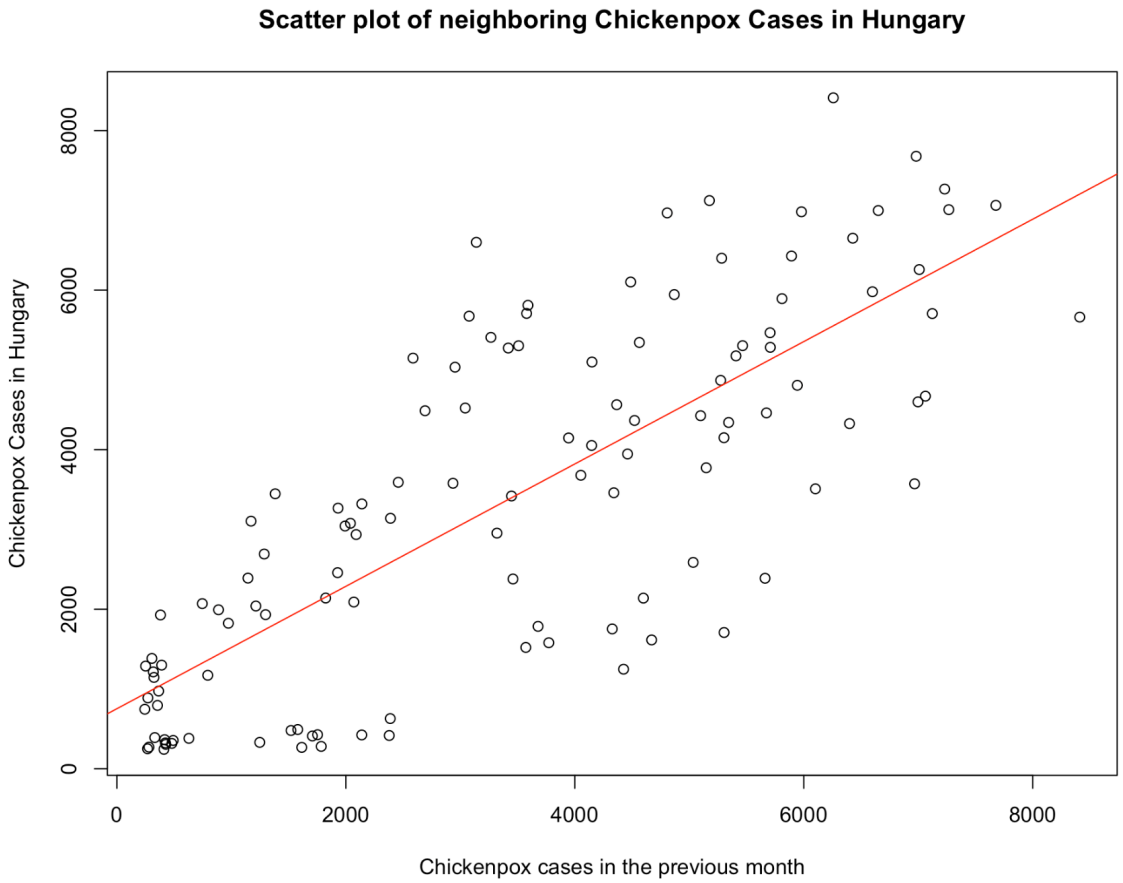

Figure 3 – Scatter plot of the number of chickenpox cases in consecutive months

R code:

```
#impact of previous month's cases on next month's cases:
y= cpts
x = zlag(cpts)

index = 2:length(x) # Create an index to get rid of the first NA value in x
cor(y[index],x[index]) #high positive correlation:  0.7769839

plot(y=cpts,x=zlag(cpts),ylab='Chickenpox Cases in Hungary', xlab= 'Chickenpox cases in the previous month',
     main= "Scatter plot of neighboring Chickenpox Cases in Hungary")
abline(lm(cpts~zlag(cpts)), col= "red")
```

Figure 3 shows a strong upward trend between the number of chickenpox cases and its lag of one month. This implies a strong autocorrelation between the neighbouring monthly changes in the number of chickenpox cases. The result is further evident by the strong correlation of 0.78 between the number of chickenpox cases and its succeeding month.

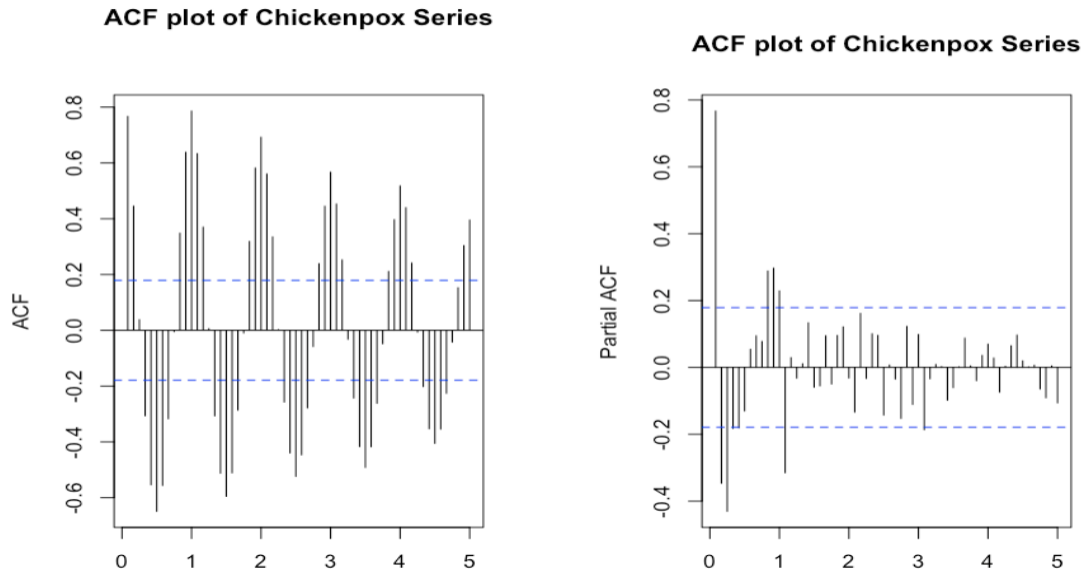

Figure 4 – ACF/ PACF plot of the Chickenpox series

R Code:

```
par(mfrow=c(1,2))
acf(cpts,lag.max = 12*5,main = "ACF plot of Chickenpox Series")
pacf(cpts,lag.max = 12*5, main = "ACF plot of Chickenpox Series")
par(mfrow=c(1,1))
```

In Figure 4, a slowly decaying sinusoidal pattern and significant seasonal lags are observed on the ACF plot. Additionally, a high correlation in the first lag of the PACF is also seen. These indicate the existence of seasonality and trend, making the series non-stationary.

The Dickey-Fuller Unit-Root Test is then used to test the null hypothesis if the process is differenced nonstationary.

```
        Augmented Dickey-Fuller Test

data:  cpts
Dickey-Fuller = -9.2913, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Figure 5– ADF Test of the Chickenpox Series

R Code:

```
adf.test(cpts) #confirms stationarity.
```

The p-value of the ADF test gives 0.01, therefore we reject the null hypothesis at the 95% confidence interval that the Chickenpox series is nonstationary.

```
        Phillips-Perron Unit Root Test

data:  cpts
Dickey-Fuller Z(alpha) = -41.538, Truncation lag parameter = 4, p-value = 0.01
alternative hypothesis: stationary
```

Figure 6– PP Test of the Chickenpox Series

R Code:

```
pp.test(cpts) #confirms stationarity.
```

The Phillips-Perron Unit Root Test (PP) rejects the null hypothesis at the 95% confidence interval that the series is nonstationary.

The ADP and PP test performed on the series both returned p-values less than alpha, rejecting the null hypothesis which states that the series is non-stationary. However, the ACF/PACF as seen in Figure 4 displays a sinusoidal decaying seasonal trend, contradicting the ADF and PP test results. Consolidating the information gained from the descriptive analysis of the time series and the ACF plot, it can be concluded that the series is stochastic with seasonal characteristics.

With hints of decreasing variance as seen in Figure 1, a box-cox transformation is performed to determine if it can remove the changing variance.

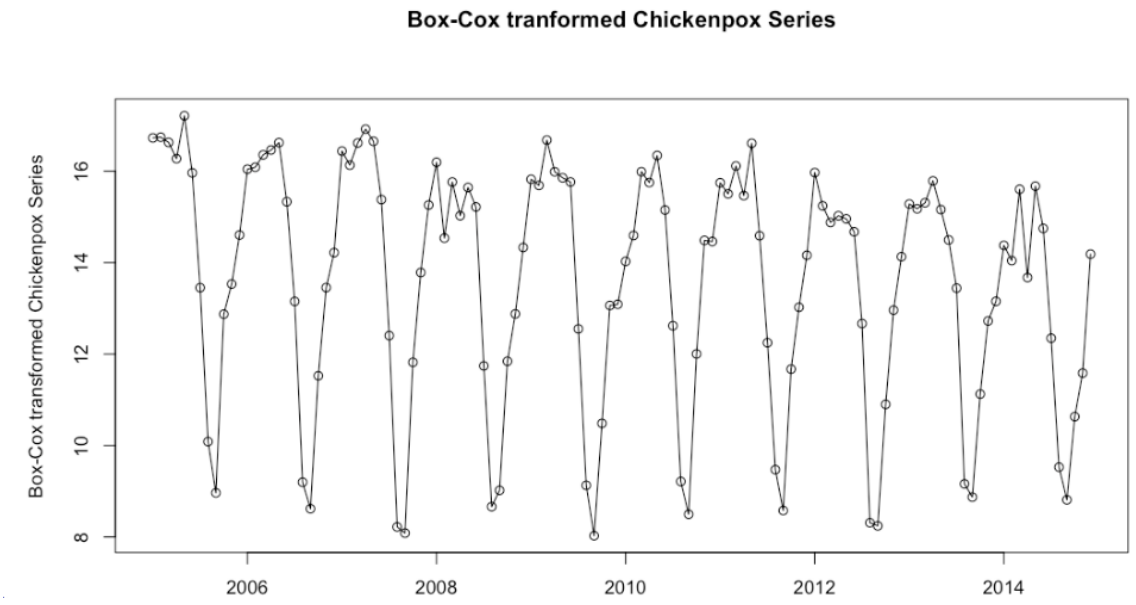

Figure 7– Time series plot of the box cox transformed Chickenpox Series

R Code:

```
#boxcox transformation
options(warn=-1)
BC <- BoxCox.ar(y = cpts , lambda = seq(-2, 2, 0.01))
# To find the optimal lambda value
lambda <- BC$lambda[which(max(BC$loglike) == BC$loglike)]
# Apply Box-Cox transformation
cptss  <- ((cpts ^lambda) - 1) / lambda
par(mfrow=c(1,2))
plot(cptss , ylab='Box-Cox transformed Chickenpox Series', xlab='Time', type='o',
     main = "Box-Cox tranformed Chickenpox Series")
```

From Figure 7, it could be seen that the variance of the series seemed to stabilize across time by the box-cox transformation.

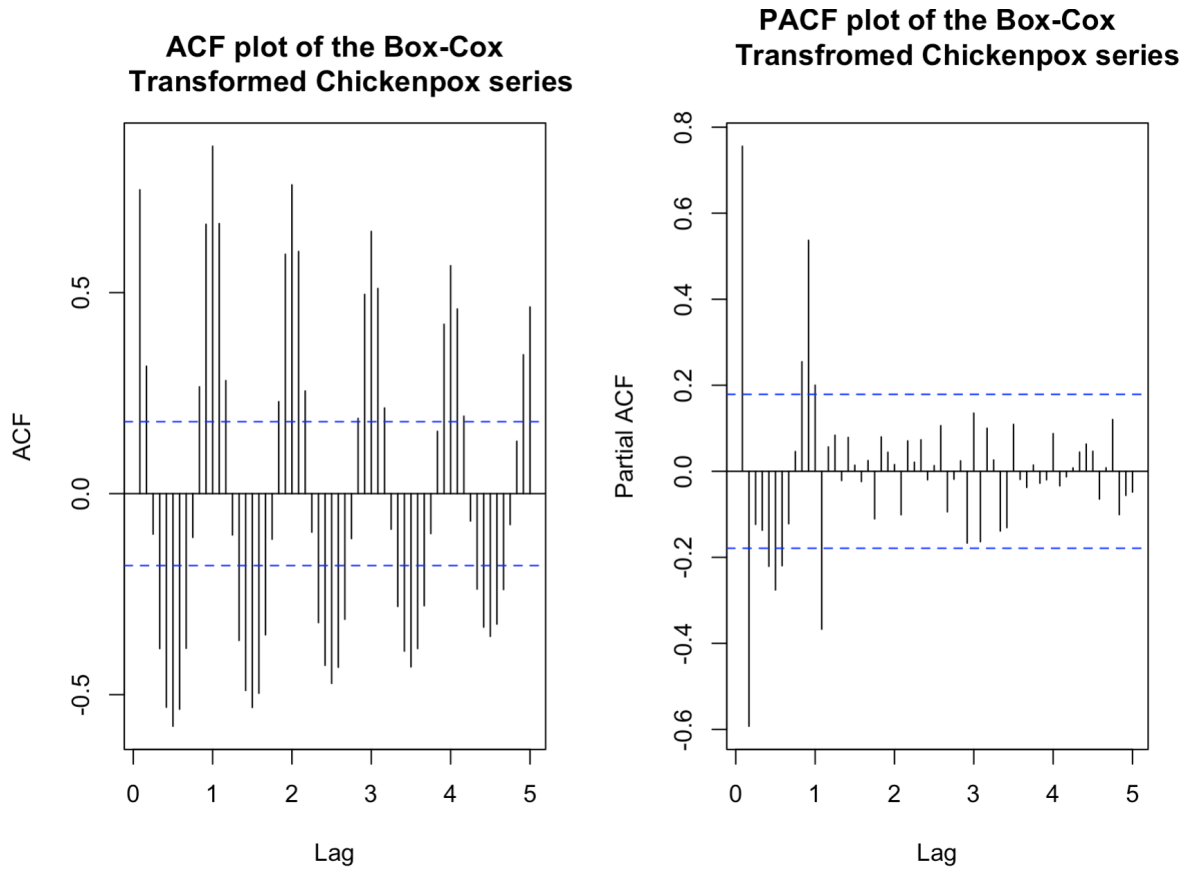

Figure 8– ACF/PACF plot of the Box-Cox transformed Chickenpox Series

R Code:

