HOMEWORK 2

Please submit your homework to xm@bu.edu. Don't forget to attach your figures and code. Feel free to ask me if you have any question. GLHF! -Sean.

Problem 1: autocorrelation

For this problem you will need the high-frequency trading data of companies in S&P 500. From the course website you can find sample data which were extracted on Feb 10, 2017, with frequency of five seconds (500×4679 data points). Now we can do something real!

1. Choose one company from the S&P 500 components. Use either buy price or sell price to find the log returns ($\Delta t = 5$ sec). Plot the autocorrelation function of the log returns. Can you find any useful short-time autocorrelation other than noise?

2. To describe the non-trivial autocorrelation of log returns, you are required to use an autoregressive (AR) model. Try to fit the following equation

$$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_s x_{t-s} + \epsilon_t$$

   with x_t the log return at time t. The autocorrelation function is determined by the parameters $\{a_0, a_1, \cdots\}$, while the noise is determined by the residual ϵ_t, a Gaussian random variable with variance σ^2. Choose $s = 5$ and estimate $\{a_0, a_1, \cdots\}$ and σ for your data and verify their validity. (To verify your estimation, simply find the mean μ of your data and check if it satisfies the equality $\mu = a_0 + (a_1 + a_2 + \cdots + a_s)\mu$)

3. Generate a random time series from your AR(5) model. Plot the autocorrelation function and compare it with that of your log returns.

4. In your AR(5) model we made an assumption that the variance σ^2 is constant at any time. You already know that this is *not* true. Indeed, we have to use an autoregressive conditional heteroskedasticity (ARCH) model to do our fitting. But first, we have to make sure that our data are unbiased. The first step is to subtract μ from your old data points to generate a new data set $y_t = x_t - \mu$ which has mean 0. Plot

the autocorrelation function of the square of y_t (i.e., volatility). Can you find any long-time autocorrelation feature?

5. Try to fit the following equation

$$y_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \cdots + \alpha_p y_{t-p}^2$$

where y_t^2 is actually the observed variance σ_t^2. The parameters $\{\alpha_0, \alpha_1, \cdots\}$ thus tell you the autocorrelation of squared residuals ϵ_t^2 in your AR(5) model. Choose $p = 5$ and estimate $\{\alpha_0, \alpha_1, \cdots\}$.

6. Generate a random time series from your ARCH(5) model. Plot the autocorrelation function and compare it with that of your volatility. Now you have gained enough knowledge of the high-frequency trading market by combining your AR(5) and ARCH(5) models!