Recurrence Network Analysis of Financial Data

Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Amey S. Amanagi

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2018

Supervisor: Prof. G. Ambika

Certificate

This is to certify that this dissertation entitled Recurrence Network Analysis of Financial Data, submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents the study/work carried out by Amey S. Amanagiat Indian Institute of Science Education and Research under the supervision of Prof. G. Ambika, Professor, Department of Physics , during the academic year 2017-2018.

Prof. G. Ambika

Data : March 30, 2018.

Committee:

Prof. G. Ambika

Dr. Anindya Goswami

This thesis is dedicated to my family.

Declaration

I hereby declare that the matter embodied in the report entitled Recurrence Network Analysis of Financial Data are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. G. Ambika and the same has not been submitted elsewhere for any other degree.

Data : March 30, 2018.

Amey S. Amanagi

Acknowledgments

My sincere thanks to prof. G. Ambika for giving me this wonderful opportunity to work on this special topic and giving her guidance and time which are crucial for this project. I would like to thank Dr. Anindya Goswami for being the TAC member for this project. My special thanks to our group Sandip, Kajri, Kashyap and Yamini for their help. I also thank IISER Pune for giving all the facility for this research project.

Amey S. Amanagi

Abstract

In the project we have used methods developed in the field of nonlinear dynamics to investigate financial markets especially near 2008 financial crisis. We have used the framework of fallibility and reflexivity suggested by George Soros to interpret our results. We start with establishing a reasonable assurance of the existence of underlying dynamics in financial market and extend our study to comment on the overall behaviour. First we remove the long term variation in the time series by polynomial detrending method, followed by taking uniform deviates to get all the data sets in a comparable scale. Power spectrum analysis for data sets gives slope of log-log plots of power vs. frequency in the range of -1.68 to -1.85, while that for red noise turns out to be -1.95, which suggests that financial systems that we are dealing with have some underlying dynamics but not totally dominated by red noise. D_2 correlation dimension based surrogate analysis suggests that data has underlying dynamics. Recurrence network based analysis by reconstruction of phase space also suggests existence of underlying deterministic dynamics, since degree distribution saturates after embedding dimension of 4. Other network measures such as clustering coefficient and characteristic path length also suggest red noise like characteristic added to underlying dynamics. To study the change in underlying dynamics near 2008 financial crisis, it is important to study time evolution of these network measures. To do that, we use windowed analysis that gives insights into temporal evolution of the dynamical characteristic which is crucial in exploring system near 2008 financial crisis. We have also found a reasonable possibility of forecasting stock market crash, although it will require detailed further study to conclude decisively.

Contents

Chapter 1

Introduction

Financial data analysis is one of the most rapidly expanding disciplines in the world today. Emergence of Quant investment funds that use computer algorithmic trading based on technical analysis on stock market charts have increased the demand for new developments in financial data analysis [1]. The fact that new financial data analysis technique can directly be applied on historic or real time data makes it easy to course correct, making the field quite dynamic. Another factor responsible for the growth is the development in the field of big data analysis in general, thanks to machine learning algorithms [2]. These developments have given new insights into the dynamics of financial markets.

Econophysics is an interdisciplinary field that uses the methods of statistical mechanics to study economic systems. This approach has given new insight into understanding of economic systems. First notable achievement in quantitative study of financial market done by Bachelier in his PhD thesis, which was an attempt to understand speculation in the market [3]. The important factor in this study was that Bachelier assigned probabilities to price changes in financial markets. In 1973, Black and Scholes put forward option pricing theory, that assumes geometric brownian motion in financial markets and provides rational price for European option [4]. Another very important idea in understanding financial markets is that price changes follow Levy stable distribution proposed by Mandelbrot [5]. A ground assumption in many theories is the efficient market hypothesis, which means that all the public information about an asset in the market is instantly processed and immediately reflected in the price of the asset in the market [6]. Despite efficient market hypothesis, Algorithmic complexity theory suggested by Kolmogorov and Chaitin allowed for unpredictability in a financial market time series [7]. The scaling behaviour, can be of two types, either in volatility data which leads to fractional analysis and Hurst exponent or the behaviour of the tails of the distribution of returns as a function of the size of the movement which leads to tail index [9] [8]. For financial modelling, GARCH model has been very important development that uses autoregressive moving average models on a time series [11]. Understanding of correlations between pairs of stocks has been crucial in designing efficient financial portfolio [12] [13]. An interesting development in geometric visualization based understanding was done by studying distance between a synchronous evolving pair of assets in a particular financial market, first introduced by Mantegna [14]. With all these developments giving new insights in understanding financial markets as well as giving applicable new strategies of investments, Econophysics has become a very important interdisciplinary field in recent years, that involves principles of physics applied to economics.

Our approach is based on treating financial market as a dynamical system. According to Jerome Singer a complex system is "a system that involves numerous interacting agents whose aggregate behaviors are to be understood. Such aggregate activity is nonlinear, hence it cannot simply be derived from summation of individual components behavior." The nature of agent is quite crucial in understanding the dynamics of the system. Here the agent or group of agents or one agent over multiple measurements in the system are following a previously known mechanics or some reasonable condition. This is where natural systems and socio-economic systems are different. The framework we have employed in this project is one suggested by George Soros [15]. The agent in a socio-economic system is intelligent as opposed to a natural system. Here what we mean by intelligence is that agent takes decisions, interacts with other agents based on his own observation, understanding, judgment and interest. According to Soros agent's understanding of the system is incomplete or flawed and yet he is forced to take decision regardless [15]. This decision in case of financial data is doing a transaction like buying a stock or gold, in case of population data it is voting. This decision making based on flawed understanding is called fallibility [15]. Now once the decision is made, its outcome is going to affect the system and it is highly likely that it will further deteriorate the agents understanding of the system, because the only way of avoiding further deteriorate the understanding is brutally objective analysis of decisions taken, which is a rare trait especially in case of a financial market bubble. For example in financial market in case of an overvalued stock, agents continuously keep buying it leading to unreasonable rise in its price making it more wrongly priced, and the fact that people are still buying it means they think the stock is worth that price meaning their judgment worsening. This phenomenon is called reflexivity [15]. This continuous rise in an overpriced asset is called a bubble in financial market. After the bubble becomes large beyond sustainability, it bursts meaning that the price sharply falls, leading to panic in the market. The complete process of creation and bursting of a bubble is called a boom-burst process. Important thing to note is that fallibility and reflexivity are intrinsic in a socio-economic system and they inevitably give rise to boom-burst processes [15]. Another important characteristic of socio-economic systems is low signal to noise ratio. In the project we try to detect signs of this peculiar dynamics in stock market data near 2008 crash which one of the most well known bubble in recent history.

Financial crisis of 2008 is one of the most important events in recent history, one that has caused serious damage to both US and global economy. It has also raised serious questions to the regulatory framework required for a financial market. In US, Federal reserve decides the monetary policy and the pace of the economic activity by controlling the interest rates on lending. After the 2001 recession, the federal reserve kept the interest rates low so as to stimulate the economic activity, allowing businesses to borrow more from banks, take more risks, expand their businesses. With low cost lending bank started more and more credit risk meaning higher risk of default. Now the number of people seeking housing loans increase as a result and with more buyers available, housing prices started rising. At this point fallibility became dominant as people made crucial misjudgement that the housing price will continue rising indefinitely thinking that real estate will always be in demand. There was oversupply of retail and commercial real estate causing prices to drop. People defaulted on their housing loan making banks realize their credit risk. Even if banks get the houses and sell them, they won't recover the money they lent when the prices were high. This way leading American banks and financial services companies started declaring bankruptcy. When the leading investment bank Lehman brothers filed for bankruptcy, financial sector of US economy started failing causing whole stock markets to crash. Given the importance of US economy in global economy, other stock markets in other countries crashed as well, causing global financial crisis.

We look at the financial market as a dynamical system and try to understand its complexity from the complexity of its underlying dynamics. In this context, we start with standard nonlinear dynamical systems like Lorenz [16] and Rössler [17] that exhibit chaos and hence complex and unpredictable dynamics. We try to relate the complexity and unpredictability of financial data to the nonlinear nature of its dynamics. We use the techniques of nonlinear time series analysis to characterize the complexity of the dynamics. However, price data sets are not large enough, so we can not use multifractals analysis effectively. Hence use the recent technique of recurrence networks for the analysis.

Irregular structures of N nodes connected by arbitrary number of edges, constructed in abstract space are called complex networks. Recurrence networks are a type of complex networks selected using the recurrence of dynamics on the reconstructed phase space. Network based methods are used in wide variety of fields giving a new, structural perspective for understanding of different systems. We have dealt with undirected complex networks in our study. First remarkable work done on networks was by Erdos and Renyi [18], that explained properties of random networks or historically, random graphs. Then different measures that can be used to quantify those properties were developed. The most important of these measures are degree distribution, clustering coefficient and characteristic path length, discussed in detailed elsewhere [19–22]. Dynamical systems are studied successfully by using network based measures [23] [24]. The main idea behind this kind of analysis is trying to get information about underlying chaotic attractor by constructing a complex network from time series of the system according to a suitable scheme [25]. To quantify the nature of attractor we use the complex network measures. There multiple methods suggested in the literature that construct complex network out of time series, such as cycle networks [26], visibility graphs [27], transition networks [28] and recurrence networks (RN) [29]. There are different methods to construct recurrence networks such as the correlation networks [30], k-nearest neighbour networks [31] and ϵ - recurrence networks [32]. A comparative study of the different characteristics of the attractor measured by these different schemes of constructing recurrence networks is done elsewhere [33, 34]. Among all these methods, we have used ϵ - recurrence networks since the project mainly involves studying real world data of financial markets and ϵ - recurrence networks are used successfully in studying dynamical transitions in palaeo-climate data, which is also a real world system [35–37] along with life sciences [38, 39], earth sciences [40], and astrophysics [41].

The main objective of the thesis is to study the dynamics of the stock market near this global economic crisis and try to see if there are some early warning signals about the coming crash. We have used methods of nonlinear dynamics and networks to study the underlying dynamics in financial markets. We compared our results with the results of standard deterministic chaotic systems Lorenz and Rössler systems along with the results of white and red noise. Based on this comparative study we try to see if the behaviour of stock market is like a standard chaotic system or dominated by noise. Given the long term variations present in the time series, we first detrend it and then take uniform deviate to get them in a comparable scale. Then we decide the parameters for phase space reconstruction such as embedding dimension by false nearest neighbor method and time delay by auto-correlation function method. We do surrogate analysis and compare D_2 correlation dimension for data and surrogates to get an idea of underlying determinism in the system. For recurrence networks, the embedding threshold criterion is, the minimum epsilon required for the giant component to include 95% of the total nodes. After the phase space reconstruction we construct recurrence networks from the reconstructed trajectory of the standard systems and noise. We calculate measures on recurrence networks such as degree distribution, clustering coefficient and characteristic path length. We compare these recurrence network based measures for data, standard chaotic systems and noise to get insights into dynamics of the financial markets. For quantification of the underlying determinism of the data, we do recurrence plots and recurrence quantification analysis(RQA), which includes DET measures on data and standard systems. To understand time evolution of dynamics of the system, we do windowed analysis. Here we use a time window of 2000 data points, calculate recurrence quantification measures, recurrence network measures and study how they change with sliding the window by 50 points. We also do this analysis on surrogates to see how determinative nature in underlying system is evolving. Our main focus in this sliding window analysis is to understand the changes near the 2008 crash. This may help us to identify early warning signals of stock market crashes or bubbles.

Chapter 2

Data and preliminary analysis

Stock market data is one of the most studied systems when it comes to financial data analysis. A stock market is a financial market where different companies raise capital by selling a part of ownership of the company. This fractional ownership is called a share or a stock. Like any financial market, price of a stock is dependant on the judgment of the people who are willing to buy it. If a company has good prospects, meaning high profitability as compared to its competition, more people will be willing to buy its share making the share price go up. If a company is not performing well or has an uncertain future, less people will be ready to buy its share lowering the price. Hence a stock market is an extremely dynamic and vibrant system whose accurate data is available. Stock market price is also considered to be one of the important indicator of a county's economic condition. In the project we are looking for understanding the underlying dynamics of this complex system by studying the price of the whole index. We first try to understand the implications of price of a stock market index.

2.1 Financial data

It is important to understand how a stock market index price is determined and what it implies about the underlying economy. When a company decides to raise capital by going public i.e. getting listed on a stock exchange, it authorizes the availability of its fractional ownership for a price. Market capitalization of a publicly traded company at a certain time is total number of stocks issued by company times price per stock. Different listed companies have different market capitalization. Stock market index is basically a weighted average of stock prices of several top companies listed on a stock exchange. Based on the scheme for assigning these weights there are two types of indices. Most widely used is capitalization-weighted (or "cap-weighted") indices that assign weights based on market capitalization of a company as a fraction of total value that is available for trade on the stock exchange. Another type of indices is price-weighted indices, that assign weights based on price. A stock at price of $100 will have 10 times more weight than stock at price $10 in a price weighted index. Out of the 10 markets studied, only Nikkei is price weighted index. The rest are capitalization-weighted indices. Next it is important to know how the stock exchanges are symbolized for reference in the literature.

2.2 Specification of Data

We collected daily closing indices from 7th of January 1991 to 17th of March 2017 of the 10 major stock exchanges, viz. Dow Jones Industrial average, S & P 500, NASDAQ, Shanghai Stock Exchange, Nikkei, DAX, CAC40 , Sensex, Torronto Stock Exchange, New York Stock Exchange. The source of the data in yahoo finance.

List below gives list of data sets and their symbols:
NYSE = New York stock exchange (USA)
DJI = Dow Jones Industrial Average (USA)
GSPC = Standard & Poor 500 (USA)
NASDAQ = National Association of Securities Dealers Automated Quotations (USA)
SSE = Shanghai Stock Exchange (China)
Nikkei = Nikkei index (Japan)
DAX = Performance Index (Germany)
CAC = Cotation Assiste en Continu (France)
Sensex = Sensitive Index (India)
TSX = Torronto Stock Exchange (Canada)
Next section focuses on the graphical representation of the price of the stock market index over the period of 27 years that we are interested in.

2.3 Time series plots

In this section we present the time series plots of the 10 data sets that are studied in the project. Following are the time series plots of all the data sets used for the study. The data represented in the figures is the stock market index prices in USD for the 10 stock exchanges.

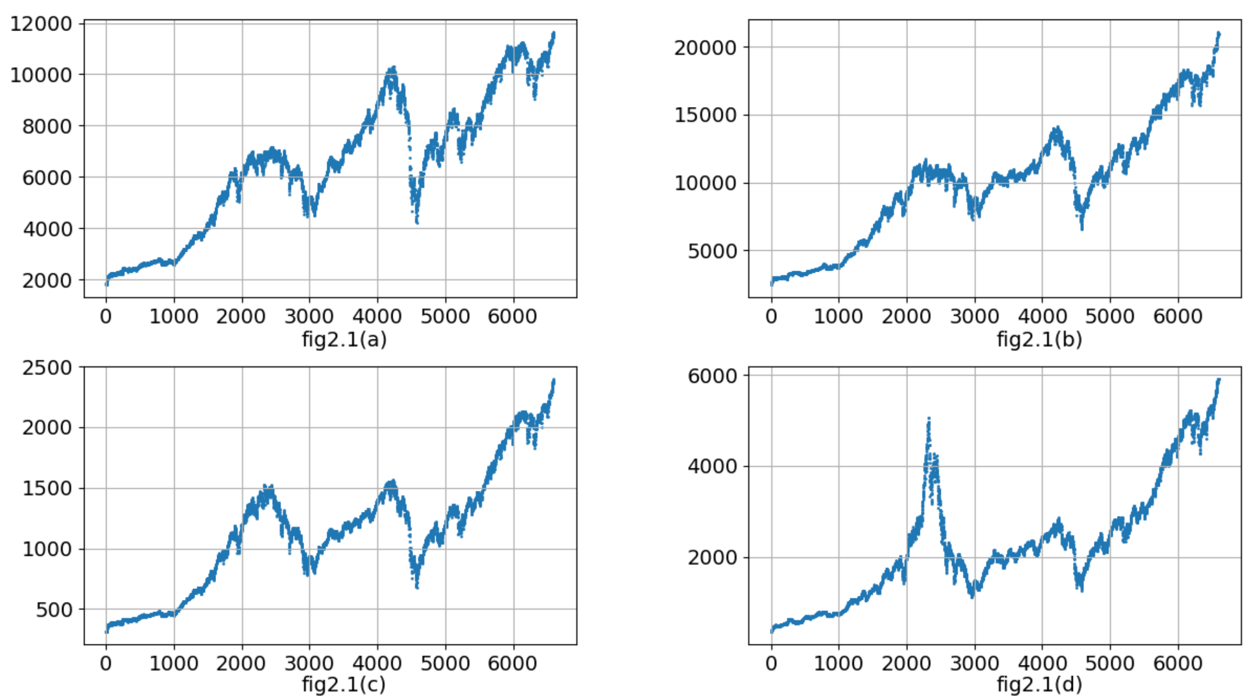

Figure 2.1: Time series plots for NYSE(a), DJI(b), GSPC(c), NASDAQ(d)

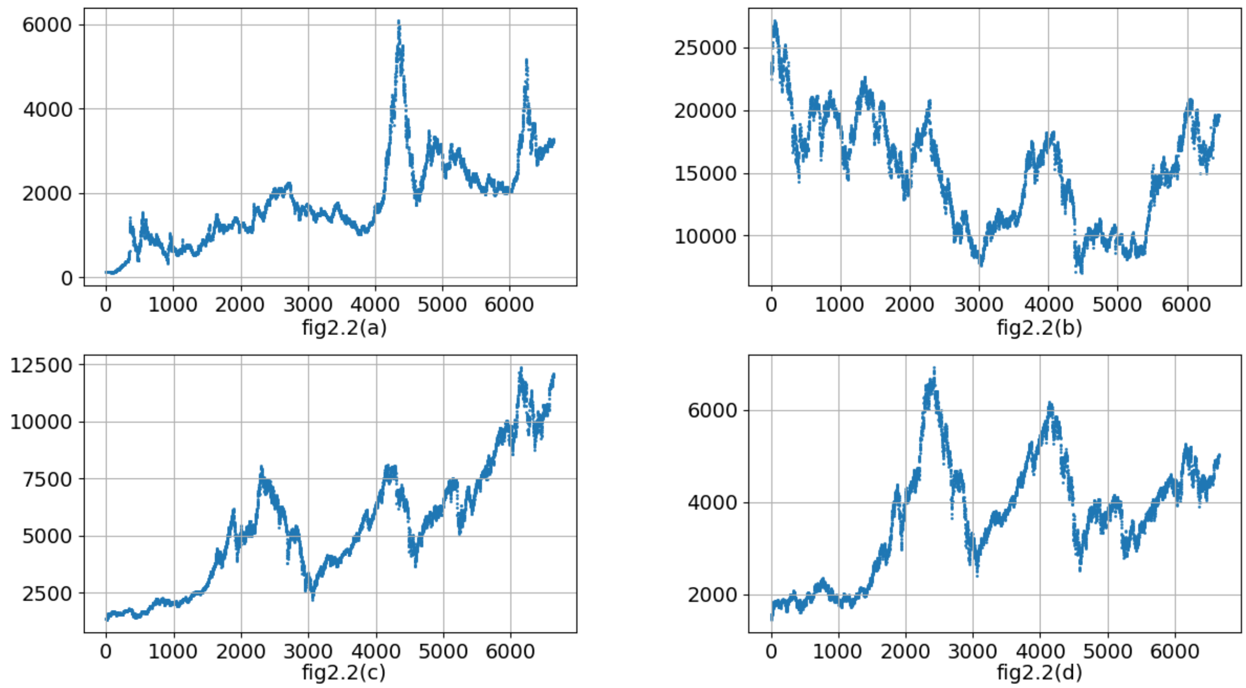

Figure 2.2: Time series plots for SSE(a), Nikkei(b), DAX(c), CAC(d)

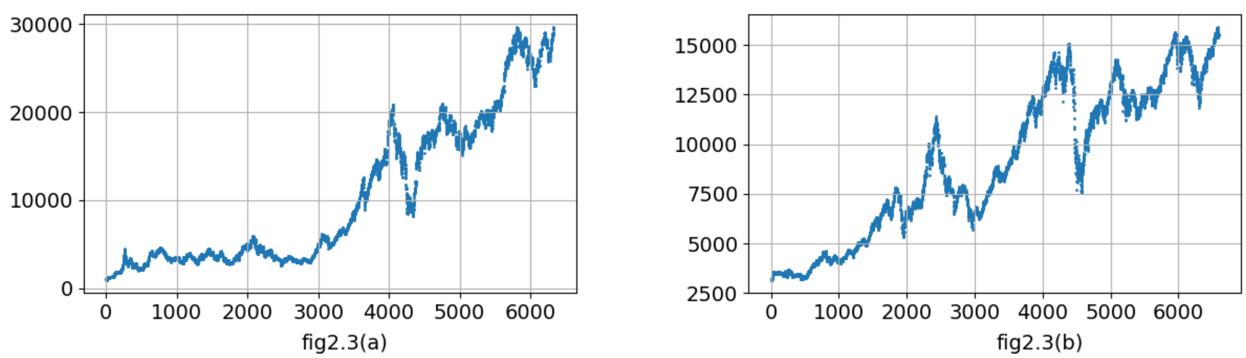

Figure 2.3: Time series plots for Sensex(a), TSX(b)

Even if it is well known that these stock markets are considerably related to each other, it can be seen that they have behaved quite differently over the years. Some important features in the time series plots are :

1. The prices of almost all the stock market indices have a long term upward variation. It is because the top companies whose stock prices are used to calculate price of an index generally become more valuable by expanding their business operations over time. So their prices increase.