```
par(mfrow=c(1,2))
acf(cptss,lag.max = 12*5,main = "ACF plot of the Box-Cox
    Transformed Chickenpox series")
pacf(cptss,lag.max = 12*5, main = "PACF plot of the Box-Cox
     Transfromed Chickenpox series")
par(mfrow=c(1,1))
```

However, the ACF plot of the box-cox transformed series still showed a slowly decaying sinusoidal pattern with significant seasonal lags. There is also a high correlation in the first lag of the PACF. This indicates the existence of a trend and seasonality.

Model Specification

To handle the seasonality, the residual analysis approach was used to determine the order of P, D and Q. The series will be seasonally differenced by fitting a SARIMA(0,0,0)x$(0,1,0)_{12}$.

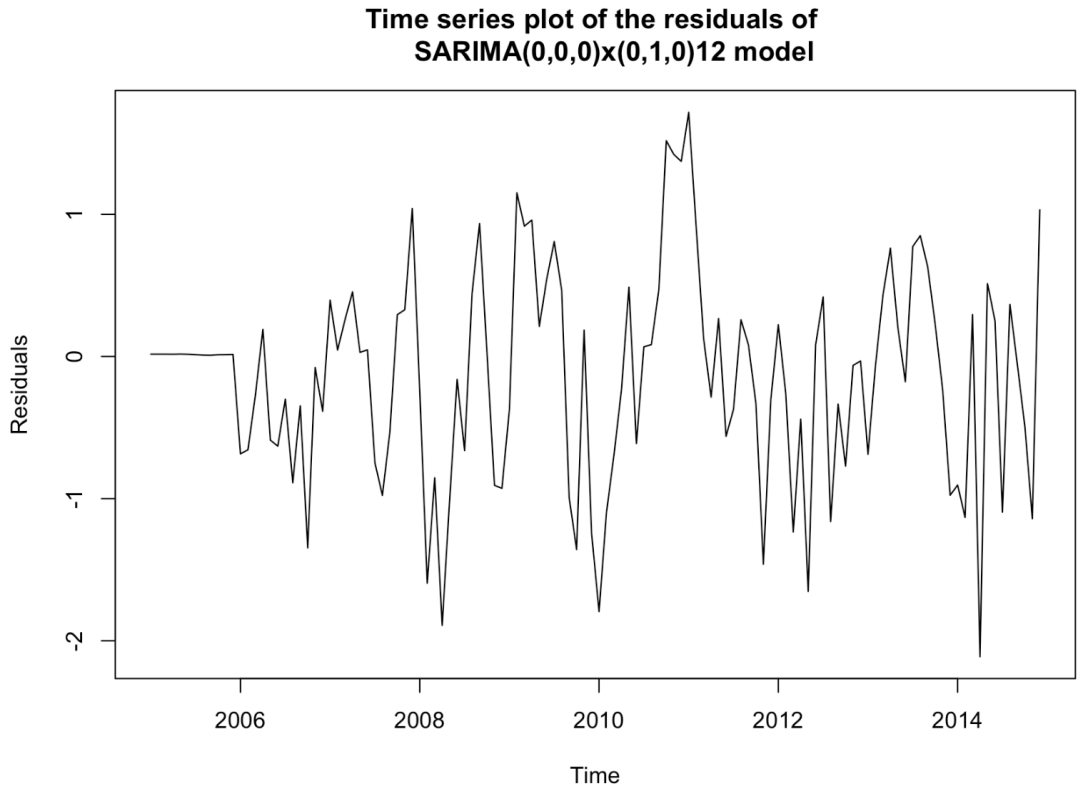

Figure 9– Time series plot of seasonally differenced residual

R Code:

```
m1 = Arima(cptss,order=c(0,0,0),seasonal=list(order=c(0,1,0), period=12))
res.m1 = residuals(m1);
par(mfrow=c(1,1))
plot(res.m1,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
     SARIMA(0,0,0)x(0,1,0)12 model")
```

The residual plot of the seasonally differenced model looks stationary however it does not look random. It suggests that the model did not capture some of the deterministic components. However, the standardised residuals are between the values of -3 and 3, thus suggesting normality.

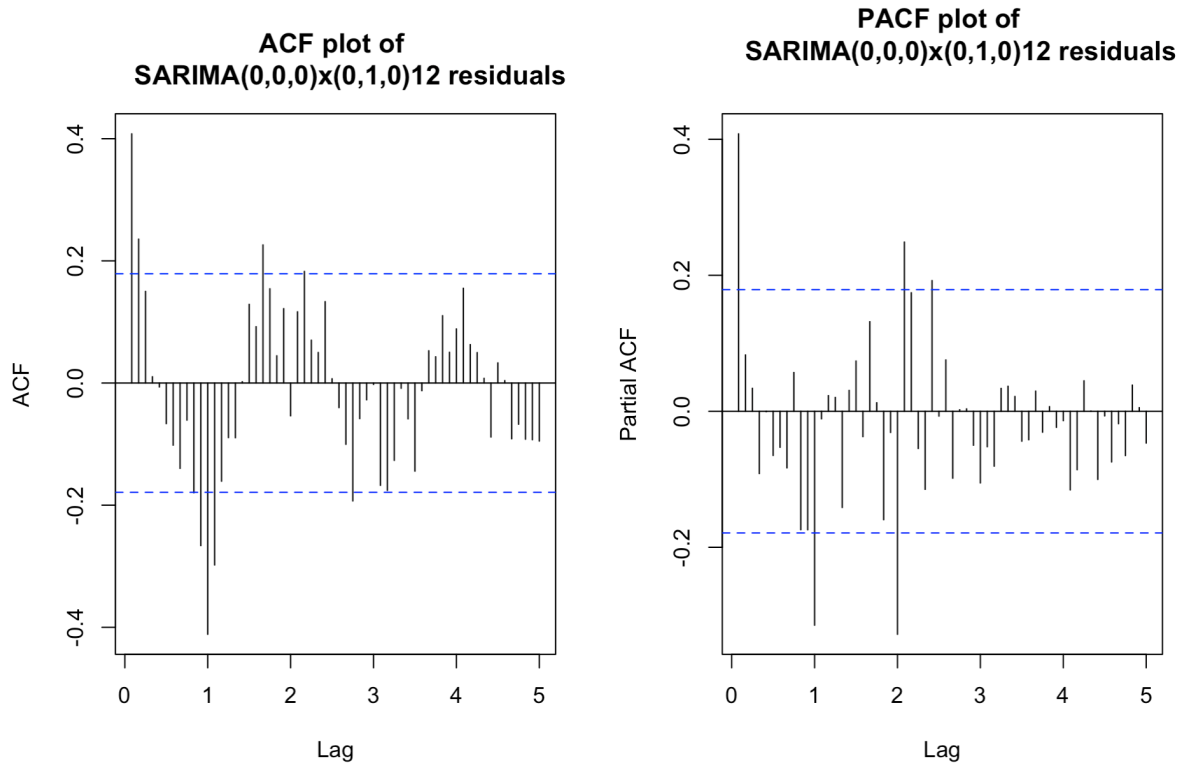

Figure 10– ACF / PACF of the residuals after fitting the first seasonal difference

R Code:

```
par(mfrow=c(1,2))
acf(res.m1, lag.max = 12*5, main = "ACF plot of
    SARIMA(0,0,0)x(0,1,0)12 residuals")
pacf(res.m1, lag.max = 12*5, main = "PACF plot of
     SARIMA(0,0,0)x(0,1,0)12 residuals")
par(mfrow=c(1,1))
```

Concerning the ACF plot, there is still a slowly decaying pattern within the first period. This suggests that there still exists a trend. Focusing on the autocorrelations at seasonal lags, we see that there are significant autocorrelations at the first seasonal lag in the ACF plot and two seasonal lag in the PACF plot. Therefore, the orders P = 2 and Q = 1 were chosen.

Based on the orders of P and Q, a SARIMA(0,0,0)x$(2,1,1)_{12}$ model will be fitted.

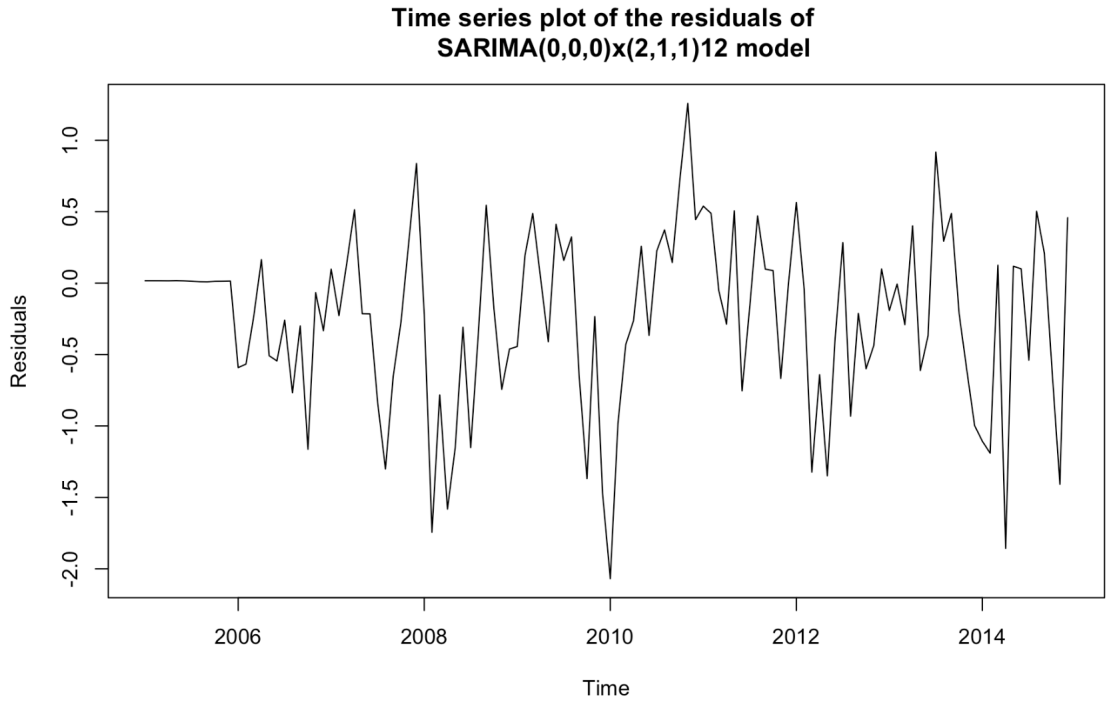

Figure 11– Time series plot of seasonally adjusted plot

R Code:

