STAT 621 HW 5

Yingying Xu

3/21/2018 (Due)

3.28 Verify IMA(1,1) model

$x_t = x_{t-1} + w_t - \lambda w_{t-1}$ as $|\lambda| < 1$.

Defining $y_t = w_t - \lambda w_{t-1} = x_t - x_{t-1}$, because $|\lambda| < 1$, then it is invertible.

$y_t = w_t - \lambda w_{t-1} = \sum_{j=1}^{\infty} \lambda^j y_{t-j} + w_t$

Substituting $y_t = x_t - x_{t-1}$,

then $y_t = x_t - x_{t-1} = \sum_{j=1}^{\infty} \lambda^j y_{t-j} + w_t = \sum_{j=1}^{\infty} \lambda^j (x_{t-j} - x_{t-j-1}) + w_t$

$x_t = x_{t-1} + \sum_{j=1}^{\infty} \lambda^j (x_{t-j} - x_{t-j-1}) + w_t = x_{t-1} + \lambda(x_{t-1} - x_{t-2}) + \lambda^2(x_{t-2} - x_{t-3}) ... + w_t = \sum_{j=1}^{\infty}(1 - \lambda)\lambda^{j-1} x_{t-j} + w_t$

3.30

```
set.seed(666)
y=varve[1:100]
x=log(y)
(x.ima1=HoltWinters(x,alpha=0.75,beta=FALSE,gamma=FALSE)) #lambda=0.25
```

```
## Holt-Winters exponential smoothing without trend and without seasonal component.
##
## Call:
## HoltWinters(x = x, alpha = 0.75, beta = FALSE, gamma = FALSE)
##
## Smoothing parameters:
##  alpha: 0.75
##  beta : FALSE
##  gamma: FALSE
##
## Coefficients:
##       [,1]
## a 3.066419
```

```
(x.ima2=HoltWinters(x,alpha=0.50,beta=FALSE,gamma=FALSE)) #lambda=0.50
```

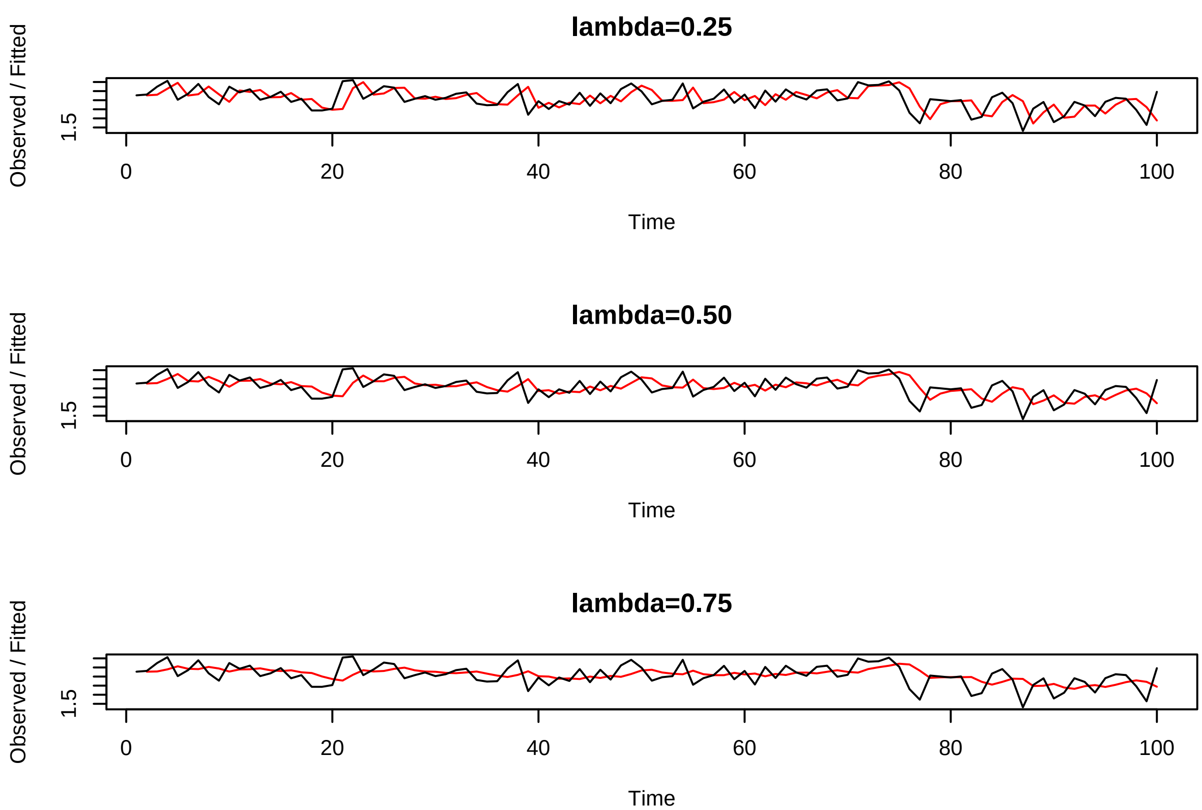
```
## Holt-Winters exponential smoothing without trend and without seasonal component.
##
## Call:
## HoltWinters(x = x, alpha = 0.5, beta = FALSE, gamma = FALSE)
##
## Smoothing parameters:
##  alpha: 0.5
##  beta : FALSE
##  gamma: FALSE
```

```
##
## Coefficients:
##       [,1]
## a 2.821237
```

```r
(x.ima3=HoltWinters(x,alpha=0.25,beta=FALSE,gamma=FALSE)) #lambda=0.75
```

```
## Holt-Winters exponential smoothing without trend and without seasonal component.
##
## Call:
## HoltWinters(x = x, alpha = 0.25, beta = FALSE, gamma = FALSE)
##
## Smoothing parameters:
##  alpha: 0.25
##  beta : FALSE
##  gamma: FALSE
##
## Coefficients:
##       [,1]
## a 2.696357
```

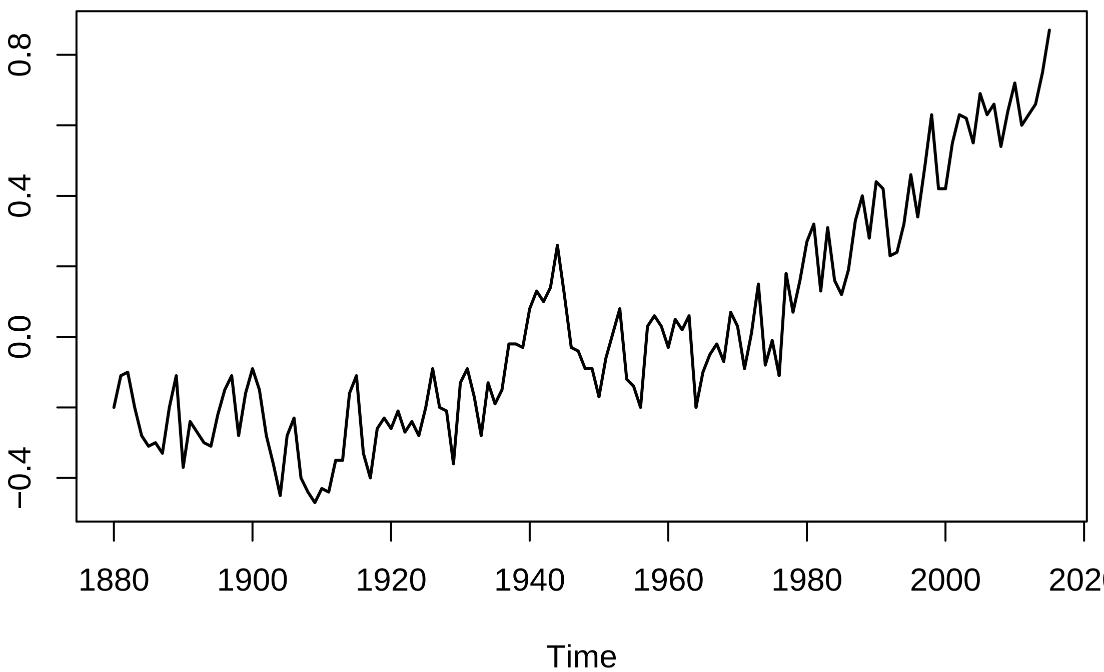
```r
par(mfrow=c(3,1))
plot(x.ima1,main="lambda=0.25")
plot(x.ima2,main="lambda=0.50")
plot(x.ima3,main="lambda=0.75")
```



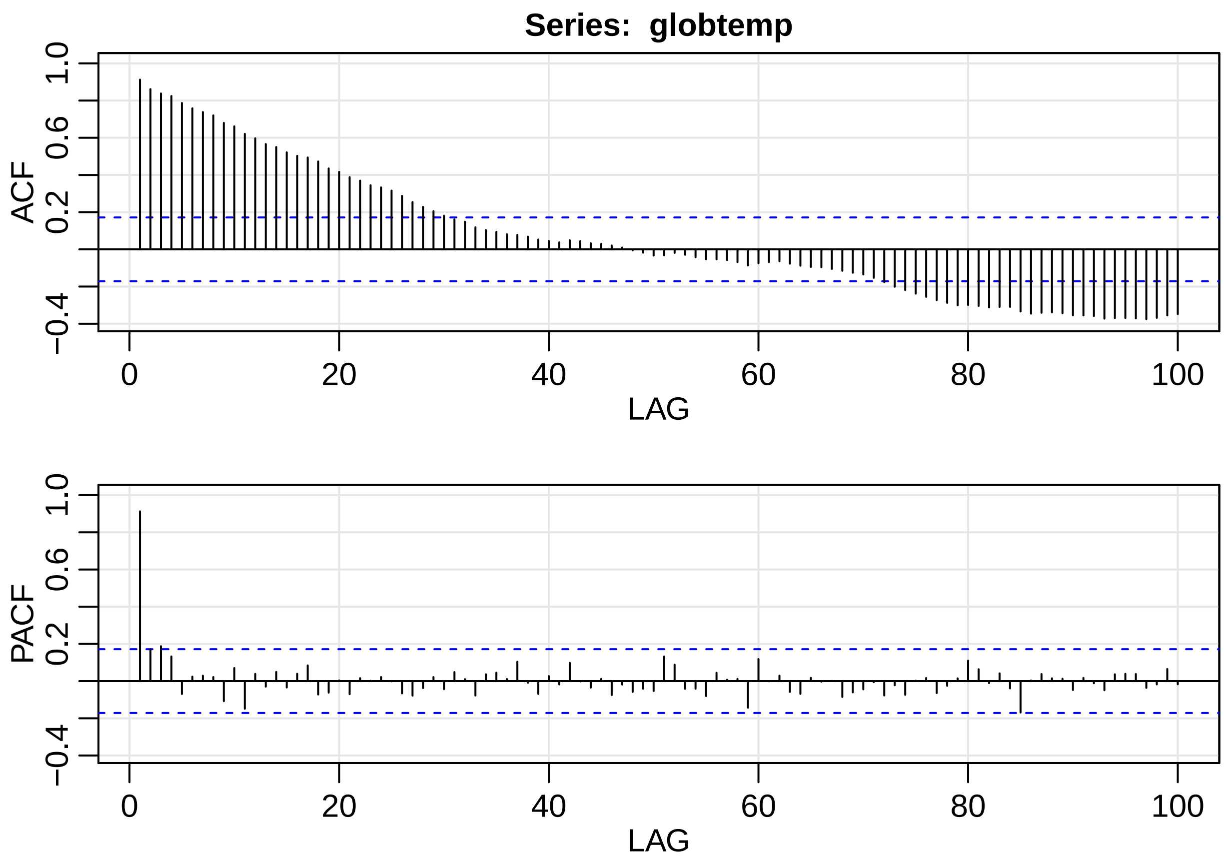
The blck lines are true value and red lines are forecasts with different lambda. As shown above, large values of λ lead to smoother forecasts.

3.33 Fit an ARIMA(p,d,q) model to globtemp

```r
plot.ts(globtemp,lwd=1.5,ylab='')
```

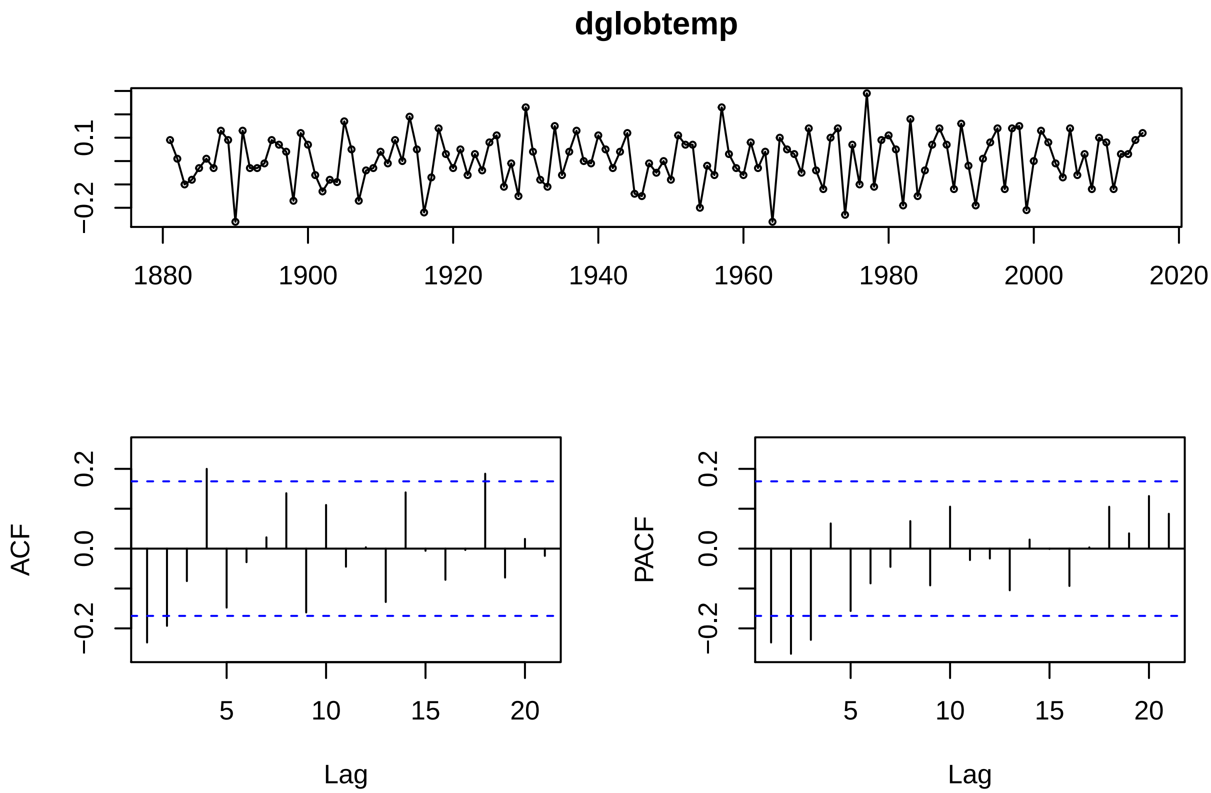



```r
invisible(astsa::acf2(globtemp,max.lag = 100))
```

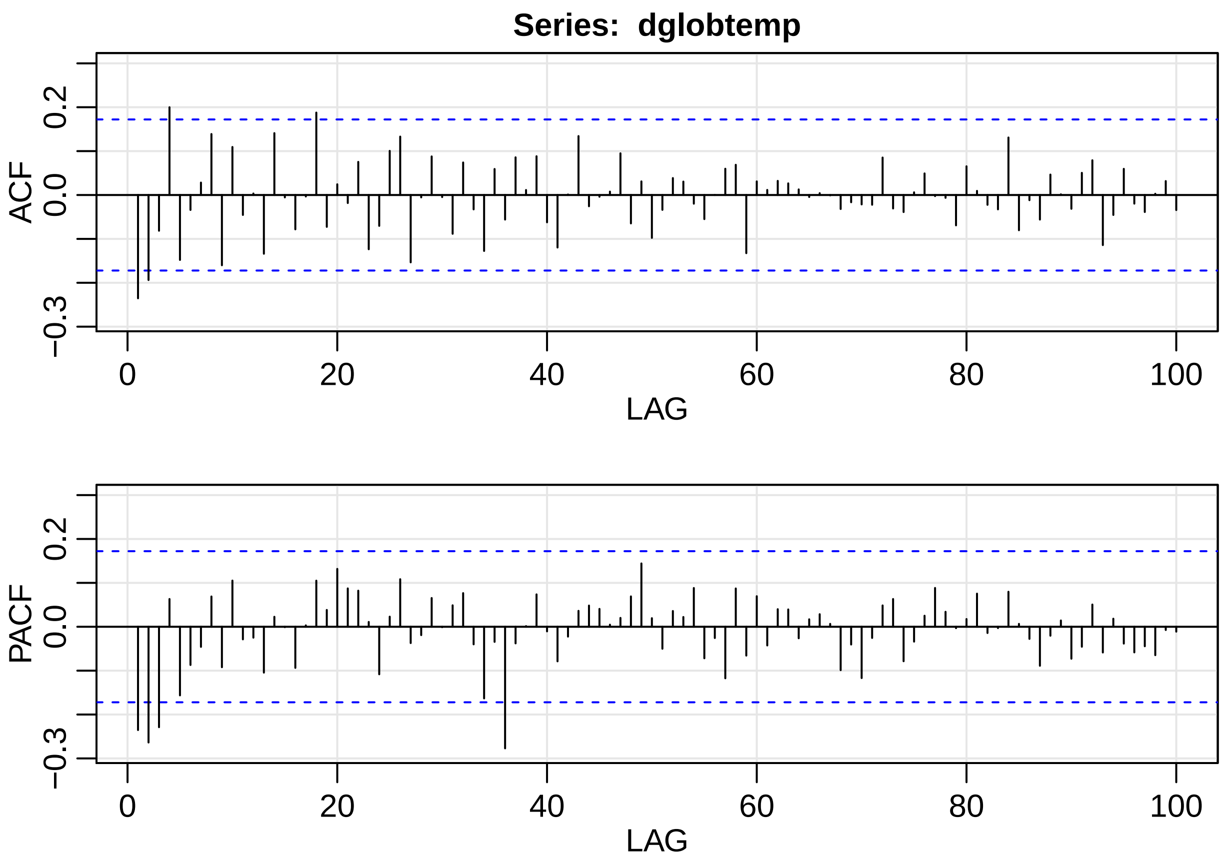


The ACF is linearly decaying and PACF has a peak in lag(1). May use first differencing.

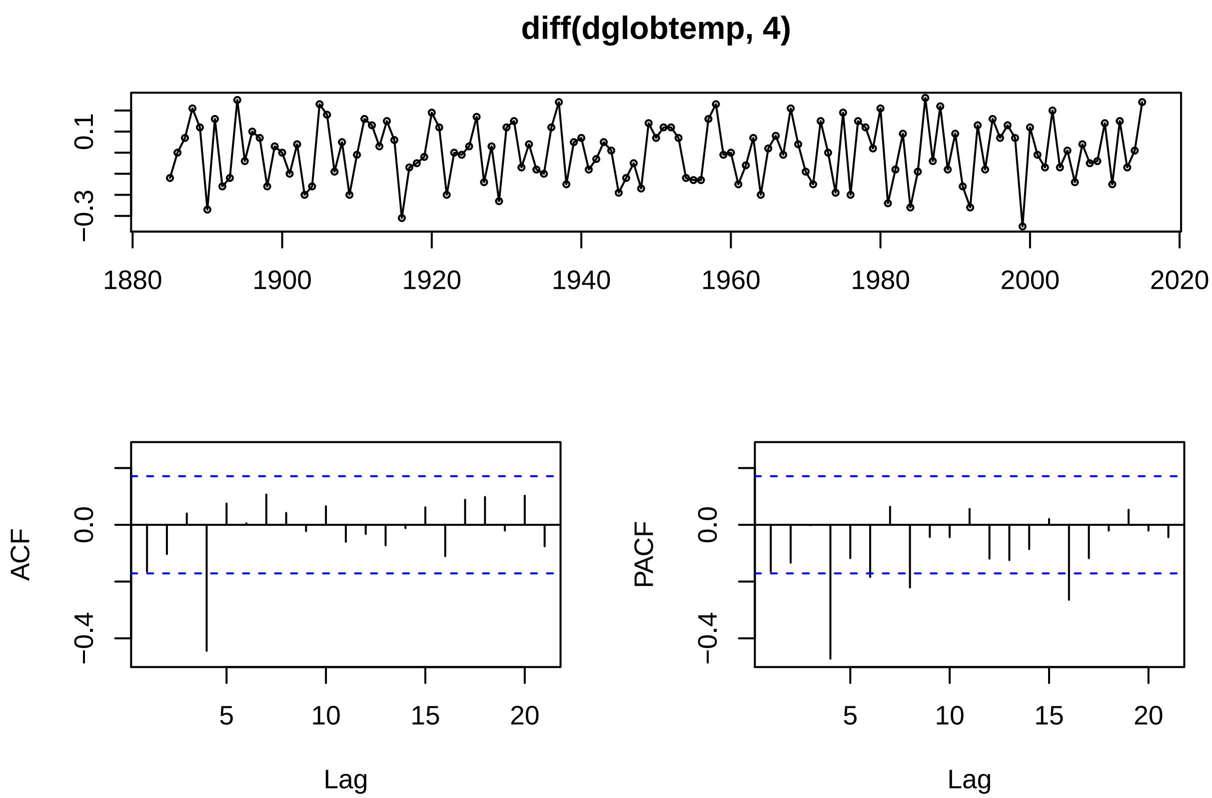
```
# Remove the trend by diff
dglobtemp=diff(globtemp)
tsdisplay(dglobtemp)
```




```
invisible(astsa::acf2(dglobtemp,max.lag = 100))
```




```
tsdisplay(diff(dglobtemp,4))
```



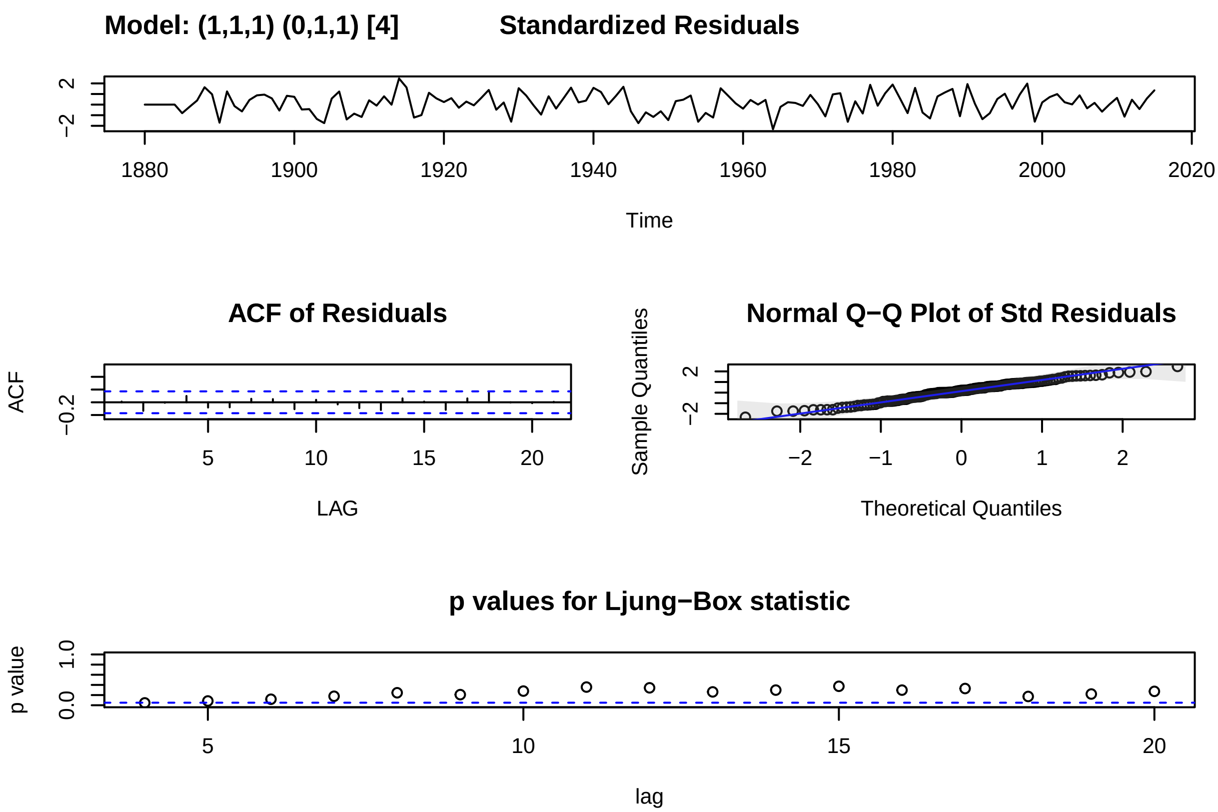

```
#invisible(astsa::acf2(diff(dglobtemp,4),max.lag = 100))
```

The PACF is exponentially deaying in seasonal pattern, the ACF has a peak in lag(4). Use SARIMA(0,1,1)[4].Try different combination in ARIMA model and examine the AICc, BIC, residuals, and p values. Finally I choose $SARIMA(1,1,1) \times (0,1,1)[4]$.

```
astsa::sarima(globtemp,p=1,d=1,q=1,P=0,D=1,Q=1,S=4,no.constant=T, details=T)
```

```
## initial  value -1.978691
## iter   2 value -2.135805
## iter   3 value -2.189593
## iter   4 value -2.201033
## iter   5 value -2.203928
## iter   6 value -2.207761
## iter   7 value -2.211609
## iter   8 value -2.213256
## iter   9 value -2.213474
## iter  10 value -2.213870
## iter  11 value -2.214121
## iter  12 value -2.214137
## iter  13 value -2.214138
## iter  13 value -2.214138
## iter  13 value -2.214138
## final  value -2.214138
## converged
## initial  value -2.233923
## iter   2 value -2.251358
```