2. between points 4000 and 5000, we see a sharp fall in all the markets. This is the famous crash of 2008. In the project we will study the behaviour of the time series near this point in great detail.

3. NASDAQ data set, shows a significantly distinguished rise and fall trend between points 2000 and 3000. This is the dot-com bubble that was mostly developed on NASDAQ and hit the worst on NASDAQ itself.

In the project we plan to compare the results we got on the data sets with time series of standard chaotic systems, white and red noise. The chaotic systems are the following :

1. Lorenz system :
   Lorenz system is standard system that exhibits deterministic chaos. The system is described as [16]

   $$\dot{x} = \sigma(y - x) \tag{2.1}$$
   $$\dot{y} = x(\rho - z) - y \tag{2.2}$$
   $$\dot{z} = xy - \beta z \tag{2.3}$$

   The parameters are chosen as $\sigma = 10$, $\rho = 28$, $\beta = 8/3$. So that the dynamics is chaotic
   The phase space plot of the Lorenz system with the above parameters is shown in fig.2.4. Since shortest of the data sets is of the size 6333 data points, we use 6333 x-variable values as our Lorenz data as shown in fig.2.5.

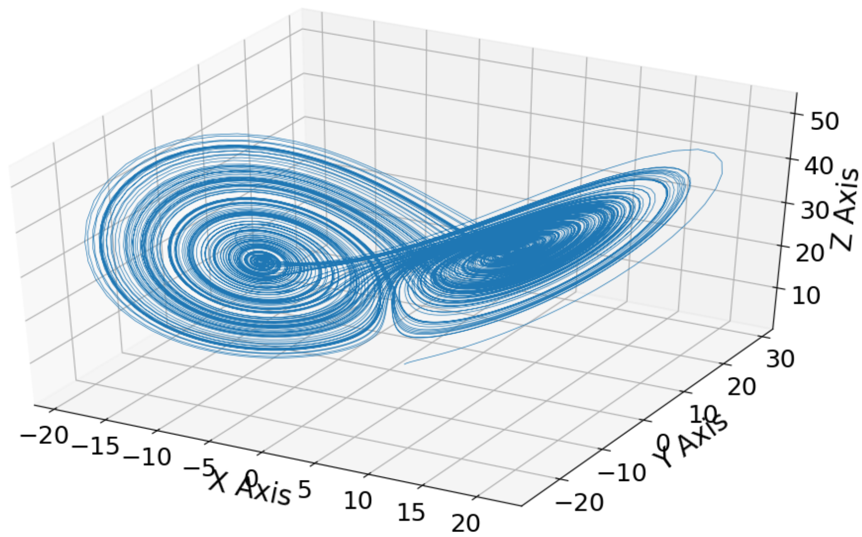

Figure 2.4: Phase space plot for Lorenz system

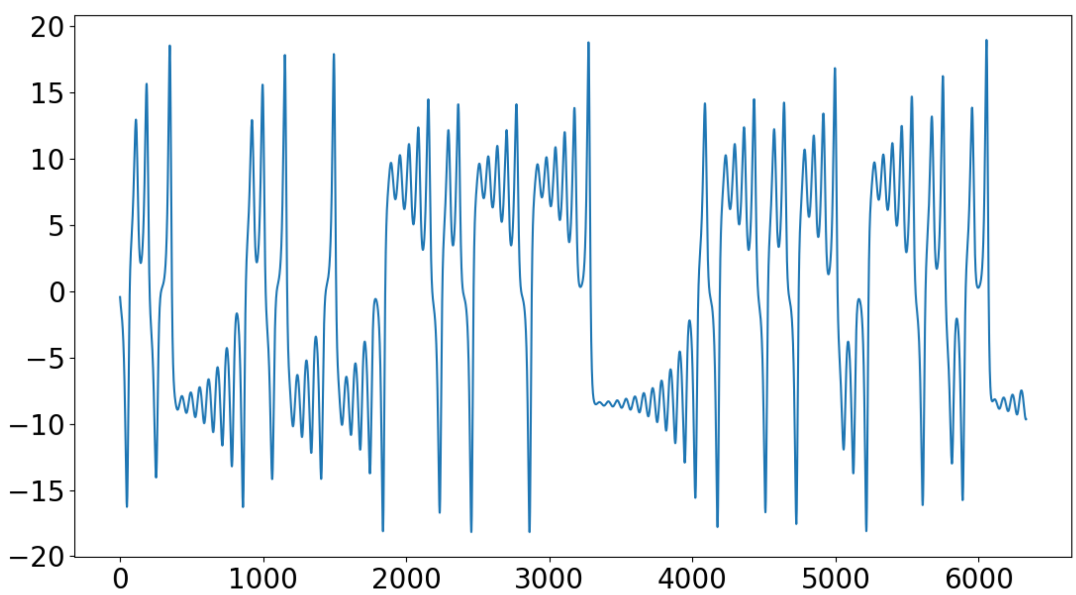

Figure 2.5: Time series for x variable of Lorenz system

2. Rössler system :

   Rössler system is another dynamical system that exhibits deterministic chaos. The system is described as [17]

$$\dot{x} = -y - z \tag{2.4}$$

$$\dot{y} = x + ay \tag{2.5}$$

$$\dot{z} = b + z(x - c) \tag{2.6}$$

   The parameters are chosen as $a = 0.2$, $b = 0.$, $c = 5.7$ such that the dynamics is chaotic as shown in fig.2.6.

   Since shortest of the data sets is of the size 6333 data points, we use 6333 x-variable values as our Lorenz data as shown in fig.2.7.

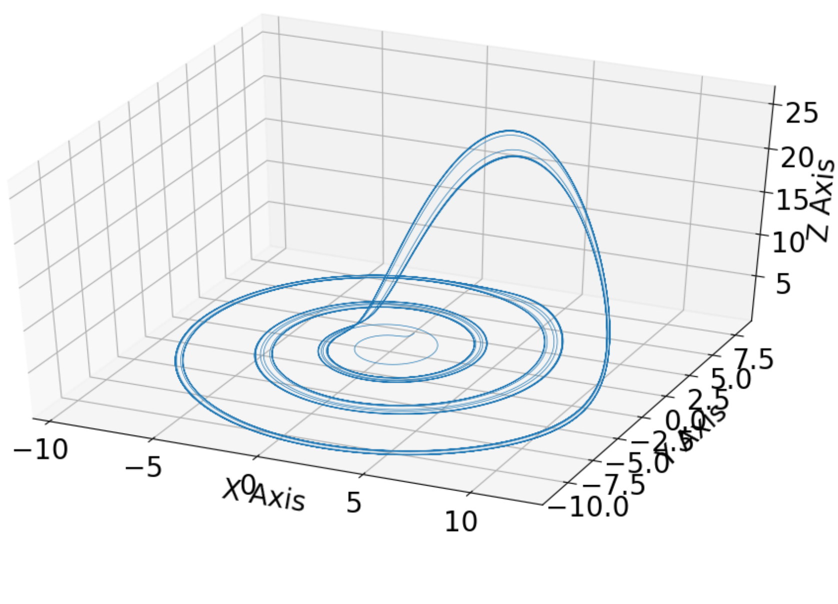

Figure 2.6: Phase space plot for Rössler system

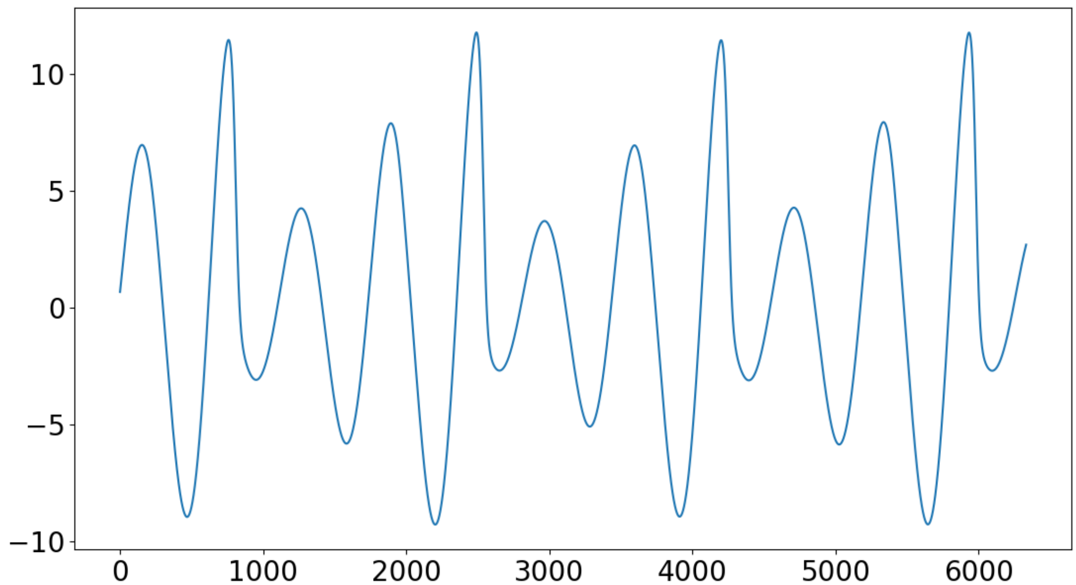

Figure 2.7: time series for x-variable of Rössler system

Red noise and white noise data :

It is possible that even data with nonlinear dynamics may be contaminated with noise or behave like red noise. Hence we also generate red noise and white noise and study their characteristic as well.

1. White noise : white noise is basically a random signal. In the project we have used Gaussian white noise, which is obtained using drawing random samples from a normal (Gaussian) distribution. We have used python package "numpy.random" to generate white noise. Data thus generated is shown in fig.2.8. as a time series.

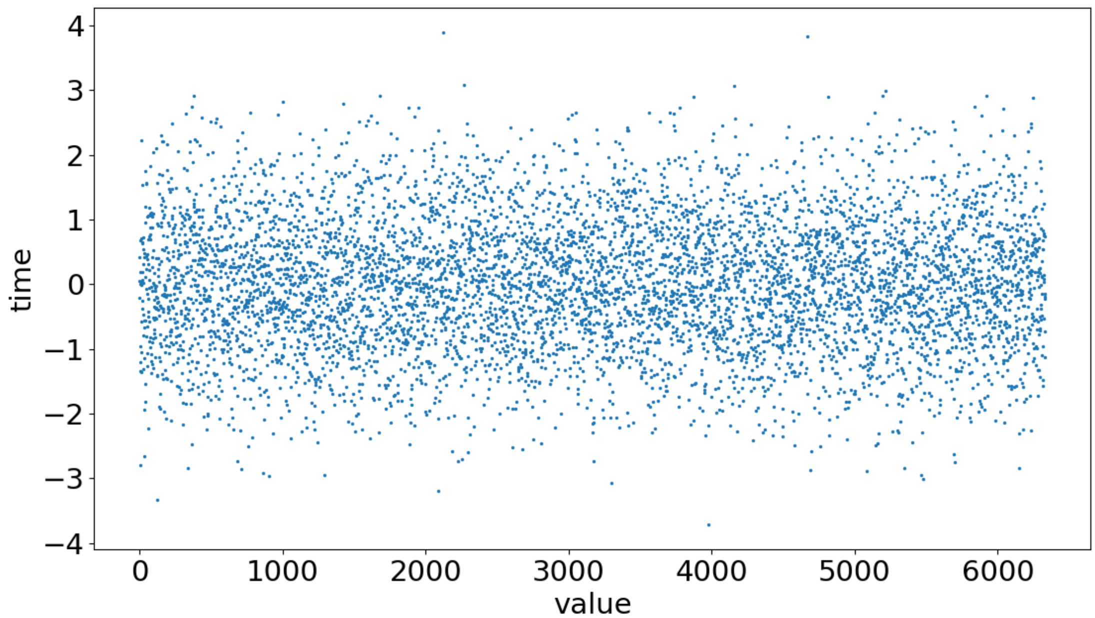

Figure 2.8: white noise data

2. Red noise : Red noise or Brownian noise is a signal produced by Brownian motion [42]. To generate red noise, we first generate power spectrum that goes as $\frac{1}{f^2}$. Then we randomize the phases and take inverse Fourier transform. A typical red noise data generated is shown in fig.2.9.

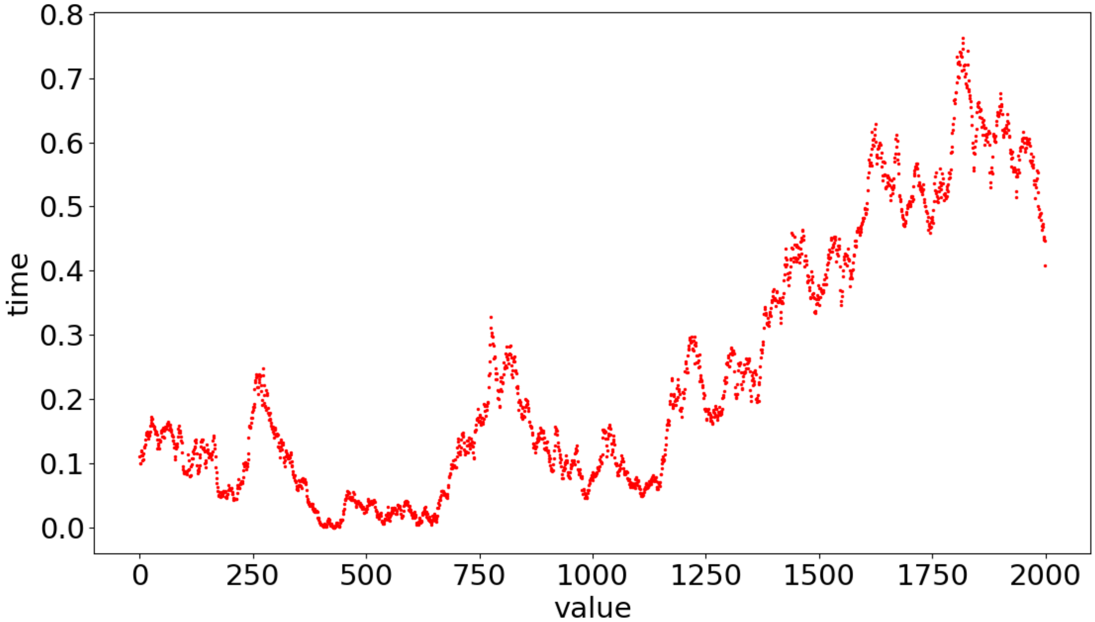

Figure 2.9: Red noise data

time series	Data size	Mean	standard deviation
Lorenz	6333	0.26	5.13
Rössler	6333	-0.25	8.26
white noise	6333	0.02	1
red noise	6333	-0.57	0.79
NYSE	6602	6521.77	2615.67
DJI	6601	9999.68	4309.48
GSPC	6602	1147.33	486.81
NASDAQ	6602	2274.16	1288.82
SSE	6660	1822.91	1045.41
Nikkei	6457	15082.94	4311.78
DAX	6641	5371.54	2685.17
CAC	6652	3693.84	1246.67
Sensex	6333	10508.19	8389.49
TSX	6591	9141.74	3761.30

Table 2.1: shows size of data, values of mean, standard deviation for standard chaotic systems, white noise, red noise and data. In next section we try to understand underlying periodic patterns in the data by studying its power spectrum.

2.4 Power Spectrum

Power spectral analysis is used to understand any periodic patterns in variations in a time series. Let x_n be nth value of the time series with total size of N. Then Discrete Fourier transform is given by [43]

$$F(n) = \sum_{k=0}^{N-1} x_n e^{-2\pi i n k/N}$$

Then power spectrum is given by,

$$P(n) = \frac{1}{N}|F(n)|^2$$

In this project we have used python package numpy.fft.fft which uses efficient Fast Fourier Transform (FFT) algorithm. [44]

We present the power spectra for the standard chaotic systems, noise and data sets.

