

LÉVY MATRICES AND FINANCIAL COVARIANCES*

ZDZISŁAW BURDA, JERZY JURKIEWICZ, MACIEJ A. NOWAK

M. Smoluchowski Institute of Physics, Jagellonian University
Reymonta 4, 30-059 Kraków, Poland

GÁBOR PAPP

Institute for Theoretical Physics, Eötvös University
Budapest, H-1518 Hungary

AND ISMAIL ZAHED

Department of Physics and Astronomy
SUNY-Stony-Brook, NY 11794, USA*(Received August 18, 2003)*

In a given market, financial covariances capture the intra-stock correlations and can be used to address statistically the bulk nature of the market as a complex system. We provide a statistical analysis of three SP500 covariances with evidence for raw tail distributions. We study the stability of these tails against reshuffling for the SP500 data and show that the covariance with the strongest tails is robust, with a spectral density in remarkable agreement with random Lévy matrix theory. We study the inverse participation ratio for the three covariances. The strong localization observed at both ends of the spectral density is analogous to the localization exhibited in the random Lévy matrix ensemble. We discuss two competitive mechanisms responsible for the occurrence of an extensive and delocalized eigenvalue at the edge of the spectrum: (a) the Lévy character of the entries of the correlation matrix and (b) a sort of off-diagonal order induced by underlying inter-stock correlations. (b) can be destroyed by reshuffling, while (a) cannot. We show that the stocks with the largest scattering are the least susceptible to correlations, and likely candidates for the localized states. We introduce a simple model for price fluctuations which captures behavior of the SP500 covariances. It may be of importance for assets diversification.

PACS numbers: 89.65.Gh, 02.10.Yn

* Presented at the Workshop on Random Geometry, Kraków, Poland, May 15–17, 2003.

1. Introduction

A number of phenomena in nature are characterized by a coexistence of different scales, usually described by power law distributions. This is the case of most phase transitions where at the critical point the correlation functions are scale invariant, as well as most fluid phases in highly developed turbulence where the velocity fluctuations are sensitive to a variety of eddies. Power law distributions are also encountered in a number of biophysical settings as well as financial markets [1].

Stable random phenomena with power law behaviors are usually described by Lévy distributions, a consequence of the central limit theorem for scale-free processes. The simplest example is a random walk with a power law distribution for single independent steps, where the relative probabilities at different times are scale free. These phenomena lead to anomalous diffusion and intermittency as encountered in charge transport in amorphous semiconductors, moving interfaces in porous media, spin glasses, turbulence [2] and phase changes in chiral QCD [3].

Recently, it was pointed out that current market covariances are Gaussian noise driven with possible consequences for the assessment of correlations in portfolio evolution and optimization [4]. In particular, it was shown that the lower part of eigenvalue distribution of the SP500 covariance matrix constructed from the daily returns normalized by the local volatility, is Gaussian noise dominated. In this paper we confirm some of these observations, but suggest that an alternative covariance constructed from the daily returns normalized to the initial price displays Lévy noise throughout the spectrum. The latter is more robust against reshuffling and certainly requires a random Lévy matrix description. This observation is overall consistent with a recent observation we made in the context of free random Lévy matrices [6].

The outline of the paper is as follows: in Section 2, we introduce the concept of financial correlation matrices, and empirically analyze their statistical content. We show that all covariances display power law tails, albeit with different indices. In Section 3 we discuss the issue of inter-stock correlations and we define reshuffling of the price series and investigate its effects on the covariance matrices. We find that the covariance with the largest tails is the least sensitive to this process. This is discussed in Section 4, where we analyze the corresponding spectral densities and show that the results of random Lévy matrices apply remarkably well to the covariance that is stable under reshuffling. In Section 5, we analyze the bulk eigenvector content of the all covariances and parallel them with the results of reshuffling and Lévy random matrix theory. The larger the tails, the stronger the localization

seen in the participation ratios and the stock scattering. In Section 6, we formulate a simple model of price fluctuations, which reproduces most of the experimentally observed features of the SP500 covariances. Our conclusions are in Section 7.

2. Financial covariances

One of the central problems in financial investment is the assessment of risk. Standard lore suggests that risk can be reduced through assets diversification, with Markowitz's portfolio analysis as one of the corner-stones in assets allocation and diversification [1]. The key to Markowitz's analysis is the concept of a covariance matrix. In this section we define and empirically analyze the distribution of entries and also correlations as captured in certain SP500 covariance matrices, with comparison to results from random Lévy matrices. Throughout, we will use price return data from the SP500 daily quotations of $N = 406$ stocks over the period of $T + 1 = 1309$ days from 01.01.1991 till 06.03.1996 (ignoring dividends).

2.1. SP500 covariances

Consider the covariance matrix constructed from the raw returns normalized by the initial price:

$$C_{ij} = \frac{1}{T} \sum_{t=1}^T M_{ti} M_{tj} = \frac{1}{T} \sum_{t=1}^T \frac{m_{ti}}{x_{0i}} \frac{m_{tj}}{x_{0j}}. \quad (1)$$

The raw returns m_{ti} of stock i (out of a total of N) at time t , labeled by an integer ($t = 1, \dots, T$) are evaluated at fixed time intervals in a given market as

$$m_{ti} = \delta x_i(t) - \delta \bar{x}_i, \quad (2)$$

where: $\delta x_i(t) = x_i(t + 1) - x_i(t)$. The mean $\delta \bar{x}_i = \sum_t \delta x_i(t)/T$ is subtracted¹.

The choice of the normalization $M_{ti} = m_{ti}/x_{0i}$ to the initial price preserves the nature of the tails and is scale invariant. The use of the relative returns instead of the logarithm of the ratio of the consecutive returns, is motivated by the additive rather than multiplicative character of the price series.

