ch7_SARIMA

2023-11-11

Summary

General multiplicative Box-Jenkins models:

$$\phi(B)\,\Phi(B^s)\,\nabla^d\,\nabla_S^D\,y_t = \theta(B)\,\Theta(B^s)\,\varepsilon_t$$

$$\boldsymbol{SARIMA(p,d,q)(P,D,Q)S}$$

for Seasonal

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_P B^{Ps})(1-B)^d(1-B^s)^D$$

$$=\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)\left(1-\Theta_1 B^s-\Theta_2 B^{2s}-\cdots-\Theta_Q B^{Qs}\right)\varepsilon_t$$

Autoregressive operator of order p:

$$\phi(B)=\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)$$

seasonal Autoregressive models of order P:

$\Phi(B^s)=(1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_P B^{Ps})$; s: seasonal period length

φ: represent the non-*seasonal* autoregressive parameter.

Φ or Φ :represent the *seasonal* autoregressive parameter.

Difference operator of order d:

$\nabla^d=(1-B)^d$

d is the minimum order of the differences that must be taken to render the series stationary.

seasonal differences at the seasonal period S

$\nabla_S^D=(1-B^s)^D$

Moving Average operator of order q:

$\theta(B)=\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)$

seasonal Moving Average operator of order Q:

$\Theta(B^s)=\left(1-\Theta_1 B^s-\Theta_2 B^{2s}-\cdots-\Theta_Q B^{Qs}\right)$

Θ : represent the seasonal moving average parameter.

θ : represent the moving average parameter.

Q: Write the mathematical formula for the following models :

1. SARIMA $(2,1,2)(1,2,2)_6$

SARIMA(2,1,2)(1,2,2)6

for Seasonal

S:Seasonal - AR: Autoregressive - I: no. of differences - MA: Moving Average - 6: seasonal periods

$(1-\phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^6)(1-B)(1-B^6)^2\, Y_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^6 - \Theta_2 B^{12})$ et

2. SARIMA $(2,0,1)(0,1,2)_3$

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^3)\, Yt = (1 - \theta_1 B)(1 - \Theta_1 B^3 - \Theta_2 B^6)\, et$$

3. SARIMA $(0,2,2)(1,0,1)_5$

$(1 - \Phi_1 B^5)(1 - B)^2 Yt = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^5)\, et$

4. SARIMA $(1,1,0)(2,3,1)_2$

$(1 - \phi_1 B)(1 - \Phi_1 B^2 - \Phi_2 B^4)(1 - B)(1 - B^2)^3\, Yt = (1 - \Theta_1 B^2)\, et$

Exercise 1: For the given data (C2 data), do the following:

1- Plot the time series and comment.
2- Is any transformation/ difference needed for this data to obtain adequate results?
3- Plot the ACF and PACF and comment, suggest a preliminary model for the data.
4- Obtain a suitable SARIMA or ARIMA model for the time series. Test the significancy of the model parameters.
5- Check the adequacy of the model? (hint residual plots and tests).
6- Write down the model using backshift operator.
7- Forecast the next **10** observations of the original series (write the prediction) . Plot the forecast.
8- 80% C.I for the next **10** forecast of the original series.

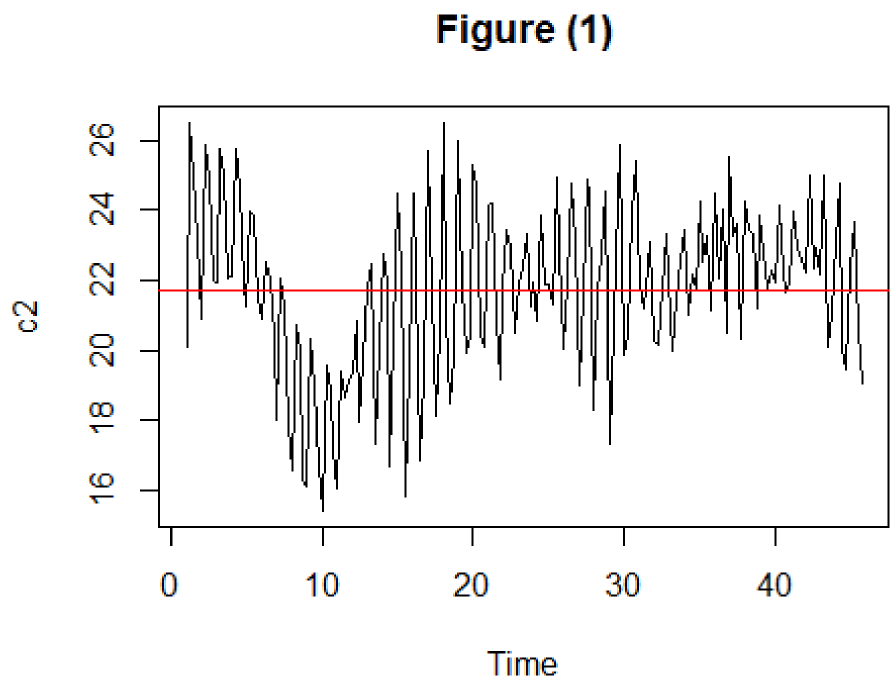
Exercise 1 using R:

The packages used in time series analysis

```
#install.packages("forecast")
#install.packages("tseries")
#install.packages("randtests")
#install.packages("astsa")
#install.packages("lmtest")
library(forecast)
library(tseries)
library(randtests)
library(astsa)
library(lmtest)
```

Checking stationary of the series:

```
## Checking Stability in the mean and variance .
data<- read.csv(file.choose(),header = T)
d=ts(data,frequency = 4)  #time-series objects with seasonal periods 4 .
#figure (1)
plot(d,main="Figure (1)") ; abline(h =mean(d),col="red")
```



The data is not stationary in the mean and in the variance.