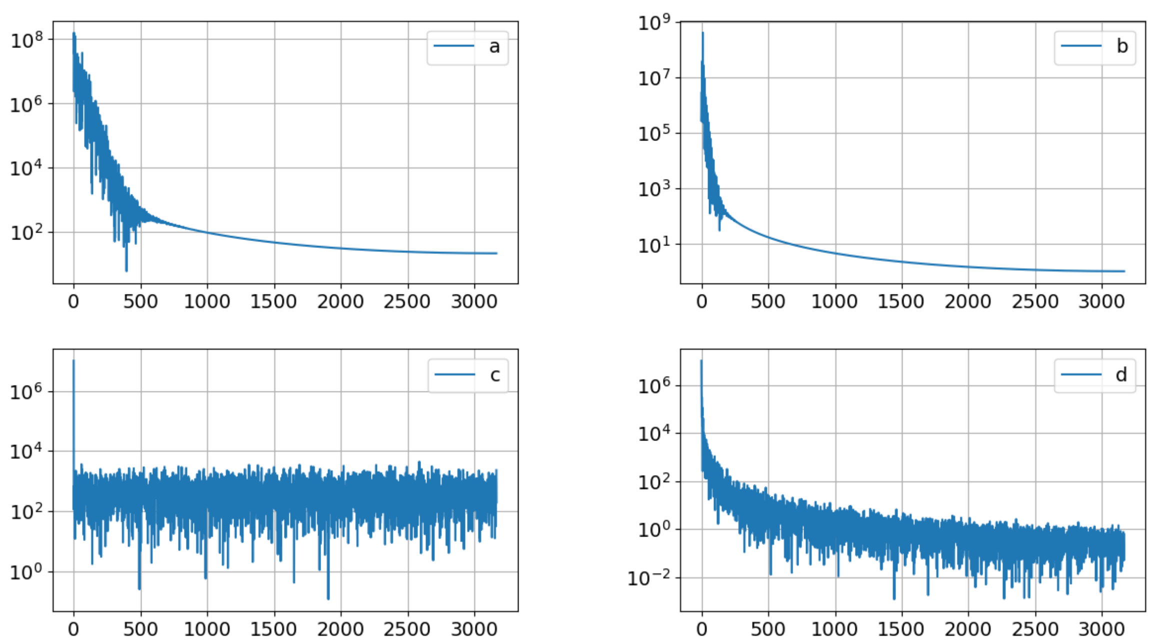

Figure 2.10: Power spectrum plots for data sets for Lorenz(a), Rössler(b), white noise(c) and red noise(d)

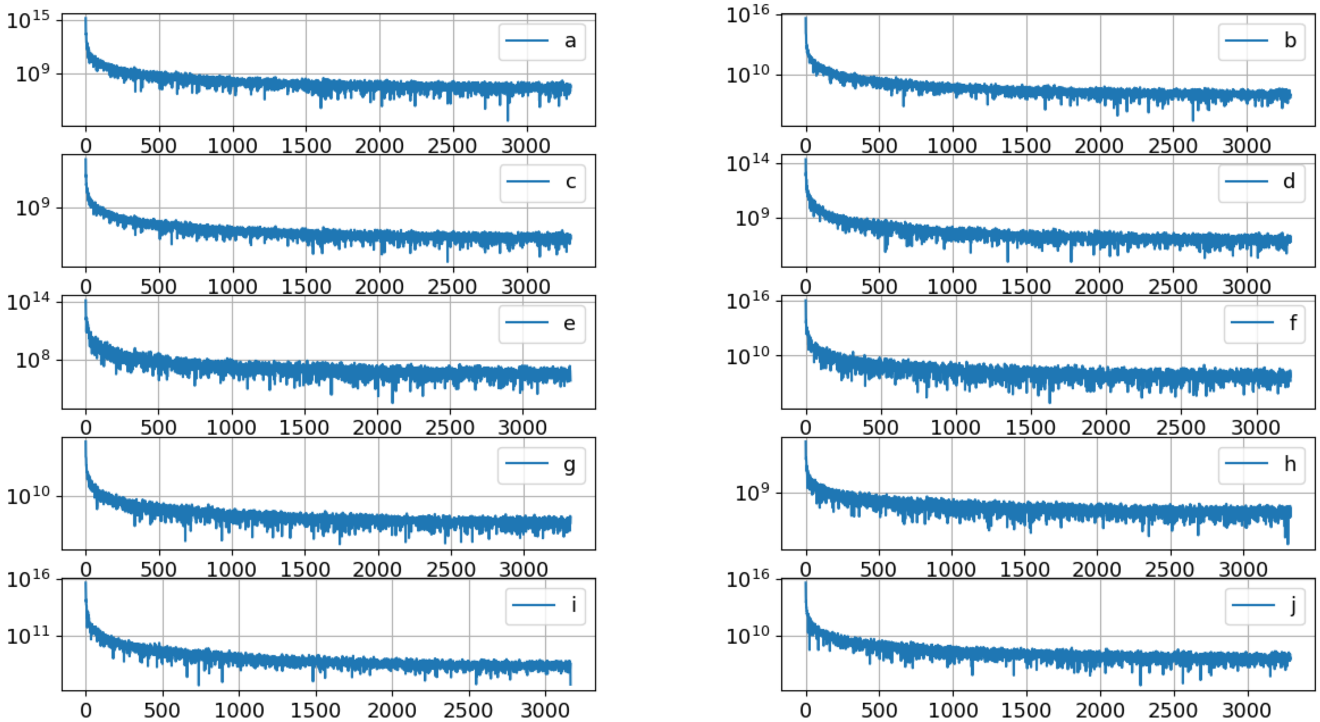

Figure 2.11: Power spectrum plots for data sets for NYSE(a), DJI(b), GSPC(c), NASDAQ(d), SSE(e), Nikkei(f), DAX(g), CAC(h), Sensex(i), TSX(j)

It can be observed from the power spectra that, the power spectra of data sets are similar to that of red noise. To investigate this further we plot a log-log plot of power spectrum which takes log of both power and frequency. We look at slope of this plot to understand the behaviour.

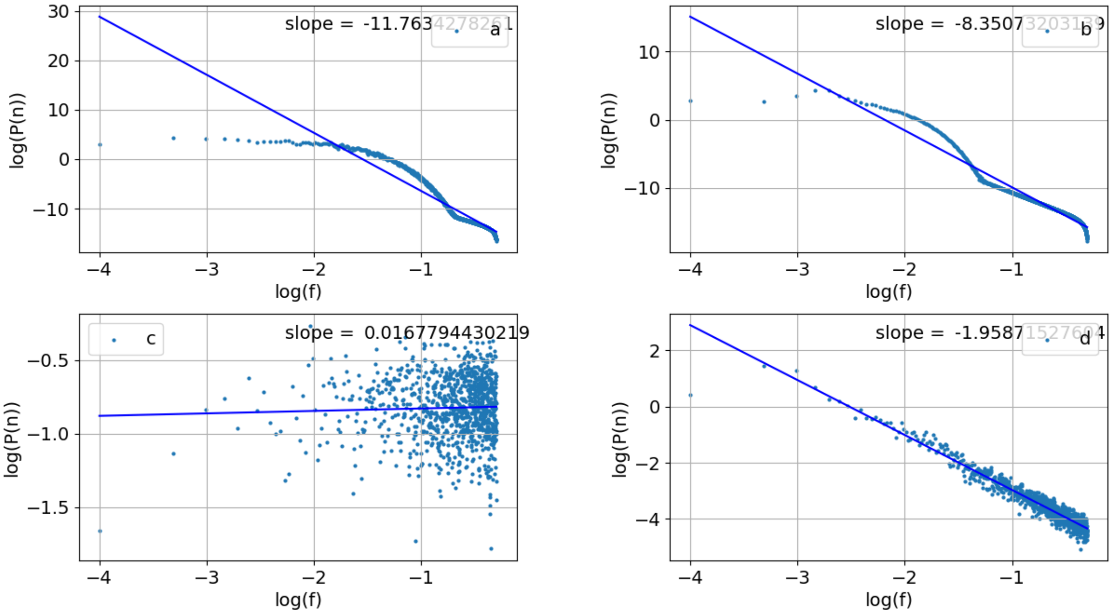

Figure 2.12: Log-log power spectrum plots for data sets for Lorenz(a), Rössler(b), white noise(c) and red noise(d)

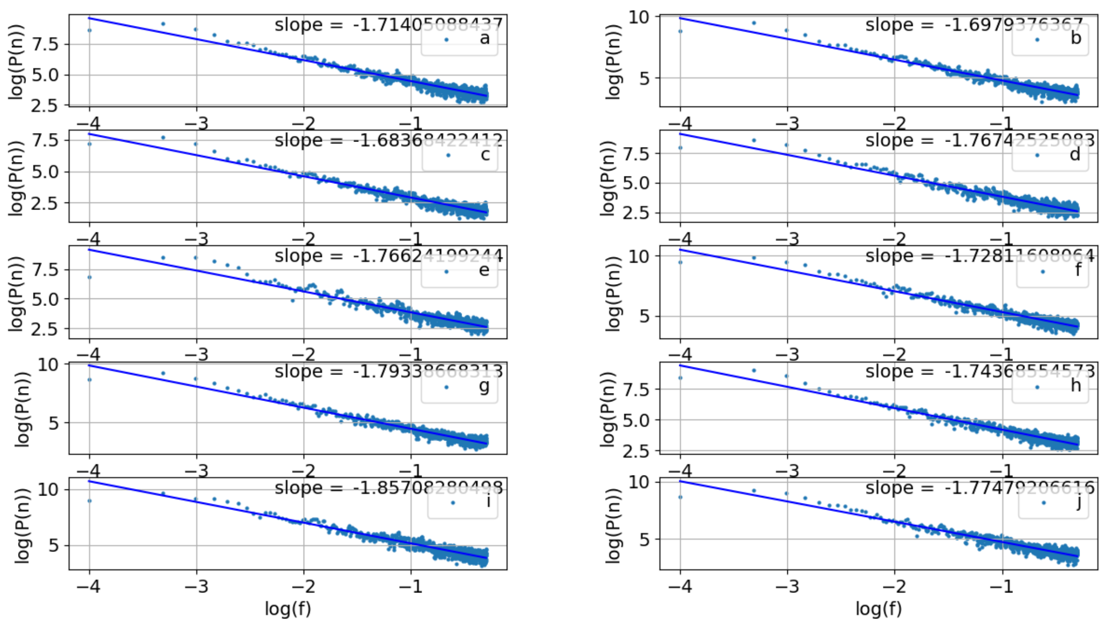

Figure 2.13: Log-log power spectrum plots for data sets for NYSE(a), DJI(b), GSPC(c), NASDAQ(d), SSE(e), Nikkei(f), DAX(g), CAC(h), Sensex(i), TSX(j)

data	slope of log-log plot	data	slope of log-log plot
Rössler	-11.76	NASDAQ	-1.77
Lorenz	-8.35	SSE	-1.77
White	0.02	Nikkei	-1.73
Red	-1.96	DAX	-1.80
NYSE	-1.71	CAC	-1.74
DJI	-1.70	Sensex	-1.86
GSPC	-1.68	TSX	-1.77

Table 2.2: Table of slopes of log-log plots for standard chaotic systems, white noise, red noise and data sets

Fig:2.12 shows log-log power spectra for standard chaotic systems, white noise and red noise. Fig:2.12 shows log-log power spectra for all the data sets. It is important to note that the slopes obtained for log-log power spectrum plots for data are in the range -1.69 to -1.85 which is quite near to that of red noise which is around -1.95. So we get an indication that the underlying dynamics might have some red noise like character or nonlinear dynamics contained with red noise. This is the first indication of red noise like behaviour in financial market time series. In later chapters we find that power spectrum is not the only measures that indicates red noise like behaviour. Still the values of slopes for data and red noise are not exactly same, which means that the dynamics is not exactly like red noise but dominated by red noise. Next we deal with the long term upward variation in the data by the method of polynomial detrending.

2.5 Detrending methods

It can be seen from the time series plots that the data has long term upward variations. Since the top companies whose weighted average is used to calculate the index price do well over time, the index moves up over time. But for our purpose we need to remove these long term trend without loosing the underlying nonlinear dynamics in the system. To achieve that we do polynomial detrending of the time series.
Algorithm for polynomial detrending :

- Polynomial fitting : This step involves fitting polynomial to a time series by least square method. let y = $(y_0, y_1, ...y_n)$ be the data. We want to fit it with n degree polynomial,

$$y_i \implies P_n(x_i) = a_0 + a_1 x_i + a_2 x_i^2 + ... + a_n x_i^n ....(1)$$

Here n is degree of polynomial and x_i is i^{th} component of vector (1,2,3.....N), where is N is size of Y.
Equation (1) can be written for multiple component as :

$$y_0 \implies a_0 + a_1 x_0 + a_2 x_0^2 + ... + a_n x_0^n ...(2)$$

$$y_i \implies a_0 + a_1 x_i + a_2 x_i^2 + ... + a_n x_i^n ....(3)$$

$$y_n \implies a_0 + a_1 x_n + a_2 x_n^2 + ... + a_n x_n^n ...(4)$$

Equations (2),(3),(4) can be written in matrix form as

$$V\vec{a} = \vec{y}$$

Here V is Vandermonde Matrix, which is a non singular matrix. So, applying V^{-1} on $\vec{y}$ gives $\vec{a}$. Also solution is unique.

- Goodness of fit(R2) : If $\bar{y}$ is mean of observed data,

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{n} y_i$$

Let SS_{tot} be total sum of squares

$$SS_{tot} = \sum_{i=1}^{N} (y_i - \bar{y})^2$$

Let $SS_{res}(n)$ be residual sum of squares

$$SS_{res}(n) = \sum_{i=1}^{N} (y_i - P_n(x_i))^2$$

The goodness of fit ($R^2(n)$) is defined as

$$R^2(n) = 1 - \frac{SS_{res}(n)}{SS_{tot}}$$

We divided each dataset into sections of 1000 points plotted $R^2(n)$ vs n, where n is from 1 to 31 and see where $R^2(n)$ saturates.
We took the degree at which $R^2(n)$ saturates for different sections and the maximum of those values of degrees is taken as the degree of polynomial with which the whole the data set will be detrended.

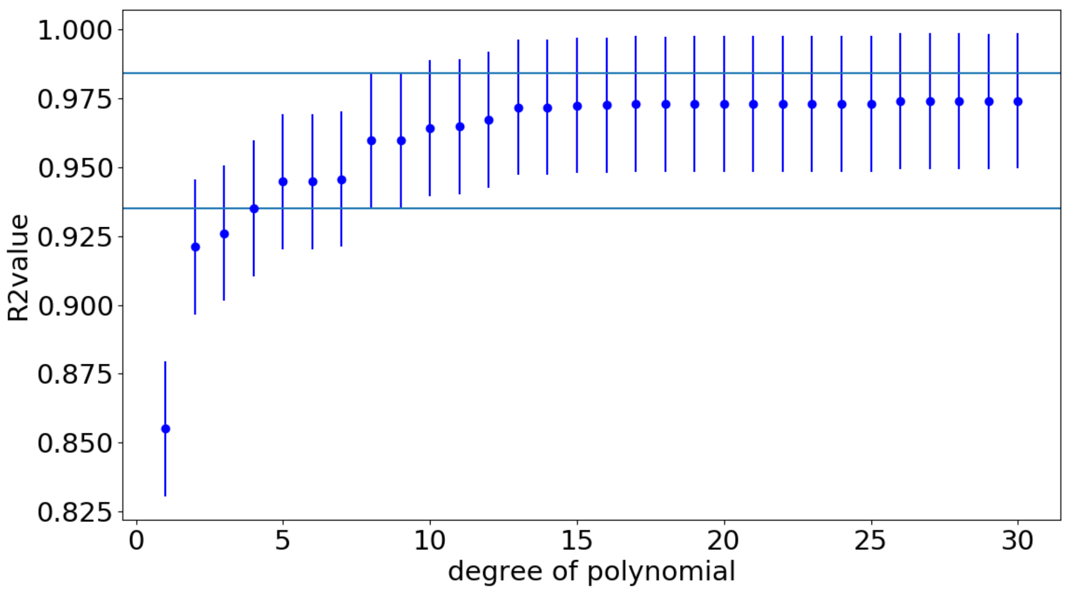

Figure 2.14: Saturation of R2 vs degree of polynomial plot for first 1000 points section of NYSE.

Error bars are from standard deviation of 30 $R^2(n)$ values. $R^2(n)$ saturates at degree 8. This 1000 points window is moved by 200 and the degree at which the R2 saturates is calculated for every window. The maximum of these degrees is chosen as the degree of polynomial with which the whole data set is detrended. This degree is different for different data sets, typically ranging between 6 to 15.

- Final detrending : Let y(t) be the time series. Let ypol(t) be the polynomial fit of the degree equal to maximum of the saturation degree. Then ydet(t) is the detrended time series we will work on from here on. ydet(t) is given by.

$$ydet(t) = y(t) - ypol(t)$$

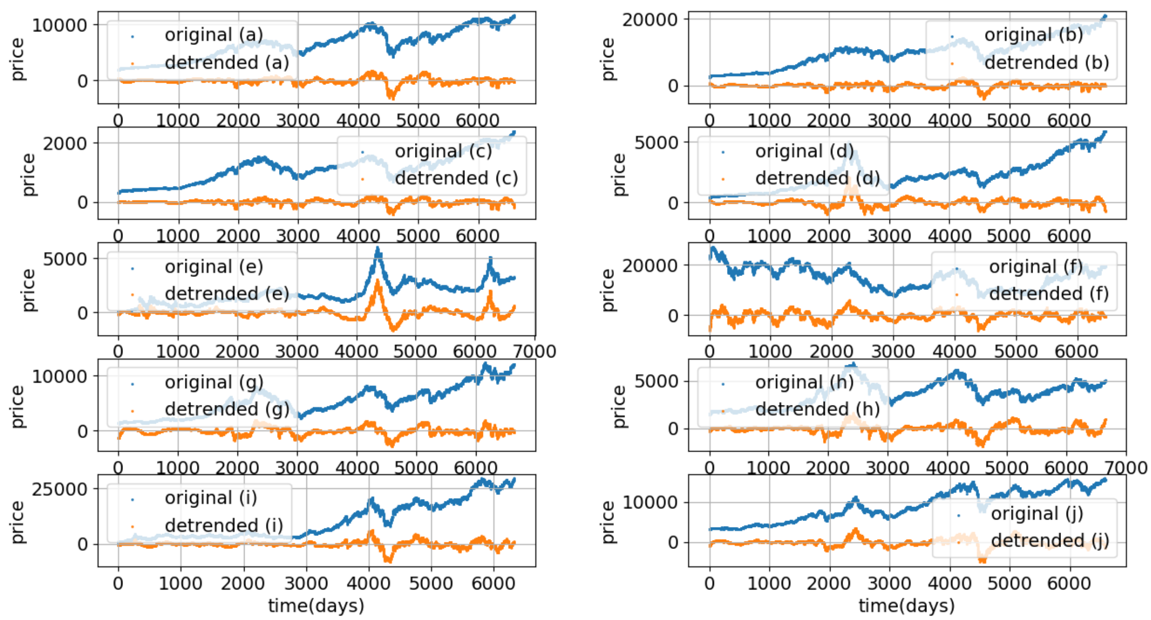

Figure 2.15: Original and detrended plots for data sets for NYSE(a), DJI(b), GSPC(c), NASDAQ(d), SSE(e), Nikkei(f), DAX(g), CAC(h), Sensex(i), TSX(j)

It is clear that detrending has not washed out the intricate variations, but only removed the general trend.

- Features of the detrended timeserieses :

  1. All detrended time serieses retain the crash of 2008. It can be observed between 4000th and 5000th points.

  2. Dot com bubble of 2000-2002 which primarily built up in NASDAQ and hurt the worst in NASDAQ. This bubble is retained in NASDAQ detrended time series. It can be observed between 2000th and 3000th points.

  3. Chinese stock market turbulence of 2015-16 which can be observed between 4000th and 5000th points is retained in the detrended data.

  These three important features that are retained in the detrended data indicate that even if the long term variations in the time series are removed, the underlying dynamics that we are interested in is not lost because of detrending.

In this section, we present the statistical measures of the data such as means, Fourier analysis, etc. In the next section, we describe the method of embedding and phase space reconstruction of dynamics from data.

Chapter 3

Embedding and phase space reconstruction

In this chapter we discuss the method of embedding and phase space reconstruction from the data sets. We also discuss criteria for choosing parameters for embedding and then analyze the nature of the embedded trajectory.

