Nature of Stock Index Interrelations

A study on connections between global stock markets using Recurrence Plots

---

Bedartha Goswami

Thesis
Submitted Towards Partial Fulfillment
of
BS MS Dual Degree Program

Under the guidance of

Dr. G. Ambika
Associate Professor, Physics
Dean, Graduate Studies
Indian Institute of Science Education and Research
Pune, India

&

Prof. Dr. Dr. h.c. Jürgen Kurths
Head of Research Domain
Transdisciplinary Concepts and Methods
Potsdam Institute for Climate Impact Research
Potsdam, Germany

Department of Physical Sciences
Indian Institute of Science Education and Research Pune

April 2011

Certificate

This is to certify that this dissertation entitled "Nature of Stock Index Interrelations: A study on connections between global stock markets using Recurrence Plots" submitted towards the partial fulfillment of the BS MS Dual Degree Program at the Indian Institute of Science Education and Research, Pune, represents original research carried out by Bedartha Goswami at the Department of Physical Sciences, Indian Institute of Science Education and Research Pune under the supervision of Dr. G. Ambika during the academic year 2010-2011.

Bedartha Goswami

Supervisor	Head, Physical Sciences
Date:	Date:
Place:	Place:

Acknowledgments

This study would not have been possible without the support and guidance of several people.

First, I would like to express my gratitude to Dr. G. Ambika, who had not only been my guide and mentor for the past five years, but had also introduced me to the exciting world of nonlinear dynamics. In the past year, she helped me negotiate a maze of dead-ends and red herrings while working on this project and patiently tolerated my irregular working hours with faith. Without her guidance and support I would not have been able to go to the Potsdam Institute of Climate Impact Research last summer where I began to work on this study.

Next, I sincerely thank Prof. Jürgen Kurths for accepting me as a summer intern at the Potsdam Institute of Climate Impact Research and for teaching me how to pursue my ideas without fear. His expertise and insights in my field of study have been the foundation stones on which I carried out my investigations. It was his enthusiastic response to the preliminary results I obtained last summer that gave me the courage to carry on with the work as my final year project. I am also grateful to him for actively guiding me as co-supervisor to this project throughout the past nine months.

I am also indebted to Dr. Norbert Marwan for his contributions to this project as well his constant encouragement at every step along the way. His sharp and matter-of-fact criticisms, always balanced with a fair amount of praise, forced me to strive to work harder and with zeal. His perceptive comments on my work has saved me from making conceptual oversights on several occasions. He always managed to find the time to go through my work and come up with interesting ideas and suggestions regarding it.

I would also like to thank the Deutscher Akademischer Austausch Dienst (DAAD) for funding my internship at the Potsdam Institute for Climate Impact Research under the Working Internships in Science and Engineering (WISE) program, wherein this study began. I am especially thankful to Sutirth Dey and Deepak Barua for helpful discussions regarding significance testing and data representation. This work was also partly supported by the Federal Ministry of Education and Research, Germany, (project PROGRESS, 03IS2191B).

I also thank our Director, Prof. K. N. Ganesh, who has been my faculty advisor, for his constant encouragement and for providing all facilities required to complete this project. Lastly, I thank my family and my close friends who have endured me continually and whose unconditional belief in my potential as a scientific investigator forever helped me to pick myself up even when the going got tough.

— Bedartha Goswami

Abstract

The main goal of this study is to present an application of recurrence-based techniques in analyzing certain aspects of interrelations among stock indices of the world. In the past, financial data has been extensively studied for correlations using Pearson's cross-correlation coefficient ρ. In the current study, an estimator based on recurrence plots — the Correlation of Probability of Recurrence (CPR) — is used to analyze connections between nine stock indices spread worldwide. A slight modification of the CPR is suggested here in order to get more robust results.

Temporal trends in the CPR are examined by moving an approximately 19-month window along the time series and the results obtained are then compared to those obtained by carrying out a similar analysis with ρ. Binning CPR into three levels of connectedness: strong, moderate and weak, temporal trends in the number of connections in each of these three bins are extracted. Furthermore, the behavior of CPR during the Dot-Com bubble is observed by shifting the time series to align their peaks. This was done to observe the manner in which different stock markets behave during financial crisis.

This analysis, using the extended CPR approach, reveals that the markets move in and out of periods of strong connectivity erratically, instead of moving monotonously towards increasing global connectivity. This is in contrast to ρ, which gives a picture of ever increasing correlation. CPR also exhibits that time shifted markets have high connectivity around the Dot-Com bubble of 2000, indicating that markets have similar dynamical nature during a crisis irrespective of the actual date when they crash.

The importance of significance testing in interpreting measures applied to field data is highlighted in this work, especially the use of Twin Surrogates as a reliable method to generate surrogate data sets required for significance testing. CPR proves to be more robust to significance testing than ρ. It also has the additional advantages of being robust to noise, as well as being reliable for short time series lengths and low and irregular frequency of sampling. Further, it is more sensitive to changes than ρ as it captures correlations between the essential dynamics of the underlying systems.

The work done in this project forms the subject of the following paper.

On interrelations of recurrences and connectivity trends between stock indices, B. Goswami, G. Ambika, N. Marwan, J. Kurths, arXiv:1103.5189v1, — submitted to Physica A, 2011.

Contents

Chapter 1

Introduction

Einmal ist Keinmal.
What occurs but once, might as well not have occurred at all.

Recurrence is a ubiquitous phenomenon. Man has forever observed the seeming repetitive patterns in changing seasons, the turning night skies, and the migrations of birds. In philosophy, the question of 'eternal return' remains an issue of interesting debate till date. From a fundamental viewpoint of science, the recurrence of a phenomenon allows us to investigate the patterns it generates and raise questions regarding the causes underlying it. Towards the end of the nineteenth century, the notion of recurrence was given a mathematical form in Poincaré's famous theorem on recurrences, wherein he stated that, for a conservative system: "... it can be shown, that the system recurs infinitely many times as close as one wishes to its initial state" (translated from [1], as given in [2]). Since then, mathematicians have extensively explored the utilities, characteristics, and implications of the idea of 'recurrence'.

Eckmann, Kamphorst and Ruelle's 1987 report [3] defined what would later become a novel tool for the analysis of time series and the characterization of dynamical systems – *the recurrence plot*. It was a simple visual tool that captured the patterns of recurrences of a sequence of vectors by graphically representing *recurring* vectors, i.e., those pairs of vectors in the sequence which are close to each other, the closeness being defined by some chosen norm and neighbourhood. Eckmann et. al. showed that these plots could extract essential information about the time scales typical to a dynamical system.

This gave rise to numerous studies that expanded on the theory of recurrence plots and the measures extractable from them. One of the first examples of the application of recurrence plots to natural systems was its use in determining the dynamical nature of cardiac data [4]. Measures based on recurrence plots were used to characterize the nature of various model systems [5–8], as well as experimental [9, 10] and natural systems [11–13], in terms of the level of determinism, stationarity, and complexity in the data sets. These measures allowed the investigator to go past the the visual impressions obtained from recurrence plots and soon came to be known as Recurrence Quantification Analysis (RQA), a data analysis technique that could extract dynamical information "without necessitating any *a* *priori* constraining mathematical assumptions" [5] on the data sets.

Studies in the past decade (such as [14]) have further extended the utility of recurrence plots by showing that they can be used to estimate dynamical invariants and that the results are quite robust to embedding choices. These studies illustrated various procedures by which topological information about the underlying systems could be deciphered, e.g., as discussed in [15]. In a 2005 study by Romano et. al. [16], recurrence plots were used to defined measures that helped estimate the extent of different regimes of chaotic synchronization, even from non-phase-coherent and non-stationary data sets. The numerous ways in which recurrences and recurrence plots can be constructed and used to gain insight into complex systems have been presented in detail in a 2007 report by Marwan et. al. [2]

In this dissertation we present the use of recurrence plots to study financial systems and understand the patterns of interrelations between various stock indices spread over the globe. Financial systems, and stock indices in particular, are well-recognized as complex systems [17, 18]. Over the past two decades, investigations have increasingly shown the underlying nonlinearities that dominate the dynamics of stock indices and thus, numerous studies have been done to characterize financial systems using nonlinear statistics (see, e.g., [19, 20]). However, in analyzing the interrelations among different financial data sets, the use of linear correlation statistics, in particular the Pearson correlation coefficient ρ, is still widespread [18, 21–37]. The primary objectives of the current work are to re-examine commonly held notions of connectedness and explore the potential of using a slightly modified extension of the Correlation of Probability of Recurrence (CPR) (presented first in [16]) to serve as a new measure that can quantify the level of connectedness between different stock indices.

The potential of recurrences in analyzing financial data has been explored in several studies, e.g., in correlation analysis among currencies [38], identification of high dimensional determinism underlying various index returns [39], in estimation of initial time of a bubble [40], identification of nature of crashes [41] and in quantifying the behavior of global stock markets during financial crises [42]. However, very few studies have attempted to study connectivities between financial markets using recurrence-based measures; only one such recent study used cross recurrence analysis on synchronicity and convergence of the GDP among member nations of the Euro region [43].

A secondary objective of this thesis is to stress on the importance of significance testing in the analysis and interpretation of measures calculated from natural systems. Most econometric studies are lacking in this regard. In particular, we apply the Twin Surrogates algorithm [2, 44], another recurrence-based algorithm, for generating surrogates of our financial time series. We then carry out statistical significance tests using these surrogate data sets for a more correct interpretation of the results.

This dissertation is organized as follows:

Chapter 2 first talks about the fundamental notion behind using a recurrence-plot-based measure to estimate interrelations between financial data sets and lists out the assumptions involved in the analysis. It then outlines the concepts essential to this study, viz., recurrence plots, probabilities of recurrence and the correlation of the probabilities of recurrence; and talks about the idea behind significance testing of measures obtained from time series and the advantages of the Twin Surrogates algorithm that was used to generate surrogate data sets for the significance tests.

In Chapter 3 the steps involved in analyzing the data are detailed, which includes a mention of the exact data sets used and their source, the pre-processing of the data sets before the actual analysis (aligning of the dates and normalization of the various time series), and the selection of the various parameters essential to recurrence analysis of the data sets.

The results obtained in the work are discussed in Chapter 4. This chapter talks about the initial static picture of the network of connections that was obtained and the subsequent sliding window analysis done to obtain temporal trends of the connections among the various stock markets; and it highlights the primary results of observed erratic cycles of long bands of strong connectedness interspersed by shorter periods of low connectivity. It also discusses the behavior of financial markets during the Dot-Com crisis of 2000 and the advantages of the methods used in this work.

Chapter 5 summarizes the primary novel features of the work and presents possible avenues of future work related to recurrence-plot-based measures and their applications to interrelations of financial data sets.

Chapter 2

From recurrences to interrelations

The first step in understanding natural dynamical systems is to look at different features of time series arising from them. Time series have the ability to reveal various characteristics of the macroscopic dynamics of the systems. In attempting to quantify the interrelations between different stock indices of the world, we look at time series data from the stock markets and ask the question: "How do connections between stock markets change over time?" To answer this, a measure of *connectedness* must first be arrived at. In earlier studies, the Pearson's cross-correlation coefficient ρ served as a proxy for 'links' between financial data sets and its extensive usage has made ρ become synonymous with the notion of correlation, and thereby *connections*, itself. Recent advances in nonlinear data analysis have, however, opened up possibilities of using newer techniques that can potentially substitute Pearson's ρ, e.g., the study of recurrences, Recurrence Plot (RP), Cross Recurrence Plot (CRP), and Joint Recurrence Plot (JRP) [2]. In this study, we estimate connections between financial data sets from the recurrences of dynamical systems. We use (a slightly modified version of) the Correlation of Probability of Recurrence (CPR), which is based on RPs and was originally devised to quantify phase synchronization between non-phase-coherent and non-stationary time series [2, 16]. As the notion of synchronization is innately bound to those of connectedness and co-movement, we argue that CPR too can serve as a measure for connectedness.

This chapter outlines the fundamentals of our approach and explains the theoretical basis for it. The last section of the chapter deals with the need of significance testing (using surrogate data sets and, in particular, Twin Surrogates (TS)) in interpretation of measures obtained from time series data.

