Econophysics - exam 28/06/2011 with solutions

(docent: Dr. Diego Garlaschelli)

Please write your name and student ID *clearly* on *every* sheet you hand in. Use only the sheets that have been distributed. You can use the textbook and printed copies of the lecture slides. Please write your solutions in English, and for each answer indicate the number of the corresponding question.

PROBLEM 1

The (synchronous) correlation coefficient $\rho_{1,2}$ between two time series $x_1(t)$ and $x_2(t)$ measures the accordance with the linear model $x_1(t) = a \cdot x_2(t) + b$. Perfect accordance is found in case of complete correlation (i.e. $\rho_{1,2} = +1$, which indicates that $a > 0$) or complete anticorrelation (i.e. $\rho_{1,2} = -1$, which indicates that $a < 0$). Values $|\rho_{1,2}| > 0.9$ already signal a very good accordance. Note that $\rho_{1,2}$ does not carry information about b.

1.1) If $x_1(t)$, $x_2(t)$ and $x_3(t)$ denote the time series of daily log-returns of three stocks traded in a financial market, determine which (if any) of the following matrices is a possible empirically measured correlation matrix (with entries $\{\rho_{i,j}\}$) among the three stocks, and explain why in each case:

$$A = \begin{pmatrix} 0 & 0.91 & -0.97 \\ 0.91 & 0 & 0.99 \\ -0.97 & 0.99 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0.91 & -0.97 \\ -0.91 & 1 & 0.99 \\ 0.97 & -0.99 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 0.91 & -0.97 \\ 0.91 & 1 & 0.99 \\ -0.97 & 0.99 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & -0.91 & 0.97 \\ -0.91 & 1 & -0.99 \\ 0.97 & -0.99 & 1 \end{pmatrix}$$

1.2) Whenever one of the above matrices is a possible correlation matrix, draw the corresponding Minimum Spanning Tree (MST) among the three stocks. Compute the degree sequence of the MST, the clustering coefficient of vertex 1, and the characteristic path length of the MST (average of shortest path lenghts over all pairs of vertices).

PROBLEM I - SOLUTION

1.1)

- A **cannot** be a correlation matrix because its diagonal elements are zero, while $\rho_{i,i} = +1 \ \forall i$.

- B **cannot** be a correlation matrix because its off-diagonal elements are antisymmetric, while $\rho_{i,j} = \rho_{j,i}$ if $i \neq j$.

- C **cannot** be a correlation matrix because $\rho_{1,2} \approx +1$ would imply that $x_1(t) \approx a \cdot x_2(t) + b$ with $a > 0$, and $\rho_{2,3} \approx +1$ would imply that $x_2(t) \approx c \cdot x_3(t) + d$ with $c > 0$. Taken together, these would imply that $x_1(t) \approx ac \cdot x_3(t) + ad + b$ with $ac > 0$, which should result in $\rho_{1,3} \approx +1$. Instead in the matrix $\rho_{1,3} = -0.97$.

- D **can** be a correlation matrix because its diagonal elements are $\rho_{i,i} = +1$, its off-diagonal elements are symmetric, and their signs are all consistent with each other. Indeed, $\rho_{1,2} \approx -1$ implies that $x_1(t) \approx a \cdot x_2(t) + b$ with $a < 0$, and $\rho_{2,3} \approx -1$ implies that $x_2(t) \approx c \cdot x_3(t) + d$ with $c < 0$. Taken together, these imply that $x_1(t) \approx ac \cdot x_3(t) + ad + b$ with $ac > 0$, which results in $\rho_{1,3} \approx +1$ as indeed is.

1.2)

- The Minimum Spanning Tree of matrix D (the only possible correlation matrix) is made of 3-1=2 links corresponding to the 2 strongest correlations (hence one between stocks 1 and 3, and one between stocks 1 and 2).

- Its degree sequence is $\vec{k} = (k_1, k_2, k_3) = (2, 1, 1)$.

- The clustering coefficient of vertex 1 is $C_1 = 0$ because there are no closed triangles to which vertex 1 belongs (this also follows in general by the definition of the MST, and is true for all vertices in a tree).

- The characteristic path length is the average of shortest path lenghts over all pairs of vertices:

$$\overline{d} = \frac{d_{12} + d_{13} + d_{23}}{3} = \frac{1 + 1 + 2}{3} = \frac{4}{3} \approx 1.33$$

  where d_{ij} denotes the shortest path lenght between vertices i and j.

PROBLEM 2

In an inter-bank network, $N = 503$ banks are linked by $L = 712$ links indicating contracts where money has been borrowed by one bank from the other. Ignoring the directionality of these links, the clustering coefficient C_i and the nearest-neighbour degree k_i^{nn} of all vertices are measured, and their average values (over all vertices) $\overline{C}$ and $\overline{k^{nn}}$ are computed.

