Problem Set 5B (due 11:59 pm, 11 November 2021)

Remark: Students should submit both Excel files and executable Python programs through Canvas. Text answers in Excel files can be written in Text Box. Text answers in Python programs can be written as comments or in Jupyter Notebook.

1. **Study of the historical data of Hang Seng Index Using Excel**
   1. Visit the Website of “Yahoo! Finance”. Enter “^HSI" for Hang Seng Index. Click “Historical Data” (歷史數據). Download the index data from the earliest date to an Excel file. There should be more than 8,800 lines.

* 1. Clean the data by removing the unwanted lines labeled by “null”.
  2. Calculate the daily returns.
  3. Calculate the normalized daily returns and separate the data into positive and negative returns.
  4. Plot the cumulative return distribution **for the positive and negative tails separately. The return values should be binned logarithmically.**
  5. Observe whether the plots obey power law distributions. If so, calculate the exponents of the power-law regime.

1. **Detrended Fluctuation Analysis of Hang Seng Index Using Python**

Suggested steps of the Python program:

* 1. Read the data from the cleaned data file using Pandas.
  2. Extract the time series of the absolute returns from the column ‘Adj Close’ using

ret = abs(np.log(hsi\_df['Adj Close'][tick + 1]/hsi\_df['Adj Close'][tick])).

* 1. Compute the average of the time series.

Remark: The average of a series can be computed from

average = np.sum(series)/len(series)

* 1. Generate the series of cumulative sum, with each term subtracted by the average.
  2. The Sliding Window Method

Next, we consider the range of the lengths of the periods to be analyzed. 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. This means that for the time series of length we will analyze the periods for increasing values of

For reliable analysis, let us consider periods containing at least 7 data points. Hence, for we start with and increase to obtaining samples of data segments for analysis using the sliding window method.

At the same time, we do not want to have too few samples for averaging. If we need at least 7 samples for averaging, the largest value of would be

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.
  2. To obtain the exponent in the relation discard the upper and lower parts of the data points that deviate from the power law and obtain the value of

Remark: Students may find it more convenient to work on parts g and h using Excel.

**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,