Regular Article

Statistical properties of cross-correlation in the Korean stock market

G. Oh[1], C. Eom[2], F. Wang[3], W.-S. Jung[3,4,5,a], H.E. Stanley[3], and S. Kim[5,6]

[1] Division of Business Administration, Chosun University, 501-759 Gwangju, Republic of Korea
[2] Division of Business Administration, Pusan National University, 609-735 Busan, Republic of Korea
[3] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
[4] Graduate Program for Technology and Innovation Managemant, Pohang University of Science and Technology, 790-784 Pohang, Republic of Korea
[5] Department of Physics, Pohang University of Science and Technology, 790-784 Pohang, Republic of Korea
[6] Asia Pacific Center for Theoretical Physics, 790-784 Pohang, Republic of Korea

Abstract. We investigate the statistical properties of the cross-correlation matrix between individual stocks traded in the Korean stock market using the random matrix theory (RMT) and observe how these affect the portfolio weights in the Markowitz portfolio theory. We find that the distribution of the cross-correlation matrix is positively skewed and changes over time. We find that the eigenvalue distribution of original cross-correlation matrix deviates from the eigenvalues predicted by the RMT, and the largest eigenvalue is 52 times larger than the maximum value among the eigenvalues predicted by the RMT. The β_{473} coefficient, which reflect the largest eigenvalue property, is 0.8, while one of the eigenvalues in the RMT is approximately zero. Notably, we show that the entropy function $E(\sigma)$ with the portfolio risk σ for the original and filtered cross-correlation matrices are consistent with a power-law function, $E(\sigma) \sim \sigma^{-\gamma}$, with the exponent $\gamma \sim 2.92$ and those for Asian currency crisis decreases significantly.

1 Introduction

Financial markets have been known as representative complex systems, which are organized by various unexpected phenomenon according to non-trivial interactions among heterogeneous agents [1,2]. The study of complex economic systems is not easy because we do not know the control parameters that govern economic systems and can not easily apply the parameters we do know to economic systems. However, much research has been conducted to understand the statistical properties of financial time series [3–11]. In particular, the analysis of financial data by various methods developed in statistical physics has become a very interesting research area for physicists and economists [12,13]. There is practical [14–16] as well as scientifically important value in analyzing the correlation coefficient between stock return time series because this contains a significant amout of information on the nonlinear interactions in the financial market and is a parameter in terms of the Markowitz portfolio theory. The cross-correlation matrix between stocks, which has unexpected properties due to complex behaviors, such as temporal non-equilibrium, mispricing, bubbles, market crashes and so on, is an important parameter to understand the interactions in the financial market [17]. To analyze the

cross-correlation matrix, previous studies presented various statistical methods, such as principal component analysis (PCA) [18], singular value decomposition (SVD) [19] and factor analysis (FA) [20]. Here, to analyze the actual cross-correlation matrix, we employ the random matrix theory (RMT), which was introduced by Wigner, Dyson and Metha [21–28]. It can explain the statistical properties of energy levels in complex nuclei well. The RMT method is a useful method for eliminating the randomness in the actual cross-correlation matrix [29–34]. Recently, Laloux et al. [35] and Plerou et al. [36] analyzed the cross-correlation matrix of financial time series by the RMT method. The authors found that 94% of the eigenvalues of cross-correlation matrix can be predicted by the RMT, while the other 6% of the eigenvalues deviated from the RMT. In addition, Plerou et al. [37] applied the RMT method to a United States stock market and observed that the cross-correlation matrix of stock markets consists of random and non-random parts, which read carry useful information in the financial market. The eigenvector deviations from the RMT show a very stable state over an entire period. We investigate the various statistical properties of the cross-correlation matrix of 473 daily stock return time series traded in the Korean stock market from 1 January 1993 to 31 May 2003. We find that the distribution of the cross-correlation matrix is positively skewed

[a] e-mail: `wsjung@postech.ac.kr`

and changes over the entire time. Using the RMT method, we show that the cross-correlation matrix contains meaningful information as well as random property. Notably, we show that for both the original, $C_{original}$, and filtered correlation, C_{filter}, matrices the entropy function, $E(\sigma)$, with the portfolio risk, σ, is consistent with a power-law function, $E(\sigma) \sim \sigma^{-\gamma}$, with an exponent $\gamma \sim 2.92$. In the following section, we describe the data and methods used in this paper. In Section 3, we present the verification results. Finally, we end with a conclusion.

