Vol. 36 (2005) *ACTA PHYSICA POLONICA B* No 8

STATISTICAL PROPERTIES OF STOCK MARKET EIGENSIGNALS*

J. Kwapień[a], S. Drożdż[a,b] and P. Oświęcimka[a]

[a]H. Niewodniczański Institute of Nuclear Physics, Polish Academy of Sciences
Radzikowskiego 152, 31-342 Kraków, Poland
[b]Institute of Physics, University of Rzeszów
Rejtana 16a, 35-959 Rzeszów, Poland

(Received April 19, 2005)

By using the correlation matrix approach, we decompose the evolution of a set of the 100 largest American companies into the components (portfolios) defined by the eigenvectors of the correlation matrix. Among the results, we show that a number of the non-random components exceeds the previous estimates based on much shorter time series of daily returns. This indicates that for short signals the bulk of random eigenvalues defined by Random Matrix Theory can comprise also a significant amount of information. We also show that the components corresponding to a few largest eigenvalues and describing the most collective part of the market evolution reveal strong nonlinear correlation structure in contrast to the other components. All the components are multifractal. Moreover, by using a modified definition of the correlation matrix, we are able to decompose the daily pattern of the German DAX30 index into components which can characterize the recurrent events occurring at precise moments of a trading day.

PACS numbers: 89.75.–k, 89.75.Da, 89.75.Fb, 89.65.Gh

1. Introduction

Temporal evolution of asset prices in a stock market, although extremely noisy [1], reveals significant cross-asset correlations, leading to formation of a hierarchy of stable asset clusters emerging out of noise [2,3]. This of course effectively reduces the number of degrees of freedom carrying information. Based on the correlation structure of the market, different theories like the Markowitz optimal portfolio theory [4] and Arbitrage Pricing Theory [5] were

* Presented at the First Polish Symposium on Econo- and Sociophysics, Warsaw, Poland, November 19–20, 2004.

developed providing the necessary mathematical basis either for the portfolio selection or for uncovering the cluster structure of the market. Those theories apply the correlation matrix formalism and allow an investor to decompose the market into a number of independent asset subsets, each corresponding to a particular correlation matrix eigenvalue; the associated eigenvectors define the potential investing portfolios. Despite the fact that in accordance with the conventional portfolio theories the portfolios associated with all the eigenvalues may be used and their information content is equivalent, the most modern approach based on Random Matrix Theory (RMT) [6] levers this assumption and groups the eigenvalues into the ones carrying genuine information about the market and the noisy ones sharing only the universal properties of the Wishart random matrix ensemble (thus unrelated to the market). Consequently, the portfolios corresponding to the "noisy" eigenvalues cannot be considered as the valid ones.

Time evolution of a portfolio can be expressed in terms of the evolution of its returns at a specific time scale. This allows one to investigate the factors of the market dynamics by decomposing the market into the components (eigensignals) that can reveal different properties. However, as the financial markets are characterized not only by the cross-asset correlations but also the temporal autocorrelations (memory), the construction of the correlation matrix from signals corresponding to different assets is not the only possibility. One can also study the market dynamics by constructing the correlation matrix from a set of mutually equivalent signals related to the same asset or to the same market index, but corresponding to different intervals of time (for example, different trading days [7]). This kind of analysis focuses entirely on the temporal correlations at a specific time scale and, by calculating the eigenvalue–eigenvector structure of the matrix and then the relevant eigensignals, it allows one to search for the existence of repeatable structures or other events which occur approximately periodically in time. Such temporally oriented approach is interesting due to the fact that there is some controversy in how to account for the so-called financial stylized facts (the fat-tailed p.d.f.'s of the stock price fluctuations at different time scales, the long-lasting memory in volatility, *etc.* [8]) that have their origin in the temporal structure of the market. The principal objective of the present work is to analyze the statistical characteristics of the market eigensignals. In the next two sections we describe the basic formalism of the correlation matrix approach and present results of our data analysis, in which we concentrate on the identification of the non-random components of the market by applying both variants of the correlation matrix described above.

