W271 Lab3 Submission

Ashesh Choudhury

Fall 2018

```
# Insert the function to *tidy up* the code when they are printed out
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)

# Clean up the workspace before we begin
rm(list = ls())

# Set working directory
wd <- "C:/Users/ashesh/Desktop/Data science prep/MIDS/MIDS Study material/W27
1/Lab3"
setwd(wd)
# Load libraries
install.packages("xts", repos="http://cloud.r-project.org")

## Installing package into 'C:/Users/ashesh/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## package 'xts' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\ashesh\AppData\Local\Temp\RtmpK4eYVJ\downloaded_packages

library(xts)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

library(forecast)
library(astsa)

##
## Attaching package: 'astsa'

## The following object is masked from 'package:forecast':
##
##     gas

library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:xts':
##
##     first, last

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(Hmisc)

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units

df <- read.csv("https://raw.githubusercontent.com/MIDS-W271/main-f18/e5dafd9a
feb57f1f862f5561a3fcb535480b364c/labs/lab3/ECOMPCTNSA.csv?token=AhtFYcA1jdroQ
1uy4ywPurn4YdW0mXdQks5cAcCUwA%3D%3D",
    header = TRUE, sep = ",")

str(df)

## 'data.frame':    69 obs. of  2 variables:
##  $ DATE      : Factor w/ 69 levels "1999-10-01","2000-01-01",..: 1 2 3 4 5
6 7 8 9 10 ...
##  $ ECOMPCTNSA: num  0.7 0.8 0.8 0.9 1.1 1.1 1 1 1.3 1.3 ...

head(df)

##         DATE ECOMPCTNSA
## 1 1999-10-01        0.7
## 2 2000-01-01        0.8
```

```
## 3 2000-04-01        0.8
## 4 2000-07-01        0.9
## 5 2000-10-01        1.1
## 6 2001-01-01        1.1
```

```r
describe(df$ECOMPCTNSA)
```

```
## df$ECOMPCTNSA
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##       69        0       50        1    3.835    2.524     0.94     1.10
##      .25      .50      .75      .90      .95
##     2.00     3.60     5.30     6.92     7.70
##
## lowest : 0.7 0.8 0.9 1.0 1.1, highest: 7.0 7.5 7.7 8.7 9.5
```

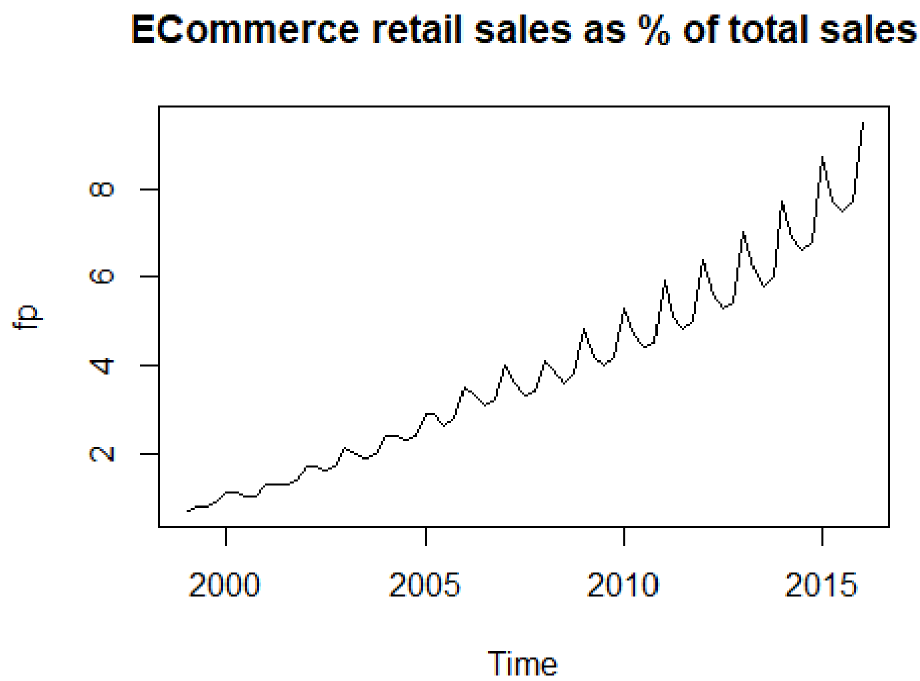
```r
# Create an R time-series object
fp <- ts(df$ECOMPCTNSA, frequency = 4, start = c(1999, 1))
str(fp)
```

```
##  Time-Series [1:69] from 1999 to 2016: 0.7 0.8 0.8 0.9 1.1 1.1 1 1 1.3 1.3
...
```

```r
head(fp)
```

```
##      Qtr1 Qtr2 Qtr3 Qtr4
## 1999  0.7  0.8  0.8  0.9
## 2000  1.1  1.1
```

```r
plot.ts(fp, main = "ECommerce retail sales as % of total sales")
```

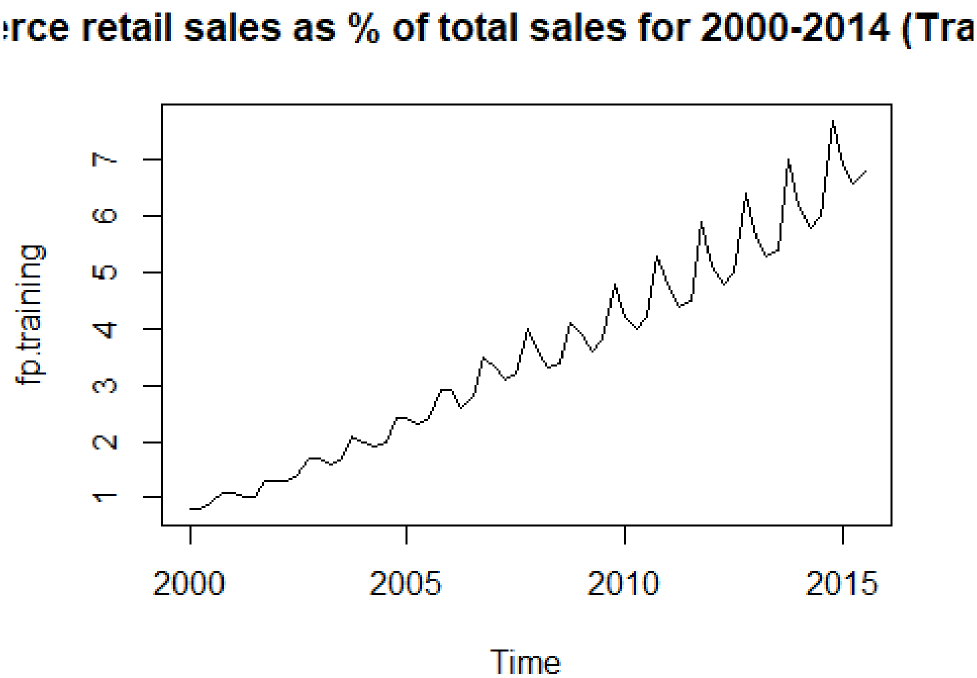



```
# Lets keep data between 2015 and 2016. Let's hold out 2015
# as test data that you can use later.

fp.training <- fp[time(fp) > 1999 & time(fp) < 2015]
fp.training <- ts(fp.training, frequency = 4, start = c(2000,
    1))
str(fp.training)

##  Time-Series [1:63] from 2000 to 2016: 0.8 0.8 0.9 1.1 1.1 1 1 1.3 1.3 1.3
...

head(fp.training, 10)

##      Qtr1 Qtr2 Qtr3 Qtr4
## 2000  0.8  0.8  0.9  1.1
## 2001  1.1  1.0  1.0  1.3
## 2002  1.3  1.3

fp.test <- fp[time(fp) >= 2015]
fp.test <- ts(fp.test, frequency = 4, start = c(2015, 1))
str(fp.test)

##  Time-Series [1:5] from 2015 to 2016: 8.7 7.7 7.5 7.7 9.5

head(fp.test, 10)
```

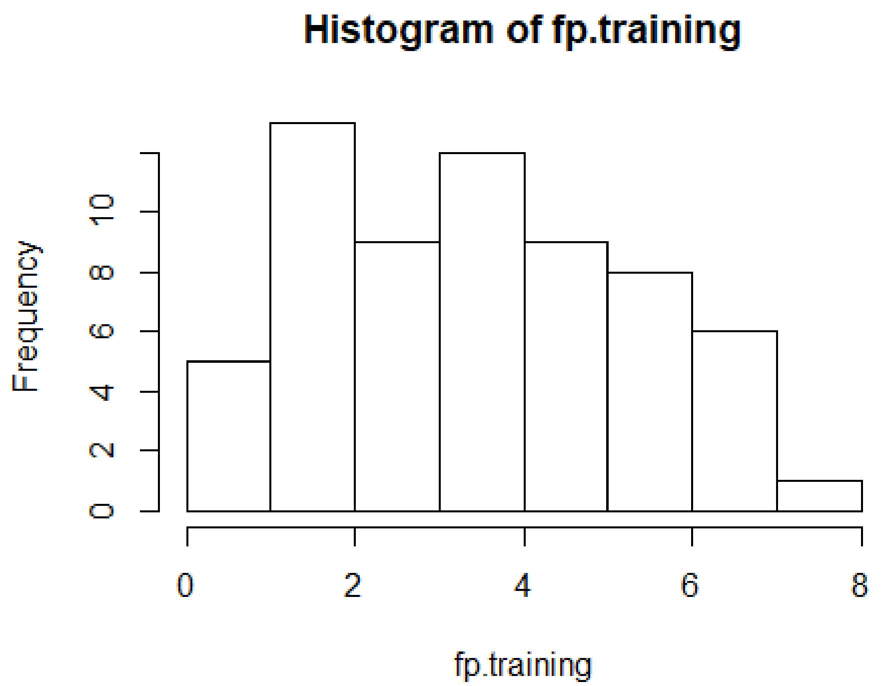
```
##      Qtr1 Qtr2 Qtr3 Qtr4
## 2015  8.7  7.7  7.5  7.7
## 2016  9.5
```

As we can see below, this time-series is clearly not stationary in the mean, and it is also pretty apparent that the time-series exhibits a lot of seasonality.

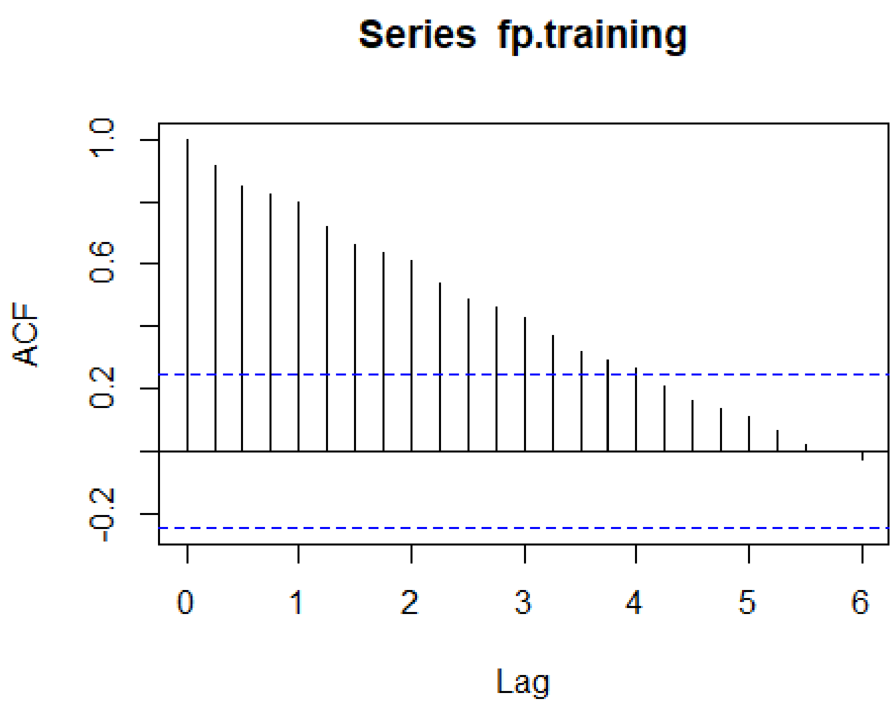
```r
plot(fp.training, main = "ECommerce retail sales as % of total sales for 2000
-2014 (Training Series)")
```