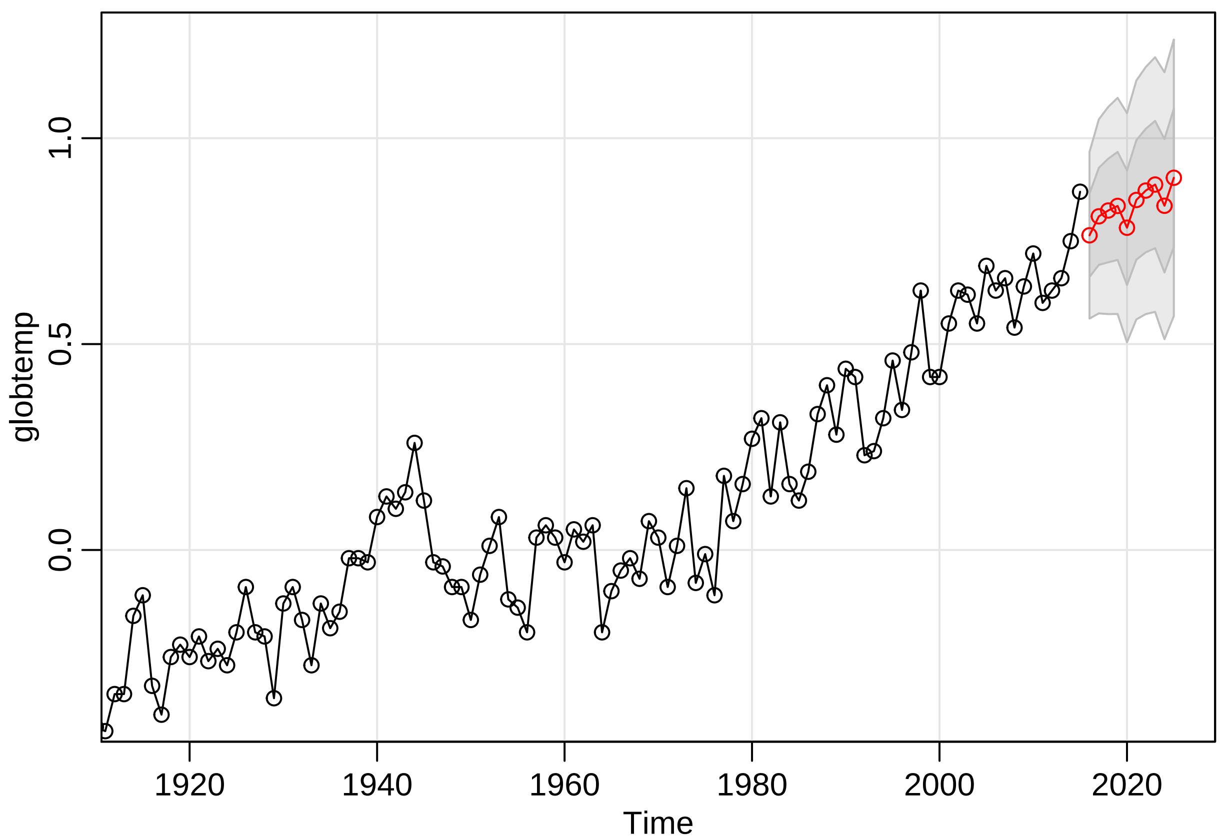
```
## iter   3 value -2.255218
## iter   4 value -2.260875
## iter   5 value -2.261839
## iter   6 value -2.262764
## iter   7 value -2.263213
## iter   8 value -2.263245
## iter   9 value -2.263246
## iter  10 value -2.263247
## iter  11 value -2.263247
## iter  12 value -2.263247
## iter  12 value -2.263247
## iter  12 value -2.263247
## final  value -2.263247
## 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, optim.control = list(trace = trc,
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ma1     sma1
##       0.4041  -0.8064  -0.8833
## s.e.  0.1339   0.0963   0.0857
```

```
##
## sigma^2 estimated as 0.01023:  log likelihood = 110.6,  aic = -213.21
##
## $degrees_of_freedom
## [1] 128
##
## $ttable
##      Estimate     SE  t.value p.value
## ar1    0.4041 0.1339   3.0171  0.0031
## ma1   -0.8064 0.0963  -8.3700  0.0000
## sma1  -0.8833 0.0857 -10.3034  0.0000
##
## $AIC
## [1] -3.538768
##
## $AICc
## [1] -3.521817
##
## $BIC
## [1] -4.474518
```

```
sarima.for(globtemp,p=1,d=1,q=1,P=0,D=1,Q=1,S=4,no.constant=T,n.ahead=10)
```



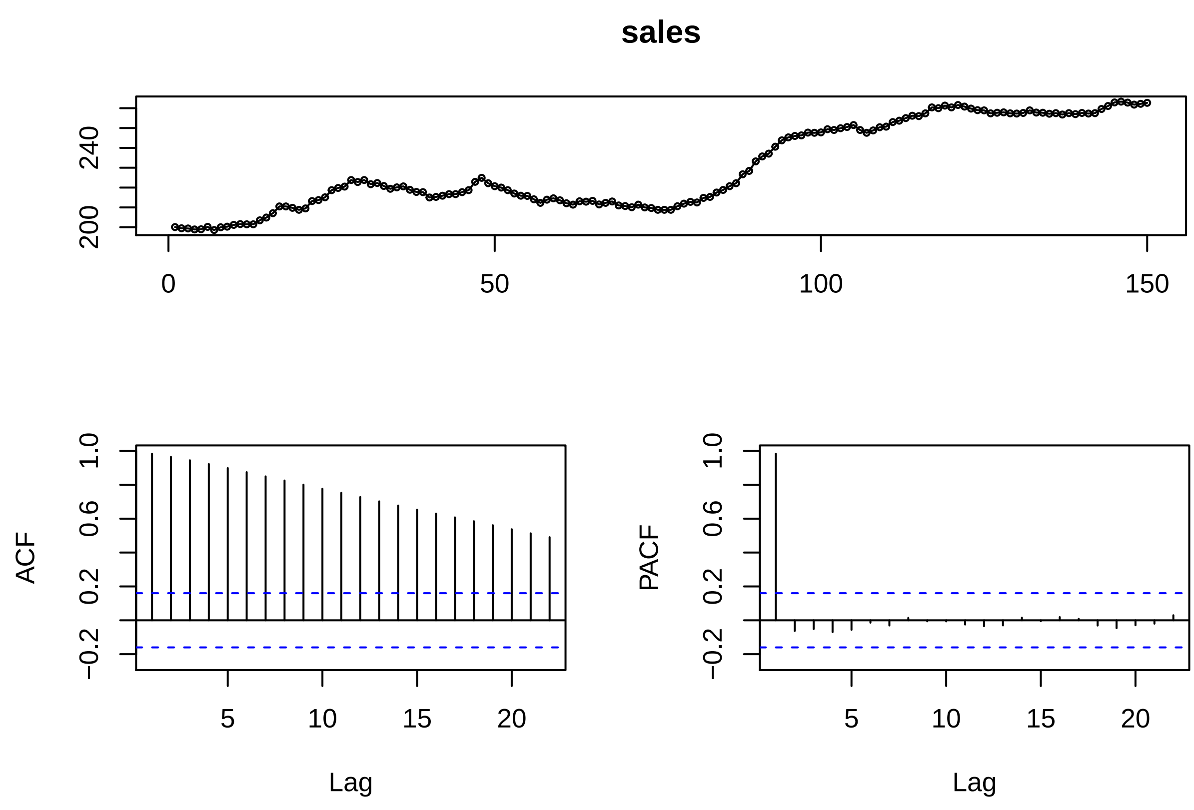

```
## $pred
## Time Series:
## Start = 2016
## End = 2025
```

```
## Frequency = 1
##  [1] 0.7641872 0.8102154 0.8244384 0.8355792 0.7826125 0.8499976 0.8728516
##  [8] 0.8874805 0.8359235 0.9038783
##
## $se
## Time Series:
## Start = 2016
## End = 2025
## Frequency = 1
##  [1] 0.1011238 0.1178127 0.1257638 0.1311971 0.1391895 0.1450041 0.1499752
##  [8] 0.1545539 0.1620642 0.1678932
```

3.35

A) Initial examination of the data

```r
tsdisplay(sales)
```

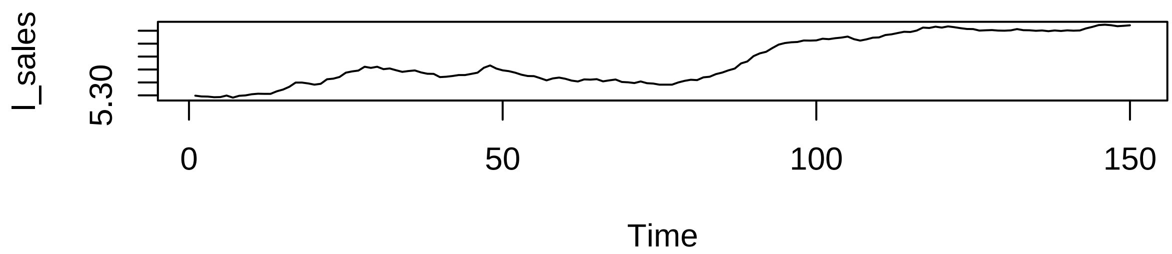


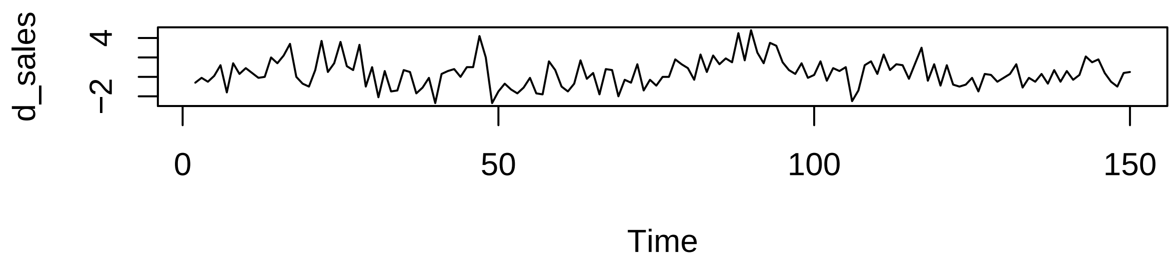

```r
#invisible(astsa::acf2(sales,max.lag = 100))
```

The ACF is linearly decaying. PACF has a peak at lag 1.

B) transformations

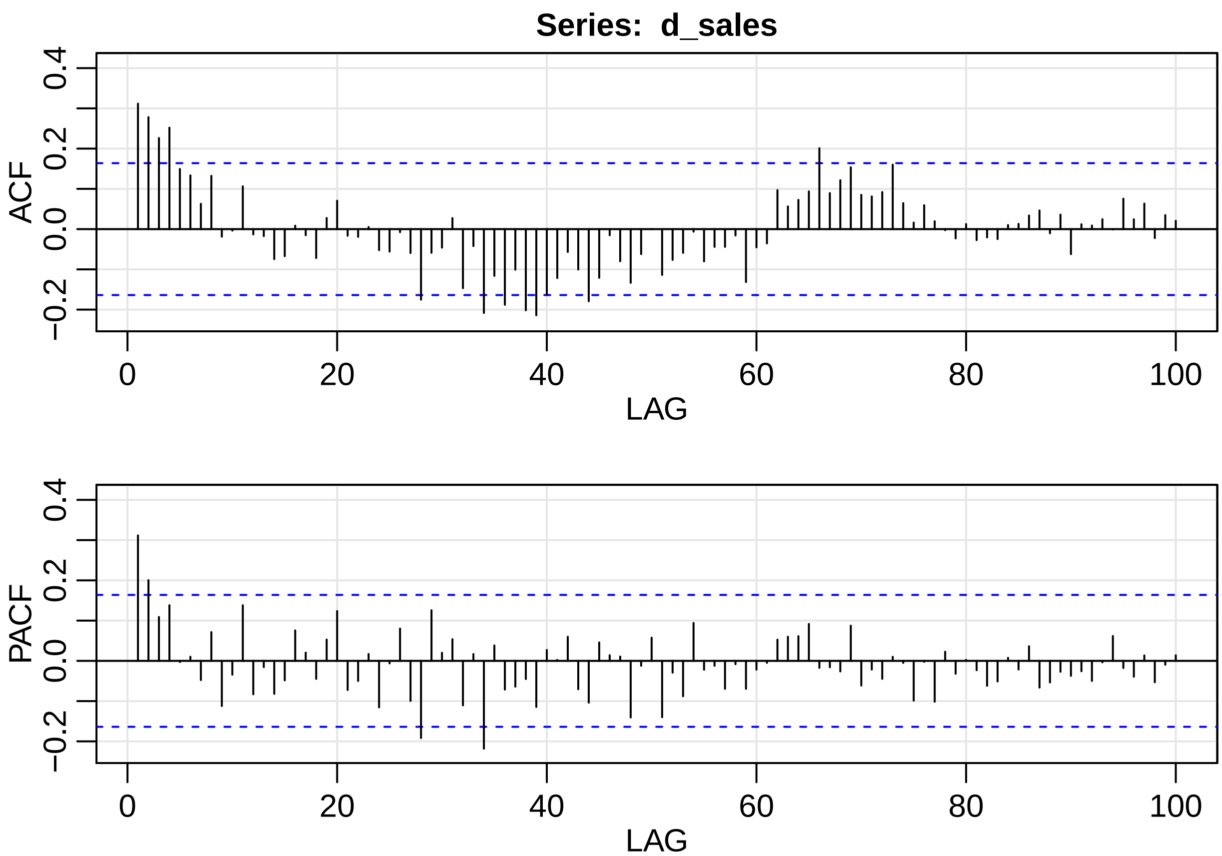
```
l_sales=log(sales)
d_sales=diff(sales)
par(mfrow=c(2,1))
ts.plot(l_sales)
ts.plot(d_sales)
```





Pick the first differencing.

```
#tsdisplay(d12_sales,max.lag = 100)
invisible(astsa::acf2(d_sales,max.lag = 100))
```



C) Initial identification of the dependence orders and differencing

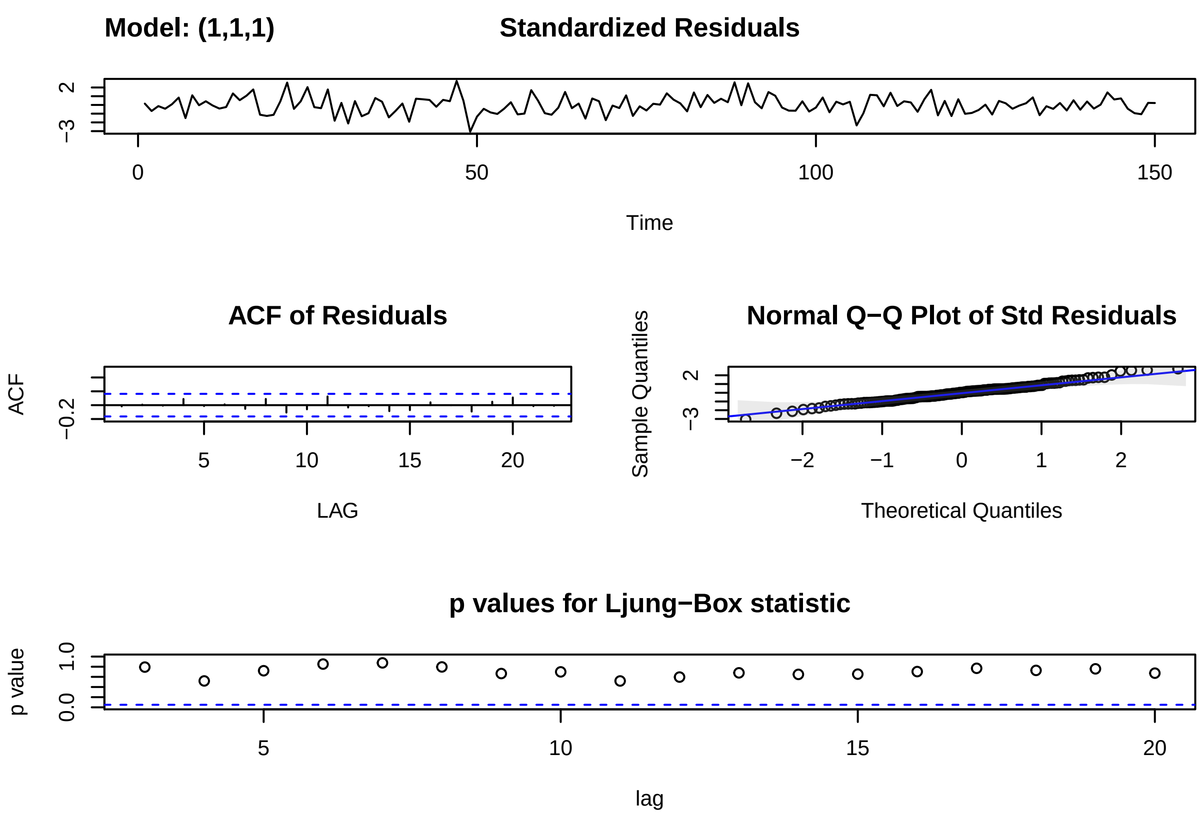
There is no need to do a logarithm transformations. We may start with $ARIMA(1, 1, 1)$.

