Assignment 05; STAT 626

Philip Anderson; panders2@tamu.edu

7/03/2018

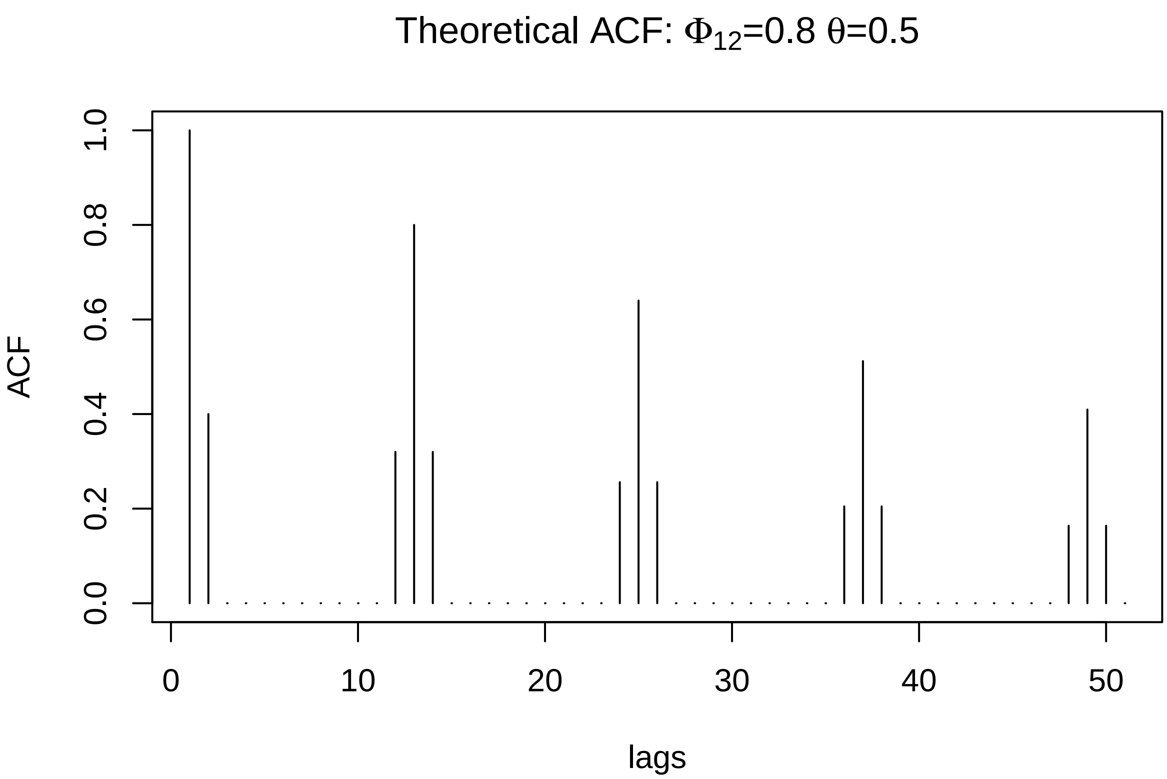
Question 3.19

Plot theoretical ACF of SARIMA(0, 0, 1) x (1,0,0)12.
PHI = 0.8, theta=0.5. lags=50

```
phi <- c(rep(0, 11), 0.8)
paste0("phi length: ", length(phi))
```

```
## [1] "phi length: 12"
```

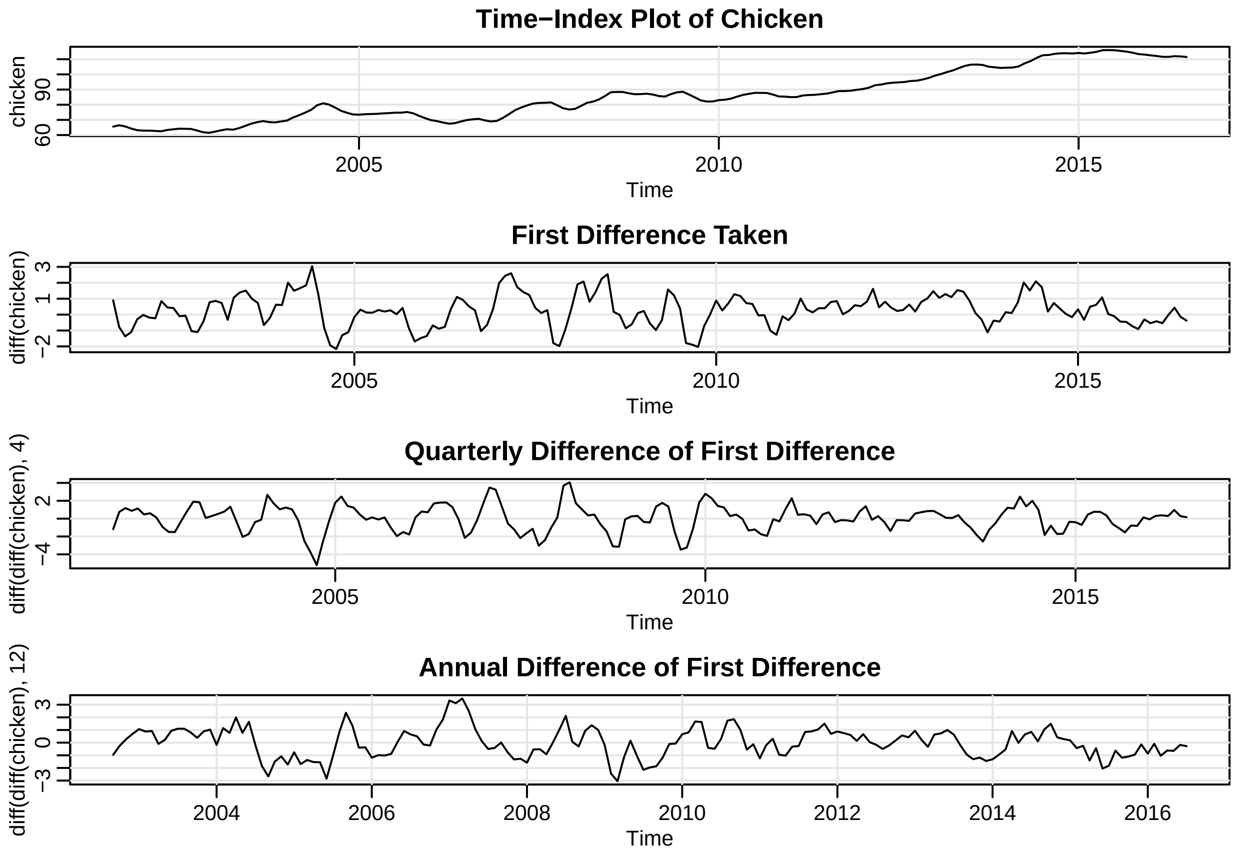
```
ACF <- ARMAacf(ar=phi, ma=0.5, lag.max=50)
plot(ACF, type="h"
     , xlab="lags"
     , main=expression(paste("Theoretical ACF: ", Phi[12], "=0.8 ", theta, "=0.5")))
```



