Model for correlations in stock markets

Jae Dong Noh

Center for Theoretical Physics, Seoul National University, Seoul 151-742, Korea

We propose a group model for correlations in stock markets. In the group model the markets are composed of several groups, within which the stock price fluctuations are correlated. The spectral properties of empirical correlation matrices reported recently are well understood from the model. It provides the connection between the spectral properties of the empirical correlation matrix and the structure of correlations in stock markets.

PACS number(s): 05.40.Ca, 05.45.Tp, 87.23.Ge, 02.10.Sp

There has been growing interest of physicists in economic systems. They exhibit various interesting complex behaviors, e.g., a power-law distribution of the variations of financial market indices [1–3], and many efforts have been devoted to understanding the temporal correlations [4–6]. The economic systems are composed of a large number of interacting units. The information of the correlations and interactions is provided by a correlation matrix [7–9]. Its matrix elements are given by the overlap between price changes of different assets. They measure a extent to which the price fluctuations of different assets are correlated.

Recently Laloux *et al.* [7] and Plerou *et al.* [8] studied spectral properties of the empirical correlation matrix $\mathbf{C}$ of stock price changes in real markets. Comparing the empirical data with the theoretical prediction of the random matrix theory, they observed a correlated aspect of the financial markets. However, the nature of the correlations is not revealed yet. In this work, we propose that the observed spectral properties reflect a specific structure of correlations in the financial markets. To this end, we postulate a model for the correlations and show that it explains all aspects of the observed spectral properties.

Following the notation of Ref. [7], the empirical correlation matrix is constructed from the time series of stock price changes $\delta x_i(t)$, where $i(i=1,\ldots,N)$ labels the assets and $1\leq t\leq T$ the time. It is assumed that δx's are shifted and rescaled to have zero mean and a unit variance [7]. The correlation matrix has elements

$$C_{ij}=(\vec{\delta x}_i)\cdot(\vec{\delta x}_j). \tag{1}$$

Here we introduced a symbolic notation that $\vec{x}\cdot\vec{y}=1/T\Sigma_{t=1}^T x(t)y(t)$.

Suppose that there exist only random correlations: $\delta x_i(t)$ are independent identically-distributed random variables. In that case the distribution of eigenvalues λ of $\mathbf{C}$ is exactly known in the limit $N,M\to\infty$ with $Q=T/N\geq 1$ fixed [10]. The density of the eigenvalues is given by

$$\rho(\lambda)=\frac{Q}{2\pi}\frac{\sqrt{(\lambda_{max}-\lambda)(\lambda-\lambda_{min})}}{\lambda}, \tag{2}$$

for $\lambda\in[\lambda_{min},\lambda_{max}]$ and zero otherwise, where $\lambda_{max}=(1+1/\sqrt{Q})^2$ and $\lambda_{min}=(1-1/\sqrt{Q})^2$. Deviations from Eq. (2) provide clues about the structure of the financial markets and the nature of the underlying interactions.

Here we briefly summarize the finding of Refs. [7] and [8]. They constructed the correlation matrix of N assets from the major stock markets and calculated the density of eigenvalues, $\rho(\lambda)$, and the inverse participation ratio, I. It is related to the reciprocal of the number of eigenvector components significantly different from zero, and defined as

$$I_k=\sum_{l=1}^N [u_{kl}]^4, \tag{3}$$

where $k(k=1,\ldots,N)$ refers to the eigenstate and u_{kl} is the lth components of its normalized eigenvector. It is observed that $\rho(\lambda)$ follows the distribution in Eq. (2) for small $\lambda\lesssim 2.0$ [7,8]. Those low-lying eigenstates reflect a random noisy part of the correlation. Besides, it is also observed that there appear several isolated eigenstates for large $\lambda>2.0$ [7,8] and that the eigenstates at both ends of the spectrum have large value of I [8]. However, an appropriate explanation for these is not presented.

To explains the observed spectral properties, we propose a model for the correlations, which will be called the group model. In the group model it is assumed that the stock price changes of companies in the same kind of industry have a strong correlation. It is in agreement with our everyday experiences and a quantitative study supports this [9]. Those sets of companies will be called the groups. The group feature is incorporated into the model by assuming that the stock price change $\delta x_i(t)$ of a company $i(=1,\ldots,N)$ at time $t(=1,\ldots,T)$ is given by a sum of two random variables:

$$\delta x_i(t)=f_{\alpha_i}(t)+\epsilon_i(t), \tag{4}$$

where α_i denotes the group to which a company i belongs, f_{α_i} represents a synchronous variation of the group α_i, and ϵ_i a random noisy part. The noise strength is parametrized by ε_{α_i} with which the variances of f_{α_i} and ϵ_i are given by $(1+\varepsilon_{\alpha_i}^2)^{-1}$ and $\varepsilon_{\alpha_i}^2(1+\varepsilon_{\alpha_i}^2)^{-1}$, respectively. The correlation matrix takes the form of $\mathbf{C}=\mathbf{C}^0+\mathbf{C}^1$ with $C_{ij}^0=\delta_{\alpha_i,\alpha_j}$ ($\delta_{\alpha,\beta}$: the Kronecker delta symbol).

