DISSECTING THE SOUTH AFRICAN EQUITY MARKETS INTO SECTORS AND STATES

Prepared by:
Bongani Mbambiso, B.Sc. in Electrical Engineering (UCT)

Supervised by:
Dr. Diane Wilcox

and:
Dr. Tim Gebbie

Dissertation
presented to the Faculty of Science of the University of Cape Town in partial fulfilment of the requirements for the
M.Sc. Degree in Mathematics of Finance

February 5, 2009

Abstract

It is common to consider financial markets as made up of stocks which can be grouped or clustered according to similar characteristics. The understanding of grouping characteristics is important for understanding the overall dynamics of the market in which such clustering takes place. The appearance of grouping or clustering behavior may be due mainly to random effects (noise) since there are bound to be overlapping properties when the number of stocks is high. Hence it is necessary to investigate methods for measuring possible noise contributions to clustering and also methods of removing or reducing the noise contributions. Grouping may change over time, therefore, it is necessary to identify the time horizons for which clusters are stable.

We identify sectors, groups of stocks which display similar behavior, and states, time periods for which the market or the market sectors behave similarly. We use maximum likelihood methods developed by Marsili (2002) [8] based on the ansatz of Noh (2000) [9]. Following Marsili's model we implement both a deterministic recursive merging algorithm as well as a simulated annealing search algorithm. These methods are use a novel non-parametric formulation of a pricing model that can be interpreted in terms of the q-state Potts model. This not only provides a tool through which to understand the structure of South African markets, but provide an alternative, non-equilibrium, framework within which to understand market structure both temporally (states) and cross-sectionally (sectors) directly in terms of market structure. A major component of this work was to write software based on the Marsili et al approach. This was written and tested as a matlab toolbox.

Keywords: Log-Likelihood, Deterministic Recursive Merging, Simulated Annealing, Hamiltonian, Potts Model, Ising Model.

Contents

List of Tables

List of Figures

Declaration

I declare that: ”Dissecting the South African Financial Markets into Sectors and States” is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

Bongani Mbambiso

....

Acknowledgement

I am most grateful to my hardworking supervisors, Dr. Diane Wilcox, and Dr. Tim Gebbie, for their guidance and helpful suggestions through the course of this research. I acknowledge Futuregrowth Asset Management, the NRF-Thuthuka program and Dr. Diane Wilcox and Dr. Tim Gebbie for financial support. I also thank my family for their support and patience throughout my period of study.

Chapter 1

Introduction

In recent years, the availability of large amounts of financial data has resulted in a rapid growth in statistical finance research. Statistical finance can be classified into three main areas [1]: (*i*) the study of universal statistical features of financial time series, e.g random matrix theory [2]; (*ii*) the use of these statistical features to design better risk models and derivatives pricing [3]; (*iii*) the study of agent-based models in order to unveil the basic mechanisms that can be used to explain statistically, the behavior of financial time series. The work described on this paper falls under (*iii*).

Financial time series are rich in information about the performance of the market and it's constituents. Studying stock price changes helps to reveal a number of statistical features about financial time series such as the cross-correlations between stocks in the market [3], the volatility, and the distribution of the price changes. It is often is assumed that stock prices behave like geometric Brownian motion. Several studies have shown that prices behave differently from GBM, and prices are not Gaussian distributed.

In the early sixties, theories that rejected the hypothesis that the probability distribution of logarithmic stock returns is Gaussian, were developed [4]. This did not prevent researchers from conducting investigations based on this hypothesis. Research work on financial markets used to be done on single assets or indices. Recent work has shown that financial markets behave as complex self-organizing critical systems. The market is made up of different agents playing different roles: producers, speculators, and noise traders [5].

In this dissertation financial markets are seen as complex multi-parts systems that exhibit similar characteristics found in model systems studied in statistical physics, the Potts Model. This was first published in Giada's PhD thesis [6], and Giada and Marsili [7]. Essentially this dissertation is an explication and implementation of a novel method of cluster analysis published by Marsili [8], and based on the statistical physics approach of Noh [9].

Financial markets are made up of defined standard economic sectors or indices listed on stock exchanges. Indices are usually groups of traded shares of companies that are in same economic business, for example, mining, banking, and retail sectors. Nowadays, the available massive flows of financial data show us that these classifications do not necessarily hold when we look at assets with similar market behaviors in terms of price returns.

Economic activities have dynamic performance patterns and differ in every country. Statistical finance has in fact shown that there are many universal features which are common across several developed markets [2]. Developing markets also have common features, but may have their own unique properties, depending on the scale, strengths and independence of economic activities in country. There are seasonal fluctuations in economic activities, and there are long-term trends in the markets sectors. In different time periods different market sectors perform differently.

In light of the information mentioned above, one may ask the following questions:

- What is the appropriate way of defining co-movement of objects in a market?
- Can markets be broken up into sectors of correlated assets?
- Can markets be broken up into sectors of correlated states?
- Which time horizons were the sectors stable?

When it comes to investing, the above questions are very important. Anyone purchasing traded shares is exposed to risk of unforseen changes or abnormal movement of stock prices. Portfolio managers try to reduce the risk by spreading their investment over a wide range of different stocks. This is done to maximize the risk-return profile of the portfolio of stocks over a specified time period [10, 11]. This forces portfolio managers to understand the behavior of different shares in relation to one another. A manager administering a portfolio of stocks with too many stocks with prices that move together face a higher risk. This is another reason why economic sector analysis is very important [12].

Since it is assumed that no trade occurs in isolation, it is necessary to estimate the interactions or the correlations of the traded shares. This can be done by studying the cross-correlations or covariances between prices of shares traded in the market. The main aim of this work is to uncover the internal structure of groups of correlated stocks (sectors) and correlated days (states) in the South African financial markets. This investigation is only concerned with co-movements of objects in the market and seeks to solve a clustering problem.

To solve a clustering problem one needs an appropriate way of defining co-movement or the measure of similarity of objects in a market. This investigation uses the analogies between two systems, namely, a physical system made up of interacting-agents and a financial market made up of correlated stocks. An unsupervised parameter free approach which was developed by Marsili and Giada based on maximum likelihood principle is introduced later in the dissertaion.

To complete the solution of the clustering problem, using Marsili and Giada's model, a deterministic recursive merging algorithm as well as a simulated annealing search algorithm are implemented to investigate South African sectors and states.

Chapter 2

Cluster Analysis

Cluster analysis is a generic name for a variety of mathematical methods that can be used to find out which objects in a set are similar [13]. Instead of using physical objects, mathematical methods for cluster analysis sort objects described as data. Objects with similar descriptions are mathematically gathered into the same cluster.

For a variety of research goals, researchers need to find out which objects in a set are similar and dissimilar. One other reason cluster analysis is so useful is that researchers in all fields need to make and revise classifications continually.

2.1 Defining Data Clustering

Clustering of data are methods by which large data sets are grouped into clusters of smaller sets of similar data. A standard definition of clustering is as follows. Partition N given objects into k groups or clusters, so that two objects in the same cluster are, in some sense, more similar than two that belong to different clusters or different groups [14]. The i = 1, 2, . . .,N data objects, $\overrightarrow{x}_i$ can be represented by D-measurable features. This means that they can be specified either in terms of their coordinates in D-dimensional space, where $x_i(d)$, d = 1, 2, 3, . . .,D, are components of the vector $\overrightarrow{x}_i$. Alternatively, the components may be specified by means of an N x N "distance matrix" whose elements d_{ij} measure the similarity of objects i and j.

Clustering is an ill-posed problem since it lacks a unique solution. The problem of organizing and analyzing large amounts of data and the concept of "data mining" has led to an introduction of clustering algorithms. Clustering algorithms attempt to find natural groups of objects based on some similarity measure. The output from a clustering algorithm is a statistical description of the cluster configuration in which each cluster contains a certain number of objects.

There are two major categories of clustering: *Partitioning* (*k-clustering*), in which every object is assigned to exactly one group, and *hierarchical clustering*, in which each group of size greater than one is composed of smaller groups.

2.2 *k*-Clustering

Suppose $X = \{x_i\}_{i=1}^N$ is a set of all objects to be clustered. A partition S of k non-empty clusters is a collection of subsets of X, given by

$$S = \{s_1, s_2, ..., s_k\}, \tag{2.1}$$

such that

$$X = s_1 \cup s_2 \cup ..... \cup s_k, \tag{2.2}$$

and

$$s_i \cap s_j = \phi, ...\forall i, j = 1, 2, ..., k, (i \neq j). \tag{2.3}$$

In partitioning clustering methods, a certain structure is assumed and a desired k number of clusters is assumed at the start. Points are then allocated among clusters so that a particular *clustering criterion* is optimized. A possible criterion is the minimization of the variability within clusters, as measured by the sum of the variances of each parameter that characterizes a point. Optimization algorithms usually assume that the elements of S are drawn from a d-dimensional metric space, usually $\mathbb{R}^d$, and define a cost function $c : \{X : X \subseteq S\} \to \mathbb{R}^+$ which associates a cost with each cluster. The goal of the algorithm is then to minimize $\sum_{i=1}^k c(s_i)$, the sum of the costs of clusters. In these methods, the clusters are mutually exclusive, and they are based in the following procedures:

- *Initiating clusters* - Choosing an initial cluster configuration depends on the algorithm being used and also on the user of the algorithm. The initial clusters could be chosen randomly, sequentially or using an algorithm.
- *Allocating entities to k clusters* - This process depends on the definition of *a measure of similarity* and the *cost function*, which is the clustering criterion, whose minimum (or maximum) corresponds to optimal clustering configuration.

Sometimes the cost function is arbitrary, but in some cases it depends on the nature of the problem. Predefining the k number of clusters is also arbitrary. Different resolutions on the data can cause different choices of k's, and different k's produce different cluster configurations. The most well-known optimization criteria which are used in k-clustering are discussed below.

2.2.1 Sum-of-Squares Criterion

Let x_f^j be the f-th element in the j-th cluster s_j, n_{s_j} be the number of elements in s_j, and $d(x_f^j, x_t^j)$ be the distance between x_f^j and x_t^j. The sum-of-squares criterion is then defined with the following cost function:

$$c(s_j) = \sum_{f=1}^{n_{s_j}} \sum_{t=1}^{n_{s_j}} (d(x_f^j, x_t^j))^2. \tag{2.4}$$

Gonzalez [15], proposed another optimization criterion where, instead of minimizing $\sum_{i=1}^k c(s_i)$, Gonzalez minimizes the

$$\max_{1 \leq j \leq k} c(s_j). \tag{2.5}$$

$c(s_j)$ is given below as:

$$c(s_j) = \max_{x_f^j, x_t^j \in s_j} d(x_f^j, x_t^j). \tag{2.6}$$

This method favors spherical clusters and it does not deal adequately with noise. This means that this approach is not stable, very sensitive to outliers, and it does not often work well when the clusters are of different size, shape and density [17].

2.2.2 Density-Based Criterion

The *density-based* k-clustering approach addresses the issues of cluster shape and noise. This approach uses a local cluster criterion in which clusters can be considered as densely populated areas in the space containing S. Suppose space G is a set of subspaces S, T and U. Furthermore, subspaces such as S are partitioned in to k areas or clusters. These areas can have arbitrary shapes and can be easily separated from one another. Clusters are defined as regions in the data space where objects are dense, and remain separated from one another by low-density regions.

Density-based clustering has an advantage over simple k-clustering in discovering clusters of arbitrary shapes and sizes. However, density-based clustering works well only on simple data set where cluster densities are similar [16]. In this approach the user supplies the parameter k and the *threshold density* for a region to be termed *dense*. The estimation of these parameters becomes more difficult high dimension data set [17].

2.3 Hierarchical Methods

Hierarchial clustering methods include those techniques where the input data sets are not partitioned into the desired number of classes or clusters in a single step. Instead, a series of successive fusions of data are performed until the final number of clusters is obtained. A set of N given objects produces a sequence of partitions (configurations of clusters) of sizes $K_1, K_2, ..., K_N$, such that

$$1 = K_0 < K_1 < K_2 < ... < K_{n-1} < K_n \leq N. \tag{2.7}$$

Each K_i represents the number of clusters of objects in the $i-th$ cluster configuration.

Hierarchical clustering methods can be classified into two general forms. *Divisive* methods recursively partition a given data set S until singleton sets are achieved; the *agglomerative* methods begin with singletons sets and merge them until S is achieved.

Hierarchical algorithms produce a tree $T(S)$ in which nodes represent subsets of S. Figure 2.1 illustrates the two hierarchical clustering method mechanisms: divisive moving top-down on the left and agglomerative moving bottom-up on the right. With divisive procedures, clusters at level $j+1$ ($j = 0, ..., n-1$) are produced by splitting up one of the clusters of level j. With agglomerative procedures, starting with $K_n = N$, clusters at level $j-1$ ($j = n, n-1, ..., 1$) are produced by merging clusters at level j. In both directions, criteria are applied. The splitting and the merging strategies are usually designed optimally at each stage. A cost function or a similarity measure $c(s_i, s_j)$ is used to find the pair of sets $\{s_i, s_j\}$, which results into s_i and s_j being replaced by $s_i \cup s_j$.

The sequence structure of clusters and their hierarchical organization can be represented by a convenient and compact tree called a *dendrogram*. The most well-known cost functions used in hierarchical clustering algorithms are discussed in the following sections.

$$
\begin{array}{c|c|c}
\downarrow & S=\{x_1,x_2,...,x_N\} & \\
 & s_1,s_2,...,s_{K_1} & \\
\text{divisive} & s_1,s_2,...,s_{K_2} & \text{agglomerative} \\
 & . & \\
 & . & \\
 & s_1,s_2,...,s_{K_n} & \uparrow
\end{array}
$$

Figure 2.1: Hierarchical Clustering Methods (*divisive* and *agglomerative*)

2.3.1 Single-Link Clustering

In single-link clustering, the similarity of two clusters is the similarity of their most similar members. This method is very sensitive to outliers, it pays attention to points that fit very well to global structure [20]. The single-link merge criterion is local and the cost function is given by

$$c(s_i,s_j)=\min_{x_f^i\in S_i,x_k^j\in S_j} d(x_f^i,x_k^j), \tag{2.8}$$

where x_f^i is the f-th element in s_i, x_k^i is the k-th element in s_i , and $d(x_f^i,x_k^j)$ is the distance between x_f^j and x_k^i.

2.3.2 Complete-Link Clustering

In complete-link clustering, the similarity of two clusters is the similarity of their most dissimilar members. This method pays too much attention to outliers, it concentrates on points that do not fit very well into the global structure [21]. The cost function is given as

$$c(s_i,s_j)=\max_{x_f^i\in S_i,x_k^j\in S_j} d(x_f^i,x_k^j). \tag{2.9}$$

2.3.3 Average-Link Clustering

This method computes the average similarity of all pairs of elements, including elements of the same clusters [22]. Self-similarities (similarity of one element with itself) are not included in the average. This method deals with outliers better than single-link and complete-link clustering methods. The cost function is given by

$$c(s_i,s_j)=\frac{1}{|s_i||s_j|}\sum_{x_f^i\in S_i}\sum_{x_k^j\in S_j} d(x_f^i,x_k^j). \tag{2.10}$$

2.4 Other Clustering Methods

Data clustering methods can alternatively be classified as *parametric* and *nonparametric* clustering techniques. In parametric techniques some knowledge of the cluster structure is assumed beforehand. For example each cluster can be represented by a center and a spread around it. This assumption is incorporated in a global criterion. The goal is to assign the data points to clusters so that the criterion is optimized. In nonparametric techniques, no further information is needed apart from that contained in the data. This is of practical importance especially when dealing with huge amounts of data sets.

Many other modern clustering methods have been proposed to deal with the nonparametric approach to clustering. These techniques include: *Expectation Minimization* clustering (EM) [23], where the density of points is modeled as a mixture of Gaussians whose centering and scale parameters can be computed by using maximum likelihood; *Super-Paramagnetic Clustering* (SPC) proposed by Blatt *et al.* [24], a method which is based on a mapping to an interacting particle system whose magnetic properties describe the cluster structure of data; *Singular Value Decomposition Technique* (SVD) [26], identifying principal components with cluster structures; *Maximum Likelihood clustering* (ML) [27], described and used in the model building process of this dissertation; the *Self-Organising Maps* (SOM) approach proposed by Kohenen *et al.* [25].

2.5 Choosing a Clustering Algorithm

Choosing a clustering algorithm for a particular problem can be a difficult task. The following criteria can be used in choosing a clustering algorithm: Input data, cluster model, scalability criterion, noise and presentation of result. Many clustering algorithms assume a certain input data format, others require the availability of a similarity measure or a distance metric for the data. Some algorithms implicitly assume that the cluster structure has certain characteristics. Input data with a lot of noise needs an algorithm that deals with noise efficiently. The scalability criterion deals with the time efficiency of the algorithm to be used. The presentation of the results affects post-processing of any practical applications of cluster analysis.

All the above mentioned criterion suggest the importance of the knowledge about the input data. This means cluster algorithms do a good job when used on appropriate data with parameters that are appropriate for the chosen clustering algorithm.

In the context of financial data we have fairly large noisy sets of data that can be easily represented by correlation matrices constructed from price return data. Because of the volume of data, algorithm efficiency is a concern. For easy inspection, it becomes expedient to first represent the clustering in terms of dendograms, before considering more abstract approaches. It is for this reason that we apply merging algorithms based on correlation matrices to obtain dendograms. We investigate the application of simulated annealing algorithms to obtain optimal configurations. The maximum likelihood method used in this dissertation is chosen in accordance with the method of Giada and Marsili to implement the Noh model for asset pricing, which is based on correlation matrices.

Chapter 3

The Potts Model

In nature there are lots of physical problems or phenomena which can be viewed as optimization processes, and the process of analyzing and understanding them led to the development of mathematical methods. It has become common to solve an optimization problem by identifying a related physical problem and using analogies to solve the optimization problem. The dissertation follows a similar process to solve a data clustering problem in the financial markets.

3.1 Combinatorial Optimization

Solving a combinatorial optimization problem amounts to finding the *best* or *optimal* solution among a finite or countably infinite number of alternative solutions. Most of these problems can be formulated in terms of mathematical terminology and notation so that it can be assumed that the quality of the solution can be quantified and can be compared with that of any other solution.

Solving combinatorial optimization problems results in the construction of appropriate algorithms. These algorithms constitute two classes [28]: 1) *Optimization algorithms*, which constitute enumeration methods such as dynamic programming; 2) *Approximation or heuristic algorithms*, with well-known examples such as local search and randomization algorithms. Some algorithms can be used for both purposes.

In both classes of algorithms, two subclasses can be distinguished. *General algorithms* are applicable to a wide variety of problems and are usually called problem independent algorithms. *Tailored algorithms* use problem-specific information and their application is limited to a restrictive set of problems. The problem with tailored algorithms is that for each type of combinatorial problem a new algorithm must be constructed, and it must be tailored to the problem.

It is desirable to have an approximation algorithm of high quality, applicable to a wide variety of combinatorial optimization problems. Algorithms based on local search are typical examples of generally applicable approximation algorithms, but they are often of low quality.

Simulated Annealing algorithm is a high quality general algorithm. It is a randomization algorithm by nature and it can be viewed as an optimization algorithm, but practically it behaves as an approximation algorithm.

3.1.1 Definition of Combinatorial Optimization Problems

A combinatorial optimization problem is either a minimization problem or maximization problem and is specified by a set of problem *instances* [29]. An instance of a combinatorial optimization problem can be formalized as a pair (S, f), where the solution space S denotes the finite set of all possible solutions and the cost function f is a mapping defined as

$$f : S \to \mathbb{R}. \tag{3.1}$$

In the case of minimization, the problem is to find a solution $i_{opt} \in S$ which satisfies

$$f(i_{opt}) \leq f(i), \forall i \in S. \tag{3.2}$$

In the case of maximization, i_{opt} satisfies

$$f(i_{opt}) \geq f(i), \forall i \in S. \tag{3.3}$$

Such a solution i_{opt} is called a global-optimal solution (either minimal or maximal), or simply an optimum (either maximum or minimum), $f_{opt} = f(i_{opt})$ denotes the optimal cost, and S_{opt} the set of optimal solutions.

3.1.2 Local Search

Local search algorithms constitute an interesting class of general approximation algorithms that are based on stepwise improvement on the value of the cost function by exploring neighborhoods.
Let (S, f) be an instance of a combinatorial optimization problem. A neighborhood structure is a mapping defined as

$$N : S \to 2^S, \tag{3.4}$$

which defines for each solution $i \in S$ a set $S_i \subset S$ of solutions that are close to i in some sense. N is the neighborhood structure. The set S_i is called the neighborhood of solution i, and each $j \in S_i$ is called neighboring solution or neighbor of i. This means that
$j \in S_i \Leftrightarrow$i $\in S_j$ [29].

3.2 Definition of the Potts Model

In 1952, Domb and Potts [30, 31] developed the Potts Model to imitate or model the behavior in which individual objects such as, atoms, animals, social behavior, etc, modify their behavior so as to conform or adjust to the behavior of other individuals in their vicinity. In general, the Potts Model is used to model a system of related or interacting-agents [32]. In finance it can be used to model the collective behavior of groups of stocks. The original model suggested that the Potts model consists of *spins* that are placed on a lattice. The lattice is taken to be a two-dimensional rectangular Euclidean lattice, where each spin can assume q possible values distributed uniformly about a circle at angles $\theta_n = \frac{2\pi n}{q}$.
The interaction ***Hamiltonian*** is given by

$$H_c = J_c \sum_{(i,j)} \cos(\theta_{s_i} - \theta_{s_j}), \tag{3.5}$$

with the sum running over the nearest neighbor pairs *(i,j)* over all lattice sites. The lattice spins take on values ranging from $1, ..., q$. The variable J_c is a coupling constant which determines the interaction strength. This model is usually known as the ***q-Vector Potts Model***. The $q = 2$ Potts model is equivalent to the 2D *Ising Model*.

Hamiltonian usually refers to a mathematical expression used to define or describe all the interactions in a given system. In classical mechanics or physics, it is a function that describes the state of a mechanical system in terms of position and momentum. In quantum mechanics, it corresponds to the total energy of a system [33]. A simpler Hamiltonian is given by

$$H_c = -\frac{1}{2}\sum_{(i,j)} J_{ij}\delta_{s_i s_j}. \tag{3.6}$$

A common generalization is to introduce an external "*magnetic field*" term g, over all the spins in the lattice,

$$H_c = -\frac{1}{2}\sum_{(i,j)} J_{ij}\delta_{s_i s_j} - \sum_i g_i s_i. \tag{3.7}$$

In equation 3.7, the first term of the right hand side is analogous to a cluster term (group of correlated assets). The second term of the RHS is analogous to the market mode.

In physics, the original Hamiltonian equation is given by $\beta H_c = -\beta\sum_{(i,j)} J_{ij}\delta_{s_i s_j} - \sum_i g_i s_i$, where $\beta = \frac{1}{kT}$, k is Boltzmann constant and T the temperature. In the model, s_i is the i-th spin and s_j is the j-th spin, with $i \neq j$.

3.3 Mapping Financial Market data to Potts Model

The Potts model can be used to explain a system of interacting agents or model the dynamics of the collective behavior of stocks. For example, one can consider a magnetic model with random interacting magnets, or a market with N interacting assets [34, 35].

We consider the latter. Let $\overrightarrow{x} = \{x_1^d, x_2^d, ..., x_N^d\}$ represent daily returns of the traded assets. N denotes the number of stocks and d denotes or indexes days.

The cross-correlations between price fluctuations of different stocks can be explained in terms of a magnetic model [35]. In terms of the Potts model, the strengths of interactions between magnets is modelled by J_c, the coupling contstant (see equations 3.5 - 3.7). In financial markets, this is similar to the empirically measured correlations in the matrix C. In financial markets, it is assumed that no trade occurs in isolation. This means that the price fluctuation of stock i is affected by price fluctuation of stock j, $i \neq j$. The families of conditional probabilities of realizing return x_j when return x_i is realized, i.e. the probability that stock j has a price fluctuation of x_j when stock i has a price fluctuation of x_i, are chosen as

$$\mathbb{P}(x_j|x_i) = A\exp[-\beta.J(x_i x_j).\delta_{x_i,x_j}]. \tag{3.8}$$

This probability comes from physics, in classical mechanics, where x_i and x_j are said to be *classical spins* and β is the inverse temperature. In financial context, $J(x_i x_j)$ measures how strongly assets respond to other assets' performances, β measures the strength of each individual asset's response and it is independent of what's going on with other assets. A is a normalization constant to be defined below. δ_{x_i,x_j} is a spin variable which becomes 1 if x_i and x_j belong in the same spin, and takes 0 otherwise.

With the assumption that the behavior of each asset in the market is affected by the behavior of all other assets, we can describe the behavior or the state of the whole market or a certain sector of the market (index). Let's assume that to quantify the state of the market we use

$$M(\overrightarrow{x}) = x_1 + x_1 + ... + x_N. \tag{3.9}$$

Firstly, one needs to define the joint probability distributions of all the market returns from a particular market, $\mathbb{P}(x_1, x_2, ..., x_N)$. This probability distribution can be written as

$$\mathbb{P}(x_1, x_2, ..., x_N) = \prod_{i}^{N-1} \mathbb{P}(x_j|x_i). \tag{3.10}$$

According to equation 3.8 and equation 3.10 we get

$$\mathbb{P}(x_1, x_2, ..., x_N) = \prod_{i}^{N-1} (A \exp[-\beta.J(x_i x_j).\delta_{x_i,x_j}]), \tag{3.11}$$

which becomes [34],

$$\mathbb{P}(x_1, x_2, ..., x_N) = A^N \exp[-\beta \sum_{\langle i,j \rangle} J(x_i x_j).\delta_{x_i,x_j}] \tag{3.12}$$

where A^N becomes the normalization constant such that

$$\sum_{x_1=v_0}^{v_{q-1}} \sum_{x_2=v_0}^{v_{q-1}} .... \sum_{x_N=v_0}^{v_{q-1}} \mathbb{P}(x_1, x_2, ..., x_N) = 1. \tag{3.13}$$

The value of the quantity $M(\overrightarrow{x})$ in equation 3.9 incorporates the fact that assets are interacting with each other, i.e. no trade occurs in isolation, but it does not take into consideration the interaction of each individual asset with the market. The interactions betweens assets are quantified by the measured empirical cross-correlations between stocks.

The aim of this work is to uncover the underlying correlation structure of the market. Essentially this means removing the market mode from the data. This is the reason equation 3.7 was not considered. Equation 3.7 includes the average spontaneous magnetic field. In a random magnet model the average spontaneous magnetic field affects every magnet object in the model. Similarly, in the financial markets the market mode affects price changes of all traded stocks and can be obtained as the first eigenmode (principal component) of the correlation matrix.

