CSC425 Time series analysis and Forecasting

Rules and guidelines for the final project
Due by 11:59 p.m. on Wednesday March 13th, 2013
Total Point: 30

The project must be completed in a group of max three students. In-class students are encouraged to create teams with online students. Students looking for group partners can post messages on our discussion forums, or send me an email. Online students have the option of completing the final project by themselves, if they so prefer.

The project consists of the analysis of a time series of your own choice and interest (Do not use any of the series analyzed in the textbook). Time series can be found for instance at
   − the Federal Reserve bank of St. Louis
      (http://research.stlouisfed.org/fred2/)
   − Time Series Data Library
      (http://datamarket.com/data/list/?q=provider:tsdl)
   − Yahoo! Finance
      (http://finance.yahoo.com/)
   − Economagic.com (need to delete advertisements in data)

The goal of this project is to apply the notions we discussed during the quarter. Use the techniques of data exploration and testing to analyze the properties of the selected time series and to identify an adequate model to explain the data. Select the "best" model to explain the time process that produced the data. Check the adequacy of the model using residual analysis and model diagnostics techniques.

Use back-testing procedures to validate the selected model on a testing set not used to fit the model. See Week 4 notes and examples for a discussion on back testing procedures. R users should use the backtest() function discussed in week 4 notes. SAS users can discuss with me alternative ways to run backtesting, if the SAS code posted under week 4 fails to converge.

REQUIRED EMAIL: Each group should send me an email at rsettimi@cdm.depaul.edu **before Friday February 22nd**, with
   1) the list of names of students in your group
   2) a short description of the selected time series.
The description should also contain information on the source where the data were taken.
Failure to do so will result in 2 points taken off the final project grade.

PRESENTATION: The final project will be presented in class at 5:45pm on Wednesday March 13th (in week 10). The project presentation is required for the inclass students, but it is optional for online students.

Project report
The group will write a single analysis of the results using a word processor. Each group member should also submit a detailed statement of the distribution of work among the group. Group

members not participating fully in the group work may receive less credit than other members of the group.

The analysis should be submitted as a report consisting of two sections:
1. **A non-technical summary** not longer than one typewritten page, describing the conclusions of your statistical analysis to a non-technical audience. It should be intelligible to a person who does not know regression analysis. Suppose you are talking to your boss who does not know statistics or to a friend who is not familiar with statistical terminology.
2. **A technical summary** not exceeding 5 pages, with details of your statistical analysis. This section is intended for a statistically literate audience and must be written in a <u>clear organized fashion</u>. For instance, you can organize the report into subsections such as
   a. Exploratory analysis of the data.
   b. Model fitting.
   c. Residual analysis and model diagnostics.
   d. Forecast analysis
   e. Analysis of the results and discussion.

You should include appropriate output and graphs from your SAS or R programs in your document in a presentable format. Further details may be relegated to an **appendix** that may contain the SAS/R code, some graphs, computer output or supplementary information about the field of study.

Clarity and synthesis in the report are important and will be rewarded. It is essential to be able to communicate your analysis to other people. You must explicitly quote any result you will be using in your analysis.

It is possible that you may not find a satisfactory model that fits adequately your data. Sometimes a data set may admit more than one satisfactory answer; sometimes there may be none. If the statistical analysis shows that no time series models are suitable for your data set, please mention what approaches you have tried and what was unsatisfactory about them. If there is more than one suitable model, mention the pros and cons and compare their performance in forecasting future observations.

If you have any questions or problems, feel free to consult me at any time.

Grading information – detailed rubric is enclosed below:

1. Received information of group members by due date (2pts)
2. The non-technical summary provides a good non-technical presentation of the results using a language that can be understood by someone who is not familiar with statistics.(3 pts)
3. The statistical analysis is complete and accurate, and includes all the appropriate steps (exploratory analysis, modeling, diagnostics, forecasting) (15 pts)
4. Quality of presentation and use of slides (5 pts)

5. The report or technical summary is well written and organized (e.g. use sections, subsections, good labels for tables and graphs…) and describes appropriately the statistical analysis. (10 pts)
6. The Appendix contains the graphs/ statistical output that did not fit in the report. The appendix is well organized. (2pts)
7. The SAS/R code is included in the Appendix, and the data file is attached to the submission (1pts)
8. Each student should submit the evaluation of the contribution of the other group members at the course website. There will be a link added to the homework submission page. You can also use the evaluation sheet posted online if you want.(2 pts)

Final Project Grading Rubric

	Excellent	Good/Fair	Poor
Written report Layout and clarity	Report is clear and neatly organized. Very good layout with appropriate use of headings and possibly subheadings Appropriate use of grammar and statistical terminology. (3pts)	Report is mostly clear with parts that are not well organized under sections. Report is well written, with a few errors in grammar. Use of appropriate statistical terminology is limited. (2pts)	Report is not clear. The layout is cluttered, confusing, and does not use spacing or headings to enhance readability. Major editing and revision are required. Errors in spellings capitalization, punctuation and grammar distract readers. (1pts)
Non technical summary	Summary describes the conclusions clearly, concisely and is written in a language that is appropriate for a non-technical audience (3pts)	Summary describes the conclusions clearly, concisely but contains statistical jargon that is inappropriate to a non-technical audience (2 pts)	Summary is unclear and uses technical terms and language that is inappropriate to a non-technical audience (1 pts)
Statistical analysis			
Exploratory data analysis	The analysis explores the distribution of time series, the serial correlation and stationarity through graphs and appropriate	The analysis of the distribution of the time series is incomplete or is incorrectly interpreted. 2pts	The analysis of the distribution of the time series contains major flaws and lacks important parts. 1pts

	statistics. 3 pts		
Modeling	Selected time series model is an adequate fit for the data. 3 pts	Selected time series model is appropriate, but there might be a better model for the data. 2pts	Time series model is a bad fit for the data, analysis contains major flaws, steps are missing. 1pts
Diagnostics/residual analysis	Model assumptions are checked appropriately through a complete residual analysis. 3pts	Model assumptions are checked but some steps in residual analysis are missing or incorrect. 2pts	Major flaws in residual analysis. 1pts
Forecasts	Forecast values and intervals are correctly computed 2pts	Forecast values and intervals contain some errors 1pts	Forecast values and intervals are not computed 0pts
Conclusions	Results are correctly interpreted and conclusions are clearly reported. 3pts	Results interpretation contains one or two errors. 2pts	Interpretation results contains three or more errors, or is missing. 1pts
Presentation (only for in class students)	Presentation is clear and well organized – excellent use of visual tools (3 pts)	Presentation is clear, but not very well organized – poor use of visual tools (2pts)	Presentation is disorganized and unclear – (1pts)
Appendix (optional) – points will be assigned elsewhere.	Well organized – graphs and output are labeled appropriately. (2pts)	Not well organized – graphs and output have missing labels (1pts)	Cluttered and confusing (0pts)
SAS/R code + data set	SAS/R code is included in appendix and data file is attached to submission(1pts)	SAS/R code is missing (0pts)	
Evaluation	Submitted (2pts)	Not submitted (0pts)	
Group members	Information received by due date (2pts)	Information not received by due date (0pts)	
Total	**30 pts**		