PAPER: Interdisciplinary statistical mechanics

The q-dependent detrended cross-correlation analysis of stock market

Longfeng Zhao[1,2], **Wei Li**[1], **Andrea Fenu**[2,3],
Boris Podobnik[2,4,5,6,8], **Yougui Wang**[2,7]
and H Eugene Stanley[2]

[1] Complexity Science Center & Institute of Particle Physics, Hua-Zhong
(Central China) Normal University, Wuhan 430079, People's Republic of
China
[2] Center for Polymer Studies and Department of Physics, Boston University,
Boston, MA 02215, United States of America
[3] Department of Economics and Management, University of Cagliari, Italy
[4] Zagreb School of Economics and Management, Zagreb, HR 10000, Croatia
[5] Luxembourg School of Business, Grand-Duchy of Luxembourg, Luxembourg
[6] Faculty of Civil Engineering, University of Rijeka, Rijeka, HR 51000,
Croatia
[7] School of Systems Science, Beijing Normal University, Beijing 100875,
People's Republic of China
[8] Faculty of Information Studies Novo Mesto, 8000 Novo Mesto, Slovenia
E-mail: zlfccnu@mails.ccnu.edu.cn (L Zhao) and liw@mail.ccnu.edu.cn (W Li)

Abstract. Properties of the q-dependent cross-correlation matrices of the stock market have been analyzed by using random matrix theory and complex networks. The correlation structures of the fluctuations at different magnitudes have unique properties. The cross-correlations among small fluctuations are much stronger than those among large fluctuations. The large and small fluctuations are dominated by different groups of stocks. We use complex network representation to study these q-dependent matrices and discover some new identities. By utilizing those q-dependent correlation-based networks, we are able to construct some portfolios of those more independent stocks which consistently perform better. The optimal multifractal order for portfolio optimization is around $q = 2$

under the mean-variance portfolio framework, and $q \in [2, 6]$ under the expected shortfall criterion. These results have deepened our understanding regarding the collective behavior of the complex financial system.

Contents

1. Introduction

The analysis of the collective behavior of cross-correlations between different financial assets has become extremely attractive since physicists started to report the violation of efficient market hypothesis [1–3]. Initially, cross-correlation analyses relied on such linear tools as the Pearson correlation, which requires data to be stationary, but real-world financial data sets are rarely stationary. To take into account the non-linearity and non-stationarity in real-world data, new methods based on detrendization have been proposed, among which the most popular has been the detrended fluctuation analysis (DFA) [4]. Motivated by the DFA that is applied for a single time series, its generalization, named detrended cross-correlation fluctuation analysis (DCCA), has been proposed to quantify the long-range cross-correlations between a pair of non-stationary

signals [5]. DFA and DCCA are subsequently extended to their multifractal versions—MFDFA and MFDCCA, respectively[6–8]. DFA, DCCA and their multifractal counterparts have been applied cross a broad range of systems, including biological, and financial to physical systems[9–17]. Recently an analog to the Pearson coefficient, the detrended cross-correlation coefficient $\rho(s)$, was introduced in [18]. This coefficient, which applies to non-stationary signals, quantifies the significance level of correlations among fluctuations of detrended non-stationary signals at a given detrending scale s [19]. More recently, the DCCA coefficient $\rho(s)$ has been widely used to study the non-linear cross-correlations among financial time series [10, 20–23]. Despite the success of the $\rho(s)$ coefficient, it has some limitations when cross-correlations are quantified among fluctuations at different magnitudes. A more recent extension of the DCCA coefficient $\rho(s)$, the q-dependent detrended cross-correlation coefficient $\rho(q,s), q \in R$, is based on the q-dependent fluctuation function F_q from MFDFA and MFDCCA [6, 8, 24]. Kwapień *et al* [24] recently indicated that this method could be applied to the analysis of empirical data from such natural complex systems as physical, biological, social and financial systems. Our focus here is on the financial market.

Here we employ the q-dependent cross-correlation coefficient to quantify the cross-correlations among the return time series of 401 constituent stocks of the S&P 500 index. For those return time series, we calculate the q-dependent cross-correlation matrices $\mathbf{C}(q,s)$. We then calculate the statistical properties of the matrices at different multifractal orders q and varying time scales s. As when analyzing the Pearson cross-correlation matrix, we also analyze the eigenvalue and eigenvector dynamics of the matrices and find that the cross-correlations of stock market fluctuations at different magnitudes exhibit unique structures and dynamics. The large fluctuations are always dominated by a few industry groups, but the small fluctuations exhibit different behaviors. We then represent the cross-correlation matrices as complex networks and use the planar maximally filtered graph (PMFG) method [25] to construct the correlation-based networks and to analyze their basic topological features. The PMFG networks for small fluctuations are more heterogeneous than those networks obtained from large fluctuations. Using a centrality metric, we classify stocks as central or peripheral according to their centrality ranking. Applying this to portfolio optimization, we find that portfolios of peripheral stocks have consistently higher returns and lower risk than those portfolios of central and randomly selected stocks.

The paper is organized as follows. In section 2 we introduce the methodologies used in this paper. In section 3 we present the data and main empirical results. In section 4, applications to portfolio optimization have been given. The last section provides our conclusion and discussion.

2. Methodology

2.1. *q*-dependent cross-correlation analysis

The q-dependent cross-correlation coefficient can be obtained from the following procedure [8]:

(i) We consider a pair of time series x_i and y_i, $i = 1 \ldots l$. We integrate these time series and obtain two new time series

$$\chi^x(k) = \sum_{i=1}^{k} x_i - \langle x \rangle, k = 1 \ldots l, \tag{1}$$

$$\chi^y(k) = \sum_{i=1}^{k} y_i - \langle y \rangle, k = 1 \ldots l. \tag{2}$$

(ii) We divide $\chi^x(k)$ and $\chi^y(k)$ into $2M_s = 2 \times \text{int}(l/s)$ non-overlapping boxes of length s from the beginning and the end of two integrated time series. We then calculate the local trends for each segment $v(v = 1, \ldots, 2M_s)$ by a least-square fit and subtract it from $\chi^x(k)$ and $\chi^y(k)$ to detrend the integrated series. We then find the residual signals X, Y equal to the differences between the integrated signals and the mth-order polynomials $P_{s,v}^{(m)}$ fitted to these signals:

$$X_s(i, v) = \sum_{i=1}^{s} \chi^x(vs + i) - P_{X,s}^{(m)}(i, v), \tag{3}$$

$$Y_s(i, v) = \sum_{i=1}^{s} \chi^y(vs + i) - P_{Y,s}^{(m)}(i, v). \tag{4}$$

The covariance and variance of X and Y in box v are defined as:

$$f_{XY;s}^2(v) = \frac{1}{s} \sum_{i=1}^{s} X_s(i, v) Y_s(i, v), \tag{5}$$

$$f_{ZZ;s}^2(v) = \frac{1}{s} \sum_{i=1}^{s} Z_s^2(i, v), \tag{6}$$

where Z represents either X or Y.

(iii) We then define the fluctuation functions at multifractal order q and detrending scale s:

$$F_{XY}^q(s) = \frac{1}{2M_s} \sum_{v=1}^{2M_s} \text{sgn}[f_{XY;s}^2(v)]|f_{XY;s}^2(v)|^{q/2}, \tag{7}$$

$$F_{ZZ}^q(s) = \frac{1}{2M_s} \sum_{v=1}^{2M_s} [f_{ZZ;s}^2(v)]^{q/2}. \tag{8}$$

Then the q-dependent cross-correlation coefficient between x_i and y_i is defined as:

$$\rho(q,s) = \frac{F_{XY}^q(s)}{\sqrt{F_{XX}^q(s)F_{YY}^q(s)}}. \tag{9}$$

When $q = 2$, we restore the detrended cross-correlation coefficient $\rho(s)$ [18]. The q-dependent cross-correlation coefficient is bounded in $[-1,1]$ when $q \geq 0$. This coefficient can be an arbitrary value when $q < 0$ [24]. Here we focus on the case when $q > 0$. The exponent q acts as a filter. When $q > 2$, the boxes with large fluctuations contribute the most to $\rho(q,s)$, but when $q < 2$, the boxes with relatively small values dominate the fluctuation function, thus contributing the most to $\rho(q,s)$.

2.2. Random matrix theory

Having introduced the q-dependent cross-correlation coefficient, we now construct the cross-correlation matrices $\mathbf{C}(q,s)$ at different multifractal orders q and detrending scales s. In previous studies, random matrix theory [26] is the most commonly used method to analyze the cross-correlation matrices. Since the exact relationship between random time series and the q-dependent cross-correlation matrix cannot be established, although the q-dependent cross-correlation matrices have obvious different identities, we still employed many concepts in random matrix theory to analyze the q-dependent cross-correlation matrices. If we assume that the correlation matrices are random, random matrix theory can be employed as a benchmark to quantify to what extent the properties of a matrix deviate from the prediction of a purely random case. Random matrix theory has been widely applied to investigate the collective phenomena in financial markets [1, 2, 27–35]. But, as mentioned in the previous context, we cannot use random matrix theory as a benchmark for the q-dependent cross-correlation matrix. A comprehensive review of the application of random matrix theory in the financial market is provided in [36]. Since we cannot derive the exact equation for the q-dependent cross-correlation matrix in a purely random situation, the shuffled and random simulation results have been used as other references. Nevertheless, the tools from random matrix theory can be quite useful to explore the structures of the cross-correlation matrices. Here we briefly introduce random matrix theory.

We consider a random correlation matrix constructed from a bunch of normal distributed uncorrelated time series R $= \{r_i(t), i = 1, \ldots, N, t = 1, \ldots, L\}$

$$\mathbf{C} = \frac{1}{L}\mathrm{RR}^T, \tag{10}$$

where R is an $N \times L$ matrix containing N time series $\{r_i(t)\}$ of length L with zero mean and unit variance, that are mutually uncorrelated. The probability distribution function of the eigenvalues for the random matrix can be written analytically in the limit $N, L \longrightarrow \infty$ with a fixed $Q = \frac{L}{N} > 1$ [26],

$$P(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{11}$$

where λ_- and λ_+ are the minimum and maximum eigenvalues of $\mathbf{C}$. λ_- and λ_+ are given by

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \tag{12}$$

If the eigenvalue distributions deviate from the prediction of equation (11), that signals the existence of mutual correlation in the time series.

We decompose the *q*-dependent cross-correlation matrices with eigenvalues $\lambda_k, k = 1, \ldots, N$ and eigenvectors $\mathrm{U}_k, k = 1, \ldots, N$ which provide information regarding the collective behavior of the stock market. The eigenvalues are sorted in ascending order with $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$. Here we use the inverse participation ratio (IPR) to quantify the reciprocal number of eigenvector components that significantly contribute to the corresponding eigenvalue. The IPR is defined as

$$I_k = \sum_{l=1}^{N} [u_k^l]^4. \tag{13}$$

Here u_k^l is the *l*th component of the eigenvector U_k corresponding to eigenvalue λ_k. The meaning of I_k can be illustrated by two limiting cases: (i) a vector with identical components $u_k^l = 1/\sqrt{N}$ has $I_k = 1/N$, whereas (ii) a vector with one component $u_k^l = 1$ and the remainder $u_k^l = 0$ has $I_k = 1$. We also define the participation ratio (PR) as $1/I_k$, which is approximately equal to the number of significant contributors for eigenvalue λ_k. In random matrix theory, the expectation of IPR for a purely random time series is

$$\langle I_k \rangle = N \int_{-\infty}^{\infty} [u_k^l]^4 \frac{1}{\sqrt{2\pi N}} \exp(-\frac{[u_k^l]^2}{2N}) \mathrm{d}u_k^l = \frac{3}{N}. \tag{14}$$

2.3. Planar maximally filtered graph

As suggested in [24], we use the complex network approach to analyze the *q*-dependent cross-correlation matrix. We employ the PMFG method [25] to construct networks based on cross-correlation matrices $\mathbf{C}(q, s)$. The PMFG algorithm converts the dense correlation matrix into a sparse representation. The algorithm is implemented as follows:

(i) Sort all of the $\rho_{ij}(q, s)$ in descending order to obtain an ordered list $l_{\mathrm{sort}}(q, s)$.

(ii) Add an edge between nodes i and j based on the order in $l_{\mathrm{sort}}(q, s)$ only if the graph remains planar after the edge is added.

(iii) A graph $G(q, s)$ is formed with $N_e = 3(N - 2)$ edges under the constraint of planarity and N is the number of stocks.

As described in [25], PMFGs not only maintain the hierarchical organizations of the minimum spanning trees (MST), but also generate cliques, especially three and

four node cliques. We calculate the basic topological parameters, such as clustering coefficient C, the shortest-path length L and the assortativity A [37]. We also adopt a heterogeneity index γ to measure the heterogeneity of PMFGs which is defined by [38]

$$\gamma = \frac{N - 2 \sum\limits_{ij \in \{e\}} (k_i k_j)^{-1/2}}{N - 2\sqrt{N - 1}}. \tag{15}$$

Here k_i and k_j are the degrees of nodes i and j connected by edge $\{e_{ij}\}$.