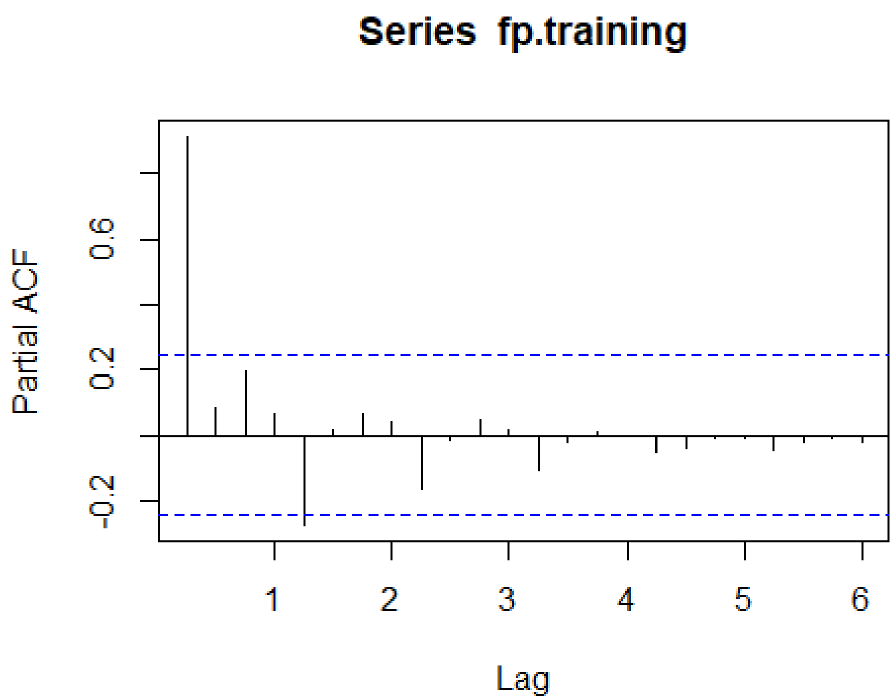

```r
hist(fp.training)
```



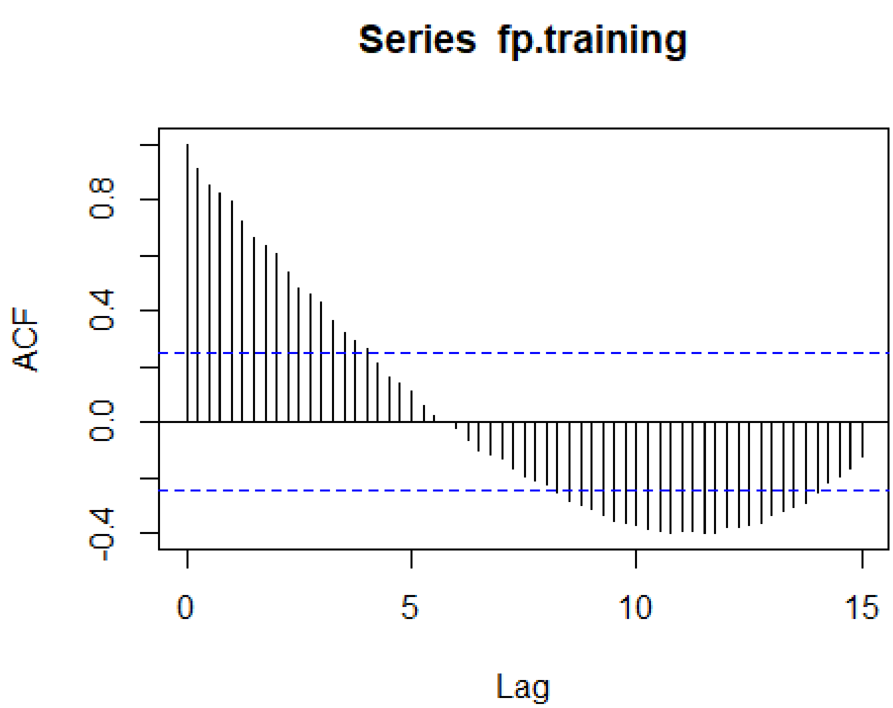

```r
acf(fp.training, lag.max = 24)
```



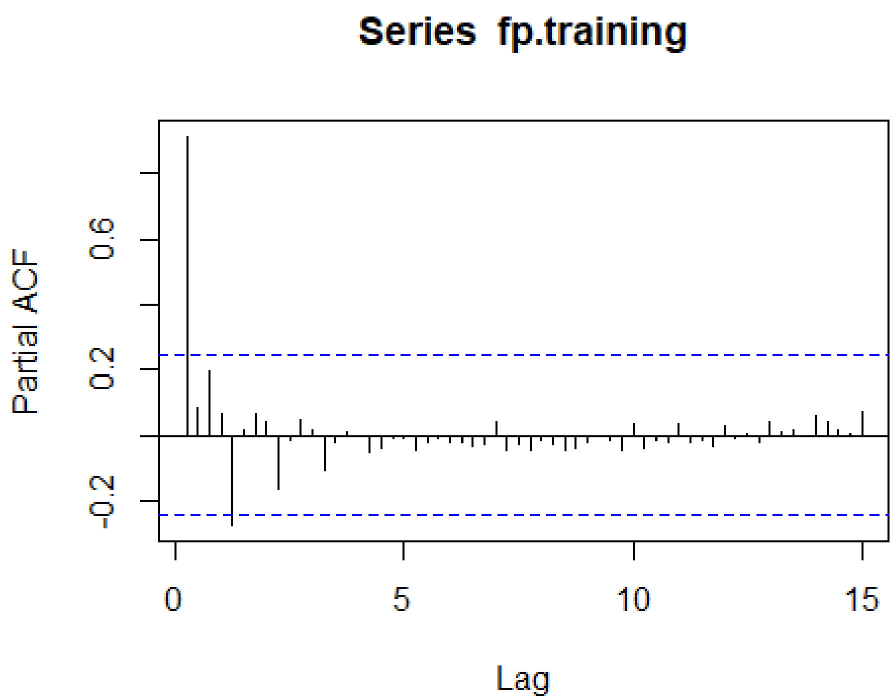

```r
pacf(fp.training, lag.max = 24)
```



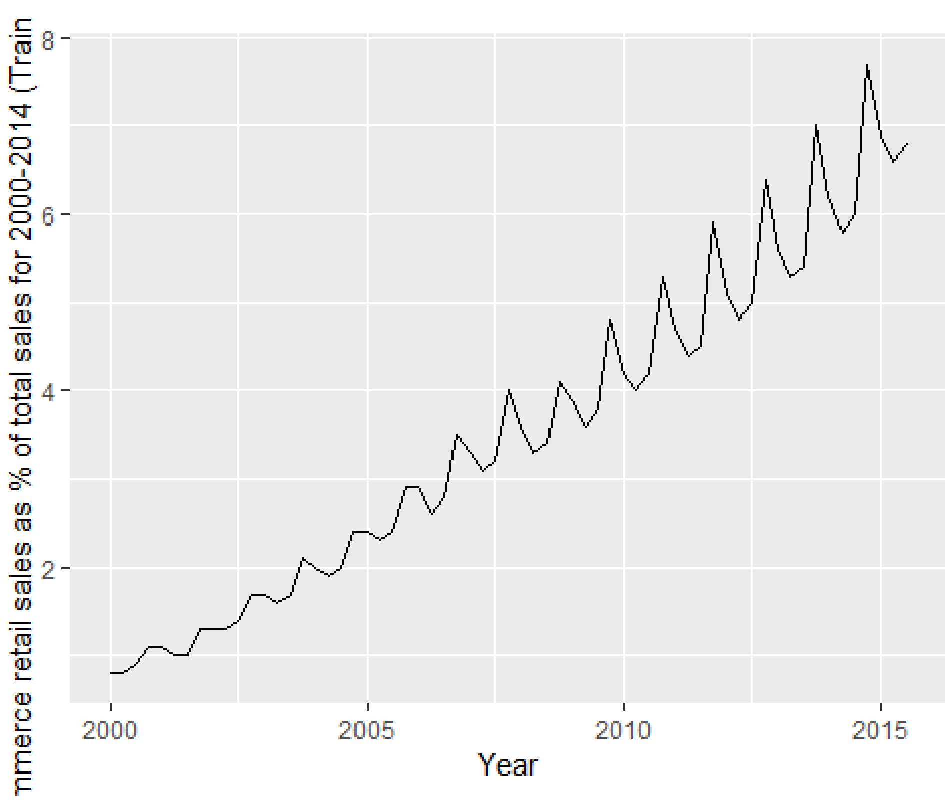

```r
acf(fp.training, lag.max = 60)
```



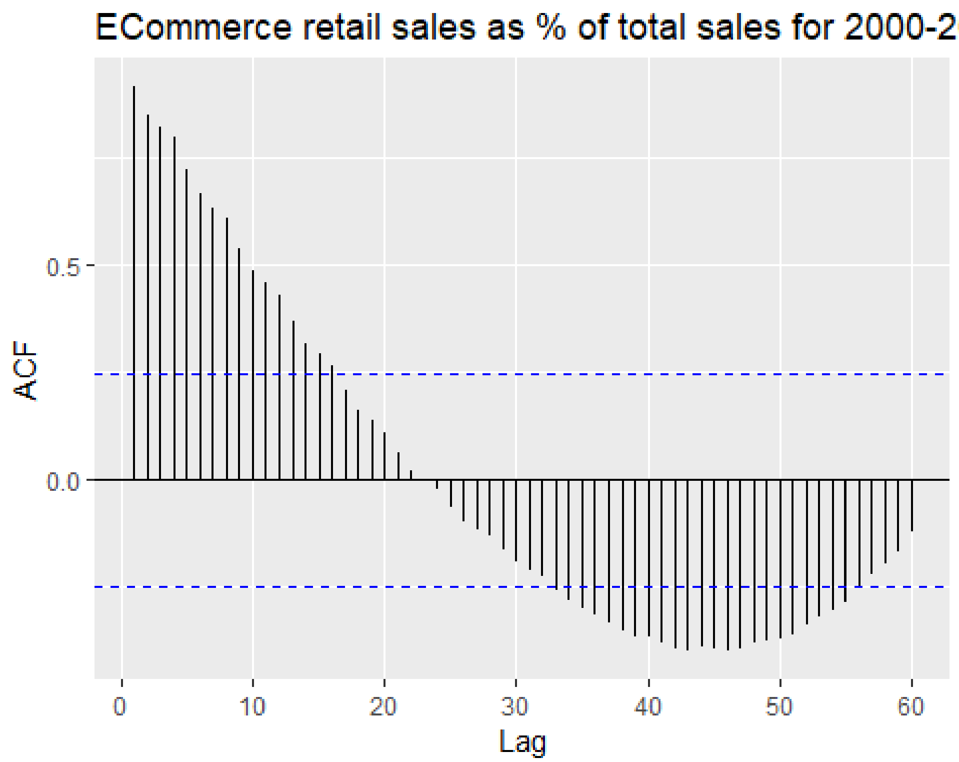

```r
pacf(fp.training, lag.max = 60)
```



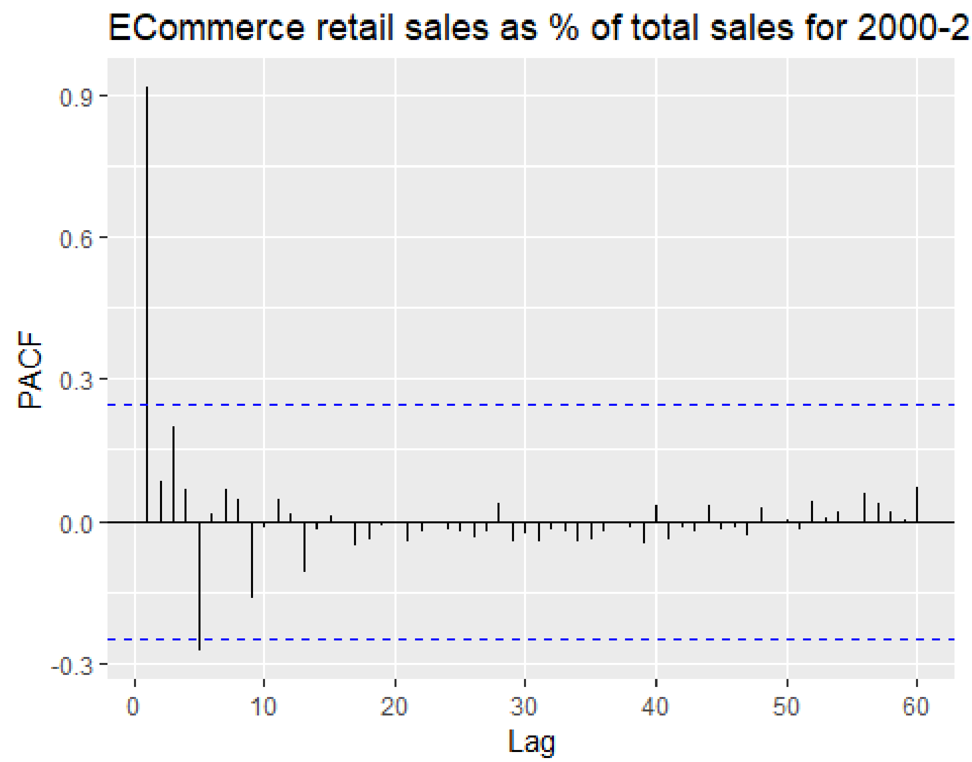

```r
# We can also use autoplot, ggAcf, and ggPacf
autoplot(fp.training) + xlab("Year") + ylab("ECommerce retail sales as % of t
otal sales for 2000-2014 (Training Series)")
```