D) Parameter estimation

```
sarima(sales,p=1,d=1,q=1,xreg=1:150,no.constant = T, details = T)
```

```
## initial  value 0.365727
## iter   2 value 0.335483
## iter   3 value 0.321346
## iter   4 value 0.319524
## iter   5 value 0.317869
## iter   6 value 0.308489
## iter   7 value 0.308211
## iter   8 value 0.303814
## iter   9 value 0.297719
## iter  10 value 0.291898
## iter  11 value 0.287205
## iter  12 value 0.282796
## iter  13 value 0.282585
## iter  14 value 0.282534
## iter  15 value 0.282517
## iter  16 value 0.282504
```

```
## iter  17 value 0.282503
## iter  18 value 0.282502
## iter  19 value 0.282502
## iter  20 value 0.282502
## iter  20 value 0.282502
## final  value 0.282502
## converged
## initial  value 0.281832
## iter   2 value 0.281791
## iter   3 value 0.281682
## iter   4 value 0.281680
## iter   5 value 0.281678
## iter   6 value 0.281678
## iter   7 value 0.281678
## iter   7 value 0.281678
## iter   7 value 0.281678
## final  value 0.281678
## converged
```




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

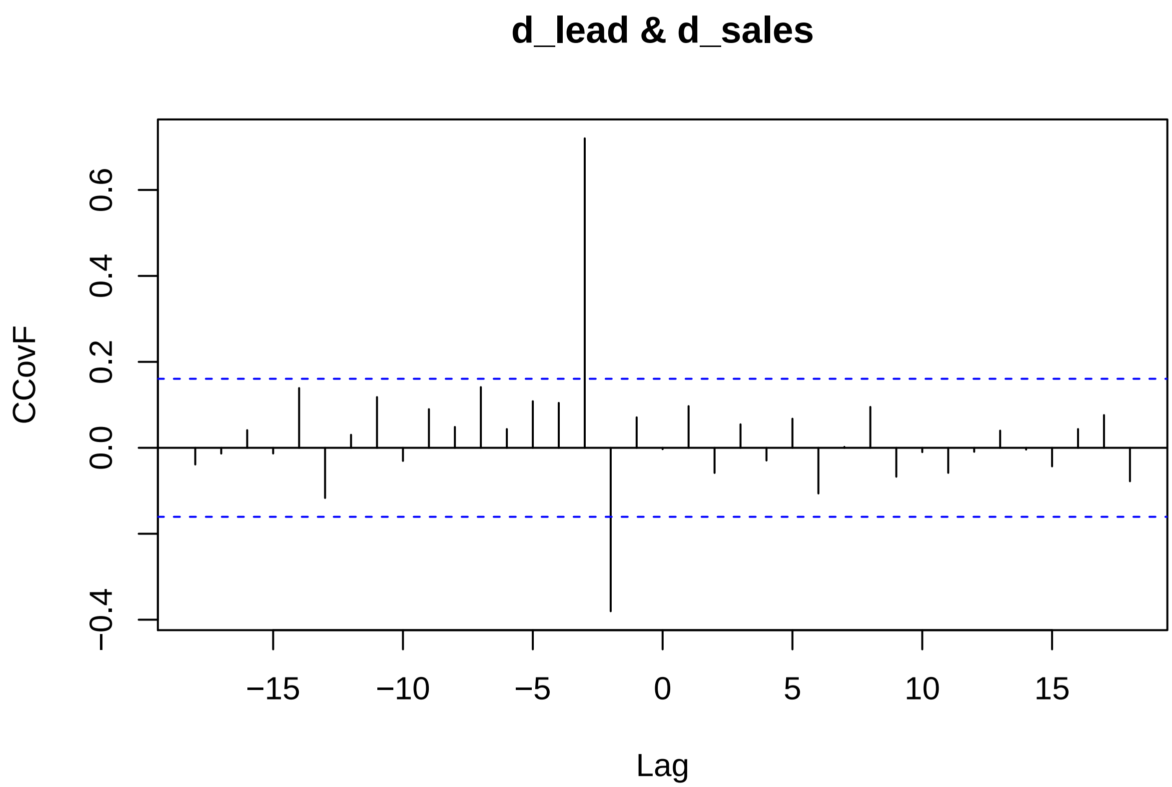
```
## Coefficients:
##          ar1      ma1    xreg
##       0.8381  -0.6097  0.4001
## s.e.  0.0834   0.1180  0.2557
##
## sigma^2 estimated as 1.754:  log likelihood = -253.39,  aic = 514.78
##
## $degrees_of_freedom
## [1] 146
##
## $ttable
##      Estimate     SE t.value p.value
## ar1    0.8381 0.0834 10.0528  0.0000
## ma1   -0.6097 0.1180 -5.1651  0.0000
## xreg   0.4001 0.2557  1.5644  0.1199
##
## $AIC
## [1] 1.601704
##
## $AICc
## [1] 1.616876
##
## $BIC
## [1] 0.6619163
```

E) Residual diagnostics and model choice

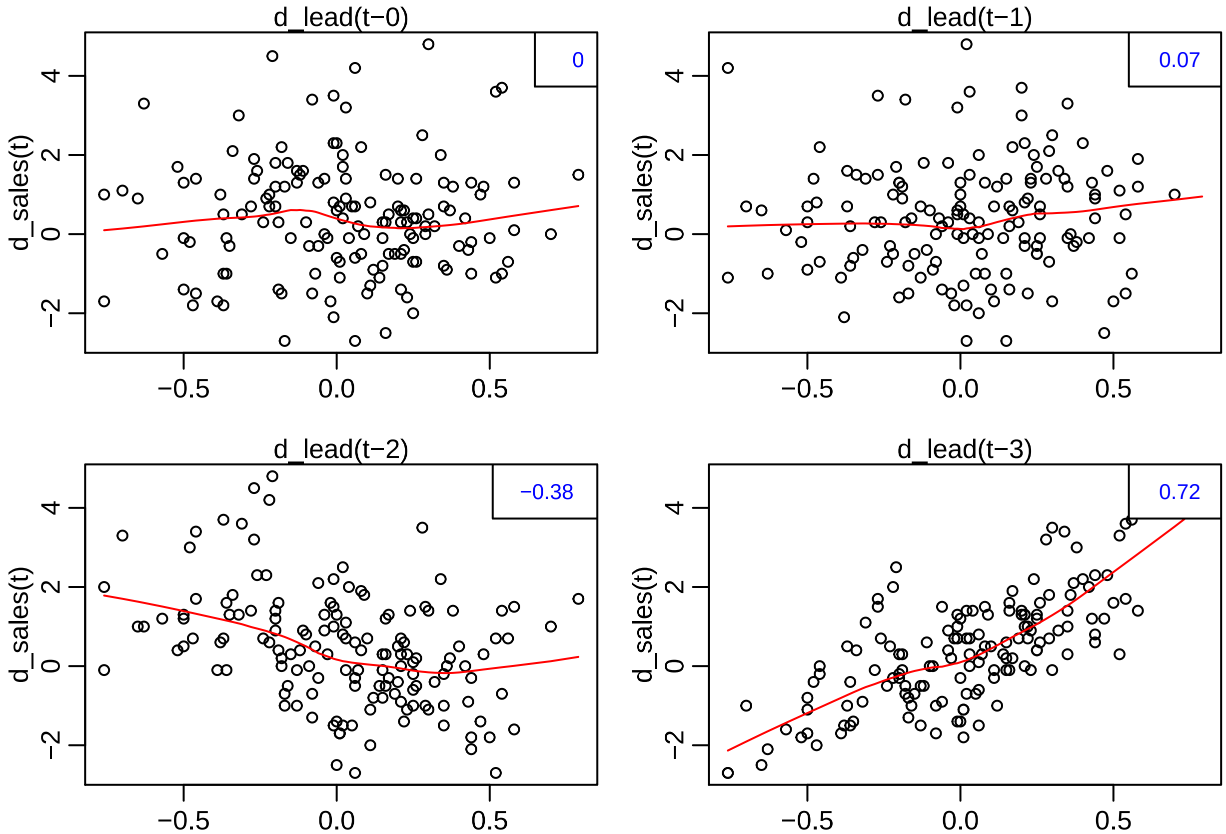
The ACF of Residuals are all in bounds and p_values are all out of the significant level.ARIMA(1,1,1) looks good.

b) Use the CCF and lag plots

```
d_sales=diff(sales)
d_lead=diff(lead)
ccf(d_lead,d_sales,ylab="CCovF", type ='correlation')
```




```
lag2.plot(d_lead,d_sales, max.lag = 3, corr = TRUE, smooth = TRUE)
```



A regression of ΔS_t on ΔL_{t-3} is reasonable. The correlation between ΔS_t on ΔL_{t-3} is 0.72, ΔS_t on ΔL_{t-2} is -0.38. Lag(3) is a good fit.

c) Fit the Regression Model

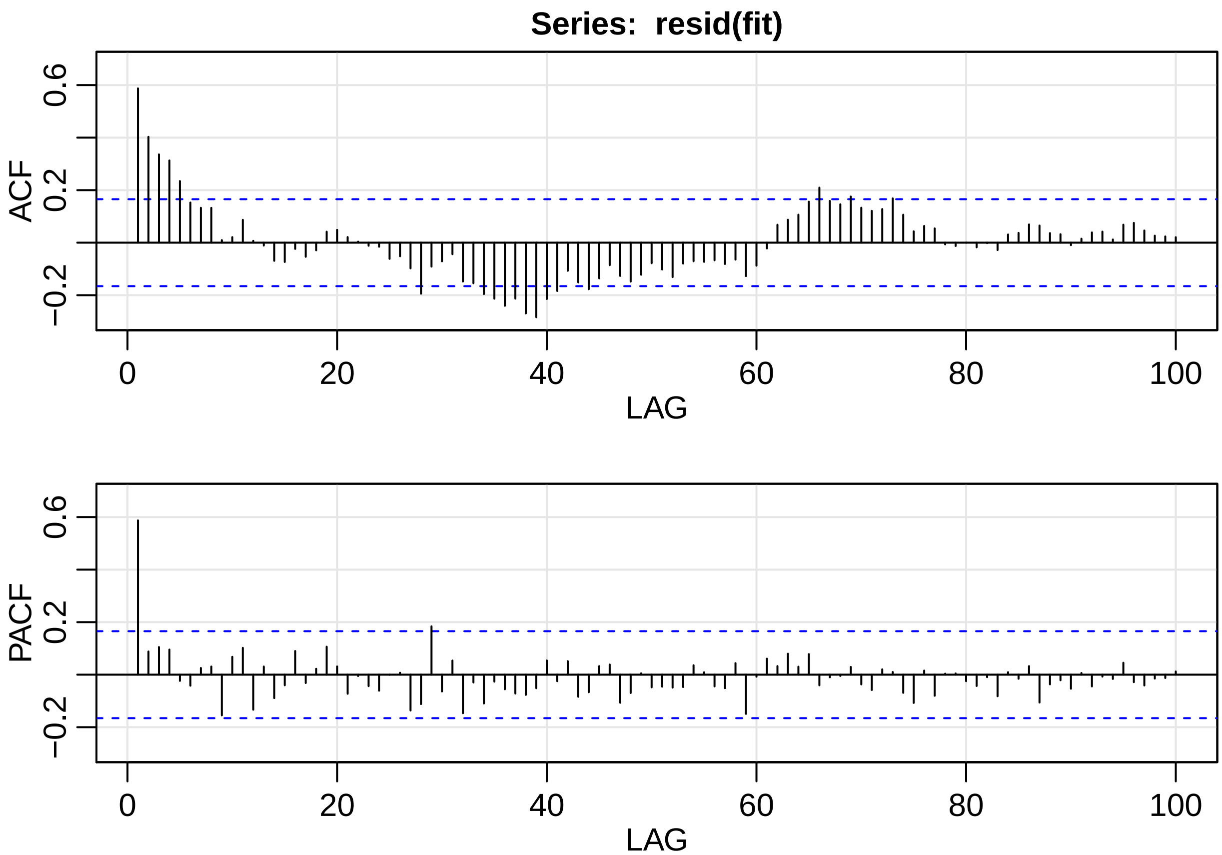
$\Delta S_t = \beta_0 + \beta_1 \Delta L_{t-3} + x_t$ where x_t is an ARMA process.

```
fit=lm(d_sales[4:149]~d_lead[1:146])
summary(fit)
```

```
##
## Call:
## lm(formula = d_sales[4:149] ~ d_lead[1:146])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.18876 -0.65502 -0.07291  0.60347  2.84546
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.35538    0.08301   4.281 3.38e-05 ***
## d_lead[1:146]  3.33733    0.26190  12.743  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1 on 144 degrees of freedom
## Multiple R-squared:   0.53,  Adjusted R-squared:  0.5267
```

```
## F-statistic: 162.4 on 1 and 144 DF,  p-value: < 2.2e-16
```

```
invisible(acf2(resid(fit),100))
```

