Applications of Statistical Physics to economics: distributions of wealth and correlations of financial data

by

Ricardo Coelho

A Thesis Submitted to
The University of Dublin
for the Degree of

Doctor of Philosophy

School of Physics
Trinity College
University of Dublin

August 2008

Declaration

This thesis has not been submitted as an exercise for a degree at any other University. Except where otherwise stated, the work described herein has been carried out by the author alone. This thesis may be borrowed or copied upon request with the permission of the Librarian, Trinity College, University of Dublin. The copyright belongs jointly to the University of Dublin and Ricardo Coelho.

Signature of Author ....Ricardo Coelho....

Ricardo Coelho

August 2008

Summary

Over the last decade the study of economic and financial problems by the physics community has become very popular. Econophysics describes the application of tools from statistical physics to the study of problems in economics, such as correlations in stock prices or the distribution of wealth in society.

In this thesis we have considered two separate topics. First, the distribution of wealth in societies where the emphasis is on examining the power laws that characterise the distribution of wealth and income. And second, the correlations in stock prices, which is linked to asset and portfolio risk management.

Since Vilfredo Pareto in 1896, it is well known that the distributions of wealth and income in societies are described by a power law. Nowadays, the analysis is refined. Power laws apply only for the rich end of the distribution and other kind of distributions such as the log-normal, the exponential or Gamma distribution explain the lower and middle parts of the distribution of wealth and income.

The power-law distribution of wealth is reproduced by various physical models, based on the analogy with collisions of particles or Langevin type equations. We review some models and empirical results found in the past and point to the existence of double power laws in the distribution of wealth. These double power laws have two different exponents, one for the millionaires of society and another one for the billionaires. A model of money exchange between agents where agents are divided in two groups, one with a higher saving propensity than the other, is presented and it reproduces the double power law characteristic. Also some analytical tools about double power laws are introduced and compared with our numerical results.

For the study of correlations in stock prices, which are of extreme importance in the construction of optimal portfolios by investors, we present an analysis of financial data from stocks that belong to some of the main indices around the world, the London Stock Exchange, FTSE100, the Dow Jones Industrial Average index, DJIA, the *Cotation Assistée en Continu*, CAC40, the Belgium index, BEL20 and the Amsterdam Exchange Index, AEX. Using the concept of random matrix theory, which is a theory developed for applications in Nuclear Physics, and

minimal spanning trees, which help us in the visualisation of the affinities between stocks, showing stocks with higher correlations next to each other, we study different characteristics of our portfolios of stocks from different indices. Our studies reveal a division of the stocks in industrial sub-sectors, mostly in good agreement with empirically derived groupings, but also indicating possible refinements, important for the use in portfolio optimisation.

A similar analysis of market indices of different countries shows that despite globalisation strong regional geographical correlations still exist.

A study of cross-correlations of stocks from different indices also show that depending on the location of the markets there are different behaviours for the correlations between stocks. For a study of cross-correlations between stocks from FTSE100 and DJIA indices we see a segregation between stocks from each market. But for a study of cross-correlations between European stocks, most of the stocks cluster in groups of the same industrial sector.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Stefan Hutzler and Prof. Peter Richmond for the support over the past three and a half years. Without our group meetings with Przemek, Joe and Stephen, most of this work wouldn't be possible. So, I'd like to thank all of them for their feedback and for all the ideas (good and bad) that we constructed together. I also would like to thank Prof. Brian Lucey from the School of Business and Prof. Claire Gilmore from the McGowan School of Business for the discussions that we have about the economic topics in my work.

I would like to thank my two sources of financial funding. For the first two years I'd like to thank the Science Foundation Ireland for their financial support. And for the past year and a half I'd like to thank the support of Fundação para a Ciência e Tecnologia.

I cannot forget my office mates, first the ones from the Sniam building: Rafa, Floriano, Sumesh and Barry where we shared a good cultural experience and then the foamy ones: Vincent Carrier, Vincent Langlois, James, Roman, Mohammad, Antje, Joe and Stephen. I also would like to thank everybody that I meet during this great time in Ireland, from the drum people to the volleyball friends and all the others that I will not forget. I have to mention Marina, Martin, Felipe, Hélder and Paulo for all their friendship and support during these past three and a half years, without them life wouldn't be the same in Dublin.

Finally, I would like to thank my family, *pai e mãe*, for their amazing support and my lovely Rita for everything, but mostly for the support in these past years living more than 1600km away from each other.

List of Publications

1. P. Richmond, P. Repetowicz, S. Hutzler and R. Coelho *Comments on recent studies of the dynamics and distribution of money*, Physica A **370**, 43 (2006)

2. P. Richmond, S. Hutzler, R. Coelho and P. Repetowicz, *A review of empirical studies and models of income distributions in society*, in *Econophysics and Sociophysics of Wealth Distributions: Trends and Perspectives.* (eds. B. K. Chakrabarti, A. Chakraborti and A. Chatterjee), Wiley-VCH, Berlin (2006)

3. R. Coelho, S. Hutzler, P. Repetowicz and P. Richmond, *Sector analysis for a FTSE portfolio of stocks*, Physica A **373**, 615 (2007)

4. R. Coelho, C. G. Gilmore, B. Lucey, P. Richmond and S. Hutzler, *The evolution of interdependence in world equity markets-Evidence from minimum spanning trees*, Physica A **376**, 455 (2007)

5. M. A. Santos, R. Coelho, G. Hegyi, Z. Néda and J. Ramasco, *Wealth distribution in modern and medieval societies*, Eur. Phys. J. Special Topics **143**, 81 (2007)

6. R. Coelho, P. Richmond, J. Barry and S. Hutzler, *Double power laws in income and wealth distributions*, Physica A **387**, 3847 (2008)

7. R. Coelho, P. Richmond, B. Lucey and S. Hutzler *A Random Matrix Theory based analysis of stocks of markets from different countries*, Advances in Complex Systems, in press (2008)

Contents

Chapter 1

Introduction

"I think the next century will be the century of complexity." (Stephen Hawking)

For the past few decades, many physicists have been exploring many different fields, that have in common one characteristic, *complexity*. There are many complex systems in nature, from grains of sand in an avalanche to the return price of a stock in a market [1].

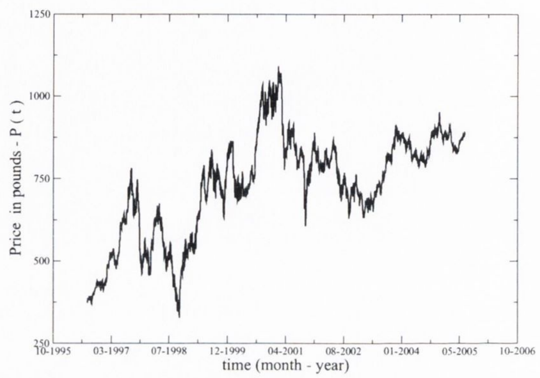
A field of study in complex systems that has received a lot of attention and which has a large number of works published, is the applications of physics to finance and economics, also baptised as Econophysics by Eugene Stanley [2].

The introduction of physics to the economical science has a historical background. For example, most of the people know Copernicus as the founder of modern astronomy, but he achieved a great reputation as an adviser on economic matters to the King of Poland. On his duties as adviser, Copernicus became aware of the economic and social distress caused by wartime inflation and wrote a report on the subject in 1522. During the next few years he revised it into a short treatise on the economic evils of a debased currency and made specific proposals for monetary reform [3].

Before Copernicus, most of the writings on the subject of money were passages in Aristotle's *Politics* as interpreted and applied by churchmen. They were very theological and were concerned more with what ought to happen than with what actually did happen. Copernicus' report was purely empirical and pragmatic, in place of appealing to a priori principles, he appeals to the observed facts and he supports each step to specific factual evidence [3].

Isaac Newton, most known as the father of classical mechanics, was also a Warden of the Mint and Master of the Royal Mint for 31 years in the Royal Mint in London, which is the body permitted to manufacture coins in the United Kingdom [4].

The mathematician Louis Bachelier, who's PhD supervisor was Henri Poincaré, was the first to publish a formalisation of a random walk in his PhD thesis, *Théorie de la spéculation* [5], where he discussed the use of Brownian motion to evaluate stock options. In Figure 1.1 we represent the evolution of the price of a stock and a random walk.



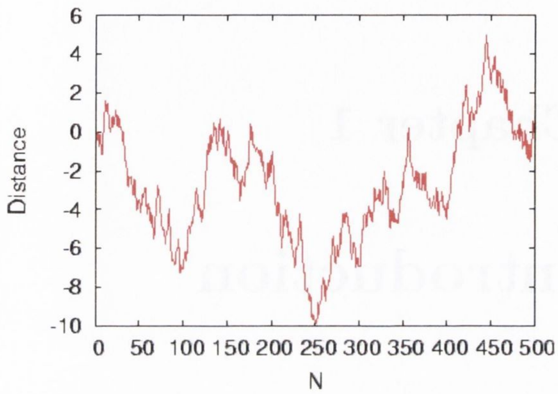


Figure 1.1: Evolution of the price of a stock (left figure) compared with the evolution of a random walk (right figure).

Benoît Mandelbrot, a mathematician famous for his achievements in fractal theory, found that price changes in financial markets did not follow a Gaussian distribution. For the high end of the distribution of price changes, Mandelbrot stated that they are fatter than a Gaussian distribution and are better explained by another distribution, the Lévy stable distribution [6, 7].

In the last 20 years the number of physicists that applied some of their techniques and knowledge to the fields of economic and finance has significantly increased. This can be observed in the amount of articles uploaded to the Los Alamos arxiv [8] related with the econophysics field and also in the number of books published about this topic. In the main physics literature: Physical Review Letters [9], Physical Review E [10] and The European Physical Journal B [11] there are plenty of publications in each volume about this field. Most of the publications are in Physica A [12].

The first books in the field were published by Rosario Mantegna and Eugene Stanley [2] and Jean-Philippe Bouchaud and Marc Potters [13], both in 2000. Next we give reference to 5 books that also laid out and reviewed the field. In 2001, Johannes Voit published his *The Statistical Mechanics of Financial Markets* [14]. Two years later Didier Sornette published *Why Stock Markets Crash: Critical Events in Complex Financial Systems* [15]. In 2004 there appeared *Dynamics of Markets: Econophysics and Finance* by Joe McCauley [16] and the Kolkata group of Bikas Chakrabarti published some reviews about econophysics in *Econophysics of Wealth Distributions* [17] in 2005 and *Econophysics and Sociophysics of Wealth Distributions: Trends and Perspectives* [18] in 2006. A review of these books and some articles can be found in the main econophysics website, the Econophysics Forum [19].

The number of conferences exclusive to the topic of econophysics has also increased in the last decade. The main conferences are: Applications of Physics in Financial Analysis (APFA), the Econophysics Colloquium, the Nikkei Econophysics Symposium and the Econophys-Kolkata Workshop.

In summary, while individual physicists were interested in problems in economy for a long time, the field of econophysics only properly emerged in the last decade when physicists started to look at open problems in economics and with the large amount of financial data available started to see many differences between the economic theory and empirical results.

The area of econophysics has many sub-areas but in this thesis we just focus on two of them, the study of wealth distributions and the correlations between time series of stock price returns.

1.1 Wealth

The area of study of wealth distributions is not only related with the study of wealth or income distributions in societies [17], but also with the size of companies in a country [20], or the GDP (Gross Domestic Product) of countries [21]. The study of wealth distributions has attracted great interest since the work of the socio-economist Vilfredo Pareto, who wrote a book about economical politics 100 years ago [22], studying a large amount of economical data. Pareto suggested that the distribution of wealth from different cities and countries follow a power law distribution with similar exponents α (between 1 and 2), known nowadays as Pareto's index:

$$p(w) \sim w^{-(1+\alpha)}, \text{ for large } w. \tag{1.1}$$

In Figure 1.2 we show the cumulative distribution of income in Japan for 1998 presented by Souma [23].

Apart from the study of the empirical data, physicists are very interested in modelling wealth distributions [17, 18]. A detailed review of some models and open problems in the study of wealth distribution [24] was published by us in a chapter of one of the econophysics books [18] and it is summarised in Chapter 2.

Models used in biological systems, such as Lotka-Volterra models, were used by physicists to explain the economic trade relations in communities [25, 26, 27]. Gas models of collisions were transformed into economic models where agents substitute molecules, money substitutes energy and trade substitutes collisions [28, 29, 30, 31, 32, 33, 34]. A model of dynamical network of families [35], where each node is one family and links between nodes indicates family relations, was used to implement money exchange between different families due to payments of new links (like weddings), payments to the society (to rear a child) and distributions of money from nodes that will disappear (inheritance).

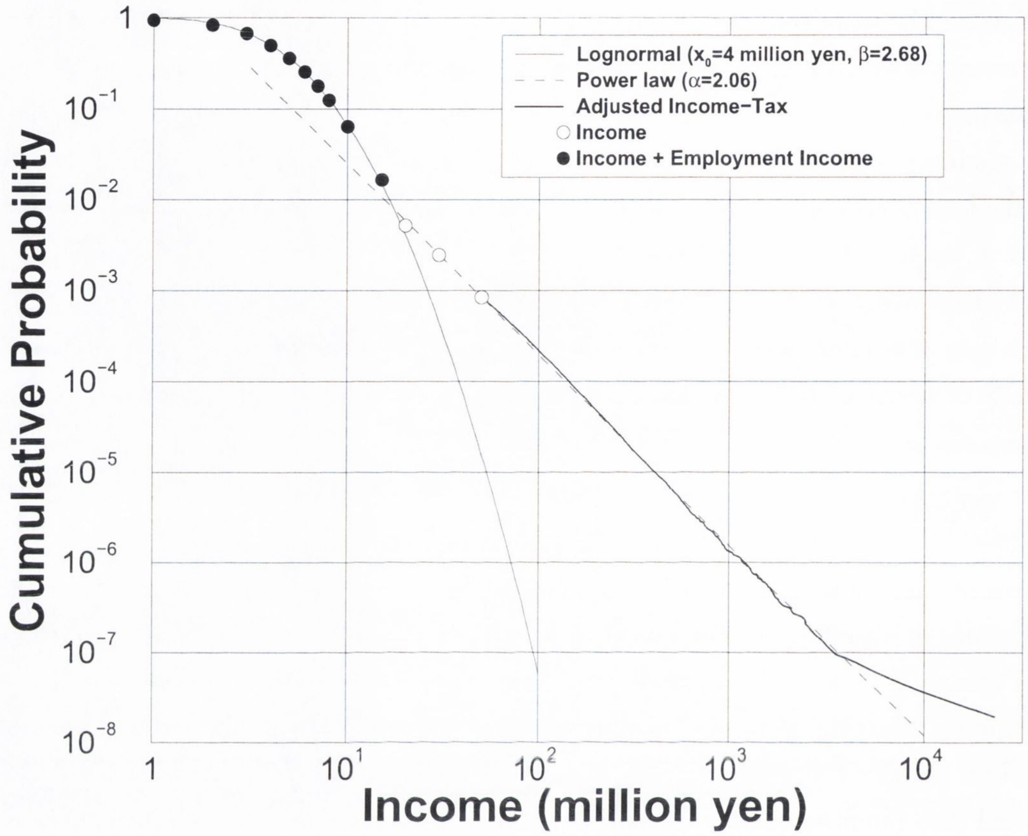


Figure 1.2: Cumulative distribution of income in Japan for 1998 presented by Souma [23]. This data is related with more than 50 million individuals, about 80% of all workers in Japan. The income-tax data is available from a list of the 84515 individuals who paid income tax of 10 million yen or more.

The appeal in using such models is that they are simple, with analytical solutions with few parameters that describe the empirical data of wealth distributions quite well. However, issues remain. For example, the power law distribution of wealth just appears for the richer part of society (5-10% of the population). The wealth of the other part of society is normally defined to have a log-normal or Gibbs distribution. But even the power law in the end of the distribution seems to have more than one Pareto exponent. A model able to explain an exponent for the millionaires and other exponent for the billionaires that normally appear on the list of World Top Richest (like Forbes [36]) is presented in Chapter 3.

1.2 Correlations between time series of stock price returns

The area of study of correlations between time series tries to understand the nature of these time series, how they evolve in time, and if it is possible to predicted movements of the market. It is also very popular to study the price returns and their distribution. For a long time, the distributions of price returns were treated as Gaussian distributions, but this model is known to provide only a first approximation of what is observed in the empirical data. To explain the empirical evidence of fatter tails of the measured distributions, the proposed Lévy stable distributions for the price changes has had much success among the physics community.

A better knowledge of market movements and the correlations of stock returns is essential for investors in terms of construction of portfolios. A matrix of correlations tell us about the affinity between different stocks. This affinity is measure in terms of how two different time series move in relation to each other. Studying the matrix of correlations between time series and its properties became very popular among the physicist working in economics problems.

The two main techniques to analyse the correlation matrix are through the random matrix theory and the visualisation of the minimal spanning tree.

Random Matrix Theory was previously used in Nuclear Physics to study the statistical behaviour of energy levels of nuclear reactions [37]. According to quantum mechanics, the energy levels are given by the eigenvalues of a Hermitian operator, the Hamiltonian which was postulated to have independent random elements. However, analysis of the eigenvalues of real data showed deviations from the spectra of fully random matrices, thus indicating non-random properties, useful for an understanding of the interactions between nuclei. This approach is nowadays applied to the study of correlations of time series of returns in the stock market, where physicists try to find the non-random properties of the matrix of correlations [38, 39, 40]. With the prediction of the eigensystem of a random matrix, compared with the eigensystem of the matrix of empirical data of stocks, we can see eigenvalues far from the prediction spectrum, that have a lot of information about the market [41, 42], the index of the market, or the clustering in industrial sectors in markets. The index of the market can be calculated as the simple mean of the prices of all the stocks that belong to the market, or the weighted mean, where some stocks contribute more to the index, related with the size of the company. The industrial sectors can be different for different classifications, but normally the industrial sector indicates which kind of business the company is engaging in.

To visualise the hierarchical structure of financial markets, Mantegna defined a distance metric between stocks, using the correlations between them [43]. Using this distance, he constructed a network of stocks (Minimal Spanning Tree), where nodes are stocks and links are the distances between them. An example of a network of companies is represented in Figure 1.3.

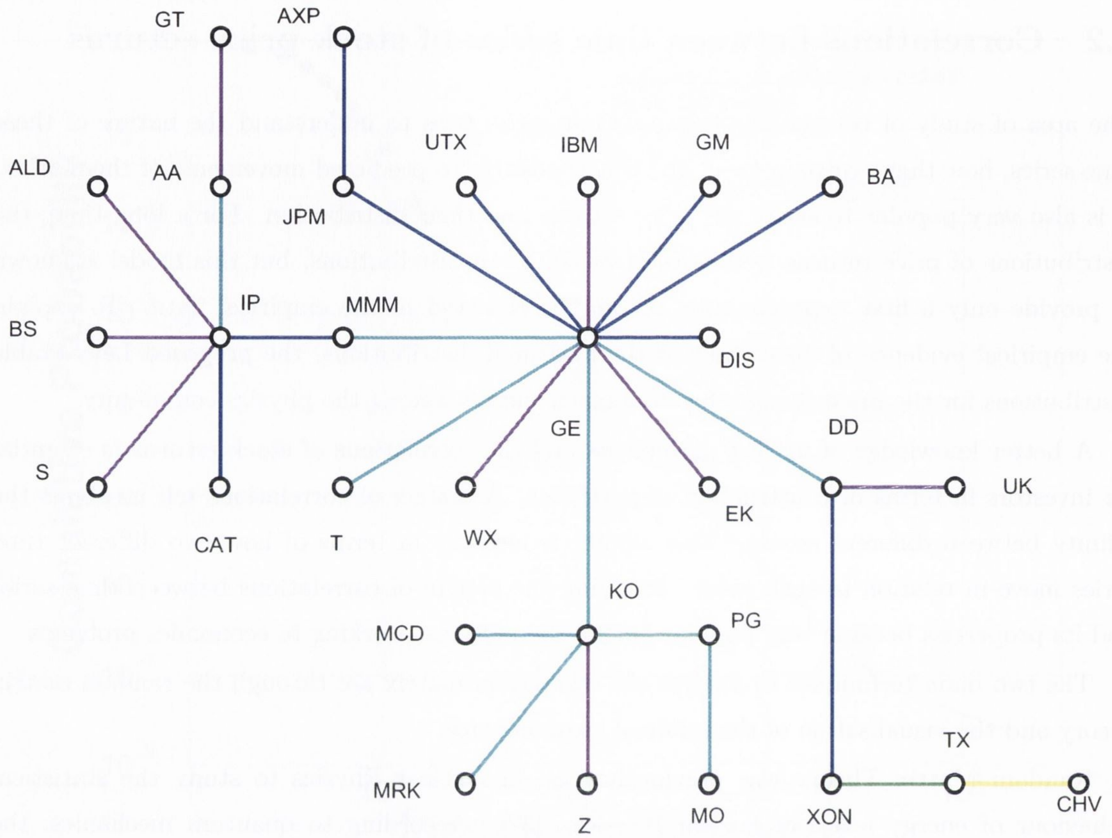


Figure 1.3: Minimal spanning tree connecting the 30 stocks used to compute the Dow Jones Industrial Average presented by Mantegna in his article [43].

The relations between companies show the formation of clusters of sectors. Properties of the trees, like topology for different time scales [44, 45], degree distribution of the nodes [46, 47, 48, 49, 50], time evolution of moments of the distances of the trees [48, 49, 51, 52, 53, 54, 55], spread of nodes in the tree [48, 52, 53], robustness of the tree [48, 49, 50, 52, 53, 54, 56], topology before and after financial crashes [54] and others are presented in Chapters 4, 5, 6, 7 and 8.

Chapter 2

Wealth Distributions

2.1 Introduction

The statement that wealth is not distributed uniformly in society appears obvious. However, this immediately leads to several non-trivial questions. How is wealth distributed? What is the form of the distribution function? Is this distribution universal or does it depend on the individual country? Does it depend on time or history? These questions were first studied by Vilfredo Pareto in 1896/97 who noticed that the rich end of the distribution was well described by a power law. Ever since these early studies of Pareto, economists, and more recently physicists, have tried to first of all infer the exact shape of the entire distribution from economic data, and secondly, to design theoretical models that can reproduce such distributions.

In this chapter we review both historical and current concepts and data which support the thesis that certain features of the wealth distribution are indeed universal and also review some models that try to explain this universality. In chapter 3 we set out to explain a feature of such distributions which is often neglected, namely double power laws..

Progress has been made, but questions remain. How can wealth be measured? Gross salary income may be a good indicator for low-to-medium income earners, but how do the super-rich, many of whom are not employees, fit into this picture. The wealth calculated for the super-rich in lists of top individuals seems to be calculated from the value of assets that each billionaire owns at each year, plus some of their possessions and income.

The definitions of wealth and income show the difference between these two quantities. The wealth is the amount of money and possessions that an agent has at a specific time. The income is the money that an agent receives regularly from payments or investments. So from the definitions we see that agents trade money in the sense of income, but don't exchange wealth. In most of the models studied, the exchanges between agents are given by a term which is proportional to their wealth, but it's not the wealth itself.

In our discussion of different models we shall restrict ourselves to income, which is what more often we use in the daily basis to trade with other agents. We shall only concern ourselves with the question of wealth when we calculate the final amount that belongs to each agent.

We begin this chapter by giving some background on the thinking of Pareto, a researcher who was ahead of his time with regards to scientific research, but who also mirrored the zeitgeist of his period. We then summarise early work by Gibrat, Champernowne, Lydall and Mandelbrot. Before reporting more recent approaches, we present some empirical results and compare with the probability distribution functions more often used to explain the distribution of wealth in societies. In particular we feature a number of different agent based models: the Family Network model, the Generalised Lotka Volterra model, the Slanina Model and other models where agents exchange money by pairwise transactions by analogy with the exchange of momentum by colliding molecules in a gas. These collision models which are conceptually extremely simple, are also accessible to analytical theory and thus currently *en vogue*.

2.2 Pareto and early models of wealth distribution

2.2.1 Pareto law

The distribution of wealth in society has proved to be of great interest for many years. Based on the numerical analysis of an impressive amount of economic data, Italian economist Vilfredo Pareto [22] was the first to suggest that it followed a natural law now often simply termed Pareto law. A sketch of an income distribution as seen in Pareto's *Manual of Political Economy* [57] is shown in Fig. 2.1. However there are a number of different forms of Pareto law quoted in the literature. Mandelbrot [6] distinguishes between two different versions:

- the strong Pareto law;
- the weak Pareto law.

If $P(u)$ is the percentage of individuals with an income U greater than u (cumulative distribution), the strong Pareto law states that:

$$P(u) = \begin{cases} (u/u_0)^{-\alpha} & \text{for } u > u_0 \\ 1 & \text{for } u < u_0 \end{cases} \tag{2.1}$$

The density $p(u) = -dP(u)/du$ is thus given by:

$$p(u) = \begin{cases} \alpha\,(u_0)^{\alpha}\, u^{-(\alpha+1)} & \text{for } u > u_0 \\ 0 & \text{for } u < u_0 \end{cases} \tag{2.2}$$

Here u_0 is a scale factor and the value of the exponent α is not determined. In the strongest form of the Pareto law, $\alpha = 3/2$, which is the average value of α in Pareto's original data (see Table 2.1). Because the strong Pareto law doesn't apply to the empirical data in the whole range of income, but just for large values of u, the weak Pareto law states that the power law only holds in the limit $u \to \infty$:

$$\frac{P(u)}{(u/u_0)^{-\alpha}} \to 1, \text{ as } u \to \infty \tag{2.3}$$

where the value of α remains unspecified. In the remainder of this chapter we shall mean this form of the law when we write Pareto law.

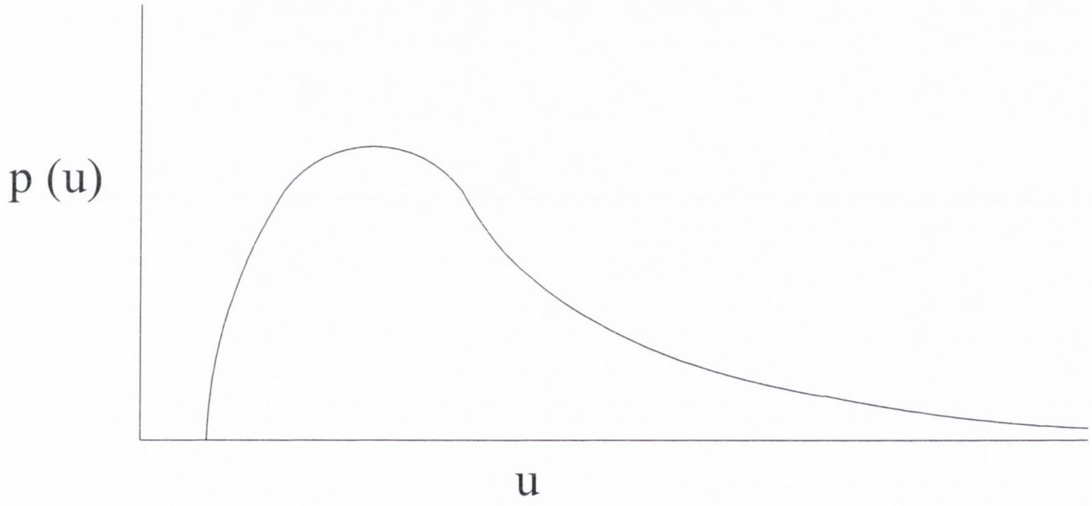


Figure 2.1: Sketch of the distribution of income as seen in Pareto's Manual [57]. For large values of income this follows a power law.

Today, Pareto's law is usually quoted in terms of the distribution density function $p(u)$, rather than the cumulative distribution function, $P(u) = \int_u^\infty p(u')du'$, viz:

$$p(u) \sim u^{-(1+\alpha)}, \text{ for large } u. \tag{2.4}$$

In Figure 2.2 we represent this distribution density function $p(u)$, for a Pareto exponent α equal to 1.5.

2.2.2 Pareto's view of society

It is interesting to read the original writings of Pareto from 1906, since these reveal more about the process by which he arrived at his conclusions. He opens the chapter on *Population* in his

Table 2.1: Table, taken from Pareto's book [22], showing the exponent α for a number of different data sets. Note that this is only a small extract of all the data that Pareto analysed.

Country	Year	α	Country	Year	α
England	1843	1.50	Perouse, village		1.69
	1879-80	1.35	Perouse, campagne		1.37
Prussia	1852	1.89	Italian cities		1.32
	1876	1.72	Italian villages		1.45
	1881	1.73	Basle	1887	1.24
	1886	1.68	Paris		1.57
	1890	1.60	Augsburg	1471	1.43
	1894	1.60		1498	1.47
Saxony	1880	1.58		1512	1.26
	1886	1.51		1526	1.13
Florence		1.41	Peru	end of 18^{th} century	1.79

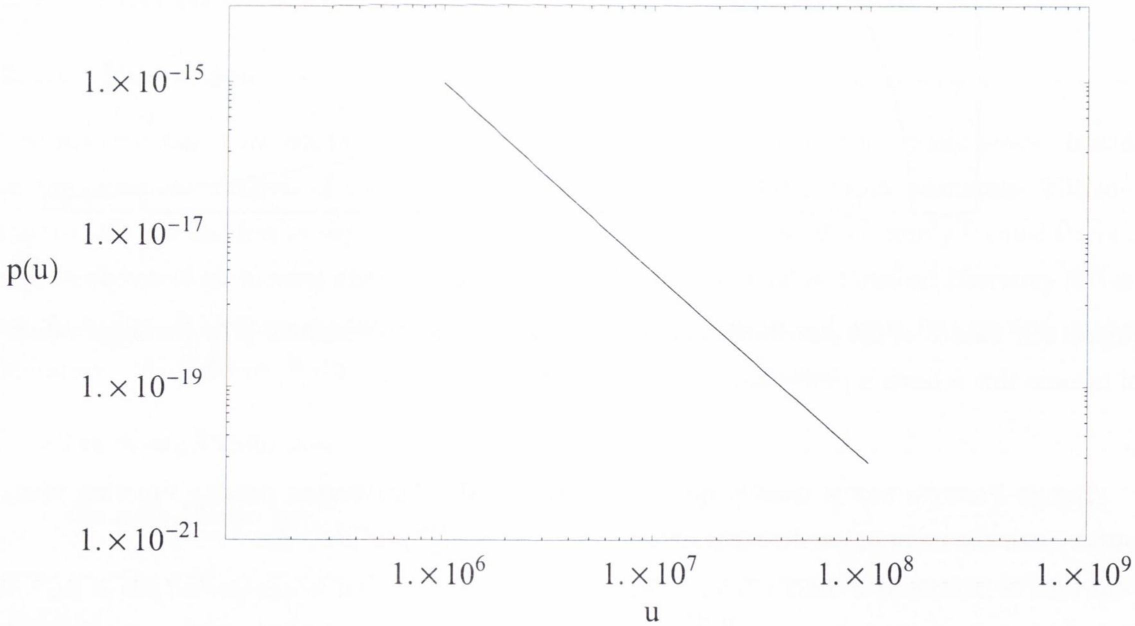


Figure 2.2: Representation of the weak Pareto law (eq. 2.4) for $\alpha = 1.5$, in a $\log - \log$ scale. The straight line is the characteristic of this power law distribution.

Manual of Political Economy [57], which summarises his findings and thoughts in a mainly non-mathematical fashion, by stating that "society is not homogeneous" ([57], p. 281). The existing "economic and social inequalities" correspond to the "inequalities of human beings *per se*" with respect to "physical, moral, and intellectual viewpoints" ([57], p. 281). Thus an excellent indicator of non-homogeneity in society is, according to Pareto, the distribution of income in society, as sketched in Fig. 2.1.

Pareto's main achievement, as seen from the perspective of today's econophysicists and economists, is the observation that this distribution is *universal*, i.e. that it "varies very little in space and time; different peoples and different eras yield very similar curves. There is a remarkable stability of form in this curve." ([57], p. 285). Furthermore Pareto discovered that the form of the curve "does not correspond at all to the curve of errors, that is, to the form the curve would have if the acquisition and conservation of wealth depended only on chance".

This non-Gaussian character of the curve is obvious from its lack of symmetry about its peak and the pronounced tail at the rich end of the distribution, although Pareto does not dwell on the concept of power laws in his *Manual* [57].

Pareto notes that the poor end of the wealth distribution cannot be fully characterised, due to a lack of data. He stresses, however the existence of a "minimum income... below which men cannot fall without perishing of poverty and hunger" ([57], p. 286).

Finally he notes the stability of the distribution: "If, for example, the wealthiest citizens were deprived of all their income [...] sooner or later [the curve] would reestablish itself in a form similar to the initial curve. In the same way, if a famine [...] were to wipe out the lower parts of the population [...] the figure [...] would return to a form resembling the original one." ([57], p. 292).

For Pareto the wealth distribution of Fig. 2.1 "gives a picture of society" ([57], p. 286) and thus forms the basis of his theory of society. In using arguments based on Darwin's ideas of social selection, in common with many of his contemporaries, and calling feminism a *malady* and referring to women as "objects of luxury who consume but do not produce" ([57], p. 297), he also paints a picture of society at his time.

In Pareto's view people are in principle free to move along the wealth axis in the course of their lifetime, in both directions, but this movement is determined by "whether they are or are not well fitted for the struggle of life" ([57], p. 287). If they drop below the minimum income they "disappear" ([57], p. 286) or are "eliminated" ([57], p. 287). In the region of low incomes "people cannot subsist, whether they are good or bad; in this region selection operates only to a very small extent because extreme poverty debases and destroys the good elements as well as the bad." ([57], p. 287). Pareto views the process of selection to be most important in the area around the peak of the distribution. Here the incomes are "not low enough to dishearten the best elements". He continues with the following statement that reflects views that were probably widely held at that time: "In this region, child mortality is considerable, and this mortality is probably a powerful means of selection" ([57], p. 287-8). In Pareto's ideology this region forms a future aristocracy which will eventually rise to the rich end of the distribution and form the leadership of the country. Since selection does not apply to the rich, this will, however, lead to degeneration in this "social stratum". If this is paired with an "accumulation in the lower strata of superior elements which are prevented from rising" ([57], p. 288), a revolution is unavoidable.

Pareto's ideas for changes in society, based on *social Darwinism*, are today no longer acceptable to the majority of people. However, Pareto's idea that a static distribution of wealth does not imply a static society holds true. People are able to move along the wealth axis in both directions, although in some societies, this movement appears not to be too prevalent. Indeed, it is often found that being born to parents at the poor end of the wealth distribution greatly reduces the chances of obtaining a university education, which may form the basis for a high income in later life.

Since the distribution of wealth appears fixed, the main indicator for the degree of development of a society, according to Pareto, is the amount of wealth per person. If this increases, as in Pareto's example of England in the 19^{th} century, it provides "individuals with good opportunities to grow rich and rise to higher levels of society" ([57], p. 296). Others, who are often in close contact with those at the lower reaches of society, such as the homeless, dispute this argument and point to other studies that suggest some forms of intervention are required in order to provide social justice across society.

The kind of thinking and explanation based on opinion is not one followed by those physicists who have begun to examine income distribution data. Physicists are basically driven by empiricism, an approach exemplified by Kepler who as a result of rather painstaking observations of the motion of planets, proposed his law of planetary motion. In similar vein, physicists and some economists have begun to construct models based on some underlying mechanisms that allow money to flow throughout a system and, in so doing, link these microscopic mechanisms to the overall distribution of income. Some of the models may be criticised as *naïve* by the economics community, however, as we shall see, at least the models that do emerge seem capable of predicting distributions of money that are observed to a greater or lesser degree. These advances allow for a rational debate, and through further research, advances in prediction to be made.

2.2.3 Robert Gibrat and rules of proportionate growth

The French economist, Robert Gibrat realised that the power law distribution did not fit all the data and proposed a law of proportionate effect, *la loi de l'effet proportionnel* [58]. This states that a small change in a quantity is independent of the quantity itself. The quantity $dz = dx/x$

should therefore be Gaussian distributed. Hence x should be distributed according to a log normal distribution. As a result of studying the empirical data he generalises the statement and concludes that $z = a \ln(x - x_0) + b$. This leads him to the Gaussian distribution:

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right) \tag{2.5}$$

which gives from the relation $f(z)dz = p(x)dx$:

$$p(x) = \frac{1}{\sqrt{2\pi}x\sigma} \exp\left[-\frac{(a\ln(x - x_0) + b)^2}{2\sigma^2}\right] \tag{2.6}$$

Gibrat's distribution can be represented as sketched in Figure 2.3.

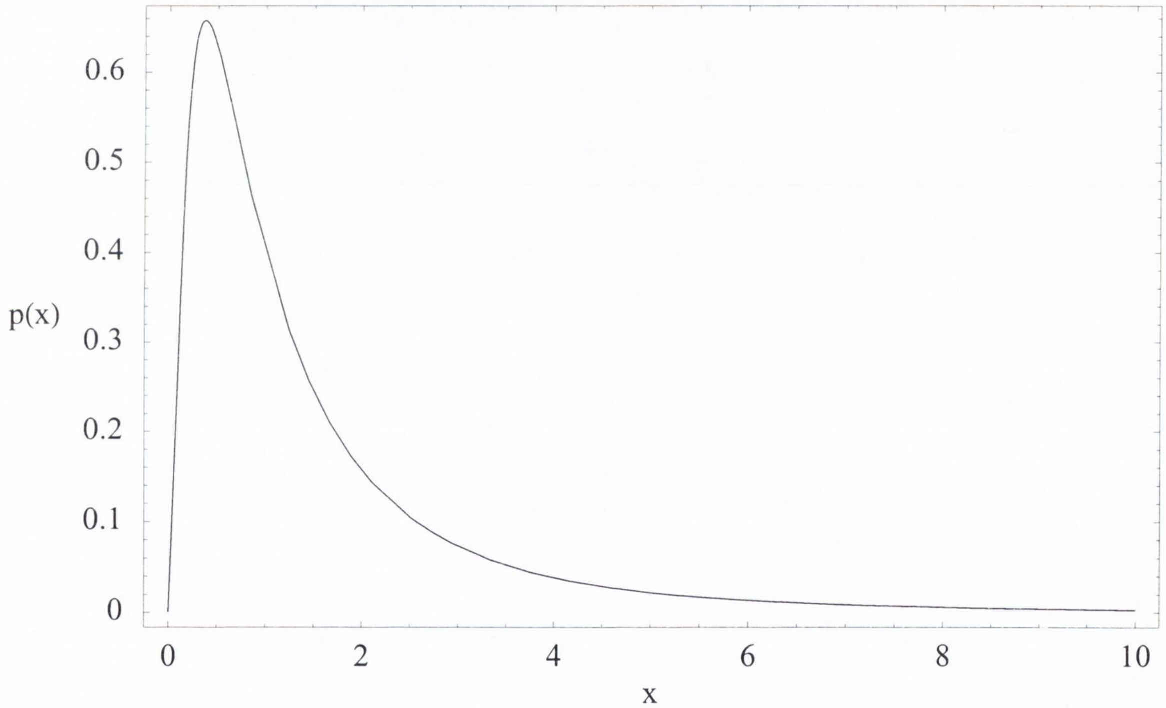


Figure 2.3: Representation of the Gibrat's distribution (eq. 2.6).

Since his argument was based on the statistics of Gauss, Gibrat felt it was a better approach than that of Pareto. Gibrat defined $100/a$ to be an inequality index. The parameter a is today related to the Gibrat index.

Recently, Fujiwara [59] has shown, using very detailed Japanese data where the variation of individual incomes can be identified over time, how in the power law region, Gibrat's law and the condition of detailed balance, i.e, $P_{12}(x_1, x_2) = P_{12}(x_2, x_1)$, both hold. The probability $P_{12}(x, y)$ represents the joint distribution of an agent to have a quantity x at time 1 and a quantity y at time 2. The condition of detailed balance shows that the probability of changing from x to y is the same as that for its reverse process. He then shows that under the condition of detailed balance, Gibrat's law implies Pareto-Zipf law, but not vice versa.

2.2.4 The stochastic model of Champernowne

An early stochastic model which reproduces Pareto law is due to Champernowne [60]. Champernowne studied different functions to fit the empirical income data such as the distribution of incomes of USA in 1929 [61]. The cumulative function used in that work was:

$$F(t) = \frac{N}{\theta} \tan^{-1} \left[\frac{\sin \theta}{\cos \theta + \left(\frac{t}{t_0} \right)^{\alpha}} \right] \tag{2.7}$$

where t denotes the income and N, α, t_0 and θ are fitted parameters. For high income this formula approximates to the Pareto's law:

$$F(t) \sim Ct^{-\alpha}. \tag{2.8}$$

The purpose of the stochastic model of Champernowne was to seek theoretical reasons for the Pareto behaviour in the high end of income distributions.

The basic idea is that an individuals' income in one year "may depend on what it was in the previous year and on a chance process" ([60], p. 319). Based on the definition of certain ranges of income, Champernowne specifies the probability for the income of an individual to change with time from one income range to another. Mathematically such a process may be expressed in terms of a vector $X_r(t)$, specifying the number of income receivers in the income range r at time t, and a set of stochastic transition matrices $p'_{rs}(t)$ that represent the proportion of individuals in income range r at time t which move to income range s in time $t+1$:

$$X_s(t+1) = \sum_{r=0}^{\infty} X_r(t) p'_{rs}(t) \tag{2.9}$$

Champernowne was able to show that provided "the stochastic matrix is assumed to remain constant throughout time [...] the distribution will tend towards a unique equilibrium distribution depending on the stochastic matrix but not on the initial distribution" ([60], p. 318) of income. These equilibrium distributions are described by a Pareto law. Obviously, the details of the transition matrix $p'_{rs}(t)$ are crucial. In the simplest case considered by Champernowne, income increases are allowed only by one range each year whereas decreases may occur over n income ranges. Furthermore, the transition probabilities are treated as independent of the present income. With these assumptions and empirical data Champernowne determined a transition matrix and deduced that his model exhibited an equilibrium solution. The form was given by the Pareto law for the whole range of incomes. Because this is not even approximately obeyed for low incomes Champernowne introduce two generalisations of the model to achieve the weak form of the Pareto law (Pareto law as the asymptotic result of the distribution for high incomes). First he allowed annual income increases over more than one range, and secondly he linked the possible range of jumps to the income before the jump, except for high incomes. Further modifications by Champernowne that took account of the age of an individual, and occupation dependent prospects did not change the result. Champernowne concludes his paper with the observation that his models "do not throw much light on the mechanism that determines the actual observed values for Pareto's alpha" ([60], p. 349). This makes it impossible to draw "any simple conclusions about the effect on Pareto's alpha of various redistribution policies" ([60], p. 351). This remark is interesting in the sense that it is clear that Champernowne was already thinking how to engineer specific income distributions within society.

2.2.5 The model for distribution of incomes within an enterprise by Lydall

After the work of Champernowne [60], H. F. Lydall introduced some criticism to the previous stochastic model [62], arguing that different sources of income showed different distributions. Lydall explained that the hypothesis of Champernowne's stochastic model is acceptable in the case of income from capital, because gains and losses of investment are dependent on quantity of money invested, but the author cannot see any explanation for the income from employment to change from one year to the next with the mechanism introduced by Champernowne [60]. So, Lydall introduce a model for the distribution of incomes within an enterprise. One observation of the empirical data for income from employment is that the slope, of the distribution of income, of the high end is steeper than the slope for total income. The main reason is that this type of income is more equally distributed than income from capital. Lydall suggested that the structure of employees in a company is arranged in the form of a pyramid with a managing director on the top and a large group of employees at the bottom. The basic hypothesis of this model [62] are the notation y_i for the number of employees in the grade i (where $i = 1$ is the lowest) and assumption that the ratio between the number of supervisors (y_{i+1}) and the number of persons supervised (y_i) is fixed:

$$\frac{y_i}{y_{i+1}} = n \tag{2.10}$$

where n is constant. Also, representing the income of each grade as x_i, Lydall assume that the income of a supervisor is related to the *aggregate income* of the employees on the grade below him in this simple way:

$$\frac{x_{i+1}}{nx_i} = p \tag{2.11}$$

where p is constant. With this simple model, Lydall is able to compute a distribution of incomes within an enterprise as a Pareto law, with a Pareto exponent that only depends on the constants n and p. The cumulative distribution of incomes is then given by:

$$Y = bx^{\log p/\log np} \tag{2.12}$$

where Y is the number of incomes exceeding any level of income x and b is a constant. From empirical Pareto exponents for different countries, Lydall was able to find the relation between np and n for each country as for example, United Kingdom and Poland [62], and from here conclude about the degree of inequality and the structured of the pyramid of incomes for each example.

2.2.6 Benoit Mandelbrot's weighted mixtures and maximum choice

While Champernowne presents a particular model with some variations that reproduces an income distribution which follows Pareto law, Benoit Mandelbrot comes to a more general conclusion in his 1960 article *The Pareto-Lévy Law and the Distribution of Income* [6] and his informal and non-mathematical paper *New Methods in Statistical Economics* [7]. According to Mandelbrot: "random variables with an infinite population variance are indispensable for a workable description of price changes, the distribution of income, and firm sizes etc." ([7], p. 421). This statement is based on Mandelbrot's observation that "essentially the same law continues to be followed by the distribution of income, despite changes in the definition of this term." ([6], p. 85)

To understand the relevance of the statement one needs to consider the so-called stability of a probability distribution. In the case of a Gaussian distribution it is known that numbers made up from the sum of independent Gaussian variables are again Gaussian distributed. This stability under summation (or invariance under aggregation) is, however, not restricted to Gaussian. It also holds for Pareto-Lévy distributions with index α between 1 and 2 (Mandelbrot introduces the term Pareto-Lévy in honour of his former supervisor Lévy, who studied the properties of stable distributions). Such distributions had generally not been considered in the economical literature. The fact that their moments may be infinite led the scientific community to ignore them for many years, partly on the grounds that they did not seem to correspond to physical reality. However, Mandelbrot dedicates a long section [7] to scale invariance of such distributions. Specifically he considers the stability of the Pareto-Lévy distribution under three different transformations:

- Linear aggregation or simple addition of various quantities in their common natural scale;

- Weighted mixture;

- Maximising choice, the selection of the largest quantity in a set.

Invariance under aggregation may be met by using Gaussian as well as Pareto-Lévy distributions. The example used is the one of various kinds of income separately or the aggregates of all the sources of income. Mandelbrot notes that the common belief that only the Gaussian is invariant under aggregation is correct only if one excludes random variables with infinite moments. The idea behind his concept of so-called weighted mixtures covers the case where the origin of an income data set is not known. If this is the case, one may consider that it was picked at random from a number of possible basic distributions. The distribution of observed incomes would then be a mixture of the basic distributions. With companies or firms "the very notion of a firm is to some extent indeterminate, as one can see in the case of almost wholly owned, but legally distinct subsidiaries" ([7], p. 424). The third property, that Mandelbrot termed maximising choice, is related with the extreme events that happen in a time series of stock market, or in the World history: "it may be that all we know about a set of quantities is the size of the one chosen by a profit maximiser. Similarly, if one uses historical data, one must often expect to find that the only fully reported events are the exceptional ones, such as droughts or floods, famines, [...]" ([7], p.424). Mandelbrot also refer that many data are just a mixture of these extreme cases with the full report.

Such transformations need not be the only ones of interest, however, they are so important that they should characterise the laws they leave invariant. In this sense, the observation that income distributions are the same whatever the definition of income, is used by Mandelbrot to support the claim that they are Paretian. Mandelbrot summarises: "It is true that incomes (or firm sizes) follow the law of Pareto; it is not true that the distributions of income are very sensitive to the methods of reporting and of observation." ([7], p. 425).

2.3 Empirical studies

Numerous recent empirical studies have all shown that the power law tail is an ubiquitous feature of income distributions. The value of the exponent may vary with time and depends on the source of the data. However, over 100 years after Pareto's observation, a complete understanding of the shape and dynamics of wealth distribution is still evasive. This is partly due to incomplete data, but may also reflect the fact that there might indeed be two distributions, one for the super rich, one for the low to medium rich, with some intermediate region in-between. For example, for U.S.A. only the top three percent of the population follow Pareto's weak law, the vast majority of people appear to be governed by a completely different law. The distribution function for the majority of the population seems to fit a different curve. The main functions used in the literature for fitting the income or wealth distributions are:

- the log-normal distribution function:

$$p_{LN}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] \tag{2.13}$$

$$P_{LN}(x) = \frac{1}{2}\left[1 - erf\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)\right] \tag{2.14}$$

where μ and σ are parameters for the mean $\exp\left(\mu + \sigma^2/2\right)$ and the variance $\left(\exp\left(\sigma^2\right) - 1\right)\exp\left(2\mu + \sigma^2\right)$. The function $\mathrm{Erf}(x) = \left(2/\sqrt{\pi}\right)\int_0^x e^{-t^2}dt$ is the error function. The log-normal distribution becomes Gibrat's distribution (equation 2.6) for $\mu = 0$ and $\sigma = 1$.

- the Gamma distribution function:

$$p_\Gamma(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} \exp\left[-\frac{x}{\theta}\right] \tag{2.15}$$

$$P_\Gamma(x) = \frac{\Gamma\left[\alpha, x/\theta\right]}{\Gamma(\alpha)} \tag{2.16}$$

where α and θ are parameters for the mean $\alpha\theta$ and variance $\alpha\theta^2$. The function $\Gamma\left[a, x\right] = \int_x^\infty t^{a-1}e^{-t}dt$ is known as the upper incomplete Gamma function. The Gamma distribution becomes the Boltzmann-Gibbs distribution for $\alpha = 1$.

- the generalised Lotka-Volterra distribution function:

$$p_{GLV}(x) = \frac{(\alpha-1)^\alpha}{\Gamma(\alpha)} \frac{1}{x^{1+\alpha}} \exp\left[-\frac{\alpha-1}{x}\right] \tag{2.17}$$

$$P_{GLV}(x) = 1 - \frac{\Gamma\left[\alpha, (\alpha-1)/x\right]}{\Gamma(\alpha)} \tag{2.18}$$

where α is a parameter for the variance $1/(\alpha - 2)$ and the mean is equal 1. The parameter α is also present in the exponent of the power law distribution for $x \to \infty$.

Sketches of these distributions are shown in Figure 2.4 for some parameters, where we can see the difference between all of them for different plot scales. The only distribution that will allow the presence of a strong power law in the limit of high values is the generalised Lotka-Volterra.

Another distribution presented by Reed [63] is the double Pareto log-normal distribution (dPlN) which is characterised by a power law for high and low values of the distribution and a log-normal behaviour in the middle range of the distribution. The probability in each case can be represented by:

$$\begin{aligned} P(X \geq x) &\sim \quad x^{-\alpha} \quad \text{for } x \to \infty \\ P(X \leq x) &\sim \quad x^{\beta} \quad \text{for } x \to 0 \end{aligned} \tag{2.19}$$

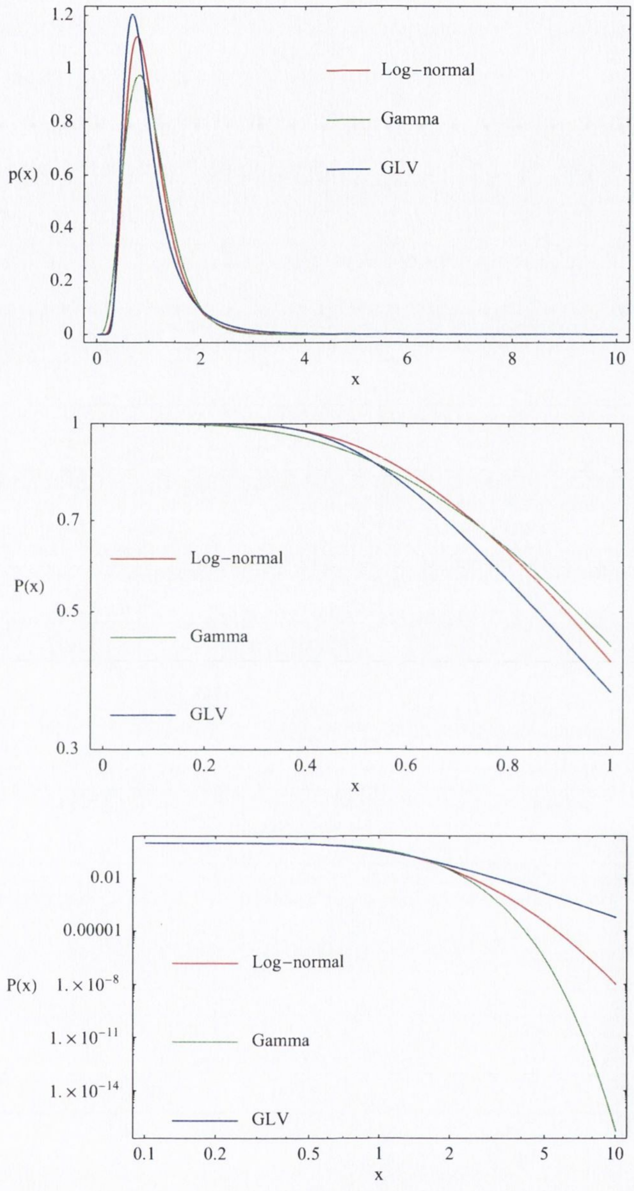


Figure 2.4: Representation of the probability distribution functions (top figure) (Eq. 2.13, 2.15 and 2.17) for log-normal (red), gamma (green) and generalised Lotka-Volterra (blue), respectively. Representation of the cumulative distribution functions (centre and bottom figures) (Eq. 2.14, 2.16 and 2.18) for log-normal (red), gamma (green) and generalised Lotka-Volterra (blue), respectively. The sketches are for values of the parameters that allow the three distributions to have similar mean and variance. The centre figure is in a log-linear plot and the bottom figure in a log-log plot.

where the probabilities represent cumulative distributions and the exponents α and β are the Pareto exponents for the upper and lower limits of the distribution, respectively.

Tables 2.2 and 2.3 summarises recent empirical studies according to the type of data, the distribution function used to fit the data and the value of Pareto exponent in the case of fitting to a power law distribution. While in the empirical studies made by Pareto, α is around 1.5 (Table 2.1), in the later studies exponents less than 1.5 almost only feature for the studies of wealth statistics from top wealthiest in the society and are generally obtained from lists of super rich people, published in magazines (Table 2.3). For all the other studies, which are mainly obtained from income distributions, the exponent is bigger than this.

Table 2.2: Table of empirical data for income distributions. In column Source, Tax means data from income tax statistics. In column Distributions: Par. - Pareto tail; LN - Log-normal; Exp. - Exponential; dPlN. - Double Pareto Log-normal; G. - Gamma.

Country	Source	Distributions	Pareto Exponents	Ref.
Japan	Tax (1992)	Par.	$\alpha = 2.057 \pm 0.005$	[23, 64, 65]
	(1998)	Par.	$\alpha = 1.98$	
	Tax (1998)	Par.	$\alpha = 2.05$	
	Tax (1998)	LN / Par.	$\alpha = 2.06$	
	(1887-2000)	LN / Par.	$\bar{\alpha} \sim 2.0$[a]	
U.S.A.	(1992)	Exp.		[66, 67, 68, 69, 70]
	(1996)	Exp.		
	Tax (1997)	Exp. / Par.		
	(1998)	Exp. / Par.	$\alpha = 1.7 \pm 0.1$	
	(1983-2001)	Exp. / Par.	$\alpha \sim 1.4 - 1.8$	
U.K.	(1994-98)	Exp. / Par.	$\alpha \sim 2.0 - 2.3$	
Australia	(1989-2000)	Exp. / LN / Par.		
U.S.A.	(1997)	dPlN.	$\alpha = 22.43$ / $\beta = 1.43$	[63]
Sri-Lanka	(1981)	dPlN.	$\alpha = 2.09$ / $\beta = 3.09$	
Australia	(1993-97)	Par.	$\alpha \sim 2.2 - 2.6$	[71]
U.S.A.	(1980)	G. / Par.	$\alpha = 2.2$	[72]
	(1989)	G. / Par.	$\alpha = 1.63$	
	(2001)	G.		
U.K.	(1998-99)	G. / Par.	$\alpha = 1.85$	
Portugal	Tax (1998-2000)	Par.	$\alpha \sim 2.30 - 2.46$	[73]
Italy	(1977-2002)	LN / Par.	$\alpha \sim 2.09 - 3.45$	[74, 75]
U.S.A.	(1980-2001)	LN / Par.	$\alpha \sim 1.1 - 3.34$	
U.K.	(1991-2001)	LN / Par.	$\alpha \sim 3.47 - 5.76$	
Germany	(1990-2002)	LN / Par.	$\alpha \sim 2.42 - 3.96$	
India	(2003)	Par.	$\alpha = 1.51$	[76]
U.K.	(1992-2002)	Par.	$\alpha \sim 2.68 - 3.34$	[24]

[a] This value is an average Pareto exponent.

In Figure 2.5 we show the cumulative distribution of wealth for the top richest in the world, for two different years, 2003 and 2006. The data was downloaded from the list of super rich

Table 2.3: Table of empirical data from wealth lists. All the distributions are Pareto like.