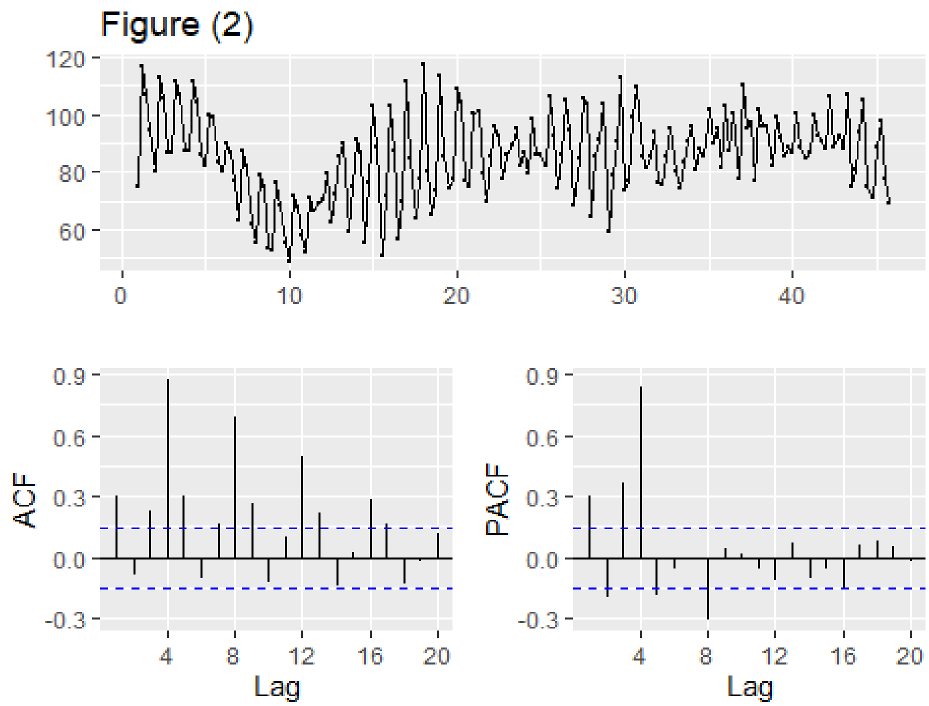
- **Normality test**

```
shapiro.test(d)
    Shapiro-Wilk normality test
data:  d
W = 0.98235, p-value = 0.02239
```

The data don't follow Normal distribution based on Shapiro test, which indicate to nonstationary in the variance. Also, we note that from Figure (1).

- **Apply Box Cox transformation to stabilize the variance.**

```
(lambda <-BoxCox.lambda( d ))

[1] 1.598283

d.B<-BoxCox( d ,lambda)
shapiro.test(d.B)


    Shapiro-Wilk normality test
data:  d.B
W = 0.98797, p-value = 0.1293

ggtsdisplay(d.B,lag.max=20, main="Figure (2)" )
```

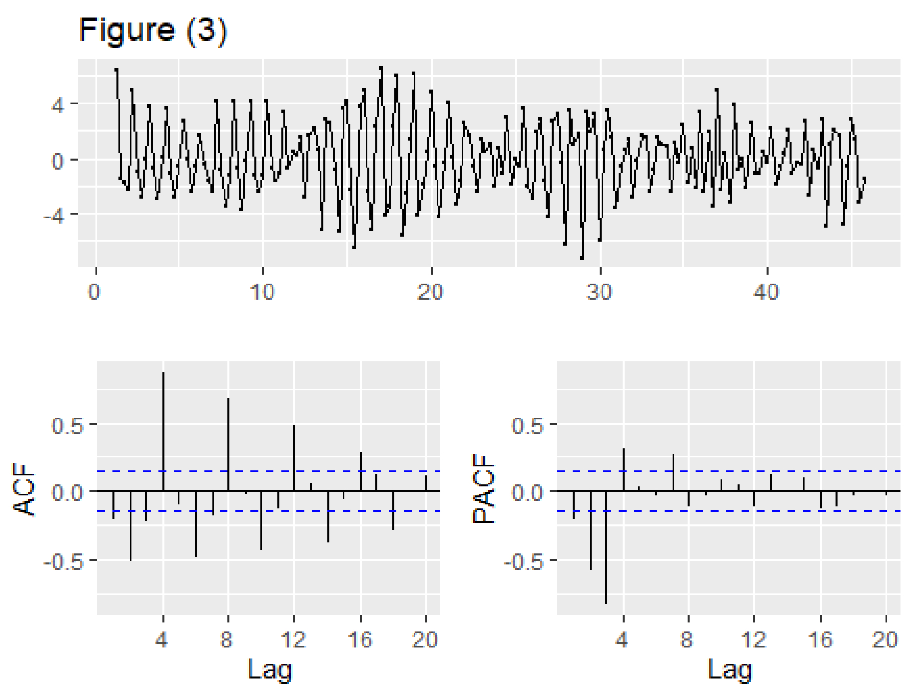



```
#plot(d.B) ; abline(h =mean(d.B),col="red")
```

After taking Box Cox transformation, the data become stationary in the variance but not stationary in the mean.

- ***Taking First difference to stabilize the mean and variance***

```
d.diff<-diff(d,differences =1 )
ggtsdisplay(d.diff,lag.max=20 , main="Figure (3)")
```



```
shapiro.test(d.diff)
    Shapiro-Wilk normality test
data:  d.diff
W = 0.99382, p-value = 0.6583

#plot(d.diff) ; abline(h =mean(d.diff),col="red")
```

After taking first difference, the data become stationary in the mean and in the variance.

