**Stat 510 Week 4 Homework Solutions**

**1**. Problem 3.36 of the book (p. 170) asks students to sketch the ACF of a seasonal ARIMA (0,1)×(1,0)12 model with the non-seasonal MA = .5 and the seasonal AR . (They left out the differencing location, which means there is no differencing.)

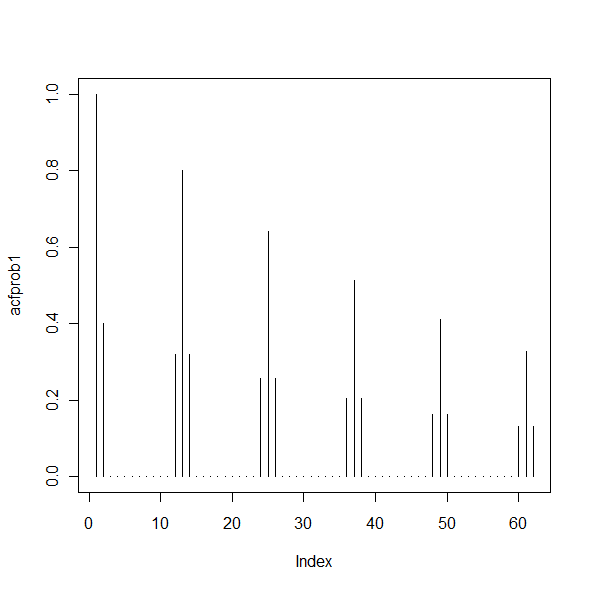
Here’s how you can do it in R:

acfprob1=ARMAacf(ar = c(0,0,0,0,0,0,0,0,0,0,0,.8), ma=c(.5),lag.max=61)

plot(acfprob1, type="h")

The first command asks for 61 lags. The ar list gives coefficients that are 0 up to the 12th lag where the coefficient = .8. The ma list gives the information that the first lag coefficient = .5. In the plot command, the type = “h” causes projections from the value to the axis as we usually do in an ACF.

a. Use R to sketch the ACF in this situation. You may either type the plot command in R rather than copy and paste or you must paste as plain text. R doesn’t like Words’ quote marks. Give the ACF graph as the answer to this part.

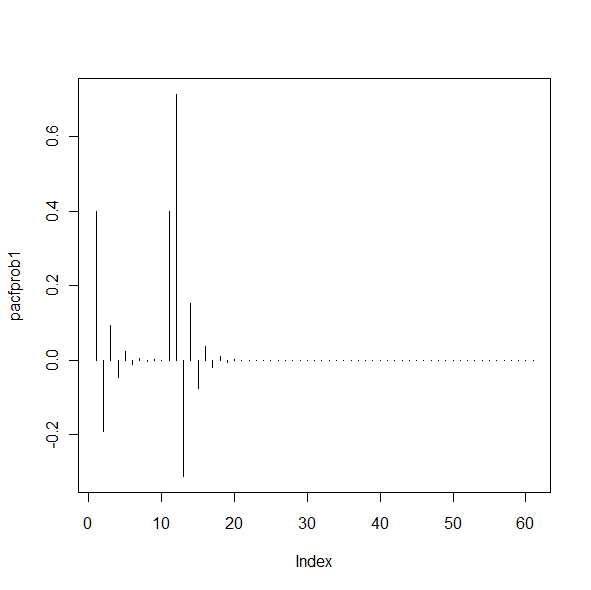


b. To get the PACF of this model, modify the commands in the following way:

pacfprob1=ARMAacf(ar = c(0,0,0,0,0,0,0,0,0,0,0,.8), ma=c(.5),lag.max=61, pacf=TRUE)

plot(pacfprob1, type="h")

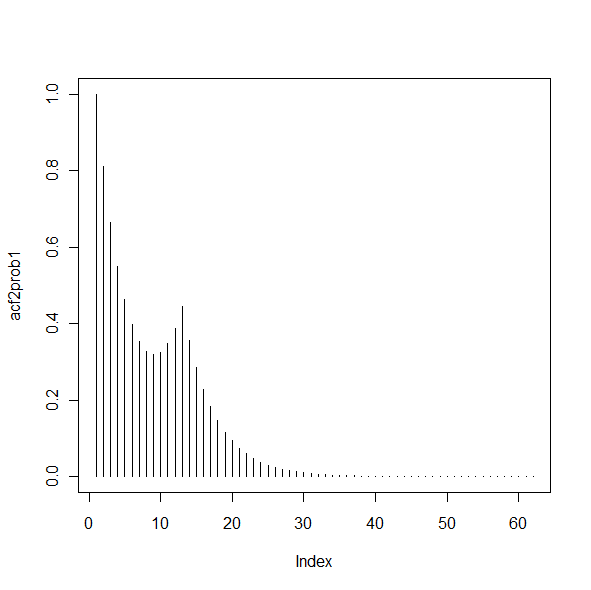
Use these commands to get the PACF of the model. Give the PACF as the answer to this problem.