Country	Source	Pareto Exponents	Ref.
U.S.A.	1996[a]	$\alpha = 1.36$	[77, 78, 79]
	1997[b]	$\alpha = 1.35 \pm 0.005$	
U.K.	1970		
	1997[c]	$\alpha = 1.06 \pm 0.004$	
Sweden	1965	$\alpha = 1.66$	
France	1994[d]	$\alpha = 1.83 \pm 0.030$	
U.K.	1996[e]	$\alpha = 1.9$	[67]
Egypt	S.H. (14^{th}B.C.)[f]	$\alpha = 1.59 \pm 0.19$	[80]
U.K.	1996	$\alpha = 1.85$	[72]
U.K.	2001[g]	$\alpha = 1.78$	[35, 81]
India	2002-2004[h]	$\alpha \sim 0.81 - 0.92$	[76]
World	2003/2006[i]	$\alpha \sim 1.37 - 1.38$	[24]
Hungary	1550[j]	$\alpha = 0.92$	[81, 82]
India	1991/2002[k]	$\alpha \sim 1.8 - 2.4$	[83]

[a] 400 wealthiest people, by Forbes.
[b] Top wealthiest people, by Forbes.
[c] Top wealthiest people, by Sunday Times.
[d] French almanac Quid list.
[e] Inheritance tax
[f] Related to the size of houses found in an archaeological study.
[g] Inheritance tax
[h] 125 wealthiest individuals in India by Business Standard magazine and 40 richest Indians by Forbes.
[i] Top wealthiest people, by Forbes.
[j] Number of serf families living on a nobleman's land.
[k] All India Debt and Investment Survey.

people in Forbes magazine [36].

2.4 Current theoretical studies

From the work of Pareto, Gibrat, Champernowne and Mandelbrot until few years ago, there was not many improvements in the study of wealth and income distributions, apart from some different versions of the same models presented in the previous sections. Most of these versions were based on adding more and more parameters to the simple models to fit a function to the overall empirical distribution of incomes. So we had to wait almost one century from the first work of Pareto to welcome an impressive amount of empirical and analytical studies from the Physics community. The main reason for this could be that a huge amount of data became available by the government authorities, which need a statistical approach based on the kind of

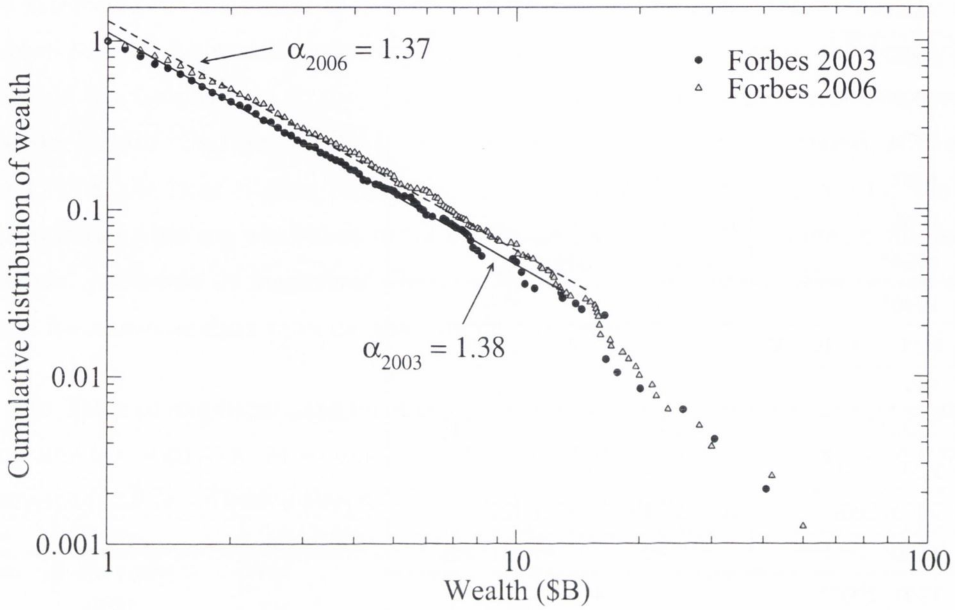


Figure 2.5: Cumulative distribution of wealth for the top richest people in the World. The data was taken from Forbes Magazine for the years 2003 (476 individuals) and 2006 (793 individuals). The wealth is in billions of dollars.

one used in Physics. Another reason is that Physics is the science which tries to explain the mechanisms of Nature, so what better science to explain the mechanisms of exchange of money than Physics?

In this section we discuss some recent models for wealth distributions in greater detail.

2.4.1 Generalised Lotka-Volterra Model

The Generalised Lotka-Volterra model, based on an ecological model with the same name, brought new ideas of the meaning of the trading terms in a model of exchange of money [84]. If all agents in a system are characterised by a term w that changes in time (e.g. income, wealth) and one agent, i, is chosen at random at time t to exchange some money with all the others, then the term w_i of this agent will be updated. The governing equation is given by:

$$w_i(t+1) = \lambda(t)w_i(t) + a\bar{w}(t) - bw_i(t)\bar{w}(t) \tag{2.20}$$

$$w_j(t+1) = w_j(t), \text{ for } j \neq i \tag{2.21}$$

where $w_i(t)$ is the system component i at time t. This system component can be analysed as wealth of agent i. The average value $\bar{w}(t)$, which is not constant in time because there is non-conservation of money, is represented as:

$$\bar{w}(t) = \frac{1}{N}\sum_{i=1}^{N} w_i(t) \tag{2.22}$$

The first term on the rhs of equation 2.20 describes the effect of autocatalysis for individual i and can be interpreted as investments return. The second term describes the same autocatalysis effect but for all the agents. One interpretation of this is that it simulates in a simplistic way the effect of a tax or social security policy. The third term describes competition for limited resources and it represents external limiting factors: finite amount of resources and money in the economy, technological inventions, wars, disasters, etc. With proper values for the constants a and b a power law is produced in the high end of the distribution of w:

$$P(w) \sim w^{-1-\alpha} \tag{2.23}$$

with an exponent α in agreement with empirical values found by Pareto. The power law is shown from numerical analysis [84].

An analytical solution of the complete model was achieved [25, 26, 85, 86], by reformulating equation 2.20 as a Langevin equation. This is done rewriting the equation 2.20 as:

$$w_i(t+1) - w_i(t) = \lambda(t)w_i(t) - b\bar{w}(t)w_i(t) + a\bar{w}(t) - w_i(t) \tag{2.24}$$

and summing both sides of equation over i:

$$\bar{w}(t+1) - \bar{w}(t) = \sum_i \lambda(t)w_i(t) + \left[-b\bar{w}(t) + a - 1\right]\bar{w}(t) \tag{2.25}$$

It can be shown that the first term of the rhs is represented by $\bar{\lambda}\bar{w}(t)$, so in the long time limit, assuming a stationary state, can be written as $b\bar{w} = \bar{\lambda} + a - 1$. Considering a normalised wealth, $w_i(t) \to w_i(t)/\bar{w}(t)$, and writing the difference between wealth as first derivatives gives the following Langevin equation:

$$\frac{dw_i}{dt} = \left[\lambda(t) - \bar{\lambda} - a\right] w_i + a \tag{2.26}$$

The solution for the probability distribution $P(w)$ is given as:

$$P(w) = \frac{1}{Zw^{2+a/D}} \exp\left[-\frac{a}{Dw}\right] \tag{2.27}$$

where Z is a normalisation factor. This solution represents a distribution with a power law for $w \to \infty$ with an exponent $2+a/D$ that does not depend on the constant b, so the competition for limited resources doesn't have a direct effect on the power law distribution of w. This solution is correct for an infinite system of agents. For finite systems the following corrections can be derived for the probability distribution:

$$P(w) = \frac{1}{w^{1+\alpha}} \exp\left[-\frac{a}{Dw}\right] \exp\left[-\frac{a}{D(1-w/N)}\right] \tag{2.28}$$

and the exponent:

$$\alpha = 1 + \frac{a/D - K}{1+K} \tag{2.29}$$

where $K = N^{-2+2/\alpha} \sim N^{-(2a/D(1+a/D))}$ and N is the total number of agents.

The generalised Lotka-Volterra model was studied by Bouchaud and Mézard [27] with a different analytical approach, but with the same kind of result for the distribution of w:

$$P(w) = \frac{(\alpha-1)^{\alpha}}{\Gamma(\alpha)} \frac{1}{w^{1+\alpha}} \exp\left[-\frac{\alpha-1}{w}\right] \tag{2.30}$$

with $\alpha = 1 + a/D$. The authors also show that there is a phase transition between a system where the wealth is evenly distributed between agents and a system where just a few agents have almost all the wealth. This phase transition is around $\alpha = 1$, where this wealth condensation appear for $\alpha < 1$. Other features of this model are studied, as for example: the inclusion of taxation terms - an income tax, over the change of the wealth $\frac{dw}{dt}$ and a capital tax, over the total wealth w; the incorporation of a network, where instead of treating the model in a mean field approach (all agents trade with all the others), each agent as a specified number of neighbours with who it is able to interact. In Figure 2.6 a representation of this distribution of equation 2.30 is shown.

2.4.2 Collision Models

In 1960, Mandelbrot wrote "There is a great temptation to consider the exchanges of money which occur in economic interaction as analogous to the exchanges of energy which occur in physical shocks between gas molecules. In the loosest possible terms, both kinds of interactions *should* lead to *similar* states of equilibrium. That is, one *should* be able to explain the law of income distribution by a model similar to that used in statistical thermodynamics: many authors have done so explicitly, and all the others of whom we know have done so implicitly." ([6], pg. 83). Unfortunately Mandelbrot does not provide any references to that specific body of work.

In analogy to two-particle collisions with a resulting change in their individual momenta, income exchange models may be based on two-agent interactions. Here two randomly picked agents exchange money by some pre-defined mechanism. Assuming the exchange process does

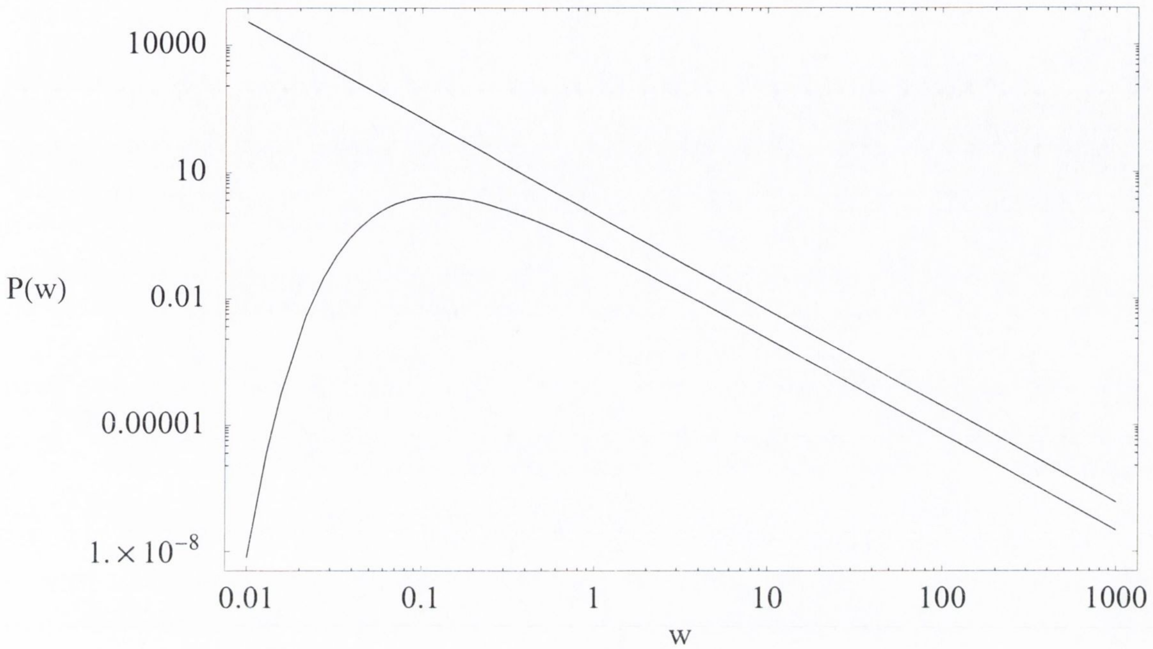


Figure 2.6: Representation of probability density distribution (eq. 2.30), in a log-log scale, with parameters $a = 0.00023$ and $D = 0.00083$ as used in [26]. The straight line is a power law distribution with exponent $-1-\alpha$, where α has the value $1 + a/D \simeq 1.277$.

not depend on previous exchanges, the dynamics follows a Markovian process as follow:

$$\begin{pmatrix} m_i(t+1) \\ m_j(t+1) \end{pmatrix} = \mathbf{M} \begin{pmatrix} m_i(t) \\ m_j(t) \end{pmatrix} \tag{2.31}$$

where $m_i(t)$ is the income of agent i at time t and the collision matrix $\mathbf{M}$ defines the exchange mechanism.

The first model of that kind appears to be that of Angle [87] who proposed that exchange occurs by one agent getting a percentage of the money of another agent, which means that what one party gains, the other loses. The model is explained as two agents that come together and put a percentage of their money, $1-\lambda$, in a bag and one randomly chosen agent takes everything. In the formulation of equation 2.31 this gives:

$$\begin{pmatrix} m_i(t+1) \\ m_j(t+1) \end{pmatrix} = \begin{pmatrix} 1-(1-\epsilon)(1-\lambda(t)) & \epsilon(1-\lambda(t)) \\ (1-\epsilon)(1-\lambda(t)) & 1-\epsilon(1-\lambda(t)) \end{pmatrix} \begin{pmatrix} m_i(t) \\ m_j(t) \end{pmatrix} \tag{2.32}$$

where $\lambda(t)$ is a random variable, from a uniform distribution between $[0,1]$, related with the percentage of money that agents save before the exchange and ϵ is a variable with two possible values, zero or one, taken from a specific probability distribution. The amount exchanged in each trade is:

$$\Delta m(t) = (\epsilon - 1)\,(1 - \lambda(t))\, m_i(t) + \epsilon\,(1 - \lambda(t))\, m_j(t) \tag{2.33}$$

which takes the value $\Delta m = (1 - \lambda(t))\, m_j(t)$ if $\epsilon = 1$ and $\Delta m = -\,(1 - \lambda(t))\, m_i(t)$ if $\epsilon = 0$. So depending on the value of ϵ each agent has the possibility of taking some money from the other. In this exchange process the money is conserved, because:

$$m_i(t+1) - m_i(t) = -\left[m_j(t+1) - m_j(t)\right] = \Delta m(t) \tag{2.34}$$

Numerical simulations of this model showed that the distribution of incomes, $p(m)$ can be fitted by a Gamma probability density function:

$$p(m) = \frac{m^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp\left[-\frac{m}{\beta}\right] \tag{2.35}$$

with fit parameters α and β.

Another study of models of asset exchange was presented by Ispolatov *et al.* [88] where the authors study both additive and multiplicative asset exchange numerically and analytically. Comparisons are drawn with the model of an ideal gas, where exchanges of momenta between collisions are substituted by exchanges of money.

Combining the model of Angle [87] with the idea of a closed economic system where money is conserved at each trade, Yakovenko *et al.* [89] performed simulations and found that the equilibrium probability distribution of money $p(m)$ should follow a Boltzmann-Gibbs law:

$$p(m) = \frac{1}{T} \exp\left[-\frac{m}{T}\right] \tag{2.36}$$

where T is the temperature of the system, which in this case is the mean value of money $T = M/N$, where M is the total money and N is the number of agents in the system. The Boltzmann-Gibbs law is equal to the Gamma probability density function (eq. 2.35) for $\alpha = 1$ as can be seen in Figure 2.7.

Chakraborti *et al.* [28, 29, 30, 31] considered the same kind of pairwise exchange models but included a constant saving propensity term for all agents, λ and changed the possibility of one of the agents to win everything. The governing equation for the evolution of money of agents i and j is:

$$\begin{pmatrix} m_i(t+1) \\ m_j(t+1) \end{pmatrix} = \begin{pmatrix} \lambda + \epsilon(t)(1-\lambda) & \epsilon(t)(1-\lambda) \\ (1-\epsilon(t))(1-\lambda) & \lambda + (1-\epsilon(t))(1-\lambda) \end{pmatrix} \begin{pmatrix} m_i(t) \\ m_j(t) \end{pmatrix} \tag{2.37}$$

where $\epsilon(t)$ is a random number between [0, 1] and λ is the saving parameter (same for all agents). So, when two agents meet, they will save some percentage of their money $\lambda m_i(t)$ and $\lambda m_j(t)$,

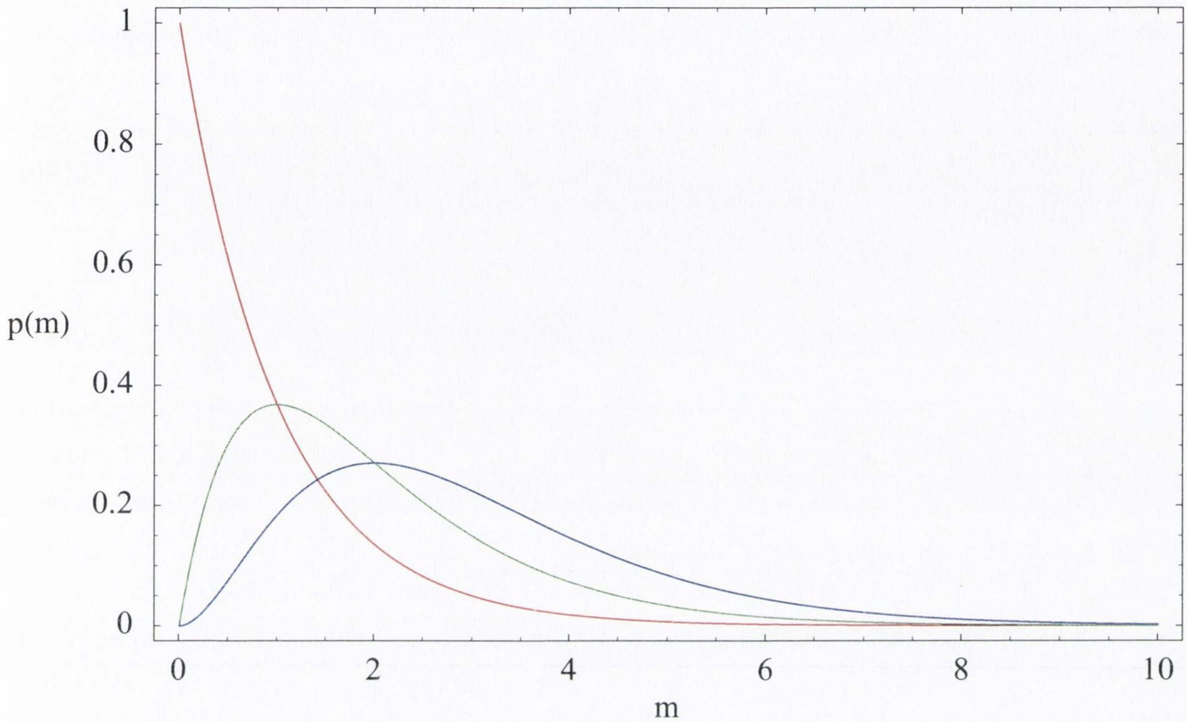


Figure 2.7: Representation of probability density distribution (eq. 2.35), with parameters $\beta = 1.0$ and α equals 1.0 (red), 2.0 (green) and 3.0 (blue), respectively. The case of $\alpha = 1.0$ is the Boltzmann-Gibbs distribution (for $T = 1$ in this case).

respectively, and throw the rest to, what we will call, a bag. The total amount of money in the bag is then $(1 - \lambda)(m_i(t) + m_j(t))$ and will be split between the two agents in percentages defined by the value of the random variable $\epsilon(t)$. If $\epsilon(t) > 0.5$, agent i will get more percentage than agent j and vice-versa. If λ is equal to zero we return to the model of Yakovenko [89]. This model is the same as the one considered by Angle with different features for the parameters, λ and ϵ. The distribution of money, $p(m)$ is no longer a Boltzmann-Gibbs distribution for $\lambda \neq 0$, but was found to be well described by the function [30]:

$$f_n(m) = \frac{1}{\Gamma(n)} \left(\frac{n}{T}\right)^n m^{n-1} \exp\left(-\frac{nm}{T}\right) \tag{2.38}$$

where $\Gamma(n)$ is the Gamma function of n, T is the money per agent in the system and n is defined as:

$$n(\lambda) = 1 + \frac{3\lambda}{1 - \lambda} \tag{2.39}$$

Chatterjee *et al.* [32, 33] improved the previous model including an inhomogeneous parameter λ, because in a society the interest of saving may vary from individual to individual, so λ_i is the saving parameter for agent i and the equation for the evolution of money of agents i and j, m_i and m_j is then given by:

$$\begin{pmatrix} m_i(t+1) \\ m_j(t+1) \end{pmatrix} = \begin{pmatrix} \lambda_i + \epsilon(t)(1-\lambda_i) & \epsilon(t)(1-\lambda_j) \\ (1-\epsilon(t))(1-\lambda_i) & \lambda_j + (1-\epsilon(t))(1-\lambda_j) \end{pmatrix} \begin{pmatrix} m_i(t) \\ m_j(t) \end{pmatrix} \tag{2.40}$$

where the amount exchanged is given by:

$$\Delta m = (\epsilon(t) - 1)(1-\lambda_i)m_i(t) + \epsilon(t)(1-\lambda_j)m_j(t). \tag{2.41}$$

The saving propensity parameters λ_i are fixed over time and are distributed independently from a random distribution between $[0, 1]$. This new dynamics also gives a different distribution of money in the steady state. In particular it was found that the high end of distribution of money follows a Pareto law with exponent α around 1.

The problem for the situation where all agents save the same fixed percentage of their money was solved analytically by Repetowicz *et al.* [34], who using the Boltzmann equation, obtained a relation between the probability distribution of wealth of one agent and the joint probability distribution of the wealth of two agents. Invoking a mean field approximation this approach allowed the solution to be solved via a moment expansion of the one agent distribution function. It was demonstrated that to third order, the moments agreed with the solution proposed as a result of numerical calculations (eq. 2.38). For the case when the savings are not equal but determined by some distribution, $\rho(\lambda)$, analytic calculations performed by Repetowicz *et al.* [34] demonstrated clearly that the Pareto exponent for this model is exactly one. This result appeared to hold regardless of the form of the savings distribution function, $\rho(\lambda)$.

Another analytical solution was found by Chatterjee *et al.* [90] also using the Boltzmann equation, but in this case, just restricted to the value of $\epsilon = 1/2$. The results are in agreement with the numerical results of Chatterjee *et al.* [32, 33] and the ones found by Repetowicz *et al.* [34].

A further variant of the collision model has been proposed by Slanina [91]. Slanina's model involves the pairwise interaction of agents, which at every exchange process also receive some money from outside. The time evolution of trades is represented as:

$$\begin{pmatrix} m_i(t+1) \\ m_j(t+1) \end{pmatrix} = \begin{pmatrix} 1-\beta+\epsilon & \beta \\ \beta & 1-\beta+\epsilon \end{pmatrix} \begin{pmatrix} m_i(t) \\ m_j(t) \end{pmatrix} \tag{2.42}$$

Thus agent i gives a fraction β of its money to agent j and vice versa. In addition it is assumed that additional money, $\epsilon(m_i(t) + m_j(t))$, is created in the exchange via some sort of wealth creating process. In the simplest case, the values of β and ϵ are kept constant for all trades. Since money is not conserved in this model, there is no stationary solution for the distribution of money, $p(m)$. However, as with the generalised Lotka-Volterra model, there is a stationary distribution function for the relative value: $m_i(t) \to m_i(t)/\langle m(t)\rangle$. The solution is then obtained by solving the associated Boltzmann equation within a mean field approximation. Taking the limits $\beta \to 0$ and $\epsilon \to 0^+$ whilst keeping the assumed power law α constant, yields:

$$p(m) = \frac{1}{m^{1+\alpha}} \frac{(\alpha-1)^\alpha}{\Gamma(\alpha)} \exp\left[-\frac{\alpha-1}{m}\right] \tag{2.43}$$

where $\alpha - 1 \approx 2\beta/\epsilon^2$. It is interesting to see that this result is identical in form to that given by the generalised Lotka-Volterra model (eq. 2.30).

To check the accuracy of this approximation, we performed some simulations of Slanina's model [92] for 10^4 agents trading $10^3 \times N$ times and averaged over 10^3 realisations. The percentage of wealth exchanged (β) was set to 0.005 and the percentage of wealth injected in the system (ϵ) to 0.1. A fit of the cumulative distribution of equation 2.43, with a power law exponent of 2.0, to the results of the numerical simulation for the steady state is shown in Figure 2.8. This exponent of 2.0 is in excellent agreement with the value of 2.0 of equation $\alpha - 1 \approx 2\beta/\epsilon^2$.

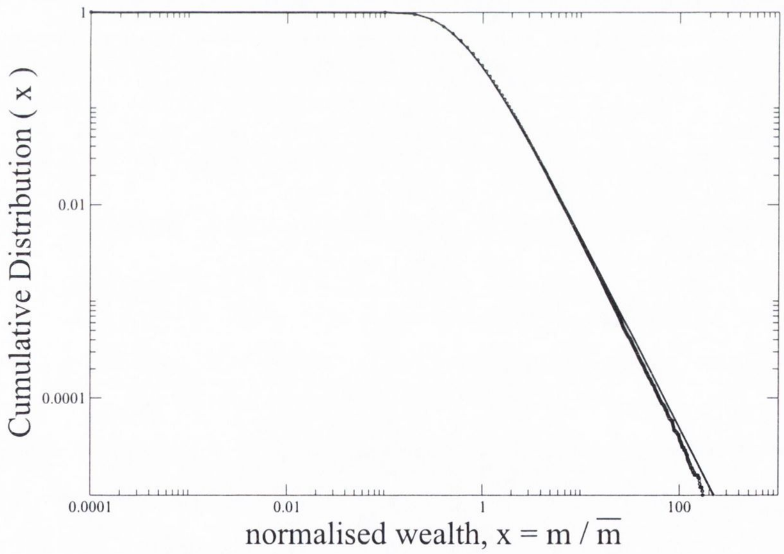


Figure 2.8: Cumulative distribution of wealth in a simple Slanina model, for 10^4 agents trading $10^3 \times N$ times and averaged over 10^3 realisations. The percentage of wealth exchanged (β) is equal to 0.005 and the percentage of wealth injected in the system (ϵ) is 0.1. The numerical simulations are represented by black dots and the cumulative distribution of the analytical solution (eq. 2.43) by the black line. The Pareto exponent for the higher end is ~ 2.0.

A review of collision models can be found in Refs. [93, 94].

2.4.3 Wealth models on networks

Coelho *et al.* [35, 81] have recently introduced the so-called family network model, which not only yields realistic wealth distributions, but also a topology of wealth. This is a model of asset exchange where the main mechanisms of wealth transfer are inheritance and social costs associated with raising a new family. The structure of the network of social (economic) interactions is not predefined but emerges from the asset dynamics. These evolve in discrete time steps in the following manner. For each time step:

- From the initial configuration (step I in Fig. 2.9), the oldest family (node) is taken away, and its assets are uniformly distributed between the families linked to it (neighbours) (step II in Fig. 2.9);
- A new family (node) is added to the system and linked to two existing families (nodes), that have wealth greater than a minimum value q (step III in Fig. 2.9);
- The small amount q, is subtracted from the wealth of the selected families (nodes) and redistributed in a preferential manner in the society. This process aims to model the wealth needed to raise a child. The preferential redistribution is justified by the fact that wealthier families control more business and benefit more from the living costs of a child;
- A portion p of the remaining wealth of each of the two families is donated to the new family as start-up money (step IV in Fig. 2.9).

The total wealth and the number of families are conserved after each time-step. Numerical calculations yield the cumulative wealth distribution in line with empirical data. For reasonable values of p and q we observe, for the upper 10% of the society, the scale-free Pareto distribution with Pareto exponents that lie between 1.8 and 2.7 (Fig. 2.10). The Pareto tail forms relatively quickly, usually after less than two generations. The degree-distribution of the family-network converges also rapidly to an exponential form. Interesting correlations between wealth, connectivity, and wealth of the first neighbours are revealed. Wealthiest families are linked together and have a higher number of links compared with the poorest families. These correlations yield new insights into the way the Pareto distribution arises in society.

2.5 Conclusions

A renewed interest in studying the distribution of income in society has emerged over the last 10 years, driven principally by the new interest of physicists in the areas of economy and sociology. This has resulted in the development of a number of theoretical models, based on concepts of

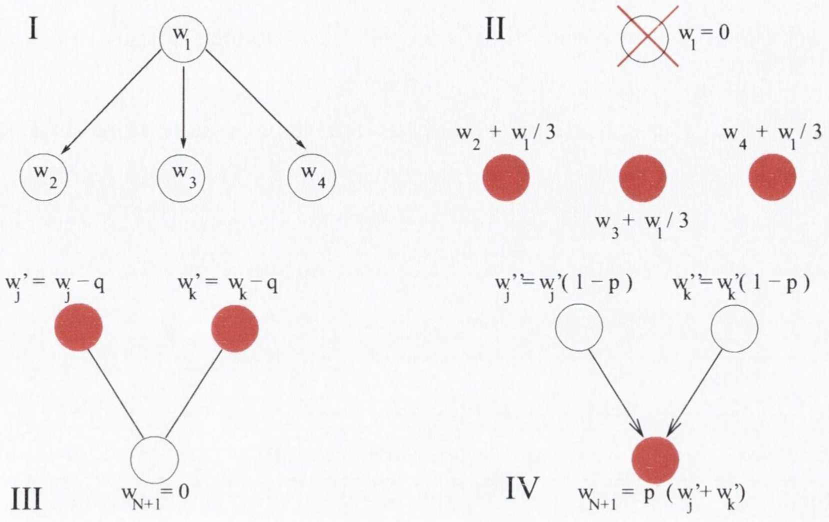


Figure 2.9: Schematic representation of a time-step in the Family Network Model. Circles represent families, edges represent the link between families and arrows the direction of transfer of money.

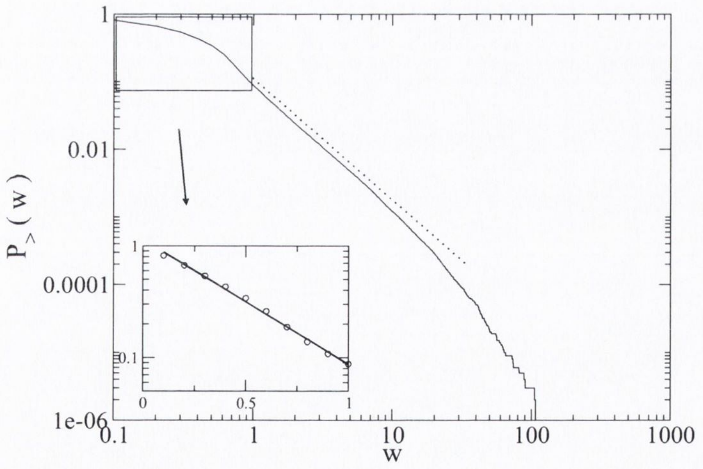


Figure 2.10: Cumulative wealth distribution function obtained from the Family Network Model ($p = 0.3$, $q = 0.7$, $N = 10000$ and results after 10 MCS). The tail is approximated by a power-law with exponent $\alpha = 1.80$, and the initial part of the curve follows an exponential. The inset shows this initial trend on a log-linear scale.

statistical physics. It now seems clear that some of these models, despite their simplicity, can reproduce key features of income distribution data.

There is still some discussion about the expressions that explain more satisfatory the low-middle end behavior of the wealth or income distributions. The more common distributions in the literature that try to explain this low end are the log-normal, the exponential and the Gamma distributions. For the higher end of the distribution of wealth or income there is an agreement about the power law behavior.

In our study we go a bit further and we try to explain the higher end of the wealth distribution with two power laws, because we saw from Tables 2.2 and 2.3 that there are two regimes of power law exponents. In next chapter we introduce a theory of double power laws and some mechanisms able to reproduce the double power law in a simple model of money exchange.

Chapter 3

Double power law in wealth distributions

3.1 Introduction

Motivated by our review of empirical data in Chapter 2, we introduce a model of wealth exchange that produces double power laws and then we try to explain it with a theory of double power laws.

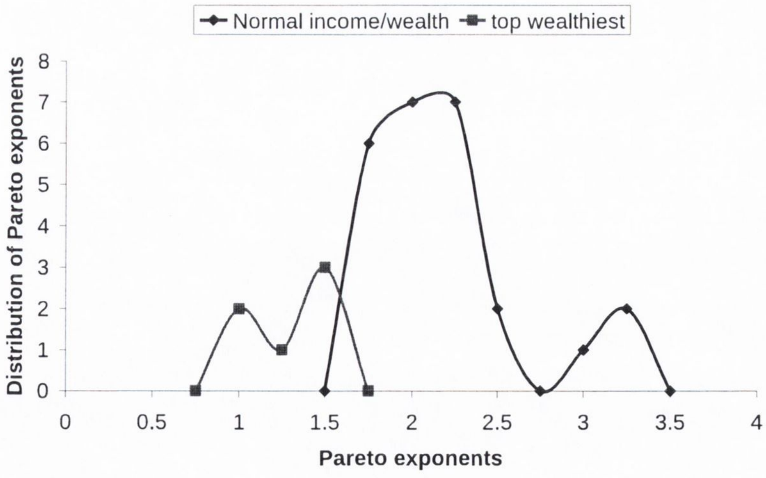


Figure 3.1: Distribution of Pareto exponents from two different sources: distributions of normal income/wealth (black) and distributions of wealth from top wealthiest lists (gray). The exponents used in this figure were taken from Tables 2.2 and 2.3.

We believe that the studies of income/wealth that are based on tax/income generally do not include the very rich people, as we can see from the study of the tables of the previous chapter and from Figure 3.1. From Figure 3.1 it seems that there is a power law exponent for millionaires that have high income/wealth but also another power law exponent for the top billionaires as found by studies from the top lists in magazines. This second exponent will be called α_2 and it is normally lower than the first one that we call α_1. A further indication of two power law regimes is the study of Souma [23] for the income in Japan. In Figure 1 of his paper [23], Souma found a Pareto exponent of 2.06 in the high end of the distribution of incomes for 1998. However, we see an indication of a second power law for the top richest (higher than 3000 million yen) which we estimate as an exponent below 1.0 based on his figure. Scarfone [95] also analysed this set of data but with a different point of view. Scarfone "claims to see" three power law regimes in this Figure from Souma. For Scarfone there are also other income distributions with two power law exponents, as the case of Japan in 1975 and U.S.A. in 2000.

In his study, the distribution is said to have an ankle when $\alpha_2 < \alpha_1$ and a knee otherwise. In Figure 3.2, we show a sketch of an ankle and a knee.

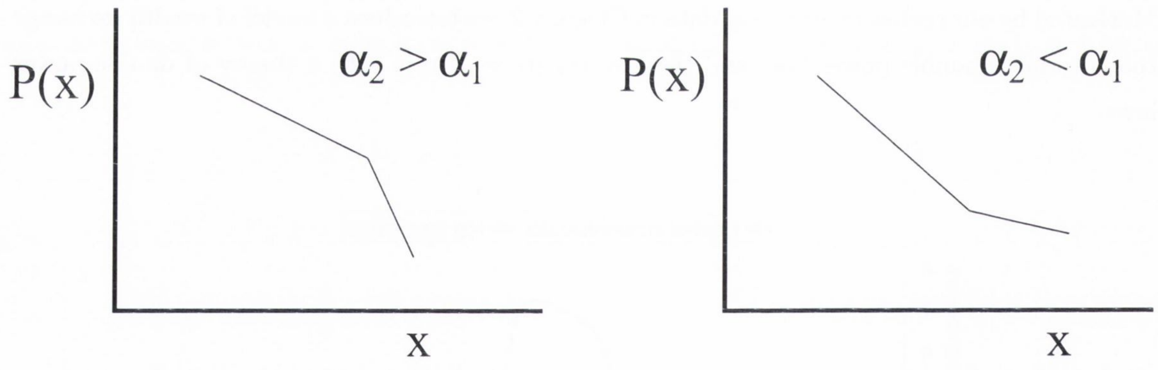


Figure 3.2: Sketch of a knee (left figure) and an ankle (right figure). The distribution of income/wealth has a knee when $\alpha_2 > \alpha_1$ where α_2 is the power law exponent of the top billionaires and α_1 is the power law exponent for the millionaires. An ankle occur when $\alpha_2 < \alpha_1$.

With the purpose of describe these distributions with only one simple analytical function, Scarfone used the concepts of generalised exponential and logarithmic functions which we will introduce in section 3.3.

The existence of two power law regimes has been observed in other fields such as dielectric relaxation [96], protein folding [97], returns of S&P500 index [98, 99], linguistics [100], cosmic-rays [101] and the value of land in Japan [102].

In the next section we introduce the expansion of Slanina model that produces double power laws.

3.2 Expansion of Slanina Model

The Slanina model [91] is a very simple model from the family of the Collision models presented in section 2.4.2. In this model the agents are chosen at random and exchange some part of their money. At every exchange process they also receive some money from outside.

The distribution of money in the stationary state gives a power law distribution for the high end of the distribution and the solution can be treated analytically. In this section we expand the Slanina model. The purpose of this is to try to create a simple collision model that will reproduce a distribution of wealth with a double power law. Adopting the Slanina theory for the case of a singular power law (section 2.4.2), we expand it allowing the presence of two regimes that will have two different values for the fraction of money that agents transfer.

Our expansion of Slanina's model is given changing the rules of equation 2.42 by making the fraction of money exchanged by each agent, β, a function of the money that the agent has at that time, m ($\beta(m)$). The main conclusion that we can take from this wealth dynamic is that a double power law arises from the difference between the percentage of money that agents put into the society for trade. This difference can be related with different levels of fear to risk or from some economical issues related with taxation. This results in the following update rule:

$$\begin{pmatrix} m_i(t+1) \\ m_j(t+1) \end{pmatrix} = \begin{pmatrix} 1+\epsilon-\beta(m_i) & \beta(m_j) \\ \beta(m_i) & 1+\epsilon-\beta(m_j) \end{pmatrix} \begin{pmatrix} m_i(t) \\ m_j(t) \end{pmatrix}. \tag{3.1}$$

Here we consider the simplest case, i.e.:

$$\beta(m) = \begin{cases} \beta_1, & m < n\bar{m}(t) \\ \beta_2, & m \geq n\bar{m}(t) \end{cases} , \beta_1 > \beta_2 \tag{3.2}$$

If an agent has wealth higher than a threshold (n times the average wealth, $\bar{m}(t)$), the second parameter (β_2) will be used. The threshold adopted in these simulations is $10\bar{m}(t)$.

For example, to simulate a society like the U.K., where two Pareto exponents exist, one for the top earners around 3.0 and another one for the super-rich around 1.5 [92], we choose the parameters for the percentage of money exchange, β, and for the percentage of money injected in the system, ϵ, according to equation $\alpha - 1 \approx 2\beta/\epsilon^2$, i.e. $\beta_1 = 0.01$, $\beta_2 = 0.00125$ and $\epsilon = 0.1$. Figure 3.3 shows the result of our simulations.

Two distinct power laws are visible, one in the regime between $\bar{m}(t)$ and $10\bar{m}(t)$ and another one for wealth larger than $10\bar{m}(t)$. The Pareto exponents are 2.51 and 1.29, respectively, and thus differ from the prediction of equation $\alpha - 1 \approx 2\beta/\epsilon^2$ where we expected 3.0 and 1.25, respectively.

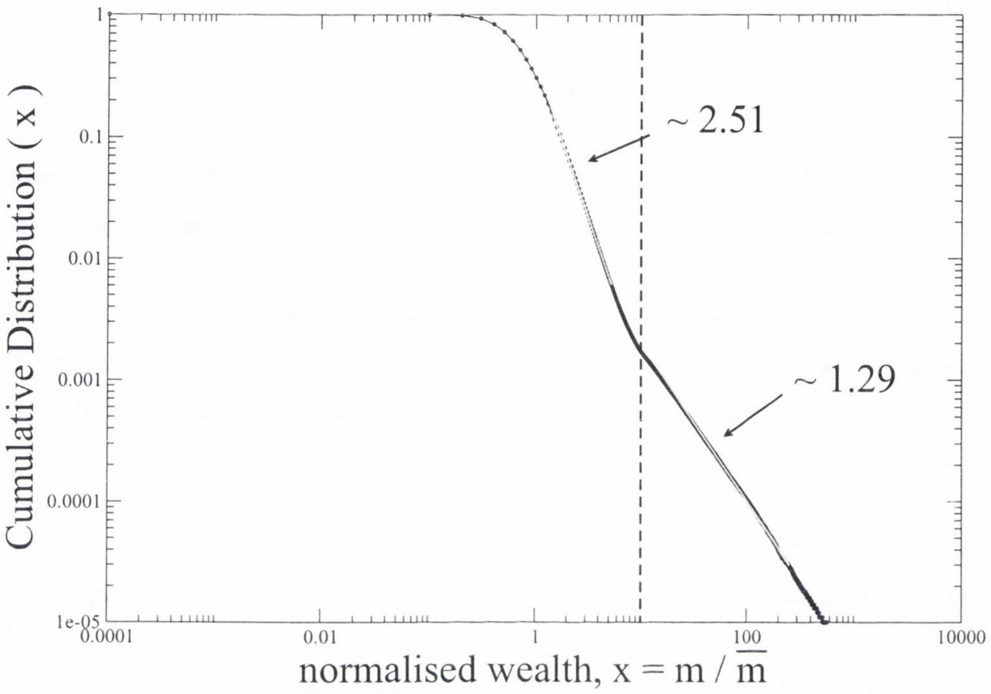


Figure 3.3: Cumulative distribution of the wealth in expanded Slanina model. The values for number of agents, time steps, realisations and percentage of wealth injected in the system (ϵ) are the same as used in the simulations shown in Figure 2.8. The percentage of wealth exchanged (β) if the agent has wealth smaller than $10\bar{m}(t)$ is 0.01 and if the agent has wealth higher or equal to $10\bar{m}(t)$ is 0.00125. Two different Pareto exponents appear in different parts of the distribution. One for what we call rich people is around $\sim$ 2.5 and a second one for the top richest is around $\sim$ 1.3. The vertical dashed line shows the threshold that we choose for different β values.

However in our case, this prediction should only be taken as a first order approximation, since we are essentially dealing with two societies (each specified by its respective β values) which are interacting. Agents switch between their interaction parameters according to their relative wealth.

3.3 Double power law from generalised functions

An analytical analysis of the double power law is necessary to explain the nature of exponents. Following Scarfone [95], in this section we introduce a theory of double power laws.

The generalised functions arise naturally from the study of non-extensive statistical mechanics [103] where the systems are described by power law features. If the distribution of a quantity x, $p(x)$ is of the exponential type:

$$\frac{dp(x)}{dx} = -\lambda_1 p(x) \tag{3.3}$$

then the solution is $p(x) = \exp(-\lambda_1 x)$. If some fractality is involved in the distribution of quantity x:

$$\frac{dp(x)}{dx} = -\lambda_q \left[p(x)\right]^q \text{ for } (\lambda_q \geq 0 \text{ and } q \geq 1) \tag{3.4}$$

then the solution is shown to be:

$$p(x) = \left[1 - (1-q)\lambda_q x\right]^{1/(1-q)} \tag{3.5}$$

which is also known as Tsallis distribution function or q-exponential function ($\exp_q(-\lambda_q x)$). This generalised exponential function recovers the normal exponential function when $q \to 1$ and behaves as a power law for $x \to +\infty$. In order to allow a second power law behaviour in this function Tsallis *et al.* [97] included another term in eq. 3.4:

$$\frac{dp(x)}{dx} = -\mu_r \left[p(x)\right]^r - (\lambda_q - \mu_r)\left[p(x)\right]^q \text{ for } (r \leq q) \tag{3.6}$$

This function with two power law exponents was the one fitted to the different data sets refered in the end of the introduction of this chapter. But this function cannot be used in the analysis of many income distributions that have an ankle ($\alpha_2 < \alpha_1$) as it is only valid when we have a knee ($\alpha_2 > \alpha_1$) [104].

Scarfone *et al.* [95, 105] obtained a class of two-parameter deformed logarithms and exponentials, and constructed a function featuring two or more power law regimes.

A deformed logarithm or (k, r)-logarithm can be represented as:

$$\log_{\{k,r\}}(x) = x^r \frac{x^k - x^{-k}}{2k} \tag{3.7}$$

which recovers the standard logarithm in the limit $(k, r) \to (0, 0)$. The inverse function of this logarithm is the deformed exponential function $\exp_{\{k,r\}}(x)$. The asymptotic behaviours of both deformed functions are:

- $\log_{\{k,r\}}(x) \to x^{r+|k|}/|2k|$ for $x \to +\infty$;
- $\log_{\{k,r\}}(x) \to -x^{r-|k|}/|2k|$ for $x \to 0$;
- $\exp_{\{k,r\}}(x) \to |2kx|^{1/(r\pm|k|)}$ for $x \to \pm\infty$;
- $\exp_{\{k,r\}}(x) \to 1 + x$ for $x \to 0$.

With the deformed logarithm and the deformed exponential functions, Scarfone [95] introduced a quantity able to reproduce a double-power law effect:

$$\Pi_{\sigma_1}(x) = \exp_{\{k_1,r_1\}}\left(a_1 \log_{\{k_1,r_1\}}(x)\right) \tag{3.8}$$

where σ_1 denotes the set of parameters (k_1, r_1, a_1). Using this quantity Scarfone [95] creates a function with the following construction:

$$f(x) = \Pi_{\sigma_1}\left(\exp_{\{k_2,r_2\}}(-x)\right) = \exp_{\{k_1,r_1\}}\left(a_1 \log_{\{k_1,r_1\}}\left(\exp_{\{k_2,r_2\}}(-x)\right)\right) \tag{3.9}$$

This function reduces to simple forms for some choice of parameters:

- if $a_1 = 1$ then $f(x) = \exp_{\{k_2,r_2\}}(-x)$;
- if $(k_1, r_1) \to (0, 0)$ then $f(x) = \left(\exp_{\{k_2,r_2\}}(-x)\right)^{a_1}$;
- if $(k_1, r_1) = (k_2, r_2)$ then $f(x) = \exp_{\{k_1,r_1\}}(-a_1 x)$.

Taking into account the asymptotic behaviour of the deformed functions, we can conclude the following about the asymptotic behaviour of $f(x)$:

- $f(x) \to 1 - a_1 x$ for $a_1 x \ll 1$;
- $f(x) \to x^{-s_1}$ for $x \ll 1 \ll a_1 x$;
- $f(x) \to x^{-s_2}$ for $a_1 x \gg 1$.

The slopes s_1 and s_2 are equal to $1/(|k_1| - r_1)$ and $1/(|k_2| - r_2)$, respectively.

So this function has the double power law behaviour in the middle region, $x \ll 1 \ll a_1 x$ and the far region, $a_1 x \gg 1$. The constant a_1 gives approximately the width of the middle region.

Scarfone [95] simplified equation 3.9, setting the values of r_1 and r_2 equal to zero. In this case, the generalised logarithm is given by:

$$\log_{\{k\}}(x) = \frac{x^k - x^{-k}}{2k} \tag{3.10}$$

and the generalised exponential:

$$\exp_{\{k\}}(x) = \left(kx + \sqrt{1 + k^2 x^2}\right)^{1/k} \tag{3.11}$$

so, the function $f(x)$ is given by:

$$\begin{aligned} f(x) &= \exp_{\{k_1\}}\left(a_1 \log_{\{k_1\}}\left(\exp_{\{k_2\}}(-x)\right)\right) = \\ &= \left[\frac{a_1}{2}\left(\sqrt{1 + k_2^2 x^2} - k_2 x\right)^{k_1/k_2} - \frac{a_1}{2}\left(\sqrt{1 + k_2^2 x^2} - k_2 x\right)^{-k_1/k_2} + \right. \\ &+ \left. \sqrt{1 + \frac{a_1^2}{4}\left[\left(\sqrt{1 + k_2^2 x^2} - k_2 x\right)^{k_1/k_2} - \left(\sqrt{1 + k_2^2 x^2} - k_2 x\right)^{-k_1/k_2}\right]^2}\ \right]^{1/k_1} \end{aligned} \tag{3.12}$$

Scarfone also showed [95] some fits of this function to empirical data of wealth distributions. We use this function to fit the results of the simulations of the expanded Slanina model from section 3.2.

Using the same results of Figure 3.3 we performed a fit of this distribution using a function of the form of equation 3.9 presented by Scarfone [95]:

$$P(x) = f(-\beta|x|^{\mu}) \tag{3.13}$$

where β and μ are fitting parameters. Using the fitting tool from GNUPLOT software [106] we plot the results in Figure 3.4. We find that the cumulative distribution function is indeed well described by the proposed functional form of Scarfone.

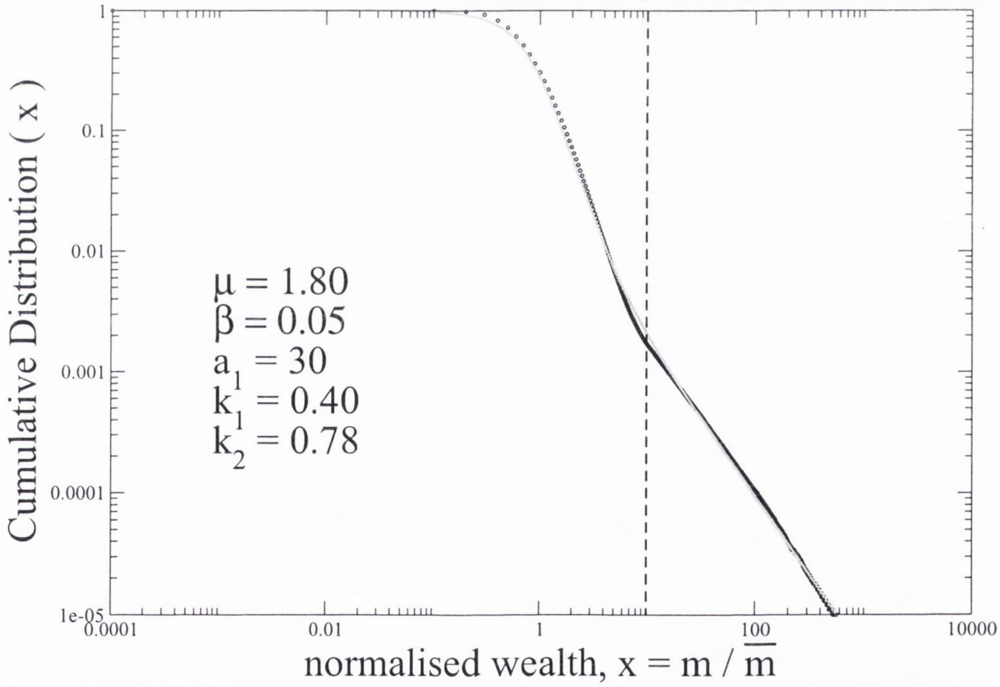


Figure 3.4: The same cumulative distribution of wealth as in Figure 3.3, but with a fit to the double power law equation 3.9. The values of the fitting parameters are shown inside the figure.

Future work should thus include the derivation of an analytical solution for our modified Slanina model. It could profit from a trial solution as given by equation 3.9.

3.4 Conclusions

The main success of the modified Slanina model is thus the reproduction of two power laws regimes. We have sought to modify the Lotka Volterra approach in an attempt to model this double power law however thus far our efforts have been unsuccessful.

As was discussed in [18], progress in understanding the details of wealth distribution is invariably linked to obtaining data sets that encompass the entire population of a country. It appears that at present, this information is only available for a few countries, for example Japan (Souma [23]). Generally, the super-rich are not included in income data. Published wealth lists are estimates, but for the moment might well remain the only public source for the information on these top earners. We hope that analyses of the kind we have made in this section encourage the release of more detailed income data over the entire income range. Only with more complete datasets we would be able to properly understand these complex economic systems.

Chapter 4

Correlations of Financial data

4.1 Introduction

In this chapter we introduce the toolkits recently developed by statistical physicists, for the analysis of financial data, including correlations. We will make use of these tools in the proceeding chapters. To illustrate the theoretical concepts we have analysed daily data from the Thomson Datastream [107]. We created a MySQL database [108] where we upload it and we wrote a C program to analyse different characteristics of the data.

An example of these characteristics is the nature of the returns of the price of a stock market. For many years, economists have treated the price changes of a stock as a normal distribution, which was the original proposal of Bachelier [5], but around the 1950s this theory was replaced by a model in which the stock prices perform a geometric Brownian motion, i.e. the stock prices are log-normal distributed [2]. But for a geometric Brownian motion the differences of the logarithms of the prices, which we call logarithm return, are Gaussian distributed. And this distribution is not able to explain the whole range of the distribution of returns, for example, for the higher values of returns the distribution has fatter tails than the Gaussian distribution, as we show in the next section when we try to fit different probability distribution functions to the distribution of logarithm returns of a stock.

The study of correlations between stocks is very important for portfolio analysis. Even for the common sense is known that we cannot put all the eggs in one basket, so if we are investing in the stock market, we need to have a diversification of our portfolio to avoid higher risk of losses. The correlations between the stocks included in our portfolio will give the weight, according to Markovitz theory [130], of each stock for a portfolio optimisation. Another important topic related with the correlation between stocks is the industrial classification of sectors in a market, and the update of these lists of industrial classification in order to help the investors.

In section 4.2 we introduce the definition of returns of the price of a stock and some probability distribution functions that are used to fit these returns. After this, we explain the importance of the correlations in financial data (section 4.3) and introduce different mechanisms to study these correlations (sections 4.4 and 4.5).

4.2 Analysing returns

Calling the value of a time series i at time t as $P_i(t)$, where for a financial time series this value is the price of a stock i at time t, the log-return of a time series is defined as the difference between the logarithms of two consecutive values:

$$R_i(t) = \ln P_i(t) - \ln P_i(t-1). \tag{4.1}$$

The definition of logarithmic return has two purposes: firstly there is a common belief that the price of stocks, $P_i(t)$ increases exponentially in time on average; secondly the difference between two consecutive values is very small, so:

$$R_i(t) = \ln\left[\frac{P_i(t)}{P_i(t-1)}\right] = \ln\left[1 + \frac{P_i(t) - P_i(t-1)}{P_i(t-1)}\right] \approx \frac{P_i(t) - P_i(t-1)}{P_i(t-1)} \tag{4.2}$$

i.e. the log-return has approximately the same value as the quotient return [2].

Figure 4.1 shows the evolution of the value of a time series, $P_i(t)$ in time. In this example the value of the time series is the daily closing price of the stock of The Hongkong and Shanghai Banking Corporation Limited (HSBC), in pounds. This company belongs to the main index in the London Stock Exchange market, the FTSE100. The closing price is shown for the period July 9^{th} 1992 until March 24^{th} 2008 (4098 market days). Figure 4.2 shows the evolution of the logarithmic returns, $R_i(t)$ of the stock HSBC, in time.

4.2.1 Gaussian distribution

The distribution of the logarithmic returns of the price of a stock generally does not follow a Gaussian distribution [98, 109, 110], as assumed by some economists, but as we can see from Figure 4.3 the tails of the distribution are more enhanced than a Gaussian distribution. The strength of the tails depends on the time scale at which returns are evaluated and scales with time [2].

A Gaussian distribution is a distribution with two parameters, the mean (μ) and the standard deviation (σ). To fit this function to any distribution we will need to calculate these two parameters of the distribution and plot the probability distribution function:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{4.3}$$

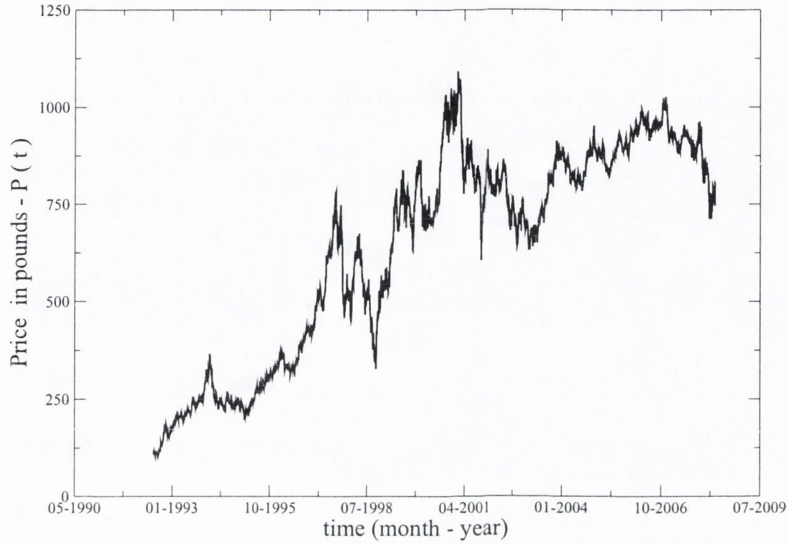


Figure 4.1: Price of the company HSBC (Bank with the tick symbol HSBA) in time from 1992-07-09 to 2008-03-24 (4098 market days).

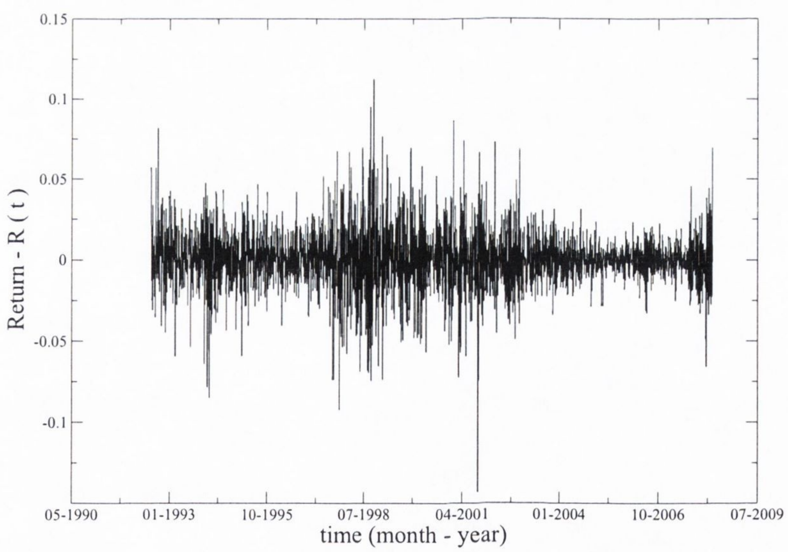


Figure 4.2: Daily return of price of company HSBC (Bank with the tick symbol HSBA) in time from 1992-07-10 to 2008-03-24 (4097 market returns).

