Alternative Risk Management: Correlation and Complexity

Thomas Conlon

BA (Mod) Theoretical Physics

MSc Financial Mathematics

A thesis submitted in fulfilment of the

requirements for the award of

Doctor of Philosophy (Ph.D.)

to the

Dublin City University

Faculty of Engineering and Computing, School of Computing

Supervisors: Prof. Heather J. Ruskin, Dr. Martin Crane

August 26, 2009

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed _____

Thomas Conlon

Student ID **55150209**

Date **August 26, 2009**

ACKNOWLEDGEMENTS

I would like to express my gratitude to everyone who helped make it possible to complete this thesis.

First, I would like to thank my supervisors, Prof. Heather Ruskin and Dr. Martin Crane for their guidance over the past three and a half years. A visit to them was always enough to inspire and reveal new paths worthy of exploration. They have made this journey far less difficult by imparting their knowledge, experience and advice.

I am forever indebted to my parents for their support, encouragement and patience over the years. I'm sure when I started school in Donegal 26 years ago, they didn't think they'd be retired before I finished! Many thanks also to my brothers, sister and my friends for all their words of encouragement and lots of interesting conversations.

Last but not least, I would like to thank my wife-to-be, Susie, for everything but most especially for her words of encouragement, her patience and support. We're getting ever closer to a day when we don't have any study to do!

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In recent years, spatiotemporal synchronisation within systems with multiple interacting components (Complex Systems), such as financial data, electroencephalographic (EEG) recordings and magnetoencephalographic (MEG) recordings, has been studied extensively, using the equal-time cross-correlation matrix. These Complex Systems are characterised by events such as Market Crashes or Seizures, which are associated with periods of hyper-synchronisation.

In this Thesis, the Risk Characterisation and Reduction of Complex Systems is studied, using the Cross-Correlation matrix to condense the system complexity. The systems studied display interactions between multivariate time series of varying granularities, including low frequency (Hedge Fund returns), medium frequency (Daily Stock returns) and high frequency (Intraday Stock returns & EEG seizure data).

The information content of the correlation matrix between low-frequency Hedge Fund returns is investigated for the first time using Random Matrix Theory (RMT). The RMT filtered correlation matrix is shown to improve the risk-return profile of a portfolio of Hedge Funds. Through the use of the Wavelet transform, scaling properties of correlations are then investigated, with correlations calculated over longer horizons found to result in a better risk-return profile for a portfolio of Hedge Funds.

Characterisation of market risk is then assessed, through the dynamics of the correlation structure and associated eigenspectrum for daily equity returns (medium frequency data), using a moving window approach. This novel characterisation, dependent on both large and small eigenvalue behaviour, is shown to be consistent across different time scales. Further, frequency dependent correlations were examined for medium and high-frequency intra-day stock returns using Wavelet multiscaling.

Investigation of a comparative system example, specifically correlation scaling characteristics of high-frequency EEG Seizure data, revealed novel frequency dependent changes in the correlation structure between channels, which may be indicative of seizures. Large correlations were found between channels at high frequencies and conversely, smaller correlations at low frequencies during Seizures, with a corresponding switch in system energy.

Our findings suggest that, even for the limited set of examples chosen, diverse applications demonstrate commonality, in terms of the interpretative power of time-series correlation structure. Through the integration of tools such as Random Matrix Theory, Wavelet multiscaling and eigenvalue analysis, we have shown the importance of the correlation matrix in risk characterisation and reduction. The potential for wider application of these methods in the detection of subtle triggers, giving advance warning of risky events, has also been demonstrated.

CHAPTER 1

INTRODUCTION

1.1 Motivation

Vast losses across multiple markets, spanning various countries, sectors and instruments over the past two years have resulted in a necessary focus on Financial Risk Management. By *Risk Management*, we refer to the *quantitative assessment and mitigation* of unnecessary risks. However, Risk Management is not limited to financial applications. In fact, there is a need to reduce risk associated with many so-called Complex Systems, including but not limited to the risk of deaths arising from natural disasters such as earthquakes and hurricanes, the risk of injury during epileptic seizures and the risk of financial loss from trading.

These systems, although diverse, have many common features, including multiple interacting components, dynamical system changes and an emergence of new properties as the system evolves. In all cases, it is possible to use the principle of diversification to reduce the risk of the system, due to the multiple components present. In the case of earthquakes and hurricanes, relocating people to many different locations makes it possible to reduce the risk of massive loss of life in one location; equally, the use of medicines and surgical treatments for seizures can reduce the chances of hypersynchronisation between channels, said to be the underlying cause of seizures. Financial risks can also be mitigated by investing in a number of *uncorrelated* assets.

In order to achieve this diversification it is necessary to measure the common dynamical properties between the various interacting components. One simple measure of this

synchronisation is the Pearson cross-correlation coefficient between the components. However, there are a number of problems with the calculation of this coefficient. The data available may be sparse, which increases the influence of noise in the correlations. Also, the interactions between the components may occur at frequencies other than that at which they are measured, resulting in hidden correlations.

In addition to its contribution to diversification, the synchronisation between various components may, itself, be a characteristic of changes in the risk profile of a system. Further, there may also be subtle changes in the synchronisation structure that may be used as an early warning to possible adverse changes in the system, such as the occurrence of an earthquake or a financial crash.

1.2 Objectives

The purpose of the research presented here is twofold. First, to demonstrate the use of the correlation matrix for diversification purposes, in particular for sparse data. Second, to show how the correlation matrix can be used to characterise the risk of extreme events through the analysis of the associated eigenspectrum. Throughout the research we analyse data with varying granularity, from low-frequency Hedge Fund data to high-frequency equity data. To demonstrate the cross-applicability of risk management methods to other complex problems, we also study high-frequency EEG seizure data. To achieve our objective of improving risk-management within Complex Systems, we set out the following specific objectives:

- To demonstrate the application of the correlation matrix for diversification purposes in a low-frequency environment, using Hedge Fund returns data.
- To investigate the effects of granularity on the calculation of the cross-correlation matrix, by testing the effects on risk-management for a portfolio of hedge-funds over different scales.
- To test the use of the correlation matrix as a technique to characterise changes in risk within complex interacting systems. To this end, we examine dynamics of the

eigenspectrum associated with the correlation matrix for medium-frequency equity returns.

- To examine the dependence of dynamic correlation changes on the granularity of the data used in the calculation of the matrix.

- To investigate the use of a combination of the above techniques, such as correlation dynamics and scaling to an alternative system, that of high-frequency EEG seizure data. We seek to characterise the behaviour of the system over different frequencies to manage the risk inherent in epileptic seizures and explore early warning potential.

1.3 Outline of Thesis

The thesis is organised as follows: In **Chapter 2**, we outline the background to the research, review previous work and introduce some of the ideas followed in the thesis. The methods, both known and developed, which are used throughout are introduced in **Chapter 3**.

In **Chapter 4**, we investigate the use of the correlation matrix in the risk management of a portfolio of Hedge Funds. Random Matrix theory is applied to filter noise from the correlation matrix and the information present in the large eigenvalues is investigated. Using a classical portfolio optimisation, the benefits of filtering a correlation matrix, constructed using low-frequency Hedge Fund data, are determined.

Wavelet multiscaling techniques are applied in **Chapter 5**, to establish the effects of coarse/fine graining on the correlation matrix. Using low-frequency Hedge Fund returns, we examine changes in the correlation structure between funds and the S&P 500 over different time frames. The scaled correlation matrices are then used as inputs to a portfolio optimisation, to judge the effect of time granularity on Risk Management.

The dynamical changes in correlation between medium frequency financial returns are studied in **Chapter 6**, to investigate whether changes in the correlation structure can be applied in risk characterisation. The relationship between index returns and relative eigenvalue size is examined, to provide insight on the *collective behaviour of traders*. A simple

'toy model' of correlations is proposed to shed light on the formation of correlation structure.

The effects of coarse/fine graining on dynamical changes in correlation structure are examined in **Chapter 7**, to determine risk characteristics of scaling. This provides insight on the time horizons which apply to the numerous constituents involved in the interactions and gives some understanding of the scales involved during financial crashes.

In **Chapter 8** we examine a broadly comparative non-financial system made up of interacting components, specifically EEG Seizure data. Using a combination of methods applied in previous Chapters, such as correlation dynamics and wavelet multiscaling, we seek to identify frequency dependent changes in the correlation structure between EEG channels. The characteristic changes found may, potentially, then be used in the Risk Management of seizure events.

Finally, in **Chapter 9** we provide a summary of the work, our conclusions and discuss future improvements.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

A well known adage among traders refers to diversification as the only "Free Lunch" in finance. By this, it is meant that by spreading one's available capital among a number of different assets it is possible to reduce the risk of a portfolio. Consider an economy with only two companies, an umbrella company and a suncream company. When the weather is wet the umbrella company performs well; conversely, when the weather is sunny, it will perform badly. As the suncream company will have the reverse performance, one can minimise the weather dependent risk by splitting the investment between the two companies.

Given time-series data of company returns, one measure of the synchronisation between the performance of two companies is the Pearson correlation coefficient, [Crawley, 2005]. The Pearson correlation is merely a *linear measure* of the synchronisation and takes no account of any higher order relationships. However, as demonstrated below, methodology based upon this measure provides a firm basis for the risk management of financial assets and other complex interactions such as Neurological systems.

In this Chapter, we first introduce the scientific field of Complex Systems by describing some of the common characteristics found in nature, society and science. We then provide evidence of the interdisciplinary nature of Complex Systems research, by focusing on the application of ideas from Statistical Physics to Economic systems, namely *Econophysics*. The noise reduction of correlation matrices, using Random Matrix Theory, is described in combination with Eigenvalue analysis to demonstrate how genuine correlation information

can be separated from noise in the cross-correlation matrix. The use of the cross-correlation matrix to characterise the risk of a Complex System is subsequently considered for both financial and EEG Seizure data. Risk decomposition provides a critical assessment of the scaling effects within financial systems and we review its potential using Multiscale Wavelet analysis. Finally, we discuss in detail properties associated with one of the datasets examined in this study, Hedge Fund returns data. In the current financial climate Hedge Funds are of particular interest, due to their optimistic claim of *capital preservation in all market conditions*.

2.2 Complex Systems

In this Section, we attempt to elucidate the idea of a Complex Dynamic System. One of the difficulties with this is the sheer number of definitions of Complex Systems that exist. To overcome this problem, we define a Complex Dynamical System as one that displays many if not all of the following properties, [Bar-Yam, 2003; Johnson *et al.*, 2003; Corning, 2003; Miller and Page, 2007]:

- *Agent-Based*

  The basic building blocks are the characteristics and activities of individual agents

- *Dynamic*

  Properties change over time, often in a nonlinear way

- *Interactions*

  Agents interact, often in a non-linear fashion

- *Organisation*

  Group or Hierarchical structures are displayed

- *Feedback*

  Changes are often as a result of feedback from the environment

- *Heterogeneous*

  Agents may differ in significant characteristics

- *Emergence*

  The arising of novel and coherent structures, patterns and properties during the process of evolution and/or self-organisation

Systems with the above characteristics are found in diverse disciplines, including Seismology, Neuroscience, Economics, Meteorology, Molecular Biology and Social Sciences, [Aki and Richards, 1997; Kelso, 1997; Bouchaud and Potters, 2003; Holton, 2004; Kaneko, 2006; Miller and Page, 2007]. The cross-fertilisation of ideas between these disparate fields means that the interdisciplinary nature of Complex System analysis is well established. In this thesis, we examine two particular examples of Complex System, arising respectively in Economics and Neuroscience where many of the characteristics outlined above are evident. We focus, in particular, on Risk Management, associated with Financial markets, and EEG seizures, through the measurement of the level of interaction between 'agents'[1]. To this end, we examine the use of the linear cross-correlation matrix as a measure of the degree of synchronisation within complex interacting systems.

2.3 Econophysics

The application of theories and methods developed by Physicists to other disciplines is not a new idea. The fields of Biophysics and Geophysics, for example, are mature and well-researched areas where physical processes are studied in Geological or Biological systems [Lowrie, 1997; Glaser, 2000]. *Econophysics* is an interdisciplinary research field, applying physical methods to problems in Economics. The first explicit use of the word dates to a "Workshop on Econophysics" held in Budapest in 1997, [Kertesz and Kondon, 1998]. In the relatively short time since this conference a variety of books have been written on the topic [Roehner, 2002; Voit, 2003; Bouchaud and Potters, 2003; McCauley, 2004; Mantegna and

[1] In this context of this thesis, the term 'agents' refers either to the various participants or traders involved in financial markets or to the neurons of the brain

Stanley, 2005; Malevergne and Sornette, 2005; Kleinert, 2006], with many more in preprint. A simple google search for 'Econophysics' (30th January 2009) brings up $165,000$ results, [Google Search, 2009].

To consider the underpinning Mathematical theory of finance in more detail, we go back much earlier, [Davis, 2001]. As early as 1900, Bachelier, [Bachelier, 1900], attempted to explain the fluctuations of financial markets by introducing the theory of random walks, an idea later developed by Einstein to explain Brownian motion. This idea was revisited by Mandelbrot in 1963, when it was shown that fluctuations in cotton prices followed a distribution that differed from a Gaussian. Mandelbrot's 1963 paper, [Mandelbrot, 1963], is now regarded as one of the crucial precursors to the field of Econophysics. Physicists, working in the field of Statistical Mechanics, first began publishing papers in Econophysics in the early 1990's. One of the driving forces behind the application of ideas from statistical physics to Economic systems was the large amount of financial data available, starting from the 1980's. A general treatise of the origin and a basic introduction to some of the ideas pursued to date in Econophysics can be found in [Gligor and Ignat, 2001; Bouchaud, 2002; De Liso and Filatrella, 2002; Yakovenko, 2003; Burda *et al.*, 2003; Wang *et al.*, 2004; Vasconcelos, 2004].

While Economists start with a few fundamental assumptions and construct a model to explain observations, Econophysicists tend to begin with the empirical evidence and extract perceived patterns from the data. The applicability of power law distributions to Economic data is probably the most studied example of this. Physicists observe and have reported on power law distributions in many problems in statistical mechanics as well as phenomena as varied as city populations, internet sites and the levels of ocean tides, [Blank and Solomon, 2000; Barbosa *et al.*, 2006]. Power laws have been applied, with very interesting results, to financial data also. The distribution of relative price changes has, for example, been shown to be non-Gaussian and various authors have attempted to characterise these using distributions such as Truncated Levy and the Student's-t [Bouchaud and Potters, 2003]. The fact that relative price changes are non-Gaussian distributed has a profound effect on

the pricing of options[2], [Bouchaud, 2002]. The Black-Scholes theory of option pricing is underpinned by the assumption of Gaussian returns, while behaviour unexplained by this model, such as the 'volatility smile' [3], may be described by power laws, [Bouchaud, 2002]. The distribution of returns for financial crashes associated with speculative bubbles has also been studied in some detail, [Sornette, 2002; Rotundo and Navarra, 2007]

2.4 Correlation Matrices and Noise Reduction

A further early application of ideas from Physics to Finance, was that of noise reduction in the correlation matrix between asset returns. Correlations of returns between financial assets play a central role in Finance, particularly as inputs to Markowitz's classical portfolio optimisation problem, (Chapter 3.3.1), [Markowitz, 1958; Elton and Gruber, 2002]. However, the calculation of a correlation matrix for a large portfolio of size N is poorly determined unless the length of the time-series, T, is much greater than N. The finite length of the sample of returns data used in the calculation of correlation matrices means that these are 'Noise Dressed', [Laloux *et al.*, 1999], implying that care needs to be taken in applications. Various methods have been suggested to construct a pure correlation matrix with no measurement noise, such as the Single-Index Model and the Multi-Index Model, [Elton and Gruber, 2002]. Another approach, which has recently been explored by Physicists working in the area of Econophysics, involves the application of Random Matrix Theory.

Random matrix theory (RMT) was developed by Wigner, Dyson, Mehta and others during the 1960's, [Dyson, 1962; Dyson and Mehta, 1963; Edelman, 1988; Mehta, 2004], in an effort to understand the energy levels of complex atomic nuclei, which previous theories failed to explain. For these systems, RMT predictions represent an average over all possible interactions. Deviations from the universal predictions of RMT are of interest, as they describe specific non-random properties of the system studied. RMT was later found to

[2] An option gives the holder the right, but not the obligation, to buy, (for a call option) or sell, (for a put option), a specific amount of a given stock, commodity, currency, index, or debt, at a specified price, (the strike price), during a specified period of time.

[3] In finance, the Volatility Smile is the empirical observation that at-the-money options tend to have lower implied volatilities than in- or out-of-the-money options for the same underlying asset and expiry.

have applicability to a wide range of problems, from quantum chaos and gravity through mesoscopic and random systems. In the application of RMT, the spectral properties of an empirical correlation matrix are compared to those of a 'random' Wishart matrix[4], where eigenvalues that are outside the theoretical bounds of Wishart matrix are said to deviate from RMT. These deviating eigenvalues are said to contain information about the system under consideration, which can be uncovered through eigenvector analysis, [Gopikrishnan *et al.*, 2000].

One of the earliest applications of RMT to finance demonstrated that the optimisation of a margin account in futures markets, (where the constraint on the weights is nonlinear), is equivalent to finding the ground state configuration of a spin glass[5] in statistical physics, [Galluccio *et al.*, 1998]. This problem is known to be NP-complete, with an exponentially large number of solutions.

Two papers, published simultaneously, used RMT methods to analyse the properties of financial cross-correlation matrices (C) and showed that $\approx 94\%$ of the eigenvalues of C agree with the predictions of RMT, [Laloux *et al.*, 1999; Plerou *et al.*, 1999]. The remaining $\approx 6\%$ of eigenvalues were shown to deviate from RMT predictions. Later studies, [Gopikrishnan *et al.*, 2000; Plerou *et al.*, 2000a,b, 2002], compared statistical properties of the eigenvalue spectrum with the predictions of RMT in greater detail and found these to be in very good agreement. Further analysis, of the deviating eigenvalues, revealed clustering *corresponding to distinct business sectors*.

More recently RMT has been shown to improve risk management of a portfolio of equities. By incorporating the information contained in the deviating eigenvalues, [Laloux *et al.*, 2000], the difference between the predicted and realised risk of an optimal portfolio was reduced substantially. The correlations between stock fluctuations were compared to those of Random Magnets, [Rosenow *et al.*, 2002a,b]. In this case, the Random Magnets were shown to provide a framework to explain the origin of the correlations, together with the occurrence of power-law correlations in the time-series of highly correlated eigenmodes.

[4]The Wishart matrix W can be obtained from $W = \frac{1}{M} GG^T$, where G is a $N \times M$ random matrix with independent, zero mean and unit variance elements.

[5]A spin glass is a disordered material exhibiting high magnetic frustration. Frustration refers to the inability of the system to remain in a single lowest energy state (the ground state).

Additionally, RMT was used to improve the correlation estimate, compared to the standard form. The dynamics of the average cross-correlation and the index volatility were compared to those of the largest eigenvalue, [Rosenow *et al.*, 2003], for moving time-windows. It was demonstrated that the fluctuations in the average correlation, (and largest eigenvalue), are themselves correlated with the volatility of the index. The information contained, in the eigenvalues deviating from the RMT predictions, was then used to forecast future correlations between assets.

Given the level of noise apparent in empirical correlation matrices, highlighted previously, this leads to some concern over the validity of correlation matrices used by industry in the calculation of optimal portfolios and in risk management. Some of these concerns were alleviated, [Pafka and Kondor, 2002], where it was shown that, in the case of a linear portfolio optimisation, solutions are determined by the large stable eigenvalues, hence the disturbance due to noise is of the order of 5% to 15%. This was reinforced by the demonstration, [Pafka and Kondor, 2003], that noise depended on the ratio of time-series length, T, to portfolio size, N, ($Q = \frac{T}{N}$). For small values of $\frac{1}{Q}$ the levels of noise were found to be rather modest. A simulation based approach was subsequently used, [Pafka and Kondor, 2004], to demonstrate the usefulness of various correlation estimators such as the Single-Index Model and the RMT filtered matrix. Additionally, exponential weighting applied to financial returns, [Pafka *et al.*, 2004], demonstrated its superiority for calculating the RMT spectrum in comparison with uniformly weighted RMT based filtering.

These early studies on RMT applications in finance led to a proliferation of activity. The memory effect in financial time-series was investigated using lagged time-series, [Drozdz *et al.*, 2001a], and was shown to be shorter than previously found from autocorrelation analysis, due to increasing market efficiency. The study of stock market correlation dynamics, [Kwapien *et al.*, 2002], through RMT have shown that two periods of synchronous bursts of activity exist in the DAX index and further that consecutive returns carry essentially *no common information*. A more recent paper, [Kwapien *et al.*, 2005], proposed that even the bulk of the spectrum of the correlation matrix contains correlations, (possibly non-linear), that are masked by measurement noise due to short time-series. The evidence provided was

based on the multi-fractal character of the eigensignals, (even in the noisy RMT zone), and the fat-tailed distributions of returns.

These initial studies focused predominantly on the application of RMT to US stock markets. RMT was further shown to hold for the Japanese market, [Utsugi *et al.*, 2004], where the authors also studied the effect of randomness on deterministic correlations. They found that randomness causes a *repulsion* between deterministic and random eigenvalues, which explains the deviations from RMT sometimes found for small eigenvalues. The South African market was also shown, unexpectedly, to conform to RMT predictions in the bulk, [Wilcox and Gebbie, 2004]. For an emerging market such as South Africa, the stock liquidity is an issue, (since stocks trade at very different frequencies), and it was shown that greater conformity with RMT predictions can be achieved, if correlations are only calculated when data exist for all stocks, (Measured-data correlations).

Further, the cleaning technique, [Laloux *et al.*, 2000; Bouchaud and Potters, 2003], was shown, [Sharifi *et al.*, 2004], to decrease the stability of the cross-correlation matrix. Subsequently, an alternative technique, [Krzanowski, 1984], was shown to eliminate the noise but preserve the stability of the matrix, while reducing the overall risk of the portfolio. This work was extended, [Daly *et al.*, 2008], with the Krzanowski stability based filter applied to the correlation matrix, the covariance matrix and to an exponentially weighted covariance matrix between S&P 500 stocks. Results showed that RMT filtering, while reducing realised risk out-of-sample, on average, actually increased realised risk on a significant number of days.

A 'group model' for correlations in stock markets proposed [Noh, 2000], assumes that the returns of companies in the same sector are highly correlated. The spectral properties of the empirical correlation matrix were explained by this model, together with the behaviour of the Inverse Participation Ratio distribution, [Gopikrishnan *et al.*, 2000; Plerou *et al.*, 2000a]. The identification of group behaviour in stock markets was further examined, [Kim and Jeong, 2005], with market and random noise filtered from the correlation matrix using RMT. Stock groups were found to be identifiable (without additional knowledge of individual stocks), both by optimising the representation of the group correlation matrix and

by use of a percolation approach. An interesting treatise of some of the current results in RMT can be found, [Potters *et al.*, 2005], which includes discussions on the stability of the top eigenvalue and its associated eigenvector, together with implications of insights from analysis of exponentially weighted and frequency dependent correlation matrices.

2.5 Eigenvalue Analysis and Correlation Dynamics

The work, previously described, reviews techniques to improve the amount of genuine correlation information found in the equal-time cross-correlation matrix. The possible uses of a cross-correlation matrix are many, with applications in finance, [Laloux *et al.*, 1999; Plerou *et al.*, 1999; Gopikrishnan *et al.*, 2000; Plerou *et al.*, 2000a; Laloux *et al.*, 2000; Bouchaud and Potters, 2003; Utsugi *et al.*, 2004; Wilcox and Gebbie, 2004; Sharifi *et al.*, 2004; Podobnik and Stanley, 2008], electroencephalographic (EEG) recordings, [Schindler *et al.*, 2007a,b] and magnetoencephalographic (MEG) recordings, [Kwapien *et al.*, 2000], amongst others. Additional approaches to filter the large volumes of data in the correlation matrix have been explored, in particular ideas from network theory such as the *Minimum Spanning Tree* and the *Planar Maximally Filtered Graph*, [Onnela *et al.*, 2004; Tumminello *et al.*, 2007; Pozzi *et al.*, 2007]. In a financial context, alternative relationships such as those between stock price changes and liquidity or trading volume, [Ying, 1966; Karpoff, 1987; LeBaron *et al.*, 1999], have also been studied.

Several authors have recently suggested that there may, in fact, be some real correlation information hidden in the RMT defined 'random part' of the eigenvalue spectrum. A technique, involving the use of a power mapping[6] to identify and estimate the noise in financial correlation matrices, has been described, [Guhr and Kälber, 2003]. This allows the suppression of those eigenvalues, associated with the noise, in order to reveal different correlation structures buried underneath. Derivation of the relationship, between the eigenvalue density of the true correlation matrix, c, and that of the empirical correlation matrix, C, showed that correlations can be measured in the random part of the spectrum, [Burda *et al.*, 2004; Burda

[6]The Correlation matrix C is mapped to the matrix C^q, with q $+ve$ and $C^q_{kl} = sign(C_{kl})\,|C_{kl}|^q$ for element kl.

and Jurkiewicz, 2004] and was used, [Papp *et al.*, 2005], in the case of portfolio optimisation of financial portfolios. Here, the Authors reconstructed the true correlation matrix using the cleaned eigenvalues and the empirical eigenvectors for shrinkage models, both for the 'one factor' and a 'market plus sectors' models[7].

A Kolmogorov test was applied, [Malevergne and Sornette, 2004], to demonstrate that the bulk of the spectrum is not in the Wishart RMT class. In this paper, the authors demonstrate that the existence of factors such as an overall market effect, firm size and industry type is due to collective influence of the assets. More evidence that the RMT fit is not perfect was provided, [Kwapien *et al.*, 2006], where it was shown that the dispersion of "noise" eigenvalues is inflated, indicating that the bulk of the eigenvalue spectrum contains correlations masked by measurement noise.

The behaviour of the largest eigenvalue of a cross-correlation matrix for small windows of time, has been studied, [Drozdz *et al.*, 2000], for the DAX and Dow Jones Industrial Average Indices (DJIA). Evidence of a time-dependence between 'drawdowns' ('draw-ups') and an increase (decrease) in the largest eigenvalue was obtained, resulting in an increase of the *information entropy*[8]of the system. Similar techniques were used, [Drozdz *et al.*, 2001b], to investigate the dynamics between the stocks of two different markets (DAX and DJIA). In this case, two distinct eigenvalues of the cross-correlation matrix emerged, corresponding to each of the markets. By adjusting for time-zone delays, the two eigenvalues were then shown to coincide, implying that one market leads the dynamics of the other.

A new technique, applying the equal-time cross-correlation matrix, to characterise dynamical changes in nonstationary multivariate time-series was described, [Müller *et al.*, 2005]. It was shown that, as the synchronisation of k time-series within an $M-$dimensional multivariate time-series increases, this causes a repulsion between eigenstates of the correlation matrix, in which k levels participate. Through the use of artificially created time-series with pre-defined correlation dynamics, it was demonstrated that there exist situations, where

[7]The 'one factor' model assumes that the co-movement between stocks is due to a single common factor or index. The 'market plus sectors' model assumes that the co-movement between stocks is due to both market and industry/sector factors

[8]In information theory, the Shannon or Information Entropy is a measure of the uncertainty associated with a random variable.

the relative change of eigenvalues from the lower edge of the spectrum is greater than that of the large eigenvalues, implying that information drawn from the smaller eigenvalues is also highly relevant.

A first application of this technique, [Müller *et al.*, 2005], was to the dynamic analysis of the eigenvalue spectrum of the equal time cross-correlation matrix of multivariate epileptic seizure time-series, using sliding windows. The authors demonstrated that information about the correlation dynamics is visible in both the lower and upper eigenstates. A further detailed study of equal-time correlations between EEG signals, [Schindler *et al.*, 2007a], investigated temporal dynamics of focal onset epileptic seizures[9]. It was shown that the zero-lag correlations between multichannel EEG signals tend to decrease during the first half of a seizure and increase gradually before the seizure ends. This work was extended to the case of *Status epilepticus*, [Schindler *et al.*, 2007b], where the equal-time correlation matrix was used to assess neuronal synchronisation prior to seizure termination.

An alternative form of the correlation measure, proposed [Müller *et al.*, 2006b], was shown to be more sensitive for weak cross-correlations. For particular examples, information on cross-correlations was shown to be found in the RMT bulk of eigenvalues, with the information extracted at the lower edge statistically more significant than that extracted from the larger eigenvalues, [Müller *et al.*, 2006a]. The authors introduced a *method of unfolding the eigenvalue level density*[10], through the normalisation of each of the level distances by its ensemble average, and used this to calculate the corresponding individual nearest-neighbour distance. Through this unfolding, those parts of the spectrum, dominated by noise, could be distinguished from those containing information about correlations. Application of this technique to multichannel EEG data showed the smallest eigenvalues to be more sensitive to detection of subtle changes in the brain dynamics than the largest.

[9] A partial or focal onset seizure affects only a part of the brain at onset. They may often be a precursor to a larger seizure, such as a generalised seizure.

[10] In one dimension, unfolding is a local rescaling of the eigenvalue density, such that the density on the unfolded scale is equal to unity

2.6 EEG Correlation Dynamics

As discussed in the previous Section, the equal-time cross-correlation matrix has been applied to measure the linear synchronisation in many diffuse Complex Systems such as the financial systems described previously and electroencephalographic (EEG) recordings [Quian Quiroga *et al.*, 2002; Seba, 2003; Müller *et al.*, 2005; Ansari-Asl *et al.*, 2006; Müller *et al.*, 2006a; Schindler *et al.*, 2007a,b], as well as magnetoencephalographic (MEG) recordings [Kwapien *et al.*, 1998, 2000] and others. The dynamics of such systems are characterised by a continuously varying level of synchronisation between different subsets of the system. The degree of synchronisation is dependent on the length of time-series studied, the *granularity* or time-interval length of the data and the amount of noise in the system, amongst other things. Hence, in the case of EEG seizure data for example, the equal-time correlation matrix has been used to analyse the changes in the synchronisation structure, prior to seizure events, with the aim of *predicting epileptic seizures*.

The predictability of these seizures has been studied in great detail by many authors. A comparative analysis of 30 different measures, both univariate and bivariate, provided evidence for a preictal[11], state, [Mormann *et al.*, 2005]. Univariate analysis such as variance, skewness and kurtosis showed preictal changes $5 - 30$ minutes before seizures, while bivariate analysis showed changes up to 240 minutes prior to seizure events. Interestingly, linear techniques for seizure prediction were found to perform comparably or better than non-linear techniques. Application of the linear cross-correlation to a small number of EEG channels, [Wendling *et al.*, 2003], led to the conclusion EEG signals decorrelate at seizure onset at high frequencies, followed by an abnormal level of recoupling as the seizure develops.

The use of Wavelet techniques, (Section 2.7), in the analysis of EEG data is widespread, allowing a time-frequency decomposition of these non-stationary signals. The Wavelet transform, [Clark *et al.*, 1995; Senhadji and Wendling, 2001; Adeli *et al.*, 2003; Indirdevi *et al.*, 2008], has been used, for example, to determine localisation of transient signals

[11]The preictal state is the period prior to the start of an epileptic seizure, while the ictal state refers to the period during the seizure

(spikes) during ictal periods[11], vital in preoperative evaluation of the foci and propagation of ictal events. A technique, based upon the evolution of the accumulated energy at each wavelet scale, [Gigola *et al.*, 2004], was shown to predict epileptic seizure onset accurately in 12 out of 13 cases.

Wavelet energy analysis, [Ursino *et al.*, 2004], of changes in scalp EEG signals during epileptic seizures, showed significant changes in energy distribution at seizure inception. The redistribution of energy was, however, found to be inconsistent across patients and channels. Further, wavelet energies and entropies were used to characterise EEG signals from Secondary generalised tonic-clonic seizures[12], [Rosso *et al.*, 2006]. The authors demonstrated that the epileptic recruitment rhythm, [Gastaut and Broughton, 1973], is described by the relative wavelet energy. Furthermore, a wavelet-based similarity method across frequencies was described, [Ouyang *et al.*, 2007], using ideas from nonlinear dynamics to predict epileptic seizures.

Various authors have also attempted to measure the interdependencies between different regions of the brain. A number techniques were compared, [Quian Quiroga *et al.*, 2002; Ansari-Asl *et al.*, 2006], with relative performance shown to be dependent on the form of the underlying signals. In the case of Magnetoencephalographic (MEG) signals, [Mizuno-Matsumoto *et al.*, 2005], a limited study consisting of three patients was performed, using wavelets to determine cross-correlations over different frequencies and calculate the time lag between different brain regions. A *coherence function* and cross-correlation between different frequency bands, defined by a continuous filter bank, [Ansari-Asl *et al.*, 2005], enabled exploration of the time-frequency dependence of epileptic seizure data.

The EEG and MEG data, used in the above studies has many properties common to that studied in a financial context, such as non-stationarity, numerous interacting constituents, (traders versus neurons), and a continuously changing degree of synchronisation. However, real differences also exist. Recordings of Brain activity are continuous unlike equity prices, (as Markets are not continuously open). Moreover, financial data is available in non-synchronous 'ticks', while EEG data is recorded synchronously. The availability of large

[12]Formerly known as grand mal seizures, tonic-clonic seizures are a type of generalised seizure affecting the entire brain

quantities of data in both fields, allows a detailed comparison of the applicability of the techniques described above to varied data types, (as discussed later).

2.7 Multiscale Analysis

Scaling laws play a vital role across a range of Complex Systems and disciplines, with scaling effects found across Biological, Physical, Social and Economic systems. The importance of time-scale information in a financial context arises from the view that stock markets consist of heterogeneous investors operating at different intervals and looking towards different time horizons. Through examination of financial data at different granularities, more insight into the nature of volatility and correlation dynamics can be gained. In the context of EEG time-series, for example, interactions between various neurons naturally occur at different frequencies, depending on the neuronal functions involved. While time-scale dependence can be examined using a Fourier Analysis, [Bracewell, 1999], it is unsuitable for the study of *non-stationary* data such as financial time-series. In examining scale features, the Wavelet Transform, localised in both time and frequency, is more appropriate for non-stationary signals.

The expression "Wavelet" evolved from *ondelette* meaning "Small Wave". Applications using wavelets in disciplines other than finance are extensive, with many papers published in Astronomy, Medicine, Forensics, Engineering and Physics, [Aldroubi and Unser, 1996; Jaffard *et al.*, 2001; Mix and Olenjniczak, 2003]. While rooted in Fourier Theory, there are important differences. Fourier Analysis uses a combination of sine and cosine functions at different wavelengths to represent a given function. Such periodic functions are non-local, (ie. go to plus and minus infinity), and can not be used to deal with localised time-series, whereas wavelets are localised in both time and scale. Scaling in Fourier Analysis is typically expressed in terms of frequency, whilst in wavelet analysis, is typically referred to in terms of time. Moving along the signal, the spectrum of a scalable modulated window is calculated for every position. The process results in a collection of time-scale representations of the function, all with different resolutions. This collection of representations is

known as a *multiresolution analysis*. An introduction to the mathematical theory with some examples of applications of wavelets, can be found in [Daubechies, 1992; Kaiser, 1994; Bruce and Gao, 1996; Burrus *et al.*, 1998].

The wavelet transform has been applied to many distinct problems in finance and economics, [Gençay *et al.*, 2001b; Ramsey, 2002; Schleicher, 2002; Crowley, 2005], allowing a time frequency decomposition of the underlying function. This multiresolution analysis allows the decomposition of a time-series into several *layers of orthogonal sequences* corresponding to different frequencies. Each of these scales can then be analysed individually and, also, compared across different series, allowing comprehensive characterisation, analysis and comparison of Market dynamics and types.

The characterisation of stock markets into Emerging and Mature Markets has been studied by various authors. Using eigenvalue analysis, [Sharkasi *et al.*, 2006a], showed that Mature Markets respond in a different way to Emerging Markets during crashes. It was also shown that the second largest eigenvalue contains information about market dynamics, in addition to the largest. The Wavelet transform has also been applied to measure the recovery time of both emerging and mature markets, [Sharkasi *et al.*, 2006b], by studying the ratio of the largest eigenvalues for time-series reconstructed from selected wavelet components. This analysis, over different time scales, has confirmed that Mature Markets recover more quickly from crashes and that they exhibit antipersistent behaviour, while Emerging Markets display persistent behaviour[13].

By breaking a time-series down into constituent components, it is possible to remove high-frequency 'noise' and characterise scale-dependent properties. The wavelet multiscaling approach was used, [Gençay *et al.*, 2001c], to decompose a time-series by frequency scale, in order to remove seasonalities in Foreign Exchange data, (the dominant source of problems in various volatility models). Scaling laws in Foreign Exchange markets were also identified, [Gençay *et al.*, 2001a], using this approach, with exchange rate volatility shown to have different scaling properties at different time horizons and the correlation between

[13]The persistence of a time-series is a description of the bias in fractional Brownian motion and can be measured using the Hurst Exponent, with $H = 0.50$ for Brownian motion, $0.50 < H < 1.00$ for persistent, or trend-reinforcing series and $0 < H < 0.50$ for an anti-persistent, or mean-reverting system.

two volatility series also shown to be considerably stronger at longer time scales.

Using unfiltered[14] time-series to measure the cross-correlation can lead to a misunderstanding of the true correlation structure. In an investigation of correlation for equity markets at longer frequencies, application of wavelets showed that even when high correlation between markets was expected, (Bombay and National stock exchanges), in fact correlation varied markedly, depending on the scale considered, demonstrating that unfiltered data may mask the correlation structure, Razdan [2004].

The Beta[15] of an asset, [Gençay *et al.*, 2003, 2005], at different scales has been extensively studied. By calculating the wavelet variance and covariance, the Beta of an asset was estimated at each scale. For various markets, including the US, Germany and the United Kingdom, it was shown that the relationship between the return of a portfolio and its Beta was stronger for longer wavelet scales. This connection between asset returns and the returns on the market portfolio was further examined, [Norsworthy *et al.*, 2000], where it was shown that the markets' influence on asset returns was principally in high frequency movements. In an examination of the scale dependency of Japanese stocks, Yamada [2005], showed that the standard Beta estimate is an "average" of the wavelet based multiscale estimates.

Scaling phenomena have also been found to have application in both Macroeconomics and Economic contagion. A non-orthogonal variant of the discrete wavelet transform, the maximal overlap discrete wavelet transform (MODWT), was used, [Gallegati and Gallegati, 2007], to decompose the variance of the industrial production index of G-7 countries. Similar techniques, applied to emerging markets, [Gallegati, 2005], demonstrated their level of integration with developed markets. Wavelet variance and cross-correlation analysis between Middle East and North African (MENA) stock markets, (Egypt, Israel, Jordan, Morocco and Turkey), as well as S&P and Eurostoxx indices were studied and it was shown that *emerging markets are neither regionally nor internationally integrated*. Through wavelet decomposition of data for 7 stock markets, evidence was found for intra-continental rela-

[14]The original data, before a filtering technique, such as wavelets, is applied to decompose it into constituent frequencies.

[15]The Beta of an asset is a measure of the volatility, or systematic risk, of a security or a portfolio in comparison to the market as a whole, see Chapter 3.3.2

tionships with an increase in market contagion since the mid-1990s.

The measurement of variance and covariance over different time frames, [Gençay *et al.*, 2001b], is not restricted to the discrete wavelet transform. Two different techniques to measure variance were described, [Percival, 1995], using both the discrete wavelet transform and one of the first applications of the maximal overlap discrete wavelet transform (MODWT). Through Monte Carlo simulations it was shown that the MODWT estimator is reasonable even for small sample sizes of 128 observations. An early application of the technique involved observations of vertical shear in the ocean. A mathematical framework providing central limit theorems for MODWT estimators of the wavelet covariance and correlation was established, [Whitcher *et al.*, 1996].

An extensively studied characteristic of stock market behaviour, is the increase of stock return cross-correlations as the sampling time scale increases, a phenomenon known as the *Epps* effect, [Epps, 1979]. More recently, analysis of time-dependent correlations between high-frequency stocks, [Tóth and Kertész, 2006], demonstrated, however, that market reaction times have increased due to greater efficiency. A diminution of the Epps effect with time is one consequence of increased market efficiency. Trading asynchronicity was demonstrated to be not solely responsible for the effect, [Tóth and Kertész, 2007b], with the characteristic time apparently independent of the trading frequency. Further analysis using a toy model of Brownian motion and memoryless renewal process, [Tóth *et al.*, 2007], found an exact expression for the Epps frequency dependence, with reasonable fitting also for empirical data. In fact, the effect was shown, [Tóth and Kertész, 2007a], not to scale with market activity but to be due to reaction times, rather than market activity. A new description of the Epps effect was provided, based on decomposition of cross-correlations.

2.8 Hedge Funds

The core initial data, studied in this thesis, are low-frequency financial time-series, namely Hedge Fund returns. Given the huge growth in this area and the historical evidence that managers can help to preserve capital in adverse markets, Hedge Funds are of particular

interest given the current difficult market conditions. A brief overview of Hedge Funds, their strategies and some of the issues surrounding their data is provided in what follows. A Hedge Fund is a lightly-regulated private investment vehicle that may utilise a wide range of investment strategies and instruments. These funds may use short positions, derivatives, leverage and charge incentive-based fees. Normally, they are structured as limited partnerships or offshore investment companies. Hedge Funds pursue positive returns in all markets and hence are described as "absolute return" strategies.

Hedge Funds are utilised by pension funds, high net-worth individuals and institutions, due to their low correlation to traditional long-only investment strategies. The incentive-based performance fees, earned by Hedge Fund managers, align the interest of the Hedge Fund manager with that of the investor. The performance of Hedge Funds has been impressive, with the various Hedge Fund indices providing higher returns, with lower volatility, than traditional assets over many years. As of the end of the third quarter of 2008 the total assets managed by Hedge Funds world wide was estimated at $1.72 trillion, [Barclays, 2009]. Hedge Funds generally only report their returns on a monthly basis, however, and this means that very limited amounts of data are available for study, as databases of Hedge Fund returns have only been in operation for about 15 years. This is in keeping with the highly secretive, proprietary nature of Hedge Fund investing. The amount of information reported by a Hedge Fund about how and where it is producing its returns is often limited to sectoral overviews and strategy allocations. For an introduction to hedge funds see [Lhabitant, 2002, 2004]; for an overview of their strategies see Appendix A.

In addition to attractive returns, many Hedge Funds claim to provide significant diversification benefits when combined with traditional assets such as equities and bonds. Moderate correlations between Hedge Funds and traditional asset classes, using monthly returns data, have been reported, [Schneeweis and Martin, 2001; Liang, 2001]. Hence, the market exposure of Hedge Funds to traditional assets classes has also been found to be low, although these results were based upon monthly returns data and may be misleading. Hedge Funds may hold illiquid exchange traded assets or over the counter[16] (OTC) secu-

[16]An over the counter security is one that is not listed or traded on an organised exchange and instead is traded directly between parties

rities, which may be priced using the last traded price, (which may not have traded at or even near the end of the month), resulting in non-synchronous pricing. Furthermore, OTC securities may not have publicly available prices but instead rely on broker prices. Due to the large performance fees charged by hedge funds, using stale or non-synchronous data may be of great benefit to the manager, at the expense of the investor.

Additional diversification benefits may be gained by investing in a variety of Hedge Fund strategies, due to the presence of low and even negative correlations between different strategies. Such strategies can be broken up into two general categories: *directional* and *market neutral*. Directional strategies, (for example Long/Short Equity, Emerging markets, Macro and Managed Futures), have a high risk, high return profile and act as return enhancers to a traditional portfolio. Market Neutral strategies, (for example Convertible Arbitrage, Equity Market Neutral and Fixed Income Arbitrage), deploy a low risk profile and act as a substitute for some proportion of the fixed income holdings in an investor's portfolio, [Lhabitant, 2002, 2004]. However, as mentioned, the inter-strategy correlations, calculated using the original data, may be misleading.

A Fund of Hedge Funds is a strategy that invests in other funds, rather than investing directly in Stocks, Bonds or other securities. These allow investors to have access to a large and diverse portfolio of Hedge Funds without having to carry out due diligence on each individual manager. The diversification benefits provided by Fund of Funds are brought about by investing in a number of funds that have a low correlation to each other with a view to spreading risk across many different strategies. These correlations, however, are often calculated by using equally weighted fund returns and can contain a significant amount of noise, due to the very small amount of returns data available for hedge funds, [Lhabitant, 2004].

Risk control techniques for portfolios of hedge funds have been studied by various authors. The threshold point of *self-organised criticality* was used as a control parameter of risk exposure, [Nishiyama, 2001], for a portfolio of Hedge Funds. The correlation matrix between the returns of Hedge Funds was used, [Miceli, 2004], to characterise Hedge Fund strategies using Random Matrix Theory (RMT) and a minimum spanning tree. A downside-

risk[17] framework was used, [Perello, 2007], for the study of Hedge Fund risk, taking into account the inherent asymmetry in Hedge Fund return distributions. Survivorship bias in Hedge Funds was studied for the period $1990 - 1999$, where it was found that the level of survivorship bias was significantly positive, [Liang, 2001].

2.9 Summary

In this Chapter, we have given a broad introduction to Complex Systems, with particular emphasis on previous research in two sample systems of particular interest to us, that is Financial and Neurological. Common to these systems is a need to characterise and reduce risk associated with behaviour under 'stresses', such as seizures and market crashes. In the following, we show that the risk involved is associated with interactions between agents and look at ways to characterise, reduce and predict the level of risk.

[17] Downside-risk measures look at how much money an investor stands to lose during adverse market conditions

CHAPTER 3

METHODOLOGY

3.1 Introduction

In this chapter, we present some of the techniques used in the analysis of the equal-time cross-correlation matrix between time-series. The calculation of the cross-correlation for general time-series is introduced and the issue of noise reduction in the cross-correlation matrix, using Random Matrix Theory (RMT), is then addressed, with financial applications in Portfolio Optimisation and Sector Identification. In order to build up a methodology suitable for risk characterisation, we then discuss an approach that allows us to study the dynamics of correlations over time. In developing a framework for the group structures found in complex interacting systems, we propose an 'ab-initio' model for the correlation structure between Equities. Finally, we discuss the use of Wavelet multiscale analysis, to investigate the effect of granularity on the cross-correlation matrix.

3.2 The Cross-Correlation Matrix and Eigenvalues

Correlation is a statistical measure of the strength and direction of a linear relationship between two random variables. The most common form of correlation is the *Pearson* or *Product-Moment* correlation[1], which is obtained by dividing the covariance of two variables by the product of their standard deviations.

Given a series of measurements, $x(t)$, $t = 1, 2, \ldots, T$ of a random variable X, we first

[1]The concept was actually first introduced by Francis Galton, a half-cousin of Charles Darwin, Bulmer [2003]

normalise each $x\,(t)$ with respect to the series standard deviation σ as follows:

$$\tilde{x}\,(t) = \frac{x\,(t) - \widehat{x}}{\sigma} \tag{3.1}$$

Where σ and $\widehat{x}$ are the standard deviation and time average of $x\,(t)$ over all the measurements, $t = 1, 2, \ldots, T$.

Then, the equal-time cross-correlation matrix between measurements of two such random variables X and Y can be expressed in terms of $\tilde{x}\,(t)$ and $\tilde{y}\,(t)$,$t = 1, 2, \ldots, T$

$$C_{ij} \equiv \langle \tilde{x}\,(t)\,\tilde{y}\,(t) \rangle \tag{3.2}$$

The elements of the cross-correlation matrix, C_{ij}, are limited to the domain $-1 \leq C_{ij} \leq 1$, where $C_{ij} = 1$ defines perfect positive correlation, $C_{ij} = -1$ corresponds to perfect negative correlation and $C_{ij} = 0$ corresponds to no correlation. In matrix notation, the correlation matrix can be expressed as

$$\mathbf{C} = \frac{1}{T}\mathbf{Z}\mathbf{Z}^{\tau} \tag{3.3}$$

Where $\mathbf{Z}$ is an $N \times T$ matrix with elements z_{it}.

The N eigenvalues λ_i and eigenvectors $\hat{\mathbf{v}}_i$ of the correlation matrix $\mathbf{C}$ are found from the following

$$\mathbf{C}\hat{\mathbf{v}}_i = \lambda_i \hat{\mathbf{v}}_i, \qquad i = 1, \ldots, N. \tag{3.4}$$

The eigenvalues are then ordered according to size, such that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$.

3.3 Modern Portfolio Theory

Modern portfolio theory is concerned with models of security and portfolio analysis, in particular with how rational investors use diversification to optimise their portfolios and a risky asset should be priced. Given the importance of mean-variance portfolio theory and the Capital Asset Pricing Model (CAPM), we describe briefly their assumptions and form.

3.3.1 Portfolio Optimisation

The *diversification* of an investment into independently fluctuating assets reduces its risk. However, since cross-correlations between asset prices exist, accurate calculation of the cross-correlation matrix is vital. The return on a portfolio with N assets is given by

$$\Phi = \sum_{i=1}^{N} w_i G_i, \tag{3.5}$$

where $G_i(t)$ is the return on Fund i, w_i is the fraction of wealth invested in asset i and $\sum_{i=1}^{N} w_i = 1$. The risk of holding this portfolio is then given by

$$\Omega^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j C_{ij} \sigma_i \sigma_j \tag{3.6}$$

where σ_i is the volatility (Standard Deviation) of G_i, and C_{ij} are the elements of the cross-correlation matrix. In order to find an optimal portfolio, using the Markowitz theory of portfolio optimisation [Markowitz, 1958; Elton and Gruber, 2002; Bouchaud and Potters, 2003], we minimise Ω^2 under the constraint that the return on the portfolio, Φ, is some fixed value. This minimisation can be implemented by using two Lagrange multipliers, which leads to a set of N linear equations which can be solved for w_i. If we minimise Ω^2 for a number of different values of Φ, we obtain a region bounded by an upward-sloping curve, called the *efficient frontier*, which reflects the highest level of expected return possible for a given amount of risk.

In the special case of optimisation with a portfolio containing only Hedge Funds, the additional constraint of no *short-selling*[2] is natural due to the difficulties involved; (note that short-selling may be achievable by the use of swaps[3] but is uncommon), [Lhabitant, 2002, 2004].

Throughout this thesis, we focus on the standard deviation of returns as a measure of

[2]The selling of a security not currently owned by the seller, where the security is 'borrowed' from a third party. The seller benefits if the price of the security falls.

[3]A derivative security in which two counterparties agree to exchange one stream of cashflows against another stream.

financial portfolio risk. However, other methods of calculating the risk of a portfolio are available such as expected semi-variance (a measure of the variation in risk below certain threshold value), the shortfall risk measure (this measure the probability of an asset dipping below a certain target) and the maximum drawdown (the maximum loss that an investor could suffer within a specific time horizon), [Lhabitant, 2004]. These various methods capture different aspects of risk and may be more appropriate in particular cases. However, these techniques do not lend themselves to simple portfolio optimisation (in particular using the correlation matrix) and so, we choose the standard deviation of returns as a measure of risk.

3.3.2 The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) was developed, [Sharpe, 1964], building on earlier work, [Markowitz, 1958], as a model for pricing an individual asset or a portfolio. The derivation of the standard CAPM is based upon a number of assumptions, (summarised see [Elton and Gruber, 2002]), and including:

- Risk-averse investors who aim to maximize economic utility
- Infinite divisibility of financial assets
- Absence of personal income tax
- Unlimited short sales
- Unlimited borrowing and lending at the risk free rate
- Axiom that all assets are marketable
- All information is available to all investors at the same time

The CAPM states that the expected return of a security or a portfolio equals the rate on a risk-free security plus a risk premium. The latter depends linearly on the market risk exposure (i.e. the Beta of an asset, Equation 3.8) and the market risk premium. Hence, the

expected return on a risky asset is given by:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f], \qquad i = 1, \ldots, N \tag{3.7}$$

where $E(R_i)$ is the expected return on each of N risky assets, $E(R_m)$ is the expected return of the market, R_f denotes the risk free rate and β_i is the Beta of each asset with respect to the market portfolio M, (in excess of the risk free rate return), and is given by

$$\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)}, \qquad i = 1, \ldots, N. \tag{3.8}$$

A full derivation of the CAPM and the Beta of a portfolio, is given, [Elton and Gruber, 2002].

Equation 3.7 can be rewritten in terms of the risk premium by simply subtracting the risk-free rate from both sides of the equation giving,

$$E(R_i) - R_f = \beta_i[E(R_m) - R_f], \qquad i = 1, \ldots, N. \tag{3.9}$$

The empirical version of equation 3.9 is given by

$$R_i - R_f = \alpha_i + \beta_i[R_m - R_f] + \epsilon_i, \qquad i = 1, \ldots, N \tag{3.10}$$

where α_i is the expected firm-specific return and ϵ_i is a random error term. For a more detailed treatise of the CAPM see [Sharpe, 1964; Elton and Gruber, 2002; Bouchaud and Potters, 2003; Gençay *et al.*, 2003, 2005] .

3.4 Random Matrix Theory

Random Matrix Theory originates from work by Wigner in nuclear Physics in the 1950's and later developed by Dyson and Mehta, and many theoretical and empirical results are known [Dyson, 1962; Dyson and Mehta, 1963; Edelman, 1988; Mehta, 2004]. In the original application of the theory, the problem was to explain the observed structures of the

energy levels of heavy nuclei, when model calculations failed to explain the experimental data. Wigner made the bold assumption that the complexity of the interactions between constituents of the nucleus was such that these could, in fact, be modelled as random.

The spectral properties of a correlation matrix, $\mathbf{C}$, may be compared to those of a "random" Wishart correlation matrix or Laguerre ensemble, $\mathbf{R}$, [Plerou *et al.*, 2000a; Laloux *et al.*, 2000],

$$\mathbf{R} = \frac{1}{T}\mathbf{A}\mathbf{A}^{\tau} \tag{3.11}$$

where $\mathbf{A}$ is an $N \times T$ matrix with each random element having zero mean and unit variance. Statistical properties of random matrices have been known for many years in the Physics literature, [Mehta, 2004], and have been applied to financial problems relatively recently [Laloux *et al.*, 1999; Plerou *et al.*, 1999; Gopikrishnan *et al.*, 2000; Plerou *et al.*, 2000a; Laloux *et al.*, 2000; Rosenow *et al.*, 2002a; Bouchaud and Potters, 2003; Wilcox and Gebbie, 2004; Sharifi *et al.*, 2004; Burda *et al.*, 2004; Burda and Jurkiewicz, 2004; Conlon *et al.*, 2007].

In particular, the limiting property for the sample size $N \to \infty$ and sample length $T \to \infty$, providing that $Q = \frac{T}{N} \geq 1$ is fixed, has been examined to show analytically, [Sengupta and Mitra, 1999], that the distribution of eigenvalues λ of a random correlation matrix $\mathbf{R}$ is given by:

$$P_{rm}(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{3.12}$$

for λ within the region $\lambda_- \leq \lambda_i \leq \lambda_+$, where λ_- and λ_+ are given by

$$\lambda_\pm = \sigma^2\left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right), \tag{3.13}$$

Where σ^2 is the variance of the elements of $\mathbf{G}$; (for $\mathbf{G}$ normalised this is equal to unity).

$\lambda_\pm$ are the bounds of the theoretical eigenvalue distribution. Eigenvalues that are outside this region are said to deviate from Random Matrix Theory. Hence, comparing the empirical distribution, $P(\lambda)$, of the eigenvalues of the correlation matrix to the distribution

for a random matrix, $P_{rm}(\lambda)$, Eqn. 3.12, deviating eigenvalues can be identified. These deviating eigenvalues are said to contain genuine correlation information about the system under consideration and we can use eigenvector analysis to identify this in a given case.

3.4.1 Eigenvector Analysis

The deviations of the empirical eigenvalue distribution, $P(\lambda)$, from the theoretical RMT result, $P_{rm}(\lambda)$, implies that these deviations should also be displayed in the statistics of the corresponding eigenvector components, [Laloux *et al.*, 2000]. In order to interpret the meaning of the deviating eigenvectors, we note that the largest of N eigenvalues is of an order of magnitude larger than the others, which constrains the remaining $N-1$, since the trace of the correlation matrix, $Tr\left[C\right]=N$. Thus, to analyse the contents of the remaining eigenvectors, we first remove the effect of the largest eigenvalue. To do this we use linear regression, [Plerou *et al.*, 2002]

$$G_i\left(t\right)=\alpha_i+\beta_i G^{large}\left(t\right)+\epsilon_i\left(t\right), \quad i=1,\ldots,N \tag{3.14}$$

Where $G^{large}=\sum_1^N u_i^{large}G_i\left(t\right)$ and N is the number of different times-series in our sample. Here u_i^{large} corresponds to the components of the eigenvector associated with the largest eigenvalue. We then calculate the correlation matrix using the residuals, $\epsilon_i\left(t\right), i=1\ldots N$. This matrix then takes the place of matrix C in porfolio optimisation. If we quantify the variance of the part not explained by the largest eigenvalue as $\sigma^2=1-\lambda_{large}/N$, [Laloux *et al.*, 2000], we can use this value to recalculate our values of $\lambda_{\pm}$.