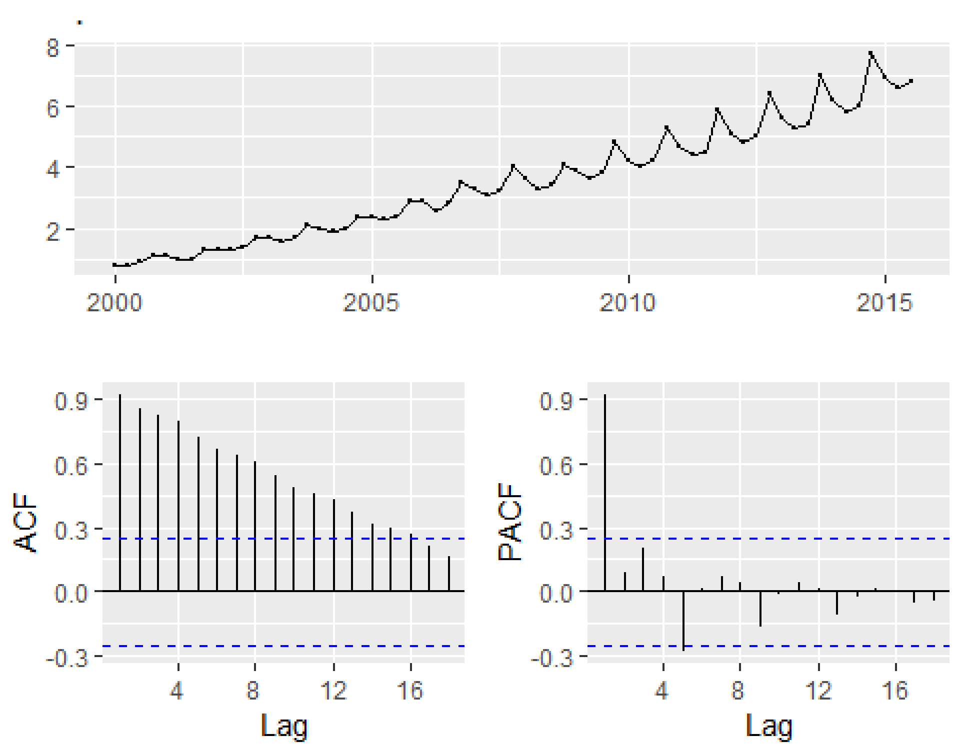
```r
ggAcf(fp.training, lag.max = 60, main = "ECommerce retail sales as % of total
sales for 2000-2014 (Training Series)")
```



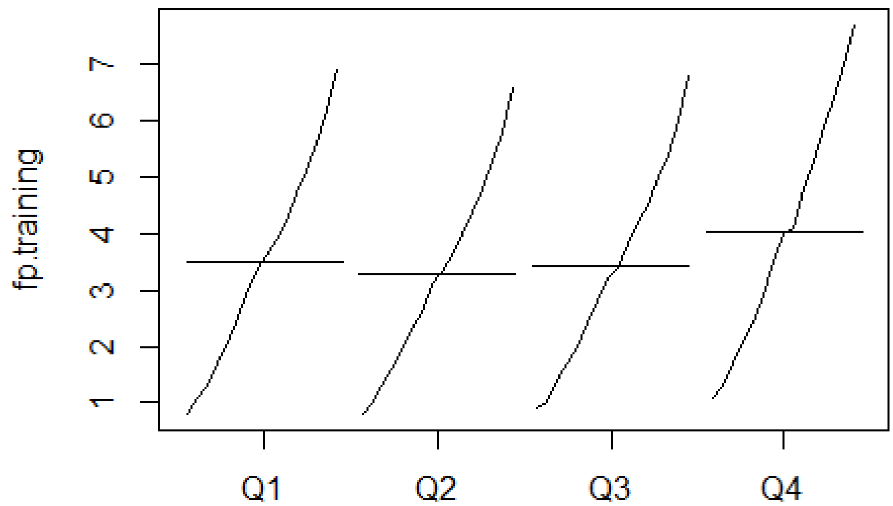

```r
ggPacf(fp.training, lag.max = 60, main = "ECommerce retail sales as % of tota
l sales for 2000-2014 (Training Series)")
```




```r
# Another (more concise) way is to use ggtsdisplay()
fp.training %>% ggtsdisplay()
```

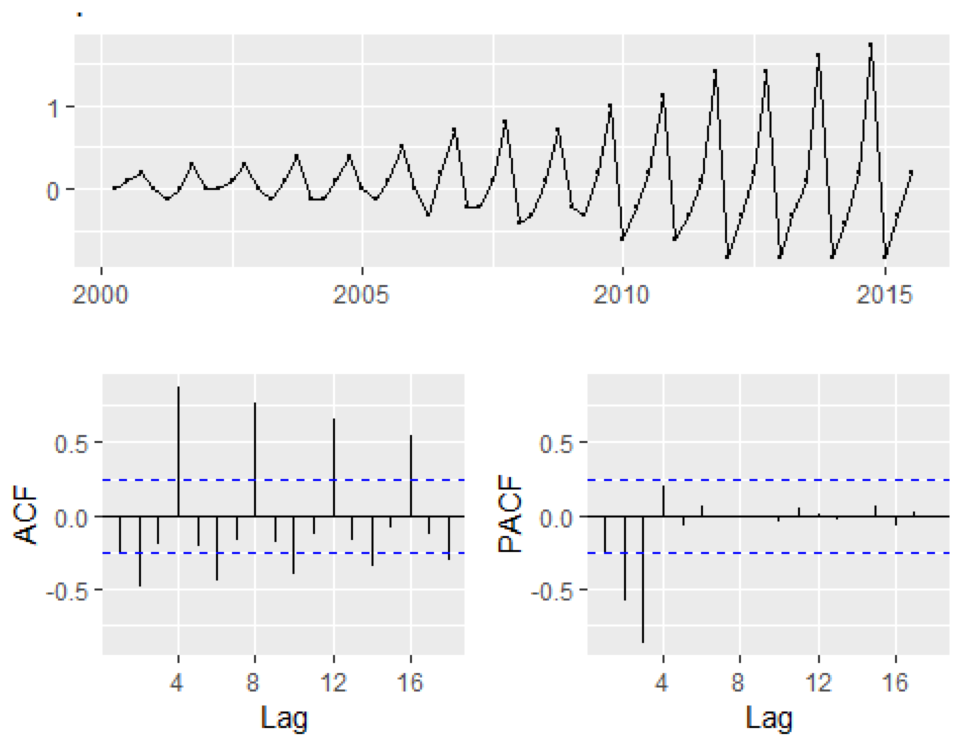


```r
monthplot(fp.training)
```

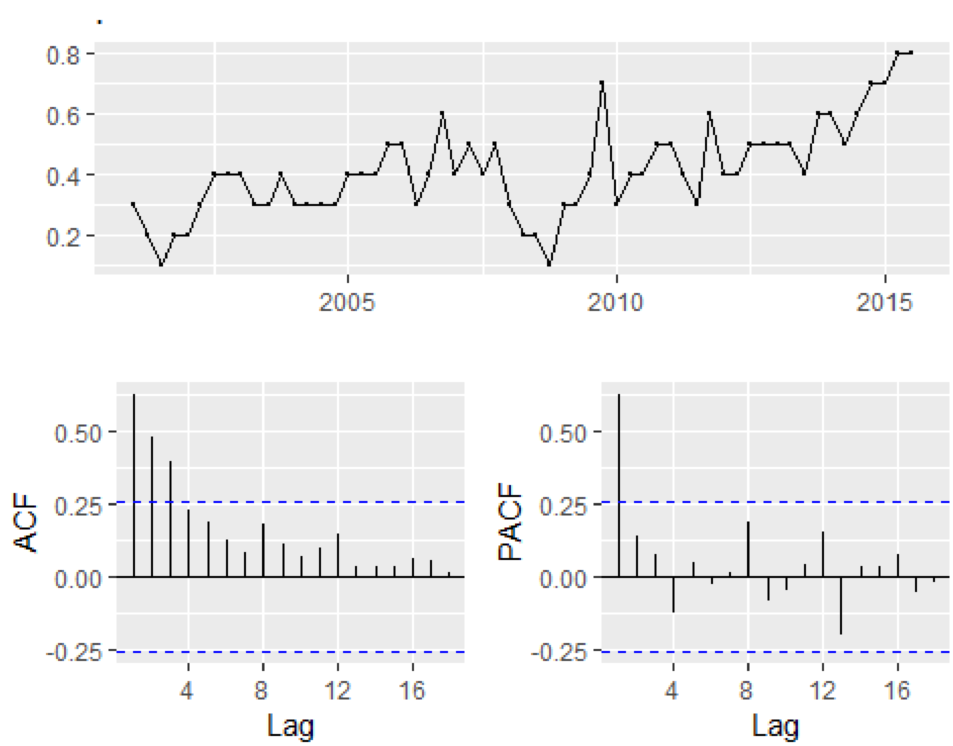


Let’s examining some differencing-transformation of the series: - seasonal differencing - non-seasonal differencing - non-seasonal differencing on top of seasonal differencing

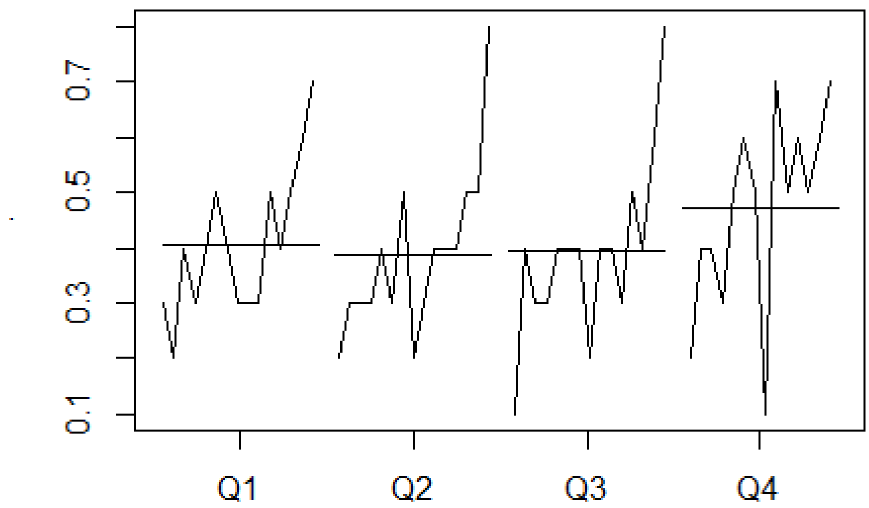
```r
fp.training %>% diff(lag = 1) %>% ggtsdisplay()
```



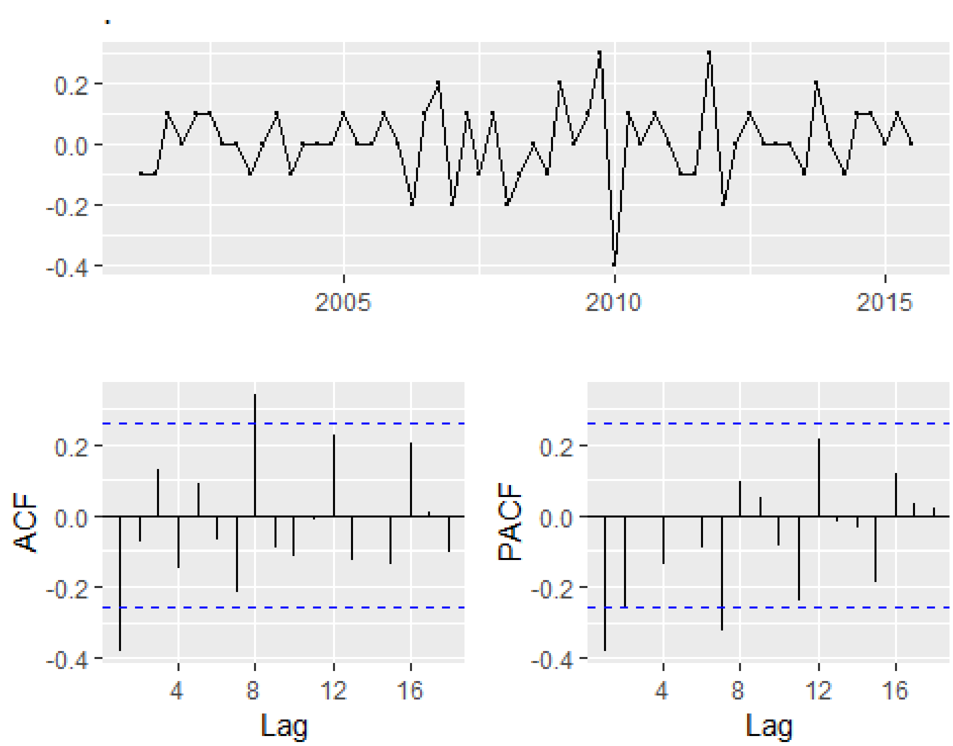

```
fp.training %>% diff(lag = 4) %>% ggtsdisplay()
```