4.2.2 T-student distribution

More appropriate distributions that will feature the fat tails are for example the T-student or Tsallis distributions [111]. The probability distribution function of a T-student distribution is

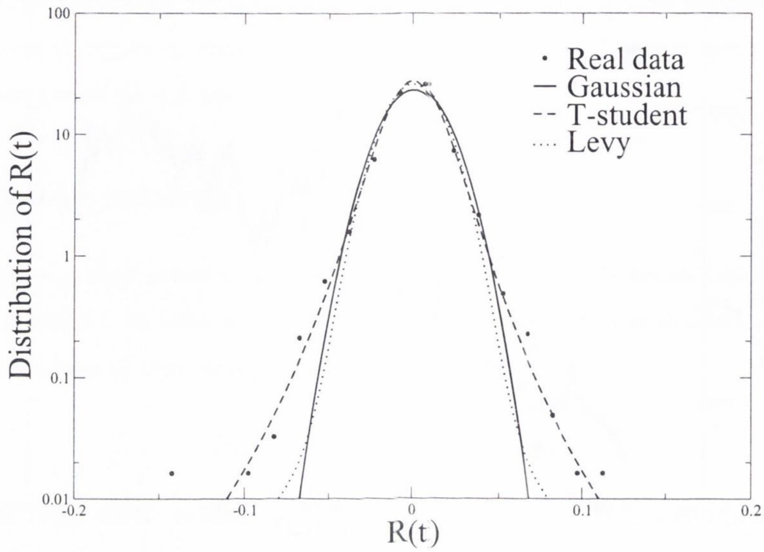


Figure 4.3: Distribution of returns of price of company HSBC (Bank with the tick symbol HSBA) represented as black circles. The solid, broken and dotted lines represent fits to Gaussian, T-student and Levy distributions, respectively. Note the log-normal scale of the plot.

given by:

$$P_k(x) = N_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e_k^{-x^2/2\sigma_k^2} \tag{4.4}$$

where N_k is a normalisation factor:

$$N_k = \frac{\Gamma(k)}{\sqrt{k}\Gamma\left(k - \frac{1}{2}\right)} \tag{4.5}$$

and $\Gamma(k)$ is the Gamma function. The factor $\sigma_k = \sigma\sqrt{(k - 3/2)/k}$ is related to the standard deviation of the distribution (σ) and with the degree of distribution (k). The function e_k^z is an approximation of the exponential function called *k-exponential*,

$$\exp_k(z) = \left(1 - \frac{z}{k}\right)^{-k} \tag{4.6}$$

which in the limit $k \to \infty$ reduces to the ordinary exponential function. The probability distribution function may be written as:

$$P_k(x) = \frac{\Gamma(k)}{\Gamma\left(k - \frac{1}{2}\right)} \frac{1}{\sigma\sqrt{\pi(2k-3)}} \left[1 + \frac{x^2}{\sigma^2(2k-3)}\right]^{-k} \tag{4.7}$$

The parameter k is related to the Tsallis parameter q by $k = 1/(q - 1)$.

To compute the parameters of a T-student distribution it is necessary to take into account the fact that some moments of the distribution might not exist, i.e. diverge. Using fractional moments we can avoid these kind of problems. Considering the fractional moment for the empirical distribution, m_f of returns,

$$m_f = \frac{1}{T-1} \sum_{t=t_0+1}^{t_0+T-1} |R(t)|^f \tag{4.8}$$

where f is a fractional number, less than the moment for which the sum diverges and $R(t)$ is the return. The rate of the moments

$$r_f = \frac{m_{f-1}}{m_{f+1}} \tag{4.9}$$

can be computed and compared with the rate of moments for the analytical distribution of returns,

$$R_f = \frac{M_{f-1}}{M_{f+1}} \tag{4.10}$$

which is given by (see Appendix A for the derivation):

$$R_f = \frac{1}{(2k-3)\sigma^2}\left[\frac{2(k-1)}{f}\right] - \frac{1}{(2k-3)\sigma^2} \tag{4.11}$$

So, it can be seen that the rate of the moments, R_f is inversely proportional to the fractional number, f. Calculating the rate of the moments for the empirical distribution for different fractional exponents, f, a plot of r_f versus $1/f$ can be used to calculate the parameters, σ and k of the T-student distribution (Fig. 4.4):

$$r_f = a\frac{1}{f} + b \tag{4.12}$$

Using different values of the fractional number, f: $\frac{2}{3} < f < 1$, the values of the parameters σ and k are calculated by linear regression:

$$\sigma^2 = \frac{1}{a+b} \tag{4.13}$$

$$k = \frac{2b-a}{2b} \tag{4.14}$$

For all the stocks of the London Stock Exchange that we studied, the minimum value of k is 1.7 and the maximum 9.0, but most of the values are between 2.0 and 4.0, which means values of q in the interval between 1.25 and 1.5, which is around the values found by Tsallis [112] (1.40, 1.37 and 1.38) for one, two and three minutes return, respectively, for the NYSE in 2001. For example the value of k found for the HSBC data of Figure 4.3 is approximately ~ 2.90.

4.2.3 Levy distribution

Another type of distributions with fat tails that is very popular amongst the physics community is the Levy stable distribution. Like the Gaussian distribution it is stable, i.e. when we sum

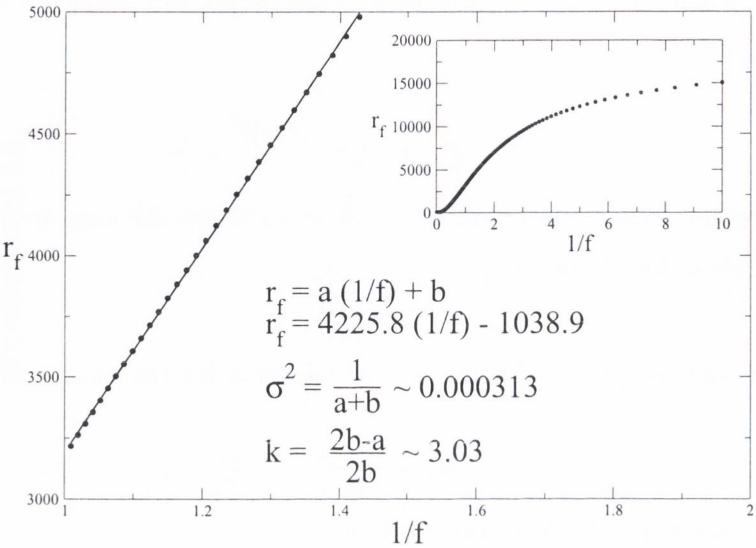


Figure 4.4: Ratio of fractional moments (r_f) versus the inverse of the exponent f for the company HSBC (Bank with the tick symbol HSBA) represented as black circles. The inset figure shows the ratio for a large range of exponents f and it can be seen that the linear part is around $1/f \sim 0.5 - 2.0$. In the main figure, we used values of f between 0.7 and 1.0. Calculating the linear regression in this region we get values a and b for the line equation, $r_f = a(1/f) + b$. With these values we computed the parameters of the T-student distribution, $\sigma \sim 0.0177$ and $k \sim 3.03$.

variables from two independent stable distributions with the same exponent α, the resulting variables will also be distributed according to a stable distribution with the same exponent [6]. If u' and u'' are distributed according to the same stable distribution and u is a linear combination of u' and u'':

$$u = c_1 u' + c_2 u'' + c_3 \tag{4.15}$$

there exist values of c_4 and c_5 such that $c_4 u + c_5$ is also distributed according to the same stable distribution.

There is no analytical expression for the probability density function of the Levy distribution. However, the Fourier transform of its characteristic function [2] is given by:

$$L(x;\alpha,\beta,\gamma,\mu) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \varphi(q)\exp(-qx)dq \tag{4.16}$$

and the general expression for the characteristic functions is:

$$\varphi(q) = \begin{cases} \exp\left\{\imath\mu q - \gamma|q|^{\alpha}\left[1 - \imath\beta\frac{q}{|q|}\tan\left(\frac{\pi}{2}\alpha\right)\right]\right\} & [\alpha \neq 1] \\ \exp\left\{\imath\mu q - \gamma|q|\left[1 + \imath\beta\frac{q}{|q|}\ln|q|\right]\right\} & [\alpha = 1] \end{cases} \tag{4.17}$$

where α is the exponent of the distribution, γ is a normalisation factor, μ is related to the mean of the distribution and β is related to its asymmetry. For the case $\alpha = 2$ this distribution reduces to the Gaussian distribution.

Assuming a symmetric distribution ($\beta = 0$) centred at the origin ($\mu = 0$) we are left with only two parameters, α and γ and then the characteristic function is given by:

$$\varphi(q) = \exp\left(-\gamma|q|^{\alpha}\right) \tag{4.18}$$

As happen for the T-student distribution, some moments of the Levy distribution do not exist, for example the second-order moment diverges for $0 < \alpha < 2$, but all the moments of order less than α do exist and are called the fractional lower-order moments (FLOM) [113]. The FLOM of a Levy random variable can be found from the parameters α and γ [113]:

$$E\left(|X|^{p}\right) = C(p,\alpha)\gamma^{\frac{p}{\alpha}} \text{ for } 0 < p < \alpha \tag{4.19}$$

where

$$C(p,\alpha) = \frac{2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(-\frac{p}{\alpha}\right)}{\alpha\sqrt{\pi}\Gamma\left(-\frac{p}{2}\right)} \tag{4.20}$$

So, computing the FLOM of the distribution of returns give us the values for the parameters α and β for each time series. For a practical computation of these parameters we proceed as follows.

The log-price can be assumed to be composed of two terms, a drift term D and a fluctuation term that follows a Levy distribution [114],

$$\ln P_i(t) = Dt + \gamma L(t) \tag{4.21}$$

where γ is the dispersion of the fluctuations and $L(t)$ is a random variable from a stable distribution with dispersion t and exponent α. In Fig. 4.5 we represent the log-price of the stock HSBC in time with the correspondent drift calculated from a linear regression. The drift adjusted prices, $\tilde{P}_i(t)$ are calculated as:

$$\ln \tilde{P}_i(t) = \ln P_i(t) - Dt - \ln P_i(0) = \gamma L(t) - \ln P_i(0), \tag{4.22}$$

where $P_i(0)$ is the price at $t = 0$.

Using the fact that a Levy process is homogeneous in time [114] the log-return of the drift adjusted price can be represented by stable variables as:

$$\ln \frac{\tilde{P}_i(t+\delta t)}{\tilde{P}_i(t)} = \ln \tilde{P}_i(t+\delta t) - \ln \tilde{P}_i(t) = \gamma\left[L(t+\delta t) - L(t)\right] = \gamma L(\delta t) \tag{4.23}$$

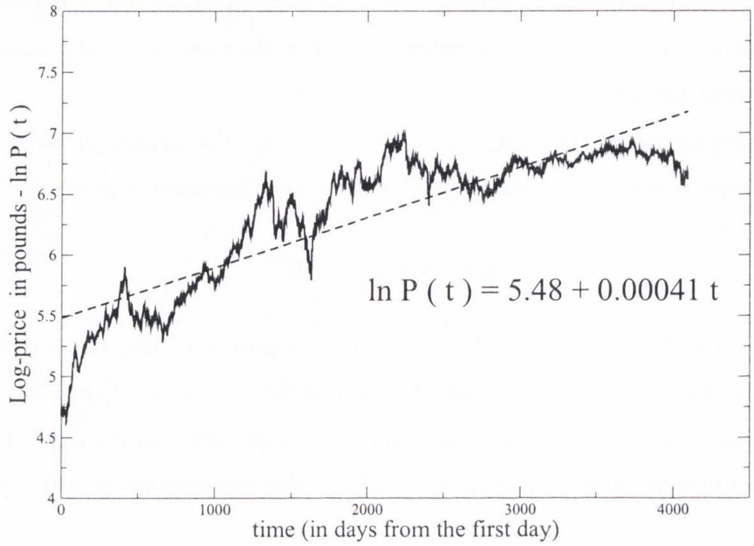


Figure 4.5: Log-price of the company HSBC (Bank with the tick symbol HSBA) in time from 1992-07-09 to 2008-03-24 (4098 market days) in solid line and the respective drift, calculated from the linear regression, in broken line.

Calculating the FLOM for the log-return of the drift adjusted price for a constant value of $p < \alpha$ and for variable values of δt we can estimate the moments of the distribution:

$$m_p(\delta t) = \frac{1}{T}\sum_{t=1}^{T}\left|\ln\frac{\tilde{P}_i(t+\delta t)}{\tilde{P}_i(t)}\right|^p = E\left(|\gamma L(\delta t)|^p\right) \tag{4.24}$$

where $E(...)$ is the expected value of the fractional moment and from equation 4.19 is given by [113]:

$$E\left(|\gamma L(\delta t)|^p\right) = \gamma^p E\left(|L(\delta t)|^p\right) = \gamma^p C(p,\alpha)\delta t^{\frac{p}{\alpha}} \tag{4.25}$$

From this equation we can see that the logarithm of $E(...)$ is proportional to the logarithm of δt:

$$\ln\left[E\left(|\gamma L(\delta t)|^p\right)\right] = p\ln\gamma + \ln C + \frac{p}{\alpha}\ln\delta t \tag{4.26}$$

and so if we plot the moments of the returns, $m_p(\delta t)$ versus δt we obtain an estimation of the values of α and γ (Figure 4.6). The linear equation for the moments is given by:

$$\ln\left[m_p(\delta t)\right] = a\ln\delta t + b \tag{4.27}$$

where a is the slope of the curve and b the interception with the y-axis.

$$\alpha = \frac{p}{a} \tag{4.28}$$

$$\gamma = C^{-\frac{1}{p}}\exp\left[\frac{b}{p}\right] \tag{4.29}$$

In Fig. 4.6, we represent the moments, $m_p(\delta t)$ versus δt and the linear regression that give the estimated values of α and γ parameters of the Levy distribution for the returns of HSBC. For a value of p equal to 1.1 the value estimated for α is around 1.97 and for γ it is around 0.01. If we increase the value of p the values estimated for α and γ also increase, sometimes even reaching values of α that are not expected for a Levy stable distribution. This value of α is in agreement

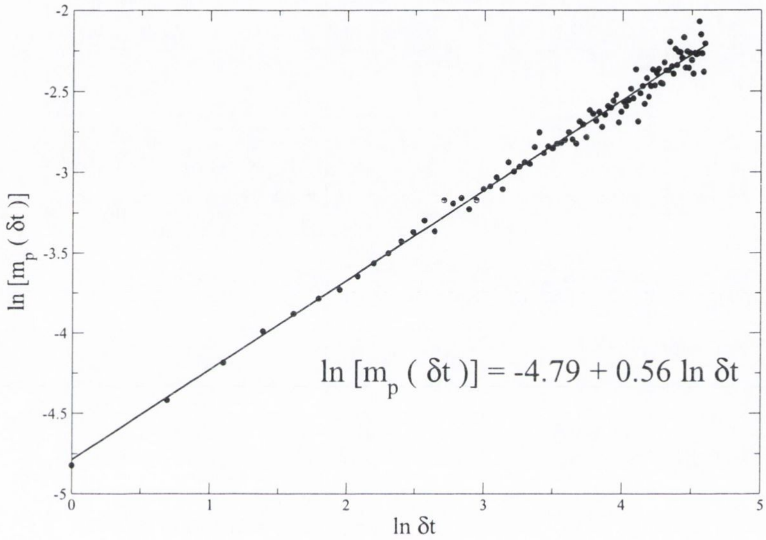


Figure 4.6: Fractional moments of the returns of company HSBC (Bank with the tick symbol HSBA) versus δt represented as black circles and the respective linear regression represented as a solid line.

with the values expected for a Levy stable distribution ($\alpha < 2.0$). The values of α and γ were used to produce the Levy fit to the distribution of returns of HSBC in Fig. 4.3. For the returns of HSBC the fit of a Levy distribution appears very similar to the fit of a Gaussian distribution, probably because the α parameter is almost equal to 2.0, so we can conclude that the best fit for this particular set of data is the one performed with a T-student distribution. For other stocks this is not the case, performing the same method to estimate the values of the parameters for other stocks we can see that some values of α are outside the permitted region of $\alpha < 2$ and the smallest value of α is equal to 1.74. In Table 4.1 we show all the values used to fit the three probability distribution functions presented before.

Table 4.1: Name, symbol and values used to fit the three probability distribution functions to the distribution of logarithm returns: μ_G and σ_G for Gaussian; k_T and σ_T for T-Student; α_L and γ_L for Levy. The stocks presented here belong to a portfolio of 85 stocks from the FTSE100 index with time series length of 2146 days.

Name	Symbol	μ_G	σ_G	k_T	σ_T	α_L	γ_L
3I GROUP	III.L	0.000	0.02	2.90	0.02	2.08	0.01
ALLIANCE & LEICESTER	AL.L	0.000	0.02	2.07	0.02	2.20	0.01
ALLIANCE TRUST	ATST.L	0.000	0.01	3.22	0.01	2.12	0.01
AMEC	AMEC.L	0.000	0.02	3.00	0.02	1.74	0.01
ANGLO AMERICAN	AAL.L	0.000	0.02	5.45	0.02	2.27	0.02
ANTOFAGASTA	ANTO.L	0.001	0.02	3.18	0.02	2.14	0.02
ASSOCIATED BRIT.FOODS	ABF.L	0.000	0.02	2.71	0.02	2.21	0.01
ASTRAZENECA	AZN.L	0.000	0.02	3.08	0.02	2.11	0.01
AVIVA	AV.L	0.000	0.02	2.97	0.02	2.37	0.02
BAE SYSTEMS	BA.L	0.000	0.02	2.64	0.02	1.90	0.01
BARCLAYS	BARC.L	0.000	0.02	3.30	0.02	2.42	0.02
BG GROUP	BG.L	0.001	0.02	6.38	0.02	2.51	0.02
BHP BILLITON	BLT.L	0.001	0.02	5.92	0.02	2.25	0.02
BP	BP.L	0.000	0.02	7.21	0.02	2.66	0.02
BRITISH AIRWAYS	BAY.L	0.000	0.03	3.59	0.03	1.86	0.02
BRITISH AMERICAN TOBACCO	BATS.L	0.001	0.02	2.67	0.02	2.86	0.02
BRITISH LAND	BLND.L	0.000	0.02	3.80	0.02	2.05	0.01
BRITISH SKY BCAST.GROUP	BSY.L	0.000	0.02	2.39	0.02	2.19	0.02
BT GROUP	BT-A.L	-0.001	0.02	2.97	0.02	2.24	0.02
BUNZL	BNZL.L	0.000	0.01	3.92	0.01	2.09	0.01
CABLE & WIRELESS	CW.L	-0.001	0.03	2.10	0.03	1.75	0.01
CADBURY	CBRY.L	0.000	0.02	3.78	0.02	2.19	0.01
CAIRN ENERGY	CNE.L	0.001	0.02	2.28	0.03	1.90	0.01
CAPITA GROUP	CPI.L	0.000	0.02	2.50	0.03	2.42	0.02
CENTRICA	CNA.L	0.000	0.02	4.35	0.02	2.36	0.01
COBHAM	COB.L	0.000	0.01	3.68	0.02	2.57	0.02
DIAGEO	DGE.L	0.000	0.02	2.76	0.02	2.49	0.01
ENTERPRISE INNS	ETI.L	0.001	0.02	3.15	0.02	1.83	0.01
FIRST GROUP	FGP.L	0.000	0.02	2.45	0.02	1.94	0.01
G4S	GFS.L	0.000	0.02	3.11	0.02	2.15	0.02

Continue on next page

– continue from the previous page

Name	Symbol	μ_G	σ_G	k_T	σ_T	α_L	γ_L
GLAXOSMITHKLINE	GSK.L	0.000	0.02	5.10	0.02	2.63	0.02
HAMMERSON	HMSO.L	0.000	0.02	3.55	0.02	2.13	0.01
HBOS	HBOS.L	0.000	0.02	2.69	0.02	2.46	0.02
HOME RETAIL GROUP	HOME.L	0.000	0.02	3.02	0.02	2.20	0.01
HSBC HDG. (ORD $0.50)	HSBA.L	0.000	0.01	2.74	0.02	2.16	0.01
ICAP	IAP.L	0.001	0.02	2.75	0.02	1.98	0.01
IMPERIAL TOBACCO GP.	IMT.L	0.001	0.02	3.14	0.02	2.66	0.02
INTERNATIONAL POWER	IPR.L	0.000	0.02	4.18	0.02	1.79	0.01
ITV	ITV.L	-0.001	0.03	5.47	0.03	2.15	0.02
JOHNSON MATTHEY	JMAT.L	0.000	0.02	3.37	0.02	2.34	0.01
KINGFISHER	KGF.L	-0.001	0.02	3.08	0.02	2.15	0.02
LAND SECURITIES GROUP	LAND.L	0.000	0.01	4.21	0.01	2.05	0.01
LEGAL & GENERAL	LGEN.L	0.000	0.02	15.77	0.02	2.31	0.02
LIBERTY INTL.	LII.L	0.000	0.01	2.84	0.01	2.08	0.01
LLOYDS TSB GROUP	LLOY.L	0.000	0.02	2.77	0.02	2.55	0.02
LONMIN	LMI.L	0.001	0.02	3.49	0.02	2.05	0.02
MAN GROUP	EMG.L	0.001	0.02	3.65	0.02	2.04	0.01
MARKS & SPENCER GROUP	MKS.L	0.000	0.02	2.63	0.02	1.90	0.01
MORRISON(WM)SPMKTS.	MRW.L	0.000	0.02	3.71	0.02	2.15	0.01
NATIONAL GRID	NG.L	0.000	0.01	3.93	0.01	2.32	0.01
NEXT	NXT.L	0.000	0.02	2.67	0.02	1.99	0.01
OLD MUTUAL	OML.L	0.000	0.02	4.66	0.02	2.14	0.02
PEARSON	PSON.L	0.000	0.02	2.46	0.02	2.35	0.02
PERSIMMON	PSN.L	0.000	0.02	4.49	0.02	2.01	0.01
PRUDENTIAL	PRU.L	0.000	0.02	2.96	0.02	2.35	0.02
RECKITT BENCKISER	RB.L	0.001	0.02	2.79	0.02	2.85	0.02
REED ELSEVIER	REL.L	0.000	0.02	2.88	0.02	2.83	0.02
REXAM	REX.L	0.000	0.02	2.74	0.02	2.35	0.01
RIO TINTO	RIO.L	0.001	0.02	5.43	0.02	2.29	0.02
ROLLS-ROYCE GROUP	RR.L	0.000	0.02	2.99	0.02	1.89	0.01
ROYAL & SUN ALL.IN.	RSA.L	-0.001	0.03	2.96	0.03	1.81	0.01
ROYAL BANK OF SCTL.GP.	RBS.L	0.000	0.02	2.61	0.02	2.28	0.02
ROYAL DUTCH SHELL B	RDSB.L	0.000	0.02	4.37	0.02	2.70	0.02
SABMILLER	SAB.L	0.000	0.02	3.57	0.02	2.07	0.01
SAGE GROUP	SGE.L	-0.001	0.03	2.72	0.03	2.23	0.02
SAINSBURY (J)	SBRY.L	0.000	0.02	2.48	0.02	2.09	0.01
SCHRODERS	SDR.L	0.000	0.03	3.06	0.03	1.89	0.01
SCHRODERS NV	SDRC.L	0.000	0.03	2.69	0.03	1.99	0.01

Continue on next page

– concluded from previous page

Name	Symbol	μ_G	σ_G	k_T	σ_T	α_L	γ_L
SCOT.& SOUTHERN ENERGY	SSE.L	0.000	0.01	4.48	0.01	2.84	0.01
SEVERN TRENT	SVT.L	0.000	0.01	3.50	0.02	2.80	0.02
SHIRE	SHP.L	0.000	0.03	2.16	0.03	2.00	0.02
SMITH & NEPHEW	SN.L	0.001	0.02	3.34	0.02	2.15	0.01
SMITHS GROUP	SMIN.L	0.000	0.02	2.87	0.02	2.11	0.01
STANDARD CHARTERED	STAN.L	0.000	0.02	3.38	0.02	2.59	0.02
TATE & LYLE	TATE.L	0.000	0.02	2.22	0.02	1.75	0.01
TESCO	TSCO.L	0.000	0.02	6.51	0.02	2.25	0.01
THOMSON REUTERS	TRIL.L	0.000	0.03	2.44	0.03	1.83	0.01
TUI TRAVEL	TT.L	0.000	0.02	3.43	0.02	2.09	0.02
TULLOW OIL	TLW.L	0.001	0.02	4.56	0.03	2.09	0.02
UNILEVER (UK)	ULVR.L	0.000	0.02	3.06	0.02	2.52	0.02
UNITED UTILITIES	UU.L	0.000	0.01	3.55	0.01	2.79	0.01
VODAFONE GROUP	VOD.L	0.000	0.02	4.43	0.02	2.16	0.02
WHITBREAD	WTB.L	0.000	0.02	3.23	0.02	1.95	0.01
WOLSELEY	WOS.L	0.000	0.02	3.27	0.02	1.94	0.01
WPP GROUP	WPP.L	0.000	0.02	3.22	0.02	2.35	0.02

In Figure 4.7 we plot the distribution of the values of α parameter from the Lévy distribution. We can see that most of the values are higher than 2.0 which contradicts the Lévy theory. This fact helps us in our conclusion that the T-student distribution is a better fit for this portfolio of stocks.

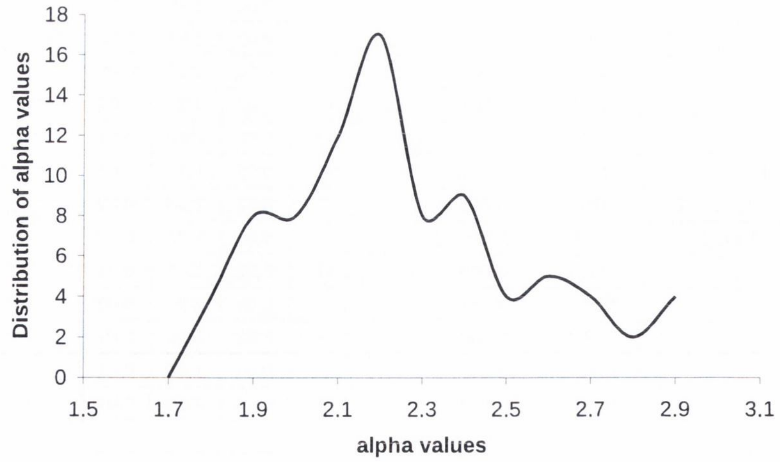


Figure 4.7: Distribution of α values from the Lévy distribution for a portfolio of 85 different stocks of the FTSE100. This values were taken from table 4.1. Most of the values are higher than 2.0 which contradicts the Lévy theory.

4.3 The correlation of time series

During the past decade, many physicists have used techniques of statistical physics and complexity to study financial problems. In particular the concept of networks proved valuable, whereby the networks are set up to represent correlations between stocks. Studying the characteristics of these networks can prove very valuable for portfolio optimisation [48, 115].

A challenging problem is the nature of a stock time series and, in particular, the nature of their randomness [45, 47, 48]. Recently the theory of random matrices has proved helpful to characterise the time series [38, 39]. In this section we introduce the concepts of a minimal spanning tree (MST) proposed by Mantegna [43] and some issues of random matrix theory (RMT) studied by Mehta [37] to examine the correlations between time series.

Important information about financial data is obtained by studying the eigensystem of the correlation matrix. In particular the spectrum of eigenvalues differs markedly from the one for random matrices [41, 42].

To analyse the correlations between time series, we computed the correlation coefficient, ρ_{ij} for the time series of log-returns R_i and R_j (see equation 4.1):

$$\rho_{ij} = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{\left(\langle R_i^2 \rangle - \langle R_i \rangle^2\right)\left(\langle R_j^2 \rangle - \langle R_j \rangle^2\right)}} \tag{4.30}$$

where:

$$\begin{aligned} \langle R_i R_j \rangle &= \frac{1}{T-1} \sum_{t=t_0+1}^{t_0+T-1} R_i(t) R_j(t) \\ \langle R_i \rangle &= \frac{1}{T-1} \sum_{t=t_0+1}^{t_0+T-1} R_i(t) \end{aligned} \tag{4.31}$$

and t_0 and T are the first time and length of the time series, respectively. Since the first value of the time series of log-returns is $R_i(t_0+1)$, which means that has one less value than the time series of values, the sum is divided by $T-1$. Each time series of log-returns can be normalised by subtracting the mean, $\langle R_i \rangle$ and dividing by the standard deviation, $\sqrt{\langle R_i^2 \rangle - \langle R_i \rangle^2}$:

$$\tilde{R}_i(t) = \frac{R_i(t) - \langle R_i \rangle}{\sqrt{\langle R_i^2 \rangle - \langle R_i \rangle^2}} \tag{4.32}$$

for every time $t = t_0+1, \ldots, t_0+T-1$. The correlation coefficient can then be written as:

$$\rho_{ij} = \langle \tilde{R}_i \tilde{R}_j \rangle = \frac{1}{T-1} \sum_{t=t_0+1}^{t_0+T-1} \tilde{R}_i(t) \tilde{R}_j(t) \tag{4.33}$$

This coefficient can vary between $-1 \leq \rho_{ij} \leq 1$, where -1 corresponds to a completely anti-correlated time series and $+1$ to a completely correlated time series. If $\rho_{ij} = 0$, the time series i

and j are uncorrelated. The coefficients for all the pairs of time series form a symmetric matrix with diagonal elements equal to unity. The correlation matrix with elements ρ_{ij} can then be represented in matrix form as:

$$\mathsf{C} = \frac{1}{T-1}\mathsf{G}\mathsf{G}^{Tr} \tag{4.34}$$

where G represents the matrix with elements $\tilde{R}_i(t)$. The size of this matrix depends on the number, N and length, T of the time series. So, it will be an $N \times (T-1)$ matrix. The matrix G^{Tr} denotes the transpose of G.

The distribution of correlation coefficients is an important aspect of our study because one can show how the changes in time of the moments of this distribution are related with each other. Following Onnela *et al.* [48, 53], we analysed the moments of the distribution of correlations coefficients in time. The first moment is the mean correlation:

$$\overline{\rho} = \frac{2}{N(N-1)}\sum_{i<j}\rho_{ij} \tag{4.35}$$

Other moments are similarly defined, the variance:

$$\lambda_2 = \frac{2}{N(N-1)}\sum_{i<j}(\rho_{ij}-\overline{\rho})^2, \tag{4.36}$$

the skewness:

$$\lambda_3 = \frac{2}{N(N-1)\lambda_2^{3/2}}\sum_{i<j}(\rho_{ij}-\overline{\rho})^3, \tag{4.37}$$

and the kurtosis:

$$\lambda_4 = \frac{2}{N(N-1)\lambda_2^{2}}\sum_{i<j}(\rho_{ij}-\overline{\rho})^4. \tag{4.38}$$

Just the elements of the upper triangle of the matrix are used to compute the matrix, because it is a symmetric matrix with diagonal elements equal to unity. If we divide our time series in small windows and we move these windows in small steps, we create different correlation matrices. If we compute the moments of each of these matrices, we can study these moments in time. Evaluation of these moments for time windows of width T reveals the dynamics of the time series. The higher moments explain how the variance of correlation coefficients increase or decrease and how the skewness and kurtosis of the distribution changes.

The decomposition of the matrix C in terms of eigenvalues and eigenvectors can be represented as:

$$\mathsf{C} = \mathsf{U}\mathsf{D}\mathsf{U}^{-1} \tag{4.39}$$

where D is a diagonal matrix with the eigenvalues, λ_i as elements:

$$\mathsf{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{bmatrix} \tag{4.40}$$

The matrix U is a matrix of the eigenvectors, u^k:

$$\mathsf{U} = \begin{bmatrix} u_1^1 & u_1^2 & \cdots & u_1^N \\ u_2^1 & u_2^2 & \cdots & u_2^N \\ \vdots & \vdots & \ddots & \vdots \\ u_N^1 & u_N^2 & \cdots & u_N^N \end{bmatrix} \tag{4.41}$$

where u_i^k is the i^{th} element of the eigenvector k. The matrix U^{-1} is the inverse of U.

4.4 Random Matrix Theory

If we study the eigensystem of a matrix we can search some conclusions about the origins of that matrix. In our case, if we analyse the correlation matrix we can obtain further information in the time series from which it originated. Is known that the distribution of eigenvalues of a random matrix is well characterised [116]. If the random matrix is defined as:

$$\mathsf{C}' = \frac{1}{T'}\mathsf{G}'\mathsf{G}'^{Tr} \tag{4.42}$$

where G' represents a $N \times T'$ matrix with independent and identically distributed elements, the distribution of eigenvalues can be calculated analytically. In the limit $N \to \infty$ and $T' \to \infty$, where $Q = T'/N$ is fixed and bigger than 1, the probability density function of the N eigenvalues, λ of the random matrix is:

$$P_{RM}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} \tag{4.43}$$

where

$$\lambda_{min}^{max} = \left(1 \pm \frac{1}{\sqrt{Q}}\right)^2 \tag{4.44}$$

limits the interval where the probability density function is different from zero.

We define the spectrum of eigenvalues as all the values of eigenvalues from a matrix and the distribution of eigenvalues as the distribution of these spectrum. The spectrum of eigenvalues, will be confine to same limits.

One characteristic eigenvalue is the one with the highest value. For the study of correlations between stock prices, the eigenvector related with the highest eigenvalue has all his elements with the same sign and is interpreted as the influence of the entire market that is common for all stocks [41]. This value depends on the portfolio studied, this can be related with the size of the portfolio and also with the mean correlation. More correlated portfolios will have a higher value for the highest eigenvalue.

4.4.1 Inverse Participation Ratio

The information contained in the eigenvalues is retrieved by looking at the corresponding eigenvectors. A quantitative factor calculated from the eigenvectors is the Inverse Participation Ratio (IPR) [42]. The IPR of an eigenvector u^k is given by:

$$I^k = \sum_{i=1}^{N} \left[u_i^k\right]^4 \tag{4.45}$$

where u_i^k is the i^{th} element of the eigenvector k. This quantity has two limits, one when all the elements of the eigenvector have the same value ($I^k \to 1/\sqrt{N}$) and another one when one element has value one and all other elements are zero ($I^k \to 1$).

4.4.2 Mean Value of Eigenvectors

Studying the elements of each eigenvector, we can determine which stocks contribute more to each eigenvector and also if there is a common market or industrial sector in these eigenvectors that contributes more. To study the influence of groups of stocks in each eigenvector, we group our stocks in markets and industrial sectors. We compute the mean value and variance for each market/industrial sector. The mean value of eigenvector k, for a group of stocks of market m and sector s is given by:

$$\left\langle \mu^k \right\rangle_{m,s} = \frac{1}{N_{m,s}} \sum_{i \in M_m, i \in S_s} u_i^k \tag{4.46}$$

where u_i^k is again the i^{th} element of the eigenvector k, M_m represents the market m, S_s represents the sector s and $N_{m,s}$ is the number of stocks that belong to sector s of market m. The symbol $\langle \cdots \rangle_{m,s}$ represents the mean over the elements that belong to market m and sector s. The respective variance of elements of market m and sector s is given by:

$$\left\langle \left[\mu^k\right]^2 \right\rangle_{m,s} - \left\langle \mu^k \right\rangle_{m,s}^2 = \frac{1}{N_{m,s}} \sum_{i \in M_m, i \in S_s} \left(u_i^k - \left\langle \mu^k \right\rangle_{m,s} \right)^2 \tag{4.47}$$

4.5 Minimum Spanning Trees

Another way to study the correlation of time series is to create a matrix of distances between time series from the correlation coefficients. With this matrix of distances we can create a network where each node represents a time series and links between nodes represent the distances between pairs of the time series. If two time series are highly correlated, the distance between them is small. The network that we use to study these properties is the Minimum Spanning Tree (MST).

4.5.1 Distances

The metric distance, introduced by Mantegna [43], is determined from a function of the Euclidean distance between vectors, $d_{ij} = \frac{1}{\sqrt{T-1}}|\tilde{\mathbf{R}}_i - \tilde{\mathbf{R}}_j|$, where $\tilde{\mathbf{R}}_i$ represents a vector for time series i with $T-1$ elements $\tilde{R}_i(t)$ for t from t_0+1 to t_0+T-1. Taking in consideration that the vector $\tilde{\mathbf{R}}_i$ is normalised, but it is not unitary $|\tilde{\mathbf{R}}_i| = \sqrt{T-1}$ it follows that:

$$d_{ij}^2 = \frac{1}{T-1}|\tilde{\mathbf{R}}_i - \tilde{\mathbf{R}}_j|^2 = \frac{1}{T-1}|\tilde{\mathbf{R}}_i|^2 + \frac{1}{T-1}|\tilde{\mathbf{R}}_j|^2 - \frac{2}{T-1}\tilde{\mathbf{R}}_i \cdot \tilde{\mathbf{R}}_j = 2 - 2\rho_{ij} \tag{4.48}$$

This relates the distance between two time series to their correlation coefficient:

$$d_{ij} = \sqrt{2(1-\rho_{ij})} \tag{4.49}$$

This distance varies between $0 \leq d_{ij} \leq 2$, where small values imply strong correlations between time series. A distance matrix D with elements d_{ij} is formed. Following the procedure of Mantegna [43], this distance matrix can be used to construct a network with the essential information of the time series.

The Minimum Spanning Tree is a sub-network of a major network. If we consider the network as all the $N(N-1)/2$ links possible between all time series, a Minimum Spanning Tree is a sub-network of this network with only $N-1$ links from the $N(N-1)/2$. The $N-1$ links are chosen to minimise the total length of the network taking in consideration that all the N nodes will be connected to the network and that no loops will occur during the construction of the network. If the distances, d_{ij} are unique, there is only one Minimum Spanning Tree. To chose the $N-1$ links for the MST many algorithms can be used, one of them is the Prim's algorithm [117]. The Prim's algorithm is given by the following steps:

1. Choose the minimum distance of the matrix D, $d_{i'j'}^{min}$ and connect the nodes i' and j' with a link;

2. Choose the next minimum distance of the matrix (for example $d_{k'l'}$);

    - If one of the nodes k' or l' has a link and the other not, **connect the nodes** k' and l' with a link and continue to 2);
    - If k' and l' both have a link, **don't connect** them and continue to 2);
    - If k' and l' both don't have a link, **don't connect** them and continue to 2);

3. After choosing $N-1$ pairs and connecting all of them, a Minimum Spanning Tree is created.

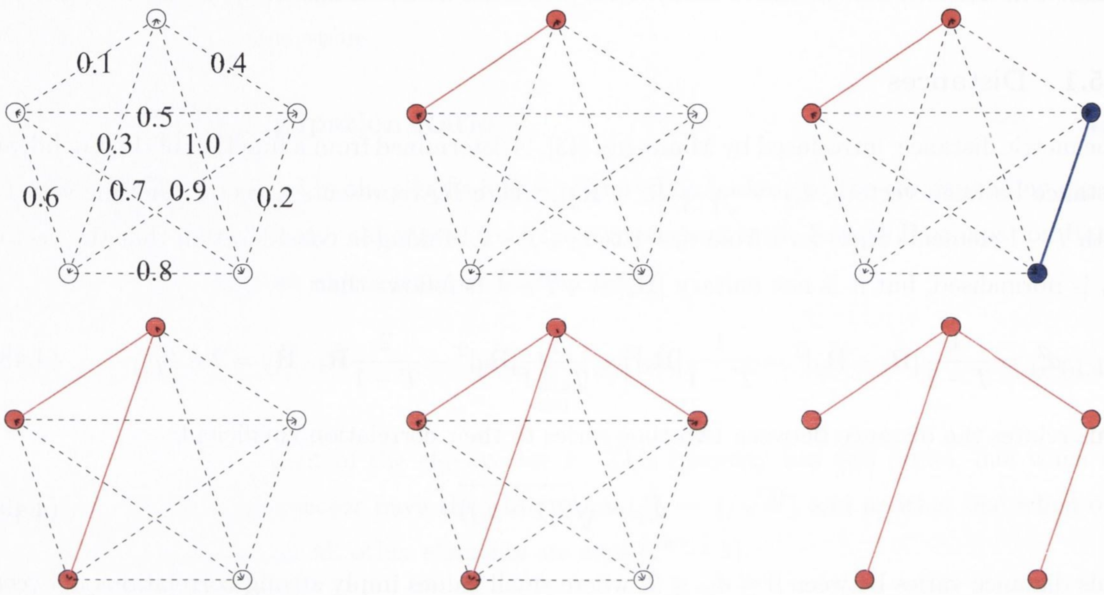


Figure 4.8: Schematic representation of the Prim's algorithm. A simple network with 5 nodes and 10 possible links. The number in each link represents the distance between pairs of nodes. Starting by choosing the minimum distance (0.1), we add a link to the nodes related with this distance. Choosing the next minimum distance (0.2), and checking the conditions of the pair, we can see that both nodes don't have a link, so we continue to the next minimum distance. And so on, until we choose 4 links and we have a fully connected network without loops, that we call Minimum Spanning Tree.

To visualise the Minimum Spanning Tree we used the Pajek software [118]. This software uses the Kamada-Kawai algorithm [119] to display the links and nodes. The algorithm introduces a dynamic system in which every two nodes are connected by a "spring" with the respective distance between two time series. The optimal layout of vertices is when the total spring energy is minimal.

With the distances from the Minimum Spanning Tree, we studied the distribution of distances in the network and the main moments (sections 5.5, 5.7 and 6.3), as the mean or normalised tree length:

$$L = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} d_{ij} \tag{4.50}$$

where Θ represents the set of distances that belong to the Minimum Spanning Tree. The other moments are the variance:

$$\nu_2 = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^2, \tag{4.51}$$

the skewness:

$$\nu_3 = \frac{1}{(N-1)\nu_2^{3/2}} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^3, \tag{4.52}$$

and the kurtosis:

$$\nu_4 = \frac{1}{(N-1)\nu_2^2} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^4. \tag{4.53}$$

Again we can divide our time series in small windows and move those windows in small steps, creating different Minimum Spanning Trees. If we compute the moments of each MST, we can study these moments in time.

4.5.2 Mean Occupation Layer

Changes in the density, or spread, of the Minimum Spanning Tree can be examined through calculation of the mean occupation layer, as defined by Onnela *et al.* [48]:

$$l(t, v_c) = \frac{1}{N} \sum_{i=1}^{N} L(v_i^t), \tag{4.54}$$

where $L(v_i^t)$ denotes the level of a node, or vertex, v_i^t in relation to the central node, whose level is defined as zero. The level of one node is the minimum number of links that separate that node with the central node.

The central node can be defined as the node with the highest number of links or as the node with the highest sum of correlations of its links. Both criteria produce similar results. When we perform a time analysis, the mean occupation layer can then be calculated using either a fixed central node for all windows, or with a continuously updated node.

4.5.3 Single and Multi Step Survival Rates

The robustness of links over time can be examined by calculating survival ratios of links, or edges in successive MST. The single-step survival ratio is the fraction of links found in two consecutive MST in common at times $t-1$ and t and is defined by Onnela *et al.* [48] as:

$$\sigma(t) = \frac{1}{N-1} |E(t) \cap E(t-1)| \tag{4.55}$$

where $E(t)$ is the set of edges of the MST at time t, $\cap$ is the intersection operator, and $|\cdots|$ gives the number of elements in the set. A multi-step survival ratio can be used to study the longer-term evolution [48]:

$$\sigma(t, k) = \frac{1}{N-1} |E(t) \cap E(t-1) \cdots E(t-k+1) \cap E(t-k)| \tag{4.56}$$

in which only the connections that continue for the entire period without any interruption are counted.

4.6 Outline of Chapters 5, 6, 7 and 8

In this chapter we saw how difficult is to conclude which probability distribution function better represents the distribution of returns, but the best fit seems to be given by the T-student distribution. One of the reasons for our conclusion is the amount of values higher than 2.0 for the α parameter of the Lévy distribution.

We also introduced some techniques used to better understand some properties of the portfolios studied. In the next chapters we will show the results for different portfolios. In Table 4.2 we represent each portfolio of stocks that we study further.

Table 4.2: Index, number of time series (N), date period, time series length (T) and chapter where those portfolios are studied.

Index	N	**date period**	T	**Chapter**
FTSE100	67	August 2^{nd} 1996 - June 27^{th} 2005	2322 days	5
	85	January 3^{rd} 2000 - March 24^{th} 2008	2146 days	5, 7 and 8
MSCI indices	44	January 8^{th} 1997 - February 1^{st} 2006	475 weeks	6
Stock market indices	9	January 8^{th} 1997 - February 1^{st} 2006	475 weeks	6
	56	January 3^{rd} 2000 - March 24^{th} 2008	2146 days	6
DJIA	30	January 3^{rd} 2000 - March 24^{th} 2008	2146 days	7 and 8
CAC40	34	January 3^{rd} 2000 - March 24^{th} 2008	2146 days	8
BEL20	17	January 3^{rd} 2000 - March 24^{th} 2008	2146 days	8
AEX	21	January 3^{rd} 2000 - March 24^{th} 2008	2146 days	8

We start with the analysis of a portfolio of stocks from only one market. The first study is of 67 stocks from the FTSE100, in Chapter 5, and then in the same chapter we also study a bigger portfolio of 85 stocks from the FTSE100. We conclude that the stocks group in terms of industrial sectors for both portfolios. The same conclusion is given for a portfolio of 30 stocks of DJIA index in Chapter 7. In Chapter 6 we analyse two portfolios of indices, first a portfolio of 53 indices, 44 MSCI indices and 9 stock market indices, and second a portfolio of 56 stock market indices. For both portfolios we concluded that the indices grouped in terms of geographical location. The next step was to analyse how the stocks from different stock market will group, if in terms of industrial sectors of in terms of geographical location. First we analyse a portfolio of stocks from the two best known indices in the world, FTSE100 and DJIA, in Chapter 7, then, in Chapter 8, we analyse a portfolio of stocks from three different indices, FTSE100, DJIA and CAC40. We choose these indices, because two of them come from the previous study in Chapter 7 and the third one is the central index of the study of Chapter 6. Because these study of these three indices main have some time-mismatch we also study a portfolio of three different indices with the same geographical location, CAC40, BEL20 and AEX, in Chapter 8.

Chapter 5

FTSE Analysis

5.1 Introduction

Our main goal is to detect any underlying structure of a portfolio, such as clustering, or identification of key stocks. We start by computing the correlation coefficient between the time series of log-returns of pairs of stocks. From these correlations we can compute a distance, for each pair, which is used for the construction of a network with links between stocks. This network is called the minimal spanning tree (MST). For some portfolios, this MST shows clustering of stocks in terms of industrial sector. In this chapter, we will show that the stocks from the main index of the FTSE100 follow this behaviour and cluster in terms of industrial sector with the stocks from the financial sector to be the backbone of the tree.

The distribution of the coefficients of the correlation matrix and the moments of this distribution are studied in this chapter. The main achieve here is the evolution of these values in time and some insights about how the correlations between stocks change for different time periods.

We also analyse the eigensystem of the correlation matrix. The higher values of the eigenvalues, and their corresponding eigenvectors show information about the portfolio that we studied. For example, the eigenvector related with the highest eigenvalue shows all the elements with the same sign. This eigenvalue is known to be related with the index of the market. The other eigenvectors related with highest eigenvalues show some segregation between stocks from different industrial sectors. For example, for some eigenvectors there is elements of one industrial sector with one sign and all the other elements with the opposite sign or also there is a bid discrepancy of the magnitudes of some elements that belong to the same industrial sector.

In this chapter we introduce a random model to simulate time series of stocks and compare the results of this random model with the ones from the real data.

We choose the stocks from the FTSE100 index, because this is one of the most popular indices among investors and it is the main index in U.K. The 100 most highly capitalised companies in the U.K. that comprise the London Stock Exchange main index, the FTSE100, represent approximately 80% of the U.K. market [120]. The first portfolio found to start our study of financial data [51] comprise the daily closing price of 67 stocks over a period of almost 9 years, starting in August 2^{nd} 1996 until June 27^{th} 2005, which equals 2322 trading days per stock (list of stocks in Appendix B).

5.2 Analysis of a global portfolio of 67 stocks from FTSE100

Starting with the analysis of the correlations of the 67 stocks, ρ_{ij} (eq. 4.30) has values between -1 and $+1$ for the 67×67 entries of the correlation matrix. The values of the elements in the diagonal of the matrix are one, and the elements in the upper triangle of the matrix mirror the ones in the lower triangle. So taking into account just the elements of one of these triangles ($67 \times 66/2$ different elements), we can compute the distribution of coefficients of the correlation matrix (Figure 5.1). The distribution of these coefficients is not symmetric, has a mean different from zero and changes according to the size of the time series chosen as we show in section 5.4.

If we shuffle a time series of stocks, where we change the value of the price of stock i at time t with the value at time t' for all times, we create a new time series with the same probability distribution of returns but with a different evolution in time. The shuffling algorithm is done as follows:

- For a time series i we pick two random times t and t' where the values of the price are $P_i(t)$ and $P_i(t')$, respectively;

- We give the value $P_i(t')$ to the new value of the price at time t and vice-versa;

- We perform this change for many pairs of times (10^4) for all the time series.

Shuffling the time series of the stocks, the correlation over time is destroyed and the value of the correlation coefficient between different time series decreases, as seen in Figure 5.1, where the dashed line shows a symmetric distribution with mean zero for the coefficients of the correlation matrix.

The eigensystem of the correlation matrix, C (eq. 4.34) is computed and the distribution of the 67 eigenvalues, $P_{real}(\lambda)$ is shown in Figure 5.2.

The analytical spectrum of eigenvalues of a random matrix, $P_{RM}(\lambda)$ can be calculated using equation 4.43. For 67 stocks with 2322 days each, the value of Q in equation 4.43 is equal to 34.6, where N is the number of stocks and T' the size of the time series minus one (taking into account the number of returns and not the number of prices of the stocks). In Figure 5.2 we

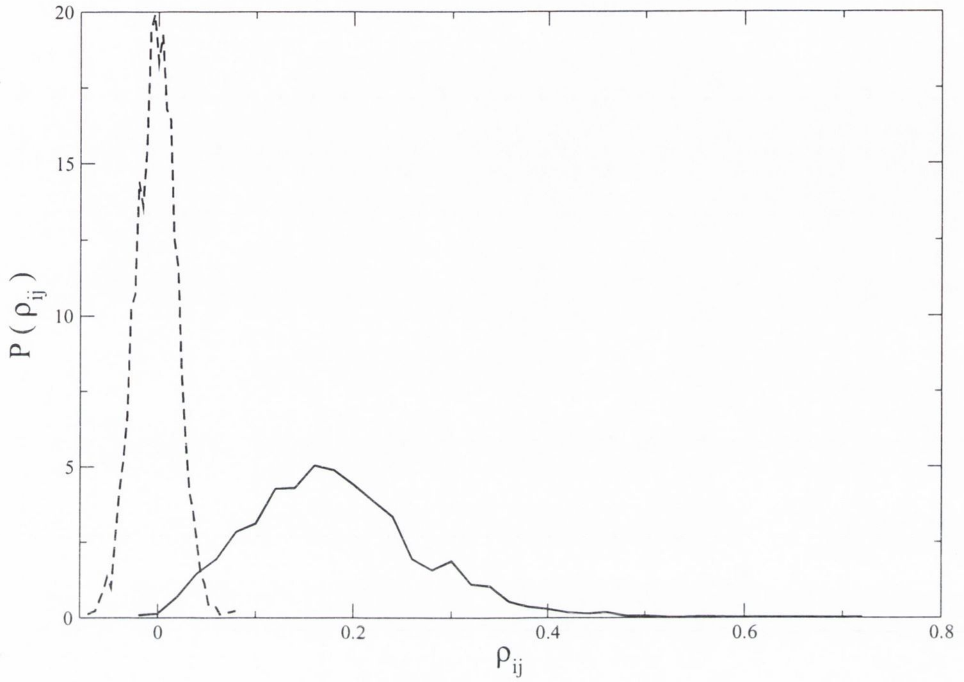


Figure 5.1: The solid line shows the distribution of coefficients of correlations ρ_{ij} between 67 stocks of the FTSE100 for the overall time series of 2322 days. The dashed line shows the distribution of coefficients of correlations between the same 67 stocks of the FTSE100, but after shuffling the time series. The correlations have now been destroyed, resulting in a symmetric distribution with mean zero.

see that some of the eigenvalues of matrix C stay outside the region predicted by the random matrix theory.

Studying the eigensystem of the correlation matrix constructed after shuffling the time series, C_{sh} we can also compare the spectrum of eigenvalues, $P_{sh}(\lambda)$ with the analytical spectrum of the eigenvalues of a random matrix, $P_{RM}(\lambda)$ as shown in Figure 5.3. All the eigenvalues from the matrix constructed after shuffling the time series stay inside the region predicted by the random matrix theory, showing that by maintaining the distribution of returns for the time series, but breaking the correlations between stocks, the financial information that is supposed to appear in the higher eigenvalues is gone.

The eigenvalues can be sorted by value from the smallest, λ_1, to the highest one, λ_{67}. There are five eigenvalues that stay outside the region predicted by random matrix, λ_{67}, λ_{66}, λ_{65}, λ_{64} and λ_{63}. The respective value of each one of these eigenvalues is: 14.58, 3, 40, 1.62, 1.54 and 1.47. These eigenvalues can be studied by looking at the respective eigenvectors. For example,

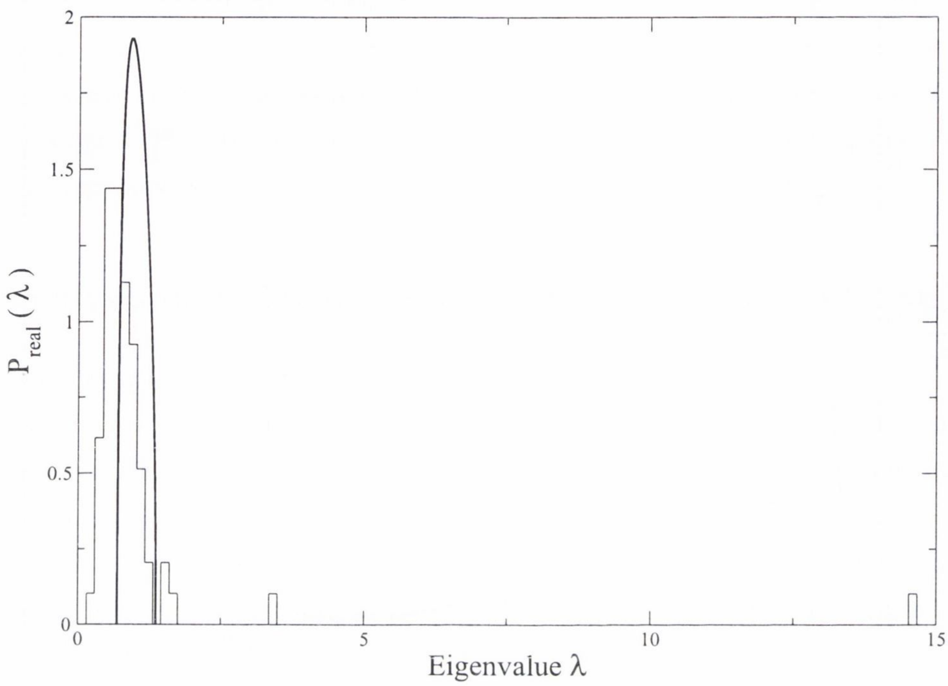


Figure 5.2: Spectrum of the 67 eigenvalues of correlation matrix of FTSE100 portfolio, $P_{real}(\lambda)$ compared with the analytical spectrum of eigenvalues of a random matrix in bold, $P_{RM}(\lambda)$.

the eigenvector that corresponds to the largest eigenvalue, λ_{67}, is shown in Figure 5.4 and it can be seen that all 67 elements have positive sign. In accordance with the Industry Classification Benchmark [121] as listed in table B.1 of Appendix B the elements are divided into different industrial sectors, each sector represented by a different colour.

The largest eigenvalue, λ_{67} can be interpreted as the collective response of the market to any external factors. Some authors [41, 42] link it to the market index. Comparing the actual index of the market with the projection of the time series in the eigenvector related to the largest eigenvalue can be helpful to understand how similar these two quantities are. The projection of the time series in the largest eigenvector, u^{67} is given by:

$$R^{67}(t) = \sum_{i=1}^{67} u_i^{67} R_i(t) \tag{5.1}$$

where the elements of the eigenvector are identified as u_i^{67} and $R_i(t)$ is the return of stock i at time t. This quantity is a weighted average of the returns of each stocks and we call it the *market mode*.