Equation 3.5 can also describe the total interaction between the spins or random magnets, without the magnetic field effect. This fact allows us to directly transpose the techniques and ideas of statistical mechanics into a finance framework. Doing this, equation 3.12, i.e the probability density function to present a specific market performance value M, can be written as

$$\mathbb{P}(M) = Z_N^{-1} \exp[-\beta \; H(\overrightarrow{x})], \tag{3.14}$$

where $M = \sum_{i=1}^{N} x_i$. In physics, Z_N is called a *partition function*. It is the normalization constant which is the sum over all possible configurations:

$$Z_N(\beta) = \sum_{x_i=v_0,.....v_q} \exp[-\beta \; H(\overrightarrow{x})]. \tag{3.15}$$

Direct evaluation of the partition function for the market model is not known to be tractable at this time. Hence, an indirect method is used, as set out in the next chapter.

Chapter 4

The Maximum Likelihood Data Clustering Algorithms

4.1 The Maximum Likelihood (M.L.) Model

This model was developed by J. D. Noh to model market returns [9]. The M.L. model models single stock price dynamics, while the Potts model models the collective stocks behavior dynamics. The main aim of this research is to classify similar objects into clusters. In practice, the meaning of similar objects of the same cluster is not clear nor is the natural number of clusters. The M.L. model helps to solve this problem.

This model is a fully unsupervised parameter free data clustering technique which is derived from the maximum likelihood principle. The main statistical hypothesis of this method is that objects are similar, or objects belong in the same cluster or group, if they have something in common.

Let's suppose we have a data set of N objects, $\{\overrightarrow{x}_i\}$, $i = 1, 2, ..., N$, which is defined in terms of D measurable features so that each i-th object is represented by a vector $\overrightarrow{x}_i = \{x_i(d)\}_{d=1}^{D}$. For simplicity, it is assumed that the data is *normalized* to have a zero mean, and a unit variance, that is [9]:

$$\frac{1}{D}\sum_{d} x_i(d) = 0, \tag{4.1}$$

and

$$||\overrightarrow{x_i}||^2 = \frac{1}{D}\sum_{d} x_i^2(d) = 1. \tag{4.2}$$

In this investigation, the objects are assets, i.e. we have N assets or stocks, and their features are the daily returns, i.e. across D days in the data set. i is indexing stocks, d is indexing days. The component of each object is assumed to be derived by the model [9]:

$$x_i(d) = \frac{\sqrt{g_{s_i}}\eta_{s_i}(d) + \epsilon_{s_i}(d)}{\sqrt{1 + g_{s_i}}}. \tag{4.3}$$

A configuration of clusters is represented by a set $S = \{s_i\}$, $s = 1, 2, ..., K$. K is the number of clusters in the data set with N assets[1]. If $s_i = s_j = s$, then object i and object j reside in the same cluster. If each cluster is allowed to contain one object, then $K = N$, and i is allowed to take integer values from 1 through N. If each cluster s_i contains n_{s_i} objects, then $\sum_{i=1}^{K} n_{s_i} = N$.

The variable $\overrightarrow{\eta}_{s_i}$ is the vector of features of cluster s_i, the common component shared by all objects in cluster s_i. The measure of similarity of the objects within cluster s_i is given by g_{s_i}, i.e. $g_{s_i} = 1$ means all objects within s_i are identical whereas a g_{s_i} close to zero means objects are very different.

Here $\overrightarrow{\epsilon_{s_i}}$ refers to the deviation of the features of object i from the features of cluster s_i. This means g_{s_i} weights the common component against the individual error measurement $\overrightarrow{\epsilon_{s_i}}$. For the simple statistical hypothesis equations of this model, it is further assumed that both $\overrightarrow{\eta}_{s_i}$ and $\overrightarrow{\epsilon_{s_i}}$ are Gaussian vectors with zero average and unit variance, i.e:

$$\overrightarrow{\eta_{s_i}} \sim N(0, 1), \tag{4.4}$$

$$\overrightarrow{\epsilon_{s_i}} \sim N(0, 1). \tag{4.5}$$

Equations 4.3 can be regarded as a *stochastic generalization of multi-factor pricing models*. One can compare equation 4.3 with the following pricing models (*Capital Asset Pricing Model*, *Factor Model*, and *Arbitrage Pricing Model*):

$$\overrightarrow{x}_i = \alpha_i + \beta_i x_m + \epsilon_i, \tag{4.6}$$

$$\overrightarrow{x}_i = \sum_j F_j x_m^j + \epsilon_i, \tag{4.7}$$

$$\overrightarrow{x}_i = a_i + \sum_j b_{ij} I_j + \epsilon_i. \tag{4.8}$$

For any given set of parameters $(G, S) = (\{g_s\}, \{s_i\})$, it is possible to compute the probability of observing the data set $\{\overrightarrow{x}_i\}$, $i = 1, 2, ..., N$, as a realization of the model in equations 4.3, that is the likelihood:

$$\mathbb{P}(\{\overrightarrow{x_i}\}|\mathcal{G}, \mathcal{S}) = \prod_{d=1}^{D} \langle \prod_{i=1}^{N} \delta(x_i(d) - \frac{\sqrt{g_{s_i}}\eta_{s_i}(d) + \epsilon_{s_i}(d)}{\sqrt{1 + g_{s_i}}}) \rangle. \tag{4.9}$$

$\langle .... \rangle$ denotes the mathematical expectation and δ is the Dirac delta function. Equation 4.9 describes a type of Potts model, but does not require the specification of topology (on the lattice). Solving this likelihood function results in

$$\mathbb{P}(\{\overrightarrow{x_i}\}|\mathcal{G}, \mathcal{S}) = e^{-DL\{\mathcal{G}, \mathcal{S}\}}. \tag{4.10}$$

For any given configuration structure S, the likelihood $\mathbb{P}(\{\overrightarrow{x_i}\}|\mathcal{G}, \mathcal{S})$ described by equation 4.10, is maximal when $g_s = g_s^*$ where, [7]:

$$g_s^* = \frac{c_s - n_s}{n_s^2 - n_s}. \tag{4.11}$$

[1] This notation is not necessarily the same as in section 2.2.

Here n_s refers to the number of stocks identified to be part of a given correlated group of stocks, cluster s:

$$n_s = \sum_{i=1}^{N} \delta_{s_i s}. \tag{4.12}$$

The sectors are assumed to only have positive internal correlations c_s, and to be uncorrelated with other sectors. The internal correlation, c_s, or the total correlation of objects in cluster s is defined as:

$$c_s = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{C}_{i,j} \delta_{s_i,s} \delta_{s_j,s}. \tag{4.13}$$

Here C_{ij} is the Pearson's correlation and is defined as:

$$\mathcal{C}_{i,j} = \frac{\overrightarrow{x_i}.\overrightarrow{x_j}}{\sqrt{||\overrightarrow{x_i}||^2 ||\overrightarrow{x_j}||^2}}. \tag{4.14}$$

The maximum likelihood per configuration structure S, $\mathbb{P}(\{\overrightarrow{x_i}\}|\mathcal{G}^*, \mathcal{S}) = e^{-DL_c\{\mathcal{S}\}}$, has a maximum log-likelihood of the structure S which is given by

$$L_c(\mathcal{S}) = \frac{1}{2} \sum_{s:n_S>1} [\log \frac{n_S}{c_S} + (n_S - 1) \log \frac{c_S - n_S}{n_S^2 - n_S}]. \tag{4.15}$$

The maximum log-likelihood does not depend on any parameter, it depends only on the original data through the Pearson's correlation coefficient, which is contained in c_s; hence the model is classified as an unsupervised nonparametric technique.

The model classifies objects according to their similarity, not dissimilarity. In the algorithm it must therefore be included that, if objects in cluster s are dissimilar or uncorrelated, then $g_s^* = 0$ or $c_s = n_s$ and hence $L_c(\mathcal{S}) = 0$. Also if every cluster contains a single object, that is $n_s = 1$ for all s, then $L_c(\mathcal{S}) = 0$. This essentially means that, the maximum log-likelihood measures the compatibility of the configuration structure S with the observed data set $\{\overrightarrow{x}_i\}$, $i = 1, 2, ..., N$. In maximum likelihood data clustering method, the log-likelihood in equation 4.15 is used as a measure of similarity between objects, and also between clusters.

4.2 The connection between the Potts Model and M.L. Model

We have considered two systems of interacting-agents, the random asset prices of the market and random magnets. Analogies can be made between the standard formulation of the Potts model and the Maximum Likelihood formulation. The M.L. formulation of the Potts model given in equation 4.9, which uses the measure of similarity in equation 4.11, allows the following interpretation:

- Hamiltonian Function $\Leftrightarrow$ Log-likelihood function.
- Coupling constant $\Leftrightarrow$ Pearson's Correlation Coefficient.
- Potts Spins $\Leftrightarrow$ Stock Membership in a given cluster.

4.3 The Merging Algorithm

The Merging Algorithm is a simple algorithm which was proposed to be used in hierarchical clustering using as a cost function the log-likelihood L_c, derived from the Maximum Likelihood Clustering technique [27]. It is described as follows:

0. Start with N clusters composed of single object clusters , *i.e* $n_s = 1, \forall\ s$,
1. At each step of the algorithm, merge two clusters into single one in such way that the cost function L_c of the current configuration S is minimal,
2. Repeat step 1. $N - 1$ times until the configuration with a single cluster is reached.

The output of this algorithm is represented in a form of a *dendrogram tree*. In our investigation, a clear and rich structure in the tree will provide evidence that there is an underlying structure of correlated asset returns (sectors) and correlated market-wide activities (states) in the SA financial markets.

4.4 Simulated Annealing

4.4.1 The Metropolis Scheme

In condensed matter physics, *annealing* is known as a thermal process for obtaining low energy states of a solid in a *heat bath*[2]. An algorithm to simulate the cooling of material in a heat bath was first published by Metropolis *et al.* (1953) [36]. Later, Kirkpatrick *et al.* (1983), [37], demonstrated that this type of simulation could be used to search the feasible solutions of a combinatorial optimization problem, with the objective of converging to an optimal solution. The annealing process contains two steps:

- Increase the temperature of the heat bath to a maximum value at which the solid melts.
- Decrease the temperature of the heat bath until the particles arrange themselves in the ground state of the solid.

In the liquid phase all the particles of the solid arrange themselves randomly. In a ground state the particles are arranged in a highly structured lattice and the energy of the system is globally minimal. The ground state of the solid is obtained only if the maximum temperature is sufficiently high and cooling is done sufficiently slowly, so that the system is approximately at thermodynamic equilibrium at any time. Otherwise the solid will be frozen into a meta-stable state rather than a ground state.

The converse of annealing is referred to as quenching, where the temperature of the heat bath is instantaneously lowered. This also results in a meta-stable state. If the initial temperature of the system is too low or the cooling is done too quickly, the system may be quenched forming defects or freezing in a meta-stable state, i.e., trapped in a local minimum energy state, instead of a global minimal energy state. The Metropolis algorithm is based on Monte Carlo techniques and generates a sequence of states of the solid in the following way:

Given the current state i of the solid with energy E_i and temperature T, a subsequent state j is generated by holding T constant and applying a perturbation mechanism which transforms the current state into the next state by a small distortion, for instance by replacing a single particle. The energy of the next state is E_j. If the *energy difference* , $dE = E_j - E_i$, is less than 0, the state j is accepted as the current state. If the energy difference is greater than 0, the j state is accepted with a certain probability which is given by

$$P = exp(-\frac{E_j - E_i}{kT}), \tag{4.16}$$

[2] A heat bath is a large system that is in thermal contact with some other system of interest [29].

where T denotes the temperature of the heat bath and k a physical constant known as the *Boltzmann Constant*. The acceptance criterion above is called a *Metropolis criterion*, and the algorithm that goes with it is called the *Metropolis Algorithm*.

This process is repeated sufficient times to give a good sampling statistic for the current temperature. The temperature is then decreased and the entire process is repeated at each temperature, until a frozen state is achieved at $T = 0$.

4.4.2 Simulated Annealing Algorithm

The simulated annealing algorithm, given on page 17, applies the Metropolis algorithm to generate a sequence of solutions of a combinatorial optimization problem. For this purpose, an analogy between a physical many-particle system and a combinatorial optimization problem is assumed based on the following equivalences [29].

- The infinitely many solutions in a combinatorial optimization problem are equivalent to states of a physical system.
- The cost of a solution is equivalent to the energy of a state.

Another parameter, which plays an important role of the temperature, is introduced. This parameter is called the *control parameter*. Essentially, the simulated annealing algorithm can be viewed as an iteration of the Metropolis algorithm, evaluated at decreasing values of the control parameter. The control parameter, let's say cf is defined to be the inverse temperature of the Boltzmann's distribution, kT. Boltzmann's probability becomes

$$exp(-\frac{E_j - E_i}{cf}). \tag{4.17}$$

The generation mechanism of an new solution S_j corresponds to the Metropolis algorithm, whereas the acceptance criterion corresponds to the Metropolis criterion.

A transition is a combined action resulting in the transformation of a current solution into a subsequent one. The action consists of the following two steps: (i) application of the generation mechanism, (ii) application of the acceptance criterion. According to the algorithm in page 17, (i) corresponds to the outer while loop and (ii) corresponds to the inner while loop.

Essentially, simulated annealing is a random search method that avoids getting trapped in local optima by accepting, in addition to solutions that improve on the value of an objective function, also solutions corresponding to a deteriorated cost function value.

In the course of the simulated annealing optimization process, the probability of accepting the deteriorated solutions descends slowly as the temperature drops. These deteriorations make it possible for simulated annealing algorithm to move away from local optima and explore the feasible solution space in its entirety.

```
Simulated Annealing Algorithm

0.  Starting condition values:
    T0 - initial temp.; T_limit - limit temp.; E0 - initial energy;
    cf - control parameter(0 < cf < 1);  S0 - initial configuration;
    SPT- number of Sweeps Per Temperature)
1.  S_best  = S0; (the best configuration solution so far)
    T = T0;      S_i = S0;
2.  While STOP = false (conditionanl loop - Annealing cycle)
         STOP = true;
         sweep_counter = 1;
         While sweep_counter <= SPT (Sweep cycle)
               sweep_counter ++;
               Find S_try;  (sweep or randomly choose new configuration)
               dE = E(S_try) - E(S_best); (cost or likelihood)
               if dE < 0, then
                    S_best = S_try; (Accept improvement)
                    STOP = false;
               else,
                    p = exp(-dE/T); (Boltzman)
                    u = random U(0,1);
                    if u < p, then
                         S_i = S_try; (Accept non-improvement)
                         (but S_best stays the same)
                         STOP = false;
                    end if;
               end if
        end while
        if T  has reached T_limit, then
            STOP = true;
        end if
        if STOP = false
             T = T x cf;
        end if
    end while
```

Chapter 5

Analytical Methods and Results

5.1 Data Manipulation

The data considered in this investigation is a 10 year data set of daily close prices of stocks traded in the Johannesburg Stock Exchange (JSE) from 01 January 1993 to 31 December 2003. For both clustering algorithms investigated in this paper (Recursive Merging Algorithm and Simulated Annealing Algorithm), data clustering is applied on different time windows or epochs of the original data given below.

1. From 1993 to 2003, the original data sets,
2. Epoch #1: From 1993 to 1997,
3. Epoch #2: From 1994 to 1998,
4. Epoch #3: From 1995 to 1999,
5. Epoch #4: From 1996 to 2000,
6. Epoch #5: From 1997 to 2001,
7. Epoch #6: From 1998 to 2002.

This was done to compare the consistency of the results for both algorithms.

5.1.1 Data Filtering

The original data contains 442 stocks, but the number of stocks traded in the JSE varies with time. Therefore the focus is on the most actively traded stocks in the JSE. All stocks which traded less than 75 percent of the ten year period were not considered, they were filtered out. This resulted into $N = 136$ most actively traded stocks.

5.1.2 Data Normalization

The main aim of the project is to uncover the internal structure of the correlations across stocks (sectors) and across days (states). Before computing the correlation matrix, the returns were normalized to get rid of the annual, monthly or even daily common trends or patterns of market activities. We used a recursive averaging method to normalize the data and in that way we eliminated the constant correlations of individual asset returns with the market return. Marsili *et al.* [8] use the same method to remove the market mode. One can also use Random Matrix Theory (RMT) to eliminate the eigenvalue associated with the market to remove the market influence [38]. This is the market mode which captures the collective response of the entire market to stimuli such as interest rate changes or currency shocks.

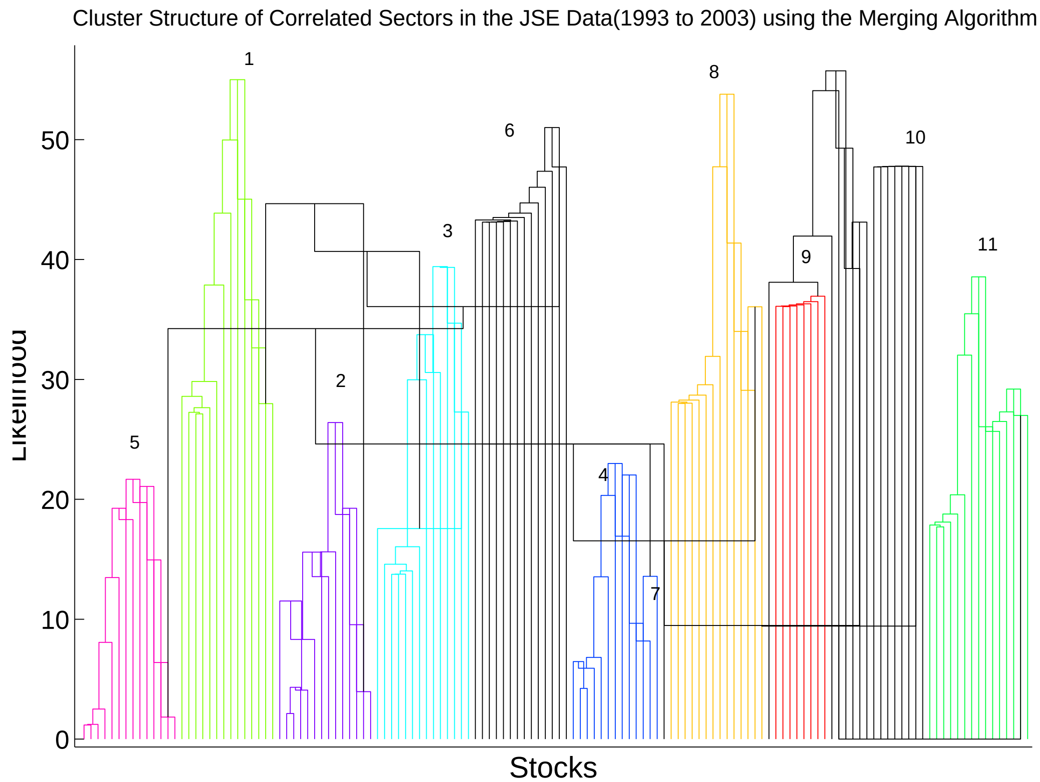


Figure 5.1: This figure shows a dendrogram with market sectors during the period 1993-2003. The list of stocks found in all other clusters of this dendrogram is found in Table 5.1. Most clusters, which are numbered 1-11, are made up of a mixture of all sorts of stocks from different economic sectors.

5.2 Market Sectors Clustering Results

5.2.1 Recursive Merging Algorithm (MA) Results

The recursive merging algorithm was applied to find the underlying hierarchical structure of groups of stocks with similar economic dynamics, with an output in the form of a dendrogram. A dendrogram with market sectors during the period 1993-2003 is shown in Figure 5.1. On the dendrogram the labels give numbers to each cluster of stocks. The list of all stocks found in each cluster are listed in Table 5.1. It can be seen from Table 5.1 that all sectors or clusters are made up of a mixture of assets from different market sectors. Cluster 1 is a mixture of all sorts of stocks from Banks to Mining. Cluster 2 consists of mainly Mining companies. Cluster 3 is made up of insurance, banks and general retailers.

Dendrograms for other epochs are shown in Figure 5.2, and the lists of all stocks found in each cluster, for all epochs, are listed in appendix D.

Table 5.1: Market Sectors or Clusters from the dendrogram in Figure 5.1.

Cluster 1	AFN.BK.INVS.; ANGLOVAAL INDUSTRIES; BOWLER METCALF; BRANDCORP; BIDVEST GROUP; ELLERINE; FIRSTRAND; IMPERIAL HDG.; INVICTA; JD GROUP; NEW AF.CAP.; RMB HDG.; WOOLTRU N; WINHOLD;
Cluster 2	AFRIKANDER LEASE; ANGLOGOLD; BARPLATS INVS.; CORPCAPITAL; DURBAN ROOD.DEEP; GOLD FIELDS; HARMONY GOLD MINING; JOHNNIC HDG.; KAIROS INDUSTRIAL; NORTHAM PLAT.; NU-WORLD; STH.AFN.CHROME & ALS.; SOFTLINE; TRANS HEX GROUP;
Cluster 3	AECI; ABSA GP; BEV. & CONSUMER IND. DEAD; EDGARS CONS.STORES; LIBERTY HDG; LIBERTY GP; NEDCOR; PEPKOR; PIK N PAY; SABMILLER (JSE); STD.BK.GP; SANTAM; VENFIN; WOOLTRU;
Cluster 4	AFRICAN OXYGEN; ALLIED TECHS.; ALLIED ELECTRONICS; GRINDROD; ALLAN GRAY PR.TST; HIGHVELD STEEL & VANADIUM; LA GROUP; ELT.MEDIA & SUPERSPORT; PANGBOURNE PROPS; SEARDEL INV; SPESCOM; SYCOM PROPERTY FD; TIGER BRANDS; UNITRANS;
Cluster 5	AFROX HEALTHCARE; ASPEN PHARMACARE; BRAIT SA.(JSE); BARNATO EXPLORATION; CITY LODGE HOTELS; CONTROL INSTRUMENT GROUP; DON GROUP; JASCO ELECTRONICS; KWV BELEGGINGS BEPERK; RELYANT RETAIL; RENTSURE HDG; SASFIN; SASANI; ZARARA EN;
Cluster 6	AFRICAN MEDIA ENTM.; BASIL READ; BYTES TECH.GP; DS & WAREHOUSING NTWRK; DNA SUPPLY CHAIN INV; ELB GP; GRINTEK; HUDACO; PALABORA MINING; RAND LEASES PROPS.; SALLIES; TRENCOR; CONCOR;
Cluster 7	S AND J LAND;
Cluster 8	ANGLOVAAL MINING; CULLINAN; DIMENSION DATA HDG (JSE); DORBYL; DISTELL GP; FOSCHINI; ADCORP; AVGOLD; CAXTON CTP PUBLISH PRINT; CASHBUILD; DELTA ELECTRICAL INDS.; FREE STATE DEV. & INV.;MEDICLINIC; PSG GROUP;
Cluster 9	BARLOW (ISE); AMAL.BEVERAGES IND; BARLOWORLD; CHEMICAL SERVICES; ALLIANCE; GILBOA PROPERTIES; KERSAF INVESTMENTS; MUTUAL; MOBILE INDUSTRIES; MR PRICE GP; MURRAY & ROBERTS; RAINBOW CHICKEN; REUNERT; SUN INTL (JSE);
Cluster 10	ISCOR; LONMIN (JSE); MATODZI RESOURCES; PIK N PAY STORES; MSASOL; UNITED SERVICE TECHS; PRETORIA PORT.CMT.; SUPER GROUP; WILSON BAY HLM-OVC;
Cluster 11	ANGLO AMERICAN (JSE); ANGLO AMERICAN PLAT; IMPALA PLATINUM; INVESTEC; INMINS; JOHNNIC COMMS.; METOREX; MVELAPHANDA RES.; NAMPAK; OCEANA GP; RICHEMONT SECS. (JSE); SAPPI; SUB NIGEL GDMNG.; TONGAAT-HULETT GROUP; WESTERN AREAS

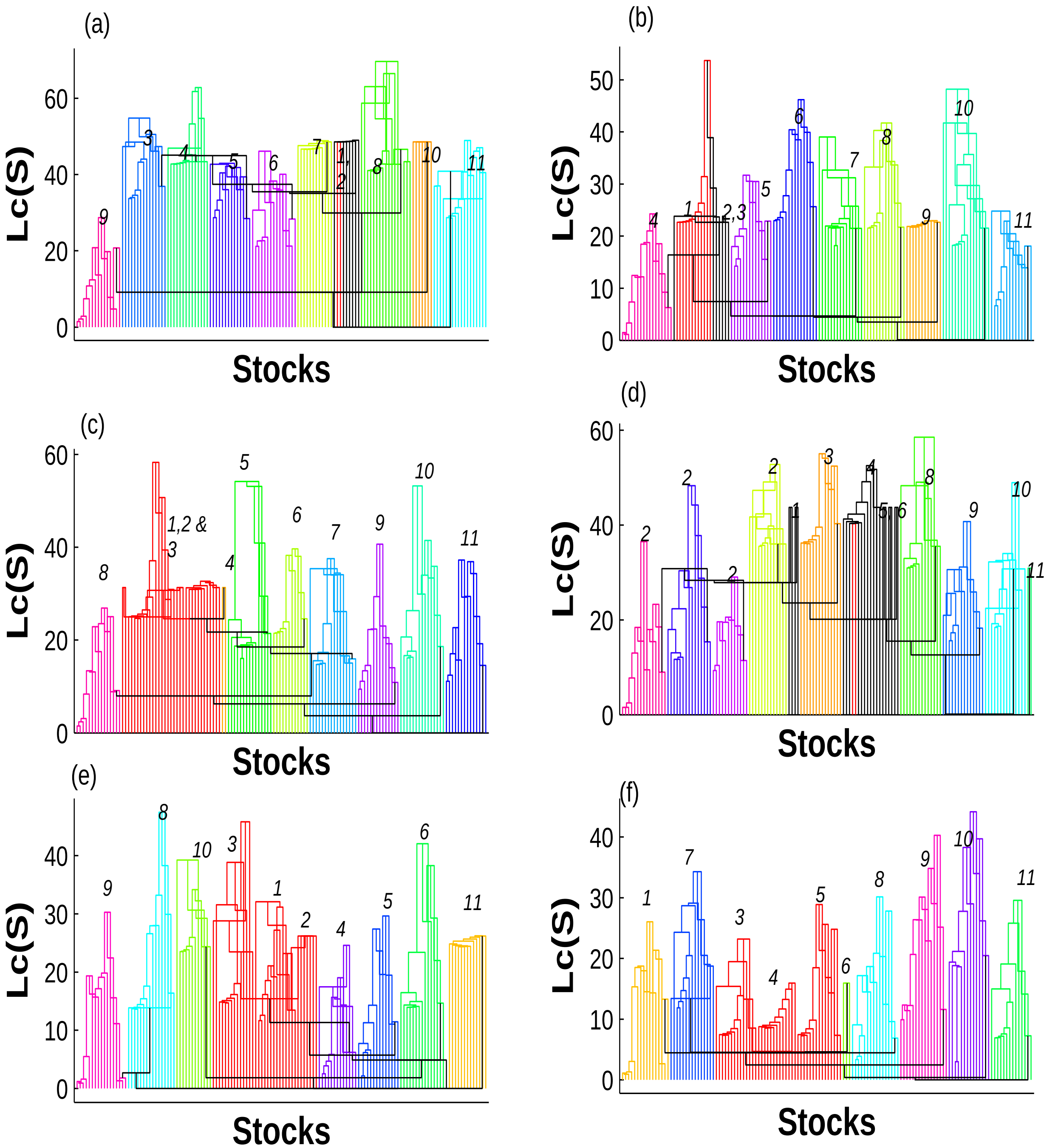


Figure 5.2: This figure shows dendrograms of market sectors during the following time period windows: a) 1993-1997: Most mining companies are in clusters 5 and 6. b) 1994-1998. Cluster 5 is a mixture of Mining and Resources. c) 1995-1999. d) 1996-2000. e) 1997-2001. f) 1998-2002. (See appendix D for the list of stocks found in each sector.)

