Time Series Forecasting for a Inbound Call Center

Student Number: s3398979 and s3730405

Student Name: Vijeta Tulsiyan & Reshma Taruni Gadala

1 June 2019

Aim of the Project

We analysed the total inbound calls received by a call center of Health Service System and forecasted the monthly calls volume for the next ten years.

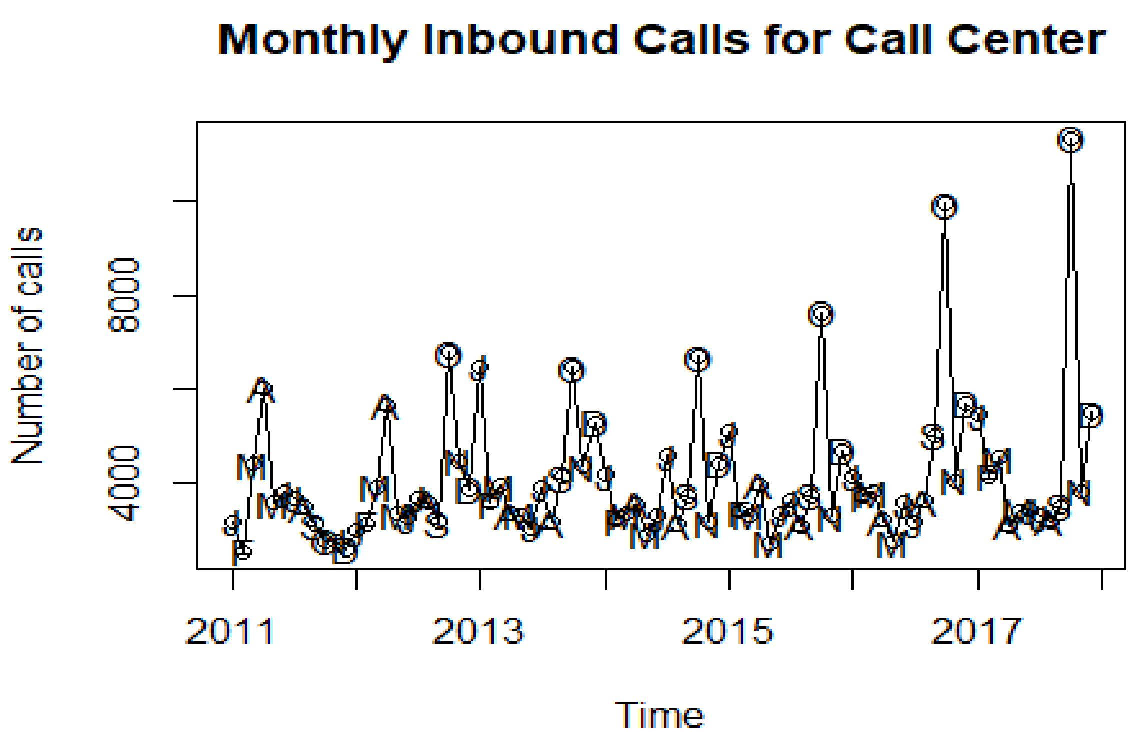
Data Description and Visual Analysis

The data for the total inbound calls received is available on a monthly basis. The time series ranges from January 2011 to December 2017. There are 6 columns and 84 observations. We are analysing & forecasting column "Inbound Calls" from the dataset.

We plotted the time series graph for the total inbound calls received and noticed the following:
- Existence of somewhat upward trend.
- Seasonality is very clear as the calls are highest for the month of October and lowest for May between 2013 - 2017.
- Changing variance is not very obvious from the plot.
- There is no intervention point.
- There is autoregressive behaviour as neighbouring values are close to each other suggesting possible auto correlation.
- Moving average may be possible as seen from the movement of line plot around the mean for the given data.

```
# Read the data
Call_Center <- read_csv("Call_Center_Metrics.csv")
#Call_Center <- read_csv("C:/Users/Vijeta Tulsiyan/Desktop/RMIT/Semester
4/Time Series Analysis/Final
Project/Call_Center_Metrics_for_the_Health_Service_System.csv")
head(Call_Center, 5)

## # A tibble: 5 x 6
##   Month `Inbound Calls` `Average Speed ~ `Abandoned Call~ `Call Abandonme~
##   <chr>           <int>            <int>            <int>            <dbl>
## 1 10/0~           11924               28              217            1.79
## 2 09/0~            5205               18               63            1.20
## 3 08/0~            4953               14               37            0.740
## 4 07/0~            4361               13               43            0.980
```

```
## 5 06/0~              4507                11               40           0.880
## # ... with 1 more variable: `In-person visits` <int>

Call_Center <-  Call_Center[order(as.Date(Call_Center$Month,
format="%m/%d/%Y")),]

inbound <- as.vector(Call_Center$`Inbound Calls`)
class(inbound)

## [1] "integer"

dim(Call_Center)

## [1] 94  6

inbound_calls <- ts(inbound, start = c(2011,1), end=c(2017,12), frequency =
12)
class(inbound_calls)

## [1] "ts"

plot(inbound_calls,type='o',ylab='Number of calls', main = 'Monthly Inbound
Calls for Call Center')
points(y=inbound_calls,x=time(inbound_calls),pch=as.vector(season(inbound_cal
ls)))
```



Dealing with Seasonality

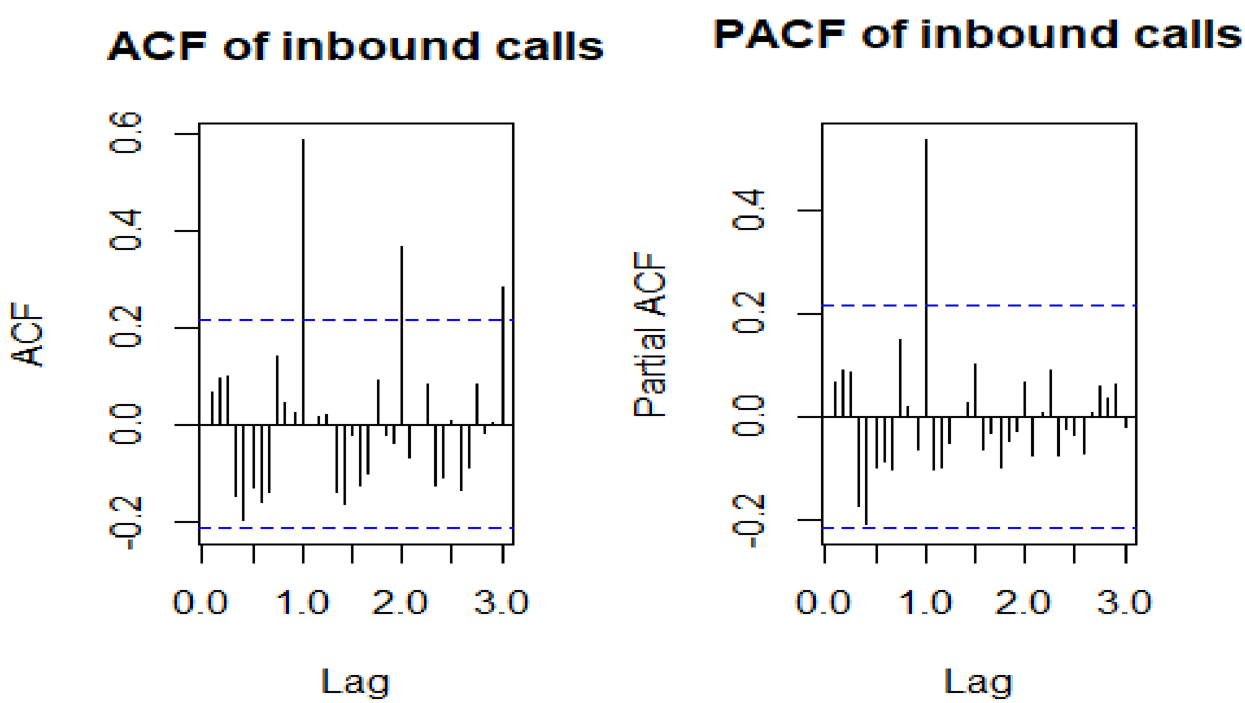
We plotted the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) of the time series plot for the total inbound calls. We found that ACF has clear downward pattern at seasonal lags 1, 2 and 3. This suggested existence of seasonality.

We applied first seasonal differencing D = 1 and plotted the ACF and PACF of the residuals.Both ACF & PACF has no pattern at seasonal lags 1, 2 and 3. ACF has no siginificant seasonal lag, so we took Q = 0. PACF has one seasonal lag at 1, so we took P = 1.

We applied SARIMA (0,0,0)*(1,1,0)_12 model to the given time series and plotted the ACF and PACF of the residuals after seasonal differencing D = 1 and P =1. We observed, ACF & PACF both had no seasonal pattern or seasonal lags at 1, 2 and 3. As, there are no significant correlation at seasonal lags of both ACF and PACF, we can conclude that the seasonality is filtrated out.

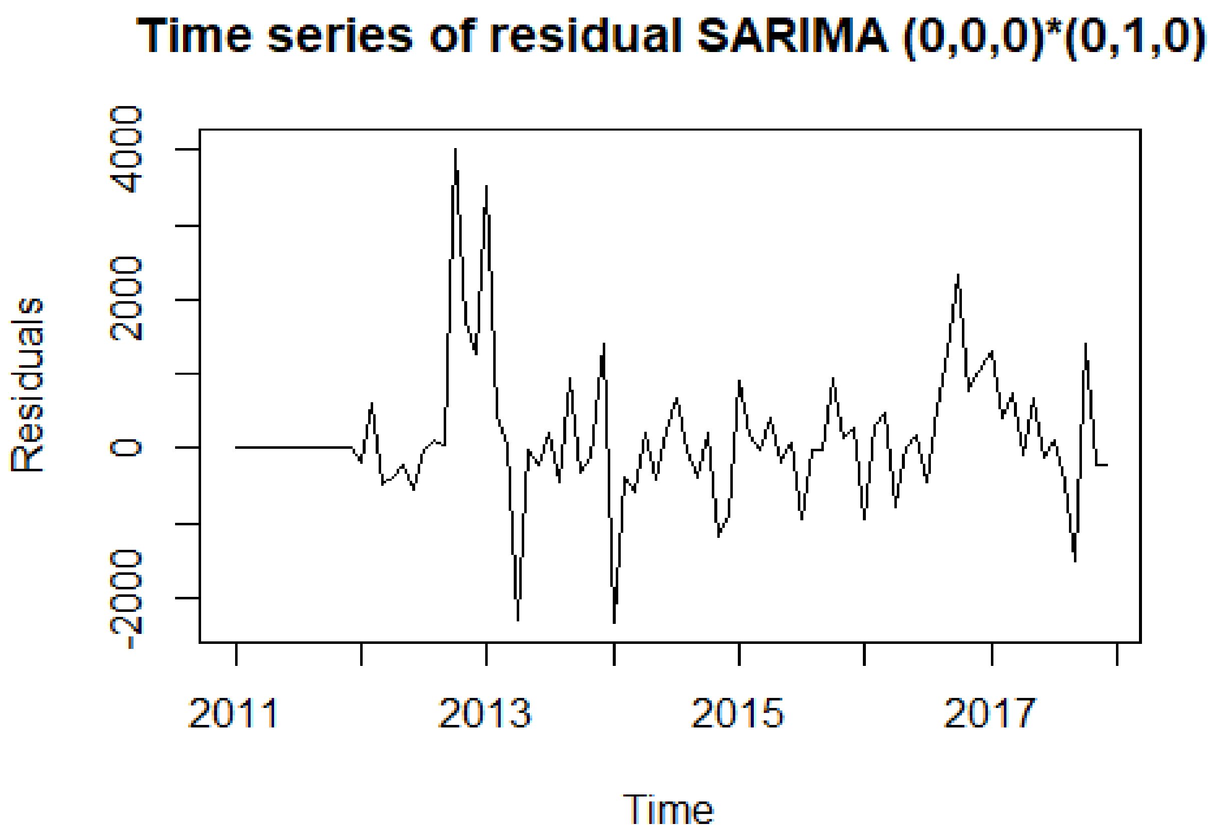
However, the ordinary part has a downward pattern in ACF so we will apply transformation and ordinary differencing and study the ACF and PACF for residuals.

```
par(mfrow=c(1,2))
acf(inbound_calls,  lag.max = 36,main="ACF of inbound calls")
pacf(inbound_calls,  lag.max = 36,main="PACF of inbound calls")
```

