International Journal of Theoretical and Applied Finance
Vol. 3, No. 3 (2000) 399–403

APPLICATION OF RANDOM MATRIX THEORY TO STUDY CROSS-CORRELATIONS OF STOCK PRICES

BERND ROSENOW, VASILIKI PLEROU, PARAMESWARAN GOPIKRISHNAN,
LUÍS A. NUNES AMARAL and H. EUGENE STANLEY

Center for Polymer Studies and Department of Physics,
Boston University, Boston, MA 02215, USA

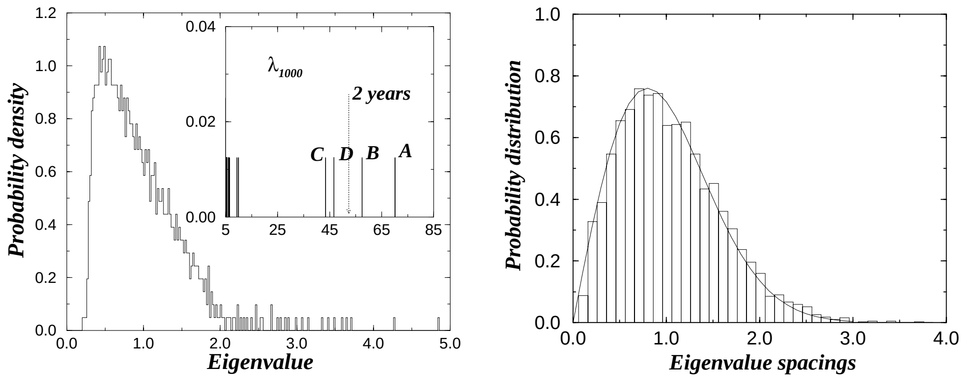
We address the question of how to precisely identify correlated behavior between different firms in the economy by applying methods of random matrix theory (RMT). Specifically, we use methods of random matrix theory to analyze the cross-correlation matrix $\mathbf{C}$ of price changes of the largest 1000 US stocks for the 2-year period 1994–1995. We find that the statistics of most of the eigenvalues in the spectrum of $\mathbf{C}$ agree with the predictions of random matrix theory, but there are deviations for a few of the largest eigenvalues. To prove that the rest of the eigenvalues are genuinely random, we test for universal properties such as eigenvalue spacings and eigenvalue correlations. We demonstrate that $\mathbf{C}$ shares universal properties with the Gaussian orthogonal ensemble of random matrices. In addition, we quantify the number of significant participants, that is companies, of the eigenvectors using the inverse participation ratio, and find eigenvectors with large inverse participation ratios at both edges of the eigenvalue spectrum — a situation reminiscent of results in localization theory.

Keywords: Cross-correlations, random matrix theory, stock portfolio, risk.

1. Introduction

Aside from scientific interest, the study of correlations between the returns of different stocks is also of practical relevance in quantifying the risk of a given stock portfolio [1]. Recently, the problem of understanding the correlations among the returns of different stocks has been addressed by applying methods of random matrix theory to the cross correlation matrix [2, 3]. Consider, e.g., the equal-time correlation of stock price changes for a given pair of companies. Since the market conditions may not be stationary, and the historical records are finite, it is not clear if a measured correlation of price changes of two stocks is just due to "noise" or genuinely arises from the interactions among the two companies.

In some ways, the problem of interpreting the correlations between individual stock-price changes is reminiscent of the difficulties experienced by physicists in the fifties, in interpreting the spectra of complex nuclei. With the minimal assumption of a random Hamiltonian, given by a real symmetric matrix with independent random elements, a series of remarkable predictions were made and successfully tested on the spectra of complex nuclei [4]. Deviations from the *universal* predictions of



(a) (b)

Fig. 1. (a) The probability density of the eigenvalues of the normalized cross-correlation matrix **C** for the 1000 largest stocks in the TAQ database for the 2-year period 1994–1995 [7]. The largest eigenvalue for the 2-year period (inset) is approximately 30 times larger than the maximum eigenvalue $\lambda_m = 1.94$ predicted for uncorrelated time series. The inset also shows the largest eigenvalue for the cross-correlation matrix for 4 half-year periods — denoted A, B, C, D. The arrow in the inset corresponds to the largest eigenvalue for the entire 2-year period, $\lambda_{1000} \approx 50$. The largest eigenvalue describes correlations within the entire market [2] as all companies contribute to it with equal weight. (b) Comparison of the RMT predictions for the spacing distributions with results for empirical cross-correlation matrix, the solid line is the GOE prediction. At the 80% confidence level, the Kolmogorov–Smirnov test cannot reject the hypothesis that the GOE is the correct description.

