Homework 10

Joshua Ingram
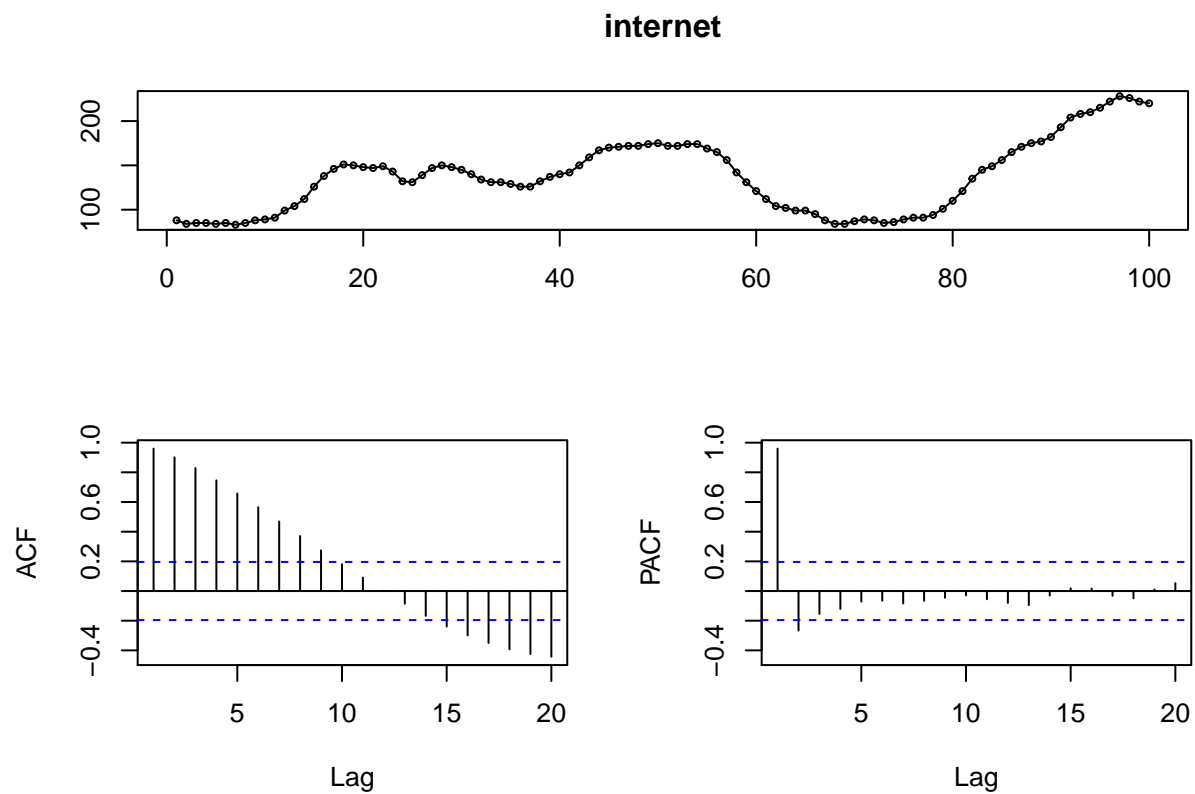
11/19/2020

Problem 1

1.

a.

```
tsdisplay(internet)
```

internet

```
#tsdisplay(log(internet))
#tsdisplay(BoxCox(internet,BoxCox.lambda(internet)))
```

After checking the original time series and testing some transformations, it doesn't seem that it is necessary to make any transformations on the data to stabalize the variance. (commented out log/BoxCox so I don't flood the pages with graphs)

b.

```
tsdisplay(diff(internet), lag.max = 100)
```

diff(internet)
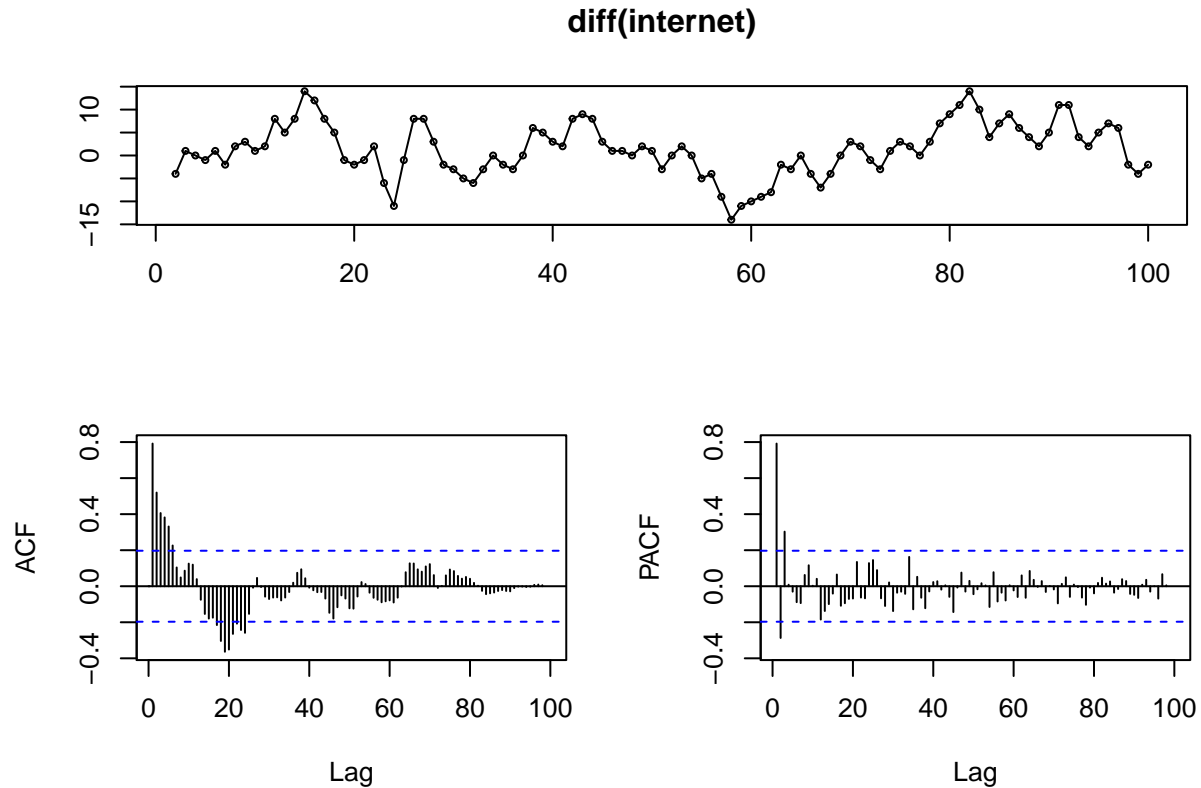
The mean does not appear to be constant, so a first-order differencing appears to stabalize the mean and the data has pretty stable variance.

c.

(looking at the ACf/PACF plots above)

The ACF plot decays exponentially with a sinusoidal pattern and here is a "spike" in the PACF at a lag 1. For these reasons, I would suggest p = 1, q = 0.

d.

For ARMIMA(1,1,0) and given $y'_t = y_t - y_{t-1}$, the modeling equation is:

$$y'_t = \mu + \phi_1 y'_{t-1} + \epsilon_t, \quad \epsilon_t \sim_{i.i.d.} N(0, \sigma^2)$$

```
fit1 <- Arima(internet, order=c(1,1,0))
summary(fit1)
```
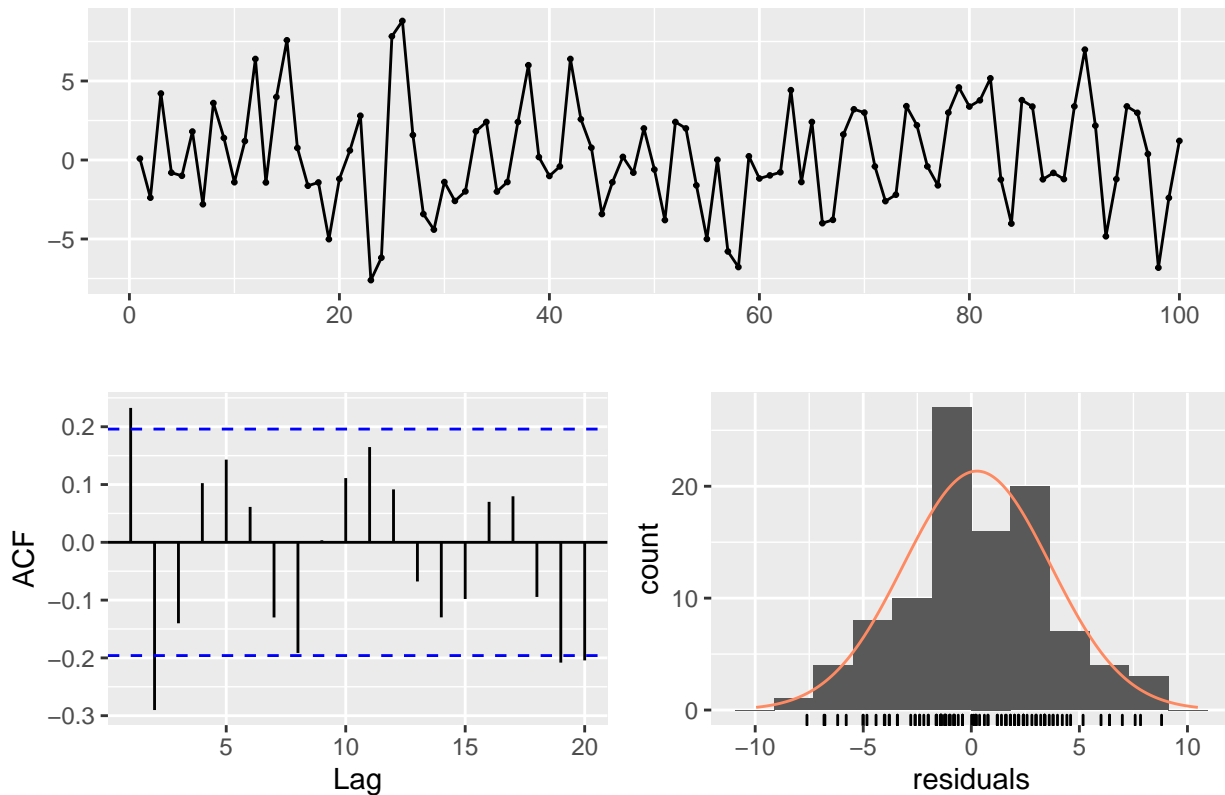
```
## Series: internet
## ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##       0.8026
## s.e.  0.0580
##
## sigma^2 estimated as 11.79:  log likelihood=-262.62
## AIC=529.24   AICc=529.36   BIC=534.43
##
## Training set error measures:
##                    ME      RMSE      MAE       MPE      MAPE       MASE       ACF1
## Training set 0.2615066 3.399188 2.73737 0.2947748 2.136749 0.6049099 0.2325748
```

Fitted Equation:

$$\hat{y}'_t = 0.8026 y'_{t-1}$$

```
checkresiduals(fit1)
```


Residuals from ARIMA(1,1,0)

3

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,0)
## Q* = 27.445, df = 9, p-value = 0.00118
##
## Model df: 1.   Total lags used: 10
```

Based upon the Ljung-Box test, we receive a small p-value, the residuals are serially correlated (not white noise) and the model is not a good fit.

2.

a.

```r
arima.param <- list(c(1,1,0),c(2,1,0),c(3,1,0), c(4,1,0), c(1,1,1), c(1,1,2))

#AICc
arima.aicc <- sapply(arima.param,function(x) Arima(internet, order=x)$aicc)
print(arima.aicc)
```

```
## [1] 529.3628 522.4308 512.4195 514.5749 514.5521 516.6774
```

```r
print(arima.param[[which.min(arima.aicc)]])
```

```
## [1] 3 1 0
```

```r
# RMSE for test data
train <- window(internet,start=c(1), end=c(90))
test <- window(internet, start=c(91))
test.rmse <- c()
for (param in arima.param){
  fit <- Arima(train, order=param[1:3])
  test.rmse <- c(test.rmse, accuracy(forecast(fit,24), test)[2,"RMSE"])
}

# printing RMSE and best ARIMA
print(test.rmse)
```

```
## [1] 20.54748 22.81439 17.41143 18.41400 21.35967 20.10155
```

```r
print(arima.param[[which.min(test.rmse)]])
```

```
## [1] 3 1 0
```

The ARIMA(3,1,0) model was the best fit for both the AICc and RMSE (for test data).

b.

```
fit2 <- Arima(internet, order=c(3,1,0))
summary(fit2)
```

```
## Series: internet
## ARIMA(3,1,0)
##
## Coefficients:
##          ar1      ar2     ar3
##       1.1513  -0.6612  0.3407
## s.e.  0.0950   0.1353  0.0941
##
## sigma^2 estimated as 9.656:  log likelihood=-252
## AIC=511.99   AICc=512.42   BIC=522.37
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 0.230588 3.044632 2.367157 0.2748377 1.890528 0.5230995
##                   ACF1
## Training set -0.003095066
```

```
checkresiduals(fit2)
```

Residuals from ARIMA(3,1,0)


```
##
##  Ljung-Box test
```
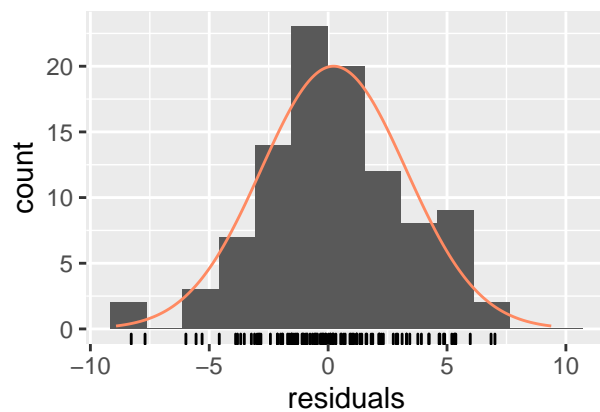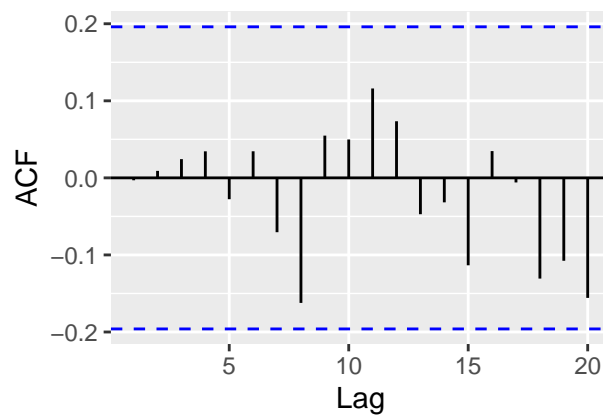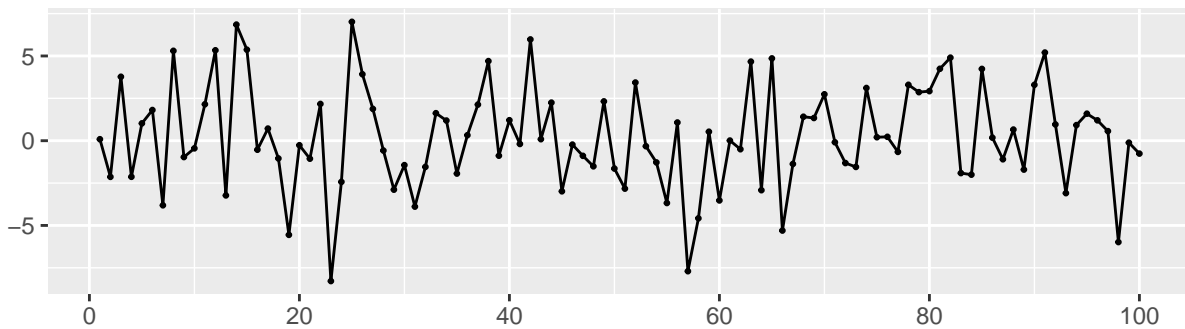
```
##
## data:  Residuals from ARIMA(3,1,0)
## Q* = 4.4913, df = 7, p-value = 0.7218
##
## Model df: 3.    Total lags used: 10
```

After checking the residuals and conducting the Ljung-Box test, we may conclude that the ARIMA(3,1,0) model is a good fit. The residuals are normally distributed, the p-value from our test was large, the AC value in the ACF plot are all within the bands, and the mean is constant.
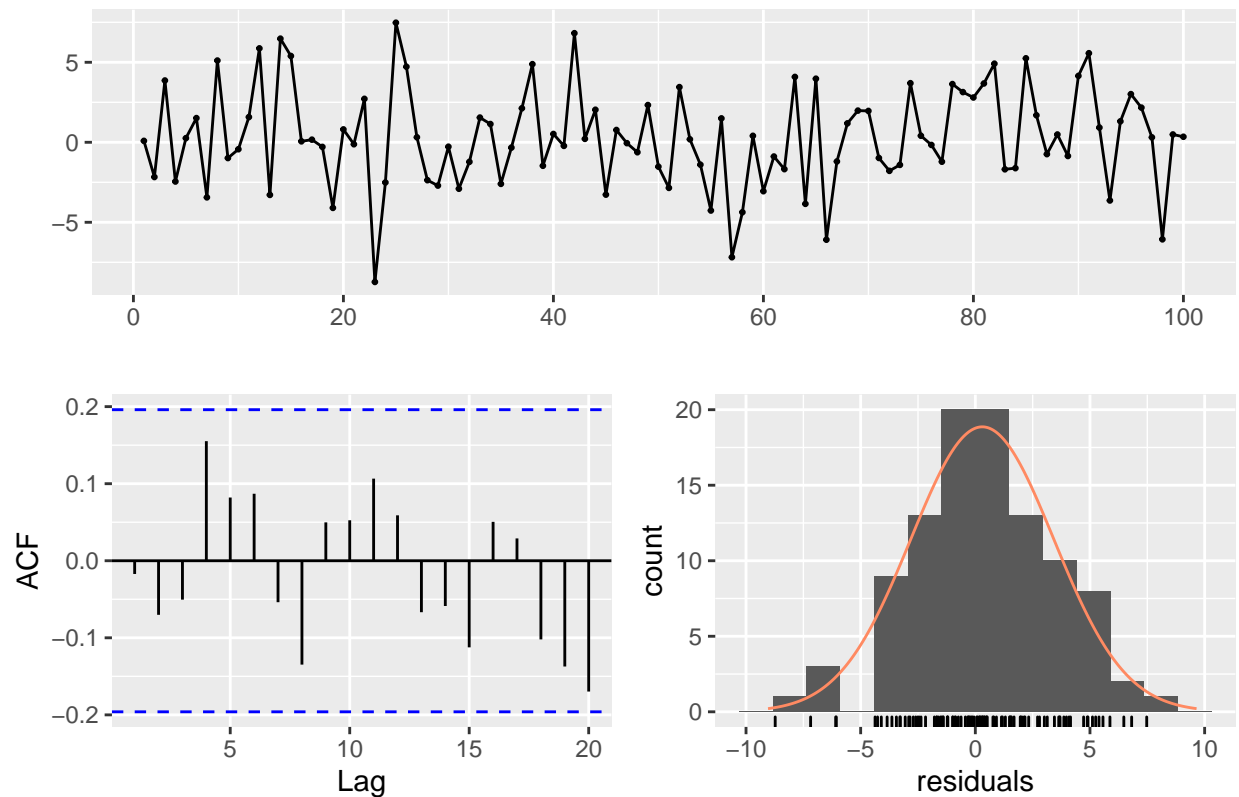
3.