2 Data and method

In this paper, we investigate the statistical properties of the cross-correlation matrix of the 473 daily stock returns traded on the Korean stock market from 3 January 1993 to 31 May 2003. The data obtained from the Korea Stock Exchanges cover 2845 days. To understand the non-trivial interactions, we calculated the cross-correlation matrix between stocks for the whole period as well as sub-periods by shifting 21 days with 250 data points. We propose a verification process to analyze the statistical properties of the correlation matrix between stock returns. First, we estimated the statistical properties of the cross-correlation matrices using the RMT method. Second, we calculate the entropy of the portfolio weights using the Markowitz portfolio theory. Before demonstrating the verification process, we introduce the RMT, which was proposed by Wigner, Dyson, and Metha et al. and Markowitz portfolio theory (MPT) [38] introduced by Markowitz in 1952. We created N (number of company) data sets with L data points following iid$(0, 1)$. Let the created data be denoted by the symbol G. Here, the G is a matrix $(N \times L)$ with the random elements and the cross-correlation matrix is defined by

$$C_{random} = \frac{1}{L}GG^T, \tag{1}$$

where G^T is the transpose of G, and the correlation between elements is approximately zero. If $N \to \infty$ and $L \to \infty$, the eigenvalue spectrum of RMT is calculated by using

$$P_{random}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{2}$$

where the eigenvalues λ lie within $\lambda_- \leq \lambda \leq \lambda_+$, $Q \equiv \frac{L}{N}$, and the maximum and minimum eigenvalue of RMT, C_{random}, are given by

$$\lambda \pm \equiv 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \tag{3}$$

If L and N have a finite length, then the eigenvalue spectrum shows gradual decrease from the theoretical values of the largest eigenvalue predicted by the RMT.

We next extend the MPT to select the optimal portfolio sets among all stocks. The MPT method introduced by Markowitz in 1952 is known as the mean-variance theory.

The purpose of MPT is to minimize the portfolio risk in a given portfolio return, which can be quantified by the variance and defined as follows.

$$\Omega = \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j C_{ij} \sigma_i \sigma_j, \tag{4}$$

where ω_i is the portfolio weight of stock i, which can be calculated using two Lagrange multipliers, σ_i is the standard deviation of stock i, and C_{ij} is the correlation coefficient between stock i and stock j. In this work, we use the no short-selling constraint for portfolio weights [12,13], i.e. we assume that all the weights are non negative numbers ($\omega_i > 0, \forall\, i = 1, \ldots, N$). We also normalize portfolio weights in such a way that $\sum_{i=1}^{N} \omega_i = 1$. The portfolio return, μ, also is calculated by

$$\mu = \sum_{i=1}^{N} \omega_i \mu_i, \tag{5}$$

where μ_i is the mean value of stock i. We next considered the portfolio weights because these could determine the portfolio efficiency frontier lines. We used Shannon's entropy method to quantify the statistical properties of the portfolio weights since $\sum_{i=1}^{N} \omega_i = 1$, defined by

$$E = \sum_{i=1}^{N} -P_i \ln(P_i), \tag{6}$$

where P_i is the portfolio weight w_i.

Using the eigenvalue distribution predicted by equation (2), we estimated a random part from the original cross-correlation matrix and as the previous paper [37], divided it two parts as follows.

