Time Series Analysis - Project

Divya Chandrashekar

April 25, 2019

```
install.packages("astsa")

install.packages("devtools")

devtools::install_github("nickpoison/astsa")

library(astsa)
```

1. Data Description

a) Data Source:U.S. Unemployment. Astsa package. Dataset - Unemp

```
unemp<-astsa::unemp
```

b) Variable Description

Monthly U.S. Unemployment series (1948-1978, n = 372)

```
summary(unemp)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   148.0   279.8   366.1   393.0   467.6   856.9

#checking class to confirm it is a ts
class(unemp)

## [1] "ts"
```

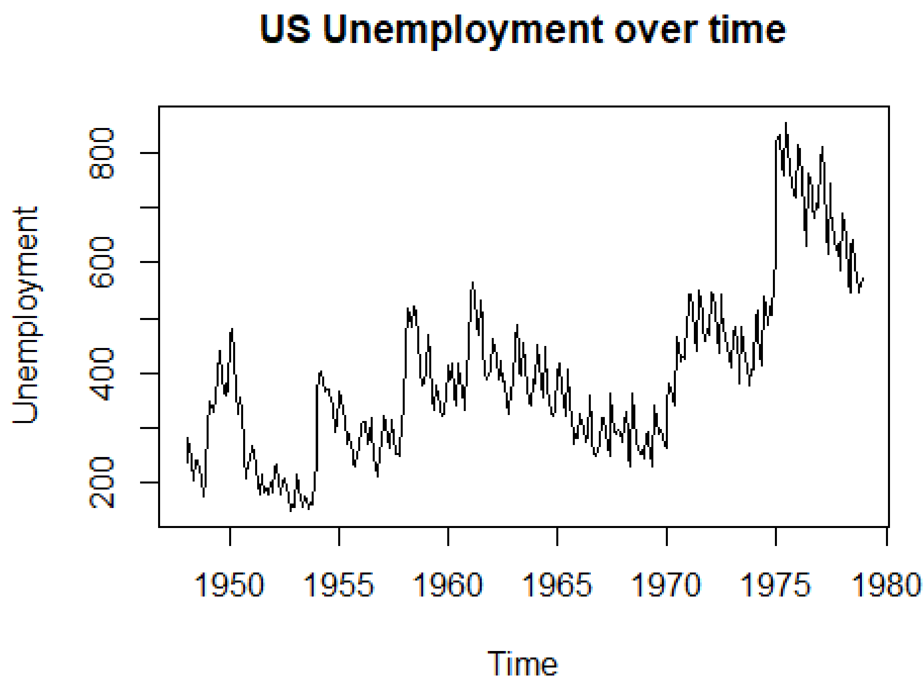
c) ts description

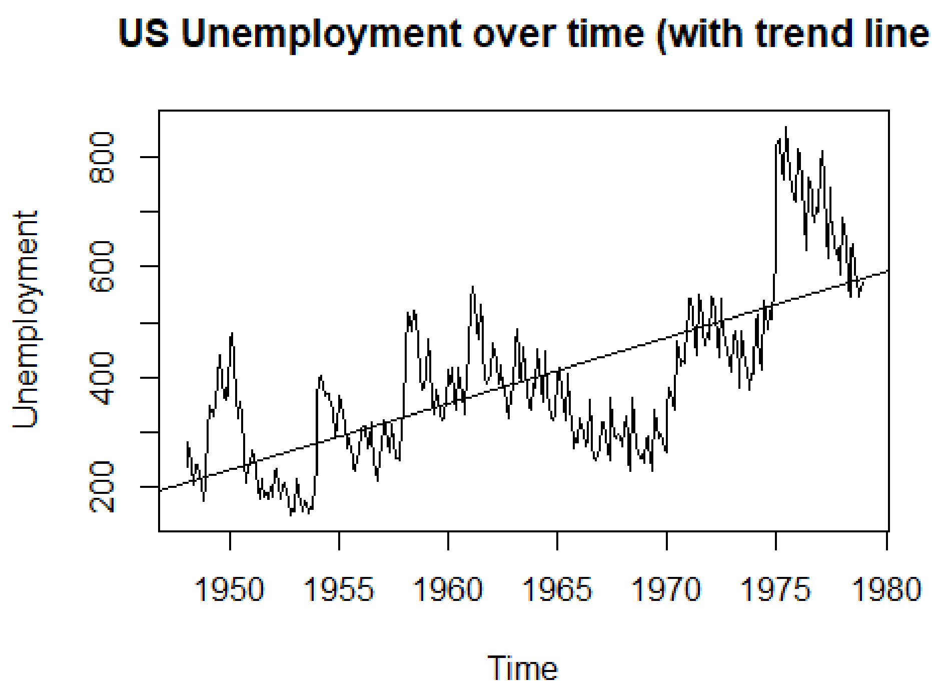
```
start(unemp)

## [1] 1948    1

end(unemp)

## [1] 1978   12

frequency(unemp)

## [1] 12
```

From these results it can be see that the series starts with 1948 January and ends at 1978 December, once cycle is 12 months.

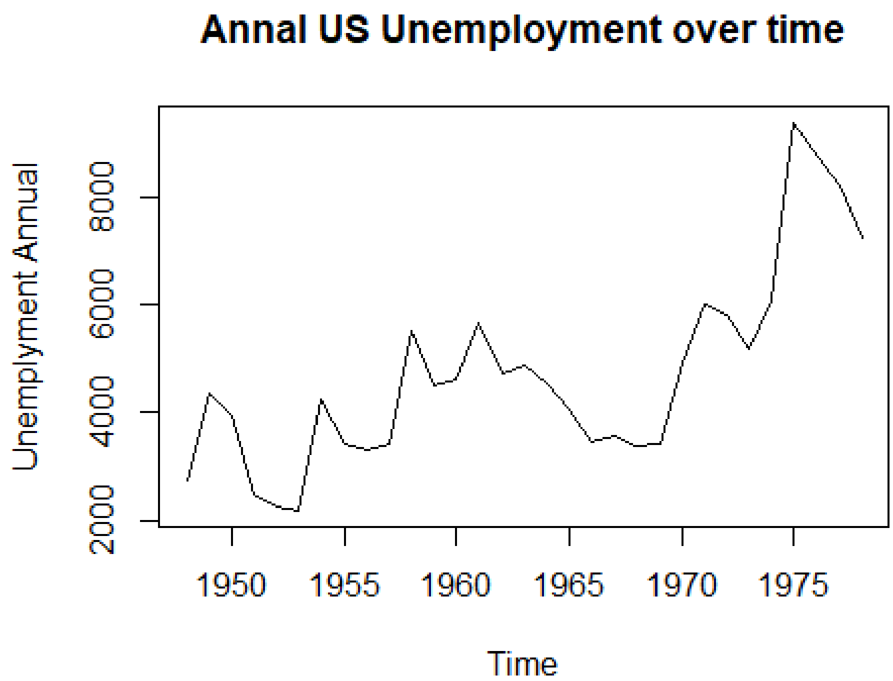
2. Data Exploration

a) Plotting data





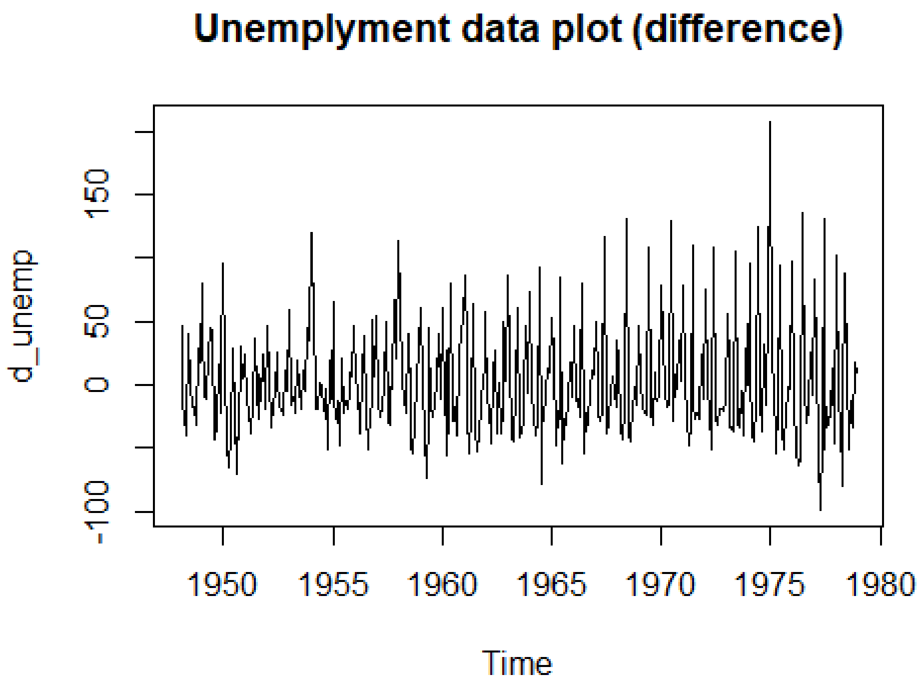
For further EDA, Box plot across months to explore seasonal effects (Also aggregating annualy)



From the preliminary plot of raw data, an increasing trend can be observed. From the boxplot across months, highest unemployment is observed in June.

