

Spectral Properties of Financial Correlation Matrices

Maxim Kazakov and Valery A. Kalyagin

Abstract Random matrix theory (RMT) is applied to investigate the cross-correlation matrix of a financial time series in four different stock markets: Russian, American, German, and Chinese. The deviations of distribution of eigenvalues of market correlation matrix from RMT global regime are investigated. Specific properties of each market are observed and discussed.

Keywords Random matrices • Wishart–Laguerre ensemble • Stock market • Market network • Random Graphs

1 Introduction

The study of correlation (or covariance) matrices has a long history in finance and it is an important aspect of risk management and one of the cornerstone of Markowitz's theory of optimal portfolios [6, 10]. Besides, equal-time cross-correlation matrices play a major role in market network analysis when it comes to constructing different network structures, such as maximum spanning tree or market graph [1, 2].

When stock market consists of several hundreds of individual stocks, it becomes a high-dimensional and complex system. To study these systems some methods of statistical physics have been employed, in particular, random matrix theory (RMT) [3, 8, 11, 13]. The idea is to compare the properties of an empirical correlation matrix to the ones of purely *random* matrix. In case of covariance matrix ensemble of such random matrices is called Wishart–Laguerre ensemble [5]. Possible deviations from the random case may reveal some peculiarities of empirical correlation matrices and it may give some insight into the market structure.

M. Kazakov (✉)

National Research University Higher School of Economics, N. Novgorod, Russia
e-mail: max.a.kazakov@gmail.com

V.A. Kalyagin

Lab LATNA, National Research University Higher School of Economics, N. Novgorod, Russia
e-mail: vkalyagin@hse.ru

The main goal of the present paper is a comparative investigation of empirical correlation matrices for different markets. The main problem addressed is to understand whether the markets are different from RMT point of view and to point out these differences. We investigate four different markets corresponding to different levels of economic development: the US, German, Russian, and Chinese. We analyze spectral properties of empirical correlation matrices and compare them to global regimes provided by RMT. In addition, we test the stability of observed deviations and their dependence on the distribution of the data.

The paper is organized as follows. In Sect. 2 we remind the main facts from RMT and discuss the results of previous related studies. In Sect. 3 we present our methods and describe the data used in numerical experiments. In Sect. 4 we conduct a comparative analysis of correlations matrices for indicated markets. Section 5 is devoted to a stability analysis of observed phenomena. Section 6 contains concluding remarks.

2 Theoretical Background

2.1 Random Matrix Theory

We want to compare spectral properties of empirical correlation matrices of stock market with the spectral properties of random matrices. In case of covariance (or correlation) matrices this is so-called Wishart–Laguerre ensemble [5]. Consider rectangular ($N \times T$) matrix H whose elements $H_{i,t}$ are independent, identically distributed random variables. Then the product $W = (1/T) \cdot H \cdot H^*$ is a positive definite symmetric ($N \times N$) matrix that represents the normalized covariance matrix of the data. When elements $H_{i,t}$ are drawn from a Gaussian distribution, the product matrices $W = \frac{1}{T} \cdot H \cdot H^*$ constitute Wishart–Laguerre ensemble of random matrices.

For the case when $T \geq N$ (the number of samples is larger than the dimension) the spectral properties of these matrices are well studied and it is known that in limit ($N \rightarrow \infty$ and $T \rightarrow \infty$ and $Q = T/N \geq 1$ fixed) all eigenvalues are positive and density distribution of the eigenvalues is given by the Marchenko–Pastur function [9, 14]:

$$\rho_{WL}(\lambda) = \frac{Q}{2\pi} \cdot \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \quad \lambda_- < \lambda < \lambda_+, \quad (1)$$

where the lower and upper bounds of eigenvalues are calculated as follows:

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \quad (2)$$

Note that above results are valid only in limit when $N \rightarrow \infty$.

2.2 Related Works

Recently series of studies has been conducted [3, 8, 11, 13] to analyze spectral properties of empirical correlation matrices and compare them to RMT global regime discussed in previous section. Following observations have been made:

- There is one largest eigenvalue λ_{max} , which is significantly higher than the upper bound λ_+ . It is also tends to be relatively close to the product $N \cdot \overline{C}$, where \overline{C} is the average of non-diagonal elements of correlation matrix C . The associated eigenvector is connected with global market index.
- There are also several eigenvalues slightly greater than λ_+ . They may reflect sector behavior.
- There are a number of eigenvalues below the lower bound λ_- , which can be explained by repulsion effect which we will talk about later. It may also correspond particularly to highly correlated pair of stocks.
- Finally, most of the eigenvalues fall within a range predicted by RMT. These eigenvalues are called bulk of eigenvalue spectrum. Nonetheless, it was shown that these eigenvalues also may contain useful information [7].

These results may differ for emerging markets [4, 12]. Such as, in emerging markets the largest eigenvalue appears to be higher with respect to λ_+ and there are fewer eigenvalues above the edge. At the same time, there is a large proportion of eigenvalues below λ_- and, consequently, less number of eigenvalues in the bulk. Also, average value of non-diagonal elements of correlation matrix is higher and fluctuates more dynamically.

3 Method and Data

3.1 Method

We consider a set of N stocks over a period of T trading days. Let $P_i(t)$ be a closing price of stock i ($i = 1, \dots, N$) in the day t ($t = 1, \dots, T$). Then the daily log return $R_i(t)$ of stock i is defined by

$$R_i(t) = \ln \frac{P_i(t)}{P_i(t-1)}. \quad (3)$$

We normalize R_i with respect to its standard deviation σ_i as follows:

$$r_i(t) = \frac{R_i(t) - \overline{R_i}}{\sigma_i}, \quad (4)$$

where \overline{R}_i denotes the average return over the period studied and standard deviation (or volatility) defined as $\sigma_i = \sqrt{\overline{R_i^2} - \overline{R}_i^2}$.

Then, the equal-time cross-correlation matrix C is expressed it terms of $r_i(t)$:

$$C_{i,j} = \sum_{t=1}^T r_i(t) \cdot r_j(t). \quad (5)$$

The element $C_{i,j}$ of matrix C denotes correlation coefficient between stock i and stock j . Correlation matrix C also can be expressed in matrix notation as

$$C = \frac{1}{T} \cdot R \cdot R^T, \quad (6)$$

where R is an $(N \times T)$ matrix with elements $r_i(t)$.