```
m2 = arima(cptss,order=c(0,0,0),seasonal=list(order=c(2,1,1), period=12))
res.m2 = residuals(m2);
par(mfrow=c(1,1))
plot(res.m2,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
     SARIMA(0,0,0)x(2,1,1)12 model")
```

The residual plot of model 2 looks exactly like the seasonally differenced model, with the residuals being stationary and not random. This suggests that the model did not capture some of the deterministic components. However, the standardised residuals are between the values of -3 and 3, thus suggesting normality.

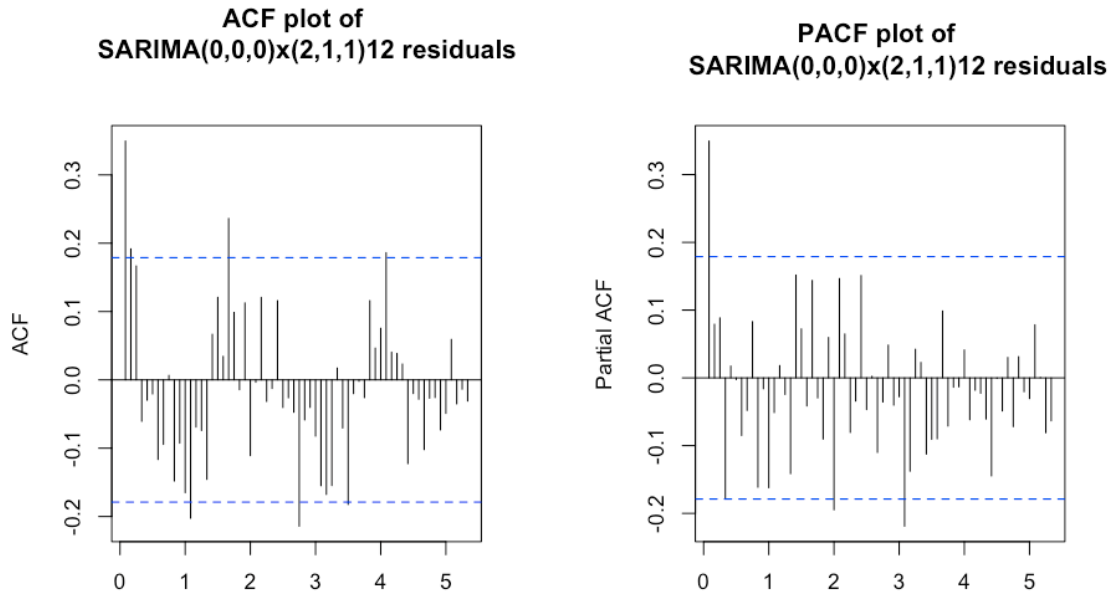

Figure 12– ACF / PACF residuals from SARIMA(0, 0, 0)x(2,1,1)_12 model

R Code:

```
par(mfrow=c(1,2))
acf(res.m2, lag.max = 64, main = "ACF plot of
    SARIMA(0,0,0)x(2,1,1)12 residuals ")
pacf(res.m2, lag.max = 64, main = "PACF plot of
     SARIMA(0,0,0)x(2,1,1)12 residuals")
par(mfrow=c(1,1))
```

After fitting a SARIMA(0,0,0)x(2,1,1)$_{12}$ model, most of the seasonality is removed except for a slightly significant correlation at seasonal lag 2 in PACF. We will go on with dealing with the ordinary series (by specifying the orders of ARIMA) and inspect the residuals to see if there is still evidence of seasonal components left.

A slight decaying pattern is seen in the ACF plot, indicating a possible ordinary trend in the series. An ADF test is done on the residuals to determine if the series is differenced nonstationary.

```
        Augmented Dickey-Fuller Test

data:  res.m2
Dickey-Fuller = -4.3566, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Figure 13– ADF test on SARIMA(,0,0,0)x(2,1,1)$_{12}$ residuals

R Code:

```
adf.test(res.m2)
```

The p-value of the ADF test is 0.01, therefore we reject the null hypothesis at the 95% confidence interval that the residuals of the SARIMA(0,0,0)x(2,1,1)$_{12}$ is nonstationary.

```
        Phillips-Perron Unit Root Test

data:  res.m2
Dickey-Fuller Z(alpha) = -80.402, Truncation lag parameter = 4, p-value = 0.01
alternative hypothesis: stationary
```

Figure 14– PP test on SARIMA(,0,0,0)x(2,1,1)$_{12}$ residuals

R Code:

```
pp.test(res.m2)
```

The Phillips-Perron Unit Root Test (PP) rejects the null hypothesis at the 95% confidence interval that the residuals of the SARIMA(0,0,0)x(2,1,1)$_{12}$ is nonstationary.

Since the ADF test, PP test and residual plot all stated that the residuals of the seasonally adjusted plot are stationary, we will not ordinary difference the residuals and would just determine the ARMA components of the model instead.

To determine the ordinary orders for the ARMA component of the model, the lags between 0 and 1 are assessed. The ACF plot in Figure 12 suggests there are two significant ordinary autocorrelations, while the PACF plot in the figure indicates possibly one to two significant lags. Therefore, the order of p = 1, 2, and q = 2 for the ARMA model will be analysed.

SARIMA(1,0,2)x(2,1,1)$_{12}$

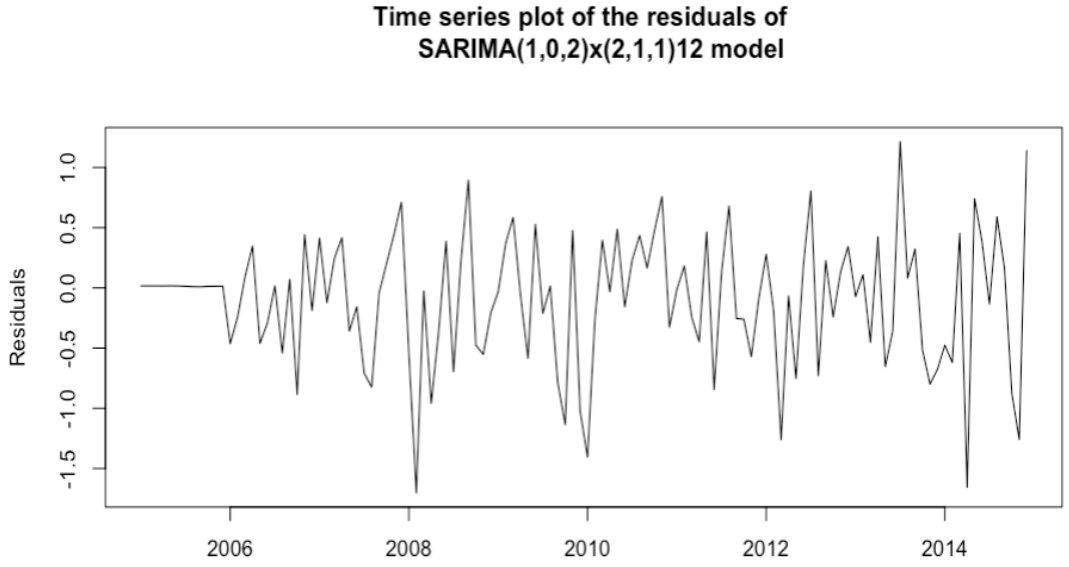

Figure 15– Time series of residuals from SARIMA(1, 0 2)x(2,1,1)_12 model

R Code:

```
m3 = arima(cptss,order=c(1,0,2),seasonal=list(order=c(2,1,1), period=12))
res.m3 = residuals(m3);
par(mfrow=c(1,1))
plot(res.m3,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
     SARIMA(1,0,2)x(2,1,1)12 model")
```

Figure 15 shows an improvement in the randomness of the residuals, a decrease in autoregressive and moving average characteristics is seen when compared to the previous model (Figure 12). This suggests the ARMA model with the order (1, 0, 2) has captured some of these behaviours in the series.

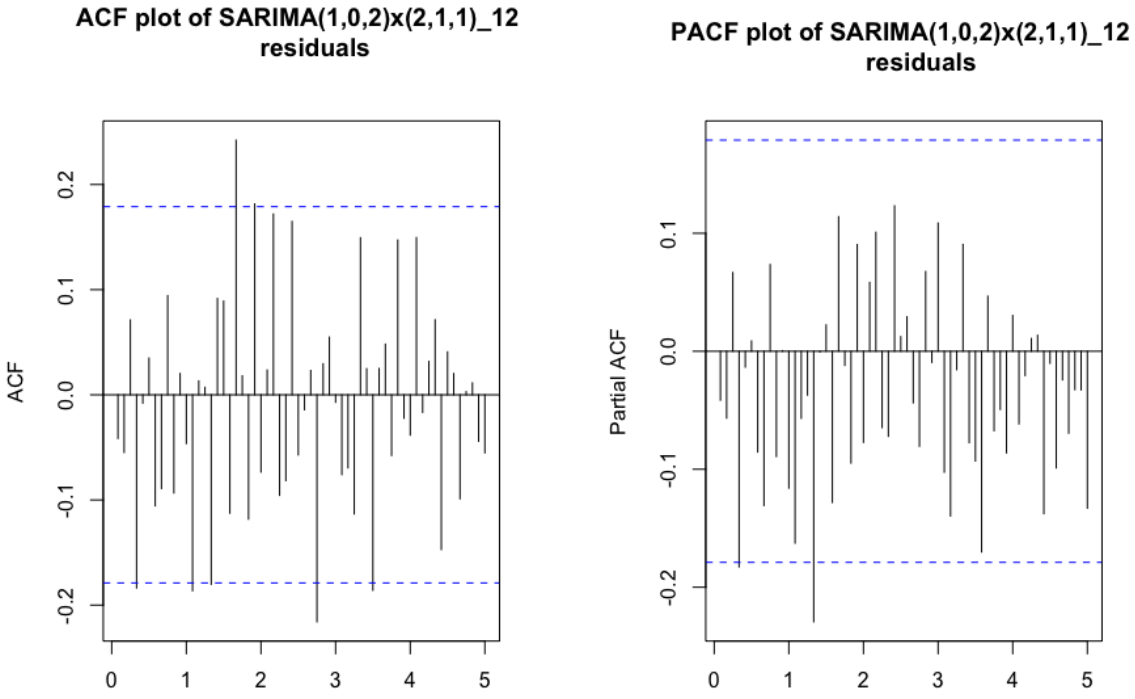

Figure 16– ACF / PACF residuals from SARIMA(1, 0, 2)x(2,1,1)_12 model

R Code:

```
par(mfrow=c(1,2))
acf(res.m3, lag.max = 12*5, main = "ACF plot of SARIMA(1,0,2)x(2,1,1)_12
    residuals")
pacf(res.m3, lag.max = 12*5, main = "PACF plot of SARIMA(1,0,2)x(2,1,1)_12
     residuals")
par(mfrow=c(1,1))
```

The ACF/PACF plots produced from this model helps to affirm the model's ability to capture the behaviour. There is no longer a slowly decaying pattern seen in the ACF plot, while the sizes of the significant ordinary lags are also smaller. The plots also help confirm the series is now stationary and does not require ordinary differencing. Moreover, this is confirmed by the PP and ADF test which both produced p-values less than alpha.

SARIMA(2,0,2)x(2,1,1)$_{12}$

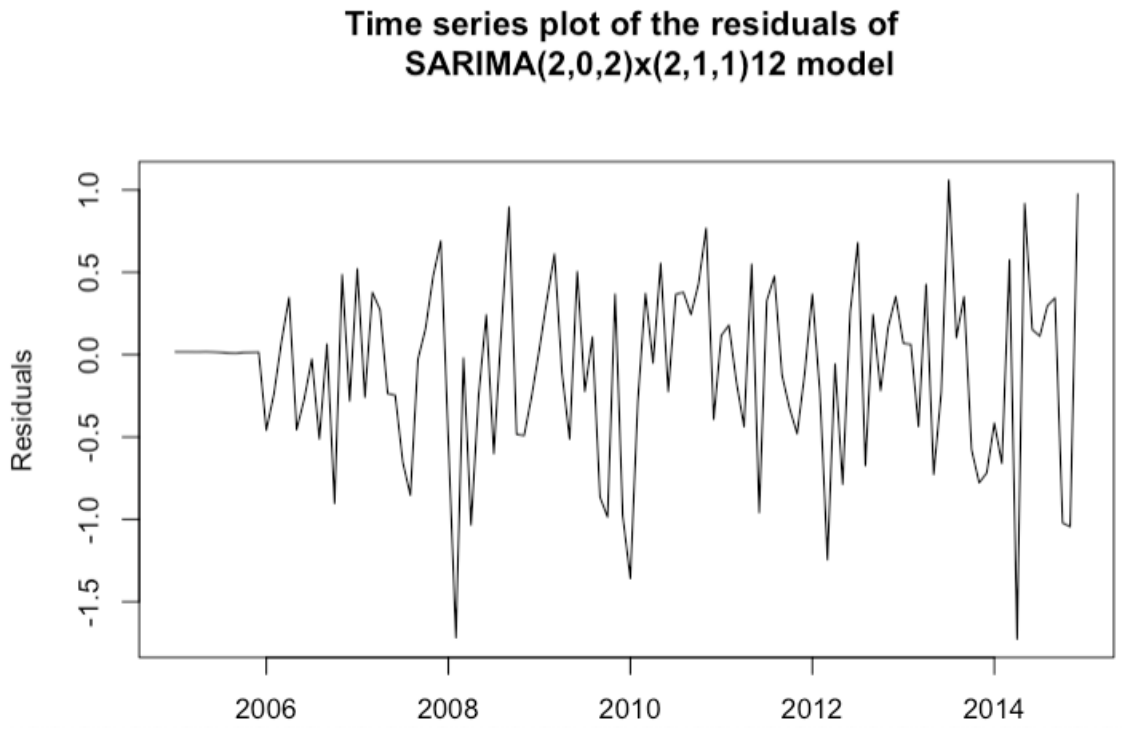

Figure 17– Time series of residuals from SARIMA(2, 0 2)x(2,1,1)_12 model

R Code:

```
m4 = arima(cptss,order=c(2,0,2),seasonal=list(order=c(2,1,1), period=12))
res.m4 = residuals(m4);
par(mfrow=c(1,1))
plot(res.m4,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
     SARIMA(2,0,2)x(2,1,1)12 model")
```

Similar to Figure 15, the residuals of SARIMA(2,0,2)x(2,1,1)$_{12}$ show improved randomness. This suggests the ARMA model with the order (2, 0, 2) has captured some of these behaviours in the series. The residuals are still stationary and within the range of -3 and 3 which suggest normality.