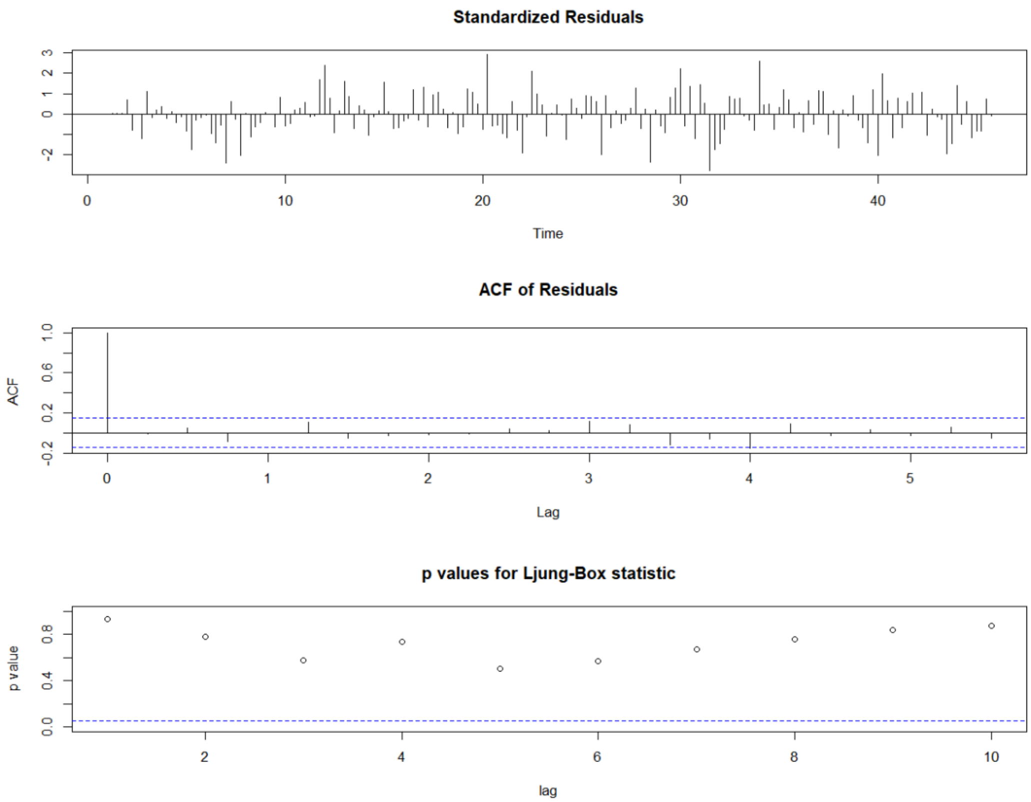
- **Finding the appropriate model using ACF and PACF plot:**

  From Figure (3) we suggest the models **SARIMA(1,0,0)(1,1,0)4 SARIMA(1,1,1)(2,0,1)4** , SARIMA(2,0,1)(1,1,1)4 …

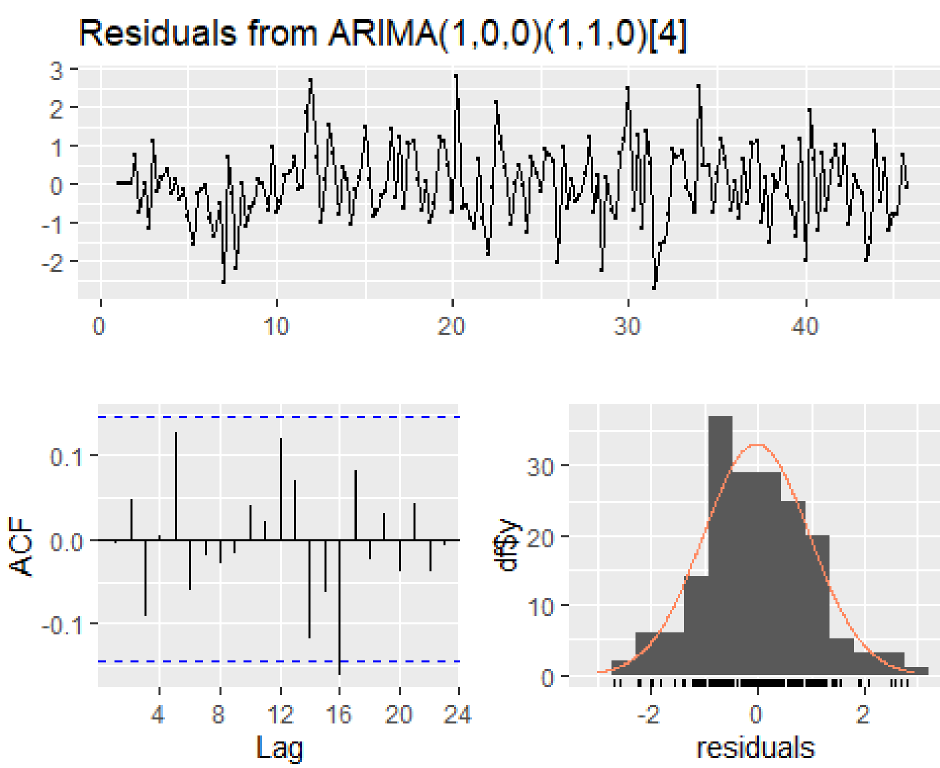
  Or if we consider the data stationary in the mean from Figure (2) , we suggest the models SARIMA(1,0,0)(2,0,1)4 , SARIMA(0,0,1)(2,0,1)4 , …

Model 1: SARIMA(1,0,0)(1,1,0)4 model

```
model1<-Arima(d,order = c(1,0,0),seasonal = c(1,1,0))
#Testing the coefficients & Diagnosing Residuals
coeftest(model1)


z test of coefficients:
     Estimate Std. Error z value  Pr(>|z|)
ar1  0.248590   0.072804  3.4145 0.0006389 ***
sar1 0.290884   0.072085  4.0353 5.453e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tsdiag(model1)
```




```r
checkresiduals(model1, lag= 12)
```



