ELSEVIER

Physica A 287 (2000) 339–361

PHYSICA A

www.elsevier.com/locate/physa

Quantifying fluctuations in economic systems by adapting methods of statistical physics

H.E. Stanley*, P. Gopikrishnan, V. Plerou, L.A.N. Amaral

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

Received 3 September 2000

Abstract

The emerging subfield of econophysics explores the degree to which certain concepts and methods from statistical physics can be appropriately modified and adapted to provide new insights into questions that have been the focus of interest in the economics community. Here we give a brief overview of two examples of research topics that are receiving recent attention. A first topic is the characterization of the dynamics of stock price fluctuations. For example, we investigate the relation between trading activity – measured by the number of transactions $N_{\Delta t}$ – and the price change $G_{\Delta t}$ for a given stock, over a time interval $[t, t + \Delta t]$. We relate the time-dependent standard deviation of price fluctuations – volatility – to two microscopic quantities: the number of transactions $N_{\Delta t}$ in Δt and the variance $W^2_{\Delta t}$ of the price changes for all transactions in Δt. Our work indicates that while the pronounced tails in the distribution of price fluctuations arise from $W_{\Delta t}$, the long-range correlations found in $|G_{\Delta t}|$ are largely due to $N_{\Delta t}$. We also investigate the relation between price fluctuations and the number of shares $Q_{\Delta t}$ traded in Δt. We find that the distribution of $Q_{\Delta t}$ is consistent with a stable Lévy distribution, suggesting a Lévy scaling relationship between $Q_{\Delta t}$ and $N_{\Delta t}$, which would provide one explanation for volume-volatility co-movement. A second topic concerns cross-correlations between the price fluctuations of different stocks. We adapt a conceptual framework, random matrix theory (RMT), first used in physics to interpret statistical properties of nuclear energy spectra. RMT makes predictions for the statistical properties of matrices that are universal, that is, do not depend on the interactions between the elements comprising the system. In physics systems, deviations from the predictions of RMT provide clues regarding the mechanisms controlling the dynamics of a given system, so this framework can be of potential value if applied to economic systems. We discuss a systematic comparison between the statistics of the cross-correlation matrix $\mathbf{C}$ – whose elements C_{ij} are the correlation-coefficients between the returns of stock i and j – and that of a random matrix having the same symmetry properties. Our work suggests that RMT can be used to distinguish random and non-random parts of $\mathbf{C}$; the non-random part of $\mathbf{C}$, which deviates from RMT results provides information regarding genuine cross-correlations between stocks.

Keywords: Volatility; Econophysics; Random matrix theory; Cross-correlations

* Corresponding author.
E-mail address: hes@bu.edu (H.E. Stanley).

0378-4371/00/$ - see front matter
PII: S0378-4371(00)00473-8

1. Introduction

The interest of physicists in economic systems has roots that date back at least as far as 1936, when the Italian physicist Majorana wrote a paper on the analogy between statistical laws in physics and in the social sciences [1]. Majorana's intriguing point of view was initially considered of marginal interest since, until recently, not many professional physicists performed research associated with social or economic systems (for exceptions see, e.g., Refs. [2–13]).

Physics research activity in this field has become less episodic and a research community is beginning to emerge (see Refs. [14–18] for details). The hope of physicists is that their efforts could in time provide an approach complementary to the approaches in economics, particularly since a number of economists are working along parallel lines [5,13,15,19–23].

Our group's recent econophysics research focuses broadly on two lines of study. The first focus relates to the statistical characterization of the "microscopic" dynamics of stock returns [2,24–54]. The second focus relates to the study of cross-correlations between the returns of stocks [55–72].

2. Scaling and universality: two concepts of modern statistical physics

Statistical physics deals with systems comprising a very large number of interacting subunits, for which predicting the exact behavior of the individual subunit would be impossible. Hence, one is limited to making statistical predictions regarding the collective behavior of the subunits. Recently, it has come to be appreciated that many such systems which consist of a large number of interacting subunits obey universal laws that are independent of the microscopic details. The finding, in physical systems, of universal properties that do not depend on the specific form of the interactions gives rise to the intriguing hypothesis that universal laws or results may also be present in economic and social systems [2,14]. [1]

2.1. Background

Suppose we have a small bar magnet made up of, say, 10^{12} strongly interacting subunits called "spins". We know it is a magnet because it is capable of picking up thumbtacks, the number of which is called the order parameter M. As we heat this system, M decreases and eventually, at a certain critical temperature T_c, it reaches zero. In fact, the transition is remarkably sharp, since M approaches zero at T_c with

[1] An often-expressed concern regarding the application of physics methods to the social sciences is that physical laws are said to apply to systems with a very large number of subunits (of order of $\approx 10^{20}$) while social systems comprise a much smaller number of elements. However, the "thermodynamic limit" is reached in practice for rather small systems. For example, in early computer simulations of gases or liquids reasonable results are already obtained for systems with 20–30 atoms.

infinite slope and hence M is not an analytic function. Such singular behavior is an example of a "critical phenomenon". Recently, the field of critical phenomena has been characterized by several important conceptual advances, two of which are scaling and universality.

2.1.1. Scaling

The scaling hypothesis has two categories of predictions, both of which have been remarkably well verified by a wealth of experimental data on diverse systems. The first category is a set of relations, called *scaling laws*, that serve to relate the various critical-point exponents characterizing the singular behavior of functions such as M.

The second category is a sort of *data collapse*, which is perhaps best explained in terms of our simple example of a uniaxial magnet. We may write the equation of state as a functional relationship of the form $M = M(H, \tau)$, where M is the order parameter, H is the magnetic field, and $\tau \equiv (T - T_c)/T_c$ is a dimensionless measure of the deviation of the temperature T from the critical temperature T_c. Since $M(H, \tau)$ is a function of two variables, it can be represented graphically and M vs. τ for a sequence of different values of H. The scaling hypothesis predicts that all the curves of this family can be "collapsed" onto a single curve provided one plots not M vs. τ but rather a *scaled* M (M divided by H to some power) vs. a *scaled* τ (τ divided by H to some different power).

The predictions of the scaling hypothesis are supported by a wide range of experimental work, and also by numerous calculations on model systems. Moreover, the general principles of scale invariance used here have proved useful in interpreting a number of other phenomena, ranging from elementary particle physics [73] and galaxy structure [74] to finance [2,75,76].

2.1.2. Universality

The second theme goes by the name "universality". It was found empirically that one could form an analog of the Mendeleev table if one partitions all critical systems into "universality classes". Consider, e.g., experimental *MHT* data on five diverse magnetic materials near their respective critical points. The fact that data for each material collapse onto a scaling function supports the scaling hypotheses, while the fact that the scaling function is the *same* (apart from two material-dependent scale factors) for all five diverse materials is truly remarkable. This apparent universality of critical behavior motivates the following question: "*Which features of this microscopic interparticle force are important for determining critical-point exponents and scaling functions, and which are unimportant?*"

Two systems with the same values of critical point exponents and scaling functions are said to belong to the same universality class. Thus, the fact that the exponents and scaling functions are the same for all five materials implies they all belong to the same universality class. Hence, we can pick a tractable system to study and the results we obtain will hold for all other systems in the same universality class.

2.2. Scaling and universality in systems outside of physics

At one time, many imagined that the "scale-free" phenomena are relevant to only a fairly narrow slice of physical phenomena [77–79]. However, the range of systems that apparently display power law and hence scale-invariant correlations has increased dramatically in recent years, ranging from base pair correlations in noncoding DNA [80,81], lung inflation [82] and interbeat intervals of the human heart [83–87] to complex systems involving large numbers of interacting subunits that display "free will", such as city growth [88–91], business firm growth [92–98], and even populations of birds [99].

Moreover, recent studies report evidence for scaling and universality in price fluctuations of financial assets [2,27–32,100–105]. Specifically, it appears that the *cumulative* distribution of returns for both individual companies and the S&P 500 index can be well described by a power law asymptotic behavior, characterized by an exponent $\alpha \approx 3$, well outside the stable Lévy regime $0 < \alpha < 2$ [12,27,32,101,103–105]. We have also found evidence for scaling: the distribution of returns, although not a stable distribution, retains its functional form for time scales from 5 min up to approximately 16 days for individual stocks [103–105].

These results were found for 1000 US stocks during the 1994–1995 period, suggesting *universality* in the dynamics of the return. Also suggestive of universality is the fact that identical results were found for the returns of the 30 German stocks comprising the DAX index [101], and similar results are found for currency exchange data [12].

3. Databases analyzed

Our empirical results are based on the analysis of different databases covering securities traded in the three major US stock exchanges, namely (i) the New York Stock Exchange (NYSE), (ii) the American Stock Exchange (AMEX), and (iii) the National Association of Securities Dealers Automated Quotation (Nasdaq).