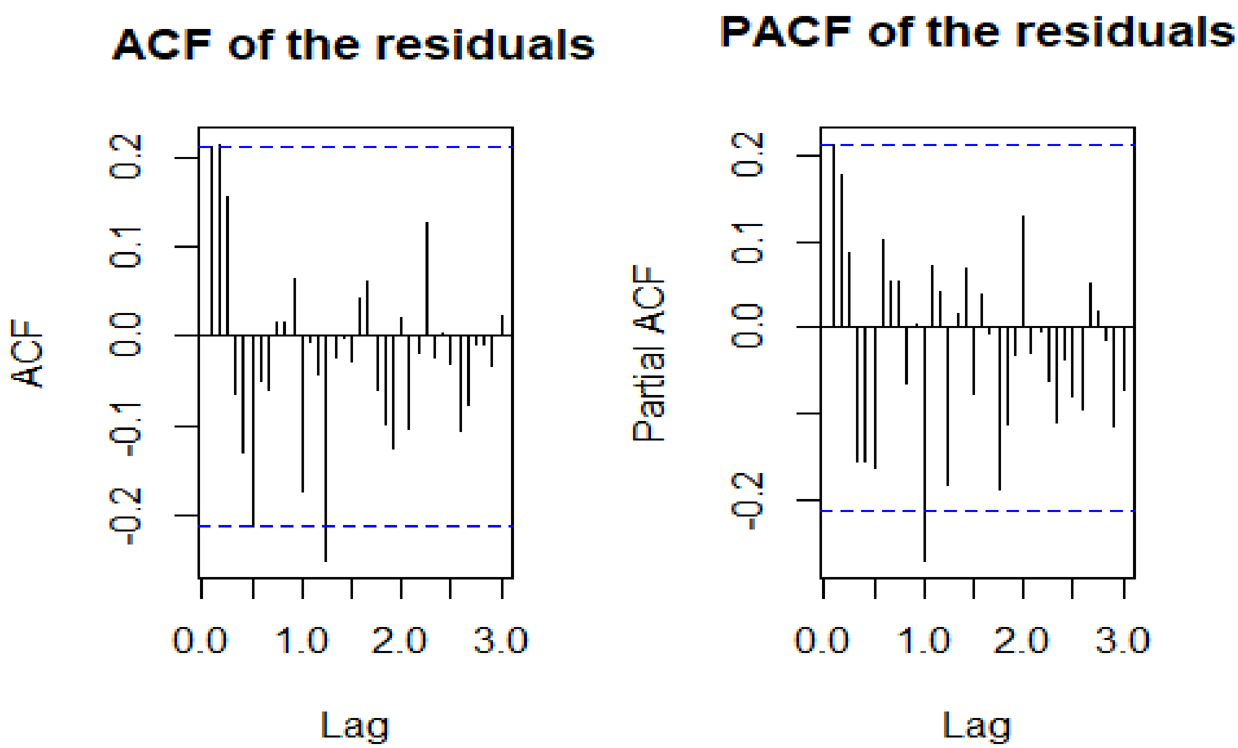



```
par(mfrow=c(1,1))

#The effect of seasonal trend is seen in the ACF. So, we will fit a plain
```

```r
model only with D = 1 and inspect the residuals.
m1.inbound <-
arima(inbound_calls,order=c(0,0,0),seasonal=list(order=c(0,1,0), period=12))
res.m1 <-  residuals(m1.inbound)
plot(res.m1,xlab='Time',ylab='Residuals',main="Time series of residual SARIMA
(0,0,0)*(0,1,0)")
```

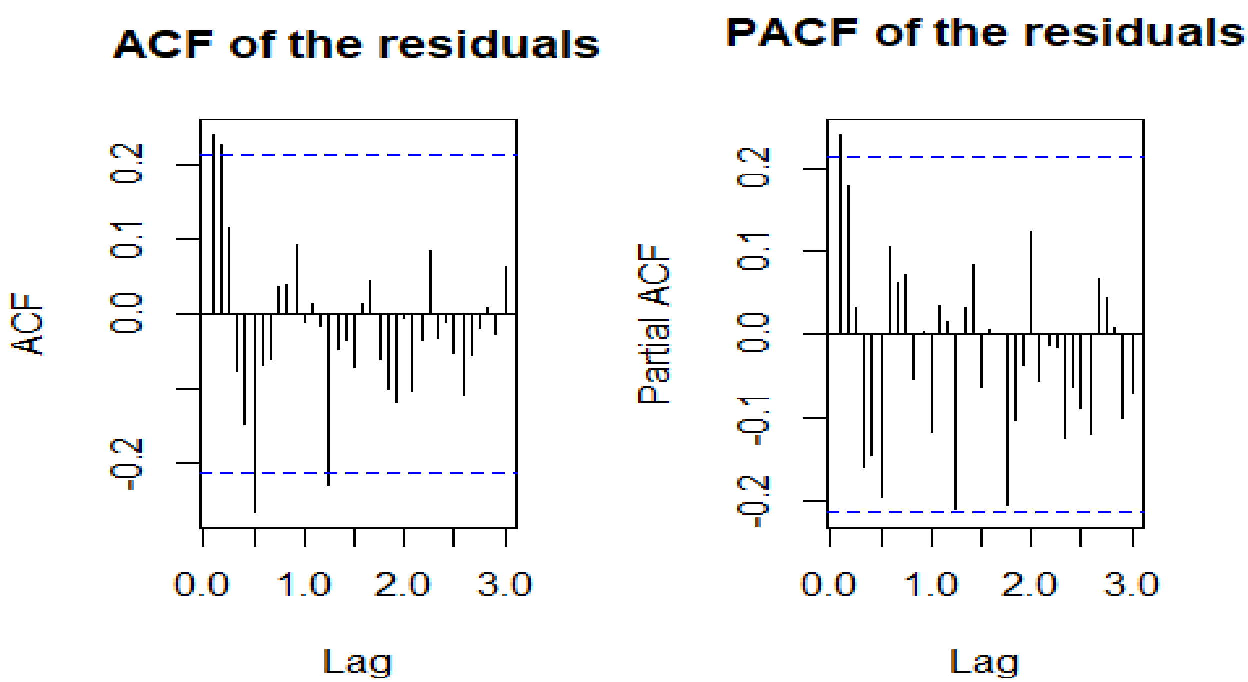



```r
par(mfrow=c(1,2))
acf(res.m1, lag.max = 36, main = "ACF of the residuals")
pacf(res.m1, lag.max = 36, main = "PACF of the residuals")
```




```
# ACF and PACF no pattern, Q =0 as no significant lag, P=1 as one significant
lag at 1.0
#The effect of seasonal trend has been filtrated out and the behaviour of a
seasonal component is clear

# From the time series plot, we can conclude that we got rid of the trend.
Seasonal autocorrelations are seen clearly in
# ACF and PACF now at the lags corresponding to the periods.
# We have one significant correlation at the first seasonal lag in only PACF.

#P=1, D=1, Q=0
m2.inbound <-
arima(inbound_calls,order=c(0,0,0),seasonal=list(order=c(1,1,0), period=12))
res.m2 = residuals(m2.inbound)
par(mfrow=c(1,1))
plot(res.m2,xlab='Time',ylab='Residuals',main="Time series plot of the
residuals")
```


```r
par(mfrow=c(1,2))
acf(res.m2, lag.max = 36, main = "ACF of the residuals")
pacf(res.m2, lag.max = 36, main = "PACF of the residuals")
```



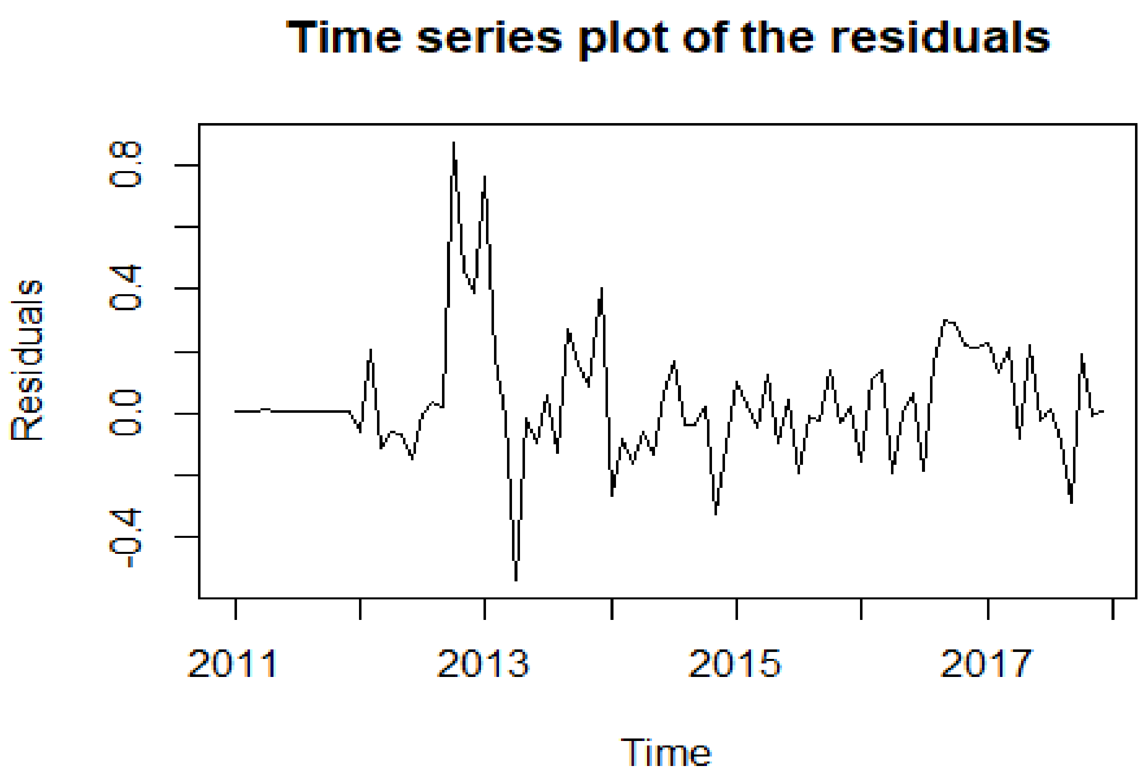

```r
par(mfrow=c(1,1))
```

Finding and dealing Non-stationarity

We took log transformation of the inbound calls time series and applied ordinary differencing d = 1. We visualised the residual for model SARIMA(0,1,0)x(1,1,0)_12 and found that the series seems to be stationary.