```
fit3 <- auto.arima(internet, approximation = FALSE)
fit3
```

```
## Series: internet
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1      ma1
##       0.6504   0.5256
## s.e.  0.0842   0.0896
##
## sigma^2 estimated as 9.995:  log likelihood=-254.15
## AIC=514.3   AICc=514.55   BIC=522.08
```

ARIMA(1,1,1) was selected by the auto.arima function.

```
checkresiduals(fit3)
```

Residuals from ARIMA(1,1,1)


```
## 
##  Ljung-Box test
## 
## data:  Residuals from ARIMA(1,1,1)
## Q* = 7.8338, df = 8, p-value = 0.4499
## 
## Model df: 2.    Total lags used: 10
```

Based upon the p-value from the Ljung-Box test, we don't reject the null hypothesis so there is no serial correlation and the data is white noise. We can also see this in the data, as it has constant mean, the residuals are normally distributed, and the ACF values are within the bands.
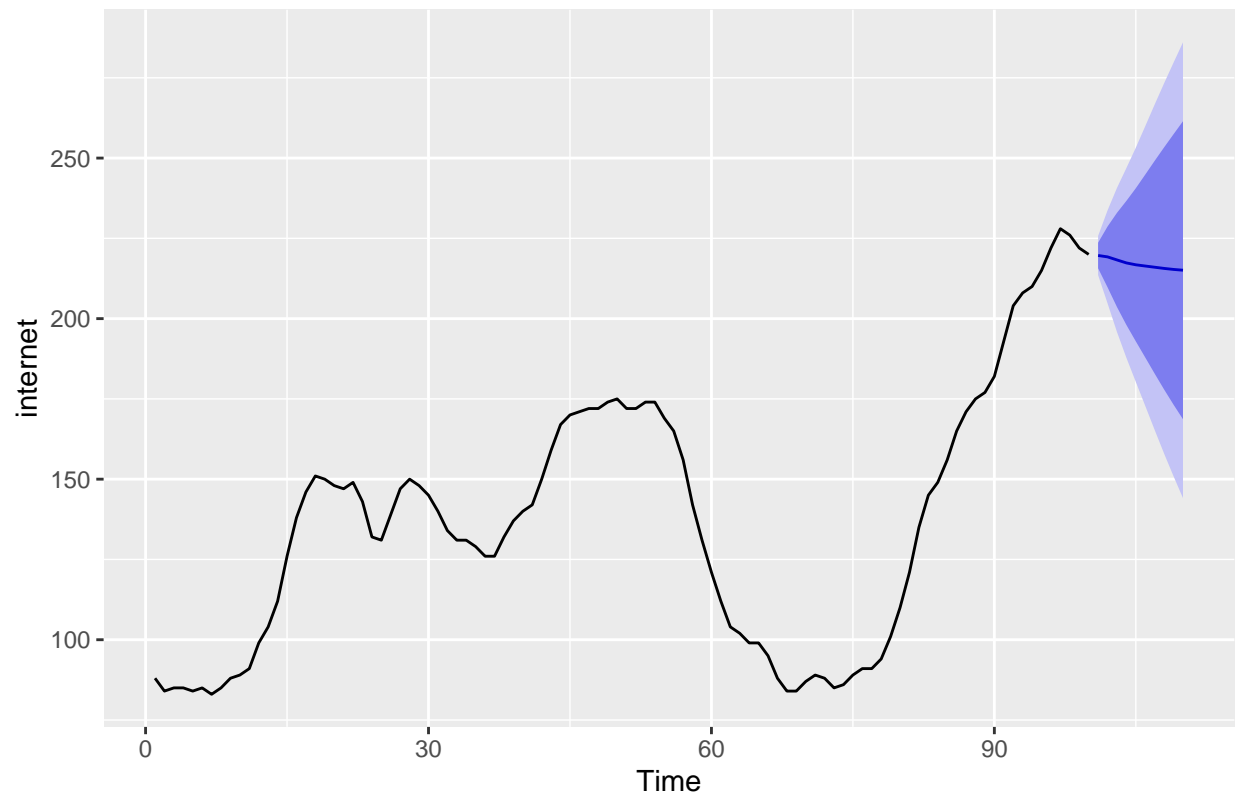
This constitutes a good fit.

4.

The ARIMA(3,1,0) model was the best fit by the AICc (512) and residual diagnostics (largest p-value in the Ljung-Box test and passes other assumptions).