The N eigenvalues λ_i and their corresponding eigenvectors u_i are calculated by diagonalizing C . One has

$$C \cdot u_i = \lambda_i \cdot u_i, \quad i = 1, \dots, N. \quad (7)$$

Note that $\sum \lambda_i$ is always equal to sum of the diagonal elements of C (the trace), which is always constant and equal to N since for all elements $C_{i,i} = 1$. Hence, if some eigenvalues increase, then some others must decrease to compensate, and vice versa. This is called *eigenvalue repulsion* [3].

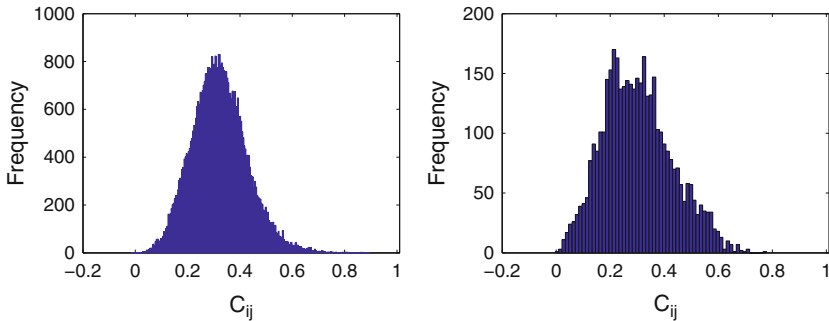
3.2 Data

In order to analyze spectral properties of empirical financial correlation matrices we consider four different stock markets, representing different types of economies: the US, Russian, German, and Chinese stock market. For Russian market we consider stocks traded on The Moscow Interbank Currency Exchange (MICEX). For American market we consider equities of S&P 500 traded on The New York Stock Exchange (NYSE). For German market we consider equities of HDAX traded on The Frankfurt Stock Exchange (FWB). And for Chinese market we consider stocks traded on The Hong Kong Stock Exchange (HKEx).

We want $Q = T/N$ to be relatively equal for all markets and we eliminate stocks if they haven't been traded long enough. For Russian market we also apply cleaning procedure in order to eliminate stocks with low liquidity. One exception here is an American market. In this case we allow Q to be essentially smaller than in other markets so we can apply our method to larger data set. Dates and the number of chosen stocks of each market are summarized in Table 1.

Table 1 Characteristics of considered markets

Market	Number of stocks, N	Length of time series, T	$Q = T/N$	Starting date	Ending date
Russia	101	1418	14.04	10/01/2008	06/06/2014
USA	316	3008	9.52	01/03/2003	12/12/2014
Germany	90	1282	14.24	01/05/2010	12/12/2014
China	78	1016	13.03	01/03/2011	12/12/2014

**Fig. 1** Distribution of correlations. *Left*—the US market. *Right*—German market

4 Comparative Analysis of Different Markets

In this section we present the results of the analysis of empirical correlation matrices for four different stock markets. We compare the empirical distribution of eigenvalue with predictions of RMT and discuss some deviations.

4.1 Distribution of the Correlation Coefficients

First, we take a look at the statistical properties of empirical cross-correlation matrices. Figures 1 and 2 show histograms of correlation coefficients (i.e., non-diagonal elements of correlation matrix C) for all four markets. Other comparative characteristics are given in Table 2.

We notice that average value \bar{C} is quite large for all cases. The interesting fact here is that it is almost the same and around 0.3 for all considered markets, except Russian. Furthermore, standard deviations are also relatively high and close to each other, this time including Russian market. For American, German, and Chinese markets almost all elements of correlation matrix are positive.

Next we test the assumption of normal distribution of cross-correlation matrix elements. Histograms on Figs. 1 and 2 don't show distribution similar to normal (or maybe just for American market). Lilliefors test rejected hypothesis of normal distribution at the 5% significance level for all markets. We also use skewness and

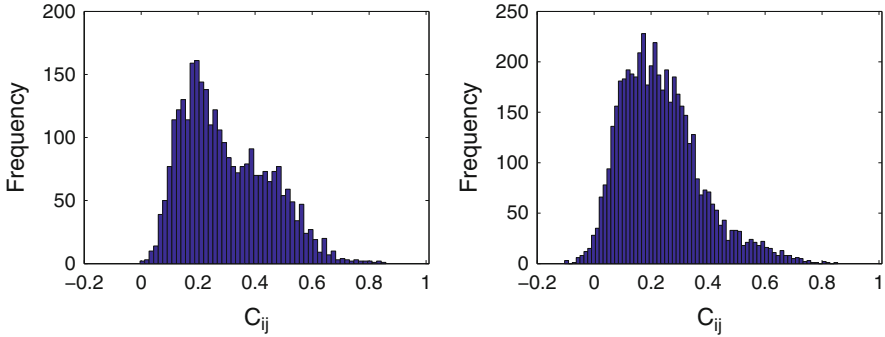


Fig. 2 Distribution of correlations. *Left*—Chinese market. *Right*—Russian market

Table 2 Statistics for cross-correlation

Market	Mean	Standard deviation	Skewness	Kurtosis
USA	0.3220	0.1060	0.3957	3.4929
Germany	0.3021	0.1282	0.3846	2.7806
China	0.2984	0.1514	0.5968	2.6833
Russia	0.2349	0.1408	0.8391	3.8482

kurtosis measures to see how much the deviations are. Skewness is a measure of the asymmetry of the data around the sample mean and kurtosis is a measure of how outlier-prone a distribution is (respectively, 0 and 3 for the normal distribution). As shown in Table 2 for all markets skewness is positive which indicates that correlations are skewed right meaning that the right tail is long with respect to the left tail. Kurtosis measure, in contrast, deviates in different directions, indicating more peaked distribution for American and Russian markets, and more flat distribution for German and Chinese. The deviations are relatively small though.

4.2 Eigenvalue Distribution

In this section we analyze spectral properties of empirical cross-correlation matrices and compare them to the predictions of RMT given by formulas (1) and (2). The eigenvalue spectrum is shown in Figs. 3b, 4b, 5b, and 6b with the spectrum predicted by RMT in Figs. 3a, 4a, 5a, and 6a. Table 3 presents the more detailed characteristics.