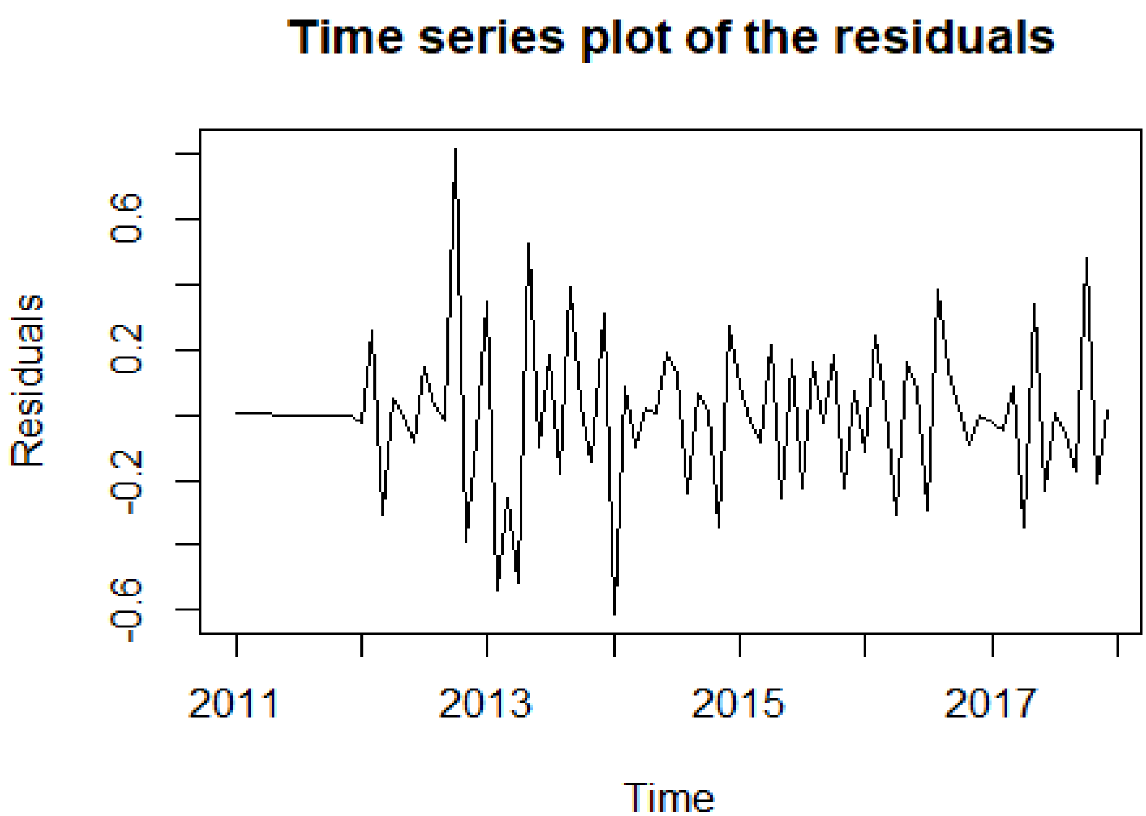
Before selecting the set of possble models, we ensured that the time series is stationary by running Augmented Dickey-Fuller Test (ADF test) on the residuals of the model SARIMA(0,1,0)x(1,1,0)_12. The p-value at 0.01 is less than 0.05, so we rejected the null hypothesis. ADF test confirmed that the series is stationary.

```
# Using log transformation
log.inbound =
arima(log(inbound_calls),order=c(0,0,0),seasonal=list(order=c(1,1,0),
period=12))
res.log = residuals(log.inbound)
par(mfrow=c(1,1))
plot(res.log,xlab='Time',ylab='Residuals',main="Time series plot of the
residuals")
```




```
#we need to take the first ordinary difference to get rid of this trend
effect before going on with the specification of ARMA orders.
m3.inbound =
arima(log(inbound_calls),order=c(0,1,0),seasonal=list(order=c(1,1,0),
```

```
period=12))
res.m3 = residuals(m3.inbound)
par(mfrow=c(1,1))
plot(res.m3,xlab='Time',ylab='Residuals',main="Time series plot of the
residuals")
```




```
# ADF Test
adf.test(res.m3)

##
##  Augmented Dickey-Fuller Test
##
## data:  res.m3
## Dickey-Fuller = -5.1237, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

Model Specification

We then analysed the ACF and PACF plot for the residual of the model SARIMA(0,1,0)x(1,1,0)_12. We found no pattern in ACF and PACF. There was 1 siginificant ordinary lag in ACF, so q = 1. For PACF, there is 1 significant lag and 2, 3 ordinary lags in PACF, so p = 1,2,3.

The set of possible models from ACF and PACF plot for the residuals of the model SARIMA(0,1,0)x(1,1,0)_12: SARIMA(1,1,1)x(1,1,0)_12 SARIMA(2,1,1)x(1,1,0)_12 SARIMA(3,1,1)x(1,1,0)_12

We applied the extended autocorrelation function (EACF) on the residual of model SARIMA(0,1,0)x(1,1,0)_12 to identify the orders of ordinary autoregressive and moving average components of this model. We took lower orders for pand q: p = 1, 2 q = 2, 3

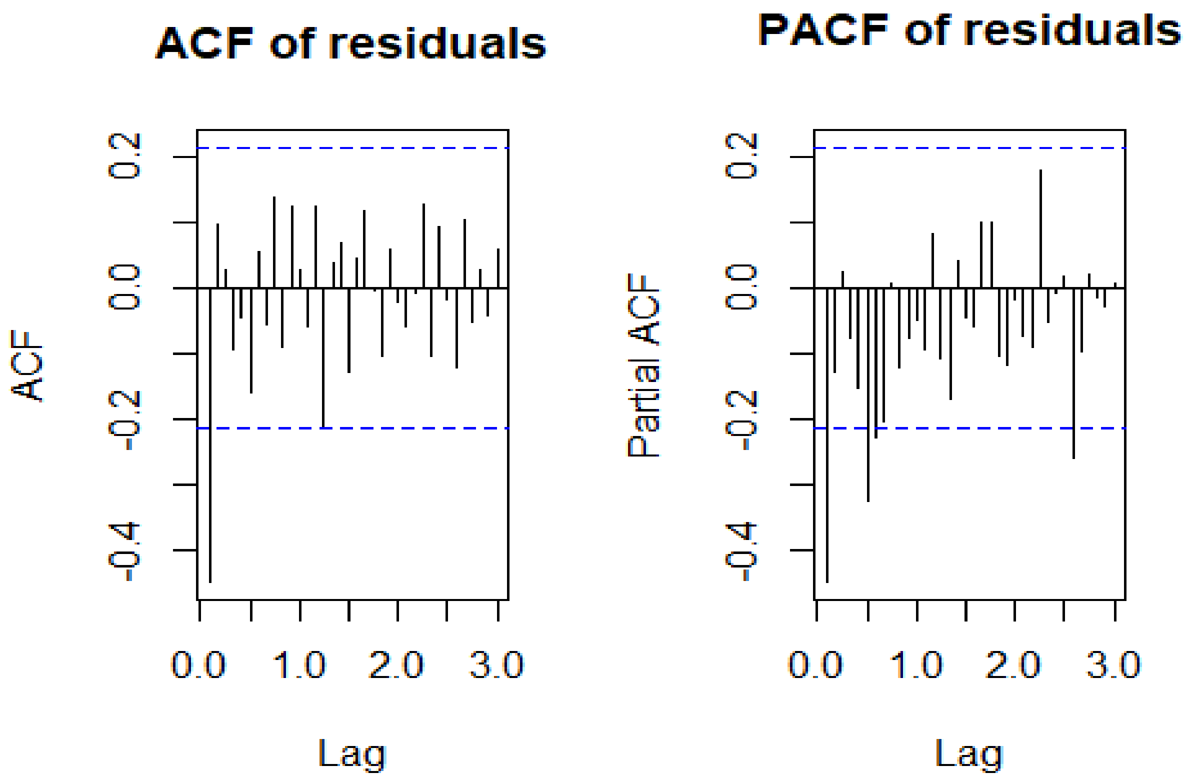
So, the set of possible modes from EACF are: SARIMA(1,1,2)x(1,1,0)_12 SARIMA(1,1,3)x(1,1,0)_12 SARIMA(2,1,3)x(1,1,0)_12

Bayesian Information Criterion (BIC) for model selection was applied on the residuals of SARIMA(0,1,0)x(1,1,0)_12 to determine the order of p and q. p = 1, 4, 6 q = 1, 5

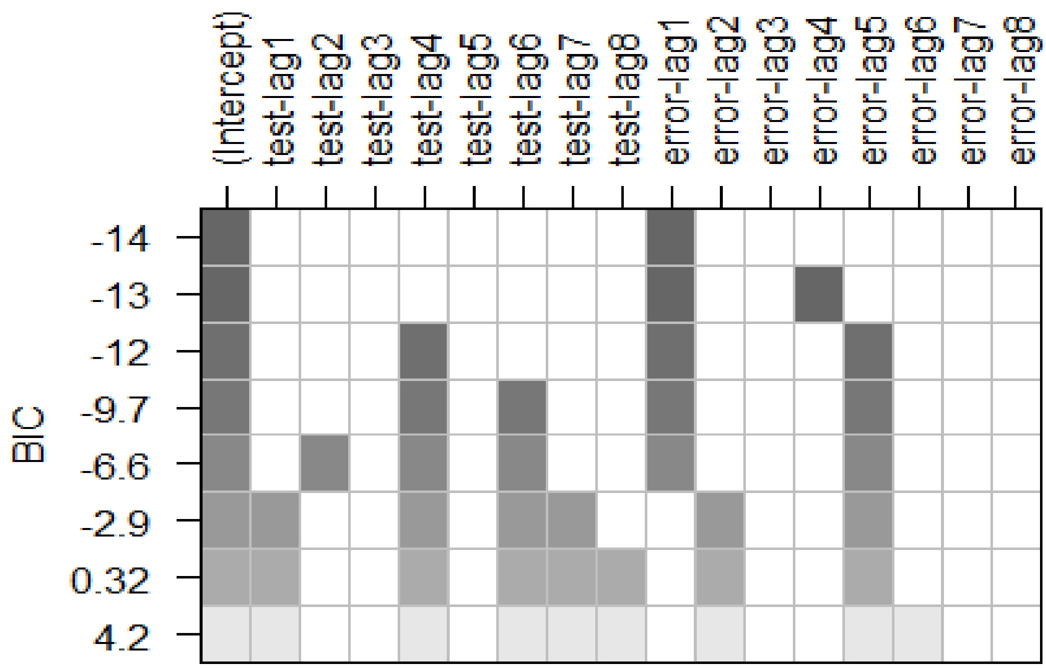
So, the set of possible modes from BIC are: SARIMA(1,1,1)x(1,1,0)_12 SARIMA(1,1,5)x(1,1,0)_12 SARIMA(4,1,1)x(1,1,0)_12 SARIMA(4,1,5)x(1,1,0)_12 SARIMA(6,1,1)x(1,1,0)_12

There are set of 9 unique possible models from ACF & PACF, EACF and BIC: SARIMA(1,1,1)x(1,1,0)_12 SARIMA(1,1,2)x(1,1,0)_12 SARIMA(1,1,3)x(1,1,0)_12 SARIMA(1,1,5)x(1,1,0)_12 SARIMA(2,1,1)x(1,1,0)_12 SARIMA(2,1,3)x(1,1,0)_12 SARIMA(3,1,1)x(1,1,0)_12 SARIMA(4,1,1)x(1,1,0)_12 SARIMA(4,1,5)x(1,1,0)_12 SARIMA(6,1,1)x(1,1,0)_12 SARIMA(6,1,5)x(1,1,0)_12