For studying short time-scale dynamics, we are analyzing the Trades and Quotes (TAQ) database, from which we select the 4-year period January 1994 to December 1997. Nasdaq and AMEX have merged on October 1998, after the end of the period studied in this work. The TAQ database, which is published by NYSE since 1993, covers *all* trades at the three major US stock markets. This huge database is available in the form of CD-ROMs. The rate of publication was 1 CD-ROM per month for the period studied, but recently has increased to 2–4 CD-ROMs per month. The total number of transactions for the largest 1000 stocks is of the order of 10^9 in the 4-year period studied. We analyze the largest 1000 stocks, by capitalization on January 3, 1994, which survived through December 31, 1995. From the set of these 1000 stocks, we select a subset consisting of 880 stocks which survive through the further two years 1996–1997.

The data are adjusted for stock splits and dividends. The data are also filtered to remove spurious events, such as occur due to the inevitable recording errors. The most common error is missing digits which appears as a large spike in the time series of returns. These are much larger than usual fluctuations and can be removed by choosing an appropriate threshold. We tested a range of thresholds and found no effect on the results.

To study the dynamics at longer time horizons, we analyze the Center for Research and Security Prices (CRSP) database. The CRSP Stock Files cover common stocks listed on NYSE beginning in 1925, the AMEX beginning in 1962, and the Nasdaq Stock Market beginning in 1972. The files provide complete historical descriptive information and market data including comprehensive distribution information, high, low and closing prices, trading volumes, shares outstanding, and total returns. In addition to adjusting for stock splits and dividends, we have also detrended the data for inflation. The CRSP Stock Files provide monthly data for NYSE beginning December 1925 and daily data beginning July 1962. For the AMEX, both monthly and daily data begin in July 1962. For the Nasdaq Stock Market, both monthly and daily data begin in July 1972.

We also analyze the S&P 500 index, which comprises 500 stocks chosen for market size, liquidity, and industry group representation in the US. In our study, we first analyze high-frequency data that covers the 13-year period 1984–1996, with a recording frequency of less than 1 min. The total number of records in this database exceeds 4.5×10^6. To investigate longer time scales, we also study daily records of the S&P 500 index for the 35-year period 1962–1996, and monthly records for the 71-year period 1926–1996.

4. The distribution of stock price fluctuations

The nature of the distribution of price fluctuations in financial time series has been a topic of interest for over 100 years [24]. A reasonable a priori assumption, motivated by the central limit theorem, is that the returns are independent, identically Gaussian distributed (*i.i.d.*) random variables, which results in a Gaussian random walk in the logarithm of price [25].

Empirical studies [2,75,27,101–104,106–110] show that the distribution of returns has pronounced tails, in striking contrast to that of a Gaussian. In addition to being non-Gaussian, the process of returns shows another interesting property: "time scaling" – that is, the distributions of returns for various choices of Δt, ranging from 1 day up to 1 month have similar functional forms [2]. These results together would suggest that the distribution of returns is consistent with a Lévy stable distribution [2,106,111–113], the rationale for which arises from the generalization of the central limit theorem to random variables which do not have a finite second moment. Empirical studies suggest, however, that the tails of the return distribution are inconsistent with the stable Paretian hypothesis [29–33,75,101–105]. In particular, alternative hypotheses for modeling the return distribution were proposed, which include a log-normal mixture of Gaussiansg [33], Student t-distributions [29–31], and exponentially truncated Lévy distributions [75,114,115].

4.1. "Universality" of the distribution of returns

Conclusive results on the distribution of returns are difficult to obtain, and require a large amount of data to study the rare events that give rise to the tails. We analyze approximately 40 million records of stock prices sampled at 5 min intervals for the 1000 leading US stocks for the 2-year period 1994–1995 and 30 million records of daily returns for 6000 US stocks for the 35-year period 1962–1996.

The basic quantity studied for individual companies is the price $S_i(t)$. The time t runs over the working hours of the stock exchange – removing nights, weekends and holidays. For each company, we calculate the return

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

For small changes in $S_i(t)$, the return $G_i(t, \Delta t)$ is approximately the forward relative change, $G_i(t, \Delta t) \approx [S_i(t + \Delta t) - S_i(t)]/S_i(t)$. For time scales shorter than 1 day, we analyze the data from the TAQ database.

We then calculate the cumulative distributions – the probability of a return larger than or equal to a threshold – of returns G_i for Δt=5 min. For each stock i=1,…,1000, the asymptotic behavior of the functional form of the cumulative distribution is consistent with a power-law,

$$P\{G_i > x\} \sim \frac{1}{x^{\alpha_i}} \,, \tag{2}$$

where α_i is the exponent characterizing the power-law decay. In order to compare the returns of different stocks with different volatilities, we define the normalized return $g_i \equiv (G_i - \langle G_i \rangle_T)/v_i$, where $\langle \ldots \rangle_T$ denotes a time average over the 40 000 data points of each time series, for the 2-year period studied, and the time-averaged volatility v_i of company i is the standard deviation of the returns over the 2-year period $v_i^2 \equiv \langle G_i^2 \rangle_T - \langle G_i \rangle_T^2$. Values of the exponent α_i can be estimated by a power-law regression on each of these distributions $P\{g > x\} \sim x^{-\alpha}$, whereby we obtain the average value for the 1000 stocks,

$$\alpha = \begin{cases} 3.10 \pm 0.03 & \text{(positive tail)}\,, \\ 2.84 \pm 0.12 & \text{(negative tail)}\,, \end{cases} \tag{3}$$

where the fits are performed in the region $2 \leqslant g \leqslant 80$. In Fig. 1(a) we show the histogram for α_i, obtained from power-law regression-fits to the positive tails of the individual cumulative distributions of all 1000 companies studied, which shows an approximate Gaussian spread around the mean value $\alpha = 3.10 \pm 0.03$. These estimates of the exponent α are well outside the stable Lévy range, which requires $0 < \alpha < 2$, and is therefore consistent with a finite variance for returns. However, moments larger than 3, in particular the kurtosis, seem to be divergent [27,32]. Our results are consistent

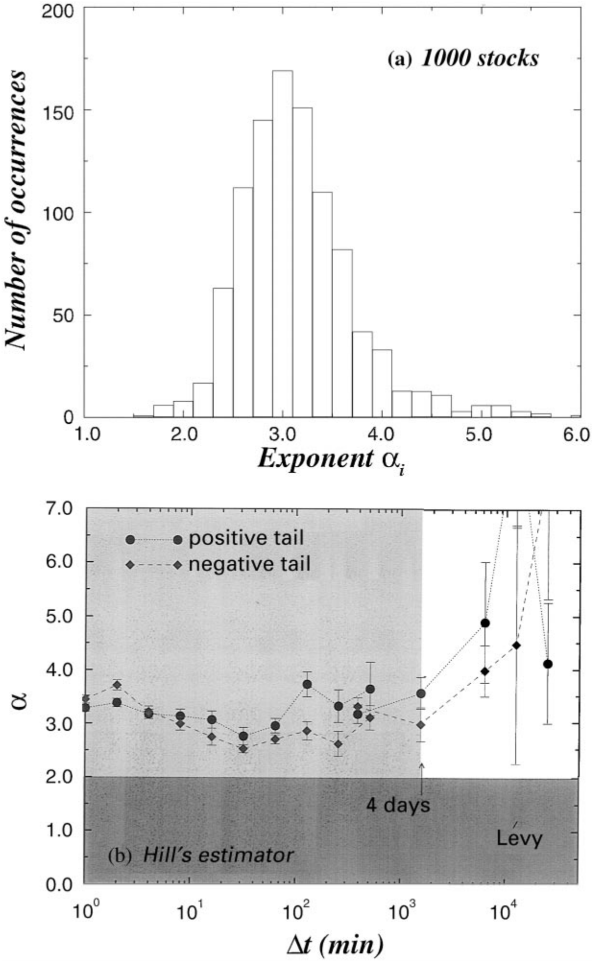

Fig. 1. (a) The histogram of the power-law exponents obtained by power-law regressions to the individual cumulative distribution functions, where the fit is for all x larger than 2 standard deviations. This histogram is not normalized – the y-axis indicates the number of occurrences of the exponent. (b) The values of the exponent α characterizing the asymptotic power-law behavior of the distribution of returns as a function of the time scale Δt obtained using Hill's estimator. The values of α for $\Delta t < 1$ day are calculated from the TAQ database while for $\Delta t \geqslant 1$ day they are calculated from the CRSP database. The unshaded region, corresponding to time scales larger than $(\Delta t)_\times \approx 16$ days (6240 min), indicates the range of time scales where we find results consistent with slow convergence to Gaussian behavior.

with the results of the analysis of the daily returns of 30 German stocks comprising the DAX index [101], daily CRSP returns [27], and foreign exchange rates [102].

