

# Chapter 2

## Complex Market Dynamics in the Light of Random Matrix Theory



Hirdesh K. Pharasi, Kiran Sharma, Anirban Chakraborti  
and Thomas H. Seligman

**Abstract** We present a brief overview of random matrix theory (RMT) with the objectives of highlighting the computational results and applications in financial markets as complex systems. An oft-encountered problem in computational finance is the choice of an appropriate epoch over which the empirical cross-correlation return matrix is computed. A long epoch would smoothen the fluctuations in the return time series and suffers from non-stationarity, whereas a short epoch results in noisy fluctuations in the return time series and the correlation matrices turn out to be highly singular. An effective method to tackle this issue is the use of the power mapping, where a non-linear distortion is applied to a short epoch correlation matrix. The value of distortion parameter controls the noise-suppression. The distortion also removes the degeneracy of zero eigenvalues. Depending on the correlation structures, interesting properties of the eigenvalue spectra are found. We simulate different correlated Wishart matrices to compare the results with empirical return matrices computed using the S&P 500 (USA) market data for the period 1985–2016. We also briefly review two recent applications of RMT in financial stock markets: (i) Identification of “market states” and long-term precursor to a critical state; (ii) Characterization of catastrophic instabilities (market crashes).

---

H. K. Pharasi

Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México,  
62210 Cuernavaca, México

e-mail: [hirdeshpharasi@gmail.com](mailto:hirdeshpharasi@gmail.com)

K. Sharma · A. Chakraborti (✉)

School of Computational and Integrative Sciences, Jawaharlal Nehru University,  
New Delhi 110067, India

e-mail: [anirban@jnu.ac.in](mailto:anirban@jnu.ac.in)

K. Sharma

e-mail: [kiransharma1187@gmail.com](mailto:kiransharma1187@gmail.com)

T. H. Seligman

Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México,  
62210 Cuernavaca, México

e-mail: [seligman@icf.unam.mx](mailto:seligman@icf.unam.mx)

Centro Internacional de Ciencias, 62210 Cuernavaca, México

© Springer Nature Switzerland AG 2019

F. Abergel et al. (eds.), *New Perspectives and Challenges in Econophysics and Sociophysics*, New Economic Windows,

[https://doi.org/10.1007/978-3-030-11364-3\\_2](https://doi.org/10.1007/978-3-030-11364-3_2)

## Introduction

With the advent of the “Big Data” era [9, 13], large data sets have become ubiquitous in numerous fields—image analysis, genomics, epidemiology, engineering, social media, finance, etc., for which we need new statistical and analytical methods [3, 5, 6, 15, 29]. Empirical correlation matrices are of primal importance in big data analyses, since various statistical methods strongly rely on the validity of such matrices in order to isolate meaningful information contained in the “observational” signals or time series [2]. Often the time series are of finite lengths, which can lead to spurious correlations and make it difficult to extract the signal from noise [11, 26]. Hence, it is very important to understand quantitative effects of finite-size time series in determination of empirical correlations [8, 11, 26, 33].

Random matrix theory (RMT) tries to describe statistics of eigenvalues of random matrices, often in the limit of large dimensions. The subject came up first in a celebrated paper of Wishart [39] in 1929 where he proposed that the correlation matrix of white noise time series was an adequate prior for correlation matrices. E. Cartan proposed the classical random matrix ensembles in an important but little known paper [4]. After that there was increasing interest in the subject among which it is important to mention work by L.G. Hua, who published the first monographs on the subject in 1952; an English translation is available [12].

Wigner introduced RMT to physics, based on the assumption that the interactions between the nuclear constituents were so complex that they could be modeled as random fluctuations in the framework of his R-matrix scattering theory [36]. This culminated in the presentation of the Hamiltonian  $\hat{H}$  as a large random matrix, such that the energy levels of the nuclear system could be approximated by the eigenvalues of this matrix, and indeed the spacings between the energy levels of nuclei could be modeled by the spacing of eigenvalues of the matrix [37, 38]. The use of RMT has spread over many fields from molecular physics [14] to quantum chromodynamics [28]. Lately, RMT has become a popular tool for investigating the dynamics of financial markets using cross-correlations of empirical return time series [25, 30].

In this chapter, we present recent techniques of random matrix theory (RMT) mainly focused on computational results and applications of correlations in financial markets viewed as complex systems [1, 10, 30, 31]. A central problem that often arises in computational finance is the choice of the epoch size over which the empirical cross-correlation return matrix needs to be computed. A very long epoch would smoothen the fluctuations in return time series and also the time series suffers from the problem of non-stationarity [19], whereas a short-time epoch would result in noisy fluctuations in return time series and the correlation matrix turns out to be highly singular (with many zero eigenvalues) [8]. Among others, an effective method to tackle this issue has been the use of the power mapping [8, 11, 26, 33], where a non-linear distortion is applied to a short epoch correlation matrix. Here, we demonstrate how the value of distortion parameter controls the noise-suppression. It also removes the degeneracy of the zero eigenvalues (which for very small values of the distortion

parameter leads to a well separated “emerging spectrum” near zero). Depending on the correlation structures, interesting properties of the eigenvalue spectra are found. Correlation matrices constructed from white noise were introduced by Wishart and their eigenvalue spectrum gets a shape of Marčenko-Pastur distribution [16]; there are significant deviations when a correlation structure is introduced [7]. We simulate different correlated Wishart matrices [18, 39] to compare the results with empirical return matrices computed using S&P 500 (USA) market data for the period 1985–2016 [8]. We also briefly review two recent applications of RMT in financial stock markets: (i) Identification of “market states” and long-term precursor to a critical state [23]; (ii) Characterization of catastrophic instabilities (market crashes) [8].

This chapter is described as follows. Section “Data Description, Methodology and Results” discusses the data description, methodology and results in details. Section “Recent Applications of RMT in Financial Markets” contains applications of RMT in financial markets. Finally, section “Concluding Remarks” contains concluding remarks.

## Data Description, Methodology and Results

### *Data Description*

We have used the database of Yahoo finance [40], for the time series of adjusted closure prices for S&P 500 (USA) market, for the period 02/01/1985–30/12/2016 ( $T = 8068$  days); number of stocks  $N = 194$ , where we have included the stocks that are present in the index for the entire duration. The sectoral abbreviations are given in Table 2.1.

### *Methodology and Results*

Correlations between different financial assets play fundamental roles in the analyses of portfolio management, risk management, investment strategies, etc. However, one only has finite time series of the assets prices; hence, one cannot calculate the exact

**Table 2.1** Abbreviations of ten different sectors for S&P 500 index

Labels	Sectors	Labels	Sectors
CD	Consumer discretionary	ID	Industrials
CS	Consumer staples	IT	Information technology
HC	Health care	MT	Materials
EG	Energy	TC	Technology
FN	Financials	UT	Utilities

correlation among assets, but only an approximation. The quality of the estimation of the true cross-correlation matrix strongly depends on the ratio between the length of the financial price time series  $T$  and the number of assets  $N$ . The larger the ratio  $Q = T/N$ , the better the estimation is; though for practical limitations, the ratio can be even smaller than unity. However, such correlation matrices are often too noisy, and thus need to be filtered from noise. To build the correlation matrices, we first calculate the return  $r_i$  from the daily price  $P_i$  of stocks  $i = 1, \dots, N$ , at time  $t$  (trading day):

$$r_i(t) = \ln P_i(t) - \ln P_i(t-1), \quad (2.1)$$

where  $P_i(t)$  denotes the price of stock  $i$  at time  $t$ . Since different stocks have varying levels of volatility, we define the equal-time Pearson cross-correlation coefficient as

$$C_{ij}(\tau) = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sigma_i \sigma_j}, \quad (2.2)$$

where  $\langle \dots \rangle$  denotes the time average and  $\sigma_k$  denotes the standard deviation of the return time series  $r_k$ ,  $k = 1, \dots, N$ , computed over an epoch of  $M$  trading days ending on day  $\tau$ . The elements  $C_{ij}$  are restricted to the domain  $-1 \leq C_{ij} \leq 1$ , where  $C_{ij} = 1$  corresponds to perfect correlations,  $C_{ij} = -1$  to perfect anti-correlations, and  $C_{ij} = 0$  to uncorrelated pairs of stocks. The difficulties in analyzing the significance and meaning of the empirical cross-correlation coefficients  $C_{ij}$  are due to several reasons, which include the following:

1. Market conditions change with time and the cross-correlations that exist between any pair of stocks may not be stationary if an epoch chosen is too long.
2. Too short epoch, for estimation of cross-correlations, introduces “noise”, i.e., fluctuations.

For these reasons, the empirical cross-correlation matrix  $\mathbf{C}(\tau)$  often contains “random” contributions plus a part that is not a result of randomness [22, 24]. Hence, the eigenvalue statistics of  $\mathbf{C}(\tau)$  are often compared against those of a large random correlation matrix—a correlation matrix constructed from mutually uncorrelated time series (white noise) known as Wishart matrix.

