Random Matrix Theory of Dynamical Cross Correlations in Financial Data

Yasuhiro Nakayama and Hiroshi Iyetomi*)

Department of Physics, Niigata University, Niigata 950-2181, Japan

A new method taking advantage of the random matrix theory is proposed to extract genuine dynamical correlations between price fluctuations of different stocks. One-day returns of 557 Japanese major stocks for the 11-year period from 1996 to 2006 are used for this study. We carry out the discrete Fourier transform of the returns to construct a correlation matrix at each frequency. Also we prepare series of random numbers which are mutually uncorrelated and hence serve as a reference. Comparison of the eigenvalues of the empirical correlation matrix with the reference results of the random one enables us to distinguish between information and noise involved in complicated behavior of the stock returns. It is thus demonstrated that there exist collective motions of the stock prices with periods well over days. Finally we indicate a possible application of the present finding to the risk evaluation of portfolios.

§1. Introduction

Recently cross correlations between price movements of different stocks have been studied very actively, and the results are applied to risk management of portfolio in financial engineering.[1)–3)] Since the length of financial time-series data is inevitably finite, it is a central issue how to extract genuine correlations in such limited data for improving the risk control. The random matrix theory (RMT), combined with the principal component analysis, has been used successfully to this end.[4)–7)] Distributions of the eigenvalues and components of the eigenvectors for random matrices have various universal properties including Wigner's semicircle law.[8)] One can thus take advantage of this fascinating feature of random matrices to distinguish between information and noise involved in complicated behavior of the stock returns. Meaningful information as regards cross correlations may be obtained by observing the largest eigenvalues of an empirical correlation matrix which are far away from the RMT prediction.

It has been elucidated that the autocorrelation of stock prices has very short lifetime; it disappears beyond 30 minutes.[1),2)] Cross correlations in the financial time series have been thereby studied mainly at equal time so far. However, it is well-known in many-body physical problems that collective motions are totally separated from individual motions. For instance, we would never reach an idea of sound wave only through observation of individual molecules moving so randomly in air. Here we thus pay our attention to not only equal-time (static) cross correlations but also time-lagged (dynamic) cross correlations, and propose a new RMT method to extract collective motions in financial time series. Some efforts[9),10)] have already made to analyze such dynamical correlations, but research in this direction has just been initiated.

*) E-mail: hiyetomi@sc.niigata-u.ac.jp, http://www.sc.niigata-u.ac.jp/

In the following section we first confirm the previous results based on the RMT for the static cross correlations using one-day returns of Japanese major stocks. In §3 we present our extension of the RMT method for dynamical correlations along with results of the new analysis on the real data; there truly exist collective motions of the stock prices with periods well over days. Section 4 is devoted to removal of market trend. Then we conclude by indicating a possible application of the present finding to the risk evaluation of portfolios.

§2. Static cross correlations

2.1. *Stock prices as stochastic variables*

We regard temporal variation of stock prices as a stochastic process. To characterize correlation properties of the financial data, we use the logarithmic return of stock:

$$r_i(t) \equiv \ln S_i(t+1) - \ln S_i(t), \tag{2·1}$$

where $S_i(t)$ is the price of stock i $(= 1, \cdots, N)$ at time t $(= 1, \cdots, T)$. If the simple return, namely, the relative increment of stock is small, e.g., 5%, there is no much difference between the two returns numerically. We then normalize the log return as

$$w_i(t) \equiv \frac{r_i(t) - \langle r_i \rangle}{\sigma_i}, \tag{2·2}$$

where $\sigma_i \equiv \sqrt{\langle (r_i - \langle r_i \rangle)^2 \rangle}$ is the standard deviation of r_i, and $\langle \cdots \rangle$ denotes a time-average over the period T. The normalized return $w_i(t)$ has thus zero mean and unit variance.

