Chapter 11
Sectoral Co-movements in the Indian Stock Market: A Mesoscopic Network Analysis

Kiran Sharma, Shreyansh Shah, Anindya S. Chakrabarti, and Anirban Chakraborti

Abstract In this article, we review several techniques to extract information from large-scale stock market data. We discuss recurrence analysis of time series, decomposition of aggregate correlation matrices to study co-movements in financial data, stock level partial correlations with market indices, multidimensional scaling, and minimum spanning tree. We apply these techniques to daily return time series from the Indian stock market. The analysis allows us to construct networks based on correlation matrices of individual stocks on one hand, and on the other, we discuss dynamics of market indices. Thus, both microlevel and macrolevel dynamics can be analyzed using such tools. We use the multidimensional scaling methods to visualize the sectoral structure of the stock market and analyze the co-movements among the sectoral stocks. Finally, we construct a mesoscopic network based on sectoral indices. Minimum spanning tree technique is seen to be extremely useful in order to group technologically related sectors, and the mapping corresponds to actual production relationship to a reasonable extent.

11.1 Introduction

In this paper, we present a coherent analysis of the Indian stock market employing several techniques recently proposed in the econophysics literature. Stock market is a fascinating example of a rapidly evolving multi-agent interacting system that

K. Sharma (✉) • A. Chakraborti
School of Computational and Integrative Sciences, Jawaharlal Nehru University, 110067, New Delhi, India
e-mail: kiransharma1187@gmail.com; anirban@jnu.ac.in

S. Shah
Indian Institute of Technology, Banaras Hindu University, 221005, Varanasi, India
e-mail: shreyansh.shah.mec13@iitbhu.ac.in

A.S. Chakrabarti
Economics Area, Indian Institute of Management, 380015, Ahmedabad/Vastrapur, Gujarat, India
e-mail: anindyac@iima.ac.in

Y. Aruka, A. Kirman (eds.), *Economic Foundations for Social Complexity Science*, Evolutionary Economics and Social Complexity Science 9,
DOI 10.1007/978-981-10-5705-2_11

generates an enormous amount of very well-defined and well-documented data. Due to the sheer volume of data, it becomes possible to construct large-scale correlation matrices across stocks that contain information about the aggregate market. Thus, the loss of information due to aggregation can be minimized to a great extent. Several useful techniques to analyze such large-scale data have been proposed, and there are multiple resources reviewing them. Interested readers can refer to Mantegna and Stanley (1999) and Bouchaud and Potters (2009) for excellent and quite extensive textbook expositions.

We present a series of analysis on the Bombay stock exchange, using both macroscale and microscale data. Even though there are separate attempts in a few other papers that presented analysis on similar datasets, this probably is the first attempt to systematically analyze Indian stock market data in a comprehensive manner. At the beginning of the discussion on every technique, we point out the papers that proposed the techniques and subsequent analyses, if any, on Indian or any other emerging market data.

India, being an emerging market, is an interesting example. Several papers Pan and Sinha (2007) and Bastos and Caiddo (2011) have pointed out that there are systematic differences between the dynamic behaviors of developed economies and emerging economies. Earlier hypothesis was that, as the financial market develops in a country, the market dynamics changes monotonically or at least in a clearly discernible way. Although it sounds intuitive, there is no clear demonstration of changing dynamical structure of the markets along with the process of development (Kuyyamudi et al. 2015). Instead, what we find is that, in a cross-sectional sense, such differences in the market behavior exist across countries.

In the following, we focus only on the Indian stock market. To summarize the findings, we see that the recurrence analysis is not very useful for the present dataset. Correlation decomposition techniques do not show any strong group correlation structure, which is consistent with the literature (Pan and Sinha 2007). Different clustering algorithms have been applied to understand sectoral concentration. Finally, we end with a section on sectoral correlation networks. A nontrivial finding is that technologically related sectors show very similar kind of fluctuations in the stock market. We quantify the relationship using network theoretic tools.

11.2 Nonlinear Dynamics: Recurrence Plot Analysis

For a very long time, it had been conjectured that the stock market indices may have certain features of highly nonlinear dynamical system. It originated from certain speculations that economic systems, in general, may show chaotic behavior (see, e.g., Baumol and Benhabib 1989). Brock and Sayers (1988) considered an idea that the aggregate macro dynamics of an economy may show chaotic behavior. By and large, such theories are no longer considered to be useful descriptions of economic dynamics. However, in recent times there have been some attempts to analyze the stock index behavior, by using recurrence analysis based on phase space reconstruction.

In general, the technique's usefulness comes from the fact that it is nonparametric, does not make any assumptions about the data, and can work with nonstationary data. In particular, the technique is useful for detecting sudden large change in a time series. A stock market crash has often been thought as a phase transition indicating a large abrupt change in the behavior (Sornette 2004). However, the technique is useful for recovering patterns in potentially highly nonlinear but recursive systems, an assumption that is not satisfied by the stock market. We follow the mode of analysis presented in details in Guhathakurta et al. (2010) and Bastos and Caiddo (2011).

Here, we describe the construction of recurrence plots. It is based on the idea of recurrence within a phase space, and the plot exhibits times, when a nonlinear system revisits the same phase space during the process of evolution. Consider a time series $\{x(i)\}_{i=1}^{N}$ representing an index of a stock market. We know, from Takens' theorem (Takens 1981), that it is possible to extract information about the phase space from the time series (see also Bastos and Caiddo 2011). We start by embedding $\{x\}$ into an m dimensional space given by

$$y(i) = [x(i), x(i+\delta), x(i+2\delta), \ldots, x(i+(m-1)\delta)] \tag{11.1}$$

where δ is the time delay. Together these two parameters constitute the set of embedding parameters. Thus $y(i)$ is a point in the m dimensional Euclidean space, representing the evolution of the system in the reconstructed phase space. We collect all such $y(i)$'s and present element-by-element difference with Euclidean norm to create a two-dimensional plot. Such a plot exhibits if there is any recurrence as explained below.

Let us define a matrix R such that its i,j-th elements ($i,j = 1,\ldots,n$, with $n = N-(m-1)\delta$) are expressed as

$$R_{ij}(\varepsilon) = \begin{cases} 0 & \text{if } |y(i) - y(j)| > \epsilon \\ 1 & \text{if } |y(i) - y(j)| \leq \epsilon \end{cases}$$

where $\|.\|$ is the Euclidean norm and ϵ is the threshold applied which is a positive real number. Recurrence plots are exactly symmetric along the diagonal.

Inference based on structures: In recurrence plots, we see multiple patterns including dots, as well as diagonal, vertical and horizontal lines, and all possible combinations of them.

- Isolated points exist if states are rare, or persistence is low, or if they represent high fluctuations.
- Existence of a diagonal line $R_{i+m,j+m} = 1$ (for $m = 1,\ldots,l$ where l is the length of the diagonal line) indicates presence of recurrence, i.e., a segment of the time

series revisits the same area in the phase space at a lag. If there are lines parallel to the line of identity, it represents the parallel evolution of trajectories.