2.1 Measuring connectivity

The common thread in connectivity studies of financial data has been *co-movement*, and it has been referred to with varied terms like correlation [21, 22], synchronization [45] and cointegration [46, 47]. Each of them incorporates a mathematical formulation that captures some aspect of co-movement for a time series pair. In measuring connectivity we look at another feature that can imply connectedness: *similarity*. If two data sets are *similar* in an aspect that can be measured, then they are closer to each other than, say, with a third data set with which neither of them share as much common ground. Similarity is a more general feature, one of whose manifestations might be co-movement. In our analysis based on a recurrence approach, we take into account the real evolution typical for financial markets better than the classic correlation analysis. Therefore, we make the following basic assumptions:

i. The time series that are dealt with, in this study, are the output of *black boxes*, i.e., those systems whose dynamics and model equations are unknown.

ii. The dynamics of such systems may change over time (non-stationarity).

iii. The change in dynamical nature is itself a characteristic feature of the system.

iv. The time series may have features that are common to all of them (e.g. power spectra and clustered volatilities) and further, these similarities are quantifiable.

v. The quantifiable features are representative of the underlying dynamical nature of the system.

The *quantifiable feature* studied here is the probability of recurrence (see Sec. 2.3) and the *similarity* is captured by the cross-correlation between the probabilities of recurrence of pairs of time series (see Sec. 2.4).

2.2 Recurrence Plots

A *Recurrence Plot* (RP) is a visual tool that shows the recurrence patterns of a dynamical system [3]. A *recurrence* is defined as the return of the trajectory of a system to an earlier state. In practice, a recurrence is said to occur when the system returns to the neighborhood of an earlier point in the phase space. Mathematically, given a point $\vec{x}_i \in \mathbb{R}^m$ of a trajectory $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$, the recurrence matrix $\mathbf{R}$ is estimated as:

$$\mathbf{R}_{i,j}(\varepsilon) = \Theta(\varepsilon - \parallel \vec{x}_i - \vec{x}_j \parallel), \quad i, j = 1, \dots, N, \tag{2.1}$$

where N is the number of points, ε is an appropriate threshold distance, $\Theta(\cdot)$ is the Heaviside function (i.e., $\Theta(a) = 0$ if $a < 0$, and 1 if $a \geq 0$) and $\parallel \cdot \parallel$ is an appropriate norm. $\mathbf{R}$ is a matrix of 0s and 1s and an RP is a graphical representation of $\mathbf{R}$ obtained by, e.g., marking a black dot for every 1 and a white dot for every 0.

RPs capture the essential dynamical features of a system [2, 3]. RPs of three different types of data sets, viz., uniform white noise, chaotic Lorenz system, and daily financial data, are shown in Fig. 2.1. All three plots are distinct from each other and characteristic of the system.

Next, we present some measures to characterize RPs.

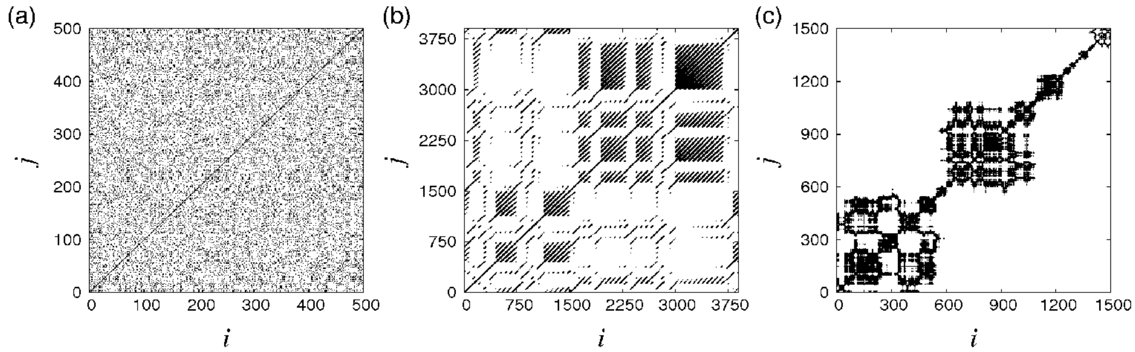

Figure 2.1: **Recurrence plots for three different types of data**. (a) Uniform white noise. (b) The Lorenz System with $\sigma = 10$, $\rho = 28$ and $\beta = 10/3$. (c) Daily data from DAX. Here, i and j are the index values for the time series entries.

2.3 Probability of Recurrence

The *Probability of Recurrence* $p(\tau)$, also called the τ-*recurrence rate*, is the recurrence rate of a diagonal line situated at τ steps from the main diagonal, i.e., $\mathbf{R}_{i,i+\tau}\,\forall\, i = 1, \ldots, N-\tau$ [2]. It is a probabilistic measure that gives the probability of an $(i+\tau)^{\text{th}}$ point falling in the ε-neighborhood of the i^{th} point:

$$p(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \mathbf{R}_{i,i+\tau}. \tag{2.2}$$

It can be considered as a generalized form of an autocorrelation function that statistically reflects the time scales of the system in which it tends to return to a previous configuration. For instance, the uniform white noise time series of Fig. 2.1(a) has almost the same probability of recurring to an earlier state for all values of τ (Fig. 2.2(a)), while the chaotic Lorenz system of Fig. 2.1(b) has periodic tendencies for high recurrences but with decreasing intensity (Fig. 2.2(b)), and the probabilities of recurrence for the daily DAX data from Fig. 2.1(c) decreases (without periodicities) with increase in τ, indicating the chances of a drift in the data set (Fig. 2.2(c)).

2.4 Correlation of Probability of Recurrence (CPR)

The *Correlation of Probability of Recurrence* (CPR) is defined as the cross-correlation coefficient between the probabilities of recurrence of two trajectories $\vec{x}$ and $\vec{y}$ [2, 16]:

$$CPR = \langle\, \bar{p}_{\vec{x}}(\tau)\, \bar{p}_{\vec{y}}(\tau)\, \rangle, \tag{2.3}$$

where $\langle\cdot\rangle$ represents the expectation value and $\bar{x}$ is the series x *normalized* to zero mean and standard deviation one (henceforth, 'normalization' refers to this particular way of normalizing a time series).

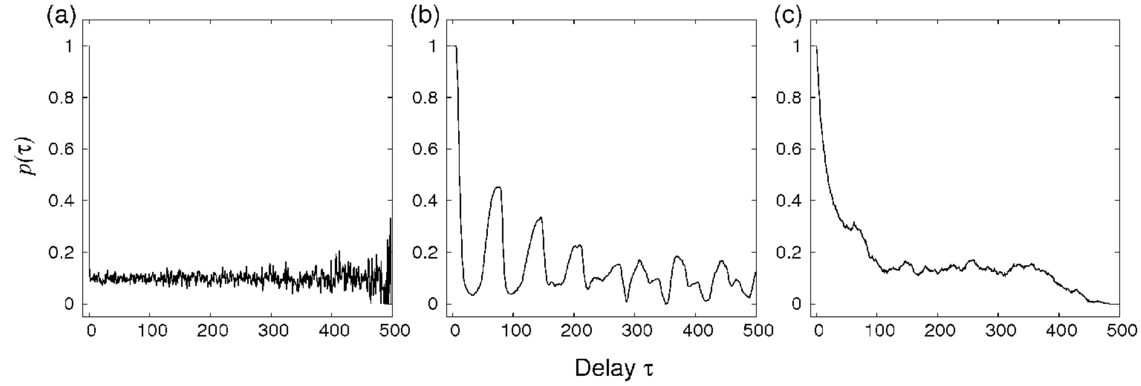

Figure 2.2: $p(\tau)$ **curves for the three different types of data given in Fig. 2.1**. (a) Uniform white noise. (b) The Lorenz System with $\sigma = 10$, $\rho = 28$ and $\beta = 10/3$. (c) Daily data from DAX.

However, all the $p(\tau)$ curves in Fig. 2.2 start from $p(0) = 1$, because the recurrence rate is always 1 at $\tau = 0$, the main diagonal. This initial portion of the $p(\tau)$ curve, common to all trajectories, introduce a bias towards a high CPR value. To evaluate CPR correctly, we suggest to consider only $p(\tau)$ for such τ larger than the autocorrelation time τ_c of the system (defined as the delay τ at which the autocorrelation function of the system falls to $1/e$):

$$CPR = \langle\, \bar{p}_{\vec{x}}(\tau > \tau_c)\, \bar{p}_{\vec{y}}(\tau > \tau_c)\, \rangle, \tag{2.4}$$

where

$$\tau_c = max\,\{\, \tau_c(\vec{x}),\ \tau_c(\vec{y})\,\}. \tag{2.5}$$

This is shown in Fig. 2.3 which illustrates the steps involved in estimating CPR. The CPR between the two time series according to Eq. (2.3) is 0.892, whereas it is 0.575 according to Eqs. (2.4) and (2.5).

CPR characterizes the degree of phase synchronization between two time series, with $CPR \approx 1$ implying that the two systems are phase synchronized [2, 16]. However, it can be interpreted more generally as a measure denoting the level to which two trajectories $\vec{x}$ and $\vec{y}$ have similar time scales of recurrence. In this study, this means that two financial time series with a high CPR tend to recur at similar times, suggesting some similarity in their underlying dynamics.

2.5 Pearson correlation

The *Pearson correlation coefficient* (ρ) is commonly used to analyze correlations between financial data. It is then used to define the *distance* between the data sets as given in [21]. It is simply the expectation value of the product of the normalized time series:

$$\rho_{\vec{x},\vec{y}} = \langle\, \bar{\vec{x}}\, \bar{\vec{y}}\, \rangle, \tag{2.6}$$

where $\langle\cdot\rangle$ and $\bar{x}$ are the same as before.

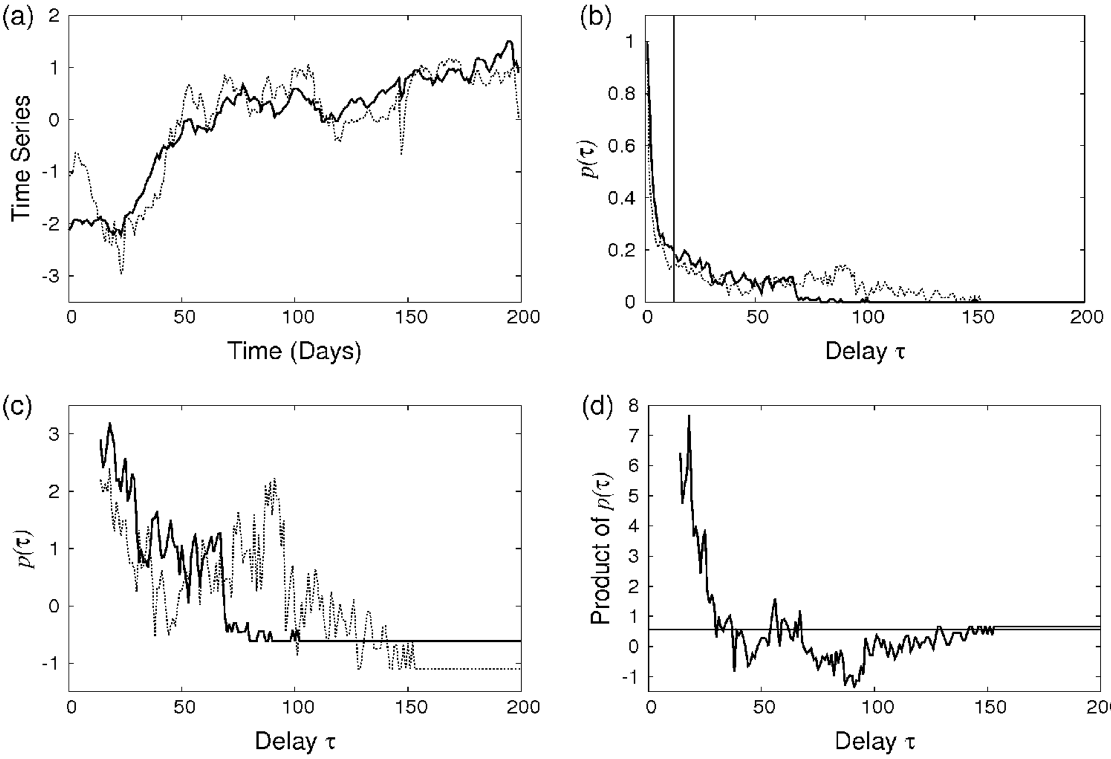

Figure 2.3: **Calculating** CPR **from** $p(\tau)$. (a) Normalized daily data for 200 days from two stock indices. (b) $p(\tau)$ curves for these time series. $\tau_c = 13$ is shown with a vertical line. (c) Normalized $p(\tau)$ curves beyond τ_c . (d) The product of the two series in (c). The horizontal line is the mean of this series, which is the CPR.