We first reproduce the basic results of RMT, e.g., the Marčenko-Pastur distribution, or Marčenko-Pastur law, which describes the asymptotic behavior of eigenvalues of square random matrices [16]. Then, we present a study of time evolution of the empirical cross-correlation structures of return matrices for  $N$  stocks and the eigenvalues spectra over different time epochs, and try to extract some new properties or information about the financial market [8, 23].

## Wishart and Correlated Wishart Ensembles

Let us construct a large random matrix  $\mathbf{B}$  arising from  $N$  random time series each of length  $T$ , where the entries of a time series are real independent random variables

drawn from a standard Gaussian distribution with zero mean and variance  $\sigma^2$ , such that the resulting matrix  $\mathbf{B}$  is  $N \times T$ . Then the Wishart matrix can be constructed as

$$\mathbf{W} = \frac{1}{T} \mathbf{B} \mathbf{B}'. \quad (2.3)$$

In RMT, the ensemble of Wishart matrices is known as the *Wishart orthogonal ensemble*. In the context of a time series,  $\mathbf{W}$  may be interpreted as the *covariance matrix*, calculated over  $N$  stochastic time series, each with  $T$  statistically independent variables. This implies that on average,  $\mathbf{W}$  does not have cross-correlations.

A correlated Wishart matrix can be constructed as

$$\mathbf{W} = \frac{1}{T} \mathbf{G} \mathbf{G}', \quad (2.4)$$

where  $\mathbf{G} = \boldsymbol{\zeta}^{1/2} \mathbf{B}$ , is a  $N \times T$  matrix;  $\mathbf{G}'$  is the  $T \times N$  transpose matrix of  $\mathbf{G}$ , and the  $N \times N$  positive definite symmetric matrix  $\boldsymbol{\zeta}$  controls the actual correlations. If  $\boldsymbol{\zeta}$  is a diagonal matrix with the diagonal entries as unity and off-diagonal entries as zero (i.e.,  $\boldsymbol{\zeta} = \mathbb{1}$ , the identity matrix), then the resulting matrix  $\mathbf{W}$  reduces to one of the former *Wishart orthogonal ensemble*. If the diagonal entries of  $\boldsymbol{\zeta}$  are unity and off-diagonal elements are non-zero and real, then the resulting matrices form the *correlated Wishart orthogonal ensemble*. For simplicity, in this chapter, we have generated and used  $\boldsymbol{\zeta}$  for which all the off-diagonal elements are same (equal to a constant  $U$ , which lies between zero and unity).

The spectrum of eigenvalues for the Wishart orthogonal ensemble can be calculated analytically. For the limit  $N \rightarrow \infty$  and  $T \rightarrow \infty$ , with  $Q = T/N$  fixed (and bigger than unity), the probability density function of the eigenvalues is given by the Marčenko-Pastur distribution:

$$\bar{\rho}(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \quad (2.5)$$

where  $\sigma^2$  is the variance of the elements of  $\mathbf{G}$ , while  $\lambda_{min}$  and  $\lambda_{max}$  satisfy the relation:

$$\lambda_{min}^{max} = \sigma^2 \left( 1 \pm \frac{1}{\sqrt{Q}} \right)^2. \quad (2.6)$$

For  $Q \leq 1$ , positive semi-definite matrices  $\mathbf{W}$ , the density  $\bar{\rho}(\lambda)$  in the above Eq. 2.5 is normalized to  $Q$  and not to unity. Therefore, taking into account the  $(N - T)$  zeros, we have

$$\bar{\rho}(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} + (1 - Q)\delta(\lambda). \quad (2.7)$$

First, we have generated a Wishart matrix  $\mathbf{W}$  (with  $\boldsymbol{\zeta} = \mathbb{1}$ ) of size  $N \times N$  constructed from  $N$  time series of real independent Gaussian variables, each of finite

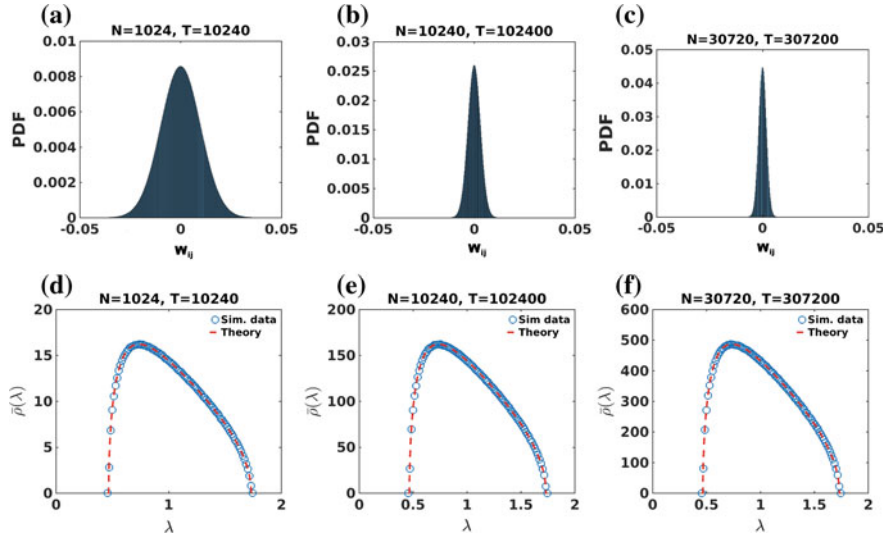
length  $T$ , zero mean and unit variance ( $\sigma^2 = 1$ ). Figure 2.1 shows the effect of finite sizes of the sets of parameters  $N$  and  $T$  on the probability distributions of the elements  $W_{ij}$  of the Wishart ensemble and the corresponding eigenvalue spectra. Figure 2.1a shows the probability distribution of the elements of the Wishart matrix of dimensions, where  $N = 1024$  and  $T = 10240$ . Figure 2.1d shows the corresponding density of eigenvalues  $\bar{\rho}(\lambda)$ , which takes the shape of the theoretical Marčenko-Pastur distribution (red dashed line) [16]. Similarly, Fig. 2.1b, c show the respective probability distributions of the elements of Wishart matrices generated using the sets of parameters  $N = 10240$  and  $T = 102400$ , and  $N = 30720$  and  $T = 307200$ . We can see that with increase in system size (both  $N$  and  $T$ ) the shape of the distribution becomes narrower, implying that the amount of spurious cross-correlations decreases. Ideally, the distribution should be a Dirac-delta at zero, since true cross-correlations do not exist. The eigenvalue spectra are less sensitive to the parameters  $N$  and  $T$ , as can be seen in Fig. 2.1e, f, which show the corresponding eigenvalue spectra. For all of the above simulations, we find the simulated data agree closely with the theoretical Marčenko-Pastur distributions (red dashed lines) with  $\lambda_{max} = 1.732$  and  $\lambda_{min} = 0.468$  (theoretically calculated using Eq. 2.6, and  $Q = 10$ ).

As we have mentioned earlier, the assumption of stationarity fails for a very long return time series, so it is often useful to break one long time series of length  $T$  into  $n$  shorter epochs, each of size  $M$  (such that  $T/M = n$ ). The assumption of stationarity then improves for each of the shorter epochs. However, if there are  $N$  return time series, such that  $N \gg M$ , then the corresponding cross-correlation matrices are highly singular with  $N - M + 1$  zero eigenvalues, which lead to poor eigenvalue statistics. We use the power map technique [11, 34] to break the degeneracy of eigenvalues at zero. In this method, a non-linear distortion is given to each element ( $W_{ij}$ ) of the Wishart matrix  $\mathbf{W}$  (or later in each correlation coefficient  $C_{ij}$  of the empirical cross-correlation matrix  $\mathbf{C}$ ) of short epoch by:

$$W_{ij} \rightarrow (\text{sign } W_{ij})|W_{ij}|^{1+\varepsilon}, \quad (2.8)$$

where  $\varepsilon$  is a noise-suppression parameter. For very small distortions, e.g.,  $\varepsilon = 0.001$  (as used here), we get an ‘‘emerging spectrum’’ of eigenvalues, arising from the degenerated eigenvalues at zero which is well separated from the original spectrum. The power mapping method suppresses noise present in the correlation structure of short-time series (see e.g., Refs. [8, 17, 21, 23, 32] for recent studies and applications). Later in this chapter, we study different aspects of the power mapping method by varying the value of distortion  $\varepsilon$  from 0 to 0.8.

