Fall 2016 MSAN 604

Assignment 2

Due Date:

Submit a hard copy in person to me by **5:00pm on Thursday, November 17th**

Instructions:

Indicated below are a series of problems from *Introduction to Time Series and Forecasting*, 2nd edition. A teaching assistant and I will grade a random subset of these problems. The "Additional Questions", on the other hand, will both be graded. To facilitate efficient grading, your weekly homework should satisfy the following properties:

1. Turn in the problems in the order in which they were assigned.
2. Clearly indicate your name and student ID number on the upper right-hand corner of the first page.
3. Staple your homework assignment in the upper left-hand corner.
4. In general, do not print out entire data sets.
5. In general, do not print out reams of R output. Everything should be orderly and easy for me to read.

Textbook Problems:

3.1, 1.10, 1.15

Additional Problems:

1. ARIMA fitting with the `LakeHuron` dataset.

The `LakeHuron` dataset contains annual measurements of the level, in feet, of Lake Huron between 1875 and 1972.

(a) Take ordinary differences of the data until the resulting time series is suitably stationary, using the Augmented Dickey-Fuller Test to aid in your choice of d.

(b) Fit an $AR(1)$ model to the d-differenced `LakeHuron` data, using maximum likelihood estimation.

(c) Fit an $AR(2)$ model to the d-differenced `LakeHuron` data, using maximum likelihood estimation.

(d) Perform a likelihood ratio test comparing the $AR(1)$ and $AR(2)$ models and state whether or not the null hypothesis is rejected.

(e) Using your results from (d) and an assessment of AIC and $\hat{\sigma}^2$ for each model, state which model you deem to be "optimal" and briefly explain this choice.

(f) Using appropriate formal and informal residual diagnostics, investigate whether the "optimal" model you've chosen in (e) satisfies the following assumptions
    i. Zero-Mean
    ii. Homoscedasticity
    iii. Zero-Correlation
    iv. Normality

2. SARIMA fitting with the `beers.csv` dataset.

The `beers.csv` dataset contains information on monthly Australian beer production, in millions of litres, from January 1956 to August 1995.

(a) Using techniques discussed in class, fit (using maximum likelihood estimation) an "optimal" $SARIMA(p,d,q)\times(P,D,Q)_s$ model to the `beers.csv` time series, and justify your choice of p, d, q, P, D, Q and s.

(b) Using appropriate formal and informal residual diagnostics, investigate whether the "optimal" model you've found in (a) satisfies the following assumptions
    i. Zero-Mean
    ii. Homoscedasticity
    iii. Zero-Correlation
    iv. Normality