As in the previous studies, we found that there is one largest eigenvalue λ_{max} in every case which exceeds significantly the upper bound λ_+ . We also noticed the similarity between λ_{max} and $N \cdot \bar{C}$ presented in Table 3 by their ratio close to the value of 1. This explains the exceptionally large value of λ_{max} for American market with respect to the others: since average value of correlation is similar for each market

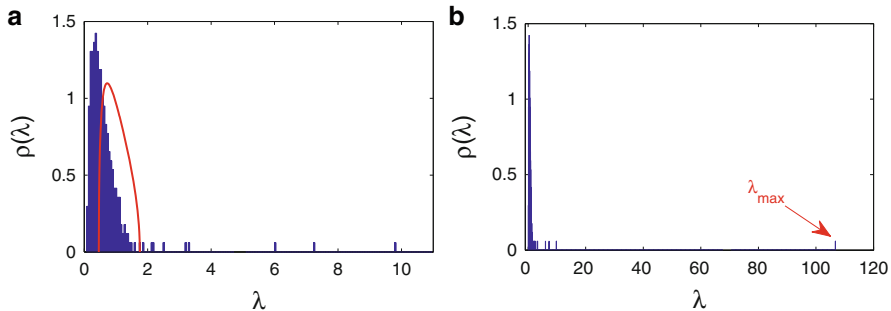


Fig. 3 American market. (a) Probability density of λ in comparison with RMT density (the red solid line) and (b) including the largest eigenvalue λ_{max}

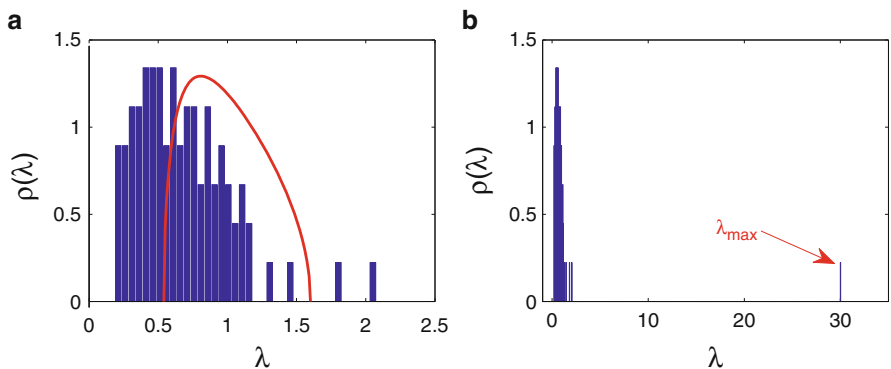


Fig. 4 German market. (a) Probability density of λ in comparison with RMT density (the red solid line) and (b) including the largest eigenvalue λ_{max}

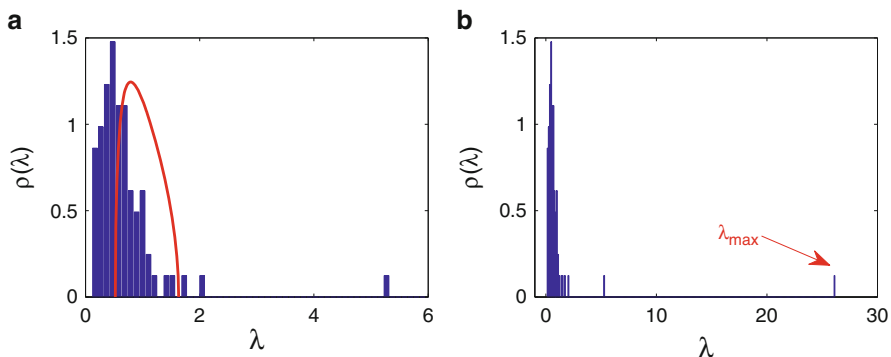


Fig. 5 Chinese market. (a) Probability density of λ in comparison with RMT density (the red solid line) and (b) including the largest eigenvalue λ_{max}

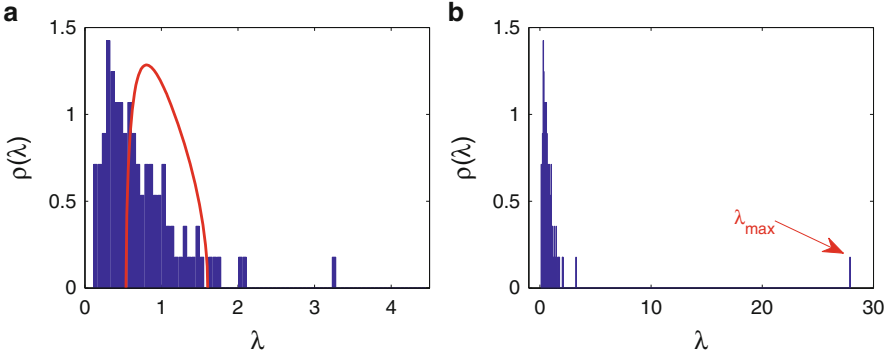


Fig. 6 Russian market. (a) Probability density of λ in comparison with RMT density (the red solid line) and (b) including the largest eigenvalue λ_{max}

Table 3 Eigenvalues statistics

Market	N	\bar{C}	λ_-	λ_+	λ_{max}	$\frac{\lambda_{max}}{\lambda_+}$	$\frac{\lambda_{max}}{N \cdot \bar{C}}$	$\lambda > \lambda_+$
USA	316	0.3220	0.4568	1.7533	106.7823	60.9039	1.0495	10
Germany	90	0.3021	0.5403	1.6001	30.0333	18.7694	1.1046	3
China	78	0.2984	0.5226	1.6309	26.1665	16.0440	1.1240	4
Russia	101	0.2349	0.5375	1.6050	27.9230	17.3975	1.1772	7

and American market is presented by data set greater by 3–4 times (with respect to the number of stocks), the value of the largest eigenvalue is also greater by 3–4 times.

The number of eigenvalues above the edge λ_+ differs for considered markets. In German and Chinese markets there are, respectively, 2 and 3 such eigenvalues, besides λ_{max} , which is small and in accordance with previous studies [11]. In American and Russian markets this number is relatively high (9 and 6, respectively) and for American market it is greater than what was observed before [8].

Furthermore, we noticed that about half of the eigenvalues falls into the range $[\lambda_-, \lambda_+]$ predicted by RMT. A little less number of eigenvalues fall below the edge λ_- . Most of this may be explained by eigenvalue repulsion effect we talked about in Sect. 3. These observations also support some previous results [11].