In Fig. 2.2, we have studied the effect of non-linear distortion on the behavior of Wishart ensemble ( $U = 0$ ), where  $N \gg M$ . The top row of Fig. 2.2 shows semi-log plots of the ensembles with parameters: (a)  $N = 1024$  and  $M = 512$ , and (b)  $N = 1024$  and  $M = 64$ . Then small non-linear distortions with  $\varepsilon = 0.001$  are given to the ensembles to display the emerging spectra, shown in Fig. 2.2c, d. Interestingly, the shape of the emerging spectrum changes from a semi-circle to a strongly distorted one, as  $M$  becomes shorter. Also, note that emerging spectrum shifts towards the left

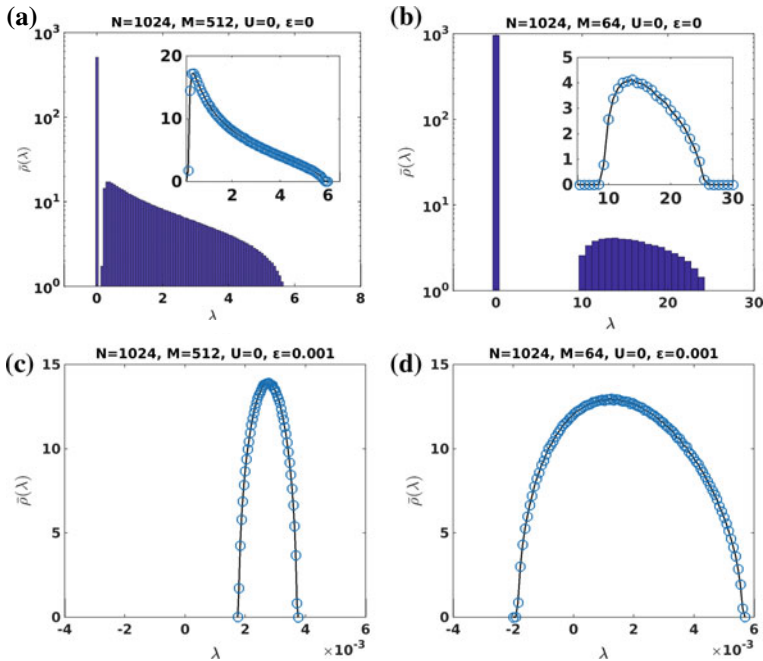


**Fig. 2.1** a–c show the effect of finite size on true correlations with the dimensions of  $\mathbf{B}$  ( $N = \text{finite}$ ,  $T = \text{finite}$  and  $Q = (T/N) = 10$ ). The probability distribution of elements ( $W_{ij}$ ) of the Wishart ensemble of size, constructed from  $N$  time series, each with real independent Gaussian random variables of length  $T$  with zero mean and variance  $\sigma^2$ . The variance of the distribution of  $W_{ij}$  decreases with the increase of  $N$  and  $T$  and reduces to zero for  $N \rightarrow \infty$  and  $T \rightarrow \infty$  with  $\frac{T}{N} = \text{finite}$ . d–f show the density of eigenvalues  $\bar{\rho}(\lambda)$  of Wishart ensemble, which are numerically fitted with the Marčenko-Pastur distributions [16] (red dash lines) for all  $N$  and  $T$ . The numerical values of  $\lambda_{\max} = 1.732$  and  $\lambda_{\min} = 0.468$  of the spectra also match exactly with the results theoretically calculated from Eq. 2.6. Numerical results for the probability distributions of the elements ( $W_{ij}$ ) and densities of the eigenvalues ( $\bar{\rho}(\lambda)$ ) have been generated using averages up to 200 ensembles

side as  $M$  becomes shorter. For smaller values of  $M$ , some of the eigenvalues of emerging spectrum become negative. The number of negative eigenvalues depend on the size of the epoch  $M$ , the distortion parameter  $\varepsilon$  and the mean correlation in the case of a correlated Wishart ensemble [21].

Figure 2.3 shows the effect of a constant correlation with strength  $U$  on the eigenvalue spectra and the emerging spectra of correlated Wishart ensembles with parameters  $N = 1024$  and  $M = 64$ . Figure 2.3a–c show the eigenvalue distributions, on the semi-log scales, for the correlated Wishart ensembles with correlations  $U = 0.1$ ,  $U = 0.3$ , and  $U = 0.8$ , respectively. Insets show the densities of non-zero eigenvalues, which are closely described by the Marčenko-Pastur distributions in all cases. In the bottom row, Fig. 2.3d–f show the densities of the corresponding emerging spectra arising from non-linear distortion of the degenerate eigenvalues at zero. The shapes of the emerging spectra change from distorted semi-circle to Lorentzian-like, as the constant correlation values increase for the correlated Wishart ensembles.

Next, we present the effect of the distortion (or noise-suppression) parameter  $\varepsilon$  on the eigenvalue spectra in Fig. 2.4. Figure 2.4a–f show the distributions of eigenvalues for the correlated Wishart ensembles with parameters  $N = 1024$  and  $M = 64$ , and



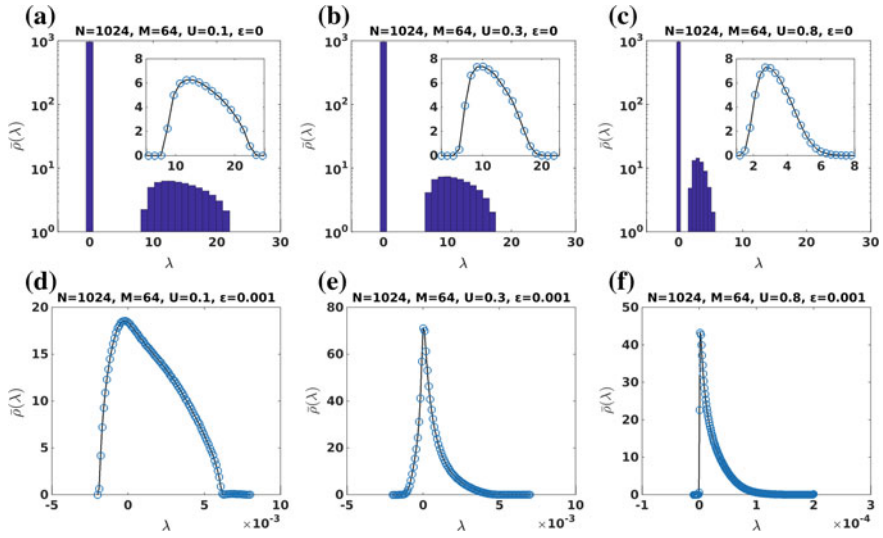
**Fig. 2.2** Semi-log plot of the eigenvalue distribution of Wishart matrix  $W$ , using the set of parameters **a**  $N = 1024$  and  $M = 512$ ; **b**  $N = 1024$  and  $M = 64$ . For short epochs ( $N > M$ ), the eigenvalue spectra have  $N - M + 1$  zero eigenvalues and the remaining eigenvalues of the spectra show a distributions similar to the Marčenko-Pastur distribution. Insets show the zoomed in views of remaining  $M - 1$  eigenvalues. **c** and **d** show the emerging spectra, generated using the power map technique with  $\varepsilon = 0.001$ , which are deformed semi-circular. Numerical results for densities of eigenvalues have been generated using the averages over 1000 ensemble members. Note that the emerging spectrum shifts towards left for smaller values of  $M$ , and also some of its eigenvalues become negative at smaller values of  $M$

varying distortion parameter values:  $\varepsilon = 0.0, 0.1, 0.2, 0.4, 0.6$  and  $0.8$ , keeping a constant correlation ( $U = 0.1$ ) among all off-diagonal elements in  $\zeta$ . The densities of non-zero eigenvalues are closely described by the Marčenko-Pastur distributions, but the emerging spectra move towards the main spectra as the value of  $\varepsilon$  increases. The emerging spectra is absent at  $\varepsilon = 0$ , while it merges with the main spectrum at high values of distortion parameter, e.g.,  $\varepsilon = 0.8$ .

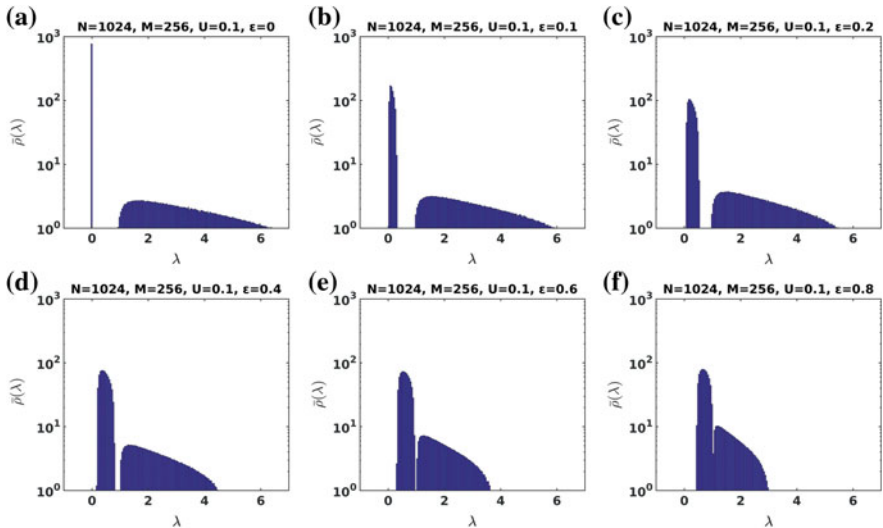
### Eigenvalue Decomposition of the Empirical Cross-Correlation Matrix

We also analyze  $N = 194$  adjusted daily closure price time series of the stocks of S&P 500 (USA) index from the Yahoo finance database [40]. As discussed in the methodology subsection, we construct the empirical cross-correlation matrix  $C(\tau)$  for an epoch of  $M = 200$  trading days, ending on trading day  $\tau$ . In Fig. 2.5a, e, we choose two correlation matrices for the time series from 07/03/2011 to 16/12/2011 (high