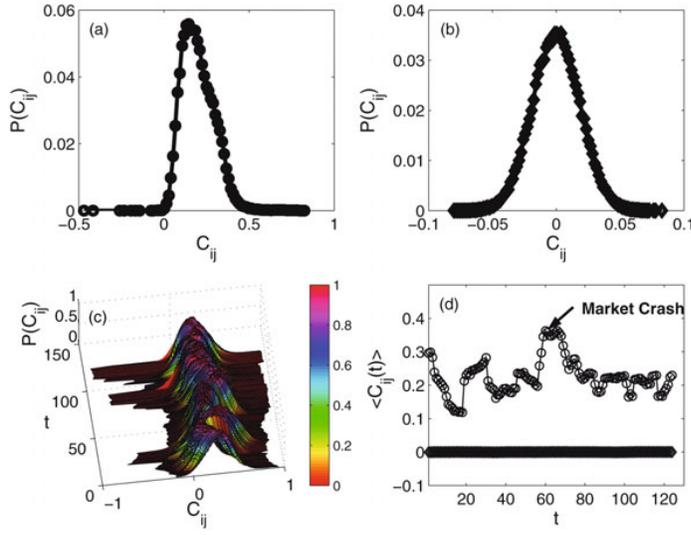$$C_{original} = C_{random} + C_{filter}. \tag{7}$$

Based on how many random elements existed in the cross-correlation matrix, we analyzed the non-trivial interactions between stocks. In addition, to estimate the eigenvalue properties, we created the data sets by using each eigenvector element.
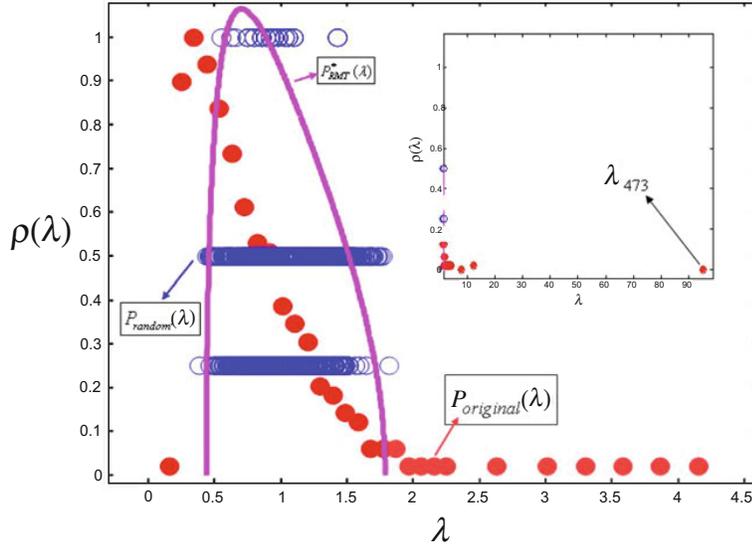
$$R(t) \equiv \sum_{i=1}^{N} V_i r_i(t), \tag{8}$$

where $r_i(t)$ is the ith stock return at time t, and V_i is the ith eigenvector. To observe the eigenvalue properties divided by the RMT method, we created the data sets, $R^{Random}(t)$ and $R^{Largest}(t)$, reflecting the eigenvalue properties of both C_{random} and C_{filter}, respectively, and, by the one-factor model, widely acknowledged in the financial literature as a pricing model, we calculated the relationship between the created time series and the market factor, which influences all stocks in the market and is defined by

$$r_i(t) = \alpha_i + \beta_i R_{Market}(t) + \epsilon_i(t), \tag{9}$$

where R_{Market} is the KOSPI market index, α_i and β_i are the regression coefficients of stock i and we use the β coefficient as the measurement to quantify the relationship between created data sets and market index.