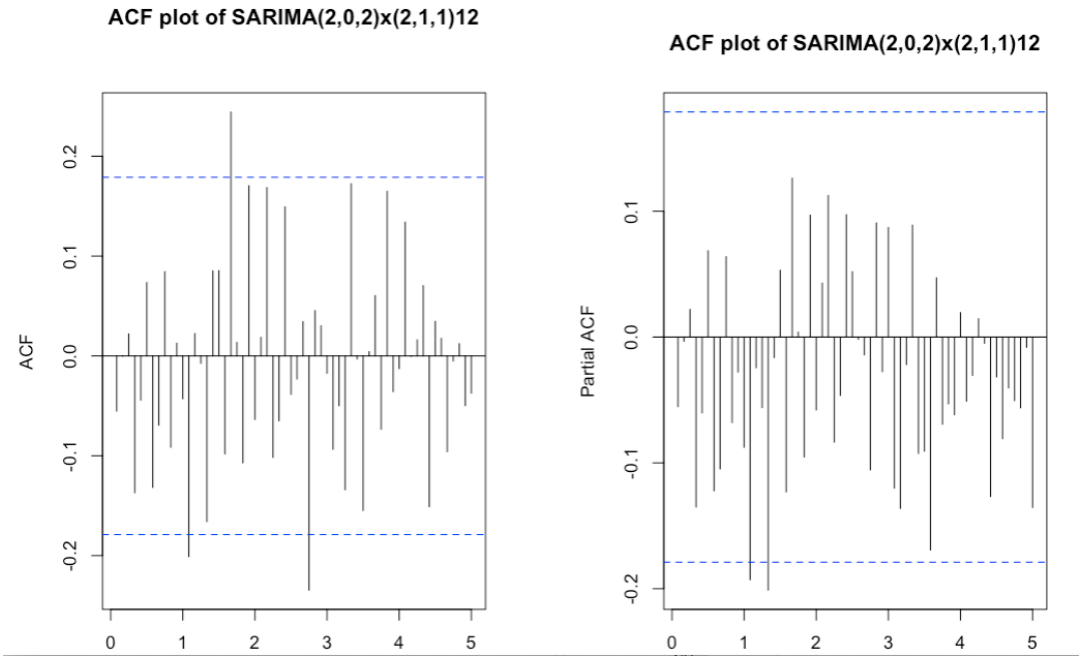

Figure 18– ACF/PACF plot of SARIMA(2, 0 2)x(2,1,1)_12 model

R Code:

```
par(mfrow=c(1,2))
acf(res.m4, lag.max = 12*5, main = "ACF plot of SARIMA(2,0,2)x(2,1,1)12")
pacf(res.m4, lag.max = 12*5, main = "ACF plot of SARIMA(2,0,2)x(2,1,1)12")
par(mfrow=c(1,1))
```

The ACF/PACF plot of SARIMA(2,0,2)x(2,1,1)$_{12}$ shows no sign of decaying pattern and no significant seasonal lags. This shows that the SARIMA(2,0,2)x(2,1,1)$_{12}$ model has captured most of the behaviour in the series. We also almost achieve white noise series with only a few significant autocorrelations.

Therefore we can include both SARIMA(1,0,2)x(2,1,1)$_{12}$ and SARIMA(2,0,2)x(2,1,1)$_{12}$ as suitable models for this series.

To get other alternative models, we can use EACF and the BIC table for the specification of ordinary lags after dealing with seasonal orders.

EACF

```
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 o o  o  x  o
1 x o o o o o o o o o o  o  o  o
2 x o x o o o o o o o o  o  o  o
3 x x x o o o o o o o o  o  o  o
4 o o x o o o o o o o o  o  o  o
5 o o x o o o o o o o o  o  o  o
6 o x o x o o o o o o o  o  o  o
7 x x o x o x o o o o o  o  o  o
```

Figure 19– EACF Plot of the Chickenpox series after dealing with seasonal orders

R Code:

```
eacf(res.m2)
```

Based on the EACF plot, the suggested SARIMA(p,0,q)x(2,1,1)$_{12}$ are:

1. SARIMA(0,0,2)x(2,1,1)$_{12}$
2. SARIMA(0,0,3)x(2,1,1)$_{12}$
3. SARIMA(1,0,2)x(2,1,1)$_{12}$
4. SARIMA(1,0,3)x(2,1,1)$_{12}$
5. SARIMA(1,0,1)x(2,1,1)$_{12}$

BIC

The models of the BIC plot are sorted based on their BIC, with the lower BIC models placed in the higher rows with darker shades. The lower the BIC, the better the model.

As the ACF/PACF and EACF plots generally suggest small models with autocorrelation less than 5, the AR order (nar) and MA order (nma) parameters are set to 5. The method used to fit the AR model will be the Ordinary Least Square (OLS) method, where the AR model will be determined by the AIC.

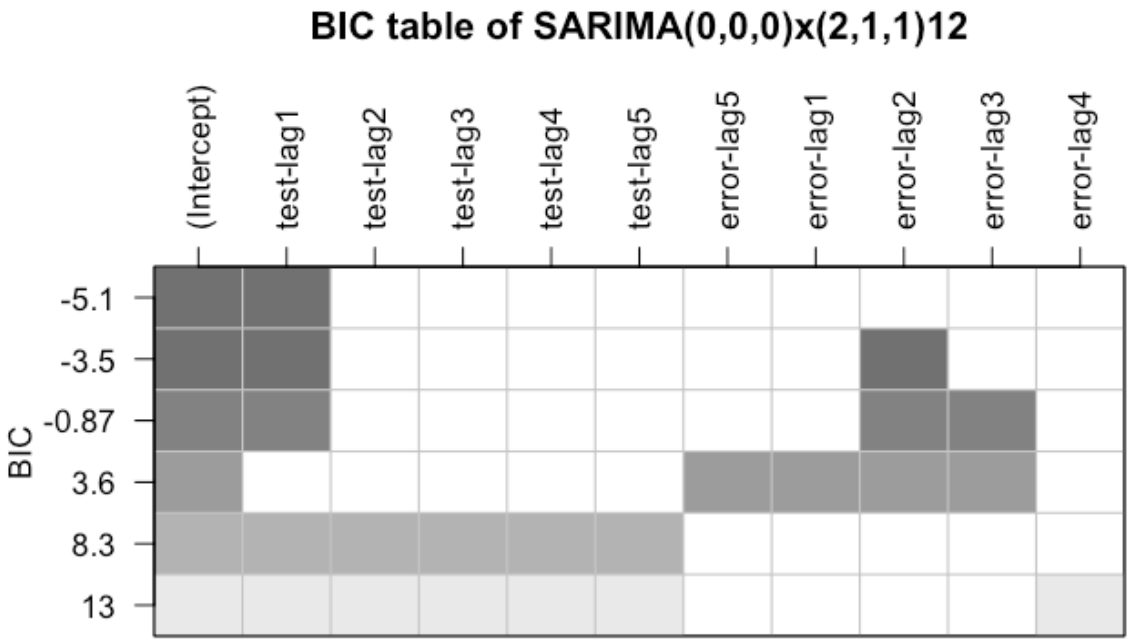

Figure 20– BIC Table of SARIMA(0,0,0)X(2,1,1)$_{12}$

Based on the BIC model, the best model contains lag 1 of the time series at lag 0 of the errors, while the next best model contains lag 1 of the time series and lag 2 of the errors.

As lag 1 of the time series model is most frequently found in the various subset models summarized in the exhibit, this suggests that it is a more important variable.

The SARIMA(p,0,q)x(2,1,1)$_{12}$ deduced from the BIC table are as follows:

1. SARIMA(1,0,0)x(2,1,1)$_{12}$
2. SARIMA(1,0,2)x(2,1,1)$_{12}$

The overall proposed set of possible SARIMA(p,d,q)x(2,1,1)$_{12}$models using all the suitable model specification tools such as the ACF/PACF plot, EACF plot and the BIC are as follows:

1. SARIMA(1,0,2)x(2,1,1)$_{12}$
2. SARIMA(2,0,2)x(2,1,1)$_{12}$
3. SARIMA(0,0,2)x(2,1,1)$_{12}$
4. SARIMA(0,0,3)x(2,1,1)$_{12}$
5. SARIMA(1,0,0)x(2,1,1)$_{12}$
6. SARIMA(1,0,1)x(2,1,1)$_{12}$

Model Fitting

Model : SARIMA(1,0,2) x $(2,1,1)_{12}$

```
z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1   0.833344   0.157022  5.3072 1.113e-07 ***
ma1  -0.404042   0.175095 -2.3076   0.02102 *
ma2  -0.062782   0.175488 -0.3578   0.72052
sar1 -0.157801   0.204929 -0.7700   0.44128
sar2 -0.327031   0.138132 -2.3675   0.01791 *
sma1 -0.505902   0.243669 -2.0762   0.03788 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 21– Coefficient Test of SARIMA(1,0,2)X(2,1,1)$_{12}$

R Code:

```
residual.analysis <- function(model, std = TRUE,start = 2, class = c("ARIMA","GARCH","ARMA-GARCH", "fGARCH")[1]){
  library(TSA)
  library(FitAR)
  if (class == "ARIMA"){
    if (std == TRUE){
      res.model = rstandard(model)
    }else{
      res.model = residuals(model)
    }
  }else if (class == "GARCH"){
    res.model = model$residuals[start:model$n.used]
  }else if (class == "ARMA-GARCH"){
    res.model = model@fit$residuals
  }else if (class == "fGARCH"){
    res.model = model@residuals
  }else {
    stop("The argument 'class' must be either 'ARIMA' or 'GARCH' ")
  }
  par(mfrow=c(3,2))
  plot(res.model,type='o',ylab='Standardised residuals', main="Time series plot of standardised residuals")
  abline(h=0)
  hist(res.model,main="Histogram of standardised residuals")
  qqnorm(res.model,main="QQ plot of standardised residuals")
  qqline(res.model, col = 2)
  acf(res.model,main="ACF of standardised residuals")
  print(shapiro.test(res.model))
  k=0
  LBQPlot(res.model, lag.max = 30, StartLag = k + 1, k = 0, SquaredQ = FALSE)
  par(mfrow=c(1,1))
}
```

