**Stat 510 Week 5 Homework Solutions**

1. Use the “pamilk.dat” dataset in the Datasets folder. The data are *n* = 160 simulated **monthly** amounts of milk produced by dairies in Pennsylvania beginning with the month of June. To read in monthly data that does not begin in January, use the following adjustment to the ts command:

milk = ts(scan("pamilk.dat"), start = c(1, 6), frequency = 12)

A. Fit an additive decomposition model to the data. See Lesson 5.1 for R code. (Use the decompose command). Don’t forget to identify the frequency of the seasonal span (see Lesson 5.1). Give the estimated values of the 12 monthly (seasonal effects)

**Seasonal adjustment factors are:**

**$figure**

**[1] 21.341185 12.054448 -1.347684 -16.384667 -18.285220 -16.565895**

**[7] -13.730689 -3.889224 0.700346 3.013180 11.619068 21.475152**

B. Suppose that in the month of October following the end of the series, the observed production is 735. Calculate the seasonally adjusted value (de-seasonalized value) for this month.

**Jun Jul Aug Sep Oct Nov**

**21.341185 12.054448 -1.347684 -16.384667 -18.285220 -16.565895**

**Dec Jan Feb Mar Apr May**

**-13.730689 -3.889224 0.700346 3.013180 11.619068 21.475152**

**Seasonally adjusted value = Observed value – seasonal adjustment = 735 – -18.28 = 753.28**

C. Create the plot that graphs the components of the additive decomposition and give that plot as the answer to this part. (See Lesson 5.1.)



D. Fit a multiplicative decomposition model to the data. Give the estimated values of the 12 monthly (seasonal) effects.

**Seasonal effects are:**

**$figure**

**[1] 1.0307231 1.0174601 0.9981005 0.9763770 0.9735935 0.9760219**

**[7] 0.9801239 0.9943379 1.0009940 1.0044064 1.0168267 1.0310350**



E. Refer to part B. Repeat the calculation using the seasonal effect based on the multiplicative decomposition.

**Jun Jul Aug Sep Oct Nov**

**1.0307231 1.0174601 0.9981005 0.9763770 0.9735935 0.9760219**

**Dec Jan Feb Mar Apr May**

**0.9801239 0.9943379 1.0009940 1.0044064 1.0168267 1.0310350**

**Seasonally adjusted value = Observed value/seasonal adjustment = 735/.9736 = 754.9302.**

F. Use the stl command in R to do a Lowess decomposition. (See Lesson 5.1.) Give the estimated values of the 12 monthly effects.

**seasonal**

**Jun 1 20.6791761**

**Jul 1 12.1330646**

**Aug 1 -1.1349357**

**Sep 1 -15.8765761**

**Oct 1 -18.1797691**

**Nov 1 -16.2615529**

**Dec 1 -13.7407654**

**Jan 2 -3.8039441**

**Feb 2 0.8684402**

**Mar 2 3.2213592**

**Apr 2 11.0559521**

**May 2 21.0395505**



G. On the basis of a time series plot of the data (part of the graph for part C, by the way), do you think that an additive decomposition will be a suitable description of the data. (Hint: See page 1 of Lesson 5.1.)

**Additive should be okay. The seasonal variation looks to be about the same from year to year.**

2. This problem gives more practice on identifying ARIMA models. Use the same dataset that you did for question 1. In R you might want to restart the session as the command you used in question 1 to identify the frequency might cause confusion when looking at lags in this question. Also, load the astsa library so you can use the acf2, sarima, and sarima.for commands.

A. Identify a possible ARIMA model (or models) for the series. Here’s the process and some hints.

Start with a time series plot of the data, or for that matter, you can rely on various results of question 1. Is there a trend? If so, you need a first difference. Is there seasonality? If so, you need a 12th difference (monthly data here). If both trend and seasonality are present, apply a 1st and 12th difference in sequence. Something like:

diff1 = diff(milk, 1)

diff1and12=diff(diff1,12)

Examine and interpret the acf and pacf of the appropriately differenced data.

As an answer to this part, briefly explain what differencing was necessary. Then describe the appearance of the acf and pacf (you could just give the plots) and briefly discuss what models are suggested.

Look at the first few lags for the non-seasonal part. Look at lags 12, 24, .. for the seasonal part. This one isn’t too bad. You might just see one spike in the seasonal part.