Fig. 1. (Color online) (a) and (b) show the distribution of the cross-correlation coefficients between stocks of 473 companies of taken from the Korean stock market and random data, respectively. (c) displays the distribution of the cross-correlation matrices of the sub-periods by shifting 21 days with 251 data points and (d) shows the average values of each cross-correlation matrix in (c).
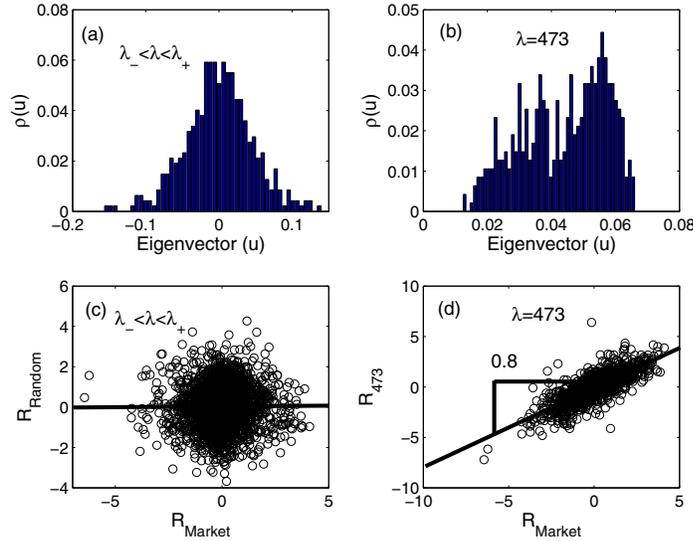

Fig. 2. (Color online) The distribution of the eigenvalues for cross-correlation matrix estimated using the 473 companies listed on the Korean stock market, random data following the iid$(0, 1)$ process, and that predicted by the RMT method. The red circles, blue circles, and pink solid-line indicate the original time series, random data, and theoretical lines, respectively.

3 Results

In this section, we analyze the various statistical features of the cross-correlation matrix of 473 daily stock returns listed on the Korean stock markets from 3 January 1993 to 31 May 2003 using the random matrix theory and Markowitz portfolio theory. We present the results on the statistical properties of the cross-correlation matrix, such as its distribution, eigenvalue spectrum and entropy of portfolio weights calculated by MPT. Figures 1a and 1b show the distribution of the cross-correlation matrices of the original and random data sets. Figure 1c shows the distribution of cross-correlation matrices calculated by shifting 21 days with 250 data points. Figure 1d displays the average value of cross-correlation matrices of Figure 1c. In Figure 1a, we find that the distribution of the cross-correlation matrix between stocks for a whole period is

positively skewed and shows a significant difference from that for the random interaction in Figure 1b. In Figure 1c, we show that the distribution of the cross-correlation matrix changes considerably over the whole time. Especially, in Figure 1d, during the Asian currency crisis, the mean values of the correlation coefficients significantly increased. In other words, the dynamically changes were caused by the complex behavior of the market crash, unlike the case of random interactions. Our findings confirm that all the possible interactions in the Korean stock market deviated from those for the random interaction.
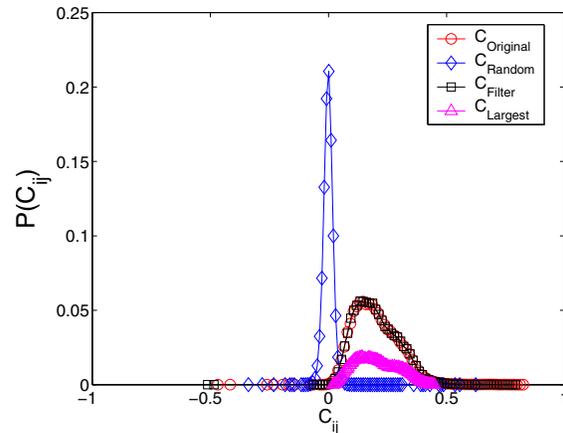
We next decompose the original cross-correlation matrix into the random C_{random} and filter C_{filter} parts using the RMT method to extract the meaningful information from the original cross-correlation matrix. Figure 2 shows the eigenvalue distribution of the cross-correlation matrix in the Korean stock market. In Figure 2, the