2.6 Significance testing and Twin Surrogates

To use the CPR alone for interpretations might be misleading. In an *active experiment* (such as laboratory experiments or numerical simulations of model systems) where the parameters of the system can be controlled, the CPR for different parameter sets can be compared and thus, a consistent interpretation is possible. However, the financial time series used here are the only realizations of the black boxes generating them. In such a *passive experiment*, where the parameters of the system cannot be changed or controlled, or its dynamics are unknown, it is crucial to generate surrogate time series and check for the statistical significance of the observed CPR against the distribution of CPR obtained from the surrogate data sets. This is done using a statistical test and an appropriate null hypothesis.

The *Twin Surrogates* (TS) algorithm is a recurrence-based method of generating surrogate time series [44]. A pair of points $\vec{x_i}$ and $\vec{x_j}$ of a trajectory $\vec{x}$ (of length N) are called *twins* if, for $k = 1, 2, \ldots, N$; $\mathbf{R}_{k,i} = \mathbf{R}_{k,j}$. This means that, barring their exact positions in the trajectory, these two points have the same neighborhood in phase space. The TS method is an iterative algorithm that involves:

i. Identifying twins from the recurrence plot of the trajectory $\vec{x}$.

ii. Taking any arbitrary point $\vec{x_k} \in \vec{x}$ as the starting point of the surrogate trajectory $\vec{s}$.

iii. Iteratively adding subsequent points to $\vec{s}$ as: if $\vec{x_l} \in \vec{x}$ is the previous point of $\vec{s}$, and $\vec{x_l}$ has no twins, then the next point of $\vec{s}$ is simply $\vec{x_{l+1}}$; whereas if $\vec{x_l}$ belongs to a set of n twins, the next point of $\vec{s}$ is any one of the remaining $n-1$ twins of that set, chosen with equal probability.

Although there are several methods to generate surrogate time series, it is natural to use TS in this study. In the widespread *iterative Amplitude Adjusted Fourier Transform* (iAAFT) method, surrogates are created with the assumed null hypothesis that the observed time series is the result of a Gaussian process seen through a static linear filter whereas, in TS, each surrogate is an independent realization of the observed time series differing only in the initial conditions. iAAFT surrogates do not preserve nonlinear characteristics such as mutual information, whereas TS preserves linear as well as nonlinear properties of a system [44]. Fig. 2.4(a) shows the original time series from DAX and both the surrogates via iAAFT and TS. Although the errors in autocorrelation for both methods are comparable, the iAAFT surrogate has much larger error in mutual information than the TS (Figs. 2.4(b)-(e)).

We use TS to generate a test distribution using which the significance of the observed measure M (which can be either CPR or ρ) is tested. The null hypothesis for this significance test is that each surrogate is an independent trajectory of the same dynamical system which gave rise to the observed time series. This means that when we test for significance, we simply check for the probability that an independent realization of one of the time series (as given by its TS) can give a similar value of M. The steps involved in the test are:

i. The value of M between time series A and B is estimated and designated as M_o (say).

ii. TS are generated from the series B.

iii. M is calculated between each surrogate and the series A. This gives the test distribution of M. We assume that this distribution is roughly normal as each TS corresponds to a distinct trajectory of system B starting from an independent initial condition.

iv. The mean μ and standard deviation σ are estimated for this test distribution.

v. The test statistic Z is then evaluated as:

$$Z = \frac{M_o - \mu}{\sigma}, \tag{2.7}$$

which can be used to infer the probability with which Z belongs to a standard normal distribution of mean μ and standard deviation σ.

A prefixed cut-off for the probability is decided below which M is said to be *significantly different* from the test distribution, and this is the *significance level* of

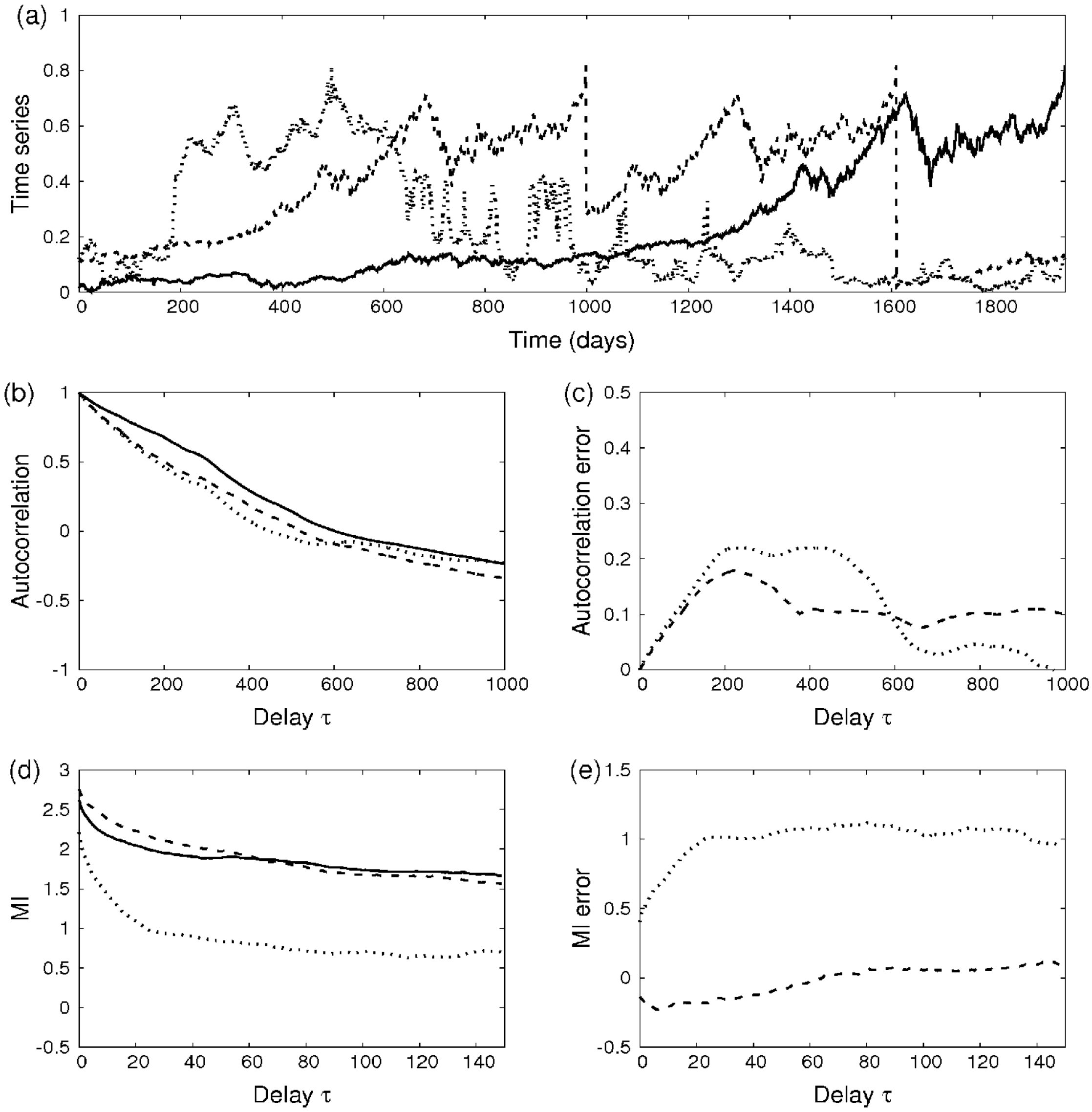

Figure 2.4: **Comparison of Twin Surrogates and the iAAFT methods**. (a) Daily data for 1942 days from DAX (bold) (scaled to lie between 0 and 1), and a realization of its corresponding Twin Surrogate (dashed) and iAAFT surrogate (dotted). (b) The autocorrelation functions of the three sets of data in (a). (c) The error in autocorrelation of the Twin Surrogate (dashed) and the iAAFT surrogate (dotted). (d) Mutual information of the data sets in (a). (e) The error in mutual information of the Twin Surrogate and iAAFT surrogate. Keys for (c) and (e) are the same as in (b) and (d) respectively.

the test. The probability value (or *p-value*) obtained from the standard normal table for the test statistic Z represents the probability that M might actually be from the test distribution. In terms of the null hypothesis — that the two time series are independent — the p-value represents the probability that the null hypothesis cannot be rejected.

Once a measure is obtained during an analysis of field data, any interpretation based on its value has to be done in the light of the results of a corresponding statistical test done using surrogates and only after considering the corresponding p-values. This is a safe way to avoid pitfalls and oversights in investigations done on natural systems.

Chapter 3

Analysis of stock market data

The overall connectivity analysis done in this study is divided into three major components: (a) creating a static picture of the network of connections by estimating CPR for the entire length of time series, (b) analyzing the temporal evolution of CPR by sliding a window of fixed size over the time series, and finally (c) looking at the behavior of the connections (defined by CPR) during the crisis period of the Dot-Com bubble of 2000.

However, before the actual estimation of CPR (and ρ), the various time series had to be aligned so that their dates were lined up, then they had to be normalized and further, several parameters required for the analysis had to be figured out. This chapter lists out the data sets that were used in the analysis and reports the steps involved in the preliminary analysis (of preprocessing the data and selection of parameters).

3.1 The data set

The daily close values ranging from 3$^{\text{rd}}$ December 1990 to 30$^{\text{th}}$ April 2010 of nine stock indices (given in Table 3.1) from around the world were used in the analysis. Three were from Asia, three from Europe and three from the U.S.A. The data was obtained from `http://finance.yahoo.com/`. Each data set contained a numerical value (representing the close value of the index on a particular date) and its corresponding date.

3.2 Preprocessing the data set

The data sets had unequal lengths because of the unequal distribution of holidays for stock indices in different regions of the world. To align them temporally, mismatched dates (and the corresponding close values) were deleted, i.e., any date of a particular market not present in any one (or more) other market(s) of the remaining eight resulted in the deletion of that date (and the corresponding close value) from all the markets. Simply put, a common intersect of all the nine data sets, in terms of dates, was obtained, which was of length 4238 time points.

Table 3.1: Test data: Market indices and their locations

Label	Stock Index	Location
A	CAC 40	France
B	FTSE 100	U.K.
C	DAX	Germany
D	NASDAQ	U.S.A
E	DJIA	U.S.A
F	S&P 500	U.S.A
G	Nikkei 225	Japan
H	Hang Seng	China
I	Strait Times	Singapore

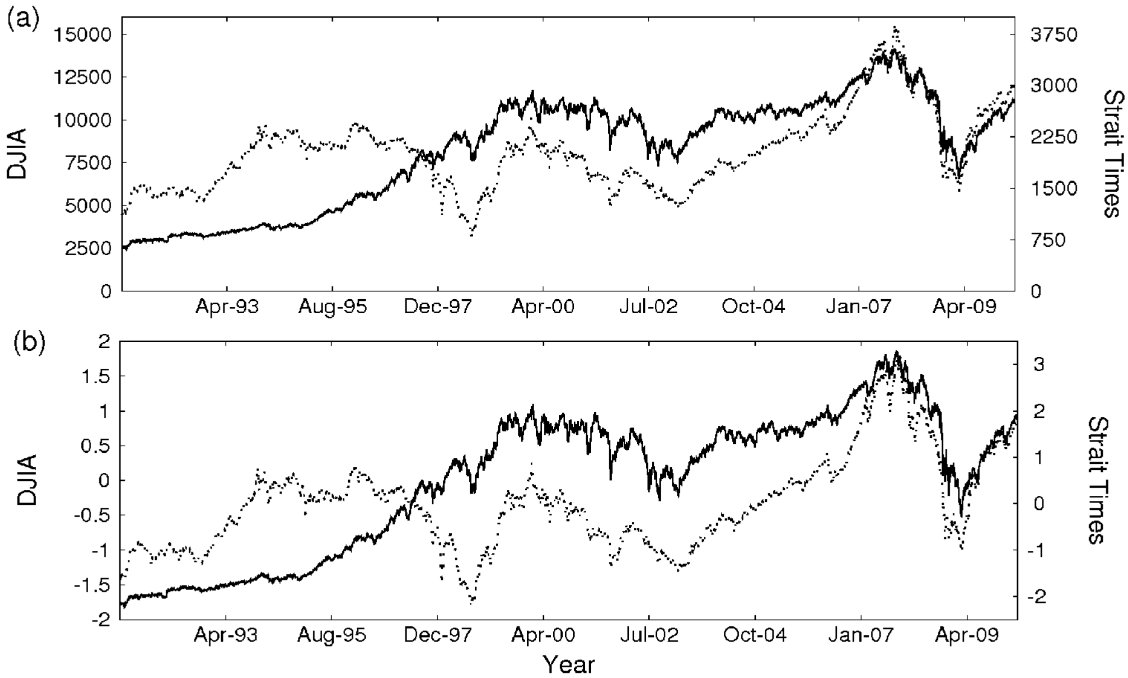

Figure 3.1: **Normalizing the data**. (a) Daily close values for Dow Jones Industrial Average (DJIA) (bold)and Strait Times (dotted). (b) The time series in (a) after normalization (note the difference in the vertical axes in both figures). The left vertical axis is for DJIA while the right vertical axis is for Strait Times.