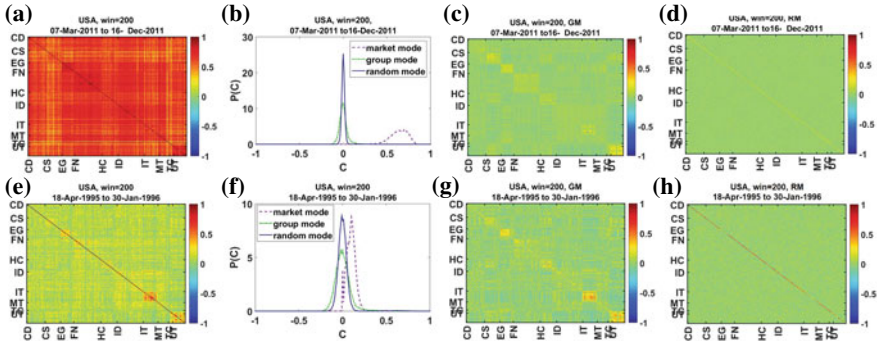




**Fig. 2.3** Eigenvalue spectra of correlated Wishart ensembles with parameters,  $N = 1024$  and  $M = 64$ , shown on semi-log scales with constant correlations: **a**  $U = 0.1$ , **b**  $U = 0.3$ , and **c**  $U = 0.8$ . Insets show the corresponding densities of non-zero eigenvalues, which are closely described by the Marčenko-Pastur distributions. **d-f** show the densities of the emerging spectra, when non-linear distortions (with  $\varepsilon = 0.001$ ) are applied to the same matrices. Note that the shape of the emerging spectrum changes from distorted semi-circle to a Lorentzian-like with the increase of constant correlation strength  $U$



**Fig. 2.4** Semi-log plots of the eigenvalue spectra for the correlated Wishart ensemble  $W$  with parameters  $N = 1024$  and  $M = 256$  at a constant correlation with  $U = 0.1$ , and distortion parameters of: **a**  $\varepsilon = 0$ , **b**  $\varepsilon = 0.1$ , **c**  $\varepsilon = 0.2$ , **d**  $\varepsilon = 0.4$ , **e**  $\varepsilon = 0.6$ , and **f**  $\varepsilon = 0.8$ . For  $\varepsilon = 0.1$ , the emerging spectrum is well separated from non-zero eigenvalues but with the increase of the distortion parameter  $\varepsilon$  the emerging spectrum starts moving towards the remaining non-zero eigenvalues spectra, and eventually merges with it at higher values, e.g.,  $\varepsilon = 0.8$



**Fig. 2.5** **a** and **e** show the cross-correlation matrices of 194 stocks of S&P 500 for  $M = 200$  days during: **a** 07/03/2011 to 16/12/2011; **b** 18/04/1995 to 30/01/1996. The stocks are arranged according to their industrial groups (abbreviations are given in Table 2.1). The blocks along the diagonal show the correlations within the same industrial groups; the color-bar shows the amount of correlation among the stocks. **a** shows the correlation matrix with high mean correlation and **(e)** with low mean correlation. **b** and **f** show the eigenvalue decomposition of the correlation matrix into the market mode, group modes and random modes. The market mode captures the mean market correlation, which corresponds to the dominant eigenvalue of the matrix. The group modes give the sectoral behavior of the market characterized by the subsequent 15 eigenvalues for a correlation matrix **(a)** and the next 62 eigenvalues for a correlation matrix **(e)** of the market. The rest of the eigenvalues show random behavior. **c** and **g** are the correlation matrix after removing the market mode and random modes from the correlation matrix; thus the matrix is composed of group modes only. We can visualize the block structure which shows the correlation among sectors. **d** and **h** show the correlation matrix after removing the market mode and group modes from the correlation matrix; so the matrix is composed of random modes only

mean correlation) and 18/04/1995 to 30/01/1996 (low mean correlation), respectively. The color-bar shows the amount of correlation among the stocks. The stocks are arranged according to their industrial groups (abbreviations are given in Table 2.1). The blocks along the diagonal show the correlations within the same industrial group. Figure 2.5b, f show the eigenvalue decomposition of the correlation matrices into the respective market mode, the group modes and the random modes. From such a segregation/decomposition, it is also possible to reconstruct the contributions of different modes to the aggregate correlation matrix as we show below.

The largest eigenvalue of the correlation matrix, corresponds to a market mode reflects the aggregate dynamics of the market common across all stocks, and strongly correlated to the mean market correlation. The group modes capture the sectoral behavior of the market, which are 15 eigenvalues subsequent to the largest eigenvalue of the correlation matrix of Fig. 2.5c, and the 62 subsequent eigenvalues for correlation matrix of Fig. 2.5g. Remaining eigenvalues capture the random modes behavior of the market (see Fig. 2.5d, h). By using the eigenvalue decomposition, we can thus filter the true correlations (coming from the signal) and the spurious correlations (coming from the random noise). For this, we first decompose the aggregate correlation matrix as

$$C = \sum_{i=1}^N \lambda_i a_i a_i', \quad (2.9)$$

where  $\lambda_i$  and  $a_i$  are the eigenvalues and eigenvectors, respectively, of the correlation matrix  $C$ . An easy way of handling the reconstruction of the correlation matrix is to sort the eigenvalues in descending order, and then rearranging the eigenvectors in corresponding ranks. This allows one to decompose the matrix into three separate components, viz., market, group and random

$$C = C^M + C^G + C^R, \quad (2.10)$$

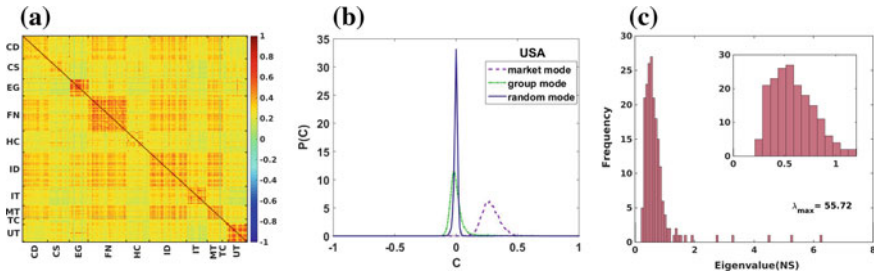
$$= \lambda_1 a_1 a_1' + \sum_{i=2}^{N_G} \lambda_i a_i a_i' + \sum_{i=N_G+1}^N \lambda_i a_i a_i', \quad (2.11)$$

where  $N_G$  is taken to be 15 for the high mean correlated matrix (Fig. 2.5a) and 62 for the low mean correlation (Fig. 2.5e), i.e., corresponding to the 15 (or 62) eigenvalues after the largest one, for two chosen correlation matrices. It is worth noting that the result is not extremely sensitive to the exact value of  $N_G$ . As mentioned above, the eigenvectors from 2 to  $N_G$  describe the sectoral dynamics.

Figure 2.5c, g show the correlation matrices after removing the market mode and random modes from the respective correlation matrices; so the matrices show group modes only. We can see the block structures, which exhibit the correlations among the sectors. Figure 2.5d, h show the correlation matrices after removing the market mode and group modes; so the matrices display the random modes only.

An important observation is that the market mode shifts towards the right with the increment of the mean correlation. The group modes almost coincide with the random modes but with higher variance. Thus, the sectoral dynamics are almost absent whereas the market mode is very strong (similar to what was observed in Ref. [27]).

Figure 2.6a shows the average cross-correlation matrix of  $N = 194$  stocks of S&P 500 for the entire duration 1985–2016 ( $T = 8068$  trading days). We decomposed the average cross-correlation matrix into the market mode, group modes and random modes. As usual, the market mode captures the mean market correlation corresponding to the maximum eigenvalue, which is separate from rest of the eigenvalues (see Ref. [35] for the comparison of the behavior of maximum eigenvalues in correlated Wishart ensembles). The group modes, which tell about the sectoral behavior of the market, largely coincide with the random modes and correspond to the random behavior of the stocks. The resulting eigenvalue distribution (shown in Fig. 2.6c) thus has part that is a Marčenko-Pastur distribution [16] (see Fig. 2.6c and its inset) and some deviations. As  $N \ll T$  so we do not get any zero eigenvalues. The maximum eigenvalue ( $\lambda_{max} = 55.72$ ) of the spectra dominates the whole market. The next 19 eigenvalues correspond to the group modes, and the rest behave as random modes. The smallest eigenvalue of the spectrum  $\lambda_{min} = 0.22$ .