```
m2_102.landing = arima(cptss,order=c(1,0,2),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_102.landing)
```

Based on the coefficient test, the coefficient AR1, MA1, SAR2, SMA1 are significant at the 95% confidence interval. On the other hand, the coefficient of MA2 and SAR1 are both not significant at the 95% confidence level. As SAR2 is significant at the 95% confidence level, we can deem SAR1 as significant.

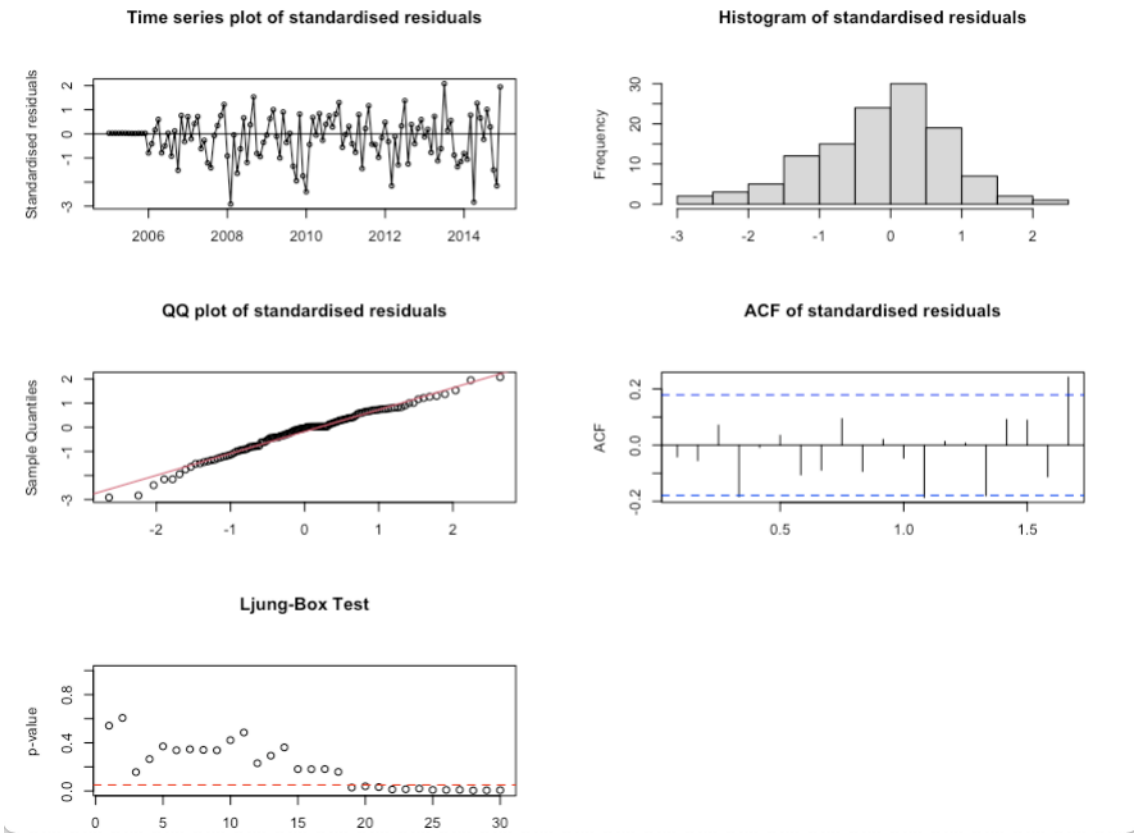

Figure 22– Residual Analysis of SARIMA(1,0,2)X(2,1,1)$_{12}$

R Code:

```
residual.analysis(model = m2_102.landing)
```

Based on the time series plot of the residuals of the standardised residuals, the plot does not suggest any major irregularities with the model as it looks randomly distributed and stationary.

The ACF plot of the standardised residuals shows signs of violation of the independence of residuals with slightly significant autocorrelation at a few lags. The Ljung-Box test confirms this, where the points after lag 19 of the series fail the test, therefore, we can conclude that the residuals are not white noise.

The histogram of the residuals appears symmetrical, suggesting normality. Additionally, the QQ plot looks approximately straight supporting the assumption of a normally distributed residual in the model.

```
        Shapiro-Wilk normality test

data:  res.model
W = 0.98302, p-value = 0.1357
```

Figure 23– Residual Analysis of SARIMA(1,0,2)X(2,1,1)$_{12}$

With a p-value of 0.1357, the Shapiro-Wilk normality test further confirms that the model is normally distributed as we can conclude not to reject the null hypothesis at the 95% confidence interval that the stochastic component of the model is normally distributed.

Model : SARIMA(2,0,2) x $(2,1,1)_{12}$

```
z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1   0.034372   0.193917  0.1772  0.859313
ar2   0.645033   0.124511  5.1805 2.212e-07 ***
ma1   0.434666   0.208175  2.0880  0.036799 *
ma2  -0.434399   0.152689 -2.8450  0.004441 **
sar1 -0.266532   0.227109 -1.1736  0.240561
sar2 -0.376796   0.143938 -2.6178  0.008851 **
sma1 -0.378297   0.269246 -1.4050  0.160013
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 24– Coefficient Test of SARIMA(2,0,2)X$(2,1,1)_{12}$

R Code:

```
m2_202.landing = arima(cptss,order=c(2,0,2),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_202.landing)
```

Based on the coefficient test, the coefficients AR2, MA1, MA2 and SAR2 are significant at the 95% confidence interval. On the other hand, the coefficients of AR1, SAR1 and SMA1 are not significant at the 95% confidence level. As AR2 and SAR2 are significant at the 95% confidence level, we can deem AR1 and SAR1 as significant.

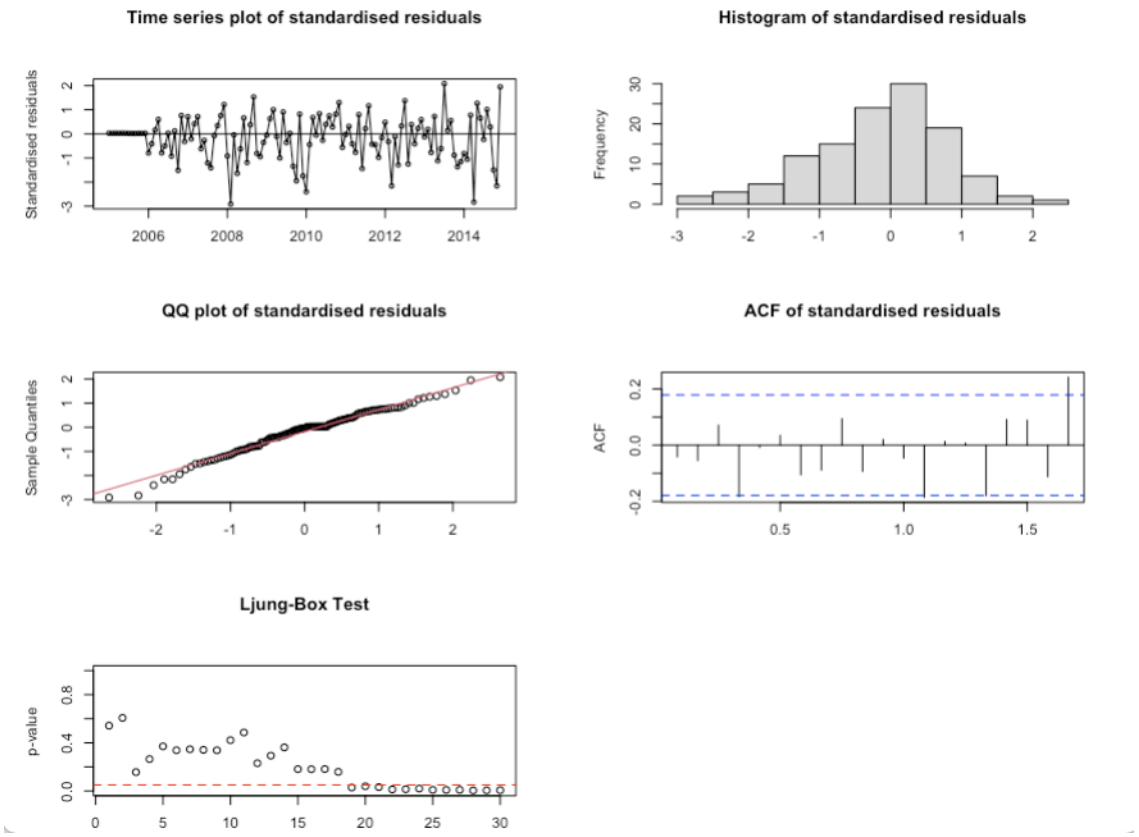

Figure 25– Residual Analysis of SARIMA(2,0,2)X(2,1,1)$_{12}$

R Code:

```
residual.analysis(model = m2_202.landing)
```

Similar to that of SARIMA(1,0,2)x(2,1,1)$_{12}$ the time series plot of the SARIMA(2,0,2)x(2,1,1)$_{12,}$ standardised residuals does not suggest any major irregularities with the model as it appears rather random and stationary.

The ACF plot of the standard residuals indicates a violation of the independence of residuals with slightly significant autocorrelations. The Ljung box test confirms this point where the points after lag 19 of the series fail the test, therefore, we can conclude that the residuals are not white noise.

The histogram does not show normality as it seems to be slightly left-skewed. The QQ plot, too, has a considerable amount of points departing from the reference line, suggesting that the normality assumption does not hold.

```
        Shapiro-Wilk normality test

data:  res.model
W = 0.97738, p-value = 0.0407
```

Figure 26– Shapiro-Wilk Normality Test of SARIMA(2,0,2)X(2,1,1)$_{12}$

With a p-value of 0.0407, the Shapiro-Wilk normality test further confirms that the residual of the SARIMA(2,0,2)x(2,1,1)$_{12}$ model is not normally distributed as we can conclude to reject the null hypothesis that the stochastic component of the linear model is normally distributed, at the 95% confidence interval.

Model : SARIMA(0,0,2) x (2,1,1)$_{12}$

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ma1   0.39709    0.11767  3.3747 0.0007389 ***
ma2   0.17548    0.11740  1.4948 0.1349731
sar1 -0.12210    0.21954 -0.5562 0.5781059
sar2 -0.24985    0.13938 -1.7926 0.0730395 .
sma1 -0.46705    0.22993 -2.0313 0.0422293 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure27– Coefficient Test of SARIMA(0,0,2)X(2,1,1)$_{12}$

R Code:

```
m2_002.landing = arima(cptss,order=c(0,0,2),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_002.landing)
```

Based on the coefficient test, the coefficients MA1 and SMA1 are significant at the 95% confidence interval. On the other hand, the coefficients of MA2, SAR1 and SAR2 are not significant at the 95% confidence level.

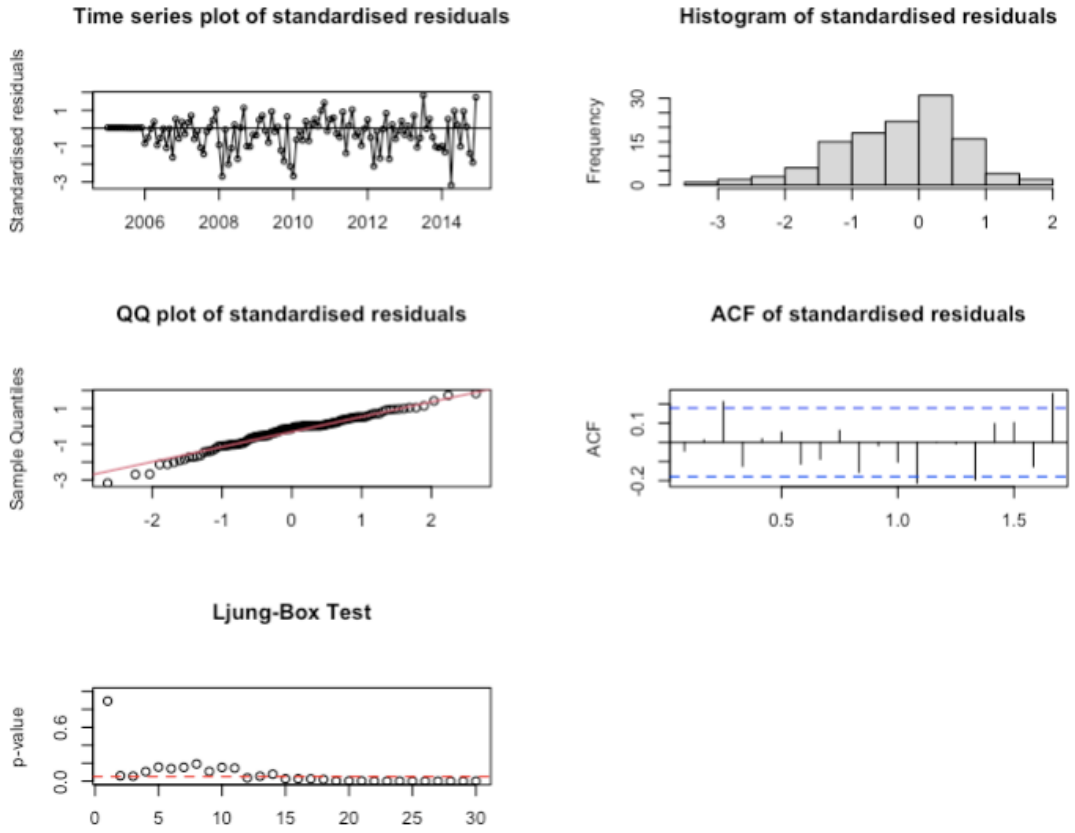

Figure 28– Residual Analysis of SARIMA(0,0,2)X(2,1,1)$_{12}$

R Code:

```
residual.analysis(model = m2_002.landing)
```

The time series plot of the SARIMA(0,0,2)x(2,1,1)$_{12,}$ standardised residuals, too, does not suggest any major irregularities with the model as it appears fairly random and stationary.