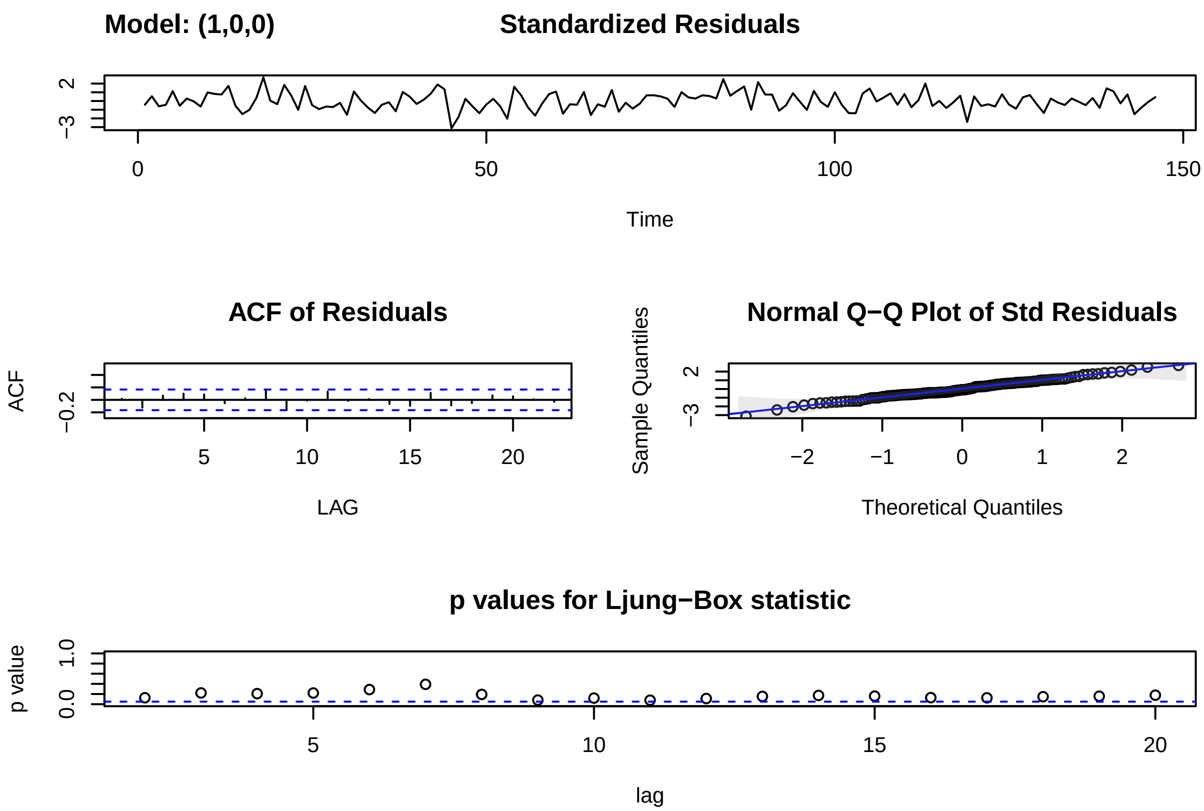


There is a lag(1) peak in PACF. Try AR(1) model.

```
adjfit=sarima(d_sales[4:149],1,0,0,xreg=d_lead[1:146])
```

```
## initial  value -0.004149
## iter   2 value -0.216173
## iter   3 value -0.243435
## iter   4 value -0.259915
## iter   5 value -0.262337
## iter   6 value -0.262396
## iter   7 value -0.262409
## iter   8 value -0.262410
## iter   9 value -0.262410
## iter  10 value -0.262410
## iter  11 value -0.262410
## iter  12 value -0.262410
## iter  12 value -0.262410
## iter  12 value -0.262410
## final  value -0.262410
## converged
## initial  value -0.263313
## iter   2 value -0.263328
## iter   3 value -0.263337
## iter   4 value -0.263341
```

```
## iter   5 value -0.263344
## iter   6 value -0.263345
## iter   6 value -0.263345
## iter   6 value -0.263345
## final  value -0.263345
## converged
```

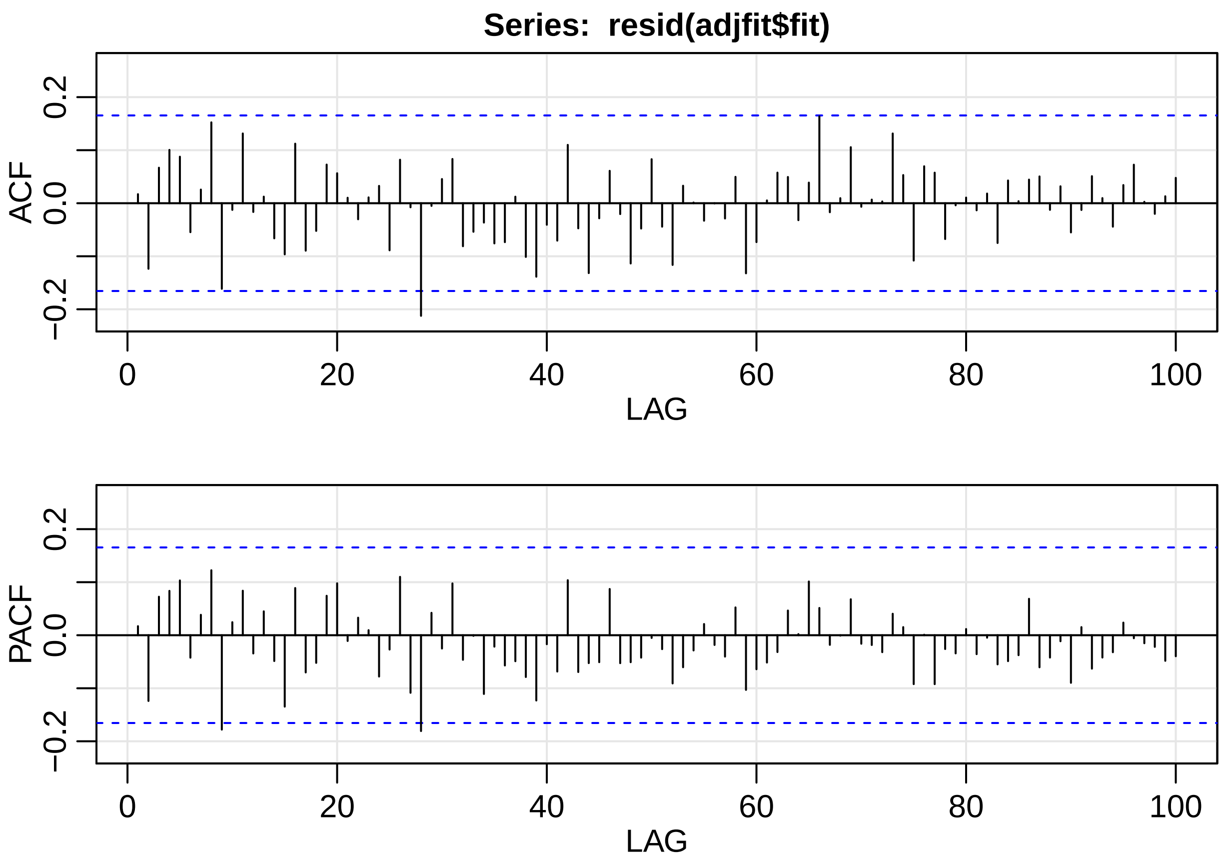



```
adjfit
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, optim.control = list(trace = trc, REPORT = 1,
##     reltol = tol))
##
## Coefficients:
##          ar1  intercept    xreg
##       0.6451     0.3624  2.7876
## s.e.  0.0628     0.1767  0.1432
##
## sigma^2 estimated as 0.5884:  log likelihood = -168.72,  aic = 345.43
##
## $degrees_of_freedom
## [1] 143
##
## $ttable
```

```
##            Estimate     SE t.value p.value
## ar1          0.6451 0.0628 10.2663  0.0000
## intercept    0.3624 0.1767  2.0509  0.0421
## xreg         2.7876 0.1432 19.4610  0.0000
##
## $AIC
## [1] 0.5107203
##
## $AICc
## [1] 0.526362
##
## $BIC
## [1] -0.4279727
```

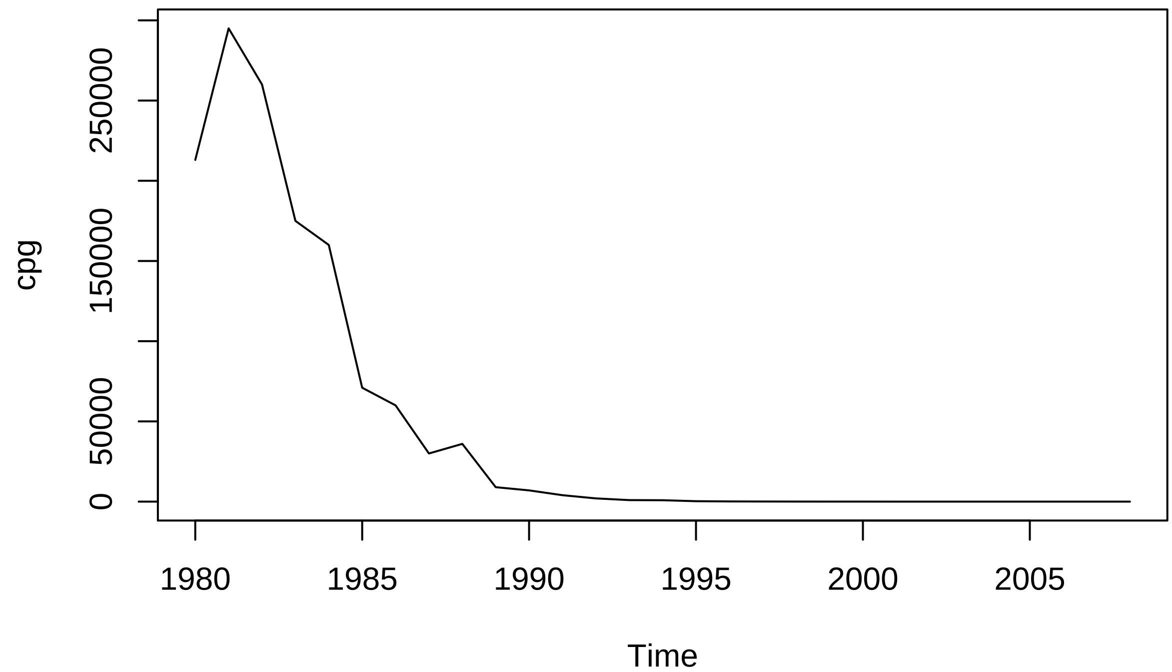
```
invisible(acf2(resid(adjfit$fit),max.lag = 100)) # show residual, it should be white noise
```



The ACF/PACF of residuals are following white noise distribution. The model is good.

3.36

```
#tsdisplay(cpg)
plot(cpg)
```

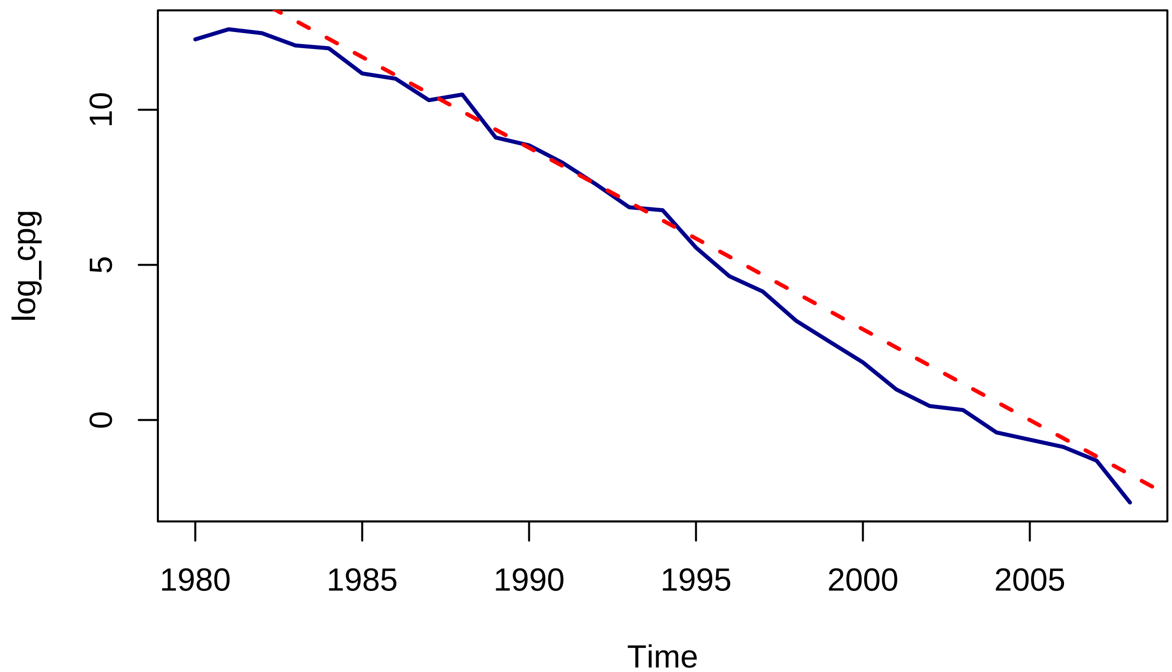


The retail price per GB reached a peak around 1982 and then start exponentially decay during 1982 to 1990. It becomes more and more constant after 1990, close to 0 after 1995.

```r
log_cpg=log(cpg)
t=(1:length(cpg))
t=t+1980
m=lm(log_cpg~t)
summary(m)
```