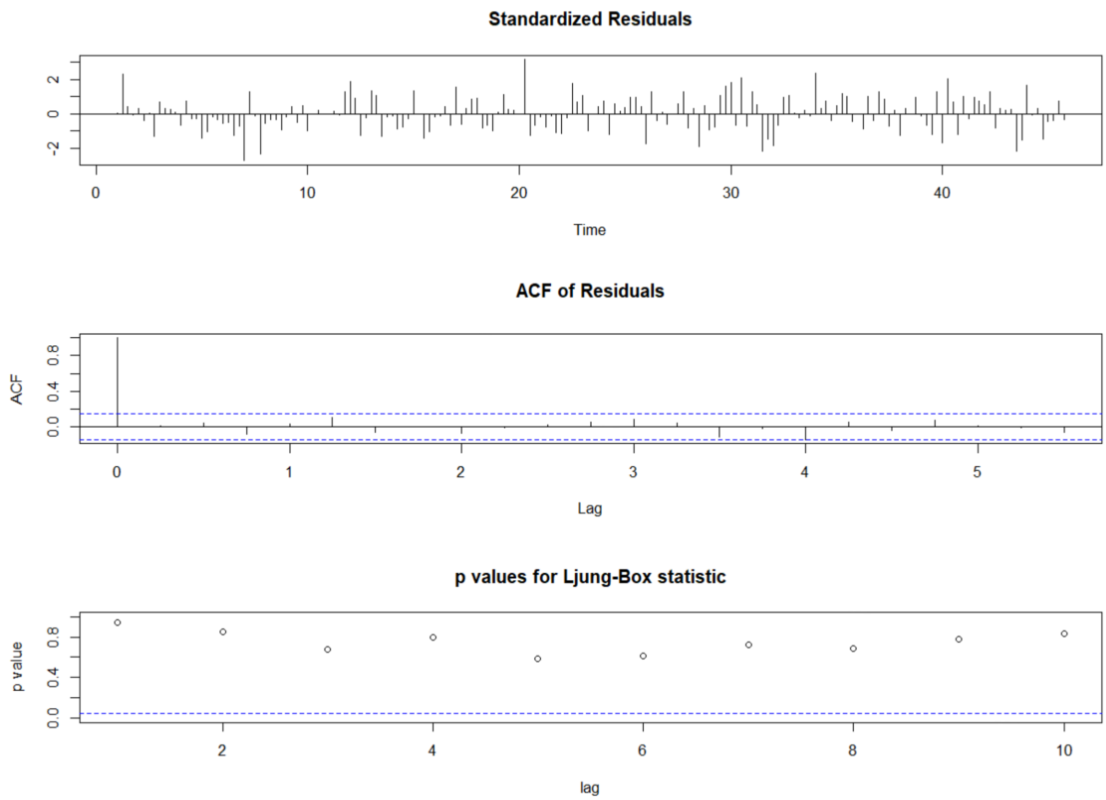
```
    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(1,1,0)[4]
Q* = 9.2283, df = 10, p-value = 0.5106

Model df: 2.   Total lags used: 12

checkresiduals(model1, lag= 24,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(1,1,0)[4]
Q* = 21.669, df = 22, p-value = 0.4798

Model df: 2.   Total lags used: 24

checkresiduals(model1, lag= 36,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(1,1,0)[4]
Q* = 48.15, df = 34, p-value = 0.05464

Model df: 2.   Total lags used: 36

checkresiduals(model1, lag= 48,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(1,1,0)[4]
Q* = 60.148, df = 46, p-value = 0.07865

Model df: 2.   Total lags used: 48

runs.test(model1$r)


    Runs Test
data:  model1$r
statistic = 1.9434, runs = 104, n1 = 90, n2 = 90, n = 180, p-value =
0.05197
alternative hypothesis: nonrandomness

shapiro.test(model1$residuals)


    Shapiro-Wilk normality test
data:  model1$residuals
W = 0.99169, p-value = 0.3883
```

```
t.test(model1$r)


    One Sample t-test
data:  model1$r
t = -0.39531, df = 179, p-value = 0.6931
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.1750647  0.1166304
sample estimates:
  mean of x
-0.02921718
```

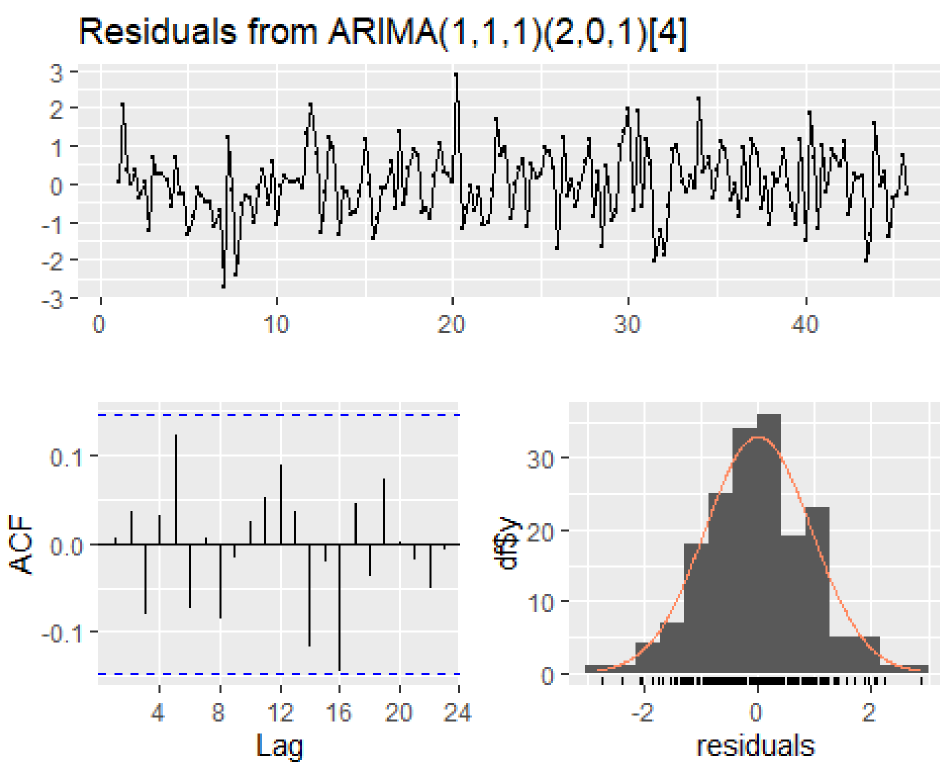
- all coefficients of model are significant .
- Plot of residuals with time: The residuals are random around the zero.
- All p-values of the Ljung-Box test > 0.05. The residuals are uncorrelated.
- The ACF of the Residuals are zeros.
- The residuals follow normal dist.
- The residuals are random.