- Existence of a vertical/horizontal line $R_{i,j+m} = 1$ (for $m = 1, \ldots, v$ where v is the length of the line) indicates a stage during evolution where the system gets trapped for some time and does not evolve fast. This can be an intermittent behavior.

Now we conduct recurrence quantification analysis (RQA) by studying the structure of the plots numerically. Such an analysis is based essentially on densities of isolated points, diagonal lines, as well as vertical lines. We borrow the discussion presented below from Bastos and Caiddo (2011). The measures, which we have considered, are as follows:

- RR: Recurrence rate.
- DET: Fraction of points in the plot forming diagonal lines. This indicates determinism and hence predictability.
- $\langle L \rangle$: Average lengths of the diagonal lines.
- LMAX: Length of the longest diagonal line (except the line of identity). Its inverse is associated with the divergence of the trajectory in phase space.
- ENTR: Shannon entropy defined over the distribution of lengths of diagonal lines indicates diversity of the diagonal lines.
- LAM: Fraction of points forming vertical lines indicates existence of laminar states in the system.
- TT: Average length of the vertical lines. This value estimates the trapping time.

We have computed the RQA measures for the BSE index, under a range of embedding dimension. In each case, we have set the delay equal to 1.

In Figs. 11.1 and 11.2, we present recurrence analysis on logarithmic return series ($r_\tau = \ln P_\tau - \ln P_{\tau-1}$) constructed from BSE index data. As it is apparent, there is no clearly discernible pattern in the data. Next, we follow the standard approach and use the level of the price data upon normalization by the maximum value of the time series ($\tilde{P}_\tau = (P_\tau - P_{\min})/P_{\max}$). Table 11.1 contains the RQA measures. Most of the prior literatures on stock market data consider high values of embedding parameter. It is evident that, in general, recurrence rates are very low and determinism is very high. However, this approach has an inherent problem that it is not particularly good at differentiating non-recursive series from recursive series. In general, we found that when we construct similar measures for standard recursive series, it is not clear from such RQA measures that they can be easily separated from a stochastic series. Thus, it does not really shed much light on the problem, as the primary focus is to figure out determinism or lack thereof. Bastos and Caiddo (2011) discusses a possible application that these measures still retain some usefulness for cross-country analysis. Since in this case we are focusing on one country only, it is not very helpful. So, we consider a fully stochastic framework in the rest of the analysis.

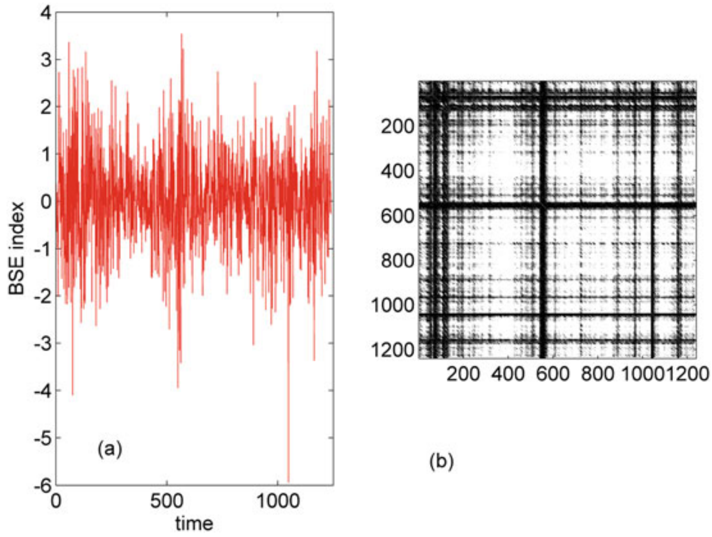

Fig. 11.1 *Left panel*: Normalized daily return series constructed from BSE index data for five years (June 6, 2011, to June 6, 2016). *Right panel*: Recurrence plot constructed from the same data with an embedding dimension equals to 11 and time delay 1

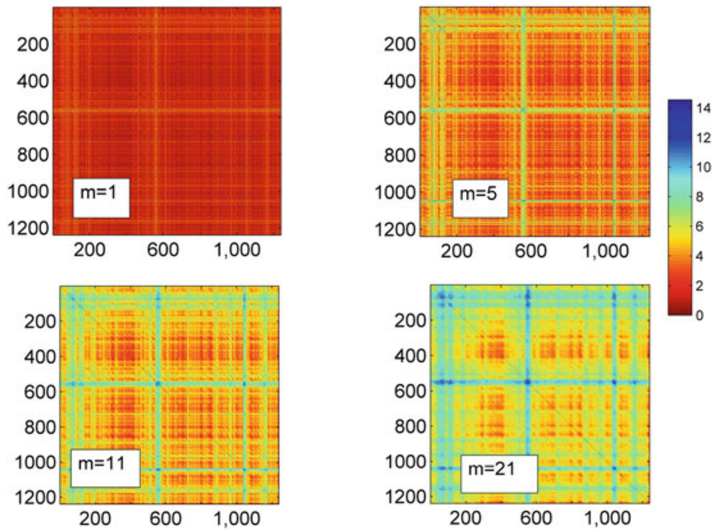

Fig. 11.2 Distance plots constructed from the BSE index for different values of the embedding dimensions ($m = 1, 5, 11, 21$)

Table 11.1 Measures based on recurrence analysis of normalized BSE data. Generated by the CRP toolbox (Marwan et al. 2002, 2007). Threshold for calculating neighbors set at the default value 0.1

Quantity	$m = 1$	$m = 2$	$m = 5$	$m = 11$
RR	0.0758	0.0442	0.0075	3.1049×10^{-4}
DET	0.9029	0.8817	0.8516	0.9211
$\langle L \rangle$	4.3854	3.9302	3.8841	4.6667
LMAX	146	88	58	18
ENTR	2.0319	1.8575	1.8095	1.8527
LAM	0.9479	0.9074	0.7234	0.2763
TT	5.7416	4.4451	3.2874	2.2703

11.3 Empirical Study of the Correlation Structure of the Indian Stock Market

In this section, we analyze the empirical cross-correlation matrices constructed from the stock market data.

11.3.1 Data Specification, Notations, and Definitions

In order to study correlations and co-movements in the stock price time series, the popular Pearson correlation coefficient was commonly used. However, with the electronic markets producing data at different frequencies (low to high), it is now known that several factors, viz., the statistical uncertainty associated with the finite-size time series, heterogeneity of stocks, heterogeneity of the average inter-transaction times, and asynchronicity of the transactions, may affect the applicability/reliability of this estimator. In this article, we have mainly focused on the daily returns computed from closure prices, for which the Pearson coefficient works well.

11.3.1.1 Dataset

We have used the freely downloadable daily adjusted closure prices from Yahoo finance for $N = 199$ companies in the Bombay stock exchange (BSE) SENSEX (https://in.finance.yahoo.com/q/hp?s=%5EBSESN), for 5 years, over a period spanning from June 6, 2011, to June 6, 2016. Also, we have downloaded 199 stock prices of companies chosen randomly from the BSE and 13 sectoral indices of the BSE, for the period May 27, 2011, to May 27, 2016. The lists are given in the Tables 11.2 and 11.3, in Appendices A and B, respectively.

