**Stat 510 Week 3 Homework Solutions**

**1.** An AR(1) model is estimated to be with the usual assumptions about . There were *n* = 80 observations, , and 3. Psi-weights  to  as given by R are:

[1] −0.70000 0.49000 −0.34300 0.24010 −0.16807

A. Forecast the value of and determine a 95% prediction interval for.



**s.e. = **

**Interval is 47.9 ± 1.96(1.581), which is 47.9 ± 3.099, or 44.801 to 50.999**

B. Forecast the value of , the value at time 83, and determine a 95% prediction interval for .

****

****

**so, **

**s.e.===2.0797**

**Interval is 38.471 ± 1.96(2.0797), which is 38.471 ± 4.076, or 34.395 to 42.547**

2. For this problem, use the dataset sim1week3.dat from the Datasets folder. Download it to the directory you’ve set up for course datasets. It’s a simulated series with n = 240 values. Note to people who aren’t using R – the time series order in the data file is across rows.

In R, change the working directory to your directory that contains the dataset. Use these commands to load the scripts (make them available) and read the data:

library(astsa)

series1 = scan ("sim1week3.dat")

These commands assume that both files are in the working directory. If not, you’ll need to specify the full path for finding the file(s) on your computer.

A. Plot the “series1” data. Comment on any important features. This command will do the trick: plot (series1, type= "b")

**There’s no trend. There’s a relatively fast up and down to the series. There are some periods of time where the variation looks to be greater, however none of the other important features to consider are present.**



B. Examine the ACF and PACF of the data. Briefly indicate a possible model (or, models) and explain why you think this model (or models) will work.

The command acf2 (series1, 20) will give an ACF and PACF up to 20 lags.

**The ACF has one clearly significant value at lag 1 and a barely significant value at lag 2. The PACF has a tapering pattern. Either an MA(1) or MA(2) model may work.**



C. After the astsa library is loaded, the command

sarima (series name, order of AR, differencing order, MA order) will give the type of output described in Lesson 3.1 for this week. For example, sarima (series1, 0, 0, 1) will fit a MA(1) to the data.

Fit a MA(1) to the data. Using the diagnostics given in the output graph, discuss whether this seems to be a suitable model.



**The diagnostics indicate this is not a suitable model. There’s a significant residual autocorrelation at lag 1 and the Ljung-Box statistics indicate significant residual autocorrelations.**

D. Fit a MA(2) to the data. Using the diagnostics given in the output graph, discuss whether this seems to be a suitable model.



**All diagnostics look good. There are no significant residual autocorrelations.**

E. Refer to the model in part D. Explain whether the two MA coefficients are statistically significant or not.

**-0.9597/0.0612 = -15.68**

**0.2532/0.0610 = 4.15**

**Both MA coefficients are statistically significant. For each, the ratio coeff./s.e. is greater than 1.96 in absolute value.**

F. Write out the estimated equation for the MA(2) model.



G. Fit an ARIMA (1,0,1) to the data. This might also be called an ARMA (1,1). There’s one AR term, one MA term and no differencing.

Using the diagnostics given in the output graph, discuss whether this seems to be a suitable model.



**All diagnostics look good. There are no significant residual autocorrelations.**

H. Compare the MA(2) and ARMA(1,1) models. Does either seem better than the other? Explain.

**It’s a close call. The ARMA(1,1) may give a slightly better fit. The variance estimate is lower as are the information criterion.**

I. Use the MA(2) to forecast values for the next 6 time periods past the end of the series. With astsa loaded, the command is sarima.for(series1,6,0,0,2). The parameters being supplied, in order, are the series name, the number of times for forecasts, the AR order, the differencing and the MA order).

As an answer give, the forecasts and their standard errors (technically, the standard errors are for the forecast errors).

**Forecasts: 11.48616 9.54875 10.00182 10.00182 10.00182 10.00182**

**SE: 1.006774 1.395420 1.418518 1.418518 1.418518 1.418518**

J. Refer to the forecasts in part I. Explain why all forecasts after the second time after the series are the same as each other.

