PHYSICAL REVIEW E, VOLUME 64, 035106(R)

Quantifying and interpreting collective behavior in financial markets

Parameswaran Gopikrishnan,[1] Bernd Rosenow,[1,2] Vasiliki Plerou,[1,3] and H. Eugene Stanley[1]
[1]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215*
[2]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138*
[3]*Department of Physics, Boston College, Chestnut Hill, Massachusetts 02167*

(Received 22 August 2000; revised manuscript received 2 February 2001; published 30 August 2001)

Firms having similar business activities are correlated. We analyze two different cross-correlation matrices $\mathbf{C}$ constructed from (i) 30-min price fluctuations of 1000 US stocks for the two-year period 1994–95 and (ii) one-day price fluctuations of 422 US stocks for the 35-year period 1962–96. We find that the eigenvectors of $\mathbf{C}$ corresponding to the largest eigenvalues allow us to partition the set of all stocks into distinct subsets. These subsets are similar to business sectors, and are stable for extended periods of time. We find that price fluctuations of these subsets are characterized by power-law decaying time correlations, reminiscent of strongly interacting systems.

DOI: 10.1103/PhysRevE.64.035106 PACS number(s): 89.90.+n, 05.45.Tp, 05.40.Fb

The internal structure of a complex system manifests itself in correlations among its constituents. In complex physical systems, interactions between constituents cause "collective modes" having special statistical properties which reflect the underlying dynamics. Can we quantify collective movement of stock prices in analogous terms?

To address this question, we analyze the equal-time correlation matrix $\mathbf{C}$ constructed from the price fluctuations of a large number of stocks. First, we find that the "collective modes" for the stock market problem partition the set of all stocks studied, into distinct subsets. Typically, these subsets are formed by combinations of related industries, and in some cases, they go beyond grouping by industry. Due to company diversification, the traditional partitioning of firms into subsets by products and services is difficult and sometimes arbitrary, and thus our results could be viewed as a "statistical alternative to traditional industry classification" [1]. Furthermore, we find that the price fluctuations of the collective modes display long-range power-law time correlations, in sharp contrast to individual stocks [2]. Collective modes in physical systems display time correlations which persist on much larger time scales than any individual unit. Motivated by this analogy, we start with an interacting-stocks framework and outline one possible mechanism that could prove useful in understanding the distinct statistics of collective price fluctuations.

We first define the cross-correlation matrix $\mathbf{C}$ with elements $C_{ij} \equiv [\langle G_i G_j \rangle - \langle G_i \rangle \langle G_j \rangle]/\sigma_i \sigma_j$, where σ_i is the standard deviation of price fluctuations $G_i(t) \equiv \ln S_i(t+\Delta t) - \ln S_i(t)$ (returns), $S_i(t)$ denotes the price of stock $i = 1, \ldots, N$, and $\langle \ldots \rangle$ denotes a time average over the period studied. To investigate correlations on different time scales, we analyze (i) 30-min returns of $N=1000$ largest stocks for the two-year period 1994–95 and (ii) daily returns of $N=422$ stocks for the 35-year period 1962–96 [3].

Next, we diagonalize $\mathbf{C}$ and rank-order its eigenvalues λ_k such that $\lambda_{k+1} > \lambda_k$; the corresponding eigenvectors are denoted $\mathbf{u}^k$. We then analyze the components of those *deviating eigenvectors* whose eigenvalues are larger than the upper bound for uncorrelated time series [4,5]. A direct examination of these eigenvectors, however, does not yield a straightforward interpretation of their economic relevance. To interpret their meaning, we note that the largest eigenvalue is an order of magnitude larger than the others, which constrains the remaining $N-1$ eigenvalues since $\mathrm{Tr}\,\mathbf{C} = N$. Thus, in order to analyze the contents of the deviating eigenvectors, we first remove the effect of the largest eigenvalue [6].

To analyze the information contained in the eigenvectors $\mathbf{u}^k$, we partition the 1000 stocks into groups labeled $l = 1 \ldots, 75$ (comprising N_l stocks each) according to the first two digits of their Standard Industrial Classification (SIC) code, which classifies major industry groups. We define a projection matrix $\mathbf{P}$, with elements $P_{li} = 1/N_l$ if stock i belongs to group l and $P_{li} = 0$ otherwise. For each deviating eigenvector $\mathbf{u}^k$, we compute the contribution $X_l^k \equiv \sum_{i=1}^{N} P_{li} (u_i^k)^2$ of each industry group l [7]. The above procedure of computing X_l^k is analogous to the analysis of wave functions in disordered systems, where one calculates the probability of finding a particle in a given region.