2. Methodology

Let us consider a set of N time series (indexed by s) of normalized logarithmic returns $\{g_s(i)\}_{i=1,\dots,T}$ at a fixed time scale Δt, which represent different assets or different intervals of time *e.g.* different trading days (in the latter case T denotes the number of the returns in a trading day). From these time series we form an $N \times T$ data matrix $\boldsymbol{M}$ and then we calculate the correlation matrix $\boldsymbol{C} = (1/T)\boldsymbol{M}\boldsymbol{M}^T$. By diagonalizing $\boldsymbol{C}$, we obtain the set of eigenvalues λ_j and the corresponding eigenvectors $\boldsymbol{v}_j$

$$\boldsymbol{C}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j, \quad j = 1, \dots, N\,. \tag{1}$$

In the case of the correlation matrix constructed from different assets, each eigenvector $\boldsymbol{v}_j$ can be considered a representation of an N-asset portfolio P_j with the weights equal to the eigenvector components $\nu_j^{(k)}$. The ith portfolio return can be expressed by

$$z_j(i) = \sum_{k=1}^{N} \nu_j^{(k)} g_k(i), \quad i = 1, \dots, T\,. \tag{2}$$

For a non-degenerate correlation matrix, P_i and P_j are independent for each pair of their indices. The portfolios can be associated with the time series of their returns $Z_j \equiv \{z_j(i)\}_{i=1,\dots,T}$ that we shall call eigensignals. In the case of the correlation matrix derived from time series representing different time intervals, the analogously-defined eigensignals can be interpreted as the components associated with repeatable structures in the price or the index fluctuations.

3. Results

Our study was entirely based on high-frequency data from the American (NYSE and NASDAQ) and the German (Deutsche Börse) stock market. For the study of the cross-asset correlations we used time series of the stock returns of $N = 100$ highly-capitalized American companies (capitalization $> 10^{10}$\$) spanning a time interval between 1 Dec 1997 and 31 Dec 1999. We carried out our analysis at a time scale of $\Delta t = 5$ min ($T = 40638$); at this time scale the correlations between the highly-capitalized stocks are already significant. As our previous study showed [9], the correlation structure among the stocks of the largest companies is well-developed at time scales shorter than 1 day. Figure 1 shows the functional dependence $\lambda_1(\Delta t)$ for the 100 stocks under study; λ_1 describes the most collective component reaching for $\Delta t = 5$ min over 60% of its saturation level. The second part of our analysis, involving the temporal correlation matrix, was performed on the

DAX30 index recorded during the trading hours 8:45–17:00 with $\Delta t = 15$s frequency and covering the interval 1 Dec 1997–17 Sep 1999 ($N = 451$ trading days of length $T = 1980$ returns); the opening 15 min (8:30–8:45) were omitted because of an undefined index value.

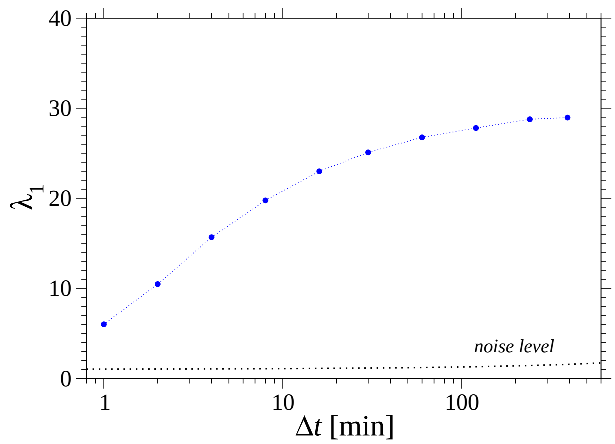

Fig. 1. $\lambda_1(\Delta t)$ for a set of 100 stocks representing the companies of highest capitalization; $\Delta t = 390$ min corresponds to daily time scale.

3.1. Cross-asset correlations

The first issue we shall discuss is the eigenvalue–eigenvector structure of $\boldsymbol{C}$, because it can be straightforwardly compared with the universal predictions of RMT. Theory defines the lower and upper bounds for the spectra of the sample correlation matrices (Wishart matrices), calculated from the random signals with a Gaussian distribution of their values [10]

$$\lambda_{\min}^{\max} = 1 + \frac{1}{Q} \pm \frac{2}{\sqrt{Q}}, \tag{3}$$

where $Q = T/N > 1$. In our case $Q = 406$ and thus the RMT spectrum is very narrow as the shaded region in figure 2 exhibits.