In the following, we will argue that the C covariance matrix is able to reveal a fat-tail nature of the price change fluctuations, contrary to the

¹ In fact, all results presented in the paper would not change almost at all, and would not affect any conclusion, if the mean were not subtracted.

covariances \mathbf{G} and \mathbf{J} , defined below. From the point of view of preserving power law tails, as a normalization one could alternatively use the prices, $x_{\tau i}$, at any random time τ . However, from the calculational point of view involving the integrated return or the portfolio risk, the choice of the initial price is most convenient.

Commonly, the following covariance is used in the analysis:

$$\mathbf{G}_{ij} = \frac{1}{T} \sum_{t=1}^T \frac{m_{ti}}{\sigma_i} \frac{m_{tj}}{\sigma_j}, \quad (3)$$

where now the normalization is given by the volatility (variance) σ_i with $\sigma_i^2 = \sum_t m_{ti}^2/T$. In the presence of fat tails, *i.e.* if the distribution has a power-law behavior $p(\xi) \sim \xi^{-1-\alpha}$, $\alpha < 2$, the variance itself has a fat tail distribution with an index $\alpha/2$ and the average variance does not exist. Obviously, in this case the use of σ as a normalization will bias the analysis. Alternatively, in place of σ_i , one may use the quantity $r_i = \sum_t |m_{ti}|/T$:

$$\mathbf{J}_{ij} = \frac{1}{T} \sum_{t=1}^T \frac{m_{ti}}{r_i} \frac{m_{tj}}{r_j}. \quad (4)$$

For $1 < \alpha < 2$, r_i has a well defined large T limit. However, in practice, for finite T , this normalization, similarly as the one for \mathbf{G} , obscures the effects of large price-change fluctuations.

A common feature of the quantities $M_{ti} = m_{ti}/x_{0i}$ for \mathbf{C} and $M_{ti} = m_{ti}/\sigma_i$, $M_{ti} = m_{ti}/r_i$ for \mathbf{G} and \mathbf{J} , is that they are invariant with respect to change in the monetary unit. They exhibit, however, a different scale behavior for large ξ . The raw cumulative probability $P_{<}(\xi)$ ($P_{>}(\xi)$) defined as a probability that M_{ti} is less (greater) than ξ , calculated for all i and t , is expected to pick up the smallest power of the tail, $\sim \pm A_{\pm} \xi^{-\alpha_{\pm}}$ (for gains/losses), present in the sample. In Fig. 1 we show the cumulative probabilities (a) $P_{<}(\xi)$ and (b) $P_{>}(\xi)$ for \mathbf{C} . The data in the figure are compared with the power laws with: $\alpha_- = 1.78$ and $A_- = 2.4 \cdot 10^{-3}$ and $\alpha_+ = 2.18$ $A_+ = 1.35 \cdot 10^{-3}$. The values are given without errors. The SP500 data set does not allow for an accurate determination of the fit parameters. The given values have a qualitative meaning. The fact that they are close to 2 signals the presence of fat tails of the underlying distribution. Indeed, the presence of fat-tails will be confirmed by the following analysis of the covariance matrices.

Repeating the same for \mathbf{G} we find much thinner tails with the following exponents: $\alpha_- = 3.8$ and $\alpha_+ = 4.5$, and for \mathbf{J} : $\alpha_- = 3.5$ and $\alpha_+ = 4.3$.

Clearly, the normalization to either the variance (volatility), σ_i , or the range r_i tends to affect the raw tail distributions, with a quenching towards the Gaussian distribution. This is expected, since the fluctuations are

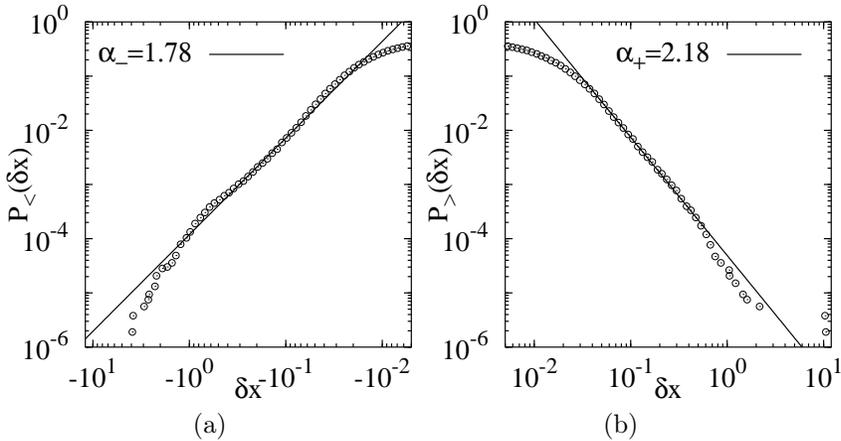


Fig. 1. (a) — The cumulative distribution $P_{<}(\xi)$ for $M_{ti} = m_{ti}/x_{0t}$ as in \mathbf{C} and the best fit to the power law. (b) — The same for $P_{>}(\xi)$.

roughly normalized to the typical fluctuation. This is not the case in \mathbf{C} , where the raw tail is retained.

Let us introduce yet another covariance matrix which will be convenient in the further analysis. We will construct it from the signs $s_{ti} = \text{sgn } m_{ti}$:

$$S_{ij} = \frac{1}{T} \sum_{t=1}^T s_{ti}s_{tj}. \tag{5}$$

We use in our analysis a three-valued sign function: $\text{sgn} = -1, 0, 1$. For all assets the average

$$\langle s_i \rangle = \frac{1}{T} \sum_{t=1}^T s_{ti} \approx 0 \tag{6}$$

and the successive entries in the historically ordered row s_{ti} are essentially uncorrelated.

3. Correlations

By construction, the financial covariance matrix is composed of intra-assets (here stocks) correlations, and therefore tells us how closely assets move in time-evolving market. The microscopic nature of these correlations is so far unknown. However, a quantitative understanding can still be achieved statistically. The source of the correlations is two-fold: real correlations between assets and statistical fluctuations. The statistical fluctuations

disappear in the $T = \infty$ limit. For finite T , however, even in the absence of any real correlations, the non-diagonal entries are non-zero. In the Gaussian universality they fall-off as $1/\sqrt{T}$.

As a measure of correlations between distinct companies $i \neq j$ one can use a distribution $\rho(\mathbf{C}_{ij})$. Similar distributions can be constructed for \mathbf{G} and \mathbf{S} .

Consider first correlations of pure signs. In Fig. 2(a), we show the distribution (solid line) of \mathbf{S}_{ij} 's histogrammed over all pairs $i \neq j$. We observe a strong asymmetry towards positive correlations with a maximum around 0.1, indicating that assets have a tendency to move collectively in the same trend: up or down.

This pronounced asymmetry is present in correlations for all other covariances \mathbf{C} , \mathbf{J} , \mathbf{G} . We show in Fig. 2(b) the distribution of correlations for \mathbf{C} . The correlations between signs are inherited by all other covariances. We shall discuss possible consequences of this behavior in the Section 6. The inter-stock (inter-sign) correlations can be easily destroyed by a procedure of reshuffling described below. Indeed, we see in Fig. 2 (dashed line) that after reshuffling the spectra become symmetric.

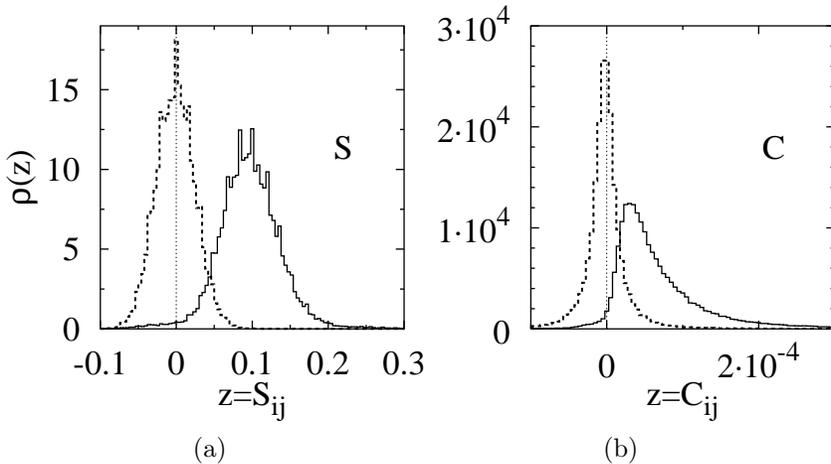


Fig. 2. (a) — The distribution of correlations of signs \mathbf{S} before (solid line) and after reshuffling (dashed line). (b) — The same for \mathbf{C} .

3.1. Reshuffling

Let us introduce the abovementioned procedure to remove the inter-stock correlations from the data. Having done this, we will be able to concentrate on the issue of the stochastic nature of the price fluctuations.

The price changes $\delta x_i(t)$ enter the covariance matrix in the historical order. This order, in particular preserves inter-stock correlations. We can suppress the inter-stock correlations in the data, by introducing a random time ordering to the time history for each asset. More precisely, for each asset, i , we can generate a random permutation of t -indices $P^{(i)} : t \rightarrow t' = P^{(i)}(t)$ and instead of the historical ordered rows of returns we can use: $\delta x'_i(t) = \delta x_i(t')$, to define: $m'_{ii} = \delta x'_i(t) - \delta \bar{x}_i$ and the corresponding covariance matrices \mathbf{C}' , \mathbf{J}' , \mathbf{G}' , and \mathbf{S}' . The permutations for different rows, $P^{(i)}$, $P^{(j)}$, are random and mutually independent. Such a reshuffling does not change the information content of the individual asset rows, because successive entries in the historical ordered row, $\delta x_i(t)$ and $\delta x_i(t + 1)$, are uncorrelated for typical time intervals on a market. Thus, the reshuffling affects only the inter-row information content destroying any correlations. Hence we expect that the reshuffled data set should reflect pure stochastic nature. Indeed, it does. In Fig. 2 we show for example the effect of reshuffling on the inter-stock correlations. The asymmetry disappears.

4. Spectra

In this section we discuss the spectral density associated with the covariance matrices defined above. The spectral density plays an important role in risk assessment [1, 4].

The results will be presented as histograms of eigenvalues λ . We will sort of unify the scale on the λ -axis by plotting histograms as a function of a quantity λ/Γ where Γ is defined as:

$$\Gamma = \frac{1}{N} \text{Tr } \mathbf{C} = \frac{1}{NT} \sum_{ti} M_{ti}^2, \quad (7)$$

and analogously for \mathbf{G} , \mathbf{J} and \mathbf{S} . For a covariance of Gaussian numbers, the constant Γ approaches a T -independent constant in the limit $T \rightarrow \infty$ while for a covariance of power-law distributed numbers with $1 < \alpha < 2$, it behaves as $\Gamma = \Gamma T^{2/\alpha-1}$, where Γ is a T -independent constant. We will explain this scaling in more detail in the section about random Lévy matrices. It is easy to see that by construction $\Gamma = 1$ for \mathbf{G} and \mathbf{S} . The histograms of eigenvalues λ/Γ for the SP500 data are presented in Fig. 3. Let us make here two points: the histograms for \mathbf{G} and \mathbf{S} are almost identical. All spectra have a few large eigenvalues. In the next section using Random Gaussian Matrices (RGM) we will define the scale which will tell us which eigenvalues can be treated as large.