Question 3.20

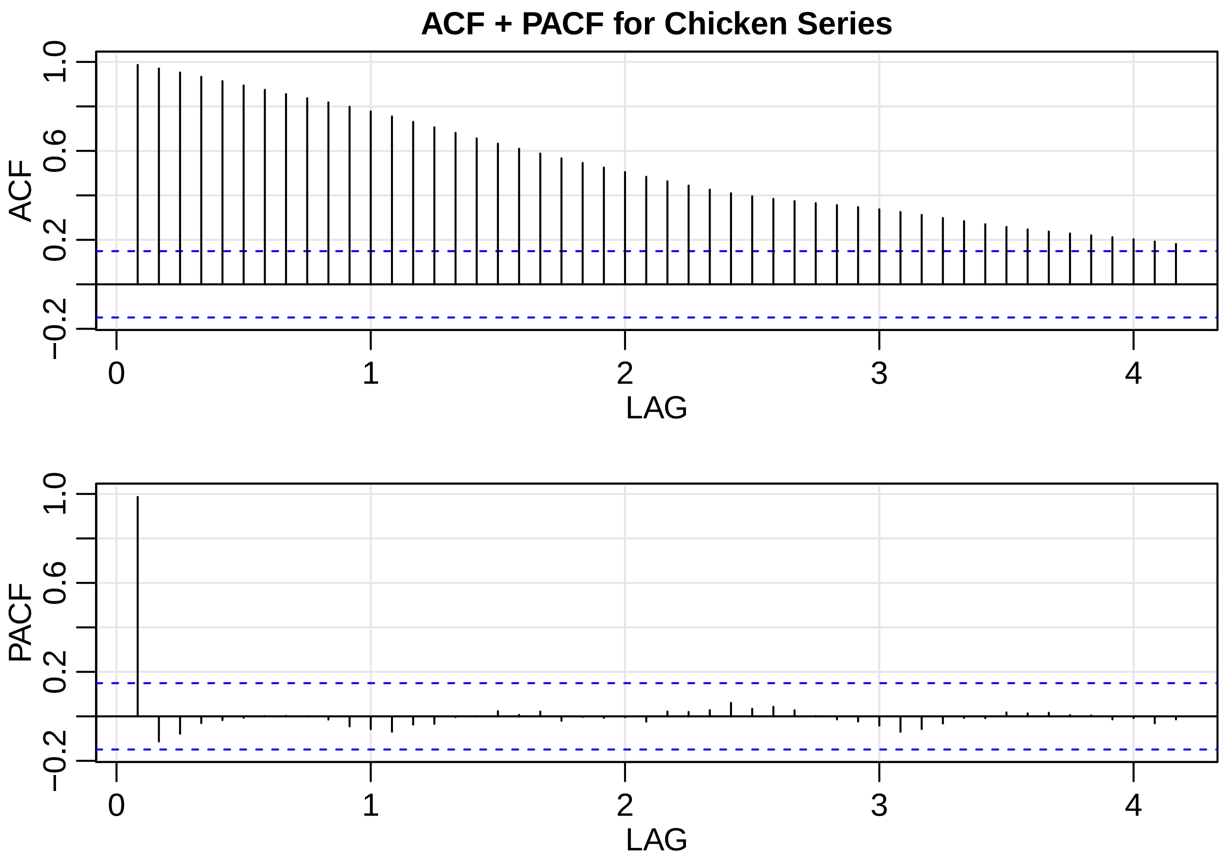
```
# first plot the time series for several variations of our series
par(mfrow=c(4,1))
astsa::tsplot(chicken, main="Time-Index Plot of Chicken")
astsa::tsplot(diff(chicken), main="First Difference Taken")
```

```r
astsa::tsplot(diff(diff(chicken), 4), main="Quarterly Difference of First Difference")
astsa::tsplot(diff(diff(chicken), 12), main="Annual Difference of First Difference")
```

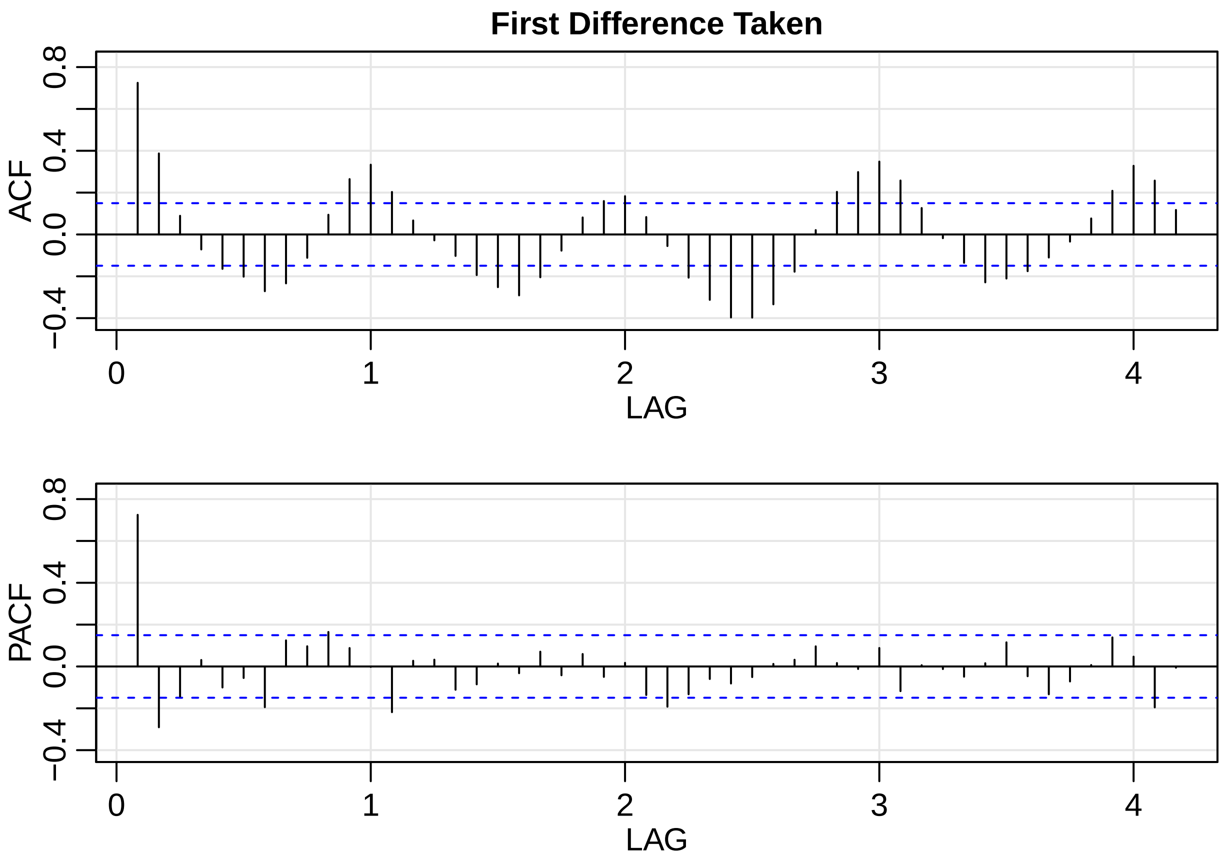


Now, plot the ACF and PACF for each of these to gather more info about them.

