ELSEVIER

Physica A 299 (2001) 175–180

PHYSICA A

www.elsevier.com/locate/physa

Collective behavior of stock price movements—a random matrix theory approach

V. Plerou[a,*], P. Gopikrishnan[a], B. Rosenow[b], L.A.N. Amaral[a], H.E. Stanley[a]

[a] *Center for Polymer Studies, Department of Physics, Boston University, Boston, MA 02215, USA*
[b] *Institut für Theoretische Physik, Universität zu Köln, D-50937 Köln, Germany*

Abstract

We review recent work on quantifying collective behavior among stocks by applying the conceptual framework of random matrix theory (RMT), developed in physics to describe the energy levels of complex systems. RMT makes predictions for "universal" properties that do not depend on the interactions between the elements comprising the system; deviations from RMT provide clues regarding system-specific properties. We compare the statistics of the cross-correlation matrix C—whose elements C_{ij} are the correlation coefficients of price fluctuations of stock i and j—against a random matrix having the same symmetry properties. It is found that RMT methods can distinguish random and non-random parts of C. The non-random part of C which deviates from RMT results, provides information regarding genuine collective behavior among stocks.

Keywords: Random matrix theory; Collective phenomena; Financial time series

The problem of quantifying cross-correlations between the price movements of different stocks is important not only from the point of view of understanding collective behavior between the constituents of a complex system, but also from the point of view of estimating the risk of a investment portfolio. The usual way of quantifying cross-correlations is either by estimating the relevant "factors"or by principal component analysis [1]. Here we review some results of a different approach to this problem applying methods of random matrix theory [2–7].

In order to quantify correlations, we first calculate the price change ("return") of stock $i = 1,\dots,N$ over a time scale Δt

$$G_i(t) \equiv \ln S_i(t + \Delta t) - \ln S_i(t)\,, \tag{1}$$

* Corresponding author.
E-mail address: plerou@cgl.bu.edu (V. Plerou).

0378-4371/01/$ - see front matter © 2001 Published by Elsevier Science B.V.
PII: S0378-4371(01)00293-X

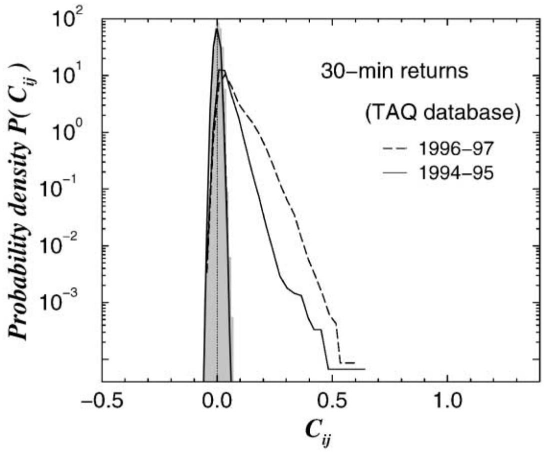

Fig. 1. $P(C_{ij})$ for C calculated using 30-min returns of 1000 stocks for the 2-y period 1994–95 (solid line) and 881 stocks for the 2-y period 1996–97 (dashed line). For the period 1996–97 $\langle C_{ij}\rangle = 0.06$, larger than the value $\langle C_{ij}\rangle = 0.03$ for 1994–95. The shaded region shows the distribution of correlation coefficients for the control $P(R_{ij})$ of Eq. (5), which is consistent with a Gaussian distribution with zero mean.

where $S_i(t)$ denotes the price of stock i. We analyze $L = 6448$ records 30-min price changes $G_i(t)$ for $N = 1000$ stocks (largest by market capitalization on 1 January 1994) for the 2-y period 1994–95. Since different stocks have varying levels of volatility (standard deviation), we define a normalized return

$$g_i(t) \equiv \frac{G_i(t) - \langle G_i\rangle}{\sigma_i}\,, \tag{2}$$

where $\sigma_i \equiv \sqrt{\langle G_i^2\rangle - \langle G_i\rangle^2}$ is the standard deviation of G_i, and $\langle\cdots\rangle$ denotes a time average over the period studied. We then compute the equal-time cross-correlation matrix C with elements

$$C_{ij} \equiv \langle g_i(t)\, g_j(t)\rangle\,. \tag{3}$$

By construction, the elements C_{ij} are restricted to the domain $-1 \leqslant C_{ij} \leqslant 1$, where $C_{ij} = 1$ corresponds to perfect correlations, $C_{ij} = -1$ corresponds to perfect anti-correlations, and $C_{ij} = 0$ corresponds to uncorrelated pairs of stocks. In matrix notation, the correlation matrix can be expressed as

$$\mathsf{C} = \frac{1}{L}\,\mathsf{G}\,\mathsf{G}^{\mathrm{T}}\,, \tag{4}$$

where G is an $N \times L$ matrix with elements $\{g_{i\,m} \equiv g_i(m\Delta t);\ i = 1,\ldots,N;\ m = 0,\ldots,L-1\}$, and G^{T} denotes the transpose of G.