Fig. 3. (Color online) (a) and (b) show the distribution of both eigenvectors corresponding to λ_{100} and λ_{473}, respectively. (c) and (d) display the β coefficients between the normalized market index and the time series created by equation (9) for the eigenvalues λ_{100} and λ_{473}. The value of both β_{100} and β_{473} are zero and 0.8, respectively.

solid-line (orange) is the eigenvalue spectrum predicted by the RMT, and the red circles and blue circles indicate the eigenvalue distributions of the original time series and random data sets, respectively. In Figure 2, we find that the eigenvalue distribution of the RMT method is very similar to one from the random data, while that for the real time series significantly shows different behavior. Moreover, the largest eigenvalue is 52 times larger than the largest eigenvalue of the RMT. The large values are greater than 25 times those in the United States stock market [37].
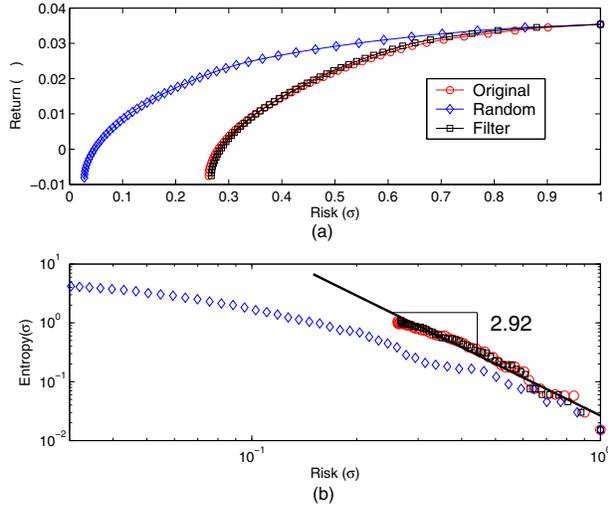
To characterize the statistical properties of each eigenvalues, we created the return time series using equation (8) and calculated the slopes β between those and the KOSPI market index using equation (9). Figures 3a and 3b shows the distribution of the eigenvector elements corresponding to both the largest eigenvalue, λ_{473} and λ_{100}, one of eigenvalues of the RMT, respectively. Figures 3c and 3d show the β coefficient between the KOSPI market index and the time series created. We find that the β_{473} between the market index and time series is 0.8, while one from the time series created using the eigenvector elements predicted by the RMT is approximately zero. We argue that the largest eigenvalue can explain the market properties well, but one from the ranges predicted by the RMT is uncorrelated to the market index. We also decomposed the original cross-correlation matrix according to each eigenvalue divided by the RMT method. Figure 4 shows the distribution of various cross-correlation matrices. The red circles, blue diamonds, black squares and pink triangles indicate the cross-correlation matrices of the original, random, filter and largest eigenvalues, respectively. Through the above findings, we can expect that the distribution of the random cross-correlation matrix C_{random} follows a Gaussian distribution, while the cross-correlation matrix C_{filter} estimated after removing the random components from the original cross-correlation matrix by the RMT

Fig. 4. (Color online) The distribution of the original cross-correlation matrix, $C_{original}$, and those created by the random matrix theory, C_{random}, C_{filter}, and $C_{largest}$, respectively. The red circles, blue diamonds, black squares and pink triangles indicate the cross-correlation matrices corresponding to the original, random, filter and largest eigenvalue, respectively.
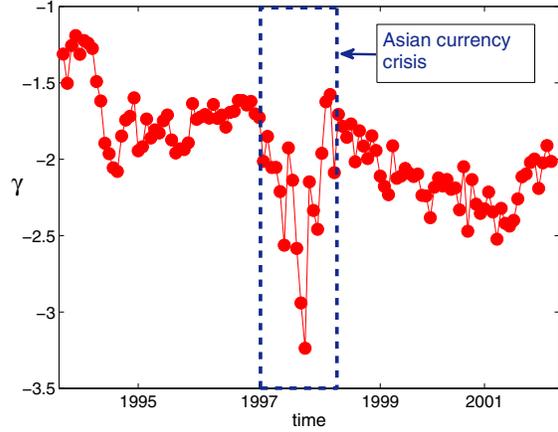
method has a similar distribution as the original time series. We found that the cross-correlation matrix reflecting the largest eigenvalue property has an obvious difference from that of the original time series.