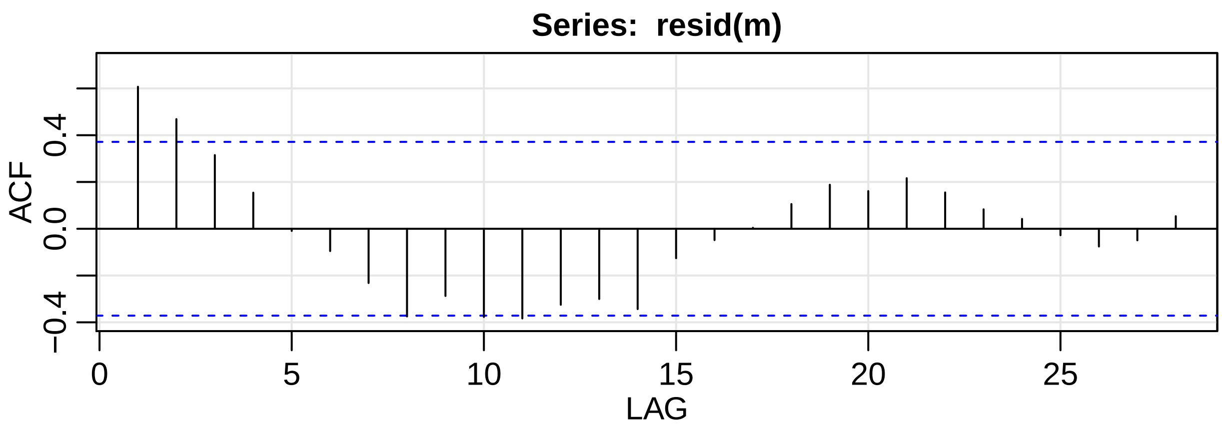
```
##
## Call:
## lm(formula = log_cpg ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77156 -0.39840  0.04726  0.42186  1.13129
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1173.07939   27.59176   42.52   <2e-16 ***
## t             -0.58508    0.01383  -42.30   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6231 on 27 degrees of freedom
## Multiple R-squared:  0.9851, Adjusted R-squared:  0.9846
## F-statistic:  1790 on 1 and 27 DF,  p-value: < 2.2e-16
```

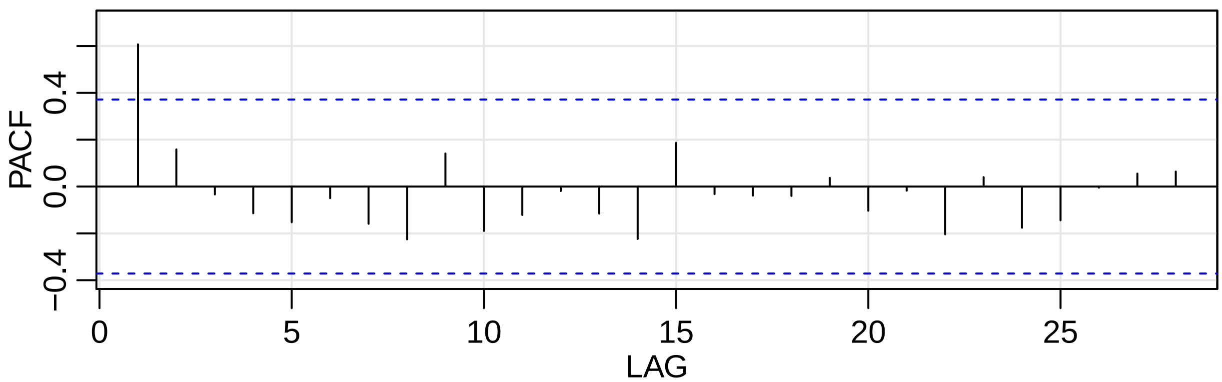
```r
plot(log_cpg,lwd=2, col="darkblue")
abline(a = coef(m)[1], b = coef(m)[2], lty=2, lwd=2, col="red")
```



The transformed log(cpg) fits the linear regression well. The coefficient -0.5851 with constant 1173.079 (since we need to start in 1980). Therefore, $log(cpg) = log(\alpha) + \beta t = 1173.079 - 0.5851$.

```r
invisible(acf2(resid(m),28))
```



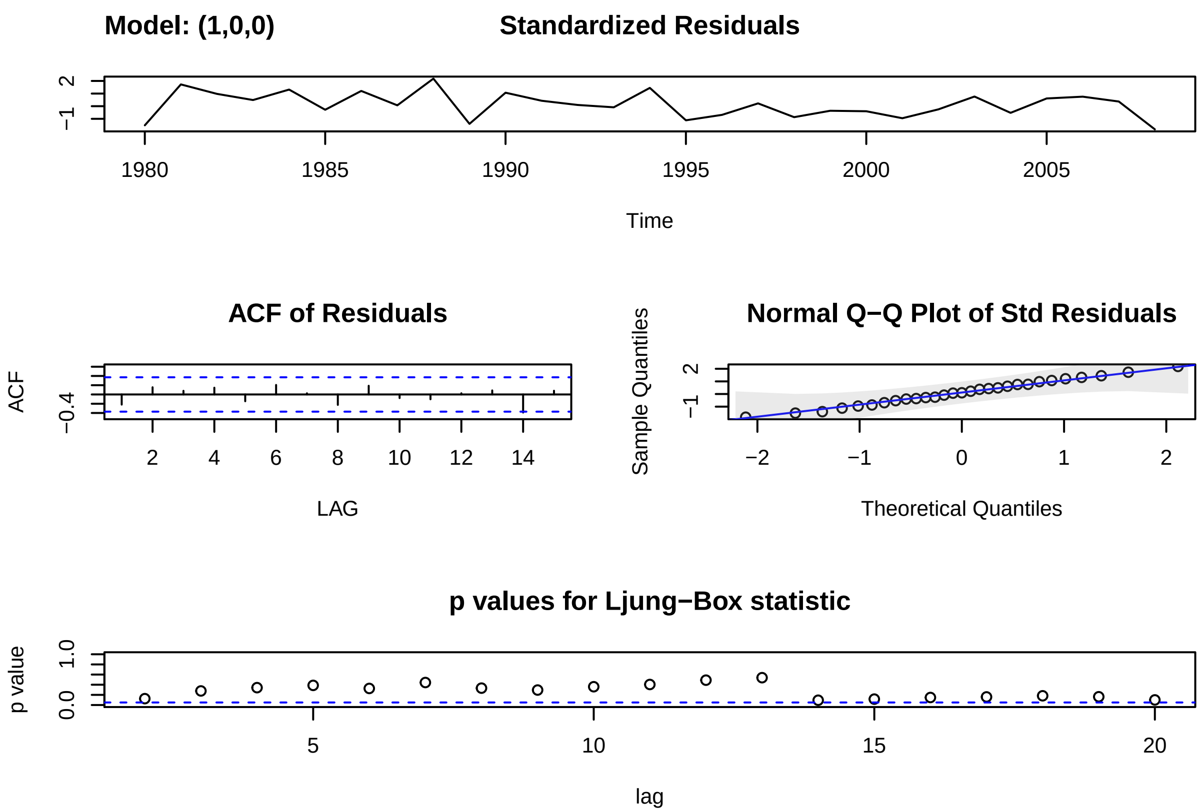


The R-squared is 0.985, t explains 98.5% of the total log(cpg). The resid(m) suggests it might be an AR(1) model.

```
sarima(log_cpg,1,0,0,xreg=t,details=T)
```

```
## initial  value -0.669056
## iter   2 value -0.999488
## iter   3 value -1.088763
## iter   4 value -1.102248
## iter   5 value -1.128914
## iter   6 value -1.131945
## iter   7 value -1.132479
## iter   8 value -1.132525
## iter   9 value -1.132540
## iter  10 value -1.132543
## iter  11 value -1.132545
## iter  12 value -1.132545
## iter  12 value -1.132545
## iter  12 value -1.132545
## final  value -1.132545
## converged
## initial  value -0.701381
## iter   2 value -0.882862
## iter   3 value -0.886699
## iter   4 value -0.888651
## iter   5 value -0.888966
## iter   6 value -0.889035
```

```
## iter   7 value -0.889043
## iter   8 value -0.889045
## iter   9 value -0.889045
## iter  10 value -0.889045
## iter  10 value -0.889045
## iter  10 value -0.889045
## final  value -0.889045
## converged
```




```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, optim.control = list(trace = trc, REPORT = 1,
##     reltol = tol))
##
## Coefficients:
##          ar1  intercept     xreg
##       0.8297  1113.5659  -0.5554
## s.e.  0.1190    73.6035   0.0368
##
## sigma^2 estimated as 0.1623:  log likelihood = -15.37,  aic = 38.73
##
## $degrees_of_freedom
## [1] 26
##
```

```
## $ttable
##           Estimate      SE  t.value p.value
## ar1         0.8297  0.1190   6.9740       0
## intercept 1113.5659 73.6035  15.1292       0
## xreg       -0.5554  0.0368 -15.0716       0
##
## $AIC
## [1] -0.6114079
##
## $AICc
## [1] -0.4849711
##
## $BIC
## [1] -1.469963
```

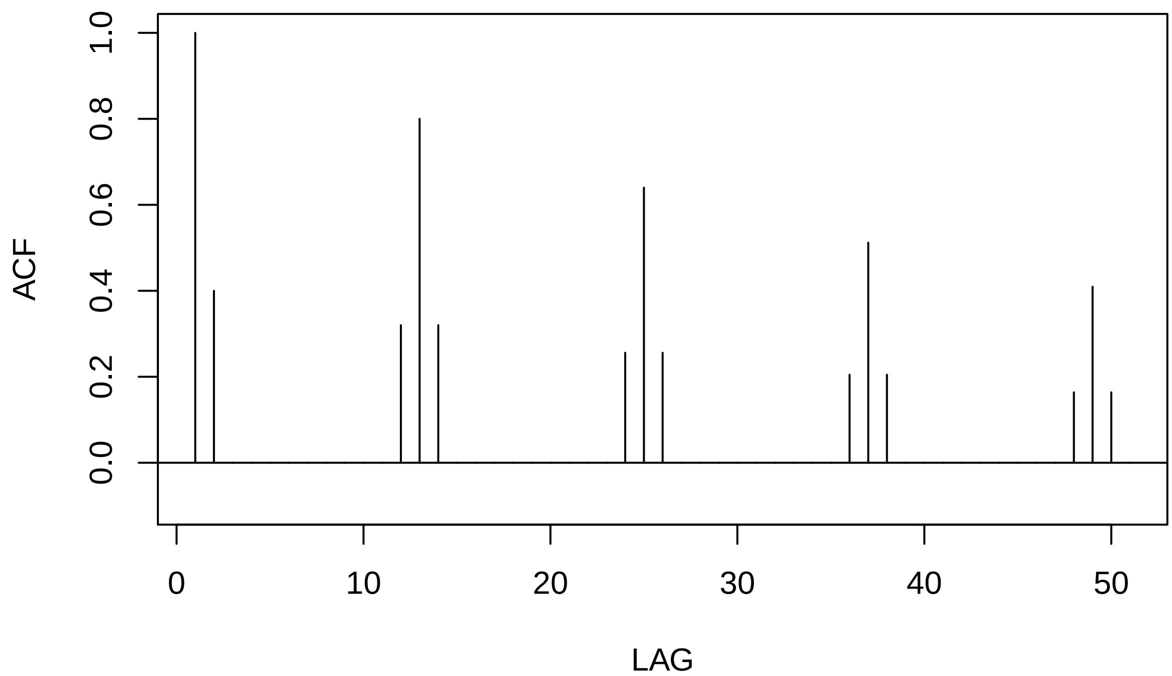
The residual diagnositics shows the residuals are normal and within the bounds. p values are out of significant levels. The model is OK. $x_t = 1113.5659 + 0.8297 * x_{t-1} - 0.5554 * t$.