```r
par(mfrow=c(1,2))
acf(res.m3, lag.max = 36, main = "ACF of residuals")
pacf(res.m3, lag.max = 36, main = "PACF of residuals")
```




```r
par(mfrow=c(1,2))

#EAC
eacf(res.m3)

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

#BIC
par(mfrow=c(1,1))
res = armasubsets(y=(res.m3),nar=8,nma=8,y.name='test',ar.method='ols')
plot(res)
```



Parameter Estimation

We applied Maximum Likelihood estimation method for estimating the parameters of the set of possible models. From the coefficents results, we checked the significance of AR and MA components for various models.

SARIMA(1,1,1)x(1,1,0)_12

For SARIMA(1,1,1)x(1,1,0)_12 model both ar(1) and ma(1) components are significant,so we have done residuals check for this model . Time series plot of the residuals show the series is stationary , ACF and PACF plots show white noise.

SARIMA(1,1,2)x(1,1,0)_12

In this model ar1 and ar2 are siginificant but ma2 is not significant. So, we can ignore this model.

SARIMA(1,1,3)x(1,1,0)_12

In SARIMA(1,1,3)x(1,1,0)_12 model only ma2 and ma3 components are slightly significant ,so we are ignoring this model.

SARIMA(1,1,5)x(1,1,0)_12

For this model ma5 component didnt give any significant result.

SARIMA(2,1,1)x(1,1,0)_12

For this model ar2 component is not significant.

SARIMA(2,1,3)x(1,1,0)_12

For SARIMA(2,1,3)x(1,1,0)_12 model as all ar and ma components are siginificant. Time series plot of the residuals show the series is stationary , ACF and PACF plots show white noise.

SARIMA(3,1,1)x(1,1,0)_12

Coefficients were not significant for this model.

SARIMA(4,1,1)x(1,1,0)_12

ar(4)component in this model was not that significant.

SARIMA(4,1,5)x(1,1,0)_12

All the coefficients are significant for this model.Time series plot of the residuals show the series is stationary. There are no significant lags in ACF ,but PACF plot has one slightly significant lag.

SARIMA(6,1,1)x(1,1,0)_12

ar6 component in this model is not significant.

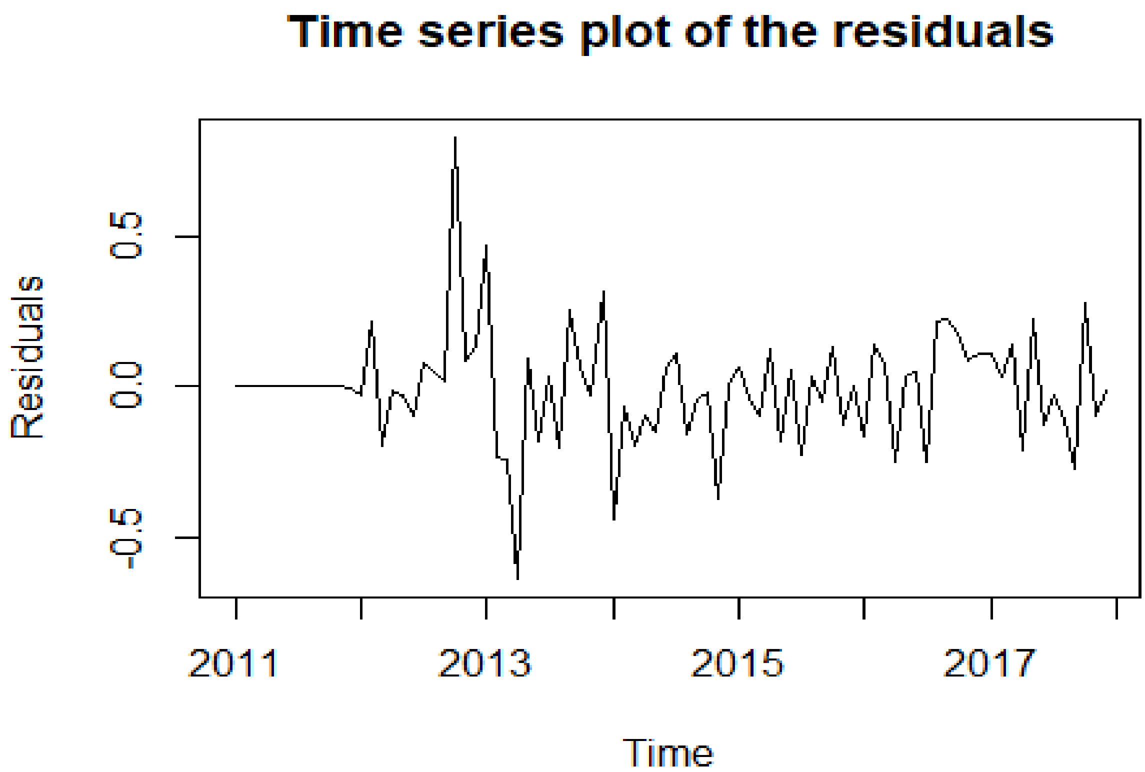
SARIMA(6,1,5)x(1,1,0)_12

Coefficients were not significant for this model.

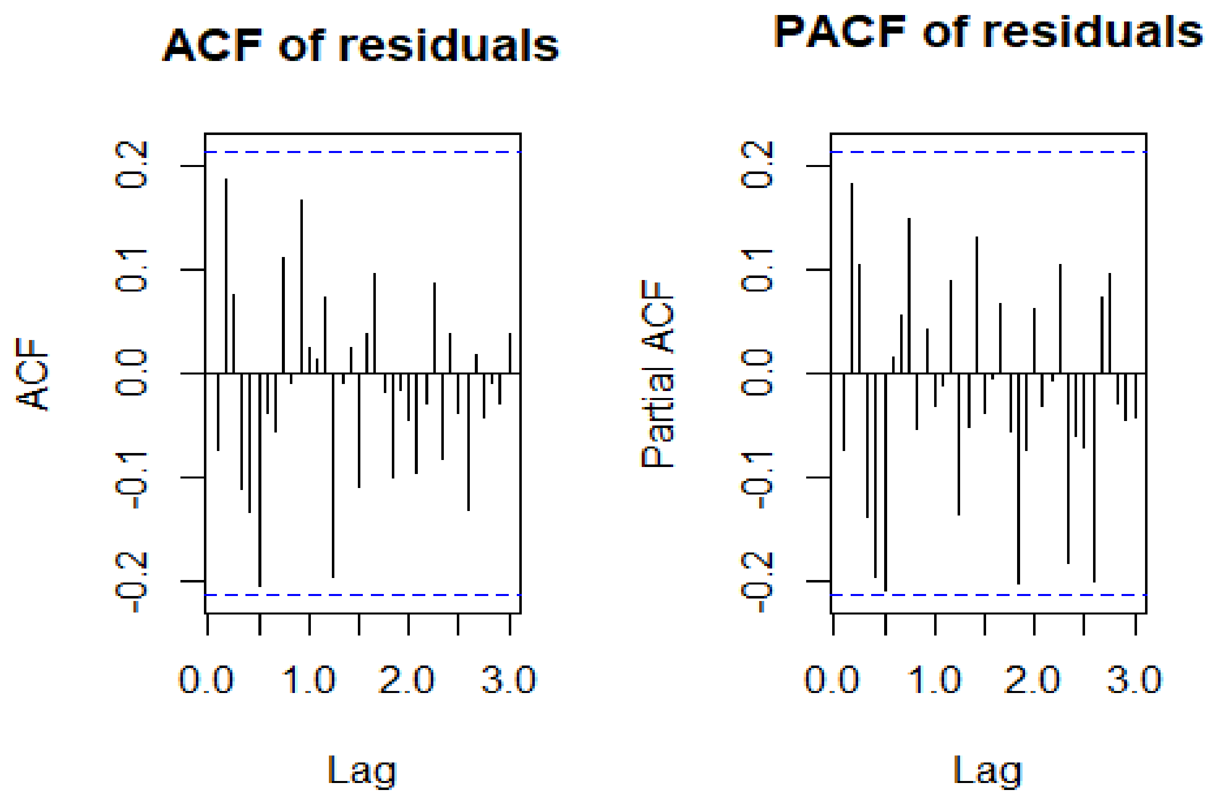
Using sort.score() AIC and BIC values are checked for models were all the coefficients are sigificant. AIC shows SARIMA(2,1,3)x(1,1,0)_12 is the best model and BIC shows SARIMA(1,1,1)x(1,1,0)_12 is the best model. We got mixed results.

