Econophysics - retake exam 16 March 2015 with solutions

Docent: Dr. Diego Garlaschelli

You should return these pages completed. Please write your answers to all questions by filling the boxes. Motivate each of your answers. Attach all the additional sheets required to prove your results. Answers are ignored if there is no explanation or proof. Please write clearly and in English. Please don't forget to write your name and student ID below. Good luck!

NAME: . ...

SURNAME: ....

STUDENT ID: ....

PROBLEM 1

Consider a very simple market with only 3 stocks, and denote the corresponding time series of log-returns as $x_1(t)$, $x_2(t)$, $x_3(t)$ $(t = 1, \dots T)$. Moreover, as a simple Index of the market, consider the quantity

$$I(t) \equiv x_1(t) + x_2(t) + x_3(t) \qquad \forall t$$

The time variances of the 3 stocks and of the Index are measured, over the entire period $[1, T]$, as

$$Var[x_1] = 1.2 \qquad Var[x_2] = 2.3 \qquad Var[x_3] = 0.7 \qquad Var[I] = 4.8$$

Question 1.1

On the basis of the above measurements, what can be concluded about the correlation among the 3 stocks?

☐ The stocks are all mutually uncorrelated;

☐ At least 1 pair of stocks is correlated;

☐ At least 2 pairs of stocks are correlated;

☐ All the 3 pairs of stocks are correlated;

☐ Nothing can be concluded.

Why?..

Question 1.2

Given the logical relationship between statistical correlation and probabilistic dependence, what can be concluded about the (in)dependence among the 3 stocks?

☐ All the 3 pairs of stocks are dependent;

☐ At most 1 pair of stocks is independent;

☐ At most 2 pairs of stocks are independent;

☐ All the 3 pairs of stocks are independent;

☐ Nothing can be concluded.

Why?..

Question 1.3

The second moments of the log-returns of stocks 1, 3 and of the Index are measured as

$$\overline{x_1^2} = 1.36 \qquad \overline{x_3^2} = 0.95 \qquad \overline{I^2} = 4.84$$

Find all the average log-returns $\overline{x_1}$, $\overline{x_2}$, $\overline{x_3}$, and $\overline{I}$ knowing that $\overline{x_1} > 0$, $\overline{x_3} > 0$, and $\overline{I} < 0$.

$\overline{x_1} = \dots\dots\dots$ $\qquad$ $\overline{x_2} = \dots\dots\dots$ $\qquad$ $\overline{x_3} = \dots\dots\dots$ $\qquad$ $\overline{I} = \dots\dots\dots$

Question 1.4

Knowing that

$$\overline{x_1 x_2} = 0.06 \qquad \overline{x_1 x_3} = 0.2$$

find the 3×3 covariance matrix among the three stocks.

$$C \equiv \begin{pmatrix} \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \end{pmatrix}$$

Question 1.5

Find the 3×3 correlation matrix among the three stocks.

$$\rho \equiv \begin{pmatrix} \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \end{pmatrix}$$

Question 1.6

Draw the (correlation-based) Minimum Spanning Tree connecting the three stocks:

PROBLEM 1 - SOLUTION

Recall that the *time variance* of a time series $x_i(t)$ is defined as

$$Var[x_i] \equiv \overline{x_i^2} - \overline{x_i}^2 \tag{1}$$

where $\overline{x_i} \equiv \sum_{t=1}^{T} x_i(t)/T$ is the *first moment* of $x_i(t)$ and $\overline{x_i^2} \equiv \sum_{t=1}^{T} x_i^2(t)/T$ is the *second moment*. Also recall that the *time covariance* between two time series $x_i(t)$ and $x_j(t)$ is defined as

$$Cov[x_i, x_j] \equiv \overline{x_i x_j} - \overline{x_i} \cdot \overline{x_j} \tag{2}$$

where $\overline{x_i x_j} \equiv \sum_{t=1}^{T} x_i(t) x_j(t)/T$. Finally, recall that the variance of a sum of N random variables $x_1, \dots x_N$ is the sum of the covariances between all possible pairs of the N variables:

$$Var\left[\sum_{i=1}^{N} x_i\right] = \sum_{i=1}^{N}\sum_{j=1}^{N} Cov[x_i, x_j] = \sum_{i=1}^{N} Var[x_i] + 2\sum_{i=1}^{N}\sum_{j<i} Cov[x_i, x_j]$$

1.1)

In this particular case,

$$\begin{aligned} Var\,[I] &= Var\,[x_1 + x_2 + x_3] \\ &= Var[x_1] + Var[x_2] + Var[x_3] + 2(Cov[x_1, x_2] + Cov[x_1, x_3] + Cov[x_2, x_3]) \end{aligned} \tag{3}$$

and the measured values imply that

$$\begin{aligned} Cov[x_1, x_2] + Cov[x_1, x_3] + Cov[x_2, x_3] &= \frac{Var\,[I] - Var[x_1] - Var[x_2] - Var[x_3]}{2} \\ &= \frac{4.8 - 1.2 - 2.3 - 0.7}{2} = 0.3 \end{aligned} \tag{4}$$

This implies that at least one of the covariances must be different from zero, i.e. *at least one pair of stocks is correlated.*

1.2)

Recall that if two random variables x and y are independent (i.e. their joint probability $P(x, y)$ is the product $P_x(x) \cdot P_y(y)$ of their two marginal probabilities) they are also uncorrelated (i.e. their covariance and correlation are zero), while the inverse is in general not true (dependent variables can be uncorrelated). This implies that if two random variables are correlated (nonzero covariance and nonzero correlation) they are necessarily not independent. Since in this case at least 1 pair of stocks are correlated, we can conclude that at least 1 pair of stocks are dependent, i.e. *at most 2 pairs of stocks are independent.*