Moreover, the index values of the different indices were arbitrary (Fig. 3.1(a)). To make qualitative comparisons possible, they were *normalized* to mean zero and a standard deviation of one (Fig. 3.1(b)). This enabled us to compare the scaled time series and all the recurrence-based measures obtained from them using the same value of the parameters, e.g., the same value of the recurrence threshold ε.

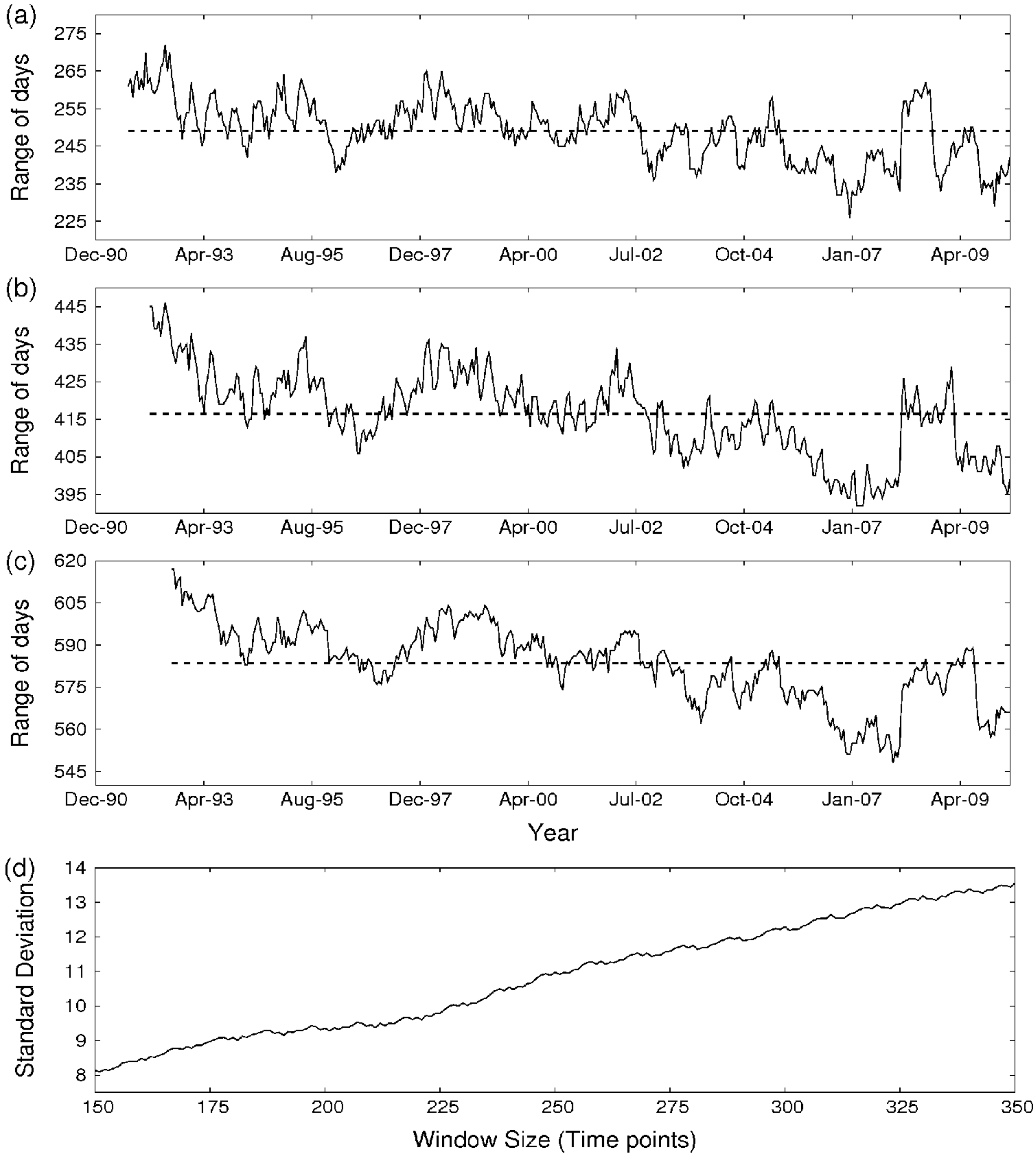

Figure 3.2: **Selecting the window size**. (a)-(c) Range of days contained in windows of sizes 150, 250 and 350 time points as they move along the time series in steps of 10 time points. The horizontal dashed lines are the mean number of days. (d) Standard deviation of the distribution of range of days contained in windows sized between 150 to 350 time points.

3.3 Selection of parameters: Preliminary analysis

3.3.1 Window size

A primary step in this study is to slide a window of fixed width along the time series and then estimate CPR or ρ. However, because of the deletion of dates (see

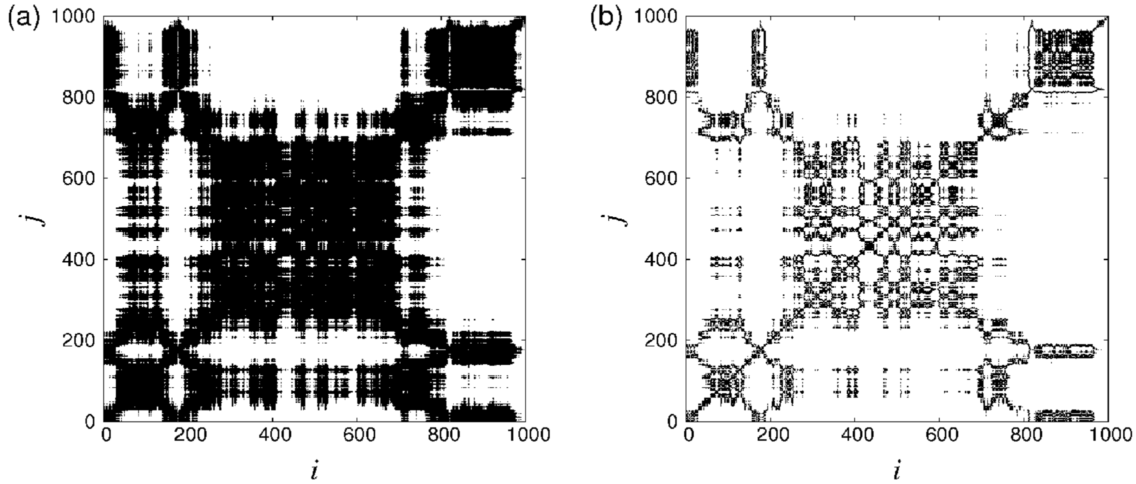

Figure 3.3: **Recurrence plots for different thresholds**. (a) RP for 1000 days of FTSE 100, $\varepsilon = 10\%$ of maximum distance. (b) RP obtained by normalizing the data in (a), and $\varepsilon = 0.1$, which is around 2% of the maximum distance. This gives a clearer visualization of the finer structure. For both RPs, the time series were not embedded.

Sec. 3.2), if a window of, say, 500 consecutive time points is chosen, the actual *range* of days contained in it is larger than 500 and further, this range changes as the window moves along the time series (Fig. 3.2(a)-(c)). In order to get an understanding of the variation in the range of days contained in a window with respect to the number of time points in the window, the window size was varied from 150 to 350 time points and the standard deviation of the range of days contained in each window was estimated: the standard deviation increases (almost) monotonically with increasing window size (Fig. 3.2(d)).

Ideally, the window should have a negligible standard deviation. However, this would mean reducing the window size which would reduce the effectiveness of the RP and, in turn, the measures estimated from it. As a compromise, a window of 250 time points is chosen, which is partly arbitrary but it is (reasonably) assumed that the qualitative features of the results are not severely effected by increasing or decreasing the window size from 250 by a small margin. The mean range of days contained in a 250 time point window is approximately 416 days, which is roughly equal to 19 months (considering a 5-day week).

3.3.2 Recurrence Plot parameters

There are several ways of constructing an RP from data [2], e.g., with a fixed threshold ε or a varying one, with or without embedding. In the current analysis, the time series are not embedded. Also, there is no fixed rule to select ε. It depends on the objectives of the study and the nature of the data set. The choice of ε should ensure that the recurrence matrix $\mathbf{R}$ represents the dynamics of the system. It should neither be too large (to avoid counting spurious recurrences) nor be too small (to

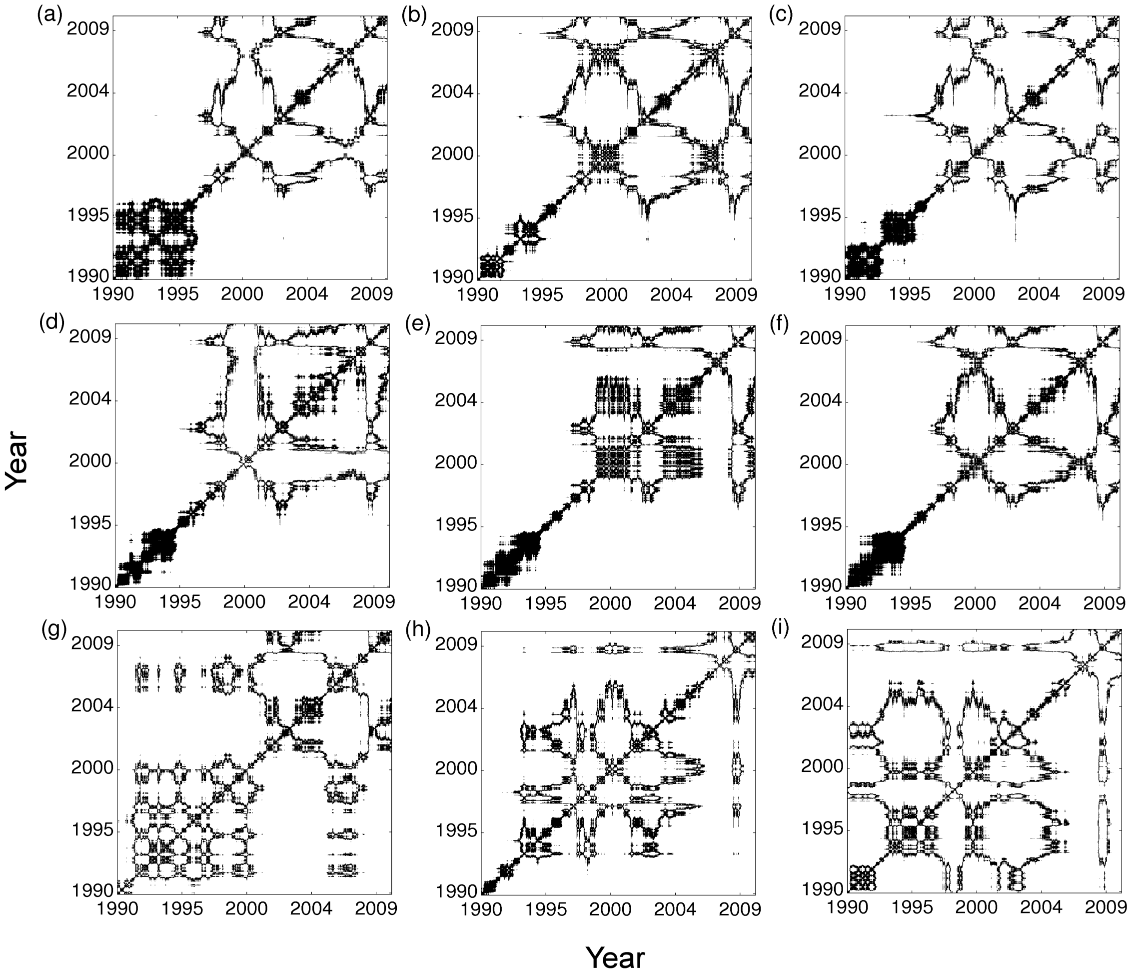

Figure 3.4: **Recurrence plots for the nine stock indices**. (a) CAC 40. (b) FTSE 100. (c) DAX. (d) NASDAQ. (e) DJIA. (f) S&P 500. (g) Nikkei 225. (h) Hang Seng. (i) Strait Times. Recurrence threshold $\varepsilon = 0.1$; RPs obtained without embedding.

avoid excluding crucial recurrences). One rule of thumb is to ensure that ε does not exceed 10% of the maximum distance between the points in the time series [48]. Another approach is to consider the recurrence-based measure as a signal detector and then choose the ε value that yields the maximum power from the signal for the detector being considered [49].