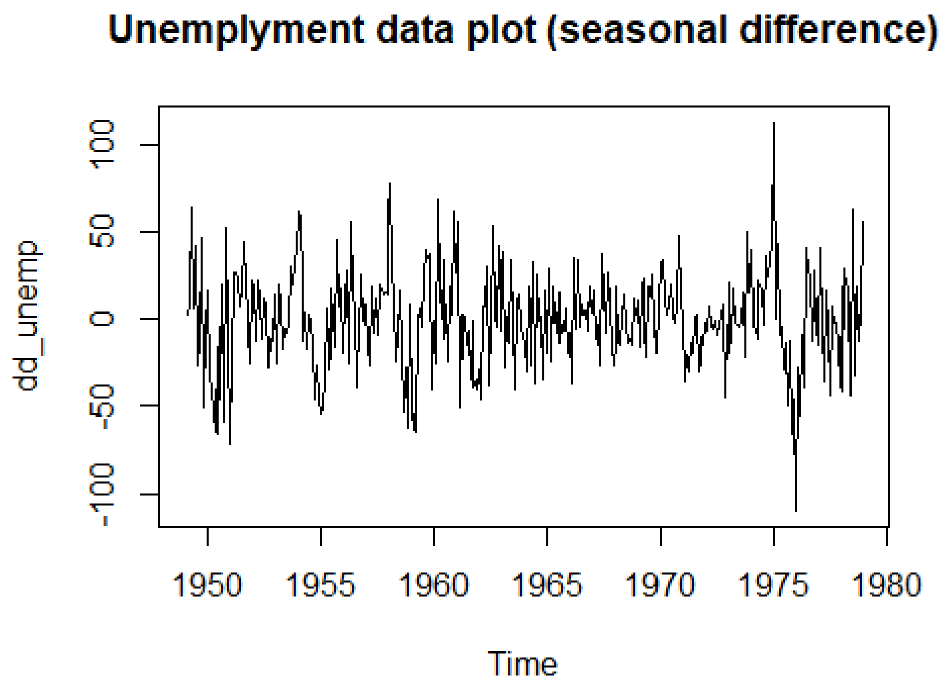
b)Transforming data

Differencing the unemployment data and plotting it



Seasonally differencing d_unemp and plotting it

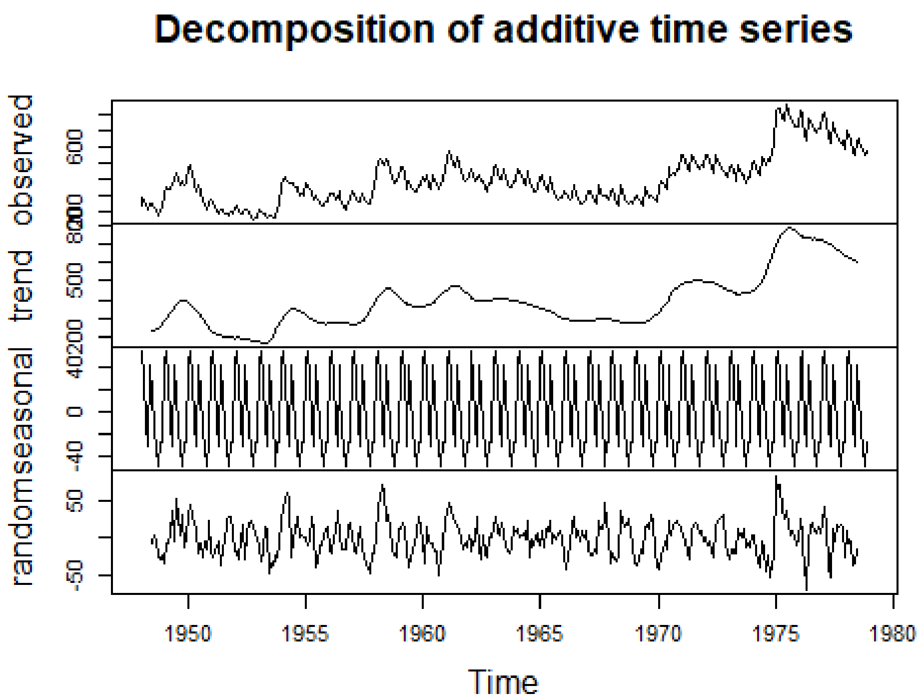
```
dd_unemp <- diff(d_unemp, lag = 12)
plot(dd_unemp,main="Unemplyment data plot (seasonal difference)")
```

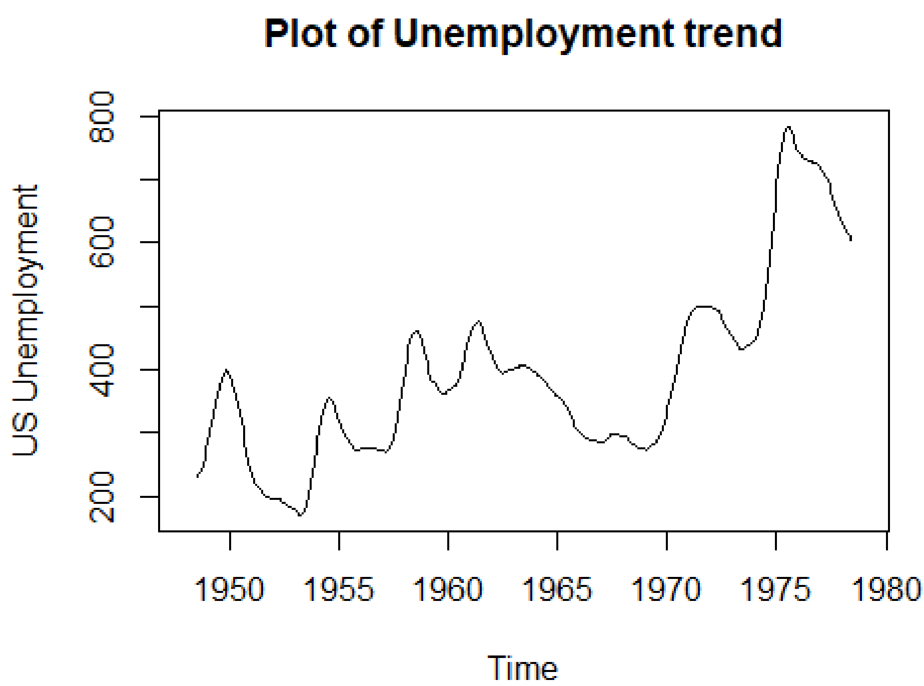


By differencing, the trend and seasonal variation in unemployment have been removed. The series now appears to be stationary.