```
# SARIMA(1,1,1)x(1,1,0)_12
m1_111.calls =
arima(log(inbound_calls),order=c(1,1,1),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m1_111.calls)

##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1   0.349580   0.114405   3.0556  0.002246 **
## ma1  -0.999995   0.050625 -19.7531 < 2.2e-16 ***
## sar1 -0.306249   0.125281  -2.4445  0.014505 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res.m1 = residuals(m1_111.calls)
par(mfrow=c(1,1))
plot(res.m1,xlab='Time',ylab='Residuals',main="Time series plot of the
residuals")
```




```
par(mfrow=c(1,2))
acf(res.m1, lag.max = 36, main = "ACF of residuals")
pacf(res.m1, lag.max = 36, main = "PACF of residuals")
```




```
# Result : Both ar(1) and ma(1) are significant here.

# SARIMA(1,1,2)x(1,1,0)_12
m2_112.calls =
arima(log(inbound_calls),order=c(1,1,2),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m2_112.calls)

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1   0.60799    0.17655  3.4438 0.0005736 ***
## ma1  -1.28083    0.19612 -6.5307 6.546e-11 ***
## ma2   0.28083    0.19117  1.4690 0.1418391
## sar1 -0.29141    0.12552 -2.3217 0.0202506 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Result: ar1 and ar2 are siginificant but ma2 is not significant. So, we can
ignore this model.

# SARIMA(1,1,3)x(1,1,0)_12
m3_113.calls =
```

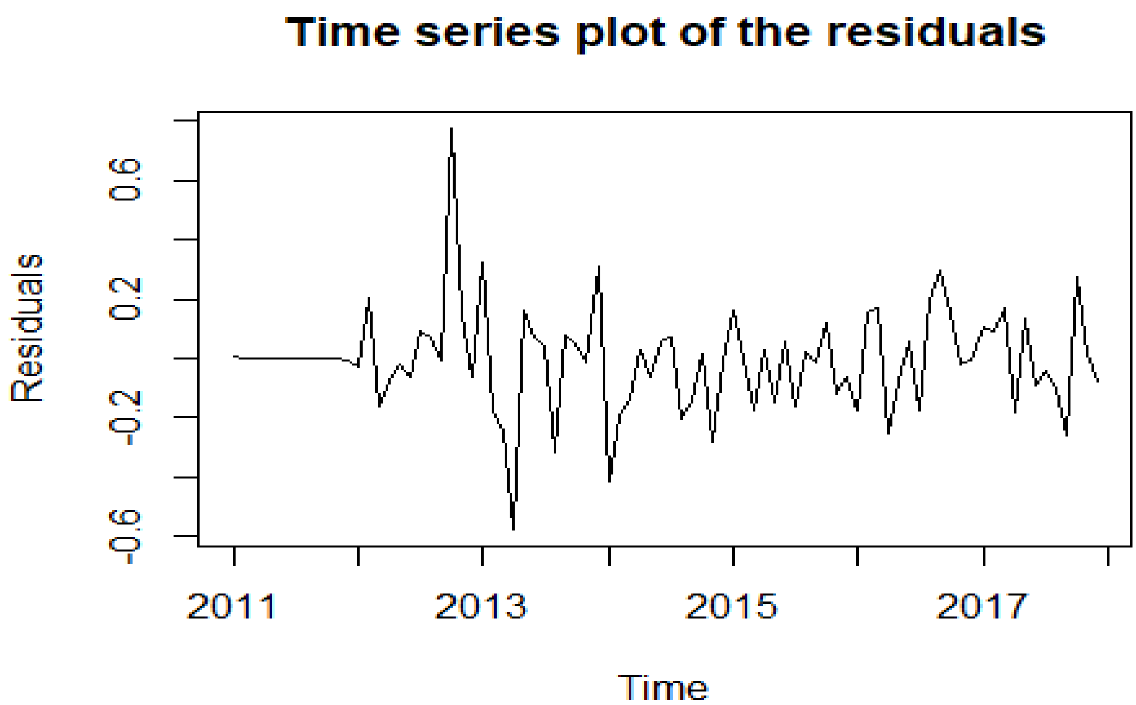
```
arima(log(inbound_calls),order=c(1,1,3),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m3_113.calls)

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1   0.48897    0.19402  2.5202   0.01173 *
## ma1  -1.24525    0.21108 -5.8995 3.647e-09 ***
## ma2   0.51445    0.27543  1.8678   0.06179 .
## ma3  -0.26918    0.14126 -1.9056   0.05670 .
## sar1 -0.28529    0.12751 -2.2374   0.02526 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Result: ma2 and ma3 are only slightly significant so we ignore this model

# SARIMA(1,1,5)x(1,1,0)_12
m4_115.calls =
arima(log(inbound_calls),order=c(1,1,5),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m4_115.calls)

##
## z test of coefficients:
##
##         Estimate  Std. Error z value Pr(>|z|)
## ar1   0.03634422  0.47610326  0.0763  0.93915
## ma1  -0.78389739  0.46369627 -1.6905  0.09092 .
## ma2   0.15948390  0.39402297  0.4048  0.68566
## ma3   0.00027971  0.17807319  0.0016  0.99875
## ma4  -0.23368801  0.16766721 -1.3938  0.16339
## ma5  -0.14217620  0.21118096 -0.6732  0.50079
## sar1 -0.25741065  0.12311791 -2.0908  0.03655 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Result: ma5 didnt give any significant result. SO we ignore this model.

# SARIMA(2,1,1)x(1,1,0)_12
m5_211.calls =
arima(log(inbound_calls),order=c(2,1,1),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m5_211.calls)

##
## z test of coefficients:
##
##       Estimate Std. Error  z value Pr(>|z|)
## ar1   0.284729   0.118333   2.4062  0.01612 *
```

```
## ar2   0.199961   0.117289   1.7049  0.08822 .
## ma1  -0.999999   0.044944 -22.2500  < 2e-16 ***
## sar1 -0.289845   0.124434  -2.3293  0.01984 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

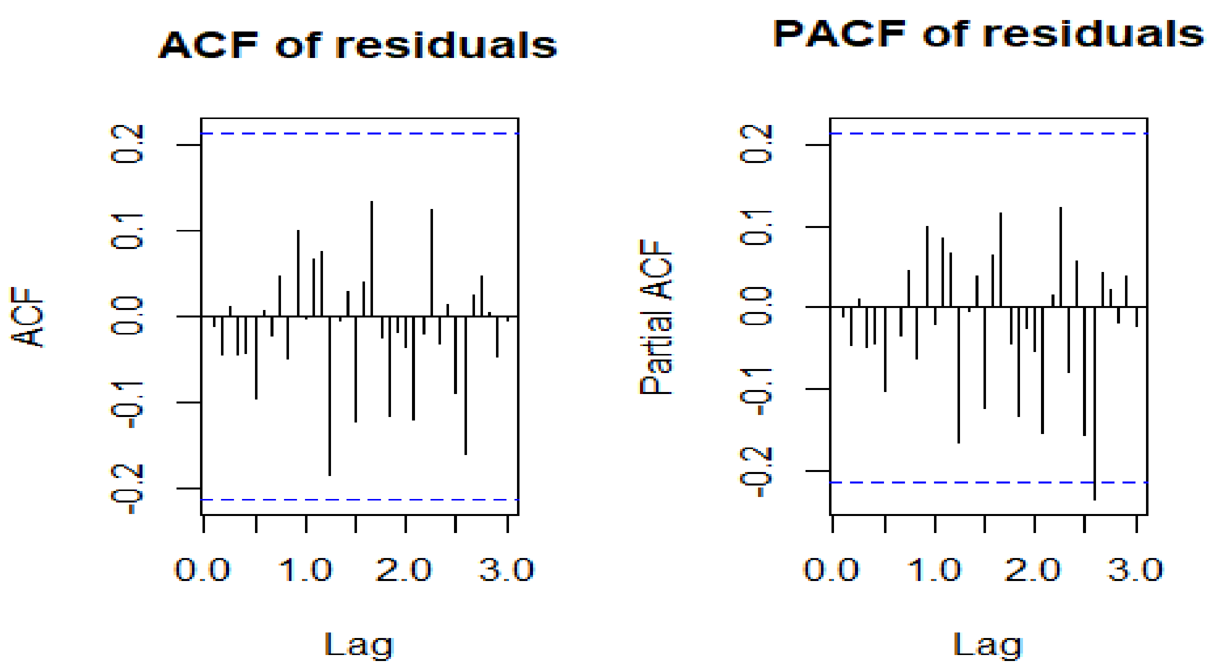
```
#Result: ar2 is not significant. So we ignore this model

# SARIMA(2,1,3)x(1,1,0)_12
m6_213.calls =
arima(log(inbound_calls),order=c(2,1,3),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m6_213.calls)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1   1.08395    0.26336  4.1159 3.857e-05 ***
## ar2  -0.58382    0.17120 -3.4101 0.0006494 ***
## ma1  -1.85822    0.19867 -9.3535 < 2.2e-16 ***
## ma2   1.57246    0.24524  6.4120 1.437e-10 ***
## ma3  -0.71423    0.16770 -4.2590 2.053e-05 ***
## sar1 -0.25896    0.13116 -1.9744 0.0483350 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res.m6 = residuals(m6_213.calls)
par(mfrow=c(1,1))
plot(res.m6,xlab='Time',ylab='Residuals',main="Time series plot of the
residuals")
```




```
par(mfrow=c(1,2))
acf(res.m6, lag.max = 36, main = "ACF of residuals")
pacf(res.m6, lag.max = 36, main = "PACF of residuals")
```

