**Stat 510 Week 14 Homework Solutions**

1. The R command arima.sim can be used to simulate sample values from a time series process. The basic structure is

y= arima.sim (n = , list (ar=c( , , …), ma=c(, ,..)))

Give the sample size after n = , list numerical values of ar coefficients within ar = c(, ,.) and list numerical values of ma coefficients within ma = c(, ,..). If there are not any terms of a type (ar or ma), you don’t need to include that portion of the code. Be sure to get the number of left and right parentheses to be equal.

Examples:

y = arima.sim (n = 200, list (ar=c(0.9,-0.5))) ## n = 200 for an AR(2)

y = arima.sim(n = 400, list (ar = c(0.7), ma=c(0.6))) ## n = 400 for an ARMA(1,1)

The variable (object) name y is arbitrary. You can name your simulated series whatever you wish.

The default is a series with mean 0 and standard deviation = 1. To change the mean and standard deviation, follow arima.sim command with a command with the following structure:

y = mean + std-dev \*y

Give specific numerical values for the mean and standard deviation. For example,

X = 120 + 6\*y ## series will have mean 120 and standard deviation = 6.

A. Describe the general theoretical appearance of the ACF and PACF of an MA(2) model.

**In the ACF, we should see a spike at lags 1 and 2 followed by generally non-significant values for lags past 2.  The PACF should taper to 0.**

B. In R, simulate n = 400 values from an MA(2) with coefficients  and . You can leave the mean and standard deviation at 0 and 1, respectively. Then, create the sample ACF and PACF of your simulated series. Give those plots and write a brief interpretation. Were the patterns what they should be? Are there any other possible interpretations?

**The pattern in the following ACF and PACF are what is expected for a MA(2).**



C. Describe the general theoretical appearance of the ACF and PACF of an AR(3) model.

**In the PACF, we should see a spike at lags 1 and 2 and 3 followed by generally non-significant values for lags past 3.  The ACF should taper to 0.**

D. In R, simulate *n* = 400 values from an AR(3) with coefficients  , , and You can leave the mean and standard deviation at 0 and 1, respectively. Then, create the sample ACF and PACF of your simulated series. Give those plots and write a brief interpretation. Were the patterns what they should be? Are there any other possible interpretations?

**The pattern in the following ACF and PACF are what is expected for an AR(3).**



E. Describe the general theoretical appearance of the ACF and PACF of an ARMA(1,1) model.

**The ACF and PACF should both taper to 0.**

F. In R, simulate *n* = 400 values from an ARMA(1,1) with coefficients  , You can leave the mean and standard deviation at 0 and 1, respectively. Then, create the sample ACF and PACF of your simulated series. Give those plots and write a brief interpretation. Were the patterns what they should be? Are there any other possible interpretations?

**The pattern in the following ACF and PACF are what is expected for an ARMA, though the order is not obvious as is expected.**



2. Use the metalsemploy.dat dataset from the Week 14 folder. The data are *n* =132 consecutive measures of the number of people employed in metals industries in the state of Wisconsin.

A. Create a time series plot of the data. Give the plot and write a brief interpretation.

**The data meanders somewhat as it follows long streaks up and down throughout the observed time period. There are no issues with non-constant variance, outliers, cyclical behavior, or abrupt changes.**



B. Determine the sample ACF and PACF of the series. Give the plots and write a brief interpretation.

**The following PACF shows a clear AR(1) pattern, though the ACF disagrees as it tapers so slowly to 0. Because the lag 1 autocorrelation is close to 1, taking first differences seems like a suitable approach.**



C. Determine the sample ACF and PACF of the first differences of the series. Give the plots and write a brief interpretation. What model is suggested?

**The first differences are white noise suggesting the data are a random walk.**



3. Use the buffsnow.dat dataset in the Week 14 folder. The data are n = annual values of snowfall (inches) in Buffalo, New York.

Determine an ARIMA model for these data. Give the sample ACF and PACF, a brief interpretation of these plots, output for your estimated model, and proof that the model is suitable (i.e., the diagnostic plots).



**The ACF and PACF show just barely significant lag 1 and lag 2 terms suggesting either an AR(1), AR(2), MA(1), or a MA(2).**

**We chose a MA(2) as the best model for these data. The diagnostic plots look fine and all terms are significant. The AR(2) model also satisfied the diagnostic checks though the 2nd AR term was not significant and the AR(1) model did not satisfy all assumptions.**

**Coefficients:**

**ma1 ma2 xmean**

**0.2971 0.2989 80.8990**

**s.e. 0.1267 0.1136 4.3118**

**sigma^2 estimated as 467.6: log likelihood = -283.16, aic = 574.32**

**$AIC**

**[1] 7.242747**

**$AICc**

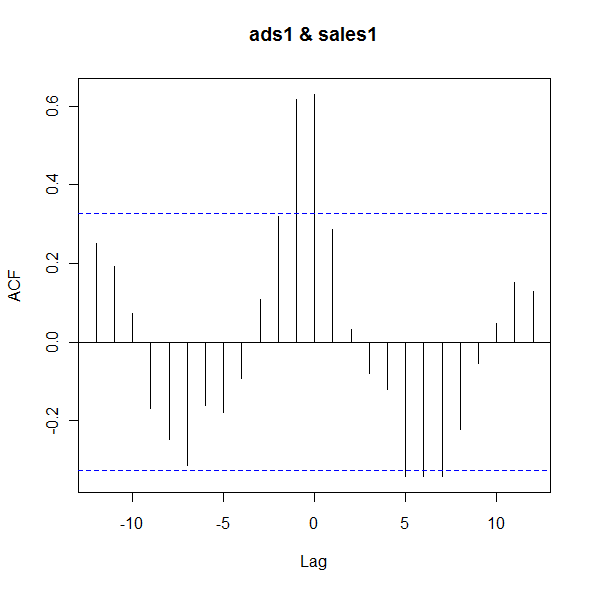
**[1] 7.28544**

**$BIC**

**[1] 6.344801**

****

4. The following CCF from R give cross correlations between x = annual advertising expenditures and y = annual sales for a product. Propose a possible lagged regression model (or more than one if you wish) for predicting sales. Briefly explain. Remember that in R the negative lags give the lags of the x-variable.



**The most dominant cross correlations occur at lags 0 and -1 suggesting that we should use the current, *xt*, and previous year, *xt-1*, advertising expenditures to predict sales, *yt*.**