Econophysics - written exam 18 December 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 additional sheets only if required to prove your results. Answers are ignored if there is no explanation. 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

Assume that the log-returns of a stock are recorded at three different frequencies: $dt = 1$ min, $\delta t = 1$ day, and $\Delta t = 1$ month. The corresponding time series are denoted as $x(t) \equiv \ln[p(t)/p(t-dt)]$, $y(t) \equiv \ln[p(t)/p(t-\delta t)]$, and $z(t) \equiv \ln[p(t)/p(t-\Delta t)]$, where $p(t)$ is the price of the stock at time t. The log-returns $\{x(t)\}$ are distributed according to the probability distribution $P_{dt}(x)$, with mean μ_{dt} and standard deviation σ_{dt}. The log-returns $\{y(t)\}$ are found to be i.i.d. and distributed according to a stable distribution $P_{\delta t}(y)$ with mean $\mu_{\delta t}$ and standard deviation $\sigma_{\delta t}$. The distribution of $\{z(t)\}$, denoted as $P_{\Delta t}(z)$, is found to be Gaussian with mean $\mu_{\Delta t}$ and standard deviation $\sigma_{\Delta t}$.

1. Recalling that there are ≈ 20 market days in a month, what can be concluded about $y(t)$?

$P_{\delta t}(y)$: note that $z(t) = \sum_{\tau \in t} y(\tau)$ where τ runs over the 20 days of month t. Since $P_{\delta t}(y)$ is stable and $\{y(t)\}$ are i.i.d., $P_{\delta t}(y)$ must have the same distribution of $P_{\Delta t}(z)$ with adjusted parameters; therefore $P_{\delta t}(y)$ is Gaussian with parameters $\mu_{\delta t}$ and $\sigma_{\delta t}$, i.e. $P_{\delta t}(y) = \frac{\exp\left(-(y-\mu_{\delta t})^2/2\sigma_{\delta t}^2\right)}{\sqrt{2\pi\sigma_{\delta t}^2}}$

$\mu_{\delta t} = \mu_{\Delta t}/20$ (since $\mu_{\Delta t} = 20\mu_{\delta t}$) $\qquad$ $\sigma_{\delta t} = \sigma_{\Delta t}/\sqrt{20}$ (since $\sigma_{\Delta t}^2 = 20\sigma_{\delta t}^2$ given that $\{y(t)\}$ are i.i.d.)

Autocorrelation of $y(t)$:
this is necessarily zero, given that $\{y(t)\}$ are i.i.d.

Does $y(t)$ exhibit aggregational normality? Why?
Yes, because we know that its sums $\{z(t)\} = \{\sum_{\tau \in t} y(\tau)\}$ are normally distributed.

Can the Central Limit Theorem be applied to $\sum_{t=1}^{T} y(t)$ for T large? Why?
Yes, because $\{y(t)\}$ are i.i.d. with finite mean and variance ($\mu_{\delta t} < \infty$ and $\sigma_{\delta t}^2 < \infty$ follow from $\mu_{\Delta t} < \infty$ and $\sigma_{\Delta t}^2 < \infty$, which in turn follow from the gaussianity of $\{z(t)\}$)

2. Considering that one trading day is ≈ 390 min, what can be concluded about $x(t)$?

$P_{dt}(x)$: we don't know: if we knew that $\{x(t)\}$ are i.i.d., we would know that $P_{dt}(x)$ may be any distribution attracted to the (Gaussian) $P_{\delta t}(y)$; but $\{x(t)\}$ may be <u>not</u> i.i.d.

$\mu_{dt} = \mu_{\delta t}/390$ (similar to above) $\qquad$ σ_{dt} : we don't know, as we ignore the (in)dependence of $\{x(t)\}$

Autocorrelation of $x(t)$:
we know that it must vanish after at most $\delta t = 390dt = 1$day, because we know that $y(t) = \sum_{\tau \in t} x(t)$ (where τ runs over all minutes in day t) are i.i.d. and therefore not correlated.

Does $x(t)$ exhibit aggregational normality? Why?
Yes, because we know that its sums $\{y(t)\} = \{\sum_{\tau \in t} x(\tau)\}$ are normally distributed (see above).

Can the Central Limit Theorem be applied to $\sum_{t=1}^{T} x(t)$ for T large? Why?
We don't know: $\{x(t)\}$ may be dependent and with infinite variance.