Model 2: SARIMA(1,1,1)(2,0,1)4 model

```
model2<-Arima(d,order = c(1,1,1),seasonal = c(2,0,1))
#Testing the coefficients & Diagnosing Residuals
coeftest(model2)


z test of coefficients:


      Estimate Std. Error  z value  Pr(>|z|)
ar1   0.252169   0.072700   3.4686 0.0005231 ***
ma1  -1.000000   0.026591 -37.6066 < 2.2e-16 ***
sar1  1.680004   0.121438  13.8343 < 2.2e-16 ***
sar2 -0.770436   0.104849  -7.3480 2.011e-13 ***
sma1 -0.575903   0.179969  -3.2000 0.0013742 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tsdiag(model2)
```




```r
checkresiduals(model2, lag= 12)
```

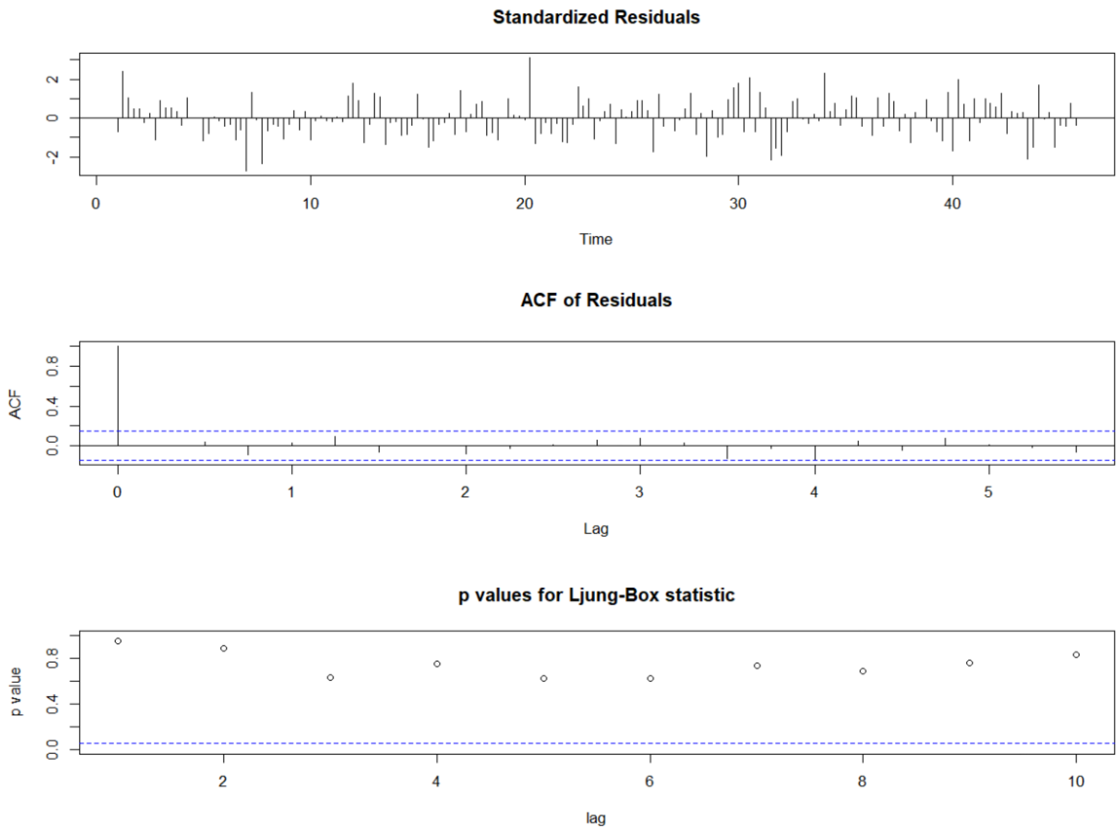


```
	Ljung-Box test
data:  Residuals from ARIMA(1,1,1)(2,0,1)[4]
Q* = 9.1607, df = 7, p-value = 0.2413

Model df: 5.   Total lags used: 12

checkresiduals(model2, lag= 24,plot=FALSE)


	Ljung-Box test
data:  Residuals from ARIMA(1,1,1)(2,0,1)[4]
Q* = 20.869, df = 19, p-value = 0.3441

Model df: 5.   Total lags used: 24

checkresiduals(model2, lag= 36,plot=FALSE)


	Ljung-Box test
data:  Residuals from ARIMA(1,1,1)(2,0,1)[4]
Q* = 39.107, df = 31, p-value = 0.1505

Model df: 5.   Total lags used: 36

checkresiduals(model2, lag= 48,plot=FALSE)


	Ljung-Box test
data:  Residuals from ARIMA(1,1,1)(2,0,1)[4]
Q* = 51.87, df = 43, p-value = 0.1664

Model df: 5.   Total lags used: 48

runs.test(model2$r)


	Runs Test
data:  model2$r
statistic = -0.44847, runs = 88, n1 = 90, n2 = 90, n = 180, p-value =
0.6538
alternative hypothesis: nonrandomness

shapiro.test(model2$residuals)


	Shapiro-Wilk normality test
data:  model2$residuals
W = 0.99728, p-value = 0.9886
```