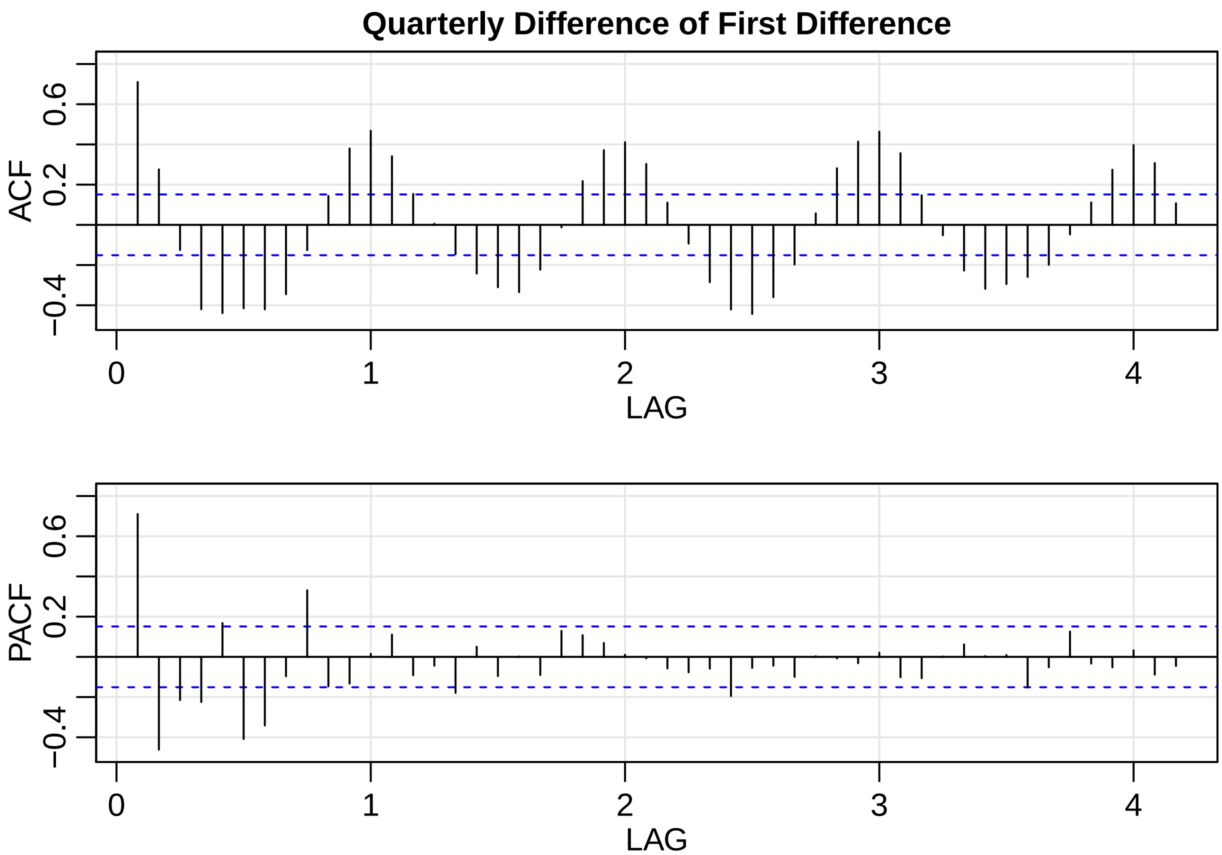
```r
acf2(chicken, max.lag=50, main="ACF + PACF for Chicken Series")
```



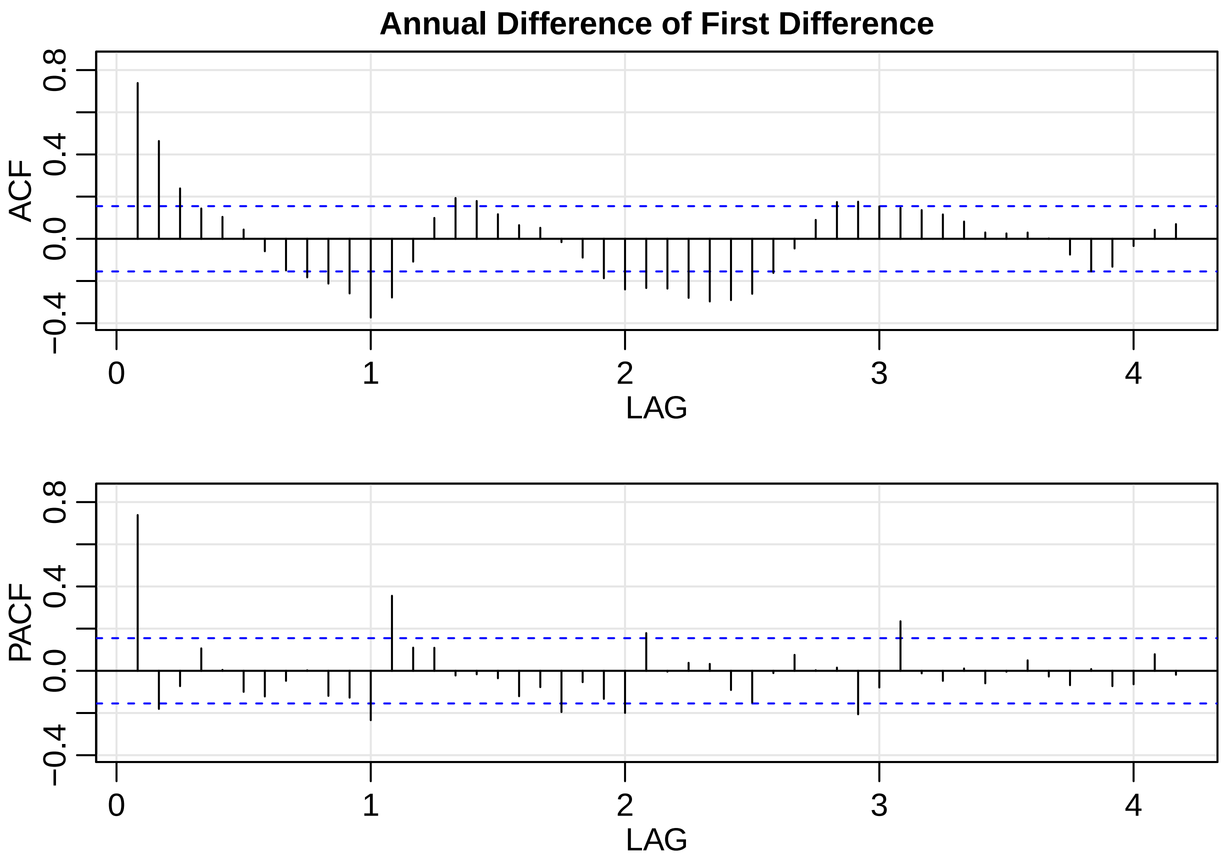

```r
acf2(diff(chicken), max.lag=50, main="First Difference Taken")
```




```
acf2(diff(diff(chicken), 4), max.lag=50, main="Quarterly Difference of First Difference")
```



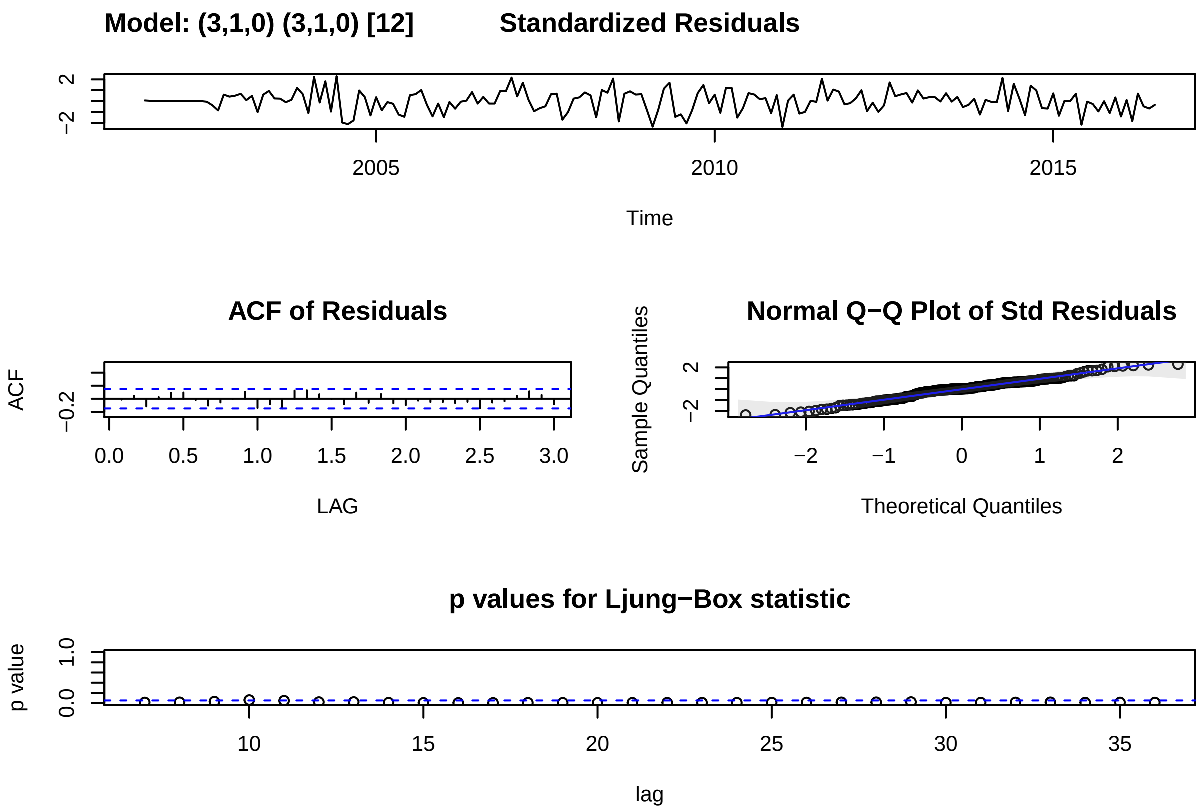

```r
acf2(diff(diff(chicken), 12), max.lag=50, main="Annual Difference of First Difference")
```