2.1) Determine in which (if any) of the following cases the inter-bank network is consistent with an Erdős-Rényi random graph model, and explain why in each case:

- Case A: $\overline{C} = 0.53139$, $\overline{k^{nn}} = 4.27$
- Case B: $\overline{C} = 0.00721$, $\overline{k^{nn}} = 4.27$
- Case C: $\overline{C} = 0.00564$, $\overline{k^{nn}} = 2.84$
- Case D: $\overline{C} = 0.00043$, $\overline{k^{nn}} = 2.84$

2.2) In terms of N and L, write down the probability $P(p)$ to generate the real inter-bank network using the Erdős-Rényi model with parameter p.

2.3) Show that the value of p that maximizes the 'log-likelihood' $\mathcal{L}(p) \equiv \ln P(p)$ to obtain the real network (*maximum likelihood principle*) is

$$p^* = \frac{2L}{N(N-1)}$$

2.4) Write down the explicit form of the degree distribution of the inter-bank network predicted by the random graph model when $p = p^*$. Determine the numerical values of the mean and variance of the distribution using the parameters specified in the problem.

PROBLEM 2 - SOLUTION

2.1)

In the Erdős-Rényi random graph model (RGM) with parameter p, the expected number of links is $\langle L\rangle = pN(N-1)/2$, the expected value of $\overline{C}$ is $\langle\overline{C}\rangle = p$ and the expected value of $\overline{k^{nn}}$ is $\langle\overline{k^{nn}}\rangle \approx pN$. Since N and L are common to all cases, one can use p to match L, which gives $p = 2L/N(N-1) \approx 0.00564$. One must then check whether the values of $\overline{C}$ and $\overline{k^{nn}}$ are consistent with this value of p, i.e. whether $\overline{C} \approx p \approx 0.00564$ and $\overline{k^{nn}} \approx pN \approx 2.84$. Thus one finds that only case C is consistent with the RGM.

2.2)

The probability to generate the real inter-bank network with N banks and L links under the RGM is

$$P(p) = p^L(1-p)^{N(N-1)/2-L} = \left(\frac{p}{1-p}\right)^L (1-p)^{N(N-1)/2}$$

because there are L successful link creations (each with probability p) and $N(N-1)/2-L$ unsuccessful link creations (each with probability $1-p$).

2.3)

The log-likelihood is

$$\mathcal{L}(p) \equiv \ln P(p) = L\ln\left(\frac{p}{1-p}\right) + \frac{N(N-1)}{2}\ln(1-p)$$

and its maximum is attained by the value p^* that makes the derivative vanish:

$$0 = \frac{d\mathcal{L}(p)}{dp}\bigg|_{p^*} = L\left(\frac{1}{p^*}+\frac{1}{1-p^*}\right) - \frac{N(N-1)}{2}\frac{1}{1-p^*} = \frac{L}{p^*(1-p^*)} - \frac{N(N-1)}{2}\frac{1}{1-p^*}$$

Multiplying by $1-p^*$ and rearranging, one finds

$$L = p^*\frac{N(N-1)}{2}$$

which leads to the desired result.

2.4)

For $p = p^*$, the RGM predicts a Binomial degree distribution

$$P(k) = \binom{N-1}{k}(p^*)^k(1-p^*)^{N-1-k} = \binom{N-1}{k}\left[\frac{2L}{N(N-1)}\right]^k\left[1-\frac{2L}{N(N-1)}\right]^{N-1-k}$$

whose mean μ and variance σ^2 read

$$\mu \equiv \langle k\rangle = p^*(N-1) = \frac{2L}{N} \approx 2.83$$

$$\sigma^2 \equiv \langle k^2\rangle - \langle k\rangle^2 = p^*(1-p^*)(N-1) \approx 2.81$$

PROBLEM 3

In the Cont-Bouchaud model of financial markets, the log-return $r(t)$ of an asset traded by N agents forming a coalition network is

$$r(t) = \sum_{A=1}^{N_C} s_A \phi_A(t)$$

where A labels the connected components of the underlying coalition network, $N_C \leq N$ is the number of connected components, s_A is the size (number of vertices) in the connected component A, and $\phi_A(t)$ is a random variable representing the choice of the coalition A at time t, i.e.

$$\phi_A(t) = \begin{cases} +1 & \text{if A 'buys'} \quad \text{(with probability } a\text{)} \\ 0 & \text{if A 'waits'} \quad \text{(with probability } 1-2a\text{)} \\ -1 & \text{if A 'sells'} \quad \text{(with probability } a\text{)} \end{cases}$$

The values of $\phi_A(t)$ at different times t and for different components A are statistically independent.