The ACF plot of the standard residuals shows signs of violation of the independence with significant autocorrelation at a few lags. The Ljung box test confirms this point where almost all the points of the series fail the test, therefore, we can conclude that the residuals still contain uncaptured characteristics. It also has more significant autocorrelation as compared to the previous 2 models.

The histogram does not show normality as it seems to be left-skewed. Furthermore, the QQ plot has a considerable amount of points departing from the reference line, suggesting that the normality assumption does not hold.

```
        Shapiro-Wilk normality test

data:  res.model
W = 0.97412, p-value = 0.02046
```

Figure 29– Shapiro-Wilk Normality Test of SARIMA(0,0,2)X(2,1,1)$_{12}$

With a p-value of 0.02046, the Shapiro-Wilk normality test further confirms that the residual of the SARIMA(0,0,2)x(2,1,1)$_{12}$ model is not normally distributed as we can conclude to reject the null hypothesis that the stochastic component of the linear model is normally distributed, at the 95% confidence interval.

Model : SARIMA(0,0,3) x (2,1,1)$_{12}$

```
z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ma1   0.481796   0.105229  4.5786 4.682e-06 ***
ma2   0.286442   0.120794  2.3713 0.0177243 *
ma3   0.342963   0.094573  3.6264 0.0002874 ***
sar1 -0.418859   0.239928 -1.7458 0.0808513 .
sar2 -0.426420   0.146215 -2.9164 0.0035411 **
sma1 -0.240422   0.278869 -0.8621 0.3886143
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 30– Coefficient Test of SARIMA(0,0,3)X(2,1,1)$_{12}$

R Code:

```
m2_003.landing = arima(cptss,order=c(0,0,3),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_003.landing)
```

Based on the coefficient test, the coefficients MA1, MA2, MA3, and SAR2 are significant at the 95% confidence interval. On the other hand, the coefficients of SAR1 and SMA1 are not statistically significant. As SAR2 is significant at the 95% confidence level, we can deem SAR1 as significant.

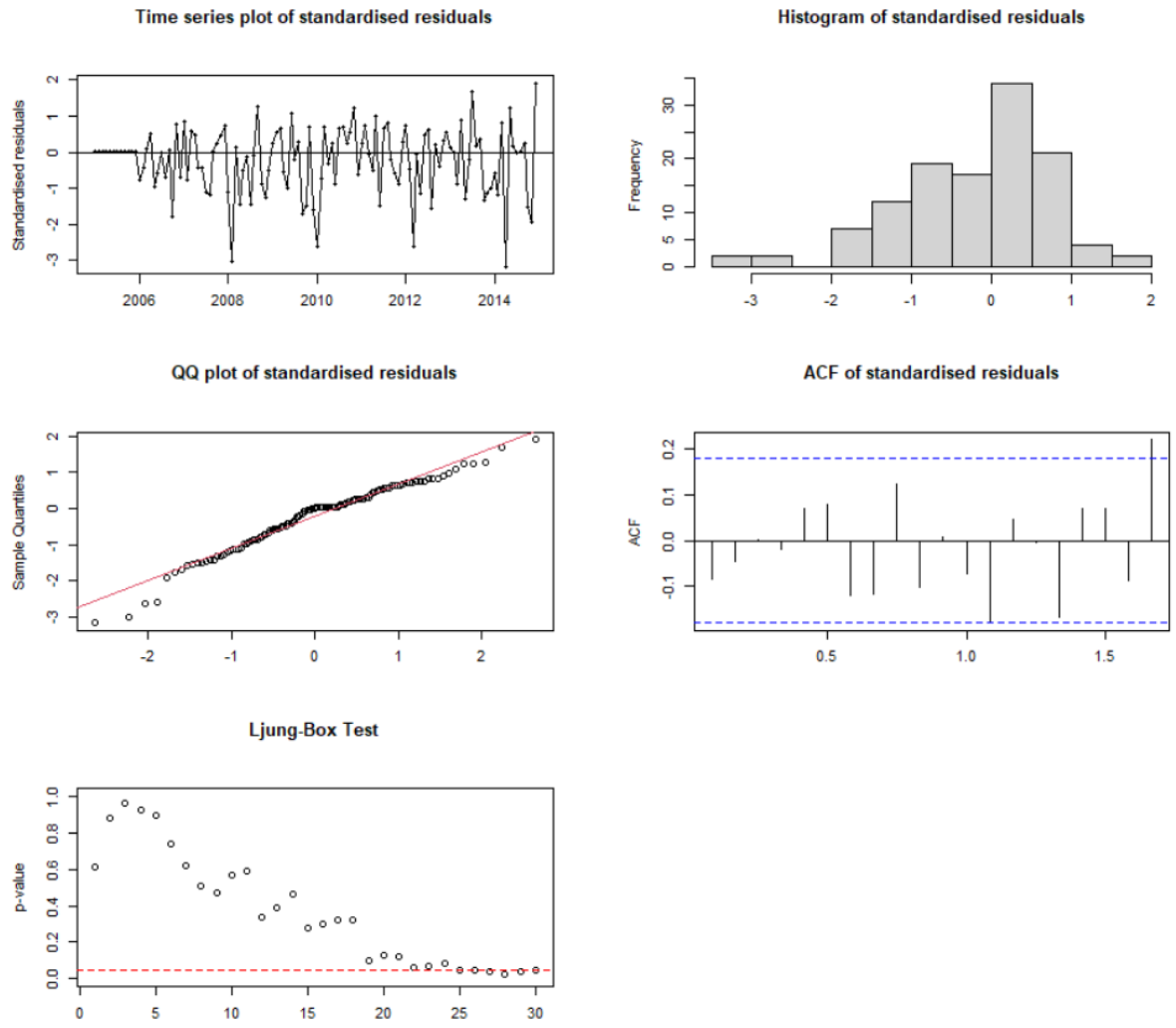

Figure 31– Residual Analysis of SARIMA(0,0,3)X(2,1,1)$_{12}$

R Code:

```
residual.analysis(model = m2_003.landing)
```

The time series plot of the SARIMA(0,0,3)x(2,1,1)$_{12,}$ standardised residuals, too, does not suggest any major irregularities with the model as it seems random and stationary.

The ACF plot of the standard residuals shows minimal signs of violation of the independence of residuals with significant autocorrelation at only two lags. The Ljung box test shows that points till 25 passes the test. However, there points towards the end fails the test and we can therefore conclude that the residuals are not white noise. This model also has less significant autocorrelation as compared to the previous models.

The histogram shows little signs of normality with a left skewness. Moreover, the QQ plot also has a considerable amount of points departing from the reference line, suggesting that the normality assumption does not hold.

```
        Shapiro-Wilk normality test

data:  res.model
W = 0.96627, p-value = 0.004173
```

Figure 32– Shapiro-Wilk Normality Test of SARIMA(0,0,3)X(2,1,1)$_{12}$

With a p-value of 0.004173, the Shapiro-Wilk normality test further confirms that the residual of the SARIMA(0,0,3)x(2,1,1)$_{12}$ model is not normally distributed as we can conclude to reject the null hypothesis that the stochastic component of the linear model is normally distributed, at the 95% confidence interval.

Model : SARIMA(1,0,0) x (2,1,1)$_{12}$

```
z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1   0.524033   0.091556  5.7237 1.043e-08 ***
sar1 -0.121269   0.211731 -0.5728   0.56681
sar2 -0.308549   0.139512 -2.2116   0.02699 *
sma1 -0.499552   0.241946 -2.0647   0.03895 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 33– Coefficient Test of SARIMA(1,0,0)X(2,1,1)$_{12}$

R Code:

```
m2_100.landing = arima(cptss,order=c(1,0,0),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_100.landing)
```

Based on the coefficient test, the coefficients AR1, SAR2, SMA1 are significant at the 95% confidence interval. On the other hand, the coefficient of SAR1 is not significant at the 95% confidence level. As SAR2 is significant at the 95% confidence level, we can deem SAR1 as significant.

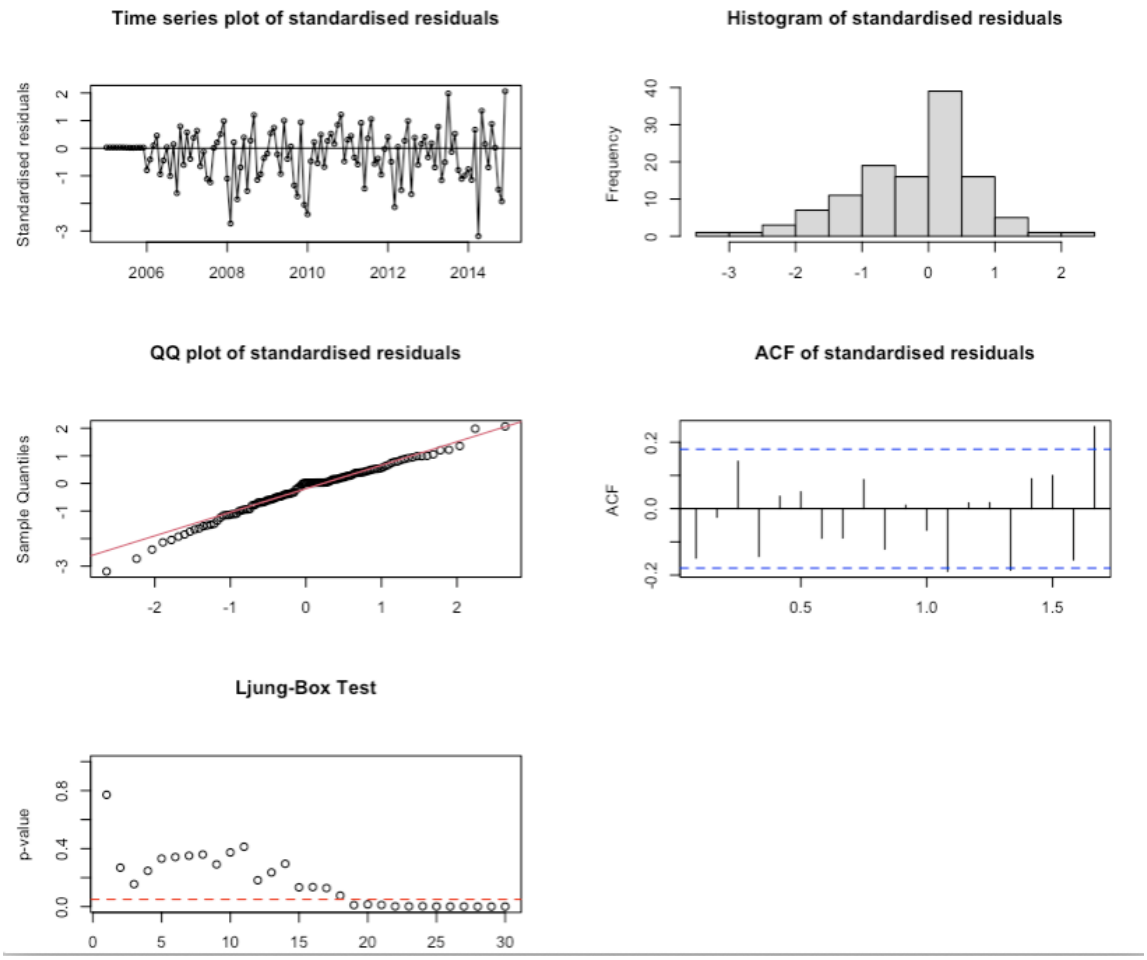

Figure 34– Residual Analysis of SARIMA(1,0,0)X(2,1,1)$_{12}$

R Code:

```
residual.analysis(model = m2_100.landing)
```

Similar to the previous plots, the time series plot of the SARIMA(1,0,0)x(2,1,1)$_{12,}$ standardised residuals, too, does not suggest any obvious trends as it appears fairly random and stationary.

The ACF plot of the standardised residuals shows slight signs of violation of the independence of residuals with slightly significant autocorrelation at 3 lags. The Ljung-Box test confirms this, where the points after lag 19 of the series fail the test, therefore, we can conclude that the residuals are not white noise.

