arXiv:cond-mat/9902283 v1 20 Feb 1999

Universal and non-universal properties of cross-correlations in financial time series

Vasiliki Plerou[1,2], Parameswaran Gopikrishnan[1], Bernd Rosenow[3],

Luís A. Nunes Amaral[1], and H. Eugene Stanley[1]

[1] *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215*

[2] *Department of Physics, Boston College, Chestnut Hill, MA 02167*

[3] *Institut für Theoretische Physik, Universität zu Köln, D–50937 Köln, Germany*

(Last modified: Feb 14, 1999; Printed: July 10, 2004)

Abstract

We use methods of random matrix theory to analyze the cross-correlation matrix C of price changes of the largest 1000 US stocks for the 2-year period 1994-95. We find that the statistics of most of the eigenvalues in the spectrum of C agree with the predictions of random matrix theory, but there are deviations for a few of the largest eigenvalues. We find that C has the universal properties of the Gaussian orthogonal ensemble of random matrices. Furthermore, we analyze the eigenvectors of C through their 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.

PACS numbers: 87.23.Ge, 02.50.Ey, 05.40.-a

Typeset using REVTEX

There has been much recent work applying physics concepts and methods to the study of financial time series [1–14]. In particular, the study of correlations between price changes of different stocks is both of scientific interest and of practical relevance in quantifying the risk of a given stock portfolio [1,2]. Consider, for example, 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. Moreover, unlike most physical systems, there is no "algorithm" to calculate the "interaction strength" between two companies (as there is for, say, two spins in a magnet). The problem is that although every pair of companies should interact either directly or indirectly, the precise nature of interaction is unknown.

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. Large amounts of spectroscopic data on the energy levels were becoming available but were too complex to be explained by model calculations because the exact nature of the interactions were unknown. Random matrix theory (RMT) was developed in this context, to deal with the statistics of energy levels of complex quantum systems [15,16]. 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 [15]. RMT predictions represent an average over all possible interactions [16]. Deviations from the *universal* predictions of RMT identify system-specific, non-random properties of the system under consideration, providing clues about the nature of the underlying interactions [17,18].

In this letter, we apply RMT methods to study the cross-correlations [10] of stock price changes. First, we demonstrate the validity of the universal predictions of RMT for the eigenvalue statistics of the cross-correlation matrix. Second, we calculate the deviations of the empirical data from the RMT predictions, obtaining information that enables us to identify cross-correlations between stocks not explainable purely by randomness.

We analyze a data base [20] containing the price $S_i(t)$ of stock i at time t, where $i = 1, \ldots, 1000$ denotes the largest 1000 publicly-traded companies and the time t runs over the 2-year period 1994-95. From this time series, we calculate the price change $G_i(t, \Delta t)$, defined as

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

where $\Delta t = 30$ min is the sampling time scale. The simplest measure of correlations between different stocks is the equal-time cross-correlation matrix C which has elements

$$C_{ij} \equiv \frac{\langle G_i G_j \rangle - \langle G_i \rangle \langle G_j \rangle}{\sigma_i \sigma_j}\,, \tag{2}$$

where $\sigma_i \equiv \sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}$ is the standard deviation of the price changes of company i, and $\langle \cdots \rangle$ denotes a time average over the period studied [20].

We analyze the statistical properties of C by applying RMT techniques. First, we diagonalize C and obtain its eigenvalues λ_k —with $k = 1, \cdots, 1000$—which we rank-order from the smallest to the largest. Next, we calculate the eigenvalue distribution [10] and compare it with recent analytical results for a cross-correlation matrix generated from finite uncorrelated time series [21]. Figure 1 shows the eigenvalue distribution of C, which deviates from the predictions of Ref. [21], for large eigenvalues $\lambda_k \geq 1.94$ (see caption of Fig. 1). This result is in agreement with the results of Ref. [10] for the eigenvalue distribution of C on a daily time scale.

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 [17–19]. Figure 2(a) 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,

$$P_{\rm GOE}(s) = \frac{\pi s}{2} \exp\left(-\frac{\pi}{4} s^2\right) . \tag{3}$$

A second independent test of the GOE is the distribution of *next*-nearest-neighbor spacings between the rank-ordered eigenvalues [17]. This distribution is expected to be identical to the distribution of nearest-neighbor spacings of the Gaussian symplectic ensemble (GSE) as verified by the empirical data [Fig. 2(b)].