3. Data and results

3.1. Data description

Our data sets include $N = 401$ S&P 500 constituent stocks from 4 January 1999 to 31 December 2014 with 4025 trading records for each stock. We use the logarithm return defined as

$$r_i(t) = \ln(p_i(t+1)) - \ln(p_i(t)), \tag{16}$$

where $p_i(t)$ is the daily adjusted closure price of stock i at time t. We then use the previous method to compute the q-dependent cross-correlation coefficients between any pair of return time series $r_i(t)$ and $r_j(t)$ and obtain the $N \times N$ matrix $\mathbf{C}(q, s)$. The matrix entries of $\mathbf{C}(q, s)$ are the correlation coefficients $\rho_{ij}(q, s)$ between all pairs of stocks. We set the multifractal orders $q \in [0.2, 5]$ with $\delta q = 0.2$ and the detrending scales $s \in [30, 600]$ trading days with $\delta_s = 40$ d. We also perform the same calculation on the shuffled return time series and the simulated uncorrelated random time series and use them as reference models.

3.2. Cross-correlation matrix analysis

With those q-dependent cross-correlation matrices $\mathbf{C}(q, s)$ at different multifractal orders q and detrending scales s, we analyze the probability distributions of the cross-correlation values, i.e. the upper triangle entries of the correlation matrices and the eigen dynamics.

First we show the plot of matrices for different multifractal orders q and detrending scales s in figures 1. The strength of the average correlation will increase slightly as the scale s increases, but will decrease dramatically as the multifractal order q increases. We sort the rows and columns of the correlation matrices according to the official sector and sub-sector partitions of S&P 500. Note the distinct sector and sub-sector structures in the cross-correlation matrices. In particular, when $q < 2$, the sector structures are much more pronounced.

Figure 2 shows the probability distribution function P(ρ) of the matrices elements for different values of $q = 0.4, 1.0, 1.4, 2.0, 4.0$ and different values of scales $s = 70, 110, 230, 310, 430$. We can observe that the distribution of the matrices becomes increasingly skewed to the left, and the widths of the distributions become narrow as the multifractal order q increases. The probability distribution of the q-dependent
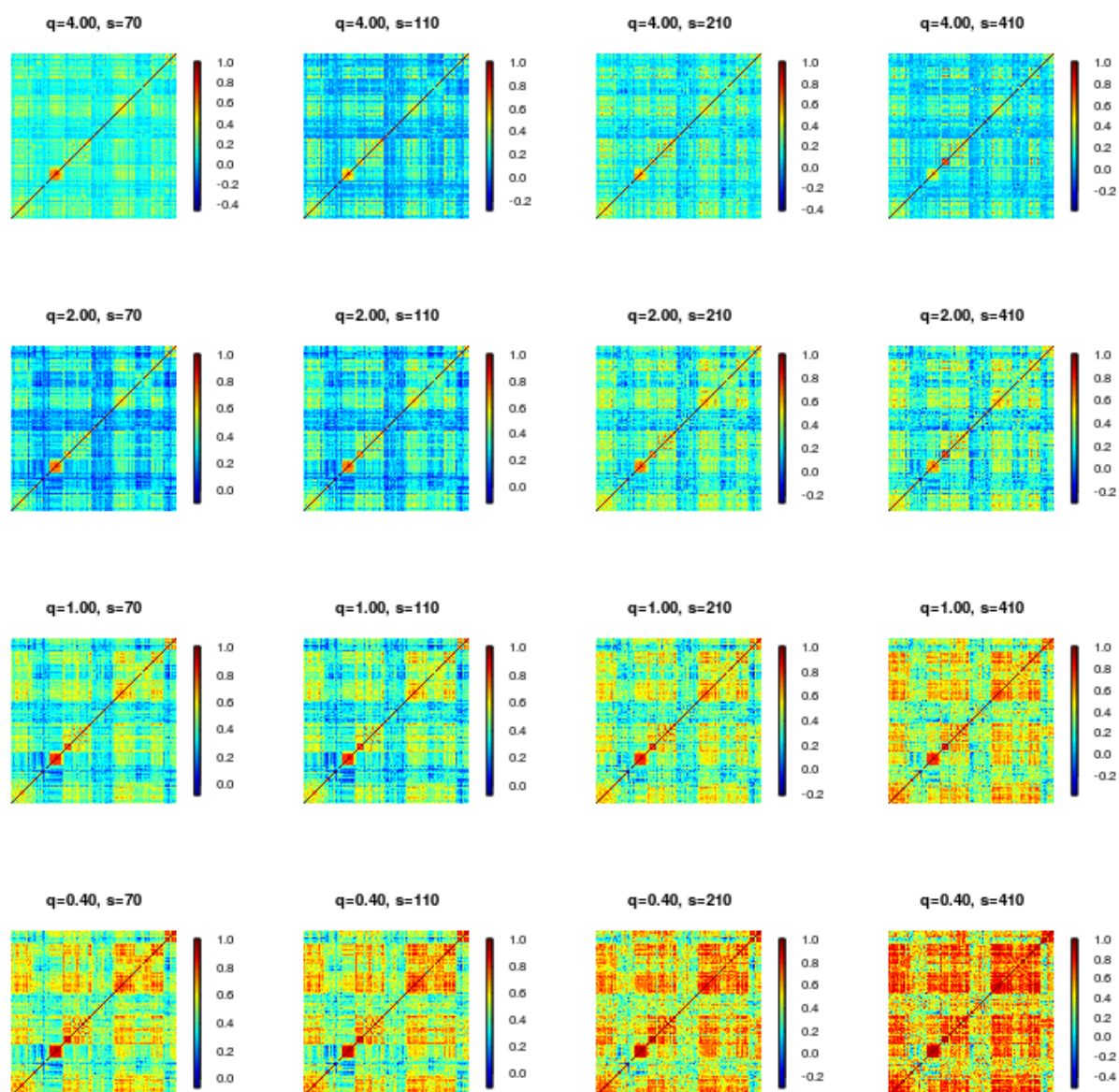
Figure 1. The cross-correlation matrices $\mathbf{C}(q,s)$ between 401 S&P 500 constituent stocks for different orders $q = 0.4, 1.0, 2.0, 4.0$ and scales $s = 70, 110, 210, 410$ d. We have sorted the matrix entries according to the official sector classification of those stocks. Thus very clear intra sector and inter sector structures have been presented.

cross-correlation coefficients for the return time series deviates significantly from the shuffled and simulated distributions, and this may provide genuine information regarding the cross-correlation among different magnitudes of fluctuations. The shuffled and simulated distributions coincide with each other, which means that the probability distribution of the return time series is not responsible for those deviated cross-correlations. At the same instant, the shuffled and simulated distributions are symmetric around zero means. Thus the different cross-correlation structures are the result of non-linear correlations among different magnitudes of fluctuations. In addition, when $q > 2$, the distributions become relatively close to the shuffled case. We also calculate the first

four order moments of the correlation matrices to better illustrate the variation of the cross-correlation distributions.

Figure 3 shows the first four order moments of the cross-correlation coefficient distributions at different multifractal orders q and detrending scales s. The average cross-correlation decreases as the multifractal order increases, indicating that the cross-correlations between large fluctuations are relatively weak. From the variance, skewness and kurtosis, we see an obvious transition in the shapes of the distributions around $q = 2$. For small q and large s, the variance of the distribution is large. The shape of the distribution changes from a right skew to a left skew shape when q decreases and s increases. The distribution becomes increasingly peaked with kurtosis larger than three when q increases and s decreases. The cross-correlation coefficients for large and small multifractal orders q are largely different, which indicates disparate correlation structures among different magnitudes of fluctuations.

To analyze the genuine information carried by the q-dependent cross-correlation matrices, we decompose the cross-correlation matrices and sort the eigenvalues $\lambda_k, k = 1, \ldots, 401$ in ascending order with their corresponding eigenvectors $U_k, k = 1, \ldots, 401$. Figures 4 and 5 show the distributions of the bulk eigenvalues and deviating eigenvalues, respectively. Figure 4 only provides eigenvalues smaller than two to better present the bulk eigenvalue distributions. The black and blue lines are the eigenvalue distributions for the original q-dependent cross-correlation matrices and the shuffled scenario. We also simulate 401 random Gaussian distributed uncorrelated time series. The green lines are the bulk eigenvalues from the q-dependent cross-correlation matrices calculated by using the simulated Gaussian time series. We find that the bulk eigenvalue distributions of the shuffled time series and the simulated time series are approximately the same. This confirms that the deviations of the eigenvalue distributions are the result of non-linear cross-correlations. The distributions of the bulk eigenvalues for the original q-dependent cross-correlation matrices differ from two reference models. Note that when $q < 2$, many eigenvalues of the q-dependent cross-correlation matrices are negative. But for $q \geqslant 2$, the q-dependent cross-correlation matrices are positive definite, and we can see that the magnitude of the negative eigenvalues will increase gradually as q decreases. For the Pearson cross-correlation matrices, if the data length $L \geqslant N$, the cross-correlation matrices are positive definite. Although our data length is more than ten times the system size, there is no such property that guarantees the positive definite of the q-dependent cross-correlation matrices. Except for the empirical q-dependent cross-correlation matrices, there are also some negative eigenvalues for the shuffled and simulated matrices when $q < 2$.

Figure 5 shows the deviating eigenvalues for the original cross-correlation matrices (black), the shuffled results (blue), and the simulated results (green). The behavior of those deviating eigenvalues differ as the values of q and s differ. Large q values and small s values tend to cause a greater amount of large deviating eigenvalues. In figure 5, note that the deviating eigenvalues for large $q = 4$ and small $s = 70$ are especially clear. In contrast, when $q = 0.4$ and $s = 430$, only the largest and second largest eigenvalues continue to deviate from the shuffled and simulated eigenvalues. This indicates that the small fluctuations only have very short characteristic time scales. The long-term average effect of small fluctuations equals the noise level. Meanwhile, we can observe that the largest eigenvalue will decrease dramatically as q increases, which means that
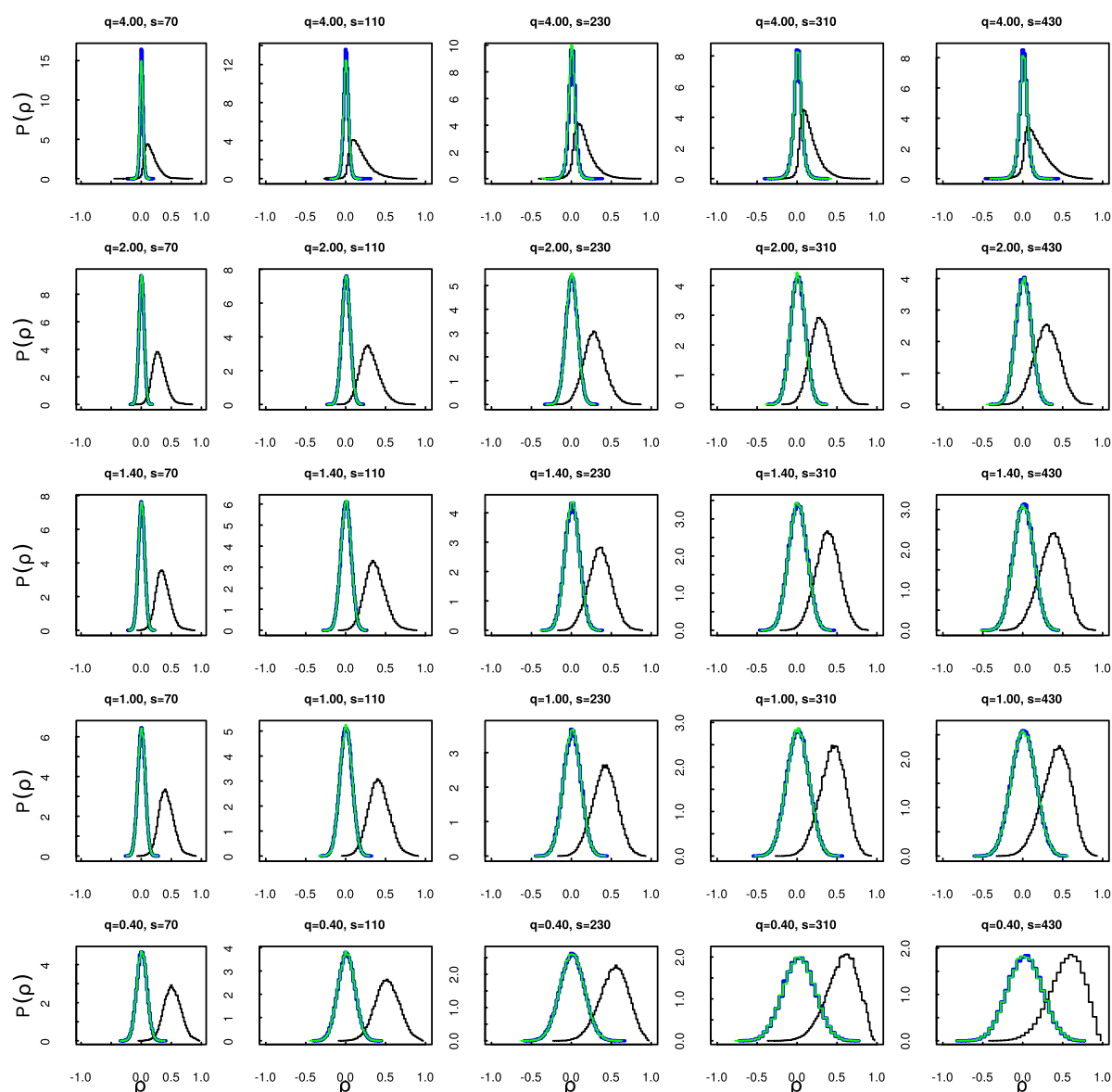
Figure 2. The distributions of the non-diagonal elements of the cross-correlation matrices $\mathbf{C}(q, s)$. The black lines are the distributions of the q-dependent cross-correlation coefficients for the return time series. The blue lines and green lines are the same distributions for the shuffled return time series and simulated random time series, respectively. $q = 0.4, 1.0, 1.4, 2.0, 4.0$ from bottom to top and $s = 70, 110, 230, 310, 430$ d from left to right. The distributions of the matrice elements for shuffled and simulated time series largely overlap each other.

