STAT 2321:
Applied Time Series
Final Exam

Due by 5 pm on Saturday 12/17/11
in Dr. Wallace's mailbox.

Instructions: You may use course notes and the text book. All computing should be done in R. Please insert your R output and code into your document. When you are finished, put the exam in my mailbox or give it to me in person.

Expectations: 1) thoughtful investigation and discussion surrounding model selection, 2) choice of appropriate plots, models, and statistics, 3) insightful interpretations of all plots, models, and statistics, and 4) organization and clarity.

Important: This exam must be your independent work. You are not to discuss or share the material in this exam with ANYONE before 5 pm on Saturday, December 17th, 2011. Sharing and/or discussing your work will result in a failing grade for the final.

1. The data set `alc.csv` contains the monthly amount of alcohol wholesale (in $1,000,000$'s) in the U.S. from January 2002 until December 2008.

   (a) Plot the data and discuss what you observe.

   (b) Build and fit a stochastic model to the alcohol data. Document and discuss the process that led you to chose this model and report all appropriate regression diagnostics.

   (c) Plot the data along with the predicted monthly alcohol sales for each month in 2009. Also include 95% prediction intervals. Discuss whether the predictions seem reasonable based on the past data trends. Which months are expected to have the lowest and highest alcohol sales?

2. The data set `homicide.csv` contains the monthly count of homicides in Cali, Colombia from January 1999 through August 2008.

   (a) Plot the data and discuss what you observe.

   (b) Build and fit a stochastic model to the homicide data. Document and discuss the process that led you to chose this model and report all appropriate regression diagnostics.

(c) Plot the data along with the predicted homicide for the first 12 months of 2009 and 95% point-wise prediction intervals. Discuss whether predictions seem reasonable based on the past data trends and your selected model.

3. The data set `erie.dat` contains monthly Lake Erie levels from 1921 through 1970.

   (a) Plot the data and discuss what you observe.

   (b) Build and fit a stochastic model to the data. Document and discuss the process that led you to chose this model and report all appropriate regression diagnostics.

   (c) What are the predicted Lake Erie levels for March of 1971, 1972, and 1973? How confident are you in each of these predictions?

4. In a gas furnace, air and methane are combined to form a mixture of gasses, including co2. The data `methane.dat` contains the methane input (cu. ft/min) into a gas furnace; the data `co2.dat` contains the contemporaneously measured percentage of carbon dioxide measured in the output. Sampling occurred every 9 seconds.

   (a) Plot the time series data sets over time and discuss what you observe.

   (b) Plot and interpret the cross-correlation function for the methane and co2 time series. From this plot, how would you summarize the relationship between methane input and co2 output?

   (c) Use the cross correlation function results as well as any other helpful plots and/or statistics to build and fit a lagged regression model to predict co2 output. Document and discuss the process that led you to chose your model. Report all appropriate regression diagnostics.

   (d) Linear models based on time series data often do not meet regression assumptions because they have correlated residuals. If your model in part (c) had correlated residuals, describe how you could adjust it to account for the fact that the error term is not actually white noise. Please do this regardless of whether or not your model actually had correlated residuals. (No need to fit any additional models.)

   (e) Discuss and interpret the model results from part c, taking care to consider whether or not regression assumptions were met. In particular, summarize the relationship between methane input and co2 output based on the model results. How does this relationship compare to your answer in part b?