**Stat 510 Week 5 Homework**

**Submit answers in a Word document or a pdf to the designated ANGEL Drop Box in the Drop Boxes folder.**

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)

$figure

[1] 21.341185 12.054448 -1.347684 -16.384667 -18.285220 -16.565895 -13.730689

[8] -3.889224 0.700346 3.013180 11.619068 21.475152

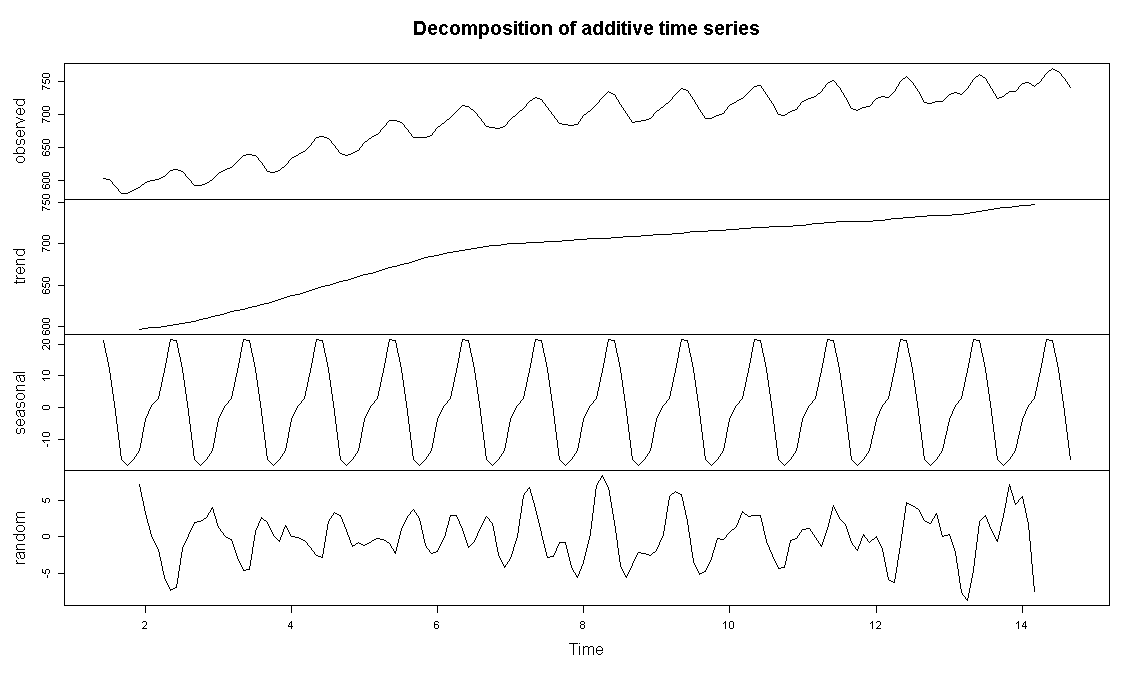
The #1 label is June and so on till the end

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.

The seasonal effect for Oct is: -18.285220

De-seasonalized value = 735 – (-18.285220) = 753.28522

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.

$figure

[1] 1.0307231 1.0174601 0.9981005 0.9763770 0.9735935 0.9760219 0.9801239 0.9943379

[9] 1.0009940 1.0044064 1.0168267 1.0310350

The #1 label is June and so on till the end

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

The seasonal effect for Oct is: 0.9735935

De-seasonalized value = 735/0.9735935 = 754.9351963

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.

The estimated values of seasonal effects are:

|  |  |
| --- | --- |
| Jun | 20.67918 |
| Jul | 12.13306 |
| Aug | -1.13494 |
| Sep | -15.8766 |
| Oct | -18.1798 |
| Nov | -16.2616 |
| Dec | -13.7408 |
| Jan | -3.80394 |
| Feb | 0.86844 |
| Mar | 3.221359 |
| Apr | 11.05595 |
| May | 21.03955 |

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.)

Yes, since an additive model is useful when the seasonal variation is relatively constant over time. We see constant seasonal variation combined with trend in the plot.