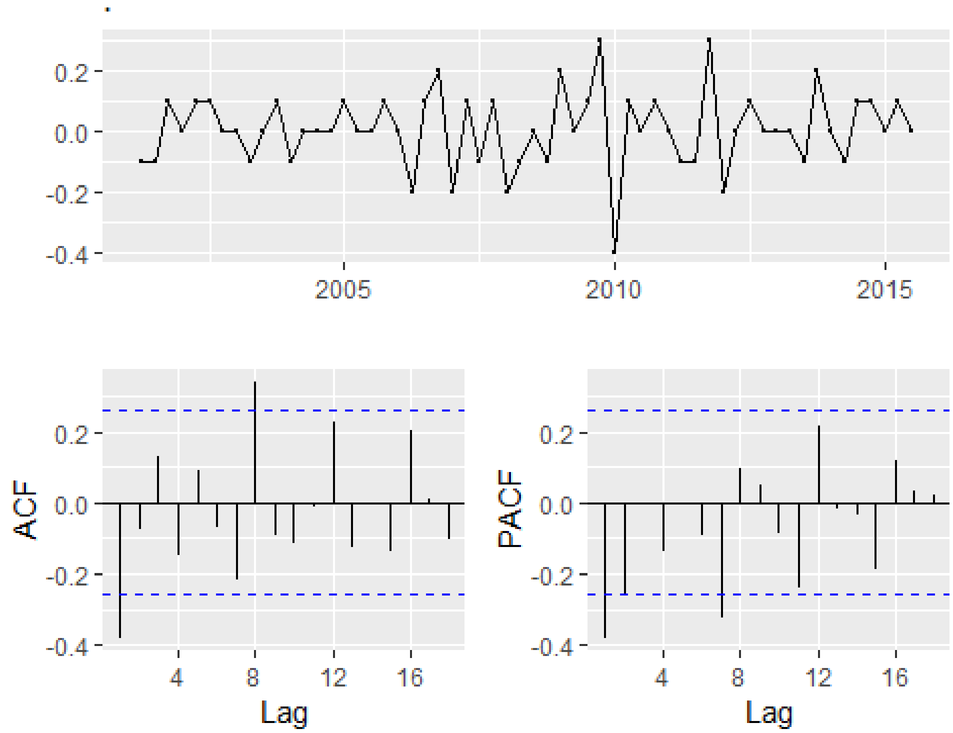

```
fp.training %>% diff(lag = 4) %>% monthplot()
```




```
fp.training %>% diff(lag = 1) %>% diff(lag = 4) %>% ggtsdisplay()
```




```
fp.training %>% diff(lag = 4) %>% diff(lag = 1) %>% ggtsdisplay()
```



Modeling the non-seasonal component

First, let's model the non-seasonal component of the raw series. In order to do that, we are going to use the `Arima` function in the forecast package. I am making the extra steps of modeling the non-seasonal component as pure AR and MA processes first, for illustrative purposes. Based on the ACF and PACF charts, I expect that we can model the non-seasonal component with an ARIMA(0,1,1) or ARIMA(0,1,2).

```
# Let's start by modeling it as a pure AR process
for (p in 0:5) {
    mod <- Arima(fp.training, order = c(p, 0, 0), seasonal = list(order = c(0
,
        0, 0), 4), method = "ML")
    print(c(p, mod$aic, mod$bic))
}

## [1]   0.0000 260.6982 264.9845
## [1]   1.0000 109.4732 115.9026
## [1]   2.0000 111.4936 120.0661
## [1]   3.00000  95.83995 106.55562
## [1]  4.00000 72.91859 85.77740
## [1]   5.00000 -70.61632 -55.61438
```

The AIC is minimized when p = 5 and BIC is minimized when p = 5. Let's examine the residuals of each model.

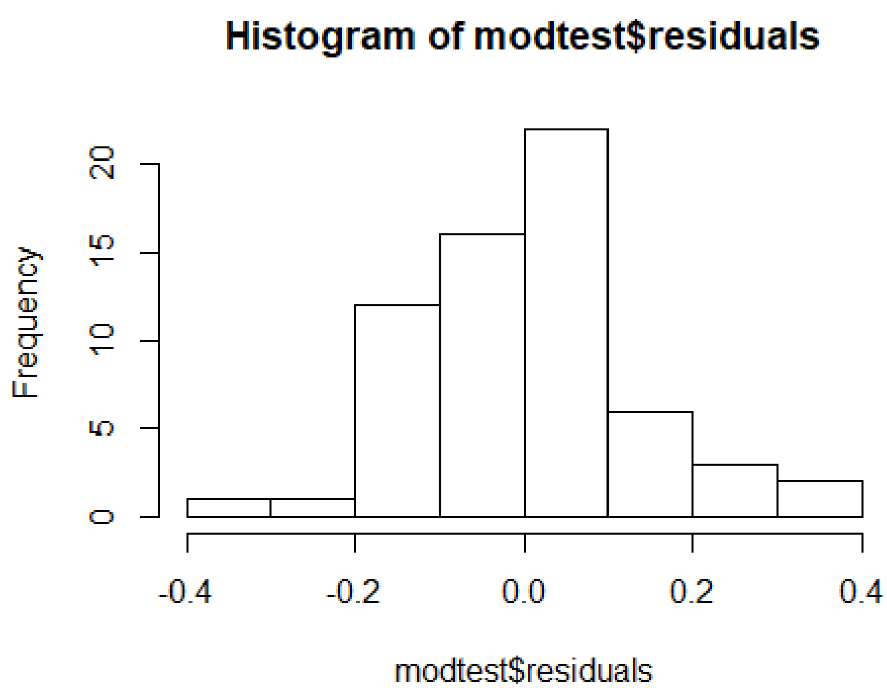
```
modtest <- Arima(fp.training, order = c(5, 0, 0), seasonal = list(order = c(0
,
    0, 0), 4), method = "ML")
modtest
```

```
## Series: fp.training
## ARIMA(5,0,0) with non-zero mean
##
## Coefficients:
```

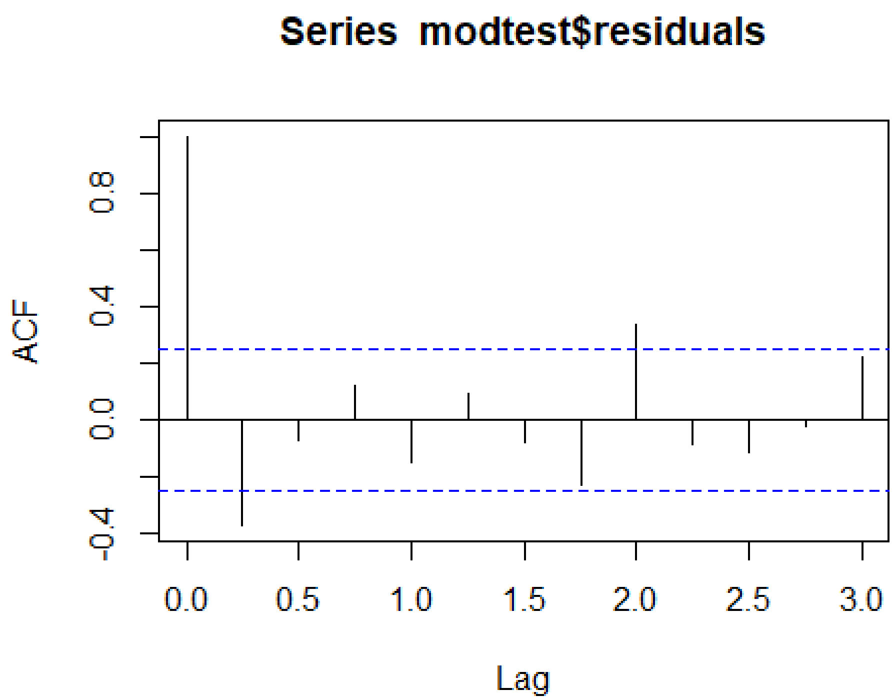
```
## Warning in sqrt(diag(x$var.coef)): NaNs produced
```

```
##            ar1     ar2     ar3     ar4  ar5    mean
##         0.9917  0.0083  0.0083  0.9917   -1  3.7476
## s.e.    0.0165  0.0165  0.0165  0.0165    0     NaN
##
## sigma^2 estimated as 0.01555:  log likelihood=42.31
## AIC=-70.62   AICc=-68.58   BIC=-55.61
```

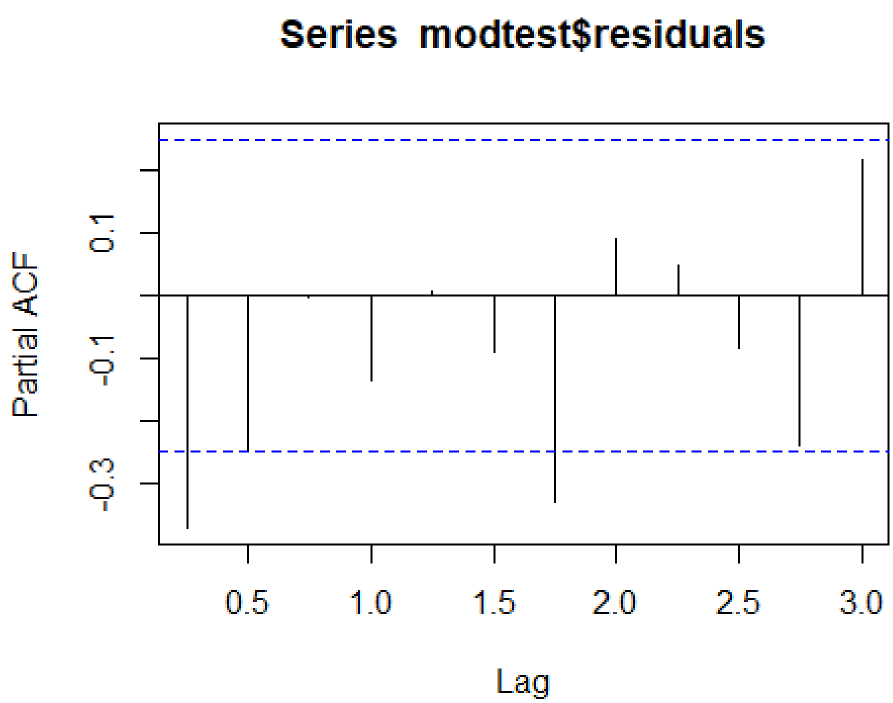
```
hist(modtest$residuals)
```




```
acf(modtest$residuals, lag.max = 12)
```



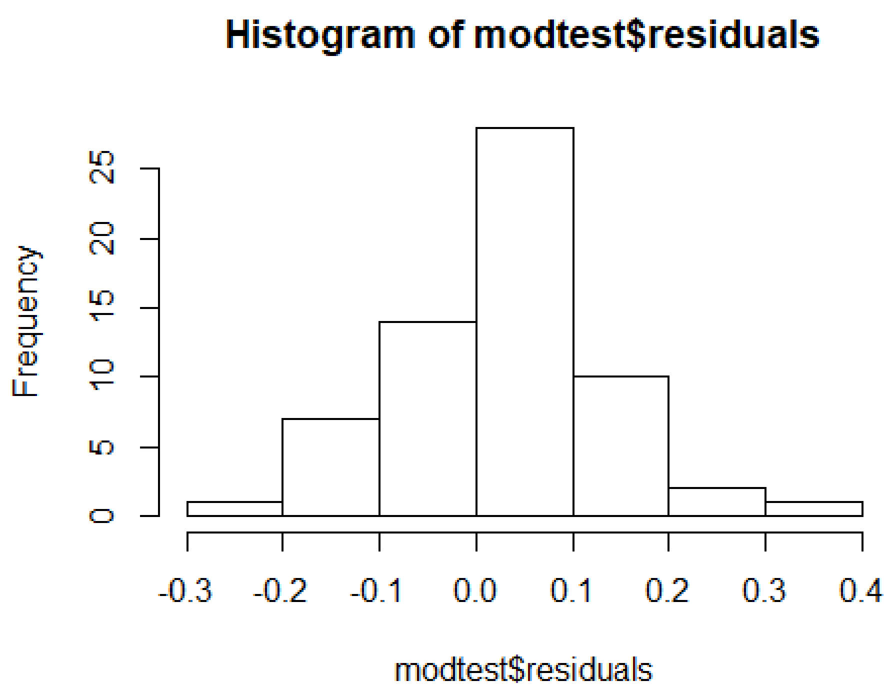

```
pacf(modtest$residuals, lag.max = 12)
```



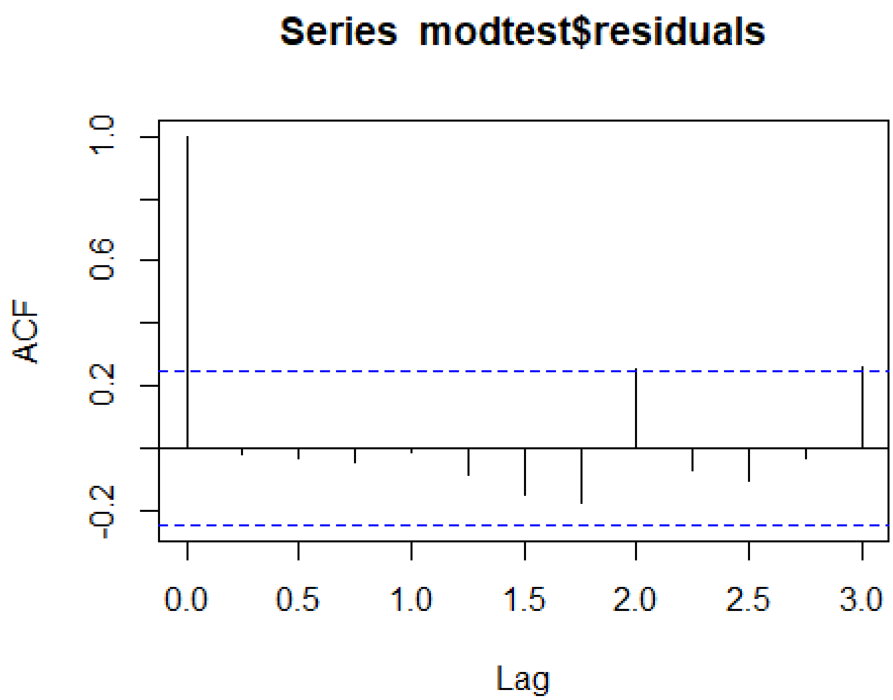
```
modtest <- Arima(fp.training, order = c(5, 0, 0), seasonal = list(order = c(0
,
    1, 0), 4), method = "ML")
modtest
```

```
## Series: fp.training
## ARIMA(5,0,0)(0,1,0)[4]
##
## Coefficients:
##          ar1     ar2     ar3      ar4     ar5
##       0.5364  0.1869  0.2274  -0.0771  0.1151
## s.e.  0.1306  0.1478  0.1466   0.1463  0.1307
##
## sigma^2 estimated as 0.0133:  log likelihood=44.57
## AIC=-77.15   AICc=-75.53   BIC=-64.68
```