11.3.2 Correlation Matrices

We construct the correlation matrix from individual stock returns in the following way.

11.3.2.1 Pearson Correlation Coefficient

In order to study the equal time cross correlations between N stocks, we first denote the adjusted closure price of stock i in day τ by $P_i(\tau)$ and determine the logarithmic return of stock i as $r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1)$. For the window of T consecutive trading days, these returns form the return vector r_i. We use the equal time Pearson correlation coefficients between stocks i and j defined as

$$C_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{[\langle r_i^2 \rangle - \langle r_i \rangle^2][\langle r_j^2 \rangle - \langle r_j \rangle^2]}}, \quad (11.2)$$

where $\langle \ldots \rangle$ indicates an average over the window of T successive trading days in the return series. Naturally, such correlation coefficients satisfy the usual condition of $-1 \leq C_{ij} \leq 1$, and we can create an $N \times N$ correlation matrix C by collecting all values (Chakraborti 2006; Tilak et al. 2012). By construction, the matrix is symmetric, and it serves as the basis of the rest of the present article.

11.3.3 Decomposition Analysis

For the present section, we are following the sequence of methods discussed by Pan and Sinha (2007) which is one of the first few papers that applied this technique. Suppose we have N return time series of length T that are pairwise uncorrelated. The correlation matrix generated by collecting all pairwise correlations for N of such series is called the Wishart matrix. In the limits $N \to \infty$ and $T \to \infty$, such that the ratio $Q \equiv T/N > 1$, the eigenvalue distribution of this matrix has a specific distributional form,

$$f(\lambda) = (Q/2\pi)\frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}, \quad (11.3)$$

for $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$ and 0 otherwise. This distribution is clearly bounded by $\lambda_{\max,\min} = [1 \pm (1/\sqrt{Q})]^2$. In the BSE data, we considered $Q = 5$. Thus, the Wishart matrix should have the following bounds: $\lambda_{\min} = 0.3056$ and $\lambda_{\max} = 2.0944$. The distribution of eigenvalues unexplained by the Wishart matrix sheds light on the interaction structures and the coevolution process of the stocks in the market.

The largest eigenvalue corresponds to the market mode which captures the aggregate dynamics of the market that is common across all stocks. The eigenvectors associated with the next few eigenvalues (we took the next five dominant eigenvalues) describe the sectoral dynamics. The rest of the eigenvectors correspond to the random mode. From such a segregation, it is possible to reconstruct the contributions of different modes to the aggregate correlation matrix.

Following the literature to filter the data to remove market mode and the random noise, we first decompose the aggregate correlation matrix as

$$C = \sum_{i=0}^{N-1} \lambda_i a_i a_i^T, \tag{11.4}$$

where λ_i are the eigenvalues of the correlation matrix C. An easy way to handle the reconstruction of the correlation matrix is to sort the eigenvalues in descending order. Then we rearrange the eigenvectors a_i in corresponding ranks. This allows us to decompose the matrix into three separate components, viz., market, group, and random:

$$C = C^M + C^G + C^R, \tag{11.5}$$

$$= \lambda_0 a_0 a_0^T + \sum_{i=1}^{N_G} \lambda_i a_i a_i^T + \sum_{i=N_G+1}^{N-1} \lambda_i a_i a_i^T, \tag{11.6}$$

where N_G is taken to be 5, i.e., it corresponds to the five largest eigenvalues except the first one. It is worth noting that the exact value of N_G is not crucial for the result as long as it is kept within the same ballpark. The decomposition is shown in Fig. 11.3.

An important finding is that the group mode almost coincides with the random mode, whereas the market mode is segregated by a large margin from the rest. Thus, the sectoral dynamics are almost absent, whereas the market mode is very strong. This is in line with the prior literature (see, e.g., Pan and Sinha 2007).

Following standard procedure (see, e.g., Pan and Sinha 2007), we also calculate the inverse participation ratio (IPR) to extract information about contribution of different stocks to the eigenvalues. IPR is defined for the k-th eigenvector as the sum of fourth power of all individual components of the corresponding eigenvector, $I_k \equiv \sum_{i=1}^{N} [a_{ki}]^4$, where a_{ki} are the components of eigenvector k. The result is presented in Fig. 11.4. Intuitively, if a single stock dominates in terms of contribution to any particular eigenvector, then the IPR would go to 1. For example, consider a limiting case of $a_{k1} = 1$ and $a_{ki} = 0$ for $i \neq 1$. On the other hand, if all elements were equal to $1/\sqrt{N}$, then we would get IPR $= 1/N$. Thus by considering IPR, we can understand if there is significant contribution coming from specific stocks or a more diversified bundle of stocks.

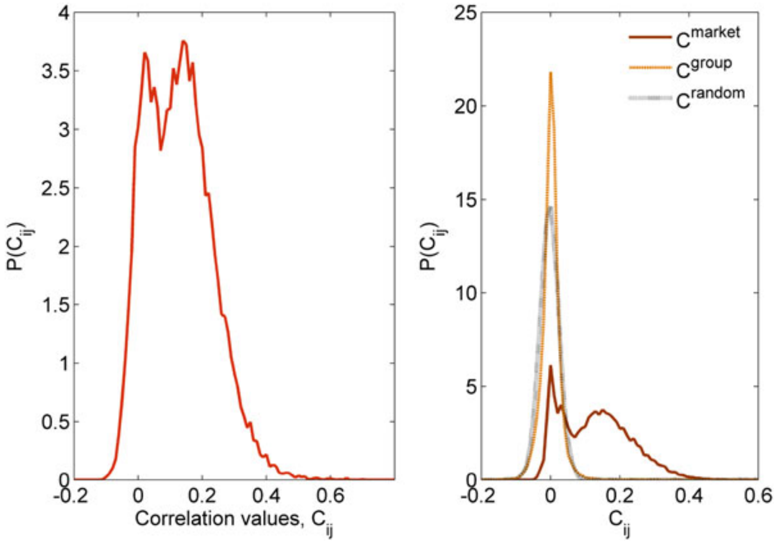

Fig. 11.3 *Left panel*: Probability density function of the cross-correlation coefficients of 199 BSE stocks. *Right panel*: Decomposition of the correlation matrix into market mode, group mode, and random mode