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)

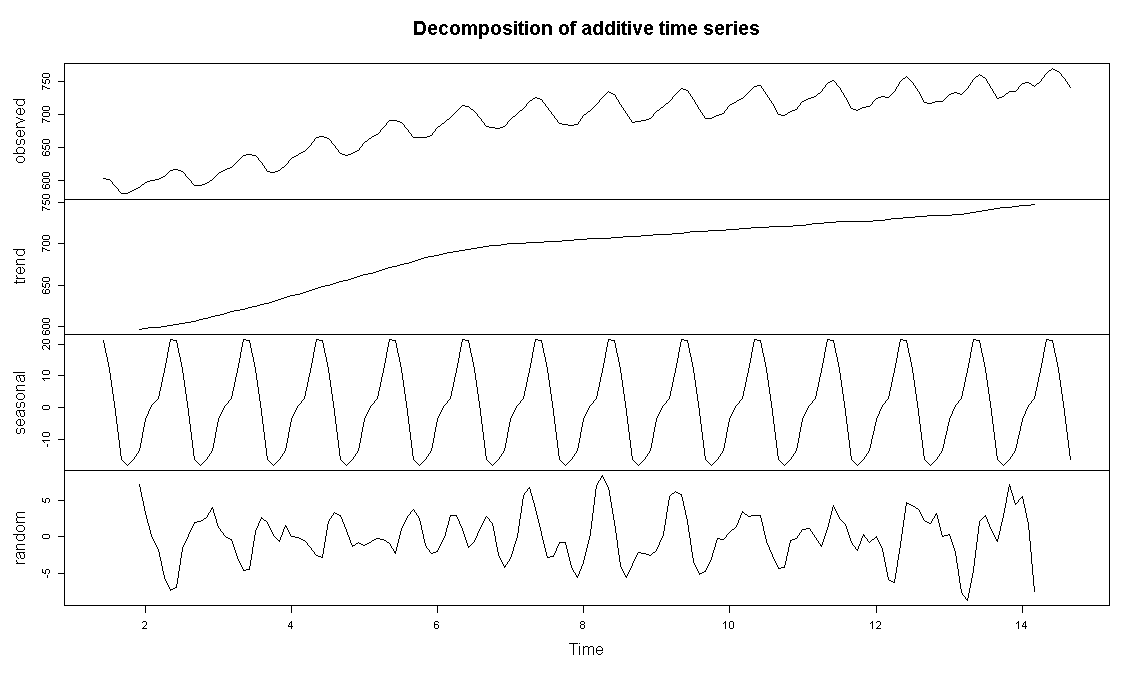
diff1and12=diff(diff1,12)

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

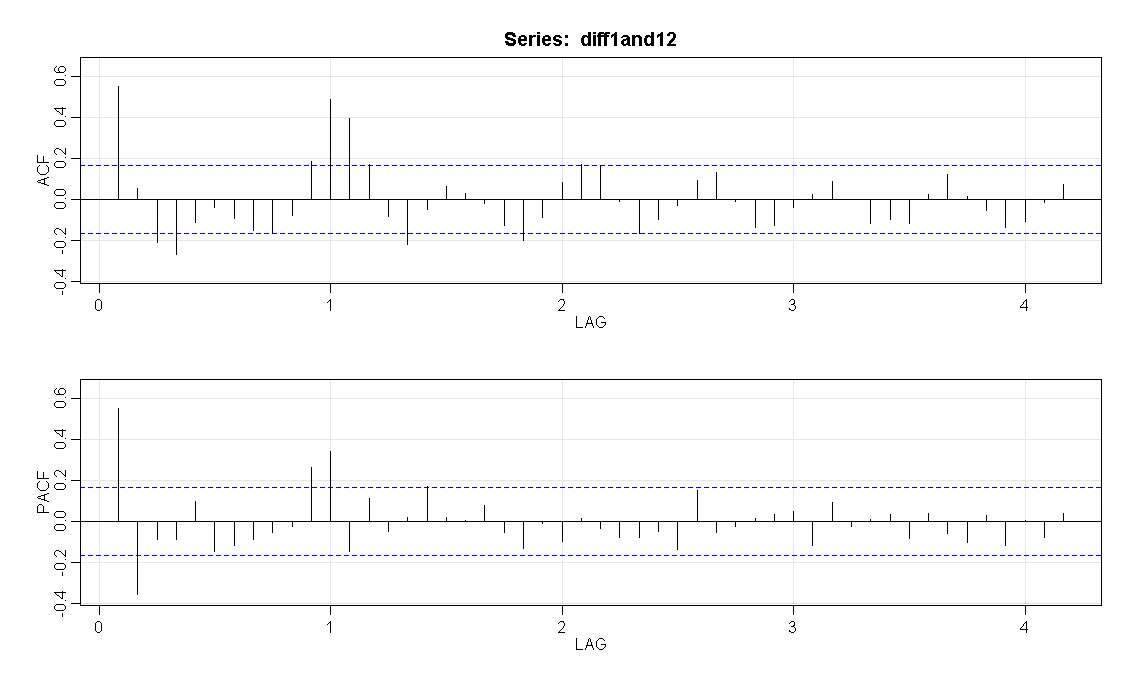
As an answer to this part, briefly explain what differencing was necessary. Then describe the appearance of the acf and pacf (you could just give the plots) and briefly discuss what models are suggested.