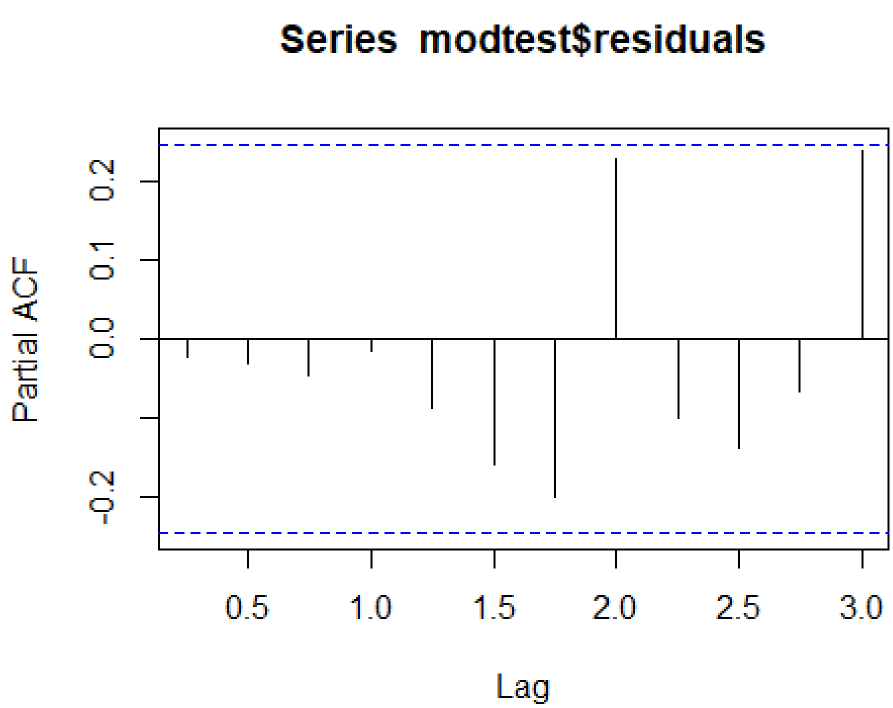
```
hist(modtest$residuals)
```




```
acf(modtest$residuals, lag.max = 12)
```




```r
pacf(modtest$residuals, lag.max = 12)
```



Both models do a decent job of eliminating lower-ordered correlated residuals, though they do not eliminate any of the seasonality, which is to be expected given that we have taken no steps to model the seasonal component (yet)!

Now, let's model the data as a pure MA process.

```
for (q in 0:5) {
    mod <- Arima(fp.training, order = c(0, 1, q), seasonal = list(order = c(0
,
        1, 0), 4), method = "ML")
    print(c(q, mod$aic, mod$bic))
}

## [1]   0.00000 -75.66779 -73.60734
## [1]   1.00000 -85.53246 -81.41158
## [1]   2.00000 -83.67860 -77.49727
## [1]   3.00000 -81.69831 -73.45654
## [1]   4.00000 -81.13321 -70.83099
## [1]   5.00000 -79.49482 -67.13216
```

Both the AIC and BIC are minimized when q = 1. Note that the AIC and BIC values for an ARIMA(0,1,1) are lower than any of the pure AR models we examined above. Again, this is consistent with what we expected given our visual examination of the ACF and PACF charts.

Now, we need to find an appriorate ARIMA(p,1,q) model. Based on the principle of parsimony, I expect that p = 0 and q = 1 (p + q < min(p',q') where p' and q' are the orders of a pure AR and MA process respectively.)

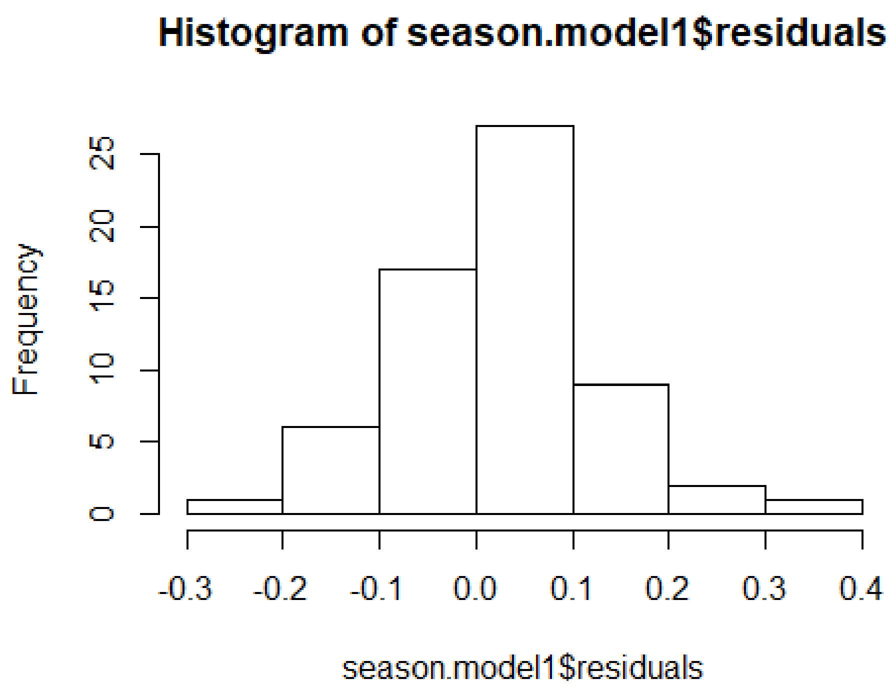
An ARIMA(0,1,1) does a decent job of removing dependency in the non-seasonal component of the data. We might choose to explore a more complicated model to see if does an even better job of generating well behaved residuals. As you can see below, the residuals look very similar to those generated by the simpler model.

```
for (P in 0:1) {
    for (Q in 0:1) {
        mod <- Arima(fp.training, order = c(0, 0, 1), seasonal = list(order =
c(P,
            0, Q), 4), method = "ML")

        print(c(P, Q, mod$aic, mod$bic))
    }
}

## [1]   0.0000   0.0000 207.9311 214.3605
## [1]   0.0000   1.0000 146.5273 155.0999
## [1]  1.00000  0.00000 39.38736 47.95990
## [1]  1.00000  1.00000 14.46116 25.17683

season.model1 <- Arima(fp.training, order = c(0, 1, 1), seasonal = list(order
= c(0,
```

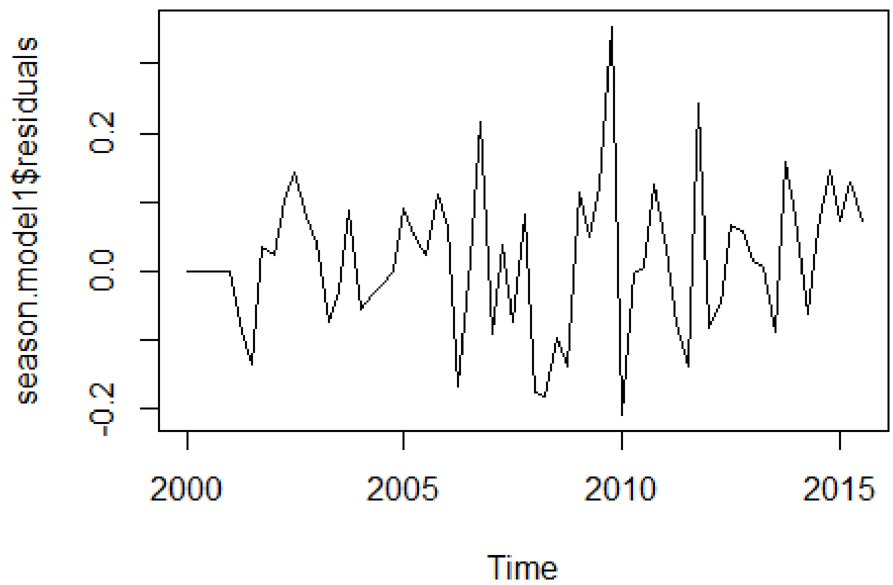
```r
    1, 1), 4), method = "ML")

season.model1
```

```
## Series: fp.training
## ARIMA(0,1,1)(0,1,1)[4]
##
## Coefficients:
##           ma1      sma1
##       -0.4988   -0.0807
## s.e.   0.1237    0.1049
##
## sigma^2 estimated as 0.01276:  log likelihood=45.05
## AIC=-84.1   AICc=-83.66   BIC=-77.92
```

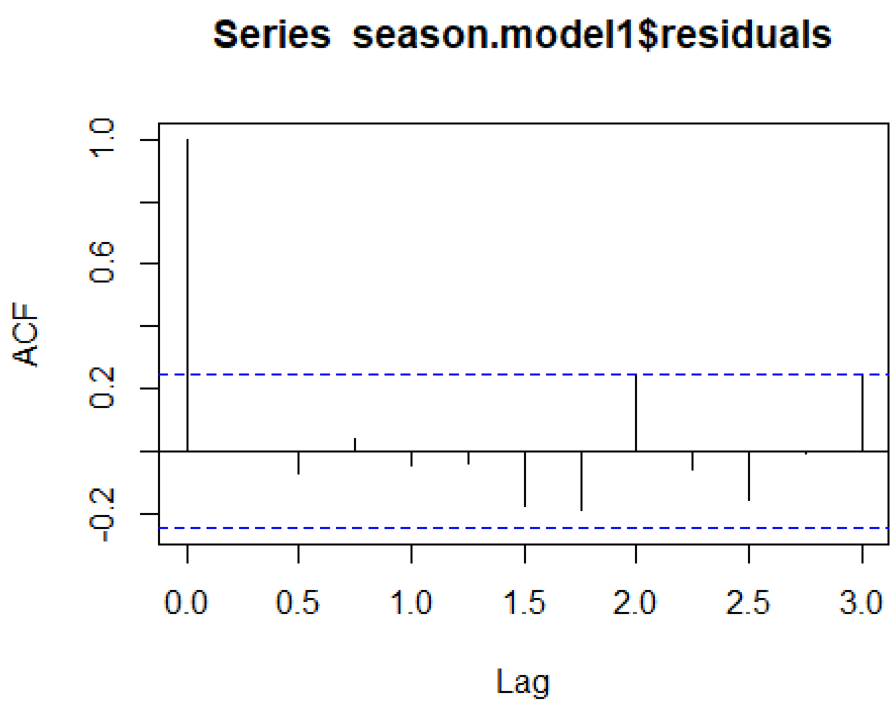
```r
hist(season.model1$residuals)
```



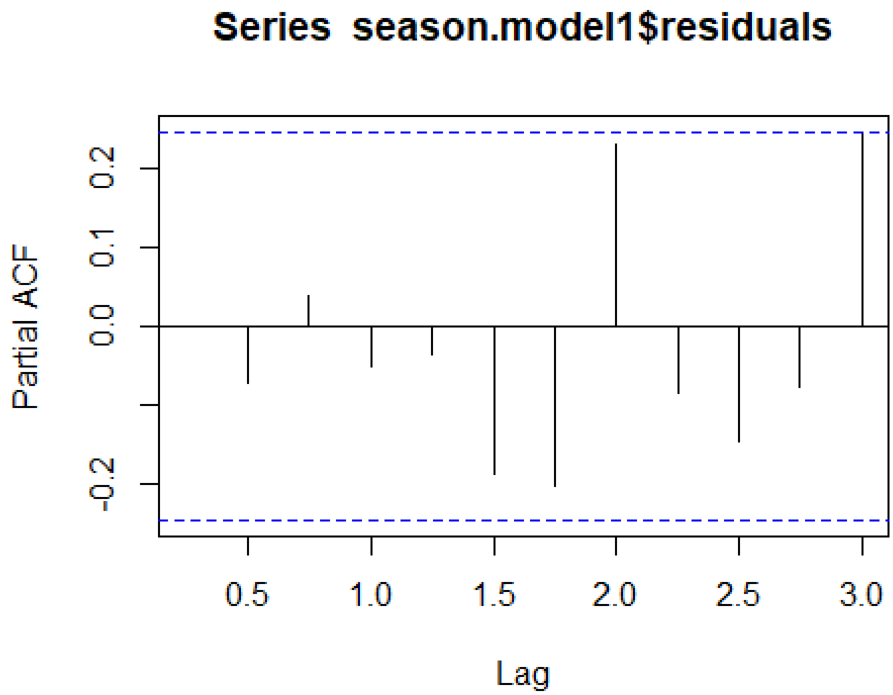

```r
plot.ts(season.model1$residuals)
```




```r
acf(season.model1$residuals, lag.max = 12)
```



