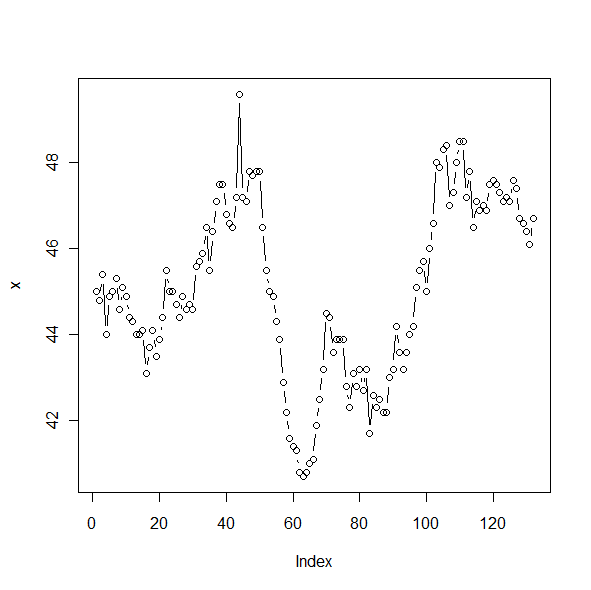
**Stat 510 Week 11 Homework Solutions**

1. Use the “metalsemploy.dat” dataset in the Week 11 folder. The data are proportional to monthly numbers of people employed in metals processing work in a state of the United States. There are *n* = 132 data values. The file can be read in R as x = scan ("metalsemploy.dat")

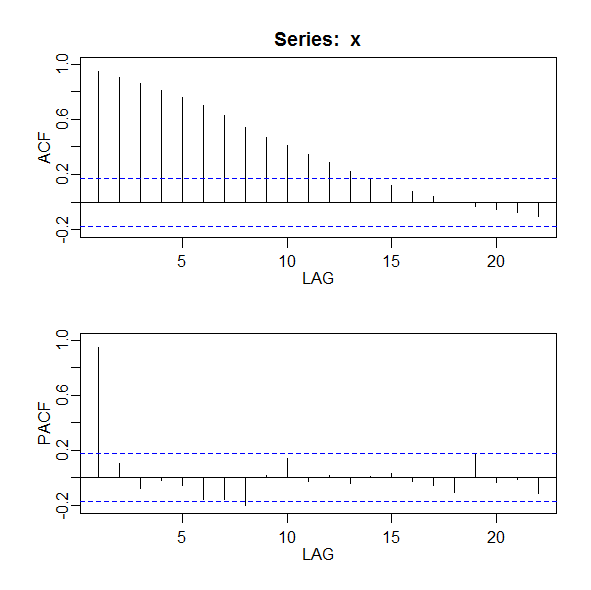
**a.** Create a time series plot of the series. Give the plot and a brief interpretation.

**The time series plot shows a very slow oscillation – perhaps just a random wandering. There are no issues with non-constant variance, outliers, cyclical behavior, or abrupt changes.**



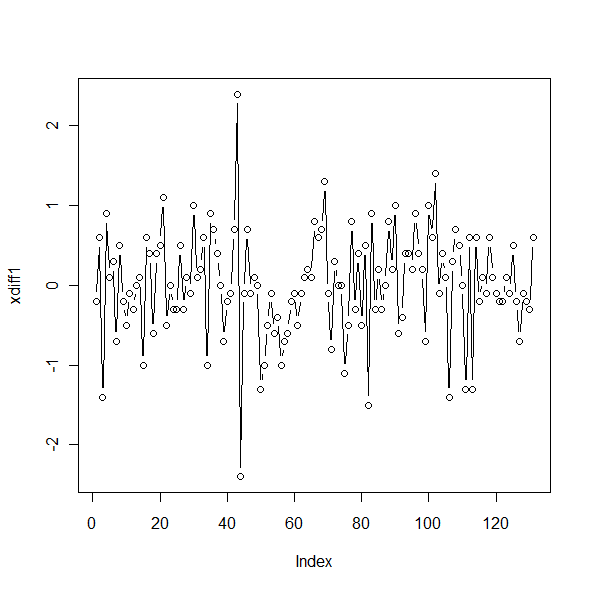
**b**. Determine the ACF and PACF of the series. Write a brief interpretation.

**The ACF and PACF show an AR(1) pattern with a first lag autocorrelation close to 1. Thus the series could be non-stationary, perhaps a random walk given the appearance of the plot in part a.**



**c.** Here’s a hint to parts A and B. The series slowly oscillates, almost in a wandering fashion, and the first lag autocorrelation is very close to 1. This suggests a possible random walk or . Take first differences of the series and create a time series plot of the series of first differences. Give the plot and write a brief interpretation. (Hint: See Lesson 11.1 for this week.)