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.

From the previous decomposed plot reproduced below, we can see that we have both trend and seasonality and therefore will use 1st and 12th difference.



The ACF and PACF are:



We can see the following possibilities:

**Seasonal**: We see ACF spike at 1st seasonal lag and cutting off after that. Pointing to an MA(1) model.

However we also see PACF spike at 1st seasonal lag and also tailing off of the ACF for the seasonal lags. So seasonal AR(1) is also a possibility.

**Non-seasonal**: We can potentially check MA(1) (because of ACF peak followed by a cut off) or an AR(2) model (because the PACF shows a clear spike at lag 1, 2 and not much else until about lag 11.

So the potential list of models to check are:

ARIMA (0,1,1)×(1,1,0)12

ARIMA (2,1,0)×(1,1,0)12

ARIMA (2,1,1)×(1,1,0)12

ARIMA (0,1,1)×(0,1,1)12

ARIMA (2,1,0)×(0,1,1)12

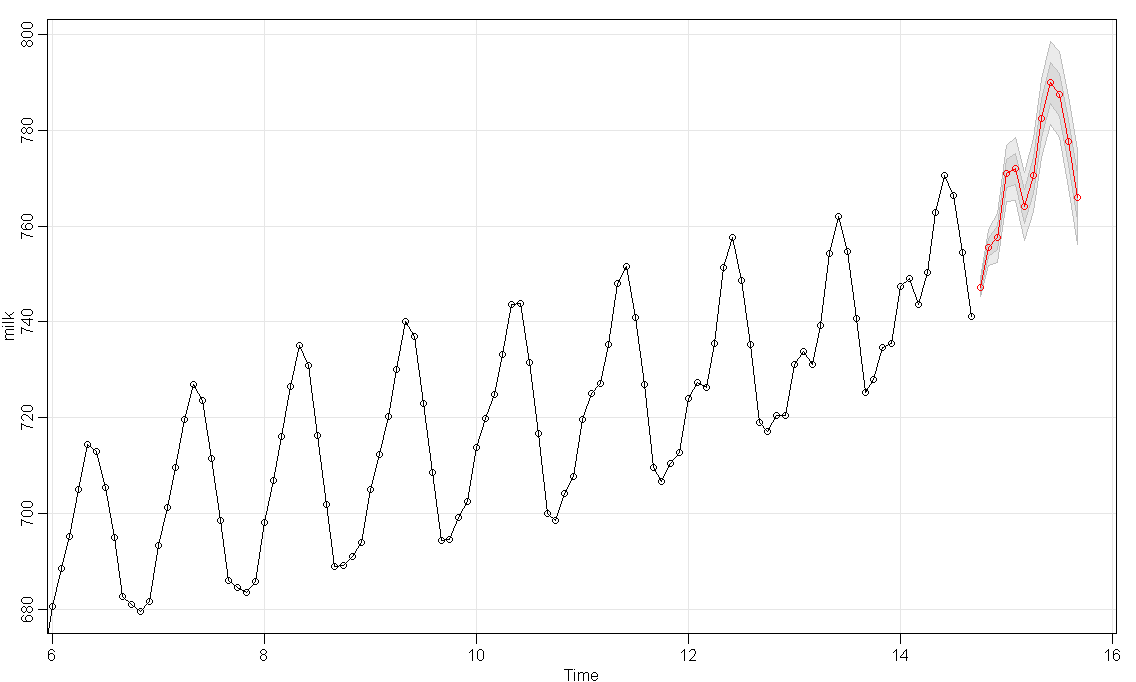
B. Estimate the model (or models) in part a. Examine the diagnostic statistics and decide on a “best” model. As an answer, briefly describe why you think this is a good model for these data.