**Because there is a trend and seasonality, it is necessary to take both the first difference and the twelfth difference. The first few lags suggest the possibility of a MA(1) model for the differenced data. There’s a spike at lag 1 in both the ACF and PACF followed by non-significant values. The seasonal model might either be a seasonal AR(1) or seasonal MA(1) because there’s a spike at lag 12 in both the ACF and PACF. We might try all of the following:**

**ARIMA (0,1,1)×(0,1,1)12**

**ARIMA (0,1,1)×(1,1,0)12**

**ARIMA (0,1,1)×(1,1,1)12**

B. Estimate the model (or models) in part a. Examine the diagnostic statistics and decide on a “best” model. As an answer, briefly describe why you think this is a good model for these data.

**Following are some stats for the 3 models suggested in part A. The graphical diagnostics look pretty good for all of them.**

**ARIMA (0,1,1)×(0,1,1)12  variance = 0.8744 AIC = 0.8908**

**ARIMA (0,1,1)×(1,1,0)12 variance = 1.156 AIC = 1.1699**

**ARIMA (0,1,1)×(1,1,1)12 variance = 0.8968 AIC = 0.9286**

**ARIMA (0,1,1)×(0,1,1)12 appears to be the best option from this output with ARIMA (0,1,1)×(1,1,1)12 as a close second. The seasonal AR term in the third model is not significant.**

C. Use the model identified in part b to forecast the milk production for the next 12 months past the end of the series. Give the forecasts. The sarima.for command would work for this.

**The forecasts for ARIMA (0,1,1)×(0,1,1)12 are:**

**Oct Nov Dec Jan Feb Mar**

**747.3941 756.1584 758.6248 772.4197 773.3226 765.3113**

**Apr May Jun Jul Aug Sep**

**771.7231 783.5965 791.0373 788.6217 778.8444 767.3216**

****

3. Use the dataset flourminn.dat for this problem. It gives n = 100 average monthly prices at a Minneapolis commodities exchange. (We saw similar data for Buffalo in a previous assignment).

A. Use a centered moving average of an appropriate length to smooth out the series in order to see trend. Plot the smoothed trend line, either alone or superimposed on a plot of the data. See Lesson 5.2 of this week for guidance (and the first couple of pages of section 2.4 in the book). As the answer to this part, give the plot and a brief description of how you determined the moving averages (an R command could do it).

**The plot is below. You get this (approximately) from either**

**pricetrend = filter(price, filter = c(1/13, 1/13, 1/13, 1/13,1/13, 1/13, 1/13, 1/13, 1/13, 1/13, 1/13, 1/13, 1/13), sides = 2)**

**plot(pricetrend)**

**OR**

**pricetrend = filter(price, filter = c(1/24, 1/12, 1/12, 1/12,1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/24), sides = 2)**

**plot(pricetrend)**

**A shortcut for the first command given in this answer is**

**pricetrend = filter (price, rep(1,13)/13, sides = 2)**

****

B. Subtract the moving average trend from the data series and then plot this difference. This will be a plot of the monthly effects (with trend removed). See Lesson 5.2 for guidance. Give the plot as the answer to this part.



C. Plot a Lowess smoothed trend line using f = 1/3 as the fraction of the data that contributes to the determination of each smoothed value. See Lesson 5.2 for guidance. Give the plot as the answer to this part.



D. Try each of these fractions (on different commands) as the f = in the plot of the Lowess smoother: f = 1/10, f = 1/5, f = 1. Describe what happened to the smoothed pattern as the value of f increased. You don’t have to give the plots – just describe what happened.

**As the value of f is increased, the plot becomes smoother and closer to a straight line. That may not be good. If you plot the original series, you see that the “trend” is more like the moving average plot of part A or the lowess with f = 1/4 in part C.**



E. Fit an ARIMA(0,1,1) model (with no constant) to this data. See Lesson 5.2 for guidance. The R arima command automatically does not include a constant when there’s differencing.

(i) What is the value of the MA coefficient?

**MA coefficient is **

(ii) For an exponential smoothing procedure, what is the value of ? See the ‘ARIMA(0,1,1) and Exponential Smoothing Equivalence’ section of Lesson 5.2 for guidance.

** = 1+0.2891 = 1.2891**

F. See Lesson 5.2. Make necessary changes in the last 6 lines or so of code in order to get a plot of the data and the predicted (smoothed) values in this situation.



G. Keep referring to Lesson 5.2 – in particular, Example 3. Change it to get the forecast of the price at time 101 (next time past the series).

** = 1+0.2891; = −0.2891. The equation is** 

**The data are called “price” and the predicted values are called “smoothed” in the following command to calculate** **: 1.2891\*price[100]-0.2891\* smoothed[100]**

**The forecast of the price at time 101 is** **= 181.7673.**