3.39

Plot the ACF of the seasonal $ARIMA(0,1) \times (1,0)_{12}$ model with $\Phi = 0.8$ and $\theta = 0.5$

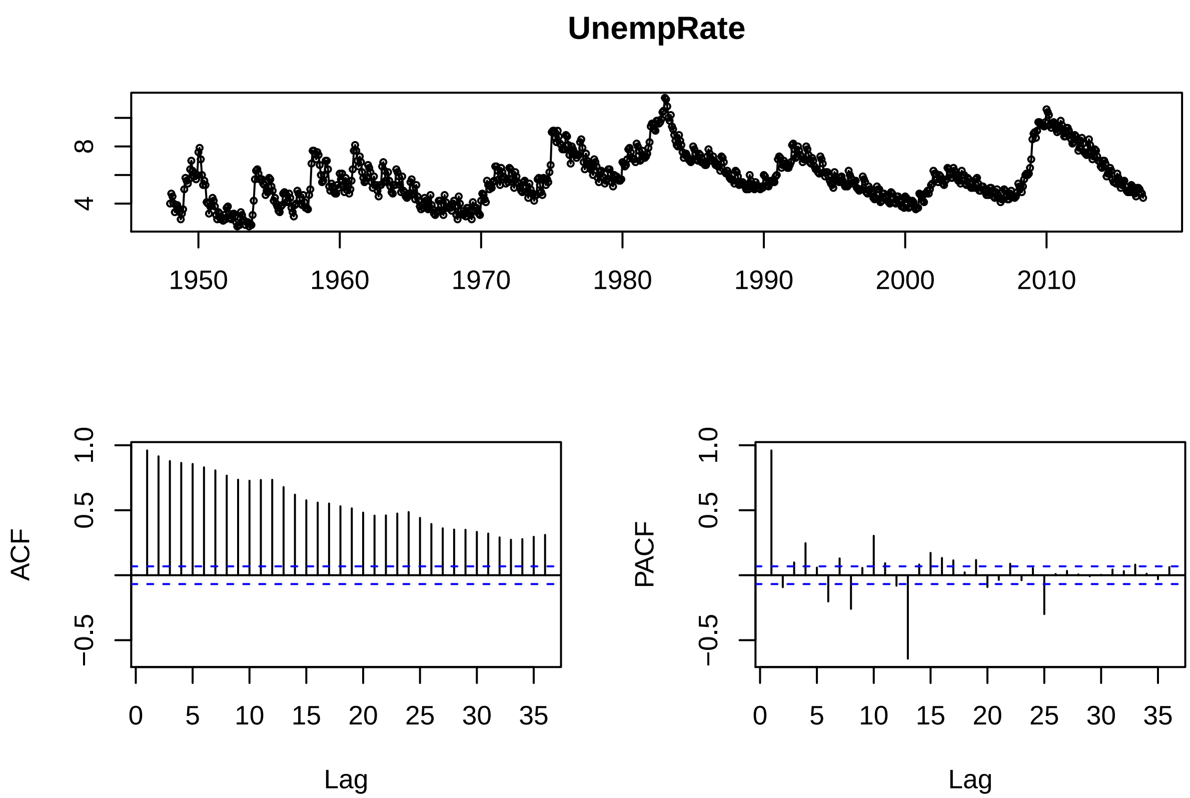
$$x_t = \Phi x_{t-12} + w_t + \theta w_{t-1}$$

```r
set.seed(666)
phi=c(rep(0,11),0.8)
ACF=ARMAacf(ar=phi,ma=0.5,50)
plot(ACF, type="h",xlab="LAG",ylim=c(-.1,1));abline(h=0)
```



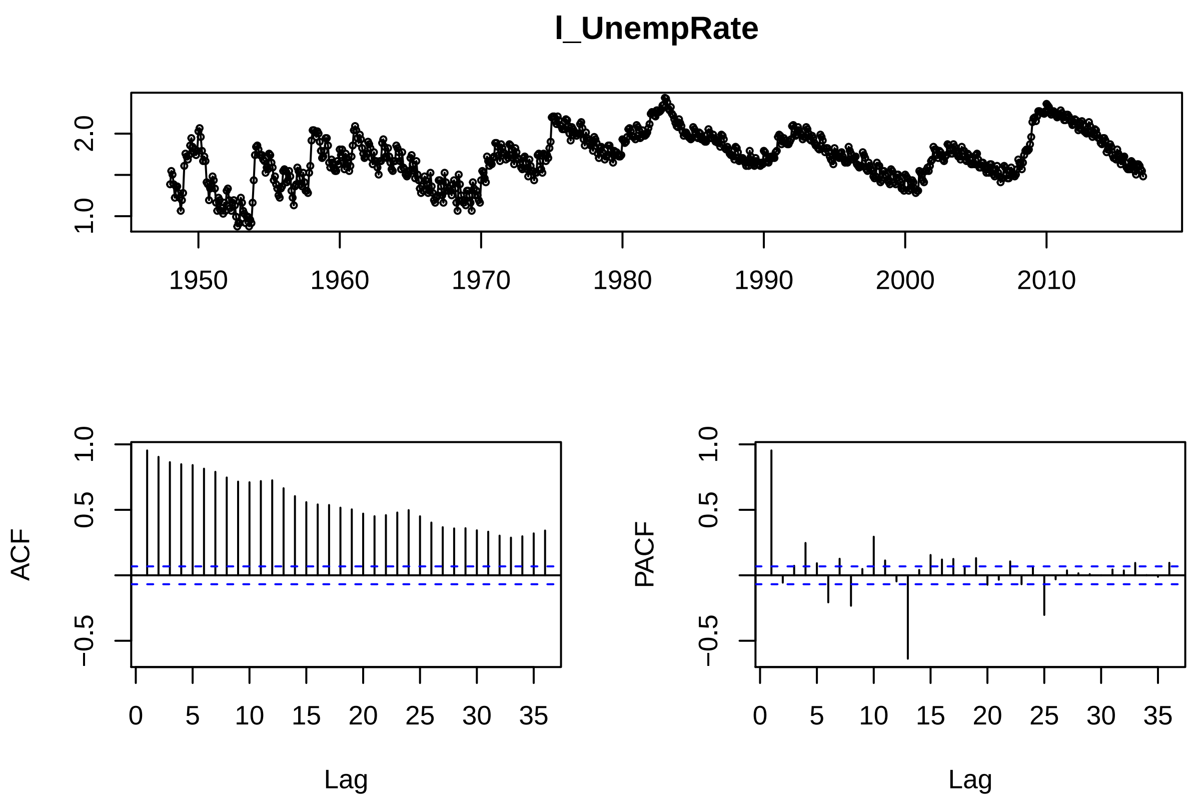
3.42

```
tsdisplay(UnempRate)
```

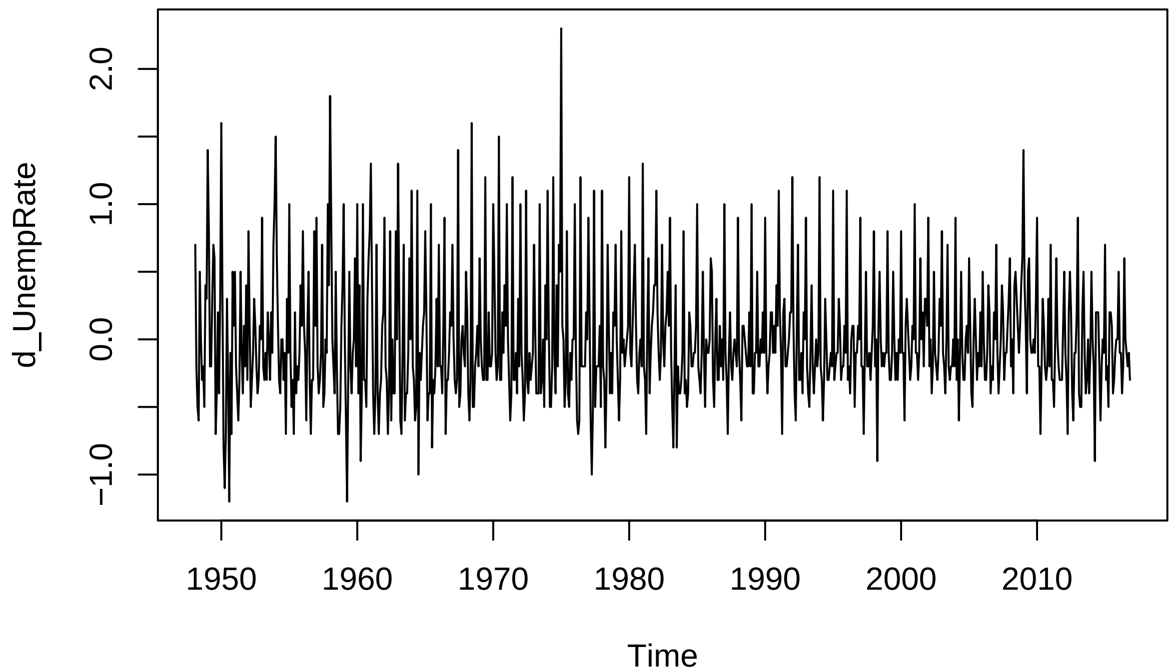


The lag 1,13, 25... is significant. Suggest it is a AR(1) with seasonal pattern.

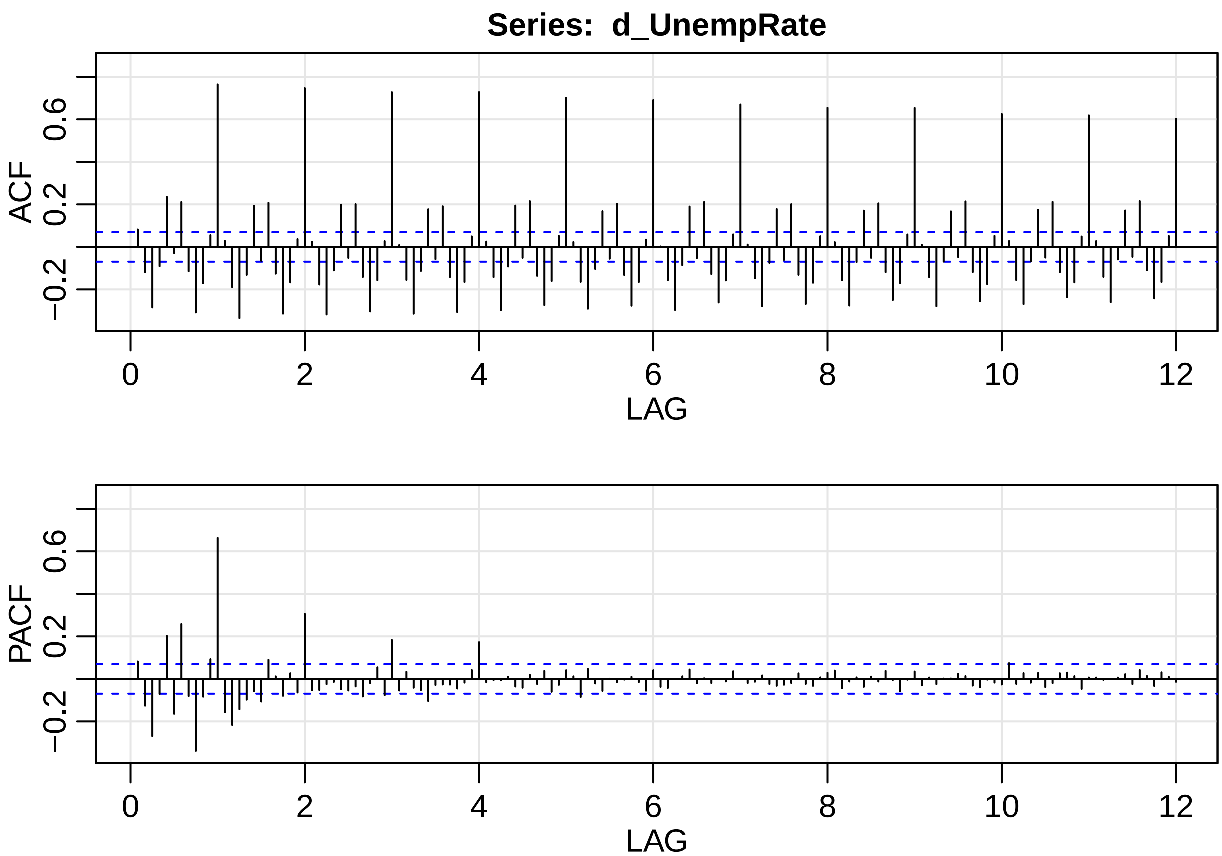
```
l_UnempRate=log(UnempRate)
tsdisplay(l_UnempRate)
```




```
d_UnempRate=diff(UnempRate)
plot(d_UnempRate)
```

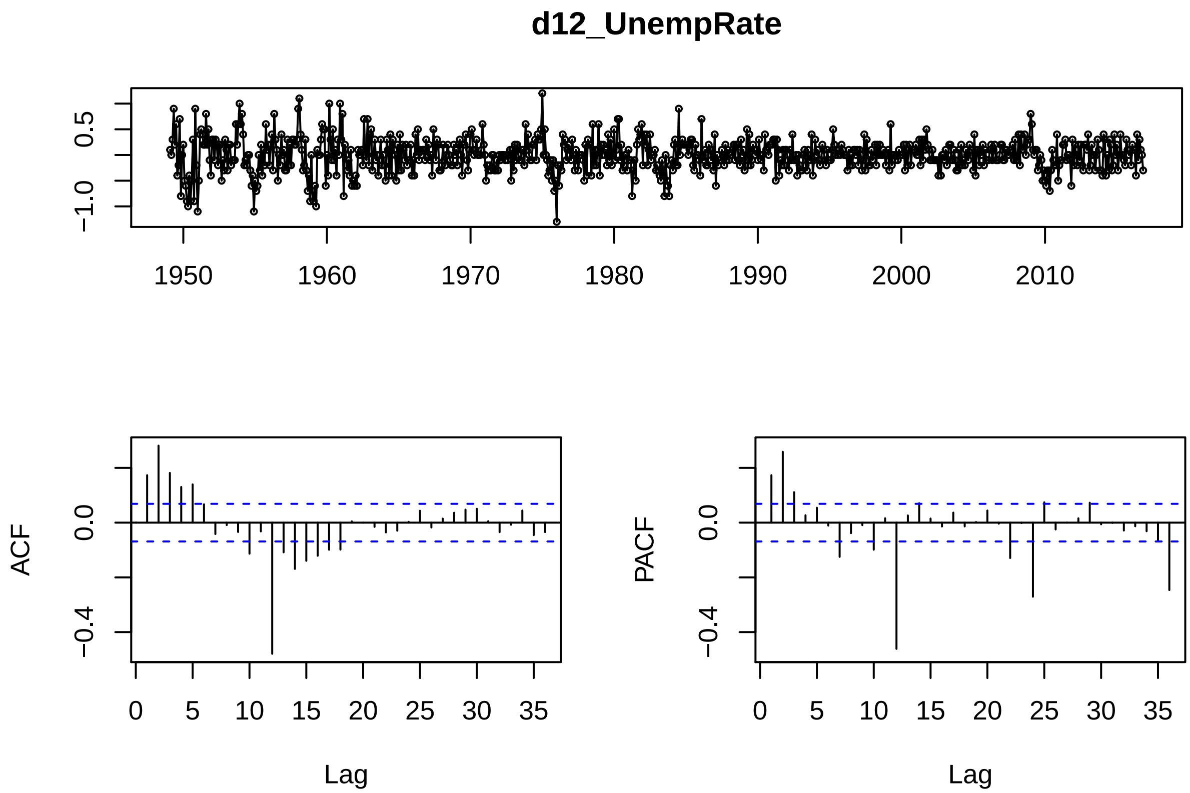



```
invisible(acf2(d_UnempRate,144))
```