1.3)

From eq.(1) it follows that

$$\overline{x_i} = \pm\sqrt{\overline{x_i^2} - Var[x_i]}$$

Selecting the signs as given in the text, we have

$$\overline{x_1} = +\sqrt{\overline{x_1^2} - Var[x_1]} = +\sqrt{1.36 - 1.2} = 0.4$$

$$\overline{x_3} = +\sqrt{\overline{x_3^2} - Var[x_3]} = +\sqrt{0.95 - 0.7} = 0.5$$

$$\overline{I} = -\sqrt{\overline{I^2} - Var[I]} = -\sqrt{4.84 - 4.8} = -0.2$$

Then, since $\overline{I} = \overline{x_1} + \overline{x_2} + \overline{x_3}$, we have

$$\overline{x_2} = \overline{I} - \overline{x_1} + \overline{x_3} = -0.2 - 0.4 - 0.5 = -1.1$$

1.4)

From eq.(2) we have

$$Cov[x_1, x_2] = \overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2} = 0.06 - 0.4 \cdot (-1.1) = +0.5$$

$$Cov[x_1, x_3] = \overline{x_1 x_3} - \overline{x_1} \cdot \overline{x_3} = 0.2 - 0.4 \cdot 0.5 = 0$$

The missing covariance $Cov[x_2, x_3]$ can be found from eq.(4):

$$Cov[x_2, x_3] = 0.3 - Cov[x_1, x_2] - Cov[x_1, x_3] = -0.2$$

Therefore the 3×3 covariance matrix is

$$C \equiv \begin{pmatrix} Cov[x_1, x_1] & Cov[x_1, x_2] & Cov[x_1, x_3] \\ Cov[x_2, x_1] & Cov[x_2, x_2] & Cov[x_2, x_3] \\ Cov[x_3, x_1] & Cov[x_3, x_2] & Cov[x_3, x_3] \end{pmatrix} = \begin{pmatrix} Var[x_1] & Cov[x_1, x_2] & Cov[x_1, x_3] \\ Cov[x_1, x_2] & Var[x_2] & Cov[x_2, x_3] \\ Cov[x_1, x_3] & Cov[x_2, x_3] & Var[x_3] \end{pmatrix}$$
$$= \begin{pmatrix} 1.2 & 0.5 & 0 \\ 0.5 & 2.3 & -0.2 \\ 0 & -0.2 & 0.7 \end{pmatrix}$$

where we have used the symmetry property $Cov[x_i, x_j] = Cov[x_j, x_i]$ and the fact that $Cov[x_i, x_i] = Var[x_i]$.

1.5)

Recall that the *correlation coefficient* between two random variables x_i and x_j is defined as

$$\rho_{i,j} \equiv \frac{Cov[x_i, x_j]}{\sqrt{Var[x_i]Var[x_j]}}$$

Therefore the 3×3 correlation matrix can be obtained from the covariance matrix and reads

$$\rho \equiv \begin{pmatrix} \rho_{1,1} & \rho_{1,2} & \rho_{1,3} \\ \rho_{2,1} & \rho_{2,2} & \rho_{2,3} \\ \rho_{3,1} & \rho_{3,2} & \rho_{3,3} \end{pmatrix} = \begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} \\ \rho_{1,2} & 1 & \rho_{2,3} \\ \rho_{1,3} & \rho_{2,3} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & -0.16 \\ 0 & -0.16 & 1 \end{pmatrix}$$

where we have used the facts that $\rho_{i,j} = \rho_{j,i}$ and $\rho_{i,i} = 1$.

1.6)

The correlation-based Minimum Spanning Tree (MST) connecting the three stocks must contain 2 links (not 1 otherwise it wouldn't be spanning, and not 3 otherwise it wouldn't be a tree). Those 2 links correspond to the 2 largest correlation coefficients. So in this case the MST contains the links (1,2) and (1,3). Note that, curiously, one of the two pairs of stocks is actually uncorrelated ($\rho_{1,3} = 0$). Still, the MST prescription requires to include it.

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.

Question 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:

☐ $\overline{C} = 0.53139, \overline{k^{nn}} = 4.27$ Why?..

☐ $\overline{C} = 0.00721, \overline{k^{nn}} = 4.27$ Why?..

☐ $\overline{C} = 0.00564, \overline{k^{nn}} = 2.84$ Why?..

☐ $\overline{C} = 0.00043, \overline{k^{nn}} = 2.84$ Why?..

Question 2.2

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

$P(p) =$ ....

Question 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

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

....

....

....

....

Question 2.4

Write down, with only L and N as parameters, 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.

$P(k) =$ ....

$\mu \equiv \langle k \rangle =$ .... $\sigma^2 \equiv \langle k^2 \rangle - \langle k \rangle^2 =$ ....

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.

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

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

Question 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 $A \neq B$.

$\langle \phi_A(t) \rangle =$ ....

$\langle \phi_A(t)\phi_A(t) \rangle =$ ....

$\langle \phi_A(t)\phi_B(t) \rangle =$ ....

Question 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* $(\sum_A x_A)^2 = \sum_{A,B} x_A x_B = \sum_A x_A^2 + \sum_{A \neq B} x_A x_B$.

$\mu =$ ....

$\sigma^2 =$ ....

Question 3.3

Under which conditions (in terms of the structure of the coalition network, recalling it is not necessarily a random graph) the Central Limit Theorem can be applied to the sum $\sum_{A=1}^{N_C} s_A \phi_A(t)$?

....

....

Assuming these conditions hold, write down the approximate form of the distribution $P(r)$ of log-returns, with N as an explicit parameter.

$P(r) =$ ....

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

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