3.1 Uniform Deviate

It can be seen from time series plots that the range of prices vary widely from one market to another. To deal with this problem we take uniform deviate of the time series. If n_i is the number of points in the time series that are less than or equal to y_i and N is the size of the time series.
Uniform deviate of a point y_i time series :

$$y_i \implies \frac{n_i}{N}$$

So uniform deviate transforms the whole time series in range of 0 to 1. Despite transforming the time series, uniform deviate does not change underlying dynamics of time series [25].
Following are the uniform deviate time series plots for all the data sets.

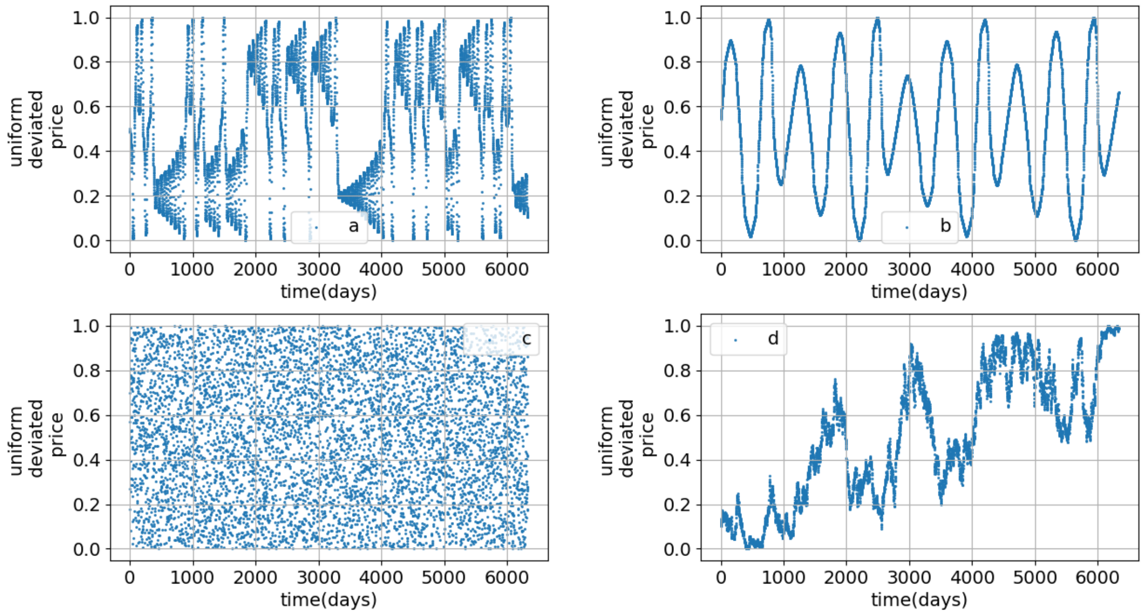

Figure 3.1: Uniform deviated time series plots for data sets Lorenz(a), Rössler(b),white noise(c), red noise(d)

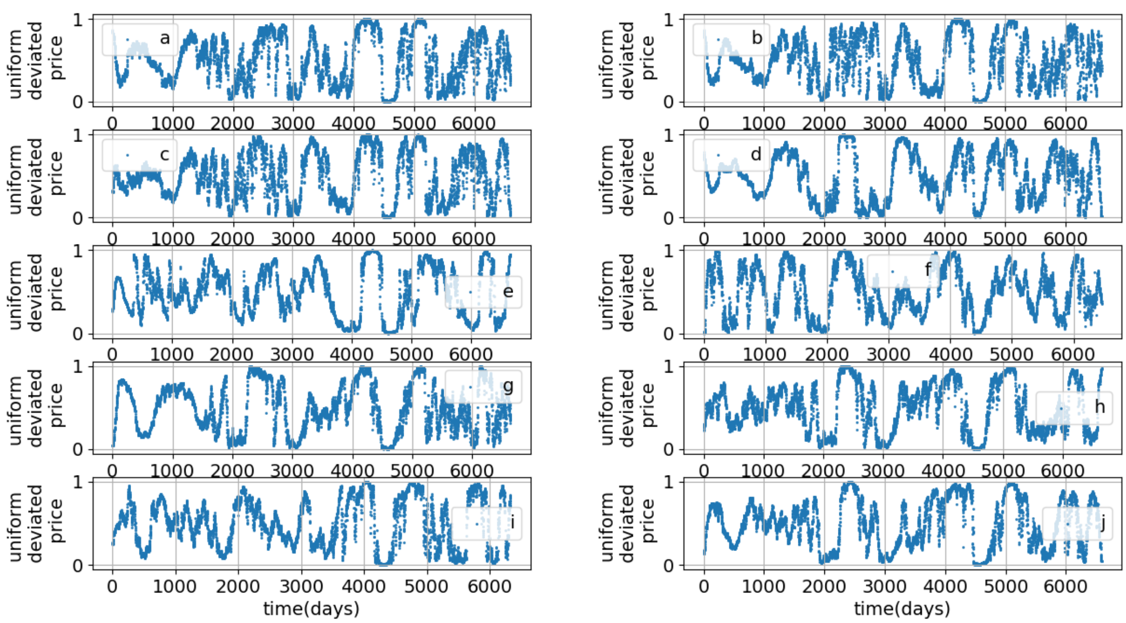

Figure 3.2: Uniform deviated time series plots for data sets for NYSE(a), DJI(b), GSPC(c), NASDAQ(d), SSE(e), Nikkei(f), DAX(g), CAC(h), Sensex(i), TSX(j)

In the next section we discuss the process of embedding and phase space reconstruction.

3.2 Embedding and phase space reconstruction

Every bound system when studied in phase space, shows phenomenon of recurrence which means that the system keeps coming back to the same neighborhood, given the system is allowed to evolve for enough time. The pattern in these recurrences involve important information about the underlying dynamics of the system. Recurrence network is a framework that involves understanding the underlying dynamics of the system by studying recurrence in phase space with the use of phase space reconstruction. [45] [46]. For this we have to first set the trajectory in the phase space from the data. This is done by Taken's embedding theorem. Consider dynamical system having states $x(t) \in \mathbb{R}^N$ that evolves through a differential equation,

$$\dot{x} = \Psi(x)$$

Here, Ψ is a vector field such that, $\Psi : \mathbb{R}^N \to \mathbb{R}^N$. For this, we use following steps : Firstly M dimensional delay vectors are created from the time series by using time delay τ of the form

$$\vec{x}_i = [s(i), s(i+\tau), ...s(i+(M-1)\tau)]$$

This step is called phase space reconstruction [47]. For that, it is crucial to fix the values of embedding dimension M and time delay τ. Then the number of delay vectors created are

$$N(RN) = N - (M-1)\tau$$

The trajectory in the embedded space can then be plotted

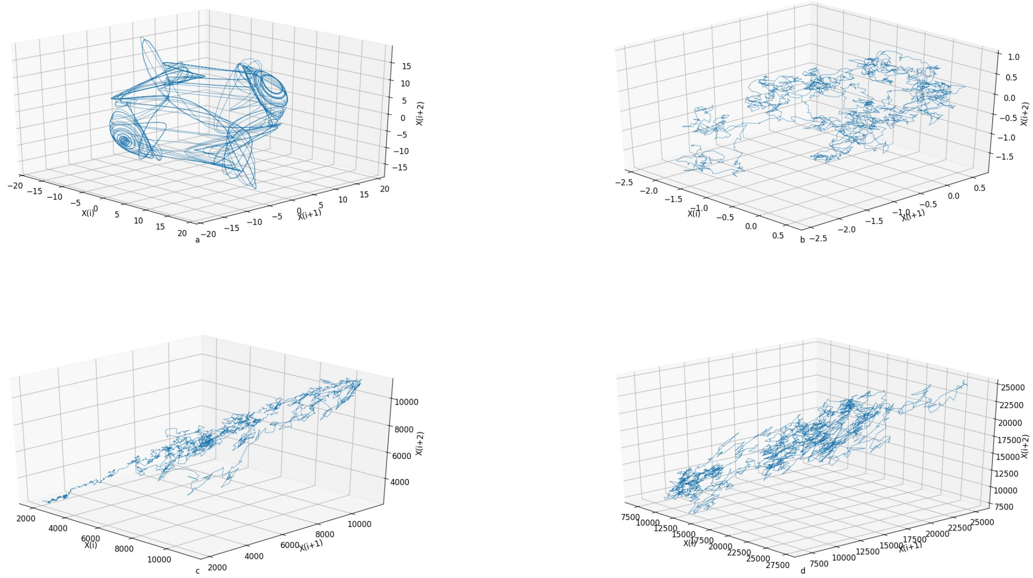

Figure 3.3: Phase space reconstruction plots for respective auto-correlation time delay and embedding dimension 3 for Lorenz(a), Red noise(b), NYSE(c) and Nikkei(d)

3.3 Choice of embedding parameters

It is clear from the embedding process that the choice of the parameters such as embedding dimension M,time delay τ of phase space.

3.3.1 Method of false nearest neighbours

For deciding embedding dimension, we use false nearest neighbours method [53]. The idea of false nearest neighbours comes from requirement that embedding dimension should be large enough that it gives fairly good resolution of the dynamics in the phase space.
Following are the steps to calculate embedding dimension by false nearest neighbors method,

- The main idea is that if two points are neighbors in lower dimensional space, they may not be neighbors in higher dimensional space. So, if such a case is found, the system is said to be more resolved than before. So, we start from embedding dimension 1 and go on increasing embedding dimension.

- Let $\vec{x_i}$ be a point in the time series and let $\vec{x_j}$ be its nearest neighbor in embedding space of dimension M.

- Let $\vec{x_i}$ be a point in the time series and let $\vec{x_j}$ be its nearest neighbor in embedding space of dimension M. We calculate the distance $||\vec{x_i} - \vec{x_j}||$. Then we compute the distance for dimension m+1 and the quantity R_i given by,

$$R_i = \frac{||\vec{x}_{i+1} - \vec{x}_{j+1}||}{||\vec{x_i} - \vec{x_j}||}$$

- If R_i exceeds some heuristic threshold R_t, then $\vec{x_i}$ is said to be false nearest neighbor. [53]

- When M is high enough that fraction of points for which $R_i > R_j$ is zero or significantly small, we fix that M as embedding dimension for the study.

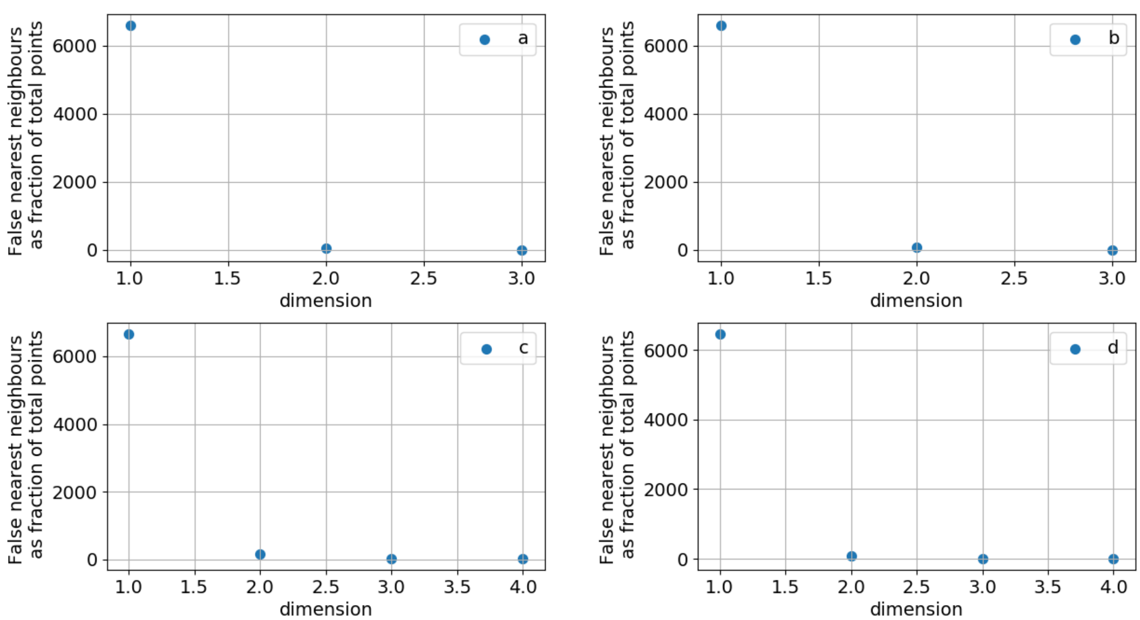

Figure 3.4: False nearest neighbors as function of dimension for NYSE(a), GSPC(b), SSE(c) and Nikkei(d).

Here we see that maximum embedding dimension turns out to be 4. The data set whose attractor is resolved at dimension of 3 will still have a well resolved attractor at embedding dimension 4. So to get all the data sets in same phase space restructuring scheme, we take embedding dimesnion to be 4 for all the data sets.

3.3.2 Time delay

For a time series x(t), auto-correlation function is given by,

$$C(\tau) = \frac{\langle (X_t - \mu) * (X_{t+\tau} - \mu) \rangle}{\sigma^2}$$

Here μ is mean and σ^2 is the variance of the data.
Time delay is the time steps taken by the time series for its auto correlation function to fall bellow 1/e.

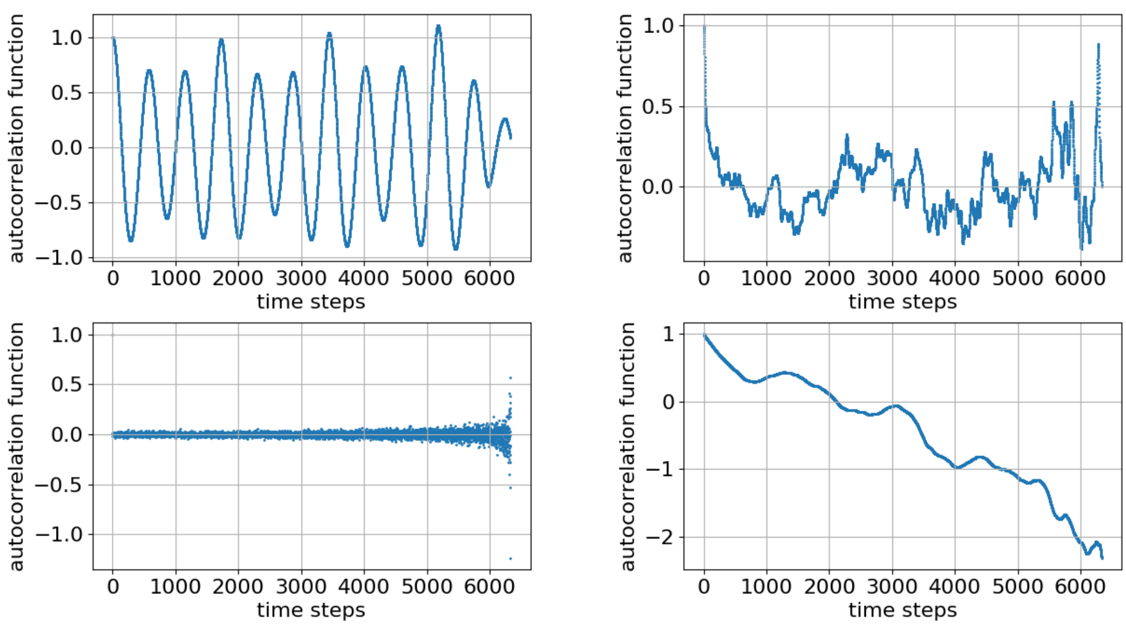

Figure 3.5: Auto-correlation function vs time plots for Rössler(a), Lorenz(b), white noise(c) and red noise(d).

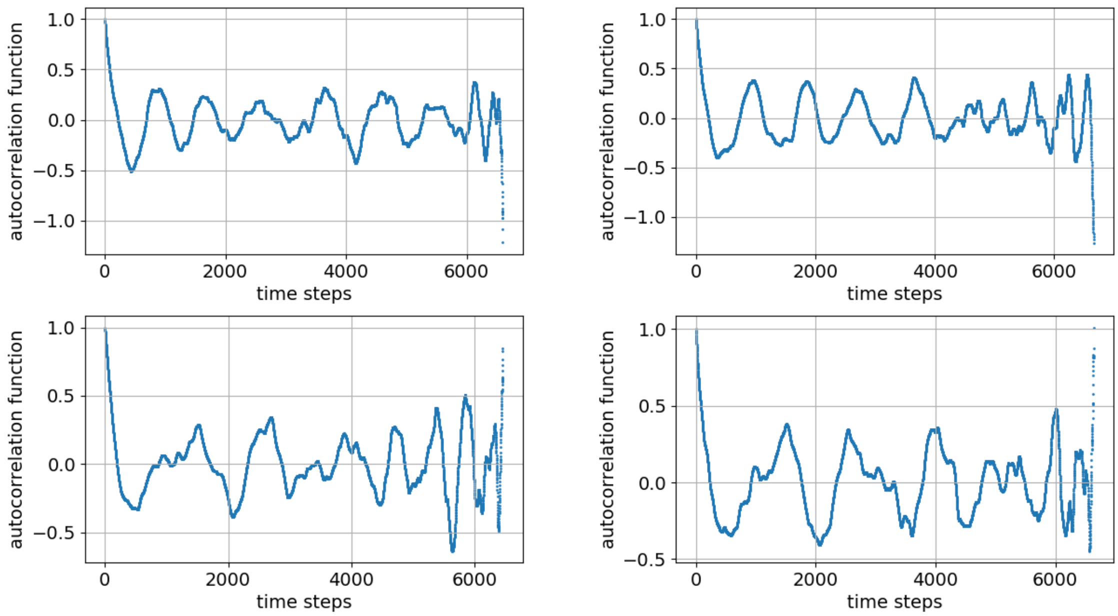

Figure 3.6: Auto-correlation function vs time plots for NYSE(a), GSPC(b), SSE(c) and Nikkei(d).

data	fnn dimension	time delay	data	fnn dimension	time delay
Rössler	3	108	NASDAQ	4	125
Lorenz	3	46	SSE	4	106
White	none	1	Nikkei	4	111
Red	none	608	DAX	4	116
NYC	3	104	CAC	3	134
DJI	3	95	Sensex	3	99
GSPC	3	99	TSE	3	124

Table 3.1: Table of embedding dimensions by false nearest neighbours method and time delay by auto-correlation time method.

The important thing to note is that data sets have time delay of around 100 but red noise has time delay of 608, also the auto-correlation function behaviour as a function of time steps is significantly different in case of red noise as compared to data sets. So, we conclude that even if there are several measures that indicate red noise like behaviour in financial time series, in auto-correlation function sense data is quite different from red noise meaning that there are features in the dynamical system that are different from those of red noise. White noise given time delay of 1 which is expected. We can conclude that data is deterministic with added red noise.

3.3.3 Correlation dimension and surrogates analysis

Surrogate testing is a test for detecting determinism in underlying dynamics for a data set. Surrogates are generated by shuffling time series, keeping Fourier amplitudes and distribution of values constant. We generate 10 surrogates from every data set both original and detrended by Amplitude adjusted Fourier transform (AAFT) method. Then we compute a discriminating measure on both data and surrogates. If the measures happen to be significantly different for data and surrogates then underlying non linearity is confirmed [54].
The discriminating measure used in our study is called correlation dimension D_2. The relative number of points within a distance R from a particular point i in the time series is

$$p_i(R) = \lim_{N_\nu \to \infty} \frac{1}{N_\nu} \sum_{i=1, i \neq j}^{N_\nu} \Theta(R - |\vec{x_i} - \vec{x_j}|)$$

Here, N_ν is the total number of vectors in the embedding(reconstructed) space.
Correlation sum is defined as

$$C_m(R) = \frac{1}{N_c} \sum_{i}^{N_c} p_i(R)$$

Correlation dimension is given by

$$D_2 = \lim_{R \to 0} \frac{dlogC_m(R)}{dlog(R)}$$

Correlation dimension is dependent upon embedding dimension. We plot correlation dimension vs embedding dimension for both data and surrogates and see if they differ significantly. Generally surrogates based D_2 correlation dimension study is done on original data set before any transformation applied on it, we have done the same.