2.2. *Correlation matrix*

The equal-time cross correlation matrix $\boldsymbol{C}$ is defined as

$$\boldsymbol{C} \equiv \frac{1}{T} \boldsymbol{W} \boldsymbol{W}^t, \tag{2·3}$$

where $\boldsymbol{W}$ is a $N \times T$ rectangular matrix with elements $\{W_{it} \equiv w_i(t)\}$, and $\boldsymbol{W}^t$ denotes the transpose of $\boldsymbol{W}$. So the component of cross correlation matrix $\boldsymbol{C}$ can be written as

$$C_{ij} = \langle w_i(t) w_j(t) \rangle. \tag{2·4}$$

If we replace $\{w_i(t)\}$ by sequences of random numbers with zero mean and unit variance, then we obtain a random correlation matrix $\boldsymbol{R}$ corresponding to $\boldsymbol{C}$ for the real data. The RMT has elucidated a number of notable statistical properties of random matrices.[8] According to the RMT, in the limit $N \to \infty$, $T \to \infty$, such that $Q = \frac{T}{N} (\geq 1)$ is fixed, the probability density $\rho(\lambda)$ of eigenvalues λ of $\boldsymbol{R}$ is given[2),11),12)] by

$$\rho(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{2·5}$$

where the upper bound λ_+ and the lower bound λ_- of the eigenvalues are given by

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \tag{2·6}$$

The i-th component of the eigenvector associated with the eigenvalue λ_α ($\alpha = 1, \ldots, N$) for the correlation matrices is denoted as $u_{\alpha,i}$. We normalize it such that $\sum_{i=1}^{N} {u_{\alpha,i}}^2 = N$. The distribution of eigenvector components of the random correlation matrix should conform[13)] to a Gaussian distribution with mean zero and unit variance,

$$\rho(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right). \tag{2·7}$$

2.3. *Application of the RMT*

Here we analyze a database from the Tokyo Stock Exchange. We use one-day return for an 11-yr period of from 1996 to 2006 ($T = 2707$ days). We extract from this database 557 stocks that are priced for all business days in that period. Figure 1(a) shows $\rho(\lambda)$ for the eigenvalues of $\boldsymbol{C}$. The eigenvalue distribution is remarkably different from the RMT prediction, Eq. (2·5), depicted by the solid curve. There are 13 eigenvalues larger than λ_+. Especially, the largest eigenvalue $\lambda \simeq 132$ is 60 times larger than the maximum eigenvalue predicted for uncorrelated time series. Figure 1(b) shows $\rho(\lambda)$ of the correlation matrix calculated with the real data but randomly shuffled in the temporal direction. The results agree well with the RMT prediction, Eq. (2·5). This agreement ascertains existence of true cross correlations involved in the stock price fluctuations.

As shown in Fig. 2(a), the density distribution $\rho(u)$ of eigenvector components for the largest eigenvalue, much larger than λ_+, deviates remarkably from the RMT prediction, Eq. (2·7). The eigenvector components are well localized with the same sign, suggesting coherent movement of all of the stock prices. This mode is thus referred to as "market trend". Figures 2(b) and (c) show $\rho(u)$ for the second- and third-largest eigenvalues, respectively. Those eigenvalues are also far away from the RMT bulk region $\lambda_- < \lambda < \lambda_+$, and accordingly the associated $\rho(u)$'s deviate systematically from the RMT prediction. As has been already shown,[6)] the deviating eigenvectors divide the set of all stocks studied into business sectors of different characteristics. Figure 2(d) shows $\rho(u)$ for the 300th-largest eigenvalue which is buried in the RMT bulk. We observe no appreciable deviation from the RMT prediction.

