Final_Exam_TS

Ali Alghaithi

4/27/2019

1

1A

$$
\begin{aligned}
E(Y_t) &= c_1 EX_{t-1} + ... + c_s EX_{t-s+1} \\
&= (c_1 + c_2 + ... + c_s)\mu_Y \\
Cov(Y_t, Y_{t-k}) &= Cov(c_1 X_t + c_2 X_{t-1} + ... + c_s X_{t-s+1}, c_1 X_{t-k} + c_2 X_{t-1-k} + ... c_s X_{t-s+1-k}) \\
&= \sum_{j=0}^{s} \sum_{i=0}^{s} c_j c_i Cov(X_{t-j}, X_{t-k-i}) \\
&= \sum_{j=0}^{s} \sum_{i=0}^{s} c_j c_i \gamma_{j-k-i}
\end{aligned}
$$

- Therefore the equation is free from t and is stationary since the mean is constantas and free of t as well.

1B

ARMA(p,q) ψ weights:

$$
\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= -\theta_1 + \phi_1 \\
\psi_2 &= -\theta_2 + \phi_2 + \phi_1 \psi_1 \\
\psi_j &= -\theta_j + \phi_p \psi_{j-p} + \phi_{p-1} \psi_{j-p+1} + ... + \phi_1 \psi_{j-1}
\end{aligned}
$$

We have that X_t to be ARMA(p,q) process where it seems to us that as MA(q) process.

We have that,

$$
\begin{aligned}
\phi_1 &= \phi_2 \\
&= \phi_3 \\
&= 0
\end{aligned}
$$

$$
\psi_0 = 1
$$

$$
\begin{aligned}
\psi_1 &= -\theta + \phi_1 \\
&= c_2 + 0 \\
&= c_2
\end{aligned}
$$

$$
\begin{aligned}
\psi_2 &= -\theta_2 + \phi_2 + \phi_1\psi_1 \\
&= c_3 + 0 + (0)(c_2) \\
&= c_3
\end{aligned}
$$

where then we have that,

$$
\psi_j = -\theta_j + \phi_p\psi_{j-p} + \phi_{p-1}\psi_{j-p+1} + ... + \phi_1\psi_{j-1}
$$

2

2A

Confirm stationarity with AR characteristic equation:

$$
\phi(x) = 1 - 2.8x + 2.6x^2 - 0.8x^3 = 0
$$

The root will be $x = 1.25$ where we see it is greater than 1,then stationarity. where we can show invertibility with MA characteristic equation:

$$
\theta(x) = 1 - 0.76x - 0.2x^2 = 0
$$

The roots are $x = -4.8323$ and $x = 1.0343$ and we can see that they are greater than 1 in modulus, where this show invertibility.

Then,

$$
\begin{aligned}
ARIMA(p,d,q) &= ARIMA(2,1,2) \\
\phi_1 &= 1.8 \\
\phi_2 &= -0.8 \\
\theta_0 &= 0 \\
\theta_1 &= 0.76 \\
\theta_2 &= 0.2
\end{aligned}
$$

where that,

$$
\begin{aligned}
ARMA(p+d,q) &= ARMA(3,2)\\
\phi_1 &= 2.8\\
\phi_2 &= -2.6\\
\phi_3 &= 0.8\\
\theta_0 &= 0\\
\theta_1 &= 0.76\\
\theta_2 &= 0.2
\end{aligned}
$$

2B

$$
\begin{aligned}
\text{Equation} &: \hat{Y}_t = e_t + Y_{t-4} + (-0.4)(Y_{t-1} - Y_{t-5}) - 0.6e_{t-1} - 0.5e_{t-4} + (0.6)(0.5)e_{t-5}\\
\text{Seasons} &= 4\\
ARIMA(p,d,q) \times (P,D,Q)_s &= ARIMA(1,0,1) \times (0,1,1)_4\\
\phi &= -0.4\\
\theta &= 0.6\\
\Phi &= 0\\
\Theta &= 0.5
\end{aligned}
$$

3

3A

$$
\begin{aligned}
\gamma_0 &= \frac{4}{3}\sigma_e^2 \\
\gamma_1 &= 0.4\sigma_e^2 \\
\gamma_2 &= -\frac{23}{75}\sigma_e^2 \\
\rho_1 &= \frac{\gamma_1}{\gamma_0} = 0.3 \\
\rho_2 &= \frac{\gamma_2}{\gamma_0} = -0.23 \\
\rho_3 &= 0.9\rho_2 - 0.5\rho_1 = -0.357 \\
\phi_{11} &= \rho_1 = 0.3 \\
\phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{-0.23 - (0.3)^2}{1 - (0.3)^2} \\
&= -0.3516 \\
\phi_{21} &= \phi_{11} - \phi_{22}\phi_{11} = 0.4055 \\
\phi_{33} &= \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_1 - \phi_{22}\rho_2} \\
&= \frac{-0.357 - (0.4055 \cdot -0.23) - (-0.3516 \cdot 0.3)}{1 - (0.4055 \cdot 0.3) - (-0.3516 \cdot -0.23)} \\
&= -0.1984
\end{aligned}
$$

Then, we have that:

$$
\begin{aligned}
\phi_{11} &= 0.3 \\
\phi_{22} &= -0.3516 \\
\phi_{33} &= -0.1984
\end{aligned}
$$

3B

$$
\begin{aligned}
\hat{\mu} &= \bar{Y} \\
&= \frac{1}{n}\sum_{i=1}^{n} y_i \\
&= \frac{1}{5}(21+19+18+20+21) = 19.8 \\
\hat{\gamma}_0 &= s^2 \\
&= \frac{1}{n-1}\sum_{t-1}^{n}(y_t-\bar{Y})^2 \\
&= \frac{1}{4}((21-19.8)^2+(19-19.8)^2+(18-19.8)^2+(20-19.8)^2+(21-19.8)^2) = 1.7 \\
\hat{\rho}_1 &= \frac{((y_2-\bar{Y})\cdot(y_1-\bar{Y}))+((y_3-\bar{Y})\cdot(y_2-\bar{Y}))+((y_4-\bar{Y})\cdot(y_3-\bar{Y}))+((y_5-\bar{Y})\cdot(y_4-\bar{Y}))}{(y_1-\bar{Y})^2+(y_2-\bar{Y})^2+(y_3-\bar{Y})^2+(y_4-\bar{Y})^2+(y_5-\bar{Y})^2} \\
&= \frac{((19-19.8)\cdot(21-19.8))+((18-19.8)\cdot(19-19.8))+((20-19.8)\cdot(18-19.8))+((21-19.8)\cdot(20-19.8))}{(21-19.8)^2+(19-19.8)^2+(18-19.8)^2+(20-19.8)^2+(21-19.8)^2} \\
&= 0.05294117647 \\
\hat{\rho}_2 &= \frac{((y_3-\bar{Y})\cdot(y_2-\bar{Y}))+((y_4-\bar{Y})\cdot(y_3-\bar{Y}))+((y_5-\bar{Y})\cdot(y_4-\bar{Y}))}{(y_2-\bar{Y})^2+(y_3-\bar{Y})^2+(y_4-\bar{Y})^2+(y_5-\bar{Y})^2} \\
&= \frac{((18-19.8)\cdot(19-19.8))+((20-19.8)\cdot(18-19.8))+((21-19.8)\cdot(20-19.8))}{(19-19.8)^2+(18-19.8)^2+(20-19.8)^2+(21-19.8)^2} \\
&= 0.24626865671 \\
\hat{\rho}_3 &= \frac{((y_4-\bar{Y})\cdot(y_3-\bar{Y}))+((y_5-\bar{Y})\cdot(y_4-\bar{Y}))}{(y_3-\bar{Y})^2+(y_4-\bar{Y})^2+(y_5-\bar{Y})^2} \\
&= \frac{((20-19.8)\cdot(18-19.8))+((21-19.8)\cdot(20-19.8))}{(18-19.8)^2+(20-19.8)^2+(21-19.8)^2} \\
&= -0.02542372881
\end{aligned}
$$