5 Stability Analysis

In this section we present the results of analysis of stability of observed phenomena. We want to see whether the observed deviations from RMT predictions are specific for a certain market or not. In order to do this we use bootstrap method. We

Table 4 Characteristics

Market	\bar{C}	λ_{max}	$\frac{\lambda_{max}}{N \cdot \bar{C}}$	$\lambda > \lambda_+$
USA	0.3220	106.7823	1.0495	10
Germany	0.3021	30.0333	1.1046	3
China	0.2984	26.1665	1.1240	4
Russia	0.2349	27.9230	1.1722	7

also test dependence of the deviations on type of distribution of the data using multivariate normal distribution and multivariate Student distribution. We test following characteristics:

- The mean value of correlation coefficient \bar{C}
- The value of the largest eigenvalue λ_{max}
- The ratio $\lambda_{max}/(N \cdot \bar{C})$, where N denotes the number of stocks
- The number of eigenvalues above the upper bound λ_+ predicted by RMT

These characteristics are summarized for all four considered markets in Table 4.

5.1 Bootstrapping

To test the stability of observed characteristics and, consequently, their deviations from predictions of RMT we apply the bootstrap method. First, we resample the data with replacement, saving the size of the resample ($N \times T$) the same as it was in the original data set. Note that sample here is a vector R_t corresponding to a trade day t characterized by daily returns of N stocks. Next we apply our method, defined by formulas (4)–(7), to compute characteristics of interest. We repeat this routine 10,000 times for each considered market.

Figures 7, 8, 9, and 10 present histograms of analyzed characteristics which provide an estimate of the shape of the distribution. We found that almost all of them are stable for each market indicating that considered deviations from RMT predictions are specific for empirical correlation matrices. One exception here is the number of eigenvalues above the upper bound λ_+ . For German and Chinese markets the value is quite robust (Fig. 10b, c), but for American and Russian cases results show that the observed values are not reliable (Fig. 10a, d). The surprising result is that the test revealed a greater number of those eigenvalues (about 12 for the USA and 9 for Russia in average).

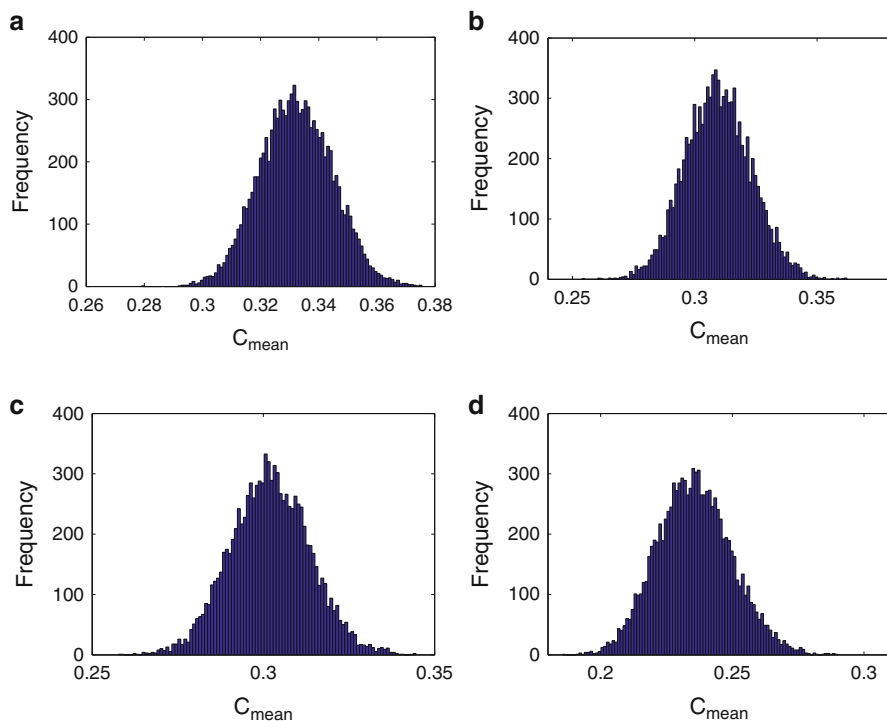


Fig. 7 Estimated distribution of \bar{C} for (a) American, (b) German, (c) Chinese, and (d) Russian markets

5.2 Multivariate Normal Distribution

In this section we test how the cross-correlation matrix will change (with respect to observed characteristics) if we let the distribution of the data be Gaussian. We generate new data of size $(N \times T)$ (the same as original) from multivariate normal distributions with zero means and the empirical correlation matrix C as covariance matrix. Next we apply our method, defined by formulas (4)–(7), to compute new correlation matrix and its characteristics. We repeat this routine 10,000 times for each considered market.

Figures 11, 12, 13, and 14 present histograms of analyzed characteristics. The results of the analysis show that this approach keeps the main characteristics the same except one. All characteristics saved their observed values in average and estimated shape of distribution is similar with the one provided by bootstrapping for most of the characteristics in each market. The number of eigenvalues above the edge hasn't saved its observed value in American and Russian markets but it appeared to be less than for bootstrapping in average (11 and 8, respectively) with very small probability for other values (Fig. 14a, d).

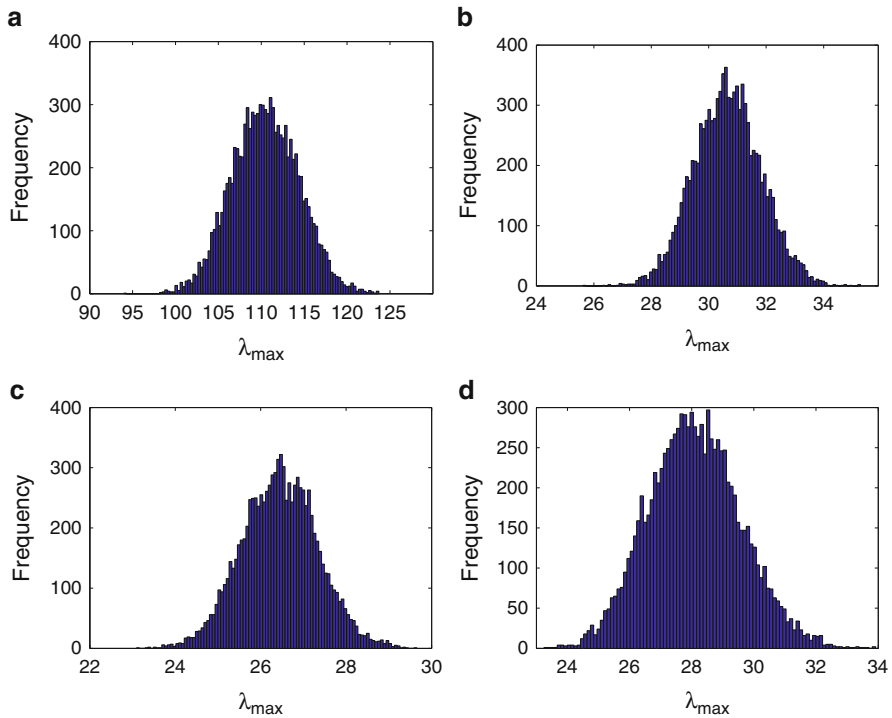


Fig. 8 Estimated distribution of λ_{max} for (a) American, (b) German, (c) Chinese, and (d) Russian markets

5.3 Multivariate Student Distribution

As in the previous section, we simulate our data ($N \times T$ time series), but this time using multivariate Student distribution with the empirical correlation matrix C as covariance matrix and 3 degrees of freedom. Then again we apply our method, defined by formulas (4)–(7), to compute new correlation matrix and its characteristics. We repeat this routine 10,000 times for each considered market.