There are potentially two sources of large eigenvalues in the spectrum: inter-stock correlations and fat-tails. In the next sections we will discuss methods to pinpoint the two effects. In the right block of Fig. 4 we plot

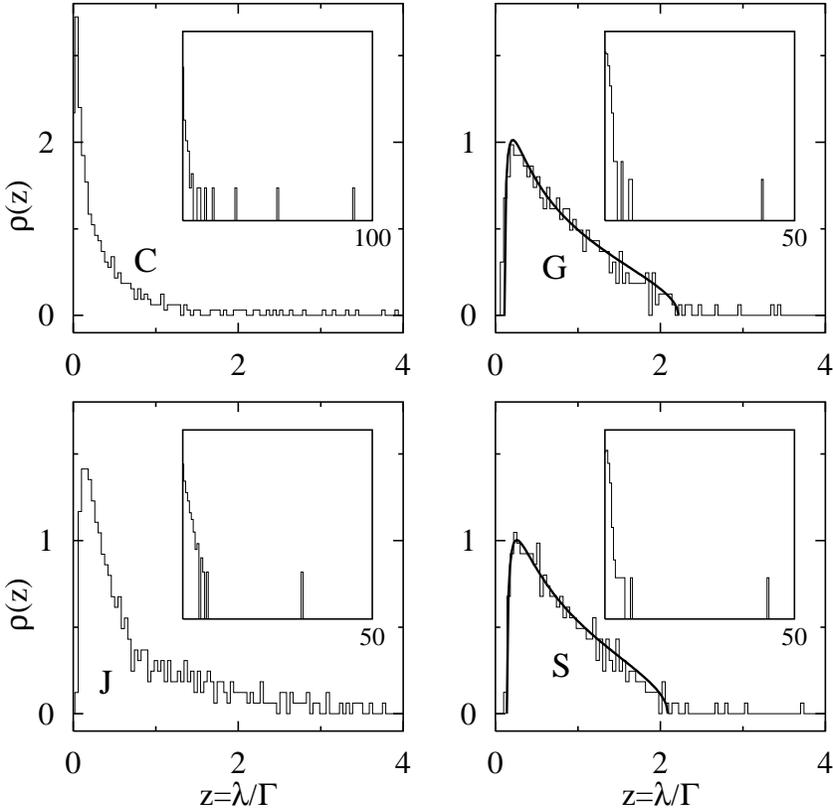


Fig. 3. The eigenvalue histograms for C, G, J and S . For G and S we present also the Gaussian fit (8). In the inlets we show the same histograms but in a range embracing all eigenvalues. To make the smallest picks visible we artificially enhanced them by setting in inlets the logarithmic scale on the vertical axis. Notice that except the one for C all plots have the same ranges.

the eigenvalue density of the G' and S' covariances. The spectra for G' and S' are again almost identical. In comparison with the spectra for the historically ordered data set, right block of Fig. 3, we see that the large eigenvalues disappear: the largest eigenvalue of G was 41.95, and for G' after reshuffling 2.77. Moreover, the G' and S' spectra fit very well to the curve:

$$\rho(\lambda) \sim \frac{1}{\lambda} \sqrt{(\lambda - \lambda_{\min})(\lambda_{\max} - \lambda)} \tag{8}$$

with $\lambda_{\min} = 0.20$, $\lambda_{\max} = 2.43$, predicted by Random Gaussian Matrices (RGM) in the large N limit [7]. In general for the asymmetry parameter

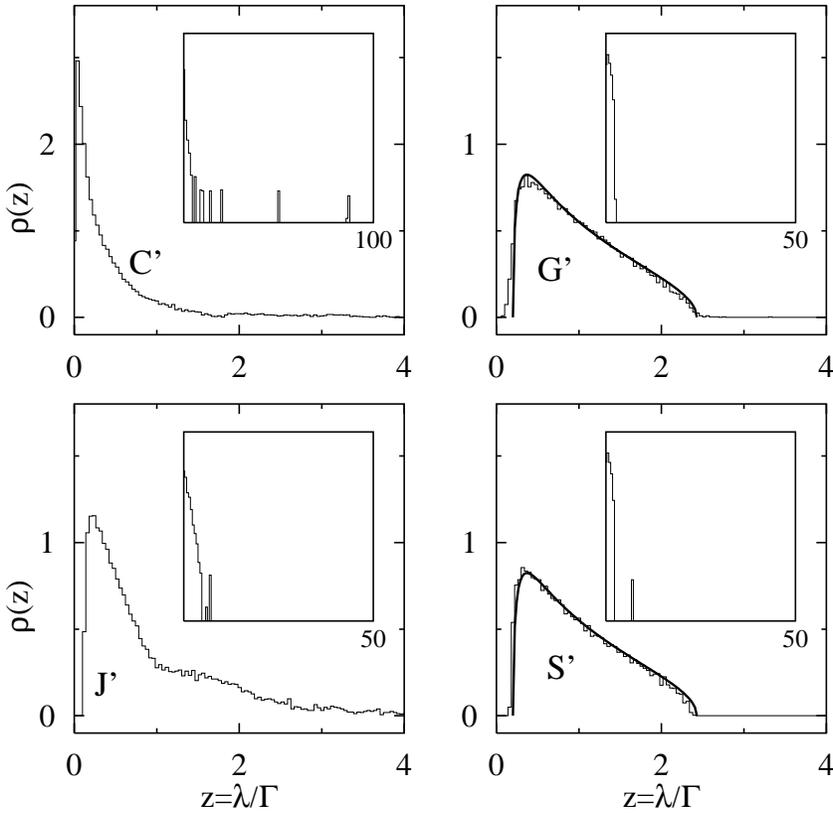


Fig. 4. The same as in Fig. 3 but for reshuffled data set. The histograms are averaged over 20 random reshufflings.