Using techniques for group/sector identification, [Gopikrishnan *et al.*, 2000], we attempt to analyse the information contained in the eigenvectors. We first partition the time-series into predefined groups $l=1,\ldots M$, (for example, in the case of Hedge Funds we use strategies such as Equity Long/Short, Managed Futures, Convertible Arbitrage. See Appendix A for a complete strategy breakdown and description). We define a projection matrix $P_{li}=\frac{1}{n_l}$, if series i belongs to group l and $P_{li}=0$ otherwise. For each deviating eigenvector u^k, $k=(N-d),\ldots,N$ and d the number of deviating eigenvalues, we then

compute the contribution,

$$X_l^k \equiv \sum_{i=1}^{N} P_{li}^k \left[u_i^k\right]^2, \quad k = (N-d), \ldots, N \tag{3.15}$$

of each grouping, where this represents the product of the projection matrix and the square of the eigenvector components. This allows us to measure the contribution of the different groupings or sectors to each of the eigenvalues.

3.4.2 Inverse Participation Ratio

As suggested, [Plerou *et al.*, 2002], we also aim to assess how random properties diminish as we move further from the RMT upper boundary, λ_+. To do this, we use the Inverse Participation Ratio (IPR). The IPR allows quantification of the number of components that participate significantly in each eigenvector and reveals more about the level and nature of deviation from RMT. The IPR of the eigenvector u^k is given by

$$I^k \equiv \sum_{l=1}^{N} \left(u_l^k\right)^4 \tag{3.16}$$

and allows us to compute the inverse of the number of eigenvector components that contribute significantly to each eigenvector. There are two limiting cases for the IPR: (i) When the eigenvector has identical components $u_l^k = \frac{1}{\sqrt{N}}$, then IPR $= \frac{1}{N}$, (ii) If one component $u_1^k = 1$ and all others zero, then IPR $= 1$. The IPR quantifies the *reciprocal of the number of eigenvector components* that contribute significantly to each eigenvalue.

3.4.3 RMT and Portfolio Optimisation

One particular application of Random Matrix Theory is in the reduction of noise in the correlation matrix calculated for returns of financial assets, ([Bouchaud and Potters, 2003; Sharifi *et al.*, 2004] and references therein), a vital input in financial portfolio optimisation. In order to examine randomness effects on the cross-correlation matrix between financial

assets, we first divide the time-series of returns into two equal parts. We assume that we have *perfect knowledge* of the future average returns m_i by taking the observed returns on the second sub-period. We calculate

1. the predicted efficient frontier using the correlation matrix for the first sub-period and the expected returns m_i

2. the realised efficient frontier using the correlation matrix for the second sub-period and the expected returns m_i

The portfolio risk due to the noise in the correlation matrix can then be calculated using

$$R_p = \frac{\Omega_r^2 - \Omega_p^2}{\Omega_p^2} \tag{3.17}$$

Where Ω_r^2 is the risk of the realised portfolio and Ω_p^2 is the risk of the predicted portfolio.

It was shown, [Burda *et al.*, 2004; Burda and Jurkiewicz, 2004], that correlations may also be measured in the random part of the eigenvalue spectrum. However, since our aim here is to demonstrate how Random Matrix Theory can be used to improve the risk/return profile for a portfolio of Hedge Funds, we assume in our early work that the eigenvalues corresponding to the noise band in RMT, $\lambda_- \leq \lambda \leq \lambda_+$, are not expected to correspond to real information, (following [Rosenow *et al.*, 2002a; Sharifi *et al.*, 2004] and references therein). We then use this assumption to remove some of the noise from the correlation matrix. Although the technique used in [Laloux *et al.*, 2000; Bouchaud and Potters, 2003] has been shown to lead to problems with the *stability* of the correlation matrix, [Sharifi *et al.*, 2004], we apply it here as a simple test case to demonstrate how noise can be removed from the cross-correlation matrix formed from Hedge Fund returns. This technique involves separating the noisy and non-noisy eigenvalues and keeping the non-noisy eigenvalues the same. The noisy eigenvalues are then replaced by their average and the correlation matrix is reconstructed using

$$C_{clean} = \hat{\mathbf{v}}^{\tau} \lambda_{\mathbf{clean}} \hat{\mathbf{v}} \tag{3.18}$$

where λ_{clean} are the cleaned eigenvalues and $\hat{\mathbf{v}}$ are the eigenvectors of the original correlation matrix. After this, we can calculate the efficient frontiers and compare the risk of both the predicted and realised portfolios for the original and cleaned correlation matrices.

Conventionally, RMT has been applied, in a financial context to high and medium-frequency data, such as equity returns, [Bouchaud and Potters, 2003]. There is little reported in the literature on attempts to subject sparse data, such as Hedge Fund returns data, to such rigors, a common perception being that such data are unlikely to yield a great deal of information. In fact, due to the level of noise inherent in sparse data sets, the application of filtering techniques such as RMT is of paramount importance.

3.5 Correlation Dynamics

The correlations between non-stationary multivariate time-series are characterised by an ever-changing degree of synchronisation within the system. By considering the eigenspectrum of the cross-correlation matrix, we are able to reduce the complexity of the various synchronisations within a system. Then, through the study of the eigenvalue dynamics, we can consider the temporal evolution of the system correlations.

The sum of the diagonal elements of a matrix, (the Trace), must always remain constant under linear transformation, [Jolliffe, 2002]. Thus, the sum of the eigenvalues must always equal the Trace of the original correlation matrix. Hence, if some eigenvalues increase then others must decrease, to compensate, and vice versa (*Eigenvalue Repulsion*).

There are two limiting cases for the distribution of the eigenvalues, [Müller *et al.*, 2005; Schindler *et al.*, 2007a]. When all the time-series are perfectly correlated, $C_i \approx 1$, the largest eigenvalue is maximised with a value equal to N, while for time-series consisting of random numbers with average correlation $C_i \approx 0$, the corresponding eigenvalues are distributed around 1, (where any deviation is due to spurious random correlations, Figure 3.1).

For cases between these two extremes, the eigenvalues at the lower end of the spectrum can be much smaller than λ_{max}. To study the dynamics of each of the eigenvalues, using a

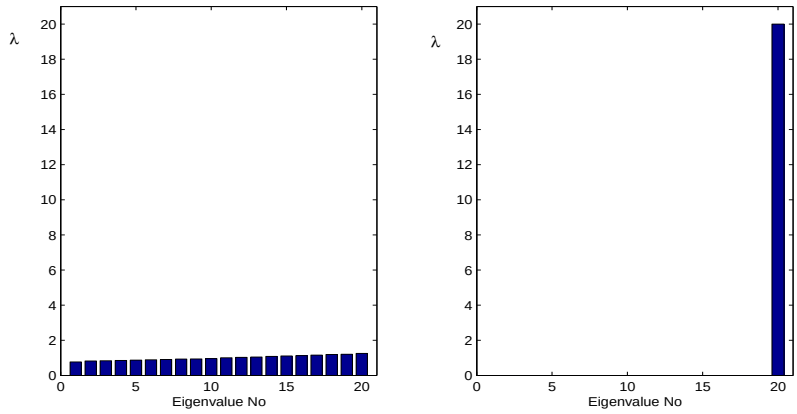


Fig. 3.1: Limiting Cases for eigenvalue distributions

sliding window, we normalise each eigenvalue in time using

$$\tilde{\lambda}_i(t) = \frac{\left(\lambda_i(t) - \widehat{\lambda_i(\tau)}\right)}{\sigma^{\lambda_i(\tau)}} \tag{3.19}$$

where $\widehat{\lambda}_i(\tau)$ and $\sigma^{\lambda_i(\tau)}$ are the mean and standard deviation of the eigenvalue i over a particular reference period, τ. This normalisation allows us to visually compare eigenvalues at both ends of the spectrum, even if their magnitudes are significantly different. The reference period, used to calculate the mean and standard deviation of the eigenvalue spectrum, can be chosen to be either a low volatility sub-period, (which helps to enhance the visibility of high volatility periods), or the full time-period studied.

3.5.1 One-factor Model

The behaviour of the largest eigenvalue of the correlation matrix can be accurately described by a simple one-factor correlation model, Section 2.5. In the *one-factor model* of stock returns, we assume a *global correlation* with the cross-correlation between all stocks the same, ρ_0, [Elton and Gruber, 2002]. The spectrum of the associated correlation matrix consists of only two values, a large eigenvalue of order $(N-1)\rho_0 + 1$, associated with the market, and an $(N-1)-$fold degenerate eigenvalue of size $(1-\rho_0) < 1$. Any deviation

from these values is due to the finite length of time-series used to calculate the correlations. In the limit $N \to \infty$, (even for *small correlation*, i.e. $\rho \to 0$), a large eigenvalue appears, which is associated with the eigenvector $v_1 = \left(\frac{1}{\sqrt{N}}\right)(1, 1, 1, \ldots, 1)$, and which dominates the correlation structure of the system.

3.5.2 'Market plus sectors' model

To expand the above to a 'market plus sectors' model, Section 2.5, we perturb a number of pairs N of the correlations $\rho_0 + \rho_n$, where $-1 - \rho_0 \leq \rho_n \leq 1 - \rho_0$. Additionally, we impose a constraint $\sum_N \rho_n = 0$, ensuring that the average correlation of the system remains equal to ρ_0. These perturbations allow us to introduce *groups of stocks* with similar correlations, (corresponding to Market Sectors).

Using the correlation matrix from the "one-factor model" and the "market plus sectors model", we can construct correlated time-series using the Cholesky decomposition $\mathbf{A}$ of a correlation matrix $\mathbf{C} = \mathbf{A}\mathbf{A}^{\tau}$, [Press *et al.*, 2007]. We can then generate finite correlated time-series of length T,

$$x_{it} = \sum_j A_{ij} y_{jt} \qquad t = 1, \ldots, T \tag{3.20}$$

where y_{jt} is a random Gaussian variable with mean zero and variance unity at time t. Using Eqn. 3.2, we can then construct a correlation matrix using the simulated time-series. The finite size of the time-series introduces 'noise' into the system and hence empirical correlations will vary from sample to sample. This 'noise' can be reduced through the use of longer simulated time-series or through averaging over a large number of series.

In order to compare the eigenvectors from each of the model correlation matrices to those constructed from the equity returns time-series, we can use the Inverse Participation Ratio (IPR), (Section 3.4.2).

Consideration of the above techniques in isolation will unmask changes in correlation dynamics for the system in question, an idea not previously examined in the literature for financial time-series. Additionally, by first decomposing the signal into component frequen-

cies, further dynamical features may emerge, allowing a multivariate time and scale analysis of correlations for the first time. In order to first break the signal up, wavelet multiscaling techniques are applied.

3.6 Wavelet Multiscaling

The characteristics of a time-series may be decomposed into different frequency components, a process known as multiscaling. Wavelets provide an efficient means of studying these multiresolution properties, as they can be used to decompose a signal into different time horizons or frequency components. The Discrete Wavelet Transform (DWT), [Bruce and Gao, 1996; Burrus *et al.*, 1998; Percival and Walden, 2000], in particular, allows the decomposition of a signal into components of different frequency. There are two basic wavelet functions, the father wavelet ϕ and the mother wavelet ψ. The formal definitions of the father and mother wavelets are the functions:

$$\phi_{j,k}(t) = 2^{\frac{j}{2}}\phi\left(2^j t - k\right) \tag{3.21}$$

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi\left(2^j t - k\right) \tag{3.22}$$

where $j = 1, , \ldots, J$ in a J-level decomposition and k ranges from 1 to the number of coefficients in the specified level. The father wavelet integrates to unity and reconstructs the longest time-scale component of the series, whereas the mother wavelet integrates to zero and is used to describe the deviations from the trend. The wavelet representation of a discrete signal $f(t)$ in $L^2(R)$ is given by:

$$\begin{aligned} f(t) &= \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\phi_{J,k}(t) + , \ldots, \\ &+ \sum_k d_{1,k}\phi_{1,k}(t) \end{aligned} \tag{3.23}$$

where J is the number of multiresolution levels (or scales) and k ranges from unity to the number of coefficients in the specified level. The coefficients $s_{J,k}$ and $d_{J,k}$ are the smooth

and detail component coefficients respectively and are given by

$$s_{J,k} = \int \phi_{J,k} f(t) dt \tag{3.24}$$

$$d_{j,k} = \int \psi_{j,k} f(t) dt \qquad (j = 1, , \ldots, J) \tag{3.25}$$

Each of the coefficient sets $S_J, d_J, d_{J-1}, , \ldots, d_1$ is called a *crystal*.

The Maximum Overlap Discrete Wavelet Transform (MODWT), [Percival and Walden, 2000; Gençay *et al.*, 2001b], is a linear filter that transforms a series into coefficients related to variations over a set of scales. Like the Discrete Wavelet Transform (DWT), it produces a set of time-dependent detail and scaling coefficients with basis vectors associated with a location t and a unitless scale $\tau_j = 2^{j-1}$ for each decomposition level, $j = 1, , \ldots, , J_0$. The MODWT, unlike the DWT, has a high level of redundancy, however, and is nonorthogonal. It retains downsampled[4] values at each level of the decomposition that would be discarded by the DWT. The MODWT can also handle any sample size N, whereas the DWT restricts the sample size to a multiple of 2^j. One of the benefits of the redundancy present in the MODWT is the ability to line up features in a multiresolution analysis with the original time-series, something not possible using the DWT. In this study, we apply the MODWT as it helps reduce the errors associated with the calculation of the Wavelet Correlation at different scales, (due to the availability of greater amounts of data at longer scales).

Decomposing a signal using the MODWT to J levels theoretically involves the application of J pairs of filters, [Percival and Walden, 2000]. The filtering operation at the j^{th} level consists of applying a re-scaled father wavelet to yield a set of *detail coefficients*

$$\tilde{D}_{j,t} = \sum_{l=0}^{L_j - 1} \tilde{\psi}_{j,l} f_{t-l} \tag{3.26}$$

and a re-scaled mother wavelet to yield a set of *scaling coefficients*

$$\tilde{S}_{j,t} = \sum_{l=0}^{L_j - 1} \tilde{\phi}_{j,l} f_{t-l} \tag{3.27}$$

[4]Downsampling or decimation of the wavelet coefficients retains half of the number of coefficients that were retained at the previous scale. Downsampling is applied in the Discrete Wavelet Transform

for all times $t =, \ldots, -1, 0, 1, \ldots$, where f is the function to be decomposed [Percival and Walden, 2000]. The rescaled mother, $\tilde{\psi}_{j,l} = \frac{\psi_{j,l}}{2^j}$, and father, $\tilde{\phi}_{j,t} = \frac{\phi_{j,l}}{2^j}$, wavelets for the j^{th} level are a set of scale-dependent localised differencing and averaging operators and can be regarded as rescaled versions of the originals. The j^{th} level equivalent filter coefficients have a width $L_j = (2^j - 1)(L - 1) + 1$, where L is the width of the $j = 1$ base filter. In practice the filters for $j > 1$ are not explicitly constructed because the detail and scaling coefficients can be indirectly calculated, (using an algorithm that involves the $j = 1$ filters operating recurrently on the j^{th} level scaling coefficients, to generate the $j + 1$ level scaling and detail coefficients), [Percival and Walden, 2000].

The MODWT is also an energy conserving decomposition, [Percival and Walden, 2000]:

$$\|\mathbf{f}\|^2 = \sum_{j=1}^{J} \left\|\tilde{\mathbf{D}}_j\right\|^2 + \left\|\tilde{\mathbf{S}}_J\right\|^2 \tag{3.28}$$

This decomposition allows the measurement of the contribution to the total energy due to changes at scale 2^{j-1}. A fractional energy, $\tilde{E}_j$, associated with each scale can be calculated using

$$\tilde{E}_j = \frac{E_j}{E_{tot}},\ with\ E_{tot} = \sum_{j=1}^{J} E_j. \tag{3.29}$$

for scales $j = 1, , \ldots, , J$.

3.6.1 Wavelet Variance

The wavelet variance $\nu_f^2(\tau_j)$ is defined as the expected value of $\tilde{D}_{j,t}^2$, if we consider only the non-boundary coefficients. An *unbiased* estimator of the wavelet variance is formed by removing all the coefficients that are affected by the boundary conditions[5] and is given by:

$$\nu_f^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^2 \tag{3.30}$$

[5]The MODWT treats the time-series as if it were periodic using "circular boundary conditions", so preventing double-counting in the calculation. There are L_j wavelet and scaling coefficients that are influenced by the extension, which are referred to as the boundary coefficients.

where $M_j = N - L_j + 1$ is the number of non-boundary coefficients at the j^{th} level, [Percival and Walden, 2000]. The wavelet variance decomposes the variance of a process on a scale by scale basis and allows us to explore how a signal behaves over different time horizons.

3.6.2 Wavelet Covariance and Correlation

The wavelet covariance between functions $f(t)$ and $g(t)$ is similarly defined to be the covariance of the wavelet coefficients at a given scale. The *unbiased* estimator of the wavelet covariance at the j^{th} scale is given by

$$\nu_{fg}(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^{f(t)} \tilde{D}_{j,t}^{g(t)} \tag{3.31}$$

where all the wavelet coefficients affected by the boundary are removed, [Percival and Walden, 2000], and $M_j = N - L_j + 1$.

The MODWT estimate of the wavelet cross-correlation, between functions $f(t)$ and $g(t)$, may then be calculated using the wavelet covariance and the square root of the wavelet variance of the functions at each scale j. The MODWT estimator, [Gençay *et al.*, 2001b], of the wavelet correlation is then given by:

$$\rho_{fg}(\tau_j) = \frac{\nu_{fg}(\tau_j)}{\nu_f(\tau_j)\nu_g(\tau_j)} \tag{3.32}$$

In the CAPM model, Section 3.3.2, the wavelet Beta estimator, at scale j, [Gençay *et al.*, 2001b], is defined as

$$\beta_{fm}(\tau_j) = \frac{\nu_{fm}(\tau_j)}{\nu_m^2(\tau_j)} \tag{3.33}$$

where $\nu_{fm}(\tau_j)$ is the covariance between an asset f and the market m, and $\nu_m(\tau_j)$ is the variance of the return on the market portfolio at scale j.

3.7 Summary

In this chapter, we outlined the use of the cross-correlation matrix as a measure of synchronisation between time-series. We then introduced a number of techniques designed to increase the amount of information gleaned from, as well as detect changes in the structure of the correlation matrix. While these methods are known, the techniques have not been previously integrated or used to explore the correlation structure in data sets from diverse applications such as Hedge Fund returns and Brain signals. A combination of techniques allows us to demonstrate the use of the correlation matrix for diversification purposes, even in sparse data-sets. It also provides a test of robustness of the methodologies considered in, for example, the characterisation of extreme events in Complex Systems through the analysis of the eigenspectrum.

CHAPTER 4

RANDOM MATRIX THEORY AND FUNDS OF FUNDS PORTFOLIO OPTIMISATION

As described, (Section 2.4), Random Matrix Theory (RMT) has been applied to the correlation matrix between returns of stocks from markets of various countries. The data examined in these studies varied from medium frequency *interday* returns [Laloux *et al.*, 1999; Plerou *et al.*, 1999; Laloux *et al.*, 2000; Plerou *et al.*, 2002; Utsugi *et al.*, 2004; Sharifi *et al.*, 2004; Conlon *et al.*, 2007], to high frequency *intraday*, [Plerou *et al.*, 2000a, 2002]. Here, we extend the application of Random Matrix Theory in financial markets through the study of sparse, low-frequency Hedge Fund returns. The purpose is to demonstrate the application of the correlation matrix to portfolio diversification, even in a low-frequency data environment. Given the current difficulties facing financial markets, perhaps the Hedge Fund focus is particularly relevant due to their stated aim of *absolute returns*.

The proprietary nature of Hedge Fund investing means that it is common practise for managers to release minimal information about their returns, often reporting only once per month. The construction of a Fund of Hedge Funds portfolio, (Section 2.8), thus requires correlation matrix estimation, typically based on a relatively small sample of monthly returns data and this means that high levels of noise can have significant impact on portfolio optimisation.

In this Chapter, we establish that information is retained within the correlation matrix for low-frequency Hedge Fund returns. Further, through the application of Random Matrix Theory, (Section 3.4), we test the ability of RMT-based filtering to clean a correlation

matrix constructed from low-frequency data. We investigate, in a financial context what diversification benefits can be found using the information contained in eigenvalues of the cross-correlation matrix deviating from Random Matrix Theory. Using classical portfolio optimisation theory, (Chapter 3.3.1), we consider how the cleaned correlation matrix can be used in Hedge Fund risk management, specifically to improve the risk-return profile of a portfolio of Hedge Funds.

4.1 Data and Implementation

The dataset studied here is a collection of 49 Hedge Funds with varying strategies, over a synchronous period from January 1997 to September 2005, $(T = 105)$. The original dataset was much larger, (approximately 1500 funds), but since the length of data series available was much less than the number of funds we were forced to choose a subset. The subset chosen contained the 49 funds with the longest track records giving us a fund to data ratio $Q = 2.143$. Reducing the dataset in this way is not unrealistic as a typical fund of hedge funds would monitor a subset of funds and choose a portfolio from these, [Lhabitant, 2002, 2004]. Often, one of the criteria used in choosing this subset of investable funds would be the completion of a minimum track record. (Other data sets, such as a portfolio made up of Hedge Fund strategy indices, were also studied with similar although less obvious results. The smaller number of strategies available means the amount of correlation information available decreases and there are fewer deviating eigenvalues than will be shown below for Fund returns).

In order to explore the information content of a correlation matrix, we calculated the *empirical Correlation Matrix* between the funds using equally weighted returns and from this found the spectrum of eigenvalues, (Section 3.2). This was then compared with the theoretical spectrum for random Wishart matrices, (as per Eqn. 3.12). To demonstrate the stability of the returns across different time periods and for different series lengths, we divided the data into two segments and reperformed the above experiment. The information found in the deviating eigenvalues was subsequently analysed using eigenvector analysis

and the inverse participation ratio. Finally, the RMT filtered correlation matrix was used as an input to a portfolio optimisation for a collection of Hedge Funds, (Section3.4.3).

4.2 Results

4.2.1 Equally weighted Correlation Matrix

The minimum and maximum theoretical eigenvalues of the correlation matrix corresponding to the full data set were found to be $\lambda_- = 0.1$ and $\lambda_+ = 2.83$, respectively, (Section3.4). As can be seen, from Fig. 4.1, the bulk of the eigenvalues conformed to those of the random matrix. There were three deviating eigenvalues, at 10.9886, 8.2898 and 2.944, corresponding to spikes outside the bulk envelope. This means that 6.1% of eigenvalues deviated from the RMT prediction, consistent with the findings of [Laloux *et al.*, 2000] for equity markets, where the authors argue that *at most* 6% of eigenvalues are non-noisy.

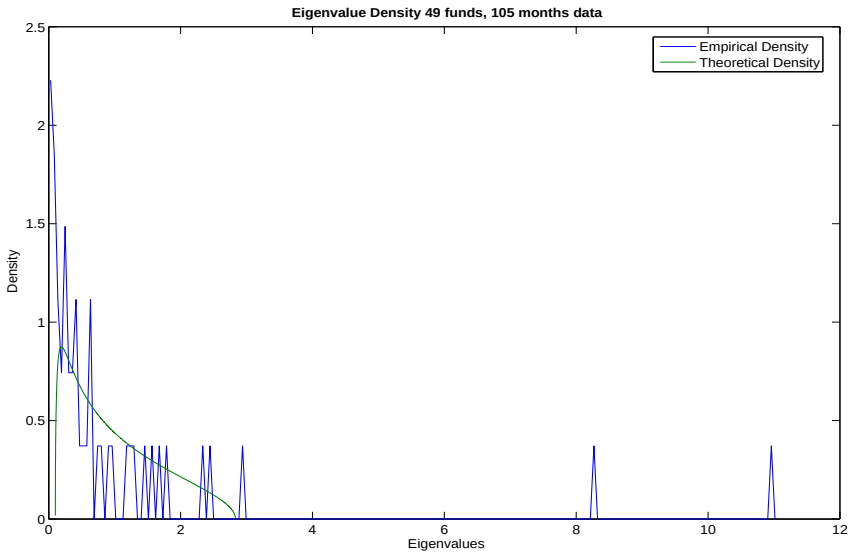


Fig. 4.1: Spectral density for equally weighted Hedge Fund correlation matrix

4.2.2 Bootstrapping

Dividing the time-series into two segments, we found just two eigenvalues that deviated from the RMT prediction, for both time periods. As can be seen in Fig. 4.2, the anomalous eigenvalue contributions are very similar for both periods chosen, which implies these are independent of the choice of time period, (ie. stationarity of the data). The values of the deviating eigenvalues are shown in Table 4.1.

Eigenvalue Rank	*105 Months Returns*	1^{st} *53 Months*	2^{nd} *52 Months*
1	10.9886	11.334	11.6874
2	8.2898	8.1375	8.9312
3	2.944		

Tab. 4.1: Eigenvalues deviating from RMT predictions

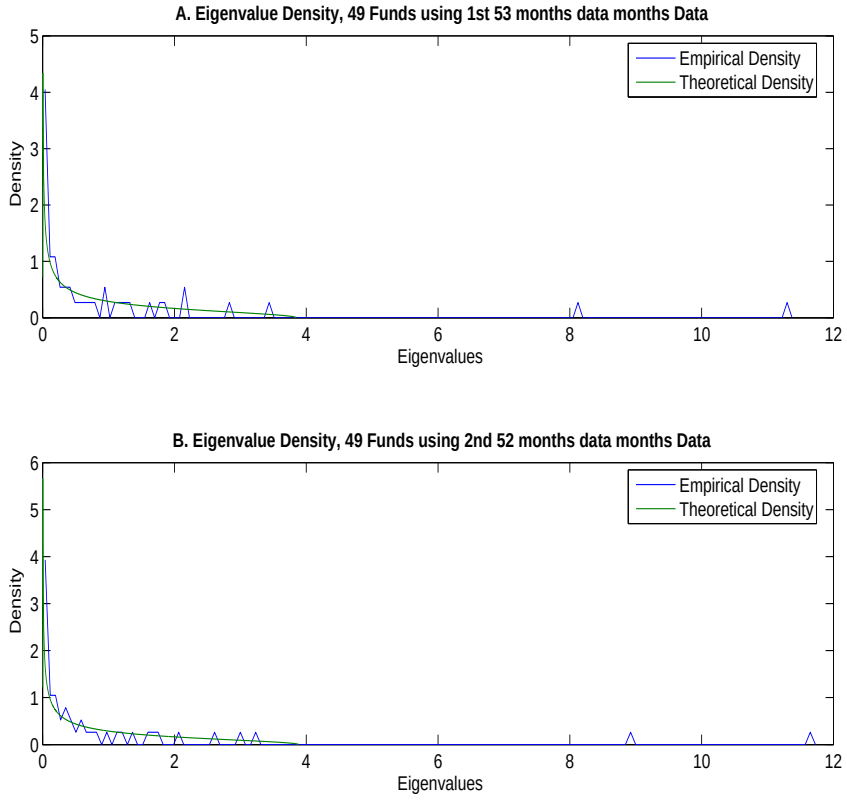


Fig. 4.2: Bootstrapped spectral density for consecutive periods.

4.2.3 Eigenvector Analysis

Fig. 4.3 shows the density of the components of the largest eigenvector and also the components of a typical eigenvector from the region predicted by RMT. As can be seen from this graph, the distribution of the components of the largest eigenvector is significantly different from that of an eigenvector chosen from the random region. The average value is greater and the variance of the components much smaller for the largest eigenvector, which is in agreement with [Plerou *et al.*, 2002; Sharifi *et al.*, 2004], where the largest eigenvector was interpreted as the 'market'. A Kolmogorov-Smirnov test rejected the hypothesis that the two eigevectors came from the same distribution, with $P < e^{-8}$. In this case the 'market' is the set of external stimuli that affect most Hedge Funds, (eg. Interest rate changes, large market (ie S&P 500 etc) moves, margin changes etc).

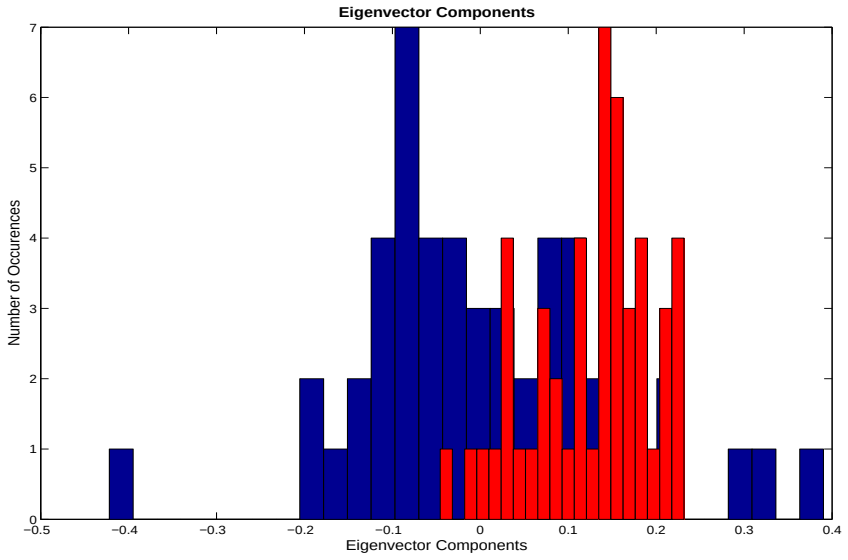


Fig. 4.3: Comparison of Eigenvector Components, showing the largest Eigenvector (Red), Eigenvector from the bulk (Blue)

We then removed the effects of the largest eigenvalue, (using techniques described in Section 3.4.1). This caused a shift in λ_{max} from 2.8329 to 2.1975 which means that 4 of the remaining largest eigenvalues are now outside the RMT region, (Fig. 4.4).

The distribution of the components of largest remaining deviating eigenvector showed

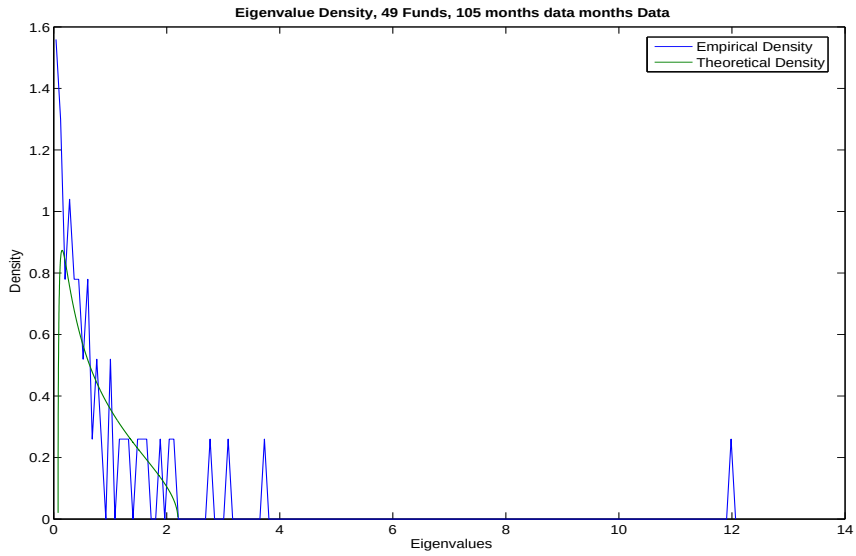


Fig. 4.4: Eigenvalue spectrum after the removal of the effects of the largest Eigenvalue

distinctive clustering, (Fig. 4.5). In particular Managed Futures, Emerging Markets and European Long/Short Equity strategies were major contributors here.

A similarly-clustered distribution also featured for other deviating eigenvalues. An analysis of the eigenvector components for the 2nd largest remaining eigenvector, after removing the effects of the market eigenvector, (Fig. 4.6), showed distinctive sectoral grouping, especially for managed futures and long/short equity. However, the components corresponding to the managed futures strategy did not deviate much from zero and contribute little. These findings for clustering of the deviating eigenvalues are in agreement with those of Sharkasi *et al.* [2006b], where the authors showed that, in addition to the largest, other deviating eigenvalues of the correlation matrix of asset returns also contain information about the risk associated with these assets.

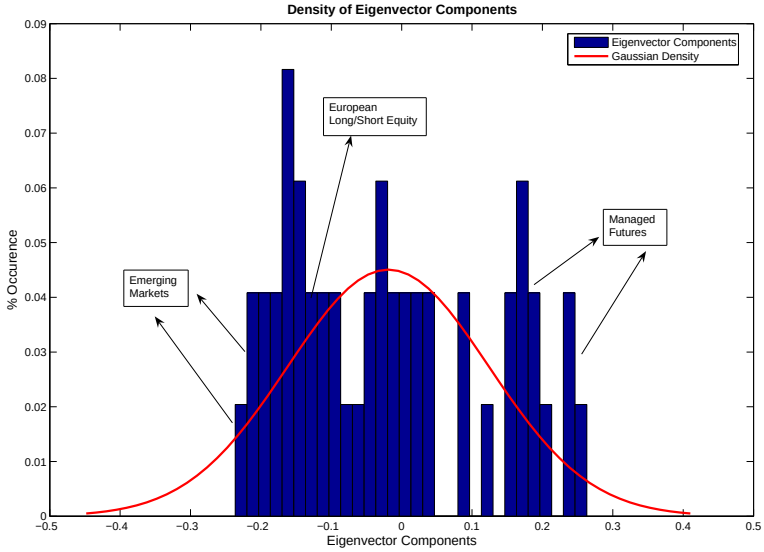


Fig. 4.5: Density of eigenvector components, largest remaining eigenvalue

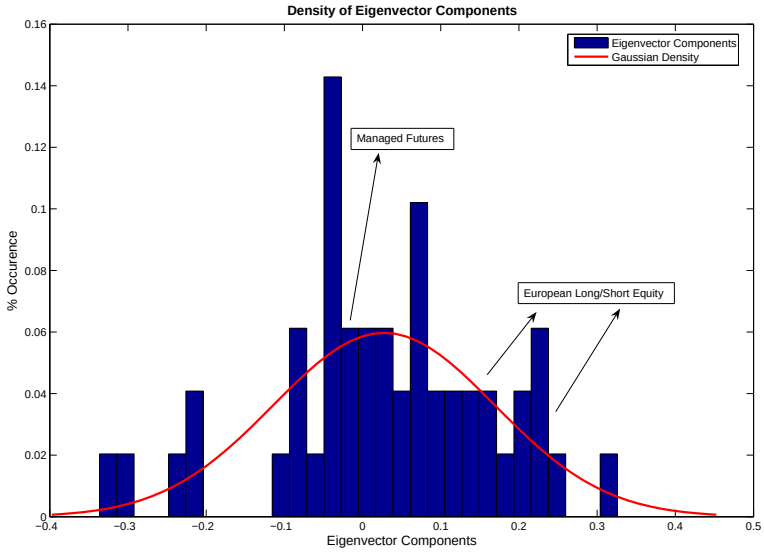


Fig. 4.6: Density of eigenvector components, 2nd largest remaining eigenvalue

4.2.4 Strategy Identification

Following on from the analysis described in Section 4.2.3, we looked also at the contribution of each strategy grouping, $X_l^k \equiv \sum_{i=1}^{N} P_{li}^k \left[u_i^k\right]^2$, to each of the deviating eigenvalues from the different strategies employed, (Section 3.4.1). Fig. 4.7 shows X_l^k for the largest remaining Eigenvector once the effects of the market eigenvalue was removed. The largest contributors were clearly Managed Futures and Emerging Markets, although the strategy contribution for Managed Futures was only around twice that for many of the other sectors. Hence, Managed Futures and Emerging Markets clearly seemed to be the dominant strategies. However, care in interpretation is needed, since neither contributor was dominant overall and there may be overlaps. In particular, as Managed Futures managers may trade the currencies of Emerging Markets, it may also be that there is a closer relationship between these two strategies than is obvious.

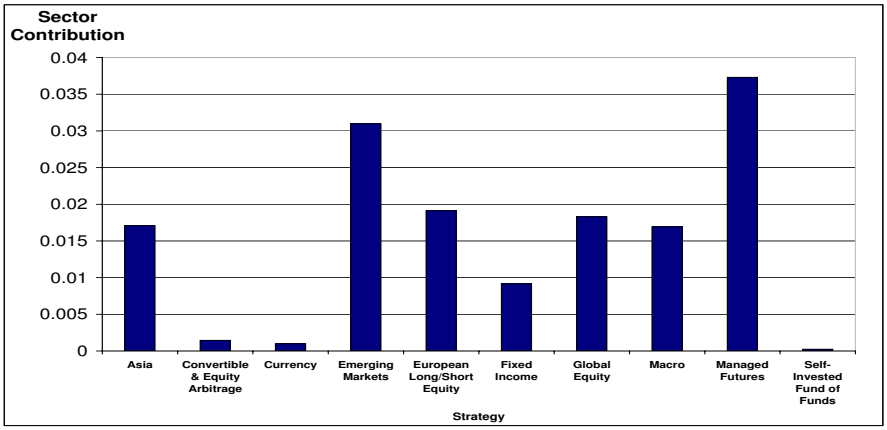


Fig. 4.7: Strategy Contribution, largest eigenvalue

Fig. 4.8 shows the strategy contributions for the second largest eigenvalue. These are interesting, since three of the four dominant strategies (Asia, Global Equity & European Long/Short Equity) are *equity strategies* and are all affected by events in world equity markets. As funds appearing in the same eigenvalue are *highly correlated*, this means that

it is *difficult to diversify by investing in equity funds*. In addition, the fourth strategy, Self Invested Fund of Funds, may well also consist of equity funds. However, information is limited on exactly what type of funds the managers were invested, although there is reason to believe that a majority of them would be equity based, (as equity based funds account for the largest proportion of Hedge Fund assets under management, [Barclays, 2009; CreditSuisse, 2009]). This implies, albeit tentatively, that this eigenvalue contains information primarily on equity funds.

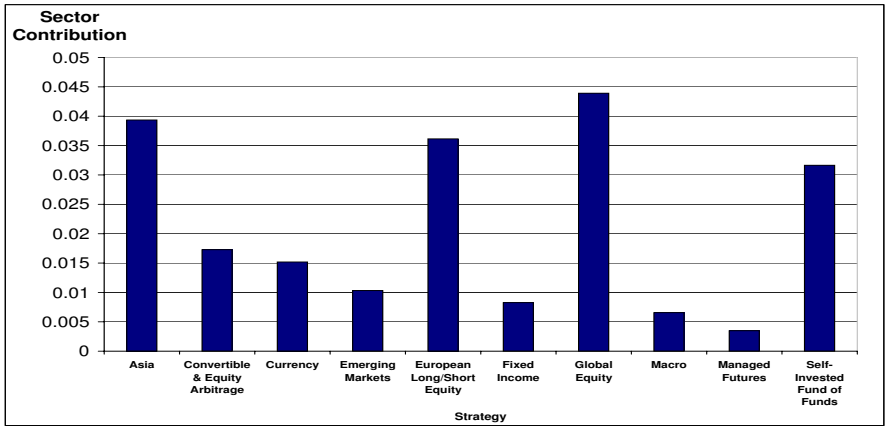


Fig. 4.8: Strategy Contribution, 2nd largest eigenvalue

Fig. 4.9 contains the strategy contributions for the third largest remaining eigenvalue. Clearly, the dominant strategy here was Currency. It was notable in the above, that analysis of the eigenvalues from *within the random matrix region* appeared to reveal no dominant strategies. The evidence of strategy information in the deviating eigenvalues, coupled with a lack of dominant strategies within the RMT region, strongly supports the idea that information in the correlation matrix is chiefly contained within the deviating eigenvalues. Additionally, as these are orthogonal to each other, diversification is accomplished by investing in funds from each of the deviating eigenvalues. Focus on the deviating eigenvalues, enables the creation of a portfolio with an improved risk-return profile, (described section 4.2.6). An alternative method of investigating strategy clustering was also examined through the

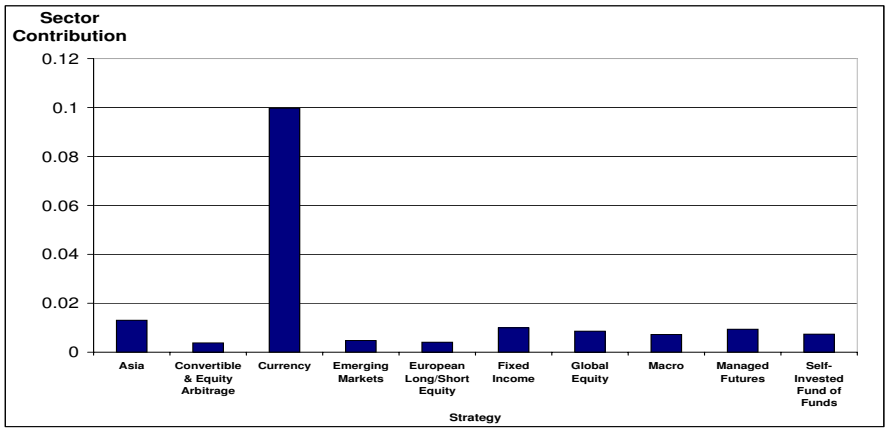


Fig. 4.9: Strategy Contribution, 3rd largest eigenvalue

minimum spanning tree, [Miceli, 2004], but given the small amount of data available within each cluster the results were found to be unstable.

4.2.5 Inverse Participation Ratio

The value of the inverse participation ratio (IPR) for Hedge Fund data is the additional insight into the number of components contributing to the eigenvectors of both large and small eigenvalues. Fig. 4.10 shows the IPR calculated for the eigenvectors of the Hedge Fund cross-correlation matrix studied. The average IPR value was around 0.06, larger than would be reasonably expected ($\frac{1}{N} \approx 0.02$) if all components contributed to each eigenvector. The IPR of the largest eigenvector is smaller, as expected, since it's corresponding eigenvalue reflects the Market or the influences that affect all Funds. However, the deviating eigenvalues, highlighted earlier, are not obvious in the eigenvector components, as is found for equity markets, [Plerou *et al.*, 2002]. The small sample size, may mean that the IPR was not particularly effective in terms of assessing by how much larger eigenvectors deviate from the random region, due to finite size limitations, (expected IPR relies on a sample size that tends to infinity).

A significant deviation from this average IPR value was found for the first few eigenvectors. In particular, for the IPR of the smallest eigenvector this was due to the inclusion of two funds, identical apart from being in different base currencies, with correspondingly high correlation, (≈ 1). This lends support, even for sparse data, to the decoupling of funds with a correlation coefficient much greater than the average, as described Plerou *et al.* [2002].

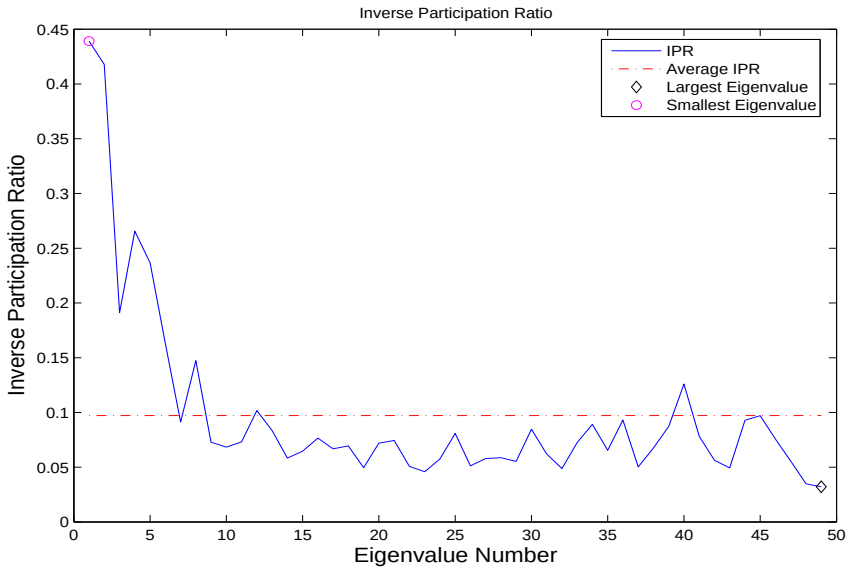


Fig. 4.10: Inverse Participation Ratio as a function of eigenvalue λ

4.2.6 Noise Removal and Portfolio Optimisation

It was noted earlier that where the time-series available to estimate cross-correlation matrices are of limited length, noise effects can be exaggerated. This problem is particularly prevalent with sparse Hedge Fund data, since only monthly returns are available. The return of a portfolio of N Hedge Funds is given given by

$$\Phi = \sum_{i=1}^{N} w_i G_i, \tag{4.1}$$

where $G_i(t)$ is the return on asset i, w_i is the fraction of wealth invested in asset i and $\sum_{i=1}^{N} w_i = 1$. The risk of holding this portfolio is then given by

$$\Omega^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j C_{ij} \sigma_i \sigma_j \tag{4.2}$$

where σ_i is the volatility of G_i, and C_{ij} are the elements of the cross-correlation matrix.

Using the methodology described in Section 3.4.3, we found realised risk to be, on average, 292% of the predicted risk, Fig. 4.11. This large difference between predicted and realised risk has obvious and serious consequences for risk management. Cleaning the correlation matrix reduced the difference by more than a third, (realised risk was 190% of predicted risk). This huge improvement was brought about by limiting the correlation matrix to the information band prescribed by RMT. Some efforts to enhance the cleaning techniques have been reported by [Sharifi *et al.*, 2004; Daly *et al.*, 2008], however these methods may be less effective on Hedge Fund data due to the difficulty in fitting given the small number of eigenvalues.

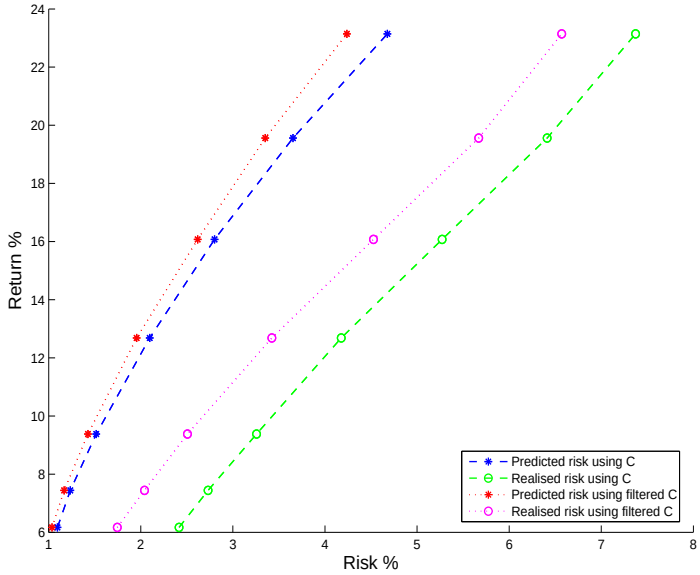


Fig. 4.11: Efficient Frontiers using original and cleaned correlation matrices

It can be seen in Fig. 4.11 that, for some return values, the predicted risk, (using the filtered correlation matrix), was actually less than that of the original correlation matrix. This may have been due to the constraints imposed on the portfolio, in particular the restriction of no 'short-selling', (Section 3.3.1). This constraint is natural in the context of Hedge Funds since, given their illiquid nature, it may be difficult to secure the borrow necessary to implement the short-sale.

The use of the cleaned correlation matrix, led to a 35% improvement in the difference between the realised risk and the predicted risk for the optimal portfolio, even with the imposed 'short-selling' restrictions. This demonstrates that, even for sparse Hedge Fund data the correlation matrix has a role in risk reduction through diversification.

4.3 Conclusions

We have illustrated that, even with limited low-frequency Hedge Fund data (105 months of returns data for 49 Hedge Funds), useful information can be extracted from the cross correlation matrix constructed. Significant deviations from Random Matrix Theory predictions were observed, while further analysis showed that there was real strategy information contained within the deviating eigenvalues. Eigenvector analysis revealed distinct *strategy clustering* in the deviating eigenvectors. These strategy effects included Emerging Markets and Managed Futures in the largest eigenvector, Equity funds in the second, Currency and Fund of Funds in the final two deviating eigenvectors. The strategy information in the deviating eigenvalues was then used to clean the correlation matrix, to alleviate noise effects due to the small sample. Construction of a portfolio using classical portfolio optimisation techniques, (Section 3.3.1), demonstrated that diversification benefits could be found using the correlation matrix, even for low-frequency data. Further, application of the RMT filtering technique, showed how financial risk management of a portfolio of Hedge Funds can be addressed, with a 35% improvement achieved between the risk of the predicted and realised portfolios.

CHAPTER 5

WAVELET MULTISCALE ANALYSIS FOR HEDGE FUNDS: SCALING AND STRATEGIES

In Chapter 4, we demonstrated that an equal-time cross-correlation matrix constructed using low-frequency Hedge Fund returns, contains real strategy information. We also showed, using Random Matrix Theory, that it is possible to clean the correlation matrix in order to create a portfolio with an improved risk-return profile, [Conlon *et al.*, 2007]. However, as mentioned earlier (Chapter 2), the diversification benefits found from Hedge Funds with low or negative correlation to more traditional investments such as equities, may be misleading. This is due to the fact that Hedge Funds may hold illiquid exchange traded assets or over-the-counter (OTC) securities, which are difficult to price accurately. Problems may also arise with intra-strategy correlations for Hedge Fund portfolio optimisation, due to misleading correlations calculated using unfiltered data.

In this Chapter, we develop a novel methodology to overcome this difficulty. Application of the Wavelet Transform to the correlation between Hedge Funds and the Market and, also, to inter-strategy correlations across different time-scales is presented. The Wavelet Transform allows the study of scaling in the correlation matrix, (see remarks on emergence in Complex Systems, Section 2.2). By examining scaling properties, we demonstrate how the correlation matrix can be used to improve the risk management for portfolios of Hedge Funds.

5.1 Data

To enable us to study the market risk and correlations of Hedge Funds at different time scales, we examine returns from the Credit Suisse/Tremont Hedge Fund indices from April 1994 to October 2006, a total of 151 months, [CreditSuisse, 2009]. These indices use the Credit Suisse/Tremont database, which tracks over 4500 funds and consists only of funds with a minimum USD 50 million under management, a 12 month track record and audited financial statements. The indices are calculated and rebalanced on a monthly basis and are net of all performance and management fees. The data used here differs from that in Section 4, as Index data was available over a longer time-frame and the helps in the calculation of wavelet coefficients to longer scales.

The funds in the Credit Suisse/Tremont database are separated into 10 primary categories based on their investment style. From this universe, Credit Suisse/Tremont selects a subset of funds for inclusion in the sub-indices such that each sub-index represents at least 85% of the assets under management in that respective category. The sub-indices are Convertible Arbitrage, Dedicated Short Bias, Emerging Markets, Equity Market Neutral, Event-Driven, Fixed Income Arbitrage, Global Macro, Long/Short Equity, Managed Futures and Multi-Strategy, (in Appendix A we give a full description of each strategy). Each of these strategies is distinct both in the instruments used and the types of markets traded and detailed descriptions of each can be found in Lhabitant [2002, 2004]; CreditSuisse [2009].

The S&P 500 index is chosen to represent the "*Market*", [Plerou *et al.*, 2002]. The S&P 500 is a cap-weighted equity index consisting of large publicly held companies that generally trade on major US stock exchanges such as the New York Stock Exchange and NASDAQ. Monthly returns data for the S&P 500 are available to download publicly on the internet, [Yahoo, 2009].

5.2 Methods

For the present study, we selected the least asymmetric (LA) wavelet (known as the *Symlet*) which exhibits near symmetry about the filter midpoint. LA filters are available in even widths, with optimal filter width dependent on the characteristics of the signal and the length of the data. The filter width chosen for this study was the LA8 (where the 8 refers to the width of the scaling function), since it enabled us to accurately calculate wavelet Betas and Correlations to the 4^{th} scale, given the length of data series available. Although the MODWT, Section 3.6, can accommodate any level J_0, in practise the largest level is chosen so as to prevent decomposition at scales longer than the total length of the data series, hence the choice of the 4^{th} scale. The MODWT was implemented using the *WMTSA Wavelet Toolkit* for Matlab, also featured in Percival and Walden [2000].

5.3 Results

5.3.1 Scalogram

In order to demonstrate graphically the dependence of the S&P 500 and the Credit Suisse/Tremont indices on both scale and time, we examine a scalogram produced using the wavelet transform, Fig. 5.1. The Scalogram shows the size of each wavelet detail coefficient across both scale and time. Examination of the scalogram also helps to reveal information about the scale dependence of features in the time-series examined. For example, Fig. 5.1(a), shows a scaleogram of the S&P 500 for the period studied, Section 5.1, with distinct behaviour found across the 20 scales studied. In particular at time-period 95, corresponding to market behaviour after the events of September 11^{th} 2001, coefficients up to the 12^{th} cofficient are raised. This means that the events of September 11^{th} influenced market movements for a period of 12 months after.

Looking at the returns of the Credit Suisse/Tremont Composite Index over the same period, coefficients are found to be raised at much longer scales, with little evidence of raised coeffients found for the Equity Market Neutral strategies, (so the strategy was little effected

by market events at this time). Another benefit of this Scalogram analysis is the ability to examine concurrent changes for different strategies. The example described, for September 11^{th}, shows an increase in synchronisation between wavelet coefficients between the S&P 500 and the Hedge Fund Composite Index. However, little change in synchronisation is evident between either the S&P or Composite Index and the Equity Market Neutral strategy, implying that this strategy would have diversification benefits for a portfolio. We examine this scale dependent synchronisation in greater detail below.

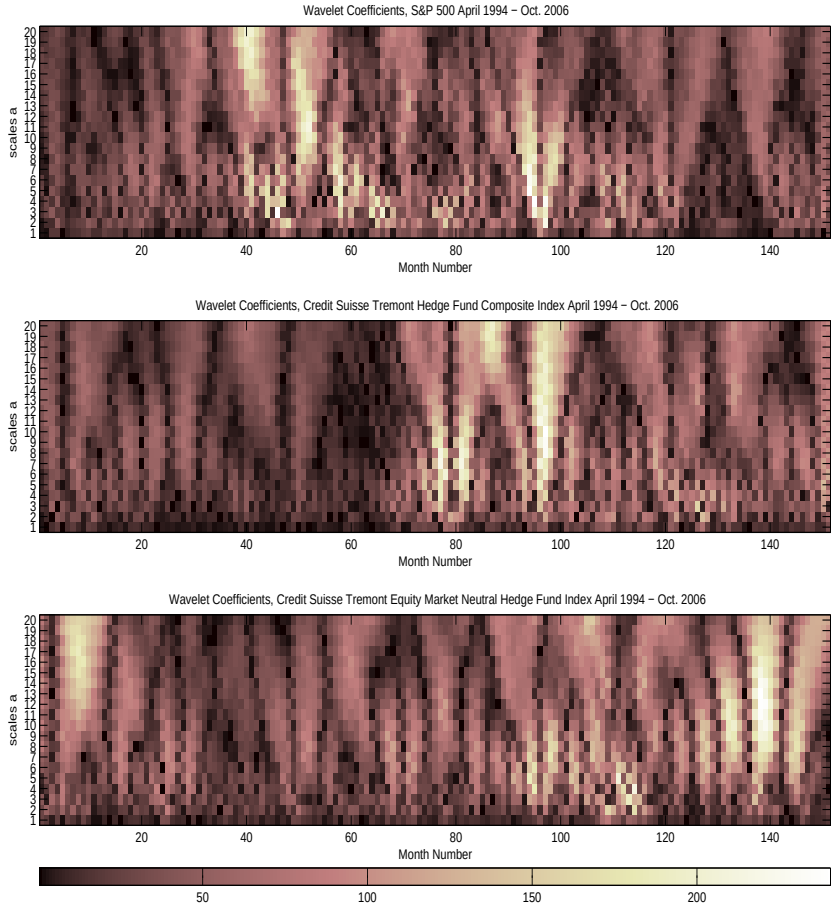


Fig. 5.1: Scalogram showing scaling behaviour over time for the S&P 500, CS/Tremont Composite and Equity Market Neutral Indices.

5.3.2 Correlation Analysis

Applying the MODWT, (Chapter 3), we examined the correlation between the S&P 500 index and the Credit Suisse/Tremont Hedge Fund composite and sub-indices for different time horizons. The correlations were calculated for scales 1, 2, 3 and 4 corresponding to 2-4 month, 4-8 month, 8-16 month and 16-32 month periods respectively, as well as for the original monthly returns data. The results are shown in Table 5.1.

Scale	*Original Data*	*1*	*2*	*3*	*4*
Hedge Fund Index	0.47	0.50	0.40	0.50	0.36
Convertible Arbitrage	0.13	0.07	0.12	0.35	0.31
Dedicated Short Bias	-0.75	-0.78	-0.76	-0.80	-0.88
Emerging Markets	0.48	0.59	0.53	0.47	0.39
Equity Market Neutral	0.37	0.43	0.41	0.30	0.13
Event Driven	0.55	0.56	0.58	0.71	0.51
Fixed Income Arbitrage	0.00	0.05	-0.19	0.08	0.23
Global Macro	0.22	0.31	0.16	0.16	0.00
Long/Short Equity	0.58	0.59	0.54	0.60	0.53
Managed Futures	-0.14	0.00	-0.25	-0.19	-0.61
Multi-Strategy	0.10	0.03	0.03	0.16	0.25

Tab. 5.1: Correlations between the S&P 500 and the Credit Suisse/Tremont Hedge Fund indices, (Increasing correlations - Blue, decreasing correlations - Red)

The correlations for the Hedge Fund composite index with the S&P were found to vary according to the time period considered, in a similar fashion to those of the Long/Short Equity strategy. This may be, at least partly, accounted for by the fact that Long/Short Equity was the component with the largest weighting in the index (28.8%, [CreditSuisse, 2009]). The correlation for Convertible Arbitrage, Fixed Income Arbitrage and Multi-Strategy with the S&P was found to increase as the time horizon increases which may be evidence that these strategies hold less liquid or hard-to-price securities (resulting in a “smoothing” of their returns). The correlations for Dedicated Short Bias, Equity Market Neutral and Managed Futures were found to decrease significantly as the time horizon increases. Hence, these strategies may be *good diversifiers* over a longer time period and would be a useful addition to a market portfolio. In particular, those strategies with negative correlation would be expected to product positive returns when S&P returns are negative, a result which has

been evident throughout the difficult period suffered by the S&P recently. Although some of these correlations may not vary much, (such as Long/Short Equity, 0.58 to 0.53), some variation by scale may have a marked effect on the optimal weights within a portfolio, particularly if two strategies in the portfolio have similar correlations.

