**Module 2 Assignment Exercises: Regression and ARMA models**

The assignment for this module is a mixture of programming and written work. Complete this assignment in R Markdown. You will need to include the question and number that you are answering within your submitted assignment. For programming answers using R, answers should be written in R Markdown and ‘knitted’ to a Word/PDF file.

The written work can be answered by filling in your answers on the Module 2 Assignment Word Doc. For math questions this includes the numerical answer, and may also include multiple-choice questions about how you solved the problem.

**Textbook Exercises (Pages 64 and 65):**

**3.1 [10 points]** (Structural Regression Model). For the Johnson and Johnson data, say yt , shown in Figure 1.1, let . In this problem, we are going to fit a special type of structural model, , where T is a trend component, S is a season component, and N is the noise. In our case, time t is in quarters (1960.00, 1960.25) so one unit of time is a year.

1. Fit the regression model:where  if time t corresponds to quarter i - 1,2,3,4 and zero otherwise. The ’s are called indicator variables. We will assume for now that wt is a Gaussian white noise sequence. Hint: Detailed code is given in Appendix A, near the end of Section A.5.
2. If the model is correct, what is the estimated annual increase in the logged earnings per share?
3. If the model is correct, does the average logged earnings rate increase or decrease from the third quarter to the fourth quarter. And, by what percentage does it increase or decrease?
4. What happens if you include an intercept term in the model in 1.? Explain why there was a problem.
5. Graph the data,  and superimpose the fitted values, say , on the graph. Examine the residuals, , and state your conclusions. Does it appear that the model fits the data well (do the residuals look white?).

**3.2. [10 points]** For the mortality data examined in Example 3.5:

1. Add another component to regression in (3.17) that accounts for the particulate count four weeks prior; that is add  to the regression in (3.17). State your conclusion.
2. Using AIC and BIC, is the model in (a) an improvement over the final model in Example 3.5?

**3.3. [10 points]** In this problem we explore the difference between a random walk and a trend stationary process.

1. Generate *four* series that are random walk, with drift, (1.4), of length n=500 with  and . Call the data  for t=1,...,500. Fit the regression  using least squares. Plot the data, the true mean function (i.e., ) and the fitted line, , on the same graph.
2. Generate *four* series t of length n=500 that are linear trend plus noise, say , where t and w are as in part 1. Fit the regression  using least squares. Plot the data, the true mean function, (i.e. ) and the fitted line .
3. Comment on the differences between the results of part 1) and part 2).

**3.4. [10 points]** Consider a process consisting of a linear trend with an additive noise term consisting of independent random variables  with zero means and variances , that is, , where  are fixed constants.

1. Prove  is nonstationary.
2. Prove that the first difference series  is stationary by finding its mean and autocovariance function.
3. Repeat part (2) if  is replaced by a general stationary process, say , with mean function  and autocovariance function .

**3.8. [10 points]** In section 3.3 we saw that the El Nino/La Nina cycle was approximately 4 years. To investigate whether there is a strong 4-year cycle, compare a sinusoidal (one cycle every four years) fit to the Southern Oscillation Index to a lowess fit (as in Example 3.18). In the sinusoidal fit, include a term for the trend. Discuss the results.