รายงานโครงการวิจัยฉบับสมบูรณ์

(Full report)

เรื่อง : Project Title:

Physics of complex systems: Economic, Finance and social systems

โดย

Dr. Pradeep Bhadola

รายงานวิจัยฉบับสมบูรณ์เป็นส่วนหนึ่งของการรายงานผลงานวิจัย
ที่ได้รับทุนสนับสนุนจากกองบริหารการวิจัย
เลขที่สัญญา R2562C002 ประจำปีงบประมาณ พ.ศ. 2562
มหาวิทยาลัยนเรศวร

ข้อมูลสรุปโครงการวิจัย (Summary and Conclusion)

รหัสโครงการ (Grant Code) : THB 200,000/-

ชื่อโครงการ (Project Title) : Physics of complex systems: Economic, Finance and social systems

ชื่อผู้ทำวิจัย (Name of researcher) : Dr. Pradeep Bhadola

ชื่อผู้ร่วมวิจัย (Name of Co-researcher): NA

Email: pradeepb@nu.ac.th

Abstract

The study is aimed at understanding the structure, function and evolution of the complex financial, and biological systems. We have developed tools from the physics, mathematics and information theory that can be used to acquire profound and new insights into topology, structure, function and evolution of these systems. This study forms a substantial ties between the field of physics, mathematics, economics, sociology, information theory and network theory.

For the financial system we study the global financial indices, foreign exchange (world currencies) and the Thai stock market SET index. In this work, we develop statistical, spectral, and network methods for the financial time series analysis. The focus is mainly on the correlation structure, its statistical and spectral properties. We find the cross correlation matrix between different financial commodities such as currencies, stock and indices. New information theoretic measures such as "eigenvector entropy" is developed which quantify the eigenvector localization as well as the information content of the eigenvector. A detailed statistical and spectral analysis is performed on the correlation matrix, created from a moving window method to study the dynamics and evolution of the system. We use different method to create network from the correlation matrix. The main network studied during the project are the minimum spanning tree (MST) and the correlation threshold network (CTN). We calculate and compare the results of two different system (1) global financial indices of 31 countries and (2) foreign exchange rates of 21 different currencies. By comparing their properties, dynamics, structure and evolution using a rolling window method.

Lastly, we study the structure, dynamics and evolution of the Thai stock market SET index. In this problem, we study more than 100 stocks from the Thai stock market to understand how the Thai stock market evolves from 2000 to 2018. The analysis is mainly based on, spectral analysis and network theory. From our analysis, we can cluster stock into sectors such as agricultural, healthcare, banking etc. The detailed analysis reveals how the properties changes when the market is undergoing crash or in the growth phase.

The next system under consideration is the protein family. A physio-chemical property based approach is proposed for the analysis of protein sequences. In this new approach, we represent a protein sequence as a multidimensional time-series and we estimate the positions in the protein that can control or affect its function. With the use of entropy measures such as property based Shannon entropy and Kullback-Leibler divergence, we find that certain substitutions preserve some physio-chemical properties. Doing a detailed analysis, we find that these are crucial for working of the protein family. The entropic measures shows that the physio-chemical properties values are more conserved than the type of amino acids during the evolutionary history of the protein family. The protein family under study is beta-lactamase which are the enzymes secreted by bacteria against the antibiotics such as penicillin. The study of the family shed lights on the working of the family. Finally we perform a network analysis to create the network of positions for different physio-chemical properties, which shows how different positions are interacting with each other.

Chapter 1 Introduction

1.1 Background and Research Problem

Background

Statistical mechanics have provided a theoretical framework for studying systems from diverse fields which not only includes physical, or chemical systems, but also been extended to the interdisciplinary fields of complex systems such as economical, biological and social systems. The application of the tools and models of statistical mechanics to the complex financial and biological systems had led to the development of the interdisciplinary fields known as "Econophysics" [1], "Biophysics" and "Sociophysics"[2]. The recent years have witnessed the growth of these branches to develop new and innovative approaches for understanding the behavior of complex economic, biological and social systems which are characterized by large number of interacting degrees of freedom giving rise to a non-trivial collective and emergent behavior. This have attracted many physicists to investigate and study these systems not only for the immediate practical application such as efficient allocation of the resources, determining the risk in financial investment, estimating the bubble and crashes in the economy or understanding the formation of consensus or biases in society but also for the pure scientific interest for having a deep understanding of the structure and dynamics of these complex systems. The growth of the field is further facilitated by the two main factors: (1) availability of large electronically accessible data sets, and (2) advancement in computational capabilities and high-throughput technologies for simulation and processing of data. These two factors have enabled researchers to analyze and process large data sets for the patterns and features that characterizes the underlying principles and laws manifesting the dynamics and evolution of these systems. The financial/stock markets and social system comes under the category of the evolving complex systems where the evolution results in the change in the market structure and components by introducing new products, strategies, alliances, partnerships, technologies along with many other factors. Stock market are very good in recording and maintaining. The availability of rich and wide range of data sets make them best suited for modeling and then validating the developed tools, methods and theories through the empirical data. There are many open question that are under active investigation such as statistical characterization of the stochastic process of price changes of a financial commodity, identify the reasons and factors for the tipping of the financial systems, search for the universal laws in these systems, develop new techniques to evaluate market risk etc. The stock and financial markets are of prime interest for an individual as well as for the society or country as a whole. Their interests in the market may be due to a variety of different reasons including the individual financial gain due to the investment in the market or above all, studying the disastrous impact on the country's economy due to global crash which may lead to situation of the confusion in society or social tension. Human social system is also an example of complex adaptive system, where the individuals have interactions with each other like friendship, economic ties, political relation etc. These interactions leads to the formation of a collective property of the society as a whole like the development of common culture and language, the emergence of consensus, or biases towards a specific group or organization.

Research Problem

we propose to study the statistical and nonlinear properties of the complex systems including biological systems, financial markets linking with the social biases. The first aim is to develop and formulate simple and understandable theory and measures to characterize the biological system and its state. The second aim of the proposed research is to understand the structure and dynamics of the biological and financial markets, understanding the reasons for the changes and evolution in the systems and how the reasons (such as market sentiments, social biases, political changes, socioeconomic factors) effects the market dynamics.

During the course of this project, we are working towards the understanding of structure, function, and evolution of the complex systems including biological financial and social systems. In the project, we devise computational tools to understand these systems based on the data analysis. In this project we mainly address the following on research problems:

1) Understand the dynamics and interaction of foreign exchange (FX) rates. As FX market directly or indirectly affects all other financial exchanges or markets therefore peeks the interest of academic community to study the structure, statistical properties and topology of the FX network.
2) We will try to investigate the local as well as global financial indices before, during and after a crisis, trying to find out the reasons behind the crisis, by the comparing the structure and properties of the market at the different stages of crisis (before, during and after).
3) Understanding the structure and the connection of system involving the global financial indices for 31 different countries and foreign exchange rates of 21 different currencies across the globe.
4) Since the community structure and correlations among the financial commodities are of crucial importance for the study of financial dynamics and portfolio risk estimation, therefore the clustering and the community structure of the financial commodities will be studied and new tools will be developed to efficiently address this problem using the information theory, random matrix theory and graph theory.
5) Investigating the structure, evolution, and dynamics of the Thai stock market (SET index). We will study how to cluster stocks into sectors such as agricultural, banking, consumer products etc.

6) Studying the working mechanism of enzyme family beta-lactamase responsible for the drug resistance in the bacteria against antibiotics. In this work, other enzymes and protein families are also considered. We study the evolution of the protein sequence how the physio-chemical properties affect the structure, function and dynamics of the families. The work also explores the possibility to find the crucial interaction and positions within the family that decides and control the function of the family.

1.2 Research Objective

The objectives of the research are listed below.

1) Find ways to manage, extract and use the information encoded in large datasets generated from various complex systems.
2) To explore the properties and behavior of the complex system and the relation with randomness, long-range interaction among agents, tolerance to noise, evolution and adaptability of system along with the systems topological and hierarchical structure.
3) Use the electronically available large data sets, to make data based computational models to make predictions and find the universal principles that govern the structure and evolution of all financial market.
4) Investigate the statistical characterization of the stochastic process of price changes of a financial commodity, identification of the reasons and factors for the tipping of the financial systems and to search for universal laws in financial and social systems.
5) To understand the evolution and adaptation of financial system. As these systems are prone to fatal crashes or abrupt changes which are the outcome of simple changes in the system.
6) Working and mechanism of action of the biological molecules such as protein and enzymes from the sequence data .

1.3 Theoretical / Assumptions (if any) and Concept Framework

Econo-physicists rely primarily on the empirical data to construct models of the financial system and validate it, thus majorly dealing with the analysis of the empirical data in the form a of financial time series. This empirical analysis have led the researchers to make the realistic models of the financial systems which are capable of addressing and explaining features in the observed data and system. To address the question quantitatively, we intended to integrate the advanced computational methods with the predictive modeling techniques like random matrix theory, graph theory, information theory and then rigorously verifying with comparing with the experimental results or real data sets. The research methodology will include a broad range of both the computational and the analytical techniques. The analytical work consists of developing the appropriate theoretical models based on statistical mechanics, random matrix and graph theory. Since the financial and social systems are the systems that have well organized and large electronically available data sets, which make it possible to make data based computational models and prediction about the working of the each system. The lost lasting goal will be to find the universal principle that govern the structure and evolution of all these systems.

Since it is believed that the price of a financial commodity includes all the fundamental factors that affects it, therefore we intended to use the advance information theoretic methods to extract the useful information contained in it. We will random matrix theory which have served as an effective and useful tool in the important and diverse fields of physics (disordered condensed matter systems and quantum chromodynamics), information theory, and more recently in interdisciplinary studies with applications to bio-physics and econo-physics [3-5] for the noise dressing of the financial time series. Our plan is to then exploit the information contained in the multiple time series to study the interaction and effect of the price change of various financial assets by studying the correlations then linking these factors with the structural and functional constraints to determine the important group position and sectors. Correlation matrices have been extensively studied in physics and forms is the major component of the Markowitz's theory of optimal portfolios [6]. The spectral analyses of the correlation matrices will be done for the identification of the important ties, motifs and clusters of the interacting assets.

The number of protein sequences are growing at an exponential rate. Therefore, it is very important to develop efficient tools and techniques to determine the structure and function encoded in the protein sequence data. The existing theoretical models and computational methods are unable to correctly determine the function and structure of protein sequence. The ability of proteins to perform important biochemical function is the result of interactions and dynamics of the constituent amino acids. Determining the residues and interactions responsible for a structure and function of a protein requires challenging theory and experiments. The existing methods treats protein sequences as a sequence of letters and the analysis is based on the occurrence of amino acids and their frequencies, estimating mutual information (MI), without taking the physio-chemical properties of amino acid into account. Literature studies have shown physio-chemical properties play a crucial role in determining the native state of a protein as well as the evolution and functional properties, by influencing the mutation rate. To address the problem of linking the sequence to structure and function, we propose a physio-chemical property based approach in which we represent a protein sequence using different physio-chemical properties as a multidimensional time series, where each dimension represents a property value. Secondly, most existing methods are focused on the prediction of functional sites but do not address the problem of the functional motifs and interactions responsible for the protein function. This research aims to achieve this by using the spectral and network approach.

1.4 (Information) Literature / Information related

Internationally, there is a great deal of interest in the problems from econophysics and sociophysics, both computationally and theoretically. Several groups around the globe are studying the financial markets, their properties, dynamics, structure, topology by various tools than spans mathematics, physics, information theory, statistics and computer science. Large number of monographs [11-14], journal research articles and reviews [15-17], books [1, 2, 6, 18] and edited volumes [19-27] have been published in the field.

From a pure statistical point of view a lot work have been done which explains the price change of a financial commodity by simple decision making process by investors [28] or using a pure statistical based model such as Ising model to identify the correlation among the financial assets [29]. One of the major milestone in the economics was done by Bachelier [30] where the financial time series was modeled using the random walk concept and was before the seminal work of Einstein on Brownian motion. One of the crucial factors in the financial market is the identification of the non random and strong correlation among different assets, industries or commodities. These interactions can then be used for clustering of assents into sectors and identification of community structure which is very vital for the portfolio risk estimation and financial dynamics. Therefore, many studies is being dedicated to the estimation of the interaction and correlations among the stocks and assets [5, 7, 31-40]. Many tools have been used to identify filter the non random correlation, where random matrix theory [5] is of prime importance to filter non random effects. The spectral analysis of the correlation matrices together with the random matrix theory estimates the highly interacting assets and classify them as a single community or sector [41-42]. This information is proved very useful in the systematic risk estimation and portfolio optimization[18, 43]. Many studies have shows that the financial market has underlying network structure. The first attempt [7] to create a network of the stock prices were done by Mantagna and the idea was then extended to a vast degree of research papers [44-50] which proved very useful in understanding the structure and topology of financial market and explaining many of its properties.

In the subject of socio-economic inequality among citizen of a county, from the long time have grown as a subject of prime importance. Not only economist or sociologist but researcher from other disciplines as well as philosophers have keenly devoted their time to understand the issue of the socioeconomic inequality [51-54]. The issue is further explored by understanding how socio-economic inequality effect over all economic growth of the country [55-58] and it has been established that socioeconomic inequality have a very deep impact on the political scenario of the country [59-62]. The question raised in the field such as the structure of wealth distribution of the society have led researchers to do extensive studies and a century ago in the seminal work of Fredo Pareto [63], it was found that the wealth distribution follow a power law tail now known as Pareto law. The research is going at a very fast rate with huge amount of literature and research activities in this field, and the references are too many to be possibly included here but few key and important references are included it the proposal.

1.5 (Reference)

1) P. Bhadola, N Deo, Network Theory and Agent-Based Modeling in Economics and Finance, 331-351, (2019).
2) P. Bhadola, N. Deo, New Perspectives and Challenges in Econophysics and Sociophysics, 133-145, (2019).
3) P. Bhadola, N. Deo, Journal of Physics: Conference Series 1144 (1), 012083(2018)
4) P Bhadola, N Deo, Econophysics and Sociophysics: Recent Progress and Future Directions, 59-69, (2017)
5) P Bhadola, N Deo, Physical Review E 94 (4), 042409 (2016),
6) B.M. Roehner, "Patterns of Speculation: A Study in Observational Econophysics", (Cambridge University Press, Cambridge, 2005).
7) U. Garibaldi, E. Scalas, Finitary probabilistic methods in econophysics (Cambridge University Press, Cambridge, 2010).
8) M. L. Mehta, Random matrices (Academic Press, 1991); T. Guhr, A. Mueller Groeling and H. A. Weidenmueller, Phys. Rep. 299, 189 (1998).
9) Pradeep Bhadola, and N. Deo. Physical Review E 88.3, 032706. (2013); P. Bhadola, I. Garg, and N. Deo, Nuclear Physics B 870.2 384-396 (2013). Pradeep Bhadola and Nivedita Deo. Physical Review E 94.4, 042409 (2016).

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

11) H. Markowitz, Portfolio Selection: Efficient Diversification of Investments (J. Wiley and Sons, New York, 1959); E. J. Elton and M. J. Gruber, Modern Portfolio Theory and Investment Analysis (J. Wiley and Sons, New York, 1995).

12) R. N. Mantegna, The European Physical Journal B-Condensed Matter and Complex Systems, 11(1), 193–197 (1999), M. Tumminello, T. Aste, T. Di Matteo, R.N. Mantegna, A tool for filtering information in complex systems. Proceedings of the National Academy of Sciences of the United States of America 102(30), 10421–10426 (2005).

13) F. Allen, A. Babus, Networks in finance. Wharton Financial Institutions Center Working Paper; 2008, S. H. Strogatz, Exploring complex networks. Nature 410(6825), 268–276 (2001).

14) B.M. Roehner, "Patterns of Speculation: A Study in Observational Econophysics", (Cambridge University Press, Cambridge, 2005).

15) U. Garibaldi, E. Scalas, Finitary probabilistic methods in econophysics (Cambridge University Press, Cambridge, 2010).

16) H. Aoyama, Y. Fujiwara, Y. Ikeda, H. Iyetomi, and W. Souma, Econophysics and Companies: Statistical Life and Death in Complex Business Networks, (Cambridge University Press, New York, 2011).

17) B. K. Chakrabarti, A. Chakraborti, S. R. Chakravarty, and A. Chatterjee, Econophysics of Income and Wealth Distributions, (Cambridge University Press, Cambridge, 2013).

18) V. Yakovenko, J. Barkley Rosser Jr., Colloquium: Statistical mechanics of money, wealth and income, Reviews of Modern Physics 81, 1703 (2009).

19) A. Chakraborti, I. Muni Toke, M. Patriarca and F. Abergel, Quantitative Finance 11(7), 991 (2011),

20) R. N. Mantegna and H.E. Stanley, An Introduction to Econophysics : Correlations and Complexity in Finance, Cambridge University Press (2000); J. P. Bouchaud and M. Potters, Theory of Financial Risk and Derivative Pricing, Cambridge University Press (2000); S. Sinha, A. Chatterjee, A. Chakraborti, B. K. Chakrabarti, Econophysics: An Introduction, Wiley-VCH, Berlin, 2011.

21) B.K. Chakrabarti, A. Chakraborti, and A. Chatterjee (Eds) Econophysics and Sociophysics: Trends and Perspectives, (Wiley-VCH, Weinheim, 2006), Eds. F. Abergel, H. Aoyama, B.K. Chakrabarti, A. Chakraborti and A. Ghosh, Econophysics of Agent-based models, to be published by Springer-Verlag (Italia), Milan, 2013.

22) Eds. F. Abergel, B.K. Chakrabarti, A. Chakraborti and A. Ghosh, Econophysics of systemic risk and network dynamics (Springer-Verlag (Italia), Milan, 2012), Eds. F. Abergel, B.K. Chakrabarti, A.

Chakraborti and M. Mitra, Econophysics of order driven markets (Springer- Verlag (Italia), Milan, 2011).

23) Eds. L. Pareschi, G. Naldi and G. Toscani, Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences (Birkhauser, Berlin, 2010).

24) Eds. A. Chatterjee and B.K. Chakrabarti, Econophysics of Markets and Business Networks (Springer-Verlag (Italia), Milan, 2007), Eds. A. Chatterjee and B.K. Chakrabarti, Econophysics of Stock and other Markets (Springer- Verlag (Italia), Milan, 2006), Eds. A. Chatterjee, S. Yarlagadda and B.K. Chakrabarti, Econophysics of Wealth Distributions (Springer-Verlag (Italia), Milan, 2005).

25) T. Kaizoji, Physica A 287, 493 (2000), T. Ibuki, S. Higano, S. Suzuki, J. Inoue and A. Chakraborti, to appear in J. Phys. Conf. Ser. (2013), Z.-Y. Zheng, B. Podobnik, L. Feng, and B.-W. Li, Sci. Rep. 2, 888 (2012), M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, J. Financial Econ. 104, 535 (2012) , M. Kritzman, Y.-Z. Li, S. Page, and R. Rigobon, J. Portf. Manag. 37, 112 (2011).

26) A. Garas and P. Argyrakis, EPL (Europhys. Lett.) 86, 28005 (2009), S. M. Cai, Y.-B. Zhou, T. Zhou, and P.-L. Zhou, Int. J. Mod. Phys. C 21, 433 (2010), M. Tumminello, F. Lillo, and R. N. Mantegna, J. Econ. Behav. Org. 75, 40 (2010).

27) F. Ren and W. X. Zhou, PLoS One 9, e97711 (2014), X. F. Jiang, T. T. Chen, and B. Zheng, Sci. Rep. 4, 5321 (2014), H. Chen, Y. Mai, and S.-P. Li, Physica A 414, 360 (2014), C.-X. Yang, X.-H. Zhu, Q. Li, Y.-H. Chen, and Q. Deng, Physica A 415, 1 (2014).

28) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. NunesAmaral, T. Guhr, and H. E. Stanley, Phys. Rev. E 65, 066126 (2002), J. Shen and B. Zheng, EPL (Europhys. Lett.) 86, 48005 (2009), L. Kullmann, J. Kert´esz, R. Mantegna, Physica A 287, 412 (2000), J. P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, Eur. Phys. J. B 30, 285 (2002)

29) J. P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Phys. Rev. E 68, 056110 (2003), J. P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, Physica A 324, 247 (2003)

30) G. Bonanno, F. Lillo, R.N. Mantegna, Quantitative Finance 1, 96 (2001), G. Bonanno, N. Vandewalle, R.N. Mantegna, Phys. Rev. E 62, R7615 (2000), L. Kullmann, J. Kert´esz, K. Kaski, Phys. Rev. E 66, 026125 (2002)

Chapter 2 Methodology

The first step will be to extract the available data from the various sources which is electronically stored in the form of Administrative data (from public bodies) or social media data such as Facebook or twitter or Private sector data such as company financial records, customer databases, service delivery records, Internet search activity etc. The work will involve the random matrix theory which laid its foundation nearly 50 years back [Wigner 1967], since then it has been successfully applied to various fields including disordered systems, finance, biological systems, communication etc. which will be used here to isolate the non-random information from the statistical and phylogenetic noise that may have arisen due to the finite size of the data or over sampling. The processed data will be used to create a network of the financial commodities or social interactions where we will use the concepts of network theory which will provide a framework for analyzing these systems from the perspective of complex systems. In the network or graph representation of the financial system, the nodes will represent the financial assets whereas the links (edges) represents the interaction between the assets. The different graph theoretic techniques will be used to extract information about crucial points, conserved patterns, sectors, motifs present in the system and will be further used to construct the structural and functional properties of the system. We are planning to explore the properties dealing with the collective behavior of the system and relation with randomness, long-range interaction among agents, tolerance to noise, evolution and adaptability of system along with the system topological and hierarchical structure. The topological and hierarchical structure will be studied using the minimum spanning tree approach. This hierarchical tree provides information useful to investigate the number and nature of factors associated with the system taxonomy. We will also try to look for changes in the hierarchical structure of the system over time.

For the financial analysis, the data set comprises of

1) the daily foreign exchange rates for 21 different currencies from 2000 to 2018 for a period of 18 years expressed in term of a base currency which in the current case is United States Dollar (USD). The currency is denoted according to the ISO 4217 standards using three letter code. The list of the countries and their currencies used for the analysis is shown in Table.
2) the daily adjusted closing stock price of the 31 financial market across the world comprising the European market, Asian market and American market etc. The filtering is done by removing the date for which more than 50% of the market are closed on that day.
3) the daily adjusted closing stock price of more than 100 stocks from the Thai stock market that comprises the SET index.

Table 1 List of 31 financial indices symbol (SYM) used for the analysis

SYM	Country	SYM	Country	SYM	Country
^MERV	Argentina	^BVSP	Brazil	^BSESN	India
^KLSE	Malaysia	^MXX	Mexico	^KS11	South Korea
^AORD	Australia	^ATX	Austria	^FCHI	France
^HSI	Hongkong	^TA100	Israel	^N225	japan
^SSMI	Switzerland	^FTSE	UK	^GSPC	USA
^BUX	Hungary	^NZ50	New Zealand	^KSE	Pakistan
^IBEX	Spain	^CSE	Sri Lanka	^AEX	Netherland
^GSPTSE	Canada	^STOXX50E	Euro Zone	GD.AT	Greece
^TWII	Taiwan	^JKSE	Indonesia	^GDAXI	Germany
^STI	Singapore	^SSEC	China	RTS.RS	Russia
^BFX	Belgium				

Table 2 List of 21 currencies with the respective countries used for the analysis

Country	Currency	Country	Currency	Country	Currency
UK	GBP	Brazil	BRL	Switzerland	CHF
Mexico	MXN	China	CNY	Taiwan	TWD
Norway	NOK	Thailand	THB	Canada	CAD
Sweden	SEK	Hong Kong	HKD	Australia	AUD
Denmark	DKK	South Africa	ZAR	Euro	EUR
Japan	JPY	Singapore	SGD	New Zealand	NZD
India	INR	Sri Lanka	LKR	South Korea	KRW

The Pearson correlation coefficients are used to estimate the correlation between different exchange rates, global indices or stocks. The study of the correlation between different agents in a system is very crucial for risk management, stability analysis, portfolio management, etc. The correlations are often plagued with the randomness and noise due to multiple factors such as finite size effects, historical noise, statistical noise. We use random matrix theory to filter the randomness from the system. The dynamics of the correlation with time is studied by rolling window method. We first investigate the statistics of the correlation matrix for each year along with the probability density function (PDF) of the independent elements of the correlation matrix. We performed a detailed spectral analysis on the correlation matrix. The eigenvalue distribution is compared and studied with the random matrix results. To evaluate the information content of each eigenvector we use the eigenvector entropy. The Shannon entropy is also a measure of the eigenvector localization, low entropy eigenvectors are highly localized, where only a few components have significant contribution other components have a negligible contribution. At last a detailed network analysis is performed by creating a threshold network and then a minimum spanning tree. Different network properties such as degree, degree distribution, clustering coefficient, centrality measures etc are calculated and analyzed.

In the context of the protein sequences, we propose an approach to detect functional sites and motifs which consists of the spectral and network analysis of the correlation generated by the physiochemical properties of amino acids, integrated with the conservation measure based on the physiochemical properties. Each sequence is represented by a vector based on a physiochemical property. Comparing the sequence vectors gives the similarities and dissimilarities between the protein sequences which depends on the physiochemical property under consideration i.e. two sequences may be very similar for a physiochemical property but diverge significantly for other properties. We use information theoretic measures followed by the techniques from the random matrix theory (RMT) as a null hypothesis to to isolate non-random information from statistical and phylogenetic noise. Spectral properties are calculated and studied in detail. A network is created for each property and its properties are studied which gives the details of how positions are interacting with each other.

Chapter 3 Research Result

During the course of this project, we have worked towards the understanding of structure, function, and evolution of the complex systems including biological financial and social systems. In the project, we devise computational tools to understand these systems based on the data that they generate.

The first system comprises of the daily adjusted close price of the 31 global financial indices from the European, Asian and American markets downloaded from yahoo finance and the second system consists of the daily foreign exchange (FX) rates for 21 different currencies from the Federal Reserve Bank of St. Louis. In this work, we use the statistical, spectral and network methods for the financial time series analysis to study the interaction and the strength of the important commodities in the network. The main focus of the work is on spectral analysis of the eigenvalues and eigenvector of the cross-correlations between the different financial commodities to identify the most influential commodity and interaction within the system. The dynamics of the system is studied on a yearly basis from 2006 to 2015. The correlation and the spectral properties is estimated for each window we find that the probability density function (PDF) is estimated which shows that for global indices the mass of PDF is concentrated on the positive side whereas for the FX rates the weight is evenly divided among the complete range. For each window, the eigenvalues and eigenvectors are estimated, which differed from the random system. The distribution of eigenvalues can be classified into two categories namely the calm period and the period of the financial crisis. An information theoretic measure on the eigenvector is used to detect the sector or highly interacting pairs of financial commodities in a system that have a functional role. We find that the eigenvectors have important physical meaning, for example, the second largest eigenvector clusters the commodities based on their geographical location. A threshold method is used to analyze the structure and evolution of the interaction between the constituents element of the system which gives a clear picture of the interaction within the system. The evolution of the network properties is studied for each window, we find that the network is denser at the time of crisis hence there is an increase in the correlation between the commodities during the crisis.

In the **second part**, we are trying to detect functional sites and motifs which consists of the spectral and network analysis of the correlation generated by the physiochemical properties of amino acids, integrated with the conservation measure based on the physiochemical properties. In this context, we propose an approach to detect functional sites and motifs which consists of the spectral and network analysis of the correlation generated by the physiochemical properties of amino acids, integrated with the conservation measure based on the physiochemical properties. he proposed method converts the protein sequences into a vector in three-dimensional physiochemical property space, from which residue positions controlling protein function can be extracted. A new graphical method to represent protein sequences using physiochemical properties is devised which gives a fast, easy and informative way of comparing the evolutionary distances between protein sequences. In this space, the evolutionary conservation of a position is specified by the conservation in the physiochemical property values. We observe that the favorable substitutions at a position are the ones which preserve the property crucial for the functioning of that site which is quantified by the entropy and Kullback-Leibler divergence. In this work gives an indication that certain physiochemical properties play a crucial role in determining the function and structure and our analysis can not only find the important property but can test the importance of property on each site. The method can also predict the interaction between residues and cluster them based on the eigenvectors into sectors which have a well defined structural or functional importance. The functional and evolutionary interactions between residue positions within the protein family are visualized using the graph theoretical techniques. The functional sites found for the beta-lactamase family using the proposed analysis have experimental verification which can be used as targets to change the functional properties such as catalytic or binding affinities to control, deactivate or activate the enzymatic action. The analysis also identifies certain positions which have remarkable interactions with other crucial positions but have not been studied experimentally. We suggest that during the course of evolution these sites may have become significant for the enzyme family and are therefore important for future experimental investigations. We are also in the process of understanding the working of other enzymes. The same procedure is applied to different protein families including serine protease, HSP70, G protein, and HTH_1 families. The crucial positions are extracted and identified.

We are also involved in the study of the structure, evolution, and dynamics of the **Thai stock market (SET index)**. In this problem, we study more than 100 stocks from the Thai stock market to understand how the Thai stock market evolves from 2000 to 2018. The analysis is mainly based on, spectral analysis and network theory. A detailed network analysis is performed by creating a threshold network and then a minimum spanning tree. Different network properties such as degree, degree distribution, clustering coefficient, centrality measures etc are calculated and analyzed. From our analysis, we can cluster stock into sectors such as agricultural, healthcare, banking etc. The detailed analysis reveals how the properties changes when the market is undergoing crash or in the growth phase

Chapter 4 Conclusion

We have made a significant progress in the project titled "Physics of complex systems: Economic, Finance and social systems". We have analyzed many complex systems from various fields such as Proteins sequences, financial systems, stock markets, world trade network etc. The data for all the systems are collected from various sources. A automated python scripts are made to mine the massive datasets. Global Stock indices data, Thai stock market data, and global currency data are collected from yahoo finance, world bank, world trade organization etc. Data is formatted and filtered to remove the noise and randomness. We have made use of the information theory and random matrix theory for the filtering. To get the information encoded by the data sets we devise and formulate the methods based on the information theory, statistical analysis, cross-correlational analysis, spectral analysis of the correlation matrix and the network theory. The methods are successfully applied to the data.

In this work, we try to develop the method based on correlation matrix to analyze the financial time series. We use different independent systems namely global financial indices, foreign exchange rates, Thai SET index for a span of 10 years from 2006 to 2015 to study using the spectral and network methods. We create the correlation matrix for yeah year from the time series data and compared its dynamics and evolution. The properties of yearwise correlation matrix are also compared with the random shuffled matrix to formulate a null model. From the statistics on the correlation matrix, we find that the financial markets have mainly positive correlation with a very less negative correlation whereas the currency market shows both high positive correlation as well as negative correlations (anti correlations). The eigenvalues and eigenvectors of the yearwise correlation matrix are studied and compared with the random matrix results. The probability density function (PDF) of correlation matrix are different for the crisis and the calm period. The eigenvalue and eigenvector analysis gives insight about the patterns in the data. Such as the second largest eigenvalues of the global financial indices is linked to the geographical location of countries and the structure of the second largest eigenvector components remains the same for all windows only with a change in sign. The RMT analysis with the eigenvector entropy shows that most of the information is located in the lower side of the eigen value spectra. These eigen modes are highly localized and gives the financial ties that changes during the 2008 crash. The study of the variation of the number of eigenvalues outside the RMT bounds with time shows a significant increase in the number of eigenvalues on the lower side of the spectra during the crisis period. At last, we created the threshold based network form the correlation matrix. The evolution of some of the network properties with time is studied which shows that the network is more connected at the time of crisis hence indicating an increase in the interaction between the commodities during the crisis. We also observe that for the world currency network the components of second smallest eigenvector, changes significantly during the crisis and the calm period. The dominating currencies in the calm period suddenly lose their contribution at the start and during the crisis period and a new group of interaction between currencies emerges at the start and during the crisis.

Also we devise a method to represent the protein sequence as multidimensional time series based on the physiochemical properties is devised. The method we devised is fast and reliable to align sequences and find the evolutionary divergence between sequences taking into account for the physiochemical properties of amino acids. The entropic measure such as Shannon entropy or Kullback-Leibler divergence shows that at the functional positions certain physiochemical properties are more conserved than the type of amino acid. The positions with low physio-chemical based entropy and the KL divergence are the sites that have important structural, evolutionary and functional role. Correlation matrix and its spectral properties are used to cluster positions into important groups termed as "sectors", that have functional and structural significance as active site, catalytic site or ligand binding site. Finally the network analysis reveals how different positions are interacting with each other forming a network of interacting positions and how these interaction give rise to the structural and functional properties.

We analyze different protein familes but the main work is for the beta-lactamase enzyme family which are the enzymes responsible for the drug resistance in the bacteria against antibiotics. In this work, we devised tools and methods to understand the working mechanism of enzyme family beta-lactamase responsible for the drug resistance in the bacteria against antibiotics. The work involves the use of various physiochemical properties to study the evolution, interactions, and correlations between the amino acids in a protein family. The network analysis is used to identify the highly interacting positions and important interactions which are crucial for the working of the protein family. The research can identify the important interactions and positions within the protein that can be used to activate or deactivate its enzymatic action.

Appendix

The project also leads to important academic collaborations one of which is with **Prof. Nivedita Deo, Professor, Department of Physics and Astrophysics, University of Delhi, India** on the field of complex systems (financial and biological). I have been on a research visit to Department of Physics and Astrophysics, University of Delhi, India during the April - May 2019. The visit was fruitful and results in the work on two different fields firstly on econophysics and biophysics.

Published Peer reviewed Articles:

1. **Evolution and Dynamics of the Currency Network,** Pradeep Bhadola, N Deo, New Perspectives and Challenges in Econophysics and Sociophysics, 133-145, (2019). **Springer Nature Switzerland AG 2019.**
2. **Spectral and Network Method in Financial Time Series Analysis: A Study on Stock and Currency Market,** Pradeep Bhadola, N Deo, Network Theory and Agent-Based Modeling in Economics and Finance, 331-351, (2019), **Springer Nature Singapore Pte Ltd.**
3. **Physiochemical property based approach for protein sequence analysis,** Pradeep Bhadola, N Deo, Journal of Physics: Conference Series 1144 (1), 012083, December 2018

To be Published (In progress)

1. **Structure, evolution and dynamics of Thai Stock market.**
   Pradeep Bhadola and Sompoach Saichaemchan
2. **Spectral and network analysis of protein families,** Pradeep Bhadola and N. Deo,
3. **Topology and dynamics of world trade network,** Pradeep Bhadola

Spectral and Network Method in Financial Time Series Analysis: A Study on Stock and Currency Market

Pradeep Bhadola and Nivedita Deo

Abstract In this work, we summarize some of the recent statistical, spectral, and network methods for the financial time series analysis. The main focus is first on spectral analysis of the eigenvalues and eigenvector of the cross-correlations between the different financial commodities, and second, the threshold networks derived from the cross-correlation matrix. We used an information theoretic measures called eigenvector localization defined as the "eigenvector entropy" to derive the community structure and interactions between the financial commodities in a system. Lastly, network analysis is performed on the system which shows how the interaction within the system changes with time. In this article, we present and compare the results of two different systems: (1) global financial indices of 31 countries and (2) foreign exchange rates of 21 different currencies using a rolling window method.

1 Introduction

Financial systems have been considered as highly complex evolving systems having many unknown agents and interactions that govern its dynamics. Understanding this complex nature of financial markets is emerged as a challenging task and in recent decades has been attracting considerable interest around the globe (Zhu et al. 2017; Kenett et al. 2015; Junior and Franca 2012; Wang et al. 2018; Ren and Zhou 2014). The complex interaction among the constituents of the system is sometimes highly stochastic in nature which is usually quantified by calculating the cross-correlations among the agents in the systems. In the current digital age, the financial data is growing at an enormous rate which is electronically accessible, thus making it possible to use the enormous financial data to predict the behavior of the financial

P. Bhadola
The Institute for Fundamental Study, Naresuan University, Phitsanulok, Thailand
e-mail: bhadola.pradeep@gmail.com

N. Deo (✉)
Department of Physics and Astrophysics, University of Delhi, Delhi 110007, India
e-mail: ndeo@physics.du.ac.in; ndeo007@gmail.com

A. S. Chakrabarti et al. (eds.), *Network Theory and Agent-Based Modeling in Economics and Finance*, https://doi.org/10.1007/978-981-13-8319-9_17

system. Most of the research in understanding the behavior of the financial time series over the last two decades is based mainly on the calculation of the cross-correlation between the price change of the financial commodities (stock, currency, goods, etc.) (Mantegna and Stanley 1999; Junior and Franca 2012; Wang et al. 2018; Ren and Zhou 2014). The is a large quantity of work on cross-correlation that includes random matrix theory (RMT) (Kumar and Deo 2012; Ren and Zhou 2014; Junior and Franca 2012), principal component analysis (PCA) (Billio et al. 2012), or the construction of the network (Nobi et al. 2014; Birch et al. 2016) not only in the field of econophysics but also in the field of biology (Bhadola and Deo 2016; Cocco et al. 2013), wireless communication (Tulino et al. 2004), medical science (Batushansky et al. 2016).

In the current article, to extract the information from the cross-correlation matrix, we first present a detailed spectral analysis of the eigenvalues and eigenvectors of the cross-correlation matrix; we introduce an entropic measure on the eigenvectors to quantify the information content and localization of the eigenvectors. This measure is used to identify the clusters of highly interacting agents within the system. A threshold method is used to analyze the structure and evolution of the interaction between the constituents element of the system.

2 System and Data

The current analysis is performed on the two diverse systems which are given below:

1. The first system comprises the daily adjusted close price of the 31 global financial indices from the European, Asian, and American markets downloaded from yahoo finance (2019) given in Table 1.
2. The second system consists of the daily FX rates for 21 different currencies from the Federal Reserve Bank of St. Louis (2019) and is shown in Table 2 denoted according to the ISO 4217 standards three-letter code. All the rates are expressed in terms of USD which is the base currency.

These two systems, namely, global financial indices (Bhadola and Deo 2017; Kumar and Deo 2012; Nobi et al. 2014) and foreign exchange rates (Bhadola and Deo 2019), are well studied in the literature. The common factor between the two systems is that in both the systems, the constituent elements (global indices and currencies) are represented as a time series. The data for both the systems are taken from the beginning of 2006 to the end of 2015 which spans a period of 10 years.

After a series of preprocessing steps on the data, the filtering is carried out to eliminate any numerical artifacts. The analysis is done on the daily logarithmic returns $R_i(t)$, which for a commodity i is calculated as

$$R_i(t) = ln(S_i(t + \Delta t)) - ln(S_i(t)) \quad (1)$$

Table 1 List of 31 financial indices symbol (SYM) used for the analysis

SYM	Country	SYM	Country	SYM	Country
^MERV	Argentina	^BVSP	Brazil	^BSESN	India
^KLSE	Malaysia	^MXX	Mexico	^KS11	South Korea
^AORD	Australia	^ATX	Austria	^FCHI	France
^HSI	Hongkong	^TA100	Israel	^N225	japan
^SSMI	Switzerland	^FTSE	UK	^GSPC	USA
^BUX	Hungary	^NZ50	New Zealand	^KSE	Pakistan
^IBEX	Spain	^CSE	Sri Lanka	^AEX	Netherland
^GSPTSE	Canada	^STOXX50E	Euro Zone	GD.AT	Greece
^TWII	Taiwan	^JKSE	Indonesia	^GDAXI	Germany
^STI	Singapore	^SSEC	China	RTS.RS	Russia
^BFX	Belgium				

Table 2 List of 21 currencies with the respective countries used for the analysis

Country	Currency	Country	Currency	Country	Currency
UK	GBP	Brazil	BRL	Switzerland	CHF
Mexico	MXN	China	CNY	Taiwan	TWD
Norway	NOK	Thailand	THB	Canada	CAD
Sweden	SEK	Hong Kong	HKD	Australia	AUD
Denmark	DKK	South Africa	ZAR	Euro	EUR
Japan	JPY	Singapore	SGD	New Zealand	NZD
India	INR	Sri Lanka	LKR	South Korea	KRW

where $S_i(t)$ is the price of commodity i (country index or currency) at time t and $\Delta t = 1$ day is the time lag. The normalized logarithmic returns are then given by

$$r_i(t) = \frac{R_i(t) - < R_i >}{\sigma_i} \tag{2}$$

where $< R_i >$ is the time average of the returns over the time period and σ_i is the standard deviation of $R_i(t)$ given by $\sigma_i =< R_i^2 > - < R_i >^2$.

3 Correlation Matrix

The study of the correlation between different commodities in a system is very crucial for risk management, stability analysis, portfolio management, etc. But the correlations are often plagued with the noise, and the quality of information depends on the ratio of the length of time series L and N number of commodities. This ratio should be greater than 1; otherwise, the resulting correlation matrix will be very noisy and is not good for analysis.

We use the Pearson cross-correlation coefficient which is calculated from the normalized log returns for any commodity pair i and j and defined as $C_{i,j} =< r_i(t)r_j(t) >$. The elements of the cross-correlation lie in the range $-1 \leq C_{i,j} \leq 1$, where $C_{i,j} = 0$ corresponds to no correlation between the commodity i and j, $C_{i,j} = 1$ is the perfect correlation, and $C_{i,j} = -1$ represents perfect anticorrelation. The correlation matrix for the period complete period 2006–2015 is calculated and shown in Fig. 1 for the 31 global financial indices and Fig. 2 for the foreign exchange rates which shows that the financial indices are positively correlated as compared to the foreign exchange rates. Foreign exchange rates show both high positive and negative correlation spanning the complete range of the possible correlations from $[-1, 1]$ whereas for the global financial indices, this range is $[-0.25, 1]$ whereas most of the weight (more than 90%) is concentrated on the positive side.

Next, we study the dynamics of both the systems on the yearly basis from 2006 to 2015 a total of 10 years. The correlation matrix is calculated for each year (window) and the properties are studied. First, we study the statistical properties of the cross-correlation for each window. First, we investigate the probability density function for each system (global indices and currencies) for the complete system and each window which is shown in Fig. 3. The PDF for both the systems global indices and foreign exchange rates is a nonsymmetrical distribution with a positive mean. The global financial indices for all window shows that the mass is concentrated on the positive side whereas for the foreign exchange rates the weight is evenly divided

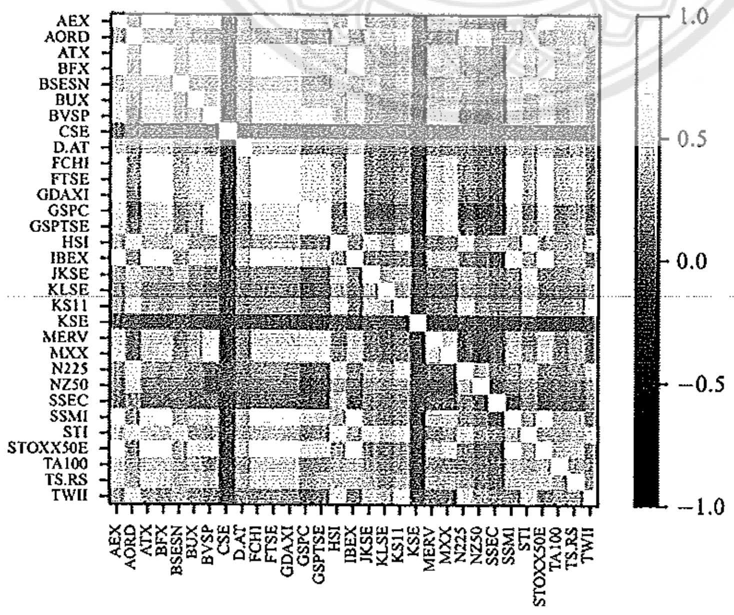


Fig. 1 Heatmap of C_{ij} for the 31 global financial indices from 2006 to 2015

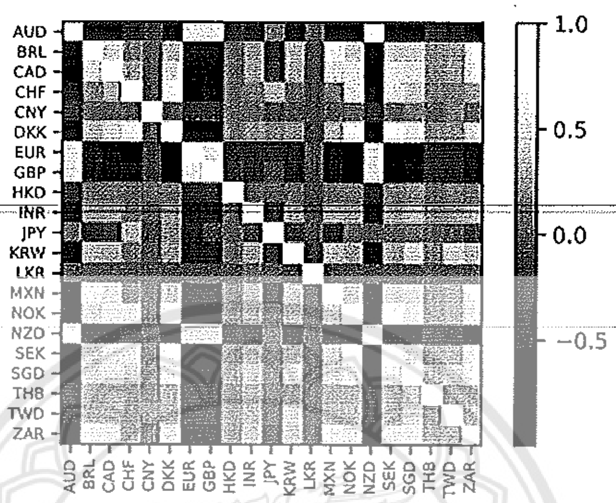


Fig. 2 Heatmap of C_{ij} for the foreign exchange rates from 2006 to 2015

among the complete range. The global indices show only high positive correlation with a very little anticorrelation whereas for the foreign exchange rates both high correlation and anticorrelation are possible.

Figure 4 shows the average of the off-diagonal elements of the cross-correlation matrix for global financial indices and foreign exchange rates. The average correlation between the global financial indices increases in the years 2007, 2008, and 2011. These three years have the highest average correlations and correspond to subprime mortgage crisis (2007), global financial crisis (2008), and European sovereign debt crisis (2011). For foreign exchange rate, $< C >$ shows a decrease average correlation in 2007 and 2008 but the highest average correlation among all the years considered. We also look into the average anticorrelation, i.e., negative off-diagonal elements of the correlation matrix which is shown in Fig. 5 for global financial indices (left) and foreign exchange rates (right). The Fig. 5 shows that anticorrelation are negligible for global financial markets and all the markets positively correlated. Whereas for the foreign exchange rates, there are significant anticorrelations which are lowest in 2011. For the complete period from 2006 to 2015, global financial markets are more correlated than the foreign exchange rates, whereas the foreign exchange rate how high degree of anticorrelation (10 times higher) than global financial markets.

4 Spectral Properties

The correlation structure can be studied by the eigenspectra of the correlation matrix C. For each year, the eigenvalues λ_i and eigenvectors v_i are determined from the corresponding correlation matrix. The eigenvalues are then arranged in ascending order such that $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_N$.

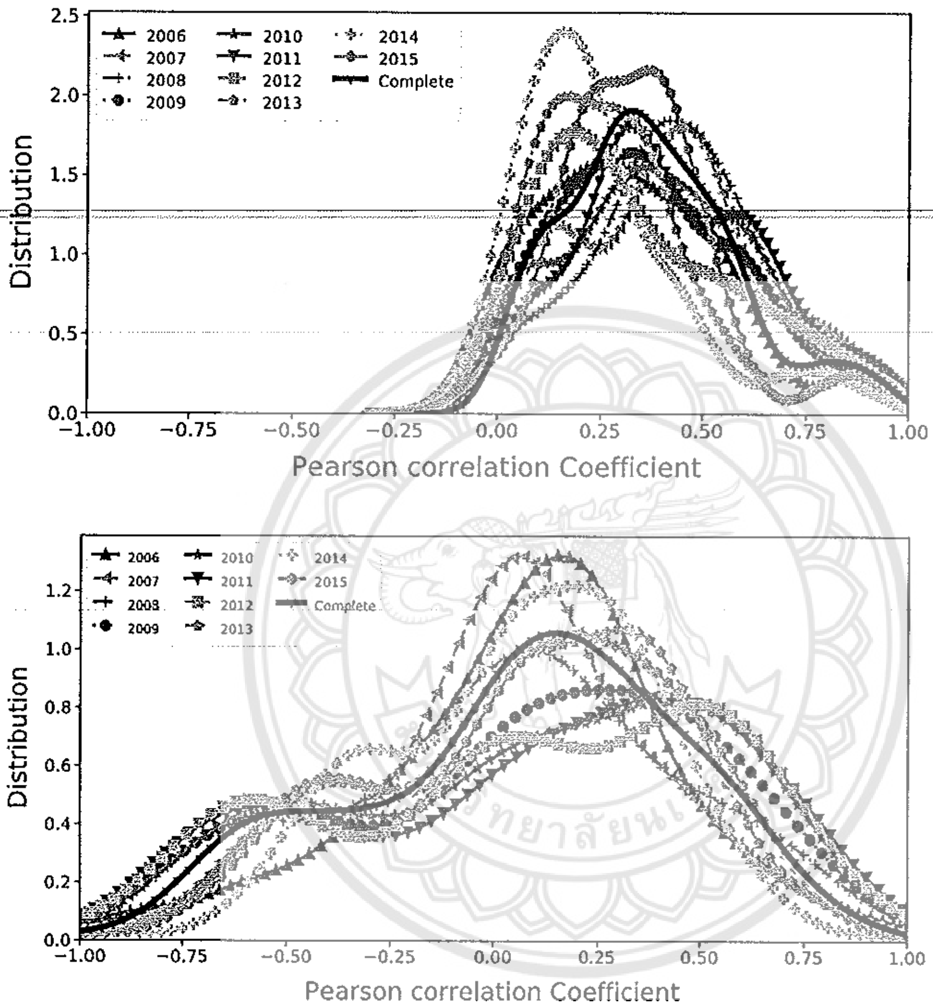


Fig. 3 First figure shows the PDF of C_{ij} for the global financial indices and second figure shows PDF of C_{ij} between the FX rates of different currencies for the complete interval from 2006 to 2015 and for the each window

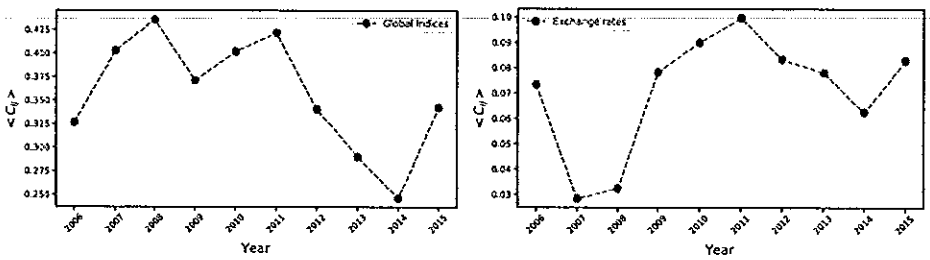


Fig. 4 Average off-diagonal elements of cross-correlation coefficients C_{ij} of the global financial indices (left) and foreign exchange rates (right)

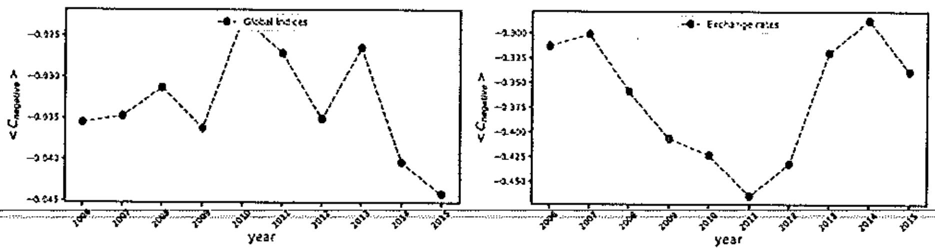


Fig. 5 Average negative off-diagonal elements of cross-correlation coefficients C_{ij} of the global financial indices (Left) and foreign exchange rates (right)

The cross-correlation matrix often contains random noise (Pandey et al. 2010) in addition to the actual information. We construct a null model by random shuffling of the data set for each system. Numerically, resultant correlation matrices from the shuffled data set are equivalent to the Wishart matrices. We use the analytical results of the Wishart matrices, which have well-defined spectral properties (Bowick and Brézin 1991) where the eigenvalue density is given by

$$P_W(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{3}$$

with $Q = \frac{W}{L} \geq 1$ and $\sigma = 1$ the standard deviation.

The abovementioned distribution is known as Marcenko–Pastur distribution which gives the theoretical bounds for the eigenvalues as

$$\lambda_{\pm} = \sigma^2\left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right) \tag{4}$$

For the current analysis, $W = 250$ is the size of the window and $N = 31$ for global financial indices and $N = 21$ for foreign exchange rates. The details of the bounds for both the system are given in Table 3.

Table 3 Details of the RMT bound and eigenvalue statistics for the two systems (global financial indices and foreign exchange rates)

System	W	N	Q	λ_+	λ_-	$\lambda < \lambda_-$	$\lambda_+ < \lambda$
Global indices	250	31	8.0645	1.8282	0.4197	15	2
FX rates	250	21	11.9047	1.6636	0.5043	11	2

4.1 Eigenvalue Distribution

In this work, we first compare the eigenvalue distribution of the correlation matrix corresponding to each year with the correlation matrix for the complete data set and then with the randomly shuffled data. As the financial markets are dynamical and change continuously with time, comparing the yearly distribution with the distribution of the complete period can give insights how the system at a small time scale differs from the overall behavior.

Figure 6 shows the complete eigenvalue distribution of the global financial indices matrices for each year from 2006 to 2015, the complete data set of 10 years and random shuffled distribution. We also analyze the behavior of the lower side of the eigenvalue spectrum for the global financial indices shown in Fig. 7. The distribution on both sides of the spectra differs significantly from random distribution as well as from the complete period. The differences between the complete data are higher for the crisis period (2007, 2008, 2011) where there is an increase in the magnitude of the largest eigenvalue as shown in Table 4. However, there is not much change in the number of eigenvalues outside the RMT bounds for each year which is nearly 2 or 3, but there is a significant increase in the number of eigenvalues outside the RMT bound on the lower side of eigenvalue spectra during the crisis period. For 2007, there are 19; 2008, we have 18; 2010, there are 18; and for 2011, a total of 19 eigenvalues are outside the lower RMT bound $\lambda < lambda_{-}$. There the number of eigenvalues of the lower side of the spectra is an indication of the crisis period.