The decrease in correlation over longer time scales for the latter strategies is in direct contrast to the *Epps* effect, [Epps, 1979; Tóth and Kertész, 2007b], where the correlation between financial time series was shown to increase over long time scales. However, in the case of the *Epps* effect the time scales studied were much shorter ($0 - 2\frac{1}{2}$ hours) and the data much more plentiful, [Tóth and Kertész, 2007b].

Using the Credit Suisse/Tremont Hedge Fund Composite index, we also measured the correlation between this and each of the sub-indices over different time horizons. Again, the correlations were calculated for scales 1, 2, 3 and 4 as above. The results are shown in Table 5.2.

Scale	*Original Data*	*1*	*2*	*3*	*4*
Convertible Arbitrage	0.40	0.23	0.36	0.53	0.63
Dedicated Short Bias	-0.48	-0.50	-0.48	-0.50	-0.39
Emerging Markets	0.66	0.71	0.65	0.66	0.64
Equity Market Neutral	0.32	0.36	0.34	0.12	0.07
Event Driven	0.68	0.61	0.70	0.72	0.69
Fixed Income Arbitrage	0.41	0.28	0.33	0.57	0.57
Global Macro	0.85	0.88	0.87	0.85	0.79
Long/Short Equity	0.79	0.76	0.83	0.85	0.82
Managed Futures	0.17	0.35	0.22	-0.01	-0.58
Multi-Strategy	0.22	0.01	0.31	0.30	0.50

Tab. 5.2: Correlations between the Credit Suisse/Tremont Hedge Fund Composite index and the sub-indices, (Increasing correlations - Blue, decreasing correlations - Red)

The results for the inter-strategy correlations with the Hedge Fund Composite Index were somewhat similar to those with the S&P 500. The correlation of Convertible Arbitrage, Fixed Income Arbitrage and Multi-Strategy with the Hedge Fund composite index was found to increase as the time period increased. The correlation of Dedicated Short Bias, Managed Futures, Equity Market Neutral and Global Macro to the Composite Index decreased significantly as the time horizon increased. Hence, these strategies may offer *real diversification* benefits over a longer time horizon, while those with increasing correlations

may actually *increase the portfolio risk*. We, therefore, examine the effects of the change in correlations at different time horizons on the efficient frontier of a portfolio made up of the Hedge Fund strategies in Section 5.3.4.

5.3.3 Systematic Risk Analysis

It was noted, Chapter 2, that systemic risk or Beta for equities is dependent on the scale studied. We measured the Beta or market risk of various Hedge Fund strategies to the S&P 500 over different time horizons using the MODWT, (Chapter 3). Once again, the correlations were calculated for scales 1, 2, 3 and 4 as defined in Section 5.3.2. The results are displayed in Table 5.3.

Scale	*Original Data*	*1*	*2*	*3*	*4*
Hedge Fund Index	0.25	0.25	0.21	0.31	0.26
Convertible Arbitrage	0.04	0.01	0.03	0.16	0.19
Dedicated Short Bias	-0.90	-0.85	-0.94	-1.22	-1.49
Emerging Markets	0.53	0.52	0.59	0.77	0.63
Equity Market Neutral	0.07	0.07	0.08	0.08	0.04
Event Driven	0.21	0.17	0.22	0.38	0.39
Fixed Income Arbitrage	0.00	0.01	-0.05	0.02	0.08
Global Macro	0.16	0.23	0.11	0.12	-0.01
Long/Short Equity	0.41	0.37	0.39	0.55	0.46
Managed Futures	-0.11	0.00	-0.23	-0.22	-0.61
Multi-Strategy	0.03	0.01	0.01	0.05	0.10

Tab. 5.3: Betas of the Credit Suisse/Tremont Hedge Fund indices, (Increasing beta - Blue, Decreasing beta - Red)

The Beta of the Composite Hedge Fund Index appears to be reasonably static but the *market risk* of its components varies according to strategy. Convertible Arbitrage had a Beta of 0.04 using monthly data, as opposed to 0.19 using a $16-32$ month time horizon. This difference may be due to the fact that there is no exchange for Convertible Bonds and they are traded "over the counter" and hence can be illiquid and difficult to price.

The Beta of the Emerging Markets strategy increased from 0.53 to a maximum of 0.77 over an 8-16 month period. This may be due to liquidity constraints in emerging markets with light markets in equities causing instruments to be marked off non-synchronous

prices, since the final traded price for an asset may have been before month-end. The effect of liquidity issues in Emerging Markets has been studied in detail, [Wilcox and Gebbie, 2004]. Another possible factor in the example here might be the difficulty in short-selling instruments in many emerging markets; hence these funds often employ a more traditional long-only type strategy.

The market risk of the Event Driven strategy was shown in Table 5.3 to increase from 0.21 to 0.39 over a longer time horizon, as was the case for Equity Long/Short strategy which increases from 0.41 to a maximum of 0.55. In both cases these increases could be due to a number of factors, such as liquidity, the holding period of the manager or the time for the positions in the portfolios to reach a "fair value". Another significant consideration might be the use of *options* which have a non-linear returns profile, [Hull, 2000], and hence might lead to an overall increase in Beta. Furthermore Long/Short Equity managers may also hold significant positions in small capitalization stocks or illiquid private securities which may trade infrequently and hence are difficult to price.

Equity Market Neutral, Fixed Income Arbitrage and Multi-Strategy all appear to have very small and static levels of market risk over all time horizons. Equity Market Neutral, as the name suggests, seeks to exploit pricing efficiencies between equity securities while simultaneously neutralising exposure to market risk. Fixed Income Arbitrage managers invest solely in Bonds, (whether sovereign or corporate), and hence have little or no equity market risk. Multi-Strategy funds implement a dynamic strategy allocation to various Hedge Fund strategies. All of these strategies should provide good diversification to a portfolio made up of S&P equities.

The market risk of the Dedicated Short Bias and Managed Futures strategies was found to decrease by a considerable amount as the time horizon increases. Dedicated Short-Bias funds are, in a sense, mirrors of traditional long-only managers. The increase in the absolute value of their Beta may be due to excessive leverage in *Bear Markets*[1] or the use of *Put Options*[2] with a non-linear return profile. Managed Futures managers tend to use fore-

[1] A Bear Market is considered to be a prolonged period in which asset prices fall, accompanied by widespread pessimism

[2] A put is an option contract that gives the holder the right to sell a certain quantity of an underlying security

casting models that depend heavily on past price movements. The time horizons of their trades can vary greatly depending on the individual model's time horizon; hence this may have a large impact on the aggregate amount of market risk held at different time intervals. In fact, both of these strategies proved to be among the best performers during 2008, with Short Bias and Managed Futures returning 14.87% and 18.33% respectively, [CreditSuisse, 2009]. This was against a background of severe market losses, with the S&P 500 losing 38.5% over the same period, [Yahoo, 2009], demonstrating the real diversification benefits of the strategies.

In summary, the level of market risk held by different strategies varied greatly according to strategy type and time horizon considered. Given the fact that many funds have minimum lock-up periods, during which investors cannot liquidate their capital, it is more appropriate to measure the market risk over longer time intervals. This has obvious consequences for the allocation strategies of Fund of Funds[3] and Institutional investors alike, as we demonstrate in the following Section.

5.3.4 Efficient Frontier

To enable us to demonstrate graphically the effects of calculating the correlation matrix of Hedge Funds using different time horizons, we show, (Figure 5.2), the efficient frontier for each time horizon studied. In the calculation of the efficient frontier we have imposed constraints on the weights allowed, to prevent short-selling of Hedge Funds. This constraint is natural in the context of Hedge Funds due to the difficulties in short-selling of funds; (note that short-selling may be achievable by the use of swaps but is uncommon due to liquidity constraints in borrowing funds), [Lhabitant, 2002, 2004]. Each of these efficient frontiers was calculated as described, Section 3.3.1, using the annualised returns and variance from the original data. Hence, the results we see in Figure 5.2 were purely dependent on the correlation matrix calculated at each time horizon. The efficient frontiers were constructed using just the Credit Suisse/Tremont sub-indices.

to the writer of the option, at a specified price up to a specified date

[3]A Hedge Fund which invests in other hedge funds

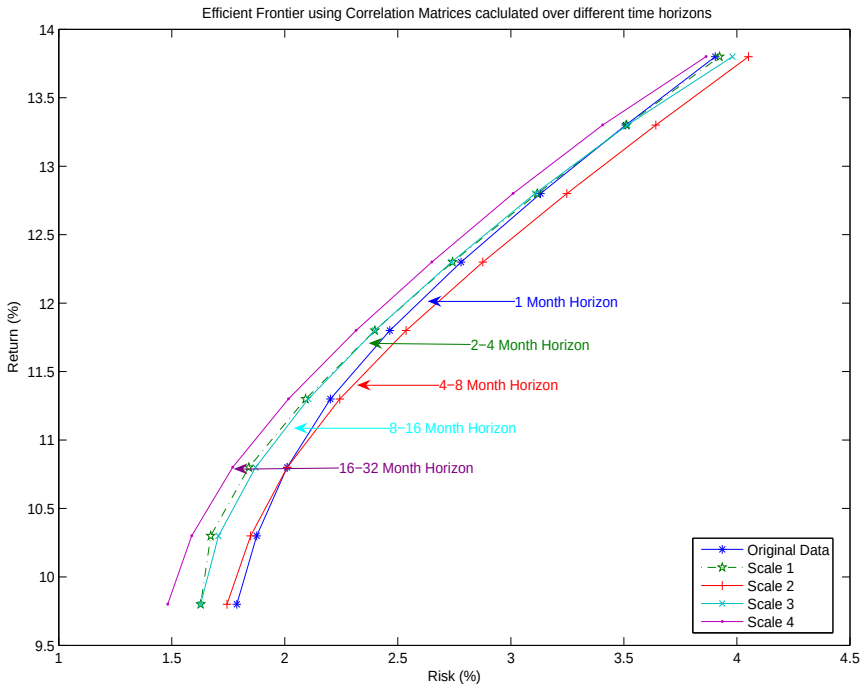


Fig. 5.2: Efficient Frontier using Correlation Matrices calculated over different time scales

At lower return values, the level of risk calculated for the original data was higher than that calculated at all other scales. As the return level increased, the risk of the portfolio (created using the correlation matrix calculated using the monthly returns), decreased relative to the others. However, the lowest risk portfolio overall was found using a correlation matrix calculated over a $16-32$ month horizon (Scale 4). The implication for Hedge Fund investors using monthly unfiltered correlations, is an *unnecessary increase in portfolio risk*. By using correlations calculated over longer horizons, investors can capture subtle changes in the relationships between funds and overcome the problem of stale pricing.

These findings have significant implications for investors allocating capital to Hedge Funds. For investors looking for diversification through investments in Hedge Funds, the expected correlation benefits may not be found by using unfiltered data. In fact, the expected risk reduction from investing in different Hedge Fund strategies may result in an increase in risk due to misleading correlations found using the original data. By filtering

the data, the effects of stale pricing are mitigated and a realistic view of the true correlation structure emerges. By adopting a novel filtering technique such as this, Fund of Funds could enhance their client offerings, lowering portfolio risk while maintaining their expected returns profile.

5.4 Conclusions

Maintaining our novel focus on Hedge Funds and the risk exposure which can be determined from an examination of the correlation structure for coarse/fine grained time series, we investigated strategies in detail to determine diversification benefits over different scales. Correlation and market risk were found to be scale dependent between Hedge Fund substrategies and both the S&P 500 and the Hedge Fund composite index. In particular, the correlation between Convertible Arbitrage, Fixed Income Arbitrage and Multi-Strategy and both the S&P and Hedge Fund Composite index was found to increase as the time scale increased, while Dedicated Short Bias, Equity Market Neutral, Global Macro and Managed Futures strategies were found to decrease.

The level of market risk held by different Hedge Fund strategies was also found to be highly dependent on the time horizon studied. For Convertible Arbitrage, Emerging Markets, Event-Driven and Long/Short Equity it was found to increase as the time scale increased, while a decrease in market risk was found for Dedicated Short Bias, Global Macro and Managed Futures.

Given the minimum lock-up times applicable to many Hedge Funds, this means that it may be more suitable when assessing risk to use correlation matrices calculated using longer time scales. To demonstrate this, we calculated the efficient frontier for four different time scales, as well as for the original, (unfiltered data), for a portfolio constructed from the substrategies. Specifically, the level of risk for a portfolio of funds attained it's lowest value over $16 - 32$ months. This scale is significant as Hedge Funds often have lock-up period of $12 - 24$ months during which investors cannot withdraw their invested capital, [Lhabitant, 2002], so that calculation of correlations over a similar horizon is more appropriate.

CHAPTER 6

CROSS-CORRELATION DYNAMICS IN FINANCIAL TIME SERIES

Previous Chapters have dealt with correlation structures from low-frequency time-series, in particular those formed from Hedge Fund returns. We have shown that it is possible to extract real correlation information from a matrix constructed using low-frequency returns, while also developing a technique to examine the scaling effects on Hedge Fund correlations, [Conlon *et al.*, 2007, 2008]. These approaches were applied to the risk management of a portfolio of Hedge Funds.

In this Chapter, we turn our attention to the dynamical changes in correlations between financial assets, [Conlon *et al.*, 2009]. Matrices formed from medium-frequency data (daily equity returns) are analysed, to investigate whether changes in the correlation structure of a complex interacting system can be applied to risk characterisation. This analysis is performed for varying time windows and number of stocks in the sample, in order to demonstrate the results for different levels of granularity.

A basic one-factor model is then proposed, to develop some insight into the formation of the correlation structure in financial markets. Perturbations are added to the one-factor model, (leading to a 'market plus sectors' model), in order to determine which eigenspectrum features are a result of sectoral interactions. The Inverse Participation Ratio is also compared to that found for empirical data.

Further, we examine the relationship between index returns and relative eigenvalue size, to provide some insight on the collective behaviour of traders with varying strategies and

consider how changes, in collective behaviour, are dependent on the level of perceived risk in the market.

6.1 Data

In order to study the dynamics of the empirical correlation matrix over time, we chose to analyse two different data sets. The first data set comprises the 384 equities of the Standard & Poors (S&P) 500, [Standard and Poor's, 2009], where full price data is available from January 1996 to August 2007 resulting in 2938 daily returns, Fig. 6.1(a). The S&P 500 is an index consisting of 500 large capitalisation equities, which are predominantly from the US. The advantage of this first dataset is that the components are well-regarded, frequently traded and liquid, giving us a robust framework to study correlation dynamics.

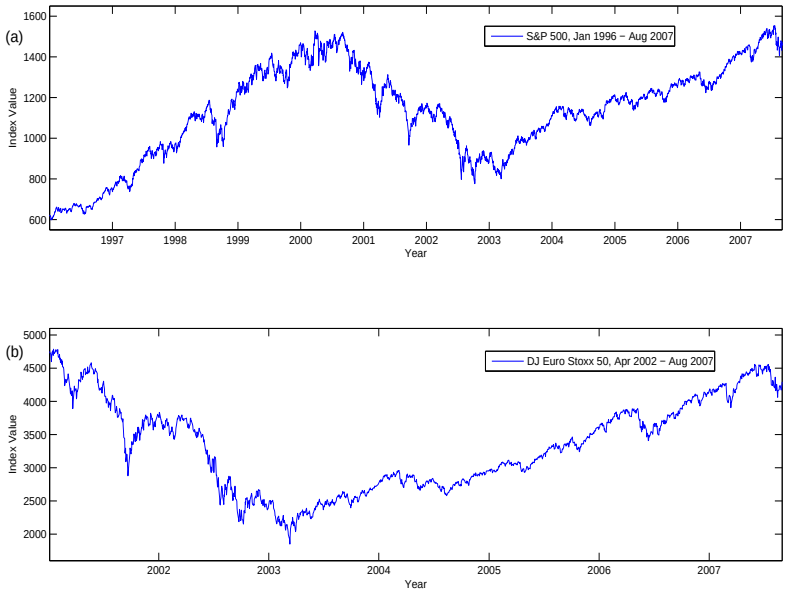


Fig. 6.1: (a) S&P500 Index, Jan. 1996 to Aug. 2007 (b) Dow Jones Euro Stoxx 50 Index, Jan. 2001 - Aug. 2007

In order to demonstrate that our results are not market specific, however, we also examined a second data set, made up of the 49 equities of the Dow Jones Euro Stoxx 50,

[Dow Jones, 2009], where full price data is available from January 2001 to August 2007 resulting in 1619 daily returns, Fig. 6.1(b). The Euro Stoxx 50 is a stock index of Eurozone equities designed to provide a blue-chip representation of supersector[1] leaders in the Eurozone. In contrast to the first data-set, the components of this index are from various nations, introducing an additional factor of interest, namely country dependent interactions.

6.2 Methods

Initially, the correlation matrix and associated eigenvalues are calculated using a sliding window approach, Section 3.2. By repeatedly calculating the eigenspectrum using a sliding window approach, eigenvalue time-series are formed. Changes in the relative size of the eigenvalues are indicative of changes within the correlation structure, Section 3.5.

In order to examine the dependence of the eigenvalue dynamics on the sample size, N, the length of the time series, T, and hence the ratio $Q = \frac{T}{N}$, we perform a number of experiments. Initial analysis focuses on changes in the size of the eigenspectrum, with more detailed studies then looking at the eigenvalues normalised in time. The normalisation, Section 3.5, is carried out using the mean and standard deviation of each of the eigenvalues over the entire time-period. Note that other choices for these, such as during a low-volatility sub-period, would have resulted in a greater emphasis during volatile periods.

To compare the empirical results presented here to those of a single factor model, we construct a correlation matrix with each non-diagonal element equal to the average of the empirical correlation matrix in each sliding window, using Cholesky decomposition. We then calculate and normalise the eigenvalues of this matrix over each sliding window, (Section 3.5).

Drawdown analysis requires the calculation of returns, correlation matrix and eigenvalue spectrum time-series for overlapping windows and these are, again, normalised using the mean and standard deviation over the entire series, (Eqn. 3.19). Representing nor-

[1] Supersector companies are the largest in the sector by market capitalisation. For example, in the European Energy Sector, *British Petroleum*, *Total* and *Royal Dutch Shell* account for 62% of the total 'large cap' market capitalisation, [Yahoo, 2009], and are considered representative of the sector.

malised eigenvalues in terms of standard deviation units (SDU) allows partitioning according to their magnitude.

6.3 Results

Eigenvalue dynamics of the correlation matrix of a subset of 100 S&P equities, chosen randomly, were analysed using a sliding window of 200 days. This subsector was chosen such that $Q = \frac{T}{N} = 2$, to ensure that data would be stationary in each sliding window. Figure 6.2(a) shows broadly similar sample dynamics from the 5^{th}, 15^{th} and 25^{th} largest eigenvalues over each of these sliding windows. The sum of the 80 smallest eigenvalues are shown in Figure 6.2(b), while the dynamics of the largest eigenvalue are displayed in Figure 6.2(c). The repulsion between the largest eigenvalue and the small eigenvalues are evident here, (comparing 6.2(b) and 6.2(c)), with the dynamics of the small eigenvalues contrary to those of the largest eigenvalue. As noted, (Section 3.5), this is a consequence of the fact that the trace of the correlation matrix must remain constant under transformations and any change in the largest eigenvalue must be reflected by a change in one or more of the other eigenvalues. Similar results were obtained for different subsets of the S&P and also for the members of the Euro Stoxx 50.

6.3.1 Normalised Eigenvalue Dynamics

The use of time-normalised eigenvalues to highlight the dynamics of the smaller eigenvalues was described, by Eqn. 3.19. Using these normalised eigenvalues, we performed a number of experiments to investigate the dynamics of a set of small eigenvalues versus the largest. For each of the experiments described below, we plot the largest and the average of a number of the small eigenvalues, in addition to a heat map of the entire normalised spectrum over time.

1. The dynamics for the same subset of 100 equities were analysed using a sliding window of 200 days, as above, with similar normalisation criteria using the mean and

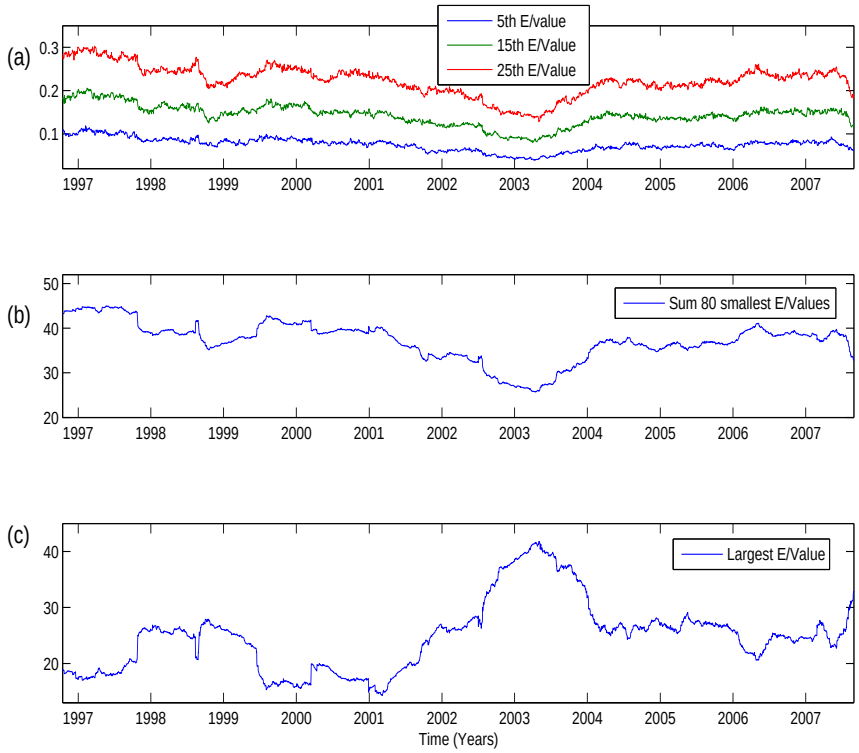


Fig. 6.2: Time Evolution of (a) Three small eigenvalues (b) Sum of the 80 smallest eigenvalues (c) The largest eigenvalue for a random subset of 100 companies of the S&P500 from Jan. 1996 to Aug. 2007

standard deviation of each of the eigenvalues over the entire time-period. Figure 6.1(a) shows the value of the S&P index from 1997 to mid-2007.

The normalised largest eigenvalue is displayed, Figure 6.3(a), together with the average of the 80 normalised small eigenvalues. The compensatory dynamics mentioned earlier are shown clearly here, with these showing opposite movements. The normalised eigenvalues for the entire eigenvalue spectrum are shown, Figure 6.3(b), where the colour indicates the number of standard deviations from the time average for each of the eigenvalues over time. As shown, there appears little to differentiate the dynamics of the $80 - 90$ or so smallest eigenvalues. In contrast, the behaviour of the largest eigenvalue was clearly opposite to that of the smaller eigenvalues. However, from the 90^{th} and subsequent eigenvalue there was a marked change in the

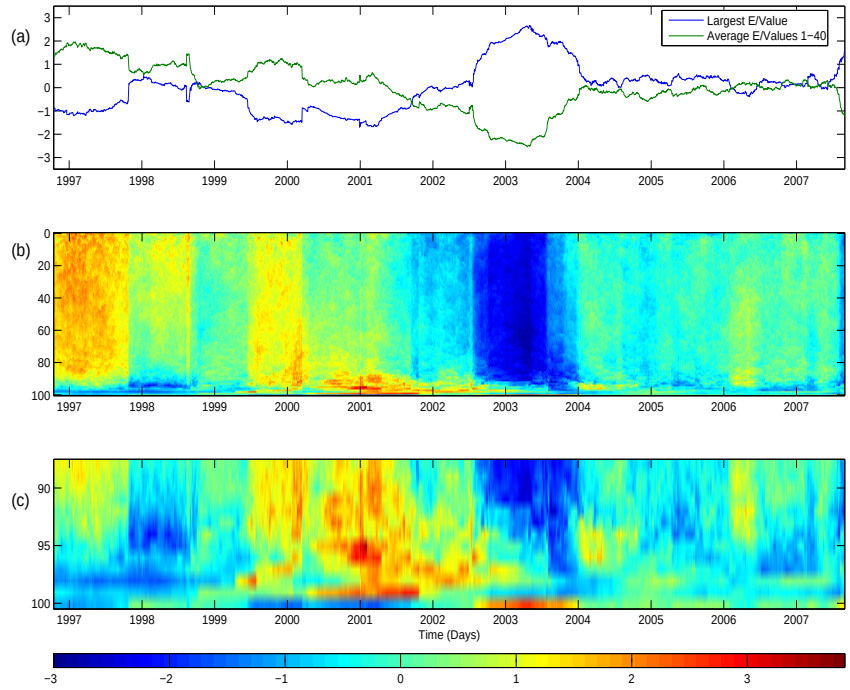


Fig. 6.3: (a) Normalised Largest Eigenvalue vs. Average of 80 smallest normalised eigenvalues (b) All Normalised Eigenvalues (c) Largest 12 Normalised Eigenvalues for a random subset of 100 companies of the S&P500 from Jan. 1996 to Aug. 2007

behaviour, (Figure 6.3(c)), and the eigenvalue dynamics were distinctly different. This may correspond to the area outside the "Random Bulk" in RMT. Similar to [Drozdz *et al.*, 2000, 2001a], we also found evidence of an increase/decrease in the largest eigenvalue with respect to 'drawdowns'/'draw-ups'. Additionally, we found the highlighted *compensatory dynamics* of the small eigenvalues. These results were tested for various time windows and normalisation periods, with shorter durations found to be better able to capture and emphasise additional features.

To illustrate the value of the above results, we look at the dynamics of the largest normalised eigenvalue over time, Fig. 6.3(a). In particular, the largest normalised eigenvalue increased from the beginning of 2001 to mid 2003, corresponding to the bursting of the 'tech' bubble and the start of the second gulf war. This increase in the largest eigenvalue corresponds to an increase in the global system correlation, Sec-

tion 3.5, with more agents acting in the same fashion during times of market stress. During times of market stress, there tends to be a greater synchronisation between the behaviour of traders (in particular, traders tend to sell at the same time), while during normal periods equities become less correlated. Results using the minimum spanning tree have shown similar effects with the size of the tree shown to shrink during market drawdowns, [Onnela *et al.*, 2003].

2. To demonstrate the above result for a *different level of granularity*, we randomly chose 50 equities with a time window of 500 days, giving $Q = \frac{T}{N} = 10$. The results obtained, (Figure 6.4), were in agreement with those for $Q = 2$ earlier, with a broad-band increase (decrease) of the 40 smallest eigenvalues concurrent with a decrease (increase) of the largest eigenvalue.

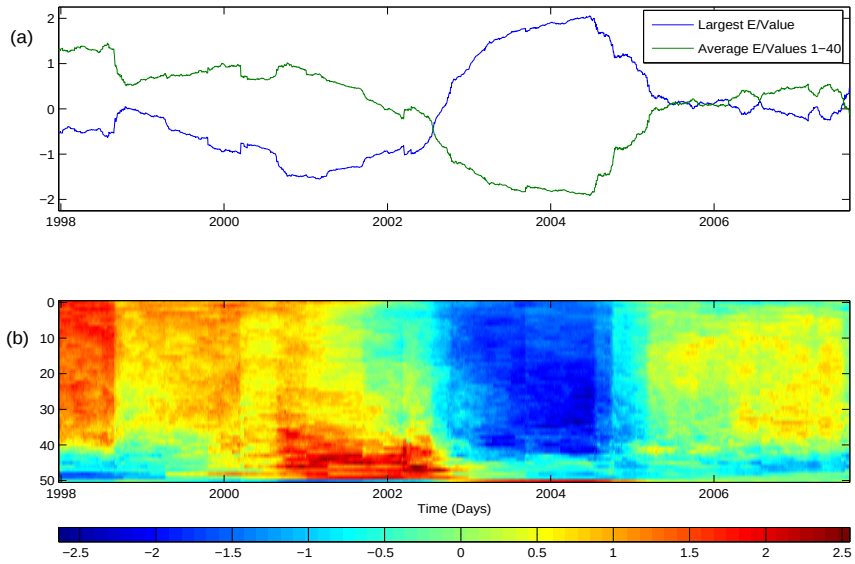


Fig. 6.4: (a) Normalised Largest Eigenvalue vs. Average of 40 smallest normalised eigenvalues (b) All Normalised Eigenvalues for a random subset of 50 companies of the S&P500 from Jan. 1996 to Aug. 2007

3. The previous examples used random subsets of the S&P universe in order to keep $Q = \frac{T}{N}$ as large as possible. To demonstrate that the above results were not sampling artifacts, we also looked at the full sample of 384 equities, (ie those that survived the

entire 11 year period), with a time window of 500 days, ($Q = 1.30$). The results, as shown in Figure 6.5, were similar to before, (Figs. 6.3-6.4), with the majority of the small eigenvalues compensating for changes in the large eigenvalue. As indicated previously, however, there are a few large eigenvalues which exhibited anomalous behaviour compared to both the small and largest eigenvalues.

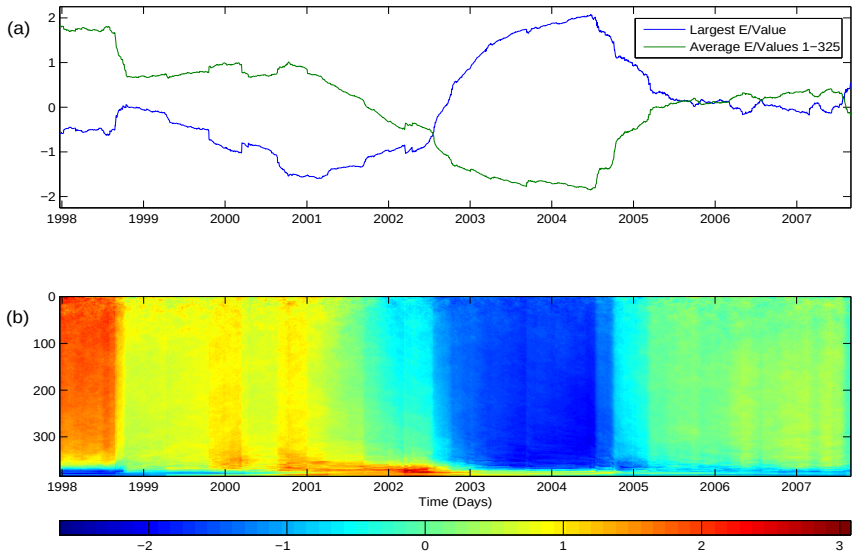


Fig. 6.5: (a) Normalised Largest Eigenvalue vs. Average of 325 smallest normalised eigenvalues (b) All Normalised Eigenvalues for a subset of 384 companies of the S&P500 from Jan. 1996 to Aug. 2007

4. All examples discussed so far have focused on the universe of equities from the S&P 500 that have survived since 1997. To ensure that the results obtained were not exclusive to the S&P 500, we also applied the same technique to the 49 equities of the Euro Stoxx 50 index that survived from January 2001 to August 2007, Fig. 6.1(b). The sliding window used was 200 days, such that $Q = 4.082$. The results, (Figure 6.6), were broadly confirmatory, with a wide band of small eigenvalues "responding to" movements in the largest eigenvalues. In this case, the band of deviating large eigenvalues (ie. those which correspond to the area outside the "Random Bulk" in RMT), (Figure 6.6(c)), was not as marked as in the previous example. This effectively

implies that equities in this index were dominated by “the Market”. This occurs due to the small number of stocks in the data set, which effectively reduces the amount of inter-sector interaction within the system.

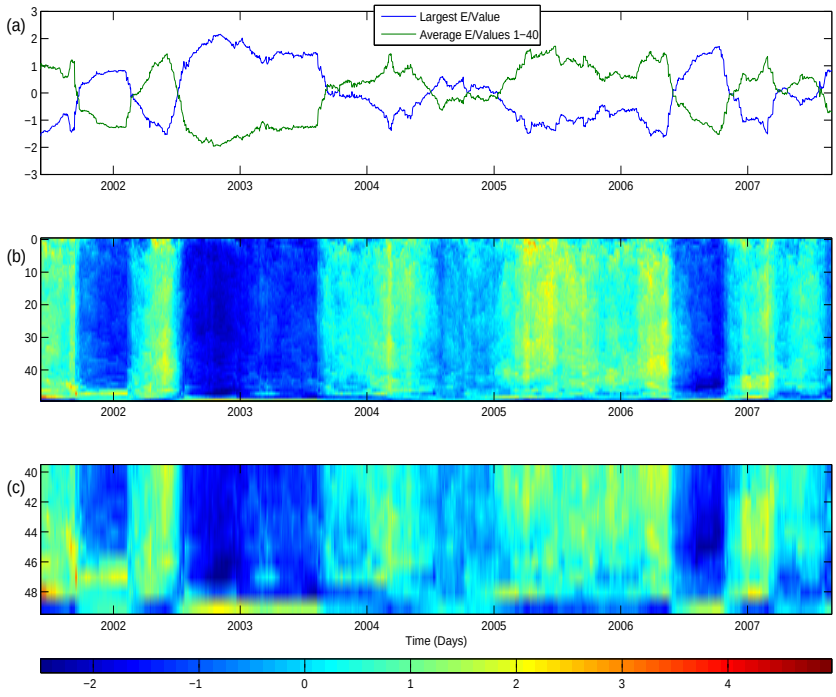


Fig. 6.6: (a) Normalised Largest Eigenvalue vs. Average of 40 smallest normalised eigenvalues for Euro Stoxx 50 (b) All Normalised Eigenvalues (c) The 9 Largest Normalised Eigenvalues from Jan. 2001 to Aug. 2007

6.3.2 Model Correlation Matrix

The results, described, demonstrate that the *time dependent dynamics of the small eigenvalues of the correlation matrix of stock returns move counter to those of the largest eigenvalue*. While not a new idea, this is important in our context because, by analysing changes in the small eigenvalues, we are able to determine changes in the global correlation structure. The value of modelling such features is that we gain some insight into the group behaviour of agents during different market types. Here, we look at possible suitable model forms to

understand this behaviour, Section 3.5.1. We show how a simple one-factor model of the correlation structure reproduces much of the empirical behaviour. Furthermore, we show how additional features can be captured by including perturbations to this model, essentially a *"market plus sectors"* model, [Noh, 2000; Malevergne and Sornette, 2004; Papp *et al.*, 2005].

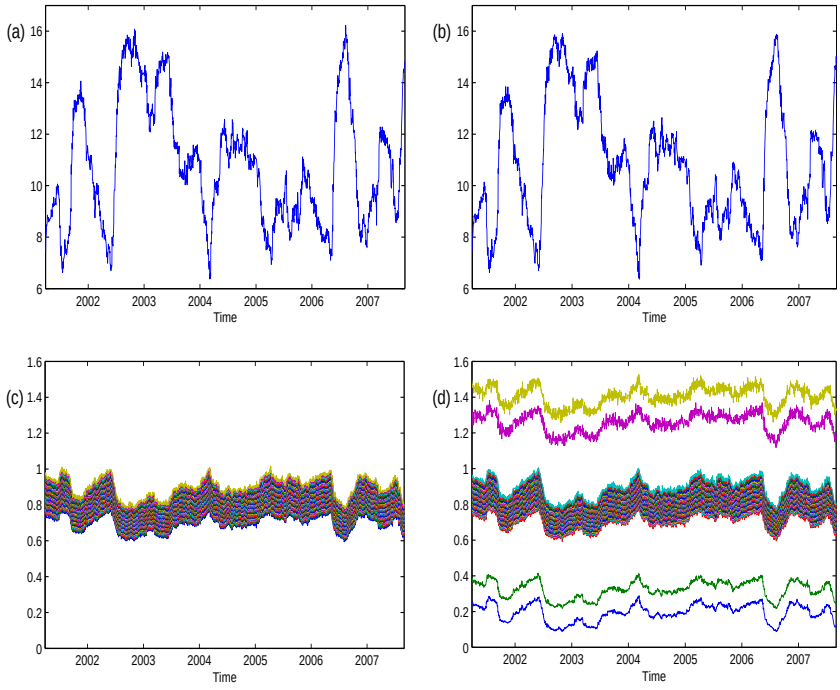


Fig. 6.7: (a) Largest Eigenvalue, one factor model (b) Largest Eigenvalue, Market plus sectors model (c) Eigenvalues 1-48, one factor model (d) Eigenvalues 1-48, Market plus sectors model using average correlation for the Euro Stoxx 50, Jan. 2001 - Aug. 2007

The dynamics of the largest eigenvalue for the single-factor model are displayed in Figure 6.7(a) for the Euro Stoxx 50 index using a sliding window of 200 days. As can be seen, the main features of the dynamics were in agreement with those of Figure 6.6 for the empirical data. The dynamics of the remaining eigenvalues, shown in Figure 6.7(c), were found to be within a narrow range, with any change in time due to compensation for change in the largest eigenvalue.

For the 'market plus sectors' model, we introduced perturbations with two groups of 5 stocks having correlation $\rho_0 - 0.15$ and $\rho_0 + 0.15$, with the average correlation at each time window remaining the same. The largest eigenvalue dynamics for the 'market plus sectors' model are shown in 6.7(b), with these almost identical to those for the largest eigenvalue of the 'one-factor' model, (any differences are due to random fluctuations induced in the Cholesky decomposition). However, looking at the remaining eigenvalues, 6.7(d), we see that a number were found to deviate significantly from the bulk. These deviations were ascribable to the additional sectoral information included in the market plus sectors model. This agrees with previous results, [Plerou *et al.*, 2000a; Utsugi *et al.*, 2004], where small eigenvalues, corresponding to highly correlated stocks as well as large eigenvalues containing sectoral information were found to deviate from the random bulk.

Inverse Participation Ratio

To examine properties of the eigenvector components themselves, we use the Inverse Participation Ratio (IPR), Chapter 3.4.2. Figure 6.8(a), displays the IPR found using the empirical data from the Euro Stoxx 50. This has similar characteristics to those found for different indices, [Plerou *et al.*, 2000a], with the IPR (i) much smaller than the mean for the largest eigenvalue, (ii) large corresponding to sectoral information in the 2^{nd} or 3^{rd} largest eigenvalues and (iii) increased for the small eigenvalues. These features correspond to real correlation information found in the eigenvectors, similar to those found for Hedge Funds, Chapter 4. A small IPR indicates that all the eigenvector components contribute equally, while an IPR larger than the average corresponds to an eigenvector where only a small number of components contribute.

For the single factor model, we created a *synthetic* correlation matrix using Eqn. (3.20), with average correlation (0.204) equal to that of the Euro Stoxx 50 over the time period studied. Using an average value captures the global correlation within the system. As shown in Figure 6.8(b), the IPR retains some of the features found for empirical data, [Plerou *et al.*, 2000a; Noh, 2000], with that corresponding to the largest eigenvector having a much

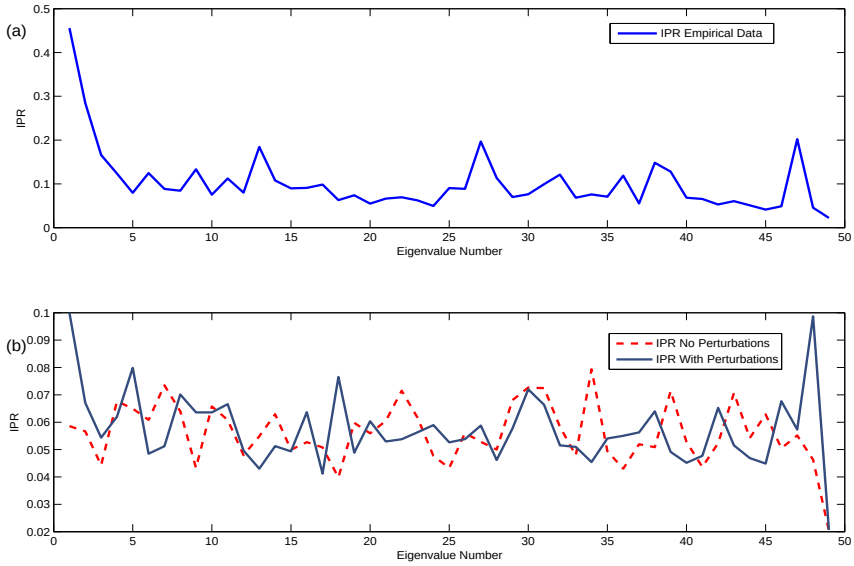


Fig. 6.8: (a) IPR Empirical Data, Euro Stoxx 50, Jan. 2001 - Aug. 2007 (b) IPR Simulated Data

smaller value than the mean. This corresponds to an eigenvector to which many stocks contribute, (effectively the market eigenvector), [Plerou *et al.*, 2000a; Noh, 2000].

For the 'market plus sectors' model, Figure 6.8(b), additional features were found, with larger IPR for both smallest and second largest eigenvalue. This agrees with [Plerou *et al.*, 2000a] where, for empirical data, the group structure resulted in a number of small and large eigenvalues with larger IPR than that of the bulk of eigenvalues. These large eigenvalues were shown, [Plerou *et al.*, 2000a], to be associated with correlation information related to the group or sectoral structure.

6.3.3 Drawdown Analysis

As discussed above, drawdowns, (*periods of large negative returns*), tend to reflect an increase of one eigenstate of the cross-correlation matrix. The opposite occurs during drawups, (*periods of predominantly positive returns*). In this section, we attempt to characterise the market according to the relative size of both the largest and small eigenvalues,

using stocks of the Euro Stoxx 50. Due to the small number of stocks in this index, it was unnecessary to choose a subset to ensure $Q = \frac{T}{N}$ was large.

Eigenvalues	*No. Std*	*Index Return*
Large	$<$-1	18.46%
	$>$1	-16.80%
Average 40 Smallest	$<$-1	-23.90%
	$>$1	19.32%

Tab. 6.1: Drawdown/Drawup analysis, average index returns for eigenvalue partitions in SDU.

Using overlapping windows of 200 days, returns, correlations and eigenvalue spectrum were calculated. Table 6.1 shows the average return of the index, (during periods when the largest eigenvalue is ± 1 SDU), for both the largest eigenvalue and the average of the (normalised) 40 smallest eigenvalues.

The results, Table 6.1, demonstrate that when the largest eigenvalue is > 1 SDU, the average index return over a 200 day period is found to be -16.80%. When it is small (< -1 SDU), the average index return is 18.46%. Hence, *the largest eigenvalue can be used to characterise the risk of markets, with it's value peaking during periods of negative returns (Drawdowns) and bottoming out when the market is rising (Drawup).* For the average of the 40 smallest eigenvalues, the partition also reflected drawdowns and drawups, but with opposite signs. This indicates that *information about the correlation dynamics of financial time series is visible in both the lower and upper eigenstates*, in agreement with [Müller *et al.*, 2005; Schindler *et al.*, 2007a] for both synthetic data (and interestingly, for EEG seizure data). In fact, the returns found using the small eigenvalues were more marked than those of the largest eigenvalue, perhaps suggesting that the small eigenvalues changes are a better indicator of changes in the synchronisation structure between equities.

6.4 Conclusions

The correlation structure between medium-frequency multivariate financial time-series was studied by investigation of the eigenvalue spectrum associated with the equal-time cross-

correlation matrix. Using sliding windows to filter the correlation matrix, we have examined behaviour of the largest eigenvalue over time. The largest eigenvalue is shown to move counter to that of a band of small eigenvalues, due to *eigenvalue repulsion*. A decrease in the largest eigenvalue, with a corresponding increase in the small eigenvalues, corresponds to a redistribution of the correlation structure across more dimensions of the vector space spanned by the correlation matrix. Hence, additional eigenvalues are needed to explain the correlation structure in the data. Conversely, when the correlation structure is dominated by a smaller number of factors (eg. the "single-factor model" of equity returns), the number of eigenvalues needed to describe the correlation structure in the data is reduced. Building upon previous work, [Drozdz *et al.*, 2000, 2001a], this means that fewer eigenvalues are needed to describe the correlation structure of 'drawdowns' than that of 'draw-ups'.

By introducing a simple 'one-factor model' of the cross-correlation in the system, we were able to reproduce the main results of our empirical study. The compensatory dynamics, described, were clearly seen for a correlation matrix with all elements equal to the average of the empirical correlation matrix. The one-factor model was then adapted, by the addition of perturbations to the correlations, with the average correlation remaining unchanged. This 'markets plus sectors' type model was able to reproduce additional features of the empirical correlation matrix with eigenvalues deviating from below and above the bulk. The Inverse Participation Ratio of the "markets plus sectors" model was also shown to have group characteristics typically associated with known Industrial Sectors, with a larger than average value for the smallest eigenvalue and for the second largest eigenvalue. Through a partition of the time-normalised eigenvalues, it was then shown quantitatively that the largest eigenvalue is greatest/smallest during drawdowns/drawups, and vice versa for the small eigenvalues, *demonstrating the potential of the correlation matrix in risk characterisation*. Further, the increase in the largest eigenvalue and corresponding increase in correlation, demonstrates the *universal collective behaviour of traders during drawdowns*. When the market is increasing, traders tend to pursue less universal strategies, exemplified by the reduction in correlation explained by the largest eigenvalue.

CHAPTER 7

MULTISCALED CROSS-CORRELATION DYNAMICS IN FINANCIAL TIME-SERIES

In previous chapters, we applied the cross-correlation matrix, calculated using low and medium frequency financial data, both for risk diversification and characterisation purposes. In Chapter 6, we demonstrated how the correlation dynamics between stocks could be detected by analysing both the largest and small eigenvalues over time. Using this, we then demonstrated how the correlation matrix could be used in risk characterisation, by showing that the largest eigenvalue peaks during periods of market distress, (drawdowns), with the smallest eigenvalues peaking during boom periods (draw-ups), [Conlon *et al.*, 2009]. In this Chapter, drawing on some of our previous discussed methods and results, (Chapters 5 & 6), we use the Maximum Overlap Discrete Wavelet Transform (MODWT), to calculate correlation matrices over different time scales for both medium and high-frequency data and then explore the *eigenvalue spectrum* over sliding time windows. The dynamics of the eigenvalue spectrum at different scales provides insight into the time horizons of the numerous constituents involved in the interactions and this is important because, by analysing relative changes across time-scales we obtain an insight into the 'herd' behaviour, (and associated shortening of time horizons), especially during extreme financial events.

The entire eigenspectrum is intrinsically important in terms of understanding the role of the correlation matrix in risk characterisation. On partitioning the eigenvalue time-series, we investigate the behaviour of the largest eigenvalue, measured at different scales, during drawdowns. Further, we examine the effect of different market types on the dynamics of

the small eigenvalues, to determine whether correlation changes at a particular scale might be visible at this end of the spectrum.

7.1 Data

The first of our two data sets comprises the 49 equities of the Dow Jones Euro Stoxx 50 (SX5E) where full price data is available from May 1999 to August 2007, resulting in 2183 daily returns. The Euro Stoxx 50 is a market capitalisation weighted index of 50 European blue-chip stocks from countries participating in the European monetary union. This index was chosen for this analysis as the number of stocks is small, allowing us to calculate cross-correlation matrices for small time windows, without reducing the rank of the matrix.

The second data set studied involved high-frequency equity returns and comprised all 50 equities of the SX5E from $8am$ on April 2^{nd} 2008 to $4.30pm$ on October 20^{th} 2008, with each day made up of approximately 450 minutes of data, Fig. 7.1. All times are in terms of GMT and equities from outside of this time-zone were aligned to coincide. The total dataset consisted of 51376 one-minute returns and were of particular interest with respect to the insight offered into the current extreme market events.

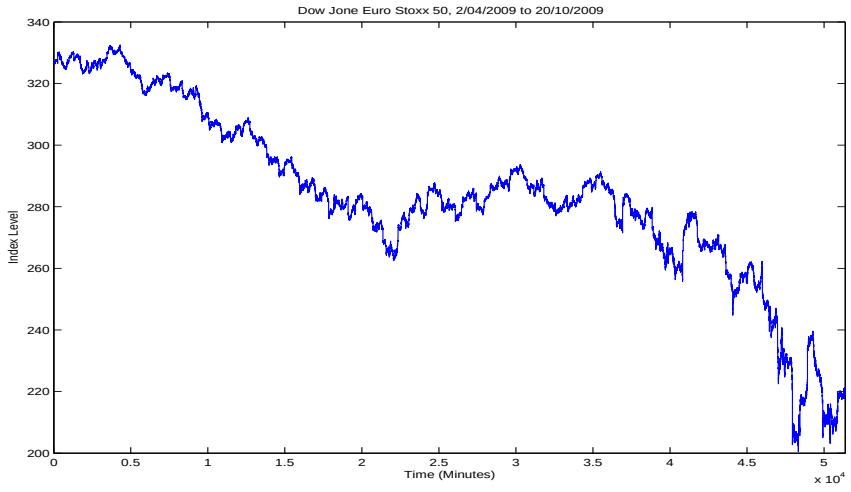


Fig. 7.1: DJ Euro Stoxx Index, April 2^{nd} 2008 to October 20^{th} 2008

7.2 Methods

As in Chapter 5, we used the asymmetric (LA) wavelet (known as the Symlet) which exhibits near symmetry about the filter midpoint. The MODWT was again implemented using the *WMTSA Wavelet Toolkit* for Matlab, [Percival and Walden, 2000]. The filter width chosen for this study was the LA4, (where the 4 refers to the width of the scaling function), since it enabled us to accurately calculate wavelet Betas and Correlations to the 3^{rd} scale given the length of data available. Although the MODWT, Section 3.6, can accommodate any level J_0, in practise the largest level is chosen so as to prevent decomposition at scales longer than the total length of the data series, hence the choice of the 4^{th} scale.

For the medium frequency data, the eigenvalue spectrum was found, using a sliding window approach with a window of 100 days. This window was chosen such that $Q = \frac{T}{N} = 2.04$, thus ensuring that the data would be close to stationary in each sliding window (Different values of Q were examined, with little variation in the results).

Using the high-frequency SX5E returns, the correlation matrix and associated eigenspectrum dynamics were found using a moving window of 1500 minutes. This time-window was chosen in order to reduce the effects of non-stationarities due to non-continuous time, by including market changes over 3 days. As markets are only open for 8.5 hours per day, new information can result in large jumps at market open. In fact, for smaller window choices, we found that the correlation dynamics were dominated by interday 'jumps'. The choice of $T = 1500$ resulted in $Q = \frac{T}{N} = 30$, resulting in an interesting comparative to the range of $1.3 \leq Q \leq 10$ studied previously.

The eigenvalue normalisation was carried out as for Chapter 6, expressed in terms of standard deviation units (SDU).

7.3 Results

7.3.1 Medium Frequency Eigenvalue Dynamics

We analysed the dynamics of the cross-correlations between stocks, calculated at different time scales, through the use of the eigenvalue spectrum. We first considered, Fig. 7.2, the dynamics of the largest eigenvalue at each time scale, calculated using a sliding time window of 100 days. Fig. 7.2(a) shows the value of the DJ Euro Stoxx 50 Index over the period studied. Fig. 7.2(b) displays the largest eigenvalue, calculated using the unfiltered (one day) time-series data for different time windows. As shown, the largest eigenvalue was far from static, rising from a minimum of 7.5 to a maximum of 30.5 from early 2001 to late 2003, (coinciding with the bursting of the "tech" bubble). This corresponds to an increase in the influence of the "Market", with the behaviour of traders becoming more correlated. The next major increase occurred in early 2006, followed by a relatively quick decline until the beginning of the "Credit Crunch" in 2007. Similar to Drozdz *et al.* [2000], we note an increase in the value of the largest eigenvalue during times of market stress, with lower values during more "normal" periods.

We next calculated, using the MODWT (Section 3.6), the value of the largest eigenvalue of the cross-correlation matrix over longer time horizons of 3, 6 and 11 days (Fig. 7.2(c-e)). The rational is that certain traders, (such as Hedge Fund managers), may have very short trading horizons while others, (such as Pension Fund managers), have much longer horizons. In looking at the value of the largest eigenvalue at different scales, we are attempting, for the first time, to characterise the impact of these different trading horizons on the cross-correlation dynamics between large capitalisation stocks. In Fig. 7.2(c-e), we see that the main features found in the unfiltered data were preserved over longer time scales. However, certain features, such as the sizeable drop in the largest eigenvalue at the longest scale, (Fig. 7.2(e)), in late 2003, were not seen at shorter scales. The aggregate impact for the unfiltered data is, nevertheless, a moderate drop. Other features, such as the increase in 2006, were not preserved across all scales. The different features, found at various scales, suggest that the *correlation matrix is made up of interactions between stocks, traded by*

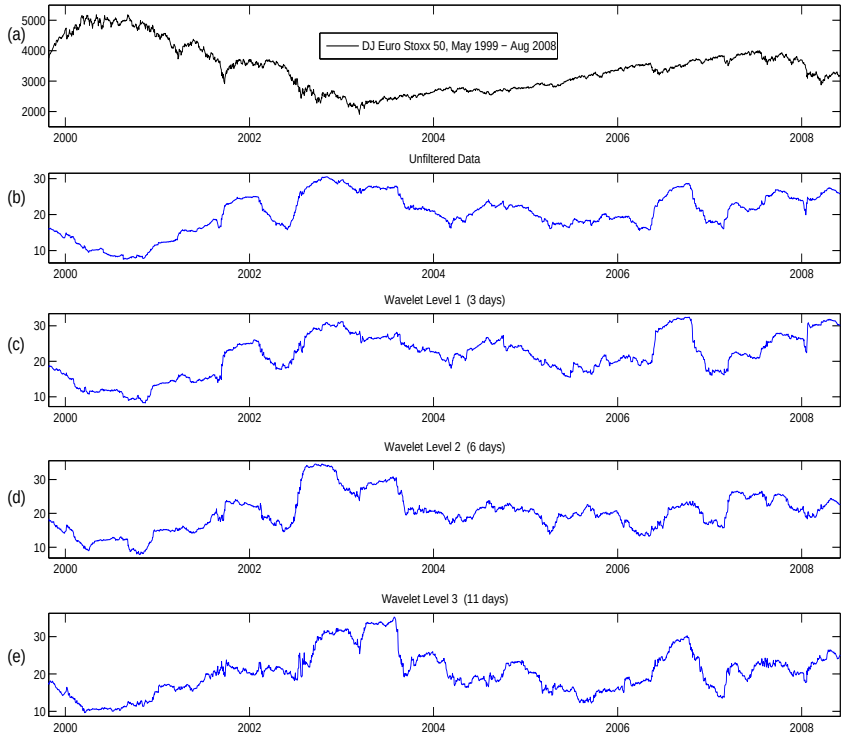


Fig. 7.2: (a) DJ Euro Stoxx 50 Index (b) Largest eigenvalue dynamics original data (c) 3 day scale (d) 6 day scale (e) 11 day scale

investors with different time horizons. This has implications for risk management, as the correlation matrix used for input in a portfolio optimisation should be appropriate to the investor's time horizon. By first decomposing the time-series, we have developed a novel method to consider these interactions in both time and scale.

The repulsion, (between eigenstates of the cross-correlation matrix as the level of synchronisation between time-series increases), has been demonstrated previously for artificial and EEG time-series, [Müller *et al.*, 2005], and by us for financial data, [Conlon *et al.*, 2009] and in the previous Chapter. In Fig. 7.3, we compare the normalised largest eigenvalue for sliding windows of 100 days with the average of the normalised 40 smallest eigenvalues over different time scales. The normalisation was carried out as previously, to allow comparison of eigenvalues at both ends of the spectrum. The normalised largest eigenvalue

preserved features, such as the increase from early 2001 to late 2003, previously identified for the largest eigenvalue, with eigenvalue repulsion also demonstrated. The large/small eigenvalues were shown to increase/decrease respectively from 2001 to 2003 across all scales, demonstrating that changes in the correlation structure can also be detected in the small eigenvalues.