In order to obtain an alternative estimate for α, we are using the methods of Hill [27,101,103,104,116]. We calculate the inverse local slope of the cumulative distribution function $P(g)$, $\gamma \equiv -(d \log P(g)/d \log g)^{-1}$ for the negative and the positive tail. We obtain an estimator for γ, by sorting the normalized increments by their size, $g^{(1)} > g^{(2)} > \cdots > g^{(N)}$. The cumulative distribution can then be written as $P(g^{(k)}) = k/N$, and we obtain for the local slope

$$\gamma = \left[(N-1) \sum_{i=1}^{N-1} \log g^{(i)} \right] - \log g^{(N)} , \tag{4}$$

where N is the number of tail events used. We use the criterion that N does not exceed 10% of the sample size, simultaneously ensuring that the sample is restricted to the tail events [27]. We thereby obtain the average estimates for 1000 stocks,

$$\alpha = \begin{cases} 2.84 \pm 0.12 & \text{(positive tail)} , \\ 2.73 \pm 0.13 & \text{(negative tail)} . \end{cases} \tag{5}$$

Removing overnight events yield the average values of $\alpha = 3.11 \pm 0.15$ for the positive tail and $\alpha = 3.03 \pm 0.21$ for the negative tail. Currently, we are also investigating the dependence of the exponent α on the time of day, by splitting a trading day into three equal parts of 130 min each. A parallel analysis on the S&P 500 index shows consistent asymptotic behavior [105], although the central part of the distribution seems to display Lévy behavior for short time scales (< 30 min) [75]. One reason for a different behavior at the central part of the distribution of S&P 500 returns is the discreteness [25,28] of the prices of individual stocks (which causes a cut-off for low values of returns) that comprise the S&P 500 index.

4.2. *Scaling of the distributions of returns and correlations in the volatility*

Since the values of α we find are inconsistent with a statistically stable law, we expect the distribution of returns $P(G)$ on larger time scales to converge to Gaussian. In contrast, our analysis of daily returns from the CRSP database suggests that the distributions of returns retain the same functional form for a wide range of time scales Δt, varying over 3 orders of magnitude, 5 min$\leqslant \Delta t \leqslant$6240 min$=$16 days [Fig. 1(b)]. The *onset* of convergence to a Gaussian starts to occur only for $\Delta t > 16$ days [104,105]. In contrast, n-partial sums of computer-simulated time series of the same length and probability distribution display Gaussian behavior for $n \geqslant 256$ [17,105]. Thus, the rate of convergence of $P(G)$ to a Gaussian is remarkably slow, indicative of time dependencies [25,52] which violate the conditions necessary for the central limit theorem to apply.

To test for time dependencies, we analyzed the autocorrelation function of returns, which we denote $\langle G(t)G(t+\tau)\rangle$, using 5 min returns of 1000 stocks. Our results show pronounced short-time (< 30 min) anti-correlations, consistent with the bid-ask bounce [25,117]. For larger time scales, the correlation function is at the level of noise (for some portfolios of common stocks Lo [54] has reported long memory), consistent with the efficient market hypothesis [25,107,118,119]. Lack of linear correlation does not imply independent returns, since there may exist higher-order correlations. Our recent studies [120] show that the amplitude of the returns measured by the absolute value or the square has long-range correlations with persistence [121–123] up to several months,

$$\langle |G(t)||G(t+\tau)|\rangle \sim \tau^{-a}\,, \tag{6}$$

where a has the average value $a = 0.34 \pm 0.09$ for the 1000 stocks studied. In order to detect genuine long-range correlations, the effects of the U-shaped intra-day pattern [124,125] for $|G|$ has been removed [120]. This result is consistent with earlier studies [25,126–128] which also noted long-range correlations. In addition to analyzing the correlation function directly, we are applying power spectrum analysis and the recently-developed detrended fluctuation analysis [120,129]. Both of these methods yield consistent estimates of the exponent a. We are also applying estimators such as those developed in Ref. [130] to obtain accurate estimates of the exponent a.

4.3. Statistics of trading activity

In order to understand the reasons for slow decaying tails in the return distribution and long-range correlations in volatility, we follow an approach in the spirit of models of time deformation proposed by Clark [33], Tauchen and Pitts [34], Stock [35], Lamoureux and Lastrapes [36], Ghysels and Jasiak [37], and Engle and Russell [41].

Returns G over a time interval Δt can be expressed as the sum of several changes δp_i due to the $i = 1,\ldots,N_{\Delta t}$ trades in the interval $[t, t+\Delta t]$,

$$G_{\Delta t} = \sum_{i=1}^{N_{\Delta t}} \delta p_i\,. \tag{7}$$

If Δt is such that $N_{\Delta t} \gg 1$, and δp_i have finite variance, then one can apply the classic version of the central limit theorem, whereby one would obtain the result that the unconditional distribution $P(G)$ is Gaussian [33,131]. It is implicitly assumed in this description that $N_{\Delta t}$ has only *narrow* Gaussian fluctuations i.e., has a standard deviation much smaller than the mean $\langle N_{\Delta t}\rangle$.

Our investigation of $N_{\Delta t}$ suggests stark contrast with a Gaussian time series with the same mean and variance – there are several events of the magnitude of tens of standard deviations which are inconsistent with Gaussian statistics [33–35,38,41,132–136]. For each stock analyzed, we choose sampling time intervals Δt such that it contains sufficient $N_{\Delta t}$; for actively-traded stocks $\Delta t = 15$ min, and for stocks with the least frequency of trading, $\Delta t = 390$ min (1 day) [135]. We find that the distribution of $N_{\Delta t}$ appears to display an asymptotic power-law decay

$$P\{N_{\Delta t} > x\} \sim x^{-\beta} \quad (x \gg 1)\,. \tag{8}$$

For the 1000 stocks that we analyze, we estimate β using Hill's method [116] and obtain a mean value $\beta = 3.40 \pm 0.05$. Note that $\beta > 2$ is outside the Lévy stable domain $0 < \beta < 2$ and is inconsistent with a stable distribution for $N_{\Delta t}$, and with the log-normal hypothesis of Clark [33].

4.4. Price fluctuations and trading activity

Since we find that $P\{G_{\Delta t} > x\} \sim x^{-\alpha}$, we can ask whether the value of β we find for $P\{N_{\Delta t} > x\}$ is sufficient to account for the fat tails of returns. To test this possibility, we implement, for each stock, the ordinary least-squares regression

$$\ln |G_{\Delta t}(t)| = a + b \ln N_{\Delta t}(t) + \psi(t)\,, \tag{9}$$

where $\psi(t)$ has mean zero and the equal time covariance $\langle N_{\Delta t}\psi(t)\rangle = 0$. Our results on 30 actively traded stocks yield the average value of $b = 0.57 \pm 0.09$.

Values of $b \approx 0.5$ are consistent with what we would expect from Eq. (7), if δp_i are *i.i.d.* with finite variance. In other words, suppose δp_i are chosen *only from the interval* $[t, t+\Delta t]$, and let us hypothesize that *these* δp_i are mutually independent, with a common distribution $P(\delta p_i \,|\, t \in [t, t+\Delta t])$ having a finite variance $W^2_{\Delta t}$. Under this hypothesis, the central limit theorem, applied to the sum of δp_i in Eq. (7), implies that the ratio

$$\varepsilon \equiv \frac{G_{\Delta t}}{W_{\Delta t}\sqrt{N_{\Delta t}}} \tag{10}$$

must be a Gaussian-distributed random variable with zero mean and unit variance [131]. We can test this hypothesis by analyzing the distribution $P(\varepsilon)$ and the correlations in ε.

Our results on 30 actively traded stocks seem to indicate that the distribution $P(\varepsilon)$ is consistent with a Gaussian, with mean values of excess kurtosis ≈ 0.1. This is noteworthy, since, for the unconditional distribution $P(G_{\Delta t})$, the kurtosis is divergent (empirical estimates yield mean values ≈ 80 for 1000 stocks).

If our hypothesis that $P(\varepsilon)$ is consistent with Gaussian is borne out by the data, this would imply that the fat tails of $P\{G_{\Delta t} > x\} \sim x^{-\alpha}$ cannot be caused solely due to $P\{N_{\Delta t} > x\} \sim x^{-\beta}$, because by conservation of probabilities $P\{\sqrt{N_{\Delta t}} > x\} \sim x^{-2\beta}$ with $2\beta \approx 6.8$. Eq. (10) then implies that $N_{\Delta t}$ alone cannot explain the value $\alpha \approx 3$.

Since $N_{\Delta t}$ is not sufficient to account for the fat tails in $G_{\Delta t}$, one other possibility is that it arises from $W_{\Delta t}$. By definition $W_{\Delta t}$ is the variance of all δp_i in Δt, which is difficult to estimate when one does not have sufficient $N_{\Delta t}$. We can investigate the statistics of $W_{\Delta t}$ and examine if the distribution of $W_{\Delta t}$ is sufficient to explain the value of α found for $P\{G_{\Delta t} > x\}$. Our results on 30 actively traded stocks suggest that $P\{W_{\Delta t} > x\} \sim x^{-\gamma}$, where we obtain rough estimates $\gamma = 2.85 \pm 0.20$, consistent with the estimates of α for the same 30 stocks. Estimates of γ are obtained by choosing $\Delta t = 15$ min for these stocks, at the same time ensuring that $N_{\Delta t} > 20$.