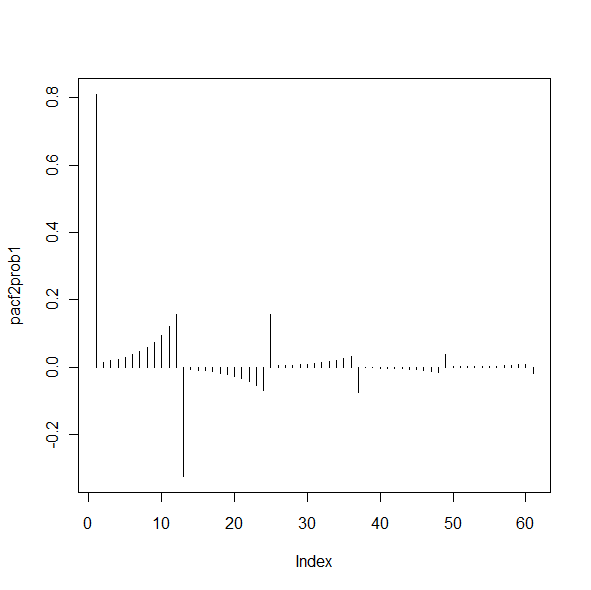
c. Write a brief description of the ACF and PACF patterns for this model.

**The ACF has a spike at lag 1 and then cuts off for all lags up to 11. At lags 12, 24, and so on, there is a tapering pattern related to the seasonal AR(1). The autocorrelations at lags 12, 24, etc. are surrounded by smaller autocorrelations for lags 1 before and 1after the lags 12, 24, and so on. The PACF has a tapering pattern in the non-seasonal lags (and the pattern alternates between positive and negative). It also has a large spike at lag 12. The lag 12 partial autocorrelation is preceded by a non-zero value at lag 11 and followed by a tapering pattern to 0 for a few lags that follow 12.**

d. Now consider the ARIMA (1,0)×(0,1)12 model with non-seasonal AR and seasonal MA =.5. Graph the ACF of this model. (Reverse the seasonality options in the acfprob1 = command.) Give the graph as the answer to this part.



e. Refer to part d. Graph the PACF for the model. Give the graph as the answer to this part.



f. Write a brief description of the ACF and PACF patterns in parts d and e.

**Overall, the ACF has a tapering pattern characteristic of a non-seasonal AR(1). There is a bump upward in the pattern at lag 12, the place where the seasonal MA(1) comes into play. In the PACF, there is a large spike a lag 1 and there is a tapering pattern through lags 12, 24, 36, and so on. In between the seasonal lags of 12, 24, and so on, and between lag 1 and 12 there are slowly increasing (in absolute magnitude) patterns of relatively small partial autocorrelations.**

**2.** This is a more structured version of problem 3.39 in the book (p. 170) which asks students to fit a model of their choice to the log transformed Johnson and Johnson earnings data series. In Homework 1, we saw that this series of 84 quarterly measures of the earnings of this company had both trend and seasonality and that a log transform made the trend linear and the variance stable.

The data are in the dataset jj.dat. (Might already be on your computer from Homework 1, If not, it’s in the Datasets folder.)

Load the astsa library, read the dataset, and create the log transform with something like:

library(astsa)

earnings=scan("jj.dat")

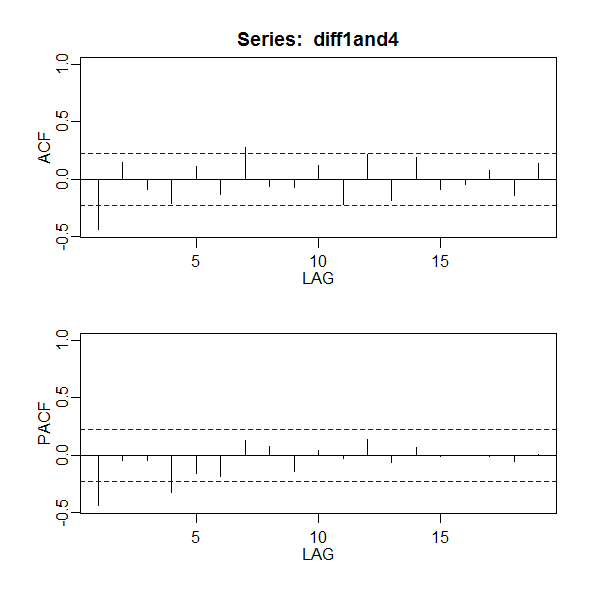
earnings = log(earnings)

Then, because we have linear trend and seasonality (quarterly effects), apply both first and fourth differences. Something like:

diff1 = diff(earnings,1)

diff1and4=diff(diff1,4)

a. Use acf2 to determine the ACF and PACF of the **differenced** data. Be sure that you supply the differenced series (diff1and4) as a parameter of the acf2 command. Give the ACF and PACF of the differenced data.



b. Consider the ACF and PACF values for the first two or three lags. Explain why they suggest either an MA(1) or AR(1) as a possible non-seasonal component.