5.2.2 Simulated Annealing (SA) Clustering Results

In terms of the final solutions, the major difference between the recursive merging algorithm and the simulated annealing algorithm is that simulated annealing gives one optimal solution for a clustering problem while with MA we get several solutions depending on where the dendrogram tree or the linkage matrix is cut. Table 5.2 shows market sectors obtained using simulated annealing for the period 1993-2003. The rest of the simulated annealing clustering results are shown in appendix E.

In the simulated annealing results, economic sectors like gold mining, platinum mining, banks, insurance, transport and resources came out as distinct clusters. Nevertheless there is a lot of overlapping across sectors of different economic activities. This implies that there are significant correlations between assets from different economic sectors.

From table 5.2 we notice the existence of singleton clusters. We conjecture that assets that are more sensitive or more correlated to the market (e.g. retailers) came out in singleton clusters. One can check this by estimating betas of these assets. Beta is defined as $\beta_i = \frac{\sigma_{iM}}{\sigma_M^2}$, where σ_M^2 is the variance of the market or market index and σ_{iM} is the covariance of stock i and the market. Assets that have larger betas are more sensitive to the general market movement. The market mode was removed from the data and this made all the stocks that are highly correlated to the market to contribute to noise. To prove this we undressed the noise from the correlation matrix of epoch #6 (time period 1998-2002). This is discussed in the next section.

5.2.3 Dealing with noise on the Market Sectors Clustering

To show that small or singleton clusters can be the result of noise, we cleaned the covariance matrix of the data and calculated a new correlation matrix as follows. Random Matrix Theory (RMT) can be used to distinguish between random and non-random part of the correlations. RMT establishes that all the eigenvalues within a computable range $[\lambda_{min}, \lambda_{max}]$ may be attributed to random effects [2, 3, 7]. A simple denoising algorithm employs this idea by deleting eigenvalues in the noise range $[\lambda_{min}, \lambda_{max}]$ from the diagonalised correlation matrix [2, 3, 7]. After applying this method we used the cleaned correlation matrix to run the SA algorithm. The data set used in this simulation is epoch #6 (time period 1998-2002) and we present the configuration of clusters for SA results in table 5.3. The number of clusters was reduced significantly in table 5.3. This provides the empirical motivation that most of the arbitrary groupings and singletons of stocks in table 5.2 can be attributed to noise.

Table 5.2: Market Sectors using Simulated Annealing(SA) for the period 1993-2003.

```
1.     ANGLOGOLD;  GOLD FIELDS;    HARMONY GOLD MINING;    DURBAN ROOD.DEEP;
       WESTERN AREAS;----
2.     RICHEMONT SECS. (JSE);     SANTAM;  STD.BK.GP. ;----
3.     ANGLO AMERICAN PLAT.;    IMPALA PLATINUM ;
4.     S & J LAND;----
5.     FOSCHINI;----
6.     HUDACO;----
7.     RAINBOW CHICKEN ;----
8.     CONCOR;    PEPKOR;----
9.     AFN.BK.INVS;   GILBOA PROPERTIES;----
10.    ANGLO AMERICAN (JSE) ;----
11.    MVELAPHANDA RES.;   AFROX HEALTHCARE;----
12.    FIRSTRAND;   RMB HDG;----
13.    BEV.& CONSUMER IND;----
14.    NEDCOR ;----
15.    PSG GROUP;   ISCOR ;----
16.    CASHBUILD;----
17.    RAND LEASES PROPS;   ALLIED TECHS;----
18.    UNITED SERVICE TECHS; WOOLTRU N; TONGAAT-HULETT GROUP; CITY LODGE HOTELS; ANGLOVAAL MINING;--
19.    JOHNNIC HDG;----
20.    NEW AF.CAP.;  JASCO ELECTRONICS;----
21.    INVICTA;----
22.    AFRICAN MEDIA ENTM.;  DON GROUP;  RENTSURE HDG.;  ELB GP.;  DNA SUPPLY CHAIN INV;----
23.    NORTHAM PLAT;----
24.    GRINDROD;  UNITRANS;----
25.    BASIL READ;    ZARARA EN;----
26.    MOBILE INDUSTRIES;     KWV BELEGGINGS BEPERK
27.    ABSA GP.;   HIGHVELD STEEL&VANADIUM;   JOHNNIC COMMS.
28.    SUPER GROUP;----
29.    PIK N PAY;----
30.    PIK N PAY STORES;----
31.    BOWLER METCALF;----
32.    LIBERTY GP;   LIBERTY HDG;----
33.    CONTROL INSTRUMENT GROUP ;----
34.    SEARDEL INV:  JD GROUP;----
35.    PRETORIA PORT.CMT;   REUNERT;----
36.    ELLERINE; ----
37.    SPESCOM;----
38.    CORPCAPITAL;    RELYANT RETAIL;    MATODZI RESOURCES
39.    DELTA ELECTRICAL INDS.;    AFRICAN OXYGEN;----
40.    KAIROS INDUSTRIAL;     MURRAY & ROBERTS;     DS.& WAREHOUSING NETWORK
41.    BARNATO EXPLORATION;    LA GROUP;----
42.    EDGARS CONS.STORES;----
43.    SAPPI;    SUN INTL.(JSE) ;----
44.    BARLOW (ISE);----
45.    INMINS;----
```

Table 5.2 continues.......

46. ALLIED ELECTRONICS; SASANI; BYTES TECH GP; ANGLOVAAL INDUSTRIES;----
47. AECI;----
48. MR PRICE GP;----
49. METOREX;----
50. CAXTON CTP PUBLISH PRINT;----
51. SUB NIGEL GDMNG; PANGBOURNE PROPS;----
52. MEDI CLINIC;----
53. BRANDCORP; KERSAF INVESTMENTS
54. VENFIN; SOFTLINE; SASFIN
55. CAPITAL ALLIANCE; OCEANA GP; SASOL; SABMILLER (JSE); FREE STATE DEV & INV.
56. DIMENSION DATA HDG(JSE) ;----
57. MUTUAL & FEDERAL IN;----
58. TRANS HEX GROUP;----
59. ALLAN GRAY PR.TST;----
60. STH.AFN.CHROME & ALS;----
61. CHEMICAL SERVICES;----
62. LONMIN (JSE) ;----
63. CULLINAN;----
64. IMPERIAL HDG; NU - WORLD; GRINTEK;----
65. BARLOWORLD;----
66. BIDVEST GROUP; NAMPAK;----
67. BARPLATS INVS.; WOOLTRU ;----
68. ELT.MEDIA NETWORK & SUPERSPORT;----
69. SALLIES; PALABORA MINING;----
70. WINHOLD; TIGER BRANDS;----
71. TRENCOR;----
72. AVGOLD;----
73. AFRIKANDER LEASE;----
74. WILSON BAY HLM - OVC;----
75. ADCORP;----
76. ASPEN PHARMACARE; BRAIT SA. (JSE) ;----
78. SYCOM PROPERTY FD; DISTELL GP;----
79. AMAL.BEVERAGES IND;----
80. NVESTEC; DORBYL;----

Table 5.3: Market Sectors during the period 1998-2002 using SA with noise-undressed correlations.

1. AFRICAN OXYGEN; FOSCHINI;
2. NAMPAK; NEW AF.CAP; LIBERTY HDG; BARLOWORLD; SABMILLER; (JSE) JOHNNIC HDG; TIGER BRANDS; VENFIN; LIBERTY GP; TONGAAT - HULETT GROUP; DIMENSION DATA HDG.(JSE); PEPKOR;
3. DURBAN ROOD.DEEP; ANGLOGOLD; GOLD FIELDS; HARMONY GOLD MINING; AVGOLD; WESTERN AREAS; AFRIKANDER LEASE;
4. RAND LEASES PROPS; CONCOR; CITY LODGE HOTELS; ALLAN GRAY PR.TST;
5. FIRSTRAND; ABSA GP; STD BK GP; RMB HDG; NEDCOR ; INVESTEC; BIDVEST GROUP; IMPERIAL HDG; JD GROUP;
6. ZARARA EN; DNA SUPPLY CHAIN INV; ELB GP; DON GROUP; RENTSURE HDG; AFRICAN MEDIA ENTM.
7. NORTHAM PLAT.; ALLIED TECHS.
8. SASOL; ANGLO AMERICAN (JSE); ANGLO AMERICAN PLAT; RICHEMONT SECS. (JSE); SAPPI; IMPALA PLATINUM; ISCOR; JOHNNIC COMMS
9. PRETORIA PORT.CMT; SANTAM;
10. ASANI;
11. BASIL READ;
12. PIK N PAY STORES; BARPLATS INVS.;
13. REUNERT;
14. LONMIN (JSE);
15. MR PRICE GP.;
16. AECI;
17. DELTA ELECTRICAL INDS.;
18. PIK N PAY; ELLERINE; CORPCAPITAL;
19. CAPITAL ALLIANCE; GRINDROD; WOOLTRU N
20. JASCO; ELECTRONICS;
21. ANGLOVAAL MINING; PALABORA MINING; DORBYL; RAINBOW CHICKEN; SUN INTL.(JSE); TRENCOR; HUDACO; MOBILE INDUSTRIES; BRAIT SA. (JSE); RELYANT RETAIL; ALLIED ELECTRONICS; SYCOM PROPERTY FD; KERSAF INVESTMENTS; KWV BELEGGINGS BEPERK
22. MUTUAL & FEDERAL IN.
23. AFN.BK.INVS; UNITED SERVICE TECHS; CHEMICAL SERVICES; SUB NIGEL GDMNG; MATODZI RESOURCES; WINHOLD; CAXTON CTP; PUBLISH PRINT; NU - WORLD; MVELAPHANDA RES; OCEANA GP; GILBOA PROPERTIES; INMINS ADCORP; WILSON BAY HLM
24. AMAL.BEVERAGES IND.;
25. LA GROUP;
26. FREE STATE DEV.& INV.;
27. CONTROL INSTRUMENT GROUP;
28. HIGHVELD STEEL&VANADIUM; BEV.& CONSUMER IND. DEAD;
30. ANGLOVAAL INDUSTRIES; SOFTLINE; WOOLTRU;
31. SASFIN; CASHBUILD;
32. METOREX; TRANS HEX GROUP; CULLINAN;
33. SPESCOM;
34. BARLOW (ISE)
35. MEDI CLINIC; DS & WAREHOUSING NETWORK;
36. SEARDEL INV.
37. BYTES TECH.GP; AFROX HEALTHCARE;
38. GRINTEK; ELT.MEDIA NETWORK & SUPERSPORT

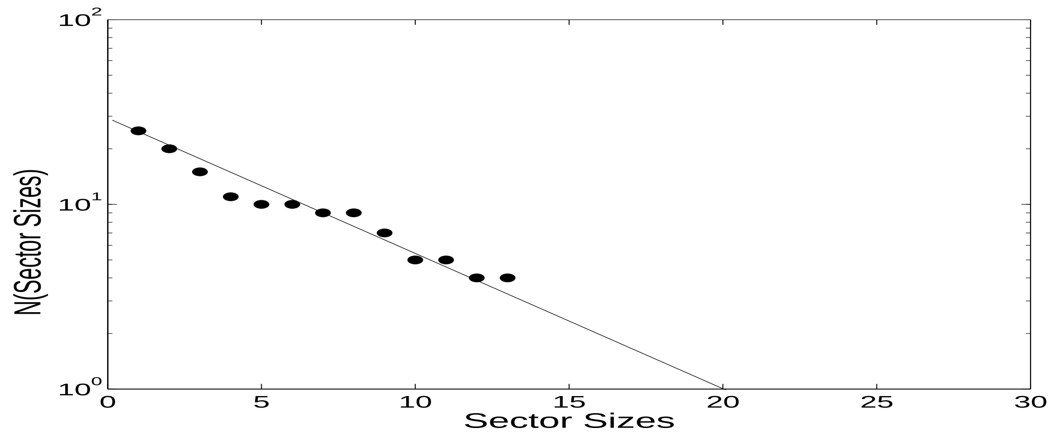


Figure 5.3: A linear-log graph that shows the distribution of the correlated market sectors sizes (n_s) as defined according to the standard industrial classifications. It is evident that the distribution of sectors sizes ($N(n_s)$) seems to be following the power law, $N(n_s) \sim n_s^{\Delta l}$, with $N(n_s) > n_s$.

5.2.4 Analysis of Market Sectors Results

The sector size n_s gives a number of stocks in cluster s, and as described in equation 4.12, this refers to the number of stocks identified to be part of a given correlated group of stocks. The sectors are assumed to only have positive internal correlations c_s, and to be uncorrelated with other sectors. The number of sector sizes $N(n_s)$ refers to a number of sectors of a particular size n_s.

The distribution of sectors sizes has been claimed to follow the power law [40]: $N(n_s) \sim n_s^{\Delta l}$, $N(n_s) > n_s$. This conclusion is consistent with the results of the work done by Marsili [8] and has been heuristically corroborated in this work as shown in figures 5.3 and 5.4, where Δl is the slope of the lines in both graphs. Essentially this means that the number of sectors of a given size, denoted $N(n_s)$, greater than n_s is given as being inversely proportional to n_s.

Figure 5.3 shows the distribution of sector sizes as defined according to the standard industrial classifications (SICs) in South Africa. Figure 5.4 shows distribution of sectors sizes from the results of the SA clustering algorithms, for all time-windows considered, and they confirm the robustness of the above analysis. The same behavior is found in all the time-windows.

The internal correlation c_s refers to the correlation inside cluster s, which is described by Eq. 4.13. The distribution of the internal correlation of cluster s is shown Figure 5.5. With the power law, the distribution is $c_s \sim n_s^{\Delta l}$. Δl is the slope of the line in the picture.

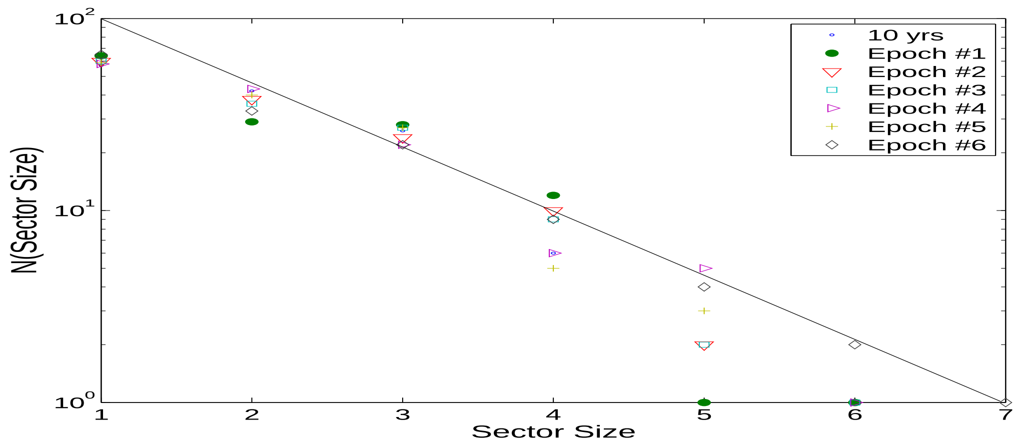


Figure 5.4: This linear-log graph shows the distribution of the correlated market sectors sizes (n_s) for all the time-windows considered using the simulated annealing algorithm. For all epochs, the distribution of sectors sizes ($N(n_s)$) seems to be following the power law, $N(>n_s) \sim n_s^{\Delta l}$.

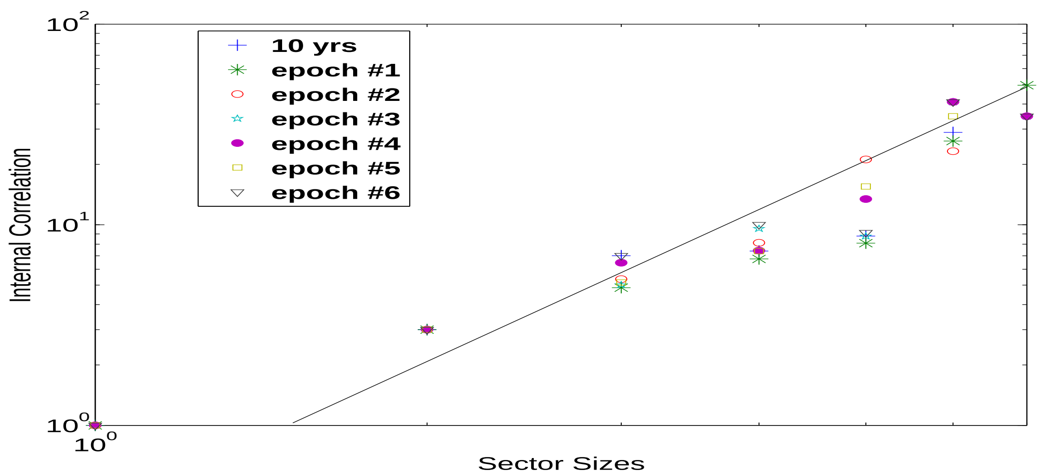


Figure 5.5: A log-log graph that shows the distribution of the internal correlation of cluster s for all the time-windows considered using the simulated annealing algorithm. The distribution of the internal correlation of cluster s is given by $c_s \sim n_s^{\Delta l}$. The condition, $N(n_s) > n_s$, also applies here .The plot shows that there is an underlying internal structure of correlated sectors.

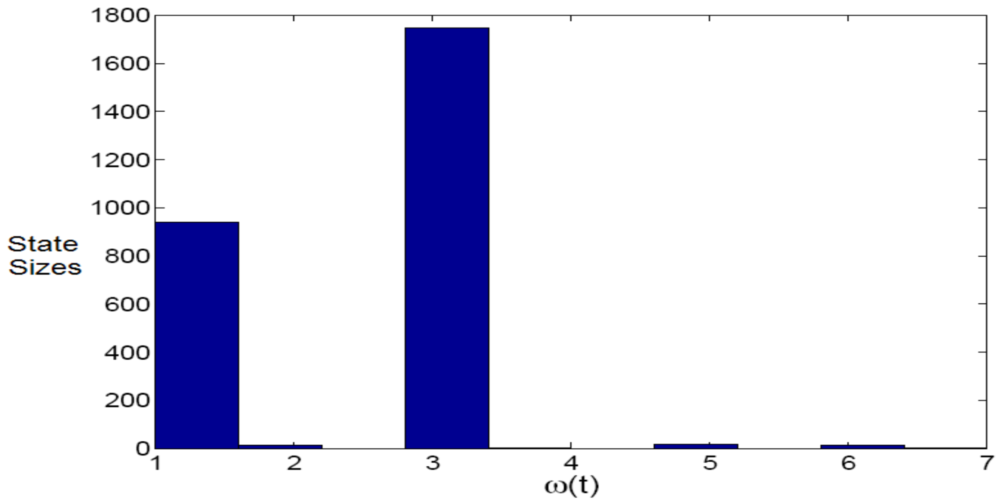


Figure 5.6: This histogram shows states vs the number of days in each state. We have seven states, $\omega(t)$ = 1,2,....7, over t = 1,2,...,T = 2736. A time series plot of states that occurred over the entire data set is shown in figure 5.7. There are two big states on the graph. These two states correspond to the market states that resulted from random effects. This means that there are very few meaningful or most frequent market patterns in South Africa. The time series can be used with other processes like an AR(1) process for further analysis.

5.3 Market States Clustering Results

In this part of the dissertation we are trying to answer the question of whether there are well-defined patterns of daily market-wide economic activities. In this section the data clustering process classifies days, instead of stocks, according to the performance of different stocks.

5.3.1 Recursive Merging Algorithm (MA) Results

Firstly, the data of stocks price returns is transposed and the correlation matrix computed represents the correlations across days. Since there are 2736 days in that ten year period, the correlation matrix C is a 2736 x 2736 matrix. The output of the MA is a linkage matrix which can be used to draw a dendrogram. The dendrograms could not be drawn in matlab. Matlab's recursive limit is not enough to draw the dendrograms of the data set of that size. Instead of drawing the dendrogram, histograms of states vs the number of days appearing in each state were drawn. The histogram is created from a cluster matrix which is created from the linkage matrix.

Market states for all the epochs or time windows are shown in Figure 5.8. For the whole ten year period, seven clusters or market states, $\omega(t)$ = 1,2,....7, were picked up from the correlated clusters, these states are shown in Figure 5.6 and there are two big states and few small ones. Time series plots of these seven states are shown in Figure 5.7.

For simplicity, the seven states from the whole ten year data set are used to analyze the market dynamics under section 5.5 and section 5.6.

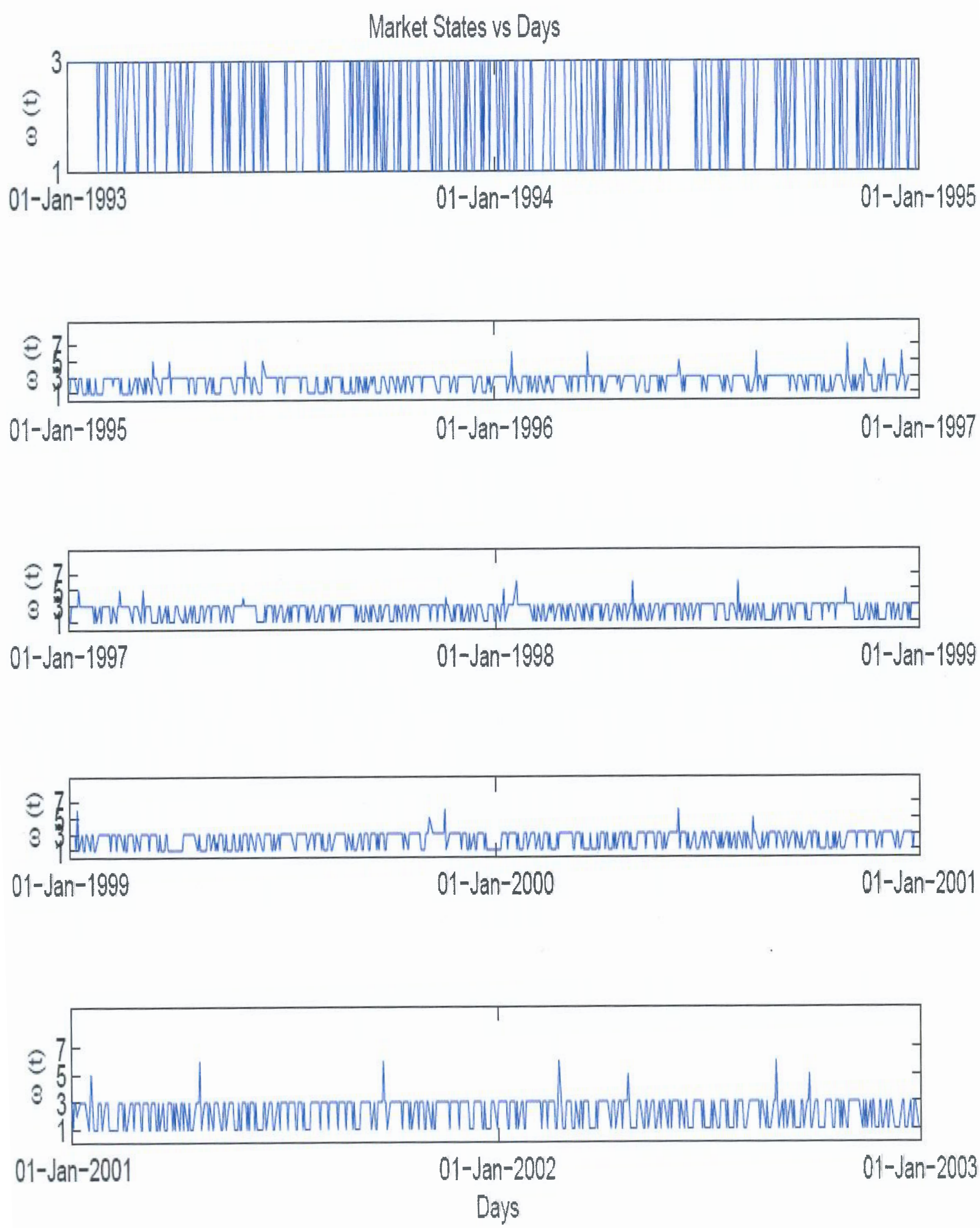


Figure 5.7: The time series of Market States over the entire data set. For two full years, during 01-January-1993 and the 01-January-1995, the market never experienced any other states above state 3, i.e. during these two years $\omega(t)$ = 1,2,3. States 1 and 3 in Figure 5.6 correspond to random market-wide activities.