4.5. Volatility correlations and trading activity

Thus far we discussed Eq. (10) from the point of view of distributions. Next, we analyze time correlations in $N_{\Delta t}$, and relate them to the time correlations of $|G_{\Delta t}|$. Our studies on the same 30 actively traded stocks indicate that the autocorrelation function $\langle N_{\Delta t}(t) N_{\Delta t}(t+\tau)\rangle \sim \tau^{-\nu}$, with a mean value of the estimates of $\nu = 0.32 \pm 0.09$

using the detrended fluctuation analysis method [129]. To detect genuine long-range correlations, the marked U-shaped intra-daily pattern [124,125] in $N_{\Delta t}$ is removed [120]. We substantiate this analysis using semi-parametric estimators such as those due to Robinson [130].

Our long-term goal is to relate the exponent v of the autocorrelation function of $N_{\Delta t}$ to that of $|G_{\Delta t}|$. To this end, we also estimate, in parallel, the time correlations in $W_{\Delta t}$ and $|\varepsilon|$. Since our investigations on the 30 stocks seem to indicate the absence of long-range correlations in $W_{\Delta t}$, the above investigation of correlations could yield the interesting statement that the long-range correlations in volatility are due to those of $N_{\Delta t}$. Together with the above discussion on distribution functions, these results suggest an interesting result – that the fat tails of returns $G_{\Delta t}$ arise from $W_{\Delta t}$ and the long-range volatility correlations arise from trading activity $N_{\Delta t}$.

4.6. Statistics of share volume traded

Understanding the equal-time correlations between volume and volatility and, more importantly, understanding how the number of shares traded impacts the price has long been a topic of great interest [25,28,33,34,39,134–138]. The number of shares traded in Δt is the sum

$$Q_{\Delta t} \equiv \sum_{i=1}^{N_{\Delta t}} q_i \,, \tag{11}$$

where q_i traded for all $i = 1, \ldots, N_{\Delta t}$ transactions in Δt. So it is clear that $Q_{\Delta t}$ must be positively correlated with $N_{\Delta t}$.

Our results on 30 actively traded stocks suggest that the probability distributions $P\{Q_{\Delta t} > x\}$ are consistent with a power-law asymptotic behavior

$$P\{Q_{\Delta t} > x\} \sim x^{-\lambda} \,. \tag{12}$$

Using Hill's estimator, we obtain an average value $\lambda = 1.7 \pm 0.2$, within the Lévy stable domain $0 < \lambda < 2$. This result suggests that $Q_{\Delta t}$ can be effectively described using a one-sided (fully asymmetric) stable distribution. A parallel analysis for $P\{q_i > x\}$ (from Eq. (11)) yields consistent values of exponents within the Lévy stable domain, suggesting a divergent second moment. We will ultimately extend this result to all 1000 stocks and test the dependency of λ on the type of stock analyzed.

As a further test for Lévy stability of $Q_{\Delta t}$, we can investigate the scaling behavior of the sum $Q_n \equiv \sum_{i=1}^{n} q_i$, where n is a fixed number of trades. We first analyze the asymptotic behavior of $P(Q_n)$ for increasing n. For a Lévy stable distribution, $n^{1/\lambda} P([Q_n - \langle Q_n \rangle]/n^{1/\lambda})$ should have the same functional form as $P(q)$, where $\langle Q_n \rangle = n \langle q \rangle$ and $\langle \cdots \rangle$ denotes average values. We can also perform an independent test and estimate λ by analyzing the scaling behavior of the moments $\mu_r(n) \equiv \langle |Q_n - \langle Q_n \rangle|^r \rangle$, where $r < \lambda$. For a Lévy stable distribution $[\mu_r(n)]^{1/r} \sim n^{1/\lambda}$. Hence, by regressing

$[\mu_r(n)]^{1/r}$ as a function of n, we obtain an inverse slope which would yield an estimate of λ.

4.7. Share volume traded and number of trades

If our hypothesis is true that $Q_{\Delta t}$ (and q_i) are consistent with a one-sided Lévy stable process, then from Eq. (11), $N_{\Delta t}^{1/\lambda} P([Q_{\Delta t} - \langle q \rangle N_{\Delta t}]/N_{\Delta t}^{1/\lambda})$ should, from Eq. (11), have the same distribution as any of the q_i. Thus, we hypothesize that the dependence of $Q_{\Delta t}$ on $N_{\Delta t}$ can be separated by defining

$$\chi \equiv \frac{Q_{\Delta t} - \langle q \rangle N_{\Delta t}}{N_{\Delta t}^{1/\lambda}}, \tag{13}$$

where χ is a one-sided Lévy-distributed variable with zero mean and exponent λ. To test this hypothesis, we first analyze $P(\chi)$ for consistent asymptotic behavior to $P(Q_{\Delta t})$.

4.8. Time correlations in share volume traded

We also study extensively the time correlations in $Q_{\Delta t}(t)$. A difficulty arises due to the divergent second moment of the distribution $P(Q_{\Delta t})$. To circumvent this problem, we consider the family of correlation functions $\langle [Q_{\Delta t}(t)]^a [Q_{\Delta t}(t+\tau)]^a \rangle$, where the parameter a $(< \lambda/2)$ is required to ensure that the correlation function is well defined. Instead of analyzing the correlation function directly, we apply detrended fluctuation analysis [129], which has been successfully used to study long-range correlations in a wide range of complex systems. Our results suggest that $Q_{\Delta t}(t)$ has strong long-range correlations, while the number of shares traded in each transaction q_i (Eq. (11)) displays only short-range correlations, suggesting that long-range correlations in $Q_{\Delta t}$ can in turn be related to those of $N_{\Delta t}$, if Eq. (13) is found to be valid.

4.9. Returns and share volume traded

An interesting implication is an explanation for the previously observed [39,137,138] equal-time correlations between $Q_{\Delta t}$ and volatility $V_{\Delta t}$, which is the local standard deviation of price changes $G_{\Delta t}$. Now, $V_{\Delta t} = W_{\Delta t}\sqrt{N_{\Delta t}}$ from Eq. (10). Consider the equal-time correlation, $\langle Q_{\Delta t}\, V_{\Delta t} \rangle$, where the means are subtracted from $Q_{\Delta t}$ and $V_{\Delta t}$. Since $Q_{\Delta t}$ depends on $N_{\Delta t}$ through $Q_{\Delta t} = \langle q \rangle N_{\Delta t} + N_{\Delta t}^{1/\zeta}\chi$, and if the equal-time correlations $\langle N_{\Delta t}\, W_{\Delta t} \rangle$, $\langle N_{\Delta t}\, \chi \rangle$, and $\langle W_{\Delta t}\, \chi \rangle$ are small (correlation coefficients ≈ 0.1), it follows that the equal-time correlation $\langle Q_{\Delta t}\, V_{\Delta t} \rangle \propto \langle N_{\Delta t}^{3/2} \rangle - \langle N_{\Delta t} \rangle \langle N_{\Delta t}^{1/2} \rangle$, which is positive due to the Cauchy–Schwartz inequality.

5. Random matrix theory and correlation matrices

The correct determination of cross-correlations among stocks is of importance of both practical and scientific reasons. On the practical side, identifying cross-correlations permits better portfolio selection [25,139,140]. On the scientific side, identifying correlations enables us to investigate their origin which may help improve our understanding of the mechanisms governing stock price dynamics.

One approach to describe cross-correlations between the returns of different stocks is to consider returns as being composed of common and idiosyncratic components [25,55–72]. We follow a different approach, in the spirit of the method of principal components [25], but may provide a more rigorous way to estimate significant cross-correlations. We start with the cross-correlation matrix $\mathbf{C}$ of returns $G_i(t)$, with elements

$$C_{ij} \equiv \frac{\langle G_i G_j \rangle - \langle G_i \rangle \langle G_j \rangle}{\sigma_i \sigma_j}, \tag{14}$$

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.

For the high-frequency data from the TAQ database, we have 1000 stocks, which yields a 1000×1000 matrix. Since correlations between stocks might not be stationary, and because of the finite length of time series used to estimate C_{ij}, there is considerable degree of randomness in the measured C_{ij}. Thus, it is a difficult problem in general to estimate correlations from $\mathbf{C}$ that are not an effect of "randomness". We start with a "null hypothesis" that $\mathbf{C}$ is a random matrix – a correlation matrix constructed from mutually uncorrelated time series. Deviations of the properties of $\mathbf{C}$ from that of a random matrix would show genuine correlations. Statistical properties of matrices with independent random elements – random matrices – has a long history in physics since 1950s and their properties are well studied [141,142].

5.1. Brief overview of random matrix theory

The physics study of random matrices was initiated by the physicist E. Wigner, but the history of random matrices within mathematics is older. The problem that prompted Wigner to develop RMT was the explanation of the energy spectra of heavy 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. Although several models were developed in the 1950s to explain the nuclear spectra, they were largely unable to account for the exact positions of energy levels. RMT was developed in this context, to deal with the statistics of energy levels of complex quantum systems. In matrix notation, the Hamiltonian would be a matrix $\mathbf{H}$ with random elements H_{ij} drawn from a probability measure [141,142]. Based on this assumption, a series of remarkable predictions were made and were found to be in remarkable agreement with the experimental data [143–145].