Figure 8 shows the complete eigenvalue distribution and the lower part of eigenvalue distribution of the foreign exchange rates on a yearly basis from 2006 to 2015, with the complete analysis period of 10 years with random shuffled data. The FX

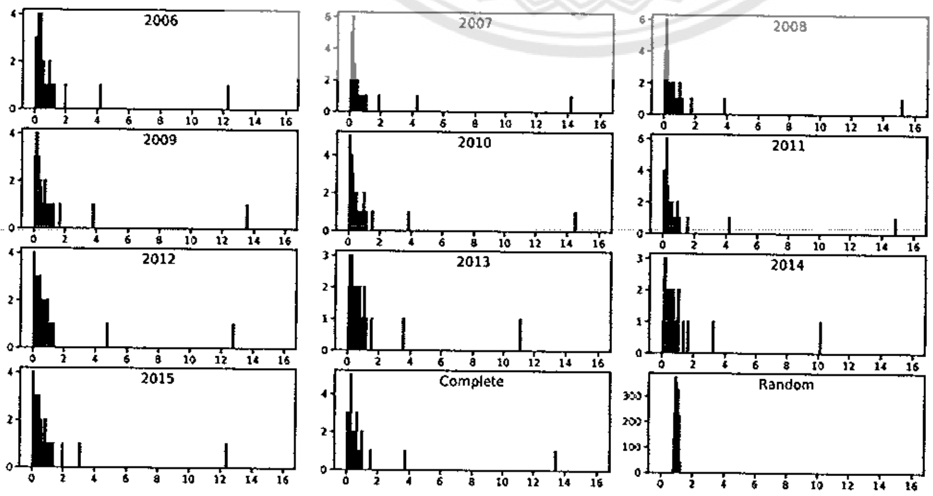


Fig. 6 Comparison of eigenvalue distribution of 31 global financial indices for each year from 2006 to 2015, for the complete period of 10 years and random shuffled system

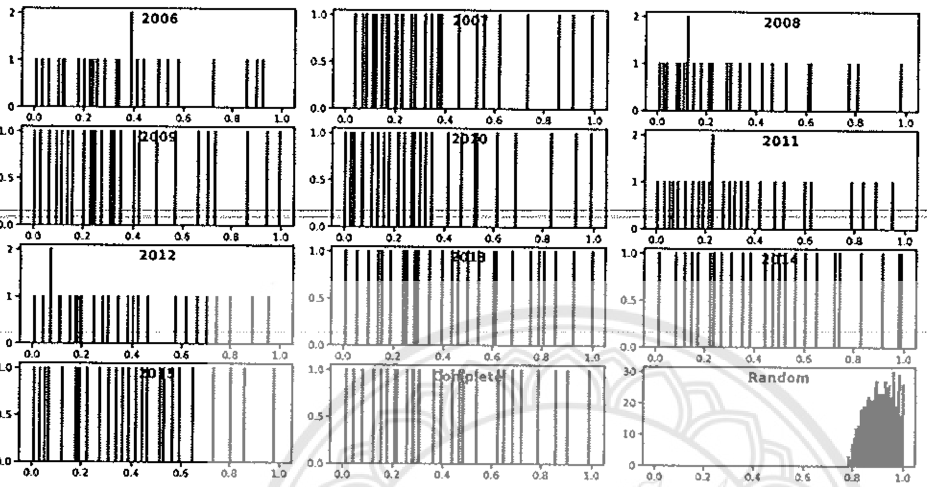


Fig. 7 Comparison of eigenvalue distribution on the lower side of the spectra of 31 global financial indices for each year from 2006 to 2015, for the complete period of 10 years and random shuffled system

Table 4 Comparison of the eigenvalue statistics of the global financial markets and foreign exchange rates

	Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Global	$\lambda < \lambda_-$	17	19	18	17	18	19	16	12	11	14
	$\lambda_+ < \lambda$	3	3	2	2	2	2	2	2	2	3
	λ_{max}	12.31	14.23	15.19	13.52	14.53	14.81	12.75	11.11	10.09	12.39
	λ_{min}	0.007	0.038	0.008	0.005	0.005	0.005	0.005	0.007	0.011	0.007
FX	$\lambda < \lambda_-$	10	9	12	12	13	11	12	10	9	10
	$\lambda_+ < \lambda$	3	2	2	2	2	1	1	2	2	2
	λ_{max}	7.77	7.48	8.91	9.82	10.06	10.55	10.44	8.03	7.09	8.08
	λ_{min}	0.0013	0.0019	0.0013	0.0007	0.0006	0.0004	0.0005	0.0019	0.0011	0.0009

rates also show a sudden increase in the largest eigenvalue during the crisis period 2007–2012, there is also an increase in the number of eigenvalues outside the lower bound. However, on the higher side of the spectra, there is not much change in the number of eigenvalues outside RMT bounds which is on average is nearly 2.

Analyzing both the systems indicates that most of the number of eigenvalues which are outside the RMT bound are located on the lower side of the spectrum. Thus, the essence of the interaction among agents is indicated mostly by the eigenvalues which are outside the lower RMT bound, most of the information is concentrated there (Figs. 7 and 9).

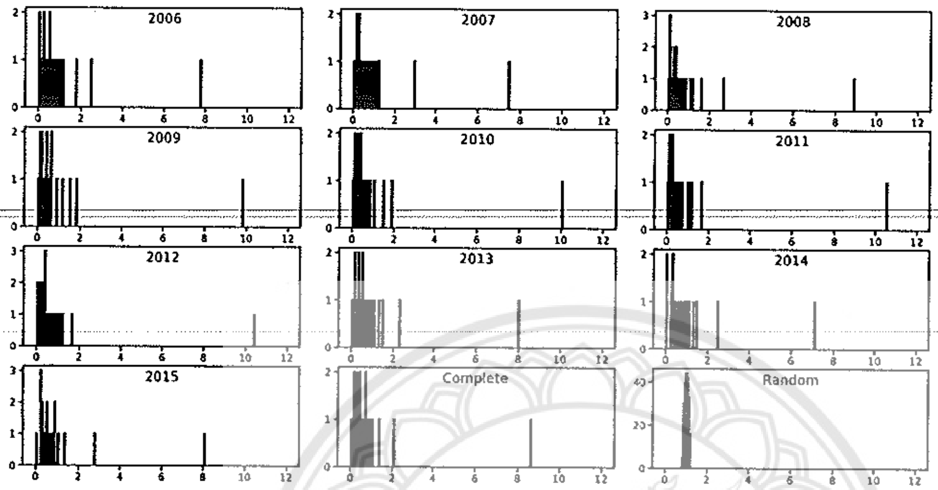


Fig. 8 Comparison of eigenvalue distribution of foreign exchange rates each year from 2006 to 2015, for the complete period of 10 years and random shuffled system

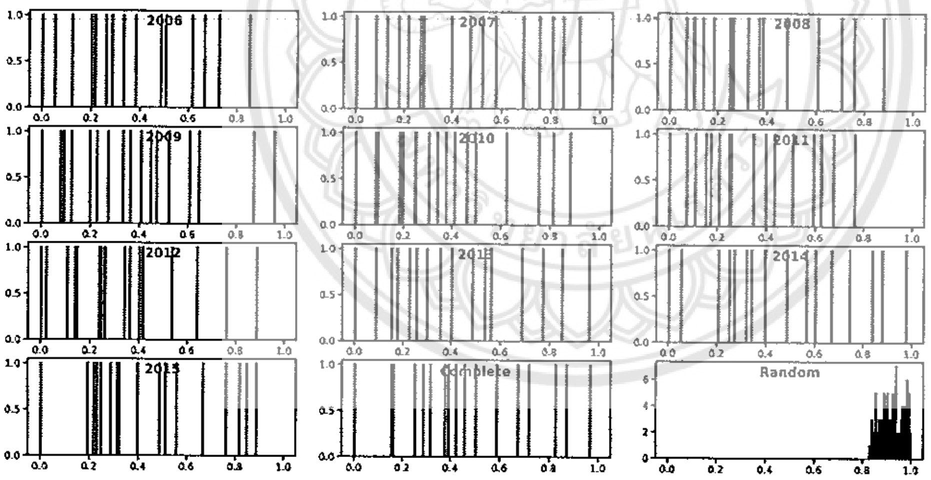


Fig. 9 Comparison of eigenvalue distribution on the lower side of the spectra of foreign exchange rates each year from 2006 to 2015, for the complete period of 10 years and random shuffled system

4.2 Localization of Eigenvector: Eigenvector Entropy

Localization of an eigenvector is estimated by the information content of the eigenvector. Higher the information content, more localized the eigenvector is. We use Shannon's entropy to assess the information contained in an eigenvector. Shannon's entropy of an eigenvector v_i is defined as

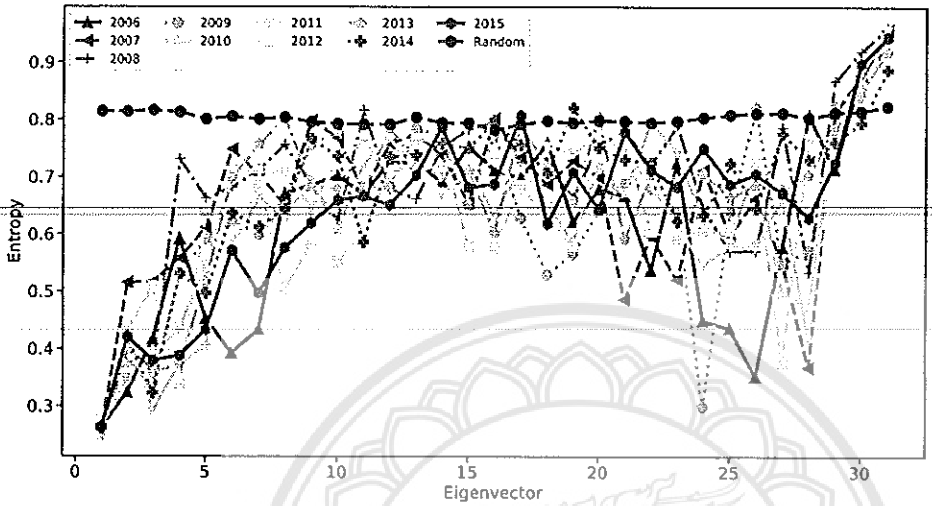


Fig. 10 Comparison of eigenvector entropy for the global financial market from 2006 to 2015

$$H_i = -\sum_{j=1}^{N} u_i(j) log_L(u_i(j)) \tag{5}$$

where N is the number of eigenvector components and $u_i(j) = (v_i(j))^2$ is the square of the jth component of the ith normalized vector v_i.

Eigenvector which has low entropy is highly localized with significant contribution from only some components. We find that in both the systems Figs. 10 and 11, we find that the eigenvector corresponding to the lower side of the spectrum has low entropy thus high localization as compared to the eigenvector on the higher side of spectra. The components of these eigenvectors have significant information about the different aspects of the system. The highly localized eigenvector gives the highly interacting group of commodities. It is established in many systems, for example, in protein families, the small eigenvectors of the correlation matrix have high information content can predict the functionally and structurally important groups of amino acids. The components of these eigenvector form highly interacting pairs that can be grouped into sectors or motif having an important functional role (Bhadola and Deo 2016). Also in finance, in Markowitz's theory of portfolio optimization (Markowitz 1971), low eigenvectors correspond to the least risky portfolios.

4.3 Eigenvector

As eigenvalues that diverge significantly from RMT predictions, the corresponding eigenvectors contains significant information. We perform a detailed analysis on the

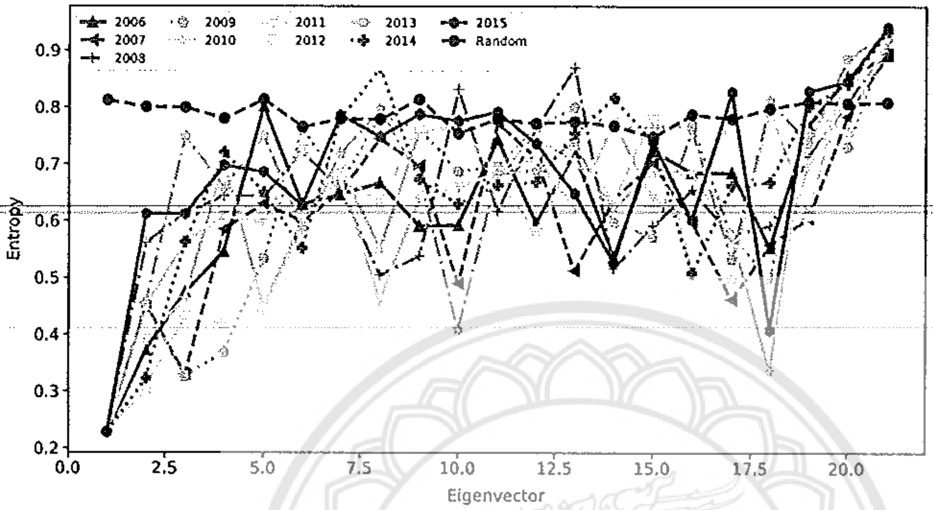


Fig. 11 Comparison of the eigenvector entropy for each year from 2006 to 2015 for the foreign exchange rate

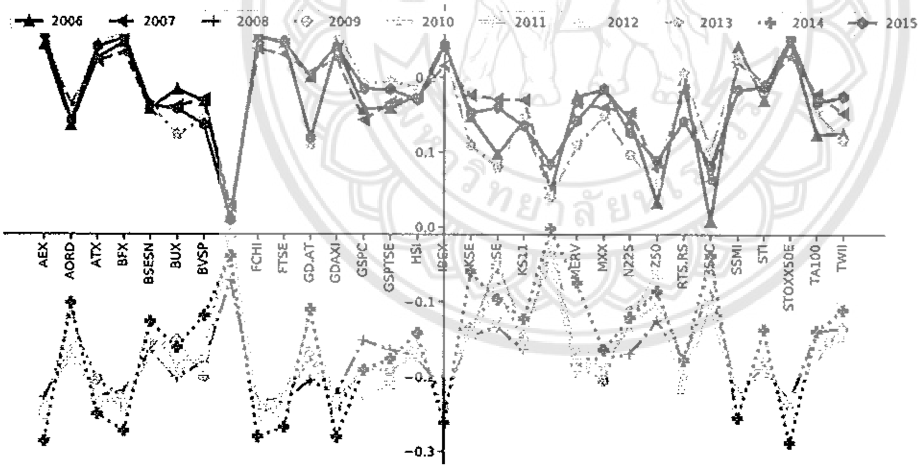


Fig. 12 Eigenvector components for the largest eigenvectors for the global financial indices for years 2006 to 2015

eigenvector components of the eigenvalues that are significantly far from the RMT bounds. The largest eigenvector shows the state of the system; in our analysis, it represents the global financial conditions.

Figures 12 and 13 show the components of the largest and the second largest eigenvector for each year from 2006 to 2015. As the largest eigenvector, gives the overall state of the system, it can be seen that the most of the Asia-Pacific countries contribute very little towards the state of the world financial condition. The global economic conditions are mainly decided by the performance of the European

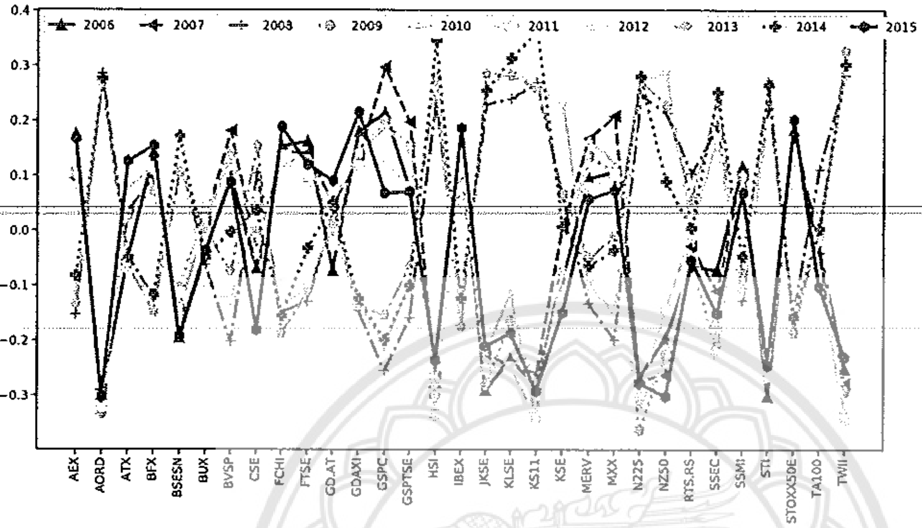


Fig. 13 Eigenvector components for the second largest eigenvectors for the global financial indices for years 2006 to 2015

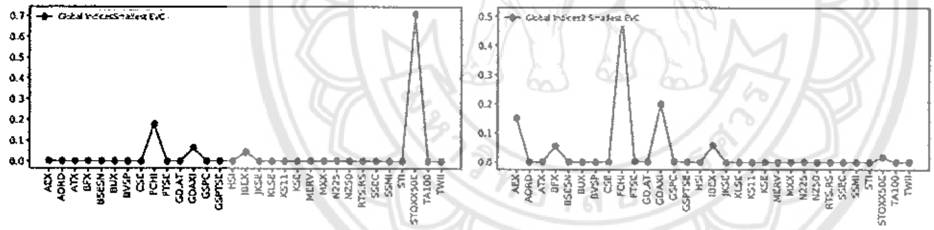


Fig. 14 Square of the eigenvector components for the smallest (left) and second smallest (right) eigenvectors for the global financial indices for years 2006 to 2015

markets. The second largest eigenvector divides the system into two categories: first those markets for which the components are positive, and second, the markets for which the components are negative. The components change signs in the groups with the change in time; the overall groups remain the same. Upon closer look, these two categories are based on the geographical location of the financial market. The two categories are first is the European and American market and the other group with opposite sign are the countries from Asia-Pacific region. The groups remain intact with time; all the countries change sign simultaneously. The groups are given by group 1 (European and American) that consists of 16 countries including Argentina, Austria, Belgium, Brazil, Canada, Euro Zone, France, Germany, Greece, Switzerland, Hungary, Spain, Netherland, Mexico, UK, and US). Group 2 consists of Australia, China, India, Indonesia, Israel, Japan, Malaysia, South Korea, Hong Kong, Singapore, New Zealand, Pakistan, Russia, Sri Lanka, and Taiwan from the Asia-Pacific region. Next, we investigate the eigenvector component corresponding to the smallest eigenvalues and the second smallest eigenvalues which are shown

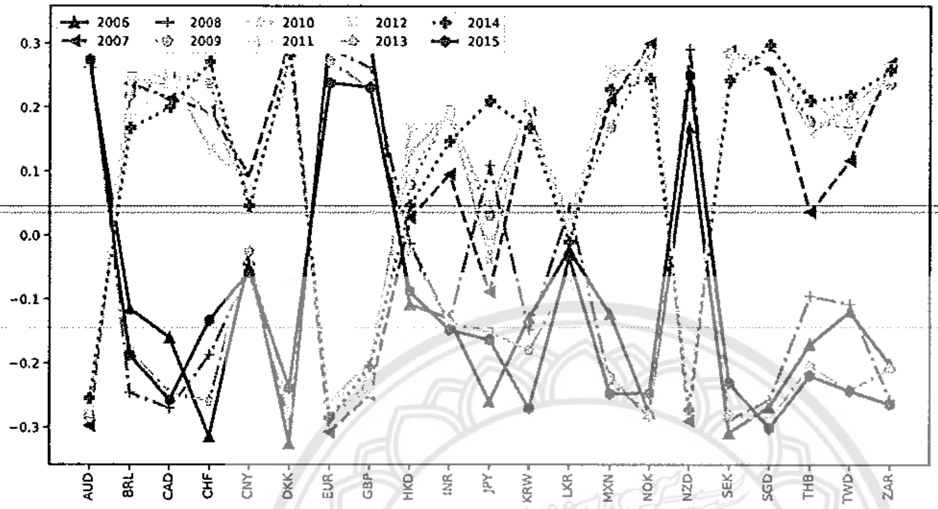


Fig. 15 Eigenvector components for the largest eigenvectors for the foreign exchange rates each year from 2006 to 2015

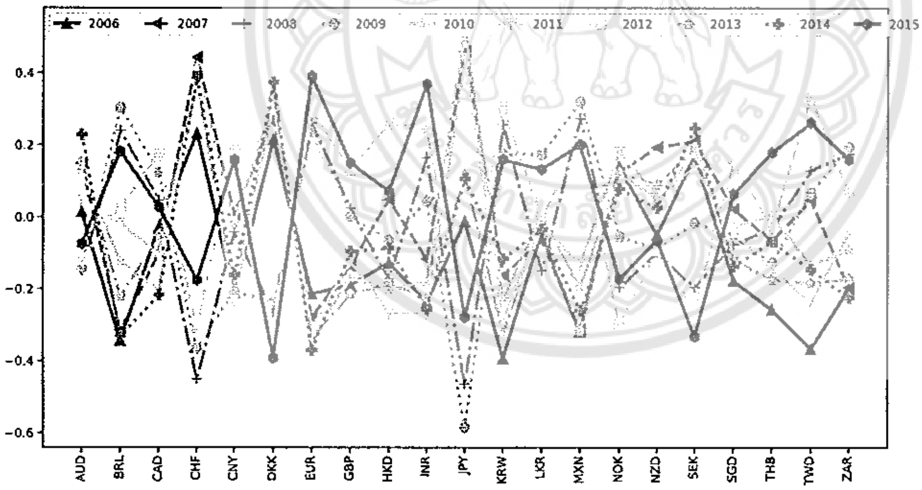


Fig. 16 Eigenvector components for the second largest eigenvectors for the foreign exchange rates each year from 2006 to 2015

in Fig. 14. These eigenvectors are highly localized with contribution from only a few components. These components form a highly interacting pair of markets that can be clustered together into sectors. For the smallest eigenvector Fig. 14, the main contribution is from France, Germany, Spain, and European Union; this forms a very highly interacting pair of markets, with high interdependence. European Union is most influential in the cluster. The second smallest eigenvector gives a cluster comprising Netherland, Belgium, France, Germany, Spain, and the European Union. In this cluster France is the most influential market.