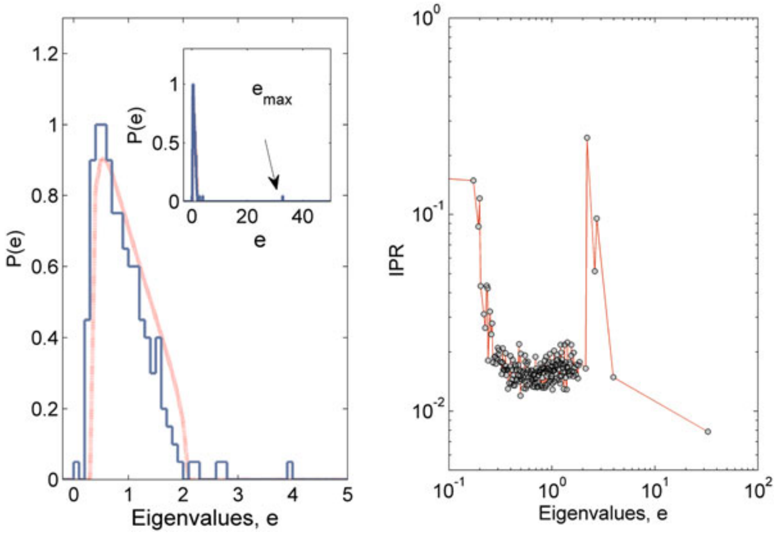

Fig. 11.4 Eigenvalue decomposition of the correlation matrix. *Left panel*: Probability density function of eigenvalues. Inset shows the full distribution. *Right panel*: Inverse participation ratio with respect to the corresponding eigenvalues

11.3.4 Partial Correlation Analysis

Partial correlation is a newly introduced tool to investigate the effects of third stock on the correlation between pairs of stocks. Kenett et al. (2015) introduced this analysis for multiple stock markets. In the present paper, we apply their technique to the Bombay stock exchange data. To describe its usefulness, consider three stocks, i, j, and k, with significant correlation between all three pairs of the stocks. Suppose we think that the high value of C_{ij} is the result of their own correlations with k, i.e., part of C_{ij} might be spurious correlation arising from a third variable effect (in this case k), we should remove such effects to figure out the actual correlation across i and j. Then, we can recalculate C_{ij}, after controlling for the effect of k. The resultant correlation value is called the partial correlation. The difference between the raw correlation value for a pair of stocks and the corresponding partial correlation tells us how much third variable effect was there.

For this purpose, we again use the same daily log return $r_i(t)$. However, we need to adjust for one more factor. From the preceding analysis, we already know that there is a significant market mode. Therefore, that will act as a common driving factor. Hence, the market mode should also be controlled in order to extract the actual correlation values for the exact same reason. In this case, the market mode is given by a market index. Note the difference from the earlier analysis. For constructing the market mode from the eigenvalue analysis, the market mode arises endogenously from the panel data itself, whereas in this case, we take the market mode to be given by an exogenous index time series. Hence, these two types of analysis complement each other.

Following the notation of Kenett et al. (2015), let x and y be two time series and let M be the BSE index for the same time frame. The partial correlation, $(x, y|M)$, is defined as the standard Pearson correlation coefficient (described above) between x and y after controlling for M. More technically, this is the correlation between the residuals of x and y which are unexplained by the market index represented by M. So first, we need the residuals of the two time series. A simple way to do it would be to regress both on M. Then we can work with the resulting variables. Formally, the correlation is given as

$$C_{x,y|M} = \frac{(C_{x,y} - C_{x,M}.C_{y,M})}{\sqrt{[1 - C_{x,M}^2].[1 - C_{y,M}^2]}} \tag{11.7}$$

In the same way, when the same two stocks x and y are affected by a common stock z, we can control for that effect as well. Given a third stock z, the partial correlation between x and y after controlling for both the market factor and the third stock z is given by the following formula:

$$C_{x,y|M,z} = \frac{C_{x,y|M} - C_{x,z|M}.C_{y,z|M}}{\sqrt{[1 - C_{x,z|M}^2].[1 - C_{y,z|M}^2]}} \tag{11.8}$$

If it is found that the third stock has an important effect on pairs of stocks, then it is useful to define the "influence quantity" (see Kenett et al. 2015)

$$d(x, y|z) = C_{x,y|M} - C_{x,y|M,z}. \tag{11.9}$$

Magnitude of this quantity will reflect how much influence the third stock has on a certain pair of stocks. A natural extension of this idea is to consider the average influence $d(x|z)$ of stock z on the correlations between a given stock x and all other stocks except x itself and z. Kenett et al. (2015) defined this index as the following:

$$d(x|z) = \langle d(x, y|z)\rangle_{y\neq x}. \tag{11.10}$$

This quantity captures the average influence from stock z to stock x through the third variable effect after controlling for the market index. We present all results of our analysis in Fig. 11.5. Panel (a) shows the correlation coefficients of all stocks after controlling for the market index. Since the bulk of it is below the 45° line, we conclude that the market index has a positive effect on pairwise correlations. This is consistent with the results from the eigenvalue analysis and also with Kenett et al. (2015). Similarly, in panel (b), we show the data for the same correlation coefficients, after controlling for all possible third variable effects. In panel (d),

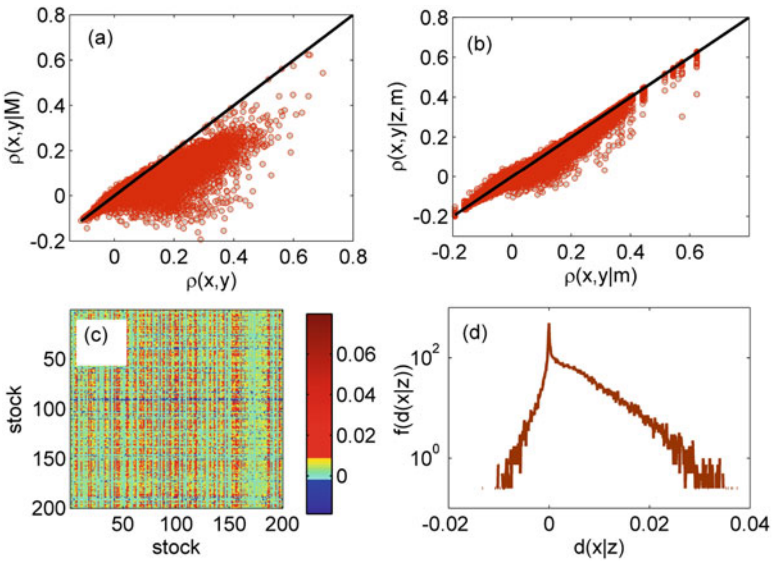

Fig. 11.5 Correlation matrix after controlling for market mode (BSE index). Panel (**a**): Partial correlation after controlling for the market mode as a function of raw correlation coefficients. Panel (**b**): Same after controlling for the third variable effect. Panel (**c**): Influence of all stocks as the third variable (on x-axis) on all other stocks (on y-axis). Panel (**d**): Probability density function of average influence quantity

we present the probability density function of the *influence quantity*. Again bulk of the distribution is in the positive quadrant implying positive effect on average.

11.4 Network Analysis

In this section, we present network analysis based on the empirical correlation matrix.

11.4.1 Distance Metric

To obtain "distances," the following transformation

$$d_{ij} = \sqrt{2(1 - C_{ij})} \tag{11.11}$$

is used, which clearly satisfies $2 \geq d_{ij} \geq 0$. Collecting all distances, one can form an $N \times N$ distance matrix D, such that all elements of the matrix are "ultrametric" (Rammal et al. 1986). The concept of ultrametricity appears in multiple papers. Interested readers can refer to the detailed discussions by Mantegna (1999), Onnela et al. (2003a,b), and Chakraborti (2006) among others. There are multiple possible ultrametric spaces. We opt for the subdominant ultrametric, as it is simple to work with, and its associated topological properties. The choice of the nonlinear function is again arbitrary, as long as all the conditions of ultrametricity are met.

11.4.2 Multidimensional Scaling (MDS)

Multidimensional scaling is a method to analyze large-scale data that displays the structure of similarity in terms of distances, given by Eq. (11.11), as a geometrical picture or map, where each stock corresponds to a set of coordinates in a multidimensional space. MDS arranges different stocks in this space according to the strength of the pairwise distances between stocks, – two similar stocks are represented by two sets of coordinates that are close to each other, and two stocks behaving differently are placed far apart (see Borg and Groenen 2015) in the space. We construct a distance matrix consisting of $N \times N$ entries from the N time series available, defined using Eq. (11.11):

$$\begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{21} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{bmatrix}$$

Given D, the aim of MDS is to generate N vectors $x_1, \dots, x_N \in \Re^D$, such that

$$\|x_i - x_j\| \approx d_{ij} \qquad \forall i, j \in N, \tag{11.12}$$

where $\|.\|$ represents vector norm. We can use the Euclidean distance metric as is done in the classical MDS. Effectively, through MDS we try to find a mathematical embedding of the N objects into $\Re^D$ by preserving distances. In general, we choose the embedding dimension D to be 2, so that we are able to plot the vectors x_i in the form of a map, representing N stocks. It may be noted that x_i are not necessarily unique under the assumption of the Euclidean metric, as we can arbitrarily translate and rotate them, as long as such transformations leave the distances $\|x_i - x_j\|$ unaffected. Generally, MDS can be obtained through an optimization problem, where $(x_1, \dots, x_N)$ is the solution of the problem of minimization of a cost function, such as

$$\min_{x_1,\dots,x_N} \sum_{i<j} (\|x_i - x_j\| - d_{ij})^2. \tag{11.13}$$

In order to capture the sectoral behavior of the market visually, we have generated the MDS plot of 199 stocks as described before, for the time window of 250 trading days between May 2015 and May 2016. As before, using the correlation matrix as input, we computed the distance matrix using the transformations (given by Eq. (11.11)). The distance matrix was then used as an input to the inbuilt MDS function in MATLAB (http://in.mathworks.com/help/stats/cmdscale.html). The outputs of the MDS were the sets of coordinates, which were plotted as the MDS map as shown in Fig. 11.6.

The coordinates are plotted in a manner such that the centroid of the map coincides with the origin $(0, 0)$. It is interesting to follow the positions of certain sectors, (i) sugar, (ii) textiles, and (iii) pharmaceuticals, which will be discussed in details in Sect. 11.5.

11.4.3 Dendrogram

Dendrogram is basically a tree diagram. This is often used to depict the arrangement of multiple nodes through hierarchical clustering. We have used the inbuilt function in MATLAB (http://in.mathworks.com/help/stats/dendrogram.html) to generate the

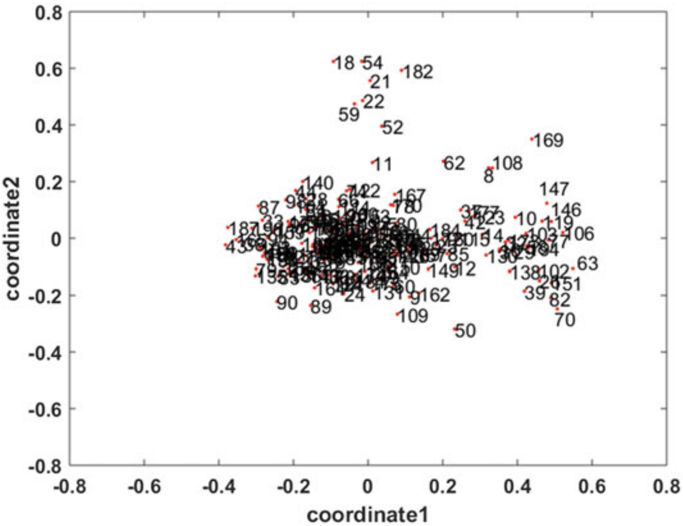

Fig. 11.6 Multidimensional scaling of the sample data for the time window May 2015–May 2016. There is a cluster of stocks with identifiers 18, 54, 21, etc. at the top, all of which belong to the sugar industry (Appendix A)

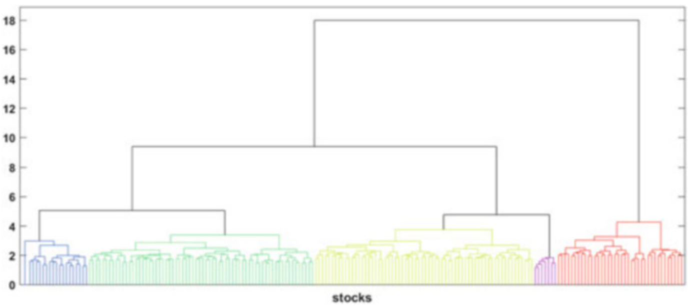

Fig. 11.7 Dendrogram of 199 stocks

hierarchical binary cluster tree (dendrogram) of N stocks connected by many U-shaped lines (as shown in Fig. 11.7), such that the height of each U represents the distance (given by Eq. (11.11)) between the two data points being connected. Thus, the vertical axis of the tree captures the similarity between different clusters, whereas the horizontal axis represents the identity of the objects and clusters. Each joining (fusion) of two clusters is represented on the graph by the splitting of a vertical line into two vertical lines. The vertical position of the split, shown by the short horizontal bar, gives the distance (similarity) between the two clusters. We set the property "linkage type" as "Wards minimum variance," which requires the distance method to be Euclidean that results in group formation such as the pooled within-group sum of squares which would be minimized. In other words, at every iteration, two clusters in the tree are connected such that it results in the least possible increment in the relevant quantity, i.e., pooled within-group sum of squares. Figure 11.7 shows the dendrogram of all the 199 stocks clustered in five different colors (by using "color threshold" property in MATLAB). The magenta color represents the cluster of "sugar industries."

11.4.4 Minimum Spanning Tree

A minimum spanning tree is a spanning tree of a connected, undirected graph such that all the N vertices are connected together with the minimal total weighting for its $N-1$ edges (total distance is minimum). The distance matrix defined by Eq. (11.11) was used as an input to the inbuilt MST function in MATLAB (http://in.mathworks.com/help/bioinfo/ref/graphminspantree.html). See MATLAB documentation for all details. Here we state Kruskal and Prim algorithms for the sake of completeness of the present article. Description of the two algorithms (source: see http://in.mathworks.com/help/bioinfo/ref/graphminspantree.html):

- Kruskal – *This algorithm extends the minimum spanning tree by one edge at every discrete time interval by finding an edge, which links two separate trees in a spreading forest of growing minimum spanning trees*.
- Prim – *This algorithm extends the minimum spanning tree by one edge at every discrete time interval by adding a minimal edge, which links a node in the growing minimum spanning tree with one other remaining node*.