§3. Dynamical cross correlations

3.1. *Correlation matrix in Fourier space*

We define a time-lagged correlation function as

$$C_{ij}(\tau) = \langle w_i(t+\tau) w_j(t) \rangle, \tag{3·1}$$

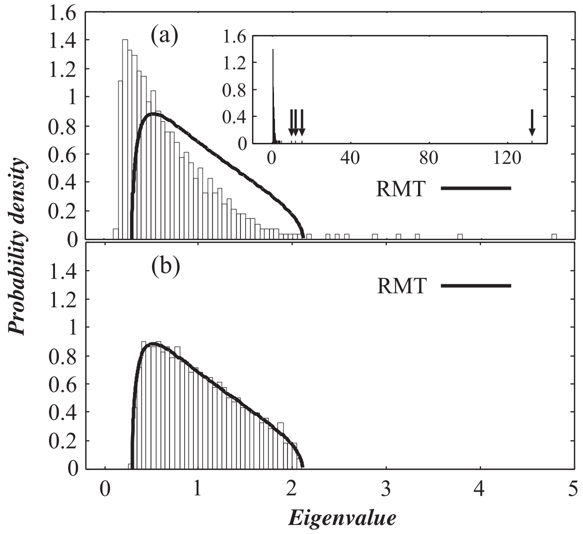

Fig. 1. (a) The probability density function of eigenvalues of $\boldsymbol{C}$, which is extracted from $N = 557$ stocks of the Tokyo Stock Exchange during the years 1996–2006. In the present case $T = 2707$, then $Q = \frac{T}{N} = 4.86$. If the returns are random, the eigenvalues have to be distributed in the interval $0.30 \leq \lambda \leq 2.11$. (b) The probability density function of eigenvalues of the correlation matrix constructed from the real data, but randomly shuffled in the temporal direction.

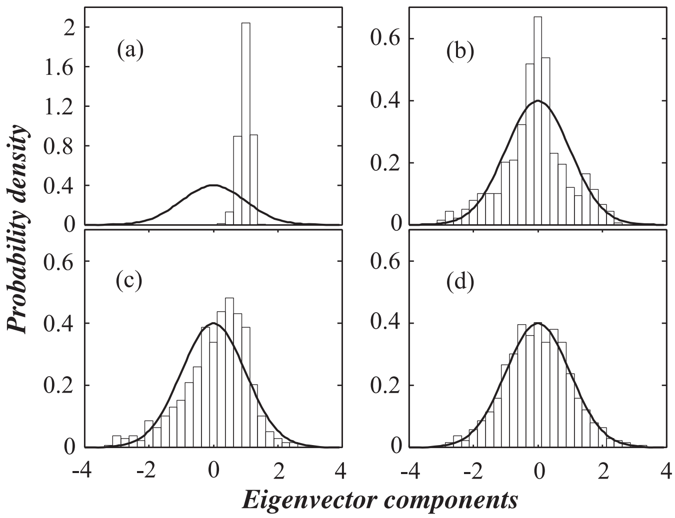

Fig. 2. (a) Distribution $\rho(u)$ of eigenvector components for the largest eigenvalue. (b) $\rho(u)$ for the second-largest eigenvalue. (c) $\rho(u)$ for the fifth-largest eigenvalue. (d) $\rho(u)$ for the 300th-largest eigenvalue.

where τ is the time-lag. The discrete Fourier transform of $C_{ij}(\tau)$ is

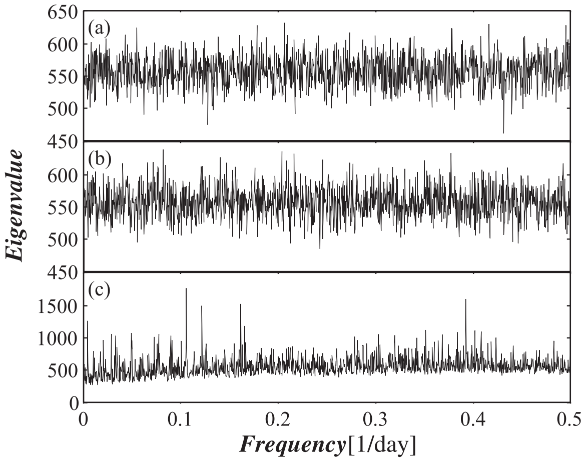

Fig. 3. Frequency dependence of the eigenvalues for random data (a); that for the real data but randomly shuffled (b); that for the real data (c). Note that the scale of the vertical axis in the panel (c) is remarkably different from that in the others.