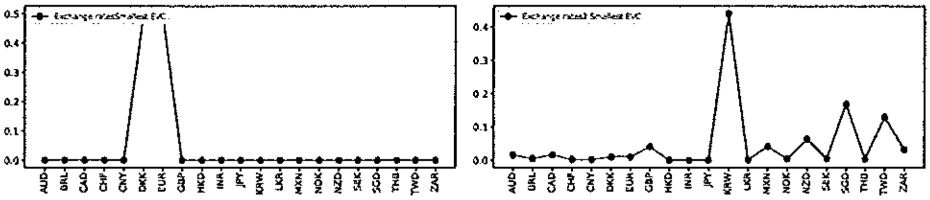


Fig. 17 Square of the eigenvector components for the smallest (left) and second smallest (right) eigenvectors for the foreign exchange rates each year from 2006 to 2015

Figures 15 and 16 show the variation of the largest and second largest eigenvectors for each currency with time. Figure 17 shows that the lowest eigenvectors are highly localized, and there are only a few contributing components. The largest eigenvector is not in one direction has both positive and negative component which is different from the largest eigenvector for the global financial markets where the components are completely positive or completely negative. The smallest eigenvector components show a very high interaction between the Denmark and Euro. The second smallest eigenvector however includes a large group of currencies including UK, South Korea, Mexico, New Zealand, Singapore, Taiwan, and South Africa.

5 Network Analysis

In this section, we use the threshold method to create a network from the correlation matrix. The correlation matrix can be converted into a graph based on the threshold θ. The resulting network $G^{\alpha}(N, E)$ with nodes (N) and edges $E^{\alpha}_{i,j}$ between two nodes i and j given by

$$E^{\alpha}_{i,j} = \begin{cases} 1 & if \quad i \neq j, C^{\alpha}_{i,j} \geq \theta \\ 0 & otherwise \end{cases} \tag{6}$$

where θ is the threshold value. The network is constructed for different threshold values ranging from [0, 1] and is shown in Fig. 18 for global financial indices and Fig. 19 for foreign exchange rates at different threshold values (0.1, 0.3, 0.5, 0.7).

Different threshold values generate different networks with the same set of nodes and different sets of edges. Using different properties results in a network with the same nodes but different links.

5.1 Network Properties

In this section, we discuss some of the basic properties of the correlation threshold network.

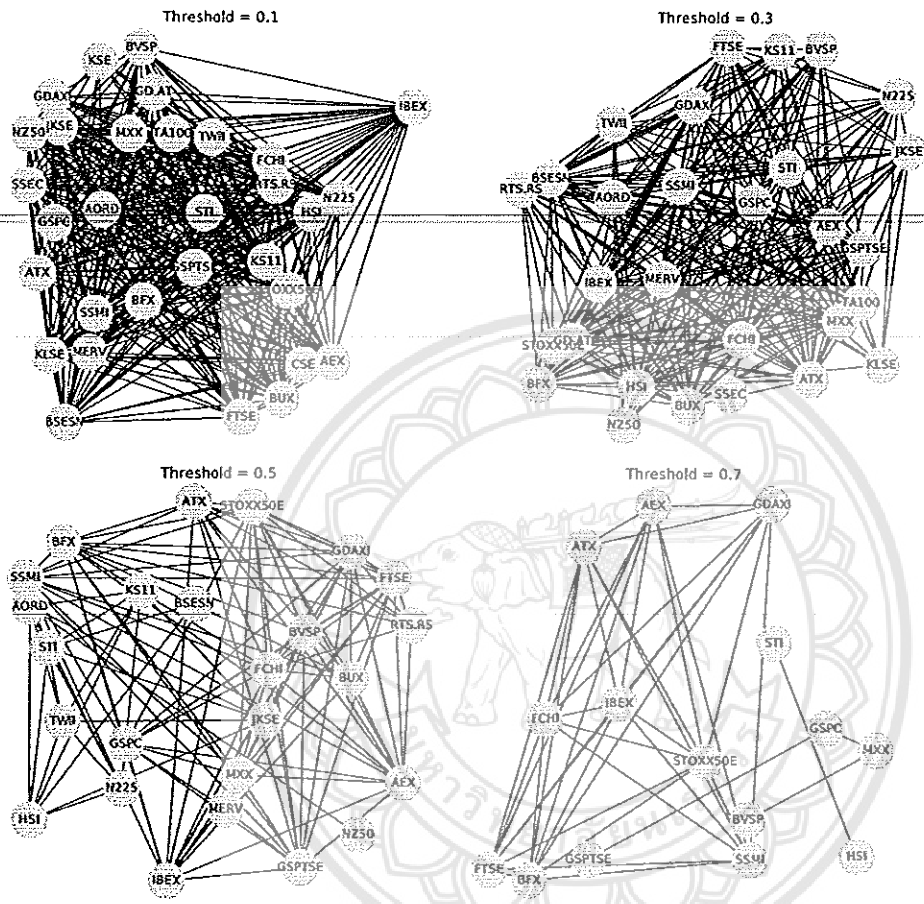


Fig. 18 Global financial network at different threshold values

5.1.1 Adjacency Matrix

Adjacency matrix or sometimes called as connection matrix is another representation of the connections in a network. The element of the adjacency matrix A_{ij} is 1 if there is a direct connection between node i and node j and A_{ij} is zero otherwise. In the case of correlation threshold network which is undirected network, the adjacency matrix is symmetric matrix which says that if there is a link from i to j, then there exists a reverse link j to i.

5.1.2 Network Density

Network density is defined as the ratio of the actual number of links in the networks to the maximum possible link in the network defined as $\rho = M/(N(N-1))$, where N is the number of nodes in the network and M is the connection in the network. We

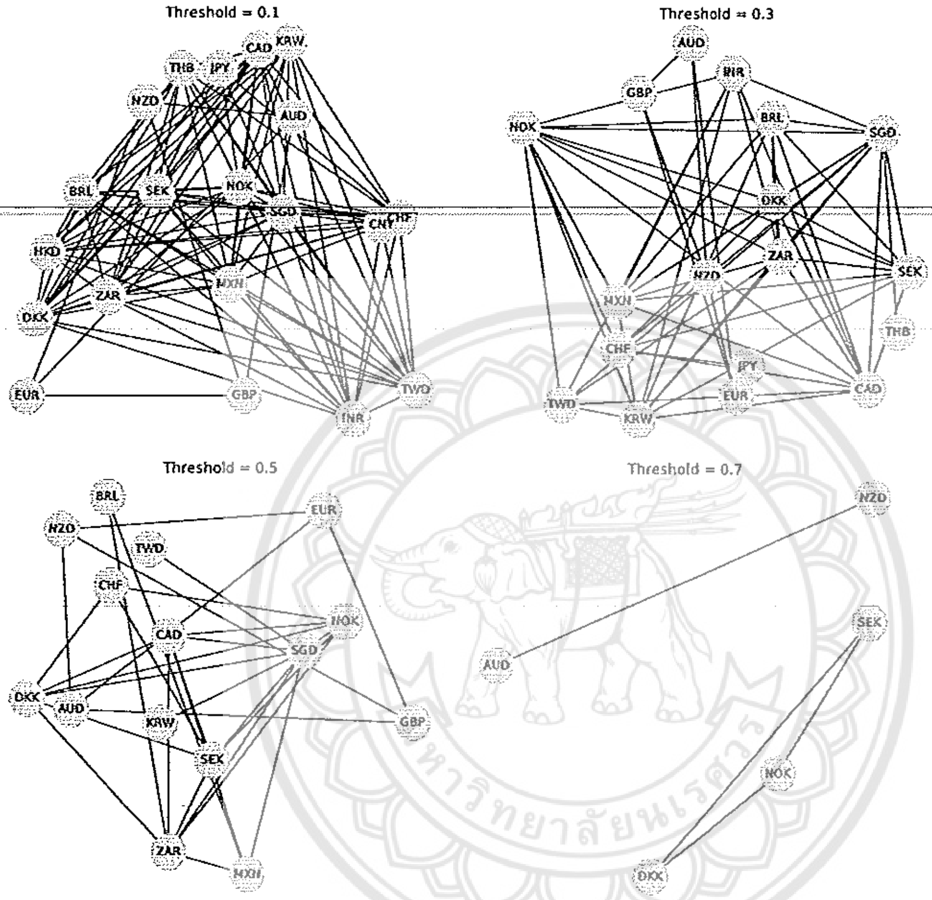


Fig. 19 Network of foreign exchange rates at different threshold values

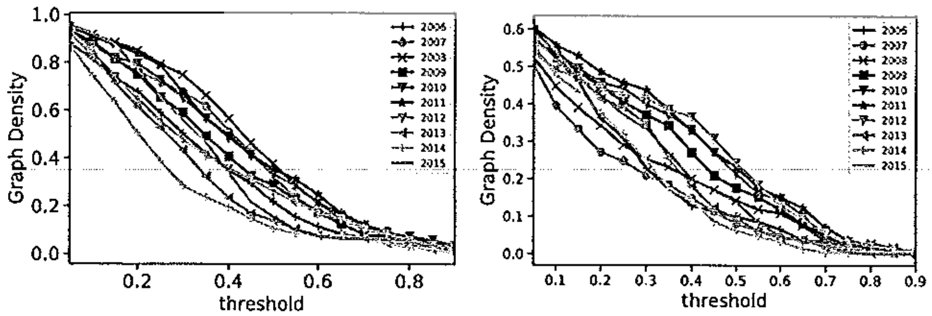


Fig. 20 Variation of network density with threshold for different windows for global financial markets (left) and .foreign exchange rates (right)

study change in network density at different threshold values with each year and i shown in Fig. 20 for global financial markets (left) and .foreign exchange rates (right). For the global financial markets, we observe an increase in the network density for the crisis period (subprime mortgage crisis (2007), global financial crisis (2008) and European sovereign debt crisis (2011)).

The foreign exchange network does not show any increase in network density for 2007 and 2008 and only shows a sudden increase in the density during the European sovereign debt crisis (2011).

5.2 *Average Degree*

Degree of a node gives the total number of interactions of that node with all other nodes. Mathematically the degree K_i of a node i is defined as

$$K_i = \sum_{j=1}^{N} A_{ij} \tag{7}$$

where N is the number of nodes in the network and A is the adjacency matrix of the network.

First, for each threshold, the degree of every node in the network is calculated for all windows. Then estimate the average degree of the network as $K_{avg} = \frac{1}{N}\sum_{i=1}^{N} K_i$, which gives the on an average what is the connectivity of each node. The variation of average degree with the threshold for both network is shown in Fig. 21. Figure 21 shows that there is an increase in the average degree in the network during the crisis period. This implies the increase in the number of connection within the network hence higher interaction.

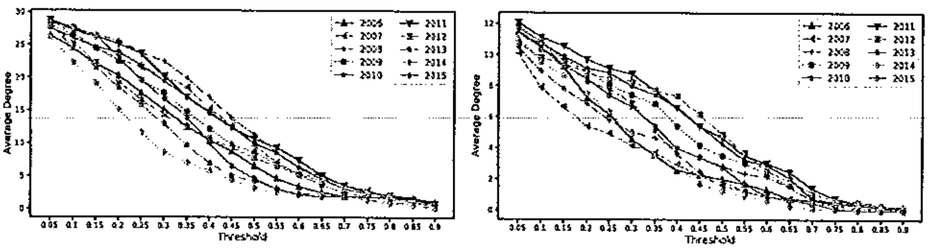


Fig. 21 Average degree with threshold for different windows for global financial markets (left) and foreign exchange rates (right)

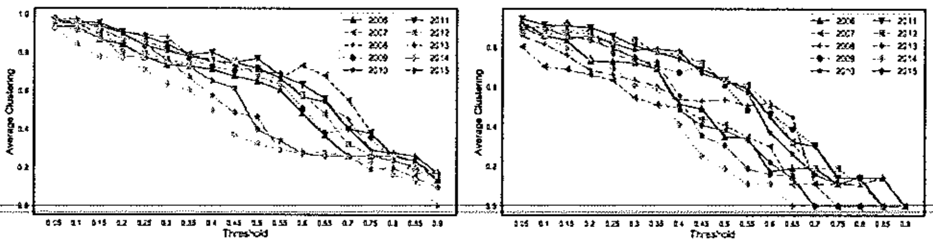


Fig. 22 Average clustering coefficient with threshold for different windows for global financial markets (left) and .foreign exchange rates (right)

6 Average Clustering Coefficient

Clustering coefficient is the measure of local connectivity in the network. It quantifies the interconnectivity between the neighbors of a node. The clustering coefficient of a node i is given as

$$C_i = \frac{2m_i}{n_i(n_i - 1)} \tag{8}$$

where m_i are the actual connection among the neighbors of i, and n_i is the number of neighbors of i. Clustering coefficient C_i is zero if $n_i \leq 2$. The average clustering coefficient is the average of clustering coefficient over all nodes in the network $C_{avg} = \frac{1}{N}\sum_{i=1}^{N} C_i$. We study the variation of the clustering coefficient with threshold and is shown in Fig. 22, which shows that there are significant changes in network with time. The clustering is high during the crisis for both the networks. The global financial indices 2007, 2008, and 2009 show high clustering. During the 2007 crisis, the foreign exchange network shows a higher value for the average clustering coefficient at large thresholds (0.5–0.8). Whereas at an intermediate threshold only sovereign debt crisis 2011 shows maximum clustering.

7 Conclusion

In this work, we discuss the spectral and network method based on correlation matrix to analyze the financial time series. Two independent systems, namely, global financial indices and foreign exchange rates, are studied in detail for a span of 10 years from 2006 to 2015. A correlation matrix is created from the time series data for each year from 2006 to 2015 and compared with the complete correlation matrix and the random shuffled matrix. Simple statistics on the correlation matrix shows that the financial markets are only positively correlated whereas the currency market shows both high positive correlation and negative anticorrelation. The eigenvalues and eigenvectors of the correlation matrices for each year and eigenvalue distribution are created. The eigenvalue distribution is compared with the random matrix results.

The distribution of eigenvalues reveals the distribution is different for the calm and the crisis period. An information theoretic measure on the eigenvector defined as eigenvector is calculated which quantifies the information content as well as the localization of the eigenvector. Eigenvector entropy shows that for both the systems, global financial indices and foreign exchange rates, the smallest eigenvectors are highly localized as compared to the eigenvalues on the upper side of the spectra. The eigenvector outside the random matrix bounds contains significant information for instance, for the global financial indices second largest eigenvector components can cluster the markets based on their geographical location. We study the variation of a number of eigenvalues out the RMT bounds with time indicate that during the crisis, there is a significant increase in the number of eigenvalues on the lower side of the spectra. The increase in the number of eigenvalues may be used as a measure to identify the crisis period. Analysis of these eigenvector gives the clusters of highly interacting pairs of commodities. At last, we created the threshold-based network form the correlation matrix. The evolution of some of the network properties with time is studied which shows that the network is more connected at the time of crisis hence indicating an increase in the interaction between the commodities during the crisis.

Acknowledgements ND thanks Department of Science and Technology (DST), India (SERB-DST No- EMR/2016/006536) for financial support, and PB acknowledges NU research grant, Naresuan University, Thailand for financial support.