However, we find that keeping $\varepsilon \approx 10\%$ of the maximum distance often results in a 'dense' RP (Fig. 3.3(a)). Therefore, to capture the finer recurrence structures while still allowing for sufficient statistics in the CPR, we choose $\varepsilon = 0.1$ as the threshold for all RPs and normalize the time series (or any segments thereof) before the recurrence-based calculations. This effectively reduces the recurrence threshold to about 2% of the maximum distance and gives a 'clearer' RP (Fig. 3.3(b)). The RPs given in Fig. 3.4 were obtained with $\varepsilon = 0.1$, where it corresponds to around 2-3% of the maximum distance.

We find that the qualitative features of recurrences are robust to this choice

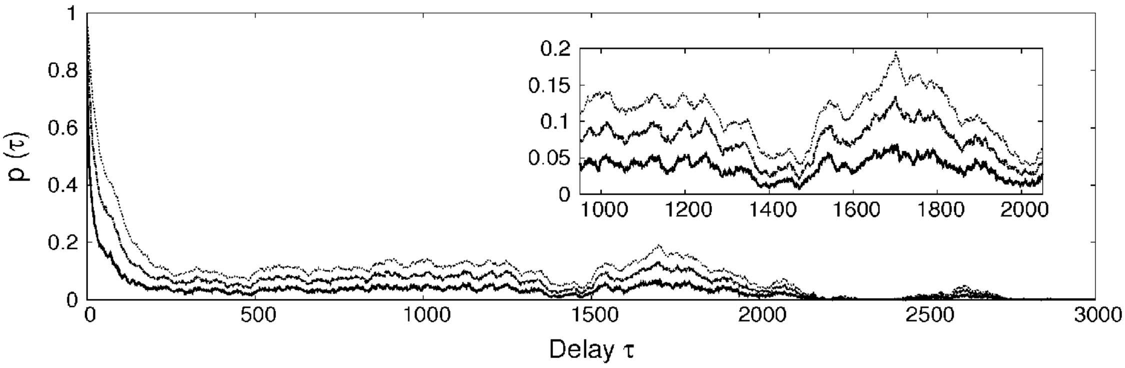

Figure 3.5: $p(\tau)$ **curves for Hang Seng index**. $\varepsilon = 1\%$ (bottom curve), $\varepsilon = 2\%$ (middle curve), and $\varepsilon = 3\%$ (top curve) of the maximum distance. All three curves have that same pattern. *Inset:* The $p(\tau)$ curves for $\tau = 950$ to $\tau = 2050$ to give a better view of the similar qualitative nature of the three curves.

of ε. In Fig. 3.5, the $p(\tau)$ curves for the Hang Seng index obtained from its RP (Fig. 3.4(h)) are shown for $\varepsilon = 1\%$, 2% and 3% of the maximum distance. Although the magnitude of recurrences increases with increasing threshold, the qualitative nature of the curves remains the same throughout.

The results obtained in this study are thus robust to small changes in ε.

Chapter 4

Cycles of connectivity

As mentioned in Chapter 3, we first visualize the network of connections given by CPR among all the nine stock markets (Sec. 4.1). The CPR values are thus estimated for the entire time series spanning almost two decades and the values obtained are binned in three categories:

i. *Strong connectedness* ($|CPR| > 0.8$),

ii. *Moderate connectedness* ($0.5 < |CPR| < 0.8$), and

iii. *Weak connectedness* ($|CPR| < 0.5$).

In the second part, a window of 250 time points was moved along pairs of time series and the CPR and ρ values were estimated (Secs. 4.2 and 4.5). This was done for all 36 pairs possible between the nine time series. Next, the number of connections (out of the 36) in each of the three bins mentioned above were counted for every window (Sec. 4.3).

Finally, in the last part, the behavior of CPR during the Dot-Com bubble of 2000 was considered (Sec. 4.4), in which: (a) taking a pair of time series, the dates when they peaked during the Dot-Com were made to coincide, (b) a window of 250 time points was moved from 500 time points before the peak to 250 time points after, and (c) the respective CPR (or ρ) was obtained.

At every step, all CPR (or ρ) values were tested for significance with 100 Twin Surrogates at 10% significance.

4.1 Network of connections

Tables 4.1 and 4.2 list the matrices of CPR and Pearson correlation respectively (labels from Table 3.1) for the entire time series from Dec-1990 to Apr-2010. Statistically non-significant values are listed as “NS”. We note that several values in Table 4.2 are non-significant, unlike those in Table 4.1.

The data contained in Tables 4.2 and 4.2 are visualized in Fig. 4.1, which shows the different connections that fall in each of the three bins of connectivity (strong, moderate and weak). NASDAQ has the highest number of connections in

Table 4.1: Matrix of CPR values

Label	A	B	C	D	E	F	G	H	I
A	1	0.492	0.716	0.927	0.898	0.878	0.685	0.700	0.493
B	0.492	1	0.766	0.689	0.444	0.799	0.746	0.764	0.242
C	0.716	0.766	1	0.867	0.583	0.792	0.768	0.746	0.455
D	0.927	0.689	0.867	1	0.863	0.922	0.819	0.862	0.467
E	0.898	0.444	0.583	0.863	1	0.717	0.592	0.769	0.647
F	0.878	0.799	0.792	0.922	0.717	1	0.781	0.779	0.412
G	0.685	0.746	0.768	0.819	0.592	0.781	1	0.795	0.283
H	0.700	0.764	0.746	0.862	0.769	0.779	0.795	1	0.428
I	0.493	0.242	0.455	0.467	0.647	0.412	0.284	0.428	1

Table 4.2: Matrix of Pearson correlation values

Label	A	B	C	D	E	F	G	H	I
A	1	0.931	0.954	0.922	0.905	0.949	NS	0.718	0.432
B	0.931	1	0.952	0.908	0.920	0.965	NS	0.754	NS
C	0.954	0.952	1	0.907	0.923	0.947	NS	0.838	0.535
D	0.922	0.908	0.907	1	0.861	0.925	NS	0.716	NS
E	0.905	0.920	0.923	0.861	1	0.980	-0.545	0.813	NS
F	0.949	0.965	0.947	0.925	0.980	1	-0.449	0.771	0.441
G	NS	NS	NS	NS	-0.545	-0.449	1	NS	NS
H	0.718	0.754	0.838	0.716	0.813	0.771	NS	1	0.828
I	0.432	NS	0.535	NS	NS	0.441	NS	0.828	1

the strong connectedness scenario (Fig. 4.1(a)). Also, Nikkei and Hang Seng are connected to NASDAQ strongly (Fig. 4.1(a)). Strait Times shares weak connections with all indices (Fig. 4.1(c)) except for DJIA, with which it has a moderate link (Fig. 4.1(b)).

Most of the connections fall in the moderate category (Fig. 4.1(b)), and it might be argued that this is not surprising as having a moderate connection might be a good strategy that markets have evolved over the years to insulate themselves against crises. However, one thing that must be kept in mind is that the picture in Fig. 4.1 is a static one — it does not say anything about how the connectivity scenario was before this time.

Moreover, this picture involves information for almost twenty years and might thus incorporate critical events of the past that might not (or, if we assume that they should not) have direct bearing on the connectivity trends of the present. This is why, in order to get a handle on the dynamic evolution of the connections between the stock markets, we slide a window over the entire length of the time series to construct temporal trends for both CPR and ρ, the results of which are discussed

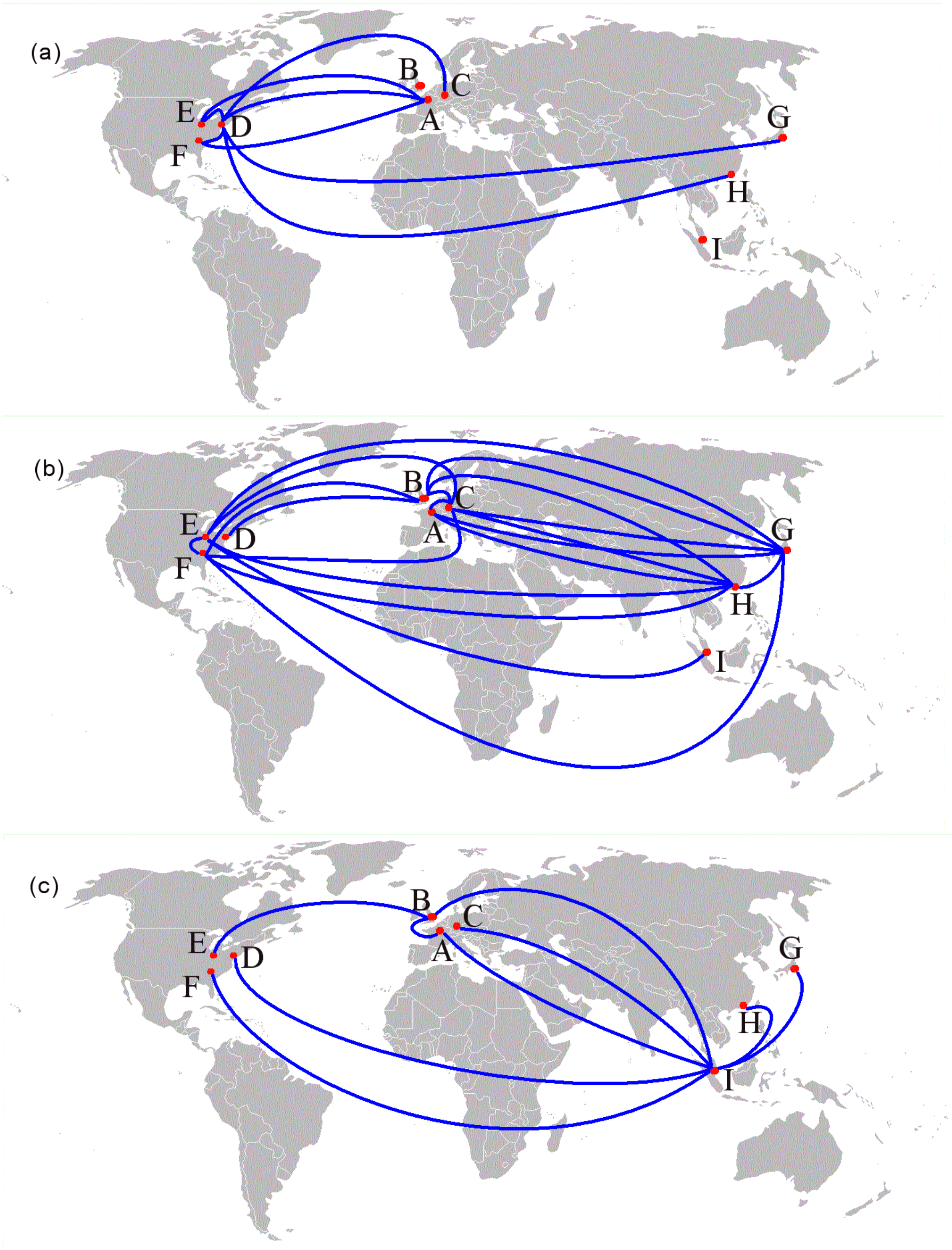

Figure 4.1: **Network of connections for stock indices using CPR**. (Labels A, B, ..., I are as in Table 3.1). (a) *Strong connectedness* ($CPR > 0.8$), (b) *Moderate connectedness* ($0.5 < CPR < 0.8$), and (c) *Weak connectedness* ($CPR < 0.5$).

in the following sections of this chapter.