It was later proved by Dyson and Metha [146] that RMT predictions represent an average over all possible interactions. Hence RMT predictions are *universal* predictions that will apply to wide classes of systems. Moreover, deviations from the universal predictions of RMT identify system-specific, non-random properties of the system under consideration, providing clues about the underlying interactions [141,142].

5.2. Comparison of the eigenvalue statistics of **C** with a random correlation matrix

Let us consider a random correlation matrix **A** constructed from random time series **X** that are uncorrelated $\mathbf{A}=(1/M)\ \mathbf{X}\ \ \mathbf{X}^{\mathrm{T}}$, where **X** is an $N \times M$ matrix containing N time series of M random elements each (with zero mean and unit variance), that are mutually uncorrelated. The properties of random matrices **A** are well studied [147,148], particularly, in the limit $N \to \infty$, $M \to \infty$, such that $Q \equiv M/N$ is held fixed, it was shown analytically that the distribution $\rho(\lambda)$ of eigenvalues λ of **A** is given by

$$\rho(\lambda)=\begin{cases} \dfrac{Q}{2\pi}\dfrac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} & [\lambda_- \leqslant \lambda \leqslant \lambda_+]\,, \\ 0 & [\lambda > \lambda_+, \lambda < \lambda_-]\,, \end{cases} \tag{15}$$

where λ_+ and λ_- are the maximum and minimum eigenvalues of **A**, respectively, given by

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\,. \tag{16}$$

For finite M and N, the abrupt cut-off of $\rho(\lambda)$ is replaced by a rapidly-decaying edge. We compare the eigenvalue distribution $P(\lambda)$ of $\mathbf{C}=(1/M)\ \mathbf{G}\ \ \mathbf{G}^{\mathrm{T}}$, where **G** denotes the time series of returns (normalized to zero mean and unit variance) of N stocks, with $\rho(\lambda)$ [149–152].

We first examine returns at time scale $\Delta t = 30$ min for $N = 1000$ stocks, each containing $M = 6448$ records. We compute the eigenvalues λ_i of the empirical correlation matrix **C**, where λ_i are rank ordered $(\lambda_{i+1} > \lambda_i)$. Fig. 2(a) compares the probability distribution $P(\lambda)$ with $\rho(\lambda)$ calculated for $Q = 6.448$. We note the presence of a well-defined “bulk” of eigenvalues which fall within $\rho(\lambda)$. We also note deviations for some of the large eigenvalues. In particular, the largest eigenvalue $\lambda_{1000} \approx 50$, for the 2-year period, which is approximately 25 times larger than λ_+. To confirm that the deviations for large eigenvalues are genuine, we compare $P(\lambda)$ for a correlation matrix generated from $N = 1000$ uncorrelated time series with the same length $M = 6448$ and find perfect agreement [Fig. 2(b)] – suggesting that the deviations from RMT found for the large eigenvalues in Fig. 2(a) are genuine. An analysis of $P(\lambda)$ for **C** calculated using $M = 1737$ daily returns of 422 stocks for the 7-year period 1990–1996 reveals a well defined bulk of eigenvalues that agree with $\rho(\lambda)$, and deviations from $\rho(\lambda)$ for large eigenvalues – similar to what we find for $\Delta t = 30$ min.

5.3. Testing the eigenvalue statistics of **C** for universal properties of random matrices

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

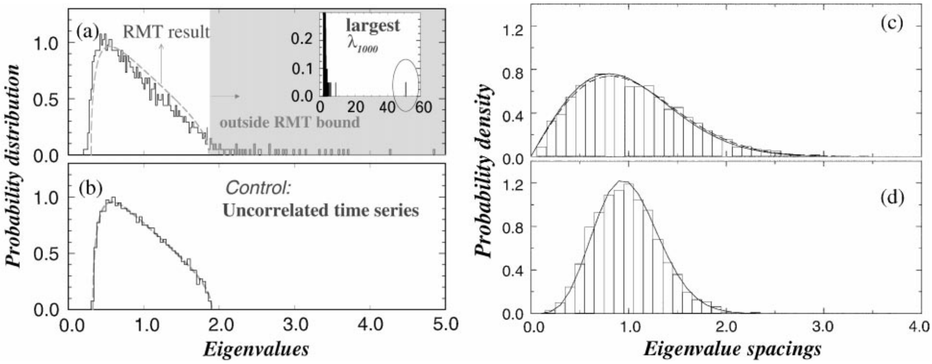

Fig. 2. (a) Eigenvalue distribution $P(\lambda)$ of **C** compared with the RMT result for a random correlation matrix shows the presence of a random "bulk" with falls within the RMT bound $\lambda_+ < \lambda < \lambda_-$. We also note the presence of several eigenvalues above the upper bound λ_+ of RMT; in particular, the largest eigenvalue $\lambda_{1000} \approx 50$ – approximately 25 times larger than λ_+. (b) $P(\lambda)$ for a correlation matrix of the same size generated from mutually uncorrelated time series shows good agreement with the RMT result. (c) Nearest-neighbor spacing distribution of the eigenvalues of **C** after unfolding. The results shown are obtained using the Gaussian broadening procedure for unfolding. 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. The solid line is the GOE prediction, 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 $B \equiv [\Gamma((\beta+2)/(\beta+1))]^{1+\beta}$. 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. (d) Next-nearest-neighbor 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.

transforming the eigenvalues in such a way that their distribution becomes uniform – a procedure known as unfolding [142]. Fig. 2(c) 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 real symmetric 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,

$$P_{\mathrm{GOE}}(s) = \frac{\pi s}{2}\exp\left(-\frac{\pi}{4}s^2\right) . \tag{17}$$

A second independent test of the GOE is the distribution of *next*-nearest-neighbor spacings between the rank-ordered eigenvalues [142]. This distribution is expected to be identical to the distribution of nearest-neighbor spacings of the Gaussian symplectic ensemble (GSE), which is consistent with the empirical data [150] [Fig. 2(d)].

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 [142],

$$\Sigma^2(L) \equiv \left\langle \left[N\left(\lambda + \frac{L}{2}\right) - N\left(\lambda - \frac{L}{2}\right) - L \right]^2 \right\rangle_\lambda , \tag{18}$$

where $N(\lambda) \equiv \sum_i \theta(\lambda - \lambda_i)$ is the integrated density of eigenvalues and $\langle \cdots \rangle_\lambda$ denotes an average over λ [142]. 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 [150].

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 [142],

$$\Delta(L) \equiv \frac{1}{L} \left\langle \min_{A,B} \int_{\lambda - L/2}^{\lambda + L/2} (N(\lambda_1) - A\lambda_1 - B)^2 \, \mathrm{d}\lambda_1 \right\rangle_\lambda , \tag{19}$$

where $\langle \cdots \rangle_\lambda$ denotes an average over λ and $N(\lambda) \equiv \sum_i \theta(\lambda - \lambda_i)$ is the integrated density of eigenvalues [142]. 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 [150].

The agreement of the eigenvalue statistics of $\mathbf{C}$ with RMT results implies that $\mathbf{C}$ has entries that contain a considerable degree of noise. Such noise could be the result of either nonstationary correlations or a result of the finite time series used. To test that finiteness of time series alone cannot be the reason for RMT agreement, we increase the length of the time series M used to compute $\mathbf{C}$ by a factor of 4. We still find agreement of the eigenvalue spacing distribution with RMT predictions, suggesting that RMT agreement is also due to non-stationary correlations. From the practical side, RMT agreement of the statistics of $\mathbf{C}$ argues *against* the wide use of empirically measured C_{ij} in a variety of applications.

5.4. Deviations from RMT predictions

The results presented in the previous section regard universal properties of the cross-correlation matrix that agree well with RMT predictions. Deviations from RMT indicate properties that are specific to the system and arise from the presence of collective modes. For example, deviations of the level spacings of certain nuclei from the Wigner distribution were found to be connected to collective modes of the nucleus. For the stock market, an interesting question one may ask is “how can one detect collective behavior?” Following the methods of RMT, we know that one approach is to study the eigenvalue distribution of $\mathbf{C}$.

Since our aim is to extract information about cross-correlations from $\mathbf{C}$, we need to compare the properties of $\mathbf{C} = \frac{1}{M}\, \mathbf{G}\ \mathbf{G}^{\mathrm{T}}$ with those of a random matrix with the same structure. Thus, in order to separate genuine correlations from randomness, we seek to group the content of $\mathbf{C}$ into two disjoint classes: (a) the part of $\mathbf{C}$ that agrees with the properties of $\mathbf{A}$ and (b) the part of $\mathbf{C}$ that deviates from the properties of $\mathbf{A}$.

Our work suggests that the eigenvalue distribution of $\mathbf{C}$ includes several eigenvalues outside the upper bound (λ_+) predicted for random matrices. In order to interpret their meaning, we must analyze the eigenvectors of $\mathbf{C}$. We will analyze the statistics of the eigenvectors [149,150]. RMT predicts that the distribution of eigenvector components for a random matrix is a Gaussian with zero mean and unit variance. Our examination of the eigenvectors corresponding to the eigenvalues which deviate from the random-matrix bound show systematic deviations from the Gaussian prediction.