References

Batushansky A, Toubiana D, Fait A (2016) Correlation-based network generation, visualization, and analysis as a powerful tool in biological studies: a case study in cancer cell metabolism. BioMed Res Int

Bhadola P, Deo N (2016) Targeting functional motifs of a protein family. Phys Rev E 94(4):042409

Bhadola P, Deo N (2017) Extreme eigenvector analysis of global financial correlation matrices. In: Econophysics and sociophysics: recent progress and future directions. Springer, pp 59–69

Bhadola P, Deo N (2019) Evolution and dynamics of the currency network. In:New perspectives and challenges in econophysics and sociophysics. Springer, pp 133–145

Billio M, Getmansky M, Lo AW, Pelizzon L (2012) Econometric measures of connectedness and systemic risk in the finance and insurance sectors. J Financ Econ 104(3):535–559

Birch J, Pantelous AA, Soramäki K (2016) Analysis of correlation based networks representing dax 30 stock price returns. Comput Econ 47(4):501–525

Cocco S, Monasson R, Weigt M (2013) From principal component to direct coupling analysis of coevolution in proteins: low-eigenvalue modes are needed for structure prediction. PLoS Comput Biol 9(8):e1003176

Federal reserve bank of St. Louis. https://fred.stlouisfed.org/

Junior LS, Franca IDP (2012) Correlation of financial markets in times of crisis. Phys A Stat Mech Appl 391(1–2):187–208

Kenett DY , Huang X, Vodenska I, Havlin S, Stanley HE (2015) Partial correlation analysis: applications for financial markets. Quant Financ 15(4):569–578

Kumar S, Deo N (2012) Correlation and network analysis of global financial indices. Phys Rev E 86(2):026101

Mantegna RN, Stanley HE (1999) Introduction to econophysics: correlations and complexity in finance. Cambridge University Press

Mark J Bowick and Édouard Brézin. Universal scaling of the tail of the density of eigenvalues in random matrix models. Phys Lett B 268(1):21–28,

Markowitz H (1971) Portfolio selection diversification of investments. By Harry M. Markowitz. YALE University Press

Nobi A, Lee S, Kim DH, Lee JW (2014) Correlation and network topologies in global and local stock indices. Phys Lett A 378(34):2482–2489

Pandey A et al (2010) Correlated wishart ensembles and chaotic time series. Phys Rev E 81(3):036202

Ren F, Zhou W-X (2014) Dynamic evolution of cross-correlations in the chinese stock market. PloS one 9(5):e97711

Tulino AM, Verdú S, et al (2004) Random matrix theory and wireless communications. Found Trends® Commun Inf Theory 1(1):1–182

Wang G-J, Xie C, Stanley HE (2018) Correlation structure and evolution of world stock markets: evidence from pearson and partial correlation-based networks. Comput Econ 51(3):607–635

Yahoo finance database. https://finance.yahoo.com/

Zhu B, Han D, Wang P, Zhanchi W, Zhang T, Wei Y-M (2017) Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Appl Energy 191:521–530

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Physiochemical property based approach for protein sequence analysis

IOP Conf. Series: Journal of Physics: Conf. Series 1144 (2018) 012083 doi:10.1088/1742-6596/1144/1/012083

Physiochemical property based approach for protein sequence analysis

Pradeep Bhadola[1] **and Nivedita Deo**[2]

[1]The Institute for Fundamental Study, Naresuan University, Phitsanulok, Thailand

[2] Department of Physics and Astrophysics, University of Delhi, India

E-mail: [1]bhadola_pradeep@yahoo.co.in, [2]ndeo@physics.du.ac.in

Abstract. We propose a physiochemical property based analysis, that represent a protein sequence as a multidimensional time-series from which residue positions controlling protein function can be extracted. We observe that the favorable substitutions at a position are the ones which preserve the property crucial for functioning of that site which is quantified by the entropy and Kullback-Leibler divergence. The entropic measures shows that during the evolutionary history of the protein family, it is the certain physiochemical properties that are conserved rather than the type of amino acids. For each physiochemical property, the correlation matrix between positions is calculated, and using an ensemble of Wishart matrices from the random matrix theory for the noise estimation and information filtering. The spectral properties of correlation matrices are calculated and compared with the analytical results for the Wishart matrices.

1. Introduction

The protein sequence space is growing at an exponential rate, it is extremely important to develop efficient and accurate techniques and tools to draw inferences from the sequence data. Even with existing theoretical models and computational techniques the function and structure encoded by a protein sequence is not completely established [1–4]. Most of the existing methods mainly treat protein sequences as strings of characters and anazlyze the sequence based on the frequencies of amino acids (will be referred as string method) by estimating entropy or mutual information (MI) [1, 2, 4], without looking at the chemical, physical or biological properties of amino acid. There are studies that show that the evolution and functional properties greatly depends on the physiochemical properties, by influencing the mutation rate or determining the native state [5]. To address the above problem we propose a physiochemical property based analysis that represent a protein sequence as a multidimensional time series, where each dimension represents a property value. The approach is mainly based on two conjectures: (1) structurally and functionally important sites show a high degree of conservation, if not in amino acid type, it will be reflected in the amino acid physiochemical properties, which is the outcome of the evolutionary constraint. (2) Functionality is a manifestation of structural interactions (direct or indirect mediated by series of intermediate amino acids) between positions, which is a result of the physical, chemical and biological coupling. The work involves information theoretic measure followed by the tools from the random matrix theory (RMT) which has been very successful in application to diverse areas of research including nuclear physics, disordered systems, information theory, string theory, statistical physics, biological sciences, finance, social systems [3, 6]. RMT is used as a null hypothesis to to isolate non-random information from statistical and phylogenetic noise.

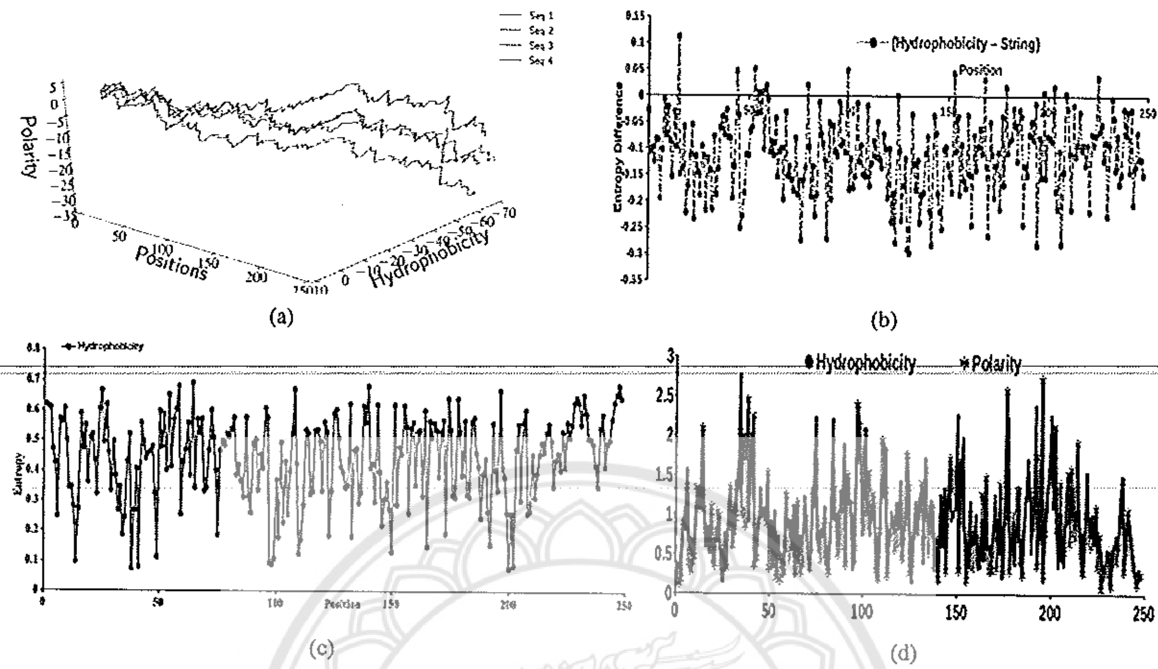


Figure 1. (a) Protein sequence vector based on physiochemical properties. (b)The difference in values of hydrophobicity and string based entropy for each positions in MSA. (c) Entropy of each position based on hydrophobicity. (d) The KL divergences based on hydrophobicity (blue) and polarity (red) for betalactamse family.

2. Data and Method

The system under study is class A β-lactamase enzyme family (Interpro entry IPR000871 consisting of 5447 proteins) which are the enzymes secreted by bacteria in response to β-lactam antibiotics like penicillin [7, 8]. These sequence are first filtered and then aligned in a multiple sequence alignment. Different physiochemical properties of amino acid are extracted from various sources and then rescaled between -1 to 1. Each amino acid in MSA is replaced by the physiochemical properties to get a 3 dimensional matrix $D^{\alpha}_{s,i}$, where s represents the sequence in MSA, i represents the amino acid position, and α represents the property (see [3] for details). We define a sequence vector as $U^{\alpha}_{s}(i) = \sum_{k=1}^{i} D^{\alpha}_{s,k}$ to see the trend in the change of values of a property across a sequence. The plot of sequence vector for 4 different protein sequence is using hydrophobicity and polarity is show in figure 1(a). The evolutionary divergence between two protein sequence can be estimated by comparing the two sequence vector using any metric distance measure. This evolutionary divergence score not only depends on the type of amino acid but also takes the physical, chemical and biological properties into account unlike the string based method which relies solely on the amino acid type.

3. Entropy

The diversity of amino acid at a position for a given family is measured by the Boltzman-Shannon entropy defined as

$$E(i) = -\sum_{j=1}^{n} P_j(i) log_n(P_j(i)) \tag{1}$$

The Boltzman-Shannon entropy quantifies the expected value of the information contained at position i with the probability distribution $P_j(i)$ as the frequency of occurrence ($j = 1 \cdots n$ with j labeling the type of amino acids and $n = 20$ the total number of distinct residues). For a fixed property α, the frequency distribution of physiochemical values is constructed using m equal sized bins in the interval [-1,1] and binning the physiochemical property value (column of data matrix D^{α}) into them ($m = 21$ for current analysis). The normalized probability distribution (P_j^{α}) at each column j is estimated from this frequency distribution (normalized). P_j^{α} gives the probability that a amino acid at a given position j has a value in the specified bin for a given property α. The entropy at position i considering property α with normalized frequency distribution $P_b^{\alpha}(i)$ is given by

$$E^{\alpha}(i) = -\sum_{b=1}^{m} P_b^{\alpha}(i) log_m(P_b^{\alpha}(i)) \tag{2}$$

The string based entropy $[E(i)]$ results in larger values for less conserved positions (large variations) as changes do not take into account the similarity of some of the amino acids imposing a severe limitation to the method. The amino acids with identical values for a given property will be placed in the same bin for property based entropy $[E^{\alpha}(i)]$ even the type of amino acid is different. This sets a range for which such substitutions can be treated as identical by taking the similarity between amino acid with respect to the property into account. The estimated entropy is lower, with increased accuracy for the property based approach as compared to string based approach as shown in figure 1(b) which gives the difference between the values of hydrophobicity and string based entropy ($E^{\alpha}(i) - E(i)$) where α is hydrophobicity. For all position the hydrophobicity based entropy is lower implying property conservation is more than the amino acid type conservation. For the given enzyme family the analysis is robust if the number of sequence exceeds 200. The entropy estimate is better if the number of bins are close to 20, increasing the number of bins (≈ 50) makes the property based entropy similar to the string based entropy. Th positions with the lowest entropy for hydrophobicity figure 1(c) estimated as (38, 41, 97, 98, 99, 199, 200, 202), are the conserved motifs for the family. These are most functional positions in the protein family which are the part of the active and catalytic sites as specified by the experimental studies [9, 10] .

4. Kullback-Leibler (KL) Divergence

The information content of a position is quantified by the Kullback-Leibler (KL) divergence or relative entropy defined as the divergence of a probability distribution at a position from the background distribution. The normalized probability distribution (P_j^{α}) for the property is used and compared with the background probability to calculate the divergence in the properties values. The background probability estimated in [4] (that is the mean frequency of amino acids constructed from all proteins) is used to construct a background column of amino acids having frequencies equivalent to the background frequencies of occurrence. Then substituting each amino acid by the rescaled physiochemical property value and using the same binning procedure as discussed above (*in the Entropy section*) for the background column, the background property based distribution Q is estimated. The Kullback-Leibler (KL) divergence or relative entropy between two probability distribution P_j^{α} and Q^{α} is defined as

$$K(P_j^{\alpha}||Q^{\alpha}) = \sum_{i=1}^{n} P_j^{\alpha}(i) log_n \left(\frac{P_j^{\alpha}(i)}{Q^{\alpha}(i)} \right) \tag{3}$$

The figure 1(d) shows the KL divergence score for the hydrophobicity and polarity for the betalactamse family where some positions show high divergence for hydrophobicity while other positions for polarity. We propose that each position is characterized by one or more physiochemical property, *i.e.* the property which has a high divergence score at a given position is responsible for the function of that position. Larger the KL divergence value for a property the more important is the property at that position for functioning. The divergence score can be used to theoretically and numerically assign properties responsible for the normal working of the positions.

properties of amino acids. The entropic measure shows that at functional position within the protein sequence some of the physiochemical properties are more conserved than the type of amino acid. The entropic and the KL divergence based on the physiochemical properties identifies positions that have important structural and functional role. The correlation based on the physiochemical properties, are calculated and compared with the random system. The eigenvalue distribution of the system is compared with null model and analytic results of the RMT. The Comparison reveals that the eigenvalues on both sides of the spectrum deviates significantly from the RMT results. The eigenvalues outside the RMT bound contains useful information. When the eigenvector corresponding to these eigenvalues are analyzed it was found that eigenvector on lower side are more localized and informative, which can group positions into a collective functional role. The eigenvalue spectra of different properties uncovers different set of information.

Acknowledgments

We acknowledge the Department of Science & Technology (DST), India, (SERB-DST No-EMR/2016/006536) for financial support.

References

[1] Chakraborty A and Chakrabarti S 2014 *Brief. Bioinfo.* 16(1) 71 – 88.
[2] Radivojac P, Clark W T, Oron T R, Schnoes A M, Wittkop T, Sokolov A, Graim K, Funk C, Verspoor K, Ben-Hur A and Pandey G 2013 Nat. methods 10(3) p.221. ; Zakeri P, Jeuris B, Vandebril R and Moreau Y 2014 *Bioinfo.* **30** 1850 – 1857 (2014).
[3] Bhadola P and Deo N 2016 *Phys. Rev. E* 94(4) 042409.
[4] Halabi N, Rivoire O, Leibler S and Ranganathan R 2009 Cell 138(4) 774.
[5] Xia X and Li W H 1998 J. Mol. Evol. 47 (5) 557; Yang Z, Nielsen R and Hasegawa M 1998 Mol. Biol. Evol. **15(12)** 1600; Anfinsen C 1973 Science 181 223.
[6] Efetov K *Supersymmetry in Disorder and Chaos* (Cambridge University Press 1997 Cambridge); Guhr T, Muller A and Weidenmuller H A, 1998 Phys. Rep. 299 189; Laloux L, Cizeau P, Bouchaud J P, and Potters M 1999 Phys. Rev. Lett. **83** 1467; Bhadola P and Deo N 2013, Phys. Rev. E 88(3) 032706; Bhadola P, Garg I and Deo N 2013 Nucl. Phys. B **870** 384.
[7] Frere J M 1995 Mol. Microbiol. 16 385; Massova I and Mobashery S 1998 Antimicrob. Agents Chemother. **42 (1)** 1 – 17.
[8] Bush K and Fisher J F 2011 Ann. Rev. Microbio. 65 455 – 478.
[9] Brasseur J L, Dive G, Dideberg O, Charlier P, Frere J M and Ghuysen J M 1991 Biochem. J. 279 213.
[10] Hermann J C, Ridder L, Mulholland A H and Holtje H D 2003 J. AM. CHEM. SOC. 125 9590.
[11] Bowick M J and Brezin E 1991 Phys. Lett. B 268 21.; Feinberg J and Zee A 1997, J. Stat.Phys. 87 473.

Metadata of the chapter that will be visualized in SpringerLink

Book Title	New Perspectives and Challenges in Econophysics and Sociophysics	
Series Title		
Chapter Title	Evolution and Dynamics of the Currency Network	
Copyright Year	2019	
Copyright HolderName	Springer Nature Switzerland AG	
Author	Family Name	**Bhadola**
	Particle	
	Given Name	**Pradeep**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	The Institute for Fundamental Study, Naresuan University
	Address	Phitsanulok, Thailand
	Email	bhadola.pradeep@gmail.com
Corresponding Author	Family Name	**Deo**
	Particle	
	Given Name	**Nivedita**
	Prefix	
	Suffix	
	Role	
	Division	Department of Physics and Astrophysics
	Organization	University of Delhi
	Address	Delhi, 110007, India
	Email	ndeo@physics.du.ac.in

Abstract

We study the statistical and spectral properties of the foreign exchange of 21 different currencies from January 4, 1999 to March 30, 2018. The correlation matrix is calculated for different periods with a rolling window method and the properties are studied for each window. The basic statistics on the correlation matrix shows that the currencies are more and more correlated with times. The distribution of the correlation matrix was very asymmetric with non zero skewness which shows a fat tail behavior for the initial years but approach Gaussian distribution for the later time. The spectral properties of the correlation matrices for each window when compare with the properties of the correlation matrix formed for the complete period and with analytical results for Wishart matrices shows that the distribution is different for the windows comprising the calm and the crisis period. The study of the number of eigenvalues which are outside the random matrix bounds for each window on both sides of spectrum reveals that for the crisis period, the number of eigenvalues outside the lower bound increases as compared to the calm period. This increase in the number of eigenvalues on the lower side of the spectrum for a window also implies a crisis in the near future. The lower end of the spectra contains more information than the higher side as revealed by the entropic measures on the eigenvalues. This entropic measure shows that the eigenvectors on the lower side are more informative and localized. In this work, the analysis of individual eigenvector captures the evolution of interaction among different currencies with time. The analysis shows that the set of most interacting currencies that are active during the calm period and the crisis period are different. The currencies which was dominating in the calm period suddenly lose all weight and new set of currencies become active at the onset and during the crisis. The largest eigenvector of the correlation matrix can separate currencies based on their geographical location.

Chapter 10
Evolution and Dynamics of the Currency Network

Check for updates

Pradeep Bhadola and Nivedita Deo

Abstract We study the statistical and spectral properties of the foreign exchange of 21 different currencies from January 4, 1999 to March 30, 2018. The correlation matrix is calculated for different periods with a rolling window method and the properties are studied for each window. The basic statistics on the correlation matrix shows that the currencies are more and more correlated with times. The distribution of the correlation matrix was very asymmetric with non zero skewness which shows a fat tail behavior for the initial years but approach Gaussian distribution for the later time. The spectral properties of the correlation matrices for each window when compare with the properties of the correlation matrix formed for the complete period and with analytical results for Wishart matrices shows that the distribution is different for the windows comprising the calm and the crisis period. The study of the number of eigenvalues which are outside the random matrix bounds for each window on both sides of spectrum reveals that for the crisis period, the number of eigenvalues outside the lower bound increases as compared to the calm period. This increase in the number of eigenvalues on the lower side of the spectrum for a window also implies a crisis in the near future. The lower end of the spectra contains more information than the higher side as revealed by the entropic measures on the eigenvalues. This entropic measure shows that the eigenvectors on the lower side are more informative and localized. In this work, the analysis of individual eigenvector captures the evolution of interaction among different currencies with time. The analysis shows that the set of most interacting currencies that are active during the calm period and the crisis period are different. The currencies which was dominating in the calm period suddenly lose all weight and new set of currencies become active at the onset and during the crisis. The largest eigenvector of the correlation matrix can separate currencies based on their geographical location.