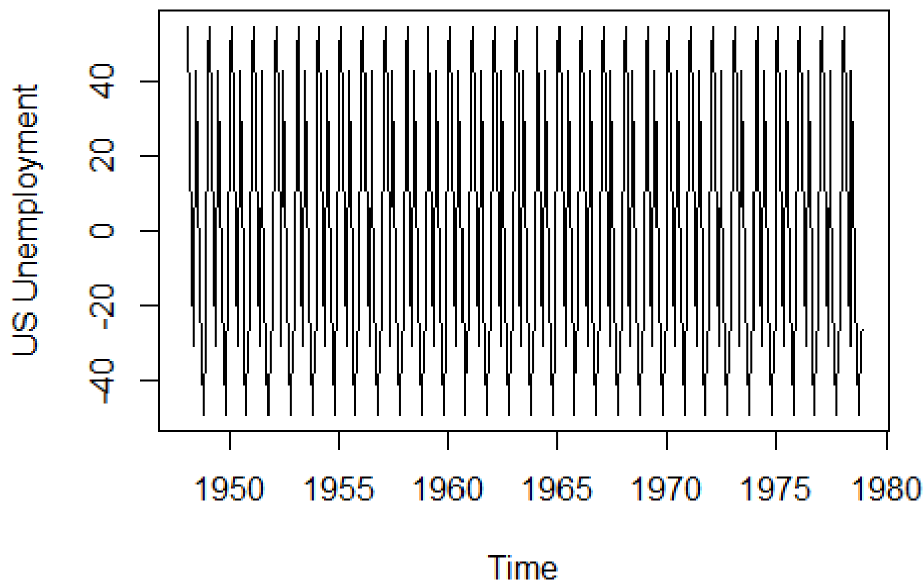
3. Data Decomposition

Decomposing series into trend, seasonal and residual components and plotting these

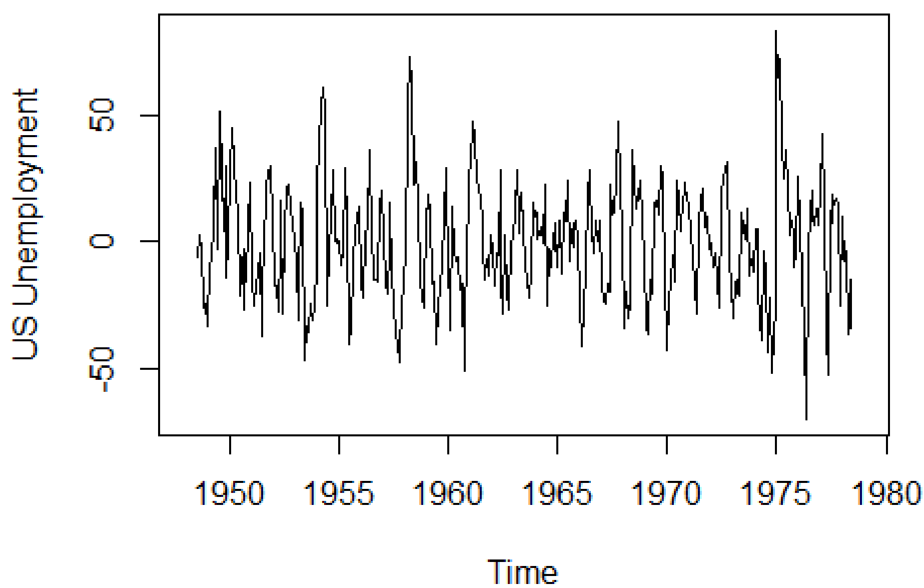

Decomposition of additive time series
observed
trend
seasonal
random
600
200
800
500
40
0
-40
50
-50
1950
1955
1960
1965
1970
1975
1980
Time

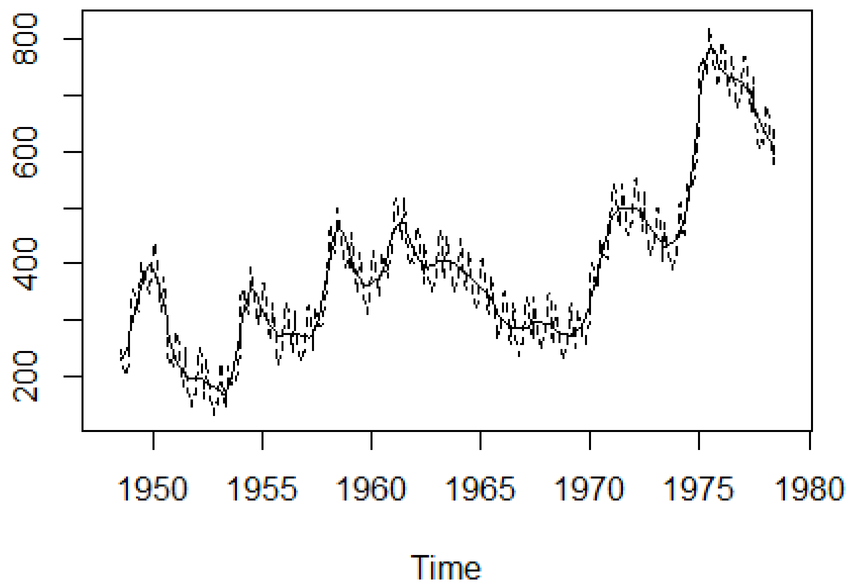

Plot of Unemployment trend
US Unemployment
200
400
600
800
1950
1955
1960
1965
1970
1975
1980
Time

Plot of Unemployment seasonal



Plot of Unemployment residuals





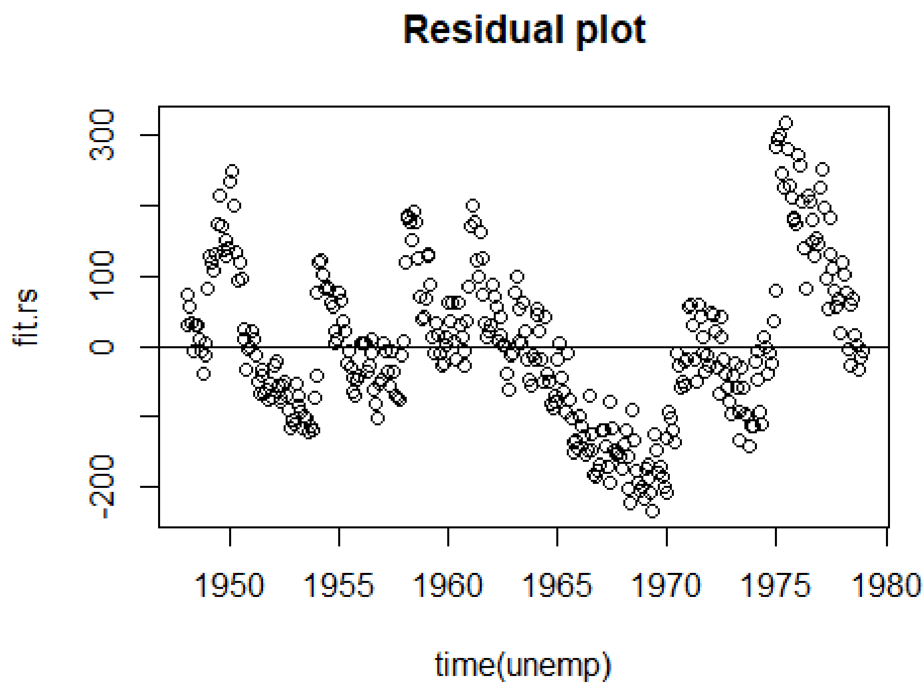
Since the series shows an increasing trend, but seasonal effect does not, an additive decomposition model looks appropriate. The residuals seem to show an increasing trend, so a transformation (log) can be considered.