The distribution of eigenvalue spacings reflects correlations only of consecutive eigenvalues but does not contain information about correlations of longer range. To probe any "long-range" correlations, we first calculate the number variance Σ^2 which is defined as the variance of the number of unfolded eigenvalues in intervals of length L around each of the eigenvalues [17–19,22]. If the eigenvalues are uncorrelated, $\Sigma^2 \sim L$. For the opposite case of a "rigid" eigenvalue spectrum, Σ^2 is a constant. For the GOE case, we find the "intermediate" behavior $\Sigma^2 \sim \ln L$, as predicted by RMT [Fig. 2(c)].

A second way to measure "long-range" correlations in the eigenvalues is through the spectral rigidity Δ, defined to be the least square deviation of the unfolded cumulative eigenvalue density from a fit to a straight line in an interval of length L [17–19,23]. For uncorrelated eigenvalues, $\Delta \sim L$, whereas for the rigid case Δ is a constant. For the GOE case we find $\Delta \sim \ln L$ as predicted by RMT [Fig. 2(d)].

Having demonstrated that the eigenvalue statistics of C satisfies the RMT predictions, we now proceed to analyze the eigenvectors of C. RMT predicts that the components of the normalized eigenvectors of a GOE matrix are distributed according to a Gaussian probability distribution with mean zero and variance one. In agreement with recent results [10], we find that eigenvectors corresponding to *most* eigenvalues in the "bulk" ($\lambda_k \leq 2$) follow this prediction. On the other hand, eigenvectors with eigenvalues outside the bulk ($\lambda_k \geq 2$) show marked deviations from the Gaussian distribution. In particular, the vector corresponding to the largest eigenvalue λ_{1000} deviates significantly from the Gaussian distribution predicted by RMT.

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 [17,24]

$$I_k \equiv \sum_{\ell=1}^{1000} [u_{k\ell}]^4 \ , \tag{4}$$

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.

Figure 3 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* [24,25]—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 50 companies contributing to these eigenvectors. The corresponding eigenvalues are well outside the bulk, suggesting that these companies are correlated [18].

Surprisingly, we also find that there are I_k values as large as 0.35 for vectors corresponding to the smallest eigenvalues $\lambda_i \approx 0.25$ [26]. These deviations from the average are two orders of magnitude larger than $\langle I \rangle$, which suggests that the vectors are *localized* [24,25]—i.e., only a few companies contribute to them. The small values of the corresponding eigenvalues suggests that these companies are uncorrelated with each other.

The presence of vectors with large I_k also arises in the theory of Anderson localization [27]. In the context of localization theory, one frequently finds "random band matrices" [24] containing extended states with small I_k in the middle of the band, whereas edge states are localized and have large I_k. Our finding of localized states for small and large eigenvalues

of the cross-correlation matrix C is reminiscent of Anderson localization and suggests that C may be a random band matrix [28]

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 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 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 [29]. 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.

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

REFERENCES

[1] I. Kondor and J. Kertesz (eds) *Econophysics: An Emerging Science* (Kluwer, Dordrecht, 1999); J.-P. Bouchaud and M. Potters, *Theorie des Risques Financiéres* (Alea-Saclay, Eyrolles, 1998); B. B. Mandelbrot, *Fractals and Scaling in Finance* (Springer, New York, 1997).

[2] G. W. Kim and H. M. Markowitz, J. Portfolio Management **16**, 45 (1989); E. J. Elton and M. J. Gruber, *Modern Portfolio Theory and Investment Analysis* (J. Wiley, New York, 1995); P. Bak, M. Paczuski and M. Shubik, Physica A **246**, 430 (1997); R. Chatagny and B. Chopard, *International Conference on High Performance Computing and Networks*, Vienna (1997); R. G. Palmer *et. al.*, Physica D **75**, 264 (1994); K. Steiglitz, M.L. Honig, L.M. Cohen in *Market-Based Control: A Paradigm for Distributed Resource Allocation*, S. Clearwater (ed) (World Scientific, Hong Kong, 1996).