The largest eigenvalue λ_1 in figure 2(a), commonly identified with the market factor is significantly repelled from the rest of the spectrum and describes the average behavior of the whole market (its magnitude $\lambda_1 \simeq 18$ can be compared with a "rigid" market with $\lambda_1 = 100$). The majority of λ_j's escapes, however, the RMT region in both directions. A much better agreement with the theoretical predictions for the random correlations can be reached after removing the market factor from the data by means of the least square fitting of Z_1 to each of the signals $\{g_s\}$: $g_s(i) = \alpha_s + \beta_s z_1(i) + \epsilon_s(i)$, where α, β are free parameters. A new correlation matrix $\boldsymbol{C}'$ can be

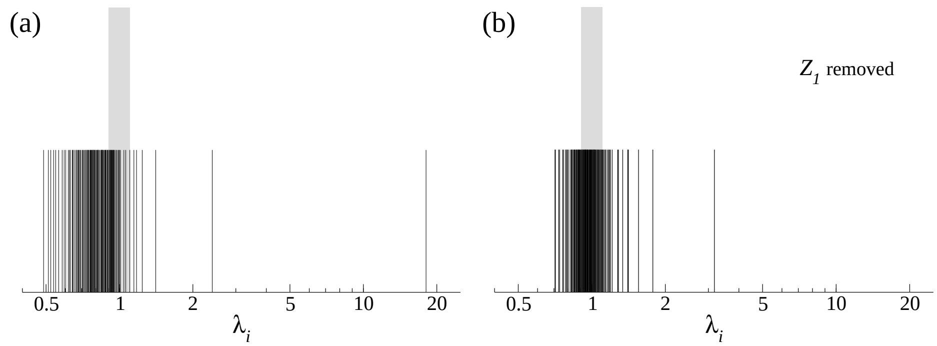

Fig. 2. (a) Eigenvalue spectrum of the correlation matrix $\boldsymbol{C}$ (vertical lines), calculated for the 100 highly capitalized American companies; the shaded vertical region comprises the eigenvalues of a random Wishart matrix with the same Q. (b) Eigenvalue spectrum after effective rank reduction of $\boldsymbol{C}$, *i.e.* after subtracting the contribution of the most collective component Z_1.

then evaluated from the time series of residuals $\epsilon_s(i)$ [1,3]. Figure 2(b) shows the so-decollectified eigenspectrum with $N-1$ non-zero eigenvalues. Now almost half of the total number of λ_j's fall within the interval $(\lambda_{\min}, \lambda_{\max})$. If one compares this fraction of RMT-like eigenvalues with the results of other studies (*e.g.* [1] where this fraction was 94%), one realizes that in our case much more non-random correlations exist. This is due to the fact that we construct $\boldsymbol{C}$ from long time series, which for a fixed N leads to a high value of Q and a narrow RMT spectrum, while in the other analyses much shorter signals (and thus a smaller Q and a wider interval $(\lambda_{\min}, \lambda_{\max})$) were used. Short signals enhance noise and make the identification of the weak non-random correlations impossible [9]. This observation suggests that the market evolution comprises more informative components than it might seem from the studies based on short signals recorded *e.g.* at the daily time scale that is typically used in such analyses.

Another quantity which can serve as a tool allowing us to differentiate between the random and the non-random eigenstates is the distribution of the eigenvector components $\nu_j^{(k)}$ and the related inverse participation ratio (IPR). For a random matrix, the eigenvector components are distributed according to the Gaussian distribution, while in the opposite case various deviations from this distribution occur, for instance a localization (a few large components dominate the p.d.f.) or a collective delocalization (the components fluctuate around the average non-zero value). In the left panel of figure 3 p.d.f.'s of the eigenvector components associated with λ_1 and λ_2 show signatures of delocalization. Almost all the stocks participate in $\boldsymbol{v}_1$ which indeed represents the market factor. This is better visualized in the

