**Problem Set 5B** (due 11:59 pm, 19 November 2023)

Remark: Students should submit executable Python programs through Canvas. Text answers in Python programs can be written in Jupyter Notebook.

1. **Autocorrelation of the Absolute Return of NYSE**

Download the historical data of ^NYA from yahoo!finance. The time range should be set to MAX.

Suggested steps of the Python program:

* 1. Read the data from the data file using Pandas.
  2. The data contains non-numerical data. Clean the data using the Pandas command “dropna()”.
  3. Extract the time series of the absolute returns from the column ‘Adj Close’.
  4. Subtract the time series by its average.
  5. Plot the autocorrelation of the absolute return using the command “plt.acorr()”. The time lag should extend to around three years.
  6. For sufficiently large time lags, the autocorrelation curve becomes noisy due to fluctuations, and negative values even exist. Suggest a value of the maximum lag such that the truncated time series can be used to compute the power spectral density in Question 2.

1. **Power Spectral Density of NYSE**
   1. Prepare the non-negative branch of the autocorrelation obtained from Question 1.
   2. Plot the power spectral density of the above autocorrelation function using logarithmic scales for both axes.
   3. Obtain the value of the exponent in the relation using regression (for example, LinearRegression in scikit-learn).
2. **Detrended Fluctuation Analysis of NYSE**
   1. Generate the series of cumulative sum.
   2. The Sliding Window Method

Next, we consider the range of time lags of the autocorrelation. The method of repeatedly dividing the time series into non-overlapping periods cannot generate sufficient data points for accurate determination of the exponent. So, we will use the sliding window method instead.

In the sliding window method for a time series of length , we will compute the averaged detrended fluctuations of all data segments of duration within the time series. The starting instants of these segments run from to . This provides samples of data segments for analysis using the sliding window method.

For reliable analysis, let us consider data segments containing at least 7 data points. Hence, the lengths of the data segments run from to . is the maximum time lag suggested in Question 1. Lengths greater than will not be considered as the autocorrelation deviates from the power law.

However, it is not necessary to increase linearly. As we are studying a power law, it is sufficient to increase exponentially. We will select the exponents that increase from to in steps of 0.1, and assign

* 1. For each segment starting at compute the root-mean-square difference between the data points and the local trend. No fitting is required because the following result can be used directly (see Appendix):

Remark: The correlation can be computed by applying np.dot to the arrays of x and y.

Remark: Since and are the same for all samples, they can be computed before entering the for loop.

Remark: For and

* 1. For each value of average over all samples of This yields the result Plot versus using logarithmic scales for both axes. Observe whether the power law is obeyed and whether the short time and long time behaviors are different. If they are, find the crossover time.
  2. Obtain the exponent(s) in the relation by regression.
  3. Verify whether the relation is valid.

**Appendix: The Root-Mean-Square Difference between the Data Points and the Local trend**

Suppose the local trend is given by Then the difference is given by

To minimize the difference, we take the derivatives with respect to and

This leads to the following two equations

The solution is

Substituting into the expression of

The first two terms vanish because they are expressions that have appeared in the derivatives. The last term is the remaining one. Hence,