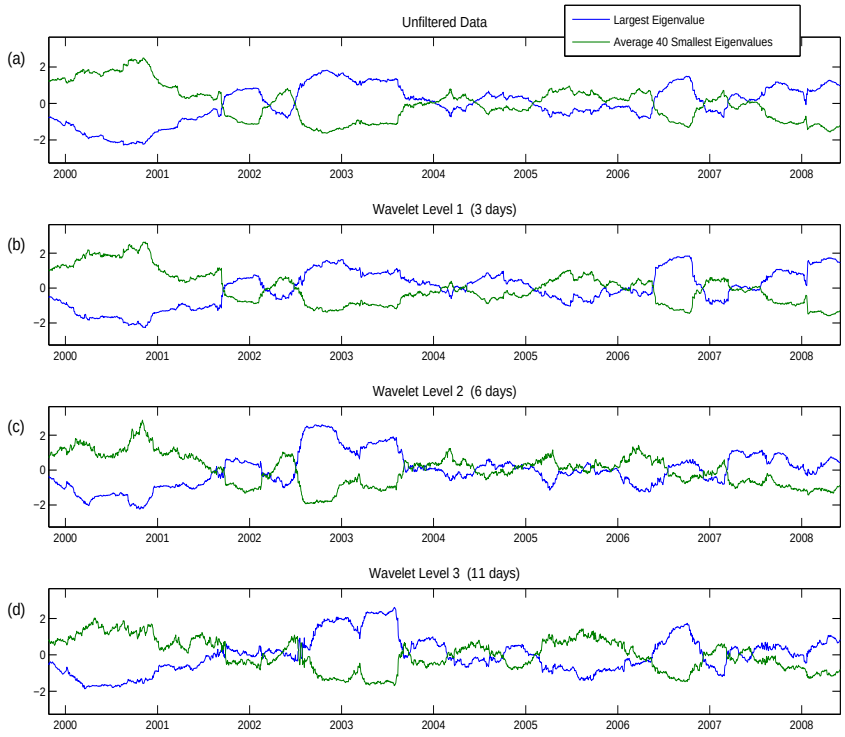


Fig. 7.3: (a) DJ Euro Stoxx 50 Index (b) Eigenvalue Dynamics at 3 day scale (c) Eigenvalue Dynamics at 6 day scale (d) Eigenvalue Dynamics at 11 day scale

Eigenvalue Ratios

Redistribution of the correlation structure across the eigenvalue spectrum was also captured using *ratios of eigenvalues*. Fig. 7.4 shows the ratio of the largest eigenvalue to the second largest eigenvalue, (again calculated using sliding windows of 100 days), for the original unfiltered data and for correlation matrices calculated with data corresponding to 3, 6 and

11 days. For unfiltered data, the ratio was found to gradually increase for the period from 2000 to 2003, corresponding to greater importance of the largest eigenvalue, followed by a large decrease, corresponding to the bursting of the 'tech bubble' followed by the relatively sanguine period that followed. At higher scales (6 and 11 days), the ratio was found to increase more abruptly from mid 2002 implying that long and short timescales capture different features of the correlation structure. Importantly, this behaviour may be a flag of structural correlation changes that could act as a barometer of a change in investor perception. However, this interpretation can only be tentative due to the relatively small amount of data studied and the analysis of high-frequency data might reveal more.

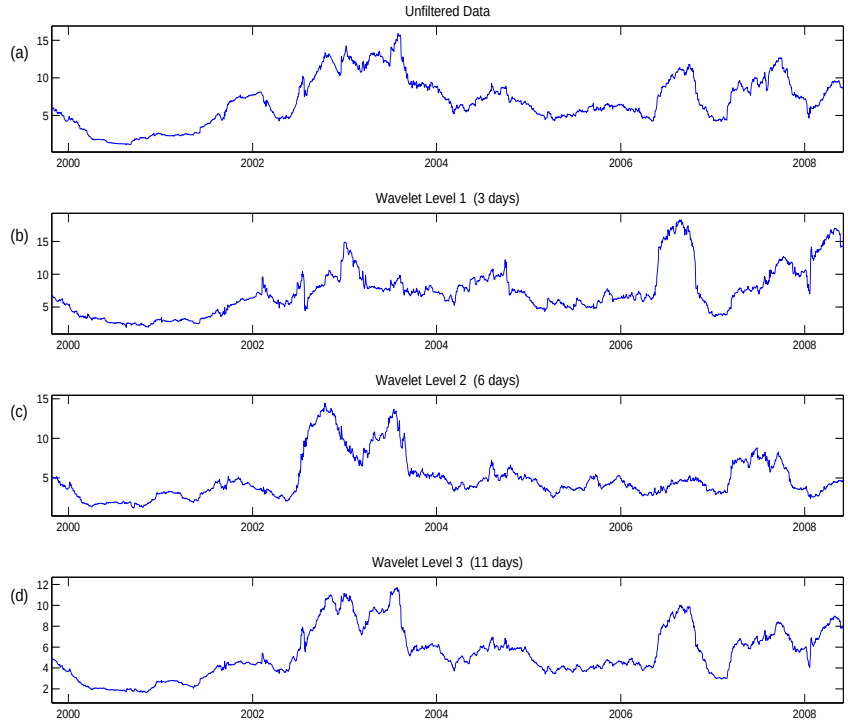


Fig. 7.4: Ratio largest eigenvalue to 2^{nd} largest eigenvalue for (a) original data (b) 3 day scale (c) 6 day scale (d) 11 day scale

The distinct features and abrupt changes visible in the dynamics of eigenvalue ratios, can be explained by the variation in the 2^{nd} largest eigenvalue over time. In the context

of the analysis that we performed using deviating eigenvalues from RMT, the information contained in the second largest eigenvalue may correspond to distinctly different *sectors* at the various scales. This has important implications for portfolio optimisation, where correlation between assets is a fundamental input, Section 3.3. If the average correlation between a given asset and others in a portfolio is raised at longer scales, then this asset may not be such a good diversifier and, in fact, may increase the risk of the portfolio. Examples of this were seen in Chapter 5, where the correlation of strategies such as Equity Long/Short increased over longer scales.

7.3.2 High Frequency Eigenvalue Dynamics

The eigenvalue dynamics, found using medium-frequency data, demonstrated that additional features of the correlation between equities are revealed as the time-scale used in the calculation is adjusted. High-frequency eigenvalue dynamics for the largest normalised eigenvalue, (normalised over the entire period), and the average of the normalised 40 smallest eigenvalues are shown in Fig. 7.5. The dynamics using the original data are given, Fig 7.5(a), while Fig 7.5(b-d) were constructed using longer scales. It is evident that additional features emerge for longer scales, with much larger variation at the 2807 minute or ≈ 6 day scale.

Of particular interest is the small disparity between the original data and that calculated using the 7 minute scale. In fact, this apparent disparity occurred for a variety of scales, up to daily. Beyond this daily scale, distinct dynamics emerged, Fig 7.5(c-d), highlighting the changes in correlation structure that occur at longer time-frames. This characteristic time-period for settled pattern emergence (approximately 1 day), may however, be attributable to the discontinuities in the data. As evident in Fig 7.5(a-b), changes in the eigenvalue structure tends to occur in 'jumps', (discontinuites in trading time are one factor). However, at scales corresponding to 3 and 6 days, Fig 7.5(c-d), the jumps are less evident as the discontinuities are smoothed over the extended period.

Looking more closely at the results, we found that the discontinuities at low scales

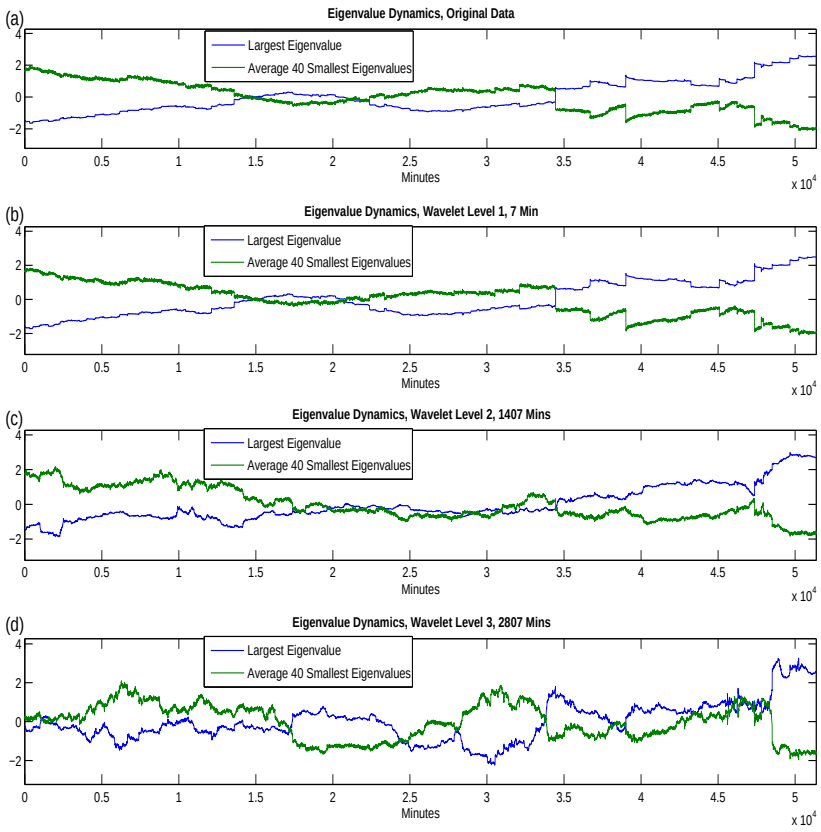


Fig. 7.5: Largest eigenvalue for (a) original data (b) 7 min scale (c) ≈ 3 day scale (d) ≈ 6 day scale, calculated using high-frequency Euro Stoxx 50 data, April 2^{nd} to October 20^{th} 2008

corresponded exactly with jumps in the market, such at those that occur when markets open or when news is released. At longer scales, the relationship was less obvious. However, the distinct rise in the largest eigenvalue at longer scales, corresponded to a large increase in correlation as the market dropped consistently, (see Fig. 7.1). At scale 3, (6 days), the sharp upward movement near the end lagged behind a similar rise at scale 2, (3 days), but was more abrupt. In practical terms traders are known to choose positions based on their trading strategies, acceptable risks and market conditions. The differences in eigenvalue dynamics evident at these scales may be the result of the competing time horizons of different agents involved in the market.

7.3.3 Drawdown analysis

As indicated earlier, drawdowns, (or periods of large negative returns), and drawups, (periods of large positive returns), tend to be accompanied by an increase in different eigenstates of the cross-correlation matrix. Similar to the analysis of Chapter 6, we again look for evidence of market behaviour in terms of fluctuations in the eigenspectrum. The average return of the index is shown in Table 7.1, (during periods when the largest eigenvalue is ± 1 Standard Deviation Unit (SDU)), for both the largest eigenvalue and the average of the normalised 40 smallest eigenvalues.

Eigenvalues	*No. Std*	*Unfiltered*	*1*	*2*	*3*
Large	$<$-1	6.4%	2.5%	8.5%	10.4%
	$>$1	-9%	-10%	-9.3%	-7.6%
Average 40 Smallest	$<$-1	-9.6%	-9.8%	-12%	-8.4%
	$>$1	10%	9.6%	8.2%	8%
Ratio 1	$<$-1	1.5%	0.4%	7.7%	6.4%
	$>$1	-6.2%	-7.2%	-9.0%	-7.2%
Ratio 2	$<$-1	3.2%	3.5%	-1%	6.9%
	$>$1	-5.6%	-4.7%	-5.9%	-3.8%

Tab. 7.1: Drawdown/Drawup analysis. Average Index Returns when various eigenvalue partitions in SDU are > 1 and < -1. Ratio 1 is that of the largest eigenvalue to the 2^{nd} largest, while Ratio 2 is the largest eigenvalue to the sum of the 40 smallest eigenvalues.

Looking first at the original or unfiltered data, we found that when the largest eigenvalue is > 1 SDU, the average index return was found to be -9%. In contrast, when it was < -1 SDU, the average index return over each of the time windows was 6.4%. This allowed us to characterise the market into drawdowns and drawups by examining the relative size of the eigenvalues. For the average of the 40 smallest eigenvalues, the partition reflected more marked drawdowns and drawups but with opposite signs. This indicates that information on the interaction between traders, captured in the correlation dynamics, is visible in both the lower and upper eigenstates.

Table 7.1 also shows the average index returns for partitioned eigenvalues, calculated using data at the longer scales (3, 6 and 11 days). The results show that characterisation of the largest eigenvalue, (well above average during drawdowns and well below during

drawups) was *consistent across scales*. The opposite was found to occur for the small eigenvalues. The implication is that the *correlation structure between stocks is less dependent on the trader's time horizon than on the current state of the market*. This is in contrast to the results found at high-frequency, where eigenvalue dynamics varied markedly across scales. However, the high-frequency sample was from a unusual period of rapid market decrease and may not have captured the full agent behaviour.

In Table 7.1, we also present the ratio of the largest eigenvalue to the second largest, (normalised over time as above). For the unfiltered data, the ratio was found to be largest during periods of negative returns and smallest during periods of positive return, again implying that the *market effect dominates during drawdowns*. This was seen across all scales with larger absolute values of the returns found at larger scales, implying that correlations, calculated using longer time scales, may be more sensitive to market behaviour. The ratio of the largest eigenvalue to the sum of the forty smallest eigenvalues is also shown in Table 7.1. This ratio was less sensitive than that shown in Figure 7.4, with values found to vary from 0.22 to 2.02. However, there is still evidence that negative returns occur during periods when the largest eigenvalue is much greater than the sum of the small eigenvalues. The opposite, though less marked for positive returns, is also indicated.

7.4 Conclusions

The multiscale correlation structure of both medium and high-frequency multivariate financial time-series was studied by investigation of the eigenvalue spectrum of the equal-time cross-correlation matrix. This analysis revealed some of the risk characteristics, associated with the correlation structure between stocks, and shed some light on the various time horizons involved in the underlying interactions. Observations may be summarised as:

1. Using the MODWT and a sliding window, the dynamics of the largest eigenvalue of the correlation matrix were examined and shown to be time-dependent (Fig. 7.2). Similar dynamics were visible across all scales, but with particular features, such as the sizeable drop in the largest eigenvalue in 2003, markedly apparent at certain

scales. This suggests that the correlation matrix between stocks comprises interactions between traders with different time horizons.

2. Large and small eigenvalues demonstrated the expected repulsion across all scales (Fig. 7.3). It is possible to characterise the correlation structure of the system with just the small eigenvalues, because of this repulsion.

3. An initial study of high-frequency data, revealed little change in the eigenvalue dynamics over shorter scales, with additional dynamical features emerging at a characteristic time of approximately one day. This characteristic time may be due to the discontinuities associated with 'trading hours'. A more detailed study is needed to confirm this behaviour.

4. A partition of the time-normalised eigenvalues demonstrated quantitatively that the largest eigenvalue is greatest/smallest during drawdowns/drawups. *Small eigenvalues* were, in general, found to be *more sensitive than large eigenvalues to behaviour of the market*, with this persisting across all scales. This suggests that the underlying state of the market is more important to the correlation structure than the time horizons of different traders.

CHAPTER 8

MULTISCALE EEG CORRELATION DYNAMICS

Risk management is not unique to financial markets. In fact, it is intrinsic to a wide range of fields including engineering, industrial processes and medicine. In particular, the monitoring and control of certain medical and neurophysiological conditions is dependent upon 'early warning' systems, often employing signal processing techniques. In previous Chapters, we demonstrated the application of methods based on the correlation matrix for risk reduction and characterisation in complex financial systems, [Conlon *et al.*, 2007, 2008, 2009]. We showed how a correlation matrix, cleaned according to Random Matrix Theory principles, could be used to reduce realised risk in a portfolio optimisation. The scaling characteristics of low-frequency Hedge Fund correlations were then explored with applications in risk management for a Portfolio of Hedge Funds. Using medium frequency financial returns, we then showed how the correlation matrix and associated eigenspectrum could be used in the characterisation of markets according to their risk. By the application of wavelet multiscaling techniques we investigated the scaling effects on this risk characterisation. In some instances, early targets for the ideas explained, particularly for risk characterisation, were Neurological systems and, in this Chapter, we explore this comparative example. The aim is to build on earlier analysis of these systems, using the scale dependent correlation analysis we developed above to characterise the risk of epileptic seizures. This application serves also to demonstrate the inter-disciplinary nature of Complex Systems research and emphasises the value of the cross-correlation matrix formalism in areas outside

of finance. It also raises interesting questions on which commonality of features is essential for informative applicability of techniques across disciplines.

The characterisation of epileptic seizures has the potential to assist in the design of targeted pharmaceutical seizure prevention techniques and pre-surgical evaluations. In this Chapter, we use the cross-correlation dynamics between electroencephalographic (EEG) channels in order to develop an early warning system for epileptic seizures. The Maximum Overlap Discrete Wavelet Transform (MODWT) is applied in order to separate the EEG channels into their underlying frequencies. The dynamics of the cross-correlation matrix between channels, at each frequency, are then analysed in terms of the eigenspectrum. By examination of changes in the eigenspectrum structure, we investigate the possibility of identifying frequency-dependent changes in the correlation structure between channels, which may be indicative of Seizure activity. Further, the energy associated with each wavelet scale is examined, to help distinguish the scales important during seizure activity. Additional features are highlighted and results discussed in the comparative context.

8.1 Background

Seizure occurrence is associated with changes in the level of interaction between neurons (agents). By reducing or countering the level of interaction between neurons, it may be possible to reduce the risk of an epileptic seizure occurring. Neurophysiological systems have many features in common with other Complex Systems, such as interacting, dynamic agents with heterogeneous characteristics. In previous chapters, we have applied the linear correlation measure to examine the synchronisation between different agents within a complex dynamic financial system. The primary risk in the case of seizures is clearly very different from our previous examples in finance. However, a financial crash can also be an equally catastrophic outcome on a personal level. The study of comparative systems like these, allow us to identify features common to both systems. The emergence of common features, such as correlation changes prior to seizures and crashes, helps us to identify techniques previously applied in one field, which may prove fruitful in another.

8.1.1 Data

In this study, EEG signals from eight patients of different ages ($28 - 78$), determined by a neurologist to be suffering from focal[1], generalised[2] or secondary generalised[3] seizures are examined. The data was obtained from the Australian EEG database, [Hunger *et al.*, 2005], and all signals were recorded using standard international system $10 - 20$ electrode placements. These electrodes are placed so as to provide for a repeatable, uniform coverage of the entire scalp, [Fisch and Spehlmann, 1999].

The database contains patient details including medical history, technician's comments, medication and a full EEG report by a neurologist. The clinical interpretation from the neurologist was used to classify the seizures. In this analysis bipolar derivations, (referred to as a channel from here on), between nearest-neighbour channels were used, [Fisch and Spehlmann, 1999], as the use of monopolar signals, (where all electrodes have a common reference), could result in the introduction of unwanted correlations into the system.

One problem with securing financial data is, of course, their confidential nature, which results in little public availability of information (in particular, for Hedge Funds). Securing suitable seizure data also proved to be difficult, due to their similarly confidential nature. A number of databases were explored, included the University of Freiburg EEG database, [University of Freiburg, 2008], but these proved to be more appropriate for univariate prediction methods. Availability of a more extensive database, containing additional examples of various forms of epileptic seizures, would have further enhanced the results detailed below.

[1] A partial or focal onset seizure affects only a part of the brain at onset. They may often be a precursor to a larger seizure, such as a generalised seizure

[2] Generalised seizures affect both cerebral hemispheres (sides of the brain) from the beginning of the seizure and produce loss of consciousness. They are divided into several sub-types: generalised tonic-clonic, myoclonic, absense and atonic

[3] Secondary generalised seizures start as a partial seizure and spread throughout the brain, becoming generalised

8.1.2 Implementation

Defining Seizure Onsets and Endings

In order to identify seizure activity in a clear and reproducible fashion, we use a technique previously described, [Schindler *et al.*, 2007a], as follows: Given time-series, $S_i(t)$, corresponding to readings from an EEG, the absolute slope, $S_i(t) = \left|\frac{\Delta EEG_i(t)}{\Delta t}\right|$, was computed over each channel i. The slope was normalised, $\tilde{S}_i(t) = \frac{S_i(t)}{\sigma_i}$, with σ_i, the standard deviation of the signal over a reference period containing no seizure activity. $\tilde{S}_i(t)$ was then smoothed using a moving average over a time window of 5 seconds. *Epiletiform activity* or periods of extreme neuronal activity, were then defined by finding a slope greater then 2.5 standard deviations from the mean. The time of seizure onset was determined from the number of EEG channels displaying epileptiform activity. For the following analysis, this was set at 5, ensuring that the seizure had spread to a minimum number of channels and was not specific to one region, allowing us to isolate those periods corresponding to epileptic events.

Specific Methods

To decompose the EEG signal into component frequencies, we selected the least asymmetric (LA) wavelet, (known as the Symmlet, [Bruce and Gao, 1996]), which exhibits near symmetry about the filter midpoint. These are defined in even widths and the optimal filter width is dependent on the characteristics of the signal and the length of the data series. Here, the filter width chosen was the LA8 (where the 8 refers to the width of the scaling function) and this enabled us to calculate accurately wavelet correlations to the 5^{th} scale, given the length of data series available, while still containing enough detail to capture subtle changes in the signal. Although the MODWT can accommodate any level, J_0, in practise the largest level is chosen so as to prevent decomposition at scales longer than the total length of the data series, hence the choice of the 5^{th} scale, [Percival and Walden, 2000].

The equal-time cross-correlation dynamics between EEG channels using a sliding win-

dow of length $5s$, (chosen so that the signal would be close to stationary during each window), was analysed. First, the MODWT of each EEG channel was calculated within each window and the correlation matrix between channels at each scale found, (Eqn. 3.32). The eigenvalues of the correlation matrix in each window were determined, (Eqn. 3.4), and eigenvalue time-series were normalised in time, (Eqn. 3.19).

8.2 Results

8.2.1 Single Patient Analysis

The seizure definition, (Section 8.1.2), and eigenvalue dynamics for a 30 year old patient suffering from focal epilepsy with possible secondary generalisation are shown, Fig. 8.1(a). This reveals 3 main periods that display epileptiform activity. Initial examination of the eigenvalue dynamics for each of the wavelet scales (corresponding to frequencies from 4Hz to 60Hz), reveals that the *eigenvalue repulsion* found using the equal-time cross-correlation matrix on unfiltered data, [Schindler *et al.*, 2007b,a], is also repeated across the different frequencies. Fig. 8.1(b-f) show the dynamics of both the largest eigenvalue and that of the average across the 15 smallest eigenvalues. The dynamics of the smallest eigenvalues complement, across all scales, those of the largest, (given the trace of the correlation matrix must remain constant under linear transformation), with increases in the latter and decreases in the former, when average correlation increases, [Müller *et al.*, 2005, 2006a; Schindler *et al.*, 2007a,b].

In Fig. 8.1, we see that the largest eigenvalue of the cross-correlation matrix, calculated at the highest frequency ($60Hz$), increases during epileptiform activity. However, as we move to lower frequencies, the largest eigenvalue tends to decrease during epileptiform activity. In the example studied, this is particularly evident at levels 3 and 4 (corresponding to $15Hz$ and $7Hz$ respectively). The increase in the largest eigenvalue at the highest frequency corresponds to an increase in the average, or global, system correlation at this frequency, with the opposite occurring at lower frequencies, (Section 3.5).

The wavelet energy, (Eqn. 3.29), measured in a sliding window of $5s$ for each of the

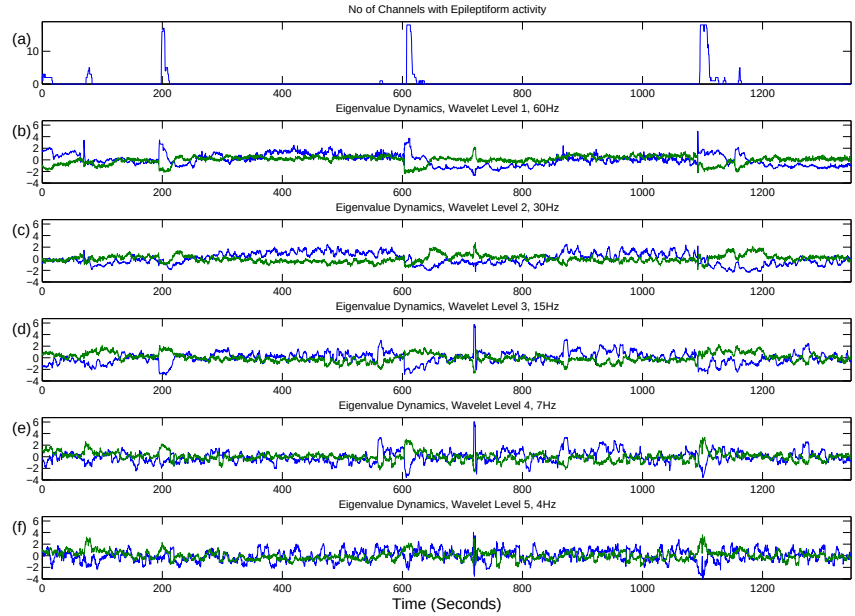


Fig. 8.1: Patient 1, (a) Seizure definition (b-f) eigenvalue dynamics across scales. Blue lines indicate largest eigenvalue dynamics, green l average of 15 small normalised eigenvalues

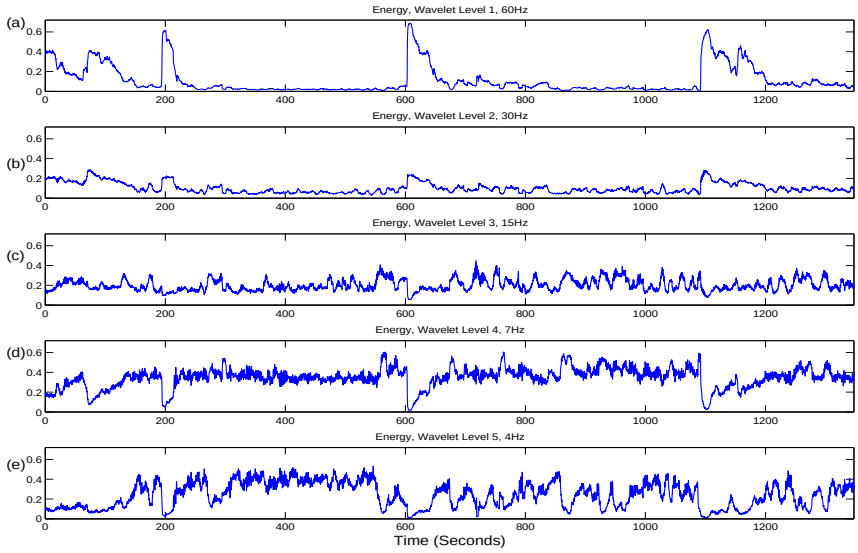


Fig. 8.2: Patient 1, (a-e) Fractional wavelet energy associated with each of the first 5 wavelet scales, measured using a sliding window of 5s .

wavelet scales, is shown in Fig. 8.2 for the same patient as previously, (Fig. 8.1). The energy at the highest frequencies is negligible, *except during periods corresponding to epileptiform activity*, when it increases greatly, corresponding to more than 60% of the total system energy. During non-epileptiform activity, the energy at low frequencies makes up most of the system energy. However, during epileptiform activity the energy at low frequencies drops to negligible levels compensating for the increase at high frequencies implying that most of the system energy is involved in high frequency events such as spikes.

These preliminary results seem to indicate increased levels of correlation between EEG channels at the highest frequencies during epileptiform activity, with corresponding increases in energy. In contrast, the average system correlation at low frequencies, (measured against the dynamical behaviour of the largest eigenvalue), decreases with corresponding decrease in energy. Since the associated energy is higher, high frequencies seem of more relative importance during epileptiform activity, so that the correlation structure at these frequencies may be of more relevance in seizure characterisation.

8.2.2 Multiple Patient Analysis

In order to investigate further questions posed by these initial results, we examined the eigenvalue dynamics and associated energy for eight patients, (described, Section 8.1.1); individual results given Table 8.1. Average eigenvalue size and associated energy during active periods[4] and during normal periods[5] are presented. The average eigenvalue size and energy are measured across scales $1-5$, as before. The results for Patient 1 are also shown, (Fig. 8.1).

Fig. 8.3 shows the distribution of the largest eigenvalue for *all patients* across each scale. For the highest frequencies, the average eigenvalue is considerably higher during active, compared to normal periods. Using a Wilcoxon signed-rank test across all samples, the probability of the median difference between eigenvalue pairs being zero is less than

[4]Periods when the largest eigenvalue, at the highest frequency, is greater than 1.5 standard deviations units (SDU) from the mean. This means that we look at the 6.7% largest readings.

[5]Periods when the largest eigenvalue, at the highest frequency, is between -1.5 and 1.5 SDU from the mean.

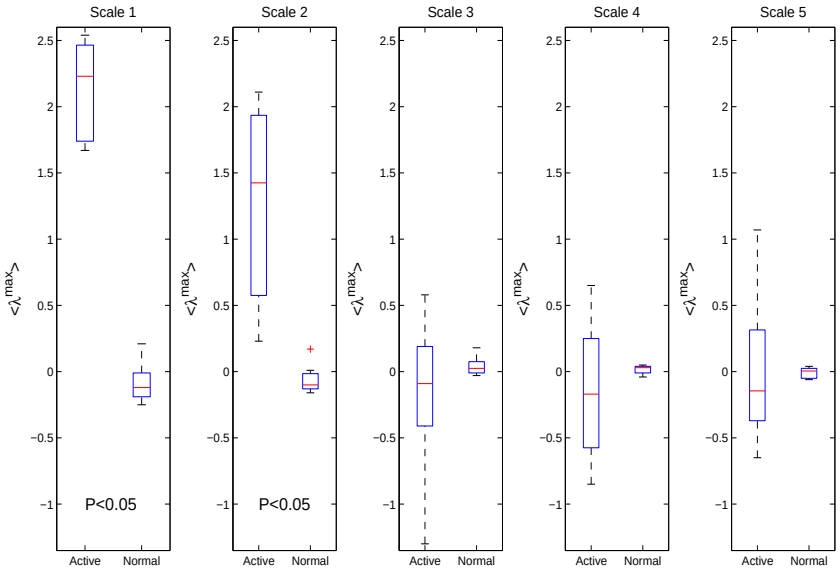


Fig. 8.3: Largest eigenvalue for all patients across first 5 scales. The red line shows the median value, while the quartiles are shown in blue and outliers as 'whiskers'. The **P** value is the Wilcoxon signed-rank probability (see text).

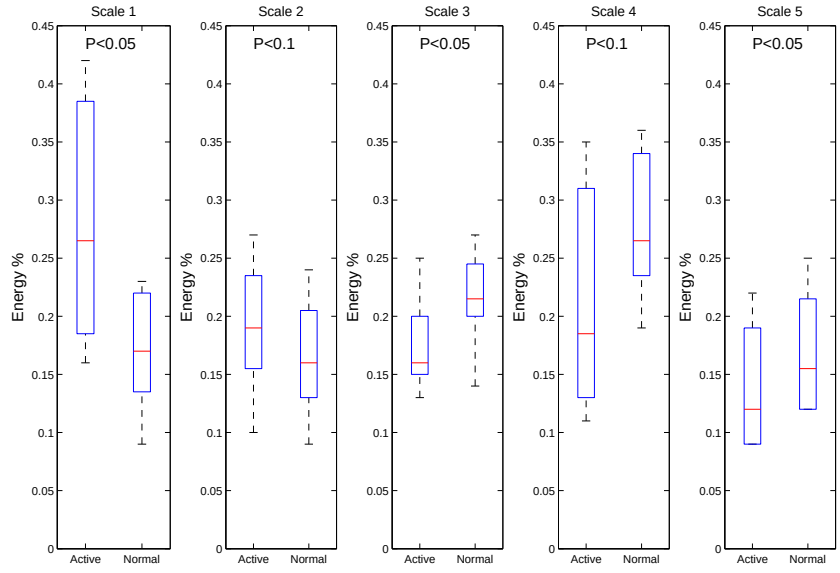


Fig. 8.4: % Energy for all patients across first 5 scales. The red line shows the median value, while the quartiles are shown in blue and outliers as 'whiskers'. The **P** value is the Wilcoxon signed-rank probability (see text).

0.05 for these scales. The increase in the largest eigenvalues corresponds to an increase in the global or average correlation at these high frequencies. For the lower frequencies, there is a definite overlap between the eigenvalues but with a lower median and much larger variance during the active period.

The energy at each of the scales, (shown Fig. 8.4), has an obvious increase at the highest frequency during active periods (Wilcoxon signed-rank test, Probability< 0.05 of distribution about same median). At the next frequency, the energy is also higher but not markedly so. For the lower frequencies, the energy decreases during active periods to compensate for the increase at higher frequencies. In particular, for scales 3 and 5, the changes are significant with P< 0.05, (Wilcoxon signed-rank test, Kanji [2006]). This behaviour indicates that the high frequency behaviour is of greatest importance during active periods, with corresponding correlation increase.

Examining the differing overlap between the two periods, (Figs. 8.3,8.4), the active period appears to have two major manifestations, namely a variance increase and a clear non-overlap for the two highest frequencies. Further, there is a large increase in variance of the energy across those same frequencies. This further suggests that the large changes in synchronisation at certain scales may act as a barometer of seizure activity.

8.3 Discussion

By analysis of the eigenvalue spectrum of EEG epileptiform signals, filtered using the wavelet transform, we were able to examine changes in the cross-correlation matrix between channels. The largest eigenvalue corresponds to the average correlation between all channels and is orthogonal to the other eigenvalues. Previously, a number of studies examined the *univariate* time-frequency behaviour using a number of different linear and non-linear techniques, [Clark *et al.*, 1995; Senhadji and Wendling, 2001; Adeli *et al.*, 2003; Indirdevi *et al.*, 2008]. However, this single channel approach ignores the interactions between neurons involved in brain activity, which are particularly prominent during seizures.

Various *bivariate methods* to examine dynamic changes in these interactions have been

Patient No.	*Patient Age*	*Neurologist Diagnosis*		*Scale 1*	*2*	*3*	*4*	*5*
1	30	Focal With Possible Sec. Generalisation	Scale 1 Active	2.20	0.23	-1.30	-0.71	-0.33
			Energy	0.38	0.18	0.15	0.17	0.12
			Scale 1 Normal	-0.12	0.01	0.09	0.05	0.03
			Energy	0.09	0.09	0.21	0.36	0.25
2	47	Benign Focal	Scale 1 Active	2.54	1.65	-0.56	0.13	0.52
			Energy	0.39	0.26	0.15	0.11	0.09
			Scale 1 Normal	-0.25	-0.16	0.06	-0.02	-0.05
			Energy	0.23	0.18	0.22	0.19	0.12
3	50	Focal	Scale 1 Active	2.50	2.09	0.58	0.37	1.07
			Energy	0.32	0.27	0.16	0.13	0.12
			Scale 1 Normal	-0.10	-0.08	0.00	0.00	-0.05
			Energy	0.23	0.23	0.19	0.23	0.12
4	48	Generalised	Scale 1 Active	1.72	1.78	-0.24	-0.44	-0.11
			Energy	0.42	0.20	0.16	0.13	0.09
			Scale 1 Normal	-0.12	-0.12	0.02	0.03	0.01
			Energy	0.21	0.15	0.26	0.24	0.14
5	28	Generalised	Scale 1 Active	2.26	1.20	0.18	-0.31	-0.41
			Energy	0.16	0.16	0.24	0.27	0.17
			Scale 1 Normal	-0.24	-0.13	-0.02	0.03	0.04
			Energy	0.17	0.17	0.23	0.26	0.17
6	78	Generalised	Scale 1 Active	2.43	2.11	-0.26	0.65	0.11
			Energy	0.21	0.21	0.16	0.20	0.22
			Scale 1 Normal	-0.14	-0.13	0.03	-0.04	0.00
			Energy	0.12	0.18	0.21	0.27	0.22
7	48	Sec. Generalised	Scale 1 Active	1.67	0.75	0.06	-0.03	-0.18
			Energy	0.16	0.15	0.25	0.35	0.09
			Scale 1 Normal	0.21	0.17	0.13	0.04	-0.06
			Energy	0.15	0.14	0.27	0.32	0.12
8	52	Sec. Generalised	Scale 1 Active	1.76	0.40	0.20	-0.85	-0.65
			Energy	0.21	0.10	0.13	0.35	0.21
			Scale 1 Normal	0.08	-0.04	-0.03	0.04	0.02
			Energy	0.17	0.12	0.14	0.36	0.21

Tab. 8.1: EEG Analysis; For each patient the first row shows average size of eigenvalues when largest normalised Eigenvalue at Scale $1 > 1.5$, while row two shows the associated energy. The third row shows average eigenvalue at each level during normal behaviour (ie. when largest eigenvalue, $-1.5 < E < 1.5$) and the fourth row shows the associated Energy.

suggested. The linear cross-correlation was measured at different frequencies, for the channels with the maximum power at high frequencies, [Wendling *et al.*, 2003]. In contrast to the work above, the correlation between these channels was shown to decrease during the seizure. This, however, may be due to the bivariate approach used, which only captured the correlation between pairs of particular channels. This activity may possibly be captured in the second or third largest eigenvalue of the multivariate technique, where correlations orthogonal to those in the largest eigenvalue are found, (corresponding to correlations between certain subsystems, [Müller *et al.*, 2005]). The linear cross-correlation for two band-filtered channels were examined for different time lags, with a strong relationship found for a frequency band around 30Hz, [Ansari-Asl *et al.*, 2005].

The bivariate methods described concentrated on correlations between a small number of channels, chosen specifically for the studies involved. Using multivariate EEG data, changes in the global correlation structure were shown to be visible at both ends of the eigenvalue spectrum, [Müller *et al.*, 2005]. For a limited study of a single seizure, a sudden system-wide change from a relatively uncorrelated to highly correlated state was found to take place, reflected in an increase in the largest eigenvalue. Analysis of the changes in the eigenvalue spectrum for a large number of seizures showed a generic change in the correlation structure during focal onset seizures, [Schindler *et al.*, 2007a]. The seizure recordings in this data consisted of $58 - 94$ channels and the changes in the eigenvalue spectrum were shown to occur for a number of the largest eigenvalues. Contrary to the behaviour found previously, [Müller *et al.*, 2005], these large eigenvalues were shown to decrease during the first half of the seizure, indicating decreased correlation, with an increase in correlation found before seizure end. It was suggested that this increase in correlation may be related to seizure termination.

Time-Frequency decompositions of EEG signals have been studied for many years, [Nakata and Mukawa, 1989]. In this initial work, we extend the previous multivariate techniques, [Müller *et al.*, 2005; Schindler *et al.*, 2007a], by examining changes in the eigenvalues of the correlation matrix between EEG time-series across various frequencies. Results for low frequencies were similar to those found previously, (see Fig. 8.1), with

a decrease in global correlation at seizure onset, reflected by the decrease in the largest eigenvalue. Those, at higher frequencies (Fig. 8.1(b)), demonstrated an increase in the largest eigenvalue at seizure onset, corresponding to an increase in global correlation. This was confirmed across seizures, with just a small variation across seizures for the correlation at lower frequencies. The limited number of channels available for study, meant that the effects of the global or average correlation were only found in the largest eigenvalue, in contrast to the work by Schindler *et al.* [2007a].

Previous studies have used wavelet energy for prediction of seizures, [Gigola *et al.*, 2004]. We perused this further to examine the fractional energy of each of the frequencies and their importance over time. Again, the highest frequencies seemed of most relevance during active periods. This development of the previous work adds insight on contributing frequencies to correlation structure. By focusing on these particular frequency breakdowns, a simple multivariate technique such as that described may be applicable to seizure prevention. Additionally, as very high frequency oscillations have been found during seizures, [Bragin *et al.*, 1999], a study of correlation dynamics at this scale may reveal further insights.

8.4 Conclusions

Wavelet multiscaling has been used to expand on the previous multivariate analysis, [Müller *et al.*, 2005, 2006a; Schindler *et al.*, 2007a,b], to explore the frequency-dependence of the correlation dynamics between EEG channels for patients suffering with various forms of epilepsy. By analysis of the eigenvalue spectrum of EEG epileptiform signals, filtered using the wavelet transform, we were able to examine the correlation dynamics over different frequencies. Detailed analysis, for a patient with focal epilepsy indicated that EEG time-series reveal an increase in the largest eigenvalue at high frequencies, (corresponding to an increase in correlation between channels), during epileptiform activity. This increased interdependence between channels was not found at low frequencies, where correlation decreased during activity. Decomposition revealed an increase in the wavelet energy at higher

frequencies during epileptiform activity, with a corresponding decrease at lower frequencies. This implies that *high frequency activity is more significant during epileptiform activity* and so the correlation dynamics at these frequencies are of greater relative importance in terms of activity displayed. The correlation activity at lower frequencies was larger when abnormal activity was low, with higher levels of associated wavelet energy. This suggests that low frequencies are of greater relative importance during normal activity.

This approach was then applied to a variety of EEG signals, (Section 8.2.2), for different types of epileptic event. *Correlation dynamics were found to be dependent upon the frequency examined, with the correlation structure acting as a barometer of EEG activity*. Clearly, the data available are limited, but evidence of clear crossover in eigenvalue energy does suggest that monitoring correlation structure in EEG signals at different frequencies can provide a more subtle gauge of incipient imbalance at pre-seizure stage, than was found by previous researchers using unfiltered signals alone.

A comparison between these results and those found for medium and high-frequency financial data leads to some obvious differences. The discontinuities in financial data, particularly apparent for high-frequency returns, leads to 'jumps' in the correlation structure at very high-frequencies. For correlations between EEG channels, the readings have smoother dynamics. Also, the gradual correlation changes found in EEG signals around seizures are contrary to the sharp jumps that occur in financial data, particularly at shorter scales. These sharp jumps are due to the sudden release of information to the market, while for a complex neuronal system there is more gradual information dissemination. The more gradual release of information in EEG signals means that the potential ability to forecast changes in correlation is better. However, by applying the energy analysis described above to financial data we will be able to identify the scales important during different market events, which may assist in the prediction of market crashes. Some additional similarities between both data types exist, however, with definite changes in correlation in times of stress. For EEG data at high frequencies, correlations were shown to increase during seizures, while for financial data correlations increase during market 'draw-downs'.

CHAPTER 9

CONCLUSIONS AND FUTURE WORK

In this work, we have investigated the synchronisation between agents in Complex Dynamical Systems, using the linear cross-correlation measure. We demonstrated how this measure could be used in an innovative fashion, both in *risk reduction and characterisation* for such systems. These properties were studied extensively, with data from a number of systems examined, including novel low-frequency Hedge Fund returns, medium and high-frequency equity returns and high-frequency EEG seizure data.

In **Chapter 4**, we considered, for the first time, the effects of noise on low-frequency Hedge Fund returns. The novel application of Random Matrix Theory to reduce the level of noise in the correlation matrix, revealed new insight into the relationship between Hedge Fund strategies. The RMT 'cleaned' correlation matrix was explored using eigenvector analysis, with useful information found in the largest 4 eigenvalues. Using the 'cleaned' correlation matrix as an input to a classical portfolio optimisation, we showed how the difference between predicted and realised risk could be improved substantially, (by 35%).

The information contained in low-frequency Hedge Fund returns can be tainted by artifacts such as 'stale prices' for instruments held by managers. In **Chapter 5**, we developed a novel technique to overcome this problem, using the maximum overlap discrete wavelet transform to decompose the signal into different time horizons. The correlations between Hedge Funds and the S&P 500 were calculated using the MODWT and found to vary according to scale. Using classical portfolio optimisation, we showed how a correlation matrix calculated over longer scales, enhanced the risk profile for a portfolio of Hedge Funds.

We then turned our attention to medium-frequency equity returns and demonstrated

how changes in the global correlation structure, captured by the dynamics of the largest eigenvalue, can be used to characterise the risk of financial markets, **Chapter 6**. When the largest eigenvalue was raised, this was found to correspond to 'drawdown' periods with the opposite occurring during 'drawups'. Due to eigenspectrum level repulsion, opposite behaviour was observed for small eigenvalues. Finally, a 'toy model' of correlations was developed, which captured many of the features observed in the empirical data and enhanced our understanding of correlation dynamics and the relationship with financial risk.

In **Chapter 7**, we developed this work further through the analysis of both temporal and scale dependencies of the correlation matrix. Using the maximum overlap discrete wavelet transform, we constructed the correlation matrix between both medium and high-frequency data over a variety of scales, using a sliding window approach. This revealed additional insights, with the correlation dynamics found to depend upon the scale studied. Further, the eigenvalues were also found to be raised during times of market stress *at all scales*, demonstrating further the ability to characterise financial risk according to changes in the correlation structure. Using high-frequency data, we then demonstrated the gradual build-up of the correlation dynamics up to a characteristic time-scale of approximately one day, with more obvious scale dependent changes occurring thereafter.

Interested in similarities with other multiscale complex data models, we considered EEG seizure data, **Chapter 8**. For this, we showed how changes in the correlation dynamics between channels could be used to characterise the risk of seizures. The maximum overlap discrete wavelet transform was again used to calculate correlation matrices at different scales and to find the associated wavelet energy at each scale. The correlation dynamics were found to be scale dependent, with increased global correlation (and associated energy) during seizures at the highest frequencies. At lower frequencies, a compensating decrease in energy was found, with decreased global correlation. The very distinct and even opposite behaviour at different scales was in contrast to that found for financial data. The continuous nature of information processing was advanced as one factor which might explain these differences but data was too limited to confirm this.

Throughout this thesis, we have explored both the spatiotemporal interaction within sys-

tems of varying granularity and the application of the correlation measure to reduce the risk inherent in such systems. In the context of financial markets, the reduction and characterisation of risk is hugely topical at the present time, against a backdrop of large banking and personal losses. Risk management is not exclusive to financial systems, however, and we have also explored, in a preliminary way, similarities and contrasts for an apparently comparable system, namely EEG seizure patterns. Although integration of techniques offers some new insight, the dynamic, evolving nature of finanical markets means that managing the risk involved will continue to present truly complex challenges.

9.1 Future Work

Given a larger set of Hedge Funds, (or a set with daily returns), it is reasonable to suggest that additional sectoral features might emerge from analysis of the correlation matrix, **Chapter 4**. Additionally, the application of improved cleaning techniques which ensure the stability of the matrix, [Sharifi *et al.*, 2004; Daly *et al.*, 2008], may further improve the risk-return profile for a portfolio of Hedge Funds. Building on the analysis of **Chapter 5**, the measurement of the Alpha, (Equation 3.10), of different strategies over different time scales is also of potential future interest. In the case of Hedge Funds, the Alpha is a measure of the abnormal return, which is the value added by the manager, [Lhabitant, 2002, 2004]. A Hedge Fund manager who *appears to add Alpha*, (using monthly returns), may actually be found to just hold market risk over longer time scales and hence be less interesting to investors, (diversification benefits may be limited).

The study of the relationship between the direction of the market and magnitude of the eigenvalues of the correlation matrix, **Chapters 6-7**, indicates the need for more detailed insight. In particular, the analysis of high-frequency data may be useful in early warning of future market turmoil. The interactions between high-frequency data may also help to integrate trading strategies into the analysis, developing our understanding of the roles of cooperation and competition in financial markets. Moreover, study of the possible relationship between the dynamics of the small eigenvalues may reveal additional correlation

information hidden in that part of the eigenvalue spectrum, normally classified as noise. Such an investigation could be achieved through analysis of the relative dynamics of the small and large eigenvalues at times of extreme volatility, (e.g. during market crashes). For a review for this question to date, see Guhr and Kälber [2003]; Burda *et al.* [2004]; Burda and Jurkiewicz [2004]; Malevergne and Sornette [2004]; Müller *et al.* [2005, 2006a]; Kwapien *et al.* [2006]. Another interesting possbile are of future study is the persistence of correlations, using for example, the Hurst exponent.

The analysis of multiscale correlation dynamics in EEG seizure data, **Chapter 8**, was restricted by the difficulties in obtaining large data sets. The work by Schindler *et al.* [2007a] finds a decrease in overall correlation after seizure start, followed by an increase in correlation as the seizure ends. Our initial results suggest similar behaviour at lower frequencies, while the opposite occurs at higher frequencies. This suggests that the examination of correlation dynamics across various frequencies prior to seizure beginning may reveal pre-seizure characteristics, which can be used to calibrate seizure prevention strategies. An in-depth study on different seizure types may reveal further distinct correlation structures specific to the seizure type. Analysis of inter-frequency correlations may well shed light on the lead-lag relationship across different frequencies. Furthermore, investigation of the second and subsequent eigenvalues in detail could expose additional sub-system behaviour not revealed by the largest.

The correlation technique used in this thesis to measure the interaction between agents suffers from the drawback of being linear and hence neglects any higher-order relationships. Application of non-linear information-theory based dependence measures, will allow the detection of complex changes in synchronisation behaviour around extreme financial events. These non-linear techniques will result in additional features emerging to those highlighted above and may result in signals that warn of likely future market turmoil. The interaction of various agents with competing strategies and operating over different time scales will be analysed, resulting in an enhanced understanding of the phenomena underlying events such as financial crashes. Additionally, these techniques would have application to other Complex Systems, such as the EEG seizure data described.

BIBLIOGRAPHY

Adeli, H., Zhou, Z., Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *J. Neurosci. Meth.*, **123**, 69–87.

Aki, K., Richards, P.G. (1997). *Quantitative Seismology: Theory and Methods*. University Science Books; 2nd edition.

Aldroubi, A., Unser, M. (1996). *Wavelets in Medicine and Biology*. CRC-Press.

Ansari-Asl, K., Bellanger, J-J., Bartolomei, F., Wendling, F., Senhadji, L. (2005). Time-frequency characterization of interdependencies in nonstationary signals: application to epileptic EEG. *IEEE T. Bio.-Med. Eng.*, **52 (7)**, 1218–1226.

Ansari-Asl, K., Senhadji, L., Bellanger, J-J., Wendling, F. (2006). Quantitative evaluation of linear and nonlinear methods characterizing interdependencies between brain signals. *Phys. Rev. E*, **74**, 031916.

Bachelier, L. (1900). *Théorie de la Speculation*. Ph.D. thesis, University of Paris.

Bar-Yam, Y. (2003). *Dynamics of Complex Systems*. Westview Press.

Barbosa, S.M., Fernandes, M.J., Silva, M.E. (2006). Long-range dependence in North Atlantic sea level. *Physica A*, **371 (2)**, 725–731.

Barclays. (2009). *BarclayHedge, Alternative investment databases*. www.barclayhedge.com, [Last accessed: 12th January 2009].

Blank, A., Solomon, S. (2000). Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Physica A*, **287 (1-2)**, 279–288.

Bouchaud, J. P. (2002). An introduction to statistical finance. *Physica A*, **313**, 238–251.

Bouchaud, J. P., Potters, M. (2003). *Theory of Financial Risk and Derivative Pricing*. Cambridge University Press.

Bracewell, R. (1999). *The Fourier Transform and its applications*. McGraw-Hill.

Bragin, A., Engel Jr., J., Wilson, C.L., Fried, I., Mathern, G.W. (1999). Hippocampal and entorhinal cortex high-frequency oscillations ($100-500$Hz) in human epileptic brain and in kainic acid-treated rats with chronic seizures. *Epilepsia*, **40 (2)**, 127–137.

Bruce, A., Gao, H. (1996). *Applied Wavelet Analysis with S-PLUS*. Springer.

Bulmer, M. (2003). *Francis Galton: Pioneer of Heredity and Biometry*. Johns Hopkins University Press.

Burda, Z., Jurkiewicz, J. (2004). Signal and Noise in Financial Correlation Matrices. *Physica A*, **344**, 67–72.

Burda, Z., Jurkiewicz, J., Nowak, M. A. (2003). Is Econophysics a Solid Science? *Acta Physica Polonica B*, **34**, 87.

Burda, Z., Görlich, A., Jarosz, A., Jurkiewicz, J. (2004). Signal and noise in correlation matrices. *Physica A*, **343**, 295–310.

Burrus, C. S., Gopinath, R. A., Guo, H. (1998). *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice Hall.

Clark, I., Biscay, R., Echeverria, M., Virues, T. (1995). Multiresolution decomposition of non-stationary EEG signals: a preliminary study. *Comput. Biol. Med.*, **25 (4)**, 373–382.

Conlon, T., Ruskin, H. J., Crane, M. (2007). Random matrix theory and funds of funds portfolio optimisation. *Physica A*, **382 (2)**, 565–576.

Conlon, T., Crane, M., Ruskin, H. J. (2008). Wavelet Multiscale Analysis for Hedge Funds: Scaling and strategies. *Physica A*, **387 (21)**, 5197–5204.

Conlon, T., Ruskin, H. J., Crane, M. (2009). Cross-Correlation Dynamics in Financial Time Series. *Physica A*, **388 (5)**, 705–714.

Corning, P. A. (2003). The Re-Emergence of "Emergence": A venerable concept in search of a theory. *Complexity*, **7 (6)**, 18–30.

Crawley, M.J. (2005). *Statistics: An Introduction using R*. WileyBlackwell.

CreditSuisse. (2009). *Credit Suisse/Tremont Hedge Fund Indices*. www.hedgeindex.com, [Last accessed: 7th April 2009].

Crowley, P. M. (2005). An intuitive guide to wavelets for economists. *Bank of Finland Discussion paper*, **1**.

Daly, J., Crane, M., Ruskin, H. J. (2008). Random matrix filters in portfolio optimisation: A stability and risk assessment. *Physica A*, **387**, 4248–4260.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM Vol. 61.

Davis, M. (2001). *Mathematics of financial markets*. Edited by Enqquist, B. and Schmid, W., Springer.

De Liso, N., Filatrella, G. (2002). Econophysics: The Emergence of a New Field? *Journal of analytical and institutional economics*, **2**, 297–332.

Dow Jones. (2009). *Euro Stoxx 50*. $www.stoxx.com/indices/index_information.html?symbol = SXXE$, [Last accessed: 2nd March 2009].

Drozdz, S., Gruemmer, F., Ruf, F., Speth, J. (2000). Dynamics of competition between collectivity and noise in the stock market. *Physica A*, **287**, 440–449.

Drozdz, S., Kwapien, J., Grümmer, F., Ruf, F., Speth, J. (2001a). Quantifying the dynamics of financial correlations. *Physica A*, **299 (1-2)**, 144–153.

Drozdz, S., Gruemmer, F., Ruf, F., Speth, J. (2001b). Towards identifying the world stock market cross-correlations: DAX versus Dow-Jones. *Physica A*, **294**, 226–234.

Dyson, F. (1962). Statistical theory of the energy levels of complex systems i-iii. *J. Math. Phys.*, **3**, 140–175.

Dyson, F., Mehta, M. (1963). Statistical theory of the energy levels of complex systems iv & v. *J. Math. Phys.*, **4 (5)**, 701–712.

Edelman, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. App*, **9**, 543–560.

Elton, E. J., Gruber, M. J. (2002). *Modern Portfolio Theory and Investment Analysis, 6th Edition*. John Wiley & Sons.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *J. Am. Stat. Assoc*, **74**, 291–298.

Fisch, B.J., Spehlmann, R. (1999). *Fisch and Spehlmann's EEG primer: Basic principles of digital and analog EEG*. Elsevier Health Sciences.

Gallegati, M. (2005). A Wavelet Analysis of MENA Stock Markets. *Finance*, **0512027**.

Gallegati, M., Gallegati, M. (2007). Wavelet variance Analysis of Output in G-7 Countries. *Stud. Nonlinear Dyn. E.*, **11 (3)**, 1435.

Galluccio, S., Bouchaud, J. P., Potters, M. (1998). Rational Decisions, Random Matrices and Spin Glasses. *Physica A*, **259**, 449–456.

Gastaut, H., Broughton, R. (1973). *Epileptic Seizures: Clinical and electrographic features, diagnosis and treatment*. C.C. Thomas.

Gençay, R., Selcuk, F., Whitcher, B. (2001a). Differentiating intraday seasonalities through wavelet multi-scaling. *Physica A*, **289**, 543–556.

Gençay, R., Selcuk, F., Whitcher, B. (2001b). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press.

Gençay, R., Selcuk, F., Whitcher, B. (2001c). Scaling properties of foreign exchange volatility. *Physica A*, **289**, 249–266.

Gençay, R., Selcuk, F., Whitcher, B. (2003). Systematic risk and time scales. *Quantitative Finance*, **3 (2)**, 108–116.

Gençay, R., Selcuk, F., Whitcher, B. (2005). Multiscale systematic risk. *Journal of International Money and Finance*, **24**, 55–77.

Gigola, S., Ortiz, F., D'Attellis, C. E., Silva, W., Kochen, S. (2004). Prediction of epileptic seizures using accumulated energy in a multiresolution framework. *J. Neurosci. Meth.*, **138**, 107–111.

Glaser, R. (2000). *Biophysics*. Springer-Verlag.

Gligor, M., Ignat, M. (2001). Econophysics: A new field for Statistical Physics? *Interdisciplinary Science Reviews*, **26 (3)**, 183–190.

Google Search. (2009). *Econophysics*. www.google.com, [Last accessed: 30th January 2009].

Gopikrishnan, P., Rosenow, B., Plerou, V., Stanley, H. E. (2000). Identifying business sectors from stock price fluctuations. *Phys. Rev. E*, **64**, 035106R.

Guhr, T., Kälber, B. (2003). A new method to estimate the noise in financial correlation matrices. *J. Phys. A*, **35**, 3009–3032.

Holton, J.R. (2004). *An introduction to Dynamic Meteorology*. Academic Press.

Hull, J.C. (2000). *Options, Futures and Other Derivatives, 4th Edition*. Prentice Hall.

Hunger, M., Smith, R.L.L., Hyslop, W., Rosso, O.A., Gerlach, R., Rostas, J.A.P., Williams, D.B., Henskens, F. (2005). The Australian EEG Database. *Clin. EEG Neurosci.*, **36 (2)**, 76–81.

Indirdevi, K. P., Elias, E., Sathidevi, P. S., Dinesh Nayak, S., Radhakrishnan, K. (2008). A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalgram. *Comput. Bio. Med.*, **38**, 805–816.

Jaffard, S., Meyer, Y., Ryan, R. D. (2001). *Wavelets: Tools for Science and Technology*. Society for industrial maths.

Johnson, N. F., Jefferies, P., Hui, P. M. (2003). *Financial Market Complexity: What Physics Can Tell Us About Market Behaviour*. Oxford University Press.

Jolliffe, I.T. (2002). *Principal Component Analysis, 2nd edition*. Springer series in statistics.