$a = T/N$, the formula predicts $\lambda_{\max,\min} = (1 \pm 1/\sqrt{a})^2$, which in particular for $a = T/N = 1308/406 \approx 3.22$, yields the values given above. These values are used in the curve plotted in Fig. 4. This agreement clearly indicates that the normalization to the volatility σ_i used in \mathbf{G} brings the signal to the Gaussian universality.

It is worth noting that if one attempts to fit the RGM result to the eigenvalue histograms for a correlated data set one generally obtains values for λ_{\min} , λ_{\max} which deviate from the predicted ones. For example, for \mathbf{G} and \mathbf{S} for the historically ordered we get $\lambda_{\max} = 2.22$, $\lambda_{\min} = 0.11$ and $\lambda_{\max} = 2.09$, $\lambda_{\min} = 0.14$, respectively, This deviation comes as a compensation for the appearance of large eigenvalues which lie far away and corresponds to asymmetry parameters $a = 2.5$ and $a = 2.9$, respectively.

As shown in Fig. 4, the large eigenvalues survive reshuffling in \mathbf{C} covariances. Moreover, the position of the larger eigenvalues is relatively stable

under reshuffling. In fact, in each of 20 random reshufflings used in the plot, the three larger eigenvalues normalized by the scale Γ , always land in the same histogram bin of the size 0.015.

The fact that there are large eigenvalues in the reshuffled spectrum can be attributed to the presence of heavy tails in the probability distributions for price fluctuations. The behavior of the spectra can be understood, as we will see below, in terms of Random Lévy Matrix (RLM) theory.

4.1. \mathbf{C} versus RLM

We want to compare eigenvalue spectra of \mathbf{C} and \mathbf{C}' with the ones obtained from an ensemble of Random Lévy Matrices (RLM) [8]. We generate Lévy matrices (RLM) as follows. We choose $N \times T$ elements of a matrix M_{ti} as independent random numbers from a Lévy distribution. We find the eigenvalues of the matrix $C_{ij} = 1/T \sum_t M_{ti} M_{tj}$. We can repeat this many times collecting the eigenvalues in a common histogram.

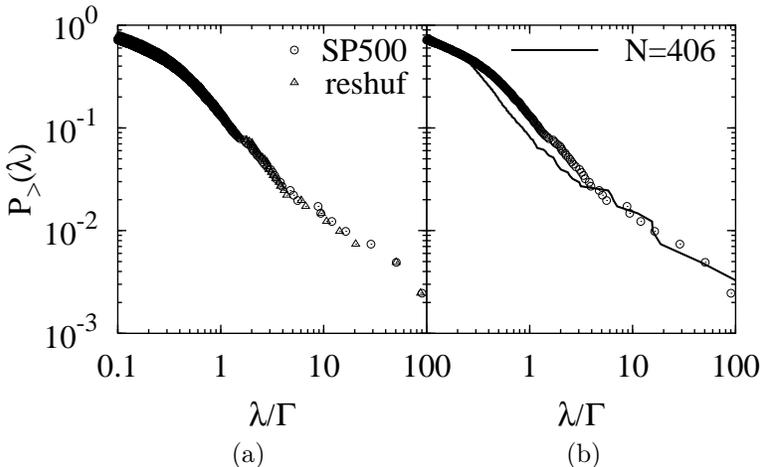


Fig. 5. (a) — Cumulative eigenvalue distribution for the SP500 data for the historically ordered data set compared with the one for reshuffled data and (b) — with the one of a randomly generated Lévy matrix.

Here we would like to mention, that there exists an alternative construction of random matrix ensembles, based on the concept of free random variables [9]. We call this realization Free Lévy Matrices (FLM) [5]. Ensembles of FLM, are, contrary to RLM, invariant under rotations. They are more easily tractable using analytical methods. On the other side, due to the invariance of the measure, the eigenvectors do not show interesting cor-

relations like in the case of RLM, which we discuss in the next sections. In the rest of this paper, we do not discuss FLM, and we refer for comparison between the FLM and RLM ensembles to [6].

As an illustration, in our numerical experiment we generated RLM with an asymmetry $a = T/N = 3.22$ adjusted to the asymmetry of the SP500 data set being presently considered. The choice of the asymmetry will become clear in the next subsection. As the distribution for the matrix elements M_{ti} we took a symmetric Lévy distribution with $\alpha = 1.7$ close to the value emerging from the analysis of the tail behavior of the cumulative probability $P_{<}(\xi)$.

In Fig. 5 we compare the cumulative distributions of eigenvalues for \mathbf{C} and \mathbf{C}' , and the random matrix result for fixed asymmetry $a = 3.22$ and size $N = 406$. We see that the large eigenvalues in the spectrum \mathbf{C} survive reshuffling. The spectrum of RLM and of the reshuffled SP500 data set exhibit similar large eigenvalue behavior.

4.2. Scaling in RLM

Assume that M_{ti} are power-law distributed: $p(\xi) \sim \xi^{-1-\alpha}$. Define:

$$\mathbf{C}_{ij} = \frac{1}{T^\sigma} \sum_t M_{ti} M_{tj} \tag{9}$$

with the normalization factor $1/T^\sigma$ whose exponent σ may differ from 1. We will argue that a natural candidate for σ is $\sigma = 2/\alpha$.