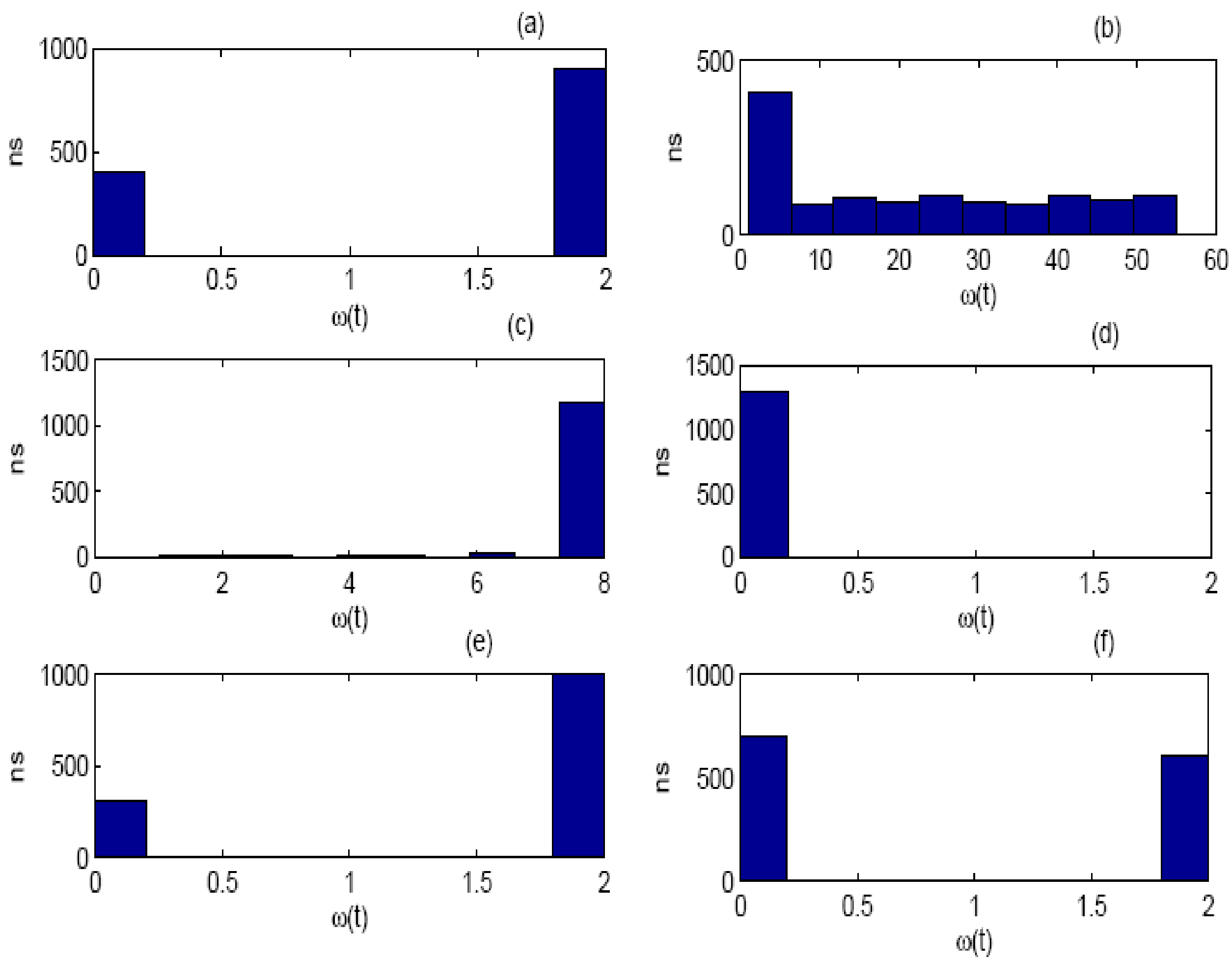


Figure 5.8: This figure shows histograms which show the states vs the number of days in each state plots. The histograms show market states during the following time period windows:
a) 1993-1997: two states were found. b) 1994-1998: consists of more than 10 states. c) 1995-1999: consists of 8 states. d) 1996-2000: every object or day belongs in one big group or state. e) 1997-2001: consists of 2 states. f) 1998-2002: consists of 2 states.

5.3.2 Simulated Annealing (SA) Results

When using the simulated annealing algorithm to solve a clustering problem, the final solution comes out as a single optimal solution. For this exercise, we applied the SA algorithm to the whole ten year data made up of 2736 days, and the outcome was a list of dates divided into 13 groups of correlated days or states. For each epoch, the simulated annealing algorithm resulted in to a different configuration of clusters with a different number of correlated clusters (states).

For the following explanations, we use the 13 states or groups of correlated days picked up by the model from the ten year data set to try to explain more characteristics about the market data.

The distribution of correlated cluster (states) sizes is shown in Figure 5.9. To pick up points on the diagram that correspond to the most frequent daily market patterns, one can fit a line with a negative slop on the points on the diagram to check any evidence of the power law. The distribution of the points or cluster sizes of correlated days along or close the fitted line would be $N(n_s) \sim n_s^{\Delta l}$, where Δl is the slope of the fitted line. In this diagram few points seem to follow the power law. This shows that most daily market patterns are random market activities or most daily market patterns in the South African market are the results of noise.

The plot of the correlation inside each cluster vs the cluster size is shown in Figure 5.10 and this diagram is drawn as an alternative method to discern patterns that are not the results of the noise. In this case one can fit a line with a positive slope through points that correspond to points on the line with a negative slope in Figure 5.9. Those few points that are along the line or those that are very close to the line are the most frequent patterns of the meaningful market patterns. Compared with the results in [8], in our results there are very few points along the line in Figure 5.9. This is because, in the South African financial markets there are very few highly traded shares. Hence, there aren't so many market activity patterns.

In conclusion, most of the market activities are results of the arrival of general random information into the market. From this we can conclude and say that, the South African equity markets are more risky than the markets of the data used in [8].

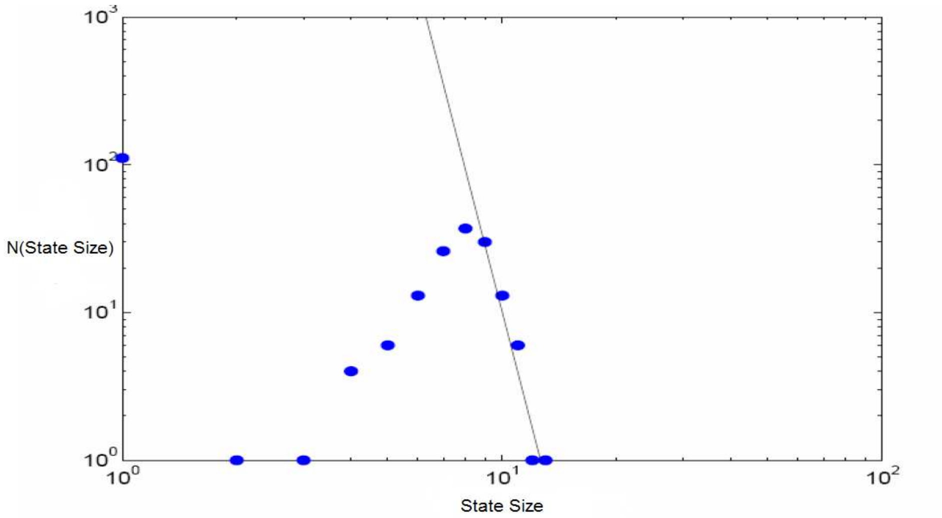


Figure 5.9: This graph shows the distribution of sizes of clusters of correlated days, i.e. of the frequency at which states occur.

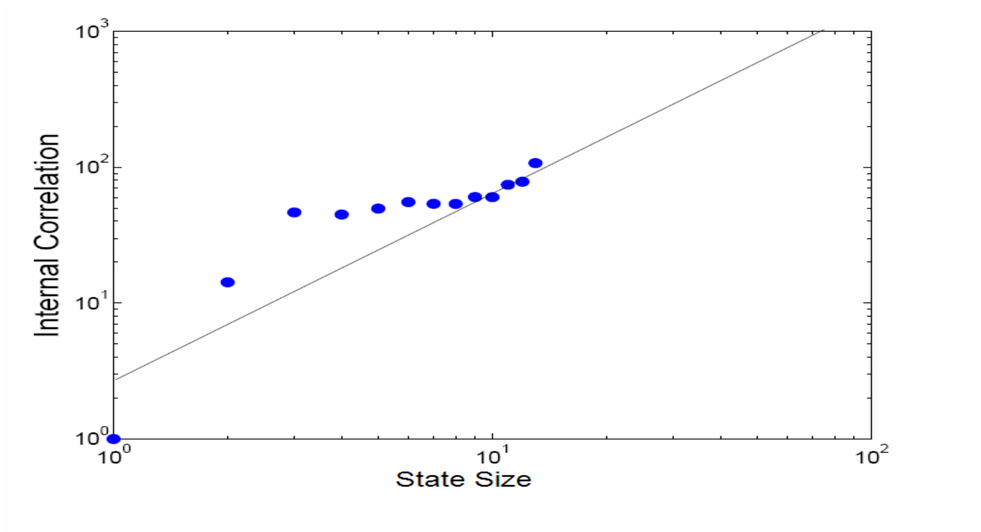


Figure 5.10: This graph shows the distribution of correlations c_s inside each cluster of correlated days. Most of the market patterns are the results of the random patterns. Points along the line or close to the line are the most meaningful market patterns.

5.4 The Effect of Data Sampling on Clustering

The main aim of this section is to show that the models work better with financial data with a lower sampling frequency and a smaller time-window. This reduces the random noise contribution, especially when clustering the market into market states. In this section the data used comes from the Johannesburg Stock Exchange All Share Index (JSE ALSI), which includes asset prices sampled *weekly* and *monthly*. The data set time-window runs from 01-Jan-2001 up to 01-Jan-2006.

Sector Clustering Results

As with the ten years of JSE daily data, we were able to identify market sectors. The underlying structure of correlated sectors comes out clearly for both weekly and monthly data. The dendrograms obtained from the *recursive merging algorithm* are shown in Figure 5.11.

The list of stocks in each correlated sector for both weekly and monthly market sectors, obtained from the *simulated annealing algorithm* are shown in Table 5.4 and Table 5.5, respectively.

Analysis of Sectors Results

The sectors in the five-year JSE-ALSI data are not so different from the sectors on the ten-year JSE data. Weekly cluster 1 is similar to monthly cluster 8 (Consumer Services), weekly cluster 6 is similar to monthly cluster 28 (All Financials), and weekly cluster 9 is similar to monthly cluster 18 (Mixed SIC sectors). The weekly results seem to have more noise than monthly result. There are more clusters in the weekly results.

The distribution of the sizes of economic sectors in Figure 5.12 seem to follow the power law, and this confirms our previous results with the ten-year JSE results.

States Clustering Results and Analysis

With the ten-year daily JSE stocks returns data, it was difficult to identify the market states. With the weekly and monthly five-year JSE returns of stocks this was not the case (see Figure 5.14. Looking at Figure 5.14, by inspection of the dendrograms, we estimate that for monthly data there are about 5 states and for the weekly sampled data there are about 17 states.

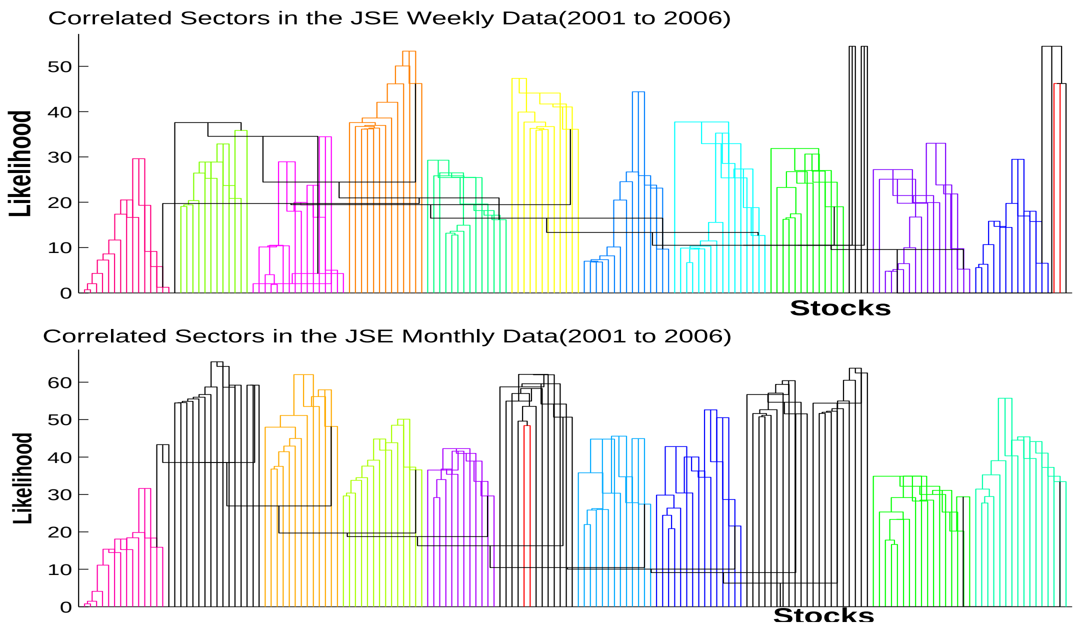


Figure 5.11: The dendrograms show the underlying structure of the correlated stocks of the JSE All-Shared Index, weekly and monthly data during the period 01-Jan-2001 to 01-Jan-2006.

Table 5.4: JSE-ALSI Correlated Market Sectors for weekly sampled data set during the period 2001-2006.

```
1.  'ECO'   'FOS'   'TRU'   'WHL'   'MPC'
2.  'TBS'   'DRD'
3.  'ANG'   'HAR'   'GFI'
4.  'BAT'   'VNF'
5.  'INP'   'INL'
6.  'RES'   'ACP'   'RDF'   'GRY'   'SYC'   'GRT'   'IFR'   'APA'   'EMI'   'PAP'
    'MTP'   'MPL'   'APB'   'VKE'   'HYP'   'CPL'   ;
7.  'DEL'
8.  'JNC'   'MTN'
9.  'SAB'   'OML'   'RCH'   'LBT'   'BIL'   'AGL'   'SOL'
10  'BVT'   'BAW'   'IPL'
11. 'PRA'
12. 'SUI'   'BCX'
13. 'ASA'   'SBK'   'SLM'
14. 'IMP'   'AMS'   'NHM'   'ARI'
15. 'NCL'   'WBO'   'TRT'
16. 'APK'
17. 'ILV'   'MET'   'TSX'
18  'PPC'
19. 'KGM'   'HDC'
20. 'WES'   'ALT'   'SHP'   'SPP'
21. 'OCT'   'SRL'
22. 'MTX'   'JCM'   'NPK'
23. 'FSP'
24. 'SHFF'
25. 'SIM'
26. 'NBKP'  'CDZ'
27. 'RMH'   'FSR'
28. 'MDC'
29. 'PIK'   'BEL'
30. 'AFE'
31. 'APN'
32. 'SPG'   'DAW'
33. 'LGL'   'LBH'
34. 'DDT'   'MRF'
35. 'AFB'
36. 'PMN'
37. 'DTC'
38. 'ATS'
39. 'WAR'
40. 'PSG'   'PTG'   'SHF'
41. 'RAH'   'RLO'   ;
42. 'MLA'
43. 'TNT'
44. 'BTG'
45. 'MVL'   'GRF'
46. 'AFR'   'CRM'
```

Table 5.4 continues....

47.	'SAP'	'CPI'	
48.	'AEG'		
49.	'MVG'	'NTC'	
50.	'REM'	'GND'	
51.	'CSL'		
52.	'HVL'		
53.	'SPE'	'MUR'	
54.	'JDG'	'ELH'	
55.	'OCE'		
56.	'ADR'		
57.	'MSM'		
58.	'CLH'		
59.	'ART'	'TIW'	'AMA'
60.	'AFO'	'NED'	
61.	'UTR'		
62.	'BRC'	'CSB'	
63.	'SUR'	'MKL'	'OMN'
64.	'ILA'		
65.	'ADH'	'NPN'	
66.	'CMH'		
67.	'ARL'		
68.	'NWL'		
69.	'KMB'	'ENL'	
70.	'TKG'	'FBR'	
71.	'GDF'	'MST'	
72.	'ATN'	'ATNP'	
73.	'LEW'	'SNT'	'RBW'
74.	'INPR'		
75.	'DSY'		
76.	'KAP'		
77.	'PMM'	'BDE'	
78.	'TRE'	'MBN'	
79.	'AFX'		
80.	'PGR'	'ABL'	
81.	'SCN'		
82.	'AVI'		

Table 5.5: Correlated Market Sectors for Monthly sampled data set from the JSE-ALSI during the period 2001-2006.

1. 'CSL' 'VKE' 'LEW' 'SPP'
2. 'OCT' 'PMM'
3. 'CDZ' 'KAP'
4. 'AFB' 'SHF'
5. 'ILA' 'CSB' 'TRT' 'MSM'
6. 'PTG' 'NBKP'
7. 'OML' 'RCH' 'DDT' 'DTC'
8. 'TRU' 'ECO' 'FOS' 'NCL' 'MPC' 'APK' 'WHL'
9. 'SUI' 'CMH' 'UTR' 'JCM' 'PMN'
10. 'HAR' 'GFI' 'ANG'
11. 'HVL' 'BDE'
12. 'NPK' 'NPN' 'PIK'
13. 'ART' 'DAW' 'PPC'
14. 'AFR' 'GRF' 'MST'
15. 'ARL' 'RBW'
16. 'BAW' 'BVT' 'IPL' 'TBS'
17. 'SHFF' 'INPR'
18. 'LBT' 'BIL' 'SOL' 'AGL' 'IMP' 'AMS' 'SAP' 'SAB' 'MLA'
    'TNT' 'NHM' 'ARI'
19. 'JDG' 'ELH'
20. 'MRF' 'WAR' 'KMB' 'DRD' 'ILV' 'SIM'
21. 'AFO' 'SUR'
22. 'MUR' 'AEG'
23. 'ATNP' 'ATN'
24. 'PGR' 'RLO' 'AFE' ;
25. 'PRA' ;
26. 'MVG'
27. 'BTG' 'ASA' 'MET' 'SLM' 'CPI' 'ABL' 'NED'
28 'APB' 'SYC' 'CPL' 'APA' 'RDF' 'MTP' 'PAP' 'MPL' 'IFR'
    'RES' 'HYP' 'GRY' 'GRT' 'EMI' 'ACP' 'SPE'
29. 'OMN' 'BEL'
30. 'TIW' 'CRM' 'AVI' ;
31. 'SPG'
32. 'AMA' 'BCX'
33. 'DSY' 'CLH'
34. 'OCE' 'BRC'
35. 'ADH' 'RAH'
36. 'RMH' 'FSR' 'SBK'
37. 'PSG' 'REM' ;
38. 'SNT' ;
39. 'TSX'
40. 'KGM' 'GDF' ;
41. 'TKG'
42. 'LBH' 'LGL'
43. 'DEL' 'MTX' 'VNF'
44. 'JNC' 'MTN'
45. 'APN' 'SHP' 'ADR'

```
Table 5.5 continues....

46. 'TRE'     'MBN'
47. 'FBR'     'SCN'     'SRL'
48. 'FSP'     'WBO'
49. 'ALT'     'MDC'
50. 'WES'     'NTC'     'NWL'   ;
51. 'BAT'
52. 'AFX'     'MKL'     ;
53. 'ATS'
54. 'GND'     'HDC'     'MVL'
55. 'INP'     'INL'     ;
56. 'ENL'
```

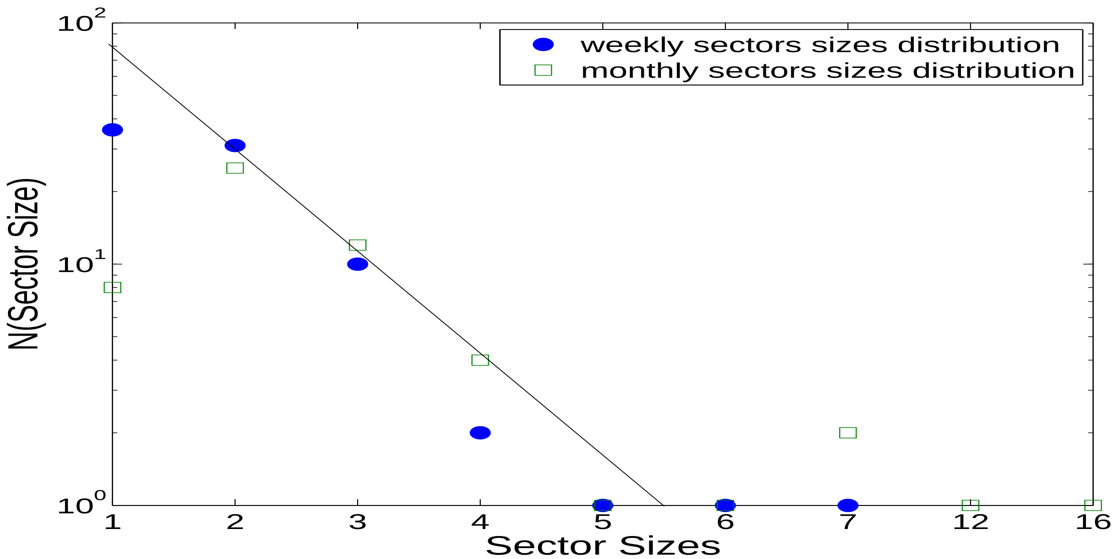


Figure 5.12: This graph shows the distribution of the sizes of economic sectors found in the All-Share Index during the period 2001-2006.

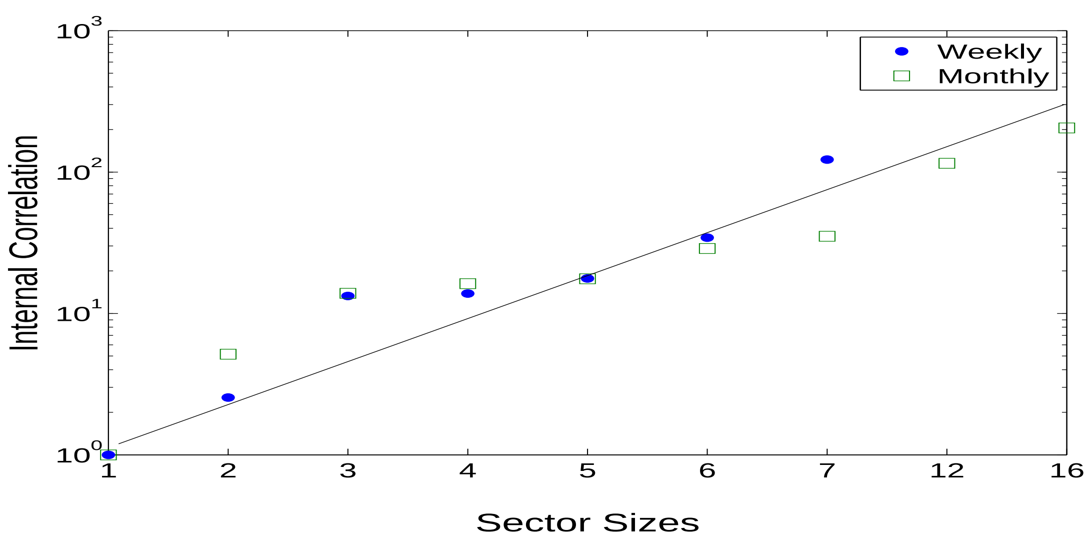


Figure 5.13: This graph shows the distributions of the correlation, c_s, inside the market sectors of the ALSI during the period 2001-2006.

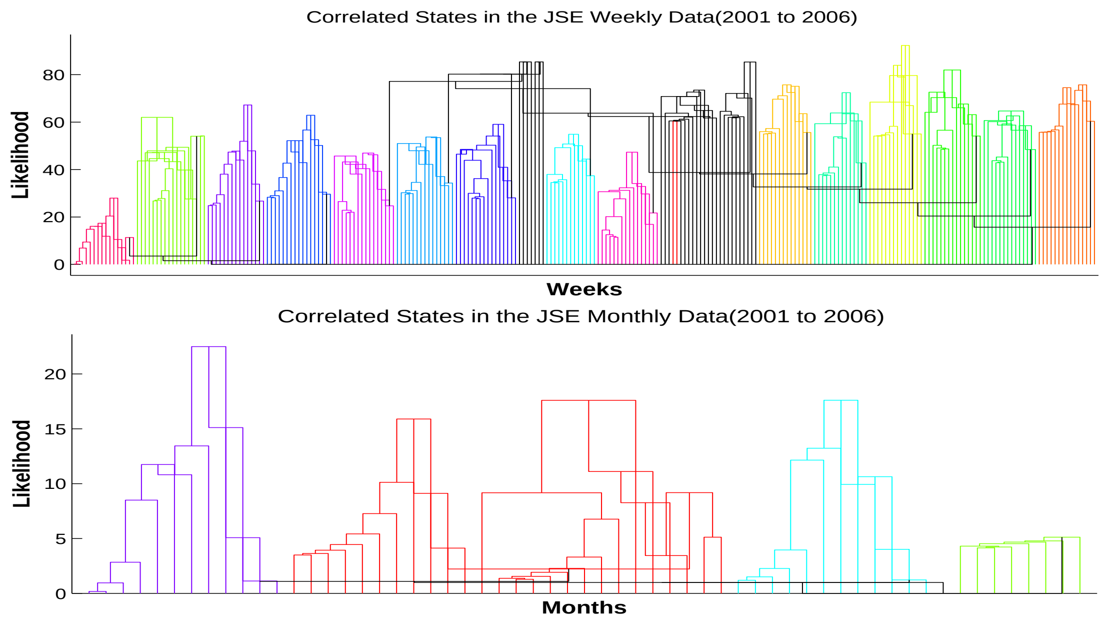


Figure 5.14: The dendrograms shows the underlying structure of the correlated stocks of the JSE All-Shared Index, weekly and monthly data during the period 01-Jan-2001 to 01-Jan-2006.

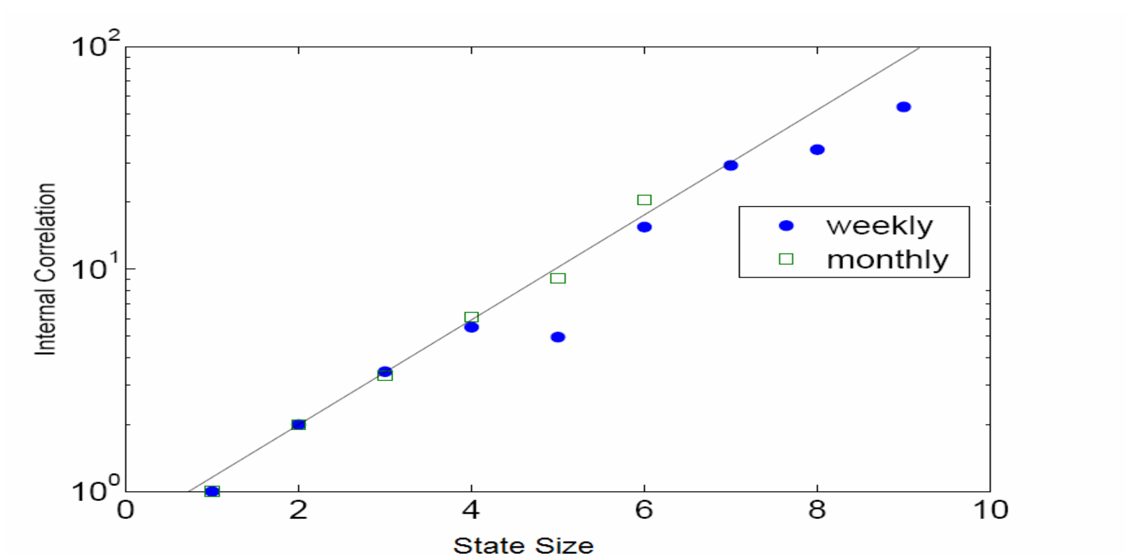


Figure 5.15: This graph shows the distributions of the correlation, c_s, inside the market states of the ALSI (weekly and monthly states during 2001-2006).