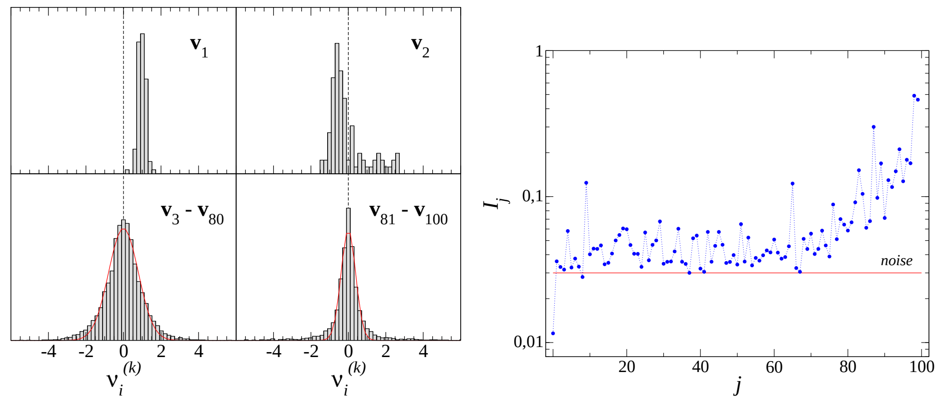

Fig. 3. (Left) Histograms of eigenvector components $\nu_j^{(k)}$ for different eigenvectors $\boldsymbol{v}_j$. A Gaussian is fitted to the empirical histograms in bottom panels. Vertical scale is different in each panel. (Right) Inverse participation ratio for the correlation matrix eigenvectors. Dashed line denotes noise level for a random case.

right panel of figure 3 in terms of IPR defined by

$$I_j = \sum_{k=1}^{N} \left(\nu_j^{(k)}\right)^4 \tag{4}$$

that estimates the inverse number of the components contributing significantly to the eigenvector $\boldsymbol{v}_j$. The market factor comprises over 90 stocks while the eigenvectors $\boldsymbol{v}_{j>80}$ are strongly localized; this remains in agreement with the outcomes of earlier works [1, 3].

Let us now pass to a description of the statistical properties of the eigensignals (2). In practical applications, variance of the eigensignal Z_j serves as a measure of the risk associated with the portfolio P_j; it can also be shown that $\sigma^2(Z_j) \sim \lambda_j$. The portfolios corresponding to the largest eigenvalues are much more risky than those corresponding to the ones of a moderate size and such a difference can manifest itself in the $z_j(i)$ p.d.f. In fact, as figure 4 documents, the cumulative distribution of $|z_1(i)|$ is much wider than its counterparts for the other eigensignals. Despite this, the tails of all the distributions show scaling that is close to the inverse cubic power law, similarly to c.d.f. of the stock returns [11–13]. Curiously, the extreme part of the tail for Z_1 looses its scaling. This behavior can be explained in a spirit of the Central Limit Theorem: the large delocalization of $\boldsymbol{v}_1$ causes Z_1 to be a sum of almost 100 stocks while the other eigensignals on average consist of only 1/3 of this number (this is not a rule, though, because in a more strongly correlated German market the situation is different). Other calculations for the eigensignals Z_j show [9] that the autocorrelation function $c(Z_j, Z_j; \tau)$ drops down quickly and reaches noise level immediately after one data point; this serves as a verification of the market efficiency. There is no difference between different eigensignals. However, the nonlinear correlations are Z_j-dependent [9]. For example, the volatility memory (a slow decay of $c(|Z_j|, |Z_j|; \tau)$ [14]) is by an order of magnitude stronger for Z_1 than for the other signals; also the leverage effect ($c(Z_j, |Z_j|; \tau) < 0$ for $\tau > 0$) [15] is present only in Z_1, Z_2 and Z_3. Generally, the eigensignals corresponding to the collective eigenvalues, and to λ_1 in particular, show considerably stronger nonlinear correlations.

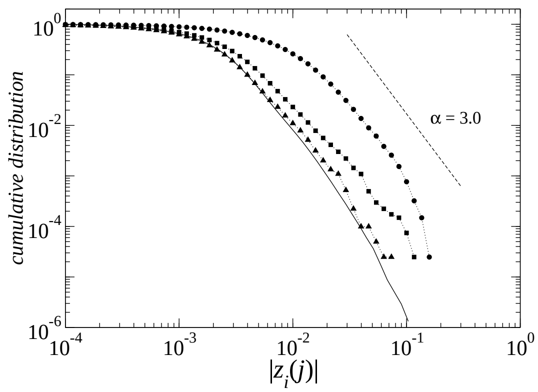

Fig. 4. Cumulative distribution functions for the eigensignals Z_1 (circles), Z_2 (squares), Z_3 (triangles), and the average c.d.f. for the other 97 eigensignals (solid line, no symbols). Dashed line denotes the inverse cubic power law.