To apply the RMT method to a portfolio optimization problem, we analyzed the portfolio weights estimated by the MPT through various cross-correlation matrices. The important parameters are the return, μ_i, standard deviation, σ and cross-correlation matrix, C_{ij}, of the original stock returns, which are needed to calculate the portfolio weights of each stock. To calculate the effects of the cross-correlation matrix among the input parameters, we apply the cross-correlation matrices, C_{filter},

Fig. 5. (Color online) (a) shows the efficient portfolio frontier lines for the original, $C_{original}$, random, C_{random}, and filter cross-correlation matrix, C_{filter}, respectively. (b) displays the relationship between the entropy of the portfolio weights and the portfolio risk.

Fig. 6. (Color online) The exponent of power-law function, $E(\sigma) \sim \sigma^{\gamma}$ estimated by the relationship between the portfolio risks, σ, and the entropy of portfolio weights.

4 Conclusions

We investigated the statistical properties of the cross-correlation matrix between the return time series of individual stocks traded in the Korean stock market using the RMT method and observed the effect of the cross-correlation matrix applied to the Markowitz portfolio theory. We found that the distribution of the cross-correlation matrix between stocks showed a positive skew and dynamically changed over time. We found that the eigenvalue distribution of the cross-correlation matrix deviated from those of the RMT, and the largest eigenvalue was 52 times larger than the eigenvalues predicted by the RMT. The slopes β between market index and the time series corresponding to the largest eigenvalue were 0.8, while those for the RMT were approximately zero. Notably, we found that the entropy function $E(\sigma)$ of portfolio weights with the portfolio risk σ was consistent with a power-law function, $E(\sigma) \sim \sigma^{-\gamma}$, with the exponent $\gamma \sim 2.92$, while the relationship between the entropy and portfolio return μ is not a power-law function. We find that while for all sub-periods the exponents calculated from the relationship between entropy of each portfolio weight and portfolio risk follow a power-law function, those for sub-periods changed over time and lie within $1.19 \leq \gamma \leq 3.23$. Especially, the exponent γ decreases significantly during Asian currency crisis. In the next step, we must rigorously study the portfolio weights of other stock markets because these play an important role in terms of the portfolio risk and return.

and C_{random} divided by the RMT method. Figure 5a shows the efficient portfolio lines created using the various cross-correlation matrices, such as $C_{original}$, C_{random}, and C_{filter}. In Figure 5a, we found that the efficient frontier lines calculated with both the original $C_{original}$ and filtered cross-correlation matrices C_{filter} show very similar behavior, while that of the random cross-correlation matrix C_{random} shows significant difference from the original correlation. In addition, the efficient portfolio frontier line of the random cross-correlation matrix C_{random} at a given portfolio risk σ overestimates the portfolio return, μ, by a greater amount than one of the original cross-correlation matrix. We next calculated the entropy of the portfolio weights with each cross-correlation matrix, such as $C_{original}$, C_{filter} and C_{random}. Figure 5b shows the relationship between the portfolio risk, σ, and the entropy of the portfolio weights for three types of cross-correlation matrices according to a log-log plot. We found that the entropy(σ) for both the original and filtered cross-correlation matrices was approximately consistent with a power-law function, $E(\sigma) \sim \sigma^{-\gamma}$ with the exponent $\gamma \sim 2.92$, while there is no the power-law function in the relationship between the entropy and the portfolio return, μ and presented in Figure 5b. We also calculated the exponents for each sub-periods by shifting 20 days with 500 data points to verify the stability over time the result observed in Figure 5. We find that while the relationship between entropy of each portfolio weight and portfolio risk follow a power-law function, the exponent values, γ, calculated from each sub-periods changes over time and lie within $1.19 \leq \gamma \leq 3.23$. Especially, the γ value calculated during the Asian currency crisis decreases significantly (Fig. 6).