Computing the correlation between the actual index of the market and the *market mode*, it

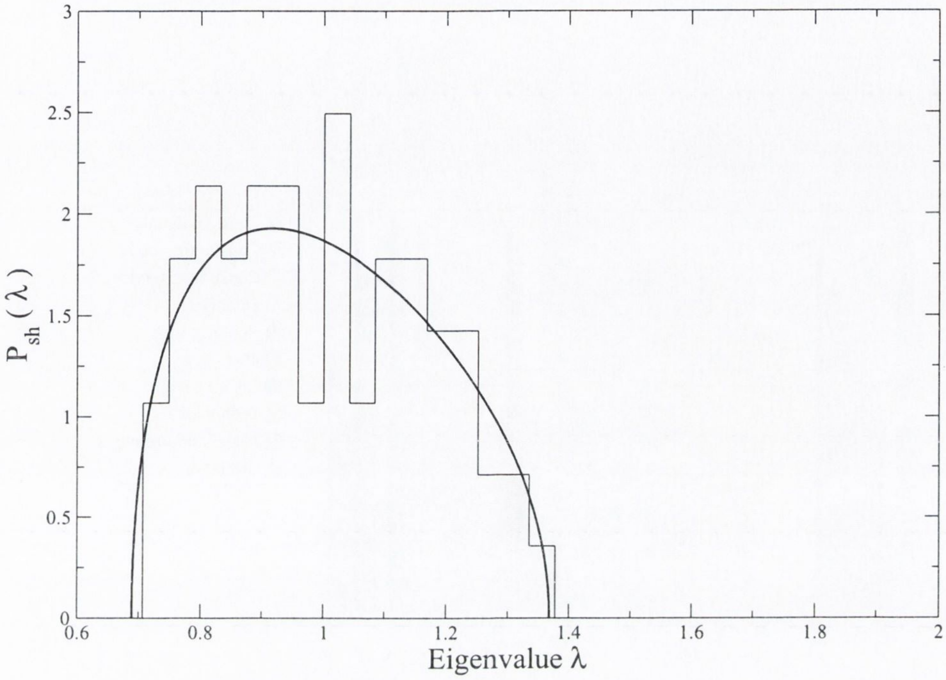


Figure 5.3: Spectrum of the 67 eigenvalues of correlation matrix of FTSE100 portfolio calculated after shuffling the time series, $P_{sh}(\lambda)$ compared with the analytical spectrum of eigenvalues of a random matrix in bold, $P_{RM}(\lambda)$. Note that some eigenvalues of the correlation matrix are much larger than predicted by random matrix theory.

can be seen in Figure 5.5 that these two quantities are strongly related with each other with a high value for correlation of 0.95.

The 67 elements of the eigenvector, u_i^{66} belonging to the second highest eigenvalue, λ_{66}, are shown in Figure 5.6. For this case not all the stocks follow the same trend (sign), some stocks have positive values and others negative ones, but it can be seen that most stocks from the same industrial sector follow the same trend. For example, for the second highest eigenvalue, λ_{66}, all the stocks from the Consumer goods sector have the same positive sign as the ones from Oil and Gas and Utilities. The stocks from Telecommunications have all negative sign. For the third highest eigenvalue, λ_{65}, all the stocks from Oil and Gas and Telecommunications have positive sign and all the stocks from Basic materials and Consumer services have negative sign. For the fourth highest eigenvalue, λ_{64}, all the stocks from Basic materials have positive sign and all the stocks from Oil and Gas, Telecommunications and Utilities have negative sign. There are some cases where stocks from the same industrial sector follow different trends, but it can be easily seen that they belong to different supersectors according to ICB [121], and so

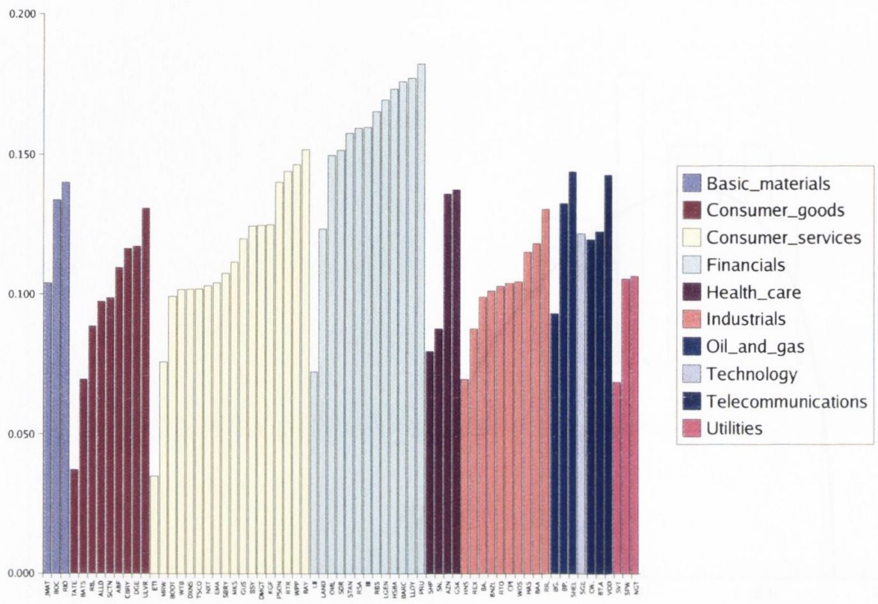


Figure 5.4: The 67 elements u_i^{67}, for $i = 1, \ldots, 67$, of the eigenvector related with the highest eigenvalue, λ_{67} of the correlation matrix C.

the segregation is made at the level of supersectors and not sectors. For example, for the second highest eigenvalue, λ_{66} there are 9 stocks from the Consumer services sector with positive sign and 9 with negative sign. All the stocks with negative sign belong to the supersectors Media and Travel and Leisure, apart from one that belongs to the Retail supersector which is the main supersector of the stocks with positive sign. The only exception for the stocks with positive sign is a stock from the Travel and Leisure supersector.

5.3 Minimal Spanning Trees

For a topological view of the market we plot the MST with all the nodes (stocks) and links between them (distances). We have studied two different classifications. First we consider the old classification for the London Exchange FTSE100, the FTSE Global Classification System [120], that was in use from 2003 until the end of 2005. This classification groups the stocks into 102 Subsectors, 36 Sectors and 10 Economic Groups.

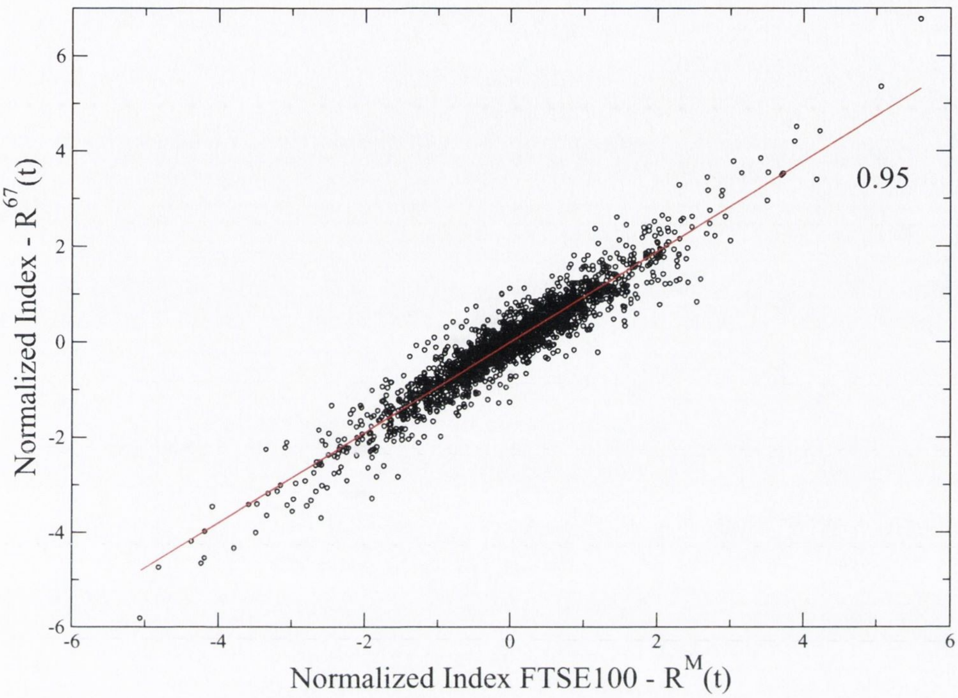


Figure 5.5: Returns of the *market mode* (equation 5.1) plotted against the actual index of FTSE100. Each point represents the value of the two quantities at the same day. The straight line shows a linear regression, where the value of the slope, computed as 0.95, represents the correlation between the indices.

Companies are divided into Economic Groups if follow a general economic theme, into Sectors if follow a general industrial theme and into Subsectors, which describe the nature of the company business. This nature is determined by the proportion of profit arising from each business areas.

Our portfolio of 67 stocks is composed of 9 economic groups and 27 sectors as shown in table 5.1.

The second classification studied is the new classification adopted by the FTSE since the beginning of 2006, the Industry Classification Benchmark [121] created by the Dow Jones Indexes and the FTSE. This classification is divided into 10 Industries, 18 Supersectors, 39 Sectors and 104 Subsectors.

The Industries can be compared with the definition of Economic Groups in the previous classification, companies from the same general economic theme. The Supersectors follow a more generic economic theme than Industries. The Sectors have the same definition from the previous classification and the Subsectors describe the nature of the company business. A company will be allocated to a Subsector whose definition most closely fits the business that

Table 5.1: Economic groups and sectors presented in the portfolio of 67 stocks in accordance with GCS [120].

Economic Groups	Sectors
Resources	Mining Oil and Gas
Basic Industries	Chemicals Construction and Building Materials
General Industrials	Aerospace and Defense
Non-cyclical Consumer Goods	Beverages Food Producers and Processors Health Personal Care and Household Products Pharmaceuticals and Biotechnology Tobacco
Cyclical Services	General Retailers Leisure and Hotels Media and Entertainment Support Services Transport
Non-cyclical Services	Food and Drug Retailers Telecommunications Services
Utilities	Electricity Utilities-others
Financials	Banks Insurance Life Assurance Investment Companies Real Estate Speciality and Other Finance
Information Technology	Software and Computer Services

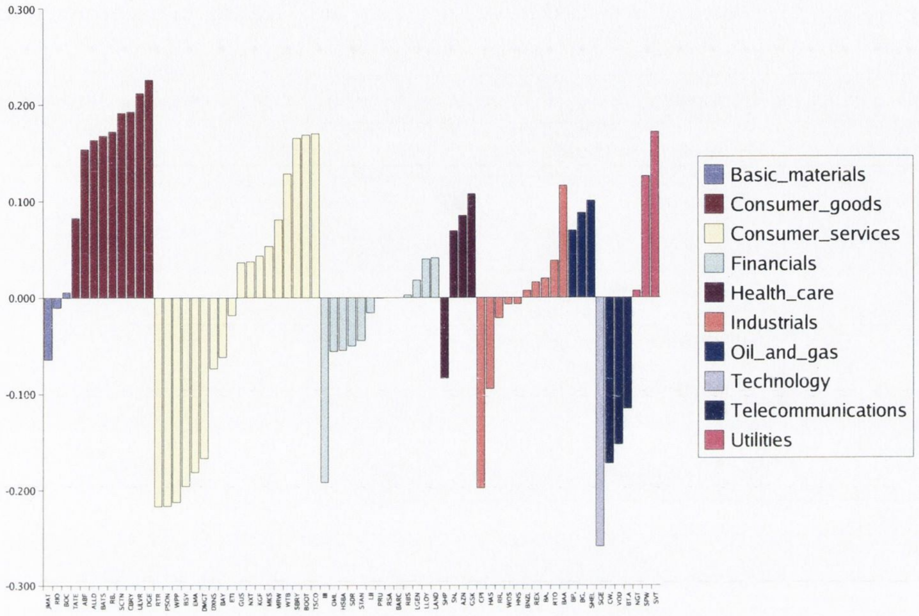


Figure 5.6: The 67 elements u_i^{66}, for $i = 1, \ldots, 67$, of the eigenvector related with the second highest eigenvalue, λ_{66} of the correlation matrix C. Most elements of the same sector have the same sign.

accounts for the primary source of the company's revenue.

Our portfolio of 67 stocks is composed of 10 industries and 30 sectors as shown in table 5.2.

For each classification we analyse the cluster formation of different economic groups (FTSE Global Classification System) or industries (ICB).

Starting with the analysis according to the old classification we represent each economic group by a different symbol: Resources (■), Basic Industries (△), General Industrials (♦), Non-cyclical Consumer Goods (□), Cyclical Services (▲), Non-cyclical Services (◊), Utilities (•), Financials (•) and Information Technology (○).

Figure 5.7 shows that the constructed MST features clusters of specific economic groups. Stocks from the Financial group are the backbone of this tree to which it seems that all other groups are connected to. The Financials, Resources, Utilities and General Industrials groups have all their stocks connected together. However for other groups divisions of stocks in sectors are apparent. For example, in the Non-cyclical Services, the Food & Drug Retailers are com-

Table 5.2: Industries and sectors presented in the portfolio of 67 stocks in accordance with ICB [121].

Industries	Sectors
Oil and Gas	Oil and Gas Producers
Basic Materials	Chemicals Mining
Industrials	Construction and Materials Aerospace and Defense General Industrials Industrial Transportation Support Services
Consumer Goods	Beverages Food Producers Household Goods Tobacco
Health Care	Health Care Equipment and Services Pharmaceuticals and Biotechnology
Consumer Services	Food and Drug Retailers General Retailers Media Travel and Leisure
Telecommunications	Fixed Line Telecommunications Mobile Telecommunications
Utilities	Electricity Gas, Water and Multiutilities
Financials	Banks Nonlife Insurance Life Insurance Real Estate General Financial Equity Investment Instruments Nonequity Investment Instruments
Technology	Software and Computer Services

pletely separated from the Telecommunication Services. Within Cyclical Services, the General Retailers, Media & Entertainment and Transports are three different clusters and the Support Services are isolated stocks connected to the Financial branch. In Non-cyclical Consumer Goods, the Health and Pharmaceuticals & Biotechnology form one cluster whereas Beverages, Tobacco, Food Producers & Processors and Personal Care & Household Products form another.

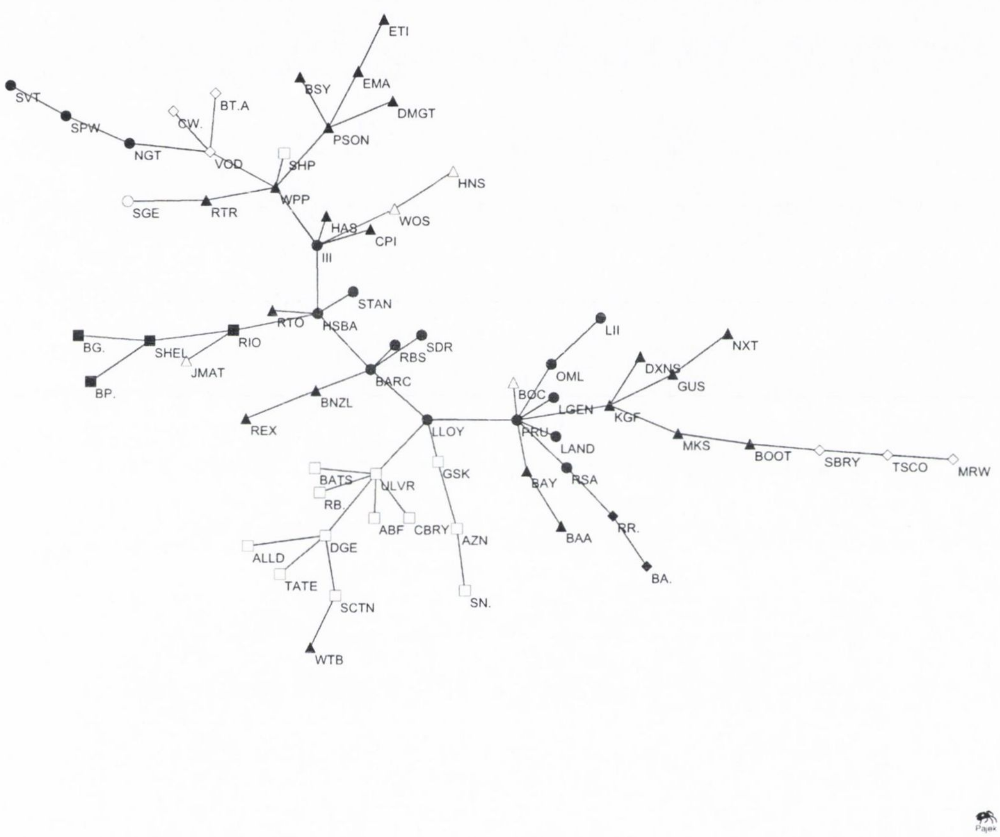


Figure 5.7: Minimal Spanning Tree for 67 stocks of the FTSE100. The time series of each stock are composed by 2322 daily closing prices. Each symbol correspond to a specific economic group from the FTSE Global Classification System: Resources (■), Basic Industries (△), General Industrials (♦), Non-cyclical Consumer Goods (□), Cyclical Services (▲), Non-cyclical Services (◊), Utilities (•), Financials (•) and Information Technology (○).

For the new classification in place since January 2006, we represent each industry by the following symbol: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◊), Utilities (•), Financials (•) and Technology (○). The MST is shown in Figure 5.8. The Financial industry has the same stocks as the one in the old classification, so it still works as the backbone of the tree. Financials, Oil & Gas, Utilities, Telecommunications and Consumer Goods have all their stocks connected together. In the Consumer Services, the supersectors Retail and Media are two big clusters but they are not connected together. The other supersector from this industry, the Travel & Leisure is disperse in the tree. Health Care industry is almost one cluster, but the stock SHP is not connected to the others. In the Industrials industry all stocks from the Support Services sector are connected to the Financial industry. The other stocks in this sector are located in isolation at other points within the tree.

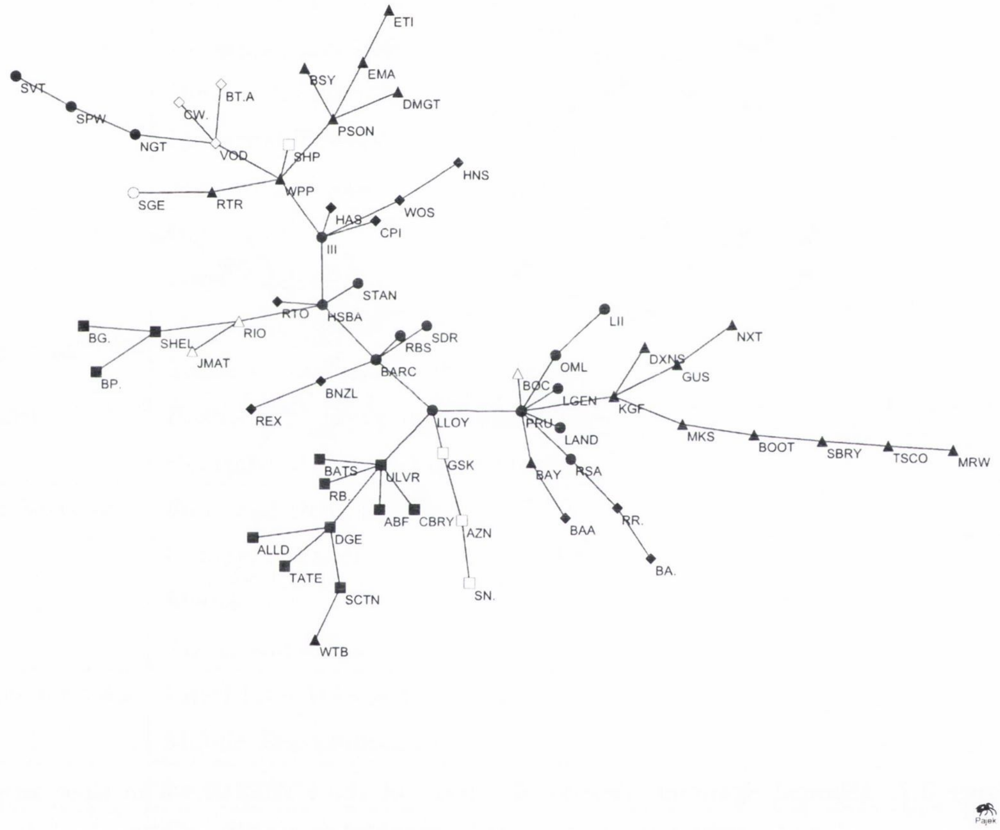


Figure 5.8: Minimal Spanning Tree for 67 stocks of the FTSE100. The time series of each stock are composed by 2322 daily closing prices. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (◆), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

The new classification adopted by the FTSE in January 2006 clearly mimics much more closely the MST results as we can see from Figures 5.7 and 5.8. The implementation of the new supersector groups ensures that apart from some notable exceptions stocks from the same supersector are now connected. It is possible that the few stocks separated from their main cluster are isolated by chance and over time they will join the appropriate clusters. However there could be other more fundamental reasons for their separation. Nevertheless it seems clear from this analysis that the MST approach is one that should complement current approaches to the development of stock taxonomy.

Coronnello *et al.* [122] have studied the topology of a portfolio of stocks from the London Stock Exchange using daily and intra-day data for 92 stocks, from year 2002. The MST for daily data looks quite different from the one shown in Figure 5.7. Using our data and studying the MST for each year, we can see that for 2002, the main hubs of the MST are the stocks from Barclays, Royal Bank of Scotland and Shell, each of them with 11, 8 and 7 links, respectively. The simple inclusion of the stock from Barclays in our study (not included in the portfolio of [122]) gives a quite different network. But the main clusters are the same in both studies.

5.4 Determination of time parameters

Apart from studying the correlations of the stocks for the overall time, we also divide the time series in small time windows for an analysis of the time dependence of correlations and distances. These time windows have a width T and overlap each other. The total number of windows depends on the window step length parameter, δT. A sketch of different time windows with width T from the same time series is shown in Figure 5.9.

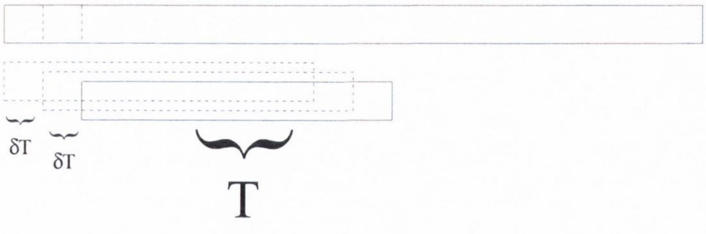


Figure 5.9: Sketch of different time windows with width T from the same time series. The time windows are moved over time by a step length δT.

Depending on the length of the time series, the correlation coefficient between two stocks changes. Thus the distance between the two stocks will be different and the MST constructed will have different characteristics. In order to select appropriate values for the size of time windows (T) and window step length parameter (δT) we looked at earlier studies in this field. As shown previously [48], the first and second moment of the correlations (mean correlation and variance) are strongly correlated. We thus computed the value of this correlation as a function of T and δT (Figure 5.10).

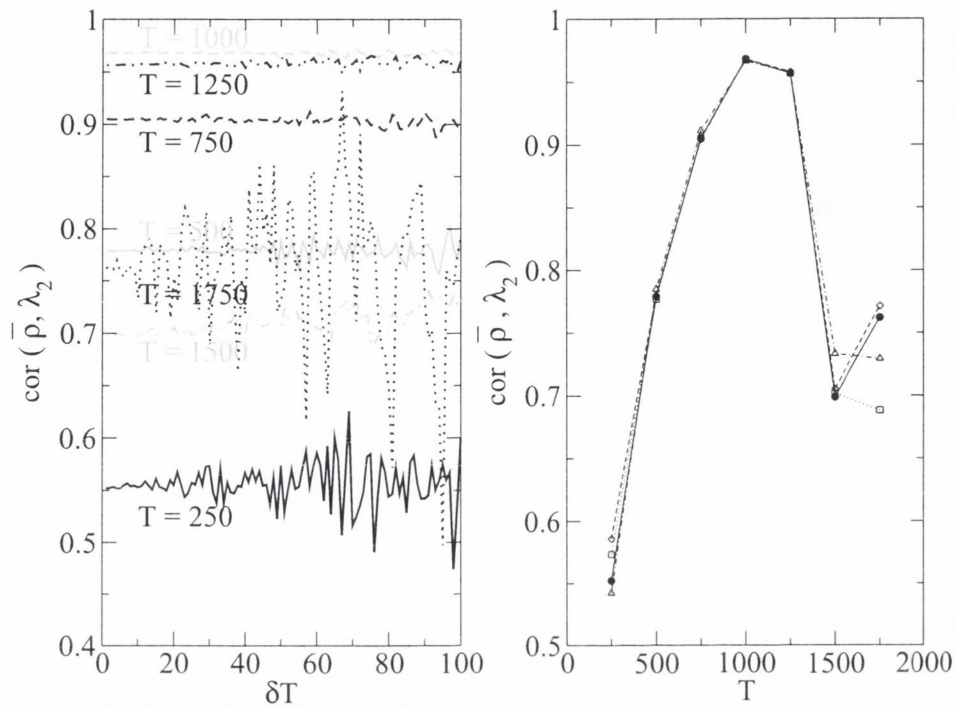


Figure 5.10: Correlation between the first two moments of the correlation coefficients, mean $\overline{\rho}$ (eq. 4.35) and variance λ_2 (eq. 4.36) as a function of T and δT. The left figure shows the correlation for different T as a function of δT. The right figure shows the correlation for $\delta T = 1$ (•), $\delta T = 30$ (□), $\delta T = 60$ (◊) and $\delta T = 90$ (△), as function of T.

Clearly, for all T, the correlation between the two moments is not only positive but strong, above 0.9 for $T = 750$, $T = 1000$ and $T = 1250$. Apart from $T = 250$ and $T = 1750$ there are only very small fluctuations for the correlation value, when we vary δT. Since when we increase δT, we are essentially removing points from our data, we decided to use the smallest value of δT (1 day) in all of the following. At this stage it is not easy to understand the non-monotonic behaviour observed in Figure 5.10. It could be associated with the reduction in data but it requires further study.

5.5 Analysis of Global Portfolio of the FTSE100 index

Some events such as wars or crashes occurred during the period of study and are noted in Figure 5.11 that shows the absolute return of the FTSE100 index. After these occurrences, which have a negative effect on stock values, all the stocks seems to follow each other, and both the correlation between them and mean correlation increase [54].

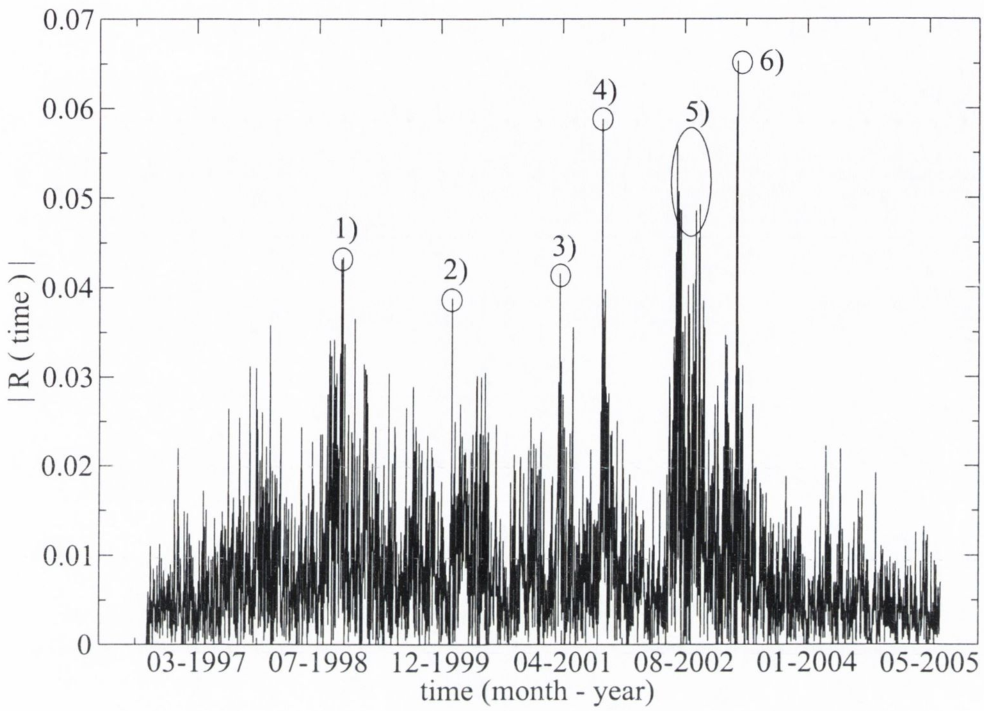


Figure 5.11: Absolute return of the FTSE100 index. Higher values indicate special days like beginning of wars or crashes. 1) Russian crash; 2) NASDAQ crash; 3) Beginning of US recession; 4) September 11^{th} 2001; 5) Stock Market downturn of 2002; 6) Beginning of Iraq War.

The time dependence of the mean correlation, the normalised tree length and the higher moments associated with these two quantities were studied for a time window of 500 days and window step length of 1 day. Figure 5.12 shows that the mean and variance of the correlation coefficients are highly correlated (0.779), the skewness and kurtosis are also highly correlated and the mean and skewness are anti-correlated. This implies that when the mean correlation increases, usually after some negative event in the market, the variance increases. Thus the dispersion of values of the correlation coefficient is higher. The skewness is almost always different from zero, which means that the distribution is asymmetric, but after a negative event the skewness moves towards zero, and the distribution of the correlation coefficients becomes more symmetric.

From Figure 5.13, we see how the normalised tree length changes with time. As expected from equation 4.49, when the mean correlation increases, the normalised tree length decreases and vice versa. Here, the mean and the variance of the normalised length of the tree are anti-correlated but the skewness and the mean continue to be anti-correlated. This means that after some negative event impacts the market, the tree shrinks, so the mean distance decreases [54], the

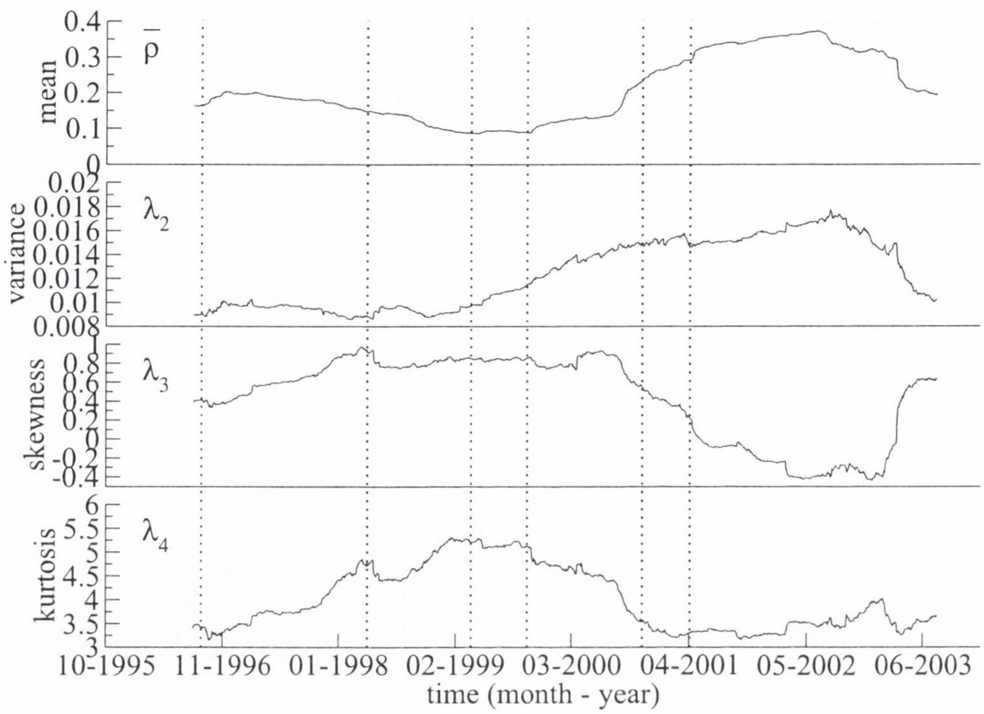


Figure 5.12: Mean (eq. 4.35), variance (eq. 4.36), skewness (eq. 4.37) and kurtosis (eq. 4.38) of the correlation coefficients. Time windows of length 500 days are moved with a window step length parameter of 1 day. The vertical lines show the external events that affect the market. 1) Russian crash; 2) NASDAQ crash; 3) US recession; 4) September 11^{th} 2001; 5) Stock Market Downturn of 2002; 6) Iraq War.

variance increases implying a higher dispersion of the values of distance and the skewness, that is almost always negative, increases towards zero showing that the distribution of the distances of the MST gets more symmetric.

From Figures 5.12 and 5.13 we can see that the external events correspond to dates different from the ones indicated in Figure 5.11, because in these figures we are using time windows of 500 days, and the changes in the values of the moments occur when these external events are introduced in the time window, normally when they are in the last day of the time window. So if we sum the size of the time windows to the dates in Figures 5.12 and 5.13 we will get the dates for external events of Figure 5.11.

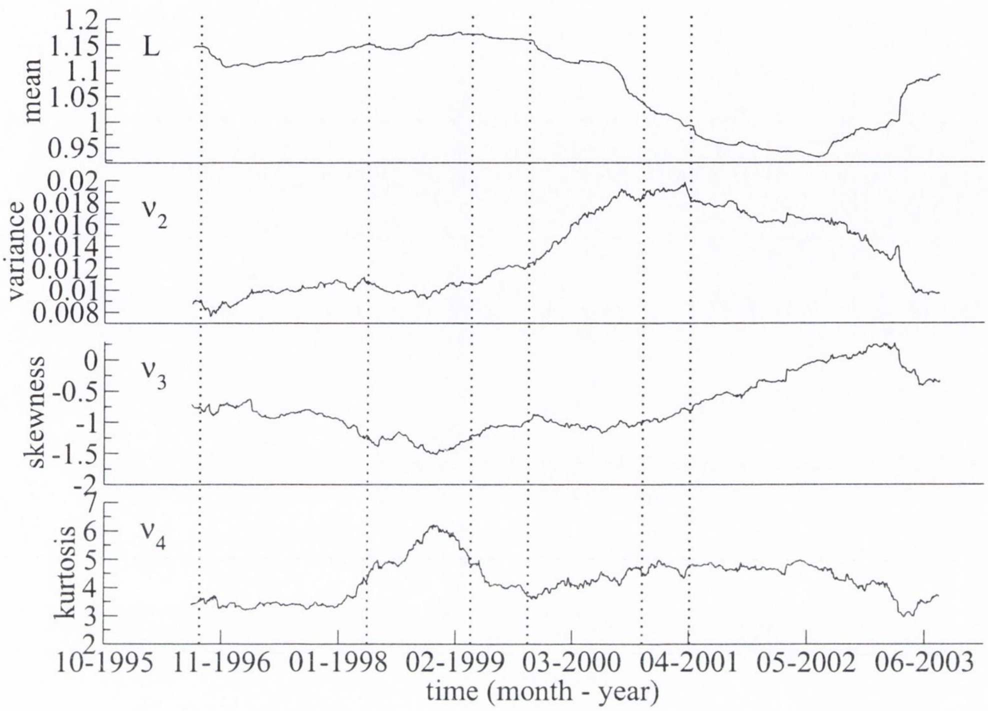


Figure 5.13: Mean (eq. 4.50), variance (eq. 4.51), skewness (eq. 4.52) and kurtosis (eq. 4.53) of the normalised tree length. Time windows of length 500 days are moved with a window step length parameter of 1 day. The vertical lines show the external events that affect the market. 1) Russian crash; 2) NASDAQ crash; 3) US recession; 4) September 11^{th} 2001; 5) Stock Market Downturn of 2002; 6) Iraq War.

5.6 Numerical Simulations of MST

5.6.1 Random market and one-factor model

In order to examine further the underlying nature of the time series we now use random time series computed from two different models. Modelling the log-returns as random numbers from a specific distribution, we can compute the correlations, distances and trees for this random series. As in [45, 47], our first approach was to consider the returns as random variables derived from a Gaussian distribution. So, using the real mean value, μ_i of each real time series of 2322 days and the specific real variance, σ_i we compute random series for our *random market*:

$$r_i(t) = \mu_i + \epsilon_i(t) \tag{5.2}$$

where $\epsilon_i(t)$ is the stochastic variable from a Gaussian distribution with variance σ_i. The MST for this random time series is represented in Figure 5.14.

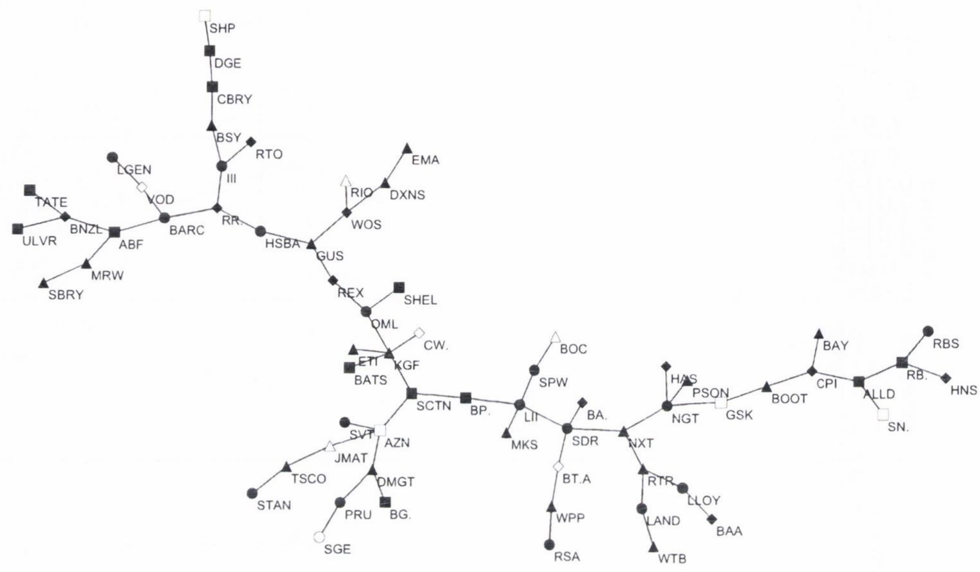


Figure 5.14: Minimal Spanning Tree for 67 random time series using random variables from a Gaussian distribution. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (▦), Health Care (□), Consumer Services (▲), Telecommunications (◊), Utilities (•), Financials (•) and Technology (○).

This MST shows no clustering according to the fact that there is no grouping of stocks of the same industrial sector, the stocks are distributed randomly in the network. To create random time series with more real characteristics we introduce a control term (the return of the FTSE100 index) and we compute one-factor model [45, 47], also known as a market model:

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \epsilon'_i(t) \tag{5.3}$$

where α_i and β_i are parameters estimated by the least square method from our data as shown below, $R_m(t)$ is the market factor (return of the FTSE100 index) and $\epsilon'_i(t)$ is the stochastic variable from a Gaussian distribution with variance σ'_i. The value of σ'_i in equation 5.3 is different from the value of σ_i in equation 5.2. The variance σ'_i is calculated from the time series $R_i(t) - \alpha_i - \beta_i R_m(t)$ where $R_i(t)$ is the real time series of returns of stock i. The two factors α_i

and β_i are calculated as:

$$\begin{aligned} \alpha_i &= \langle R_i(t)\rangle - \beta_i\langle R_m(t)\rangle \\ \beta_i &= \frac{cov(R_i(t), R_m(t))}{\sigma^2_{R_m}} \end{aligned} \quad (5.4)$$

where $cov(\ldots,\ldots)$ is the covariance:

$$cov(R_i(t), R_m(t)) =< R_i(t)R_m(t) > - < R_i(t) >< R_m(t) > \quad (5.5)$$

$\sigma^2_{R_m}$ is the variance of the returns of the FTSE100 index:

$$\sigma^2_{R_m} =< (R_m(t))^2 > - < R_m(t) >^2 \quad (5.6)$$

and $R_i(t)$ is the returns of real time series of returns of stock i. The MST for random time series created using this model is shown in Figure 5.15. Again we use data from the whole time series available (2322 days).

This network is completely different from the previous random network, apart from the random distribution of stocks from the same sector along the network, we can see that the number of links of some stocks is much higher than in the MST of Figure 5.14. However, the presence of 6 nodes with up to 13 links differs from the topology of real data. MST's computed using 5-minute data [122] have a greater similarity to the trees based on the one-factor model than trees computed using daily data. Coronnello *et al.* [122] concluded that the MSTs are more hierarchically structured when they used daily returns confirming that the networks computed using 5-minute data are not as fully formed as they are with daily data. The MSTs for 5-minute data show that the clustering is less pronounced than in the MSTs constructed from daily data and the stocks are organised around two main hubs with a large amount of links, 29 and 17 links, respectively.

But as we showed in chapter 4, the time series of stock returns are better fitted by a T-student distribution then by a Gaussian distribution. So we created random time series where instead of using stochastic variables from a Gaussian distribution we used random variables from T-student distributions.

To create the T-student stochastic variables we used a numerical routine from the GNU Scientific Library [123] that generates T-student variables from the distribution:

$$p_\nu(w)dw = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi\nu}} \left[1 + \frac{w^2}{\nu}\right]^{-\frac{\nu+1}{2}} dw \quad (5.7)$$

which is different from the distribution previous presented in Chapter 4 (eq. 4.7). But considering $k = (\nu + 1)/2$ and the assumption:

$$x = \sigma\sqrt{\frac{2k-3}{2k-1}}w \quad (5.8)$$

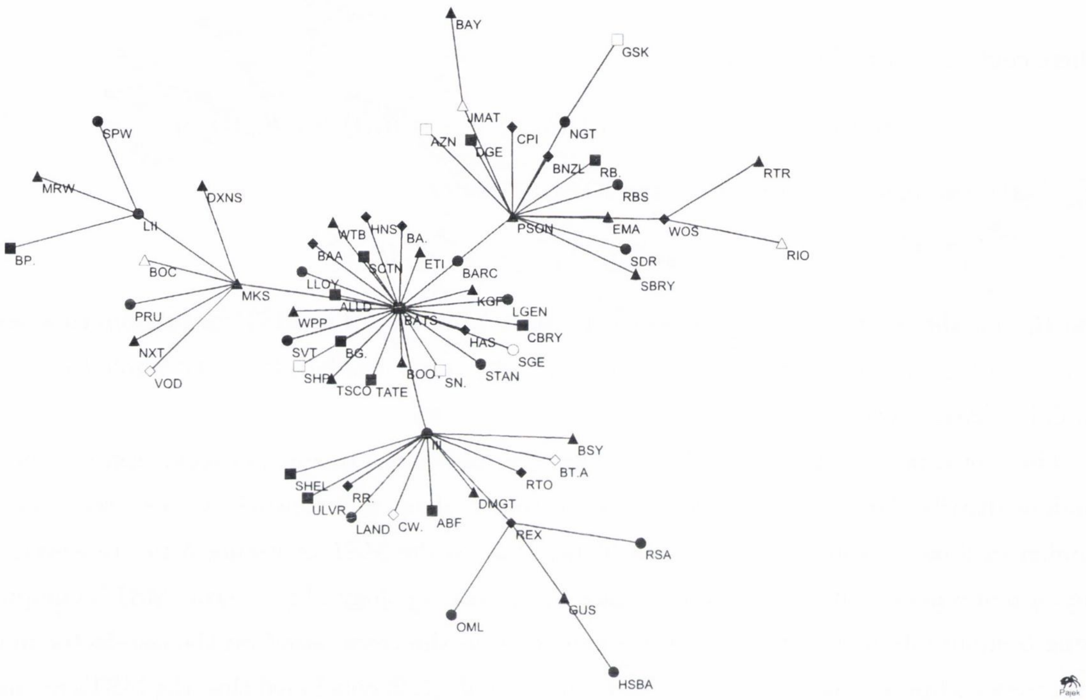


Figure 5.15: Minimal Spanning Tree for 67 random time series using the one-factor model with stochastic variables from a Gaussian distribution. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

we can transform the variables from one distribution into variables from the other.

So, if we generate a stochastic variable w with $\sigma = 2k - 1$ from the distribution of equation 5.7, we can transform this variable into a variable x from the distribution of equation 4.7, using the transformation presented in equation 5.8.

In Figures 5.16 and 5.18 we represent the MST for the random market and market model, respectively.

The MST of random time series created using the random market with T-student random variables shows some similarities with the real MST in Figures 5.7 and 5.8. The structure of the MST or the degree distribution, which is the distribution of number of links of a node in a tree, shows that the maximum number of links in the MST created using the random market with

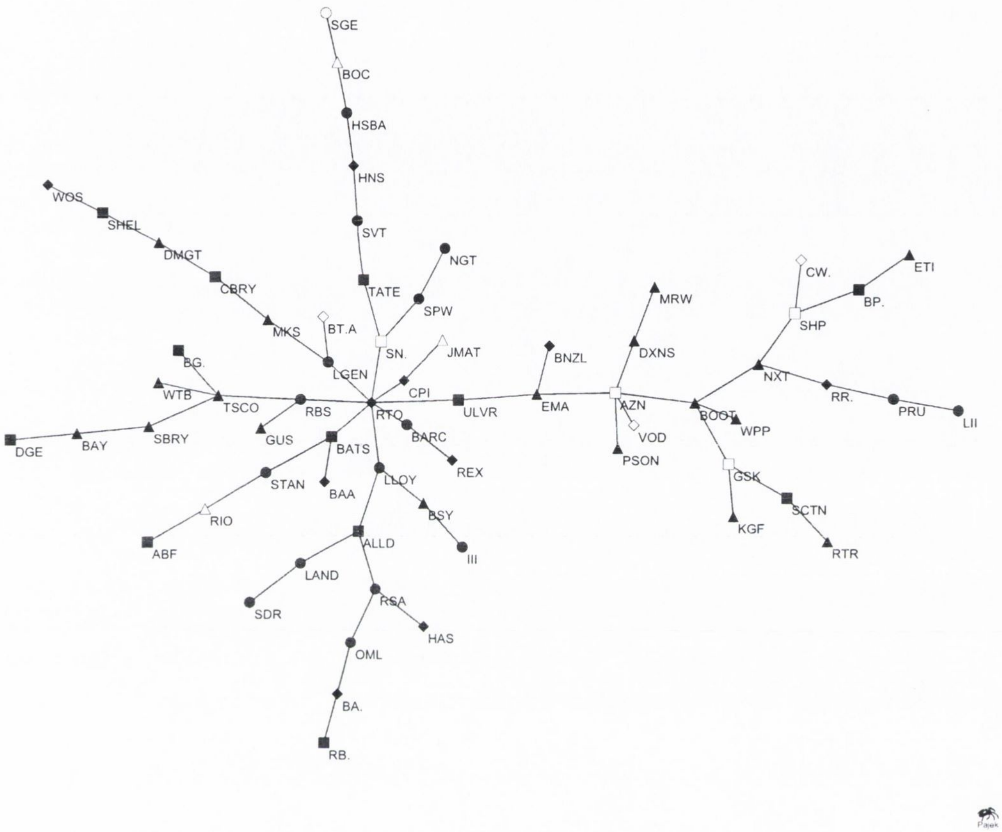


Figure 5.16: Minimal Spanning Tree for 67 random time series using random variables from a T-student distribution. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◊), Utilities (•), Financials (•) and Technology (○).

Gaussian variables is 5, where for the real case and random market with T-student variables is 8 (Figure 5.17).

The MST created using the market model with T-student random variables is like the one for Gaussian random variables very different from the real MST in Figures 5.7 and 5.8. Again this MST is more similar with the topology of real MST computed using high frequency data [122]. The one-factor model a very simple model that cannot mimic all the features and correlations that exist in the stock market. Even when we introduce random variables from T-student distribution that fit better the data of the return of the price, the one-factor model does not solve our problem.

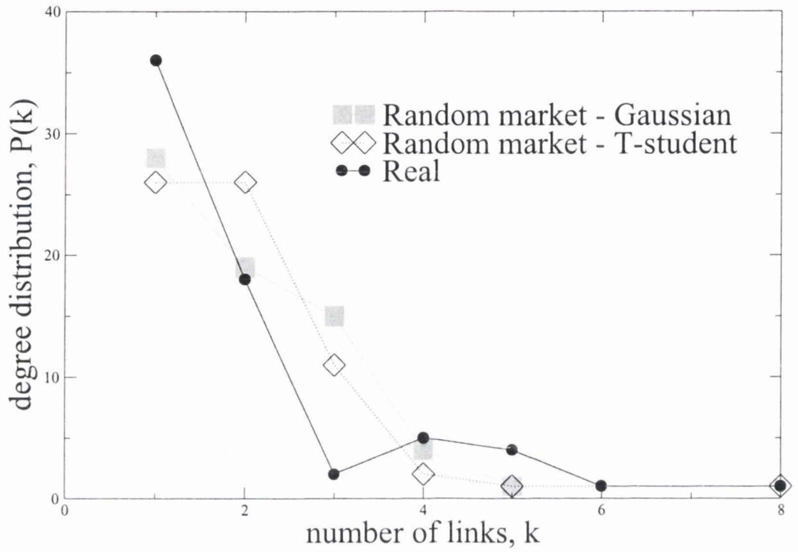


Figure 5.17: Degree distribution for the case of real MST (black circle), random MST from a market model with Gaussian variables (white diamond) and random MST from a market model with T-student variables (grey square).

5.6.2 Evolution with time

After creating the random time series using the market model, we can compute their behaviour in time as we did for the real case in Figure 5.12, where we compute the main moments of the distribution of correlation coefficients over time. In Figure 5.19 we compare these moments for the real case (black lines) and for the random case (grey lines) using stochastic variables from T-student distributions.

From Figure 5.19, we can see that using the market model with stochastic variables from T-student distributions, we can mimic the evolution of the mean and the skewness of the correlation coefficients, in time. The variance of the correlation coefficients from the random case follow the same trend but the values are not the same as in the real case. From the evolution of the moments of the random case we can also see the correlations and anti-correlations between different moments as we stated previous, in section 5.5, for the real case.

Studying the evolution of the moments of distances presented in each MST, we can see if the market model can at least be trusted as a model to mimic the distances of the MST, because we already saw that we cannot use it as a good model for the structure of the MST.

In Figure 5.20, we compare the moments of the distances in the tree for the real case (black lines) with those from the random case (grey lines) using stochastic variables from T-student

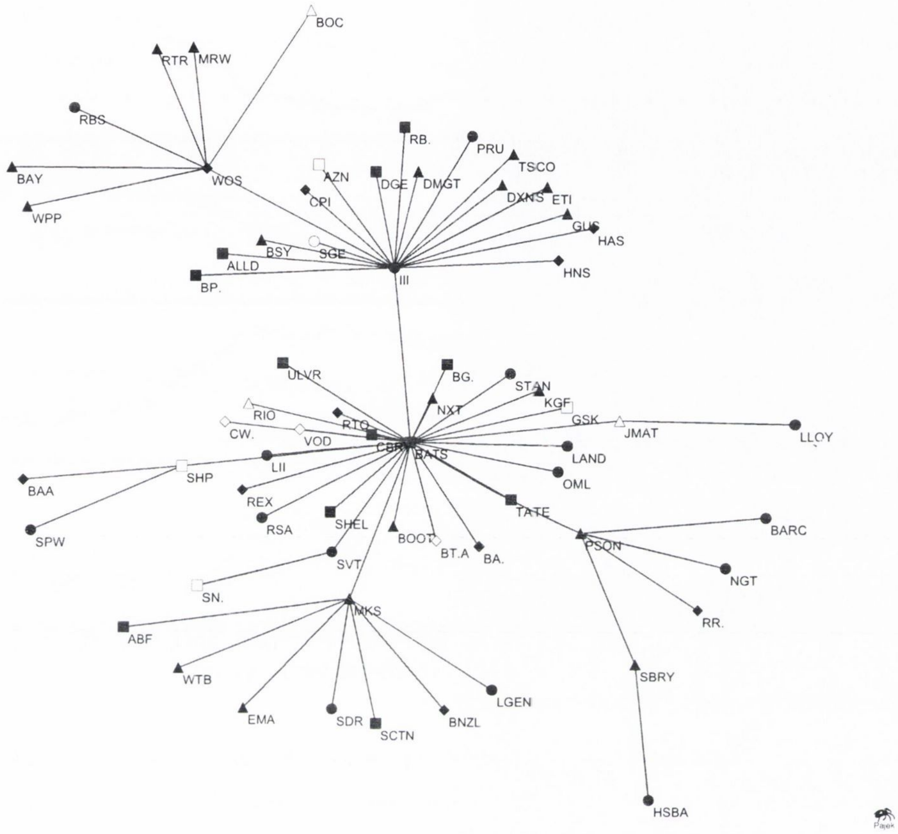


Figure 5.18: Minimal Spanning Tree for 67 random time series using the one-factor model with stochastic variables from a T-student distribution. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

distributions.

From Figure 5.20, we can see that apart from the similar trend in the evolution of the mean distance of the tree for the random case, the other moments are very different for the real and random cases. So again we see the shortcoming of our simple models in describing the data.

5.7 Study of portfolio of 85 stocks of FTSE100

A portfolio of stocks from the FTSE100 index (as it is in April 1^{st} 2008) was also studied. The main reason for this new study was the facility of new data from the Thomson Database Datastream. The main index of the London Stock Exchange has 102 stocks quoted as shown in

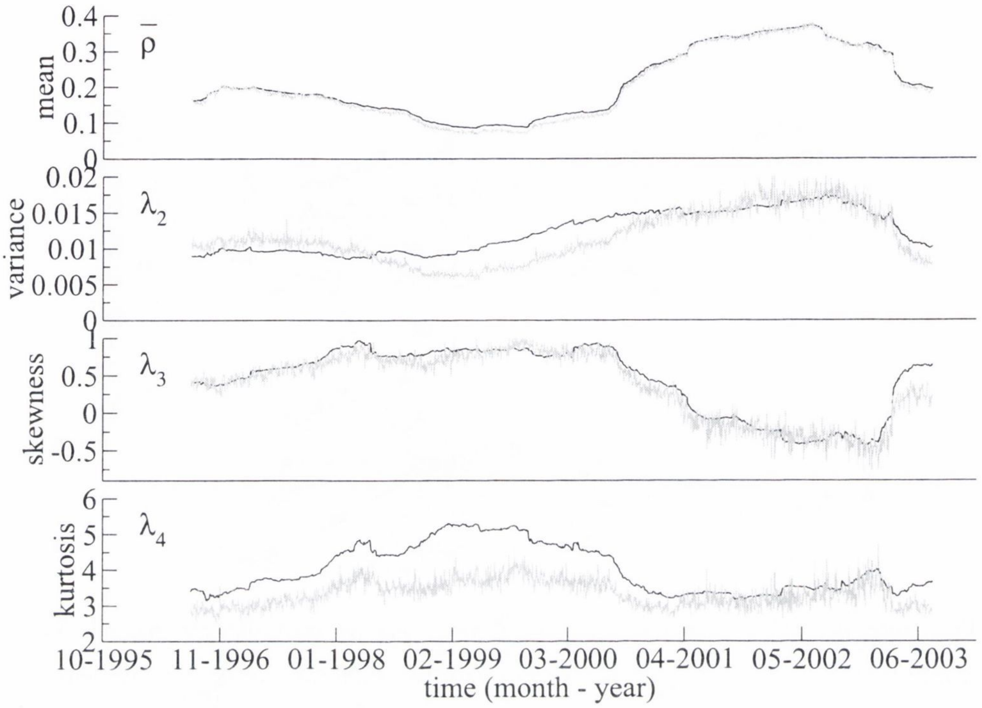


Figure 5.19: Mean (eq. 4.35), variance (eq. 4.36), skewness (eq. 4.37) and kurtosis (eq. 4.38) of the correlation coefficients. Time windows of length 500 days are moved with a window step length parameter of 1 day. The black lines represent the results from the real FTSE data and the grey lines the results from the random data using stochastic variables from T-student distribution.

table C.1 of Appendix C. But not all of these stocks are listed in the database from the same date, so we decided to choose a part of these stocks that have non-zero data from January 3^{rd} 2000 until March 24^{th} 2008, which gives us more than 8 years of data. The total number of stocks in our portfolio is 85.

5.7.1 Minimal Spanning Tree

The minimal spanning tree of this portfolio constructed as laid out in section 4.5, is shown in Figure 5.21.

In the MST of Figure 5.21 the clustering is very similar to the MST of Figures 5.7 and 5.8. As the backbone of the MST we have some of the major Banks that belong to the Financial industrial sector. One characteristic of this portfolio is the inclusion of Alliance Trust (ATST), an Equity Investment Instruments company. This stock is the main hub of the MST with 21 links to other stocks from a variety of different sectors. Apart from this fact, the MST continues

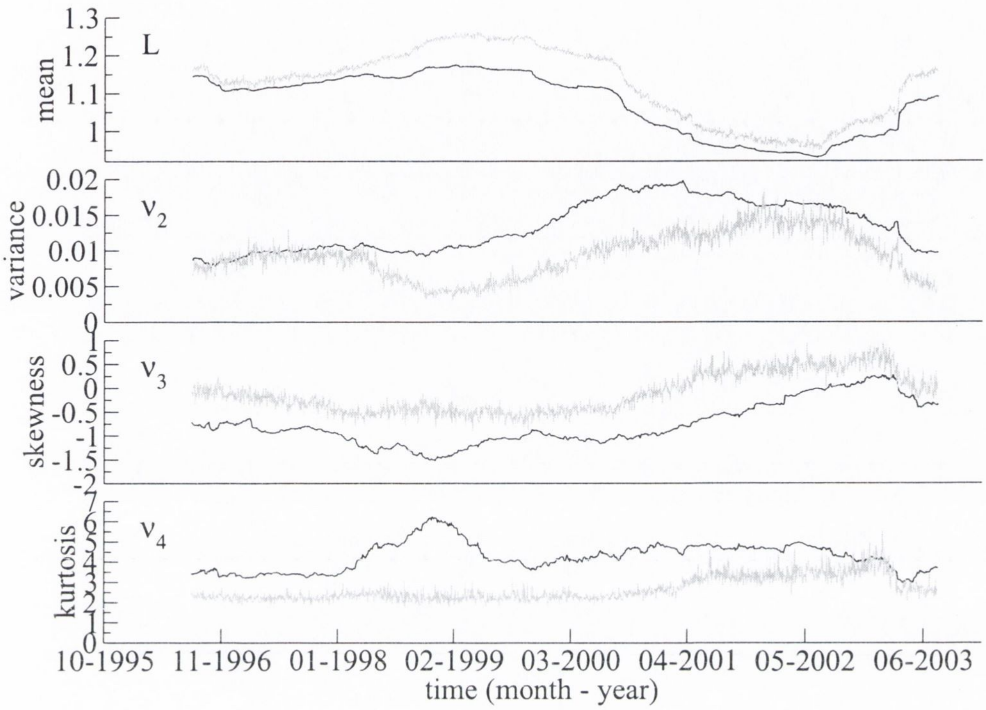


Figure 5.20: Mean (eq. 4.35), variance (eq. 4.36), skewness (eq. 4.37) and kurtosis (eq. 4.38) of the normalised tree length. Time windows of length 500 days are moved with a window step length parameter of 1 day. The black lines represent the results from the real case and the grey lines the results from the random case using stochastic variables from T-student distribution.

to have different clusters such as Basic Materials, the Media (part of Consumer Services), the Health and Care, the Retail (part of Consumer Services), the Oil and Gas, the Utilities, the Telecommunications, the Consumer Goods and also a Real Estate cluster inside the Financial cluster. Again all the stocks that belong to the Financial sector are linked together.