P. Bhadola
The Institute for Fundamental Study, Naresuan University, Phitsanulok, Thailand
e-mail: bhadola.pradeep@gmail.com

N. Deo (✉)
Department of Physics and Astrophysics, University of Delhi, Delhi 110007, India
e-mail: ndeo@physics.du.ac.in

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_10

463809_1_En_10_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

Introduction

The foreign exchange (FX) are the market which are global, decentralized, over-the-counter which includes all aspects of exchanging, buying or selling currencies determining the foreign exchange rates. The financial markets are deemed as one of the most complex systems that captures human financial activity on a global scale with a very high trading volume leading to high liquidity. The FX market directly or indirectly affects all other financial exchanges or markets therefore peeks the interest of academic community to study the structure, statistical properties and topology of the FX network. The main reasons to study the structure and nature of the world foreign network is first, the world foreign exchange (currency) market is considered to be the largest financial market, which according to the Triennial Central Bank Survey show a daily average trading for over $5.09 trillion per day in April 2016 [1] extending over all countries. Second, the inability to express the absolute price of a given currency i.e. the absence of a reference frame, makes it more complex than any other financial system. In FX, one has to represents a currency in terms of the other base currency. Since internal dynamics of the base currency greatly depends on multiple factors such as political or social changes, economy, inflation as well as and sensitive to events in any part of the world, the local events can have a global effects in the FX network. With the digital age, there is an enormous growth of the electronically recorded financial data. Even with these huge data sets together with the modern day high-throughput computational methods, the understanding of the complex nature, structure, interaction, dynamics and behavior of the foreign exchange rate remains a challenge. One of the important features of the financial systems is the existence of the correlations between different financial commodities or agents. The study of the cross correlations at different time scales between the financial data is been widely study [2–4] and are used for portfolio optimization or asset risk management [5, 6].

System, Data and Nomenclature

For the analysis, the data set comprises of the daily FX rates for 21 different currencies from January 4, 1999 to March 30, 2018 which spans a period of over 18 years. The currency is denoted according to the ISO 4217 standards using three letter code. The list of the countries and their currencies used for the analysis is shown in Table 10.1 The FX rates are expressed in term of a base currency, which in the current case is United States Dollar (USD). The base currency serves as a frame of reference for all other currencies. The dynamics of other currencies by using the same base currency is equivalent to study the dynamics from the perspective of the base currency (USD). In other words, the evolution of the all other currencies are studied in the frame in which the base currency is at rest.

463809_1_En_10_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

Table 10.1 Countries and their respective currencies used for the analysis

Country	Currency	Country	Currency	Country	Currency
Brazil	BRL	South Korea	KRW	Switzerland	CHF
Canada	CAD	Mexico	MXN	Taiwan	TWD
China	CNY	Norway	NOK	Thailand	THB
Denmark	DKK	Sweden	SEK	Australia	AUD
Hong Kong	HKD	South Africa	ZAR	Euro	EUR
India	INR	Singapore	SGD	New Zealand	NZD
Japan	JPY	Sri Lanka	LKR	United Kingdom	GBP

The daily FX exchange rates data was first preprocessed for filtering to remove any numerical artifacts. Let $S_i(t)$ is daily FX rate of a currency i on day t expressed in terms of USD. The logarithmic returns $R_i(t)$ of the currency i on day t is defined as

$$R_i(t) = ln(S_i(t)) - ln(S_i(t - \Delta)) \tag{10.1}$$

where $\Delta t = 1$ day. The normalized logarithmic returns is then given by

$$r_i(t) = \frac{R_i(t) - < R_i >}{\sigma_i} \tag{10.2}$$

where $< R_i >$ is the time average of the returns over the time period and σ_i is the standard deviation of $R_i(t)$ defined as $\sigma_i =< R_i^2 > - < R_i >^2$.

Correlation Coefficients

The Pearson correlation coefficients are used to estimate the correlation between different FX rates. The correlation coefficient between currency i and j is given by

$$C_{i,j} =< r_i(t) r_j(t) > . \tag{10.3}$$

The correlation coefficients are obtained, such that $-1 \leq C_{i,j} \leq 1$ where $C_{i,j} = 1$ represents perfect correlation and $C_{i,j} = -1$ represents perfect anti-correlation. The correlation matrix is a $N \times N$ symmetric matrix where $N = 21$ in this case. These correlation are correlation as viewed in the frame of base currency USD.

To check the evolution of the FX exchange rates the correlation matrix is calculated with a rolling window of size 250 days with a shift of 50 days. Our data spans the period from January 4, 1999 to March 30, 2018, which results in a 97 windows. Various statistical and spectral properties of correlation matrices for each window is studied and compared.

463809_1_En_10_Chapter ☑TYPESET ☐DISK ☐LE ☑CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

Statistics Correlation Matrix

Before analyzing the spectral properties of the correlation matrices, we investigate the statistical properties of the correlation matrix of each window along with the probability density function (PDF) of the independent elements of the correlation matrix i.e. for C_{ij} for $i < j$. This results in the $N(N-1)/2 = 210$ elements. Figure 10.1 shows the PDFs of the correlation coefficients C_{ij}, calculated for the complete interval from 1999–2018 and the correlation coefficients C_{ij}, calculated for the each window. Where the window is represented by the starting date of the window.

From Fig. 10.1, it is evident that the PDFs of the correlation coefficients for the complete dataset (1999–2018) is a non symmetrical distribution with a positive mean (0.289). This distribution (differs from the Gaussian distribution) having a a right-skewed distribution with skewness 0.65. The distribution is right tailed and with a kurtosis of 3.10 (for Gaussian the kurtosis is 3.0). From Fig. 10.1 we can conclude that for the FX exchange rates the positive cross-correlations are more common than negative cross-correlations. With window (time) the PDFs for of the correlation coefficients shifts towards right (the larger positive correlations) and amount of the negative correlations decreases significantly with window (time).

For each window, we calculated and study the descriptive statistics (i.e., the mean, standard deviation, skewness, and kurtosis) of the cross-correlation coefficients $C_{ij}; i > j$, which is shown in Fig. 10.2. For first window (which starts at 04–01–1999), the mean of the correlation coefficient is low (0.10), which start to increase with rolling windows (time). This implies that with time, the FX currency exchange network becomes more and more correlated. An interesting observation in plot of skewness with window Fig. 10.2, is that during the crisis period there is a decrease in the magnitude of skewness. The skewness which was positive for the period far from the crisis (2008), changes sign and becomes negative just before and during the crisis. The absolute value of skewness decreases with window form 1.96 in 1999 to 0.40 in 2017, the same trend is seen in kurtosis which also shows a

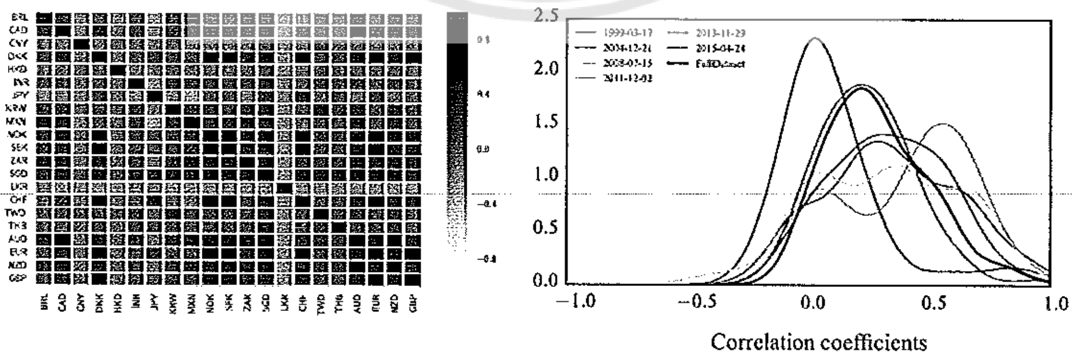


Fig. 10.1 First figure shows the color map of the correlation coefficients between the FX rates of different currencies for the complete interval from 1999–2018. The second figure shows PDFs of C_{ij} for the each window, where the date indicates the start of the window

463809_1_En_10_Chapter ☑TYPESET ☐DISK ☐LE ☑CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

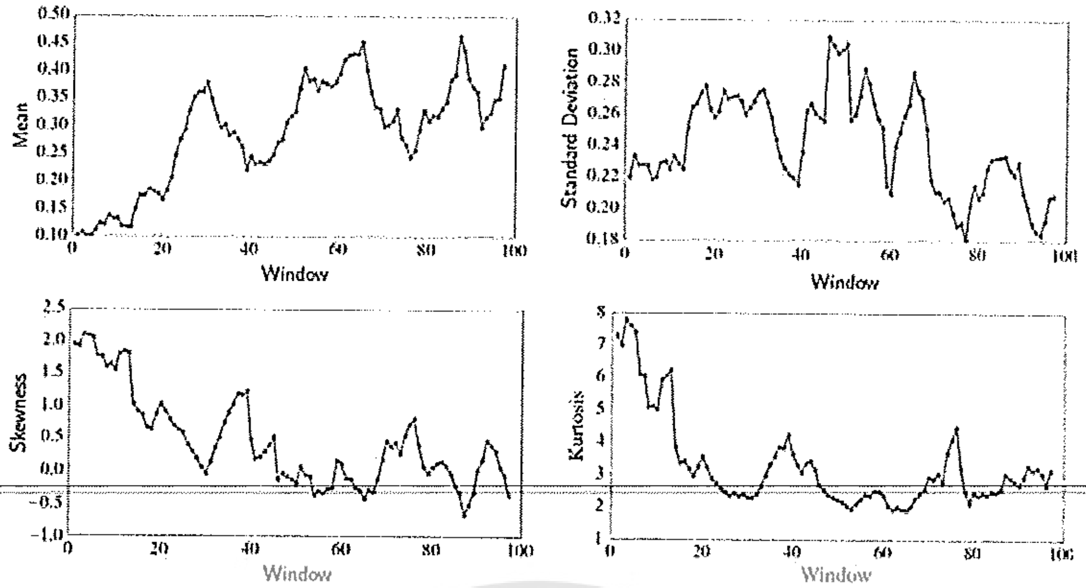


Fig. 10.2 First figure descriptive statistics (mean, standard deviation, skewness and kurtosis) of the cross correlation coefficients C_{ij}; $i > j$ for each window

decrease in value with time (window), in 1999 the kurtosis was 7.8 which decreases to 3.18 in 2017.

For the PDFs and the basic statistics we can conclude that the FX currency network becomes more and more positively correlated with time. Also, as seen with the skewness, the distribution of the cross correlation is becoming more and more symmetric and Gaussian like with the passage of time as the skewness approaching close to zero and the kurtosis approaching close to 3 which are the standard for the Gaussian distribution.

Spectral Properties

The FX exchange correlation structure, in the reference of the base currency USD, can be described by the eigen spectra of the correlation matrix C. For every correlation matrix C, (correlation matrix corresponding to each window), the complete set of the eigenvalues λ_i and eigenvectors v_i, are determined from the eigenvalue equation $C\mathbf{v_i} = \lambda_i \mathbf{v_i}$ where $i = 1 \ldots N$. These eigenvalues are arranged in an ascending order such that $\lambda_1 \leq \lambda_2 \leq \lambda_3 \cdots \leq \lambda_N$. The spectral properties of the correlation matrix for each window is studied.

In the current work, we first compare the eigenvalue distribution of correlation matrix corresponding to each window with the correlation matrix for the complete data set. This may result in the useful insights on how the dynamics for a small scale (window) differs from the over all system dynamics (full data set).

463809_1_En_10_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

To distinguish noise from the information contained in the system, we constructed a null model by randomly shuffling the FX exchange rates data for each currency. We find that the results from the random shuffling are numerically equivalent to the analytical results for the Wishart matrices. The eigen-spectra of Wishart matrices is well studied [7], where the density function for the eigenvalues $P(\lambda)$ is defined as

$$P(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}. \tag{10.4}$$

with $Q = \frac{N}{L} \geq 1$. Where N the number of currencies and L is the number of days for which the exchange rates are observed (used) and $\sigma = 1$ the standard deviation. Equation (10.4) is known as Marcenko–Pastur distribution where the the upper and lower bounds for the eigenvalue λ are defined as

$$\lambda_\pm = \sigma^2\left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right) \tag{10.5}$$

We use the above analytical results, to established the bounds for noise in the eigen-spectrum of the system. For the full data set $N = 21$, number of currencies and the exchange rate data is taken for $L = 4835$ days. This gives $\lambda_+^c = 1.14$ and $\lambda_-^c = 0.87$ where c in the superscript implies that this is for the complete data set. For each window, the number of days are fixed to 250, therefore $L = 250$, the size of the window and $N = 21$ is the number of currencies used for the analysis giving $\lambda_+ = 1.66$ and $\lambda_- = 0.50$.

For each window, we estimate the number of eigenvalues outside the RMT bounds on each side of the eigenvalue spectra ($\lambda \leq \lambda_-$) and ($\lambda \geq \lambda_+$) and is plotted. Figure 10.3 shows that the number of eigenvalues outside the RMT upper bound is very less, (two in most cases but the second largest eigenvalue is very close to λ_+). On the other hand, the number of eigenvalues on the lower side of the spectra for which $\lambda \leq \lambda_-$ are higher in number. On an average 10 out of 21, eigenvalues are outside the lower RMT bounds, which are nearly 50% of the total eigenvalues. This indicates that most of the information about the correlation and interactions between different currencies is located on the lower side of the spectrum.

On of the interesting observation is the number of eigenvalues smaller than the lower RMT bounds, increases at the time of the stress in the global economy. Just before the 2008 crisis, the number of eigenvalues outside the lower bound increases from 9 to 11, which further increase to 13 during the 2008–2009 crisis. After the crisis, the number of eigenvalues outside the lower RMT bounds drops. But further increase in the number (13 eigenvalues $\leq \lambda_-$) is observed for the window corresponding to the period 2011–09–20 to 2012–09–17, this period corresponds to European sovereign debt crisis and the United States debt ceiling crisis [8, 9]. In Fig. 10.3, this was followed a decrease in the number of eigenvalues $\leq \lambda_-$, and again during the 2015–16 Chinese stock market turbulence following slowdown in China and its currency devaluation [9], we observe a increase in number of eigenvalues $\leq \lambda_-$ (again the

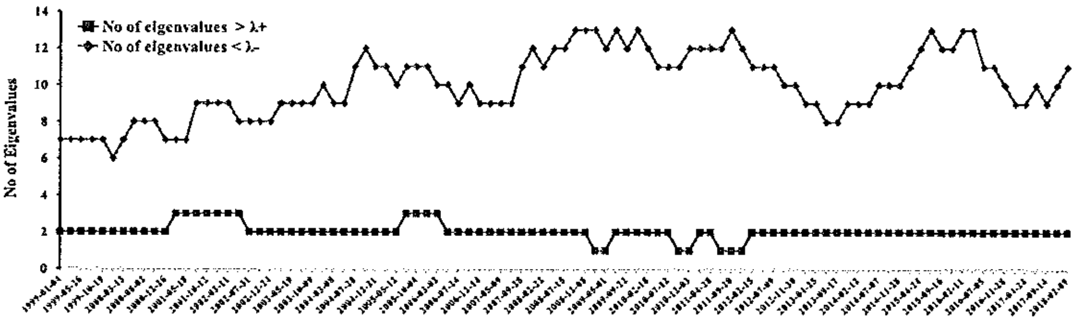


Fig. 10.3 Number of eigenvalues outside the RMT bound on each side of the spectra for every window, where the date indicates the start of the window

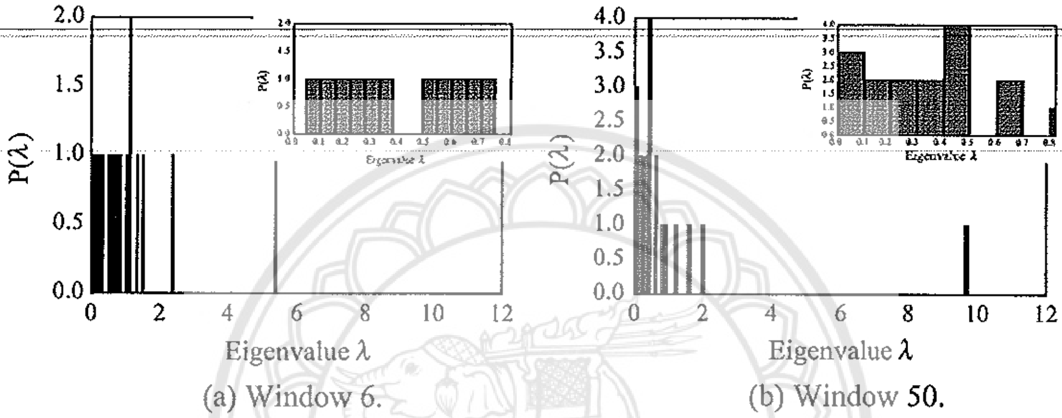
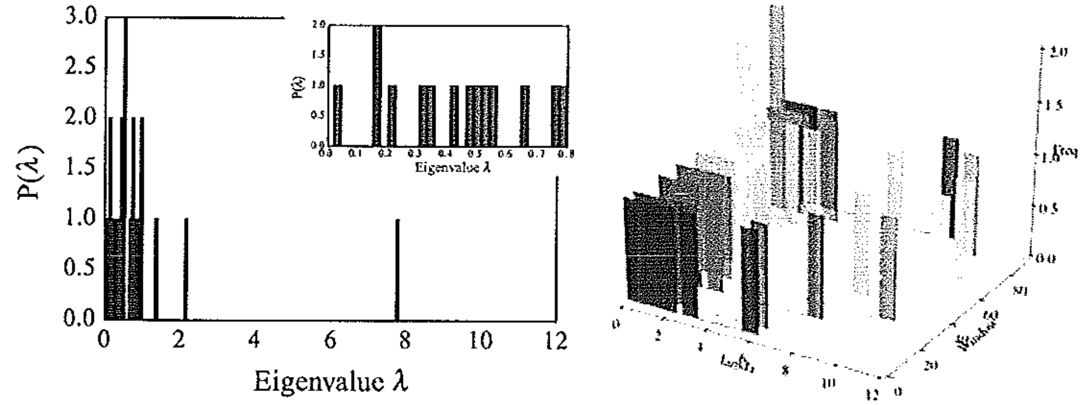


Fig. 10.4 Eigenvalue distributions for different windows. Insets show eigenvalues outside the lower bound

number is 13). The increase in number of eigenvalues on the lower side of spectra that are less than the lower RMT bounds correlates very well with the time of the financial crisis.