The histograms on Figs. 15, 16, and 17 show that again the characteristics saved their observed values in average in each market. But this time variance is much less and estimated shape of distribution is not reminiscent of the one provided by bootstrapping. For the number of eigenvalues above λ_+ the picture is completely different from previous two tests. The average value is significantly higher for all markets and estimated shape of distribution also differs (Fig. 18). It means that this characteristic is sensitive to distribution of returns.

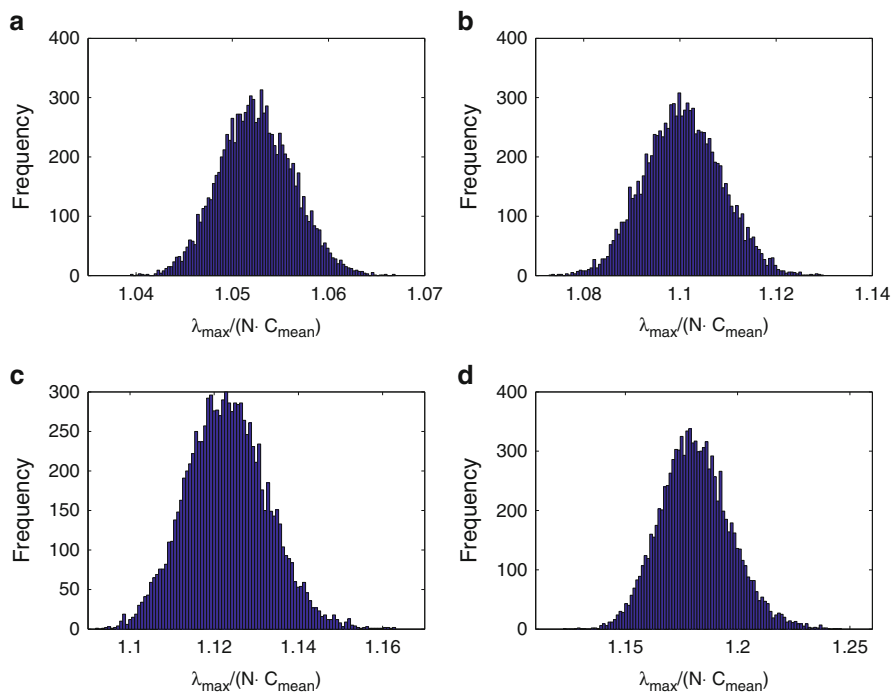


Fig. 9 Estimated distribution of $\frac{\lambda_{\max}}{N \cdot \bar{C}}$ for (a) American, (b) German, (c) Chinese, and (d) Russian markets

6 Concluding Remarks

Four different stock markets (Russian, American, German, and Chinese) are compared with respect to deviation of spectral properties of correlation matrix to predictions provided by RMT. It is observed that (like in the previous studies), there is one largest eigenvalue significantly higher than upper bound λ_+ of RMT range, and it is very close to the product $N \cdot \bar{C}$, where N denotes the number of stocks and \bar{C} —the average value of correlation. Average value of correlation is about 0.3 for all markets, except Russian, which is surprisingly high. In contrast, the number of eigenvalues above λ_+ , and the value of these numbers, differs from one market to another one. It can be related with sectors interconnections in different markets. Stability of observed phenomena was tested using bootstrapping method to see whether they are specific for considered markets or not. The analysis showed that the most of observed deviations from RMT are stable, the exception is the number of eigenvalues above the upper bound in American and Russian markets. This characteristic is not stable with respect to distribution of returns too.

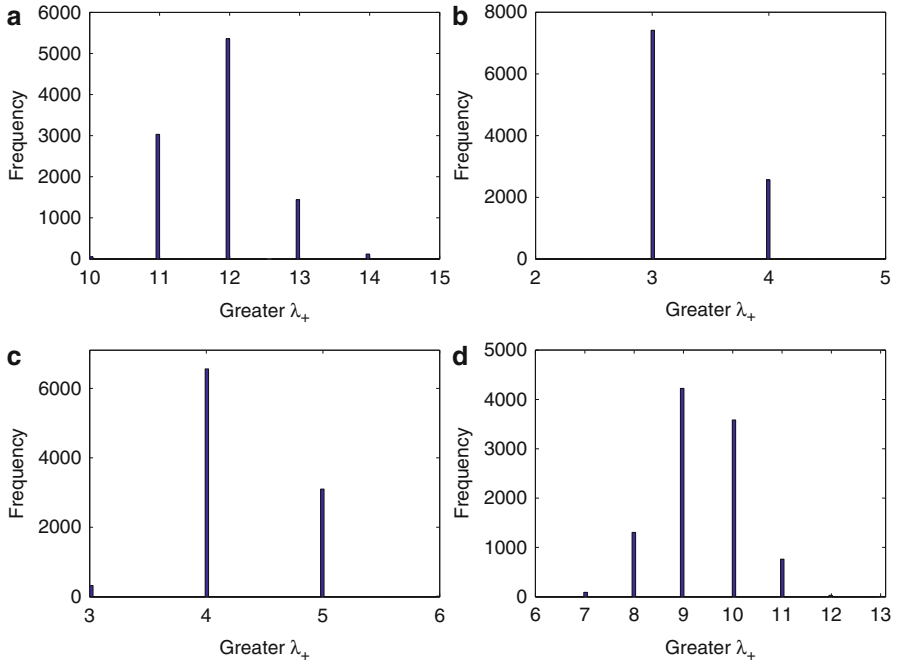


Fig. 10 Estimated distribution of the number of λ above λ_+ for (a) American, (b) German, (c) Chinese, and (d) Russian markets

Acknowledgements This work is partly supported by LATNA Laboratory, NRU HSE, RF government grant 11.G34.31.0057 and RFHR grant 15-32-01052.