Random Matrix Theory (RMT) identify system-specific, non-random properties of the system under consideration, providing clues about the nature of the underlying interactions [5].

We analyze here the cross-correlation matrix $\mathsf{C} \equiv C_{ij} \equiv \langle G_i G_j \rangle - \langle G_i \rangle \langle G_j \rangle / \sigma_i \sigma_j$ of the returns at 30-minute intervals of the largest 1000 US stocks for the 2-year period 1994–1995. First, we diagonalize **C** and obtain its eigenvalues λ_k — with $k = 1, \ldots, 1000$ — which we rank-order from the smallest to the largest. Next, we calculate the eigenvalue distribution [2] and compare it with recent analytical results for a cross-correlation matrix generated from finite uncorrelated time series [6]. Figure 1 shows the eigenvalue distribution of **C**, which deviates from the predictions of Sengupta and Mitra [6] for large eigenvalues $\lambda_k \geq 1.94$ (see caption of Fig. 1). This result is in agreement with the results of Laloux *et al.* [2] for the eigenvalue distribution of **C** on a daily time scale.

2. Universal Properties: Eigenvalue Spacings

To test for universal properties, we first calculate the distribution of the nearest-neighbor spacings $s \equiv \lambda_{k+1} - \lambda_k$. The nearest-neighbor spacing is computed after transforming the eigenvalues in such a way that their distribution becomes uniform — a procedure known as unfolding [5]. Figure 1(b) shows the distribution of nearest-neighbor spacings for the empirical data, and compares it with the RMT predictions for real symmetric random matrices. This class of matrices shares universal properties with the ensemble of matrices whose elements are distributed according to a Gaussian probability measure — the Gaussian orthogonal ensemble (GOE). We find good agreement between the empirical data and the GOE prediction, $P_{\mathrm{GOE}}(s) = (\pi s/2)\exp(-\pi s^2/4)$.

3. Nonuniversal Properties: Eigenvector Statistics

Having demonstrated that the eigenvalue statistics of $\mathbf{C}$ satisfies the RMT predictions, we proceed to analyze the eigenvectors of $\mathbf{C}$. The component ℓ of a given eigenvector relates to the contribution of company ℓ to that eigenvector. Hence, the distribution of the components contains information about the number of companies contributing to a specific eigenvector. In order to distinguish between one eigenvector with approximately equal components and another with a small number of large components we define the inverse participation ratio $I_k \equiv \sum_{\ell=1}^{1000} [u_{k\ell}]^4$, where $u_{k\ell}$, $\ell = 1, \ldots, 1000$ are the components of eigenvector k. The physical meaning of I_k can be illustrated by two limiting cases: (i) a vector with identical components $u_{k\ell} \equiv 1/\sqrt{N}$ has $I_k = 1/N$, whereas (ii) a vector with one component $u_{k1} = 1$ and all the others zero has $I_k = 1$. Therefore, I_k is related to the reciprocal of the number of vector components significantly different from zero [5].

Figure 2 shows I_k for eigenvectors of a matrix generated from uncorrelated time series with a power law distribution of price changes [8]. The average value of I_k is $\langle I\rangle \approx 3\times 10^{-3} \approx 1/N$ indicating that the vectors are *extended* [5] — i.e., almost all companies contribute to them. Fluctuations around this average value are confined to a narrow range. On the other hand, the empirical data show deviations of I_k from $\langle I\rangle$ for a few of the largest eigenvalues. These I_k values are approximately 4–5 times larger than $\langle I\rangle$ which suggests that there are groups of approximately