```
autoplot(forecast(fit2, h = 10))
```
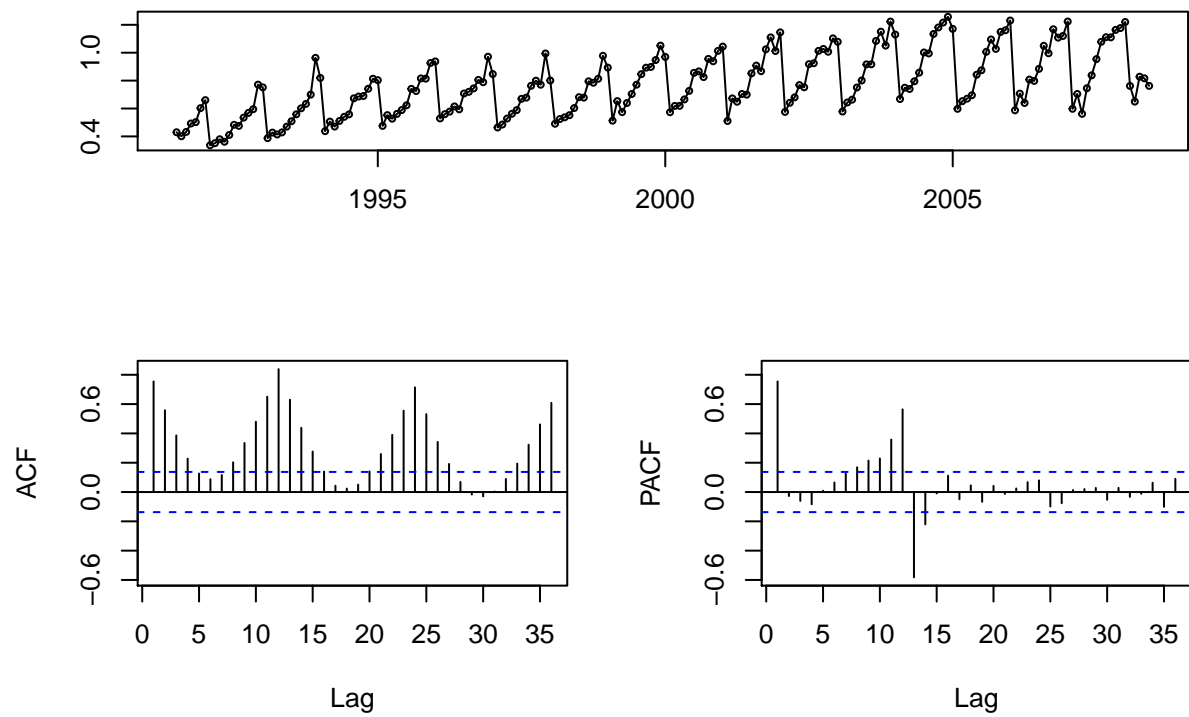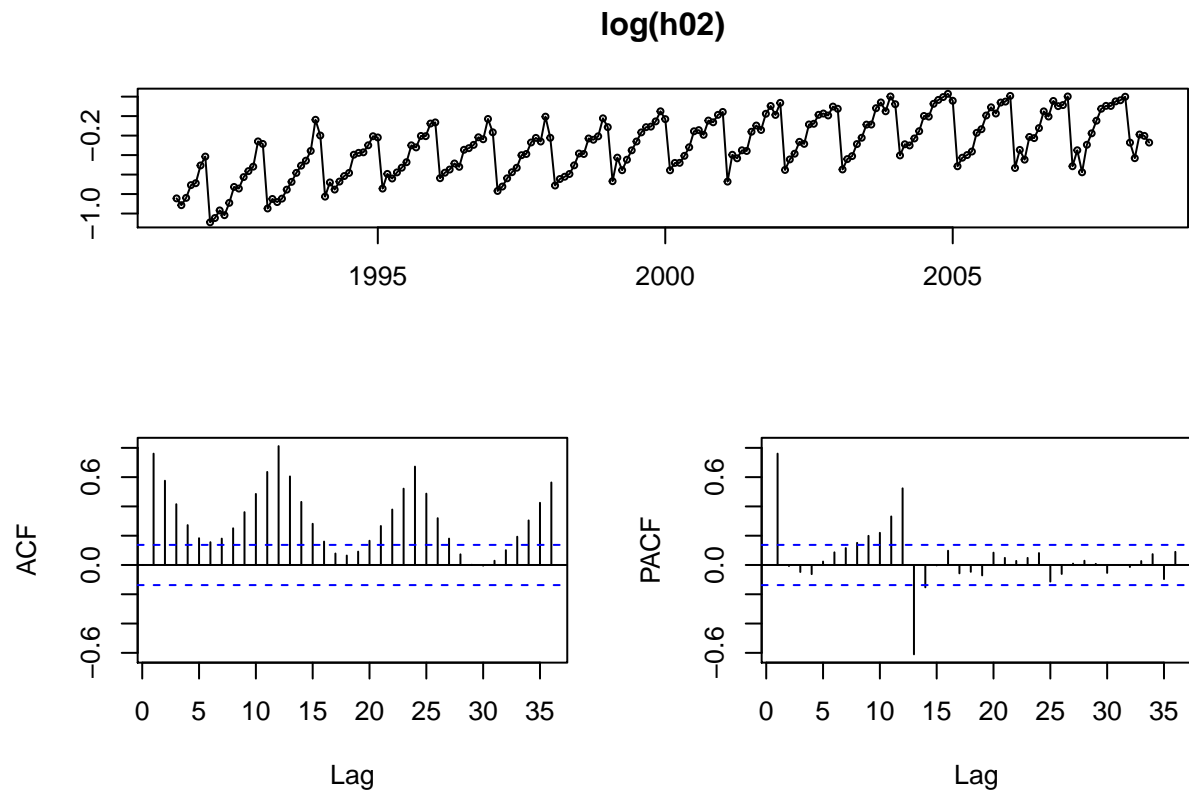
Forecasts from ARIMA(3,1,0)

Problem 2

1.

a.

```
tsdisplay(h02)
```

h02


```
tsdisplay(log(h02))
```
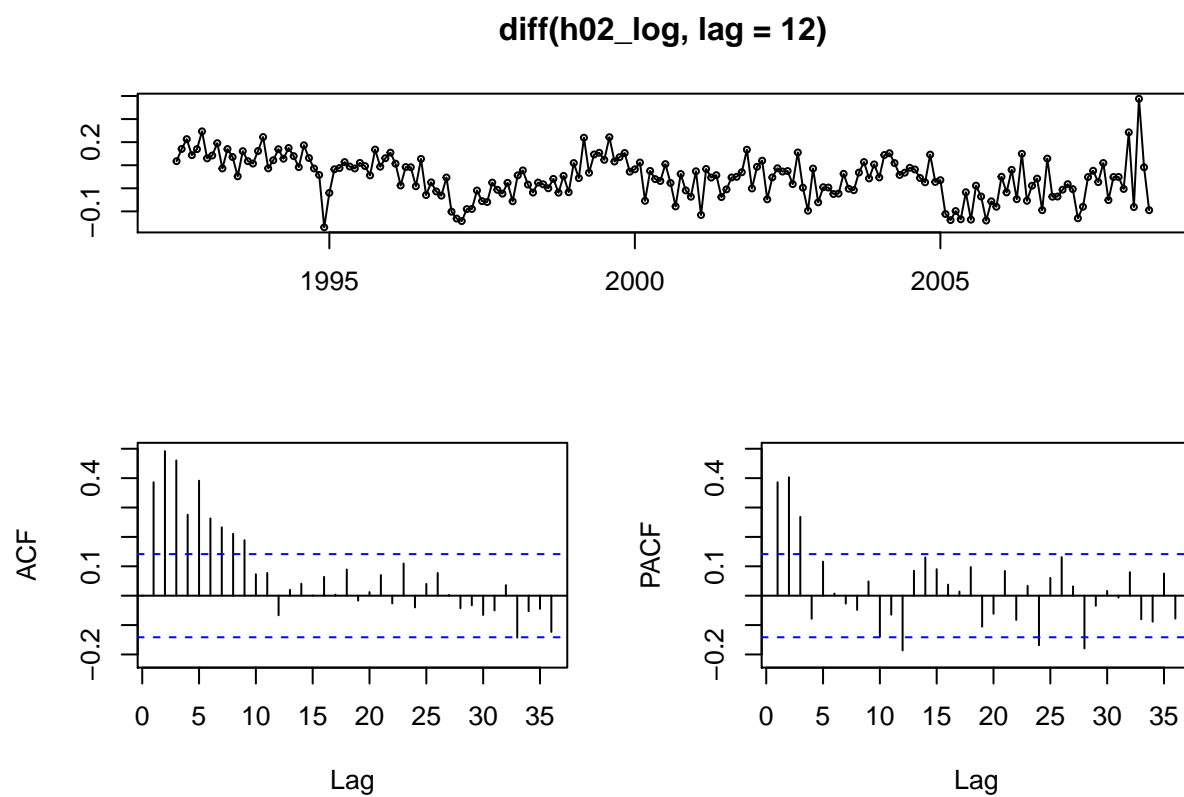
log(h02)