Kaiser, G. (1994). *A friendly guide to Wavelets*. Birkhäuser.

Kaneko, K. (2006). *Life: An introduction to Complex Systems Biology*. Springer.

Kanji, G. K. (2006). *100 Statistical Tests*. Sage Publications, 3rd Edition.

Karpoff, J. J. (1987). The relationship between price changes and trading volumes: A survey. *J. Financ. Quant. Anal.*, **22**, 109–126.

Kelso, J.A.S. (1997). *Dynamic Patterns: The Self-Organisation of Brain and Behaviour*. MIT Press.

Kertesz, J., Kondon, I. (1998). *Econophysics: An Emerging Science, Proceedings 1997 Budapest Conference*. Kluwer, Dordrecht.

Kim, D. H., Jeong, H. (2005). Systemic analysis of group identification in stock markets. *Phys. Rev. E*, **72**, 0462133.

Kleinert, H. (2006). *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*. World Scientific Publishing.

Krzanowski, W. (1984). Sensitivity of principal components. *J. Royal Stats. Soc. B.*, **46 (3)**, 558–563.

Kwapien, J., Drozdz, S., Liu, L. C., Ionannides, A. A. (1998). Cooperative dynamics in auditory brain responce. *Phys. Rev. E*, **58**, 6359–6367.

Kwapien, J., Drozdz, S., A., Ionannides A. (2000). Temporal correlations versus noise in the correlation matrix formalism: an example of the brain auditory response. *Phys. Rev. E*, **62**, 5557–5564.

Kwapien, J., Drozdz, S., Grümmer, F., Ruf, F., Speth, J. (2002). Decomposing the stock market intraday dynamics. *Physica A*, **309 (1-2)**, 171–182.

Kwapien, J., Drozdz, S., Oswieimka, P. (2005). The bulk of the stock market correlation matrix is not pure noise. *Physica A*, **359**, 589–606.

Kwapien, J., Drozdz, S., Oswiecimka, P. (2006). The bulk of the stock market correlation matrix is not pure noise. *Physica A*, **359**, 589–606.

Laloux, L., Cizeau, P., Potters, M., Bouchaud, J. P. (1999). Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, **83 (7)**, 1467–1470.

Laloux, L., Cizeau, P., Potters, M., Bouchaud, J. P. (2000). Random matrix theory and financial correlations. *Int. J. Theoret. & Appl. Finance*, **3 (3)**, 391–397.

LeBaron, B., Arthur, W. B., Palmer, R. (1999). Time series properties of an artificial stock market. *J. Econ. Dyn. Control*, **23**, 1487–1516.

Lhabitant, F. S. (2002). *Hedge Funds*. Wiley.

Lhabitant, F. S. (2004). *Hedge Funds - Quantitative Insights*. Wiley.

Liang, B. (2001). Hedge fund performance: 1990-1999. *Financ. Anal. J.*, **57 (1)**, 11–18.

Lowrie, W. (1997). *Fundamentals of Geophysics*. Cambridge University Press.

Malevergne, Y., Sornette, D. (2004). Collective origin of the coexistence of apparent random matrix theory noise and of factors in large sample correlation matrices. *Physica A*, **331 (3-4)**, 660–668.

Malevergne, Y., Sornette, D. (2005). *Extreme Financial Risks: From Dependence to Risk Management*. Springer.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, **36**, 394–419.

Mantegna, R. N., Stanley, H. E. (2005). *An introduction to Econophysics: Correlations and complexity in finance*. Cambridge University Press.

Markowitz, H. (1958). *Portfolio Selection: Efficient diversification of investments*. John Wiley & sons.

McCauley, J. L. (2004). *Dynamics of Markets: Econophysics and Finance*. Cambridge University Press.

Mehta, M. (2004). *Random Matrices*. Academic Press.

Miceli, M. A. (2004). Ultrametricity in fund of funds diversification. *Physica A*, **344**, 94–99.

Miller, J. H., Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princton Universtiy Press.

Mix, D. F., Olenjniczak, K. (2003). *Elements of Wavelets for Engineers and Scientists*. Wiley-Interscience.

Mizuno-Matsumoto, Y., Ukai, S., Ishii, R., Date, S., Kaishima, T., Shinosaki, K., Shimojo, S., Takeda, M., Tamura, S., Inouye, T. (2005). Wavelet cross-correlation analysis: non-stationary analysis of neurophyiological signals. *Brain Topogr.*, **17 (4)**, 237–252.

Mormann, F., Kreuz, T., Rieke, C., Andrzejak, R. G., Kraskov, A., David, P., Elger, C. E., Lehnertz, K. (2005). On the predictibility of epileptic seizures. *Clin. Neurophysiol.*, **116**, 569–587.

Müller, M., Baier, G., Galka, A., Stephani, U., Muhle, H. (2005). Detection and characterization of changes of the correlation matrix in multivariate time series. *Phys. Rev. E*, **71**, 046116.

Müller, M., Jiménez, Y. L., Rummel, C., Baier, G., Galka, A., Stephani, U., Muhle, H. (2006a). Localized short-range correlations in the spectrum of the equal-time correlation matrix. *Phys. Rev. E*, **74**, 041119.

Müller, M., Wegner, K., Kummer, U., Baier, G. (2006b). Quantification of cross-correlations in complex spatiotemporal systems. *Phys. Rev. E*, **73**, 046106.

Nakata, M., Mukawa, K. (1989). Fourier analysis of broad spectrum EEG from a fluctuation point of view. *Integr. Psychiat. Beh. Sci.*, **24 (3)**, 90–97.

Nishiyama, N. (2001). One idea of portfolio risk control for absolute return strategy risk adjustments by signals from correlation behavior. *Physica A*, **301 (1-4)**, 457–472.

Noh, J. D. (2000). Model for correlations in stock markets. *Phys. Rev. E*, **61**, 5981–5982.

Norsworthy, J. R., Li, D., Gorener, R. (2000). Wavelet-based analysis of time series: An export from engineering to finance. *Engineering Management Society, Proceedings of the 2000 IEEE*, 126–32.

Onnela, J.-P., Kaski, K., Kertesz, K. (2004). Clustering and information in correlation based financial networks. *Eur. Phys. J. B*, **38 (2)**, 353–362.

Onnela, J.P., Chakraborti, A., Kaski, K., Kertesz, J. (2003). Dynamic asset trees and Black Monday. *Physica A*, **324**, 247–252.

Ouyang, G., Li, X., Li, Y., Guan., X. (2007). Application of wavelet-based similarity analysis to epileptic seizures prediction. *Comput. Bio. Med.*, **37**, 430–437.

Pafka, S., Kondor, I. (2002). Noisy covariance matrices and portfolio optimization. *Eur. Phys. J.*, **B27**, 277–280.

Pafka, S., Kondor, I. (2003). Noisy covariance matrices and portfolio optimization II. *Physica A*, **319**, 487–494.

Pafka, S., Kondor, I. (2004). Estimated correlation matrices and portfolio optimization. *Physica A*, **343**, 623–634.

Pafka, S., Potters, M., Kondor, I. (2004). Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization. *cond-mat/0402573, accessed 29/09/2006*.

Papp, G., Pafka, Sz., Nowak, M., Kondor, I. (2005). Random Matrix Filtering in Portfolio Optimization. *Acta Physica Pol. B*, **36 (9)**, 2757–2765.

Percival, D. B., Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University press.

Percival, D.B. (1995). On estimation of the wavelet variance. *Biometrika*, **82 (3)**, 619–631.

Perello, J. (2007). Downside Risk analysis applied to the Hedge Funds universe. *Physica A*, **383 (2)**, 480–496.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Guhr, T., Stanley, H. E. (1999). Universal and non-universal properties of cross-correlations in financial time series. *Phys. Rev. Lett.*, **83 (7)**, 1471–1474.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Stanley, H. E. (2000a). Econophysics: financial time series from a statistical point of view. *Physica A*, **279**, 443–456.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Stanley, H. E. (2000b). A random matrix theory approach to financial cross-correlations. *Physica A*, **287 (3-4)**, 374–382.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Guhr, T., Stanley, H. E. (2002). A random matrix approach to cross-correlations in financial data. *Phys. Rev. E*, **65**, 066126.

Podobnik, B., Stanley, H. E. (2008). Detrended cross-correlation analysis: A new method for analysing two non-stationary time series. *Phys. Rev. Lett.*, **100**, 084102.

Potters, M., Bouchaud, J. P., Laloux, L. (2005). Financial Applications of Random Matrix Theory: Old Laces and New Pieces. *Acta Physica Pol. B*, **36 (9)**, 2767.

Pozzi, F., Aste, T., Rotundo, G., Di Matteo, T. (2007). Dynamical correlations in financial systems. *Proc. SPIE*, **6802**, 68021E.

Press, W. H., Teukolsky, S. A., Vetterling, W.T., Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press.

Quian Quiroga, R., Kraskov, A., Kreuz, T., Grassberger, P. (2002). Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Phys. Rev. E*, **65**, 041903.

Ramsey, J. B. (2002). Wavelets in economics & finance: Past and Future. *Stud. Nonlinear Dyn. E.*, **6 (3)**, Article 1.

Razdan, A. (2004). Wavelet Correlation of 'strongly correlated' financial time series. *Physica A*, **333**, 335–342.

Roehner, B. M. (2002). *Patterns of Speculation: A study in observational econophysics*. Cambridge University Press.

Rosenow, B., Plerou, V., Gopikrishnan, P., Stanley, H. E. (2002a). Portfolio Optimization and the Random Magnet Problem. *Eur. Phys. Lett.*, **59**, 500–506.

Rosenow, B., Gopikrishnan, P., Plerou, V., Stanley, H. E. (2002b). Random Magnets and correlations of stock price fluctuations. *Physica A*, **314**, 762–767.

Rosenow, B., Gopikrishnan, P., Plerou, V., Stanley, H. E. (2003). Dynamics of cross-correlations in the stock market. *Physica A*, **324**, 241–246.

Rosso, O. A., Martin, M. T., Figliola, A., Keller, K., Plastino, A. (2006). EEG analysis using wavelet-based information tools. *J. Neurosci. Meth.*, **153**, 163–182.

Rotundo, G., Navarra, M. (2007). On the maxiumum drawdown during speculative bubbles. *Physica A*, **382**, 235–246.

Schindler, K., Leung, H., Elger, C. E., Lehnertz, K. (2007a). Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG. *Brain*, **130**, 65–77.

Schindler, K., Elger, C. E., Lehnertz, K. (2007b). Increasing synchronization may promote seizure termination: Evidence from Status Epilepticus. *Clin. Neurophysiol.*, **118 (9)**, 1955–1968.

Schleicher, C. (2002). An introduction to wavelets for economists. *Bank of Canada working paper*, **02-3**.

Schneeweis, T., Martin, G. (2001). The benefits of hedge funds: Asset allocation for the institutional investor. *Journal of Alternative Investments*, **4**, 7–26.

Seba, P. (2003). Random matrix analysis of human EEG data. *Phys. Rev. Lett.*, **91 (19)**, 198104.

Sengupta, A. M., Mitra, P. P. (1999). Distributions of singular values for some random matrices. *Phys. Rev. E*, **60**, 3389–3392.

Senhadji, S., Wendling, F. (2001). Epileptic transient detection: wavelets and time-frequency approaches. *Clin. Neurophyiol.*, **32**, 175–192.

Sharifi, S., Crane, M., Shamaie, A., Ruskin, H. J. (2004). Random matrix theory for portfolio optimization: a stability approach. *Physica A*, **335 (3-4)**, 629–643.

Sharkasi, A., Crane, M., Ruskin, H. J. (2006a). *Apples and Oranges: the Difference Between the Reaction of Emerging and Mature Markets to Crashes*. Proceedings of the Third Nikkei Econophysics Symposium - Practical Fruits of Econophysics, Tokyo, Japan, Springer-Verlag.

Sharkasi, A., Crane, M., Ruskin, H. J., Matos, J. A. (2006b). The Reaction of Stock Markets to Crashes and Events: A Comparison Study between Emerging and Mature Markets using Wavelet Transforms. *Physica A*, **368 (2)**, 511–521.

Sharpe, W. F. (1964). Capital asset prices: a theory of capital market equilibrium under conditions of risk. *J. Finance*, **35**, 910–919.

Sornette, D. (2002). *Why stock markets crash: Critical events in complex financial systems*. Princeton University Press.

Standard and Poor's. (2009). *S&P 500*. $www2.standardandpoors.com$, [Last accessed: 2nd March 2009].

Tóth, B., Kertész, J. (2006). Increasing market efficiency: Evolution of cross-correlations of stock returns. *Physica A*, **360**, 505–515.

Tóth, B., Kertész, J. (2007a). The Epps effect revisited. *Accepted: Quantitative Finance*.

Tóth, B., Kertész, J. (2007b). On the origin of the Epps effect. *Physica A*, **383**, 54–84.

Tóth, B., Tóth, B., Kertész, J. (2007). Modelling the Epps effect of cross correlations in asset prices. *Proceedings of SPIE (Fluctuations and Noise 2007)*, **6601**.

Tumminello, M., Aste, T., Di Matteo, T., Mantegna, R.N. (2007). Correlation based networks of equity returns sampled at different time horizons. *Eur. Phys. J. B*, **55**, 209–217.

University of Freiburg. (2008). *EEG Database*. $https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database$, [Last accessed: 1st February 2008].

Ursino, M., Magosso, E., Gardella, E., Rubboli, G., Tassinari, C. A. (2004). A wavelet based analysis of energy distribution in scalp EEG during epileptic seizures. *Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA*, 255–258.

Utsugi, A., Ino, K., Oshikawa, M. (2004). Random matrix theory analysis of cross correlations in financial markets. *Phys. Rev. E*, **70**, 026110.

Vasconcelos, G. L. (2004). A Guided Walk Down Wall Street: an Introduction to Econophysics. *Braz. J. Phys.*, **34**, 1039–1065.

Voit, J. (2003). *The Statistical Mechanics of Financial Markets*. Springer.

Wang, Y., Wu, J., Di, Z. (2004). Physics of Econophysics. *Cond-mat/0401025, accessed 29/09/2006*.

Wendling, F., Bartolomei, F., Bellanger, J. J., Bourien, J., Chauvel, P. (2003). Epileptic fast intracerebral EEG activity: evidence for spatial decorrelation at seizure onset. *Brain*, **126**, 1449–1459.

Whitcher, B., Guttorp, P., Percival, D.B. (1996). Mathematical background for wavelet estimators of cross-covariance and cross-correlation. *NRCSE Technical Report Series*, **38**.

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

Yahoo. (2009). *S&P 500 Index returns data*. finance.yahoo.com, [Last accessed: 7th April 2009].

Yakovenko, V. M. (2003). Research in Econophysics. *Cond-mat/0302270v2, accessed 29/09/2006*.

Yamada, H. (2005). Wavelet-based beta estimation and Japanese industrial stock prices. *Appl. Econ. Lett.*, **12 (2)**, 85–88.

Ying, C.C. (1966). Stock market prices and volumes of sales. *Econometrica*, **34**, 676–686.

List of Publications & Presentations

Conlon, T., Ruskin, H.J., Crane, M. Random Matrix Theory and Hedge Funds Strategy Identification. Presented at the 4^{th} *INFINITI Conference on International Finance*, Trinity College Dublin, 12-13 June 2006.

Conlon, T., Ruskin, H.J., Crane, M. (2007). Random Matrix Theory and fund of funds portfolio optimisation. *Physica A* **382 (2)** 565 - 576.

Conlon, T., Crane, M., Ruskin, H.J. (2008). Wavelet Multiscale Analysis for Hedge Funds: Scaling and strategies. *Physica A* **387 (21)** 5197 - 5204.

Conlon, T., Ruskin, H.J., Crane, M. Multiscaled Cross-Correlation Dynamics in Financial Time Series. Presented at the 5^{th} *European Conference on Complex Systems*, Jerusalem, Israel, Sept 10-19 2008.

Conlon, T., Ruskin, H.J., Crane, M. (2009). Cross-Correlation Dynamics in Financial Time Series. *Physica A* **388 (5)** 705 - 714.

Conlon, T., Ruskin, H.J., Crane, M. Cross-Correlation Dynamics in Financial Time Series. *Computers in Biology and Medicine* **39** 760 - 767.

Conlon, T., Ruskin, H.J., Crane, M. Multiscaled Cross-Correlation Dynamics in Financial Time Series. Accepted for Publication: *Advances in Complex Systems*, (June 2009).

APPENDIX A

HEDGE FUND STRATEGIES

Strategies employed by the managers in the sample considered in Chapter 4:

Strategies	**Number of Funds**
Asia excluding Japan Long/Short Equities	2
Convertible & Equity Arbitrage	2
Currency	7
Emerging Markets	6
European Long/Short Equity	10
Fixed Income	1
Global Equity	5
Japan Market Neutral	1
Macro	3
Managed Futures	11
Self-Invested Fund of Funds	1

Tab. A.1: Hedge Fund strategies from sample considered in Chapter 4

A.1 Strategy Overview

In the following, we describe some of the investment styles used by Hedge Fund managers, covering those used in the calculation of the Credit Suisse/Tremont indices, Chapter 5.

Convertible Arbitrage

Convertible Arbitrage managers seek to exploit pricing anomolies between a convertible bond and it's underlying equity. These anomolies occur as the embedded options in a convertible bond are often undervalued when compared with their theoretical value and is exploited by buying the bond and shorting either stock or options against it.

Dedicated Short Bias

The portfolio of a Dedicated Short Bias manager has more short positions than long positions in Equity securities (and derivatives of), leaving them with net short exposure to the market. They are considered a hedge against Bear markets.

Emerging Markets

Emerging Markets Managers invest in currencies, debt instruments and equities of developing countries. Examples of such markets include China, India, Russia and Brazil.

Equity Market Neutral

Equity Market Neutral is a style of investment that applies quantitative statistical models to exploit pricing inefficiencies between equity securities. The portfolio is formed to minimise exposure to the systematic market risk and this is achieved by offsetting long positions by short positions on a dollar neutral or zero beta basis.

Event Driven

Event Driven managers focus on debt and equity of companies where a specific corporate or market event is taking place. Examples include mergers, restructurings, spin-offs, bankrupcies, re-capitalisations and share buybacks.

Fixed Income Arbitrage

Fixed income arbitrage exploits inefficiencies within and across global fixed income markets. Strategies include yield curve and sovereign debt arbitrage, Treasury versus corporate and municipal yield spreads, basis trading and mortgage-backed security arbitrage.

Global Macro

Global macro managers make directional, leveraged, opportunistic bets on anticipated price movements in currencies, equities, bonds and comodities. The style is based upon top-down global analysis and focuses on fundamental economic, political and market factors.

Long/Short Equity

Equity Long/Short managers combine long investments with short sales to reduce but not elimate market exposure. Funds tend to have net long market exposure, so often suffer drawdowns at the same time as equity markets. Managers can also trade equity futures and options and use leverage to increase their exposures.

Managed Futures

Managed futures managers (or Commodity trading advisors (CTA's)) trade listed commodity, currency, bond and equity futures. Managers often employ systematic trading programs that rely on computer-generated trading signals to produce returns. Managers do not have a particualar bias towards a particular market or to being net long or short.

APPENDIX B

SOFTWARE

The code used in the studies above were all written using Matlab and related toolboxes such as the Wavelet and Statistics toolboxes. The calculations for the Maximum Overlap Discrete Wavelet Tranform (MODWT), Chapter 3.6, was implemented using the *WMTSA Wavelet Toolkit* for Matlab, which is the toolbox associated with Percival and Walden [2000]. The scripts involved are reproduced below:

```
function hfRMT
% THIS FUNCTION COMPARES THE EMPIRICAL CORRELATION MATRIX TO THE
% THEORETICAL DISTRIBUTION FROM RANDOM MATRIX THEORY

% Load Hedge Funds returns data
g = load('C:\PhD\49 funds 105 returns normalised.txt');

t = size(g,1);
n = size(g,2);

% Ratio of time to number of funds
q = t/n

% Data has been pre-normalised, so ....
sigma = 1;

% Work out the min and max RMT eigenvalues
lampda_max = (sigma^2)*(1 + (1/q) + 2*(sqrt(1/q)))
lampda_min = (sigma^2)*(1 + (1/q) - 2*(sqrt(1/q)))

% Calculate the empirical correlations and eigenvalues
```

```
corrFunds = corrcoef(g);
e = eig(corrFunds)

j = 1;
totP = 0;
for i = lampda_min+0.00001:0.001:lampda_max
    eval(j) = i;
    P(j) = (q/(2 * pi*sigma^2)) * (sqrt((lampda_max -i)*(i - lampda_min))/i);
    j = j + 1;
end

% Specify how many buckets you want to break the data up into
bkt = 1000;

% Here I get the normalisation constant so the area under the curve is 100%
dist = e(end,1) - e(1,1);
dist = dist/bkt;

% This outputs the buckets and the number of e/values in each bucket
[n, xout] = hist(e,bkt);

n(1,end + 1) = 0;
xout(1, end + 1) = (xout(1,end) - xout(1,end-1)) + xout(1,end);
n = n/sum(n);
n = n/(dist);

% Now plot the results
plot(xout,n,eval,P)
title('\fontsize{12}\bf Eigenvalue Distribution 49 funds, 105 months data');
set(gca, 'FontSize',12);
xlabel('\fontsize{12} Eigenvalues'); ylabel('\fontsize{12} Distribution');
legend('Empirical Distribution', 'Theoretical Distribution');
```

```
function RemoveLargEVec

% THIS FUNCTION REMOVES THE EFFECTS OF THE LARGEST EIGENVALUE FROM THE
% DATA; IE. REMOVES THE EFFECT OF THE 'MARKET'

g = load('C:\PhD\49 funds 105 returns Normalised.txt');
returns = load('C:\PhD\49 funds 105 returns Unnormalised.txt');
```

```
savefile = 'C:\PhD\Residuals.mat';

EVnumber = 49;

corrFunds = corrcoef(g);

[vec,val] = eig(corrFunds);

% Find the eigensignal corresponding to the largest eigenvalue
G_large =  returns*vec(:,EVnumber);

% Regress the returns for each stock against the largest eigenvalue
% and find the residuals
for i = 1:49
    [b(i),bint(:,i),r(:,i)] = regress(returns(:,i),G_large);
end

save(savefile, 'b', 'bint', 'r');
```

```
function evectorAnal

function evectorAnal

% THIS FUNCTION IS USED TO ANALYSE THE DISTRIBUTION OF EIGENVECTOR
% COMPONENTS, AS WELL AS THE INVERSE PARTICIPATION RATIO

% Pull in data
g = load('C:\PhD\49 funds 105 returns.txt');

% Number of buckets in graphs
bucket = 200;

%specify the eigenvalue to analyse
EVnumber = size(g,2);

% Calculate correlations and eigenspectrum
corrFunds = corrcoef(g);

[vec,val] = eig(corrFunds,'nobalance');

% Plot the eigenvector components
```

```
histfit(vec(:,EVnumber),25)
title('Distribution of Eigenvector Components');
xlabel('Eigenvector Components'); ylabel('Number of Occurences');

% Calculate the IPR for each eigenvalue
IPR = sum(vec.^4);

% Plot the IPR
figure;
plot(IPR)
title('Inverse Participation Ratio');
xlabel('Eigenvalue Number'); ylabel('Inverse Participation Ratio');
```

```
function portfundOpt
% PORTFOLIO OPTIMISATION FOR PREDICTED AND REALISED CORRELATIONS USING
% ORIGINAL AND CLEANED CORRELATION MATRIX

% Define portfolio returns
PortReturn = [0.005,0.006,0.0075, 0.01,0.0125,0.015,0.0175];

numPorts = length(PortReturn);

% Call data
fid = fopen('C:\PhD\Results\Portfolio Weights.txt', 'wt');
r = load('C:\PhD\Unnormalised Data.txt');
normData = load('C:\PhD\49 funds 105 returns Normalised.txt');

% Break up data into 2 and calculate returns, std, cov and correlations
r1 = normData(1:53,:); r2 = normData(54:end,:);

ExpReturn1 = mean(r(1:53,:)); ExpReturn2 = mean(r(54:end,:));

r1_std = std(r(1:53,:)); r2_std = std(r(54:end,:));

corrFunds1 = corrcoef(r1); corrFunds2 = corrcoef(r2);

ExpCovariance1 = corr2cov(r1_std, corrFunds1);
ExpCovariance2 = corr2cov(r2_std, corrFunds2);

% Set constraints for the portfolio optimisation
```

```
ConSet = portcons('PortValue', 1, 49,'AssetLims',0, 1, 49);

% Optimise portfolio for original data & get predicted and realised
% risk/return
[PortRisk, PortReturn, PortWts] = portopt(ExpReturn2, ...
ExpCovariance1, [],PortReturn,ConSet)
[ActPortRisk, ActPortReturn] = portstats(ExpReturn2, ExpCovariance2,PortWts);

% number of outlying eigenvalues
n = 3;
[V,D] = eig(corrFunds1);

% Clean the first correlation matrix
C_clean_1 = Clean_RMT_Bouchaud(V,D,n);
C_clean_1 = corr2cov(r1_std, C_clean_1);

[V2,D2] = eig(corrFunds2);

% Clean the 2nd correlation matrix
C_clean_2 = Clean_RMT_Bouchaud(V2,D2,n);
C_clean_2 = corr2cov(r2_std, C_clean_2);

% Optimise portfolio for cleaned data & find predicted and realised
% risk/return
[PortRisk_clean, PortReturn_clean, PortWts_clean] = ...
portopt(ExpReturn2, C_clean_1, [],PortReturn,ConSet);
[ActPortRisk_clean, ActPortReturn_clean] = portstats(ExpReturn2, ...
C_clean_2,PortWts_clean);

% Annualise the results for graphing
PortRisk = sqrt(12)*PortRisk; ActPortRisk = sqrt(12)*ActPortRisk;
PortRisk_clean = sqrt(12)*PortRisk_clean; ....
ActPortRisk_clean = sqrt(12)*ActPortRisk_clean;

for i = 1:numPorts
    PortReturn(i,1) = ((1+PortReturn(i,1))^12)-1;
    ActPortReturn(i,1) = ((1+ActPortReturn(i,1))^12)-1;
    PortReturn_clean(i,1) = ((1+PortReturn_clean(i,1))^12)-1;
    ActPortReturn_clean(i,1) = ((1+ActPortReturn_clean(i,1))^12)-1;
end
```

```
% Plot the results
hold on
plot(100*PortRisk, 100*PortReturn,'*--','LineWidth',1, 'color', [0,0,0.25098])
plot(100*ActPortRisk, 100*ActPortReturn,'o--','LineWidth',1, ...
'color', [0,0,0.25098])
plot(100*PortRisk_clean, 100*PortReturn_clean,'*:','LineWidth',1, ...
'color', [0,0,0.25098])
plot(100*ActPortRisk_clean, 100*ActPortReturn_clean,'o:','LineWidth',1, ...
 'color', [0,0,0.25098])
legend('Predicted risk using C', 'Realised risk using C','Predicted risk ....
  using filtered C', 'Realised risk using filtered C', 'location', 'SouthEast')
xlabel('Risk %'); ylabel('Return %');
```

```
function C_cl = Clean_RMT_Bouchaud(V,D,n)
% CLEAN CORRELATION MATRIX USING METHOD DESCRIBED BY BOUCHAUD ET AL

D_size=size(D, 1);

% D_mean is the mean of the noisy eigenvalues
D_mean=mean(diag(D(1:D_size-1, 1:D_size-n)));

% D_cl is the cleaned eigenvalue matrix
D_cl=zeros(D_size);

% Non-noisy part
for j=D_size-n:D_size
    for i=D_size-n:D_size
        D_cl(i, j)=D(i, j);
    end
end

% Noisy part is replaced by the identity matrix * mean of the noisy eigenvalues
for j=1:D_size-n
    for i=1:D_size-n
        if (j==i)
            D_cl(i, j)=D_mean;
        end
    end
end
```

```
Trace(D)
% Cleaning C
C_cl=V*D_cl*V';
```

```
function StatWaveletComps
% FUNCTION TO CALCULATE THE VARIANCES, COVARIANCES AND CORRELATIONS BETWEEN
% HEDGE FUNDS AT DIFFERENT SCALES USING THE MODWT.  USES THE WMTAS TOOLBOX
% FOR MATLAB

clear;

% Load data
s = load('C:\Program Files\MATLAB704\work\Thesis Code\Index Data SandP.txt');

n_c = size(s,2);

levelNo = 4;

Coefs = [];
% First Decompose the Data using the MODWT
for i = 1:n_c
    modwtCoefs = modwt(s(:,i),'LA8',levelNo);
    sizeCoefs = size(modwtCoefs);
    Coefs = [Coefs; modwtCoefs];
end

% Find Variances, Covariances & Correlations

for i = 1:levelNo

    for j = 0:n_c-1
           flag = sizeCoefs*j;
           Coefs1 = Coefs(flag + 1:flag + sizeCoefs, i);

           [waveVar,CI_wvar] = modwt_wvar(Coefs1,'gaussian','unbiased','LA8');
           varCoef(j+1, i) = waveVar;
           varCI1(j+1,i) = CI_wvar(:,1);
           varCI2(j+1,i) = CI_wvar(:,2);

       for k = 0:n_c-1
```

```
            flag2 = sizeCoefs*k;
            Coefs2 = Coefs(flag2 +1:flag2 + sizeCoefs, i);

            [waveCov, CI_wcov] = modwt_wcov(Coefs1, Coefs2,'gaussian', ...
             'unbiased','LA8');
            [waveCorr, CI_wcor] = modwt_wcor(Coefs1, Coefs2);

            corrCoef(j+1+n_c*(i-1), k+1) = waveCorr;
            covCI1(j+1+n_c*(i-1), k+1) = CI_wcov(:,1);
            covCI2(j+1+n_c*(i-1), k+1) = CI_wcov(:,2);
            covCoef(j+1+n_c*(i-1), k+1) = waveCov;

        end

    end

end


savefile = 'C:\PhD\Wavelets\Measure Alpha and Beta\Correlation SandP ...
db8 MODWT.txt';
save(savefile, 'corrCoef', '-ASCII');

savefile = 'C:\PhD\Wavelets\Measure Alpha and Beta\covCI1 SandP ...
db8 MODWT.txt';
save(savefile, 'covCI1', '-ASCII');

savefile = 'C:\PhD\Wavelets\Measure Alpha and Beta\covCI2 SandP ...
db8 MODWT.txt';
save(savefile, 'covCI2', '-ASCII');

savefile = 'C:\PhD\Wavelets\Measure Alpha and Beta\Covariance SandP ...
db8 MODWT.txt';
save(savefile, 'covCoef', '-ASCII');

savefile = 'C:\PhD\Wavelets\Measure Alpha and Beta\Variance SandP ...
db8 MODWT.txt';
save(savefile, 'varCoef', '-ASCII');

savefile = 'C:\PhD\Wavelets\Measure Alpha and Beta\Variance CI1 SandP ...
```

```
db8 MODWT.txt';
save(savefile, 'varCI1', '-ASCII');

savefile = 'C:\PhD\Wavelets\Measure Alpha and Beta\Variance CI2 SandP ...
db8 MODWT.txt';
save(savefile, 'varCI2', '-ASCII');
```

```
function portOptDiffScales
% PORTFOLIO OPTIMISATION USING CORRELATION MATRICES CALCULATED OVER
% DIFFERENT SCALES.  EFFICIENT FRONTIERS ARE FOUND FOR EACH SCALE

PortReturn = [0.098:0.005:0.14]

returnIndex = load('C:\PhD\Wavelets\Measure Alpha and Beta\...
 Port Opt\Index Returns.txt');
vol = load('C:\PhD\Wavelets\Measure Alpha and Beta\Port Opt\Index Stand Dev.txt');

OrigCorr = load('C:\PhD\Wavelets\Measure Alpha and Beta\Port Opt\...
Correlation Original Data.txt');
corr1 = load('C:\PhD\Wavelets\Measure Alpha and Beta\Port Opt\...
Correlation LA8 scale 1.txt');
corr2 = load('C:\PhD\Wavelets\Measure Alpha and Beta\Port Opt\...
Correlation LA8 scale 2.txt');
corr3 = load('C:\PhD\Wavelets\Measure Alpha and Beta\Port Opt\...
Correlation LA8 scale 3.txt');
corr4 = load('C:\PhD\Wavelets\Measure Alpha and Beta\Port Opt\...
Correlation LA8 scale 4.txt');

expCovOrig = corr2cov(vol, OrigCorr);
expCov1 = corr2cov(vol, corr1); expCov2 = corr2cov(vol, corr2);
expCov3 = corr2cov(vol, corr3); expCov4 = corr2cov(vol, corr4);


[PortRiskOrig, PortReturnOrig, PortWtsOrig] = portopt(returnIndex, ...
expCovOrig,[],PortReturn);
[PortRisk1, PortReturn1, PortWts1] = portopt(returnIndex, ...
expCov1,[],PortReturn);
[PortRisk2, PortReturn2, PortWts2] = portopt(returnIndex, ...
expCov2,[],PortReturn);
```

```
[PortRisk3, PortReturn3, PortWts3] = portopt(returnIndex, ...
expCov3,[],PortReturn);
[PortRisk4, PortReturn4, PortWts4] = portopt(returnIndex, ...
expCov4,[],PortReturn);

PortRisk1 = 100*PortRisk1; PortRisk2 = 100*PortRisk2;
PortRisk3 = 100*PortRisk3; PortRisk4 = 100*PortRisk4;
PortReturn1 = 100*PortReturn1; PortReturn2 = 100*PortReturn2;
PortReturn3 = 100*PortReturn3; PortReturn4 = 100*PortReturn4;

PortRiskOrig = 100*PortRiskOrig; PortReturnOrig = 100*PortReturnOrig;

plot(PortRiskOrig, PortReturnOrig,'*-',PortRisk1,PortReturn1,'p-.', ...
PortRisk2, PortReturn2,'+-', PortRisk3, PortReturn3,'x-', ...
PortRisk4, PortReturn4,'.-');
legend('Original Data','Scale 1', 'Scale 2','Scale 3','Scale 4', ...
'location', 'SouthEast');
xlabel('Risk (%)'); ylabel('Return (%)');
title('Efficient Frontier using Correlation Matrices caclulated ...
over different time horizons');
```

```
function eegPearsonCorrel
% Function used to plot normalised eigenvalues vs the Index

load('C:\PhD\Wavelets\Eigvalue Analysis\SX5Edata.mat','indexDates', ...
'indexPX', 'rtData');

% Set the returns data to be examined
 coef = rtData;

% select 100 stocks randomly (if needed, if not set > size(coef)

for i = 1:130

   R(i) = round(unifrnd(1,384));

end

Rleft = unique(R);  Rleft = Rleft(1:100);
```

```
tic;
signal = coef';

% Define Variables
blockSize = 100; blockSpace = 1; sig_freq = 1;

indexPXPLot = indexPX(blockSize:end-1,:);
sig_time_segment = 1/sig_freq;

% Find the Pearson Correlation & eigenvalues for each time window

j = 1;

for i = 1:blockSpace:(size(signal,2)-blockSize)

    signal_seg = signal(:,(i):(blockSize+i -1));

    sig_std = std(signal_seg,0,2);

    sig_mean = mean(signal_seg,2);

    A = repmat(sig_mean, 1, size(signal_seg,2));
    B = repmat(sig_std, 1,size(signal_seg,2));

    sig_norm = (signal_seg-A)./B;

    eegCorrel = corrcoef(sig_norm');

    eegEig = eig(eegCorrel);

    Eigs(:,j) = eegEig(:);

    j = j + 1;
end

% Here we normalise each of the eigenvalue by dividing by its mean &
% dividing by its STD

    eigs_std = std(Eigs,0,2);
    eigs_mean = mean(Eigs,2);
```

```
    A = repmat(eigs_mean, 1, size(Eigs,2));
    B = repmat(eigs_std, 1,size(Eigs,2));

    eigs_norm = (Eigs-A)./B;

eigsSmall = (eigs_norm(1,:));
eigsBulk = mean(eigs_norm(1:40,:));
eigsLarge = (eigs_norm(end,:));

%  Now plot the data

xData = indexDates(blockSize+2:end)';

clims = [0 100];

 subplot(4,1,1)
    plot(xData, indexPXPLot)
    axis([xData(1),xData(end),1500,5000]);
    text(729010,1500,'(a)','FontSize',14)
    legend('DJ Euro Stoxx 50, Apr 2002 - Aug 2007');
    datetick('x',10,'keeplimits');

maxLarge = max(eigsLarge); maxSmall = max(eigsSmall);
minLarge = min(eigsLarge); minSmall = min(eigsSmall);

maxEig = max([maxLarge,maxSmall]); minEig = min([minLarge,minSmall]);

 subplot(4,1,2)
    plot(xData, eigsLarge, xData, eigsBulk)
    axis([xData(1),xData(end),-3.5,3.5]);
    text(729010,2.25,'(b)','FontSize',14)
    legend( 'Largest E/Value','Average E/Values 1-40');
    datetick('x',10,'keeplimits');

    cmin = min(min(eigs_norm));
    cmax = max(max(eigs_norm));

 subplot(4,1,3)

    imagesc(min(xData):max(xData),0:size(eigs_norm,1), eigs_norm, clims)
    colormap(jet); set(gca, 'CLim', [cmin, cmax]);
```

```
    text(729010,10,'(c)','FontSize',14)
    datetick('x',10,'keeplimits');

 subplot(4,1,4)
    imagesc(min(xData):max(xData),40:size(eigs_norm,1), ...
     eigs_norm(40:end,:), clims);
    colormap(jet);  set(gca, 'CLim', [cmin, cmax])
    colorbar('location','southoutside');
    text(729010,91,'(d)','FontSize',14)
    datetick('x',10,'keeplimits');  xlabel('Time (Days)');

t = toc
```

```
function findDrawdowns
%% FIND MARKET RETURNS ASSOCIATED WITH DRAWDOWNS OR DRAWUPS

load('C:\PhD\Wavelets\Eigvalue Analysis\SX5Edata.mat','indexDates', ....
'indexPX', 'rtData');

indexReturns = indexPX(2:end)./indexPX(1:end-1)-1;

Returns = rtData;

blockSize = 200; blockSpace = 1; sig_freq = 1;
sig_time_segment = 1/sig_freq;

% Set the Normalisation Period
norm_period = [1:size(rtData,1)-blockSize];

% Find 200 day returns
for i = 1:1:size(indexPX)-blockSize
    cumReturn(i) = indexPX(i+blockSize)/indexPX(i)-1;
end

% Correlations and Eigenspectrum Original Data

for i = 1:1:size(Returns,1)-blockSize
    correl = corrcoef(Returns(i:i+blockSize,:));
    eigSXXP = eig(correl);
    Eigs(:,i) = eigSXXP;
```

```
end

     eigs_std = std(Eigs(:,norm_period),0,2);
     eigs_mean = mean(Eigs(:,norm_period),2);

     A = repmat(eigs_mean, 1, size(Eigs,2));
     B = repmat(eigs_std, 1,size(Eigs,2));

     eigs_norm = (Eigs-A)./B;

eigsSmallOrig = mean(eigs_norm(1:40,:));
eigsLargeOrig = (eigs_norm(end,:));

%%% Find eigenvalues < -1 & > 1 STD
drawUpSTD = 1;

eigsLargeDownOrig = find(eigsSmallOrig<-drawUpSTD);
eigsLargeUpOrig = find(eigsSmallOrig>drawUpSTD);

disp('Orig Data, Large Eigs, >1 std');
mean(cumReturn(eigsLargeUpOrig))
disp('Orig Data, Large Eigs, <-1 std');
mean(cumReturn(eigsLargeDownOrig))
```

```
unction modwtCorrCoefs
% FUNCTION TO CALCULATE THE CORRELATIONS BETWEEN STOCKS AT DIFFERENT SCALES
% USING THE MODWT.  USES THE WMTAS TOOLBOX FOR MATLAB.  WITH SLIGHT CHANGES
% (DATA INPUT) CAN ALSO BE USED FOR EEG EPILEPTIC DATA. WARNING: CAN TAKE
% A LONG TIME TO RUN

clear;

% Load data
load('C:\PhD\Eigenvalue Analysis\SX5E Data\SX5P.mat', 'Returns');
s = Returns;
n_c = size(s,2);

levelNo = 4;
blockSize = 300;
blockSpace = 250;
```

```
Coefs = [];
m = 1;

% Find correlations for each moving window
for l = 1:blockSpace:(size(s,1)-blockSize)

    signal_seg = s((l):(blockSize+l -1),:);


    % First Decompose the Data using the MODWT
    for n = 1:n_c
        modwtCoefs = modwt(signal_seg(:,n),'LA8',levelNo);
        sizeCoefs = size(modwtCoefs);
        Coefs = [Coefs; modwtCoefs];
    end

    % Find Correlations & Eigenvalues

    for i = 1:levelNo

        for j = 0:n_c-1
               flag = sizeCoefs*j;
               Coefs1 = Coefs(flag + 1:flag + sizeCoefs, i);

            for k = 0:n_c-1

               flag2 = sizeCoefs*k;
               Coefs2 = Coefs(flag2 +1:flag2 + sizeCoefs, i);

               [waveCorr, CI_wcor] = modwt_wcor(Coefs1, Coefs2);

               corrCoef(j+1, k+1) = waveCorr; %+n_c*(i-1)

            end

        end

        eigenV(m,:,i) = eig(corrCoef);
```

```
    end
    Coefs = [];
m = m + 1;
end

savefile = 'C:\PhD\Eigenvalue Analysis\SX5E Data\SX5Eeigs.mat';
save(savefile, 'eigenV');
```

```
function plotEigsScaled
% FUNCTION TO PLOT EIGENVALUES FOR EACH SCALE STUDIED IN THE FUNCTION
% MODWTCORRCOEFS.  THIS FUNCTION CAN BE USED FOR FINANCIAL DATA (AS BELOW)
% OR WITH MINOR MODIFICATIONS FOR EEG EPILEPTIC DATA

load('C:\PhD\Eigenvalue Analysis\SX5E Data\IndexPrice.mat', 'indexPX');
load('C:\PhD\Eigenvalue Analysis\SX5E Data\IndexDates.mat', 'indexDates');
load('C:\PhD\Eigenvalue Analysis\SX5E Data\IndexReturns.mat', 'IndexReturns');
load('C:\PhD\Eigenvalue Analysis\SX5E Data\SX5P.mat', 'Returns');
signal = Returns';

load('C:\PhD\Eigenvalue Analysis\SX5E Data\SX5Eeigs.mat', 'eigenV');
eigCC1 = eigenV(:,:,1)'; eigCC2 = eigenV(:,:,2)'; eigCC3 = eigenV(:,:,3)';

% Set Variables
blockSize = 300; blockSpace = 250;
sig_freq = 1; sig_time_segment = 1/sig_freq;
wname = 'sym4';

amax = 5; a = 2.^[1:amax];

% Calculate the approximate frequencies
f = scal2frq(a,wname,1/sig_freq);
scale = 1./f;

%%%%%%%%%%%%% Unfiltered Data
j = 1;
for i = 1:blockSpace:size(Returns,1)-blockSize

    correl = corrcoef(Returns(i:i+blockSize,:));
    Eigs(:,j) = eig(correl);
```

```
    j = j +1;
end

    eigs_std = std(Eigs,0,2);
    eigs_mean = mean(Eigs,2);
    A = repmat(eigs_mean, 1, size(Eigs,2));
    B = repmat(eigs_std, 1,size(Eigs,2));
    eigs_norm = (Eigs-A)./B;

eigsSmallOrig = mean(eigs_norm(1:40,:));
eigsLargeOrig = (eigs_norm(end,:));

%%%%%%%%%%%% Level 1

Eigs = eigCC1;

    eigs_std = std(Eigs,0,2);
    eigs_mean = mean(Eigs,2);
    A = repmat(eigs_mean, 1, size(Eigs,2));
    B = repmat(eigs_std, 1,size(Eigs,2));
    eigs_norm = (Eigs-A)./B;

eigsSmall1 = mean(eigs_norm(1:40,:));
eigsLarge1 = (eigs_norm(end,:));

%%%%%%%%%%%% Level 2

Eigs = eigCC2;

    eigs_std = std(Eigs,0,2);
    eigs_mean = mean(Eigs,2);
    A = repmat(eigs_mean, 1, size(Eigs,2));
    B = repmat(eigs_std, 1,size(Eigs,2));
    eigs_norm = (Eigs-A)./B;

eigsSmall2 = mean(eigs_norm(1:40,:));
eigsLarge2 = (eigs_norm(end,:));

%%%%%%%%%%%% Level 3
```

```
Eigs = eigCC3;

    eigs_std = std(Eigs,0,2);
    eigs_mean = mean(Eigs,2);
    A = repmat(eigs_mean, 1, size(Eigs,2));
    B = repmat(eigs_std, 1,size(Eigs,2));
    eigs_norm = (Eigs-A)./B;

eigsSmall3 = mean(eigs_norm(1:40,:));
eigsLarge3 = (eigs_norm(end,:));

%%% Find Time Frame
xData = indexDates';
xData = xData(1:blockSize:size(xData,2));

%% Plot Results
maxLarge = max(eigsLarge1); maxSmall = max(eigsSmall1);
minLarge = min(eigsLarge1); minSmall = min(eigsSmall1);

maxEig = max([maxLarge,maxSmall]); minEig = min([minLarge,minSmall]);

subplot(4,1,1)
     plot(xData, eigsLargeOrig, xData, eigsSmallOrig);
     axis([xData(1),xData(end),minEig-1,maxEig+1]);
     title(['Unfiltered Data']); text(730256,2.5,'(a)','FontSize',14);
     legend('Largest Eigenvalue', 'Average 40 Smallest Eigenvalues');
     datetick('x',10,'keeplimits');

clims = [0 100];

subplot(4,1,2)
     plot(xData, eigsLarge1, xData, eigsSmall1);
     axis([xData(1),xData(end),minEig-1,maxEig+1]);
     title(['Wavelet Level 1  (', num2str(round(scale(1,1))), ' days)']);
     text(730256,2.5,'(b)','FontSize',14); datetick('x',10,'keeplimits');

subplot(4,1,3)
     plot(xData, eigsLarge2, xData, eigsSmall2);
     axis([xData(1),xData(end),minEig-1,maxEig+1]);
     title(['Wavelet Level 2  (', num2str(round(scale(1,2))), ' days)']);
     text(730256,2.5,'(c)','FontSize',14); datetick('x',10,'keeplimits');
```

```
subplot(4,1,4)
     plot(xData, eigsLarge3, xData, eigsSmall3);
     axis([xData(1),xData(end),minEig-1,maxEig+1]);
     title(['Wavelet Level 3  (', num2str(round(scale(1,3))), ' days)']);
     text(730256,2.5,'(d)','FontSize',14); datetick('x',10,'keeplimits');
```

APPENDIX C

PAPERS PUBLISHED

A number of papers have been published to date, in the course of research towards this Thesis. These are included in the pages that follow.

Available online at www.sciencedirect.com
ScienceDirect
Physica A 382 (2007) 565–576
PHYSICA A
www.elsevier.com/locate/physa

Random matrix theory and fund of funds portfolio optimisation

T. Conlon*, H.J. Ruskin, M. Crane

Dublin City University, Glasnevin, Dublin 9, Ireland

Received 3 October 2006; received in revised form 11 April 2007
Available online 4 May 2007

Abstract

The proprietary nature of Hedge Fund investing means that it is common practise for managers to release minimal information about their returns. The construction of a fund of hedge funds portfolio requires a correlation matrix which often has to be estimated using a relatively small sample of monthly returns data which induces noise. In this paper, random matrix theory (RMT) is applied to a cross-correlation matrix C, constructed using hedge fund returns data. The analysis reveals a number of eigenvalues that deviate from the spectrum suggested by RMT. The components of the deviating eigenvectors are found to correspond to distinct *groups of strategies* that are applied by hedge fund managers. The inverse participation ratio is used to quantify the number of components that participate in each eigenvector. Finally, the correlation matrix is cleaned by separating the noisy part from the non-noisy part of C. This technique is found to greatly reduce the difference between the predicted and realised risk of a portfolio, leading to an improved risk profile for a fund of hedge funds.

Keywords: Random matrix theory; Hedge funds; Fund of funds; Correlation matrix; Portfolio optimisation

C-2

1. Introduction

A hedge fund is a lightly regulated private investment vehicle that may utilise a wide range of investment strategies and instruments. These funds may use short positions, derivatives, leverage and charge incentive-based fees. Normally, they are structured as limited partnerships or offshore investment companies. Hedge funds pursue positive returns in all markets and hence are described as "absolute return" strategies.

Hedge funds are utilised by pension funds, high net-worth individuals and institutions, due to their low correlation to traditional long-only investment strategies. The incentive-based performance fees, earned by hedge fund managers, align the interest of the hedge fund manager with that of the investor. The performance of hedge funds has been impressive, with the various hedge fund indices providing higher returns, with lower volatility, than traditional assets over many years. As of the end of the first quarter 2006 the total assets managed by hedge funds world wide is estimated at $1.25 trillion [1]. Hedge funds generally only report their returns on a monthly basis and this means that there is a very limited amount of data available to study as databases of hedge fund returns have only been in operation for about 15 years. This is in keeping with the highly secretive, proprietary nature of hedge fund investing. The amount of information reported by a hedge fund about how and where it is producing its returns is often limited to sectoral overviews and strategy allocations. For an introduction to hedge funds see Refs. [2,3].

Significant diversification benefits can be gained by investing in a variety of hedge fund strategies, due to the presence of low and even negative correlations between different hedge fund strategies. Such strategies can be broken up into two general categories: directional and market neutral. Directional strategies, (for example long/short equity, emerging markets, macro and managed futures) have a high risk, high return profile and act as return enhancers to a traditional portfolio. Market neutral strategies, (for example convertible arbitrage, equity market neutral and fixed income arbitrage) deploy a low risk profile and act as a substitute for some proportion of the fixed income holdings in an investors portfolio [2,3].

A fund of hedge funds allows investors to have access to a large diverse portfolio of hedge funds without having to carry out due diligence on each individual manager. The diversification benefits provided by fund of funds are brought about by investing in a number of funds that have a low correlation to each other. These correlations are often calculated by using equally weighted fund returns and can contain a significant amount of noise due to the very small amount of returns data available for hedge funds [3].

In this paper we apply random matrix theory (RMT) to hedge fund returns data with the aim of reducing the levels of noise in these correlation matrices formed from this data and hence constructing a fund of hedge funds with an improved risk profile. Previous studies have used the information found in the RMT defined deviating eigenvalues of a correlation matrix as inputs into a minimum spanning tree [4] to enable characterisation of hedge fund strategies. In this paper, the components of the deviating eigenvectors are shown to correspond to distinct groups of strategies that are applied by hedge fund managers and this is exploited to construct a portfolio with reduced levels of risk.

This paper is organised as follows: in Section 2 we review RMT and discuss its use in the extraction of information from a correlation matrix of hedge fund returns using RMT techniques. In Section 3, we look at the results obtained applying RMT to hedge funds and, in the final section, we draw our conclusions.

2. Methods

2.1. Random matrix theory

Given returns $G_i(t)$, $i = 1, \ldots, N$, of a collection of hedge funds we define a normalised return in order to standardise the different fund volatilities. We normalise G_i with respect to its standard deviation σ_i as follows:

$$g_i(t) = \frac{G_i(t) - \widehat{G_i(t)}}{\sigma_i}, \tag{1}$$

where σ_i is the standard deviation of G_i for assets $i = 1, \ldots, N$ and $\widehat{G_i}$ is the time average of G_i over the period studied.

Then the equal time cross-correlation matrix is expressed in terms of $g_i(t)$

$$C_{ij} \equiv \langle g_i(t) g_j(t) \rangle. \tag{2}$$

The elements of C_{ij} are limited to the domain $-1 \leqslant C_{ij} \leqslant 1$, where $C_{ij} = 1$ defines perfect correlation between funds, $C_{ij} = -1$ corresponds to perfect anti-correlation and $C_{ij} = 0$ corresponds to uncorrelated funds. In matrix notation, the correlation matrix can be expressed as

$$\mathbf{C} = \frac{1}{T}\mathbf{G}\mathbf{G}^t, \tag{3}$$

where $\mathbf{G}$ is an $N \times T$ matrix with elements g_{it}.

*Corresponding author.
E-mail addresses: tconlon@computing.dcu.ie (T. Conlon), hruskin@computing.dcu.ie (H.J. Ruskin), mcrane@computing.dcu.ie (M. Crane).

0378-4371/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2007.04.039

The spectral properties of **C** may be compared to those of a "random" Wishart correlation matrix [8,11],

$$\mathbf{R} = \frac{1}{T}\mathbf{A}\mathbf{A}^t, \tag{4}$$

where **A** is an $N \times T$ matrix with each element random with zero mean and unit variance. Statistical properties of random matrices have been known for many years in the physics literature [5] and have been applied to financial problems relatively recently [6–16].

In particular, the limiting property for the sample size $N \to \infty$ and sample length $T \to \infty$, providing that $Q = T/N \geqslant 1$ is fixed, has been examined to show analytically that the distribution of eigenvalues λ of the random correlation matrix **R** is given by

$$P_{rm}(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{5}$$

for λ within the region $\lambda_- \leqslant \lambda_l \leqslant \lambda_+$, where λ_- and λ_+ are given by

$$\lambda_\pm = \sigma^2\left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right), \tag{6}$$

where σ^2 is the variance of the elements of **G** (for **G** normalised this is equal to unity).

$\lambda_\pm$ are the bounds of the theoretical eigenvalue distribution. Eigenvalues that are outside this region are said to deviate from RMT. Hence by comparing the empirical distribution of the eigenvalues of the funds correlation matrix to the distribution for a random matrix as given in Eq. (5), we can identify the deviating eigenvalues. These deviating eigenvalues are said to contain information about the system under consideration and we use eigenvector analysis to identify specifically the information present.

2.2. Eigenvector analysis

The deviations of $P(\lambda)$ from the RMT result $P_{rm}(\lambda)$ implies that these deviations should also be displayed in the statistics of the corresponding eigenvector components [8]. In order to interpret the meaning of the deviating eigenvectors, we note that the largest eigenvalue is of an order of magnitude larger than the others, which constrains the remaining $N-1$ eigenvalues since $\mathrm{Tr}[C] = N$. Thus, in order to analyse the contents of the remaining eigenvectors, we first remove the effect of the largest eigenvalue. To do this we use the linear regression (as discussed in Ref. [11])

$$G_i(t) = \alpha_i + \beta_i G^{large}(t) + \varepsilon_i(t), \tag{7}$$

where $G^{large} = \sum_1^N u_i^{large} G_i(t)$ and $N = 49$ is the number of funds in our sample. Here u_i^{large} corresponds to the components of the largest eigenvector. We then recalculate the correlation matrix C using the residuals $\varepsilon_i(t)$. If we quantify the variance of the part not explained by the largest eigenvalue as $\sigma^2 = 1 - \lambda_{large}/n$, [8], we can use this value to recalculate our values of $\lambda_\pm$.

Using techniques for sector identification, [10], we try to analyse the information contained in the eigenvectors. We partition the funds into groups $l = 1, \ldots, 10$ as defined by the managers strategy. (e.g., equity long/short, managed futures. See Appendix A for a complete strategy breakdown for the sample.) We define a projection matrix $P_{li} = 1/n_l$, if fund i belongs to group l and $P_{li} = 0$ otherwise. For each deviating eigenvector u^k we compute the contribution $X_l^k \equiv \sum_{i=1}^N P_{li}^k \left[u_i^k\right]^2$ of each strategy group, where this represents the product of the projection matrix and the square of the eigenvector components. This allows us to measure the contribution of the different hedge fund strategies to each of the eigenvectors.

As suggested in Ref. [11], we also aim to assess how the effects of randomness lessen as we move further from the RMT upper boundary, λ_+. To do this we use the inverse participation ratio (IPR). The IPR allows quantification of the number of components that participate significantly in each eigenvector and tells us more

about the level and nature of deviation from RMT. The IPR of the eigenvector u^k is given by $I^k \equiv \sum_{l=1}^N \left(u_l^k\right)^4$ and allows us to compute the inverse of the number of eigenvector components that contribute significantly to each eigenvector.

2.3. Application to portfolio optimisation

The diversification of an investment into independently fluctuating assets reduces its risk. However, since cross-correlations between asset prices exist, the accurate calculation of the cross-correlation matrix is vitally important. The return on a portfolio with N assets is given by $\Phi = \sum_{i=1}^N w_i G_i$ where $G_i(t)$ is the return on asset i, w_i is the fraction of wealth invested in asset i and $\sum_{i=1}^N w_i = 1$. The risk of holding this portfolio is then given by

$$\Omega^2 = \sum_{i=1}^N \sum_{j=1}^N w_i w_j C_{ij} \sigma_i \sigma_j, \tag{8}$$

where σ_i is the volatility of G_i and C_{ij} are the elements of the cross-correlation matrix. In order to find an optimal portfolio, using the Markowitz theory of portfolio optimisation [6,17,18], we minimise Ω^2 under the constraint that the return on the portfolio, Φ, is some fixed value. This minimisation can be implemented by using two Lagrange multipliers, which leads to a set of N linear equations which can be solved for w_i. If we minimise Ω for a number of different values of Φ, then we can obtain a region bounded by an upward-sloping curve, called the *efficient frontier*.

In the case of optimisation with a portfolio containing only hedge funds, the additional constraint of no short-selling is natural due to the difficulties in short-selling of funds (note that short-selling may be achievable by the use of swaps but is uncommon) [2,3].