**A time series plot of the first differences has a rather fast oscillation with some periods of greater variance; no other features are noteworthy.**



**d**. Determine the ACF and PACF of the first differences. What model (if any) is suggested?

**The ACF and PACF of the first differences suggest a random walk. No correlations are statistically significant.**

**e**. Square the first differences. In R this can be done with a command like diffsqrd = diff1\*diff1.

(Exact command will depend upon what you called the first differences.) Determine the ACF and PACF of the squared first differences. What model (or models) is suggested?

**The first differences might have either an MA(1) or an AR(1). For the purpose of an ARCH model we will want to go in the direction of the AR(1).**

**f.** Parts d and e suggest the possibility of an ARCH(1) model for the first differences. One way to estimate the parameters is to utilize the fact that the coefficients for an AR(1) model for the squared values (differences here) are the same as for the ARCH(1) model for the variance. Thus we can fit an AR(1) model to the squared differences to estimate the parameters of the ARCH(1) variance equation. Fit an AR(1) model to the squared differences. What are the estimated coefficients? (Intercept and “slope”.) Note: If you use arima, it mistakenly labels the mean as the intercept. If you use sarima, it correctly labels the mean. You must then calculate intercept = mean \*(1-).

**Coefficients:**

**ar1 xmean**

**0.3150 0.4384**

**s.e. 0.0825 0.0990**

**The slope (phi for the AR(1)) is 0.315**

**The intercept is mean = .4384(1−.3150) = 0.300**

**g**. Part f generated estimates of and for the variance model . Estimate the variance at time *t* = 133. At time 132 (end of the series), the value of = 46.7 and at time 131, =46.1. Thus, the first difference at time 132 is = 0.6. (Note that there are 131 first differences, however the start parameter of the differences vector is set to 2 and so the 131st entry of the differences vector is actually the first difference at t=132.)

**= 0.300+0.315(0.6)2 = 0.4134**

2. Use the files y1hw11.2.dat, y2hw11.2.dat, and y3hw11.2.dat from the Week 11 folder. The files simulate food and beverage price index values for (the same) 400 consecutive months at commodities exchanges in 3 different US cities.

Refer to Lesson 11.2 of this week for guidance on R code for this problem.

A. Estimate the coefficients of a VAR(1) model for the three series. Check the significance of the intercept and trend; include just what is necessary in your final analysis by searching the help file for the VAR command (?VAR). Write the three resulting regression equations (one for each series).

**Trend is not significant. VAR(1) regression equations for the series are:**

**predicted y1t= 187.6 + 0.2670 y1t-1 - 0.1497 y2t-1 + 0.1780 y3t-1**

**predicted y2t = 144.7 − 0.1654 y1t-1 + 0.0781 y2t-1 + 0.1953 y3t-1**

**predicted y3t = 160.9 + 0.2256 y1t-1 + 0.3789 y2t-1 + 0.2900 y3t-1**

B. Estimate the coefficients of a VAR(2) model for the three series; include both, constant, trend, or neither as necessary. Write the equation for City 3 only.

**Trend is not significant; VAR(2) equation for City 3:**

**predicted y3t = 164.8 + 0.2258 y1t-1 + 0.3905 y2t-1 + 0.3009 y3t-1 + 0.0265 y1t-2 − 0.0085y2t-2 - 0.0414 y3t-2**

**Note that none of the lag 2 coefficients are significant, so you should suspect that the VAR(1) model is appropriate at this point.**

C. Use formula 5.90 at the top of page 306 in the text to calculate the BIC value for the VAR(1) model and for the VAR(2) model. Give both values. On the basis of these two BIC values, which model is better? (Lower is better.)

Here’s the format of R commands for the VAR(1) model:

n = 400

k = 3

p = 1

log(det(summary(fitvar1)$covres)) + (k^2\*p\*log(n)/n)

In the formula, k = number of variables and n = sample size, so those will stay the same from model to model. The “fitvar1” within the summary command is the object name that you specify in the VAR(1) fit. That will have to change for the VAR(2) fit (as does the value of p, the order of the AR).

**For VAR(1) BIC = 0.3866**

**For VAR(2), BIC = 0.5402**

**VAR(1) is slightly better in terms of BIC.**

D. For the VAR(1) model, examine the ACF of the residuals for each series. Discuss whether each is white noise (completely random) or not.

**Following Lesson 11.2, all residual series appear to be white noise. There are no significant autocorrelations. acf(residuals(fitvar1)) shows each residual series with itself and each of the other residual series:**

****