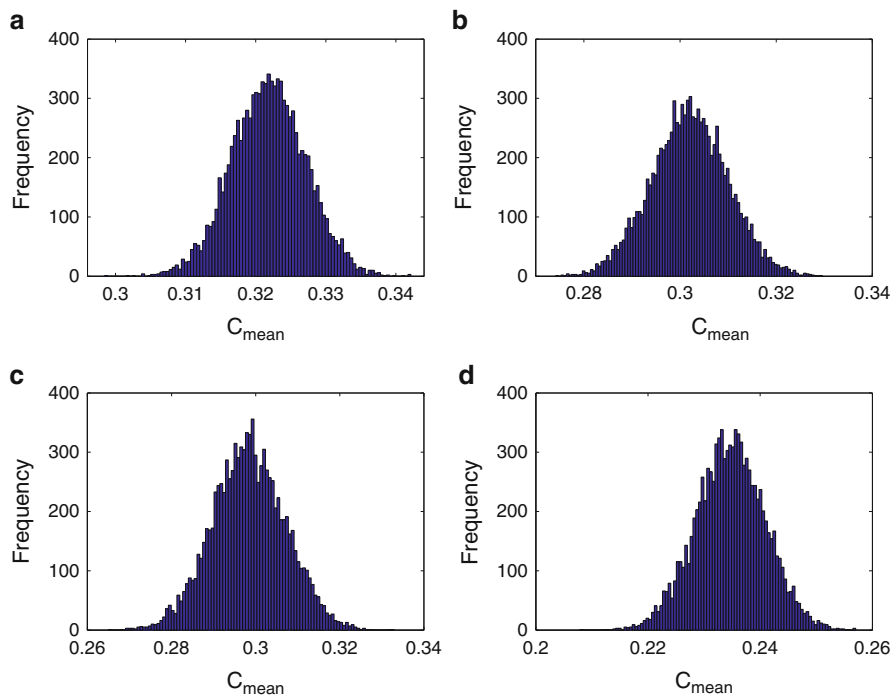


Fig. 11 Distribution of \bar{C} for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVN

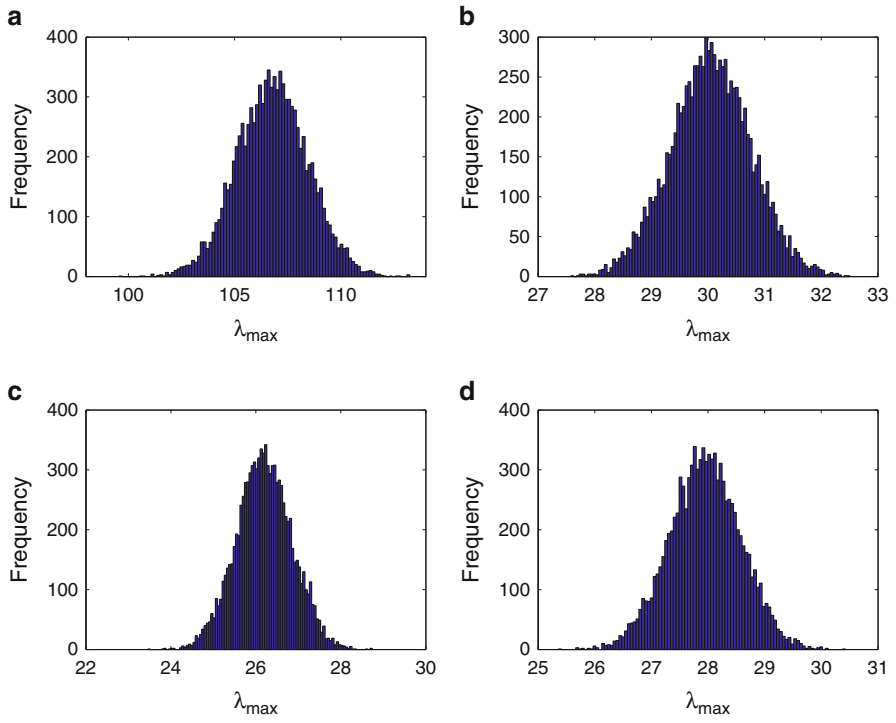


Fig. 12 Distribution of λ_{max} for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVN

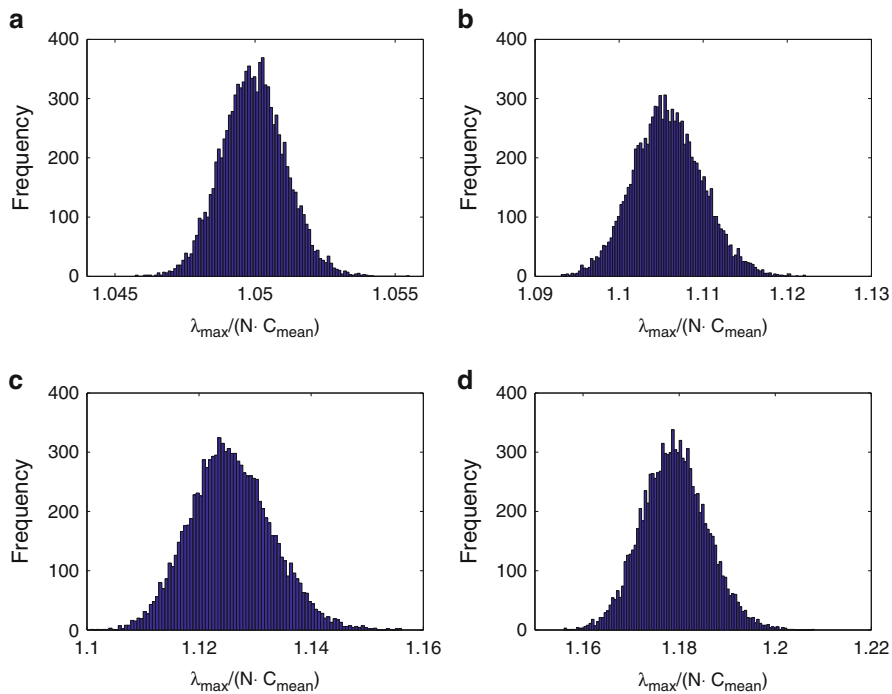


Fig. 13 Distribution of $\frac{\lambda_{\max}}{N \cdot C}$ for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVN

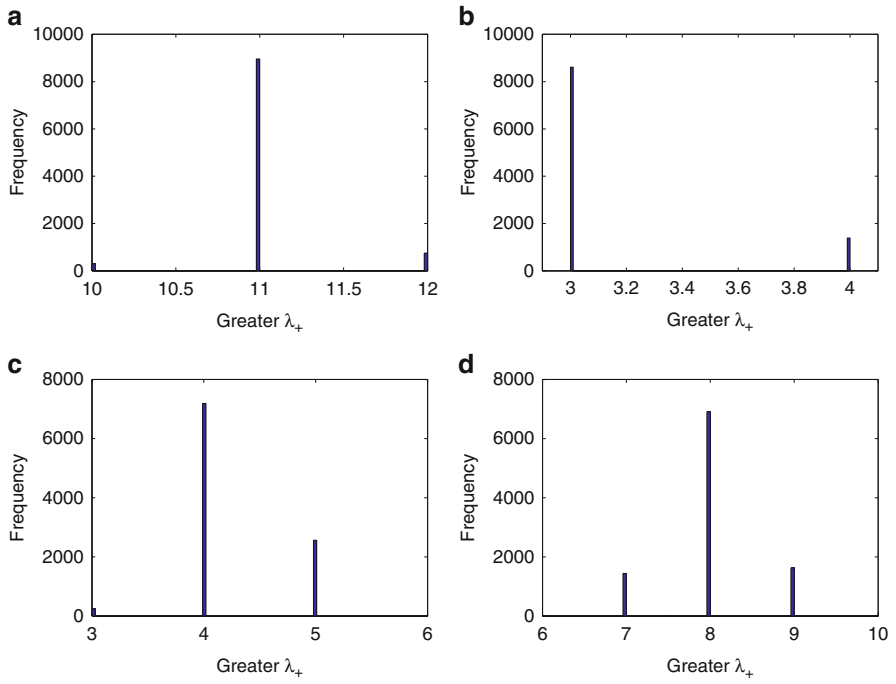


Fig. 14 Distribution of the number of λ above λ_+ for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVN

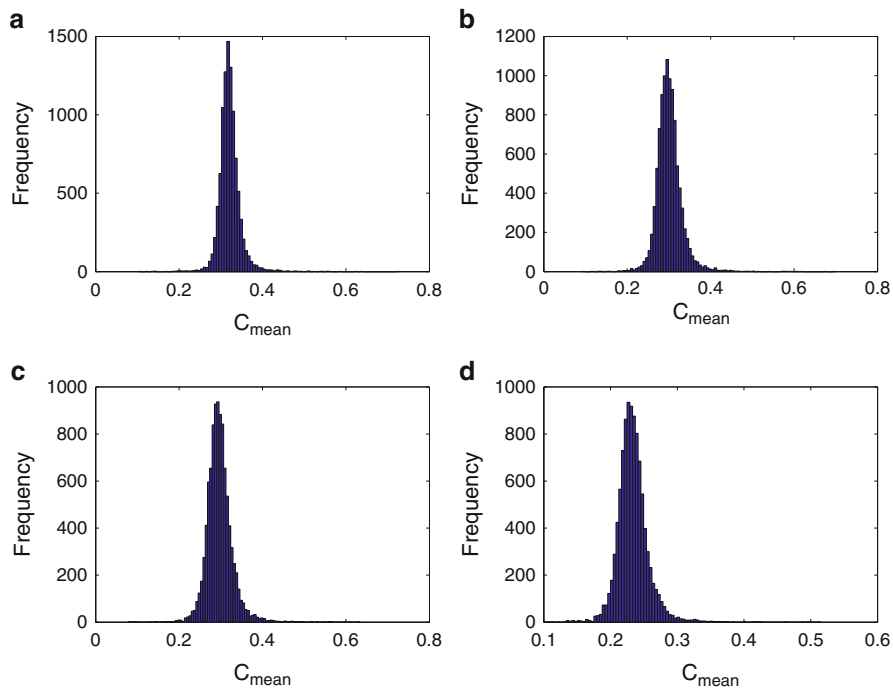


Fig. 15 Distribution of \bar{C} for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVStudent

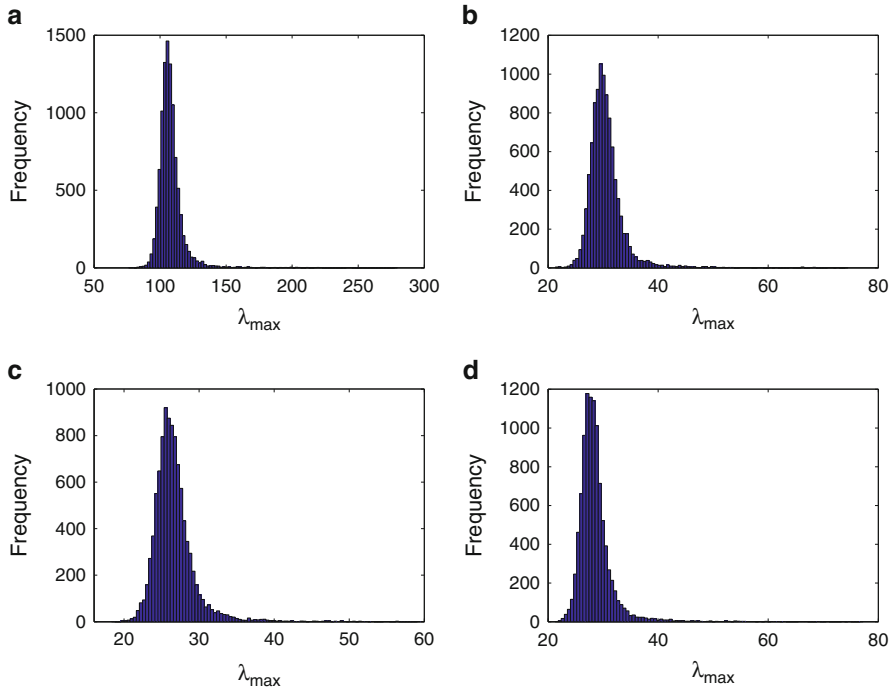


Fig. 16 Distribution of λ_{max} for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVStudent

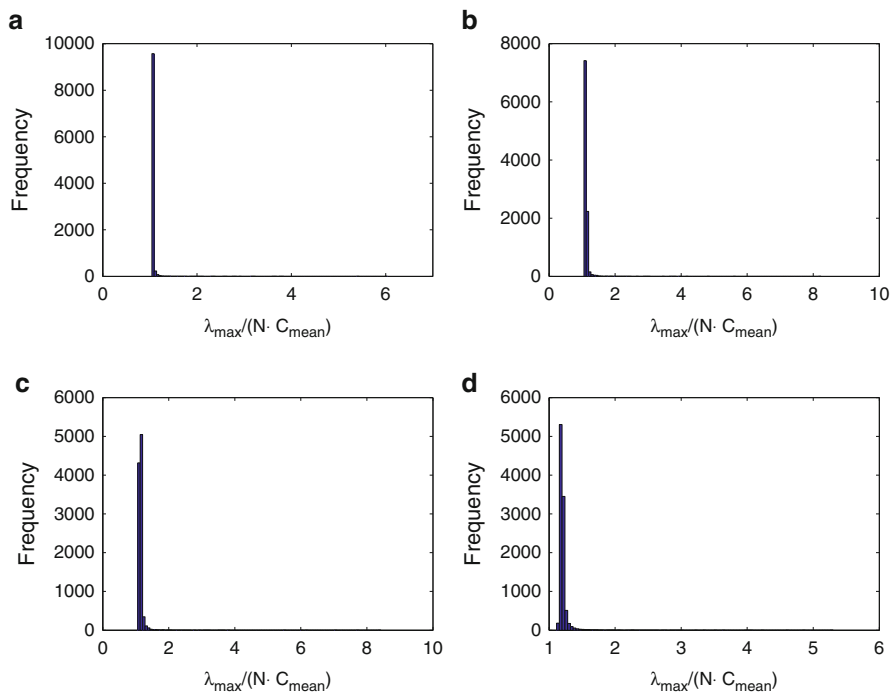


Fig. 17 Distribution of $\frac{\lambda_{\max}}{N \cdot \bar{C}}$ for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVStudent

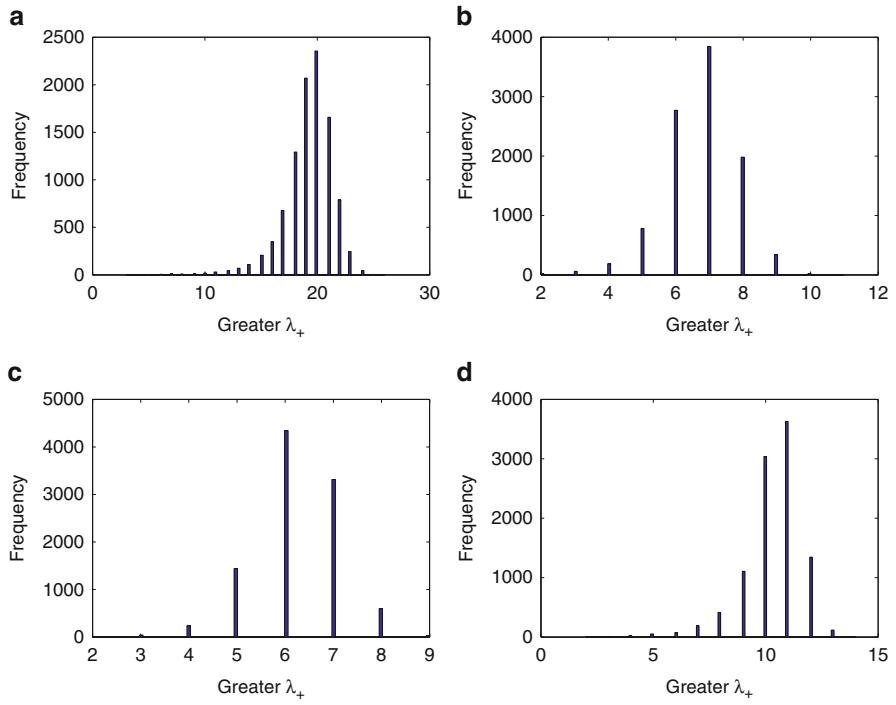


Fig. 18 Distribution of the number of λ above λ_+ for (a) American, (b) German, (c) Chinese, and (d) Russian markets from data generated by MVStudent

References

1. Boginski, V., Butenko, S., Pardalos, P.M.: On structural properties of the market graph. In: *Innovations in Financial and Economic Networks* (Anna Nagurney, Editor), pp. 29–45, Edward Elgar Publishing, (2003)
2. Boginski, V., Butenko, S., Pardalos, P.M.: Mining market data: a network approach. *Comput. Oper. Res.* **33**(11), 3171–3184 (2006)
3. Conlon, T., Ruskin, H.J., Crane, M.: Cross-correlation dynamics in financial time series. *Physica A* **388**(5), 705–714 (2009)
4. Cukur, S., Eryigit, M., Eryigit, R.: Cross correlations in an emerging market financial data. *Physica A* **376**, 555–564 (2007)
5. Edelman, A.: Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9**(4), 543–560 (1988)
6. Elton, E.J., Gruber, M.J.: *Modern Portfolio Theory and Investment Analysis*. Wiley, New York (1995)
7. Kwapien, J., Oswiecimka, P., Drozd, S.: The bulk of the stock market correlation matrix is not pure noise. *Physica A* **359**, 589–606 (2006)
8. Laloux, L., Cizeau, P., Bouchaud, J.P., Potter, M.: Noise dressing of financial correlation matrices. *Phys. Rev. Lett.* **83**(7), 1467–1470 (1999)
9. Marchenko, V., Pastur, L.: Distribution of some sets of random matrices. *Math. USSR-Sb* **1**, 457–483 (1967)
10. Markowitz, H.M.: *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York (1959)
11. Nguyen, Q.: One-factor model for the cross-correlation matrix in the Vietnamese stock market. *Physica A* **392**(13), 2915–2923 (2013)
12. Oh, G., Eom, C., Wang, F., Jung, W.S., Stanley, H.E., Kim, S.: Statistical properties of cross-correlation in the Korean stock market. *Eur. Phys. J. B* **79**, 55–60 (2011)
13. Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Stanley, H.E.: Universal and nonuniversal properties of cross correlations in financial time series. *Phys. Rev. Lett.* **83**(7), 1471–1474 (1999)
14. Sengupta, A.M., Mitra, P.P.: Distributions of singular values for some random matrices. *Phys. Rev. E* **90**(3), 3389–3392 (1999)