The histogram of the residuals seems symmetrical suggesting normality. Additionally, the QQ plot appears to be straight, supporting the assumption of a normally distributed residual in the model.

```
        Shapiro-Wilk normality test

data:  res.model
W = 0.98034, p-value = 0.07666
```

Figure 35– Shapiro-Wilk Normality Test of SARIMA(1,0,0)X(2,1,1)$_{12}$

With a p-value of 0.07666, the Shapiro-Wilk normality test further confirms that the residual of the SARIMA(1,0,0)x(2,1,1)$_{12}$ model is normally distributed as we can conclude not to reject the null hypothesis that the stochastic component of the linear model is normally distributed, at the 95% confidence interval.

Model : SARIMA(1,0,1) x (2,1,1)$_{12}$

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ar1   0.79008    0.11697  6.7547 1.431e-11 ***
ma1  -0.37872    0.17147 -2.2086   0.02720 *
sar1 -0.16917    0.21069 -0.8030   0.42200
sar2 -0.33128    0.13887 -2.3855   0.01705 *
sma1 -0.48412    0.24314 -1.9911   0.04647 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 36– Coefficient Test of SARIMA(1,0,1)X(2,1,1)$_{12}$

R Code:

```
m2_101.landing = arima(cptss,order=c(1,0,1),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_101.landing)
```

Based on the coefficient test, the coefficients AR1, MA1, SAR2, SMA1 are significant at the 95% confidence interval. On the other hand, the coefficient of SAR1 is not significant at the 95% confidence level. As SAR2 is significant at the 95% confidence level, we can deem SAR1 as significant. Based on the coefficient test of all the models, both the SARIMA(1,0,0)x(2,1,1)$_{12}$ and SARIMA(1,0,1)x(2,1,1)$_{12}$ are deemed as the better models amongst those discussed as they have all of their coefficients deemed significant.

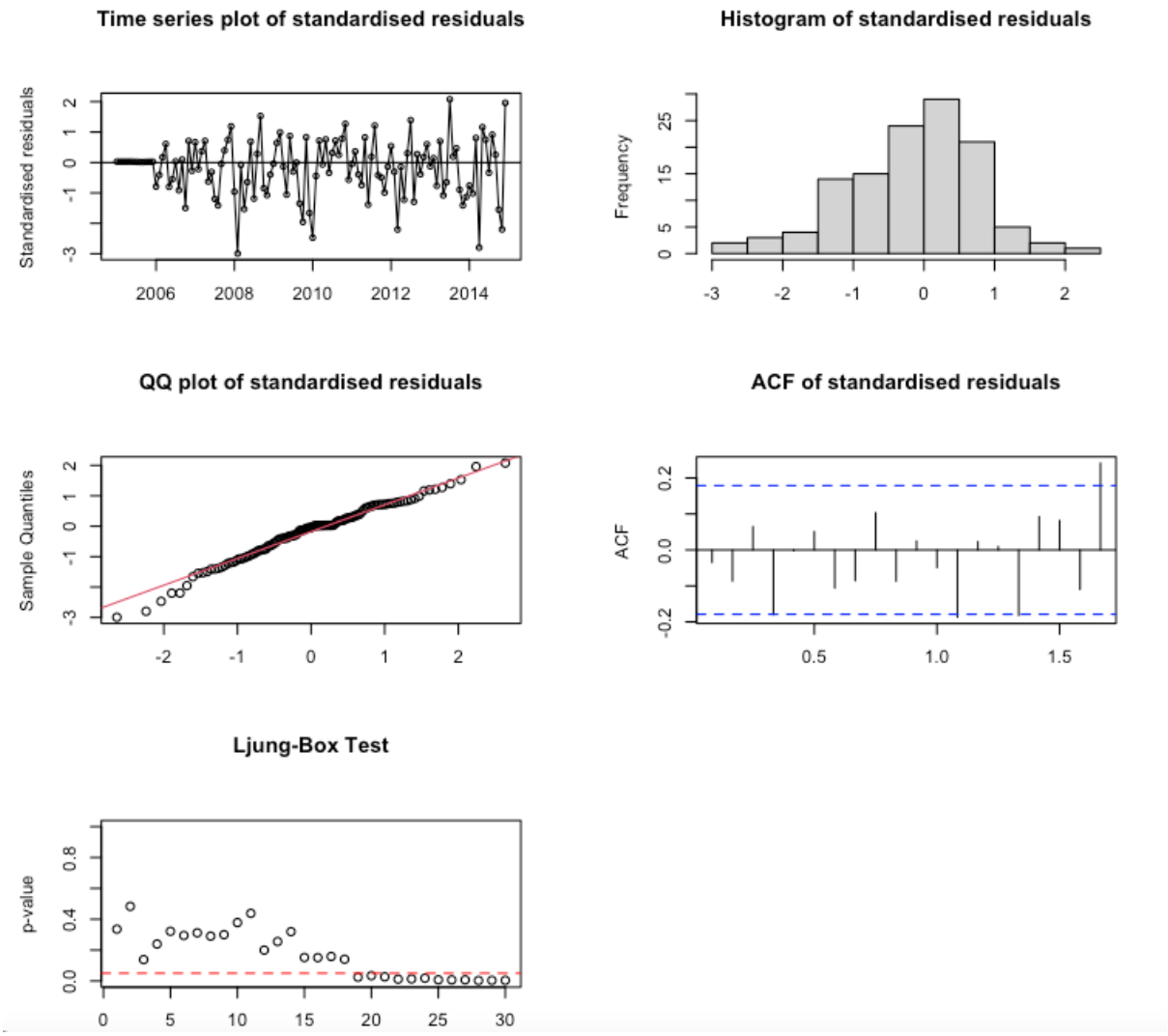

Figure 37– Residual Analysis of SARIMA(1,0,1)X(2,1,1)$_{12}$

R Code:

```
residual.analysis(model = m2_101.landing)
```

The time series plot of the SARIMA(1,0,1)x(2,1,1)$_{12,}$ standardised residuals, too, does not suggest any obvious trends as it appears fairly random and stationary.

The ACF plot of the standardised residuals shows slight signs of violation of the independence of residuals with slightly significant autocorrelation at 4 lags. The Ljung-Box test confirms this, where the points after lag 19 of the series fail the test, therefore, we can conclude that the residuals are not white noise.

The histogram of the residuals seems symmetrical suggesting normality. In addition, the QQ plot appears like a straight line, supporting the assumption of a normally distributed residual in the model.

```
        Shapiro-Wilk normality test

data:  res.model
W = 0.98187, p-value = 0.1061
```

Figure 38– Shapiro-Wilk Normality Test of SARIMA(1,0,1)X(2,1,1)$_{12}$

With a p-value of 0.1061, the Shapiro-Wilk normality test further confirms that the residual of the SARIMA(1,0,1)x(2,1,1)$_{12}$ model is normally distributed as we can conclude not to reject the null hypothesis that the stochastic component of the linear model is normally distributed, at the 95% confidence interval.

With regards to the coefficient tests and residual analysis, both the SARIMA(1,0,0)x(2,1,1)$_{12}$ and SARIMA(1,0,1)x(2,1,1)$_{12}$ would be deemed as the better models as they both have all their coefficients deemed significant. Its residuals, although not white noise, are normal, random and stationary. Significant p-values in the Ljung-Box test in the later lags are seen for all models, therefore, it would be deemed acceptable as there are no other models that would give a better result.

AIC

```
                df      AIC
m2_003.landing  7 212.4004
m2_101.landing  6 213.0146
m2_202.landing  8 213.8632
m2_100.landing  5 214.6438
m2_102.landing  7 214.8850
m2_002.landing  6 221.9672
```

Figure 39– AIC of models

R Code:

```
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")')
  }
}
```

```
sc.AIC=AIC(m2_102.landing , m2_002.landing,m2_003.landing, m2_100.landing, m2_202.landing, m2_101.landing)
sort.score(sc.AIC, score = "aic")
```

Based on AIC, the model with the lowest AIC would be SARIMA(0,0,3)x(2,1,1)$_{12}$, followed by SARIMA(1,0,1)x(2,1,1)$_{12}$, SARIMA(2,0,2)x(2,1,1)$_{12}$, SARIMA(1,0,0)x(2,1,1)$_{12}$.

Although having the lower AIC, the SARIMA(0,0,3)x(2,1,1)$_{12}$ and the SARIMA(2,0,2)x(2,1,1)$_{12}$ model has some of its coefficients insignificant at the 95% confidence level, the residuals of the model are also not normally distributed nor white noise. Also, despite the SARIMA(1,0,1)x(2,1,1)$_{12}$ having a lower AIC score as compared to SARIMA(1,0,0)x(2,1,1)$_{12,}$ the differences between the AICs are minor.

BIC

	df	AIC
m2_100.landing	5	228.5813
m2_101.landing	6	229.7395
m2_003.landing	7	231.9128
m2_102.landing	7	234.3975
m2_202.landing	8	236.1631
m2_002.landing	6	238.6922

Figure 40– BIC of models

R Code:

```
sc.BIC=AIC(m2_102.landing , m2_002.landing,m2_003.landing, m2_100.landing, m2_202.landing, m2_101.landing, k = log(length(cpts)))
sort.score(sc.BIC, score = "aic")
```

On the other hand, based on BIC, the model with the lowest BIC would be SARIMA(1,0,0)x(2,1,1)$_{12}$, followed by SARIMA(1,0,1)x(2,1,1)$_{12}$ then SARIMA(0,0,3)x(2,1,1)$_{12.}$

Since the SARIMA(1,0,0)x(2,1,1)$_{12}$ and SARIMA(1,0,1)x(2,1,1)$_{12}$ are incomparable in terms of their coefficient, residual, AIC and BIC analysis, the smaller model, SARIMA(1,0,0)x(2,1,1)$_{12}$ will be picked according to the law of parsimony as the best fitting model for the time series data.

Over Parameterization: SARIMA(2,0,0)x(2,1,1)$_{12}$

Since SARIMA(1,0,1)x(2,1,1)$_{12}$ has already been compared with SARIMA(1,0,0)x(2,1,1)$_{12,}$ to over parameterise the model, SARIMA(2,0,0)x(2,1,1)$_{12}$ will be run to determine if SARIMA(1,0,0)x(2,1,1)$_{12}$ would be the best model of choice.

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ar1   0.44086    0.10189  4.3267 1.514e-05 ***
ar2   0.16247    0.10151  1.6005   0.10949
sar1 -0.14292    0.21109 -0.6771   0.49836
sar2 -0.31742    0.13908 -2.2823   0.02247 *
sma1 -0.49626    0.24363 -2.0370   0.04165 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 41– Coefficient Test of SARIMA(2,0,0)X(2,1,1)$_{12}$

R Code:

```
m2_200.landing = arima(cptss,order=c(2,0,0),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_200.landing)
residual.analysis(model = m2_200.landing)
```

Based on the coefficient test, the coefficients AR1, SAR2 and SMA1 are significant at the 95% confidence interval. On the other hand, the coefficient of AR2 and SAR1 is not significant at the 95% confidence level. Therefore when an additional AR parameter is added (AR2), the parameter is insignificant at the 95% confidence level. This concludes the suitability of SARIMA(1,0,0)x(2,1,1)$_{12}$ with this diagnostic check.