5.5 Implications for prediction

We can now analyze the market performance in different states. For simplicity, in this exercise we use 7 states which were picked up in section 5.3.1, using the MA algorithm. Figure 5.16 shows the average daily returns of different assets in different states from the ten year market data. Few plots are shown in this diagram. In Figure 5.16, each point has coordinates $(\langle r_i|\omega\rangle, \langle r_i|\omega'\rangle)$, where ω and ω' can take values between 1 and 7 on the axes.

Most of the plots or diagrams that didn't show a certain pattern are not shown on the picture. In some states, assets of the same market sector do not behavior similarly. The market behavior in state 1 vs state 2, state 3 vs state 4 and state 3 vs state 6, exhibits a negative correlation in each case. This means assets that have positive returns in state 1 have negative returns state 2.

Figure 5.16 also shows that there might be a linear relation between state 1 & 2, state 3 & 4, and state 3 & 6. This tells us that markets are not perfectly efficient, they can be predicted. In state 1, most stocks in the financial market sector have negative returns, but positive returns in state 2. In state 7, the technology sector performed very poorly, the returns are very small compared to other market sectors. This pattern is also followed by the basic industries. This could spell correlation between the technological and the basic industries. In states 4 and 5, the market seems to show little performance, all industries have returns which are closer to zero.

In almost all the states, one or two venture capital assets and some non-cyclical consumer goods seem to be performing in a similar behavior, in every graph there's at least one black 'x' and one blue triangle are in a similar position.

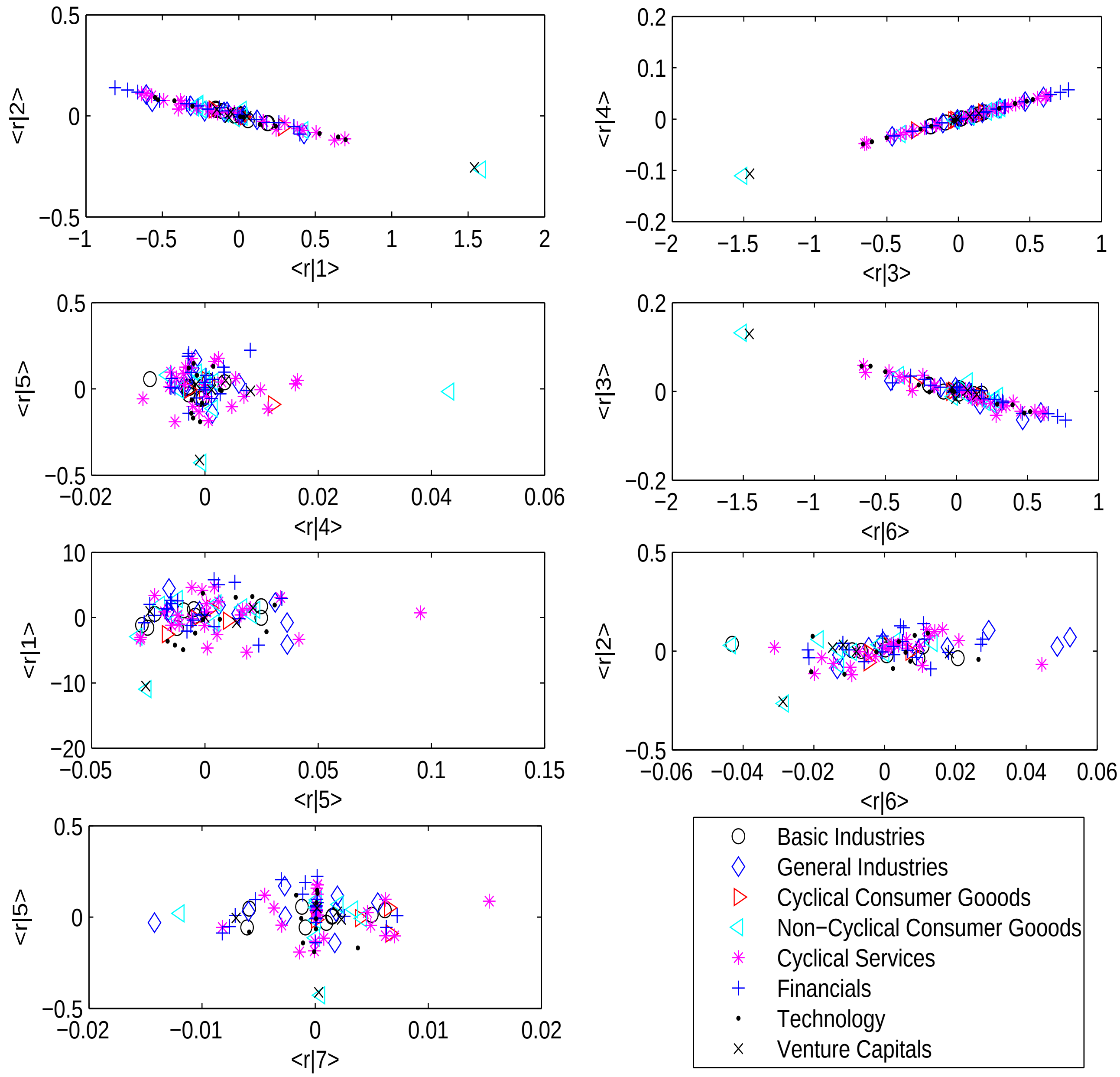


Figure 5.16: This figure shows the performance of the market in different states. Each point or coordinate $(\langle r_i|\omega\rangle, \langle r_i|\omega'\rangle)$ corresponds to a single asset i in the market in states ω and ω'. Assets from the same economic group are plotted with the same symbol.

5.6 Market Predictability and Efficiency

Minimization of risk is a difficult task. This is due to the fact that no model works 100% perfectly. Aside guessing, every risk management strategy is based on some probability when predicting the market returns or dynamics. When the market dynamics are classified in terms of sectors and states, it is possible to look at the market dynamics as a sequence of states $\omega(t)$ in different days $t = 1, 2, ..., T$.

From the previous section, there were signs of some relationships between certain market states, this suggests that the market might be predictable. Since we have seven states, we can find the probability of moving from one state to another, using the formula in equation 5.1. This equation was also used by Marsili *et al.* [8]. This is the probability of the transition from ω state to state ω'. With seven states, the transition probability matrix is a 7 x 7 matrix.

$$\mathbb{P}_1(\omega'|\omega) = \sum_{t=1}^{T-1} \delta_{\omega(t),\omega}\delta_{\omega(t+1),\omega'} \Big/ \sum_{t=1}^{T-1} \delta_{\omega(t),\omega} \tag{5.1}$$

Both the classification in states and in transition probability matrix are not consistent or stable across different epochs. The small states found in Figure 5.6 die out in each epoch. All the epochs end up with two large states or one very lager state and a small one. Those large states correspond to noise contributions. This means that, with the kind of data being used in this work (daily asset prices), the model can not pick up the market states because of too much noise. A different data set with asset prices sampled at different frequencies is used and it is discussed in section 5.4. In this paper, for the market predictability and efficiency analysis, for simplicity we use the states classification of the whole data set with a ten year time window from 1993 to 2003.

If we want to predict the market returns in the future, we can only use the information we have today. Since we have the transition probability matrix, we can estimate the expected return of asset i tomorrow given today's state ω. The equation below was used [8]:

$$\langle r_i(t+1)|\omega(t)\rangle = \sum_{\omega'} \langle r_i(t+1)|\omega(t+1)\rangle \mathbb{P}_1(\omega'|\omega). \tag{5.2}$$

Using the expected return of asset i tomorrow given today's state ω, we can calculate the measure of predictability using equation 5.3 below as [1]

$$H_i(t'|t) = \sqrt{\sum_{\omega} \rho_\omega \frac{\langle \delta r_i(t')|\omega(t) = \omega\rangle^2}{\langle \delta r_i^2|\omega\rangle}}, \tag{5.3}$$

where $\delta r_i(t) = r_i(t) - \langle r_i\rangle$ are the deviations of the returns from the mean, and ρ_ω is the frequency at which state ω occurs. The distribution of H_i across assets is shown in Figure 5.17. The diagram also shows the probability distribution of a random noise and the probability distribution of a Markov sequence. This diagram tells us that markets do have a certain level of inefficiency. It can be seen from the graph that the distribution of predictability H_i can be explained by a Markov process more than it can be explained by noise.

[1]By M. Marsili, D. Challet, J. Berg et al [8, 41, 42, 43, 44].

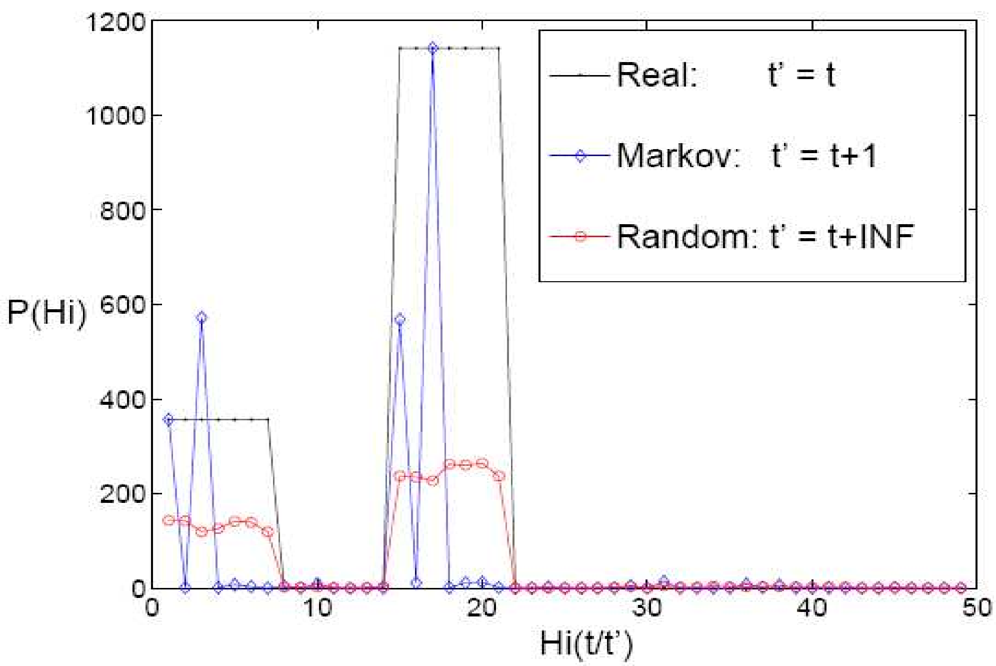


Figure 5.17: This figure shows the Distribution of the Predictability of the Market, $H_i(t'|t)$. The distribution of predictability H_i can be explained better by a discrete Markov process.

Chapter 6

Conclusions

Our results show that, because of company diversification, clusters of stocks go beyond grouping by related industries. In these groups of stocks, only a few groups are similar to the traditional SIC classifications (mining, financial and resources). The grouping is done only for positively correlated objects. Negatively correlated object might result into noise. The correlated sectors from the model results have small overlapping with the standard defined sectors of the economic activities. During the ten year period studied in this work, the economy of South Africa was heavily dominated by the mining sector and the volatility of the currency. All the stocks which are positively correlated with these two economic factors are in bigger and distinct clusters, e.g Financial and Resources.

In our results, most of the stocks are in small or singleton clusters and they don't show any consistent cluster pattern. Stocks which have larger betas in the South African market were found in these random clusters. This means these stocks are more sensitive to the market stimuli, and they are not highly positively correlated to individual stocks in the market. This could also mean that they are negatively correlated to most of the traded stocks in the market. When the market mode is removed, these stocks behave like noise when uncovering the underlying internal structure of correlated stocks. Those distinct correlated clusters which were identified, i.e. economic sectors like mining and banks, showed stability through all epochs. Most of the small clusters are stable for a short period of time.

Uncovering market states helps us to quantify the market predictability. Figure 5.17 tells us that that markets do have a certain level of inefficiency.

Compared to [38] and [39], we have used a different approach to solve the problem of quantifying collective behavior of stock price movements or the cross-correlations between price movements of different stocks. Mapping financial data to a physical system of interacting objects provides a good framework for studying the market microstructure origins of cross-correlations between stocks traded in a market. It allows us to discover and interpret the occurrence of the power-law distribution in the financial time series of highly correlated eigenmodes.

Noise reduction in data can also be achieved by simply reducing the sampling frequency of asset prices. A different data set with different sampling frequencies was used in section 5.4, where more definitive results were obtained.

Appendix A

Sectors and States Toolbox Help Files: Recursive Merging Algorithm

NB: Only help files are presented on this paper, to see full codes that we have developed the reader must visit the website: http://www.financialcybernetics.co.za/downloads.html

A.1 Contents Script

```
% Sector and States Toolbox
% Version 1.0 (R14) 01-Sep-2005
%
% Copyright (C) 2004  Bongani Mbambiso, Tim Gebbie, Diane Wilcox, University of Cape Town
%
%    This program is free software; you can redistribute it and/or modify
%    it under the terms of the GNU General Public License as published by
%    the Free Software Foundation; either version 2 of the License, or
%    (at your option) any later version.
%
%    This program is distributed in the hope that it will be useful,
%    but WITHOUT ANY WARRANTY; without even the implied warranty of
%    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
%    GNU General Public License for more details.
%
%    You should have received a copy of the GNU General Public License
%    along with this program; if not, write to the Free Software
%    Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA  02111-1307
%    USA
%
% General information.
%   Comprises functions to generate clusters of objects using the
%   covariance matrix and a utility function. Clusters are defined by the
%   their linkaged structure. There are two approaches followed here based
%   on Giada & Marsili 2002:
%
%       1. Recursive Merging:
```

```
%           1.1. Merge 2 clusters with maximal likelihood
%           1.2. Repeat 1.1 until there is only one cluster
%
%       2. Simulated annealing:
%           2.1 Sweep the lattice (switch configurations of pairs) using metropolis
%           2.2 Merge 2 clusters on the lattice using metropolis conditions
%           2.3 Split a cluster on the lattice into 2 clusters using metropolis.
%           2.4 Repeat 2.1, 2.2 and 2.3 until likelihood is maximized.
%
%   See the matlab LINKAGE, CLUSTER and DENDROGRAM functions for the
%   conventions used here. The linkage structure Z is output of the
%   functions provided here based on the input covariance matrix or
%   correlation matrix. To compute the covariance or correlation matrix use
%   the functions provided in covariance toolbox.
%
% Also Uses: COVTOOLS/AVERAGE, COVTOOLS/SPEARMAN, COVTOOLS/KENDALL
%
% General Functions:
%   energy         - internal energy of a cluster based on number of
%                    elements and the internal correlation.
%   likelihood     - Likelihood of the configuration based on the number of
%                    clusters and their internal correlations.
%   determine      - (rm) Clusters by deterministic recursive cluster
%                    merging (principle).
%   parameters     - (rm) Find new clusters parameters from configuration
%                    and correlations.
%   minimal        - (rm) Find the next deterministic minimal combination
%                    of clusters.
%   annealing      - (sa) Find clusters by simulated annealing.
%   configuration  - (sa) Compute the initial configuration structure to be
%                    used with annealing.
%   change         - (sa) Change configuration of lattice using: sweep,
%                    merge and split.
%   cfsweep        - (sa) Randomly swap pairs of elements between pairs
%                    of clusters.
%   cfmerge        - (sa) Randomly merge 2 clusters that have at least two
%                    elements.
%   cfsplit        - (sa) Split a cluster, with at atleast 4 elements,
%                    according to the correlation distances of the elments.
%                    variables for a new configuration structure.
%   ground         - (sa) Reset the ground state variables of the
%                    configuration structure.
%   AnalysisOutputs- (sa) Calculate and output analysis variables.
%
% Test Code
%   sectors_test_determine_001.m - deterministic merging with small
%                                  synthetic data sets.
%   sectors_test_determine_002.m - deterministic merging with large
%                                  synthetic data sets.
%   sectors_test_annealing_001.m - simulated annealing with small
```

```
%                                  synthetic data sets.
%   sectors_test_annealing_002.m - simulated annealing with large
%                                  synthetic data sets.
%
% Obsolete functions.
%   None at this time.
%
% Future Development
%   sectors_test_determine_003.m - deterministic merging J203 data sectors
%   sectors_test_determine_004.m - deterministic merging J203 data states
%   sectors_test_annealing_003.m - simulated annealing J203 data sectors
%   sectors_test_annealing_004.m - simulated annealing J203 data states
%
%
% Others
%   (every graph in paper Marsili 2002)
%
% GUI Utilities.
%   None at this time.

% Copyright 2004, Tim Gebbie, (Futuregrowth Asset Management), Diane Wilcox
% (University of Cape Town)
%
% Terms and Conditions of Use
%
% THIS SOFTWARE IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER
% EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED
% WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.
% THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS
% WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF
% ALL NECESSARY SERVICING, REPAIR OR CORRECTION.
%
% IN NO EVENT WILL THE COPYRIGHT HOLDERS OR THEIR EMPLOYERS, OR ANY OTHER
% PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THIS SOFTWARE, BE LIABLE TO
% YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR
% CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE
% PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING
% RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A
% FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF
% SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH
% DAMAGES.

% $Revision: 1.11 $ $Date: 2006-05-23 10:35:18+02 $ $Author: bmbambiso $
```

A.2 MA Driver:

```
function [Z] = determine(varargin)
% DETERMINE Clusters by deterministic recursive cluster merging
%
```

```
% [Z] = DETERMINE(X) Compute the linkage for the data X, and MxN
% matrix with M objects and N realizations. The data is first averaged
% using AVERERAGING which removes the bias from the data i.e. the
% market mode is removed.
%
% Cluster information will be returned in the matrix Z with size m-1
% by 3, where m is the number of observations in the original data.
% Column 1 and 2 of Z contain cluster indices linked in pairs
% to form a binary tree. The leaf nodes are numbered from 1 to
% m. They are the singleton clusters from which all higher clusters
% are built. Each newly-formed cluster, corresponding to Z(i,:), is
% assigned the index m+i, where m is the total number of initial
% leaves. Z(i,1:2) contains the indices of the two component
% clusters which form cluster m+i. There are m-1 higher clusters
% which correspond to the interior nodes of the output clustering
% tree. Z(i,3) contains the corresponding linkage distances between
% the two clusters which are merged in Z(i,:), e.g. if there are
% total of 30 initial nodes, and at step 12, cluster 5 and cluster 7
% are combined and their distance at this time is 1.5, then row 12
% of Z will be (5,7,1.5). The newly formed cluster will have an
% index 12+30=42. If cluster 42 shows up in a latter row, that means
% this newly formed cluster is being combined again into some bigger
% cluster.
%
% [Z] = DETERMINE(X,METHOD) Compute the linkage for the data X, and MxN
% matrix with M objects and N realizations. The correlation matrix isdata
% is first averaged computed using the method in METHOD. METHOD can take
% on values 1, 2 or 3, which imply 'pearson', 'kendall' or 'spearman',
% respectively. The data is first averaged using AVERERAGING which removes
% the bias from the data i.e. the market mode is removed.
%
% Example 1:
%       c = cov(randn(10,1000));
%       Z = determine(c);
%
% Example 2:
%       x = randn(10,1000);
%       x = average(x,5);
%       Z1 = determine(x,1);  %for 'pearson'
%       Z2 = determine(x,2);  %for 'kendall'
%       Z3 = determine(x,3);  %for 'spearman'
%
% Note 1: To plot the dendrogram if the tree is simply connected use
% see DENDROGRAM(Z,0).
%
% Note 2: The algorithm iterates MINIMAL until convergence see
% MINIMAL and PARAMETERS to find out which data structures are iterated.
% Based on the method by L. Giada & M. Marsili (2005).
%
% See Also : ANNEALING, PEARSON, KENDALL, SPEARMAN, LIKELIHOOD, MINIMAL,
```

```
%            DENDROGRAM, AVERAGING

% Authors: Bongani Mbambiso, Tim Gebbie
```

A.3 Parameters:

```
function [s, cs, ns, index] = parameters(a, c)
% PARAMETERS Find new cluster parameters of a new configuration.
%
% [S,CS,NS,INDEX] = PARAMETERS(A, C) From the previous configuration’s
% index vector A and correlation matrix C, computes S, the configuration
% logical cluster matrix with columns as clusters and rows as objects.
% ’1’ (True) indicates that an object is in that cluster, false or ’0’ if
% it is not.
%
% Also computes CS, the pearson coefficient for the new configuration of
% clusters. NS is the umber of objects per cluster in the new
% configuration. The cluster number is given by the position in INDEX.
%
% NOTE 1: The index matrix, I(i,k)=j, j-th object in the original data is
% the k-th object in the in i-th cluster of the current configuration.
%
% NOTE 2: j - is indexing original data objects, i - is indexing clusters,
% and k - objects of the i-th / currrent cluster. This gives an index of
% all the k objects in i-th cluster. Rows (i’s) are clusters and values
% are elements in the cluster. This give membership.
%
% See Also : UNIQUE, REPMAT, SPARSE, FULL.

% Authors: Bongani Mbambiso, Tim Gebbie
```

A.4 Energy:

```
function [e,gs] = energy(ns,cs)
% ENERGY Calculates the energy of each configuration
%
% [E, GS] = ENERGY(NS,CS) Compute E the energy of each cluster or
% configuration of clusters, and the optimal factor coefficient. The
% inputs NS and CS are, the number of objects in each cluster S of the
% configuration, and the pearson coefficient for the cluster of objects,
% respectively.
%
% NOTE: The method is based on the method by L. Giada & M. Marsili (2005).

% Authors: Bongani Mbambiso, Tim Gebbie
```

A.5 Minimal:

```
function [e0,c0,n0,ij,de] = minimal(e,c,n)
% MINIMAL Find the next deterministic minimal combination of clusters
%
% [E0, C0, N0, IJ, DE] = MINIMAL(E,C,N) This function computes, E0 the
% energy of the new configuration, C0 the new pearson coefficient, N0 the
% new number of objects in a given cluster, IJ the merged and minimal
% cluster index, and DE the change in energy of the configuration.
% I is the cluster configuration index. DE is the change in energy of the
% configuration. C is the new pearson coefficient. N is the new number of
% objects in a given cluster.
%
% NOTES: In IJ, I and J are the merged clusters, the j-th cluster is eliminated
% from the set of clusters after being merged with the i-th cluster,
% reducing the previous number of cluster by 1. Brute force search.
%
% See Also : ENERGY

% Authors: Bongani Mbambiso, Tim Gebbie
```

A.6 Likelihood:

```
function lc = likelihood(ns,cs)
% LIKELIHOOD The log-likelihood of the configuration.
%
% LC = LIKELIHOOD(NS,CS) This function uses NS, the number of
% objects per cluster, and CS, the internal correlations of clusters
%
% NOTE: The functions sums up all internal energies in a given
% configuration algorithm (see L. Giada & M. Marsili (2005)).

% Authors: Bongani Mbambiso, Tim Gebbie
```

Appendix B

Sectors and States Toolbox Help Files: Simulated Annealing Algorithm

B.1 SA Driver:

This code is the main function that drives the simulated annealing process.

```
function [A] = annealing(varargin)
% ANNEALING Find the maximum likelihood cluster configuration
%
% [A] = ANNEALING(VARARGIN) Find the Maximum likelihood configuration,
% using the Metropolis Scheme or Simulated annealing method and
% the J.D. Noh's Model. The algorithm finds the ground state,which is the
% best configuration at the end of the algorithm simulation.
%
% [A] = ANNEALING(DATA,IB,FB,T_STEPS,N_CYCLES,CF,CORR_M )  Uses the data
% to compute the correlation matrix, which is used to find the maximum
% likelihood configuration of the data. CORR_M specifies the method used
% to compute correlation matrix. It takes values 1, 2, and 3, which
% represent 'Pearson', 'Kendal' and 'Spearman', respectively. All the other
% variables are annealing variables and they are described below.
%
% [A] = ANNEALING(C,IB,FB,T_STEPS,N_CYCLES,CF)  Takes the correlation
% matrix, and use it to find the maximum likelihood configuration of
% the data. Other variables are Annealing input varaibles (see below).
%
% A is the configuration data structure and has structure described below.
% A(i)=j gives that the i-th object is in the j-th cluster.
% The configuration index I, where I(i,j)=k gives k object as the
% j-th element of the i-th cluster. E(i) is the likelihood
% per element of the i-th cluster. C is the similarity the correlation
% matrix. Here it can be taken as the Pearson matrix or Kendal's
% Coefficient. This is taken to be a square matrix of the size of the
% number of objects.
%
% ANNEALING VARIABLES:
```

```
% IB       - initial beta
% FB       - final beta
% T_STEPS  - Steps of Temperature changes
% CF       - factor for change in temperature
%
%
% DATA STRUTURE OF THE CONFIGURATION [A]
%
% Status of the system during the annealing process:
% These variables keep track of the current solution of the maximum
% likelihood problem.
% A.N        - Original data size
% A.a        - the current solution configuration
% A.nc       - number of (non-empty) clusters in the current configuration
% A.nec      - Index array of Non-empty clusters
% A.b        - current inverse temperature (beta).
% A.c        - internal correlations
% A.C        - Correlation matrix
% A.e        - energy (minus log of likelihood), which is being minimized.
% A.n        - number of elements per cluster
% A.I        - configuration index
% A.t        - time, counts the number of steps.
% A.cf       - cooling factor for delta_temperature
% A.updates - no. of spin updates per sweep
% A.merges  - number of merges per sweep
% A.splits  - number of splits per sweep
% A.t_steps - number of sweeps between changes in temperature
% A.cycle   - number of annealing/ temperature clycles
% A.n_cycles- maximum number of annealing/ temperature clycles
%
% Ground state parameters:
% These paramters represent the final solution, the best configuration that
% gives the best maximum likelihood.
% A.gs.a       - the best configuration
% A.gs.nc      - number of clusters in the final solution configuration
% A.gs.nec     - Index array of Non-empty clusters
% A.gs.b       - Ground state energy (per spin).
% A.gs.c       - internal correlations
% A.gs.C       - Correlation matrix
% A.gs.e       - energy (minus log of likelihood),which is being minimized.
% A.gs.I       - configuration index
% A.gs.n       - number of ground state elements per cluster cluster
% A.gs.t       - time, counts the number of steps.
% A.gs.updates - no. of spin updates, per spin, per sweep, or
%                the no. of times a better configuration was found in
%                the same temperature or cooling schedule
% A.gs.merges  - number of mergings, per sweep
% A.gs.splits  - number of splitting, per sweep
% A.gs.cycle   - number of annealing / temperature cycles
%
```

```
%
% See Also : PEARSON, KENDALL, SPEARMAN, LIKELIHOOD, AVERAGING
%            CHANGE, ENERGY.
%
% References:
%
% 1) J. D. Noh  (2000)
% 2) Marsili M. (2002)
% 3) L. Giada & Marsili M (2005)
% 4) Metropolis, N., Rosenbluth. A. W. (1953)
%

% Authors: Bongani Mbambiso, Tim Gebbie
```