$$\sum_\tau C_{ij}(\tau)\exp(-i\omega\tau) = \frac{1}{T}\sum_\tau\sum_t w_i(t+\tau)w_j(t)\exp(-i\omega\tau)$$
$$= \frac{1}{T}W_i(\omega)W_j^*(\omega) \equiv C_{ij}(\omega), \tag{3·2}$$

where $W_i(\omega) \equiv \sum_t w_i(t)\exp(-i\omega t)$ and we have imposed a cyclic boundary condition on the returns in temporal direction with the period T. Then the correlation matrix in Fourier space is given by

$$\boldsymbol{C}(\omega) = \frac{1}{T}\begin{bmatrix} W_1(\omega) \\ W_2(\omega) \\ \vdots \\ W_N(\omega)\end{bmatrix}\begin{bmatrix} W_1^*(\omega) & W_2^*(\omega) & \cdots & W_N^*(\omega)\end{bmatrix}. \tag{3·3}$$

The dynamical correlation matrix $\boldsymbol{C}(\omega)$ is related to the equal-time correlation matrix $\boldsymbol{C}$ through the sum rule:

$$\boldsymbol{C} = \frac{1}{T}\sum_\omega \boldsymbol{C}(\omega). \tag{3·4}$$

3.2. *Frequency dependence of eigenvalues*

The Fourier decomposition of stock price fluctuations reduces the rank of the correlation matrix to 1, so that $\boldsymbol{C}(\omega)$ has only a single eigenvalue $\lambda_\omega(\neq 0)$. The eigenvalue is explicitly given by

$$\lambda_\omega = \frac{1}{T}\sum_j |W_j(\omega)|^2$$
$$= \frac{1}{T}\sum_j\sum_t\sum_{t'} w_j(t)w_j(t')\exp\left\{i(t'-t)\omega\right\}, \tag{3·5}$$

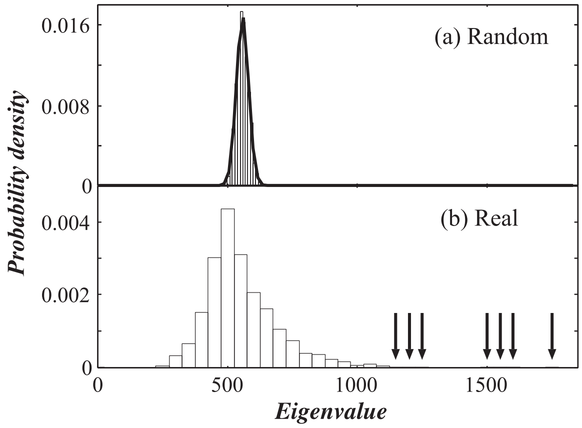

Fig. 4. Probability density for the eigenvalues of the random correlation matrix (a); that of the empirical correlation matrix over all frequencies (b). The solid line in the panel (a) refers to the RMT result, Eq. (3·6).

which amounts to total sum as regards the power spectrum of each of stock price fluctuations. In the case when $w_j(t)$ are random numbers with zero mean and unit variance, the eigenvalues at any frequency are distributed according to the normal distribution with mean and variance both of N for large N:

$$\rho(\lambda) = \frac{1}{\sqrt{2\pi N}} \exp\left(-\frac{(\lambda - N)^2}{2N} \right). \tag{3·6}$$

This is readily proved by invoking the central limit theorem.

We calculated λ_ω's for $\boldsymbol{C}(\omega)$ using the same real data as for $\boldsymbol{C}$. Also we repeated the same calculation by replacing the real data with random time series or the randomly-shuffled real data. Frequency dependence of the eigenvalues so obtained with the three different data is compiled in Fig. 3. The two results based on random time series and the randomly-shuffled real data agree with each other as shown in the panels (a) and (b) of Fig. 3. The panel (c), in contrast, demonstrates a number of spikes in the frequency spectrum of the eigenvalues for the real data. Comparison between the panels (b) and (c) really confirms that dynamical correlations between price fluctuations are also present behind the stock market.