It has been shown [16,18] that the evolution of the stock prices is strongly multifractal on different time scales ranging from the high-frequency to the daily one. This multifractality comes either from the broad probability distributions or from the existence of nonlinear correlations in data [17,18]. It can be expected that the eigensignals made up of the sums of stock returns are also of multifractal nature, but in general their multifractality expressed in terms of *e.g.* the singularity spectra $f(\alpha)$ can differ among Z_j's as a result of their different content. We applied the Multifractal Detrended Fluctuation Analysis (MF-DFA) [19] to all the eigensignals and calculated the $f(\alpha)$ spectra for each Z_j. In this procedure one removes trends from the integrated data and calculates the functional dependence of the average qth-order fluctuation function $F_q(n)$ (related to the qth moment of the signal) on the length of a signal window n. If the signal under study has a fractal structure, the fluctuation function shows the power-law behavior $F_q(n) \sim n^{h(q)}$, where $h(q)$ denotes a family of the generalized Hurst exponents. For an actual multifractal signal $h(q)$ forms a decreasing function of q. The singularity spectrum can then be obtained easily by using the following formula: $f(\alpha) = q[\alpha - h(q)] + 1$, where $\alpha = h(q) + qh'(q)$ is called the Hölder exponent.

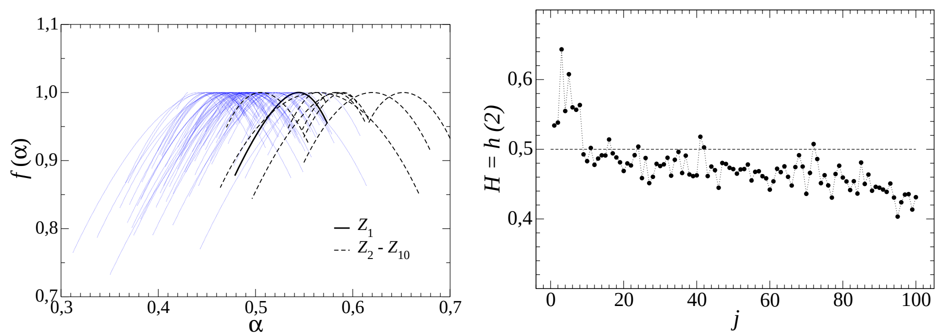

Fig. 5. (Left) Singularity spectra $f(\alpha)$ for eigensignals Z_j. Spectrum for the market factor Z_1 (thick solid) and a few other collective eigensignals $Z_2 - Z_{10}$ (thick dashed) are distinguished. (Right) The corresponding Hurst exponents $h(2)$; note the transition between $h(q) > 0.5$ for $j \leq 8$ and $h(q) < 0.5$ for $j > 8$.

Surprisingly, all the spectra in the left panel of figure 5 show clear multifractality (broad curves) and their widths are comparable; even the most collective Z_1 does not deviate much from the rest of the eigensignals. Such multifractal structure of all the eigensignals indicates that despite the fact that from a point of view of the information content some Z_j's can be considered as random and meaningless (the corresponding λ_j's fall within the RMT bounds and the eigenvector components are Gaussian-distributed), they still carry correlations and nonstationarity inherited from the stock returns. This can be regarded as an important observation also for practical reasons, because the eigensignals are associated with the specific portfolios. However, that these correlations leading to the multifractality are highly nonlinear testifies the apparent conflict between indications of the two panels of figure 5. In the right panel, where the Hurst exponents for $q = 2$ are presented, we observe a transition from a persistent behavior of Z_j for $j \leq 8$ to a slightly antipersistent one for $j > 8$. Thus, according to this measure, a majority of the eigensignals apparently do not reveal any significant linear autocorrelations.

3.2. Temporal correlations

The correlation matrix derived from the time series representing different trading days has different properties than its counterpart for a set of stocks. The main difference is that the time series are much shorter and that the ratio T/N is small ($Q = 4.39$). As a result, the spectrum of the eigenvalues for a Wishart matrix is relatively wide (figure 6). In spite of this fact, the largest eigenvalue λ_1=3.45 carries some amount of collectivity, which in the present case can be interpreted as a signature of the existence of repeatable events in the market evolution which occur at precisely the same moments of different trading days. The rest of the eigenvalues except a few ones form a bulk which perfectly agrees with RMT. This perfect agreement might suggest that nothing more interesting is present in the data, but keeping in mind the results of Section 3.1 it is more secure to conclude that at least no other easily-detectable strong correlations exist. The magnitude of λ_1 is not impressive, however, if compared with the matrix rank of 451. This means that even the events which are responsible for this eigenvalue are rather subtle.