the large fluctuations are localized. Generally speaking, large multifractal orders q and small detrending scales s make the sector structures (deviating eigenvalues) separate from the noise level. However, small multifractal orders q and large detrending scales s correspond to the strong market mode (the largest eigenvalue).

The first four eigenvalues for different multifractal orders q and detrending scales s are shown in figure 6. The largest eigenvalues λ_{401} for $q < 2$ are approximately equal to the order of the system size. The behavior of the largest eigenvalue is similar to the
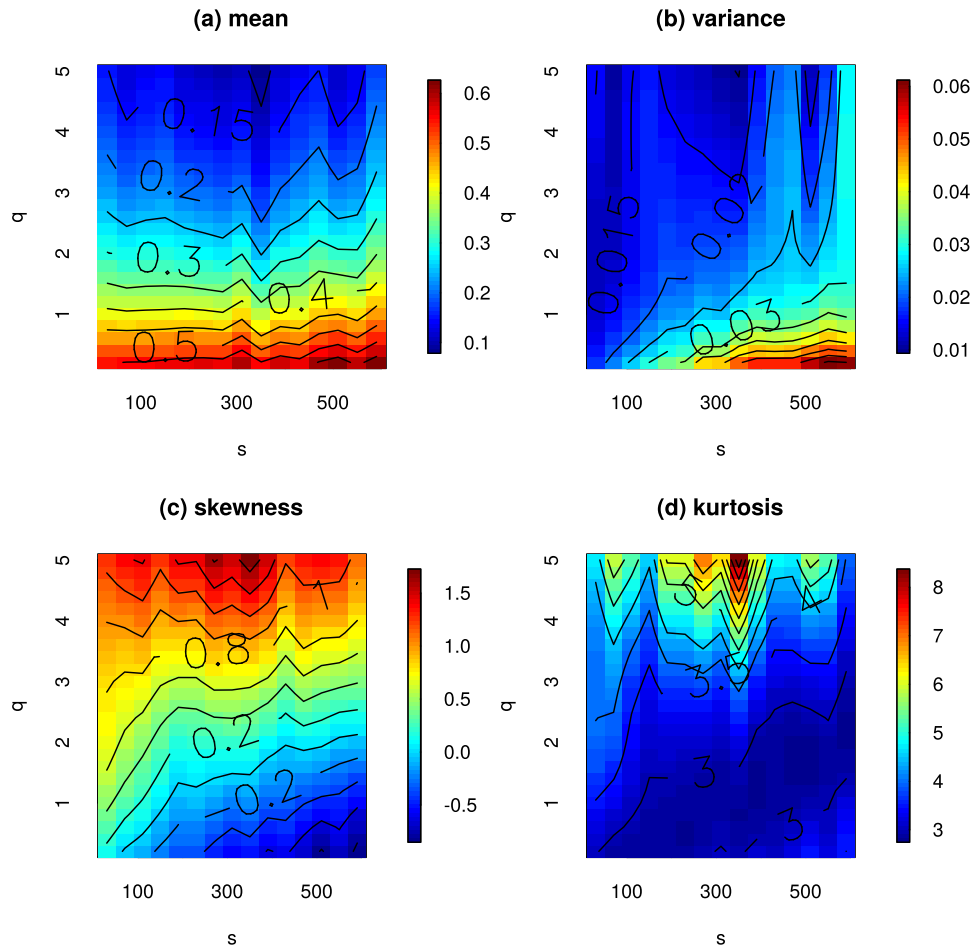
Figure 3. The first four order moments—mean, variance, skewness, kurtosis—of the *q*-dependent cross-correlation matrices $\mathbf{C}(q,s)$ at different multifractal orders *q* and detrending scales *s*. $q \in (0.2, 5)$ with step size $\delta_q = 0.2$ and $s \in (30, 600)$ with step size $\delta_s = 40$ d. The black lines inside the heat map are contour lines for the first four order moments.

average cross-correlation in figure 3(a). This supports the conclusion that the largest eigenvalue corresponds to the market mode described by numerous studies [1, 27] and it decreases when the value of *q* increases. Thus the market mode at small *q* is extremely strong, which seems counterintuitive. It means that the whole market fluctuations only influence the constituent stocks in a relatively small fluctuation level. We also observe that the first four eigenvalues increase as the detrending scale *s* increases.

It is believed that those eigenvalues which deviated from the bulk contain some genuine information related to the sector or industry, as described in [27, 29]. To uncover the hidden information carried by those deviating eigenvalues at different multifractal orders and detrending scales, we first partition those 401 stocks into industry groups labeled $l = 1 \ldots 24$ (N_l stocks each) according to the industry group code of the stocks supplied by GICS [39]. We then construct a projection matrix P, with elements $P_{li} = 1/N_l$ if stock *i* belongs to industry group *l* and $P_{li} = 0$ otherwise. For
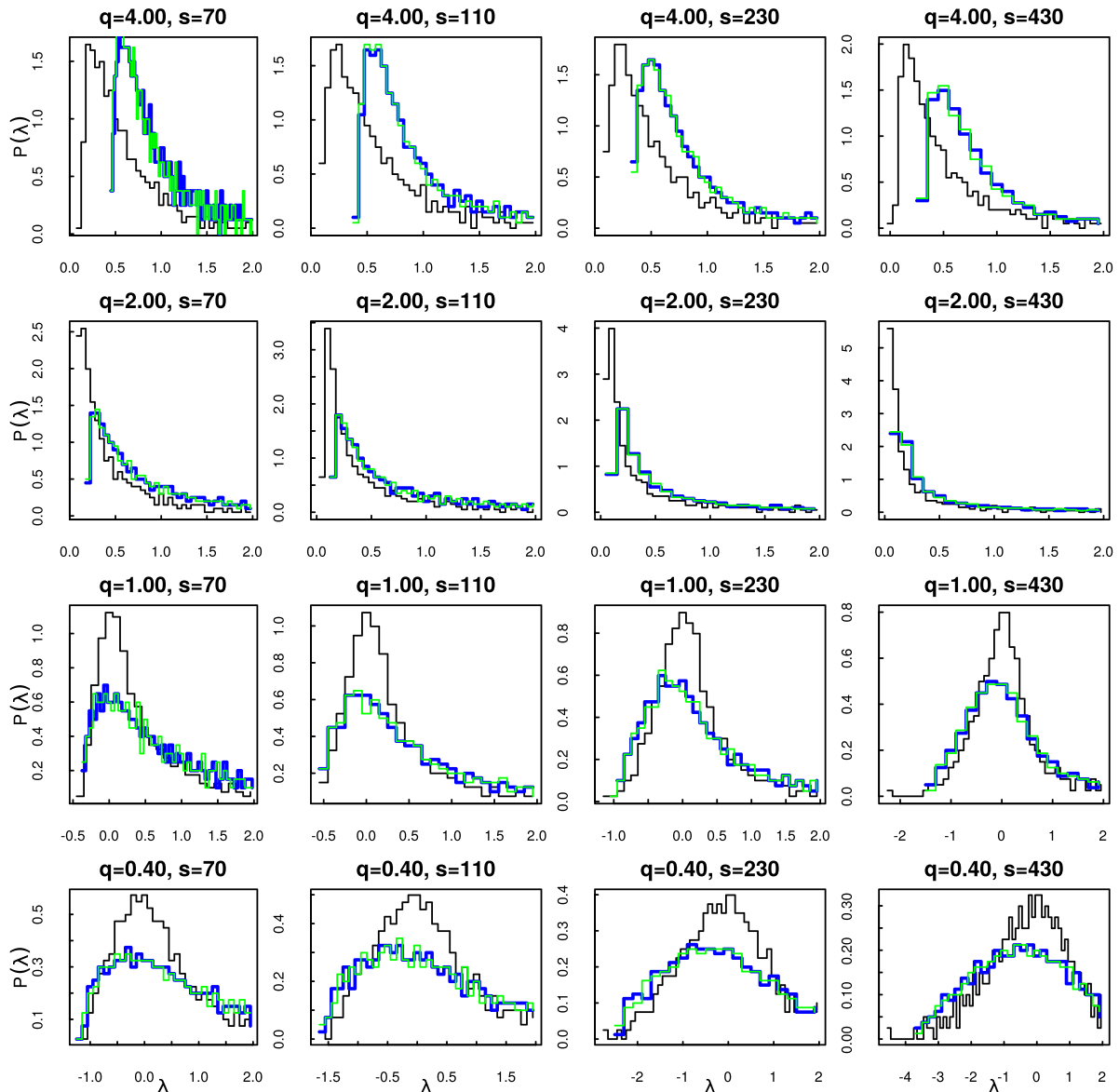
Figure 4. The eigenvalue distributions P(λ) of the *q*-dependent cross-correlation matrices $\mathbf{C}(q,s)$ inside the eigenvalue bulk region. We only show the distributions of those eigenvalues that are smaller than two. The black, blue and green lines are the eigenvalue distributions of the *q*-dependent cross-correlation matrices for the original, shuffled and simulated time series, respectively. The multifractal orders $q = 0.4, 1.0, 2.0, 4.0$ from bottom to top and the detrending scales $s = 70, 110, 230, 430$ d from left to right.