[3] T. Lux and M. Marchesi, Nature **297**, 498 (1999); T. Lux, J. Econ. Behav. Organizat. **33**, 143 (1998); J. Econ. Dyn. Control **22**, 1 (1997); Appl. Econ. Lett. **3**, 701 (1996).

[4] J.-P. Bouchaud and R. Cont, Eur. Phys. J. B **6**, 543 (1998); M. Potters, R. Cont, and J.-P. Bouchaud, Europhys. Lett. **41**, 239 (1998); J.-P. Bouchaud and D. Sornette, J. Phys. I (France) **4**, 863 (1994).

[5] M. Marsili and Y.-C. Zhang, Phys. Rev. Lett. **80**, 2741 (1998); G. Caldarelli, M. Marsili, and Y.-C. Zhang, Europhys. Lett. **40**, 479 (1997); S. Galluccio *et al*, Physica A **245**, 423 (1997).

[6] H. Takayasu, A. H. Sato, and M. Takayasu, Phys. Rev. Lett. **79**, 966 (1997); A. H. Sato and H. Takayasu, Physica A **250**, 231 (1998); H. Takayasu *et al*, Physica A **184**, 127 (1992).

[7] D. Chowdhury and D. Stauffer, Eur. Phys. J. B (in press); I. Chang and D. Stauffer, Physica A **264**, 1 (1999); D. Stauffer and T. J. P. Penna, Physica A **256**, 284 (1998);

D. Stauffer, P. M. C. de Oliveria and A. T. Bernardes, Int. J. Theor. Appl. Finance (in press).

[8] R. N. Mantegna and H. E. Stanley, Nature **376**, 46 (1995); Y. Liu *et. al.* Physica A **245**, 437 (1997); P. Cizeau *et. al.* Physica A **245**, 441 (1997); P. Gopikrishnan *et. al.* Eur. Phys. J. B **3**, 139 (1998).

[9] A. Johansen and D. Sornette, Int. J. Mod. Phys. C **10** (1999); D. Sornette, A. Johansen, and J.-P. Bouchaud, J. Phys. I (France) **6**, 167 (1996); A. Arnoedo, J.-F. Muzy and D. Sornette, Eur. Phys. J. B **2**, 277 (1998).

[10] L. Laloux *et. al.*, Risk Magazine, to appear (cond-mat/9810255); S. Galluccio *et. al.* Physica A **259**, 449 (1998).

[11] R. N. Mantegna, cond-mat/9802256.

[12] O. Biham *et al*, Phys. Rev. B **58**, 1352 (1998); M. Levy, H. Levy, and S. Solomon, Economics Letters **45**, 103 (1994).

[13] S. Ghashghaie *et. al.* Nature **381**, 767 (1996); R. N. Mantegna and H. E. Stanley, Nature **383**, 587 (1996); Physica A **239**, 255 (1997); A. Arnoedo *et. al.*, cond-mat/9607120.

[14] N. Vandewalle and M. Ausloos, Int. J. Mod. Phys. C **9**, 711 (1998); Eur. Phys. J. B **4**, 257 (1998); N. Vandewalle *et al*, Physica A **255**, 201 (1998).

[15] E.P. Wigner, Ann. Math. **53**, 36 (1951); E.P. Wigner, in *Conference on Neutron Physics by Time–of–flight* (Gatlinburg, Tennessee, 1956).

[16] F. J. Dyson, J. Math. Phys. **3**, 140 (1962); F. J. Dyson and M. L. Mehta, J. Math. Phys. **4**, 701 (1963); M. L. Mehta and F. J. Dyson, J. Math. Phys. **4**, 713 (1963).

[17] T. Guhr, A. Müller–Groeling, and H. A. Weidenmüller, Phys. Rep. **299**, 190 (1998); M. L. Mehta, *Random Matrices* (Academic Press, Boston, 1991).

[18] T. A. Brody *et. al.*, Rev. Mod. Phys **53**, 385 (1981).

[19] H. Bruus and J.-C Anglés d'Auriac, Europhys. Lett. **35**, 321 (1996).

[20] *The trades and quotes* (TAQ) database. This database is published by the New York Stock Exchange and consists of 24 CD-ROMS for the period 1994-95.