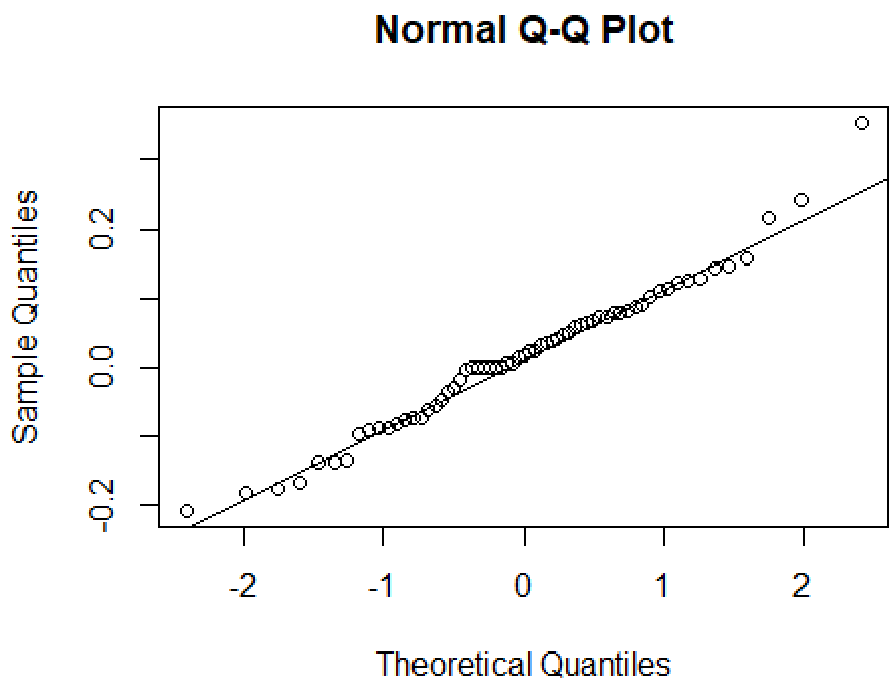

```r
pacf(season.model1$residuals, lag.max = 12)
```




```r
# Let's conduct some formal tests.

# Let's examine normality
shapiro.test(season.model1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  season.model1$residuals
## W = 0.98052, p-value = 0.4174
```

```r
qqnorm(season.model1$residuals)
qqline(season.model1$residuals)
```



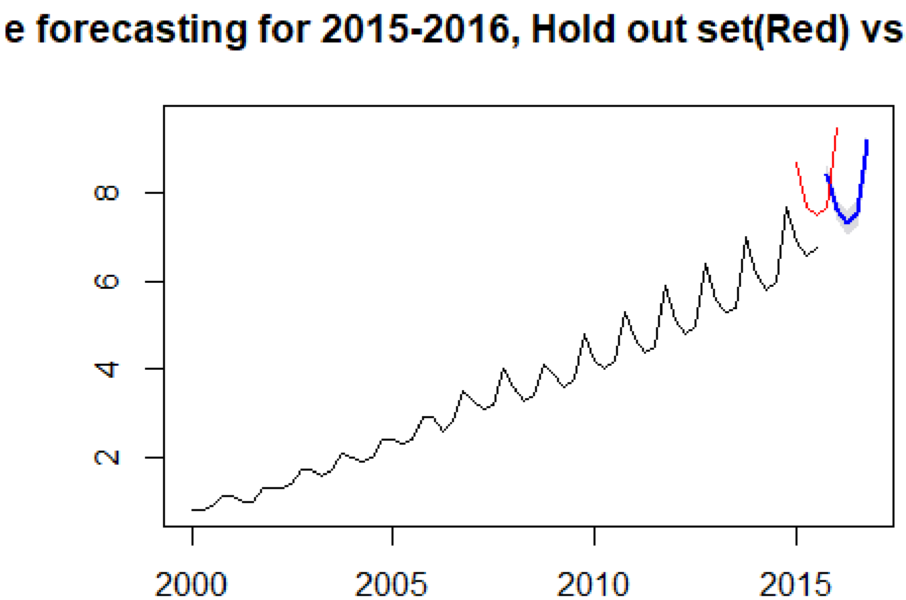

```
# Box-Ljung test
Box.test(season.model1$residuals, type = "Ljung-Box")

##
##  Box-Ljung test
##
## data:  season.model1$residuals
## X-squared = 5.2229e-06, df = 1, p-value = 0.9982
```

The residuals generated from this model rejects the null hypothesis that they are generated from a normal distribution. Bear in mind that this test is really sensitive so we should not solely base our evaluation of the model on this test.

In sample forecasting for 2015 and 2016

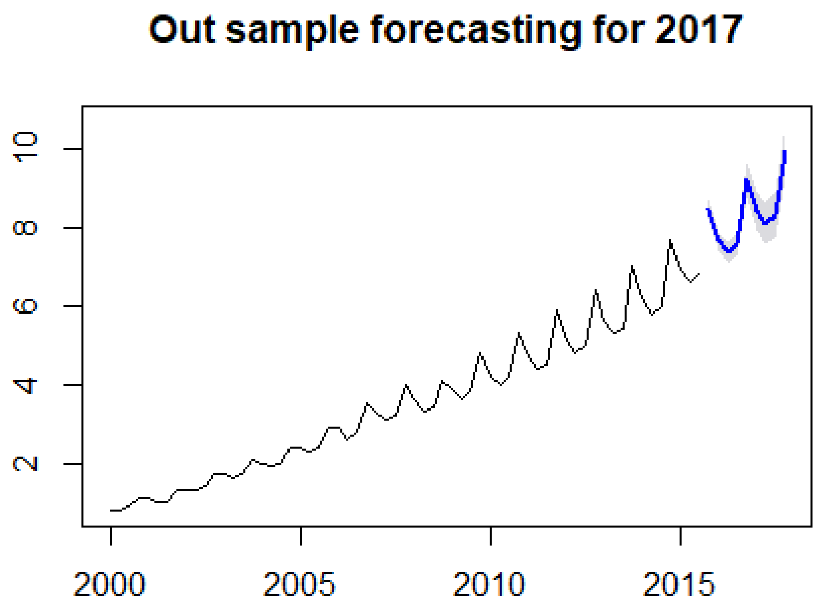
```
par(mfrow = c(1, 1))
futurVal <- forecast(season.model1, h = 5, level = c(95))
plot(futurVal, main = "In sample forecasting for 2015-2016, Hold out set(Red) vs. Forecast (Blue)",
    xatn = "n")

## Warning in plot.window(xlim, ylim, log, ...): "xatn" is not a graphical
## parameter

## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "xatn" is no
t
## a graphical parameter
```

```
## Warning in axis(1, ...): "xatn" is not a graphical parameter

## Warning in axis(2, ...): "xatn" is not a graphical parameter

## Warning in box(...): "xatn" is not a graphical parameter

lines(fp.test, col = "red")
```



Out sample forecasting for 2017

```
par(mfrow = c(1, 1))
futurVal1 <- forecast(season.model1, h = 9, level = c(95))
plot(futurVal1, main = "Out sample forecasting for 2017")
```



1. Read AMAZ.csv and UMCSENT.csv into R as R DataFrames

2. Convert them to xts objects

```
df1 <- read.csv("https://raw.githubusercontent.com/MIDS-W271/main-f18/e5dafd9
afeb57f1f862f5561a3fcb535480b364c/labs/lab3/AMAZ.csv?token=AhtFYQ_40dG2ZWsetc
VyOu47H2a_F_twks5cAuVcwA%3D%3D",
    header = TRUE, stringsAsFactors = FALSE)
str(df1)

## 'data.frame':    1179 obs. of  6 variables:
##  $ Index      : chr  "2007-01-03" "2007-01-04" "2007-01-08" "2007-01-09" .
..
##  $ AMAZ.Open  : num  20 20 19.2 22 20.8 20.8 22 21.6 22 23.2 ...
##  $ AMAZ.High  : num  20 20 22 22 20.8 21.6 22 21.6 22 23.2 ...
##  $ AMAZ.Low   : num  16 20 19.2 20.8 20.8 20.8 22 21.2 21.6 22.8 ...
##  $ AMAZ.Close : num  16 20 22 20.8 20.8 21.6 22 21.2 21.6 22.8 ...
##  $ AMAZ.Volume: int  650 67 1801 356 438 2318 306 925 2138 527 ...

names(df1)

## [1] "Index"       "AMAZ.Open"   "AMAZ.High"   "AMAZ.Low"    "AMAZ.Close"
## [6] "AMAZ.Volume"

head(df1, 5)
```

```
##        Index AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume
## 1 2007-01-03      20.0      20.0     16.0       16.0         650
## 2 2007-01-04      20.0      20.0     20.0       20.0          67
## 3 2007-01-08      19.2      22.0     19.2       22.0        1801
## 4 2007-01-09      22.0      22.0     20.8       20.8         356
## 5 2007-01-10      20.8      20.8     20.8       20.8         438
```

```
tail(df1, 5)
```

```
##           Index AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume
## 1175 2013-01-07      0.80      1.00     0.80       1.00        2715
## 1176 2013-01-08      0.80      0.80     0.68       0.68        4668
## 1177 2013-01-09      0.88      0.88     0.80       0.80        2750
## 1178 2013-01-11      0.80      0.80     0.80       0.80        3000
## 1179 2013-01-15      0.68      0.68     0.68       0.68        1000
```

```
df2 <- read.csv("https://raw.githubusercontent.com/MIDS-W271/main-f18/e5dafd9
afeb57f1f862f5561a3fcb535480b364c/labs/lab3/UMCSENT.csv?token=AhtFYeW-O3b8agq
FzIzVFjRTbg9mhFxzks5cAuWXwA%3D%3D",
    header = TRUE, stringsAsFactors = FALSE)
str(df2)
```

```
## 'data.frame':    477 obs. of  2 variables:
##  $ Index  : chr  "1978-01-01" "1978-02-01" "1978-03-01" "1978-04-01" ...
##  $ UMCSENT: num  83.7 84.3 78.8 81.6 82.9 80 82.4 78.4 80.4 79.3 ...
```

```
names(df2)
```

```
## [1] "Index"   "UMCSENT"
```

```
head(df2, 5)
```

```
##        Index UMCSENT
## 1 1978-01-01    83.7
## 2 1978-02-01    84.3
## 3 1978-03-01    78.8
## 4 1978-04-01    81.6
## 5 1978-05-01    82.9
```

```
tail(df2, 5)
```

```
##          Index UMCSENT
## 473 2017-05-01    97.1
## 474 2017-06-01    95.1
## 475 2017-07-01    93.4
## 476 2017-08-01    96.8
## 477 2017-09-01    95.1
```

```
unemp_idx2 <- seq(as.Date("1978/1/1"), by = "month", length.out = length(df2[
,
    1]))
head(unemp_idx2)
```

```
## [1] "1978-01-01" "1978-02-01" "1978-03-01" "1978-04-01" "1978-05-01"
## [6] "1978-06-01"
```

```
df2_xts <- xts(df2$UMCSENT, order.by = unemp_idx2)
str(df2_xts)
```

```
## An 'xts' object on 1978-01-01/2017-09-01 containing:
##   Data: num [1:477, 1] 83.7 84.3 78.8 81.6 82.9 80 82.4 78.4 80.4 79.3 ...
##   Indexed by objects of class: [Date] TZ: UTC
##   xts Attributes:
##  NULL
```

```
head(df2_xts)
```

```
##             [,1]
## 1978-01-01 83.7
## 1978-02-01 84.3
## 1978-03-01 78.8
## 1978-04-01 81.6
## 1978-05-01 82.9
## 1978-06-01 80.0
```

```
unemp_idx1 <- seq(as.Date("2007/1/3"), by = "day", length.out = length(df1[,
    1]))
df1_xts <- xts(df1, order.by = unemp_idx1)
```