PROBLEM 2

The time series of daily closing prices of a set of $N = 136$ stocks are recorded for $T = 204$ days. The set of stocks is ordered according to their sector as follows: Technology ($0 < i \leq 17$), Financials ($17 < i \leq 41$), Energy ($41 < i \leq 69$), Services ($69 < i \leq 103$) and Automobile ($103 < i \leq 136$). From the time series of log-returns, the cross-correlation matrix $\mathbf{C}$ is calculated. The four largest eigenvalues $\{\lambda_1, \ldots, \lambda_4\}$ of $\mathbf{C}$ are:

$$\lambda_1 = 20.7, \quad \lambda_2 = 8.2, \quad \lambda_3 = 4.1, \quad \lambda_4 = 2.9.$$

The entries of the corresponding eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_4\}$ are found to be:

$$v_{1,j} = \begin{cases} a & 0 < j \leq 17 \\ b & 17 < j \leq N \end{cases}; \qquad v_{2,j} = \begin{cases} c & 0 < j \leq 41 \\ d & 41 < j \leq N \end{cases}; \qquad v_{3,j} = \begin{cases} e & 0 < j \leq 69 \\ f & 69 < j \leq N \end{cases}; \qquad v_{4,j} = \begin{cases} g & 0 < j \leq 103 \\ h & 103 < j \leq N \end{cases},$$

where a, b, c, e, h are all positive, while d, f, g are all negative. The group correlation matrix $\mathbf{C}^{group}$ is calculated, and a threshold at the value zero is applied to project this matrix into a network of stocks, i.e. a link is drawn between stock i and stock j if and only if $C_{ij}^{group} > 0$.

Calculate the total number of links L in the projected network, the connectance (link density) c of the projected network, the number n_{cc} of connected components, the list of sizes $\{S_{cc}\}$ of the connected components, and the composition (in terms of stock sector) of the connected components, in the following four cases:

1. $\dfrac{cd}{e^2} > -\dfrac{1}{2}$ and $\dfrac{ef}{d^2} > -2$:

$L = M + N_1 N_2 + N_2 N_3$ (see below) $\qquad c = \frac{2L}{N(N-1)} \qquad n_{cc} = 1$

$\{S_{cc}\} = \{136\}$

Composition of the connected components: all sectors are in a single component (T+F+E+S+A).

Explanation: the largest eigenvalue consistent with random matrix theory is $\lambda_+ = (1 + \sqrt{N/T})^2 \approx 3.29$; therefore $\lambda_4 < \lambda_+$ is in the random bulk, while λ_1 is the market mode (since $v_{1,j} > 0 \; \forall j$). Thus $C_{ij}^{group} = \lambda_2 v_{2,i} v_{2,j} + \lambda_3 v_{3,i} v_{3,j}$, which implies the 3×3 block structure

$$\mathbf{C}^{group} = \begin{pmatrix} (\lambda_2 c_2 + \lambda_3 e^2)\mathbf{U}(N_1, N_1) & (\lambda_2 cd + \lambda_3 e^2)\mathbf{U}(N_1, N_2) & (\lambda_2 cd + \lambda_3 ef)\mathbf{U}(N_1, N_3) \\ (\lambda_2 cd + \lambda_3 e^2)\mathbf{U}(N_2, N_1) & (\lambda_2 d^2 + \lambda_3 e^2)\mathbf{U}(N_2, N_2) & (\lambda_2 d^2 + \lambda_3 ef)\mathbf{U}(N_2, N_3) \\ (\lambda_2 cd + \lambda_3 ef)\mathbf{U}(N_3, N_1) & (\lambda_2 d^2 + \lambda_3 ef)\mathbf{U}(N_3, N_2) & (\lambda_2 d^2 + \lambda_3 f^2)\mathbf{U}(N_3, N_3) \end{pmatrix}$$

where $\mathbf{U}(n, m)$ denotes a $n \times m$ matrix with all unit entries, $N_1 = 41$ (Technology+Financials), $N_2 = 69 - 41 = 28$ (Energy), and $N_3 = 136 - 69 = 67$ (Services+Auto). The rest follows from selecting only the entries $C_{ij}^{group} > 0$ and noting that $\lambda_2 = 2\lambda_3$. The 3 diagonal blocks are always positive, so a number $M \equiv \sum_{k=1}^{3} N_k(N_k - 1)/2$ of links is always present in the projected network. Depending on the value of the parameters, there can be additional positive entries in the off-diagonal blocks connecting the k-th block with the k'-th one, thus generating $N_k N_{k'}$ additional links (note: the entries between blocks 1 and 3 are however always negative).
With this first parameter choice, the entries $\lambda_2 cd + \lambda_3 e^2$ connecting blocks 1 and 2 and the entries $\lambda_2 d^2 + \lambda_3 ef$ connecting blocks 2 and 3 are positive. This creates a single component with $L = M + N_1 N_2 + N_2 N_3$.

