Econophysics - solutions of written exam 20 January 2014

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 simple model of a market with only one stock and N agents trading it. Time is discrete and, at each time step t, the agent i has a certain demand $x_i(t)$ for the stock ($i = 1, \dots N$ and $t = 1, \dots T$ where T is the total number of time steps). The demand $x_i(t)$ is assumed to be a real number ($-\infty < x_i(t) < +\infty$) where $x_i(t) > 0$ means that at time t the agent i wants to *buy* a quantity $x_i(t)$ of the stock, while $x_i(t) < 0$ means that at time t the agent i wants to *sell* a quantity $x_i(t)$ of the stock. The case $x_i(t) = 0$ means that at time t the agent i does not want to take action.

We assume that, for different values of i and t, the quantities $\{x_i(t)\}$ are *independent and identically distributed* (i.i.d.) random variables $\{X_i(t)\}$, each with probability density function (PDF) given by

$$p(x) = p(X = x) = \begin{cases} C \cdot x^{-\alpha} & x > +1 \\ p_0(x) & -1 \leq x \leq +1 \\ C \cdot (-x)^{-\alpha} & x < -1 \end{cases} \tag{1}$$

with $C > 0$, $\alpha > 1$ and where $p_0(x)$ is symmetric around $x = 0$, i.e. $p_0(x) = p_0(-x)$, and such that

$$\int_{-1}^{+1} p_0(x) dx = A \tag{2}$$

where A is positive and finite (i.e. $0 < A < +\infty$).

Question 1.1

Find the value of C (as a function of α and A) such that $p(x)$ is a correctly normalized PDF.

Normalization requirement: $1 = \int_{-\infty}^{+\infty} p(x)dx = \int_{-\infty}^{-1} p(x)dx + \int_{-1}^{+1} p(x)dx + \int_{+1}^{+\infty} p(x)dx = A + 2\int_{+1}^{+\infty} p(x)dx =$
$= A + 2C\int_{+1}^{+\infty} x^{-\alpha}dx = A + \frac{2C}{\alpha-1}$

This implies $C = \frac{(1-A)(\alpha-1)}{2}$

Question 1.2

Calculate the mean value μ of $X_i(t)$. If necessary, distinguish different regimes depending on the value of α.

From the symmetry of $p(x)$ around zero, i.e. $p(x) = p(-x)$, it follows trivially that

$\mu \equiv \int_{-\infty}^{+\infty} xp(x)dx = \int_{-\infty}^{0} xp(x)dx + \int_{0}^{+\infty} xp(x)dx = -\int_{0}^{+\infty} xp(x)dx + \int_{0}^{+\infty} xp(x)dx = 0$

Note that this holds irrespective of the value of α.

Question 1.3

Calculate the variance σ^2 of $X_i(t)$ as a function of α, A and B, where $B \equiv \int_{-1}^{+1} x^2 p_0(x)dx$ is a positive and finite constant ($0 < B < +\infty$). If necessary, distinguish different regimes depending on the value of α.

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 p(x)dx = B + 2\int_{+1}^{+\infty} x^2 p(x)dx = \begin{cases} +\infty & \alpha < 3 \\ B + \frac{2C}{\alpha-3} = B + \frac{(\alpha-1)(1-A)}{\alpha-3} & \alpha > 3 \end{cases}$$

PROBLEM 2

With reference to Problem 1, assume that at time t the log-return $r(t)$ of the traded stock is determined by the demand of all agents as follows (where $1 \ll N < +\infty$ and $0 < D < +\infty$):

$$r(t) \equiv D \sum_{i=1}^{N} x_i(t) \tag{3}$$

Question 2.1

For each of the regimes you have identified in Problem 1, what can be concluded about the distribution of $r(t)$ and its parameters?