Figure 1 shows X_l^k for the ten largest eigenvectors after excluding the influence of the largest eigenvalue. The contribution X_l^{999} shows several industries. We examine the significant contributors and find mainly stocks with large market capitalization (Fig. 2). We analyze X_l^k for the remainder of the deviating eigenvectors and find a significant 'peak' at distinct values of the SIC code, suggesting that these eigenvectors correspond to distinct industry groups [8].

One deviating eigenvector $\mathbf{u}^{995}$ displays large values of X_l^k for the heavy construction and telecommunication industries. An examination of these firms shows significant business activity in Latin America [9]. Another interesting case corresponds to eigenvectors $\mathbf{u}^{996}$ and $\mathbf{u}^{997}$, both of which contain a mixture of stocks of gold-mining firms and banking firms, which separate when we compute the symmetric and antisymmetric combinations $1/\sqrt{2}(\mathbf{u}^{996} \pm \mathbf{u}^{997})$. The other deviating eigenvectors display technology, metal mining, banking, petroleum refining, auto manufacturing, drug manufacturing, and paper manufacturing firms (Fig. 1).

We next focus on the interpretation of the largest eigenvalue λ_{1000}. Using the eigenvector $\mathbf{u}^{1000}$, we construct a time series $G^{(1000)}(t) \equiv \sum_{i=1}^{1000} u_i^{1000} G_i(t)$. We then compare

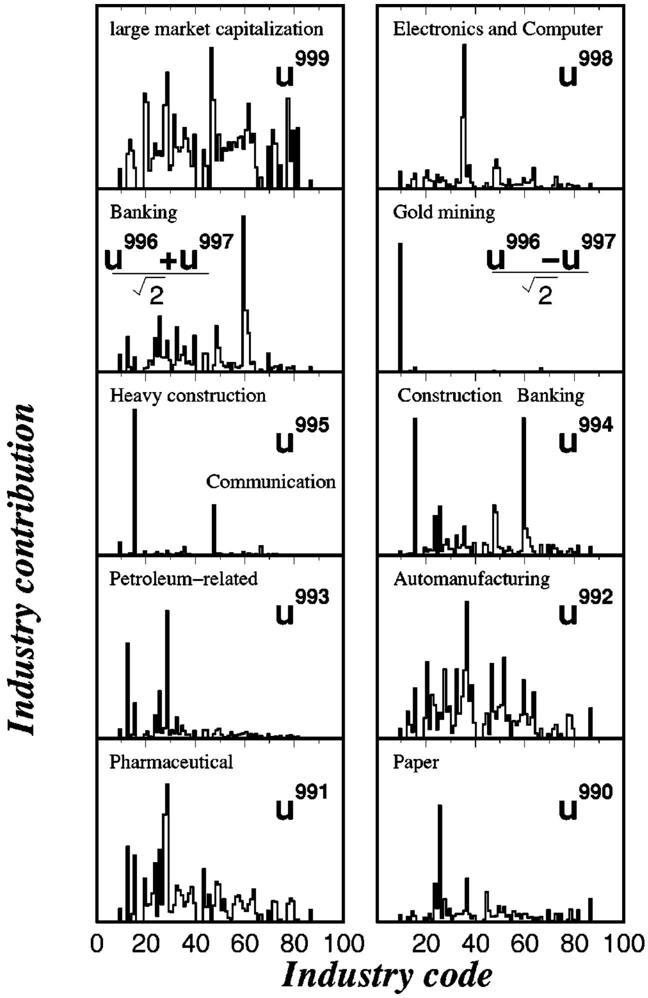

FIG. 1. Contribution X_l^k to industry sector l of eigenvector $\mathbf{u}^k$ for the deviating eigenvectors shows marked peaks at distinct values of SIC code, for all but $\mathbf{u}^{999}$ which contains stocks with large capitalizations as significant contributors.

