Objective: To gain experience in using alternative approaches to volatility calculation, neural network, and analysis of high-frequency data.

Due Date: May 29 (campus class) and May 31 (Weekend class), 2014.

1. The goal of this question is to study the daily log return and its volatility of Netflix stock. Focus on the period from January 3, 2006 to April 30, 2014. You may download the daily open, high, low, close prices and related information from Yahoo via the quantmod package. The tick symbol is NFLX.

Use the data to construct the (conditional) variance estimates $\sigma^2_{t,i}$ of Section 3.15.2 of the textbook for $i = 0, 1, 2, 3, 5, \text{and } 6$. These conditional variances are for the stock price. Take the square-root transformation to obtain the volatility series. Multiple the volatility by $\sqrt{252}$ to obtain annualized volatility. Obtain mean, median, maximum, and minimum of each of the six annualized volatility series.

2. Again, consider the Netflix stock, but focus on the volatility of the daily log returns. Use the data and $n = 63$ to compute the Yang and Zhang (2000) variance estimate $\hat{\sigma}^2_{yz}$ of Section 3.15.2 of the textbook. Obtain a time plot of the estimated (and annualized) volatility series (square-root of annualized variance). Build a time series model to produce 1-step to 5-step ahead volatility forecasts at the forecast origin April 30, 2014.

Hint: You can consider the log series of the volatility in modeling.

3. Consider the monthly stock returns of IBM and the S&P composite index from January 1941 to December 2013. Data are in the file m-ibmsp4113.txt. Define the direction of price movement of IBM stock as follows:

$$ M_t = \begin{cases} 
1 & \text{if } r_t > 0 \\
0 & \text{otherwise}, 
\end{cases} $$

where $r_t$ is the simple returns. Similarly, we can define the direction for market movement using the S&P composite index returns. Denote the direction by the market by $S_t$.

(a) Fit a linear logistic regression model for $P(M_t = 1)$ using $M_{t-1}, S_{t-1}, M_{t-2}, S_{t-2}$ as input. Based on the model, can past price movements of either the stock or the market predict the future PG price movement? Why?
(b) Use \((M_{t-i}, S_{t-i})\) for \(i = 1\) and \(2\) to build a 4-3-1 look-forward network with direct link for \(P(M_t = 1)\). Write down the fitted model.

(c) Divide the sample into modeling and forecasting subsamples with the latter consisting of the last 84 observations. Use 1-step ahead prediction to compare the previous two models. You can treat \(\hat{M}_t = 1\) if \(P(M_t = 1) > 0.5\).

4. Consider the tick-by-tick trade data of Exxon-Mobil stock from September 1 to September 30, 2013. The sample size is large with 1,081,154 transactions.

- Use the data within the normal trading hours only, i.e. from 9:30 am to 4:00 pm Eastern time, to construct a series of intra-day 5-minute log returns. If there is no trading within a 5-minute interval, assume that the log return is zero. If there are multiple trades in a 5-minute interval, use the last trade to obtain the price for that interval. Plot the log return series.

- Are there any serial correlations in the intra-day 5-minute log return series? Use \(Q(10)\) to perform the test.

- Use 5-minute intra-day log returns to compute the realized volatility for each of the trading days.

- Use 1-minute intra-day log returns to compute the realized volatility for each of the trading days.

5. Again, consider the tick-by-tick trade data of Exxon-Mobil stock from September 01 to September 30, 2013.

- Construct the series of the number of trades within a 5-minute intervals. Use data in the normal trading hours only.

- Compute the ACF of the constructed time series, say from lag 1 to lag 310. Is there any evidence of diurnal pattern? [No formal test is needed. Simply comment on the ACF plot.]

Reading assignments: Sections 4.1, 5.1 to 5.4 of the textbook.