4.Regression

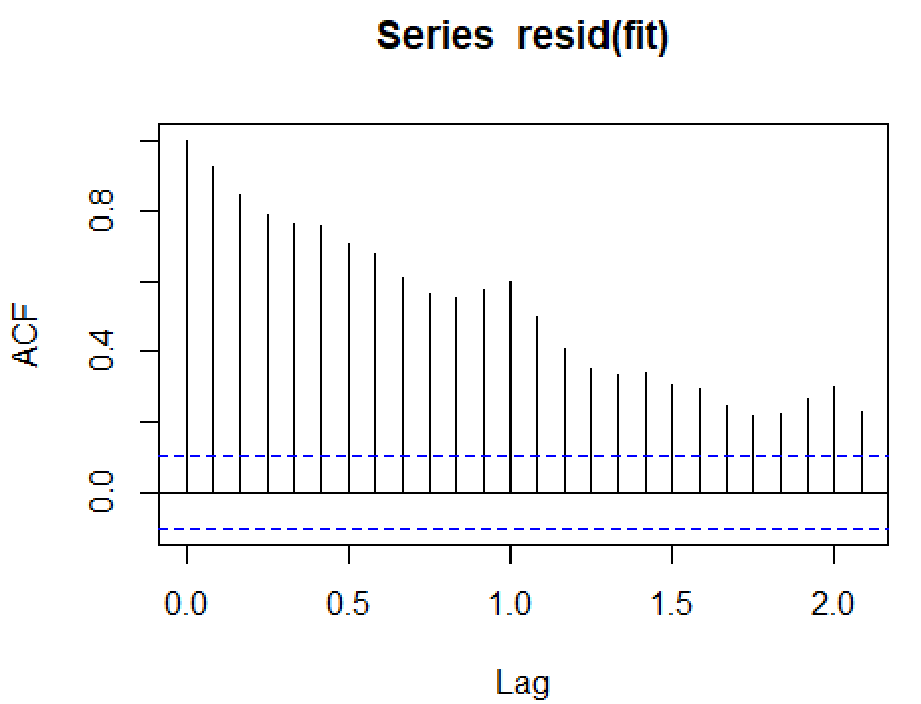
Regressing unemployment data on time

```
fit <- lm(unemp~time(unemp), na.action=NULL)

summary(fit)

##
## Call:
## lm(formula = unemp ~ time(unemp), na.action = NULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -234.32  -72.71  -10.25   63.84  318.77
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.343e+04  1.259e+03  -18.61   <2e-16 ***
## time(unemp)  1.213e+01  6.413e-01   18.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

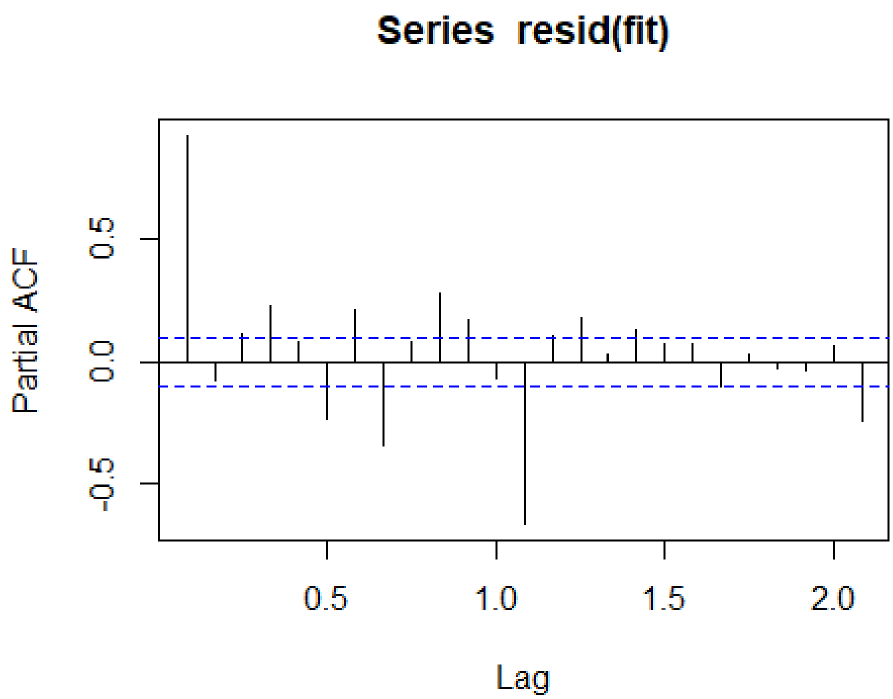
```
## Residual standard error: 110.7 on 370 degrees of freedom
## Multiple R-squared:  0.4917, Adjusted R-squared:  0.4903
## F-statistic: 357.9 on 1 and 370 DF,  p-value: < 2.2e-16

summary(aov(fit))

##               Df  Sum Sq Mean Sq F value Pr(>F)
## time(unemp)    1 4385584 4385584   357.9 <2e-16 ***
## Residuals    370 4533528   12253
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

num = length(unemp)


AIC(fit)/num - log(2*pi) # AIC

## [1] 10.42425

BIC(fit)/num - log(2*pi) # BIC

## [1] 10.45585

(AICc = log(sum(resid(fit)^2)/num) + (num+5)/(num-5-2)) # AICc

## [1] 10.44099

#Preparing Residual Plot

fit.rs=resid(fit)