Let us split \mathbf{C} into a diagonal \mathbf{D} and off-diagonal \mathbf{A} parts

$$\mathbf{C}_{ij} = \mathbf{D}_i \delta_{ij} + \mathbf{A}_{ij}. \tag{10}$$

We can use the Central Limit Theorem for Lévy universality to obtain the distribution of the entries in \mathbf{C} in the large T limit. We get

$$\mathbf{D}_i \sim T^{2/\alpha-\sigma} d_i, \quad \mathbf{A}_{ij} \sim T^{1/\alpha-\sigma} a_{ij}, \tag{11}$$

where d_i and a_{ij} are T independent constants, distributed with the stable Lévy distributions. For the diagonal elements d_i we expect a distribution with an index $\alpha/2$ and the skewness parameter $\beta = 1$, while the off-diagonal elements a_{ij} will be distributed with a symmetric distribution with an index α . To assess the importance of the off-diagonal entries on the spectrum, we use the standard perturbation theory. For that, we write

$$\mathbf{C}_{ij} = T^{2/\alpha-\sigma} c_{ij} = T^{2/\alpha-\sigma} \left(d_i \delta_{ij} + T^{-1/\alpha} a_{ij} \right) \tag{12}$$

and expand $c_{ij} = d_i \delta_{ij} + \epsilon a_{ij}$ in $\epsilon = 1/T^{1/\alpha}$. In zeroth order the eigenvalues of \mathbf{C}_{ij} are just d_i . The first order corrections are zero because the matrix \mathbf{A}_{ij}

is off-diagonal. Generically, for a random matrix, d_i 's are not degenerate, so up to the second order, the eigenvalues of \mathbf{C}_{ij} are

$$\lambda_i = d_i + \epsilon^2 \sum_{j(\neq i)} \frac{a_{ij}^2}{d_j - d_i} = d_i + T^{-2/\alpha} \sum_{j(\neq i)} \frac{a_{ij}^2}{d_j - d_i}. \tag{13}$$

There are $N - 1$ terms in the sum, each of order unity. Thus the sum contributes a factor proportional to N , say $\approx s_i N$, and we have:

$$\lambda_i = d_i + s_i N T^{-2/\alpha}. \tag{14}$$

The off-diagonal terms compete with the diagonal ones for $N \approx T^{2/\alpha}$. In our case, $\alpha = 1.7$, $N/T^{2/\alpha} \approx 1/3$. The range of the spectrum of \mathbf{C} will not grow with T for $\sigma = 2/\alpha$.

The normalization constant $\Gamma = \langle \text{Tr } \mathbf{C} \rangle / N$, which we have introduced previously for the experimental covariances, behaves for RLM with $1 < \alpha < 2$ as $\Gamma = \Gamma T^{2/\alpha - \sigma}$. Again it simplifies for the choice $\sigma = 2/\alpha$.

5. States

In this section we analyze the eigenvector content of the three covariances using the inverse participation ratios and the stock scattering. We show that the covariances with larger tails are more stable under reshuffling the SP500 data, with localized states at the edge of the spectrum.

5.1. Inverse participation ratios

To better understand the nature of large-eigenvalues in the SP500 data, we now turn to the eigenvector content. For that we use the inverse participation ratio

$$\mathbf{Y}_\lambda = \sum_{i=1}^N V_{\lambda i}^4, \tag{15}$$

where V_λ is a normalized ² eigenvector:

$$\sum_{i=1}^N V_{\lambda i}^2 = 1 \tag{16}$$

of \mathbf{C} to the eigenvalue λ . We can distinguish the ‘mixed’ states with $\mathbf{Y}_\lambda \approx 1/N \approx 0$ and ‘pure’ states with $\mathbf{Y}_\lambda \approx 1$.

² This implicitly assumes that the entries $V_{\lambda i}$ are at least power law distributed with index $\alpha > 1$.

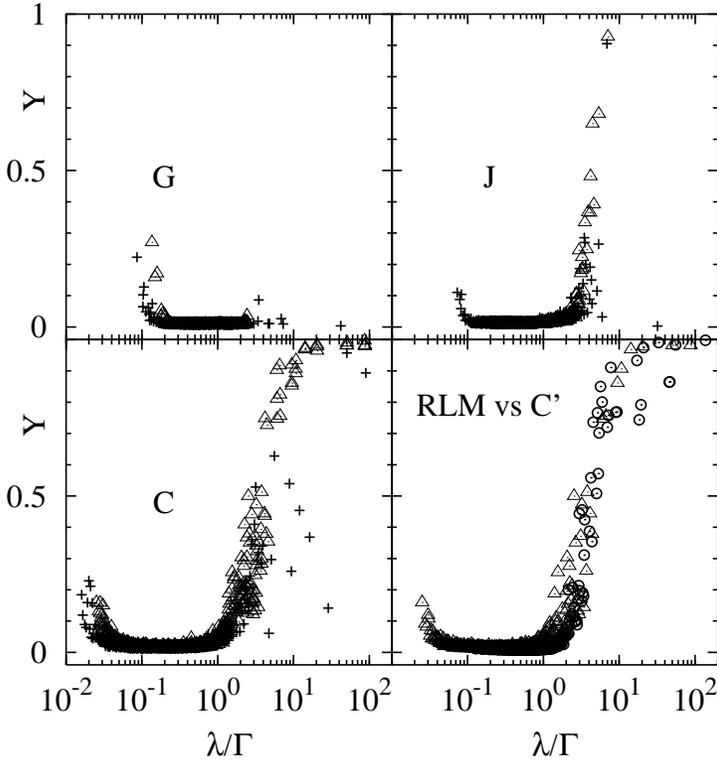


Fig. 6. Inverse participation ratios for for the SP500 covariances, for the historically ordered data (crosses) and for the reshuffled data (triangles). In the lower right figure we compare the distribution of the reshuffled data (triangles) and of Random Lévy Matrix (circles).