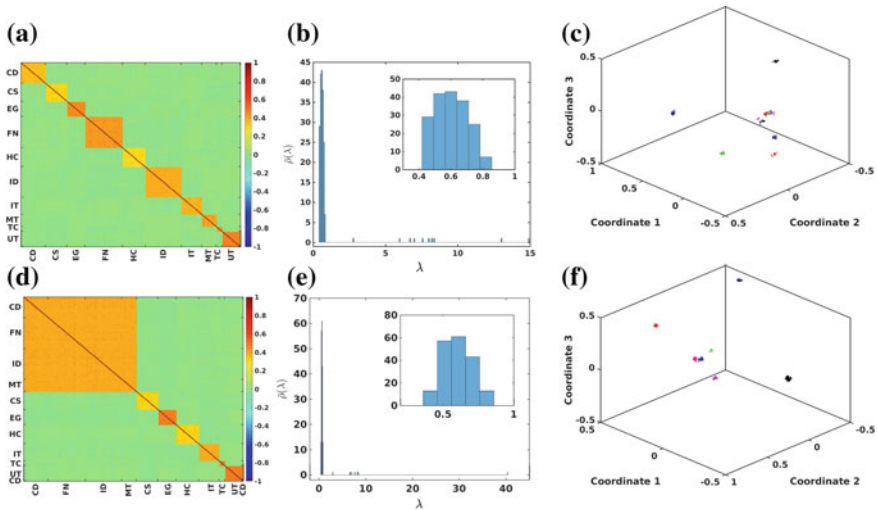
**Fig. 2.6** **a** Average cross-correlation matrix of 194 stocks of S&P 500 in 32-years period from 1985 to 2016. The stocks are arranged according to their industrial groups (abbreviations are given in Table 2.1). The diagonal blocks show the correlations within the same industrial groups and off diagonal elements show correlations with other industrial groups. **b** Eigenvalue decomposition of the average correlation matrix into market mode, group modes and random modes. The market mode captures the mean market correlation. The group modes give the sectoral behavior of the market. The random modes of the correlation matrix yield the Marčenko-Pastur distribution. **c** Eigenvalue spectrum of the correlation matrix, evaluated for the *long* return time series for the entire period of 32-years, with the maximum eigenvalue of the normal spectrum  $\lambda_{max} = 55.72$ . The largest eigenvalue is well separated from the ‘bulk’. Inset shows the random part of the spectrum, with the smallest eigenvalue of the normal spectrum  $\lambda_{min} = 0.22$

Figure 2.7a shows the cross-correlation matrices constructed from *surrogate* data ( $N = 194$  correlated Gaussian noises, each of length  $T = 10000$ ) such that the matrix has 10 diagonal blocks of different correlations (equal to the average correlations of different sectors of the S&P 500 market). Figure 2.7d shows the *surrogate* cross-correlation matrix ( $N = 194$ ;  $T = 10000$ ) but now with one big block and 6 smaller blocks. The mean correlation of the big block is equal to the mean correlation of four sectors (CD, FN, ID and MT of Fig. 2.6a) and they show high inter-sectorial correlation in S&P 500 market in 32 years. Eigenvalue spectra of the correlation matrices are shown in Fig. 2.7b, e, each of which consists of the Marčenko-Pastur distributions (see insets), followed by 10 (and 7) eigenvalues corresponding to 10 (and 7) blocks (similar to sectors), respectively. Figure 2.7c, f show the 3D MDS plots, where the points (representing stocks) are scattered based on the correlations among the 10 and 7 blocks, respectively. In the MDS maps, more correlated stocks are placed nearby and anti-correlated are placed far apart (see also Ref. [23]). The  $k$ -means clustering performed on the surrogate data matrices, with  $k = 10$  and  $k = 7$ , yield 10 and 7 different clusters (represented in different colors), respectively.

## Dynamics of the Correlation Structure of US Market

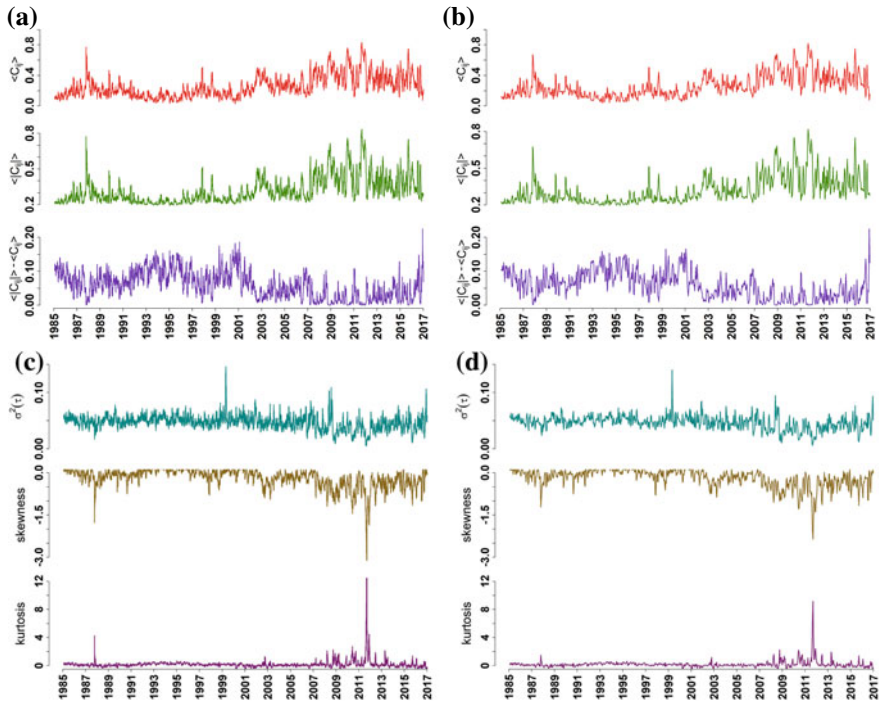
Next, we study the time evolution of the market correlations computed with the daily returns of  $N = 194$  stocks of S&P 500 over the period of 32-year (1985–2016, with  $T = 8068$  trading days).

Figure 2.8a, b show plots of mean of correlation coefficients ( $\langle C_{ij} \rangle$ ), mean of absolute values of correlation coefficients ( $\langle |C_{ij}| \rangle$ ) and the difference of the



**Fig. 2.7** **a** Cross-correlation matrices constructed from the correlated Gaussian time series with 10 diagonal blocks of different correlations (equal to the average correlation of each sector in Fig. 2.6a). **d** shows the same cross-correlation matrix but with one big block and 6 smaller blocks. The mean correlation of the big block is equal to the mean correlation of four sectors (CD, FN, ID and MT of Fig. 2.6a). They have high inter-sectorial correlation over the last 32 years in S&P 500 market. **b** and **e** show the eigenvalue spectra of the correlation matrices, which consist of the Marčenko-Pastur distributions followed by 10 group modes corresponding to 10 sectors and 7 group modes corresponding to 7 sectors, respectively. Insets show the enlarged pictures of the random part of the spectrum. **c** and **f** show plots of 10 and 7 different clusters, respectively, drawn in different colors using 3-dimensional  $k$ -means clustering technique. The clustering was performed on 3- $D$  multidimensional scaling (MDS) map of 194 stocks. Each point on the MDS map represents a stock of the market. The points are scattered in the map, based on the cross-correlations among the stocks—more correlated stocks are placed nearby and less correlated are placed far apart (see also Ref. [23])

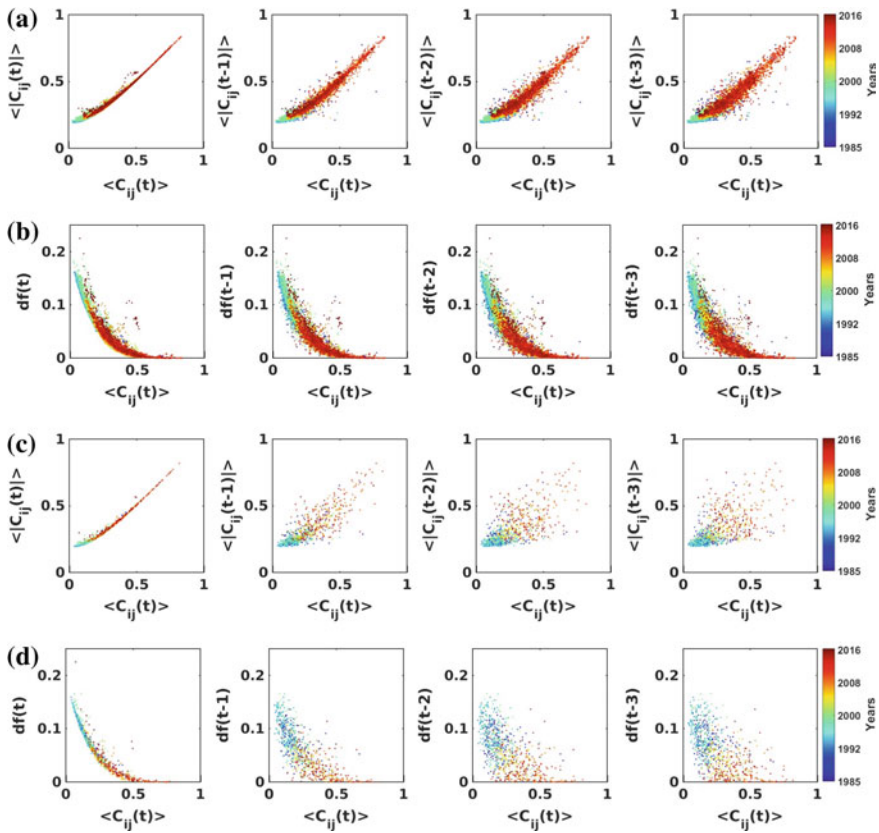
absolute mean and the mean of correlation coefficients  $df = \langle |C_{ij}| \rangle - \langle C_{ij} \rangle$  for short epochs of  $M = 20$  days, with shifts of:  $\Delta\tau = 1$  day (95% overlap) and  $\Delta\tau = 10$  days (50% overlap), respectively. Shifts toward the positive side of correlations are pointing toward periods of market crashes (with very high mean correlation values). The values of  $df$  are anti-correlated with the values of the mean of correlation coefficients. During a market crash when mean of correlation coefficient is high, there are very little anti-correlations among the stocks, then the value of  $df$  is extremely small, indeed near to zero (see Ref. [17]). It may act as an indicator of a market crash, as we observe that there is a high anti-correlation between the values of  $df$  and  $\langle C_{ij} \rangle$ , with leads of one or two days (ahead of the market crashes). Similarly, Fig. 2.8c, d show the plots of variance, skewness, and kurtosis of the correlation coefficients  $C_{ij}$  as functions of time with shifts of  $\Delta\tau = 1$  day and  $\Delta\tau = 10$  days, respectively. The mean correlation is anti-correlated to variance and skewness of  $C$ , i.e., when the mean correlation is high then both variance and skewness are low. Kurtosis is highly correlated to the mean correlation. These observations are seen in