```
#tsdisplay(BoxCox(h02,BoxCox.lambda(h02)))
h02_log <- log(h02)
```
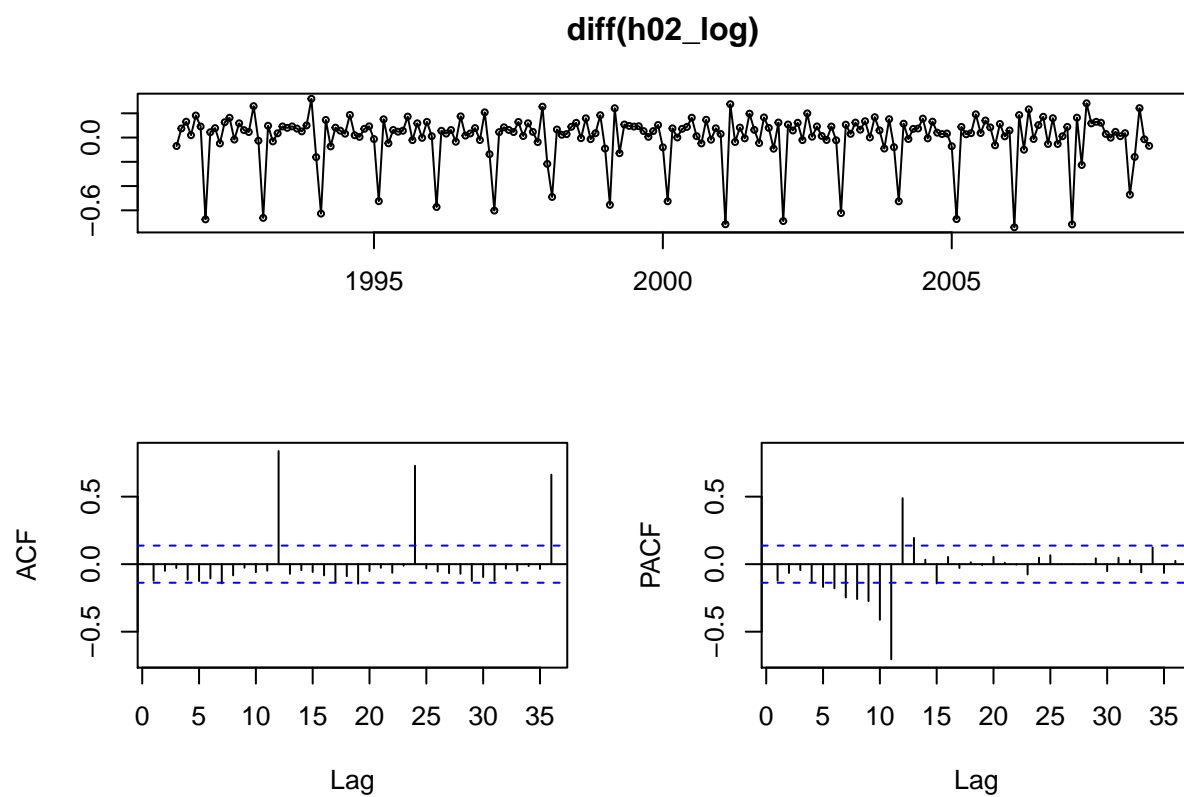
The variance is non-constant in time so a transformation is needed. After comparing the BoxCox estimate and the log transformation, I decided to go with the log transformation.
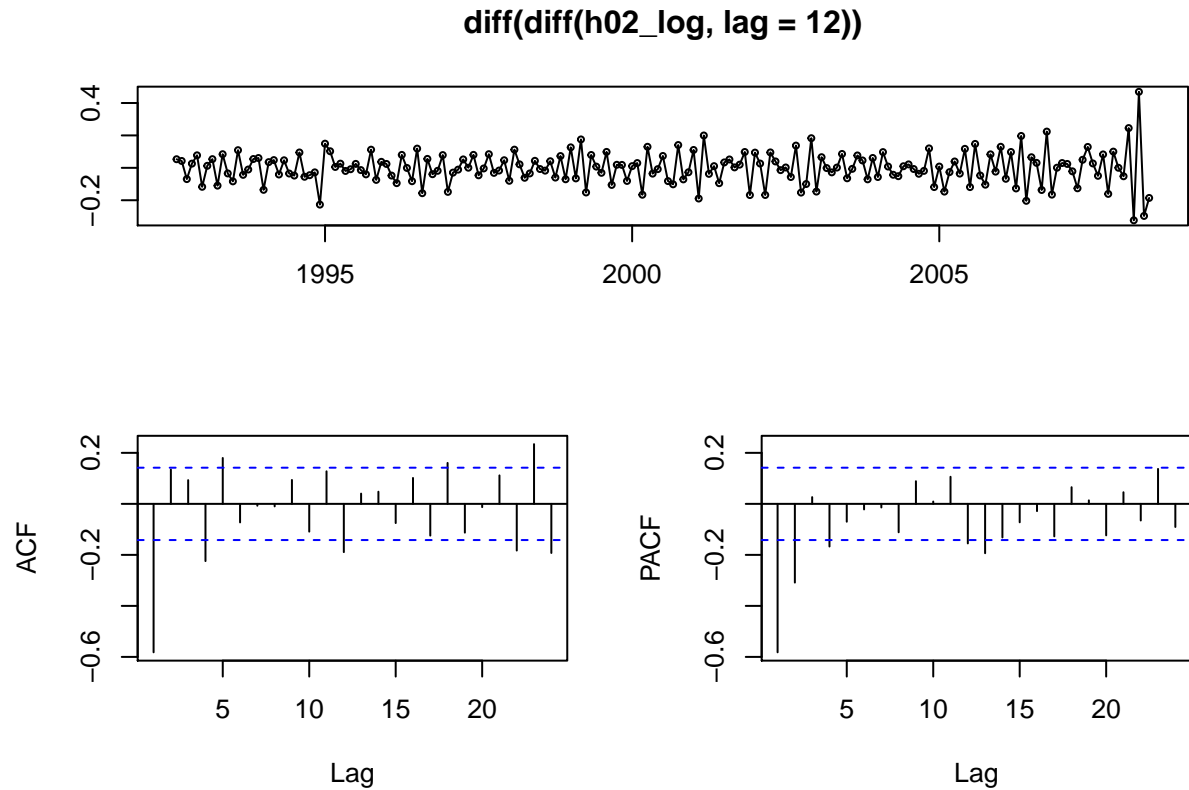
b.

```
tsdisplay(diff(h02_log, lag = 12))
```

diff(h02_log, lag = 12)


```
tsdisplay(diff(h02_log))
```

11

diff(h02_log)


```
tsdisplay(diff(diff(h02_log, lag = 12)), lag.max = 24)
```

diff(diff(h02_log, lag = 12))

Differencing is needed to stabalize the mean. I first attempted seasonal differencing (due to the seasonal trend shown in the ACF plot), but it was still not stable. I continued on to add first-order differencing to the seasonal difference and the time series now has a stable mean.

c.

There does not appear to be any need for the seasonal ARIMA parameters since there is no clear pattern seen at lags 12, 24, etc in the ACF/PACF plots. This could be because the seasonal and first-order differencing took care of it. However, there appears to be a subtle exponenetial decay in the PACF plot, with a spike at lag 1 in the ACF plot. This woul suggest an ARIMA(0,1,1)(0,1,0) model would be appropriate.

d.

Given

$$y'_t = (y_t - y_{t-12}) - y_{t-1}$$

$$y'_t = \mu + \theta\epsilon_{t-1} + \epsilon_t, \ \ \epsilon_t \sim_{i.i.d.} N(0, \sigma^2)$$

```
fit4 <- Arima(h02, lambda = 0, order=c(0,1,1), seasonal=c(0,1,0))
fit4
```

13

```
## Series: h02
## ARIMA(0,1,1)(0,1,0)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1
##       -0.7219
## s.e.   0.0493
##
## sigma^2 estimated as 0.005919:  log likelihood=218.99
## AIC=-433.98   AICc=-433.91   BIC=-427.47
```

$$y'_t = -0.7219\epsilon_{t-1}$$

summary(fit4)

```
## Series: h02
## ARIMA(0,1,1)(0,1,0)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1
##       -0.7219
## s.e.   0.0493
##
## sigma^2 estimated as 0.005919:  log likelihood=218.99
## AIC=-433.98   AICc=-433.91   BIC=-427.47
##
## Training set error measures:
##                         ME       RMSE        MAE        MPE     MAPE      MASE
## Training set -0.001576765 0.06056938 0.04290171 -0.4143405 5.372761 0.7077353
##                    ACF1
## Training set -0.1622259
```
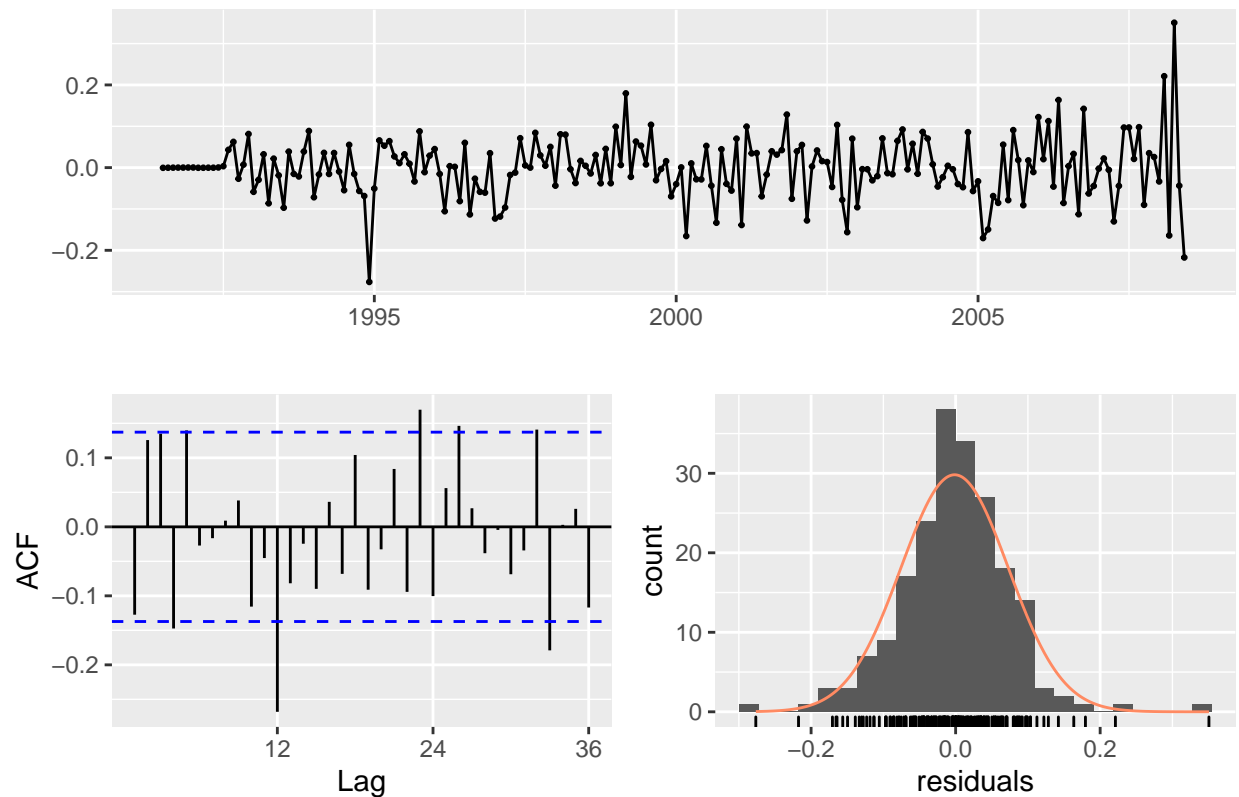
checkresiduals(fit4)

Residuals from ARIMA(0,1,1)(0,1,0)[12]


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)(0,1,0)[12]
## Q* = 60.78, df = 23, p-value = 2.945e-05
##
## Model df: 1.    Total lags used: 24
```