4.2 Trends in CPR

Three points are evident from Fig. 4.2 (which shows trends in CPR for six pairs of indices along with the corresponding p-values, viz., (*a*) CAC 40 and FTSE 100, (*b*) DAX and NASDAQ, (*c*) DAX and Nikkei 225, (*d*) DJIA and S&P 500, (*e*) S&P 500 and Strait Times, and (*f*) Nikkei 225 and Hang Seng).

i. The CPR does not have a monotonous trend over time. Rather it oscillates erratically with short periods of low CPR interspersing broader bands of high CPR.

ii. The regions of low CPR values have high p-values (and vice versa), meaning that the lower ranges of CPR tend to be less statistically significant in comparison to the higher values.

iii. These patterns in the CPR are same for all index pairs, of which only six are shown in Fig. 4.2.

Thus, if CPR were to be interpreted as a measure of similarity, and hence connectedness, it means that the financial world is not moving monotonously to a globally connected scenario, but is rather oscillating between long periods of strong connectedness and short spans of low connectivity. (Nothing conclusive can be said about the nature of connections in the short periods of low CPR as the estimates therein are not statistically significant.) In a sense, what it says is that the more things change the more they remain the same.

4.3 Patterns of connectivity

The patterns in the number of connections in each of the three bins, summed up among all pairs of indices, reinforce the picture put forth in Fig. 4.2. This is shown in Fig. 4.3. The strong and moderate connections are always statistically significant except for a few instances. The weak connections are clearly more prone to be statistically non-significant. Again, we see that there is no global trend in Fig. 4.3, i.e., there is no indication for an increasing global connectivity. Instead, we see that the number of strong and moderate connections oscillate erratically, implying that these nine indices come close to each other, and then move apart, and then come close again, and so on. Had all of them moved to higher connectivity over time, the number of connections in the strong category would have increased progressively, and the numbers in the other two bins would have gone down with it.

The strong and moderate connections move (loosely) in phase with each other (see Fig. 4.4). Starting from the dip in Jan-1999, consecutive dips occur roughly at Dec-2000, Jul-2002, Sep-2003, Feb-2005, Sep-2006, Feb-2008, and Oct-2009, meaning that the periods of these dips lie between 14 to 20 months. This might be

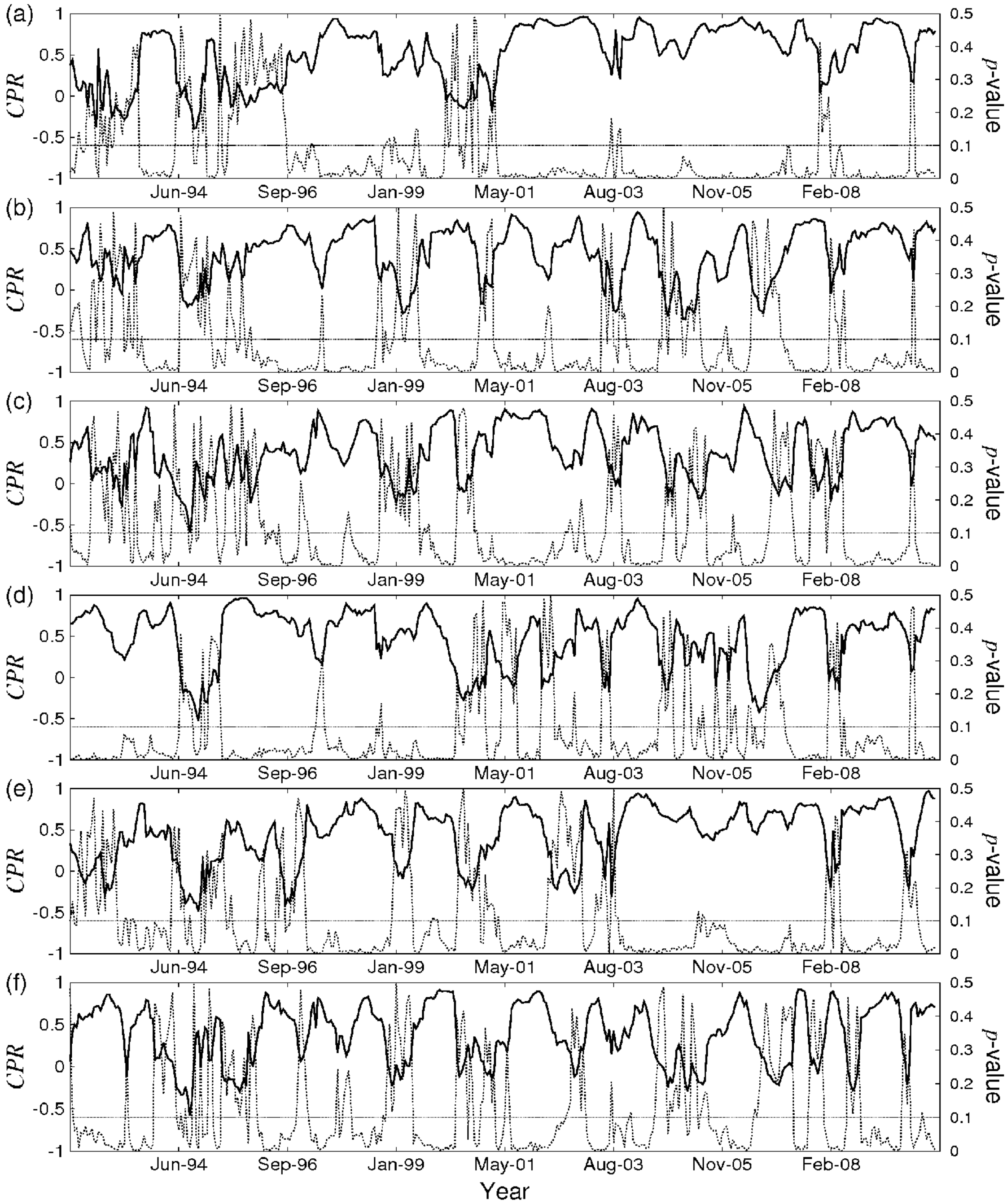

Figure 4.2: **Trends in CPR.** CPR (bold curve) and corresponding p-values (dotted curve) for six pairs. (a) CAC 40 and FTSE 100. (b) DAX and NASDAQ. (c) DAX and Nikkei 225. (d) DJIA and S&P 500. (e) S&P 500 and Strait Times. (f) Nikkei 225 and Hang Seng. Window size = 250 time points. Step size = 10 time points. The horizontal dotted line denotes the test significance level, $p = 0.1$.

indicative of the Kitchin business cycle [50].

Also, the number of 'weak' connections (Fig. 4.3(c)) is anti-phase to the number of 'strong' and 'moderate' connections (Figs. 4.3(a)-(b) and Fig. 4.4), because

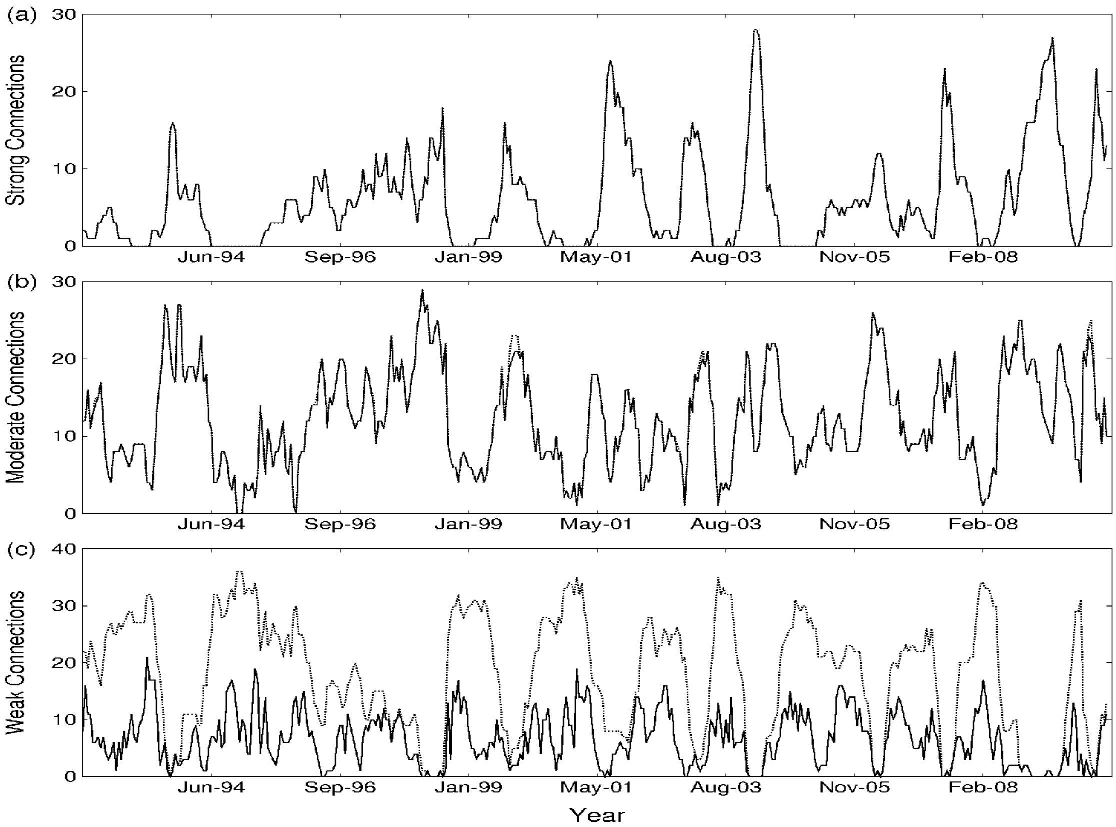

Figure 4.3: **Patterns of connectivity.** Number of connections in each bin along time. (a) Strong, (b) Moderate, and (c) Weak. The bold curves represent the number by counting statistically significant CPR values alone, while the lighter curves represent the number by counting all CPR values in each bin. Window size = 250 time points. Step size = 10 time points.

the total number of connections has to be conserved while the number in each of the three levels of connectedness go through cyclical patterns.

4.4 CPR in the Dot-Com bubble

Figs. 4.5 and 4.6 give an insight into the way stock indices approach a crisis and then recede from it. Even though the indices peaked and crashed on distinct days (sometimes months apart), once the time series is shifted to align the peak-off dates (Fig. 4.5(a)), the CPR increases around the peak-off date and decreases thereafter. For CAC 40 and Strait Times, the decrease in CPR starts almost at the peak-off date (Fig. 4.5(b)). On the other hand, for DAX and NASDAQ it occurs after the peak-off date (Fig. 4.6(a)) and for NASDAQ and DJIA it happens before (Fig. 4.6(b)). The increase-and-decrease of CPR is common to all of them. Thus, the probabilities of recurrence have strong correlation around the peak, i.e., irrespective of the actual date on which a particular index may peak, all indices approach and recede from a crisis similarly.

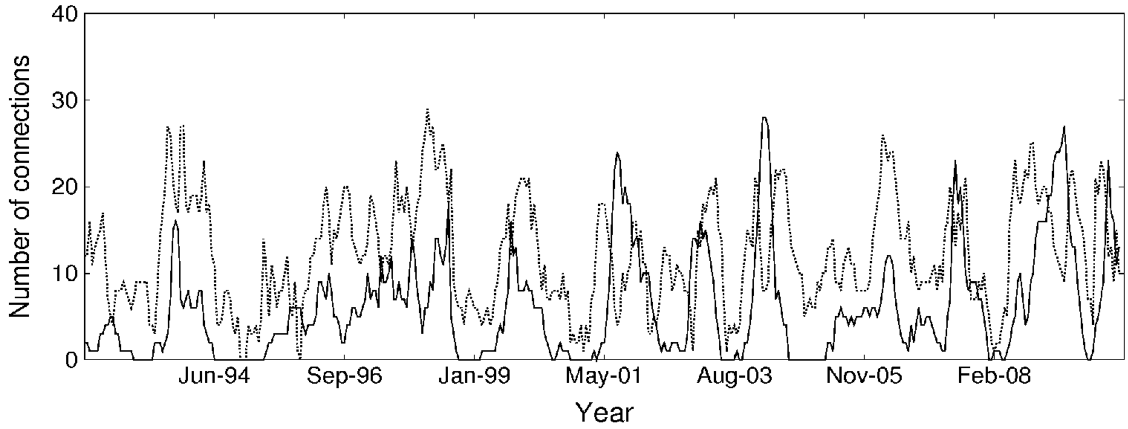

Figure 4.4: **Strong and moderate connections.** Strong (solid) and moderate (dotted) connections (from Fig. 4.3 (a) and (b)) shown together. Only statistically significant CPR values were counted.