plot(time(unemp), fit.rs,main="Residual plot")
abline(0, 0)
```




```
#Plotting acf and pacf
acf(resid(fit))
```



```r
pacf(resid(fit))
```



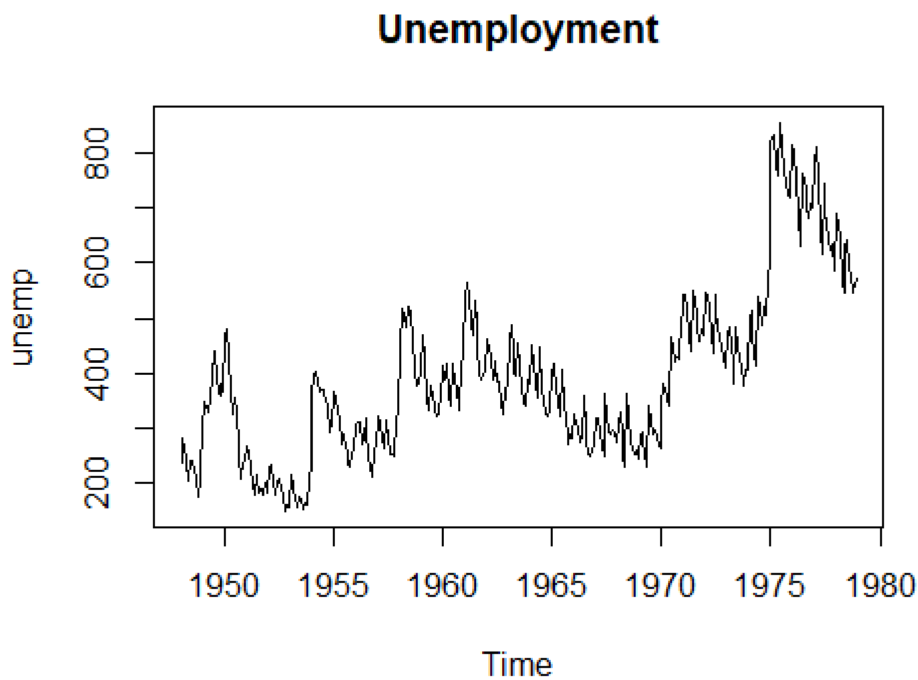
Interpretation of linear regression model:

The adjusted R square of the model is 0.49 the closer it gets to 1 better the model fits the data.Transformations such as log or taking a difference may result in a model with higher R squared value.

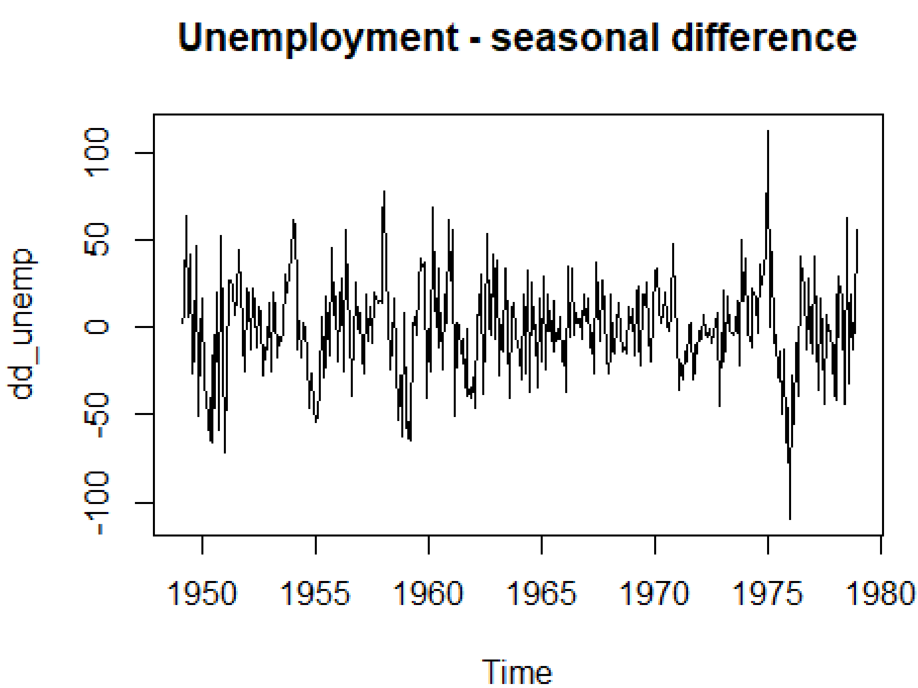
The slow linear decay of the ACF plot is typical of a non-stationary time series.

5.ARIMA model

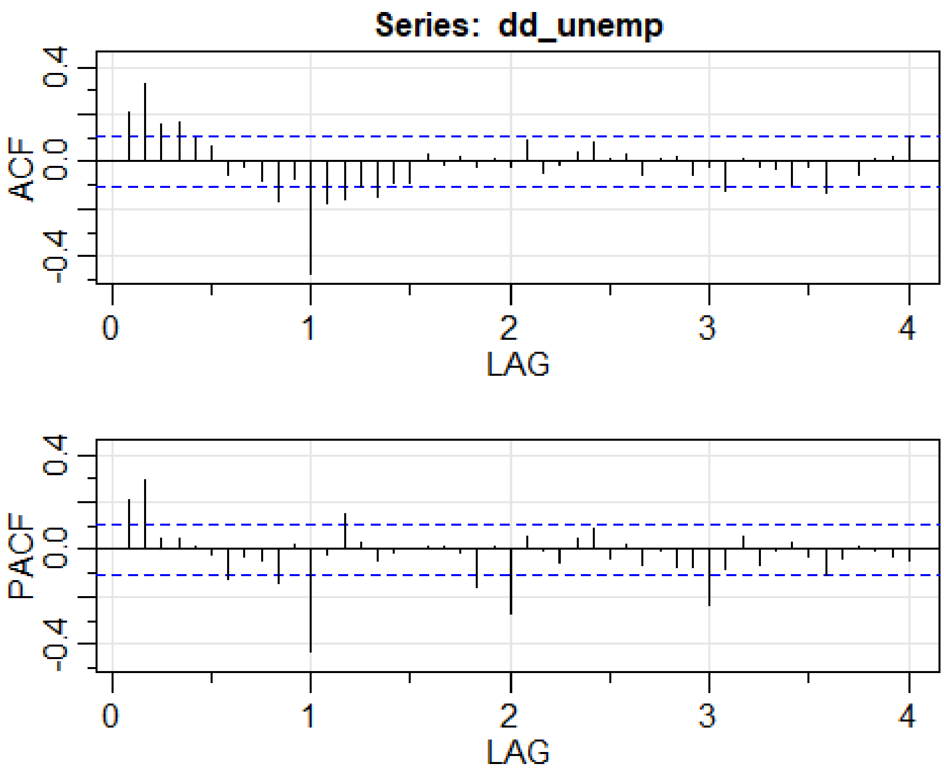
Differencing the data and plotting it



Seasonally differencing d_unemp and plotting it.



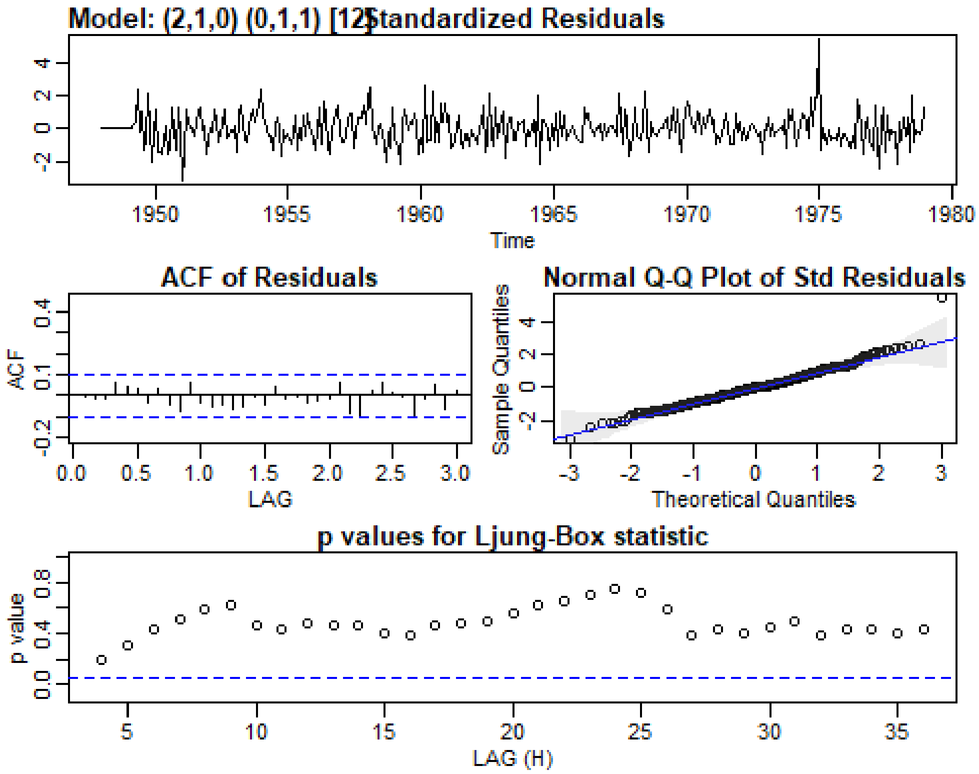
The data appears to be stationary now .Plotting P/ACF pair of fully differenced data to lag 60




```
##          ACF  PACF
##  [1,]  0.21  0.21
##  [2,]  0.33  0.29
##  [3,]  0.15  0.05
##  [4,]  0.17  0.05
##  [5,]  0.10  0.01
##  [6,]  0.06 -0.02
##  [7,] -0.06 -0.12
##  [8,] -0.02 -0.03
##  [9,] -0.09 -0.05
## [10,] -0.17 -0.15
## [11,] -0.08  0.02
## [12,] -0.48 -0.43
## [13,] -0.18 -0.02
## [14,] -0.16  0.15
## [15,] -0.11  0.03
## [16,] -0.15 -0.04
## [17,] -0.09 -0.01
## [18,] -0.09  0.00
## [19,]  0.03  0.01
## [20,] -0.01  0.01
## [21,]  0.02 -0.01
## [22,] -0.02 -0.16
## [23,]  0.01  0.01
```

```
## [24,] -0.02 -0.27
## [25,]  0.09  0.05
## [26,] -0.05 -0.01
## [27,] -0.01 -0.05
## [28,]  0.03  0.05
## [29,]  0.08  0.09
## [30,]  0.01 -0.04
## [31,]  0.03  0.02
## [32,] -0.05 -0.07
## [33,]  0.01 -0.01
## [34,]  0.02 -0.08
## [35,] -0.06 -0.08
## [36,] -0.02 -0.23
## [37,] -0.12 -0.08
## [38,]  0.01  0.06
## [39,] -0.03 -0.07
## [40,] -0.03 -0.01
## [41,] -0.10  0.03
## [42,] -0.02 -0.03
## [43,] -0.13 -0.11
## [44,]  0.00 -0.04
## [45,] -0.06  0.01
## [46,]  0.01  0.00
## [47,]  0.02 -0.03
## [48,]  0.11 -0.04
```