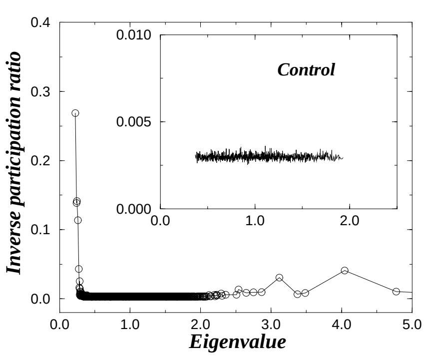


Fig. 2. Inverse participation ratio I_k for each of the 1000 eigenvectors. As a control, we show in the inset the values for the eigenvectors of a cross-correlation matrix computed from uncorrelated independent power-law distributed time series [8] of the same length as the data. Empirical data show marked peaks at both edges of the spectrum, whereas the control shows only small fluctuations around the average value $\langle I\rangle = 3 \times 10^{-3}$.

50 companies contributing to these eigenvectors. The corresponding eigenvalues are well outside the bulk, suggesting that these companies are correlated.

Surprisingly, we also find that there are I_k values as large as 0.35 for vectors corresponding to the smallest eigenvalues $\lambda_i \approx 0.25$.[a] The deviations from $\langle I \rangle$ for the smallest eigenvalues are about 10^2 to 10^3 times larger than the standard deviation of the fluctuations for the control, which suggests that the vectors are *localized* [5] — i.e., only a few companies contribute to them. The small size of the corresponding eigenvalues suggests that these companies are uncorrelated with one another. The appearance of localized states is a phenomenon well-known in RMT and suggests that $\mathbf{C}$ may be a random band matrix.[b]

4. Conclusion

In summary, we find that the most eigenvalues in the spectrum of the cross-correlation matrix of stock price changes agree surprisingly well with the *universal* predictions of random matrix theory. In particular, we find that $\mathbf{C}$ satisfies the universal properties of the Gaussian orthogonal ensemble of real symmetric random matrices. We find through the analysis of the inverse participation ratio of its eigenvectors that $\mathbf{C}$ may be a random band matrix, which may support the idea that a metric can be defined on the space of companies and that a distance can be defined between pairs of companies.[c] Hypothetically, the presence of localized states may allow us to draw conclusions about the "spatial dimension" of the set of stocks studied here and about the "range" of the correlations between the companies. On the practical side, our findings imply that the random part of the cross-correlation matrix must be removed before it can be used for designing financial instruments.

Acknowledgments

We thank M. Barthélémy, N. V. Dohkolyan, X. Gabaix, U. Gerland, S. Havlin, R. N. Mantegna, Y. Lee, C.-K.-Peng, D. Stauffer, and H. A. Weidenmüller for helpful discussions. LANA thanks FCT/Portugal for financial support. The Center for Polymer Studies is supported by NSF.

References

[1] E. J. Elton and M. J. Gruber, *Modern Portfolio Theory and Investment Analysis* J. Wiley, New York (1995).

[a] The large values of I_k for small eigenvalues indicate that the distribution of eigenvector components for the eigenvalues at the lower edge of the spectrum deviate from Gaussian prediction.
[b] A random band matrix B has elements B_{ij} independently drawn from different probability distributions. These distributions are often taken to be Gaussian and to be parameterized by their variance, which depends on i and j. Although such matrices are random, they still contain probabilistic information regarding the fact that a metric can be defined on their set of indices i.
[c] A related idea for a hierarchical structure of the financial cross-correlation matrix was recently put forward in Ref. [9].

[2] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, Phys. Rev. Lett. **83** (1999) 1467.
[3] V. Plerou, P.Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, Phys. Rev. Lett. **83** (1999) 1471.
[4] E. P. Wigner, Ann. Math. **53** (1951) 36.
[5] T. Guhr, A. Müller-Groeling and H. A. Weidenmüller, Phys. Rep. **299** (1998) 190.
[6] A. M. Sengupta and P. P. Mitra, Phys. Rev. **E60** (1999) 3389.
[7] *The trades and quotes* (TAQ) database. This database is published by the New York Stock Exchange and comprises 24 CD-ROMS for the period, 1994–1995.
[8] T. Lux, Appl. Fin. Econ. **6** (1996) 463; V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer and H. E. Stanley, Phys. Rev. **E60** (1999) 6519; U. A. Müller, M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz and C. Morgenegg, J. Bank. Fin. **14** (1990) 1189.
[9] R. N. Mantegna, Europ. Phys. J. **B11** (1999) 193.