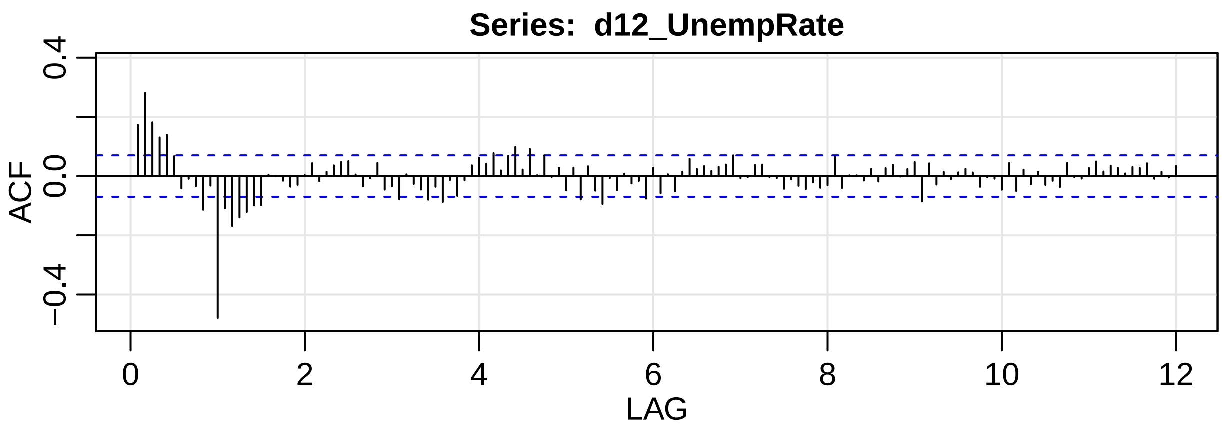


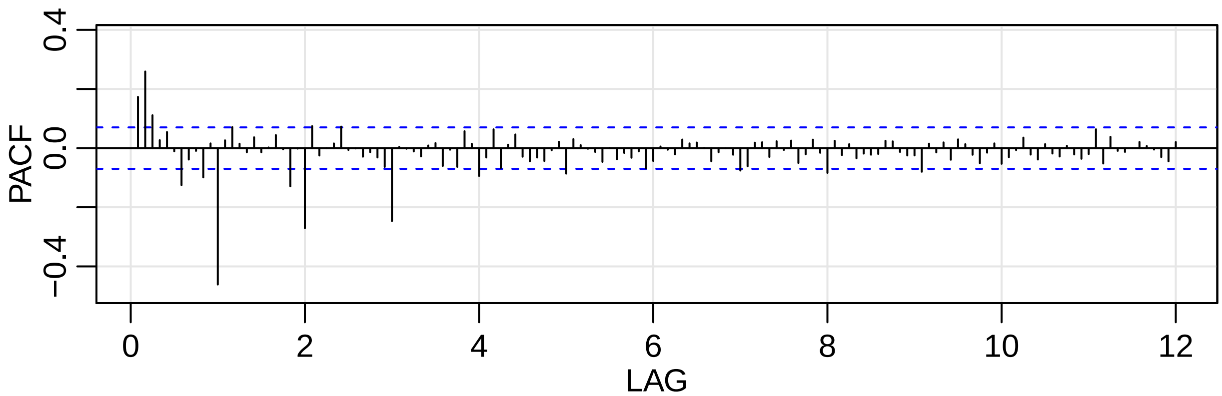
It is exponentially decaying in PACF at lag 12,24,36… The ACF is linearly decaying.

```
d12_UnempRate=diff(d_UnempRate,12) #seasonal differencing the diff(UnempRate)
tsdisplay(d12_UnempRate)
```




```
invisible(acf2(d12_UnempRate,144))
```



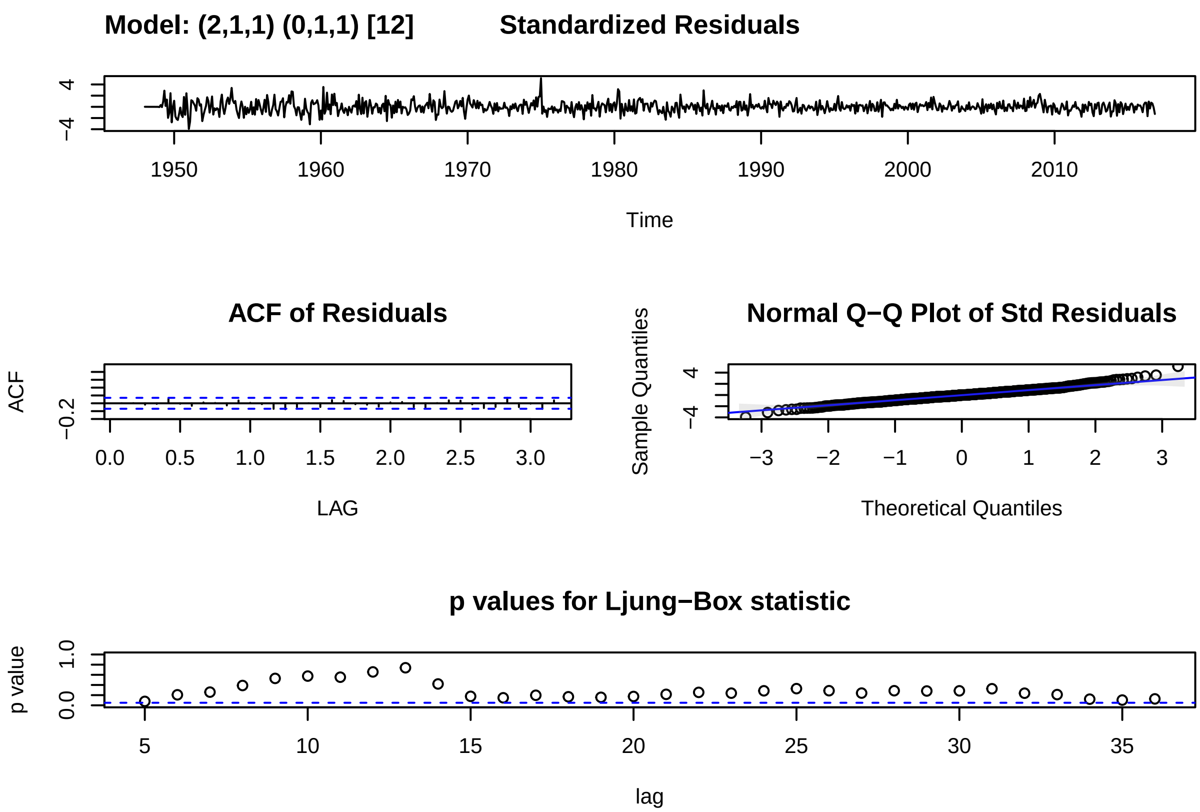


It is clear that the PACF is exponentially decay in lag(12), and one peak in ACF lag(12). Use SARIMA(0,1,1)[12] and try other numbers in ARIMA. We may try different combinations of ARIMA.

```
t=length(l_UnempRate)
sarima(UnempRate,p=2,d=1,q=1,P=0,D=1,Q=1,S=12,xreg=1:t,no.constant = T, details = T)
```

```
## initial  value -1.155932
## iter   2 value -1.357012
## iter   3 value -1.391185
## iter   4 value -1.406765
## iter   5 value -1.420077
## iter   6 value -1.422661
## iter   7 value -1.424201
## iter   8 value -1.424561
## iter   9 value -1.424834
## iter  10 value -1.428478
## iter  11 value -1.429019
## iter  12 value -1.429737
## iter  13 value -1.430052
## iter  14 value -1.430200
## iter  15 value -1.430291
## iter  16 value -1.430303
## iter  17 value -1.430326
## iter  18 value -1.430395
## iter  19 value -1.430400
## iter  20 value -1.430400
## iter  20 value -1.430400
```

```
## iter  20 value -1.430400
## final  value -1.430400
## converged
## initial  value -1.437696
## iter   2 value -1.437906
## iter   3 value -1.438202
## iter   4 value -1.438207
## iter   5 value -1.438209
## iter   6 value -1.438209
## iter   7 value -1.438211
## iter   8 value -1.438212
## iter   9 value -1.438212
## iter  10 value -1.438212
## iter  10 value -1.438212
## iter  10 value -1.438212
## final  value -1.438212
## converged
```

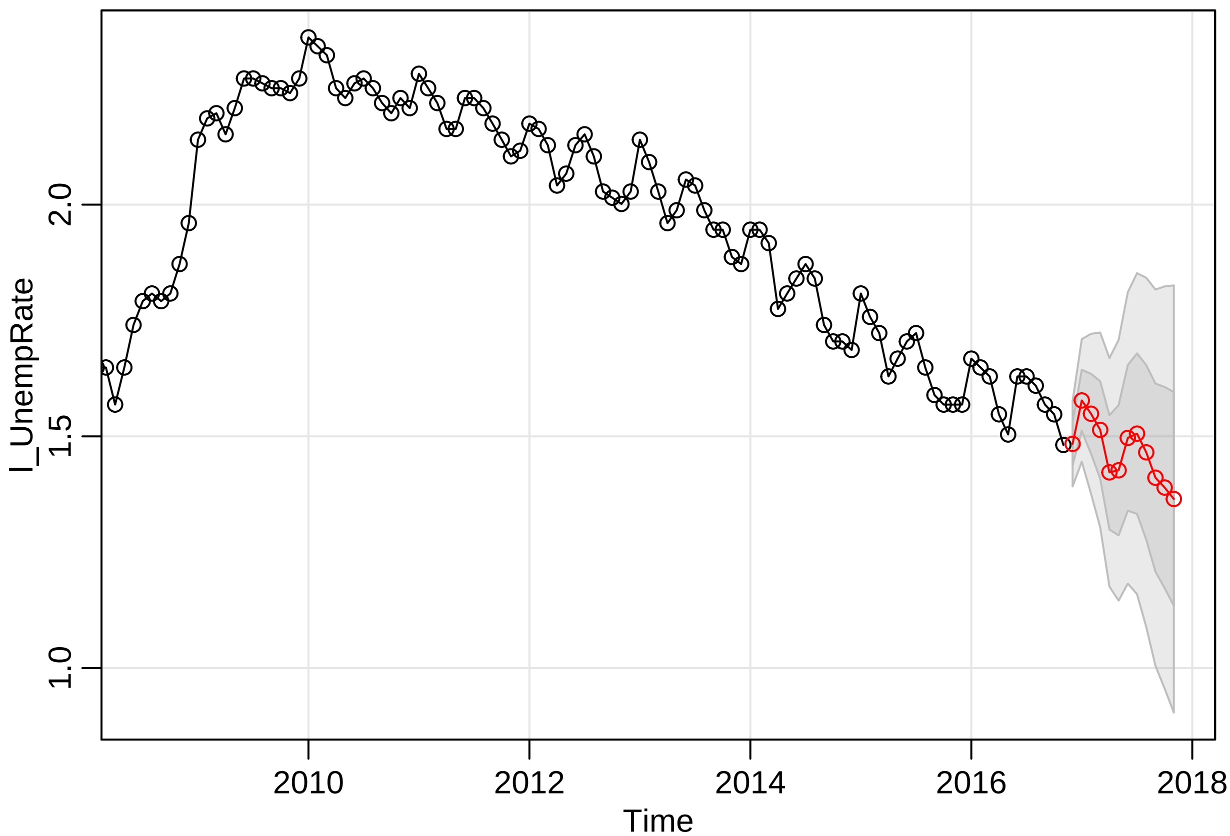



```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, optim.control = list(trace = trc, REPORT = 1,
##     reltol = tol))
##
## Coefficients:
```

```
##          ar1     ar2      ma1     sma1      xreg
##       0.5896  0.1342  -0.4831  -0.7675    0.0111
## s.e.  0.1105  0.0465   0.1090   0.0254  152.3660
##
## sigma^2 estimated as 0.0556:  log likelihood = 15.69,  aic = -19.38
##
## $degrees_of_freedom
## [1] 809
##
## $ttable
##      Estimate       SE  t.value p.value
## ar1    0.5896   0.1105   5.3345  0.0000
## ar2    0.1342   0.0465   2.8831  0.0040
## ma1   -0.4831   0.1090  -4.4316  0.0000
## sma1  -0.7675   0.0254 -30.2424  0.0000
## xreg   0.0111 152.3660   0.0001  0.9999
##
## $AIC
## [1] -1.87757
##
## $AICc
## [1] -1.875028
##
## $BIC
## [1] -2.849046
```

By trying different combinations, I choose $SARIMA(2,1,1) \times (0,1,1)_{12}$. The residual diagnosis shows the residuals are all within the bounds.

```
sarima.for(l_UnempRate,p=2,d=1,q=1,P=0,D=1,Q=1,S=12,no.constant = T, n.ahead=12)
```




```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 2016
## 2017 1.577518 1.548868 1.514026 1.422309 1.426817 1.496744 1.505940
##           Aug      Sep      Oct      Nov      Dec
## 2016                                    1.483859
## 2017 1.465536 1.410861 1.389672 1.364933
##
## $se
##             Jan        Feb        Mar        Apr        May        Jun
## 2016
## 2017 0.06614233 0.08614282 0.10506860 0.12321148 0.14057710 0.15719995
##             Jul        Aug        Sep        Oct        Nov        Dec
## 2016                                                          0.04601404
## 2017 0.17311176 0.18834929 0.20295174 0.21695938 0.23041216
```