[21] A. M. Sengupta and P. P. Mitra, cond-mat/9709283.

[22] The number variance is defined as $\Sigma^2(L) \equiv \langle [N(\lambda + \frac{L}{2}) - N(\lambda - \frac{L}{2}) - L]^2 \rangle_\lambda$, where $N(\lambda) \equiv \sum_i \theta(\lambda - \lambda_i)$ is the integrated density of eigenvalues and $\langle \ldots \rangle_\lambda$ denotes an average over λ [17,19].

[23] The spectral rigidity is defined as $\Delta(L) \equiv \frac{1}{L} \left\langle \min_{A,B} \int_{\lambda-L/2}^{\lambda+L/2} (N(\lambda_1) - A\lambda_1 - B)^2 d\lambda_1 \right\rangle_\lambda$, where $\langle \ldots \rangle_\lambda$ denotes an average over λ and $N(\lambda) \equiv \sum_i \theta(\lambda - \lambda_i)$ is the integrated density of eigenvalues [17,19].

[24] Y. V. Fyodorov and A. D. Mirlin, Phys. Rev. Lett. **69**, 1093 (1992); **71**, 412 (1993); Int. J. Mod. Phys. B **8**, 3795 (1994); A. D. Mirlin and Y. V. Fyodorov, J. Phys. A: Math. Gen. **26**, L551 (1993); E. P. Wigner, Ann. Math. **62**, 548 (1955).

[25] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys **57**, 287 (1985).

[26] 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.

[27] Metals or semiconductors with impurities can be described by Hamiltonians with random-hopping integrals [F. Wegner and R. Oppermann, Z. Physik **B34**, 327 (1979)]. Electron-hopping between neighboring sites is more probable than hopping over large distances, leading to a Hamiltonian that is a random band matrix.

[28] 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.

[29] A related idea for a hierarchical structure of the financial cross-correlation matrix was recently put forward in Ref. [11].