This model is not a great fit. The residuals are normally distributed and the mean is constant, but the ACF is not looking great with several spikes. The Ljung-Box test returned a small p-value, so the residuals are serially correlated (not white noise).

2.

a.

```r
arima.param <- list(c(0,1,1,0,1,0),
                     c(1,1,1,0,1,0),
                     c(1,1,2,0,1,0),
                     c(1,1,1,1,1,0),
                     c(1,1,1,1,1,1),
                     c(1,1,2,1,1,1),
                     c(0,1,2,0,1,0),
```

```
                     c(0,1,2,1,1,1),
                     c(0,1,1,1,1,0),
                     c(0,1,1,0,1,1))

#AICc
arima.aicc <- sapply(arima.param,function(x) Arima(h02, order=x[1:3], seasonal=x[4:6])$aicc)
print(arima.aicc)
```

```
##  [1] -518.7784 -523.5061 -521.4654 -542.1942 -554.3542 -552.2238 -522.5806
##  [8] -552.8773 -535.5129 -552.2006
```

```
print(arima.param[[which.min(arima.aicc)]])
```

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

```
# RMSE for test data
train <- window(h02,start=c(1991,7), end=c(2006,6))
test <- window(h02, start=c(2006,7))
test.rmse <- c()
for (param in arima.param){
  fit <- Arima(train, lambda=0, order=param[1:3], seasonal = param[4:6])
  test.rmse <- c(test.rmse, accuracy(forecast(fit,24), test)[2,"RMSE"])
}

# printing RMSE and best ARIMA
print(test.rmse)
```

```
##  [1] 0.07243801 0.07348127 0.07347407 0.07107656 0.06499490 0.06487137
##  [7] 0.07319178 0.06446402 0.07343985 0.06597533
```

```
print(arima.param[[which.min(test.rmse)]])
```

```
## [1] 0 1 2 1 1 1
```

According to the AICc, the ARIMA(1,1,1)(1,1,1) model was best, while the model with the lowest RMSE was ARIMA(0,1,2)(1,1,1).

b.

```
# ARIMA(1,1,1)(1,1,1)
fit5 <- Arima(h02, lambda = 0, order=c(1,1,1), seasonal = c(1,1,1))
summary(fit5)
```
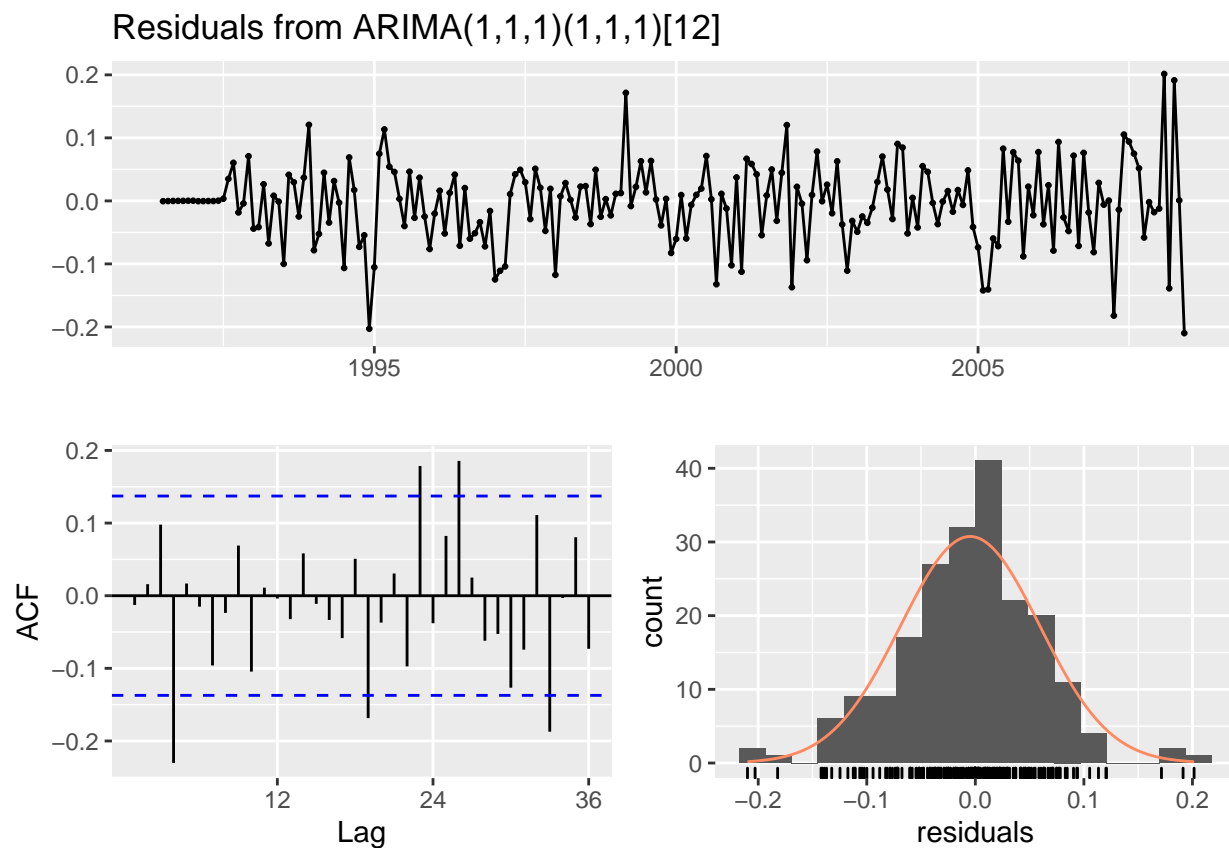
```
## Series: h02
## ARIMA(1,1,1)(1,1,1)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
```

```
##            ar1      ma1     sar1     sma1
##        -0.2758  -0.5533   0.1938  -0.7520
## s.e.    0.1066   0.0988   0.1200   0.0873
##
## sigma^2 estimated as 0.004479:  log likelihood=243.8
## AIC=-477.6   AICc=-477.28   BIC=-461.34
##
## Training set error measures:
##                        ME       RMSE        MAE        MPE      MAPE       MASE
## Training set -0.004304311 0.05227024 0.03786129 -0.6821375 4.787773 0.6245853
##                     ACF1
## Training set -0.03370971
```

```
checkresiduals(fit5)
```