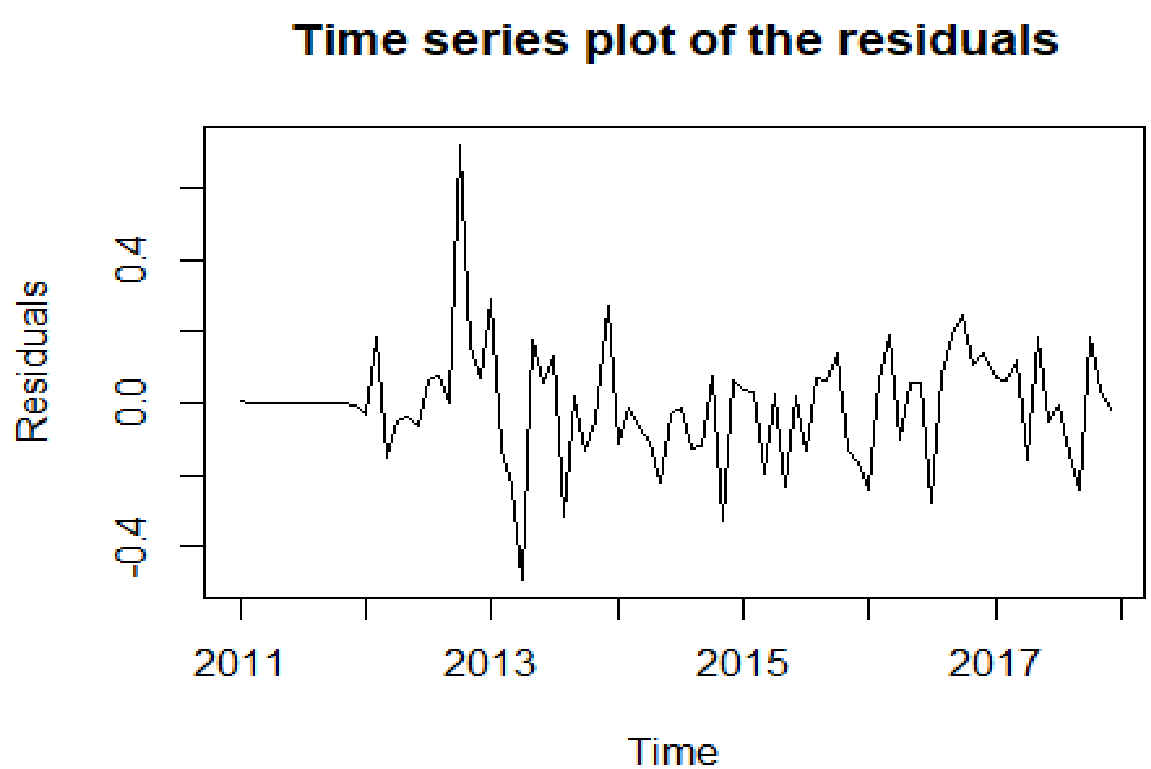


```r
# This is siginificant model as all AR and ma components are siginificant

# SARIMA(3,1,1)x(1,1,0)_12
m7_311.calls =
arima(log(inbound_calls),order=c(3,1,1),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m7_311.calls)

##
## z test of coefficients:
##
##       Estimate Std. Error  z value Pr(>|z|)
## ar1   0.292496   0.121735   2.4027  0.01627 *
## ar2   0.208059   0.121090   1.7182  0.08576 .
## ar3  -0.033465   0.126914  -0.2637  0.79202
## ma1  -0.999999   0.045894 -21.7891  < 2e-16 ***
## sar1 -0.301234   0.131539  -2.2901  0.02202 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Not significant. We ignore this model.


# SARIMA(4,1,1)x(1,1,0)_12
m8_411.calls =
arima(log(inbound_calls),order=c(4,1,1),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m8_411.calls)

##
## z test of coefficients:
##
##       Estimate Std. Error  z value Pr(>|z|)
## ar1   0.281229   0.115818   2.4282  0.01517 *
## ar2   0.261056   0.120577   2.1651  0.03038 *
## ar3   0.030892   0.124333   0.2485  0.80378
## ar4  -0.249647   0.118783  -2.1017  0.03558 *
## ma1  -0.999998   0.058963 -16.9596  < 2e-16 ***
## sar1 -0.315256   0.125115  -2.5197  0.01174 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Result: adding ar4 component was not that significant,so we ignore this
model


# SARIMA(4,1,5)x(1,1,0)_12
m9_415.calls =
arima(log(inbound_calls),order=c(4,1,5),seasonal=list(order=c(1,1,0),
```

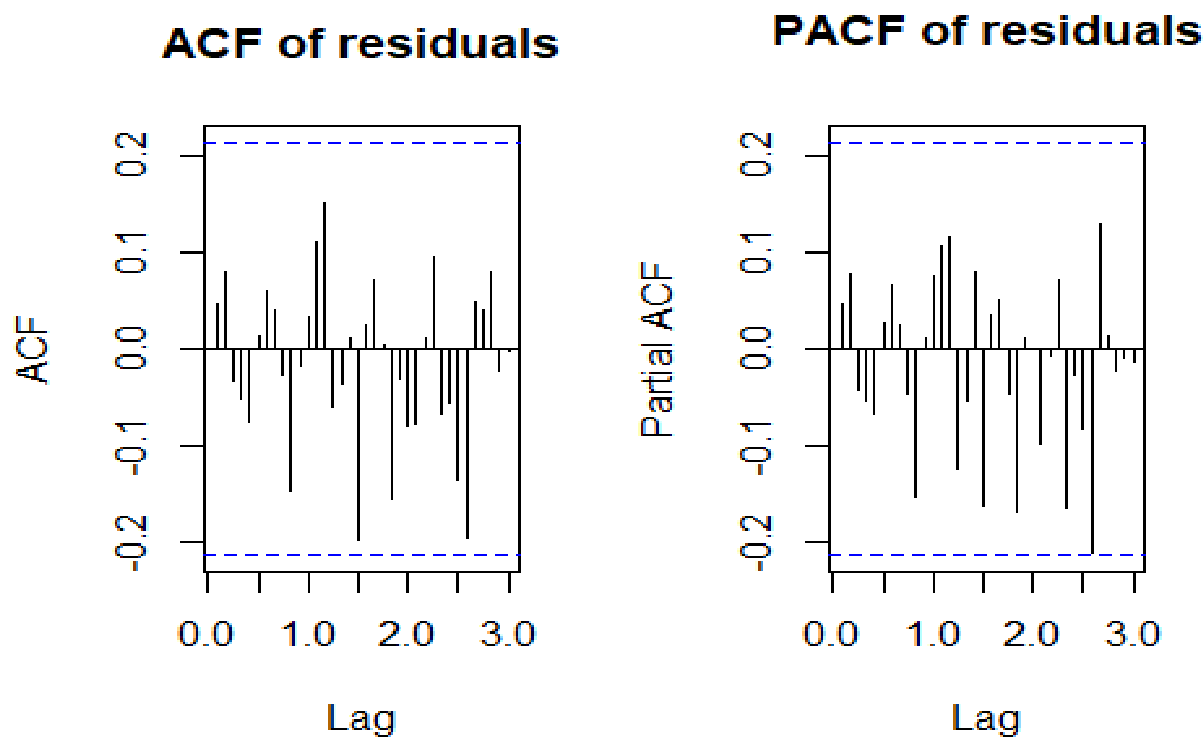
```r
period=12),method = "ML")
coeftest(m9_415.calls)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error  z value  Pr(>|z|)
## ar1  -0.385608   0.068176  -5.6561 1.549e-08 ***
## ar2   0.945973   0.068783  13.7530 < 2.2e-16 ***
## ar3  -0.318437   0.071502  -4.4536 8.445e-06 ***
## ar4  -0.886766   0.069670 -12.7281 < 2.2e-16 ***
## ma1  -0.466141   0.099219  -4.6981 2.626e-06 ***
## ma2  -1.269393   0.096330 -13.1775 < 2.2e-16 ***
## ma3   1.269306   0.120008  10.5769 < 2.2e-16 ***
## ma4   0.466224   0.094073   4.9560 7.196e-07 ***
## ma5  -0.999929   0.103381  -9.6723 < 2.2e-16 ***
## sar1 -0.352219   0.133953  -2.6294  0.008553 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
res.m9 = residuals(m9_415.calls)
par(mfrow=c(1,1))
plot(res.m9,xlab='Time',ylab='Residuals',main="Time series plot of the
residuals")
```



```
par(mfrow=c(1,2))
acf(res.m9, lag.max = 36, main = "ACF of residuals")
pacf(res.m9, lag.max = 36, main = "PACF of residuals")
```




```
#Result:This model is quite siginificant

# SARIMA(6,1,1)x(1,1,0)_12
m10_611.calls =
arima(log(inbound_calls),order=c(6,1,1),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m10_611.calls)

##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ar1   0.098487   0.158149  0.6227  0.533450
## ar2   0.182735   0.127682  1.4312  0.152381
## ar3   0.066199   0.120778  0.5481  0.583620
## ar4  -0.199358   0.116999 -1.7039  0.088396 .
## ar5  -0.208093   0.125564 -1.6573  0.097465 .
## ar6  -0.209365   0.144817 -1.4457  0.148257
## ma1  -0.823007   0.134482 -6.1198 9.368e-10 ***
## sar1 -0.353692   0.124982 -2.8300  0.004655 **
```

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

```
# ar6 is not significant. We ignore this model.


# SARIMA(6,1,5)x(1,1,0)_12
m11_615.calls =
arima(log(inbound_calls),order=c(6,1,5),seasonal=list(order=c(1,1,0),
period=12),method = "ML")
coeftest(m11_615.calls)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value Pr(>|z|)
## ar1   0.256660   0.812527  0.3159   0.7521
## ar2   0.114388   0.708843  0.1614   0.8718
## ar3  -0.331377   0.442531 -0.7488   0.4540
## ar4   0.121014   0.517598  0.2338   0.8151
## ar5  -0.077605   0.182650 -0.4249   0.6709
## ar6  -0.180227   0.220181 -0.8185   0.4131
## ma1  -1.033667   0.805418 -1.2834   0.1994
## ma2   0.181832   1.226504  0.1483   0.8821
## ma3   0.348754   0.856596  0.4071   0.6839
## ma4  -0.647121   0.710832 -0.9104   0.3626
## ma5   0.150216   0.527589  0.2847   0.7759
## sar1 -0.300066   0.165960 -1.8081   0.0706 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# No significant result. We ignore this model.


sc.AIC=AIC(m1_111.calls, m6_213.calls, m9_415.calls )
sc.BIC=AIC(m1_111.calls, m6_213.calls, m9_415.calls, k =
log(length(inbound_calls)))

sort.score(sc.AIC, score = "aic")
```

```
##              df        AIC
## m6_213.calls  7 -13.086235
## m9_415.calls 11 -11.299179
## m1_111.calls  4  -9.316711
```

```
sort.score(sc.BIC, score = "aic")
```

```
##              df        AIC
## m1_111.calls  4  0.4065566
## m6_213.calls  7  3.9294822
## m9_415.calls 11 15.4398058
```

Diagnostic Checking (Residual Analysis and Overfitting)

We have done residual analysis for 3 models were the coefficients are significant.

For model SARIMA(1,1,1)x(1,1,0) time series plot of standardised residuals show that the series is stationary ,histogram is normally distributed,but there is one significant lag in ACF. In QQplot few points are away from the line at the tails .Ljung-Box Test failed as there is one significant point below the line .In Shapiro-Wilk normality test p-value is less than 0.05.