5.7.2 Eigensystem analysis

Analysing the eigenspectrum of the correlation matrix of this portfolio of 85 stocks we can see which eigenvalues correspond to different industrial sectors. In Figure 5.22 we can see the spectrum and distribution of eigenvalues (section 4.4). There are 6 eigenvalues with values higher than λ_{max} which is the maximum value predicted by the RMT for a random matrix. The study of the structure of the eigenvectors that correspond to these eigenvalues gives more insights about the clustering in sectors of this portfolio.

In Figure 5.23 we show the eigenvectors that correspond to the 6 eigenvalues with values

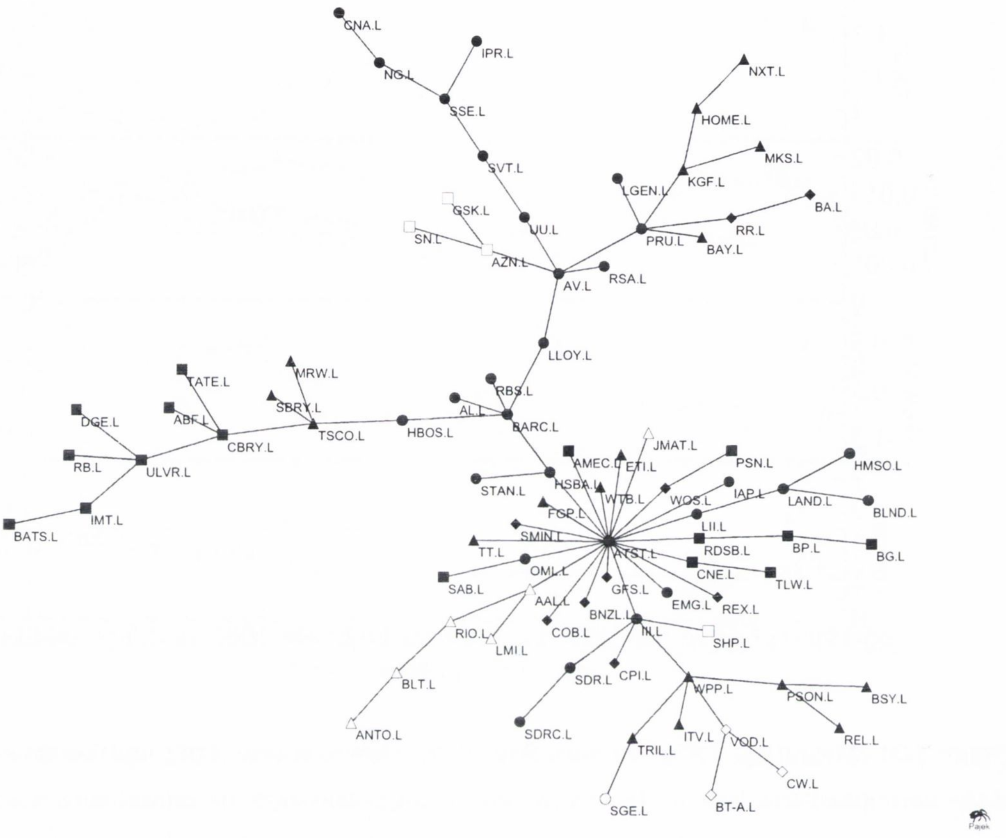


Figure 5.21: Minimal Spanning Tree for 85 stocks of the FTSE100 index. The time series start at January 3^{rd} 2000 and run until March 24^{th} 2008 for 2146 days. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

higher than λ_{max}. For each eigenvector we grouped the stocks in terms of industrial sectors: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. We also included error bars that represent the variance of each industrial sector (section 4.4.2).

For the eigenvector belonging to the highest eigenvalue, λ_{85} we can see that all the elements have a positive sign as shown in Figure 5.4.

For the eigenvector of the second highest eigenvalue, λ_{84}, the stocks from the Telecommunications industrial sector all have a negative sign and all the stocks from Utilities and Consumer Goods sectors have positive sign. This behaviour was already shown in section 5.2 also for the second highest eigenvalue and in Figure 5.6.

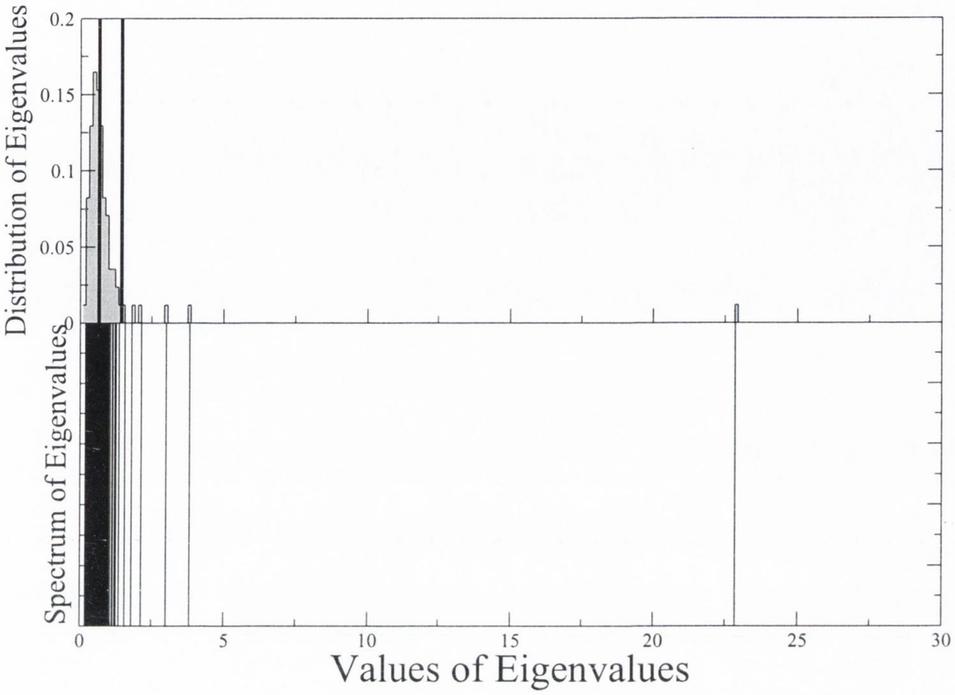
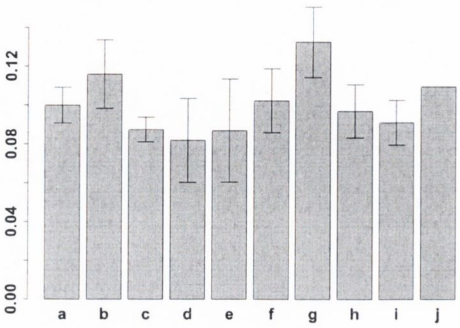


Figure 5.22: Spectrum and distribution of eigenvalues for a portfolio of 85 stocks from the FTSE100 index. The thick vertical lines in the upper figure show the limits λ_{min}^{max} (eq. 4.44). Just 40% of the eigenvalues are inside the region predicted by the RMT.

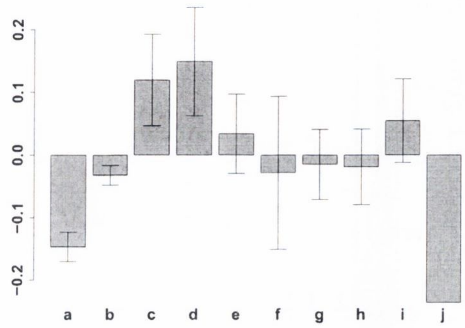
For the eigenvector related with the third highest eigenvalue, λ_{83}, the stocks from Telecommunications and Health Care all have a positive sign and all the stocks from Basic Materials have a negative sign. For the third eigenvalue of the previous portfolio of section 5.2 the Telecommunications and Basic Materials also follow this trend.

For the eigenvector related with the fourth highest eigenvalue, λ_{82}, the stocks from Basic Materials and Oil and Gas have a negative sign, only the stocks from Oil and Gas follow the same trend shown for the portfolio of section 5.2.

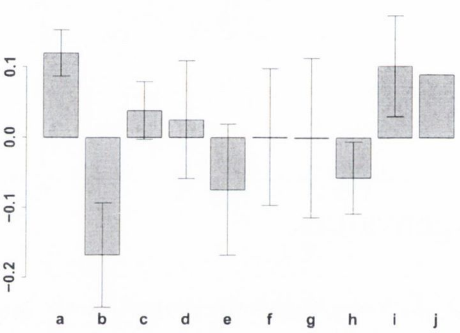
For the eigenvector related with the fifth highest eigenvalue, λ_{81}, the main sectors are Telecommunications and Utilities with positive sign and for the eigenvector related with the sixth highest eigenvalue, λ_{80} the main sectors are Consumer Services and Industrials with positive sign and Utilities and Oil and Gas with negative sign.



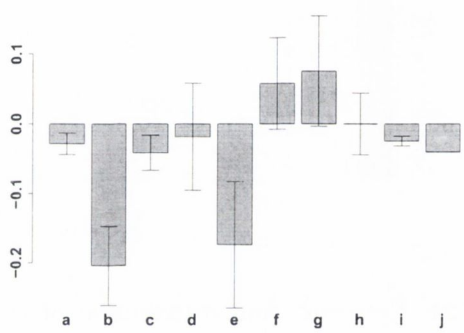
$\langle\mu^{85}\rangle_{FTSE100,s}$



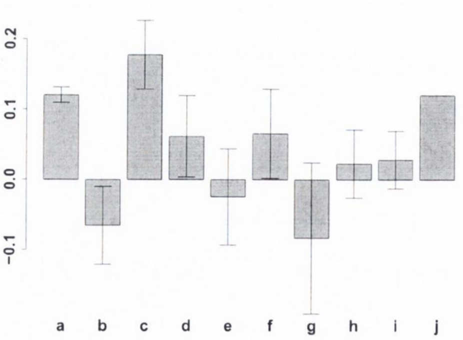
$\langle\mu^{84}\rangle_{FTSE100,s}$



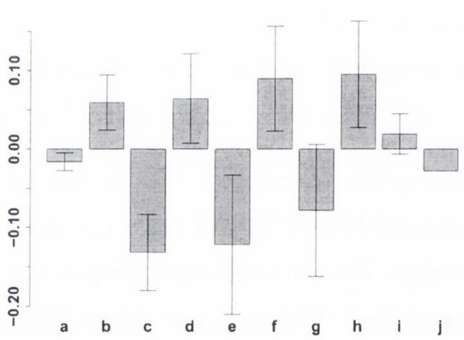
$\langle\mu^{83}\rangle_{FTSE100,s}$



$\langle\mu^{82}\rangle_{FTSE100,s}$



$\langle\mu^{81}\rangle_{FTSE100,s}$



$\langle\mu^{80}\rangle_{FTSE100,s}$

Figure 5.23: Mean value of eigenvector elements, for each industrial sector, of the six highest eigenvalues, λ_{85}, λ_{84}, λ_{83}, λ_{82}, λ_{81} and λ_{80} for a portfolio of 85 stocks from the FTSE100 index. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector.

5.7.3 Time analysis

We also performed a study in time for the correlations, MST and values of the eigenvalues. There are very surprising results for the curve of the mean correlation in time and the curve of the value of the highest eigenvalue also in time. We choose to perform the temporal analysis for different time windows length, T.

In Figure 5.24 we show the four moments of the coefficients of correlation matrix.

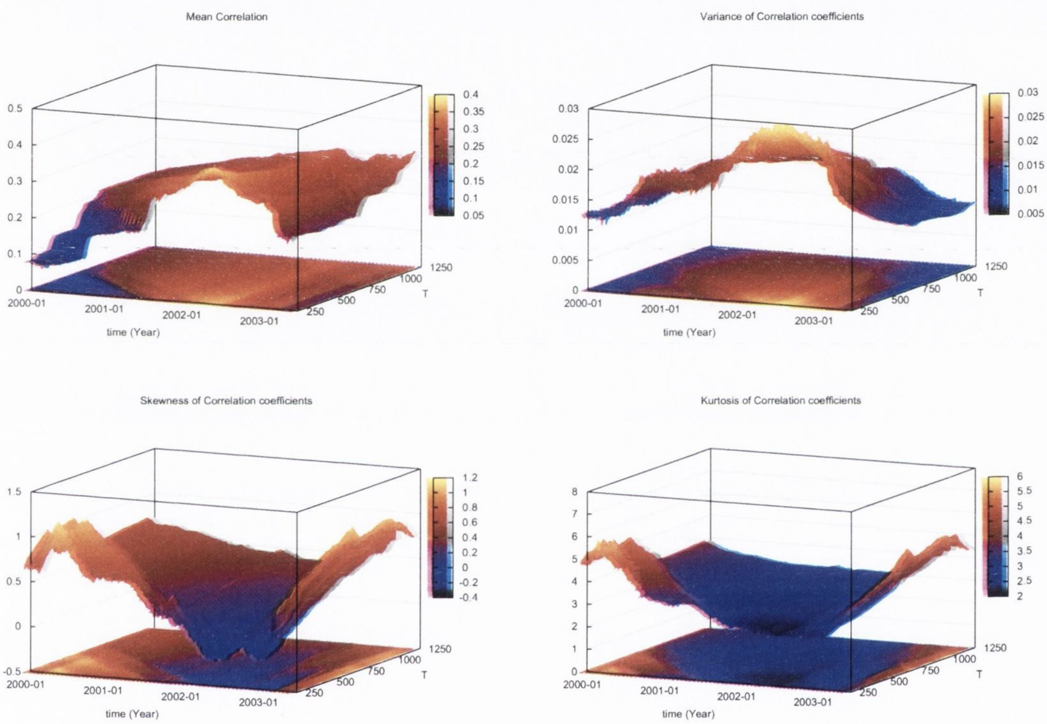


Figure 5.24: Mean (eq. 4.35), variance (eq. 4.36), skewness (eq. 4.37) and kurtosis (eq. 4.38) of the correlation coefficients. Time windows of length from 250 to 1250 days are moved with a window step length parameter of 1 day.

For each moment we can see the evolution in time for different sizes of the time window, T from 250 to 1250 days. For the mean correlation, the values are low around January 2000 and are high around the beginning of the year 2002. The same behaviour was stated for the analysis of the previous portfolio, where the mean correlation in time was shown in Figure 5.12. The variance of the correlation coefficients shows the same feature as the mean, showing some correlation in time between these two moments. The third moment, the skewness, shows the opposite behaviour of the mean correlation with high values in January 2000 and low values around the beginning of year 2002. Here we can see some anti-correlation between the first and the third moments. The same happen with the fourth moment, the kurtosis. The evolution of this moment in time shows high values in the beginning of 2000 and low values around year 2002. The same behaviour was shown in Figure 5.12 for these four moments.

In Figure 5.25 we present the four moments of the distribution of distances in the MST.

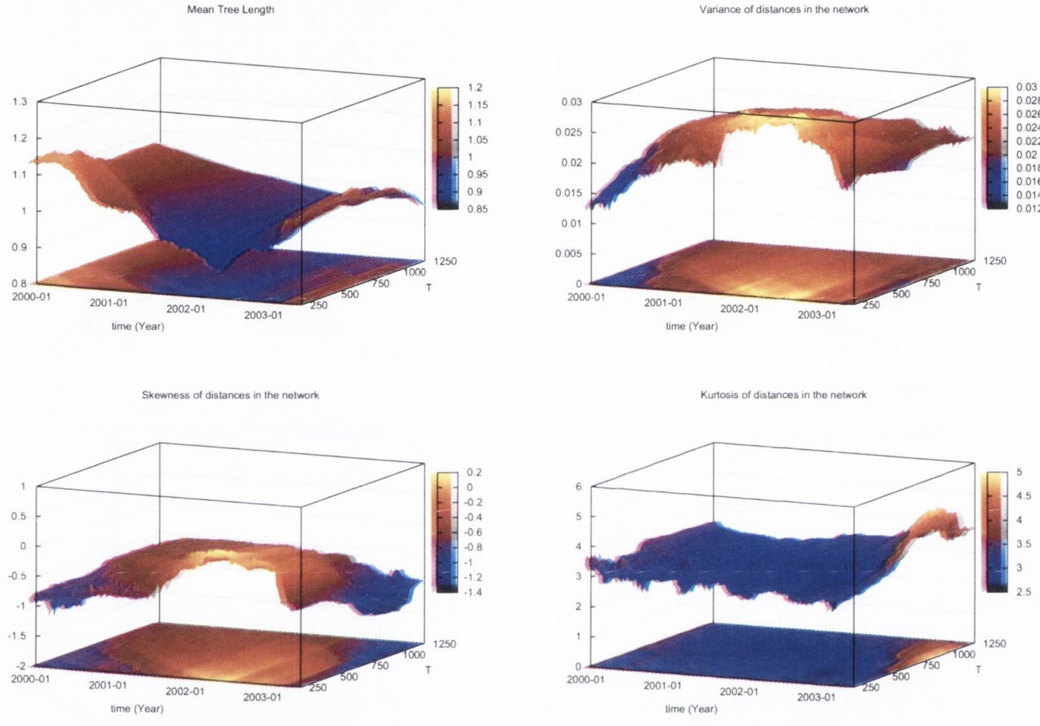


Figure 5.25: Mean (eq. 4.50), variance (eq. 4.51), skewness (eq. 4.52) and kurtosis (eq. 4.53) of the normalised tree length. Time windows of length from 250 to 1250 days are moved with a window step length parameter of 1 day.

For the mean tree length, the values are high around January 2000 and low around the beginning of year 2002, the opposite of what we saw for the mean correlations but the same behaviour that we saw for the previous portfolio in Figure 5.13. The variance of distances in the MST shows the same behaviour as the variance of the correlation coefficients of Figure 5.24. The third moment of this distribution, the skewness, shows low values for the beginning of January 2000 and high values around the year 2002, this is the same pattern shown for the variance of the distribution but it is the opposite of the skewness of the distribution of correlation coefficients. About the fourth moment, the kurtosis, there are not many conclusions that we can take from Figure 5.25.

Another important property of the MST is how this network changes its structure with time. Computing the mean occupation layer (MOL) in time (section 4.5.2) we can see if the MST shrinks or extends. Decreasing values of the MOL correspond to a shrinking of the MST, the stocks will be closer together. In Figure 5.26 we show the MOL in time for three different ways of calculating the centre of the MST (section 4.5.2). In a dynamic way we can choose the centre vertex as the one with maximum correlation value (top left picture of Figure 5.26) or the one with the maximum number of links (top right picture of Figure 5.26). We can also choose a static central vertex, for this case we choose the company Alliance Trust because it is the one with most links.

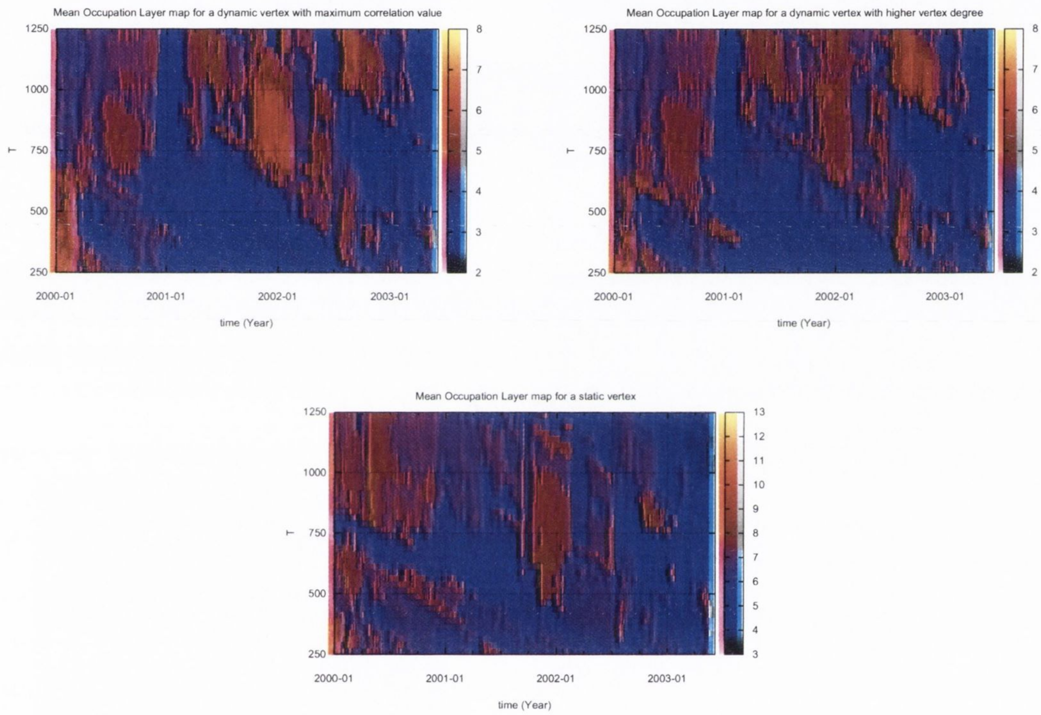


Figure 5.26: Contour map of the mean occupation layer as a function of time for three different methods of choosing the centre vertex: a dynamic central vertex based on maximum correlation value (top left), a dynamic central vertex based on maximum number of links (top right) and a static central vertex (bottom). Time windows of length from 250 to 1250 days are moved with a window step length parameter of 1 day.

For the three cases we can see that a very low value of the mean occupation layer occurs when the length of the time windows is small and around the year 2002.

Another important property of the MST is how the structure of the MST and the number of links survive from time step to time step (section 4.5.3). The single step survival ratio (SSSR) shows the percentage of links that are maintained from one step to the other. The multi step survival ratio (MSSR) shows the number of links that are maintained from the first MST until the last one. In Figure 5.27 we present the single and multi step survival ratio.

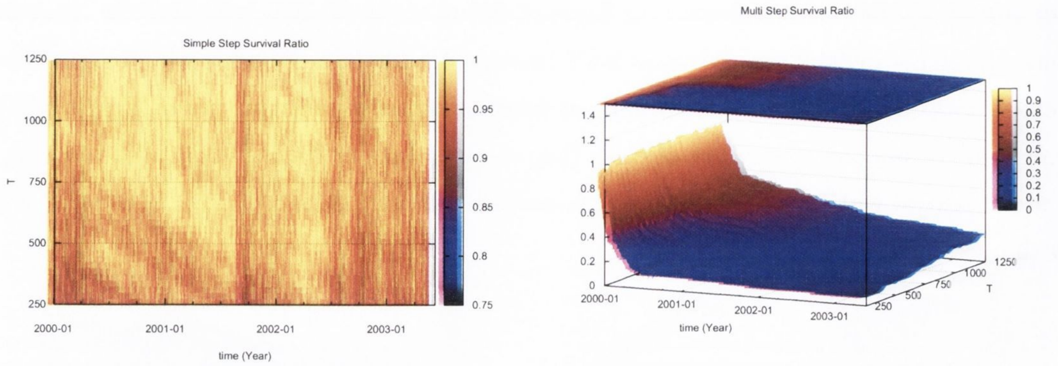


Figure 5.27: Single and multi step survival ratio as a function of time. The SSSR is represented in a contour map graphic. The MSSR is represented on the right. Time windows of length from 250 to 1250 days are moved with a window step length parameter of 1 day.

For the SSSR we can see that there are not many changes from time step to time step in the MST, but for the MSSR there are some changes. For the MSSR, in the first steps there are a lot of links that are missed, but after some time the links that remain will be there almost until the end. These are the strongest links in the MST.

We can also study the evolution of the value of the highest eigenvalue of the correlation matrix. In Figure 5.28 we plot the time analysis of the highest eigenvalue, and we can see that low values appear in the beginning of January 2000 and the high values appear around the year 2002. This is the same feature observed for the variation of mean correlation in Figure 5.24.

Comparing the contour map of the evolution of both quantities, the mean correlation and the highest eigenvalue, we can see that the two values have the same evolution over time. They are completely correlated with each other, as we can see in Figure 5.29.

5.8 Conclusions

In summary, we have studied the correlations between time series of log-returns of stocks from two FTSE100 portfolios and examined how these change with both the size of the time series and time. The mean correlation increases after external crises, and different moments feature correlations or anti-correlations as a result.

From the MST we can see that some stocks from the same industrial sector cluster together. This does not happen with all stocks from specific economic groups or industries. It would seem from the MST analysis that the new FTSE classification (ICB) introduced in January 2006 offers a more logical clustering of the different stocks as opposed to the previous classification scheme (GCS). However from the MST it is clear that anomalies are still present that could affect the

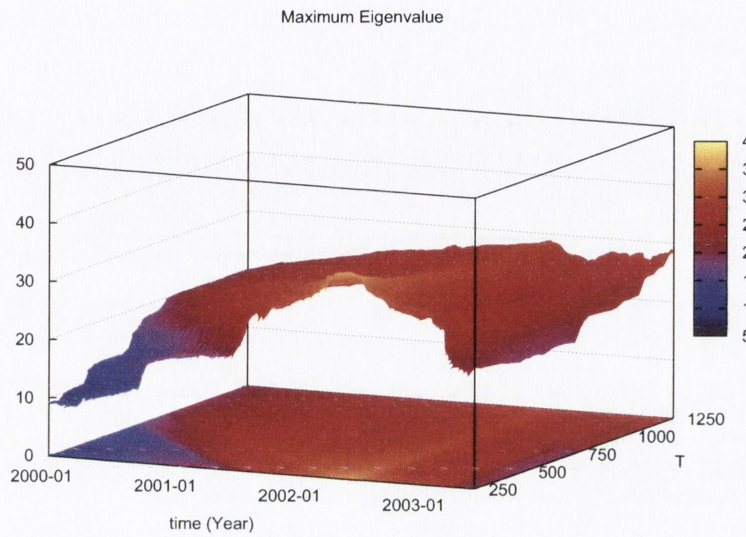


Figure 5.28: Highest eigenvalue of the correlation matrix in time. Time windows of length from 250 to 1250 days are moved with a window step length parameter of 1 day.

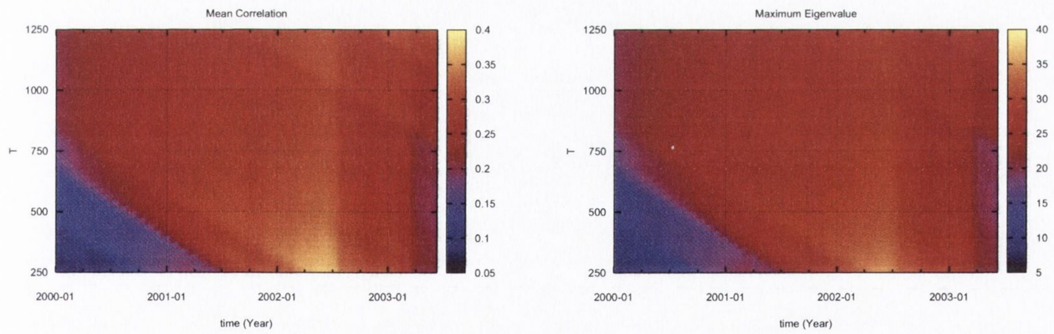


Figure 5.29: Contour map of the mean correlation (left) and highest eigenvalue (right) over time. Time windows of length from 250 to 1250 days are moved with a window step length parameter of 1 day.

building of optimum portfolios.

Studying the mean occupation layer and the single and multi step survival ratios, we conclude that there are some causal changes in the MST, but in the overall the MST maintained their structure in time.

The structure of trees generated from random time series differs significantly from real markets. Furthermore there appears to be no obvious hub node. On the other hand the one-factor model produces a MST where we can see hubs with many links. This kind of structure is close to that obtained using intra-day data. The MST created with random time series with T-student stochastic variables are more similar with the MST of real data than the ones created with Gaussian stochastic variables.

The eigensystem analysis shows that the eigenvector related with the highest eigenvalue has all its elements positive. Some eigenvectors related with other highest eigenvalues show that a segregation between stocks from different industrial sectors occur. This fact is in agreement with the conclusions from the MST analysis.

The conclusions from this chapter, about the segregation of stocks in terms of industrial sectors, can help investors in the optimisation of portfolios taking in consideration the diversity of stocks from different industrial sectors.

Chapter 6

Indices of markets around the world

6.1 Introduction

This chapter examines the extent and evolution of interdependence between world equity markets [52] over a 10-year period using the Minimum Spanning Tree (MST) approach of Mantegna [43] that we detailed in Chapter 4.

To our knowledge only one study has been published to date applying the MST approach to groups of national equity markets [124]. There is one other earlier study [125], based on numerical taxonomy analysis of weekly stock market index returns, for a period between 1962 and 1973, of 12 major international equity markets: Australia, Austria, Belgium, Canada, France, Italy, Japan, Netherlands, Switzerland, U.K., West Germany and U.S.A. Using hierarchical clustering techniques, similar with the MST approach, but where they used dendograms instead of representing the similarities of time series in a network they conclude that a core of international markets have higher degrees of similarity than others: U.S.A., Canada, Netherlands, Switzerland, West Germany and Belgium, i.e. these indices have higher correlation coefficients between them.

The study that applied the MST approach [124] is a simple dynamic analysis based on partially overlapping windows of indices for 20 countries for the years 1988-1996 and finds that markets group according to a geographical principle, as is also the case for an overall average examination of 51 world indices for the years 1996-1999 in the same study [124]. The temporal evolution showed the stability of the North-America cluster and the increase in the size of the European and Asian-Pacific clusters, in time. Our research significantly extends this work by applying dynamic MST methods to examine the time-varying behaviour of global equity market co-movements for a group of 53 developed, emerging and developing countries over the years 1997-2006. This period includes major market events such as the Asian and Russian economic crises, the introduction of the euro, and the enlargement of the European Union (EU). In addition to confirming the earlier evidence of a geographical organising principle we document a tendency of the MST toward higher density over time, indicating an increasing degree of integration of international equity markets. Such a finding is of interest to portfolio managers and investors, as the implication is of decreased potential for diversification benefits and thus perhaps decreased returns for international investors.

After the conclusions in the previous Chapter 5, that the stocks from the FTSE100 index grouped in terms of industrial sector, the main reason for the study presented in this Chapter is to get an insight in how the indices cluster in a MST. After this study we will analyse portfolios of stocks from the indices study here and we will try to understand the behaviour behind the interdependence in stocks from different markets.

6.2 Data

We analyse the returns of the equity markets of 53 countries. The data consists of Morgan Stanley Capital International (MSCI) daily closing price indices for 44 countries, for the period January 8^{th} 1997 to February 1^{st} 2006. An additional nine countries, which data is not available from the MSCI indices, are also included in the sample, resulting in a total of 53 indices. These countries and indices are: Croatia (Nomura), the Czech Republic (PX 50), Hungary (BUX), Iceland (ICEX 15 Cap), Lithuania (Nomura), Malta (HSBC Bank), Romania (Nomura), Slovakia (SAX) and Slovenia (HSBC Bank). We included these indices to have a better view about the correlations between indices of the East European markets and the relation of these indices with other indices around the world. All series are expressed in US dollar terms as the reference currency, thus reflecting the perspective of an international investor. All data are sourced from DataStream, Thomson Financial [107]. One issue that needs to be addressed is the non-synchronous nature of the data, i.e. the fact that equity markets open at different times. Recent research suggests that the use of daily data may lead to significant underestimation of equity market integration [126] because the time-mismatch have to be handle carefully. The use of weekly returns can help with the problem of time-mismatch but it leads to significant loss of information. As a consequence, to minimise the problem of non-synchronous trading the daily index level data were converted to weekly (Wednesday) returns for this portfolio and we maintained the daily returns for a second portfolio that we will present further. The resulting number of weekly observations is 475. The 53 countries in our study and the respective symbols are given in Table 6.1.

The reliance for the most part on MSCI indices allows for significant confidence in the findings, as these indices are designed explicitly to allow for cross market consideration of returns by investors. By contrast, studies that rely on indices from the individual equity markets, indices

Table 6.1: Countries and respective symbol.

Symbol	Country	Symbol	Country	Symbol	Country
ARG	Argentina	HUN	Hungary	PHI	Philippines
AUS	Australia	ICE	Iceland	POL	Poland
AUT	Austria	IDO	Indonesia	PRT	Portugal
BEL	Belgium	IND	India	ROM	Romania
BRZ	Brazil	IRE	Ireland	RUS	Russia
CAN	Canada	ISR	Israel	SAF	South Africa
CHF	Switzerland	ITA	Italy	SGP	Singapore
CHL	Chile	JAP	Japan	SOK	South Korea
COL	Colombia	JOR	Jordan	SVK	Slovakia
CRT	Croatia	LTU	Lithuania	SVN	Slovenia
CZK	Czech Republic	MAL	Malaysia	SWE	Sweden
DNK	Denmark	MEX	Mexico	THI	Thailand
ESP	Spain	MTA	Malta	TUK	Turkey
FIN	Finland	NEZ	New Zealand	TWA	Taiwan
FRA	France	NLD	Netherlands	UK	United Kingdom
GER	Germany	NOR	Norway	USA	United States
GRC	Greece	PAK	Pakistan	VEZ	Venezuela
HK	Hong Kong	PER	Peru		

such as the NIKKEI225, the DJIA or the FTSE100, run the risk of non-comparability due to differences in construction, coverage and completeness. We will illustrate this difference in the results in section 6.4 when we use the indices for each country constructed at each national stock exchange, instead of an index constructed by MSCI and also we use daily returns instead of the weekly returns in this case.

The differences in construction of an index can be related with the weighted given to each stock that belongs to it. For example, if the index is constructed with an arithmetic mean of the values of the price of the stocks, it will be given as follow:

$$I(t) = \sum_{i=1}^{N} P_i(t) \tag{6.1}$$

but if instead of an arithmetic mean, the construction is made with a weighted arithmetic mean, we have to take into account for example the volume of transitions of each stock:

$$I(t) = \sum_{i=1}^{N} w_i P_i(t) \tag{6.2}$$

Most of the indices follow these two ways of construction.

6.3 Results

We present the findings of our MST analysis in two sub-sections. We first show the overall average MST, derived from an analysis of the entire sample of data. Following that, a number of dynamic approaches are applied.

6.3.1 Analysis of MST constructed from averaging data from 1997-2006

Shown in Figure 6.1 is the average MST for the 1997-2006 period. The clusters which we observe appear to be organised principally according to a geographical criterion (possibly also reflecting political and trade criteria). This is similar to the results in [124]. To analyse the graph we identify a "central" node, the market most strongly connected to its nearest neighbours in the tree. With the highest number of linkages, France can be considered the central node. Surprisingly, the U.S.A., whose equity market is globally dominant in terms of market value, exhibits a somewhat looser linkage to the other markets. Closely connected to France are a number of the more developed European countries in the European Monetary Union (EMU) and in the EU. This European grouping forms a set of markets that are highly correlated with each other, with France at its centre. We can also identify several "branches" which form the major subsets of the MST and these can then be broken down into "clusters" that may or not be completely homogeneous. The Netherlands heads a branch that includes clusters of additional European countries (along with Jordan, anomalously). The U.S.A. links a cluster of North and South American countries, except for Peru, to France via Germany. Not surprisingly, the three members of the North American Free Trade Association (NAFTA) - the U.S.A., Canada and Mexico - are directly connected, with Mexico forming the link to the South American countries. Australia heads a branch with several groupings: all the Asian-Pacific countries form two clusters, one of more developed and the other of less advanced countries; most of the Central and East European (CEE) countries, that joined the EU in 2004, form an incomplete link to Australia through South Africa, along with Turkey and Peru. Jordan, which appears in a European clustering, is an apparent anomaly. This is likely due to the fact that Jordan is the last node connected to the network and has correlations with other countries close to zero, which means a relatively high distance. We can conclude that Jordan is an outlier of our study that does not have any close relation to any of the other countries represented here. The reason why Jordan is connected with Norway and not another country may reflect the fact that many companies in Norway belong to the Oil and Gas industrial sector.

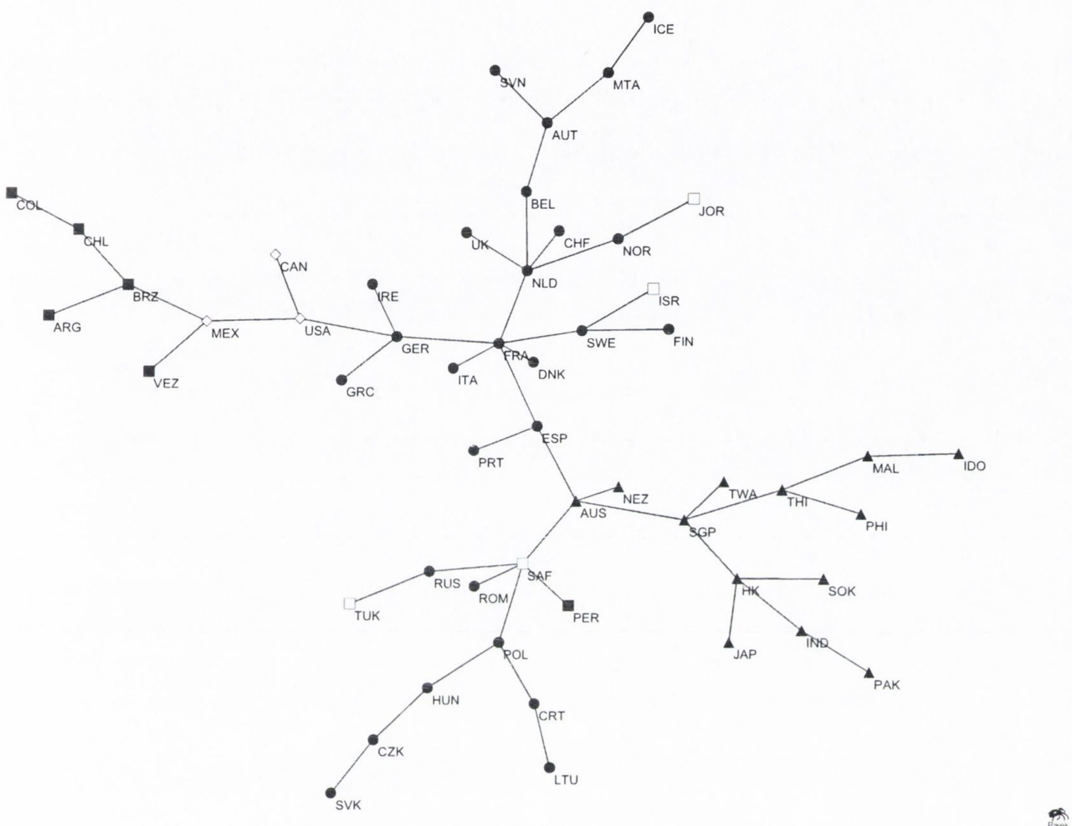


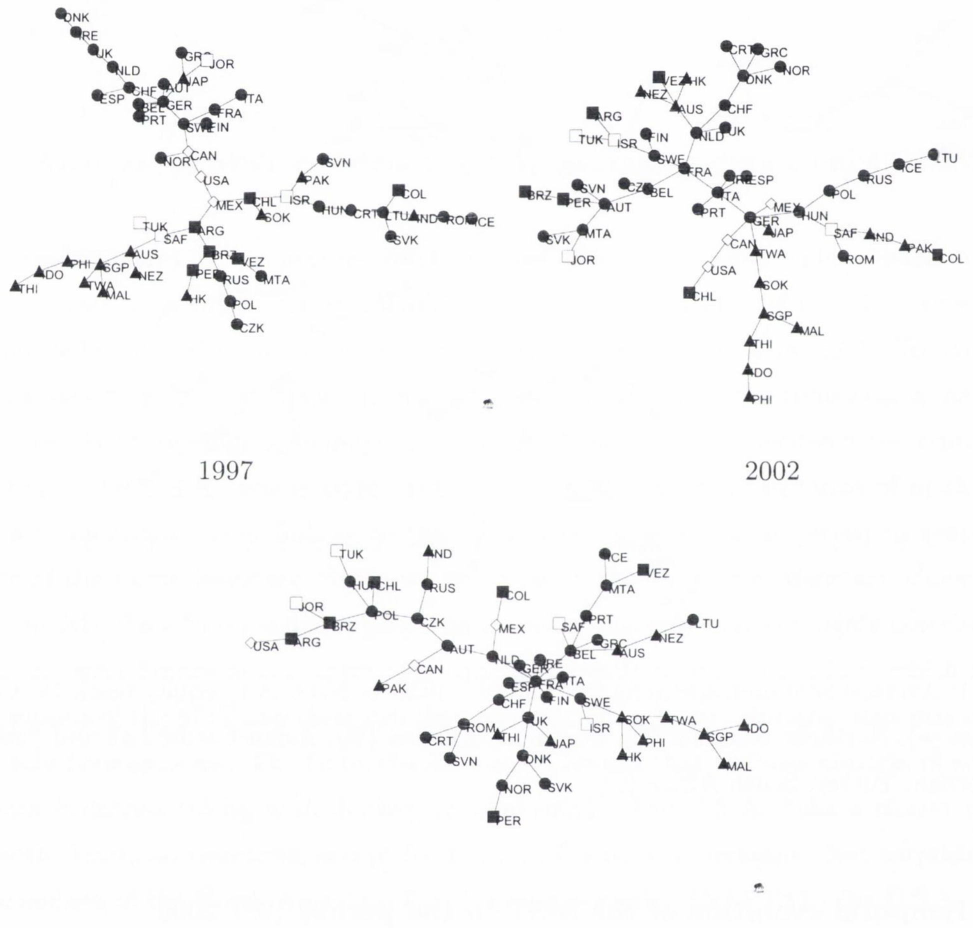
Figure 6.1: Average minimum spanning tree for 1997-2006 for 53 country equity markets. Coding is: Europe (•), Northern America (◇), Southern America (■), Asian-Pacific (▲) and "other" - Israel, Jordan, Turkey, South Africa (□).

6.3.2 Temporal evolution of the MST in the period 1997-2006

The MST of Figure 6.1 presents a static analysis of the relationships between the markets for the time period as a whole. It is possible, however, also to examine the time-dependent properties of the tree to provide insights on the changing relationships between the equity markets over time. To this end several techniques are used. First, we construct what we call a rolling and recursive MST. Second, we show the evolution of the four moments of the mean correlations and mean tree lengths of the MST (sections 4.3 and 4.5.1). Third, calculation of the mean occupation layer (section 4.5.2) reveals changes in the compactness of the MST over time, the degree of compactness being interpretable as the extent of overall equity market integration. Finally, the single-step and multi-step survival ratios (section 4.5.3) for market linkages provide an indication of the stability of linkages between markets over time.

Rolling MST

The dynamic evolution of the MST can be examined by looking at a series of MSTs created from non-overlapping rolling windows, each with width 1 year, or 52 (53) weeks. The MST shown in Figure 6.2 are those for 1997, 2002, and 2005.



1997 2002

2005

Figure 6.2: Rolling one-year window MST for 1997, 2002 and 2005. Coding is: Europe (•), Northern America (◇), Southern America (■), Asian-Pacific (▲) and "other" - Israel, Jordan, Turkey, South Africa (□).

We detect several consistent relationships as well as a number of less stable arrangements. One clear consistency is that the developed European countries form the central structure of the MST. Initially, Germany is the central node; however, in more recent years France has taken over this role. The CEE countries do not form a single cluster but tend to fragment into several subgroups, with changing composition year by year. However, perhaps reflecting the growing economic and political ties with the developed EU members, they tend to move slightly closer to those countries over time in terms of levels away from the central node.

With respect to the Asian markets there is usually a link between Australia and New Zealand, which often head a branch connecting most of the remaining Asian markets to Europe. The coherence of the Asian countries is particularly evident in 1998, possibly reflecting increased correlations in the region in the aftermath of the Asian crisis. This particular clustering does not continue as strongly in subsequent years. The main exception in this group is Japan, which does not fit into the Asian cluster but is generally linked directly to Western markets.

For the North American markets the U.S.A., Canada, and Mexico are usually closely linked, reflecting most likely the ongoing effects of both geography and NAFTA trade ties. An apparent exception is the year 2005. An explanation of the disconnect in this cluster lies in examining the construction of the MST. In 2005 relatively higher correlations between European countries almost completely dominated the formation of the MST as a cluster first formed around France, followed by a group of CEE countries (the Czech Republic, Hungary, and Poland).

The South American markets have not formed a complete cluster in any of the years under examination; however, a sub-cluster of Argentina, Brazil, and Chile, the largest, most developed and most liquid Latin American markets, can occasionally be observed. This cluster is usually directly linked to the European grouping via Mexico.

Recursive MST

To further examine the stability of the relationships we constructed recursive MST by forming the MST for the first year and then successively adding one year's data at a time. These are shown in Figure 6.3 cumulatively through 1998, cumulatively through 2001 and cumulatively through 2005.

The first issue that emerge from this analysis is that the MST appears to have become somewhat more compact in comparison to the rolling window MST for 1997 in Figure 6.2. In the 1997 rolling window, the maximum number of levels was twelve (central node Germany to Iceland), while it is consistently smaller in the recursive MST beginning with 1998. For the period 1997-2005, it is seven (central node France to Jordan).

Not surprisingly, given the results from the rolling MST, the recursive graphs also reflect the dominance of the developed European grouping and the shift of central node to France from Germany, with Australia, the Netherlands, and the U.S.A. at the head of the branches. The German-U.S.A. link persists, even as the centre of the European cluster shifts to France. The CEE countries continue to reflect some tendency to split into different clusters, although by 2002 six of them have settled into one group, leaving out only Russia, Slovenia, and Romania. Hungary and Poland, among the more developed CEE equity markets, alternate the role of node

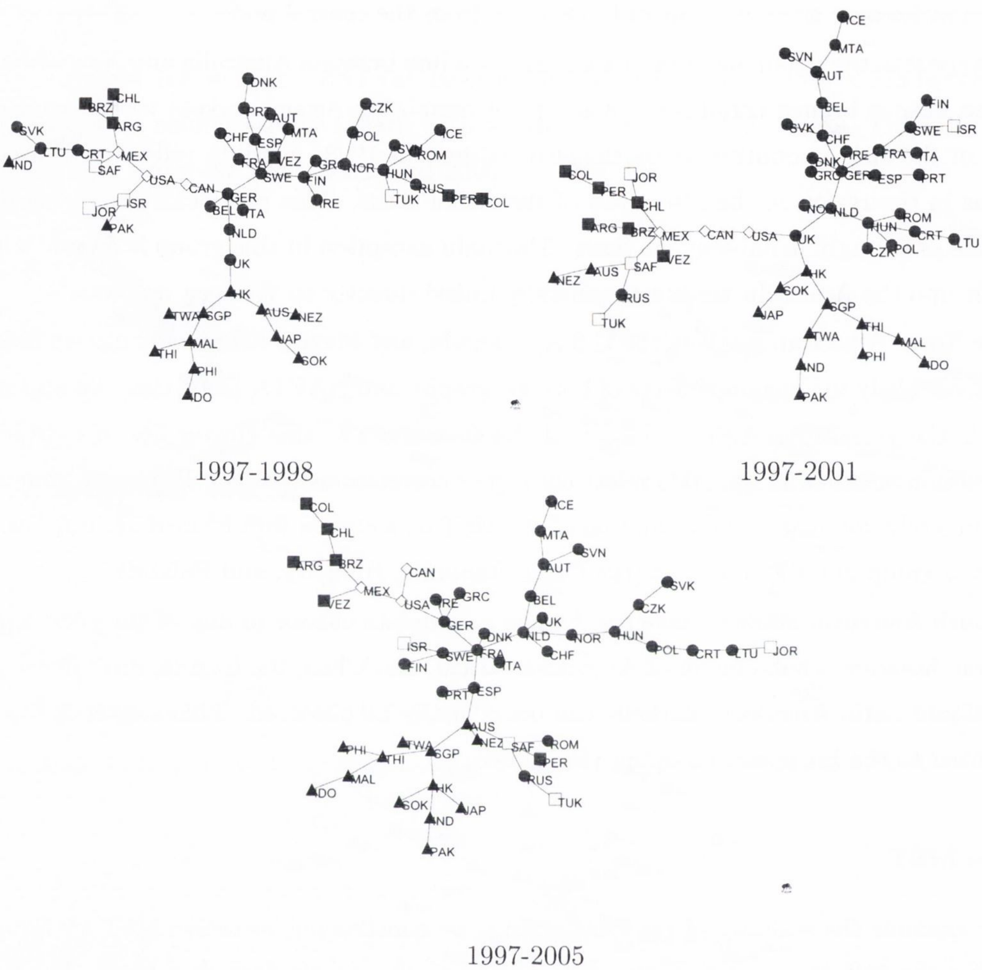


Figure 6.3: Recursive MST for cumulative through 1998, cumulative through 2001 and cumulative through 2005. Coding is: Europe (•), Northern America (◇), Southern America (■), Asian-Pacific (▲) and "other" - Israel, Jordan, Turkey, South Africa (□).

linking the CEE countries to the developed EU members. This cluster also gradually moves to a closer attachment to France as that country becomes the central node.

The clustering of the Latin American markets, except for Peru, becomes more consistent as the time period is increased in the recursive graphs, with Mexico generally the link to the European core. Similarly, as the time period increases a more consistent pattern is established for the Asian markets. Pakistan and India join the other Asian countries in 2001, and Australia and New Zealand in 2002. Japan, whose behaviour year by year appeared to be largely disconnected from the other Asian markets, is now seen as tied into the Asian cluster via Hong Kong or South Korea consistently since 2001. Finally, a Turkey-Russia-South Africa cluster emerges in 2000 and stays reasonably stable.

Correlation and mean tree length analysis

The temporal analysis of changes in the first four moments of the mean correlations, ρ_{ij} (eq. 4.30) and of the distances in the MST, d_{ij} (eq. 4.49) are presented in Figures 6.4 and 6.5, where the window length is 52 weeks and the window step length is 1 week.

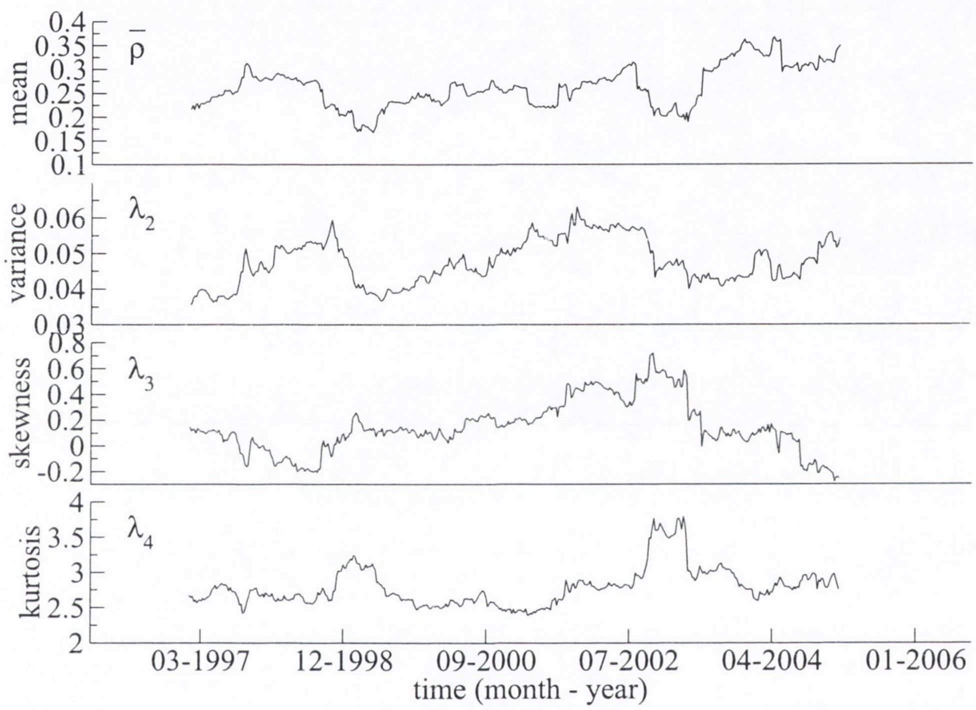


Figure 6.4: Mean (eq. 4.35), variance (eq. 4.36), skewness (eq. 4.37) and kurtosis (eq. 4.38) of the correlation coefficients. Time windows of length 52 weeks are moved with a window step length parameter of 1 week. Results are plotted according to start date of window.

The mean correlation and its variance increase over the initial period covered by the data, roughly corresponding to the era of the Asian and Russian crises. In times of market uncertainty and downturns these measures tend to increase [48, 51, 127]. The tendency of these measures to increase together has significant implications for standard econometric methodology [128], i.e. these cases show that dramatic movements in one market can have strong implications in other markets with different sizes and structures around the world. We also note that the skewness decreases toward zero, implying that the distribution of the correlations becomes more normal. A similar pattern has been observed for London stocks in the FTSE index [51] (Figure 5.12 of section 5.5). This initial period is followed by declining correlations as global markets move past the 1997-1998 crisis events. Correlations rise again, however, possibly reflecting the

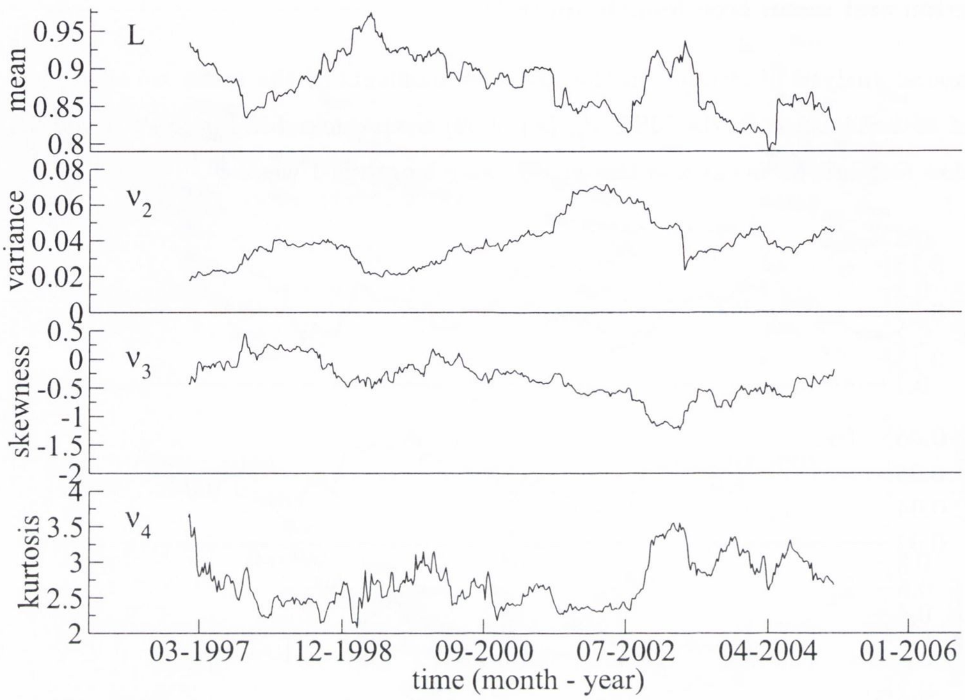


Figure 6.5: Mean (eq. 4.35), variance (eq. 4.36), skewness (eq. 4.37) and kurtosis (eq. 4.38) of the normalised tree length. Time windows of length 52 weeks are moved with a window step length parameter of 1 week. Results are plotted according to start date of window.

broad market declines that begin in 2000. An upward spike occurs in the fall of 2001, which corresponds to the entry into the rolling window of the steepest downturns of global markets as measured by the MSCI world index. Recovery is accompanied, once again, by declining mean correlations. A second, larger, upward spike is observed as the window begins to include the early 2004 period, which occurs in the context of a broader trend toward higher correlations. The mean correlation, for the year ending May 13^{th} 2004, is 0.24223 while that for the year ending May 20^{th} 2004, is 0.30222. Subsequent correlations remain relatively high. Interestingly, this spike coincides with the entry of new members into the European Union (EU) on May 1^{st} 2004. A breakdown of rolling correlations shows a strong, abrupt increase in correlations for the European group of countries at this point, as well as a consistent tendency over the entire time period for their correlations to be higher than for the set of 53 markets as a whole. This event has introduced a new element of uncertainty as well as the prospects for closer economic ties, both of which could tend to increase correlations. In contrast to these larger movements the introduction of the euro on January 1^{st} 1999, was not accompanied by major changes in correlation structure.

Essentially the same information provided by the correlation matrix can be obtained also from the moments of the normalised tree length, as shown in Figure 6.5. The mean distance is negatively correlated with the mean correlations, tending to fall, for example, in times of market crisis. This underlines the ability of the MST as a strongly reduced representative of the entire correlation matrix to convey relevant market information. Overall, the mean distance shows a tendency to decrease over the ten years, indicating a "tighter" composition of the MST.

Mean occupation layer

The structure of the MST can also be analysed by the distance of the nodes to the central node, the mean occupation layer (equation 4.54). Using the two definitions of a central node (sub-section 4.5.2), we identify the central node for our rolling MST. The two criteria produce similar results. Germany is the central node in the early years, but France takes its place for most of the subsequent periods. Using the highest number of links criterion, France is the central node 41.5% of the time and Germany 27.3%. The highest correlation sum criterion identifies France as the central node 53.8% of the time and Germany 30.2%. Other countries occasionally assume the position of central node.