The largest eigenvalue is strongly non-Gaussian, tending to uniform – suggesting that all companies participate equally. This observation can be tested by comparing the returns of the portfolio defined by $\mathbf{u}^{1000}$, $G^{1000}(t) \equiv \sum_{i=1}^{N} u_i^{1000} G_i(t)$, with a commonly-used indicator of market performance, the S&P 500 index. An ordinary least-squares regression between the two suggests a large degree of dependence indicated by an equal time correlation coefficient $= 0.85 \pm 0.03$ for the two years analyzed. Thus, the largest eigenvalue corresponds to the entire market influence that is common to all stocks, which is consistent with the commonly-used one factor market model [25].

One explanation for eigenvalues that deviate from the RMT upper bound is the commonly used multifactor models [25,69,140]. One sensitive test to determine the number of significantly deviating eigenvalues from RMT would be to examine the agreement of the correlation matrix to the universal properties of random matrices discussed above, as a function of the number of eigenvalues excluded.

In particular, we have examined the value of the largest eigenvalue as a function of the sample size. An asymptotic extrapolation of this dependence (finite-size scaling) suggests that in the infinite size limit this eigenvalue tends to infinity. Such phenomena occur in the physics of systems near the vicinity of a critical point, and are suggestive of a collective mode.

5.5. *Quantifying the number of significant participants*

To analyze the remainder of the deviating eigenvectors in a systematic way, we introduce the concept of inverse participation ratios (IPR), which is commonly used in localization theory [142]. The IPR for an eigenvector $\mathbf{u}^k$ quantifies the reciprocal of the number of its significant contributors and is defined as $I^k \equiv \sum_{i=1}^{N} [u_i^k]^4$. The meaning of IPR can be illustrated by two limiting cases: (i) a vector with identical components $u_\ell^k \equiv 1/\sqrt{N}$ has $I^k = 1/N$, whereas (ii) a vector with one component $u_1^k = 1$ and all the others zero has $I^k = 1$. Therefore, IPR quantifies the reciprocal of the number of eigenvector components that participate significantly.

An examination of the IPR values indicate that all components participate approximately equally to the largest eigenvector [150]. The remainder of the eigenvalues that deviate from RMT upper bound have varying degrees of participation. In order to interpret their meaning, we extract $1/I^k$ significant components of each of these deviating eigenvectors and identify common features such as type of stock, type of industry, or geographic region of activity.

Both "edges" of the eigenvalue spectrum of $\mathbf{C}$ show significant deviations of I^k from the average value $\langle I \rangle$, i.e., a plot of IPR as a function of eigenvalue displays a U shape. For the largest eigenvalues which deviated from the RMT bulk, I^k values are approximately 4–5 times larger than $\langle I \rangle$ which suggests that there are varying numbers of stocks participating to these eigenvectors. The corresponding eigenvalues are well outside the random bulk, suggesting that these companies are correlated. In addition, we also find that there are I^k values as large as 0.35 for vectors corresponding to the smallest eigenvalues $\lambda_i \approx 0.25$. These deviations are considerably larger than $\langle I \rangle$, which suggests that the vectors have only a few companies stocks contributing significantly.

We also note that the presence of vectors with large I^k at the edges of the eigenvalue spectrum (U-shaped dependence of IPR on eigenvalue) also arises in the theory of Anderson localization [142]. In the context of localization theory, one frequently finds "random band matrices" [142] which give rise to eigenvectors with small I^k in the middle of the band, whereas the eigenvectors at the edge have large I^k. A random band matrix $\mathbf{B}$ has elements B_{ij} independently drawn from different probability distributions. These distributions are often taken to be Gaussian, 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. Our finding of localized states for small and large eigenvalues of the cross-correlation matrix $\mathbf{C}$ is reminiscent of Anderson localization and suggests that $\mathbf{C}$ may be a random band matrix.

5.6. Time stability of RMT deviations

Another question to be investigated to validate our results is the that of the stability in time of the eigenvectors corresponding to the eigenvalues that deviate from RMT bounds. To test the time stability, we first split the entire two year period into four six-month sub-periods A, B, C, and D. For each sub-period, we calculate a cross-correlation matrix, and compute its eigenvalues and eigenvectors. We can then identify, from each sub-period, the p largest eigenvectors that deviate from the RMT bounds. Let us denote by a_i; $i = 1, \ldots, p$, the p eigenvectors of period A (in ascending order of eigenvalue), and similarly b_j; $j = 1, \ldots, p$ for period B. One may measure time stability by the scalar product

$$O_{ij}(\tau) \equiv \sum_{\ell=1}^{N} a_{i\ell}(t) b_{\ell j} \qquad (20)$$

where $\mathbf{O}$ is a $p \times p$ matrix, and $N = 1000$ is the number of components of each eigenvector. If the vectors are perfectly stable, then we expect O_{ij} to be diagonal with elements $O_{ij} = \delta_{ij}$, where δ_{ij} is the Kronecker delta. No stability would mean all elements of O_{ij} have values close to zero. Our results suggest that the eigenvectors corresponding to the largest 4–5 eigenvalues show large values of O_{ij}. As we move toward the RMT bound, the eigenvectors show decreasing amounts of stability.

References

[1] E. Majorana, Il valore delle leggi statistiche nella fisica e nelle scienze sociali, Scientia 36 (1942) 58–66.

[2] B.B. Mandelbrot, The variation of certain speculative prices, J. Business 36 (1963) 394–419.

[3] P.W. Anderson, More is different, Science 177 (1972) 393–396.

[4] E.W. Montroll, W.W. Badger, Introduction to Quantitative Aspects of Social Phenomena, Gordon and Breach, New York, 1974.

[5] P.W. Anderson, J.K. Arrow, D. Pines (Eds.), The Economy as an Evolving Complex System, Addison-Wesley, Redwood City, 1988.

[6] W. Weidlich, Physics and social science: the approach of synergistics, Phys. Rep. 204 (1992) 1–163.

[7] B.M. Roehner, Theory of Markets, Springer, Berlin, 1995.

[8] R.N. Mantegna, Lévy Walks and enhanced diffusion in Milan stock exchange, Physica A 179 (1991) 232–242.

[9] H. Takayasu, H. Miura, T. Hirabayashi, K. Hamada, Statistical properties of deterministic threshold elements: the case of market price, Physica A 184 (1992) 127–134.

[10] P. Bak, K. Chen, J. Scheinkman, M. Woodford, Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics, Ricerche Economiche 47 (1993) 3–30.

[11] J.-P. Bouchaud, D. Sornette, The black & scholes option pricing problem in mathematical finance: generalization and extensions for a large class of stochastic processes, J. Phys. I (France) 4 (1994) 863–881.

[12] U.A. Müller, M.M. Dacorogna, R.B. Olsen, O.V. Pictet, M. Schwarz, C. Morgenegg, Statistical study of foreign exchange rates, empirical evidence of a price change scaling law and intraday analysis, J. Banking Finance 14 (1990) 1189–1208.

[13] W.B. Arthur, S. Durlauf, D. Lane, The Economy as a Complex Evolving System II, Addison-Wesley, Redwood City, 1997.

[14] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, 1999.

[15] J.D. Farmer, Physicists attempt to scale the ivory towers of finance, Comput. Sci. Eng. 1 (1999) 26–39.

[16] S. Moss De Oliveira, P.M.C de Oliveira, D. Stauffer, Evolution, Money, War and Computers, Teubner, Stuttgart, 1999.

[17] J.P. Bouchaud, M. Potters, Theory of Financial Risk, Cambridge University Press, Cambridge, 2000.

[18] H. Levy, M. Levy, S. Solomon, Microscopic Simulation of Financial Markets, Academic Press, New York, 2000.

[19] S.N. Durlauf, How can statistical mechanics contribute to social science? Proc. Natl. Acad. Sci. USA 96 (1999) 10 582–10 584.

[20] W. Brock, C. Hommes, Rational routes to randomness, Econometrica 65 (1998) 1059–1096.

[21] S. Durlauf, Statistical mechanics approaches to socioeconomic behavior, in: W.B. Arthur, S. Durlauf, D. Lane (Eds.), The Economy as a Complex Evolving System II, Addison-Wesley, Redwood City, 1997.

[22] P. Krugman, The Self-Organizing Economy, Basil Blackwell, Oxford, 1996.

[23] L. Blume, The statistical mechanics of strategic interaction, Games Econom. Behavior 5 (1993) 387–424.

[24] L. Bachelier, Théorie de la spéculation, Ph.D. Thesis, Annales Scientifiques de l'Ecole Normale Supérieure III-17, 1900, p. 21.

[25] J. Campbell, A.W. Lo, A.C. MacKinlay, The Econometrics of Financial Markets, Princeton University Press, Princeton, 1997.

[26] M. O'Hara, Market Microstructure Theory, Blackwell, Cambridge MA, 1995.

[27] A. Pagan, The econometrics of financial markets, J. Empirical Finance 3 (1996) 15–102.