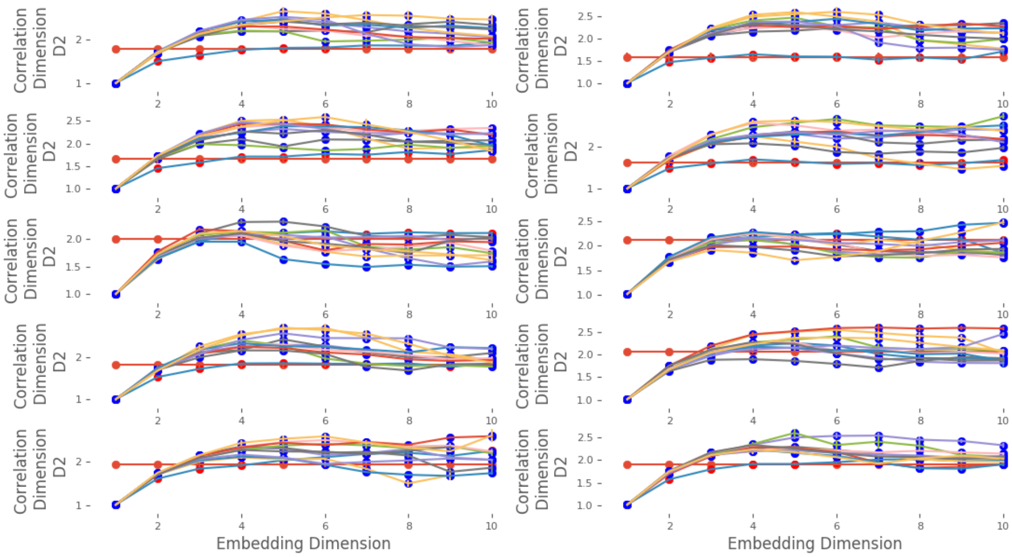

Figure 3.7: Correlation dimension vs m for data sets for NYSE(a), DJI(b), GSPC(c), NASDAQ(d), SSE(e), Nikkei(f), DAX(g), CAC(h), Sensex(i), TSX(j)

Figures 3.11, 3.12, 3.13 show D_2 correlation dimension vs. embedding dimension plots for all the time series before detrending or uniform deviate.In the figures,red scatter dots indicate D_2 values for original data sets and blue ones indicate D_2 measures for surrogates. The horizontal red line indicates D_2 saturation value. We have used "Nonlinear time series analysis tools ", developed by K. P. Harikrishnan [55]. 7 out of 10 data sets show significant deviation from surrogates. This indicates the data has underlying dynamics. More importantly, it is important to note at this point that American stock exchanges NYSE, DJI, GSPC and NASDAQ are showing significant deviation from surrogates indicating underlying dynamics, since these are the stock markets that we are mainly interested in since these are stock exchanges on which the housing bubble built up.

In this chapter, we presented the embedding technique to reconstruct the phase space trajectory from the data. The required dimension M and delay τ are computed. Surrogate analysis using D_2 as discriminating measure is also presented for all the data sets.
In the next chapter, we study how recurrence network can be constructed from the embedded attractor.

Chapter 4

Recurrence Networks from data

From the phase space trajectories obtained by embedding data in a properly chosen phase space, we now construct networks that reveal the recurrence patterns in the dynamics.

4.1 Recurrence plots and recurrence quantification analysis(RQA)

Recurrence plot is a visual tool for analysis of recurrence based dynamics of a system from its time series [56].
We define Heaviside function $\Theta(x)$,which is given by,

$$\Theta(x) = 0, x < 0 \tag{4.1}$$

$$1, x > 0 \tag{4.2}$$

Then recurrence matrix $R_{i,j}$ is given by,

$$R_{i,j}(\epsilon) = \Theta(\epsilon - ||x_i - x_j||)$$

Here, since we are not embedding, x_i and x_j are just time series points and $\Theta(x)$ Recurrence plot is graphical representation of this recurrence matrix.

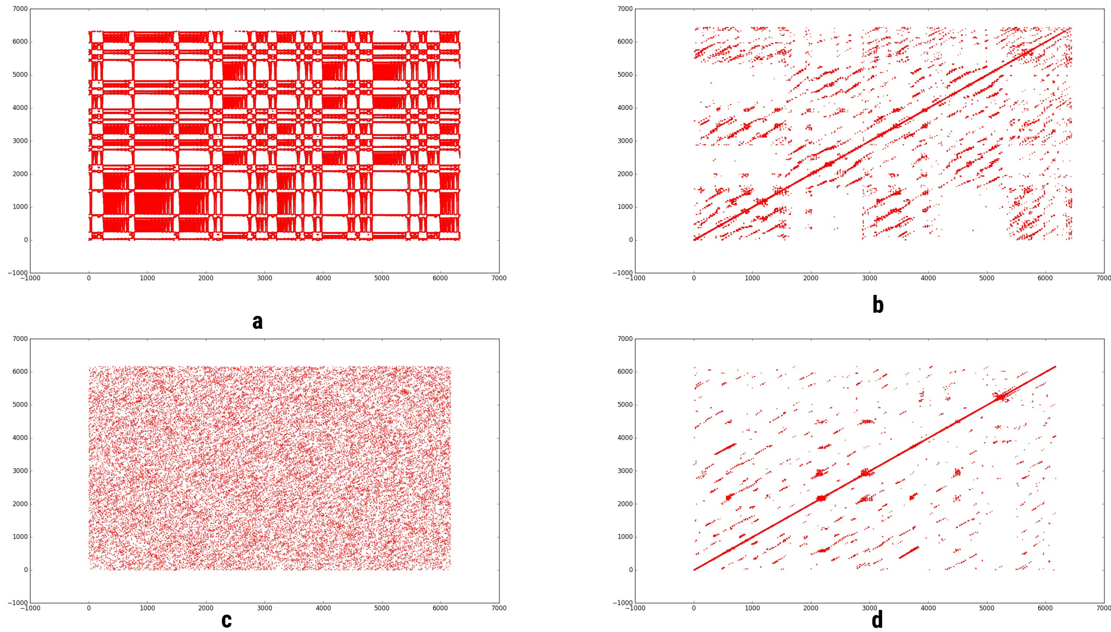

Figure 4.1: Recurrence plots for Lorenz(a), CAC(b), white noise(c), red noise(d).

Recurrence quantification analysis:

For recurrence quantification analysis we use different criteria for ϵ threshold for recurrence matrix. First we define recurrence rate RR as,

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}$$

We decide ϵ as the smallest ϵ that gives RR as 0.01.

Now we use Determinism measure DET for recurrence quantification analysis. DET is the percentage of recurrence points which form diagonal lines, and given by,

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)}$$

P(l) is the histogram of the lengths l of the diagonal lines.

Data	DET	Data	DET
NYSE	0.24	Nikkei	0.23
DJI	0.23	DAX	0.28
GSPC	0.22	CAC	0.26
NASDAQ	0.29	Sensex	0.28
SSE	0.37	TSE	0.31

Table 4.1: DET measures for data sets

4.2 Construction of recurrence networks

- Firstly M dimensional delay vectors are created from the time series by using time delay τ of the form

  $$\vec{x_i} = [s(i), s(i+\tau), ...s(i+(M-1)\tau)]$$

  These delay vectors are considered as nodes in the recurrence network to be constructed.

- In this case Recurrence matrix $R_{i,j}$ is given by,

  $$R_{i,j}(\epsilon) = \Theta(\epsilon - ||\vec{x_i} - \vec{x_j}||)$$

  Here, $\Theta(x)$ is the same Heaviside function discussed in the previous section. Then the adjacency matrix $A_{i,j}(\epsilon)$ is given by,

  $$A_{i,j}(\epsilon) = R_{i,j}(\epsilon) - \delta_{i,j}$$

  Here $\delta_{i,j}$ is the Kronecker delta.

- Finally edges are added according to this adjacency matrix $A_{i,j}(\epsilon)$ to construct the recurrence network from the data.

Some ambiguities in this seemingly simple algorithm have been raised in the past [48]. These ambiguities are mainly about ability of recurrence networks to capture the underlying chaotic attractor. If ϵ is too large, small scale properties of the attractor will not be captured and if it is too large we will have a fairly disconnected network which won't be enough to get some quantitative measures. While choice of ϵ does not have a specific set criterion that is well accepted in the literature, choice of M should be such that it sufficiently resolves the attractor [49–51].

Choice of ϵ dictates the number and distribution of edges in the network and hence the information we seek to extract from the network. ϵ is chosen as per the criterion that the ϵ for which the giant component in the network contains at least 95% of the total nodes. Here the giant component implies the components of network that has maximum nodes in it.

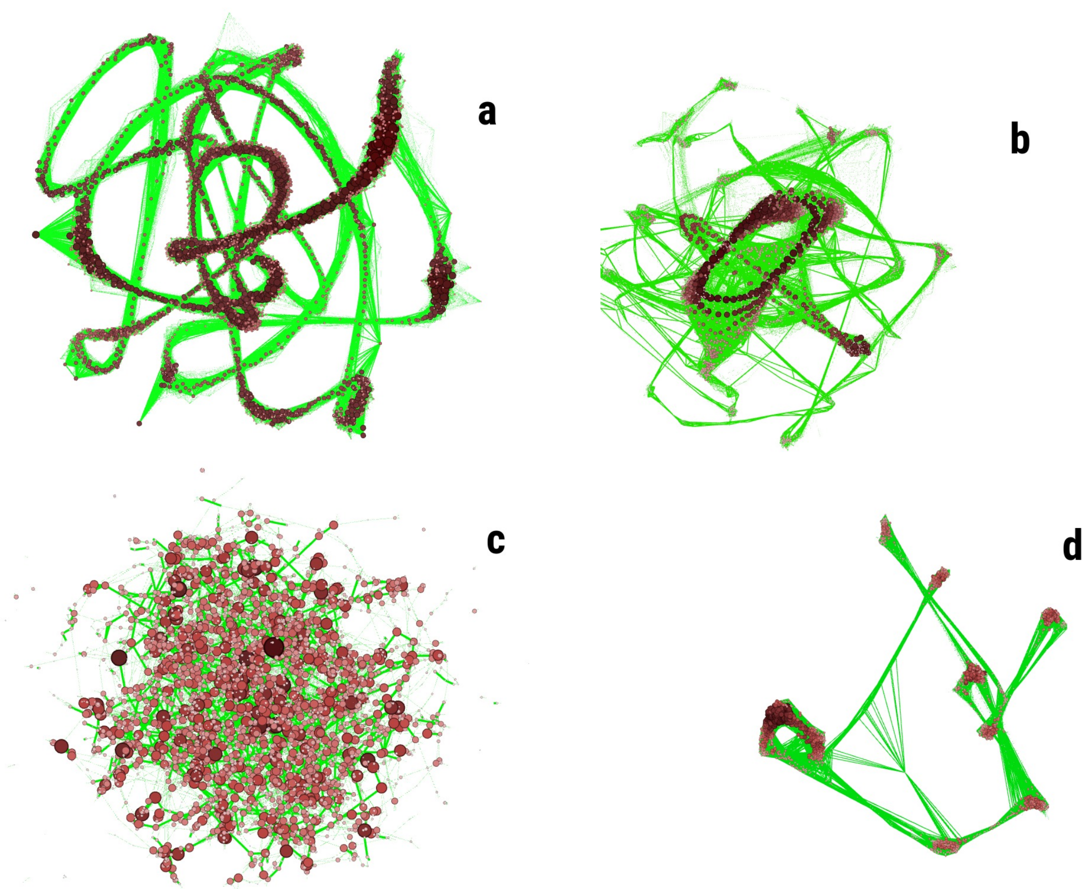

Figure 4.2: Recurrence networks for data sets of Rössler(a), Lorenz(a), white noise(c), red noise(d)

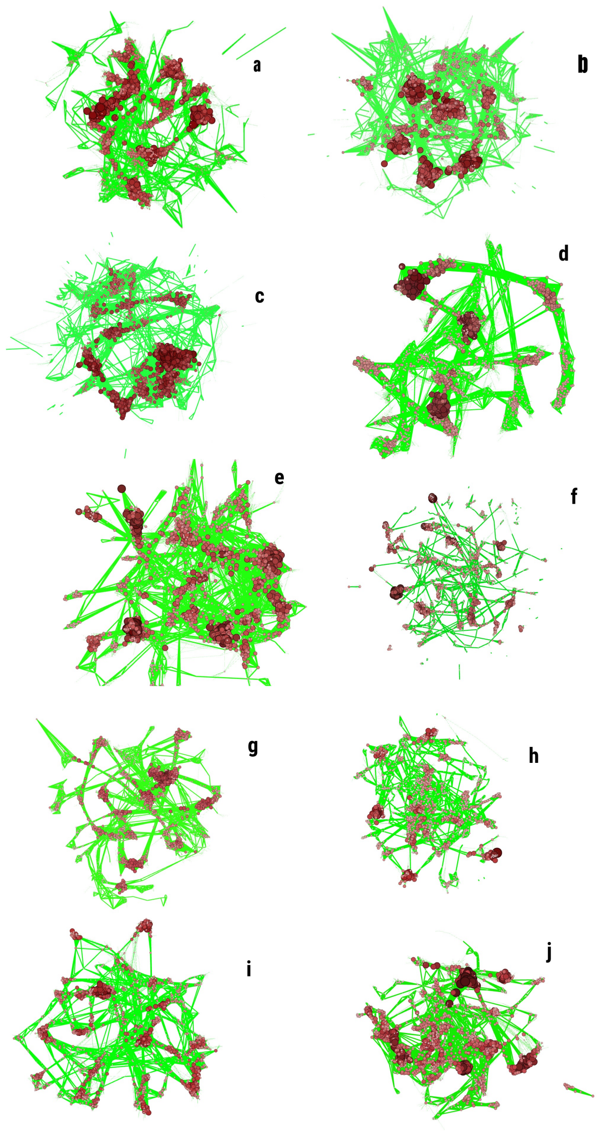

Figure 4.3: Recurrence networks for data sets of NYSE(a), DJI(b), GSPC(c), NASDAQ(d), SSE(e), Nikkei(f), DAX(g), CAC(h), Sensex(i), TSX(j)

Fig:4.1 and Fig:4.2 show typical recurrence networks constructed by ϵ-RN method for the data sets we have worked on. Since ϵ is chosen such that resulting recurrence network has a giant component containing at least 95% of the total nodes, network is quite dense as can be confirmed from the figure. In next chapter we characterize the underlying attractor by using standard network measures and compare with standard chaotic systems, white noise and red noise.

4.3 Network measures

In this chapter we discuss different measures done on networks that to quantify structural properties of a network.

4.3.1 Degree distribution

We define adjacency matrix A_{ij}, as a N $\times$ N matrix , where N is the number of nodes in the network. An element of A_{ij} is assigned value 1 if these is an edge between two nodes and 0 if two nodes are disconnected. Every node i in the network is assigned a degree k_i, where

$$k_i = \sum_j A_{ij}$$

Degree distribution is the probability distribution of degrees of nodes over the whole network. We plot the degree distribution of the constructed RNs from each data and then look at the variations with the embedding dimension.

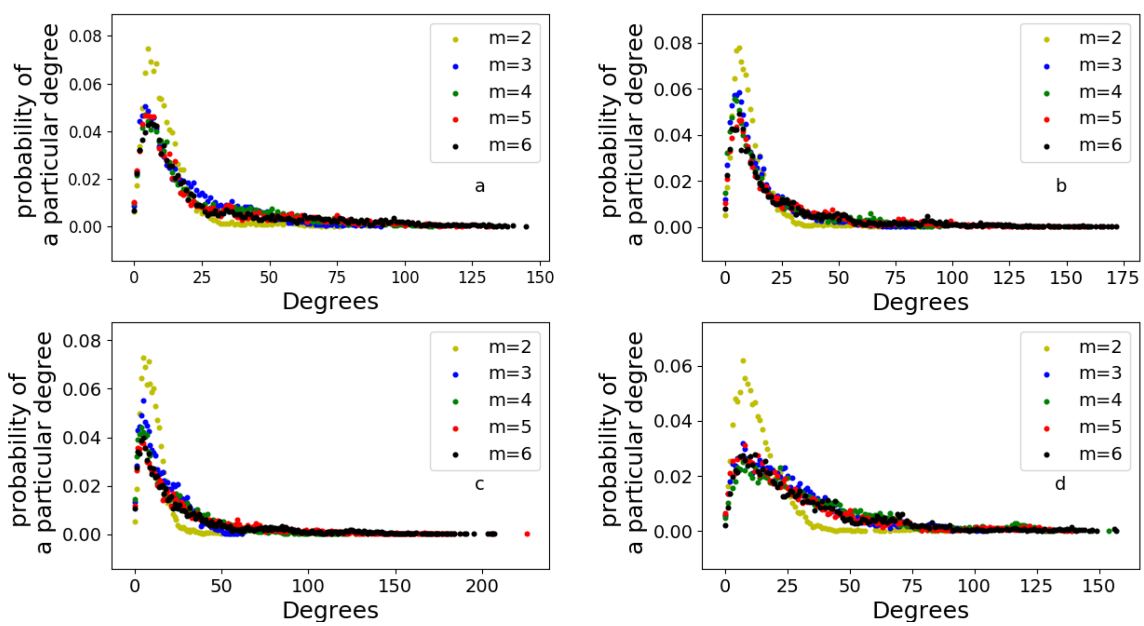

Figure 4.4: Degree distributions with different embedding dimensions for data sets of NYC(a), DJI(a), GSPC(c), NASDAQ(d).

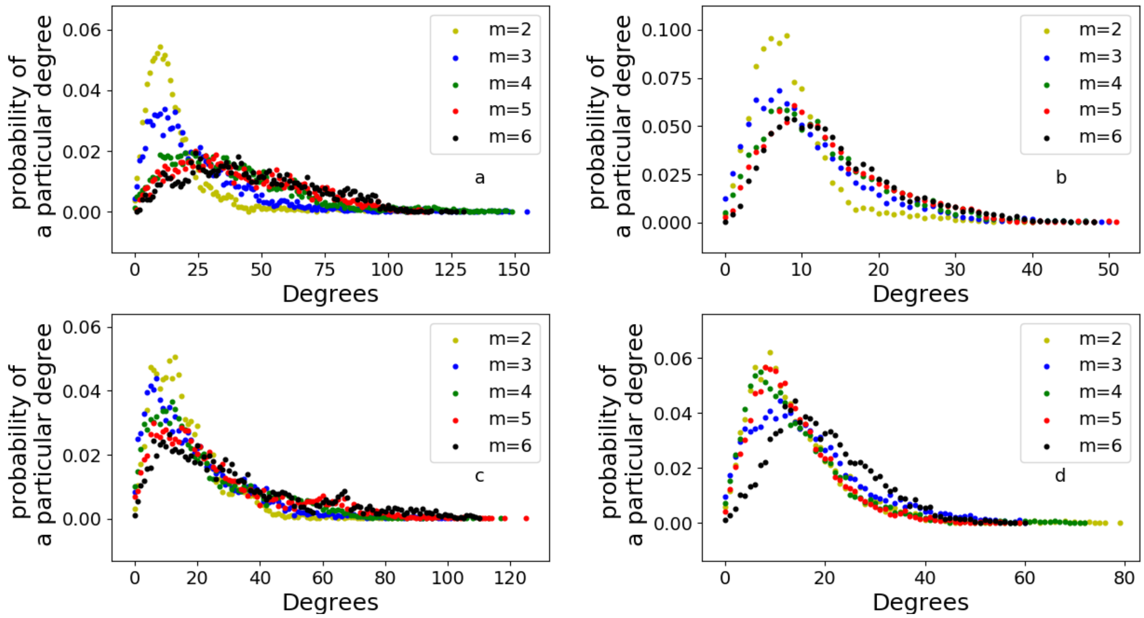

Figure 4.5: Degree distributions with different embedding dimensions for data sets of SSE(a), Nikkei(b), DAX(c), CAC(d).