The mean occupation layer can then be calculated using either a fixed central node for all windows, i.e., France, or a continuously updated node. In Figure 6.6 the results are shown for France as the fixed central node (black line), the dynamic maximum vertex degree node (dotted line) and the dynamic highest correlation vertex (grey line). The three sets of calculations are roughly consistent. The mean occupation layer fluctuates over time as changes in MST occur due to market forces. There is, however, a broad downward trend in the mean occupation layer, indicating that the MST over time is becoming more compact.

Single and Multi Step Survival Rates

The robustness of links over time is the final analysis that we did for our MST (sub-section 4.5.3). Figure 6.7 presents the single-step survival ratios for the MST (equation 4.55). The average is about 0.85, indicating that a large majority of links between markets survives from one window to the next. As might be expected, the ratio increases with increases in window length.

Figure 6.8 shows the multi-step survival ratio. In both cases the length of time series was 52 weeks and the window step length 1 week. Here, as might be expected, the connections disappear quite rapidly, but a small proportion of links remains intact, creating a stable base for construction of the MST. Again the evidence here is of importance for the construction of portfolios, indicating that while most linkages disappear in the relatively short to medium term

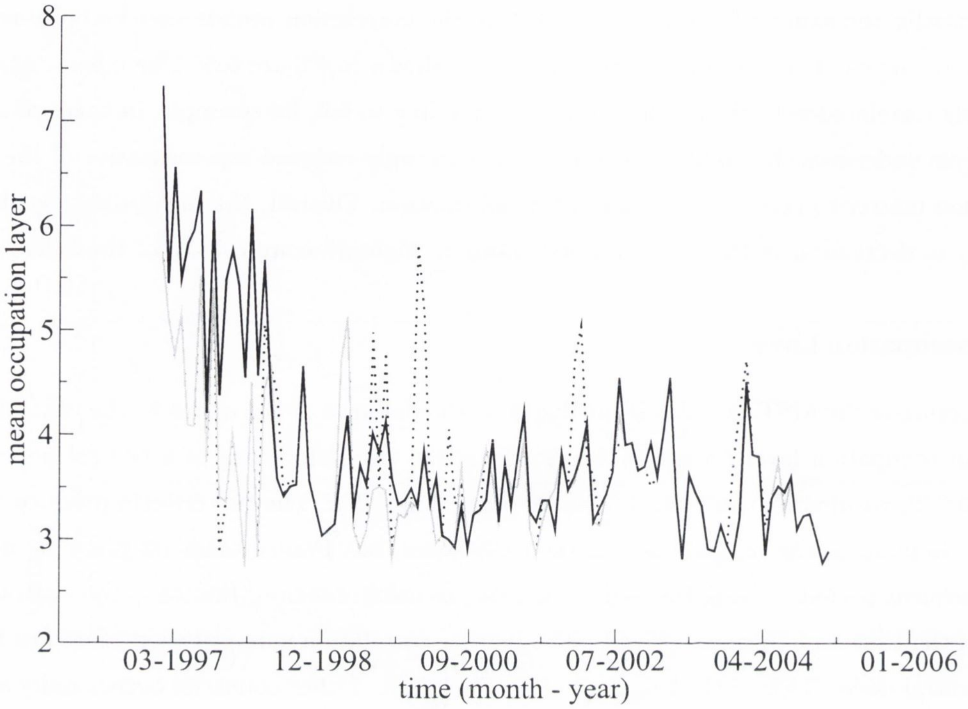


Figure 6.6: Plot of mean occupation layer as a function of time for time windows of length 52 weeks and window step length of 4 weeks). The black line is for a static central vertex (France), the dotted line uses dynamic central vertex based on maximum number of links, while the grey line shows dynamic central vertex based on maximum correlation value.

there are islands of stability where the dynamics are consistent.

The behaviour of these two measures is similar to what has been observed for individual stocks within a single equity market [48]. These results may understate the stability of the global system of markets since some of the linkage shifts appear to take place within relatively coherent geographical groups.

6.4 Stock Exchange Indices for different countries

As stated before there is differences between the study of correlations between the indices constructed from MSCI data and the indices constructed from individual stock market data. First the indices constructed from MSCI data as explained in their website [129] are weighted indices designed to measure the equity market performance of developed and emerging markets. The other indices from individual stock markets have a different method of construction for each stock market, some times they are calculated as weighted indices, other times they are an

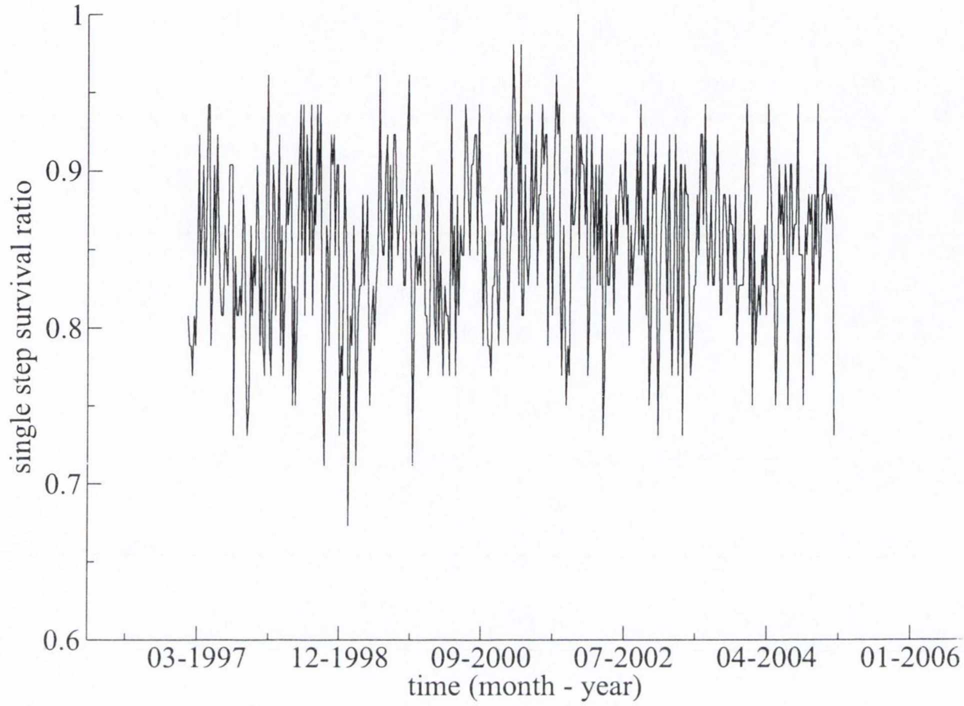


Figure 6.7: Single-step survival ratio as a function of time. Time windows of length 52 weeks are moved with a window step length parameter of 1 week.

average of the prices of each stock that belongs to the index.

In this section we use the main indices of some countries around the world. The whole portfolio is composed of 71 different countries with one index each. In Appendix D, the tables show the structure of this portfolio of indices and how we divided them in different groups. This data was downloaded from the Datastream server [107].

All the prices of the indices are changed to dollars according to the currency exchange at specific day and we choose time series of daily closing price from January 3^{rd} 2000 until March 24^{th} 2008. In this period the total number of indices with non-zero values is equal to 56. In Figure 6.9 we show the MST for this portfolio of indices.

As in Figure 6.1 the main hub of the MST is the index of France (CAC40). Around this main hub there are 9 links, seven of them to another European index and two of them to an American index (IGPA from Chile) and to a Middle East index (MAOF25 from Israel). The North American cluster of Canada and U.S.A. is linked through Germany to this main hub.

In this MST of Figure 6.9 there is also a second main hub represented by the Austrian index (ATX Prime) with 8 links to other European indices. This hub is of extreme importance to the East European indices because it links many of the East European indices to the main hub

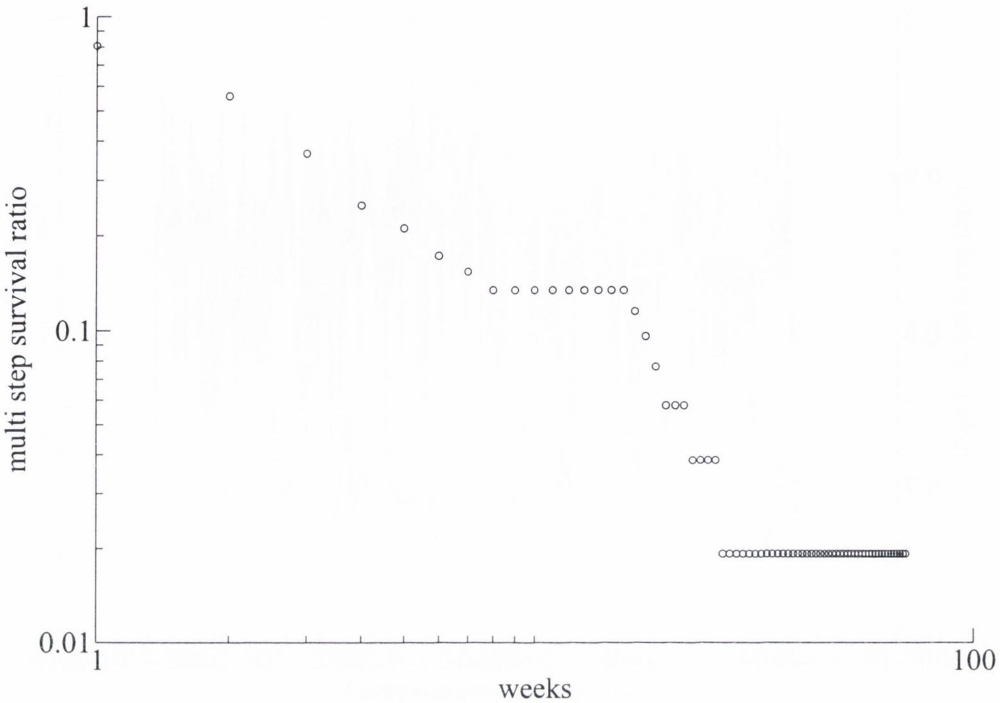


Figure 6.8: Multi-step survival ratio as a function of the number of weeks in a log-log scale. Time windows of length 52 weeks are moved with a window step length parameter of 1 week.

of France. The inclusion of some Middle East indices shows a small cluster of these countries with the indices of Egypt, Kuwait, Jordan and Oman linked together. The Asian-Pacific cluster maintain the same structure shown in Figure 6.1.

From these results we can conclude that the use of MSCI indices or indices from each stock market have similar results in terms of the correlations between the indices. Also when we use daily returns instead of weekly returns the correlations between indices maintain their structure and there seems to be no problem related with time-mismatch. The second portfolio of indices can be more useful for further work because we know each of the stocks that belong to the indices, so when we study the correlations of stocks from different indices in Chapters 7 and 8 we should compare with the MST of the second portfolio (Figure 6.9) and not to the MST of the first one (Figure 6.1).

6.5 Conclusions

The use of the MST provides a way to extract a manageable amount of information from a large correlation matrix of global stock returns to reveal patterns of links between different markets.

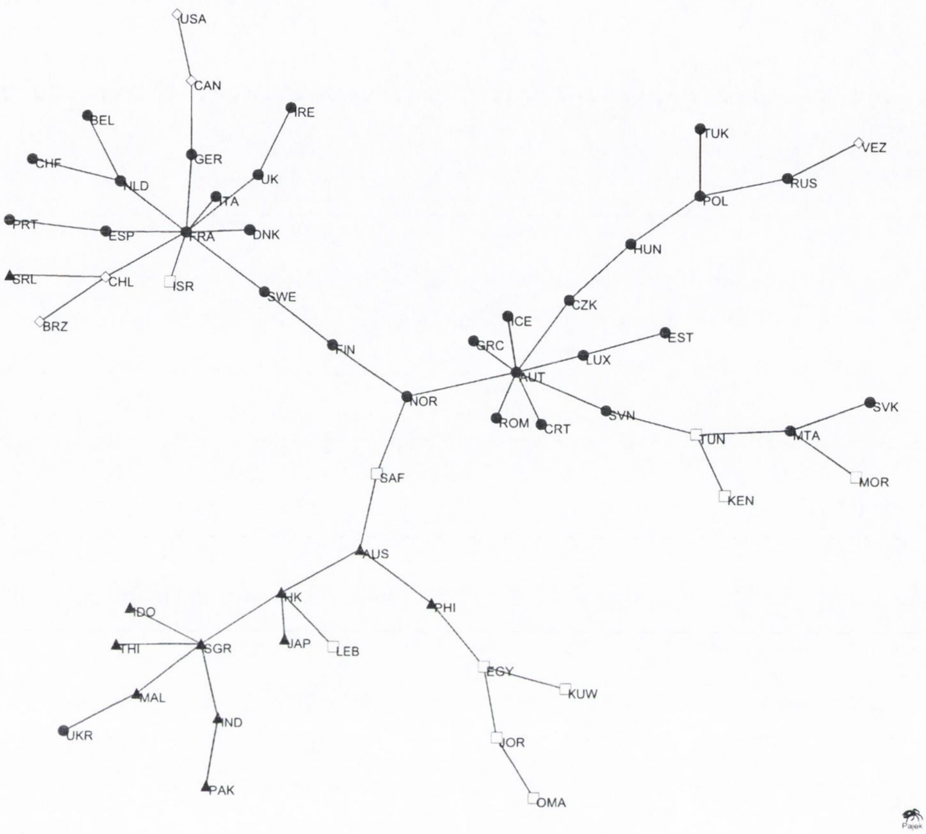


Figure 6.9: Minimum spanning tree for 56 different indices of different countries. The time series are from January 3^{rd} 2000 until March 24^{th} 2008, which means 2146 days. Coding is: European indices ($\bullet$), American indices ($\Diamond$), Asian-Pacific indices ($\blacktriangle$) and African and Middle East indices ($\Box$).

It provides an insight into market behaviour that is not as easily obtained from the correlation matrix as a whole. Applied dynamically, the analysis lets us observe consistencies as well as evolutions in patterns of market interactions over time. As would be expected, there is a strong tendency for markets to organise by geographical location, although other, related factors such as economic ties, may also play important roles. Developed European countries, with France and Germany at their centre, have consistently constituted the most tightly linked markets within the MST. There has also been a limited tendency of the CEE accession countries to link more closely with the more developed EU countries.

The study of 56 different indices around the world showed similar results with the study for the 53 MSCI indices. The introduction of some Middle East indices showed a cluster of indices with Jordan included, which was a kind of an outlier in the previous study. The Austrian index, ATX Prime is the second hub of the MST with a strong linkage with the Eastern European indices.

We have seen that the mean correlations show a tendency to increase over the period as a whole, while mean distances in the MST and the mean occupation layers have been trending downward. These dynamic measures point to a compression of the MST over time, meaning a tighter degree of interaction, or integration, between markets. These findings have implications for the international investor. International diversification under standard Markowitz portfolio construction relies on the existence of a set of assets which display consistent and persistent differences in correlations. These correlations form the basis of the MST. From a Markowitz portfolio perspective [130], or any portfolio perspective which relies on a spread of (relatively low) correlations, the compression which we have observed implies reduced diversification benefits over the time period we have examined. Finally, the multi-step survival ratio also indicates that while clusters of any given period may be homogeneous, the likelihood of these remaining stable over a reasonable portfolio period is small. This points to a need for frequent restructuring to make maximum use of diversification benefits.

Chapter 7

Cross correlations between stocks from DJIA and FTSE100 indices

7.1 Introduction

While Chapter 5 addressed the issue of correlations in stock prices in the same market, in the current chapter we analyse correlations of stocks traded in two different markets. In particular we will find that the clustering of stocks in a combined MST is primarily by market.

We computed correlations between the main stocks on the London Stock Exchange main index (the FTSE100) and the main stocks on the Dow Jones Industrial Average index (the DJIA). We selected these sets of stocks because they belong to the main indices used by investors and are also classified by the same industrial classification, namely Industry Classification Benchmark (ICB) [121]. The DJIA is the oldest continuing U.S. market index and comprises 30 of the largest companies in U.S. [131] The companies are chosen by the editors of The Wall Street Journal [132]. It is called an average because it originally was computed by adding up stock prices and dividing by the total number of stocks [131].

From our previous study of Chapter 6, we concluded that the indices grouped in terms of geographical location. In Figure 6.9 we could see that these two main indices (FTSE100 and DJIA) are not linked together. One reason for this can be the time-mismatch. In this chapter we study the correlation when we use the same day for both markets and also with one day delay for the stocks that belong to one of the markets, and we conclude that this artificial feature doesn't change anything in the results of both Random Matrix analysis and Minimal Spanning Tree which show that stocks from the DJIA and the FTSE100 remain separated.

The results for the portfolio of 30 stocks from the DJIA are shown in section 7.3. Section 7.4 shows the results for the cross-correlations between stocks from the DJIA and FTSE100 indices.

Finally the conclusions are presented in section 7.5.

7.2 Data

In this portfolio combined of stocks from DJIA and FTSE100, all the companies listed in one index are different from the companies listed in the other. This may be a cause for the segregation of stocks. In the next chapter 8 we will study portfolios with the same company listed in different indices.

The total number of stocks in this portfolio is 115 (the total 30 stocks from DJIA and 85 stocks from the 102 stocks of the FTSE100 index) starting from January 3^{rd} 2000 until March 24^{th} 2008. The stocks from FTSE100 and DJIA indices used in this study are shown in tables C.1 and C.2 of Appendix C, respectively.

The total number of days is 2146. The set of stocks chosen are the stocks that belong to these indices at April 1^{st} 2008. The data from the 17 stocks that we excluded from the FTSE100 index did not have non-zero closing prices for the entire period of our study. The main reason for this is probably that the companies enter the stock market after January 3^{rd} 2000.

7.3 Analysis of correlations of stocks in the DJIA index

The distributions of the eigenvalues of the correlation matrix for the 30 stocks of the DJIA index is shown in Figure 7.1 and can be compared with the results for 85 stocks from the FTSE100 index in Figure 5.22 (Chapter 5). The value of the largest eigenvalue for each market seems to depend on the size of the portfolio or probably in the correlation of the stocks in the portfolio.

Figure 7.1 shows that most of the eigenvalues are located outside the region predicted by Random Matrix Theory (eq. 4.44). Just one sixth of the eigenvalues are inside this predicted region. The main cause for this can be the small amount of stocks (30) that comprise the DJIA index. The theory of Random matrices should just work in the case of a large number of stocks and the large length of time series. There are four eigenvalues with values higher than the maximum predicted by the RMT, λ_{max} but there are many more with values lower than the minimum value predicted, λ_{min}. Again, this might be related to the fact that the portfolio is very small.

The eigenvalues that have higher values are the ones that we believe contain non-random information about the market [38, 39]. The mean value of the eigenvector elements, for each industrial sector, for the eigenvalues that have values higher than λ_{max} are represented in Figure 7.2.

Each eigenvector shows different industrial sectors that drive it. For example, as shown

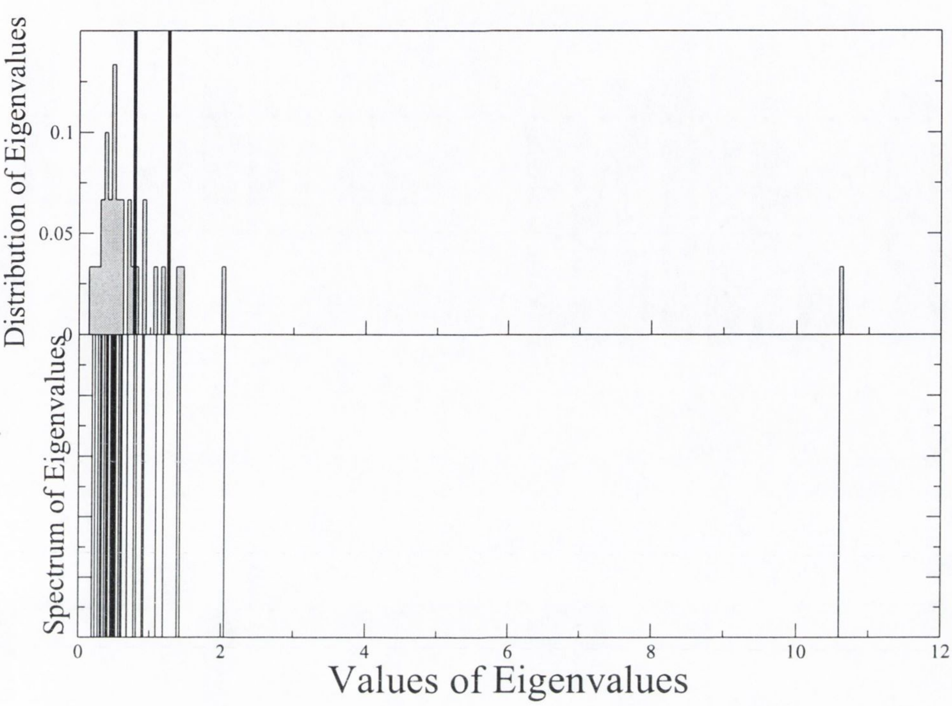


Figure 7.1: Spectrum and distribution of eigenvalues for a portfolio of 30 stocks from the DJIA. The vertical lines, in the upper figure, show the limits λ_{min}^{max} (eq. 4.44). Only $1/6^{th}$ of the eigenvalues are inside the region predicted by the RMT.

by other authors [38, 39, 133], for the eigenvector related with the highest eigenvalue, λ_{30}, all elements have the same sign, which means that all stocks contribute almost the same. This is known as the market mode and can be compared with the return index of the market that we are studying. For the eigenvector related with the 2^{nd} highest eigenvalue, λ_{29}, the stocks from different industrial sectors have different behaviours, but we can see that some industrial sectors are positive and other are negative. For example, Oil and Gas and Health Care are positive and Technology is negative for all the stocks that belong to these industrial sectors. The other two eigenvalues, λ_{28} and λ_{27} have different sectors driving them. For the eigenvector related with the 3^{rd} highest eigenvalue, λ_{28}, the main sector is Telecommunications and for the eigenvector related with the 4^{th} highest eigenvalue, λ_{27}, the main sector is Oil and Gas.

The eigenvectors related with third and fourth highest eigenvalues can be compared with the same eigenvectors from the FTSE100 portfolio in Figure 5.23 (Chapter 5), where the main industrial sectors that drive the eigenvalues are the same.

Some of these strong sectorial correlations can be seen in Figure 7.3, which shows the visualisation of the correlations between stocks using the MST for the portfolio of stocks from the

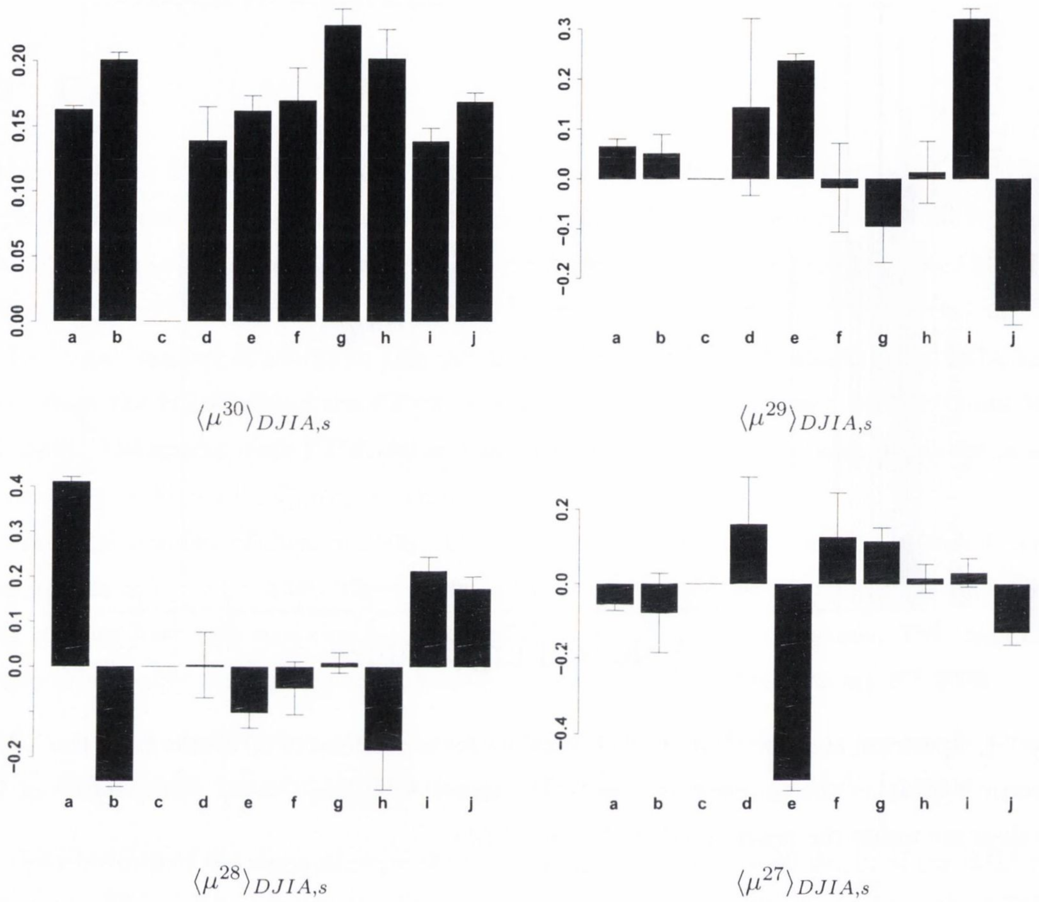


Figure 7.2: Mean value of eigenvector elements, for each industrial sector, of the four highest eigenvalues, λ_{30}, λ_{29}, λ_{28} and λ_{27} for a portfolio of 30 stocks from the DJIA index. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector.

DJIA index. The symbol used for each stock is the same used before in Chapter 5 corresponding to a specific industrial sector from the ICB classification [121] as: Oil and Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (▦), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (∙) and Technology (○).

Almost all the stocks from the DJIA index, that belong to the same industrial sector are linked together, thus illustrating the correlations between stocks of the main U.S.A. index. A similar behaviour was found in Chapter 5 (Figures 5.7, 5.8 and 5.21) for the MST of the FTSE100 portfolios.

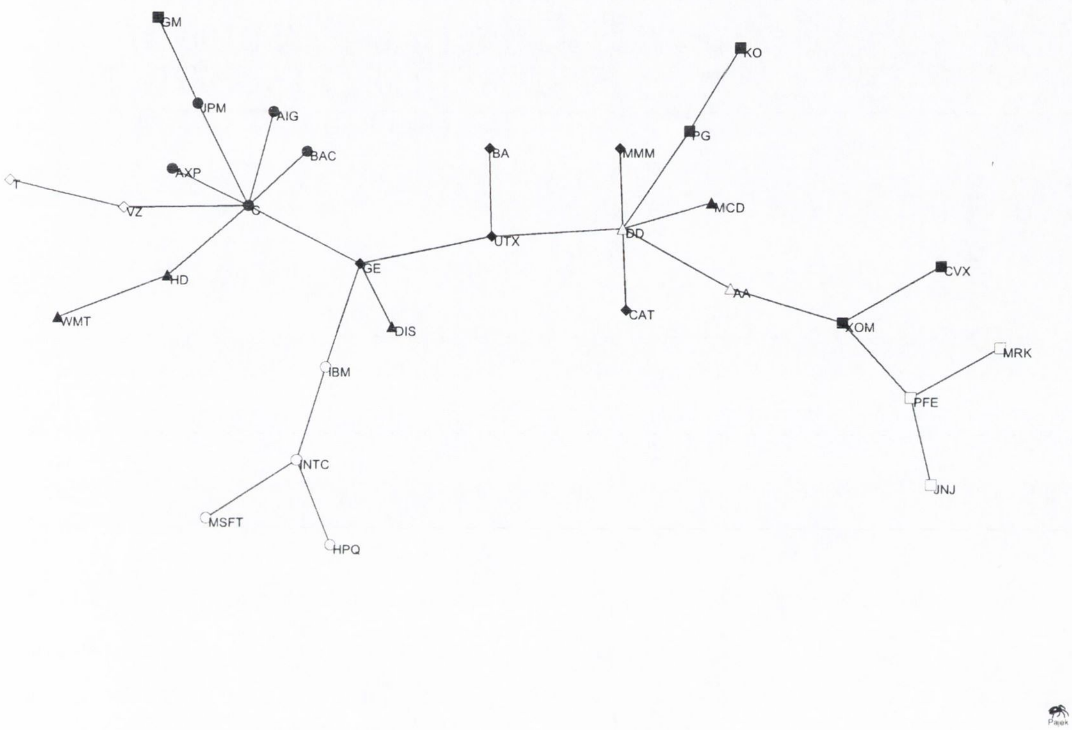


Figure 7.3: Minimal Spanning Tree for 30 stocks of the DJIA index. The time series of each stock are composed by 2146 daily closing prices. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◊), Utilities (•), Financials (•) and Technology (○).

7.4 Cross correlations between stocks of the DJIA and the FTSE100 indices

Next step in our analysis is the study of cross correlations between both sets of stocks. Taking into account the fact that the data we use is the daily closing price of stocks, and knowing that the two stock markets close at different times, we also studied the cross correlations when the return of one set is one day ahead of the return of the other (Figure 7.4).

From Figure 7.4 we can see that for values of correlations, ρ_{ij} higher than 0.3 the distribution is almost the same for the three different cases studied. This correlations are the same because they are the correlations between stocks of the same market, and that does not change from case

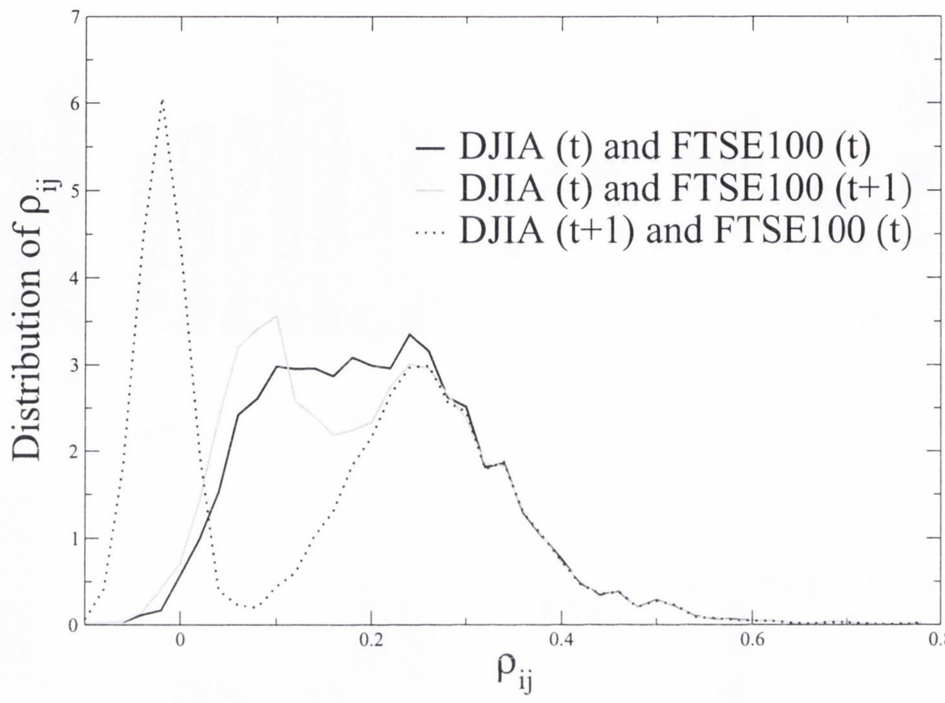


Figure 7.4: Distribution of the coefficients of the correlation matrix ρ_{ij} for the case of stocks from the DJIA and the FTSE100 at the same day (black solid line), the FTSE100 one day ahead of the DJIA (grey solid line) and the DJIA one day ahead of the FTSE100 (black dotted line).

to case. The difference is in the lower correlations, related with the cross correlations between stocks of different markets. When correlations are calculated with the stocks of the DJIA one day ahead of the ones from the FTSE100 there are many coefficients next to zero, showing a break in the correlations between both markets. This happens because the closing price of the DJIA is $4:30$ hours later than the closing price of FTSE, so when we used the closing price of the stocks of DJIA one day ahead of the closing price of the stocks from FTSE100, there is a difference of $28:30$ hours. For the case when correlations are calculated with the stocks of the FTSE100 one day ahead of those from the DJIA there is also a small break around 0.2. The difference in this case is of $19:30$ hours. From these results we concluded that the best approach is when the correlations are calculated for all the stocks at the same time, t.

The distribution of eigenvalues of the cross correlations can be seen in Figure 7.5. Just 41% of the eigenvalues stay inside the region predicted by the RMT. There are 8 eigenvalues higher than λ_{max}. These eigenvalues seems to be a mix between the higher eigenvalues of the portfolio of 30 stocks of the DJIA index in Figure 7.1 and the portfolio of 85 stocks of the FTSE100 index in Figure 5.22 (Chapter 5).

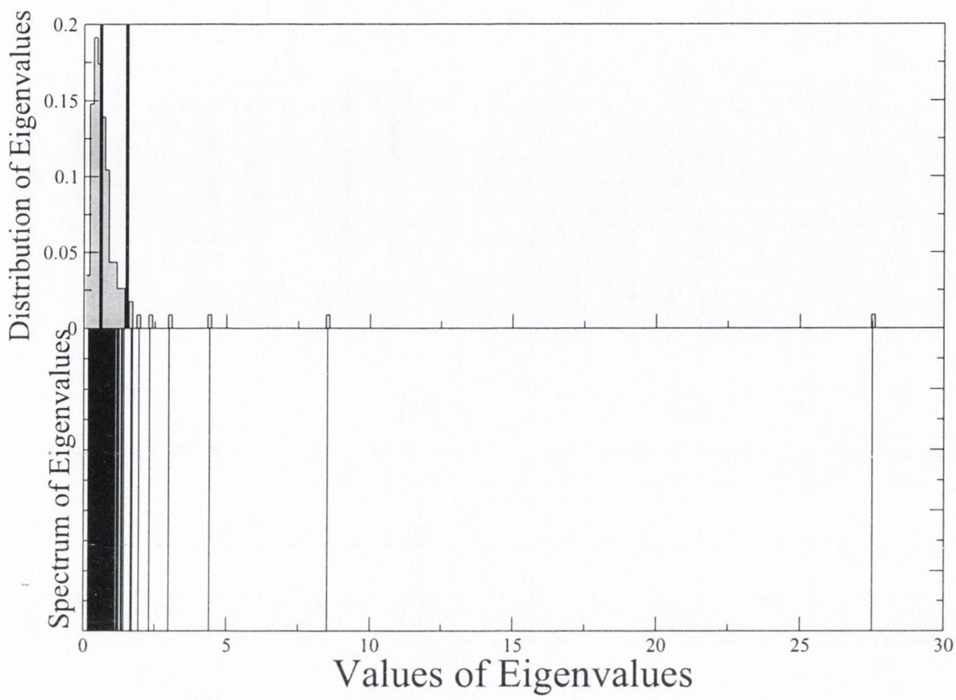


Figure 7.5: Spectrum and distribution of eigenvalues for a portfolio of 115 stocks from the DJIA and the FTSE100 indices. The vertical lines, in the upper figure, show the limits λ_{min}^{max} (eq. 4.44). Less than half (41%) of the values of the eigenvalues are inside the region predicted by the RMT.

The highest eigenvalue has almost the same value as the highest eigenvalue of the FTSE100 portfolio. The second highest can be compared with the highest from the DJIA portfolio. The third, fourth and fifth highest can be compared with the second, third and fourth highest, respectively, for the FTSE100 portfolio. The sixth highest can be compared with the second highest of the DJIA portfolio and the seventh and eighth highest with the fifth and sixth highest of the FTSE100 portfolio.

The information contained in these eigenvalues show us how stocks from different markets are related to each other. Figure 7.6 shows the eigenvectors of the four highest eigenvalues, λ_{115}, λ_{114}, λ_{113} and λ_{112}.

The mean value of eigenvector elements in Figure 7.6 shows that for the eigenvector related with the highest eigenvalue, λ_{115}, all the elements have the same sign, as we saw for the individual markets. The eigenvector related with the second highest eigenvalue, λ_{114}, shows a segregation between the stocks of the two different markets, where all the elements of the DJIA index have a positive sign and all the elements of the FTSE100 index have a negative sign, apart from the

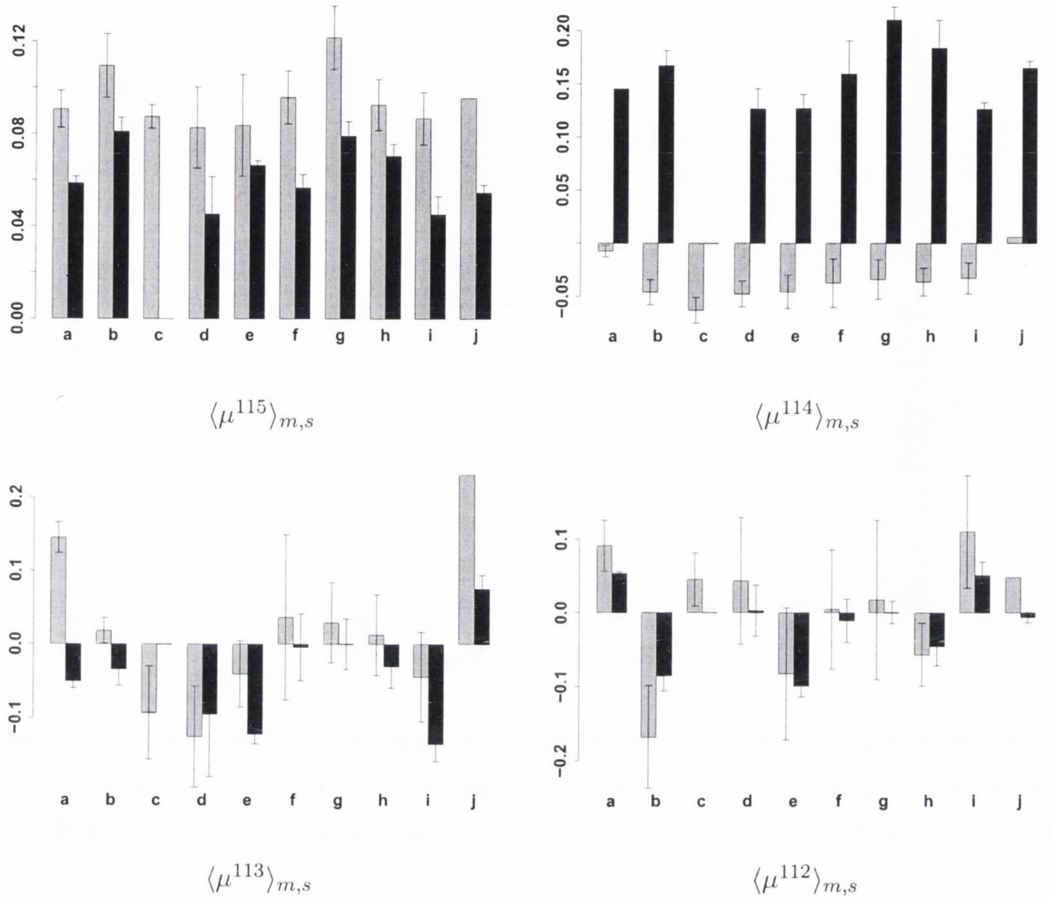


Figure 7.6: Mean value of eigenvector elements, for each industrial sector, of the four highest eigenvalues, λ_{115}, λ_{114}, λ_{113} and λ_{112} for a portfolio of 115 stocks from the DJIA and the FTSE100 indices. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector. The columns in grey are respected to the FTSE100 stocks and the columns in black to the DJIA stocks.

FTSE100 stock that belong to the Technology industrial sector, that also has a positive sign like the stocks from the DJIA index.

The eigenvector related with the third highest eigenvalue, λ_{113} shows similarities with the second highest eigenvalue for both the DJIA and the FTSE100 indices and the eigenvector related with the fourth highest eigenvalue, λ_{112} shows similarities with the third highest eigenvalue for both the DJIA and the FTSE100 indices.

We have also performed an eigenvector analysis for the cases where the values of stocks of one market are one day ahead of the others. When the stocks of the FTSE100 index are one day ahead of the stocks of the DJIA index, the results are the same as in Figure 7.6 but when the stocks of the DJIA index are one day ahead of the stocks of the FTSE100 index, the eigenvector elements show very different results as we can see from Figure 7.7.

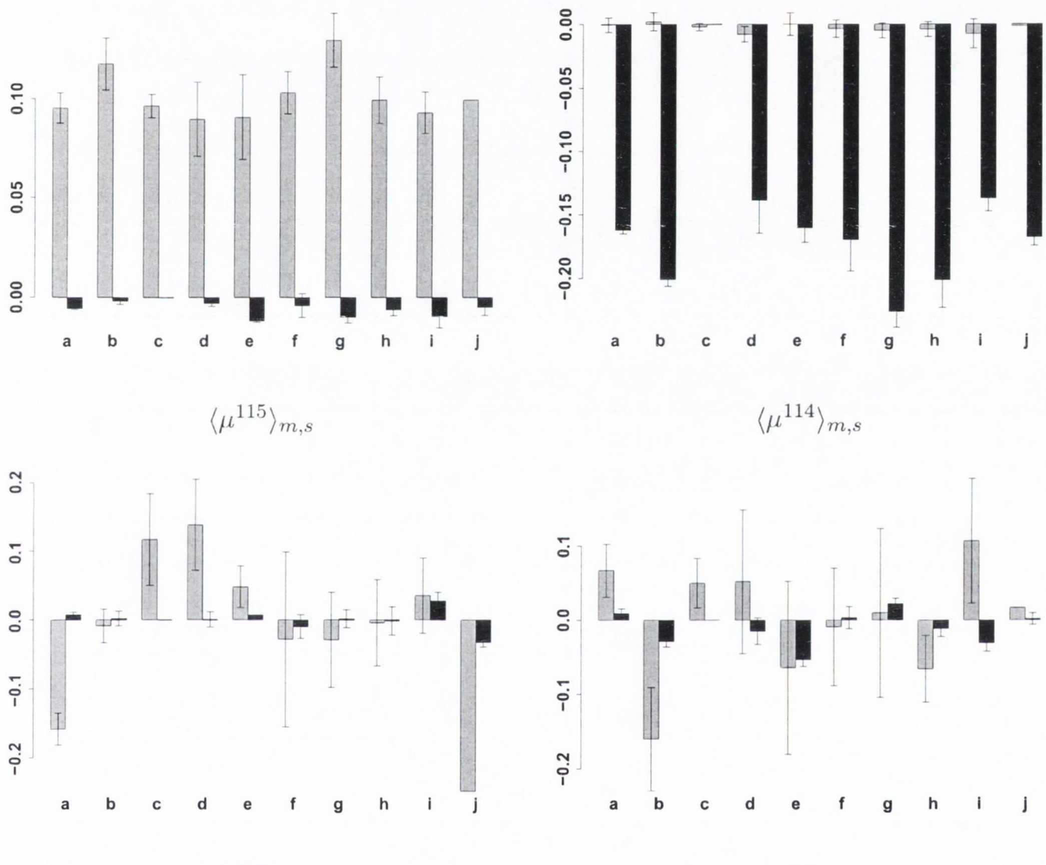


Figure 7.7: Mean value of eigenvector elements, for each industrial sector, of the four highest eigenvalues, λ_{115}, λ_{114}, λ_{113} and λ_{112} for a portfolio of 115 stocks from the DJIA and the FTSE100 indices when the correlations are calculated for the stocks of the DJIA index one day ahead of the stocks of the FTSE100 index. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector. The columns in grey are respected to the FTSE100 stocks and the columns in black to the DJIA stocks.

The mean value of eigenvector elements in Figure 7.7 shows that the eigenvector related to the highest eigenvalue, λ_{115} no longer has all its elements positive. It seems that this eigenvalue just shows the influence of the stocks of the FTSE100 index in the portfolio. On the other hand, the eigenvector related with the second highest eigenvalue, λ_{114} shows just the influence of the stocks from the DJIA index. The eigenvector related with the third and fourth highest eigenvalues mimic the composition of the eigenvectors related with the second and third highest eigenvalues of the individual stocks of the FTSE100 index.

The segregation between the stocks of the FTSE100 and the DJIA indices can be seen when we analyse the MST of the portfolio of 115 stocks. In Figure 7.8 we can see that on the top right of the MST all the stocks of the FTSE100 are linked together with the same industrial sector cluster that we saw previously when we studied the MST for the individual portfolio of stocks of the FTSE100 in Figure 5.21 (Chapter 5). At the bottom left of the MST all the stocks of the DJIA index are also linked together with the same configuration as in Figure 7.3. The link is between BP from the FTSE100 index and Chevron from the DJIA index both from the industrial sector Oil and Gas.

For the other two cases where the stocks of one market are calculated one day ahead of the other, the structure of the MST does not change. There is always a cluster of stocks from the FTSE100 index on one side and another cluster of stocks from the DJIA index on the other side. The only feature that changes is the link between both markets. If the correlations are calculated for the stocks of the FTSE100 one day ahead of the DJIA, the link is between GlaxoSmithKline from the FTSE100 and AT&T from the DJIA. If the correlations are calculated for the stocks of the DJIA one day ahead of the FTSE100, the link is between Alliance Trust from the FTSE100 and JP Morgan Chase from the DJIA.

7.5 Conclusions

We have used two different methods to study correlations between stocks of the FTSE100 and the DJIA. Our results using Random Matrix Theory show that the markets remain largely separate even when cross-correlations between stocks across the two markets are included. The results for the Minimal Spanning Trees broadly reflect the results from the Random Matrix Theory. But it is not as easy to see the detail provided by the Random Matrix analysis. This of course is not too surprising since the Minimal Spanning Trees approach only uses partial information from the correlation matrix.

Much research in finance has addressed the issue of whether or not stocks ultimately cluster by market or by industry. There is no consensus on this. Some [134] suggest that the clustering is primarily industrial, while others [135] contend that the split is primarily geographical. The evidence here is that geographical (more correctly, market) location is the most important element in determining the cluster into which a stock falls. The implication for portfolio managers

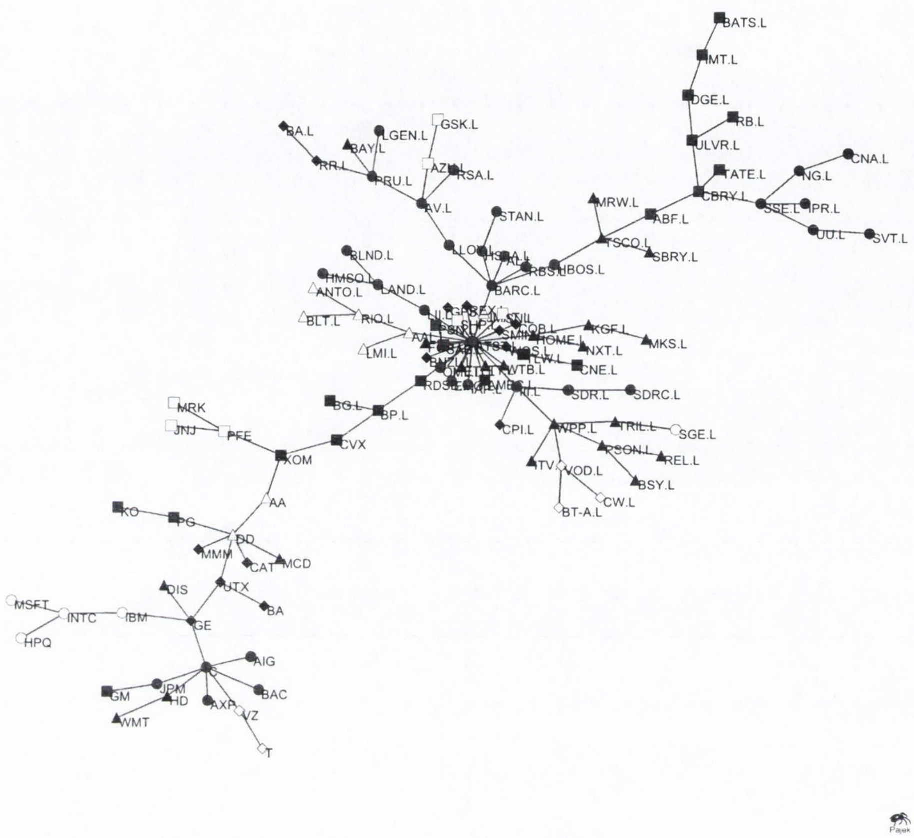


Figure 7.8: Minimal Spanning Tree for 115 stocks of the FTSE100 and the DJIA indices. The time series of each stock is composed of 2146 daily closing prices. Each symbol corresponds to a specific industry from the ICB classification. All the stocks from the FTSE100 index have a *.L* after the code. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (⧫), Consumer Goods (▩), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (●), Financials (•) and Technology (○).

is that, at least at a first level, they should consider diversification along market lines, and only subsequently along industrial or sectoral lines.

It is an important consideration that we are studying stocks from two markets that operate at different time. Further investigation with stocks from markets that operate at the same time is essential as we show further in Chapter 8. If we want to upgrade our study of stocks from the FTSE100 and the DJIA indices we should use intraday data from both markets and study the correlation at the same minute, but only for minutes when both markets are in operation.

Chapter 8

Cross correlations between portfolios of stocks from different geographical locations

8.1 Introduction

In this chapter we use Random Matrix Theory to examine the correlation between stocks traded on markets of different countries and compare with the results obtained from a simple market model as done before in section 5.6.1. Based on the study of correlations of indices from different countries in Chapter 6 [52] where we analysed the Minimum Spanning Trees of 56 market indices around the world, we want to study the cross-correlations between stocks that are quoted in different stock markets. In this previous study, the indices cluster in terms of geographical location and they were localised in a central cluster of West and Central Europe markets and three other clusters around: American indices; Asia-Pacific indices; Eastern European indices.

Studying the correlations between stocks that belong to different markets give us a better understand of the correlation between the indices. For example, if the stocks of DJIA segregate from the stocks of FTSE100, as we saw in Chapter 7, this show us why the two main indices in the study of Chapter 6 were so separate from each other.

A preliminary extension of this work was shown in Chapter 7 where we analysed the cross correlation between stocks listed in the FTSE100 and DJIA indices, two of the most common indices for investors. But as we saw from Figure 6.9, these two indices are not the main hubs of the MST of the indices around the world. The central index is France.

For our study of stocks from different indices we chose three main world indices from three different countries: France, U.K. and U.S.A. (U.K. and U.S.A. are separated from France by one and three links, respectively in the MST of Figure 6.9). The three indices that we study here are the CAC40, the FTSE100 and the DJIA and the stocks from each index are represented in tables C.3, C.1 and C.2, respectively. As shown in Chapter 7 the stocks from the FTSE100 segregate from the stocks from the DJIA. With the inclusion of the stocks from the CAC40 index, we want to test if these stocks also segregate from the previous two groups, or if they cluster with some group of stocks.

To test the time-mismatch problem between stocks that are listed in indices from different markets, we also present a second study with European stocks from three indices with close geographical locations, namely France, Netherlands and Belgium (France and Belgium are separated from Netherlands by only one link in the MST of Figure 6.9). With this second portfolio, we want to check the differences between markets that have geographical and trade affinity and markets that don't. The three indices study here are the CAC40, the BEL20 and the AEX and the stocks from each index are represented in tables C.3, C.4 and C.5. Important differences between this second portfolio and the first one is that for the European portfolio all the stocks are quoted with the same currency and all the markets also work at the same time. In this second portfolio there are some companies listed in different indices.

In the next section we present the data analysed here and then our results for the Minimal Spanning Tree analysis. In section 8.4 we present a factor model for random time series and the results of the eigensystem analysis of both correlation matrices of real and random data.

8.2 Data

Our data is the daily closing price for 2146 days, from January 3^{rd} 2000 until March 24^{th} 2008. Our first portfolio contains 149 stocks from the CAC40, the FTSE100 and the DJIA indices, with 34, 85 and 30 stocks, respectively. We changed all the currencies of these stocks to the same, US dollars, using the currency exchange for each specific day.

The second portfolio is composed of the closing price in Euros, as it is the currency established by these three countries. The portfolio contains 72 stocks from the CAC40, the BEL20 and the AEX indices, with 34, 17 and 21 stocks, respectively.

We divided the stocks in groups of industrial sectors using the ICB classification [121] and use all the stocks that belong to each index that we have data available for the period January 3^{rd} 2000 until March 24^{th} 2008.

8.3 Minimal Spanning Trees

8.3.1 Analysis of data from CAC40, FTSE100 and DJIA indices

We have computed the correlation matrix, and analysed the eigensystem of this matrix for the portfolio consisting of stocks from the CAC40, the FTSE100 and the DJIA indices called portfolio A. The Minimal Spanning Tree constructed from the matrix of distances is shown in Figure 8.1.

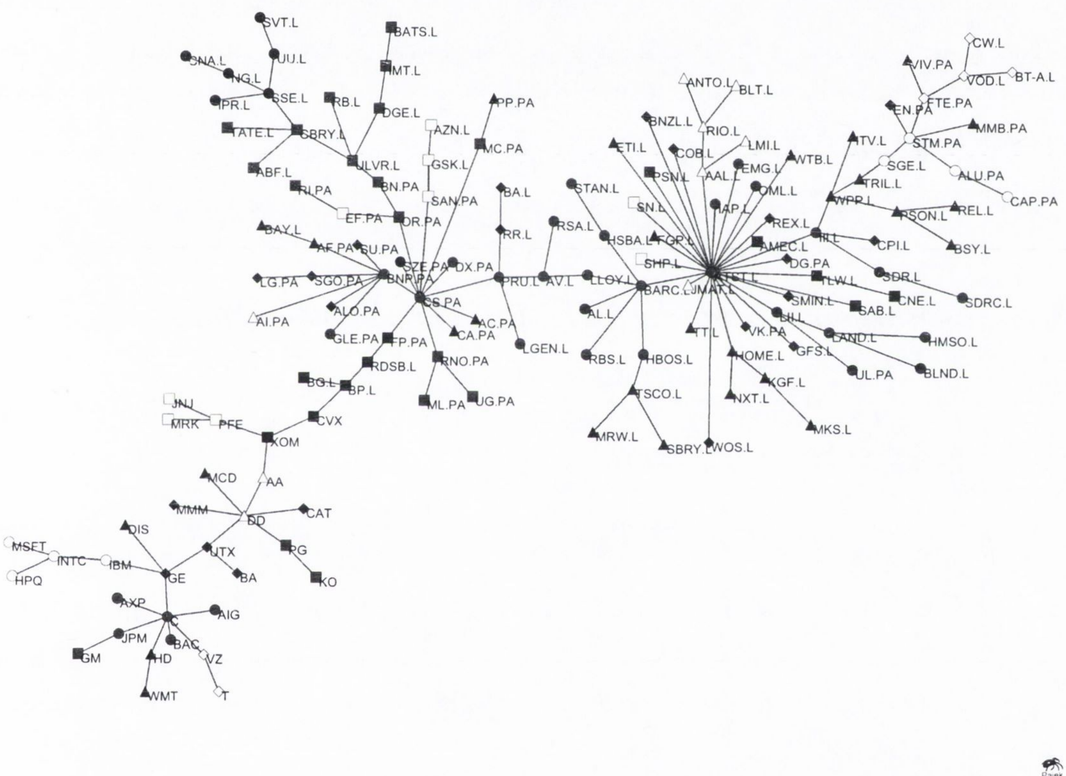


Figure 8.1: Minimal Spanning Tree for 149 stocks of the CAC40, the FTSE100 and the DJIA indices. The time series of each stock is composed of 2146 daily closing prices. The stocks from the CAC40 have a *.PA* after the code and the ones from the FTSE100 have a *.L*. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

From the MST we can see that there is a segregation from the stocks of the DJIA index in the bottom left corner of the MST. All 30 stocks from the DJIA are linked together and maintained the structure shown in Figure 7.3 for these stocks when studied alone. The link to this cluster with the rest of the portfolio is made through Chevron from the DJIA index and BP from the FTSE100 index as in Figure 7.8. These stocks form an Oil and Gas cluster that consist of 6 stocks from the three different markets.

As shown in previous figures the main hub of the MST is the stock Alliance Trust from the FTSE100 index. There is also interdependence between some stocks of the CAC40 and the FTSE100 indices. All stocks from Telecommunications are linked together, with France Telecom joining the previous cluster of the FTSE100. The Technology stocks also form another cluster with stocks from the CAC40 and the FTSE100 indices. The same also happen with the stocks from Consumer Goods, Health and Care and Real Estate (sub-sector of Financials). Another curious result is the link between British Airways and Air-France as part of the Consumer Services industrial sector.

8.3.2 Analysis of data from CAC40, BEL20 and AEX indices

The MST of the second portfolio, which we call portfolio B, consisting of stocks from CAC40, BEL20 and AEX indices, is shown in Figure 8.2.

From this MST we can see many clusters with stocks from different markets. For example, the Telecommunications cluster has one stock from each market. The same happens with the industrial sectors Technology and Oil and Gas. The sub-sector Real Estate from the industrial sector Financials is also composed of three stocks, one from each market. Another cluster of stocks from different markets is formed from stocks of the super-sector Food and Beverage, part of the Consumer Goods industrial sector.

The main hubs of this MST are AXA (with symbol $CS.PA$) from the CAC40 index, an insurance company with 13 links, one of these links is to the other hub of the MST, ING Groep (with symbol $INGA.AS$) from the AEX index, another insurance company with 13 links. We can see that part of the stocks from the BEL20 index cluster together, or are directly linked to the same company that is also quoted in a different market. This is the case of Fortis, a bank that is quoted in the BEL20 index with the symbol $FORB.BR$ and is quoted in the the AEX index with the symbol $FORA.AS$. The same also happens for two other companies, Suez, an utilities stock quoted in the BEL20 and the CAC40 indices with symbols $SZEB.BR$ and $SZE.PA$, respectively and Dexia, a bank quoted in the BEL20 and the CAC40 indices with symbols $DEXB.BR$ and $DX.PA$, respectively.