For model SARIMA(2,1,3)x(1,1,0) time series plot of standardised residuals show that the series is stationary ,histogram is nearly normally distributed,there is no significant lag in ACF which indicates the series is white noise.In QQplot few points are away from the line at the tails . Ljung-Box Test looks good as there are no significant points below the line .In Shapiro-Wilk normality test p-value is less than 0.05.

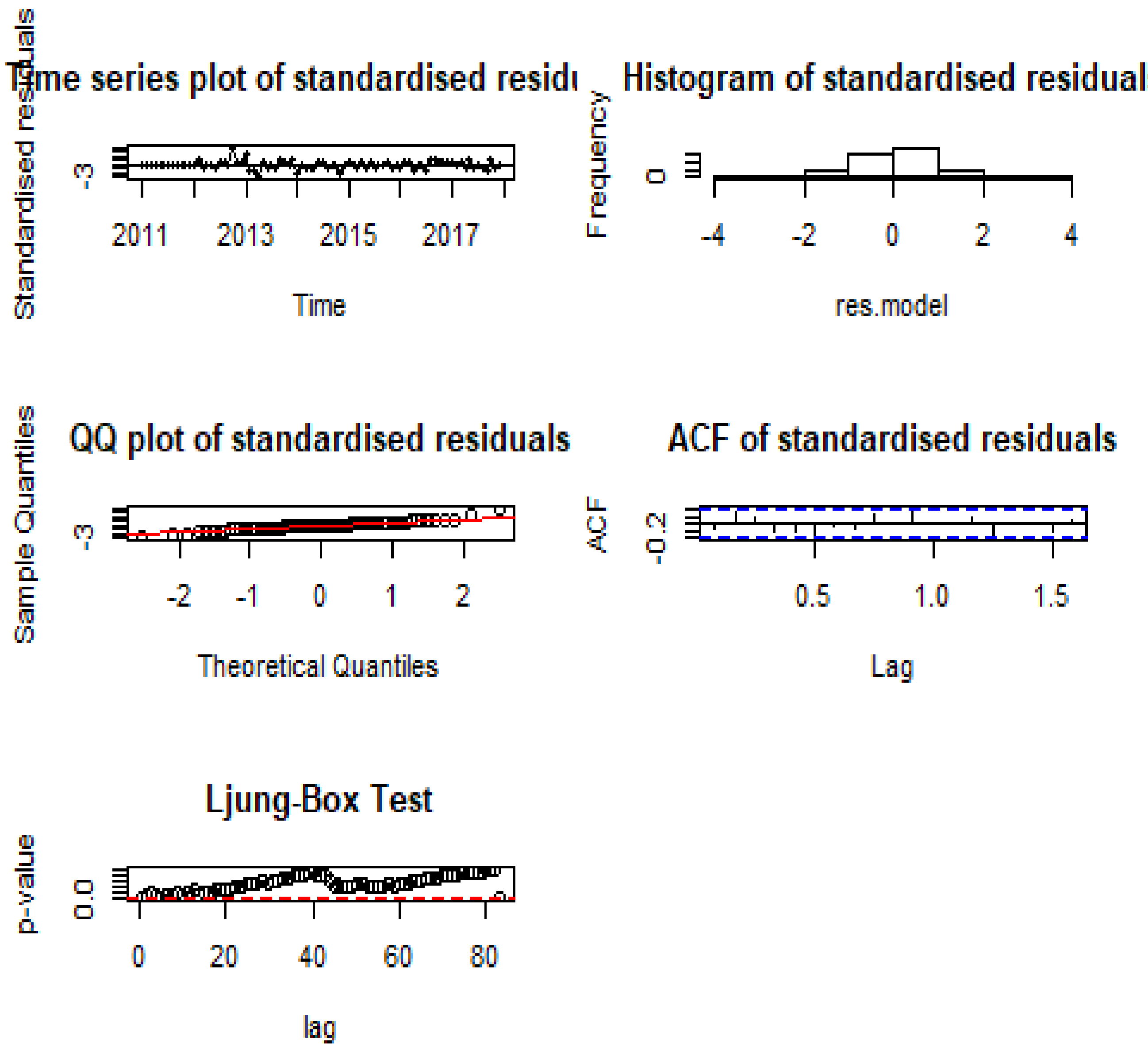
For model SARIMA(4,1,5)x(1,1,0) time series plot of standardised residuals show that the series is stationary ,histogram is nearly normally distributed,but there is one significant lag in ACF .In QQplot few points are away from the line at the tails .Ljung-Box Test failed as there is one significant point touching the line .In Shapiro-Wilk normality test p-value is less than 0.05.

Seems SARIMA(2,1,3)x(1,1,0) is doing better compared to the other 2 models.So we are doing overfitting for this model. Our Overfitting models are SARIMA(3,1,3)x(1,1,0) and SARIMA(2,1,4)x(1,1,0),but coefficients are insignificant for these models.

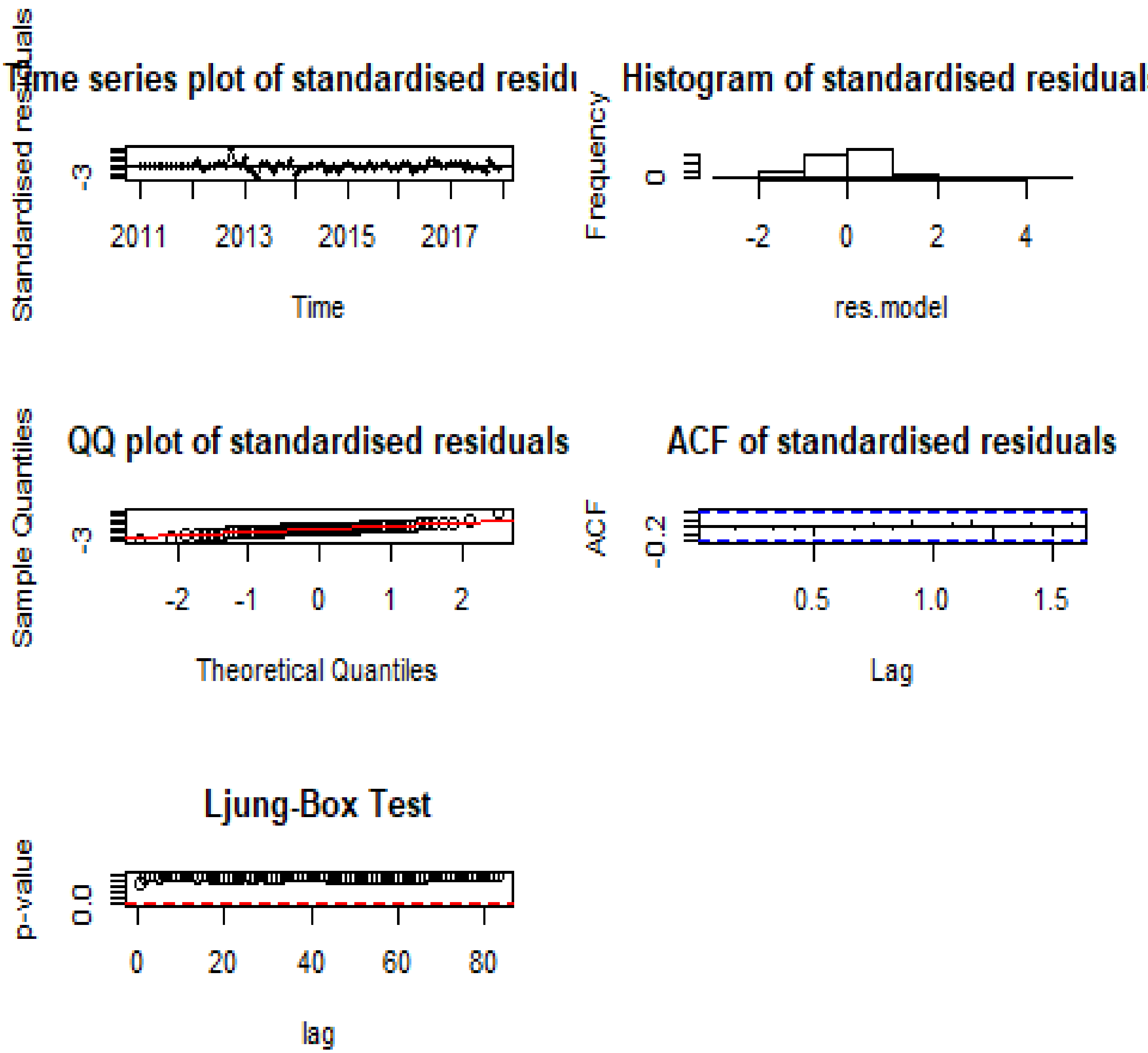
```
residual.analysis(model = m1_111.calls)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.92597, p-value = 0.0001263
```

```
residual.analysis(model = m6_213.calls)
```