3C

$$
\begin{aligned}
\theta_0 &= \mu(1-\phi_1-\phi_2) \\
&= 19.8\cdot(1-0.9-(-0.5)) = 11.88 \\
\hat{Y}_5(1) &= \phi_1 Y_5+\phi_2 Y_4+\theta_0+\theta e_5 \\
&= (0.9\cdot 21)+((-0.5)\cdot 20)+11.88+(0.6\cdot 1) = 21.38 \\
\hat{Y}_5(2) &= \phi_1\hat{Y}_5(1)+\phi_2 Y_5+\theta_0 \\
&= (0.9\cdot 21.38)+((-0.5)\cdot 21)+11.88 = 20.622 \\
\hat{Y}_5(3) &= \phi_1\hat{Y}_5(2)+\phi_2\hat{Y}_5(1)+\theta_0 \\
&= (0.9\cdot 20.622)+((-0.5)\cdot 21.38)+11.88 = 19.7498 \\
\hat{Y}_5(4) &= \phi_1\hat{Y}_5(3)+\phi_2\hat{Y}_5(2)+\theta_0 \\
&= (0.9\cdot 19.7498)+((-0.5)\cdot 20.622)+11.88 = 19.34382 \\
\hat{Y}_5(5) &= \phi_1\hat{Y}_5(4)+\phi_2\hat{Y}_5(3)+\theta_0 \\
&= (0.9\cdot 19.34382)+((-0.5)\cdot 19.7498)+11.88 = 19.414538
\end{aligned}
$$

3D

$$
\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \phi_1\psi_0 = 0.9 \cdot 1 = 0.9 \\
\psi_2 &= \phi_1\psi_1 + \phi_2\psi_0 = (0.9 \cdot 0.9) + ((-0.5) \cdot 1) \\
&= 0.31 \\
\psi_3 &= \phi_1\psi_2 + \phi_2\psi_1 \\
&= (0.9 \cdot 0.31) + ((-0.5) \cdot 0.9) \\
&= -0.171 \\
\psi_4 &= \phi_1\psi_3 + \phi_2\psi_2 \\
&= (0.9 \cdot (-0.171)) + ((-0.5) \cdot 0.31) \\
&= -0.3089 \\
\text{Then, now we can find :} & \\
Var(e_t(1)) &= \sigma_e^2 = 1 \\
Var(e_t(2)) &= \sigma_e^2(1 + \psi_1^2) = 1(1 + (0.9)^2) \\
&= 1.81 \\
Var(e_t(3)) &= \sigma_e^2(1 + \psi_1^2 + \psi_2^2) = 1(1 + (0.9)^2 + (0.31)^2) \\
&= 1.9061 \\
Var(e_t(4)) &= \sigma_e^2(1 + \psi_1^2 + \psi_2^2 + \psi_3^2) \\
&= 1(1 + (0.9)^2 + (0.31)^2 + (-0.171)^2) \\
&= 1.935341 \\
Var(e_t(5)) &= \sigma_e^2(1 + \psi_1^2 + \psi_2^2 + \psi_3^2 + \psi_4^2) \\
&= 1(1 + (0.9)^2 + (0.31)^2 + (-0.171)^2 + (-0.3089)^2) \\
&= 2.03076021
\end{aligned}
$$

3E

- The 80% confidence intervals are

$$
\begin{aligned}
\hat{Y}_5(1) &= \hat{Y}_5(1) \pm (\sqrt{Var(e_t(1))} \cdot 1.28) \\
&= 21.38 \pm (\sqrt{1} \cdot 1.28) \\
&= 21.38 \pm 1.28 \\
\hat{Y}_5(2) &= \hat{Y}_5(2) \pm (\sqrt{Var(e_t(2))} \cdot 1.28) \\
&= \hat{Y}_5(2) \pm (\sqrt{1.81} \cdot 1.28) \\
&= 20.622 \pm 1.72206387803 \\
\hat{Y}_5(3) &= \hat{Y}_5(3) \pm (\sqrt{Var(e_t(3))} \cdot 1.28) \\
&= \hat{Y}_5(3) \pm (\sqrt{1.9061} \cdot 1.28) \\
&= 19.7498 \pm 1.76718822993 \\
\hat{Y}_5(4) &= \hat{Y}_5(4) \pm (\sqrt{Var(e_t(4))} \cdot 1.28) \\
&= \hat{Y}_5(4) \pm (\sqrt{1.935341} \cdot 1.28) \\
&= 19.34382 \pm 1.78069163372 \\
\hat{Y}_5(5) &= \hat{Y}_5(5) \pm (\sqrt{Var(e_t(5))} \cdot 1.28) \\
&= \hat{Y}_5(5) \pm (\sqrt{2.03076021} \cdot 1.28) \\
&= 19.414538 \pm 1.82406072488
\end{aligned}
$$

3F

- Then we need to find $\hat{Y}_6(1)$,$\hat{Y}_6(2)$,$\hat{Y}_6(3)$,and $\hat{Y}_6(4)$ so we can the required forcasts

$$
\begin{aligned}
\hat{Y}_6(1) &= \hat{Y}_5(2) + \psi_1[Y_6 - \hat{Y}_5(1)] \\
&= 20.622 + (0.9)[19 - 21.38] \\
&= 18.48 \\
\hat{Y}_6(2) &= \hat{Y}_5(3) + \psi_2[Y_6 - \hat{Y}_5(1)] \\
&= 19.7498 + (0.31)[19 - 21.38] \\
&= 19.012 \\
\hat{Y}_6(3) &= \hat{Y}_5(4) + \psi_3[Y_6 - \hat{Y}_5(1)] \\
&= 19.34382 + (-0.171)[19 - 21.38] \\
&= 19.7508 \\
\hat{Y}_6(4) &= \hat{Y}_5(5) + \psi_4[Y_6 - \hat{Y}_5(1)] \\
&= 19.414538 + (-0.3089)[19 - 21.38] \\
&= 20.14972
\end{aligned}
$$