In Figs. 6 we display the inverse participation ratios for the three covariances C , G and J for the raw SP500 data (triangles) and reshuffled SP500 data (pluses). Additionally, in the fourth insert we compare the inverse participation ratio for C' (triangles) and for RLM with index $\alpha = 1.5$. The large eigenvalues are localized for J and C with intermediate and large participation ratio, respectively. For G and J the largest eigenvalue states are ‘mixed’ while for C they are ‘pure’. Large eigenvalue pure states are present also in the C' covariance! This is clearly displayed by the data. The inverse participation ratio as a function of eigenvalue for C' has the same character as RLM.

The Lévy randomness has an equally strong effect on the large part of the distribution as the inter-stock correlations. Also, in the G and J covariances reshuffling ‘breaks’ the clustering, removing from the spectrum large eigenvalues.

Concerning the lower part of the spectrum, it is interesting to note that it is characterized by smaller participation ratios than the large eigenvalue part. The exception is \mathbf{G} for which the part with small eigenvalues has larger inverse participation ratio. On the other hand, the low eigenvalue parts of the \mathbf{G} and \mathbf{C} show a similar behavior. In fact, one expects that the shape of the lower part of the spectra for RLM and RGM to be strongly related to the asymmetry of the matrices rather than to the type of randomness. Indeed, for matrices with asymmetry $a < 1$, the spectra exhibit exact zero eigenvalue states (zero modes).

5.2. Stock scattering

In analogy to the inverse participation ratio we define a quantity:

$$\mathbf{P}_i = \sum_{\lambda} V_{\lambda i}^4 \quad (17)$$

which measures how many eigenstates are mixed in a pure stock state i . We will refer to it as the stock scattering. Again we have the normalization:

$$\sum_{\lambda} V_{\lambda i}^2 = 1. \quad (18)$$

The inverse $1/\mathbf{P}_i$ tells us how many eigenstates are influenced by the stock i .

In Fig. 7 we show the value of \mathbf{P}_i for the consecutive 406 stocks (horizontal) with $0 \leq \mathbf{P}_i \leq 1$ (vertical), for each of the three covariances discussed

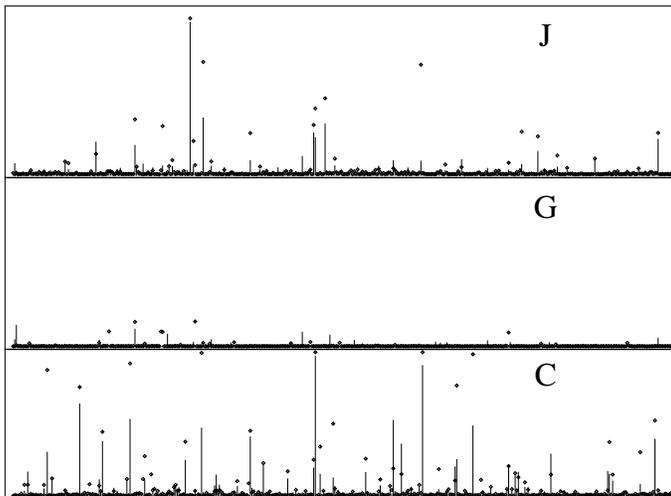


Fig. 7. Stock scattering for the SP500 covariances for the historically ordered data (lines) and the reshuffled data (dots).

in this work, at the same scale to facilitate the comparison. Clearly, the normalization to unit volatility drives \mathbf{P}_i towards Gaussian noise. We have checked that the effects of reshuffling is to enhance the localization of certain stocks (without much affecting the original ones), in agreement with the participation ratio analysis. The stock scattering of \mathbf{C} provides a relatively simple filter for those stocks that localize in a market, and are likely to drive the large tail behavior of the covariance matrix.

6. Model

Collecting all experimental evidence for the SP500 data we are led to the conjecture, that the returns normalized to the initial price $M_{ti} = m_{ti}/x_{0i}$ undergo fluctuations which can be well described by a randomness of the type:

$$M_{ti} = \text{sgn}_{ti} \cdot \xi_{it}, \quad (19)$$

where sgn_{ti} is a random matrix of correlated signs and ξ_{it} are identical independent distributed numbers from the Lévy universality. Indeed, as shown in Fig. 2 the sign correlations are present in all covariances. Additionally, one can check that the substitution of signs of $\delta x_i(t)$ by random signs has the same effect on spectra of the SP500 covariances as reshuffling.

Moreover, the spectra of eigenvalues of \mathbf{G} and \mathbf{S} are almost identical and can be seen from the Fig. 3 and from the comparison of a few largest eigenvalues which are 6.89, 7.27 and 41.95 for \mathbf{G} and 5.48, 7.43 and 43.25 for \mathbf{S} . This tells us that the information about the inter-stock correlations present in \mathbf{G} is already present in \mathbf{S} . Thus, as long as the long tails are suppressed, as in \mathbf{G} , the absolute value of the fluctuations does not matter.

The absolute value of the changes matters, however, if we do not introduce any superfluous normalizations and expose the covariance³ to large price fluctuations as in \mathbf{C} . In this case, the correlations of signs play a secondary role, since even when one decorrelates them by reshuffling, the large eigenvalues stay in the spectrum.