B.2 Configuration:

```
function [A] = configuration(varargin)
% [A] = CONFIGURATION(VARARGIN)Initialize the configuration matrix.
%
% [A] = CONFIGURATION(C,A, SWEEP_TYPE) This function intializes the
% configuration of clusters in the data, Randomly, from the similarity or
% correlation matrix, C, of the data. SWEEP_TYPE takes values 'random' or
% 'sequential'. Its default value is
%
% [A] = CONFIGURATION(C, A) The default value for SWEEP_TYPE is
% 'random'. [A] is the configuration of clusters, and it is data
% structure with a format described ANNEALING (See Example below).
%
%
% Example 1:
%       A.C = [[1  0  0];
%              [0  1  0];
%              [0  0  1]];
%
% then, initial configuration paramters are:-
% A.a       = [1,2,3];  - configuration
% A.c       = [1,1,1];  - internal correlations
% A.n       = [1,1,1];      - no. of objects per cluster
% A.I       = [1,2,3]   - cluster index
% A.e       = [0,0,0];  - internal energies
% A.updates = 0         - no. of spin updates
% A.b       = ib        - inverse temperature = initial beta
% A.splits  = 0         - number of splits,
%
%
% Example 2:
%     A.I = [1  8 5 9 0 0 0];
%           [0 0 0 0 0 0 0 ];
%           [6  7 0 0 0 0 0];
```

```
%           [0 0 0 0 0 0 0 ];
%           [4  2 0 0 0 0 0];
%           [0 0 0 0 0 0 0 ];
%           [0 0 0 0 0 0 0 ];
%           [3 10 0 0 0 0 0];
%           [0 0 0 0 0 0 0 ];
%           [0 0 0 0 0 0 0 ];
%
%   A.a   = [1 5 8 5 1 3 3 1 1 8];
%   A.a(5)= 1, means that the 5-th object is in the 1st cluster.
%
%   A.nc  = 4; There are 4 non-empty clusters in the configuration
%
%   A.nec = [1 3 5 8]; An array index of non-empty clusters

% Author:  Bongani Mbambiso, Tim Gebbie
```

B.3 Change:

```
function A = change(A)
% CHANGE Find a new configuration.
%
% [A] = CHANGE(A) Change the configuration A by cycling through the current
% cluster configuration, with a constant temperature, find a new
% configuration A, with a better clustering solution.
%
% This function rearranges a lattice configuration A, and finds a
% configuration which increased energy, at a constant temperation or
% cooling schedule. Configuration
% membership index, INDEX, initial internal energies, E, at
% a given inverse temperature BETA. NR is the number of
% configuration changes. At time T.
%
% First a random lattice configuration change is attempted
% by swapping two elements between different clusters via a
% lattice sweep. A random cluster merge is attempted. A
% random cluster split is attempted.
%
% See Also: CFSWEEP, CFMERGE, CFSPLIT

% Author: Bongani Mbambiso, Tim Gebbie
```

B.4 Sweep:

```
function A = cfsweep(A, sweep_type)
% CFSWEEP Randomly or sequential sweep through the lattice.
%
% [A] = CFSWEEP(A, SWEEP_TYPE) This function sweeps through the lattice by
% randomly swapping elements or spins between pairs of clusters, check if
```

```
% that new pair of clusters improves the energy of the system i.e dE <=0.
% If the formation of a new pair improves the system energy, the new
% configuration is kept as a  new solution configuration. If dE > 0, it
% takes a solution with a probability p = exp(-dE/bT)
% (Boltzmann probability) as new solution. b - is the inverse temperature.
% In this way, the function tries to find a new configuration with a better
% energy level, at the same constant temperature (Metropolis scheme), then
% then recomputes the configuration variables.
%
% [A] is a new configuration with changed pairs of spins or elements.
% This  function also counts, with A.UPDATES, how many spin interactions
% were broken or how many new configurations resulted in an improvement in
% the energy of the entire system.
%
% See Also: ENERGY

% Author: Bongani Mbambiso, Tim Gebbie
```

B.5 Merge:

```
function A = cfmerge(A)
% CFMERGE Two clusters are merged.
%
% [A] = CFMERGE(A) Merge two randomly chosen clusters that have at least 2
% elements. Checks if energy permits configuration change, and then
% recomputes or updates the configuration variables. It also updates the
% ground state variables, only if the merge improves the energy of the
% configuration.
%
% [A] is a new configuration of clusters, with two previous clusters joint
% to form a new cluster. This means that, the previous number of clusters
% drops by 1, i.e A.n_new = A.n -1. (See ANNEALING for the data
% structure format of the configuration A).
%
% See Also: ENERGY,

% Author: Bongani Mbambiso, Tim Gebbie
```

B.6 Split:

```
function A = cfsplit(A)
% CFSPLIT Split a randomly chosen cluster from configuration A.
%
% [A] = CFSPLIT(A) Split a cluster with atleast four elements into two
% cluster of  different nubmer of elements, depending on the correlation
% distances between the elements in that.

% If there is not cluster with four elements in the configuration,
```

```
% the configuration is left as it is, no splitting will occur. If the
% splitting does not improve the temperature of the system, the
% configuration remains the same.
%
% See Also: ENERGY,

% Author: Bongani Mbambiso, Tim Gebbie
```

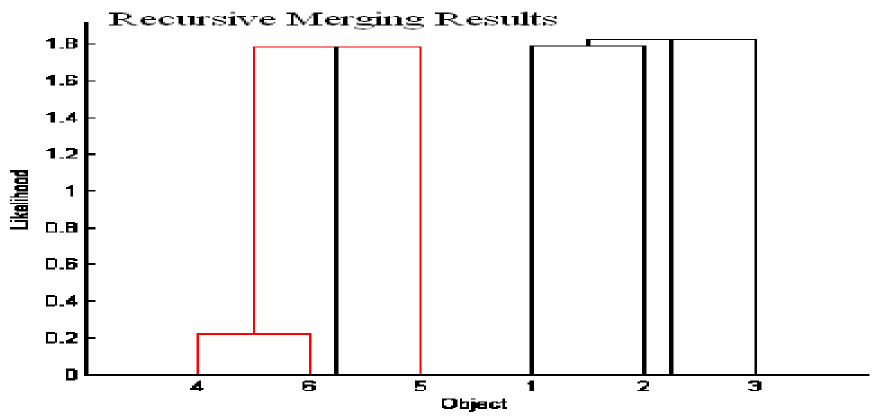
Appendix C

Toolbox Test Examples

Test Case #1

Correlation matrix =

```
[   1,  0.1,   0.1,      0,      0,   -0.3]; ...
[ 0.1,    1,   0.1,      0,      0,      0]; ...
[ 0.1,  0.1,     1,      0,      0,      0]; ...
[   0,    0,     0,      1,   0.45,   0.6]; ...
[   0,    0,     0,   0.45,      1,   0.6]; ...
[-0.3,    0,     0,    0.6,    0.6,     1]];
```



Simulated Annealing Output:

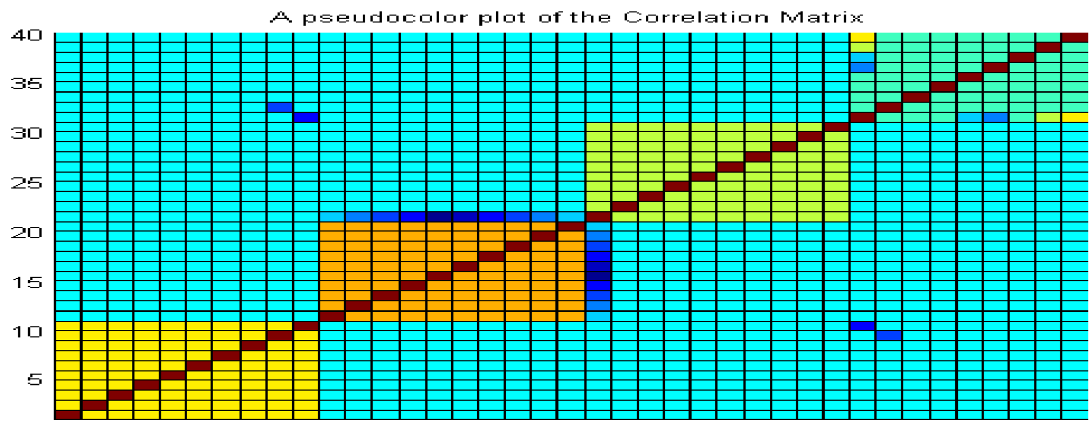
The ground state has 2 clusters, with energy = -0.883478.

Cluster 1 has an energy of -0.855078 & 3 element(s) listed below:-
5, 4, 6

Cluster 2 has an energy of -0.028399 & 3 element(s) listed below:-
2, 3, 1

Figure C.1: This is a test example of the toolbox with small synthetic data. Both algorithms where tested, the input was a 6 x 6 synthetic correlation matrix. In the correlation matrix there are two clusters of correlated objects. These clusters formed two separate branches of the dendrogram plot. The simulated annealing output gives a distinct list of objects for each cluster.

Test Case #2



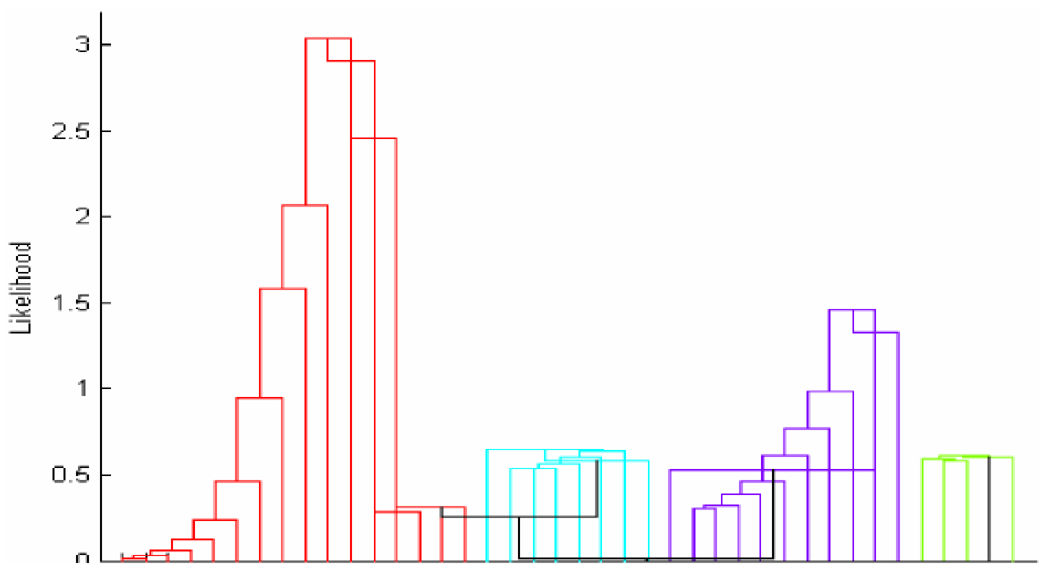


Figure C.2: This test example shows the results of the Recursive Merging algorithm. The input is a 40 X 40 correlation matrix, which is represented in a form of a pseudocolor plot. The pseudocolor plot shows four clusters in the correlation matrix, the dendrogram also picked up those four clusters, hence there are four different colors used in the dendrogram plot.

Test Case #3

Input Correlation Matrix

```
C  =  [[   1,   0.4,   0.3,    0,    0,   0.2]; ...
    [ 0.4,      1,   0.1,    0,    0,      0]; ...
    [ 0.3,   0.1,      1,    0,    0,      0]; ...
    [    0,      0,      0,    1,    0,      0]; ...
    [    0,      0,      0,    0,    1,      0]; ...
    [ 0.2,      0,      0,    0,    0,      1]];
```

Simulated Annealing Output:

The ground state has 2 clusters, with energy = -0.192866.
Cluster 1 has an energy of -0.192866 & 3 element(s) which are:-
1, 3, 2

Cluster 2 has an energy of 0.000000 & 3 element(s) which are:-
6, 4, 5

Test Case #4

Input Correlation Matrix

```
C  =  [ [    1,  0.65,      0,       0,        0,      0]; ...
   [ 0.65,       1,      0,       0,        0,      0]; ...
   [      0,       0,      1,    0.2,        0,      0]; ...
   [      0,       0,   0.2,       1,        0,      0]; ...
   [      0,       0,      0,       0,        1,  -0.1]; ...
   [      0,       0,      0,       0,    -0.1,      1]];
```

Simulated Annealing Output:

The ground state has 4 clusters, with energy = -0.589869.
Cluster 1 has an energy of -0.549047 & 2 element(s) listed below:-
1, 2
Cluster 2 has an energy of 0.000000 & 1 element(s) listed below:-
6
Cluster 3 has an energy of 0.000000 & 1 element(s) listed below:-
5
Cluster 4 has an energy of -0.040822 & 2 element(s) listed below:-

Appendix D

Merging Algorithm Sectors

Table D.1: Epoch # 1 Clusters or Market Sectors.

Cluster 1
BOWLER METCALF; BARNATO EXPLORATION; DELTA ELECTRICAL INDS; ALLAN GRAY PR.TST; JOHNNIC COMMS; LONMIN (JSE); PSG GROUP; SASANI;
Cluster 2
TRANS HEX GROUP
Cluster 3
AMAL.BEVERAGES IND; AFRICAN OXYGEN; BARLOWORLD; BASIL READ; NEW AF.CAP; NAMPAK; PIK N PAY STORES; PRETORIA PORT. CMT; RICHEMONT SECS. (JSE); SABMILLER (JSE); STD.BK GP; SOFTLINE; SASOL; VENFIN; WOOLTRU N;
Cluster 4
AFROX HEALTHCARE; BRANDCORP; BIDVEST; CAXTON CTP PUBLISH PRINT; CHEMICAL SERVICES; CAPITAL ALLIANCE; CULLINAN; INMINS; IMPERIAL HDG; JD GROUP; MEDI CLINIC; MATODZI RESOURCES; SUPER GROUP; WINHOLD;
Cluster 5
AFRIKANDER; ANGLOGOLD; ABSA GP; AVGOLD; BARPLATS INVS; ISCOR; JOHNNIC HDG; MVELAPHANDA RES; NORTHAM PLAT; OCEANA GP; SAPPI; STH.AFN.CHROME & ALS; WESTERN;
Cluster 6
ANGLOVAAL MINING; AFRICAN MEDIA ENTM; CASHBUILD; DIMENSION DATA HDG.(JSE); DNA SUPPLY CHAIN INV; ELLERINE; FOSCHINI; MOBILE INDUSTRIES; METOREX; SALLIES; SEARDEL INV; SASFIN; S & J LAND; TRENCOR; ZARARA EN;
Cluster 7
ALLIED TECHS; CONTROL INSTRUMENT GROUP; DORBYL; DISTELL GP; GRINDROD; INVICTA;KWV BELEGGINGS BEPERK; MUTUAL & FEDERAL IN; PANGBOURNE PROPS; SYCOM PROPERTY FD; WILSON BAY HLM - OVC; WOOLTRU;
Cluster 8
BYTES TECH. GP; BEV. & CONSUMER IND; CONCOR; EDGARS CONS.STORES; ELB GP; HUDACO; KERSAF INVESTMENTS; LA GROUP; MR PRICE GP; NEDCOR; PEPKOR; PIK N PAY; RAINBOW CHICKEN; REUNERT; RAND LEASES PROPS; SUN INTL.(JSE); SPESCOM;
Cluster 9
BARLOW (ISE); ASPEN PHARMACARE; ALLIED ELECTRONICS; BRAIT SA (JSE); CITY LODGE HOTELS; DS & WAREHOUSING NETWORK; DON GROUP; HIGHVELD STEEL & VANADIUM; JASCO ELECTRONICS; KAIROS INDUSTRIAL; LIBERTY HDG; PALABORA MINING; RELYANT RETAIL; RENTSURE HDG; UNITRANS;
Cluster 10
‘ AECI; INVESTEC; ELT.MEDIA NETWORK & SUPERSPORT; SANTAM; TIGER BRANDS; TONGAAT - HULETT GROUP;
Cluster 11
AFN. BK INVS; ADCORP; ANGLO AMERICAN (JSE); ANGLO AMERICAN PLAT; ANGLOVAAL INDUSTRIES; CORPCAPITAL; DURBAN ROOD.DEEP; FREE STATE DEV & INV; FIRSTRAND; GILBOA PROPERTIES; GRINTEK; HARMONY GOLD MINING; IMPALA PLATINUM; LIBERTY GP; MURRAY & ROBERTS; NU - WORLD; SUB NIGEL GDMNG; UNITED SERVICE TECHS;

Table D.2: Epoch # 2 Clusters or Market Sectors.

Cluster 1
AMAL.BEVERAGES IND; ADCORP; AFRIKANDER LEASE; CHEMICAL SERVICES; CAPITAL ALLIANCE; ELLERINE; INMINS; INVICTA; JD GROUP; KAIROS INDUSTRIAL; LONMIN (JSE); PSG GROUP; STD. BK. GP; SUPER GROUP; SASANI; WILSON BAY HLM;
Cluster 2
CITY LODGE HOTELS; SOFTLINE;
Cluster 3
CULLINAN;
Cluster 4
AFRICAN OXYGEN; ANGLOVAAL MINING; BARNATO EXPLORATION; BYTES TECH. GP; CONTROL INSTRUMENT GROUP; CASHBUILD; DS & WAREHOUSING NETWORK; DON GROUP; FOSCHINI; HIGHVELD STEEL & VANADIUM; MURRAY & ROBERTS; NORTHAM PLAT; PEPKOR; ELYANT RETAIL; RENTSURE HDG; SUN INTL. (JSE); ZARARA EN;
Cluster 5
ANGLO AMERICAN (JSE); ANGLOGOLD; ANGLOVAAL INDUSTRIES; BARLOWORLD; BARPLATS INVS; BASIL READ; CAXTON CTP PUBLISH PRINT; FREE STATE DEV. & INV; GOLD FIELDS; JOHNNIC HDG; MEDI CLINIC; OCEANA GP; RICHEMONT SECS. (JSE); WESTERN AREAS;
Cluster 6
ANGLO AMERICAN PLAT; AVGOLD; BOWLER METCALF; DURBAN ROOD. DEEP; FIRSTRAND; HARMONY GOLD MINING; IMPALA PLATINUM; METOREX; MATODZI RESOURCES; MVELAPHANDA RES; NU - WORLD; PIK N PAY STORES; STH. AFN. CHROME & ALS; UNITED SERVICE TECHS; VENFIN;
Cluster 7
BEV. & CONSUMER IND; DIMENSION DATA HDG (JSE); DISTELL GP; EDGARS CONS.STORES; HUDACO; INVESTEC; IMPERIAL HDG; PRETORIA PORT. CMT; REUNERT; SAPPI; SANTAM; SASOL; TONGAAT - HULETT GROUP; WOOLTRU N; WOOLTRU;
Cluster 8
BARLOW (ISE); AFN. BK INVS; BIDVEST GROUP; GILBOA PROPERTIES; GRINTEK; ALLAN GRAY PR. TST; ISCOR; LA GROUP; MR PRICE GP; RMB HDG; SUB NIGEL GDMNG; SASFIN; TRANS HEX GROUP; WINHOLD;
Cluster 9
DORBYL; GRINDROD; KERSAF INVESTMENTS; KWV BELEGGINGS BEPERK; MUTUAL & FEDERAL IN; ELT. MEDIA NETWORK & SUPERSPORT; MOBILE INDUSTRIES; PANGBOURNE PROPS; RAND LEASES PROPS; SEARDEL INV; S & J LAND; SPESCOM;
Cluster 10
AECI; ALLIED TECHS; ABSA GP; ALLIED ELECTRONICS; CORPCAPITAL; LIBERTY HDG; LIBERTY GP; NEW AF. CAP; NEDCOR; NAMPAK; PALABORA MINING; PIK N PAY; SABMILLER (JSE); SYCOM PROPERTY FD; TIGER BRANDS; UNITRANS;
Cluster 11
AFROX HEALTHCARE; AFRICAN MEDIA ENTM; ASPEN PHARMACARE; BRAIT SA. (JSE); BRANDCORP; CONCOR; DELTA ELECTRICAL INDS; DNA SUPPLY CHAIN INV; ELB GP; JOHNNIC COMMS; JASCO ELECTRONICS; RAINBOW CHICKEN; SALLIES; TRENCOR;

Table D.3: Epoch # 3 Clusters or Market Sectors.

Cluster 1
BARLOW (ISE); AFRIKANDER LEASE; ANGLO AMERICAN (JSE); ANGLO AMERICAN PLAT; BOWLER METCALF; CHEMICAL SERVICES; HIGHVELD STEEL & VANADIUM; INVICTA; MVELAPHANDA RES; OCEANA GP; STH.AFN.CHROME & ALS; SANTAM ; SASOL; SASANI; TRANS HEX GROUP; WESTERN AREAS; WILSON BAY HLM; WINHOLD;
Cluster 2
DELTA ELECTRICAL INDS; MEDI CLINIC; METOREX;
Cluster 3
AFROX HEALTHCARE; CITY LODGE HOTELS; CULLINAN; DORBYL; DISTELL GP; JOHNNIC COMMS; KWV BELEGGINGS BEPERK; RAND LEASES PROPS; SUN INTL (JSE); S & J LAND; SYCOM PROPERTY FD; UNITRANS;
Cluster 4
NEW AF.CAP; VENFIN;
Cluster 5
AMAL.BEVERAGES IND; ALLIED TECHS; BRANDCORP; BEV. & CONSUMER IND; DIMENSION DATA HDG (JSE); EGARS CONS.STORES; GRINTEK; ALLAN GRAY PR.TST; KERSAF INVESTMENTS; MOBILE INDUSTRIES; PEPKOR; SPESCOM; TONGAAT-HULETT GROUP; WOOLTRU N; WOOLTRU;
Cluster 6
AFN.BK.INVS; ADCORP; CAXTON CTP PUBLISH PRINT; CONTROL INSTRUMENT GROUP; CASHBUILD; GILBOA PROPERTIES; ISCOR; LONMIN (JSE); PIK N PAY STORES; RICHEMONT SECS (JSE); SUPER GROUP; UNITED SERVICE TECHS;
Cluster 7
AECI; ANGLOVAAL MINING; ALLIED ELECTRONICS; BARLOWORLD; ELLERINE; JD GROUP; LIBERTY HDG; LIBERTY GP; NEDCOR; PIK N PAY; RAINBOW CHICKEN; REUNERT; SABMILLER (JSE); STD. BK. GP; SASFIN; TIGER BRANDS;
Cluster 8
AFRICAN OXYGEN; ASPEN PHARMACARE; AVGOLD; BRAIT SA (JSE); BARNATO EXPLORATION; BASIL READ; CONCOR; DON GROUP; MUTUAL & FEDERAL IN; ELT. MEDIA NETWORK & SUPERSPORT; MURRAY & ROBERTS; PALABORA MINING; PANGBOURNE PROPS; RENTSURE HDG; ZARARA EN;
Cluster 9
AFRICAN MEDIA ENTM; BYTES TECH GP; DS & WAREHOUSING NETWORK; DNA SUPPLY CHAIN INV; ELB GP; FOSCHINI; GRINDROD; HUDACO; JASCO ELECTRONICS; KAIROS INDUSTRIAL; RELYANT RETAIL; SALLIES; SEARDEL INV; TRENCOR;
Cluster 10
ABSA GP; ANGLOVAAL INDUSTRIES; BIDVEST GROUP; CAPITAL ALLIANCE; FIRSTRAND; INVESTEC; INMINS; IMPERIAL HDG; MATODZI RESOURCES; NAMPAK; PRETORIA PORT. CMT; PSG GROUP; RMB HDG; SAPPI; SOFTLINE;
Cluster 11
ANGLOGOLD; BARPLATS INVS; CORPCAPITAL; DURBAN ROOD.DEEP; FREE STATE DEV. & INV; GOLD FIELDS; HARMONY GOLD MINING; IMPALA PLATINUM; JOHNNIC HDG; LA GROUP; MR PRICE GP; NORTHAM PLAT; NU - WORLD; SUB NIGEL GDMNG;

Table D.4: Epoch # 4 Clusters or Market Sectors.

Cluster 1
AMAL. BEVERAGES IND; MEDI CLINIC;
Cluster 2
ANGLOVAAL MINING; ALLIED TECHS; AFRICAN MEDIA ENTM; ASPEN PHARMACARE; ABSA GP; ALLIED ELECTRONICS; BRAIT SA (JSE); BARPLATS INVS; BASIL READ; BYTES TECH GP; CHEMICAL SERVICES; DS & WAREHOUSING NETWORK; DELTA ELECTRICAL INDS; DORBYL; DNA SUPPLY CHAIN INV; DON GROUP; DISTELL GP; ELLERINE; ELB GP; FOSCHINI; FREE STATE DEV. & INV; GRINTEK; HIGHVELD STEEL & VANADIUM; ISCOR; JD GROUP; JOHNNIC HDG; KAIROS INDUSTRIAL; LIBERTY HDG; LIBERTY GP; ELT.MEDIA NETWORK & SUPERSPORT; MR PRICE GP; MURRAY & ROBERTS; NEDCOR; NAMPAK; PALABORA MINING; PEPKOR; PIK N PAY STORES; PIK N PAY; REUNERT; RELYANT RETAIL; RENTSURE HDG; SABMILLER (JSE); SALLIES; STD.BK. GP; SOFTLINE; SUN INTL (JSE); S & J LAND; SANTAM; SUPER GROUP; SASANI; SYCOM PROPERTY FD; TIGER BRANDS; TONGAAT-HULETT GROUP; TRENCOR; UNITRANS; VENFIN; ZARARA EN;
Cluster 3
ADCORP; AECI; BOWLER METCALF; CAXTON CTP PUBLISH PRINT; CONCOR; CORPCAPITAL; CULLINAN; INVICTA; JOHNNIC COMMS; MUTUAL & FEDERAL IN; NU - WORLD; OCEANA GP; SASFIN; UNITED SERVICE TECHS;
Cluster 4
AFRICAN OXYGEN; DIMENSION DATA HDG (JSE); EDGARS CONS. STORES; ALLAN GRAY PR TST; HUDACO;KERSAF INVESTMENTS; KWV BELEGGINGS BEPERK; MOBILE INDUSTRIES; NORTHAM PLAT; PANGBOURNE PROPS; PRETORIA PORT. CMT; RAINBOW CHICKEN; SPESCOM;
Cluster 5
AFROX HEALTHCARE; CITY LODGE HOTELS;
Cluster 6
CASHBUILD; SEARDEL INV;
Cluster 7
LONMIN (JSE); TRANS HEX GROUP;
Cluster 8
BARLOW (ISE); AFN BK INVS; ANGLO AMERICAN (JSE); AVGOLD; BARNATO; GRINDROD; IMPALA PLATINUM; MVELAPHANDA RES; PSG GROUP; RICHEMONT SECS (JSE); STH.AFN. CHROME & ALS; SASOL; WESTERN AREAS; WILSON BAY HLM;
Cluster 9
AFRIKANDER LEASE; ANGLO AMERICAN PLAT; ANGLOGOLD; CAPITAL ALLIANCE; DURBAN ROOD.DEEP; GOLD FIELDS; GILBOA PROPERTIES; HARMONY GOLD MINING; INMINS; LA GROUP; METOREX; RAND LEASES PROPS; SUB NIGEL GDMNG; WINHOLD;
Cluster 10
ANGLOVAAL INDUSTRIES; BARLOWORLD; BRANDCORP; BEV. & CONSUMER IND; BIDVEST GROUP; CONTROL INSTRUMENT GROUP; FIRSTRAND; INVESTEC; IMPERIAL HDG; JASCO ELECTRONICS; MATODZI RESOURCES; NEW AF. CAP; RMB HDG; SAPPI;
Cluster 11
WOOLTRU N; WOOLTRU;

Table D.5: Epoch # 5 Clusters or Market Sectors.