**An AR(1) interpretation may be the more obvious interpretation as there is a clear spike in the PACF at lag 1 followed by values that are nearly 0. On the other hand, the ACF could be interpreted in the same way, thus suggesting an MA(1).**

c. Examine the ACF and PACF at lags 4, 8 and 12. What possibilities are suggested for a seasonal component of the model?

**One possibility is a seasonal AR(1) if we interpret the PACF as a significant spike at lag 4 followed by non-significant partial autocorrelations. Another interpretation is that there is a seasonal MA(1) as there is a significant spike at lag 4 in the ACF and we could see the PACF as having a tapering pattern through the seasonal lags. The ACF is a bit confusing as there is a barely significant spike at lag 7 – not a seasonal lag. This could be generated by the combination of the non-seasonal and seasonal components. The combination usually creates correlations at the times immediately before and after some seasonal lags.**

d. The ACF and PACF don’t actually give a clear pointer to the best model in this case, but here some possibilities (Note that this list is not all-inclusive):

ARIMA (1,1,0)×(0,1,1)4

ARIMA (0,1,1)×(0,1,1)4

ARIMA (0,1,1)×(1,1,0)4

ARIMA (1,1,0)×(1,1,0)4

Assuming that you’ve loaded the astsa library, these commands will estimate the four models above.

sarima ( name of log series, 1,1,0,0,1,1,4)

sarima ( name of log series, 0,1,1,0,1,1,4)

sarima ( name of log series, 0,1,1,1,1,0,4)

sarima ( name of log series, 1,1,0,1,1,0,4)

Estimate the four models above. For each give the variance estimate and a brief description of the model diagnostics. (You don’t have to give all the graphs.)

**ARIMA (1,1,0)×(0,1,1)4**

**sigma^2 estimated as 0.008472, AIC = -3.72339. Most diagnostics okay, might be a significant Ljung-Box stat early and a significant residual autocorrelation at lag 7.**

**ARIMA (0,1,1)×(0,1,1)4**

**sigma^2 estimated as 0.00793, AIC = -3.789403. Diagnostics basically okay, although might be a significant Ljung-Box statistic early.**

**ARIMA (0,1,1)×(1,1,0)4**

**sigma^2 estimated as 0.007913: AIC = -3.791572. All diagnostics look okay**

**ARIMA (1,1,0)×(1,1,0)4**

**sigma^2 estimated as 0.008468. AIC = - 3.723852. Most diagnostics okay, might be a significant Ljung-Box stat early and a significant residual autocorrelation at lag 7.**

e. Which of the four models in part d do you think is the best model? Explain.

**It’s a close call, but ARIMA (0,1,1)×(1,1,0)4 gives the lowest variance estimate and the lowest AIC. The diagnostics are okay for the model.**

**3.** This is a more structured version of problem 3.37 on page 170 of the book. The problem asks students to determine a model for a monthly series of number of unemployed persons in the United States. There are n = 372 data values, the monthly numbers for 1948-1978.

The data are in the file unemp.dat in the Datasets folder.

There is an upward trend because these are numbers of people rather than rates, so the number increases as the population size increases.

There also is seasonality. Unemployment numbers are affected by such things as seasonal agricultural production, holiday seasons, and so on.

a. Calculate a variable that is the result of applying a 1st difference and then a 12th difference. Something like the following should do the trick:

diff1=diff(x,1)

diff1and12=diff(diff1,12)

Then determine the ACF and PACF of the differenced data series (diff1and12). Be sure to use the differenced series, not the original data for this step.

(i) Discuss why the ACF and PACF suggest that an AR(2) may be good for the non-seasonal component. (Look at what’s happening for the first few lags in the ACF and PACF.)

**In the PACF there are two clear spikes followed by essentially 0 values for the next few lags. This is accompanied by a tapering sinusoidal pattern in the ACF.**

(ii) Discuss why the ACF and PACF suggest that a seasonal MA(1) may be good for the seasonal component. Look at what’s happening around lags 12, 24, and so on.

**There’s a clear spike at lag 12 in the ACF, and essentially 0 values at lags 24, 36, and so on. The PACF has what may be a tapering pattern through lags 12, 24 and so on.**

b. Fit an ARIMA (2,1, 0,)× (0,1,1)12 model to the original series. In the sarima command, specify the original series, not the differenced series because the differencing will be specified in the command itself. Discuss whether the diagnostics for the model suggest that the model is suitable. (Look at ACF of residuals, Ljung-Box-Pierce tests, and so on.)

**The diagnostics are good. All values in the residual ACF are non-significant as are the Ljung-Box values. The normal probability plot does show one outlier, although that doesn’t negate the model. Some might also add that the coefficients are all significant.**



c. Use the model fit in part b to forecast the unemployment values for the next 12 months following the end of the series. One possible command structure is

sarima.for (name of data series, 12, 2,1,0,0,1,1,12)

Give the 12 forecasts as the answer to this problem.

**676.4664, 685.1172, 653.2388, 585.6939, 553.8813, 664.4072, 647.0657, 611.0828,**

**594.6414, 569.3997, 587.5801, 581.183**

**Problem didn’t call for graph of forecasts in answer, but here it is anyway 🡪**