Figure 11.8 shows the MST for all the 199 stocks. MATLAB algorithms set the root node as the first node in the largest connected component, which in our case is node 43.

11.5 Sectoral Co-movements: Mesoscopic Network

After quantifying the general cross-correlation structure of the market, we probe deeper into the sectoral co-movements. There are multiple ways to analyze the data. First, we can impose a threshold on the group cross-correlation matrix and

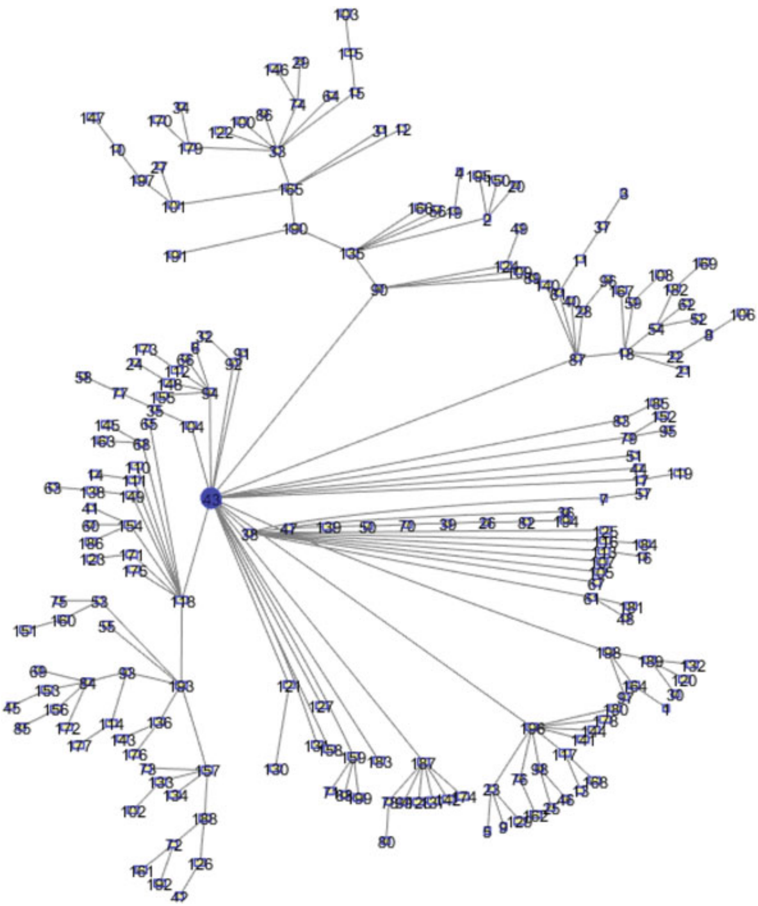

Fig. 11.8 Minimum spanning tree (MST) of the sample data. For such microlevel data, there is no clear pattern. However, the pattern becomes much clearer with sectoral data (Sect. 11.5)

construct a network of stocks which move closely. This is the approach that is followed in Pan and Sinha (2007), for example. This approach has some problems. First, the threshold has to be exogenous, hence basically arbitrary. Second, even with such networks, it is difficult to identify clusters that match with actual industry classifications. An alternative way is to follow the industry classifications first, and then try to see if they form clusters.

To study the sectoral behavior in the market, we have selected stocks from the list of BSE from the industries: (i) sugar, (ii) textiles, and (iii) pharmaceuticals. Following the same methodologies, as described in the previous Sects. 11.4.2, 11.4.3, and 11.4.4, we have generated the plots given in Figs. 11.9, 11.10, and 11.11. By looking at the diagram, it becomes clear that the method is partially successful to segregate the market into clusters, but not fully. Therefore, we construct a new

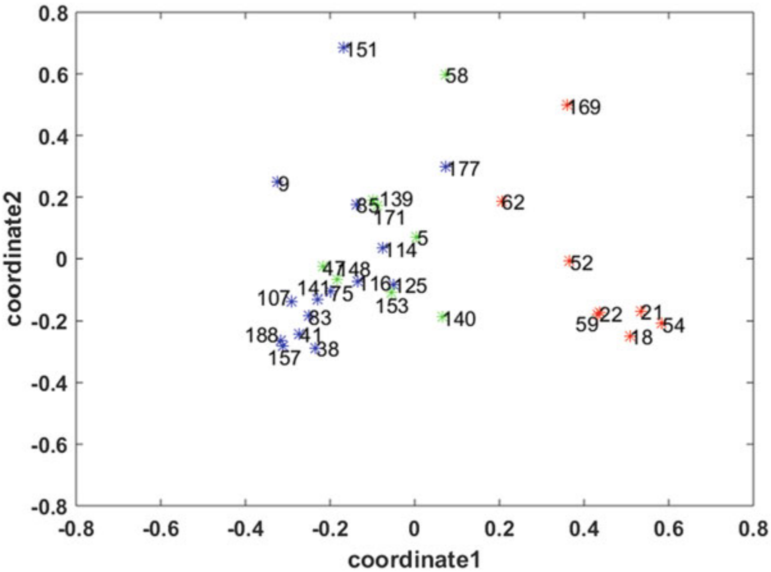

Fig. 11.9 Plot of MDS for stocks within (i) sugar (*red*), (ii) textile (*blue*), and (iii) pharmaceuticals (*green*) industries. Stock details are given in Appendix B

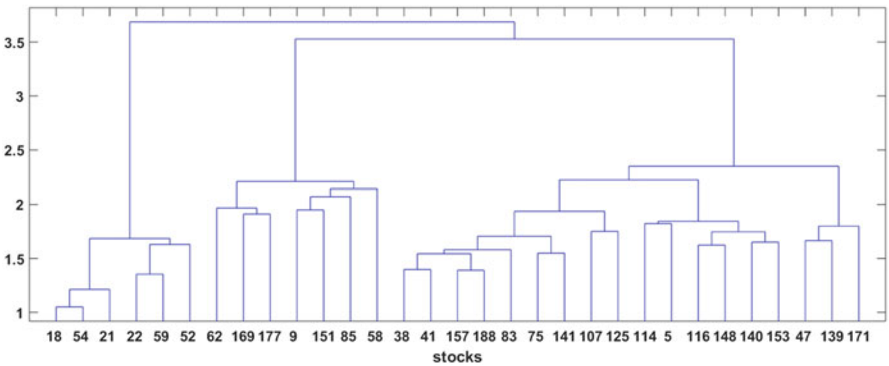

Fig. 11.10 Plot of dendrogram for stocks within (i) sugar, (ii) textile, and (iii) pharmaceuticals industries. Stock details are given in Appendix B