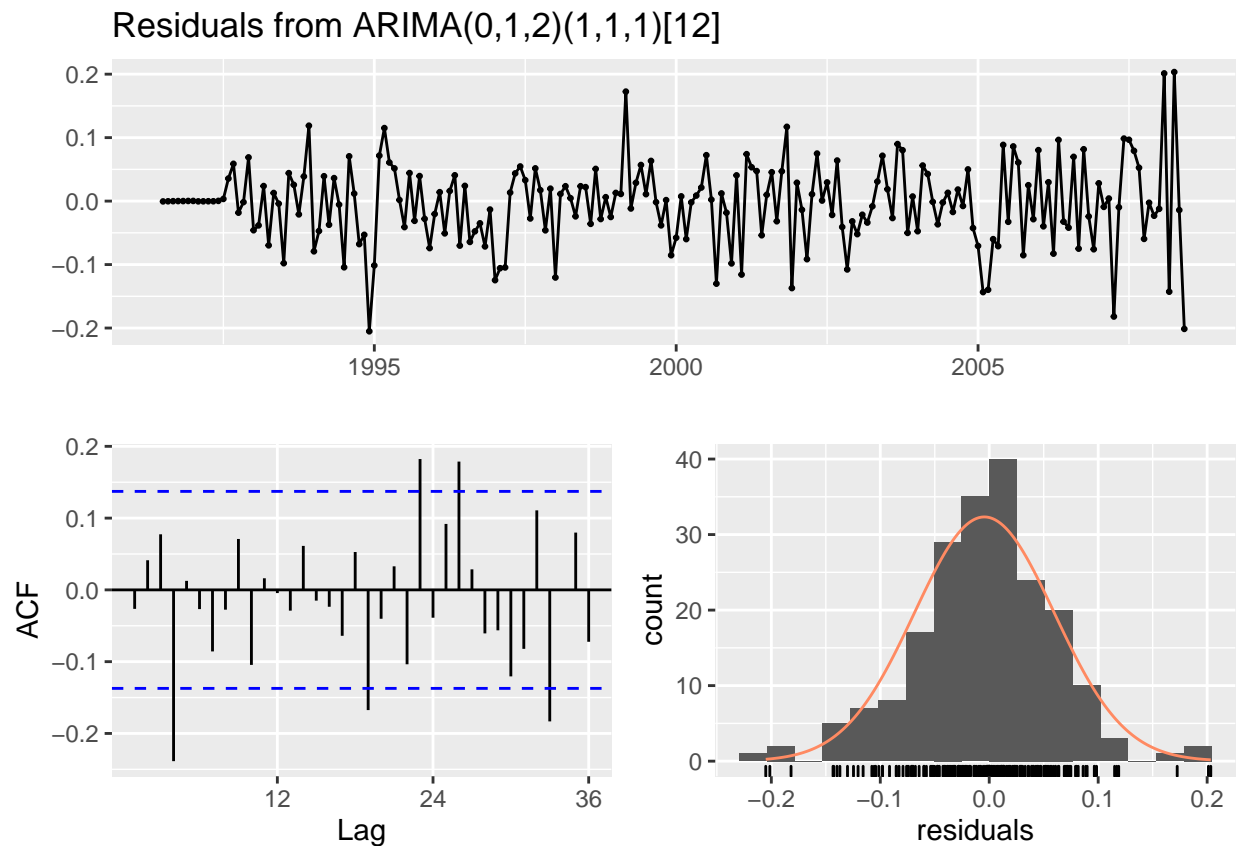## Residuals from ARIMA(1,1,1)(1,1,1)[12]


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(1,1,1)[12]
## Q* = 38.364, df = 20, p-value = 0.00799
##
## Model df: 4.   Total lags used: 24
```

17

```
# ARIMA(0,1,2)(1,1,1)
fit6 <- Arima(h02, lambda = 0, order=c(0,1,2), seasonal = c(1,1,1))
summary(fit6)
```

```
## Series: h02
## ARIMA(0,1,2)(1,1,1)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ma1     ma2     sar1     sma1
##        -0.8079  0.1736  0.1922  -0.7480
## s.e.    0.0722  0.0754  0.1207   0.0886
##
## sigma^2 estimated as 0.004503:  log likelihood=243.35
## AIC=-476.7   AICc=-476.38   BIC=-460.44
##
## Training set error measures:
##                          ME       RMSE        MAE        MPE      MAPE      MASE
## Training set -0.004131724 0.05237381 0.03826365 -0.6565241 4.831516 0.6312228
##                      ACF1
## Training set -0.04676228
```

```
checkresiduals(fit6)
```


Residuals from ARIMA(0,1,2)(1,1,1)[12]

```
##
```

18

```
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,2)(1,1,1)[12]
## Q* = 39.401, df = 20, p-value = 0.005942
##
## Model df: 4.   Total lags used: 24
```

These models are better than the ARIMA(0,1,1)(0,1,0) model, but do not constitute goods fits as their residuals are not white noise by the Ljung-Box test.
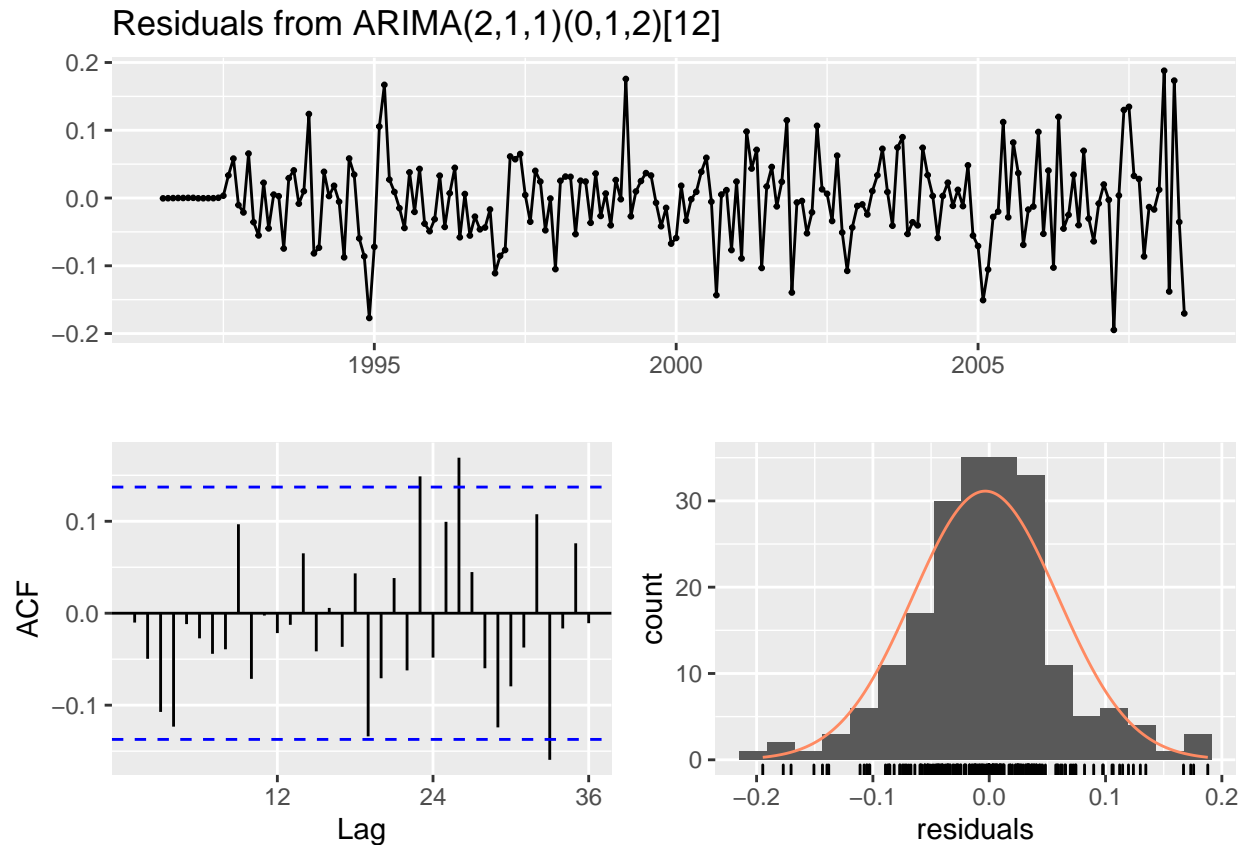
3.

```
fit7 <- auto.arima(h02, lambda = 0)
fit7
```

```
## Series: h02
## ARIMA(2,1,1)(0,1,2)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ar1      ar2     ma1     sma1     sma2
##       -1.1358  -0.5753  0.3683  -0.5318  -0.1817
## s.e.   0.1608   0.0965  0.1884   0.0838   0.0881
##
## sigma^2 estimated as 0.004278:  log likelihood=248.25
## AIC=-484.51   AICc=-484.05   BIC=-465
```

ARIMA(2,1,1)(0,1,2) was selected by auto.arima.

```
checkresiduals(fit7)
```

Residuals from ARIMA(2,1,1)(0,1,2)[12]


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,1)(0,1,2)[12]
## Q* = 24.514, df = 19, p-value = 0.1772
##
## Model df: 5.    Total lags used: 24
```

This model selected does constitute a good fit, as the residuals are normally distributed, the mean and variance are stable, and the residuals are white noise by the Ljung-Box test.

4.