From the above ACFs+PACFs, it appears that annual differencing of the first differenced series is a good option. The final ACF+PACF combination shows a tailing off in the ACF, but a cut-off in the PACF after seasonal lag 3. This suggests that a suitable model may be SAR(3), P=3, Q=0 in S=12, or SARIMA(3,1,0)x(3,1,0)_12. Let's fit that here:

```
smod <- sarima(xdata=chicken, p=3, d=1, q=0, P=3, D=1, Q=0, S=12)
```

```
## Warning in sqrt(diag(fitit$var.coef)): NaNs produced

## Warning in sqrt(diag(fitit$var.coef)): NaNs produced
```




```
print(smod)
```

```
## Warning in sqrt(diag(x$var.coef)): NaNs produced
```

The above plots look ok, but the Ljung-Box p-values are highly significant, and the model fit procedure failed, which is not a good sign. Let's loop through some potential variations of the model and see if we can improve on the BIC.

I am going to do a nested for-loop. AR differences of [0,1,2,3], MA differences of [0,1,2,3], at both the immediate and seasonal levels. Also going to try Seasonal = 4, 12, in case the chicken prices fluctuate quarterly.

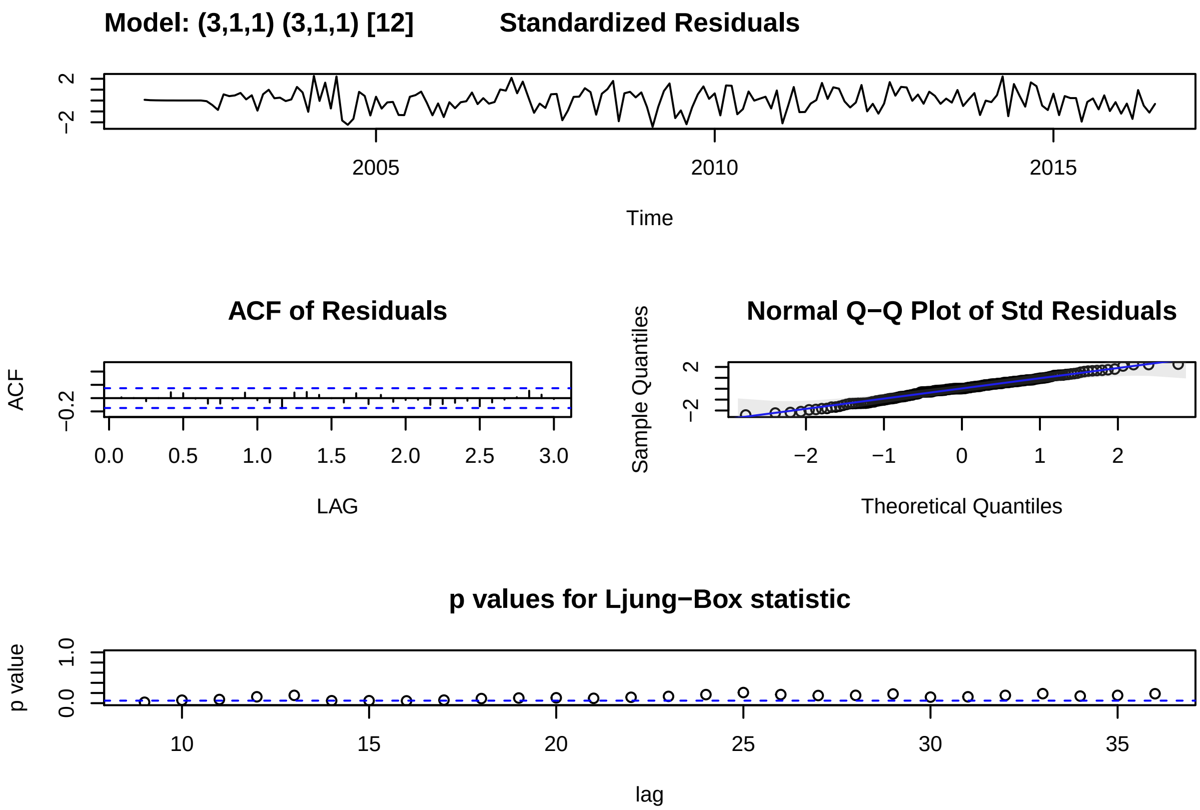
```
puzzle <- cbind(0, 0, 0, 0, 0)

for (k in c(12, 4)) {
  for (i in 0:3) {
    for (j in 0:3) {
      smod <- try(sarima(xdata=chicken, p=i, d=1, q=j, P=i, D=1, Q=j, S=k), silent=T)

        if(class(smod)=="list") {
            piece <- cbind(i, j, k, smod$AIC, smod$BIC)
                                }
        else {
            piece <- cbind(i, j, k, 0, 0)
             }

      puzzle <- rbind(puzzle, piece)
    }
  }
}
```

```r
# print our results, in order of ascending BIC and AIC
puzzle2 <- data.frame(puzzle)
names(puzzle2) <- c("i", "j", "k", "AIC", "BIC")
puzzle2 %>% dplyr::arrange(BIC, AIC)
```

```
##    i j  k         AIC        BIC
## 1  0 2 12 -0.10455512 -1.0336005
## 2  1 1 12 -0.06553306 -0.9945785
## 3  0 3 12 -0.08842420 -0.9819923
## 4  2 1 12 -0.07255410 -0.9661222
## 5  1 2 12 -0.07107875 -0.9646469
## 6  1 3 12 -0.07719480 -0.9352856
## 7  3 1 12 -0.07396589 -0.9320567
## 8  3 3  4 -0.12987864 -0.9170148
## 9  0 1 12  0.05551807 -0.9090046
## 10 3 3 12 -0.12008079 -0.9072170
## 11 2 3 12 -0.07144800 -0.8940615
## 12 3 2 12 -0.04069674 -0.8633103
## 13 2 3  4 -0.03129321 -0.8539067
## 14 3 0 12  0.04056751 -0.8530006
## 15 3 1  4  0.03820038 -0.8198904
## 16 1 1  4  0.12210880 -0.8069366
## 17 2 1  4  0.09818935 -0.7953788
## 18 0 2  4  0.13389282 -0.7951526
## 19 0 3  4  0.10493279 -0.7886353
## 20 2 0 12  0.14190984 -0.7871356
## 21 2 2  4  0.07274329 -0.7853475
## 22 1 2  4  0.12466217 -0.7689059
## 23 3 2  4  0.05834919 -0.7642643
## 24 1 3  4  0.10219198 -0.7558988
## 25 2 0  4  0.25185771 -0.6771877
## 26 3 0  4  0.22343956 -0.6701285
## 27 1 0 12  0.33788675 -0.6266360
## 28 0 1  4  0.37272369 -0.5917990
## 29 1 0  4  0.82863849 -0.1358842
## 30 0 0  0  0.00000000  0.0000000
## 31 2 2 12  0.00000000  0.0000000
## 32 0 0 12  1.30021681  0.3002168
## 33 0 0  4  1.76832436  0.7683244
```