Figure 4(a) shows the probability density $\rho(\lambda)$ of eigenvalues for the random data, which is in good agreement with the RMT prediction, Eq. (3·6). The average of the eigenvalue is 556.8 with the standard deviation of 23.5. As shown in Fig. 4(b), on the other hand, $\rho(\lambda)$ for the real market data significantly deviates from the RMT result. The average of the eigenvalue is 557.0 and its standard deviation is 149.0. The largest eigenvalue $\lambda = 1769.7$ is about 3.18 times larger than the average. Then the averages of eigenvalues are almost identical for both data, but the standard deviation of the eigenvalue distribution for the real data is much larger than that for the random data. According to Eq. (3·5), the large eigenvalues far apart from Eq. (3·6) imply that stock prices move coherently in a statistically meaningful way at periods corresponding to these eigenvalues.

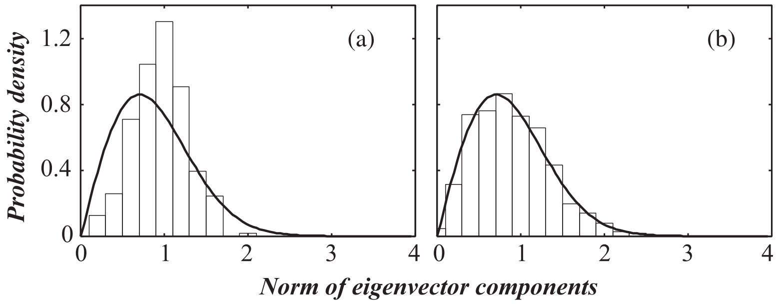

Fig. 5. (a) Probability density $\rho(|u|)$ of the norm of eigenvector components for the largest eigenvalue $\lambda \approx 1769.7$ (period $P \approx 9.47$ days). (b) $\rho(|u|)$ for the third-smallest eigenvalue $\lambda \approx 264.0$ ($P \approx 142.5$ days). The solid curve depicts Eq. (3·7).

3.3. *Distribution of eigenvector components*

The i-th component of the eigenvector corresponding to the eigenvalue λ_ω is denoted as $u_{\omega,i}$. The normalization is determined by $\sum_{i=1}^{N} |u_{\omega,i}|^2 = N$. The central limit theorem again proves that the distribution of norm $|u|$ of eigenvector components for the random correlation matrix should approach

$$\rho(|u|) = 2|u| \exp\left(-|u|^2\right), \tag{3·7}$$

as N is increased. Figure 5(a) confirms that $\rho(|u|)$ for the largest eigenvalue of the real correlation matrix $\boldsymbol{C}(\omega)$ also deviates from the RMT prediction, Eq. (3·7). On the other hand, the panel (b) of Fig. 5 demonstrates that $\rho(|u|)$'s for small eigenvalues agree well with Eq. (3·7).

3.4. *Dynamics of market trend*

As has been shown, the components coherently participate in the eigenvector associated with the largest eigenvalue for the static correlation matrix. It thereby represents an influence that is common to all stocks. Such market trend is also reflected in the eigenvectors with large eigenvalues for the dynamical correlation matrix. The fact is demonstrated in Fig. 6, where the eigenvector components corresponding to the first ten largest eigenvalues of the dynamical correlation matrix are displayed in the complex plane. Those outliers in fact represent the market trend.

Figure 7 highlights the top five dominant components of the market trend eigenvector for the static correlation matrix in Fig. 6. Tracing their behavior with varied frequency, we see that the dominant components in the market trend do not always lead in the dynamical correlations. We have thus been able to resolve the market trend into dynamically inter-correlated components each of which has a characteristic frequency (period).

§4. Removal of market trend

Finally we try to remove the market trend out of the real stock price data following the procedure adopted in Ref. 6). We assume the market return common