**Fig. 2.8** Plots of mean of correlation coefficients ( $\langle C_{ij} \rangle$ ), mean of absolute values of correlation coefficients ( $\langle |C_{ij}| \rangle$ ) and the difference ( $df = \langle |C_{ij}| \rangle - \langle C_{ij} \rangle$ ) as functions of time, for short epochs of  $M = 20$  days, and shifts of: **a**  $\Delta\tau = 1$  day and **b**  $\Delta\tau = 10$  days. We find that during crashes (when mean correlation is very high), the difference  $df = \langle |C_{ij}| \rangle - \langle C_{ij} \rangle$  show minima (close to zero) (see Ref. [17]). Plots of variance ( $\sigma^2$ ), skewness, and kurtosis of the correlation coefficients as functions of time, for short epochs of  $M = 20$  days, and shifts of: **c**  $\Delta\tau = 1$  day and **d**  $\Delta\tau = 10$  days

the dynamical evolution of the market with epochs of  $M = 20$  days, and shifts of  $\Delta\tau = 1, 10$  day(s).

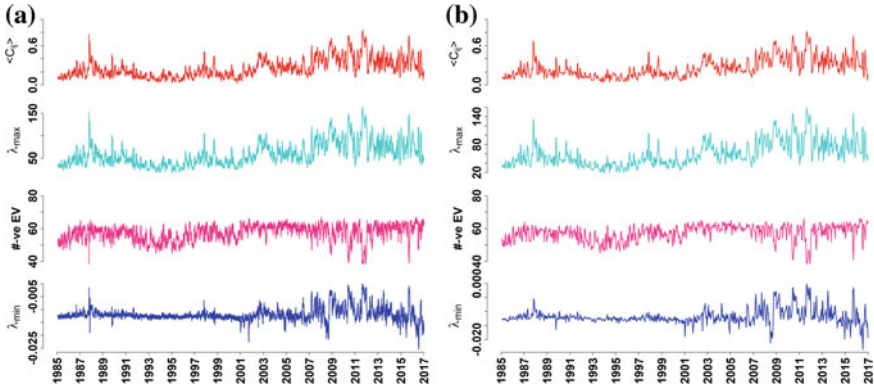
The scatter plots between  $\langle C_{ij} \rangle$  and  $\langle |C_{ij}| \rangle$ , and  $\langle C_{ij} \rangle$  and  $df (= \langle |C_{ij}| \rangle - \langle C_{ij} \rangle)$  for different time lags (no-lag, lag-1, lag-2, and lag-3) of empirical correlation matrices  $C(\tau)$ , with 194 stocks of S&P 500 and epochs of  $M = 20$  days, and shift of  $\Delta\tau = 1$  day, are shown in Fig. 2.9a, b, respectively. Here lag-1, lag-2, and lag-3 represent time lags of 1 day, 2 days, and 3 days, respectively. The color-bar shows the time period from 1985 to 2016 in years. The scatter plots show the correlations among  $\langle C_{ij} \rangle$  versus  $\langle |C_{ij}| \rangle$  and  $\langle C_{ij} \rangle$  versus  $df$ , at different time lags. The variances of the scatter plots increase with the increase of time lag, keeping the value of linear correlation coefficient almost similar. The strong linear correlation between  $\langle C_{ij} \rangle$  and  $\langle |C_{ij}| \rangle$  may give us an early information about a crash up to 3 days ahead (from the result of lag-3). Similar linear correlations are also visible in Fig. 2.9c, d, between  $\langle C_{ij} \rangle$  and  $\langle |C_{ij}| \rangle$ , and  $\langle C_{ij} \rangle$  and  $df$ , at different time lags (no-lag, lag-1, lag-2, and lag-3) for a shift of  $\Delta\tau = 10$  days. Here,



**Fig. 2.9** Scatter plots of  $\langle C_{ij} \rangle$  versus  $\langle |C_{ij}| \rangle$  and  $\langle C_{ij} \rangle$  versus  $df = \langle |C_{ij}| \rangle - \langle C_{ij} \rangle$ , for different time lags (No lag, 1-day, 2-days and 3-days) for the correlation matrix of epoch 20 days, with shifts of: **a–b**  $\Delta\tau = 1$  day; **c–d**  $\Delta\tau = 10$  days. The color-bar shows the time period in years

obviously lag-1, lag-2, and lag-3 represent time lags of 10 days, 20 days, and 30 days, respectively. The large variances in scatter plots indicate that it is hard to detect and extract information about a crash, e.g., 30 days in advance.

Figure 2.10a shows the temporal variation of mean correlation ( $\langle C_{ij} \rangle$ ), maximum eigenvalue ( $\lambda_{max}$ ), number of negative eigenvalues ( $\# -ve EV$ ) and smallest eigenvalue ( $\lambda_{min}$ ) of the emerging spectra with a shift of  $\Delta\tau = 1$  day. Using a small distortion ( $\varepsilon = 0.01$ ), we break the degeneracy of eigenvalues at zero and get the “emerging spectra” of eigenvalues which contain some interesting information about the market. The effect of the small distortion parameter  $\varepsilon = 0.01$  is negligible on non-zero eigenvalues of the spectrum including  $\lambda_{max}$ . We observed high correlation between  $\langle C_{ij} \rangle$  and  $\lambda_{max}$ . But the other properties of emerging spectrum ( $\# -ve EV$  and  $\lambda_{min}$ ) are less correlated with mean correlation  $\langle C_{ij} \rangle$  [21]. Figure 2.10b shows the same for the shift of  $\Delta\tau = 10$  days.



**Fig. 2.10** Plots for mean of correlation coefficients ( $\langle C_{ij} \rangle$ ), maximum eigenvalue ( $\lambda_{max}$ ), number of negative eigenvalues ( $\# - ve EV$ ) and smallest eigenvalue ( $\lambda_{min}$ ) of the spectrum as a function of time for an epoch of 20 days at  $\varepsilon = 0.01$  with shifts of: **a**  $\Delta\tau = 1$  day and **b**  $\Delta\tau = 10$  days. The correlation between  $\langle C_{ij} \rangle$  and  $\lambda_{max}$  is high, but two other properties of the “emerging spectrum” ( $\# - ve EV$  and  $\lambda_{min}$ ) are less correlated to mean correlation  $\langle C_{ij} \rangle$

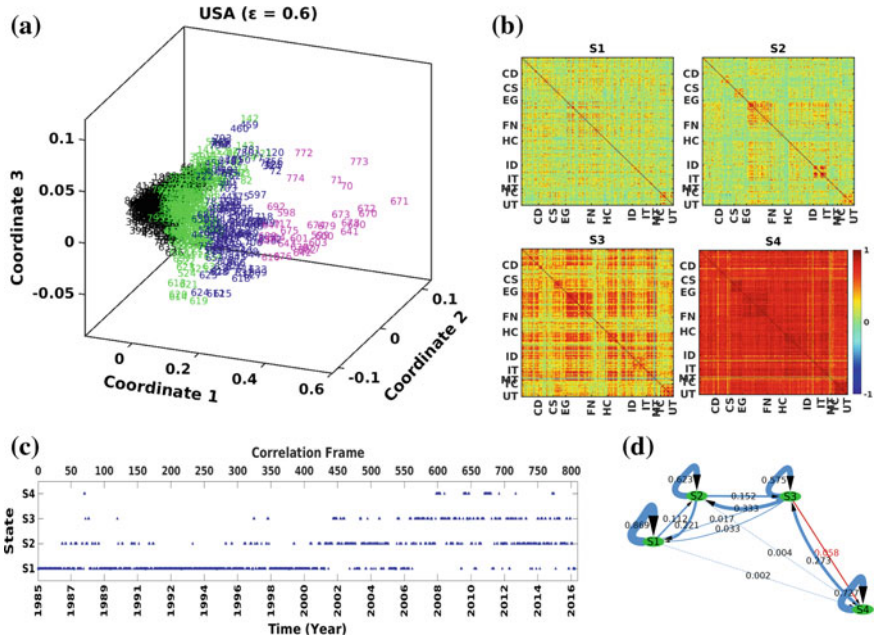
## Recent Applications of RMT in Financial Markets

### *Identification of Market States and Long-Term Precursors to a Crash State*

The study of the critical dynamics in any complex system is interesting, yet it can be very challenging. Recently, Pharasi et al. [23] presented an analysis based on the correlation structure patterns of S&P 500 (USA) data and Nikkei 225 (JPN) data, with short time epochs during the 32-year period of 1985–2016. They identified “market states” as clusters of similar correlation structures which occurred more frequently than by pure chance (randomness).