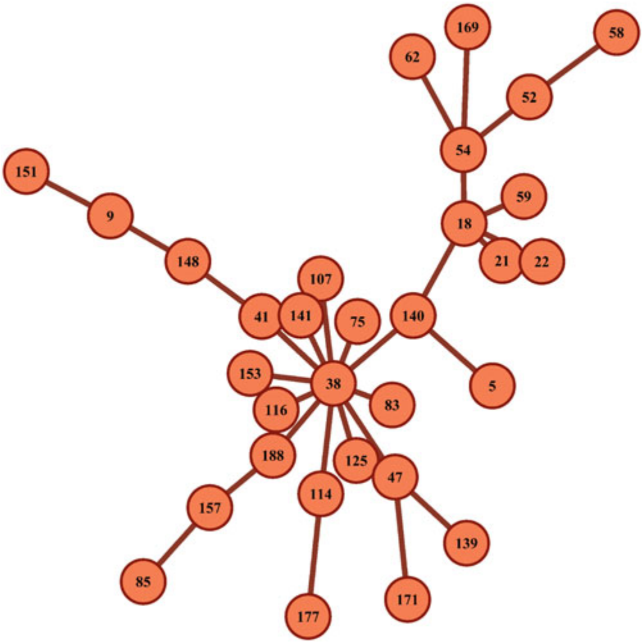

Fig. 11.11 Plot of minimum spanning tree (MST) for stocks within (i) sugar, (ii) textile, and (iii) pharmaceuticals industries. Stock details are given in Appendix B

network. Rather than working with actual stock returns, we work with sectoral index returns. This marks a prominent departure from the usual mode of analysis. Typically, most studies focus on either an aggregate macrolevel market index like S&P 500 or consider collective dynamics of microlevel individual stock returns. Here, we consider a *mesoscopic* network to characterize correlations.

Empirically, we used the 13 sectoral indices from the BSE (list given in Appendix B) for the time window May 2015–May 2016. The resulting multidimensional scaling results have been plotted in Fig. 11.12, dendrogram in Fig. 11.13, and minimum spanning tree in Fig. 11.14.

The MDS algorithm cannot segregate the markets into clusters in a way that corresponds to the industry classifications. Dendrogram produces better results than that. Finally, the minimum spanning tree corresponds to a fairly intuitive market structure. Note that the only information used was sectoral returns' correlations. The MST shows that the banks and realty sectors are most closely related to the finance sector. Energy sector is most closely associated with oil and gas sector and so on. Thus we see that the sectoral MST approximates the industrial relations in a fairly correct manner.

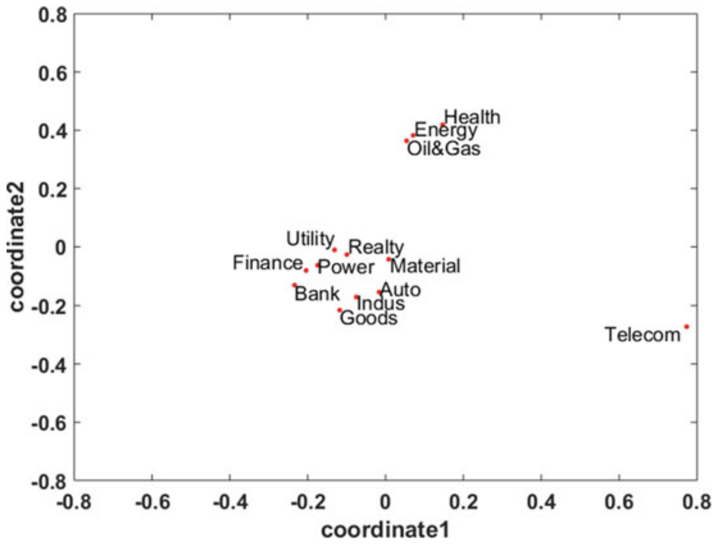

Fig. 11.12 Plot of clustering with multidimensional scaling (MDS) algorithm is shown for the BSE indices. As it is clear from the figure, it is not very useful for segregating sectors, at least with the present sample. See Appendix A for sector details

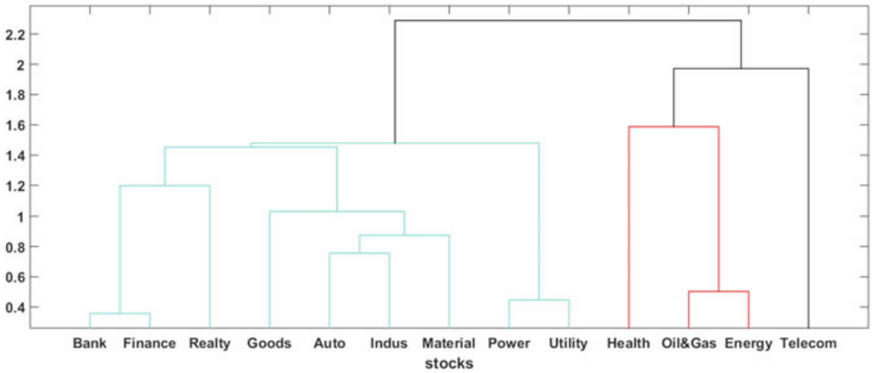

Fig. 11.13 Plot of dendrogram for the BSE indices. This algorithm clusters related sectors. But MST (Fig. 11.14) presents a clearer picture. See Appendix A for sector details

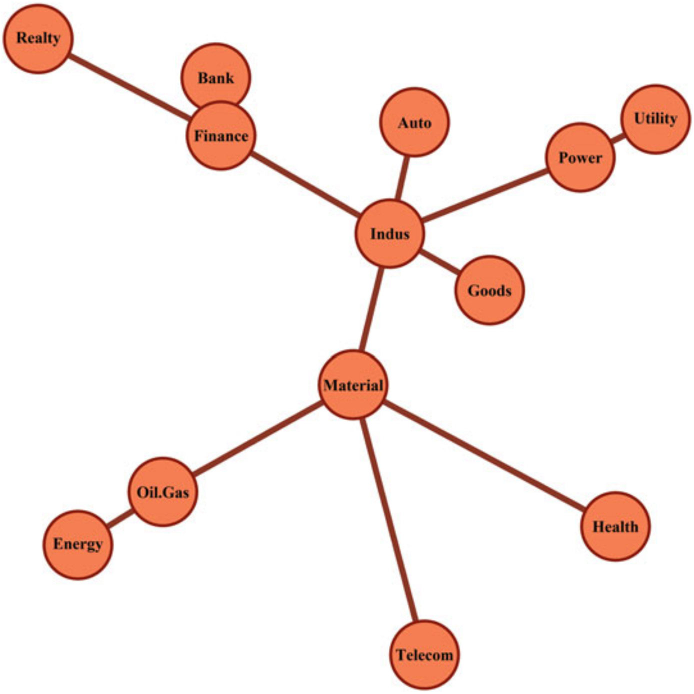

Fig. 11.14 Plot of the minimum spanning tree (MST) of BSE indices. Technologically related sectors are closer to each other, e.g., bank and realty are related to finance only. Thus return fluctuations of technologically related sectors co-move significantly more than other sectors. See Appendix A for sector details

11.6 Summary

In this article, we have applied multiple techniques to analyze daily data from Bombay stock exchange. Our analyses cover a large spectrum of tools proposed in the econophysics literature in the last two decades. Using eigen decomposition method, we show that the market cross-correlation structure shows a very prominent market mode. Consistent with the literature, we show that the group mode is not very strong for emerging countries and, in fact, is very difficult to differentiate from the random mode. Then we carry out partial correlation analysis, which is a newly proposed method, on the Indian data. This helps us to explicitly characterize and quantify the average “third variable” effect in the cross correlations.