ARIMA(2,1,1)(0,1,2), selected by auto.arima, was the best overall fit. It had one of the lower AICc, but most importantly the residuals were white noise (passed the Ljung-Box test), unlike the other models.

```
autoplot(forecast(fit7, h = 24))
```
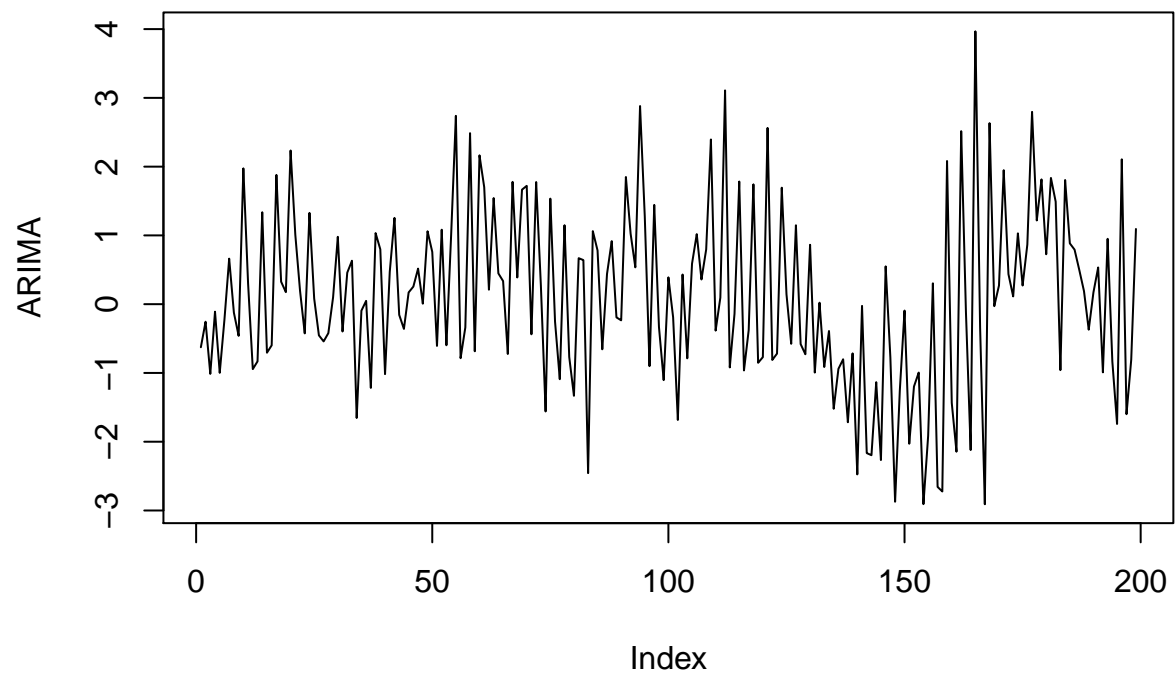
Forecasts from ARIMA(2,1,1)(0,1,2)[12]
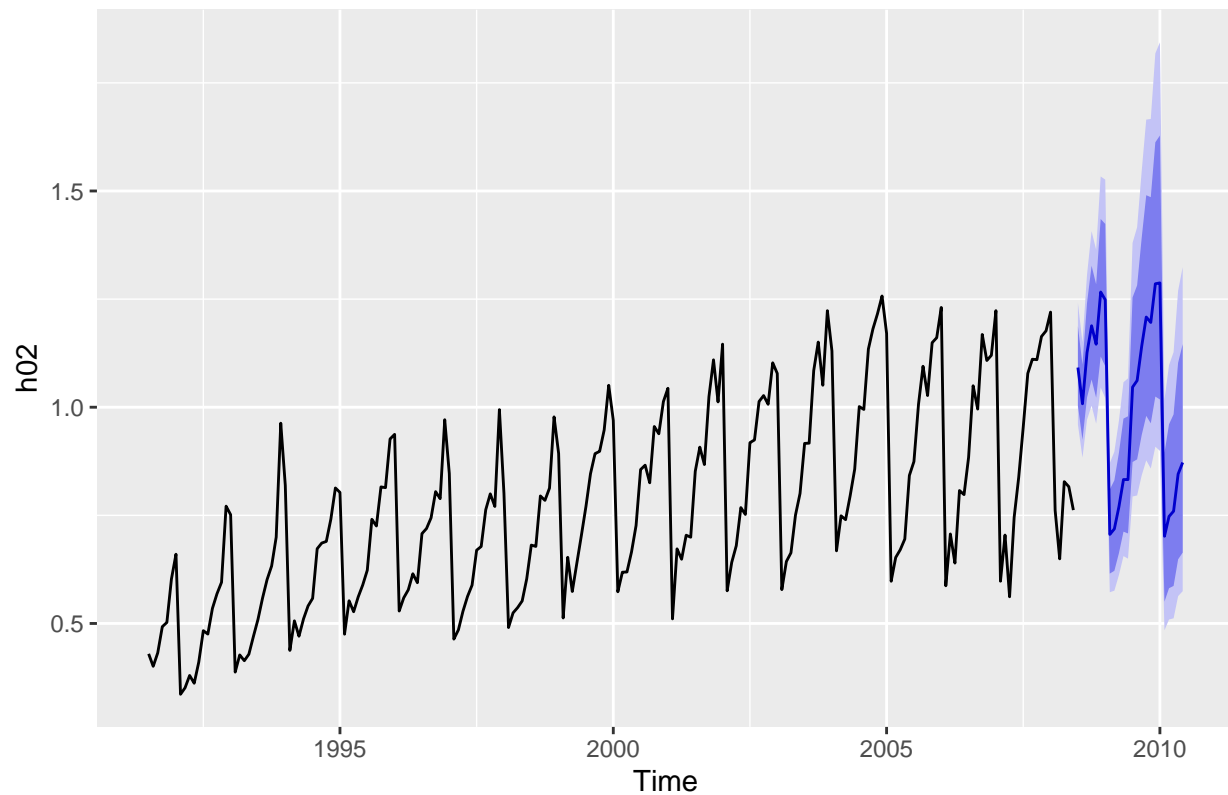
Problem 3

$$y_t = 0.7 y_{t-4} + \epsilon_t, \ \ \epsilon_t \sim_{i.i.d.} N(0,1)$$

```r
ts.length <- 200
set.seed(1)
err <- rnorm(ts.length, sd=1)
ARIMA <- err[1]

for (i in 2:ts.length){
  ARIMA <- c(ARIMA, 0.7*ARIMA[i-4] + err[i])
}

plot(ARIMA, type = "l")
```
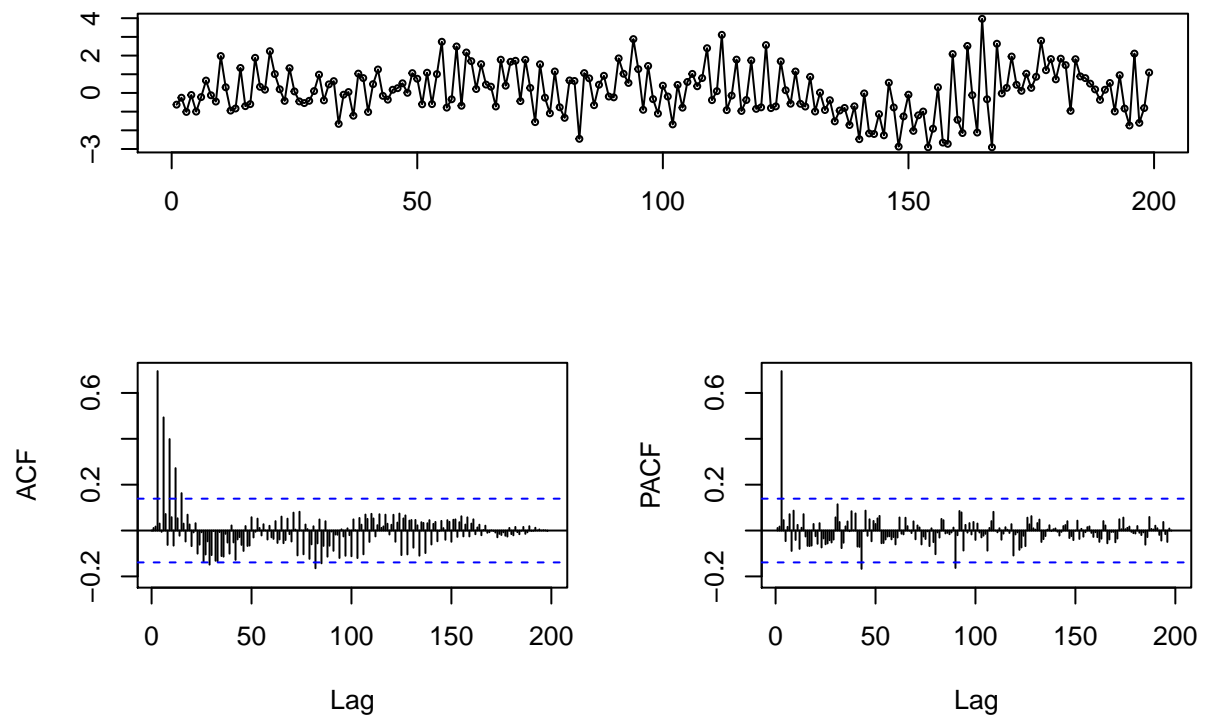
```
tsdisplay(ARIMA, lag.max = 200)
```

ARIMA

The ACF plot is decaying while the PACF has a spike at lag 4. This is as expected from the way we generated the quarterly data with ARIMA(0,0,0)(1,0,0)