In order to demonstrate the effects of randomness on the cross-correlation matrix of hedge funds we first divide the time series studied into two equal parts. We assume that we have *perfect knowledge* on the future average returns m_i by taking the observed returns on the second sub-period. We calculate:

(1) the predicted efficient frontier using the correlation matrix for the first sub-period and the expected returns m_i;
(2) the realised efficient frontier using the correlation matrix for the second sub-period and the expected returns m_i.

The portfolio risk due to the noise in the correlation matrix can then be calculated using

$$R_p = \frac{\Omega_r^2 - \Omega_p^2}{\Omega_p^2}, \tag{9}$$

where Ω_r^2 is the risk of the realised portfolio and Ω_p^2 is the risk of the predicted portfolio.

It was shown in [13,14] that correlations may also be measured in the random part of the eigenvalue spectrum. However, since our aim here is to demonstrate how RMT can be used to improve the risk/return profile for a portfolio of hedge funds, we assume that the eigenvalues corresponding to the noise band in RMT, $\lambda_- \leqslant \lambda \leqslant \lambda_+$, are not expected to correspond to real information following [6–12,15,16]. We then use this assumption to remove some of the noise from the correlation matrix. Although the technique used in [6,8] has been shown to lead to problems with the *stability* of the correlation matrix [16], we apply it here as a simple test case to demonstrate how noise can be removed from the cross-correlation matrix formed from hedge fund returns. This technique involves separating the noisy and non-noisy eigenvalues and keeping the non-noisy eigenvalues the same. The noisy eigenvalues are then replaced by their average and the correlation matrix is reconstructed from the cleaned eigenvalues. We can then compare the risk of both the predicted and realised portfolios for the original and cleaned correlation matrices.

3. Results

3.1. Equally weighted correlation matrix

The data set studied here is a collection of 49 hedge funds with varying strategies over a synchronous period from January 1997 to September 2005 ($T = 105$). The original data set was much larger (approximately 1500 funds) but since the length of data available was much less than the number of funds we were forced to choose a subset. The subset chosen were the 49 funds with the longest track records giving us a fund to data ratio $Q = 2.143$. Reducing the data set in this way is not unrealistic as a typical fund of hedge funds would monitor a subset of funds and choose a portfolio from these [2,3]. Often one of the criteria used in choosing this subset of investable funds would be the completion of a minimum track record. Other data sets, such as a portfolio made up of hedge fund strategy indices, were also studied by the authors with notably similar results.

First we calculated the empirical correlation matrix between the funds using equally weighted returns and from this found the spectrum of eigenvalues. This is then compared with the theoretical spectrum for random Wishart matrices (as per Eq. (5)), using $\lambda_- = 0.1$ and $\lambda_+ = 2.83$. As can be seen from Fig. 1, the bulk of the eigenvalues conform to those of the random matrix. There are three deviating eigenvalues, at 10.9886, 8.2898 and 2.944. This means that 6.1% of eigenvalues deviate from the RMT prediction which is consistent with the findings of Ref. [8], where the authors argue that at most 6% of eigenvalues are non-noisy.

3.2. Bootstrapping

In order to show that there was no dependence on the choice of time period or the length of the series we broke the time series up into two segments and again compared the eigenvalue spectrum of the correlation matrix with that of a random Wishart matrix. For both time periods studied we found just two eigenvalues that deviated from the RMT prediction. As can be seen in Fig. 2, the anomalous eigenvalue contributions are very similar for both periods chosen, which implies independence from the choice of time period and stationarity of the data. The values of the deviating eigenvalues are shown in Table 1.

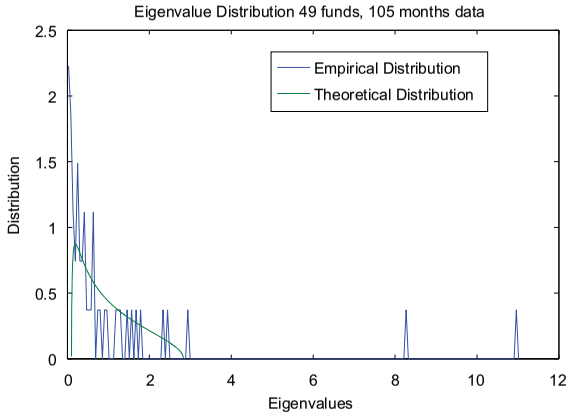


Fig. 1. Spectral distribution for equally weighted hedge fund correlation matrix.

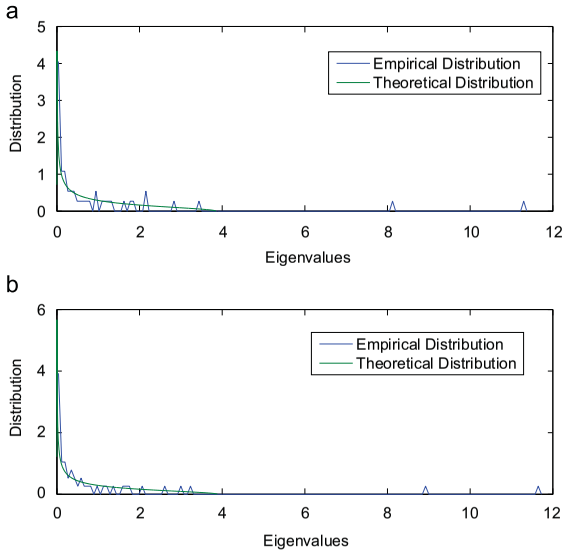


Fig. 2. Bootstrapped spectral distribution for consecutive periods.

Table 1
Deviating eigenvalues

Eigenvalue rank	105 months returns	First 53 months	Second 52 months
1	10.9886	11.334	11.6874
2	8.2898	8.1375	8.9312
3	2.944		

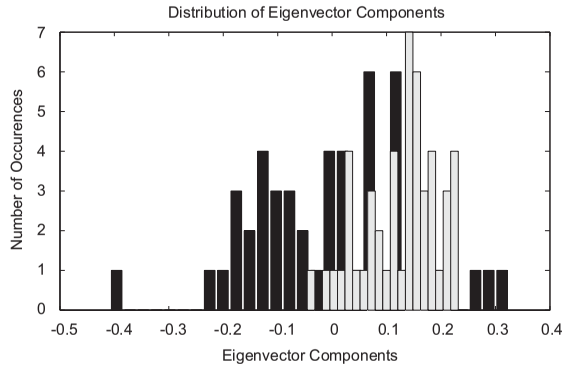


Fig. 3. Comparison of eigenvector components, largest eigenvector (grey), eigenvector from the bulk (black).

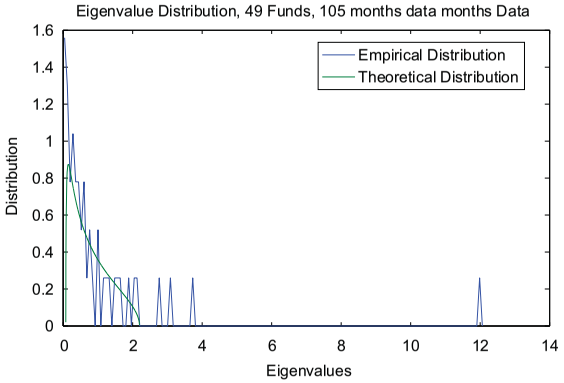


Fig. 4. Eigenvalue spectrum after the removal of the effects of the largest eigenvalue.

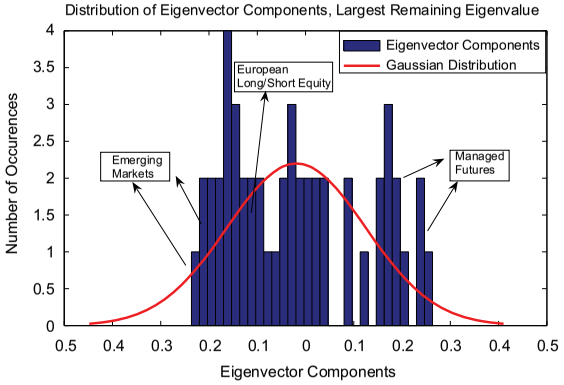


Fig. 5. Distribution of eigenvector components, largest remaining eigenvalue.

3.3. Eigenvector analysis

Fig. 3 shows the distribution of the components of the largest eigenvector and also the components of a typical eigenvector from the region predicted by RMT. As can be seen from this graph, the distribution of the components of the largest eigenvector are significantly different from that of an eigenvector chosen from the random region. The average value is much larger and the variance of the components much smaller than for the largest eigenvector, which is in agreement with Refs. [11,16] (the largest eigenvector can be interpreted as the 'market'). In this case the 'market' is the set of external stimuli that affect most hedge funds. (e.g., Interest rate changes, large market (i.e., S&P 500, etc.) moves, margin changes, etc.)

We then remove the effects of the largest eigenvalue using the techniques described in Section 2.2. This changes the value of λ_{max} from 2.8329 to 2.1975 which means that 4 of the remaining largest eigenvalues are now outside the RMT region (Fig. 4).

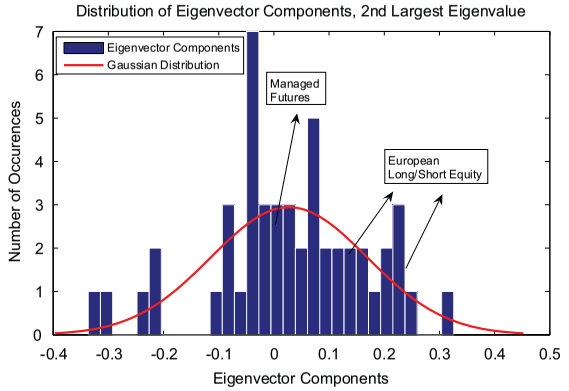


Fig. 6. Distribution of eigenvector components, second largest remaining eigenvalue.

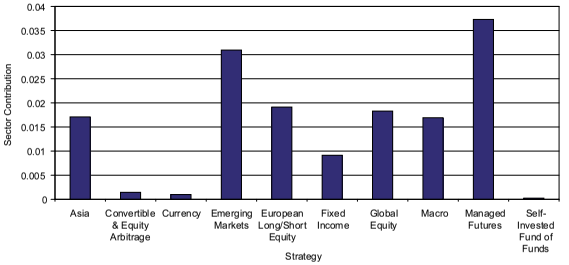


Fig. 7. Strategy contribution, largest eigenvalue.

The distribution of the components of largest remaining deviating eigenvector shows some distinctive clustering (Fig. 5). In particular, the managed futures, emerging markets and European long/short equity strategies are the major contributors here.

A similarly clustered distribution also emerges for the other deviating eigenvalues. An analysis of the eigenvector components for the second largest remaining eigenvector, after removing the effects of the market eigenvector (Fig. 6), shows distinctive clusters emerging, especially for the managed futures and long/short equity sectors. However, the components corresponding to the managed futures strategy do not deviate much from zero and hence make little contribution. These findings for clustering of the deviating eigenvalues agree with Ref. [19] where the authors show that, in addition to the largest, the other deviating eigenvalues of the correlation matrix of asset returns also contain information about the risk associated with the assets.

3.4. Strategy identification

From the analysis described in Section 3.3, we look at the contribution, $X_l^k \equiv \sum_{i=1}^{N} P_{li}^k [u_i^k]^2$, to each of the deviating eigenvalues from the different strategies employed. Fig. 7 shows X_l^k for the largest remaining

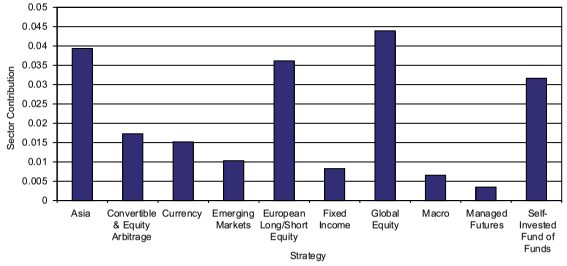


Fig. 8. Strategy contribution, second largest eigenvalue.

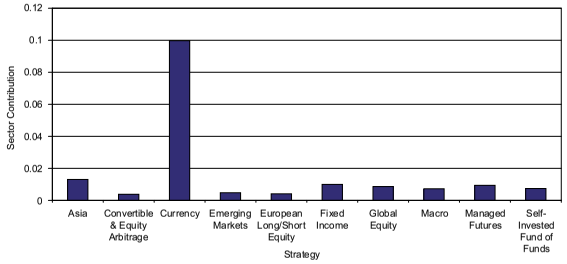


Fig. 9. Strategy contribution, third largest eigenvalue.

eigenvector once the effects of the market eigenvalue are removed. The largest contributors are clearly managed futures and emerging markets. However, the strategy contribution for managed futures is only around twice that for many of the other sectors. Hence managed futures and emerging markets are the dominant strategies but care in interpretation is needed, since neither contributor is dominant overall.

Fig. 8 shows the strategy contributions for the second largest eigenvalue. These are interesting, since three of the four dominant strategies (Asia, global equity & European long/short equity) are *equity strategies* and would all be affected by events in world equity markets. Also the fourth strategy, self-invested fund of funds, may well also consist of equity funds. However, there is limited information available on exactly what type of funds the managers were invested, although there is reason to believe that a majority of them would be equity based. This implies that this eigenvalue seems to contain information just on equity funds.

Fig. 9 contains the strategy contributions for the third largest remaining eigenvalue. Clearly the dominant strategy here is currency. The final deviating eigenvalue also has one dominant strategy which is self-invested fund of funds. It is notable in the above that analysis of the eigenvalues from within the random matrix region revealed no dominant strategies. The evidence of strategy information in the deviating eigenvalues, coupled with a lack of dominant strategies within the RMT region, supports the idea that information in the correlation matrix is chiefly contained within the deviating eigenvalues. We show how this information can be used to create a portfolio with an improved risk–return profile in Section 3.6.

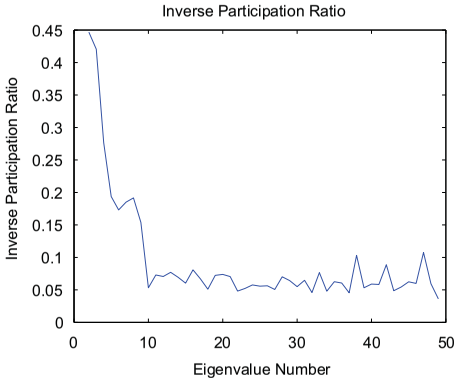


Fig. 10. Inverse participation ratio.

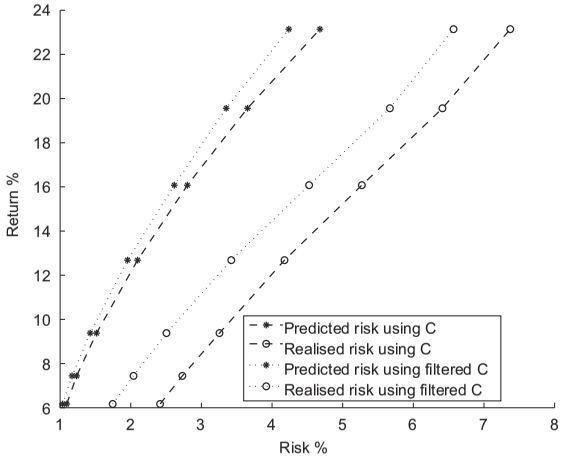


Fig. 11. Efficient frontiers using original and cleaned correlation matrices.

3.5. Inverse participation ratio

Fig. 10 shows the IPR calculated for the eigenvectors of the hedge fund cross-correlation matrix studied. The average IPR value is around 0.06, larger than would be reasonably expected ($1/N \approx 0.02$) if all components contributed to each eigenvector. We would also expect that the largest eigenvectors contributed much more markedly. However, they appear to have a similar IPR to eigenvectors corresponding to the random section. Part of the reason may be that the sample size is small, so the IPR is not particularly effective in terms of assessing by how much larger eigenvectors deviate from the random region, since the average value of the IPR relies on a sample size that tends to infinity.

A significant deviation from the average IPR value is found for the first few eigenvalues. This seems to have been caused by the inclusion of four or five funds, identical apart from being either in different base currencies or leveraged versions of each other with correspondingly high correlation (≈ 1). As mentioned in Ref. [11] funds with a correlation coefficient much greater than the average effectively decouple from the other funds.

3.6. Noise removal and portfolio optimisation

It was noted earlier that where the time series available to estimate cross-correlation matrices are of finite length this leads to measurement noise. This problem is particularly prevalent with hedge fund data, since only monthly returns are available. As seen in Fig. 11 the realised risk is, on average, 292% of the predicted risk. This has obvious consequences for risk management. However, by cleaning the correlation matrix as described in Section 2.3, the realised risk is 190% of the predicted risk. This huge improvement is brought about by limiting the correlation matrix to the information band prescribed by RMT.

It can be seen in Fig. 11 that, for some return values, the predicted risk (using the filtered correlation matrix) is actually less than that of the original correlation matrix. This is due to the constraints imposed on the portfolio, in particular the restriction of no 'short-selling' (Section 2.3).

The use of the cleaned correlation matrix leads to a 35% improvement in the difference between the realised risk and the predicted risk for the optimal portfolio. So we have shown that, in the case of hedge fund returns data, the cleaned correlation matrix is a good choice for portfolio optimisation.

4. Conclusions

We have illustrated that, even with limited data (105 months of returns data for 49 hedge funds), useful information can be extracted from a cross-correlation matrix constructed from hedge fund returns. Significant deviations from random matrix theory predictions are observed, with further analysis showing that there is real strategy information contained within the deviating eigenvalues. Eigenvector analysis revealed distinct *strategy clustering* in the deviating eigenvectors. These included emerging markets and managed futures in the largest eigenvector, equity funds in the second, currency and fund of funds in the final two deviating eigenvectors. The strategy information in the deviating eigenvalues was then used to clean the correlation matrix, by flattening the eigenvalues from the bulk to their average and holding the deviating eigenvalues the same. A 35% improvement between the risk of the predicted and realised portfolios was found using this filtering technique.

Appendix A. Hedge fund strategies

Strategies employed by the managers in the sample considered are given in Table A1.

Table A1
Hedge fund strategies

Strategies	Number of funds
Asia excluding Japan long/short equities	2
Convertible and equity arbitrage	2
Currency	7
Emerging markets	6
European long/short equity	10
Fixed income	1
Global equity	5
Japan market neutral	1
Macro	3
Managed futures	11
Self-invested fund of funds	1

References

[1] The Barclay Group, ⟨www.barclaygrp.com⟩, accessed on 1st August 2006.
[2] F.S. L'habitant, Hedge Funds, Wiley, New York, 2002.
[3] F.S. L'habitant, Hedge Funds—Quantitative Insights, Wiley, New York, 2004.
[4] M.A. Miceli, G. Susinno, Ultrametricity in fund of funds diversification, Physica A 344 (2004) 94–99.
[5] M.L. Mehta, Random Matrices, third ed., Elsevier, Amsterdam, 2004.
[6] J.P. Bouchaud, M. Potters, Theory of Financial Risk and Derivative Pricing, Cambridge University Press, Cambridge, 2003.
[7] L. Laloux, P. Cizeau, J. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (7) (1999) 1467–1470.
[8] L. Laloux, P. Cizeau, M. Potters, J. Bouchaud, Random matrix theory and financial correlations, Int. J. Theor. Appl. Finance 3 (3) (2000) 391–397.
[9] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, H.E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Phys. Rev. Lett. 83 (7) (1999) 1471–1474.
[10] P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, Quantifying and interpreting collective behaviour in financial markets, Phys. Rev. E 64 (2001) 035106R.
[11] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, T. Guhr, H.E. Stanley, A random matrix approach to cross-correlations in financial data, Phys. Rev. E 65 (2002) 066126.
[12] B. Rosenow, V. Plerou, P. Gopikrishnan, H.E. Stanley, Portfolio optimization and the random magnet problem, Europhys. Lett. 59 (2002) 500–506.
[13] Z. Burda, A. Görlich, A. Jarosz, J. Jurkiewicz, Signal and noise in correlation matrix, Physica A 343 (2004) 295–310.
[14] Z. Burda, J. Jurkiewicz, Signal and noise in financial correlation matrices, Physica A (1–2) (2004) 295–310.
[15] D. Wilcox, T. Gebbie, On the analysis of cross-correlations in South African market data, Physica A 344 (1–2) (2004) 294–298.
[16] S. Sharifi, M. Crane, A. Shamaie, H.J. Ruskin, Random matrix theory for portfolio optimization: a stability approach, Physica A 335 (3–4) (2004) 629–643.
[17] H. Markowitz, Portfolio Selection: Efficient Diversification of Investments, Wiley, New York, 1959.
[18] E.J. Elton, M.J. Gruber, Modern Portfolio Theory and Investment Analysis, sixth ed., Wiley, New York, 2002.
[19] A. Sharkasi, M. Crane, H.J. Ruskin, J.A. Matos, The reaction of stock markets to crashes and events: a comparison study between emerging and mature markets using wavelet transforms, Physica A 368 (2) (2006) 511–521.

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Wavelet multiscale analysis for Hedge Funds: Scaling and strategies

T. Conlon *, M. Crane, H.J. Ruskin

Dublin City University, Glasnevin, Dublin 9, Ireland

ARTICLE INFO

Article history:
Received 10 April 2008
Available online 5 June 2008

PACS:
89.65.Gh
89.65.-s
89.75.-k

Keywords:
Scaling
Cross-correlation
Wavelets
Hedge Funds
Econophysics

ABSTRACT

The wide acceptance of Hedge Funds by Institutional Investors and Pension Funds has led to an explosive growth in assets under management. These investors are drawn to Hedge Funds due to the seemingly low correlation with traditional investments and the attractive returns. The correlations and market risk (the *Beta* in the Capital Asset Pricing Model) of Hedge Funds are generally calculated using monthly returns data, which may produce misleading results as Hedge Funds often hold illiquid exchange-traded securities or difficult to price over-the-counter securities. In this paper, the Maximum Overlap Discrete Wavelet Transform (MODWT) is applied to measure the scaling properties of Hedge Fund correlation and market risk with respect to the S&P 500. It is found that the level of correlation and market risk varies greatly according to the strategy studied and the time scale examined. Finally, the effects of scaling properties on the risk profile of a portfolio made up of Hedge Funds is studied using correlation matrices calculated over different time horizons.

1. Introduction

A Hedge Fund is a lightly-regulated private investment vehicle that may utilise a wide range of investment strategies and instruments. These funds may use short positions, derivatives, leverage and charge incentive-based fees. Normally, they are structured as limited partnerships or offshore investment companies. Hedge Funds pursue positive returns in all markets and hence are described as "*Absolute Return*" strategies.

Hedge Funds are utilised by pension funds, high net-worth individuals and institutions, partly due to their low correlation with traditional long-only investment strategies. The incentive-based performance fees, earned by hedge fund managers, align the interest of the hedge fund manager with that of the investor. The performance of Hedge Funds has been impressive, with the various Hedge Fund indices providing higher returns, with lower volatility, than traditional assets over many years. As of the end of the fourth quarter of 2007, the total assets managed by hedge funds world-wide is estimated at $2.065 trillion [1].

In addition to their attractive returns, many hedge funds claim to provide significant diversification benefits. Moderate correlations between Hedge Funds and traditional asset classes, using monthly returns data, have been reported [2,3]. Hence, the market exposure of Hedge Funds to traditional assets classes has also been found to be low. However, these results were based upon monthly returns data and may be misleading. Hedge Funds may hold illiquid exchange traded assets or over the counter (OTC) securities. These may be priced using the last traded price, which may not have traded at or even near the end of the month, resulting in non-synchronous pricing. Furthermore, OTC securities may not have publicly available prices but instead rely on broker prices. Due to the large performance fees charged by hedge funds, using stale or non-synchronous data may be of great benefit to the manager, at the expense of the investor.

Risk control techniques for portfolios of hedge funds have been studied by various authors. The threshold point of self-organised criticality was used as a control parameter of risk exposure [4] for a portfolio of Hedge Funds. The correlation matrix between the returns of Hedge Funds was used [5] to characterise Hedge Fund strategies using Random Matrix Theory (RMT) and a minimum spanning tree. Random Matrix Theory was applied to the correlation matrix between Hedge Funds [6] where the components of the eigenvalues deviating from RMT were shown to correspond to distinct groups of strategies that are applied by Hedge Fund managers. The information in these deviating eigenvalues was then applied to construct a portfolio with reduced levels of risk. A downside-risk[1] framework was used [7] for the study of Hedge Fund risk, taking into account the inherent asymmetry in Hedge Fund returns distributions.

Multiscale analysis of the covariance between time series has been applied to the physical sciences in numerous cases (eg. Refs. [8–10]). Wavelet multiscale analysis has also been used by various authors to decompose economic and financial time series into orthogonal time-scale components of varying granularities. Wavelet decomposition was used [11] to analyse economic relationships in order to investigate anomalies that were previously unexplained. The scaling properties of foreign exchange volatility were examined [12] using wavelet multiscaling techniques where it was shown that the correlation between foreign exchange volatilities increases for scales from 20 min to daily. Wavelet multiscaling was applied as a filtering technique [13] to remove intraday seasonalities from data in the estimation of volatility models. The systematic risk in a Capital Asset Pricing Model was estimated over different granularities [14] where it was shown that the return of a portfolio and its beta becomes stronger as the scale increases for the S&P 500. This technique was then applied to the markets of various other countries [15] where similar results were also found. The dependence of stock return cross-correlations on the sampling time scale is known as the *Epps* effect [16]. This decrease of correlations between stock returns for shorter time scales was studied [17] where it was shown that asynchronicity between stock returns was not solely responsible for the *Epps* effect.

In this paper we use wavelet multiscaling techniques to show that the correlation between Hedge Fund strategies and the market, calculated using monthly returns data, can be significantly different to that calculated using longer time horizons from 2–4 months to 16–32 months. Also, the systematic market risk of different Hedge Fund strategies is different using monthly data to that calculated over longer horizons. We then show, by calculating an efficient frontier (see Section 2.1) at each time scale, the effect of scaling on the risk profile of a portfolio of Hedge Funds.

This paper is organised as follows. In the next section, we present the main results of the Capital Asset Pricing Model and Portfolio Optimisation. Section 3 introduces the Maximum Overlap Discrete Wavelet transfer and the calculation of correlation and variance at different scales. The Hedge Fund data used is described in Sections 4 and 5 shows the empirical results obtained. Section 6 concludes the paper and gives some directions for further research.

2. The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) was developed [18], building on earlier work [19], as a model for pricing an individual asset or a portfolio. The derivation of the standard CAPM is based upon a number of assumptions such as risk-averse investors, infinite divisibility of financial assets, absence of personal income tax, unlimited short sales, unlimited borrowing and lending at the risk free rate together with the axiom that all assets are marketable (see Ref. [20]).

The CAPM states that the expected return of a security or a portfolio equals the rate on a risk-free security plus a risk premium. The latter depends linearly on the market risk exposure (i.e. the Beta of an asset, Eq. (2)) and the market risk premium. Hence, the expected return on a risky asset is given by:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] \tag{1}$$

where $E(R_i)$ is the expected return on the risky asset, $E(R_m)$ is the expected return of the market, R_f denotes the risk free rate and β_i is the Beta of the portfolio with respect to the market portfolio M (in excess of the risk free rate return) and is given by

$$\beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)}. \tag{2}$$

Eq. (1) can be rewritten in terms of the risk premium by simply subtracting the risk-free rate from both sides of the equation giving,

$$E(R_i) - R_f = \beta_i[E(R_m) - R_f]. \tag{3}$$

The empirical version of Eq. (3) is given by

$$R_i - R_f = \alpha_i + \beta_i[R_m - R_f] + \epsilon_i \tag{4}$$

where α_i is the expected firm-specific return and ϵ_i is a random error term. For a more detailed treatise of the CAPM see Refs. [14,15,18,20,21].

* Corresponding author. Tel.: +353 863718053.
E-mail addresses: tconlon@computing.dcu.ie (T. Conlon), mcrane@computing.dcu.ie (M. Crane), hruskin@computing.dcu.ie (H.J. Ruskin).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.05.046

[1] Downside-risk measures look at how much money an investor stands to lose during adverse market conditions.

2.1. Portfolio optimisation

The diversification of an investment into independently fluctuating assets reduces its risk (see Refs. [18,20,21]). However, since cross-correlations between asset prices exist, accurate calculation of the cross-correlation matrix is vitally important. The expected return on a portfolio with N assets is given by

$$\Phi = w_0 R_f + \sum_{i=1}^{N} w_i R_i \tag{5}$$

where $R_i(t)$ is the expected return on asset i at time t, w_0 is the fraction of wealth invested in the risk-free security R_f, w_i is the fraction of wealth invested in asset i and $\sum_{i=0}^{N} w_i = 1$. The risk associated with holding this portfolio is then given by

$$\Omega^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j C_{ij} \sigma_i \sigma_j \tag{6}$$

where σ_i is the volatility of R_i and C_{ij} are the elements of the cross-correlation matrix. In order to find an optimal portfolio, using the Markowitz theory of portfolio optimisation [19–21], we minimise Ω^2 under the constraint that the return on the portfolio, Φ, is some fixed value. This minimisation can be implemented by using two Lagrange multipliers and leads to a set of N linear equations which can be solved for w_i. If we minimise Ω^2 for a number of different values of Φ, then we can obtain a region bounded by an upward-sloping curve, called the *efficient frontier* with the portfolio risk on the x-axis and the expected portfolio return on the y-axis.

3. Wavelet analysis

The term Wavelet was coined by Morlet and Grossman in the 1980's, some of the earliest researchers in the field. The word "Wavelet" derives from the French *Ondelette* meaning "Small Wave". These "Small Waves" provide an efficient means of studying the multiresolution properties of a signal, as they can be used to decompose a signal into different time horizons. Wavelet filters have been applied to many different applications such as meteorology, image analysis, signal processing and financial time series [8–15,22–24].

The Discrete Wavelet Transform (DWT) [25–28], in particular, allows the decomposition of a signal into components of different frequency. There are two basic wavelet functions, the father wavelet ϕ and the mother wavelet ψ. The formal definition of the father and mother wavelets are the functions:

$$\phi_{j,k}(t) = 2^{\frac{j}{2}} \phi\left(2^j t - k\right) \tag{7}$$

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi\left(2^j t - k\right) \tag{8}$$

where $j = 1, \ldots, J$ in a J-level decomposition. The father wavelet integrates to 1 and reconstructs the longest time-scale component of the series, whereas the mother wavelet integrates to 0 and is used to describe the deviations from the trend. The wavelet representation of a discrete signal $f(t)$ in $L^2(R)$ is given by:

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\phi_{J,k}(t) + \cdots + \sum_k d_{1,k}\phi_{1,k}(t) \tag{9}$$

where J is the number of multiresolution levels (or scales) and k ranges from 1 to the number of coefficients in the specified level. The coefficients $s_{J,k}$ and $d_{J,k}$ are the smooth and detail component coefficients respectively and are given by

$$s_{J,k} = \int \phi_{J,k} f(t) \mathrm{d}t \tag{10}$$

$$d_{j,k} = \int \psi_{j,k} f(t) \mathrm{d}t \quad (j = 1, \ldots, J). \tag{11}$$

Each of the sets of the coefficients $S_J, d_J, d_{J-1}, \ldots, d_1$ is called a crystal.

The MODWT [8,12,22,28] is a linear filter that transforms a series into coefficients related to variations over a set of scales. Like the Discrete Wavelet Transform (DWT) it produces a set of time-dependent wavelet and scaling coefficients with basis vectors associated with a location t and a unitless scale $\tau_j = 2^{j-1}$ for each decomposition level $j = 1, \ldots, J_0$. The MODWT has, unlike the DWT, a high level of redundancy and is nonorthogonal. It retains downsampled[2] values at each level of the

[2] Downsampling or decimation of the wavelet coefficients retains half of the number of coefficients that were retained at the previous scale. Downsampling is applied in the Discreate Wavelet Transform.

decomposition that would be discarded by the DWT. The MODWT can also handle any sample size N, whereas the DWT restricts the sample size to a multiple of 2^j.

Decomposing a signal using the MODWT to J levels theoretically involves the application of J pairs of filters. The filtering operation at the jth level consists of applying a rescaled father wavelet to yield a set of detail coefficients

$$\tilde{D}_{j,t} = \sum_{L_j-1}^{l=0} \tilde{\psi}_{j,l} f_{t-l} \tag{12}$$

and a rescaled mother wavelet to yield a set of scaling coefficients

$$\tilde{S}_{j,t} = \sum_{L_j-1}^{l=0} \tilde{\phi}_{j,l} f_{t-l} \tag{13}$$

for all times $t = \ldots, -1, 0, 1, \ldots$, where f is the function to be decomposed [28]. The rescaled mother $\tilde{\psi}_{j,l} = \frac{\psi_{j,l}}{2^j}$ and father $\tilde{\phi}_{j,t} = \frac{\phi_{j,l}}{2^j}$ wavelets for the jth level are a set of scale-dependent localised differencing and averaging operators and can be regarded as rescaled versions of the originals. The jth level equivalent filter coefficients have a width $L_j = (2^j - 1)(L-1) + 1$, where L is the width of the $j = 1$ base filter. In practice the filters for $j > 1$ are not explicitly constructed because the detail and scaling coefficients can be calculated using an algorithm that involves the $j = 1$ filters operating on the jth level scaling coefficients to generate the $j + 1$ level scaling and detail coefficients [28].

3.1. Wavelet variance

The wavelet variance $\nu_f^2(\tau_j)$ is defined as the expected value of $\tilde{D}_{j,t}^2$ if we consider only the non-boundary coefficients. An *unbiased* estimator of the wavelet variance is formed by removing all the coefficients that are affected by the boundary conditions[3] and is given by:

$$\nu_f^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^2 \tag{14}$$

where $M_j = N - L_j + 1$ is the number of non-boundary coefficients at the jth level [28]. The wavelet variance decomposes the variance of a process on a scale by scale basis and allows us to explore how a signal behaves over different time horizons.

3.2. Wavelet covariance and correlation

The wavelet covariance between functions $f(t)$ and $g(t)$ is similarly defined to be the covariance of the wavelet coefficients at a given scale. The *unbiased* estimator of the wavelet covariance at the jth scale is given by

$$\nu_{fg}(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^{f(t)} \tilde{D}_{j,t}^{g(t)} \tag{15}$$

where all the wavelet coefficients affected by the boundary are removed [28] and $M_j = N - L_j + 1$.

The MODWT estimate of the wavelet cross correlation between functions $f(t)$ and $g(t)$ may then be calculated using the wavelet covariance and the square root of the wavelet variance of the functions at each scale j. The MODWT estimator [22] of the wavelet correlation is then given by:

$$\rho_{fg}(\tau_j) = \frac{\nu_{fg}(\tau_j)}{\nu_f(\tau_j)\nu_g(\tau_j)}. \tag{16}$$

In the CAPM model, the wavelet Beta estimator, at scale j, is defined as

$$\beta_{fm}(\tau_j) = \frac{\nu_{fm}(\tau_j)}{\nu_m^2(\tau_j)} \tag{17}$$

where $\nu_{fm}(\tau_j)$ is the covariance between an asset f and the market m, and $\nu_m(\tau_j)$ is the variance of the return on the market portfolio at scale j.

[3] The MODWT treats the time series as if it were periodic using "circular boundary conditions". There are L_j wavelet and scaling coefficients that are influenced by the extension, which are referred to as the boundary coefficients.

Table 1
Correlations between the S&P 500 and the Credit Suisse/Tremont Hedge Fund indices

Scale	Original data	1	2	3	4
Hedge fund index	0.47	0.50	0.40	0.50	0.36
Convertible arbitrage	0.13	0.07	0.12	0.35	0.31
Dedicated short bias	−0.75	−0.78	−0.76	−0.80	−0.88
Emerging markets	0.48	0.59	0.53	0.47	0.39
Equity market Neutral	0.37	0.43	0.41	0.30	0.13
Event driven	0.55	0.56	0.58	0.71	0.51
Fixed income arbitrage	0.00	0.05	−0.19	0.08	0.23
Global macro	0.22	0.31	0.16	0.16	0.00
Long/short equity	0.58	0.59	0.54	0.60	0.53
Managed futures	−0.14	0.00	−0.25	−0.19	−0.61
Multi-Strategy	0.10	0.03	0.03	0.16	0.25

4. Data

To enable us to study the market risk and correlations of Hedge Funds at different time scales, we use returns from the Credit Suisse/Tremont indices [29] from April 1994 to October 2006, a total of 151 months. These indices use the Credit Suisse/Tremont database, which tracks over 4500 funds and consists only of funds with a minimum of USD 50 million under management, a 12 month track record and audited financial statements. The indices are calculated and rebalanced on a monthly basis and are net of all performance and management fees.

The funds in the Credit Suisse/Tremont database are separated into 10 primary categories based on their investment style. From this universe, Credit Suisse/Tremont selects a subset of funds for inclusion in the sub-indices such that each sub-index represents at least 85% of the assets under management in that respective category. The sub-indices are Convertible Arbitrage, Dedicated Short Bias, Emerging Markets, Equity Market Neutral, Event-Driven, Fixed Income Arbitrage, Global Macro, Long/Short Equity, Managed Futures and Multi-Strategy. Each of these strategies is distinct both in the instruments used and the types of markets traded and detailed descriptions of each can be found in Refs. [29–31].

The S&P 500 index is chosen to represent the "*Market*". The S&P 500 is a cap-weighted equity index consisting of large publicly held companies and trade on major US stock exchanges such as the New York Stock Exchange and NASDAQ. Monthly returns data for the S&P 500 is available to download publicly on the internet [32].

5. Results

For the present study, we selected the least asymmetric (LA) wavelet (known as the Symmlet) which exhibits near symmetry about the filter midpoint. LA filters are available in even widths and the optimal filter width is dependent on the characteristics of the signal and the length of the data. The filter width chosen for this study was the LA8 (where the 8 refers to the width of the scaling function) since it enabled us to accurately calculate wavelet Betas and Correlations to the 4th scale given the length of data available. Although the MODWT can accommodate any level J_0, in practise the largest level is chosen so as to prevent decomposition at scales longer than the total length of the data series, hence the choice of the 4th scale.

5.1. Correlation analysis

Using wavelets we examined the correlation between the S&P 500 index and the Credit Suisse/Tremont Hedge Fund composite and sub-indices for different time horizons using the MODWT as described in Section 3. The correlations were calculated for scales 1, 2, 3 and 4 corresponding to a 2–4 month period, a 4–8 month period, a 8–16 month period and a 16–32 month period respectively, as well as for the original monthly returns data. The results are shown in Table 1.

The correlations for the Hedge Fund composite index with the S&P vary according to the time period looked at, in a similar fashion to those of the Long/Short Equity strategy. This may be at least partly accounted for by the fact that Long/Short Equity is the component with the largest weighting in the index (28.8%, [29]). The correlation for Convertible Arbitrage, Fixed Income Arbitrage and Multi-Strategy with the S&P is found to increase as the time horizon increases which may be evidence that these strategies hold less liquid or hard-to-price securities (resulting in a "smoothing" of their returns). The correlations for Dedicated Short Bias, Equity Market Neutral and Managed Futures are found to decrease significantly as the time horizon increases. Hence, these strategies may be good diversifiers over a longer time period and would be a useful addition to a market portfolio.

Using the Credit Suisse/Tremont Hedge Fund Composite index, we also measured the correlation between this and each of the sub-indices over different time horizons. Again, the correlations were calculated for scales 1, 2, 3 and 4 as above. The results are shown in Table 2.

The results for the inter-strategy correlations with the Hedge Fund Composite Index were somewhat similar to those with the S&P 500. The correlation of Convertible Arbitrage, Fixed Income Arbitrage and Multi-Strategy with the Hedge

Table 2
Correlations between the Credit Suisse/Tremont Hedge Fund Composite index and the sub-indices

Scale	Original data	1	2	3	4
Convertible arbitrage	0.40	0.23	0.36	0.53	0.63
Dedicated short bias	−0.48	−0.50	−0.48	−0.50	−0.39
Emerging markets	0.66	0.71	0.65	0.66	0.64
Equity market neutral	0.32	0.36	0.34	0.12	0.07
Event driven	0.68	0.61	0.70	0.72	0.69
Fixed income arbitrage	0.41	0.28	0.33	0.57	0.57
Global macro	0.85	0.88	0.87	0.85	0.79
Long/short equity	0.79	0.76	0.83	0.85	0.82
Managed futures	0.17	0.35	0.22	−0.01	−0.58
Multi-Strategy	0.22	0.01	0.31	0.30	0.50

Table 3
Betas of the Credit Suisse/Tremont Hedge Fund indices

Scale	Original data	1	2	3	4
Hedge fund index	0.25	0.25	0.21	0.31	0.26
Convertible arbitrage	0.04	0.01	0.03	0.16	0.19
Dedicated short bias	−0.90	−0.85	−0.94	−1.22	−1.49
Emerging markets	0.53	0.52	0.59	0.77	0.63
Equity market neutral	0.07	0.07	0.08	0.08	0.04
Event driven	0.21	0.17	0.22	0.38	0.39
Fixed income arbitrage	0.00	0.01	−0.05	0.02	0.08
Global macro	0.16	0.23	0.11	0.12	−0.01
Long/short equity	0.41	0.37	0.39	0.55	0.46
Managed futures	−0.11	0.00	−0.23	−0.22	−0.61
Multi-Strategy	0.03	0.01	0.01	0.05	0.10

Fund composite index was found to increase as the time period increased. Hence, these strategies may not offer such significant diversification benefits in a portfolio of hedge funds over a longer period. The correlation of Dedicated Short Bias, Managed Futures, Equity Market Neutral and Global Macro to the Composite Index decreased significantly as the time horizon increased. Hence these strategies may offer real diversification benefits over a longer time horizon. We examine the effects of the change in correlations at different time horizons on the efficient frontier of a portfolio made up of the Hedge Fund strategies in Section 5.3.

5.2. Systematic risk analysis

We measured the Beta or market risk of various hedge fund strategies to the S&P 500 over different time horizons using the MODWT. Once again, the correlations were calculated for scales 1, 2, 3 and 4 as defined in Section 5.1. The results are displayed in Table 3.

The Beta of the Composite Hedge Fund Index appears to be reasonably static but the market risk of its components varies according to strategy. Convertible Arbitrage has a Beta of 0.04 using monthly data but using a 16–32 month time horizon it has Beta of 0.19. This difference may be due to the fact that there is no exchange for Convertible Bonds and they are traded "over the counter" and hence can be illiquid and difficult to price.

The Beta of the Emerging Markets strategy increases from 0.53 to a maximum of 0.77 over an 8–16 month period. This may be due to liquidity constraints in emerging markets with light markets in equities causing instruments to be marked off non-synchronous prices, since the final traded price for an asset may have been before month end. The effect of liquidity issues in Emerging Markets has been studied in more detail in Ref. [33]. Another possible factor might be the difficulty in short-selling instruments in many emerging markets, hence these funds often employ a more traditional long-only type strategy.

The market risk of the Event Driven strategy is shown in Table 3 to increase from 0.21 to 0.39 over a longer time horizon, as is the case for Equity Long/Short strategy which increases from 0.41 to a maximum of 0.55. In both cases these increases could be due to a number of factors, such as liquidity, the holding period of the manager or the time for the positions in the portfolios to reach a "fair value". Another significant consideration might be the use of *options* which have a non-linear returns profile and hence might lead to an increase in Beta. Furthermore Long/Short Equity managers may also hold significant positions in small capitalization stocks or illiquid private securities.

Equity Market Neutral, Fixed Income Arbitrage and Multi-Strategy all appear to have very small and static levels of market risk over all time horizons. Equity Market Neutral, as the name suggests, seeks to exploit pricing efficiencies between equity securities while simultaneously neutralising exposure to market risk. Fixed Income Arbitrage managers invest solely in Bonds (whether sovereign or corporate) and hence have little or no equity market risk. Multi-Strategy funds implement a

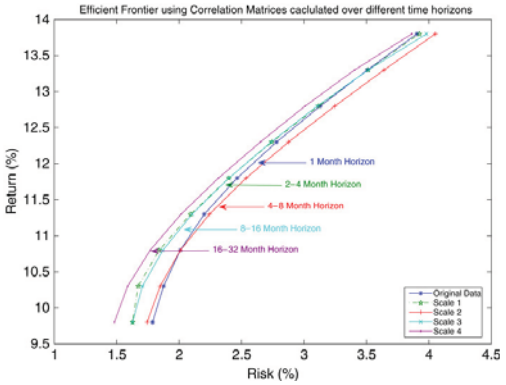


Fig. 1. Efficient frontier using correlation matrices calculated over different time scales.

dynamic strategy allocation to various Hedge Fund strategies. All of these strategies should provide good diversification to a portfolio made up of S&P equities.

The market risk of the Dedicated Short Bias and Managed Futures strategies is found to decrease by a considerable amount as the time horizon increases. Dedicated Short-Bias funds are, in a sense, mirrors of traditional long-only managers. The increase in the absolute value of their Beta may be due to excessive leverage in *Bear Markets*[4] or the use of *Put Options*[5] with a non-linear return profile. Managed Futures managers tend to use forecasting models that depend heavily on past price movements. The time horizons of their trades can vary greatly depending on the time horizon examined in their models and hence this may have a large impact on the aggregate amount of market risk held at different time intervals.

In summary, the level of market risk held by different strategies varies greatly according to strategy and the time horizon studied. Given the fact that many funds have minimum lock-up periods, during which investors cannot liquidate their capital, it is more appropriate to measure the market risk over longer time intervals. This has obvious consequences for the allocation strategies of Fund of Funds[6] and Institutional investors alike.

5.3. Efficient frontier

To enable us to demonstrate graphically the effects of calculating the correlation matrix of Hedge Funds using different time horizons, we show (Fig. 1) the efficient frontier for each time horizon studied. In the calculation of the efficient frontier we have imposed constraints on the weights allowed, to prevent short-selling of Hedge Funds. This constraint is natural in the context of Hedge Funds due to the difficulties in short-selling of funds; (note that short-selling may be achievable by the use of swaps but is uncommon) [30,31]. Each of these efficient frontiers is calculated using the annualised returns and variance from the original data. Hence, the results we see in Fig. 1 are purely dependent on the correlation matrix calculated at each time horizon. The efficient frontiers are calculated using the techniques described in Section 2.1 and consist of just the Credit Suisse/Tremont sub-indices.

At lower return values, the level of risk calculated for the original data is higher than that calculated at all other scales. As the return level increases, the risk of the portfolio (created using the correlation matrix calculated using the monthly returns), decreases relative to the others. However, the lowest risk portfolio overall is found using a correlation matrix calculated over a 16–32 month horizon (Scale 4). These findings have significant implications for investors allocating capital to Hedge Funds.

6. Conclusions

The MODWT was applied to demonstrate the dependence of the correlation and market risk of Hedge Funds on the coarse/fine graining of Hedge Fund time series. This dependence was found to apply for correlations between Hedge Fund

[4] A Bear Market is considered to be a prolonged period in which asset prices fall, accompanied by widespread pessimism.

[5] A put is an option contract that gives the holder the right to sell a certain quantity of an underlying security to the writer of the option, at a specified price up to a specified date.

[6] A Hedge Fund which invests in other Hedge Funds.

sub-strategies and both the S&P 500 and the Hedge Fund composite index. In particular, the correlation between Convertible Arbitrage, Fixed Income Arbitrage and Multi-Strategy and both the S&P and Hedge Fund Composite index was found to increase as the time scale increased. The correlation between Dedicated Short Bias, Equity Market Neutral, Global Macro and Managed Futures strategies with the S&P and the Hedge Fund Composite index was found to decrease as the time scale increases. This decrease in correlation over longer time scales for certain strategies is in contrast to the *Epps* effect [16,17] where the correlation between financial time series was shown to increase over long time scales. However, in the case of the *Epps* effect the time scales studied were much shorter [17] ($0–2\frac{1}{2}$ h).

The level of market risk held by different Hedge Fund strategies was also found to vary according to the time horizon studied. The level of market risk of Convertible Arbitrage, Emerging Markets, Event-Driven and Long/Short Equity was found to increase as the time scale increased. The market risk of Dedicated Short Bias, Global Macro and Managed Futures was found to decrease as the time scale increased.

The minimum lock-up times applicable to many Hedge Funds mean that it may be more suitable to use correlation matrices calculated using longer time scales. To demonstrate this, we calculated the efficient frontier for four different time scales, as well as for the original data for a portfolio constructed from the sub-strategies. It was found that the level of risk of a portfolio of funds varies according to time-scale with the lowest risk portfolio attained using a scale of 16–32 months. This scale is significant as Hedge Funds often have lock-up period of 12–24 months inside of which investors cannot withdraw their invested capital [30].

Interesting future work worth considering include the measurement of the Alpha (Eq. (4)) of different strategies over different time scales. In the case of Hedge Funds the Alpha is a measure of the abnormal return, which is the value added by the manager. A Hedge Fund manager who appears to add Alpha, using monthly returns, may actually be found to just hold market risk over longer time scales and hence be less interesting to investors. Additional insights could be expected through the analysis of the levels of correlation and market risk of Hedge Funds with, for example, a bond index, a world equity index and a small-cap equity index over different scales using a multi-factor regression model.

References

[1] The Barclay Group. www.barclayhedge.com, 2008 (accessed on 4.04.08).
[2] T. Schneeweis, G. Martin, J. Altern. Invest. 4 (2001) 7–26.
[3] B. Liang, Financ. Anal. J. 57 (2001) 1.
[4] N. Nishiyama, Physica A 301 (1–4) (2001) 457–472.
[5] M.A. Miceli, G. Susinno, Physica A 344 (1–2) (2004) 95–99.
[6] T. Conlon, H.J. Ruskin, M. Crane, Physica A 382 (2) (2007) 565–576.
[7] J. Perello, Physica A 383 (2) (2007) 480–496.
[8] B.J. Whitcher, P. Guthorp, D.B. Percival, J. Geophys. Res. 105 (D11) (2000) 14941–14962.
[9] C.R. Cornish, C.S. Bretherton, D.B. Percival, Bound-Lay. Meteorol. 119 (2006) 339–374.
[10] R. Benzi, L. Biferale, F. Toschi, Phys. Rev. Lett. 80 (1998) 3244–3247.
[11] J.B. Ramsey, C. Lampart, Macroecon. Dyn. 2 (1) (1998) 49–71.
[12] R. Gencay, F. Selcuk, B. Whitcher, Physica A 289 (2001) 249–266.
[13] R. Gencay, F. Selcuk, B. Whitcher, Physica A 289 (2001) 543–556.
[14] R. Gencay, F. Selcuk, B. Whitcher, Quant. Financ. 3 (2003) 108–116.
[15] R. Gencay, F. Selcuk, B. Whitcher, J. Int. Money Financ. 24 (2005) 55–77.
[16] T.W. Epps, J. Am. Stat. Assoc 74 (1979) 291–298.
[17] B. Tóth, J. Kertész, Physica A 383 (2007) 54–84.
[18] W.F. Sharpe, J. Finance, 35, 910–919.
[19] H. Markowitz, Portfolio Selection: Efficient Diversification of Investments, John Wiley & Sons, 1959.
[20] E.J. Elton, M.J. Gruber, Modern Portfolio Theory and Investment Analysis, 6th ed., John Wiley & Sons, 2002.
[21] J.P. Bouchaud, M Potters, Theory of Financial Risk and Derivative Pricing, Cambridge University Press, 2003.
[22] R. Gencay, B. Selcuk, B. Whitcher, An Introduction to Wavelets and other Filtering Methods in Finance and Economics, Academic Press, 2001.
[23] A. Sharkasi, H.J. Ruskin, M. Crane, Int. J. Theor. Appl. Finance 8 (5) (2005) 603–622.
[24] A. Sharkasi, M. Crane, H.J. Ruskin, J.A. Matos, Physica A 368 (2) (2006) 511–521.
[25] I. Daubechies, Ten lectures on Wavelets, SIAM, 1992.
[26] A. Bruce, H. Gao, Applied Wavelet Analysis with S-Plus, Springer-Verlag, 1996.
[27] C.S. Burrus, R.A. Gopinath, H. Gao, Introduction to Wavelets and Wavelets Transforms, Prentice Hall, 1997.
[28] D.B. Percival, A.T. Walden, Wavelet Methods for Time Series Analysis, Cambridge University press, 2000.
[29] Credit Suisse/Tremont Hedge Fund Indices. www.hedgeindex.com, 2008 (accessed 11.01.08).
[30] F.S. L'habitant, Hedge Funds, Wiley, 2002.
[31] F.S. L'habitant, Hedge Funds — Quantitative Insights, Wiley, 2004.
[32] S&P 500 Index returns data, finance.yahoo.com, 2008 (accessed 11.01.08).
[33] D. Wilcox, T. Gebbie, Physica A 344 (1–2) (2004) 294–298.

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Cross-correlation dynamics in financial time series

T. Conlon *, H.J. Ruskin, M. Crane

Dublin City University, Glasnevin, Dublin 9, Ireland

ARTICLE INFO

Article history:
Received 8 July 2008
Received in revised form 4 October 2008
Available online 17 November 2008

PACS:
89.65.Gh
89.65.-s
89.75.-k

Keywords:
Correlation matrix
Eigenspectrum analysis
Econophysics

ABSTRACT

The dynamics of the equal-time cross-correlation matrix of multivariate financial time series is explored by examination of the *eigenvalue spectrum* over sliding time windows. Empirical results for the S&P 500 and the Dow Jones Euro Stoxx 50 indices reveal that the dynamics of the small eigenvalues of the cross-correlation matrix, over these time windows, oppose those of the largest eigenvalue. This behaviour is shown to be independent of the size of the time window and the number of stocks examined.

A basic one-factor model is then proposed, which captures the main dynamical features of the eigenvalue spectrum of the empirical data. Through the addition of perturbations to the one-factor model, (leading to a 'market plus sectors' model), additional sectoral features are added, resulting in an Inverse Participation Ratio comparable to that found for empirical data. By partitioning the eigenvalue time series, we then show that negative index returns, (*drawdowns*), are associated with periods where the largest eigenvalue is greatest, while positive index returns, (*drawups*), are associated with periods where the largest eigenvalue is smallest. The study of correlation dynamics provides some insight on the collective behaviour of traders with varying strategies.

C-12

1. Introduction

In recent years, the analysis of the equal-time cross-correlation matrix for a variety of multivariate data sets such as financial data, [1–12], electroencephalographic (EEG) recordings, [13,14], magnetoencephalographic (MEG) recordings, [15], and others, has been studied extensively. Other authors have investigated the relationship between stock price changes and liquidity or trading volume, [16–18]. In particular, Random Matrix Theory, (RMT), has been applied to filter the relevant information from the statistical fluctuations, inherent in empirical cross-correlation matrices, for various financial time series, [1–10]. By comparing the eigenvalue spectrum of the correlation matrix to the analytical results, obtained for random matrix ensembles, significant deviations from RMT eigenvalue predictions provide genuine information about the correlation structure of the system. This information has been used to reduce the difference between predicted and realised risk of different portfolios.

Several authors have suggested recently that some real correlation information may be hidden in the RMT defined random part of the eigenvalue spectrum. A technique, involving the use of power mapping to identify and estimate the noise in financial correlation matrices, [19], allows the suppression of those eigenvalues, associated with the noise, in order to reveal different correlation structures buried underneath. The relationship, between the eigenvalue density c of the true correlation matrix, and that of the empirical correlation matrix C, was derived to show that correlations can be measured in the random part of the spectrum, [20,21]. A Kolmogorov test was applied to demonstrate that the bulk of the spectrum is not in the Wishart RMT class, [22], while the existence of factors, such as an overall market effect, firm size and industry type, is due to collective influence of the assets. More evidence that the RMT fit is not perfect was provided, [23], where it was

* Corresponding author. Tel.: +353 863718053.
E-mail addresses: tconlon@computing.dcu.ie (T. Conlon), hruskin@computing.dcu.ie (H.J. Ruskin), mcrane@computing.dcu.ie (M. Crane).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.10.047

shown that the dispersion of "noise" eigenvalues is inflated, indicating that the bulk of the eigenvalue spectrum contains correlations masked by measurement noise.

The behaviour of the largest eigenvalue of a cross-correlation matrix for small windows of time, has been studied, [24], for the DAX and Dow Jones Industrial average Indices (DJIA). Evidence of a time-dependence between 'drawdowns' ('drawups') and an increase (decrease) in the largest eigenvalue was obtained, resulting in an increase of the *information entropy*[1] of the system. Similar techniques were used, [25], to investigate the dynamics between the stocks of two different markets (DAX and DJIA). In this case, two distinct eigenvalues of the cross-correlation matrix emerged, corresponding to each of the markets. By adjusting for time-zone delays, the two eigenvalues were then shown to coincide, implying that one market leads the dynamics in the other.

Equal-time cross-correlation matrices have been used, [26], to characterise dynamical changes in nonstationary multivariate time-series. It was specifically noted that, for increased synchronisation of k series within an M-dimensional multivariate time series, a repulsion between eigenstates of the correlation matrix results, in which k levels participate. Through the use of artificially-created time series with pre-defined correlation dynamics, it was demonstrated that there exist situations, where the relative change in eigenvalues from the lower edge of the spectrum is greater than that for the large eigenvalues, implying that information drawn from the smaller eigenvalues is highly relevant.

The technique was subsequently applied to the eigenvalue spectrum of the equal time cross-correlation matrix of multivariate Epileptic Seizure time series and information on the correlation dynamics was found to be visible in *both* the lower and upper eigenstates. Further studies of the equal-time correlation matrix of EEG signals, [13], investigated temporal dynamics of focal onset epileptic seizures[2] and showed that zero-lag correlations between multichannel EEG signals tend to decrease during the first half of a seizure and increase gradually before the seizure ends. A further extension (to the case of *Status Epilepticus*, [14]), used the equal-time correlation matrix to assess neuronal synchronisation prior to seizure termination.