- Now we can use them to find $\hat{Y}_7(1)$,$\hat{Y}_7(1)$, and $\hat{Y}_7(1)$

Then we can find:

$$
\begin{aligned}
\hat{Y}_7(1) &= \hat{Y}_6(2) + \psi_1[Y_7 - \hat{Y}_6(1)] \\
&= 19.012 + (0.9)[18 - 18.48] \\
&= 18.58 \\
\hat{Y}_7(2) &= \hat{Y}_6(3) + \psi_2[Y_7 - \hat{Y}_6(1)] \\
&= 19.7508 + (0.31)[18 - 18.48] \\
&= 19.602 \\
\hat{Y}_7(3) &= \hat{Y}_6(4) + \psi_3[Y_7 - \hat{Y}_6(1)] \\
&= 20.14972 + (-0.171)[18 - 18.48] \\
&= 20.2318
\end{aligned}
$$

4

4A

```r
dat = read.csv("/Users/alialghaithi/Downloads/Nebraska_LaborForce_UnemploymentRate.csv")
labor = ts(dat$LaborForce,start=1976,frequency=12)
unemployment = ts(dat$UnemploymentRate,start=1976,frequency=12)
```

- Fit an appropriate model to labor. Forecast the next 3 years with 95% prediction limits

labor Data

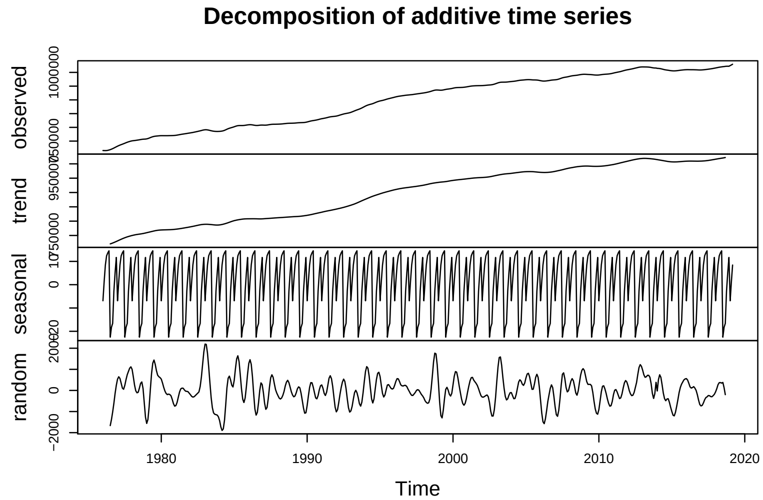
labor: data cleaning stage and understanding

```r
library(TSA)
library(cowplot)
library(forecast)
set.seed(38180987)
#Reading the data
dat = read.csv("/Users/alialghaithi/Downloads/Nebraska_LaborForce_UnemploymentRate.csv")
labor = ts(dat$LaborForce,start=1976,frequency=12)
unemployment = ts(dat$UnemploymentRate,start=1976,frequency=12)

#checking if we have missing values
any(is.na(labor))
```

```
## [1] FALSE
```

- After looking at the data, we can see that our data is cleaned

```r
plot(decompose(labor))
```



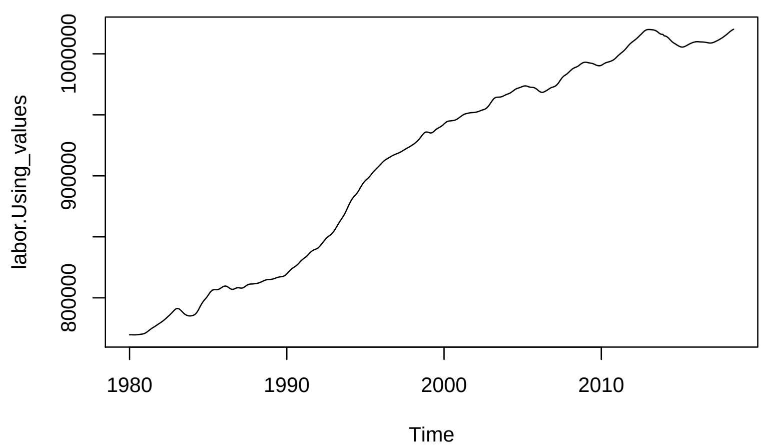
- Summary of the Data by using decompose function, we can now see and know the data better. It seems that we can see some seasonal trend. We might want to consider a seasonal model to this data. the Data is increasing over time.

```
summary(labor)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  715419  808372  922210  898169  988966 1029708
```

- The summary of the data help us understadn the data better, and what do the value inside the time series as well.

```
labor.Real_values <- window(labor,start = c(2018,7),end=c(2019,3))
labor.Using_values <- window(labor,start=1980,end =c(2018,6))
plot(labor.Using_values)
```



- I decide to use the data until the end of December of 2018, and will try to see if we can forcast the three months of 2019 to see if our model is good. We see that we need to try few transformation methods.

labor: transformation,deterministic trend model, and seasonal model

```r
# Seasonal Differencing
nsdiffs(labor.Using_values)  # number for seasonal differencing needed
```

```
## [1] 0
```

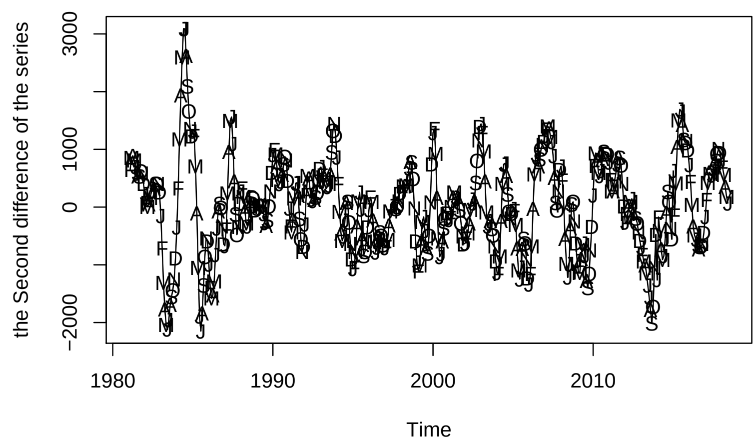
- Then we do not need to have any Differencing

```r
ndiffs(labor.Using_values)  # number of differences need to make it stationary
```

```
## [1] 2
```

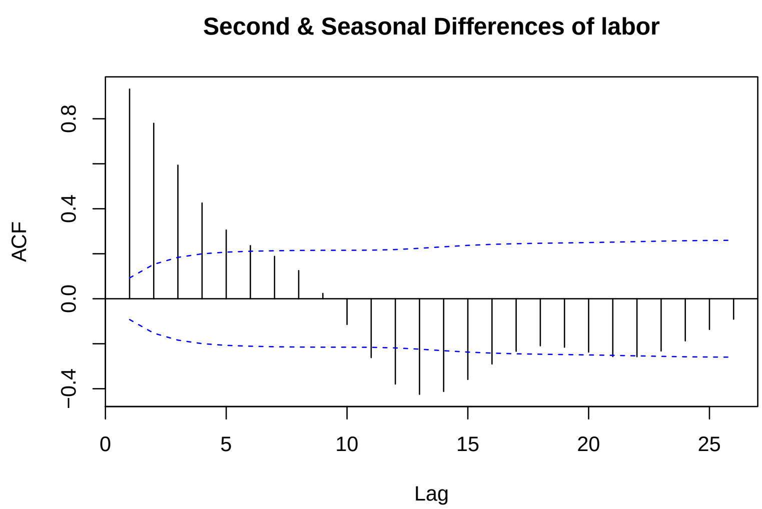
- Then we are going to have to Differencing the data 2 times

```r
 labor.transformed <- diff(diff(labor.Using_values),lag = 12)
plot(labor.transformed,type='l',ylab='the Second difference of the series')
points(labor.transformed,x=time(labor.transformed), pch=as.vector(season(labor.transformed)))
```