Cluster 1
ADCORP; ANGLOGOLD; AVGOLD; DURBAN ROOD. DEEP; GILBOA PROPERTIES; GRINDROD; HARMONY GOLD MINING; ISCOR; METOREX; MVELAPHANDA RES; PSG GROUP; RAND LEASES PROPS; SUB NIGEL GDMNG; SEARDEL INV;
Cluster 2
AFROX HEALTHCARE; BEV. & CONSUMER IND; ALLAN GRAY PR. TST; HIGHVELD STEEL & VANADIUM; MUTUAL & FEDERAL IN; SYCOM PROPERTY FD; UNITRANS;
Cluster 3
AMAL.BEVERAGES IND; AFRICAN OXYGEN; ANGLOVAAL MINING; ANGLOVAAL INDUSTRIES; CULLINAN; DISTELL GP; JASCO ELECTRONICS; KERSAF INVESTMENTS; KWV BELEGGINGS BEPERK; MOBILE INDUSTRIES; RAINBOW CHICKEN; REUNERT; SPESCOM; TRENCOR;
Cluster 4
ALLIED TECHS; AFRICAN MEDIA ENTM; BARNATO EXPLORATION; CITY LODGE HOTELS; DELTA ELECTRICAL INDS; DNA SUPPLY CHAIN INV; EDGARS CONS.STORES; ELB GP; HUDACO; MURRAY & ROBERTS; NORTHAM PLAT; PALABORA MINING; S & J LAND;
Cluster 5
ABSA GP; ALLIED ELECTRONICS; CHEMICAL SERVICES; DORBYL; ELLERINE; JOHNNIC COMMS; LIBERTY HDG; LIBERTY GP; NEDCOR; NAMPAK; RMB HDG; STD BK GP; SASOL; SUPER GROUP;
Cluster 6
BIDVEST GROUP; CAPITAL ALLIANCE; GRINTEK; MEDI CLINIC; NEW AF. CAP; PEPKOR; PIK N PAY STORES; PRETORIA PORT. CMT; PIK N PAY; SABMILLER (JSE); SASFIN; TONGAAT-HULETT GROUP; VENFIN; WOOLTRU N; WOOLTRU;
Cluster 7
IMPERIAL HDG;
Cluster 8
AFRIKANDER LEASE; ANGLO AMERICAN (JSE); ANGLO AMERICAN PLAT; BARLOWORLD; CONCOR; IMPALA PLATINUM; INVESTEC; INMINS; JD GROUP; JOHNNIC HDG; LONMIN (JSE); RICHEMONT SECS (JSE); SAPPI; TIGER BRANDS; WESTERN AREAS; WILSON BAY HLM;
Cluster 9
BRAIT SA (JSE); BARPLATS INVS; BASIL READ; BYTES TECH GP; CONTROL INSTRUMENT GROUP; DS & WAREHOUSING NETWORK; DON GROUP; FOSCHINI; GOLD FIELDS; KAIROS INDUSTRIAL; ELT. MEDIA NETWORK & SUPERSPORT; MR PRICE GP; RELYANT RETAIL; RENTSURE HDG; SOFTLINE; SUN INTL (JSE); ZARARA EN;
Cluster 10
BARLOW (ISE); AFN. BK. INVS; AECI; CAXTON CTP PUBLISH PRINT; CASHBUILD; DIMENSION DATA HDG (JSE); FIRSTRAND; NU - WORLD; STH. AFN CHROME & ALS; TRANS HEX GROUP; UNITED SERVICE TECHS; WINHOLD;
Cluster 11
ASPEN PHARMACARE; BOWLER METCALF; BRANDCORP; CORPCAPITAL; FREE STATE DEV. & INV; INVICTA; LA GROUP; MATODZI RESOURCES; OCEANA GP; PANGBOURNE PROPS; SALLIES; SANTAM; SASANI ;

Table D.6: Epoch # 6 Clusters or Market Sectors.

Cluster 1
AFRIKANDER LEASE; AFRICAN OXYGEN; BRAIT SA (JSE); DON GROUP; FREE STATE DEV & INV; LA GROUP; MUTUAL & FEDERAL IN; PIK N PAY; RELYANT RETAIL; RENTSURE HDG; SASFIN; SOFTLINE; S & J LAND; SASANI; ZARARA EN;
Cluster 2
ELLERINE;
Cluster 3
AMAL. BEVERAGES IND; AFROX HEALTHCARE; ANGLOVAAL MINING; ALLIED TECHS; CULLINAN; DORBYL; GRINDROD; KWV BELEGGINGS BEPERK; NORTHAM PLAT; PANGBOURNE PROPS; RAINBOW CHICKEN; RAND LEASES PROPS; SUN INTL (JSE); TRENCOR;
Cluster 4
BARNATO EXPLORATION; BARPLATS INVS; BRANDCORP; BASIL READ; DELTA ELECTRICAL INDS; LONMIN (JSE); MEDI CLINIC; NU-WORLD; OCEANA GP; SALLIES; SANTAM; UNITRANS; WILSON BAY HLM;
Cluster 5
ADCORP; AECI; CITY LODGE HOTELS; CONCOR; CORPCAPITAL; GRINTEK; ALLAN GRAY PR. TST; INMINS; KAIROS INDUSTRIAL; PSG GROUP; RICHEMONT SECS (JSE); SUB NIGEL GDMNG; STH. AFN CHROME & ALS; TRANS HEX GROUP; WINHOLD;
Cluster 6
HIGHVELD STEEL & VANADIUM; ELT. MEDIA NETWORK & SUPERSPORT; TONGAAT - HULETT GROUP;
Cluster 7
BARLOW (ISE); ANGLO AMERICAN (JSE); ANGLOGOLD; AVGOLD; HARMONY GOLD MINING; IMPALA PLATINUM; INVICTA; JOHNNIC COMMS; METOREX; MVELAPHANDA RES; SEARDEL INV; SASOL; SUPER GROUP; UNITED SERVICE TECHS; WESTERN AREAS;
Cluster 8
AFN. BK INVS; ANGLO AMERICAN PLAT; BEV & CONSUMER IND; CAXTON CTP PUBLISH PRINT; CONTROL INSTRUMENT GROUP; DIMENSION DATA HDG (JSE); FIRSTRAND; ISCOR; MATODZI RESOURCES; NEDCOR; PEPKOR; RMB HDG; SAPPI; STD BK GP; WOOLTRU N; WOOLTRU;
Cluster 9
AFRICAN MEDIA ENTM; ASPEN PHARMACARE; BOWLER METCALF; CASHBUILD; DNA SUPPLY CHAIN INV; DURBAN ROOD. DEEP; ELB GP; FOSCHINI; GOLD FIELDS; HUDACO; JASCO ELECTRONICS; MR PRICE GP; MURRAY & ROBERTS; PALABORA MINING; PRETORIA PORT. CMT; SYCOM PROPERTY FD;
Cluster 10
ABSA GP; ALLIED ELECTRONICS; BARLOWORLD; DISTELL GP; INVESTEC; JOHNNIC HDG; KERSAF INVESTMENTS; LIBERTY HDG; LIBERTY GP; MOBILE INDUSTRIES; NEW AF. CAP; PIK N PAY STORES; REUNERT; SPESCOM;
Cluster 11
ANGLOVAAL INDUSTRIES; BYTES TECH.GP; BIDVEST GROUP; CHEMICAL SERVICES; CAPITAL ALLIANCE; DS & WAREHOUSING NETWORK; EDGARS CONS. STORES; GILBOA PROPERTIES; IMPERIAL HDG; JD GROUP; NAMPAK; SABMILLER (JSE); TIGER BRANDS; VENFIN;

Appendix E

Simulated Annealing Market Sectors

E.1 SA: Epoch #1 Sectors

1. SASOL; SOFTLINE;----
2. SUPER GROUP; MEDI CLINIC; FOSCHINI;----
3. HARMONY GOLD MINING; WESTERN AREAS; GOLD FIELDS; ANGLOGOL; ANGLO AMERICAN (JSE);--
4. ANGLO AMERICAN PLAT.; IMPALA PLATINUM ;----
5. DON GROUP; RENTSURE HDG; AFRICAN MEDIA ENTM;
   DNA SUPPLY CHAIN INV; ZARARA EN; ELB GP;----
6. NEDCOR; AVGOLD;----
7. STD.BK.GP; PALABORA MINING
8. SASANI;----
9. WOOLTRU;----
10. NEW AF.CAP;----
11. GILBOA PROPERTIES;----
12. GRINDROD; UNITRANS; KERSAF INVESTMENTS;----
13. NORTHAM PLAT;----
14. CITY LODGE HOTELS; BARLOW (ISE) ;----
15. LIBERTY HDG;----
16. SAPPI; INVESTEC;----
17. DIMENSION DATA HDG.(JSE); ALLIED ELECTRONICS;----
18. RAND LEASES PROPS;----
19. SALLIES;----
20. CONCOR; CONTROL INSTRUMENT GROUP;----
21. DORBYL; TRANS HEX GROUP; MR PRICE GP;----
22. AFROX HEALTHCARE;----
23. DS.& WAREHOUSING NETWORK ;----
24. ISCOR; HIGHVELD STEEL&VANADIUM; EDGARS CONS.STORES
25. WOOLTRU N;----
26. JASCO ELECTRONICS;----
27. SUB NIGEL GDMNG.
28. INMINS;----
29. UNITED SERVICE TECHS.; MUTUAL & FEDERAL IN.; LONMIN (JSE) ;----
30. PSG GROUP;----
31. SANTAM;----
32. RMB HDG;----

33. GRINTEK; BEV.& CONSUMER IND. DEAD;----
34. CULLINAN; FIRSTRAND; PANGBOURNE PROPS;----
35. DELTA ELECTRICAL INDS; MATODZI RESOURCES;----
36. SPESCOM; AFN.BK.INVS;----
37. METOREX; HUDACO;----
38. JD GROUP; AMAL.BEVERAGES IND;----
39. BASIL READ; SUN INTL.(JSE); DISTELL GP;----
40. CORPCAPITAL; RICHEMONT SECS. (JSE) ;----
41. PIK N PAY;----
42. MVELAPHANDA RES.; PEPKOR ;----
43. SASFIN;----
44. PRETORIA PORT.CMT;----
45. VENFIN; SABMILLER (JSE); MURRAY & ROBERTS ;----
46. ANGLOVAAL MINING; INVICTA;----
47. BRANDCORP;----
48. LA GROUP; AFRIKANDER LEASE;----
49. ALLIED TECHS.; ANGLOVAAL INDUSTRIES; ASPEN PHARMACARE;----
50. TRENCOR; MOBILE INDUSTRIES; BRAIT SA. (JSE) ;----
51. ARPLATS INVS;----
52. KAIROS INDUSTRIAL; AECI ;----
53. WILSON BAY HLM - OVC; BOWLER METCALF ;----
54. CAPITAL ALLIANCE; SEARDEL INV;----
55. PIK N PAY STORES; JOHNNIC COMMS;----
56. AFRICAN OXYGEN;----
57. CASHBUILD; NAMPAK; BARLOWORLD ;----
58. RAINBOW CHICKEN; BARNATO EXPLORATION;----
59. RELYANT RETAIL ;---
60. FREE STATE DEV & INV; KWV BELEGGINGS BEPERK;----
61. TIGER BRANDS; ADCORP; OCEANA GP;----
62. ABSA GP;----
63. BIDVEST GROUP;----
64. S & J LAND;----
65. ALLAN GRAY PR.TST.; REUNERT;----
66. SYCOM PROPERTY FD.
67. IMPERIAL HDG.; BYTES TECH.GP; DURBAN ROOD.DEEP; CAXTON CTP PUBLISH PRINT;----
68. LIBERTY GP.; ELLERINE;----
69. WINHOLD; TONGAAT - HULETT GROUP;----
70. ELT.MEDIA NETWORK & SUPERSPORT; JOHNNIC HDG;----
71. NU - WORLD;----
72. CHEMICAL SERVICES; STH.AFN.CHROME&ALS;----

E.2 SA: Epoch #2 Sectors

1. WESTERN AREAS; GOLD FIELDS; HARMONY GOLD MINING; ANGLOGOLD;----
2. ANGLO AMERICAN (JSE); BARLOWORLD; SABMILLER (JSE); JOHNNIC HDG.; SASOL; VENFIN;----
3. STD.BK.GP; NEDCOR; BYTES TECH.GP;----
4. INMINS; GILBOA PROPERTIES;----

5. ELB GP; DON GROUP; RENTSURE HDG; DNA SUPPLY CHAIN INV; AFRICAN MEDIA ENTM; ZARARA EN;----
6. SPESCOM; BASIL READ;----
7. SOFTLINE; DURBAN ROOD.DEEP;----
8. RAINBOW CHICKEN; GRINDROD;----
9. HIGHVELD STEEL&VANADIUM ;----
10. ANGLO AMERICAN PLAT; IMPALA PLATINUM;----
11. MB HDG;----
12. OCEANA GP;----
13. JASCO ELECTRONICS;----
14. ANGLOVAAL INDUSTRIES;----
15. MEDI CLINIC;----
16. AFROX HEALTHCARE;---
17. CASHBUILD; AFRICAN OXYGEN; CONCOR;----
18. MOBILE INDUSTRIES; TRENCOR;----
19. PIK N PAY STORES; PIK N PAY;----
20. SUPER GROUP; BARLOW (ISE) ;----
21. EDGARS CONS.STORES;----
22. SYCOM PROPERTY FD; BARNATO EXPLORATION;----
23. NAMPAK;----
24. CAXTON CTP PUBLISH PRINT; FIRSTRAND; ADCORP;----
25. ASPEN PHARMACARE;----
26. ELLERINE;----
27. IMPERIAL HDG.; MATODZI RESOURCES; BIDVEST GROUP;----
28. RELYANT RETAIL;----
29. STH.AFN.CHROME&ALS.; LA GROUP;----
30. JD GROUP;----
31. CONTROL INSTRUMENT GROUP; NEW AF.CAP;--
32. TRANS HEX GROUP; DORBYL; RICHEMONT SECS. (JSE) ;----
33. ABSA GP.; TIGER BRANDS;----
34. S & J LAND;----
35. MVELAPHANDA RES;----
36. UNITRANS;----
37. ALLAN GRAY PR.TST;----
38. ONMIN (JSE); SASFIN;----
39. MR PRICE GP;----
40. AVGOLD;----
41. KWV BELEGGINGS BEPERK; SUN INTL.(JSE) ;----
42. SANTAM;----
43. PRETORIA PORT.CMT; PEPKOR;----
44. BARPLATS INVS;----
45. BEV.& CONSUMER IND. DEAD; WOOLTRU; ISCOR;----
46. HUDACO; MURRAY & ROBERTS; METOREX;----
47. AFN.BK.INVS.; UNITED SERVICE TECHS.; SUB NIGEL GDMNG.
48. PSG GROUP;--- -
49. MUTUAL & FEDERAL IN;----
50. DISTELL GP; RAND LEASES PROPS;----
51. DS.& WAREHOUSING NETWORK ;----
52. CULLINAN ;----
53. SEARDEL INV; PANGBOURNE PROPS;----

54. AECI; AFRIKANDER LEASE;----
55. DELTA ELECTRICAL INDS;----
56. WILSON BAY HLM-OVC;----
57. ALLIED ELECTRONICS;----
58. WOOLTRU N;----
59. CAPITAL ALLIANCE; NU-WORLD; AMAL.BEVERAGES IND; BRAIT SA. (JSE) ;----
60. INVESTEC; CORPCAPITAL; CHEMICAL SERVICES;----
61. GRINTEK;----
62. NORTHAM PLAT; DIMENSION DATA HDG.(JSE); BRANDCORP
63. WINHOLD; ALLIED TECHS;----
64. KAIROS INDUSTRIAL;----
65. SASANI;----
66. CITY LODGE HOTELS ;----
67. FOSCHINI;----
68. SAPPI; REUNERT;----
69. LIBERTY HDG; LIBERTY GP;----
70. TONGAAT - HULETT GROUP; JOHNNIC COMMS;----
71. INVICTA;----
72. PALABORA MINING; ANGLOVAAL MINING ;----
73. FREE STATE DEV & INV;----
74. ELT.MEDIA NETWORK & SUPERSPORT;----
75. BOWLER METCALF;----
76. KERSAF INVESTMENTS; SALLIES

E.3 SA: Epoch #3 Sectors

1. TIGER BRANDS; NEW AF.CAP;---
2. INVICTA;---
3. DON GROUP; RENTSURE HDG; ELB GP; AFRICAN MEDIA ENTM; DNA SUPPLY CHAIN INV;---
4. SUB NIGEL GDMNG; AFN.BK.INVS; UNITED SERVICE TECHS;---
5. SUN INTL(JSE); CULLINAN;---
6. ANGLO AMERICAN PLAT; IMPALA PLATINUM;---
7. DURBAN ROOD.DEEP; WESTERN AREAS; GOLD FIELDS; ANGLOGOLD; HARMONY GOLD MINING; AVGOLD;---
8. SOFTLINE;---
9. SASOL;---
10. SPESCOM; METOREX;---
11. HUDACO;---
12. ASPEN PHARMACARE; BASIL READ; CAPITAL ALLIANCE;---
13. VENFIN;---
14. NORTHAM PLAT.; EDGARS CONS.STORES; GILBOA PROPERTIES;---
15. WOOLTRU N;---
16. LIBERTY HDG.; LIBERTY GP;---
17. ALLIED TECHS; ALLIED ELECTRONICS;---
18. AMAL.BEVERAGES IND; PIK N PAY STORES;---
19. BRANDCORP; AECI;---
20. STD.BK.GP;---
21. BARPLATS INVS.

22. SALLIES; MURRAY & ROBERTS;--
23. RELYANT RETAIL; BRAIT SA. (JSE);---
24. WINHOLD;---
25. TONGAAT - HULETT GROUP; RICHEMONT SECS. (JSE); ANGLOVAAL MINING; INMINS;---
26. PALABORA MINING; TRENCOR;---
27. RAND LEASES PROPS;---
28. CORPCAPITAL; JASCO ELECTRONICS;---
29. WILSON BAY HLM - OVC;---
30. BARLOWORLD;---
31. BIDVEST GROUP;---
32. DISTELL GP;---
33. CONCOR; ALLAN GRAY PR.TST;---
34. MEDI CLINIC;---
35. MATODZI RESOURCES; AFRICAN OXYGEN;---
36. WOOLTRU; CONTROL INSTRUMENT GROUP;---
37. REUNERT; TRANS HEX GROUP;---
38. SABMILLER (JSE); CASHBUILD; ADCORP;---
39. DELTA ELECTRICAL INDS.; RAINBOW CHICKEN;---
40. SASANI;---
41. FIRSTRAND; RMB HDG;---
42. INVESTEC;---
43. HIGHVELD STEEL & VANADIUM; SUPER GROUP;---
44. BYTES TECH.GP;---
45. ABSA GP;---
46. SYCOM PROPERTY FD;---
47. KAIROS INDUSTRIAL; STH.AFN.CHROME&ALS;---
48. MR PRICE GP;---
49. EPKOR; FROX HEALTHCARE; BARNATO EXPLORATION;---
50. ISCOR; SAPPI; LONMIN (JSE) ;---
51. ANGLOVAAL INDUSTRIES;---
52. KWV BELEGGINGS BEPERK; UNITRANS;---
53. ELT.MEDIA NETWORK & SUPERSPORT;---
54. KERSAF INVESTMENTS; CITY LODGE HOTELS;---
55. SEARDEL INV;---
56. NU - WORLD;---
57. SASFIN; IMPERIAL HDG;---
58. CHEMICAL SERVICES; ANGLO AMERICAN (JSE); S & J LAND;---
59. SANTAM;---
60. PRETORIA PORT.CMT;---
61. DS.& WAREHOUSING NETWORK; AFRIKANDER LEASE;---
62. PIK N PAY;---
63. FOSCHINI;---
64. NAMPAK; MOBILE INDUSTRIES;---
65. DIMENSION DATA HDG(JSE) ;---
66. JOHNNIC COMMS;---
67. GRINDROD; PANGBOURNE PROPS;---
68. GRINTEK;---
69. BARLOW (ISE) ;---
70. PSG GROUP; MUTUAL & FEDERAL IN.; LA GROUP;---
71. CAXTON CTP PUBLISH PRINT;---

72. ZARARA EN.; DORBYL; JOHNNIC HDG; BOWLER METCALF;---
73. ELLERINE; OCEANA GP; MVELAPHANDA RES;---
74. BEV& CONSUMER IND;---
75. NEDCOR;---
76. FREE STATE DEV.& INV;---

E.4 SA: Epoch #4 Sectors

1. JOHNNIC HDG.; STD.BK.GP.; ABSA GP.; NEDCOR;---
2. WOOLTRU N; WOOLTRU;---
3. IMPALA PLATINUM; ANGLO AMERICAN PLAT;---
4. LIBERTY GP; LIBERTY HDG;---
5. RMB HDG; FIRSTRAND;---
6. GOLD FIELDS; WESTERN AREAS; ANGLOGOLD; AVGOLD;
HARMONY GOLD MINING; DURBAN ROOD.DEEP;---
7. WINHOLD; SALLIES; JD GROUP;---
8. BOWLER METCALF ; CHEMICAL SERVICES; VENFIN;---
9. SASANI; AFROX HEALTHCARE;---
10. BARNATO EXPLORATION;---
11. DORBYL;---
12. BASIL READ;---
13. CONTROL INSTRUMENT GROUP; NAMPAK; BARPLATS INVS.; SANTAM;---
14. AECI;---
15. UNITED SERVICE TECHS;---
16. SUB NIGEL GDMNG.; MR PRICE GP;---
17. RAND LEASES PROPS;---
8. DELTA ELECTRICAL INDS;---
19. UNITRANS; FOSCHINI;---
20. NU - WORLD; AFN.BK.INVS;---
21. BRAIT SA. (JSE) ;---
22. LA GROUP;---
23. RELYANT RETAIL;---
24. BARLOWORLD; AFRICAN OXYGEN; DIMENSION DATA HDG.(JSE) ;---
25. PIK N PAY STORES; SPESCOM;---
26. PRETORIA PORT.CMT;---
27. ALLIED ELECTRONICS; ANGLO AMERICAN (JSE) ;---
28. METOREX; OCEANA GP;---
29. CITY LODGE HOTELS;---
30. BIDVEST GROUP;---
31. STH.AFN.CHROME&ALS; CONCOR;---
32. KERSAF INVESTMENTS;---
33. CORPCAPITAL;---
34. ALLIED TECHS;---
35. CAXTON CTP PUBLISH PRINT; MVELAPHANDA RES;---
36. DS & WAREHOUSING NETWORK;---
37. BRANDCORP;---
38. AFRIKANDER LEASE; ISCOR; SUN INTL.(JSE) ;---
39. MURRAY & ROBERTS;---
40. PEPKOR; ELT.MEDIA NETWORK & SUPERSPORT;---

41. PSG GROUP; BARLOW (ISE); FREE STATE DEV & INV; PIK N PAY;---
42. CAPITAL ALLIANCE; TRANS HEX GROUP;---
43. KWV BELEGGINGS BEPERK;---
44. SEARDEL INV; SAPPI; ANGLOVAAL INDUSTRIES;---
45. NORTHAM PLAT;---
46. AMAL.BEVERAGES IND;---
47. INMINS;---
48. ALLAN GRAY PR.TST; PALABORA MINING; BEV.& CONSUMER IND. DEAD;---
49. RAINBOW CHICKEN; ELLERINE;---
50. JOHNNIC COMMS;---
51. MATODZI RESOURCES;---
52. NEW AF.CAP;---
53. GILBOA PROPERTIES;---
54. SOFTLINE;---
55. TONGAAT - HULETT GROUP;---
56. REUNERT;---
57. CASHBUILD;---
58. EDGARS CONS.STORES;---
59. ADCORP;---
60. ASPEN PHARMACARE; GRINTEK; IMPERIAL HDG; JASCO ELECTRONICS;---
61. DNA SUPPLY CHAIN INV; DON GROUP; ELB GP; AFRICAN MEDIA ENTM ; RENTSURE HDG;---
62. S & J LAND; INVESTEC;---
63. INVICTA;---
64. DISTELL GP; SASOL;---
65. KAIROS INDUSTRIAL; HUDACO;---
66. CULLINAN;---
67. LONMIN (JSE); BYTES TECH.GP; ANGLOVAAL MINING; SASFIN;---
68. ZARARA EN; PANGBOURNE PROPS;---
69. HIGHVELD STEEL&VANADIUM;---
70. RICHEMONT SECS (JSE);---
71. MUTUAL & FEDERAL IN;---
72. WILSON BAY HLM - OVC;---
73. TRENCOR; MOBILE INDUSTRIES;---
74. GRINDROD;---
75. SUPER GROUP;---
76. TIGER BRANDS;---
77. SABMILLER (JSE); SYCOM PROPERTY FD;---
78. MEDI CLINIC;---

E.5 SA: Epoch #5 Sectors

1. JOHNNIC COMMS;---
2. ANGLOGOLD; AVGOLD; HARMONY GOLD MINING; GOLD FIELDS;
   DURBAN ROOD.DEEP; WESTERN AREAS;---
3. ELT.MEDIA NETWORK & SUPERSPORT; CITY LODGE HOTELS;---
4. PIK N PAY; PIK N PAY STORES;---
5. SUB NIGEL GDMNG.; AFN.BK.INVS.; PSG GROUP; NU - WORLD;
   MATODZI RESOURCES;---
6. RAINBOW CHICKEN; BYTES TECH.GP; ANGLOVAAL MINING;---