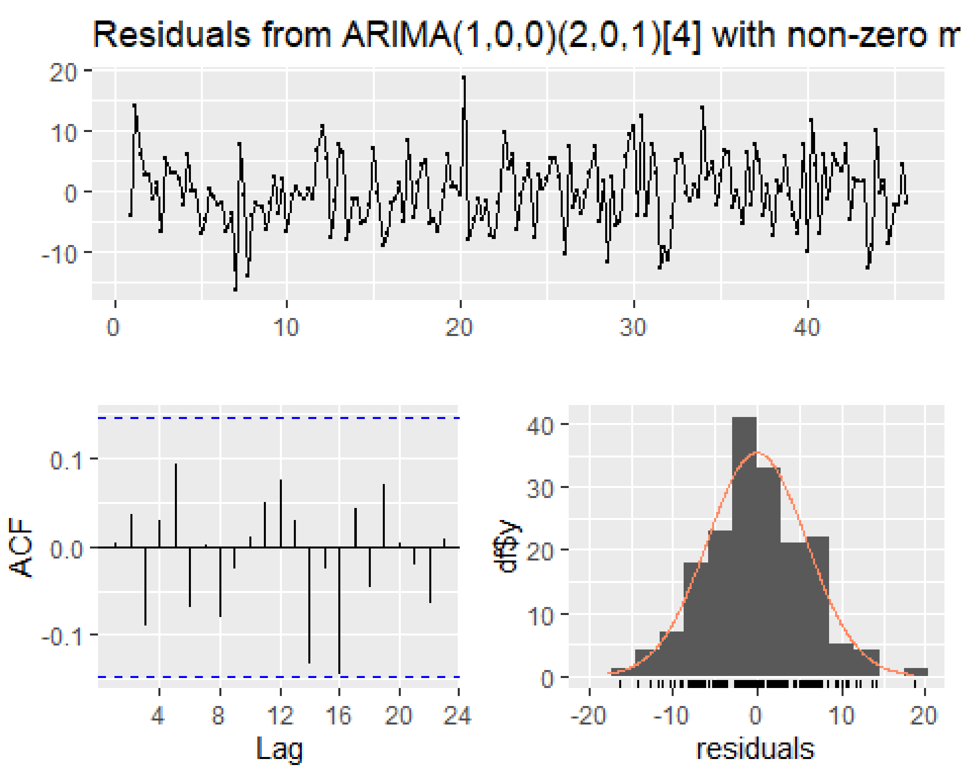
```
t.test(model2$r)
    One Sample t-test

data:  model2$r
t = 0.073613, df = 179, p-value = 0.9414
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.1323098  0.1425638
sample estimates:
  mean of x
0.005126972
```

Model 3: SARIMA(1,0,0)(2,0,1)4 model

```
model3<-Arima(d,order = c(1,0,0),seasonal = c(2,0,1),lambda = lambda, biasadj
=TRUE)

#biasadj =TRUE: return to raw data, after using BoxCox transformation
#Testing the coefficients & Diagnosing Residuals
coeftest(model3)


z test of coefficients:
           Estimate Std. Error z value  Pr(>|z|)
ar1        0.230383   0.072670  3.1702 0.0015231 **
sar1       1.724466   0.106796 16.1473 < 2.2e-16 ***
sar2      -0.809126   0.089803 -9.0100 < 2.2e-16 ***
sma1      -0.651400   0.168277 -3.8710 0.0001084 ***
intercept 85.332720   2.371989 35.9752 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tsdiag(model3)
```




```r
checkresiduals(model3, lag= 12)
```