- This plot have been used with seasonal plotting symbols. The seasonality is much less obvious now. Some Decem- bers are high and some low. Similarly, some Octobers are high and some low.

```r
acf(as.vector(labor.transformed),ci.type='ma',
   main='Second & Seasonal Differences of labor')
```



this plots strongly suggest an ARIMA seasonal model for this series. An ARIMA(0,1,0)(0,2,0)[12] might also be tried.

labor: model specifcation AND parameter estimation

```r
#choosign a model using auto.arima
auto.arima(labor.transformed)
```

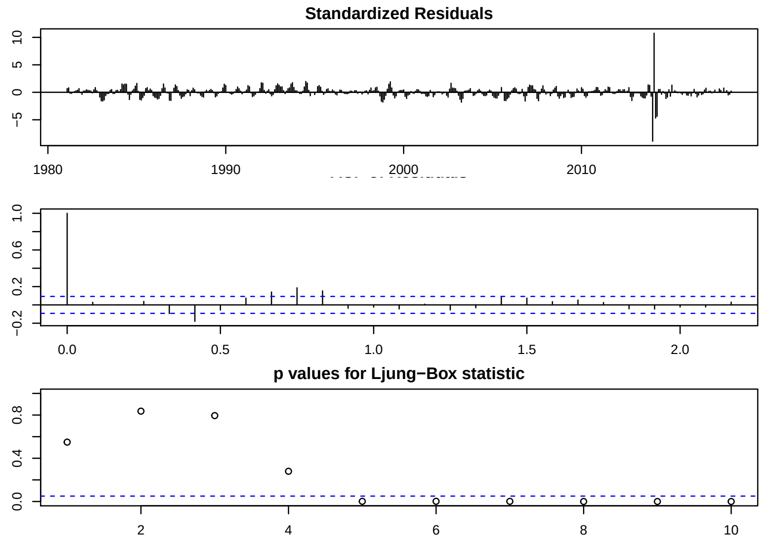
```
## Series: labor.transformed
## ARIMA(2,0,0)(2,0,0)[12] with zero mean
##
## Coefficients:
##          ar1      ar2     sar1     sar2
##       1.5609  -0.6470  -0.7193  -0.3724
## s.e.  0.0357   0.0358   0.0438   0.0427
##
## sigma^2 estimated as 28898:  log likelihood=-2946.56
## AIC=5903.12   AICc=5903.26   BIC=5923.66
```

- Using auto.arima to see the suggested model for this series. But I am also thinking of other 2 models that might fit the data well.
- ARIMA(2,0,2)(2,0,2)[12] from auto.arima
- ARIMA(0,1,0)(0,1,0)[12]
- ARIMA(0,1,0)(0,2,0)[12]

```r
###Models that I tried:

#ARIMA (2,0,1)
#labor_model=arima(labor.transformed,order=c(2,0,1))

#ARIMA(1,4,0):seasonally integrate, performing a model
#labor_model=arima(labor.transformed,order=c(1,4,0))

#ARIMA(0,1,4)
#labor_model=arima(labor.transformed,order=c(0,1,4))




#ARIMA(2,0,2)(2,0,2)[12] with non-zero mean
labor_model <-arima(labor.transformed,order = c(2,0,2),seasonal = list(order=c(2,0,2),period=12))

#ARIMA(1,0,0)(2,0,0)[12]
# labor_model <-arima(labor.transformed,order = c(0,1,0),seasonal = list(order=c(0,1,0),period=12))

#ARIMA(0,1,0)(0,2,0)[12]
# labor_model <-arima(labor.transformed,order = c(0,1,0),seasonal = list(order=c(0,2,0),period=12))
```

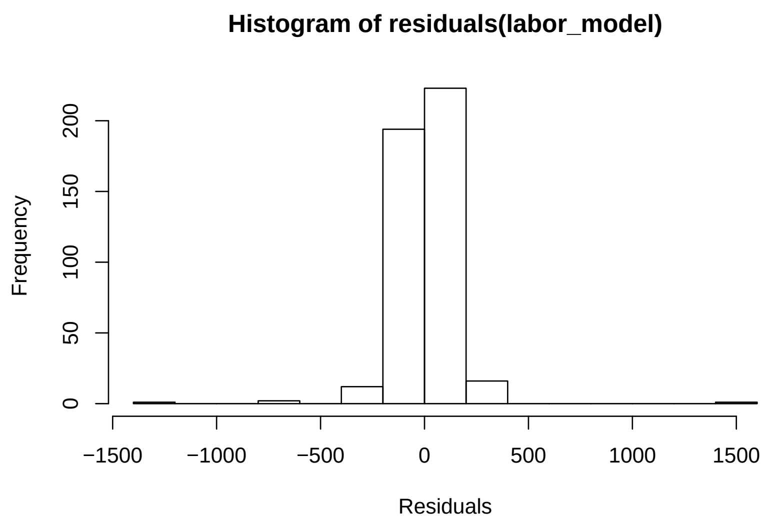
- Then now will try ARIMA(2,0,2)(2,0,2)[12] from auto.arima to see if it is good or no.

```r
##cheaking the model
par(mar = rep(2, 4))

library(grDevices)
tsdiag(labor_model)
```

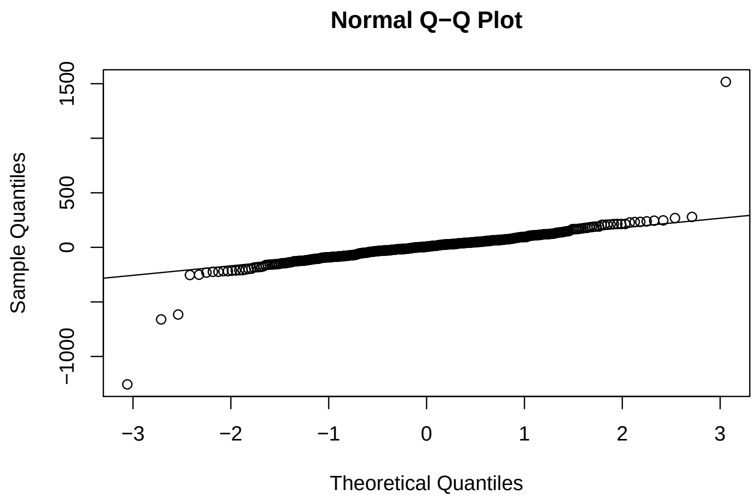


There are also around 4 residual acf values outside the critical limits. This Model diagnostics is yelling us that the Ljung-Box statistics are better as shown in the bottom display.

```
#normality
hist(residuals(labor_model),xlab = 'Residuals')
```




```
qqnorm(residuals(labor_model),plot.it = TRUE)
qqline(residuals(labor_model))
```