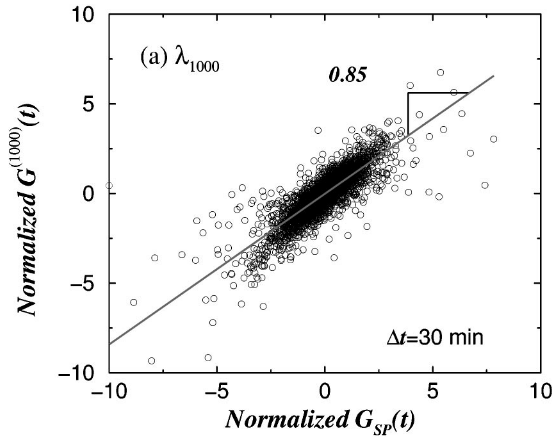

FIG. 3. S&P 500 returns $G_{SP}(t)$ regressed against the return $G^{(1000)}(t)$ of the portfolio defined by the eigenvector $\mathbf{u}^{1000}$. Both axes are scaled by their respective standard deviations. A linear regression yields a slope 0.85 ± 0.09, showing a large degree of correlation.

$G^{(1000)}(t)$ with the returns $G_{SP}(t)$ of the S&P 500 index, a benchmark for gauging the performance of the entire US stock market. Regressing $G^{(1000)}(t)$ against $G_{SP}(t)$ shows a scatter around a linear fit with slope 0.85 ± 0.09 (Fig. 3). Thus, we interpret the eigenvector $\mathbf{u}^{1000}$ as the influence of the entire market that is common for all stocks [4,5].

Next, we examine whether the eigenvectors $\mathbf{u}^k$ corresponding to business sectors remain stable in time [10]. Partitioning the year 1994 into two six-month periods, A and B, we calculate the corresponding eigenvectors $\mathbf{u}_A$ and $\mathbf{u}_B$ of the cross-correlation matrices and quantify the time stability by calculating the magnitude of the scalar products $O_{ij} \equiv |\mathbf{u}_A^i \mathbf{u}_B^j|$ for the 20 largest eigenvalues. Perfect time stability would mean $O_{ij} = \delta_{ij}$. For $i = 1000$, we find $O_{ii} = 0.93$, indicating almost perfect stability. We find that O_{ii} decreases as i decreases from 1000 (Fig. 4). Extending this analysis to

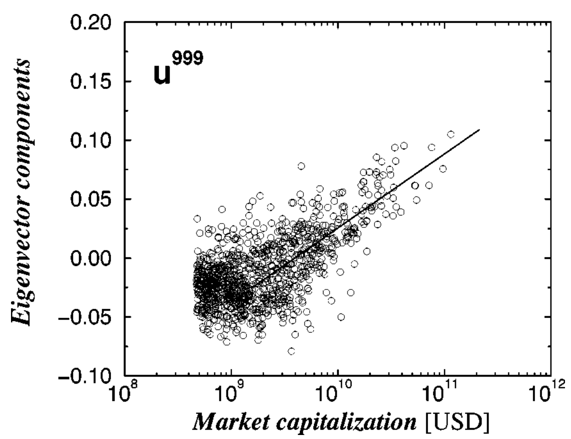

FIG. 2. All 10^3 eigenvector components of $\mathbf{u}^{999}$ plotted against market capitalization (in units of US Dollars) shows that large firms contribute more than small firms. The straight line, which shows a logarithmic fit, is a guide to the eye.

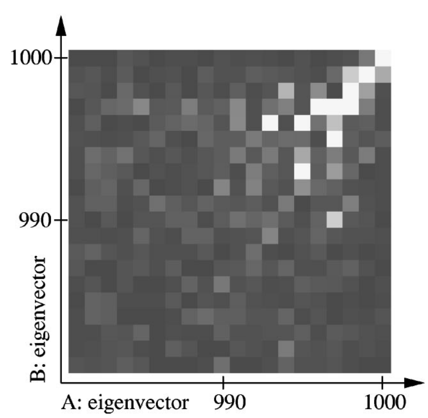

FIG. 4. Comparison of eigenvectors for different time periods A (first half of 1994) and B (second half of 1994) by means of their scalar product O_{ij}, represented on a greyscale, where zero (black) corresponds to no overlap, and white (one) to perfect overlap. Note that the eigenvectors corresponding to the four largest eigenvalues have a large degree of time stability.

daily returns using database (ii) shows that the eigenvectors corresponding to the largest three eigenvalues are stable for as many as ten years.