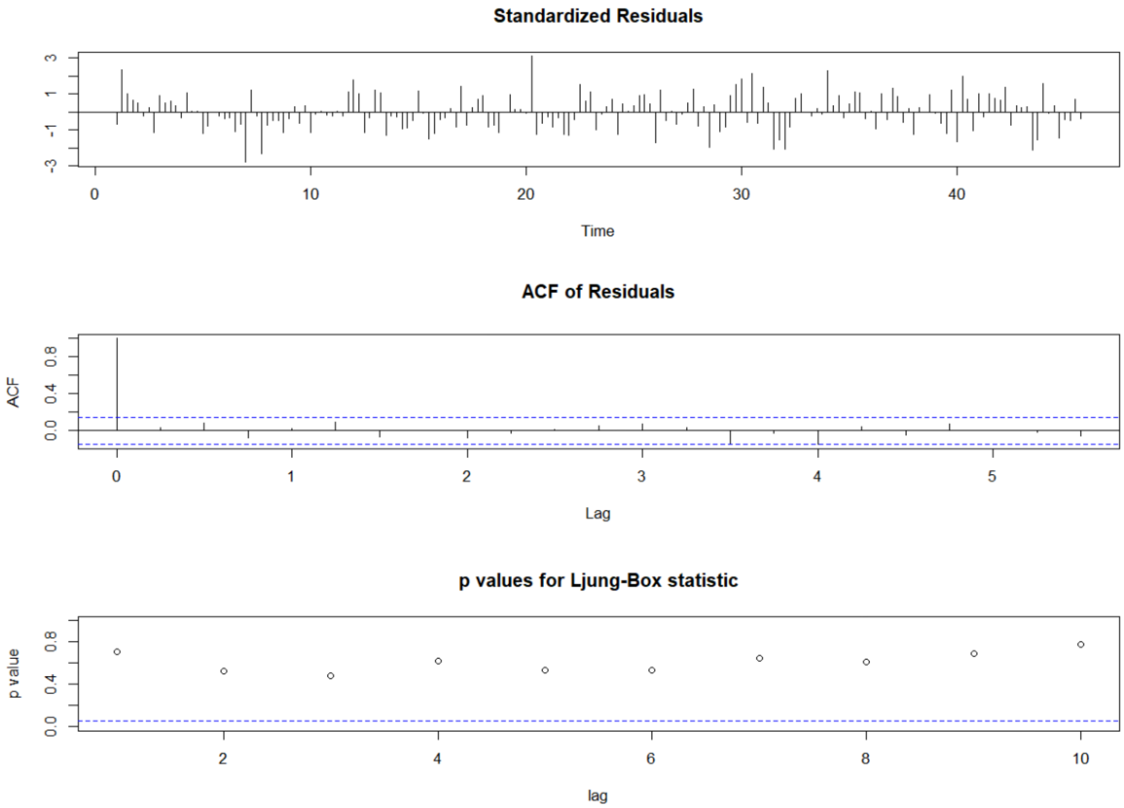


```
    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(2,0,1)[4] with non-zero mean
Q* = 7.3702, df = 8, p-value = 0.4973

Model df: 4.   Total lags used: 12

checkresiduals(model3, lag= 24,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(2,0,1)[4] with non-zero mean
Q* = 20.251, df = 20, p-value = 0.4423

Model df: 4.   Total lags used: 24

checkresiduals(model3, lag= 36,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(2,0,1)[4] with non-zero mean
Q* = 37.515, df = 32, p-value = 0.231

Model df: 4.   Total lags used: 36

checkresiduals(model3, lag= 48,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(1,0,0)(2,0,1)[4] with non-zero mean
Q* = 48.92, df = 44, p-value = 0.2821

Model df: 4.   Total lags used: 48

runs.test(model3$r)


    Runs Test
data:  model3$r
statistic = 0, runs = 91, n1 = 90, n2 = 90, n = 180, p-value = 1
alternative hypothesis: nonrandomness

shapiro.test(model3$residuals)


    Shapiro-Wilk normality test
data:  model3$residuals
W = 0.99657, p-value = 0.9593
```

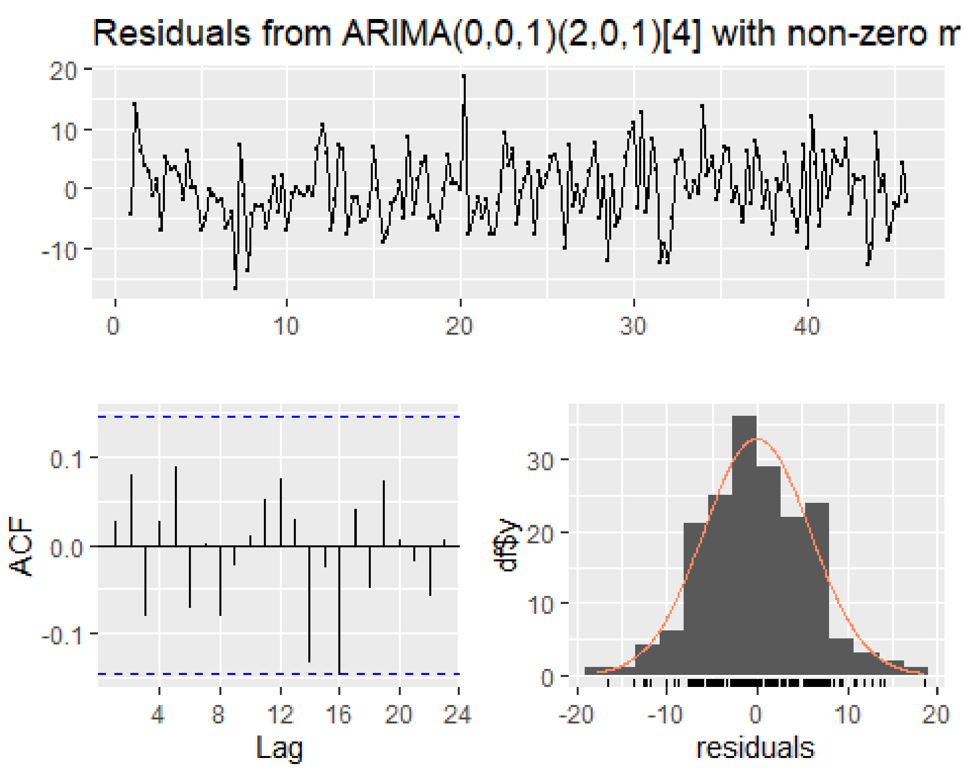
```
t.test(model3$r)
    One Sample t-test

data:  model3$r
t = -0.1511, df = 179, p-value = 0.8801
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.9292902  0.7970975
sample estimates:
 mean of x 
-0.0660963
```

