Financial Econometrics
FINN/ECON 6219, Fall 2017

Problem Set 1
Getting started with data analysis in MATLAB
Due September 6 (hard copy, beginning of class)

Problem 1.

The file *sizeDecileReturns.xls* contains daily returns (in *decimal* format, i.e., 0.01 indicates a 1 percent return) for the Center for Research in Security Prices (CRSP) size portfolios for a 5-year period. To construct these returns, CRSP sorts all NYSE, AMEX, and NASDAQ firms into ten portfolios on the basis of their market capitalization (size). The first portfolio (decile 1) contains the first 10% of the firms in the sorted list (i.e., the small firms), the second portfolio (decile 2) contains the next 10% of the firms in the sorted list, etc.

Read these data into MATLAB and perform the following analysis.

a. MATLAB has many pre-defined statistical functions. Use the mean, max, and min functions to compute the mean, maximum, and minimum of the *percentage returns* for each of the ten portfolios. Report the results.

b. Some pre-defined functions return two or more outputs. One example is sort, which can be used to sort a vector or matrix (column-by-column). It delivers the sorted data as the first output and the indices of the sorted data as the second output. Sort the *percentage returns* for decile 1. Report the value and index of the largest *percentage return*. Sort the absolute values of the *percentage returns* for decile 10. Report the value and index of the smallest absolute *percentage return*.

c. Many pre-defined functions take two or more inputs. Consider corr, which computes the correlation between variables. Find and report the correlation between the returns for decile 10 and the returns for each of the other deciles (you might want to look at the help file for corr).

d. Sometimes we call a function using only some of its possible inputs. Consider std, which computes the standard deviation of variables. It has the data as its first argument, a flag for bias correction as its second argument, and the dimension to be analyzed as its third argument. Find the standard deviation of (i) the squared *percentage returns* for each decile without bias correction, and (ii) the *percentage returns* across each of the days in the sample with bias correction (you might want to look at the help file for std). In part (ii), you only need to report the standard deviation for the first day in the sample.

e. User-defined functions play an important role in model fitting and parameter estimation. Construct a function that takes three scalar arguments, x, μ and σ^2, and computes the log likelihood of a normal random variable, i.e., <u>the natural logarithm of</u>

$$f(x;\ \mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Save this function as normal_log_likelihood.m and compute the log likelihood of $x = 0$, $\mu = 0$, and $\sigma^2 = 1$.

f. The log likelihood function of a vector of N independent Normal random variables drawn from the same distribution is the natural logarithm of

$$f(x_1, x_2, \dots, x_N;\ \mu, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Modify the function of part *e* to treat x as a vector of independent observations instead of a scalar. Save this function as modified_normal_log_likelihood.m and compute the log likelihood of the *percentage returns* for decile 5 with μ set equal to the sample mean of the *percentage returns* and σ^2 set equal to the sample variance of the *percentage returns*. Report the result.

.

Problem 2.

Standard normal random variables can be simulated using the command randn(*N*,*M*) where *N* and *M* denote the dimensions of the desired matrix of pseudo random numbers.

a. Generate 10000 realizations of a standard normal random variable and store these data in a 10000 x 1 vector named e. Initialize a 10000 x 1 vector named Y that contains zeros using the function zeros. Use a for loop to simulate a process of the form

$$Y(t) = (1 - \alpha) * Y(t - 1) + e(t)$$

for t=1,2,...,10000 with $\alpha = 0.1$ using the initial condition Y(0)=0. Use the plot function to plot the resulting time series.

b. Use a for loop along with an if statement to simulate a process of the form:

$$Y(t) = (1 - \alpha) * Y(t - 1) + \left(1 + I_{[e(t-1)<0]}\right) e(t)$$

for t=1,2,...,10000 with $\alpha = 0.1$ using the initial conditions Y(0)=0 and e(0)=0. Note that $I_{[e(t-1)<0]}$ denotes the indicator function (i.e., it takes a value of one if $e(t-1) < 0$ and zero otherwise). Use the plot function to plot the resulting time series.

c. Find and report the sample kurtosis of the time series in parts *a* and *b* using the kurtosis function. Suppose that we increase the length of the time series in part *a*. Do you think that the sample kurtosis will converge to a specific value as the length of the series gets very long? If so, then what value? Briefly explain how you arrived at your answer.

Problem 3.

This problem uses the data in *sizeDecileReturns.xls* from Problem 1.

a. Find the number of returns for decile 5 that are less than zero, equal to zero, and greater than zero.

b. Find the number of times that the returns in decile 5 are greater, in absolute value, than two times the standard deviation of the returns in that decile.

c. Use find to indentify the indices of the negative returns in decile 5 and store these returns in rNeg. Compute the sample mean and sample standard deviation of these negative returns.

d. Use all to determine the number of days on which the returns for all decile portfolios were positive. Use any to determine the number of days on which there was (i) at least one positive return and (ii) no positive returns (Hint: use negation (~)).