How can we understand correlations between stocks? In physical systems, one starts from the interactions between the constituents, and then relates interactions to correlated ''modes'' of the system. Here, we ask if an analogous mechanism involving ''interactions'' can give rise to the correlated behavior that we find. Interactions arise when two companies are doing business together, compete for the same market, or when they are perceived by investors to be linked.

One generic model for interacting physical systems is the soft spin model [11], which we apply to describe the dynamics of ''instantaneous'' returns $g_i(t) \equiv d/dt \ln S_i(t)$; we write a stochastic differential equation for $g_i(t)$,

$$\tau_o \partial_t g_i(t) = -r_i g_i(t) - \kappa g_i^3(t) + \sum_j J_{ij} g_j(t) + \frac{1}{\tau_o} \xi_i(t), \tag{1}$$

where $\xi_i(t)$ are Gaussian random variables with correlation function $\langle \xi_i(t) \xi_j(t') \rangle = \delta_{ij} \tau_o \delta(t-t')$, and τ_o sets the time scale of the problem. In the context of a soft spin model, the first two terms on the right-hand side of Eq. (1) arise from the derivative of a double-well potential, enforcing the soft-spin constraint. The interaction among soft spins is given by the couplings J_{ij}. In the absence of the cubic term, and without interactions, τ_o / r_i are relaxation times of the $\langle g_i(t) g_i(t+\tau) \rangle$ correlation function. A similar differential equation, without the couplings, was derived in a financial context to describe the dynamics of returns, using a quadratic instead of a cubic term [12].

As the coupling strengths increase, the soft-spin system undergoes a transition to an ordered state with permanent local magnetizations [11]. At the transition point, the spin dynamics are very ''slow'' as reflected in a power law decay of the spin autocorrelation function in time. To test whether this signature of strong interactions is present for the stock market problem, we analyze the autocorrelation functions $c^{(k)}(\tau) \equiv \langle G^{(k)}(t) G^{(k)}(t+\tau) \rangle$, where $G^{(k)}(t) \equiv \Sigma_{i=1}^{1000} u_i^k G_i(t)$ is the time series defined by eigenvector $\mathbf{u}^k$. Instead of analyzing $c^{(k)}(\tau)$ directly, we apply the detrended fluctuation analysis (DFA) method [13]. Figure 5 shows that the correlation functions $c^{(k)}(\tau)$ indeed decay as power laws for the deviating eigenvectors $\mathbf{u}^k$, which is in sharp contrast to the behavior of $c^{(k)}(\tau)$ for the rest of the eigenvectors and the autocorrelation functions of individual stocks, which show only short-range correlations. We interpret this as evidence for strong interactions.

In the absence of the cubic term, we obtain only exponentially decaying correlation functions for the ''modes'' corresponding to the large eigenvalues, which is inconsistent with our finding of power-law correlations.

In summary, given only the change in price of a stock, and no additional information about that stock, we can partition the set of all 10^3 stocks studied into subsets whose identities correspond well to conventionally identified sectors of economic activity. Motivated by the concept of critical slowing down in correlated physical systems, we analyze the time evolution of ''collective modes'' corresponding to these sectors, and find that they are characterized by power-law decaying correlation functions, which is consistent with the possibility that cross-correlations in the stock market arise not just from common influences such as relevant news-breaks (the common view), but also from interactions between stock price fluctuations.