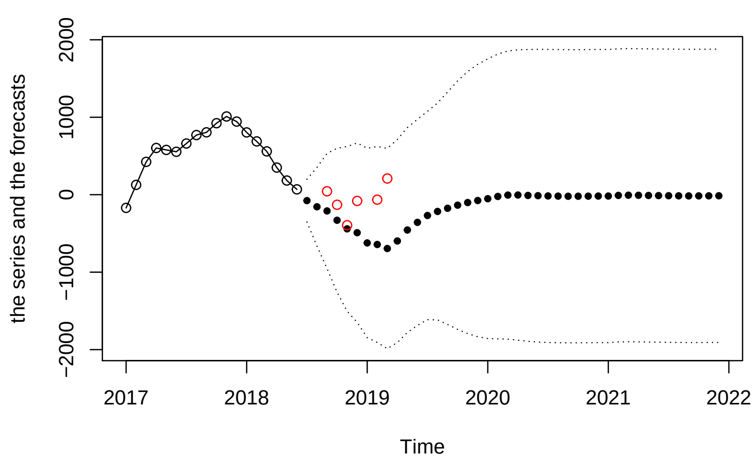

```r
shapiro.test(residuals(labor_model))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(labor_model)
## W = 0.72671, p-value < 2.2e-16
```

- Both the Q-Q plot and the results of the Shapiro-Wilk test indicate that we should reject of normality for the error terms in this model but this could be caused by the suspected outliers in the series. So, we will try to go with this model.

labor: forecasting

```r
# for th next 3 years \
library(TSA)
outputs_plot = TSA:::plot.Arima(labor_model,n1=2017,n.ahead=42
                                ,ylab='the series and the forecasts',pch=20)
points(diff(diff(labor.Real_values)),col="red")
```



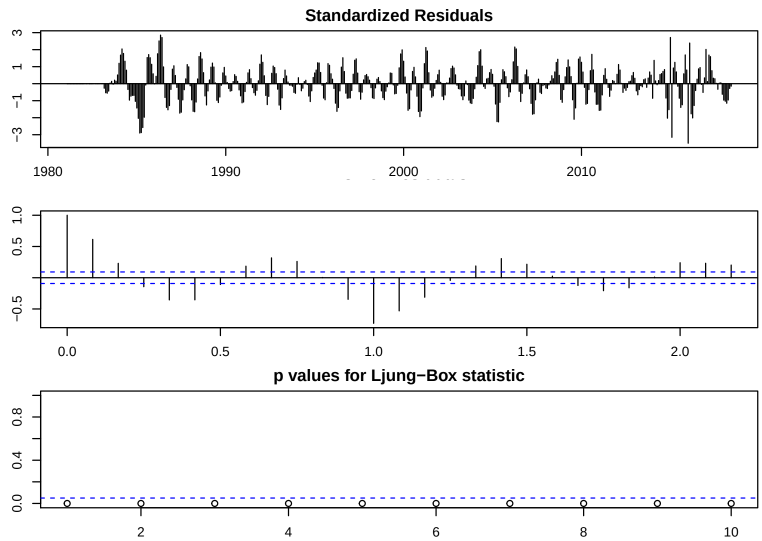
- The forecat limits contain all of the actual values but they are quite wide In addition,since, the model does not contain a lot of autocorrelation or other structure, the forecasts, plotted as solid circles, quickly settle down to the mean of the series after 2020.

labor Data: Model Number 2 after trying many model

- We will try to use ARIMA(0,1,0)(0,2,0)[12]

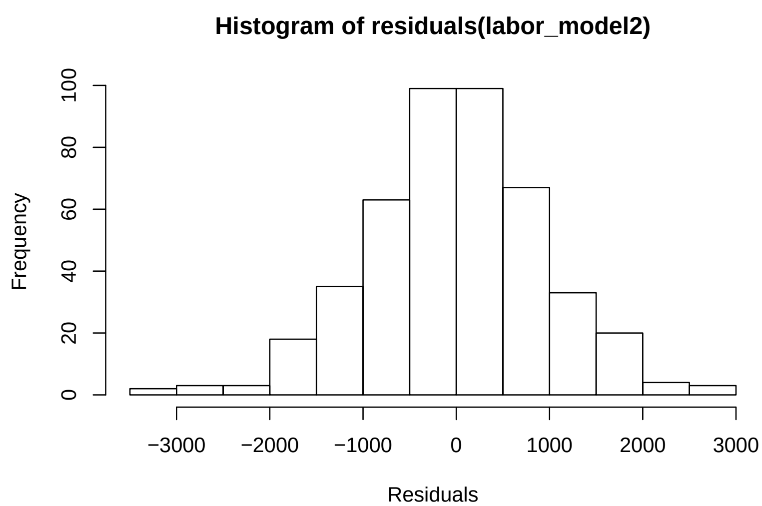
```
labor_model2 <-arima(labor.transformed,order = c(0,1,0),seasonal = list(order=c(0,2,0),period=12))
```

```
##cheaking the model
par(mar = rep(2, 4))

library(grDevices)
tsdiag(labor_model2)
```

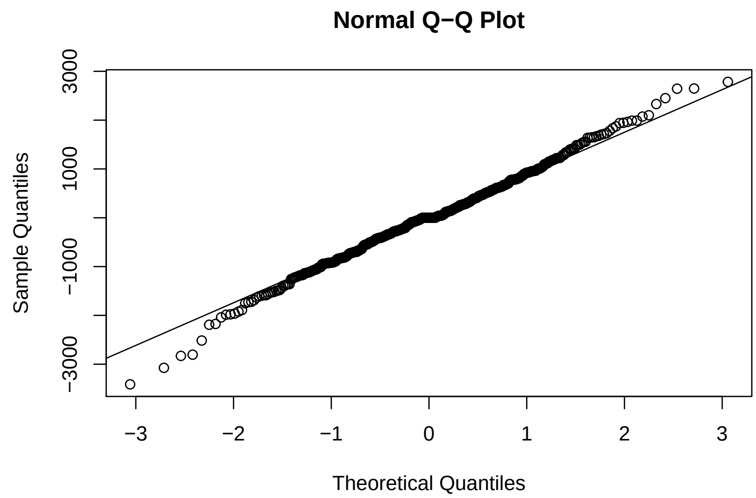


-The bottom display shows the p-values of the Ljung-Box test for a variety of values of the "K" parameter—the highest lag used in the sum. The top display will flag potential outliers, if any, using the Bonferroni criteria.

```
#normality
hist(residuals(labor_model2),xlab = 'Residuals')
```




```
qqnorm(residuals(labor_model2),plot.it = TRUE)
qqline(residuals(labor_model2))
```



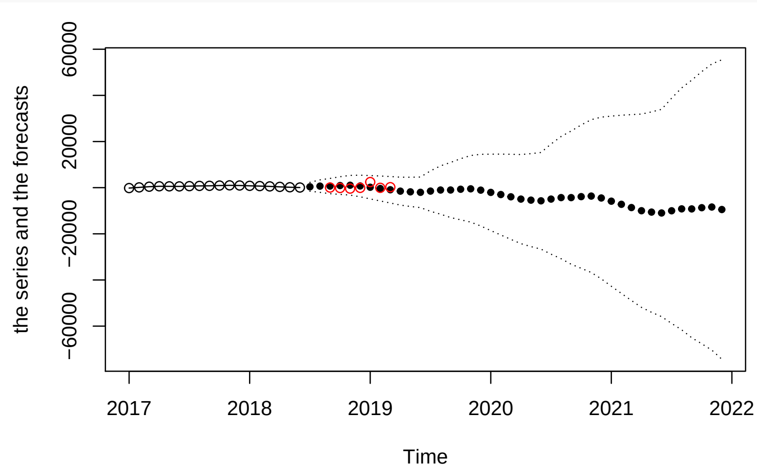

```
shapiro.test(residuals(labor_model2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(labor_model2)
## W = 0.99446, p-value = 0.104
```