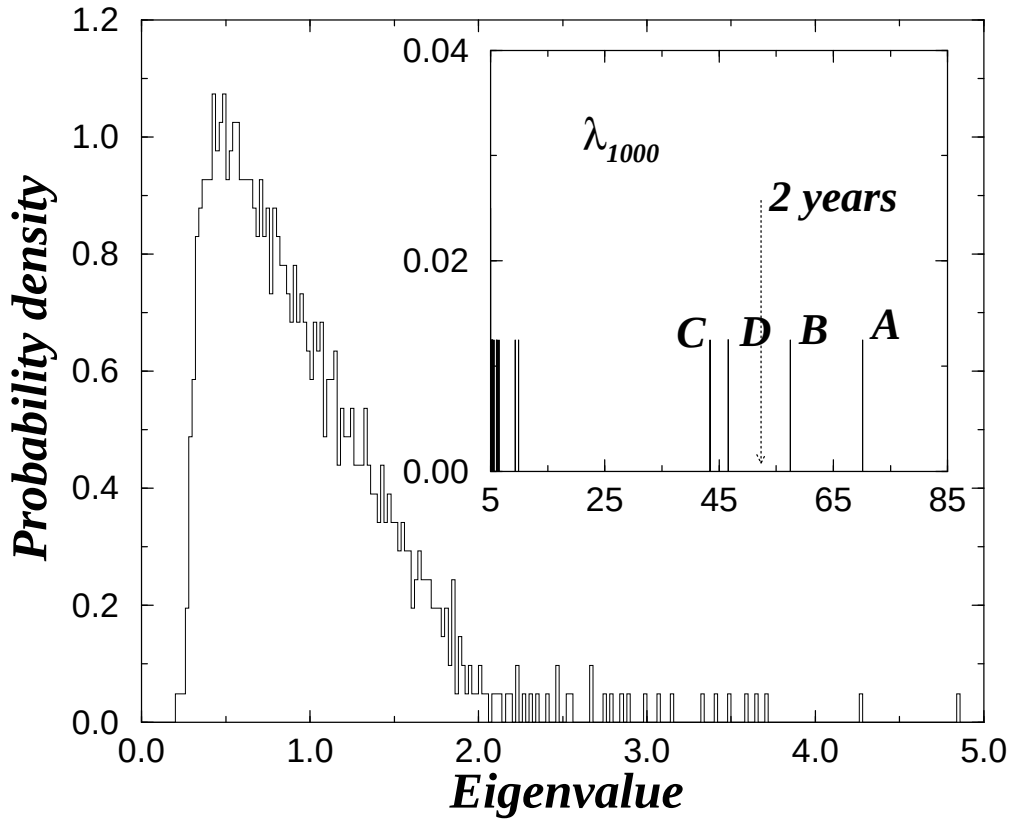
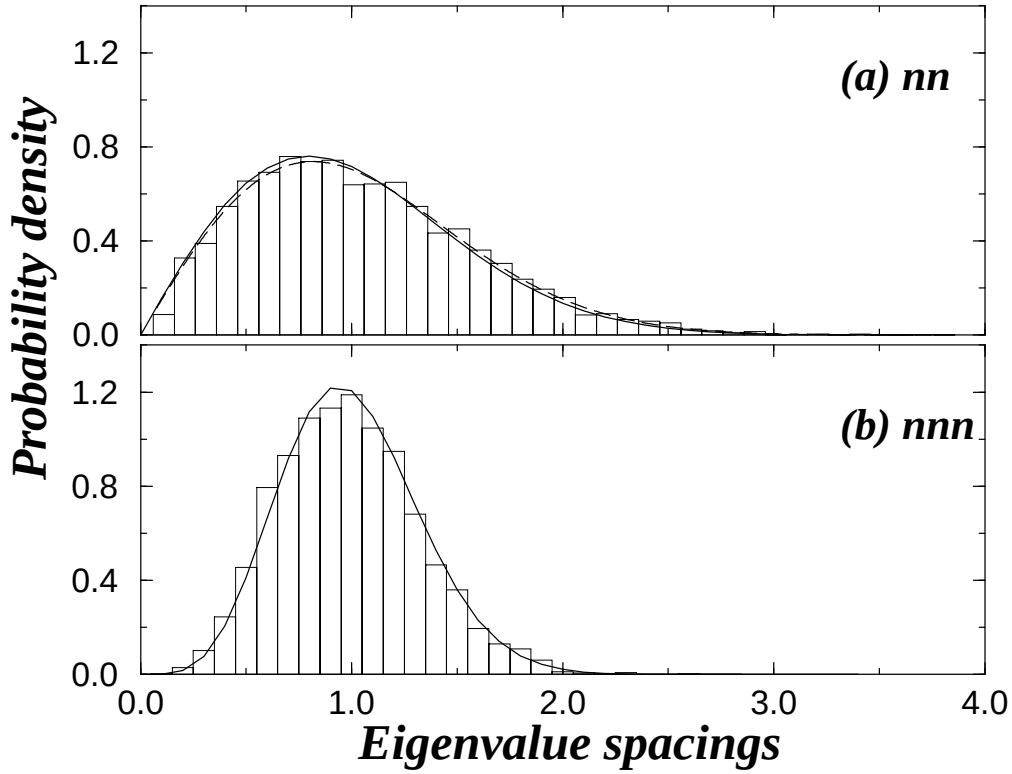


FIG. 1. The probability density of the eigenvalues of the normalized cross-correlation matrix C for the 1000 largest stocks in the TAQ database for the 2-year period 1994-95 [20]. Recent analytical results [21] for cross-correlation matrices generated from uncorrelated time series predict a finite range of eigenvalues depending on the ratio R of the length of the time series to the dimension of the matrix [10]. In our case $R = 6.448$ corresponding to eigenvalues distributed in the interval $0.37 \leq \lambda_k \leq 1.94$ [21]. However, the largest eigenvalue for the 2-year period (inset) is approximately 30 times larger than the maximum eigenvalue 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 distribution of eigenvector components for the large eigenvalues, well outside the bulk show significant deviations from the Gaussian prediction of RMT, which suggests "collective" behavior or correlations [18] between different companies. The largest eigenvalue would then correspond to the correlations within the entire market [10].