Case $\alpha > 3$:
all the variables $\{X_i(t)\}$ are i.i.d. with finite mean ($\mu = 0$) and finite variance ($\sigma^2 < 0$). (See Question 1.3.)
Since N is large, we can therefore apply the Central Limit Theorem (CLT) to eq.(3) and conclude
that the PDF of $r(t)$ is asymptotically a Normal distribution with mean

$$E[r(t)] = D \sum_{i=1}^{N} E[X_i(t)] = DN\mu = 0$$

and, since the $\{X_i(t)\}$ are i.i.d., variance

$$Var[r(t)] = D^2 Var\Big[\sum_{i=1}^{N} X_i(t)\Big] = D^2 \sum_{i=1}^{N} Var[X_i(t)] = D^2\sigma^2 N = D^2 N \left[B + \frac{\alpha + A - \alpha A}{\alpha - 3}\right]$$

Case $\alpha < 3$:
all the variables $\{X_i(t)\}$ are i.i.d. with finite mean ($\mu = 0$) and *infinite* variance ($\sigma^2 = +\infty$). (See Question 1.3.)
Since N is large, the PDF of $r(t)$ is asymptotically a Levy-stable distribution with infinite variance, mean

$$E[r(t)] = D \sum_{i=1}^{N} E[X_i(t)] = DN\mu = 0$$

and power-law tail
$P[r] \propto r^{-\alpha}$ where α is the same exponent as the tail of $p(x)$, so $1 < \alpha < 3$.

Question 2.2

For each of the regimes you have identified in Problem 1, what does the distribution of $r(t)$ have in common with the empirical log-return distributions and what differs?

Case $\alpha > 3$:
What differs: the PDF of $r(t)$ is asymptotically a Normal distribution, while real log-return distributions have power-law tails $P[r] \propto r^{-\beta}$ with exponent $\beta > 3$.
What is similar: aggregational normality (normality of the distribution of longer and longer sums of log-returns) is reproduced.

Case $\alpha < 3$:
What is similar: the PDF of $r(t)$ is asymptotically a distribution with power-law tail $P[r] \propto r^{-\alpha}$, i.e. *qualitatively* similar to real log-return distributions that have power-law tails $P[r] \propto r^{-\beta}$.
What differs: here the exponent is $1 < \alpha < 3$, while in most empirical log-return distributions it is $\beta > 3$.
Moreover, aggregational normality is *not* reproduced since here the Levy-stable distribution with exponent $1 < \alpha < 3$ is preserved at all frequencies.

PROBLEM 3

With reference again to Problem 1, imagine that the correlation coefficient C_{ij} between the time series of the demands of agents i and j is calculated for each i and j, and that the resulting $N \times N$ correlation matrix $\mathbf{C}$ is obtained. Assume that N is very large and that $T = 4N$.

(Note: $\boldsymbol{C}$ is the correlation matrix between different agents' demands for the same stock, not between price returns of different stocks. Still, the same definitions for stock return correlation matrices can be applied.)

Question 3.1

Write the PDF of the eigenvalues $\{\eta\}$ of $\mathbf{C}$. Explain your result.

Since the $\{X_i(t)\}$ are i.i.d., the correlation matrix $\mathbf{C}$ is a Wishart matrix consistent with Random Matrix Theory (RMT). Since N is very large and $T = 4N$, the eigenvalues of $\mathbf{C}$ are asymptotically distributed according to the Sengupta-Mitra distribution with parameters:
$Q \equiv T/N = 4, \quad \eta_{max} = [1 + 1/\sqrt{Q}]^2 = [3/2]^2 = 9/4, \quad \eta_{min} = [1 - 1/\sqrt{Q}]^2 = [1/2]^2 = 1/4.$

So $P(\eta) = \begin{cases} \frac{2}{\pi}\frac{\sqrt{(9/4-\eta)(\eta-1/4)}}{\eta} & 1/4 \leq \eta \leq 9/4 \\ 0 & else \end{cases}$

Question 3.2

Write in a compact form the market-mode component $\mathbf{C}^m$ and the group-mode component $\mathbf{C}^g$ of $\mathbf{C}$.