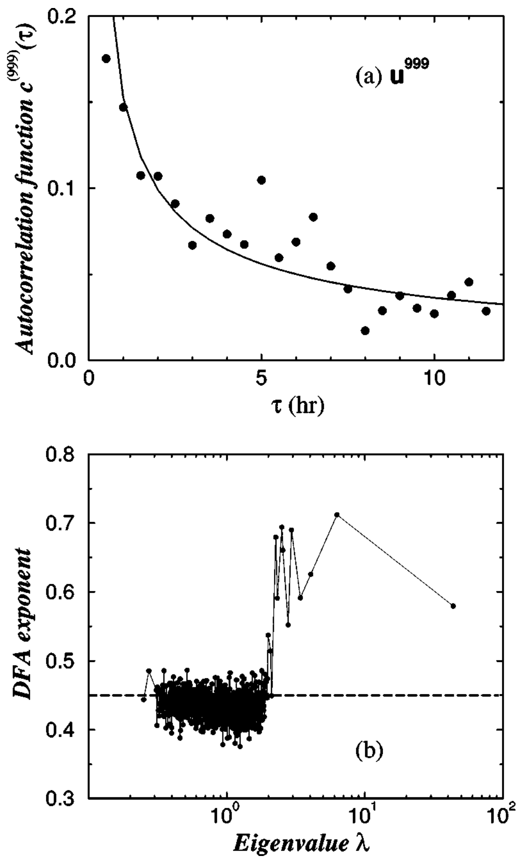

FIG. 5. (a) Autocorrelation function $c^{(k)}(\tau)$ of the time series defined by the eigenvector $\mathbf{u}^{999}$. The solid line shows a fit to a power-law functional form $\tau^{-\gamma_k}$, whereby we obtain values $\gamma_k = 0.61 \pm 0.06$ for $k=999$. (b) To quantify the exponents γ_k for all $k=1, \ldots, 1000$ eigenvectors, we use the method of DFA analysis [13] often used to obtain accurate estimates of power-law correlations. We plot the detrended fluctuation function $F(\tau)$ as a function of the time scale τ for each of the 1000 time series. Absence of long-range correlations would imply $F(\tau) \sim \tau^{0.5}$, whereas $F(\tau) \sim \tau^{\nu}$ with $0.5 < \nu \leqslant 1$ implies power-law decay of the correlation function with exponent $\gamma = 2 - 2\nu$. We plot the exponents ν as a function of the eigenvalue and find values exponents ν significantly larger than 0.5 for all the deviating eigenvectors. In contrast, for the remainder of the eigenvectors, we obtain the mean value $\nu = 0.44 \pm 0.04$, comparable to the value $\nu = 0.5$ for the uncorrelated case.

We thank P. Cizeau, E. Derman, T. Guhr, J. Hill, L. Viceira, and especially L. A. N. Amaral and X. Gabaix for stimulating discussions. B.R. is supported by Forschungsstipendium of the DFG. The Center for Polymer Studies is supported by the National Science Foundation and British Petroleum.

[1] E. J. Elton and M. J. Gruber, *Modern Portfolio Theory and Investment Analysis* (Wiley and Sons, New York, 1996), p. 445.

[2] For a recent review, see J. D. Farmer, Comput. Sci. Eng. **1**, 26 (1999).

[3] The 30-min data analyzed are from the Trades and Quotes database published by the New York Stock Exchange. The daily data are from the Center for Research in Securities Prices (CRSP) database. In both cases, only those companies that survive the entire period are considered in our analysis.

[4] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).

[5] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).

[6] The presence of a very large eigenvalue, by preservation of the trace $\mathrm{Tr}(\mathbf{C})=N$, tends to cause the other large eigenvalues and eigenvectors to be influenced by randomness. Therefore, in order to remove the effect of the "market mode" on other large eigenvalues, we implement the linear regression $G_i(t)=\alpha_i+\beta_i G^{(1000)}(t)+\epsilon_i(t)$, where $G^{(1000)}\equiv\sum_{i=1}^{1000}u_i^{1000}G_i(t)$. We then recompute the correlation matrix $\mathbf{C}$ using the residuals $\epsilon_i(t)$. Alternatively, removing the intradaily patterns in absolute values of $G_i(t)$ using the procedure of Y. Liu *et al.*, Phys. Rev. E **60**, 1390 (1999) also yields similar results.

[7] The choice of weights $(u_i^k)^2$ is because of the normalization $\sum_{i=1}^{N}(u_i^k)^2=1$.

[8] An interesting classification of stocks into sectors using the concept of ultrametric distances was obtained by R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).

[9] At first sight, there seems to be no obvious reason why this particular geographical region should play a significant role, so we examine news archives and find, interestingly, that there was a large currency devaluation in Mexico in the second half of 1994.

[10] Time evolution of the largest eigenvalues are analyzed in S. Drozdz *et al.*, Physica A **287**, 440 (2000).

[11] P. C. Hohenberg and B. I. Halperin, Rev. Mod. Phys. **49**, 435 (1977); K. H. Fisher and J. A. Hertz, *Spin Glasses* (Cambridge University Press, New York, 1991).

[12] J. D. Farmer, e-print adap-org/9812005; R. Cont and J.-P. Bouchaud, Eur. Phys. J. B **6**, 543 (1998).

[13] C. K. Peng, *et al.*, Phys. Rev. E **49**, 1685 (1994).