(a) nn
(b) nnn
Probability density
Eigenvalue spacings
0.0
0.4
0.8
1.2
1.0
2.0
3.0
4.0

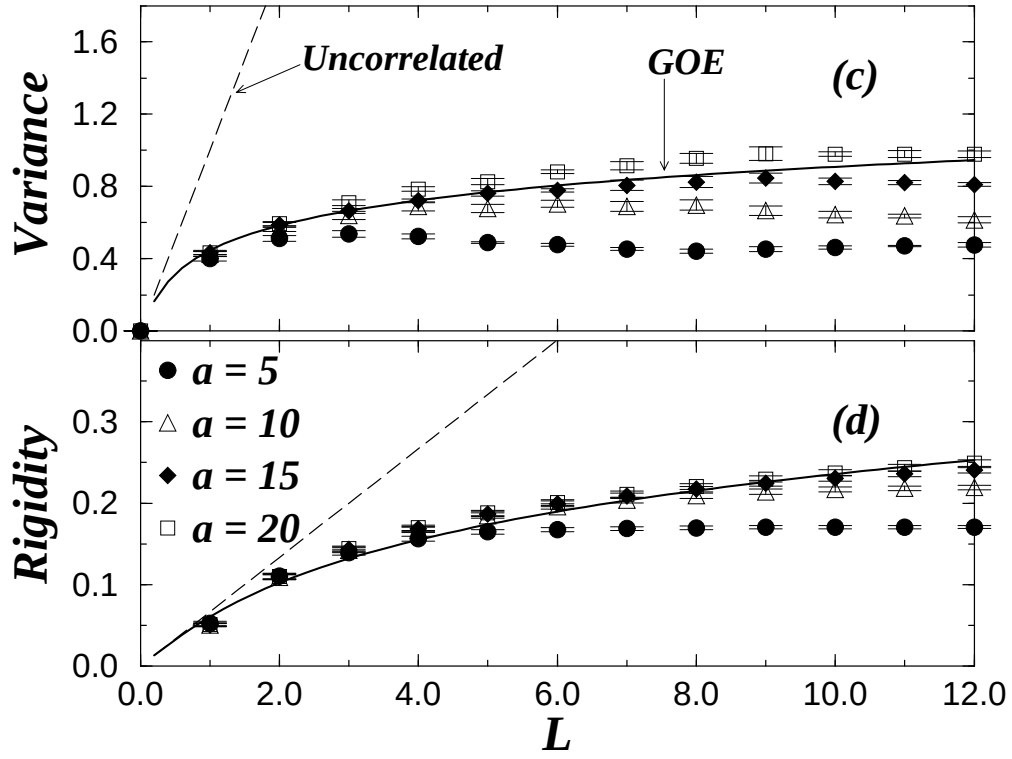

Uncorrelated
GOE
(c)
(d)
a = 5
a = 10
a = 15
a = 20
Variance
Rigidity
L
0.0
0.4
0.8
1.2
1.6
0.1
0.2
0.3
2.0
4.0
6.0
8.0
10.0
12.0

FIG. 2. Comparison of the RMT predictions for the spacing distributions with results for empirical cross-correlation matrix . (a) Nearest-neighbor (nn) spacing distribution of the eigenvalues of C after unfolding. We use the Gaussian broadening procedure [19]. The eigenvalue distribution can be considered as a sum of delta functions about each eigenvalue, λ_k, each of which is then "broadened" by choosing a Gaussian distribution with standard deviation $(\lambda_{k+a} - \lambda_{k-a})/2$, where $2a$ is the size of the window used for broadening [19]. Here, $a = 15$, the optimum value obtained from Fig. 2(d). The solid line is the GOE prediction, Eq. (3), and the dashed line is a fit to the one parameter Brody distribution $p(s) \equiv B\,(1+\beta)\,s^{\beta}\,\exp(-Bs^{\beta+1})$, with $B \equiv [\Gamma(\frac{\beta+2}{\beta+1})]^{1+\beta}$. The fit yields $\beta = 0.99 \pm 0.02$, in good agreement with the GOE prediction $\beta = 1$. A Kolmogorov-Smirnov test suggests that the GOE is 10^5 times more likely to be the correct description than the Gaussian unitary ensemble, and 10^{20} times more likely than the GSE. Furthermore, at the 80% confidence level, the Kolmogorov-Smirnov test cannot reject the hypothesis that the GOE is the correct description. (b) Next-nearest-neighbor (nnn) spacing distribution of C. RMT predicts that, for the GOE, the distribution of next-nearest-neighbor spacing should follow the same distribution as the nearest-neighbor spacing for the GSE. This prediction is confirmed for the empirical data both visually and by a Kolmogorov-Smirnov test that at the 40% confidence level cannot reject the hypothesis that the GSE is the correct distribution. (c) Number variance and (d) spectral rigidity of C for different values of the unfolding parameter a, as compared to the exact expression for the GOE (solid line) and the uncorrelated case (dashed line) . As a increases, both the number variance and the spectral rigidity approach the theoretical curve for the GOE while the spacing distribution remains essentially unchanged. We choose $a = 15$ as the optimal-value.