```
## initial  value 3.340809
## iter   2 value 3.105512
## iter   3 value 3.086631
## iter   4 value 3.079778
## iter   5 value 3.069447
## iter   6 value 3.067659
## iter   7 value 3.067426
## iter   8 value 3.067418
## iter   8 value 3.067418
## final  value 3.067418
## converged
## initial  value 3.065481
## iter   2 value 3.065478
## iter   3 value 3.065477
## iter   3 value 3.065477
## iter   3 value 3.065477
## final  value 3.065477
## converged
```



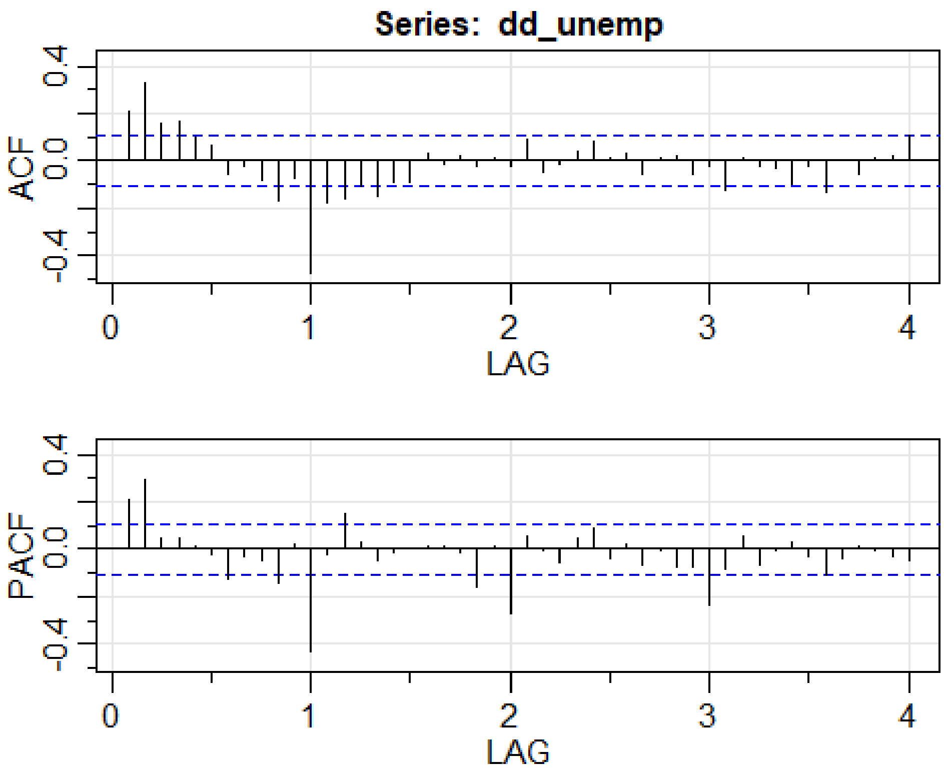

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P,
D,
##     Q), period = S), include.mean = !no.constant, transform.pars = trans,
fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1     ar2     sma1
##       0.1351  0.2464  -0.6953
## s.e.  0.0513  0.0515   0.0381
##
## sigma^2 estimated as 449.6:  log likelihood = -1609.91,  aic = 3227.81
##
## $degrees_of_freedom
## [1] 356
##
## $ttable
##      Estimate     SE  t.value p.value
## ar1    0.1351 0.0513   2.6326  0.0088
## ar2    0.2464 0.0515   4.7795  0.0000
## sma1  -0.6953 0.0381 -18.2362  0.0000
##
## $AIC
## [1] 8.723811
```

```
##
## $AICc
## [1] 8.723988
##
## $BIC
## [1] 8.765793
```

For the nonseasonal component: PACF cuts off at lag 2 and the ACF tails off and in the seasonal component: the ACF cuts off at lag 12 and the PACF tails off at lags 12, 24, 36 etc.

6.Model Diagnostics

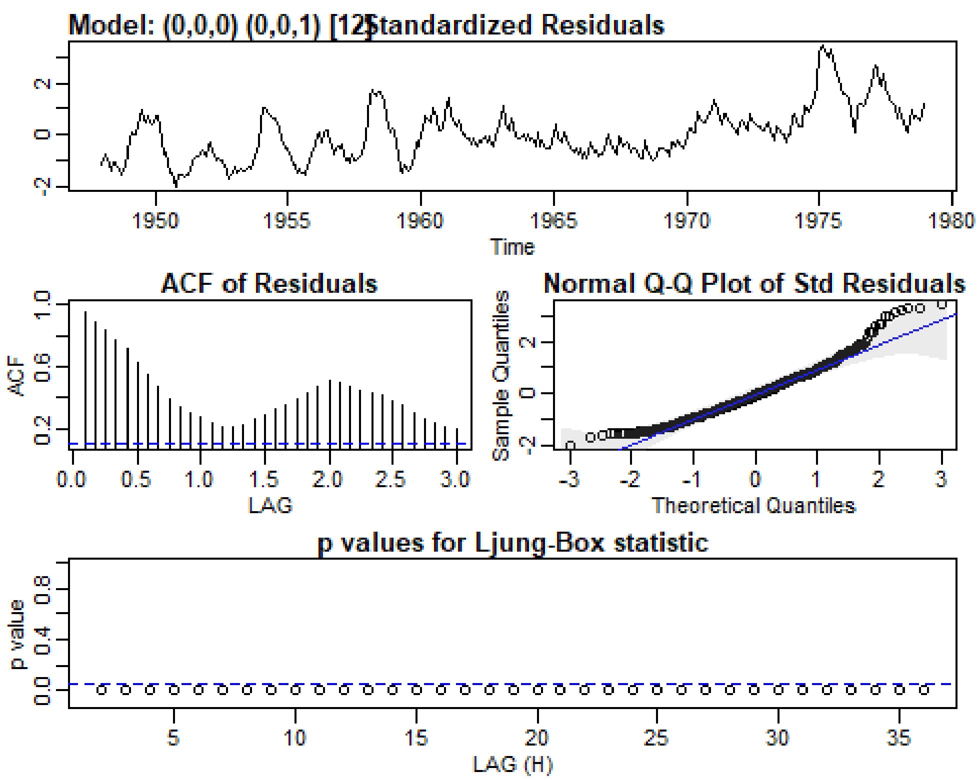
Ploting P/ACF pair of fully differenced data to lag 60




```
##         ACF  PACF
##  [1,]  0.21  0.21
##  [2,]  0.33  0.29
##  [3,]  0.15  0.05
##  [4,]  0.17  0.05
##  [5,]  0.10  0.01
##  [6,]  0.06 -0.02
##  [7,] -0.06 -0.12
##  [8,] -0.02 -0.03
##  [9,] -0.09 -0.05
## [10,] -0.17 -0.15
## [11,] -0.08  0.02
## [12,] -0.48 -0.43
```

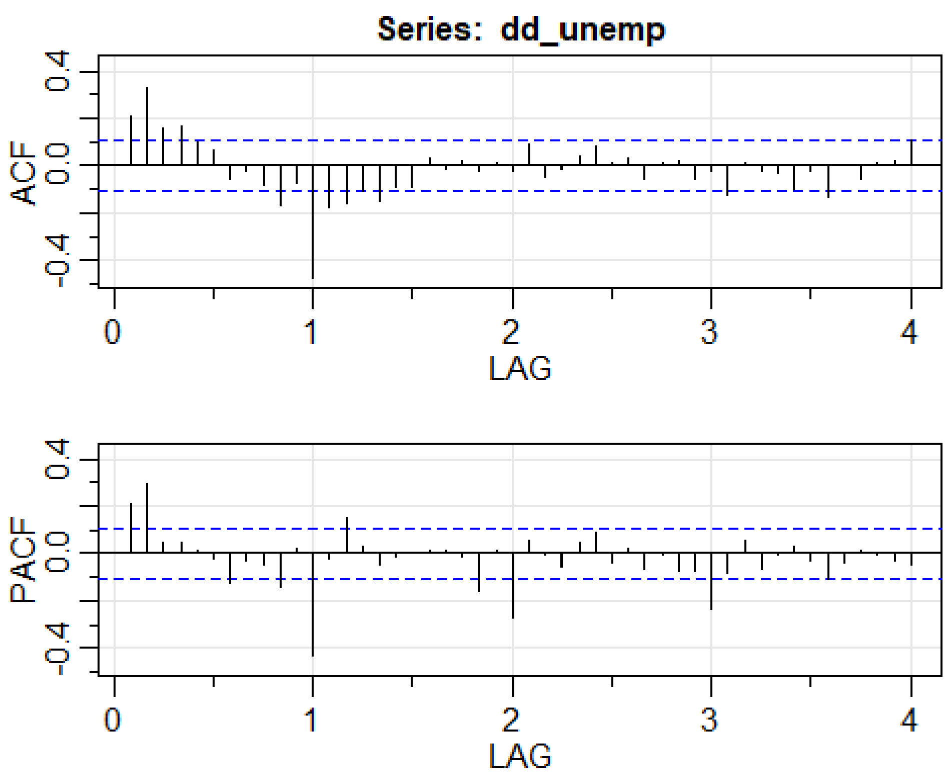
```
## [13,] -0.18 -0.02
## [14,] -0.16  0.15
## [15,] -0.11  0.03
## [16,] -0.15 -0.04
## [17,] -0.09 -0.01
## [18,] -0.09  0.00
## [19,]  0.03  0.01
## [20,] -0.01  0.01
## [21,]  0.02 -0.01
## [22,] -0.02 -0.16
## [23,]  0.01  0.01
## [24,] -0.02 -0.27
## [25,]  0.09  0.05
## [26,] -0.05 -0.01
## [27,] -0.01 -0.05
## [28,]  0.03  0.05
## [29,]  0.08  0.09
## [30,]  0.01 -0.04
## [31,]  0.03  0.02
## [32,] -0.05 -0.07
## [33,]  0.01 -0.01
## [34,]  0.02 -0.08
## [35,] -0.06 -0.08
## [36,] -0.02 -0.23
## [37,] -0.12 -0.08
## [38,]  0.01  0.06
## [39,] -0.03 -0.07
## [40,] -0.03 -0.01
## [41,] -0.10  0.03
## [42,] -0.02 -0.03
## [43,] -0.13 -0.11
## [44,]  0.00 -0.04
## [45,] -0.06  0.01
## [46,]  0.01  0.00
## [47,]  0.02 -0.03
## [48,]  0.11 -0.04

## initial  value 5.042407
## iter   2 value 4.787644
## iter   3 value 4.781286
## iter   4 value 4.772186
## iter   5 value 4.771781
## iter   6 value 4.771721
## iter   7 value 4.771689
## iter   8 value 4.771626
## iter   9 value 4.771617
## iter   9 value 4.771617
## iter   9 value 4.771617
## final  value 4.771617
## converged
```

```
## initial  value 4.762540
## iter   2 value 4.762263
## iter   3 value 4.762038
## iter   4 value 4.761972
## iter   5 value 4.761889
## iter   6 value 4.761887
## iter   7 value 4.761887
## iter   7 value 4.761887
## iter   7 value 4.761887
## final  value 4.761887
## converged
```




```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P,
D,
##     Q), period = S), xreg = xmean, include.mean = FALSE, transform.pars =
trans,
##     fixed = fixed, optim.control = list(trace = trc, REPORT = 1, reltol =
tol))
##
## Coefficients:
##         sma1     xmean
##       0.6503  391.8956
## s.e.  0.0344    9.7965
##
```

```
## sigma^2 estimated as 13441:  log likelihood = -2299.27,  aic = 4604.53
##
## $degrees_of_freedom
## [1] 370
##
## $ttable
##       Estimate     SE t.value p.value
## sma1    0.6503 0.0344 18.9035       0
## xmean 391.8956 9.7965 40.0035       0
##
## $AIC
## [1] 12.37778
##
## $AICc
## [1] 12.37787
##
## $BIC
## [1] 12.40938
```