**In general, forecasts from a stationary ARIMA model will converge to the mean. For an MA model this occurs after a number of time periods equal to the order of the model. At this point all estimated (unobserved) errors are 0.**

3. For this problem, use the dataset lakehuron.dat from the Week 3 folder. The series is *n* = 98 consecutive annual measures of the level of Lake Huron. The years are 1875-1972. Again, the time series order is across rows.

If you’re doing this in a new R session (compared to problem 2), start with

library(astsa) to gain access to those scripts.

A. Create a time series plot of the data. Comment on the important features.

****

**There might be a downward trend. There are no obvious issues with seasonality, non-constant variance, outliers, or abrupt changes.**

B. Determine the ACF and PACF of the series. Propose a possible model (or models). Explain.



**There are two clear spikes at lags 1 and 2 in the PACF and a tapering pattern in the ACF. This indicates an AR(2).**

C. Fit an AR(2) model to the series and provide the model output. Using the diagnostics given in the output graph, discuss whether this seems to be a suitable model.

**Following is the R output for the AR(2) model. If we let where  = Lake Huron level at time t, the forecast model is .**

**Coefficients:**

**ar1 ar2 xmean**

**1.0436 -0.2495 9.0473**

**s.e. 0.0983 0.1008 0.3319**

**sigma^2 estimated as 0.4788: log likelihood = -103.63, aic = 215.27**

**$AIC $AICc $BIC**

**[1] 0.3247953 [1] 0.3495923 [1] -0.5960731**

**The following diagnostics for the AR(2) were given by the “sarima” command. All looks good, although we still might wonder a bit about the almost significant spike at lag 9.**



D. Refer to part C. Explain whether the two AR coefficients are statistically significant or not.

**1.0436/ 0.0983= 10.62**

**-0.2495/ 0.1008= 2.48**

**For both coefficients, the ratio, coeff/se, is greater than 1.96 in absolute value. Thus both coefficients are statistically significant.**

E. The output includes an estimate of the mean μ. For an AR(2), the intercept of the usual regression type equation is . Calculate an estimate of the intercept and then write the estimate AR(2) equation in the usual regression type fashion.

**= 1.862839.**

****

F. The final two observations in the series are 9.89 and 9.96 (in that time order). Use these values and the equation written in part E to forecast the values for the next three times past the series. Show the calculations. (But, you can use sarima.for as explained in question 2 part I to check your answers.)







**Check:**

**Forecasts: 9.789559 9.594219 9.432885**

**SE: 0.6919687 1.0001591 1.1566667**

*Comment: You may have thought that the time series plot showed a possible trend. This is a reasonable interpretation, though results without differencing also seem to be reasonable.*

**4**. Use the flourbuffalo.dat dataset in the Datasets folder for this problem. The series gives the index value for monthly flour prices in a Buffalo, NY commodity exchange. Again, the series time order is across rows in the dataset.

If this is a new R session, load the astsa library at the beginning of the session.

A. Plot the flour price series. Comment on any important features.



**There’s an upward trend for about 30 time periods followed by a downward trend for about 30 time periods, and then there’s an upward trend for about 40 time periods. We do not see a repeating pattern to suggest seasonality even though the data are monthly. No other important features are of concern in the plot.**

B. Determine the ACF and PACF for the series. Propose a possible model (or models). Explain.



**There’s a significant spike in the PACF at lag 1 and the ACF is tapering. Before we consider an AR(1), we should examine first differences because the lag 1 ACF is close to 1.**

C. Calculate the first difference for the series. The basic structure of the command is

New variable name = diff( series name, 1)

For example, if the series is named prices, you might use the command diff1=diff(prices,1).

Then, determine the ACF and PACF for the first differences. (For example, acf2(diff1) might work.) Discuss the result.



**The ACF and PACF of the first differences have no significant values. This suggests that the first differences may be white noise. Thus prices may be a random walk because changes from one time to the next are random.**

**Note: In part b, the first order autocorrelation is quite high. You weren’t asked to estimate the AR(1) model. You would have found that the estimated AR coefficient is 0.97. The AR(1) model is estimated to be  which is . This is quite close to the model suggested in part C, that the first differences are random.**