They first used the power mapping to reduce noise of the singular correlation matrices and obtained distinct and denser clusters in three dimensional MDS map (as shown in Fig. 2.11 a). The effects of noise-suppression were found to be prominent not only on a single correlation matrix at one epoch, but also on the similarity matrices computed for different correlation matrices at different short-time epochs, and their corresponding MDS maps. Using 3D-multidimensional scaling maps, they applied  $k$ -means clustering to divide the clusters of similar correlation patterns into  $k$  groups or market states. One major difficulty of this clustering method is that one has to pass the value of  $k$  as an input to the algorithm. Normally, there are several proposed methods of determining the value of  $k$  (often arbitrary). Pharasi et al. [23] showed that using a new prescription based on the cluster radii and an optional choice of the noise suppression parameter, one could have a fairly robust determination of the “optimal” number of clusters.





**Fig. 2.11** **a** Classification of the US market into four typical market states.  $k$ -means clustering is performed on a MDS map constructed from noise suppressed ( $\epsilon = 0.6$ ) similarity matrix [20]. The coordinates assigned in the MDS map are the corresponding correlation matrices constructed from short-time series of  $M = 20$  days and shifted by  $\Delta\tau = 10$  days. **b** shows the four different states of the US market as S1, S2, S3 and S4, where S1 corresponds to a calm state with low mean correlation, and S4 corresponds to a critical state (crash) with high mean correlation. **c** Temporal dynamics of the US market in four different states (S1, S2, S3 and S4) for the period of 1985 – 2016. **d** Network plot for transition probabilities of paired market states (MS). The transition probability of paired market states going from S1 and S2 to S4 is much lesser than 1% but from S3 to S4 is 6%. Figure adapted from Ref. [23]

In the new prescription, they measured the mean and the standard deviation of the intra-cluster distances using an ensemble of fairly large number (about 500) of different initial conditions (choices of random coordinates for the  $k$ -centroids or equivalently random initial clustering of  $n$  objects); each set of initial conditions usually results in slightly different clustering of the  $n$  objects representing different correlation matrices. If the clusters of points are very distinct in the coordinate space, then even for different initial conditions, the  $k$ -means clustering method yields same results, producing a small variance of the intra-cluster distance. However, the problem of allocating the matrices into the different clusters becomes problematic, when the clusters are very close or overlapping, as the initial conditions can then influence the final clustering of the different points; so there is a larger variance of the intra-cluster distance for the ensemble of initial conditions. Therefore, a minimum variance or standard deviation for a particular number of clusters implies the robustness of the clustering. For optimizing the number of clusters, Pharasi et al. proposed that one should look for *maximum*  $k$ , which has the *minimum* variance or standard

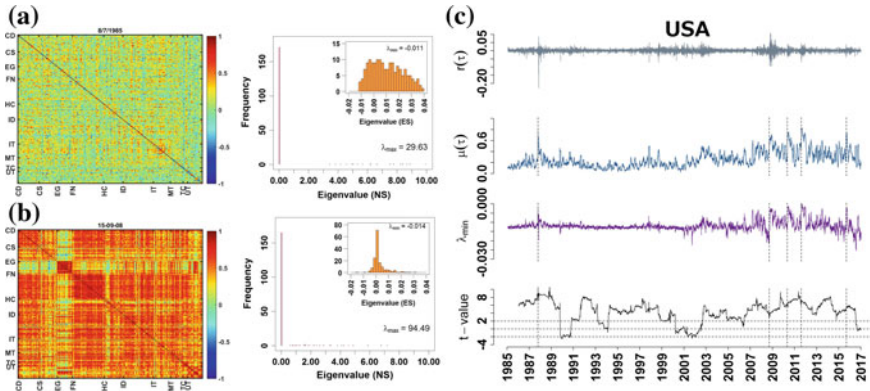
deviation in the intra-cluster distances with different initial conditions. Thus, based on the modified prescription of finding similar clusters of correlation patterns, they characterized the market states for USA and JPN.

Here, in Fig. 2.11b, we reproduce the results for the US market, showing four typical market states. The evolution of the market can be then viewed as the dynamical transitions between market states, as shown in Fig. 2.11c. Importantly, this method yields the correlation matrices that correspond to the critical states (or crashes). They correspond to the well-known financial market crashes and clustered in market state S4. They also analyzed the transition probabilities of the paired market states, and found that (i) the probability of remaining in the same state is much higher than the transition to a different states, and (ii) most probable transitions are the nearest neighbor transitions, and the transitions to other remote states are rare (see Fig. 2.11d). Most significantly, the state adjacent to a critical state (crash) behaved like a long-term “precursor” for a critical state, serving an early warning for a financial market crash.

### *Characterization of Catastrophic Instabilities*

Market crashes, floods, earthquakes, and other catastrophic events, though rarely occurring, can have devastating effects with long term repercussions. Therefore, it is of primal importance to study the complexity of the underlying dynamics and signatures of catastrophic events. Recently, Sharma et al. [8] studied the evolution of cross-correlation structures of stock return matrices and their eigenspectra over different short-time epochs for the US market and Japanese market. By using the power mapping method, they applied the non-linear distortion with a small value of distortion parameter  $\varepsilon = 0.01$  to correlation matrices computed for any epoch, leading to the *emerging spectrum* of eigenvalues.

Here, we reproduce some of the significant findings of the paper [8]. Interestingly, it is found that the statistical properties of the emerging spectrum display the following features: (i) the shape of the emerging spectrum reflects the market instability (see Fig. 2.12a, b), (ii) the smallest eigenvalue (in a similar way as the maximum eigenvalue, which captured the mean correlation of the market) indicated that the financial market had become more turbulent, especially from 2001 onward (see Fig. 2.12c), and (iii) the smallest eigenvalue is able to statistically distinguish the nature of a market turbulence or crisis—internal instability or external shock (see Fig. 2.12c). In certain instabilities the smallest eigenvalue of the emerging spectrum was positively correlated with the largest eigenvalue (and thus with the mean market correlation) while in other cases there were trivial anti-correlations. They proposed that this behavioral change could be associated to the question whether a crash is associated to intrinsic market conditions (e.g., a bubble) or to external events (e.g., the Fukushima meltdown). A lead-lag effect of the crashes was also observed through the behavior of  $\lambda_{min}$  and mean correlation, which could be examined further.



**Fig. 2.12** **a** Non-critical (normal) period of the correlation matrix and its eigenspectrum, evaluated for the *short* return time series for an epoch of  $M = 20$  days ending on 08-07-1985, with the maximum eigenvalue of the normal spectrum  $\lambda_{max} = 29.63$ . Inset: Emerging spectrum using power map technique ( $\varepsilon = 0.01$ ) is a deformed semi-circle, with the smallest eigenvalue of the emerging spectrum  $\lambda_{min} = -0.011$ . **b** Critical (crash) period of the correlation matrix and its eigenspectrum, evaluated for an epoch of  $M = 20$  days ending on 15-09-2008, with the maximum eigenvalue of the normal spectrum  $\lambda_{max} = 94.49$ . Inset: Emerging spectrum using power map technique ( $\varepsilon = 0.01$ ) is Lorentzian, with the smallest eigenvalue of the emerging spectrum  $\lambda_{min} = -0.014$ . **c** USA (i) market return  $r(t)$ , (ii) mean market correlation  $\mu(t)$ , (iii) smallest eigenvalue of the emerging spectrum ( $\lambda_{min}$ ), and (iv) t-value of the t-test, which tests the statistical effect over the lag-1 smallest eigenvalue  $\lambda_{min}(t-1)$  on the mean market correlation  $\mu(t)$ . The mean of the correlation coefficients and the smallest eigenvalue in the emerging spectra are correlated to a large extent. Notably, the smallest eigenvalue behaves differently (sharply rising *or* falling) at the same time when the mean market correlation is very high (crash). The vertical dashed lines correspond to the major crashes, which brewed due to internal market reactions. Note that, the smallest eigenvalue of the US market indicates that the financial market has become more turbulent from 2001 onward. Figure adapted from Ref. [8]

## Concluding Remarks

We have presented a brief overview of the Wishart and correlated Wishart ensembles in the context of financial time series analysis. We displayed the dependence of the length of the time series on the eigenspectra of the Wishart ensemble. The eigenspectra of large random matrices are not very sensitive to  $Q = T/N$ ; however, the amount of spurious correlations is dependent on it. To avoid the problem of non-stationarity and suppress the noise in the correlation matrices, computed over short epochs, we applied the power mapping method on the correlation matrices. We showed that the shape of the emerging spectrum depends on the amount of the correlation  $U$  of the correlated Wishart ensemble. We also studied the effect of the non-linear distortion parameter  $\varepsilon$  on the emerging spectrum.