7. SALLIES;---
8. OCEANA GP;---
9. RELYANT RETAIL; JASCO ELECTRONICS; REUNERT;---
10. MOBILE INDUSTRIES; NEW AF.CAP;---
11. METOREX; ALLIED ELECTRONICS; MUTUAL & FEDERAL IN;---
12. INVICTA; ISCOR;---
13. RICHEMONT SECS. (JSE); CAPITAL ALLIANCE;---
14. ALLIED TECHS.; DORBYL; ADCORP;---
15. ELLERINE;---
16. GRINDROD;---
17. ANGLO AMERICAN PLAT; IMPALA PLATINUM;---
18. PEPKOR;---
19. SANTAM;---
20. NORTHAM PLAT;---
21. BARLOWORLD; SUN INTL.(JSE) ;---
22. SAPPI; ASPEN PHARMACARE; BARLOW (ISE) ;---
23. LA GROUP;---
24. STH.AFN.CHROME&ALS;---
25. ALLAN GRAY PR.TST.; RAND LEASES PROPS;---
26. MR PRICE GP;---
27. VENFIN; EDGARS CONS.STORES;---
28. FREE STATE DEV & INV;---
29. KAIROS INDUSTRIAL;---
30. BOWLER METCALF;---
31. HUDACO; SYCOM PROPERTY FD;---
32. BIDVEST GROUP;---
33. MVELAPHANDA RES;---
34. AFROX HEALTHCARE;---
35. AMAL.BEVERAGES IND; CORPCAPITAL;---
36. JD GROUP;---
37. INMINS;---
38. PALABORA MINING;---
39. AECI; ---
40. SASOL; ---
41. CASHBUILD; ---
42. LONMIN (JSE); TRENCOR ;---
43. ZARARA EN.; KWV BELEGGINGS BEPERK; SEARDEL INV;
DS.& WAREHOUSING;--- NETWORK; CONTROL INSTRUMENT GROUP;---
44. BARNATO EXPLORATION;---
45. DNA SUPPLY CHAIN INV; ELB GP; RENTSURE HDG;
DON GROUP; AFRICAN MEDIA ENTM;---
46. KERSAF INVESTMENTS; UNITED SERVICE TECHS;---
47. DISTELL GP; CHEMICAL SERVICES;---
48. WOOLTRU N; WOOLTRU;---
49. GRINTEK;---
50. S & J LAND;---
51. CONCOR; DIMENSION DATA HDG.(JSE) ;---
52. TIGER BRANDS;---
53. CULLINAN; AFRICAN OXYGEN;---
54. FOSCHINI;---

55. MEDI CLINIC;---
56. IMPERIAL HDG; BRAIT SA (JSE);---
57. PANGBOURNE PROPS; GILBOA PROPERTIES;---
58. SPESCOM; BEV.& CONSUMER IND. DEAD;---
59. NAMPAK;---
60. BASIL READ; CAXTON CTP PUBLISH PRINT;---
61. WINHOLD;---
62. RMB HDG.; FIRSTRAND; NEDCOR; STD.BK.GP.; ABSA GP; INVESTEC; BARPLATS INVS;---
63. SUPER GROUP; DELTA ELECTRICAL INDS;---
64. PRETORIA PORT.CMT;---
65. ANGLO AMERICAN (JSE); BRANDCORP;---
66. HIGHVELD STEEL&VANADIUM;---
67. AFRIKANDER LEASE;---
68. TONGAAT - HULETT GROUP;---
69. JOHNNIC HDG;---
70. SASANI; SABMILLER (JSE) ;---
71. UNITRANS; WILSON BAY HLM - OVC;---
72. LIBERTY GP; LIBERTY HDG;---
73. SOFTLINE; MURRAY & ROBERTS;---
74. SASFIN;---
75. ANGLOVAAL INDUSTRIES;---
76. TRANS HEX GROUP;---

E.6 SA: Epoch #6 Sectors

1. ABSA GP; NEDCOR; STD.BK.GP;---
2. ALLIED TECHS; SASANI; CHEMICAL SERVICES;---
3. IMPALA PLATINUM; ANGLO AMERICAN PLAT;---
4. MR PRICE GP; NU - WORLD; ADCORP; BYTES TECH.GP;---
5. DURBAN ROOD.DEEP; HARMONY GOLD MINING; GOLD FIELDS; ANGLOGOLD; AVGOLD; WESTERN AREAS;---
6. RMB HDG; FIRSTRAND;---
7. REUNERT; JASCO ELECTRONICS;---
8. CAPITAL ALLIANCE; HIGHVELD STEEL &VANADIUM;---
9 . WINHOLD;---
10. SPESCOM; MVELAPHANDA RES;---
11. TONGAAT-HULETT GROUP;---
12. SOFTLINE;---
13. PANGBOURNE PROPS;---
14. CAXTON CTP PUBLISH PRINT;---
15. FOSCHINI;---
16. INVICTA;---
17. ELLERINE;---
18. ANGLO AMERICAN (JSE) ;---
19. FREE STATE DEV.& INV;---
20. KWV BELEGGINGS BEPERK; CONCOR; SANTAM; INVESTEC;---
21. BRAIT SA. (JSE) ;---

```
22.     JOHNNIC COMMS;   PSG GROUP;     NORTHAM PLAT;     NAMPAK;---
23.     KERSAF INVESTMENTS;  UNITED SERVICE TECHS;  BOWLER METCALF;  BIDVEST GROUP;---
24.     ELT.MEDIA NETWORK  & SUPERSPORT;  PRETORIA PORT.CMT;---
25.     SALLIES;---
26.     NEW AF.CAP;---
27.     CULLINAN;---
28.     HUDACO;---
29.     WILSON BAY HLM-OVC;    AFRICAN OXYGEN;     ANGLOVAAL INDUSTRIES;---
30.     ANGLOVAAL MINING;---
31.     LONMIN (JSE);     BARLOWORLD;      ASPEN PHARMACARE;---
32.     WOOLTRU N;     WOOLTRU;---
33.     VENFIN;   JOHNNIC HDG;---
34.     SUN INTL.(JSE) ;---
35.     MEDI CLINIC;---
36.     SUB NIGEL GDMNG;     BARLOW (ISE);      MURRAY & ROBERTS;---
37.     RENTSURE HDG;   DNA SUPPLY CHAIN INV;   ELB GP;  DON GROUP;
        AFRICAN MEDIA ENTM;---
38.     MOBILE INDUSTRIES;    RAND LEASES PROPS;     DISTELL GP;
        SABMILLER; (JSE);    LA GROUP;---
39.     AECI;---
40.     ALLAN GRAY PR.TST;      AMAL.BEVERAGES IND;---
41.     GILBOA PROPERTIES;---
42.     KAIROS INDUSTRIAL;---
43.     BARPLATS INVS;      SEARDEL INV;     IMPERIAL HDG;---
44.     DIMENSION DATA HDG.(JSE) ;---
45.     ALLIED ELECTRONICS;---
46.     BEV & CONSUMER IND DEAD;   DORBYL;---
47.     GRINDROD;  PALABORA MINING;---
48.     UNITRANS;---
49.     BASIL READ;---
50.     TIGER BRANDS;     MUTUAL & FEDERAL IN;---
51.     INMINS;     AFN.BK.INVS.;     TRANS HEX GROUP;---
52.     SAPPI;      SASOL;     ISCOR;---
53.     MATODZI RESOURCES;---
54.     SYCOM PROPERTY FD;     TRENCOR;---
55.     CORPCAPITAL;     SUPER GROUP;---
56.     RICHEMONT SECS (JSE);    DS.& WAREHOUSING NETWORK;---
57.     PIK N PAY STORES;  PIK N PAY;---
58.     CASHBUILD;    PEPKOR;---
59.     AFROX HEALTHCARE ;   AFRIKANDER LEASE;---
60.     EDGARS CONS.STORES;---
61.     ZARARA EN ;---
62.     RAINBOW CHICKEN;---
63.     CITY LODGE HOTELS;   JD GROUP;---
64.     BARNATO EXPLORATION;---
65.     DELTA ELECTRICAL INDS;    SASFIN;---
66.     OCEANA GP;---
67.     RELYANT RETAIL;       METOREX;     CONTROL INSTRUMENT GROUP;---
68.     GRINTEK;---
69.     S & J LAND;      STH.AFN.CHROME & ALS;---
```

```
70.     LIBERTY HDG;    LIBERTY GP;---
71.     BRANDCORP;---
```

Appendix F

Standard Industrial Classification of the JSE Stocks

STOCK NAME	Sub-sector	Sector	Economic Group
ABL ABIL	Consumer Finance	General Financial	Financials
ACP Acucap	Real Estate Holding & Development	Real Estate	Financials
ADH Advtech	Specialised consumer services	General Retailers	Consumer Services
ADR Adcorp	Bus. Training & Emplymnt Agencies	Support Services	Industrials
AEG Aveng	Heavy construction	Constr. & Materials	Industrials
AFB Alexfbs	Insurance Brokers	Nonlife Insurance	Financials
AFE AECI	Speciality Chemicals	Chemicals	Basic Materials
AFR Afgri	Farming and Fishing	Food Producers	Consumer Goods
AFX Afrox	Speciality Chemicals	Chemicals	Basic Materials
AGL Anglos	General Mining	Mining	Basic Materials
ALT Altech	Telecommunication Equipment	Tech. Hardware & Equip.	Technology
AMA Amap	Consumer Electronics	Leisure Goods	Consumer Goods
AMS Angloplat	Platinum and Precious Metals	Mining	Basic Materials
ANG AnglAshanti	Gold Mining	Mining	Basic Materials
APA Apexhi-A	Real Estate Holding & Development	Real Estate	Financials
APB Apexhi-B	Real Estate Holding & Development	Real Estate	Financials
APK Astrapak	Containers and Packaging	General Industrials	Industrials
APN Aspen	Pharmaceutical	Pharma.& Biotechnology	Health Care
ARI AfrRainbow	General Mining	Mining	Basic Materials
ARL Astral	Farming and Fishing	Food Producers	Consumer Goods
ART Argent	Diversified Industrials	General Industrials	Industrials
ASA ABSA	Banks	Banks	Financials
ATN Altron	Electrical Components & Equipment	Electrnc& Elect.Equip.	Industrials
ATNP Altron	Electrical Components & Equipment	Electrnc& Elect.Equip.	Industrials
ATS Atlas	Real Estate Holding & Development	Real Estate	Financials
AVI AVI	Food Products	Food Producers	Consumer Goods
BAT Brait	Investment Services	General Financial	Financials
BAW Barloworld	Diversified Industrials	General Industrials	Industrials
BCX BusConnex	Computer Services	Software & Comp. Serv.	Technology
BEL Bell Equip	Commercial Veh. & Trucks	Industrial Engineering	Industrials

BIL	BHPBilliton	General Mining	Mining	Basic Materials
BPL	Barplats	Platinum and Precious Metals	Mining	Basic Materials
BRC	Brandco	Broadline Retailers	General Retailers	Consumer Services
BRN	Brimston-N	Equity Investment Instruments	Equity Invest. Instr.	Financials
BTG	BTG	Computer Services	Software & Comp. Serv.	Technology
BVT	Bidvest	Business Support Services	Support Services	Industrials
CAT	Caxton	Publishing	Media	Consumer Services
CDZ	Cadiz	Investment Services	General Financial	Financials
CLE	Clientle	Life Insurance	Life Insurance	Financials
CLH	City Lodge	Hotels	Travel & Leisure	Consumer Services
CMH	Comb.Motors	Speciality Retailers	General Retailers	Consumer Services
CML	Coro-FM	Asset Manager	General Financial	Financials
CPI	Capitec	Banks	Banks	Financials
CPL	CaptProp	Real Estate Investment Trust	Real Estate	Financials
CRM	Ceramics	Building Materials & Fixtures	Constr. & Materials	Industrials
CSB	Cashbuild	Home Improvement Retailers	General Retailers	Consumer Services
CSL	Consol	Containers and Packaging	General Industrials	Industrials
DAW	Dawn	Building Materials & Fixtures	Constr. & Materials	Industrials
DDT	Didata	Computer Services	Software & Comp.Serv.	Technology
DEL	Delta	Electrical Components & Equipment	Electrnc& Elect.Equip.	Industrials
DRD	DRDGold	Gold Mining	Mining	Basic Materials
DSY	Discovery	Life Insurance	Life Insurance	Financials
DTC	Datatec	Computer Services	Software & Comp.Serv.	Technology
ECO	Edcon	Apparrel Retailers	General Retailers	Consumer Services
ELH	Ellerine	Home Improvement Retailers	General Retailers	Consumer Services
EMI	Emira	Real Estate Investment Trust	Real Estate	Financials
FBR	FamBrands	Restaurants & Bars	Travel & Leisure	Consumer Services
FOS	Foschini	Apparrel Retailers	General Retailers	Consumer Services
FRO	Frontrange	Software	Software & Comp.Serv.	Technology
FSP	Freestone	Real Estate Holding & Development	Real Estate	Financials
FSR	Firstrand	Banks	Banks	Financials
GDF	Goldreef	Gambling	Travel & Leisure	Consumer Services
GFI	Gold Fields	Gold Mining	Mining	Basic Materials
GND	Grindrod	Marine Transportation	Industrial Transport.	Industrials
GRF	Group 5	Heavy construction	Constr. & Materials	Industrials
GRT	Growthpoint	Real Estate Holding & Development	Real Estate	Financials
GRY	Grayprop	Real Estate Investment Trust	Real Estate	Financials
HAR	Harmony	Gold Mining	Mining	Basic Materials
HDC	Hudaco	Industrial Machinery	Industrial Engineering	Industrials
HVL	Hiveld	Steel	Industrial Metals	Basic Materials
HYP	Hyprop	Real Estate Holding & Development	Real Estate	Financials
IFR	I-Four	Real Estate Holding & Development	Real Estate	Financials
MLA	Mittal	Steel	Industrial Metals	Basic Materials
ILA	Iliad	Industrial Suppliers	Support Services	Industrials
ILV	Illovo	Food Products	Food Producers	Consumer Goods
IMP	Implats	Platinum and Precious Metals	Mining	Basic Materials
INL	Invest Ltd	Investment Services General	Financial	Financials
INP	Invest plc	Investment Services General	Financial	Financials
IPL	Imperial	Transportation Services	Industrial Transport.	Industrials
IVT	Invicta	Industrial Machinery	Industrial Engineering	Industrials

JCM	JohnComm	Publishing	Media	Consumer Services
JDG	Jd Group	Home Improvement Retailers	General Retailers	Consumer Services
KAP	Kap	Diversified Industrials	General Industrials	Industrials
KGM	Kgmedia	Broadcasting & Entertainment	Media	Consumer Services
KMB	Kumba	General Mining	Mining	Basic Materials
LBT	Lib-Int	Real Estate Holding & Development	Real Estate	Financials
LEW	Lewis	Home Improvement Retailers	General Retailers	Consumer Services
LGL	LibertyLife	Insurance	Life Insurance	Financials
MDC	MediClinic	Health care Providers	HealthCare Equip.&Serv.	Health Care
MET	Metlife	Life Insurance	Life Insurance	Financials
MKL	Makalani	Equity Investment Instruments	Equity Investmnt Instr.	Financials
MPC	Mr Price	Apparrel Retailers	General Retailers	Consumer Services
MPL	Metprol	Real Estate Holding & Development	Real Estate	Financials
MRF	Merafe	General Mining	Mining	Basic Materials
MSM	Massmart	Broadline Retailers	General Retailers	Consumer Services
MST	Mustek	Computer Hardware	Techn. & Equip.	Technology
MTA	Metair	Auto Parts	Automobiles & Parts	Consumer Goods
MTN	MTN	Mobile Telecommunications	Mobile Telecomms.	Consumer Services
MTP	Marprop	Real Estate Holding & Development	Real Estate	Financials
MTX	Metorex	General Mining	Mining	Basic Materials
MUR	M&R	Heavy construction	Constr. & Materials	Industrials
MVG	Mvela	Business Support Services	Support Services	Industrials
MVL	Mvela	General Mining	Mining	Basic Materials
NCL	Nuclicks	Broadline Retailers	General Retailers	Consumer Services
NED	Nedbank	Banks	Banks	Financials
NHM	Northams	Platinum and Precious Metals	Mining	Basic Materials
NPK	Nampak	Containers and Packaging	General Industrials	Industrials
NPN	Naspers-N	Broadcasting & Entertainment	Media	Consumer Services
NTC	Netcare	Health care Providers	Health Care Equipment & Services	Health Care
OCE	Oceana	Farming & Fishing	Food Producers	Consumer Goods
OML	Old Mutual	Life Insurance	Life Insurance	Financials
OMN	Omnia	Speciality Chemicals	Chemicals	Basic Materials
PAM	Palamin	Non-ferrous metals	Industrial Metals	Basic Materials
PAP	Panprop	Real Estate Holding & Development	Real Estate	Financials
PGR	Peregrine	Investment Services	General Financial	Financials
PIK	Piknpay	Food Retailers & Wholesalers	Food & Drug Retailers	Consumer Services
PMA	Primedia	Media Agencies	Media	Consumer Services
PMM	Premium	Real Estate Holding & Development	Real Estate	Financials
PMN	Primedia-N	Media Agencies	Media	Consumer Services
PPC	PPC	Building Materials & Fixtures	Construction & Materials	Industrials
PRM	Prima	Real Estate Holding & Development	Real Estate	Financials
PSG	PSG	Investment Services	General Financial	Financials
PTG	Peermont	Gambling	Travel & Leisure	Consumer Services
RAH	Rahold	Equity Investment Instruments	Equity Investmnt Instr.	Financials
RBW	Rainbow	Farming & Fishing	Food Producers	Consumer Goods
RCH	Richemont	Clothing & Accessories	Personal Goods	Consumer Goods
RDF	Redefine	Real Estate Holding & Development	Real Estate	Financials
REM	Remgro	Specialty Finance	General Financial	Financials
RES	Resilient	Real Estate Holding & Development	Real Estate	Financials
RLO	Reunert	Electrical Componenets & Equipment	Electronic & Electrical Equipment	Industrials

RMH	RMB Hold	Banks	Banks	Financials
SAB	SAB Miller	Brewers	Beverages	Consumer Goods
SAP	Sappi	Paper	Forestry & Paper	Basic Materials
SBK	Stanbank	Banks	Banks	Financials
SCN	Scharrig	General Mining	Mining	Basic Materials
SHF	Steinhoff	Furnishings	Household Goods	Consumer Goods
SHP	Shoprite	Food Retailers & Wholesalers	Food & Drug Retailers	Consumer Services
SLM	Sanlam	Life Insurance	Life Insurance	Financials
SNT	Santam	Property & Casualty Insurance	Nonlife Insurance	Financials
SOL	Sasol	Intergrated Oil and Gas	Oil & Gas Producers	Oil & Gas
SPE	Spearhead	Real Estate Holding & Development	Real Estate	Financials
SPG	SuperGroup	Trucking	Industrial Transport.	Industrials
SPP	Spar	Food Retailers & Wholesalers	Food & Drug Retailers	Consumer Services
SRL	Sa Retail	Real Estate Holding & Development	Real Estate	Financials
SUI	SunInt	Gambling	Travel & Leisure	Consumer Services
SUR	Spurcorp	Restaurants & Bars	Travel & Leisure	Consumer Services
SYC	Sycom	Real Estate Holding & Development	Real Estate	Financials
TBS	Tigbrands	Food Products	Food Producers	Consumer Goods
TIW	Tiwheel	Auto Parts	Automobiles & Parts	Consumer Goods
TKG	Telkom	Fixed Line Telecommunications	Fixed Line Telecomms.	Consumer Services
TNT	Tongaat	Food Products	Food Producers	Consumer Goods
TRE	Trencor	Transportation Services	Industrial Transport.	Industrials
TRT	Tourvst	Travel & Tourism	Travel & Leisure	Consumer Services
TRU	Truwths	Apparrel Retailers	General Retailers	Consumer Services
TSX	Trans Hex	Diamonds and Gemstones	Mining	Basic Materials
UTR	Unitran	Speciality Retailers	General Retailers	Consumer Services
VKE	Vukile	Real Estate Holding & Development	Real Estate	Financials
VNF	Venfin	Specialty Finance	General Financial	Financials
WAR	W Areas	Gold Mining	Mining	Basic Materials
WBO	Wlsn Bailey	Heavy construction	Constr. & Materials	Industrials
WES	Wesco	Auto Parts	Automobiles & Parts	Consumer Goods
WHL	Woolies	Broadline Retailers	General Retailers	Consumer Services

Bibliography

[1] Bouchaud J.P., (2002), *An introduction to statistical finance*, Physica A, Vol. 313, p. 238-251.

[2] Mehta M., (1995), *Random Matrices*, Academic Press: New York.

[3] Wilcox D., and Gebbie T., (2007), *An analysis of cross-correlations in an emerging market*, Physica A, Vol. 375, Issue 2, p. 584-598.

[4] Mandelbrot B.B., (1963), *The Variation of Certain Speculative Prices*, J. Business, Vol. 36, p. 394-419.

[5] Challet D., Marsili M., and Zhang Y-C, (2000), *Modeling Market Mechanism with Minority Game*, Physica A, p. 276-286.

[6] Giada L., (2001), Phd Thesis: *From microscopic noise to macroscopic structures: A direct and an inverse problem*, International School for Advanced Studies, Trieste, Italy.

[7] Giada L., Marsili M., (2001), *Data clustering and noise undressing of correlation matrices* , Phys. Rev. E 63.

[8] Marsili M., (2002), *Dissecting financial markets: Sectors and States*, Quantitative Finance, Vol. 2, p. 297.

[9] Noh J.D., (2000), Phys. Rev. E, Vol. 61, p. 5981.

[10] Markowitz, (1959), *Portfolio Selection: Efficient Diversification of Investments*, J. Wiley and Sons.

[11] Samuelson P., (1969), *Lifetime portfolio selection by dynamic programming*, Rev.Econ. Stat. 51, p. 239-246.

[12] Coelho R., Hutzler S., Repetowicz P., and Richmond R., (2006), *Sector Analysis for a FTSE Portfolio of stocks*, Physica A, Vol. 373, p. 615.

[13] Aldenderfer M.S., and Blashfield R.K., (1984), *ClusterAnalysis*, Quantitative Applications in the Social Sciences, Vol. 44.

[14] Jain A.K., and Dubes R.C., (1988), *Algorithms of Data Clustering*, Prentice Hall, Englewood Cliffs, .

[15] Gonzalez T.F., (1985), *Clustering to minimize the maximum intercluster distance.*, Theoretical Computer Science, p. 38:293-306.

[16] Ankerst M., Breuning M. M., and Kriegel H-P, (1999), *OPTICS: Ordering Points to identify the Clustering Structure*, Proceedings ACM SIGMOD, p. 49-60.

[17] Ertoz L., Steinbach M., and Kumar V., (2002), *A new shared nearest neighbor clustering algorithm and its application*, AHPCRC, Tech. Rep., Vol. 134.

[18] Ben-Dor A., and Yakhini Z., (1999), *Clustering gene expression patterns*, Proceedings of the 3rd Annual International Conference on Computational Molecular Biology (RECOMB 99).

[19] Matula D. W., (1977), *Classification and Clustering, chapter Graph Theoretic Techniques for Cluster Analysis Algorithms*. Academic Press.

[20] Sibson R., (1973), *SLINK: An optimally efficient algorithm for complete link method*. The Computer Journal 16, p. 30-34.

[21] Defays D., (1977), *An effeicient Algorithm for a complete link method*. The Computer Journal 20, p. 364-366.

[22] Voorhees E. M., (1986), *Implementing agglomerative hierarchical clustering for use in document retrieval*. Information Processing Management, 22(6), p. 465-476.

[23] Dellaert F., (2002), *The Expectation Maximization Algorithm*, College of Computing, Georgia Institute of Technology Technical Report number GIT-GVU-02-20.

[24] Blatt M., Wiseman S., Domany E., (1995), *Superparamagnetic Clustering of Data*, Department of Physics of Complex Systems, The Weizmann Institute of Science, Israel.

[25] Vesanto J., Alhoniemi E., (2000), *Clustering of the Self-Organizing Map*, IEEE Transaction on Neural Networks, Vol. 11 (3).

[26] Drinea P., Frieze A., Kannan R., Vempala S., Vinay V., (2004), *Clustering Large Graphs via the Singular Value Decomposition*, Machine Learning 56, p. 933.

[27] Giada L., Marsili M., (2005), *Algorithms of maximum likelihood data clustering with applications*, Phys. Rev. E, Vol. 63.

[28] Papadimitriou C.H., Steiglitz K., (1998), *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall.

[29] Aarts E., Korst J., (1988), *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*, Weily - Interscience Series in Discrete Mathematics and Optimization.

[30] Potts R.B., (1952), *Some generalized orderdisorder transformations*, Proceedings of the Cambridge Philosophical Society 48, p. 106109.

[31] Grimmett G., (1994), *The Random-Cluster Model*, Wiley, p. 49-63.

[32] Kaizoji T., (2006), *An interacting-agent model of financial markets from the viewpoint of nonextensive statistical mechanics*, Physica A.

[33] Toda M., Kubo R., Sait$\widehat{o}$ N., (1983), *Statistical Physics*, Springer-Verlag: Berlin.

[34] da Silva R., Baraviera A., Dahmen S.R., (2005), *Modelling Investment in Artificial Stock Markets*, Working Papers of GRES 2004-04.

[35] Rosenow B., Gopikrishnan P., Plerou V., Stanley H.E., (2002), *Random magnets and correlations of stcok price fluctuations*, Physica A, Vol. 314, Issue 1-4, p. 762-767.

[36] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., (1953), *Equation of state calculations by fast computing machines*, J. Chem. Phys., Vol. 21, p. 10871092.

[37] Kirkpatrick S., Gelatt C.D., and Vecchi M.P., (1983), *Optimization by simulated annealing*, Science Magazine no. 4598, Vol. 220, p. 671680.

[38] Gopikrishnan P., Rosenow B., Plerou V., and Stanley H.E., (2001), *Quantifying and interpreting collective behavior in financial markets*, Phys. Rev. E, Vol. 64.

[39] Tuncay C., Stauffer D., (2006), *Power Laws and Gaussians for Stock Market Fluctuations*, Physica A: Statistical Mechanics and its Applications, Vol. 374, Issue 1, p. 325-330 .

[40] Kaizoji T., (2006), *Power laws and market crashes*, Progress of Theoretical Physics Supplement, Vol. 162, p. 165.

[41] Berg J., Marsili M., Rustichini A., Zecchina R., (2001), *Statistical mechanics of asset markets with private information*, Quantitative Finance 1, p. 203-211.

[42] Challet D., Chessa A., Marsili M., Zhang Y-C., (2001), *From Minority Games to real markets*, Quantitative Finance, Vol. 1, p. 168-176(9).

[43] Challet D., Marsili M., (1999), *Phase transition and symmetry breaking in the minority game*, Phys. Rev. E, Vol. 60., p. 6271-6274.

[44] Challet D., Marsili M., Zecchina R., (2000), *Statistical mechanics of systems with heterogeneous agents: Minority games*, Phys. Rev. Lett., Vol. 84, p. 1824 - 1827.