each eigenvector U_k, the contribution $X_k^l = \sum_{i=1}^{N} P_{li}(u_k^i)^2$ of each industry group can be obtained. Figure 7 shows the contribution of each industry group to the smallest and second smallest eigenvalues $\lambda_k, k = 1$ and $\lambda_k, k = 2$. The red ($k = 1$) and sky blue ($k = 2$) lines are the contribution values of the industry groups to these two eigenvalues. The blue lines are the average contribution values X_k^l for the correlation matrices calculated by using the shuffled time series. This reference model tells us how much the X_k^l deviates from the noise level. There are 24 major industry groups for
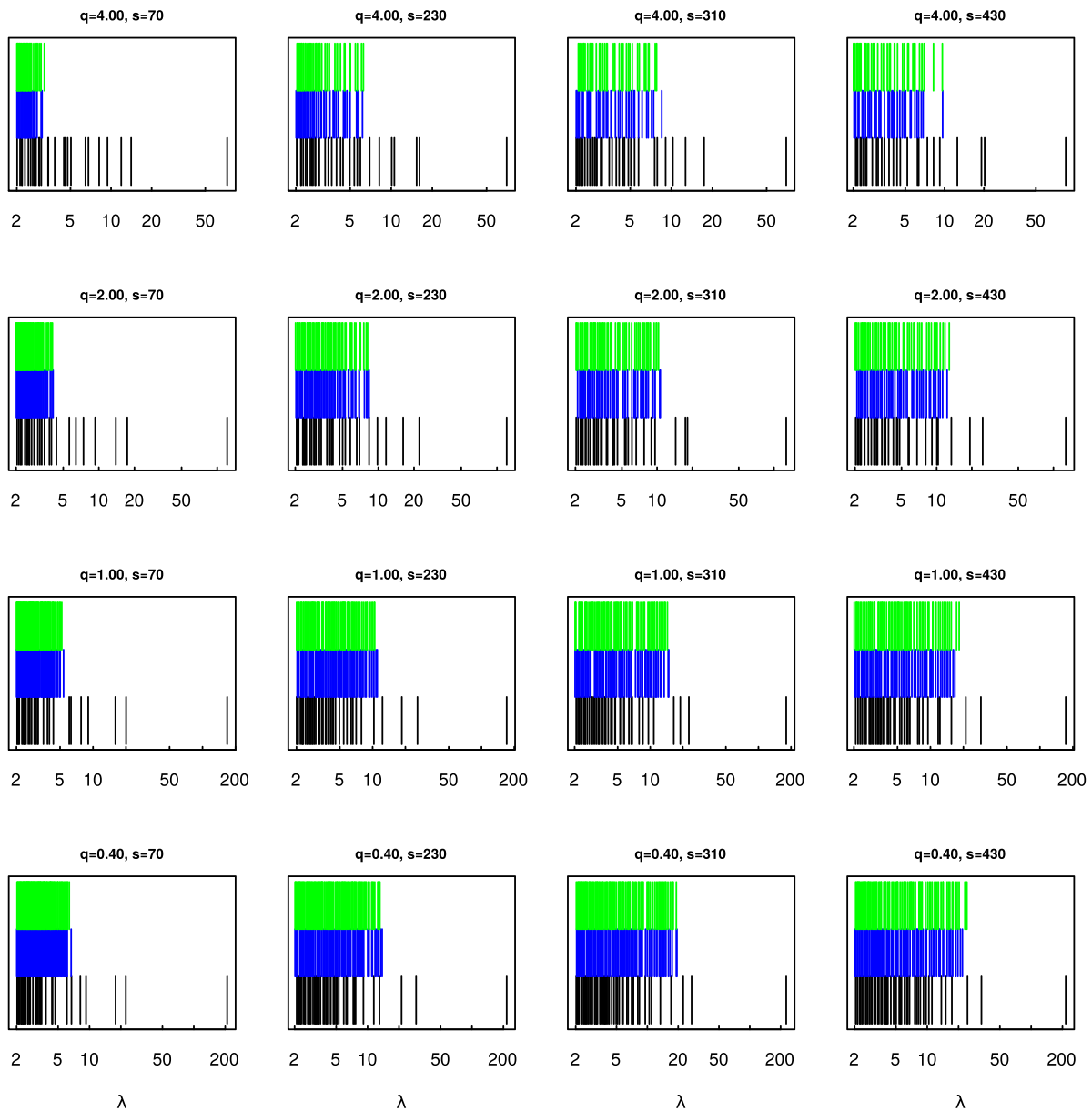
Figure 5. The eigenvalues of the cross-correlation matrices $\mathbf{C}(q, s)$ deviating from the bulk ($\lambda > 2$). The meaning of the color is the same as in figure 4. Each vertical line indicates that there is an eigenvalue with a corresponding magnitude. The multifractal orders $q = 0.4$, 1.0, 2.0, 4.0 from bottom to top and the detrending scales $s = 70$, 230, 310, 430 d from left to right.

the 401 stocks, from left to right: Media, Retailing, Consumer Durables & Apparel, Automobiles & Components, Consumer Services, Food & Beverage & Tobacco, Food & Staples Retailing, Household & Personal Products, Energy, Diversified Financials, Banks, Insurance, Real Estate, Pharmaceuticals, Biotechnology & Life Sciences, Health Care Equipment & Services, Capital Goods, Transportation, Software & Services, Commercial & Professional Services, Materials, Technology Hardware & Equipment, Semiconductors & Semiconductor Equipment, Telecommunication Services, Utilities. It is shown that for the smallest and second smallest eigenvalues, λ_1 and λ_2, the

contributions are from a few industry groups such as Household & Personal Products, Pharmaceuticals and Utilities. The X_k^l for these industry groups are much stronger than the noise level. One explanation for this phenomena is that those three industries are relatively insensitive to the micro economics status. So the fluctuation of those sectors is extremely small. Meanwhile, the industry contributions of two large eigenvalues $\lambda_k, k = 399, 400$ are presented in figure 8. The contributions to large eigenvalues also come from a few industry groups and are much stronger than the noise level. For λ_{399}, there are multiple industries which contribute significantly with a mixed pattern. The main contributions are from Diversified Financials, Banks and Real Estate. Those sectors are all related to the financial industry with very high risk and the contributions come from different sectors for different multifractal orders q. Meanwhile, the contribution to λ_{400} is always from the Energy sector for $q < 2$. But for $q \geq 2$, the contributions to λ_{400} are from Energy, Banks, Semiconductors & Semiconductor Equipment and Utilities.

As shown in figure 4, there are many small eigenvalues within the eigenvalue ranges of the shuffled and simulated matrices. Figure 9 shows the contributions of each industry group to the eigenvalues $\lambda_k, k = 200, 250$ deep inside the eigenvalue bulk region. As expected, in this region the eigenvalues do not exhibit a significant pattern. The contribution levels X_k^l of each industry group are the same as for the shuffled time series. For both $\lambda_k, k = 200$ and $\lambda_k, k = 250$, there is no clear contributing industry group.

As explained above, the IPR quantifies the reciprocal number of the eigenvector components that contribute significantly. Here we give the IPRs of the q-dependent cross-correlation matrices at different multifractal orders and detrending scales in figure 10. We present the IPR without the largest eigenvalues for better visualization. Note that there is a transition in the IPR I_k between small and large multifractal orders q. When $q \leqslant 2$, the small eigenvalues are dominant by a relatively small proportion of stocks with larger IPR. It can be validated by using the PR $1/I_k$ in figure 12, which shows that the PRs for those small eigenvalues are less than 50, but for medium and large eigenvalues, the PR is larger than 200.

Figure 11 shows the PR $1/I_k$ for the largest eigenvalue. The largest PR for $q < 2$ is 376, which approaches the system size $N = 401$. When $q \geqslant 2$, the PR for the largest eigenvalue decreases rapidly with a value of 200. The striking difference in the contribution number of the largest eigenvalues for different fluctuations implies that the collective behavior of small fluctuations ($q < 2$) is more homogeneous (large PR). Figure 12 shows the heat map of the PR $1/I_k$ at different multifractal orders q when $s = 50$, 210, 410, 810. $k = 1, \ldots, 400$ are the labels of the eigenvalues λ_k. Note that the PR for the largest eigenvalue λ_{401} is not shown in figure 12. When $q \geqslant 2$, the PRs for small eigenvalues (small k) are very small, suggesting that the small eigenvalues contain useful information. Only a very small set of stocks contributed to the smallest eigenvalue. We can verify this from the eigenvector component contributions in figure 7. When $q \geqslant 2$, the small and large eigenvalues are dominated by a few sectors with PR values smaller than 80. Meanwhile, the PR values for those medium eigenvalues are around 120 which are very close to the purely random situation $N/3 \approx 133.7$. But for $q < 2$, the PR values are all around $N/3 \approx 133.7$ for most of the eigenvalues, which means for small fluctuations the contributions are homogeneous and very close to the random case except the PR values for those very large eigenvalues.
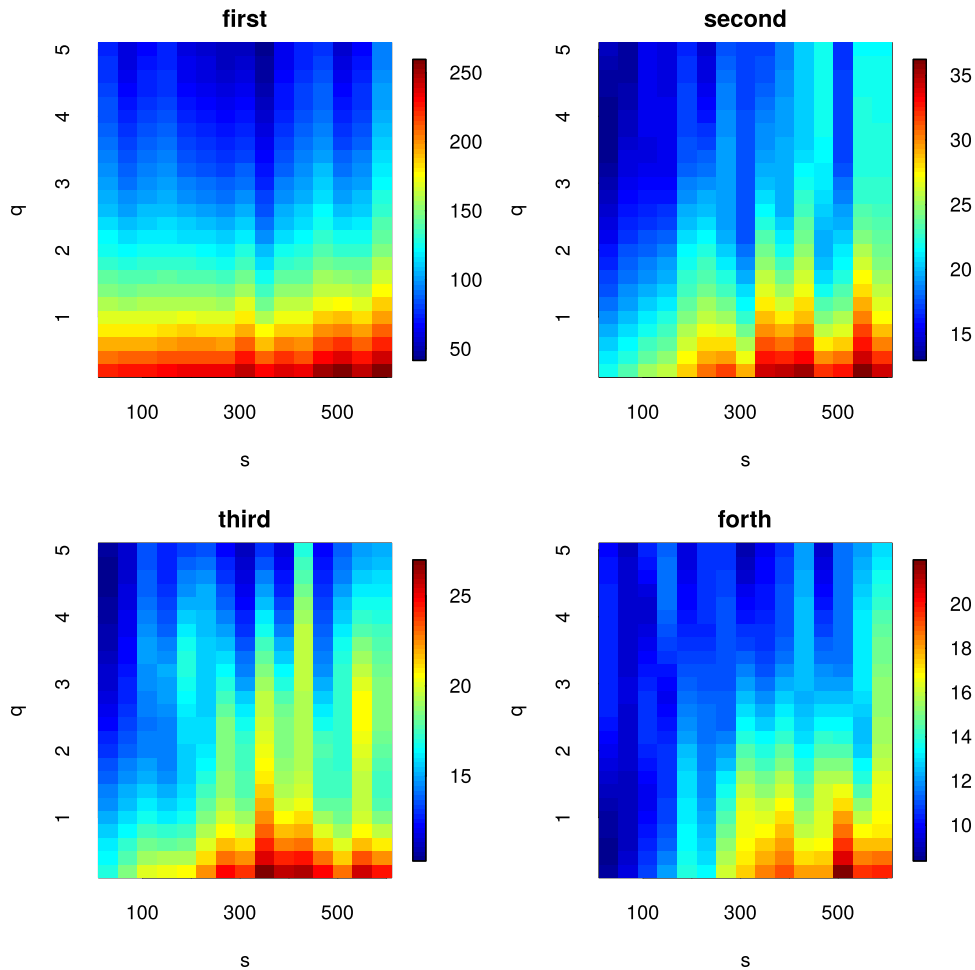
Figure 6. The heat maps of the first four eigenvalues $\lambda_k, k = 401, 400, 399, 398$ as a function of multifractal orders $q \in (0.2, 5)$ with step size $\delta_q = 0.2$ and detrending scales $s \in (30, 600)$ with step size $\delta_s = 40$ d.

Those results above have implications relevant to portfolio optimization. In general, the pattern of collective behavior for small fluctuations is quite different from that for large fluctuations.

3.3. PMFG analysis

The complex network has become a standard and powerful way to analyze the cross-correlation among subunits with mutual interactions. Particularly for the financial market, rich dynamics of the cross-correlation matrix has inspired plenty of correlation-based network methods and enormous applications[40–44] The PMFG has been used to analyze the structure and dynamics of the stock market in times of crisis [45, 46], and it can effectively capture the sector structures. Here we construct the PMFG networks by using the q-dependent cross-correlation matrices, namely, q-PMFG.