3. Merge the two set of series together, perserving all of the obserbvations in both set of series.

a. fill all of the missing values of the UMCSENT series with -9999

```
unemp01 <- merge(df1_xts, df2_xts, join = "outer", fill = -9999)
```

```
## Warning in merge.xts(df1_xts, df2_xts, join = "outer", fill = -9999): NAs
## introduced by coercion
```

```
str(unemp01)
```

```
## An 'xts' object on 1978-01-01/2017-09-01 containing:
##   Data: num [1:1618, 1:7] -9999 -9999 -9999 -9999 -9999 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr [1:7] "Index" "AMAZ.Open" "AMAZ.High" "AMAZ.Low" ...
##   Indexed by objects of class: [Date] TZ: UTC
##   xts Attributes:
##  NULL
```

```
head(unemp01)
```

```
##             Index AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume
## 1978-01-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-02-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-03-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-04-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-05-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-06-01 -9999     -9999     -9999    -9999      -9999       -9999
##             df2_xts
## 1978-01-01    83.7
## 1978-02-01    84.3
## 1978-03-01    78.8
## 1978-04-01    81.6
## 1978-05-01    82.9
## 1978-06-01    80.0
```

b. then create a new series, named UMCSENT02, from the original UMCSENT series replace all of the -9999 with NAs

```
unemp02 <- unemp01
head(unemp02)
```

```
##             Index AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume
## 1978-01-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-02-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-03-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-04-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-05-01 -9999     -9999     -9999    -9999      -9999       -9999
## 1978-06-01 -9999     -9999     -9999    -9999      -9999       -9999
##             df2_xts
## 1978-01-01    83.7
## 1978-02-01    84.3
## 1978-03-01    78.8
## 1978-04-01    81.6
## 1978-05-01    82.9
## 1978-06-01    80.0
```

```
describe(unemp02$df2_xts)
```

```
## unemp02$df2_xts
##
##  1  Variables      1618  Observations
## ------------------------------------------------------------------------
-
## df2_xts
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1618        0      292    0.649    -7026     4197  -9999.0  -9999.0
##      .25      .50      .75      .90      .95
##  -9999.0  -9999.0     70.4     92.6     96.2
##
## Value      -10000    100
```

```
## Frequency    1141    477
## Proportion  0.705  0.295
## ---------------------------------------------------------------------------
-
```

```r
unemp02[unemp02 <= -9999] <- NA
head(cbind(unemp01$df2_xts["2001-01-01/2018-12-15"], unemp02$df2_xts["2001-01
-01/2018-12-15"]),
    100)
```

```
##            df2_xts df2_xts.1
## 2001-01-01    94.7      94.7
## 2001-02-01    90.6      90.6
## 2001-03-01    91.5      91.5
## 2001-04-01    88.4      88.4
## 2001-05-01    92.0      92.0
## 2001-06-01    92.6      92.6
## 2001-07-01    92.4      92.4
## 2001-08-01    91.5      91.5
## 2001-09-01    81.8      81.8
## 2001-10-01    82.7      82.7
## 2001-11-01    83.9      83.9
## 2001-12-01    88.8      88.8
## 2002-01-01    93.0      93.0
## 2002-02-01    90.7      90.7
## 2002-03-01    95.7      95.7
## 2002-04-01    93.0      93.0
## 2002-05-01    96.9      96.9
## 2002-06-01    92.4      92.4
## 2002-07-01    88.1      88.1
## 2002-08-01    87.6      87.6
## 2002-09-01    86.1      86.1
## 2002-10-01    80.6      80.6
## 2002-11-01    84.2      84.2
## 2002-12-01    86.7      86.7
## 2003-01-01    82.4      82.4
## 2003-02-01    79.9      79.9
## 2003-03-01    77.6      77.6
## 2003-04-01    86.0      86.0
## 2003-05-01    92.1      92.1
## 2003-06-01    89.7      89.7
## 2003-07-01    90.9      90.9
## 2003-08-01    89.3      89.3
## 2003-09-01    87.7      87.7
## 2003-10-01    89.6      89.6
## 2003-11-01    93.7      93.7
## 2003-12-01    92.6      92.6
## 2004-01-01   103.8     103.8
## 2004-02-01    94.4      94.4
## 2004-03-01    95.8      95.8
```

```
## 2004-04-01    94.2      94.2
## 2004-05-01    90.2      90.2
## 2004-06-01    95.6      95.6
## 2004-07-01    96.7      96.7
## 2004-08-01    95.9      95.9
## 2004-09-01    94.2      94.2
## 2004-10-01    91.7      91.7
## 2004-11-01    92.8      92.8
## 2004-12-01    97.1      97.1
## 2005-01-01    95.5      95.5
## 2005-02-01    94.1      94.1
## 2005-03-01    92.6      92.6
## 2005-04-01    87.7      87.7
## 2005-05-01    86.9      86.9
## 2005-06-01    96.0      96.0
## 2005-07-01    96.5      96.5
## 2005-08-01    89.1      89.1
## 2005-09-01    76.9      76.9
## 2005-10-01    74.2      74.2
## 2005-11-01    81.6      81.6
## 2005-12-01    91.5      91.5
## 2006-01-01    91.2      91.2
## 2006-02-01    86.7      86.7
## 2006-03-01    88.9      88.9
## 2006-04-01    87.4      87.4
## 2006-05-01    79.1      79.1
## 2006-06-01    84.9      84.9
## 2006-07-01    84.7      84.7
## 2006-08-01    82.0      82.0
## 2006-09-01    85.4      85.4
## 2006-10-01    93.6      93.6
## 2006-11-01    92.1      92.1
## 2006-12-01    91.7      91.7
## 2007-01-01    96.9      96.9
## 2007-01-03 -9999.0        NA
## 2007-01-04 -9999.0        NA
## 2007-01-05 -9999.0        NA
## 2007-01-06 -9999.0        NA
## 2007-01-07 -9999.0        NA
## 2007-01-08 -9999.0        NA
## 2007-01-09 -9999.0        NA
## 2007-01-10 -9999.0        NA
## 2007-01-11 -9999.0        NA
## 2007-01-12 -9999.0        NA
## 2007-01-13 -9999.0        NA
## 2007-01-14 -9999.0        NA
## 2007-01-15 -9999.0        NA
## 2007-01-16 -9999.0        NA
## 2007-01-17 -9999.0        NA
## 2007-01-18 -9999.0        NA
```

```
## 2007-01-19 -9999.0        NA
## 2007-01-20 -9999.0        NA
## 2007-01-21 -9999.0        NA
## 2007-01-22 -9999.0        NA
## 2007-01-23 -9999.0        NA
## 2007-01-24 -9999.0        NA
## 2007-01-25 -9999.0        NA
## 2007-01-26 -9999.0        NA
## 2007-01-27 -9999.0        NA
## 2007-01-28 -9999.0        NA
## 2007-01-29 -9999.0        NA
```

c. then create a new series, named UMCSENT03, and replace the NAs with the last observation

```
unemp03 <- unemp02
describe(unemp03$df2_xts)
```

```
## unemp03$df2_xts
##
##  1  Variables      1618  Observations
## ---------------------------------------------------------------------------
-
## df2_xts
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      477     1141      291        1    85.69    14.32    63.24    66.90
##      .25      .50      .75      .90      .95
##    76.10    89.30    94.30    99.82   105.84
##
## lowest :  51.7  52.7  55.3  55.8  56.3, highest: 109.2 110.4 110.7 111.3 1
12.0
## ---------------------------------------------------------------------------
-
```

```
unemp03 <- na.locf(unemp02, option = "locf", na.remaining = "rev")
head(cbind(unemp01$df2_xts["2001-01-01/2018-12-30"], unemp02$df2_xts["2001-01
-01/2018-12-30"],
    unemp03$df2_xts["2001-01-01/2018-12-30"]), 100)
```

```
##            df2_xts df2_xts.1 df2_xts.2
## 2001-01-01    94.7      94.7      94.7
## 2001-02-01    90.6      90.6      90.6
## 2001-03-01    91.5      91.5      91.5
## 2001-04-01    88.4      88.4      88.4
## 2001-05-01    92.0      92.0      92.0
## 2001-06-01    92.6      92.6      92.6
## 2001-07-01    92.4      92.4      92.4
## 2001-08-01    91.5      91.5      91.5
## 2001-09-01    81.8      81.8      81.8
## 2001-10-01    82.7      82.7      82.7
```

```
## 2001-11-01    83.9      83.9      83.9
## 2001-12-01    88.8      88.8      88.8
## 2002-01-01    93.0      93.0      93.0
## 2002-02-01    90.7      90.7      90.7
## 2002-03-01    95.7      95.7      95.7
## 2002-04-01    93.0      93.0      93.0
## 2002-05-01    96.9      96.9      96.9
## 2002-06-01    92.4      92.4      92.4
## 2002-07-01    88.1      88.1      88.1
## 2002-08-01    87.6      87.6      87.6
## 2002-09-01    86.1      86.1      86.1
## 2002-10-01    80.6      80.6      80.6
## 2002-11-01    84.2      84.2      84.2
## 2002-12-01    86.7      86.7      86.7
## 2003-01-01    82.4      82.4      82.4
## 2003-02-01    79.9      79.9      79.9
## 2003-03-01    77.6      77.6      77.6
## 2003-04-01    86.0      86.0      86.0
## 2003-05-01    92.1      92.1      92.1
## 2003-06-01    89.7      89.7      89.7
## 2003-07-01    90.9      90.9      90.9
## 2003-08-01    89.3      89.3      89.3
## 2003-09-01    87.7      87.7      87.7
## 2003-10-01    89.6      89.6      89.6
## 2003-11-01    93.7      93.7      93.7
## 2003-12-01    92.6      92.6      92.6
## 2004-01-01   103.8     103.8     103.8
## 2004-02-01    94.4      94.4      94.4
## 2004-03-01    95.8      95.8      95.8
## 2004-04-01    94.2      94.2      94.2
## 2004-05-01    90.2      90.2      90.2
## 2004-06-01    95.6      95.6      95.6
## 2004-07-01    96.7      96.7      96.7
## 2004-08-01    95.9      95.9      95.9
## 2004-09-01    94.2      94.2      94.2
## 2004-10-01    91.7      91.7      91.7
## 2004-11-01    92.8      92.8      92.8
## 2004-12-01    97.1      97.1      97.1
## 2005-01-01    95.5      95.5      95.5
## 2005-02-01    94.1      94.1      94.1
## 2005-03-01    92.6      92.6      92.6
## 2005-04-01    87.7      87.7      87.7
## 2005-05-01    86.9      86.9      86.9
## 2005-06-01    96.0      96.0      96.0
## 2005-07-01    96.5      96.5      96.5
## 2005-08-01    89.1      89.1      89.1
## 2005-09-01    76.9      76.9      76.9
## 2005-10-01    74.2      74.2      74.2
## 2005-11-01    81.6      81.6      81.6
## 2005-12-01    91.5      91.5      91.5
```

```
## 2006-01-01    91.2      91.2      91.2
## 2006-02-01    86.7      86.7      86.7
## 2006-03-01    88.9      88.9      88.9
## 2006-04-01    87.4      87.4      87.4
## 2006-05-01    79.1      79.1      79.1
## 2006-06-01    84.9      84.9      84.9
## 2006-07-01    84.7      84.7      84.7
## 2006-08-01    82.0      82.0      82.0
## 2006-09-01    85.4      85.4      85.4
## 2006-10-01    93.6      93.6      93.6
## 2006-11-01    92.1      92.1      92.1
## 2006-12-01    91.7      91.7      91.7
## 2007-01-01    96.9      96.9      96.9
## 2007-01-03 -9999.0        NA      96.9
## 2007-01-04 -9999.0        NA      96.9
## 2007-01-05 -9999.0        NA      96.9
## 2007-01-06 -9999.0        NA      96.9
## 2007-01-07 -9999.0        NA      96.9
## 2007-01-08 -9999.0        NA      96.9
## 2007-01-09 -9999.0        NA      96.9
## 2007-01-10 -9999.0        NA      96.9
## 2007-01-11 -9999.0        NA      96.9
## 2007-01-12 -9999.0        NA      96.9
## 2007-01-13 -9999.0        NA      96.9
## 2007-01-14 -9999.0        NA      96.9
## 2007-01-15 -9999.0        NA      96.9
## 2007-01-16 -9999.0        NA      96.9
## 2007-01-17 -9999.0        NA      96.9
## 2007-01-18 -9999.0        NA      96.9
## 2007-01-19 -9999.0        NA      96.9
## 2007-01-20 -9999.0        NA      96.9
## 2007-01-21 -9999.0        NA      96.9
## 2007-01-22 -9999.0        NA      96.9
## 2007-01-23 -9999.0        NA      96.9
## 2007-01-24 -9999.0        NA      96.9
## 2007-01-25 -9999.0        NA      96.9
## 2007-01-26 -9999.0        NA      96.9
## 2007-01-27 -9999.0        NA      96.9
## 2007-01-28 -9999.0        NA      96.9
## 2007-01-29 -9999.0        NA      96.9
```

d. then create a new series, named UMCSENT04, and replace the NAs using linear interpolation.

```
unemp04 <- unemp02
unemp04 <- na.approx(unemp04, maxgap = 31)
head(round(cbind(unemp01$df2_xts["2001-01-01/2018-12-30"], unemp02$df2_xts["2
001-01-01/2018-12-30"],
    unemp03$df2_xts["2001-01-01/2018-12-30"], unemp04$df2_xts["2001-01-01/201
```

```
8-12-30"]),
    1), 100)
```

