Lab Assignment 1, ORFE 569
Due Feb. 15, 2007

Objective: The objective of this assignment is first to get you familiar with tick data. You will learn where to get data and what variables are available, filter the data, put the data into usable format, explore the data via graphs and descriptive statistics.

Where to get data Go to *Wharton Research Data Service* at `https://wrds.wharton.upenn.edu/`. A class account has been set up. Login names: ***** and password: *******. Choose **NYSE TAQ** database (namely, New York Security Exchange, Trades and Quotes database).

Two Data Sets NYSE TAQ has five types of data. We will focus on the first four "Consolidated Trades", "Consolidated Quotes", "NBBO" and "NBBO+Trades". You are required to obtain **two** data sets. The first data set must be from "Consolidated Trades" for a company in a period. For example, APPLE for July 15 - 17, 2005. (Note that Princeton does not subscribed 2006 TAQ data, but 1993 - 2005 available.) "Trade" data set (with trading volume) must contain at least 30,000 data, but not more than 100,000. You can choose one of the other three types for the second data set. (Send a group email to let others know which company you choose so that no one else will choose the same company). Or, if you have other tick data set such as bonds or exchange rates, you may use them also.

A "clean" bond tick data from GovPX, which some published papers (e.g. in *Journal of Finance*) used, is available for a particularly interested student. Please contact me.

Process Data to Suitable Format The main task is to convert the time stamp to a suitable time in seconds (after setting an initial time as zero). Sometimes, you may have to filter or correct some "wrong" data. A SAS program on the course website is available for your reference. You can use whatever software such as R, or others to accomplish this task.

For "Trade" data set, the format is three rows: Time, Price, Volume.
For "Quote" data set, the format is: Time Bid offer (for example).

Describe Each Data Set with tables and graphs. A sample S program is available on the course website. The following is minimal and write short comments for each table or graph you produce.

1. A table to present basic descriptive statistics: number of data, min, max, Q1, Median, Q3, mean, standard deviation, skewness, and kurtosis for price, volume, bid and offer quote.

2. Plots of price (quote, volume) with time for the whole time frame and for a small time frame to show discreteness.

3. Histograms for price (quote) change and for the fractional part of price (quote).

4. Histograms for duration (between trades) (or intertrade-occurrence time)

5. Histograms for trading volume.

6. For one day (9:30am - 4:00pm) and every 5-minute interval, calculate the volatility (or standard deviation) of price, or average offer-bid spread (namely, offer - bid). Plot them to see whether there is a "U" shape.

Be creative to produce extra pictures! If a picture is impressive and make a good point, you may get **extra credits**!