HOMEWORK 3

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

From the course website (http://polymer.bu.edu/hes/PY538Materials.html), you can find a sample dataset, which contains the buy/sell prices of all components of the S&P500 index on Feb 10, 2017. The data were collected with a frequency of five seconds.

1. Choose one company from the five hundred components and pre-process your data. Use either buy price or sell price to find the log returns ($\Delta t = 5$ sec). Plot the autocorrelation function of log returns. Can you find any useful short-term correlation, or is it just noise?

2. To quantify the autocorrelation of log returns, you should use an autoregressive (AR) model. Try to fit the following equation

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

   where x_t is the log return at time t. The autocorrelation function is determined by the parameters $\{a_0, a_1, \cdots, a_s\}$, while the noise is determined by the residual ϵ_t, a Gaussian random variable with variance σ^2. Choose $s = 5$ and find the best fit parameters $\{a_0, a_1, \cdots, a_s\}$ and σ for your data. Check the validity of your result. (To verify your estimation of the parameters, simply find the mean μ of the log returns 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 the autocorrelation of the log returns.

4. In your AR(5) model we made an assumption that the strength of fluctuation σ^2 is constant at any time. You already know that this is *not* true. Indeed, an autoregressive conditional heteroskedasticity (ARCH) model is the next step to finish our fitting. First, we need to make sure that our data is unbiased. The first step is to generate

a new data set from the residuals of your AR(5) model by letting $y_t = x_t - (a_0 + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_s x_{t-s})$. One can easily see that y_t has mean zero. Plot the autocorrelation function of the square of y_t (or to say, volatility). Can you find any long-term autocorrelation feature?

5. Try to fit the following equation

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

The best fit parameters $\{\alpha_0, \alpha_1, \cdots\}$ should tell you the autocorrelation between the 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 the autocorrelation of the squared log returns.