When $\varepsilon_\alpha=0$ and T is infinite, $\mathbf{C}=\mathbf{C}^0$ becomes a block diagonal matrix; a direct product of $N_\alpha\times N_\alpha$ matrices with all elements being unity (N_α: the size of the group α). Each block has a nondegenerate eigenstate with the eigenvalue $\lambda=N_\alpha$ and $(N_\alpha-1)$-fold degenerate eigenstates with $\lambda=0$.

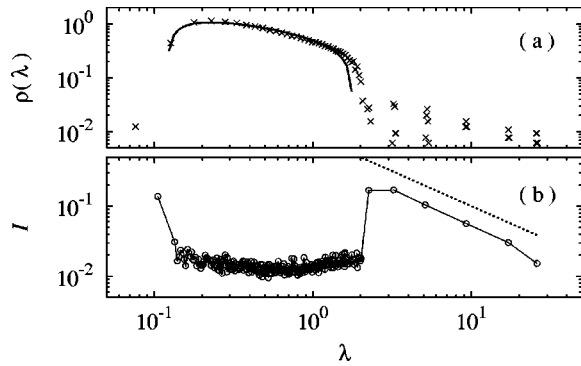FIG. 1. (a) $\rho(\lambda)(\times)$ and a fitting curve to the form in Eq. (2) (solid line). (b) Inverse participation ratio with a plot of $1/\lambda$ against λ (dotted line).

The inverse participation ratio of the nondegenerate eigenstates is the inverse of the group size, $I=1/N_\alpha$.

For nonzero ε_α and finite T, the random overlap part $\mathbf{C}^1$ becomes nonzero. Its effect can be treated as a perturbation to $\mathbf{C}^0$. Up to the first order perturbation, the eigenvalues of the nondegenerate states are scaled down by a factor $(1+\varepsilon_\alpha^2)^{-1}$ and the degenerate states split into the states whose eigenspectrum are given by those of $\mathbf{C}^1$. Therefore, the perturbed degenerate eigenstates with small λ will result in the distribution of the eigenvalues of Eq. (2), while the nondegenerate states isolated peaks at $\lambda \simeq (1+\varepsilon_\alpha^2)^{-1}N_\alpha$ as found in Ref. [7] and [8].

The information of the correlation is mainly contained in the isolated eigenstates. Each one represents a correlated group, whose size and participating companies are obtained from the eigenvector and its eigenvalue. The eigenstates with small eigenvalues, which are approximately the eigenstates of $\mathbf{C}^1$, may also provide the correlation information. Since $\mathbf{C}^1$ generates random overlaps between groups, its eigenvector components spread over the whole market in general. However, the overlap may be confined mainly within the group, provided that the noise strength of a group is small. Then there may appear an eigenstate with the eigenvalue near the unperturbed value, zero, and with relatively large inverse participation ratio, as observed in Ref. [8].

We demonstrate the validity of the group model by examining the spectral properties of an artificial correlation matrix constructed from a time series of Eq. (4) with $N=254$ and $T=800$. A part of them is separated arbitrarily into groups of size $N_\alpha=2,4,8,16,32,64$ and the remaining ones are assumed to behave independently. The noise strength is assigned to $\varepsilon_\alpha=(1-1/N_\alpha)$. Figure 1 shows $\rho(\lambda)$, ensemble-averaged over 50 realizations of the time series, in (a) and the inverse participation ratio I for a realization in (b). They are plotted in the log-log scale to show the large and small λ regions on an equal footing. They are in excellent agreement with the empirical data shown in Refs. [7] and [8]. $\rho(\lambda)$ for $\lambda<2.0$ is well fitted to Eq. (2). It deviates from Eq. (2) for $\lambda>2.0$. The isolated peaks represents the groups of size $N_\alpha=4,\ldots,64$. A peak for the group of $N_\alpha=2$ is smeared with the continuous distribution. The inverse participation ratio is large at both ends of the spectrum. For large λ, λ and I are inversely proportional to each other since λ is proportional to N_α, and I to $1/N_\alpha$. The eigenstate with the smallest eigenvalue has also large value of I. It has large components within the group of size $N_\alpha=2$, which has smallest noise strength.

In summary, we proposed the group model to explain the spectral properties of the empirical correlation matrices. According to the group model, the stock markets are composed of several correlated groups of companies. The stock price fluctuations of companies within a group have a synchronous component. The groups manifest themselves as the eigenstates of $\mathbf{C}$ with large eigenvalues. The sizes of groups and composing companies are known from the eigenvalues and the eigenvectors. It will be an interesting work to reveal the group structure of the real markets from the empirical correlations matrix.

[1] R. N. Mantegna and H. E. Stanley, Nature (London) **376**, 46 (1995).

[2] P. Gopikrishnan, M. Meyer, L. A. N. Amaral, and H. E. Stanley, Eur. Phys. J. B **3**, 139 (1998); P. Gopikrishnan *et al.*, Phys. Rev. E **60**, 5305 (1999).

[3] V. Plerou *et al.*, e-print cond-mat/9907161.

[4] P. Bak, M. Paczuski, and M. Shubik, Physica A **246**, 430 (1997).

[5] T. Lux and M. Marchesi, Nature (London) **397**, 498 (1999).

[6] V. M. Eguíluz and M. G. Zimmerman, e-print cond-mat/9908069; R. D'Hulst and G. J. Rodgers, e-print cond-mat/9908481.

[7] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).

[8] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).

[9] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).

[10] A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).