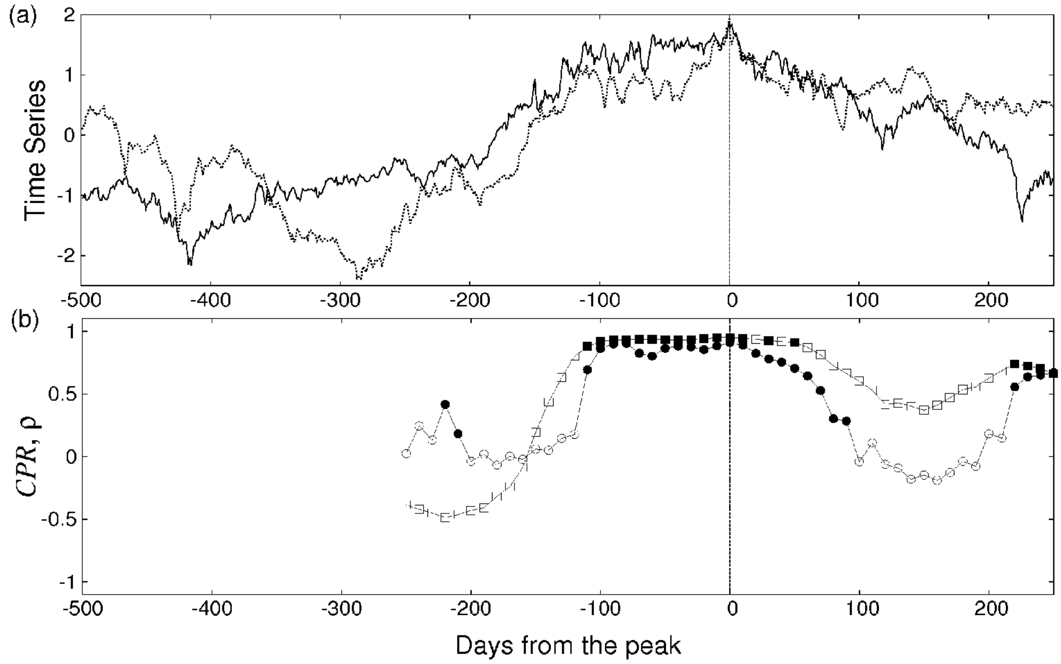

Figure 4.5: CPR **in the Dot-Com bubble (around the year 2000).** (a) The series CAC 40 (bold) and Strait Times (dotted) with their peak-off date during the bubble coinciding. The vertical dotted line marks the peak-off date. (b) Corresponding CPR (circles) and Pearson's ρ (squares) values. Statistically significant and non-significant values are represented by filled and empty markers respectively.

Pearson's ρ also captures this behavior, but CPR is a more sensitive measure, as ρ does not change as sharply as CPR around the peak-off date. Also, the CPR values tend to pass the statistical significance test more often than ρ (see Fig. 4.5(b) and Fig. 4.6). This point is further elaborated in Sec. 4.6.

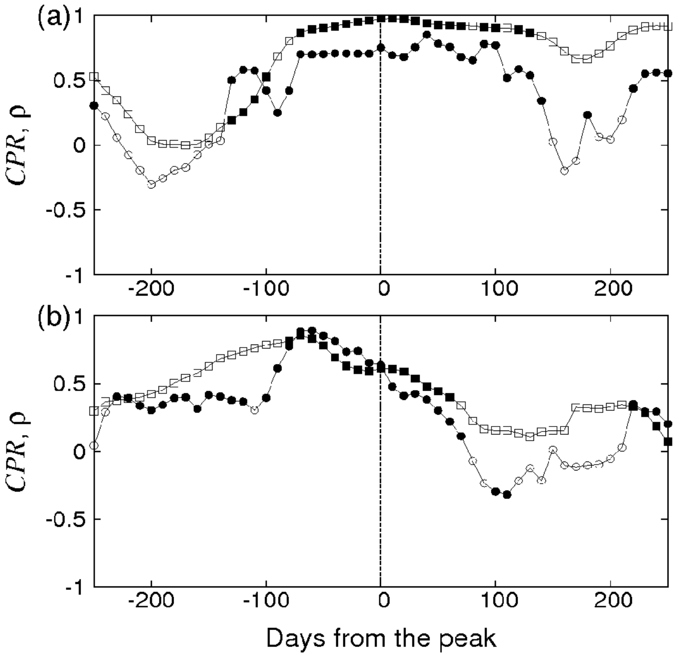

Figure 4.6: CPR **in the Dot-Com bubble.** (a) DAX vs. NASDAQ. (b) NASDAQ vs. DJIA. Figure keys and markers are the same as in Fig. 4.5.

4.5 Trends in Pearson's ρ

The trend in Pearson's ρ is strongly different from CPR, as shown for S&P 500 and Strait Times in Fig. 4.7 (which is the same pair as in Fig. 4.2(e)). Fig. 4.7 shows an overall movement towards higher correlation as time progresses, in contrast to the oscillating pattern of Fig. 4.2(e). The question then arises: *which is the correct picture?* Here, it is crucial to emphasize that these two measures capture different aspects of the time series. While ρ measures the tendency of the time series values to move together in one direction (or opposite directions), CPR measures the tendency

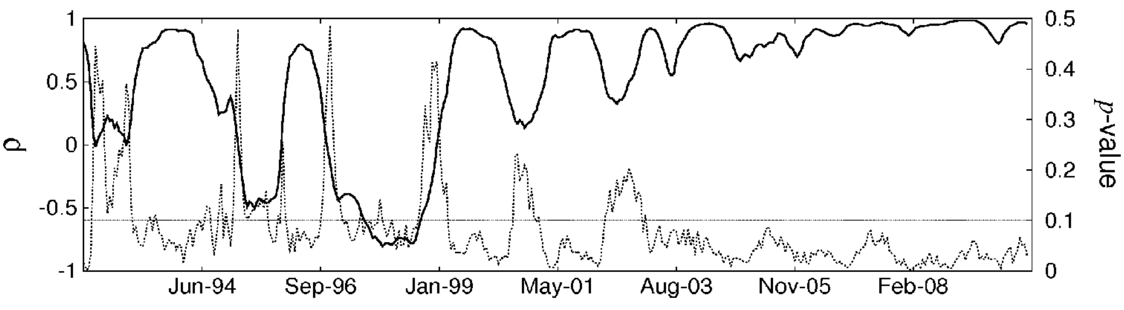

Figure 4.7: **Trends in Pearson's ρ.** Pearson's ρ (bold curve) and its corresponding p-values (dotted curve) for the pair S&P 500 and Strait Times. The horizontal dotted line denotes the test significance level, $p = 0.1$.

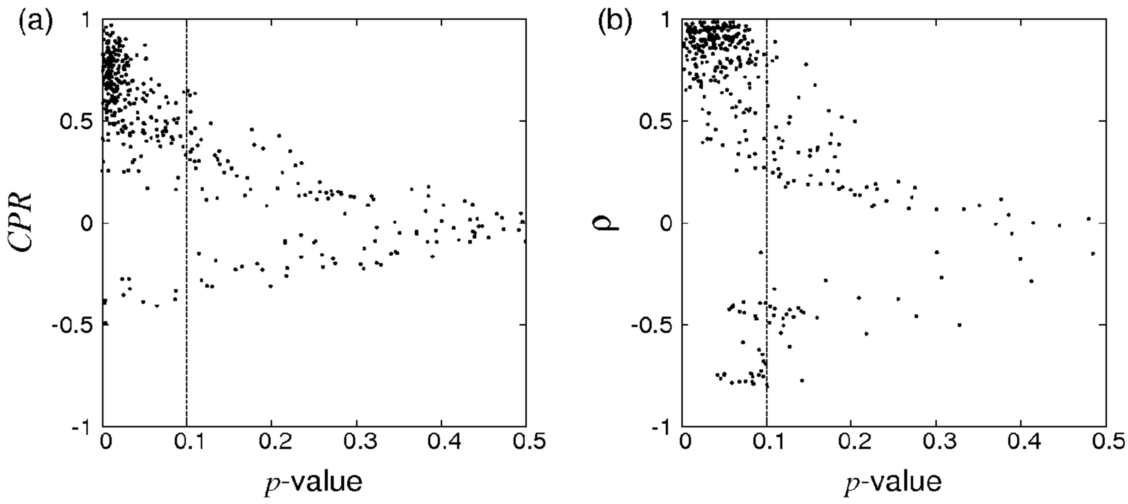

Figure 4.8: **Patterns of significance.** (a) CPR, and (b) Pearson's ρ, along with p-values.

of the time series values to return to earlier values at similar time scales. Hence, it is expected that their results are different. However, the case for CPR can be interpreted as follows: being based on recurrence rates it captures the essential dynamical nature of the system. This is in contrast to Pearson's ρ which is simply a statistical comparison of co-evolution of states. Pearson's ρ is less sensitive to changes in the time series as pointed out earlier. Thus, while it may be true that stock indices tend to show more co-movement towards the latter half of our data sets, this does not necessarily imply that they will continue to do so as the corresponding system dynamics move in and out of strong periods of correlation (see Sec. 4.2).

4.6 Advantages of CPR

We highlight the following advantages of CPR as a measure for estimating connections between financial data sets.

i. It is able to extract patterns even from noisy data sets, as is the case with most financial data. Moreover, it does not require that the data are distributed normally, as is required for the proper usage of ρ.

ii. The data sets need not necessarily be embedded for its estimation.

iii. It can be estimated for short time series as well, as is done in the current work.

iv. It does not require high-frequency sampling of the data, which is common in most financial analyses. Our analysis was done with daily sampled data which was freely available on the internet.

v. It tends to have lower p-values than ρ, as seen from the spread of points in the top-left corners of the plots in Fig. 4.8. Most CPR values are grouped close to the $p = 0$ axis, whereas the ρ values have a broader spread. This implies that the probability of rejecting the null hypothesis tends to be lower with CPR.

CPR not only has several advantages over Pearson's ρ as a measure that captures the similarity between two financial time series, but it also tells us that Pearson's ρ is not the only way to look at correlations between markets, along with the critical fact that the stock markets of the world might not be moving in any general direction at all. This provides a crucial understanding of the financial world and reinforces the notion that financial markets are, in fact, complex systems that are not amenable to trivial conclusions.

Chapter 5

Conclusion

The primary goal of this study had been to extend the domain of application of recurrence plots as a tool for nonlinear data analysis by using it to analyze connections among financial data sets, based on recent developments in recurrence-plot-based theory. In doing so it has been shown that recurrence-plot-based measures have the potential to extract crucial dynamical information about the interrelations between financial systems. It is hoped that in the light of these results, newer possibilities for data analysis in the financial world shall be opened up.

5.1 Primary findings

5.1.1 Connectivity cycles

The main result obtained in this study is that the stock markets of the world are not necessarily progressing towards higher connectivity monotonically, as might be trivially assumed by most people in light of the perception that the financial world is getting increasingly globalized with each passing day. The CPR provides an alternative view of the interrelations among the stock indices which shows that the global markets mostly co-exist in periods of high connectivity which are interspersed irregularly by short periods of low connectivity. This has been the characteristic feature of all pairs of stock indices chosen and for the entire length of the time span considered in the analysis. This was contrasted to the image of connectivity trends given by Pearson's correlation coefficient, which showed market pairs having a trend towards higher and higher correlation as time progressed from 1990 to 2010. The contrast among these two pictures tells us that Pearson's correlation does not capture all aspects of similarity and co-movement and thus, for a more wholesome perspective of the connections among the financial markets, other measures are necessary.