We will check the following models

|  |  |
| --- | --- |
| Model | Diagnostics |
| ARIMA (0,1,1)×(1,1,0)12 | sigma^2 estimated as 1.156  $AIC  [1] 1.169901  $AICc  [1] 1.183363  $BIC  [1] 0.2083411  Diagnostics:   * Residuals appear fine with constant variance * Many ACF are almost significant – not a desirable result * Normality assumption is reasonable * Ljung statistic is not desirable for many lags |
| ARIMA (2,1,0)×(1,1,0)12 | sigma^2 estimated as 1.103  $AIC  [1] 1.135393  $AICc  [1] 1.149506  $BIC  [1] 0.1930528  Diagnostics:   * Residuals appear fine with constant variance * Many ACF are almost significant – not a desirable result * Normality assumption is reasonable * Ljung statistic is not desirable for many lags |
| ARIMA (2,1,1)×(1,1,0)12 | sigma^2 estimated as 1.103  $AIC  [1] 1.147737  $AICc  [1] 1.162672  $BIC  [1] 0.2246164  Diagnostics:   * Residuals appear fine with constant variance * Many ACF are almost significant – not a desirable result * Normality assumption is reasonable * Ljung statistic is not desirable for many lags |
| ARIMA (0,1,1)×(0,1,1)12 | sigma^2 estimated as 0.8744  $AIC  [1] 0.8907561  $AICc  [1] 0.9042176  $BIC  [1] -0.07080426  Diagnostics:   * Residuals appear fine with constant variance * ACF are not significant – desirable result * Normality assumption is reasonable * Ljung statistic is desirable for all lags except lag 3 |
| ARIMA (2,1,0)×(0,1,1)12 | sigma^2 estimated as 0.8909  $AIC  [1] 0.9220017  $AICc  [1] 0.9361146  $BIC  [1] -0.02033875  Diagnostics:   * Residuals appear fine with constant variance * ACF are not significant – desirable result * Normality assumption is reasonable * Ljung statistic is desirable for all lags |

From the above diagnostics the ARIMA (2,1,0)×(0,1,1)12 is the most desirable.

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.



$pred

Jan Feb Mar Apr May Jun Jul Aug Sep

14

15 770.9576 771.9099 764.1094 770.6028 782.4433 789.8505 787.4010 777.5770 766.0332

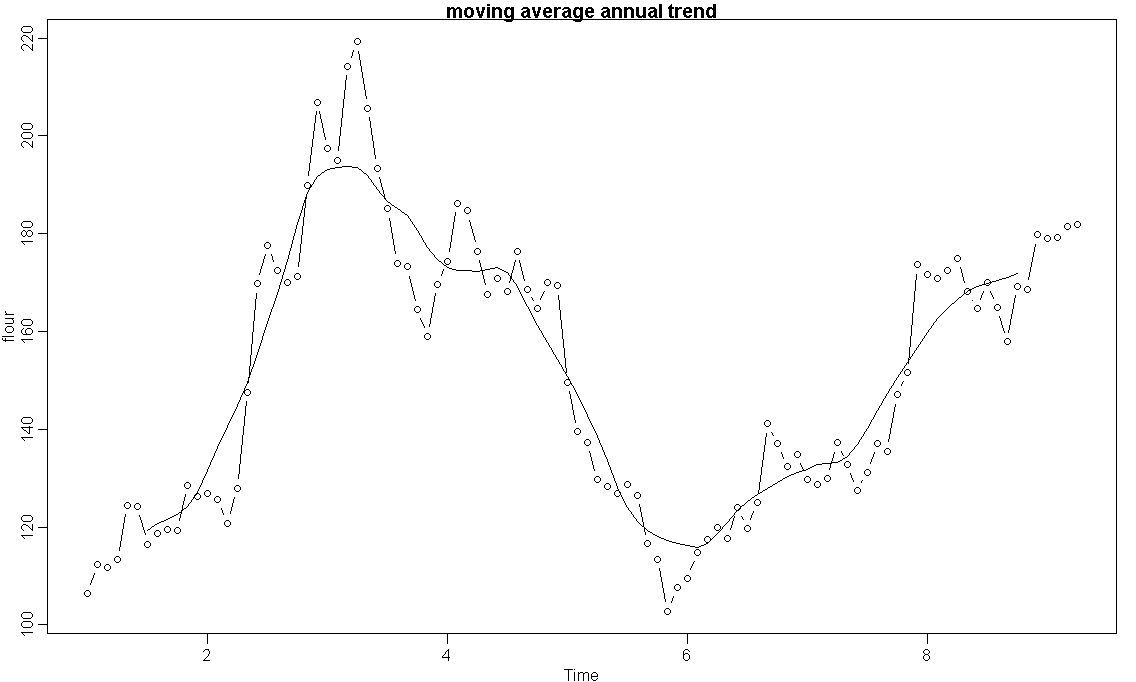
Oct Nov Dec

14 747.1718 755.5728 757.5157

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:



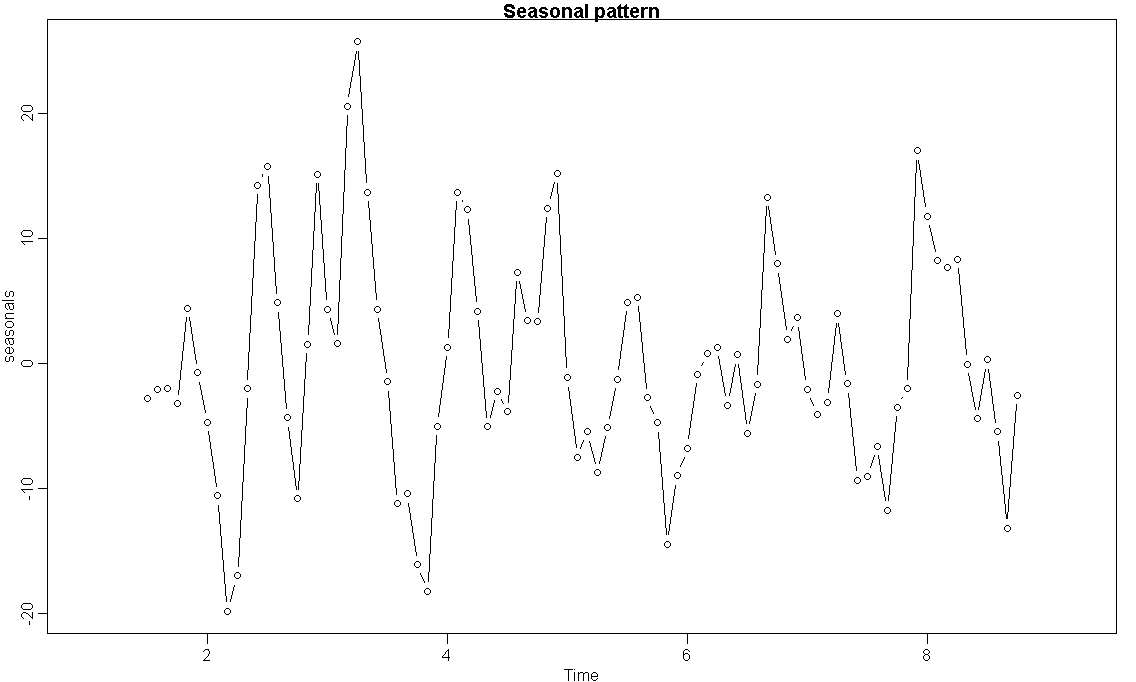
The moving average was determined with R using:

trendpattern =filter(flour, 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)

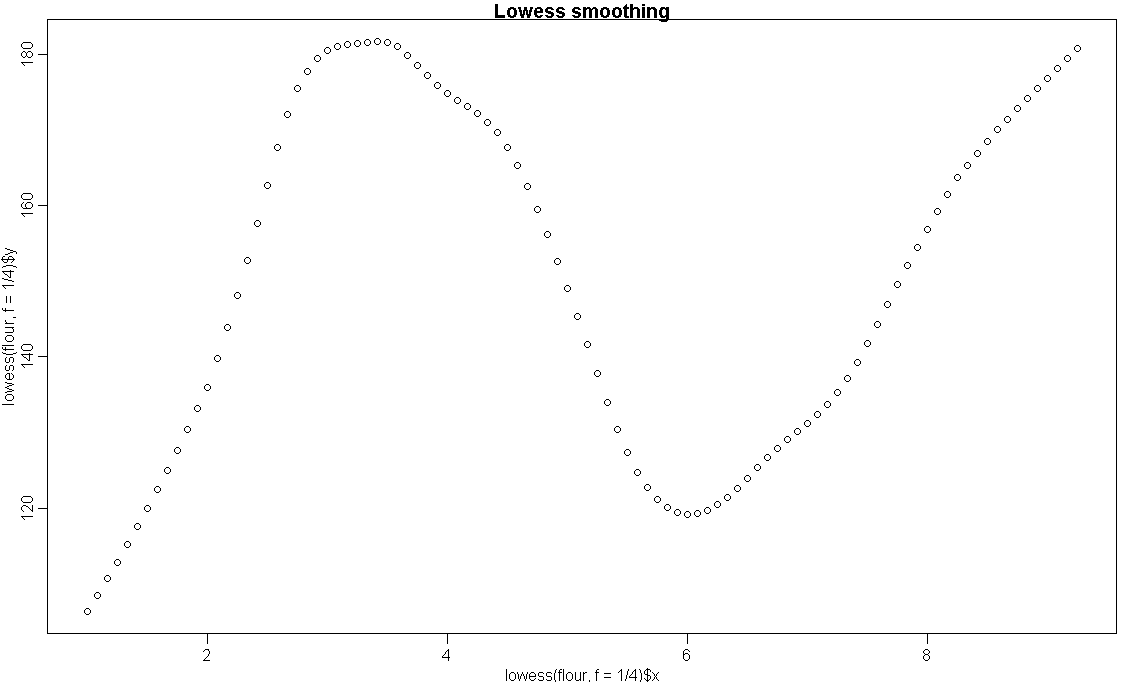
B. Subtract the moving average trend from the data series and then plot this difference. This will be a plot of the monthly effects (with trend removed). See Lesson 5.2 for guidance. Give the plot as the answer to this part.