- Both the Q-Q plot and the results of the Shapiro-Wilk test indicate that we should reject of normality for the error terms in this model but this could be caused by the suspected outliers in the series as well. So, we will try to go with this model. But we see that now this model more normal than the old model so it is better than the old model.

labor: forecasting

```
# for th next 3 years \
library(TSA)
outputs_plot = TSA:::plot.Arima(labor_model2,n1=2017,n.ahead=42
                                ,ylab='the series and the forecasts',pch=20)
points(diff(diff(labor.Real_values)),col="red")
```



- The forecat limits contain all of the actual values but they are quite wide. The forecasts mimic the pseudo periodic nature of the series but also decay toward the series mean as they go further into the future. We will use this model than the other models that We have tried.

unemployment Data

unemployment: data cleaning stage and understanding

```
library(TSA)
library(cowplot)
library(forecast)
set.seed(38180987)
#Reading the data
dat = read.csv("/Users/alialghaithi/Downloads/Nebraska_LaborForce_UnemploymentRate.csv")
labor = ts(dat$LaborForce,start=1976,frequency=12)
unemployment = ts(dat$UnemploymentRate,start=1976,frequency=12)

#checking if we have missing values
any(is.na(unemployment))
```

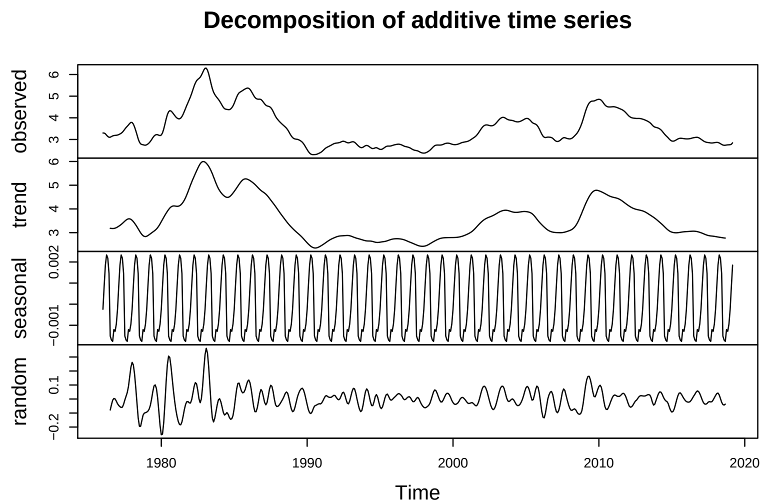
```
## [1] FALSE
```

```
summary(unemployment)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.300   2.830   3.196   3.515   3.989   6.292
```

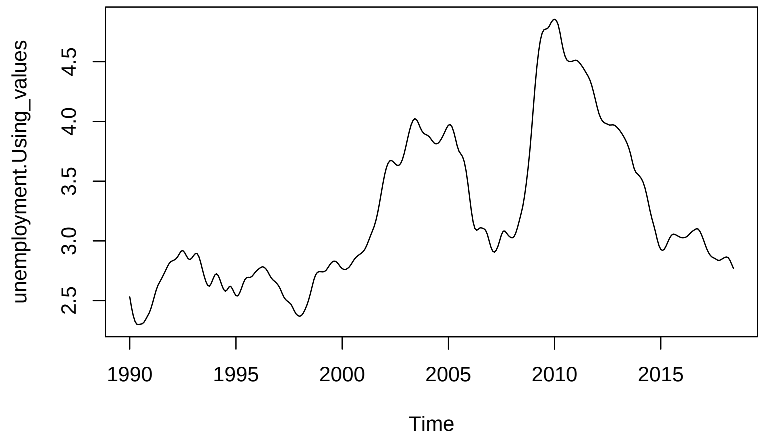
- After looking at the data, we can see that our data is cleaned

```
plot(decompose(unemployment))
```



- Summary of the Data by using decompose function, we can now see and know the data better. It seems that we can see seasonal but not very clear to us. It shows us also that we might need to consider to use the data after 1990 since that there a big diffreance between the data before 1990 and after.

```
unemployment.Using_values <-  window(unemployment,start=c(1990,1),end = c(2018,6))

unemployment.real_values <- window(unemployment,start=c(2018,7))
plot(unemployment.Using_values)
```



- I decide to use the data until the start of Jan of 2019, and will try to see if we can forcast the three months of 2019 to see if our model is good. We see that we need to try few transformation methods to this data.

unemployment: transformation,deterministic trend model, and seasonal model

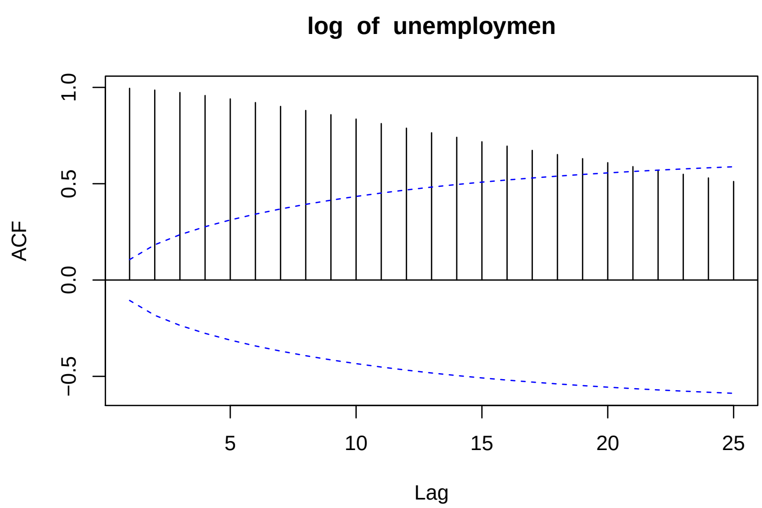
```
# Seasonal Differencing
nsdiffs(unemployment.Using_values)  # number for seasonal differencing needed
```

```
## [1] 0
```

- Then we do not need to have any Differencing
- But taking the log it will help and better and best transforation to this data

```
unemploymen.transformed <- log(unemployment.Using_values)
```

```
acf(as.vector(unemploymen.transformed),ci.type='ma',
   main='log  of  unemploymen')
```



- the roots are real and that one root is very close to the stationarity boundary (+1). This explains the slow decay of the autocorrelation function. We might now cosider our models.