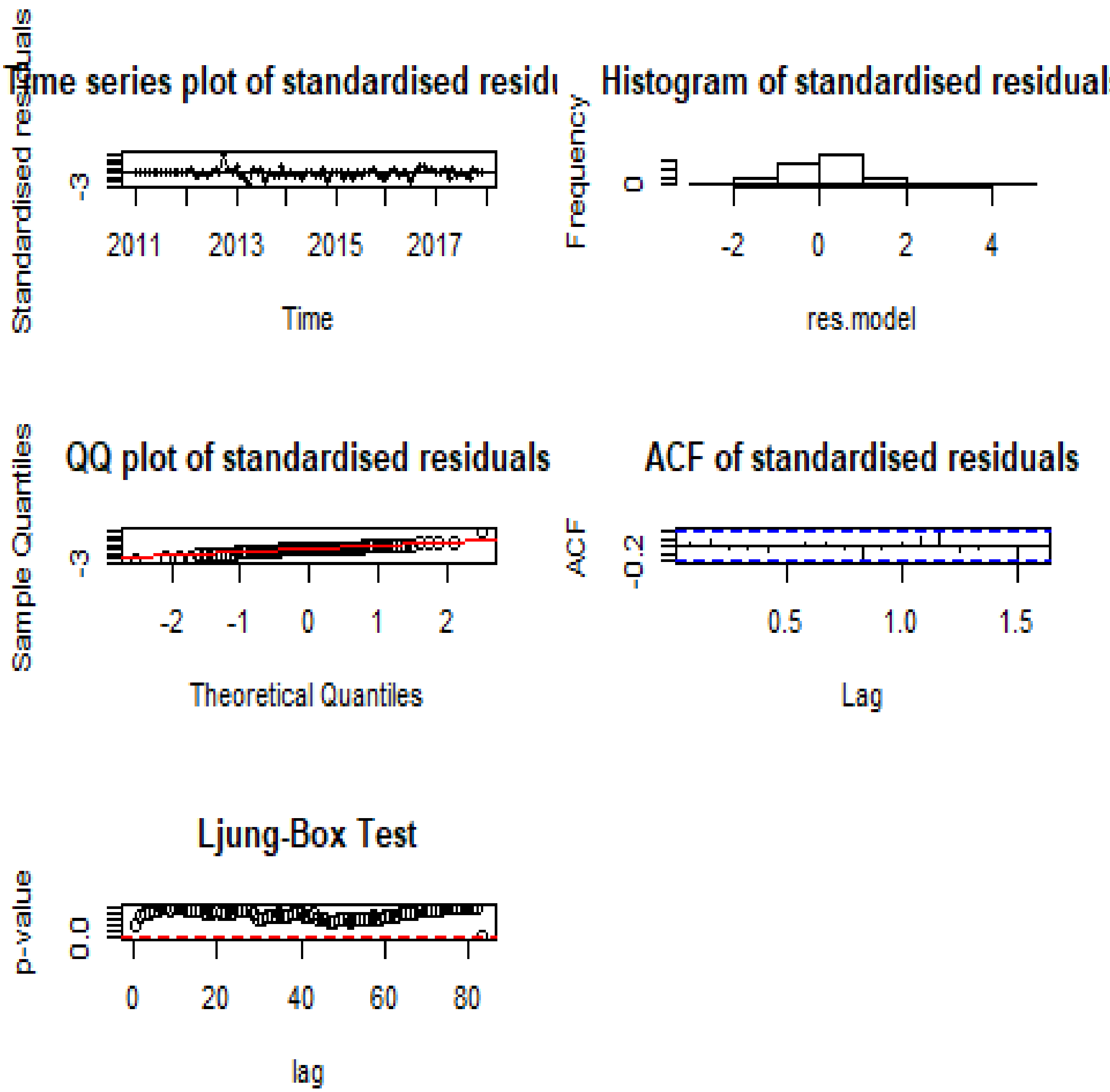



```
##
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.9321, p-value = 0.0002604

residual.analysis(model = m9_415.calls)
```




```
##
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.93522, p-value = 0.0003808

# We go with model m6_213.calls among the above 3 models
# We got the white noise series with SARIMA(2,1,3)x(1,1,0) component.
```

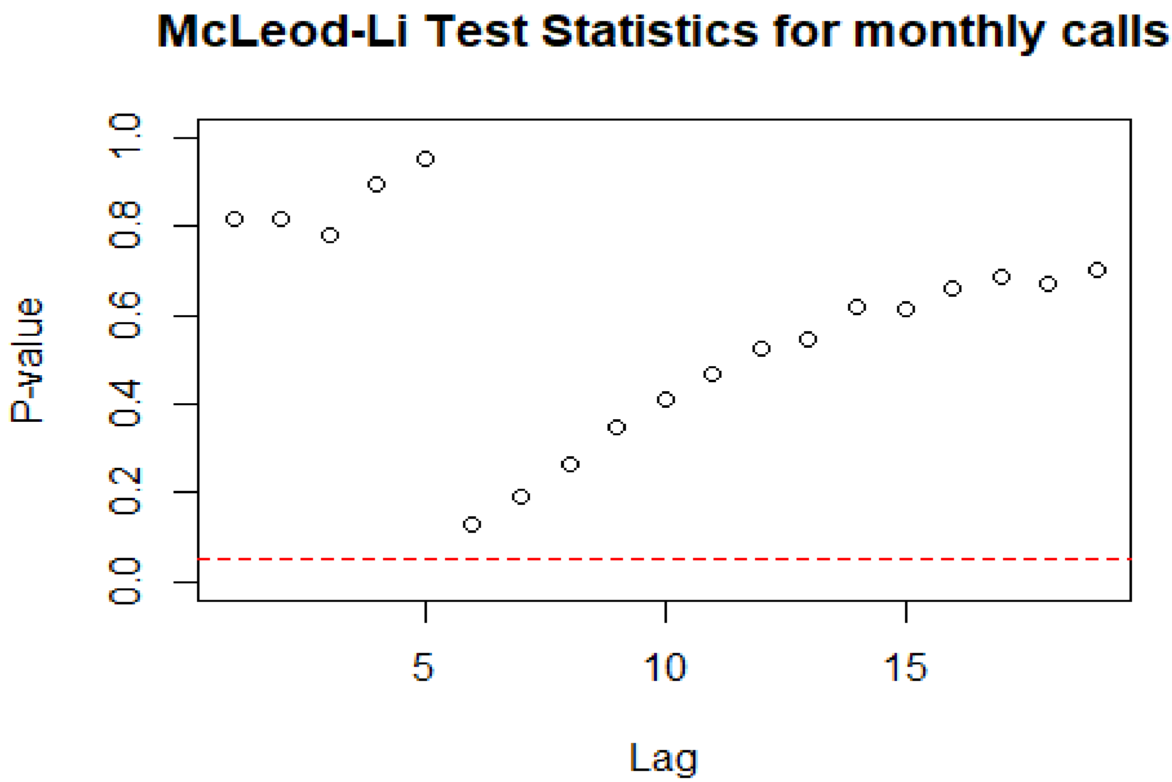
```r
# For overfitting, we will consider SARIMA(3,1,3)x(1,1,0) and
SARIMA(2,1,4)x(1,1,0).

m10_313.calls =
arima(inbound_calls,order=c(3,1,3),seasonal=list(order=c(1,1,0), period=12))
m11_214.calls =
arima(inbound_calls,order=c(2,1,4),seasonal=list(order=c(1,1,0), period=12))

coeftest(m10_313.calls)

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  -0.92349    0.81971 -1.1266   0.25990
## ar2   0.11380    0.59855  0.1901   0.84921
## ar3   0.31517    0.18557  1.6984   0.08943 .
## ma1   0.16223    0.88763  0.1828   0.85498
## ma2  -0.82429    0.16139 -5.1074 3.265e-07 ***
## ma3  -0.33793    0.82161 -0.4113   0.68085
## sar1 -0.25174    0.17830 -1.4119   0.15798
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

coeftest(m11_214.calls)

##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  -0.14245    1.08526 -0.1313   0.8956
## ar2   0.20890    0.73602  0.2838   0.7765
## ma1  -0.67358    1.12678 -0.5978   0.5500
## ma2  -0.25337    1.61897 -0.1565   0.8756
## ma3   0.24436    0.42658  0.5728   0.5668
## ma4  -0.31740    0.19926 -1.5929   0.1112
## sar1 -0.17749    0.12887 -1.3772   0.1685



# For overfitting the added coefficients are insignificant at 5% level of
significant.
# We will keep SARIMA(2,1,3)x(1,1,0) model
```



Checking for Variance using McLeod- Li test

McLeod-Li tests for the residuals of model SARIMA(2,1,3)x(1,1,0) shows the results are all significant at the 5% significance level.

```
#McLeod test for variance
par(mfrow=c(1,1))
# Residual
McLeod.Li.test(y=res.m6,main="McLeod-Li Test Statistics for monthly calls")
```

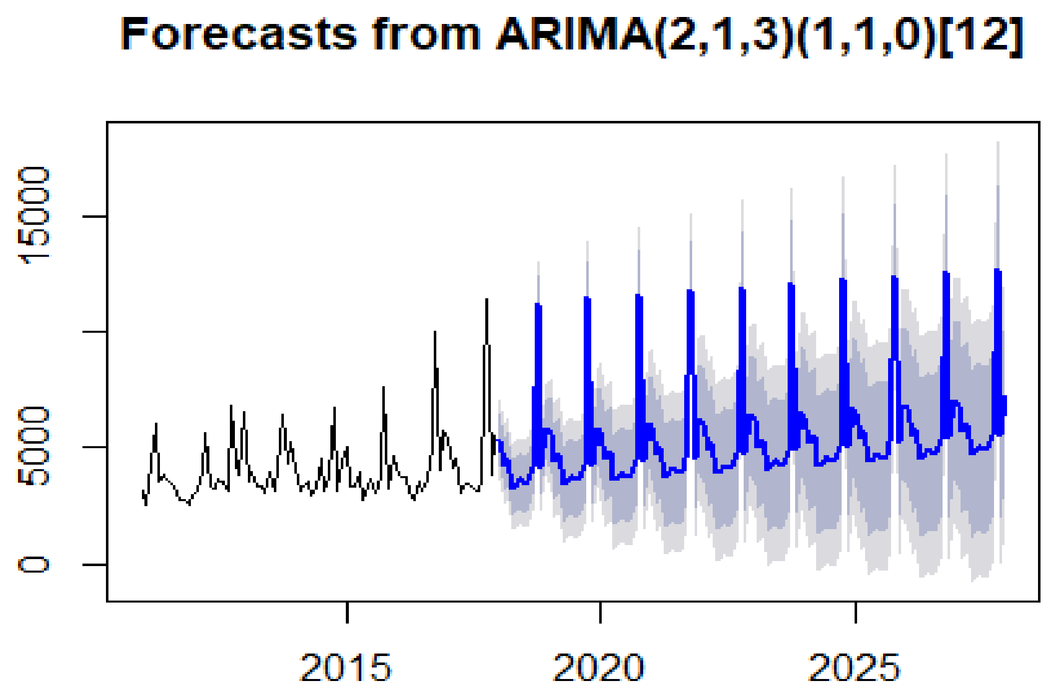



```
#SARIMA
#McLeod.Li.test(arima(log(inbound_calls),order=c(2,1,3),seasonal=list(order=c
(1,1,0), period=12),method = "ML"))
```

Forecasting

We used model SARIMA(2,1,3)x(1,1,0) to produce predictions for next 10 units for the series.As our original series is small and we are forecasting Inbound calls for next 10 years , our prediction became wider as we look forward from the prediction point.

```
m6_213 = Arima(inbound_calls,order=c(2,1,3),seasonal=list(order=c(1,1,0),
period=12))
future = forecast(m6_213, h = 120)
par(mfrow=c(1,1))
plot(future)
```



Summary

After analysing the time series for the total inbound calls we found that there exists trend and seasonality. We dealt with trend, seasonality, non-stationarity and auto-correlation to ensure the series is stationary before fitting model.We also performed McLeod- Li test and found there is no changing variance. Parameter estimation and diagnostic checking helped us to find the best models from the set of 9 possible models. The two best models were SARIMA(2,1,3)x(1,1,0) and SARIMA(4,1,5)x(1,1,0), however we went with lower order model SARIMA(2,1,3)x(1,1,0) for forecating the inbound calls time series for the next 10 years.

References

1. U.S. Department of Health & Human Services,(2018), Call Center Metrics for the Health Service System: Data and Resources, Available at: https://healthdata.gov/dataset/call-center-metrics-health-service-system