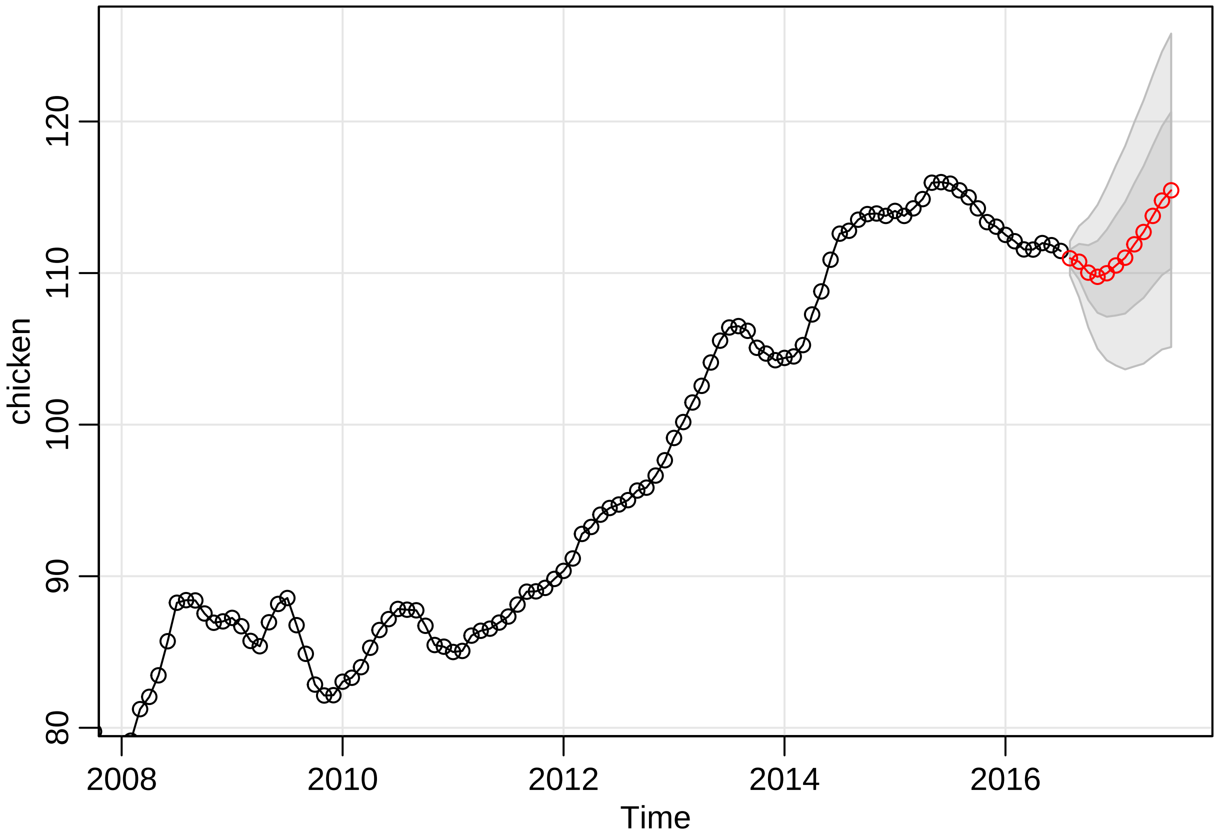
Of the models evaluated, it looks like SARIMA(3,1,1)x(3,1,1)12 is the strongest combination of what I had originally hypothesized and what is borne out by fit statistics.

```r
smod <- sarima(xdata=chicken, p=3, d=1, q=1, P=3, D=1, Q=1, S=12)
```



Forecast this model 12 months into the future.

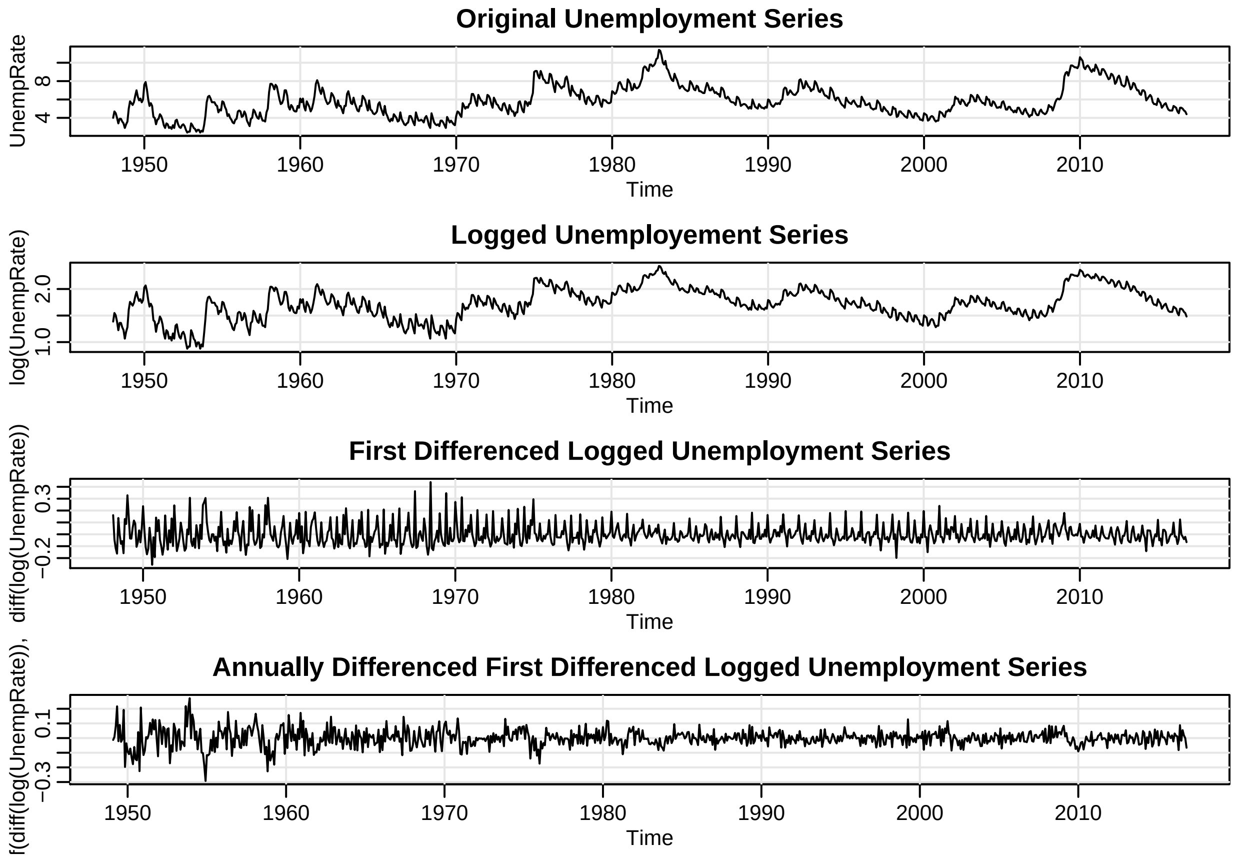
```
sarima.for(xdata=chicken, p=3, d=1, q=1, P=3, D=1, Q=1, S=12, n.ahead=12)
```




```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 2016
## 2017 110.5020 111.0135 111.8971 112.7097 113.7723 114.7813 115.4591
##           Aug      Sep      Oct      Nov      Dec
## 2016 110.9760 110.7482 110.0279 109.7537 109.9881
## 2017
##
## $se
##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 2016
## 2017 3.3007139 3.6839357 4.0281565 4.3436162 4.6361881 4.9108763 5.1706327
##            Aug       Sep       Oct       Nov       Dec
## 2016 0.5591012 1.1773626 1.8078250 2.3716907 2.8686388
## 2017
```