Figure 13 shows the q-PMFG networks constructed by using the PMFG algorithm. The sector structures for small q are clearer than those for large q. Recently Kawpeń *et al* [16] constructed an MST by using q-dependent cross-correlation matrices.
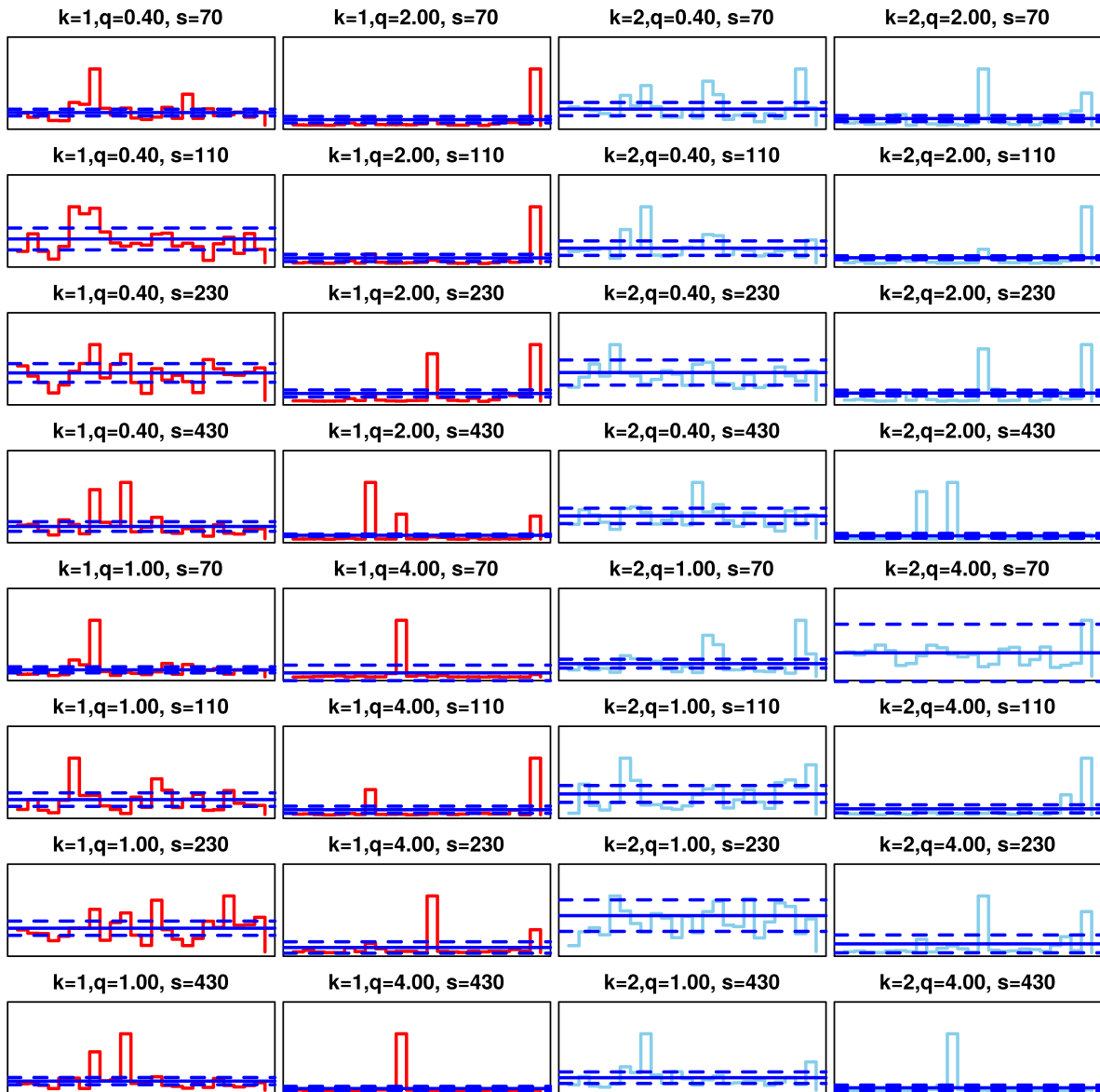
Figure 7. The contributions $X_k^l, l = \dots 24$ of each industry group to the smallest eigenvalue $\lambda_k, k = 1$ (red lines) and second smallest eigenvalue $\lambda_k, k = 2$ (sky blue lines) at different multifractal orders $q = 0.4, 1.0, 2.0, 4.0$ and detrending scales $s = 70, 110, 230, 430$ d. The blue solid and dashed lines are the mean X_k^l with one standard deviation for the cross-correlation matrices calculated from the shuffled return time series.

Some hidden structures were found by using minute datasets. It is known that the PMFG network includes the MST. Here we find that when $q \leqslant 2$, a hub stock emerges, but when $q > 2$, the degree heterogeneity becomes weak. In particular, when $q \leqslant 2$, the dark green nodes (the stocks from the financial sector) are very close to each other. However, when $q > 2$, the links between the financial sector stocks loosen. Those characteristics qualitatively agree with the results from [16] where they discovered a star-like MST structure when $q \leqslant 2$.
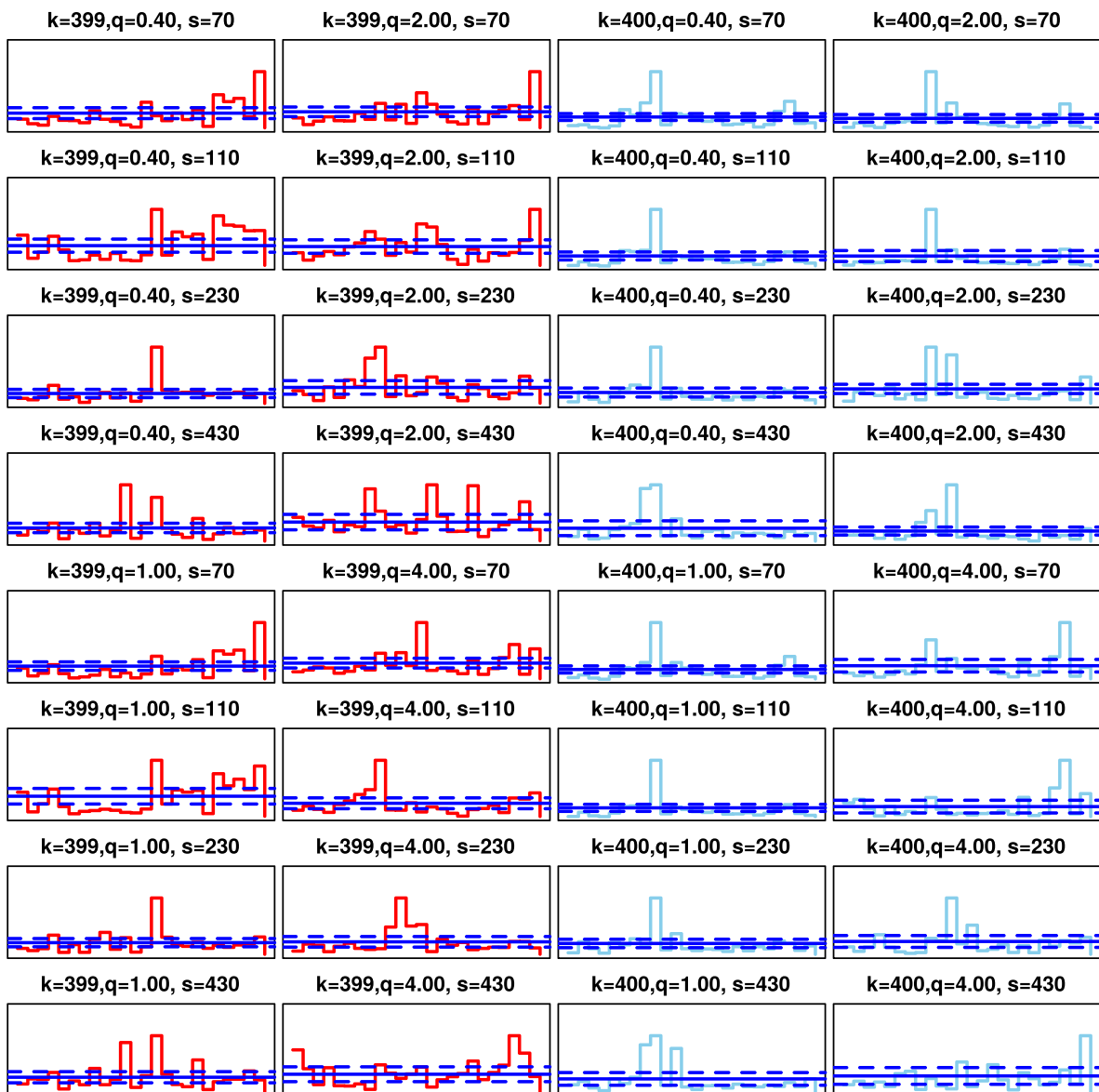
Figure 8. The contributions $X_k^l, l = 1, \ldots, 24$ of each industry group to the third largest eigenvalue $\lambda_k, k = 399$ (red lines) and second largest eigenvalue $\lambda_k, k = 400$ (sky blue lines) at different multifractal orders $q = 0.4, 1.0, 2.0, 4.0$ and detrending scales $s = 70, 110, 230, 430$ d. The blue solid and dashed lines are the mean X_k^l with one standard deviation for the shuffled correlation matrices.

To quantify the influence of the fluctuations on the *q*-PMFGs at different multifractal orders *q* and detrending scales *s*, we calculate the topological quantities of the *q*-PMFGs. The topological quantities of the *q*-PMFGs are presented in figure 14. Figure 14(a) shows that the clustering coefficient C of *q*-PMFG increases as the multifractal order *q* increases. The shortest path length L has been shown in figure 14(b). The shortest path length is longer for large *q* and short *s*. Figure 14(c) shows the heterogeneity indexes H [38], which quantifies the heterogeneity level of the *q*-PMFGs. It is analogous to the power law index of the scale-free network. It is known that
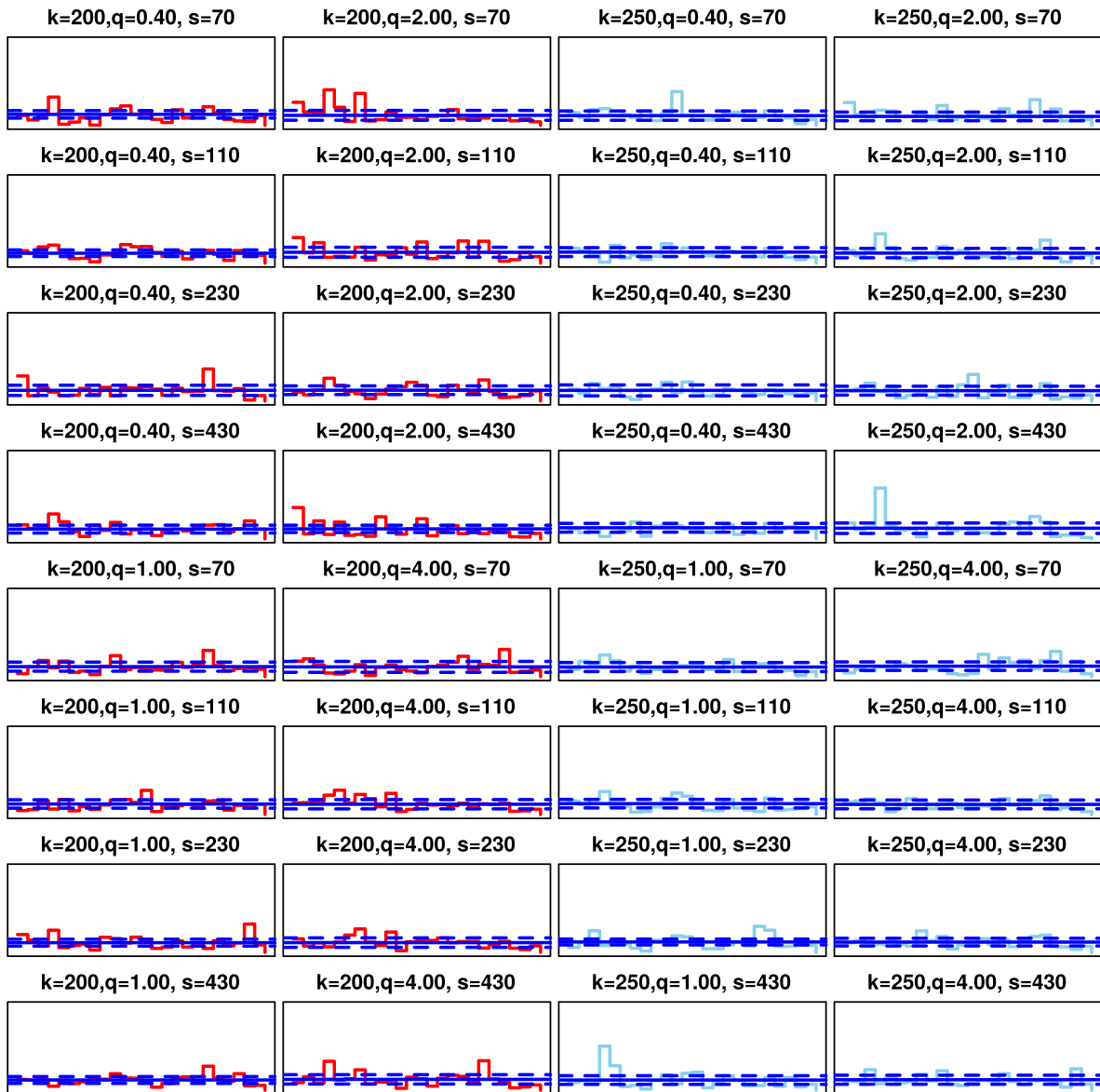
Figure 9. The contributions $X_k^l, l = \dots 24$ of each industry group to two eigenvalues fall deep inside the bulk region $\lambda_k, k = 200$ (red lines) and $\lambda_k, k = 250$ (sky blue lines) at different multifractal orders $q = 0.4, 1.0, 2.0, 4.0$ and detrending scales $s = 70, 110, 230, 430$ d. The blue solid and dashed lines are the mean X_k^l with one standard deviation for the cross-correlation matrices calculated from the shuffled return time series.