We analyze the distribution $P(C_{ij})$ of the elements $\{C_{ij};\ i \neq j\}$ of the cross-correlation matrix C. We first examine $P(C_{ij})$ for 30-min returns from the TAQ database for the 2-y periods 1994–95 and 1996–97 (Fig. 1). First, we note that $P(C_{ij})$ is asymmetric and centered around a positive mean value ($\langle C_{ij}\rangle > 0$), implying that positively correlated behavior is more prevalent than negatively correlated (anti-correlated) behavior.

Secondly, we find that $\langle C_{ij} \rangle$ depends on time, e.g., the period 1996–97 shows a larger $\langle C_{ij} \rangle$ than the period 1994–95. We contrast $P(C_{ij})$ with a control—a correlation matrix R with elements R_{ij} constructed from $N = 1000$ mutually uncorrelated time series, each of length $L = 6448$, generated using the empirically found distribution of stock returns [8,9]. Fig. 1(a) shows that $P(R_{ij})$ is consistent with a Gaussian with zero mean, in contrast to $P(C_{ij})$. In addition, we see that the part of $P(C_{ij})$ for $C_{ij} < 0$ (which corresponds to anti-correlations) is within the Gaussian curve for the control, suggesting the possibility that the observed negative cross-correlations in C may be an effect of randomness.

Although by construction the elements of C are supposed to express the pairwise correlations that exist in the system, in practice, their meaning is not clear because of the time average involved in their calculation. Time averaging over a finite time series introduces measurement "noise" whereas the use of long time series amounts to averaging over possibly changing correlations. This raises the following problem: how can we extract from C, the cross-correlations that are significant?

The approach followed here is to compare the empirical cross-correlation matrix C against the "null hypothesis" of a random matrix of the same type ("symmetry"). Therefore, we consider a random correlation matrix

$$\mathsf{R} = \frac{1}{L} \mathsf{A}\,\mathsf{A}^{\mathrm{T}} , \tag{5}$$

where A is an $N \times L$ matrix containing N time series of L random elements with zero mean and unit variance, that are mutually uncorrelated. By construction R belongs to the type of matrices often referred to as Wishart matrices in multivariate statistics [10].

The comparison between C and R is performed in the diagonal basis. Thus, we first compute the eigenvalues λ_k and eigenvectors u^k, where $k = 1, \ldots, N$ is arranged in order of increasing eigenvalues. Statistical properties of the eigenvalues of random matrices such as R are known [11,12] in the limit of very large dimensions. Particularly, in the limit $N \to \infty$, $L \to \infty$, such that $Q \equiv L/N$ is fixed, it was shown analytically [12] that the distribution $P_{\mathrm{rm}}(\lambda)$ of eigenvalues λ of the random correlation matrix R is given by

$$P_{\mathrm{rm}}(\lambda) = \frac{Q}{2\pi} \sqrt{\frac{(\lambda_+ - \lambda)(\lambda - \lambda_-)}{\lambda}} \tag{6}$$

for λ within the bounds $\lambda_- \leqslant \lambda_i \leqslant \lambda_+$, where λ_- and λ_+ are the minimum and maximum eigenvalues of R, respectively, given by

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2 \sqrt{\frac{1}{Q}} . \tag{7}$$

We first compare the eigenvalue distribution of C and compare against $P_{\mathrm{rm}}(\lambda)$ (Fig. 2). Our observations are two-fold. First, we observe that the "bulk" of the eigenvalues of C are consistent with $P_{\mathrm{rm}}(\lambda)$ [2,3]. This suggests the randomness of the bulk which can be tested more rigorously by comparing against universal features of eigenvalue correlations of real symmetric random matrices. Specifically, our examination

