Homework 4 - Midterm

Ma641 Time Series I

due by 10:00pm, Monday July 13, 2009

You can hand in the assignments either in class at the beginning of the lecture or using the elearn page. If you chose elearn please convert the report to pdf format before submitting.

This assignment counts as the Midterm too. It is worth 300 total points.

1. (50 points) Solve problems 2.1 and 2.2 on page 93.

2. (100 points) For this problem we refer to the attached files: *"SPclose6monthsJune24.csv"*, *"SPclose18monthsJune24.csv"*, and *"SPclose30monthsJune24.csv"*. These files contain daily data on S&P500 for the periods Jan 2009-June 2009, Jan 2008 - June 2009 and Jan 2007 - June 2009. A great concern for time series estimation as you have discovered in the previous assignments is that one needs a large number of observations from the time series for a reliable estimation and forecast. On the other hand the data has to be stationary for a reliable estimation and in general it is believed that data stays stationary only for short periods of time.

   Here we will study the performance of various models and the effect of the data length on the estimation.

   Repeat each of the following for all the three data sets.

   (a) Work with continuously compounded returns. Pay attention to the date order in your data.

(b) First test the specific type of non-stationarity provided by the unit root models (i.e. perform a unit root test). Compare for all 3 data periods.

(c) Build an AR, an MA and a straight ARMA model after you remove non-stationarity (if needed) for each time period.

(d) Build a seasonal model for your data using whatever lag you feel approximates the data best.

(e) Using each of the models that you constructed forecast the next seven observations. 24 June is a Wednesday so you are forecasting the daily adjusted closing data for the current week and the next. Use the three data sets separately when building the models. All in all you should have three categories of models. The models in each category should correspond to the respective time period length.

(f) At the time when you write this report look back at the adjusted closing price for the respective days and compare with the forecasted values by each of the models that you constructed. Calculate the squared error of your forecast for each model and the absolute value of your forecast. What was the better model in retrospective?

(g) Was there a difference in the three time periods (i.e., is there a type of model that consistently outperformed the others for all the datasets under consideration)?

(h) Are the perceived differences between models due to chance or there is a model that performed significantly better? (Hint: Use the standard error of the forecast as given by the R output.)

3. (150 points) For the final problem use the high frequency data provided *"INTCdata.xls"* (minute data for Intel Corporation). The data is from two years ago for the period June 19-June 30 2006. The data format is identical with what we used in the first assignment. I suggest you use the column 6 of the data (closing minute data) or column 9 (the average data within the minute). These are two weeks (notice that there are only 10 days worth of data due to the weekend). Construct the best model you can with the material taught until now. You should discuss a lot. Pay attention to the fact that the data contains a weekend. It

is up to you how to divide the data to evaluate the performance of the models.

Notice that this problem is worth half the total points of this assignment.