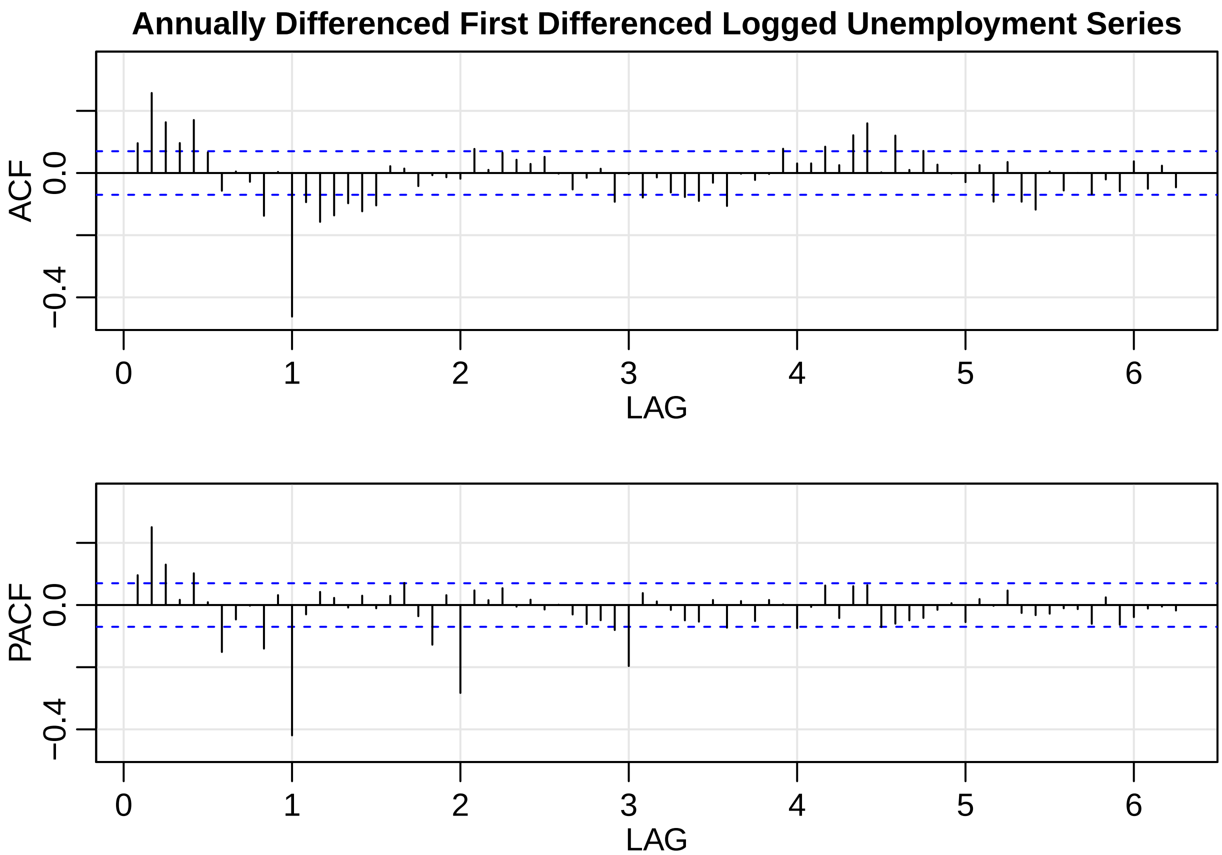
Question 3.21

```
# first, fit several time series plots to variations of this data
par(mfrow=c(4,1))
astsa::tsplot(UnempRate, main="Original Unemployment Series")
astsa::tsplot(log(UnempRate), main="Logged Unemployement Series")
astsa::tsplot(diff(log(UnempRate)), main="First Differenced Logged Unemployment Series")
astsa::tsplot(diff(diff(log(UnempRate)), 12), main="Annually Differenced First Differenced Logged Unempl
```



The only one of these that looks potentially stationary is the final attempt. Let's plot the ACF and PACF to get additional information.

```
astsa::acf2(diff(diff(log(UnempRate)), 12), 75, main="Annually Differenced First Differenced Logged Unem
```



The ACF cuts off after the first season, but the PACF tails off. This could suggest a SMA model, or SARIMA(0,1,1)x(0,1,1)12, but we will work through several iterations to make sure.

```
puzzle <- cbind(0, 0, 0, 0, 0)

for (k in c(12)) {
  for (i in 0:3) {
    for (j in 0:3) {
      smod <- try(sarima(xdata=UnempRate, p=i, d=1, q=j, P=i, D=1, Q=j, S=k), silent=T)

        if(class(smod)=="list") {
            piece <- cbind(i, j, k, smod$AIC, smod$BIC)
                                   }
        else {
            piece <- cbind(i, j, k, 0, 0)
             }

      puzzle <- rbind(puzzle, piece)
    }
  }
}
```

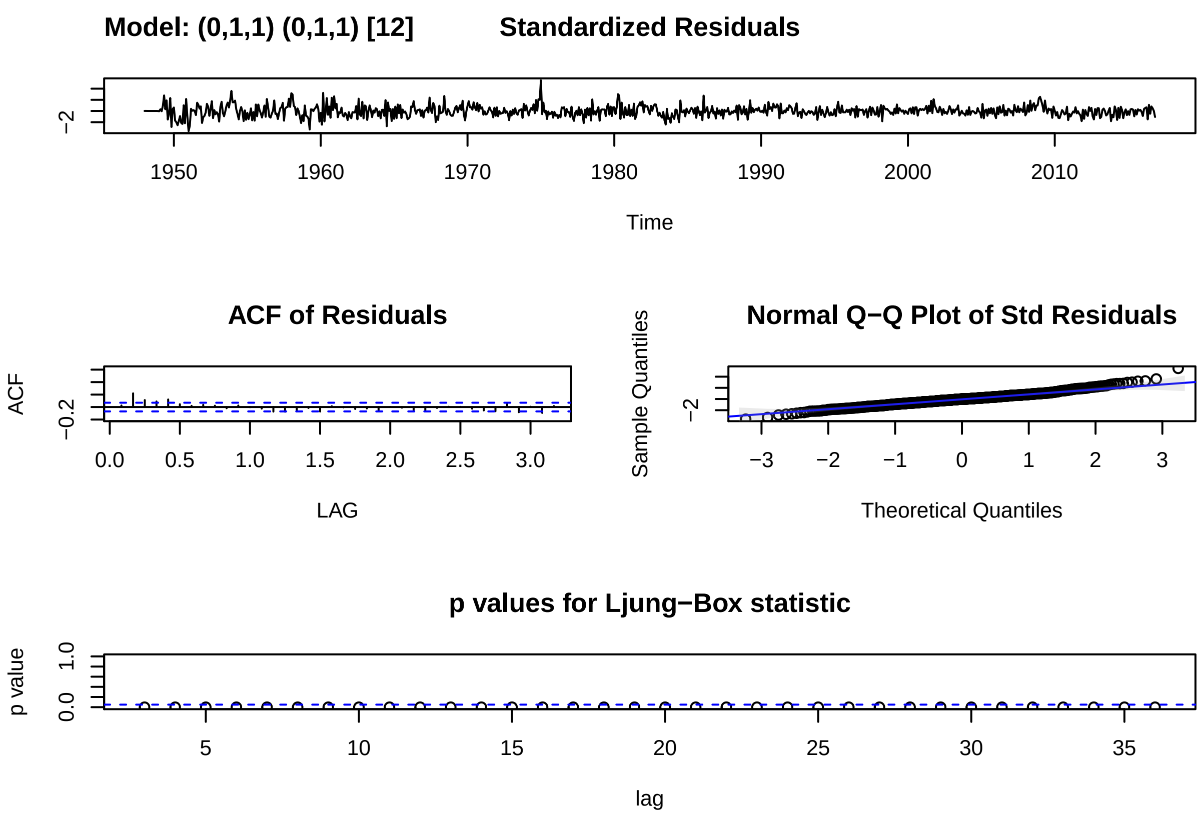
```
puzzle2 <- data.frame(puzzle)
names(puzzle2) <- c("i", "j", "k", "AIC", "BIC")
puzzle2 %>% dplyr::arrange(BIC, AIC)
```

```
##    i j  k       AIC       BIC
## 1  1 1 12 -1.870598 -2.847779
```

```
## 2  1 2 12 -1.877796 -2.843567
## 3  2 1 12 -1.875346 -2.841118
## 4  0 2 12 -1.853916 -2.831097
## 5  1 3 12 -1.872174 -2.826536
## 6  3 1 12 -1.869982 -2.824345
## 7  2 2 12 -1.868760 -2.823122
## 8  0 3 12 -1.856527 -2.822298
## 9  2 3 12 -1.865338 -2.808291
## 10 3 3 12 -1.875498 -2.807041
## 11 0 1 12 -1.818442 -2.807032
## 12 3 0 12 -1.833010 -2.798781
## 13 2 0 12 -1.765740 -2.742921
## 14 1 0 12 -1.622175 -2.610765
## 15 0 0 12 -1.314200 -2.314200
## 16 0 0  0  0.000000  0.000000
## 17 3 2 12  0.000000  0.000000
```