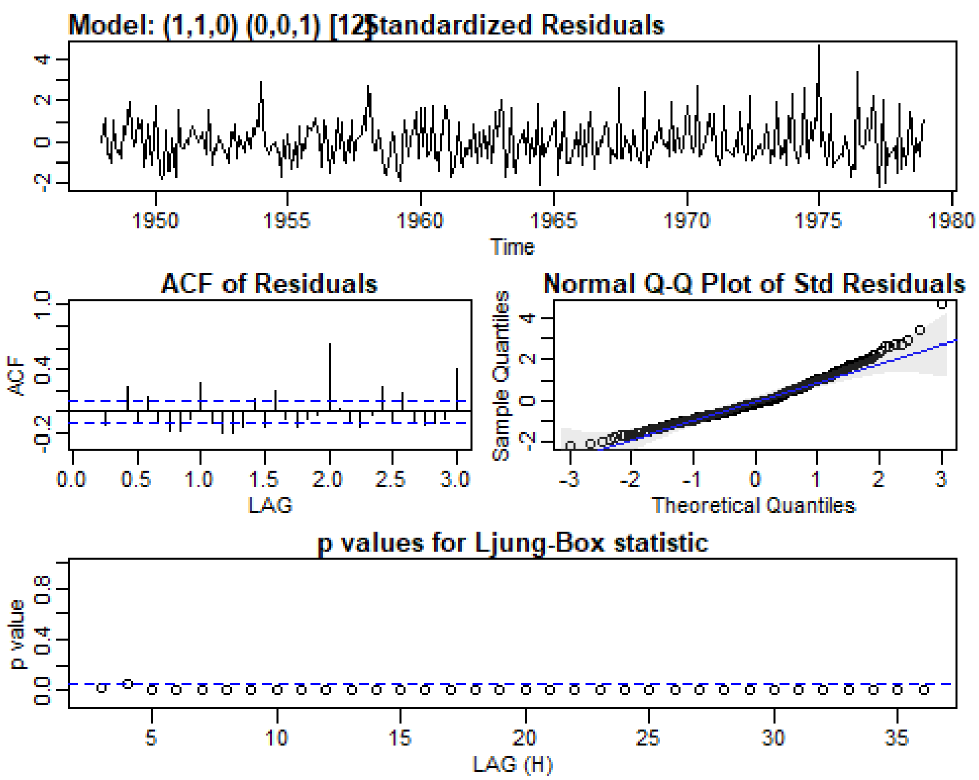



```
##         ACF  PACF
##  [1,]  0.21  0.21
##  [2,]  0.33  0.29
##  [3,]  0.15  0.05
##  [4,]  0.17  0.05
##  [5,]  0.10  0.01
##  [6,]  0.06 -0.02
##  [7,] -0.06 -0.12
```

```
##  [8,] -0.02 -0.03
##  [9,] -0.09 -0.05
## [10,] -0.17 -0.15
## [11,] -0.08  0.02
## [12,] -0.48 -0.43
## [13,] -0.18 -0.02
## [14,] -0.16  0.15
## [15,] -0.11  0.03
## [16,] -0.15 -0.04
## [17,] -0.09 -0.01
## [18,] -0.09  0.00
## [19,]  0.03  0.01
## [20,] -0.01  0.01
## [21,]  0.02 -0.01
## [22,] -0.02 -0.16
## [23,]  0.01  0.01
## [24,] -0.02 -0.27
## [25,]  0.09  0.05
## [26,] -0.05 -0.01
## [27,] -0.01 -0.05
## [28,]  0.03  0.05
## [29,]  0.08  0.09
## [30,]  0.01 -0.04
## [31,]  0.03  0.02
## [32,] -0.05 -0.07
## [33,]  0.01 -0.01
## [34,]  0.02 -0.08
## [35,] -0.06 -0.08
## [36,] -0.02 -0.23
## [37,] -0.12 -0.08
## [38,]  0.01  0.06
## [39,] -0.03 -0.07
## [40,] -0.03 -0.01
## [41,] -0.10  0.03
## [42,] -0.02 -0.03
## [43,] -0.13 -0.11
## [44,]  0.00 -0.04
## [45,] -0.06  0.01
## [46,]  0.01  0.00
## [47,]  0.02 -0.03
## [48,]  0.11 -0.04
```