Examples have demonstrated, [27], that information about cross correlations can be found in the RMT bulk of eigenvalues and that the information extracted at the *lower* edge is statistically *more significant* than that extracted from the larger eigenvalues. The authors introduced a method of unfolding the eigenvalue level density, through the normalisation of each of the level distances by its ensemble average, and used this to calculate the corresponding individual nearest-neighbour distance. Those parts of the spectrum, dominated by noise, could be distinguished from those containing information about correlations. Application of this technique to multichannel EEG data showed the smallest eigenvalues to be more sensitive to detection of subtle changes in the brain dynamics than the largest.

In this paper, we examine eigenvalue dynamics of the cross-correlation matrix for multivariate financial data with a view to characterising market behaviour. Methods are reviewed in Section 2, the data described in Section 3 and in Section 4 we look at the results obtained both for the empirical correlation matrix and the model correlation matrices described.

2. Methods

2.1. Empirical dynamics

The equal-time cross-correlation matrix, between time series of equity returns, is calculated using a sliding window, where the number of assets, N, is smaller than the window size T. Given returns $G_i(t)$, $i = 1, \ldots, N$, of a collection of equities, we define a normalised return, within each window, in order to standardise the different equity volatilities. We normalise G_i with respect to its standard deviation σ_i as follows:

$$g_i(t) = \frac{G_i(t) - \widehat{G_i(t)}}{\sigma_i} \tag{1}$$

where σ_i is the standard deviation of G_i for assets $i = 1, \ldots, N$ and $\widehat{G_i}$ is the time average of G_i over a time window of size T.

Then, the equal-time cross-correlation matrix is expressed in terms of $g_i(t)$

$$C_{ij} \equiv \langle g_i(t)\, g_j(t) \rangle . \tag{2}$$

The elements of C_{ij} are limited to the domain $-1 \leq C_{ij} \leq 1$, where $C_{ij} = 1$ defines perfect positive correlation, $C_{ij} = -1$ corresponds to perfect negative correlation and $C_{ij} = 0$ corresponds to no correlation. In matrix notation, the correlation matrix can be expressed as

$$\mathbf{C} = \frac{1}{T}\mathbf{G}\mathbf{G}^{\tau} \tag{3}$$

where $\mathbf{G}$ is an $N \times T$ matrix with elements g_{it}.

The eigenvalues $\boldsymbol{\lambda}_i$ and eigenvectors $\hat{\mathbf{v}}_i$ of the correlation matrix $\mathbf{C}$ are found from the following

$$\mathbf{C}\hat{\mathbf{v}}_i = \boldsymbol{\lambda}_i \hat{\mathbf{v}}_i. \tag{4}$$

[1] In information theory, the Shannon entropy or information entropy is a measure of the uncertainty associated with a random variable.
[2] A partial or focal onset seizure occurs when the discharge starts in one area of the brain and then spreads over other areas.

The eigenvalues are then ordered according to size, such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. The sum of the diagonal elements of a matrix, (the Trace), must always remain constant under linear transformation. Thus, the sum of the eigenvalues must always equal the Trace of the original correlation matrix. Hence, if some eigenvalues increase then others must decrease, to compensate, and vice versa (*Eigenvalue Repulsion*).

There are two limiting cases for the distribution of the eigenvalues [13,26]. When all of the time series are perfectly correlated, $C_i \approx 1$, the largest eigenvalue is maximised with a value equal to N, while for time series consisting of random numbers with average correlation $C_i \approx 0$, the corresponding eigenvalues are distributed around 1, (where any deviation is due to spurious random correlations).

For cases between these two extremes, the eigenvalues at the lower end of the spectrum can be much smaller than $\lambda_{\max}$. To study the dynamics of each of the eigenvalues using a sliding window, we normalise each eigenvalue in time using

$$\tilde{\lambda}_i(t) = \frac{\left(\lambda_i(t) - \widehat{\lambda_i(\tau)}\right)}{\sigma^{\lambda_i(\tau)}} \tag{5}$$

where $\widehat{\lambda}$ and σ^{λ} are the mean and standard deviation of the eigenvalue i over a particular reference period, τ. This normalisation allows us to visually compare eigenvalues at both ends of the spectrum, even if their magnitudes are significantly different. The reference period, used to calculate mean and standard deviation of the eigenvalue spectrum, can be chosen to be a low volatility sub-period, (which helps to enhance the visibility of high volatility periods), or the full time period studied.

2.2. One-factor model

In the *one-factor model* of stock returns, we assume a *global correlation* with the cross-correlation between all stocks the same, ρ_0. The spectrum of the associated correlation matrix consists of only two values, a large eigenvalue of order $(N-1)\rho_0 + 1$, associated with the market, and an $(N-1)$-fold degenerate eigenvalue of size $1 - \rho_0 < 1$. Any deviation from these values is due to the finite length of time series used to calculate the correlations. In the limit $N \to \infty$ (even for *small correlation*, i.e. $\rho \to 0$) a large eigenvalue appears, which is associated with the eigenvector $v_1 = \left(\frac{1}{\sqrt{N}}\right)(1, 1, 1 \ldots 1)$, and which dominates the correlation structure of the system.

2.3. 'Market plus sectors' model

To expand the above to a 'market plus sectors' model, we perturb a number of pairs N of the correlations $\rho_0 + \rho_n$, where $-1-\rho_0 \leq \rho_n \leq 1-\rho_0$. Additionally, we impose a constraint $\sum_N \rho_n = 0$, ensuring that the average correlation of the system remains equal to ρ_0. These perturbations allow us to introduce *groups of stocks* with similar correlations, (corresponding to Market Sectors).

Using the correlation matrix from the "one-factor model" and the "market plus sectors model", we can construct correlated time series using the Cholesky decomposition A of a correlation matrix $C = AA^{\tau}$. We can then generate finite correlated time series of length T,

$$x_{it} = \sum_j A_{ij} y_{jt} \quad t = 1, \ldots, T \tag{6}$$

where y_{jt} is a random Gaussian variable with mean zero and variance 1 at time t. Using Eq. (2) we can then construct a correlation matrix using the simulated time series. The finite size of the time series introduces 'noise' into the system and, hence, empirical correlations will vary from sample to sample. This 'noise' could be reduced through the use of longer simulated time series or through averaging over a large number of time series.

In order to compare the eigenvectors from each of the Model Correlation matrices to that constructed from the equity returns time series, we use the Inverse Participation Ratio (IPR) [4,29]. The IPR allows quantification of the number of components that participate significantly in each eigenvector and tells us more about the level and nature of deviation from RMT. The IPR of the eigenvector u^k is given by $I^k \equiv \sum_{l=1}^{N} \left(u_l^k\right)^4$ and allows us to compute the inverse of the number of eigenvector components that contribute significantly to each eigenvector.

3. Data

In order to study the dynamics of the empirical correlation matrix over time, we analyse two different data sets. The first data set comprises the 384 equities of the Standard & Poors (S&P) 500 where full price data is available from January 1996 to August 2007, resulting in 2938 daily returns. The S&P 500 is an index consisting of 500 large capitalisation equities, which are predominantly from the US. In order to demonstrate that our results are not market specific, however, we examine a second data set, made up of the 49 equities of the Dow Jones Euro Stoxx 50, where full price data is available from January 2001 to August 2007, resulting in 1619 daily returns. The Dow Jones Euro Stoxx 50 is a stock index of Eurozone equities designed to provide a blue-chip representation of supersector leaders in the Eurozone.

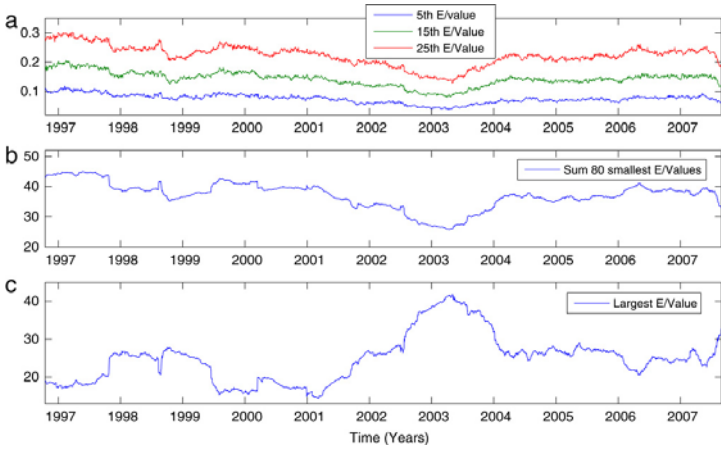


Fig. 1. Time Evolution of (a) Three small eigenvalues (b) Sum of the 80 smallest eigenvalues (c) The largest eigenvalue.

4. Results

We analyse the eigenvalue dynamics of the correlation matrix of a subset of 100 S&P equities, chosen randomly, using a sliding window of 200 days. This subsector was chosen such that $Q = \frac{T}{N} = 2$, thus ensuring that the data would be close to non-stationary in each sliding window. Fig. 1(a) shows broadly similar sample dynamics from the 5th, 15th and 25th largest eigenvalues over each of these sliding windows. The sum of the 80 smallest eigenvalues are shown in Fig. 1(b), while the dynamics of the largest eigenvalue is displayed in Fig. 1(c). The repulsion between the largest eigenvalue and the small eigenvalues is evident here, (comparing Fig. 1(b) and (c)), with the dynamics of the small eigenvalues contrary to those of the largest eigenvalue. As noted earlier (Section 2.1), this is a consequence of the fact that the trace of the correlation matrix must remain constant under transformations and any change in the largest eigenvalue must be reflected by a change in one or more of the other eigenvalues. Similar results were obtained for different subsets of the S&P and also for the members of the Dow Jones Euro Stoxx 50.

4.1. Normalised eigenvalue dynamics

Using normalised eigenvalues as described above, (Eq. (5)), we performed a number of experiments to investigate the dynamics of a set of small eigenvalues versus the largest eigenvalue. The various experiments are described below:

(1) As in Section 4, the dynamics for the same subset of 100 equities are analysed using a sliding window of 200 days. The normalisation is carried out using the mean and standard deviation of each of the eigenvalues over the entire time-period. Fig. 2(a) shows the value of the S&P index from 1997 to mid −2007.

The normalised largest eigenvalue is shown in Fig. 2(b) together with the average of the 80 normalised small eigenvalues. The compensatory dynamics mentioned earlier are shown more clearly here, with the largest and average of the smallest 80 eigenvalues having opposite movements. The normalised eigenvalues for the entire eigenvalue spectrum are shown in Fig. 2(c), where the colour indicates the number of standard deviations from the time average for each of the eigenvalues over time. As shown, there is very little to differentiate the dynamics of the 80–90 or so smallest eigenvalues. In contrast, the behaviour of the largest eigenvalue is clearly opposite to that of the smaller eigenvalues. However, from the 90th and subsequent eigenvalue there is a marked change in the behaviour, (Fig. 2(d)), and the eigenvalue dynamics are distinctly different. This may correspond to the area outside the "Random Bulk" in RMT. Similar to Refs. [24,25], we also find evidence of an increase/decrease in the largest eigenvalue with respect to 'drawdowns'/'draw-ups'. Additionally, we find the highlighted *compensatory dynamics* of the small eigenvalues. These results were tested for various time windows and normalisation periods, with smaller windows and normalisation periods found to capture and emphasise additional features.

(2) To demonstrate the above result for a *different level of granularity*, we chose 50 equities randomly with a time window of 500 days, giving $Q = \frac{T}{N} = 10$. The results obtained, (Fig. 3), are in keeping with those for $Q = 2$ earlier, with a broad-band increase (decrease) of the 40 smallest eigenvalues concurrent to a decrease (increase) of the largest eigenvalue.

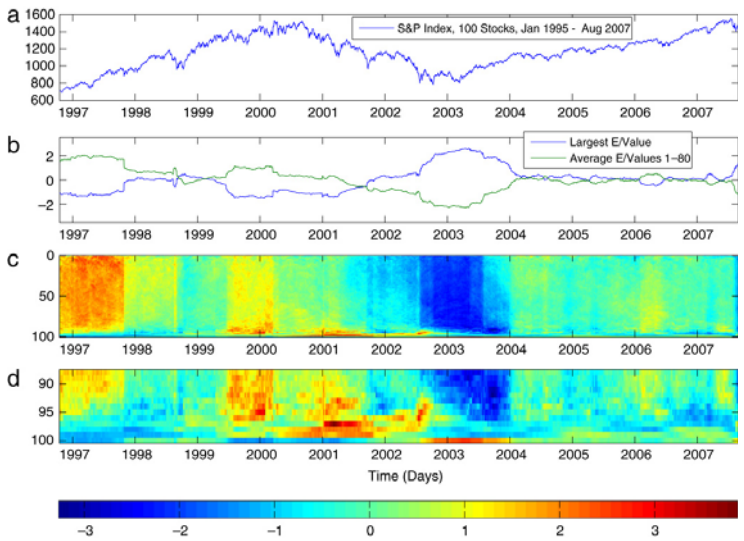


Fig. 2. (a) S&P Index (b) Normalised largest eigenvalue vs. average of 80 smallest normalised eigenvalues (c) All normalised eigenvalues (d) Largest 12 normalised eigenvalues.

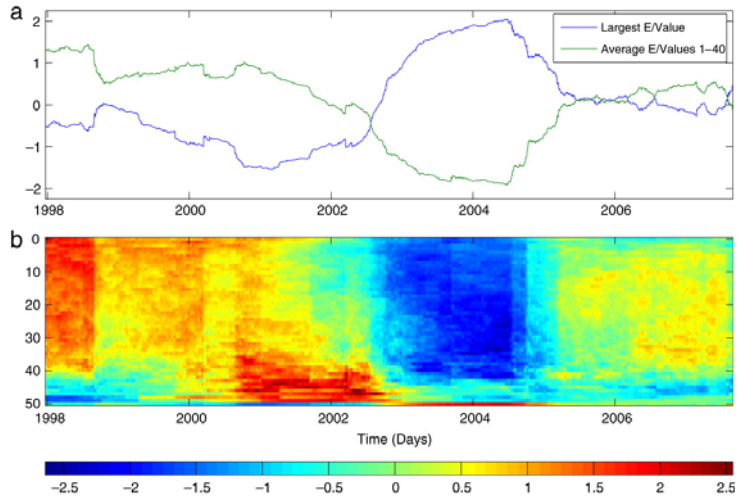


Fig. 3. (a) Normalised largest eigenvalue vs. Average of 40 smallest normalised eigenvalues (b) All normalised eigenvalues.

(3) The previous examples used random subsets of the S&P universe in order to keep $Q = \frac{T}{N}$ as large as possible. To demonstrate that the above results were not sampling artifacts, we also looked at the full sample of 384 equities, (that survived the entire 11 year period), with a time window of 500 days ($Q = 1.30$). The results, as shown in Fig. 4, are similar to those above, with the majority of the small eigenvalues compensating for changes in the large eigenvalue. As indicated previously, however, there is a small band of large eigenvalues, for which behaviour is different to that of both the band of small eigenvalues and the largest eigenvalue.

(4) All examples discussed so far have focused on the universe of equities from the S&P 500 that have survived since 1997. To ensure that the results obtained were not exclusive to the S&P 500, we also applied the same technique to the 49 equities of the EuroStoxx 50 index that survived from January 2001 to August 2007. The sliding window used was 200 days, such that $Q = 4.082$. The results, (Fig. 5), were again similar to before, with a wide band of small eigenvalues "responding to" movements in the largest eigenvalues. In this case, the band of deviating large eigenvalues (ie. those which correspond to the area outside the "Random Bulk" in RMT), (Fig. 5(d)), is not as marked as in the previous example. This effectively implies that equities in this index are dominated by "the Market".

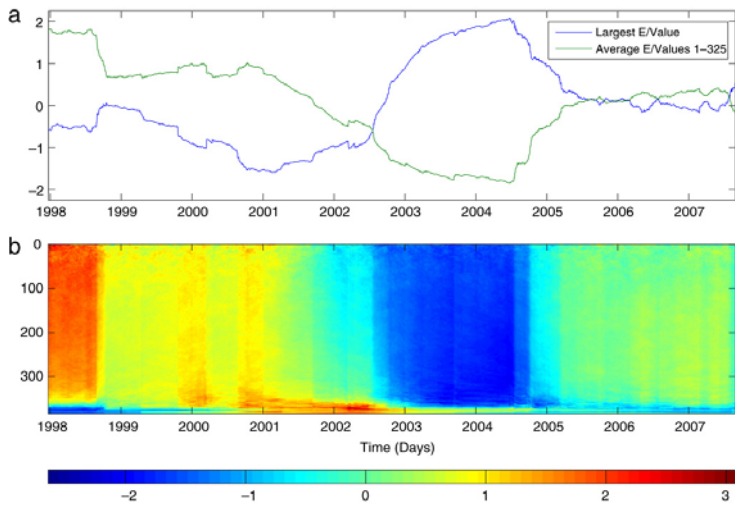


Fig. 4. (a) Normalised largest eigenvalue vs. Average of 325 smallest normalised eigenvalues (b) All normalised eigenvalues.

4.2. Model correlation matrix

The results, described, demonstrate that the time dependent dynamics of the small eigenvalues of the correlation matrix of stock returns move counter to those of the largest eigenvalue. Here, we show how a simple one-factor model, Section 2.2, of the correlation structure reproduces much of this behaviour. Furthermore, we show how additional features can be captured by including perturbations in this model, essentially a *"market plus sectors"* model, Section 2.3, [22,28–30].

In order to compare the empirical results, Section 4, to those of the single factor model, we first constructed a correlation matrix where each non-diagonal element was equal to the average correlation of the empirical matrix in each sliding window. We then calculated the eigenvalues of this matrix over each sliding window and normalised these as before, (Section 2). The dynamics of the largest eigenvalue for the single-factor model are displayed in Fig. 6(a) for the EuroStoxx 50 index using a sliding window of 200 days. As can be seen, the main features of the dynamics are in agreement with those of Fig. 5 for the empirical data. The dynamics of the remaining eigenvalues, shown in Fig. 6(c), are found to be within a narrow range, with any change in time due to compensation for change in the largest eigenvalue.

For the 'market plus sectors' model, we introduced perturbations with two groups of 5 stocks having correlation $\rho_0 - 0.15$ and $\rho_0 + 0.15$, with the average correlation at each time window remaining the same. The largest eigenvalue dynamics for the 'market plus sectors' model is shown in Fig. 6(b), with the dynamics almost identical to those for the largest eigenvalue of the 'one-factor' model, (any differences are due to random fluctuations induced in the Cholesky decomposition, Section 2.3). However, looking at the remaining eigenvalues, Fig. 6(d), we see that a number are found to deviate significantly from the bulk. These deviations are due to the additional sectorial information included in the market plus sectors model. This agrees with previous results, [4,6], where small eigenvalues were found to deviate from the random bulk, in addition to large deviating eigenvalues containing sectorial information.

To examine properties of the eigenvector components themselves, we use the Inverse Participation Ratio (IPR). Fig. 7(a), displays the IPR found using the empirical data from the Eurostoxx 50. This has similar characteristics to those found for

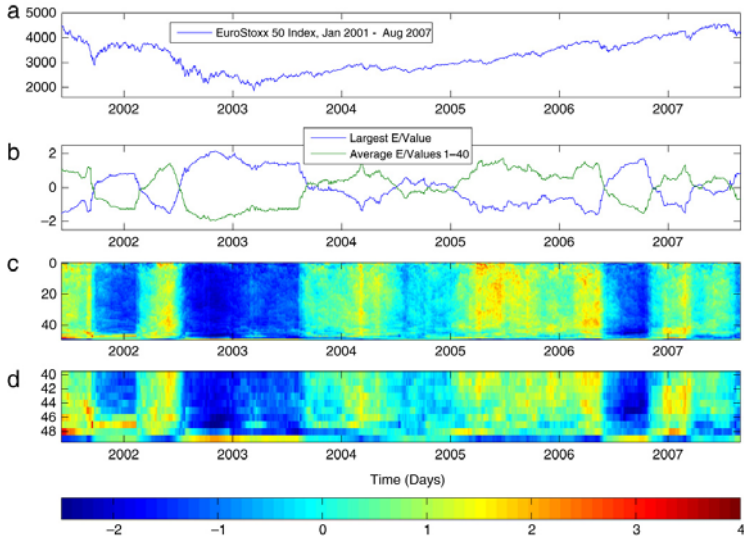


Fig. 5. (a) EuroStoxx 50 Index, Jan 2001–Aug 2007 (b) Normalised largest eigenvalue vs. Average of 40 smallest normalised eigenvalues for EuroStoxx 50 (c) All normalised eigenvalues (d) The 9 largest normalised eigenvalues.

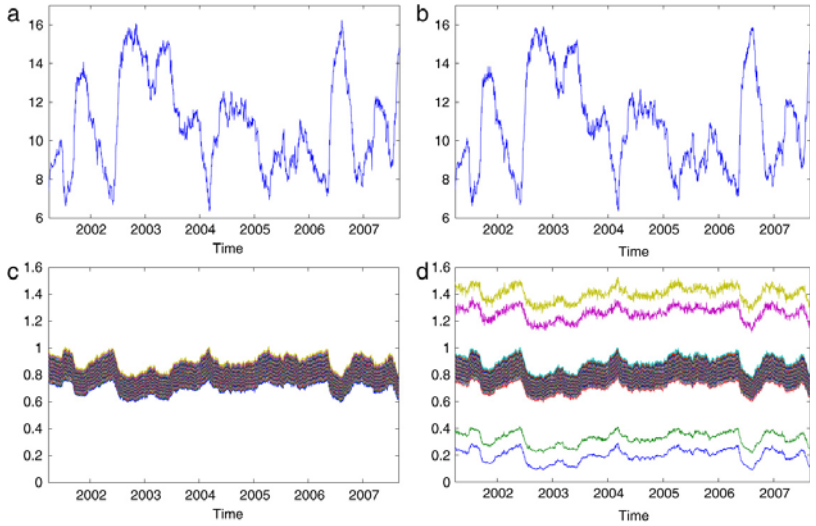


Fig. 6. (a) Largest eigenvalue, one factor model (b) Largest eigenvalue, Market plus sectors model (c) Eigenvalues 1-48, one factor model (d) Eigenvalues 1-48, Market plus sectors model.

different indices, [4], with the IPR for the largest eigenvalue much smaller than the mean, a large IPR corresponding to sectorial information in the 2nd or 3rd largest eigenvalues and the IPR for the small eigenvalues raised. For the single factor

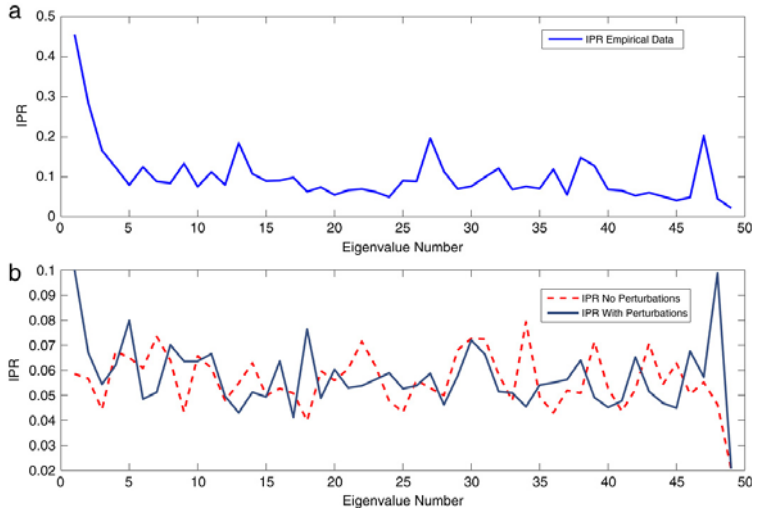


Fig. 7. (a) IPR Empirical Data (b) IPR Simulated Data.

model, we created a *synthetic* correlation matrix using Eq. (6), with average correlation (0.204) equal to that of the Euro Stoxx 50 over the time period studied. As shown in Fig. 7(b), the IPR retains some of the features found for empirical data, [4,29], with that corresponding to the largest eigenvector having a much smaller value than the mean. This corresponds to an eigenvector to which many stocks contribute, (effectively the market eigenvector), [4,29].

In an attempt to include additional empirical features, such as the band of deviating large eigenvalues between the bulk and the largest eigenvalue, we also considered a perturbation, with two groups of stocks having correlation $\rho_0 - 0.15$ and $\rho_0 + 0.15$, and the same average correlation. In this case, Fig. 7(b), additional features are found, with larger IPR for both smallest and second largest eigenvalue. This agrees with Ref. [4] where, for empirical data, the group structure resulted in a number of small and large eigenvalues with larger IPR than that of the bulk of eigenvalues. These large eigenvalues were shown, [4,10], to be associated with correlation information related to the group structure.

4.3. Drawdown analysis

As described, [24], and discussed above, drawdowns (periods of large negative returns) tend to reflect an increase of one eigenstate of the cross-correlation matrix, with the opposite occurring during drawups. In this section, we attempt to characterise the market according to the relative size of both the largest and small eigenvalues, using stocks of the Dow Jones Euro Stoxx 50.

Returns, correlation matrix and eigenvalue spectrum time series for overlapping windows of 200 days were calculated and normalised using the mean and standard deviation over the entire series, (Eq. (5)). Representing normalised eigenvalues in terms of standard deviation units (SDU) allows partitioning according to the magnitude of the eigenvalues. Table 1 shows the average return of the index, (during periods when the largest eigenvalue is ± 1 SDU), for both the largest eigenvalue and the average of the normalised 40 smallest eigenvalues.

The results, Table 1, demonstrate that when the largest eigenvalue is >1 SDU, the average index return over a 200 day period is found to be -16.80%. When it is small (<-1 SDU), the average index return is 18.46%. Hence, the largest eigenvalue can be used to characterise markets, with the eigenvalue peaking during periods of negative returns (Drawdowns) and bottoming out when the market is rising (Drawup). For the average of the 40 smallest eigenvalues, the partition also reflected drawdowns and drawups, but with opposite signs. This indicates that information about the correlation dynamics of financial time series is visible in both the lower and upper eigenstates, in agreement with the results found by Refs. [13,26] for both synthetic data and EEG seizure data.

5. Conclusions

The correlation structure of multivariate financial time series was studied by investigation of the eigenvalue spectrum of the equal-time cross-correlation matrix. By filtering the correlation matrix through the use of a sliding window we have

Table 1
Drawdown/Drawup analysis, average index returns for eigenvalue partitions in SDU.

Eigenvalues	No. Std	Index return (%)
Large	<-1	18.46
	>1	−16.80
Average 40 smallest	<-1	−23.90
	>1	19.32

examined the behaviour of the largest eigenvalue over time. As shown, Figs. 2–5, the largest eigenvalue moves counter to that of a band of small eigenvalues, due to *eigenvalue repulsion*. A decrease in the largest eigenvalue, with a corresponding increase in the small eigenvalues, corresponds to a redistribution of the correlation structure across more dimensions of the vector space spanned by the correlation matrix. Hence, additional eigenvalues are needed to explain the correlation structure in the data. Conversely, when the correlation structure is dominated by a smaller number of factors (eg. the "single-factor model" of equity returns), the number of eigenvalues needed to describe the correlation structure in the data is reduced. In the context of previous work, [24,25], this means that fewer eigenvalues are needed to describe the correlation structure of 'drawdowns' than that of 'draw-ups'.

By introducing a simple 'one-factor model' of the correlation in the system (Section 4.2), we were able to reproduce the main results of the empirical study. The compensatory dynamics, described, were clearly seen for a correlation matrix, with all elements equal to the average of the empirical correlation matrix. The model was then adapted, by the addition of perturbations to the correlations, with the average correlation remaining unchanged. This 'markets plus sectors' type model was able to reproduce additional features of the empirical correlation matrix, with eigenvalues deviating from below and above the bulk. The Inverse Participation Ratio of the "markets plus sectors" model was also shown to have group characteristics typically associated with known Industrial Sectors, with a larger than average value for the smallest eigenvalue and for the second largest eigenvalue. Through a partition of the time-normalised eigenvalues, it was then shown quantitatively that the largest eigenvalue is greatest/smallest during drawdowns/drawups, and vice versa for the small eigenvalues.

Suggested future work includes a more detailed study of the relationship between the direction of the market and magnitude of the eigenvalues of the correlation matrix. Studying the multiscaled correlation dynamics over *different granularities* may shed some light on the different collective behaviour of traders with different strategies and time horizons. Additional analysis of high frequency data may also be useful in the characterisation of correlation dynamics, especially prior to market crashes. Study of the possible relationship between the dynamics of the small eigenvalues and additional correlation information which, according to some authors, [19–23,26,27], may be hidden in the part of the eigenvalue spectrum normally classified as noise, could be achieved through analysis of the relative dynamics of the small and large eigenvalues at times of extreme volatility (such as during market crashes).

References

[1] L. Laloux, P. Cizeau, J. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (7) (1999) 1467–1470.
[2] Plerou Gopikrishnan, Rosenow Amaral, T. Gurh, H.E. Stanley, Universal and non-universal properties of cross-correlations in financial time series, Phys. Rev. Lett. 83 (1999) 1471–1474.
[3] L. Laloux, P. Cizeau, M. Potters, J. Bouchaud, Random matrix theory and financial correlations, Int. J. Theoret. Appl. Finance 3 (3) (2000) 391–397.
[4] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, H.E. Stanley, A random matrix approach to cross-correlations in financial data, Phys. Rev. E 65 (2000) 066126.
[5] P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, Identifying business sectors from stock price fluctuations, Phys. Rev. E 64 (2001) 035106R.
[6] A. Utsugi, K. Ino, M. Oshikawa, Random matrix theory analysis of cross-correlations in financial markets, Phys. Rev. E 70 (2004) 026110.
[7] J.P. Bouchaud, M. Potters, Theory of Financial Risk and Derivative Pricing, Cambridge University Press, 2003.
[8] D. Wilcox, T. Gebbie, On the analysis of cross-correlations in South African market data, Physica A 344 (1–2) (2004) 294–298.
[9] S. Sharifi, M. Crane, A. Shamaie, H.J. Ruskin, Random matrix theory for portfolio optimization: A stability approach, Physica A 335 (3–4) (2004) 629–643.
[10] T. Conlon, H.J. Ruskin, M. Crane, Random matrix theory and fund of funds portfolio optimisation, Physica A 382 (2) (2007) 565–576.
[11] T. Conlon, H.J. Ruskin, M. Crane, Wavelet multiscale analysis for Hedge Funds: Scaling and strategies, Physica A 387 (2008) 5197–5204.
[12] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: A new method for analysing two non-stationary time series, Phys. Rev. Lett 100 (2008) 084102.
[13] K. Schindler, H. Leung, C.E. Elger, K. Lehnertz, Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG, Brain 130 (2007) 65–77.
[14] K. Schindler, C.E. Elger, K. Lehnertz, Increasing synchronization may promote seizure termination: Evidence from status epilepticus, Clin. Neurophysiol. 118 (9) (2007) 1955–1968.
[15] J. Kwapien, S. Drozdz, A.A. Ioannides, Temporal correlations versus noise in the correlation matrix formalism: An example of the brain auditory response, Phys. Rev. E 62 (2000) 5557.
[16] C.C. Ying, Stock market prices and volumes of sales, Econometrica 34 (1966) 676.
[17] J.J. Karpoff, The relationship between price changes and trading volumes: A survey, Financ. Quant. Anal. 22 (1987) 109.
[18] B. LeBaron, W.B. Arthur, R. Palmer, Time series properties of an artificial stock market, J. Econom. Dynam. Control 23 (1999) 1487.
[19] T. Guhr, B. Kälber, A new method to estimate the noise in financial correlation matrices, J. Phys. A 35 (2003) 3009–3032.
[20] Z. Burda, A. Görlich, A. Jarosz, J. Jurkiewicz, Signal and noise in correlation matrices, Physica A 343 (2004) 295–310.
[21] Z. Burda, J. Jurkiewicz, Signal and noise in financial correlation matrices, Physica A 344 (1–1) (2004) 67–72.
[22] Y. Malevergne, D. Sornette, Collective origin of the coexistence of apparent random matrix theory noise and of factors in large sample correlation matrices, Physica A 331 (3–4) (2004) 660–668.
[23] J. Kwapien, S. Drozdz, P. Oswiecimka, The bulk of the stock market correlation matrix is not pure noise, Physica A 359 (2006) 589–606.
[24] S. Drozdz, F. Gruemmer, F. Ruf, J. Speth, Dynamics of competition between collectivity and noise in the stock market, Physica A 287 (2000) 440–449.

[25] S. Drozdz, F. Gruemmer, F. Ruf, J. Speth, Towards identifying the world stock market cross-correlations: DAX versus Dow-Jones, Physica A 294 (2001) 226.
[26] M. Müller, G. Baier, A. Galka, U. Stephani, H. Muhle, Detection and characterization of changes of the correlation matrix in multivariate time series, Phys. Rev. E 71 (2005) 046116.
[27] M. Müller, Y.L. Jiménez, C. Rummel, G. Baier, A. Galka, U. Stephani, H. Muhle, Localized short-range correlations in the spectrum of the equal-time correlation matrix, Phys. Rev. E 74 (2006) 041119.
[28] E.J. Elton, M.J. Gruber, S.J. Brown, W. Goetzmann, Modern Portfolio Theory and Investment Analysis, 7th ed., Wiley, 2006.
[29] J.D. Noh, A model for stock correlations for stock markets, Phys. Rev. E 61 (2000) 5981–5982.
[30] G. Papp, Sz. Pafka, M.A. Nowak, I. Kondor, Random matrix filtering in portfolio optimization, Acta Phys. Polon. B 36 (2005) 2757.

Advances in Complex Systems, Vol. 12, No. 4 (2009) 1–16

World Scientific
www.worldscientific.com

MULTISCALED CROSS-CORRELATION DYNAMICS IN FINANCIAL TIME-SERIES

T. CONLON*, H. J. RUSKIN and M. CRANE

School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland
**tconlon@computing.dcu.ie*

Received
Revised

The cross-correlation matrix between equities comprises multiple interactions between traders with varying strategies and time horizons. In this paper, we use the Maximum Overlap Discrete Wavelet Transform to calculate correlation matrices over different time–scales and then explore the *eigenvalue spectrum* over sliding time-windows. The dynamics of the eigenvalue spectrum at different times and scales provides insight into the interactions between the numerous constituents involved.

Eigenvalue dynamics are examined for both medium, and high-frequency equity returns, with the associated correlation structure shown to be dependent on both time and scale. Additionally, the *Epps* effect is established using this multivariate method and analyzed at longer scales than previously studied. A partition of the eigenvalue time-series demonstrates, at very short scales, the emergence of negative returns when the largest eigenvalue is greatest. Finally, a portfolio optimization shows the importance of time–scale information in the context of risk management.

Keywords:

C-17

1. Introduction

In recent years, the equal-time cross-correlation matrix has been studied extensively for a variety of multivariate data sets across different disciplines such as financial data [1–11], electroencephalographic (EEG) recordings [12, 13, 15], magnetoencephalographic (MEG) recordings [16], and others. In particular, Random Matrix Theory (RMT) has been applied to filter the relevant information from the statistical fluctuations inherent in empirical cross-correlation matrices, constructed for various types of financial [1–9], and MEG [16] time-series. In this paper, we extend previous Complex Systems research, by examining time and scale dependent dynamics of the correlation matrix, in order to better understand the ever-changing level of synchronization between financial time-series.

The dynamics of the largest eigenvalue of a cross-correlation matrix over small windows of time was studied [17] for the Dow Jones Industrial Average (DJIA)

1st Reading

August 21, 2009 20:16 WSPC/169-ACS 00232

and DAX indices. This analysis revealed evidence of time-dependence between "drawdowns"/"drawups" and an increase/decrease in the largest eigenvalue. The interactions between stocks of two different markets (DAX and DJIA) were then investigated [18] revealing two distinct eigenvalues of the combined cross-correlation matrix, corresponding to each market. Adjusting for time-zone delays, the two eigenvalues coincided, implying that one market leads the dynamics in the other.

It has been suggested recently by several authors [19–23] that there may, in fact, be some real correlation information hidden in the RMT defined part of the eigenvalue spectrum. A multivariate technique involving the equal-time cross-correlation matrix has been shown [12] to characterize dynamical changes in nonstationary multivariate time-series. Using this technique, the authors also demonstrated how correlation information can be observed in the noise band. As the synchronization of k time-series within an M-dimensional multivariate time-series increases, a repulsion between eigenstates of the correlation matrix results, in which k levels participate. Through the use of artificially created time-series with pre-defined correlation dynamics, it was demonstrated that there exist situations, where the relative change in eigenvalues from the lower edge of the spectrum is greater than that for the large eigenvalues, implying that information drawn from the smaller eigenvalues is highly relevant.

This technique was subsequently applied to the dynamic analysis of the eigenvalue spectrum of the equal-time cross-correlation matrix of multivariate Epileptic Seizure time-series, using sliding windows. The authors demonstrated that information about the correlation dynamics is visible in *both* the lower and upper eigenstates. A further study [15] which investigated temporal dynamics of focal onset epileptic seizures,[a] showed that the zero-lag correlations between multichannel EEG signals tend to decrease during the first half of a seizure and increase gradually before the seizure ends. Information about cross correlations was also found in the RMT bulk of eigenvalues [13] with that extracted at the *lower* edge statistically *more significant* than that from the larger eigenvalues. Application of this technique to multichannel EEG data showed small eigenvalues to be more sensitive to detection of subtle changes in the brain dynamics than the largest. A time- and frequency-based approach, has recently demonstrated the importance of correlation dynamics at high-frequencies during seizure activity [14].

Although originally applied to complex EEG seizure time-series [12, 13, 15], these techniques were adapted to financial time-series [11] where the correlation dynamics for medium-frequency data were calculated using short time-windows. Changes in the correlation structure were shown to be visible in both large and small eigenvalue dynamics. Further, "drawdowns" and "drawups" were shown to be market dependent, with corresponding characteristic changes in relative eigenvalue size found at both ends of the eigenvalue spectrum. This analysis was carried out using daily

[a]A focal onset or partial seizure occurs when the discharge starts in one area of the brain and then spreads over other areas.

data, with the effects of varying data granularity not examined. In this paper, we build upon this analysis by studying the multiscale effects on correlation dynamics using medium- and high-frequency data.

The correlation structure for various markets was studied in detail [24] for data encompassing a number of time horizons ranging from 5 to 255 min. By removing the center of mass (the market), the authors found that the correlation structure is well defined at high-frequency, while for the original data the structure emerges at longer time horizons (using stocks from the New York Stock Exchange).

Wavelet multiscale analysis has been used by numerous authors to decompose economic and financial time-series into orthogonal time-scale components of varying granularities. Gençay and co-workers have variously examined the scaling properties of foreign exchange volatility [25] and volatility models without intraday seasonalities [26] using wavelet multiscaling techniques. The systematic risk in a Capital Asset Pricing Model was estimated over different granularities [27] where it was shown that the return of a portfolio and its Beta[b] become stronger as the scale increases for the S&P 500.

An extensively studied characteristic of stock market behavior, is the increase of stock return cross correlations as the sampling time–scale increases, a phenomenon known as the *Epps* effect [28]. More recently, analysis of time-dependent correlations between high-frequency stocks demonstrated, however, that market reaction times have increased due to greater efficiency [29]. A diminution of the Epps effect with time is one consequence of increased market efficiency. Trading asynchronicity was demonstrated to be not solely responsible for the effect [30] with the characteristic time apparently independent of the trading frequency. Further, analysis using a toy-model of Brownian motion and memoryless renewal process [31] found an exact expression for the Epps frequency dependence, with reasonable fitting also for empirical data. In fact, the effect was shown [32] not to scale with market activity but to be due to reaction times, rather than market activity.

In this paper, we extend the multivariate correlation technique, first applied to complex EEG seizure data [12, 13, 15] and later to medium-frequency financial data [11]. This interdisciplinary approach, allows characterization of correlation changes in time. By decomposing equity returns into their component scales in short time-windows, we are able to study the correlation and associated eigenspectrum at various scales, allowing a time–scale analysis of the eigenspectrum. Using both medium- and high-frequency stock returns, we examine the dynamics of the large eigenvalues and demonstrate the time–scale dependence of the correlation structure. We show how the Epps effect [28] can be demonstrated through a multivariate approach and expand previous work by examining considerably longer scales. Finally, we look at some applications of the work, with scale dependent risk characterization and portfolio optimization examined. Sections 2 and 3 describe the

[b]The Beta of an asset is a measure of the volatility, or systematic risk, of a security or a portfolio in comparison to the market as a whole.

techniques and data used, Sec. 4 details the results obtained and conclusions are given in Sec. 5.

2. Methods

2.1. *Wavelet multiscale analysis*

Wavelets provide an efficient means of studying the multiresolution properties of a signal, allowing decomposition into different time horizons or frequency components (Discrete Wavelet Transform, DWT) [33,34]. The definitions of the two basic wavelet functions, the father ϕ and mother ψ wavelets are given as:

$$\phi_{j,k}(t) = 2^{\frac{j}{2}}\phi(2^j t - k), \tag{1}$$

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^j t - k), \tag{2}$$

where $j = 1, \ldots, J$ in a J-level decomposition. The father wavelet integrates to 1 and reconstructs the longest time–scale component of the series, while the mother wavelet integrates to 0 and is used to describe the deviations from the trend. The wavelet representation of a discrete signal $f(t)$ in $L^2(R)$ is:

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\phi_{J,k}(t) + \cdots + \sum_k d_{1,k}\phi_{1,k}(t), \tag{3}$$

where J is the number of multiresolution levels (or *scales*) and k ranges from 1 to the number of coefficients in the specified level. The coefficients $s_{J,k}$ and $d_{J,k}$ are the smooth and detail component coefficients, respectively, and given by

$$s_{J,k} = \int \phi_{J,k} f(t) dt, \tag{4}$$

$$d_{j,k} = \int \psi_{j,k} f(t) dt \quad (j = 1, \ldots, J). \tag{5}$$

Each of the coefficient sets $S_J, d_J, d_{J-1}, \ldots, d_1$ is called a *crystal*.

In this paper, we apply the Maximum Overlap Discrete Wavelet Transform (MODWT) [33,34], a linear filter that transforms a series into coefficients related to variations over a set of scales. Like the DWT it produces a set of time-dependent wavelet and scaling coefficients with basis vectors associated with a location t and a unitless scale $\tau_j = 2^{j-1}$ for each decomposition level $j = 1, \ldots, J_0$. The MODWT, unlike the DWT, has a high level of redundancy, however, is nonorthogonal and can handle any sample size N. It retains downsampled[c] values at each level of the decomposition that would be discarded by the DWT. This reduces the tendency for larger errors at lower frequencies when calculating frequency dependent variance and correlations (Sec. 2.3) as more data is available. Having decomposed the equity returns into their component time–scales, we then calculate the correlation between these components.

[c]Downsampling or decimation of the wavelet coefficients retains half of the number of coefficients that were retained at the previous scale and is applied in the DWT.

2.2. *Correlation dynamics*

The equal-time correlation matrix between time-series of stock returns is calculated using a sliding time-window where the number of stocks, N, is smaller than the window size T. Given time-series of stock returns $R_i(t)$, $i = 1, \ldots, N$, we normalize the time-series within each window as follows:

$$r_i(t) = \frac{R_i(t) - \widehat{R_i(t)}}{\sigma_i}, \tag{6}$$

where σ_i is the standard deviation of stock $i = 1, \ldots, N$ and $\widehat{R_i}$ is the time average of R_i over a time-window of size T. Then, the equal-time cross-correlation matrix, expressed in terms of $r_i(t)$, is

$$C_{ij} \equiv \langle r_i(t) r_j(t) \rangle. \tag{7}$$

The elements of C_{ij} are limited to the domain $-1 \leq C_{ij} \leq 1$, where $C_{ij} = 1$ defines perfect positive correlation, $C_{ij} = -1$ corresponds to perfect negative correlation and $C_{ij} = 0$ corresponds to no correlation. In matrix notation, the correlation is expressed as $\mathbf{C} = \frac{1}{T}\mathbf{R}\mathbf{R}^t$, where $\mathbf{R}$ is an $N \times T$ matrix with elements r_{it}.

The eigenvalues $\boldsymbol{\lambda_i}$ and eigenvectors $\hat{\mathbf{v}}_\mathbf{i}$ of the correlation matrix $\mathbf{C}$ are found using $\mathbf{C}\hat{\mathbf{v}}_\mathbf{i} = \boldsymbol{\lambda_i}\hat{\mathbf{v}}_\mathbf{i}$ and then ordered according to size, such that $\boldsymbol{\lambda_1} \leq \boldsymbol{\lambda_2} \leq \cdots \leq \boldsymbol{\lambda_N}$. Given that the sum of the diagonal elements of a matrix (the Trace) remains constant under a linear transformation, $\sum_i \lambda_i$ must always equal the trace of the original correlation matrix. Hence, if some eigenvalues increase then others must decrease, to compensate, and vice versa (*Eigenvalue Repulsion*).

There are two limiting cases for the distribution of the eigenvalues [12, 15], with perfect correlation, $C_i \approx 1$, when the largest is maximized with value N (all others taking value zero). When each time-series consists of random numbers with average correlation $C_i \approx 0$, the corresponding eigenvalues are distributed around 1 (where deviations are due to spurious random correlations). Between these two extremes, the eigenvalues at the lower end of the spectrum can be much smaller than $\lambda_{\max}$. By calculating the eigenspectrum associated with the correlation matrix at different scales in each time-window, we get time-series of eigenvalues $\lambda_{mn}(\tau)$, where m denotes the eigenvalue number and n is the scale studied in a particular time-window, τ.

To study the market behavior when eigenvalues take extreme values, we need to partition the later depending on their size. This is achieved by first normalizing each eigenvalue in time using

$$\bar{\boldsymbol{\lambda}}_\mathbf{i}(t) = \frac{(\boldsymbol{\lambda_i} - \bar{\boldsymbol{\lambda}}(\boldsymbol{\tau}))}{\sigma^{\lambda(\tau)}}, \tag{8}$$

where $\bar{\boldsymbol{\lambda}}(\boldsymbol{\tau})$ and $\sigma^{\lambda(\tau)}$ are the mean and standard deviation of the eigenvalues over a particular reference period, τ. By expressing the eigenvalue time-series in terms of standard deviation units (SDU), we can then easily partition each according to its relative size.

Combining the techniques of multivariate correlation analysis and multiscaling, we are able to provide some novel insight into the dependence of correlation on both time and scale. To enable this analysis, we first show how correlations are calculated using wavelet coefficients.

2.3. *Wavelet covariance and correlation*

The wavelet covariance between functions $f(t)$ and $g(t)$ is defined to be the covariance of the wavelet coefficients at a given scale. The *unbiased* estimator of the wavelet covariance at the jth scale is given by

$$\nu_{fg}(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^{f(t)} \tilde{D}_{j,t}^{g(t)}, \tag{9}$$

where all the wavelet coefficients affected by the boundary are removed [33] and $M_j = N - L_j + 1$. The wavelet variance at a particular scale $\nu_f^2(\tau_j)$ is found similarly.

The MODWT estimate of the wavelet cross correlation between functions $f(t)$ and $g(t)$ may then be calculated using the wavelet covariance and the square root of the wavelet variance of the functions at each scale j. The MODWT estimator [34] of the wavelet correlation is then given by:

$$\rho_{fg}(\tau_j) = \frac{\nu_{fg}(\tau_j)}{\nu_f(\tau_j)\nu_g(\tau_j)}. \tag{10}$$

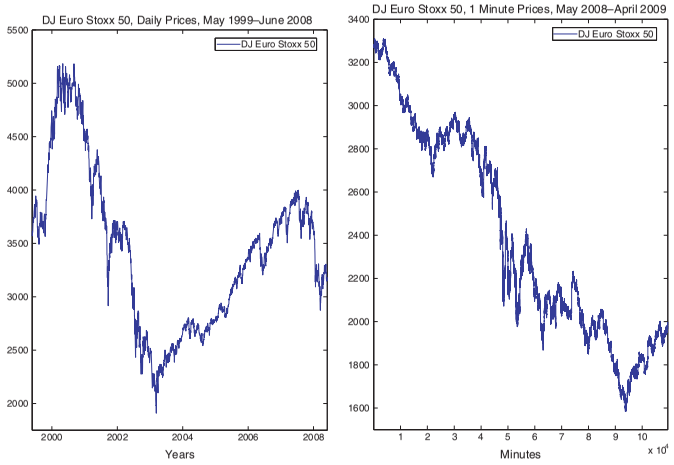
3. Data

The initial data set comprises the 49 equities of the Dow Jones (DJ) Euro Stoxx 50 where full price data is available from May 1999 to August 2007, resulting in 2183 daily returns, Fig. 1(a). The DJ Euro Stoxx 50 is a stock index of Eurozone equities, designed to provide a blue-chip representation of Eurozone supersector leaders. The small number of stocks in the index allows calculation of the cross-correlation matrices for small time-windows, without reducing the matrix rank.

The second data set studied consists of high-frequency returns for the DJ Euro Stoxx 50 from May 2008 to April 2009, resulting in 109,540 one-minute Returns, Fig. 1(b). High-frequency data allows us to examine additional features and analyze whether the correlation dynamics previously found [11] are inherent at all time-scales or if there is a gradual emergence. The time-frame studied is of particular interest, with sustained market drops and high levels of volatility. Understanding market interactions in a period such as this may assist in forming portfolios that are robust to large downwards moves.

4. Results

To decompose the data into component time–scales, we selected the least asymmetric (LA) wavelet (known as the Symmlet), which exhibits near symmetry about



(a) (b)

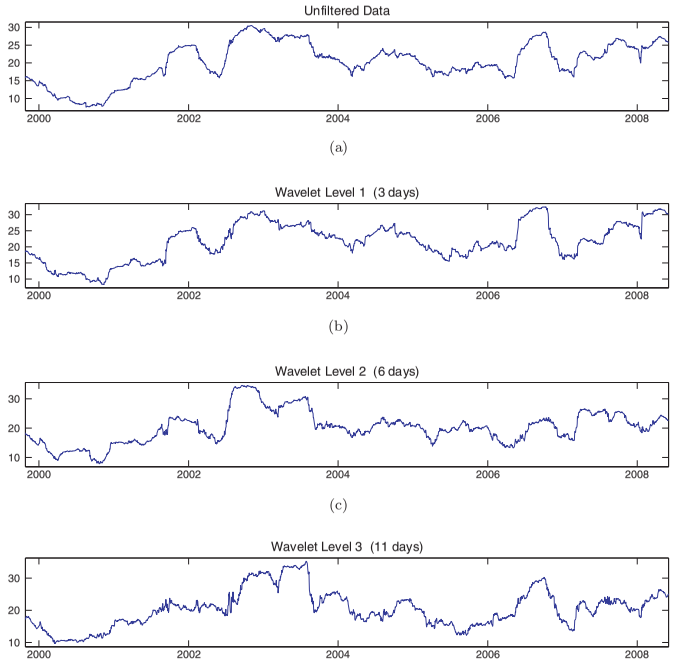
Fig. 1. (a) DJ Euro Stoxx 50, daily prices May 1999–June 2008. (b) DJ Euro Stoxx 50, one-minute prices May 2008–April 2009.

the filter midpoint. LA filters are available in even widths and the optimal filter width is dependent on the characteristics of the signal and the length of the data. The filter width chosen for this study was the LA8 (where the 8 refers to the width of the scaling function) since it enabled us to accurately calculate wavelet correlations to relatively long scales, dependent upon the time-window used. Although the MODWT can accommodate any level J_0, in practice the largest level is chosen so as to prevent decomposition at scales longer than the total length of the data series.

4.1. *Medium-frequency eigenvalue dynamics*

We first focus on medium-frequency equity returns measured at a one-day horizon. For each stock, using a sliding time-window of 100 days, we decompose the returns into their component scales in each window, using the wavelet transform. The correlation between the wavelet coefficients, corresponding to each equity, is then calculated at each scale, as described (Sec. 2.3). Time-series of eigenvalues are found at each scale, by calculating the associated eigenspectrum. This approach of calculating the eigenspectrum in sliding time-windows, allows analysis of the dependence of correlation dynamics on granularity.

The time-window was chosen such that $Q = \frac{T}{N} = 2.04$, thus ensuring that the data would be close to stationary in each sliding window (different values of Q were examined [11] with little variation in the results). Figure 1(a) shows the value of the DJ Euro Stoxx Index over the period studied. Figure 2(a) displays the largest eigenvalue, calculated using the unfiltered (one day) time-series data for sliding time-windows. As shown, the largest eigenvalue is far from static, rising from a minimum of 7.5 to a maximum of 30.5 from early 2001 to late 2003 (coinciding with the bursting of the "tech" bubble). This corresponds to an increase in the influence of the "Market," with the behavior of traders becoming more correlated. The next major increase occurred in early 2006, followed by a relatively marked



(a) (b) (c) (d)

Fig. 2. (a) Largest eigenvalue dynamics original data, (b) 3 day scale, (c) 6 day scale, and (d) 11 day scale.

decline until the beginning of the "Credit Crunch" in 2007. Similar to Ref. 17, we note an increase in the value of the largest eigenvalue during times of market stress, with lower values during more "normal" periods.

We next calculate, using the MODWT (Sec. 2.1), the value of the largest eigenvalue of the cross-correlation matrix over longer time horizons of 3, 6, and 11 days (Figs. 2(b)–2(d)). Certain traders (such as Hedge Fund managers) may have very short trading horizons while others (such as Pension Fund managers) have much longer horizons. By looking at the value of the largest eigenvalue at different scales, we try to characterize the impact of these different trading horizons on the cross-correlation dynamics between large capitalization stocks. In Figs. 2(b)–2(d), we see that the main features found in the unfiltered data are preserved over longer time–scales. However, certain features, such as the sizeable drop in the largest eigenvalue at the longest scale in late 2003, are not seen at shorter scales, but the aggregate impact for the unfiltered data is a moderate drop. Other features, such as the increase in 2006, are not preserved across all scales. The different features, found at various scales, suggest that the correlation matrix is made up of interactions between stocks, traded by investors with different time horizons. This has implications for risk management, as the correlation matrix used for input in a portfolio optimization should depend on the investor's time horizon. We look more closely at this in Sec. 4.5.

4.2. *High-frequency correlation analysis*

Using the entire high-frequency data set for the Euro Stoxx 50, described earlier, we calculated the average correlation and eigenvalue spectrum at a number of scales, Table 1. As expected, the largest eigenvalue increases as the average correlation between stocks increases, in keeping with the "one-factor" toy-model of correlations described previously [11]. Analysis of Table 1 gives some intuition for the results shown, Fig. 2, with increases in the largest eigenvalue in time corresponding to an increase in the average correlation between stocks ("single factor" model [11]).

As mentioned (Sec. 1), the Epps effect is the gradual build-up of correlation as the scale at which stock returns are measured, increases. In previous work [29, 30],

Table 1. Correlation and eigenspectrum analysis.

Scale	Time Horizon	Average Correlation	1st Eigenvalue	2nd Eigenvalue	3rd Eigenvalue
1	1 min	0.35	20.24	6.66	1.085
2	5 min	0.39	20.66	5.53	1.26
3	30 min	0.45	23.33	4.229	1.64
5	120 min	0.47	24.7	4.18	1.74
8	480 min	0.52	27.72	3.54	2.42
10	1100 min	0.51	27.15	3.83	1.97
12	2000 min	0.41	18.42	2.78	2.39

The 1st eigenvalue is the largest eigenvalue of the correlation matrix of Euro Stoxx 50 high-frequency data from May 2008 to April 2009.

the Epps effect has been studied using either the bivariate correlation between pairs of stocks or the average of a number of such pairs. However, as seen in Table 1, this effect is also visible using a multivariate approach, with the largest eigenvalue increasing from 20.24 at the smallest scale to 27.72 at the 480 min ($\approx$ 1 day) scale, with corresponding increase in average system correlation (from 0.35 to 0.52).

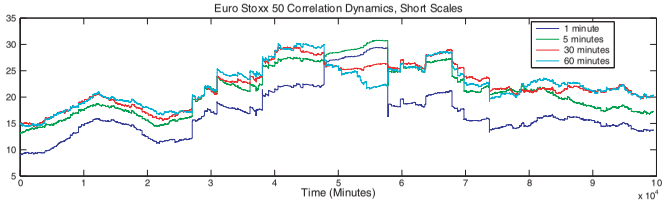
In previous studies, authors have calculated the correlation at different scales by recalculating the returns, resulting in reduced quantities of available data, and constraining the maximum possible scale. However, using Wavelet Multiscaling to decompose the data into component scales, we can calculate the correlations and eigenvalues at much longer scales. Of particular interest is the increase in the average correlation to a maximum at a scale of 480 min and the subsequent decrease at longer scales of 2000 min ($\approx$ 4 days) to 0.41. This decrease in correlation corresponds to a decrease in the "Market Effect," with traders having longer time-horizons acting in a less synchronous fashion. The decrease may be due to the additional discontinuities incorporated in the data at scales longer than one day as additional information is incorporated as stock markets open for the day (visible in discontinuous jumps in the equity prices).