[28] J. Hausman, A. Lo, A.C. MacKinlay, An ordered probit analysis of stock transaction prices, J. Financial Econom. 31 (1992) 319–379.
[29] R.R. Officer, The distribution of stock returns, J. Am. Statist. Assoc. 67 (1972) 807–812.
[30] P.D. Praetz, The distribution of share price changes, J. Business 45 (1972) 49–55.
[31] R.C. Blattberg, N. Gonedes, A Comparison of the stable paretian and student distributions as statistical models for prices, J. Business 47 (1974) 244–280.
[32] M. Loretan, P.C.B. Phillips, Testing the covariance stationarity of heavy-tailed time series: An overview of applications to several financial data sets, J. Empirical Finance 1 (1994) 211–248.
[33] P.K. Clark, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41 (1973) 135–155.
[34] G. Tauchen, M. Pitts, The price variability-volume relationship on speculative markets, Econometrica 57 (1983) 485–505.
[35] J. Stock, Estimating continuous time processes subject to time deformation, J. Am. Statist. Assoc. 83 (1988) 77–85.
[36] C.G. Lamoureux, W.D. Lastrapes, Heteroskedasticity in stock return data: volume versus GARCH effects, J. Finance 45 (1990) 221–229.
[37] E. Ghysels, J. Jasiak, Stochastic volatility and time deformation: an application to trading volume and leverage effects, CRDE, Université de Montreal, preprint.
[38] E. Ghysels, A. Harvey, E. Renault, Stochastic volatility, in: G.S. Maddala (Ed.), Handbook of Statistics, Vol. 14, Statistical Methods in Finance, North-Holland, Amsterdam, 1996, pp. 119–191.
[39] C. Jones, K. Gautam, M. Lipson, Transactions, volumes and volatility, Rev. Financial Stud. 7 (1994) 631–651.
[40] R.F. Engle (Ed.), ARCH: Selected Readings, Oxford University Press, Oxford, 1995.
[41] R.F. Engle, J.R. Russell, Autoregressive conditional duration: a new model for irregularly spaced transaction data, Econometrica 66 (1998) 1127–1162.
[42] T. Bollerslev, R. Chou, K. Kroner, ARCH modelling in finance: a review of the theory and empirical evidence, J. Econom. 52 (1992) 5–59.
[43] T. Bollerslev, R. Engle, D. Nelson, ARCH Models, in: R. Engle, D. McFadden (Eds.), Handbook of Econometrics, Vol. IV, Elsevier, Amsterdam, 1994.
[44] W. Brock, P.J.F. De Lima, Nonlinear time series, complexity theory and finance, in: G.S. Maddala, C.R. Rao (Eds.), Handbook of Statistics Vol. 14, Elsevier, Amsterdam, 1996, pp. 317–361.
[45] B. Le Baron, Some relations between volatility and serial correlations in stock returns, J. Business 65 (1992) 199–219.
[46] B. Le Baron, Chaos and nonlinear forecastability in economics and finance, Philos. Trans. Roy. Soc. London, Ser. A 348 (1994) 397–404.
[47] J. Hasbrouck, Trades, quotes, inventories and information, J. Financial Econom. 22 (1988) 229–252.
[48] J. Hasbrouck, Measuring the information content of stock trades, J. Finance 46 (1991) 179–207.
[49] A.S. Kyle, Continuous auctions and insider trading, Econometrica 53 (1985) 1315–1336.
[50] L.R. Glosten, P.R. Milgrom, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, J. Financial Econom. 14 (1985) 71–100.
[51] D. Easley, M. O'Hara, Time and the process of security price adjustment, J. Finance 47 (1992) 577–606.
[52] A. Lo, A.C. MacKinlay, Stock prices do not follow random walks: Evidence from a simple specification test, Rev. Financial Stud. 1 (1988) 41–66.
[53] A. Lo, A.C. MacKinlay, An econometric analysis of nonsynchronous trading, J. Econom. 45 (1990) 181–212.
[54] A. Lo, Long term memory in stock market prices, Econometrica 59 (1991) 1279–1313.
[55] W. Sharpe, Portfolio Theory and Capital Markets, McGraw-Hill, New York, NY, 1970.
[56] W. Sharpe, G. Alexander, J. Bailey, Investments, 5th Edition, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[57] W. Sharpe, Capital asset prices: a theory of market equilibrium under conditions of risk, J. Finance 19 (1964) 425–442.
[58] J. Lintner, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, Rev. Econom. Statist. 47 (1965) 13–37.
[59] S. Ross, The arbitrage theory of capital asset pricing, J. Econom. Theory 13 (1976) 341–360.
[60] S. Brown, M. Weinstein, Derived factors in event studies, J. Financial Econom. 14 (1985) 491–495.

[61] F. Black, Capital market equilibrium with restricted borrowing, J. Business 45 (1972) 444–454.
[62] M. Blume, I. Friend, A new look at the capital asset pricing model, J. Finance 28 (1973) 19–33.
[63] E. Fama, K. French, The cross-section of expected stock returns, J. Finance 47 (1992) 427–465.
[64] E. Fama, K. French, Common risk factors in the returns on bonds and stocks, J. Financial Econom. 33 (1993) 3–56.
[65] E. Fama, J. Macbeth, Risk, return, and equilibrium: empirical tests, J. Political Econom. 71 (1973) 607–636.
[66] R. Roll, S. Ross, On the cross-sectional relation between expected returns and betas, J. Finance 49 (1994) 101–122.
[67] N. Chen, R. Roll, S. Ross, Economic forces and the stock market, J. Business 59 (1986) 383–403.
[68] R. Merton, An intertemporal capital asset pricing model, Econometrica 41 (1973) 867–887.
[69] B. Lehmann, D. Modest, The empirical foundations of the arbitrage pricing theory, J. Finance Econom. 21 (1988) 213–254.
[70] J. Campbell, Understanding risk and return, J. Political Econom. 104 (1996) 298–345.
[71] J. Campbell, J. Ammer, What moves the stock and bond markets? a variance decomposition for long-term asset returns, J. Finance 48 (1993) 3–37.
[72] G. Connor, R. Korajczyk, A test for the number of factors in an approximate factor structure, J. Finance 48 (1993) 1263–1291.
[73] R. Jackiw, Introducing scale symmetry, Phys. Today 25 (1) (1972) 23–27.
[74] P.J.E. Peebles, The Large-Scale Structure of the Universe, Princeton University Press, Princeton, NJ, 1980.
[75] R.N. Mantegna, H.E. Stanley, Scaling behavior in the dynamics of an economic index, Nature 376 (1995) 46–49.
[76] R.N. Mantegna, H.E. Stanley, Physics investigation of financial markets, in: F. Mallamace, H.E. Stanley (Eds.), Proceedings of the International School of Physics Enrico Fermi, Course CXXXIV, IOS Press, Amsterdam, 1997.
[77] H.E. Stanley, Introduction to Phase Transitions and Critical Phenomena, Oxford University Press, Oxford, 1971.
[78] A.-L. Barabasi, H.E. Stanley, Fractal Concepts in Surface Growth, Cambridge University Press, Cambridge, 1995.
[79] H.E. Stanley, Scaling, universality, and renormalization: three pillars of modern critical phenomena, Rev. Mod. Phys. 71 (1999) S358–S366.
[80] C.K. Peng, S. Buldyrev, A. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Long-range correlations in nucleotide sequences, Nature 356 (1992) 168–171.
[81] R.N. Mantegna, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, H.E. Stanley, Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics, Phys. Rev. E 52 (1995) 2939–2950.
[82] B. Suki, A.-L. Barabási, Z. Hantos, F. Peták, H.E. Stanley, Avalanches and power law behaviour in lung inflation, Nature 368 (1994) 615–618.
[83] C.K. Peng, J. Mietus, J. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, Non-equilibrium behavior of the heartbeat, Phys. Rev. Lett. 70 (1993) 1343–1346.
[84] P.Ch. Ivanov, M.G. Rosenblum, C.-K. Peng, J. Mietus, S. Havlin, H.E. Stanley, A.L. Goldberger, Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis, Nature 383 (1996) 323–327.
[85] L.A.N. Amaral, A.L. Goldberger, P.Ch. Ivanov, H.E. Stanley, Scale-independent measures and pathologic cardiac dynamics, Phys. Rev. Lett. 81 (1998) 2388–2391.
[86] P.Ch. Ivanov, L.A.N. Amaral, A.L. Goldberger, H.E. Stanley, Stochastic feedback and the regulation of biological rhythms, Europhys. Lett. 43 (1998) 363–368.
[87] P.Ch. Ivanov, L.A.N. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z. Struzik, H.E. Stanley, Multifractality in human heartbeat dynamics, Nature 399 (1999) 461–465.
[88] H. Makse, S. Havlin, H.E. Stanley, Modeling urban growth patterns, Nature 377 (1995) 608–612.
[89] H.A. Makse, J.S. Andrade, M. Batty, S. Havlin, H.E. Stanley, Modeling urban growth patterns with correlated percolation, Phys. Rev. E 58 (1998) 7054–7062.
[90] X. Gabaix, Zipf's law for cities: an explanation, Quarterly J. Econom. 114 (1999) 739–767.
[91] G. Malescio, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, Hierarchical organization of cities and nations, preprint (2000).