Figure 8.2: Minimal Spanning Tree for 72 stocks of the CAC40, the BEL20 and the AEX indices. The time series of each stock is composed of 2146 daily closing prices. The stocks from the CAC40 have a $.PA$ after the code, the stocks from the BEL20 have a $.BR$ after the code and the ones from the AEX have a $.AS$. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

8.4 Simple Market Model

To try to mimic the time series of returns for each stock we need to create a market model. So we use the one-factor model already presented in section 5.6.1:

$$r_i(t) = \alpha_i + \beta_i R_{m_j}(t) + \epsilon_i(t) \tag{8.1}$$

where R_{m_j} is the total return of the stocks of the market j (m_j, $j = 1, 2, 3$) and $\epsilon_i(t)$ is a Gaussian distributed random number. The total return is calculated as:

$$R_{m_j}(t) = \sum_{i \in m_j} R_i(t) \tag{8.2}$$

where $R_i(t)$ is the return of stock i at time t and we will call this total return the index of market j.

The parameters α_i and β_i are estimated by the least square method between the real returns and the index of the market to which the stock belongs:

$$\alpha_i = \langle R_i \rangle - \beta_i \langle R_{m_j} \rangle \tag{8.3}$$

$$\beta_i = \frac{\langle R_i R_{m_j} \rangle - \langle R_i \rangle \cdot \langle R_{m_j} \rangle}{\langle (R_{m_j})^2 \rangle - \langle R_{m_j} \rangle^2} \tag{8.4}$$

After creating the random time series we compute the correlation matrix and then we perform an eigensystem analysis. We also create a MST from the random market.

The MST for both portfolio A and B are represented in Figures 8.3 and 8.4, respectively.

The MST of Figure 8.3 is not similar with the real MST from Figure 8.1, because there are some hubs with a huge amount of links, which is not normal in the MST of real data. This behaviour was already stated in Chapter 5 for Figure 5.15 where we also used a market model to create some random time series.

The main feature of MST of Figure 8.3 is that there are three cluster of stocks. The cluster in the bottom left of MST has all the stocks from the DJIA index, the middle cluster has all the stocks from the CAC40 index and the cluster on the top right of MST has all the stocks from the FTSE100 index.

The fact that all the stocks that belong to the DJIA index are linked with each other and form a cluster is a mimic of the behaviour of the real MST in Figure 8.1, but the absence of interdependence between the stocks of the CAC40 and the FTSE100 shows that the simple model cannot mimic the whole figure of the MST.

The main hubs in the Figure 8.3 are some of the stocks from the Financials industrial sector, between them there are the two main hubs from the MST of real data (Figure 8.1), AXA and Alliance Trust, which are also the main hubs for their two market indices, the CAC40 and the FTSE100, respectively.

The MST of Figure 8.4 also shows a division in three main clusters. The bottom right cluster has all the stocks from the BEL20 index, the middle cluster has all the stocks from the AEX index and the cluster in the top left of MST has all the stocks from the CAC40 index.

The main hubs of the MST continue to be the stocks from Financials industrial sector with the insurance companies AXA and ING Groep as two of these main hubs like it was for the MST

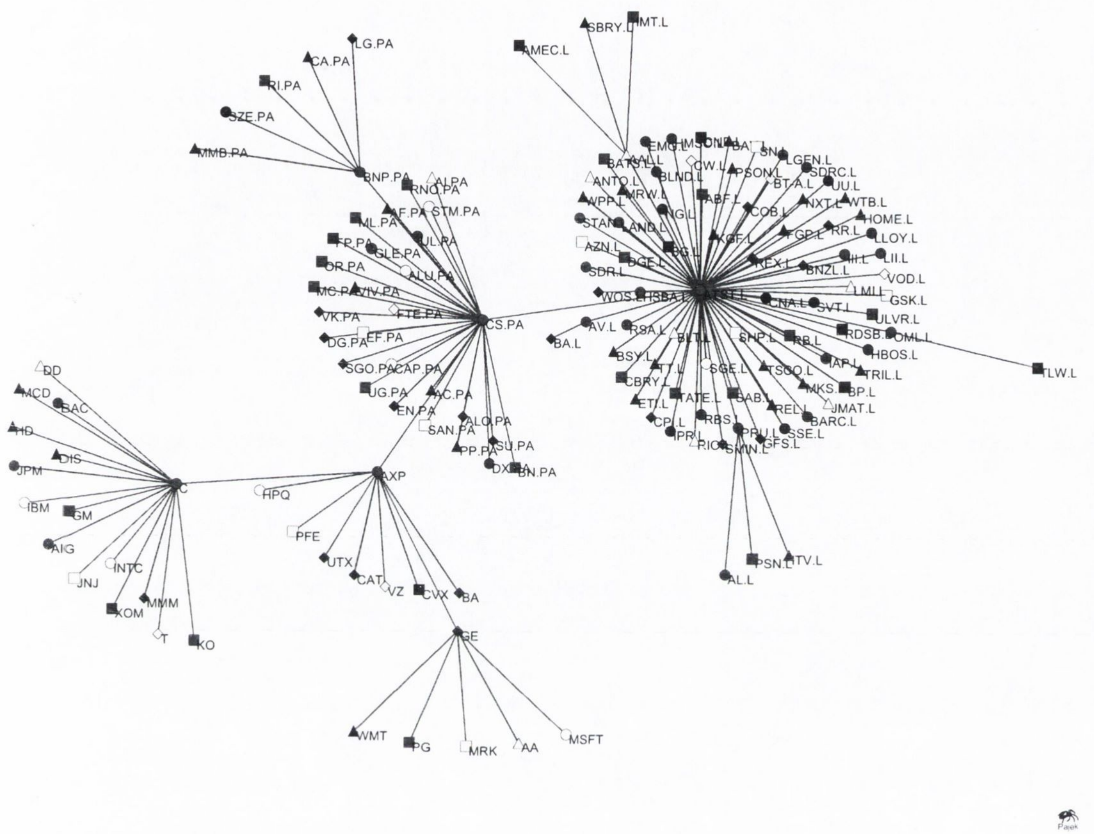


Figure 8.3: Minimal Spanning Tree for 149 stocks of the CAC40, the FTSE100 and the DJIA indices created using a random market model. The time series of each stock is composed of 2146 daily closing prices. The stocks from the CAC40 have a $.PA$ after the code and the ones from the FTSE100 have a $.L$. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

of real data in Figure 8.2. The absence of interdependency between stocks of different market indices shows that a simple market model is not enough to mimic the structure of a MST.

8.5 Market indices

The index for each market calculated from the random time series ($r_{m_j}(t) = \sum_{i \in m_j} r_i(t)$, $j = 1, 2, 3$) is compared with the index calculated from the real time series ($R_{m_j}(t)$, eq. 8.2). The correlation between two indices is computed from eq. 4.30, where we use the return of each index, instead of the return of stocks. Tables 8.1 and 8.2 show the values for these correlations.

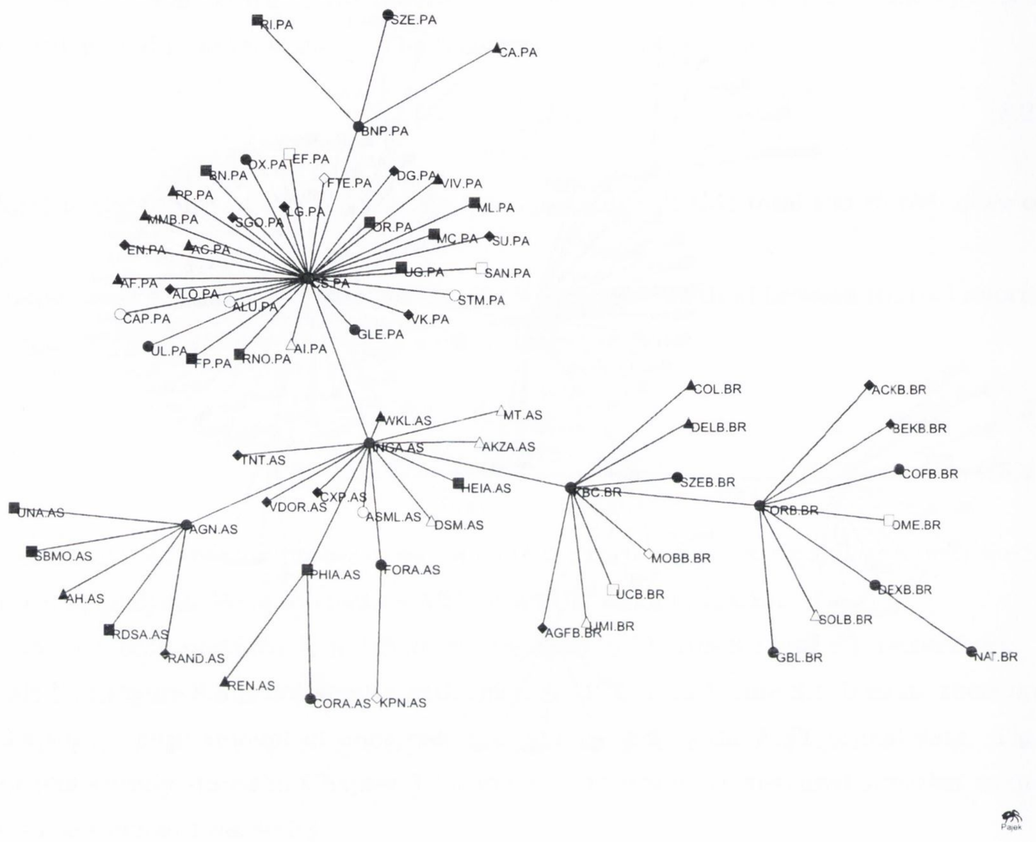


Figure 8.4: Minimal Spanning Tree for 72 stocks of the CAC40, BEL20 and AEX indices created using a random market model. The time series of each stock is composed of 2146 daily closing prices. The stocks from the CAC40 have a $.PA$ after the code, the stocks from the BEL20 have a $.BR$ after the code and the ones from the AEX have a $.AS$. Each symbol correspond to a specific industry from the ICB: Oil & Gas (■), Basic Materials (△), Industrials (♦), Consumer Goods (■), Health Care (□), Consumer Services (▲), Telecommunications (◇), Utilities (•), Financials (•) and Technology (○).

Table 8.1 shows that the higher correlation is between the CAC40 and the FTSE100 indices, as it is expected since the markets are on the same geographical cluster and are also two of the main indices in the world. The correlations between the DJIA returns and any other index are very low. These results are in agreement with what we have shown for the MST in Figure 8.1 with the stocks from the DJIA separated from the rest and in MST of Figure 6.9 where the DJIA is 3 links away from the CAC40 but the FTSE100 and the CAC40 are directly linked.

Table 8.2 shows that the higher correlation is between the CAC40 and the AEX indices and the second highest correlation is between the CAC40 and the BEL20 indices. This last result is

Table 8.1: Correlations between market returns for portfolio A: the CAC40, the FTSE100 and the DJIA indices. The cells in light grey represent the correlation between the return of real market i, $R_{m_i}(t)$ and the return of real market j, $R_{m_j}(t)$. The cells in dark grey represent the correlation between the return of random market i, $r_{m_i}(t)$ and the return of random market j, $r_{m_j}(t)$. The cells in white represent the correlation between the return of real market i, $R_{m_i}(t)$ and the return of random market i, $r_{m_i}(t)$ and those deviate from 1.

	CAC40	FTSE100	DJIA
CAC40	0.97	0.81	0.43
FTSE100	0.78	0.98	0.38
DJIA	0.41	0.37	0.97

Table 8.2: Correlations between market returns for portfolio B: the CAC40, the BEL20 and the AEX indices. The cells in light grey represent the correlation between the return of real market i, $R_{m_i}(t)$ and the return of real market j, $R_{m_j}(t)$. The cells in dark grey represent the correlation between the return of random market i, $r_{m_i}(t)$ and the return of random market j, $r_{m_j}(t)$. The cells in white represent the correlation between the return of real market i, $R_{m_i}(t)$ and the return of random market i, $r_{m_i}(t)$ and those deviate from 1.

	CAC40	BEL20	AEX
CAC40	0.97	0.79	0.89
BEL20	0.73	0.93	0.78
AEX	0.82	0.70	0.95

in disagreement with the results of MST of Figure 6.9 where the CAC40 and the BEL20 indices are directly linked with the AEX index but are separated by two links, so we should expect a higher value of correlation between the BEL20 and the AEX indices than between the CAC40 and the BEL20 indices. If we look at the correlation matrix that produces the MST of Figure 6.9 we can see that the value of correlation between the BEL20 and the AEX indices is 0.81 and it is higher than the correlation between the CAC40 and the BEL20 indices which is 0.77. The difference of values with our table 8.2 is related with the fact that when we calculate the returns $R_{m_j}(t)$, we compute an arithmetic mean with the values of returns of stocks that we have in our portfolio which is not always the total number of stocks that compose the index. It can also be the case that some indices are calculated with a weighted arithmetic mean.

The higher value of correlation between the CAC40 and the AEX indices is also in agreement with the results of MST of Figure 8.2 where many stocks of the CAC40 and the AEX indices cluster together.

8.6 Eigensystem analysis

8.6.1 Distribution of eigenvalues

The distribution of eigenvalues for the random correlation matrix and the real correlation matrix for portfolio A is shown in Figure 8.5.

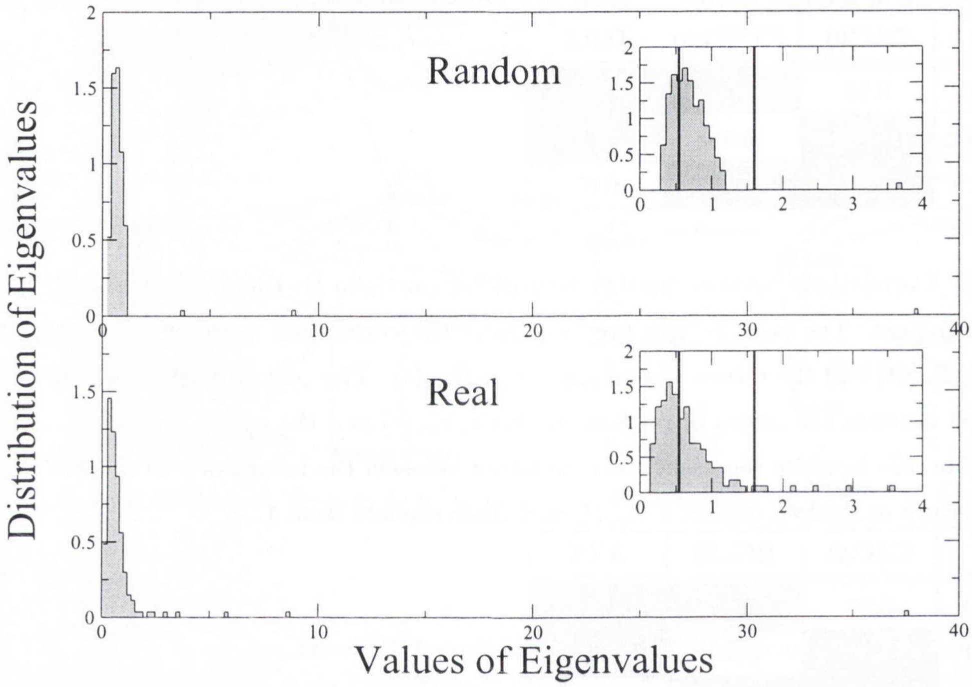


Figure 8.5: Distribution of eigenvalues of the correlation matrix computed from the random time series (top) and from the real time series (bottom) for the portfolio A. In both cases the highest eigenvalue has almost the same value. The vertical lines, in the inset figures, indicate the region predicted by random matrix theory, λ_{min}^{max}.

We can see that there is a high number of eigenvalues that stay outside the region predicted for a random matrix, λ_{min}^{max}. The percentage of eigenvalues that stay inside this region is 45% for the real case and 66% for the random case. The random time series leads to three eigenvalues outside the predicted region as shown by F. Lillo *et al.* for multifactor models [136], while there are more than three eigenvalues outside this region for the real data. For the real case there are 9 eigenvalues higher than λ_{max}.

The two highest eigenvalues for the random case ($\hat{\lambda}_{149} = 38.02$ and $\hat{\lambda}_{148} = 8.89$) have almost the same value as the two highest for the real data ($\lambda_{149} = 37.55$ and $\lambda_{148} = 8.68$).

In Figure 8.6 we show the distribution of eigenvalues for the random and real correlation matrices for portfolio B.

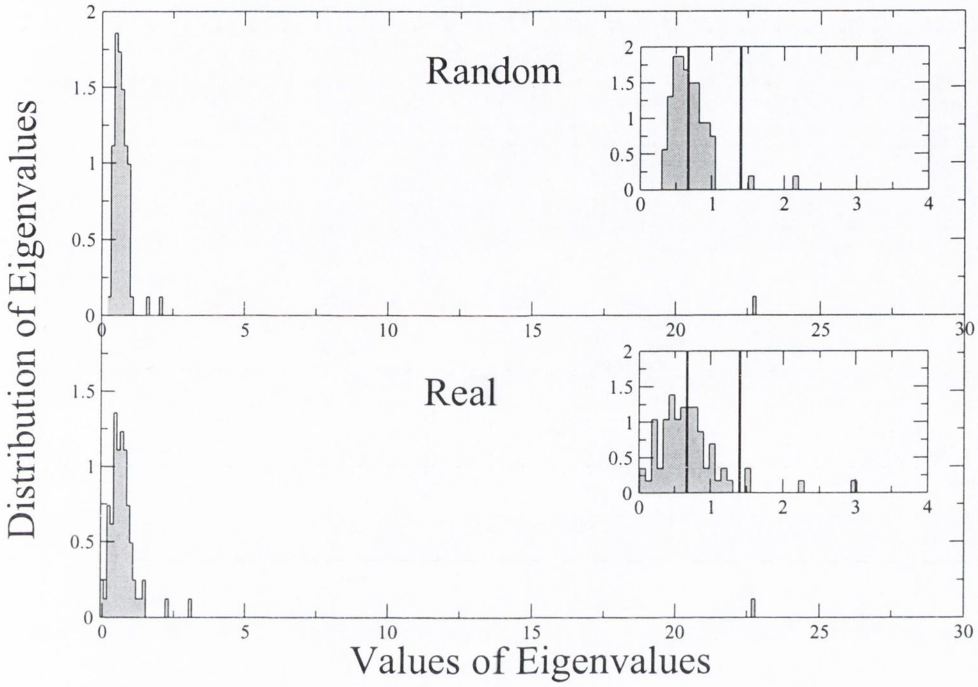


Figure 8.6: Distribution of eigenvalues of the correlation matrix computed from the random time series (top) and from the real time series (bottom) for the portfolio B. In both cases the two highest eigenvalue have almost the same value. The vertical lines, in the inset figures, indicate the region predicted by random matrix theory, λ_{min}^{max}.

Also for portfolio B there are many eigenvalues that stay outside the region predicted for a random matrix, λ_{min}^{max}. The percentage of eigenvalues that stay inside this region is 40% for the real case and 43% for the random case. As in Figure 8.5 for the random case there are only 3 eigenvalues with values higher than λ_{max} [136]. For the real case there are 5 eigenvalues higher than this value.

The highest eigenvalue for the random case ($\hat{\lambda}_{72} = 22.70$) has almost the same value as the highest one for the real data ($\lambda_{72} = 22.69$).

8.6.2 Inverse Participation ratio

Looking at the eigenvectors that correspond to each eigenvalue we can get some information about the nature of time series that we are studying. Comparing the Inverse Participation ratio (subsection 4.4.1) for the real and random cases we can almost see which eigenvalues from the random case correspond to eigenvalues for the real case. In Figure 8.7 we present the IPR for portfolio A.

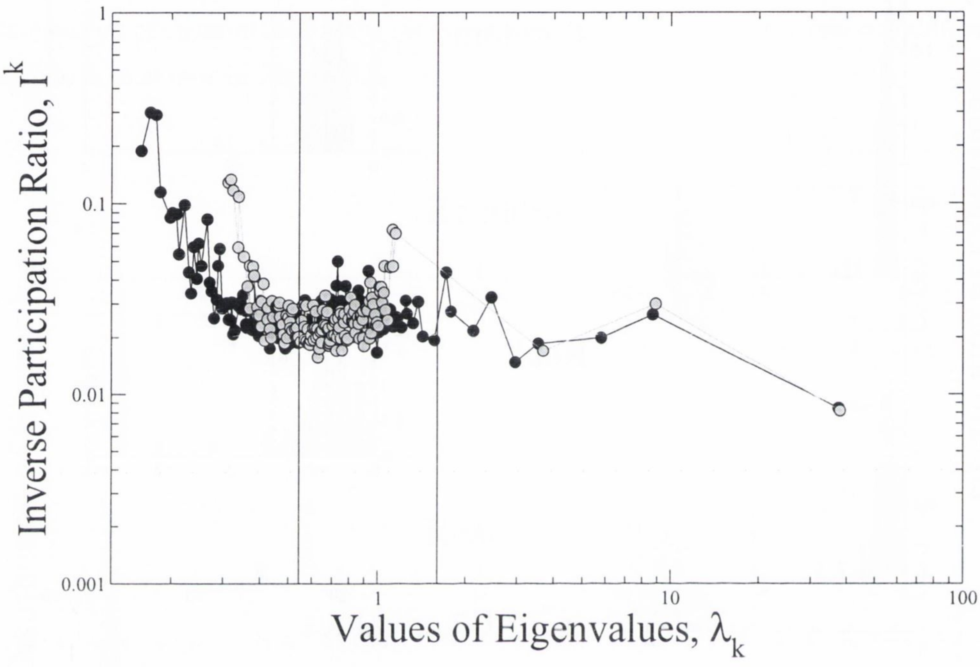


Figure 8.7: Inverse Participation Ratio from random time series (grey circles) and real time series (black circles) for the portfolio A on a $log - log$ scale. The vertical lines indicate the region predicted by random matrix theory, λ_{min}^{max}.

From Figure 8.7 we can see that the highest eigenvalue from the random case, $\hat{\lambda}_{149}$ has almost the same value of IPR as the highest eigenvalue for the real case, λ_{149}. The same behaviour can be seen with the second highest eigenvalue, but the third highest eigenvalue of the random case, $\hat{\lambda}_{147}$ seems to be very similar with the fourth highest eigenvalue of the real case, λ_{146}. A better comparison can be made when we study each element of the eigenvectors related with these eigenvalues.

The comparison of the Inverse Participation ratio between the real and random cases for portfolio B is presented in Figure 8.8.

For the case of portfolio B, the correspondence between an eigenvalue for the random case and another from the real case is not so evident than for portfolio A. The only evident case is for the highest eigenvalue of random case, $\hat{\lambda}_{72}$ which has almost the same value of IPR as the highest eigenvalue for the real case, λ_{72}.

For both portfolios, the IPR of the highest eigenvector is low (Figures 8.7 and 8.8), which

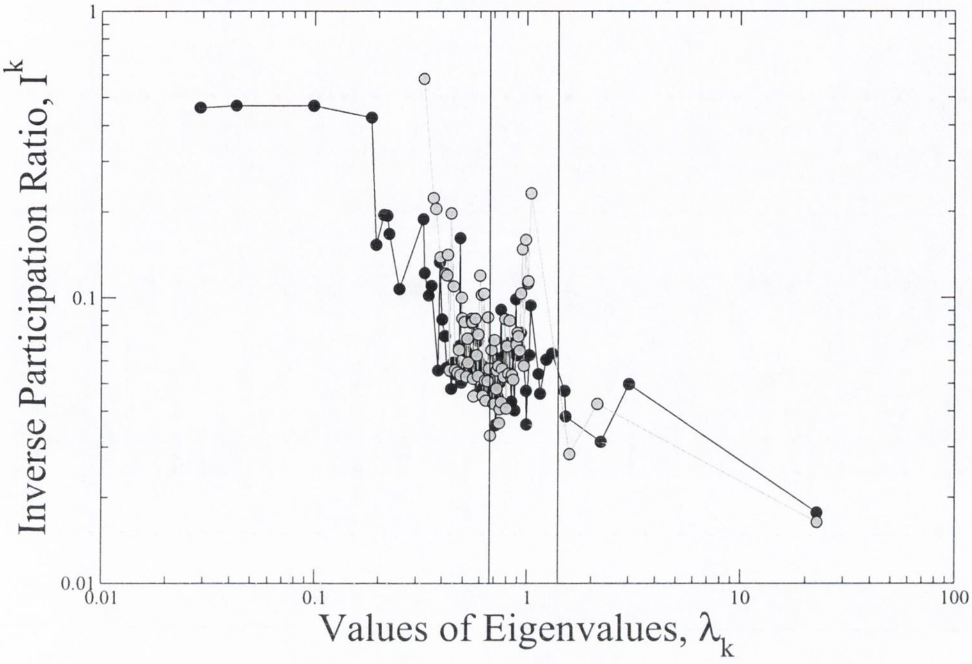


Figure 8.8: Inverse Participation Ratio from random time series (grey circles) and real time series (black circles) for the portfolio B on a $log - log$ scale. The vertical lines indicate the region predicted by random matrix theory, λ_{min}^{max}.

means that the stocks contribute in the same way. We can see this behaviour when we analyse each eigenvector and its elements in the Figures 8.9 and 8.11 where all elements are positive. This corresponds to the general trend of all stocks and it also specifies how each stock contributes to the overall index.

8.6.3 Eigenvector elements of the highest eigenvalues

To check if these aparent correspondences of the IPR between eigenvalues of the random case and eigenvalues of the real case are real we compute the mean value of the eigenvector elements (subsection 4.4.2) for the real and random cases.

In the following Figures, we will show the comparison between the eigenvectors from the real and random cases, for portfolios A and B. All the elements of the eigenvector that corresponds to the highest eigenvalue have the same positive sign. This is shown for both portfolios and for the real and random cases. Other correspondence is between the real and random elements of the eigenvector related with the second highest eigenvalue for portfolio A. In both cases, the elements related with the DJIA index segregate from the other elements.

In Figure 8.9 we present these elements, for each industrial sector and index, of the two highest eigenvalues, λ_{149} and λ_{148}, for the study of portfolio A.

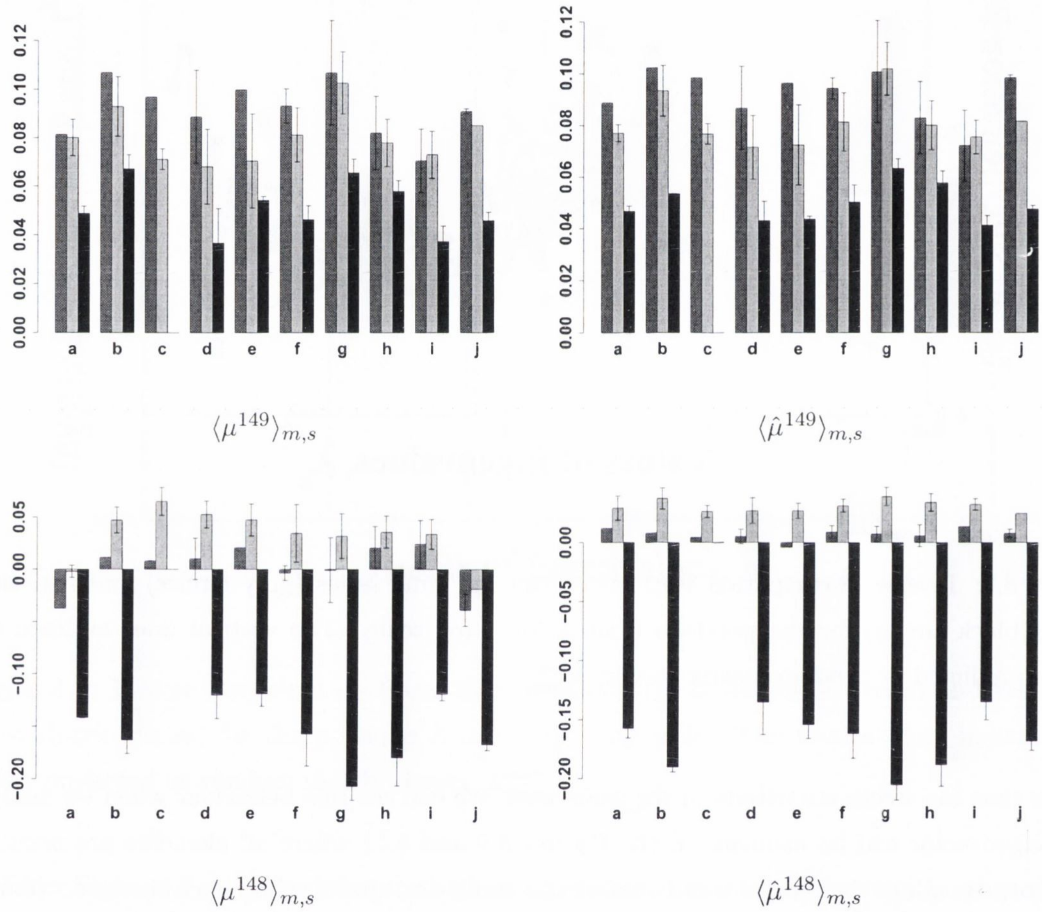


Figure 8.9: Mean value of eigenvector elements, for each industrial sector, of the two highest eigenvalues, λ_{149} and λ_{148}, for the study of portfolio A. The left figures correspond to the real time series and the right figures to the random time series. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector. Each colour column represent one index: dark grey for the CAC40 index, light grey for the FTSE100 index and black for the DJIA index.

For portfolio A, the eigenvector corresponding to the highest eigenvalue, λ_{149} (top left picture of Figure 8.9) shows all the elements of different sector and different indices with the same positive sign, showing that all the elements of this eigenvector contribute positively to the highest eigenvalue. The same behaviour is mimic by the random case where the eigenvector corresponding to the highest eigenvalue, $\hat{\lambda}_{149}$ (top right picture of Figure 8.9) shows all the stocks from different sectors and indices with a positive sign.

The eigenvector corresponding to the second highest eigenvalue, λ_{148} (bottom left picture of Figure 8.9 shows a segregation between the stocks from the DJIA index (all negative) and the stocks from the CAC40 and the FTSE100 (almost all of them positive). The eigenvector corresponding to the second highest eigenvalue of the random case, $\hat{\lambda}_{148}$ (bottom right picture of Figure 8.9) mimic the behaviour of the real case, with all the stocks from the DJIA index with a negative sign and almost all the stocks from the CAC40 and the FTSE100 indices with a positive sign. This segregation between the stocks of the DJIA index and the stocks left from the other two indices was already spotted in the MST of Figures 8.1 and 8.3 for the real and random cases, respectively.

In Figure 8.10 we present the eigenvector elements, for each industrial sector and index, of the third and fourth highest eigenvalues, λ_{147} and λ_{146}, for the study of portfolio A.

For portfolio A, the eigenvector corresponding to the third highest eigenvalue, λ_{147} (top left picture of Figure 8.10) shows almost all the stocks from the DJIA index with positive sign, again showing some kind of clustering between the stocks of this index. For the random case, the eigenvector related with the third highest eigenvalue, $\hat{\lambda}_{147}$ (top right picture of Figure 8.10) is very different from that of the real case. This one shows another segregation between stocks of one index and the stocks of the other two, where in this case, the stocks that segregate correspond to stocs from the CAC40 index, all of them with a positive sign, where almost all the stocks from the FTSE100 and the DJIA indices show a negative sign. This behaviour has been shown in the MST for the random case (Figure 8.3) where the stocks from the three different indices were divided in three different groups, each one related with one different index. But this behaviour is not at all what we see in the real case.

For the eigenvector that corresponds to the fourth highest eigenvalue, λ_{146} (bottom left picture of Figure 8.10) almost all stocks from the CAC40 index are positive. This behaviour is very similar with that of the eigenvector related with the third highest eigenvalue for the random case, $\hat{\lambda}_{147}$ (top right picture of Figure 8.10). This result is in agreement with what we stated after looking at the comparison between the IPR, for real and random cases, for the portfolio A.

The eigenvector corresponding to the fourth highest eigenvalue, for the random case, $\hat{\lambda}_{146}$ (bottom right picture of Figure 8.10) shows a random distribution of the values of each stock. There is no pattern for a specific sector or index.

In Figure 8.11 we present the eigenvector elements, for each industrial sector and index, of the two highest eigenvalues, λ_{72} and λ_{71} of both real and random cases, for portfolio B.

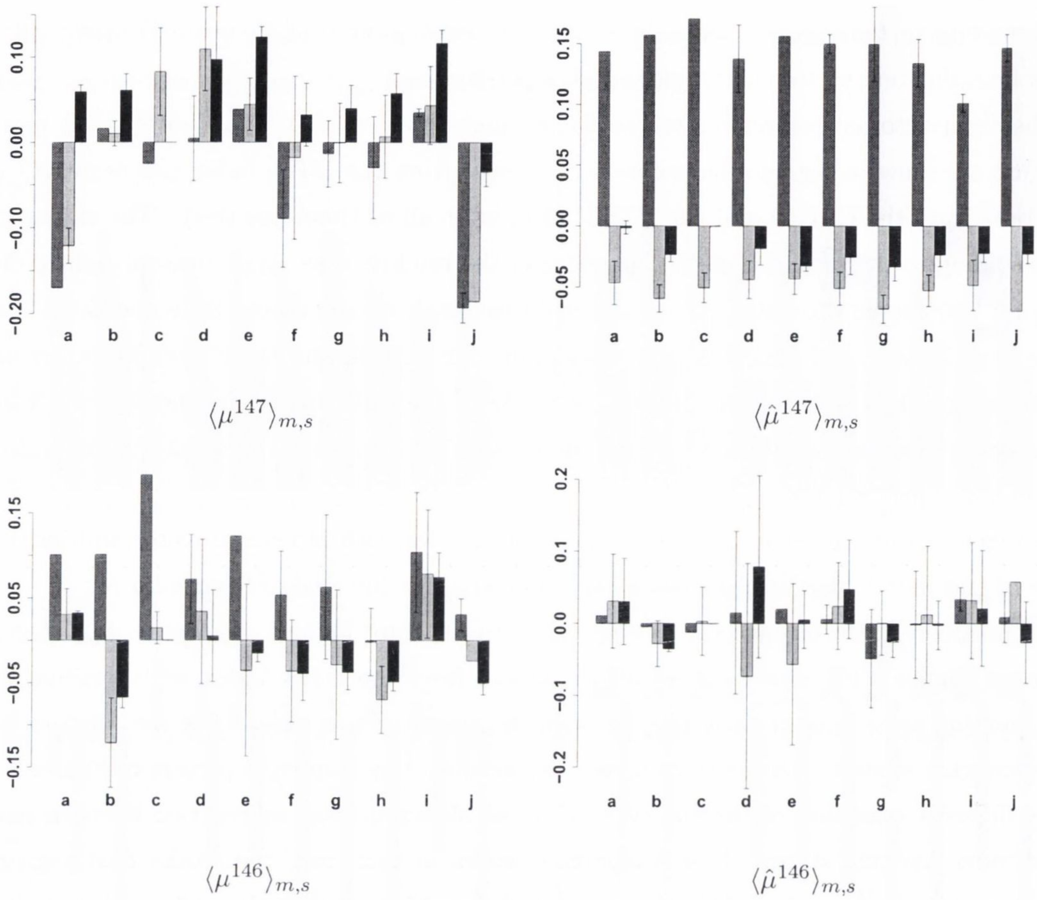


Figure 8.10: Mean value of eigenvector elements, for each industrial sector, of the third and fourth highest eigenvalues, λ_{147} and λ_{146}, for the study of portfolio A. The left figures correspond to the real time series and the right figures to the random time series. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector. Each colour column represent one index: dark grey for the CAC40 index, light grey for the FTSE100 index and black for the DJIA index.

For the eigenvector related with the highest eigenvalue, λ_{72} (top right picture of Figure 8.11) all the elements of different sectors and indices have the same positive sign as shown before for other portfolios. The same behaviour is also presented for the random case, where the eigenvector related with the highest eigenvalue, $\hat{\lambda}_{72}$ (top right picture of Figure 8.11) has all the elements with a positive sign.

For portfolio B, the eigenvector corresponding to the second highest eigenvalue, λ_{71} (bottom

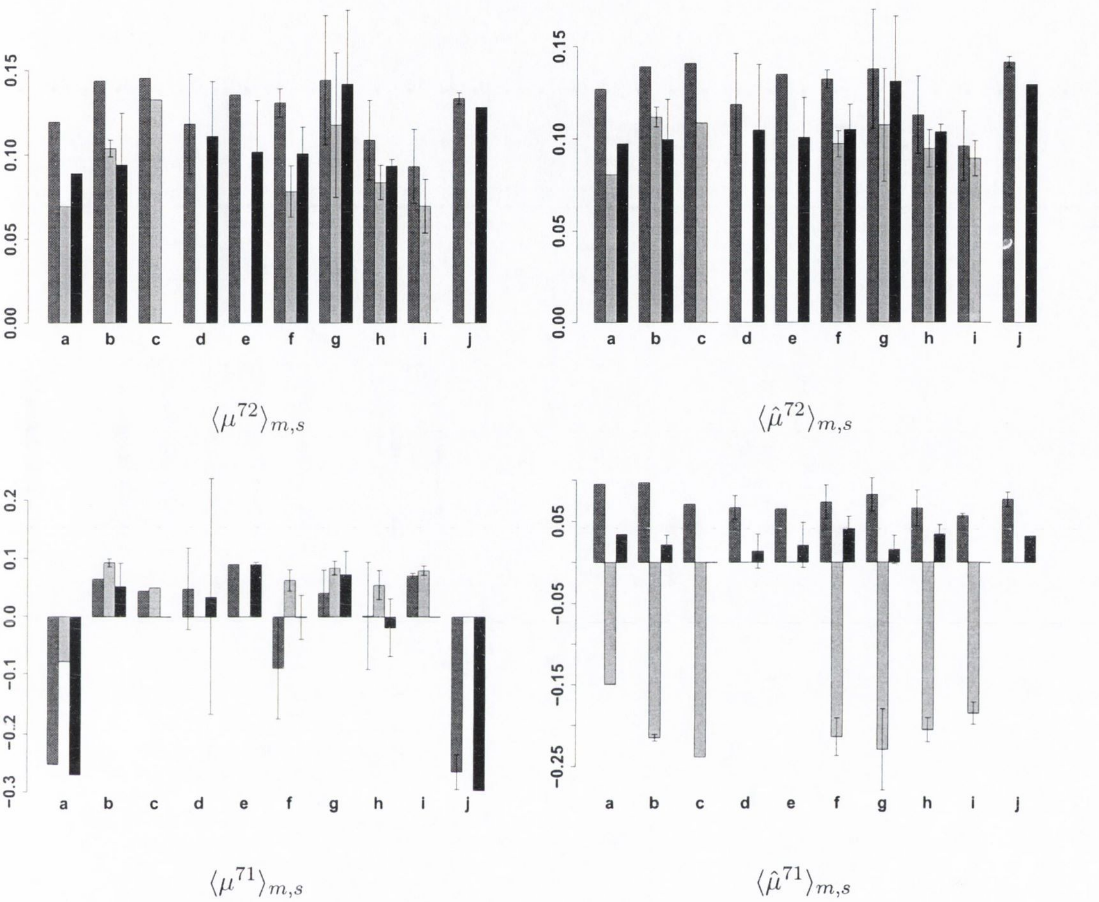


Figure 8.11: Mean value of eigenvector elements, for each industrial sector, of the two highest eigenvalues, λ_{72} and λ_{71}, for the study of portfolio B. The left figures correspond to the real time series and the right figures to the random time series. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector. Each colour column represents one index: dark grey for the CAC40 index, light grey for the BEL20 index and black for the AEX index.

left picture of Figure 8.11) does not show any kind of segregation in terms of index, as shown for portfolio A, but it shows some patterns in terms of industrial sectors. For different indices, almost all the elements for the same industrial sector follow the same sign. For the random case, the eigenvector corresponding to the second highest eigenvalue, $\hat{\lambda}_{71}$ (bottom right picture of Figure 8.11) shows the same behaviour as the second highest eigenvalue of the portfolio A, where all the elements of one index segregate from the elements of the other two indices. In this case, the stocks that segregate are those that belong to the BEL20 index. All these stocks have a negative sign, and almost all the stocks from the CAC40 and the AEX indices have a positive sign. This segregation was already spotted in the MST of Figures 8.2 and 8.4 where the stocks from the BEL20 index usually cluster together.

In Figure 8.12 we present the eigenvector elements, for each industrial sector and index, of the third and fourth highest eigenvalues, λ_{70} and λ_{69}, for the study of portfolio B.

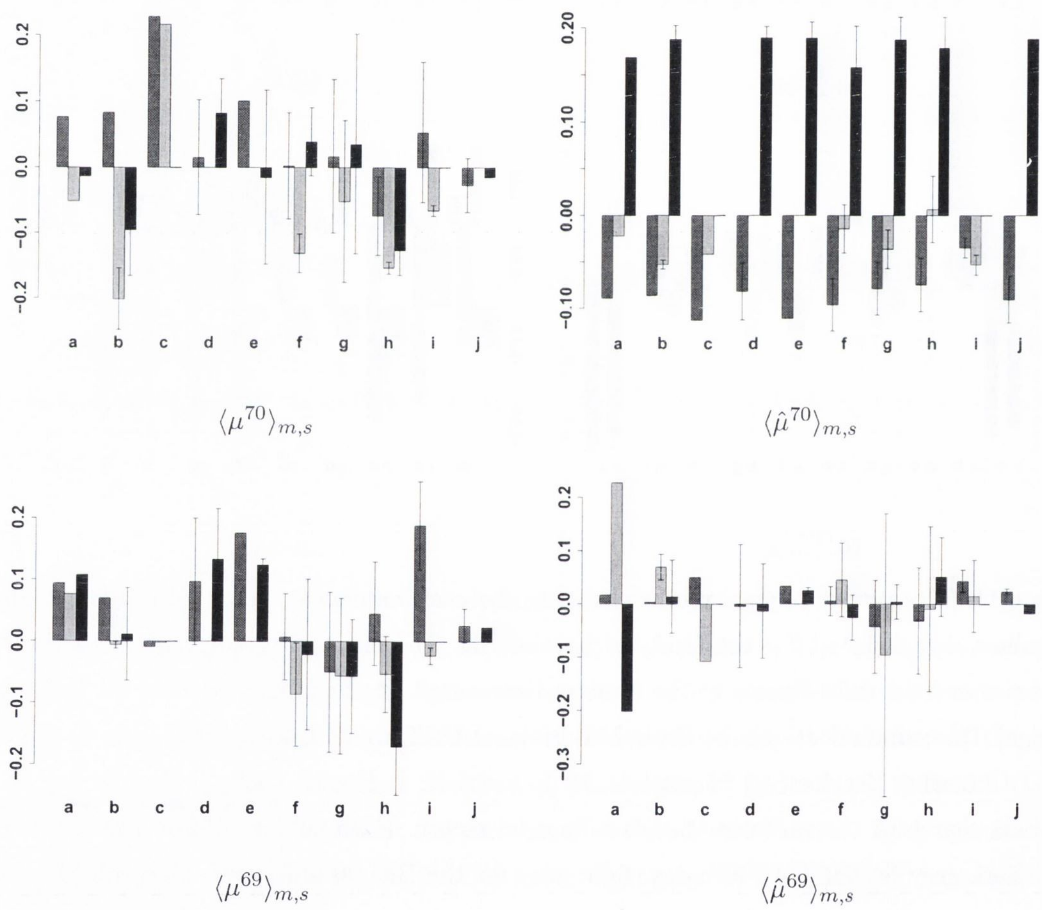


Figure 8.12: Mean value of eigenvector elements, for each industrial sector, of the third and fourth highest eigenvalues, λ_{70} and λ_{69}, for the study of portfolio B. The left figures correspond to the real time series and the right figures to the random time series. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector. Each colour column represents one index: dark grey for the CAC40 index, light grey for the BEL20 index and black for the AEX index.

For the eigenvectors related with the third and fourth highest eigenvalues, λ_{70} (top left picture of Figure 8.12) and λ_{69} (bottom left picture of Figure 8.12), respectively, there seems to be no segregation or clustering in indices or industrial sectors.

The eigenvector corresponding to the third highest eigenvalue, for the random case, $\hat{\lambda}_{70}$ (top right picture of Figure 8.12) has the same behaviour shown for portfolio A with the stocks from the AEX index segregating from the stocks of the CAC40 and the BEL20 indices. All the stocks from the AEX index have a positive sign and almost all the stocks from the CAC40 and the BEL20 indices have a negative sign.

The eigenvector related with the fourth highest eigenvalue, for the random case $\hat{\lambda}_{69}$ (bottom right picture of Figure 8.12) shows a random distribution of the values of each stock. There is no pattern for a specific sector or index.

For portfolio B, the only eigenvector corresponding to an eigenvalue that is higher than λ_{max} and that we haven't presented yet is the one related with the fifth highest eigenvalue. In Figure 8.13, we show the elements of this eigenvector and can see that all the stocks from the BEL20 index have the same feature, a negative sign.

8.7 Multifactor model

A realistic market model needs to incorporate correlations between stocks as shown by J. D. Noh [137]. We use a multifactor market model with terms for each market:

$$r_i(t) = \alpha_{0_i} + \alpha_{1_i} R_{m_1}(t) + \alpha_{2_i} R_{m_2}(t) + \alpha_{3_i} R_{m_3}(t) + \epsilon_i(t) \tag{8.5}$$

where R_{m_j} is the return of index of market m_j and $\epsilon_i(t)$ is a Gaussian distributed random number. The parameters α are estimated by the least square method for multivariate data analysis [138] between the real returns R_i and the indices of each market R_{m_j}:

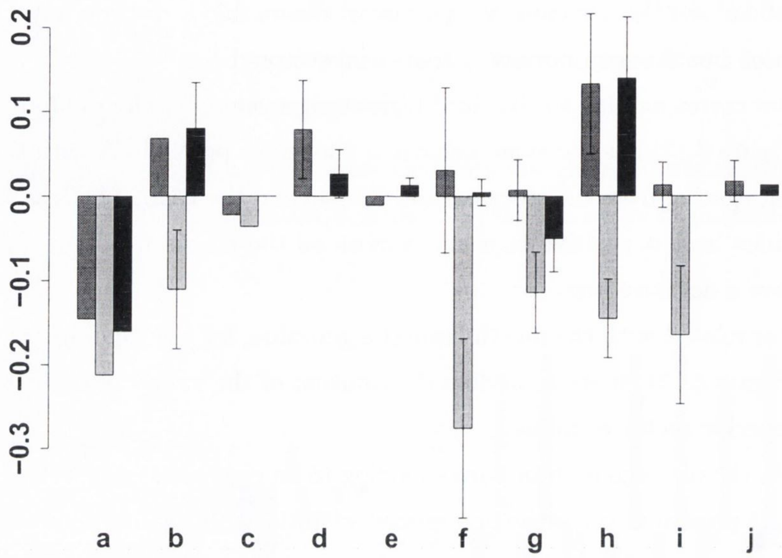
$$\begin{aligned}
\alpha_0 &= \langle R\rangle - \alpha_1\langle R_{m_1}\rangle - \alpha_2\langle R_{m_2}\rangle - \alpha_3\langle R_{m_3}\rangle \\
\alpha_1 &= \frac{\sigma_{X1} - \alpha_2\sigma_{12} - \alpha_3\sigma_{13}}{\sigma_{11}} \\
\alpha_2 &= \frac{\sigma_{X2}\sigma_{11} - \sigma_{X1}\sigma_{12} + \alpha_3\left(\sigma_{12}\sigma_{13} - \sigma_{23}\sigma_{11}\right)}{\sigma_{11}\sigma_{22} - \sigma_{12}{}^2} \\
\alpha_3 &= \frac{\left(\sigma_{11}\sigma_{X3} - \sigma_{X1}\sigma_{13}\right)\left(\sigma_{11}\sigma_{22} - \sigma_{12}{}^2\right) + \left(\sigma_{X2}\sigma_{11} - \sigma_{X1}\sigma_{12}\right)\left(\sigma_{12}\sigma_{13} - \sigma_{23}\sigma_{11}\right)}{\left(\sigma_{11}\sigma_{33} - \sigma_{13}{}^2\right)\left(\sigma_{11}\sigma_{22} - \sigma_{12}{}^2\right) - \left(\sigma_{12}\sigma_{13} - \sigma_{23}\sigma_{11}\right)^2}
\end{aligned}$$

where σ_{ij} is the covariance between indices of two markets:

$$\sigma_{ij} = \langle R_{m_i} R_{m_j}\rangle - \langle R_{m_i}\rangle\langle R_{m_j}\rangle \tag{8.6}$$

and σ_{Xj} is the covariance between return of stock i and index of market j:

$$\sigma_{Xj} = \langle R_i R_{m_j}\rangle - \langle R_i\rangle\langle R_{m_j}\rangle \tag{8.7}$$



$\langle \mu^{68} \rangle_{m,s}$

Figure 8.13: Mean value of eigenvector elements, for each industrial sector, of the fifth highest eigenvalues, λ_{68} for the study of portfolio B. In the x axis we have the industrial sector: a) Telecommunications; b) Basic Materials; c) Utilities; d) Consumer Goods; e) Oil and Gas; f) Consumer Services; g) Financials; h) Industrials; i) Health Care; j) Technology. The error bars represent the variance of each industrial sector. Each colour column represents one index: dark grey for the CAC40 index, light grey for the BEL20 index and black for the AEX index.

We found that the values of α_0 are around zero as expected by the zero mean of stock returns. The values of αs are shown in Tables E.1 and E.2. If the indices of each market were orthogonal each value of α should represent the covariance between the stock and that index of the market. For each stock, depending on which market it belongs to, the highest value of α is the one for the market to which the stock belongs, there are just a few cases for portfolio B where this rule does not apply. The cases are for companies Dexia ($DX.PA$) quoted in the CAC40 index and Corio ($CORA.AS$), Fortis ($FORA.AS$), KPN ($KPN.AS$) and Royal Dutch Shell ($RDSA.AS$) quoted in the AEX index. Probably there are many reasons why these examples don't follow the rule. For example, Dexia is a French-Belgian financial company that results from a merge of two other financial companies, one from Belgium and the other from France. Fortis is also a financial company listed in both the BEL20 and the AEX indices with a strong market in the Benelux region.

Because the exceptions to the rule are few, that is the reason why we decided to use a simple market model that gives the same results as the multifactor model.

8.8 Conclusions

To the best of our knowledge, this is the first time correlations between stocks from three different countries have been studied and compared with the outcome of a market model where the main factor is the mean return of each of the three markets.

The study of the eigensystem of correlations between stocks shows some particular behaviour for each eigenvalue outside the region predicted by the Random Matrix theory. The highest eigenvalue shows the general trend of the market, because all the elements of the eigenvector corresponding to this eigenvalue have positive sign. Other eigenvalues outside the region predicted by Random Matrix theory show some kind of segregation between sectors of different markets. But we see from two different portfolios that there is not a rule about which eigenvalue should represent which market. For the case of a portfolio of stocks from the CAC40, the BEL20 and the AEX indices the segregation between stocks of different markets is very weak. But for a portfolio of stocks from the CAC40, the FTSE100 and the DJIA indices there is a strong segregation between the stocks of the DJIA index and the stocks of the CAC40 and the FTSE100 indices. This same segregation is reproduced in our simple market model. Our market model is also able to reproduce some values of the Inverse Participation Ratio, that serves to quantify the distribution of elements of an eigenvector. For the case of stocks from the CAC40, the FTSE100 and the DJIA indices a good agreement between real and random cases for the values of the first two eigenvalues and the values of the IPR of these eigenvalues is achieved. This might be one of the reasons why the segregation is not so visible for portfolio B, as the IPR for the real and random cases only looks identical for the highest eigenvalue.

Most of the conclusions about the segregation of stocks taken from the analysis of the eigensystem of the correlation matrix was also achieved by the visualisation of the MST and the clusters formed in the MST.

A model with more correlations between stocks of different markets will be necessary to achieve the behaviour that some eigenvalues showed in our study.

Outlook

From Chapters 2 and 3, we saw that the distributions of wealth for the rich people in the society has a different power law exponent than for the billionaires. This conclusion comes from the analysis of different sets of data, but should be important to have access to a data set of the distribution of wealth that includes these two groups of the society.

In Chapter 3 we introduced a model able to recreate a double power law for the wealth distributions. This double power law regime is due to the exchange rules adopted in this model where the richest agents exchange a less percentage of their money compared with the poorest agents. Using an analytical solution derived from a generalised function we fitted the distribution of wealth from the simulations of our model successfully. A better understanding of the relation between the values of parameters chosen in the model and the exponents of the power laws should be a goal for the future. In addition, the search for an analytical solution of this model should form the basis for further research.

In Chapters 5, 6, 7 and 8 we conclude that companies from the same stock market cluster in terms of industrial sectors. The correlation between stocks from the same industrial sectors is very high. Studying various portfolios of stocks we concluded that, depending on the market, the stocks cluster first in terms of market and then in terms of industrial sector. For example, in the study of the cross-correlations between stocks from the FTSE100 and DJIA indices, the stocks from one market segregated from the stocks of the other market, but in each one of these clusters, the stocks cluster in terms of industrial sectors. For a portfolio of stocks from European markets, the segregation of stocks in terms of market is not so clear. There are some stocks from different markets with higher correlation with stocks from other markets than with stocks from the same market. From a portfolio of stocks from CAC40, BEL20 and AEX indices we saw that many stocks from different markets cluster together in the same industrial sector. A further study should include intraday data, to explain better the time-mismatch problem of stocks from markets with different closing times, as for example, FTSE100 and DJIA.

The results are qualified based on the measures used in this thesis. Based on correlation map further information can be obtained from complex network theory.

Appendix A

Computation of parameters of T-student distribution

The fractional moment of a T-student distribution is given as:

$$M_f = 2\int_0^\infty x^f P_k(x)dx \tag{A-1}$$

where $P_k(x)$ is the T-student probability distribution function of equation 4.4. We can write the equation of moments as:

$$M_f = \frac{2N_k}{\sqrt{2\pi\sigma_k^2}}\int_0^\infty x^f \left[\frac{\frac{x^2}{2\sigma_k^2 k}}{1+\frac{x^2}{2\sigma_k^2 k}}\right]^k \left[\frac{2\sigma_k^2 k}{x^2}\right]^k dx \tag{A-2}$$

Using a changing of variables where:

$$z = \frac{\frac{x^2}{2\sigma_k^2 k}}{1+\frac{x^2}{2\sigma_k^2 k}} \tag{A-3}$$

$$x = \sqrt{2\sigma_k^2 k\frac{z}{1-z}} \tag{A-4}$$

$$dx = \sqrt{2\sigma_k^2 k}(1-z)^{-3/2}z^{-1/2}dz \tag{A-5}$$

the fractional moments can be written as:

$$M_f = \frac{2N_k}{\sqrt{2\pi\sigma_k^2}}\left[2\sigma_k^2 k\right]^{\frac{f+1}{2}}\int_0^1 z^{\frac{f-1}{2}}(1-z)^{k-\frac{f+3}{2}}dz \tag{A-6}$$

where the integral can be solved using the Beta function:

$$B(m+1,n+1) = \int_0^1 u^m(1-u)^n du = \frac{m!n!}{(m+n+1)!} = \frac{\Gamma(m+1)\Gamma(n+1)}{\Gamma(m+n+2)} \tag{A-7}$$

In the case of the rate of moments (equation 4.10), dividing the fractional moments, M_{f-1} by M_{f+1} we now obtain:

$$R_f = \frac{M_{f-1}}{M_{f+1}} = \frac{1}{2\sigma_k^2 k} \frac{B\left[\frac{f}{2}, k - \frac{f}{2}\right]}{B\left[\frac{f}{2} + 1, k - \frac{f}{2} - 1\right]} = \frac{1}{2\sigma_k^2 k} \left[\frac{2(k-1)}{f} - 1\right] \tag{A-8}$$

Appendix B

Stocks that belong to FTSE100 index at June 30^{th} 2005

Table B.1: Name, code and sector classification for the Industry Classification Benchmark (ICB) and the Global Classification System (GCS) for 102 stocks that belonged to FTSE100 index at June 30^{th} 2005.