Model 4: SARIMAARIMA(0,0,1)(2,0,1)[4]

```
model4<-Arima(d,order = c(0,0,1),seasonal = c(2,0,1),lambda = lambda, biasadj
=TRUE)

#Testing the coefficients & Diagnosing Residuals
coeftest(model4)
z test of coefficients:
           Estimate Std. Error z value  Pr(>|z|)    
ma1        0.206488   0.067637  3.0529 0.0022665 ** 
sar1       1.715405   0.109710 15.6358 < 2.2e-16 ***
sar2      -0.803230   0.091820 -8.7479 < 2.2e-16 ***
sma1      -0.639457   0.170578 -3.7488 0.0001777 ***
intercept 85.340044   2.201998 38.7557 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tsdiag(model4)
```




```
checkresiduals(model4, lag= 12)
```



```
    Ljung-Box test
data:  Residuals from ARIMA(0,0,1)(2,0,1)[4] with non-zero mean
Q* = 8.1499, df = 8, p-value = 0.419

Model df: 4.   Total lags used: 12

checkresiduals(model4, lag= 24,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(0,0,1)(2,0,1)[4] with non-zero mean
Q* = 21.055, df = 20, p-value = 0.3939

Model df: 4.   Total lags used: 24

checkresiduals(model4, lag= 36,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(0,0,1)(2,0,1)[4] with non-zero mean
Q* = 38.555, df = 32, p-value = 0.1973

Model df: 4.   Total lags used: 36

checkresiduals(model4, lag= 48,plot=FALSE)


    Ljung-Box test
data:  Residuals from ARIMA(0,0,1)(2,0,1)[4] with non-zero mean
Q* = 49.806, df = 44, p-value = 0.2534

Model df: 4.   Total lags used: 48

runs.test(model4$r)


    Runs Test
data:  model4$r
statistic = 0.29898, runs = 93, n1 = 90, n2 = 90, n = 180, p-value =
0.765
alternative hypothesis: nonrandomness

shapiro.test(model4$residuals)


    Shapiro-Wilk normality test
data:  model4$residuals
W = 0.99615, p-value = 0.9304
```

```
t.test(model4$r)
    One Sample t-test
data:  model4$r
t = -0.16351, df = 179, p-value = 0.8703
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.9375985  0.7941074
sample estimates:
  mean of x
-0.07174551
```

Using AIC or BIC to choose best model

```
model1$aic

[1] 506.0465

model2$aic

[1] 505.3966

model3$aic

[1] 1167.146

model4$aic

[1] 1168.165
```

The best model has the lowest AIC value , which is model 2.

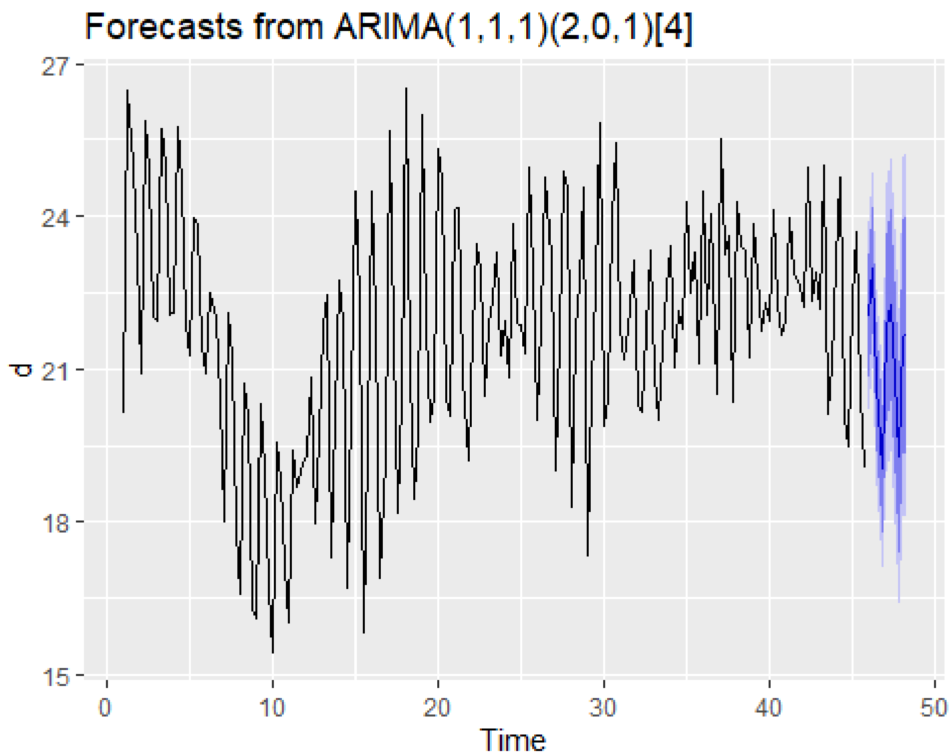
$$(1-0.2522B)(1-1.68B^4-(-0.7704)B^8)(1-B)^1\, y_t = (1-(-1)B)(1-(-0.5759)B^4)\, \varepsilon_t$$

$$(1-0.2522B)(1-1.68B^4+0.7704B^8)(1-B)\, y_t = (1+B)(1+0.5759B^4)\, \varepsilon_t$$

Forecasting using SARIMA(1,1,1)(2,0,1)4:

```
best.model<-Arima(d,order = c(1,1,1),seasonal = c(2,0,1))
(f=forecast(best.model, h=10,level = c(80, 95)))

      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
46 Q1       22.04153 20.82358 23.25947 20.17884 23.90421
46 Q2       22.95799 21.70029 24.21569 21.03450 24.88147
46 Q3       20.66510 19.40450 21.92569 18.73719 22.59301
46 Q4       19.03367 17.77279 20.29455 17.10532 20.96202
47 Q1       21.81451 19.96105 23.66797 18.97988 24.64913
47 Q2       22.28156 20.39419 24.16894 19.39507 25.16806
47 Q3       20.86137 18.97122 22.75153 17.97064 23.75211
47 Q4       19.27213 17.38165 21.16262 16.38088 22.16338
48 Q1       21.62871 19.30870 23.94873 18.08056 25.17687
48 Q2       21.70723 19.35954 24.05492 18.11675 25.29772
```

```r
autoplot(f)
```



Forecasts from ARIMA(1,1,1)(2,0,1)[4]
27
24
21
18
15
d
0
10
20
30
40
50
Time