[92] M.H.R. Stanley, S.V. Buldyrev, S. Havlin, R. Mantegna, M.A. Salinger, H.E. Stanley, Zipf plots and the size distribution of firms, Econom. Lett. 49 (1995) 453–457.
[93] J. Sutton, Gibrat's legacy, J. Econom. Lit. 35 (1997) 40–59.
[94] M.H.R. Stanley, L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, Scaling behavior in the growth of companies, Nature 379 (1996) 804–806.
[95] L.A.N. Amaral, S.V. Buldyrev, S. Havlin, M.A. Salinger, H.E. Stanley, Power law scaling for a system of interacting units with complex internal structure, Phys. Rev. Lett. 80 (1998) 1385–1388.
[96] Y. Lee, L.A.N. Amaral, D. Canning, M. Meyer, H.E. Stanley, Universal features in the growth dynamics of complex organizations, Phys. Rev. Lett. 81 (1998) 3275–3278.
[97] V. Plerou, L.A.N. Amaral, P. Gopikrishnan, M. Meyer, H.E. Stanley, Similarities between the growth dynamics of university research and of competitive economic activities, Nature 400 (1999) 433–437.
[98] J. Sutton, The variance of firm growth rates: the 'scaling' puzzle, Working Paper, London School of Economics, 2000.
[99] T. Keitt, H.E. Stanley, Scaling in the dynamics of North American breeding-bird populations, Nature 393 (1998) 257.
[100] D. Jansen, C. de Vries, On the frequency of large stock returns: putting booms and busts into perspective, Rev. Econom. Statist. 73 (1991) 18–24.
[101] T. Lux, The stable paretian hypothesis and the frequency of large returns: an examination of major German stocks, Appl. Finance Econom. 6 (1996) 463–475.
[102] U.A. Muller, M.M. Dacorogna, O.V. Pictet, Heavy tails in high-frequency financial data, in: R.J. Adler, R.E. Feldman, M.S. Taqqu (Eds.), A Practical Guide to Heavy Tails, Birkhäuser, Basel, 1998, pp. 83–311.
[103] P. Gopikrishnan, M. Meyer, L.A.N. Amaral, H.E. Stanley, Inverse cubic law for the distribution of stock price variations, Eur. Phys. J. B 3 (1998) 139–140.
[104] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, Scaling of the distribution of price fluctuations of individual companies, Phys. Rev. E 60 (1999) 6519–6529.
[105] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, H.E. Stanley, Scaling of the distributions of fluctuations of financial market indices, Phys. Rev. E 60 (1999) 5305–5316.
[106] E.F. Fama, Mandelbrot and the stable paretian distribution, J. Business 36 (1963) 420–429.
[107] E.F. Fama, The behavior of stock market prices, J. Business 38 (1965) 34–105.
[108] S. Ghashgaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Turbulent cascades in foreign exchange markets, Nature 381 (1996) 767–770.
[109] R.N. Mantegna, H.E. Stanley, Turbulence and exchange markets, Nature 383 (1996) 587–588.
[110] R.N. Mantegna, H.E. Stanley, Stock market dynamics and turbulence: parallel analysis of fluctuation phenomena [Proceedings of the International Conference on Pattern Formation in Fluids and Materials], Physica A 239 (1997) 255–266.
[111] V. Pareto, Cours d'Economie Politique, Marcel Giard, Lausanne and Paris, 1897.
[112] P. Lévy, Théorie de l'Addition des Variables Aléatoires, Gauthier-Villars, Paris, 1937.
[113] E. Fama, R. Roll, Parameter estimates for symmetric stable distributions, J. Am. Statist. Assoc. 66 (1971) 331–338.
[114] R.N. Mantegna, H.E. Stanley, Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight, Phys. Rev. Lett. 73 (1994) 2946–2949.
[115] B. Podobnik, P.Ch. Ivanov, Y. Lee, A. Chessa, H.E. Stanley, Truncated Lévy flights in the presence of variance correlations, Europhys. Lett. 50 (2000) 711–717.
[116] B. Hill, A simple general approach to inference about the tail of a distribution, Ann. Math. Statist. 3 (1975) 1163–1174.
[117] R. Roll, A simple implicit measure of the effective bid-ask spread in an efficient market, J. Finance 39 (1984) 1127–1140.
[118] E.F. Fama, Efficient capital markets: a review of theory and empirical work, J. Finance 25, 383–420 (1984) (1970).
[119] E.F. Fama, Efficient capital markets: II, J. Finance 46 (1991) 1575–1617.
[120] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, The statistical properties of the volatility of price fluctuations, Phys. Rev. E 60 (1999) 1390–1400.
[121] C.W.J. Granger, The typical spectral shape of an economic variable, Econometrica 34 (1966) 150–161.
[122] C.W.J. Granger, Long memory relationships and the aggregation of dynamic models, J. Econometrics 14 (1980) 227–238.

[123] J. Beran, Statistics for Long-Memory Processes, Chapman & Hall, New York, 1994.
[124] R.A. Wood, T.H. McInish, J.K. Ord, An investigation of transactions data for NYSE stocks, J. Finance 40 (1985) 723–739.
[125] A. Admati, P. Pfleiderer, A theory of intraday patterns: Volume and price variability, Rev. Financial Stud. 1 (1988) 3–40.
[126] Z. Ding, C.W.J. Granger, R.F. Engle, A long memory property of stock market returns and a new model, J. Empirical Finance 1 (1993) 83–105.
[127] C.W.J. Granger, Z. Ding, Varieties of long memory models, J. Econometrics 73 (1996) 61–77.
[128] T. Anderson, T. Bollerslev, F.X. Diebold, P. Labys, The distribution of exchange rate volatility, NBER Working Paper WP6961 (1999).
[129] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (1994) 1685–1689.
[130] P.M. Robinson, Gaussian semiparametric estimation of long range dependence, Ann. Statist. 23 (1995) 1630–1661.
[131] W. Feller, An Introduction to Probability Theory and its Applications, Wiley, New York, 1966.
[132] B.B. Mandelbrot, H. Taylor, On the distribution of stock price differences, Oper. Res. 15 (1962) 1057–1062.
[133] D.M. Guillaume, O.V. Pictet, U.A. Muller, M.M. Dacorogna, Unveiling non-linearities through time scale transformations, Olsen group preprint OVP.1994-06-26 (1995), available at `http://www.olsen.ch`.
[134] T. Ane, H. Geman, Order flow, transaction clock, and normality of asset returns, J. Finance (2000), to appear.
[135] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, X. Gabaix, H.E. Stanley, Diffusion and economic fluctuations, Phys. Rev. E 62 (2000) 3023–3026.
[136] P. Gopikrishnan, V. Plerou, X. Gabaix, H.E. Stanley, Statistical properties of share volume traded in financial markets, Phys. Rev. E 62 (2000).
[137] A.R. Gallant, P.E. Rossi, G. Tauchen, Stock-prices and volume, Rev. Financial Stud. 5 (1992) 199–242.
[138] J. Karpoff, Price variability and volume: a review, J. Finance Quantitative Anal. 22 (1987) 109.
[139] H. Markowitz, Portfolio Selection: Efficient Diversification of Investment, Wiley, New York, 1959.
[140] E.J. Elton, M.J. Gruber, Modern Portfolio Theory and Investment Analysis, Wiley, New York, 1995.
[141] M.L. Mehta, Random Matrices, Academic Press, Boston, 1991.
[142] T. Guhr, A. Müller-Groeling, H.A. Weidenmüller, Random-matrix theories in quantum physics: common concepts, Phys. Rep. 299 (1998) 189–425.
[143] E.P. Wigner, On a class of analytic function from the quantum theory of collisions, Ann. Math. 53 (1951) 36–65.
[144] E.P. Wigner, On the statistical distribution of the widths and spacings of nuclear resonance levels, Proc. Cambridge Philos. Soc. 47 (1951) 790–798.
[145] E.P. Wigner, Results and theory of resonance absorption, in: Conference on Neutron Physics by Time-of-flight, Gatlinburg, TN, 1956, Oakridge National Laboratory, pp. 59–70.
[146] M.L. Mehta, F.J. Dyson, Statistical theory of the energy levels of complex systems. V, J. Math. Phys. 4 (1963) 713–719.
[147] F.J. Dyson, The distribution of singular values of certain random matrices, Revista Mexicana Física 20 (1971) 231.
[148] A.M. Sengupta, P.P. Mitra, Distributions of singular values for some random matrices, Phys. Rev. E 60 (1999) 3389–3392.
[149] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (1999) 1469–1482.
[150] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Universal and nonuniversal properties of financial cross-correlation matrices, Phys. Rev. Lett. 83 (1999) 1471–1475.
[151] P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, Identifying business sectors from stock price fluctuations, Phys. Rev. Lett., submitted for publication.
[152] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, T. Guhr, H.E. Stanley, A random matrix approach to financial cross-correlations, Phys. Rev., submitted for publication.