References

1. T. Lux, M. Marchesi, Nature **397**, 498 (1999)
2. T. Lux, J. Econ. Behav. Organizat. **33**, 143165 (1998)
3. R.N. Mantegna et al., Nature **376**, 46 (1995)
4. R.N. Mantegna et al., Nature **383**, 587 (1996)
5. V. Plerou et al., Nature **421**, 130 (2003)

6. X. Gabaix et al., Nature **423**, 267 (2003)
7. G. Oh et al., J. Korean, Phys. Soc. **48**, 197 (2006)
8. Y. Liu et al., Phys. Rev. E **60**, 1390 (1999)
9. P. Cizeau et al., Physica A **245**, 441 (1997)
10. K. Yamasaki et al., Proc. Natl. Acad. Sci. **102**, 9424 (2005)
11. W.C. Jun et al., Phys. Rev. E **73**, 066128 (2006)
12. R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlation and Complexity in Finance* (Cambridge University Press, Cambridge, UK, 1999)
13. J.-P. Bouchaud, M. Potters, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridge, USA, 2004)
14. E. Elton, M. Gruber, *Modern Portfolio Theory and Investment Analysis* (Wiley, New York, 1981)
15. Y. Okhrin, W. Schmid, J. Econom. **134**, 235 (2006)
16. D. Sornette, P. Simonetti, J.V. Andersen, Phys. Rep. **335**, 19 (2002)
17. J.D. Noh, Phys. Rev. E **61**, 5981 (2000)
18. J.E. Jackson, *A User's Guide to Principal Components*, 2nd edn. (Wiley-Interscience, New York, 2003)
19. J.E. Gentle, *Singular Value Factorization* (Springer-Verlag, Berlin, 1998)
20. D.F. Morrison, *Multivariate Statistical Methods* (New York, McGraw-Hill, 1990)
21. M.L. Mehta, *Random Matrices* (Academic Press, Boston, 1991)
22. E.P. Wigner, Ann. Math. **53**, 36 (1951)
23. E.P. Wigner, Proc. Cambridge Philos. Soc. **47**, 790 (1951)
24. F.J. Dyson, J. Math. Phys. **3**, 140 (1962)
25. F.J. Dyson, M.L. Mehta, J. Math. Phys. **4**, 701 (1963)
26. M.L. Metha, Nucl. Phys. **18**, 395 (1960)
27. M.L. Mehta, F.J. Dyson, J. Math. Phys. **4**, 713 (1963)
28. M.L. Metha, Commun. Math. Phys. **20**, 245 (1971)
29. T. Guhr, A. Müller-Groeling, H.A. Weidenmüller, Phys. Rep. **299**, 190 (1998)
30. A.M. Sengupta, P.P. Mitra, Phys. Rev. E **60**, 3 (1999)
31. P. Sěba, Phys. Rev. Lett. **91**, 19 (2003)
32. A. Utsugi, K. Ino, M. Oshikawa, Phys. Rev. E **70**, 026110 (2004)
33. T. Guhr, B. Kälberzk, J. Phys. A: Math. Gen. **36**, 3009 (2003)
34. S. Shari, M. Crane, A. Shamaie, H. Ruskin, Physica A **335**, 629 (2004)
35. L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Phys. Rev. Lett. **83**, 1467 (1999)
36. V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999)
37. V. Plerou, P. Gopikrishnan, B. Rosennow, L.A.N. Amaral, T. Guhr, H.E. Stanley, Phys. Rev. E **64**, 066126 (2002)
38. H.M. Markowitz, J. Finance **7**, 77 (1952)
39. B. Rosenow, V. Plerou, P. Gopikrishnan, L.A.N. Amaral, H.E. Stanley, Int. J. Theoret. Appl. Finance **3**, 399 (2000)