the heterogeneity of the BA network is 0.11. We notice that for small q, the heterogeneity index of the q-PMFG network is larger than the BA network. This means the structure of the q-PMFG networks for small multifractal orders q are extremely heterogeneous. We also show the assortativity A of the q-PMFGs at figure 14(d). The negative assortativity values for $q < 2$ provide a hint regarding the disassortative structures in which hub stocks tend to connect with small degree stocks. When $q > 2$, the assortativity approaches 0. This indicates that for large q, the connections
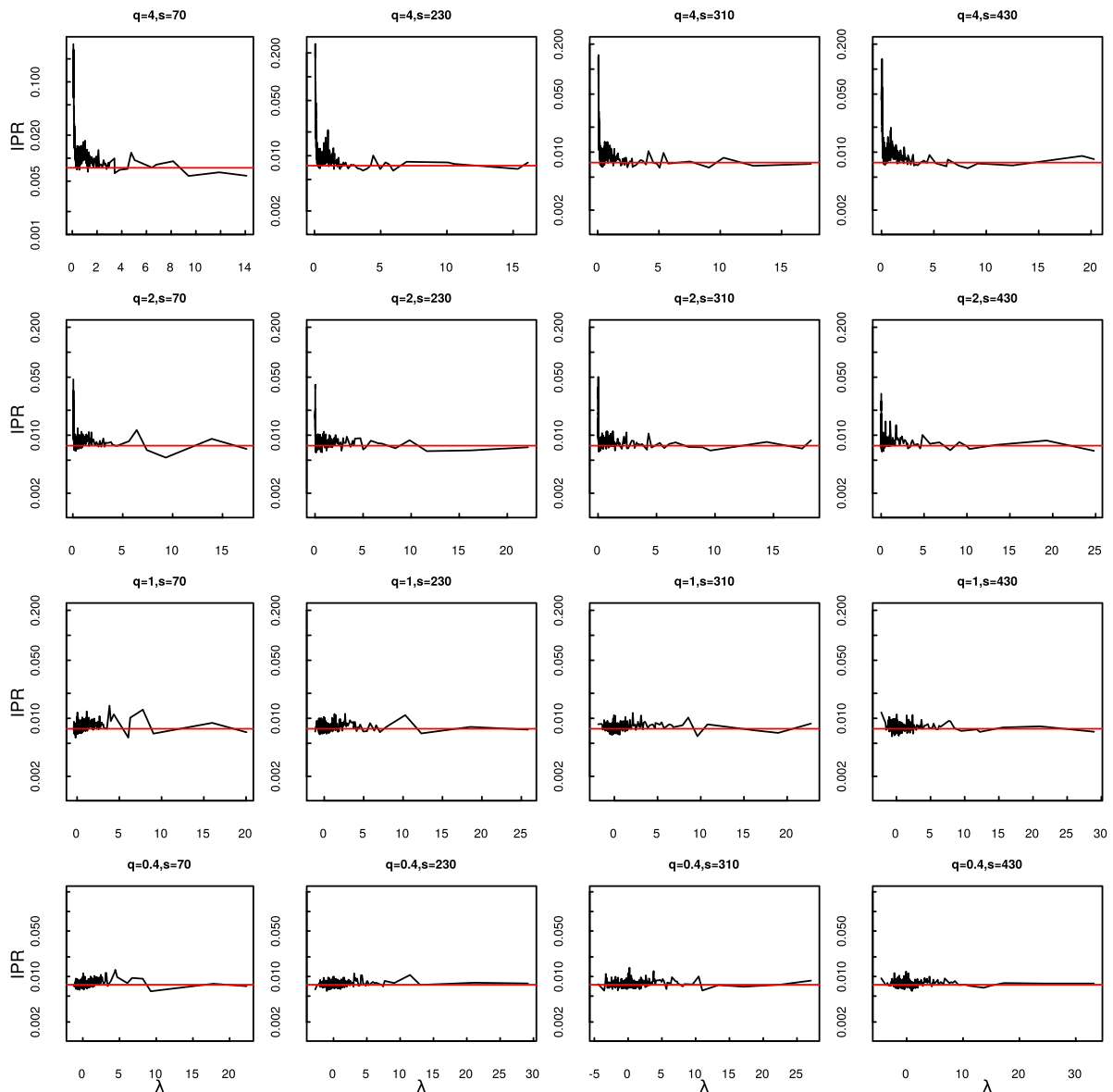
Figure 10. The inverse participation ratio (IPR) as a function of eigenvalues without the largest eigenvalue for different multifractal orders q and detrending scales s. The multifractal orders $q = 0.4, 1.0, 2.0, 4.0$ from bottom to top and the detrending scales $s = 70, 230, 310, 430$ d from left to right. The red line is the IPR for the random matrix with value $\langle I_k \rangle = 3/N$.

are more evenly distributed (see figure 13). In networks with $q > 2$, the degree of the hub stocks are smaller than those hubs in networks with $q < 2$. From the variation of topological quantities, we can infer that for small fluctuations (small q) at short time scales (small s), some leading stocks exist. But for large fluctuations (large q) and long time scales (large s), stocks are correlated uniformly. To sum up, from those topological quantities, an obvious structure transition is evident which provides an indication of the collective behavior differences among fluctuations of different magnitudes at varying time scales.
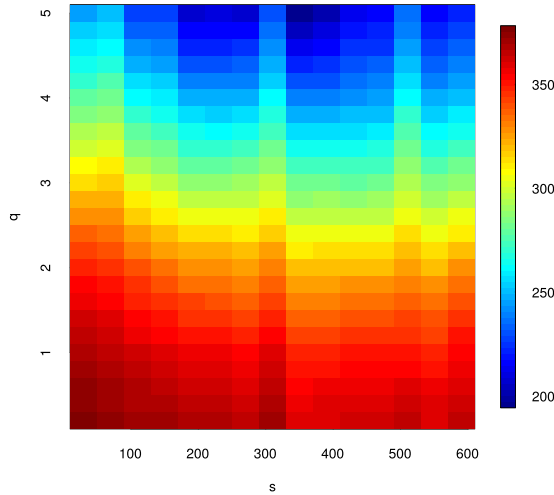
Figure 11. The participation ratio (PR) $1/I_k$ of the largest eigenvalue λ_{401} for different multifractal orders $q \in (0.2, 5)$ with step size $\delta_q = 0.2$ and detrending scales $s \in (30, 600)$ with step size $\delta_s = 40$ d.

4. Application

4.1. Mean-variance portfolio optimization

We now explore the possibility of using the q-PMFG networks to improve the performance of portfolio optimization under the Markowitz portfolio framework [47]. First, we briefly introduce the Markowitz portfolio theory and then we use a centrality metric to construct portfolios from the q-PMFG networks. Considering a portfolio $\Pi(t)$ of stocks with return $\{r_i(t), i = 1, \ldots, m; t = 1, \ldots, L\}$, m is the portfolio size, i.e. the number of stocks in the portfolio. The return $\Pi(t)$ of the portfolio is

$$\Pi(t) = \sum_{i=1}^{m} \omega_i r_i(t), \tag{17}$$

where ω_i is the fraction of wealth invested in stock i. The fractions ω_i are normalized such that $\sum_{i=1}^{m} \omega_i = 1$. The risk in holding the portfolio $\Pi(t)$ can be quantified by the return variance of the portfolio

$$\Omega^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \omega_i \omega_j C_{ij} \sigma_i \sigma_j, \tag{18}$$

where C_{ij} is the Pearson cross-correlation between the return series r_i and r_j, and σ_i and σ_j are the standard deviations of r_i and r_j. To find the optimal portfolio weights, we maximize the return of the portfolio $\Phi = \sum_{t=1}^{T} \Pi(t)$ under the constraint that the risk on the portfolio is some fixed value Ω^2. Maximizing Φ subject to those two constraints is equivalent to a quadratic optimization problem

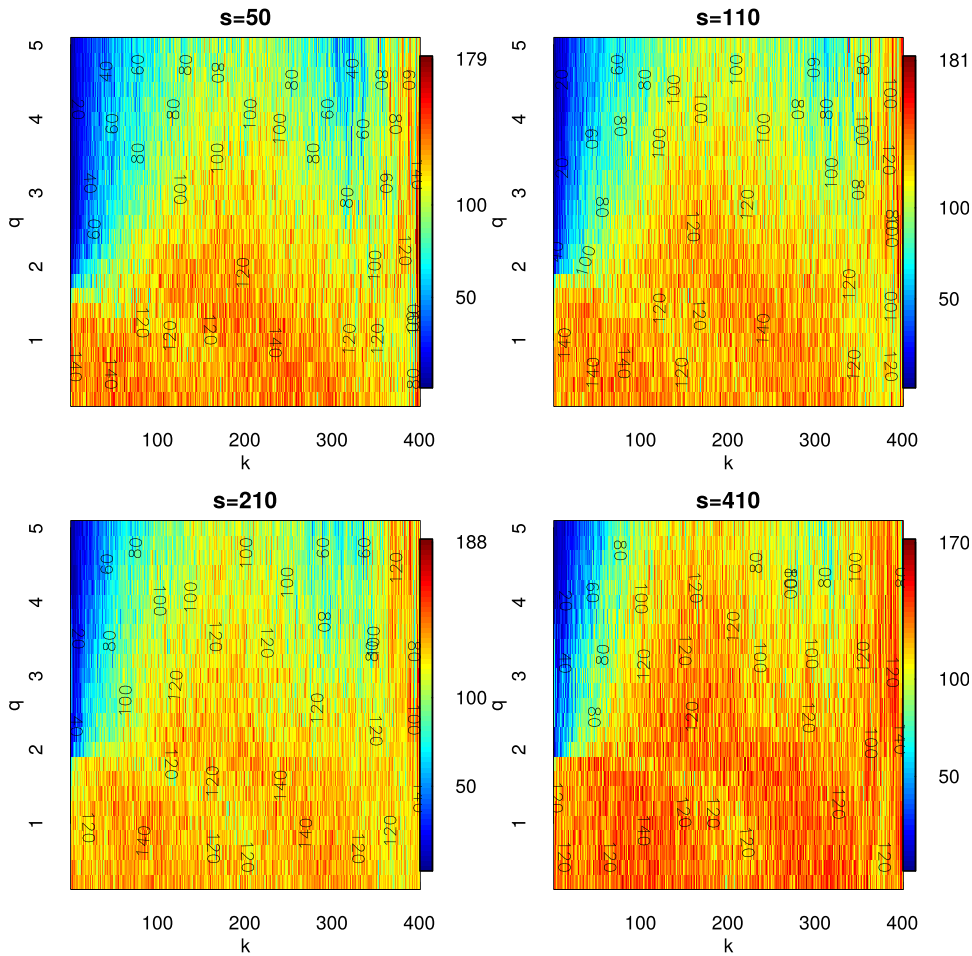$$\omega^T \Sigma \omega - q * \mathrm{R}^T \omega. \tag{19}$$

Figure 12. The PR $1/I_k$ as a function of $q \in (0.2, 5)$ with step size 0.2 and $k \in (1, 400)$ from left to right. k is the label of the eigenvalues λ_k. The PR of the largest eigenvalue is not shown here. We set the detrending scales $s = 50, 110, 210, 410$ d. The numbers inside the heat maps are marked to better distinguish the magnitudes of the corresponding PR values.

Here Σ is the covariance matrix of the return matrix R (mentioned in the previous context, now with dimension $L \times m$). The parameter q is the risk tolerance with $q \in [0, \infty)$. If we set large q we have strong tolerance to the risk, which will result in a large expected return. The optimal portfolios can be represented as a plot of the return Φ as a function of risk Ω^2 which is known as the efficient frontier. Because the positive definite property of the q-dependent covariance matrix Σ (or q-dependent cross-correlation matrix $\mathbf{C}(q, s)$) cannot be guaranteed when $q < 2$, we do not use the q-dependent cross-correlation matrices (or covariance matrices) in the risk metric Ω^2. Thus we try to use the q-dependent PMFG network topological properties to help us select m stocks, and then the traditional Markowitz portfolio theory (with the Pearson cross-correlation matrix $\mathbf{C}$ of those selected m return time series) is used to quantify the performance of the portfolio. In this sense, the constituent stocks for portfolios at different multifractal orders q and detrending scales s are very likely to be different. It has shown that a portfolio selected from the PMFG networks (with the Pearson cross-correlation matrix $\mathbf{C}$),
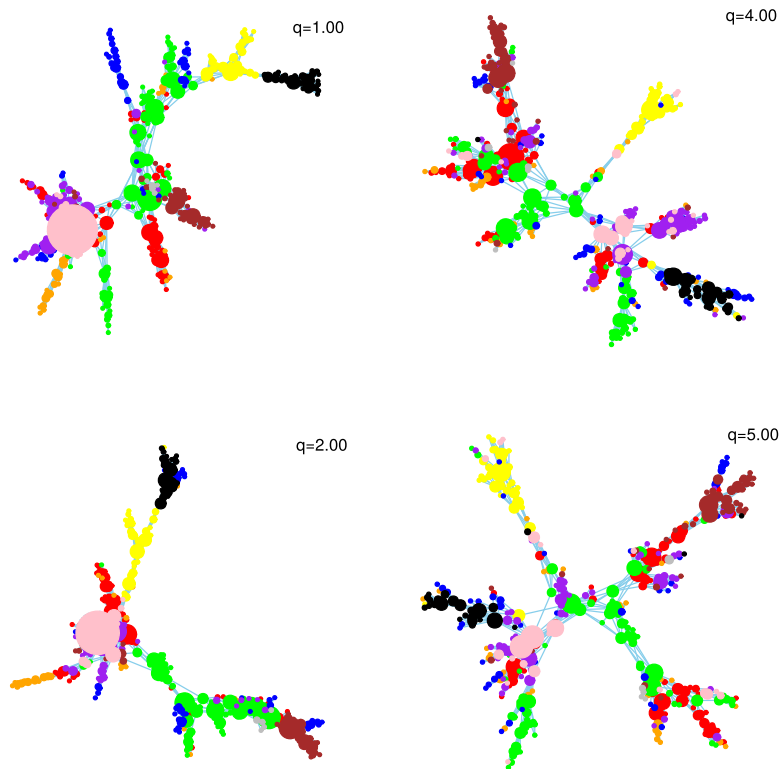
Figure 13. The q-PMFG networks for different multifractal orders $q = 1.0$, 2.0, 4.0, 5.0. We set the detrending scale $s = 110$ d here. Different vertex colors represent different sectors. The color table related to the sectors are: consumer discretionary—red, consumer staples—orange, energy—yellow, financials—green, health care—blue, industrials—purple, information technology—brown, materials—pink, telecommunications services—gray, utilities—black. The vertex size is proportional to the degree of the vertex.