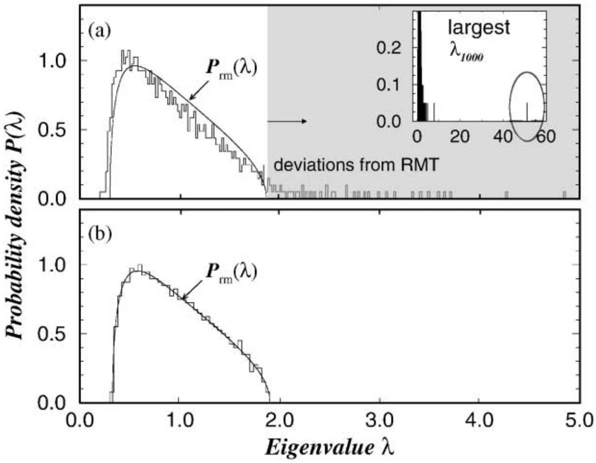

Fig. 2. (a) Eigenvalue distribution $P(\lambda)$ for C constructed from the 30-m returns for 1000 stocks for the 2-y period 1994–95. The solid curve shows the RMT result $P_{\mathrm{rm}}(\lambda)$ of Eq. (6). We note several eigenvalues outside the RMT upper bound λ_+ (shaded region). The inset shows the largest eigenvalue $\lambda_{1000} \approx 50 \gg \lambda_+$. (b) $P(\lambda)$ for the random correlation matrix R, computed from $N = 1000$ computer-generated random uncorrelated time series with length $L = 6448$ shows good agreement with the RMT result, Eq. (6) (solid curve).

of the eigenvalue spacing distribution shows good agreement with the results for real symmetric (GOE-type) random matrices.

Secondly, in Fig. 2, we find deviations from RMT for the largest few eigenvalues [2,3]. These deviations are also evident when one examines the distribution of eigenvector components [2,3]. Fig. 3(a) shows that $\rho(u)$ for a typical u^k from the bulk shows good agreement with the RMT result $\rho_{\mathrm{rm}}(u)$. Similar analysis on the other eigenvectors belonging to eigenvalues within the bulk yields consistent results, in agreement with the results of the previous sections that the bulk agrees with random matrix predictions. Consider next the "deviating" eigenvalues λ_i, larger than the RMT upper bound, $\lambda_i > \lambda_+$. Fig. 3(b) and (c) show that, for deviating eigenvalues, the distribution of eigenvector components $\rho(u)$ deviates systematically from the RMT result $\rho_{\mathrm{rm}}(u)$. Finally, we examine the distribution of the components of the eigenvector u^{1000} corresponding to the largest eigenvalue λ_{1000}.

Fig. 3(d) shows that $\rho(u^{1000})$ deviates remarkably from a Gaussian. We observe from $\rho(u^{1000})$ that all stocks contribute almost equally, and the distribution is rather narrow, suggesting that this eigenvector represents a collective mode in which all stocks participate. This notion can be quantified by comparing the price fluctuations of the portfolio defined by the u^{1000} against a standard measure of the fluctuations of the entire market—the fluctuations of the S&P 500 index. This comparison shows an equal-time correlation coefficient of 0.85 showing good agreement [7]. Thus, the eigenvector corresponding to the largest eigenvalue represents a collective mode in which all companies participate.

The magnitude of the largest eigenvalue itself seems to reflect the degree of collective behavior, as can be seen by examining the time evolution of the largest eigenvalue.

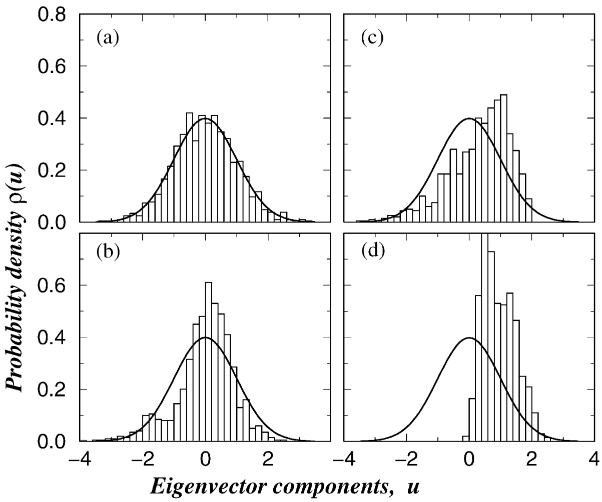

Fig. 3. (a) Distribution $\rho(u)$ of eigenvector components for one eigenvalue in the bulk $\lambda_- < \lambda < \lambda_+$ shows good agreement with the RMT prediction of a Gaussian with zero mean (solid curve). Similar results are obtained for other eigenvalues in the bulk. $\rho(u)$ for (b) u^{996} and (c) u^{999}, corresponding to eigenvalues larger than the RMT upper bound λ_+ (shaded region in Fig. 2). (d) $\rho(u)$ for u^{1000} deviates significantly from the Gaussian prediction of RMT. The above plots are for C constructed from 30-min returns for the 2-y period 1994–95. We also obtain similar results for C constructed from daily returns.