We use: seasonals = flour - trendpattern

The plot is:



C. Plot a Lowess smoothed trend line using f = 1/4 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 we increased f, we observed that smoothing increases and finer patterns disappear.

The plots are:

|  |  |  |
| --- | --- | --- |
| f=1/10 | f=1/5 | f=1 |
|  |  |  |

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.

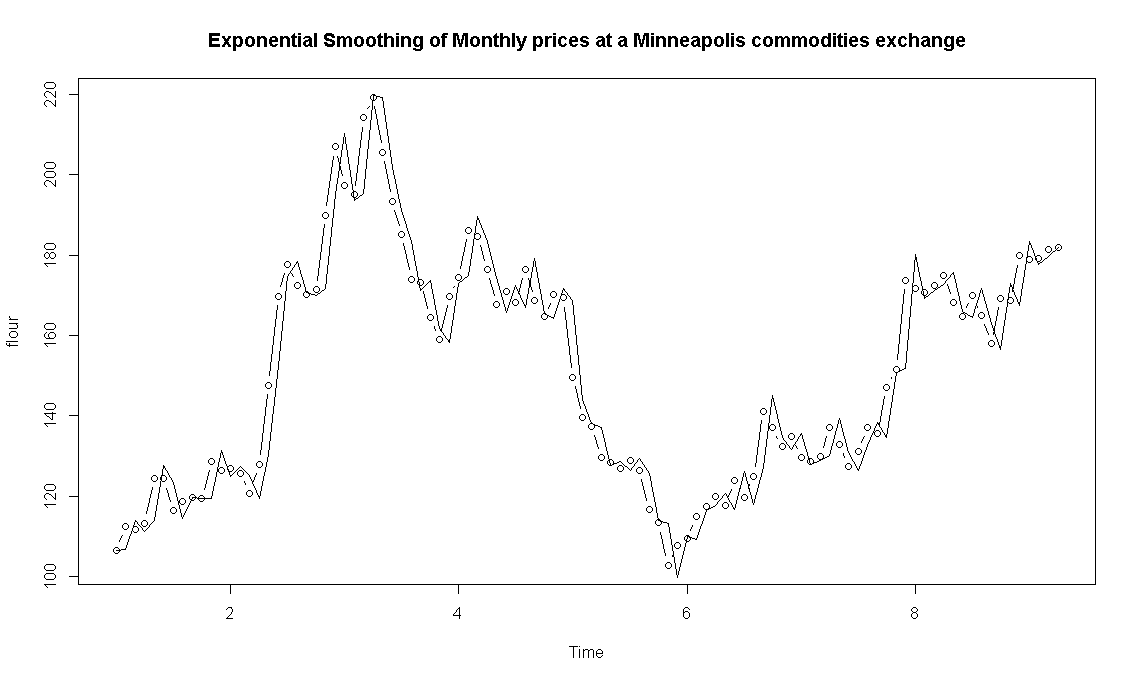
(i) What is the value of the MA coefficient?

(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. An ARIMA(0,1,1) fit in R gave an MA(1) coefficient = 0.2891
2. α = (1+ θ1) = 1.2891 and 1- α = -0.2891

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.

The plot is:



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).

> 1.2891\*flour[100]-0.2891\*predicteds[100]

[1] 181.7673