Similarly as under reshuffling, the spectrum of \mathbf{C} does not change its character and the largest eigenvalues are quite stable, if one uses randomly generated signs instead of those inherited from the historical data.

The randomness given by the formula (19) captures many experimentally observed features of the real data. It should be treated, however, as zeroth order approximation. In a more involved analysis, one should introduce

³ One can define other covariances, like for example, the one constructed from instantaneous returns: $M_{ti} = m_{ti}/x_{ti}$, or $M_{ti} = \log \delta x_i(t+1)/\delta x_i(t)$, which similarly to \mathbf{C} , would preserve information about the power-law tails. As a consequence, for instance, their spectra would behave in the same way under the reshuffling as the one of \mathbf{C} .

some corrections to the conjecture which for example can take into account the possibility of correlations between sign and absolute value of the price changes. Indeed, the data show the existence of such correlations.

7. Conclusions

We have carried an empirical analysis of the covariance matrices characterizing daily price returns from the SP500 market. We have shown that a specific covariance (returns normalized to the stock initial price) exhibits matrix entries with almost stable Lévy tails. A comparative study shows that only this covariance is stable under reshuffling, with a spectrum in remarkable agreement with the one extracted from an ensemble of random Lévy matrices with commensurate sizes and asymmetry. An analysis of the corresponding participation ratio shows large localized and almost ‘pure’ states. This is not the case of the other covariances (returns normalized to the stock mean volatility or range), which are characterized by ‘mixed’ states with one characteristically large and delocalized eigenvalue reminiscent of Yang’s ODLRO [10]. The stock content of the localized states is best displayed using the stock scattering.

In nearly Gaussian markets, the risk is usually assessed by minimizing the variance of a pertinent market policy, say an investment portfolio, using the empirical market covariance as suggested by Markowitz [11]. Recently, it was pointed out that the low-lying eigenvalues of the empirical market covariance are Gaussian noise dominated (information free), implying that standard Markowitz’s theory for risk assessment is flawed [4]. In non-Gaussian markets, the potential for large asset fluctuations may require using an alternative to Markowitz’s theory through the use of value-at-risk or tail-covariance [1], each of which requiring the covariance matrix.

In the present work we have shown that eigenvalues of a market covariance follow the theoretical distribution of eigenvalues of almost randomly generated Lévy matrices. The empirical market covariance reflects on a state of maximum entropy in the generalized sense of Dyson for random Lévy matrices. Our observations may be relevant for assets diversification and risk management.

This work was supported in part by the US DOE DE-FG02-88ER40388, the Polish State Committee for Scientific Research (KBN) 2 P03B 096 22 (2002-2004), the Hungarian FKFP 220/2000, and by the EC IHP HPRN-CT-1999-00161 grants.

REFERENCES

- [1] R. Mantegna, H. Stanley, *An Introduction to Econophysics*, Cambridge University, 2000; J. Bouchaud, M. Potters, *Theory of Financial Risks*, Cambridge University, 2000.
- [2] For a review, see *e.g.* *Lévy Flights and Related Topics*, Eds. M. Shlesinger, G. Zaslavsky, U. Frisch, Springer, 1995.
- [3] R.A. Janik, M.A. Nowak, G. Papp, I. Zahed, *Acta Phys. Pol. B* **28**, 2949 (1997).
- [4] L. Laloux, P. Cizeau, J. Bouchaud, M. Potters, *Phys. Rev. Lett.* **83**, 1467 (1999); V. Plerou, P. Gopikrishnan, B. Rosenow, L. Nunes Amaral, H. Stanley, *Phys. Rev. Lett.* **83**, 1471 (1999). For a recent applications, see *e.g.* P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, *Phys. Rev.* **E64**, 035106R (2001); S. Drożdż, J. Kwapien, F. Gruemmer, F. Ruf, J. Speth, *Physica A* **299**, 144 (2001) and references therein.
- [5] Z. Burda, R.A. Janik, J. Jurkiewicz, M.A. Nowak, G. Papp, I. Zahed, *Phys. Rev.* **E65**, 21106 (2002).
- [6] Z. Burda, J. Jurkiewicz, M.A. Nowak, G. Papp, I. Zahed, *cond-mat/0103109*.
- [7] This formula was obtained by several authors using various techniques, see *e.g.* A. Crisanti, H. Sompolinsky, *Phys. Rev.* **A36**, 4922 (1987); A. Edelman, *SIAM J. Matrix Anal. Appl.* **9**, 543 (1988); M. Opper, *Europhys. Lett.* **8**, 389 (1989); J. Feinberg, A. Zee, *cond-mat/9609190*; R.A. Janik, M.A. Nowak, G. Papp, J. Wambach, I. Zahed, *Phys. Rev.* **E55**, 4100 (1997); A.M. Sengupta, P.P. Mitra, *Phys. Rev.* **E60**, 3389 (1999).
- [8] P. Cizeau, J.P. Bouchaud, *Phys. Rev.* **E50**, 1810 (1994).
- [9] D.V. Voiculescu, *Invent. Math.* **104**, 201 (1991); D.V. Voiculescu, K.J. Dykema, A. Nica, *Free Random Variables*, Am. Math. Soc., Providence, RI (1992); for new results see also A. Nica, R. Speicher, *Amer. J. Math.* **118**, 799 (1996); H. Bercovici, D. Voiculescu, *Ind. Univ. Math. J.* **42**, 733 (1993).
- [10] C.N. Yang, *Rev. Mod. Phys.* **34**, 694 (1962).
- [11] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, J. Wiley and Sons, New York 1959.