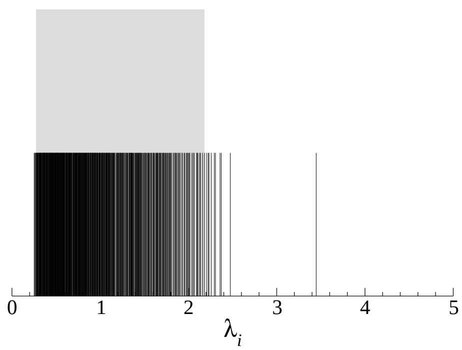

Fig. 6. Empirical eigenvalue spectrum of the correlation matrix $\boldsymbol{C}$ (vertical lines), calculated for DAX30 divided into $N = 451$ trading days, and compared with eigenvalue range for a random Wishart matrix with the same $Q = 4.39$ (shaded). Deviating eigenvalues represent repeatable events occurring at specific moments of a trading day.

Figure 7 shows p.d.f.'s for the components of different eigenvectors. In agreement with the above, this figure does not indicate any delocalization, but a visible degree of localization in $\boldsymbol{v}_1$ instead. Indeed, an inspection of the components of this eigenvector show large $\nu_j^{(k)}$ for certain trading days, while the rest assumes small values (not shown). It is very interesting that the largest components in $\boldsymbol{v}_1$ occur almost periodically for a long period of time: the events related to this eigenvector could happen with about monthly interval [7]. The nature of the events underlying the eigenvector $\boldsymbol{v}_1$ can be better understood if we look at figure 8 displaying the time series representing Z_1 and Z_2. Apart from the varying variance of the Z_1 and Z_2 fluctuations and from their non-Gaussian character (see [7]), a particularly extreme jump can be seen at 14:30 in Z_1 and there is a period of volatile

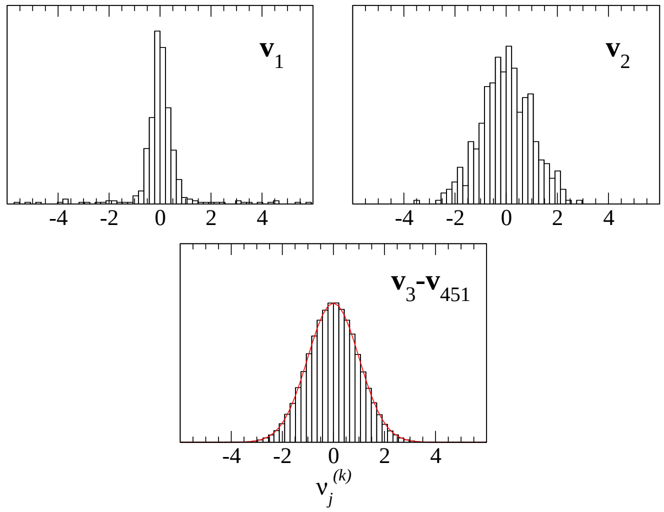

Fig. 7. Distributions of eigenvector components $\nu_j^{(k)}$ for j=1, j=2, and an average distribution for j=3,...,451 (histograms) together with a Gaussian (solid line).

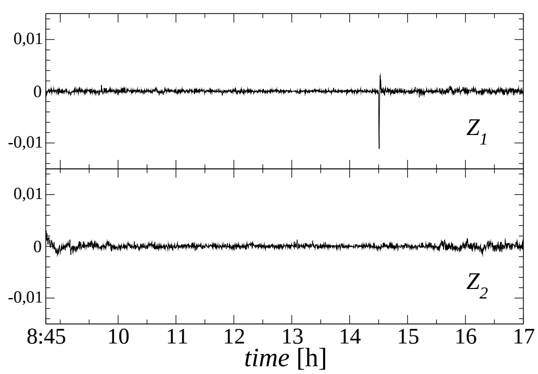

Fig. 8. DAX30 intraday eigensignals corresponding to two largest eigenvalues. The first 15 min of trading time were omitted because of an undefined value of index.

trading in early moments (before 9:15) in Z_2. The former can be associated with the market-related news releases in America (8:30 local time) which usually have an impact on some European markets [7]. The eigensignals corresponding to the smaller random eigenvalues do not show such large jumps and p.d.f.'s of their fluctuations does not deviate from Gaussian.