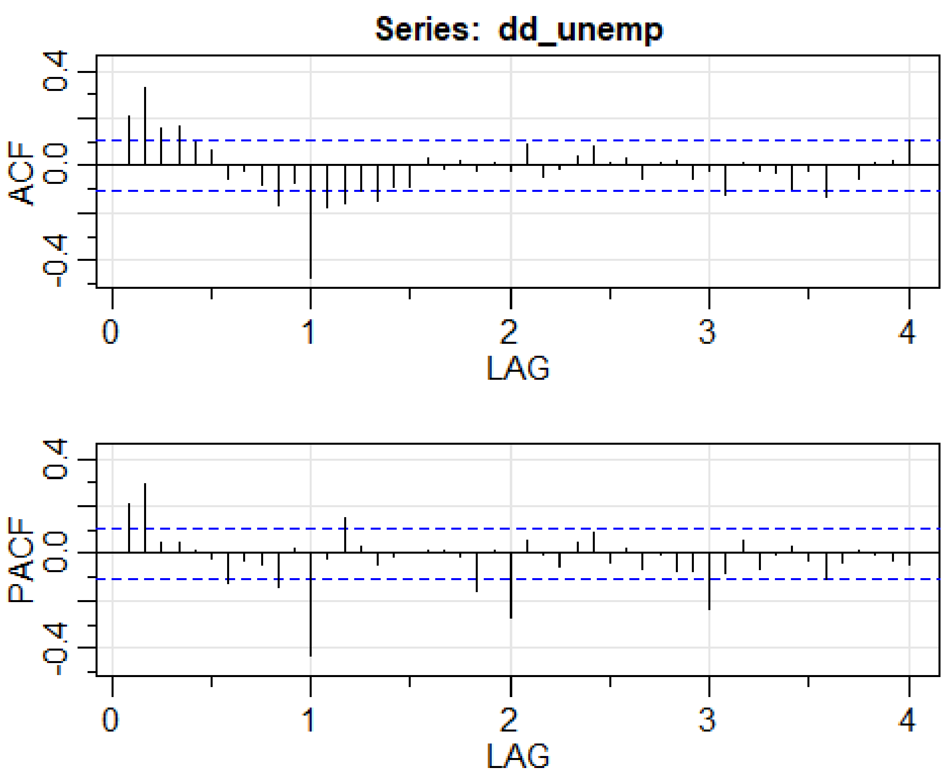
```
## initial  value 3.761438
## iter   2 value 3.608787
## iter   3 value 3.559781
## iter   4 value 3.536239
## iter   5 value 3.532296
## iter   6 value 3.531998
## iter   7 value 3.531995
## iter   8 value 3.531994
```

```
## iter   9 value 3.531994
## iter  10 value 3.531993
## iter  11 value 3.531993
## iter  11 value 3.531993
## iter  11 value 3.531993
## final  value 3.531993
## converged
## initial  value 3.530618
## iter   2 value 3.530579
## iter   3 value 3.530554
## iter   4 value 3.530551
## iter   5 value 3.530546
## iter   6 value 3.530546
## iter   6 value 3.530546
## iter   6 value 3.530546
## final  value 3.530546
## converged
```




```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P,
D,
##     Q), period = S), xreg = constant, transform.pars = trans, fixed =
fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
```

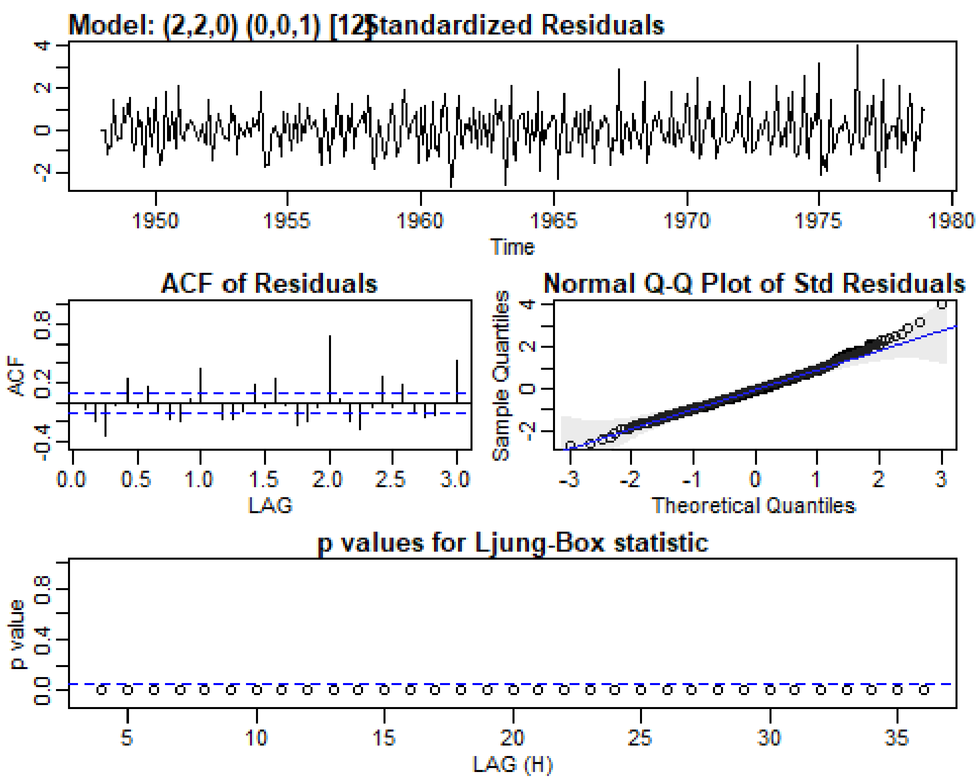
```
## Coefficients:
##          ar1    sma1  constant
##       0.1086  0.5543    0.8540
## s.e.  0.0525  0.0357    3.0369
##
## sigma^2 estimated as 1152:  log likelihood = -1836.26,  aic = 3680.52
##
## $degrees_of_freedom
## [1] 368
##
## $ttable
##          Estimate     SE t.value p.value
## ar1        0.1086 0.0525  2.0680  0.0393
## sma1       0.5543 0.0357 15.5144  0.0000
## constant   0.8540 3.0369  0.2812  0.7787
##
## $AIC
## [1] 9.920532
##
## $AICc
## [1] 9.920709
##
## $BIC
## [1] 9.962756
```




```
##        ACF  PACF
##  [1,]  0.21  0.21
```

```
##  [2,]  0.33  0.29
##  [3,]  0.15  0.05
##  [4,]  0.17  0.05
##  [5,]  0.10  0.01
##  [6,]  0.06 -0.02
##  [7,] -0.06 -0.12
##  [8,] -0.02 -0.03
##  [9,] -0.09 -0.05
## [10,] -0.17 -0.15
## [11,] -0.08  0.02
## [12,] -0.48 -0.43
## [13,] -0.18 -0.02
## [14,] -0.16  0.15
## [15,] -0.11  0.03
## [16,] -0.15 -0.04
## [17,] -0.09 -0.01
## [18,] -0.09  0.00
## [19,]  0.03  0.01
## [20,] -0.01  0.01
## [21,]  0.02 -0.01
## [22,] -0.02 -0.16
## [23,]  0.01  0.01
## [24,] -0.02 -0.27
## [25,]  0.09  0.05
## [26,] -0.05 -0.01
## [27,] -0.01 -0.05
## [28,]  0.03  0.05
## [29,]  0.08  0.09
## [30,]  0.01 -0.04
## [31,]  0.03  0.02
## [32,] -0.05 -0.07
## [33,]  0.01 -0.01
## [34,]  0.02 -0.08
## [35,] -0.06 -0.08
## [36,] -0.02 -0.23
## [37,] -0.12 -0.08
## [38,]  0.01  0.06
## [39,] -0.03 -0.07
## [40,] -0.03 -0.01
## [41,] -0.10  0.03
## [42,] -0.02 -0.03
## [43,] -0.13 -0.11
## [44,]  0.00 -0.04
## [45,] -0.06  0.01
## [46,]  0.01  0.00
## [47,]  0.02 -0.03
## [48,]  0.11 -0.04

## initial  value 4.094456
## iter   2 value 3.765216
```

```
## iter   3 value 3.740103
## iter   4 value 3.697745
## iter   5 value 3.692276
## iter   6 value 3.685320
## iter   7 value 3.685056
## iter   8 value 3.685050
## iter   8 value 3.685050
## final  value 3.685050
## converged
## initial  value 3.685822
## iter   2 value 3.685797
## iter   3 value 3.685782
## iter   4 value 3.685781
## iter   4 value 3.685781
## iter   4 value 3.685781
## final  value 3.685781
## converged
```




```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P,
D,
##     Q), period = S), include.mean = !no.constant, transform.pars = trans,
fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
```

```
## Coefficients:
##           ar1      ar2    sma1
##       -0.5467  -0.2188  0.6197
## s.e.   0.0509   0.0510  0.0331
##
## sigma^2 estimated as 1564:  log likelihood = -1888.75,  aic = 3785.49
##
## $degrees_of_freedom
## [1] 367
##
## $ttable
##      Estimate     SE  t.value p.value
## ar1   -0.5467 0.0509 -10.7384       0
## ar2   -0.2188 0.0510  -4.2940       0
## sma1   0.6197 0.0331  18.7049       0
##
## $AIC
## [1] 10.23106
##
## $AICc
## [1] 10.23124
##
## $BIC
## [1] 10.27337
```

Based on the above models, my first arima model (p=2,d=1,q=0,P=0,D=1,Q=1,S=12) has the best fit.The AIC and BIC values for it are lowest (AIC=7.12457 and BIC=6.156174) From this model, residuals are small in magnitude, appear to be uncorrelated, close to normal and have a mean near zero.