The model hypothesized does not have dramatically different results from the others.

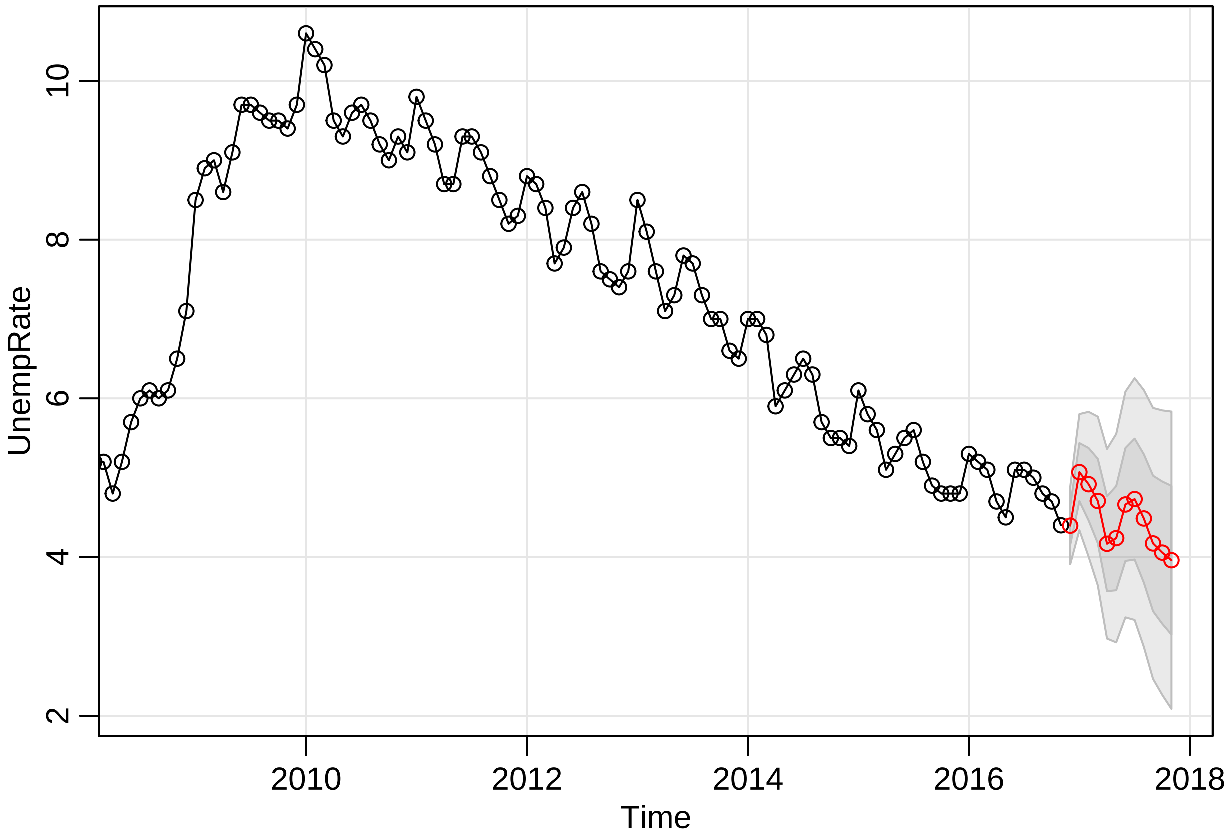
```
smod <- sarima(xdata=UnempRate, p=0, d=1, q=1, P=0, D=1, Q=1, S=12)
```




```
smod
```

Forecast this series 12 months into the future.

```
sarima.for(xdata=UnempRate, p=0, d=1, q=1, P=0, D=1, Q=1, S=12, n.ahead=12)
```




```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 2016
## 2017 5.070169 4.917907 4.706359 4.167457 4.238349 4.662388 4.730676
##           Aug      Sep      Oct      Nov      Dec
## 2016                                     4.395719
## 2017 4.486049 4.171379 4.055880 3.960063
##
## $se
##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 2016
## 2017 0.3657912 0.4562844 0.5315899 0.5974782 0.6567895 0.7111714 0.7616805
##            Aug       Sep       Oct       Nov       Dec
## 2016                                         0.2437434
## 2017 0.8090423 0.8537809 0.8962891 0.9368705
```