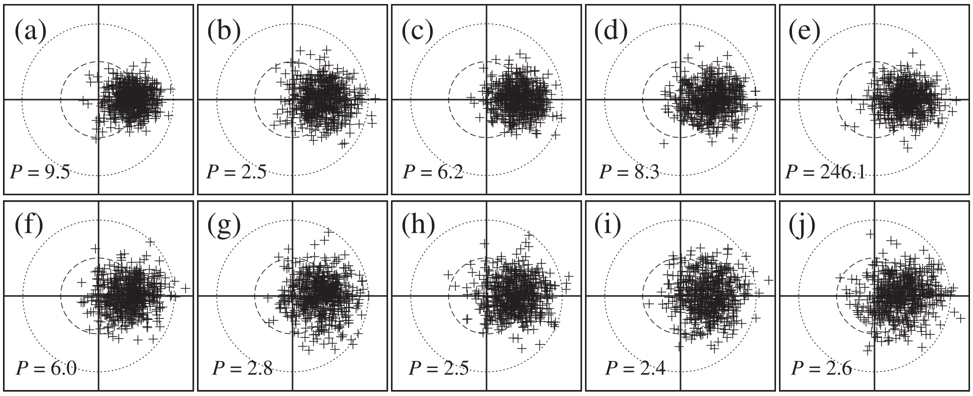

Fig. 6. Distribution of the eigenvector components of the dynamical correlation matrix in the complex plane for the first ten largest eigenvalues. The alphabetical assignment of the panels is in accordance with the order of the eigenvalues; namely, the panel (a) is the result for the largest eigenvalue. The period P corresponding to each eigenvalue is given in units of day.

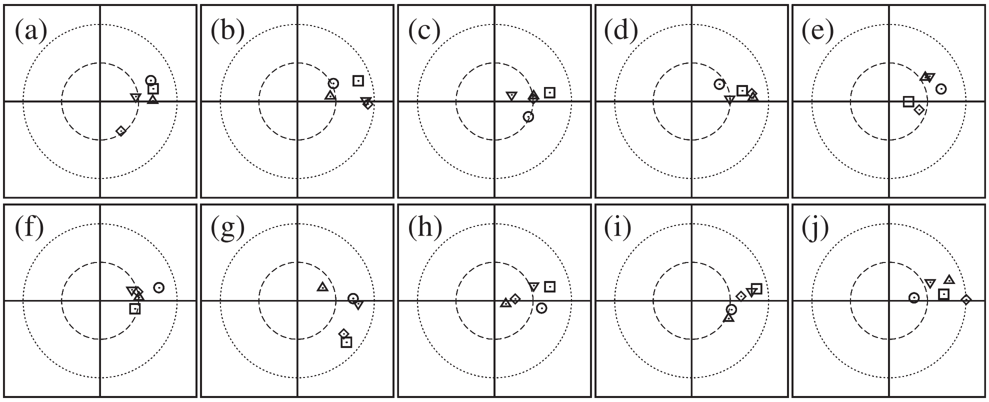

Fig. 7. Same as Fig. 6, but only the top five stocks as regards the static market trend are illuminated. The symbols, square, circle, triangle, inverse triangle, and diamond are ordered according to the rank of stocks; the square displays where the greatest trend follower at each characteristic period is.

to all stocks is given by

$$r_{market}(t) = \frac{1}{N} \sum_{i=1}^{N} v_{1,i} r_i(t), \tag{4·1}$$

where $v_{1,i}$ is the eigenvector corresponding to the largest eigenvalue λ_1. To remove this market trend from the data, we simply use the linear regression formula:

$$r_i(t) = \alpha_i + \beta_i r_{market}(t) + \varepsilon_i(t), \tag{4·2}$$

where α_i and β_i are stock-specific constants, and the residuals $\varepsilon_i(t)$ carry information on cross correlations inherent at deeper level in the data. Thus we adopt $\varepsilon_i(t)$ instead of $r_i(t)$ to reconstruct the static correlation matrix $\boldsymbol{C}'$.

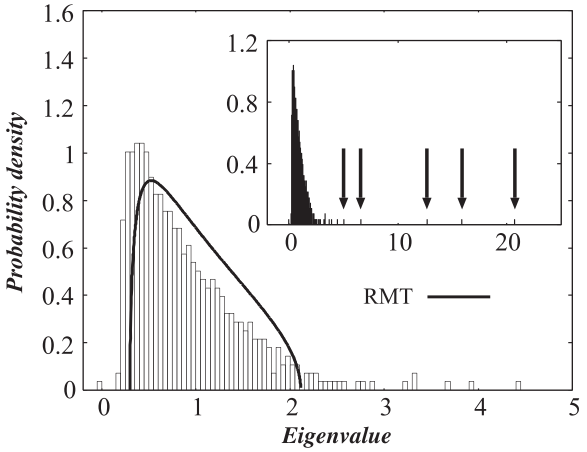

Fig. 8. The probability density of eigenvalues for the static cross-correlation matrix $\boldsymbol{C}'$, where the market trend was removed using Eq. (4·1).