5.1.2 Similarity during crises

The results obtained from the analysis of the behavior of CPR during the Dot-Com bubble of 2000 showed the way in which financial markets approach and recede from a financial crisis. In shifting the time series to align their peak-off dates and then measuring the temporal change in CPR, it was found that CPR had high values around the peak-off date. This showed that irrespective of the historic nature of the stock index, its global location, and the exact date at which it finally starts to crash, the way in which any two stock markets behaved during a financial crisis was greatly similar, implying that their underlying dynamical systems share similar properties during these critical periods. This feature was also captured by Pearson's correlation but CPR turned out to be a more sensitive measure.

5.2 Methodological aspects

5.2.1 Extension of CPR

A crucial aspect of this study is that the definition of the CPR had been modified slightly before being used in the analysis. This novel modification was motivated by the fact that all probability of recurrence curves start from 1 and this introduces a bias towards a higher correlation value between two probability of recurrence curves. In considering the probability of recurrence for only those values of the time delay which lay beyond the maximum autocorrelation time of the system, the correlation CPR among them was made more robust and a better representative of the underlying similarities of the time scales of recurrence of the two sub-systems.

5.2.2 Significance tests with Twin Surrogates

Another objective of this study was to stress on the need to do significance tests on the measures that are obtained from financial time series. This fact was highlighted because significance tests are not frequently used in financial time series analysis. Also, the use of the Twin Surrogates method of generating surrogate data sets for the significance tests was pointed out as a better way of generating surrogates. This was because Twin Surrogates retain linear as well as nonlinear characteristics of the original time series and correspond to different trajectories of the same system differing only in the initial conditions; and thus it is equivalent to sampling data sets from the same dynamical system from nature itself, which is the fundamental idea behind using a surrogate data set in the first place.

5.2.3 Advantages of using Recurrence Plots

CPR, being a recurrence-plot-based measure, has some of the inherent advantages that come with the use of recurrence-based data analysis. It does not have any limitations on the distribution of the time series values, nor does it need the data sets to be stationary. It works well with noisy data sets as well. The analysis can be done with or without embedding depending on the need and objectives of the study and the exact nature of the data sets. CPR also works well with data series that are very short and that have low frequency of sampling. It does not even need the sampling interval to be constant throughout the length of the times series. In fact, in our case, the data sets finally used in the analysis were short, irregularly sampled, had a minimal sampling frequency of one day, contained systemic and observational noise and were not embedded. We carried out the entire study with freely available data sets obtained from the internet. That the CPR manages to extract patterns even in such adverse conditions is a commendable feature of this measure.

5.3 Future perspectives

To the best of our knowledge, this is the first study in which the modified version of the CPR has been applied to stock indices. This opens up newer possibilities of extending this measure to other financial data sets such as individual stocks within a particular index, foreign exchange rates, and commodities indices. Considering this only as a test run for the CPR, a next step would be to include a larger number of stock indices that spans all the continents and then create a dynamic picture of the interrelations among them. This study also sets the stage for the use of other recurrence-based synchronicity measures to be also applied in the financial world for a better understanding of its interrelations.

Another interesting aspect would be to use CPR to define a 'distance' among the different financial data sets and construct minimum spanning trees from them, as has been done in numerous econometric studies based on the Pearson's correlation coefficient. The network based on these distances would clearly be different from the ones obtained from Pearson's correlation and their temporal evolution might throw new light on the nature and structure of financial data sets.

Lastly, it would be also be interesting to try and devise model systems that would replicate the behavior of the financial time series such as those used in this study, i.e., the model should give rise to time series for which the Pearson's ρ shows an increasing trend over time and the CPR shows erratic cycles of connectivity. In the ideal case, one would want a single system, which under different parameter values, give rise to the different dynamics observed in all of the nine time series considered here. If this were achieved, it would grant us a better understanding of financial systems on the whole.

References

[1] H. Poincaré, Sur la probleme des trois corps et les équations de la dynamique, Acta Mathematica 13 (1890) 1–271.

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

[3] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, Europhys. Lett. 4 (9) (1987) 973–977.

[4] J. P. Zbilut, M. Koebbe, H. Loeb, G. Mayer-Kress, Use of recurrence plots in the analysis of heart beat intervals, in: Proceedings of the IEEE Conference on Computers in Cardiology, Chicago,1990, IEEE Computer Society Press, 1991, pp. 263–266.

[5] L. L. Trulla, A. Giuliani, J. P. Zbilut, C. L. Webber, Jr., Recurrence quantification analysis of the logistic equation with transients, Phys. Lett. A 223 (4) (1996) 255–260.

[6] J. P. Zbilut, A. Giuliani, C. L. Webber, Jr., Recurrence quantification analysis and principal components in the detection of short complex signals, Phys. Lett. A 237 (3) (1998) 131–135.

[7] J. P. Zbilut, A. Giuliani, C. L. Webber, Jr., Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification, Phys. Lett. A 246 (1-2) (1998) 122–128.

[8] J. P. Zbilut, A. Giuliani, C. L. Webber, Jr., Recurrence quantification analysis as an empirical test to distinguish relatively short deterministic versus random number series, Phys. Lett. A 267 (2-3) (2000) 174–178.

[9] J. P. Zbilut, J.-M. Zaldivar-Comenges, F. Strozzi, Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data, Phys. Lett. A 297 (3-4) (2002) 173–181.

[10] K. Shockley, M. Butwill, J. P. Zbilut, C. L. Webber, Jr., Cross recurrence quantification of coupled oscillators, Phys. Lett. A 305 (1-2) (2002) 59–69.

[11] J. P. Zbilut, N. Thomasson, C. L. Webber, Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals, Med. Engg. & Phys. 24 (1) (2002) 53–60.

[12] C. L. Webber, Jr., J. P. Zbilut, Recurrence quantification analysis of nonlinear dynamical systems, in: M. A. Riley, G. V. Orden (Eds.), Tutorials in contemporary nonlinear methods for the behavioral sciences, Retrieved April 6, 2010, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.pdf, pp. 26–94.

[13] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence-plot-based measures of complexity and their application to heart-rate-variability data, Phys. Rev. E 66 (2) (2002) 026702.

[14] M. Thiel, M. C. Romano, P. L. Read, J. Kurths, Estimation of dynamical invariants without embedding by recurrence plots, Chaos 14 (2) (2004) 234–243.

[15] M. Thiel, M. C. Romano, J. Kurths, How much information is contained in a recurrence plot?, Phys. Lett. A 330 (5) (2004) 343–349.

[16] M. C. Romano, M. Thiel, J. Kurths, I. Z. Kiss, J. L. Hudson, Detection of synchronization for non-phase-coherent and non-stationary data, Europhys. Lett. 71 (3) (2005) 466–472.

[17] R. N. Mantegna, H. E. Stanley, Physics investigation of financial markets, in: F. Mallamace, H. E. Stanley (Eds.), The Physics of Complex Systems [Proceedings of the International School of Physics "Enrico Fermi" Course CXXXIV], IOS Press, Amsterdam, 1997, pp. 473–490.

[18] G. Bonanno, F. Lillo, R. N. Mantegna, Levels of complexity in financial markets, Physica A 299 (1-2) (2001) 16–27.

[19] L. Cao, A. S. Soofi, Nonlinear deterministic forecasting of daily dollar exchange rates, Int. J. Forecast. 15 (4) (1999) 421–430.

[20] A. Carbone, G. Castelli, H. E. Stanley, Time-dependent Hurst exponent in financial time series, Physica A 344 (1-2) (2004) 267–271.

[21] R. N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1) (1999) 193–197.

[22] R. N. Mantegna, H. E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, 2000.

[23] J.-P. Bouchaud, M. Potters, More stylized facts of financial markets: leverage effect and downside correlations, Physica A 299 (1-2) (2001) 60–70.

[24] S. Drożdż, F. Grümmer, F. Ruf, J. Speth, Towards identifying the world stock market cross-correlations: DAX versus Dow Jones, Physica A 294 (1-2) (2001) 226–234.

[25] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, Dynamic asset trees and Black Monday, Physica A 324 (1-2) (2003) 247–252.

[26] G. Bonanno, G. Caldarelli, F. Lillo, S. Miccichè, N. Vandewalle, R. N. Mantegna, Networks of equities in financial markets, Eur. Phys. J. B 38 (2) (2004) 363–371.

[27] J. Kwapień, S. Drożdż, J. Speth, Time scales involved in emergent market coherence, Physica A 337 (1-2) (2004) 231–242.

[28] D. Wilcox, T. Gebbie, On the analysis of cross-correlations in South African market data, Physica A 344 (1-2) (2004) 294–298.

[29] T. Aste, T. D. Matteo, M. Tumminello, R. N. Mantegna, Correlation filtering in financial time series (Invited Paper), in: D. Abbott, J.-P. Bouchaud, X. Gabaix, J. McCauley (Eds.), Noise and Fluctuations in Econophysics and Finance, Vol. 5848 of Proc. SPIE, SPIE, Bellingham and WA, 2005, pp. 100–109.

[30] J.-P. Onnela, J. Saramäki, K. Kaski, J. Kertész, Financial market - a network perspective, in: H. Takayasu (Ed.), Practical Fruits of Econophysics, Springer Tokyo, 2006, pp. 302–306.

[31] R. Coelho, C. G. Gilmore, B. Lucey, P. Richmond, S. Hutzler, The evolution of interdependence in world equity markets—Evidence from minimum spanning trees, Physica A 376 (2007) 455–466.

[32] R. Coelho, S. Hutzler, P. Repetowicz, P. Richmond, Sector analysis for a FTSE portfolio of stocks, Physica A 373 (2007) 615–626.

[33] S. Çukur, M. Eryiğit, R. Eryiğit, Cross correlations in an emerging market financial data, Physica A 376 (2007) 555–564.

[34] C. Eom, G. Oh, S. Kim, Deterministic factors of stock networks based on cross-correlation in financial market, Physica A 383 (1) (2007) 139–146.

[35] M. Tumminello, C. Coronnello, F. Lillo, S. Micciché, R. N. Mantegna, Spanning trees and bootstrap reliability estimation in correlation-based networks, Int. J. Bifurc. Chaos 17 (7) (2007) 2319–2329.

[36] T. Conlon, H. J. Ruskin, M. Crane, Cross-correlation dynamics in financial time series, Physica A 388 (5) (2009) 705–714.

[37] D. Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R. N. Mantegna, E. Ben-Jacob, Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market, PLoS ONE 5 (12) (2010) e15032.

[38] F. Strozzi, J.-M. Zaldívar, J. P. Zbilut, Application of nonlinear time series analysis techniques to high-frequency currency exchange data, Physica A 312 (3-4) (2002) 520–538.

[39] A. Antoniou, C. E. Vorlow, Recurrence quantification analysis of wavelet pre-filtered index returns, Physica A 344 (1-2) (2004) 257–262.

[40] A. Fabretti, M. Ausloos, Recurrence analysis near the NASDAQ crash of April 2000, in: H. Takayasu (Ed.), Practical Fruits of Econophysics, Springer Tokyo, 2006, pp. 52–56.

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

[42] J. A. Bastos, J. Caiado, Recurrence quantification analysis of global stock markets, Physica A 390 (7) (2011) 1315–1325.

[43] P. M. Crowley, Analyzing convergence and synchronicity of business and growth cycles in the euro area using cross recurrence plots, Eur. Phys. J. Spec. Top. 164 (1) (2008) 67–84.

[44] M. Thiel, M. C. Romano, J. Kurths, M. Rolfs, R. Kliegl, Twin surrogates to test for complex synchronisation, Europhys. Lett. 75 (4) (2006) 535–541.

[45] J. Imbs, Trade, finance, specialization, and synchronization, Rev. Econ. Stat. 86 (3) (2004) 723–734.

[46] C. W. J. Granger, Some properties of time series data and their use in econometric model specification, J. Econom. 16 (1) (1981) 121–130.

[47] R. F. Engle, C. W. J. Granger, Co-integration and error correction: Representation, estimation, and testing, Econometrica 55 (2) (1987) 251–276.

[48] J. P. Zbilut, C. L. Webber, Jr., Embeddings and delays as derived from quantification of recurrence plots, Phys. Lett. A 171 (3-4) (1992) 199–203.

[49] S. Schinkel, O. Dimigen, N. Marwan, Selection of recurrence threshold for signal detection, Eur. Phys. J. Spec. Top. 164 (1) (2008) 45–53.

[50] J. Kitchin, Cycles and trends in economic factors, Rev. Econ. Stat. 5 (1) (1923) 10–16.