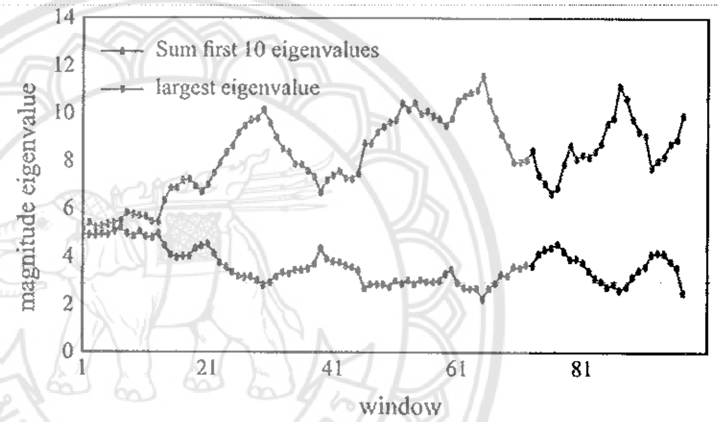
Figures 10.4 and 10.5 shows the eigenvalue distribution for different windows and for the complete period (1999–2018). For Fig. 10.4a which is the calm period (1999–12–30 to 2000–12–22), the magnitude of the largest eigenvalue is 5.4 which is very less compared to the period of high financial stress (crisis) window 50 (2008–09–24 to 2009–09–21, with largest λ equal to 9.6) as shown in Fig. 10.4b. We observe that for the less financial stress the largest eigenvalue of correlation matrix for the FX exchange is lower as compared to the period of high financial stress. The largest eigenvalue of the correlation matrix of FX exchange rates for the complete period is 7.8. We study the time evolution of each eigenvalue which is outside the RMT bound. We observe the the sum of the eigenvalues on the lower side of spectra outside the lower RMT bound is opposite to the dynamics of largest eigenvalue as shown in Fig. 10.6.



(a) For full period (1989–2018). (b) For different windows.

Fig. 10.5 Eigenvalue distributions. Insets show eigenvalues outside the lower bound

Fig. 10.6 Dynamics of the largest and the sum first few eigenvalues



Information Content of Eigenvectors

The information content of each eigenvector is estimated by estimating the Shannon's entropy. The entropy of an eigenvector v_i is defined as

$$H_i = -\sum_{j=1}^{N} u_i(j) log_L(u_i(j)), \tag{10.6}$$

where N is the total number of currencies used for the analysis (number of eigenvector components) and $u_i(j) = (v_i(j))^2$ is the square of the jth component of the ith normalized vector v_i. The entropy estimates also helps to calculate the localization of the eigenvector. Eigenvector with low entropy should be highly localized. In the current analysis, we find that the eigenvector corresponding to the small eigenvalues are very informative as well as highly localized as compared to the eigenvector corresponding to the large eigenvalues. These eigenvectors are further used to estimate the strong interacting pair of currencies. For many systems especially for correlation

463809_1_En_10_Chapter ☑TYPESET ☐DISK ☐LE ☑CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

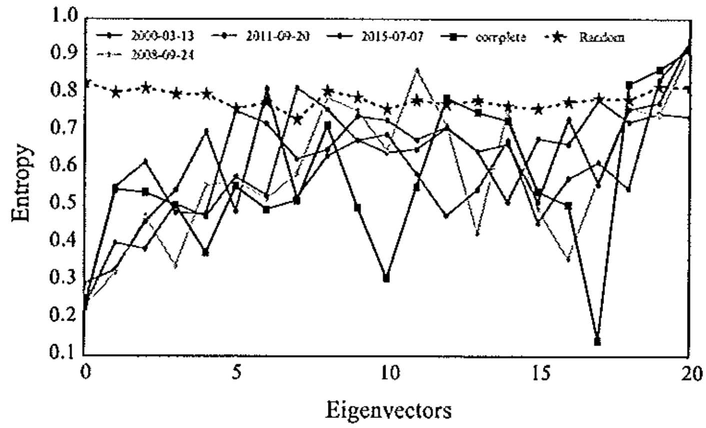


Fig. 10.7 Entropy of eigenvector for different windows, where the date indicates the start of the window

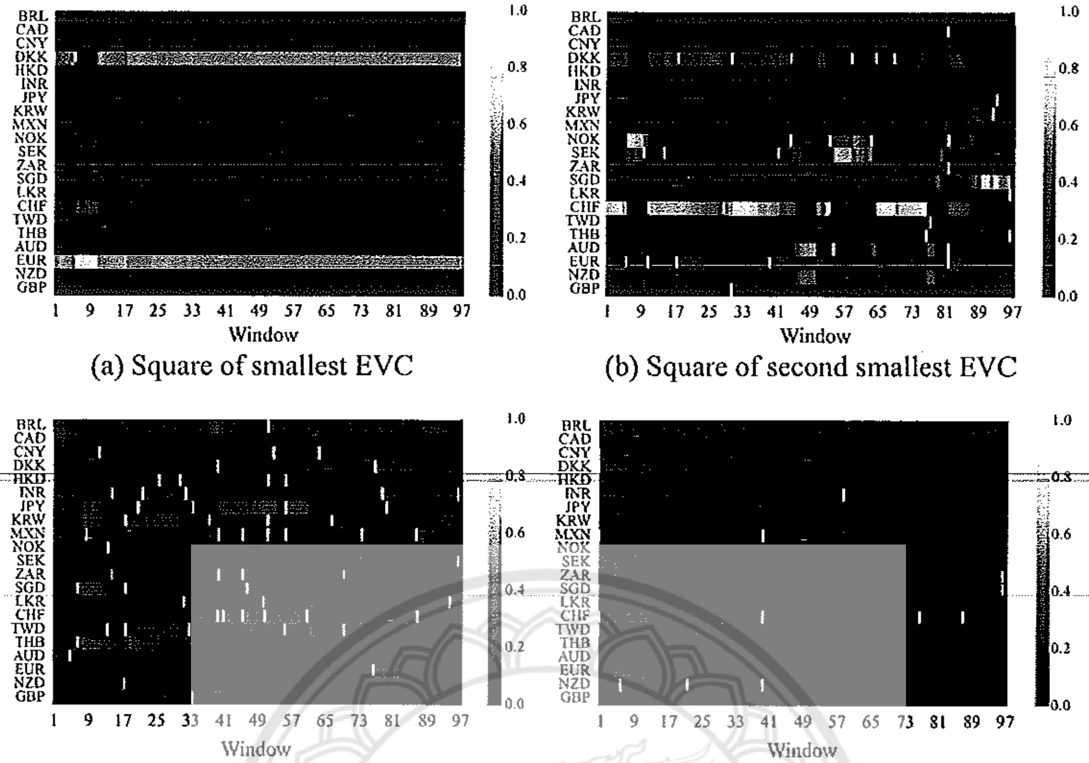
matrices between position in a protein family, it is established that the eigenvector corresponding to the small eigenvalues can identify important positions and interaction responsible for the formation of motifs and sectors [10]. In the Markowitzs theory of optimal portfolios [11], the least risky portfolios corresponds to the lowest eigenvalues and the corresponding eigenvectors. In recent work [12], which involves the analysis of the correlation structure of global financial indices, it was established that the lower side of the eigenvalue spectrum is more informative and localized which is able to capture most of the system dynamics.

Figure 10.7 compares the eigenvector entropy with for different windows with the random shuffled system and eigenvector entropy for the complete period. The eigenvectors corresponding to the large eigenvalues have entropy values very close to the random system (black dashed line). But there is a clear distinction of the entropy of the small eigenvectors with the random system. The fist few eigenvectors have very small entropy (≈ 0.25) for all windows. These eigenvectors are highly localized and gives the set of highly interacting currencies. Analyzing each eigenvector independently reveals clusters of currencies with very close ties.

Eigenvector Components

The analysis of the eigenvector corresponding to the eigenvalues outside the RMT bounds should contain the information present within the system. We use the square of the eigenvector components, to determine their contribution towards that a given eigenvalue. As it is already been discussed in the previous section, that the low eigenvector are highly localized as compared to the high eigenvector components (EVC). Figure 10.8 shows the variation of the square of the component for each currency with window, for the two lowest and two largest eigenvectors. From the Fig. 10.8a it is clear that the lowest eigenvector is highly localized over all windows and there are only a few currencies that are contributing to it. Most of the time (window), there are only two contributing currencies (Euro (EUR) and Danish Krone

463809_1_En_10_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard



(a) Square of smallest EVC

(b) Square of second smallest EVC

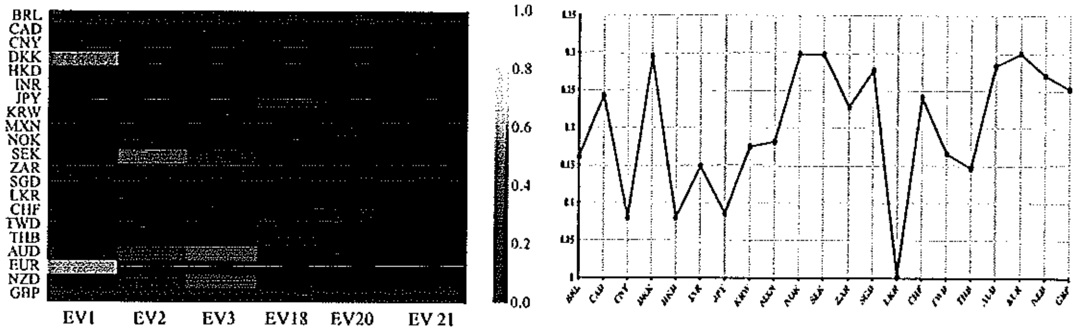
(c) Square of second largest EVC

(d) Square of largest EVC

Fig. 10.8 The square of the eigenvector components (EVC) for different windows

(DKK)) but only for a short period from window number 6–10 (period from 1999–12–30 to 2001–10–11) Swiss Franc (CHF) have high interactions with Euro. This implies that there exist strong interaction between the Euro (EUR) and Danish Krone (DKK).

For the second smallest EVC Fig. 10.8b, we find the contribution depends on the time period of observation (window). The second smallest eigenvector shows a drastic change in the contributing currencies for the period of financial stress. The early 2000s recession represented by window number 6–10 (period from 1999–12–30 to 2001–10–11), contributing components changes from Euro (EUR), Danish Krone (DKK) and Swiss Franc (CHF) to Norwegian Krone (NOK), Swedish Krona (SEK), Australian Dollar (AUD) and New Zealand Dollar (NZD). A similar change is observed for during the global 2008 financial crisis (window 46–50, period from 2007–12–10 to 2009–07–10), European sovereign debt crisis and the United States debt ceiling crisis a period from 2010–02–18 to 2012–09–17 (window 57–65). The three currencies (EUR, DKK and CHF) dominates the second smallest eigenvector forming very close ties during the calm period but at the onset and during crisis, their interaction cease to exist and a new interaction between NOK, SEK, AUD and NZD is formed. These four currencies dominates forming very strong interactions during the crisis. During the Chinese stock market turbulence captured by window 79–85, there are other currencies such as SGD, ZOR, MXN, JPY, KRW that have



(a) Square of EVC (b) Largest EVC

Fig. 10.9 Eigenvector components (EVC) for the complete time period 1999–2017

contribution. The largest and the second largest eigenvector for the windows are not localized and have contribution from all the currencies and the contribution changes with time (window) Fig. 10.8c, d.

We compare the eigenvector components obtained for each window with the eigenvector components for the complete time period (1999–2018). Figure 10.9a shows the distribution of the square of the smallest and largest three eigenvector. The small eigenvectors are localized and more informative where as the larger eigenvectors are non localized and plagued with noise. The smallest eigenvector overall shows the same behavior as the shown by the smallest eigenvector for each window where the only contribution are from Euro (EUR) and Danish Krone (DKK). For the second smallest eigenvector, there are a total of 7 contributing currencies Euro (EUR), Danish Krone (DKK), Swiss Franc (CHF), Norwegian Krone (NOK), Swedish Krona (SEK), Australian Dollar (AUD) and New Zealand Dollar (NZD). These currencies are same as obtained analyzing the second smallest eigenvector for all windows Fig. 10.8b, but there are divided into two set one set (EUR), DKK and CHF) dominates during the calm period where the other set (NOK, SEK, AUD and NZD) which is more active during the crisis period.

Analyzing the components of the largest eigenvector Fig. 10.9b, we found that the currencies can be separated by their geographical location based on the range of the value of their contribution. If we choose components by the magnitude, and we define two groups, first group which has magnitude of component greater than 0.2 and the other group with magnitude less than or equal to 0.2. The details of the currencies and their geographical location are given in Table 10.2. The Asian and the Latin American countries form a one group which corresponds to the components less than 0.2 Fig. 10.9b, where as the other group is formed by the countries whose component greater than 0.2. These are the countries belonging to Europe, Australia, one each from North America, Africa and Asia.

463809_1_En_10_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

Table 10.2 Currencies with respective countries separated based on the magnitude of the largest eigenvector of correlation matrix for the complete dataset

EVC <= 0.2			EVC > 0.2		
Currency	County	Continent	Currency	County	Continent
BRL	Brazil	Latin America	CAD	Canada	North America
CNY	China	Asia	DKK	Denmark	Europe
HKD	Hong Kong	Asia	NOK	Norway	Europe
INR	India	Asia	SEK	Sweden	Europe
JPY	Japan	Asia	ZAR	South Africa	Africa
KRW	South Korea	Asia	SGD	Singapore	Asia
MXN	Mexico	Latin America	CHF	Switzerland	Europe
LKR	Sri Lanka	Asia	AUD	Australia	Australia
TWD	Taiwan	Asia	EUR	Euro	Europe
THB	Thailand	Asia	NZD	New Zealand	Australia

Conclusions

In this work, we try to study and understand the relation and interaction between the foreign exchanges rates for a period from January 1999 to March 2018. All the foreign exchange rates are expressed in terms of a base currency which for the current case is USD. All the dynamics and properties studied in this paper will be in the reference frame where the USD is at rest. The evolution of FX rates is studied using a rolling window of size 250 days with a shift of 50 days. The statistics and the correlation between different currencies is calculated and studied. The statistics on the correlation matrix reveals that with time the currencies are getting more and more correlated. At the start of the period (1999) the mean of the correlation matrix was very less as well was very highly non symmetric (non zero skewness) and high Kurtosis but with time the distribution of correlation coefficients try to approach Gaussian distribution by approaching zero skewness 0 and Kurtosis (three) standard for Gaussian distribution. The spectral properties of the correlation matrices are studied for each window and then compared with the correlation matrix formed from the complete data set and with the analytical results for Wishart matrices. The distribution of eigenvalues reveals the distribution is different for the calm and the crisis period. The number of eigenvalues for each window which are outside the random matrix bounds are higher on the lower sides. For the period of the crisis the number of eigenvalues outside the lower bound increases as compared to the calm period. This may be to incorporate the addition information generated during the crisis. We propose that the if there is increase in the number of eigenvalues outside the rmt bound then that may indicate a crisis in the near future. The information content and localization of each eigenvector is estimated by the using an entropic measure. This measures shows that the eigenvalues on the lower side of the spectra are more localized as well as informative as compared to the eigenvalue on the higher side of the spectra. The analysis of the individual eigenvectors gives information about the interaction between different currencies. We observe in the second smallest eigenvector, at each crisis period, the contribution and interaction among the currencies changes. The currencies which was dominating in the calm period suddenly lose all their contribution during the crisis and a new group of interaction between currencies become active at the onset and during the crisis. The components of largest eigenvector of the correlation matrix for the complete period can separate the currencies based on their geographical location based on the magnitude of the components.

Acknowledgements We acknowledge the Department of Science and Technology (DST), India, (SERB-DST No- EMR/2016/006536) for financial support.

References

1. Triennial Central Bank Survey of foreign exchange and OTC derivatives markets (2016). https://www.bis.org/publ/rpfx16.htm
2. Kenett, D.Y., Huang, X., Vodenska, I., Havlin, S., Stanley, H.E.: Partial correlation analysis: applications for financial markets. Quant. Financ. 15(4), 569–578 (2015)
3. Podobnik, B., Stanley, H.E.: Detrended cross-correlation analysis: a new method for analysing two non-stationary time series. Phys. Rev. Lett. **100**, 084102 (2008)
4. Conlon, T., Ruskin, H.J., Crane, M.: Random matrix theory and fund of funds portfolio optimisation. Phys. A **382**(2), 565–576 (2007)
5. Kenett, Dror Y., Shapira, Yoash, Madi, Asaf, Bransburg-Zabary, Sharron, Gur-Gershgoren, Gitit, Ben-Jacob, Eshel: Dynamics of stock market correlations. AUCO Czech Econ. Rev. 4(3), 330–341 (2010)
6. Kenett, Dror Y., Preis, Tobias, Gur-Gershgoren, Gitit, Ben-Jacob, Eshel: Quantifying meta-correlations in financial markets. Eur. Lett. 99(3), 38001 (2012)
7. Bowick M. J., Brezin E.: Phys. Lett. B **268**, 21 (1991); Feinberg J., Zee A.J.: Stat. Phys. **87**, 473 (1997)
8. Jayech, Selma: The contagion channels of July–August-2011 stock market crash: a DAG-copula based approach. Eur. J. Oper. Res. 249(2), 631–646 (2016)
9. Clements, Adam, Hurn, Stan, Shi, Shuping: An empirical investigation of herding in the US stock market. Econ. Model. **67**, 184–192 (2017)
10. Bhadola, Pradeep, Deo, Nivedita: Targeting functional motifs of a protein family. Phys. Rev. E 94(4), 042409 (2016)
11. Elton E.J., Gruber, M.J.: Modern Portfolio Theory and Investment Analysis. Wiley, New York (1995); Markowitz, H.: Portfolio Selection: Efficient Diversification of Investments. Wiley, New York (1959). See also: Bouchaud J.P., Potters, M.: Theory of Financial Risk. Alea-Saclay, Eyrolles, Paris (1997) (in French)
12. Pradeep Bhadola, Nivedita Deo.: Extreme eigenvector analysis of global financial correlation matrices. In: Econophysics and Sociophysics: Recent Progress and Future Directions, pp. 59–69, Springer, Cham (2017)

463809_1_En_10_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:19/1/2019 Pages: 13 Layout: T1-Standard

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	/ through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	/ through letter or through characters	new character / or new characters /
Change to italics	— under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	~ under matter to be changed	~
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	/ through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	/ through character or where required	
Reduce space between characters or words	| between characters or words affected	