Then we demonstrated the eigenvalue decomposition of the empirical cross-correlation matrix into market mode, group modes and random modes, using the

return time series of 194 stocks of S&P 500 index during the period of 1985-2016. The bulk of the eigenvalues behave as random modes and give rise to the Marčenko-Pastur. We also created surrogate correlation matrices to understand the effect of the sectoral correlations. Then we studied the eigenvalue distribution of those matrices as well as  $k$ -means clustering on the MDS maps generated from the correlation matrices. Evidently, if we have 10 diagonal blocks (representing sectors) then we get 10 clusters on a MDS map. Similarly, when we merged the four blocks to one and had 7 diagonal blocks then again we got 7 clusters on the MDS map.

Further, we studied the dynamical evolution of the statistical properties of the correlation coefficients using the returns of the S&P 500 stock market. We computed the mean, the absolute mean, the difference between absolute mean and mean, variance, skewness and kurtosis of the correlation coefficients  $C_{ij}$ , for short epochs of  $M = 20$  days and shifts of  $\Delta\tau = 1$  day and  $\Delta\tau = 10$  days. We also showed the evolution of the mean of correlation coefficients, maximum eigenvalue of the correlation matrix, as well as the number of negative eigenvalues and smallest eigenvalue of the emerging spectrum, for the same epoch and shift.

Finally, we discussed the applications of RMT in financial markets. In an application, we demonstrated the use of RMT and correlation patterns in identifying possible “market states” and long-term precursors to the market crashes. In the second application, we presented the characterization of catastrophic instabilities, i.e., the market crashes, using the smallest eigenvalue of the emerging spectra arising from correlation matrices computed over short epochs.

**Acknowledgements** The authors thank R. Chatterjee, S. Das and F. Leyvraz for various fruitful discussions. A.C. and K.S. acknowledge the support by grant number BT/B1/03/004/2003(C) of Govt. of India, Ministry of Science and Technology, Department of Biotechnology, Bioinformatics division, University of Potential Excellence-II grant (Project ID-47) of JNU, New Delhi, and the DST-PURSE grant given to JNU by the Department of Science and Technology, Government of India. K.S. acknowledges the University Grants Commission (Ministry of Human Resource Development, Govt. of India) for her senior research fellowship. H.K.P. is grateful for postdoctoral fellowship provided by UNAM-DGAPA. A.C., H.K.P., K.S. and T.H.S. acknowledge support by Project CONACyT Fronteras 201, and also support from the project UNAM-DGAPA-PAPIIT IG 100616.

## References

1. Bar-Yam, Y.: General Features of Complex Systems. Encyclopedia of Life Support Systems (EOLSS). UNESCO, EOLSS Publishers, UK (2002)
2. Bendat, J.S., Piersol, A.G.: Engineering Applications of Correlation and Spectral Analysis, p. 315. Wiley-Interscience, New York (1980)
3. Bouchaud, J.P., Potters, M.: Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management. Cambridge University Press, Cambridge (2003)
4. Cartan, É.: Sur les domaines bornés homogènes de l'espace des variables complexes. In: Abhandlungen aus dem mathematischen Seminar der Universität Hamburg, vol. 11, pp. 116–162. Springer, Berlin (1935)

5. Chakraborti, A., Muni Toke, I., Patriarca, M., Abergel, F.: Econophysics review: I. empirical facts. *Quant. Financ.* **11**(7), 991–1012 (2011)
6. Chakraborti, A., Muni Toke, I., Patriarca, M., Abergel, F.: Econophysics review: II. agent-based models. *Quant. Financ.* **11**(7), 1013–1041 (2011)
7. Chakraborti, A., Patriarca, M., Santhanam, M.: Financial time-series analysis: a brief overview. In: *Econophysics of Markets and Business Networks*, pp. 51–67. Springer, Berlin (2007)
8. Chakraborti, A., Sharma, K., Pharasi, H.K., Das, S., Chatterjee, R., Seligman, T.H.: Characterization of catastrophic instabilities: market crashes as paradigm (2018). [arXiv:1801.07213](https://arxiv.org/abs/1801.07213)
9. Chen, C.P., Zhang, C.Y.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Inf. Sci.* **275**, 314–347 (2014)
10. Gell-Mann, M.: What is complexity? *Complexity* **1**, 16–19 (1995)
11. Guhr, T., Kälber, B.: A new method to estimate the noise in financial correlation matrices. *J. Phys. A: Math. Gen.* **36**(12), 3009 (2003)
12. Hua, L.: *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*, vol. 6. American Mathematical Society (1963)
13. Jin, X., Wah, B.W., Cheng, X., Wang, Y.: Significance and challenges of big data research. *Big Data Res.* **2**(2), 59–64 (2015)
14. Leviandier, L., Lombardi, M., Jost, R., Pique, J.P.: Fourier transform: a tool to measure statistical level properties in very complex spectra. *Phys. Rev. Lett.* **56**(23), 2449 (1986)
15. Mantegna, R.N., Stanley, H.E.: *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge (2007)
16. Marčenko, V.A., Pastur, L.A.: Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik* **1**(4), 457 (1967)
17. Martinez, M.M.R.: *Caracterización estadística de mercados europeos*. Master’s thesis, UNAM (2018)
18. Mehta, M.L.: *Random Matrices*. Academic (2004)
19. Mikosch, T., Stărică, C.: Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *Rev. Econ. Stat.* **86**(1), 378–390 (2004)
20. Münnix, M.C., Shimada, T., Schäfer, R., Leyvraz, F., Seligman, T.H., Guhr, T., Stanley, H.E.: Identifying states of a financial market. *Sci. Rep.* **2**, 644 (2012)
21. Ochoa, S.: *Mapeo de Guhr-Kaelber aplicado a matrices de correlación singulares de dos mercados financieros*. Master’s thesis, UNAM (2018)
22. Pandey, A., et al.: Correlated Wishart ensembles and chaotic time series. *Phys. Rev. E* **81**(3), 036202 (2010)
23. Pharasi, H.K., Sharma, K., Chatterjee, R., Chakraborti, A., Leyvraz, F., Seligman, T.H.: Identifying long-term precursors of financial market crashes using correlation patterns. *New J. Phys.* **20**, 103041 (2018). [arXiv:1809.00885](https://arxiv.org/abs/1809.00885)
24. Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Guhr, T., Stanley, H.E.: Random matrix approach to cross correlations in financial data. *Phys. Rev. E* **65**(6), 066126 (2002)
25. 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 (1999)
26. Schäfer, R., Nilsson, N.F., Guhr, T.: Power mapping with dynamical adjustment for improved portfolio optimization. *Quant. Financ.* **10**(1), 107–119 (2010)
27. Sharma, K., Shah, S., Chakraborti, A.S., Chakraborti, A.: Sectoral co-movements in the Indian stock market: a mesoscopic network analysis, pp. 211–238 (2017)
28. Shuryak, E.V., Verbaarschot, J.: Random matrix theory and spectral sum rules for the Dirac operator in QCD. *Nuclear Phys. A* **560**(1), 306–320 (1993)
29. Sinha, S., Chatterjee, A., Chakraborti, A., Chakraborti, B.K.: *Econophysics: an Introduction*. Wiley, New York (2010)
30. Utsugi, A., Ino, K., Oshikawa, M.: Random matrix theory analysis of cross correlations in financial markets. *Phys. Rev. E* **70**(2), 026110 (2004)
31. Vemuri, V.: *Modeling of Complex Systems: An Introduction*. Academic, New York (1978)

32. Vinayak, Prosen, T., Buča, B., Seligman, T.H.: Spectral analysis of finite-time correlation matrices near equilibrium phase transitions. *Europhys. Lett.* **108**(2), 20006 (2014)
33. Vinayak, Schäfer, R., Seligman, T.H.: Emerging spectra of singular correlation matrices under small power-map deformations. *Phys. Rev. E* **88**(3), 032115 (2013)
34. Vinayak, Seligman, T.H.: Time series, correlation matrices and random matrix models. In: *AIP Conference Proceedings*, vol. 1575, pp. 196–217. AIP (2014)
35. Vyas, M., Guhr, T., Seligman, T.H.: Multivariate analysis of short time series in terms of ensembles of correlation matrices (2018). [arXiv:1801.07790](https://arxiv.org/abs/1801.07790)
36. Wigner, E.: Ep wigner. *Ann. Math.* **53**, 36 (1951)
37. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. *Ann. Math.* 325–327 (1958)
38. Wigner, E.P.: Random matrices in physics. *SIAM Rev.* **9**(1), 1–23 (1967)
39. Wishart, J.: The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* 32–52 (1928)
40. Yahoo finance database. <https://finance.yahoo.co.jp/> (2017). Accessed 7 July 2017, using the R open source programming language and software environment for statistical computing and graphics