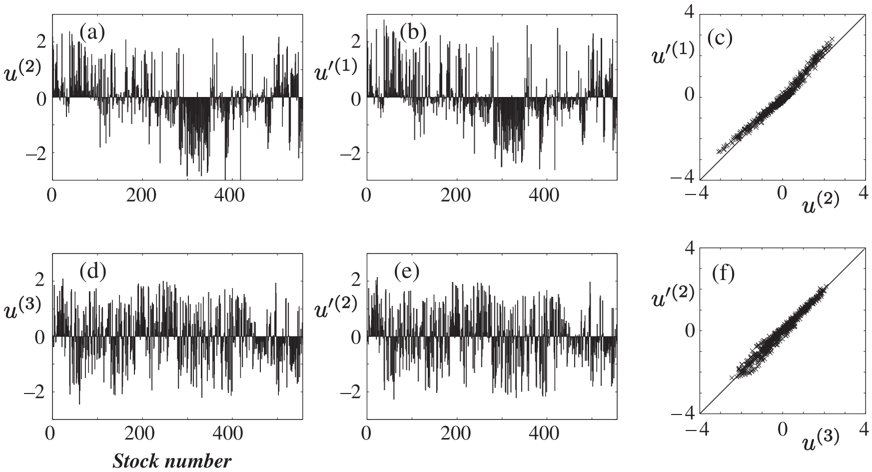

Fig. 9. (a) Eigenvector components $u^{(2)}$ for the second-largest eigenvalue of $\boldsymbol{C}$. (b) Eigenvector components $u'^{(1)}$ for the largest eigenvalue of $\boldsymbol{C}'$. (c) Comparsion of the eigenvector components in (a) and (b). (d) Eigenvector components $u^{(3)}$ for the third-largest eigenvalue of $\boldsymbol{C}$. (e) Eigenvector components $u'^{(2)}$ for the second-largest eigenvalue of $\boldsymbol{C}'$. (f) Comparison of the eigenvector components in (d) and (e).

Figure 8 shows the probability density of the eigenvalues for $\boldsymbol{C}'$. The largest eigenvalue ($\lambda \approx 20.8$) for $\boldsymbol{C}'$ is much smaller than the largest eigenvalue ($\lambda \approx 132$) for $\boldsymbol{C}$. This indicates the market trend is removed well in $\boldsymbol{C}'$. Therefore we expect that the second-largest eigenvalue of $\boldsymbol{C}$ may correspond to the largest eigenvalue of $\boldsymbol{C}'$ and the third-largest eigenvalue of $\boldsymbol{C}$, to the second-largest eigenvalue of $\boldsymbol{C}'$. This is really true as demonstrated by the panels of Fig. 9.

Then we turn our attention to eigenvalues λ_ω for the correlation matrix in Fourier space constructed using $\varepsilon_i(t)$. Figure 10 shows frequency spectrum of eigenvalues for the market trend removed data with and without random shuffling. Removal of

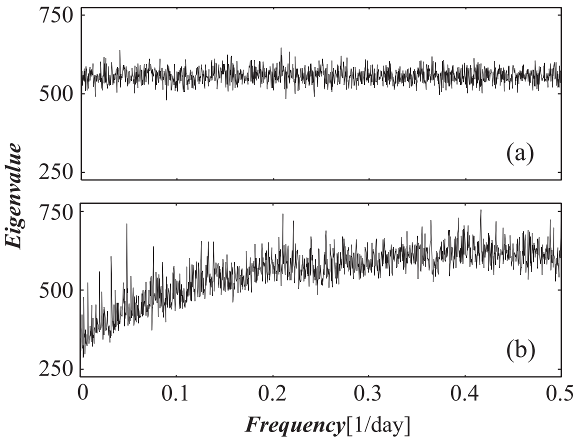

Fig. 10. Frequency spectrum of the eigenvalue for the market trend removed data (a) randomly shuffled and (b) as they are.