- ARIMA(1, 1, 0) looks a good fit for this data but we might need to use auto.arima to see if that model is good as well ##unemployment: model specifcation AND parameter estimation

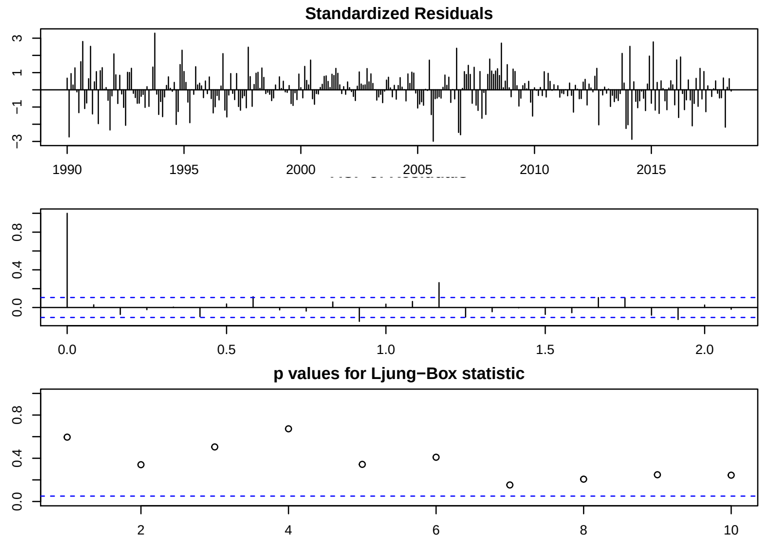
```
#choosign a model using auto.arima
auto.arima(unemploymen.transformed)
```

```
## Series: unemploymen.transformed
## ARIMA(5,1,2)(2,0,2)[12] with drift
##
## Coefficients:
##          ar1      ar2     ar3      ar4    ar5      ma1     ma2    sar1
##       2.9015  -4.2846  4.1098  -2.5036  0.758  -0.4943  0.8516  0.7891
## s.e.  0.0533   0.1314  0.1514   0.1047  0.038   0.0623  0.0490  0.2546
##          sar2     sma1    sma2    drift
##       -0.2970  -1.4050  0.5629  -0.0003
## s.e.   0.1082   0.2489  0.2322   0.0017
##
## sigma^2 estimated as 1.946e-06:  log likelihood=1755.56
## AIC=-3485.12   AICc=-3484.01   BIC=-3435.31
```

```
###Models that I tried:



#ARIMA(1, 1, 0) model and a modified version that adjusts
#unemployment_model=arima(log(unemployment.Using_values),order=c(1,1,0))

#auto.arima model.
unemployment_model <-arima(log(unemployment.Using_values),order = c(5,1,2),seasonal = list(order=c(2,0,

#we use an ARIMA (2,0,1) model for the unemployment rate.
#unemployment_model=arima(log(unemployment.Using_values),order=c(2,0,1))
```

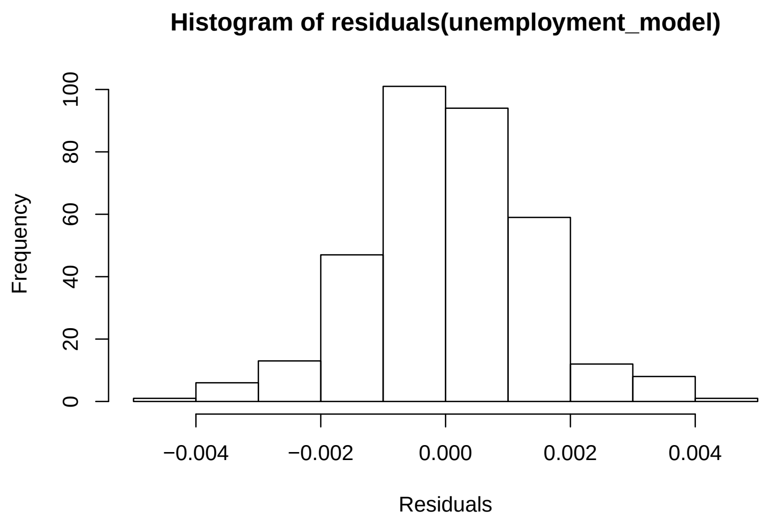
- Then now will try ARIMA(5,1,2)(2,0,1)[12] from auto.arima to see if it is good or no.

```
##cheaking the model
par(mar = rep(2, 4))

library(grDevices)
tsdiag(unemployment_model)
```

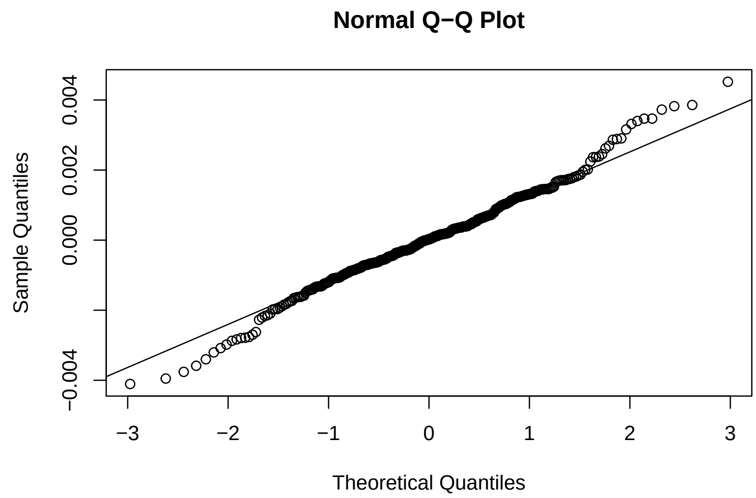


There are also around 5 residual acf values outside the critical limits. This Model diagnostics is telling us that the Ljung-Box statistics are better as shown in the bottom display where the bottom display shows the p-values of the Ljung-Box test for a variety of values of the "K" parameter—the highest lag used in the sum. The top display will flag potential outliers, if any, using the Bonferroni criteria as well.

```
#normality
hist(residuals(unemployment_model),xlab = 'Residuals')
```




```
qqnorm(residuals(unemployment_model),plot.it = TRUE)
qqline(residuals(unemployment_model))
```



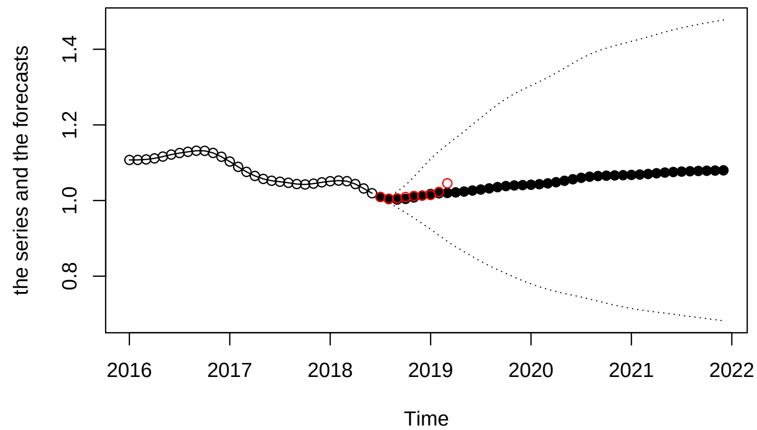

```r
shapiro.test(residuals(unemployment_model))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(unemployment_model)
## W = 0.98898, p-value = 0.01101
```