Name	code	ICB sector	GCS sector
Anglo American	AAL	Basic materials	Resources
Associated British Foods	ABF	Consumer goods	Non-cyclical consumer goods
Alliance & Leicester	AL.	Financials	Financials
Allied Domecq	ALLD	Consumer goods	Non-cyclical consumer goods
Antofagasta	ANTO	Basic materials	Resources
Aliiance Unichem	AUN	Consumer services	Non-cyclical consumer goods
Aviva	AV.	Financials	Financials
Amvescap	AVZ	Financials	Financials
AstraZeneca	AZN	Health care	Non-cyclical consumer goods
BAE Systems	BA.	Industrials	General industrials
BAA	BAA	Industrials	Cyclical services
Barclays	BARC	Financials	Financials
British American Tobacco	BATS	Consumer goods	Non-cyclical consumer goods
British Airways	BAY	Consumer services	Cyclical services
BG Group	BG.	Oil and gas	Resources
British Land Co	BLND	Financials	Financials
BHP Billiton	BLT	Basic materials	Resources
Bunzl	BNZL	Industrials	Cyclical services
			Continue on next page

– continue from the previous page

Name	code	ICB sector	GCS sector
BOC Group	BOC	Basic materials	Basic industries
Boots Group	BOOT	Consumer services	Cyclical services
BP	BP.	Oil and gas	Resources
British Sky Broadcasting Group	BSY	Consumer services	Cyclical services
BT Group	BT.A	Telecommunications	Non-cyclical services
Cadbury	CBRY	Consumer goods	Non-cyclical consumer goods
Carnival	CCL	Consumer services	Cyclical services
Centrica	CNA	Utilities	Utilities
Cairn Energy	CNE	Oil and gas	Resources
Compass Group	CPG	Consumer services	Cyclical services
Capita Group	CPI	Industrials	Cyclical services
Corus Group	CS.	Basic materials	Basic industries
Cable & Wireless	CW.	Telecommunications	Non-cyclical services
Diageo	DGE	Consumer goods	Non-cyclical consumer goods
Daily Mail & General Trust	DMGT	Consumer services	Cyclical services
Dixons Group	DXNS	Consumer services	Cyclical services
EMAP	EMA	Consumer services	Cyclical services
Man Group	EMG	Financials	Financials
Enterprise Inns	ETI	Consumer services	Cyclical services
Exel	EXL	Industrials	Cyclical services
Friends Provident	FP.	Financials	Financials
Gallagher Group	GLH	Consumer goods	Non-cyclical consumer goods
GlaxoSmithKline	GSK	Health care	Non-cyclical consumer goods
GUS	GUS	Consumer services	Cyclical services
Hays	HAS	Industrials	Cyclical services
HBOS	HBOS	Financials	Financials
Hilton Group	HG.	Consumer services	Cyclical services
Hanson	HNS	Industrials	Basic industries
HSBC Hldgs	HSBA	Financials	Financials
InterContinental Hotels Group	IHG	Consumer services	Cyclical services
3i Group	III	Financials	Financials
Imperial Tobacco Group	IMT	Consumer goods	Non-cyclical consumer goods
International Power	IPR	Utilities	Utilities
ITV	ITV	Consumer services	Cyclical services
Johnson Matthey	JMAT	Basic materials	Basic industries
Kingfisher	KGF	Consumer services	Cyclical services
Land Securities Group	LAND	Financials	Financials
Legal & General Group	LGEN	Financials	Financials

Continue on next page

– continue from the previous page

Name	code	ICB sector	GCS sector
Liberty International	LII	Financials	Financials
Lloyds TSB Group	LLOY	Financials	Financials
Marks & Spencer Group	MKS	Consumer services	Cyclical services
Morrison (Wm) Supermarkets	MRW	Consumer services	Non-cyclical services
National Grid Transco	NGT	Utilities	Utilities
Northern Rock	NRK	Financials	Financials
Next	NXT	Consumer services	Cyclical services
Old Mutual	OML	Financials	Financials
O2	OOM	Telecommunications	Non-cyclical services
Prudential	PRU	Financials	Financials
Pearson	PSON	Consumer services	Cyclical services
Reckitt Benckiser Group	RB.	Consumer goods	Non-cyclical consumer goods
Royal Bank of Scotland Group	RBS	Financials	Financials
Reed Elsevier	REL	Consumer services	Cyclical services
Rexam	REX	Industrials	Cyclical services
Rio Tinto	RIO	Basic materials	Resources
Rolls-Royce Group	RR.	Industrials	General industrials
Royal & Sun Alliance Insurance Group	RSA	Financials	Financials
Rentokil Initial	RTO	Industrials	Cyclical services
Reuters Group	RTR	Consumer services	Cyclical services
SABMiller	SAB	Consumer goods	Non-cyclical consumer goods
Sainsbury (J)	SBRY	Consumer services	Non-cyclical services
Scottish & Newcastle	SCTN	Consumer goods	Non-cyclical consumer goods
Schroders	SDR	Financials	Financials
Schroders N/V	SDRC	Financials	Financials
Sage Group	SGE	Technology	Information Technology
Shell Transport & Trading Co	SHEL	Oil and gas	Resources
Shire	SHP	Health care	Non-cyclical consumer goods
Smiths Group	SMIN	Industrials	General industrials
Smith & Nephew	SN.	Health care	Non-cyclical consumer goods
Scottish Power	SPW	Utilities	Utilities
Scottish & Southern Energy	SSE	Utilities	Utilities
Standard Chartered	STAN	Financials	Financials
Severn Trent	SVT	Utilities	Utilities
Tate & Lyle	TATE	Consumer goods	Non-cyclical consumer goods
Tesco	TSCO	Consumer services	Non-cyclical services
Unilever	ULVR	Consumer goods	Non-cyclical consumer goods
United Utilities	UU.	Utilities	Utilities

Continue on next page

– concluded from previous page

Name	code	ICB sector	GCS sector
United Utilities A Shares	UU.A	Utilities	Utilities
Vodafone Group	VOD	Telecommunications	Non-cyclical services
William Hill	WMH	Consumer services	Cyclical services
Wolseley	WOS	Industrials	Basic industries
WPP Group	WPP	Consumer services	Cyclical services
Whitbread	WTB	Consumer services	Cyclical services
Xstrata	XTA	Basic materials	Resources
Yell Group	YELL	Consumer services	Cyclical services

Appendix C

Stocks that belong to the FTSE100, the Dow Jones Industrial Average (DJIA), the *Cotation Assistée en Continu* (CAC) 40, the BEL20 and the Amsterdam Exchange Index (AEX) indices at April 1^{st} 2008

Table C.1: Name, code and ICB sector classification for 102 stocks that belonged to the FTSE100 index at April 1^{st} 2008.

Name	code	ICB sector
3I GROUP	III.L	Financials
ADMIRAL GROUP	ADM.L	Financials
ALLIANCE & LEICESTER	AL.L	Financials
ALLIANCE TRUST	ATST.L	Financials
AMEC	AMEC.L	Oil and Gas
ANGLO AMERICAN	AAL.L	Basic Materials
ANTOFAGASTA	ANTO.L	Basic Materials
ASSOCIATED BRIT.FOODS	ABF.L	Consumer Goods

Continue on next page

– continue from the previous page

Name	code	ICB sector
ASTRAZENECA	AZN.L	Health Care
AVIVA	AV.L	Financials
BAE SYSTEMS	BA.L	Industrials
BARCLAYS	BARC.L	Financials
BG GROUP	BG.L	Oil and Gas
BHP BILLITON	BLT.L	Basic Materials
BP	BP.L	Oil and Gas
BRITISH AIRWAYS	BAY.L	Consumer Services
BRITISH AMERICAN TOBACCO	BATS.L	Consumer Goods
BRITISH ENERGY GROUP	BGY.L	Utilities
BRITISH LAND	BLND.L	Financials
BRITISH SKY BCAST.GROUP	BSY.L	Consumer Services
BT GROUP	BT-A.L	Telecommunications
BUNZL	BNZL.L	Industrials
CABLE & WIRELESS	CW.L	Telecommunications
CADBURY	CBRY.L	Consumer Goods
CAIRN ENERGY	CNE.L	Oil and Gas
CAPITA GROUP	CPI.L	Industrials
CARNIVAL	CCL.L	Consumer Services
CARPHONE WHSE.GP.	CPW.L	Consumer Services
CENTRICA	CNA.L	Utilities
COBHAM	COB.L	Industrials
COMPASS GROUP	CPG.L	Consumer Services
DIAGEO	DGE.L	Consumer Goods
ENTERPRISE INNS	ETI.L	Consumer Services
EURASIAN NATRES.CORP.	ENRC.L	Basic Materials
EXPERIAN GROUP	EXPN.L	Industrials
FIRST GROUP	FGP.L	Consumer Services
FRIENDS PROVIDENT	FP.L	Financials
G4S	GFS.L	Industrials
GLAXOSMITHKLINE	GSK.L	Health Care
HAMMERSON	HMSO.L	Financials
HBOS	HBOS.L	Financials
HOME RETAIL GROUP	HOME.L	Consumer Services
HSBC HDG. (ORD $0.50)	HSBA.L	Financials
ICAP	IAP.L	Financials
ICTL.HTLS.GP.	IHG.L	Consumer Services
IMPERIAL TOBACCO GP.	IMT.L	Consumer Goods

Continue on next page

– continue from the previous page

Name	code	ICB sector
INTERNATIONAL POWER	IPR.L	Utilities
ITV	ITV.L	Consumer Services
JOHNSON MATTHEY	JMAT.L	Basic Materials
KAZAKHMYS	KAZ.L	Basic Materials
KINGFISHER	KGF.L	Consumer Services
LAND SECURITIES GROUP	LAND.L	Financials
LEGAL & GENERAL	LGEN.L	Financials
LIBERTY INTL.	LII.L	Financials
LLOYDS TSB GROUP	LLOY.L	Financials
LONDON STOCK EX.GROUP	LSE.L	Financials
LONMIN	LMI.L	Basic Materials
MAN GROUP	EMG.L	Financials
MARKS & SPENCER GROUP	MKS.L	Consumer Services
MORRISON(WM)SPMKTS.	MRW.L	Consumer Services
NATIONAL GRID	NG.L	Utilities
NEXT	NXT.L	Consumer Services
OLD MUTUAL	OML.L	Financials
PEARSON	PSON.L	Consumer Services
PERSIMMON	PSN.L	Consumer Goods
PRUDENTIAL	PRU.L	Financials
RECKITT BENCKISER	RB.L	Consumer Goods
REED ELSEVIER	REL.L	Consumer Services
REXAM	REX.L	Industrials
RIO TINTO	RIO.L	Basic Materials
ROLLS-ROYCE GROUP	RR.L	Industrials
ROYAL & SUN ALL.IN.	RSA.L	Financials
ROYAL BANK OF SCTL.GP.	RBS.L	Financials
ROYAL DUTCH SHELL A(LON)	RDSA.L	Oil and Gas
ROYAL DUTCH SHELL B	RDSB.L	Oil and Gas
SABMILLER	SAB.L	Consumer Goods
SAGE GROUP	SGE.L	Technology
SAINSBURY (J)	SBRY.L	Consumer Services
SCHRODERS	SDR.L	Financials
SCHRODERS NV	SDRC.L	Financials
SCOT.& SOUTHERN ENERGY	SSE.L	Utilities
SEVERN TRENT	SVT.L	Utilities
SHIRE	SHP.L	Health Care
SMITHS GROUP	SMIN.L	Industrials

Continue on next page

– concluded from previous page

Name	code	ICB sector
SMITH & NEPHEW	SN.L	Health Care
STANDARD CHARTERED	STAN.L	Financials
STANDARD LIFE	SL.L	Financials
TATE & LYLE	TATE.L	Consumer Goods
TESCO	TSCO.L	Consumer Services
THOMAS COOK GROUP	TCG.L	Consumer Services
THOMSON REUTERS	TRIL.L	Consumer Services
TUI TRAVEL	TT.L	Consumer Services
TULLOW OIL	TLW.L	Oil and Gas
UNILEVER (UK)	ULVR.L	Consumer Goods
UNITED UTILITIES	UU.L	Utilities
VEDANTA RESOURCES	VED.L	Basic Materials
VODAFONE GROUP	VOD.L	Telecommunications
WHITBREAD	WTB.L	Consumer Services
WOLSELEY	WOS.L	Industrials
WOOD GROUP (JOHN)	WG.L	Oil and Gas
WPP GROUP	WPP.L	Consumer Services
XSTRATA	XTA.L	Basic Materials

Table C.2: Name, code and ICB sector classification for 30 stocks that belonged to the DJIA index at April 1^{st} 2008.

Name	code	ICB sector
3M	MMM	Industrials
ALCOA	AA	Basic Materials
AMERICAN EXPRESS	AXP	Financials
AMERICAN INTL.GP.	AIG	Financials
AT&T	T	Telecommunications
BANK OF AMERICA	BAC	Financials
BOEING	BA	Industrials
CATERPILLAR	CAT	Industrials
CHEVRON	CVX	Oil and Gas
CITIGROUP	C	Financials
COCA COLA	KO	Consumer Goods
E I DU PONT DE NEMOURS	DD	Basic Materials
EXXON MOBIL	XOM	Oil and Gas
GENERAL ELECTRIC	GE	Industrials
GENERAL MOTORS	GM	Consumer Goods
HEWLETT-PACKARD	HPQ	Technology
HOME DEPOT	HD	Consumer Services
INTEL	INTC	Technology

Continue on next page

– concluded from previous page

Name	code	ICB sector
INTERNATIONAL BUS.MACH.	IBM	Technology
JOHNSON & JOHNSON	JNJ	Health Care
JP MORGAN CHASE & CO.	JPM	Financials
MCDONALDS	MCD	Consumer Services
MERCK & CO.	MRK	Health Care
MICROSOFT	MSFT	Technology
PFIZER	PFE	Health Care
PROCTER & GAMBLE	PG	Consumer Goods
UNITED TECHNOLOGIES	UTX	Industrials
VERIZON COMMS.	VZ	Telecommunications
WAL MART STORES	WMT	Consumer Services
WALT DISNEY	DIS	Consumer Services

Table C.3: Name, code and ICB sector classification for 40 stocks that belonged to the CAC40 index at April 1^{st} 2008.

Name	code	ICB sector
ACCOR	AC.PA	Consumer Services
AIR FRANCE-KLM	AF.PA	Consumer Services
AIR LIQUIDE	AI.PA	Basic Materials
ALCATEL-LUCENT	ALU.PA	Technology
ALSTOM	ALO.PA	Industrials
ARCELORMITTAL (PAR)	MTP.PA	Basic Materials
AXA	CS.PA	Financials
BNP PARIBAS	BNP.PA	Financials
BOUYGUES	EN.PA	Industrials
CAP GEMINI	CAP.PA	Technology
CARREFOUR	CA.PA	Consumer Services
CREDIT AGRICOLE	ACA.PA	Financials
DANONE	BN.PA	Consumer Goods
DEXIA (PAR)	DX.PA	Financials
EADS (PAR)	EAD.PA	Industrials
EDF	EDF.PA	Utilities
ESSILOR INTL.	EF.PA	Health Care
FRANCE TELECOM	FTE.PA	Telecommunications
GAZ DE FRANCE	GAZ.PA	Utilities
L'OREAL	OR.PA	Consumer Goods
LAFARGE	LG.PA	Industrials
LAGARDERE GROUPE	MMB.PA	Consumer Services
LVMH	MC.PA	Consumer Goods

Continue on next page

– concluded from previous page

Name	code	ICB sector
MICHELIN	ML.PA	Consumer Goods
PERNOD-RICARD	RI.PA	Consumer Goods
PEUGEOT	UG.PA	Consumer Goods
PPR	PP.PA	Consumer Services
RENAULT	RNO.PA	Consumer Goods
SAINT GOBAIN	SGO.PA	Industrials
SANOFI-AVENTIS	SAN.PA	Health Care
SCHNEIDER ELECTRIC	SU.PA	Industrials
SOCIETE GENERALE	GLE.PA	Financials
STMICROELECTRONICS (PAR)	STM.PA	Technology
SUEZ	SZE.PA	Utilities
TOTAL	FP.PA	Oil and Gas
UNIBAIL-RODAMCO	UL.PA	Financials
VALLOUREC	VK.PA	Industrials
VEOLIA ENVIRONNEMENT	VIE.PA	Utilities
VINCI (EX SGE)	DG.PA	Industrials
VIVENDI	VIV.PA	Consumer Services

Table C.4: Name, code and ICB sector classification for 20 stocks that belonged to the BEL20 index at April 1^{st} 2008.

Name	code	ICB sector
ACKERMANS	ACKB.BR	Industrials
AGFA-GEVAERT	AGFB.BR	Industrials
BEKAERT	BEKB.BR	Industrials
BELGACOM	BELG.BR	Telecommunications
COFINIMMO	COFB.BR	Financials
COLRUYT	COL.BR	Consumer Services
DELHAIZE	DELB.BR	Consumer Services
DEXIA	DEXB.BR	Financials
FORTIS (BRU)	FORB.BR	Financials
GBL NEW	GBL.BR	Financials
INBEV	INB.BR	Consumer Goods
KBC GROUPE	KBC.BR	Financials
MOBISTAR	MOBB.BR	Telecommunications
NPM-CIE.NAT.PORTEFEUILLE	NAT.BR	Financials
NYRSTAR (WI)	NYR.BR	Basic Materials
OMEGA PHARMA	OME.BR	Health Care
SOLVAY	SOLB.BR	Basic Materials
SUEZ (BRU)	SZEB.BR	Utilities
UCB	UCB.BR	Health Care

Continue on next page

– concluded from previous page

Name	code	ICB sector

Table C.5: Name, code and ICB sector classification for 24 stocks that belonged to the AEX index at April 1^{st} 2008.

Name	code	ICB sector
AEGON	AGN.AS	Financials
AHOLD KON.	AH.AS	Consumer Services
AKZO NOBEL	AKZA.AS	Basic Materials
ARCELORMITTAL	MT.AS	Basic Materials
ASML HOLDING	ASML.AS	Technology
CORIO	CORA.AS	Financials
CORPORATE EXPRESS	CXP.AS	Industrials
DSM KONINKLIJKE	DSM.AS	Basic Materials
FORTIS	FORA.AS	Financials
HEINEKEN	HEIA.AS	Consumer Goods
ING GROEP	INGA.AS	Financials
KPN KON	KPN.AS	Telecommunications
PHILIPS ELTN.KONINKLIJKE	PHIA.AS	Consumer Goods
RANDSTAD HOLDING	RAND.AS	Industrials
REED ELSEVIER (AMS)	REN.AS	Consumer Services
ROYAL DUTCH SHELL A	RDSA.AS	Oil and Gas
SBM OFFSHORE	SBMO.AS	Oil and Gas
TELE ATLAS	TA.AS	Consumer Services
TNT	TNT.AS	Industrials
TOM TOM	TOM2.AS	Technology
UNIBAIL-RODAMCO (AMS)	ULA.AS	Financials
UNILEVER CERTS.	UNA.AS	Consumer Goods
VEDIOR	VDOR.AS	Industrials
WOLTERS KLUWER	WKL.AS	Consumer Services

Appendix D

Indices from around the world

Table D.1: Country, index and symbol from european indices.

Country	Index	Symbol	Country	Index	Symbol
PORTUGAL	PSI20	PRT	GREECE	ASE20	GRC
U.K.	FTSE100	UK	POLAND	WIG20	POL
GERMANY	DAX30	GER	CZECH REPUBLIC	PX50	CZK
FRANCE	CAC40	FRA	RUSSIA	RSMTIND	RUS
SPAIN	IBEX35	ESP	HUNGARY	BUX	HUN
SWITZERLAND	SWISSMI	CHF	ROMANIA	BET	ROM
ITALY	MIB30	ITA	UKRAINE	KPDRAG	UKR
IRELAND	ISEQ20	IRE	SLOVAKIA	SAX16	SVK
ICELAND	ICEX15	ICE	CROATIA	CROBEX	CRT
NETHERLANDS	AEX	NLD	SLOVENIA	SBI20	SVN
BELGIUM	BEL20	BEL	ESTONIA	OMXT	EST
LUXEMBOURG	LUXX	LUX	LATVIA	OMXR	LAT
DENMARK	OMXC20	DNK	LITHUANIA	OMXV	LTU
FINLAND	OMXH25	FIN	BULGARIA	SOfFIX	BUL
NORWAY	OBX25	NOR	TURKEY	NAT30	TUK
SWEDEN	OMXS30	SWE	MALTA	MALTAIX	MTA
AUSTRIA	ATX	AUT			

Table D.2: Country, index and symbol from middle east and african indices.

Country	Index	Symbol	Country	Index	Symbol
SOUTH AFRICA	JSE40	SAF	KUWAIT	KIC	KUW
EGYPT	EFG	EGY	ISRAEL	MAOF25	ISR
MOROCCO	CFG25	MOR	LEBANON	BLOM	LEB
TUNISIA	TUNIN	TUN	BAHRAIN	BHSE	BAH
NIGERIA	NIGALSH	NIG	JORDAN	AMMAN	JOR
KENYA	NSE20	KEN	OMAN	OMANMSM	OMA

Table D.3: Country, index and symbol from american indices.

Country	Index	Symbol	Country	Index	Symbol
U.S.A.	DJIA	USA	CHILE	IGPAGEN	CHL
CANADA	TSESP60	CAN	VENEZUELA	VENGENL	VEZ
MEXICO	IMC30	MEX	PERU	ISP15	PER
ARGENTINA	ARGSIBI	ARG	COLOMBIA	IGBC	COL
BRAZIL	BOVESPA	BRZ			

Table D.4: Country, index and symbol from asian-pacific indices.

Country	Index	Symbol	Country	Index	Symbol
JAPAN	NIKKEI225	JAP	THAILAND	SET	THI
HONG KONG	HANGSENG	HK	INDONESIA	LQ45	IDO
CHINA	SH180	CH	INDIA	BSE200	IND
TAIWAN	TSEC50	TWA	SINGAPORE	STI	SGR
SOUTH KOREA	KRX100	SOK	MALAYSIA	KLPCOMP	MAL
AUSTRALIA	S&PASX100	AUS	PHILIPPINES	PSEi	PHI
NEW ZEALAND	NZ50	NEZ	VIETNAM	FTSEVI	VIE
PAKISTAN	PKSE100	PAK	BANGLADESH	DSE20	BAN
SRI LANKA	MILANKA	SRL			

Appendix E

Tables of parameters α of the least square method for multivariate analysis between the real returns and the indices of each market

Table E.1: Symbol, market the stock belongs to (m_j), parameters α of the least square method for multivariate analysis and market the stock has the highest correlation with (m'_j), for a portfolio of stocks from the CAC40, FTSE100 and DJIA indices.

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
AC.PA	CAC40	0.00	0.98	0.05	-0.05	CAC40
AF.PA	CAC40	0.00	1.06	0.12	0.05	CAC40
AI.PA	CAC40	0.00	0.78	0.11	0.02	CAC40
ALU.PA	CAC40	0.00	2.03	-0.54	0.12	CAC40
ALO.PA	CAC40	0.00	1.72	-0.26	-0.18	CAC40
CS.PA	CAC40	0.00	1.32	0.11	0.11	CAC40
BNP.PA	CAC40	0.00	0.97	0.14	0.11	CAC40
EN.PA	CAC40	0.00	1.25	-0.23	-0.06	CAC40
CAP.PA	CAC40	0.00	1.71	-0.29	-0.04	CAC40
CA.PA	CAC40	0.00	0.84	0.03	0.06	CAC40
BN.PA	CAC40	0.00	0.54	0.05	-0.01	CAC40
DX.PA	CAC40	0.00	0.74	0.20	0.19	CAC40
EF.PA	CAC40	0.00	0.45	0.17	-0.15	CAC40

Continue on next page

– continue from the previous page

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
FTE.PA	CAC40	0.00	1.61	-0.46	0.06	CAC40
OR.PA	CAC40	0.00	0.83	-0.01	0.02	CAC40
LG.PA	CAC40	0.00	0.85	0.15	-0.03	CAC40
MMB.PA	CAC40	0.00	1.08	-0.15	-0.07	CAC40
MC.PA	CAC40	0.00	1.06	0.06	0.05	CAC40
ML.PA	CAC40	0.00	0.73	0.22	0.05	CAC40
RI.PA	CAC40	0.00	0.52	0.03	-0.18	CAC40
UG.PA	CAC40	0.00	0.79	0.10	0.05	CAC40
PP.PA	CAC40	0.00	1.07	-0.02	0.01	CAC40
RNO.PA	CAC40	0.00	0.88	0.12	0.13	CAC40
SGO.PA	CAC40	0.00	1.10	-0.01	-0.03	CAC40
SAN.PA	CAC40	0.00	0.76	-0.01	0.04	CAC40
SU.PA	CAC40	0.00	0.83	0.12	-0.02	CAC40
GLE.PA	CAC40	0.00	1.05	0.13	0.15	CAC40
STM.PA	CAC40	0.00	1.53	-0.40	0.19	CAC40
SZE.PA	CAC40	0.00	1.18	-0.09	-0.05	CAC40
FP.PA	CAC40	0.00	0.59	0.22	0.01	CAC40
UL.PA	CAC40	0.00	0.35	0.31	-0.12	CAC40
VK.PA	CAC40	0.00	0.79	0.09	-0.17	CAC40
DG.PA	CAC40	0.00	0.52	0.19	-0.15	CAC40
VIV.PA	CAC40	0.00	1.49	-0.24	-0.12	CAC40
AL.L	FTSE100	0.00	-0.11	1.25	0.07	FTSE100
ATST.L	FTSE100	0.00	0.06	0.73	-0.02	FTSE100
AMEC.L	FTSE100	0.00	-0.09	0.91	-0.04	FTSE100
AAL.L	FTSE100	0.00	0.13	1.30	0.08	FTSE100
ANTO.L	FTSE100	0.00	-0.13	1.16	-0.11	FTSE100
ABF.L	FTSE100	0.00	-0.15	0.87	-0.03	FTSE100
AZN.L	FTSE100	0.00	0.02	0.85	0.03	FTSE100
AV.L	FTSE100	0.00	0.14	1.22	0.10	FTSE100
BA.L	FTSE100	0.00	-0.10	1.04	0.06	FTSE100
BARC.L	FTSE100	0.00	0.07	1.22	0.12	FTSE100
BG.L	FTSE100	0.00	-0.14	1.09	-0.06	FTSE100
BLT.L	FTSE100	0.00	0.09	1.26	0.01	FTSE100
BP.L	FTSE100	0.00	0.06	0.76	0.04	FTSE100
BAY.L	FTSE100	0.00	0.41	1.02	0.16	FTSE100
BATS.L	FTSE100	0.00	-0.26	0.87	-0.01	FTSE100
BLND.L	FTSE100	0.00	-0.20	1.09	0.09	FTSE100
BSY.L	FTSE100	0.00	0.38	0.80	-0.04	FTSE100

Continue on next page

– continue from the previous page

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
BT-A.L	FTSE100	0.00	0.26	0.78	0.07	FTSE100
BNZL.L	FTSE100	0.00	-0.08	0.81	-0.01	FTSE100
CW.L	FTSE100	0.00	0.37	0.95	-0.05	FTSE100
CBRY.L	FTSE100	0.00	-0.14	0.92	-0.05	FTSE100
CNE.L	FTSE100	0.00	-0.24	1.09	-0.15	FTSE100
CPI.L	FTSE100	0.00	0.34	0.79	-0.02	FTSE100
CNA.L	FTSE100	0.00	-0.20	1.08	-0.06	FTSE100
COB.L	FTSE100	0.00	-0.10	0.82	-0.05	FTSE100
DGE.L	FTSE100	0.00	-0.11	0.80	0.01	FTSE100
ETI.L	FTSE100	0.00	-0.15	0.86	-0.12	FTSE100
FGP.L	FTSE100	0.00	-0.28	1.14	-0.11	FTSE100
GFS.L	FTSE100	0.00	-0.16	1.01	-0.04	FTSE100
GSK.L	FTSE100	0.00	-0.02	0.72	0.11	FTSE100
HMSO.L	FTSE100	0.00	-0.23	1.02	-0.03	FTSE100
HBOS.L	FTSE100	0.00	-0.06	1.36	0.02	FTSE100
HOME.L	FTSE100	0.00	-0.15	1.17	0.13	FTSE100
HSBA.L	FTSE100	0.00	0.11	0.72	0.11	FTSE100
IAP.L	FTSE100	0.00	-0.09	1.02	-0.07	FTSE100
IMT.L	FTSE100	0.00	-0.28	0.89	-0.05	FTSE100
IPR.L	FTSE100	0.00	-0.17	1.10	-0.02	FTSE100
ITV.L	FTSE100	0.00	0.36	0.87	0.01	FTSE100
JMAT.L	FTSE100	0.00	0.05	0.97	-0.08	FTSE100
KGF.L	FTSE100	0.00	0.09	1.04	-0.03	FTSE100
LAND.L	FTSE100	0.00	-0.18	1.04	0.01	FTSE100
LGEN.L	FTSE100	0.00	0.05	1.24	0.07	FTSE100
LII.L	FTSE100	0.00	-0.15	0.92	-0.06	FTSE100
LLOY.L	FTSE100	0.00	0.05	1.16	0.05	FTSE100
LMI.L	FTSE100	0.00	-0.03	1.18	-0.06	FTSE100
EMG.L	FTSE100	0.00	-0.06	1.03	-0.02	FTSE100
MKS.L	FTSE100	0.00	-0.13	1.07	0.00	FTSE100
MRW.L	FTSE100	0.00	-0.21	1.07	-0.12	FTSE100
NG.L	FTSE100	0.00	-0.07	0.77	-0.10	FTSE100
NXT.L	FTSE100	0.00	-0.06	0.98	-0.07	FTSE100
OML.L	FTSE100	0.00	0.12	1.21	0.04	FTSE100
PSON.L	FTSE100	0.00	0.29	0.97	-0.08	FTSE100
PSN.L	FTSE100	0.00	-0.14	1.09	-0.04	FTSE100
PRU.L	FTSE100	0.00	0.23	1.27	0.18	FTSE100
RB.L	FTSE100	0.00	-0.26	0.93	0.00	FTSE100

Continue on next page

– continue from the previous page

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
REL.L	FTSE100	0.00	0.24	0.78	-0.01	FTSE100
REX.L	FTSE100	0.00	-0.18	1.06	0.00	FTSE100
RIO.L	FTSE100	0.00	0.13	1.19	0.02	FTSE100
RR.L	FTSE100	0.00	0.21	0.96	0.03	FTSE100
RSA.L	FTSE100	0.00	0.28	1.26	0.06	FTSE100
RBS.L	FTSE100	0.00	0.15	1.11	0.07	FTSE100
RDSB.L	FTSE100	0.00	0.14	0.70	0.08	FTSE100
SAB.L	FTSE100	0.00	-0.02	0.93	0.04	FTSE100
SGE.L	FTSE100	0.00	0.57	0.77	0.09	FTSE100
SBRY.L	FTSE100	0.00	-0.14	1.08	-0.06	FTSE100
SDR.L	FTSE100	0.00	0.08	1.40	0.12	FTSE100
SDRC.L	FTSE100	0.00	0.13	1.51	0.07	FTSE100
SSE.L	FTSE100	0.00	-0.25	0.96	-0.10	FTSE100
SVT.L	FTSE100	0.00	-0.25	0.87	-0.02	FTSE100
SHP.L	FTSE100	0.00	-0.07	1.08	-0.10	FTSE100
SN.L	FTSE100	0.00	-0.07	0.87	-0.07	FTSE100
SMIN.L	FTSE100	0.00	0.00	0.83	0.02	FTSE100
STAN.L	FTSE100	0.00	0.03	1.19	0.02	FTSE100
TATE.L	FTSE100	0.00	-0.20	0.64	0.03	FTSE100
TSCO.L	FTSE100	0.00	-0.27	1.11	-0.02	FTSE100
TRIL.L	FTSE100	0.00	0.35	1.08	0.02	FTSE100
TT.L	FTSE100	0.00	-0.14	1.07	-0.03	FTSE100
TLW.L	FTSE100	0.00	-0.14	0.92	-0.20	FTSE100
ULVR.L	FTSE100	0.00	0.05	0.71	0.00	FTSE100
UU.L	FTSE100	0.00	-0.28	0.98	-0.02	FTSE100
VOD.L	FTSE100	0.00	0.55	0.61	0.04	FTSE100
WTB.L	FTSE100	0.00	-0.14	0.92	-0.02	FTSE100
WOS.L	FTSE100	0.00	-0.05	1.12	-0.03	FTSE100
WPP.L	FTSE100	0.00	0.34	0.90	0.10	FTSE100
III.L	FTSE100	0.00	0.19	1.09	0.09	FTSE100
MMM	DJIA	0.00	-0.11	0.12	0.85	DJIA
T	DJIA	0.00	0.00	-0.01	0.90	DJIA
AA	DJIA	0.00	0.08	0.20	1.16	DJIA
AXP	DJIA	0.00	0.01	-0.01	1.36	DJIA
AIG	DJIA	0.00	0.00	0.04	1.09	DJIA
BAC	DJIA	0.00	-0.06	0.03	1.03	DJIA
BA	DJIA	0.00	-0.04	0.07	0.94	DJIA
CAT	DJIA	0.00	0.00	0.06	1.07	DJIA

Continue on next page

– concluded from previous page

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
CVX	DJIA	0.00	-0.05	0.16	0.58	DJIA
C	DJIA	0.00	0.05	0.02	1.32	DJIA
KO	DJIA	0.00	-0.10	0.03	0.59	DJIA
DD	DJIA	0.00	-0.05	0.09	0.98	DJIA
XOM	DJIA	0.00	-0.06	0.13	0.73	DJIA
GE	DJIA	0.00	0.02	-0.09	1.21	DJIA
GM	DJIA	0.00	0.03	0.07	1.18	DJIA
HPQ	DJIA	0.00	0.19	-0.23	1.26	DJIA
HD	DJIA	0.00	-0.05	-0.04	1.30	DJIA
INTC	DJIA	0.00	0.02	-0.10	1.53	DJIA
IBM	DJIA	0.00	0.15	-0.16	0.96	DJIA
JPM	DJIA	0.00	0.16	-0.15	1.41	DJIA
JNJ	DJIA	0.00	0.01	-0.09	0.55	DJIA
MCD	DJIA	0.00	-0.09	0.14	0.66	DJIA
MRK	DJIA	0.00	0.02	-0.06	0.73	DJIA
MSFT	DJIA	0.00	-0.04	-0.08	1.12	DJIA
PFE	DJIA	0.00	-0.01	0.03	0.76	DJIA
PG	DJIA	0.00	-0.07	-0.04	0.58	DJIA
UTX	DJIA	0.00	0.02	0.00	1.09	DJIA
VZ	DJIA	0.00	0.06	-0.02	0.84	DJIA
WMT	DJIA	0.00	-0.09	-0.03	0.98	DJIA
DIS	DJIA	0.00	0.09	-0.04	1.05	DJIA

Table E.2: Symbol, market the stock belongs to (m_j), parameters α of the least square method for multivariate analysis and market the stock has the highest correlation with (m'_j), for a portfolio of stocks from the CAC40, BEL20 and AEX indices.

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
AC.PA	CAC40	0.00	0.94	-0.08	0.11	CAC40
AF.PA	CAC40	0.00	1.10	0.05	0.04	CAC40
AI.PA	CAC40	0.00	0.83	0.14	-0.08	CAC40
ALU.PA	CAC40	0.00	2.22	-0.72	0.01	CAC40
ALO.PA	CAC40	0.00	2.17	-0.60	-0.43	CAC40
CS.PA	CAC40	0.00	1.06	0.23	0.34	CAC40
BNP.PA	CAC40	0.00	0.98	0.18	0.03	CAC40
EN.PA	CAC40	0.00	1.25	-0.15	-0.08	CAC40
CAP.PA	CAC40	0.00	1.82	-0.63	0.15	CAC40
CA.PA	CAC40	0.00	0.81	0.09	0.03	CAC40
BN.PA	CAC40	0.00	0.52	0.17	-0.10	CAC40
DX.PA	CAC40	0.00	0.39	0.72	0.14	BEL20
EF.PA	CAC40	0.00	0.46	0.08	-0.05	CAC40

Continue on next page

– continue from the previous page

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
FTE.PA	CAC40	0.00	1.70	-0.54	-0.01	CAC40
OR.PA	CAC40	0.00	0.87	0.03	-0.06	CAC40
LG.PA	CAC40	0.00	0.82	0.15	0.01	CAC40
MMB.PA	CAC40	0.00	1.03	-0.32	0.16	CAC40
MC.PA	CAC40	0.00	1.12	0.02	0.01	CAC40
ML.PA	CAC40	0.00	0.71	0.27	0.02	CAC40
RI.PA	CAC40	0.00	0.39	0.16	-0.14	CAC40
UG.PA	CAC40	0.00	0.83	0.07	-0.02	CAC40
PP.PA	CAC40	0.00	1.02	0.02	0.05	CAC40
RNO.PA	CAC40	0.00	1.05	0.10	-0.11	CAC40
SGO.PA	CAC40	0.00	1.14	-0.04	-0.08	CAC40
SAN.PA	CAC40	0.00	0.74	0.16	-0.08	CAC40
SU.PA	CAC40	0.00	0.97	0.04	-0.16	CAC40
GLE.PA	CAC40	0.00	1.02	0.19	0.10	CAC40
STM.PA	CAC40	0.00	1.56	-0.77	0.40	CAC40
SZE.PA	CAC40	0.00	0.90	0.53	-0.08	CAC40
FP.PA	CAC40	0.00	0.58	0.17	0.08	CAC40
UL.PA	CAC40	0.00	0.40	0.23	-0.14	CAC40
VK.PA	CAC40	0.00	0.66	0.22	-0.07	CAC40
DG.PA	CAC40	0.00	0.51	0.14	-0.09	CAC40
VIV.PA	CAC40	0.00	1.44	-0.28	0.10	CAC40
ACKB.BR	BEL20	0.00	-0.12	0.90	0.03	BEL20
AGFB.BR	BEL20	0.00	-0.19	1.24	-0.03	BEL20
BEKB.BR	BEL20	0.00	-0.16	1.30	-0.01	BEL20
COFB.BR	BEL20	0.00	-0.04	0.36	-0.07	BEL20
COL.BR	BEL20	0.00	-0.24	1.17	-0.14	BEL20
DELB.BR	BEL20	0.00	-0.20	1.30	0.07	BEL20
DEXB.BR	BEL20	0.00	0.30	0.81	0.14	BEL20
FORB.BR	BEL20	0.00	0.27	0.91	0.38	BEL20
GBL.BR	BEL20	0.00	0.17	0.83	0.02	BEL20
KBC.BR	BEL20	0.00	0.06	1.15	0.04	BEL20
MOBB.BR	BEL20	0.00	0.02	0.99	-0.11	BEL20
NAT.BR	BEL20	0.00	-0.06	0.65	0.07	BEL20
OME.BR	BEL20	0.00	-0.35	1.69	-0.21	BEL20
SOLB.BR	BEL20	0.00	0.02	0.92	-0.07	BEL20
SZEB.BR	BEL20	0.00	0.68	0.71	-0.09	BEL20
UCB.BR	BEL20	0.00	-0.11	1.07	-0.07	BEL20
UMI.BR	BEL20	0.00	-0.06	1.00	0.03	BEL20

Continue on next page

– concluded from previous page

Symbol	m_j	α_0	α_1	α_2	α_3	m'_j
AGN.AS	AEX	0.00	0.38	0.34	0.94	AEX
AH.AS	AEX	0.00	-0.18	0.26	1.26	AEX
AKZA.AS	AEX	0.00	0.03	0.15	0.70	AEX
MT.AS	AEX	0.00	-0.68	0.03	1.45	AEX
ASML.AS	AEX	0.00	1.24	-1.10	1.28	AEX
CORA.AS	AEX	0.00	-0.09	0.30	0.24	BEL20
CXP.AS	AEX	0.00	-0.68	0.05	1.99	AEX
DSM.AS	AEX	0.00	0.06	0.17	0.50	AEX
FORA.AS	AEX	0.00	0.29	0.80	0.46	BEL20
HEIA.AS	AEX	0.00	0.01	0.17	0.30	AEX
INGA.AS	AEX	0.00	0.47	0.36	0.76	AEX
KPN.AS	AEX	0.00	0.53	-0.55	0.91	AEX
PHIA.AS	AEX	0.00	1.08	-0.55	0.90	CAC40
RAND.AS	AEX	0.00	-0.20	-0.06	1.25	AEX
REN.AS	AEX	0.00	0.17	-0.05	0.72	AEX
RDSA.AS	AEX	0.00	0.34	0.10	0.33	CAC40
SBMO.AS	AEX	0.00	-0.27	0.30	0.61	AEX
TNT.AS	AEX	0.00	-0.09	0.23	0.59	AEX
UNA.AS	AEX	0.00	0.16	0.20	0.26	AEX
VDOR.AS	AEX	0.00	-0.25	-0.24	1.60	AEX
WKL.AS	AEX	0.00	-0.11	-0.03	0.83	AEX

Bibliography

[1] P. Bak, *How nature works: The science of self-organized criticality.* Oxford University Press, Oxford (1997)

[2] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press, Cambridge (2001)

[3] J. Taylor, *Copernicus on the Evils of Inflation and the Establishment of A Sound Currency,* Journal of the History of Ideas **16**, 540 (1955)

[4] G. F. Shirras and J. H. Craig, *Sir Isaac Newton and the Currency,* The Economic Journal **55**, 217 (1945)

[5] L. Bachelier, *Théorie de la spéculation*, Annales scientifiques de l'É.N.S. 3e série, **17**, 21-86 (1900)

[6] B. Mandelbrot, *The Pareto-Lévy Law and the Distribution of Income*, International Economic Review **1**, 79 (1960)

[7] B. Mandelbrot, *New Methods in Statistical Economics*, The Journal of Political Economy **71**, 421 (1963)

[8] http://xxx.lanl.gov/

[9] http://prl.aps.org/

[10] http://pre.aps.org/

[11] http://epjb.edpsciences.org/

[12] http://www.elsevier.com/locate/physa

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

[14] J. Voit, *The Statistical Mechanics of Financial Markets.* Springer-Verlag, Berlin (2001)

[15] D. Sornette, *Why Stock Markets Crash: Critical Events in Complex Financial Systems.* Princeton University Press, Princeton (2003)

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

[17] A. Chatterjee, S. Yarlagadda and B. K. Chakrabarti, *Econophysics of Wealth Distributions.* Springer-Verlag, Milan (2005)

[18] B. K. Chakrabarti, A. Chakraborti and A. Chatterjee, *Econophysics and Sociophysics of Wealth Distributions: Trends and Perspectives.* Wiley-VCH, Berlin (2006)

[19] http://www.unifr.ch/econophysics/

[20] R.L. Axtell, *Zipf Distribution of U.S. Firm Sizes*, Science **293**, 1818 (2001)

[21] C. Di Guilmi, E. Gaffeo and M. Gallegati, *Power Law Scaling in the World Income Distribution*, Economics Bulletin **15**, 1 (2003)

[22] V. Pareto, *Cours d'Économie Politique*, Libraire Droz (Genève), (1964), new edition of the original from 1897

[23] W. Souma, *Universal structure of the personal income distribution*, Fractals **9**, 463 (2001)

[24] P. Richmond, S. Hutzler, R. Coelho and P. Repetowicz, *A review of empirical studies and models of income distributions in society*, in *Econophysics and Sociophysics of Wealth Distributions: Trends and Perspectives.* (eds. B. K. Chakrabarti, A. Chakraborti and A. Chatterjee), Wiley-VCH, Berlin (2006)

[25] S. Solomon and P Richmond, *Stable power laws in variable economies: Lotka-Volterra implies Pareto-Zipf*, Eur. Phys. J. B **27**, 257 (2002)

[26] O. Malcai, O. Biham, P. Richmond, and S. Solomon, *Theoretical analysis and simulations of the generalized Lotka-Volterra model*, Phys. Rev. E **66**, 031102 (2002)

[27] J.-P. Bouchaud and M. Mézard, *Wealth condensation in a simple model of economy*, Physica A **282**, 536 (2000)

[28] A. Chakraborti and B. K. Chakrabarti, *Statistical mechanics of money: how saving propensity affects its distribution*, Eur. Phys. J. B **17**, 167 (2000)

[29] A. Chakraborti, *Distributions of money in model markets of Economy*, Int. J. Mod. Phys. C **13**, 1315 (2002)

[30] M. Patriarca, A. Chakraborti K. and Kaski, *Statistical model with a standard* Γ *distribution*, Phys. Rev. E **70**, 016104 (2004)

[31] M. Patriarca, A. Chakraborti and K. Kaski, *Gibbs versus non-Gibbs distributions in money dynamics*, Physica A **340**, 334 (2004)

[32] A. Chatterjee, B. K. Chakrabarti and S. S. Manna, *Money in gas-like markets: Gibbs and Pareto laws*, Physica Scripta T **T106**, 36 (2003)

[33] A. Chatterjee, B. K. Chakrabarti and S. S. Manna, *Pareto law in a kinetic model of market with random saving propensity*, Physica A **335**, 155 (2004)

[34] P. Repetowicz, S. Hutzler and P. Richmond, *Dynamics of money and income distributions*, Physica A **356**, 641 (2005)

[35] R. Coelho, Z. Néda, J. J. Ramasco and M. A. Santos, *A family-network model for wealth distribution in societies*, Physica A **353**, 515 (2005)

[36] http://www.forbes.com

[37] M. L. Mehta, *Random Matrices*, Elsevier-Academic Press, Netherlands (2004)

[38] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, *Noise Dressing of Financial Correlation Matrices*, Phys. Rev. Lett. **83**, 1467 (1999)

[39] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, *Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series*, Phys. Rev. Lett. **83**, 1471 (1999)

[40] Z. Burda, J. Jurkiewicz and M. A. Nowak, *Applications of Random Matrices to Economy and other Complex Systems*, Acta Physica Polonica B **36**, 2603 (2005)

[41] P. Gopikrishnan, B. Rosenow, V. Plerou and H. E. Stanley, *Quantifying and interpreting collective behavior in financial markets*, Phys. Rev. E **64**, 035106 (2001)

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

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

[44] G. Bonanno, F. Lillo and R. N. Mantegna, *High frequency cross-correlation in a set of stocks*, Quantitative Finance **1**, 96 (2001)

[45] G. Bonanno, G. Caldarelli, F. Lillo, S. Miccichè, N. Vandewalle and R. N. Mantegna, *Network of equities in financial markets*, Eur. Phys. J. B **38**, 363 (2004)

[46] N. Vandewalle, F. Brisbois and X. Tordois, *Non-random topology of stock markets*, Quantitative Finance **1**, 372 (2001)

[47] G. Bonanno, G. Caldarelli, F. Lillo and R. N. Mantegna, *Topology of correlation-based minimal spanning trees in real and model markets*, Phys. Rev. E **68**, 046130 (2003)

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

[49] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész and A. Kanto, *Asset trees and asset graphs in financial markets*, Physica Scripta T**106**, 48 (2003)

[50] M. McDonald, O. Suleman, S. Williams, S. Howison and N. F. Johnson, *Detecting a currency's dominance or dependence using foreign exchange network trees*, Phys. Rev. E **72**, 046106 (2005)

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

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

[53] J.-P. Onnela, A. Chakraborti, K. Kaski and J. Kertész, *Dynamic asset trees and portfolio analysis*, Eur. Phys. J. B **30**, 285 (2002)

[54] J.-P. Onnela, A. Chakraborti, K. Kaski and J. Kertész, *Dynamic asset trees and Black Monday*, Physica A **324**, 247 (2003)

[55] H. Situngkir and Y. Surya, *On Stock Market Dynamics through Ultrametricity of Minimum Spanning Tree*, Technical Report WPH2005 Bandung Fe Institute, Dept. Computational Sociology, Bandung Fe Institute (2005)

[56] H. Situngkir and Y. Surya, *Tree of Several Asian Currencies*, BFI Working Paper No WPI 2005 (2005)

[57] V. Pareto, *Manual of Political Economy*, Macmillan (London), (1971), translated from the original edition of 1906

[58] R. Gibrat, *Les Inégalites Économiques*, Libraire du Recueil Sirey (Paris), (1931)

[59] Y. Fujiwara, *Pareto-Zipf, Gibrat's Laws, Detailed-Balance and their Breakdown* in *Econophysics of Wealth Distributions*, A. Chatterjee, S. Yarlagadda and B.K. Chakrabarti Editors, Springer (Italy), 24 (2005)

[60] D. G. Champernowne, *A Model of Income Distribution*, The Economic Journal **63** 318 (1953)

[61] D. G. Champernowne, *The Graduation of Income Distributions*, Econometrica **20**, 591 (1952)

[62] H. F. Lydall, *The Distribution of Employment Incomes*, Econometrica **27**, 110 (1959)

[63] W. J. Reed, *The Pareto law of incomes - an explanation and an extension*, Physica A **319**, 469 (2003)

[64] H. Aoyama, W. Souma, Y. Nagahara, M. P. Okazaki, H. Takayasu and M. Takayasu, *Pareto's law for income of individuals and debt of bankrupt companies*, Fractals **8**, 293 (2000)

[65] Y. Fujiwara, W. Souma, H. Aoyama, T. Kaizoji and M. Aoki, *Growth and fluctuations of personal income*, Physica A **321**, 598 (2003)

[66] A. Dragulescu and V. M. Yakovenko, *Evidence for the exponential distribution of income in the USA*, Eur. Phys. J. B **20**, 585 (2001)

[67] A. Dragulescu and V. M. Yakovenko, *Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States*, Physica A **299**, 213 (2001)

[68] A. Dragulescu and V. M. Yakovenko, *Statistical Mechanics of Money, Income, and Wealth: A Short Survey*, AIP Conference Proceedings **661**, 180 (2003)

[69] A. C. Silva and V. M. Yakovenko, *Temporal evolution of the "thermal" and "superthermal" income classes in the USA during 1983-2001*, Europhys. Lett. **69**, 304 (2005)

[70] A. Banerjee, V. M. Yakovenko and T. Di Matteo, *A study of the personal income distribution in Australia*, Physica A **370**, 54 (2006)

[71] T. Di Matteo, T. Aste and S. T. Hyde, *Exchanges in complex networks: income and wealth distributions*, in *The Physics of Complex Systems (New Advances and Perspectives)*, (eds. F. Mallamace and H. E. Stanley), IOS Press, Amsterdam (2004)

[72] N. Scafetta, S. Picozzi and B. J. West, *An out-of-equilibrium model of the distributions of wealth*, Quantitative Finance **4**, 353 (2004)

[73] R. Coelho, *Modelos de Distribuição de Riqueza*, MSc Thesis - Universidade do Porto, Porto (2004)

[74] F. Clementi and M. Gallegati, *Power law tails in the Italian personal income distribution*, Physica A **350**, 427 (2005)

[75] F. Clementi and M. Gallegati, *Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom and the United States*, in *Econophysics of Wealth Distributions*, A. Chatterjee, S. Yarlagadda and B. K. Chakrabarti Editors, Springer-Verlag, Milan (2005)

[76] S. Sinha, *Evidence for power-law tail of the wealth distribution in India*, Physica A **359**, 555 (2006)

[77] M. Levy and S. Solomon, *New evidence for the power-law distribution of wealth*, Physica A **242**, 90 (1997)

[78] S. Levy, *Wealthy People and Fat Tails: An Explanation for the Lévy Distribution of Stock Returns*, Finance, 30 (1998)

[79] M. Levy, *Are rich people smarter?*, J. Econ. Theory **110**, 42 (2003)

[80] A. Y. Abul-Magd, *Wealth distribution in an ancient Egyptian society*, Phys. Rev. E **66**, 057104 (2002)

[81] M. A. Santos, R. Coelho, G. Hegyi, Z. Néda and J. Ramasco, *Wealth distribution in modern and medieval societies*, Eur. Phys. J. Special Topics **143**, 81 (2007)

[82] G. Hegyi, Z. Néda and M. A. Santos, *Wealth distribution and Pareto's law in the Hungarian medieval society*, Physica A **380**, 271 (2007)

[83] A. Jayadev, *A power law tail in India's wealth distribution: Evidence from survey data*, Physica A **387**, 270 (2008)

[84] O. Biham, O. Malcai, M. Levy and S. Solomon, *Generic emergence of power law distributions and Lévy-Stable intermittent fluctuations in discrete logistic systems*, Phys. Rev. E **58** 1352 (1998)

[85] P. Richmond and S. Solomon, *Power laws are disguised Boltzmann laws*, Int. J. Mod. Phys. C **12**, 333 (2001)

[86] S. Solomon and P. Richmond, *Power laws of wealth, market order volumes and market returns*, Physica A **299**, 188 (2001)

[87] J. Angle, *The Surplus Theory of Social Stratification and the Size Distribution of Personal Wealth*, Social Forces **65**, 293 (1986)

[88] S. Ispolatov, P. L. Krapivsky and S. Redner, *Wealth distributions in asset exchange models*, Eur. Phys. J. B **2**, 267 (1998)

[89] A. Dragulescu and V. M. Yakovenko, *Statistical mechanics of money*, Eur. Phys. J. B **17**, 723 (2000)

[90] A. Chatterjee, B. Chakrabarti and R. B. Stinchcombe, *Master equation for a kinetic model of a trading market and its analytic solution*, Phys. Rev. E **72**, 026126 (2005)

[91] F. Slanina, *Inelastically scattering particles and wealth distribution in an open economy*, Phys. Rev. E **69**, 046102 (2004)

[92] R. Coelho, P. Richmond, J. Barry and S. Hutzler, *Double power laws in income and wealth distributions*, Physica A **387**, 3847 (2008)

[93] P. Richmond, P. Repetowicz, S. Hutzler and R. Coelho *Comments on recent studies of the dynamics and distribution of money*, Physica A **370**, 43 (2006)

[94] A. Chatterjee and B. K. Chakrabarti, *Kinetic exchange models for income and wealth distributions*, Eur. Phys. J. B **60**, 135 (2007)

[95] A. M. Scarfone, *A mechanism to derive multi-power law functions: An application in the econophysics framework*, Physica A **382**, 271 (2007)

[96] K. Weron and A. Jurlewicz, *Two forms of self-similarity as a fundamental feature of the power-law dielectric response*, J. Phys. A **26**, 395 (1993)

[97] C. Tsallis, G. Bemski and R. S. Mendes, *Is re-association in folded proteins a case of nonextensivity*, Phys. Lett. A **257**, 93 (1999)

[98] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer and H. E. Stanley, *Scaling of the distribution of flutuations of financial market indices*, Phys. Rev. E **60**, 5305 (1999)

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

[100] M. A. Montemurro, *Beyond the Zipf-Mandelbrot law in quantitative linguistics*, Physica A **300**, 567 (2001)

[101] C. Tsallis, J. C. Anjos and E. P. Borges, *Fluxes of cosmic rays: a delicately balanced stationary state*, Phys. Lett. A **310**, 372 (2003)

[102] A. Ishikawa, *Annual change of Pareto index dynamically deduced from the law of detailed quasi-balance*, Physica A **371**, 525 (2006)

[103] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys **52**, 479 (1988)

[104] E. P. Borges, *Empirical nonextensive laws for the county distribution of total personal income and gross domestic product*. Physica A **334**, 255 (2004)

[105] G. Kaniadakis, M. Lissia and A. M. Scarfone, *Deformed logarithms and entropies*, Physica A **340**, 41 (2004)

[106] http://www.gnuplot.info/

[107] http://www.datastream.com/

[108] http://www.mysql.com/

[109] P. Gopikrishnan, M. Meyer, L. A. N. Amaral and H. E. Stanley, *Inverse cubic law for the distribution of stock price variations*, Eur. Phys. J. B **3**, 139 (1998)

[110] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer and H. E. Stanley, *Scaling of the distribution of price flutuations of individual companies*, Phys. Rev. E **60**, 6519 (1999)

[111] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics and Financial Markets*. World Scientific Publishing, Singapore (2004)

[112] C. Tsallis, C. Anteneodo, L. Borland and R. Osorio, *Nonextensive statistical mechanics and economics*, Physica A **324**, 89 (2003)

[113] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley-Interscience, New York (1995)

[114] P. Repetowicz and P. Richmond, *Option Pricing with Log-stable Lévy Processes*, in *Econophysics of Markets and Business Networks*. (eds. A. Chatterjee and B. K. Chakrabarti), Springer, Milan (2007)

[115] V. Tola, F. Lillo, M. Gallegati and R. N. Mantegna, *Cluster analysis for portfolio optimization*, Journal of Economic Dynamics and Control **32**, 235 (2008)

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

[117] R. C. Prim, *Shortest connection networks and some generalisations*, Bell System Tech. J. **36**, 1389 (1957)

[118] V. Batagelj and A. Mrvar, *Pajek - Program for Large Network Analysis*, http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[119] T. Kamada and S. Kawai, *An algorithm for drawing general undirected graphs*, Information Processing Letters **31**, 7 (1989)

[120] http://www.ftse.com

[121] http://www.icbenchmark.com/

[122] C. Coronnello, M. Tumminello, F. Lillo, S. Miccichè and R. N. Mantegna, *Sector Identification in a Set of Stock Return Time Series Traded at the London Stock Exchange*, Acta Physica Polonica B **36**, 2653 (2005)

[123] GNU Scientific Library (GSL), http://www.gnu.org/software/gsl

[124] G. Bonanno, N. Vandewalle and R. N. Mantegna, *Taxonomy of stock market indices*, Phys. Rev. E **62**, R7615 (2000)

[125] D. B. Pantom, V. P. Lessig and O. M. Joy, *Comovement of International Equity Markets: A Taxonomic Approach*, The Journal of Financial and Quantitative Analysis **11**, 415 (1976)

[126] P. Schotman and A. Zalewska, *Non-synchronous trading and testing for market integration in Central European emerging markets*, Limburg Institute of Financial Economics, LIFE Working Paper **04-017** (2004)

[127] T. Baig and I. Goldfajn, *Financial Market Contagion in the Asian Crisis*, International Monetary Fund, IMF Working Paper **98-155** (1998)

[128] K. J. Forbes and R. Rigobon, *No Contagion, Only Interdependence: Measuring Stock Market Comovements*, The Journal of Finance **57**, 2223 (2002)

[129] http://www.mscibarra.com/

[130] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New York (1959)

[131] http://www.djindexes.com

[132] http://online.wsj.com

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

[134] T. J. Flavin, *The effect of the Euro on country versus industry portfolio diversification*, Journal of International Money and Finance **23**, 1137 (2004)

[135] K. Phylaktis and L. Xia, *The Changing Roles of Industry and Country Effects in the Global Equity Markets*, The European Journal of Finance **12**, 627 (2006)

[136] F. Lillo and R. N. Mantegna, *Spectral density of the correlation matrix of factor models: A random matrix theory approach*, Phys. Rev. E **72**, 016219 (2005)

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

[138] K. V. Mardia, J. T. Kent and J. M. Bibby, *Multivariate Analysis*. Academic Press, London (1979)