by using some centrality measures, performed very well [48]. Here we first calculate the centrality scores defined by

$$\eta = \frac{C_{\mathrm{D}}^{w} + C_{\mathrm{D}}^{u} + C_{\mathrm{BC}}^{w} + C_{\mathrm{BC}}^{u} - 4}{4 \times (N-1)} + \frac{C_{\mathrm{E}}^{w} + C_{\mathrm{E}}^{u} + C_{\mathrm{C}}^{w} + C_{\mathrm{C}}^{u} + C_{\mathrm{EC}}^{w} + C_{\mathrm{EC}}^{u} - 6}{6 \times (N-1)},$$
(20)

where C_{D}^{w} is the weighted (superscript w) degree (D) centrality and C_{D}^{u} is the unweighted (superscript u) counterpart. The other centrality metrics are betweenness centrality (BC), eccentricity (E), closeness (C), and eigenvector centrality (EC). A portfolio constructed by using the central (peripheral) stocks are those with a very high (low) centrality value η. This composed centrality measure is constructed in this way because different centrality measures are not independent. With a large amount of numerical tests, ten centrality measures have been found to fall into two groups with extremely high correlation within each group. A complete description of this composed centrality metric is provided in [48]. Actually, the choice of the centrality metric does not significantly affect the final results.

In figure 15, we show the efficient frontiers calculated from those portfolios constructed by using central (black lines), peripheral (red lines) and random selected (blue lines) stocks with different multifractal orders q. We have conducted the tests using the
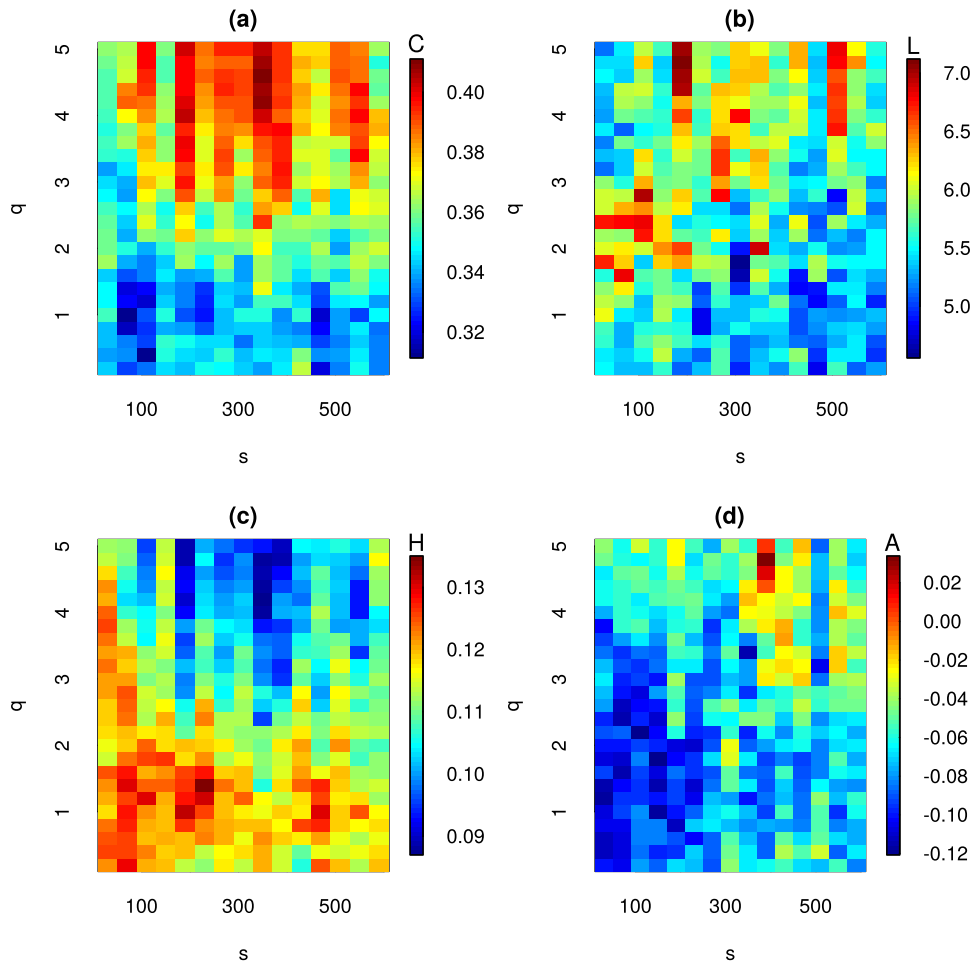
Figure 14. The topological quantities of the q-PMFGs at different multifractal orders $q \in (0.2, 5)$ with step $\delta_q = 0.2$ and scales $s \in (30, 600)$ with step $\delta_s = 40$ d. (a) The clustering coefficient C, (b) the shortest path length L, (c) the heterogeneity index H, and (d) the assortativity A.

portfolio sizes $m = 10, 20, 30, 40, 50, 60$ and calculated the average return value for all the detrending scales s at one specific risk value Ω^2 to reveal the effect of fluctuations at different magnitudes. It is very clear that for different portfolio sizes from $m = 10$ to $m = 60$ stocks, the peripheral portfolios (red lines) are always the best performing ones. In contrast, the performance of the central portfolios is even worse than the random portfolios.

We then calculate the return differences between peripheral and central portfolios $\Delta = \Phi_p - \Phi_c$ (p and c are peripheral and central portfolios, respectively) as a function of the multifractal orders q in figure 16. Here we use the multifractal orders q from 0.2–10 to identify the optimal q. It is very evident that the peripheral portfolios outperform the central portfolios most around multifractal order $q = 2$ and exhibit a superiority greater than 7%. This provides a hint that we should trade based on moderate fluctuations which will lead to higher returns and lower risks. The results above indicate the potential of utilizing the q-dependent cross-correlation matrix and q-PMFG network as new portfolio optimization tools.
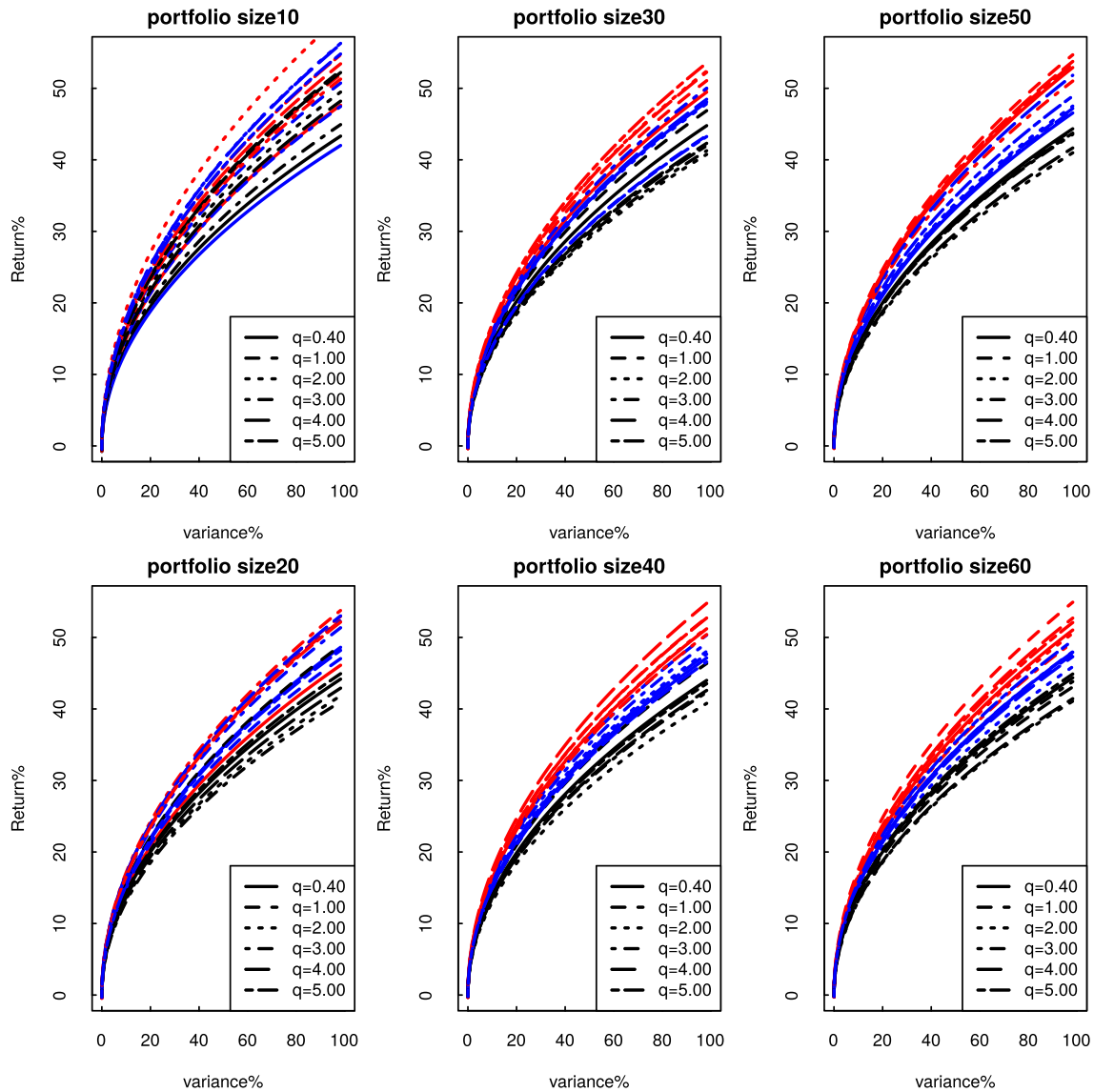
Figure 15. The efficient frontiers for different portfolio sizes $m = 10, 20, 30, 40, 50, 60$ stocks. The red, blue and black lines are efficient frontiers for those portfolios constructed by using peripheral, random selected and central stocks in the q-PMFG networks. The effect of the detrending scales $s \in (30, 600)$ with step size $\delta_s = 40$ d has been averaged during the portfolio optimization procedure.