- Both the Q-Q plot and the results of the Shapiro-Wilk test indicate that we should reject of normality for the error terms in this model but this could be caused by the suspected outliers in the series. So, we will try to go with this model Knowing that this model does not pass the Shapiro-Wilk normality test but it looks good for the other tests.

labor: forecasting

```r
# for th next 3 years = 36 months
library(TSA)
outputs_plot = TSA:::plot.Arima(unemployment_model,n1=2016,n.ahead=42,ylab='the series and the forecasts
points(log(unemployment.real_values),col="red")
```



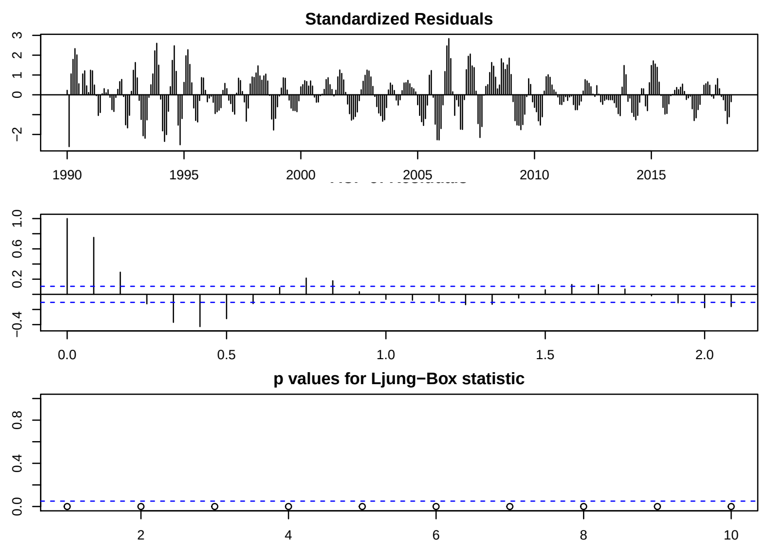
- The forecat limits does contain the actual values which means that this model is very bad to use.

unemployment Data: Model Number 2 after trying many model

- ARIMA(1, 1, 0)

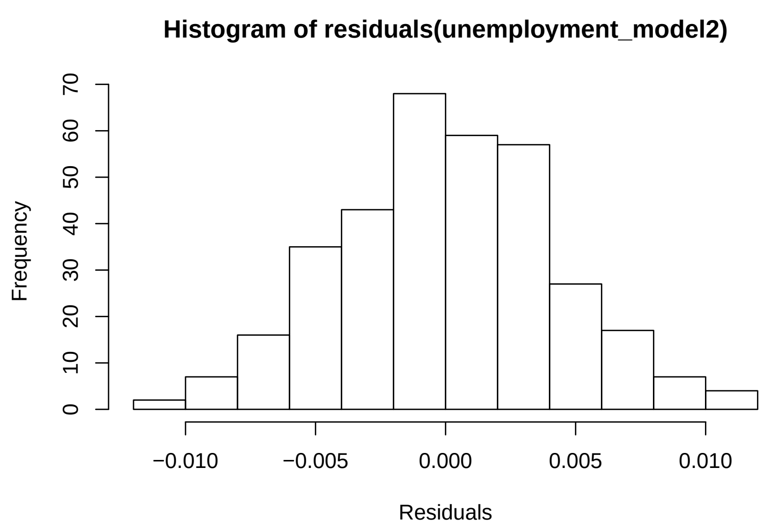
```r
#ARIMA(1, 1, 0) model and a modified version that adjusts
unemployment_model2=arima(log(unemployment.Using_values),order=c(1,1,0))
```

```r
##cheaking the model
par(mar = rep(2, 4))

library(grDevices)
tsdiag(unemployment_model2)
```

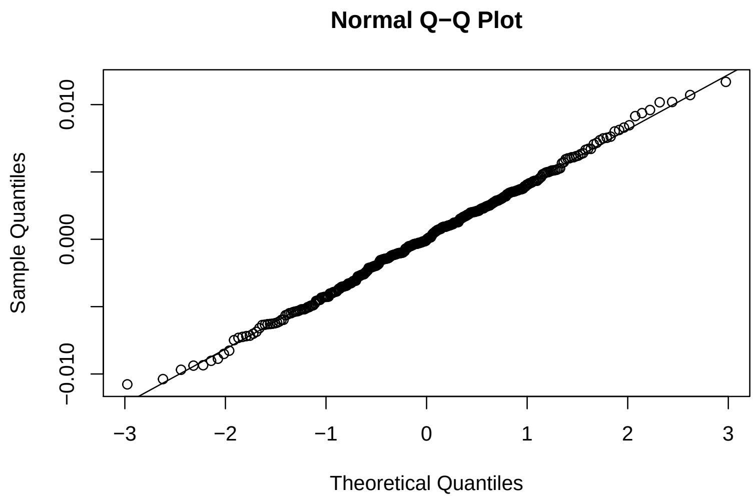


- These diagnostic plots do not show any inadequacies with the model. More evidence the error terms are independent. the residuals have statistically significant positive autocorrelation at first few lags, There are also around 8 residual acf values outside the critical limits.

```r
#normality
hist(residuals(unemployment_model2),xlab = 'Residuals')
```




```r
qqnorm(residuals(unemployment_model2),plot.it = TRUE)
qqline(residuals(unemployment_model2))
```



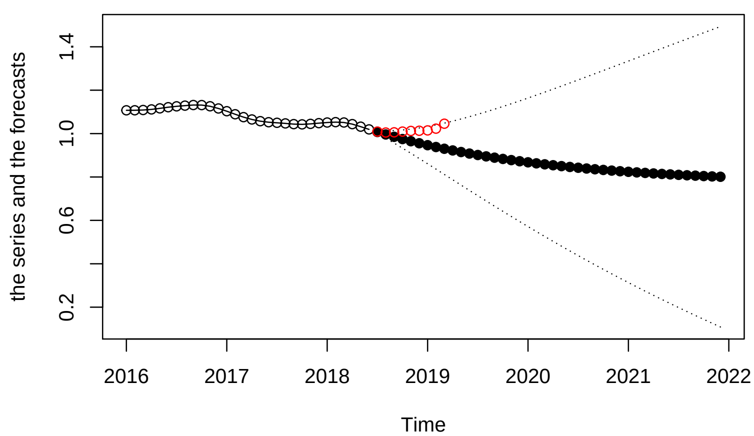

```
shapiro.test(residuals(unemployment_model2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(unemployment_model2)
## W = 0.99767, p-value = 0.9167
```

- Both the Q-Q plot and the results of the Shapiro-Wilk test indicate that we should accept of normality for the error terms in this model. It passes all the normality test. This tells us that this model might be best fit for our data coparing to the old model.

unemployment: forecasting

```
# for th next 3 years = 36 months
library(TSA)
outputs_plot = TSA:::plot.Arima(unemployment_model2,n1=2016,n.ahead=42,ylab='the series and the forecas
points(log(unemployment.real_values),col="red")
```



- The forecat limits contain the actual values which means that this model is very good to use.as well. But this model is passing the Shapiro-Wilk, but it seems the old model is better at forcasting and working more accurate than this model