The multifractal and Hurst analysis would be unreliable in this case because our time series are too short.

4. Conclusions

In our study we decomposed the evolution of the set of stocks for 100 highly capitalized American companies into the eigensignals defined by the eigenvectors of the correlation matrix. Along the same scheme we also decomposed the daily pattern of the German DAX30 index into the components which can characterize the recurrent events occurring periodically during a trading day. Our results firmly show that

(a) for the high-frequency data and long time series, the number of the components carrying genuine information about the cross-asset correlations exceeds the previous estimates based on much shorter time series of daily returns; this suggests that the fact that an eigenvalue is located in the RMT range should be interpreted with care;

(b) also for the set of 100 stocks, the eigensignals corresponding to a few largest eigenvalues and describing the most collective components of the market reveal different and more pronounced nonlinear correlation structure than the other eigensignals;

(c) despite the different correlation strength and the different statistical properties of the collective and the other eigensignals, in the case of the 100 stocks all Z_j show rich multifractal behavior; the eigensignals for a few largest eigenvalues show also a trace of persistence;

(d) inherited from the stock price fluctuations, the tails of p.d.f. for the eigensignals scale closely to the inverse cubic law with a slightly worse scaling only for Z_1;

(e) the correlations between different trading days do not show strong periodic recurrence since only one eigenvalue departs considerably from the RMT spectrum; however, there are volatile periods of a trading day in which one can identify the repeatable events which might carry a portion of the market memory independently of the volatility clustering phenomenon.

REFERENCES

[1] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, *Phys. Rev. Lett.* **83**, 1467 (1999); V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, *Phys. Rev. Lett.* **83**, 1471 (1999).

[2] G. Bonanno, F. Lillo, R.N. Mantegna, *Quant. Finance* **1**, 96 (2001).

[3] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, T. Guhr, H.E. Stanley, *Phys. Rev.* **E65**, 066126 (2002).

[4] H. Markowitz, *J. Finance* **7**, 77 (1952).

[5] R. Roll, S.A. Ross, *J. Finance* **35**, 121 (1980).

[6] M.L. Mehta, *Random Matrices*, Academic Press, Boston 1999.

[7] S. Drożdż, J. Kwapień, F. Grümmer, F. Ruf, J. Speth, *Physica A* **299**, 144 (2001); J. Kwapień, S. Drożdż, F. Grümmer, F. Ruf, J. Speth, *Physica A* **309**, 171 (2002).

[8] R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, University Press, Cambridge 2000.

[9] J. Kwapień, S. Drożdż, P. Oświęcimka, `physics/0505074`.

[10] A.M. Sengupta, P.P. Mitra, *Phys. Rev.* **E60**, 3389 (1999).

[11] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, *Phys. Rev.* **E60**, 6519 (1999).

[12] S. Drożdż, J. Kwapień, F. Grümmer, F. Ruf, J. Speth, *Acta Phys. Pol. B* **34**, 4293 (2003).

[13] J. Kwapień, S. Drożdż, J. Speth, *Physica A* **330**, 605 (2003).

[14] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, *Phys. Rev.* **E60**, 1390 (1999).

[15] J.-P. Bouchaud, M. Potters, *Physica A* **299**, 60 (2001); J. Masoliver, J. Perello, *Int. J. Th. App. Fin.* **5**, 541 (2002); Z. Eisler, J. Kertész, *Physica A* **343**, 603 (2004).

[16] M. Pasquini, M. Serva, *Economics Letters* **65**, 275 (1999); K. Ivanova, M. Ausloos, *Physica A* **265**, 279 (1999); A. Bershadskii, *Physica A* **317**, 591 (2003); N. Vandewalle, M. Ausloos, *Eur. Phys. J.* **B4**, 257 (1998); A. Bershadskii, *Eur. Phys. J.* **B11**, 361 (1999).

[17] K. Matia, Y. Ashkenazy, H.E. Stanley, *Europhys. Lett.* **61**, 422 (2003).

[18] P. Oświęcimka, J. Kwapień, S. Drożdż, *Physica A* **347**, 626 (2005); J. Kwapień, P. Oświęcimka, S. Drożdż, *Physica A* **350**, 466 (2005).

[19] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, A. Bunde, Sh. Havlin, H.E. Stanley, *Physica A* **316**, 87 (2002).