4.3. *High-frequency correlation dynamics*

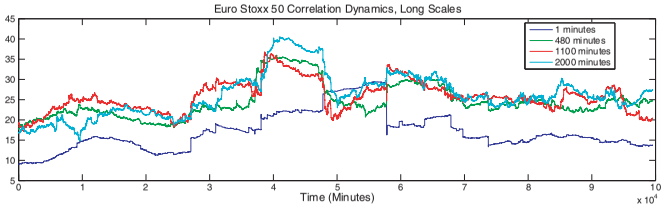
Building on the analysis of previous sections, we now examine the dynamics of the eigenvalues over time using the high-frequency data described. The methodology is the same as for daily data, described in Sec. 4.1, using a time-window of 10,000 min. The results for small scales are shown, Fig. 3(a), with a general increase in the largest eigenvalue for longer scales, corresponding to the increase in correlation shown in Sec. 4.2. However, for small scales (1–5 min), some distinct dynamics are found, with a sharp increase at around 50,000 min and a decrease at longer scales (30–60 min). These distinct dynamics emerge even more distinctly at longer scales, Fig. 3(b), with eigenvalues actually having opposite behavior at certain points (the largest eigenvalue increases at the one-minute scale but actually decreases at the longest scales, at approximately 50,000 min).

Of particular interest is the gradual increase in the largest eigenvalue moving to longer scales, in keeping with the Epps effect. As found in the previous section at longer scales (2,000 min), using the complete data set, the largest eigenvalue is smaller than for shorter scales at certain points in time. This variation in the largest eigenvalue corresponds to a change in the average correlation between stocks. The distinct behavior at longer time–scales may be due to a number of factors, with longer scales incorporating additional information and smoothing out high-frequency noise.

The dynamics of the 2nd and 3rd largest eigenvalue are also shown in Fig. 4. Again, the relative size and dynamics of the eigenvalues are found to be scale-dependent, with a broad increase in the eigenvalues for longer scales. This is in sharp contrast to the results found in Sec. 4.2, where the 2nd largest eigenvalue was found



(a)



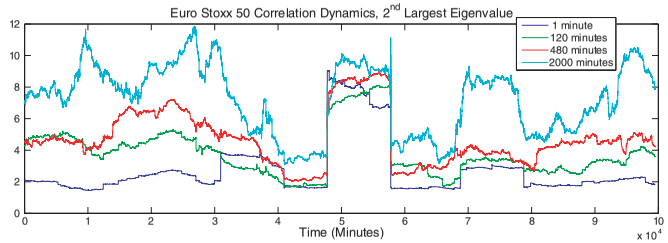
(b)

Fig. 3. Eigenvalues dynamics.

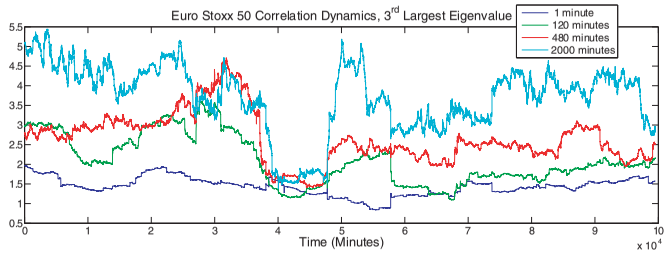
to strictly decrease as the scale increased. It appears that for small windows, the *second largest eigenvalue is of greater relative importance* (at longer scales) with a corresponding redistribution of correlation structure across the eigenspectrum, (eigenvalue repulsion). In the context of previous work done on RMT [4, 9], this may imply that different information is to be found in the 2nd and subsequent eigenvalues at longer scales. These eigenvalues are said to correspond to different sectoral groupings, with possible group changes as we move down the scales. By removing the market mode, researchers [24] found the correlation structure to be invariant across scales. However, this paper did not examine different time-windows or longer time-scales (> 255 mins). By combining the method of these authors with those proposed above, it may be possible to investigate the invariance of correlations across time as well as scale.

4.4. *Drawdown analysis*

As indicated (Sec. 1 and Refs. 11, 17) drawdowns (or periods of large negative returns) and drawups (periods of large positive returns) tend to be accompanied by an increase in different eigenstates of the cross-correlation matrix. In this section, we attempt to characterize the market according to the relative size of the eigenvalues, as well as through the use of eigenvalue ratios.



(a)



(b)

Fig. 4. Eigenvalues dynamics, 2nd and 3rd largest eigenvalues.

The returns, correlation matrix, and eigenvalue spectrum time-series for overlapping windows of 10,000 min were calculated and normalized using the mean and standard deviation over the entire series [Eq. (8)]. By representing normalized eigenvalues in terms of SDU, we can partition the eigenvalues according to their magnitude. The average return of the index is shown in Table 2 (during periods when the largest eigenvalue is ± 1 SDU). As the market trend was predominantly negative, average returns were negative both for periods when the largest eigenvalue was equal to its minimum and maximum values. However, distinct behavior still emerged across all scales.

Looking first at the original or unfiltered data, we see that when the largest eigenvalue is > 1 SDU, the return of the index was -8.06%. In contrast, when the largest eigenvalue is < -1 SDU, the return was -3.06%. Moving up the scales, similar behavior was found with large eigenvalues being characterized by more pronounced downward moves. The emergence of this characteristic at very small scales (≈ 5 min), implies that it is not a result of data granularity.

Table 2. Drawdown/drawup analysis.

Scale	No. Std	> 1	< -1
Original		− 8.06%	−3.39%
5 min		− 7.95%	−3.06%
60 min		− 9.27%	−3.64%
360 min		− 8.41%	−3.52%
1100 min		−10.79%	0.01%
2000 min		−12.06%	−6.32%

Average index returns when various eigenvalue partitions in SDU are > 1 and < -1.

4.5. *Portfolio optimization*

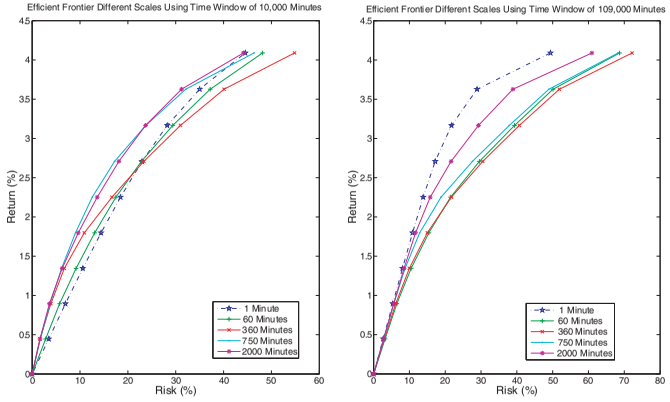
As mentioned previously, the time–scale dependence of correlation structure has implications for risk management. In this Section we demonstrate, using high-frequency one-minute stock returns, that the optimal portfolio found from classic portfolio optimization is time–scale dependent. The correlation matrix between stocks of the Euro Stoxx 50 was calculated for two time-frames, the first being a short time-window of 10,000 min (corresponding to the first time-window studied in Sec. 4.3), the second the complete data set of 109,540 min (corresponding to the data studied in Sec. 4.2). The portfolio optimization was unconstrained, allowing "short-selling" of stocks, which means that positive expected returns for the portfolio were possible, even though the majority of expected returns were negative. The expected returns and risk for the equities were the same for both optimizations, meaning that all deviations in the risk/return profile are due to changes in the correlation structure.

The portfolio optimization was performed in a standard fashion [35, 37, 38] and the results, shown in Fig. 5, are highly dependent on the time–scale studied. Interestingly, for the smaller time-window of 10,000 min, the portfolios tended to be less risky at longer scales. In contrast, using the larger time-window to calculate the correlations between stocks, the least risky portfolio was found using a one-minute scale. Portfolios calculated at the 60–750 minute scales were the riskiest, with longer-period scales found to be less risky. This is in keeping with the findings of Sec. 4.2, where the average correlation peaked at a scale of 480 min, meaning that further diversification benefits (lower correlation) can be found at the shortest and longest scales.

The optimal portfolio is clearly dependent on the time–scale used to calculate the correlation matrix, which in turn, is dependent on the time-window used. Obviously, this means that care must be taken by investors in the choice of both time-window and granularity used in the calculation of correlation matrices.

5. Conclusions

Using a multivariate technique first applied to complex EEG seizure data, the correlation structure between medium- and high-frequency financial time-series was



(a) (b)

Fig. 5. Portfolio optimization.

studied. Application of the MODWT extended the study of changes in the cross-correlation structure across both scale and time.

(1) Using the MODWT and a sliding window, the dynamics of the largest eigenvalue of the correlation matrix were examined and shown to be time-dependent at all scales (using both medium- and high-frequency equity returns). Similar dynamics were visible across all scales, but with particular features markedly apparent at certain scales. This suggests that the correlation matrix between equities consists of interactions caused by traders with different time horizons.
(2) Study of the 2nd and 3rd largest eigenvalues over both time and scale revealed a large increase in sectoral correlations for longer scales, using high-frequency data.
(3) High-frequency data were used to calculate the average correlation and largest eigenvalue for various scales. The *Epps* effect was found using this multivariate technique and analysis of longer scales revealed a drop in the average system correlation at scales beyond one day.
(4) A partition of the time-normalized eigenvalues demonstrated quantitatively the relationship between the size of the largest eigenvalue and the return of the index. This feature emerged at very small scales (< 5 min), implying that this relationship is not due to the granularity of the data studied.
(5) Finally, we looked at the problem of portfolio optimization at various scales. Results show that the optimal portfolio depends not only on scale but also on the time-window used in the calculation of the correlation matrix.

The correlation technique used in this paper to measure the interaction between agents suffers from the drawback of being linear and hence neglects any higher-order relationships. Future work includes the application of nonlinear information-theory based dependence measures, which will allow the detection of complex changes in synchronization behavior around extreme financial events. These nonlinear techniques may uncover further emergent features to those highlighted above and may result in signals that warn of likely future market turmoil. The interaction of various agents with competing strategies and operating over different time–scales, requires further investigation in order to improve understanding of complex phenomema such as financial crashes. Additionally, while this financial work has benefited from parallels drawn with EEG system analysis, insights gained here may also raise questions with regards to the behavior of other complex systems.

References

[1] Laloux, L., Cizeau, P., Bouchaud, J. and Potters, M., *Phys. Rev. Lett.* **83**(7) (1999) 1467–1470.
[2] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Guhr, T. and Stanley, H. E., *Phys. Rev. Lett.* **83** (1999) 1471–1474.
[3] Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J., *Int. J. Theoret. Appl. Finance* 3 (3) (2000) 391–397.
[4] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. and Stanley, H. E., *Phys. Rev. E* **65** (2000) 066126.
[5] Gopikrishnan, P., Rosenow, B., Plerou, V. and Stanley, H. E., *Phys. Rev. E* **64** (2001) 035106R.
[6] Utsugi, A., Ino, K. and Oshikawa, M., *Phys. Rev. E* **70** (2004) 026110.
[7] Wilcox, D. and Gebbie, T., *Physica A* **344**(1–2) (2004) 294–298.
[8] Sharifi, S., Crane, M., Shamaie, A. and Ruskin, H. J., *Physica A* **335**(3–4) (2004) 629–643.
[9] Conlon, T., Ruskin, H. J. and Crane, M., *Physica A* **382**(2) (2007) 565–576.
[10] Conlon, T., Crane, M. and Ruskin, H. J., *Physica A* **387**(21) (2008) 5197–5204.
[11] Conlon, T., Ruskin, H. J. and Crane, M., *Physica A* **388**(5) (2009) 705-714.
[12] Müller, M., Baier, G., Galka, A., Stephani, U. and Muhle, H., *Phys. Rev. E* **71** (2005) 046116.
[13] Müller, M., Jiménez, Y. L., Rummel, C., Baier, G., Galka, A., Stephani, U. and Muhle, H., *Phys. Rev. E* **74** (2006) 041119.
[14] Conlon, T., Ruskin, H. J. and Crane, M., Seizure characterisation using frequency dependent multivariate dynamics, *Comput. Bio. Med.* (2009) (to appear).
[15] Schindler, K., Leung, H., Elger, C. E. and Lehnertz, K., *Brain* **130** (2007) 65–77.
[16] Kwapien, J., Drozdz, S. and Ionannides, A. A., *Phys. Rev. E* **62** (2000) 5557–5564.
[17] Drozdz, S., Gruemmer, F., Ruf, F. and Speth, J., *Physica A* **287** (2000) 440–449.
[18] Drozdz, S., Gruemmer, F., Ruf, F. and Speth, J., *Physica A* **294** (2001) 226–234.
[19] Guhr, T. and Kälber, B., *J. Phys. A* **35** (2003) 3009–3032.
[20] Burda, Z., Görlich, A., Jarosz, A. and Jurkiewicz, J., *Physica A* **343** (2004) 295–310.
[21] Burda, Z. and Jurkiewicz, J., *Physica A* **344**(1–1) (2004) 67–72.
[22] Malevergne, Y. and Sornette, D., *Physica A* **331**(3–4) (2004) 660–668.
[23] Kwapien, J., Drozdz, S. and Oswiecimka, P., *Physica A* **359** (2006) 589–606.
[24] Borghesi, C., Marsili, M. and Micciche, S., *Phys. Rev. E* **76** (2007) 0261104.
[25] Gençay, R., Selcuk, F. and Whitcher, B., *Physica A* **289** (2001) 249–266.

C-24

1st Reading

August 21, 2009 20:16 WSPC/169-ACS 00232

[26] Gençay, R., Selcuk, F. and Whitcher, B., *Physica A* **289** (2001) 543–556.
[27] Gençay, R., Selcuk, F. and Whitcher, B., *Quant. Financ.* **3** (2003) 108–116.
[28] Epps, T. W., *J. Am. Stat. Assoc.* **74** (1979) 291–298.
[29] Tóth, B. and Kertész, J., *Physica A* **360** (2006) 505–515.
[30] Tóth, B. and Kertész, J., *Physica A* **383** (2007) 54–84.
[31] Tóth, B. and Kertész, J., *Proceedings of SPIE* (Fluctuations and Noise 2007), 6601 (2007).
[32] Tóth, B. and Kertész, J., *Quant. Finance* (Feb. 2007).
[33] Percival, D. B. and Walden, A. T., *Wavelet Methods for Time Series Analysis* (Cambridge University press, 2000).
[34] Gençay, R., Selcuk, B. and Whitcher, B., *An Introduction to Wavelets and other Filtering Methods in Finance and Economics* (Academic Press, 2001).
[35] Sharpe, W. F., *J. Finance* **35** (1964) 910–919.
[36] Markowitz, H., *Portfolio Selection: Efficient Diversification of Investments* (Wiley, 1959).
[37] Elton, E. J. and Gruber, M. J., *Modern Portfolio Theory and Investment Analysis*, 6th edn. (Wiley, 2002).
[38] Bouchaud, J. P. and Potters, M, *Theory of Financial Risk and Derivative Pricing* (Cambridge University Press, 2004).

Computers in Biology and Medicine 39 (2009) 760–767

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Seizure characterisation using frequency-dependent multivariate dynamics

T. Conlon*, H.J. Ruskin, M. Crane

Dublin City University, Glasnevin, Dublin 9, Ireland

ARTICLE INFO

Article history:
Received 16 January 2009
Accepted 4 June 2009

Keywords:
Cross-correlation
Wavelet multiscaling
EEG
Epilepsy
Eigenspectrum

ABSTRACT

The characterisation of epileptic seizures assists in the design of targeted pharmaceutical seizure prevention techniques and pre-surgical evaluations. In this paper, we expand on the recent use of multivariate techniques to study the cross-correlation dynamics between electroencephalographic (EEG) channels. The maximum overlap discrete wavelet transform (MODWT) is applied in order to separate the EEG channels into their underlying frequencies. The dynamics of the cross-correlation matrix between channels, at each frequency, are then analysed in terms of the eigenspectrum. By examination of the eigenspectrum, we show that it is possible to identify frequency-dependent changes in the correlation structure between channels which may be indicative of seizure activity.

The technique is applied to EEG epileptiform data and the results indicate that the correlation dynamics vary over time and frequency, with larger correlations between channels at high frequencies. Additionally, a *redistribution* of wavelet energy is found, with increased fractional energy demonstrating the relative importance of high frequencies during seizures. Dynamical changes also occur in both correlation and energy at lower frequencies during seizures, suggesting that monitoring frequency-dependent correlation structure can characterise changes in EEG signals during these. Future work will involve the study of other large eigenvalues and inter-frequency correlations to determine additional seizure characteristics.

1. Introduction

In recent years, the application of the equal-time correlation matrix to a variety of multivariate data sets such as financial data [1–4], electroencephalographic (EEG) recordings [5–11], magnetoencephalographic (MEG) recordings [12] and others has been studied extensively. The dynamics of such systems are characterised by a continuously varying level of synchronisation between different subsets of the system. The degree of synchronisation is dependent on the length of time-series studied, the granularity of the data and the amount of noise in the system, amongst other things.

Equal-time cross-correlation matrices have been shown to characterise dynamical changes in nonstationary multivariate time-series [8]. It was demonstrated that, as the synchronisation of k individual time-series within an M-dimensional multivariate time-series increases, this causes a repulsion between eigenstates of the correlation matrix, in which k levels participate. Through the use of artificially created time-series with pre-defined correlation dynamics, it was found that there exist situations where the relative change of eigenvalues from the lower edge of the spectrum is greater than that of the large eigenvalues, implying that the information drawn from the smaller eigenvalues is more relevant as an indicator of event activity. An examination of shifts in the eigenspectrum of the cross-correlation matrix between epileptic seizure time-series also revealed that information about the correlation dynamics is reflected in both the lower and the upper eigenstates.

The application of the equal-time correlation matrix to EEG seizure time-series was used to study temporal dynamics for a large sample of focal onset epileptic seizures [10]. It was shown that the zero-lag correlations between multichannel EEG signals tend to decrease during the first half of a seizure and increase gradually before the seizure ends. This work was further extended to the case of *status epilepticus*, where the equal-time cross-correlation matrix was used to assess neuronal synchronisation prior to seizure termination [11].

It was demonstrated, for particular examples, that information about cross-correlations can be found in the random matrix theory (RMT) defined bulk of eigenvalues and that the information extracted at the lower edge is statistically more significant than that extracted from the larger eigenvalues [9]. The method was applied to multichannel EEG data, where the dynamics of the smallest eigenvalues were shown to be more sensitive to the detection of subtle changes in the brain dynamics than those of the largest eigenvalues.

The use of wavelet techniques in the study of EEG signals is widespread, allowing a time-frequency decomposition of these non-stationary signals. Wavelets [13–16] were used to determine localisation of transient signals (spikes) during ictal periods, vital in preoperative evaluation of the foci and propagation of ictal events. Furthermore, a wavelet-based similarity method across frequencies was described, using ideas from nonlinear dynamics to predict epileptic seizures [17].

Various authors have attempted to measure the interdependencies between different regions of the brain. A number of these techniques were compared, with their relative performance shown to be dependent on the form of the underlying signals [6,7]. In the case of MEG signals, a limited study consisting of three patients was performed, using wavelets to determine cross-correlations over different frequencies and calculate the time lag between different brain regions [18]. The application of the *coherence function* and cross-correlation between different frequency bands defined by a continuous filter bank was also used to explore the time-frequency dependence of epileptic seizure data [19].

In this paper, we seek to expand these techniques through the application of time-frequency decomposition to the multivariate analysis of EEG channel interdependency. Data considered are from a number of patients suffering from *focal*, *generalised* or *secondary generalised* epileptic seizures. The frequency dependence of this multivariate technique adds an additional dimension to previous studies, with distinct behaviour of the eigenvalue spectrum and associated correlation between channels demonstrated across different frequencies. This paper is organised as follows. Methods are reviewed in Section 2. In Section 3 we describe the data studied, while in Section 4 we summarise the results obtained and discuss their importance.

2. Methods

2.1. Correlation dynamics

Equal-time cross-correlation matrices between EEG time-series are calculated using a sliding time window where the number of channels, N, is smaller than the window size T. Given EEG time-series $S_i(t)$, $i=1,\ldots,N$, we normalise these within each window as follows:

$$s_i(t) = \frac{S_i(t) - \widehat{S_i}}{\sigma_i} \tag{1}$$

with σ_i the standard deviation and $\widehat{S_i}$ the time average of S_i over a time window of size T, for channel $i=1,\ldots,N$.

The equal-time cross-correlation matrix is expressed in terms of $s_i(t)$

$$C_{ij} \equiv \langle s_i(t) s_j(t) \rangle \tag{2}$$

The elements of C_{ij} are limited to the domain $-1 \le C_{ij} \le 1$, where $C_{ij}=1$ defines perfect positive correlation, $C_{ij}=-1$ corresponds to perfect negative correlation and $C_{ij}=0$ to no correlation [1,2]. The correlation matrix can be expressed in matrix notation as

$$\mathbf{C} = \frac{1}{T}\mathbf{S}\mathbf{S}^{\tau} \tag{3}$$

where $\mathbf{S}$ is an $N \times T$ matrix with elements s_{it}.

The eigenvalues $\boldsymbol{\lambda}_i$ and eigenvectors $\hat{\mathbf{v}}_i$ of the correlation matrix $\mathbf{C}$ are found from the following:

$$\mathbf{C}\hat{\mathbf{v}}_i = \boldsymbol{\lambda}_i \hat{\mathbf{v}}_i \tag{4}$$

The eigenvalues are then ordered by size, such that $\boldsymbol{\lambda}_1 \le \boldsymbol{\lambda}_2 \le \cdots \le \boldsymbol{\lambda}_N$. The trace of a matrix, i.e. the sum of the diagonal elements, must always remain constant under linear transformation. Thus, the sum of the eigenvalues must always equal the trace of the original correlation matrix [26]. Hence, if some eigenvalues increase then others must decrease, to compensate, and vice versa.

There are two limiting cases for the distribution of the eigenvalues [8,10]. When all of the time-series are perfectly correlated, $C_i \approx 1$, the largest eigenvalue is maximised with a value equal to N, while all other eigenvalues take value zero. When each of the time-series consists of random numbers with average correlation $C_i \approx 0$, the corresponding eigenvalues are distributed around 1 (where any deviation is due to spurious random correlations).

For cases between these two extremes, we can use the eigenvalue spectrum, calculated using a moving window, to detect changes in the correlation structure. If all channels have an average cross-correlation ρ_0, then the largest eigenvalue is of order $(N-1)\rho_0 + 1$ where N is the number of channels. As this average or *global* correlation between channels increases, the largest eigenvalue increases (to a maximum N), with a corresponding decrease in the small eigenvalues (*eigenvalue repulsion*). Other cases, where correlation structure was more intricate, have been described [8], but in this paper we concentrate on the *global correlation* between channels.

As indicated, the small eigenvalues decrease when the global correlation between channels increases, resulting in a large disparity between the relative size of the small and large eigenvalues. To allow the comparison of eigenvalues from both ends of the spectrum, we normalise each eigenvalue in time using

$$\tilde{\boldsymbol{\lambda}}_i(t) = \frac{(\boldsymbol{\lambda}_i(t) - \bar{\boldsymbol{\lambda}})}{\sigma^{\lambda}} \tag{5}$$

where $\bar{\boldsymbol{\lambda}}$ and σ^{λ} are the mean and standard deviation of the eigenvalues over a particular reference period [10]. This normalisation allows us to visually compare eigenvalues at both ends of the spectrum, even if their magnitudes are significantly different. The reference period used to calculate mean and standard deviation of the eigenvalue spectrum can be chosen to be a low volatility sub-period (which helps to enhance the visibility of high volatility periods), or the full-time period studied. By normalising the eigenvalues over time, it becomes possible to identify epileptiform events, even using the small eigenvalues.

2.2. Wavelet multiscale analysis

Wavelets provide an efficient means of studying multiresolution properties, as they can be used to decompose a signal into different time horizons or frequency components. The discrete wavelet transform (DWT), in particular, allows the decomposition of a signal into components of different frequency [20–22]. There are two basic wavelet functions, the father wavelet ϕ and the mother wavelet ψ. The formal definitions of the father and mother wavelets are the functions

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \tag{6}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{7}$$

where $j=1,\ldots,J$ in a J-level decomposition. The father wavelet integrates to 1 and reconstructs the longest time-scale component of the series, whereas the mother wavelet integrates to 0 and is used to describe the deviations from the trend. The wavelet representation of a discrete signal $f(t)$ in $L^2(R)$ is given by

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\phi_{J,k}(t) + \cdots + \sum_k d_{1,k}\phi_{1,k}(t) \tag{8}$$

where J is the number of multiresolution levels (or scales) and k ranges from 1 to the number of coefficients in the specified level. The coefficients $s_{J,k}$ and $d_{J,k}$ are the smooth and detail component coefficients, respectively, and are given by

$$s_{J,k} = \int \phi_{J,k} f(t)\,dt \tag{9}$$

$$d_{j,k} = \int \psi_{j,k} f(t)\,dt \quad (j=1,\ldots,J) \tag{10}$$

Each of the coefficient sets $S_J, d_J, d_{J-1}, \ldots, d_1$ is called a *crystal*.

* Corresponding author.
E-mail addresses: tconlon@computing.dcu.ie (T. Conlon), hruskin@computing.dcu.ie (H.J. Ruskin), mcrane@computing.dcu.ie (M. Crane).

0010-4825/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2009.06.003

Author's personal copy

2.2.1. MODWT

The maximum overlap discrete wavelet transform (MODWT) is a linear filter that transforms a series into coefficients related to variations over a set of scale [22,23]. Like the DWT it produces a set of time-dependent wavelet and scaling coefficients with basis vectors associated with a location t and a unitless scale $\tau_j = 2^{j-1}$ for each decomposition level $j = 1, \ldots, J_0$. The MODWT, unlike the DWT, has a high level of redundancy, however, and is nonorthogonal. It retains downsampled[1] values at each level of the decomposition that would be discarded by the DWT. The MODWT can also handle any sample size N, whereas the DWT restricts the sample size to a multiple of 2^J. Here, we apply the MODWT as it helps reduce the errors associated with the calculation of wavelet correlation at different scales (due to the availability of greater amounts of data at longer scales).

Decomposing a signal using the MODWT to J levels theoretically involves the application of J pairs of filters. The filtering operation at the jth level consists of applying a rescaled father wavelet to yield a set of *detail coefficients*

$$\tilde{D}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{\psi}_{j,l} f_{t-l} \tag{11}$$

and a rescaled mother wavelet to yield a set of *scaling coefficients*

$$\tilde{S}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{\phi}_{j,l} f_{t-l} \tag{12}$$

for all times $t = \ldots, -1, 0, 1, \ldots$, where f is the function to be decomposed [22]. The rescaled mother, $\tilde{\psi}_{j,l} = \psi_{j,l}/2^j$, and father, $\tilde{\phi}_{j,t} = \phi_{j,l}/2^j$, wavelets for the jth level are a set of scale-dependent localized differencing and averaging operators and can be regarded as rescaled versions of the originals. The jth level equivalent filter coefficients have a width $L_j = (2^j - 1)(L-1)+1$, where L is the width of the $j=1$ base filter. In practice the filters for $j > 1$ are not explicitly constructed because the detail and scaling coefficients can be calculated, using an algorithm that involves the $j = 1$ filters operating recurrently on the jth level scaling coefficients, to generate the $j+1$ level scaling and detail coefficients [22].

The MODWT is an energy conserving decomposition [22]:

$$\|\mathbf{f}\|^2 = \sum_{j=1}^{J} \|\tilde{\mathbf{D}}_j\|^2 + \|\tilde{\mathbf{S}}_J\|^2 \tag{13}$$

This decomposition allows the measurement of the contribution to the energy due to changes at scale 2^{j-1}. A fractional energy, $\tilde{E}_j$, associated with each scale can be calculated using

$$\tilde{E}_j = \frac{E_j}{E_{tot}} \quad \text{with } E_{tot} = \sum_{j=1}^{J} E_j. \tag{14}$$

for scales $j = 1, \ldots, J$.

2.2.2. Wavelet variance

The wavelet variance $\nu_f^2(\tau_j)$ is defined as the expected value of $\tilde{D}_{j,t}^2$ if we consider only the non-boundary coefficients.[2] An *unbiased* estimator of the wavelet variance is formed by removing all coefficients that are affected by boundary conditions and given by

$$\nu_f^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^2 \tag{15}$$

where $M_j = N - L_j + 1$ is the number of non-boundary coefficients at the jth level [22]. The wavelet variance decomposes the variance of a process on a scale-by-scale basis (at increasingly higher resolutions of the signal) and allows us to explore how a signal behaves over different time horizons.

2.2.3. Wavelet covariance and correlation

The wavelet covariance between functions $f(t)$ and $g(t)$ is similarly defined to be the covariance of the wavelet coefficients at a given scale. The *unbiased* estimator of the wavelet covariance at the jth scale is given by

$$\nu_{fg}(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^{f(t)} \tilde{D}_{j,t}^{g(t)} \tag{16}$$

where all the wavelet coefficients affected by the boundary are removed [22], and $M_j = N - L_j + 1$.

The MODWT estimate of the wavelet cross-correlation between functions $f(t)$ and $g(t)$ may be calculated using the wavelet covariance and the square root of the wavelet variance of the functions at each scale j. The MODWT estimator [23] of the wavelet correlation is given by

$$\rho_{fg}(\tau_j) = \frac{\nu_{fg}(\tau_j)}{\nu_f(\tau_j)\nu_g(\tau_j)} \tag{17}$$

where $\nu_{fg}(\tau_j)$ is the covariance between $f(t)$ and $g(t)$ at scale j, $\nu_f(\tau_j)$ is the variance of $f(t)$ at scale j and $\nu_g(\tau_j)$ is the variance of $g(t)$ at scale j.

2.3. Defining seizure onsets and endings

Following [10], we attempted to isolate seizure beginning and end in a reproducible way. First, the absolute slope, $S_i(t) = |\Delta EEG_i(t)/\Delta t|$, was computed over each channel i. The slope was normalised, $\tilde{S}_i(t) = S_i(t)/\sigma_i$, with σ_i the standard deviation of the signal over a reference period containing no seizure activity. $\tilde{S}_i(t)$ was then smoothed using a moving average over a time window of 5 s. *Epiteliform activity* was defined by finding a slope greater then 2.5 standard deviations from the mean, corresponding to periods of extreme neuronal activity [10]. By calculating the number of EEG channels displaying epileptiform activity, the time of seizure onset was defined. For the following analysis the number of channels needed to signify seizure activity was set to 5, allowing us to isolate periods containing epileptic events.

3. Data

In this study, EEG signals from eight patients of different ages (28–78) determined by a neurologist to be suffering from focal, generalised or secondary generalised seizures were examined. The data set was obtained from the Australian EEG database [24], and all signals were recorded using standard international system 10–20 electrode placements. The sampling rate for the data was 167 Hz, with 23 monopolar channels recorded (all electrodes have a common reference). The database contains patient details including a history, technician's comments, medication and a full EEG report by a neurologist. The clinical interpretation from the neurologist was used to comparatively classify the seizures. In this analysis only bipolar derivations (18 bipolar derivations or channels are examined below) between nearest-neighbour channels were used [25], as the use of monopolar signals might result in the introduction of unwanted correlations into the system.

[1] Downsampling or decimation of the wavelet coefficients retains half of the number of coefficients that were retained at the previous scale. Downsampling is applied in the discrete wavelet transform.

[2] The MODWT treats the time-series as if it were periodic using "circular boundary conditions". There are L_j wavelet and scaling coefficients that are influenced by the extension, which are referred to as the boundary coefficients.

Author's personal copy

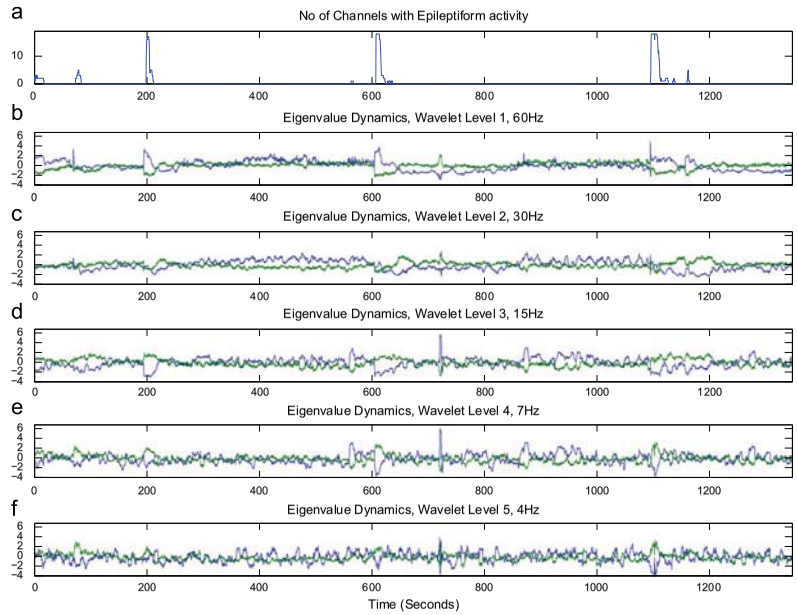


Fig. 1. Patient 1: (a) seizure definition; (b)–(f) eigenvalue dynamics across first five wavelet scales. Blue lines indicate dynamics of largest eigenvalue, while green lines show the average of 15 smallest normalised eigenvalues. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Results

For the present study, we selected the least asymmetric (LA) wavelet (known as the Symmlet [20]), which exhibits near symmetry about the filter midpoint. LA filters are defined in even widths and the optimal filter width is dependent on the characteristics of the signal and the length of the data series. The filter width chosen for this study was the LA8 (where 8 refers to the width of the scaling function), since this enables accurate calculation of wavelet correlations to the 5th scale (given the length of data series available), while still containing enough detail to capture subtle changes in the signal. Although the MODWT can accommodate any level, J_0, the largest level is chosen, in practise, so as to prevent decomposition at scales longer than the total length of the data series, hence the choice of the 5th scale (Section 2 and [22]).

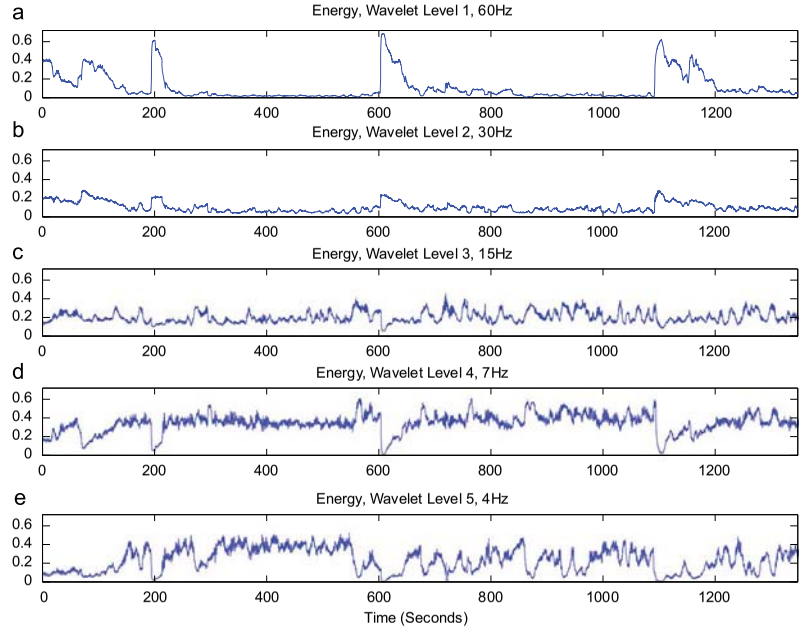
4.1. Single patient analysis

We analyse the equal-time cross-correlation dynamics between EEG channels using a sliding window of length 5 s (chosen so that the signal would be close to stationary during each window). First, the MODWT of each EEG channel was calculated within each window and the correlation matrix between channels at each scale found (Eq. (17)). The eigenvalues of the correlation matrix in each window were determined (Eq. (4)), and eigenvalue time-series were normalised in time (Eq. (5)).

The seizure definition (Section 2.3) and eigenvalue dynamics for a 30-year-old patient suffering from focal epilepsy with possible secondary generalisation are shown, Fig. 1(a). The seizure definition reveals three main periods that display epileptiform activity. An initial examination of the eigenvalue dynamics over each of the wavelet scales (corresponding to frequencies from 4 to 60 Hz) reveals that the *eigenvalue repulsion* found using the equal-time cross-correlation matrix on unfiltered data [10,11] is also repeated across the different frequencies. Fig. 1(b)–(f) shows the dynamics of both the largest eigenvalue and that of the average across the 15 smallest eigenvalues. The dynamics of the smallest eigenvalues complement those of the largest eigenvalue (since the trace of the correlation matrix must remain constant under linear transformation) [8–11] across all scales. As described earlier, the largest eigenvalue increases when the average correlation between channels increases, with a corresponding decrease in the smallest eigenvalues.

In Fig. 1, we see that the largest eigenvalue of the cross-correlation matrix calculated at the highest frequency (60 Hz) increases during epileptiform activity. However, as we move to lower frequencies, the largest eigenvalue tends to decrease during epileptiform activity. In this example, the decrease is particularly evident at levels 3 and 4 (corresponding to 15 and 7 Hz, respectively). The increase in the largest eigenvalue at the highest frequency corresponds to an increase in the average, or global, system correlation at this frequency, with the opposite occurring at lower frequencies (Section 2.1).

The wavelet energy (Eq. (14)), measured in a sliding window of 5 s for each of the wavelet scales, is shown in Fig. 2 for the same patient as previous (Fig. 1). The energy at the highest frequencies is negligible, *except during periods corresponding to epileptiform activity*, when it increases greatly, corresponding to more than 60% of the



C-27

Fig. 2. Patient 1: (a)–(e) fractional wavelet energy associated with each of the first five wavelet scales, measured using a sliding window of 5 s.

total system energy. During non-epileptiform activity, the energy at low frequencies makes up most of the system energy. However, during epileptiform activity the energy at low frequencies drops to negligible levels compensating for the increase at high frequencies implying that most of the system energy is involved in high frequency events such as spikes.

These preliminary results seem to indicate increased levels of correlation between EEG channels at the highest frequencies during epileptiform activity, with corresponding increases in energy. In contrast, the average system correlation at low frequencies (measured against the dynamical behaviour of the largest eigenvalue) decreases with corresponding decrease in energy. This analysis indicates that high frequencies are of more importance during epileptiform activity (since the associated energy is higher), and hence the correlation structure at high frequencies may be of more relevance in the characterisation of seizures.

4.2. Multiple patient analysis

In order to investigate robustness of the initial results further, we examined the eigenvalue dynamics and associated energy for all eight patients described earlier, Section 3, with the individual results shown in Table A1. For each patient we measured the average eigenvalue size and the associated energy during active periods (when the largest eigenvalue, at the highest frequency, is greater than 1.5 standard deviations units (SDU) from the mean. This corresponds to the 6.7% largest readings), as well as during normal periods (when the largest eigenvalue, at the highest frequency, is between −1.5 and 1.5 SDU from the mean). The average eigenvalue size and energy are measured across scales 1–5, as before. The results for Patient 1 correspond to those shown in Fig. 1.

Fig. 3 shows the distribution of the largest eigenvalue for all patients across each scale. For the highest frequencies, the average eigenvalue is raised considerably during active, when compared to normal periods. Using a Wilcoxon signed-rank test across all samples at each scale, the probability of zero median difference between eigenvalue pairs is less than 0.05 at these frequencies. The raised largest eigenvalues correspond to an increase in the global or average correlation at these high frequencies. For the lower frequencies, there is a definite overlap between the eigenvalues but with a lower median and much larger variance during the active period.

The energy at each of the scales is shown, Fig. 4, with an obvious increase in energy at the highest frequency during active periods (Wilcoxon signed-rank test, probability < 0.05 of distributions with the same median). At the next frequency, 30 Hz, the energy is also raised but the behaviour is not as marked. For lower frequencies, the energy decreases during active periods to compensate for the increase at higher frequencies. In particular, for scale 3, the reduction is significant with $P < 0.05$ (Wilcoxon signed-rank test). This behaviour indicates that the high frequency behaviour is of greatest importance during active periods, with corresponding correlation increase.

5. Discussion

By analysis of the eigenvalue spectrum of EEG epileptiform signals, filtered using the wavelet transform, we were able to examine changes in the eigenspectrum of the cross-correlation matrix between channels. The largest eigenvalue corresponds to the average correlation between all channels and is orthogonal to the other eigenvalues. Previously, a number of studies examined the univariate time-frequency behaviour using a number of different linear and non-linear techniques [13–16]. However, this single channel approach ignores the interactions between neurons involved in brain activity, particularly prominent during seizures.

Various bivariate methods to examine dynamic changes in these interactions have been suggested. The linear cross-correlation was measured at different frequencies, for the channels with the maximum power at high frequencies [30]. In contrast to the work above, the correlation between these channels was shown to decrease during the seizure. This, however, may be due to the bivariate approach used, which only captured the correlation between particular channels reflecting specific brain regions. This activity may possibly be reflected in the second or third largest eigenvalue of the multivariate technique, where correlations orthogonal to those in the largest eigenvalue are found (corresponding to correlations between certain subsystems [8]). The linear cross-correlation for two band-filtered channels was examined for different time lags, with a strong relationship found for a frequency band around 30 Hz [19].

The bivariate methods, described above, concentrated on correlations between a small number of channels, chosen specifically for the study. Using multivariate EEG data, changes in the global correlation structure were shown to be visible at both ends of the eigenvalue spectrum [8]. For a limited study of a single seizure, a sudden system-wide change from a relatively uncorrelated to highly correlated state was found to take place, reflected in an increase in the largest eigenvalue. Analysis of the changes in the eigenvalue spectrum for a large number of seizures showed a generic change in the correlation structure during focal onset seizures [10]. The seizure recordings in the data set studied here consisted of 58–94 channels and the changes in the eigenvalue spectrum were shown to occur for a number of the largest eigenvalues. Contrary to the behaviour found previously [8], these eigenvalues were shown to decrease during the first half of the seizure, indicating de-correlation, with an increase in correlation found before seizure end. It was suggested that this increase in correlation may be related to seizure termination.

Time-frequency decompositions of EEG signals have been studied for many years [27]. In this initial work, we extend the previous multivariate techniques [8,10], by examining changes in the eigenvalues of the correlation matrix between EEG time-series across various frequencies. The results for low frequencies were similar to those found by [10] (see Fig. 1), with a decrease in global correlation at seizure onset, reflected by the decrease in the largest eigenvalue. However, at higher frequencies (Fig. 1(b)), we found an increase in the largest eigenvalue at seizure onset, corresponding to an increase in global correlation. Additional analysis, across a number of seizures, confirmed that when eigenvalues at the highest frequencies increased, small variation only occurred in the correlation at lower frequencies. The limited number of channels available

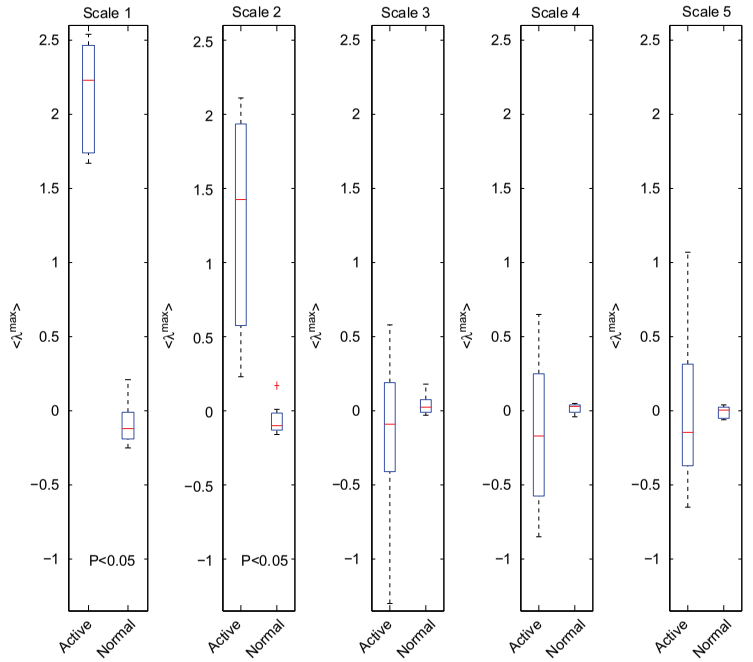


Fig. 3. Largest eigenvalue for all patients across first five scales. The red line shows the median value, while the quartiles are shown in blue and outliers as 'whiskers'. The *P* value is the Wilcoxon signed-rank probability (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

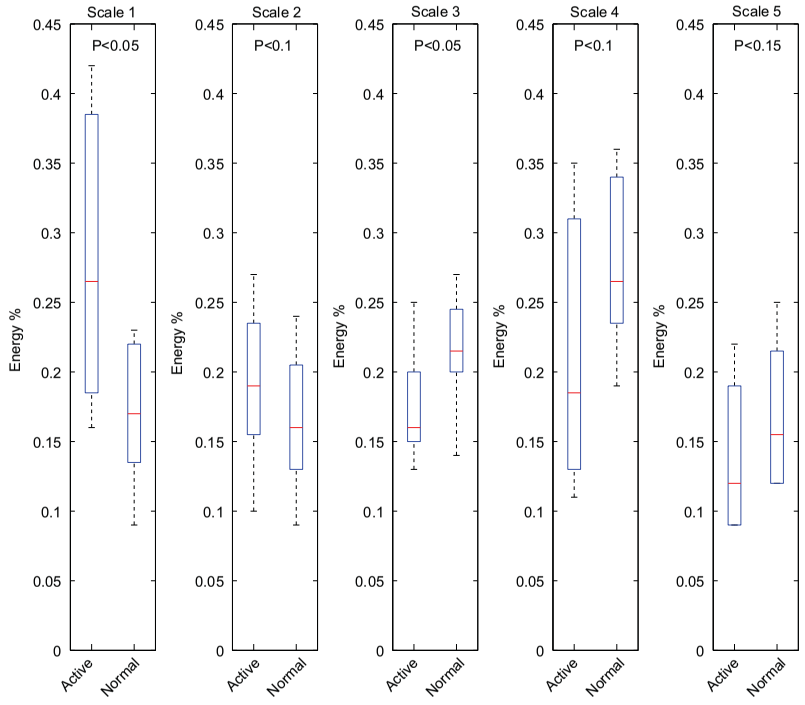


Fig. 4. % Energy for all patients across first five scales. The red line shows the median value, while the quartiles are shown in blue and outliers as 'whiskers'. The *P* value is the Wilcoxon signed-rank probability (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

C-28

for study meant that the effects of the global or the average correlation were only found in the largest eigenvalue, in contrast to the work by [10].

Previous studies have used wavelet energy in the prediction of seizures [28]. In our analysis, we examine the fractional energy of each of the frequencies, in order to determine the significance of each over time. This complements the multivariate analysis described, by emphasising the importance of the highest frequencies during active periods. Additionally, it develops the previous work [8,10], by including frequency insight on the correlation structure over time. By focusing on particular frequencies, a simple multivariate technique such as that described may be applicable to seizure prevention. Additionally, as very high frequency oscillations have been found during seizures [29], a further study of correlation dynamics at this scale is indicated.

6. Conclusions

Through the use of wavelet multiscaling, we expand the previous multivariate analysis [8–11], to explore the frequency dependence of the correlation dynamics between EEG channels for patients suffering with different forms of epilepsy. The eigenvalue spectrum of EEG epileptiform signals was filtered using the wavelet transform. Initial analysis (Section 4.1) consisted of EEG time-series from a patient with focal epilepsy, with results revealing an increase in the largest eigenvalue during epileptiform activity (corresponding to an increase in correlation between channels), at the highest frequency. In contrast, correlation at lower frequencies decreased during epileptiform activity. Wavelet energy decomposition revealed an increase in energy at higher frequencies during epileptiform activity, with a corresponding decrease at lower frequencies. This implies that *high frequency activity is more significant during epileptiform activity* and hence correlation dynamics at these frequencies assume greater relative importance. The correlation activity at lower frequencies was larger when abnormal activity was low, with higher levels of associated wavelet energy. This suggests that low frequencies assume greater importance during normal activity.

This approach was then applied to a variety of EEG signals (Section 4.2) for different types of epileptic activity. *Correlation dynamics were found to be dependent upon the frequency examined, with the correlation structure acting as a barometer of EEG activity.* Clearly, the data available are limited, but the evidence of clear crossover in eigenvalue energy does suggest that monitoring correlation structure in EEG signals at different frequencies can provide a more subtle gauge of incipient imbalance at pre-seizure stage, than is found using unfiltered signals alone.

A more detailed analysis of eigenvalue dynamics at different frequencies during seizures is certainly needed. The work by [10] finds a decrease in overall correlation after seizure starts, followed by an increase in correlation as the seizure ends. Our initial results suggest similar behaviour at lower frequencies, while the opposite occurs at higher frequencies. This suggests that the examination of correlation dynamics across various frequencies prior to seizure beginning may reveal pre-seizure characteristics, which can be used to calibrate seizure prevention strategies. An in-depth study on different seizure types may reveal further distinct correlation structures specific to the seizure type. Moreover, the analysis of inter-frequency correlations may shed light on the lead–lag relationship across different frequencies, while subsidiary eigenvalue investigation is also worthwhile.

Conflict of interest statement

None declared.

Acknowledgements

The data set used in this work was taken from the Australian EEG Database of the University of Newcastle, Australia [24]. The authors are grateful for the kind assistance of Adam Walker and others associated with the database.

Appendix A.

Table A1 shows the seizure analysis.

Table A1
Seizure analysis: row 1 shows average size of λ^{max} when $\lambda^{max} > 1.5$ at scale 1, and row 2 the associated energy at each scale.

Patient no.	Scale	1	2	3	4	5
1	Scale 1 active	2.20	0.23	−1.30	−0.71	−0.33
	Energy	0.38	0.18	0.15	0.17	0.12
	Scale 1 normal	−0.12	0.01	0.09	0.05	0.03
	Energy	0.09	0.09	0.21	0.36	0.25
2	Scale 1 active	2.54	1.65	−0.56	0.13	0.52
	Energy	0.39	0.26	0.15	0.11	0.09
	Scale 1 normal	−0.25	−0.16	0.06	−0.02	−0.05
	Energy	0.23	0.18	0.22	0.19	0.12
3	Scale 1 active	2.50	2.09	0.58	0.37	1.07
	Energy	0.32	0.27	0.16	0.13	0.12
	Scale 1 normal	−0.10	−0.08	0.00	0.00	−0.05
	Energy	0.23	0.23	0.19	0.23	0.12
4	Scale 1 active	1.72	1.78	−0.24	−0.44	−0.11
	Energy	0.42	0.20	0.16	0.13	0.09
	Scale 1 normal	−0.12	−0.12	0.02	0.03	0.01
	Energy	0.21	0.15	0.26	0.24	0.14
5	Scale 1 active	2.26	1.20	0.18	−0.31	−0.41
	Energy	0.16	0.16	0.24	0.27	0.17
	Scale 1 normal	−0.24	−0.13	−0.02	0.03	0.04
	Energy	0.17	0.17	0.23	0.26	0.17
6	Scale 1 active	2.43	2.11	−0.26	0.65	0.11
	Energy	0.21	0.21	0.16	0.20	0.22
	Scale 1 normal	−0.14	−0.13	0.03	−0.04	0.00
	Energy	0.12	0.18	0.21	0.27	0.22
7	Scale 1 active	1.67	0.75	0.06	−0.03	−0.18
	Energy	0.16	0.15	0.25	0.35	0.09
	Scale 1 normal	0.21	0.17	0.13	0.04	−0.06
	Energy	0.15	0.14	0.27	0.32	0.12
8	Scale 1 active	1.76	0.40	0.20	−0.85	−0.65
	Energy	0.21	0.10	0.13	0.35	0.21
	Scale 1 normal	0.08	−0.04	−0.03	0.04	0.02
	Energy	0.17	0.12	0.14	0.36	0.21

Row 3 shows average eigenvalue at each level when $-1.5 < \lambda^{max} < 1.5$ and row 4 the associated energy.

Author's personal copy

References

[1] L. Laloux, P. Cizeau, M. Potters, J. Bouchaud, Random matrix theory and financial correlations, Int. J. Theoret. Appl. Finance 3 (3) (2000) 391–397.
[2] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, H.E. Stanley, A random matrix approach to cross-correlations in financial data, Phys. Rev. E 65 (2000) 066126.
[3] T. Conlon, H.J. Ruskin, M. Crane, Random matrix theory and fund of funds portfolio optimisation, Physica A 382 (2) (2007) 565–576.
[4] T. Conlon, H.J. Ruskin, M. Crane, Cross-correlation dynamics in financial time series, Physica A 388 (5) (2009) 705–714.
[5] P. Seba, Random matrix analysis of human EEG data, Phys. Rev. Lett. 91 (19) (2003) 198104.
[6] R. Quian Quiroga, A. Kraskov, T. Kreuz, P. Grassberger, Performance of different synchronization measures in real data: a case study on electroencephalographic signals, Phys. Rev. E 65 (2002) 041903.
[7] K. Ansari-Asl, L. Senhadji, J.-J. Bellanger, F. Wendling, Quantitative evaluation of linear and nonlinear methods characterizing interdependencies between brain signals, Phys. Rev. E 74 (2006) 031916.
[8] M. Müller, G. Baier, A. Galka, U. Stephani, H. Muhle, Detection and characterization of changes of the correlation matrix in multivariate time series, Phys. Rev. E 71 (2005) 046116.
[9] M. Müller, Y.L. Jiménez, C. Rummel, G. Baier, A. Galka, U. Stephani, H. Muhle, Localized short-range correlations in the spectrum of the equal-time correlation matrix, Phys. Rev. E 74 (2006) 041119.
[10] K. Schindler, H. Leung, C.E. Elger, K. Lehnertz, Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG, Brain 130 (2007) 65–77.
[11] K. Schindler, C.E. Elger, K. Lehnertz, Increasing synchronization may promote seizure termination: evidence from status epilepticus, Clin. Neurophysiol. 118 (9) (2007) 1955–1968.
[12] J. Kwapien, S. Drozdz, A.A. Ioannides, Temporal correlations versus noise in the correlation matrix formalism: an example of the brain auditory response, Phys. Rev. E 62 (2000) 5557.
[13] I. Clark, R. Biscay, M. Echeverria, T. Virues, Multiresolution decomposition of non-stationary EEG signals: a preliminary study, Comput. Biol. Med. 25 (4) (1995) 373–382.
[14] S. Senhadji, F. Wendling, Epileptic transient detection: wavelets and time-frequency approaches, Clin. Neurophysiol. 32 (3) (2002) 175–192.
[15] H. Adeli, Z. Zhou, N. Dadmehr, Analysis of EEG records in an epileptic patient using wavelet transform, J. Neurosci. Meth. 123 (2003) 69–87.
[16] K.P. Indirdevi, E. Elias, P.S. Sathidevi, S. Dinesh Nayak, K. Radhakrishnan, A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram, Comput. Bio. Med. 38 (2008) 805–816.
[17] G. Ouyang, X. Li, Y. Li, X. Guan, Application of wavelet-based similarity analysis to epileptic seizures prediction, Comput. Bio. Med. 37 (2007) 430–437.
[18] Y. Mizuno-Matsumoto, S. Ukai, R. Ishii, S. Date, T. Kaishima, K. Shinosaki, S. Shimojo, M. Takeda, S. Tamura, T. Inouye, Wavelet crosscorrelation analysis: non-stationary analysis of neurophysiological signals, Brain Topogr. 17 (4) (2005).
[19] K. Ansari-Asl, J.-J. Bellanger, F. Bartolomei, F. Wendling, L. Senhadji, Time-frequency characterization of interdependencies in nonstationary signals: application to epileptic EEG, IEEE T. Bio. Med. Eng. 52 (7) (2005).
[20] A. Bruce, H. Gao, Applied Wavelet Analysis with S-Plus, Springer, New York, 1996.
[21] C.S. Burrus, R.A. Gopinath, H. Gao, Introduction to Wavelets and Wavelet Transforms, Prentice-Hall, Englewood Cliffs, NJ, 1997.
[22] D.B. Percival, A.T. Walden, Wavelet Methods for Time Series Analysis, Cambridge University Press, Cambridge, 2000.
[23] R. Gencay, B. Selcuk, B. Whitcher, An Introduction to Wavelets and Other Filtering Methods in Finance and Economics, Academic Press, New York, 2001.
[24] M. Hunger, R.L.L. Smith, W. Hyslop, O.A. Rosso, R. Gerlach, J.A.P. Rostas, D.B. Williams, F. Henskens, The Australian EEG database, Clin. EEG Neurosci. 35 (2) (2005).
[25] B.J. Fisch, R. Spehlmann, Fisch and Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG, Elsevier, Amsterdam, 1999.
[26] I.T. Jolliffe, Principal Component Analysis, second ed., Springer Series in Statistics, 2002.
[27] M. Nakata, K. Mukawa, Fourier analysis of broad spectrum EEG from a fluctuation point of view, Integr. Psychiat. Beh. Sci. 24 (3) (1989) 90–97.
[28] S. Gigola, F. Ortiz, C.E. D'attellis, W. Silva, S. Kochen, Prediction of epileptic seizures using accumulated energy in a multiresolution framework, J. Neurosci. Meth. 138 (2004) 107–111.
[29] A. Bragin, J. Engel Jr., C.L. Wilson, I. Fried, G.W. Mathern, Hippocampal and entorhinal cortex high-frequency oscillations (100–500 Hz) in human epileptic brain and in kainic acid-treated rats with chronic seizures, Epilepsia 40 (2) (1999) 127–137.
[30] F. Wendling, F. Bartolomei, J.J. Bellanger, J. Bourien, P. Chauvel, Epileptic fast intracerebral EEG activity: evidence for spatial decorrelation at seizure, Brain 126 (2003) 1449–1459.