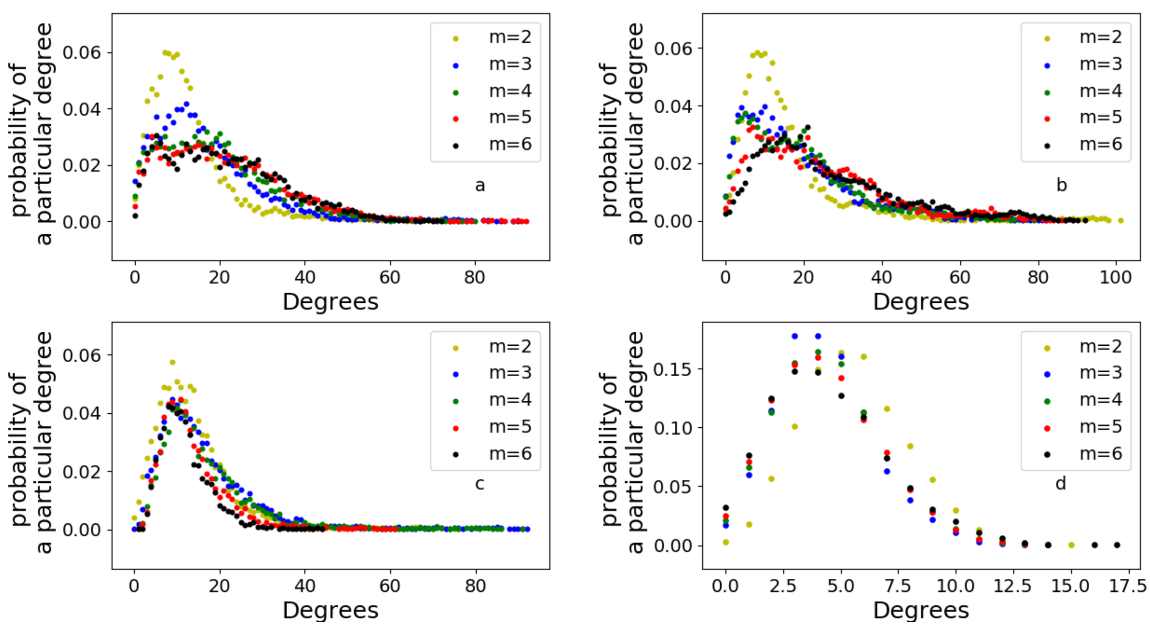

Figure 4.6: Degree distributions with different embedding dimensions for data sets of Sensex(a), TSE(b), red noise(c) and white noise(d).

There are two two important observations :

1. After embedding dimension of 4 degree distribution saturates for all the data sets and for red noise but not for white noise. This generally means that the data has underlying dynamics that is different from white noise.

2. Degree distribution pattern of data is considerably similar to red noise. This would again mean considerable red noise contamination.

4.3.2 Clustering coefficient and characteristic path length

To quantify characteristics of RNs of whole data we study clustering coefficient and characteristic path length. Local clustering for a node ν is given by,

$$C_\nu = \frac{\sum_{i,j} A_{\nu i} A_{ij} A_{j\nu}}{k_\nu (k_\nu - 1)}$$

Here $A_{\nu i}$ are elements of adjacency matrix.
Now, clustering coefficient CC is given by,

$$CC = \frac{1}{N} \sum_{\nu} C_\nu$$

Characteristic path length CPL is given by,

$$CPL = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{N-1} \sum_{i \neq j = 1}^{N-1} l_{ij}^s \right)$$

Here, N is total number of nodes, l_{ij}^s is the shortest distance between nodes i and j. Following graph shows cc-cpl plot for all the data sets, white noise, red noise and lorenz system.

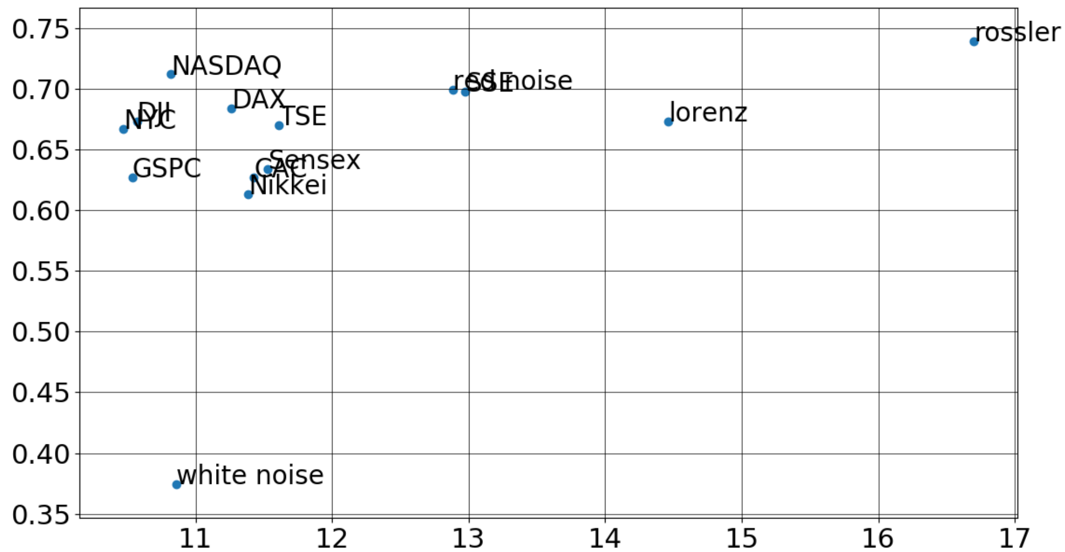

Figure 4.7: clustering coefficient vs. characteristic path length for all data sets, chaotic systems, white noise and red noise.

Here it can be observed that data sets lie nearer to red noise than white noise or Lorenz. From both degree distributions and cc-cpl plots it can be concluded that data sets show red noise like behaviour, which would imply dynamics with additive red noise.

4.4 Surrogates based analysis of network measures

In this section we take 10 surrogates for every data set and calculate ϵ threshold for every surrogate. Then we take maximum of those epsilons for one data set and construct recurrence networks for all the surrogates for that one ϵ. Then we calculate clustering coefficient and characteristic path length for surrogates, and take their mean and standard deviation. We plot measures for both data and surrogates in one plots to see if data shows significant deviation from surrogates.

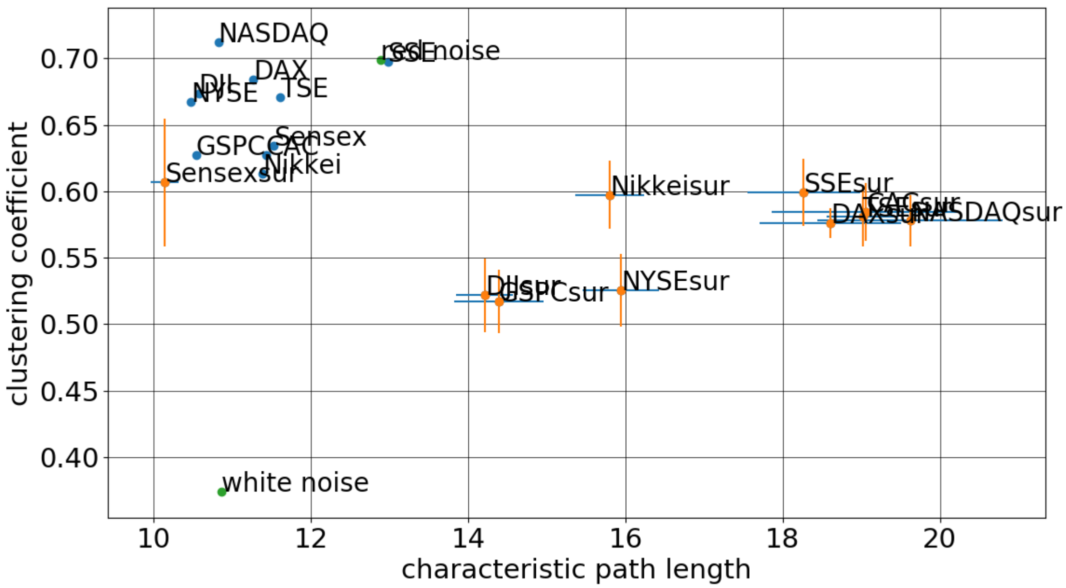

Figure 4.8: clustering coefficient vs. characteristic path length plots for both data and surrogates. Here NYC means original data and NYCsur means values for surrogates which also includes standard deviation, same representation is maintained for all the data sets.

There are two important observations from this plots :

1. Network measures for data sets and their surrogates are significantly different. This indicates the presence of underlying deterministic dynamics in the system.

2. It is clear that cc and cpl of surrogates are near to white noise

In this chapter we present the construction of recurrence network from data using carefully chosen parameters. The measures of the RNs constructed are then computed and compared with those of chaotic systems, noise and surrogates from data. In the next section, we try to investigate time evolution of recurrence network based measures by the method of windowed analysis.

Chapter 5

Windowed analysis for dynamical transitions

Dynamic recurrence networks are the recurrence networks derived from sections of the time series with a certain window with the window is movable by some time steps. The measures of network are then evaluated for each window. Thus the variations in measures characterizing the complexity of the system can be studied [57–59].

In our study, the window used is 2000 time steps long and we slide the window by 50 or 10 time steps. Then we calculate characteristic path length and clustering coefficient of those recurrence network to see if there is any change in the underlying dynamics of the time series in the window that we are dealing with. We have used a common representation scheme in this chapter, the three vertical lines indicate the time interval of the 2008 financial crisis. The first red line indicates start of 2006, black line indicates September 2008 which includes collapse of Lehman Brothers, and the second red line indicates December 2009. The horizontal dashed lines are there to indicate confidence interval. Middle yellow line indicates mean of the statistic, and the green lines above and below indicate the confidence interval. In this chapter, the main interest is the NYSE analysis because it is the main stock exchange on which the bubble was built and it is the biggest stock market in the world in terms of volumes traded so it was big enough to create the crisis. DJI and GSPC also need to be observed as they are also US based stock markets and had a lot of exposure to housing market.

5.1 Clustering coefficient

Following plots show clustering coefficients plots for all the data sets

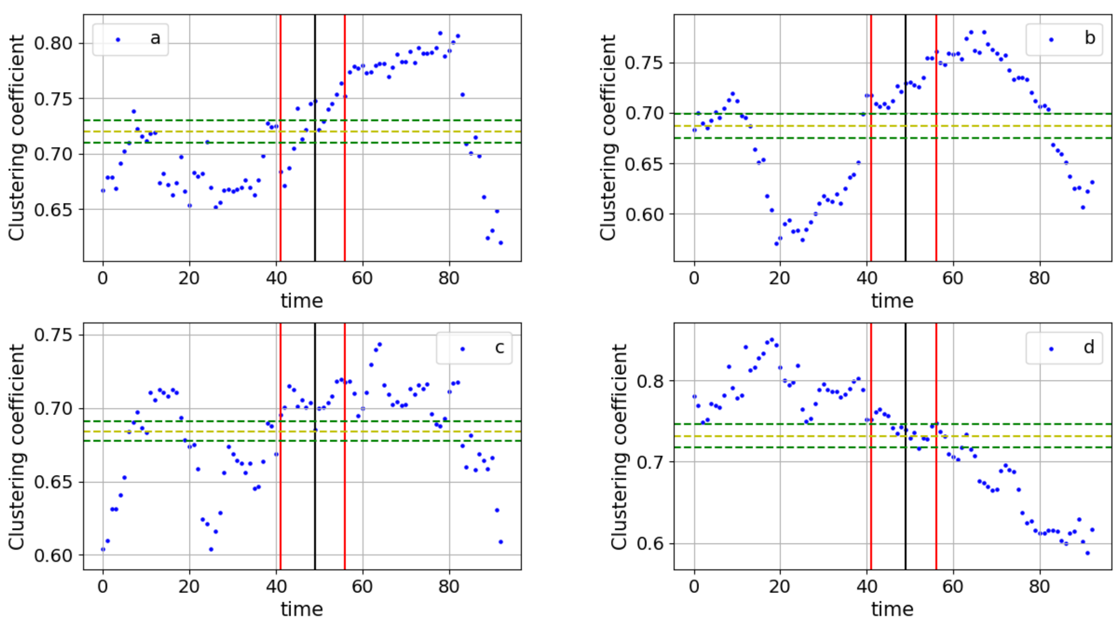

Figure 5.1: Windowed clustering coefficients for data sets of NYC(a), DJI(b), GSPC(c), NASDAQ(d).

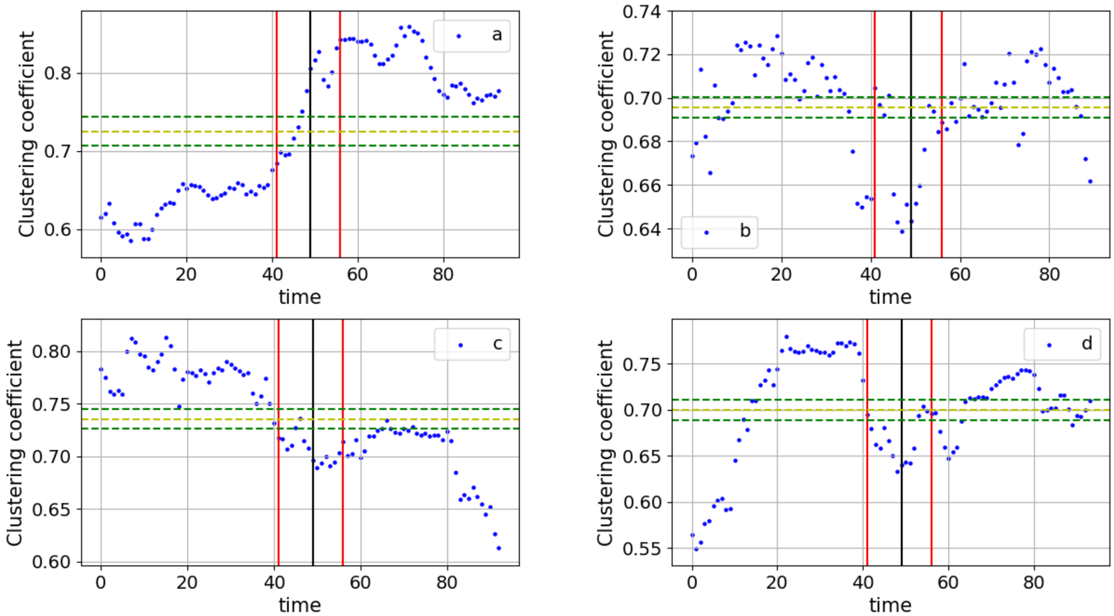

Figure 5.2: Windowed clustering coefficients for data sets of SSE(a), Nikkei(b), DAX(c), CAC(d).

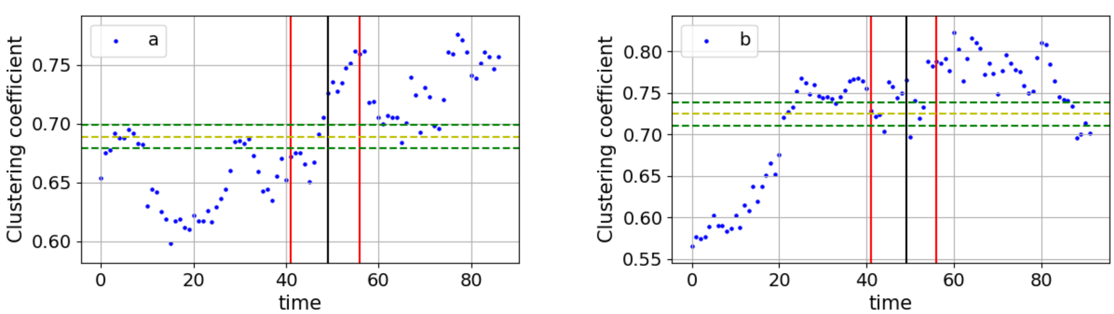

Figure 5.3: Windowed clustering coefficients for data sets of Sensex(a), TSE(b)

It can be seen that there is a sharp rise is clustering coefficient from 2006 to 2008, i.e. from first red line to black line and the time series sharply enters into the confidence interval. Immediately after the crash there is a sharp fall. But, a more important but subtle observation is that 3 of the 4 US markets, NYSE, DJI and GSPC achieve a local maxima before the first red line which belongs 2006. Also in case of NYSE, there is a stationarity observed just before 2006 red line, the time when the bubble started builing up. To further investigate this subtle behaviour, we increase resolution of the windowed analysis and slide the 2000 size windows by 10 time steps.

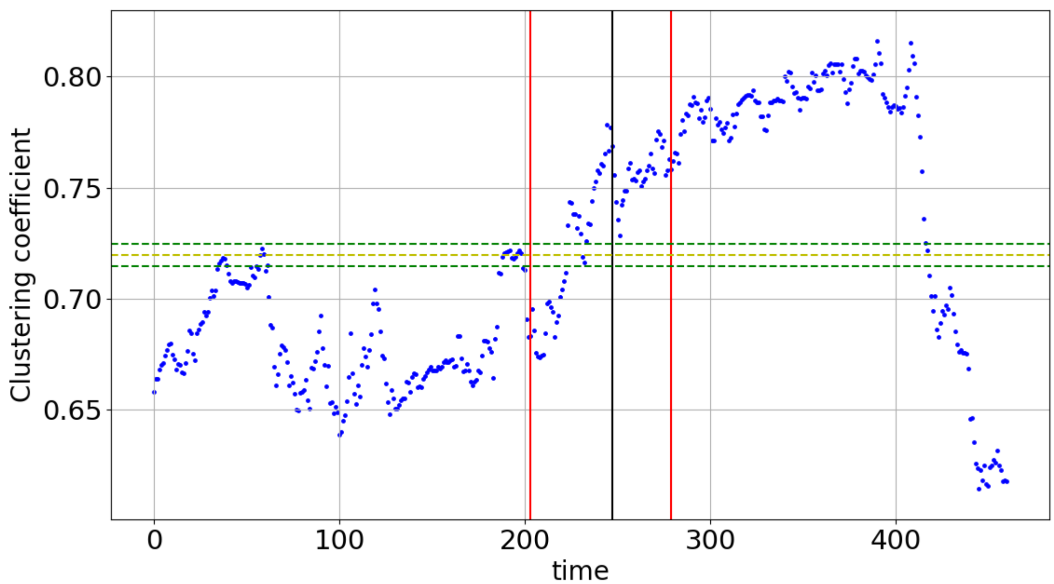

Figure 5.4: clustering coefficients of windowed analysis of NYSE with window sliding by 10 time steps.