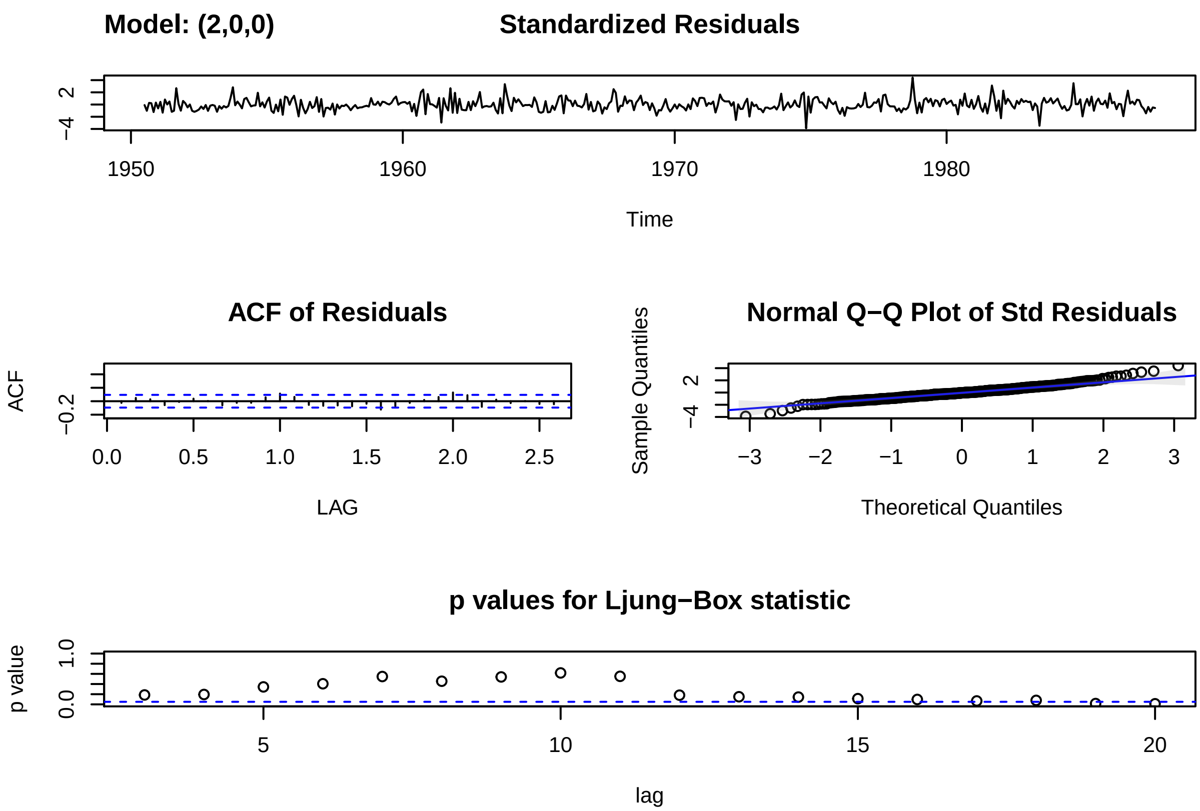
Question 3.24

Rerun the model from Example 3.33, but fit an ARIMA(2, 0, 0) × (0, 1, 1)12 to the residuals. Does this improve the fit?

```
# replicate the original example
dummy = ifelse(soi<0, 0, 1)
fish  = ts.intersect(rec, soiL6=stats::lag(soi,-6), dL6=stats::lag(dummy,-6), dframe=TRUE)
summary(fit <- lm(rec ~soiL6*dL6, data=fish, na.action=NULL))
attach(fish)
```

```r
tsplot(resid(fit))
acf2(resid(fit)) # indicates AR(2)
intract = soiL6*dL6 # interaction term
```

```r
smod <- sarima(rec,2,0,0, xreg = cbind(soiL6, dL6, intract))
```

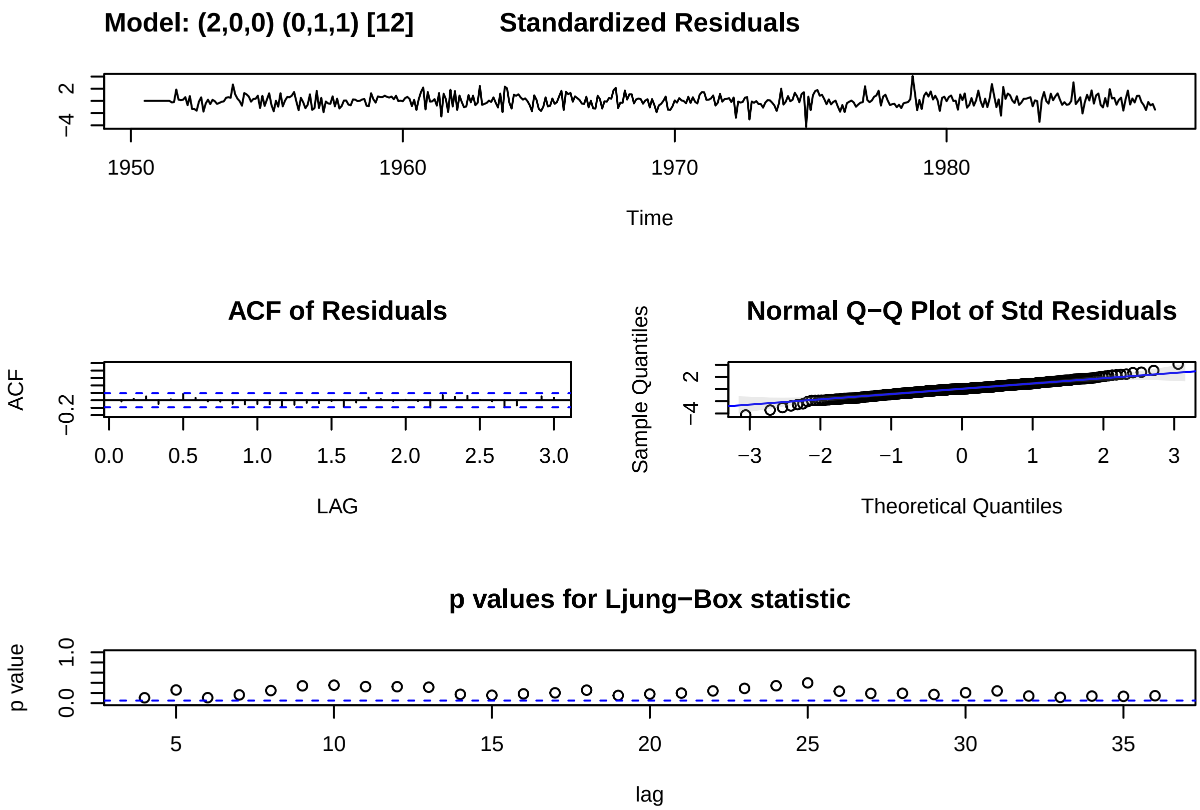



```r
smod
```

```
## $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  intercept   soiL6      dL6   intract
##       1.3624  -0.4703    64.8028  8.6671  -2.5945  -10.3092
## s.e.  0.0440   0.0444     4.1121  2.2205   0.9535    2.8311
##
## sigma^2 estimated as 86.78:  log likelihood = -1633.07,  aic = 3280.13
##
## $degrees_of_freedom
## [1] 441
##
## $ttable
##            Estimate     SE  t.value p.value
## ar1          1.3624 0.0440  30.9303  0.0000
## ar2         -0.4703 0.0444 -10.5902  0.0000
```

```
## intercept  64.8028 4.1121  15.7590  0.0000
## soiL6        8.6671 2.2205   3.9033  0.0001
## dL6         -2.5945 0.9535  -2.7209  0.0068
## intract    -10.3092 2.8311  -3.6415  0.0003
##
## $AIC
## [1] 5.490258
##
## $AICc
## [1] 5.495303
##
## $BIC
## [1] 4.545326
```

```
smod <- sarima(rec,2,0,0, 0, 1, 1, 12, xreg = cbind(soiL6, dL6, intract))
```




```
smod
```

```
## $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     sma1   soiL6      dL6   intract
##       1.2702  -0.3497  -1.0000  8.8775  -2.2932   -8.4351
## s.e.  0.0462   0.0464   0.0444  2.2269   0.9655    2.8655
##
```

```
## sigma^2 estimated as 73.46:  log likelihood = -1574.52,  aic = 3163.04
##
## $degrees_of_freedom
## [1] 429
##
## $ttable
##         Estimate     SE  t.value p.value
## ar1       1.2702 0.0462  27.4652  0.0000
## ar2      -0.3497 0.0464  -7.5390  0.0000
## sma1     -1.0000 0.0444 -22.5346  0.0000
## soiL6     8.8775 2.2269   3.9864  0.0001
## dL6      -2.2932 0.9655  -2.3752  0.0180
## intract  -8.4351 2.8655  -2.9437  0.0034
##
## $AIC
## [1] 5.323649
##
## $AICc
## [1] 5.328694
##
## $BIC
## [1] 4.378717
```

```r
detach(fish)
```

The specified SARIMA model is a marginally better fit over the ARIMA model, according to the BIC fit statistic. The residual ACF and QQ-Plots appear slightly improved as well.