Finally, we turn to network analysis to study the core-periphery structure. We use multidimensional scaling and dendrograms to identify clusters. In general, we do not find any significant pattern between such clusters and the industrial classifications. However, a much more intuitive picture emerges when we construct a mesoscopic network with the sectoral indices. We see that minimum spanning tree across the indices clearly segregates nodes according to their industrial classification, just by using the return cross correlations.

A Appendix

Table 11.2 List of all sectoral indices. The first column has the serial number, the second column has the abbreviation, the third column has the full name of the sector, and the fourth column has the category of the sector as given in the BSE

S.No.	ID	Name	Category
1	SI1900	S&P BSE AUTO	AUTO
2	SIBANK	S&P BSE BANKEX	BANK
3	SPBSBMIP	S&P BSE BASIC MATERIALS	MATERIAL
4	SI0200	S&P BSE CAPITAL GOODS	GOODS
5	SPBSEIP	S&P BSE ENERGY	ENERGY
6	SPBSFIIP	S&P BSE FINANCE	FINANCE
7	SPBSIDIP	S&P BSE INDUSTRIALS	INDUS
8	SI1400	S&P BSE OIL & GAS	OIL & GAS
9	SIPOWE	S&P BSE POWER	POWER
10	SIREAL	S&P BSE REALTY	REALTY
11	SPBSTLIP	S&P BSE TELECOM	TELECOM
12	SPBSUTIP	S&P BSE UTILITIES	UTILITY
13	SI0800	S&P BSE HEALTHCARE	HEALTH

B Appendix

Table 11.3 List of all stocks considered for the analysis. The first column has the serial number, the second column has the abbreviation, the third column has the full name of the stock, and the fourth column specifies the sector as given in the BSE

S.No.	ID	Name	Category
1	ABB	ABB India Limited	Heavy electrical equipment
2	ABIRLANUVO	Aditya Birla Nuvo Ltd.	Diversified
3	AEGISLOG	Aegis Logistics Ltd.	Oil marketing and distribution
4	AMARAJABAT	Amara Raja Batteries Ltd.	Auto parts and equipment
5	AMBALALSA	Ambalal Sarabhai Enterprises Ltd.	Pharmaceuticals
6	ANDHRAPET	ANDHRA PETROCHEMICALS LTD.	Commodity chemicals
7	ANSALAPI	ANSAL PROPERTIES and INFRASTRUCTURE LTD.	Realty
8	APPLEFIN	APPLE FINANCE LTD.	Finance (including NBFCs)
9	ARVIND	ARVIND LTD.	Textiles
10	ASIANHOTNR	ASIAN HOTELS (NORTH) LIMITED	Hotels

(continued)

Acknowledgements This research was partially supported by the institute grant, IIM Ahmedabad. KS thanks University Grants Commission (Ministry of Human Research Development, Govt. of India) for her junior research fellowship. AC acknowledges financial support from the institutional research funding IUT (IUT39-1) of the Estonian Ministry of Education and Research, grant number BT/BI/03/004/2003(C) of Govt. of India, Ministry of Science and Technology, Department of Biotechnology, Bioinformatics division, and University of Potential Excellence-II grant (Project ID-47) of the Jawaharlal Nehru University, New Delhi, India.

References

J.A. Bastos, J. Caiddo, Recurrence quantification analysis of global stock markets. Physica A **390**(7), 1315–1325 (2011)

W. Baumol, J. Benhabib, Chaos: significance, mechanism, and economic applications. J. Econ. Perspect. **3**, 77–105 (1989)

I. Borg, P. Groenen, *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York, 2005)

J.-P. Bouchaud, M. Potters, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridge, 2009)

W. Brock, C. Sayers, Is the business cycle characterized by deterministic chaos? J. Monet. Econ. **22**, 71–90 (1988)

A. Chakraborti, An outlook on correlations in stock prices, in *Econophysics of Stock and Other Markets*, ed. by A. Chatterjee, B.K. Chakrabarti (Springer, Milan, 2006). Available at arXiv:physics/0605246

K. Guhathakurta, B, Bhattacharya, A. Roy Chowdhury, Using recurrence plot analysis to distinguish between endogenous and exogenous stock market crashes. Physica A **389**(9), 1874–1882 (2010)

D. Kenett, X. Huang, I. Vodenska, S. Havlin, H.E. Stanley, Partial correlation analysis: applications for financial markets. Quant. Finance **15**(4), 569–578 (2015)

C. Kuyyamudi, A.S. Chakrabarti, S. Sinha, Long-term evolution of the topological structure of interactions among stocks in the New York stock exchange 1925–2012, in *Econophysics and Data Driven Modelling of Market Dynamics*, ed. by F. Abergel et al. (Springer, Milan, 2015)

R.N. Mantegna, Hierarchical structure in financial markets. Eur. Phys. J. B **11**, 193 (1999)

R.N. Mantegna, H.E. Stanley, *Introduction to Econophysics* (Cambridge University Press, Cambridge, 1999)

N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence plot based measures of complexity and its application to heart rate variability data. Phys. Rev. E, **66**, 026702 (2002)

N. Marwan, M.C. Romano, J. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems. Phys. Rep. **438**(5–6), 237–329 (2007)

J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Dynamics of market correlations: taxonomy and portfolio analysis. Phys. Rev. E **68**, 056110 (2003)

J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Asset trees and asset graphs in financial markets. Phys. Scr. T **106**, 48 (2003)

R.K. Pan, S. Sinha, Collective behavior of stock price movements in an emerging market. Phys. Rev. E **76**, 046116 (2007)

R. Rammal, G. Toulouse, M.A. Virasoro, Ultrametricity for physicists. Rev. Mod. Phys. **58**, 765 (1986)

D. Sornette, *Why Stock Markets Crash?* (Princeton University Press, Princeton, 2004)

F. Takens, Detecting strange attractors in turbulence, in *Dynamicl Systems and Turbuence*, ed. by D. Rand, L. Young (Springer, Berlin, 1981), pp. 366–381

G. Tilak, T. Szell, R. Chicheportiche, A. Chakraborti, Study of statistical correlations in intraday and daily financial return time series, in *Econophysics of Systemic Risk and Network Dynamics*, ed. by F. Abergel et al. (Springer, Milan, 2012)