Figure 5.4 shows clustering coefficients of windowed analysis of NYSE with window sliding by 10 time steps. Here the stationarity is remarkably achieved inside the confidence interval and then there is sharp fall. and then there is a rougly monotonic rise in clustering coefficient value as the bubble built up and a sharp fall in the interval of crisis.

5.2 Characteristic path length

Similar analysis is done for characteristic path length. Following plots show characteristic path lengths plots for all the data sets

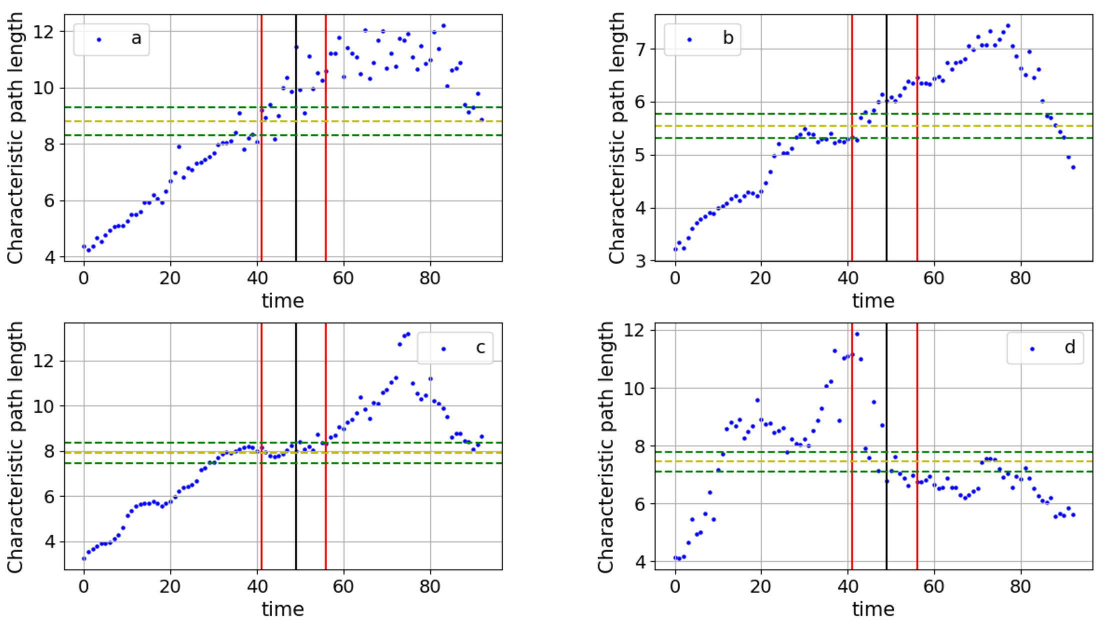

Figure 5.5: Windowed characteristic path length for data sets of NYC(a), DJI(b), GSPC(c), NASDAQ(d).

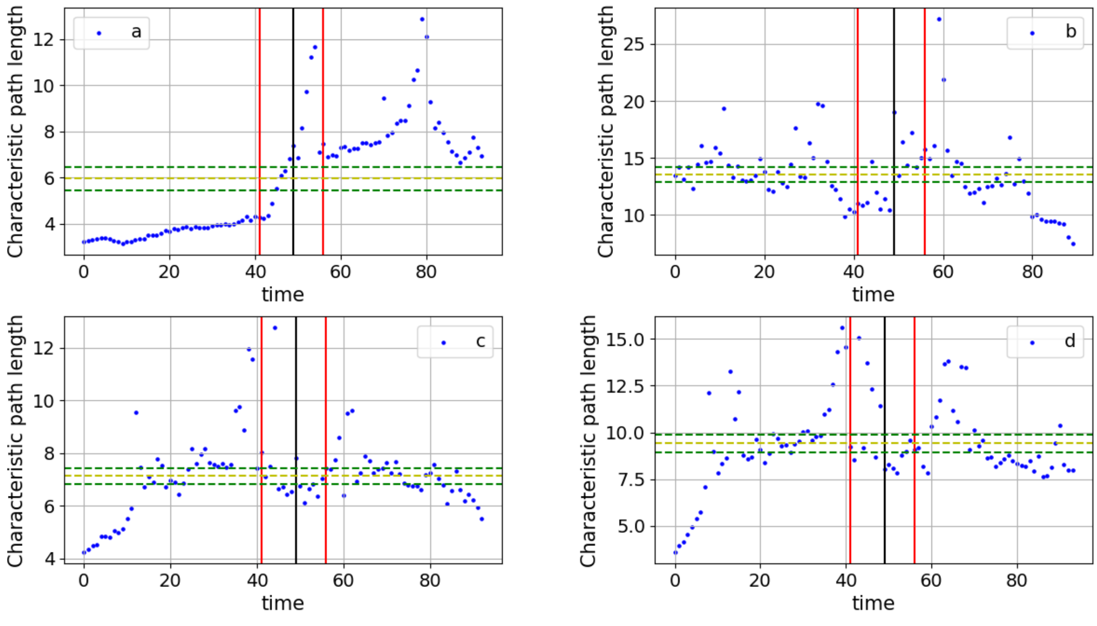

Figure 5.6: Windowed characteristic path length for data sets of SSE(a), Nikkei(b), DAX(c), CAC(d).

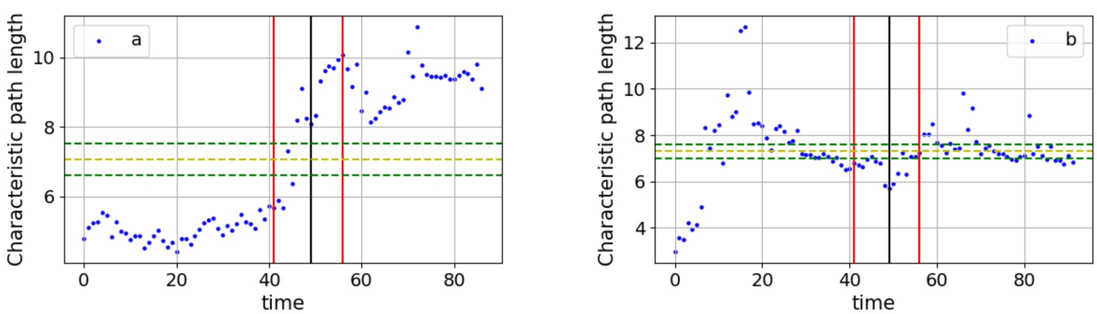

Figure 5.7: Windowed characteristic path length for data sets of Sensex(a), TSE(b)

It is important to note that between 2006 and 2008 the characteristic path length value has penetrated into confidence interval and there is sharp fall during the interval of the crash in NYSE, DJI and GSPC.

5.3 Assortativity coefficient

Assortativity coefficient indicates how vertices of different types are preferentially connected amongst themselves, and is defined by

$$r = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b}$$

where $a_x = \sum_y e_{xy}$ and $b_y = \sum_x e_{xy}$ and e_{xy} is the fraction of edges from a vertex of type x to a vertex of type y.

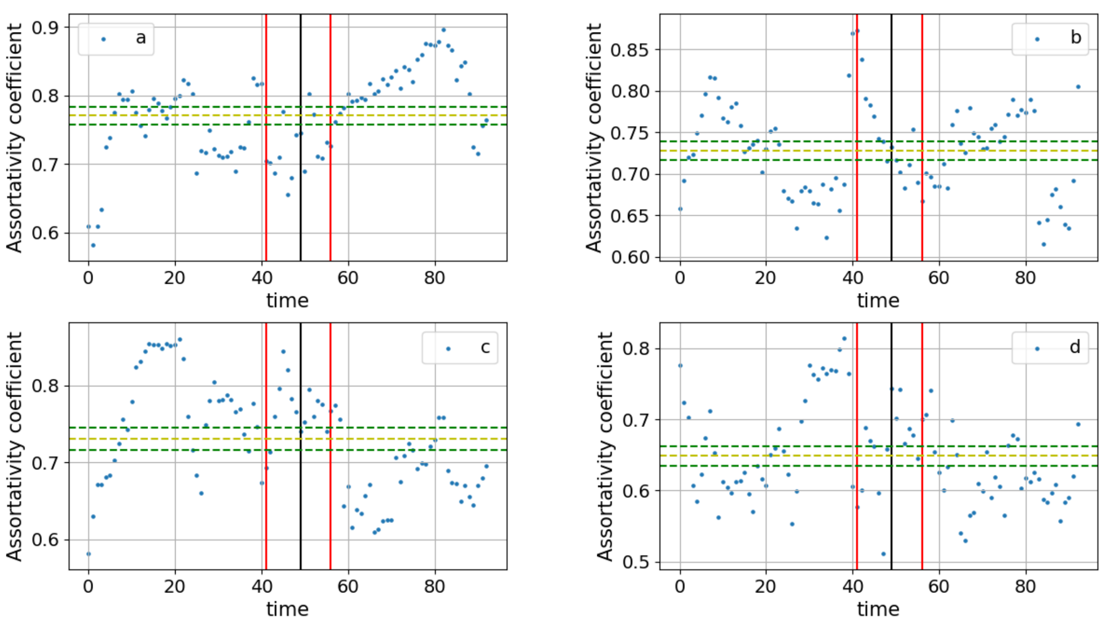

Figure 5.8: Windowed assortativity for data sets of NYC(a), DJI(b), GSPC(c), NASDAQ(d).

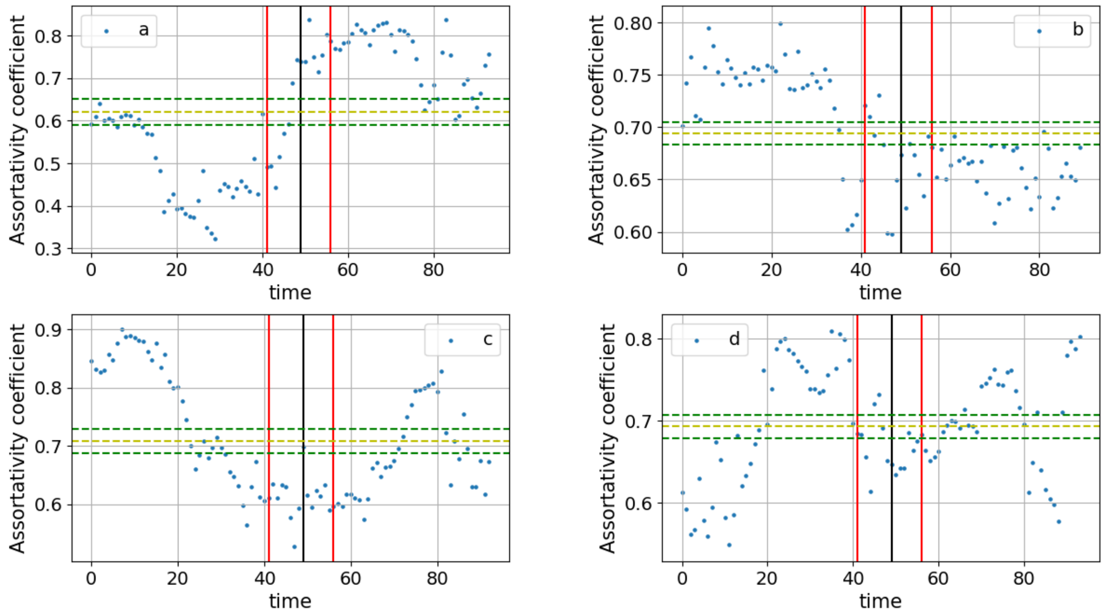

Figure 5.9: Windowed assortativity for data sets of SSE(a), Nikkei(b), DAX(c), CAC(d).

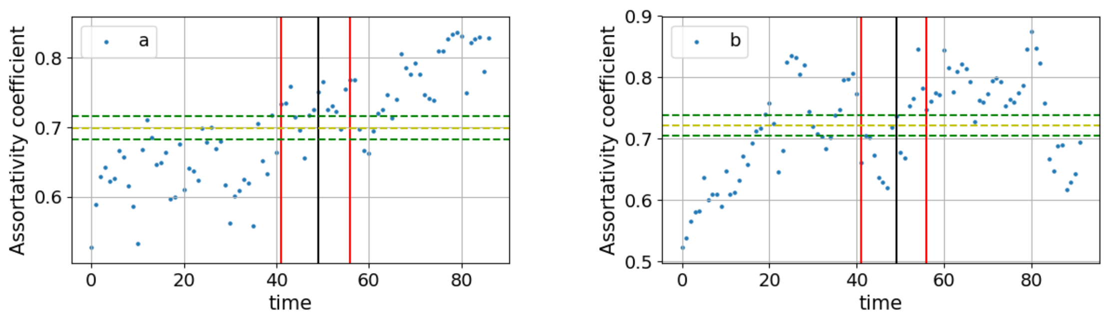

Figure 5.10: Windowed assortativity for data sets of Sensex(a), TSE(b)

Important observation is that there is a sharp fall in assortativity during the build up of the bubble and sharp rise in the interval of the crisis. This indicates changes in connectivity that comes from changes in the underlying dynamics.

5.4 Surrogates analysis for network measures

In this section we calculated 10 surrogates of every 2000 points window shifted by 50 points. Then we calculated clustering coefficients for all of the data sets. Following plots show comparison of original data set and surrogates in moving windows :

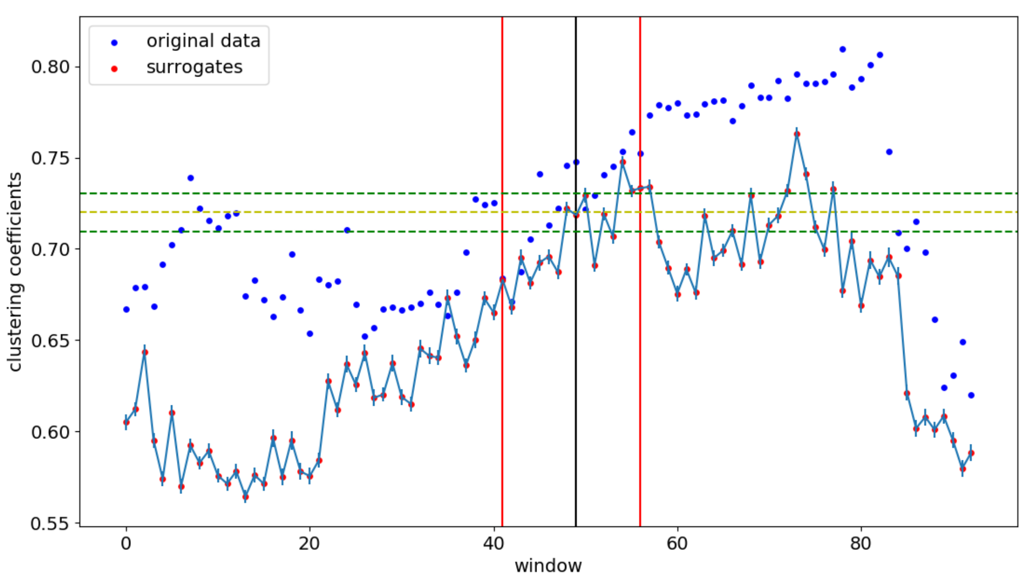

Figure 5.11: clustering coefficients for windowed analysis of surrogates for NYSE

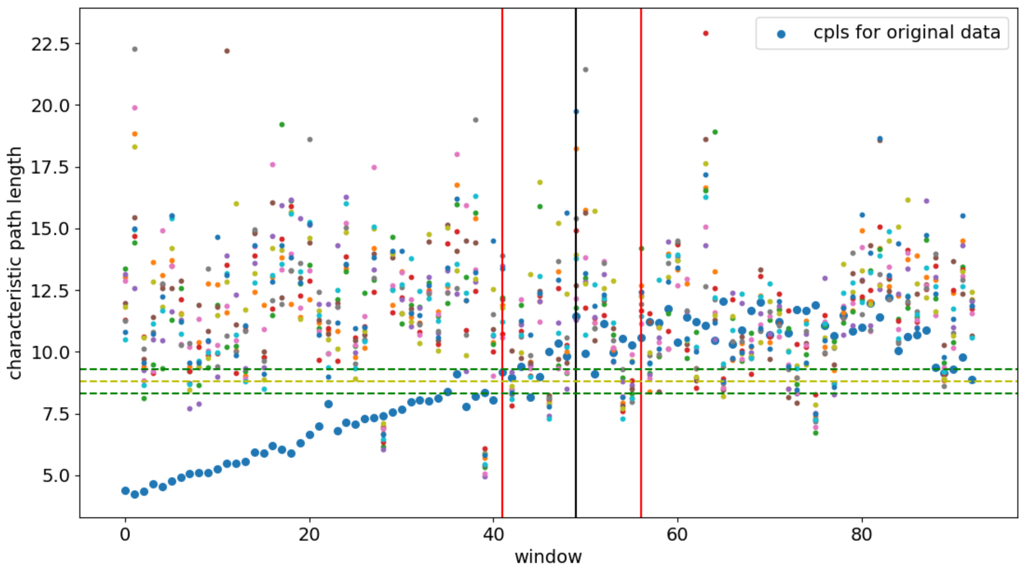

Figure 5.12: characteristic path length for windowed analysis of surrogates

It is clear that in the build up of the bubble which is the interval between first red line and the black line data behaved like surrogates in both clustering coefficient and characteristic path length sense. It is quite contrasting behaviour if the rest of the plot is considered. So from this analysis we conclude that system behaved like noise in the building of the bubble and after the second red line again dynamics started dominating at least in the clustering coefficient sense. In characteristic path length picture even if surrogate and data measures have merged even after that crash was left behind, building up of the bubble was the only time data and surrogates both lied in the confidence interval. Next we do windowed analysis of power spectra.

5.5 Power spectrum analysis

Here we calculate window wise slope of log-log plot of power spectrum for all the data set. It is important to note that red noise, by definition has this slope to be -2 of close to -2. That gives us an important criterion to characterize the behaviour with respect to red noise. If the value of the slope goes near -2, we say that at least in power spectrum sense, data is behaving like red noise.