Since $\mathbf{C}$ is described by the above Sengupta-Mitra distribution, it has no deviating eigenvalue outside the range $[\eta_{min}, \eta_{max}]$, so it does not have any market-mode or group-mode component:

$\mathbf{C}^m = 0$

$\mathbf{C}^g = 0$

Question 3.3

Now write the PDF of the components $\{\phi\}$ of one eigenvector of the matrix $\mathbf{C} - \mathbf{C}^m - \mathbf{C}^g$. Explain your result.

Since $\mathbf{C} - \mathbf{C}^m - \mathbf{C}^g = \mathbf{C}$ is a Wishart matrix consistent with RMT, the components $\{\phi\}$ of its eigenvectors are distributed according to the Porter-Thomas distribution:
$P(\phi) = \frac{e^{-\phi^2/2}}{\sqrt{2\pi}}$

Question 3.4

Compare the properties of $\mathbf{C}$ with those observed in empirical stock return correlations.

$\mathbf{C}$ does not have any market-mode or group-mode component, while real stock return correlations have a very strong market mode characterized by a dominating eigenvalue $\eta_{market} \gg \eta_{max}$ and significant group-mode correlations characterized by eigenvalues in the range $(\eta_{max}, \eta_{market})$, partly reflecting industrial sectors.

PROBLEM 4

With reference to Problem 3, filter the correlation matrix $\mathbf{C}$ to define a network among the N agents: for each pair of agents i and j (with $i \neq j$), draw an undirected link if $C_{ij} > 0$, otherwise leave the two agents not connected.

Question 4.1

Write the expected number of links L, the expected average degree $\bar{k}$ and the expected link density c in the network.

Since $\mathbf{C}$ comes from i.i.d. time series, its non-diagonal entries (note: not its eigenvalues!) are symmetrically distributed around zero. So half of the non-diagonal entries will be such that $C_{ij} > 0$ and another half will be such that $C_{ij} < 0$. Then the network described above will have half the number of possible links:

$L = \frac{1}{2}\frac{N(N-1)}{2} = \frac{N(N-1)}{4}$ $\qquad$ $\bar{k} = \frac{N-1}{2} \approx \frac{N}{2}$ $\qquad$ $c = \frac{1}{2}$

Question 4.2

Now consider an Erdos-Renyi random graph model with connection probability p. What value of p would be required to obtain the same properties you found in Question 4.1?

We would need a half-connected random graph, i.e. $p = \frac{1}{2}$

Question 4.3

Let $\mathbf{G}_1$ be the graph generated from the correlation matrix $\mathbf{C}$ and let $\mathbf{G}_2$ be a graph generated using the Erdos-Renyi model with the value of p you found in Question 4.2.

Are pairs of links statistically independent in $\mathbf{G}_1$? No

Why? Because correlations (even when generated by random i.i.d. data) must obey constraints, such as the fact that if $C_{ij} = 1$ and $C_{jk} = 1$ then $C_{ik} = 1$. Therefore if i is connected to j and j is connected to k, it is more likely that i is connected to k than it is not. Another explanation is the fact that correlations can be transformed to metric distances (e.g. when one constructs the Minimum Spanning Tree), and must therefore obey some 'transformed' triangular inequality, making triples of correlations dependent on each other.

Are pairs of links statistically independent in $\mathbf{G}_2$? Yes

Why? Because links are drawn independently by construction in the Erdos-Renyi model.

Question 4.4

If $C(\mathbf{G}_1)$ and $C(\mathbf{G}_2)$ denote the average clustering coefficients of $\mathbf{G}_1$ and $\mathbf{G}_2$ respectively, do you expect that $C(\mathbf{G}_1) = C(\mathbf{G}_2)$, $C(\mathbf{G}_1) > C(\mathbf{G}_2)$ or $C(\mathbf{G}_1) < C(\mathbf{G}_2)$?

$$C(\mathbf{G}_1) \quad > \quad C(\mathbf{G}_2)$$

Why? Because of the above dependencies, it is more likely to form triangles in $C(\mathbf{G}_1)$ than in $C(\mathbf{G}_2)$.