2. $\dfrac{cd}{e^2} > -\dfrac{1}{2}$ and $\dfrac{ef}{d^2} < -2$:

$L = M + N_1 N_2$ $c = \frac{2L}{N(N-1)}$ $n_{cc} = 2$

$\{S_{cc}\} = \{69, 67\}$

Composition of the connected components: (T+F+E) and (S+A).

Explanation: With this second parameter choice, the entries $\lambda_2 cd + \lambda_3 e^2$ connecting blocks 1 and 2 are positive. This creates two components (T+F+E) and (S+A), and implies $L = M + N_1 N_2$.

3. $\dfrac{cd}{e^2} < -\dfrac{1}{2}$ and $\dfrac{ef}{d^2} > -2$:

$L = M + N_2 N_3$ $c = \frac{2L}{N(N-1)}$ $n_{cc} = 2$

$\{S_{cc}\} = \{41, 95\}$

Composition of the connected components: (T+F) and (E+S+A).

Explanation: With this third parameter choice, the entries $\lambda_2 d^2 + \lambda_3 ef$ connecting blocks 2 and 3 are positive. This creates two components (T+F) and (E+S+A), and implies $L = M + N_2 N_3$.

4. $\dfrac{cd}{e^2} < -\dfrac{1}{2}$ and $\dfrac{ef}{d^2} < -2$:

$L = M$ $c = \frac{2L}{N(N-1)}$ $n_{cc} = 3$

$\{S_{cc}\} = \{41, 28, 67\}$

Composition of the connected components: (T+F), (E), and (S+A).

Explanation: With this fourth parameter choice, the off-diagonal block entries are all negative. This creates three components (T+F), (E) and (S+A), and implies $L = M$.

PROBLEM 3

Let us consider five countries of the world, and let us represent their mutual trade relationships as links of a binary undirected graph **G**. Assume that the fitness model with probability $p_{ij} = \frac{zx_ix_j}{1+zx_ix_j}$ (where x is the Gross Domestic Product and z is a positive constant) is used to model this graph, and that the expected degrees under the model are found to be exactly equal to the observed degrees in the real network. Moreover, assume that $p_{34} > p_{53}$, $p_{15} > p_{25}$, $p_{43} > p_{14}$, $p_{52} = p_{45}$.

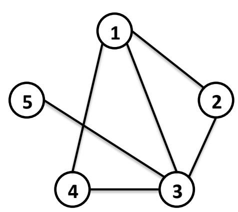
1. Find the degree k_i (number of trade partners) of each country i.

$k_1 = 3$ $k_2 = 2$ $k_3 = 4$ $k_4 = 2$ $k_5 = 1$

Explain your result: from the form of the connection probability p_{ij}, we note that $p_{il} > p_{jl}$ implies $x_i > x_j$ and $p_{il} = p_{jl}$ implies $x_i = x_j$ (for $l \neq i, j$). From the (in)equalities given in the text, we can therefore conclude that $x_3 > x_1 > x_2 = x_4 > x_5$. Moreover, we know that the expected degree $\langle k_i \rangle = \sum_{j \neq i} p_{ij}$ is an increasing function of x_i, and that it concides with the observed degree k_i. Therefore we must have $k_3 > k_1 > k_2 = k_4 > k_5$. Since the degrees must be non-negative integers, it is easy to check that the only set of numbers that satisfies the above (in)equalities in such a way that a simple undirected graph can be realized is $4 > 3 > 2 = 2 > 1$.

2. Draw the network **G** of international trade among the five countries.

G =



It is easy to check that this is the only graph consistent with the degrees found above.

3. Calculate the average nearest-neighbour degree k_i^{nn} (average number of trade partners of the neighbours of i) of each country i.

$k_1^{nn} = 8/3$ $k_2^{nn} = 7/2$ $k_3^{nn} = 2$ $k_4^{nn} = 7/2$ $k_5^{nn} = 4$

4. Calculate the clustering coefficient C_i of each country i (for countries with one or zero trade partners, conventionally set the clustering coefficient to 0).

$C_1 = 2/3$ $C_2 = 1$ $C_3 = 1/3$ $C_4 = 1$ $C_5 = 0$