the market trend significantly shrinks the variance of eignevalues based on the real data. However, we still observe notable outliners. Those should arise from more sophisticated correlations in the price movement dynamics, which are grasped by the eigenvectors corresponding to the largest eigenvalues of $\boldsymbol{C}$ but except for the largest one. In addition to the spiky structures, we find secular deviation of the real result from the reference based on the artificial data. This may imply that the cross correlations have such long lifetime as to compete with the observation period T for them or that the market itself have changed its correlation structure. Further analysis is necessary to identify the origin of the systematic trend in the real data.

§5. Concluding remarks

We have proposed a new method taking advantage of the random matrix theory to extract genuine dynamical correlations between price fluctuations of different stocks. Comparison of the eigenvalues of the empirical correlation matrix with the reference results of the random one demonstrated existence of collective motions of the stock prices with periods well over days. Detailed analysis on such dynamical cross correlations in the stock price data is in progress.

We conclude this paper by indicating a possible application of the present finding to the risk evaluation of portfolios. The risk of a portfolio is usually measured by the following variance:

$$\sigma_{\mathrm{P}}^2 = \sum_{i,j=1}^{M} p_i p_j \sigma_i \sigma_j C_{ij}, \qquad (5{\cdot}1)$$

where p_i is the weight of stock i with variance σ_i^2. If we take account of dynamical cross correlations between stock price fluctuations, however, the formula (5·1) is generalized to

$$\tilde{\sigma}_{\mathrm{P}}^2 = \sum_{i,j=1}^{M} p_i p_j \sigma_i \sigma_j \tilde{C}_{ij}\,(T_{\mathrm{P}}), \qquad (5\cdot2)$$

where T_{P} is the time interval of portfolio and the static correlation factor C_{ij} appearing in Eq. (5·1) is replaced by the dynamical one defined as

$$\tilde{C}_{ij}\,(T_{\mathrm{P}}) = \frac{1}{T_{\mathrm{P}}} \sum_{t}^{T_{\mathrm{P}}} \sum_{t'}^{T_{\mathrm{P}}} \langle w_i(t) w_j(t') \rangle. \qquad (5\cdot3)$$

Combination of Eq. (5·2) with information on the dynamical cross correlations may thus improve the predictability of financial risk in constructing portfolios.

Acknowledgements

We thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-07-16 on "Econophysics III: Physical Approach to Social and Economic Phenomena" were highly beneficial to completing this work. We also would like to thank Ken Millennium Corporation for its financial support.

References

1) R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000).
2) J. P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivate Pricing: From Statistical Physics to Risk Management*, 2nd ed. (Cambridge University Press, Cambridge, 2003).
3) J. Voit, *The Statistical Mechanics of Financial Markets*, 3rd ed. (Springer, Berlin, 2005).
4) L. Laloux, P. Cizeau, J. P. Bouchaud and M. Potters, Phys. Rev. Lett. **83** (1999), 1467.
5) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, Phys. Rev. Lett. **83** (1999), 1471.
6) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr and H. E. Stanley, Phys. Rev. E **65** (2002), 066126.
7) D.-H. Kim and H. Jeong, Phys. Rev. E **72** (2005), 046133.
8) M. L. Mehta, *Random Matrices*, 3rd ed. (Academic Press, New York, 2004).
9) C. Biely and S. Thurner, physics/0609053.
10) S. Thurner and C. Biely, Acta Phys. Pol. B **38** (2007), 4111.
11) A. Edelmann, SIMA Journal of Matrix Analysis and Applications **9** (1988), 543.
12) A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60** (1999), 3389.
13) T. Guhr, A. Müller-Groeling and H. A. Weidenmüller, Phys. Rep. **299** (1998), 189.