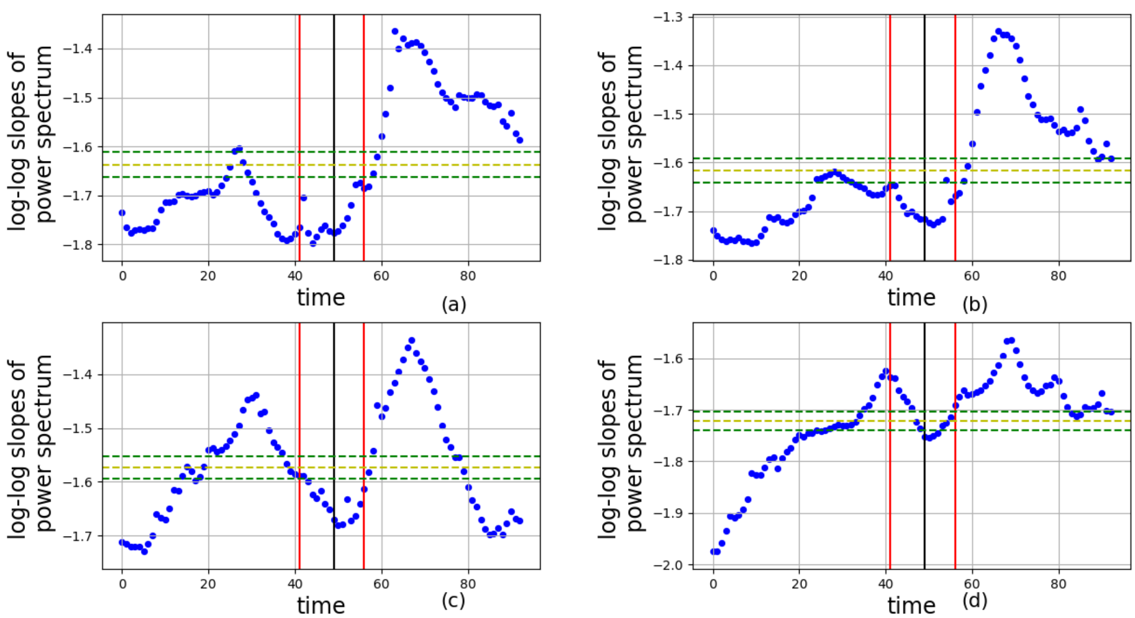

Figure 5.13: Windowed power spectrum analysis for NYSE(a), DJI(b), GSPC(c), NASDAQ(d).

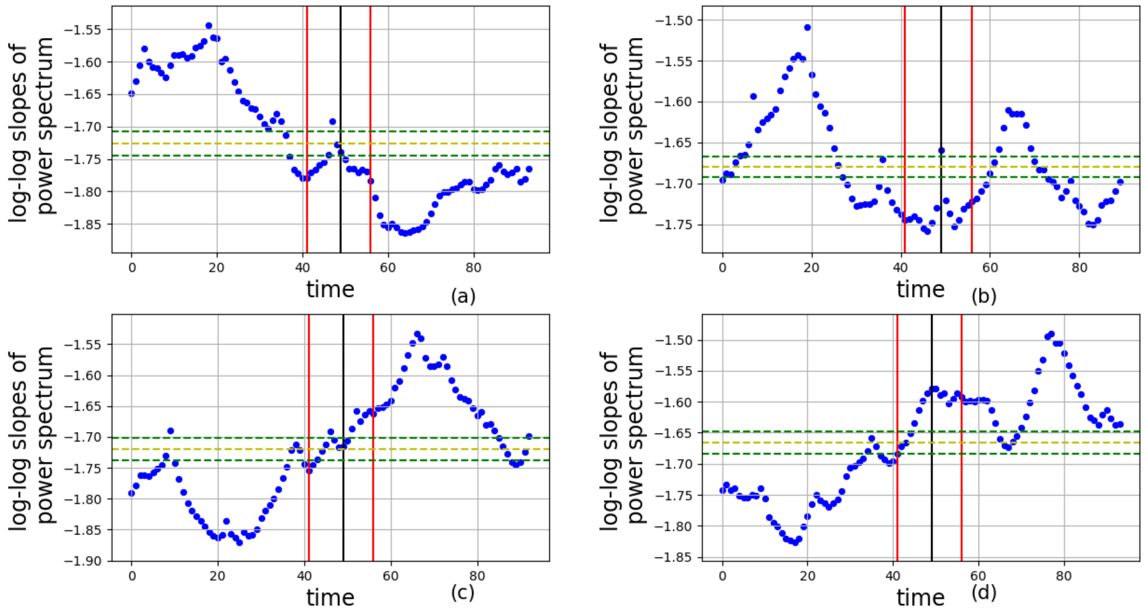

Figure 5.14: Windowed power spectrum analysis for SSE(a), Nikkei(b), DAX(c), CAC(d).

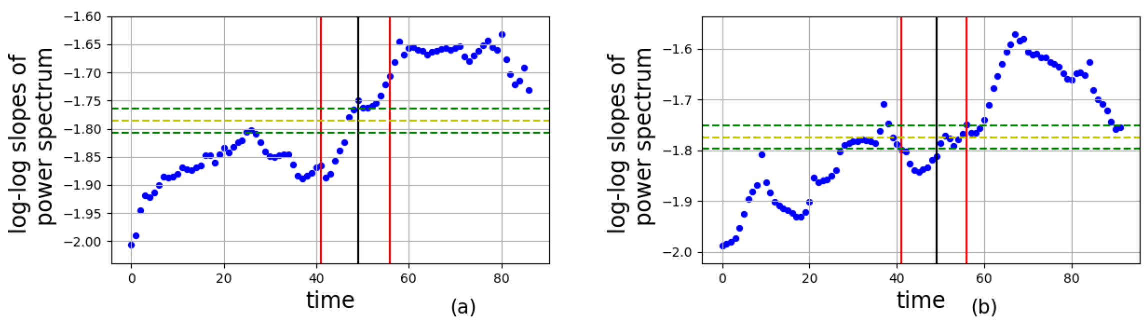

Figure 5.15: Windowed power spectrum analysis for Sensex(a), TSE(b)

If we focus on the US based stock markets, It can be seen clearly that all the US based data sets fell near -2 in the crisis period, particularly in the build up period. This indicates that during the build up of the bubble system behaved like red noise. Also It is remarkable to note that after the interval of crash, these is a monotonic rise in the slope away from -2 meaning that there is less red noise like behaviour and there is more of a dynamical behaviour. In the next chapter we conclude our work and look at the future possibilities in this way of looking at the financial markets.

Chapter 6

Discussion and future trends

The work presented in the thesis mainly centres around methods of nonlinear dynamics and time seires analysis to characterize the underlying dynamics from the data of stock markets, keeping in mind framework of fallibility and reflexivity suggested by George Soros. We first dealt with the long term variation in the data by polynomial detrending method. Then we took uniform deviates to get all the data sets in a comparable scale. We investigated presence of periodic variations by power spectrum analysis. We did phase space restructuring and calculated embedding parameters such as embedding dimension and time delay. We investigated presence of underlying dynamics by D_2 correlation dynamics. We constructed recurrence networks on the whole data and calculated network measures and compared with those on standard chaotic systems and noise. We did windowed analysis of characteristic path length, clustering coefficient, assortativity, and power spectrum. Finally we did recurrence based analysis of whole data as well as windowed analysis. Our analysis is particularly focused in the interval of the build up of the 2008 housing bubble to explore some possibility of a forecast if possible. Now we discuss the conclusions we have drawn.

- Power spectral analysis suggests that data has red noise like behaviour.
- Correlation dimension based surrogate analysis indicates that data has underlying deterministic dynamics as 7 of the 10 data sets show significant saturation of D_2.
- Degree distribution of the recurrence networks of the whole data also indicates that data has underlying dynamics as degree distribution saturates with increasing embed-

ding dimension from 4 as opposed to behaviour of white noise. However, the degree distribution of data sets resembles that of red noise.

- Characteristic path length(cpl) vs clustering coefficient(cc) plots also show similarity to red noise. Also cpl and cc based surrogate analysis suggests that data has significant difference as compared to their own surrogates.

- Windowed analysis of clustering coefficient in case of 3 of the 4 US based stock markets showed a common trend around the 2008 financial crisis, with NYSE showing distinctive stationarity behaviour in the confidence interval.

- Windowed analysis of characteristic path length showed that, the value penetrated in confidence interval during the build up of the bubble.

- Windowed analysis of assortativity coefficient showed that there is a sharp fall in the build up of the bubble.

- Surrogates based Windowed analysis suggested that US based markets behaved distinctively more like noise than a dynamical system in the build up of the bubble.

- Windowed analysis of power spectrum showed that the markets behaved significantly and distinctively more like red noise in the build of the bubble since the slope of the log log plot moved towards -2, way off the confidence interval.

Initial part of our work suggests that financial markets have underlying red noise like behaviour. This kind of characteristic is explored in the literature, but our studies have done this in recurrence network based methods. Our results indicate an underlying deterministic nonlinear dynamics. So we conclude that the data comes from dynamics with considerable red noise features added to it. The main output of our work is the effectiveness of windowed analysis with confidence interval, in the financial time series especially in the building up of a bubble, given the distinctive behaviour of the measures in the 2006 to 2008 interval. Our interpretations are as follows, in reality what happened was, Federal reserve of US kept the interest rates too low for too long. This led banks to neglect credit risk especially from 2006. Normally a financial system is said to be doing well when participants are taking well thought out decisions based of research, paying strict attention to risk. From 2006 housing prices skyrocketed with more money available to wide population because of the low interest rates. This basically means that reflexivity was in full force with agent's understanding of the

system deteriorating. Hence poorly thought out financial transactions were made, increasing noise in the system. Our study manages to pick up this important phenomenon through windowed analysis of power spectrum(value approaching -2 or red noise like behaviour) and surrogate analysis of network measures (values of data moving towards surrogates). The other network based measures also show some indications of the building up of the bubble. After 2009, well thought out decisions started being made with long term perspective. This is also picked up in our study with slopes of data deviating away from red noise and network measures moving away from surrogates. Hence recurrence network based analysis of financial data has completely analyzed dynamical transitions near the crash. Although further analysis is necessary to decisively say whether forecasting is possible with these methods. Our results point towards the possibility of forecasting a crash.
Future work in this field has 3 major directions :

- Application on sector wise data : Application of these methods on sector wise data can give a detailed view. Investigation of sectors like financial sector and real estate sector with these methods might have more significant information especially near the crash.

- Another important aspect of the financial crash is its scale. It will be interesting to see how the crash transferred throughout the global economy. This can be approached with cross recurrence and joint recurrence based analysis.

- Another very interesting financial market is the global currency market, which is the biggest market in the world. The framework of reflexivity was originally proposed and tested on this market by Soros and still there is a lot of scope for further insights in this very dynamic system.

Bibliography

[1] Gregory Zuckerman, Bradley Hope. The Quants Run Wall Street Now. The Wall Street Journal, 21 May 2017, www.wsj.com/articles/the-quants-run-wall-street-now-1495389108.

[2] Jim R. Oliver (2015) A Machine-Learning Approach to Automated Negotiation and Prospects for Electronic Commerce, Journal of Management Information Systems, 13:3, 83-112, DOI: 10.1080/07421222.1996.11518135

[3] L. Bachelier, Theorie de la speculation' [Ph.D. thesis in mathematics], Annales Scientifiques de I'Ecole Normale Superieure III-17, 21-86 (1900).

[4] F. Black and M. Scholes, 'The Pricing of Options and Corporate Liabilities', J. Polit. Econ. 81, 637-654 (1973).

[5] B. B. Mandelbrot, 'The Variation of Certain Speculative Prices', J. Business 36, 394-419 (1963).

[6] E. F. Fama, 'Efficient Capital Markets: A Review of Theory and Empirical Work', J. Finance 25, 383-417 (1970).

[7] A. N. Kolmogorov, 'Three Approaches to the Quantitative Definition of Information', Problems of Information Transmission 1, 4 (1965).

[8] Michel M. Dacorogna, Ramazan Gencay, Ulrich A. Muller, Richard B. Olsen, Olivier V. Pictet. An introduction to high frequency finance. Academic press, 2001.

[9] T. Di Matteo (2007) Multi-scaling in finance, Quantitative Finance, 7:1, 21-36, DOI: 10.1080/14697680600969727

[10] T. Bollerslev, 'Generalized Autoregressive Conditional Heteroskedasticity', J. Econometrics 31, 307-327 (1986).

[11] T. Bollerslev, 'Generalized Autoregressive Conditional Heteroskedasticity', J. Econometrics 31, 307-327 (1986).

[12] E. J. Elton and M. J. Gruber, Modern Portfolio Theory and Investment Analysis (J. Wiley and Sons, New York, 1995).

[13] H. Markowitz, Portfolio Selection: Efficient Diversification of Investment (J. Wiley, New York, 1959).

[14] R. N. Mantegna, 'Hierarchical Structure in Financial Markets', Cond.-Mat. preprint server 9802256; Eur. Phys. J. B 11, 193-197 (1999).

[15] Soros, G. (2013). Fallibility, reflexivity, and the human uncer- tainty principle. Journal of Economic Methodology, 20(4), 309-329. doi:10.1080/1350178x.2013.859415

[16] E.N. Lorenz, Deterministic non periodic flow. J. Atmosph. Sci. 20, 130141 (1963)

[17] O.E. Rssler, An equation for continuous chaos, Physics Letters A, Volume 57, Issue 5, 1976, Pages 397-398, ISSN 0375-9601,https://doi.org/10.1016/0375-9601(76)90101-8.(http://www.sciencedirect.com/science/article/pii/0375960176901018)

[18] P. Erdos and A. Renyi, On the evolution of random graphs, Publ.Math.Inst.Hung.Acad. Sci. 5, 17-61(1960)

[19] M. E. J. Newman, Networks: An introduction, (Oxford University Press, Oxford,2010)

[20] D. J. Watts, Six Degrees: The Science of a Connected Age, (Norton, New York, 2003).

[21] R. Albert and A. L. Barabasi, Statistical Mechanics of Complex Networks, Rev. Mod. Phys. 74, 47-97(2002)

[22] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D. U. Hwang, Complex Networks: Structure and Dynamics, Phys. Reports 424, 175-308(2006)

[23] S. H. Strogatz, Exploring complex networks, Nature 410, 268-276(2001)

[24] B. Barzel and A. L. Barabasi, Universality in network dynamics, Nature Phys. 9, 673-681(2013)

[25] Uniform framework for the recurrence-network analysis of chaotic time series, Rinku Jacob, K. P. Harikrishnan, R. Misra and G. Ambika-Phy Rev E 93, 012202 (2016)

[26] J. Zhang and M. Small, Complex networks from pseudoperiodic time series: topology versus dynamics, Phys. Rev. Lett. 96, 238701(2006).

[27] L. Lacasa, B. Luque, F. Ballesteris, J. Luque and J. C. Nuno, From time series to complex networks: the visibility graph, Proc. Natl. Acad. Sci. USA 105, 4972- 4975(2008)

[28] G. Nicolis, G. Cantu and C. Nicolis, Dynamical aspects of interaction networks, Int. J. Bif. Chaos 15, 3467- 3480(2005)

[29] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner and J. Kurths, Complex network approach for recurrence analysis of time series, Phys. Lett. A 373, 4246-4254(2009).

[30] Y. Yang and H. Yang, Complex network based time series analysis, Physica A 387, 1381-1386(2008)

[31] X. Xu, J. Zhang and M. Small, Super family phenomena and motifs of networks induced from time series, Proc. Natl. Acad. Sci. USA 105, 19601-5(2008)

[32] R. V. Donner, J. Heitzig, J. F. Donges, Y. Zou, N. Marwan and J. Kurths, The geometry of chaotic dynamics-A complex network perspective, European Phys. J. B 84, 653-672(2011)

[33] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan and J. Kurths, Recurrence networks: A novel paradigm for nonlinear time series analysis, New J. Phys. 12, 033025(2010)

[34] R. V. Donner, M. Small, J. F. Donges, N. Marwan, Y. Zou, R. Xiang and J. Kurths, Recurrence based time series analysis by means of complex network methods, Int. J. Bif. Chaos 21, 1019-46(2011)

[35] Y. Zou, R. V. Donner, J. F. Donges, N. Marwan and J. Kurths, Identifying complex periodic windows in continuous time dynamical systems using recurrence based methods, CHAOS 20, 043130(2010).

[36] J. F. Donges, R. V. Donner, K. Rehfeld, N. Marwan, M. H. Trauth and J. Kurths, Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis, Nonlinear Proc. Geophys. 18, 545-562(2011)

[37] Z. Gao and N. Jin, Flow pattern identification and nonlinear dynamics of gas liquid two phase flow in complex networks, Phys. Rev. E 79, 066303(2009)

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

[39] R. Avila, A. Gapelyuk, N. Marwan, T. Walther, H. Stepan, J. Kurths and N. Wessel, Classification of cardiovascular time series based on different coupling structures using recurrence network analysis, Philos. T. Roy. Soc. A 371, 20110623(2013)

[40] N. Marwan, M. H. Trauth, M. Vuille and J. Kurths, Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis, Clim. Dynam. 21, 317-326(2003)

[41] W. V. Zolotova, D. I. Ponyavin, N. Marwan and J. Kurths, Long term asymmetry in the wings of the butter- fly diagram, Astron. and Astrophys. 505, 197-201(2009)

[42] Redaelli, S., Plewczyski, D., & Macek, W. M. (2002). Influence of colored noise on chaotic systems. Physical Review E, 66(3). doi:10.1103/physreve.66.035202

[43] Anderson, T. W. (1971). The Statistical Analysis of Time Series. New York: Wiley.

[44] Cooley, James W., and John W. Tukey, 1965, An algorithm for the machine calculation of complex Fourier series, Math. Comput. 19: 297-301.

[45] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, Recurrence networks: A novel paradigm for nonlinear time series analysis, New J. Phys. 12, 033025 (2010).

[46] R. V. Donner, M. Small, J. F. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths, Recurrence based time series analysis by means of complex network methods, Int. J. Bif. Chaos 21, 1019 (2011).

[47] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, Physica D 9, 189-208(1983)

[48] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan and J. Kurths, Ambiguities in recurrence based complex network representations of time series, Phys. Rev. E 81, 015101(R)(2010)

[49] J. F. Donges, R. V. Donner, K. Rehfeld, N. Marwan, M. H. Trauth and J. Kurths, Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis, Nonlinear Proc. Geophys. 18, 545-562(2011).

[50] Z. Gao and N. Jin, Complex network from time series based on phase space reconstruction, CHAOS 19, 033137(2009).

[51] S. Schinkel, O. Dimigen and N. Marwan, Selection of recurrence threshold for signal detection, European Phys. J. ST 164, 45-53(2008).

[52] J. F. Donges, J. Heitzig, R. V. Donner and J. Kurths, Analytical framework for recurrence network analysis of time series, Phys. Rev. E 85, 046105(2012).

[53] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction", Phys. Rev. A 45, 3403 (1992).

[54] James Theiler, Stephen Eubank, Andr Longtin, Bryan Galdrikian, J. Doyne Farmer, Testing for nonlinearity in time series: the method of surrogate data, In Physica D: Nonlinear Phenomena, Volume 58, Issues 14, 1992, Pages 77-94, ISSN 0167-2789,https://doi.org/10.1016/0167-2789(92)90102-S.

[55] Nonlinear Time Series Analysis Tools - Harikrishnan Kizhakke Pisharam. (n.d.). Retrieved March 20, 2018, from https://sites.google.com/site/kphk11/home/nonlinear-time-series-analysis-tools

[56] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, Europhys. Lett. 4 (9) (1987) 973977.

[57] Ngamga, E. J., Bialonski, S., Marwan, N., Kurths, J., Geier, C., & Lehnertz, K. (2016). Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic EEG data. Physics Letters A, 380(16), 1419-1425. doi:10.1016/j.physleta.2016.02.024

[58] Marwan, N., Schinkel, S., & Kurths, J. (2013). Recurrence plots 25 years later Gaining confidence in dynamical transitions. EPL (Europhysics Letters), 101(2), 20007. doi:10.1209/0295-5075/101/20007

[59] Donges, J. F., Donner, R. V., Rehfeld, K., Marwan, N., Trauth, M. H., & Kurths, J. (2011). Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis. Nonlinear Processes in Geophysics, 18(5), 545-562. doi:10.5194/npg-18-545-2011