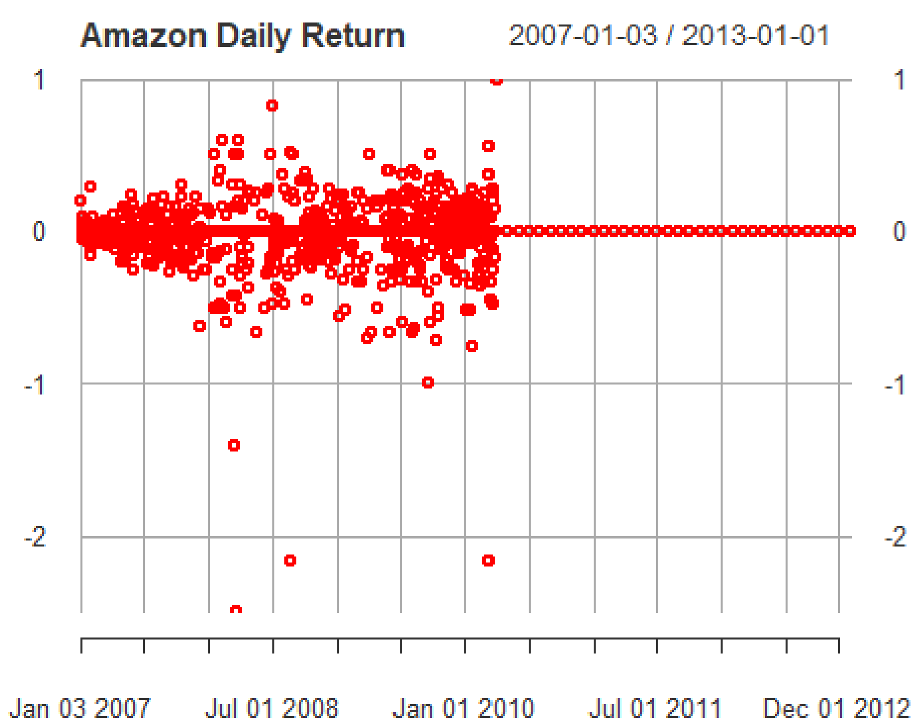
```
##            df2_xts df2_xts.1 df2_xts.2 df2_xts.3
## 2001-01-01    94.7      94.7      94.7      94.7
## 2001-02-01    90.6      90.6      90.6      90.6
## 2001-03-01    91.5      91.5      91.5      91.5
## 2001-04-01    88.4      88.4      88.4      88.4
## 2001-05-01    92.0      92.0      92.0      92.0
## 2001-06-01    92.6      92.6      92.6      92.6
## 2001-07-01    92.4      92.4      92.4      92.4
## 2001-08-01    91.5      91.5      91.5      91.5
## 2001-09-01    81.8      81.8      81.8      81.8
## 2001-10-01    82.7      82.7      82.7      82.7
## 2001-11-01    83.9      83.9      83.9      83.9
## 2001-12-01    88.8      88.8      88.8      88.8
## 2002-01-01    93.0      93.0      93.0      93.0
## 2002-02-01    90.7      90.7      90.7      90.7
## 2002-03-01    95.7      95.7      95.7      95.7
## 2002-04-01    93.0      93.0      93.0      93.0
## 2002-05-01    96.9      96.9      96.9      96.9
## 2002-06-01    92.4      92.4      92.4      92.4
## 2002-07-01    88.1      88.1      88.1      88.1
## 2002-08-01    87.6      87.6      87.6      87.6
## 2002-09-01    86.1      86.1      86.1      86.1
## 2002-10-01    80.6      80.6      80.6      80.6
## 2002-11-01    84.2      84.2      84.2      84.2
## 2002-12-01    86.7      86.7      86.7      86.7
## 2003-01-01    82.4      82.4      82.4      82.4
## 2003-02-01    79.9      79.9      79.9      79.9
## 2003-03-01    77.6      77.6      77.6      77.6
## 2003-04-01    86.0      86.0      86.0      86.0
## 2003-05-01    92.1      92.1      92.1      92.1
## 2003-06-01    89.7      89.7      89.7      89.7
## 2003-07-01    90.9      90.9      90.9      90.9
## 2003-08-01    89.3      89.3      89.3      89.3
## 2003-09-01    87.7      87.7      87.7      87.7
## 2003-10-01    89.6      89.6      89.6      89.6
## 2003-11-01    93.7      93.7      93.7      93.7
## 2003-12-01    92.6      92.6      92.6      92.6
## 2004-01-01   103.8     103.8     103.8     103.8
## 2004-02-01    94.4      94.4      94.4      94.4
## 2004-03-01    95.8      95.8      95.8      95.8
## 2004-04-01    94.2      94.2      94.2      94.2
## 2004-05-01    90.2      90.2      90.2      90.2
## 2004-06-01    95.6      95.6      95.6      95.6
## 2004-07-01    96.7      96.7      96.7      96.7
## 2004-08-01    95.9      95.9      95.9      95.9
## 2004-09-01    94.2      94.2      94.2      94.2
## 2004-10-01    91.7      91.7      91.7      91.7
```

```
## 2004-11-01    92.8      92.8      92.8      92.8
## 2004-12-01    97.1      97.1      97.1      97.1
## 2005-01-01    95.5      95.5      95.5      95.5
## 2005-02-01    94.1      94.1      94.1      94.1
## 2005-03-01    92.6      92.6      92.6      92.6
## 2005-04-01    87.7      87.7      87.7      87.7
## 2005-05-01    86.9      86.9      86.9      86.9
## 2005-06-01    96.0      96.0      96.0      96.0
## 2005-07-01    96.5      96.5      96.5      96.5
## 2005-08-01    89.1      89.1      89.1      89.1
## 2005-09-01    76.9      76.9      76.9      76.9
## 2005-10-01    74.2      74.2      74.2      74.2
## 2005-11-01    81.6      81.6      81.6      81.6
## 2005-12-01    91.5      91.5      91.5      91.5
## 2006-01-01    91.2      91.2      91.2      91.2
## 2006-02-01    86.7      86.7      86.7      86.7
## 2006-03-01    88.9      88.9      88.9      88.9
## 2006-04-01    87.4      87.4      87.4      87.4
## 2006-05-01    79.1      79.1      79.1      79.1
## 2006-06-01    84.9      84.9      84.9      84.9
## 2006-07-01    84.7      84.7      84.7      84.7
## 2006-08-01    82.0      82.0      82.0      82.0
## 2006-09-01    85.4      85.4      85.4      85.4
## 2006-10-01    93.6      93.6      93.6      93.6
## 2006-11-01    92.1      92.1      92.1      92.1
## 2006-12-01    91.7      91.7      91.7      91.7
## 2007-01-01    96.9      96.9      96.9      96.9
## 2007-01-03 -9999.0        NA      96.9      96.5
## 2007-01-04 -9999.0        NA      96.9      96.4
## 2007-01-05 -9999.0        NA      96.9      96.2
## 2007-01-06 -9999.0        NA      96.9      96.0
## 2007-01-07 -9999.0        NA      96.9      95.8
## 2007-01-08 -9999.0        NA      96.9      95.6
## 2007-01-09 -9999.0        NA      96.9      95.5
## 2007-01-10 -9999.0        NA      96.9      95.3
## 2007-01-11 -9999.0        NA      96.9      95.1
## 2007-01-12 -9999.0        NA      96.9      94.9
## 2007-01-13 -9999.0        NA      96.9      94.7
## 2007-01-14 -9999.0        NA      96.9      94.6
## 2007-01-15 -9999.0        NA      96.9      94.4
## 2007-01-16 -9999.0        NA      96.9      94.2
## 2007-01-17 -9999.0        NA      96.9      94.0
## 2007-01-18 -9999.0        NA      96.9      93.8
## 2007-01-19 -9999.0        NA      96.9      93.6
## 2007-01-20 -9999.0        NA      96.9      93.5
## 2007-01-21 -9999.0        NA      96.9      93.3
## 2007-01-22 -9999.0        NA      96.9      93.1
## 2007-01-23 -9999.0        NA      96.9      92.9
## 2007-01-24 -9999.0        NA      96.9      92.7
## 2007-01-25 -9999.0        NA      96.9      92.6
```

```
## 2007-01-26 -9999.0        NA       96.9       92.4
## 2007-01-27 -9999.0        NA       96.9       92.2
## 2007-01-28 -9999.0        NA       96.9       92.0
## 2007-01-29 -9999.0        NA       96.9       91.8
```

4. Calculate the daily return of the Amazon closing price (AMAZ.close), where daily return is defined as $(x(t) - x(t-1))/x(t-1)$. Plot the daily return series.

```
plot((diff(unemp01$AMAZ.Close["2007-01-03/2013-01-15"], Lag = 1,
    difference = 1, log = FALSE, na.pad = TRUE)/unemp01$AMAZ.Close[-nrow(unem
p01$AMAZ.Close),
    ]), main = "Amazon Daily Return", type = "p", col = "red")
```



Create a 20-day and a 50-day rolling mean series from the AMAZ.close series.

```
head(df1, 10)

##          Index AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume
## 1  2007-01-03      20.0      20.0     16.0       16.0         650
## 2  2007-01-04      20.0      20.0     20.0       20.0          67
## 3  2007-01-08      19.2      22.0     19.2       22.0        1801
## 4  2007-01-09      22.0      22.0     20.8       20.8         356
```

```
## 5  2007-01-10      20.8      20.8      20.8       20.8          438
## 6  2007-01-11      20.8      21.6      20.8       21.6         2318
## 7  2007-01-12      22.0      22.0      22.0       22.0          306
## 8  2007-01-16      21.6      21.6      21.2       21.2          925
## 9  2007-01-17      22.0      22.0      21.6       21.6         2138
## 10 2007-01-22      23.2      23.2      22.8       22.8          527
```

```r
head(cbind(df1_xts[, 5], rollapply(df1_xts[, 5], 20, FUN = mean,
    na.rm = TRUE, fill = NA)), 30)
```

```
##            AMAZ.Close AMAZ.Close.1
## 2007-01-03       16.0           NA
## 2007-01-04       20.0           NA
## 2007-01-05       22.0           NA
## 2007-01-06       20.8           NA
## 2007-01-07       20.8           NA
## 2007-01-08       21.6           NA
## 2007-01-09       22.0           NA
## 2007-01-10       21.2           NA
## 2007-01-11       21.6           NA
## 2007-01-12       22.8           NA
## 2007-01-13       22.8           NA
## 2007-01-14       22.0           NA
## 2007-01-15       23.2           NA
## 2007-01-16       24.0           NA
## 2007-01-17       24.0           NA
## 2007-01-18       24.0           NA
## 2007-01-19       25.6           NA
## 2007-01-20       24.4           NA
## 2007-01-21       23.6           NA
## 2007-01-22       23.2        22.28
## 2007-01-23       23.6        22.66
## 2007-01-24       23.6        22.84
## 2007-01-25       23.6        22.92
## 2007-01-26       22.4        23.00
## 2007-01-27       20.8        23.00
## 2007-01-28       20.4        22.94
## 2007-01-29       17.6        22.72
## 2007-01-30       16.0        22.46
## 2007-01-31       22.8        22.52
## 2007-02-01       22.0        22.48
```

```r
head(cbind(df1_xts[, 5], rollapply(df1_xts[, 5], 50, FUN = mean,
    na.rm = TRUE, fill = NA)), 70)
```

```
##            AMAZ.Close AMAZ.Close.1
## 2007-01-03      16.00           NA
## 2007-01-04      20.00           NA
## 2007-01-05      22.00           NA
## 2007-01-06      20.80           NA
## 2007-01-07      20.80           NA
```

```
## 2007-01-08      21.60          NA
## 2007-01-09      22.00          NA
## 2007-01-10      21.20          NA
## 2007-01-11      21.60          NA
## 2007-01-12      22.80          NA
## 2007-01-13      22.80          NA
## 2007-01-14      22.00          NA
## 2007-01-15      23.20          NA
## 2007-01-16      24.00          NA
## 2007-01-17      24.00          NA
## 2007-01-18      24.00          NA
## 2007-01-19      25.60          NA
## 2007-01-20      24.40          NA
## 2007-01-21      23.60          NA
## 2007-01-22      23.20          NA
## 2007-01-23      23.60          NA
## 2007-01-24      23.60          NA
## 2007-01-25      23.60          NA
## 2007-01-26      22.40          NA
## 2007-01-27      20.80          NA
## 2007-01-28      20.40          NA
## 2007-01-29      17.60          NA
## 2007-01-30      16.00          NA
## 2007-01-31      22.80          NA
## 2007-02-01      22.00          NA
## 2007-02-02      22.00          NA
## 2007-02-03      22.00          NA
## 2007-02-04      22.80          NA
## 2007-02-05      21.60          NA
## 2007-02-06      24.00          NA
## 2007-02-07      22.80          NA
## 2007-02-08      22.80          NA
## 2007-02-09      23.60          NA
## 2007-02-10      22.00          NA
## 2007-02-11      22.00          NA
## 2007-02-12      22.80          NA
## 2007-02-13      22.00          NA
## 2007-02-14      22.80          NA
## 2007-02-15      22.80          NA
## 2007-02-16      22.80          NA
## 2007-02-17      22.40          NA
## 2007-02-18      21.60          NA
## 2007-02-19      20.40          NA
## 2007-02-20      20.00          NA
## 2007-02-21      21.00     22.0520
## 2007-02-22      21.60     22.1640
## 2007-02-23      20.20     22.1680
## 2007-02-24      21.04     22.1488
## 2007-02-25      21.60     22.1648
## 2007-02-26      22.80     22.2048
```

```
## 2007-02-27      21.20      22.1968
## 2007-02-28      22.40      22.2048
## 2007-03-01      21.60      22.2128
## 2007-03-02      21.60      22.2128
## 2007-03-03      21.60      22.1888
## 2007-03-04      21.60      22.1648
## 2007-03-05      22.40      22.1728
## 2007-03-06      22.40      22.1568
## 2007-03-07      23.20      22.1408
## 2007-03-08      23.20      22.1248
## 2007-03-09      23.20      22.1088
## 2007-03-10      23.60      22.0688
## 2007-03-11      23.60      22.0528
## 2007-03-12      23.20      22.0448
## 2007-03-13      24.80      22.0768
```