Prediction With SARIMA(1,0,0)x(2,1,1)$_{12}$

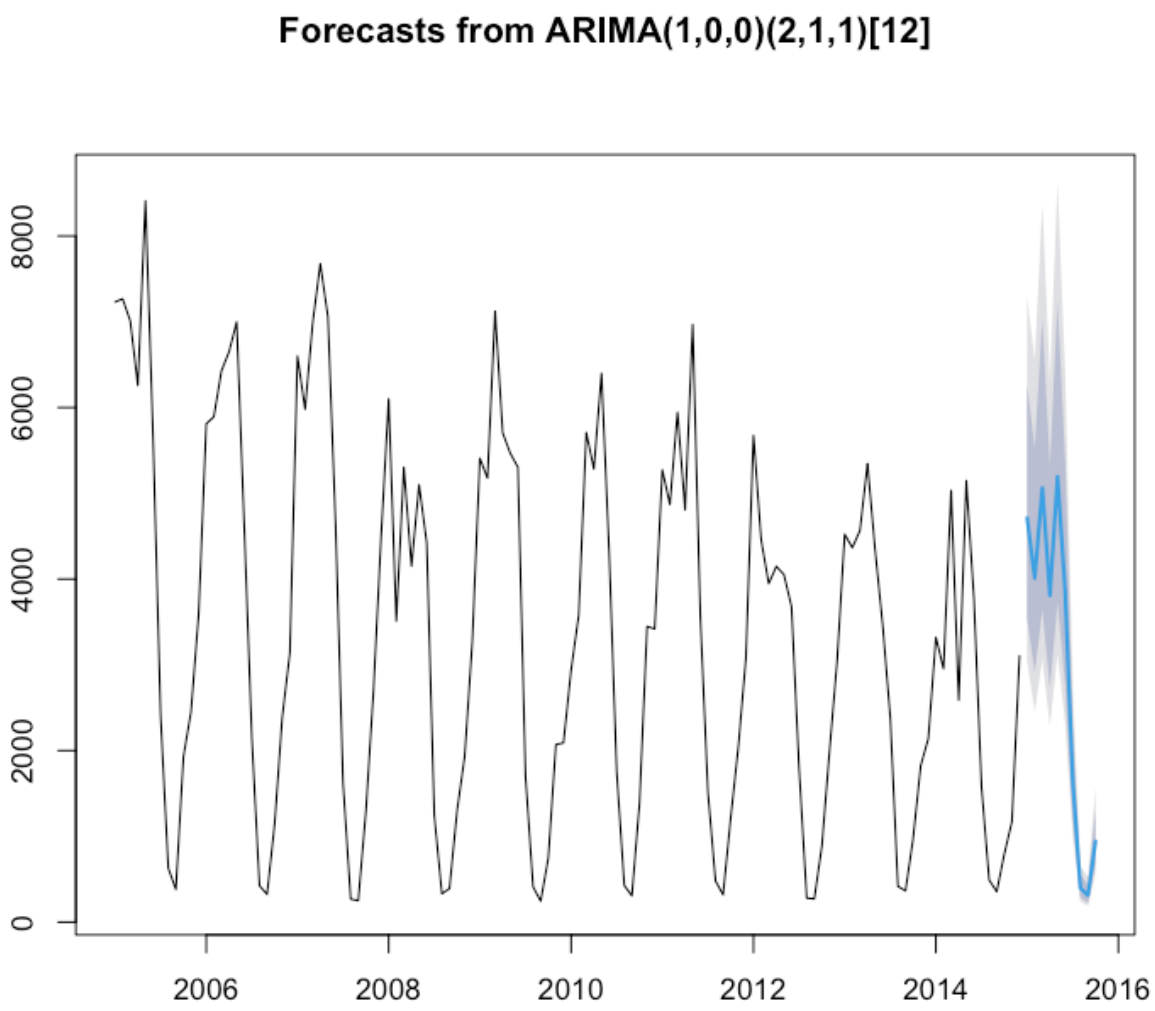

Figure 42– forecast from ARMA(1,0,0)(2,1,1)$_{12}$

R Code:

```
m2_100.pred= Arima(cpts,order=c(1,0,0),seasonal=list(order=c(2,1,1), period=12), lambda = 0)
future = forecast(m2_100.pred, h = 10)
plot(future)
```

Figure 42 shows the prediction of the monthly Chickenpox cases for the next 10 months based on the best-fitting model (SARIMA(1,0,0)x(2,1,1)$_{12}$. From the forecasted trend line (blue), we can see that the forecast mimics the stochastic periodicity in the data quite well. The forecast's dark grey areas display the 85% prediction interval, while the lighter grey plot gives the 95% prediction interval. This means that the new observations are expected to be within that range 85% or 95% of the time.

Conclusion

The Hungary chickenpox dataset indicated a stochastic trend with numerous characteristics. Firstly, a box-cox transformation was applied in an attempt to reduce the changing variance in the series. With the transformed series, seasonal models were created to deal with the seasonal trends that were discovered during the data exploration stage. Then, ARMA models were built upon the seasonal models. Several models were proposed through various tools like the ACF/PACF, EACF and BIC tools. With the set proposed models, each was fitted and their residuals analysed. It is concluded that SARIMA(1,0,0)x$(2,1,1)_{12}$ was the best model to represent the series as it not only captured many of the characteristics but also was parsimonious. However, the analysis of this model still indicated some signs of autocorrelation in the residuals, suggesting a more comprehensive model is needed to capture all the information in the series.

A prediction of the monthly Chickenpox cases for the next 10 months was run based on the best-fitting model (SARIMA(1,0,0)x$(2,1,1)_{12.}$ The forecast is seen to mimic the stochastic periodicity in the data pretty well.

Bibliography

Huber, A., Gazder, J., Dobay, O., Mészner, Z., & Horváth, A. (2020). Attitudes towards varicella vaccination in parents and paediatric healthcare providers in Hungary. *Vaccine*, *38*(33), 5249-5255.

Rettner, R. (2019). *Not Just the Flu: Gonorrhea, Chicken Pox Also Go Through Seasons*. livescience.com. Retrieved 20 May 2021, from https://www.livescience.com/64036-infectious-diseases-seasonality.html#:~:text=Some%20of%20the%20best%2Ddescribed,in%20the%20summer%20and%20autumn.

UCI Machine Learning Repository (2021), *Hungarian Chickenpox Cases Data Set,* data file, University of California, School Information and Computer Science, Irvine, CA, viewed 28 April 2021, <https://archive.ics.uci.edu/ml/datasets/Hungarian+Chickenpox+Cases>

Appendix

```
library(TSA)
library(fUnitRoots)
library(forecast)
library(CombMSC)
library(lmtest)
library(tseries)
library(lubridate)
library(fpp)
library(dplyr)

cp <- read.csv("hungary_chickenpox.csv")

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

```
  print(shapiro.test(res.model))
  k=0
  LBQPlot(res.model, lag.max = 30, StartLag = k + 1, k = 0, SquaredQ = FALSE)
  par(mfrow=c(1,1))
}


cpdata <- cp %>% summarise(cases = rowSums(cp[2:21]))
cps <- cbind(cp[,1], cpdata)
cps <- as.data.frame(cps)
cps[,1] <- dmy(cps[,1])
cps <- cps %>% group_by(year(cps[,1]),  month(cps[,1])) %>% summarise(total = sum(cases))



cpts <- ts(cps[,3],start=c(2005,1), end=c(2014,12), frequency=12)
plot(cpts ,ylab='Chickenpox cases in Hungary',xlab='Year',type='o',
    main = "Time series plot of Number of Chickenpox cases in Hungary")


#look out for seasonality
plot(cpts ,type='l', main = "Time series plot of Number of chickenpox cases in Hungary")
points(y=cpts ,x=time(cpts), pch=as.vector(season(cpts)))

#impact of previous month's cases on next month's cases:
y= cpts
x = zlag(cpts)

index = 2:length(x) # Create an index to get rid of the first NA value in x
cor(y[index],x[index]) #high positive correlation:  0.7769839

plot(y=cpts,x=zlag(cpts),ylab='Chickenpox Cases in Hungary', xlab= 'Chickenpox cases in the previous month',
    main= "Scatter plot of neighboring Chickenpox Cases in Hungary")
abline(lm(cpts~zlag(cpts)), col= "red")

par(mfrow=c(1,2))
acf(cpts,lag.max = 12*5,main = "ACF plot of the Chickenpox Series")
pacf(cpts,lag.max = 12*5, main = "ACF plot of the Chickenpox Series")
par(mfrow=c(1,1))

adf.test(cpts) #confirms stationarity.
pp.test(cpts) #confirms stationarity.


#boxcox transformation
options(warn=-1)
BC <- BoxCox.ar(y = cpts , lambda = seq(-2, 2, 0.01))
# To find the optimal lambda value
lambda <- BC$lambda[which(max(BC$loglike) == BC$loglike)]
```

```
# Apply Box-Cox transformation
cptss  <- ((cpts ^lambda) - 1) / lambda
par(mfrow=c(1,2))
plot(cptss , ylab='Box-Cox transformed Chickenpox Series', xlab='Time', type='o',
    main = "Box-Cox tranformed Chickenpox Series")

par(mfrow=c(1,2))
acf(cptss,lag.max = 12*5,main = "ACF plot of the Box-Cox
   Transformed Chickenpox series")
pacf(cptss,lag.max = 12*5, main = "PACF plot of the Box-Cox
    Transfromed Chickenpox series")
par(mfrow=c(1,1))




#####################################################################
m1 = Arima(cptss,order=c(0,0,0),seasonal=list(order=c(0,1,0), period=12))
res.m1 = residuals(m1);
par(mfrow=c(1,1))
plot(res.m1,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
    SARIMA(0,0,0)x(0,1,0)12 model")
par(mfrow=c(1,2))
acf(res.m1, lag.max = 12*5, main = "ACF plot of
   SARIMA(0,0,0)x(0,1,0)12 residuals")
pacf(res.m1, lag.max = 12*5, main = "PACF plot of
    SARIMA(0,0,0)x(0,1,0)12 residuals")
par(mfrow=c(1,1))


m2 = arima(cptss,order=c(0,0,0),seasonal=list(order=c(2,1,1), period=12))
res.m2 = residuals(m2);
par(mfrow=c(1,1))
plot(res.m2,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
    SARIMA(0,0,0)x(2,1,1)12 model")
par(mfrow=c(1,2))
acf(res.m2, lag.max = 64, main = "ACF plot of
   SARIMA(0,0,0)x(2,1,1)12 residuals ")
pacf(res.m2, lag.max = 64, main = "PACF plot of
    SARIMA(0,0,0)x(2,1,1)12 residuals")
par(mfrow=c(1,1))

adf.test(res.m2)
pp.test(res.m2)

m3 = arima(cptss,order=c(1,0,2),seasonal=list(order=c(2,1,1), period=12))
res.m3 = residuals(m3);
par(mfrow=c(1,1))
plot(res.m3,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
```

```
    SARIMA(1,0,2)x(2,1,1)12 model")
par(mfrow=c(1,2))
acf(res.m3, lag.max = 12*5, main = "ACF plot of SARIMA(1,0,2)x(2,1,1)_12
   residuals")
pacf(res.m3, lag.max = 12*5, main = "PACF plot of SARIMA(1,0,2)x(2,1,1)_12
    residuals")
par(mfrow=c(1,1))

m4 = arima(cptss,order=c(2,0,2),seasonal=list(order=c(2,1,1), period=12))
res.m4 = residuals(m4);
par(mfrow=c(1,1))
plot(res.m4,xlab='Time',ylab='Residuals',main="Time series plot of the residuals of
    SARIMA(2,0,2)x(2,1,1)12 model")
par(mfrow=c(1,2))
acf(res.m4, lag.max = 12*5, main = "ACF plot of SARIMA(2,0,2)x(2,1,1)12")
pacf(res.m4, lag.max = 12*5, main = "ACF plot of SARIMA(2,0,2)x(2,1,1)12")
par(mfrow=c(1,1))

eacf(res.m2)


res1 = armasubsets(y=res.m2, nar=5, nma=5, y.name='test',ar.method='ols')
plot(res1)
title("BIC table of SARIMA(0,0,0)x(2,1,1)12", line = 6)


############################################################
#MODEL FITTING

m2_102.landing = arima(cptss,order=c(1,0,2),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_102.landing)
residual.analysis(model = m2_102.landing)
##############
m2_202.landing = arima(cptss,order=c(2,0,2),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_202.landing)
residual.analysis(model = m2_202.landing)

##################
m2_002.landing = arima(cptss,order=c(0,0,2),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_002.landing)
residual.analysis(model = m2_002.landing)

##############
m2_003.landing = arima(cptss,order=c(0,0,3),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_003.landing)
residual.analysis(model = m2_003.landing)

##############
m2_100.landing = arima(cptss,order=c(1,0,0),seasonal=list(order=c(2,1,1), period=12),method = "ML")
```

```
coeftest(m2_100.landing)
residual.analysis(model = m2_100.landing)

##############
m2_101.landing = arima(cptss,order=c(1,0,1),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_101.landing)
residual.analysis(model = m2_101.landing)


sc.AIC=AIC(m2_102.landing , m2_002.landing,m2_003.landing, m2_100.landing, m2_202.landing,
m2_101.landing)
sort.score(sc.AIC, score = "aic")
sc.BIC=AIC(m2_102.landing , m2_002.landing,m2_003.landing, m2_100.landing, m2_202.landing,
m2_101.landing, k = log(length(cpts)))
sort.score(sc.BIC, score = "aic")

# Best model SARIMA(1,0,0)x(2,1,1)_12
# Over-parameterizations: SARIMA(2,0,0)x(2,1,1)_12 and SARIMA(1,0,1)x(2,1,1)_12

m2_200.landing = arima(cptss,order=c(2,0,0),seasonal=list(order=c(2,1,1), period=12),method = "ML")
coeftest(m2_200.landing)
residual.analysis(model = m2_200.landing)


########################
#PREDICTION

m2_100.pred= Arima(cpts,order=c(1,0,0),seasonal=list(order=c(2,1,1), period=12), lambda = 0)
future = forecast(m2_100.pred, h = 10)
plot(future)
accuracy(m2_100.pred)
```