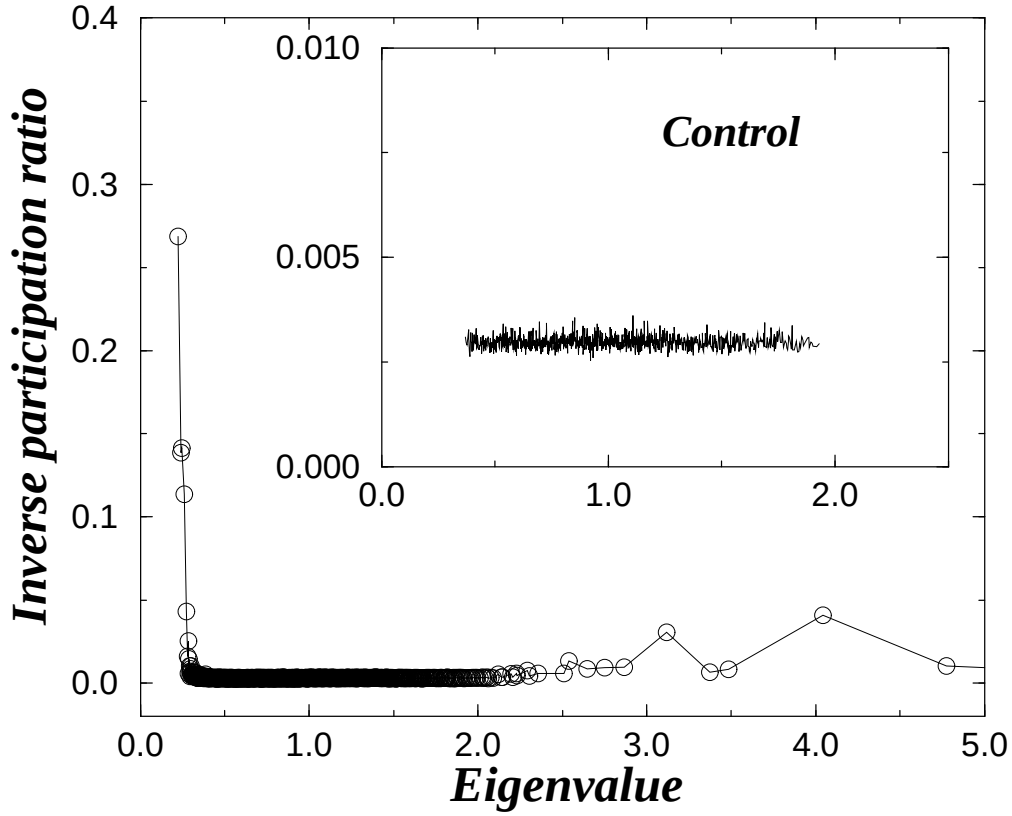


FIG. 3. Inverse participation ratio I_k for each of the 1000 eigenvectors. As a control, we show in the inset the I_k values for the eigenvectors of a cross-correlation matrix computed from uncorrelated independent power-law distributed time series [8] of the same length as the data. Empirical data show marked peaks at both edges of the spectrum, whereas the control shows only small fluctuations around the average value $\langle I \rangle = 3 \times 10^{-3}$. The large I_k values for the largest eigenvalues are to be expected from Fig. 1, but the large values of I_k for the small eigenvalues are surprising. Large I_k values at the edges of the eigenvalue spectrum is a situation often found in localization theory.