Econophysics - written exam 13 January 2017

Docent: Dr. Diego Garlaschelli

Consider a financial market consisting of N stocks. The time series of log-returns calculated from the daily closing prices of these N stocks are recorded for $T = 800$ days and then standardized. This means that, if $x_i(t)$ denotes the standardized daily log-return of the i-th stock on day t, then the temporal average $\mu_{x_i} \equiv \overline{x}_i = T^{-1}\sum_{t=1}^{T} x_i(t)$ is zero for all i and the temporal variance $\sigma^2_{x_i} \equiv \overline{x^2}_i - (\overline{x}_i)^2$ is one for all i. After calculating the $N \times N$ cross-correlation matrix $\mathbf{C}$, whose entries are the correlation coefficients $C_{i,j}$ calculated among all pairs of the N standardized time series of $x_i(t)$, the eigenvalues $\{\lambda_i\}_{i=1}^N$ and the corresponding orthonormal eigenvectors $\{\vec{v}_i\}_{i=1}^N$ of $\mathbf{C}$ are calculated. Let $v_{i,k}$ denote the k-th entry of the eigenvector $\vec{v}_i$.

It is found that the three largest eigenvalues are

$$\lambda_1 = 40, \quad \lambda_2 = 4, \quad \lambda_3 = 2$$

and that the eigenvectors $\vec{v}_1$ and $\vec{v}_2$ corresponding to the two largest eigenvalues λ_1 and λ_2 have entries

$$v_{1,k} = a > 0 \quad \text{for } 1 \le k \le N, \qquad\qquad v_{2,k} = \begin{cases} b > 0 & \text{for } 1 \le k \le M \\ -b < 0 & \text{for } M < k \le N \end{cases}$$

where $0 < M < N$. The entries of all the other eigenvectors $\{\vec{v}_i\}_{i=3}^N$ are found to be distributed according to a standard normal distribution.

Recalling that the matrix $\mathbf{C}$ admits the spectral decomposition

$$\mathbf{C} = \sum_{i=1}^{N} \lambda_i v_{i,k} v_{i,k},$$

and that the orthonormality of the N eigenvectors $\{\vec{v}_i\}_{i=1}^N$ means that their scalar products equal $\vec{v}_i \cdot \vec{v}_i = \sum_{k=1}^N v_{i,k}^2 = 1$ for all i and $\vec{v}_i \cdot \vec{v}_j = \sum_{k=1}^N v_{i,k}v_{j,k} = 0$ for all i,j (with $i \neq j$), answer all the following questions.

- **Question 1.**

Knowing that $a = 0.05$, determine the number N of stocks traded in the market. *(Hint: use the normalization of $\vec{v}_1$.)*

- **Question 2.**

Determine the value of b. *(Hint: use the normalization of $\vec{v}_2$.)*

- **Question 3.**

Determine the value of M. *(Hint: use the orthogonality of $\vec{v}_1$ and $\vec{v}_2$.)*

- **Question 4.**

Determine the value of the maximum theoretical eigenvalue λ_+ predicted by Random Matrix Theory under the null hypothesis that the log-returns of the N stocks are all independent of each other and identically distributed.

- **Question 5.**
Using the previous result and the decomposition $\mathbf{C} = \mathbf{C}^{(random)} + \mathbf{C}^{(group)} + \mathbf{C}^{(market)}$ (which separates the contributions of random, group, and market modes to the overall cross-correlation respectively), compute the matrix $\mathbf{C}^{(market)}$.

- **Question 6.**
Compute the matrix $\mathbf{C}^{(group)}$.

- **Question 7.**
Compute the *diagonal* of the matrix $\mathbf{C}^{(random)}$. *(Hint: use the properties of the diagonal of* $\mathbf{C}$*.)*

- **Question 8.**
The empirical distribution (i.e. the normalized histogram) of the *off-diagonal* entries of the matrix $\mathbf{C}^{(random)}$ is found to be symmetric around an average value of zero. If the symbol $\underline{\mathbf{A}}$ denotes the average of *all* the N^2 entries of a $N \times N$ matrix $\mathbf{A}$, i.e. $\underline{\mathbf{A}} = N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{i,j}$, calculate the values $\underline{\mathbf{C}}^{(market)}$, $\underline{\mathbf{C}}^{(group)}$, and $\underline{\mathbf{C}}^{(random)}$.

- **Question 9.**
Now let us project the matrix $\mathbf{C}^{(group)}$ to a network $\mathbf{G}^{(group)}$ whose nodes represents the N stocks and where a link between stocks i and j is established if and only if $C_{i,j}^{(group)} \geq 0$. Calculate the number n of connected components, the degree k_i of each node i and the clustering coefficient c_i of each node i in $\mathbf{G}^{(group)}$.

- **Question 10.**
Similarly, let us project the matrix $\mathbf{C}^{(random)}$ to a network $\mathbf{G}^{(random)}$ whose nodes represents the N stocks and where a link between stocks i and j is established if and only if $C_{i,j}^{(random)} \geq 0$. Calculate the (expected) number L of links, the (expected) degree k_i of each node i, and the (expected) average nearest neighbour degree k_i^{nn} of each node i in $\mathbf{G}^{(group)}$.

- **Question 11.**
Now let us define the total log-return over all stocks at time t as $y(t) \equiv \sum_{i=1}^{N} x_i(t)$. Calculate the temporal average $\mu_y \equiv \overline{y} = T^{-1}\sum_{t=1}^{T} y(t)$ of $y(t)$.

- **Question 12.**
Calculate the temporal variance $\sigma_y^2 \equiv \overline{y^2} - (\overline{y})^2$ of $y(t)$. *(Hint: calculate* $\underline{\mathbf{C}}$ *from your answer to Question 8 and recall that, for standardized time series like* $x_i(t)$*, covariances coincide with correlation coefficients.)*

- **Question 13.**
It is found that the temporal autocorrelation of each of the N stocks vanishes on a time scale much faster than one day, thus providing evidence that, for each i, the log-returns $x_i(t)$ and $x_i(t')$ (with $t \neq t'$) are independent. Can the Central Limit Theorem be used to approximate the empirical distribution $P(y)$ of the values of $y(t)$ over time, i.e. the normalized histogram of the values $(y(1), y(2), \ldots, y(T))$? Why?

- **Question 14.**
Assume the distribution $P(y)$ defined above has power-law tails of the form $P(y) \propto |y|^{-3.5}$ for large $|y|$. Can it be a Levy-stable distribution? Why?

- **Question 15.**
Does the time series $y(t)$ exhibit aggregational normality? Why?