4.2. Expected shortfall approach

Except the mean-variance framework, the expected shortfall (ES) is a modern tool to quantify the performance of a portfolio. The ES is a coherent risk measure [49–51]. It meets all the requirements of a risk measure. Thus we also employ the ES to quantify the performance of those portfolios constructed with the guidance of the q-PMFG networks. Let X be the profit loss of a portfolio on a specified time horizon T and let $\alpha = A\% \in (0, 1)$ be some specified probability level. The expected $A\%$ shortfall of the portfolio is then defined as [49]
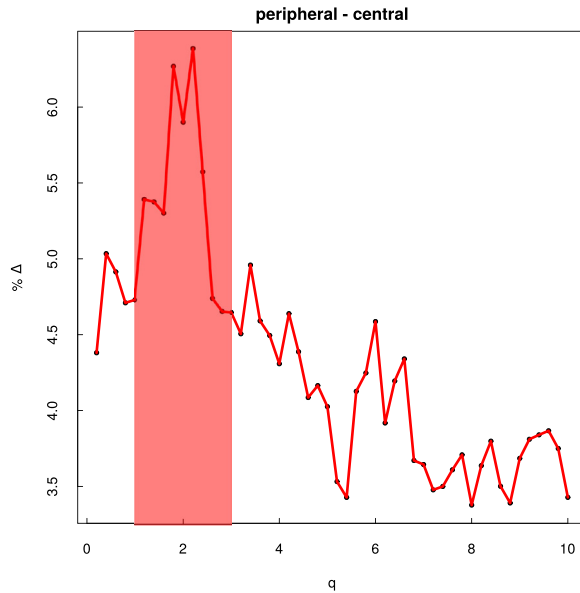
Figure 16. The differences between the returns of peripheral portfolios and the returns of central portfolios as a function of the multifractal orders $q \in (0.2, 10)$ with step size $\delta_q = 0.2$.
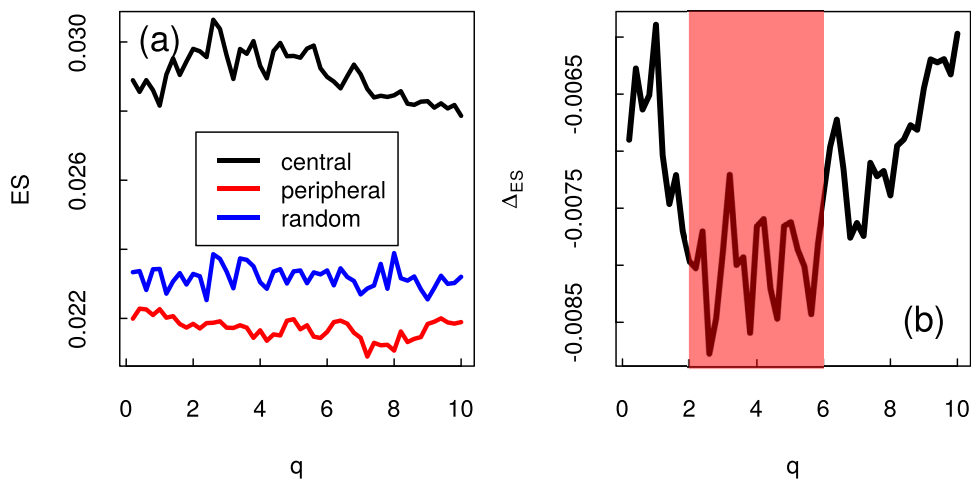

Figure 17. (a) The expected shortfall $\mathrm{ES}(\alpha)$ for central, peripheral and random portfolios at different multifractal orders $q \in (0.2, 10)$ with step size $\delta_q = 0.2$ and $\alpha = 0.95$. (b) The differences between the $\mathrm{ES}(\alpha)$ of the portfolios for central and peripheral stocks. Here the portfolio size $m = 30$ stocks and we have averaged the $\mathrm{ES}(\alpha)$ for different detrending scales $s \in (30, 600)$ with step size $\delta_s = 40$ d.

$$\mathrm{ES}^{\alpha}(X) = -\frac{1}{\alpha}\left(E[X 1_{X \leqslant x^{\alpha}}] - x^{\alpha}(\mathrm{P}[X \leqslant x^{\alpha}] - \alpha)\right). \qquad (21)$$

It gives the expected loss incurred in the $A\%$ worst cases of the portfolio. For a portfolio $\{\omega_i, i = 1, \ldots, m\}$ of m stocks with return time series $\{r_i, i = 1 \ldots m\}$, we want to minimize the $\mathrm{ES}(\alpha)$ of the portfolio under the constraint of normalization $\sum_{i=1}^{m} \omega_i = 1$. Here we use the expected shortfall of the portfolio $\mathrm{ES}(\alpha)$ with confidence level $\alpha = 0.95$,

and the short selling is prohibited. After ranking the stocks with centrality scores η described in the previous subsection, we choose the portfolio size $m = 30$, namely, 30 central (peripheral) stocks with the largest (smallest) centrality scores. Meanwhile, a portfolio with 30 randomly chosen stocks is also used here as a reference. Figure 17(a) shows the expected shortfall for central, peripheral and random portfolios. Obviously the central portfolios have the highest expected shortfalls. However, the peripheral portfolios' ES is much lower than the central ones and even lower than the random portfolios. This confirms that the peripheral stocks have relatively low risk compared to the central and random selected ones. In figure 17(b), we provide the differences between the expected shortfalls for peripheral and central portfolios $\Delta_{ES} = ES_p - ES_c$ for different multifractal order q. Under this performance measure, the optimal multifractal order q ranges from 2–6 which is different from the result given by the mean-variance framework.

5. Conclusion and discussion

In this paper, we have employed the q-dependent cross-correlation coefficient to analyze the cross-correlations among fluctuations at different magnitudes for the stock market. With the help of random matrix theory and complex network theory, we analyze the cross-correlation matrices of the stock market for different magnitudes of fluctuations. We find that the cross-correlations among small fluctuations are stronger than large ones. There are more deviating eigenvalues for the cross-correlation matrices of large fluctuations than that of small fluctuations. By analyzing the IPR and the eigenvector contribution, we find that the small eigenvalues of the cross-correlation matrices for large fluctuations are dominated by a small number of industry groups. This is similar to those large deviating eigenvalues that are also dominated by a small number of industry groups. Thus we conclude that small eigenvalues of the q-dependent cross-correlation matrices also carry some genuine information, which seems very counterintuitive. We also find that the large and small fluctuations are dominated by different groups of stocks or sectors. The structure variation of the q-dependent cross-correlation matrices reveal the different identities of the cross-correlations for small and large fluctuations. The multifractal order q acts as a filter which enlarges the effect of different magnitudes of fluctuations. In [52, 53], a similar method, named power-mapping, is used as a filter to suppress the noise which can 'prolong' the time series. The traditional Pearson cross-correlation matrix is powered by an index $q > 1$. But from the definition of the q-dependent cross-correlation coefficient, the detrended covariance has been emphasized by the multifractal order q. Checking the equivalence between these two methods is not so obvious. However, in [53], different factor models and real time series are used to prove the effectiveness of the method in which the authors find that the optimal order q ranges from 1.5–1.8. Thus it should be very interesting to test the q-dependent cross-correlation coefficient with factor models and datasets from other markets. In this work, the complex network representation has also validated the correlation difference between small and large fluctuations. The network structures are more heterogeneous and disassortative for the networks constructed from small fluctuations which means

the existence of leading stocks. We then utilize the network centrality as a portfolio selection metric. Under the mean-variance portfolio framework, we find that the portfolios of the peripheral stocks always outperform the portfolios of central stocks. Optimal multifractal order with the largest return difference approaching %7 is around $q = 2$. Then we also employ a modern and coherent risk measure named the expected shortfall to test the portfolios constructed with the guidance of the q-PMFG networks. The outcome is that the portfolios consisting of those peripheral stocks have the lowest risk compared to the portfolios of the central and randomly selected stocks. The optimal multifractal order q ranges from 2–6 under the expected shortfall measure. Those results may be used as new portfolio optimization and risk management tools. Thus our investigations regarding the cross-correlations among stocks with different magnitudes of fluctuations have demonstrated the huge difference between large and small fluctuations of the stock market. They are regulated by different non-linear correlation structures. Those results expand our understanding regarding the collective behavior of the stock market.

Acknowledgments

ORCID iDs

Longfeng Zhao ● https://orcid.org/0000-0002-0642-3798

References

[1] Plerou V, Gopikrishnan P, Rosenow B, Nunes Amaral L A and Stanley H E 1999 *Phys. Rev. Lett.* **83** 1471–4
[2] Laloux L, Cizeau P, Bouchaud J P and Potters M 1999 *Phys. Rev. Lett.* **83** 1467–70
[3] Kwapień J and Drożdż S 2012 *Phys. Rep.* **515** 115–226
[4] Peng C K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 *Phys. Rev. E* **49** 1685–9
[5] Podobnik B and Stanley H E 2008 *Phys. Rev. Lett.* **100** 084102
[6] Kantelhardt J W, Zschiegner S A, Koscielny-Bunde E, Havlin S, Bunde A and Stanley H E 2002 *Physica* A **316** 87–114
[7] Zhou W X 2008 *Phys. Rev. E* **77** 066211
[8] Oświęcimka P, Drożdż S, Forczek M, Jadach S and Kwapień J 2014 *Phys. Rev. E* **89** 023305
[9] Alvarez-Ramirez J, Rodriguez E and Echeverria J C 2009 *Phys. Rev. E* **79** 057202
[10] Podobnik B, Jiang Z Q, Zhou W X and Stanley H E 2011 *Phys. Rev. E* **84** 066118
[11] Stan C, Cristescu M T, Luiza B I and Cristescu C 2013 *J. Theor. Biol.* **321** 54–62
[12] Rak R, Drożdż S, Kwapień J and Oświęcimka P 2015 *Europhys. Lett.* **112** 48001
[13] Kristoufek L 2015 *Phys. Rev. E* **91** 022802

[14] Oświęcimka P, Livi L and Drożdż S 2016 *Phys. Rev.* E **94** 042307
[15] Salat H, Murcio R and Arcaute E 2017 *Physica* A **473** 467–87
[16] Kwapień J, Oświęcimka P, Forczek M and Drożdż S 2017 *Phys. Rev.* E **95** 052313
[17] Zhao L, Li W, Yang C, Han J, Su Z and Zou Y 2017 *PLoS One* **12** 1–23
[18] Zebende G 2011 *Physica* A **390** 614–8
[19] Kristoufek L 2014 *Physica* A **402** 291–8
[20] Wang G J, Xie C, Chen S, Yang J J and Yang M Y 2013 *Physica* A **392** 3715–30
[21] Wang G J, Xie C, Chen Y J and Chen S 2013 *Entropy* **15** 1643–62
[22] Zebende G, da Silva M and Filho A M 2013 *Physica* A **392** 1756–61
[23] Sun X and Liu Z 2016 *Physica* A **444** 667–79
[24] Kwapień J, Oświęcimka P and Drożdż S 2015 *Phys. Rev.* E **92** 052815
[25] Tumminello M, Aste T, Di Matteo T and Mantegna R N 2005 *Proc. Natl Acad. Sci. USA* **102** 10421–6
[26] Marčenko V A and Pastur L A 1967 *Math. USSR-Sb.* **1** 457
[27] Gopikrishnan P, Rosenow B, Plerou V and Stanley H E 2001 *Phys. Rev.* E **64** 035106
[28] Rosenow B, Plerou V, Gopikrishnan P and Stanley H E 2002 *Europhys. Lett.* **59** 500
[29] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N, Guhr T and Stanley H E 2002 *Phys. Rev.* E **65** 066126
[30] Podobnik B, Wang D, Horvatic D, Grosse I and Stanley H E 2010 *Europhys. Lett.* **90** 68001
[31] Zhou W X, Mu G H and Kertész J 2012 *New J. Phys.* **14** 093025
[32] Livan G, Alfarano S and Scalas E 2011 *Phys. Rev.* E **84** 016113
[33] Fenn D J, Porter M A, Williams S, McDonald M, Johnson N F and Jones N S 2011 *Phys. Rev.* E **84** 026109
[34] Wang D, Podobnik B, Horvatić D and Stanley H E 2011 *Phys. Rev.* E **83** 046121
[35] Singh A and Xu D 2016 *Quant. Finance* **16** 69–83
[36] Bun J, Bouchaud J P and Potters M 2017 *Phys. Rep.* **666** 1–109
[37] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D U 2006 *Phys. Rep.* **424** 175–308
[38] Estrada E 2010 *Phys. Rev.* E **82** 066102
[39] www.msci.com/gics (Accessed: 11 December 2017)
[40] Marti G, Nielsen F, Bińkowski M and Donnat P 2017 in preparation (arXiv:1703.00485v2)
[41] Cai S M, Zhou Y B, Zhou T and Zhou P L 2010 *Int. J. Mod. Phys.* C **21** 433–41
[42] Gao Y C, Zeng Y and Cai S M 2015 *J. Stat. Mech.* P03017
[43] Gan S L and Djauhari M A 2015 *J. Stat. Mech.* P12005
[44] Djauhari M A and Gan S L 2016 *J. Stat. Mech.* 093401
[45] Song D M, Tumminello M, Zhou W X and Mantegna R N 2011 *Phys. Rev.* E **84** 026108
[46] Zhao L, Li W and Cai X 2016 *Phys. Lett.* A **380** 654–66
[47] Markowitz H 1952 *J. Finance* **7** 77–91
[48] Pozzi F, Di Matteo T and Aste T 2013 *Sci. Rep.* **3** 1665
[49] Acerbi C and Tasche D 2002 *Econ. Notes* **31** 379–88
[50] Caccioli F, Still S, Marsili M and Kondor I 2013 *Eur. J. Finance* **19** 554–71
[51] Caccioli F, Kondor I, Marsili M and Still S 2016 *Int. J. Theor. Appl. Finance* **19** 1650035
[52] Guhr T and Kälber B 2003 *J. Phys. A: Math. Gen.* **36** 3009
[53] Schäfer R, Nilsson N F and Guhr T 2010 *Quant. Finance* **10** 107–19