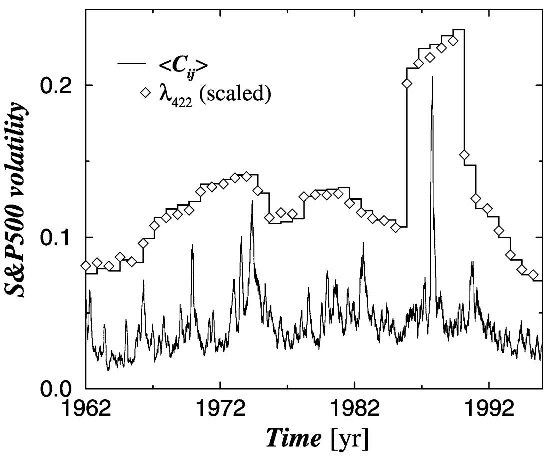

Fig. 4. The stair-step curve shows the average value of the correlation coefficients $\langle C_{ij} \rangle$, calculated from 422×422 correlation matrices C constructed from daily returns using a sliding $L = 965$ day time window in discrete steps of $L/5 = 193$ days. The circles correspond to the largest eigenvalue λ_{422} (scaled by a factor 4×10^2) for the correlation matrices thus obtained. The bottom curve shows the S&P 500 volatility (scaled for clarity) calculated from daily records with a sliding window of length 40 days. We find that both $\langle C_{ij} \rangle$ and λ_{422} have large values for periods containing the market crash of October 19, 1987.

We consider daily price fluctuations of 422 stocks for the years 1962–96. Fig. 4 shows the time evolution of the largest eigenvalue λ_{422} compared against the time evolution of the S&P 500 index and the S&P 500 volatility. The large downward movement of the index in 1987 corresponds to the 1987 crash, when all stocks in the market almost simultaneously lost value; i.e., all stocks were moving more synchronously than usual. We see that during this period, the largest eigenvalue almost doubled in magnitude.

We also examine the remainder of the eigenvalues. Our analysis [7] shows that the eigenvectors corresponding to these eigenvalues have significant participants that corresponds to major industry groups. Thus, remaining deviating eigenvectors quantify collective behavior of stocks belonging to the same or related industries. We also find that one of the deviating eigenvectors contains mainly stocks of firms having business in Latin America. It is possible that this collective behavior is related to the large currency devaluation in Mexico during the end of 1994 [7]. Similar results were obtained by using ultra-metric concepts by Refs. [13,14].

These deviating eigenvectors also have interesting dynamical features. For example, we find that the price fluctuations corresponding to the portfolios defined by the deviating eigenvectors are characterized by time correlations that decay significantly slower than that for a random eigenvector or for an individual stock [7]. This is reminiscent of the phenomenon of critical slowing down where collective modes of the system display very large relaxation times in the vicinity of a critical point [15,16].

References

[1] J. Campbell, A.W. Lo, A. MacKinlay, The Econometrics of Financial Markets, Princeton University Press, Princeton, NJ, 1997.
[2] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Phys. Rev. Lett. 83 (1999) 1467.
[3] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Phys. Rev. Lett. 83 (1999) 1471.
[4] S. Drozdz, F. Gruemmer, F. Ruf, J. Speth, Physica A 287 (2000) 440.
[5] J.D. Noh, Phys. Rev. E 61 (2000) 5981.
[6] M. Marsili, e-print cond-mat/0003241.
[7] P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, e-print cond-mat/0011145.
[8] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, Phys. Rev. E 60 (1999) 6519.
[9] T. Lux, Appl. Financial Econ. 6 (1996) 463.
[10] R. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[11] F.J. Dyson, Rev. Mexicana Fís. 20 (1971) 231.
[12] A.M. Sengupta, P.P. Mitra, Phys. Rev. E 60 (1999) 3389.
[13] R.N. Mantegna, Eur. Phys. J. B 11 (1999) 193.
[14] G. Bonanno, F. Lillo, R.N. Mantegna, e-print cond-mat/0009350.
[15] P.C. Hohenberg, B.I. Halperin, Rev. Mod. Phys. 49 (1977) 435.
[16] S.-K. Ma, Modern Theory of Critical Phenomena, Benjamin, Reading, MA, 1976.