3.1) In the simple case $a = 1/2$, determine the expected values $\langle \phi_A(t) \rangle$ and $\langle \phi_A(t)\phi_B(t) \rangle$, distinguishing the cases $A = B$ and $A \neq B$.

3.2) After further assuming that the underlying network is static (i.e. coalitions do not change in time) and that we do not know anything about its topology (in particular, it is not necessarily a random graph as in the original model), use the previous results to compute the expected log-return $\mu \equiv \langle r \rangle$ and the variance $\sigma^2 \equiv \langle r^2 \rangle - \langle r \rangle^2$ in terms of the component sizes $\{s_A\}$. *Hint: write* $x_A \equiv s_A \phi_A$ *and use the fact that* $(\sum_A x_A)^2 = \sum_{A,B} x_A x_B = \sum_A x_A^2 + \sum_{A \neq B} x_A x_B$.

3.3) Determine under which conditions (in terms of the coalition network, and remember it is not necessarily a random graph) the Central Limit Theorem can be applied to the sum $\sum_{A=1}^{N_C} s_A \phi_A(t)$, and in such a case write down the approximate form of the distribution $P(r)$ of log-returns, with the appropriate parameters.

PROBLEM 3 - SOLUTION

3.1)

- The expected value of $\phi_A(t)$ is $\langle \phi_A(t) \rangle = 0$ beacuse the values $\phi_A(t) = +1$ and $\phi_A(t) = -1$ occur with the same probability.

- In the case $A = B$, $\langle \phi_A(t)\phi_B(t) \rangle = \langle \phi_A(t)^2 \rangle = +1$ beacuse if $a = 1/2$ then $\phi_A(t)^2 = +1$ always. Note that $\phi_A(t)^2$ becomes a deterministic variable!

- In the case $A \neq B$, $\phi_A(t)$ and $\phi_B(t)$ are statistically independent, therefore $\langle \phi_A(t)\phi_B(t) \rangle = \langle \phi_A(t) \rangle \langle \phi_B(t) \rangle = 0$.

3.2)

- The expected log-return is

$$\mu \equiv \langle r \rangle = \sum_{A=1}^{N_C} s_A \phi_A(t) \rangle = 0$$

- Writing $x_A \equiv s_A \phi_A$ and using the fact that $(\sum_A x_A)^2 = \sum_{A,B} x_A x_B = \sum_A x_A^2 + \sum_{A\neq B} x_A x_B$, the variance is

$$\sigma^2 \equiv \langle r^2 \rangle - \langle r \rangle^2 = \langle r^2 \rangle = \langle \sum_{A=1}^{N_C} x_A^2 \rangle + \langle \sum_{A\neq B} x_A x_B \rangle = \sum_{A=1}^{N_C} s_A^2 \langle \phi_A^2(t) \rangle + \sum_{A\neq B} s_A s_B \langle \phi_A(t)\phi_B(t) \rangle = \sum_{A=1}^{N_C} s_A^2$$

(i.e. it is the sum of squared component sizes).

3.3)

- The Central Limit Theorem (CLT) can be applied to the sum $\sum_{A=1}^{N_C} s_A \phi_A(t)$ if the summands are independent and identically distributed (i.i.d.), if their individual variance is finite, and if the number of summands is very large. The summand $s_A \phi_A(t)$ has a variance

$$\tilde{\sigma}^2 \equiv \langle s_A^2 \phi_A^2(t) \rangle - \langle s_A \phi_A(t) \rangle^2 = s_A^2 \langle \phi_A^2(t) \rangle = s_A^2$$

(consistenlty with the expression for σ^2 above. The only possibility that ensures that all the summands are i.i.d. is that all components have the same size, i.e. $s_A = s\ \forall A$ (otherwise summands are still independent, but not identically distributed). Furthermore the number of summands (hence the number N_C of connected components) must become very large, but at the same time their individual variance $\tilde{\sigma}^2 = s^2$ (hence the size s of each component) must remain finite. This means that the size of the network (i.e. the number N of vertices) must become large, and it must be split into connected components of constant (and independent of N) size s. So the number of connected components must be $N_C = N/s$, which ideed grows as N grows and s remains finite.

- When the above conditions apply, the CLT can be applied and its predicts that the approximate form of the distribution $P(r)$ of log-returns is Gaussian with mean

$$\mu = \langle r \rangle = 0$$

(see above) and variance

$$\sigma^2 = \langle r^2 \rangle - \langle r \rangle^2 = \sum_{A=1}^{N_C} s_A^2 = N_C s^2 = Ns$$

Thus

$$P(r) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(r-\mu)^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi Ns}} \exp\left[-\frac{r^2}{2Ns}\right]$$