MSDM5058 Information Science Computational Project I: Time Series Analysis with Financial Data

This computational project consists of six parts. You need to analyse the time series of two stocks' daily closing prices and write an individual report to show your results. You have to submit your report onto Canvas at the latest on **27 April, 2024**. For you to experience a formal academic setting, your report must conform to the format of Lecture Notes in Computer Science (LNCS), on which this manual is based; otherwise, 10% of score will be deducted from your report. You may download its template from **Important downloads for authors** on LNCS's website [1]. This project accounts for 35% of your final grade.

1 Data Preprocessing

Download the time series of two stocks' daily closing prices from any credible source such as Yahoo Finance. You may select any stock whose price is available on more than 4000 days (around 15 years). Cite your source properly.

Let time start at $t = 0$ and end at $t = N$. Denote the stocks' time series by $S_1(t)$ and $S_2(t)$, then compute their daily return $X_1(t)$ and $X_2(t)$ with

$$X_i(t) = \ln\left[\frac{S_i(t)}{S_i(t-1)}\right] \quad \text{for} \quad i = 1, 2 \ . \tag{1}$$

Hereafter when a task is stated in terms of S or X without any subscripts, you need to perform it on both stocks.

- Update X by subtracting its mean $\overline{X} = \frac{1}{N}\sum_{t=1}^{N} X(t)$ from it.
- Plot $S(t)$ and $X(t)$.

2 Stationarity and Autocorrelation

We now have two daily-return time series X_1 and X_2, each with a mean equal to zero.

- Perform the augmented Dickey-Fuller test on X to judge its stationarity.
- Plot the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of X. According to the plots, what may be the orders of the ARMA model fitted to X? Compare your guess to the best fit.
- Plot the ACF of $|X|$, i.e. the absolute value of X. This ACF can be regarded as a "**nonlinear autocorrelation**" of X, while the usual ACF of X is its "linear autocorrelation".

  Compare the trends of the two autocorrelations. What may their differences suggest?

3 Fractal Behaviour of Time Series

Let us now analyse the fractality of X and S, which hints at correlation and memory of data, with various methods.

3.1 Hurst Exponent

- Compute the Hurst exponent H of X. What kind of value do you get? Comment on your result.
- Visualize your result with a plot of the rescaled range $R(n)$ of X against window size n. The plot should be able to show $R(n) \sim n^H$.

3.2 Detrended Fluctuation Analysis (DFA)

- Plot the detrended fluctuation $F(n)$ of X as a function of window size n.
- Hence, compute the scaling exponent α of X in $F(n) \sim n^\alpha$. What kind of value do you get? Is it consistent with the Hurst exponent?

3.3 Multifractality

The two methods above are empirically defined. They implicitly assume a time series can be described by a single fractal dimension, which gives constant values to H and α. However, a time series may not be fractal only but in fact "multifractal" as it may show different levels of **self-similarity** upon different ways of measurement.

We can detect multifractality via a mathematically more fundamental approach. If a time series is at least fractal, the absolute moments of the change of any of its parameters $Y(t)$ should scale according to

$$M(q,\tau) \equiv \left\langle |\delta_\tau Y(t)|^q \right\rangle \equiv \left\langle |Y(t+\tau) - Y(t)|^q \right\rangle \sim \tau^{f(q)} \ , \tag{2}$$

where $f(q)$ is some function of q. As $X = \delta_{\tau=1} \ln S$ is used in the calculation of Hurst exponent above, we will substitute $Y = \ln S$ in Eq. (2) for a meaningful comparison.

- Compute $M(q,\tau)$ for several values of q using $Y = \ln S$, then plot $M(q,\tau)^{1/q}$ against τ.
- Hence, plot $f(q)/q$ against q. Are S and thus X multifractal according to the plot? [Hint: the value of $f(1)$ should be roughly around the Hurst exponent H of X.]
- In addition to this approach, we can study multifractality with **multifractal detrended fluctuation analysis** (MDFA), which generalizes DFA.

  Perform MDFA on X, then plot the obtained scaling exponents $\alpha(q)$ against their orders q. (The original DFA sets $q = 2$.) Just as DFA should agree with Hurst exponent, does MDFA suggest a result consistent with the one obtained from Eq. (2)?

4 Granger Causality

Next, we are going to study the causal relation between X_1 and X_2 by fitting a **VARMA model** to them.

- Determine the orders of the VARMA model.
- What are the regression coefficients in the model? How significant are they?
- Perform F-tests to judge the Granger causality between X_1 and X_2.

5 Fourier Transform and Power Spectrum

Let us now study the power spectrum of a time series.

- Compute the Fourier transform of X, then plot the magnitude of its Fourier coefficients against frequency.
- Plot the power spectral density (PSD) of X against frequency. Comment on its power spectrum.

  (Hint: The Nyquist-Shannon sampling theorem forbids the frequency of a daily time series to exceed 0.5 cycles per day.)

6 Empirical Mode Decomposition

Because of its excessive components, Fourier transform is not a concise method to decompose a nonlinear time series and to extract useful information therefrom. In contrast, the method of empirical mode decomposition (EMD) can decompose a nonlinear time series into a finite and often small number of components called "**intrinsic mode functions**" (IMFs).

Let $c_j(t)$ be the jth IMF. The frequency of c_j decreases as its order j increases, meaning that the first IMF has the lowest frequency and so on.

- Decompose X into its IMFs. Plot c_1, $c_{k/4}$, $c_{k/2}$, $c_{3k/4}$, and c_k, where k is the number of IMFs.
- Compute the Hurst exponent of each of the IMFs, then plot their Hurst exponents against their orders. Explain its trend.
- Now consider the first two IMFs. Plot their PSD against frequency. What do they look like?
- Finally, plot the PSDs of $X - c_1$ and $X - c_1 - c_2$ against frequency. Compare the spectra of these "reduced" time series to that of X. Comment on your observation.

References

1. Information for Authors of Springer Computer Science Proceedings, https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines.